Experiential Perspectives on Sound and Music for Virtual Reality Technologies

Stephen Tatlow

March 2023

Department of Music & Department of Media Arts

Royal Holloway, University of London

Declaration of Authorship

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Acknowledgements

Special thanks are given to Tim Summers, who knew what he was signing up for, but agreed to be my primary supervisor anyway.

I am also thankful for the assistance of Rhys Davies and Julie Brown. Your comments and assistance have been invaluable throughout this project. Thank you also to David Howard, who stepped in on several occasions to assist. Hopefully you have all enjoyed terrorising standardised administrative forms as much as I have.

Thanks to my colleagues and peers at Royal Holloway for their willingness to engage with all aspects of my research, my friends and family for their tolerance and understanding of my negotiation with the abyss, and my perpetually absent students (without whom I managed to write significant portions of my thesis whilst nominally 'on-the-clock').

Research presented within this thesis was supported in part by funding awards from the Department of Music, the School of Performing and Digital Arts, and the Doctoral School at Royal Holloway.

Significant portions of Chapter 3 of this thesis are forthcoming as a chapter in *History as Fantasy* (ed. James Cook, Alexander Kolassa, Alex Robinson and Adam Whittaker),Abingdon, Routledge.

Abstract

This thesis examines the intersection of sound, music, and virtuality within current and next-generation virtual reality technologies, with a specific focus on exploring the experiential perspectives of users and participants within virtual experiences.

The first half of the thesis constructs a new theoretical model for examining intersections of sound and virtual experience. In Chapter 1, a new framework for virtual experience is constructed consisting of three key elements: virtual hardware (e.g., displays, speakers); virtual software (e.g., rules and systems of interaction); and virtual externalities (i.e., physical spaces used for engaging in virtual experiences). Through using and applying this new model, methodical examinations of complex virtual experiences are possible. Chapter 2 examines the second axis of the thesis through constructing an understanding of how sound is designed, implemented, and received within virtual reality. The concept of soundscapes is explored in the context of experiential perspectives, serving as a useful approach for describing received auditory phenomena. Auditory environments are proposed as a new model for exploring how auditory phenomena can be broadcast to audiences. Chapter 3 explores how inauthenticity within sound can impact users in virtual experience and uses authenticity to critically examine challenges surrounding sound in virtual reality. Constructions of authenticity in music performance are used to illustrate how authenticity is constructed within virtual experience. Chapter 4 integrates music into the understanding of auditory phenomena constructed throughout the thesis: music is rarely part of the created world in a virtual experience. Rather, it is typically something which only the audience – as external observers of the created world – can hear. Therefore, music within immersive virtual reality may be challenging as the audience is placed within the created world.

The second half of this thesis uses this theoretical model to consider contemporary and future approaches to virtual experiences. Chapter 5 constructs a series of case studies to demonstrate the use of the framework as a trans-medial and intra/inter-contextual tool of analysis. Through use of the framework, varying approaches to implementation of sound and music in virtual reality technologies are considered, which reveals trans-medial commonalities of immersion and engagement with virtual

experiences through sound. Chapter 6 examines near-future technologies, including brain-computer interfaces and other full-immersion technologies, to identify key issues in the design and implementation of future virtual experiences and suggest how interdisciplinary collaboration may help to develop solutions to these issues. Chapter 7 considers how the proposed model for virtuality might allow for methodical examination of similar issues within other fields, such as acoustics and architecture, and examines the ethical considerations that may become relevant as virtual technology develops within the 21st Century.

This research explores and rationalises theoretical models of virtuality and sound. This permits designers and developers to improve the implementation of sound and music in virtual experiences for the purpose of improving user outcomes.

Contents

Tables and Figures

Tables

Figures

Chapter 0: Introductions

0.0.0: Introductions

This research project investigates the design, implementation and analysis of sound and music within virtual experiences. Virtuality – and virtual experiences – has been the cause of a furore of activity since the late 1980s and early 1990s. This has continued through to the present day. Many existing projects have focused on the way in which virtual reality technologies are expected to change and mutate society. For example, Steve Woolgar's introduction to an edited collection reviewing 'virtual society' privileges the role of technology in virtuality:

> We are continuing to witness the burgeoning growth of new electronic information and communications technologies…These new technologies have been widely regarded as the impetus for radical changes…All aspects of the social, cultural, economic, and political life thus stand to be affected by the continued massive growth in electronic technologies…Discussion of the social consequences of new electronic technology is organized around one contentious summary claim about the likely effects of new electronic technologies: the idea that the transformative effects of these technologies amount to our transition to a virtual society. (Woolgar, 2002, pp. 1–2)

Similarly, Steve Bryson – like many others – begins their discussion of the design and implementation of virtual reality applications by establishing a purely technological framework:

> The design process for virtual reality applications has two driving requirements: the virtual environment and its interface should be tailored to the task; stringent performance constraints must be met for the benefit of virtual reality to be realised. The first requirement is in recognition of the fact that immersing the user in a three-dimensional computer generated environment presents many opportunities not easily found in conventional "desktop" three-dimensional graphics… The second requirement refers to the fact that the virtual environment must run with a certain minimal speed in order to be usable. These two requirements are often in conflict. (Bryson, 1995, pp. 4–6)

Even in more modern work, this same focus on a specifically (even exclusively) technological understanding of virtuality can be observed. Kay Stanney et al. open their introduction to the *Handbook of Virtual Environments* by quoting Descartes:

> As Descartes suggested, there is no definitive truth; reality emanates from that which is present to our senses, and these senses we trust to distinguish reality from illusion can be deceived. It is this capacity to fool the sensory systems that we attempt to capitalize on when building a virtual world… The goal of this experience is to *deceive* us and represent a *truth*

> that can educate, train, entertain, and inspire. In their ultimate form, virtual environments (VEs) immerse users in an alternate *reality* that stimulates multiple senses, providing vibrant experiences that are so veridical they fundamentally transform those exposed. (Stanney et al., 2015, pp. 3–4, emphasis original)

Given Descartes died in 1650, it is unlikely that he is speaking of the kind of virtual environments – that is to say, the kind of virtual reality[1] – that Stanney et al. almost immediately describe as being specifically technological (and, to some extent, inherently computer-based):

> Technology has been invested in, yet widespread adoption of VE technology has proven elusive… The technology used to generate VE systems has rapidly matured over the past decade, driven primarily by the gaming industry… beyond processing capability, technology advances have come in many areas, including human-machine interfaces, the hardware and software used to generate the VE, electromechanical systems used in telerobotics, as well as the communication networks that can be used to transform VE systems into shared virtual worlds. (Stanney et al., 2015, p. 4)

These discussions represent a broader – perhaps even near-ubiquitous – approach to considerations of virtual reality. Throughout many virtual reality projects of the last forty years, the systems of technology used to generate virtual experiences has been the central area of focus of scholars and practitioners; the concept of a ‘virtual experience’ has been rooted almost entirely within some technological framework.

However, the approaches represented by Woolgar, Bryson, Stanney et al., and many others are inherently limited as they focus on the way in which technology succeeds or fails at a becoming a ‘virtual experience’ rather than the construction of the experience itself. Whilst technology must remain a component of any consideration of virtuality, this project takes a broader approach to the concept of the ‘virtual’ which examines it as a mediation of human experience, rather than a technological mediation of the real; this project attempts to move past questions of how technology might change or alter our perspectives of reality and take a user-focused perspective of experiences of the virtual instead.

[1] As shall be discussed throughout this opening chapter, there are many terms for virtual reality and virtuality more broadly. Stanney et al. use “virtual environment” synonymously with “virtual reality”.

This experiential perspective has been at the periphery of research for some time – e.g., David Travis et al. argue in 1994 that “there is a wide agreement that in order to design effective systems, the systems must be user-centred” (Travis et al., 1994, p. 44) – but much of the existing scholarship takes a systems-eye, technologically deterministic point of view that too readily assumes an idealised outcome and ignores factors anchored in the user, rather than the technology of the system. This research project takes the perspective that whilst virtual experiences can be formed in collaboration with technology, they are not a specifically technological phenomena, but rather a phenomenon of constructed or mediatised human perspective which is aided by or mediatised by a variety of means including, but not limited to, technological approaches.

The research in this thesis presupposes that processes of reception have an inevitable effect on the user’s experience, especially when examining the paradigms of sound, music, and other auditory phenomena. Reframing considerations of sound within virtual experience through this specific lens of experientiality offers pragmatic benefits for those designing sound within virtual experience: as will be demonstrated throughout this thesis, a focus on the user experience of virtuality can provide meaningful answers to difficult questions by highlighting trans-medial, inter-contextual, and inter-disciplinary similarities between issues at the centre of virtual experience. This approach is valuable as many issues across the field(s) of virtual experience have similar experiential framings despite different technological approaches and/or different design intentions. By taking a holistic approach, we can have greater confidence in the applicability and usability of the research presented in this thesis and in the conclusions drawn through applying this research.

There are two overarching questions of interest to be explored. First, how does sound and music function within virtual experience? Second, how might we holistically approach the design, implementation, and analysis of sound and music within virtual experience? In order to do this, we need to both understand what is meant by the conceptual term ‘virtual experience’ and understand the purposeful use of sound within virtual experience functions. These will be established throughout the first section of this thesis. In Chapter 1, a new framework for virtual experience is constructed consisting of three key elements: virtual hardware (e.g., controllers, displays, speakers); virtual software (e.g., rules and systems of interaction); and virtual externalities (i.e., physical spaces used for engaging in virtual experiences). Chapter 2 examines the second axis of the thesis through constructing an understanding of how sound is designed, implemented, and received within virtual reality. R Murray Schafer's concept of soundscapes (Schafer, 1977b) is adapted for experiential perspectives as a useful way to describe received auditory phenomena for use in virtual experiences. Auditory environments are proposed as a new model for describing how auditory phenomena can be broadcast to audiences. Authenticity within sound is interrogated in Chapter 3 as a potentially problematic element of immersion within virtual experiences. Chapter 4 integrates music into the understanding of auditory phenomena constructed throughout the thesis: music is rarely part of the created world in a virtual experience. Rather, it is typically something which only the audience – as external observers of the created world – can hear. Therefore, music within immersive virtual reality may disrupt immersion at times as the audience is placed within the created world. Music within cinema is considered alongside music for interactive media to develop an understanding of why and how music can function in virtual reality, e.g., how might music effect immersion in virtual reality experiences?

The second half of this thesis uses this theoretical model to consider contemporary and future approaches to virtual experiences to demonstrate the applications and usefulness of the model. Chapter 5 codifies and summarises the framework, and then demonstrates the use of the framework as a trans-medial and intra/inter-contextual tool of analysis through a series of case studies. Varying approaches to the implementation of sound and music in virtual reality technologies are considered through the framework developed in the first half of this thesis, which reveals trans-medial commonalities of immersion and engagement with virtual experiences through sound. Chapter 6 examines near-future technologies, including brain-computer interfaces and other full-immersion technologies, to identify key issues in the design and implementation of future virtual experiences and suggest how interdisciplinary collaboration may help to develop solutions to these issues. Finally, Chapter 7 reviews the research contained within the thesis and considers how the proposed model for virtuality might allow for methodical trans-medial examinations of issues within other fields, such as urban planning, or to examine the other considerations that may become relevant as virtual technology further integrates with society in the 21st Century, such as transhuman ethics.

As can be inferred from the outline of thesis presented above, one of the major issues facing anyone researching in the field of virtual experience is the lack of consistency and coherence within the terminology used to explore it. The research within this thesis draws on a broad body of literature across disparate fields ranging from sound design to neuroscience. Although this literature was predominantly written throughout the last hundred years, many linguistic shifts within these fields can be observed. This is partly a consequence of the extensive technological development, but it is also due to the varied backgrounds of those working in this field. The terminology used throughout the existing literature on this topic is therefore inconsistent.

Compounding these issues, the field of virtual experience is currently in the process of rapid expansion. This means that many of those discussing virtual reality in emerging literature may not necessarily aware of the full extent of the existing modalities of engagement. For example: an on-going obsession with head-mounted devices (HMDs), such as Oculus headsets, can be seen throughout discussions focused on virtual reality. Whilst HMDs are a key part of our understanding of the landscape of virtual reality devices, engaging with ‘virtual reality’ does not necessitate use of HMDs. As technology develops, terminology must also develop to describe the new expressions of virtual experience which become possible. In pragmatic terms, this thesis is written at a time of terminological change. The definitions of terms such as “mixed reality” and “augmented reality” are variable, and the relationship between these terms and those in general usage such as “virtual reality” is also unclear.

This initial chapter serves as an introduction to many of these topics and keywords. Common ground between differing bodies of literature will be identified and developed to establish how core conceptual issues such as ‘virtuality’ and ‘immersion’ can be rationalised. This chapter also sets out how terminology such as ‘virtual reality’ will be used throughout the thesis, and reviews current technologies for virtual experiences.

0.0.1: Virtuality

A cohesive concept of *virtuality* is central to a discussion of ‘virtual experience’ or ‘virtual reality’. Rob Shields suggests in his etymological exploration of the term that the modern understanding of the term ‘virtuality’ emerges from doctrinal debates between Catholic and Protestant churches in the 16th Century: Calvinists espoused the doctrine of “virtualism” – of Christ being a “virtual” presence in the bread and wine of the Eucharist – whilst the more traditional Catholicism focused around “transubstantiation” – of the bread and wine of the Eucharist transforming miraculously into “actual” body and blood (Shields, 2002, p. 4). Shields explores several other historical examples of virtuality (and “virtualism”) but suggests an inherent contradiction in the term “virtual reality” uncovered throughout each of these examples – if something is virtual, can it be real?

A more explicit conceptualisation of virtuality can be found in Jean Baudrillard’s *Simulacra and Simulation*. He draws upon a key issue at the centre of discussions of virtualisation: “if any symptom [of an illness] can be ‘produced,’ and can no longer be taken as a fact of nature, then every illness can be considered as simultable and simulated” (Baudrillard, 1981, p. 3–4). Baudrillard expands from this medical quandary to a broader concern: as humanity becomes more capable of simulation, we may become less capable of recognising reality. Baudrillard suggests that his contemporaries faced metaphysical despair over virtuality not because of fear that representations conceal aspects of reality, but rather out of fear that the representations are equally real. For Baudrillard, simulations of representations (or “images”) can be considered to come in five successive stages, each moving away from an implied “profound reality” (pp.6–7):

- i) The reflection of a profound reality
- ii) The masking and denaturing of a profound reality
- iii) The masking of the fact that there is no profound reality
- iv) No relation to any reality
- v) A pure simulacrum

Baudrillard's philosophical examination of virtuality is situated only in the context of early forms of what we now think of as 'virtual reality' technologies in the modern understanding. Expressions of virtuality within the modern world seem to nonetheless concur with Baudrillard's stages of simulation.

Virtual worlds such as those created within Google Earth are a direct reflection of reality. We can explore the real world through the virtual world. Shields describes how the on-going digitisation of architecture has allowed for digital walkthroughs to reduce the cost of expensive prototyping, and how digitised tourist destinations allow for digital tourists who travel internationally from the safety of their own home. (Shields, 2002, p. 67) Virtual explorations of physical places have only become more sophisticated and more exotic in the twenty years since Shields began his exploration of virtuality: tourists can even embark on an adventure along the only path driven on Mars after NASA turned the path of Curiosity – a Mars rover – into a new virtual experience similar to the ones offered by Google Earth (Figures 1–5).

But virtual worlds such as Google Earth also represent a broader approach to the virtualisation of reality. Those exploring an area in the modern era using the StreetView function on Google Maps can browse pictures taken during any of the multiple surveys of roads conducted by Google: in Europe, one of Google's first mapped locations, users now have access to data and photographs across more than a decade of surveys. What began simply as a tool to virtually and directly represent the reality has now been enhanced by the ability to instantaneously view the unique history of those locations. Google Maps is no longer an exploration of space, but of space *and time*. Shops can be seen flickering in and out of existence, pedestrians come and go with each instance, each captured moment reminding us of who and what was once there (Figures 6–7).

192


193 *Figure 1: Map view of 'Explore with Perseverance' (https://mars.nasa.gov/mars2020/surface-experience/)*

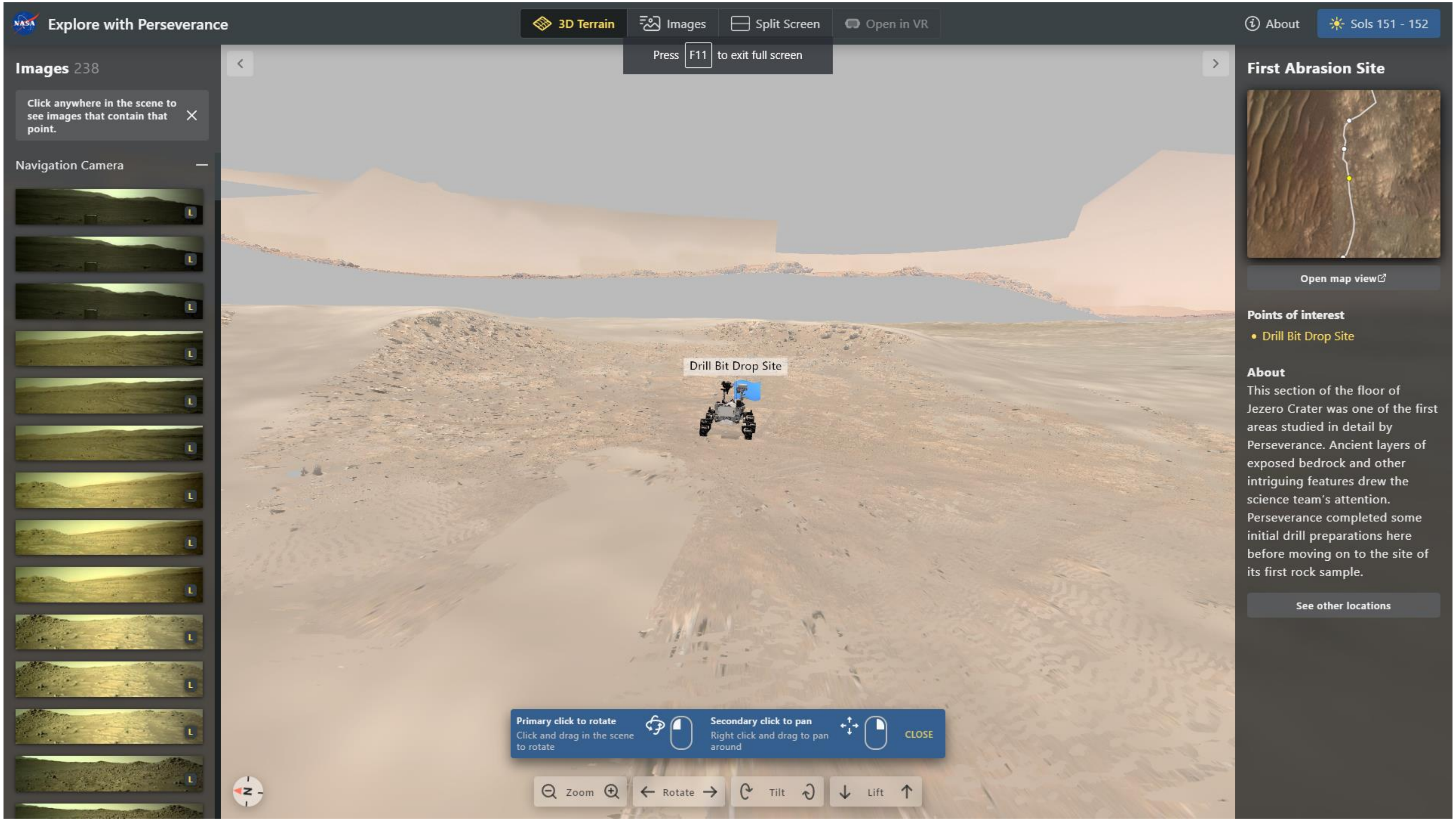

Figure 2: Explorer view of 'Explore with Perseverance' (https://mars.nasa.gov/mars2020/surface-experience/)

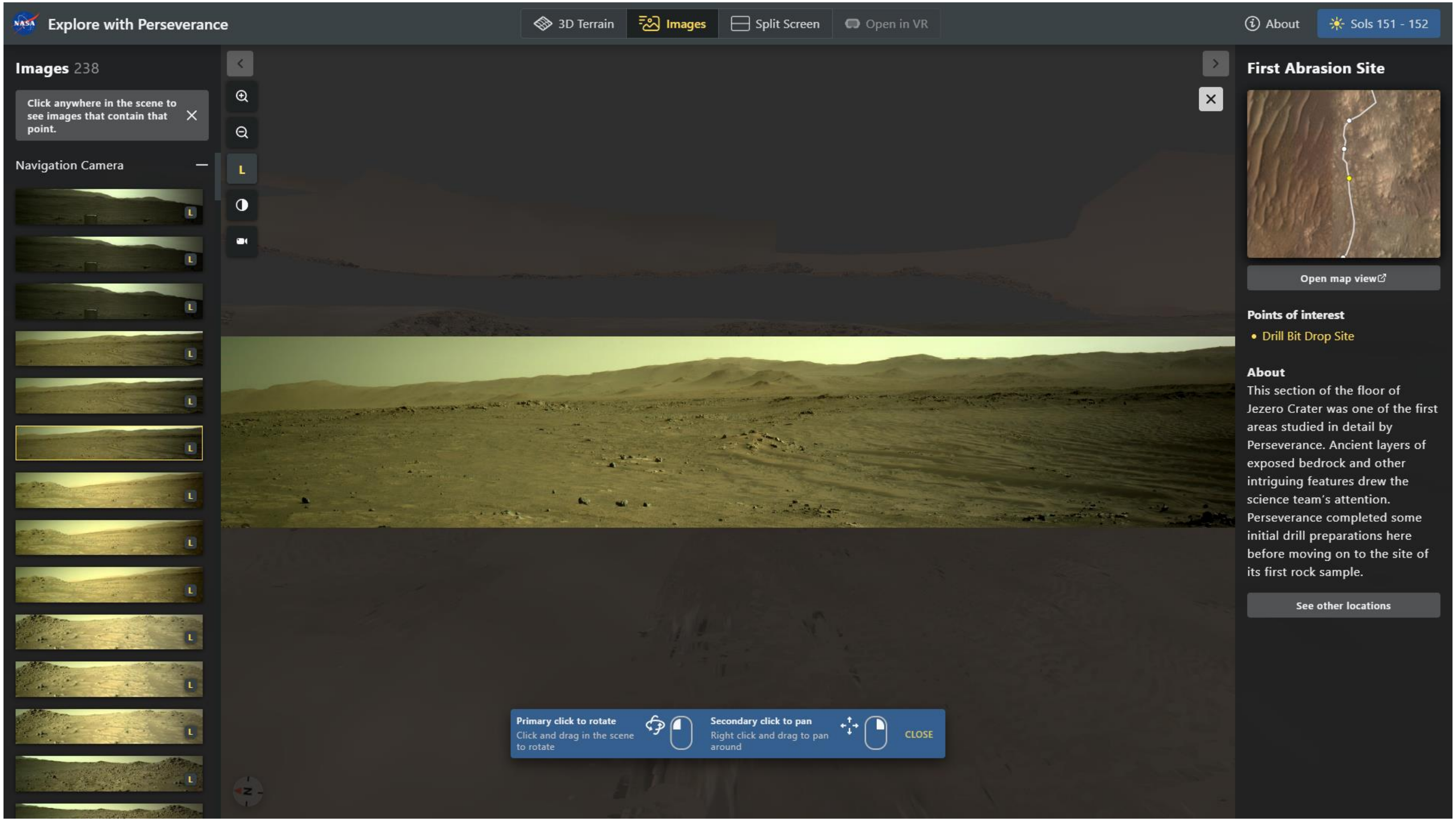

Figure 3: Camera view of 'Explore with Perseverance' (https://mars.nasa.gov/mars2020/surface-experience/)

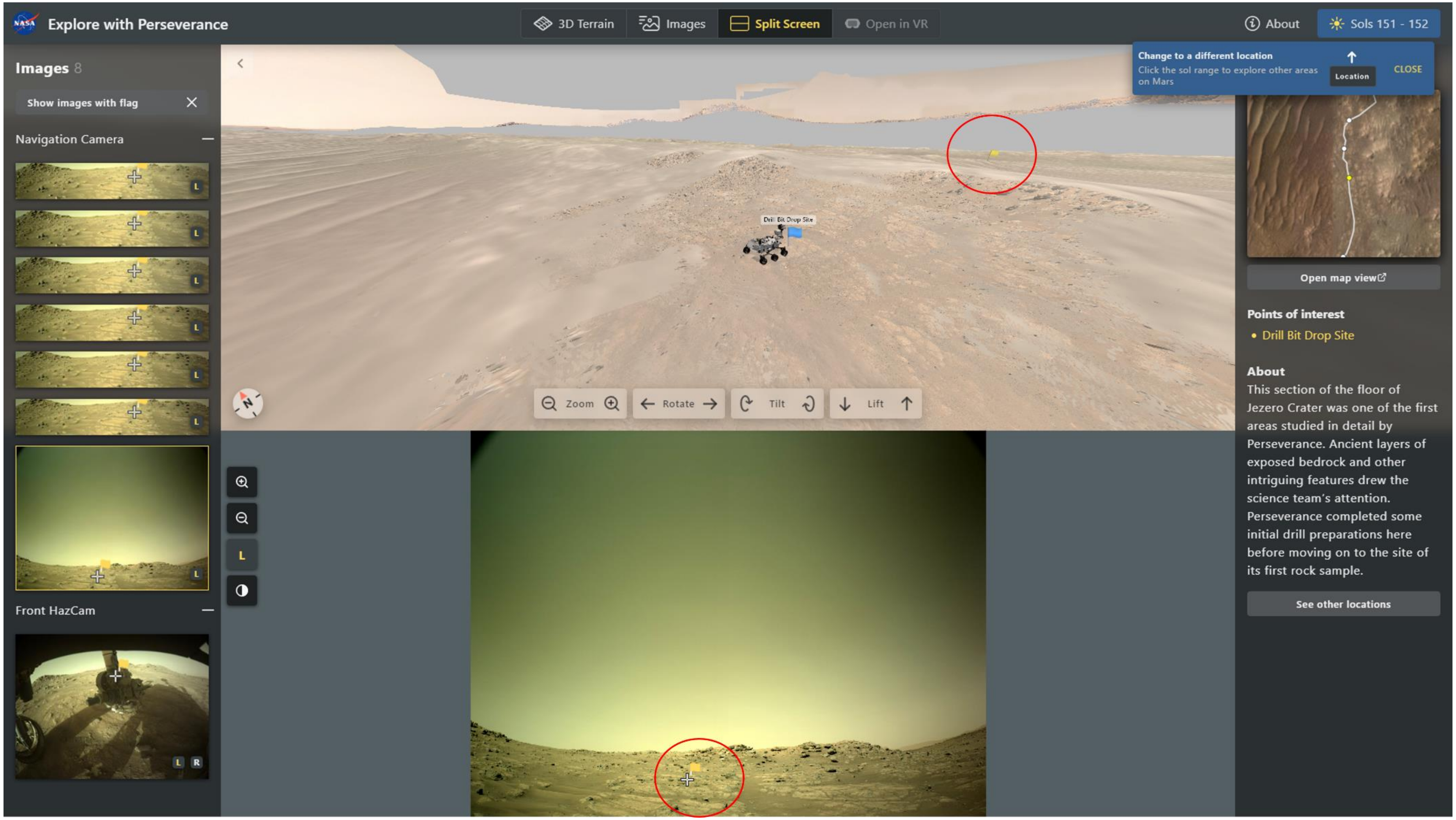

Figure 4: Hybrid view of 'Explore with Perseverance' (https://mars.nasa.gov/mars2020/surface-experience/). The yellow flag circled in both images is in the same location. Users can place these flags at any point on the Mars surface to easily locate any given point in the available images.

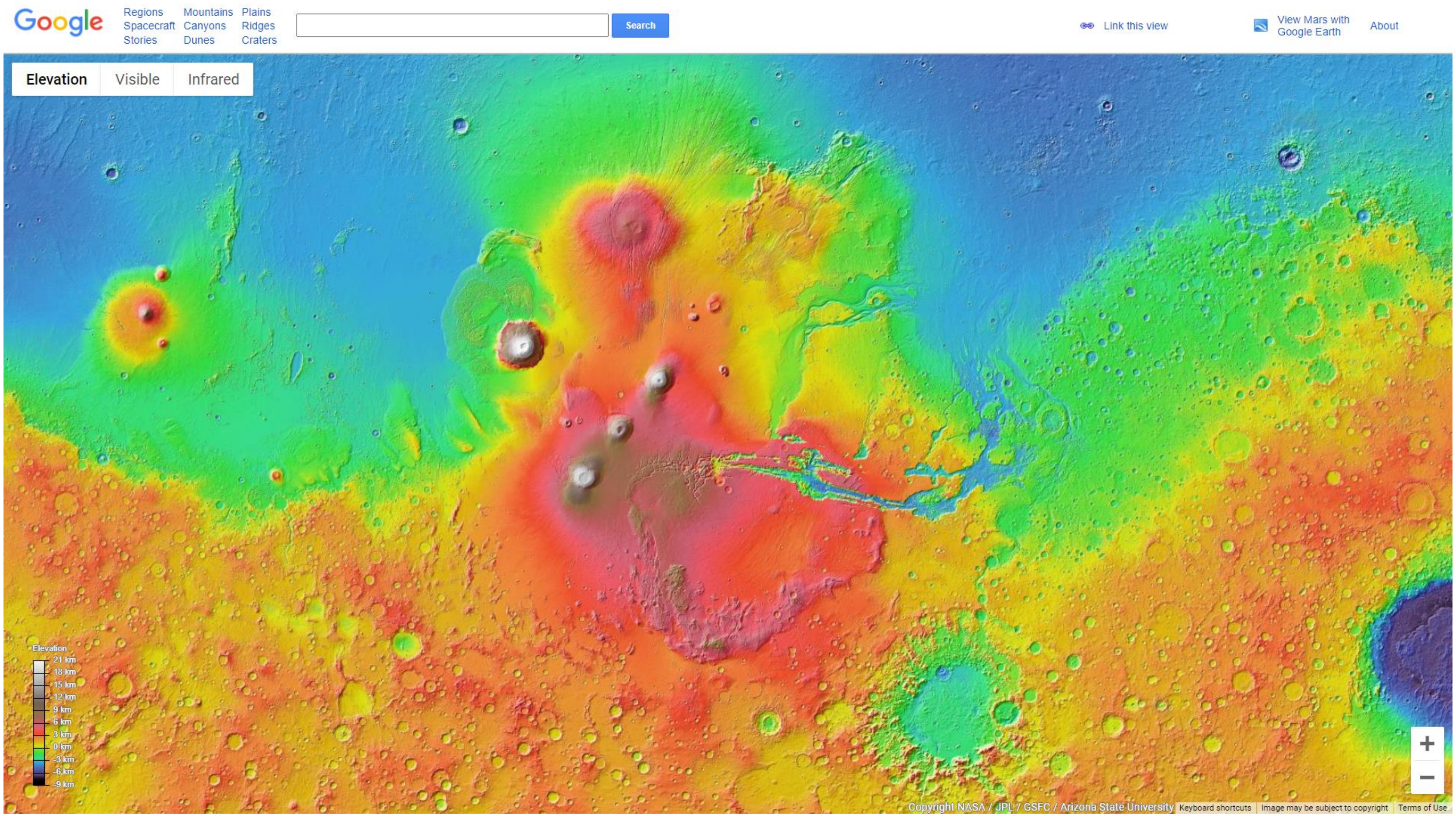

Figure 5: Google Mars (https://www.google.com/mars/) offers satellite images of Mars

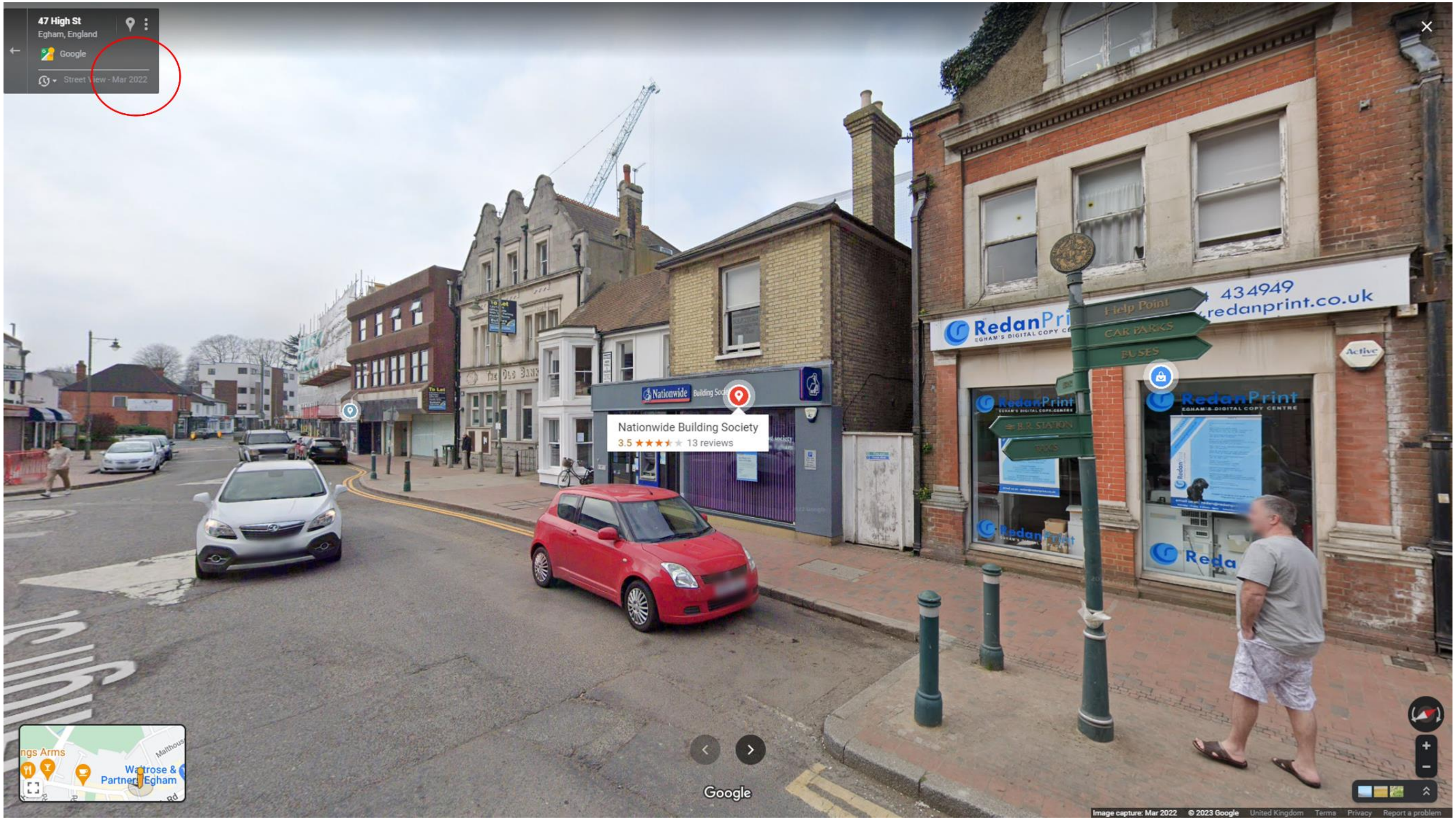

Figure 6: Picture of Egham High St from Google Street View (https://www.google.co.uk/maps) dated March 2022. The date selection mechanism is circled in the top left

Figure 7: Picture of Egham High St from Google Street View (https://www.google.co.uk/maps) dated August 2017. The date selection mechanism is circled in the top left

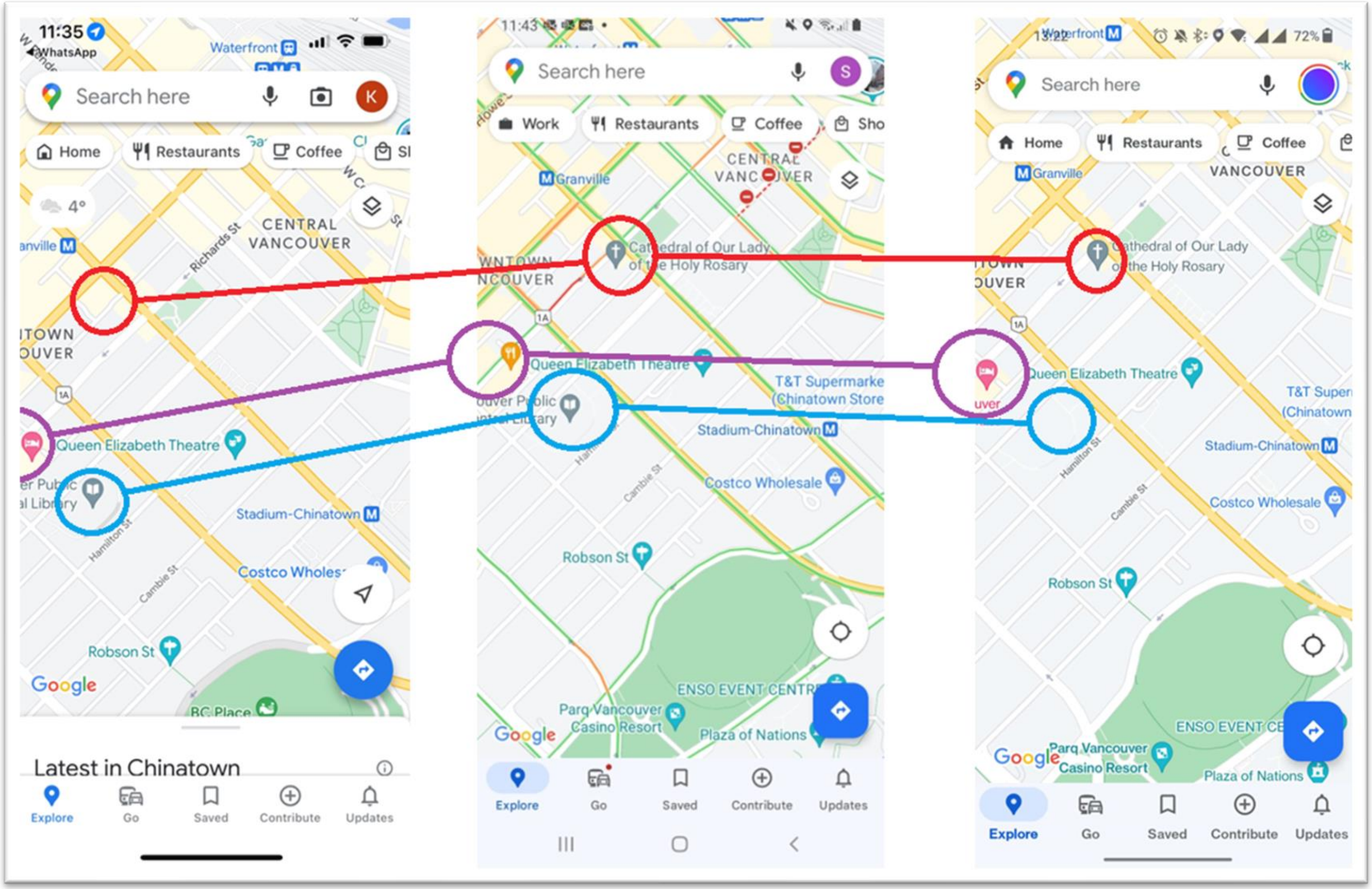

Figure 8: A comparison of three different views of the same area of downtown Vancouver (BC) from three different individuals demonstrating some of the differences which can emerge. Try it yourself by using this link to get your own map of downtown Vancouver: https://www.google.co.uk/maps/@49.2798395,-123.1131499,15.96z

We might further imagine the effect that user input has had on these virtualisations. When do we approach Baudrillard’s second stage of virtuality: the masking and denaturing of reality? Alongside photographs of physical locations, we can access information beyond representations of the physical: a subjective view specific to our own personal experience. The algorithms used to specify what buildings and locations appear as names as you zoom in and out are based on Google’s perception of what you would find interesting. Every search you make – even every step you take – is a datapoint sent to the artificial intelligence behind the user experience of Google Maps. Users are presented with a virtualised view of the world around them, designed specifically to ensure that their specific experience of the world is tailored to their satisfaction. Reality is masked and denatured: parts of distinct and separate from the real world: their own virtual realisation of the world around us (Figure 8).

As we spiral further into Baudrillard’s simulacra, profundity becomes lost to absurdity. Do locations need to exist to be placed within the virtual world? Oobah Butler became convinced that TripAdvisor presented a “false reality [of] meals [that] never took place” (Butler, 2017) and set out to prove it. As a result of his bewilderingly successful con, London’s best restaurant was – for some time – a virtual creation that had no underlying reality at all: the restaurant simply did not exist. The widescale manipulation of websites and algorithms and newspapers created, from nothing, a hugely successful restaurant. As Butler says, he had “created this reality.” And, as he observed, it had an undeniable effect. On their opening night (after several weeks of holding the number one spot on TripAdvisor), they put their dinner guests into a cold wintery garden and served them microwave-cooked ready-meals served on cheap plates. Despite the conditions, their dinner guests have had only one question: when can they come again? (Vice, 2018)

Butler seems to have successfully masked the lack of profound reality underneath his virtual creation. So successful is his virtual creation that those experiencing the depressingly awful profound reality are convinced that they are in one of London’s best restaurants. He is far from the only one to have embarked on such adventures. The perceptive effect of positive reviews and virtual chicanery on profound experiences is well-known, and there are many similar stories to be found – albeit none quite so similarly successful in execution. Are websites that serve to document an area and recreate their physicality within a digital, virtual world moving towards Baudrillard's third stage? Perhaps they are already there.

In these ways, reality seems to have become something which conflicts with virtuality: reality is in opposition to virtuality, reality is in opposition to simulation, reality is in opposition to representation. As Shields puts it: "the virtual [best contrasts] with the concretely present (which may also be called the real actual)" (Shields, 2002, p. 29). However, this raises a corollary question: what, and where, is Shields' "concretely present" or Baudrillard's "profound reality"? This is the difficulty in virtuality explored by Brian Massumi's paradox of the virtual: "[virtuality is] where what are normally opposites coexist, coalesce, and connect; where what cannot be experienced cannot but be felt – albeit reduced and contained" (Massumi, 2002, p. 30).

A dyadic conception of 'virtuality' and 'reality' is not a complete rationalisation of experience. Rather, both ends of the dichotomy exist simultaneously: reality and virtuality are not oppositional, but rather symbiotic. Virtuality is the experiential realisation of a theoretical profound reality; virtuality is the emergence of a specific event from the nebulous potential abstractions and representations that the 'real' presents us with. Within this symbiosis, the inherent conflict of a term such as 'virtual reality' – and other similar terms – is easily answered: 'reality' is *always* virtual, the symbiotic interdependence between them creating one as the consequence of the other, inevitably connected.

Within the last few decades, the spread of virtualisation of experience has been particularly noticeable due to the dissemination of digital technologies. Recent reports suggest 95% of people under 35 owned a smartphone in the UK in 2018 (Statista, 2020), providing constant instantaneous access to the internet and our constructed digital lives. Whilst digital does not always mean virtual, nor does virtuality require the presence of technology, it is not hard to find many examples of virtuality that we encounter on a regular basis that use different digital technologies. Our lives have become virtualised in almost every way through technology: we participate in virtual communities through virtual conversations, posting comments and links on social media such as Facebook and

Reddit; we share virtual documents in virtual workplaces, putting electronic files onto shared computer systems to collaborate with people we will never meet; we occupy our spare time with a menu of virtual entertainment, swiping our way through potential lovers with the same equanimity we apply to selecting food for delivery from a virtual shopfront.

For many growing up within this new ‘digital age’, virtuality becomes something which must be confronted as the lines separating virtuality and reality begin to blur. We face not only a *rationalisation* of virtuality as a reality, but the *recognition* of virtuality as a reality – perhaps even *determining* that virtuality *is* a reality, without adequate caution for the mutable subjectivity of perception. When studying virtual experience, this clearly poses problems. Philosophically, the modern interpretations of what it means to exist within a virtual world may have endless implications. Is the ability of the new generation to retrospectively alter their virtual identity connected to their desire for a ‘perfect’ political candidate with no ‘skeletons in their closet’? Is the construction of virtual communities of like-minded people driving a self-affirming tribal view on complex issues such as gender, race, and environmentalism? How is the virtualisation of ‘self’ affecting the conceptualisation of being?

This thesis does not try to answer these (staggeringly large) questions. Rather, these questions are intended to acknowledge that what it means to be virtual – or to have a virtual experience – is becoming increasingly important within modern society. To a large extent, the human experience is becoming the virtual experience. Pragmatically, a shift in how society rationalises virtuality will always affect how virtual experiences are perceived and engaged with. And, therefore, a shift in how virtuality is rationalised should – will – alter how we design, implement, and analyse these experiences.

For those attempting to design an experience which immerses the user, Baudrillard’s stages of simulation could almost be seen as a design guide to the liminal place between reality and virtuality: how might we convince someone that their experience of the ‘profoundly real’ should be sublimated in favour of a designed ‘virtually real’ experience? Often, it is the purpose of virtuality that the normal rules of reality are suspended: by suspending these rules, we can engage with new ideas, including those which could not exist within the actual-real in any pragmatic sense. As suggested by Katie Salen and Eric Zimmerman for video games, we can perhaps recognise that we enter a virtual construction when engaging with virtual experience: “the magic circle is where the game takes place. To play a game means entering into a magical circle, or perhaps creating one as a game begins. The magic circle of a game might have a physical component… but many games have no physical boundaries…the game simply begins when one or more players decide to play” (Salen & Zimmerman, 2004, p.95).

Virtuality, then, can be summarised as the mediatisation[2] of experience: symbiotically co-existing with reality to create distinct unique experiences. Within the context of deliberately constructed virtualities, such as those contained within HMD technologies such as the Oculus Quest, this can be a world constructed almost entirely out of virtualisations. Within other contexts, virtual worlds will emerge from the distinctly subjective nature of experience. In the general sense, virtuality emerges where aspects of the real-actual are supplanted, replaced, or augmented by things outside of, or beyond, the real-actual.

0.1.0: Types of Virtual Experience

The broadness of virtuality within the contemporary context poses problems of terminological framing. There are many ways of engaging with virtuality and/or engaging within a virtual experience, and designers typically self-select the terms which they use to describe each of their experiences. Rather like all other forms of ‘genre’ or ‘categorisation’, it is difficult to pin down exactly where one type of virtual experience ends and another begins.

Furnio Kishino and Paul Milgram propose a ‘virtuality continuum’ which frames virtuality and reality as extremes of a spectrum on which virtual experiences can be plotted (Kishino & Milgram, 1994). Within their research, they suggest a hierarchical approach to categorising virtual experiences: ‘augmented reality’ experiences build upon reality with additional virtual components, and ‘virtual reality’ experiences supplant and replace reality entirely within virtuality. Similar approaches have

[2] The term mediatisation is used here as defined in the Mirriam-Webster dictionary (“to put in a middle or intermediate position : make instrumental or subordinate”) rather than as used in political communication theory as a word to describe the means through which the media (e.g., TV news stations) shape and frame the process and discourse of society.

remained popular throughout the last twenty years, with many designers and academics drawing upon the key terms identified by Kishino & Milgram: the real environment (i.e.., "reality"), *augmented* reality/virtuality, *virtual* reality (or "virtuality"), and *mixed* reality. However, these terms are now found within common usage with more nuance than originally proposed by Kishino & Milgram. To rationalise the differing uses of these terms, the next section of this introductory chapter will now set out how key terms such as 'virtual reality,' 'augmented reality,' 'mixed reality,' and 'immersion' will be used throughout the thesis.

0.1.1: Augmented Reality (and Augmented Virtuality)

Kishino & Milgram propose that augmented reality consists of a real environment which has been augmented through the addition of virtual objects or items (Kishino & Milgram, 1994, p. 5). Similar views can be found contemporaneously, such as Jane Ellen Stevens explaining in 1995 that "in augmented reality, the person operating the system interacts with something in the real world" (J. E. Stevens, 1995, p. 435). Some slight variation in this can be seen in other literature. Ronald Azuma suggests in 1997 that augmented reality includes experiences which integrate three-dimensional virtual objects into a real environment in real-time (Azuma, 1997). Here, Azuma perhaps touches on what Kishino & Milgram originally proposes to be 'augmented virtuality' – the inverse of augmented reality: where a *virtual* environment is augmented through the addition of *real* objects or items. We might imagine a virtual experience which presents a virtual environment with a virtual table placed at the centre of the room. Kishino & Milgram would argue that placing a *real* table at the centre of the room would be creating an 'augmented virtuality.' Conversely, Azuma – and most work on the topic thereafter – seems to conflate Kishino's two terms of 'augmented virtuality' and 'augmented reality' into the singular 'augmented reality.' This thesis, similarly, adopts the position that augmented reality and augmented virtuality are functionally similar.[3]

[3] This conflation perhaps reflects a lack of concern for whether an experience is primarily virtual or primarily real. Whilst Kishino & Milgram value this distinction (possibly out of a belief that virtuality and reality may eventually become indistinguishable), the trend within more recent literature is to accept that there will probably always be elements of reality within virtuality. Virtuality is therefore seen primarily as building on or navigating around these realities. Some discussion of forthcoming developments in full-immersion environments can be found in section 6.3.0.

A broader definition of ‘augmented reality’ can be identified within more recent years. Donna Berryman suggests that augmented reality is created through “the process of overlaying computer-generated information on reality, whether that reality is a geographic place or an object.” (Berryman, 2012, p. 213). Similarly, Julie Carmigniana et al. define augmented reality as “a real-time direct or indirect view of a physical real-world environment that has been enhanced/augmented by adding virtual computer-generated information to it” (Carmigniani et al., 2011, p.342); reviews of modern augmented reality products include informational overlays (Nee et al., 2012) and integrated technology (Feiner, 2002); and so on. Rather than requiring a specific virtual object – such as a table or vase of flowers – be placed within a virtual experience, this broader definition of ‘information’ means that modern understandings of augmented reality include almost any experience which combines digital information with real environment.

This broader contemporary definition places ‘augmented reality’ in the hands of every single person with access to a smartphone. Shields speculates about the possibilities of creating a “virtual world” which we might inhabit (Shields, 2002, pp. 19–22) – it certainly seems true that we have now managed to create an “augmented world” which we do inhabit. Applications such as Google Maps provide users with a constant informational overlay for the world around them, enhancing their understanding of what is nearby. Many regional and national tourism boards are now developing augmented reality products for tourists visiting local areas (Jingen Liang & Elliot, 2021; N.-J. Shih et al., 2019; Wiltshier & Clarke, 2017) and/or specific applications designed for on-location use (Mata & Claramunt, 2014; Pence, 2010; Yang & Chan, 2019). Augmented reality has transitioned from referring to the specific combination of virtual objects and real objects into anything which combines virtual information with real objects.

The use of ‘augmented reality’ is perhaps problematic at times. Some suggest that augmentation suggests enhancement: to make *greater*, or to *improve*, a situation. This can be seen explicitly in the taxonomy of augmented reality set out by Oliver Hughes et al.:

> Although any increase in the quantity of information – and consequently, any increase in our understanding of reality – admitted by AR aims for greater mastery of what is real, it is clear

> that, from a technological point of view, AR can offer interfaces which propose either, more explicitly, information, or, more explicitly [sic], a better mastery of our actions with regard to real events. (O. Hughes et al., 2011, p. 48)

However, other contemporary uses of the term ‘augmented reality’ include products which focus on overlaying fictional information onto the real world. Rather than enhancing or adding to the real world, these products *modify* or *supplant* the real world in part. For example, popular products such as *Ingress* (Niantic, 2013), *Pokémon Go* (Niantic, 2016), and *Harry Potter: Wizards Unite* (Niantic, 2019) are all described as augmented reality experiences. Overlays used in these augmented reality games are primarily transformational: e.g., “Gyms” in *Pokémon Go* replace a real-world building with a virtual equivalent – at least when viewed through the frame of the game. There is a corresponding semantic debate surrounding the use of ‘augmented reality’ in these instances as a result: i.e., is such supplantation or modification necessarily also an augmentation? However, popular use of the term ‘augmented reality’ generally ignores this underlying semantic debate: the term ‘augmented reality’ does not require the presented virtual modifications to the actual physical world to be factual, or accurate, or even relevant.

Throughout this thesis, ‘augmented reality’ will be used to encapsulate this broader usage of a ‘digitally mediated reality’ being an ‘augmented reality’; the phrase ‘augmented reality’ refers to any product which begins with a real object or space and then digitally alters it or otherwise mediates it in any way. This means that the use of the term ‘augmented reality’ in this thesis will additionally include concepts such as ‘reduced reality’ – where augmented reality technologies are co-opted in order to remove objects from the virtual presentation of a real space.

0.1.2: Virtual Reality

Virtual reality is, ironically, perhaps the most problematic term within the field of virtual experience. Historically, the term has been used to describe almost every form of virtual experience. Ken Pimentel and Kevin Texeira, for example, use the term ‘virtual reality’ to refer to technologies which allow you to “step through the computer screen into a 3-D artificial world” where you can “interact with computer worlds” (Pimentel & Teixeira, 1993, p. xiii). This could include essentially all expressions of virtual experience. Similarly, 'virtual reality' has been conflated with other terms such as 'cyberspace.' For example, John W. Murphy's criticism of the 'virtualisation' of museums describes various expressions of virtual experience using the term 'cyberspace' – ranging from virtual/digital reconstructions which could be explored by users to simple informational resources placed on a computer (J. W. Murphy, 1996). This is likely due to the closeness of 'virtuality' and 'virtual reality' – in themselves, terms which are often conflated. G.M. Peter Swann and Tim P. Watts summarise the disagreement surrounding these terms as having two poles of purism and pragmatism:

> By one of the most common working definitions, VR is intuitive, interactive, and real-time 3D graphics. … But even if we accept this definition, different members of the VR community interpret it in very different ways. One critic of VR put it to us that much of what is currently marketed as VR is none of these: it is not intuitive, interactivity is theoretically possible but practically absent, any system based on a PC using a standard PC operating system cannot be real time, and VR viewed on the monitor of a PC is not even 3D. Some purists have taken a very narrow definition: VR models must be viewed through a head-mounted display, which offers genuine stereo vision. At the other end of the spectrum are those pragmatists who adopt a wider, encompassing definition of VR… most of the VR applications selling in the market of 2000 are non-immersive in the purist's sense. The VR models are viewed on PC monitors, and the inputs used are for the most part the keyboard and the mouse. We would not presume to judge whether the purists or the pragmatists are right. (G. M. P. Swann & Watts, 2002, pp. 54–55)

Whilst Swann & Watts's comments were made over twenty years ago, similar examples of 'virtual reality' being used as a general term for any virtual experience can be identified in research from the last decade: Richard A. Blade and Mary Lou Padgett attempted to define standards and terminology for use within the field of virtual reality, setting out a position that 'virtual reality' and 'virtual environment' are functionally synonymous as terms for "model[s] of reality with which a human can interact, getting information from the model by ordinary human senses such as sight, sound, and touch and/or controlling the model using ordinary human actions such as position and/or motion of body parts and voice" (Blade & Padgett, 2015, p. 33). Their definition of virtual reality therefore essentially encompasses almost all expressions of virtual experience.

It is this conflation of terms which led Jonathan Steuer to declare that the "most effective solution to the problems with the current usage of *virtual reality* would be to abandon [the term] entirely" (Steuer, 1992, p. 74, emphasis original). Instead, Steuer suggests that virtual reality is "a real or simulated environment in which a perceiver experiences *telepresence"* (Steuer, 1992, p. 76, emphasis original; i.e., virtual reality occurs when a person feels as if they were present at a place other than their current real location). By focusing on telepresence, Steuer establishes a simple test for use of the term virtual reality: do you feel like you are still where you are? Perhaps this idea underlies Kishino & Milgram's definition of a virtual reality environment as "one in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world" (Kishino & Milgram, 1994, p. 2) – although it should also be noted that Steuer does not relate virtuality to a synthetic (i.e., created) world, but rather to the transportation to a different location, thereby avoiding the primarily technological focus described in section 0.0.0.

More modern definitions of 'virtual reality' tend to define virtual reality as that which redirects or replaces the perspectives of the participants. Succinctly, virtual reality participants are isolated and separated from reality by or through the virtual experience through the redirection or replacement of objects within their perceptual horizons. For example: a head-mounted device for virtual reality (HMDVR) positions screens closely to the eyes and blocks out participants' vision of the externality they are situated within. This separates them from reality and replaces the real visual externality with an entirely virtual visual environment. This thesis uses virtual reality to describe such immersive virtual experiences, with a specific focus on those which are intended to be *full-immersion* experiences, in which the real externality is completely supplanted by a visual environment.

0.1.3: Mixed Reality

Much like *virtual reality*, *mixed reality* is an often-controversial term. Kishino & Milgram suggest that mixed reality should be used as an overarching term for any environment which combines physical and virtual elements. In their visual representation of Milgram's virtuality continuum, Kishino & Milgram suggest that mixed reality encompasses both augmented reality and augmented virtuality environments (Kishino & Milgram, 1994, p. 3) – and presumably any other combinations of real and virtual. More recent use of the term *mixed reality* sees it as something distinct from both *virtual* reality and *augmented* reality. However, it is not always clear that these differentiations are entirely meaningful. For example, Tara Brigham contrasts augmented reality – "see the real, physical world…overlaid with a layer of digital content" (Brigham, 2017, pp. 172) – with mixed reality by emphasising the use of "real-looking virtual objects" in mixed reality (Brigham, 2017, p. 175). The examples as described by Brigham would clearly fall within what is referred to as 'augmented reality' within common usage and within this thesis: e.g., a 2016 review of literature on medical uses of augmented reality focuses on the ability to see "hidden organs inside a body" (Ha & Hong, 2016, p. 242); Henry Pence suggests that "augmented reality [can] enhance the experience at cultural heritage sites by adding…missing paintings or statues" (Pence, 2010, p. 138). Brigham's differentiation between mixed reality and augmented reality is not reflected in the current common usage of the terms.

Conversely, Jean-Francois Lucas et al. focus on the use of mixed reality in non-realistic settings, exploring two "cultural products" developed between the city of Rennes (France) and digital sites in *Second Life* (Linden Lab, 2003). Rather than seeing mixed reality as a specific technology or experiential framing, they argue mixed reality as "a spatial composite made from the hybridization of a physical space and a digital space, not their addition or their superposition" (Lucas et al., 2012, p. 284). Similarly, Mikhail Ostanin et al. investigate the use of mixed reality for controlling robots in industrial settings. They view mixed reality as "bringing the real and virtual worlds together, which allows humans and robots to operate within a common coordinate system" (Ostanin et al., 2020, p. 1). This approach is the one which will be taken throughout this thesis. Mixed reality will be distinguished from augmented reality through considering the symbiotic or co-existential properties of real and virtual within an experience: mixed reality invites reality and virtuality to co-exist whereas augmented reality adds virtual information to a real space (i.e., the virtuality is subordinate to the reality).

0.1.4: Defining Virtuality

The way in which language surrounding technology changes often reveals ways in which perspectives on that technology have changed. A recent shift towards more precise meanings for each of the terms explored in this introduction exists because the underlying understanding of 'virtuality' has expanded. As the big picture grows bigger, more precise terms must be used to ensure that there is clarity as to which section of the picture we are addressing. One example of such changes in perception can be found in the attempt by Kishino & Milgram to differentiate between virtualities in their 1994 taxonomy. They raise an interesting issue: what if an experience is virtual, but a physical item is provided to enhance the experience? We might imagine, for example, that someone in a virtual racing simulator be given a physical steering wheel to hold. Kishino & Milgram focus on the specifically *technological* experience first: how does the user view the experience? Is the viewed experience predominantly virtual or predominantly real? How do they interact with the virtual world? Are there elements of the virtual world which are real? Predominantly, they focus their approach on considerations of the visual technologies used in virtuality. For example, they draw upon examples of different types of video display which might be used to facilitate virtual experiences. Both non-immersive computer monitors (such as those widely used on PCs and laptops today) and head-mounted displays (such as those used in what are commonly identified as ‘virtual reality’ headsets) are considered to be ‘mixed reality environments.’ These contrast with ‘virtuality reality environments’ – which are described as “ones in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world” (Kishino & Milgram, 1994, p. 2). For them, a virtual reality environment requires complete simulation: the presence of a physical steering wheel immediately precludes the possibility that the environment could be considered ‘virtual reality.’

Writers such as Kishino & Milgram focus on virtuality as an entity distinct and separate in its own right – a fully simulated reality. Conversely, modern taxonomic determinations place virtuality and reality in the same experiential space: the thrust of discussion surrounds the authenticity of the experience, rather than the underlying technologies for virtuality. For Kishino & Milgram, the presence of a physical steering wheel detracts from the virtual experience of being a race car driver because it is not simulated: virtuality is challenged. For many modern technologists, the presence of a physical steering wheel enhances the experience of being a race car driver because it creates a more ‘authentic’ experience: virtuality has been enhanced. These considerations underpin the need for a conceptual shift from a *technological* or *visual* approach to an *experiential* approach for defining and investigating virtuality. . Indeed, we can distinguish *augmented*, *virtual* and *mixed* reality through examining relationships between virtuality and reality:

(1) Virtual reality *supplants* reality with virtuality. Reality is sub-ordinate to virtuality: e.g., HMDVR games.
(2) Augmented reality *builds upon* reality with virtuality. Virtuality is sub-ordinate to reality: e.g., smartphone applications.
(3) Mixed reality *combines* reality and virtuality. Virtuality is symbiotic with reality: e.g., a music concert held simultaneously in an online space and a physical space.

0.2.0: Technological Overview

Throughout this terminological exploration, there has been a deliberate emphasis on moving away from technological perspectives on virtuality and moving towards a broader understanding of virtual experience which does not focus exclusively (or even predominantly) on specific technology. However, it is irrefutable that technological solutions are a key part of the modern context for various virtual experiences. This section reviews current technology used to facilitate virtual experiences and otherwise create virtuality. As with most technology, devices for virtuality have had significantly accelerated development in the 21st Century. This section contains, therefore, a guide to the common technological solutions used in the early 2020s for virtual reality, augmented reality, and mixed reality experiences. Alongside these common solutions, a few examples of lesser-known and/or complex examples will be provided, demonstrating some of the variety of technological approaches taken to virtuality in the last twenty years. This will illustrate how the terminology set out above intersects with the variety of technologies and experiences available at the time of research.

Attempting to list every possible technology used to create virtuality would be an exercise in futility: both because no such list could ever be realistically holistic, and because any such list would be outdated almost before a reader reached the end of the chapter. The effect of COVID-19 on virtuality should also not be understated. Whilst predictions of future working practices throughout the world are likely unreliable, COVID-19 led to dramatic claims about the possible expansion of technology use in the workplace. Many of these included expanding the virtual domain within the workplace. For example, it seems probable that the practices of using remote meeting tools such as Microsoft Teams will continue to replace in-person business meetings in many places, enabling a greater degree of working from home, or working across international borders. Functionally, we could view this as a virtual experience located within the workplace. Whilst the virtual meeting has not quite replaced the physical meeting, there are clearly lasting effects of COVID-19 which will place additional pressures on technology firms to develop new virtuality solutions. However, whilst not a holistic list, examples of all major solutions for virtual experiences will now be provided to contextualise the taxonomy set out above and provide a single-moment snapshot of available technology.

0.2.1: Virtual Reality Technologies

By far the most known virtual reality technologies are head-mounted devices for virtual reality (HMDVRs). In HMDVRs, a display is mounted in front of the face – typically providing an independent display for each eye. These displays present a view of a virtual world and are often controlled through some form of inertial measurement unit (IMU); when the user turns their head, their view within the virtual world also rotates. The aim of this system is to provide an experience analogous to the user being present within the virtual world. Whilst some HMDVRs are visual-only or have a visual-only configuration, most commercially available designs are intended to be used in an audio-visual configuration. In most of these cases, speakers are mounted adjacent to the ears. These speakers vary drastically in design, with commercially available models providing audio solutions ranging from closed-ear headphones to off-ear mounted speakers (e.g., Valve Index) to headband-mounted directional speakers (e.g., HTC Vive, Oculus Quest). Many commercially available HMDVRs allow users to connect their own audio devices (e.g., PSVR provides the user with a 3.5mm audio jack, allowing them to connect most personal audio systems).

HMDVRs have also been combined with other technologies to create a more immersive experience. These experiences have proven effective for training skills in controlled environments (Marler et al., 2020; Xie et al., 2021), such as in America where New York Police Department reported the addition of VR training simulators for active shootings and other difficult-to-recreate policing scenarios to their training in 2019 (Eyewitness News ABC7NY, 2019). This is part of a broader global adoption of VR training for law enforcement officers reported in the UK (Derbyshire Constabulary, 2020; ITV News, 2021), Australia (SA Police News, 2017), Canada (Government of Canada, 2017; CBC News, 2019), and elsewhere. Similar programs have also been rolled out for other government organisations, e.g., firefighting services (Chanthadavong, 2017), military personnel (Nye, 2021). These programs typically combine HMDVRs with physical gear that recruits must learn to use (e.g., fire extinguishers, firearms, tasers) within dangerous situations that cannot be safely and accurately replicated in a more traditional, less virtual, environment.

Multi-modal virtual reality experiences aimed at a consumer market have been slower to emerge. US-based Positron offers a “premium cinematic experience” which combines HMDVRs with a multi-sensory delivery system in the form of a ‘pod’ or chair. This chair can rotate, vibrate, tilt, dispense aerosolised chemicals (i.e., for olfactory engagement) and integrate with other immersive technologies. As a result, the user receives an audio-visual experience which augments a typical HMDVR experience through haptic feedback, scent dispensers and other technologies to create a “new level” of immersive experience[4] (Positron, 2023). We could recognise these as echoing approaches used in *Hale’s Tours of the World*, which attempted to construct a multi-sensory recreation of a train journey in late 19th– and early 20th-century amusement parks using sound effects, wind machines, and panoramic film screens which generated what was described as “intense realism” such that “it was difficult to realise, after such a ride, that one had not actually been in Switzerland” (B.S. Brown, 1916, p. 372).

Despite the focus on HMDs, they are not the only technology used for creating virtual reality experiences. Holistic virtual reality environments are frequently used for training environments. These

[4] Whether this creates a distinct “new level” is subjective. The term is presented as found within Positron marketing materials. Similarities can be observed between Positron’s cinematic experience and other experiences such as the “4D-film experience” popularised at theme parks and earlier commercial products such as Sensorama. These experiences (also described as “full-immersion” in *their* respective marketing), all utilised additional technology to provide a multi-faceted cinematic experience. It is most accurate to consider the use of “new level” to highlight a shift in our understanding of what “full immersion experience” *can* mean.

environments combine physical hardware (e.g., a plane cockpit) with virtual images delivered via projection or large image displays. The trainee placed within these environments does not wear a HMDVR but is similarly immersed within a simulation as they are unable to see past the boundaries of the virtual reality environment. Similar to virtual reality training environments which integrate a HMDVR, these systems are used to provide a controlled environment in which to teach users how to handle dangerous situations. Some success has been found in the use of virtual reality training environments designed to improve railway safety and critical event response (Tichon et al., 2006).[5] Similar training programs have also been designed for airplane pilots (Parsons, 2019), operators of heavy industrial machinery such as cranes (Mevea, 2019), and racing car drivers (Bower, 2018). It seems likely that virtual reality environments will become an essential part of training in many areas within the near-future, above and beyond the areas which they have already revolutionised.

0.2.2: Augmented Reality Technologies

If we understand that augmented reality is, as Berryman suggests, created through “the process of overlaying computer-generated information on reality, whether that reality is a geographic place or an object” (Berryman, 2012, p. 213) then we can recognise that virtual reality technologies will not be directly applicable to augmented reality. Virtual reality solutions aim to exclude the real from the virtual; users are intended to perceive only the virtual. Conversely, in order to overlay virtual information on the reality, augmented reality solutions must allow users to perceive the virtual and the real simultaneously as the user experience combines the two: virtual elements are placed within an experience to build upon elements of the real-actual.

Augmented reality HMDs must therefore permit users to perceive the real-world around them. Devices which allow users to see the real *alongside* the virtual are often referred to as “Optical Head-Mounted Devices” (OHMDs) to differentiate them from other HMDs such as HMDVRs which

[5] Two illustrative examples have been published by educational YouTubers: Geoff Marshall was invited to visit the Rail Operation Centre in Basingstoke, which houses railway simulators used to train South Western Railway drivers on correct operation in difficult weather conditions (‘Geoff Marshall’, *Driving the SWR Simulator in Bad Weather Conditions*, 2019); similarly, Tom Scott was invited to visit the Eisenbahnbetriebsfeld – a model railway in Darmstadt (Germany) used to simulate railway signalling for trainee train controllers (‘Tom Scott’, *The world’s most useful model railway*, 2021).

obscure the real world. One of the most successful entries to the OHMD market is the Microsoft HoloLens product line – most recently including the HoloLens 2 – which has seen extensive use in commercial environments. OHMD devices such as the HoloLens are typically self-contained (no additional hardware is required) and untethered (no wired connections required). HMDs typically designed for VR can also be adapted for use in AR. For example, the Oculus has a “pass-through” mode which uses video feeds from HMD-mounted cameras to allow users to perceive the world beyond the headset.

Modern augmented reality devices integrate digital interfaces and virtual objects into the space surrounding the user. This typically requires a scan of the surrounding area to be completed. For example, the official Microsoft documentation explains how the HoloLens 2 uses visual light to do this. Unique features of the externality around it are identified through a visual scan and mapped to a specific location using Wi-Fi fingerprinting for the device’s future reference. However, this process is not necessarily perfect: moving objects, reflective surfaces, non-unique environmental features (e.g., multiple workstations that all look essentially identical), and even specific types of lighting can confuse the device’s spatial tracking and environmental integration (Microsoft, 2022).

These devices have many potential uses, but the HoloLens 2 focuses predominantly on professional applications. In their 2019 launch trailer, Microsoft promote their vision of HoloLens 2 as being used to aid in engineering, medicine and theatre: holograms help professionals working across many fields to understand their specific situation more clearly thereby improving their work output (Microsoft HoloLens, 2019). The device (and the predecessor ‘HoloLens’) was also used as part of museum exhibits (Microsoft, 2018b, 2018c, 2019), in collaboration with NASA as part of the Curiosity programme and in public-facing experiences surrounding the Mars Programme at NASA (NASA Jet Propulsion Laboratory, 2016), and to create augmented reality art exhibitions in public spaces (Microsoft, 2018a).

Whilst OHMDs represent the current ‘cutting edge’ augmented reality technology, earlier technologies could also be described as augmented reality. For example, audio tours – or ‘ambulatory lectures’ – were first pioneered in 1952 to provide additional information and engage visitors to popular tourist destinations such as museums or galleries. Visitors were able to explore a location – often in an order of their choosing – whilst receiving additional audio commentary specific to each location they visited. In more modern iterations, audio tours now implement location-tracking technology and/or smartphone integration which reduces the inconvenience of engaging with such experiences and therefore encourages visitors to participate (Pence, 2010).

Smartphone technology also facilitates many other augmented reality experiences. As already highlighted in this chapter, informational overlays are constantly used in modern society: e.g., the virtualisation offered through apps such as Google Maps develops traditional maps through providing additional information such as historical photographs, information and reviews about locations and businesses within an area, traffic and population updates, and more. These are rudimentary augmented reality environments, presenting the 'virtual' information alongside the 'real' world (in the form of a map); in the more recent versions of these applications, it is possible to overlay this information in an OHMD which utilises GPS tracking and photo-recognition software to locate where a user is and what building a user is looking at. As discussed in section 0.1.1, augmented reality games such as *Pokémon Go* are also popular on smartphones. Similar to apps such as Google Maps, *Pokémon Go* players receive additional 'virtual' information alongside the 'real' world, with the additional caveat that the information presented in *Pokémon Go* is deliberately fictionalised and fantastical, rather than intended as an accurate source of information about the world.[6] Other mobile devices (such as wearable devices) also offer useful expressions of augmented reality, often with a specific focus on making personal lives more convenient.

0.2.3: Mixed Reality Technologies

Similarities between the technology used for mixed reality and the technology used for augmented reality can be frequently observed. This should be expected, as many expressions of both mixed reality and augmented reality require users to be able to perceive the real and the virtual simultaneously. This can also be seen in marketing materials for OHMDs such as the HoloLens 2,

[6] For more information about Pokémon Go, see section 1.3.2.

which has been described as "mixed reality ready" or as a "mixed reality device" by Microsoft (Microsoft HoloLens, 2019). Therefore, the distinction between mixed reality experiences and augmented reality experiences typically emerges from differences in how the user interacts with and perceives virtuality within the experience. Where augmented reality uses virtuality to supplant or supplement reality in a unidirectional (hierarchical) relationship, mixed reality uses virtuality in a bidirectional (symbiotic) relationship; mixed reality positions reality and virtuality as symbiotic partners, whereas augmented reality uses virtuality to impose a *new* perception of reality onto the user.

A number of completed mixed reality projects were examined in section 0.1.3. Ostanin et al. describe a mixed-reality control system for a human-robot interface to allow for intuitive and effective use of the robot device. One feature discussed in some detail was the integration of HoloLens technology in order to "allow controllers to see the path of the robot's movement [in the real-actual space], thereby reducing the number of robot control errors" (Ostanin et al., 2020, p. 11). Other uses of mixed reality projects have been less pragmatically-minded. Lucas et al. aimed to create an integrated shared space for a cultural event which permitted social interactions between audience participants. This was considered mixed reality as it combined users engaging with the experience in many different ways; in-person, through a computer screen, and through a virtual reality headset (Lucas et al., 2012, pp. 278–280). This allowed artists and audiences to create a multi-modal concert experience that existed simultaneously within the virtual world and the real-actual world.

This discussion of mixed reality experiences also exemplifies the issues in the primarily technological frameworks discussed in the introduction to this chapter: whilst the same hardware can be used for both augmented reality and mixed reality environments, the underlying virtual experience presents different virtual-real relationships. Technologically deterministic approaches can simplify virtual experience, occasionally to the point of 'one technology for one experiential framework'. As this section reveals, this is not the case. Similarities between mixed reality and augmented reality technologies suggest that the experiential perspective adopted in this thesis may offer a clearer differentiation between categorisations of virtual experience.

0.3.0: Paradigms of Immersion

Another term frequently used when discussing virtuality is immersion. Like terms used to define virtuality, significant differences between definitions of immersion can be observed and it is often hard to establish exactly what is meant by use of the word ‘immersion.’ Gordon Calleja emphasises these difficulties, and suggests that they emerge “because it is used to widely when discussing experiential facets of anything form digital games to painting, literature, and cinema” (Calleja, 2013, p.222) such as “to describe the affective properties of hardware” (p.224), to describe “a psychological state” of engagement with an experience (p.224), to describe “involvement in non-ergodic media” (p.226), and so on. Given this confusion, this section of the thesis examines some of the underlying work on ‘immersion’ which has been conducted over the last thirty years and works toward a definition of ‘immersion’ which will be utilised throughout this thesis.

In attempting to find a clear definition, with a specific focus on video games, Emily Brown and Paul Cairns empirically argue for a model of immersion which functions hierarchically across three levels: engagement, engrossment, and total immersion. In the first stage, participants must “invest time, effort, and attention” into the experience. Brown & Cairns suggest that a particular issue in overcoming this initial stage of immersion is accessibility of control schemata within an environment: “the controls and feedback [to the user’s input] need to correspond in an appropriate manner so that the user can become an expert [in using the main controls]” (Brown & Cairns, 2004, p. 1298). Once the user has invested sufficient time and effort to master the rules and controls for interacting with the environment – and shown a willingness to concentrate on interacting with the environment (i.e., attention to the environment) – then they approach the second stage: engrossment. Brown & Cairns specify that engrossment requires an *emotional* engagement with the environment: “the [environment] becomes the most important part of the [user’s] attention and their emotions are directly affected by the game. … The [user] is now less aware of their surroundings and less self-aware than previously” (p. 1299). At the final stage of total immersion, Brown & Cairns identify the need for presence within a virtual environment; they describe how users at this stage felt “cut off from reality.” Should a developer desire to create this final stage of total immersion, they identified that developers might need to create environments which require multiple elements of attention – e.g., “visual, auditory, and mental attention” – as the level of immersion “seemed to correlate to the number of attentional sources needed as well as the amount of each attentional type” (p. 1300). Brown & Cairns are describing immersion as a system of involvement; immersion is seen as deep engagement with a task, or distraction from reality.

Other models of immersion take alternate approaches to exploring the concept. Rather than categorising the *depth* of immersion, Laura Ermi & Frans Mäyrä categorise *components* of immersion based on empirical field studies and user interviews. Ermi & Mäyrä particularly focus on the sensation of immersion that players might feel whilst playing video games. However, their research could be more broadly applied to all virtual environments. In their study, immersion was argued and defined to be greatest where participants were engaged in three ways: through sensation, through imagination, and through challenge. Sensorial immersion relates to the “audio-visual execution of games.” Ermi & Mäyrä describe how “screens close to the [user]’s face and powerful sounds easily overpower the sensory information coming from the real world” (Ermi & Mäyrä, 2011, p. 101). Imaginative immersion relates to the dimension of the game experience “in which one becomes absorbed with the stories and the world, or begins to feel for or identify with a game character” (p. 101–102). Finally, challenge-based immersion relies on interaction with the virtual environment: were users faced with mental or physical challenges which they had to overcome as part of their engagement with the environment? (pp. 102–103). Similarities to Brown & Cairns should be acknowledged: all authors describe the physical challenge of controlling the virtual environment (or an avatar within the virtual environment) and the need for an emotional connection to the virtual environment. However, Ermi & Mäyrä offer a – potentially more useful — non-hierarchical structure, which offers greater flexibility in considering a variety of virtual environments than Brown & Cairns’ hierarchical structures. For example, non-interactive experiences such as a film might engage a user emotionally (and thereby immerse them within the virtual environment of the film) without requiring an investment of time to learn how to control or navigate the immersive environment. Despite a more flexible structure, however, Ermi & Mäyrä seem to concur with Brown & Cairns in that immersion is a system of engagement with virtuality.

Alongside these paradigms of immersion is another term closely related to concepts of immersion: ‘flow’. To define this term, Mihalyi Csikszentmihalyi theoretically describes a zen-like state in which the sense and awareness of self (and other external objects) “disappears from awareness” (Csikszentmihalyi, 1991, p. 62), which is sometimes accompanied by a sense of union with the environment or experience. When in this state of flow, time no longer seems to pass the way it ordinarily does, with some moments stretching out and other moments compressing hours into brief instances (Csikszentmihalyi, 1991, p. 66). Csikszentmihalyi later attempted to categorise his elements of flow into conditions and characteristics of flow. He states that the conditions of flow are the clarity of goals, immediacy of feedback and a balance between the skills of the individual and the challenges of the task they are attempting to complete. The characteristics of flow are identified as concentration on the task at hand, total absorption in the task, decreased sense of self, a sense of control over the outcome, and the transformation of time (Csikszentmihalyi, 2000, pp. 389–391). However, Csikszentmihalyi’s categorisations are not yet universally accepted. For example, Christian Swann et al. suggest a differentiation between *necessary* conditions and *sufficient* conditions for flow to be experienced (C. Swann et al., 2012, p. 817) whilst Cairns et al. argue that flow can only be experienced when all of the characteristics of flow identified by Csikszentmihalyi are present within an experience (Cairns et al., 2014, pp. 343–344). Attempts to determine the conditions of flow experimentally have also been unable to provide a definitive answer: for example, Yongxia Skadberg and James Kimmel empirically assessed the sense of flow experienced whilst browsing the internet and found that enjoyment, time distortion, and telepresence best characterised flow in this context (Skadberg & Kimmel, 2004), whilst Martin Klasen et al. experimentally assessed the sense of flow experienced during video game play and found that the predominant factors affecting flow were the balance between skills and challenges, concentration and focus, clear goals, and control over the video game (Klasen et al., 2012). Other investigations have provided similarly conflicting reports on the primary considerations and characteristics of flow – perhaps suggesting that flow is experienced differently across varying contexts.

However, notwithstanding these difficulties in precisely defining what flow is and how flow emerges, flow is frequently connected to concepts of immersion. Both immersion and flow seem to share an underlying concept of minimising a person's awareness of the (or at least: an) external reality. In a review of literature surrounding both immersion and flow, Lazaros Michailidis et al. argue that research into immersion and flow has "not been effective in diversifying the measurement of flow and immersion as separate states" and suggest through critical review that this means that "immersion and flow are not conceptually distinct, and their proposed differences are not compelling enough to set immersion apart as a different mental state" (Michailidis et al., 2018, p. 5). Whilst this may be true, it should also be noted that discussions surrounding flow and immersion also seem to emphasise a conceptual difference between the two terms. Whilst immersion is sometimes described in terms of *awareness* of the virtual, flow is generally described in terms of *lack of awareness* of the real.

Conversely, others suggest that immersion is connected to presence: "the perceptual illusion of non-mediation" (Lombard & Ditton, 1997, JCMC321). We have already heard how Steuer defines virtual reality as being an environment in which one experiences *telepresence*: the experience of relocation to another location without movement in the real-actual (Steuer, 1992, p.76). Others apply a similar logic to broader definitions of virtual experience. For example, Mel Slater et al. describe how immersion leads to "the participant sense of 'being there' in the virtual environment" (Slater et al., 1994. p. 131); i.e., (tele)presence is the result of immersion. This was emphasised in later research based on on-going empirical work as specifically *not* requiring engagement:

> In reality we might be completely uninterested in what is going on in the physical space in which we are located, but this will not destroy the sense of being there. … Now suppose we are in a virtual reality that renders this same boring place, and similarly we become uninterested and not engaged in what is happening. This actually is not incompatible with the illusion of "being there" (Slater et al., 2022, p.7).

Therefore, we can recognise that the sense of relocation (or 'teleportation') represents a paradigm of immersion distinct from an immersion of engagement: an immersion as *teleportation*. Participants can be immersed in an experience by feeling like they are truly present in another location or place; they no longer perceive the real-actual.

The different definitions of immersion reflect different forms of engagement with virtual experience. It is likely that differing forms of virtual experience will lead to one or the other of these paradigms of immersion being more prominent. When viewed holistically, however, these relationships between these different paradigms of immersion suggest that we could consider immersion as an increasing awareness of virtuality or a decreasing perception of reality. We can relate this to Milgram's virtuality continuum: at one end, virtual reality is a *complete* immersion, intended to subsume the user into a world completely devoid of the real; reality is to be avoided, discarded and disregarded. The user has been completely subsumed into virtuality and has no conceptualisation of reality during the experience. At the other, augmented reality emphasises and relies on the presence of the real: reality is to be embraced, used, involved. Reality is a constant reference point within the experience (Figure 9).

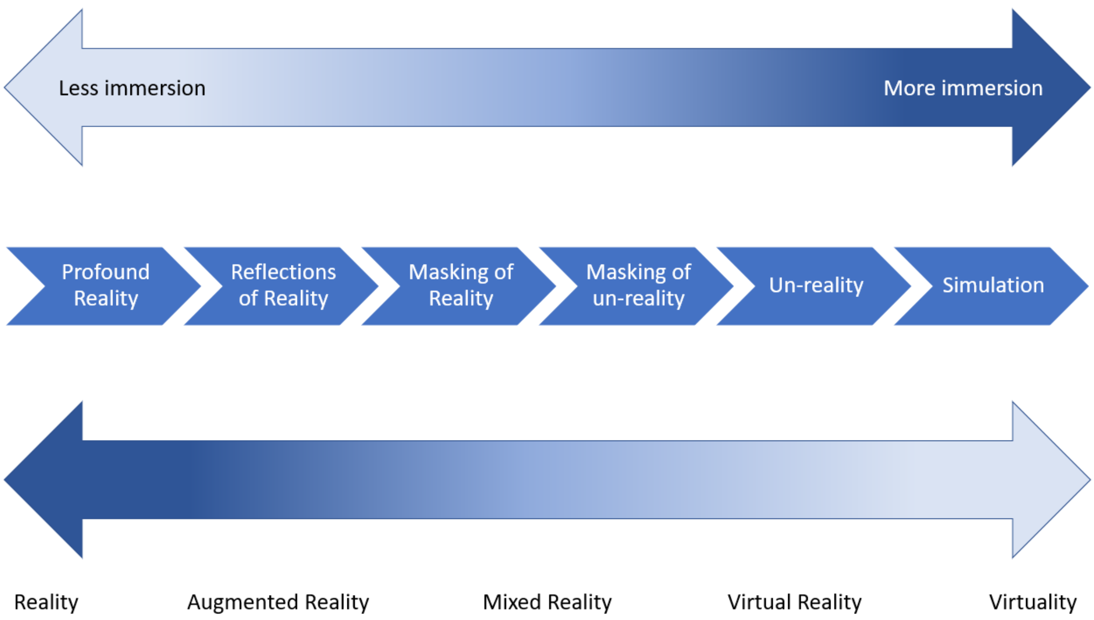

Figure 9: Baudrillard's stages of immersion (centre) compared to Milgram's virtuality continuum (bottom)

By furthering combining considerations of immersion with Baudrillard's six degrees of simulation, we can observe another relationship between the framings of flow/immersion and degrees of simulation: augmented reality experiences reflect on or mask reality, whilst mixed reality might begin to introduce 'un-real' elements. Virtual reality, in this comparison. corresponds to simulation or complete un-reality. To distil these differing perspectives on immersion and unify them together throughout this thesis, I propose greater immersion occurs when the 'un-real' or the 'virtual' is accepted by users to a greater degree, or when users reject (or fail to perceive) reality. This may emerge within a specific domain – e.g., sensorial immersion, narratival immersion, challenge-based immersion – or across several domains. Each of these domains will increase the user's overall sense of immersion (i.e., their perceived immersion) within the experience. A user who is less immersed is more likely to be conscious of reality and to therefore view an experience as being less virtual.[7]

0.3.1: Immersion and Authenticity

An important concept related to immersion is the concept of authenticity. Slater & Sylvia Wilbur suggest that immersion within a virtual experience increases as the overall user perception of the virtual experience approaches an (imagined and idealised) equivalent non-virtual experience (Slater & Wilbur, 1997). Slater later developed this into the specific notion of the plausibility illusion: "the illusion that what is happening is real even though you know that it is not real" (Slater, 2009, p. 3556) In this later development, Slater argues in particular that user immersion is dependent upon the credibility of the user experience of the virtual environment when compared to the expectations the user has of what that environment would be like should they ever enter it in reality.

In more recent literature, these relationships to (subjective) authenticity have been explored in more detail, with specific links being made between immersion and authenticity: for example, Annie

[7] Parallels can be found between this definition of 'immersion' and Calleja's suggested term 'incorporation' which "offers a more precise conceptualization of the virtual-world habitation that is media specific and multifaceted, while at the same time providing a fresh perspective on the issue which is not weighed down in the confusion arising from the exclusionary logic that underpins the immersion and presence metaphors" (2013, p.233–234). Perhaps the key difference is that Calleja focuses on specifically digital environments (e.g., video games), and the use of 'immersion' here is intended to encapsulate all virtual experience. However, those interested in further exploring debates surrounding the term 'immersion' may find it useful to also consider *In-Game: From Immersion to Incorporation* (Calleja, 2011).

Wang et al. suggest that authenticity of action, authenticity of environment, and authenticity of narrative combine to create a sense of ‘presence’ (and therefore immersion) within virtual reality environments. (A. Wang et al., 2021, p. 3). Slater et al. (2022) offers an update to the earlier research by contesting that plausibility should be assessed not through the perceived realness of the environment, but through the realness of the events happening within the experience (p.12). Slater et al. (2022) objects in particular to the recontextualization of the plausibility illusion conducted by Richard Skarbez et al. (2017), which suggests that presence (and immersion) of participants should be measured by the perceived realness of the virtual experience and the coherence (i.e., the similarity of the expected information and the received information on sensory, perceptual, and cognitive layers of virtual experience) of the virtual experience. Each of these models – and other similar models which attempt to combine or blend elements of the Slater-model and Skarbez-model, such as Marc Erich Latoschik & Carolin Weinrich’s congruence-plausibility model for XR (2022) – present a slightly different approach to considerations of authenticity. Emerging from this body of recent literature are particular concerns surrounding authenticity of fictional or fantastical experiences, which are explored in relationship to musical experience in Chapter 4 of this thesis. Throughout this thesis, the broadly agreed sense of ‘plausibility’ defined in Slater et al (2009, p. 3556: “the illusion that what is happening is real even though you know that it is not real”) will be used, rather than more contested definitions emerging in later literature.

Other relationships between authenticity and immersion at the intersection of sound and virtuality are explored in Chapter 3 of this thesis, and work such as Isabella van Elferen’s ALI model for connecting music and immersion (2016) are examined in Chapter 4. One final aspect of immersion that is particularly relevant to the experiential approach used throughout this thesis is the nature of immersion as being, in many ways, connected to the user’s knowledge of and identification of the virtual world as being ‘virtual’ – the final challenge of overcoming this barrier to immersion in order to create a ‘full-immersion’ experience is examined in Chapter 6.

0.4.0: Research Methodology and Foundations Toward a Framework for Virtual Experience

This thesis presents an experiential perspective on sound and music in virtuality. It creates a framework with two primary axes: virtuality and audio. By centralising and interrogating user experience, our understanding of virtuality is moved beyond a technology-focused approach and towards a participant-focused approach. This will allow us to better understand virtual experiences of the current day and use these virtual experiences to design and develop virtual experiences of the future. The work in this thesis is both interdisciplinary and inter-contextual: it draws on literature across different fields of study to construct a framework for virtual experience in order to allow for purposeful comparisons between different, yet similar, modalities of experience. Virtuality is, therefore, taken in the broadest sense throughout this thesis: virtual experiences exist wherever users are expected to supplant aspects of the real-actual with aspects of the virtual or enhance/integrate aspects of the virtual into the real-actual. There are many different terms for types of virtual experience – such as virtual reality, augmented reality, and mixed reality – which reflect the myriad ways in which the symbiotic relationship between actual and virtual can be established for participants. Each different term reflects a different mediation between real and actual, and therefore a different experiential perspective on sound and music within those contexts. This thesis examines virtual experience holistically and aims to provide a framework which can discuss sound and music across all such mediations.

In developing this framework, I have predominantly drawn upon previously published materials on sound and music within virtual experiences, and on virtuality more broadly, alongside publicly accessible materials such as reviews, interviews, conference talks and design notes, and autoethnographic research reflecting on my own engagement with virtual experience. My research distils these materials, which often detail specific experiences, into a broader framework for understanding sound and music within virtual experience.

In developing an experiential perspective, my approach adopts a research methodology, often applied in the field of sound for interactive media, described by Michiel Kamp as phenomenological: a “systematic account of not just the experiencer…but of the experience itself…a reflection on conscious experience in order to find logical preconditions for those experiences” (Kamp, 2021, pp.170–171). Such an approach “imagines variations on a phenomenon…[to] identify its essential characteristics” (p. 172) which, as Kamp suggests, “lends itself best to experiences widely shared, but not thoroughly understood” (p.172). Virtual experience – as the research in this thesis will demonstrate – is both widely shared and not yet thoroughly understood. As with other phenomenological research, the intent of this thesis is to create a nomothetic understanding of music and sound within virtuality that can be applied not only to understanding and analysing existing virtual experiences but also to designing and implementing emerging and future virtual experiences.

This phenomenological approach – a kind of reflection on subjective experience toward a nomothetic construct – is at the core of much existing research in the field of sound studies (even if it is not explicitly labelled as phenomenological). Prominent examples include Barry Truax’s *Acoustic Communication* which explores “phenomena involving sound from a human perspective” (Truax, 2001, p.xvii), Calleja’s work on immersion (2011; 2013), and much of the extensive and influential work of Mark Grimshaw and Tom Garner, including their *Sonic Virtuality* which examines sound as systematic phenomena of emergent perception (Grimshaw & Garner, 2015. Despite the recent popularity of these approaches, however, we should nonetheless acknowledge some limitations of a phenomenological approach.

A theme throughout much of this research is the inherent subjectivity both of individual experience and specifically of virtual experience. Given that the research drawn upon emerges from dozens of countries across several decades of time, what ‘user’ or ‘participant’ is being discussed? Already, we have discussed concepts such as immersion, and drawn upon theoreticians that fundamentally disagree with each other about the precise mechanics and outcomes of immersion. As we move into discussing sound and music, we will acknowledge the auditory information that is present within various virtual experiences. These are inherently subjective matters. As Hildegard Westerkamp puts it: “A sound occurs. And is heard. But by which person? From which culture? In what mental state? What physical state? What psychic space? With what intellectual knowledge? Which past experiences? What age? From which gender? And so on” (Westerkamp, 1995, n.p.).

In recent years, particular attention has been paid to the "auraldiversity" of listeners: the "plurality of senses of hearing, encompassing the whole of human and animal nature and extending to machine listening" (Drever & Hugill, 2023, p.1). The terms 'auraldiversity' and 'auraltypicality' is explicitly intended to echo notions of 'neurodiversity' and 'neurotypicality' (Drever, 2019, pp. 92–93), as hearing is a cognitive process (as explored in Chapter 2) and auraldiversity can often intersect with someone's neurodiversity. For example, William Renel explores notions of auraldiversity in the context of d/Deaf and disabled people (Renel, 2023), Meri Kytö explores how auditory neuroprosthetics such as cochlear implants can create auraldiversity (Kytö, 2023),[8] and William Davies explores how neurodevelopmental conditions such as autism might correlate to auraldiversity (Davies, 2023). Whilst there are many other aspects of identity that might intersect with hearing – e.g., gender or cultural background – these aspects as explored by Drever and others set out a clear example of why this may matter: a model of listening which is predicated solely on 'auraltypical' understandings may facilitate the creation of inherently exclusionary environments.

These considerations are not a new issue in the research tradition that this doctoral project is situated within. As Kamp describes, "It is hard to deduce when [phenomological approaches] stray from the universal and become 'too applied', because they are ultimately and inescapably rooted in subjective experience" (Kamp, 2021, p.210). Succinctly, the key issue may be summarised as follows: is framing the 'user' in one specific way disregarding the experiences of other 'users' in an experience? This question applies even where the 'user' is only implicitly framed in one specific way, such as in autoethnography, which implicitly accepts the author as the user within the frame of the research. For example, Jonathan Godsall objects to the "generalisation" of 'player experience' in his review of William Gibbons' *Unlimited Replays: Video Games and Classical Music*:

> Some generalization around categories of high and low, and gamers' knowledge and experience of classical music, is surely acceptable in a trade for the breadth of examples and ideas discussed within this digestible book. Where the book does find room for more detailed case-study interpretations, though, these can themselves seem somewhat at odds with its broader conceptions of gamers, given they prioritize the author's own music-historical expertise over the likely experiences of other consumers. (Godsall, 2020, p.115)

[8] Cochlear implants are explored in greater depth in section 6.2.1.

But, as Gibbons suggests in their reply to this review, Godsall's concern emerges from the methodological conflict between "social-science approaches to (real, historical) players and humanistic approaches to (imagined, abstract) players" (Gibbons, 2020, p. 116) which is delineated most clearly in Espen Aarseth's earlier work on the 'implied player':

> While the humanities and the social sciences both have formal and informal methods, it is in their empirical object that they differ most: their conception of the player. For the social scientist, whether doing qualitative or quantitative research, the player is historical, situated, flesh and blood. For the humanist game scholar … the player is a necessary but uncontrollable part of the process of creating ludic meaning, a function that is created by the gameplay as well as co-creator of it. (Aarseth, 2014, pp. 129–130)

The underlying philosophy that underpins this debate has not been engaged with in this thesis, but a position on it has, clearly, been adopted, insofar as I am presenting an 'experiential' framework which privileges user perceptions: whose perceptions am I talking about, and whose experiences?

Of course, the goal of this framework is – as with any framework approached through phenomenological methods – to allow for the consideration of *any* user: a "real, historical" user; an "imagined, abstract" user; an auraldiverse listener; an auraltypical listener; and so forth. But in the process of *creating* this framework, and as with other research which adopts a similar approach, I understand the user to be a nebulous series of potential understandings and actions, rather than attempting to generate specific observations of an actual user. That is to say: the 'user' or 'participant' or 'player' discussed throughout this research is not any specific individual, but an abstracted participant imagined within the underlying virtual experience; the user is part of a systemic process. Tension exists between the subjective experiences examined throughout this thesis and the universal (i.e., nomothetic) framework which is evolved from those examinations throughout this thesis. However, this approach allows for the mapping of multiple potential listening subjectivities to be accommodated throughout the research.

This abstract participant is imagined reacting in a manner 'typical' for that virtual experience; this means that there will always be some possibility that a specific user may or may not react in that fashion. Whilst the factors that Drever, Westerkamp, and others identify may have some effect on a specific user's understanding of sound, music, or virtual experience, the phenomenological approach offers useful advantages for the thematic work undertaken in this research. For example, a portion of this thesis examines not current and historical virtual experiences, but emerging and future virtual experiences. These experiences clearly lack actual users as they have not yet been realised. Through adopting a phenomenological approach, which isolates and discusses unchanging "essential characteristics" of the experience, we can discuss these (hypothetical) *future* users by projecting forward an expected/abstract user based on our experiences of *current* (or historic) users.

However, beyond the general issues raised by creating an abstract user – who will be auraltypical by definition – we can observe other limitations. Perhaps most importantly, this approach places additional interpretative layers between the experiential event and the analysis of that research: i.e., a research participant took part in an experiment, that experiment was analysed by other researchers, that analysis was then recontextualised and analysed in my research in aggregate to examine the underlying thematic ideas that emerge from their findings. These multi-layered processes of analysis pose challenges, as each stage of analysis abstracts the originating user experience that is under discussion. Whilst the research in this thesis works towards a framework capable of discussing these specific experiences – with a particular focus on comparative analysis as explored in Chapter 5 and 6 – an awareness of how the specifics of virtualities influence these comparisons should always be maintained.

Autoethnography features throughout this research, and online discourse and reception has been used to inform the research. However, the original plan for this research was to support the thematic distillation created through the hermeneutical/humanistic approach with empirical practice-based experiments and case studies which could investigate specific-user environments. In other terms, the original goal of this research was to develop a framework, and then demonstrate the use of this framework through direct utilisation in experimental research. These practice-based research elements proved to not be possible because of COVID-19 lockdowns and the resulting changes to research guidelines which occurred throughout 2020–2022 at the time this research took place.[9] Alternate approaches for conducting specific-user-centred research were considered when it became obvious that COVID restrictions were not likely to ease in a timeframe convenient to the doctoral research. However, it was unclear that these offered substantiative benefits. Further ethnographic analysis of online discussions surrounding virtual experience was one possibility, for example, but the terminological issues emerging in academic literature are only worsened in general discourse surrounding these issues, as enthusiasts rarely use terminology precisely or consistently. Further, the analysis of historical discussions within these communities is difficult to interpret as much context is missing: comments are deleted by privacy-conscious users, historic articles on websites and blogsites have vanished over time, videos and other peritexts are deleted by content hosts such as YouTube due to concerns about copyright or licensing, and so forth. Without sufficient context and with great ambiguity surrounding user experience, it was unclear that a COVID-friendly approach to specific-user-centric research would achieve the research goals of this project, particularly when considering that – as is discussed throughout section 1.2.2 and 3.0.2 – small differences in audio hardware can make substantial differences to the overall experience. It seemed unlikely that an experimental environment controlled sufficiently to ensure meaningful results could be maintained in remote conditions.

Alongside these concerns of creating a research environment which could output reliable results was also the consideration that much existing research utilised empirical approaches which *did* engage with specific users either quantitatively or qualitatively. Brown & Cairns' discussion of game immersion "interviewed gamers about their experiences of gaming and immersion" (Brown & Cairns, 2004, p.1297), much of the work conducted in collaboration with Slater discussed in this chapter draws upon participant experiences generated through participation in experimentation, and so on.

[9] Perhaps related to the challenges posed to research during COVID-19, researchers at Royal Holloway have informally reported difficulties in receiving timely ethical review of proposed experimental work throughout much of the course of my doctoral research. This has had direct effects on the research possibilities of this project. For example, whilst some proof-of-concept case studies did seem feasible as COVID-19 restrictions eased in 2022, this proved not to be the case: ethical review has still not been completed nearly ten months (as of time of writing in July 2023) after the details of the proposed project were submitted to the review committee and marked 'urgent' by the relevant research officers.

Later sources discussing specific experiences also draw upon specific participant experiences: e.g., Karin Bijsterveld "interviews and observes museum visitors and other users of the installation [to] acquire insights into how museum visitors and other users experience the installation" (Bijsterveld, 2015, p. 82) and much of the research cited on binaural audio in section 3.0.4 empirically and quantitatively studied user experience . Whilst no specific users were direct participants with my doctoral research, the research does make extensive use of existing specific-user-centric research nonetheless to develop and explore thematic trends in experientiality and virtual experience.

It is also important to note that conversations with those working in virtuality or adjacent fields of design were also restricted during this time, as key industry conferences and networking events were cancelled and support funding for conference travel was unavailable due to concerns surrounding COVID-19. The research has been informed by conversations with industry professionals working in these fields, such as during visits to virtual experiences, emerging from discussing my research online, and during those conferences and other events that did take place. However, as the research conducted looks ahead to new and emerging expressions of virtual experience, many of these conversations included discussions which touched on commercially sensitive areas and could not be directly published. This means that these conversations could only implicitly inform my interpretations of the publicly available documentation, rather than offering direct evidence which could be explicitly attributed.

However, where comments have been made in the public record, these are often referenced – either explicitly with citation, or implicitly as part of my broader awareness of the underlying issues. For example, in discussing the sound and music of *Assassin's Creed Odyssey* throughout sections 3.0.1–3.1.2, interviews given by Lydia Andrew (the audio director for the game), The Flight (the composition team), and other members of the creative team have been drawn upon. Whilst some concerns might be raised about the sincerity of comments given to the press during interviews regarding their creative process, it seems likely that these concerns would equally apply to testimony given for publication in a doctoral thesis.

Throughout this thesis, I have predominantly drawn upon previously published materials on sound and music within virtual experiences (and on virtuality more broadly), alongside publicly accessible materials such as reviews, interviews, conference talks and design notes, and autoethnographic research reflecting on my own engagement with virtual experience. This reflects a continued tradition of research into these matters, drawing on the same methodology as found in research conducted by other authors such as Calleja (2011), Grimshaw & Garner (2015), Kamp (2021), and Truax (2001). Though not always explicitly stated, similar methodologies are also used for many of the chapters within the *Oxford Handbook of Virtuality* (e.g., Adams, 2013; Calleja, 2013; Garner & Grimshaw, 2013; Heim, 2013; Massumi, 2013; and others) and increasingly within the field of human-computer interaction (see Gunkel, 2018; Larsen & Adu, 2021; Prpa et al., 2020; Svanæs, 2019). Whilst there are limitations which emerge from this approach, it nonetheless allows for broad thematic trends to be identified across a wide range of materials which is essential to the development of a framework intended to be used for a similarly general definition of "virtual experience" that encompasses each of the elements set out in section 0.1.0–0.1.4.

Using this phenomenological-experiential approach, the thesis aims to answer two key questions: how do sound and music function within virtual experience; and how might we approach the design, implementation, and analysis of virtual experience from a perspective with centralises user experience? In order to answer the first of these questions, the second section of this thesis examines audio – specifically sound and music – for virtual experience. Throughout Chapter 2, the construct of sound within virtual experience is examined: what purpose does sound serve in virtual experience; how might we construct soundscapes of virtual experience; and how can virtual experiences be disrupted by 'noise'? The importance of authenticity – highlighted already in section 0.3.0 – is explored further in Chapter 3, with specific emphasis on how authenticity and sound intersect within virtual experience. Chapter 4 examines music within virtual reality, and virtual experience more generally, as an example of non-diegetic sound.

The third and final section of this thesis uses the models established throughout the first two sections to examine current and future expressions of virtual experience. This will allow an answer to the second question outlined above: how might we approach the design, implementation, and analysis of virtual experience from a perspective with centralises user experience? Chapter 5 explores the inter-contextual use of the framework for designers and developers by modelling the framework's use in practice. Theoretical studies are framed through autoethnography and hermeneutical examination of key virtual experience 'texts' to demonstrate how experiential perspectives can be purposefully interrogated. Chapter 6 looks forward to next-generation and future-generation technologies, examining the potential uses of the framework beyond the analysis of existing virtual experiences. This includes the examination of virtual experiences which appear, at first, to exist outside the framework established in the first section of this thesis. Finally, Chapter 7 concludes the thesis by contemplating how a broader definition of virtuality might allow this framework to be utilised in practice across a broader range of fields, and areas for further research and development in which the framework may be useful. The framework developed through this thesis will change the way in which we approach discussion of virtual experiences of all kinds and allow us to approach the future development and design of virtual experience – particularly at the intersection of sound, music, and virtual experience – in a way which prioritises and centralises the user.

Chapter 1: Framework for Virtuality

1.0.0: Introductions Redux

Throughout the introductory chapter, three different terms for specific expressions of virtual experiences were explored: augmented reality, mixed reality, and virtual reality. The introductory chapter establishes the differences between these types of virtual experiences and provides examples of the technologies used to create virtual experiences. It also defines many of the key terms in the field such as ‘virtual reality’ and ‘immersion’ and outlined the methodological approaches which this thesis uses. Whilst the previous chapter established that differentiation cannot be solely technological and must also be experiential, this approach creates an issue in exploring approaches to virtuality throughout the remainder of this thesis: how can sound within virtual experience be discussed when the technological approaches to virtuality vary so drastically? – e.g., from head-mounted displays (HMDs) to holographic imagery to 2D monitors to solutions which do not require electronic technology at all.

The first axis of a new framework for analysing sound and music in virtual experiences will be constructed in this chapter. This axis will consider virtuality from an experiential approach: “what is the user experience within virtual reality, and how is that experience generated?” This approach allows for a cross-comparative study of virtuality as whilst technical aspects of virtuality may vary between experiences, they share fundamental similarities when viewed from an experiential perspective. To draw out these similarities, virtual experiences are framed phenomenologically as the conjunction of three symbiotic elements: hardware, software and externalities. Each experience requires each of these elements to function. Explicating these categorisations for different virtual experiences allows for a cross-examination to be conducted. A framework developed from an experiential perspective broadens the existing work on virtual reality and other expressions of virtuality. Using this framework, the chapter will then examine three different virtual experiences in detail in order to demonstrate how different virtual experiences can be situated within it. First, we will explore how someone new to virtuality might enter the created world of *Half-Life: Alyx* (Valve, 2020; hereafter, *Alyx*). This is a well-regarded example of virtual reality which uses a head-mounted device for virtual reality (HMDVR) to present the created world. This first case study will establish how the ‘typical’ virtual experience is constructed by identifying the core underlying elements that create the experience: this will formulate the axis of virtuality within the framework for sound and music in virtual experience constructed through this thesis.

Second, an exploration of the worlds created by Disney theme parks will be presented. These are full-immersion mixed realities, in which participants are surrounded by and consumed within a world that combines real and virtual elements to construct a fantasy experience. This case study examines an atypical virtual experience to see how the construction and maintenance of such a mixed reality has been sustained. Third, an augmented reality experience will be considered: *Pokémon Go* has become increasingly popular in recent years, inviting users to traverse the real world in order to make progress within a virtual world. Modes of interaction between virtual and real will be considered to construct an understanding of the necessary elements of augmented reality. Through the combination of the varied case studies across the full range of expressions of virtuality highlighted in the initial chapter of this thesis, a framework for analysing all virtual experiences will be demonstrated: this framework can be used, as discussed in Chapter 5 & 6 of this thesis, as part of the design, implementation and analysis of virtual experiences.

1.1.0: Case Study: *Half-Life: Alyx*

Placing ourselves at the beginning of a metaphorical journey into the world of *Alyx* is not as simple as it first might seem. For many users, their journey began with the release of the Valve Index in April 2019. One of the keystones of the second generation of modern commercial VR headset technology, the Valve Index launched designers Valve (a game development company known for franchises such as *Left For Dead, Portal, DOTA* and *Half-Life*, and also for their operation of Steam – a popular software distribution platform) into the VR market. The HMDVR device was a premium virtuality solution aimed at the upper end of the commercial market of HMD users. Also of interest, and released simultaneously, was Steam VR: open-source software which attempted to provide a *de facto* standard for HMDVR solutions, allowing for users of any HMDVR headset to play any HMDVR game. Prior to this stage, the divergence of HMDVR technologies caused problems when users tried to use applications designed for one VR system with a different VR system (i.e., games designed for the Oculus VR system would not function correctly on the HTC Vive VR system). Steam VR aimed to resolve some of these compatibility issues: games designed for the Index were intended to work with other HMDVR headsets. Similarly, the Index was intended to be a universally compatible device, able to play games designed for other VR systems.

The Index also came with Valve's own entrant into the VR video game market: *Half-Life: Alyx*, which was a prequel to their incredibly successful video game *Half-Life 2* (2004), a game still fondly remembered by many gamers despite being nearly twenty years old. Available only for virtual reality, and free to all Index buyers, Valve was hoping that the enthusiasm surrounding the *Half-Life* franchise[10] might encourage potential buyers to jump off the metaphorical fence and into the world of virtual reality. In order to help encourage potential buyers, *Alyx* was also marketed as one of the first 'triple-A' games released for a virtual reality platform.[11] Within the game, players take on the role of the titular character 'Alyx' who is part of the Resistance – a group of humans who are attempting to break the Earth free from the grip of the Combine (intergalactic slavers who conquered the planet in a matter of hours). This story serves as a way for the game to show an overview of many features that VR games can use. Throughout the story, players are offered a series of puzzles which must be solved using various gestural commands. A locked gate requires the player to press the correct combination of buttons to unlock the path forward: the player must reach out with their hands and input the correct combination. A door has been sabotaged: the player must trace the route of wires through the walls of the room around them until they find the sabotaged connections and repair the power grid, walking around their play space to trace wires within the virtual walls using their hands. Lasers shoot across the room: the player must align a three-dimensional orb in the space in front of them using their hands, using micro-adjustments to get the orb *just right*. These may seem like simple puzzles.

[10] *Half-Life 3* – the purported final instalment in the *Half-Life* franchise – has been keenly anticipated since the release of *Half-Life 2: Episode 2* (2007). The registering of a new *Half-Life* trademark drew almost instantaneous global attention onto Valve's upcoming projects, which helped to significantly expand the profile of the Valve Index prior to its initial release.

[11] A 'triple-A' game is an informal classification used to describe games produced and distributed by a major publisher which have a larger-than-normal budget. *Alyx*, with an estimated development budget of $30-75m, had a development budget at least one order of magnitude larger than the vast majority of virtual reality games on the market in 2020.

However, at the time of release in 2020, the ability to precisely map player gestures in this fashion was a new and exciting technology which players needed to be shown and demonstrated. Alongside these puzzle elements, the game also features a variety of path-finding challenges where players must find a way through a virtual maze full of obstacles. This requires players to become accustomed to positioning themselves within the virtual space. Further, several moments in the story feature active combat against a variety of enemies ranging from slow-moving zombies to small (and jumpy) head-eating crabs, to military forces with guns and grenades. Whilst entertaining and challenging, it can also be seen that each of these enemies requires different strategies to overcome and highlighted different aspects of the (then-new) VR technology.

We can therefore suggest that *Alyx* seems to have been intended, at least in part, to introduce new gamers to the ‘world of virtual reality’ in much the same way that games such as *Pinball*, *Minesweeper* and *Solitaire* are thought to have been intended to introduce new users to the ‘world of personal computing’ – Libby Duzan, Microsoft product manager, specified in a 1994 interview that Solitaire was intended to “soothe people intimidated by the operating system” and “taught them how to use a mouse” (Garreau, 1994). As Valve predominantly operates as a game distribution service rather than a game development company, there may also have been a desire to show off the features of the Index to attract developers to make additional virtual reality games. We can understand that the *Alyx* experience is not only intended to be entertaining, but also implicitly educational – a showcase for possible features used within virtual reality games and on the Index platform.

The question remains however: where did the construction of this experience begin, and what components constitute the overall virtual experience? Perhaps it goes without saying that the most obvious element of the *Alyx* experience is, in fact, the HMDVR. Whilst it does not have to be an Index (as noted above, *Alyx* supports all headsets which have comparable features to the Index), the game cannot be played using keyboard-and-mouse or console-style controllers: a large portion of input is gesture-based, and the specific functionality of the Index’s ‘wand’ controllers allows for very precise gesture-based interfacing[12] with the virtual world. At one stage of the game, users can even attempt to play a fully functional grand piano present within the world. However, the participant must also have a variety of other hardware: the Index is Windows-based, so a PC or Laptop with a recent Windows operating system (e.g., Windows 10) and excellent internal hardware is recommended. Some fans even recommend further purchases of external hardware. For example, the Valve Index uses off-ear mounted speakers to deliver sound to participants. Some fans suggest that closed-ear headphones to help isolate the user and sensorially immerse them within the virtual world: "Ambient noise being hearable [during a VR experience] is a pretty big reason to use different headphones. When I am in VR it would be pretty bad for my immersion to hear my PC's fans, cars passing by outside, or my girlfriend watching TV." (Reddit: "Can You Use Your Own Headphones with the Valve Index?")

Once the participant has managed to set-up the hardware, there is still further work required to enjoy the *Alyx* experience – and to construct the borders of the magic circle discussed in 0.0.1. First and foremost, the game software must be installed. Alongside the actual game application, other software may also be required, such as the Steam storefront which manages the download of the gigabytes of data involved in this process. User-facing interfaces are not the only software however: the software itself has reliance on tools to communicate between user and computer. This includes hardware 'drivers' which control the operation of a device, translating user actions into computer-readable input, and software interfaces such as DirectX which translate computer code into hardware commands that generate sound from speakers, images on a screen, and other essential elements of a virtual experience.

Some consideration should also be given to the role of the participant themselves. Within video games, participants are not only observers but also active collaborators in their entertainment experience, shaping their specific experiences through navigating implicit and explicit choices within a video game world (Bodi & Thon, 2020; S. B. Cassidy, 2011; Tanenbaum & Tanenbaum, 2010).

[12] Interfacing is used here, and elsewhere, as described by Vincent Mauger: "to enable information to be provided, accessed and applied … acting like a translator [between machine and user] making one sensible to the other." (Mauger, 2014, p.32)

Within the scope of virtual reality games such as *Alyx*, this extends further than simply making narrative choices or determining how to approach ludological challenges. In virtual reality games, participants seek to be immersed within the virtual world in many ways. Participants seek to become part of the virtual world. Whilst, in practical terms, this may not always be achieved, the philosophical precepts of immersion and becoming emerge in several ways. For example, most HMD technologies utilise participant movement as an essential interface with the virtual world: moving within the real space causes movement within a virtual space – whether that be crossing from one side of a room to another, or simply moving your hand through the air. For participants in the *Alyx* experience, Valve offers a series of modalities as to how this physical interfacing is expressed. A series of consequent experiences can be imagined. Users are allowed to determine whether they will be seated, standing or confined within a space of their own determination. Seated and standing users are provided with methods of teleporting around the game world, whilst users with sufficiently large rooms are able to simply physically walk around game spaces within the virtual world. Choices such as these control of the relationship between virtuality and reality in the hands of the user themselves, ensuring that they are confident in embarking upon the virtual reality experience in a way that maintains their comfort.

1.2.0: Towards A New Framework For Virtuality: Considering *Alyx*

The world of *Alyx* and the experiences contained therein fit within a common understanding of "virtual reality". In essence, the experience poses a 'typical' contemporary understanding of what a virtual experience might be. Whilst the definition of 'virtual experience' used within this thesis will extend beyond those specifically emerging from HMDVR technologies, *Alyx* can nonetheless be used to explore how the general understand of 'virtual experiences' can be contextualised. What are the constituent elements of *Alyx* as a virtual experience, and how might we conceptualise those elements more broadly?

We can construct an understanding of how the player experience is formed within *Alyx*. The player stands within a tangible space and uses physical objects to interact with a virtual world. Three key elements of virtual experiences therefore seem to emerge from this new understanding of *Alyx*: (1) externalities: the space within which the participants are located during the virtual experience; (2)

hardware: the tangible objects which participants will use to explore the virtual experience; and (3) software: the constituent elements of the experience which react to participant exploration. To consider these elements within a broader virtual experience, however, we must answer two questions.

First, how can each of these elements be considered within a broader context than just *Alyx*? To answer this, we must consider what the underlying aspects of each element are and how they might apply to a broader context than the 'typical' virtual reality experience. Second, how might other expressions of expanded reality situate themselves within the framework that emerges from the investigation of an archetypal virtual reality experience like *Alyx*? To answer that, we must then examine two additional case studies of expanded reality that engage with understandings of virtuality beyond simply examining 'virtual reality.'

1.2.1: Externality

The first of the three elements identified above is the virtual externality. This term is used to describe the real-actual space within which participants engage with a specific virtual experience. Any virtual experience will always be contained within a profoundly real-actual space: perhaps there are wires which cannot stretch further than a certain distance, or obstacles such as walls or tables which would prevent the participant from accessing part of a virtual space, or less obvious limitations such as the maximum distance a device can communicate across wirelessly. Whatever the underlying justification, we can recognise that the user is contained with actual-real bounds beyond which the virtual experience cannot be sustained. Within *Alyx,* this is typically a space around the size of a room: participants must remain within the detection range of the IR sensors used to capture gestures and movements.

Similarities between this concept of the virtual externality and other conceptualisations of 'virtual worlds' can be noted. Edward Castronova describes how the boundaries around virtual worlds "can be considered a shield of sorts, protecting the fantasy world from the outside world" (Castronova, 2006, p. 147). Similarly, Katie Salen & Eric Zimmerman describe a 'magic circle' of play: "a special place in time and space created by a game … the space it circumscribes is enclosed and separate from the real world” (Salen & Zimmerman, 2004, p. 95). There is an emphasis in these writings on ‘play’ – Castronova is referring to the particular circumstances in which the online virtual worlds of MMORPGs are separated from the offline actual world of reality, and Salen & Zimmerman are using notions of ‘magic circles’ to make broader points about the design of games. Within these writings, we can see direct relationships to Johann Huizinga’s description of games as being situated within a “temporary world within the ordinary world, dedicated to the performance of an act [of play]” (Huizinga, 1938, p. 10).

However, the conceptualisation of a ‘magic circle’ – and a virtual externality – – should be extended beyond games to all virtual experiences. Similarities between these concepts and concepts of immersion explored in section 0.3.0 can be seen: Laura Ermi & Frans Mäyrä’s description of “overpowering” the external world (Ermi & Mäyrä, 2011, p. 101) or Mihalyi Csikszentmihalyi’s description of the outside world “disappearing from awareness” (Csikszentmihalyi, 1991, p. 62) correlates with Salen & Zimmerman’s description of a magic circle which “separates” the virtual from the real. Following the definition of immersion proposed above in section 0.3.1 as greater acceptance of the ‘virtual’ or rejection of the ‘actual-real’, then greater immersion within a virtual experience correlates to greater power of the ‘magic circle’; users have higher perceived immersion when the boundary of the ‘magic spell’ of the virtual experience is stronger – whether that perceived immersion be generated through sensorial immersion, narrative immersion, challenge-based immersion, or any other specific form of immersion. This ‘magic circle’ extends beyond just entertainment: immersivity was identified as key to all virtual experiences, and so within ‘unplayful’ externalities we will still need a magic circle. In essence, any virtual experience must still challenge the participant’s customary sensory perception of the actual real: virtual reality supplants reality with virtuality; augmented reality builds upon reality with virtuality; mixed reality combines reality with virtuality. No matter whether we are in a state of play or not, we will always utilise some form of ‘magic circle’ within virtual experiences.

Perhaps less immediately evident within these considerations is the scope and size of these externalities. Within the concept of game or play we are used to conceptualising ‘magic circles’ as being relatively small – perhaps contained within a room (or a basement, to draw upon the recurring images of gaming emerging from the satanic panic of the 1980s). However, there is no reason why this externality cannot be large. If we accept the concept of a ‘magic circle’ as being applicable to any virtual experience, then the virtual externality encapsulated within the magic circle can be any shape or size; from small (e.g., a chair, a desk) to large (e.g., a warehouse, a parking lot, a city). The virtual experience can expand as far as the ‘rules of play’ (i.e., the rules of the virtual experience) can be maintained.

From the perspective of user experience, it is also useful to consider how the boundaries of virtuality might change the perceived world; or conversely, the ways in which the rules of play cannot be maintained, and the virtuality therefore cannot be contained. Virtual externalities are not curtailed solely by physical barriers or conceptual differentiations. Rather, from an experiential perspective, the virtual externality should be conceptualised as the limitations of our perceptions. Within any given experience, the boundaries of a user’s externality are challenged when something disrupts their sense of immersion along one of the paradigms set out in section 0.3.0–0.3.1.

For example, we could imagine that an audio-visual experience may be disrupted by olfactory or tactile inputs from the ‘actual real’ world that do not cohere to the presented virtual world. We can imagine that a participant is within a home office. During the day, they normally keep the window closed. Their experience of the virtual world is contained through rudimentary soundproofing. However, with the simple act of opening a real-actual window, some additional real-actual elements can now enter the user’s perception of the virtual experience; sound can permeate from the user’s physical surroundings into the user’s virtual experience. This lack of sensorial immersion could include sounds which conflict with narrative immersion, or which make it more difficult to navigate the virtual experience (i.e., alter challenge-based immersion). The virtual externality is altered significantly by these small changes; we could imagine that the externality is ‘larger’ as the user can perceive external sounds from a further distance.

Further demonstrating how user experiences of virtuality may be altered by the blurring of real and virtual, we can imagine a user hearing a doorbell ringing whilst engaging with a virtual experience: to the user, it may not be clear whether they hearing a real doorbell transmitted from outside the externality to within the experience or if the sound is occurring within the experience. The user may exit the experience and walk to their door, only to discover that it was a virtual doorbell ringing within the virtual world. Less humorously, a user may walk through a wall in the virtual wall where no such real actual wall exists or, conversely, walk into a real actual wall where no such wall exists in the virtual world. Efforts have been made to protect users from such interfacing issues: e.g., Oculus Quest HMDVRs now include Guardian – a software tool which allows user to define a safety zone – the user location within the physical space is tracked, and the device shows the user a barrier within the virtual world should the user approach an area that they have marked as dangerous (Byford, 2021). However, issues remain (e.g., users may make a sudden large movement through running or jumping and ergo exit the safety zone with potentially harmful speed).

Other issues emerge from where simulations cannot currently simulate events within the virtual externality. A participant's proprioceptive sense of movement may not coincide with vestibular feedback should an experience include events such as falling, flying or tilting. This can be particularly disconcerting if the externality itself has either uncontrollable movement (e.g., if using HMDs on a ship or plane) or a delay between the virtual movement and the physical movement. These areas area often of particular concern as disjuncts in vestibular feedback are known to be one cause of virtual reality sickness for some users (Chang et al., 2020).

As the user experience of the virtual externality is affected by events beyond the control of the designer, it is difficult to establish a precisely defined limit of a virtual externality. Perhaps this is most relevant when it comes to audio within virtual experiences. Audio is generally more capable of crossing boundaries and barriers – whilst light can be stopped by a simple piece of paper, preventing audio vibrations is a more involved matter. As a result, many strategies for controlling audial experiences rely on mitigation rather than isolation: whether that be physical techniques (e.g., speaker placement on VR headsets for maximal auditory screening, consideration of architectural acoustics, the redesign of internal spaces), digital techniques (e.g., noise cancellation algorithms), or design choices (e.g., including music or background sound as a noise screen to prevent audial disruption).

Whilst in conceptual terms, we can recognise the virtual externality as being infringed upon – and restricted – by the failure to isolate the virtuality from reality (i.e., external sounds disrupt the coherence of the virtual externality), we must also recognise in pragmatic terms that user experiences of virtual externalities may not always recognise these issues. It is likely that the more immersed a user feels – i.e., the less aware they are of the external world – the stronger the metaphorical walls created by the 'magic circle' of the virtual externality will seem to them. Above, greater immersion was defined as an increased acceptance of the 'un-real' or the 'virtual' and rejection or lack of perception of reality. Therefore, we could imagine that immersion is also a measure of how well the virtual externality isolates the user from reality and/or how well the virtuality is perceived as real by the user. This specific intersection is investigated further in Chapter 3.

Within most play experiences of *Alyx*, the boundaries of the virtual externality will be the perceptual limits imposed by the HMDVR. Participants are visually immersed: they cannot see the real-actual whilst within the virtual experience. They will have a separate auditory boundary dependent on the hardware they are using (i.e., closed-ear headphones will impose a small boundary to the virtual externality than open-ear headphones will) and may also be able to control their virtual externality through other hardware such as wind machines or wearable technology that controls temperature or pressure, as this sensorially immerses them within the virtual experience in other ways.

1.2.2: Hardware

Within any given externality, participants in virtual experiences must have some tools to facilitate their interaction with the experience. Within *Alyx* this is an HMDVR headset combined with peripheral devices such as hand-held controllers. Within a more general virtual experience framework, there are infinite numbers of tools that can be used for interacting with virtuality and, therefore, placed within this category of 'hardware for virtuality'. However, we can recognise three major categories of such tools, which I shall call 'Types':

1. Type 1 Hardware which receives participant input and conveys it from the user into the virtual experience (e.g., keyboards, controllers, other haptic devices).

2. Type 2 Hardware which receives software commands and outputs it from the virtual experience to the user (e.g., computer monitors, speakers)
3. Type 3 Hardware which hosts or mediates between the first two types of hardware (e.g., laptops, personal computers, consoles).

Type 1 Hardware, or user input devices, are often collectively referred to as controllers. Sheila Murphy describes controllers as a "yoke between player and game … the site of physical interactivity that links a player with his or her in-game proxy" (S. C. Murphy, 2014, pp. 19–20). Murphy draws on Andreas Gregersen & Torben Grodal (2008) to connect the physical engagement of player and controller to the psychological engagement of player and game, drawing in particular on the then-recent developments in controller technology revealed by the XBOX Kinect system, which used cameras and infrared systems to sense motion sensing and conduct gesture-recognition to capture player input. As a result of these motion-sensing systems, Murphy suggests that the player "essentially becomes the game controller [them]self" (S. C. Murphy, 2014, p. 23). The same considerations could be applied to players within virtual reality systems, many of which also use motion tracking and other sensorimeters to determine placement of the avatar.

As a result of variability in the tools of mediation, there is predictable variation in virtual experiences even between users who have otherwise seemingly identical experiences presented to them. For example, Janet Murray describes how their immersion within an arcade game "depended heavily on the heft and shape of the laser gun controller and on the way it was placed in a hip-height holster ready for quick-draw contests" (Murray, 1997, p. 146). In the same way, participants who use a steering wheel to play a driving game may find that their challenge-based immersion is increased because of the haptic similarity between their virtual experience and their expectations of what that experience of driving would be like in reality. This may relate to Mel Slater's "plausibility illusion" (see section 0.3.0), as users may find it more plausible to control a car using a steering wheel rather than a keyboard and mouse.

Much of what Murphy, and Murray, say about the relationship between player, controller, and game should also be applied to the relationship between participant, controller, and experience in a broader virtuality context. Scott Bukatman describes a conventional imagining of virtual reality control systems in some detail:

> A figure stands in a kind of high-tech bondage. Wires and cables snake from gloves and sensors to a pair of hard-crunching computers off to the side. The head is enshrouded by an elaborate apparatus that blocks the subject's eyes and ears. The figure stands in an uneasy crouch, reaching out to grasp the invisible air. This is not, however, some sensory deprivation nightmare. The subject is comfortably ensconced in virtual reality, a cybernetic paraspace comprised of real-time interactive data. … Virtual reality represents an immersion in a computer-constructed space. The character is inserted into the cybernetic field, transforming perception into subject mobility. The human is granted its own spatiality, and consequently a control, over these new vectored fields. (Bukatman, 1993, pp. 186–187 & 200–201)

Within Bukatman's framing of virtuality, users exist simultaneously in two places: an "objective" body in the real world and a "phenomenal body" in the "terminal" (i.e., virtual) reality (Bukatman, 1993, p. 188). Controllers, then, should be recognised as a crucial interactive boundary between technology and humanity: without controllers, we are unable to communicate with the virtual world. If we are unable to communicate with the virtual world, then the virtual world cannot imbue us with agency: the ability to make "meaningful" changes (Tanenbaum & Tanenbaum, 2010).

Type 2 Hardware can be viewed as almost the opposite of Type 1 Hardware. Where Type 1 Hardware converts user input to virtual impulse, Type 2 Hardware converts virtual impulse to physical event. Type 2 Hardware *broadcasts* to the user. As a result, this type of hardware has a more easily observed effect on user experiences: it controls how the user perceives the virtuality. At a fundamental level, we could imagine the effect on the user should we completely remove all speakers from a hardware set-up. Deprived of any audio whatsoever, participants would inevitably miss key elements of the virtual experience.

However, the participant experience is also mutated by less extreme changes. Even the least change may still have a significant effect on user experience. To imagine how a small change might affect the overall experience, we can compare different types of video display. In Fumio Kishino & Paul Milgram's taxonomy (1994), they define six classes of video display that might be used for virtual reality experiences. A consumer television screen, a traditional cinema screen and an IMAX theatre screen would all fall underneath "Class 1: Monitor based (non-immersive) video display" as each of them portrays a “window to another world” upon which images are electronically/digitally overlaid (Kishino & Milgram, 1994). However, despite their apparent similarity, a participant is likely to experience the least sensorial/visual immersion (i.e., the least occlusion of reality) with a consumer television screen and the most sensorial/visual immersion with an IMAX theatre screen. This is because a consumer television screen occupies a lesser proportion of the participant’s perceptual horizon – which can be broadly defined as the limits of the participant’s awareness – than an IMAX theatre screen. Other types of broadcast hardware have corollary comparisons: participants using closed-ear headphones are likely to have a more isolated virtual externality than participants using room speakers. In essence: Type 2 hardware helps define boundaries of the virtual externality set out as the first element of the virtual experience by replacing or supplementing the physical-actual for the participant.

Controllers (Type 1) and broadcasters (Type 2) require a third type of hardware to function: *platforms*. Platforms are the physical objects which act to process control input to generate broadcast output. Often referred to as “computing platforms” or “digital platforms,” they are best understood as the operating ground for the final element of virtual experiences: software. There are multiple types of platforms, which can be considered collectively (e.g., PC, XBOX, PlayStation, etc.) or individually (e.g., soundcards, videocards, etc.). As platforms function primarily as mediators between other types of hardware, a greater level of depth surrounding the particulars is not required at this stage. However, it should be noted that the platform can have extensive effects on how the user perceives the virtuality as it converts digital instructions from software into physical instructions for hardware. Changes in how these instructions are understood and generated can cause many changes: e.g., soundcards may differ significantly between platforms, causing sound to be produced at higher or lower quality irrespective of any other hardware or software limitations.

Controller hardware used for the *Alyx* experience will typically include motion-trackers used to control the avatar movement, and hand-held controllers used to interact with the virtual word. Broadcast hardware includes a HMDVR such as the Valve Index and head-mounted speakers. Additional broadcast hardware such as wearable technology designed to provide the sensation of pressure or fans connected to the virtual experience to provide the sensation of wind may also be used in some cases. These devices will all be controlled through a platform – typically a Windows PC of some kind, though *Alyx* also works on Linux operating systems. Individual devices will be connected to specific parts of the platform: e.g., HMDVRs are often connected directly to a computer's graphics card.

1.2.3: Software

So far, we have investigated two elements of virtual experiences: the externality, where the participant is located when the virtual experience takes place; the hardware, which the participant uses to interact with (and be interacted with by) the virtual experience. Software is the third and final part of the virtual experience. It would be easy to define this as the virtual world itself. However, it is more accurate to perceive software as behavioural rules: instructions as to how the virtual experience should react to specific inputs and outputs on the virtual hardware. For example, we might consider a standard expectation within many video games that the keyboard buttons "WASD" control movement: when a player presses "W," their character will move forwards within the virtual world. This is a behavioural rule: the player can control movement within virtual experiences with the "WASD" buttons.

Implicit to the understanding of software being behavioural rules presented above, however, was the concept that virtual software must therefore contain implicit behavioural *expectations.* The player presses "W" precisely because they expect that this means that their character will move forwards within the virtual world. Describing their concept of 'agency', Karen Tanenbaum & Theresa Tanenbaum establish the principle of "communicative competence": participants "need to trust that the [experience] is correctly interpreting her expressed meanings via the often-limited communication channels available to her" (Tanenbaum & Tanenbaum, 2010, p. 14). This is only possible where the participant is aware of how to engage with virtual worlds: they must be 'literate' in the rules of the world (see Aufderheide, 1993), or they must be taught what those rules are. However, the rules of virtual reality worlds are not yet completely unified in the way that rules for other expressions of virtuality might be; as discussed in our examination of *Alyx*, it seems likely that one purpose of the game was for Steam to attempt to standardise user expectations for virtual reality experiences. The *Alyx* software is the game files, downloadable from Steam, combined with the ludic knowledge of the player and the rules of interaction established by Valve for their VR titles (e.g., moving in the real-actual also moves you in the virtual world).

1.3.0: Further case studies

The underlying framework I am proposing here therefore has a tripartite structure of externality, hardware, and software. Virtual experiences are located within a real-actual space. The outer boundary of this real-actual space is determined by human perspective (i.e., the virtual externality is defined by the sensory horizon of the participant) and these boundaries may be multi-dimensional as a result of the differing technologies used to isolate the user from the real-actual (e.g., a user wearing headphones may be isolated from external auditory signals, but not from external visual signals). Virtual experiences also have systems of rules and interactions: virtual software. In order to allow participants to communicate instructions to the software and to receive sensory output from the software, they must interact through systems of hardware which are located in or adjacent to the virtual externality. As set out in the sections above, we can apply this framework to the *Alyx* experience (Table 1). We must also recognise that the three elements of virtuality within this framework are inextricably linked in a constantly shifting network of non-hierarchical inter-relationships. For example, the virtual externality is bounded by the perceptual horizons of the user: this means that the hardware choices will alter the virtual externality by altering the perceptual horizons of the user; software changes may also alter the virtual externality by making adjustments to the ‘noise floor’ of the experience and thereby increasing or decreasing perception of external noises for the user.

	Externality	Controllers	Broadcasters	Platform	Software
Definitions	The boundaries of the perceptual limits of the experience.	Hardware which receives participant input and conveys it from the user to the virtual experience	Hardware which receives software commands and outputs it from the virtual experience to the user	Hardware which mediates between controllers and broadcasters	The rules of engagement with the virtual experience; i.e., how should the virtual experience react to specific inputs/outputs?
Half-Life: Alyx (Valve, 2020)	Visual boundary defined by opacity of HMDVR: i.e., small visual boundary Auditory boundary defined by permeability of auditory hardware: e.g., closed-ear headphones provide smaller externality than open-ear headphones	Base Stations (motion-trackers); used to control avatar movement Hand-held controllers; allows players to interact with virtual world	Visual: HMDVR such as Valve Index Auditory: head-mounted speakers Other: haptic controllers (such as fans for wind conditions, or wearable technology to simulate pressure in VR)	Windows / Linux PC Some devices connected to specific parts of the platform: i.e., HMDVR connected directly to the graphics card in most cases	Game files Hardware drivers Ludic knowledge of the player

Table 1: Half-Life: Alyx *placed within the framework for virtuality established throughout this chapter*

To explore how different virtual experiences can be situated within this framework, two further case studies will now be investigated. First, the virtual experience of Disney theme parks will be examined: by re-framing Disney theme parks as a mixed reality experience, we can identify not only how the model outlined above adapts to a different scenario (and how it usefully reveals qualities of the virtual experience), but also gain a new perspective on the challenges and difficulties of mixed reality experiences. Second, the virtual experience of the *Pokémon Go* mobile game will be examined: as an augmented reality experience, a distinctly non-immersive externality is portrayed to participants. How can this non-immersive experience be represented within the framework and how does this inform broader questions of the boundaries between virtuality and reality? These two case studies will broaden the focus from virtual reality and cover the full range of expressions of virtuality set out in the initial chapter of this thesis.

1.3.1: Case Study: Disney Theme Parks

Presenting Disney theme parks (hereafter: Disney) as a virtual experience may seem counterintuitive. Participants in the Disney experience do not wear HMDs. They do not necessarily intend to step inside a technological creation. However, theme parks have attracted attention for their early adoption and later widespread use of virtual reality technology. Disneyland integrated an Aladdin-themed ride wherein participants 'rode on a magic carpet' in the early 90s (Pausch et al., 1996). Similarly, theme parks have an identifiable interest in a variety of other modern expanded reality solutions: e.g., projection based augmented reality (Mine et al., 2012), virtual reality rollercoasters (The Economist, 2017), etc. However, whilst the use of technology in theme parks is a contributing factor towards considering them as expanded reality or virtual experience, the fetishization of specifically technological worlds in discussions of expanded realities is not the only reason to consider Disney's theme parks as virtual experience.

Broader conceptualisations of the 'theme park' as a 'virtual experience' can be identified within current literature. Jean Baudrillard describes Disneyland as a "perfect model of all the entangled orders of simulacra…the religious miniaturized pleasure of real America" (Baudrillard, 1981, pp. 12–13). Similarly, Abby Waysdorf & Stijn Reijnders argue that we can “think of the entirety of the theme park experience [as an immersive experience],” using interviews to establish how “[fans] were pleased that they could have the desired embodied connection to the story-world” at the Wizarding World of Harry Potter areas created by Universal Orlando (Waysdorf & Reijnders, 2018, pp. 7–8). These descriptions help to re-frame theme parks as a *mixed reality* experience: the virtual (represented both by the technological and the phenomenological) co-exists alongside the actual (the structures and environment of the theme park space itself; the guests and visitors to the park).

We can therefore use the experience of the ‘theme park’ to contextualise the structure of mixed realities within the broader picture of expanded reality experiences. In this section, the virtual experience of Disney theme parks will be deconstructed, with a specific focus on exploring the intersection of virtual and actual – the hallmark of a specifically “mixed reality” externality. Through consideration of the factors of immersion and engagement, we can continue to develop a model for understanding expressions of virtuality in contemporary society.

In *Alyx*, users had a clearly defined moment where they entered an experience: they put on a HMDVR and pressed ‘start’ and proceeded to see the opening credits of the video game experience. The specific experience of the ‘Disney theme park’ begins when participants transition into the specific theme park virtual. Baudrillard emphasises a distinction between the “absolute solitude of the parking lot – a veritable concentration camp” and “inside, a panoply of gadgets” (Baudrillard, 1981, pp. 12–13). Alexander Moore’s framing of the journey to the ‘Magical Kingdom’ provides clearer-still delineation, observing how visitors must overcome a series of physical barriers to succeed in entering a “bounded ritual space… apart from ordinary settlements” that constitutes the Magic Kingdom:

> The first barrier is the parking ticket gate, where special booths siphon off guests desiring lodging in the hotels. After purchasing parking tickets they must then purchase admission [to the park and to various attractions]. Beyond the ticket gates, visitors enter an enormous concourse. They may go straight ahead and catch the transportation of the future, a sleek

> monorail which traverses an arc high in the air…to the Magic Kingdom itself. (Alexander Moore, 1980, p. 211)

These descriptions share commonality in presenting the fundamental gateway to the virtual experience of the theme park as the ticket barrier: participants enter the experience through the barrier (in return for a financial fee). This experience lasts until the participants exist through the gateway at the end of their stay within the park. An implication found here in these descriptions is that all contained within the ticket gates is to some extent virtual – perhaps explicated by Disney in a sign found just within the gates of the Magic Kingdom which reads: "Here you leave today and enter the world of yesterday, tomorrow and fantasy." Unlike *Alyx*, where users are constrained as a result of technological limitations to a maximum play area of around 1000 square feet – or 100 square metres – Disney theme parks frequently offer "play areas" of several hundred-thousand acres. Disney World occupies a geographical space of 250,000 acres, for example, containing miles of railways, tramways and bus routes. The space is also three-dimensional, with hidden tunnels snaking underneath dirt roads, and buildings stretching into the sky in each of the eleven distinct zones constructed across the Bay Area in Florida.

But is the implied belief presented by Alexander Moore, Baudrillard, and others accurate? Are the outer limits of the park also the limits of the virtual experience? Missing from this framing of the experience are secondary boundaries within the parks: e.g., backstage areas where actors and crew take breaks and change costumes, or maintenance areas to help keep the theme park's attractions running where technicians can be found. Accessing these areas is often done not through doors marked with immersion-breaking "staff only" signs, but rather through finding hidden entrances concealed organically within the scenery of the park. The mixed reality of Disney exists only where the virtual and the real co-exist. This co-existence is revealed in many ways, including through the many beings that inhabit the park: we can observe that the 'reality' of the cast and crew exists outside of the framing of the Disney experience – they are simultaneous real-actual humans, as well as virtual characters.[13] Beyond the boundaries of the mixed reality, real and virtual do not co-exist; the virtual is abandoned in favour of the real. The cast do not remain as their characters beyond the boundaries of the experience. Therefore, we can recognise that the scenery of the park exists not only as part of the externality of the experience, but also as one of the limitations of the experience. They provide visible, yet hidden, tangible limitations to the experiential space. We can observe spaces in which participants are intended to receive a carefully curated experience. But where users trespass beyond those spaces – if they go 'out of bounds' as some players do in technologically-facilitated experiences such as *Alyx* – they guests may stumble across elements of the virtuality out of context: cast members out of costume, or scenery whose decorations do not stand up to closer scrutiny.

Where participants remain within the curated space, however, oh – what an adventure! Susan Aronstein posits that designers of the park were challenged by Walt Disney to build narrative rides over 'cheap thrills' – something which "posed a challenge for the designers" (Aronstein, 2012, p. 63). Perhaps much of their success can be attributed to the marketing of Disneyland. Aronstein discusses how early television commercials for Disneyland promised audiences the opportunity to "enter into the adventure… fly with Peter Pan to Neverland; wander with Alice through Wonderland; ride Cinderella's pumpkin coach" (Aronstein, 2012, p. 63).

It is useful at this stage to, as with *Alyx*, recognise how these considerations may affect the typical user experience of the virtual experience of Disney. Having arrived in the car park and traversed through the gates, the participant is placed within a world that mixes the virtual and the actual. A water ride becomes an exciting pirate adventure through carefully placed animatronics and a bombastic soundtrack. The act of picking up lunch is transformed into a delightful exchange with Disney characters through careful costuming and scripting. Within a theme park experience, a participant does not need to make use of HMDs. HMDs seek to minimise the perceptive abilities of the user to engage them within a virtuality. Theme parks scale the virtuality up so that the full

[13] We can also observe that visitors import their own slice of reality into the experience, as they are entering it as themselves, absent any character or persona.

perceptive abilities of the user are still contained within the virtuality. Both users of HMDs and participants in a theme park experience cannot easily perceive beyond the framing of the intended externality. They are immersed – they cannot engage (solely) with actuality.

Co-existence with other participants is necessary within theme park experiences, which contrasts with the individualism of the *Alyx* experience: outside of rare (almost-unique) occasions, the participatory experience of a theme park is typically one shared with many other participants. Disney World has an average of 250,000 guests each day sharing a synchronous experience within a shared space. Whilst this may be an accepted part of tourist experiences, it should be emphasised that synchronous and shared experiences are known to be problematic for immersion: participants within shared experiences have the individual agency to disrupt the virtual experience for other participants. For example, William Cheng recounts an example of online 'trolls' deliberately disrupting a shared synchronous digital space in the game *Lord of the Rings Online*:

> Sure enough, at around 10:40 that night, Jimbrosil and five of his friends marched into the Auction Hall with matching uniforms. At the time, there were about a dozen other players in the hall… One can imagine their surprise when the room suddenly became filled with the sounds of six bagpipers performing six different [songs]. Jimbrosil and his kinmates had positioned themselves such that very corner of the hall fell within the musical range of at least one performer. Anyone standing near the center of the room would have been able to hear all six musicians…most players resolved that it would be easier to flee the scene altogether. Within minutes, all but one or two listeners vacated the space" (Cheng, 2014a, pp. 134–135).

This can be applied beyond multiplayer video games: simply co-existing within a shared virtual experience is enough to pose potential issues as participants may not have the same intentions within, or understanding of, the experience. Much the same issue can also be identified in shared physical spaces such as that hosting the Disney experience: participants may have conflicting goals or may conflict as a result of their co-existence. This can disrupt their immersion and/or their engagement with the experience.

Perhaps the simplest demonstration is the presence of queues throughout the park – an inevitability whilst attempting to engage with popular narratives found throughout Disney theme parks. Participants are prevented from engaging in the Disney experience because the experience has a limited capacity for delivery. As the density of participants increases, the risk of a deleterious experience for users – often referred to as "crowding" – increases.[14] By examining an aspect of the mixed reality where the experience is under the most strain, we can understand better the nature of mixed reality externalities.

Theme parks offer a conglomerate experience of rides, attractions and shows. This means it can be hard for theme park managers to predict participant demand for specific elements of the overall experience. Further, some research suggests that a low density of users may not offer the best experience to participants. For example, Veronica Thomas & Christina Saenger identify that "retailers may benefit from encouraging crowding" in some circumstances as "the human desire to avoid isolation may make crowding more desirable [than under-crowding]" (Thomas & Saenger, 2020, p. 520).[15] Crowding also seems to implicitly emphasise the popularity of an experience which, in turn, can have a positive effect on user experiences (Milman et al., 2020). Succinctly, it seems that to effectively manage "crowding" in theme park experiences, a balance must be found between positive effects (e.g., seeing others enjoying a shared experience) and negative effects (e.g., increased wait times) whilst also avoiding the appearance of 'under-crowding'.

Disney have found many ways to mitigate potential for disharmonious conflict in their theme park experiences, using strategies that Baudrillard describes as "a panoply of gadgets [which] magnetizes the crowd in directed flows" (Baudrillard, 1981, p.13). These strategies are known to include capacity-control policies, demand management strategies (e.g., ticket pricing structures, season passes), queue-management strategies (e.g., queue-skipping tickets, priority queues, virtual

[14] An important factor in crowding is the *user perception* of how other individuals sharing their co-existent space have negatively affected their *individual* experience. It is too tangential to explore in this thesis, but crowding is not simply a matter of calculating density of persons present within a space, but rather a matter of calculating the extent to which the density of persons within a space has had (or at least, is perceived to have had) a deleterious effect. Difficulties in assessing this – in particular, the extent to which crowding interferes with participant experiences – remains a topic of much research. See Eroglu & Harrel (1986), Klanjšček et al. (2018), R. Manning et al. (1999), Vaske & Shelby (2008).

[15] COVID-19 resulted in "under-crowding" at many tourist locations as a result of global social distancing policies. The total impact is not yet known, but it is expected that COVID-19 will drastically change participant perceptions of crowding at tourist locations (i.e., participants will require lower population density in order to avoid feeling 'crowded'). See Lim (2021) and Milman et al. (2020).

queues) and interactive queuing experiences (Milman, 2019). These strategies clearly point towards a conscious curation of the collaborative atmosphere of the park.

In the introductory chapter, mixed reality was identified as an expanded reality which focuses on the explicit combination of the real-virtual dyad, where reality and virtuality are symbiotic. Considering Disney's strategies for managing crowding offers can help delineate specific elements of the experience and identify how mixed reality is constructed. The prosaicism of queueing is mutated by interactive experiences: perhaps a cast member, dressed as a Disney princess, wanders near the waiting area, distracting those within from the mundanity. Here, we can see a delineation between environment, hardware, and software. The experience takes place within a bounded environmental space – the queue for a ride. The participant interacts through hardware – the cast members, and perhaps other interactive objects. And that interaction is guided through a series of rules – perhaps there is a script, or at least some staff policies that the staff member must follow.

Through this kind of reframing of the Disney experience, we can "think of the entirety of the theme park experience [as an immersive experience]" (Waysdorf & Reijnders, 2018, pp. 7–8). Similarities between going 'out of bounds' in *Alyx* and going 'out of bounds' at Disney have already been playfully drawn. But perhaps it should be emphasised in more detail how the two experiences relate to each other more broadly: *Alyx* uses level design to carefully guide the players from one curated experience to another, ensuring that they never feel lost. So too does Disney – parks are designed to always highlight the next 'adventure' to embark on: the exit of one ride leading smoothly to the entrance of another area. However, a key difference between virtual reality and mixed reality can be found here: within virtual reality, the boundaries are established through the virtual world. They do not always exist as barriers within the physical world; the edge of the externality might be discovered when a player moves beyond the range of the motion-tracking technology, as there are no walls in the real-actual. Conversely, Disney's boundaries *must* exist within the real-actual in order to prevent participants from accessing spaces beyond the virtual experience. Despite these differences, both must conceal their boundaries to some extent: both *Alyx* and Disney make clever use of scenery and worldbuilding to make it seem as if the virtual experience extends far beyond the spaces which are accessible to participants.

Differences are perhaps clearer in the hardware of the experiences. *Alyx* uses predominantly technological hardware – e.g., HMDVR, motion trackers, computer. It is perhaps too simplistic – and too contrived – to be entirely precise, but we can envisage part of the hardware of the Disney experience to be the cast and crew hired by Disney to portray their virtual experience in much the same way that Murphy suggests that the player can "become the game controller [them]self" (S.C. Murphy, 2014, p.23). 'Programmed' through careful training and rigorous scripts, staff can be seen as advanced animatronic (though non-technological) hardware designed to build immersion and enjoyment within the virtual experience. For Disney, staff are almost interchangeable: cast members who cause 'glitches' in the virtual world – e.g., through failing to engage or by breaking the fourth wall and revealing the deception – are replaced in much the same way that we might replace a broken controller for the Valve Index. Staff within Disney are subordinated to 'The Mouse' – they are provided with character sheets, park policies, handbooks for expected situations and stories of how unexpected situations were handled. Through these peritextual resources, Disney is hosted on a platform of people to be controlled through human interaction – and it is the improvisatory freedom offered to staff that maintains the virtual experience of Disney: actors are free to creatively improvise around prepared themes to avoid canned, 'inauthentic' experiences. Perhaps most revelatory of this freedom is the permission given to all staff to create "magical moments" for participants: whether that is replacing lost ice-creams or encouraging participation in performances.[16]

This improvisatory exploration is clearly not possible within the virtual technologies currently available: computer software (unlike the 'human software' used by Disney) is not capable of cohesively generating an improvisatory and interactive experience on demand given almost any input.

[16] One example of how 'magical moments' occur can be found in Hayes (2019): a girl wearing a Disney princess costume hears some music and starts dancing. The bandmaster notices her dancing and guides her into the spotlight. The girl dances until the end of the song and is then congratulated by the band and crowd. To quote: "I was just in my world and I was just living my moment. And when the moment was over, I was just like – did that just happen?" (Hayes, 2019)

Attempts to do so have proven generally unsuccessful, though some case studies into procedural software are explored in Chapter 4. But by examining how Disney 'programs' their hardware, we could begin to anticipate how we will need to program hardware for virtual reality if we wish to provide a commensurately individual experience using only computer hardware. If we want *Alyx* to have a similarly individual experience, then we need to identify methods through which the game is able to react and interact with players in a quasi-improvisatory fashion.

To some extent, exploring these differences also reflect on a given similarity between the experience of Disney and the experience of *Alyx*. *Alyx* is an individual experience, designed for enjoyment by a single player; Disney is a collective experience, designed for enjoyment by many participants simultaneously. However, both aim to create a world where the user is effectively immersed. We have already recognised that Disney immerses the user visually by creating a large-scale virtuality where the boundaries of reality are beyond (or hidden from) user perception while *Alyx* obscures visual reality through use of a HMDVR. The virtuality is also scaled audially: *Alyx* locates speakers close to the ears, helping to ensure that noises cannot trespass through the noise floor into the user's awareness; Disney enforces a boundary around its theme parks in three-dimensions to limit any potential noises from entering the virtual experience: e.g., planes are usually not permitted to fly lower than 3,000 feet over Disney World (Dapcevich, 2021) which prevents the sound of planes from disrupting the virtual experience. Disney also bombards the participants within the space with loud music at almost all times in an oversized example of Ermi & Mäyrä's suggestion that sound could be used to "overpower the sensory information coming from the real world" (Ermi & Mäyrä, 2011, p. 101). Whilst the mechanisms of the experiences differ significantly, the participant perspectives have many similarities. To contrast with these two immersive experiences, a third case study will now be considered: *Pokémon Go*. Unlike *Alyx* or Disney, *Pokémon Go* exists outside of a clearly delineated bounded space: how might the model used to explore these other expressions of virtuality provide insights into *Pokémon Go*?

1.3.2: Case Study: Pokémon Go

Before discussing how the virtual experience fits within the framework established in this chapter, it is important to explore what makes *Pokémon Go* different to other iterations of the *Pokémon* franchise and different to the other case studies in this chapter. In 2016, video game development company Niantic released what has now become one of the best-known – and probably most successful – smartphone games: *Pokémon Go*. The game itself offered a familiar ludic formula to fans of the Pokémon franchise. Players explore the game world discovering creatures – Pokémon – which can be tamed for capture or fought for rewards. Using captured Pokémon, players compete in tournaments to control gyms, with the eventual aim of becoming – as the title song of the original television series suggests – the "very best" Pokémon trainer. Despite these similarities to other video game instalments in the Pokémon franchise, *Pokémon Go* differed in one very important aspect. Rather than exploring an entirely virtualised world with fictional characters, *Pokémon Go* offered players the chance to physically explore the *real* world. Pokémon are placed on a map of their virtual surroundings, and players must physically relocate themselves within the real world to attempt to capture Pokémon. Similarly, locations within the *Pokémon Go* universe typically have real life equivalents. Schools, churches, coffee shops, even actual gyms – each of these has become a virtual battleground fought over by those occupying the nearby real space.

Pokémon Go portrays an augmented reality. Participating in this virtual world does not require an HMD. Unlike virtual reality, augmented reality has a dependence on reality. Without being able to perceive the actual world, participants would not be able to make game actions. Rather, participants must download and install an application onto their smartphone device. This application provides a reframing of actuality: a map of the local environment is provided with various points of interest to the user highlighted – specific creatures to capture, specific locations to travel to, other users to trade or fight (Figure 10). But, unlike mixed reality, augmented reality is not a cohabitation of the real and the virtual. Pokémon Go imposes a *new* reality – an alternate reality – onto the world. When viewed through the frame of the game, reality has been supplanted. The game therefore differs from both *Alyx* – which uses an HMD to block out the external world – and Disney – which scales the virtuality to completely subsume participant perceptions. Rather than attempting to immerse the user within a new virtuality, *Pokémon Go* instead accepts that the virtual experience will be perpetually invaded by reminders and objects of the real; the 'magic circle' will always contain things not conceited as within the virtual experience.

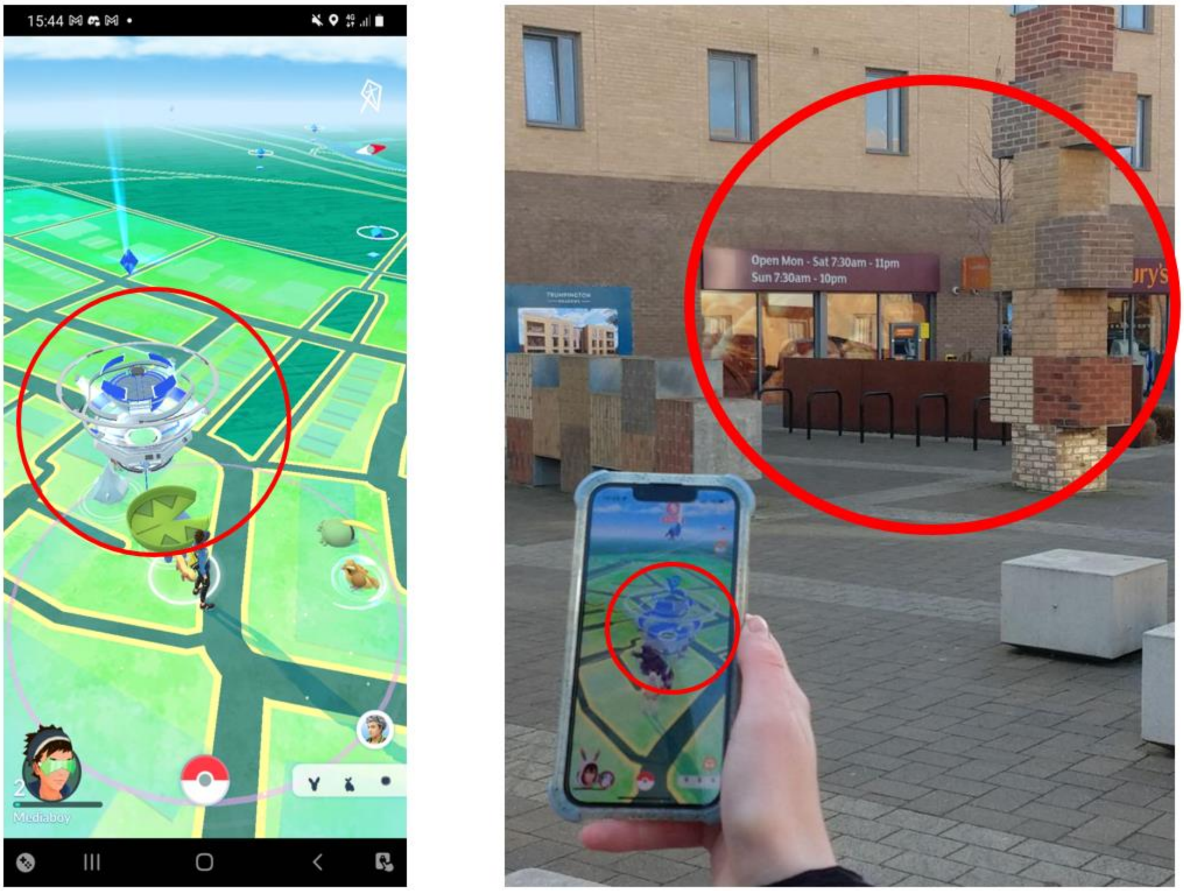

Figure 10: A local supermarket is transformed into an in-game 'Gym' – an in-game location where players can battle – within the game. The real-actual location has been augmented with additional meaning within the virtual experience.

What then, should we consider to be the hardware within this space? Evidently, excessive hardware would be problematic: users must be conscious of the fact that they are transgressing real spaces whilst participating in a virtual experience. If we were to imagine a version of *Pokémon Go* that used a HMD, then we could imagine the resultant chaos. Isolated visually from reality, participants in this hypothetical HMD-*Pokémon Go* experience would encounter reality predominantly through collisions with non-participants and physical objects.[17] In other words: ouch. Out of a desire to avoid this result, *Pokémon Go* uses minimal external hardware: players interact with the world predominantly through their smartphone, which serves both as an input device (i.e., through touch-based gesture commands) and an output device (e.g., to display where Pokémon might be found). However, *Pokémon Go* introduces a physicality to the engagement with the virtuality: players must *physically* transverse the *virtual* world (which is overlaid on the real world) in order to progress within the game. The game also offers different opportunities for different modes of travel: walking, cycling and riding in a car each offer variable paths to the user. In this way, we can consider the user's locomotive device to be an essential part of hardware for the game – whether that be their legs, their bicycle or their car.[18]

The examination of *Pokémon Go* presented in this chapter demonstrates how all expressions of virtuality established in the initial chapter can be described and discussed through the same experiential framework: we can recognise that the experience is formulated of the same three elements of externality (e.g., the use of the external reality as the play-space), hardware (e.g., the smartphone's touchscreen), and software (e.g., the mapping of the user's movement in the external reality to the internal virtual world). We have also discussed the three types of virtuality discussed in sections 0.1.0–0.1.4: virtual reality through *Half-Life*: *Alyx*; mixed reality through the examination of a Disney theme park; and augmented reality through *Pokémon Go*. We can place these virtual experiences in dialogue with each other by showing how the elements of the experience discussed throughout this chapter are situated within the overall framework (Table 2). The flexibility of the constructed framework allows for useful intermedial comparisons between differing expressions of

[17] Inattentiveness as a result of playing *Pokémon Go* has already been estimated to have cost billions of dollars in America alone (Faccio & McConnell, 2020) due to inattentive drivers causing vehicular damage, injuries and casualties whilst playing the game. This hypothetical situation imagines a more holistic immersion within the game – though it is hard to determine whether this would decrease damages (as players would be less likely to attempt to multi-task) or increase damages (as players would be *even less able* to multi-task if it was attempted).
[18] It should be noted that one popular way of playing *Pokémon Go* is to install an emulator that tricks the game into believing that you are travelling using a mode of travel that the game recognises, thereby avoiding the whole exercise. The lazy ingenuity of these players does not detract from the design intentions of the developers, however.

virtuality, as will be explored throughout the remainder of this thesis, with a specific focus on sound and music: phenomena which extend across all three elements of virtual experience.

Experience	**Hardware**	**Software**	**Externality**
Half-Life: Alyx (Valve, 2020)	Broadcaster: Virtual reality headset Controller: handheld controllers; VR Base Stations Platform: PC / Laptop	*Half-Life: Alyx* game files Drivers for hardware	Room-scale space, bounded by base stations (in virtual space) and walls/objects (in real-actual space)
Disney theme park	Broadcaster: Speakers/screens et al. within the park; cast members Controller: the user Platform: the theme park	Disney theme park materials (e.g., background music loops) Character sheets, scripts, policies, rules, etc.	Large geographic area of hundreds of thousands of acres; bounded by real-actual objects to limit participant perception horizon (e.g., walls, trees); some legal restrictions placed around parks also (e.g., flight restrictions)
Pokémon Go (Niantic, 2016)	Broadcaster: Smartphone display/speakers Controller: Smartphone touchscreen, GPS receiver, means of locomotion (e.g., feet, bike, car, train, bus) Platform: Internal smartphone hardware	*Pokémon Go* game files	Unbounded international space: users can travel (almost) anywhere on Earth and have the *Pokémon Go* experience.

Table 2: Comparison between three different virtual experiences using framework established throughout Chapter 1

1.3.3: Towards Hearing Virtuality

Throughout these three case studies in this chapter, the first axis of a framework for examining virtualities was developed. This axis of the framework situates virtuality as being created within a tripartite structure with three elements: externality, hardware, and software. The virtual externality is the real-actual space in which participants of a virtual experience are physically immersed. The virtuality externality is rationalised and limited through the boundaries of the user's perception of the real-actual and their willingness accept the virtual as their new real-actual. Virtuality hardware exists within three specific categories. Type 1 Hardware ('controllers') receives participant input and conveys it to the virtual world (e.g., keyboards). Type 2 Hardware ('broadcasters') outputs information to the user (e.g., speakers, monitors). Type 3 Hardware ('platforms') mediate between Type 1 and Type 2 Hardware (e.g., an entire PC, or specific items such as a soundcard). Virtuality software is the behavioural rules that determine how the virtual experience will react to input and when it should provide output.

This tripartite structure allows for considerations of how virtuality is created and sustained across different expressions of virtuality: e.g., across *virtual*, *mixed*, and *augmented* realities. By positing these considerations with this new model, we can understand the challenges posed in the design and implementation of a virtual reality. We can draw similarities between different approaches to both designing and engaging with virtual experiences such as the comparison between *Alyx* and Disney at the end of section 1.3.1. These comparisons can help inform the design, implementation, and analysis of other virtual experiences as highlighted throughout Chapter 5 & 6. However, virtuality is only one of the axes which must be considered in this thesis, which focuses on the intersection of virtuality with sound and music. Therefore, alongside the axis of virtuality explored in this chapter, we must now consider a second axis of audio. Sound forms a significant part in user participation within virtual experiences. Chapter 2 establishes how sound is used within virtual experience to communicate and convey information: what is sound in virtual experiences, and how (and why) might we approach the design and implementation of sound from an experiential perspective?

Chapter 2: Sound

2.0.0: What is Sound?

In the previous chapter, we theorized virtual experiences as being constructed of three elements – hardware, software, and externality. As part of this theorisation, we examined Disney as an example of a mixed reality experience in which the technological and phenomenological virtual co-existed alongside the actual structures and environment of the theme park(s) and the actuality of the guests and visitors within the park. The Disney theme park experience was used to envisage the structure of virtual experiences within a broader context, with consideration given to how the three elements of the framework for virtual experience are expressed within Disney theme park(s). A comparison of Disney theme parks and *Half-Life: Alyx* was conducted, to demonstrate how the framework allows us to place elements of different (and differing) virtual experiences into context with each other.

This thesis phenomenologically constructs a framework for analysing, designing, and implementing sound and music with virtual experience. This approach will generate a framework which allows for the consideration of both real historical users and imagined abstract users (see section 0.4.0). To be able to examine sound and music as part of this framework, we must understand how sound and music function within virtual experiences. This chapter engages with this second axis of the overall research project by examining sound within virtual experiences.[19] Since our concern is with experiential perspectives, we must first consider what is meant by ‘sound’ and how we can come to understand sound as a perceptual event. We must then understand how we can record our perceptions and experiences of sound such that we can recreate an auditory experience at a later date, or for a new audience. As a framework for recording these perceptions and experiences of sound, we will consider models of ‘soundscapes’ such as those suggested by R. Murray Schafer (1977b). To use the conceptual model of soundscapes as part of the design process for virtual experiences, we must then examine the process of ‘sound design’ with a particular focus on how we might be able to create

[19] Chapter 4 will explore *music* in greater detail to complete the examination of “sound and music within virtual experience” that is the subject of this thesis.

soundscapes for virtual experience participants – whether those experiences are intended to recreate a real-world experience and or intended to create something entirely new within a virtual space. We will then use the understanding of sound and sound design developed throughout this chapter to examine how we might construct sonic experience of the Disney theme park mixed reality experience and compare these hypothetical constructs to a pre-existing virtual recreation of a Disney theme park: the video game *Disneyland Adventures* (2011). The comparison between the computer game and the mixed reality environment of a Disney theme park is used to establish how the sound might be transformed through the process of reproducing a virtuality in a different context. By comparing the two, we can gain insight into how sound functions within virtual experience, and how we might approach the design and implementation of sound within virtual experience from a perspective that centralises user perspectives.

2.1.0: Sound

Sound is often considered to be simultaneously two entirely separate phenomenon: (a) physical vibrations travelling through a medium, and (b) the reception and interpretation of those same vibrations – often specifically by the human brain. As Mark Grimshaw-Aagaard observes, such definitions can lead to the "amusing but absurd statement that sound is evoked by sound" (Grimshaw-Aagaard, 2021, p. 271). It is this fundamentally absurd duality that leads to Robert Pasnau's declaration that the contemporary understanding of sound is "incoherent":

> Our standard view about sound is incoherent. On the one hand we suppose that sound is a quality, not of the object that makes the sound, but of the surrounding medium… On the other hand, we suppose that sound is the object of hearing…Yet these two assumptions cannot both be right – not unless we wish to concede that hearing is illusory, and we do not listen to the objects that make sounds. (Pasnau, 1999, p. 309)

Pasnau takes issue with the concept of sound as a vibration which must travel to the perceiver; he argues that sound is perceived to come from the objects that make them: "Sounds which are caused at a distance seem to be at a distance; they do not seem to be coming towards you, unless that which makes the sound is in fact coming towards you." Pasnau identifies that sound is not a transient object conveyed within the air around us. Rather, it is "among the various sensible properties of objects; among colour, shape, and size" (Pasnau, 1999, p. 324). We can correlate this to Barry Truax's later notion of communicational sound: "the exchange of information, rather than the transfer of energy. … [listening to sound] can be defined simply as the processing of sonic information that is usable and potentially meaningful to the brain" (Truax, 2001, p.11).

Mark Grimshaw & Tom A. Garner identify a selection of qualities that contribute to sound, which they define as "an emergent perception arising primarily in the auditory cortex and that is formed through spatio-temporal processes in an embodied system" (Grimshaw & Garner, 2015, p. 1). This theory of "emergent perception" describes sound as a "creative act within our mind … dependent wholly or in part on cognition and emotion … in the here and now" (Grimshaw & Garner, 2015, pp. 1–2) which leads to sound being actualised as a perceptive event through the "sensuous and nonsensuous elements that are combined in a perceptual immediacy and simultaneity to form the present occasion" (Grimshaw & Garner, 2015, p. 165).[20] Grimshaw & Garner further identify many other elements of the "sonic aggregate" which lead to "sound perception" – including cognitive and perceptual aspects of sound, such as where someone may "focus their auditory attention" to allow sound potential to be "actualise[d] by our conscious decision [to listen]" (Grimshaw & Garner, 2015, pp. 170–171).[21] We can recognise that implicit within Garner, Grimshaw, Pasnau, and Truax's work on sound is the rationalization that sound, if it is to have any meaning, must fundamentally be perceived by a listener – i.e., someone that will translate the physical vibrations into purposeful information. This is common across much other literature, which positions "sound" as a discrete

[20] The terms sensuous and nonsensuous are drawn from Alfred Whitehead's ontology of the unconscious. Sensuous elements are those which are perceived by the sensory systems (e.g,. the ears) whilst the nonsensuous is – as Grimshaw & Garner state – the "mode of perception of our past that includes uncognitive memory and experience" (Grimshaw & Garner, 2015, p.165). Similarities between the nonsensuous – as uncognitive memory and experience – and the concepts which underlie 'media literacy' explored in section 2.2.1. should be noted.

[21] Grimshaw & Garner frame this as a debate between deterministic and non-deterministic philosophy and ultimately conclude that the actualisation of sound as a perceptive event combines non-deterministic and deterministic processes as a result of the changing balance of consciously cognitive and uncognitive (alongside, presumably, unconsciously cognitive) decisions made by the listener. The implications of this are relatively unimportant insofar as the content of this thesis, as it matters here only that sound is emergent, not whether the process of emergence is inherently deterministic.

entity that exists only as a perceptible event. For example, Rick Altman describes sound as something which requires an audience: “Just as sound events remain only hypothetical sound sources until they are actualized by a hearer, so the playing of a sound record takes on meaning only in the presence of an audience” (Altman, 1992, p. 28).

To some extent, this conceptualisation of sound is easily understood: sound provides information about what objects are around us – e.g., we hear bird song, therefore there are birds – in what Grimshaw & Garner refer to as “associative meanings” (Grimshaw & Garner, 2015, p. 175). However, alongside information provided directly through semiosis and associations, sound also provides transitive information (i.e., information that emerges as a result of the sound transiting through a spatial environment before being perceived). Sounds we hear can provide information about the world around us which is not directly communicated through the associative power of that sound. Altman describes our ears as being “surprisingly attentive… even before we look out the window, our ears tell us that it has snowed during the night… They help us to get the ‘feel’ of every room we enter without ever touching any of the room’s surfaces… The fact that we come equipped with two functioning ears each makes still more information available to use... our own personal sonar gives us valuable information about our soundscape” (Altman, 1992, pp. 22–23). Pontus Larsson et al. identify that whilst “the auditory system is less accurate than the visual one in terms of spatial resolution” it can nonetheless “provide us with spatial cues from the entire surrounding space … allowing us to both locate objects and to feel surrounded by a spatial scene” (Larsson et al., 2010, p. 5) as a result of the physical changes that sound undergoes as it transits through an environment. Similarly, Grimshaw & Garner emphasise sound as having “exosonic” components which emerge because of these spatial elements of sound:

> Because air is a material medium with density, viscosity, elasticity, humidity, level of pollution, and temperature, it transforms the sound waves that propagate in it through geometric and absorbing processes. In the main, this manifests itself as decreasing amplitude of the sound wave with increasing distance from the sound wave source but there are also frequency effects…. Perhaps most important is the effect [the transformational properties of the environment] has on the sensation of a sound wave is the presence of other materials and objects in the space in which the sound wave propagates … In typical listening conditions,

> such as a living room or a concert hall, there are many surfaces of differing sizes and shapes each of which absorb and reflect acoustical energy. The cumulative effect, and therefore specific effect on sensation, depends on the location of the listener in the acoustic space. (Grimshaw & Garner, 2015, pp. 166–168)

There are several well-known examples of the kind of information that can be provided to a listener through these effects. Slight delays and modulations resulting from sound travelling around or through our head can provide us with information about the location of an object. Doppler shift tells us whether an object is travelling towards or away from us by compressing or elongating sound waves. Acoustic properties of sound also convey information about the spaces that we are within. For example, a longer sound decay (or "reverb") implies a larger space to our understanding; a small classroom might have a reverberation time of less than a second, whilst a large cathedral might have a reverberation time of several seconds. Different materials also have different acoustic properties and reflect and absorb different frequencies of sound. This can allow us to make assumptions about the composition of surfaces within a room. In each of these examples, sonic differences are perceivable to our brain, which collates them (typically subconsciously) into information about the world around us. We can also recognise that sound must therefore always have a physical presence: there is something – by necessity – whenever and wherever sound is. The differences between physical components varies dependent on the spatial and material properties of the component, and uniquely alters the sound in each case. We can also understand that as sound has a physical presence (being constructed, at least in part, of physical motion), it therefore has a spatial location; in some part, therefore, sound will always be, and contain, spatial and/or environmental information.

We can therefore see sound as being the result of conscious and subconscious processes of interpretation and mediation which happens simultaneous to the reception of physical vibrations. Information is gained through sound through both recognition of signals which are similar to those emerging from objects we have encountered before, and through recognition of various sound effects which emerge dynamically from the modulations occurring as physical vibrations traverse through the environment around us. A sound signal is neither the physical vibration, nor the interpretation of that sound signal, but rather the experiential combination of both.

2.1.1: Schafer and Soundscapes

Practical issues with the theoretical definition of sound highlighted in section 2.1.0 lie in the murky waters of reality: humans do not receive sound signals individually, but rather all together. They experience it, as Westerkamp suggests (building on Michael Stocker), as a "continuous now" (Westerkamp, 2019, p.45). Whilst a specific sound signal can be imagined as an individual physical vibration, the human perception of sound is as a series of vibrations which arrive at our ear over time as the result of many collisions between many individual sound waves (i.e., vibrations) emerging from objects around us simultaneously. That is to say, our experience of sound happens all at once, with every sound source co-mingling within our perceptual horizons. Deciphering this sound is both an active and a passive process. As Martine Huvenne describes, "when I have the impression that I am just moved by a sound in a passive way, there is also an active component to the act of listening. I am not confused by different vibrations—instead, I actively distinguish noise from voices or music. Listening from a first-person perspective, I make these choices in the act of listening itself, without reflecting on the listening or on the sound" (Huvenne, 2016, p.131). To develop our understanding of sound, and understand how to use sound within virtual experiences, we must consider how humans (sub)consciously categorise and perceive sounds around them: for example, how might specific sounds arise for a listener within their personal (i.e., first-person) perspective of that environment?

Many visualisations of sound can be found in common usage throughout different fields concerned with examining and using sound. For example, musicians may be familiar with musical scores, sound engineers may be familiar with waveforms, and linguists may be familiar with spectrograms. Each of these visualisations captures some elements of the phenomenological 'sound'. Simultaneously, however, each of these visualisations omits key parts of the phenomena. They only describe some properties of the 'sound' – in essence, falling into some of the traps identified explored earlier in this chapter by representing only specific aspects of the sound source (or qualities of the physical vibrations of the sound wave), rather than the overall experience of hearing a 'sound'.

A more holistic approach used for understanding how humans perceive sound around them is Schafer's model of the 'soundscape'. The soundscape can be broadly thought of as the acoustic environment surrounding a listener. Schafer emphasises a subjective experience, stating that a description of sound is "trustworthy only when writing about sounds directly experienced and intimately known" (Schafer, 1977b, p. 8) and emphasises the positionality of the observer throughout his classification of physical characteristics of sound: e.g., "what is the estimated distance of this sound from the observer?" (Schafer, 1977b, pp. 134–136). Rather like how a landscape painting portrays a specific visual experience of a specific location at a specific time as seen by a specific subject, a soundscape (in Schafer's use) portrays a specific audial experience of a specific location at a specific time as heard by a specific listener/subject. These soundscapes can be presented in many formats, such as written prose or as diagrams or maps.[22]

Schafer identifies three key elements of a soundscape that should be explored within such a description to ensure a holistic portrayal of what can be heard – *keynote sounds*, *signals* and *soundmarks*. *Keynote sounds* are those which identify the character of the location: "those created by its geography and climate" (Schafer, 1977b, p. 9). Schafer suggests that such sounds are often heard *subconsciously* – the listener is not consciously aware that they are there, although they may notice the *lack* of those sounds when they cease. Conversely, *signal sounds* are consciously heard; they are "foreground sounds" – particularly those which "must" be listened to. *Soundmarks* exist somewhere between the two – sounds which are the unique and distinct identifiers of a given area or community (Schafer, 1977b, pp. 9–10). By combining these different sounds, an understanding of what an experience sounds like can be communicated to someone else.

Schafer's model of soundscapes has remained popular since its proposal in 1977. In most cases, the audial experience of a listener can be understood through the three sound types that Schafer

[22] Schafer's *Five Village Soundscapes* (1977a) uses a number of diagrams, maps, and photographs to represent sonic elements within European village life, but these approaches have also been used to explore virtual environments: for example, Hambleton (2020) uses a similar approach to examine the soundscape of the virtual world of *Leaving Lyndow* (Eastshade Studios, 2017).

proposes. The model encourages analysts and sound designers to consider not only the sounds which immediately catch your attention – i.e., signal / foreground sounds – but also sounds which are omnipresent and, therefore, heard subconsciously – i.e., keynote / ambient sounds. By combining these elements, Schafer's model allows for a relatively holistic consideration of the sounds of a given environment.

2.1.2: A Soundscape of Disney

Using Schafer's model, we can now consider a hypothetical visit to a Disney theme park. As explored in the previous chapter, Disney theme parks can be seen as a complex – or messy – mixed reality experience in which many different expressions of virtuality conflict within a space which also contains many different real objects. Disney theme parks posit a conflict between the physical and the meta-physical: e.g., the real visitors and the virtual characters; the real physical sets constructed within the park and the virtual imagined worlds constructed within the Disney fantasy. This same conflict can also be observed within a soundscape of Disney. Within a visit to a Disney theme park, a typical experience of the sound heard by a visitor might be described something like this:

> The screams of children fill the air, their excitement and energy only matched by the enthusiasm with which actors deliver their lines from the stages distributed around the scene. A parade marches down the street, their big band sound drawing attention to the on-going festivities. As they pass by, recorded music begins to once again emerge from speakers nestled in the scenery. From similar speakers, placed strategically through each of the rides, come the voices of actors far too expensive to stand on a Disney stage – Johnny Depp brings Jack Sparrow to life, hinting at connections from the fantastical picture to the real world, delivering on the promise that the pilgrimage to this promised land proffered. Hidden underneath these swirling sonorities are hushed conversations between parents – is it time for ice cream? – and the whirring hums of machinery pounding away to deliver each ride. Water trickles beneath boats, brakes grind against rails.

However, these are not the only sounds which a participant might be able to hear in this experience. Alongside the sounds that someone might recall and observe, someone standing at various places throughout the park might be able to hear other noises coming to prominence – the helipad at the local hotel is not simply decorative, and once upon a time the largest parking lot was an active airport. Cars drive around the outside of the park – and even inside on occasion – their engines a distant roar. The world is not without its own wildlife ecosystem, real rats competing with Mickey Mouse for space in the storage units beneath the park as they scratch and squeak. Birds perch on benches, in trees, overhead, their songs occasionally filtering through the noise as they try to communicate with each other. And sounds may emerge from beyond the looking glass if you come too close to the edges of the virtual space. Liminal spaces can be found at the edge of each Disney theme park area, where guests exist between two distinct Disney zones, and between the guest areas and staff areas of the park. If you step too far into these liminal spaces, then you may receive an odd mixture of two separate virtual worlds, or even find that cast member voices travel surprisingly well out of backstage tunnels and into the theme park festivities. Perhaps we might even hear the sound of a family argument beginning as we pass through the gates of the park – a reminder that reality is not as fantastical at Disney intends us to believe. As a visitor to a Disney theme park, we are inundated with thousands of distinct sound sources constantly competing for our attention.

We can categorise many of these sounds within the framework suggested by Schafer. The general ambiance of the Disney theme park experience is created through the use of franchised soundmarks – the voice of Jack Sparrow (as recorded by Johnny Depp) or the distinct music used throughout the Disney franchises. Some of these sounds may also function as signal sounds to those audiences who are focused on uncovering specific aspects of the Disney experience. Perhaps a participant seeks out a favourite character or favourite location at a given theme park. But other signal sounds can also be observed: Disney is known to use music to signal when a show is about to start and therefore attract crowds. Clearly this is an example of how signal sounds can be used within an experience such as Disney. Keynotes can also be observed; we discussed the concept of (under-/over-)crowding in section 1.3.1 and should recognise that the sounds of *other guests* are therefore likely to be part of the general character of the Disney experience.

However, we should also recognise that Schafer's model is designed primarily as a mode of analysis, rather than a mode of design. To explore how Schafer's model can be evolved to inform questions of sound design, imagine now that we are not listening to this Disney soundscape as 'just' a visitor. Rather, we are listening actively for the express purpose of recreating this soundscape (at a later date) as a sound designer. We are going to ‘virtualise’ Disney, similar to the experiences provided in games such as *Disneyland Adventures*. A key part of this kind of virtualisation is presumed to be the authentic recreation of ‘the sound of Disney’. However, the question that arises is one which remains at the crux of sound design for virtuality. Which of the sounds described above should be considered the “sound of Disney”? For example, Baudrillard argues that Disneyland portrays (in part) a utopian ideal of America: a “miniaturized pleasure of real America” (Baudrillard, 1981, p. 12). Are the sounds from ‘beyond the looking glass’ identified in the Disney soundscape above part of that depiction of the perfect American dream? Or are we mishearing Disneyland when those sounds gain our attention as it arises from our surroundings?

It seems unlikely that many of these liminal sounds are intended to be part of the Disney experience when the acoustic designers were considering the construction of Disney theme parks. However, we can recognise that those sounds are clearly there and are part of our experienced soundscape. This issue extends beyond the remit of Disney theme park(s) and across all virtual experiences; many ancillary noises may be present within other experiences. During the process of recreation, we can therefore imagine that there is a clear difference between *reproducing* Disney in the virtual (i.e., creating an exact replica of an actual experience) and *simulating* Disney in the virtual (i.e., creating an idealised version of the imagined experience). To move beyond analysis of a given soundscape and the beginnings of sound design, we must draw a distinction between recreating sounds which were actually there and recreating an overall imagined experience which the creators (of the real experience) intended for us to hear. Achieving this requires us to acknowledge and address weaknesses within Schafer’s *Soundscapes*.

2.2.0: Recognising limitations within the soundscape model

There are three key areas within Schafer’s model which must be critiqued to develop his conceptual model of ‘soundscapes’ into a more useful form: (1) Schafer’s advocation for what he calls archetypal sounds; (2) Schafer’s separation of ‘noise’ within audio experiences; and (3) methods of accommodating and mitigating the limitations of relying on specific personal experiences such as those encapsulated within a soundscape. Each of these areas will now be discussed as to construct a model for utilising soundscapes for sound design: do archetypal sounds (a somewhat important element of Schafer's philosophy of sound) have any relevance to the process of sound design; how might we formulate an understanding of 'noise' (and disruptive sound) within virtual experience; and what steps must be taken to ensure that our specific listening experiences are representative of a given environment when developing sound designs for virtual recreations of an experience?

First, Schafer expounds upon the three types of sound identified above – keynote sounds, signals, and soundmarks – by suggesting the existence of 'archetypal' sounds: "those mysterious ancient sounds, often possessing felicitous symbolism, which we have inherited from remote antiquity or prehistory" (Schafer, 1977b, p. 9). Schafer is drawing predominantly upon Carl Jung's concept of archetypes as symbols which "can arise autochthonously in every corner of the earth and are none the less identical, just because they are fashioned out of the same world-wide human unconscious, whose contents are infinitely less variable than are races and individuals" (Jung, 1921, p. 153). Archetypal sounds (or perhaps more accurately: sounds which are understood through archetypal mechanics) are, to an extent, central to Schafer's model as they imply that some or all sounds are universally understood by humans. For Schafer, archetypal sounds emerge from a shared understanding of a (and inherent) collective memory of sound. These sounds, therefore, are suggested to be non-subjective: the understanding of these vibrations is not dependent upon a learned understanding or a personal subjectivity. Rather, the understanding of these vibrations is an objective reality – established globally and universally within the primordial subconscious part of the human brain where "all psychic functions are indistinguishably merged in the original and fundamental activity of the psyche" (Jung, 1921, p. 142).

The notion of such globally synchronous archetypes is, to say the least, contentious within the general domain of human experience. Within the specific domain of sound, little evidence exists to suggest that such sounds exist – at least using Schafer's description of *archetypal sounds*. Within

Schafer's own work on archetypal sound, an inherent contradiction can be found: Schafer focuses in on the use of bells, suggesting that "the sound of bells continues to evoke some deep and mysterious response in the psyche" (Schafer, 1977b, p. 176). Whilst this may seem, at first, superficially plausible, Schafer seems to challenge his own thesis of archetypal sound by suggesting that, "One day perhaps we will be able to run a test similar to that with the church bell among the oriental people, employing the gong as the test sound" (Schafer, 1977b, pp. 10–11). It seems self-evident that a perceived need to change the sound to elicit the same fundamental response from an alternate geographic domain demonstrates that such sound cannot be "primordial" or "archetypal." If the understanding of bell chimes differs through cultural contexts, through it seems more likely that the hearing of bell chimes is informed by cultural (i.e., regional rather than universal) semiotic perspectives rather than any perspective inherent to humanity as a whole. Schafer further challenges his own concept of primordiality by correctly identifying that technological progression also informs 'hearing' – "[as civilisation progresses], all acoustic symbolism slowly but steadily undergoes modification… [Humanity] has sought to tame the wind in the air-conditioner. Transformations such as these will undoubtedly change the symbolism of such archetypes." (Schafer, 1977b, p. 178) Again, it seems to a large extent self-evident that shifts in understanding emerging from societal contexts suggests that sound semiosis is learned, rather than inherent.

Similarities to arguments surrounding other 'universal' interpretations of specific sounds can be examined. An argument that music will be interpreted or understood by everyone in the same way has constantly been stated throughout most of the last few millennia – from the ancient Greek perspective that different *harmonia* contained some intrinsic elements which evoked a specific temperament or emotion, to the modern blog posts professing that music is a "universal language". Yet evidence suggests that these similarities are, at best, limited: e.g., one large-scale international study on the 'major-happy minor-sad' distinction found that it was a "learned cultural convention instead of an innate musical universal" as communicating emotions through music required at least some learned cultural specificity (Athanasopoulos et al., 2021, p. 13).[23] Succinctly, it appears that little convincing work has been done to evidence objective or 'primordial' understandings of sound in the specific sense, beyond sweeping generalisations of timbral qualities of sound. Given that any archetypal qualities of sound are therefore somewhat vague, I shall consider them essentially irrelevant to a consideration of a specific "sound of Disney" – or any other specific soundscape – and omit the concept of 'archetypal' sounds (i.e., sounds which are received in the same way across cultural/regional divides) from our development and discussion of soundscapes and recognising that sound – as explored in section 2.1.0 – is a primarily subjective and personal experience. Whilst archetypal sounds may be important within Schafer's model, they are not useful within the specific understanding of sound that this chapter develops as one axis of the framework developed throughout this thesis.

Second, Schafer's framing of different 'noises' which may be heard suggest an implicit idealisation of the audio experience. Schafer is perhaps intending to set out the way in which transformations of the soundscape can, through long periods of history, be changed by emerging technologies or ideas as can be shown by his description of soundscapes as studying "sounds that matter" (Schafer, 1977b, p. 10). This shows that Schafer is not discussing the audiological 'noise' in the sense of randomised sound (e.g., static), but rather making implicit value judgements as to what sounds 'matter'. Schafer sets out the soundscape as being disrupted – and destroyed – by certain sounds which he classifies as 'noise':

> Noise has a variety of meanings and shadings of meaning… but of the four general definitions, probably the most satisfactory is still "unwanted sound". … the new sensibility of man depended on his appetite for noise. Today, as the machines whirl in the hearts of our

[23] A limited number of counter examples do exist. For example, low-frequency sounds may inspire feelings of fear in many people. This is thought to be due to consistent relationships between low-frequency sounds and danger: whether that be large-scale natural events such as earthquakes or thunderstorms, or many predator sounds (e.g., Tigers are thought to sometimes generate infrasound frequencies whilst roaring). However, these examples seem to predominantly suggest broad relationships between timbre and reception, rather than universal associations with specifics sound across all human cultures. Similarly, research into linguistics has identified that many separate languages share common sounds (or avoid common sounds) to describe similar concepts, which has been suggested to indicate an underlying universality of phenomes (Blasi et al., 2016). However, examples of archetypal sound beyond these limited examples have not been, to my knowledge, been evidenced in published peer-reviewed research.

> cities day and night, destroying, erecting, destroying, the significant battleground of the modern world has become the neighbourhood Blitzkrieg. (Schafer, 1977b, pp. 181–183, 185)

New sounds – or at least, sounds of technology – are portrayed as disruptive or obstructive, barring access to information which was formerly available. This implicit valuation is potentially problematic – should certain audial phenomena within a soundscape be removed because they counteract the listener's subjective views on what sounds *should* be present? Do these sounds matter? If the intent is to faithfully reproduce the specific experiential audial sensations at a location, then it seems likely that a holistic soundscape would (or at least, *should*) include these sounds alongside the other categories – that is to say, these sounds matter, even though they are disruptive to the listener's expectations. It is not clear whether Schafer excludes noise from his categories – keynotes, soundmarks, signals – because he feels that any noise likely *also* belongs to one of these three existing categories, or because he wishes to exclude them.

Whatever justification we may read into Schafer's separation of noise from his central categories, the subjectivity of noise should be highlighted: sounds which some may consider "unwanted" may be appreciated by others. We can consider, for example, the number of complaints surrounding the sound of church bells in modern urban environments as a well-known area of contention between those that like and dislike their sound. We can recognise that church bells are seen as intrusive by those that do not believe that such sounds *should* be included within the given soundscape – perhaps they feel that church bells are an archaic tradition which should be abolished or find that the sound of church bells disrupts their favourite morning activities. These individuals may find that the bells are noisy. Conversely, others hearing the same church bells at the same time in the same location may enjoy them and recognise them as an important part of the soundscape. Such disagreements over whether a noise is an important feature of a soundscape or simply noise obscuring and/or disrupting the soundscape have featured prominently throughout history and continue through to the current day.[24]

[24] Recent examples can be found from Italy (Giuffrida, 2022), America (Park & Mac, 2015), Wales (Hull, 2019), England (Petherick, 2021), and undoubtedly elsewhere too.

The final issue with Schafer's model for soundscapes is that it may not be the most appropriate approach for describing environments experienced simultaneously by many users. The specificity of soundscapes – whilst a key part of the perception of sound – works against the hypothetical situation in which we are trying to recreate the experience as a virtual experience for *any* user. This should be expected. As set out above, Schafer is providing an *analytical model*, whereas we require a *design approach* to create our virtual experiences. Alongside the changes highlighted above, we can therefore recognise that we must ensure that we transform a specific experienced soundscape (i.e., the description of sounds/noises which surrounded us at a specific place and specific moment in time) into a generalised sound design (i.e., an auditory environment that allows users to experience a specific soundscape). Several perspectives of a real environment across many different contexts need to be provided to ensure that varying sounds are represented accurately throughout the experience. Research into what specific sounds represent may be required to ensure that the subjective interpretation of the listener does not affect the supposedly objective representation of those sounds within the experience: that is to say, sound designers must be aware of their own cognitive/subjective biases and ensure that their understanding of the experienced soundscape is accurate before recreating it. Returning at different times, or taking recordings from many periods, may also be required to establish the frequency and commonality of specific sounds – e.g., is a sound always part of a soundscape, or only sometimes part of a soundscape? We can imagine that a soundscape of a British urban environment would be very different on the evening of 5th November or 31st December when compared to the same urban environment on most other days of the year – or even at any other time throughout those days. But if our only perspective were our experienced soundscape from the evening of 5th November then we would have a very unusual perspective on the 'normal' activities taking place in an urban environment. To a large extent, this should be self-evident. A well-described soundscape is likely to accommodate many of these potential issues in the specificity of sound.

Returning to the specific context of Disney, our initial hypothetical soundscape included the potential for a familial argument at the gates of the park. We can examine this argument through the lens of each of the issues outlined above: is the argument simply an occurrence that happened at one instance, or is the argument frequently part of the sound of Disney? Is the sound of the argument disruptive or obstructive to the overall experience of Disney? Will the noise of the argument be interpreted in the same way by all listeners? (i.e., some listeners may not realise it is an argument and, therefore not have their narratival immersion within the virtual Disney disrupted). These investigations are both useful for considering the limitations of the soundscapes and for considering the broader process of sound design.

2.2.1: Sound Design and Meaningful Hearing

We have already described how sound implicitly has meaning – we hear birdsong, therefore there are birds. This has been discussed in brief throughout section 2.1.0. To use sound purposefully in virtual experience, however, we must determine how to tap into the auditory understandings that audiences have. A detailed understanding of how we generate semiotic connections between sounds and objects is beyond the scope of this thesis. However, a paraphrased version is provided here to contextualise further discussion: we are exposed to audial cues throughout our life. Through exposure and understanding, we learn what each different audial cue conveys. Subconsciously, we are constantly comparing the sensations of now to the sensations of the past. We can also consciously analyse sound to determine if it is familiar to other sounds which we already recognise. Through these conscious and subconscious processes, we use sound content to understand our experiences of the world around us. In essence: we understand what sounds mean because we have heard them before and can (sub)consciously associate new vibrations with old interpretations.[25]

This interpretative process of hearing is part of the communicative process of listening discussed in 2.1.0, wherein sound is derived when physical vibrations are processed by the brain. However, as outlined by Wierzbicki, a problem emerges in sound for virtual experience because “every sound has meaning” (Wierzbicki, 2016, p.153). Whilst the human brain is capable of

[25] See, for example, Peirce’s theory of signs, as explained by Albert Atkin (2022).

"screening the incoming [auditory] information … to that which may be significant" (Truax, 2001, p.19), Wierzbicki notes that within contexts such as the fictional world of cinema, this screening is "typically done *for* us" (Wierzbicki, 2016, p.153, emphasis original). This goes beyond simply understanding sound, and – instead – refocuses our attention on sound *design*. This section of the thesis examines the process of sound design. By understanding how users derive information from sound in virtual experience, with particular focus on the concept of 'media literacy,' we can work toward a systemic process of sound design that ensures users can correctly interpret meaning from designed auditory environments.

The process of sound *design* can be imagined as an attempt to synthesise and distil what is/was heard (or imagined to be heard) into a series of meanings. In the process of sound design, we are examining the sensations of the past – e.g., the sensations encapsulated within the soundscape – and creating a list of meanings. We can then identify which sounds we can use to communicate the same meanings to a future listener. Whilst these sounds may be recordings, these sounds are often constructed or engineered: e.g., the sound of a balloon popping, modified through sound engineering tools, becomes an explosion. By playing sounds that communicate the same meaning to the listener, we convey the constructed soundscape to the listener. The process of design is complete when we can imagine participants *should* be able to derive the intended communicative meanings from the auditory environment, therefore recreating the same experience that we described in our soundscape (even where the exact noises that we heard at a specific time and place are not faithfully replicated).[26]

[26] Truax describes a similar process as belonging to the "so-called applied fields of traditional [acoustic] disciplines" in which "their design goal is that the desired stimulus will arrive at the receiver and the undesired noise will not, or at least not in such quantities as to obscure the desired signal" (Truax, 2001, p.14). However, Truax calls these approaches into question stating that they "do not lead to design criteria that go beyond the question of the appropriate control of signals to ask what kind of environment is desirable, meaningful, or beneficial" (p.14). Conversely, I believe it is evident (as shown through discussion in this chapter) that the process of controlling auditory signals *inherently* asks what signals (and what acoustic environments) are meaningful as – per section 2.2.2 – a meaningful signal which is misheard by a listener will be perceived as noise. This is likely because I have adopted an experiential perspective (i.e., how does the audience hear a given auditory stimulus?) rather than a technological perspective (i.e., how is an auditory stimulus transmitted through a sound system?).

Some concerns do surround this process – Westerkamp, for example, suggests that someone composing a soundscape might create a “schizophonic” sonic experience if they “have not experienced [the soundscape] through the recording process” (Westerkamp, 2002, p.55) as such designed soundscapes are “always rooted in themes of the sound environment… never abstract” (p.53). Presumably, the construction and engineering of additional or synthetic sounds would raise further issues for Westerkamp. However, these practices are clearly widespread and welcomed broadly by audiences, possibly due to a pre-existing relationship between the audience’s expectations and the sounds that they hear. As Westerkamp describes: “if the listener knows the place, time or situation of which the [soundscape] speaks, the composer may have less of a problem communicating meaningfully to the audience, because a relationship of some kind already exists quite apart from the composition itself” (p.56). Alternately, these differences may reflect the difference between processes of acoustic ecology – “[the study of] relationships and interactions among humans and sounds in an environment” (Pijanowski et al. 2011, p. 204) – and processes of sound design for virtual experience.

A variety of different methodological approaches to sound design for virtual experience have been used over the last century. Perhaps most ascetically, Danish directors Lars von Trier and Thomas Vinterberg started the (deliberately absurd and extreme) low-budget film movement *Dogme 95*, which calls for the use of “diegetic sound only” – with an implication that sound should be captured live-on-set (“the sound must never be produced apart from the images or *vice versa*”[27]). Less extreme approaches appear throughout various realist approaches to film, such as recreating the sounds captured on set-microphones within a sound studio (e.g., dubbing) to improve the overall quality of the audio. However, many modern filmmakers use hyper-realistic approaches to sound: Tomlinson Holman describes the need for sound recordings for film and television to be “an exaggeration of reality” as “there is so much competing sound at any given moment that sound [which] must be heard has to be rather overemphatically stated, just to “read” through the clutter” (Holman, 2010, p. xii). Andrey Tarkovsky concurs that sound should prioritise effective communication over realism,

[27] This is found as the second rule in the Dogma Manifesto: *The Vows of Chastity* (Lars von Trier & Thomas Vinterberg, 1995). See Stevenson (2003), pp. 22–23.

suggesting that “if sounds of the world [were] reproduced naturalistically in cinema…there would be a cacophony” (Tarkovsky, 1989, p. 161). The need for this can also be purposeful, rather than aesthetic. Wierzbicki emphasises how filmmakers found that realistic sounds “were simply not theatrical enough for the filmmakers’ needs” and replaced them with “sonic simulations” (Wierzbicki, 2016, p.154). Specifically examining sound within virtual reality, Angela McArthur et al. express a belief that designers must consider the perceptual experience of sound to create an aesthetically appreciable world (McArthur, Sandley, et al., 2017, pp. 30–31). In each of these hyper-reality is portrayed not just as a choice, but as a requirement of purposeful sound design. Cinema is seen to rely on “audio-visual editing in which short shots are amplified by hugely exaggerated sound effects…the diegetic sounds are processed and constructed into unrealistic, striking sound effects with lives of their own” (Kulezic Wilson, 2015, p.48). The differing approaches to sound design exemplify how sound design is not just a process of distillation but also an active process of creation: sound designers must make active decisions to create an acoustic environment for their listeners and, in modern sound design, often do so through deliberate exaggerated and ‘hyper-real’ methods.

It is accepted that sound fundamentally changes our understanding of a visual scene. Michel Chion describes how sound “enriches a given image as to the definite impression, in the immediate or remembered experience one has of it, that [expressive or informative value] ‘naturally’ comes from what is seen, and is already contained in the image itself” (Chion, 1994, p. 5). An image of water falling on a window is explained further through the addition of sound: is this a rainstorm? A leaking gutter? A person throwing a bucket of water at a window? Sound can allow us to easily determine this. As Kulezic-Wilson puts it: “sound and image [in audio-visual media] should … be employed in such a way that each always brings to the synthesis what the other is not capable of” (Kulezic-Wilson, 2015, pp. 91–92).. Sound designers must understand how users experience and receive sound events. They must establish what sound events should be placed within a virtual reconstruction of an experience – e.g., through recording or imagining soundscapes – whilst also understanding the ways in which those sound events will be understood. In order to do this, sound designers must be aware of how sound can be used purposefully.

The most obvious level of this was discussed in 2.1.0 – sound implicitly conveys informational meaning. However, sound conveys information only about what can be seen, but also about what cannot be seen. These sounds are "acousmatic" – heard "without seeing their originating cause" (Chion, 1994, p. 71). By hearing objects which are not present within the visual frame, a viewer can imagine a world beyond the frame: e.g., car horns and other traffic noises might imply a busy street just outside a window (or otherwise outside the visual frame). We could also consider spatial qualities of sound to be acousmatic qualities as they convey information about things which are not present within the visual frame: e.g., reverberation may help a viewer to understand the size of a room only partially visible in the visual frame.

Accordingly, we must consider how sound is spatialised when mediatising it through technology. We frequently discuss whether sound is *mono* (no spatialisation) or *stereo* (e.g., left/right spatialisation). Other common approaches include *surround sound* (sound approaches from many angles, e.g., cinematic sound), and *binaural* (perspective-specific sound, e.g., head-mounted virtual reality).[28] We can see the determination and implementation of this decision as a process of spatial mediatisation, which is particularly useful for communicating information in virtual reality experiences. Martin Rieger describes how the spatialisation of voiceovers must be carefully considered in virtual reality: "most of the confusion I see with people is that they don't really get voiceovers. They think that there is a person in the scene and they're trying to find it… but with spatial sound you can immediately make them understand that mono[phonic] sound means there's a voiceover and spatial audio means there's a person really in this scene" (World XR Forum Crans-Montana, 2020, sec. 09:10–10:30).[29] Similarly, Larsson suggests that sound within virtual experiences

---

[28] Other approaches for spatialising sound are discussed in Begault (1994) and Rusmey (2011).
[29] Whilst Rieger – and this section of the chapter – focus on the *stereo spatialisation* of the sounds, it should be noted that other spatial elements (e.g., reverberation) also contribute to these differentiations between diegetic and non-diegetic with fully mediatised virtual experiences such as virtual reality. The combination of these

can be used to “induce both object presence (‘the feeling of something being located at a certain place in relation to myself’) and spatial presence (‘the feeling of being inside or enveloped by a particular space’) at the same time” (Larsson et al., 2010, p. 5) through this spatial mediatisation. These examples show that different spatialisation of the same sonic material can produce distinctly different understanding of a virtual experience for a given participant.

When considering interactive audio-visual media, we can recognise another benefit to the use of acousmatic audio: acousmatic audio can suggest to the user that there is something for participants to explore and/or discover beyond the part of the virtual experience that they are currently visually perceiving. Karen Collins suggests that “players [of video games] must be involved in the sound because if they fail to listen attentively and respond correctly, their gameplay will suffer” (Collins, 2013, p. 143). Similarly, Grimshaw observes that “sound is capable of depicting events and spaces beyond the confines of the screen… the importance of sound to the positioning of the player within the game world cannot be underestimated” (Grimshaw, 2014, p. 117). Whilst Collins and Grimshaw-Aagaard are specifically talking about video games, we can recognise that acousmatic audio will aid users of all virtual worlds to navigate more easily through their virtual experiences. For example, Larsson et al. establish that the auditory system can allow us to locate objects within a space as “the direct sound from the object constantly reminds us it is actually there” (Larsson et al., 2010, pp. 10–11) and Rieger suggests that acousmatic audio allows sound designers to “subtly guide viewer’s attention” to items or events that are out of frame (World XR Forum Crans-Montana, 2020, sec. 06:20–06:40). So strongly embedded is this explicit expectation of sound in virtual experiences (and specifically within video games), that – as Kevin Donnelly suggests – acousmatic audio can be used to create “uncertainty for the perceiver” through ensuring that a player is unsure “what these sounds indicate and where they are located” (Donnelly, 2016, p.76) and thereby betraying the user’s experience.

different elements can also help users to differentiate between sounds generated by the virtual software/hardware from sounds external to the virtual experience.

Sound can also communicate information about the virtual experience itself. These are described by Kristine Jørgensen as "transdiegetic" sounds: "diegetic sounds that address an entity external to the diegesis, or extradiegetic sounds that communicate to entities positioned internal to the diegesis" (Jørgensen, 2007, p. 115). That is to say: sounds which cross the boundary between the diegetic world and the nondiegetic world[30] – the 'magic circle' discussed in section 1.2.1. Jørgensen explores a specific example from within video game studies: "although [the player's avatar] does not hear extradiegetic sounds due to its position internal to the diegesis, in effect it may react to extradiegetic sounds because of the control link between the player and the avatar. In a sense, this fictional character can evaluate and act on information that it should not be able to hear" (Jørgensen, 2007, p. 108). Jørgensen is particularly concerned with how this might inform an understanding of communicative sound outside of video games: "In the natural world, meaningful auditory information occurs naturally as causes of a range of events such as doors closing…and they work to provide information about what is going on around us. If the same kind of response to events is desired in computer-based systems, however, these sound cues must consequentially be added" (Jørgensen, 2007, p. 115).

Examples of these types of transdiegetic sound can be found in many types of virtual experience outside those in video games as discussed by Jørgensen. In a 1985 technical report released by the American Office of Scientific and Technical Information, Denise Ann Sumikawa coined the use of 'earcon' (1985), which was later refined by Meera Blattner et al. as a term to describe "nonverbal audio messages used in the user-computer interface to provide information [and feedback] to the user about some computer object, operation, or interaction" (Blattner et al., pp. 12–14). Similarly, William W. Gaver examines the use of "auditory icons" in a 1986 paper: "caricatures

[30] The separation of diegetic and non-diegetic within cinematic analysis has been criticised widely: e.g., Ben Winters suggests that such separation threatens to "separate non-diegetic [sound] from the space of the narrative, denying it an active role in shaping the course of onscreen events" (Winters, 2010, pp.224–225). The concerns of Winters and others are clearly well-grounded as it is clear that many – if not all – 'non-diegetic' sounds are part of the narrative world. Therefore, the distinction between 'diegetic' and 'non-diegetic' is ultimately somewhat arbitrary. However, the use of these terms has been well-established in sound and music literature, and 'diegesis' is used as it has been widely used elsewhere.

of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes" suggesting that "auditory icons may represent conceptual objects in a computer system more clearly than other sounds" thereby allowing it to be used to provide specific information to the user of a computer system (Gaver, 1986, p.167). Both auditory icons and earcons could be seen as fitting within Jørgensen's definition of transdiegesis as they are sound emerging from the 'diegetic world' (e.g., an email inbox) to address the external entity (i.e., the user).

These sounds form an interface – something which "enables information to be provided, accessed, and applied. Acting like a translator…making one sensible to the other." (Mauger, 2014, p. 32) – which allows for communication between the virtual hardware and the users. We can see transdiegetic audio cues as an audio interface. Virtual reality experiences are likely to have a greater need for audio interfacing as other human-computer interface systems such as flashing lights or system indications (e.g., taskbar notifications) are often excluded from virtual reality interfaces as they are not immediately visible to users who are visually immersed; audio interfacing has already been shown to improve outcomes when used in situations where users are unable or unwilling to closely examine visual interfaces such as whilst driving (Jeon et al., 2015; Nadri, Chan Lee, et al., 2021; Nadri, Ko, et al., 2021), browsing websites (Cantrell et al., 2022; Rohani Ghahari et al., 2013), and in assistive technology for people with visual impairments (Machado et al., 2017).

The design of these audio interfaces has always been known to rely on an understanding or development of similarities between related cues: e.g., Blattner et al. suggest that earcons should be grouped into hierarchical families which share auditory properties such as timbre or rhythm so that users are able to easily determine what information an earcon might be indicating (Blattner et al., 1989, pp. 30–31). Blattner et al. propose that this interconnectivity may aid the understanding and interpretation of auditory interfaces because "for an earcon to be useful, listeners must learn and remember it" (Blattner et al., 1989, p. 34). The need for interconnectivity in earcons can be observed in action in the design of specific auditory cues, such as in Chia-Fen Chi et al.'s recommendation that ISO standards should be more widely adopted for auditory signals to facilitate ease of use for consumers (Chi et al., 2017, p. 162) and as part of Guillaume Lemaitre et al.'s framework for developing new auditory signals which could function as a car horn:

> Hearing a car horn sound warns road users because they recognize the sound of a car horn, they know what this sound means, and they know what they have to do as a consequence…when introducing new warning signals, care must be taken that these sounds are not too different from the already-existing ones: the more the new signals are different from the already-existing ones, the more the road users will need time to learn their meaning (Lemaitre et al., 2009, pp. 358–359).[31]

The use of underlying knowledge required to interpret auditory interfaces has broader applicability to sound within all virtual experiences. Isabella van Elferen draws upon Patricia Aufderheide's definition of literacy as "the ability to access, analyze, evaluate and communicate messages in a variety of forms" (Aufderheide, 1993, p.6) and Jose P. Zagal's concept of ludoliteracy (2010) – which van Elferen summarises as "heterogeneous skills of hardware competence such as handling a console, knowledge of generic gameplay such as picking up food or ammunition, and interpreting the virtual game world's relation to the actual day-to-day world" (van Elferen, 2016, p. 36) – to suggest a broader sensory model of media literacy. In this model, which van Elferen focuses on "musical media literacy," audiences develop an ability to "interpret film, television or advertising music through the fact of our frequent exposure to them… We know, for instance, to expect catastrophes when a horror film protagonist's descent into a dark cellar is accompanied by low, dissonant cello motifs" (van Elferen, 2016, p. 36). Further to this, van Elferen suggests that music's "crucial role in gaming interaction [means] the sonic cues that the player gets must be immediately recognizable and interpretable" (van Elferen, 2016, p. 37).

Sound allows for participants within virtual experiences to gain vital information in several domains. Sound can be used to directly portray objects within the virtual world; sound can be used to indirectly provide information about the spatial dimensions and properties of the virtual world; sound

[31] Lemaitre et al.'s framework is complicated by their recognition that potential replacement sounds for car horns are frequently not recognised as car horns because they are recognised as an alternate specific sound (p.359), which suggests a need to balance novelty with recognition when introducing replacement sounds within an existing framework of semiotic understanding for any purpose.

can encourage audiences to imagine (or look) beyond the frame of the world as it is provided to them; and sound can be used to interface with users to communicate information about the virtual experience itself (e.g., earcons, auditory icons). In all these cases, the understanding (and therefore, the use) of these sounds often relies on the prior knowledge and expectations of the viewers/participants and draw upon generally established semiotic domains in what can be referred to as 'media literacy.' The combination of sounds within a virtual experience can be referred to as an auditory environment; the purpose of an auditory environment is to allow users to experience a soundscape which fulfils the goals of the virtual experience.

2.2.2: Noise

The considerations set out above have predominantly considered what an auditory environment can do, and how sound designers can create an auditory environment which functions to do so successfully. Implicit within those considerations it the notion that if sound designers are unsuccessful in their realisation of an auditory environment, then users may experience obstacles to determining the correct interpretation of the auditory environment and – therefore – obstacles to perceiving the intended soundscape. We can, as Jacques Attali states, recognise that these interferences can be regarded as "noise … a signal that interferes with the reception of a message by a receiver, even if the interfering signal itself has a meaning for that receiver" (Attali, 1985, p. 27) Within virtual experience, we can see that noise predominantly emerges through two paradigms: (1) sounds placed intentionally within an environment by the sound designer which are misunderstood by the user; and (2) sounds which are not intended by the sound designer but which are transmitted within the user experience. We can contemplate the former as a form of audial hallucination – i.e., a sound mis-identified as belonging to one (real or virtual) environment but which actually belongs to another (real or virtual) environment[32] – emerging not only from the conflict of reality and virtuality,

[32] Audial hallucinations are used here in the sense of what Grimshaw & Garner refer to as exosonic components – i.e., they are physical vibrations which are misinterpreted by the listener. It is therefore possible for them to be shared or directly co-experienced with others. This directly contradicts Grimshaw & Garner's definition of audial hallucinations, which describes specifically pathological audial hallucinations internal to the listener which therefore cannot be shared (Grimshaw & Garner, 2015, pp.118–121). The two should not be confused.

but also from conflicts between virtual experiences happening simultaneously. We can imagine that someone playing a video game whilst listening to music may not always be able to distinguish between purposeful music from within the game and incidental music from outside the game (e.g., from a radio). Where music functions to communicate meaning within video games,[33] the resulting confusion could be said to cause the incidental music from outside the game to become 'noisy.' It obstructs the intended meaning and can cause issues for players reliant on the information communicated by the game music. These concerns may have become increasingly relevant in recent years due to the spread of what Anahid Kassabian refers to as 'ubiquitous music' which "comes out of the wall, our televisions, our video games, our computers, and even out of our clothing" (Kassabian, 2013, p.12) or – perhaps more precisely – "musics that we listen to as secondary or simultaneous activities" (p.31). It is increasingly likely that users will *also* hear music whilst they are engaging with virtual experience and, therefore, face these challenges more often and may do so, as Kassabian observes, "often without choice" (p.31).

Another form of misunderstanding can be observed when sounds do not meet with the expectations of the listener: as we are reliant on our understanding of sounds emerging from our past encounters with them – especially within the paradigm of media experiences – then we can imagine that our understanding of sound is a distinctly cultural phenomenon. In our discussion of media literacy, we established how our understanding of sound is informed by our learned knowledge gained through exposure to similar media: e.g., van Elferen suggested that intended auditory messages must be "immediately recognizable and interpretable" (van Elferen, 2016, p. 37) and Lemaitre et al. suggests that "the more the new signals are different…the more the [users] will need time to learn their meaning" (Lemaitre et al., 2009, pp. 358–359).

Two distinct possibilities for noise emerge from this reliance on learned knowledge. First, the globalisation of media will expose people to a new series of cultural phenomena outside of their normal frame of understanding. Shoma A. Chatterji, for example, argues that Indian cinema has a

[33] Chapter 4 explores this topic in greater detail.

specific and unique sound design which is reflective of Indian ethos and Indian culture (Chatterji, 1999). Similarly, Alexander Binns explores how Japanese cinema developed a distinctive model reliant on musical association and musical rhetoric specific to the history of music in pre-cinema theatre (e.g., narration theatre) within Japan. Whilst Binns avoids making specific claims about how the music 'should' be read, his conclusion seems clear: "by understanding some of the ways in which music embraced and supplemented the theatrical traditions on which much of Japanese cinema was founded…a framework emerges in which to read the function and effect of this music appropriately" (Binns, 2017, pp.437–438). A Western audience exposed to this global media may struggle to derive the intended reading of the narrative from the audio of the film. Under-exposed to these region-specific cultural frameworks, audiences may begin to find the sound and music of these other traditions confusing; they might find the sound and music of cultural traditions they are unfamiliar with to be 'noisy.'

Second, media which aims to be 'authentic' in reproducing sound may discover that existing cultural framings of specific sounds are incorrect. This will challenge audience perceptions dramatically and may pose a disruption to perceived immersion and engagement with virtual experiences. This will be explored more in Chapter 3, which explores issues of authenticity in sound design, but can perhaps be easily exemplified here through acknowledging that recent research shows that dinosaurs appear to have been unable to roar like dinosaurs found in *Jurassic Park* (Universal Pictures: 1993) and similar films but rather were more likely to chirp and trill, in a fashion similar to pigeons (Riede et al., 2016). We might expect an audience familiar only with the framing of 'dinosaur sounds' provided through existing media to have issues with 'authentic dinosaur sounds' – as a result of their confusion, they may experience them more as noise than as sound; the sound may disrupt their immersion as it reminds them of the external world.

Sounds which are not intended to be part of an experience, but which are nonetheless transmitted to within the user experience, are perhaps the more traditional perspective on noise. This would include disruptive sound. We could imagine that we are in a virtual experience when an alarm goes off within the externality from which we are exploring the experience. At first there might be momentary confusion – is the alarm going off inside the framing of the experience? We might experience an ‘audial hallucination’: our perceptions of reality have been subverted through sensorial immersion within virtuality, and therefore our perceptions of reality are conceived as happening within the immersive virtuality. But eventually we are likely to realise that the fire alarm does not follow the rules or processes of other sounds happening within the virtual world and, therefore, is likely not a sound that exists within the virtual world. For example, similar to the use of spatialisation to differentiate between (non-spatialised) narration and (spatialisation) dialogue described by Rieger, users may be able to identify whether a sound is emerging from the virtual world: as the sound does *not* react to player actions within the virtual world (e.g., turning around, tilting their head, pausing the experience), the user becomes aware that the sound must exist outside the intended virtual externality. We can see this as disrupting the overall immersion of the user, and, therefore, being noise.[34] Alongside these environmental sounds, some noise in this category may emerge from errors in hardware or software configuration. For example, electromagnetic interference with hardware may cause ‘white noise’ or other feedback, and software misconfiguration may cause sounds to be (in)audible at the wrong time.

Many of these issues are often excluded from considerations of sound design for virtual experiences as they are considered to be beyond the control of the designer. For example, users playing virtual reality games at home will have their own hardware configuration. There is an almost infinite number of possible combinations and, therefore, many designers do not consider methods of mitigating disruption from these external noises within virtual experiences. However, from an experiential perspective, sound designers can and should consider methods of addressing these external noises, especially given technological progress being made in these areas: for example, Grimshaw-Aagarad & Walter-Hansen suggest that noise cancellation, audio filtering, audio masking, and real-time audio mixing based on user eye-tracking may enhance virtual reality experiences

[34] A specific discussion of the risks posed by “full-immersion” experiences (i.e., where users are completely unaware of the real-actual world) is found in section 6.3.0.

(Walther-Hansen & Grimshaw-Aagaard, 2020). Approaches to controlling, and designing for, external sound are discussed further in sections 3.0.3, 5.3.1, and 5.4.2.

Other sounds are deliberately intended to cross the barrier between virtuality and reality, such as user voice. Voice communication allows users to communicate information to each other. In my earlier research outlining a new framework for player voice in video games (Tatlow, 2020), I explore how pre-existing views on voice prioritised (and privileged) game sound over external audio: William Cheng outlines how voice can oppose constructed player/character identities (and the resulting player-character relationships) within virtual worlds (Cheng, 2014b, p. 337), Richard Bartle seems to strongly oppose the concept of player voice entering into virtuality at all arguing that it remains a reminder of the physical reality outside of the diegetic fantasy (Bartle, 2003, pp. 145–146), and Karen Collins argues that user-generated sound is a "particularly strong force in breaking the fourth wall" (Collins, 2014, p. 354). I propose a new model that acknowledges the existence of player voices which *can* cross the barrier between virtuality and reality without necessarily breaking the overall sensation of immersion experienced by user, with a particular focus on the way in which developers can encourage a more holistically immersive approach to voice within their games. However, the fact remains that voices of other players become 'noise' in many virtual experiences as a result of inevitable conflicts with the immersive virtual world. As with the concept of crowding explored in section 1.3.1, this may more easily occur in situations where there are a large number of concurrent users simultaneously engaging within a virtual experience. Some considerations of how digital multi-user environments can be managed are presented in section 5.4.0–5.4.3.

Throughout this chapter so far, Schafer's theoretical model for soundscapes has been examined, and the weaknesses of this model as part of a purposeful process of sound design for virtual reality identified. Suggestions for changes to Schafer's model of soundscapes were made, and an exploration of the new concept of 'auditory environments' generated. In particular, soundscapes were envisaged as a specific auditory experience of a particular place (or experience) at a particular time. Auditory environments were envisaged as a specific set of sound events and rules of interaction which allowed for multiple soundscapes to emerge from the same virtual experience. The process of sound design was identified as the creation of an auditory environment with the goal of providing the overall experience of a specific soundscape.

We can understand from this discussion that sound design for virtual experience should have three key stages. Firstly, the identification of the key components of the diegetic soundscape that must be created. Secondly, the identification of any important meta-textual information which is intended to be transmitted through auditory phenomena. Finally, the consideration of aspects of noise; both in terms of limiting externalities which might disrupt the immersive world (e.g., sensorially, narratively), and in terms of ensuring that participants in the virtual experience are able to correctly identify the intended meaning from the first two stages of the design. Through this theoretical model, we can ensure that the auditory environments we generate will dynamically create the expected soundscapes for the users, allowing them to have the intended virtual experience.

2.2.3: *Disneyland Adventures*

In order to understand how a soundscape might be transformed into an auditory environment, the hypothetical Disney soundscape presented in section 2.1.2 will now be compared to the auditory environment constructed for Frontier Development’s *Disneyland Adventures*, initially published as *Kinect: Disneyland Adventures* in 2011 and later re-mastered by Asobo Studio as *Disneyland Adventures* in 2017. An open-world video game loosely set in Disneyland California – one of Disney’s many theme parks – players must explore the theme park, meet major Disney characters such as Mickey Mouse and Donald Duck, and complete quests and missions for those characters. To accomplish the overall aims of the game, players will need to succeed in a number of minigames themed around the rides at Disneyland, and also socialise with Disney characters and non-player character (NPC) park guests. They are aided in their mission by an anthropomorphic ticket and a variety of items which they can unlock or discover such as an autograph book, a camera, and a magic wand.

Disneyland Adventures is, transparently, an advertisement for Disneyland theme parks – in much the same way that the LEGO games and movies are an advertisement for the LEGO toys, as perhaps best explicated by the viral 2018 YouTube campaign in which users could elect to watch the whole 100-minute LEGO movie as an advertisement (Julia, 2018). That the game was likely intended as an advertisement is not necessarily problematic insofar as considering it as a virtual recreation of the theme park. Rather, it suggests that this recreation is an idealised (hyper-real) version of the theme park intended to highlight the best aspects in every way. We can expect that elements of the park which Disney considers untoward will have been removed – for example, all mentions of corporate sponsors of Disneyland are removed from the recreation of the park within the game, likely because corporate capitalism may come into conflict with the sense of pilgrimage that leads many guests to visit Disney theme parks discussed in section 1.3.1. Other aspects have also been removed, including mentions of franchises or characters which Disney did not own licensing rights to in 2011 such as Roger Rabbit, and all characters from the Star Wars franchise. Similarly, there are no queues within the game. The player is able to instantly engage with any mini-game (the ‘rides’ of the game) of their choosing. The experience is not a perfect recreation, but rather one distilled and adapted for its virtualisation.

Nonetheless, many of the sounds identified in section 2.1.2 are present. The sound of children laughing and screaming can be found constantly throughout the game, and snippets of conversations can be heard as you pass through crowds as characters tell each other things like “This is the life!” and “This park is really amazing!” The characteristic voices of many major Disney characters are interspersed throughout the game – although Jack Sparrow is notably missing, as the character was created for the *Pirates of the Caribbean* movies co-produced by Disney and Jerry Bruckheimer rather than as part of the original Walt Disney *Pirates of the Caribbean* concept. The player is even offered the chance to participate in an event called “Mickey’s Soundsational Parade” where they not only watch a marching band and various parade floats travel through the park but also get to participate – conducting the band, bringing the parade floats come to life through using their magic wand, and dancing with the characters on the screen (Figure 11). Some of the ancillary sounds discussed in our soundscape can also be heard throughout the game: e.g., as the player avatar passes through areas with trees and bushes, the wind rustles through the branches and the faint tweeting of birds can be heard (Figure 12)

Figure 11: Gameplay clip of Mickey's Soundsational Parade from Disneyland Adventures *(The 8-Bit Arcade, 2018).*

Figure 12: Ambient sounds recorded during a playthrough of Disneyland Adventures *(Asobo Studios, 2017)*

However, we can recognise at least one way in which Disney idealises their soundscape through hearing the *lack* of familial arguments. Suggested in section 2.1.2 as one of the many potential reminders of the disjunct between the real-actual and the Disney-fantastical, it should not be surprising that familial arguments are missing from the idealised soundscape of *Disneyland Adventures*. Disney's conceit of family is well-known – Henry A. Giroux & Grace Pollock describe a "pretense to innocence" which is "predicated on the virtues of fun [and] family values…a cultural universe that is largely conservative in its values, colonial in its production of racial differences, and middle class in its portrayal of family values" (Giroux & Pollock, 2010, pp. 127–128) – and this conceit is present throughout the virtual experience. It is notable that a majority of 'guest' characters (i.e., the crowds) within the game are Caucasian, and that most of the families which wander around the park consist of a man, woman, and one or two children. There are almost no single parent families portrayed, and where they do exist, it is normally a female adult accompanying a number of children.[35] Whilst some areas of diversity are emphasised – there are several guest characters in wheelchairs – it seems that, at least subconsciously, the Disney conceit of 'traditional' (heterosexual, nuclear) families are portrayed. Familial arguments would not have a place within this conceit, let alone within the distilled 'perfect' Disney experience that we should imagine a game/advert like *Disneyland Adventures* to aim to present. As we should therefore expect, no familial arguments can be heard within the game. The majority of parental voices that can be overheard within the game display only parental concern (e.g., "Slow down a minute!" or "Are we feeling hungry?") – adult figures within the game are reduced only to their role as caregivers for their children, reflecting once again the audience of the game.

Despite the idealisation of the intended experience, sounds also emerge from the edges of the experience from the user perspective. Whilst the liminal sounds emerging from the Disney environment discussed in 2.1.2 included familial arguments and conversations between cast members, my experience of *Disneyland Adventures* is situated in my home office. Outside my window, the local boy racers compete to see who has fitted the loudest illegal after-market exhaust onto their sports cars. My friends, watching me stream on Discord or sat next to me at my computer, laugh and joke about the ways in which *Disneyland Adventures* sometimes feels like a somewhat dystopian attempt to convince children that their lifelong dream should be to have these magical adventures in reality by conducting a real actual pilgrimage to Disney, as discussed in section 1.3.1. The doorbell rings. The smoke alarm goes off. The neighbour practices violin. Similarities between these sounds and the liminal sounds in Disneyland can be observed: both disrupt the experience through introducing reminders of the world outside the virtual externality; the sounds "interfere with the reception of [the] message" (Attali, 1985, p.27) and, therefore, can be considered 'noise.'

Audio spatialisation should also be considered throughout *Disneyland Adventures*. The game is predominantly in stereo, allowing for easy location of points of interest throughout the park.

[35] Conversely, there are a great deal of *parentless* children present within the game – presumably, this is to emphasise the child-focused nature of the Disney experience, or perhaps even to suggest to children that Disney represents freedom from the (presumably buzz-killing) oversight of their parents.

However, the spatialisation is not always connected to the visual representation of the virtual world. Sound is frequently presented in a perspective distinct from the visual perspective of the camera within the virtual world. This is common within video games, as described by Richard Stevens & Dave Raybould:

> As the [perspective] cuts between each camera, the orientation of the sounds relative to each camera is different. This gives the effect of sounds suddenly panning form one side of the screen to the other. Although this may be strictly accurate, it is distracting to the player. The problem is, of course, well known in cinema, and to avoid it the sound track is often not panned according to the camera movements but instead maintains a separate continuity. For us to avoid this distracting panning during cutscenes, we need to essentially switch off the normal game sounds, which are heard from the point of view of the camera, and play a more cinematic sound track that we have more control over instead. (R. Stevens & Raybould, 2011, p. 157)

Similarities to the design of other video games can also be observed. For example, James Boer, audio programmer for the *Guild Wars* franchise, described how the resulting compromises in that game functioned:

> In games where the player's avatar can extend a long way from the camera, especially when the camera can move around freely, putting the audio listener on the camera can cause problems with identifying important sounds that are near the player, which is, after all, the most important focal point of the game. Thus, it causes a mental disconnect with the avatar. Battle sounds feel less impactful, and enemies right next to you sound more distant and weaker … However, if you put the listener directly on the character, it sounds a bit weird to hear audio source panning from the perspective of the player character. That is, in the worst case, a sound source between the camera and the player's avatar will render that sound as though it's coming from behind. This is especially disconcerting if you're using good 3D audio. (Boer, 2020)

Similar approaches are taken in *Disneyland Adventures* to allow players to more easily connect with their on-screen avatars: sonic spatialisation does not change rapidly as the camera is rotated around the character, and the placement of the camera does not determine the overall intensity of different sound sources within the virtual world. A continuous distinction between the heard sound and the visual world should also be noted: characters can be visually observed to be talking, but the heard sound emerges from somewhere else in the stereo spatialisation. Children run past, and the sound of laughter can be heard moving, but a disjunct between the movement of the children and the stereo spatialisation is almost guaranteed. The overall effect is an uncanniness which emerges from the *almost* synchronous connections between sound and image: the game does not feel quite real. Something is just a little bit wrong.

Some parts of the *Disneyland Adventures* game are presented in a two-dimensional world, rather than a three-dimensional world. Many other virtual experiences have been presented in a two-dimensional frame – e.g., *Dune II* (1992), which uses a predominantly top-down perspective and *Super Mario Bros* (1985), which uses a predominantly side-on perspective. Despite the two-dimensional visual perspective, audio designers sometimes choose to, nonetheless, present a stereo or binaural audio world. For example, Gabe Cuzzilo's *APE OUT* (2019) – in which players must take control of an ape escaping from an experimental laboratory – uses stereo audio spatialised from the specific perspective of the escaping ape, despite having a top-down perspective:

> **Gabe Cuzzilo:** This game has a full spatialised audio system – sound reflects off different surfaces differently. You can always tell which way you're facing from the ambi-sonic ambiences that are playing in the game.
>
> **Bennet Foddy:** Yeah, unlike most games from this [top-down] perspective, the audio listener is inside the head of the gorilla so when you rotate your character it actually rotates the soundscape around you [the player]. All the positioning stuff is relative to where the ape is facing as opposed to being like on the right of the screen or whatever… it should give you more of a chance to hear people around a corner because you're more embodied in the space. (IGN, 2020, sec. 00:10:35–00:11:45)

Cuzzilo and Foddy suggest that these approaches can help users (or players) to feel embodied within a virtual world as the audio will more closely reflect the sonic experience of the player characters. The opposite choice seems to have been made throughout most of *Disneyland Adventures*: the player receives a cinematic style audio mix during many of the minigames, placing them within the role of a cinematic spectator even when their character is clearly visible on-screen as part of the event. This may be to deliberately create a sense of distance between player and on-screen narrative to heighten the spectacle or perhaps even to subconsciously create an implication that these experiences are ones which should (or could) be seen at Disneyland (or other Disney theme parks).

We can recognise that *Disneyland Adventures* only presents a very specific vision of 'the sound of Disney'. The game distils and rationalises Disneyland to work towards a particular version of Disney, and adjustments in the sound are made in order to support and enhance this version of the Disney ideal. We can see how sound within *Disneyland Adventures* has been optimised to achieve the dual functions of advertising Disneyland (and other Disney theme parks) and supporting engagement with the virtual world. Whilst we can recognise disjuncts in the designed experience – such as in the decision to omit sounds emerging from liminal spaces at Disneyland within the virtual recreation of *Disneyland Adventures*, we can nonetheless recognise that the experiential perspective of *Disneyland Adventures* will have liminal sounds emerging from the virtual externality of the user. This case study has helped us to understand how sound can function within virtual experience, and – therefore – some of the ways in which the subjective phenomenon of sound is transformed when an experience is re-mediatised as a virtual experience.

One important question about the sound of *Disneyland Adventures* remains, however. Is this really 'the sound of Disney' that has been recreated? We have already discussed in 2.1.2. that there is a clear difference between *reproducing* Disney in the virtual (i.e., creating an exact replica of an actual experience) and *simulating* Disney in the virtual (i.e., creating an idealised version of the imagined experience). It is patently obvious that elements of Disneyland are omitted within the game: this is not a reproduction, but a simulation. To what extent does this actually matter? We discussed in section 0.3.1 the concept of immersion and identified that one element of perceived immersion was authenticity: does that mean that *Disneyland Adventures* would be more authentic if it introduced these 'noisy' elements of Disneyland? The next chapter examines the importance of authenticity to virtual experience: why is authenticity important, what affects the perception of authenticity in virtual experience, and what issues might occur as we try to create and sustain authenticity within virtual experiences. This will help us to recontextualise the axes of sound and virtuality in relationship to each other, combining the framework established throughout Chapter 1 with the understanding of sound constructed in Chapter 2, in order to address the fundamental research question of how sound and music function within virtual experience.

Chapter 3: Authenticity

3.0.0: Noise, Authenticity, and Immersion[36]

The previous chapter set out what sound is, how sound can be modelled as a soundscape and the process of sound design for interactive worlds. Sound design was presented as a process through which we create and shape an auditory experience in order to generate user-specific soundscapes. As these soundscapes are experiential, considerations in the process of sound design should be considerations of how sound is (or will be) received by users within virtual experiences.

This chapter investigates the way in which sound interrelates to areas of virtuality presented in section 0.3.0. In this earlier section, we identified that perceived immersion is dependent upon the credibility of the user experience of the virtual experience (Slater, 2009; Slater & Wilbur, 1997; A. Wang et al., 2021), in what is sometimes referred to as 'authenticity'. Relationships between authenticity and sound can also be seen in the exploration of 'noise' presented in section 2.1.2: 'noisy' sounds are those which disrupt perceived immersion, or which the user finds non-credible. This includes both explicitly disruptive sound – i.e., those which the user is able to identify as being incorrect in some way – and implicitly disruptive sound – i.e., those which simply 'feel wrong' to the user, even where they cannot identify exactly why. That is to say: noise is sound which detracts from the virtual experience because it is inauthentic to the virtual experience. This general example demonstrates one of the reasons that authenticity must be considered when investigating audio for virtual experience.

The next chapter of this thesis investigates those relationships between authenticity and sound across and within the tripartite structure for virtuality developed in Chapter 2. A common area in which authenticity is particularly valued – and is perhaps particularly difficult – is virtual reality experiences intended also to be historically authentic (i.e., historically accurate) experiences: Karin

[36] Significant portions of this chapter are included as a chapter in *History as Fantasy* (ed. James Cook, Alexander Kolassa, Alex Robinson, and Adam Whittaker), Abingdon: Routledge, Forthcoming.

Bijsterveld rhetorically asks “who does not occasionally dream of stepping back into history and feeling truly immersed in the past?” (Bijsterveld, 2015, pp. 74–75). Bijsterveld has a specific focus on the need to create an ‘authentic’ historic experience to achieve this implied goal. By considering the issues and challenges experienced by those working to recreate historical experiences, we can more easily assess how issues and challenges may emerge whilst designing sound for other virtual experiences across the axes of virtuality (established in Chapter 1) and sound (outlined in Chapter 2).

This chapter will first examine sound design for historical experiences to explore how inauthenticity emerges throughout virtuality: for example, how is authenticity challenged by the virtual hardware used for sound in virtual experiences? This chapter develops an understanding of authenticity as a construct formed through the combination of multiple paradigms of authenticity extending across all elements of the virtual experience. Particular attention will also be given to the technically difficult realisation of binaural sound within virtual reality experiences; a known area of interest for future research in sound for virtual reality due to the significant issues that it poses. These areas will also be considered from an experiential perspective: how do the users perceive within virtual experiences and what effect does this have on their perceived immersion? The fundamentally subjective nature of authenticity in virtual experiences will be discussed, through the examination of the contentious field of historically informed performance and through sound for historical virtual experiences. As authenticity is key within these areas, we can use them to examine broader questions about authenticity (and plausibility): we can use them to consider how users will hear (i.e., ‘experience’) sounds within the virtuality. As authenticity is an important part of the user’s perceived immersion, understanding how authenticity is constructed and how inauthenticity can be avoided will allow us to consider how we can make experiences more immersive and therefore improve the outcomes of those experiences.

3.0.1: Authenticity in sound design for historical experiences

Creative aspects of sound design were explored in section 2.2.1. However, we can recognise that there is an implicit understanding that sound design involves – to some extent – recreating the sound of ‘what was really there’. Hyper-realistic approaches, for example, focus on creating a version of the events which recreate an exaggerated version of ‘what was really there’ in what we could perhaps justify as being some form of ‘what we remember being there’ or ‘what is important for us to realise is there’. Sound for virtual experience requires, to some extent, sonic authenticity as a consideration within the virtual software: i.e., we must understand the rules of interaction between user and experience and generate an auditory signal that provides them with the intended interaction. Correspondingly, we can imagine that, in attempting to create an authentic sound design for any experience, sound designers are likely to have spent a considerable amount of time attempting to find methods through which they can recreate specific sounds (or, at least, the sonic *affect* of those sounds).[37]

However, when creating historical sounds, an underlying absence of information about the sound can be observed as sound recordings have only existed in very recent history. Thomas Edison’s 1877 phonograph is often considered to be the beginning of modern sound recording. Although some earlier prototypes of sound recording devices exist (e.g., the phonautograph, patented 1857), it was Edison who first began to commercialise devices that could play recorded soundwaves back through another device.[38] No matter the specific technology, however, these early sound recording devices were limited – both in popularity and in capability. Furthermore, any physical media containing recordings made using these early technologies will have degraded significantly over the last 150 years. Therefore, whilst these recordings can be used to inform our understanding of the sounds of more recent times, the reality is that where an ‘authentic’ soundscape is desired, then nearly all historical sounds must be (re)created through trans-medial processes: converting non-auditory

[37] As described by Lisa Coulthard: “As any Foley artist will tell you, the object used in the image is not necessarily representative of what will sound most effective or even convincing: the squish of a melon has the correct emotional, tonal, and affective impact [as the sound of a head being crushed] and that is the most significant thing” (Coulthard, 2016, p.186)

[38] Roger Beardsley and Daniel Leech-Wilkinson (n.d) offer an excellent review of early sound recordings but appear to have been unaware of the work of Edourd-Leon Scott de Martinville – who created the phonautograph several decades before Thomas Edison created the phonograph. Information about the phonautograph is available in Patrick Feaster’s annotated discography of phonautograph recordings (Feaster, 2010). Many early sound recordings (including phonautograph recordings) have been preserved by *First Sounds* (www.firstsounds.org).

information (such as a written ‘soundscape’) to an auditory event such that sound effect and sound affect are conjoined.

Specific projects have faced significant challenges in their attempts to be ‘authentic’. Bijsterveld recounts, for example, the efforts made to create an authentic-sounding car sound for a museum exhibit: an original 1928 car was sourced along with a patch of asphalt similar to that from 1930. The car was then recorded driving over the asphalt to create an ‘authentic’ sound that would have been heard in the target time period intended for the soundscape. (Bijsterveld, 2015, p. 81). The intention was that by using an original car and original asphalt, the original sound could be re-created through simulation. Similar processes are used to create the sound of a horse-drawn tram running on rails and the sound of a handcart being pulled over cobblestones (Bijsterveld, 2015, pp. 80–81).

Similarly, work on Ubisoft’s *Assassin’s Creed Odyssey* (2018) – alongside other instalments in the franchise – attracted much attention for the claim that it was ‘historically authentic.’ Lydia Andrew, the Audio Director for the game, recalls that “Everything was researched…We wanted to do some recordings to recreate these ancient wooden boats, but most boats now are built out of metal… We did a lot of recording of foley work, we used our sound effects libraries, but we also went on a recreated 16th-century galleon… That boat was built [entirely] of wood and so it gave us the opportunity to record what a wooden boat sounds like, and how it moves through the water” (Ubisoft, 2018). The effort should be applauded. However, to what extent does the sound of a 16th-century galleon truly represent the sound of the ancient wooden Greek triremes? Galleons relied on wind-power with large numbers of masts. Whilst some triremes had sails, their primary form of propulsion was oars. Galleons were intended as long-distance travelling ships, capable of crossing the ocean for purposes of war and trade. Triremes were ramming boats intended for power projection in the relatively smaller region of the Mediterranean. Galleons were typically constructed predominantly out of heavy woods such as oak, capable of withstanding long exposure to sea water. Triremes required beaching on a regular basis – almost daily – to prevent waterlogging and, as such, were predominantly constructed from light woods such as fir and/or pine as to allow small numbers of men to bring the boat ashore. These differences also expressed themselves in their construction: triremes were long and narrow, capable of high speeds; galleons were broader and wider, offering more stability in stormy weather. Whilst the exact degree of auditory inaccuracies in *Odyssey* emerging from these differences cannot be easily known, it seems likely that the collective effect of these differences would be a substantial change in the overall sounds heard whilst travelling by boat.[39]

However, these considerations of auditory authenticity are expressions only of what should be placed within the auditory environment of the virtual experience. That is to say: what sound sources existed and what did sounds did those sources generate? These are considerations of *sound content*: which sounds should be heard? How can we ensure that users understand this sounds properly? How can we approach and utilise virtual software in auditory domains of virtual experience? As discussed in the introduction of this chapter, we should recognise that issues can also emerge from other areas. For example, inauthenticity can emerge from auditory infidelity caused by the specific hardware used for specific experiences.

Other issues with audio which can implicitly challenge authenticity in virtual experiences are inherent to the process of sound design (and the implementation of that sound design within virtual experiences). Other concerns surround external sounds which may enter into the virtual externality. These sounds, where they permeate the ‘magic circle’ of experience, challenge the authenticity – and immersion – of the experience. These are considerations of *sound implementation*: the processes through which the sounds are delivered to the participant. Unlike questions of sound *content*, which are broadly applicable to all types of (re)creation, sound *implementation* tends to have more specific considerations when considering virtual reality technologies. To some extent, this has already been hinted at in our examination of early recording devices: we identified that some recorded sounds were not ‘true to life.’ Has this changed? Do modern recording processes capture a more ‘authentic’

[39] The information provided in this paragraph are well-known general statements about these historical ships intended to illustrate some of the many differences between them. For those seeking more detailed information about triremes, introductory information on their construction, purpose and use can be found in Meijer (2014), and Morrison & Coates (1986). Similar information on galleons can be found in Kirsch (1990).

version of the real-actual sound? The next two sections further examine issues at the intersection of sound implementation and authenticity in the software, hardware and externality of a virtual experience.

3.0.2: Sonic authenticity in hardware for virtual experience

It is generally accepted that the process of recording and playing back sounds will have unavoidable artefacts of mediation. Both microphones and speakers will alter some of the acoustic properties of sound. Unlike electrical interference or background noises being captured by microphones, these issues are inherent in the recording process. Different microphones respond differently to varying frequencies of sound. As Barry Truax states: "Perfect fidelity, of course, is technically impossible to achieve because every stage of the signal transfer process inevitably adds noise or distortion to the signal, however slight (or gross, judging by what is commonly heard over cheap loudspeakers daily). Thus, like the "true satisfaction" of the advertisements or the promises of other types of slogans, fidelity is unattainable, hence safe and unquestionable" (Truax, 2001, p.10). For example, many microphones struggle to capture sounds below 40Hz. Similarly, different speakers have varying capabilities at reproducing certain frequencies – or combinations of frequencies – of sound.

These challenges to authenticity resulting from mediation in technological processes are particularly relevant when considering the types of speakers used in many Head Mounted Display Virtual Reality (HMDVR) experiences: HMDVR experiences typically use head-mounted speakers. Unlike cinema-style surround sound or hi-fi stereo sound systems, these head-mounted speakers often have a limited frequency response as only small speaker cones fit within headphones. The small size of the speaker cones mean that frequency response may be poor, especially in bass frequencies where auditory wavelength is very large. Solutions such as artificially boosting bass are ineffective at creating an authentic sound as this solution can lead to distortion and speaker feedback. The use of open-ear headphones, such as those found on many HMDVR devices such as the Valve Index, may also minimise the effectiveness of these solutions. Whilst techniques such as the "false fundamental"[40] can be used to improve sound quality, there is not currently a way to completely avoid frequency-response limitations in the implementation of sound design within virtual reality.

Broader issues with hardware can also be identified. Sounds are likely to contain some audial information as a result of the mediating processes of recording and playback. The recording location is likely to impart some acoustic properties onto any recorded sound: for example, the same noise recorded in a large room will have more reverberation than one recorded in a small room. It is easier to add, rather than remove, reverberation from a recorded sound, meaning that sounds intended to evoke a low-reverberation environment may be difficult to implement authentically.

This can be problematic as spatial properties of sound affect perceived immersion. Pontus Larsson et al. suggest that room acoustic cues improve immersion as using incongruous acoustic cues may lead to the user's sense of presence being disrupted (Larsson et al., 2010, pp. 155–156). This is likely due to a (sub)conscious awareness that incongruous spatial acoustic properties are inauthentic (or otherwise implausible), or due to an increase in the challenge of interpreting those acoustic properties (with a commensurate effect on challenged-based immersion). Therefore, an 'ideal' audio solution for virtual experiences might be considered as one in which these spatial acoustic properties are fully congruous with the presented virtuality, as this would ensure that users can draw upon the "exosonic" components of sound highlighted by Grimshaw & Garner (2015) and discussed in section 2.1.0. One approach would be to track the physical vibrations perceived by a listener as noise are between sound source and listener, with the sound being dynamically altered resulting from interactions (e.g., reflections, absorptions, etc.) with objects within the virtual scene. This would require both building an acoustical model of the room and having a version of the recorded sound that has minimal acoustic reverberation.

[40] The 'false fundamental' is a psychoacoustic effect which convinces the human brain that a bass frequency is present through playing back harmonics that the brain would expect to hear only if the bass note was present. A discussion of issues specific to 'false fundamentals 'is found in section 5.3.0 – 5.3.2)

However, creating accurate acoustic models to generate realistic audial representations of virtual spaces presents pragmatic issues. Firstly, there may be insufficient information to generate these models within the/a historical context. The overall acoustic of a room is determined by many things, from the room's shape and size to the materials used to construct it. Objects within the room may also have an effect – e.g., the presence of curtains dramatically alters the acoustics through increasing sound absorption. When examining historical environments, the required information may be incomplete. Nikos Barkas notes, for example, how scenery will "improve the acoustic comfort in most [ancient Greek] theatres" but that "it is difficult to compile an exact record of the types of scenery…because they have been completely destroyed or because they have been incorporated into later reconstructions" (Barkas, 2019, pp. 349–350). As a result, acoustic models of ancient Greek theatres – whether for physical or digital reconstruction – must make assumptions about the scenery of Greek theatre, which will lead to potential inauthenticity. Attention should also be given to the way in which these properties of historic buildings change over time. Dario D'Orazio et al. notes that "the acoustics of an historical opera house needs to change during all its life" with particular attention given to the way in which the introduction of the orchestra pit in the late 19th-century dramatically altered many performance locations (D'Orazio et al., 2019, pp. 694–695). Similarly, even small changes to the materials used to (re)construct buildings can result in drastic acoustic changes, as seen in the changes in acoustic resulting from differing types of plaster used to maintain the Süleymaniye Mosque (Sü Gül, 2019). Whilst this is only a limited selection of the ways in which available information about historical environments may be incomplete – or even incorrect! – the overall effect can be seen: an accurate acoustic model of a specific building at a specific moment in its chronology can be very difficult to re-create.

Secondly, even where estimations of acoustic models can be created, dynamically processing audio in real time places extreme demands on computer hardware. This often affects overall hardware performance within the virtual experience, which can lead to a loss of immersion as other aspects of the virtual experience such as the visuals also require computing power to maintain plausibility in presentation. Richard Stevens and Dave Raybould discuss this difficulty, with a specific focus on computer games, and ultimately conclude that – as of 2015 – it is not feasible: “The sheer complexity of calculating the scattering of sound waves across an intricate virtual scene remains out of reach of the real-time demands of video games” (R. Stevens & Raybould, 2015, p. 60).[41] Audio designers for virtual experiences must compromise between audio quality (and the authenticity of their reverberation) and a computer processing budget imposed by the limitations of the systems used for their virtual experiences. Typically, solutions revolve around pre-rendering acoustic qualities such as reverberation: i.e., the virtual experience plays different versions of the sound on demand, rather than using a single sound and dynamically modifying its acoustic properties. For example, R. Stevens and Raybould identify in their study of video game sound that “the items of key importance…rely on a system of pre-rendered sounds to evoke the acoustic response of the different spaces within the game” (R. Stevens & Raybould, 2015, pp. 60–61). Whilst this has the benefit of being significantly less demanding on computer hardware, the process may nonetheless fail to represent the acoustic environment authentically – especially where users are able to interact with the virtual experience freely to change the items contained within a virtual space. For example, sounds in the simulation may have identical acoustic responses irrespective of how their positionality within the virtual space should properly affect it.

Issues also exist when trying to record sounds without existing acoustic reverberation. The obvious solution to this issue is to record sounds in an anechoic chamber (where there is no reverberation) and then use software to additively generate synthetic reverberation to create an ‘authentic’ version of the sound for any given reverberation. This would allow for sounds to have more ‘authentic’ spatial cues. Whilst this may be possible for some sounds, it is far from practical for others: for example, we can imagine that the *Odyssey* team may have struggled to fit an authentic

[41] R. Stevens & Raybould appear to be basing their expectations of the practicality of fully rendered acoustic modelling on the work of Nikunj Raghuvanshi et al. (2010). Similar concerns arose surrounding the processing of virtual light for ‘photo-realistic’ shading in real-time graphic rendering, but sophisticated real-time environmental processing for light has now been implemented in latest generation game engines (e.g., Unreal Engine 5). Some of the research on this technology will also have helped optimise acoustic modelling methods, which may mean that we are now significantly closer to these ‘impossible’ acoustic models.

wooden sailing ship within an anechoic chamber.[42] Furthermore, if any of the recorded sounds include human performance – e.g., singing or speaking – then the use of an anechoic chamber may fundamentally alter aspects of the performance in unexpected ways.

A holistic overview of how performances are affected by the acoustics of the recording space is still an area of active research. However, one well-researched example is how singers alter their performances in response to the acoustic space within which they are performing. For example, singers often find it more challenging to perform in ensemble in low or zero reverberant conditions such an anechoic chamber (Braak et al., 2019). This may explain why singers sometimes change their performance (sub)consciously when performing in anechoic chambers (Luizard et al., 2018). Whether part of the cause of these issues, or an entirely separate issue in itself, anechoic chambers are also frequently uncomfortable: in their reflections on a recent research project, James Cook et al. describe anechoic chambers as “claustrophobic, hot, [and] lacking in [vocal] feedback” and note how close-microphone recording techniques can be “terrifying” for performers in a way that is “distinct from, and more challenging than, recording in any other space” (J Cook et al., forthcoming, pp.20–21). Therefore, singers are unlikely to perform in a way authentic to the acoustic of a recreated (or imagined) performance space when singing in an anechoic recording space.[43]

As a result of these difficulties in creating anechoic recordings, it seems unlikely that using reverberation generated through acoustic models will ever completely capture the objective reality of what it ‘would have been like’ to visit a particular historical location. These difficulties are further complicated by issues inherent to the recording process. Through each stage of mediation – recording, production, playback – fundamental changes are made to the underlying audio information, which can combine to create objectively inauthentic sounds. Whilst the resultant inauthenticity may not always

[42] Moreover, should the sailing boat be in water, then it should the chamber would no longer be “anechoic” as bodies of water are acoustically reflective.

[43] Some possibilities for resolving these issues have been investigated. J Cook et al. (forthcoming) investigated different methods of providing auditory feedback to musicians (e.g., including reverberation on studio headphone monitors, similar to methods used in Luizard et al., 2018) before settling on post-recording review. This suggests that solving issues within the anechoic chamber was not feasible.

be consciously human perceptible, it seems likely that users will be subconsciously aware that sounds are not entirely authentic and, therefore, find it more difficult to immerse themselves – sensorially or narratively – within a virtual experience.

3.0.3: Sonic authenticity in externalities for virtual experiences

Other forms of authenticity can also be considered. So far, we have considered how we might create authentic sounds and how variation in audio hardware can change the perception of authenticity within the virtual experience. Consideration should also be given to issues in the sonic authenticity of the virtual externality. We have already discussed some virtual experiences which were created using head-mounted displays for virtual reality (HMDVRs). Other virtual experiences include alternate expressions of virtual experience, such as augmented reality, where the participant is provided with additional information to further contextualise a 'real' experience, such as hearing sounds when facing a painting (as in the experience for Amsterdam Museum produced by Bijsterveld). Different formulations of the virtuality will require different audio solutions to ensure participants experience perceived immersion within the experience as the externality of these virtual experiences will interact in different ways: e.g., an augmented reality experience does not typically aim to sensorially immerse the user, whereas a virtual reality experience does typically aim to sensorially immerse the user.

As discussed in section 1.2.1, it is common to envisage immersive experiences as taking place within some form of magic circle – based loosely on fundamental work by Johan Huizinga (*Homo Ludens*, 1938) and popularised within video game studies by Katie Salen and Eric Zimmerman (*Game Design Fundamentals*, 2004). As discussed in section 0.3.0–0.3.1, these rationalisations of virtuality use immersion to describe the extent to which the virtual experience isolates the user from reality and/or how well the virtuality is perceived as real by the user – i.e., how well does the magic circle protect the user from the real-actual? In pragmatic terms, visual information can easily be blocked out by something as simple as a piece of cloth. Our visual realities are easily separated from a created virtuality. Conversely, sound (and therefore noise) conveys information through physical vibration, which means that it will pass through physical materials. This property of sound can make it hard to audially isolate the user from the real-actual (i.e., sensorially (audially) immerse the user) as isolating users from sound is therefore a much more involved process: placing a participant within a chamber encased within a vacuum may be an effective way of isolating them from the external word, but it also introduces significant technical challenges and potential safety issues.[44] Therefore, our external auditory realities converge with the created auditory environment within the virtual world.

In recent years, more attention has been given to the permeability of the boundaries that exist between reality and virtuality. For example, Edward Castronova argues that "the membrane between synthetic worlds and daily life is definitely porous, and this is by choice of the users" (Castronova, 2006, p. 166). In practical terms for virtual reality, permeability is likely a necessary part of approaches to virtuality. The danger of completely blocking out the outside world has been humorously explored in media surrounding virtual reality. In one parody commercial for the Oculus Rift, a man playing a virtual reality game is unaware that his flat is slowly being emptied by two burglars until he emerges after completing the game session to discover everything has been stolen. Shrugging, he returns back to the virtual world, with the punch line "Who Needs Real Stuff Anyway?" fading in next to the Oculus Rift logo (AudVid Bros., 2016). Less humorously, we might imagine difficulties and dangers emerging from participants being unsure whether important sounds (such as fire alarms or doorbells) are part of the virtual experience. The conflict between a need for some level of permeability and a desire to provide a nonetheless authentic audial experience to participants means that sound can be considered a particularly difficult issue within immersion: as set out in section 2.2.2, we should expect that a minimal level of 'noise' must typically be accepted within virtual experiences, even though it creates issues for the authenticity of those experiences. In order for users to engage with virtual experience safely, some sounds *must* be conducted from the external space to within the user experience; sound must be allowed to cross the boundary of the

[44] Whilst noise-cancelling headphones are a popular solution to this issue, they are far from perfect and could perhaps be more accurately described as "noise-reducing headphones."

magic circle, even if this may lead to an increased risk of audial hallucinations or experiential disruption.

Notwithstanding this, Mads Walther-Hansen and Mark Grimshaw-Aagaard argue in favour of isolating the user from external sounds where possible: "An intelligent reduction of our sensory and thus perceptual field will improve the differential processing of external stimuli (selective attention) and will allow for better concentration, focus, and ability to achieve flow and presence" (Walther-Hansen & Grimshaw-Aagaard, 2020, p. 3). Whilst they argue from theoretical grounds, it should also be noted that there are pragmatic benefits as well: simply turning up the volume to drown out external noises is likely to damage hearing (RNID, n.d.).

However, practical difficulties in using these approaches can be identified. Are there successful approaches to effective audio isolation for virtual experience participants? One widely explored technology is active noise cancellation (ANC). Frequently found on headphones, active noise cancellation uses an external microphone to identify and capture external sounds, and then use phase-inversion to calculate a signal which interferes with the signal generated by external sounds to remove them for the user.[45] By reducing external stimulus, the user is able to focus better on the intended stimulus from the virtual world. However, the viability of this technology outside of headphones is questionable: one proof of concept study was conducted in 2018 using Internet-of-Things hardware to 'intercept' sound waves and generate noise-cancelling sounds for the user of a device further away. This 'lookahead' active noise cancellation system (LANC) worked well under experimental conditions, but questions surround the capability of the technology in environments with multiple noise sources, the portability of the technology, radio frequency interference, and head mobility (Shen et al., 2018). Further concerns could also be raised: whilst LANC functions for an individual listener at a specific location, it seems likely to be ineffective at controlling external stimuli for *many* listeners across a location, or even for a single listener moving around within a room.

[45] See Triggs (2021) for a review of current ANC technologies.

For virtual experiences where headphones or other head-mounted devices are not a possibility, alternate solutions must clearly be found. Some research has shown that auditory masking is an effective technique: participants are played some form of noise (e.g., white noise, pink noise, brown noise). Walther-Hansen and Grimshaw-Aagaard suggest this may be because introducing unstructured sounds to an experience allows the user to ignore external stimuli more effectively; user's awareness of the external world is reduced as sounds are hidden by the unstructured sound (Walther-Hansen & Grimshaw-Aagaard, 2020). However, it is not clear whether this approach to audio isolation can be widely adopted as a solution for audio isolation. For example, Simon Banbury and Dianne Berry suggest that the volume of white noise needed to create audio isolation within an office may be impractical for use (Banbury & Berry, 2005), which would likely be an issue in many virtual experiences; similarly, Ravi Mehta et al. identify that white noise may have a negative effect on user cognition (Mehta et al., 2012) which would also detract from virtual experiences. It should equally be observed that the addition of noise to the virtual experience is deliberately adding an inauthentic sound to most environments: it is possible that using these additive approaches may conflict with authentic sound design. Whilst more similar frequency-range structured sounds can be introduced in some settings (e.g., flowing water, rain, etc.) to try to create a similar effect using more-authentic sound sources, auditory masking may not be an effective approach for those seeking to create authentic virtual experiences.

3.0.4: Issues with binaural implementation for virtual experiences

Three areas of authenticity in sound design have been explored so far in this chapter: the authenticity of the sound emerging from the 'software' of the experience, the authenticity of the sound emerging from the hardware used to generate the experience, and the way in which the inauthenticity of noises within the externality must be considered during an experience. We can therefore understand that authenticity emerges from and is challenged by all elements of virtual experience. Other areas of authenticity emerge from a combination of two or more elements of the framework: one such example is binaurally rendered audio.

Binaural sound solutions use two speakers mounted close to the ears (e.g., headphones) to playback algorithmically altered sounds which give a facsimile of perfect surround sound: sound coming from any direction and distance in three-dimensions. For example, a sound which is intended to seem like it is coming from the right-hand side of the listener, will be played back fractions of a second later through the left-speaker than through the right-speaker. This combination of hardware and software tricks the human brain into believing that the sound originated from their right-hand side, even though it was played through both speakers essentially simultaneously. Latest generation binaural audio implementations use these interaural time differences (ITDs; i.e., a difference in the time a sound is heard by each ear), interaural level differences (ILDs; i.e., a difference in the volume of a sound heard by each ear), and other differences to generate a variety of head-related transfer functions (HRTFs) which dynamically alter properties of the sound to model how it would be changed as a result of travelling around and through the spaces between a listener's ears.[46]

Binaural audio implementation is perceived as providing a potentially more authentic user experience than stereo audio implementation as it should provide the user with a greater sense of presence within the virtual reality experience, as it more accurately represents what a human would subjectively hear within a given virtual reality. That is to say: binaural audio implementation can increase the user's sense of *embodiment* within a virtual experience. However, realising this potential is difficult. Most HRTFs used for products released onto the consumer market are non-personalised. These non-personalised HRTFs use audio calculations measured from a dummy head mannequin or other individual (often drawing on publicly available acoustic models, i.e., LISTEN, CIPIC, and ARI). This means that users are not hearing an authentic recreation of what *they* would (expect to) hear, but rather a facsimile based on a non-specific model. The use of non-personalised HRTFs reduces the accuracy and specificity – and therefore the authenticity – of binaural audio implementations. Further, a 2020 roundtable discussion on HRTFs highlighted that the current models

[46] For more information about the specific mechanics of binaural audio, Møller (1992) reviews early binaural technology, Nowak (2018) examine developments within 21st-century binaural technology, and Pike (2019) provides an overview of current binaural technology. Other approaches to individually spatialised sound are also being developed: see Hong et al. (2017) and Zhang et al. (2017) for other spatial audio solutions.

used to generated HRTF-algorithms are (still)[47] based on head-shapes found predominantly within Europe and North America (Summers et al., 2021, pp. 75–76). Users from outside these regions may, therefore, experience issues when using non-personalised HRTFs.

The range of issues created using non-specific HRTFs is still being established by on-going research. Research in this area is difficult, as listening is – as discussed in Chapter 2 – fundamentally a subjective experience. In experimental terms, this means that many experiments assessing the specific impact of changes to HRTFs suffer from reproducibility issues, as many users struggle to consistently rate assessment factors during experiments (Andreopoulou & Katz, 2016). However, preliminary studies seem to show that non-specific HRTFs impact a user's ability to distinguish speech and other sounds in high-noise environments (Cuevas-Rodriguez et al., 2021), affect the perception of brightness and other timbral qualities of sound (Armstrong et al., 2018), and may reduce distance estimation accuracy (i.e., how far away a sound source is) in virtual environments (Arend et al., 2021). It seems likely that there are other effects that have not yet been experimentally tested. Each of these issues can disrupt a participant's experience of a virtual world. However, generating personalised (or specific) HRTFs that accurately model the exact differences that someone would perceive between their two ears is currently a time-consuming and technically difficult process which typically requires the assistance of an engineer with specialised tools. Whilst some efforts have been made to create personalised HRTFs using images and videos of a subject (e.g., Genelec's *Aural ID*), it is unclear the extent to which these systems improve upon non-specific HRTFs and, therefore, whether audio designers can create an automatic method to generate personalised HRTFs

Alongside the difficulty of creating specific HRTFs, contemporary HRTFs also struggle to accurately modify sounds for near-field spatial frames. Current binaural spatialisation frequently takes advantage of the fact that reverberation for sound signals travelling in large space is chaotic. In many cases, humans perceive these chaotic signals as general room reverberation, which can be approximated with a single reverberation layer. However, this only applies to 'far-field' audio – that

[47] Awareness of this issue has been present since the early days of binaural technology. See Rumsey (2011).

which is far enough away from a VR participant that they are unable to determine the distance of the sound accurately. For near-field audio signals (i.e., those within the peripersonal space, defined broadly as the space surrounding the body where we can reach or be reached by external entities), distance becomes an important factor in the rendering of binaural sound. This is for several reasons. First, sound intensity is inversely proportional to the square of the distance between signal and receiver. As explained by Laurent Betbeder, a member of the development team working on *Oculus* HMDVR technology, these effects mean that sounds need greater dynamic variance in near-field spatial frames. There will also be much greater difference between the initial signal sound and the resultant reverberations (i.e., the listener will perceive much less reverberation for a sound close to their ears). Second, sounds close to the head are also more affected by the presence of the head itself as low-frequency sounds diffuse around the skull and high-frequency sounds do not; there will be greater differences between the sound heard by each ear in the binaural mix (Betbeder, 2017).

Alongside these dynamic considerations, the closeness of the sound source means that the size of the sound source may also have a significant effect. Sounds within auditory environments are typically modelled as emerging from a singular point, as this is how far-field sounds are perceived in most cases. However, this is an inaccurate model: sounds actually emerge from distinct points within a system or object. For example: the sound of a car engine is created from the combination of many individual sounds, such as moving pistons, burning fuel, and flowing coolant. If you were sufficiently close to a car engine, then you might be able to identify the precise location of each of these sounds, and the dynamic balance between each of these sounds (i.e., the sound mix) would change depending on your precise location. Current sound implementation solutions rarely allow for these effects to be simulated: i.e., the recreated sound of a car engine is a singular audio file, rather than the combination of many individual audio files. As the individual sounds do not exist within the game engine as audio files, the individual sounds cannot be dynamically balanced in real-time.

Similarities to the earlier discussion surrounding accurate acoustic modelling can be drawn. The fidelity of digital images is often discussed through the lens of resolution (i.e., dots per inch /

number of pixels). We could see the aggregation of sound cues as a type of ‘sonic resolution’. Higher resolution auditory environments would use a greater number of individually spatialised sounds, whilst lower resolution auditory environments would use a greater number of aggregate sounds.[48] A major issue for near-field sound, therefore, is that current sound implementation has insufficient sound resolution to resolve near-field spatialisation accurately.

Finally, Olli Rummukainen et al. identify two modes of interaction within near-field spaces: a user may approach (or be approached by) a stationary sound source and inspect it without direct interaction; or a user may choose to move the perceived origin of the sound source, or move around the perceived sound source (Rummukainen et al., 2019, pp. 448–449). These non-stationary (‘dynamic’) sound sources are particularly problematic within near-field spaces. Lawrence Rosenblum et al. have previously identified that observers may gain additional information about a near-field sound (e.g., be better able to estimate the distance between the listener and the sound source) as a result of the changing acoustic dimensions of sound which are perceptible when the user or the sound source are dynamic (Rosenblum et al., 1996, pp. 12–14). As a result, we can imagine that near-field audio rendering is particularly problematic when the listener, the sound source, or both are dynamic; the listener may (at least subconsciously) expect to be able to determine additional information about the sound source, and not be able to do so. This disruption of expectation is likely to reduce sensorial immersion as it will make the sounds feel less genuine (i.e., less ‘authentic’). These issues could potentially introduce other issues for players. For example, we might imagine a multiplayer virtual reality game in which players must throw a ‘ticking bomb’ at each other. Within

[48] An inherent implication of this is that dynamic resolution variance may provide some solutions in minimising the computational requirements of advanced audio solutions. Current generation visual processing feeds use dynamic resolution variance to aggregate digital objects far from the perspective of the viewer: as a virtual observer approaches an object, a version of the object with a higher number of polygons (i.e., more detail) is loaded into the game engine – for more information, see Unreal Engine (2021). Similar approaches could be used within sound implementation. For example, as an observer approaches a crowd, the sound engine could transition from an low-resolution (aggregated) “crowd noise” to a high resolution (de-aggregated) “crowd noise” where each member of the crowd is individually modelled. These approaches are not currently used within game sound engines, which instead take an approach which combines ambient noise with specific spatialised signal sounds from points of interest. However, the more technically complicated approach seems likely to be an area of interest in future acoustic simulation for virtual reality.

the virtual world, players may find that they experience more difficulty than expected in catching the 'bomb' as the sound will inaccurately represent the location of the bomb whilst it is within their peri-personal space (and, potentially, due to technical difficulties in tracking physical movement precisely within virtual reality). The inauthenticity of spatialisation of sounds within peripersonal regions can make it difficult for players to judge the distance, movement, and speed of a virtual object. We can frame this inauthenticity as presenting an implausible challenge to interpreting audio in virtual experience (i.e., the user cannot rely on their subconscious understanding of spatial properties of sound) and therefore disrupting challenge-based immersion.

Joseph Tylka identifies several solutions which address these issues in binaural recordings: near-field sound sources can "be identified, localised, and isolated from the rest of the recording [so that] they can be processed and rendered separately" (Tylka, 2019, p. 7), or interpolation between multiple ambisonic microphones can be used (Tylka, 2019, pp. 7–8) although this may lead to issues as the interpolated signal may not account accurately for the positionality of objects within the soundscape (Tylka, 2019, pp. 102–103). However, Tylka is specifically referring to sounds recorded using binaural microphones, and it is unclear how these approaches can be used within experiences where the overall auditory environment is constructed within a virtual environment from a combination of distinct sounds, rather than being recorded directly in a sound studio. Camilo Arevalo et al. (2018) and Arvealo et al. (2022) suggest an alternate approach to near-field spatialisation which places near-field sounds endocentrically (i.e., in the centre of the head) which they report as more successful than other spatialisation approaches. However, we can imagine that endocentric spatialisation is unlikely to be effective at communicating information about *dynamic* sound sources in near-field regions, as an endocentric spatialisation would not adjust as an object moved within peripersonal spaces: i.e., the user would not be able to hear an object move from directly next to the left ear to directly next to their right ear with a purely endocentric spatialisation.

HRTFs – and binaural spatialisation more generally – therefore pose challenges to authenticity within sound for virtual reality and other virtual experiences. Whilst potential solutions continue to be suggested (e.g., Assenmacher et al., 2004; Lentz, 2008), these have not yet resolved all issues. Therefore, though binaural spatialisation seems likely to generally enhance the plausibility (and hence user-perceived authenticity) within virtual experiences, the current implementations of binaural spatialisation still have obstacles to overcome to ensure that users do not perceive inauthentic (i.e., immersion-disrupting) audio artifacts emerging from the techniques used to generate binaural spatialisation.

3.1.0: Faux-thenticity

Throughout the previous sections, we identified that there are many issues in authentic sound design for virtual experiences. Two questions therefore emerge. First, is achieving a completely objectively authentic experience possible? At present, it seems likely that objective authenticity – i.e., recreating what was actually there – is not a realistic goal: many of the issues outlined above are the result of contemporary limitations, technical processes, or imperfect simulation technologies, and therefore cannot be easily addressed. We can also recognise that there will always be some doubt in the accuracy of our understanding of the past. Second, would achieving this completely 'objectively authentic' experience be necessary even if we could achieve it? To a large extent, it seems questionable: Stephanie Lind, for example, references players who identify the music of *Odyssey* as authentic despite its lack of adherence to the musical tropes of the relevant period and region and resultant objective inauthenticity (Lind, 2022). Similarly, we could see Bijsterveld's discovery of museum visitors who did not want to hear the sound of a 1930s car as also being driven to some extent by authenticity: they 'knew' what a 1930s car sounded like, and so did not need to. We can imagine that if they *had* chosen to listen to the objectively authentic sound of the 1930s car, then they might be confused – they might think that this was inauthentic. Many other examples have also been identified: James Cook highlights how *The Witcher 3* "draws on modernist thought" in representing medieval Europe by using "iconic neo-medievalist aural cues" such as the rasp of metal when unsheathing a sword. Whilst common in Hollywood and other contemporary media, a well-maintained sword leaving a well-maintained sheath would likely have been silent (J Cook, 2016).

Similarly, Brendan Lamb and Barnabas Smith explore how fantasy video games “attempt to remove the player from the contemporary world through a[n inauthentic] musical concept of ‘the past’” (Lamb & Smith, 2018, p.99). We can recognise that audience perspectives of auditory authenticity are not solely determined by the considerations highlighted throughout 3.0.1–3.0.4.

We have already discussed the ways in which the noise of the dinosaur is inauthentic in section 2.2.2, and that using an authentic dinosaur noise may lead to participants being confused or viewing the sound as outside of the simulation (i.e., as ‘noise’), but it is worth revisiting them in the specific context of authenticity. Formerly thought to be roars and shrieks (e.g., as in the *Jurassic Park* franchise), recent discoveries suggest that dinosaur vocalisations may have closed-mouth vocalisations similar to trills and chirps of contemporary birds (Riede et al., 2016; Clarke et al., 2016).[49] Imagine replacing the roars in *Jurassic Park* with chicken sounds – is this as scary? Does this feel as *real*? Is this as *authentic*? Prehistoric sounds pose all kinds of additional issues to any sense of objective authenticity in any case,[50] but these questions do highlight that a sizeable portion of our personal judgement of authenticity is principally subjective. We make a value-based judgement of how authentic something is based on our expectations of an event – in this case, how close is what we are hearing to what we are expecting to hear? From an experiential perspective, we can recognise that users are more interested in ‘faux-thenticity’; the extent to which something meets their prior expectations based on either their past experiences (or imagined experiences) or reality, or their understanding of the sounds from media literacy. As Casanelles puts it, “the fidelity of a sound comes from it *being perceived* as realistic” (Casanelles, 2016, p.68, emphasis added): whilst the elements discussed throughout 3.0.1–3.0.4 clearly constitute many of the ways in which a sound can be considered realistic (or ‘authentic’), these are not the only approaches. These considerations are discussed further in section 3.1.2 and 4.2.1.

[49] A (less terrifyingly scientific) review of this topic is provided in NBCLX’s interview with sound designer Al Nelson – who worked on *Jurassic Park* (2015) – and palaeontologist Julie Clarke – who specialises in dinosaur vocalisations (LX News, 2020).
[50] Presumably any literature written by the great Mesozoic researchers on the acoustic histories of their species were lost to time. Or to a giant asteroid wiping out complex life.

What must be recognised, however, is that this plausibility includes not only the actual sounds which are heard, but also other acoustic information contained within them: e.g., specific distortions and colorations that appear when recording loud/high-intensity sounds using low-quality recording equipment were present in the audio design for *Battlefield 4* (DICE, 2013), a first-person shooter video game in which users fight their way through a war in the year 2020. The decision to use low-fidelity recordings was based on the belief that these recordings would be more similar to footage that their players might have seen previously of warzones around the world (R. Stevens & Raybould, 2015, pp. 61–62). We can see that in order for virtual experience participants to identify sounds as (subjectively) authentic, they must both hear what they expect, and hear what they expect to hear *in the "right" way* – even where these sounds or mediations are (objectively) inauthentic: we must "remain attentive to what and how people saw – their culturally mediated visual imperatives – and to how they listened, heard, and used sound" (M. Smith, 2015, p. 142). To investigate the construction of auditory authenticity further, we will now examine historically informed performance. Historically informed performance has been influenced by changes in music technology – such as changes to how instruments are constructed, and the rules by which we should perform – which have dramatically altered the possibilities of performance. As a result of these alterations, expectations of what music *should* sound like have changed. This allows us to use the debate surrounding historically informed performance to investigate how authenticity has been constructed within auditory domains despite changes to the elements that construct the overall experience of performance.

3.1.1: Authenticity in Historically Informed Performance

Historically informed performance, also known as period performance, is an approach which aims to be faithful to the approach, manner, and style of the musical era in which a musical work was originally conceived. This musical movement is at the centre of a well-known discussion examining authenticity in sound. Whilst initially conceived as referring specifically to the performance of classical music,[51] it should be noted that similar discussions are also raised in the performance of other genres and types of music: e.g., Allan Moore describes how the term ‘authentic’ can be justified “on occasion … with close reference to details of sonic design” (Allan Moore, 2002, p. 209) and sets out how it has been applied “not exclusively to popular music …. [but also] pop and rock … dance music … [music for] commercialised raves in the late 1980s … folk music tradition … and Euro-American formal music tradition” (Allan Moore, 2002, p. 210). However, whilst the considerations may be widely applied, the specific focus of this section will be historically informed performance as the ‘historically accurate’ performance of Western classical music, as many of the questions surrounding authenticity explored so far in this chapter have also been explored in historically informed performance of Western classical music.

As with the concept of ‘historically authentic sound’ set out above, historically informed performance has been a controversial area of musicological debate. For example, Richard Taruskin was already enquiring by the end of the 20th Century whether some expressions of authenticity had become “a positivistic purgatory, literalistic and dehumanizing, a thing of taboos and shalt-nots instead of the liberating expansion of horizons and opportunities it could be and was meant to be” (Taruskin, 1995, pp. 76–77). Taruskin’s distaste for ascetic purism seems to some extent to echo earlier opinions from other writers. For example, similarities between Taruskin’s arguments and Theodor W. Adorno’s contempt for those advocating for musical purism could be drawn: “The favourite argument of the purists is that [interpretation] should be left to the work itself, which need only be performed ascetically in order to speak… [Their] attempt to do justice to Bach’s objective content by directing their effort towards abolishing the subject is subtracted. Objectivity is not left over after the subject[ive] is subtracted” (Adorno, 1981 [original released 1951], p. 144). Both Taruskin and Adorno object to the idea that authenticity comes from recreating the ideas of a (dead) creator, rather than portraying an original interpretation by a (living) performer.

[51] The precise definition of classical music – or ‘western art music’, or ‘European art music’, or any other similar term – is a matter of some debate and has been for several centuries (e.g., see Salaman, 1879). This debate has been explored in far more detail and depth elsewhere and will not be revisited in this thesis.

Throughout this controversy, however, the construct of what it means to have a ‘historically authentic’ musical performance can be observed as requiring authenticity across many different vectors. Period-appropriate instruments, for example, can be seen to be widely considered an important part of historically informed performance practice – Laurence Dreyfus, for example, highlights how supporters of the Early Music performance practices “absolutizes a nineteenth-century of orchestration that did not yet exist in the early eighteenth century…the reconstruction of the original instruments is taken as essential to the meaning of the music before the idea of essential instrumentation becomes historically operable” (Dreyfus, 1983, p. 301). Whilst Dreyfus is implicitly critical of these practices, he does nonetheless highlight that the practice of demanding the use of specific instrument designs or specific methods of constructing instruments were widespread. Similarly, we can see Robert Donington’s consideration of the ‘correct’ orchestration for baroque music as another example of these aspects of instrumental authenticity in historical performance – albeit one that Donington is clearly unconvinced has any singular objectively correct answer, and which he is keen to avoid framing as a ‘moral’ issue (Donington, 1973).

Similarly, performers of historical performance must establish the means through which they interpret the musical score. Frederick Neumann – a firm supporter of authentic performance practice – argues that historical performance practices have been “beset by difficulties which are caused by questionable procedures… the use of wrong sources [and] the wrong use of sources” (Neumann, 1967, p. 315). In particular, Neumann suggests that the most important sources for those seeking historical performance information are historical treatises (Neumann, 1967, p. 316), which must be carefully read and considered, highlighting in particular the risk that a “modern commentator” might “take [treatises] at their face value …. [and become] guilty of too narrow interpretation” (Neumann, 1967, p. 322). Perhaps most notable is Neumann’s desire that musicians treat the musical scores not as infallible source of information, but as a way to “give the treatises their most valuable support by shedding light on crucial questions” (Neumann, 1967, p. 323).

Dorottya Fabian's review of the historically informed performance movement also highlights other aspects of authenticity which should be considered. Fabian highlights the work of Ludwig Finscher – who suggests that the 'ideal' historical sound can "only be reconstructed through the resource of any given historical period and considers instruments, voice-types, playing and singing techniques, proportions, acoustics, improvisation, tempo, dynamics, agogic and phrasing *all* to belong to those resources" (Fabian, 2001, p. 163) – and Wenzinger – who Fabian argues believed that "the production of the musical structure is not enough; the *Affekt* must be reproduced and everything else must flow from that [because] the Baroque era was obsessed with theatre and the theatrical" (Fabian, 2001, p. 164). In both of these cases, Fabian seems to be emphasising how the context in which the performance takes place was also considered part of the authenticity of historical performance by referring to both Finscher, who insists that the acoustics of the concert hall be considered as part of musical authenticity, and Wenzinger, who highlights affective considerations of performance as part of musical authenticity.

3.1.2: Deconstructing Authenticity

Whilst many differing approaches to the creation (and critical judgement) of historically informed performance can be identified in this literature and throughout other literature on historically informed performance, the discussion helpfully identifies that there are clearly many elements which were required to construct an 'authentically historic' performance. A relationship between such formation of an 'authentically historic' performance and the framework for virtuality set out within Chapter 2 can be observed. Performers must use specific musical instruments to perform; they must use the right hardware for the experience to be authentic. Similarly, performers must interpret musical scores in a specific way; they must have the right 'software' to be able to interpret the instructions correctly. And, for some of the most devoted adherents of 'authentically historic' performance, the performance must also take place within an authentic externality; they must be within a concert hall or other similar setting, or with the right affect. It is often not enough for *one* of these elements to be correct if a performance is to be considered ‘historically authentic’; many or all factors should be predominantly correct.

Similar discussions can be observed in other debates surrounding musical ‘authenticity’. For example, Keir Keightley suggests that for something to be recognised as ‘rock’ then it must be seen as “serious, significant and legitimate” (Keightley, 2001, p. 109) – i.e., ‘authentic’. These debates are recognised as multi-modal: Keightley highlights earlier literature that draws upon “specific historical contexts, audiences, critical discourses, and industrial practices… a rejection of those aspects of mass-distributed music which are believed to be safe, soft or trivial” (p.109) and suggests that authenticity “requires a sense of music’s external contexts … and marketing strategies, music-making technologies, or the ongoing history of music’s broader stylistic changes” (p.131). Some differences do emerge as a result of the increased technological mediation in rock and other popular music styles. However, just as in historically informed performance, we can see that the aspects of ‘authenticity’ discussed by Keightley require a knowledge of (and adherence to) specific historical contexts, drawing on specific critical discourses, utilising specific creative practices (and rejecting other musical practices), and specific stylistic practices, such that an external (and subjective) authentication of the music as ‘real’ can take place. Stan Hawkins also seems to concur, suggesting that the authenticity of Morrissey’s vocal performance is recognised through it’s “natural” (i.e., untrained) sound (Hawkins, 2002, p.86). This seems similar to the way in which authenticity of historically informed performance might be created through specific vocal timbre or techniques being (un)used.

Similarly, Simon Frith argues for a ‘authenticity’ which “is a quality not of the music as such (how it is actually made), but of the story it’s heard to tell … the narrative of musical interaction in which the listeners place themselves” (Frith, 1996, p. 124). For Frith, this authentication revolves around a continuous process of affirmation of (or recognition of) identity: “playing and hearing what sounds right… that we both express ourselves, our own sense of rightness, and suborn ourselves, lose ourselves, in an act of participation” (p. 110, emphasis original). Whilst Frith inverts the notion of music and identity by arguing that, fundamentally, music “articulates in itself an understanding of both group relations and individuality” (p.111, emphasis original), they also acknowledge that this identification revolves around a series of decisions about “what is ‘good’” (p.119) which can include cultural, social, and technical elements. In other words, authenticity must be seen across several paradigms for the overall experience also to be considered ‘authentic.’ We can recognise that this expression of authenticity is *also* constructed across all paradigms within the tripartite structure suggested in Chapter 2.

However, as with Slater’s observance that virtual reality must be ‘plausible’ to be ‘authentic’, it might also be observed that authenticity within each of the paradigms is not so much ‘historically authentic’ as ‘historically plausible’. For example, whilst the distinction between sacred and secular music may not have been as clear in the Baroque period as in more contemporary periods, much of J.S. Bach’s music is specifically liturgical. Therefore, it would likely have been intended for performance as part of a religious service. Despite this, many contemporary examples of ‘historically authentic’ performances of the music of J.S. Bach are conducted specifically as secular concerts. This is clearly a change in the ‘virtual externality’ of performance – the concert hall, and the context in which these musical performances were conducted. It appears that these performances are to, some extent, inherently inauthentic. However, they are often nonetheless accepted as authentic examples of Baroque music in a historical performance practice.

An implicit emphasis, therefore, can be seen on the value of subjective authenticity in assessing the overall authenticity of music performance practices. Moore in particular argues that authenticity is a social experience: “rather than ask *what* (piece of music, or activity) is being authenticated, I ask *who* [is authenticating it]” (Allan Moore, 2002, p. 210). Through casting authenticity as a social determination rather than a determination of historical accuracy, Moore broadly concurs with Johan Fornäs, who suggests that authenticity is “a spatially, socially, and temporally rooted position” (Fornäs, 1995b, p. 10) in which “credibility is constructed through complex cultural practices … with textual markers that imply a close relation between the text and its author, that is, that the expression of a subject is honestly meant" (Fornäs, 1995a, pp. 275–276). Both Moore and Fornäs identify several kinds of authenticity, both Fornäs and Moore are therefore arguing towards a subjective valuation of overall authenticity that depends on the people(s) receiving any given 'text.' Whilst both highlight several expressions of authenticity (e.g., Moore highlights 'first-person', 'second-person' and 'third-person' authenticity), both also state that there are not necessarily firm barriers between any categorisations or types of authenticity; an experience may be subjectively authenticated through a mixture of different formulations of authenticity. The underlying unification of an authentic text is found in the social (or subjective) concurrence of those viewing it.

Whilst Moore and Fornäs are specifically talking about *musical* authenticity – and to some extent the authenticity of musical performance – much of the same framework can also be applied to authenticity for sound within virtual experience. We discussed in section 3.1.0 the difficulties faced by those creating soundscapes or sound effects for use in historically informed experiences, such as in *Odyssey*. Coalescing with our discussion of historic performance practices in section 3.2.1, historical authenticity also could be seen to have had significant limitations in the music of *Odyssey.* Whilst composers Alexis Smith and Joe Henson (collectively known as 'The Flight') originally discussed the need for extensive research to try and authentically recreate the music of ancient Greece (*The Flight's Music Brings Ancient Greece Alive In 'Assassin's Creed Odyssey'*, 2018, sec. 01:00–03:50), post-release interviews highlight an underlying pragmatism to their eventual compositional approaches: "No one knows [what it would have sounded like]. … No one wants that anyway. … [We were told] the game is very particular about the historical stuff, but the music doesn't have to [be like that]" (Spitfire Audio, 2019, sec. 38:50–40:30).

The pragmatism surrounding (in)authenticity expressed in this interview likely emerged as the result of several factors. Firstly, *Odyssey* exists first and foremost as an interactive gaming experience: the music must serve not only serve diegetic purposes but also several communicative purposes – e.g., indicating when combat begins and ends.[52] Secondly, only very limited information about the music of ancient Greece survives. Whilst Ubisoft correctly identify the existence of many groups attempting to piece together the music of ancient Greece, these efforts are significantly handicapped by a lack of primary evidence: Egert Pohlmann and Martin L. West offer what is described as a “uniquely comprehensive and up to date” commentary on fragments of surviving music manuscripts from Ancient Greece in *Documents of Ancient Greek Music: The Extant Melodies and Fragments edited and transcribed with commentary* (2001). The book examines 61 items – an increase from the 38 items known at the time of Pohlmann’s first such commentary in 1970. Whilst other resources do exist, which provide less direct information about the music of ancient Greece,[53] the potential for a fully authentic realisation seems low as a result of this relative scarcity.

Player expectations of what ‘ancient Greek’ music should sound like may have had some prejudicial effects on The Flight’s music development process for *Odyssey*. In their examination of *Odyssey*’s music, Lind finds that some parts of the music are “surprisingly” authentic (Lind, 2022, p.34), but that the overall framing of the music is functionally inauthentic: “Modes, rhythm, and accompaniment all serve to suggest the Ancient Greek setting of the game in ways that do not necessarily correspond to historical practice” (Lind, 2022, p.25). Despite this, Lind acknowledges that players and listeners nonetheless identify the sound of the music as being authentically ‘Greek’ (Lind, 2022, p.31). This is suggested to be because authenticity within video games is not meant to be objectively real, but rather than authenticity is intended to be “that which conforms closest to the expected experience and is in the eye of the beholder” (Lind, 2022, p.40). The Flight may have made the determination that it was better to create a score which would be recognised as Greek, rather than one which was authentically *ancient Greek.* This concurs with approaches to other musical scores

[52] Further detail about the intersection of music and interactive audio-visual media, specifically virtual reality, is explored in detail in Chapter 4 of this thesis.

[53] For example, Aristotle’s *Poetics* and Plato’s *Republic* both contain sections which discuss music – either within the broader context of Greek society, or through a philosophical/theoretical lens. However, these philosophical writings provide little information about how to recreate the musical instruments, musical sounds, and musical scores of ancient Greece, which would be required for an objectively verifiably authentic recreation of ancient Greece.

discussed in literature surrounding authenticity in video games and other interactive experiences – e.g., Tim Summers suggests that inauthenticity may be a tool through which the game's blend of realism and fantasy is combined (Summers, 2016, p. 90), or part of the overall joy of playing video games (Summers, 2016, p. 111).

These observations can be applied beyond just music. We can see that sound design also seems to require subjective authenticity rather than an 'objective' authenticity. Bijsterveld recalls scepticism from museum visitors who listened to the painstakingly recreated sound of the 1920s car explored earlier in this chapter: users admitted skipping these 'mundane' elements of the soundscape because they assumed or believed that they knew what the car *should* sound like (Bijsterveld, 2015, p. 82). These assumptions and the emergent scepticism that an unexpected sound is incorrect is rooted, at least partly, in notions of subjective perception. Just as Lind seems to emphasise the concept of player expectations in how they identified authenticity in The Flight's score for *Odyssey*, we could also recognise that a key part of Slater's "plausibility illusion" are the user's expectations of what "should" be there: "how much do [the events that are taking place] conform to expectations of what would normally happen in the circumstances portrayed?" (Slater, 2009, p. 12). The users, faced with something plausible, determine whether they are willing to accept that as authentic. We can therefore understand that the process of reception – of perceiving a soundscape – depends on our experiences within, and our understanding of the authenticity of, the experience.

3.1.3: Authenticity in context

By considering how authenticity was constructed within historically informed performance, and how authenticity has been considered within experiences intended to be historically authentic (or 'historically accurate'), we have established that authenticity is a construct formed through the user's recognition that many different aspects of an experience are individually – and collectively – authentic. This suggests that sound designers aiming to create immersive experiences must consider how inauthenticity can be created through many different aspects of a virtual reality. In particular, sound design of a virtual experience requires consideration of how users will hear (or 'experience') sounds within the virtuality. This includes considering aspects of sound and authenticity across all elements of the tripartite structure for virtuality developed in Chapter 2.

In practical terms, this might mean conducting user reviews to ensure that communicative aspects of sound are communicated clearly to the user – and developing methods of teaching users to understand essential sounds in context. As sound design within many virtual reality experiences necessitates some level of compromise between audio fidelity, audio resolution, and performance of computer hardware, we can also see the importance of considering how the virtual software selects and prioritises sounds within a virtual experience. The potential importance of accurate digital modelling was also highlighted. Particular attention was given to the ways in which inauthenticity in virtual reality can emerge from the design approaches required to overcome challenges in the implementation of near-field binaural sound.

Alongside these concerns, which are well-established within sound design literature, this research also suggests that designers should consider whether appropriate hardware has been selected: e.g., will hardware allow the user to perceive frequencies in a way which they understand as authentic? Whilst perhaps most commonly understood as relevant within virtual reality experiences, this research also suggests that these issues extend beyond traditional virtual reality experiences and are a potential issue in other virtualised experiences. One area of interest is the virtualisation of experience facilitated by brain-computer implants, which will be explored in greater detail in section 6.2.0 – 6.2.3. This research also suggests that designers should consider the virtual externality of an experience and consider how they can control and shape the externality which their users will be engaging within, which will be a recurring theme throughout many of the case studies presented in Chapter 5 of this thesis.

We have now considered the two main axes of the framework for analysing and designing virtual experiences which this thesis constructs. In Chapter 1, we explored how virtuality can be constructed within a broadly applicable framework. We then considered sound throughout Chapter 2 and 3. The principle focus of these two chapters has been diegetic sound and non-diegetic sound (such as music) has not yet been discussed. Music is found within many virtual experiences and is considered to be an important part of virtual experiences. However, music is often considered to be inauthentic to virtual experiences as there is no reason why a user might expect music to emerge ‘from thin air’ as it does in – for example – cinema. In Chapter 4, we must now consider how music situates itself within sound and authenticity to function within virtual reality experiences as a widespread example of non-diegetic audio and how non-diegetic audio interrelates with the considerations of virtuality, sound, and authenticity discussed throughout Chapters 1–3.

Chapter 4: Music

4.0.0: Music in Virtuality

In Chapter 2, the framing of sound within virtual reality was investigated. Chapter 3 explored authenticity within this understanding of sound across the elements of virtuality explored in Chapter 1. Whilst authenticity was identified as being fundamentally subjective, it was suggested that designers should consider many different aspects of sonic authenticity to ensure holistic engagement with virtual experience: we can feel more narratively immersed (or experience less challenge-based disruption to immersion) when sounds accurately represent objects within the world, and when spatial cues such as reverberation are accurately recreated within the virtual experience, such that audiences believe that the perceived sounds represent a plausible experience – in some cases, sound may even be central to the concept of embodiment: to feel as if they are 'really there' within some virtual worlds, audiences must be convinced that the sounds that they are hearing represent a plausible experience.

Historically informed performance was used to consider how this model of hearing might intersect with the model of virtual experience suggested in the first chapter of the thesis. Within historically informed performance movements, musicians frequently consider the accuracy of all three elements of the tripartite model of virtuality suggested as a mode of analysis. Musicians must consider their hardware – e.g., the instruments they play – and their software – e.g., the scores that they play from – and the externalities that they play in – e.g., the concert venues. For audiences to feel that the music they are hearing from historically informed performers is authentic, they must be convinced that this combination of hardware, software and external cues accurately represents what might have once been created in a real-actual concert venue of the past. Similar approaches were also observed in other judgements of musical authenticity found in genres such as pop music and rock music.

Sound considered within the previous two chapters was principally diegetic. Diegetic sound is that which is conceptualised as emerging from within the virtual world: characters can hear (and react to) noises within the world, and the sound is understood to happen at the same time (and in the same place) as the portrayed action. Not all sounds within virtual experiences are diegetic. For example, section 2.2.1 has already discussed transdiegetic sounds, such as interface sounds, which are sounds which react to events within or about the virtual world to convey information at a meta-level to a user or player (e.g., a video game character might have a dialogue line which plays when an ability is used, to indicate to the player that the ability has been used). This chapter examines non-diegetic sound: sound which is conceptualised as existing outside the virtual world: e.g., characters are not able to hear and/or do not react to the sounds. Whilst there are several examples of non-diegetic sound (e.g., voiceover, narration, soundtrack music), this chapter focuses in particular on the use of music within virtual experiences.[54]

Music is used widely throughout virtual experiences of all kinds and is the most widely used form of non-diegetic sound. This may be considered odd: non-diegetic music is patently inauthentic to virtual experience creating what Inger Ekman refers to as “the difficult question why, and how, something as obviously constructed as the [musical] score, does not completely destroy the sense of realism within the [experience]” (Ekman, 2008, p. 5). Notwithstanding the potential difficulties arising from the (im)plausibility of music, we can also recognise that music is used for a number of specific purposes. Building on the discussions of sound literacy in Chapter 2, and models of authenticity explored in Chapter 3, this chapter examines reasons why music might be used within virtual experiences, the challenges to implementing music within virtual experiences, and the broader possibilities for music within atypical virtual experiences such as business or education contexts. Through the examination of music within virtual experience, we can also begin to establish how fictional realities can be constructed more broadly as ‘authentic’ despite their often-fantastical approaches. This will finalise the axis of sound and music which began in Chapter 2 and which

[54] Kulezic-Wilson (2017) argues that those working in audio for contemporary media use “an approach that destabilises hierarchical relationships between the film soundtrack’s constitutive elements” (p.127) and “blur the line between music and sound design” (p.130) such that there can be “ambiguity regarding the source and the nature of the sounds … it is difficult to tell whether what we hear emanates from the diegesis or the score” (p.133). Sound and music have been separated in this research solely for structural purposes and it should be recognised that contemporary audio practice is increasingly combining them.

intersects the axis of virtuality established in Chapter 1 to create the framework for analysing, designing, and implementing sound and music in virtual experience.

4.1.0: Purposeful Music in Virtual Experiences

The considerations within this chapter draw heavily on the use of music within other contexts, such as film and video games. To some extent, we should expect similarities between music for virtual experiences, music for cinema, and music for video games. Conceptualised broadly, we could consider cinema and video games as virtual experiences: they invite audiences and participants into a virtual world created through combining hardware and software within a (sometimes specialised/dedicated) externality. As considered through the framework for virtual experience suggested in Chapter 1, then, we can recognise similarities between existing media such as cinema and video game and emergent new media such as virtual reality. We should therefore expect music within many expressions of virtual experience to be used in much the same way as in cinema and video game.

Alongside the theoretical similarities in virtual experience discussed in Chapter 1 of this thesis,, we can also recognise that many audiences do not specifically distinguish between experiences on a technological level: e.g., virtual reality video games are frequently considered to be ‘just another type’ of video game. In practice, audiences seem to view new technology as offering new expressions of old media, despite differences in the creation and reception of the experience, rather than necessarily offering an entirely new form of experience. This may be because whether the game uses a virtual reality headset with hand-held controllers, or a more ‘traditional’ PC set-up with keyboard and mouse, the underlying systems of engagement and logic used within the video game context are broadly similar. Whilst the hardware may have changed, the underlying cultural knowledge and assumptions of the user are not challenged by the transition to technologies such as virtual reality.

Further, as we might imagine from the discussion of media literacy in section 2.2.1, users have expectations of what video games should sound like and how sound and music should function within those video games: e.g., Isabella van Elferen suggests that “immersion in [games] is supported by an ‘epic’ composing style that relies heavily on compositional idioms from epic and fantasy films” (van Elferen, 2016, p. 37). These expectations will remain consistent even where technology changes, as we might already have seen in the continued use of a cinematic musical lexicon within narrative video games. This means that we can examine the use of music within cinema and video games in order to consider how music will be used within virtual reality games. Given the similarities between the media, we can also expect creators of virtual experiences to draw upon approaches and methods used to score cinema and video games.

Therefore, the purposes of music within many virtual experiences will be broadly similar to the purposes of music used within cinematic and video game media. This chapter explores how music is used within virtual experiences for four key purposes: semiotics (i.e., using music to signify meaning to an audience, e.g., geographic location); emotion (i.e., using music to evoke a particular emotional state in an audience); immersion (i.e., using music to sensorially or narratively engage the user); and embodiment (i.e., using music to make the audience feel more present within the virtual world). Whilst these are not the only purposes of music – or the only way to categorise purposeful uses of music within audio-visual media – they each provide useful frameworks for exploring music within virtual experience. Throughout this chapter, the music of *Half-Life: Alyx* (hereafter: *Alyx*) will be used to demonstrate how these four purposeful uses of music are expressed within a virtual experience.

Alongside the purposeful use of music, this chapter also examines key challenges to music within virtual reality. The concept of authenticity was problematised in section 2.2.2, and – as identified above – poses specific issues to music in many cases, as music could be considered as inherently inauthentic to the virtual world. Section 4.2.1 considers authenticity and ‘fauxthenticity’ in music to examine why these issues may not always be problematic in virtual experiences. Music also relies on close synchronisation with narrative. Within interactive virtual experiences – i.e., where users control the pacing or framing of an experience – it is important to ensure that the music is correctly synchronised. Section 4.2.2 explores different approaches to synchronising music to interactive experience. A third major challenge specific to virtual reality is the difficulty of spatialising music within the virtual world: section 4.2.3 isolates the specific difficulties music faces whilst being spatialised within virtual reality and explores some solutions to these issues.

Virtual experiences extend beyond video games and cinematic experiences, however, and include a wide range of other possibilities. Whilst much of this chapter is broadly applicable, the use of *Alyx* as a case example means that the predominant focus will therefore lie on virtual reality video games. Section 4.3.1 – 4.3.3 apply the same discussions to other expressions of virtual experience, such as augmented reality, educational/professional experiences, and virtual experiences that do not use screens (e.g., theme parks). This will ensure that the applicability of this research to broader fields of interest is clear and demonstrate the relevance and importance of music to broader conceptualisations of virtual experience.

4.1.1: Musical Semiosis

Music in film has often been used to provide contextual information to viewers. Justin London begins his analysis of cinematic leitmotif by establishing the general concept that "certain portions of the musical soundtrack of a film 'refer' to characters or actions that may (or may not) be present on the imagetrack … particular melodic and/or harmonic figures serve as sonic tokens for persons, objects, and/or ideas that have a significant role in the film's narrative" (London, 2000, p. 85). It is sometimes suggested that this is similar to the Wagnerian leitmotif; Scott Paulin argues that "Wagner is cited as a model (or the model) for film-music composers and performers to follow" (Paulin, 2000, p. 58). Roy M. Prendergast similarly argues that the "formal device of the [Wagernian] leitmotiv fell naturally into use in the composition of scores for some of Hollywood's films" (Prendergast, 1992, pp. 40–41) concurring with Caryl Flinn's description of an early cinema which followed "rules that adhere to the principles of Richard Wagner and late musical romanticism" (Flinn, 1992, pp. 14–15). However, whilst some examples of some Hollywood scores may use these connections,[55] the semiotic connections discussed here are broader than the specific use of Wagnernian leitmotif.

Claudia Gorbman identifies how music in cinema communicates specific information such as "historical, geographical, and atmospheric setting" (Gorbman, 1987, pp. 53–54). In Gorbman's model of Hollywood scoring, music is subordinated to the narrative: it existed as a means to an end, in particular providing narrative cuing and unity (Gorbman, 1987, p. 73). Whilst the codified tropes and topics of 'classical Hollywood' scoring have since been challenged in contemporary composition practices,[56] Gorbman identified many historical approaches still in use in 1987: a jazzy major theme for a major city; a cocktail-lounge jazz saxophone for the seductress next door; "wrong"-sounding notes for comic characters; rumba rhythms for Latin America; and so on (Gorbman, 1987, p. 83). Gorbman asserts that music for audio-visual media is intended to allow the audience quickly to understand the context of the images by "guiding the spectator-auditor unambiguously into [compelling identifications]" (Gorbman, 1987, p. 98); music is seen as a way to signify information about the narrative events of the film.

Similar notions can also be found in video game music. Tim Summers examines how national leaders in the strategy game *Civilization IV* (Firaxis Games, 2005) are given a melodic motif "taken from pre-existing music that somehow resonates with their national identity" (Summers, 2016, p. 95); this approach was built on further in more recent instalments in the *Civilization* series, which combined these nationalistic resonances with historical resonances. Specific topoi are frequently used within video game music by developers, in an effort to guide players through the game: Stephanie Lind identifies that the "modes, rhythm, and accompaniment [of the music of *Assassin's Creed:*

---

[55] For example, Irena Paulus argues in an exploration of the 'Wagernian' approach to the music of *Star Wars* that Wagernian leitmotifs require "themes to be linked with given characters, subjects or ideas, to be seen in basic form and in variants, and thus keep up with transformations of characters, and to be akin to each other, and to give shape, through their appearance, to a unified and single score." (Paulus, 2000, p. 172). However, in highlighting how *Star Wars* does use these Wagnerian ideals, Paulus is suggesting that other film scores fail to meet all necessary criteria for the technical definition of 'Wagernian leitmotif.'

[56] Music often mis-represents the musical cultures of the geographies and communities that films represent. For general discussion surrounding the ways in which media misrepresents non-American cultures, and the importance of avoiding this misrepresentation, see Caldas-Coulthard (2003), Castañeda (2018), and Kidd (2016).

Odyssey] all serve to suggest [to the player] the Ancient Greek setting of the game” (Lind, 2022, p. 25).

Alyx’s use of music demonstrates that these established techniques continue to be effective in the new media of virtual reality. Several different approaches can be identified throughout the game. Music within opening credits has been suggested to help signal the beginning of an experience and focus the viewer (Davison, 2013). Valve does this be re-using a short clip of music which has featured in the opening credits of almost every game created by Valve. Gamers are familiar with this music, but at the time that *Alyx* released, it had been several years since the music was last heard in a single-player game. The last full-scale single-player title to be released by Valve had been *Portal 2* in 2011 – nearly a decade before *Alyx* hit the (virtual) shelves in March 2020. Hearing this music may serve as a way to remind players of the past games within the *Half-Life* franchise and, by extension, help to remind players of the fictional world these games existed within, in much the same way that Donnelly describes the music of *Silent Hill* as a “constant evocation of memories of the game and feelings associated with the game” (Donnelley, 2016, p.78). Even before our engagement with the interactive game systems begins, players begin to recall the story and world: they have heard this music before. They know what world they are entering. They are preparing to enter that world.

Other uses of semiosis can also be found throughout the music of *Alyx*. In many ways, the timbral qualities of the music of *Alyx* establish much of the thematic background for the story. The musical genre for *Alyx* is hard to pin down precisely. There are electronic elements placed prominently throughout the music: synthesisers, tone generators, and deliberately audible music production artifacts. Drum kits also feature prominently. Some tracks consistent of distorted bass synthesisers and drumbeats in a facsimile of the D’n’B genre, although the game perhaps draws more heavily on techno or electronica influences. The use of a powerful electronic guitar in some sections might draw comparisons to heavy metal tracks – although it should be noted that there is a close relationship between genre tropes of heavy metal music and video game music (see Rossetti, forthcoming; Guay & Arsenault, 2012), which may suggest that these similarities reflect creative inspiration emerging from other video game sources, rather than specifically drawing on heavy metal tropes. Other tracks defy these expectations, stretching out without any sense of time or beat, with a melody stretching oddly through the distinct sonic elements of the track. Bringing together the music under a single genre identifier is difficult – if not impossible.

However, no matter the precise genre(s) of the music of the game, the underlying semiotic language is easily identified. Music with many electronic elements is often seen as otherworldly in audio-visual contexts: Rebecca Leydon identifies that “the earliest couplings of electronic sounds and cinematic images consistently involved the uncanny and the alien” (Leydon, 2004, p. 63) and describes how the “abstruse technology of the Krel” is represented in *The Forbidden Planet* (1956) by “contrapuntally dense music…roaring, humming and grinding” (Leydon, 2004, p. 74). Similarities to the soundtrack for *Alyx* are evident. Similarly, Michael Hannan & Melissa Carey explore how the music of *Blade Runner* (1982) is “characterised by its extensive use of ambient sound effects and atmospherics” (Hannan & Carey, 2004, p. 152). The description of the music as “single deep, explosive sounds” accompanied by “a lonely single-line melody, played high in pitch [using a synthesised sound]” is also applicable to the music of *Alyx*. The music of *Alyx* therefore can be seen to establish the game as dystopian science-fiction almost immediately, drawing upon timbral conventions for genre established across the latter half of the 20th Century. Whilst existing players of the franchise are likely already aware of the genre tropes which the *Half-Life* franchise draws upon, the music helps to establish and emphasise those tropes for new players.

When the game shifts away from these timbres, it is therefore particularly noticeable. We can consider that the player may be invited to pay particular attention to these moments. In one key narrative moment, the player meets a member of the Resistance in their apartment. Throughout the apartment, a very different style of music can be heard. The track consists of a piano and an accompanying high-pitched instrument deliberately mixed and altered to sound like a low-quality and low-fidelity music file.[57] The recording is detuned and has had low-frequency sounds removed. The balance between the two instruments is not quite right, with the piano overly dominant and the other instrument – which is difficult to identify – in the background, despite it playing what could be the main melody. The overall effect is similar to amateur music videos shared widely on sites such as YouTube, where recordings have been done with a smart phone (or other low-quality microphone) placed on the piano. This means that the timbral acoustics are completely different to other pieces heard throughout the game.

Similarities could be drawn between approaches used to compose and produce this track and approaches used to compose "lo-fi beats". Adam Harper explores how lo-fi aesthetics – such as distortion and reduced prominence of higher frequencies – combined with the pastiche of "now-archaic styles from before the 1990s" combine to create "nostalgia that combined cultural and technological pasts" (Harper, 2014, pp. 316–317). Harper also identifies that "lo-fi effects … [are] closely connected with nostalgia, decay, and the passage of time" (Harper, 2014, p. 346); this emphasises what Emma Winston & Laurence Saywood describe as a "contradictory and simultaneous relationship with both a hyper-specific real past, and a vague imagined one" (Winston & Saywood, 2019, p. 48). It is these relationships with nostalgia that we can recognise as important within the score for *Alyx*: by drawing on these 'nostalgic' timbres, the music may intend to signify a suppressed cultural society within the resistance; humans who are trying to remember the pre-alien past or are trying to create a new, vague, approach to music and culture. This fits with the general setting of the game as explored in section 1.1.0 – the player exists in a world harshly repressed by the Combine (alien invaders from another universe). Similarly, the placement of the track – and the unique qualities of the track – makes it clear that the suppressed human population has their own music and culture which is distinct from the outside force. The music is not the techno/electronic music which symbolises the oppression of the Combine, and so we understand that the suppressed population are in

[57] Interestingly, the version of this track released on the OST recording has not undergone these alterations and sounds significantly different. This suggests that the semiotic connections highlighted in this section of the game are deliberate choices by the sound designers of this area; this may provide some evidence for the reading of the music presented here to be an intentional result of designer motives.

resistance *against* the Combine. Through this semiotic inference, the music supports the broader narrative of the game, which centres around humanities longing for an escape from alien subjugation. Music clearly allows for the communication of information through semiotic references; players and participants are able to determine information about the setting and narrative of experiences through the music used within them. Within virtual experiences, and virtual reality, the same well-worn techniques used within cinema, video game, and other media remain an important and relevant consideration within the purposeful use of music.

4.1.2: Musical Emotion

Emotional frameworks for music in audio-visual experiences coexist with semiotic frameworks for music in audio-visual experiences. Michel Chion describes how music "can directly express its participation in the feeling of the scene, by taking on the scene's rhythm, tone, and phrasing: obviously such music participates in cultural codes for things like sadness, happiness, and movement. In this case we can speak of *empathetic music*. [Or] on the other hand, [*anempathetic*] music can exhibit conspicuous indifference to the situation…[conjuring] up the mechanical texture of this tapestry of the emotions and senses" (Chion, 1994, pp. 8–9). Rather than signifying a specific element of the narrative (such as a character or location), Chion focuses on how music can inspire emotion and/or emotional connections to the cinematic narrative. Robert Stam et al. also define several types of music used within film on the basis of their emotive content: redundant music (which supports emotion of the narrative); contrapuntal music (which contrasts the emotion of the narrative); empathetic music (which conveys the emotions of the character/s); a-empathetic music (which is indifferent to the narrative); and didactic contrapuntal music (which tries to separate the audience from the drama) (Stam et al., 1992, pp. 60–65). Stam et al. and Chion's frameworks focus on how music functions to communicate (i.e., effect) emotions and narrative: either connecting the audience to the onscreen action, or divorcing them from it entirely.

Some crossovers between these two frameworks can be acknowledged: music rarely serves only a single purpose within an audio-visual experience of any kind. However, a key difference should be highlighted between frameworks which focus on signification (such as in Gorbman's analysis of Hollywood scoring) and frameworks which focus on emotional conveyance (such as Chion and Stam et al). For Gorbman, music principally *signifies* – it presents information to the viewer about the narrative such as location or character intentions. For Chion and Stam et al., music principally *emotes* – it helps the viewer to understand the characters onscreen, guide them towards a specific emotion that they should be feeling, or divorce them from emotion entirely. It seems likely that both approaches are used – either symbiotically or situationally – by composers of music for cinema and similar audio-visual experiences.

Emotional signalling within music for video game is also well-established as a key part of the audio-visual relationship. Stephen Baysted explores how music within *Need for Speed Shift 2: Unleashed* (Slightly Mad Studios, 2011) needed to "seek to describe the "real" racing driver's emotional and physical journey and at the same time represent and enhance the concomitant experiences of the game player through well-understood processes of aesthetic response" (Baysted, 2016, p. 158). The consideration of how the interactivity of video games alters relationships and conceptualisations of music is also important. Zach Whalen suggests that "though video games borrow from and adapt filmic musical practices, games rely on important cognitive associations between music and interpretations of causality, physicality and character" (Whalen, 2004). Jørgensen also argues that game sounds (including music) should not be considered in isolation as "the situation in which hit is heard always decides the interpretation of the informative content of the sound signal" (Jørgensen, 2008). Ekman also suggested a similar contextual understanding of music in games: "the framing of the music is bound to influence how it will be evaluated [emotionally]" (Ekman, 2014, pp. 207–208). Whilst it is not always clear whether the emotional evaluation is the result of an induced emotion or simply the user's acknowledgement that the music communicates the *intention* or *perception* of a specific emotion (Williams, 2021, p. 307–308), we can recognise that music sometimes serves as a way for player and character to become emotionally linked and share the same narrative together.

One example of this musical linking can be found early on in *Alyx*. At the end of the prologue, Alyx is walking over the rooftops to return to her apartment. She watches her father be arrested by agents of the hostile alien government. They drag him away and Alyx must (unsuccessfully) fight her way past the guards. Throughout this sequence, the music swells to include orchestral elements: horns and strings in a traditionally ‘epic’ orchestral arrangement. When the player is unsuccessful, the music abruptly ends. The player and Alyx both clearly feel that this is the moment to become a hero and fight to protect their family from the Combine but are ultimately frustrated in these ambitions. The emotional journey of the music mirrors both Alyx’s frustration that she cannot save her father and the player’s frustration that they cannot ‘be the hero’ by building towards an epic moment, which is then cut off.

This unification of player and character may, in some ways, encourage the player to feel embodied within the virtual experience: as the music reacts to both the character and the player, it may suggest that the player *is* the character. This could improve the player’s sense of ‘body ownership’ (i.e., the sense that the avatar/character within an immersive experience belongs to the player/controller) or improve the player’s sense of ‘agency’ within the virtual experience. Both body ownership and agency are considered to be essential to virtual experiences as they are key components of embodiment/presence (and, therefore, part of narrative immersion) within virtual experiences (Kilteni et al., 2012, pp. 375–377). These connections can centralise presence within virtual experiences. We can imagine that the player may feel less able to relate to Alyx’s struggles – and less able to see themselves *as*, in some ways, Alyx – if the music did not unite the player’s emotional journey with Alyx’s emotional journey. Some studies have shown connections between music and presence or embodiment. Dana Plank suggests, for example, that “aural stimuli elicit emotional, psychological and physiological responses from players, whether that response is ultimately meant to support narrative, influence gameplay decisions, foster player agency, or facilitate incorporation into the game body… a dynamic and vital bridge between the bodies of player and avatar” (Plank, 2021, pp. 284–285). With respect to the virtual experience, the emotional synchronicity between player and avatar is likely to increase the sense of immersion: e.g., by engaging them emotionally with the game in what Emily Brown & Paul Cairns described as the second stage of immersion: “the [environment] becomes the most important part of the [user’s] attention and their emotions are directly affected by the game. … The [user] is now less aware of their surroundings and less self-aware than previously” (Brown & Cairns, 2004, p. 1299). As one of the implied goals of virtual reality is often understood to be ‘boundary-free’ virtual experience – i.e., the player does not perceive the inherent restrictions placed on them by the hardware, software, and externality of the experience – then we can recognise that the ability of music to create emotional synchronicity and thereby improve embodiment will aid narrative immersion of users within the experience: they will be less aware of the real-actual.

These synchronicities can emerge out of congruence between many different kinds of structure, as identified by Annabel Cohen: “Rhythm, beat, and accent pattern are examples of musical structure. There are other kinds of structure as well, for example, harmony and the up and down contour pattern of melody. Perhaps less obvious is this type of structure in the visual domain of film, for example, an accent pattern created by objects moving on the screen, be it human limbs, fluttering leaves, boughs of branches, or, traveling vehicles that create patterns of light and dark that change across time” (Cohen, p.21). Cohen provides several examples of experiments which demonstrate the “associative influences of music on the interpretation of the moving image” (p.27) through congruencies emerging in such a way. However, other approaches to manipulating user emotions through music (or, at least, creating associations between music and visuals) have also begun to emerge in recent years.

Steven Willemsen & Miklós Kiss argue that psychoanalytic theories of screen music (such as those discussed earlier in this section) “tend to overlook [some] dynamics of communication, presuming that viewers claim all associated emotions as their own” (Willemsen & Kiss, 2013, p.174). They posit that our understanding of screen music for media emerges from “multi-modal flows of information” (p.179) – i.e., the congruence between musical structures and visual structures is only part of what determines the emotional reaction of the listener – and suggest that "the binding of sound and image achieves an interpretational influence before it would be consciously evaluated as disruptively mismatched" (p.176). These relationships between music and image go beyond those suggested by Chion as empathetic ("music that is, or appears to be, in harmony with the tone of the scene;" Chion, 1996, p.8) and anempathetic (music which "exhibits conspicuous indifference to the situation;" Chion, 1996, p.8) and includes what is often described as "incongruent music" (music which "lacks shared properties" with the narrative or visuals; Ireland, 2017, p.21).

As explored by David Ireland, 'incongruent' music can provoke strong reactions in audiences as it draws upon their intertextual expectations of how music and image will interrelate – i.e., their "cinematic literacy" (p.32). Ireland challenges the notion that incongruent music functions solely through being "surprising or unexpected" (Ireland, 2017, p.22) given the presence of such incongruence in a wide array of 'mainstream' films and instead argues that incongruence draws upon "sophisticated levels of cultural and intertextual expectations" (p.23). In essence, viewers recognise that there is an accepted approach to scoring audio-visual media which utilises many of the emotive approaches discussed toward the beginning of this section, recognise when this accepted approach is being contravened, see the incongruent music as juxtaposed against the expected music, and derive a response from that juxtaposition through a range of responses, including "invitations to reconcile the differences or to appreciate the tensions they present" (p.32). This explains why, as Willemsen & Kiss observe, "incongruent music cannot be fully habitualized through overuse of the practice" (Willemsen & Kiss, 2013 p.181): incongruent music is not a transgressive approach to portraying emotion through movies, but simply *another* approach to portrayed emotion through movies.

Despite the growth in the use of incongruent music, it has been less explored in research than congruent music (e.g., empathetic/anempathetic). This may potentially be due to the reliance that incongruent music has on a learned media literacy: much use of incongruent music relies on the viewer's understanding of what congruent music would be. Prior research has recognised the importance of musical literacy when considering how music will change the viewer's (emotional) interpretation of a portrayed narrative. For example, Cohen recognises that their congruence-association model depends upon a diachronic (i.e., non-synchronous) model of association: “early immersion in electronic multimedia primes the efficiency of the cognitive processes represented in congruence-association models, in the same way that exposure to a language or music practice early in life provides facility that cannot be easily duplicated later in life” (Cohen, 2013, p.36). More specifically, Cohen suggests that “during adolescence, some aspects of musical grammar are laid down… If one is born into a movie-going environment, then one’s facility in interpreting moving images will differ from that of someone who has no exposure to TV or movies” (pp.36–37). It therefore seems plausible that emotional affect created through incongruence has received less examination because it is only in recent years – with the emergence of easy and continuous access to digital media – that viewers have been exposed to sufficient audio-visual materials to reliably derive an intended emotional relationship from the juxtaposition of ‘incongruent’ music and visuals. However, despite this relative scarcity of examination, it is clear that incongruence is an important part of how audiences interpret the emotional messaging of music within audio-visual materials, and that it will become an important part of considering how music, narrative, and emotions interact in future virtual experiences.

4.1.3: Musical Immersion

Music within cinema is also known to engage the viewer beyond the emotional embodiment set out in section 4.1.2: such as with Kathryn Kalinak’s acceptance of Hanns Eisler and Theodor Adorno’s framing of film music as a form of “distraction” which “turns the spectator’s attention away from the technological basis of film” (Kalinak, 1992, pp. 33–35)[58]. Similar connections can also be observed within video games. An increased sense of embodiment, as discussed in section 3.1.2, is likely to also

[58] See also Eisler & Adorno, 1947, pp.59–61: “[Music]’s social function is that of a cement, which holds together elements that otherwise would oppose each other unrelated – the mechanical product and the spectators, and also the spectators themselves….[music] is the systematic fabrication of the atmosphere for the events of which it is itself part and parcel. It seeks to breathe into the pictures some of the life that photography has taken away from them. Not for nothing did music migrate from the orchestra pit to the screen, of which it has become an integral part. … The collective function of music has become transformed into the function of ensnaring the customer.”

reflect an increased sense of presence within an experience – the feeling that you are someplace other than where you are. Music has been suggested to increase the sense of presence within virtual reality.[59] It is also known that presence can help lead to the development of a sense of 'flow' which is, in itself, closely related to immersion, as discussed in section 0.3.1.[60] In particular, we could see this as a form of narratival immersion within virtual experience emerging from the sense of embodiment discussed in 3.1.2.

However, alongside the psychological immersion offered through these metrics, an additional form of immersion has perhaps been neglected in literature surrounding music in virtual experience. Gorbman draws on Kurt London's work on early film music to suggest that musical accompaniment may have been desired in early film to cover up the sound of the projector (Gorbman, 1987, pp. 36–37). London's early observations are now typically understood to have been a supplemental reasoning for music that existed alongside the other purposes for music identified throughout this section. However, whilst it was not discussed in these terms– the field of film studies being in its infancy at the time – we can now understand this as being an increase in perceived immersion resulting from increased sensorial immersion: the mechanical noises which existed outside of the diegetic world being screened off by sounds and noises conceited as part of the presentation of the diegetic world.

Similarities should be observed to the observations of Walther-Hansen & Grimshaw-Aagaard drawn upon in Chapter 2: they identify that within virtual reality experiences, "excessive noise or unwanted sound has been implicated in stress … sound masking reduces the impact of incoming sounds, thus preventing them from interfering with the user's mind-internal cognitive tasks (e.g., imagination)" (Walther-Hansen & Grimshaw-Aagaard, 2020). It seems possible, therefore, that music serves a similar function within virtual reality as Kurt London suggests it did in silent cinema: isolating the participant/viewer from externalities which would distract and detract from the experience. Laura Ermi and Frans Mäyrä in particular seem to suggest this as a key part of sensory

[59] See Fritsch & Summers (2021) for a brief review of relevant literature surrounding this topic.
[60] For more details on Mihalyi Csikszentmihalyi's concept of flow and resulting links to immersion, see section 0.3.0

immersion: “powerful sounds easily overpower the sensory information coming from the real world, and the player becomes entirely focused on the game world and its stimuli” (Ermi & Mäyrä, 2011, p. 101).

Within *Alyx*, it is notable that there are only very few moments where the game is not playing some kind of music. On those rare occasions where there is no music, the game typically provides ambient noise – the distant hum of electrical lines, wind blowing through tunnels, the rumble of passing trains, or just distant alien noises. It is likely that this music – and the corollary noises where there is no music – is intended to provide narratival immersion imagined by Kalinak, Gorbman, and others. However, alongside this narratival immersion is also experiential immersion: by raising the noise floor of the experience, minor noises and distractions from outside of the experience are moved beyond the user’s sensory horizons. Alongside the psychological immersion explored earlier in this section, we can recognise that music also serves to sonically – i.e., physically – immerse the user within a virtual experience.

Music may also serve as a form of bookends for an immersive experience, demarcating where an experience begins and end. Connections were already made in 3.1.1 to the use of Valve’s theme music to identify the beginning of the *Alyx* experience for participants. However, the continuous presence of sound may also provide a way for users to (subconsciously) identify when they are contained within the virtual experience. This may help them to distinguish between – for example – issues with the virtual hardware or virtual software, and deliberate decisions on the part of the developer to have a black-out or fade-out within the virtual experience. We could, perhaps, see similarities to the Max Rieger’s discussion of spatialisation-as-authentication in section 2.2.1: Rieger relied upon timbral qualities of sound to help players distinguish between diegetic and non-diegetic voiceovers (World XR Forum Crans-Montana, 2020, sec. 09:10–10:30), and we can suggest that providing players with a continuous source of mediatised sound may lead players to more easily recognise when non-mediatised (i.e., external / non-intentional ‘noise’) is infringing on their virtual experience.

4.1.4: Music-as-Interface

Both emotional and semiotic frameworks provide users with some information about the diegetic world. However, music can also directly provide information to participants within an interactive virtual experience. As music reacts to the player's actions in many cases, players can begin to connect musical changes to changes within the virtual software. Whilst this relies on both musical synchronicity and what Van Elferen describes as "musical media literacy" – the "fluency in hearing and interpreting film, television or advertising music through the fact of our frequent exposure to them and, subsequently, our ability to interpret their communications" (van Elferen, 2016, p. 36) – the presence of information within video game music is broadly taken as a given. It seems likely that similar purposes can be found in other interactive media. This category of musical purpose differs from those identified above as it requires active and conscious engagement with the music: the participant within these experiences is deliberately seeking to gain information from the music, rather than drawing information subconsciously (or accidentally) from the music of the game.

We can also recognise similarities between the use of music to deliver information in this way and the use of earcons discussed in section 2.2.1. Whilst non-diegetic music is (by definition) non-representational – i.e., there is no object within the virtual world which is playing that music, or which the music otherwise directly represents – it can still be used to provide abstract information: for example, music progressively changes in *PAYDAY 2* (Overkill Software, 2013) to allow the player to keep track of the intensity of the waves of enemies that they should expect to face; music for some boss fights in *Final Fantasy XIV* (Square Enix, 2010) changes for each stage of the boss fight, which allows players to prepare for changes in the underlying systems (Tatlow, 2022, pp. 216–218); many games have different music for combat encounters, which can also allow players to identify when an enemy is engaging them in combat even if it is not visually obvious (e.g., see Lamb & Smith, 2018). These examples go beyond the semiotic considerations in 4.1.1 by being explicitly trans-diegetic. That is to say: the information conveyed by these musical changes does not provide information about the narratival world but rather provides information about the ludic world. Correspondence to Ben

Winters' theoretical framework for music can be observed. Winters suggests that music exists within a "distinctive film reality … a distinctly *musical* reality where, generally unlike our reality, un-sourced music 'naturally' often pervades the space" (Winters, 2014, p. 197, emphasis original); the ludic reality for video game music suggested by music-as-interface relies on the ability of music to pervade the play space and be responded to by both players of the game and their avatars within the game.

4.1.5: Purposeful Music – A Summary

Music within virtual reality – and virtual experience more generally – can therefore be shown to provide information to the user through semiotic topics and tropes, facilitate and enhance sensorial immersion within games through screening external noises, increase the sense of embodiment (and thereby the narratival immersion) experienced within games through creating emotional synchronicity between player and avatar, and provide direct information about the virtual world as a form of auditory interface. However, achieving these purposes within virtual experiences – and specifically within virtual reality – can be more challenging than achieving these purposes within more traditional approaches to cinema and video game. In part this is because virtual reality technology is still in its infancy compared to these other fields: cinematic music has been evolving over more than a century whilst video game music dates back to at least 1980 – when Namco released the arcade game *Pac-Man* (Namco, 1980) with a short looping musical score.[61] As virtual reality has had less time to address fundamental implementation issues, then we should expect approaches to music in virtual reality to be, as yet, less sophisticated to some extent. Alongside these pragmatic concerns, however, there also exists theoretical issues within music for virtual reality. These are issues inherent to the

[61] The exact historical origins of video game music – and of computer-based music playback – is a matter of some debate. Some limitations can be suggested: e.g., the origins of computer music have been suggested to begin as early as 1951, either on the Ferranti Mark 1 general-purpose computing machine in the UK, or on the CSIRAC computer in Australia (see Michaels, 2008). We also can consider *Pac-Man* to be a definitive point at which music is integrated into video games. However, documentation surrounding music in video games across the intermediary decades is not always well-evidenced; it is possible that earlier examples of video game music do exist before *Pac-Man*. However, this historical debate extends beyond the limits of this thesis.

format of virtual reality. The next section of this thesis examines some of these issues and explores how they are expressed in unique ways in virtual reality.

4.2.0: Musical Issues for Virtual Reality

Whilst there are many issues with the composition, implementation, and reception of music within virtual reality – and more broadly within virtual experience – three major issues can be acknowledged as posing particularly difficult issues. We identified at the beginning of this chapter that most music within most virtual experiences poses a risk to the authenticity of the virtual world as non-diegetic music does not have an obvious source within the virtual world. This was acknowledged within the opening of this chapter by classifying music as an important example of *nondiegetic* sound – sound that is outside the diegetic world. We also expect, as explored in Chapter 3, that authenticity is important to virtual experiences as authentic experiences are typically perceived as more immersive along one or several of the paradigms of immersion set out in 0.3.0 – e.g., audio which is spatially plausible for a given virtual experience means users face less challenges within the virtual experience and therefore can become more immersed in the virtual experience. This echoes Slater's statement that participants must believe that a given virtual experience is plausible – "that what is happening is real even though you know that it is not real" (Slater, 2009, p. 13) – in order to experience perceived immersed. However, as set out by Ekman, music is "obviously constructed" (Ekman, 2008, p.5) – this poses a clear conflict with the exploration of authenticity set out throughout Chapter 3. This means that authenticity and music is a major conflict which must be addressed in the considerations of music for virtual experience within this chapter.

Synchronicity is another issue which has particular relevance within virtual experiences. Synchronicity between sound and image is a well-explored area which is known to require a degree of careful consideration in both composition and implementation. However, virtual reality poses specific challenges as many of the techniques used to manage synchronicity may be problematic. Section 4.2.2 examines techniques for modifying music to allow for closer musical synchronicity and discusses issues in visual synchronicity. Finally, Section 4.2.3 gives consideration to how music is spatialised within virtual reality, which can pose a number of difficulties for composers and users.

4.2.1: Musical Authenticity

Despite the “obviously constructed” (Ekman, 2008, p.5) nature of non-diegetic music, which may seem to challenge some notions of authenticity as set out in Chapter 3, audiences nonetheless expect music within film, video game, and other audio-visual media. This expectation is likely because of long-standing relationships between music and narrative, which Gorbman observes “[started] no doubt even before the Greeks, continuing through the Middle Ages and the Renaissance, and resurfacing to popularity in the late eighteenth-century French melodrama” (Gorbman, 1987, pp. 33–34).[62] Jessica Green argues that 21st-century audiences “have come to accept film music as an integral part of what it means to watch a film. Films that fail to use much music…will often have a problem involving the audience as completely as films that embrace music as a tool” (Green, 2010, pp. 81–82). [63] Similarly, Christian Metz describes music as one of the four key elements which are fused in cinema: image, speech, music, noise (Metz, 1990, p. 58). Given the audience expectation for music present within media such as film, audiences and participants within virtual reality therefore seem likely to expect music also as part of their virtual experience.

Music within virtual experiences therefore has an inherent conflict: audiences and participants want (or need) music within audio-visual experiences, but music may also create inauthenticity. Inauthenticity within the diegetic world is a vector for detracting from the participant’s perceived immersion: the participant is reminded of (and therefore less isolated from) the real-actual; the participant may reject the (inauthentic) virtuality. Further, excluding music from virtual experiences

---

[62] Much other evidence of these historical relationships can be found, whether that be the Roman musical comedy (see Gellar-Goad & Moore, 2015), musical interludes in the plays of Shakespeare and other Tudor writers (Mann, 2012; von Ende, 1965; etc.), the development of the opera (and other similar marriages of drama and music) as distinct art forms, or elsewhere.

[63] This may not have always been the case. The presence of music in early sound cinema (and silent cinema) was considered somewhat controversial. As a representative example, Harry Alan Potamkin argues that cinema music is inherently inauthentic: “there is no absolute justification for the use of music to accompany the film… In the movie [silence] is a paramount virtue, and an art can transcend its material lineaments only by the play of its virtues, or better, by virtue of its inherent characteristics” (Potamkin, 1929, p.281).

may lead audiences to feel less involved (i.e., less narratively immersed) within the virtual experience based on literature surrounding film and other audio-visual media. This is because music is the *de facto* expectation within 21st-century conventions for audio-visual experiences of any kind.

Two separate reasons can be suggested for the apparent conflict between music and authenticity. First, users may identify music as part of an 'authentic virtual experience' – whilst music is objectively inauthentic in many cases, users may expect it to be present due to its presence in other adjacent media such as cinema or video game. This subjective authenticity acknowledges that whilst music is inherently inauthentic to the virtual world, it may still form part of the virtual experiences that take place within it. This could be reframed as users recognising that virtual experiences are predominantly hyper-realistic: the plausible expectation within a hyper-realistic environment is that background music exists and relates to the virtual world through cohering to expected behaviours.[64]

Alternately, the benefits of music may outweigh any potential loss of narrative immersion caused through the presence of music within the virtual experience. To some extent, we can recognise that the loss of immersion caused by music relates to the plausibility of that music. If the music seems implausible – i.e., it does not correlate to the expectations of the audience – then it likely to be understood to be immersion-breaking. We can often see this implicitly reflected in discussions surrounding "appropriate" music for any given experience: that is to say, music which is congruent with typical user expectations is often considered appropriate, whilst music which is incongruent with typical user expectations is often considered inappropriate, irrespective of whether or not the music would be objectively authentic to the narrative world being portrayed.

In essence, we can recognise that two separate forms of authenticity can be observed within user perspectives of sound in virtual experience. The first form is that which was discussed

[64] Whilst Wierzbicki highlights that affective sound "almost never occurs whilst underscore is in progress" (Wierzbicki, 2016, p. 156), similarities can nonetheless be drawn between the authenticity of music and authenticity of the *expectation* of music and Wierzbicki's differentiation between sound *effects* (i.e., the "credible" sounds) and *affective* sound (i.e., sounds which trigger specific emotional responses) can be drawn here: i.e., whilst the effect of non-diegetic music is frequently authentic to the objective reality of the created world, it can nonetheless have an affect which audiences recognise as authentic.

throughout Chapter 3: authenticity *of* sonic material. In a musical sense, that would relate to many of the considerations proposed throughout section 3.1.1; is the sound (i.e., in this specific case, the music) faithful to the approach, manner, and style of the era which a virtual experience claims to be recreating? This is often, at least to some extent, subjective: music for *Odyssey* discussed in 3.1.2 was considered to be authentic by listeners because they perceived it as being 'Greek' – it matched the expectations that participants of the *Odyssey* experience had of 'Greek' music – despite not being a historically authentic recreation of the music. Often listeners draw upon different subjective evaluations of plausibility also. Lind describes how discussions with students surrounding the game *Cuphead* (2017) exposed this subjective authenticity:

> While many of the historical cartoons referenced in the game still aired as reruns on television during my childhood in the 1980s, their problematic racial and gender stereotyping mean that they are no longer acceptable to air today. In my gameplay, nostalgia was evoked through my specific memories of the source materials. … For my Gen Z students, on the other hand, the game function… as a retro-genre game because of its ample use of big band jazz, muted colour palette, and animation resemblances to Mickey Mouse. (Lind, 2022, p.2)

We can also observe from the discussions above that audiences and participants for virtual experience have expectations of how music should function within a virtual experience; users have expectations that music should be *authentic to the expected mechanisms of virtual experience*. As with *Odyssey*, and as suggested in the conclusions in section 3.1.3, the primary consideration of authenticity is not an objective authenticity of the event but rather the subjective valuation of authenticity by the participant. As music is a *de facto* expectation of narrative audio-visual media, participants within virtual experiences that present narrative audio-visual media (such as in virtual reality games, or augmented reality cinema, or mixed reality theme parks) will also identify that the music is authentic to the media of virtual experience. An experience which does not use music in the expected way, or which lacks music entirely, may be considered to have 'betrayed' the audiences' expectations, and thereby be seen as inauthentic because it will draw the audience's attention to the artificiality of the experience. These considerations can aid clarity when considering the function of music within non-narrative audio-visual media, as discussed in section 4.3.1.

These discussions of authenticity must also be contextualised alongside the work of Winters, who objects to the traditional notion that non-diegetic music does not issue from, or belong to, the world of the story: “Whereas David Bordwell and Kristin Thompson believe that viewers ‘understand that movie music is a convention and does not issue from the world of the story’, I am more inclined to ‘imagine’ that it does issue from that world; why else would I recall it along with the other elements of a story? Nor do I need to look for its visual ‘source’ to regard it as such” (Winters, 2010, p. 232). As Winters later explains, this is why audiences will “imagine the film’s score” when “reimagining the filmic world” (Winters, 2014, pp.182–183); Winters suggests that music is “not a cold conveyer of cold narrative information” but rather is based on a “shared experience of the music” (p. 184) wherein the audience and characters may have a conjoined (or, at least, simultaneous) emotional response to the music because they both hear it and therefore have the “potential for a shared emotional experience (p. 195). Succinctly, Winters is arguing that non-diegetic music is no less inauthentic than diegetic music as both diegetic and non-diegetic music are inherently part of filmic realities: “removing music from one side of the interaction – the film’s world – would seem to be damaging to understanding the characters in fiction… the fictional worlds encountered in cinema are ones in which musical content may be shared between audience and character” (pp. 195–197). Whilst this approaches different aspects of musical authenticity than those highlighted above by Lind, it offers a new understanding of Green’s emphasis on the importance of the music to filmic reality: music may be authentic to the expectations of audio-visual worlds because music is an inherent part of the filmic reality.

4.2.2: Musical Synchronicity in Interactive Experiences

Synchronicity between music and image has been studied extensively in both film and video game. Within film studies, Kalinak posits that music and image are now expected to interrelate by audiences, and that audiences (and participants) may feel betrayed if the music and image fail to do so: “Our tendency as spectators would be to perceive music as meaningful… We might even expect some kind of narrative complication because of [suspenseful] music. If these expectations are thwarted, most spectators would feel manipulated, even cheated … The farther music and image drift from a kind of mutual dependency, the more potential there is for the disruption of the cinematic illusion” (Kalinak, 1992, p. 15). Those studying video games broadly concur but identify that video games may pose a particular problem for synchronising music and image as video games are indeterminate texts: the player’s ability to interact with the world of the game means that they control, to some extent, the length and duration of the game. As Scott Brendan Cassidy frames it: “A particular game session produces a linear narrative when the player-author collaborates with the game’s original “programmer-author” by interacting with the text. … Pressing the “jump” button or the “shoot” button are just a few of the many adventures the player can choose. The act of pressing one button or the other produces a single actualized narrative out of the many that a video game allows” (S. B. Cassidy, 2011, pp. 295–296).

Several approaches to scoring indeterminate texts have been proposed within video game design. Historically, video game music has predominantly used dynamic music to synchronise music and image. Very basic dynamic music engines have existed since early in video game development: for example, games such as *Super Mario Bros.* (1985) changed the music heard by the player dependent on the current location of the character. We could recognise this as a form of dynamic music engine, in opposition to games such as *Tetris* (1984) which had a fixed linear score. The development of dynamic music engines (sometimes referred to as ‘adaptive music engines’ at the time of development) continued across many regions. As described by Donnelly, these engines often take the form of “music in the form of loops, which can be repeated endlessly to accompany certain sections of the game…to be sensitive to certain aspects [of audio-visuality] that have not been important in other forms of music” (Donnelley, 2016, pp.80–81).

One such engine used for some games programmed in America is the iMUSE engine designed for use in *Monkey Island 2* (LucasArts, 1991) and later distributed as audio middleware for other video game projects. iMUSE allowed composers to program the game to use musical branches (i.e., transitions between (sections of) musical cues without a fade) and musical loops (i.e., repeating (sections of) musical cues) to allow the music to adapt to the actions of the players dependent on low long it took for players to complete a section of the game (Sweet, 2015, pp. 98–99). As the engine predominantly functioned as a MIDI-controller, other possibilities were also realised throughout many iterations of the engine. For example, *Monkey Island 2* frequently used the same musical material but with different instrument voicings: an adaptation possible due to the use of MIDI synthesis within the engine. More recent dynamic music engines arrange pre-defined musical materials in real-time to generate a musical soundtrack using more complex methods than simply branching and looping. Richard Evans identifies two types of layering. First, Evans describes "interchangeable layering" in which "different layers of music replace one another by fading in and out. The replacement may contain new harmonic, melodic, or rhythmic material, or a variation of orchestration" (Evans, 2019, sec 3.2.1) Second, Evans builds on Sweet's "additive layer technique" (2015) to describe an layering technique where "different layers of music are added or subtracted from a foundational base layer of music…new layers being added do not serve as a replacement for existing layers that are already playing" (Evans, 2019, sec. 3.2.2).[65]

Other approaches to video game music generate music in real-time to accompany the narrative of a game. For example, Peter Langston's soundtrack to *Ballblazer* (Lucasfilms Games, 1984) is often considered one early example of generative music. The game draws on 32 predefined melody fragments (referred to by the composer as "riffs") which are then combined together using an algorithm that determines "how fast to play the riff, how loud to play the riff, when to omit or elide notes, [and] when to insert a rhythmic break … in a facsimile, unimaginative, (and slightly lazy) [jazz] guitarist" (Langston, 1988, p. 6). The music also has a bass line and drum part, which are also algorithmically generated. The resultant semi-randomised musical elements are then combined into a single soundtrack played back to the listener. Langston also notes that some accommodations must be made to create musical 'smoothness': "To choose the next rift to play, the program… randomly

[65] The nomenclature of these musical engines is undetermined. For example, Evans' model is similar to Elizabeth Medina-Gray's exploration of vertical and horizontal organisation of musical material, which Medina-Gray describes as a 'modular structure' for composition (Medina-Gray, 2014).

selects a possibility. From these it chooses the riff that is "easiest" and "smoothest" to play, i.e. the riff whose starting note is closest to one scale step away from the previous riff's ending note" (Langston, 1988, p. 7).

Both generative and dynamic music engines integrate musical loops to create a consonant and unified musical experience for the user. However, two key differences should also be observed between dynamic and generative music engines: adaptability and randomness. Adaptability refers to the ability of generative music engines to change the music materials on the fly. Dynamic music engines piece together predetermined musical material in real-time, whilst generative music engines can create new musical material for use in real-time. Langston describes how his generative music engine could be further expanded to be capable of modifying individual notes within each riff to allow it to fit an underlying chord progression (Langston, 1988, p. 10) or to pseudo-randomly interpolate notes within a broader melodic structure (or between melodic structures) to increase the range of riffs which the algorithm can consider selecting whilst maintaining musical 'smoothness' (Langston, 1988, p. 15). This allows musical materials to be combined in many different ways – as described above, Langston's score for *Ballbreaker* included metrical variation of the materials, meaning that 'riffs' could be played back at different speeds, which is not possible in most dynamic music engines. Randomness refers to the use of pseudo-randomisation[66] within algorithms for generative music engines. Most dynamic music engines are deterministic: if provided with the same original state, they will always produce the same musical material (if arranged in a different order). Generative music engines are nondeterministic: if provided with the same original state, they will produce similar, but different, musical material and therefore similar, but different, musical experiences.

[66] It should be noted that computers cannot generate a truly random number as computers are functionally mathematical operators, and all mathematical equations are both deterministic and reversible. Computers use pseudo-random number generators (PRNGs) as part of these "random" algorithms. PRNGs are known to be problematic: e.g., IBM's PRNG "RANDU" is thought to have disproportionately influenced scientific studies in the mid-20th century as the operation used to generate a "random" number was later found to be recursive (i.e., specific sets of values would repeat).

However, generative music does pose some challenges for synchronisation. In an interview, the composers of the original musical materials used for the generative soundtrack of *No Man's Sky* (2016), 65daysofstatic, state that "most generated [music] tends towards toward ambient and quite granular, soft, synth-y stuff" (Nast, 2016). This is framed as an issue in two ways: firstly, generative music engines might create less memorable moments as it can be hard to create, utilise, and develop distinctive melodies within a generative music engine; secondly, the generative music engine might (seem to) react too slowly to player actions – either because it can be difficult for the game to rapidly generate new musical materials in a timely manner, or because the music is slow-paced and 'ambient' and therefore changes too slowly for the musical changes to be associated to the gameplay changes. Similarly, Langston identifies that computer programs are "unable to manipulate semantics" (Langston, 1988, p. 7) – referring to the need to supply at least *some* fundamental musical material and musical rules to the generative music engine to prevent it from only generating dissonant noise. These issues can create difficulties in using music for the purposes established in the earlier sections of this chapter: emotional synchronicity between user and avatar may be less likely if the generative music engine can only provide music with certain types of emotional resonance, or if the music engine reacts too slowly to the player action. The music will be less likely to "support narrative, foster player agency, or facilitate incorporation into the game body" and thereby fail to support the "vital bridge between the bodies of player and avatar" (Plank, 2021, pp. 284–285), which will reduce the musical immersion within the game described in section 4.1.3. We can imagine that, in extreme situations, the disjunct between narratival emotion and musical emotion will draw attention to the artificiality of the experience, contravening the systems of engagement suggested by Brown & Cairns (2004, p. 1299).

Many discussions surrounding synchronicity in audio-visual media specifically claim to focus on interrelationships between image and sound – that is to say, audio-visuality. However, it should be noted that this is not precisely true. Rather, music is typically synchronised to a narratival or contextual event within the cinematic or ludographic world; music is intended to be emotionally resonant rather than audio-visually synchronised. For example, in *The Elder Scrolls V: Skyrim* (Bethesda Game Studios, 2011), the music changes to combat music the moment a hostile entity enters an invisible zone centred on the player character. There is no requirement that the player *see* the hostile entity: the synchronicity is not between sound and *image* but between sound and *diegetic event*. This is likely an area of unrealised study as the predominant focus in the last century has been on film and cinema, where the viewer has no control over the framing of the diegetic world. Whilst visual synchronicity is a well-understood issue within cinema sound, the filmmakers can determine how the audial framing should support (or contrast) the visual framing. Conversely, interactive experiences that allow the user to – in the words of Janet H. Murray – “move closer, lean into the shot and get a better view” (Murray, 1997, p. 47) are, effectively, delegating responsibility for the visual framing of the diegetic world. This includes both virtual reality cinema – especially those cinematic experience from a first-person perspective where the audience take the place of a character within the narrative – and within essentially all video game contexts. Within these interactive experiences, the ability to freely look around a three-dimensional space creates further issues for synchronicity. Whilst we might expect virtual reality films – as a fixed narrative text with little apparent interactivity – to be able to take similar approaches to synchronicity as cinematic film, the possibility that the user is looking in the wrong direction at a key narrative moment requires overcoming additional challenges. We could imagine a realisation of this issue: the users look towards a distant hill, dramatic music rises, the user waits to see what is coming over the hill, only for the music to fade out. The user looks around in confusion and finds the characters at the very edge of their sight behind them. The music as asynchronous; the player (and the music) has missed the moment. Whilst this is a potentially comic framing of the issue, this is – nonetheless – a potential result of allowing users to have control of the visual framing of the experience in virtual (reality) experiences.

Within virtual experiences where users have control of the visual framing of the experience, designers must consider whether the music reacts primarily to the viewer or to the diegetic world. If the music changes simultaneous to the emergence of diegetic events, then the viewer may not be looking in the right direction at the right moment. This can reduce the efficacy of the musical purposes. This can also cause confusion (e.g., if the user is unable to find the diegetic event). However, it also encourages users to *seek out* the diegetic event by looking around and trying to find it: e.g., as described by Kenny Young for his work on *Tethered* (2016): “Unlike the visual sense, sound you can hear all around you. That’s why this is worth looking into [music] to communicate information about things you cannot see” (Young, 2017, sec. 07:30–08:15). Synchronising diegetic events and music can also help to prevent ‘stalling’ in virtual experiences – where events do not progress because the user is unsure how to proceed – by encouraging the user to continue exploring the virtual space.

Conversely, where the music reacts to the viewer, then it will wait for the viewer to be looking in the right direction. Once a specific object or event enters the frame (or is looked at by the user) then the musical cue begins. This allows the music and event to synchronise, tying them together. This is especially the case in situations where the diegetic event will not begin until it enters the user’s field of view. In some cases, this is possible through using field-of-view algorithms: i.e., if a diegetic object is within the field-of-view of the observer, then begin the associated diegetic event. In other cases, eye-tracking hardware and software can be installed to allow developers – and the resultant virtual worlds – to know exactly what a user is looking at. HMDVRs are already beginning to provide eye-tracking capabilities for designers to utilise, and these capabilities are being used for other aspects of sound. For example, Walther-Hansen & Grimshaw-Aagaard suggest that eye-tracking technologies may play an important part in determining what sound sources should be heard, especially in augmented/mixed reality contexts that use audio-scene segregation solutions (Walther-Hansen & Grimshaw-Aagaard, 2020, p. 7). Additional use of these technologies to inform musical changes should be expected, though it should also be observed that delaying diegetic events or musical cues may also mean extending or prolonging other aspects of the experience, as suggested by R. Stevens: “If the music were truly interactive then this would raise the possibility of thresholds and triggers being altered, or game events waiting, in order to enable the synchronization of game events *to* music" (R. Stevens, 2021, p. 91).

As a new technology, generative music engines are lacking – to some extent – in the sophistication of dynamic music engines. However, the potential for these approaches to allow for more synchronicity between music and image is clear: different elements of visual-music synchronicity and diegetic-music synchronicity can be combined to create an 'audio-visual-diegetic synchronicity'. Just as dynamic music engines have adopted some elements of generative music engines to create a more flexible and reactive experience, we can expect that future virtual experiences will adopt some field-of-view tracking or eye-tracking approaches in order to manage visual synchronicity and ensure that audio-visual-diegetic synchronicity is achieved holistically. Designers and composers will need to consider carefully the different musical purposes and select appropriate approaches to synchronicity to ensure that their musical goals are achieved in each instance. For example, semiotic or leitmotivic music is likely to benefit from a visual-musical synchronicity as this connects the represented object in musical terms. Conversely, emotional music is likely to benefit from diegetic synchronicity as the user will be emotionally prepared for the diegetic events when they (eventually) find them. We can imagine that a sophisticated music engine would be able to playback or generate a musical cue which provides the correct emotional signification to the user whilst also being able to near-instantaneously interject semiotic/leitmotivic content (e.g., a melodic phrase) when the user visually locates a specific diegetic object.

4.2.3: Spatialising Music in Virtual Experiences

Screen-based media provides a two-dimensional representation of the three-dimensional real world. This means that depth can be difficult to judge in screen-based media as binocular cues (e.g., the differences between the visual signals received by each eye) are small. Audio has often been considered as part of the solution to this. In sections 2.1.0 and 2.2.1, we discussed the importance of spatialising sound. It has been shown that sound is often used to help users navigate within virtual reality: through listening to stereo or binaural audio, users can determine the location of diegetic objects within the virtual world. Whilst audio spatialisation in screen-based media maintains a continuity independent of the spatialisation of the visual framing (see section 2.2.3), these approaches to diegetic sound nonetheless allow users to navigate more confidently within virtual experiences.

However, music is not diegetic sound; music does not emanate from objects or places within the diegetic world. Despite this, music is still nominally spatialised as sound can be thought of as – inherently – spatial. This creates issues for audio designers: a decision must be made about how best to ‘spatialise’ music (and other non-diegetic sounds) in order to minimise the potential impact of these issues. This is especially relevant to virtual reality and other virtual experiences which make extensive use of binaural sound spatialisation to heighten perceived immersion; within these contexts, non-spatialised music may challenge narratival immersion as it draws attention to the external world beyond the diegesis, or increase the challenge posed to the user and thereby reduced challenge-based immersion. Winifred Phillips describes the issue succinctly: “if virtual reality is all about presence, about making players feel as if they exist in a real place – where is the music coming from? Does it also exist in the VR world, and if it doesn't, how do we introduce it so that it doesn't disconcert the player and interfere with the realism of the VR experience?” (Phillips, 2018a).

Phillips discusses a variety of approaches to spatialising music within virtual reality. Phillips first discusses the possibility of using only diegetic music: e.g., “all music comes from in-game sources that players can see in the environment around them” (Phillips, 2018c). As the music has a clear spatial source, spatialisation of the music can therefore be easily rationalised: players can receive the music as it would be spatialised within the live performance. Providing a representation of musicians within a virtual world, however, can be difficult. We can imagine that it would detract from the authenticity (and therefore reduce the naratival immersion) of the experience if, for example, a large orchestra followed the player around in *The Elder Scrolls V: Skyrim* to provide a ‘diegetic’ soundtrack. There may also be issues with games set in specific regions, cultures, or historical periods: for example, the music of medieval Europe is known to be significantly different to the music of cinematic medieval Europe – the music of cinema often uses anachronistic instrumentation (see J

Cook, 2020; K Cook et al., 2020). This can be problematic as audiences may not be able to understand the music (as set out in section 4.1.5) if it does not relate to their prior knowledge of media music, even if this is for excellent reasons. As this suggests, and as Phillips acknowledges, using *only* diegetic music requires “really specific circumstances to justify it” (Phillips, 2018c). It is clearly not a general solution to the challenge of spatialising music within virtual experiences.

Another approach suggested by Phillips is to use a non-spatial stereo mix for the music. Whilst this may reduce the player’s sense of presence – as the music draws attention to the artificiality of the digital space – it can also have benefits. Phillips suggests that this approach may be best where developers “don’t want a spatialized score to introduce any confusion [in understanding other audio cues]” such as in first-person shooters (Phillips, 2018c). In order to make this work, Phillips had to work out how to delineate the stereo music mix from the sound design of the VR world; they used dry acoustics (with no spatial processing such as reverberation) for the music to allow players to distinguish (subconsciously) between the diegetic binaural sound design of the VR world and the nondiegetic stereo music.

A third approach involves drawing on both aspects suggested above. By mixing two-dimensional stereo musical elements and three-dimensional binaural musical elements, Phillips suggests that composers can create a “stereo music mix that also integrates well into the three-dimensional soundscape” (Phillips, 2018b). Phillips suggests two particular approaches based on their work for *Dragon Front* (2016) and *Fail Factory* (2020). In *Dragon Front*, Phillips explores how the music transitioned from a nondiegetic stereo mix during ‘battle’ sequences to a diegetic binaural mix during ‘rest’ sequences (with the sound appearing to emerge from a speaker at the edge of the ‘rest’ zone). In *Fail Factory*, the music was incorporated into the sonic environment of the game itself: “bleeps and bloops and pitched to integrate with the score, and the bangs and clangs are timed to emphasize the tempo. Whilst much of the music is delivered to the player in traditional stereo, there are also lots of separate rhythmic and pitched elements that are spatially positioned on the game’s factory floor” (Phillips, 2018b). We might even imagine that this kind of spatialised music offers a direct example of Winters' notion of music being "issued" from the filmic (or, in this case: ludic) reality (Winters, 2010, p. 232).

Other composers of music for virtual reality experiences posit similar considerations. Joseph Trapanese discusses, for example, how "music [for film] is placed in a single space. Even when working in [Dolby] Atmos, there is an assumed seated position… but with VR, we can attach music and layers of music to objects within the environment. We can also go a step further and process music using ambisonics, truly playing it … around the listener's head" (Trapanese, 2020). Stephen Cox concurs, explaining that music could be remixed spatially in virtual reality: "We were all reinventing the wheel with VR sound. Does this wide thing work? Or is it better if we have instruments that are pin-point focused like a solo instrument?" (Martindale, 2017). Cox even suggests that VR offers new ways to experience the music of a game, explaining that "The biggest challenge is immersion. … The way the instruments interacted with the space is very important … [Music in VR] can be focused or spread in a way that wouldn't make sense if it were played back on speakers" (Balofsky, 2017). These composers go beyond Phillips' use of the diegetic environment as part of the score by re-spatialising the non-diegetic musical score within the virtual experience to aid the player's sense of immersion.

From these considerations, it is clear that the spatialisation of music must be considered as part of the design process for virtual reality experiences. At what points, and in what ways, will the music spatialised? How will this interact with the user's perspective and experience of the virtual world? As explored by Cox, Phillips, and Trapanese, the spatialisation of the music can distort the user's perspective of the virtual experience around them, aiding or disrupting their perceived immersion within (and understanding of) the virtual world.

4.2.4: Music Within Other Virtual Experiences

Throughout this chapter, a relatively limited perspective has been considered. Specifically, we have examined how music functions within virtual reality games and virtual reality cinema. However, virtual *experience* extends beyond virtual *reality* experiences. How does music function in other virtual experiences? Three specific contexts will be examined. First, we will consider the use of music within uncreated worlds; many of the purposes discussed thus far have been directly linked to either a constructed narrative or a constructed world with a specifically fictional focus. How does music function within virtual worlds which are created as non-fictional worlds (i.e., a recreation of a physical space)? Second, many of the examples discussed above come from experiences which are intended to be specifically *entertainment* experiences such as video games or cinema. However, there are many uses for virtual experiences which are not primarily entertainment-based (e.g., virtual experiences for training, assistive expanded/virtual reality). What purposes does music have within these educational experiences, professional experiences, and other similar experiences not focused on user entertainment? Third, many of the experiences explored within this chapter have focused on *screen-based* experiences, as can be seen in the discussion surrounding audio-visual synchronicity in section 4.2.2. However, many virtual experiences do not use screens. Does the change in the hardware configuration of the experience also change the purposeful use of music within an experience?

The primary conflict within ‘uncreated’ worlds (i.e., non-fictional/fantastical worlds or recreations of physical spaces) is that non-diegetic music is functionally inauthentic. Whilst this is a running theme throughout much of this chapter, the inauthenticity can become particularly problematic in settings which are intended to *not* have fictional elements at all. This might include virtual tourism, edutainment experiences (e.g., VR documentaries), virtual training resources (e.g., VR training), or assistive technology (e.g., augmented reality smartphone applications such as Google Maps). In many of these cases, there is no fictional reality which requires support from music.

Disneyland Adventures was discussed in section 2.3.2 as a recreation of Disneyland California. Whilst music is part of the virtual experience of Disneyland California, the way in which music functions within the virtual experience of *Disneyland Adventures* is fundamentally changed. Music within *Disneyland Adventures* is more continuous; music can be heard throughout the game at almost all times, whereas music within Disneyland California is more intermittent – some areas of the park have less music, and music stops and starts more frequently. Music within *Disneyland*

Adventures is more homogenised; music appears to emerge spontaneously from the atmosphere of the video game, rather than emerging from loudspeakers located throughout the park. Music within *Disneyland Adventures* is closely tied to the interactive state of the game; mini-games within *Disneyland Adventures* each have their own musical track which is closely synchronised to the game interactions. This emphasises the individual agency of the player within the game. Conversely, Disneyland California cannot permit every guest at the park to have the same level of individual agency enjoyed by players of *Disneyland Adventures* as this would create conflict between differing experiences of the park. Therefore, music in Disneyland California is not tied closely to interactions within the park, although it does often provide information to guests (e.g., parade music may signify that a parade is happening). The changes made to the music of Disneyland for the video game recreation emphasise the purposeful nature of music within virtual experience: since the purposes of the music within video game and the purposes of the music within theme parks are not similar, the overall musical experience of the two experiences differ significantly. This is particularly noticeable where the musical purposes do coalesce. For example, music will reflect the ‘zone’ that the player is within both within Disneyland California and *Disneyland Adventures*: e.g., when you are within the Main Street area, the game plays an arrangement of Disney songs such as Beautiful Beulah and Flitterin’ (from the Sherman & Sherman score to the 1963 movie *Summer Magic*). This is similar to the music used within the Main Street area of the real-actual theme park which arranges and mash-ups well-recognised themes from Disney canon[67] which allows both players of the game and guests of the theme park to use music as a locational device to understand where they are within the experience: i.e., they are hearing the music that plays in the Main Street area, therefore they have they arrived at the Main Street area.

Similar changes are made in other cases. The video game *Assassin’s Creed Odyssey* (Ubisoft Quebec, 2018) includes both an exciting narrative video game story set in ancient Greece and an educational experience intended to provide the opportunity to walk around an ‘authentic’ recreation of

[67] Disney provides examples of the musical loops that are designed for use within Disney attractions on YouTube (e.g., onstageDisney 2020a; 2020b; 2020c) which demonstrate different approaches.

ancient Greece, referred to as “Discovery Mode”. Whilst both of these modes use the same underlying music which was critiqued in section 3.1.0, the *implementation* of the music might be expected to differ. Whilst we can understand the use of ludic interactive frames within the video game story, we might expect that the implementation of music within the educational experience would be altered to more closely resemble real life: e.g., using only diegetic music performed by musician characters visible within the game. However, this is not the case. Rather, music is used as part of the ambiance of the educational experience. This decision by the developers prioritises the immersive and communicative aspects of music above attempting to create ‘objective authenticity’ of the music within Discovery Mode. No specific details behind this decision have been highlighted by the developers. However, there are many possible reasons why musical ambiance might be preferred: (a) users of the Discovery Mode may be implicitly aware that the experience is set within a video game recreation, and expect to have music as part of the ambiance; (b) musical ambiance may have been considered an essential way to ‘set the mood’ for the experience, helping users of Discovery Mode to recognise that the experience was intended to be a recreation of ancient Greece, or to recognise emotional beats within the educational narrative; (c) users of Discovery Mode may find the musical aspects of the experience pleasurable, either as a conscious decision that they enjoy the music, or as a subconscious rationalisation of the immersive use of music discussed in section 3.1.0. Other reasons to include music within ‘authentic’ recreations of real-actual locations could also be considered. No matter the definitive reason, we can recognise that the designers functionally elected to sacrifice authenticity in favour of a world which was considered more aesthetically enjoyable. This is not unusual: other representations of history, such as those found in museums, have also used music within exhibitions that aim to show visitors an ‘authentic’ part of a historical period.[68] As explored in section 4.2.1, various academics have noted that the challenges offered to objective authenticity by these creative practices have been broadly disregarded in favour of creating a recognisable aesthetic

[68] A number of recent projects utilising music in museums are set out in a special edition of *Music in Art*. For relevant papers, see: De Visscher (2014), Deters (2014), Gahtan (2014), Lobley (2014), Tabellini (2014), Varsányi (2014), Wyatt (2014).

for participants; the subjective authenticity experienced by the audience in placed above the objective authenticity to the original real-actual experience that is being simulated within the virtuality.

This suggests that there are two key types of authenticity within virtual experience: authenticity *of* the experience, i.e., how close the experience is to a perceived reality; and authenticity *to* the experience, i.e., how close the experience is to the user's expectations of the experience. Virtual experiences outside of video games help highlight these two types of authenticity, as we can recognise that participants still expect authenticity *to* the experience, even where that conflicts with authenticity *of* the experience. However, the other expressions of virtuality explored above are principally still entertainment experiences which take place in a virtualised world: *Assassin's Creed Odyssey* portrays a created world, and *Disneyland Adventures* is based on the mixed reality environment of Disney theme parks. Can we observe a similar determination of 'authenticity' elsewhere, where virtual experience is based on a non-virtualised real-world experience?

The use of music within virtual recreations for purposes such as training emergency responders or military personnel is not well-known and has not been widely discussed within materials on this type of virtual experience. Unfortunately, this makes it hard to determine the exact purpose of music within existing virtual experiences. It seems likely that this is because music would be considered more problematic in these instances as the potential benefits for the instantaneous experience would potentially detract from the long-term benefits of the experience: e.g., a police trainee could come to rely on musical cues to understand how to interpret simulations of real-world situations and be less able to interpret events that emerge during their real-actual work where music does not function. This situation could emerge as a result of deliberately interactive cues placed within virtual training experiences, or because of associations developed accidentally through something similar to Annabel J. Cohen's congruence association model (Cohen, 2013) or through accidental or spontaneous synchronicity (i.e., the participant believes that the music changes for a specific purpose to communicate a specific message, even though it does not).

Viewing it through the dyad of authenticity to/of virtual experience, we can recognise that music is likely not used in many cases because whilst it may increase perceived immersion (e.g., by blocking out the real world through sensorial immersion) it nonetheless decreases the subjective authenticity of the virtual world across both types of authenticity. That is to say: music would be inauthentic *to* the user’s expectations of the virtual experience and inauthentic *of* the real experience being recreated. Further research into these areas may be useful; both to determine whether music could be used purposefully within real-actual environments, and to determine whether the purposeful use of music within virtual experiences for training contributes or detracts from the overall long-term benefits. This could help establish the limitations and/or essentiality of authenticity to virtual experience, which may be useful as real-world workflows integrate augmented reality and mixed reality solutions. However, whilst these possibilities exist, it should be highlighted that no footage or reports were found to indicate that music is currently used in virtual training environments.

Consideration of how *inauthenticity* may be used by developers should also be noted. Michiel Kamp explores how the designers of *Google Earth VR* use music as a way to “afford movement through space” in order to create a “’cohesive story’ or holistic experience” (Kamp, 2021, pp. 729–730) and as a form of “uniquely conspicuous ’sonic wallpapers’ that make VR (anti-)environments very much unlike our own” (p. 731). Kamp suggests that the conspicuously inauthenticity of music allows for users to feel liberated from their human limitations and, thereby, embodied with the godlike powers offered by apps such as *Google Earth VR*, where users can fly around the world in the blink of an eye. This posits a pragmatic purpose for *inauthentic* music within virtual experiences.

Alongside the edutainment experiences such as Discovery Mode which have already been discussed, we should also consider the place of music within professional training experiences such as those designed by Google, Microsoft, and others. As with other experiences that re-create a real-actual environment, it is unclear whether music creates authenticity within the virtual experience. For example: another area in which music can be considered is in remote working environments such as Microsoft Teams. Music seems unlikely to provide a useful semiotic purpose in most cases: no matter how stern your manager may be, they are unlikely to appreciate boss music accompanying their meetings. Similarly, emotive uses of music seem out of place within most remote working environments. Consideration could be given, however, to the use of music for immersive reasons; music could help improve the overall quality of remote meeting experiences and make them more pleasurable for participants. This is especially the case in situations where colleagues have low quality hardware, or inadequate audio compression algorithms are used by the meeting software, or where noisy virtual environments collide within an online space. Remote meetings are discussed in more detail in section 5.4.0 – 5.4.3, but it seems likely that – at least in this instance – any users who would find music a useful addition to these kinds of virtual environment would choose to listen to it (e.g., an open Spotify client as a background process) without broadcasting it through the remote meeting software.

4.3.0: Virtual Experiences of Music

So does music have a place within virtual experience despite its apparent inauthenticity? Evidently yes – this was never really in doubt. Not only is music subjectively authentic to the experiences expected by audiences of virtual reality, but music also fulfils essential purposes within virtual reality: it conveys narrative information, it conveys emotion, it aids immersion (by incentivising flow, increasing presence, and masking sounds of the real-actual and thereby reducing the size of the virtual externality), and it can deliver interactive information to the participant. We can therefore expect that music could also have the potential to serve useful purposes within a variety of virtual experiences outside of entertainment. However, the role of music within *non-narrative* virtual experience nonetheless poses a challenge; music is not currently expected in those experiences and would be strikingly inauthentic to the user’s expectations of the experience. Use of music within these experiences would need to be carefully considered to ensure that users derive an intended function, and do not experience the music as noise.

Some limited examples of music being used purposefully within real-world situations can be found. For example, American soldiers deployed to Iraq were known to use music to ‘pump themselves up' before combat (Pieslak, 2007). However, information on the overall change in efficacy of soldiers is not available, perhaps because it is difficult to work out a way to test this objectively. In examining future uses of music in virtual experience, it is also difficult to establish what the proposed *benefit* to using music could be: i.e., if soldiers become more willing to kill potential combatants as a result of listening to 'violent' music, it is unclear whether this is a benefit (as they are more likely to react decisively in emerging crises) or a detriment (as they are more likely to kill people who they only *suspect* of being an enemy combatant).

Even if music could be used to, for example, raise the efficacy of remote-operating drone pilots by increasing their perceived immersion within the virtual piloting experience, the risks of adjusting the pilots' judgement would need to be considered. Similarly, we might imagine police officers hearing soothing music during stressful situations; would this increase their desire to de-escalate a situation peacefully, or make them less aware of the potential escalation risk? This is an area for much further examination, especially as equipment becomes more technologically capable. This examination may require novel approaches as some evidence has shown that the effects of music vary significantly (G. Cassidy & Macdonald, 2009; Tan et al., 2010; Yamada et al., 2001), and that the user perspective of the effects of music on their perception may not always match the observed (and experimentally tested) effects (Fritsch, 2018).[69]

Music throughout this chapter has been used to explore approaches to non-diegetic sound throughout virtual experience. It should be understood that the differentiation between music and sound is not always clear, especially within virtual reality experiences where music may be spatialised: e.g., as explained by Stephen Cox, "During the process of [musical composition], it felt like the sound design and the music were kind of one and the same" (Balofsky, 2017). Many of the discussions surrounding music in this chapter could therefore also be applied more broadly to sound within virtual experiences.

[69] This is explored in greater detail by Fritsch & Summers (2021).

We have now explored different expressions of sound and music within virtual experience. Throughout Chapter 2, the framing of sound within virtual experience was investigated. Chapter 3 investigated the specific domain of authenticity within audio. Chapter 4 has investigated music (and more broadly, non-diegetic sound) in virtual experience. This represented the second axis of the framework within this thesis, which examines experiential perspectives of sound and music within virtual experience. The understanding of sound developed through Chapters 2–4 can be combined with the theoretical groundwork laid out throughout Chapter 1 to analyse expressions of sound within virtual experience. The second half of this thesis will demonstrate how the understanding developed throughout the first half of the thesis can be used to interrogate, design, and develop virtual experiences.

Chapter 5: A Framework for Sound and Music in Virtual Experience: Overview and Applications

5.0.0: A brief summary

Throughout the thesis so far, a theoretical framework for virtual experiences has been constructed. This framework consists of two axes. The first of these is virtuality which is constructed of the three elements of virtual **externality**, virtual **hardware**, and virtual **software**. The virtual **externality** is the real-actual space in which a virtual experience takes place, which is defined by the sensory horizon of the participant. These boundaries can be multi-dimensional as a result of the affordances of the isolating technologies used: for example, a user may wear a head-mounted device for virtual reality (HMDVR) which visually isolates them almost completely from the real-actual but use open-ear speakers which do not isolate them almost completely from the real-actual. This would create differing sensory horizons and, therefore, different limits on the virtual **externality** dependent on the sensorial mode of engagement. Virtual **hardware** are the objects used by the participants/users of the virtual experience to interact with that virtual experience. These broadly fall into three categories: '**controllers**' (which receive participant input and convey it to the virtual world), '**broadcasters**' (which output information to the user from the virtual world), and '**platforms**' (which mediate between **controllers** and **broadcasters**). Whilst '**hardware**' often implies a specifically technological object, in this case it is intended to be inclusive of non-technological objects: e.g., it was suggested in section 1.3.1 that cast members in Disney theme parks may constitute a form of biological **hardware** for the Disney virtual experience. Finally, virtual **software** are the behavioural rules which determine how the virtual experience will react to input, when it should provide output, and what outputs will be provided. Again, the term '**software**' often has a technological implication (as in 'computer programming') but is intended to include all behavioural rules that might affect a virtual experience.

Whilst this model of virtuality can be used to interrogate or design any kind of virtual experience, the specific focus of the framework as discussed in this thesis also revolves around a second axis: audio.

Considerations of audio have been considered holistically as “the sounds that the user hears” throughout this thesis. Whilst two implicit categorisations can be found throughout this thesis (i.e., sound effects/music as seen in the separation of sound/music in Chapter 2 / Chapter 4; the keynote/signal/soundmark model suggested by Schafer (1977b) discussed in section 2.1.1–2.2.1), an approach which draws more heavily on communicative models of audio has been utilised. That is to say: sound is considered in the terms of what information it provides to the user, and how it provides that information. Auditory information is principally communicated through explicit signification (i.e., there is the sound of birdsong, therefore there are birds). However, additional meaning is gained from what Grimshaw & Garner refer to as “exosonic” components of sound: the way in which sound is modified by travelling through the “presence of other materials and objects in the space in which the sound wave propagates” (Grimshaw & Garner, 2015, pp.168). Users also have an implicit understanding of exosonic components of sound (i.e., there is a long reverb, therefore the room is large) which can convey information. When constructing/designing auditory environments for use in a virtual experience, we must understand what sounds will communicate the desired information to users and provide them with an auditory environment that allows them to perceive the soundscape that we intend. Sounds which prevent users from having the intended auditory experience (i.e., disruptive sounds) can be considered ‘noise’, which principally emerge as a result of intentional sounds which are misunderstood by the user or as a result of external sounds crossing the boundary of the virtual **externality** (i.e., sounds from the real-actual entering into the virtual experience).

As discussed throughout Chapter 2, our understanding of sound is often constructed through subconscious/nonconscious interpretation of sound signals surrounding us. One effect of this, as discussed in Chapter 3, is that sonic authenticity can therefore become a key part of perceived immersion in virtual experience. Perceived immersion relies, in part, on the plausibility illusion: “the illusion that what is happening is real even though you know that it is not real” (Slater, 2009, p. 3556). Plausibility emerges not from one aspect of virtual experience, but from holistic consideration of concerns emerging from all aspects of virtual experience. Disruptive sound entering from the virtual

externality has already been discussed but was explored in more detail in section 3.0.3: techniques such as LANC (Shen et al., 2018), reduced reality (Walther-Hansen & Grimshaw-Aagaard, 2020), and auditory screening (e.g., from white/pink noise) were discussed as possibilities. Inauthenticity emerging from virtual **hardware** was also discussed in 3.0.2. In particular, it was acknowledged that both speakers and microphones can create artefacts of mediation: Westerkamp highlights how microphones "alter listening" as "the microphone's ways of hearing is non-selective … limited by its technical specifications" (Westerkamp, 2002, p.53); Truax similarly argues that loudspeakers cannot achieve perfect reproduction as "every stage of the signal transfer process inevitably adds noise or distortion to the signal" (Truax, 2001, p.10). These artefacts of mediation can create challenges to authenticity such as the poor response frequency in low-frequency sounds in the small speakers frequently mounted on HMDVRs. Authenticity in sound content was also discussed (e.g., did dinosaurs *really* roar?) as was authenticity in the exosonic components of sound, such as reverberation (see section 3.0.2 for issues in modelling reverberation in virtual environments) and spatialisation (see section 3.0.4 for a discussion of binaural audio).

Further complexifying these issues is that considerations of 'plausibility' are subjectivised by participants in virtual experience. Users have learned expectations of what an auditory environment will sound like and how this auditory environment will be expressed. In Chapter 2, it was highlighted that differing cultural approaches to cinema sound can alter a viewer's sense of engagement with a film: Binns explores how Japanese cinema draws upon musical association and musical rhetoric specific to Japanese cinema Binns, 2017); Chatterji argues that Indian cinema has a specific and unique sound design reflective of Indian ethos and culture (Chatterji, 1999). In some cases, audiences who have been under-exposed to these specific cultural 'languages', such as Western audiences, may find the sound and music of these other traditions disruptive to their engagement (and thereby disruptive to their narratival immersion) as they lack the learned knowledge to correctly interpret the communicative intent of the sounds. Chapter 4 also highlighted how the user's cultural background can alter their sense of engagement with media: Lind reports differences in the authentication of music for media based on differences between Lind's personal engagement with the game *Cuphead* and their students' engagement with the same game (Lind, 2022, p. 2; see section 4.2.1).

Chapter 4 tackled the more difficult question of musical authenticity: non-diegetic music is patently inauthentic to virtual experience and creates what Inger Ekman refers to as "the difficult question why, and how, something as obviously constructed as the [musical] score, does not completely destroy the sense of realism within the [experience]" (Ekman, 2008, p. 5). It was suggested in section 4.2.1 that authenticity must therefore exist in two paradigms: authenticity *of* virtual experience (i.e., is the sound faithful to the object, approach, manner, and style of the era which a virtual experience claims to be recreating?) and authenticity *to* (audience expectations of) virtual experience (i.e., is the sound faithful to *the audience expectations* of how auditory information will be presented to them within the context of a specific virtual experience?). That is to say, whilst non-diegetic music is objectively inauthentic to experience of any kind (as music does not emerge from nowhere in any real-actual environment), non-diegetic music can nonetheless be subjectively authentic as users may identify music as part of an 'authentic virtual experience.' Users accept that hyper-realistic environments contain background music which relates to the virtual world through cohering to expected behaviours of music established through the cultural language that those users have been exposed to. It was also acknowledged that the overall benefits of sounds which are 'objectively' inauthentic (such as music) may outweigh any potential drawback or potential disruption: for example, music can convey narrative information through semiosis (section 4.1.1), support emotional embodiment or directly convey emotions (section 4.1.2), create sensorial immersion (section 4.1.3), and deliver interactive information to the user (section 4.1.4).

5.1.0: An approach to analysis

Throughout the thesis so far, various examples have been considered along one of the two axes of virtuality and audio. For example, Chapter 1 examined *Half-Life Alyx*, Disney theme parks, and *Pokémon Go* to work toward constructing a framework for virtuality. Similarly, to work toward an understanding of sound and music in virtual experience, Chapter 2 examined the audio in *Disneyland*

Adventures, Chapter 3 discussed historically informed performance, and Chapter 4 looked at the use of music in *Assassin's Creed Odyssey*. Even where examples have been discussed in multiple chapters – such as *Half-Life Alyx*, which was discussed in both Chapter 1 and Chapter 4 – they have not been considered along both axes simultaneously. This is because these case studies were specifically used to illustrate and develop the theoretical framework and model for virtual experience throughout Chapters 1–4.

The framework is intended to be used to aid the design, implementation, and analysis of sound and music in virtual experience. [70] In order to utilise the framework, an interrogative approach can be developed: through targeted questions, a given experience can be broken down into the most important constituent elements that are under analysis, or which are being considered as part of a design, or which are being implemented. To model how the framework can be applied in practice, a series of interrogative questions that can be used to investigate these elements will be explored in section 5.1.1–5.1.3 and – following the example of Royal S Brown (1994), Tim Summers (2016), and Elsie Walker (2015).[71] A simple list of these interrogative questions that might be used to analyse virtual experience through the framework are set out in Appendix A. To some extent, it should also be recognised that these elements and processes are conjoined: e.g., analysis examines the implementation of specific sound design(s), and changes to **hardware** might have effects on the **software** or **externality** of an experience. These multi-dimensional considerations, and their use a part of the framework, are discussed in section 5.1.4.

5.1.1: Applying the Framework: Audio in the Virtual Externality

Audio in (or, perhaps, *from*) the virtual **externality** of an experience must be considered in three specific ways: what sounds will be present in the virtual **externality**; will these sounds transgress into

[70] Many other aspects of virtual experience exist alongside the auditory dimension explored throughout this thesis: e.g., visuality. However, these other dimensions are beyond the bounds of this thesis' consideration.
[71] Elsie Walker's interrogative questions are provided at the end of each chapter of her book, rather than collated as an appendix as in Brown (1994) and Summers (2016). See Walker (2015), pp.19–20; p.100; p.177; etc.

the virtual experience; and how might sounds from the virtual **externality** be controlled? Key questions surrounding these considerations at the intersection of audio and virtual **externality**.

When considering what audio is present in the virtual **externality** (i.e., the real-actual space which the virtual experience takes place within), the most obvious question is what that location is. Very different sounds will be present in a private home as compared to a business office or a public mall, and virtual experiences can take place in any of these locations. One consideration that might have a significant effect on these sounds is whether someone engaging with the experience is within an individual or collective environment. For example, as discussed in section 1.3.1, Disney theme parks represent a collective environment. The presence of other users can lead to additional sound occurring within a user's experience (e.g., a familial argument, as discussed throughout Chapter 2). Conversely, an individual in an individual environment such as a home office is far less likely to have noise from other participants (or other people within the space, even those not contributing to or engaging with the experience) within their virtual **externality**.

The second consideration, therefore, is the extent to which audio in the virtual **externality** infringes on and disrupts the overall virtual experience. Following on the discussion in section 2.2.1–2.2.3, sounds which do infringe on the virtual experience might be considered as noise: "a signal that interferes with the reception of a message by a receiver" (Attali, 1985, p.27). Two predominant paradigms of noise were identified: sounds placed intentionally within an environment by the sound designer which are misunderstood by the user (which is discussed further in section 5.1.3); sounds which were not intended to be part of the virtual experience by the designer but which are transmitted within the user experience, nonetheless. It was suggested that this second paradigm of noise might be considered as a form of 'audial hallucination' – a simple example is a user hearing the sound of a doorbell and being unclear whether the sound emerged from within the **software**/**hardware** of the virtual experience or from the virtual **externality**. As the holistic user experience is a synthesis of these elements, we can imagine that these audial transgressions could challenge narrative immersion (as sound will not coalesce with narrative events within the virtual experience) or increase the difficulty of challenge-based immersion (as users may become confused by these ancillary sounds).

However, it is also important to recognise that not all sounds in the virtual **externality** will disrupt narrative immersion or make challenge-based immersion more difficult. For example, a virtual experience which is set in a busy city is unlikely to be disrupted should the sound of car horns enter the virtual **externality**. Conversely, a virtual experience set in a lush fantasy forest *is* likely to risk some form of disruption to narrative immersion should that same sound of car horns enter the virtual **externality**. Design considerations in this area must also account for the potential *need* for some external sounds to be heard within a virtual experience for the purpose of user safety: e.g., fire alarms.

The presence of audio within the virtual **externality** is, to some degree, inevitable as sensorial immersion – as outlined above – cannot always be the intended desire. Even where it is a goal, complete sensorial immersion is unlikely to occur: It is very hard – if not impossible – to audially isolate a space completely as sound consists partly of physical vibrations. But, nonetheless, considering ways in which audio in the virtual **externality** can be controlled is an important part of design processes. For example, designers can consider sound in the virtual **externality** when deciding what kind of audio **hardware** should be used, or whether noise absorption materials should be installed at a venue, or whether noise-cancellation should be utilised as part of the experience. Whilst many of these framings of the virtual **externality** emerge as a result of choices made in the **hardware** and **software**, they are nonetheless decisions made as a result of factors rooted in the virtual **externality** of the experience.

5.1.2: Applying the Framework: Considering Audio in the Virtual Hardware

Hardware, as discussed in section 1.2.2 and reiterated above, is broadly separated into three types: (1) **controllers** (i.e., devices which take user input and convey it to the **software** controlling the virtual experience); (2) **broadcasters** (i.e., devices which receive commands from the **software** controlling the virtual experience and convey it to the user); and, (3) **platforms** (i..e, devices which mediate between the first two types of **hardware**).

Whilst we rarely think of these devices as *controlling* virtual experience (although, as discussed in Tatlow (2020), voice-input has been utilised as a control interface in several games) outside of this context, we should nonetheless consider the ‘controller’ **hardware** for audio in virtual experience to principally be microphones (or microphone arrays). At the most obvious level, the primary considerations are how many microphones are in use, and where those microphones are located. For example, are the microphones mounted on the user’s person (e.g., headset microphone), or on a device (e.g., a laptop microphone), or completely external to the user (e.g., a hanging microphone)? As shall be discussed later in this chapter, these considerations have important implications for the capability to screen other sounds from the **externality** (i.e., non-user sound) from also being received by, and potentially affecting, the virtual experience.

However, alongside these obvious concerns, other properties of microphones may also impact the virtual experience. For example: what is the response-frequency and polar patterns of these microphone arrays? We can imagine that a directional microphone may become difficult for a user to control in some circumstances, especially if they cannot see it because of visual immersion in the virtual experience. Similarly, consideration should be given the capabilities both of other **hardware** such as virtual **platforms** and the underlying **software**, as it may not be possible to use many microphones within the same experience. In some cases, novel auditory control **hardware** may also be used. For example, digital audio devices such as soundboards or musical instruments can be plugged directly into a computer. For example, the game *Rocksmith* (2011) allows players to connect and use an electric guitar as an audio controller. In these circumstances, it is important to consider how this additional control **hardware** will interface with other aspects of the virtual experience.

Broadcaster **hardware** can conventionally be recognised as the speakers used within the experience. As with controller **hardware**, several immediate obvious questions come to mind: what speakers are used? Where are these speakers located? If the user is wearing headphones, are they open-ear, or closed-ear? Each of these questions has implications for the sensorial immersion of the user. As discussed in section 3.0.2, another question running parallel to this includes the response-frequency profile of the speakers being used: many speakers have poor fidelity when playing low-frequency sounds, which can disrupt the perceived authenticity of the experience by affecting both narrative and sensorial immersion. As with considerations of controller **hardware**, novel auditory **hardware** can also be considered. For example, rumble pads can be used to help synthesise haptic elements of low-frequency sounds (see section 5.3.1–5.3.2) and should therefore be considered part of the audio broadcaster **hardware**. Similarly, as explored in section 6.2.1, brain-computer implants (BCIs) utilise electrodes to directly generate the neurophysiological activity which is perceived as sound. This could also be considered a form of broadcaster **hardware** for audio, albeit one operating in a very different manner.

Mediating between these two types of **hardware** are the **platforms**. We can see this as being specifically related to the question of "how/where is the audio being processed?" For example, is there an audio interface included as part of the overall audio **hardware** for the virtual experience? What limitations does the platform have on audio **hardware**? Are there any auditory side-effects emerging from the selected platform (e.g., as a result of how it interprets signals to/instructions from the virtual **software**)? In many cases, considerations of which platform to use will have the greatest effect when considering what the affordances are possible elsewhere within the **hardware** for virtual experience. Similarly, consideration of how systems of **hardware** interact may also be important, although many of these will principally emerge from considerations of the virtual **software**. For example, head-mounted headphones intended to provide binaural audio for a virtual experience perceived through a head-mounted device such as the Valve Index must interact with the motion-tracking and head-tracking devices within the headset so that the binaural audio always accounts for the direction that the user is facing. Finally, one of the most important aspects of this may be to consider the extent to which the user can control the audio **hardware** in use, which is often a direct result of the chosen platform. For example, virtual experiences hosted on a computer are likely to allow users almost complete control over the audio **hardware** in use. Conversely, a virtual experience designed as an installation is likely to allow users almost no control over the audio **hardware** in use.

From design perspectives, considering how users may alter the audio **hardware** – and accounting for the most common options where possible, or encouraging users to engage using specific **hardware** – may improve the overall experience. Similarly, when examining user perspectives on virtual experiences, it could be that considering the designers *intended* audio **hardware** may offer some useful insights.

5.1.3: Applying the Framework: Considering Audio in the Virtual Software

Virtual **software** has been described as the rules of the virtual experience: i.e., the way in which user input is processed, and the user output is generated. In terms of audio, we can see this as understanding how sounds are listened to or by understanding how sound signals are processed by the brain. A wide variety of considerations for audio in virtual experience have been considered throughout this thesis because (as observed by Wierzbicki) “every sound has meaning” (Wierzbicki, 2016, p.153), and these meanings can be derived in many different ways. Several of the most important considerations are presented here, with reference to the section of the thesis which discusses that particular aspect of auditory understanding.

One of the most important properties of sound is the sound content: i.e., what is signified by the sound? For example, bird song represents birds. As described in section 2.2.1, this understanding emerges from a continuous process of learning: we are exposed to audial cues throughout our life. Through exposure and understanding, we learn what each different audial cue conveys. Subconsciously, we are constantly comparing the sensations of now to the sensations of the past. Through conscious and subconscious processes, we use this understanding to interpret the sounds that we hear. When implementing or analysing sound for media, however, we must consider the *design* of this listening experience: Wierzbicki emphasises that realistic sounds are often “not theatrical enough for [creators’] needs” (Wierzbicki, 2016, p.154) which is echoed by Kulezic-Wilson’s analysis of cinema sound as being “amplified by hugely exaggerated sound effects” (Kulezic-Wilson, 2015, p.48). Indeed, much modern media relies on a ‘hyper-real’ exaggeration of reality as it is believed that ‘realistic’ approaches to sound would create “a cacophony” (Tarkovsky, 1989, p.161). To some extent, this approach has created the inherent expectation that all sounds within created virtual experiences carry some form of meaning which must be attended to which is (at least partially) the cause of the ‘audial hallucinations’ created by external sounds entering virtual experience as described in section 5.1.1. And, because of the synthetic nature of some of these created sounds, it is frequently the case that these sounds rely on what Aufderheide terms media literacy: “the ability to access, analyze, evaluate and communicate messages in a variety of ways" (Aufderheide, 1993, p.6). That is to say: we derive a specific understanding through a continuous process of hearing *sound in virtual experience* as well as the continuous process of hearing *sound*.

To a large extent, this is the means through which music can be used to signify certain information in virtual experience. Section 4.2.1 explored how participants in virtual experience may expect music because of the long-standing relationship between music and narrative. As Green explains, 21st-century audiences "have come to accept film music as an integral part of what it means to watch a film. Films that fail to use much music…will often have a problem involving the audience as completely as films that embrace music as a tool" (Green, 2010, pp. 81–82). This relationship extends beyond film and can be applied to all similar virtual experiences. In particular, music is used to communicate information such as "historical, geographical, and atmospheric setting" (Gorbman, 1987, pp. 53–54) and other elements of narrative (see section 4.1.1), to help communicate or effect participant emotion by "taking on the scene's rhythm, tone, and phrasing" (Chion, 1995, p.8) or by "facilitate incorporation into the game body… a dynamic and vital bridge between the bodies of the play and avatar" (Plank, 2021, pp.284) and thereby embodying the user within the experience (see section 4.1.2), or by directly communicating information to the user (see section 4.1.3).

Alongside these understandings of sound generated from considerations of sound *content*, consideration must also be given to exosonic components of sound. Whilst these may seem, at first, to relate more to the virtual **externality**, it is important to note that these exosonic components of sound must also be generated within virtual experience and that these exosonic components must match the user's expectations of what the perceived virtual environment would sound like as incongruous acoustic cues can lead to the user's sense of presence being disrupted (Larsson et al., 2010, pp. 155–156). This is particularly problematic within binaurally spatialised audio and near-field audio, as discussed in section 3.0.4. Further issues of immersion can arise as users may be able to subconsciously distinguish between the exosonic components of sound emerging from the **software** of the virtual experience, and the exosonic components of sound within the virtual **externality**.

Consideration should also be given to the process through which users learn to interpret these sounds. As discussed in section 2.2.1, the process of sound design – which we could envisage in this context as developing the virtual audio **software** for a given virtual experience – is complete when participants can derive the intended communicative meanings from the auditory environment we create, such that each participant can experience the intended soundscape. Two primary considerations can be found here. First, how do participants derive that intended message? This has been examined in some detail in section 2.2.1, in which the concepts of 'earcons' (Blattner et al., 1985) and 'auditory icons' (Gaver, 1986) are investigated. Blattner et al. highlight the importance of interconnectivity in audio cues as "listeners must learn and remember them" (Blattner et al., 1989, p.34) which has also been echoed in more recent work such as Lemaitre et al.'s framework for warning sounds in common usage such as car horns: "when introducing new warning signals, care must be taken that these are not too different from the already-existing ones: the more the new signals are different from the already-existing ones, the more the road users will need time to learn their meaning" (Lemaitre et al., 2009, pp.358–359).

Second, we must consider what might prevent participants from deriving intended messages. As highlighted in 5.1.1, users can perceive sounds within virtual experience which "interfere with the reception of a message" (Attali, 1985, p.27). Corollary to the concerns explored in 5.1.1 surrounding sounds transgressing into the virtual **externality**, a second paradigm of noise is sounds placed intentionally within an auditory environment by the sound design which are misunderstood by the user. One example of these problems was given in 5.0.0 as found in the work of Chatterji (1999) on Indian cinema and Binns (2017) on Japanese cinema: both Chatterji and Binns explore how specific learned knowledge is required to interpret the auditory signals within these cinematic environments. Authenticity can also pose issues in this area, as discussed in section 3.0.1: tensions exist between environments which purport to be authentic (such as Ubisoft's *Assassin's Creed Odyssey* (2018) and users which are used to the *affect* of that sound being captured through other means. As a gruesome (but effective) example, Coulthard describes how, in cinema sound, the squish of a melon has come to have “the correct affective impact” to represent a head being crushed (Coulthard, 2016, p.186). We can imagine that audiences might misperceive the intended affect if an alternative sound were used even if that sound were closer to the authentic sound of a head being crushed. This further highlights the importance of work by Lemaitre et al, Blattner et al., and Gaver as being a model through which we can understand the way in which user’s understanding of sounds can be developed.

These affects, as discussed in section 3.0.2, go beyond sound content also. R. Stevens & Raybould discuss how limitations on **hardware** for virtual experience mean that video games “rely on a system of pre-rendered sounds to evoke the acoustic responses of the difference spaces within the game” (R. Stevens & Raybould, 2015, pp.60–61). This interrelates with virtual **software** in two ways. First, through acknowledging that the underlying computer **software** (which forms part of the virtual **software**) is – as yet – too resource-intensive to rely on estimations of acoustic models to parse and generate exosonic components of sound in real-time. Second, through recognising that these estimations can, nonetheless, present the correct *affect* to the user such that the estimations communicate the intended exosonic auditory components to the user despite their inauthenticity.

5.1.4: Applying the Framework: Multi-dimensional considerations of virtual experience

Whilst the presentation of virtual experience throughout this thesis has separated out elements of virtual experience, it is important to note that the construction of virtual experience is inherently inter-dependent: changes to one element of virtual experience will inevitably affect other elements. This was shown in section 3.1.1–3.1.2, in which it was shown how presentations of authenticity in musical performance require authentication of aspects across the framework of virtual experience set out in Chapter 1: Keightley highlights that rock music is authenticated through “specific historical contexts, audiences, critical discourses, and industrial practices” (Keightley, 2001, p.109) in much the same way that Fabian argues for an authentication of historically informed performance through consideration of “instruments, voice-types, playing and singing techniques, proportions, acoustics, improvisation, tempo, dynamics, agogic, and phrasing … [and] the *Affekt* [of the music]” (Fabian, 2001, pp.163–163). Another implicit example of this drawn upon throughout the thesis and highlighted against through section 5.1.1–5.1.3 is the way in which audio **hardware** such as headphones can affect the virtual experience by changing the user's perceptual horizon of sound within the virtual experience. It is therefore important not to consider elements of the virtual experience in isolation: all three elements of virtual experience have an inherent mutuality and depend on each other to create the overall virtual experience. That is to say: we could not have a virtual experience that consisted *only* of virtual **externality**, or *only* of virtual **hardware** and **software**; all virtual experiences will have virtual **externality**, virtual **hardware**, and virtual **software**, and the overall virtual experience will be formulated through the combination of all three elements and the relationships between them. This means that in order to understand the potential effect of any given change or design decision, or to analyse any specific element of a virtual experience, it is likely necessary to consider the virtual experience holistically – especially when considering things from an experiential perspective.

This framework is perhaps at its most useful when it is used in this way: an experience is broken down into the constituent elements outlined throughout Chapter 1, with specific considerations for audio highlighted through section 5.1.1–5.1.3. A phenomenological approach to analysing this can be taken by "imagining variations on a phenomenon, and considering at what point those variations would cease to be instances of that phenomenon [to] identify its essential characteristics" (Kamp, 2016, p.172). For example, we may consider a virtual experience as 'authentic' and derive the components of the experience which contribute to that authenticity by changing each element explored by the framework above and positing the question "in what ways is this experience (in)authentic?" Similarly, a phenomenological approach to design and implementation can be approach: by imagining a given variation (e.g., changing the **hardware** used in an experience) we can consider the way in which other variations will emerge across the phenomenon as a result of the intratextuality of **hardware**, **software**, and **externality**.

Finally, it should be highlighted that whilst many of the aspects of sound and music in virtual experience have been presented throughout this outline of the framework, there will undoubtedly be other considerations which emerge in the use of this framework over time. These could be the result of changes in the approaches used to construct a virtual experience within aspects of sound, music, or virtuality. Or, alternately, these could be the result of additional axes of virtual experience being considered, such as other sensorial axes (e.g., visual or olfactory dimensions of virtual experience), and the way in which these other dimensions intersect with and influence user perceptions of music and sound within virtual experience. However, the considerations outlined above provide a pragmatic approach to the analysis, design, and implementation of sound and music in virtual experiences.

5.2.0: Framework in practice

To understand how these questions and considerations might be used, we can compare three experiences of shared-**externality** virtual experiences which I engaged with during a trip to Houston, TX (USA) in March 2022.[72] These three experiences were THE INFINITE edutainment virtual experience, the Houston Battlegrounds puzzle room virtual reality experience, and the Zero Latency Houston 'shoot 'em up' virtual reality experience. Whilst all three experiences were a 'virtual reality' experience which aimed to immerse the user across the audio-visual domains of their perceptual horizons, and all three used HMDVR to allow participants to engage with a virtual reality world, these three experiences nonetheless took a different approach across the categories identified within the framework; a full comparison of each aspect of **hardware**, **software**, and **externality** is provided in Table 3. By examining these experiences in more detail, this case study demonstrates how the framework allows us to understand how virtual experiences interrelate - both in terms of highlighting similarities and accentuating differences in virtual experience.

[72] The trip discussed in this section of the thesis was only possible due to generous support from the University of North Texas, Dallas, TX (USA) and the award of a Royal Holloway Doctoral School Travel Award.

	Controllers	Broadcasters	Platforms	Software	Environment
THE INFINITE	Utilisation of smart positioning system to identify user gestures; move within virtual environment to move within virtual world.	Quest Pro HMDVR device with device-mounted speakers (open/non-enclosed) and no microphone	Self-contained VR unit with no user-visible external **hardware**	Bespoke VR experience designed in Unity; no social communication **software**	Large room (approx. 60’ x 60’) shared with up to 100 other participants. Additional VR cinema space with chairs next to area.
Houston Battlegrounds	VR ‘wand’ **controllers** held in each hand; limited movement due to small room, teleportation within virtual world primary movement	HTC Vive 2 HMDVR device with additional closed ear head-mounted speakers and external microphone	Dedicated PC for each user networked through LAN.	Commercially available puzzle room virtual experience with built-in non-spatialised social communication **software**.	Separate small room (approx. 6’ x 6’) for each participant; collective environment created through **software**
Zero Latency Houston	VR ‘gun’ controller held in both hands; experience scaled to size of room. Move within environment to move within virtual world.	HTC Vive 2 HMDVR device with additional closed ear head-mounted speakers, head-mounted microphone and vibrating body suit.	Dedicated PCs for each user, networked through LAN, with external internet connection for potential multi-national use.[73]	Bespoke VR experience designed in Unreal Engine 4 for use at all Zero Latency sites worldwide; spaatialised social communication **software**.	Large room (approx. 80’ by 30’) shared with up to 8 other participants. Bluetooth used for in-space localisation (i.e., locating users).

Table 3: a comparison of three virtual experiences

[73] Whilst technically possible, according to discussions with their technicians, this doesn’t seem to be something which they are currently offering. Alas!

5.2.1: Considering communication in virtual experiences

One purpose of the framework is to provide a structure for isolating inter-experience similarities and differences. These comparisons of historical/current experiences can help inform the design of future virtual experiences by providing insights into different approaches to solving specific problems, or into how different design goals might lead to small (but significant) changes in the overall design of the virtual experience. In the case of the three examples given in section 5.2.0, one area that particularly reveals how this framework can provide a structured examination is user communication. Communication within multiplayer video games has both ludic (i.e., communication for the purpose of ludic success) and social components (i.e., communication to develop social bonds, in-character and/or out-of-character) (Tatlow, 2020, pp. 31–32), and the relationship between the poles of the social/ludic dyad varies dependent on the goals of the specific experience and the current state of the experience. In this case, the three experiences demarcate an example of three very different relationships between ludic and social communication.

Zero Latency placed up to eight players in a single large room. Players then fought their way through a zombie 'shoot 'em up' experience, where waves of enemies would attack them in virtual reality. This was a high-intensity experience which had almost no down-time, as waves of zombies (or other enemies such as robots and helicopters) continuously tried to defeat the players. This means that players need to predominantly use ludic communication to find success; players were required to continuously focus on new ludic challenges to overcome if they wished to find success.

Conversely, Battleground Houston presented a virtual reality escape room. Players could select from a variety of commercially available VR escape rooms and had to solve a variety of mechanical and logical puzzles to make progress through the game. In many cases, players had limited or asymmetric access to information and needed to communicate with each other in order to find the correct solution to the puzzle; players were located in their own individual rooms – both always in the real-actual and, frequently in the virtual-real. The experience invited both social and ludic communication, as puzzles required players to communicate with each other but also provided periods in which players were simply fulfilling the tasks they had agreed to do (i.e., carrying items around the virtual space).

Finally, INFINITE presented an ‘edutainment’ experience, where participants explored the international space station using virtual reality and could watch interviews and video logs from on-board the space station. The experience took place in a single large room shared by up to ninety individuals simultaneously. The experience emphasised the possibility of social communication during the ‘on-boarding’ process by explaining how to safely communicate with other members of your booking group during the experience - i.e., by positioning ear-to-ear rather than face-to-face as to ensure that the headsets did not collide.[74] This reflects the stated goal of the experience: users were encouraged to freely explore the space and discover the areas of the space station that they were most interested in, the experience required no ludic communication between users and encouraged essentially only social communication.

We can use the framework to explore this difference in more depth, and establish how the **software**, **hardware**, and **externality** of the experiences facilitated and effected the goals of those experiences in regard to play communication. This means that the framework will highlight how each part of the virtual experience affects participant communication and allow designers of other experiences to consider approaches and solutions to inter-user communication.

5.2.2: Comparative analysis of three experiences

Many differences across the three experiences can be identified and discussed. This section, therefore, focuses on investigating a number of specific questions to explore how the framework might be used to create structured answers to these questions.

For example, should the voice communication **software** spatialise participant voices? Zero Latency choose to binaurally spatialise player voice. This allowed players to locate each other. It is

[74] Members of the group were presented as golden avatars in the visual display of the experience, allowing you to distinguish them easily from staff (green avatars) and other guests (blue avatars).

likely that the shared **externality** was a factor in these decisions: players are not able to perceive the shared **externality** as they are wearing HMDVRs which exclude the visual **externality** and closed-ear headphones which exclude the audial **externality**. By spatialising the audio, especially audio from other players, players are able to locate each other within the virtual experience and – as the virtual experience is spatially mirrored in the **externality** – therefore within the **externality**. This minimises the risk of participant collisions as player voices were both directional and distance-linked (i.e., voices got louder as you moved toward another player). Whilst these audial systems were supported by visual systems which caused your screen to flash red if you got too close to another player in the **externality**, the audial systems are clearly an important part of managing the shared **externality**.

Conversely, Houston Battlegrounds chose not to spatialise player voices. This is likely for several reasons: players did not have a shared **externality**, and so there was no risk of participant collision which spatialised player voices would aid. The experience was a lower intensity – players moved less, and moved more slowly and carefully, and there were few time-critical narrative events to respond to (whereas even the least-intense moments of Zero Latency had a horde of zombies to respond to). Within the narrative of the experience, players would frequently be in entirely separate locations and unable to see each other at all: spatialised player voices would therefore be diegetically nonrepresentative and perhaps even frustratingly abstract. Whereas voice spatialisation in Zero Latency added a form of sonic authenticity that could be used to more easily navigate the world (as discussed as a possibility in section 3.0.4), we can recognise that voice spatialisation in Houston Battlegrounds would be unnecessary and inauthentic.

We can also observe that other design decisions were influenced by, or influenced, the choice to spatialise player voice. Both Houston Battlegrounds and Zero Latency required players to be able to communicate ludically at all times. This meant that both experiences needed to provide microphones for participants. However, as Houston Battlegrounds used separate **externalities** for each participant, they could use microphones mounted externally (i.e., to the walls/ceiling of the play-space). Conversely, as Zero Latency used a shared **externality**, the microphone had to be mounted on the players to maximise audio fidelity and minimise extraneous noise (such as reverberation or noises from outside the intended virtual experience). This difference also likely reflects the difference in the size of the physical virtual **externality**: Houston Battlegrounds used a small room which meant that a small number of external microphones were able to provide reasonable fidelity of sound as participants were always close to a microphone; Zero Latency used a larger room which meant that a large number of external microphones would have been required to ensure that participants were always close enough to a microphone to ensure a reasonable degree of audio fidelity. The difference between a shared **externality** and an individual **externality** also plays into these factors, as Zero Latency may have needed some way to distinguish not only *where* a sound was being picked up by a microphone, but also *which player* was generating that sound so that this information could be conveyed to other users within the experience. This information can be easier to determine using player-mounted microphones as player locations are tracked, and players have individual microphones (i.e., any sound picked up by the microphone mounted on Player A is probably generated by Player A).

In contrast to both Zero Latency and Battlegrounds Houston, INFINITE required very little ludic communication: there was no essential need for participants to communicate to find success in the virtual experience as the virtual experience could be considered an unstructured documentary, rather than a ludic challenge such as a puzzle game or shooter game. Further the open-ear headphones allowed participants to communicate socially without mediating that communication through the framework of the experience. That is to say: if participants wished to communicate socially then they could approach them to talk *without needing* microphones. It is not clear from available design discussions whether this possibility of social communication was an intentional side effect of **hardware** decisions or whether social communication emerged naturalistically from the affordances of the **hardware** decisions. However, the decision to demarcate users who bought tickets together by making them appear gold to each other within the virtual experience suggests that, at the very least, this social communication was encouraged – as was explicated during the pre-experience briefing by venue staff.

Another example emerges from the use of music within the experiences. Both Zero Latency and Houston Battlegrounds clearly focused on providing an experience focused primarily on entertainment and used the accepted hyperreal conventions of music within virtual reality experiences discussed through Chapter 4. Conversely, the designers of INFINITE emphasise in their interviews that they wanted an 'authentic' experience: "We knew that we wanted this production to be 100 percent made in space, and that was not to find a marketing angle that was cool for the project. It was about honoring authenticity—that everything that you see and hear in this production was captured in that particular environment at that particular time" (Brassard et al., 2022, p.37). Despite this, the experience uses non-diegetic music for several narrative films interspersed throughout the virtual experience. Perhaps the intended feeling of authenticity could have been meant here in the sense of music being authentic to audience expectations of virtual experience. In this specific case, we can perhaps have a sense that the music is inspired by the astronauts – and the filmic characters they represent – and emerges from the void of space as a result of the filmic framing of reality present within the virtual world; the music is "issued from that [filmic] world … along with the other elements of [the] story" (Winters, 2010, p. 232). The use of music in this way (i.e., to portray spatial environments) in space-themed media is explored in greater detail by Miguel Mera who describes a "continuous evolution and negotiation between sound and music within the cinematic space vacuum" (Mera, 2016, p.108) as emerging over the last few decades. We can recognise that the musical soundscape of the INFINITE experience was inspired by the historic use(s) of music and sound to represent space (and objects within space) in narrative media, and that audiences may have implicitly expected to occur within the INFINITE experience.

Alternately, the stated goal of authenticity may have been suspended in the service of pragmatic goals such as helping to isolate the participants from noise created by other participants: with up to ninety other participants in the space, and the use of open-ear headband speakers rather than a closed-ear headphones, extraneous noise could easily be part of participant perceptions. Music could aid immersion, as suggested in section 4.1.3, by screening out some of the sounds that will arise from the virtual **externality**. The inclusion of music within the experience – i.e., as part of the virtual **software** – could be seen as part of the answer to the challenges raised by the shared **externality** and the open-ear headband speakers used throughout the experience. Other justifications can also be suggested. For example, music appeared in both entrance credits and final credits: could the music have acted as a framing device similar to the audio-visual narrative media discussed by Davison (2013; see section 4.1.1). The use of music within INFINITE demonstrates some of the tensions surrounding (in)authenticity of music in virtual experience and the resulting conflict with the pragmatic benefits that music can offer and shows how these tensions are approached in practice: music is used at specific times, for specific purposes, in specific ways.

The comparison of three similar, but different, virtual experiences demonstrates how the framework discussed throughout section 5.1.0–5.1.4 can be used to methodically and holistically examine virtual experiences. Two further case studies are investigated in this chapter to demonstrate how the framework might address specific issues within the design and implementation of sound and music for virtual experience. The next case study will examine the generation of low-frequency sounds in virtual reality experiences. This was discussed as being a particularly problematic area of auditory experience in section 3.0.2 as exosonic components of low-frequency sounds are difficult to simulate when head-mounted speakers (e.g., those found on HMDVRs) are used. The case study demonstrates intra-contextual investigations; how might the framework be used as a focusing device to analyse specific challenges to virtual experience, and identify potential changes we could make to the design of that virtual experience? The framework also provides a structured approach to discussing the holistic impact of those changes to the virtual experience. The final case study examines the management of audio in large-scale shared virtual meetings. This demonstrates the ability of the framework to facilitate inter-contextual comparisons – i.e., a comparison between management of voice in virtual meetings and management of voice in online gaming – in order to identify potential solutions and/or areas of further investigation.

5.3.0: Cinematic sound for HMDVR

This case study examines cinematic virtual reality, which has become an area of rapid development within the virtual reality market. Companies such as Positron have begun to capitalise on the possibilities of virtual reality technology to offer ‘premium cinematic experiences’ marketed under slogans like “4D-film” or “full-immersion cinema”. We can use the framework to consider the auditory experience of sound within cinematic experiences designed for HMDVR viewing with a particular focus on how the auditory experience differs from cinematic experiences designed for viewing in traditional film theatres. We can compare the different aspects of the two experiences using the framework developed throughout the earlier portion of this thesis, and the interrogative questions suggested throughout section 5.1.0–5.1.4.

Differences between HMDVR cinema experiences and theatre cinema experiences typically differ in three key areas. First, HMDVR cinema experiences typically occur within a smaller **externality** than theatre cinema experiences. Second, HMDVR cinema experiences use a bifocal display with head-mounted speakers rather than a large projector screen and surround sound speakers. Third, HMDVR cinema experiences often allow viewers to control the visual field of the experience through 3D-visuals and head-tracking whereas theatre cinema experiences typically offer only a fixed-perspective view. This is, in part, because HMDVR cinema is inherently individualistic (i.e., HMDVRs are mounted on a specific individual’s head) and theatrical cinema is not (i.e., cinema theatres are intended to accommodate large audiences). These differences are set out in Table 4.

Experience Modality	Environment	Hardware	Software
Cinematic experience viewed in a traditional film theatre	Varies from small private cinemas to large public cinemas	Visuals projected onto a large screen with high quality projecter; sound provide through surround sound external speakers	Cinematic film
Cinematic experience viewed on head-mounted display for virtual reality (HMDVR)	Typically, small-scale (e.g., private room, small room)	HMDVR worn with bifocal screen displays; sound provided through head-mounted speakers	Cinematic film; binaural sound solutions

Table 4: A comparison of cinematic experiences viewed in traditional film theatres, and cinematic experiences viewed on HMDVR devices

These differences mean that HMDVR cinema is often utilised to create 'realistic' POV experiences: e.g., the 'Sundowning' VR film discussed by Angela McArthur et al. aims to provide a first-person perspective of Alzheimers to provide an "imaginative, insightful portrayal of the condition" (McArthur, Sandler, et al., 2017, p. 2). The use of head-tracking and 3D-visuals allow viewers to explore the cinematic world in a way which is distinct from theatrical cinema – i.e., they are able to look around the cinematic world. This requires a number of changes to be made to the creative process of recording the film. For example, McArthur et al. describe the need to perform the piece in "one take" to create an authentic first-person perspective free from editorial cuts (p.3) and a need to hide microphones on-set as the 360-degree camera used for capturing HMDVR cinema has no frame outside of which booms or other microphones can be placed (p.3).

The use of a first-person perspective also necessitates, to some extent, the use of binaural audio. For example, McArthur et al. describe "spatial diegetic sound (ideally in three dimensions)" as a "requirement" (McArthur, Sandley et al., 2017, p. 3). However, as discussed in section 3.0.4, the use of binaural audio is also problematic; binaural audio relies on accurate HRTFs, with speech (a predominant form of sound within cinema) being particularly difficult to parse if HRTFs are inaccurate (Cuevas-Rodriguez et al., 2021); near-field (i.e., peri-personal; as would sometimes be the case for a speaker stood at a conversational distance from a listener) spatial frames may also pose a challenge for current-generation HRTFs; interpolation between ambisonic microphones (such as those used in recording sound on-set for HMDVR cinema) may lead to issues as the interpolated audio signals may have inaccuracies (see Tylka, 2019). These issues remain an area of active research (as discussed in 3.0.4) and the framework constructed throughout this thesis can be used to discuss many of the issues emerging from the affordances of the **hardware** used to create and manage these virtual experiences.

Sound within film theatre experiences conventionally uses a multi-channel surround sound speaker set-up. Conversely, sound for HMDVR experiences conventionally use a head-mounted speaker system. As discussed in section 3.0.2, this poses issues for audio fidelity as only small form-factor auditory **broadcast hardware** can be placed within the device; large speakers would be too unwieldy and too heavy to be comfortable attached to a head-mounted device. The small size of the speaker cones mean that frequency response may be poor, especially in bass frequencies where auditory wavelength is very large. These issues are particularly notable in sounds with low frequencies (20–100Hz) and infrasound frequencies (sub-20Hz). Some strategies for mitigating these issues were discussed briefly in section 3.0.2. We will now return to this discussion in greater depth.

One of the most common strategies for mitigating low fidelity in low-frequency sounds is to provide a ‘false fundamental’. The ‘false fundamental’ is a psychoacoustic effect wherein the harmonics of low-frequency sounds are played. The human brain fills in the sub-harmonics and therefore perceives the low-frequency sound as being present, even though the speaker is incapable of producing sounds at those frequencies with any reasonable fidelity. The ’false fundamental’ psychoacoustic strategy can extend the perceived range of the speaker significantly. However, some technical issues remain. For example, early implementations of the MaxxBass psychoacoustic system only demonstrated range enhancement of around an octave (Ben-Tzur & Colloms, 1999). Similarly, many of the audio processing methods used to generate this psychoacoustic effect create auditory artifacts such as distortion, disrupt the timbre of the original due to interference between added harmonics and original audio, or otherwise reduce the quality of the sound (Aarts et al., 2002; Hill & Hawksford, 2010; Mu et al., 2012).

Further, these psychoacoustic approaches to low-frequency sounds often fail to address known differences between the human perception of infrasound and low-frequency sound, and the human perception of other sounds. Whilst the ears remain the primary perception organ, participants in infrasound studies showed “spontaneous reactions that suggested vibrotactile sensations and a feeling of pressure may also occur in upper parts of the throat and in the throat region” which also seemed to be part of the human perception of these low frequency sounds (Moller & Pedersen, 2004). The mechanism of these vibrations is not fully understood, but seems to be predominantly a physical sensation, perhaps caused by resonances between the sound waves and the human body (Berglund et al., 1996). Psychoacoustic approaches will therefore fail to holistically and authentically recreate low-frequency sounds. Higher-frequency sounds may not generate the sensation of vibration which emerges from resonance between the human body and low-frequency sounds. Whilst psychoacoustic bass can help improve overall audio quality in the low-frequency range, the approach may not prevent experiential inauthenticity: there is a fundamental perceptual difference between sound for cinematic experiences in film theatre, and cinematic experiences for HMDVR as a result of the differing **broadcast hardware** used by the two experiences. Further, these fundamental perceptual differences cannot be easily addressed, as providing larger speakers capable of producing the resonant harmonics would not be feasible in a head-mounted device.

5.3.1: Auditory Hardware for HMDVR Cinema

We can reframe this as a simple comparison: the audio **hardware** used in theatre cinema experiences generate a sensation of bodily vibration which contributes to human perception of low-frequency sound; the audio **hardware** used in HMDVR cinema experiences cannot generate this sensation of bodily vibration as it is the result of physical properties of low-frequency sound which cannot be generated by audio **hardware** currently used in HMDVR cinema experiences. We can use the framework set out above in section 5.1.1–5.1.4 to consider how a change in the **broadcast hardware** might change the overall virtual experience.

Two obvious approaches for changing the **hardware** present in the experience can be suggested: additional subwoofers could be provided for HMDVR experiences, allowing for low-frequency sounds to be generated, or participants of cinematic HMDVR experiences could be fitted with ‘rumble pads’ – e.g., wearable vibrating technology which could directly provide the same physical sensation to audiences. By delivering auditory signals through multiple **hardware** modalities (e.g., head-mounted speakers and subwoofers; head-mounted speakers and vibrating pads), audiences can be more sensorially immersed as low-frequency sounds will seem more authentic (i.e., sounds with low-frequency components will adhere to the subconscious expectations of what those sounds should *feel* like even if it’s not authentic to the actual-real sensation of those sounds).

Where virtual reality experiences attempt to provide these simulated vibrations, the currently accepted approach is to provide the tactile sensations (which humans perceive as low frequency auditory signals) through use of a vibrating device. This has been a well-established part of the lexicography of virtual experience for some time. For example, so-called “4D-cinema” frequently integrates vibrating seats such as the TremorFX system (Garun, 2012) to enhance the participant experience by providing additional haptic elements to sound. These systems sometimes respond specifically to the audio track, in an attempt to create greater vibrations when deep sounds are played. Similarly, vibrating **controllers** have been part of console gaming since the release of the Nintendo 64 in 1997 and are thought to increase embodiment – e.g., Mauricio Orozco et al. suggest that the virtual experience of a car racing simulator is enhanced when the player is using a steering wheel with a mechanism to cause vibratory feedback that reflects the car travelling over rough surfaces in the virtual world (Orozco et al., 2012, pp. 223–224). Some modern devices include wearable vibrating devices: e.g., I was provided with a vibrating body pack whilst participating in some of the VR shoot ‘em up experiences discussed in section 5.2.0–5.2.2.

However, the addition of subwoofers to a HMDVR experience may be preferred by some cinematic producers as it is closer to the approaches used in traditional cinema. This may especially be the case for producers or directors who are unfamiliar with the relatively unknown “4D-cinema” and/or the better-known but distinct video gaming contexts that make sure of vibrating **hardware**. The addition of subwoofers to the audio **broadcast hardware** may also provider a higher fidelity auditory experience; subwoofers are able to generate lower frequencies of sound at higher levels of intensity. Conversely, most vibrating devices only have very limited modes of vibration and can therefore only provide approximations of the auditory signals they are interpreting and/or attempting to simulate. That is to say: subwoofers may provide a higher fidelity low-frequency vibrations, which adhere closer to the subconscious expectations users have of what low-frequency sounds should feel like.

Despite the potential for an improved fidelity of experience, however, this approach is not currently used in virtual reality experiences. One immediate concern is that introducing external speakers may pose issues for head-tracked binaural audio. It is unlikely that a 360-degree subwoofer can be installed in most use-cases, and therefore the subwoofer will be at a fixed location, and all sounds played by the subwoofer must emerge from that singular fixed point. It is therefore impossible to spatialise the subwoofer in most use-cases. The importance of spatialisation to sensorial immersion and authenticity has already been discussed in section 3.0.4. It may seem self-evident that having some sounds emerge from a singular point would make it harder for users to navigate the virtual experience. However, to what extent would this be problematic within this specific use-case? It is generally accepted that humans are less accurate at determining the direction which low-frequency sound originate from: e.g., Smith and Price suggest that directional acuity in low-frequency sound principally emerges from the perception of interaural time differences (ITDs) rather than interaural level differences (ILDs), which causes directional acuity in low-frequency sounds to be lower is ITDs are more difficult to perceive than ILDs (Smith & Price, 2014). Whilst further research is required to establish the extent to which low-frequency sounds are perceived non-directionally (or, at least, less-directionally), it seems reasonable to suggest that combining binaurally-spatialised ‘false fundamentals’ with non-spatialised ‘true fundamentals’ from a single subwoofer, or a limited number of subwoofers, could be perceived authentically in both the auditory and physical domains without dramatically increasing challenge-based immersion of interpreting the spatiality of the virtual environment.

5.3.2: Framework as comparative tool

As established in section 5.1.4, any alterations to **hardware** will indubitably alter the overall user experience of the virtual experience. We can use the framework to assess the impact of these changes on the overall virtual experience; what is the impact of the virtual **externality** and **software** when we change the **hardware** used for a virtual experience? As it has been less explored in commercial applications, we will focus predominantly on the changes to the virtual experience which would result from the combination of external subwoofers and head-mounted speakers.

Changes to the virtual **externality** must be grounded in an acknowledgement of the virtual **externality** which exists in current virtual reality experiences. Collisions between participants in experiences with a shared **externality** were explored in section 1.3.1: participants are able to disrupt (both deliberately and accidentally) each other’s virtual experiences. One way that auditory conflicts are avoided in virtual reality cinema at present is to localise the sound: each participant has their own set of head-mounted speakers which can only be heard by them. They are isolated from the auditory experiences of others (Figure 13).

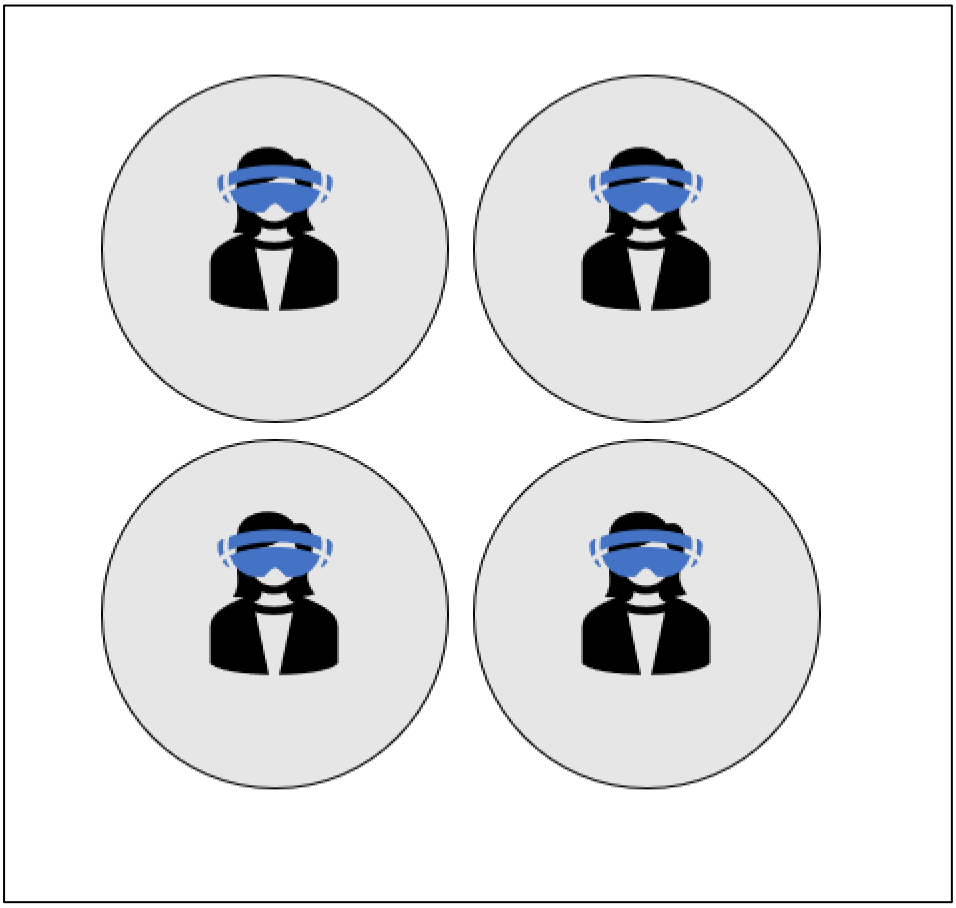

Figure 13: Individuals using only head-mounted speakers can be isolated audially

This allows for large-scale virtual reality experiences (including cinematic HMDVR experiences and experiences which include cinematic elements) to be hosted asynchronously; many present VR cinemas function on an ad-hoc basis wherein audience members within the space view a different selection of video content simultaneously. This would not be possible if the virtual **externality** was expanded to include external subwoofers. The auditory domain of the virtual **externality** would encompass everyone who is able to hear the subwoofer (Figure 14). It is self-evident that using subwoofers would necessitate a change in the framing of cinematic HMDVR experiences: audiences would need to engage synchronously. We could see this as a change in the **software** of the virtual experience; the rules governing user behaviour would need to be changed to facilitate the change in **hardware**. This may be unpopular with producers creating and/or venues hosting cinematic HMDVR experiences as this would mean the (sometimes lengthy) on-boarding process would need to be conducted simultaneously rather than concurrently and dramatically expand the time requirements of the virtual experience (i.e., reducing the number of users/hour that such an experience could be sold to).

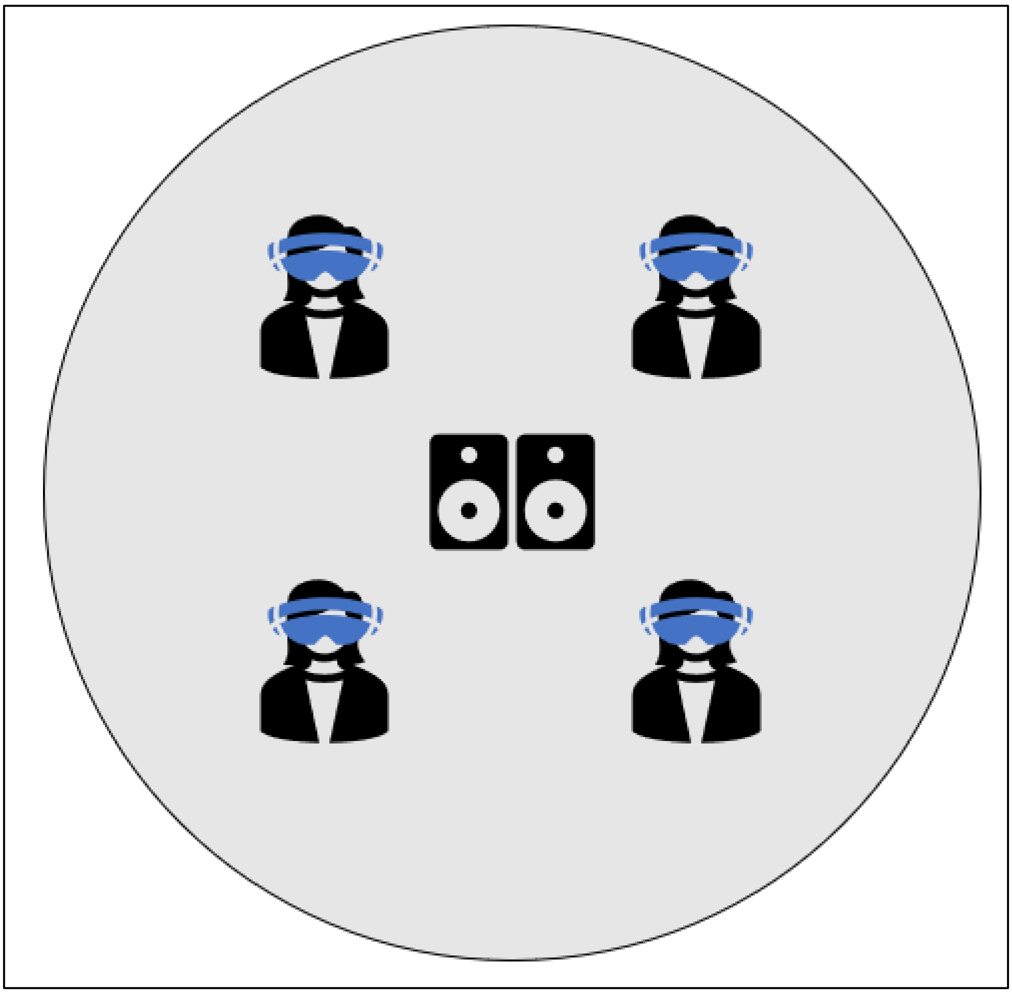

Figure 14: Participants utilising external speakers have a shared virtual environment

Even where the HMDVR experience is intended to be explored in a private basis (such as in a home), this change to the virtual **externality** may prove problematic. Consumers may also have concerns about the privacy of this approach if intended to be used as part of a cinematic HMDVR solution in their own house. Headphones provide isolation for sounds in both direction; users are unable to hear external sounds, and external individuals (e.g., housemates) are unable to hear the experiential sounds. If some sounds are broadcast through external subwoofers (thereby creating a non-isolated **externality**), this **hardware** configuration significantly alters the experiential paradigm: external parties may be able to hear the sounds (Figure 15). This may be an issue for some users or during specific cinematic experiences which are particularly personal or explicit; the required change in user behaviour (i.e., **software**) may make this change in **hardware** unpalatable.

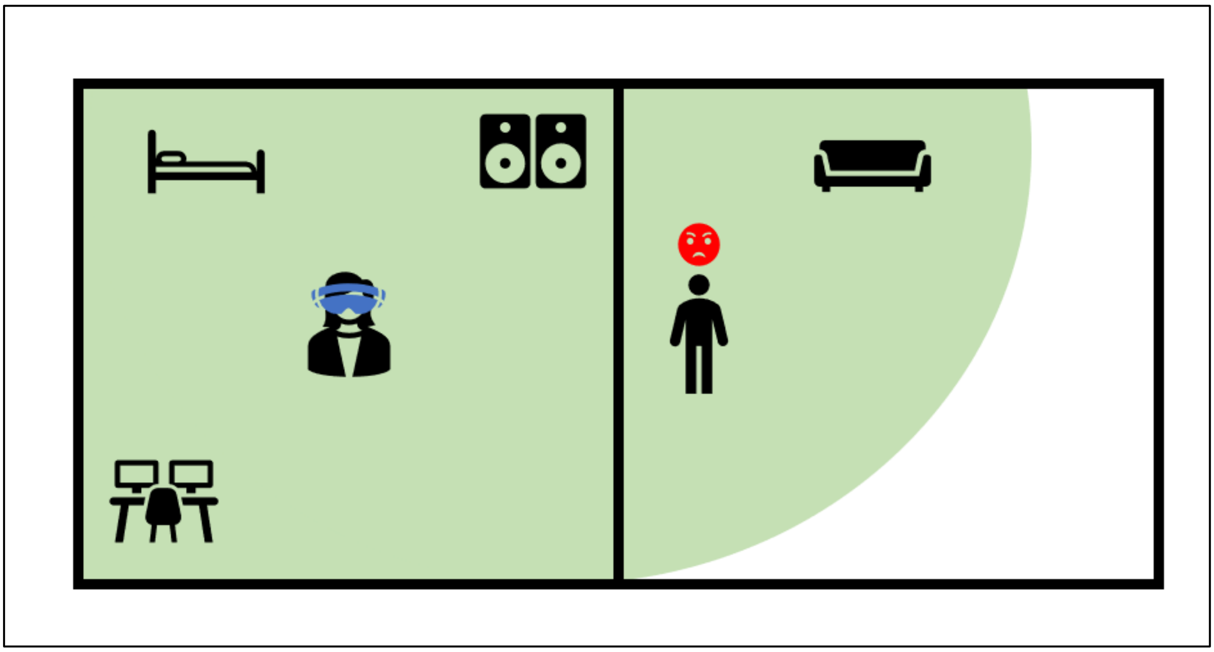

Figure 15: The use of external speakers can also cause issues in home use, as other people in the surrounding area may be able to hear sounds from within the virtual experience.

Difficulties in sourcing and using virtual **software** should also be acknowledged. Many current commercially available **software** packages do not offer configurations that would allow this approach to be used. For example, whilst both Windows 10 and Windows 11 have had – at varying points over the last few years – workarounds which allow some form of stereo or multi-channel configuration allowing applications to simultaneously send sound to multiple audio **broadcast hardware** devices, these are known to be unstable and prone to issues.[75] The issue is further complicated by the fact that the proposed change de-spatialises only low-frequency sounds: to maintain the binaural spatialisation essential to authentic experience (see 3.0.4) then the sound output

[75] Whilst not cited, as it offered little to an academic consideration of this issue, readers are nonetheless invited to pursue the Windows Help Forum to review various discussions surrounding the stereo/multi-channel audio configurations alternately implemented and removed from Windows throughout the last decade. It is somewhat entertaining; my personal favourite is a user who explains that one of the side-effects of altering the registry (which is, in itself, an ill-advised action for anyone not very familiar with troubleshooting and maintaining a Windows installation) to make use of the workaround suggested by another user was that they no longer had any audio at all and had to reinstall Windows to fix the issue. This is unlikely to be a palatable risk for the average consumer.

from any experience designed to use external subwoofers would need to be processed to separate binaurally spatialised (and dynamically rendered based on head tracking) audio channels for head-mounted speakers, and low-frequency audio channels for the external subwoofers. This would require both the **platform** and the **software** to be able to handle multi-channel audio, similar to the functionality of the Dolby Atmos surround sound systems used in many home cinemas but with additional processing to separate the processing of head-tracked binaurally spatialised audio from the processing of non-tracked non-spatialised audio.

The use of the framework allows for a targeted discussion of a specific issue emerging within the design of a virtual experience, with a specific focus on how that issue impacts the experience of the user. We can then use the framework to methodically consider and outline potential solutions (and the consequences of those solutions). The framework can therefore be useful as part of conversations surrounding design and implementation of virtual experiences. Another use of the framework is to identify potential solutions by identifying other virtual experiences which have a similar profile within the framework, and which have addressed similar issues in their development and design processes. One such example can be found in the consideration of audio in large-scale shared virtual meetings.

5.4.0: Managing audio in large-scale shared virtual experiences

Throughout the COVID-19 pandemic, the adoption of remote working accelerated significantly with many companies shifting most or all employees to a remote working basis. Remote working has become a central part of the modern work environment. One aspect of remote working has, in particular, become commonplace; the use of virtual meting **software** such as Microsoft Teams, Zoom, Skype, and Discord has increased dramatically, with a particular focus on facilitating large-scale collaborative meetings without the need for participants to travel to a central location. Meetings located in this virtual **software** can save both time and money for employees and employers alike, allowing for a more efficient workflow to be developed.

We can explore the concept of a virtual meeting through the use of the virtual framework, in order to fully understand the construct of the virtual meeting. To create the virtual experience, participants must install the necessary virtual **software** (e.g., Microsoft Teams), activate the necessary virtual **hardware** (e.g., a platform to install the **software** onto (such as a Windows PC), an audio-input device and an audio-output device, with many companies also requesting that employees have a video input device) and find an appropriate virtual **externality** free of extraneous noise (e.g., a home office) in which to engage with the experience.

One complex area of this model is in the consideration of the virtual **externality**. In section 1.2.1, the boundaries of the virtual **externality** were conceptualised as the limitations of our perceptions, and that these boundaries can be disrupted by extraneous noise. Further, these boundaries are necessarily mutable, as small changes to our physical surroundings can make significant differences to the possibility of extraneous noise entering our perceptual horizons and therefore dramatically extend the physical limits of the virtual **externality**. To visualise this, imagine a window; if a real-actual window is opened, then sounds from outside the window can more easily be perceived by the user/participant during the virtual experience. The virtual **externality** of the user has been expanded; we must consider potential auditory sensations from a larger physical space when attempting to create sensorially immersive experiences.

Virtual meeting **software** such as Microsoft Teams allows us to open virtual windows into the **externality** of another person. Whilst we have an individual virtual **externality** (i.e., the real-actual world around us), we also have a collective virtual environment which is the shared result of the combination of every individual's virtual **externality**. In simple terms, extraneous sounds can be transmitted from another person's **externality** into our own (Figure 16), and noisy collective environments created through the combination of many individual virtual **externalities** (Figure 17).

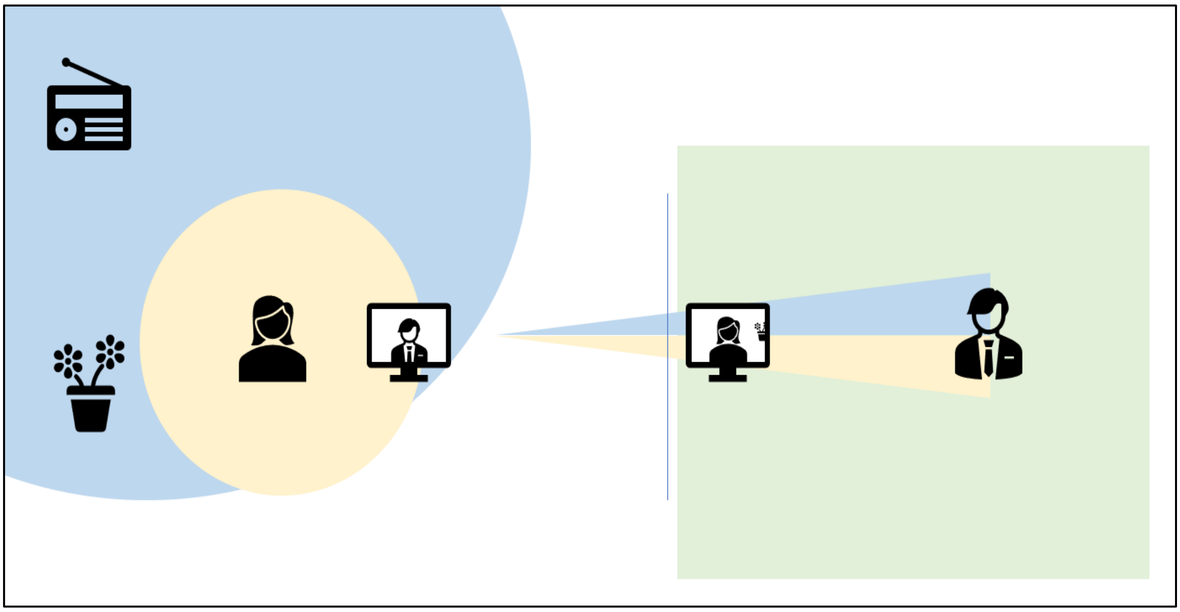

Figure 16: Users can transmit extraneous noise from within their individual environment to other participants through remote meeting software.

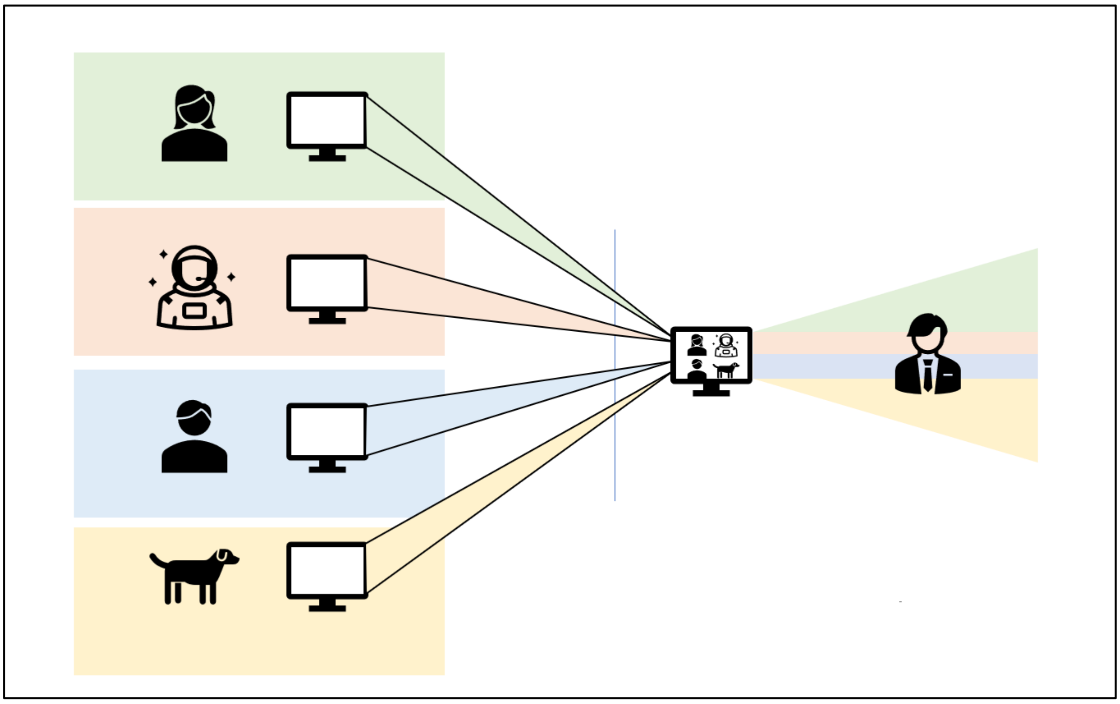

Figure 17: Collective virtual environments in remote meetings are generated by combining the individual virtual environments of all participants within the experience.

Most relevant to this thesis are considerations of how audio functions within these virtual meetings. A core (and intended) functionality of the **software** is capturing and broadcasting sound: e.g., user voice. However, as the figures above show, this can create an issue for participants if there is too much extraneous sound within the virtual **externality**: if any individual virtual **externality** being broadcast is ‘noisy’ then the collective shared virtual environment for all participants will also become ‘noisy.’ This will be disruptive. As the number of participants within a meeting grows, then the risk of this happening increases. Further, only one participant within the call must broadcast a noisy **externality** to disrupt the virtual experience for all other participants. Preventing this from happening is, therefore, an area of concern for not only the companies designing remote meeting **software**, but also for companies *using* remote meeting **software**. Holistically understanding how disruption can occur within remote meeting **software** may also be useful for academics studying the sociological context of the digital working environment, and other similar virtual experiences.

Considering challenges to audio quality resulting from each aspect of the framework for virtuality allows for the rigorous construction of rules for, and approaches to, remote meeting environments which will avoid or minimise any potential audio disruption and, therefore, maximise the efficiency of remote meeting. Further, a systemic analysis of the factors at play within remote meeting **software** allows for both better understanding of how social factors may differ in online environments and for a more specific approach to tackling issues within the design and development of modifications to, or the creation of, virtual meeting **software**.

5.4.1: Issues and Challenges for Audio in Virtual Meetings

Challenges to audio quality within virtual meetings can be grouped into four distinct categories. First, issues emerging from the **hardware** used by participants in the experience. Second, issues emerging from the configuration of the **software** used by the participant. Third, considerations of the individual virtual **externality** of each participant. Fourth, resultant issues within the collective environment generated by the networked participants.

Audio **hardware** for virtual meetings can be reduced to the combination of an audio input device (i.e., a microphone) and an audio output device (i.e., speakers, headphones) connected to a device such as a laptop, smartphone, tablet, or desktop computer. Three general types of audio **hardware** can be observed: device-mounted **hardware**, head-mounted **hardware**, and external audio **hardware**.

Device-mounted **hardware** is included with most commercial smartphones, tablets, and laptops and some single-unit desktop computers (such as Apple Macs) also include device-mounted microphone(s) and speakers. Device-mounted microphones are frequently embedded within the frame of the device and are therefore, by necessity, small-form factor microphones such as MEMS or ECM. Many functions of the microphone arrays in these devices, such as selecting an appropriate polar pattern and post-processing such as frequency filtering and compression, are hidden from the user and must be controlled by the virtual **software**; e.g., Apple only made stereo recording and polar pattern selection functions accessible to application developers with the release of iOS14 in 2020 (Apple, 2022). Further, as the microphone is located on the device itself, the microphone will be around arms-length away during virtual meetings: as many device-mounted microphones are poorly insulated from vibrations, this can lead to extraneous noises such as (tapping on a keyboard) to be picked up, increasing the noisiness of the collective virtual experience. Similarly, device-mounted speakers must also be relatively small. This can lead to poor audio quality – especially when the device is placed some distance away from the user – and also lead to the possibility of feedback if the device microphone is not programmed to exclude sound played through the device speakers.

Head-mounted audio **hardware** (i.e., a headset with microphone) provides different challenges for users. The audio quality of playback can be improved (as the speakers can be physically bigger, and psychoacoustic effects such as 'false fundamentals' – see 3.0.2 and 5.3.1 – can be used to improve audio quality). Further, the physical closeness of the microphone to the mouth can improve signal-noise ratios[76] and the known location and orientation of head-mounted **hardware** can be used to make intelligent choices for audio design (e.g., such as selecting a mono-directional microphone array, rather than an omni-directional microphone array to improve signal isolation). However, some users find head-mounted audio **hardware** to be uncomfortable, especially for long periods of use. Head-

[76] For each doubling of distance from a point source, the sound pressure level decreases by approximately 6dB. A device-mounted microphone will be approximately 100 – 150cm away, whilst a head-mounted microphone will be approximately 5–20cm away. Therefore, the sound pressure level of the user voice increases by approximately 20–40dB by switching from a device-mounted microphone to a head-mounted microphone. This allows audio filtering processes to more easily remove extraneous noise from audio signals captured by head-mounted microphones compared to device-mounted microphones.

mounted **hardware** can also physically isolate users from external sounds; isolating users from external sound may be considered beneficial (e.g., as it allows them to focus on the meeting), or be potentially problematic (e.g., as it means that they are unaware of problematic noise entering from their individual **externalities** into the collective virtual environment).

External audio **hardware** allows for the highest-quality devices; as discussed in 5.3.1, the inclusion of specific speaker systems such as subwoofers can dramatically improve audio in virtual experiences. However, one of the principal benefits of externally mounted audio-**hardware** in the context of virtual meetings is the ability to locate devices in separate places. A desk-mounted microphone can be located physically close to the user, allowing the easy removal of extraneous noise. Concurrently, external speakers avoid the discomfort associated with head-mounted speakers and the lower quality of device-mounted speakers. A major obstacle to the adoption of external audio **hardware**, however, is the additional cost of external audio **hardware** (e.g., the best-selling external microphone is the Blue Yeti, which currently retails for £119.99.). Further, some users may find it difficult to set-up (and/or troubleshoot) external audio **hardware**.

The use of external **hardware** may also lead to resultant issues in the configuration of the **software** used for the virtual meetings. Installing additional audio devices can cause (relatively) simple drop-down menus to become populated with many different devices: e.g., my work laptop has only two options for audio input devices in the Microsoft Teams drop-down (“Internal Microphone” and “AuxAudioMic”), whilst my home computer has seven audio input devices (“OnBoard SoundCard”, “Virtual Audio Output 1”, “Webcam1”, “Webcam2”, “USB Microphone”, “CaptureCardInput”, “Sound Interface”). This can increase user error in configuring **software** properly and thereby cause disrupt the virtual experience.

Virtual **software** offers a variety of other options which are relevant to virtual meetings. Many **software** offers a variety of options for audio management. One of the most common options offered in most virtual meeting **software** is some form of “push-to-talk” or “push-to-unmute” option. For example, Zoom allows you to press the SPACE key to temporarily unmute yourself during meetings

(Figure 18), Other common options include echo cancellation, noise reduction, and variable audio fidelity. Many **platforms** also include some form of ‘audio optimisation’ which combines varying options such as compression, frequency filtering, and other post-processing to improve the quality of the received and broadcast audio.

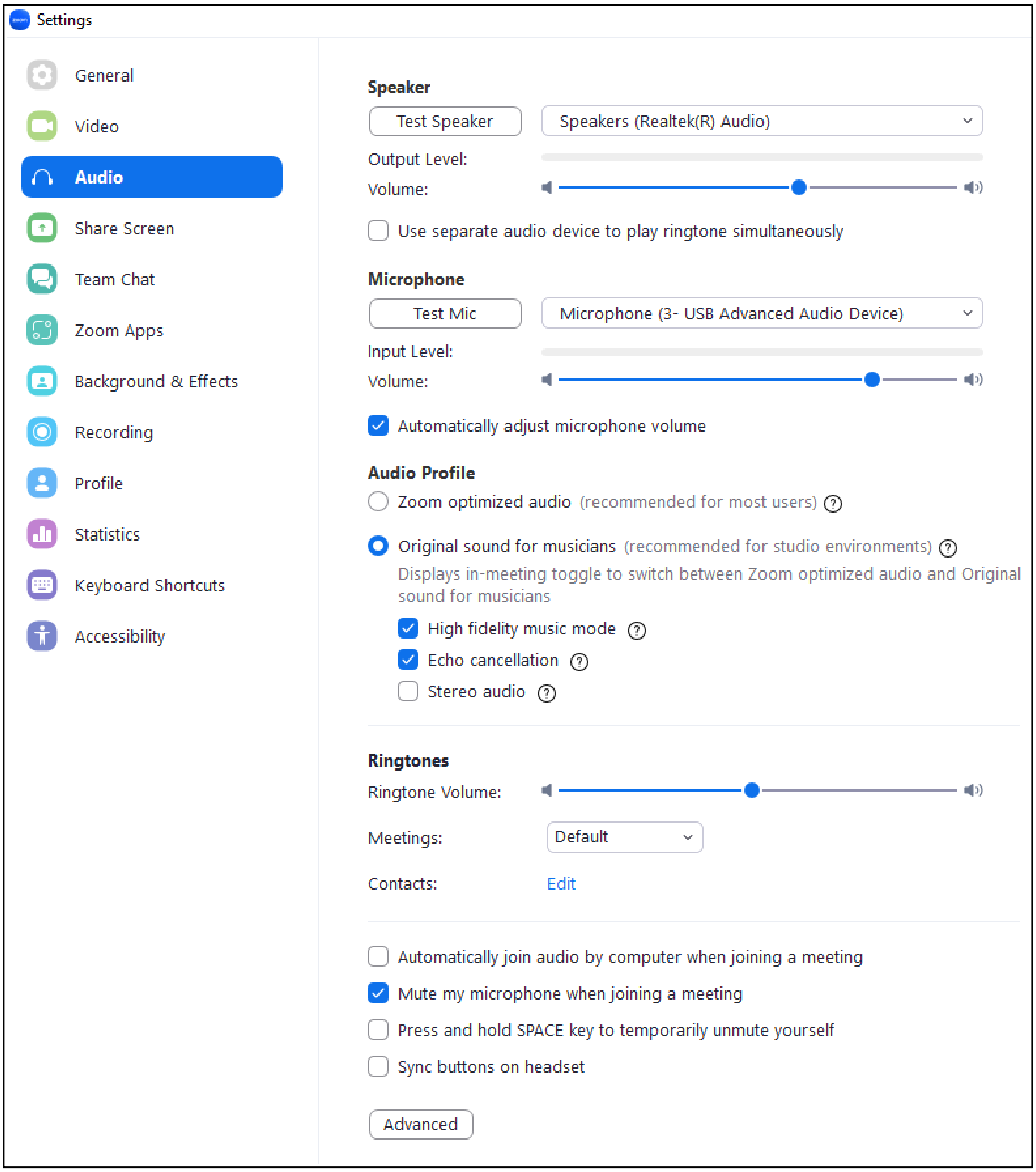

Figure 18: Audio options for Zoom in Version 5.12.2 (9281)

Some configuration options are functionally concealed from users as they are inherent to the platform selected by the user; they are determined by the developers, and users can only select which approach they wish to take by changing which platform they use.[77] One such option is the underlying nature of the network within the selected virtual **software**. Some **software** uses peer-to-peer connections (i.e., each user connects directly to every other user in the call). This has underlying issues of scalability: at the crudest level, we could propose that two users would have only two data-streams (A>B; B>A), three users would increase this to six data-streams (A>B; A>C; B>A; B>C; C>A; C>B), six users would significantly increase this to twenty-one data-streams, and so forth (Figure 19). Each additional network participant adds an ever-increasing number of new parallel connections to be maintained by every current network participant.

[77] In many professional environments, this information is also not immediately evident and is buried in technical specifications aimed at IT staff who decide which software solutions to support at an organisational level. A lack of knowledge about the *consequences* of those choices on the user experience may even be a net benefit to those IT staff: e.g., a network administrator may choose a solution which offers the maximum level of access control and security, at the cost of occasional issues in audio fidelity; users may not value the additional security, as they wish to have the highest level of audio fidelity instead, and therefore wish to use alternate software if they become aware that it may offer higher audio fidelity. The relative obscurity of these options also means that many users are unaware that some of their issues with the virtual meeting experience is the result of issues with the specific software used for the virtual meeting, rather than issues with virtual meetings as a concept. This may contribute to underlying user scepticism toward virtual meetings.

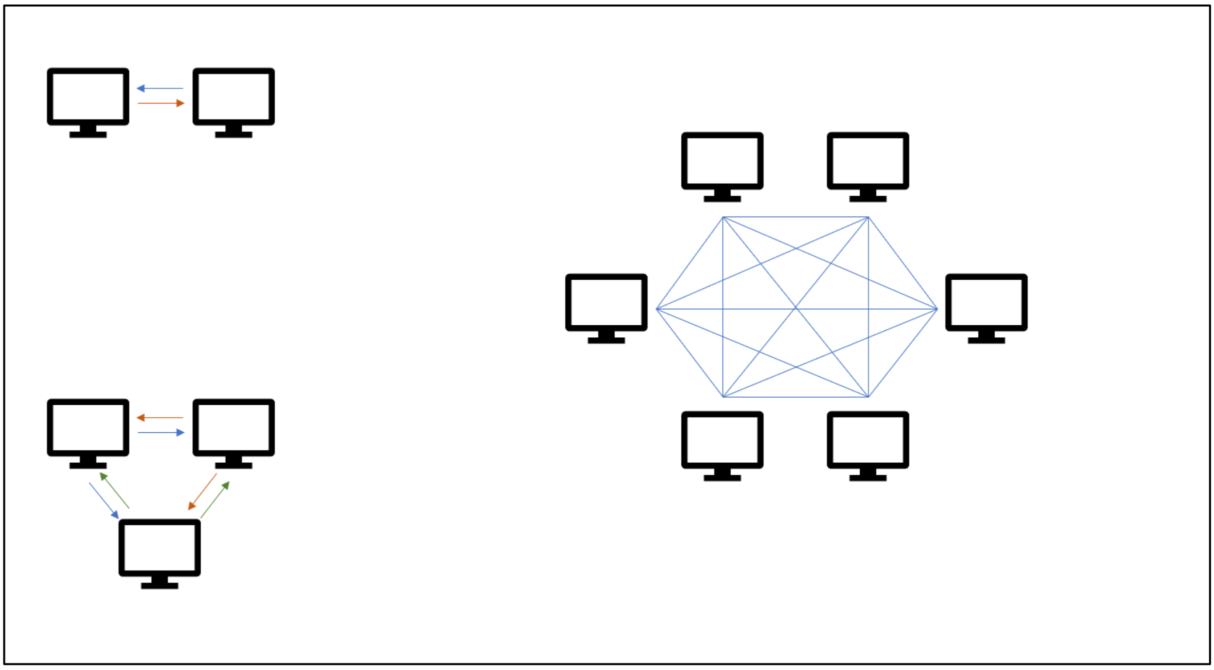

Figure 19: Demonstration of how increasing number of participants in a peer-to-peer network causes a rapid increase in the number of parallel connections required

We can therefore see that peer-to-peer connections require an exponentially increasing amount of bandwidth as the number of network participants grows. This can result in network instability – participants may struggle to remain connected as the required network bandwidth is too high, or the quality of connection may decrease. This will create issues within the audio of the call (e.g., stuttering or ‘robot voices’). Conversely, other **software** uses peer-to-server connections (i.e., users connect to a server, which receives and distributes audio to each user). This reduces the underlying issues of scalability and stability, as each user must only maintain a singular connection to the server. This requires less bandwidth and can allow users to maintain higher quality audio connections. However, this can also mean that audio processing is delegated to an external server, which can have other effects: e.g., users may be unable to selectively mute other participants in the call.

The **externality** of each participant must also be considered: i.e., how does the participant *hear* the virtual experience? As the user wishes to hear (and understand) what is occurring within the virtual meeting, then noise would be anything which disrupts the understanding of the virtual experience. Alongside potentially disruptive external noises intruding on the experience such as those discussed in section 1.2.1,[78] section 2.2.2 identifies two principal types of noise which should be considered: (1) noises which are generated intentionally within the construct of the virtual experience which are misunderstood by the user; and (2) noises which are generated outside the construct of the virtual experience, but which are perceived to be present within the virtual experience. We might imagine, for example, that the audio cue for a user joining a virtual meeting might be similar to the audio cue for an email arriving. This would be potentially confusing and therefore disruptive (i.e., noise).

Finally, the overall shared virtual environment can be seen as the net result of all the above considerations for each individual participant within the call. Extraneous noise which is disruptive to Participant A's **externality** may be transmitted to all participants and become disruptive for all of them if Participant A is using audio **hardware** which has poor isolation, combined with **software** which does not moderate or remove extraneous noise. Further, virtual meeting **software** often handles concurrent audio streams poorly; if many users broadcast audio simultaneously, this may also introduce significant noise. Small audio infidelities (such as hissing or other white noise) present as a result of the mediation between the real-actual and the virtual can combine within the shared environment, resulting in excessive levels of noise.

5.4.2: Inter-contextual analysis

The number of possible solutions to the issues identified above are significant; any given issue could have any number of potential solutions, each of which would need to be considered. However, one possibility offered by the framework developed earlier in this thesis is to draw on solutions found by those participating within or developing similar, but distinct, virtual experiences. This would allow us to examine solutions which are known to work and consider whether they are an appropriate solution for the issues and challenges identified in section 5.4.1, and concurrently avoid solutions which are similar to those known **not** to work to address identical issues in other similar virtual experiences.

[78] Imagine, for example, a neighbour mowing a lawn. The extraneous noise may prevent the user from hearing other participants.

We can use this framework to develop inter-contextual lenses and identify other virtual experiences which share core qualities and may, therefore, allow us to conduct inter-contextual analyses. If we break down the virtual experience in discussed throughout the previous two sections, then we must find another virtual experience which allows us to synchronously connect large groups of individuals over the internet for the purpose of collaborative (i.e., bi-directional) voice communication. For example, we can recognise that a virtual experience such as a Disney theme park is very different to the virtual experience of online/remote meetings; whilst a Disney theme park has a similarly collective environment (as discussed in section 1.3.1), the boundaries are predominantly defined by the physical constructs such as walls. Conversely, the collective environment of virtual meetings is contained by the limitations of the digital **software**. This difference also has a commensurate effect on the **hardware** of the experience: people communicate within virtual meetings using **hardware** such as microphones and headphones. These are not necessary within a Disney theme park.

Conversely, internet-mediated experiences do bear more similarity. We could consider, for example, the possibility of comparing virtual meetings to broadcasts on a platform such as Twitch – a video live-streaming site, with a particular focus on live-streaming video games and eSports. Here, there is a virtual collective environment: the streamer and their audience are engaging in a shared experience. However, unlike the experience of virtual meetings set out above, not all members of the experience are able to voice chat; the predominant mode of communication for Twitch audiences to communicate with the streamer is through text-chat. This means that there is little consideration of managing auditory communication in Twitch streamer communities. Further, we could also question whether this communication is bidirectional; in virtual meetings, the expectation is that all participants communicate simultaneously with each other. The rules – i.e., the **software** – of a Twitch stream virtual experience differ significantly from the '**software**' of virtual meetings. Brett Sherrick et al. suggest that livestreamers should "encourage parasociality with their audiences" (Sherrick et al., 2023, p. 67) which concurs with other explorations of creator-audience relationships on **platforms** such as Twitch and YouTube (Leith, 2021; Rihl & Wegner, 2019). Conversely, relationships in virtual meetings are conventionally (ortho)social;[79] participants respond to each other directly. Similarities exist between Twitch streams and some forms of remote work such as conferences and presentations. There are greater similarities between virtual meetings and Twitch streams than between virtual meetings and Disney theme parks. However, these experiences are still not close enough within the bounds of the framework to engender useful inter-contextual comparisons.

However, we could perhaps consider the virtual experience of playing multiplayer video games. The virtual **externality** for multiplayer gaming will be broadly similar to the virtual **externality** of virtual meetings: both utilise a collective environment which all participants must connect to using computer **software**; participants broadcast sound from their individual virtual **externality** into the shared virtual experience. Virtual gaming typically requires computer applications for the game **software** *in addition to* computer applications for voice communication (as compared to virtual meetings, where participants are often implicitly expected to focus solely on the voice communication application). Some gamers will also have extraneous sounds from additional peripheral **software**, such as music streaming **software** or assistive **software**. However, from the perspective of participant experience, the sound from the multitude of **software** sources is heard simultaneously as part of a unified 'gaming experience'. Similarly, communication between participants in multiplayer gaming experience is often voice-led; in my 2020 article on player voice, one community leader within the game *EVE Online* described voice as a "passport to play" (Tatlow, 2020, p. 26). Players who did not utilise voice communication were not able to participate in the collective experience. Many gaming communities believe that using alternate forms of communication such as text-chat are inefficient as voice communication is both faster and does not require taking hands away from **controllers** to type (Collins, 2014, p. 357). This emphasis on voice

[79] The term "orthosocial" (i.e., a portmanteau of 'orthodox social relationships') in opposition to "parasocial" is used as set out in Rihl & Wegener (2019).

communication for multiplayer gaming echoes the inherent importance of voice communication in virtual meetings.[80]

We can also recognise that voice communication within multiplayer gaming is typically collaborative; relationships developed through voice communication in gaming are typically social rather than parasocial. Collins explicates this, describing gaming as “primarily a social experience and played in groups” (Collins, 2013, p. 14). Whilst Collins is considering a social/anti-social dyad rather than a (ortho)social/parasocial dyad, the latter sense of social is nonetheless echoed in their work. Collins describes how “most players rate this social aspect of gaming as the most important factor for playing … the sense of social presence [is] provided by voices in the game space. When playing, the player is surrounded by conversations, the player is surrounded by conversations … the social mediations that go on between players through their talk with each other” (pp. 78–80). If the conversations were predominantly parasocial, then these social mediations would not go *between* the players as they talk *with each other*; the description of this communication is inherently (ortho)social.

Therefore, we can see multiplayer gaming as containing a voice-led experience similar to that found within virtual meetings. Participants share a collective experience which is, in part, defined by the use of voice-led communication, whose boundaries are established through computer applications. They use similar **hardware** – microphones and headphones – and have similarly social ‘**software**’ underpinning the implicit rules of their communication with each other. We could, therefore, conduct an inter-contextual examination. We can consider how multiplayer video games address auditory issues within their voice communication in order to inform changes to the virtual experience of remote meetings. What auditory issues do people face in voice communication during multiplayer video games, and how do they address those issues?

[80] A virtual meeting without voice communication could just be a series of emails. Some might argue that this is also the case for many meetings *with* voice communication. Unfortunately, that doesn’t seem to be the approach that companies wish to take.

As we should expect from two virtual experiences which share many underlying similarities, multiplayer video games share many of the same auditory issues as virtual meetings. In my prior research, I discussed how voice within video games is often contentious as it can result in “poor-quality or aggravating auditory experiences for players” (Tatlow, 2020, p. 24) similar to those auditory experiences emerging as a result of the issues identified in section 5.2.1: e.g., stuttering, feedback, ‘robot voices’, and background noise. It should be emphasised that not all issues are transferred between different contexts. For example, voice within video game communities has additional dimensions of player identity and character identity: Cheng recalls concerns from players of the game *Second Life* (Linden Lab, 2003) that their adoption of an alternate identity within the fantasy-reality of the virtual experience created within *Second Life* may be challenged through the introduction of their real-actual voices (Cheng, 2014b, pp. 344–345). These issues can be disregarded when considering potential approaches to voice communication in remote meetings, as most companies are formed of individuals expressing their real-actual identities, rather than a fantasy-virtual identity. Further, and perhaps related to the use of alternate identities within gaming spaces and the often-transitory nature of collaboration between gamers,[81] voice communication within collaborative video game play has many specific challenges to moderation (Jiang et al., 2019) emerging from issues such as extensive (and explicit) misogyny (Fox & Tang, 2017; McLean & Griffiths, 2018; Vossen, 2018), racism (A. Brown, 2018; Gray, 2012a; 2012b; Ortiz, 2019), “trash talk” (i.e., online toxicity) (Howiie, 2022; Kou, 2020) and political extremism (United Nations Counter-Terrorism Centre, 2022). In most cases, these issues are less prevalent and are expressed differently in professional environments as there is less pseudo-anonymity and therefore a higher level of accountability for undesirable actions. Therefore, policing these issues requires a different approach in gaming communities than professional environments. However, whilst social dimensions of voice communication for gaming are substantially different to social dimensions of professional

[81] Many modern online multiplayer games use some form of randomised group-creation system to group up players of a similar skill level together. In many cases, this means that players are grouped together with strangers for short collaborative periods and only establish lasting relationships under specific circumstances (e.g,. high-skill players in competitive games will frequently be grouped together to ensure competitive games).

voice communication, gaming communities nonetheless have many of the same technical issues in their voice communications as those found within virtual meetings.

Promisingly for using this inter-contextual comparison for finding solutions to issues which exist within virtual meetings, voice communication within video gaming is often conducted under conditions which are time-sensitive: multiplayer games rarely have a pause function, and even a small delay in responding can lead to complete ludic failure. As a result, participants must communicate efficiently. In practice, small issues with audio fidelity can have significant effects on the success of communication in video games as repetition (to ensure accuracy) is typically too inefficient to lead to ludic success. Conversely, most business use-cases value accuracy of communication over efficiency of communication, and therefore accept that users may need to repeat as-necessary. This means that considering the structuring of voice within the collaborative multiplayer gaming experience may allow us to address issues which have been under-prioritised within the consideration of virtual meeting experiences. By examining solutions to these technical issues used by gaming communities, we can therefore identify potential solutions for issues shared by virtual meeting participants.

5.4.3: Considering Voice Communications in Gaming[82]

So how do video game players address the issues identified in 5.4.1, when and where they arise within collaborative play, and how might approaches used within collaborative video game play inform changes to the remote meeting virtual experience? We can use the structure of the framework established in this thesis to holistically examine voice communication within gaming by first

[82] My knowledge of voice in video games is, in part, based on nearly two decades of engagement and leadership in multiplayer and ‘massively multiplayer’ games. At present, I am ‘guildmaster’ (i.e., supreme leader) of Going Critical, an international English-speaking gaming community which focuses on playing together in competitive multiplayer games. Observations in this section of the thesis draw upon my experiences in recent years leading this community. Games played by the community include *EVE Online*, *Albion Online*, *No Man’s Sky*, *League of Legends*, *Northgard*, *V-Rising*, *World of Warcraft* and others. Member fluctuates between 100 and 300 members, with play groups typically having around 5–30 members. A microphone is a requirement of joining the community, and members are expected to be connected to a voice call whilst playing with other community members.

examining the **hardware**, then the **software**, then the **externalities** of the virtual experience of voice communication.

My earlier research into player voices in *EVE Online* (CCP, 2003) – a space-themed MMORPG – revealed that players elected to take several approaches to minimising the same issues faced in online virtual meetings. Audio **hardware** is quasi-standardised amongst the player-base of the game, with a "now-ubiquitous requirement in almost every major *EVE Online* community that players should have a functional microphone" (Tatlow, 2020, p. 26). We can imagine that the specificity of 'headset' refers not only to the desire that all players can speak up on comms but also a desire that players do so in a way that allows for the highest fidelity of voice communication; players are expected to reduce external noise by using head-mounted or external audio **hardware**. This maximises the signal-noise ratio and reduces the audibility of extraneous noise. As a result, the norms found in *EVE Online* have become common expectations of engaging in voice-led multiplayer environments in online gaming. Similarly, companies could mandate (or, at least, expect) high-quality audio **hardware** for remote meetings which minimise the possibility of extraneous noise entering the collective environment.

Differences between business **software** and gaming **software** for remote collaborative communication should also be considered. Many large gaming communities now use the Discord platform for voice communication during collaborative play. This is for several reasons. First, the Discord platform utilises Krisp noise filtration **software** which uses machine-learning solutions to identify non-human sounds such as "dogs barking, vacuums, or doors slamming" (Discord, 2022). The latest generation of Krisp now uses machine learning to further improve the efficacy of noise cancellation by offering additional features. Krisp updated their **software** in 2022 to include "voice cancellation" features which remove voices which are distinct from the owner's voice (e.g., conversations between housemates or co-workers in the background of a call) and "echo cancellation" which removes reverberation detected within the user's **externality** to allow users to "take their call from anywhere … the office, kitchen, or bathroom" (Krisp, 2022). By using **software** which focuses on providing these features (and combined with high-quality audio **hardware**), gaming communities aim to prevent background noise from the virtual **externality** from entering the collective virtual experience. Similar approaches are already used within business **software** such as Teams, which uses AI-based noise cancellation **software** to allow users to remove sounds such as “children playing loudly nearby, construction noise outside of your home office [or] other unwanted background noise” (Aichner, 2020).

Second, Discord also offers several audio features which are useful for large-scale collaboration: each user can mute any other participant individually (i.e., User A is muted by User B; User A is still heard by User C, who has not muted them) and also individually adjust the volume of other participants (i.e., User A is turned down to 20% volume by User B. User A is still at 100% volume for User C, who has not adjusted them) which allows for a fully customisable audio experience for each user. This can then be combined with “priority speaker” – an option which allows the server moderators to select specific users who will mute all other users when they speak. This allows for leaders within the community to always speak up and be heard, no matter the size of the group or the volume of the on-going discussion.

Discord also offers a variety of options for broadcasting audio. Depending on the selection of games being played and the size of the group playing, many gaming communities require users to use ‘push-to-talk’ and ‘threshold broadcasting’. ‘Push-to-talk’ only broadcasts audio whilst a user-specified key is pressed. This allows users to select convenient keys such as mouse hotkeys or keyboard action keys. Similar push-to-talk tools are available in corporate **software**, but Discord offers greater customisability by allowing you to select your own hotkeys. ‘Threshold broadcasting’ only broadcasts audio when an automatic or user-specified level of noise is detected by a microphone. Allowing manual adjustment (Figure 20) allows users to ensure that unusual extraneous noise in their personal environment does not disrupt the collective environment. The customisability of audio experience is beyond that currently offered within leading professional remote meeting **software** such as Teams, Zoom, and Google Meet, which offer automatic threshold broadcasting detection, or continuous broadcasting, or fixed-key push-to-talk.

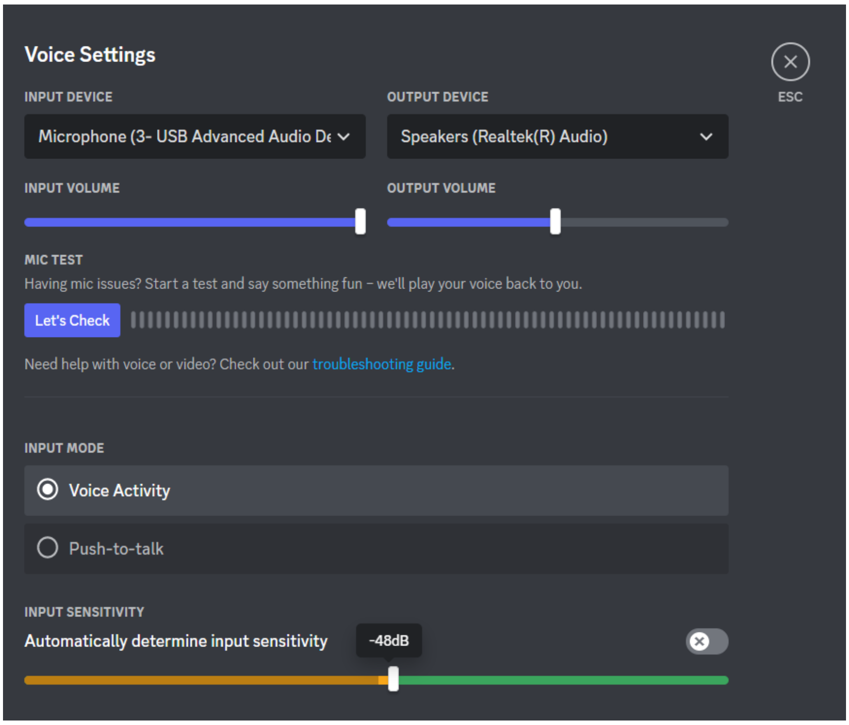

Figure 20: Discord Audio Settings: Voice Activity has a manually adjustable input sensitivity

Alongside broadcast options, Discord also offers a number of options to ensure that users receive voice communication clearly. Alongside the suite of noise mitigation and echo cancellation programs discussed above, which can improve the fidelity of the received environments, Discord also allows users to attenuate the volume of other applications when they or others are speaking (Figure 21). This allows users both to prevent feedback by muting external speakers whilst talking and to hear other users more clearly by muting other computer applications – e.g., a video game, or background music – which may be in use. Whilst corporations are unlikely to wish to encourage users to play video games whilst in a Teams Meeting, providing users with more approaches to reduce the possibility of feedback would be beneficial to the remote meeting experience. Despite this, attenuation is not currently offered by corporate remote meeting **software** such as Teams, Zoom, or Google Meet.

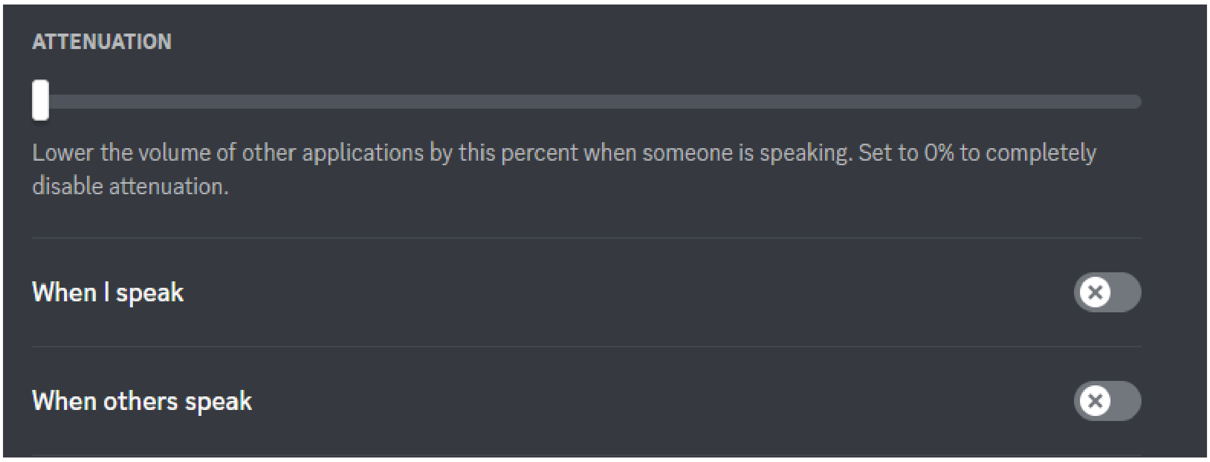

Figure 21: Discord Audio Settings: Attenuation allows users to temporarily lower the volume of other applications when they or other users are broadcasting audio.

One final consideration that may also provide an advantage in business use cases is the ability to spatialise audio. Rudimentary spatialisation is often used by gaming community which collaborate with large groups of several thousand players. Members of former gaming communities I participated in would load up multiple instances of the Teamspeak **software** client to remap outputs to different audio **hardware**. This allowed them to broadcast specific voice channels to specific **hardware**. Typically, this was used to separate out voice channels to their left ear and their right ear: e.g., channels with team leaders directed only to the left ear, whilst channels for general conversation were sent only to the right ear. This allowed users to be more conscious of who information was being distributed to, allowing them to collaborate more easily. Other users used similar strategies to create social communication channels distinct and separate from the ludic communication channels – allowing them to differentiate between (urgent, immediate) ludic instructions and (low priority) social communication.

Whilst this use-case is unlikely to be useful in a business context, where users are rarely expected to be in multiple meetings held simultaneously, spatialisation does offer other (similar) advantages. A frequently requested feature for the Discord platform is audio spatialisation for voice chat; this is not offered natively by the Discord application, as audio is only broadcast and received in mono to minimise bandwidth demands, but supporters of the idea suggest that stereo spatialised audio would allow users to more clearly hear individual voices in group chats with many members ("Chris

Pratt", 2022).[83] This is an updated, and more sophisticated, corollary to the original use-case explored by myself, and other members of the gaming communities I participate within. Stereo spatialisation may help users to correctly identify speaking individuals in large group meetings. We can also imagine that stereo spatialisation may also decrease fatigue in virtual meetings by helping to reduce cognitive load, as reported in comparisons of spatialisation approaches in other virtual experiences (Mendonça & Korshunova, 2020). This would make stereo (and, potentially, binaural) spatialisation a favourable option for business-use cases.[84]

5.5.0: Framework in review

The case studies in this chapter demonstrate that the framework allows for a purposeful examination of current-generation virtual experiences, allowing for intra-contextual and inter-contextual comparisons to be made and lessons to be adopted from one virtual experience to another. Examining cinematic sound allowed for an example of how **hardware** changes may lead to corollary consequences throughout other aspects of **externality** and **software**. Similarly, through examining the virtual experience of remote meetings using the framework and understanding of sound developed through the first half of this thesis, actionable areas of development for virtual meeting **software** can be identified. The framework allows for a systemic dissection of the existing design of a virtual experience and the possible changes that emerge from changes to a virtual experience. From design perspectives, this framework can also be seen as aiding in solving problems in design and implementation by accentuating similarities in superficially different experiences and/or in experiences which share some aspects of design and implementation.

---

[83] Discord users will frequently refer to other applications such as Mumble, which offer customisable stereo spatialisation, in these feature requests; this suggests that sophisticated stereo spatialisation is in use in other gaming communities.

[84] Between the time of submitting this thesis in early 2023 and post-defence amendments revisions being made to this thesis in July 2023, Microsoft Teams added spatialisation to their software. To some extent, this offers support to the findings of the comparative analysis such as conducted throughout sections 5.4.0–5.4.3 which suggested that it would be beneficial to offer this feature for business users given that it was considered to be beneficial in entertainment contexts.

However, the framework has use-cases beyond examining current virtual experiences. As the framework is not based on any specific technological framework but rather emphasises experiential perspectives, it can also be used to examine and consider prospective (and predicted) virtual experiences that might emerge in the future. The next chapter will examine the use of the framework to analyse other virtualisations of human experience in current and next-generation technology.

Chapter 6: The Future

6.0.0: The Framework and The Future

Throughout Chapter 5, a series of case studies were undertaken to show how the framework developed in this thesis can be used to holistically examine various virtual experiences which are in contemporary use, and how the framework might be used to discuss areas of design and implementation. However, the framework presented in this thesis has purpose beyond analysis of existing virtual experiences. As indicated in Chapter 0, the approach to examining virtual experience developed throughout this thesis – which centralises user experience – is predominantly intended to allow for trans-medial and inter-disciplinary approaches. These approaches to analysis, design, and implementation will continue to have much to offer to the development of virtual experience: whilst the underlying technologies used to engage with virtual experience may change (as has happened in recent years with the growing popularity of commercial HMDVRs), the underlying human experience and the fundamental components of virtuality are likely to retain similarities. For example, section 5.3.0–5.3.2 showed that virtual reality cinema and theatrical cinema share many experiential similarities despite the differing underlying technology and investigated an emerging issue in virtual reality cinema wherein changing technology led to potential decreased sensorial immersion. These considerations make most sense when considered trans-medially: i.e., when considering how the user experience has been changed by the differences between two different media.

This chapter, therefore, serves three purposes. First, to demonstrate how the models proposed throughout this thesis will still be relevant to the analysis of future-generation technologies and experiences. Second, to demonstrate how the models proposed in this thesis can function within the design pathways leading to these future-generation technologies, both through aiding in the systemisation of complex technology and through identifying concepts from antecedent technologies which address similar or identical issues. Third, to demonstrate how the models proposed in this thesis can be used as the basis of a discussion which extends beyond just the auditory considerations which are the primary focus of this thesis to broaden our awareness to other potential issues. Sound is a particularly effective starting point for this extension, as sound frequently extends across all elements of virtual experience. In order to do this, the chapter examines antecedent technologies and proof-of-concept technologies, and models how approaches to these technologies can be structured through the use of the framework developed within this thesis.

This chapter also considers how definitions of virtuality may continue to change as we approach a transhuman society (i.e., a society in which technology is used to fundamentally alter or otherwise improve the human experience). This chapter posits that some current-generation technologies can be re-imagined as a virtual experience in which reality is mediated through technological means to generate a unique experiential perspective for a user. For example, section 6.2.1 examines assistive devices for hard-of-hearing and deaf individuals. Whilst not commonly thought of as virtual experiences, sections 6.2.1 – 6.2.2 identify the way in which cochlear implants can and should be thought of as a virtual experience and the benefits which emerge from doing so. This reframing has implications for the development of future technologies.

6.1.0: Next-Generation Wearable Technology

Many commercially successful examples of wearable technology are in use in the present day. Smart watches have become a popular accessory, allowing people easier access to many of the functions of their smartphone and/or basic health monitoring tools that track user biometrics such as heartrate. Interest surrounds the potential of using these wearable devices for other purposes such as detecting heart attacks (Strickland, 2018) or providing diagnostic assistance through providing ECG readings of the user for paramedics (Spaccarotella et al., 2020). Integration of these tools into auditory domains is already taking place: for example, *Bring To Light* (Red Meat Games, 2018) have used biometric tracking to dynamically adjust the difficulty (and ‘creepiness’) of the game in response to the user’s heart rate. Whilst this kind of game remains a novelty, the developer suggested that “using biometric feedback will eventually become an option for games and other apps out there” (Houser, 2018). Perhaps demonstrating the accuracy of these claims, interfacing between biometric data and virtual **software** has already been used in other games to allow for dynamic adaptation of music in response to biofeedback to emphasise psychophysiological effects of music (see Williams, 2021). We can recognise that next-generation wearable technology is likely to build on these examples and find other ways to integrate biometric data into virtual experiences beyond the existing entertainment products that have already begun to use such data. Perhaps more importantly, these developments highlight an interest in and excitement for the possibilities of wearable technology.

Other wearable technology, however, has not proven successful in commercial markets to date. Major tech companies such as Google, Microsoft, and Apple have all attempted to release optical head-mounted displays (OHMDs) for augmented-reality aimed at a commercial market for regular use such as the Google Glass, Microsoft HoloLens, and the forthcoming Apple Vision. Whilst these devices have had some success in business markets such as with the Dutch police force (Hoevers, 2021; Vuzix Corporation, 2020) and American fire service (Rackspace, 2021), issues seem to surround the adoption of these technologies to a broader audience. Google Glass in particular was widely criticised in both critical and popular opinions: e.g., concerns surrounding personal privacy (and the invasion of that privacy by camera-wearing AR-users) led to the term “Glass-hole” being coined to describe someone using OHMDs in public. One of the challenges faced by designers and users of OHMDs, which may explain some of the hesitance preventing widespread adoption of OHMDs, is auditory interfacing. In this context, auditory interfacing includes both allowing information to be conveyed to the user through ‘earcons’ and auditory icons as discussed in sections 2.2.1 and 4.1.4, and the capability for users to interact with their devices through speech. Further, interest also surrounds next-generation developments in audio interfacing such as Google’s ambitious aim to provide “subtitles for the world” (Google, 2022, sec. 1:54:32–1:59:40) as part of the Google Glass 2 OHMD technology.

6.1.1: Google Glass 2: Beyond Virtual Assistants

We can use Google’s bold promise to deliver such a feature to investigate the potential mediatisation[85] of sound within the development of next-generation wearable augmented/mixed reality technology by examining how they – and others with similar projects – are aiming to achieve such a goal. Google Glass 2 bears many similarities to current-generation OHMD technology such as the HoloLens 2, which was discussed in section 0.2.2. Like the HoloLens 2, Google Glass 2 is expected to overlay digital interfaces and virtual objects into the space surrounding the user. One area that has required a great deal of research and development, however, is the “subtitles for the world” feature for Google Glass 2 mentioned above. This prospective feature can be examined through the lens of the framework established in this thesis to demonstrate how the framework can be used to explore the design of future commercial products. What **hardware** and **software** does this feature require, and how might the **externalities** that the Google Glass 2 is expected to be used within effect the experience?

To capture and process sound and display subtitles, the Google Glass 2 must have access to specific **hardware**. Perhaps most obviously, the device must be connected to a microphone. This is already the case for other OHMDs like the Google Glass series: voice commands were a feature of the Google Glass 1, and this was likely to remain a feature of the Google Glass 2 in any case: other **controllers** such as keyboards are unwieldy for use in the contexts which the Google Glass 2 is intended to function within (i.e., it is unlikely that you would be able to comfortably use a keyboard whilst walking down the street). Whilst developments in microphone technology and digital sound processing are likely to improve the overall quality of the audio signal (see, for example, discussion surrounding noise cancellation for use in VoIP in section 5.4.1 – 5.4.3), the underlying required functionality of the input **hardware** (i.e., **controller hardware**) is unchanged.

[85] As in section 0.01, the term mediatisation is used here as defined in the Mirriam-Webster dictionary (“to put in a middle or intermediate position : make instrumental or subordinate”) rather than as used in political communication theory as a word to describe the means through which the media (e.g., TV news stations) shape and frame the process and discourse of society.

But is this **hardware** sufficient to address the new functions that Google Glass 2 wishes to provide to consumers? We can recognise that there are clear differences between interpreting audio signals as voice-inputted commands and translating an audio signal into understandable – and reliably accurate – subtitles. By comparing these differences, we can begin to understand what adjustments Google Glass 2 must make to their existing audio processing systems (i.e., **platform hardware**) for voice commands to ensure that their new audio processing systems (and the underlying vision of 'subtitles for reality') can be used by those that can utilise them, such as d/Deaf/hard-of-hearing individuals.

The process of translating an audio signal into a voice command requires the combination of several fields of research, including audio signal processing (to identify distinct words and phrases spoken by a user) and natural language processing (to identify what function the user is attempting to access). Audio signal processing is a computationally difficult task and, by far, the most common solution is to distribute audio processing to a centralised server farm; virtual assistants such as Alexa, Cortana, and Siri all use such an approach to minimise processing times and maximise the convenience of the virtual assistant. This solution requires audio to be uploaded to a server in real-time and analysed by a computer algorithm. The exact approach used by each company varies but, for example, Amazon describe how devices are pre-programmed to recognise specific 'wake words' such as 'Alexa', 'Amazon' and 'Computer', and then begin recording and uploading data to remote **hardware** when the trigger word is overheard. A neural network is then used to analyse audio signals: e.g., acoustic properties of the speaker's voice during the utterance of the 'wake word' can be used to identify the speaker's voice across the entire audio clip and thereby allow for the voices and noises in the background to be filtered out (Hardesty, 2022). However, the involvement of a neural network implicitly means that the devices are incapable of interpreting the speech without access to the internet. In practice, this means that Alexa, like many virtual assistants, does not have an 'offline' mode. The implications for Google Glass 2 and the "subtitles for reality" are clear: the device will likely need to be internet-connected to provide this function, as it will be dependent upon remote audio signal processing.

However, a further issue can also be identified in Google's ambitious new feature: voice commands often require the use of specifically restricted vocabulary. This allows automated speech recognition (ASR) programs to identify potential context clues to minimise error rate in voice commands: e.g., if a user is adding an event to the calendar, then Amazon's neural network biases towards interpretations of the audio signal which are in a recognisable date-time format such as "August 20th, all day" (Hardesty, 2022).[86] Alongside these approaches, Amazon describe how a new proposed contextual ASR program resolves ambiguity emerging from audio signal processing by considering additional information:

> The acoustic signal is passed through an ASR, which generates multiple hypotheses about what the customer said. The top-ranked hypotheses then pass to a natural-language-understanding (NLU) model, which identifies the customer's intent [such as playing a video] … and the utterance slots – the entities on which the intent should act, such as [the title of a video to be played]. … hypotheses generated by our ASR and NLU models pass to a third model, called HypRank, for hypothesis ranker. HypRank combines the predictions and confidence scores of [ASR and NLU] with contextual signals, such as which skills a given customer has enabled, to produce an overall ranking of the different hypotheses. (Kim, 2021)

To put this in simplest terms: the system uses information about what the customer has previously used the virtual assistant for, the known and recent activities of the user, and other contextual information available to the server to minimise ambiguity in interpreting audio signals. Other information such as the unique 'voice print' of the user may also be used to aid in interpreting audio signals by determining the acoustic properties of the user's **externality** and using this to process the audio signal. The combination of these approaches allows relatively high success rates when using voice commands with virtual assistants.

[86] Similarities to the use of restricted speech to minimise misinterpretation within player communications in video gaming (players develop a game-specific 'language' to allow for communication between international participants with only limited understanding, see Tatlow 2020) or international air travel. By having established dialogic frameworks, misunderstandings are minimised.

The issues identified above are, predominantly, challenges which have historically been addressed through the virtual **software** used to create the augmented reality experience. This is understandable: virtual assistants are a form of computer **software**, and therefore the natural inclination is to consider how the **software** can be updated or adapted in order to facilitate changes in consumer demands. However, the user perspective on virtual assistants is, deliberately, not a **software**-perspective. Amazon makes efforts to develop a ‘conversational’ approach to its virtual assistants by “hiding the computation time under Alexa’s reply to the customer’s request” and using dynamic database management solutions to reduce read/write/search times (Hardesty, 2022). These efforts, in the long-term, are expected to humanise (or at least anthromorphise) Alexa and other virtual assistants (see Abercrombie et al, 2021; Wagner & Schramm-Klein, 2019). Users will not see an exchange with a virtual assistant as an interaction with a computer system (or any other **platform**), but rather an interaction with a digital being.

However, providing “subtitles for reality” adds many layers of complexity beyond those created through creating so-called ‘conversational virtual assistants.’ Rather than interpreting one voice (i.e., the user’s voice), Google Glass 2 must now be capable of interpreting *every* voice (i.e., anyone that the user wishes to speak to). This may also require the device to identify and separate *multiple* voices, and to identify *which voice* should be subtitled for the device user. Similarly, rather than interpreting specific command phrases and contextual words within command phrases, Google Glass 2 aims to be capable of interpreting *every* sentence. We should also expect differences between the structure of a command phrase and the structure of conversational phrases (e.g., informal language, discontinuity in speech, run-on/ungrammatical sentences, etc.), which will raise the ambiguity in (and thereby lower the reliability of) ASR as essential contextual information will be harder to identify.

We can use the framework designed within this thesis to identify methods of addressing each of these issues. For example, Google Glass 2 must be able to identify and separate multiple voices in order to provide subtitles for each speaker. To examine how this might be approached on a design level, we need an multi-dimensional approach as highlighted in section 5.1.4. Many current approaches focus on developing **software** which can correctly interpret multiple voices speaking simultaneously. However, if individual voices could be separated, then existing ASR tools such as those used by virtual assistants may continue to be useful within the next-generation of technology. One such way of separating voices would be to include additional microphones – i.e., additional **controller hardware** – within the device. For example, it is possible to more easily determine the approximate location of any sound detected by more than one microphone by identifying key components of the audio (e.g., through audio fingerprinting – the use of audio meta-data such as spectral flatness, spectral weighting, prominence of frequency bands, zero-crossing points of sound signals, etc. to identify sounds)[87] and then digitally separating sounds with different fingerprints based on their interlineal delay. With the assistance of other specialised digital sound processing **software**, this would allow the device to more accurately differentiate between stationary speakers and thereby provide more accurate subtitles for each individual speaker. In other words, designers can respond to a change in the virtual environment (i.e., more speakers to interpret) by making commensurate changes to the virtual **hardware** (i.e., more microphones) to allow existing virtual **software** (i.e., ASR/CASR) to function in the same way.[88]

It should be noted that this is only a proposed solution which emerges from a holistic consideration through the framework established in this thesis, rather than a solution which is already known to be in use by Google or other tech companies. To develop such solutions into pragmatic

[87] We could perhaps consider these to be mathematical expressions of *timbral* qualities of sound: e.g., 'spectral weighting' expresses the relative intensity of different frequencies within the overall audio spectrum. This could be considered to be the mathematical expression of the subharmonics which determine timbre of instruments.

[88] Speaker differentiation in complex auditory environments is a well-studied topic and is often referred to as 'the cocktail party problem'. The presentation of audio fingerprinting in this section is not a holistic review of the factors which allow modern solutions for this issue to be successful; modern approaches to audio source segmentation (i.e., differentiating between different speakers in a multi-way conversation) frequently use machine learning and neural networking which means that a holistic list of factors is not known, nor possible to know (as the systems generate their own criteria and are unable to share their criteria with their developers). However, for a review of what is known about the approaches to the cocktail party problem, see Hänsler & Schmidt (2008), Vincent et al (2018). Several new approaches have been found in recent years building on Hershey et al (2015) which found a novel way for machine learning to be integrated into segmentation and separation of audio signals on a single microphone: e.g., Asni et al (2018) reviews a convolutional autoencoder approach to speaker differentiation; Nachmani et al (2020) propose a new approach using gated neural networks.

technological advancements, it would need to be examined experimentally to establish whether pragmatic goals are achieved. That is to say, does this multi-microphone approach examining and comparing multiple audio recordings result in sufficiently better accuracy with little or no increase to computational latency when compared to single-microphone approaches?[89] Nonetheless, we can recognise how the framework proposed in this thesis encourages designers and developers to take a holistic approach to resolving issues in the design and implementation of features for forthcoming technologies through cross-sectional examination of the intended virtual experience. Next-generation wearable technology represents only one expression of virtual experience in the future, however. How might the framework be used in the development and design of other technologies which are less immediately present?

6.2.0: Brain-Computer Implants

Recent instalments in the cyberpunk dystopia of the video game franchise *Deus Ex* (2000–2016) have placed players in control of the life of Adam Jensen. Suffering from multiple near-fatal wounds, he is given a variety of prosthetics and machine augmentations, including replacements for his sensory organs: he can zoom in on distant objects with his machine eyes, smell lies with his machine nose, and hear through walls with his machine ears. For Adam Jensen, virtuality is not a mode of existence, it *is* existence. His technological implants irrevocably and holistically alter the world around him. To some extent, such fictions represent the imagined natural development of augmented reality devices of the modern day: the computer aids the human in the tasks the human selects. Asimov's *Bicentennial Man* (1976) offers a meditation on humanity that engages with the same ideas in a different way: what does it mean to be human? Andrew – a robot – slowly replaces his machine parts with biological parts. He surrenders his supernatural hearing and seeing in an effort to be 'more human' – only managing to achieve his goals when he finally surrenders his machine brain and

[89] Perhaps aiding the development of such an approach, most current-generation smartphones and other similar devices already have multiple microphones in any case and – as per section 6.1.1 – contemporary smartphones are now allowing users to use these arrays to record in stereo. However, it is unclear whether the interlineal delay resulting from the physical space between microphones in smartphone arrays allows for analysis of the dynamic sonic environments which emerge from real-actual experiences.

accepts a human death. Unlike the total-immersion worlds, however, humanity is not concerned with the entrapment of the human soul within unbounded virtual environments. Instead, the moral dilemmas within such fictional worlds revolve around whether *humanity* will be lost to these virtual worlds.

Whilst the moral dilemmas of ‘designer humans’ continue to be discussed in technological, biological, and sociological paradigms, rapid progress towards realising the technologies at the centre of these science-ficitonalities as science-actualities is being made. Various approaches have been explored in recent years and brain-computer interfaces (BCIs) – also sometimes known as ‘brain-machine interfaces’ (BMIs) – represent one of the most promising avenues for the near-future of next-generation technology for virtual experiences. BCIs are devices which allow for direct manipulation of the electronic charges of the body, allowing for the reading, editing, or creation of brain activity at the most fundamental level. Rather than presenting the brain with images or sounds which must then be parsed through the human perceptual system, as with current HMD, most BCI devices aim to directly provide that information to the brain *without* the necessary transit of information through the perceptual system.[90] In science-fiction, this often revolves around a desire to escape the human condition. In science-actual this process may be done for many reasons: key contemporary neuro-technologies are understood to revolve around the mitigation of injuries (e.g., interfacing with prosthetics), curing diseases which affect the ocular or aural perception systems (e.g., tinnitus), treatment of neurodegenerative disorders (e.g., Alzheimer’s), and the enhancement of human brain systems (Bansal & Mahajan, 2019, pp. 198–201). In the long-term, it seems likely that BCIs will also have entertainment and communication purposes.

[90] There are, broadly speaking, two types of BCI: invasive (i.e., implanted within the body) and non-invasive (i.e., not implanted within the body). Invasive BCI devices are known to require biocompatible materials which are not yet widely available. However, they nonetheless seem more promising for long-term development as non-invasive BCI devices have a number of key issues: e.g., those based on EEG have long set-up times, inconsistent control schemata, low-signal and/or high-noise leading to poor signal-noise ratio which can make it difficult to process information, and poor signal attenuation. (Rao, 2013, pp.279–280).

One might assume that BCIs are futuristic devices far beyond current technology and unlikely to make any realistic progress within the lifespan of anyone alive today. This is inaccurate: BCIs are already commercially available and have been for several decades. One of the first commercial BCIs is considered to be the cochlear implant (e.g., Rao, 2013, pp. 239–240). Cochlear implants are neuroprostheses which provide assistance to those with sensorineural hearing loss (i.e., issues with the mechanisms of the inner ear) by directly stimulating the auditory nerve with electrical impulses. Whilst these devices may seem superficially simplistic – perhaps not even measuring up to the expectations of 'brain-computer interfacing' established by science fiction – they clearly meet the technical definition: an external input is received by a computer, which processes it, transforms it into an electrical signal, and outputs that signal directly into the brain. But how does the cochlear implant fit within the framework established throughout this thesis?

6.2.1: Cochlear Implants[91]

A variety of devices intended to assist those with hearing issues have been developed. In recent years, the focus has shifted towards implantable sensors for hearing devices. Diego Calero et al. reviewed existing technologies in 2018 and categorised them into three types: hearing aids (HAs), middle ear implants (MEIs) and cochlear implants (Calero et al., 2018, pp. 1–2). In each of these categories, the device uses an external microphone – typically placed near or in the ear – to capture sound around the user. However, unlike hearing aids and middle ear implants,[92] cochlear implants can be considered BCIs as they interface directly with the brain: cochlear implants use an electrode array inserted into the cochlea – the part of the hearing system which transduces vibrations in the middle-ear into electrical impulses in the brain (Casale et al., 2022) – to directly interface with the auditory nerve. As the cochlear implant allows for sensorineural systems to be side-stepped, it addresses issues with these

[91] Since time of submission, a collection published by Drever & Hugill (2023) exploring the auraldiverse listening experiences of deaf/Deaf/hard-of-hearing individuals have been published. Those interested in this topic should also peruse the relevant chapters by Farmer (2023), Kytö (2023), Renel (2023), and Spring (2023).
[92] Whilst each of these devices fulfils the same purpose of capturing, processing, and then delivering sound to the user, only cochlear implants directly interface with the brain: hearing aids capture sound with a microphone, and then play it back with a directional speaker inside the ear; middle-ear implants capture sound signals and translate them into ossicular chain vibrations – the series of bones that vibrate to allow humans to perceive sound.

systems (e.g., insufficient sensory cells, damaged sensory cells, issues in the ossicular chain, etc.) for patients with sensorineural hearing loss. Early attempts at the cochlear implant date back to the 1950s, and implants began to find great success and widespread use from the 1970s.

The cochlear implant can be reimagined as part of a virtual experience within the framework established through this thesis. Whilst it may be odd to consider this a virtual experience, doing so highlights the artificiality of the received sound. Cochlear implants do not provide (nor do they aim to provide) a perfect representation of the real world. Rather, they aim to provide sufficient information about auditory events to improve the user's perception of the world around them.[93] By referring to this as a virtual experience, we can understand that users of cochlear implants are receiving a modified experience of reality, and that cochlear implants are deliberately intended to modify the received auditory input.

Three systems of **hardware** can be observed within the cochlear implant: (1) A sound-reception device (e.g., **controller**; a microphone); (2) A sound-to-impulse processing device (e.g., **platform**; a computer and/or digital sound processor (DSP); (3) An electrical impulse generation device (e.g., **broadcaster**; an electrode). Differences in this **hardware** result in consequences to the overall virtual experience: different microphones have different response profiles to differing frequencies of sound: some microphones 'hear' low-frequency sounds better, whilst others 'hear' high-frequency sounds better. Microphone arrays also have varying response patterns (e.g., unidirectional, cardioid, ultra-cardioid, etc.) which can also affect the frequency response of a given microphone array. By considering the cochlear implant through the lens of the framework developed throughout this thesis, specific issues and factors can be highlighted. For example, the frequency response of the microphones used within cochlear implants have significant effects on the user's

[93] An obvious question here is "What does a cochlear implant sound like, then?" The deeply unsatisfying (but only accurate) answer is that this is, to a large extent, an impossible question: the perceived sound from a cochlear implant is so different that users of the technology report a need to train their ears to understand the new auditory signals. A short discussion of this is provided by SciShow (2017); the featured comments on this video also include several testimonies from users of cochlear implants who attempt to answer this question by sharing their own experiences to adjusting to the technology.

perception of the virtual environment: Emily Halliwell et al. examine how frequency response and input compression can affect music perception (Halliwell et al., 2015), and Alexandra Weatherby et al. examine how frequency response can affect speech perception (Weatherby et al., 2003). We could then draw on inter-contextual approaches to identify how these solutions have been resolved in other virtual experiences, similar to the approach taken in section 5.4.0–5.4.3.

Discussion surrounding the directionality of microphones is also considered in reviews of current-day implants: directional microphones are suggested to improve interaural differences in binaural systems, allowing those with cochlear implants to more easily localise sound sources (Aronoff et al., 2011; Kurz et al., 2021). The size, shape, and number of electrodes within the cochlea can also have an effect on the virtual experience: e.g., Deniz Başkent & Robert Shannon demonstrated that multichannel cochlear implants can mimic the distribution of auditory perception across the cochlea, which improves the overall performance of the implant (Başkent & Shannon, 2004).

DSPs also vary between cochlear implants and have a significant effect on the overall virtual experience. Latest generation cochlear implants have begun to introduce complex signal processing chips which algorithmically reduce auditory input to provide the 'cleanest' output to the user.[94] Distinguishing between **hardware** and **software** may become difficult in some instances when discussing specific devices. For example, both DSPs and auditory processing **software** can be used to filter high-frequency sound from an audio signal. Whilst DSPs are not easily interchangeable at the user-level in the way the commonly accepted understanding of '**software**' may imply, DSPs appear to establish rules of behaviour for how the cochlear implant **hardware** should process inputs and provide outputs. In some cases, it may be useful to therefore consider DSPs as a form of **software** for cochlear implants in order to emphasise their effect on the virtual experience.

The importance of these considerations is clear. If the wrong **hardware** is used within cochlear implants, then the user may be unable to clearly perceive specific ranges – and specific types

[94] For specific details of approaches to DSP in cochlear implants, see Dillier et al. (1992), Tierney et al. (1994), Zeng et al (2008).

– of sound. The overall quality of their virtual experience is significantly decreased. As the virtual experience, in this case, represents the bulk of their day-to-day experiences of the world around them, this poses a significant problem. Alongside these considerations of **hardware** and **software**, we can observe considerations of how the environment surrounding the user (i.e., the virtual **externality**, see section 1.2.1) may affect users of cochlear implants. In particular, cochlear implants underperform when exposed to environments with high levels of background noise (i.e., environments with high signal-noise ratios): e.g., sentence recognition for cochlear implant users was only 64% for one brand of cochlear implant when exposed to noisy stimuli (Balkany et al., 2007). Similarly, King Chung & Nicholas McKibben note that wind noise (caused either by high wind speeds, or due to the user travelling at high speeds – e.g., driving a car with open windows) can be debilitating to cochlear implant users in outdoor environments, and make suggestions for mitigating the negative effects of 'noisy' environments for cochlear implant users such as altering compression ratios across frequency bandwidths, or varying which microphones within the array used in the device were active at any given time (Chung & McKibben, 2011). A final example of this issue can be seen in the use of cochlear implants underwater. Cochlear implants can be used underwater: one case study examines a cochlear implant user who went on over eighty scuba dives with their implants (Kompis et al., 2003). However, sound travels much more efficiently – and much faster – in water, as compared to air. In large bodies (such as oceans and lakes), the water surface is also an excellent acoustic reflector as it is relatively flat and extremely uniform. This means that underwater environments are often both noisy and reverberant. It is unsurprising, therefore, that underwater use of cochlear implants is known to pose some issues. This should be expected as the **hardware** is unlikely to have been calibrated for operation in underwater conditions.[95]

It is clear that the cochlear implant creates an experience which can be described as virtual and examined using the framework constructed within this thesis. By acknowledging and highlighting

[95] It should be noted, of course, that even people without cochlear implants struggle to interpret sounds in water as sound perception in water differs significantly from sound perception in air. See Anthony et al (2010), Montague & Strickland (1961), Norman et al (1971), Parvin & Nedwell (1995).

the artificiality of the cochlear implant experience, we can perhaps justify further considerations of how to ensure that this artificiality – this virtuality – helps users of cochlear implants to experience the least difficulties in their day-to-day lives. It also emphasises the distinction that cochlear implant users are not expecting to receive a perfect representation of the real environments that they are situated within, but rather are expecting to receive a virtualised representation of this environment that allows them to engage with what interests them most. This can be seen as an implicit part of much literature and discussion: e.g., a focus on "signal-noise" ratios implicitly suggests that there are specific sounds which are obstructive to the intended experience which should be filtered out for an 'ideal' experience; and interviews with cochlear implant users often focus on identifying which sounds (e.g., conversation) they most wish to hear in any given scenario or environment.

We can also recognise the way in which perceived audio is prioritised is likely to reflect the understanding of sound developed throughout Chapters 2–4. Signal processing approaches in cochlear implants typically aim to minimise 'noise' (and thereby maximise signal-noise ratios) by reducing extraneous auditory signals. The understanding of noise developed in section 2.2.2 also suggested that misinterpretation may be a vector for the generation of noise in virtual experience; section 2.1.0 also emphasised the subjectivity of auditory experience. These considerations emphasise the need for qualitative research into cochlear user experiences to support existing scientific study, as reported in Gfeller et al. (2019), Harris et al. (2016), and Hughes et al. (2018). The research in this thesis emphasises qualitative research as fundamentally important to the development pathway of devices such as cochlear implants: qualitative research allows for an assessment of the subjective considerations that underpin auditory perception in virtual experiences (such as those discussed throughout Chapters 2–4), including in the virtual(ised) experience created through use of cochlear implants as set out through the framework in this section.

Of course, we should also recognise that modern cochlear implants work extremely well: a review of existing technology suggests that 82% of patients who experienced hearing loss after the development of linguistic skills experienced improved speech perception with a cochlear implant

(Boisvert et al., 2020, pp. 20–21). The hypothetical next-generation auditory brain-computer interfaces are intended to tackle other hearing issues not currently addressed by cochlear implants. One area of particular interest is the mitigation of tinnitus, which is thought to have some neurological component: those suffering from tinnitus seem to show brain activity relating to the perception of sound (Dohrmann et al., 2007; Schmidt et al., 2017; Weisz et al., 2005). It seems plausible that a device intended to control electronic signals within the auditory processing centres of the brain would be able to at least mitigate – if not completely remove – symptoms of tinnitus. We might imagine that, in time, any level of hearing issue including complete loss (or absence) of hearing might be addressed through brain-computer interfacing. Whilst these next-generation devices would not, strictly speaking, be cochlear implants (as they would not be implanted solely within the cochlea), they are broadly based on the same underlying concept.

Through using the framework proposed by this thesis, we could undertake a similar process to the one set out surrounding cochlear implants in this section. This will allow us to identify technologies, strategies, and observations from other virtual experiences which may be applicable to the development of BCIs intended to facilitate auditory processing. The use of the framework may help to clarify avenues of approach for areas of BCI development: e.g., through encouraging designers to draw upon approaches and innovations used in cochlear implants when designing BCIs. Similar to the Google Glass 2, however, it should be noted that the development of BCIs represents a significant increase in the complexity of neuro-devices. Existing devices such as the cochlear implant have very specific purposes and are designed solely to do a single task well. In the case of cochlear implants, the **hardware** and **software** has been designed for the sole purpose of converting real auditory events into virtual auditory events. These future-generations will not able be only to do this, but also serve to provide any other imaginable benefit of BCIs. In essence, BCIs are expected to undergo a shift towards general computing in the 21^{st} century similar to the development of the personal computer throughout the 20^{th} century; BCIs are expected to become ‘Turing complete’ (i.e., capable of computing anything that is computable) rather than assisting with one specific purpose like cochlear implants. How might this shift in purpose have consequences for the underlying design of BCIs, and how might we use the framework to explore these consequences?

6.2.2: General-purpose BCIs

General-purpose BCIs are, in many ways, the “next ‘next-generation’” device, building on wearable technology such as Google Glass by integrating it into the human body itself. The devices will become instantly and continuously accessible; the devices must be able to handle any situation which humans can (un)reasonably expect to encounter. One of the next stages of developing a general-purpose BCI will be discovering ways in which specific-purpose BCIs can be developed. Shih et al. conducted a review of the medical uses of BCIs, suggesting that the devices can be used to help patients communicate, more intuitively control prosthetics, and facilitate and improve neurorehabilitation and motor rehabilitation (J. J. Shih et al., 2012). More recent research has shown that BCIs can also be used to improve cognitive function (Widge et al., 2019) and facilitate memory encoding and recall (Hampson et al., 2018). Other researchers are investigating the possibility of connecting human brains directly to the internet (and, potentially, to each other) (Kremenska et al., 2022) or allowing direct control of peripheral devices such as prosthetic limbs (Tariq et al., 2018), smartphones (Martínez-Cagigal et al., 2018), or ‘smart homes’ (Jeyakumar et al. 2022; Kosmyna et al., 2016). Each of these use cases represents some form of virtualisation of experience, and many of them will require consideration of auditory aspects of experience.

From an audio perspective, one of the most exciting possibilities offered by the shift to a general-purpose BCI is the possibility of extensive integration of other technologies and devices. Current cochlear implants are designed to be self-contained devices: a microphone is processed through a specific DSP, which then generates an electrical impulse using electrodes located in the cochlea. In our considerations of the implications of different types of **hardware** suggested in section 6.2.1, we identified that each of these systems of **hardware** has emergent issues for those that use cochlear implants. A general-purpose BCI may be able to avoid some or all these issues.

Microphone technology was identified as a potential issue for cochlear implants. Due to the size, shape, and position of microphones used in current hearing aids, they are prone to many issues. We might imagine that future-generation devices allow for integration of additional external devices to generate cleaner sound signals from within a given user **externality** or perhaps even to sensorially isolate them completely from the real-actual **externality** to create a completely synthetic auditory environment for the user which includes only the components of the real-actual deemed most essential. Some use of this can already be found. Audio induction loops were developed in the 20th century and became a requirement (where reasonably possible) in the UK as part of the Equality Act 2010. These systems use electromagnetic fields or infrared energy to communicate sound directly to compatible devices allowing for specific sound (i.e., the intended primary signal) to be communicated to the device user with minimal noise. Similarly, many cochlear implant manufacturers have offered Bluetooth integration since 2019, allowing users to connect their smartphones to their implant. This means that users can directly stream audio from their phone (or other compatible devices) to their cochlear implant. This also allows for the best signal-noise ratio during use. For example, one interview with a user of a smartphone-compatible cochlear implant user for the HearingLikeMe blog for people whose lives are affected by hearing loss highlighted in particular how the user "love[s] using the device on the plane to stream bilaterally and not hear the noise of the jet engines" (Büren, 2019). As BCIs move from hearing-specific devices to generic brain-computer devices, it is likely that much of this connectivity and compatibility will be maintained. This also raises the possibility that integration with developing **software** for noise cancellation (e.g., the Krisp noise reduction highlighted in section 5.4.3) will be made relatively simple. This would be likely to improve signal-noise ratios and therefore improve the overall auditory experience within BCIs.

General-purpose BCIs may, eventually, form a generalised mixed reality environment in which the mediatisation of reality symbiotically co-exists with the actualisation of virtuality. For example, we might imagine that our perception of sound might become hyper-real: when receiving directions from a GPS, the BCI could reduce the volume of the real-actual to make it easier to hear and understand the virtual directions. A future-generation device might integrate subvocal recognition devices (e.g., throat microphones) or even capture the *thought* of speaking and interpret these non-auditory signals as a transmissible signal – one which might not even need be reproduced as an auditory signal itself, but which could be translated directly into the *thought* of hearing. Using these future-generation BCI devices, conversations could happen at the speed of thought, and without a single word being spoken or heard aloud. Reality is remixed: reality becomes virtual; virtual becomes reality. In these situations, it will become important to consider the value of auditory authenticity; for example, users who experience a re-mediatisation of reality which may face challenges in navigating the world similar if the created binaural spatialisation is sufficiently inauthentic, as discussed in section 3.0.4.

Whilst the transhuman possibilities which emerge from general-purpose BCIs are exciting, ethical and moral dilemmas also emerge. One benefit of the framework development throughout this thesis is that we can project forward our current understanding of technology to examine emerging technologies. For example, we could consider how **software** hosted on existing **platforms** such as smartphones is affecting our virtual experiences. Songs can be identified using acoustic fingerprinting – e.g., demonstrated by the Shazam music-recognition app (A. L. Wang, 2003). **Software** such as YouTube's ContentID identifies when this copyright material is present within a monetised video – which could form part of a virtual experience – and removes the audio from that section of the video. These systems already work on real-time content and have been used by police officers in the US to prevent recordings and streams of them from being distributed through YouTube (e.g., Massie, 2021). If similar systems are utilised as part of real-time audio for BCI users, then we could imagine a system in which BCI users might find their hearing disabled whenever copyright music is detected. Given the ubiquity of listening in modern culture (see Kassabian, 2013), and the high rates of false positives in these automatic detection systems (Boroughf, 2015; Depoorter & Walker, 2014), discussion surrounding how these differing elements of **hardware**, **software**, and **externality** interact is key to sociological discussions of the (virtual) human experience of the future. Given the links between music and emotion (and between music and semiosis) discussed through Chapter 4, we may also wish to consider how music could be used to influence our lives. Some evidence has already shown that music can influence our political opinions or drive personal activism (Hubbard & Crawford, 2008; Prior, 2022); it seems likely that the affordances offered by devices such as BCIs may lead organisations to consider not only how to use sound and music to beneficially communicate, but also to shape the views of those using such devices. BCIs could alter the sound and music we hear to re-mediatise not only our experience of the world but also our *engagement* with the world: e.g., by muting or playing music which supports specific political messages.

The framework can also be used to consider how new inequalities of human experience might emerge from a widespread adoption of general-purpose BCIs. If we envisage the human body as part of the **hardware** of human experience, then we could suggest that our existing **hardware** typically responds well to frequencies within the 20–20,000Hz range. However, should this biological **hardware** be replaced with technological **hardware** (i.e., in what we might consider as a change of the **platform hardware** for human experience), then commensurate replacement in this biological response frequency will also take place as most consumer microphones are capable of detecting sounds beyond the normal range of human perception, and a computer platform could be programmed to deliver electrical impulses into the auditory cortex when it detects these sounds. This would allow BCI users to hear sounds beyond the normal range of human perception and potentially mean that information transmitted in ultrasound or infrasound frequencies may become ‘transhuman-interpretable.’ A straightforward demonstration of the resultant issue would be a sound art installation which uses ultrasonic frequencies. This art would only be transhuman-interpretable: those without technological augmentations would not be able to perceive the artwork in its entirety as they would be unable to perceive the auditory information beyond organic human capabilities. We can recognise that an environment designed to engage with a transhuman user could potentially be inhospitable to those without technological augmentations.

Other similar concerns have been raised by those working with and developing these technologies. As explored in sections 6.1.1 and 6.2.2, noise cancellation technologies are already developing to allow computers to determine when someone is speaking, who is speaking, and whether that person's voice should be transmitted into a virtual experience. It seems plausible that the same technology could be applied at a larger scale. For example, a user could elect to remove or 'block' someone or certain ideas. Similar ideas were suggested in a 2017 talk by YouTube educator and documentarian Tom Scott who posed the simple question: what if you could block someone from your perception of reality? (Tom Scott, 2017). Whilst this may seem initially useful – e.g., allowing people to block 'hate speech' – the potential for misuse should also be highlighted. Many digital communities online maintain blacklists of individuals who are not welcome within their community. Within virtualised communities established by users of BCIs, it would be possible for some individuals to become 'banned users' – incapable of being heard (or seen) by any within the broader community.

The use of the framework developed throughout this thesis helps to emphasise the virtualisation of reality that takes place as technologies such as BCIs and wearable technology become more sophisticated and more broadly adopted. By recontextualising human experience altered by these technologies as a virtual experience, similar to other virtual experiences, we can more holistically examine the ways in which the virtual environments may become modified by individuals, corporations, and governments. The framework encourages trans-medial considerations (e.g., how might our expectations for the possibilities of future virtual **hardware** be informed by our knowledge about the possibilities of existing virtual **hardware**?) which can help anchor discussions of these pressing matters.

6.3.0: Full-Immersion Technology

One final area of current development is full-immersion technology. One of the best-known examples of full-immersion technology is the virtual world of *The Matrix* (1999), which is based loosely on the concept of simulation explored by Baudrillard in *Simulacra and Simulation* (1981). Here, the

Wachowski Siblings engage with a fear common throughout other science fiction of the 20^{th} century: would humanity know if they were within total-immersion virtual worlds? Within *The Matrix*, humanity has been enslaved. Their physical actual selves are entombed in tanks, existing only because their machine overlords have purpose for them. Their minds, equally, are entombed in virtuality. Nolan's *Inception* (2010) posits an alternate take on a similar idea: Is Dominick Cobb trapped within a dream? How might he (or we, the audience) tell? The 2002 cinematic adaptation on Philip K. Dick's 1956 *Minority Report* posits a direct use-case for full-immersion reality. Violent criminals – or at least, those predicted to become violent criminals – are arrested and sent to a prison where they are trapped inside their minds. "Who knows what they dream of?" asks the warden, glancing across a room filled with virtual reality pods. To some extent it does not matter: as the film explains, the prisoners dream of paradise and therefore never realise their true situation. As they cannot conceive that they are imprisoned, they therefore can never escape.

Full-immersion environments differ from other environments as they fully supplant our perception of the real-actual: in our formulation of virtual experience in section 1.2.1, we described the virtual environment as being encapsulated within a 'magic circle;' a full-immersion environment is one in which the user cannot perceive the *existence* of the magic circle at all. Users are immersed completely across *all* the aspects of immersion identified in section 0.3.0. In these virtualities, users either come to completely accept the virtual environment as a real environment or forget that they are within a virtual environment at all. Users become completely and holistically unaware of the real-actual world at any level.

Full-immersion experiences are some distance away, as there are many different elements which have not yet been solved fully. For example, olfactory sensorial immersion is known to pose great challenges in both the design and implementation of virtual experiences (see Braun, 2019, pp.49–82). However, antecedent full-immersion technologies such as Positrons VR Cinematic Pod (see section 0.2.1) now provide visual, auditory, olfactory and tactile feedback to participants. However, they are only partial-immersion virtual reality experiences as participants are still aware of, and can still interact with, the external reality (i.e., participants are not fully isolated/excluded from the real-actual). As development toward full-immersion technology continues, consideration will be given to two key areas: first, how might users distinguish between virtuality and reality (i.e., what information must be controlled to create a full-immersion experience); second, what are the risks of excluding this information from our perception? For example, as discussed in section 3.0.3, we can recognise that fire alarms are inauthentic (and hence disruptive) to virtuality but, nonetheless, essential to the user's safety; users may be placed in dangerous situations if they are fully immersed within a virtual experience. Using the framework to interrogate some of the questions proposed in 5.1.1 such as "what sounds are present within the user's **externality**" will be key to finding an appropriate balance between safety and immersion in full-immersion experiences.

This chapter has explored several future technologies, such as augmented reality and mixed reality devices, brain-computer implants, and full-immersion technologies. The framework developed throughout this thesis provides a holistic approach to discuss how human experience will be virtualised – and the effects of that virtualisation – by emerging technologies. Three key areas of future technology were investigated: wearable technology such as Google Glass 2 and other augmented reality devices; implantable technology such as brain-computer interfaces; and full-immersion technology. Of particular importance, this chapter shows how the framework can develop our awareness and understanding of virtualities in unexpected or atypical guises. The framework can improve our understanding of audio and hearing within these contexts, which may become increasingly important as human experience is virtualised through increasing integration with new technologies and digitalities. The case studies presented in this chapter also demonstrate that the framework developed throughout this thesis will have relevance and importance as a tool for the design, implementation, and analysis of sound and music in virtual experiences beyond the examples of current practice in virtual experience explored in Chapter 5 and provided as case studies throughout Chapters 1–4.

Chapter 7: And the Rest

7.0.0: Conclusions

How do sound and music function within virtual experience, and how might we approach the design, implementation, and analysis of sound and music in virtual experience? As technology for virtuality continues to develop in the 21st century, these questions will become increasingly important. Sound and music clearly constitute a major part of virtual experience, but many former understandings of sound and music have tended to privilege the intentions and rationalisations of the designers, rather than the experiences and perceptions of the users of these virtual environments. Further, many other studies have focused on specifically technological expressions of virtual experience, which under-values the analytic and design benefits of examining other intersections of virtuality and sound.

This thesis has taken a broader definition of virtual experience which allows for flexible and adaptable analysis across many different expressions of virtual experience, such as virtual reality, augmented reality, and mixed reality. A new framework for the analysis and design of virtual experiences has been suggested. The framework uses a model of virtual experience which has three key elements: virtual software (i.e., behavioural rules), virtual hardware (i.e., tools of interaction), and a virtual externality (i.e., a real-actual space within which the experience is contained). Virtual hardware was broadly broken down into three types: controllers, which can be used by participants to provide information (e.g., keyboard and mouse, joystick); broadcasters, which distribute information such as images or sound to participants (e.g., computer screens); and platforms, which mediate between the two types of hardware (e.g., a computer). The key focus of this framework is experientiality: how do users understand participation within (and limitations of) virtual experience? For example, discussion surrounding virtual externalities has focused on the way in which external sounds can enter – and thereby alter – virtual experience.

Throughout Chapter 2, sound was established as the second axis of the research questions answered throughout this thesis. An understanding of how we might examine sound within virtual experience was developed using Schafer's model of 'soundscapes' This framework for virtuality was considered in correlation with a holistic understanding of sound and music within virtual experience. Chapter 3 considered how authenticity emerges throughout different elements of sound and virtuality, and the role of authenticity in developing perceived immersion. Chapter 4 further developed these findings through the examination of music and established that authenticity within virtual experience exists on two fronts: authenticity *of* virtual experience – i.e., how "plausible" the virtual experience is (Slater, 2009) – and authenticity *to* the virtual experience – i.e., how close the virtual experience is to the user's expectations of the experience. The multi-faceted nature of authenticity must be understood in both design and analysis to improve purposeful virtual experiences, as authenticity can impact on perceived immersion (e.g., sounds which conflict with the portrayed narrative may disrupt narratival immersion).

Key examples of each of different types of virtual experience have been examined through the use of this framework for sound and music within virtual experience. *Half-Life Alyx* was a recurring example used for exploring virtual reality experiences throughout the first section of this thesis, and analysis of Disney theme park experiences – both those envisaged as a mixed reality environment *in situ*, and also those recreated as specifically virtual experiences such as in *Disneyland Adventures* – formed an essential part of the consideration of sound within virtual experience in Chapter 2 and Chapter 3. Chapter 5 provided an overview of the framework and interrogative questions which can be used to analyse a virtual experience through the framework (which are reprised in Appendix A). A variety of case studies were examined through Chapter 5 and 6 using the approach outlined in section 5.1.4. This demonstrates the use of the framework not only for the analysis of current virtual experiences, but also the relevance of the framework to future considerations of virtual experience, including novel conceptualisations of virtual experience such as cochlear implants and other brain-computer implants.

Whilst there are many similarities between predecessor media and modern iterations of virtual experience, virtual experience also poses new challenges for sound designers to overcome.

Three specific challenges examined within this research were immersion, authenticity, and spatialisation. Virtual experience – and especially virtual reality – offers an innately immersive experience. Users are expected to supplant aspects of the real-actual with aspects of the virtual or enhance/integrate aspects of the virtual into the real-actual. As sound is used for a variety of purposes, as explored in Chapter 2 and Chapter 3, we must consider how to communicate within virtuality effectively, whilst being careful to navigate potential conflicts between the virtual and the real-actual.[96]

Spatialisation also poses a number of challenges within virtual reality, as explored within section 3.0.3 – 3.0.4. Sound is inherently spatial, as perhaps best explicated by Pontus Larsson et al. who suggest that spatialisation of sound within virtual experiences can be used to "induce both object presence ('the feeling of something being located at a certain place in relation to myself') and spatial presence ('the feeling of being inside or enveloped by a particular space') at the same time" (Larsson et al., 2010, p. 5). In practical terms, users subconsciously expect the spatialisation of sound within a virtual experience to accurately represent the virtual world. Where it does not do so, it can pose challenges to both immersion and authenticity. In some cases, this can be used purposefully. For example, Max Rieger explains how spatialisation can be used to indicate whether sounds are intended to be diegetic or nondiegetic to virtual reality participants: by contrasting spatialised binaural audio for diegetic sounds with non-spatialised nondiegetic sounds (e.g., narration), experiences can communicate clearly to users whether sounds are supposed to be diegetic (World XR Forum Crans-Montana, 2020). However, it is clear that spatialisation of sound within virtual experience remains an area for future investigation: specific issues in the spatialisation of sound for virtual experience were identified in section 3.0.4, including near-field binaural spatialisation, dynamic audio objects, and real-time audio models for recreating acoustics within real-actual environments.

[96] It should be emphasised that navigation of potential conflicts does not necessarily mean avoidance of potential conflicts: Rieger's notion of using spatial qualities of sound to differentiate between voiceover/narration and character voices within the virtual world (see Section 2.2.1) is an example of how inauthenticity can be used creatively for a constructive purpose. However, the awareness of inauthenticity to/of a virtual experience is essential to discovering these purposeful approaches.

This thesis also presents novel research into the design, composition, and implementation of music for virtual experience throughout Chapter 4. Whilst many functions of music for virtual experience are similar to functions of music within other audio-visual media, relationships between music and the challenges identified above were explored. Music can aid immersion through facilitating understanding of a virtual experience (i.e., through media literacy), and also through physically excluding external sounds from user perception similar to what Grimshaw-Aagaarad and Walter-Hansen refer to as 'reduced reality' within sound design for virtual reality (Walther-Hansen & Grimshaw-Aagaard, 2020). However, music can also *challenge* engagement with (and perceived immersion within) virtual experience as elements of the music are often inauthentic (i.e., implausible). Synchronicity of music and image was highlighted as particularly problematic within virtual reality in section 4.2.3: within film, music often reacts to the diegetic frame. However, within virtual reality, the diegetic frame and the visual frame are not always concurrent: i.e., events may happen behind a user in the virtual world without their awareness. This means that those composing and implementing music for virtual reality need to consider audio-visual-diegetic relationships rather than just audio-visual relationships.

The framework established throughout this thesis was also used to explore future-generation technologies in Chapter 6. Three key areas were investigated: wearable technology such as Google Glass 2 and other augmented reality devices; implantable technology such as brain-computer interfaces; and full-immersion technology. The framework established throughout the first half of the thesis was used to both highlight potential issues within these emerging technologies and find inter-contextual / trans-medial solutions from other virtual experiences. For example, similarities between the new approaches taken to minimise noise-signal ratios in online voice chats discussed in section 5.4.3 were highlighted as a potential way to improve the auditory experience of brain-computer interfaces in section 6.2.2.

Relationships between existing and future technology also exposed issues that may become prominent technological concerns as technology continues to develop. A recurring theme throughout

Chapter 6 was the concept of the transhuman listener; by considering the ways in which personal experience has been virtualised already for deaf listeners, we can identify ways in which the future virtualisation of personal experience may become problematic in the future. We could recognise the concept of 'transhuman listener' as a futuristic example of auraldiversity: the "plurality of senses of hearing… extending to machine listening" (Drever & Hugill, 2023, p.1). The framework may therefore be useful in context with much of the recent research on auraldiversity which discusses how alternate auralities are often under-considered when examining virtual experience; auraldiversity would reflect changes in the underlying models of perception discussed throughout Chapter 2 and expressed – to some extent – as part of the software of a virtual experience. As the approach taken within this thesis is primarily an experiential approach which focuses on the perceptions of the users/participants, the framework presented in this thesis is not only relevant to the analysis of past virtual experience, or the design and implementation of current and near-future virtual experience, but will also remain relevant to the design, implementation, and analysis of future virtual experiences.

7.1.0: Framework in context

The framework presented within this thesis offers a holistic approach to the design, implementation and analysis of sound and music within virtual experiences by providing a holistic approach to discussing the ways in which sound and music function within virtual experience. The purposeful use of both sound and music within virtual experience has been explored throughout the thesis and placed within the context of the framework to demonstrate its use in context.

The opening chapter of this thesis discussed the on-going focus on virtual reality hardware which permeates much of the current discourse surrounding virtual experience and established that virtual experience should be much more broadly applied. Whilst three primary forms of virtual experience were suggested – virtual reality, augmented reality, and mixed reality – it should be noted that all expressions of virtual experience exist on a multi-dimensional spectrum. It is this multi-dimensional spectrum which the tripartite model suggested in Chapter 1 is intended to help explore in a more structured way: as discussed in section 5.1.4, all virtual experiences will require some form of hardware, software, and externality as these are core elements of the conceptualised ‘virtual experience.’

At times, dimensions of this spectrum are immediately obvious. For example, we can imagine that the virtual externality exists on a spectrum of scale – ranging from small office-room virtual experiences (e.g., consumer virtual reality experiences such as *Half-Life Alyx*, discussed throughout Chapter 1–4 of this thesis), to large warehouse scale experiences (e.g., NASA *INFINITE*, see section 5.2.1), and likely everything in-between and to either side as well. Other dimensions of this spectrum may only become apparent upon further consideration: e.g., we might consider the *permeability* of the virtual externality to exist on a scale from a vacuum-isolated anechoic chamber on one end, to an open space in a public area on the other. A more permeable externality will be less isolated from the real-actual and therefore have a more mutable virtual externality from the experiential perspective: users are more likely to conflate real-actual events with the designed virtual experience. A less permeable externality will more holistically prevent transgressions from the real-actual into the virtual, or from (perception of) the virtual into (perception of) the real-actual. These additional dimensions of virtuality, such as permeability, often emerge from the intersection of the three elements of virtuality: e.g., permeability emerges from the interaction of virtual hardware and virtual externality; by changing the virtual hardware used to immerse the user, we expect a commensurate change in the virtual externality of the user. Further research would be required (as explored further in 7.2.0) to identify how other discrete elements of virtual experience may emerge from intersections throughout the framework in this thesis.

Of particular interest throughout this research has been the way in which differing virtual experiences may share aspects of the multi-dimensional framework and how this alters the way that sound is used and/or perceived within those experiences. For example, comparisons between Disneyland (envisaged as a mixed reality experience) and *Disneyland Adventures* (a virtual experience based on the mixed reality experience) were drawn in section 4.3.1 to examine how changes in the construction of the virtual experience resulted in changes to the music of the virtual experience. Similarly, similarities between current-generation cochlear implants and future-generation brain-computer interfaces were identified in section 6.2.2. It was suggested that the design of brain-computer interfaces could be informed the design of cochlear-implants. However, these similarities can also be applied more broadly: e.g., the virtual reality experience of *Half-Life Alyx* and the mixed reality experience of Disneyland were contrasted in Chapter 1. When applied more broadly, the framework established throughout this thesis can emphasise the similarities faced in the design process of different virtual experiences, and perhaps enable developers, designers, and creators to use trans-medial and inter-contextual approaches to design: i.e., using approaches from different modalities of media, or using approaches from within virtual experiences which are identified as similar through the use of the framework.

Boellstorff describes how we are on "the threshold of new banalities of virtuality: the embedding of virtuality into everyday life" (Boellstorff, 2014, p. 744) – this view of an approaching technological singularity or subsummation is increasingly common. Concerns were raised in section 0.0.1 regarding the way in which the virtualisation of society is divorcing human experience from the real-actual. It should perhaps be noted again that the framework constructed throughout this research is broadly applicable beyond technological applications. As a model of human experience, it is possible that it can be applied at the most general level to other fields. For example, we might imagine an urban planning discussion surrounding noise reduction in a public park. Functionally, we can envisage this as an issue with the virtual externality: the permeability of the virtual externality is causing noise from outside the intended experience to enter and disrupt the participant experience within the park: e.g., the sound of passing cars, unwanted (or too many) users of the public space generating ancillary noise. Solutions to this experiential issue can be examined through the framework of virtuality developed in this thesis. Noise isolation devices – such as trees, fences, or barriers – could be installed around the edge of the intended virtual externality to reduce permeability. White noise devices – such as fountains or other water features – could be implemented to reduce the perception of external sounds by raising the noise floor and thereby reducing the perception of the external reality. Virtual software could be re-examined: can access to the space be better managed, or behaviour within the space modified? These solutions are not intended to be novel, but rather to illustrate how the framework developed in this thesis can be used to reconceptualise problems in other areas. This is specifically useful where trans-medial and/or inter-disciplinary solutions can be identified: e.g., how is behaviour within other public spaces (such as the internet) managed, and can these rules be adopted for use within the specific public park?

This research can also be used by practitioners to inform sound design practices by reframing and analysing issues relating to sound and sound design, ensuring a holistic approach to sound design, and allowing for inter-contextual / trans-medial considerations to be used as part of the design process. A frequent major issue in discussing sound design issues is communicating the effects and consequences of design challenges in a way that is interpretable by non-practitioners and/or non-specialists who may be collaborating on a specific project. This framework may provide a useful model for discussing issues with these colleagues (and managers) where necessary as it allows for intermedial discussions focused on user experience rather than any technological paradigm. This is further supported by the particular emphasis on examining user experiences within virtual experience, rather technological phenomena. By focusing on user-based perspectives, sound practitioners are encouraged to focus on how users will encounter and experience their design decisions. The framework also emphasises how design approaches can be incorporated from other virtual experiences and provides a useful system for identifying and appropriating useful approaches to sound design in virtual experience.

7.2.0: Further Research

A wide variety of potential further research has been identified throughout the course of this thesis. This is expected as virtual reality, augmented reality, mixed reality, and other expressions of virtual experience remain active areas of research and development. Further, fields of study into virtuality are in the process of change; this thesis coexists alongside in this broader context of changing approaches to research by suggesting a new approach to examining virtual experience. As new technologies and new approaches to virtuality are uncovered, new experiential issues will arise that challenge the integration and use of the experiences which use them. It is therefore impossible to provide a holistic list of all future issues which must be investigated – many of these issues will only become apparent as new experiences are designed, consumed, and investigated. Instead, the further research discussed here is intended to demonstrate how the experiential framework developed throughout this thesis can be used to highlight specific areas within virtual experiences which are specifically problematic from an experiential perspective.

A recurring theme has been the ways in which hardware mediatises user experiences. The inherent limitations of audio hardware (and the resultant potential for audio infidelity to emerge during audio recording or playback) were identified as a particular challenge to authenticity in virtual reality throughout Chapter 3. This led to a case study examining the potential to adjust virtual hardware used for virtual reality experiences in section 5.1.0 – 5.1.1, which suggested that one challenge to authenticity/fidelity in virtual reality is low-frequency sounds. This was because low-frequency sounds are typically generated or supported through psychoacoustic effects as the hardware used for virtual reality is not able to generate high-fidelity and/or high-intensity low-frequency sounds. These psychoacoustic approaches lack tactile components, which can cause users to feel less immersed in virtual environments where they expect to feel vibrations from low-frequency and ultra-low-frequency sounds. By utilising alternate hardware, user immersion can be improved by introducing tactile feedback. This case study directly targeted one specific vector of inauthenticity within virtual reality experiences and may lead to the development of approaches which directly increase the authenticity of (and therefore, engagement with) virtual experience. However, it is important to emphasise that it also demonstrates how the framework can isolate specific elements of a virtual experience to guide systematic research. Given the breadth of the field of virtual experience, this framework can also be seen as an essential tool for identifying areas in which research is pragmatically feasible to guide future development of virtual technologies and virtual experiences.

Changes in the way that we understand and approach sound within virtual experience can also be expected, and this framework is well situated to aid in future discussions of sound within virtual experience. The beginning of the 20th century saw some concerns that audiences may be confused by, for example, non-diegetic music, or find such music inappropriate (Cooke, 2008, pp.55–58). These concerns now seem amusing; as discussed in Chapter 4, it is almost impossible to imagine a cinematic experience without non-diegetic music when approaching analysis with modern ears. Film is now, as Winters suggests, a "distinctly *musical* reality" (Winters, 2014, p.197, emphasis original). It is difficult to be sure exactly how our experiences of sound will change in the 21st century. However, one area of much current discussion is the ethicality of transhumanism. In these discussions, transhumanism is portrayed as the combination of technology and humanity to create a 'meta-human' or 'super-human'-like figure. Such a figure would have a vastly different experience of reality; one mediated and virtualised by the technology that they have integrated into their life. Potential concerns for transhumanist sound were highlighted and would be excellent ground for further examination and discussion; the framework presented within this thesis offers insights into how such a transhuman individual's experience of reality might be constructed and examined and allows for a dynamic approach to discussing the ways in which human experience has been, and will be, mediatised. These considerations may also offer insight into how auraldiversity might emerge from technological inequalities in the future.

This framework has promising uses across a range of fields as a communicative and analytical tool; the framework may also be able to aid predictions of, and roadmaps for, future development of virtual experience. The use-cases presented within this thesis are only the initial stages of what this framework might be used for, and how this framework might illuminate aspects of virtual experience in an increasingly virtualised world. Whilst the future is, as ever, unclear, it is definite that this framework can offer a rich variety of options and tools to those working with virtuality now and in the future.

7.3.0: Closing Thoughts

"You've got to see this one," the venue manager says with a certain level of unholy glee in their voice, "We've got so many recently, but this one is something special." They pull out their phone and hit play. On the screen, a person in a high-tech (and expensive) VR set-up – complete with haptic controllers, HMDVR, microphones, motion trackers, and rumble packs strapped to their chest – screams loud enough to distort the camera microphone, drops their peripheral hardware, and then turns around and runs away as fast as they can. They hit a wall. Their legs go through it. The rest of them does not. They wobble uneasily to their feet and lift a hand to remove their headset, tapping out from the game. "We get at least one a day," the manager continues, "it never stops being funny." There's a group chat named after the venue which they're flipping through. It's full of this kind of stuff – around one a day at least. Funny or not, it represents the kind of experience that might put someone (and any friends they tell) off virtual reality for life.

The adoption of new virtual reality technologies undoubtedly continues to be slow; users are still adapting to the new challenges posed by the new mediatisation of reality. Several popular communities on Reddit exist to gather compilations of clips and news stories that show the dangers of VR going wrong – It's the 'You've Been Framed' of the 21st-century, with prize-winners gaining internet karma instead of cold hard cash for sharing their – or their friends' – most humiliating moments. One of the most pithily named online communities that gathers such stories is titled "VR to ER" – a reflection of the expectations held by the pundits that gather there to examine the comical stupidity of the future. Other cautionary tales include (thankfully, novelty) VR headsets that literally kill you if you die in-game (Gault, 2022) and health warnings informing potential users that VR can "disrupt the sensory system and lead to symptoms such as nausea, dizziness, sweating, pallor, loss of balance … [cause] temporary change in a person's sensory, motor and perceptual abilities, affecting the manual dexterity or ability to orientate [your] body … [and] trigger epileptic seizures" (ANSES, 2021). Combined with stories such as those set out in the previous paragraph, you can, perhaps, see why some potential consumers are dubious about the underlying virtual reality technology even if (and as) others embrace the immersive possibilities of virtual reality.

But virtual experience goes beyond virtual reality, and other forms of technology have been adopted so holistically that the teenagers I teach barely consider it technology at all. I'm reminded, occasionally, of the way in which my parents despaired when I texted someone rather than called them: how could I be sure that my friends would read it? Now parents are struggling to handle the cameras perpetually attached to their children. Why text your friend when you can video call them? Reading is, after all, gauche. They aren't *nerds*. (Though, outside of their earshot, everyone seems to acknowledge that kids are adopting the latest technologies almost too fast for society to adapt around them). Why not bring your friends along with you on a trip to the museum? – a smartphone with a camera is all your need, and everyone has one of those. These habits only became more prevalent during the social revolution provoked by COVID-19. Nor is it limited solely to 'the youth' - remote working, for example, is here to stay. I can count the times I've been in the same physical space as my doctoral supervisors on my fingers. I imagine I will not be the only research student to have embraced a commute-free life. Virtual conferences are becoming more normal; occasional weekends are lost to organisers from around the world approaching me to investigate the best way to combine online spaces and offline conference rooms – or the best way to use online spaces to create the same kind of 'conference experience' they could spend triple the money to achieve in a physical space. Results vary; attempts continue, nonetheless.

This thesis has not spent too much time examining the way in which technology will change and mutate society – focusing instead on our auditory experiences of virtuality and constructing a framework for examining those experiences – but it is irrefutable that at least one use of this framework will be to examine exactly those changes. Throughout the latter half of this thesis, several examples have already been provided of exactly that – whether that be transhumanism or the driver-free cars of the future. But a constant theme throughout the thesis has been an emphasis on experiential perspective over technological impact: the framework is *not* technological and, as explored through the exploration of Disney as a mixed reality virtual experience (section 1.3.1), doesn't require there to be technological components to a virtual experience to function.

This is likely an essential viewpoint to develop as society becomes increasingly technologically-agnostic. Someone in my gaming community describes vividly how the experience of ‘meeting’ me changed their life; we don’t even know each other’s names. A book club springs up in my fanfic community; does it matter whether the book is ink and paper, pixels on a screen, or just a voice in a box? Somewhere along the way, the discussion surrounding virtual experience has deprivileged – perhaps even abandoned – a technological framework. Perhaps what matters most to the *consumers* of such virtual experiences is the facilitation of their experience, rather than the technological approaches which they use. The framework developed in this thesis, then, might reflect a more human experience of virtuality that exists outside of technological paradigms; the framework in this thesis might reflect the beginnings of a humanistic approach that considers virtual experiences in the same light as they are used, free from considerations of technological paradigms. Designers and developers are unlikely to stray too far from the approaches that consumers use to value the designed products, if only because they need to be sure of a buyer once the product is complete.

Somewhere, however, we can almost sense the metaphysical despair highlighted by Baudrillard: “One can live with the idea of distorted truth. But [the Iconoclasts’] metaphysical despair came from the idea that the image didn’t conceal anything at all, and that these images were in essence not images, such as an original model would have made them, but perfect simulacra, forever radiant with their own fascination” (1981, pp.4–5). Perhaps Baudrillard’s stages of simulation (see 0.0.1) are less of a design guide that helps us to design a convincing experience, and more of a prediction as to how the virtual will manifest itself as being reality in its own right: the virtualisations discussed throughout this thesis are not reflecting or masking reality, but rather creating a *new* reality that perpetuates itself as the new profundity. Our attentions spiral round and round the stages of simulation as we plunge down into the pits of an experiential dystopia, where capitalism reigns supreme and our very eyes and ears are controlled by the bourgeoisie.

At the start of their introductory book on next-generation brain-computer implants (BCIs), Rajesh Rao rhetorically asks: “Are we as a species ready to make such a radical jump in our evolution? Are the governments and regulatory agencies of the world willing to work together to ensure a safe, equitable, and mutually beneficial transition for all to such a future?" (Rao, 2013, p. 280). Any time we feel the need to ask this kind of question, we probably won't like the answer. However, this framework allows to understand *how* this virtualisation of experience might occur – perhaps even allowing us to mitigate the worst possibilities by providing a framework useful for engaging with the governments and regulatory agencies that will need to act to ensure an equitable virtual experience that avoids many of the moral and ethical issues discussed throughout this thesis.

The framework may also allow for more holistic consideration of the risks surrounding future technologies and future virtual experiences by highlighting how issues of accessibility and control emerge as we allow our perspective to be altered and mediatised by external parties; the framework may allow us to address issues of equality in future virtual experiences by allowing us to draw on a broader selection of current-day examples. Developers and designers will be able to use the framework to inform their design and implementation pathways, or as part of their post-mortem analysis to inform their future design. Similarly, the framework will inform future research into virtual experience by providing a tool for analysis.

There are many obstacles to overcome before humanity succumbs to technological mediatisation of reality and embraces a fully virtual world – if such subsummation will ever completely happen at a societal or cultural level at all. However, it is undeniable that throughout the last few decades, mediatisation of human experience has become more common and appears across more aspects of human experience. Therefore, it seems likely that human experience will continue to become increasingly virtualised in many ways over the next few decades. This framework represents a contribution to the toolbox of sound designers, composers, developers of virtual experience, and those with an interest in designing, implementing, or analysing virtual experiences of all kinds as part of the pathways forward. Whilst other tools will also be necessary components of the journey to the future, the framework for sound and music within virtual reality established throughout this thesis offers an essential and flexible approach to considering experiential perspectives on virtuality.

Bibliography

Aarseth, E. (2014). I Fought the Law: Transgressive Play and the Implied Player. In N. Segal & D. Koleva (Eds.), *From Literature to Cultural Literacy* (pp. 180–188). Palgrave Macmillan UK. https://doi.org/10.1057/9781137429704_13

Aarts, R. M., Larsen, E., & Schobben, D. (2002). Improving Perceived Bass and Reconstruction of High Frequencies for Band Limited Signals. *Proceedings of the 1st IEEE Benelux Workshop on Model Based Processing and Coding of Audio*, 59–71

Abercrombie, G., Cercas Curry, A., Pandya, M., & Rieser, V. (2021). Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, 24–33. https://doi.org/10.18653/v1/2021.gebnlp-1.4

Adams, P. C. (2014). Communication in Virtual Worlds. In M. Grimshaw-Aagaard (Ed.), *The Oxford Handbook of Virtuality* (pp. 239–253). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.003

Adorno, T. W. (1981). Bach Defended Against His Devotees. In *Prisms* (1st ed., pp. 133–146). MIT Press.

Adu, P., & Larsen, H. G. (2021). The Theoretical Framework in Phenomenological Research: Development and Application. Routledge. https://doi.org/10.4324/9781003084259

Aichner, R. (2020, December 16). Reduce background noise in Microsoft Teams meetings with AI-based noise suppression. *Microsoft Teams Blog*. https://techcommunity.microsoft.com/t5/microsoft-teams-blog/reduce-background-noise-in-microsoft-teams-meetings-with-ai/ba-p/1992318

Altman, R. (Ed.). (1992). *Sound Theory Sound Practice*. Routledge.

Andreopoulou, A., & Katz, B. (2016, May 26). Investigation on Subjective HRTF Rating Repeatability. Audio Engineering Society Convention 140.

ANSES. (2021). *What are the risks of virtual reality and augmented reality, and what good practices does ANSES recommend?* https://www.anses.fr/en/content/what-are-risks-virtual-reality-and-augmented-reality-and-what-good-practices-does-anses

Anthony, T., Wright, N., & Evans, M. (2010). Review of diver noise exposure. *Underwater Technology*, *29*(1), 21–39. https://doi.org/10.3723/ut.29.021

Apple. (2022). *Capturing Stereo Audio from Built-In Microphones*. Apple Developer Documentation. https://developer.apple.com/documentation/avfaudio/avaudiosession/capturing_stereo_audio_from_built-in_microphones

Arend, J. M., Liesefeld, H. R., & Pörschmann, C. (2021). On the influence of non-individual binaural cues and the impact of level normalization on auditory distance estimation of nearby sound sources. *Acta Acustica*, *5*, 10. https://doi.org/10.1051/aacus/2021001

Arevalo, C., Kariyado, Y., & Villegas, J. (2022). Virtual Reality Tool for Exploration of Three-Dimensional Cellular Automata. *Electronics*, *11*(3), 497. https://doi.org/10.3390/electronics11030497

Arevalo, C., Sarria M., G. M., & Villegas. (2018, July 30). *Accurate Spatialization of VR Sound Sources in the Near Field*. Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction - Aesthetics and Science. https://www.aes.org/e-lib/browse.cfm?elib=19647

Armstrong, C., Thresh, L., Murphy, D., & Kearney, G. (2018). A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database. *Applied Sciences*, *8*(11), 2029. https://doi.org/10.3390/app8112029

Aronoff, J. M., Freed, D. J., Fisher, L. M., Pal, I., & Soli, S. D. (2011). The effect of different cochlear implant microphones on acoustic hearing individuals' binaural benefits for speech perception in noise. *Ear and Hearing*, *32*(4), 468–484. https://doi.org/10.1097/AUD.0b013e31820dd3f0

Aronstein, S. (2012). Pilgrimage and Medieval Narrative Structures in Disney's Parks. In T. Pugh & S. Aronstein (Eds.), *The Disney Middle Ages: A Fairy-Tale and Fantasy Past* (pp. 57–74). Palgrave Macmillan US. https://doi.org/10.1057/9781137066923_4

Asimov, I. (1976). The Bicentennial Man and Other Stories. Doubleday

Asni, M., Shapiro, D., Bolic, M., Mathew, T., & Grebler, L. (2018). Speaker Differentiation Using a Convolutional Autoencoder. *2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*, 1–5. https://doi.org/10.1109/ICSEE.2018.8646307

Assenmacher, I., Kuhlen, T., Lentz, T., & Vorländer, M. (2004). *Integrating Real-time Binaural Acoustics into VR Applications*. Eurographis Sympoisum on Virtual Environments, 129–136

Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatsos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *PLOS ONE, 16*(1), e0244964. https://doi.org/10.1371/journal.pone.0244964

Atkin, A. (2022). Peirce's Theory of Signs. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/peirce-semiotics/

Attali, J. (1985). Noise: *The political economy of music*. University of Minnesota Press.

AudVid Bros. (2016, November 22). *Oculus Rift Commercial—AUDVID BROS.* https://www.youtube.com/watch?v=fqZj8hG6dSc

Aufderheide, P. (1993). *Media Literacy. A Report of the National Leadership Conference on Media Literacy*. Aspen Institute, Communications and Society Program, 1755 Massachusetts Avenue, N. https://eric.ed.gov/?id=ED365294

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

BAE Systems. (2016a). BattleView 360: Cutting Edge Technology So Soldiers Can See Through Vehicles. *BAE Systems | International*. https://www.baesystems.com/en/feature/cutting-edge-technology-so-soldiers-can-see-through-vehicles

BAE Systems. (2016b). Striker® II Digital Helmet-Mounted Display. *BAE Systems | United Kingdom*. https://www.baesystems.com/en-uk/product/striker-ii-digital-helmet-mounted-display

Balkany, T., Hodges, A., Menapace, C., Hazard, L., Driscoll, C., Gantz, B., Kelsall, D., Luxford, W., McMenomy, S., Neely, J. G., Peters, B., Pillsbury, H., Roberson, J., Schramm, D., Telian, S., Waltzman, S., Westerberg, B., & Payne, S. (2007). Nucleus Freedom North American clinical trial. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *136*(5), 757–762. https://doi.org/10.1016/j.otohns.2007.01.006

Balofsky, J. (2017, April 26). An Interview with Stephen. *Real Otaku Gamer*. https://realotakugamer.com/an-interview-with-stephen-cox/26962/

Banbury, S., & Berry, D. (2005). Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics*, *48*(1), 25–37. https://doi.org/10.1080/00140130412331311390

Bang, J., Lee, Y., Lee, Y.-T., & Park, W. (2019). AR/VR Based Smart Policing For Fast Response to Crimes in Safe City. *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 470–475. https://doi.org/10.1109/ISMAR-Adjunct.2019.00126

Bansal, D., & Mahajan, R. (2019). *EEG-based brain-computer interfaces: Cognitive analysis and control applications*. Elsevier/Academic Press.

Barkas, N. (2019). The Contribution of the Stage Design to the Acoustics of Ancient Greek Theatres. *Acoustics*, *1*(1), Article 1, 337–353. https://doi.org/10.3390/acoustics1010018

Bartle, R. (2003). *Designing Virtual World*. New Riders.

Başkent, D., & Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *The Journal of the Acoustical Society of America*, *116*(5), 3130–3140. https://doi.org/10.1121/1.1804627

Baudrillard, J. (1981). *Simulacra and Simulation* (S. F. Glaser, Trans.). University of Michigan Press.

Baysted, S. (2016). *Palimpsest, Pragmatism and the Aesthetics of Genre Transformation: Composing the Hybrid Score to Electronic Arts' 'Need for Speed Shift 2: Unleashed'* (T. Summers, M. Kamp, & M. Sweeney, Eds.; pp. 132–171). Equinox Publishing.

Beardsley, R., & Leech-Wilkinson, D. (n.d.). *A Brief History of Recording to ca. 1950*. Retrieved 22 June 2022, from https://www.charm.rhul.ac.uk/history/p20_4_1.html

Begault, D. R. (1994). *3D-sound for virtual reality and multimedia*. AP Professional.

Beinicke, A., & Muff, A. (2019). Effectiveness of Simulation-Based Learning in Basic Police Training. *European Law Enforcement Research Bulletin*, *4 SCE*, Article 4 SCE.

Ben-Tzur, D., & Colloms, M. (1999, May 1). *The Effect of MaxxBass Psychoacoustic Bass Enhancement on Loudspeaker Design*. Audio Engineering Society Convention 106. https://www.aes.org/e-lib/browse.cfm?elib=8288

Berglund, B., Hassmén, P., & Job, R. F. S. (1996). Sources and effects of low-frequency noise. *The Journal of the Acoustical Society of America*, *99*(5), 2985–3002. https://doi.org/10.1121/1.414863

Berryman, D. R. (2012). Augmented Reality: A Review. *Medical Reference Services Quarterly*, *31*(2), 212–218. https://doi.org/10.1080/02763869.2012.670604

Betbeder, L. (2017). Near-field 3D Audio Explained. *Oculus for Developers*. https://developer.oculus.com/blog/near-field-3d-audio-explained

Bijsterveld, K. (2015). Ears-on Exhibitions: Sound in the History Museum. *The Public Historian*, *37*(4), 73–90. https://doi.org/10.1525/tph.2015.37.4.73

Binns, A. (2017). 'Sounding' Japanese: Traditions of Music in Japanese Cinema. In M. Mera, R. Sadoff, & B. Winters (Eds.), *The Routledge Companion to Screen Music and Sound* (1st ed., pp. 428–439). Routledge. https://doi.org/10.4324/9781315681047-35

Blade, R. A., & Padgett, M. L. (2015). Virtual Environments Standards and Terminology. In K. M. Stanney & K. S. Hale (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (Second edition, pp. 23–35). CRC Press, Taylor & Francis Group.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, *4*(1), 11–44. https://doi.org/10.1207/s15327051hci0401_1

Bodi, B., & Thon, J.-N. (2020). Playing stories?: Narrative-dramatic agency in Disco Elysium (2019) and Astroneer (2019). *Frontiers of Narrative Studies*, *6*(2), 157–190. https://doi.org/10.1515/fns-2020-0012

Boellstorff, T. (2014). An Afterword in Four Binarisms. In M. Grimshaw-Aagaard (Ed.), The Oxford Handbook of Virtuality (pp. 739–746). Online edition. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.047

Boer, J. (2020, June 5). *3rd-person games: Should positional audio be from the perspective of the camera or the character?* [Reddit Post]. R/Gamedev. www.reddit.com/r/gamedev/comments/gwv8yw/3rdperson_games_should_positional_audio_be_from

Boisvert, I., Reis, M., Au, A., Cowan, R., & Dowell, R. C. (2020). Cochlear implantation outcomes in adults: A scoping review. *PloS One*, *15*(5), e0232421. https://doi.org/10.1371/journal.pone.0232421

Boroughf, B. (2015). The Next Great YouTube: Improving Content ID to Foster Creativity, Cooperation, and Fair Compensation. *Albany Law Journal of Science & Technology*, *25*(1).

Bower, J. (2018, September 29). How VR is Being Used in Racing Training? *Tire Meets Road*. https://tiremeetsroad.com/2018/09/28/how-vr-is-being-used-in-racing-training/

Braak, E. V. D., Marshall, H., Meyer, J., Halstead, M., & Protheroe, D. (2019). *Further Investigation of Ensemble Singers Preferred Sound Fields*. 39–46.

Brassard, M., Canty, D., Fiset, A., Greenberg, P., Ikeda, R., Janzen, E., Kluger, J., Koek, A., Lajeunesse, F., Raphael, P., TIME Studios, & Valcourt, T. (2022). *THE INFINITE: Living among the stars*. Hirmer Verlag.

Braun, M. H. (2019). *Enhancing User Experience with Olfaction in Virtual Reality* [Doctoral Thesis]. City, University of London.

Brigham, T. J. (2017). Reality Check: Basics of Augmented, Virtual, and Mixed Reality. *Medical Reference Services Quarterly*, *36*(2), 171–178. https://doi.org/10.1080/02763869.2017.1293987

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, *18*(3), 297–326.

Brown, B. S. “Hale’s Tours and Scenes of the World.” Moving Picture World 29, no. 3 (July 15, 1916): 372–73.

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1297–1300. https://doi.org/10.1145/985921.986048

Bryson, S. (1995). Approaches to the successful design and implementation of VR applications. In R. A. Earnshaw, J. A. Vince, & H. Jones (Eds.), *Virtual reality applications: Proceedings of the Conference on Applications of Virtual Reality, British Computer Society 1994* (pp. 3–18). Academic Press.

Bukatman, S. (1993). *Terminal Identity: The Virtual Subject in Postmodern Science Fiction*. Duke University Press.

Büren, J. von. (2019, May 10). New Tech Turns Cochlear Implants into Wireless Headsets. *Hearing Like Me*. https://www.hearinglikeme.com/bluetooth-cochlear-implant-wireless-headset/

Butler, O. (2017). I Made My Shed the Top-Rated Restaurant on TripAdvisor. *Vice*. https://www.vice.com/en/article/434gqw/i-made-my-shed-the-top-rated-restaurant-on-tripadvisor

Byford, S. (2021, February 24). New Oculus Quest feature lets you mark your actual couch in VR. *The Verge*. https://www.theverge.com/2021/2/24/22298629/oculus-quest-couch-mode-feature-guardian-experimental-rollout

Cairns, P., Cox, A., & Nordin, A. I. (2014). Immersion in Digital Games: Review of Gaming Experience Research. In *Handbook of Digital Games* (pp. 337–361). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118796443.ch12

Caldas-Coulthard, C. R. (2003). Cross-Cultural Representation of 'Otherness' in Media Discourse. In G. Weiss & R. Wodak (Eds.), *Critical discourse analysis: Theory and interdisciplinarity* (pp. 272–296). Palgrave Macmillan.

Calero, D., Paul, S., Gesing, A., Alves, F., & Cordioli, J. A. (2018). A technical review and evaluation of implantable sensors for hearing devices. *BioMedical Engineering OnLine*, *17*(1), 23. https://doi.org/10.1186/s12938-018-0454-z

Calleja, G. (2011). In-game: From immersion to incorporation. MIT Press.

Calleja, G. (2013). Immersion in Virtual Worlds. In M. Grimshaw-Aagaard (Ed.), *The Oxford Handbook of Virtuality* (pp. 222–236). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.012

Cantrell, S. J., Winters, R. M., Kaini, P., & Walker, B. N. (2022). Sonification of Emotion in Social Media: Affect and Accessibility in Facebook Reactions. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1–26. https://doi.org/10.1145/3512966

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, *51*(1), 341–377. https://doi.org/10.1007/s11042-010-0660-6

Casale, J., Kandle, P. F., Murray, I., & Murr, N. (2022). Physiology, Cochlear Function. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK531483/

Casanelles, S. (2016). Mixing as a Hyperorchestration Tool. In L. Greene & D. Kulezic-Wilson (Eds.), *The Palgrave Handbook of Sound Design and Music in Screen Media: Integrated Soundtracks* (pp. 57–72). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51680-0_5

Cassidy, G., & Macdonald, R. (2009). The effects of music choice on task performance: A study of the impact of self-selected and experimenter-selected music on driving game performance and experience. *Musicae Scientiae*, *13*(2), 357–386. https://doi.org/10.1177/102986490901300207

Cassidy, S. B. (2011). The Videogame as Narrative. *Quarterly Review of Film and Video*, *28*(4), 292–306. https://doi.org/10.1080/10509200902820266

Castañeda, M. (2018). The Power of (Mis)Representation: Why Racial and Ethnic Stereotypes in the Media Matter. In H. Jiménez (Ed.), *Challenging Inequalities: Readings in Race, Ethnicity, and Immigration* (pp. 200–207). Cognella.

Castronova, E. (2006). *Synthetic Worlds: The Business and Culture of Online Games*. University of Chicago Press.

CBC News. (2019, November 28). Empathy-based virtual reality training a 'game changer' for Halton police. *CBC*. https://www.cbc.ca/news/canada/toronto/halton-police-virtual-reality-1.5376107

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human–Computer Interaction*, *36*(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Chanthadavong, A. (2017). Victorian volunteer firefighters to receive training through virtual reality simulation. *ZDNet*. https://www.zdnet.com/article/victorian-volunteer-firies-to-receive-training-through-virtual-reality-simulation/

Chatterji, S. A. (1999, April). *The Culture-specific use of sound in india cinema*. International Symposium on Sound in Cinema. http://filmsound.org/india/

Cheng, W. (2014a). *Sound Play: Video Games and the Musical Imagination*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199969968.001.0001

Cheng, W. (2014b). Acoustemologies of the Closet. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality*. https://doi.org/10.1093/oxfordhb/9780199826162.013.011

Chi, C.-F., Dewi, R. S., & Huang, M.-H. (2017). Psychophysical evaluation of auditory signals in passenger vehicles. *Applied Ergonomics*, *59*, 153–164. https://doi.org/10.1016/j.apergo.2016.08.019

Chion, M. (1994). *Audio-Vision: Sound on Screen*. Columbia University Press.

Chris Pratt. (2022, December 10). *Feature request: Static spatial audio for group voice chat*. Discord. http://support.discord.com/hc/en-us/community/posts/10858010590487-Feature-request-Static-spatial-audio-for-group-voice-chat

Chung, K., & McKibben, N. (2011). Microphone Directionality, Pre-Emphasis Filter, and Wind Noise in Cochlear Implants. *Journal of the American Academy of Audiology*, *22*(09), 586–600. https://doi.org/10.3766/jaaa.22.9.4

Churchill Navigation. (2020). *ARS Augmented Reality Mapping for Airborne Law Enforcement*. https://churchillnavigation.com/ars-airborne-law-enforcement/

Clarke, J. A., Chatterjee, S., Li, Z., Riede, T., Agnolin, F., Goller, F., Isasi, M. P., Martinioni, D. R., Mussel, F. J., & Novas, F. E. (2016). Fossil evidence of the avian vocal organ from the Mesozoic. *Nature*, *538*(7626), Article 7626. https://doi.org/10.1038/nature19852

Cohen, A. J. (2013). Congruence-Association Model of music and multimedia: Origin and evolution. In S.-L. Tan, A. J. Cohen, S. D. Lipscomb, & R. A. Kendall (Eds.), *The Psychology of Music in Multimedia* (pp. 17–47). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199608157.003.0002

Collins, K. (2013). *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. MIT Press.

Collins, K. (2014). Breaking the Fourth Wall? User-Generated Sonic Content in Virtual Worlds. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality*. https://doi.org/10.1093/oxfordhb/9780199826162.013.028

Cook, J. (2016, October 3). Playing with the Past in the Imagined Middle Ages: Music and Soundscape in Video Game. *Sounding Out!* https://soundstudiesblog.com/2016/10/03/playing-with-the-past-in-the-imagined-middle-ages-music-and-soundscape-in-video-game/

Cook, J. (2020). Sonic Medievalism, World Building, and Cultural Identity in Fantasy Video Games. In K. Fugelso (Ed.), *Studies in Medievalism XXIX: Politics and Medievalism (Studies)* (pp. 217–238). Boydell & Brewer. https://doi.org/10.1017/9781787448957.015

Cook, J., Kirkman, A., McAlpine, K., & Selfridge, R. (Forthcoming). Hearing historic Scotland: Reflections on recording in virtually reconstructed acoustics. *Journal of the Alamire Foundation*.

Cook, J., Kolassa, A., Robinson, A., & Whittaker, A. (Forthcoming). *History As Fantasy*. Routledge.

Cook, K. M., Meyer, S. C., & Yri, K. (2020). Gaming the Medievalist World in Harry Potter. In *The Oxford Handbook of Music and Medievalism*. https://doi.org/10.1093/oxfordhb/9780190658441.013.45

Cooke, M. (2008). *A History of Film Music*. Cambridge University Press. https://doi.org/10.1017/CBO9780511814341

Coulthard, L. (2016). Acoustic Disgust: Sound, Affect, and Cinematic Violence. In L. Greene & D. Kulezic-Wilson (Eds.), *The Palgrave Handbook of Sound Design and Music in Screen Media: Integrated Soundtracks* (pp. 183–193). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51680-0_13

Cowper, T. J., & Buerger, M. E. (2005). *Improving Our View of the World: Police and Augmented Reality Technology* (Futures Working Group). FBI.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (1. HarperPerennial ed). HarperPerennial.

Csikszentmihalyi, M. (2000). The contribution of flow to positive psychology. In *The science of optimism and hope: Research essays in honor of Martin E. P. Seligman* (pp. 387–395). Templeton Foundation Press.

Cuevas-Rodriguez, M., Gonzalez-Toledo, D., Reyes-Lecuona, A., & Picinali, L. (2021). Impact of non-individualised head related transfer functions on speech-in-noise performances within a synthesised virtual environment. *The Journal of the Acoustical Society of America*, *149*(4), 2573–2586. https://doi.org/10.1121/10.0004220

Dapcevich, M. (2021, April 4). Is There a ‘No-Fly Zone’ Over Walt Disney World? *Snopes.Com*. https://www.snopes.com/fact-check/no-fly-zone-wdw/

Davies, W. J. (2023). Autistic Listening. In J. L. Drever & A. Hugill, *Aural Diversity* (1st ed., pp. 90–100). Routledge. https://doi.org/10.4324/9781003183624-10

Davison, A. (2013). Title Sequences for Contemporary Television Serials. In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), *Oxford Handbook of New Audiovisual Aesthetics* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199733866.013.0036

De Visscher, E. (2014). Sight and Sound: From a Museum of Instruments to a Museum of Music. *Music in Art*, *39*(1–2), 237–241.

Depoorter, B., & Walker, R. (2014). Copyright False Positives. *Notre Dame Law Review*, *89*(1), 319.

Derbyshire Constabulary. (2020). *Derbyshire police trialling cutting edge virtual reality training tool*. https://www.derbyshire.police.uk/news/derbyshire/news/news/forcewide/2020/august/derbyshire-police-trialling-cutting-edge-virtual-reality-training-tool/

Deters, S. (2014). The History and Future of Public Programming at Musical Instrument Museum Edinburgh. *Music in Art*, *39*(1–2), 257–266.

Dick, P. (1956). The Minority Report. *Fantastic Universe*, 4(6)

Dillier, N., Bögli, H., & Spillmann, T. (1992). Digital Speech Processing for Cochlear Implants. *ORL*, *54*(6), 299–307. https://doi.org/10.1159/000276319

Discord. (2022). *Krisp FAQ*. Discord Documentation. https://support.discord.com/hc/en-us/articles/360040843952-Krisp-FAQ

Dohrmann, K., Elbert, T., Schlee, W., & Weisz, N. (2007). Tuning the tinnitus percept by modification of synchronous brain activity. *Restorative Neurology and Neuroscience*, *25*, 371–378.

Donington, R. (1973). The Choice of Instruments in Baroque Music. *Early Music*, *1*(3), 131–138.

Donnelly, K. J. (2016). Emotional Sound Effects and Metal Machine Music: Soundworlds in Silent Hill Games and Films. In L. Greene & D. Kulezic-Wilson (Eds.), *The Palgrave Handbook of Sound Design and Music in Screen Media: Integrated Soundtracks* (pp. 73–88). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51680-0_6

D'Orazio, D., Rovigatti, A., & Garai, M. (2019). The Proscenium of Opera Houses as a Disappeared Intangible Heritage: A Virtual Reconstruction of the 1840s Original Design of the Alighieri Theatre in Ravenna. *Acoustics*, *1*(3), Article 3, 694–710. https://doi.org/10.3390/acoustics1030041

Drever, J. L. (2019). 'Primacy of the Ear' – But Whose Ear?: The case for auraldiversity in sonic arts practice and discourse. *Organised Sound*, *24*(1), 85–95. https://doi.org/10.1017/S1355771819000086

Drever, J. L., & Hugill, A. (2023). Aural Diversity: General Introduction. In J. L. Drever & A. Hugill, *Aural Diversity* (1st ed., pp. 1–12). Routledge. https://doi.org/10.4324/9781003183624-1

Dreyfus, L. (1983). Early Music Defended against Its Devotees: A Theory of Historical Performance in the Twentieth Century. *The Musical Quarterly*, *69*(3), 297–322. https://doi.org/10.1093/mq/LXIX.3.297

Dreyfus, L. (1997). Early Music Defended against its Devotees—Postscript 1997. *Canor*, *19*, 27–39.

Eisler, H., & Adorno, T. W. (1947). *Composing for the films*. Atlone Press.

Ekman, I. (2008). Psychologically Motivated Techniques for Emotional Sound in Computer Games. Proc. AudioMostly 2008, 20–26.

Ekman, I. (2014). A Cognitive Approach to the Emotional Function of Game Sound. In K. Collins, B. Kapralos, & H. Tessler (Eds.), *The Oxford Handbook of Interactive Audio*. https://doi.org/10.1093/oxfordhb/9780199797226.013.012

Ermi, L., & Mäyrä, F. (2011). Fundamental components of the gameplay experience: Analysing immersion. In S. Günzel, M. Liebe, & D. Mersch (Eds.), *DIGAREC Keynote-Lectures 2009/10*, (pp. 88–115). University Press.

Eroglu, S., & Harrell, G. D. (1986). Retail crowding: Theoretical and strategic implications. *Journal of Retailing*, *62*(4), 346–363.

Evans, R. (2019). *Analyzing and Designing Dynamic Music Systems for Games* [Phd, University of York]. https://etheses.whiterose.ac.uk/27365/

Eyewitness News ABC7NY. (2019, April 25). *NYPD using VR to train for active shootings and real-life scenarios*. https://www.youtube.com/watch?v=VZyhQZSTIGQ

Fabian, D. (2001). The Meaning of Authenticity and The Early Music Movement: A Historical Review. *International Review of the Aesthetics and Sociology of Music*, *32*(2), 153–167. https://doi.org/10.2307/1562264

Faccio, M., & McConnell, J. J. (2020). Death by Pokémon GO: The Economic and Human Cost of Using Apps While Driving. *Journal of Risk and Insurance*, *87*(3), 815–849. https://doi.org/10.1111/jori.12301

Farmer, P. (2023). ∼∼. In J. L. Drever & A. Hugill, *Aural Diversity* (1st ed., pp. 82–89). Routledge. https://doi.org/10.4324/9781003183624-9

Feaster, P. (2010). Édouard-Léon Scott de Martinville: An annotated Discography. *ARSC Journal*, *XLI*(i), 43–82.

Feiner, S. K. (2002). Augmented Reality: A New Way of Seeing. *Scientific American*, *286*(4), 48–55.

Flinn, C. (1992). *Strains of Utopia: Gender, Nostalgia, and Hollywood Film Music*. Princeton University Press.

Fornäs, J. (1995a). *Cultural theory and late modernity*. London: Thousand Oaks; California: Sage Publications.

Fornäs, J. (1995b). Do you see yourself?: Reflected subjectivities in youthful song texts. *YOUNG*, *3*(2), 3–22. https://doi.org/10.1177/110330889500300202

Fox, J., & Tang, W. Y. (2017). Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. *New Media & Society*, *19*(8), 1290–1307. https://doi.org/10.1177/1461444816635778

Fritsch, M. (2018). Musik. In B. Beil, T. Hensel, & A. Rauscher (Eds.), *Game Studies* (pp. 87–107). Springer Fachmedien. https://doi.org/10.1007/978-3-658-13498-3_5

Fritsch, M., & Summers, T. (2021). Realities, Perception and Psychology: Introduction. In M. Fritsch & T. Summers (Eds.), *The Cambridge Companion to Video Game Music* (pp. 263–318). Cambridge University Press.

Gahtan, M. W. (2014). Music in Museums: Strange Beginnings of the Avant-Garde. *Music in Art*, *39*(1–2), 195–200.

Garitee, L. (2022, November 22). Improving Officer Performance with Realistic Training. *VirTra News*. https://www.virtra.com/improving-officer-performance-with-realistic-training-blog/

Garner, T. A., & Grimshaw, M. (2013). Sonic Virtuality: Understanding Audio in a Virtual World. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp.364-377). https://doi.org/10.1093/oxfordhb/9780199826162.013.041

Garreau, J. (1994, March 9). Office Minefield. *Washington Post*. https://www.washingtonpost.com/archive/lifestyle/1994/03/09/office-minefield/3b74132a-5f0a-455f-a04e-6171d023149b/

Garun, N. (2012, April 19). Vibrating theater seats are shaking things up with the movie-going experience. *Digital Trends*. https://www.digitaltrends.com/cool-tech/vibrating-theater-seats-are-shaking-things-up-with-the-movie-going-experience/

Gault, M. (2022, November 7). Palmer Luckey Made a VR Headset That Kills the User If They Die in the Game. *Vice*. https://www.vice.com/en/article/dy7kbq/palmer-luckey-made-a-vr-headset-that-kills-the-user-if-they-die-in-the-game

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human–Computer Interaction*, *2*(2), 167–177. https://doi.org/10.1207/s15327051hci0202_3

Gellar-Goad, T. H. M., & Moore, T. J. (2015). Using Music in Performing Roman Comedy. *The Classical Journal*, *111*(1), 37. https://doi.org/10.5184/classicalj.111.1.0037

Geoff Marshall. (2019, October 25). *Driving the SWR Simulator in Bad Weather Conditions*. https://www.youtube.com/watch?v=OoSgU-7BiV4

Gfeller, K., Driscoll, V., & Schwalje, A. (2019). Adult Cochlear Implant Recipients' Perspectives on Experiences With Music in Everyday Life: A Multifaceted and Dynamic Phenomenon. *Frontiers in Neuroscience*, *13*. https://www.frontiersin.org/articles/10.3389/fnins.2019.01229

Gibbons, W. (2020). The Problem with Players: A Response to Jonathan Godsall. *Journal of Sound and Music in Games*, *1*(1), 116–118. https://doi.org/10.1525/jsmg.2020.1.1.116

Giroux, H. A., & Pollock, G. (2010). *The Mouse that Roared: Disney and the End of Innocence* (Updated and Expanded edition). Rowman & Littlefield Publishers.

Giuffrida, A. (2022, February 27). Turmoil in small Italian town after judge silences church bells. *The Guardian*. https://www.theguardian.com/world/2022/feb/27/turmoil-in-small-italian-town-after-judge-silences-church-bells

Godsall, J. (2020). Unlimited Replays: Video Games and Classical Music, by William Gibbons. *Journal of Sound and Music in Games*, *1*(1), 113–115. https://doi.org/10.1525/jsmg.2020.1.1.113

Goetting, B. (2022, May 19). Apple Dislikes Windows So Much Its Self Driving Car Might Not Have Any. *HotHardware*. https://hothardware.com/news/apple-car-patent-hints-vr-in-place-of-windows-hilarious-reaction

Google. (2022, May 11). *Google Keynote (Google I/O '22)*. https://www.youtube.com/watch?v=nP-nMZpLM1A

Gorbman, C. (1987). *Unheard melodies: Narrative film music*. London : BFI Pub. ; Bloomington : Indiana University Press.

Government of Canada, R. C. M. P. (2017, February 7). *Virtual reality training*. https://www.rcmp-grc.gc.ca/en/gazette/virtual-reality-training

Gray, K. L. (2012a). Intersecting Oppressions and Online Communities. *Information, Communication & Society*, *15*(3), 411–428. https://doi.org/10.1080/1369118X.2011.642401

Gray, K. L. (2012b). Deviant bodies, stigmatized identities, and racist acts: Examining the experiences of African-American gamers in Xbox Live. *New Review of Hypermedia and Multimedia*, *18*(4), 261–276. https://doi.org/10.1080/13614568.2012.746740

Green, J. (2010). Understanding the Score: Film Music Communicating to and Influencing the Audience. *The Journal of Aesthetic Education*, *44*(4), 81. https://doi.org/10.5406/jaesteduc.44.4.0081

Gregersen, A., & Grodal, T. (2008). Embodiment and Interface. In B. Perron & M. J. P. Wolf (Eds.), The Video Game Theory Reader 2 (pp. 65–83). Routledge.

Grimshaw, M. (2014). Sound. In B. Perron & M. J. P. Wolf (Eds.), *The Routledge Companion to Video Game Studies* (pp. 117–124). Routledge.

Grimshaw, M., & Garner, M. (2015). *Sonic Virtuality: Sound as Emergent Perception*. Oxford University Press.

Grimshaw-Aagaard, M. (2021). A Step Back from Reality: Sound and Presence in Computer Games and Other Worlds. In M. Fritsch & T. Summers (Eds.), *The Cambridge Companion to Video Game Music* (pp. 269–283). Cambridge University Press. https://doi.org/10.1017/9781108670289.017

Guay, L.-M., & Arsenault, D. (2012). Thumb-Bangers: Exploring the Cultural Bond between Video Games and Heavy Metal. In A. Brown & K. Fellezs (Eds.), *Heavy Metal Generations* (pp. 103–115). Brill. 10.1163/9781848881471_010

Gunkel, D. J. (2018). The Relational Turn: Third Wave HCI and Phenomenology. In M. Filimowicz & V. Tzankova (Eds.), *New Directions in Third Wave Human-Computer Interaction: Volume 1—Technologies* (pp. 11–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-73356-2_2

Ha, H.-G., & Hong, J. (2016). Augmented Reality in Medicine. *Hanyang Medical Reviews*, *36*(4), 242–247. https://doi.org/10.7599/hmr.2016.36.4.242

Halliwell, E. R., Jones, L. L., Fraser, M., Lockley, M., Hill-Feltham, P., & McKay, C. M. (2015). Effect of input compression and input frequency response on music perception in cochlear implant users. *International Journal of Audiology*, *54*(6), 401–407. https://doi.org/10.3109/14992027.2014.986689

Hambleton, E. (2020). Gray Areas: Analyzing Navigable Narratives in the Not-So-Uncanny Valley Between Soundwalks, Video Games, and Literary Computer Games. *Journal of Sound and Music in Games*, *1*(1), 20–43. https://doi.org/10.1525/jsmg.2020.1.1.20

Hampson, R. E., Song, D., Robinson, B. S., Fetterhoff, D., Dakos, A. S., Roeder, B. M., She, X., Wicks, R. T., Witcher, M. R., Couture, D. E., Laxton, A. W., Munger-Clary, H., Popli, G., Sollman, M. J., Whitlow, C. T., Marmarelis, V. Z., Berger, T. W., & Deadwyler, S. A. (2018). Developing a hippocampal neural prosthetic to facilitate human memory encoding and recall. *Journal of Neural Engineering*, *15*(3), 036014. https://doi.org/10.1088/1741-2552/aaaed7

Hannan, M., & Carey, M. (2004). Ambient Soundscapes in Blade Runner. In P. Hayward (Ed.), *Off the Planet: Music, Sound and Science Fiction Cinema* (pp. 149–164). John Libbey Publishing. https://doi.org/10.2307/j.ctt2005s0z

Hänsler, E., & Schmidt, G. (Eds.). (2008). *Speech and audio processing in adverse environments*. Springer-Verlag.

Hardesty, L. (2022, January 31). The engineering behind Alexa's contextual speech recognition. *Amazon Science*. https://www.amazon.science/latest-news/the-engineering-behind-alexas-contextual-speech-recognition

Harper, A. (2014). *Lo-fi Aesthetics In Popular Music Discourse* [Doctoral, Wadham College, University of Oxford].

Harris, M. S., Capretta, N. R., Henning, S. C., Feeney, L., Pitt, M. A., & Moberly, A. C. (2016). Postoperative Rehabilitation Strategies Used by Adults With Cochlear Implants: A Pilot Study. *Laryngoscope Investigative Otolaryngology*, *1*(3), 42–48. https://doi.org/10.1002/lio2.20

Hayes, K. T. (2019, October 16). 'Just living my moment': Girl, 8, dressed as Princess Tiana dances with Disney band in viral video. *FOX TV Digital Team*. https://www.ktvu.com/news/just-living-my-moment-girl-8-dressed-as-princess-tiana-dances-with-disney-band-in-viral-video

Heim, M. R. (2013). The Paradox of Virtuality. In M. Grimshaw-Aagaard (Ed.), *The Oxford Handbook of Virtuality* (pp. 111–126). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.002

Hershey, J. R., Chen, Z., Roux, J. L., & Watanabe, S. (2015). *Deep clustering: Discriminative embeddings for segmentation and separation* (arXiv:1508.04306). arXiv. https://doi.org/10.48550/arXiv.1508.04306

Hill, A. J., & Hawksford, M. O. J. (2010). A hybrid virtual bass system for optimized steady-state and transient performance. *2010 2nd Computer Science and Electronic Engineering Conference (CEEC)*, 5606489. https://doi.org/10.1109/CEEC.2010.5606489

Hoevers, M. (2021). *The potential of Augmented Reality Glasses in the Practice of Street Policing: An Experimental Study among Dutch (Mounted) Police Officers* [Masters, University of Twente].

Holman, T. (2010). *Sound for film and television* (3rd ed). Focal Press.

Hong, J., He, J., Lam, B., Gupta, R., & Gan, W.-S. (2017). Spatial Audio for Soundscape Design: Recording and Reproduction. *Applied Sciences*, *7*(6), 627. https://doi.org/10.3390/app7060627

Houser, K. (2018, July 22). This VR horror game is exactly as scary as your body can handle. *Futurism*. https://futurism.com/the-byte/vr-horror-game-heart-rate-monitor

Howiie, H. (2022, February 14). *The Ever Increasing Impact of Voice Chat in Games*. https://www.youtube.com/watch?v=McJBdTOzOSY

Hubbard, G. T., & Crawford, E. C. (2008). Music in Political Advertisements: Music to the Ears or Background Noise? A Study of Music's Influence on Message-relevant Thinking. Journal of Radio & Audio Media, 15(2), 167–181. https://doi.org/10.1080/19376520802397276

Huvenne, M. (2016). Intertwining Music and Sound in Film. In L. Greene & D. Kulezic-Wilson (Eds.), *The Palgrave Handbook of Sound Design and Music in Screen Media: Integrated Soundtracks* (pp. 123–137). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51680-0_9

Hughes, O., Fuchs, P., & Nannipieri, O. (2011). New Augmented Reality Taxonomy: Technologies and Features of Augmented Environment. In B. Furht (Ed.), *Handbook of Augmented Reality* (pp. 47–63). Springer New York. https://doi.org/10.1007/978-1-4614-0064-6

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., & Boisvert, I. (2018). Social Connectedness and Perceived Listening Effort in Adult Cochlear Implant Users: A Grounded Theory to Establish Content Validity for a New Patient-Reported Outcome Measure. *Ear and Hearing*, *39*(5), 922. https://doi.org/10.1097/AUD.0000000000000553

Huizinga, J. (1938). *Homo ludens: A study of the play-element in culture*. Beacon Press.

Hull, L. (2019, March 20). Vicar's plea to stop his own church bells from chiming infuriates town. *Mail Online*. https://www.dailymail.co.uk/news/article-6832595/Vicars-plea-stop-church-bells-chiming-night-infuriates-town.html

IGN. (2020, March 28). *Ape Out Developers React to 26 Minute Speedrun*. https://www.youtube.com/watch?v=Ek7dSIqrZlY&t=742s

Ireland, D. (2017). Great Expectations? The Changing Role of Audiovisual Incongruence in Contemporary Multimedia. *Music and the Moving Image*, *10*(3), 21–35. https://doi.org/10.5406/musimoviimag.10.3.0021

ITV News. (2021, August 18). The virtual reality training hoping to improve hate crime responses. *ITV News*. https://www.itv.com/news/granada/2021-08-18/gmp-officers-to-virtual-reality-training-around-hate-crime

Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu Navigation With In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload. *International Journal of Human–Computer Interaction*, *31*(1), 1–16. https://doi.org/10.1080/10447318.2014.925774

Jeyakumar, V., Krishnan, P. T., Sundaram, P., & Raj, A. N. J. (2022). Chapter 9—Brain-computer interface in Internet of Things environment. In A. K. Bhoi, V. H. C. de Albuquerque, S. N. Sur, & P. Barsocchi (Eds.), *5G IoT and Edge Computing for Smart Healthcare* (pp. 231–255). Academic Press. https://doi.org/10.1016/B978-0-323-90548-0.00012-7

Jiang, J. A., Kiene, C., Middler, S., Brubaker, J. R., & Fiesler, C. (2019). Moderation Challenges in Voice-based Online Communities on Discord. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), Article 55. https://doi.org/10.1145/3359157

Jingen Liang, L., & Elliot, S. (2021). A systematic review of augmented reality tourism research: What is now and what is next? *Tourism and Hospitality Research*, *21*(1), 15–30. https://doi.org/10.1177/1467358420941913

Jørgensen, K. (2007). On transdiegetic sounds in computer games. *Northern Lights: Film and Media Studies Yearbook*, *5*(1), 105–117. https://doi.org/10.1386/nl.5.1.105_1

Jørgensen, K. (2008). Audio and Gameplay: An Analysis of PvP Battlegrounds in World of Warcraft. *Game Studies*, *8*(2). http://gamestudies.org/0802/articles/jorgensen

Julia, A. (2018, November 20). YouTube will stream The Lego Movie free as an ad for the sequel. *The Verge*. https://www.theverge.com/2018/11/20/18105471/lego-movie-2-youtube-free-black-friday-sequel-second-part

Jung, C. (1921). *Psychological Types* (1st edition). Routledge.

Kalinak, K. (1992). *Settling the Score: Music and the Classical Hollywood Film*. University of Wisconsin Press.

Kamp, M. (2021). Autoethnography, Phenomenology and Hermeneutics. In M. Fritsch & T. Summers (Eds.), *The Cambridge Companion to Video Game Music* (pp. 159–175). Cambridge University Press. https://doi.org/10.1017/9781108670289.012

Kamp, M. (2021). Old(er) Media and New Musical Affordances in Virtual Reality Experiences. In C. Cenciarelli (Ed.), *The Oxford Handbook of Cinematic Listening* (1st ed., pp. C35–736). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190853617.013.37

Kassabian, A. (2013). *Ubiquitous Listening: Affect, Attention, and Distributed Subjectivity*. University of California Press.

Kidd, M. A. (2016). Archetypes, Stereotypes and Media Representation in a Multi-cultural Society. *Procedia - Social and Behavioral Sciences*, *236*, 25–28. https://doi.org/10.1016/j.sbspro.2016.12.007

Kilteni, K., Groten, R., & Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Kim, J.-K. (2021, December 22). Reducing unnecessary clarification questions from voice agents. *Amazon Science*. https://www.amazon.science/blog/reducing-unnecessary-clarification-questions-from-voice-agents

Kirsch, P. (1990). *Galleon: The Great Ships of the Armada Era* (1st UK edition). Naval Inst Pr.

Kishino, F., & Milgram, P. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, *E77-D*(12).

Klanjšček, J., Geček, S., Marn, N., Legović, T., & Klanjšček, T. (2018). Predicting perceived level of disturbance of visitors due to crowding in protected areas. *PLoS ONE*, *13*(6), e0197932. https://doi.org/10.1371/journal.pone.0197932

Klasen, M., Weber, R., Kircher, T. T. J., Mathiak, K. A., & Mathiak, K. (2012). Neural contributions to flow experience during video game playing. *Social Cognitive and Affective Neuroscience*, *7*(4), 485–495. https://doi.org/10.1093/scan/nsr021

Kompis, M., Vibert, D., Vischer, M. W., Senn, P., & Häusler, R. (2003). Scuba Diving with Cochlear Implants. *Annals of Otology, Rhinology & Laryngology*, *112*(5), 425–427. https://doi.org/10.1177/000348940311200507

Kosmyna, N., Tarpin-Bernard, F., Bonnefond, N., & Rivet, B. (2016). Feasibility of BCI Control in a Realistic Smart Home Environment. *Frontiers in Human Neuroscience*, *10*, 416. https://doi.org/10.3389/fnhum.2016.00416

Kou, Y. (2020). Toxic Behaviors in Team-Based Competitive Gaming: The Case of League of Legends. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 81–92. https://doi.org/10.1145/3410404.3414243

Kremenska, A. G., Lekova, A. K., & Dimitrov, G. P. (2022). EEG Brain-Computer Interfaces for Internet of Everything (IoE). *2022 International Conference on Information Technologies (InfoTech)*, 1–6. https://doi.org/10.1109/InfoTech55606.2022.9897097

Krisp. (2022). *Background Noise Cancellation & Reduction Software*. Krisp. https://www.krisp.ai/

Kulezic-Wilson, D. (2015). *The Musicality of Narrative Film*. Palgrave Macmillan UK. https://doi.org/10.1057/9781137489999

Kurz, A., Zanzinger, M., Hagen, R., & Rak, K. (2021). The impact of cochlear implant microphone settings on the binaural hearing of experienced cochlear implant users with single-sided deafness. *European Archives of Oto-Rhino-Laryngology*, *278*(6), 2067–2077. https://doi.org/10.1007/s00405-020-06450-5

Kytö, M. (2023). Soundscapes of Code. In J. L. Drever & A. Hugill, *Aural Diversity* (1st ed., pp. 73–81). Routledge. https://doi.org/10.4324/9781003183624-8

Lamb, B., & Smith, B. (2018). From Skyrim to Skellige: Fantasy Video Game Music Within a Neo-Mediaevalist Paradigm. *Musicology Australia*, *40*(2), 79–100. https://doi.org/10.1080/08145857.2018.1550140

Langston, P. S. (1988). *Six Techniques for Algorithmic Composition*.

Larsson, P., Väljamäe, A., Västfjäll, D., Tajadura-Jiménez, A., & Kleiner, M. (2010). Auditory-Induced Presence in Mixed Reality Environments and Related Technology. In E. Dubois, P. Gray, & L. Nigay (Eds.), *The Engineering of Mixed Reality Systems* (pp. 143–163). Springer London. https://doi.org/10.1007/978-1-84882-733-2_8

Latoschik, M. E., & Wienrich, C. (2022). Congruence and Plausibility, not Presence?! Pivotal Conditions for XR Experiences and Effects, a Novel Model. *Frontiers in Virtual Reality*, *3*, 694433. https://doi.org/10.3389/frvir.2022.694433

Leith, A. P. (2021). Parasocial cues: The ubiquity of parasocial relationships on Twitch. Communication Monographs, 88(1), 111–129. https://doi.org/10.1080/03637751.2020.1868544

Lemaitre, G., Susini, P., Winsberg, S., McAdams, S., & Letinturier, B. (2009). The Sound Quality of Car Horns: Designing New Representative Sounds. *Acta Acustica United with Acustica*, *95*(2), 356–372. https://doi.org/10.3813/AAA.918158

Lentz, T. (2008). Binaural technology for virtual reality. *The Journal of the Acoustical Society of America*, *124*(6), 3359–3359. https://doi.org/10.1121/1.3020604

Leydon, R. (2004). Forbidden Planet: Effects and Affects in the Electro Avant-garde. In P. Hayward (Ed.), *Off the Planet: Music, Sound and Science Fiction Cinema* (pp. 61–76). John Libbey Publishing. https://doi.org/10.2307/j.ctt2005s0z

Lim, W. M. (2021). Toward an agency and reactance theory of crowding: Insights from COVID-19 and the tourism industry. *Journal of Consumer Behaviour*, 10.1002/cb.1948. https://doi.org/10.1002/cb.1948

Lind, S. (2020). Music as temporal disruption in Assassin's Creed. *The Soundtrack*, *11*(1–2), 57–73. https://doi.org/10.1386/ts_00005_1

Lind, S. (2022). *Authenticity in the Music of Video Games*. Lexington Books.

Lobley, N. (2014). Sound Galleries: Curating the Experience of Sound and Music In and Beyond Museums. *Music in Art*, *39*(1–2), 243–255.

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

London, J. (2000). Leitmotifs and Musical Reference in the Classical Film Score. In J. Buhler, C. Flinn, & D. Neumeyer (Eds.), *Music and Cinema* (pp. 85–96). Wesleyan University Press.

Lucas, J.-F., Cornish, T., & Margolis, T. (2012). To a cultural perspective of mixed reality events: A case study of event overflow in operas and concerts in mixed reality. *New Review of Hypermedia and Multimedia*, *18*(4), 277–293. https://doi.org/10.1080/13614568.2012.746741

Luizard, P., Brauer, E., Weinzierl, S., & Henrich Bernardoni, N. (2018, October). How singers adapt to room acoustical conditions. *Auditorium Acoustics 2018*. https://hal.archives-ouvertes.fr/hal-01969126

LX News. (2020, May 4). *'Jurassic Park' Probably Didn't Get Dinosaur Sounds Right.* https://www.youtube.com/watch?v=9qJA91rRUqE

Machado, R. P., Conforto, D., & Santarosa, L. (2017). Sound chat: Implementation of sound awareness elements for visually impaired users in web-based cooperative systems. *2017 International Symposium on Computers in Education (SIIE)*, 1–6. https://doi.org/10.1109/SIIE.2017.8259677

Mann, D. (2012). Reinstating Shakespeare's Instrumental Music. *Early Theatre*, *15*(2), 67–91.

Manning, R., Valliere, W., & Wang, B. (1999). Crowding Norms: Alternative Measurement Approaches. *Leisure Sciences*, *21*(2), 97–115. https://doi.org/10.1080/014904099273174

Manning, L. (1933). The Man Who Awoke. *Wonder Stories*.

Marler, T., Straus, S. G., Mizel, M. L., Hollywood, J. S., Harrison, B., Yeung, D., Lewis, M. W., Rizzo, S., Hartholt, A., & Swain, C. (2020). Effective Game-Based Training for Police Officer Decision-Making: Linking Missions, Skills, and Virtual Content. Article 20546.

Martindale, J. (2017, May 28). Farpoint's Composer On What VR Game Audio Is All About. *Digital Trends*. https://www.digitaltrends.com/gaming/stephen-cox-farpoint-interview/

Martínez-Cagigal, V., Santamaría-Vázquez, E., & Hornero, R. (2018). Controlling a Smartphone with Brain-Computer Interfaces: A Preliminary Study. In F. J. Perales & J. Kittler (Eds.), *Articulated Motion and Deformable Objects* (pp. 34–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-94544-6_4

Massie, G. (2021, July 2). Police officer plays Taylor Swift song to prevent protest video going on YouTube. *The Independent*. https://www.independent.co.uk/news/world/americas/police-taylor-swift-protest-video-b1877311.html

Massumi, B. (2002). *Parables for the Virtual: Movement, Affect, Sensation*. Duke University Press.

Massumi, B. (2013). Envisioning the Virtual. In M. Grimshaw-Aagaard (Ed.), *The Oxford Handbook of Virtuality* (pp. 55–70). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.010

Mata, F., & Claramunt, C. (2014). A social navigation guide using augmented reality. *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 541–544. https://doi.org/10.1145/2666310.2666364

Mauger, V. (2014). Interface. In *The Routledge Companion to Video Game Studies* (pp. 32–40).

McArthur, A., Sandler, M., & Stewart, R. (2017). Distance in audio for VR: Constraints and opportunities. *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, Article 28, 1–7. https://doi.org/10.1145/3123514.3123530

McArthur, A., Stewart, R., & Sandler, M. (2017). Sounds too true to be good: Diegetic infidelity – the case for sound in virtual reality. *Journal of Media Practice*, *18*(1), 26–40. https://doi.org/10.1080/14682753.2017.1305840

McLean, L., & Griffiths, M. D. (2019). Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study. *International Journal of Mental Health and Addiction*, *17*(4), 970–994. https://doi.org/10.1007/s11469-018-9962-0

Medina-Gray, E. (2014). *Modular Structure and Function in Early 21st-Century Video Game Music* [Doctoral, Yale University].

Mehta, R., Zhu, R. (Juliet), & Cheema, A. (2012). Is Noise Always Bad? Exploring the Effects of Ambient Noise on Creative Cognition. *Journal of Consumer Research*, *39*(4), 784–799. https://doi.org/10.1086/665048

Meijer, F. (2014). *A History of Seafaring in the Classical World (Routledge Revivals)*. Routledge. https://doi.org/10.4324/9781315779980

Mendonça, C., & Korshunova, V. (2020). Surround sound spreads visual attention and increases cognitive effort in immersive media reproductions. *Proceedings of the 15th International Audio Mostly Conference*, 16–21. https://doi.org/10.1145/3411109.3411118

Metz, C. (1990). *Film Language: A Semiotics of the Cinema*. University Of Chicago Press.

Mevea. (2019, October 30). *Multiphysics simulator training improving crane operator productivity*. https://www.youtube.com/watch?v=XWicigm1IKA

Michaels, S. (2008, June 19). First recorded computer music unveiled. *The Guardian*. https://www.theguardian.com/music/2008/jun/19/news.culture2

Michailidis, L., Balaguer-Ballester, E., & He, X. (2018). Flow and Immersion in Video Games: The Aftermath of a Conceptual Challenge. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01682

Microsoft. (2018a, September 18). *Behind the scenes of Unmoored with artist Mel Chin*. https://www.youtube.com/watch?v=0hPP0cHHubM

Microsoft. (2018b, October 9). *Dr. Mae Jemison and Defying Gravity: Women in Space*. https://www.youtube.com/watch?v=kkIVhkHsV6I

Microsoft. (2018c, November 12). *This rare 17th century technology was brought to life using HoloLens at the Musée des Plans-Reliefs*. https://www.youtube.com/watch?v=MGtpe0ni_nI

Microsoft. (2019, May 17). *"Hollywood Dream Machines" brings iconic cars to life | Microsoft In Culture*. https://www.youtube.com/watch?v=b_8u5ptgcaI

Microsoft. (2022). *HoloLens environment considerations*. https://docs.microsoft.com/en-us/hololens/hololens-environment-considerations

Microsoft HoloLens. (2019, February 24). *Introducing Microsoft HoloLens 2*. https://www.youtube.com/watch?v=eqFqtAJMtYE

Milman, A. (2019). Visitor management in highly-visited attractions: Lessons that practitioners can learn from the US theme park industry and the national parks. In Pechlaner, H., Innerhofer, E., Erschbamer, G., (Eds.), *Overtourism* (pp. 104–124). Routledge.

Milman, A., Tasci, A. D. A., & Wei, W. (200). Crowded and popular: The two sides of the coin affecting theme-park experience, satisfaction, and loyalty. *Journal of Destination Marketing & Management*, *18*, 100468. https://doi.org/10.1016/j.jdmm.2020.100468

Mine, M. R., van Baar, J., Grundhofer, A., Rose, D., & Yang, B. (2012). Projection-Based Augmented Reality in Disney Theme Parks. *Computer*, *45*(7), 32–40. https://doi.org/10.1109/MC.2012.154

Møller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, *36*(3–4), 171–218. https://doi.org/10.1016/0003-682X(92)90046-U

Moller, H., & Pedersen, C. S. (2004). Hearing at low and infrasonic frequencies. *Noise and Health*, *6*(23), 37–57.

Montague, W. E., & Strickland, J. F. (1961). Sensitivity of the Water-Immersed Ear to High- and Low-Level Tones. *The Journal of the Acoustical Society of America*, *33*(10), 1376–1381. https://doi.org/10.1121/1.1908443

Moore, A. (1980). Walt Disney World: Bounded Ritual Space and the Playful Pilgrimage Center. *Anthropological Quarterly*, *53*(4), 207–218. https://doi.org/10.2307/3318104

Moore, A. (2002). Authenticity as Authentication. *Popular Music*, *21*(2), 209–223.

Morrison, J. S., & Coates, J. F. (1986). *The Athenian trireme: The history and reconstruction of an ancient Greek warship*. Cambridge University Press.

Mu, H., Gan, W.-S., & Tan, E.-L. (2012). A psychoacoustic bass enhancement system with improved transient and steady-state performance. *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 141–144. https://doi.org/10.1109/ICASSP.2012.6287837

Murphy, J. W. (1996). There Is Nothing Virtual About Virtual Reality. *ETC: A Review of General Semantics*, *53*(4), 458–463.

Murphy, S. C. (2014). Controllers. In *The Routledge Companion to Video Game Studies* (pp. 19–24).

Murray, J. H. (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace* (updated edition). MIT Press.

Nachmani, E., Adi, Y., & Wolf, L. (2020). *Voice Separation with an Unknown Number of Multiple Speakers* (arXiv:2003.01531). arXiv. https://doi.org/10.48550/arXiv.2003.01531

Nadri, C., Chan Lee, S., Kekal, S., Li, Y., Li, X., Lautala, P., Nelson, D., & Jeon, M. (2021). Effects of Auditory Display Types and Acoustic Variables on Subjective Driver Assessment in a Rail Crossing Context. *Transportation Research Record*, *2675*(9), 1457–1468. https://doi.org/10.1177/03611981211007838

Nadri, C., Ko, S., Diggs, C., Winters, M., Sreehari, V. K., & Jeon, M. (2021). Novel Auditory Displays in Highly Automated Vehicles: Sonification Improves Driver Situation Awareness, Perceived Workload, and Overall Experience. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 586–590. https://doi.org/10.1177/1071181321651071

NASA Jet Propulsion Laboratory. (2016, March 30). *Mixed-Reality Tech Brings Mars to Earth*. https://www.youtube.com/watch?v=wPOCcG33mJQ

Nast, C. (2016). Inside the infinite world of No Man's Sky and its evolving soundtrack. *Wired UK*. https://www.wired.co.uk/article/no-mans-sky-65daysofstatic-interview

Nee, A. Y. C., Ong, S. K., Chryssolouris, G., & Mourtzis, D. (2012). Augmented reality applications in design and manufacturing. *CIRP Annals*, *61*(2), 657–679. https://doi.org/10.1016/j.cirp.2012.05.010

Neumann, F. (1967). The Use of Baroque Treatises on Musical Performance. *Music & Letters*, *48*(4), 315–324.

Norman, D. A., Phelps, R., & Wightman, F. (1971). Some observations on underwater hearing. *The Journal of the Acoustical Society of America*, *50*(2), 544–548. https://doi.org/10.1121/1.1912669

Nowak, P., Zimpfer, V., & Zolzer, U. (2018). 3D Virtual Audio with Headphones: A Literature Review of the Last Ten Years. Fortschritte der Akustik – DAGA München.

Nye, L. (2021). The US military is using VR to simulate combat jumps. *Business Insider*. https://www.businessinsider.com/us-military-using-new-virtual-reality-trainer-simulate-combat-jumps-2017-5

onstageDisney. (2020a, June 13). *Music Loop—Disney's Magic Kingdom Main Entrance*. https://www.youtube.com/watch?v=kDUdRQ96AUY

onstageDisney. (2020b, June 29). *Music & Video Loop—Disneyland Today*. https://www.youtube.com/watch?v=1uXE3DJo5Oc

onstageDisney. (2020c, July 5). *Music Loop—Disneyland and California Adventure Esplanade*. https://www.youtube.com/watch?v=9E-l9qYiqxQ

Orozco, M., Silva, J., El Saddik, A., & Petriu, E. (2012). The Role of Haptics in Games. In A. El Saddik (Ed.), Haptics Rendering and Applications (pp. 217–234). InTech. https://doi.org/10.5772/32809

Ortiz, S. M. (2019). 'You Can Say I Got Desensitized to It': How Men of Color Cope with Everyday Racism in Online Gaming. *Sociological Perspectives*, *62*(4), 572–588.

Ostanin, M., R. Yagfarov, D. Devitt, A. Akhmetzyanov, and A. Klimchik. 2020. "Multi Robots Interactive Control Using Mixed Reality." International Journal of Production Research 59 (23): 7126–7138. https://doi.org/10.1080/00207543.2020.1834640

Park, A., & Mac, R. (2015). "I Don't Want to Hear It": Woman Fights to Stop Nearby Church Bells. *NBC Los Angeles*. https://www.nbclosangeles.com/news/woman-fights-to-stop-nearby-church-bells-2/114697/

Parsons, D. (2019, August 1). Full Flight Simulators Incorporate VR for Next Generation of Pilots. *Aviation Today*. https://www.aviationtoday.com/2019/08/01/training-brain-mind/

Parvin, S. J., & Nedwell, J. R. (1995). Underwater Sound Perception and the Development of an Underwater Noise Weighting Scale. *Underwater Technology*, *21*(1), 12–19. https://doi.org/10.3723/175605495783328836

Pasnau, R. (1999). What Is Sound? *The Philosophical Quarterly (1950-)*, *49*(196), 309–324.

Paulin, S. D. (2000). Richard Wagner and the Fantasy of Cinematic Unity—The Idea of the Gesamtkunstwerk in the History and Theory of Film Music. In J. Buhler, C. Flinn, & D. Neumeyer (Eds.), *Music and Cinema* (pp. 58–84). Wesleyan University Press.

Paulus, I. (2000). Williams versus Wagner or an Attempt at Linking Musical Epics. *International Review of the Aesthetics and Sociology of Music*, *31*(2), 153–184.

Pausch, R., Snoddy, J., Taylor, R., Watson, S., & Haseltine, E. (1996). Disney's Aladdin: First steps toward storytelling in virtual reality. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '96*, 193–203. https://doi.org/10.1145/237170.237257

Pence, H. E. (2010). Smartphones, Smart Objects, and Augmented Reality. *The Reference Librarian*, *52*(1–2), 136–145. https://doi.org/10.1080/02763877.2011.528281

Pereira, F., da Silva, E. A. B., & Lafruit, G. (2018). Chapter 2 - Plenoptic imaging: Representation and processing. In R. Chellappa & S. Theodoridis (Eds.), *Academic Press Library in Signal Processing, Volume 6* (pp. 75–111). Academic Press. https://doi.org/10.1016/B978-0-12-811889-4.00002-6

Petherick, S. (2021, June 10). Village church bells silenced after single complaint they are 'too noisy'. *Metro*. https://metro.co.uk/2021/06/10/church-bells-silenced-after-single-complaint-they-are-too-noisy-14751731/

Phillips, W. (2018a, April 16). Composing video game music for Virtual Reality: The role of music in VR. *Game Developer*. https://www.gamedeveloper.com/audio/composing-video-game-music-for-virtual-reality-the-role-of-music-in-vr

Phillips, W. (2018b, May 15). Composing video game music for Virtual Reality: 3D versus 2D. *Game Developer*. https://www.gamedeveloper.com/audio/composing-video-game-music-for-virtual-reality-3d-versus-2d

Phillips, W. (2018c, June 14). Composing video game music for Virtual Reality: Diegetic versus Non-diegetic. *Game Developer*. https://www.gamedeveloper.com/audio/composing-video-game-music-for-virtual-reality-diegetic-versus-non-diegetic

Pieslak, J. R. (2007). Sound Targets: Music and the War in Iraq. *Journal of Musicological Research*, *26*(2–3), 123–149. https://doi.org/10.1080/01411890701360153

Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., Gage, S. H., & Pieretti, N. (2011). Soundscape Ecology: The Science of Sound in the Landscape. *BioScience, 61*(3), 203–216. https://doi.org/10.1525/bio.2011.61.3.6

Pike, C. W. (2019). *Evaluating the Perceived Quality of Binaural Technology* [Doctoral Thesis]. University of York.

Pimentel, K., & Teixeira, K. (1993). *Virtual reality: Through the new looking glass*. Intel/Windcrest.

Plank, D. (2021). Audio and the Experience of Gaming: A Cognitive-Emotional Approach to Video Game Sound. In M. Fritsch & T. Summers (Eds.), *The Cambridge Companion to Video Game Music* (pp. 284–301). Cambridge University Press. https://doi.org/10.1017/9781108670289.018

Pohlmann, E., & West, M. L. (2001). *Documents of Ancient Greek Music: The Extant Melodies and Fragments edited and transcribed with commentary*. Oxford University Press.

Positron. (2023). The Technology. *Positron*. https://gopositron.com/the-technology/

Potamkin, H. A. (1929). Music and the Movies. *The Musical Quarterly*, *XV*(2), 281–296. https://doi.org/10.1093/mq/XV.2.281

Prendergast, R. M. (1992). *Film music: A neglected art: a critical study of music in films* (2nd ed). W.W. Norton.

Presser, H. (2021). *The constituent aspects of synchresis: A practice-led examination of audio-visual synchronization in key sites of production* [Doctor of Philosophy]. Swinburne University of Technology.

Prior, H. M. (2022). How Can Music Help Us to Address the Climate Crisis? Music & Science, 5, 20592043221075724. https://doi.org/10.1177/20592043221075725

Prpa, M., Fdili-Alaoui, S., Schiphorst, T., & Pasquier, P. (2020). Articulating Experience: Reflections from Experts Applying Micro-Phenomenology to Design Research in HCI. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376664

Rackspace. (2021, March 30). The First Ever IoT-enabled Firefighter Helmet Provides 'Superpowers' Through Augmented Reality. *Rackspace Technology*. https://www.rackspace.com/en-gb/newsroom/first-ever-iot-enabled-firefighter-helmet-provides-superpowers-through-augmented-reality

Raghuvanshi, N., Snyder, J., Mehra, R., Lin, M., & Govindaraju, N. (2010). Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes. *ACM Transactions on Graphics*, *29*(4), 68:1-68:11. https://doi.org/10.1145/1778765.1778805

Rao, R. P. N. (2013). Brain-computer interfacing: An introduction. Cambridge University Press.

reddit. (2020). *Can you use own headphones with valve index?* Reddit. https://www.reddit.com/r/ValveIndex/comments/cc9cwm/can_you_use_own_headphones_with_valve_index/

Renel, W. (2023). The Auditory Normate. In J. L. Drever & A. Hugill, *Aural Diversity* (1st ed., pp. 55–63). Routledge. https://doi.org/10.4324/9781003183624-6

Riede, T., Eliason, C. M., Miller, E. H., Goller, F., & Clarke, J. A. (2016). Coos, booms, and hoots: The evolution of closed-mouth vocal behavior in birds. *Evolution*, *70*(8), 1734–1746. https://doi.org/10.1111/evo.12988

Rihl, A., & Wegener, C. (2019). YouTube celebrities and parasocial interaction: Using feedback channels in mediatized relationships. Convergence, 25(3), 554–566. https://doi.org/10.1177/1354856517736976

RNID. (n.d.). *Listen to music safely*. Retrieved 9 May 2022, from https://rnid.org.uk/information-and-support/ear-health/protect-your-hearing/listen-to-music-safely/

Rohani Ghahari, R., George-Palilonis, J., & Bolchini, D. (2013). Mobile Web Browsing with Aural Flows: An Exploratory Study. *International Journal of Human-Computer Interaction*, *29*(11), 717–742. https://doi.org/10.1080/10447318.2013.773875

Rosenblum, L. D., Wuestefeld, A. P., & Anderson, K. L. (1996). Auditory reachability: An affordance approach to the perception of sound source distance. *Ecological Psychology*, 1–24.

Rossetti, G. (Forthcoming). Mandatory Metal Moments: The Ubiquity of Heavy Metal Culture in Video Games. Journal of Sound and Music in Games, 4(2).

Rummukainen, O. S., Schlecht, S. J., Robotham, T., Plinge, A., & Habets, E. A. P. (2019). Perceptual Study of Near-Field Binaural Audio Rendering in Six-Degrees-of-Freedom Virtual Reality. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 448–454. https://doi.org/10.1109/VR.2019.8798177

Rumsey, F. (2011). Whose Head Is it Anyway? Optimizing Binaural Audio. *Journal of the Audio Engineering Society*, *59*(9), 672–675.

SA Police News. (2017, July 19). *SA Police Firearms Training Simulator*. https://www.youtube.com/watch?v=0tls_KdUdw4

Salaman, C. K. (1879). Classical Music. *The Musical Times and Singing Class Circular*, *20*(434), 200–203. https://doi.org/10.2307/3355606

Salen, K., & Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. MIT Press.

Schafer, R. M. (1977a). *Five village soundscapes*. A.R.C. Publications.

Schafer, R. M. (1977b). *Soundscape: Our Sonic Environment and the Tuning of the World*.

Schmidt, S. A., Carpenter-Thompson, J., & Husain, F. T. (2017). Connectivity of precuneus to the default mode and dorsal attention networks: A possible invariant marker of long-term tinnitus. *NeuroImage : Clinical*, *16*, 196–204. https://doi.org/10.1016/j.nicl.2017.07.015

scientia.global. (2017, October 8). Chirping Communication – Sending Data Over Sound. *Scientia.Global*. https://www.scientia.global/chirping-communication-sending-data-sound/

SciShow. (2017, November 14). *What Does A Cochlear Implant Sound Like?* https://www.youtube.com/watch?v=lzgQrHFDNLE

Shen, S., Roy, N., Guan, J., Hassanieh, H., & Choudhury, R. R. (2018). MUTE: Bringing IoT to noise cancellation. *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 282–296. https://doi.org/10.1145/3230543.3230550

Sherrick, B., Smith, C., Jia, Y., Thomas, B., & Franklin, S. B. (2023). How Parasocial Phenomena Contribute to Sense of Community on Twitch. Journal of Broadcasting & Electronic Media, 67(1), 47–67. https://doi.org/10.1080/08838151.2022.2151599

Shields, R. (2002). *The Virtual* (1st edition). Routledge.

Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-Computer Interfaces in Medicine. *Mayo Clinic Proceedings*, *87*(3), 268–279. https://doi.org/10.1016/j.mayocp.2011.12.008

Shih, N.-J., Diao, P.-H., & Chen, Y. (2019). ARTS, an AR Tourism System, for the Integration of 3D Scanning and Smartphone AR in Cultural Heritage Tourism and Pedagogy. *Sensors (Basel, Switzerland)*, *19*(17), 3725. https://doi.org/10.3390/s19173725

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' flow experience while browsing a Web site: Its measurement, contributing factors and consequences. *Computers in Human Behavior*, *20*(3), 403–422. https://doi.org/10.1016/S0747-5632(03)00050-5

Skarbez, R., Neyret, S., Brooks, F. P., Slater, M., & Whitton, M. C. (2017). A Psychophysical Experiment Regarding Components of the Plausibility Illusion. *IEEE Transactions on Visualization and Computer Graphics*, *23*(4), 1369–1378. https://doi.org/10.1109/TVCG.2017.2657158

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., & Oliva, R. (2022). A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality*, *3*, 914392. https://doi.org/10.3389/frvir.2022.914392

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Smith, M. M. (2015). Sound—So What? *The Public Historian*, *37*(4), 132–144. https://doi.org/10.1525/tph.2015.37.4.132

Smith, R. C. G., & Price, S. R. (2014). Modelling of Human Low Frequency Sound Localization Acuity Demonstrates Dominance of Spatial Variation of Interaural Time Difference and Suggests Uniform Just-Noticeable Differences in Interaural Time Difference. *PLoS ONE*, *9*(2), e89033. https://doi.org/10.1371/journal.pone.0089033

Spaccarotella, C. A. M., Polimeni, A., Migliarino, S., Principe, E., Curcio, A., Mongiardo, A., Sorrentino, S., De Rosa, S., & Indolfi, C. (2020). Multichannel Electrocardiograms Obtained by a Smartwatch for the Diagnosis of ST-Segment Changes. *JAMA Cardiology*, *5*(10), 1176–1180. https://doi.org/10.1001/jamacardio.2020.3994

Spitfire Audio. (2019, December 1). *Behind The Soundtrack of Assassin's Creed Odyssey*. https://www.youtube.com/watch?v=-X9PTMsg33M

Spring, M. (2023). Thomas Mace. In J. L. Drever & A. Hugill, *Aural Diversity* (1st ed., pp. 168–177). Routledge. https://doi.org/10.4324/9781003183624-19

Stam, R., Burgoyne, R., & Flitterman-Lewis, S. (1992). *New Vocabularies in Film Semiotics: Structuralism, Post-structuralism, and Beyond*. Psychology Press.

Stanney, K. M., Hale, K. S., & Zyda, M. (2015). Virtual Environments in the Twenty-First Century. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (Second edition, pp. 3–22). CRC Press, Taylor & Francis Group.

Statista. (2020). *UK: Mobile phone ownership 1996–2018*. Statista. https://www.statista.com/statistics/289167/mobile-phone-penetration-in-the-uk/

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Stevens, J. E. (1995). The Growing Reality of Virtual Reality. *BioScience*, *45*(7), 435–439. https://doi.org/10.2307/1312785

Stevens, R. (2021). The Inherent Conflicts of Musical Interactivity in Video Games. In M. Fritsch & T. Summers (Eds.), *The Cambridge Companion to Video Game Music* (pp. 74–93). Cambridge University Press. https://doi.org/10.1017/9781108670289.007

Stevens, R., & Raybould, D. (2011). *The Game Audio Tutorial: A Practical Guide to Sound and Music for Interactive Games* (1st edition). Routledge.

Stevens, R., & Raybould, D. (2015). The reality paradox: Authenticity, fidelity and the real in Battlefield 4. *The Soundtrack*, *8*(1–2), 57–75. https://doi.org/10.1386/st.8.1-2.57_1

Stevenson, J. (2003). *Dogme uncut: Lars von Trier, Thomas Vinterburg, and the gang that took on Hollywood*. Santa Monica, CA : Santa Monica Press.

Stokes, A. F., & Wickens, C. D. (1988). 12—Aviation Displays. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 387–431). Academic Press. https://doi.org/10.1016/B978-0-08-057090-7.50018-7

Strickland, E. (2018, July 17). This Smart Watch Detects Cardiac Arrest, and Summons Help. *IEEE Spectrum*. https://spectrum.ieee.org/this-smart-watch-detects-cardiac-arrest-and-summons-help

Sü Gül, Z. (2019). Acoustical Impact of Architectonics and Material Features in the Lifespan of Two Monumental Sacred Structures. *Acoustics*, *1*(3), Article 3, 493–516. https://doi.org/10.3390/acoustics1030028

Sumikawa, D. A. (1985). *Guidelines for the integration of audio cues into computer user interfaces* (UCRL-53656). Lawrence Livermore National Lab., CA (USA). https://www.osti.gov/biblio/5475406

Summers, T. (2016). *Understanding Video Game Music*. Cambridge University Press.

Summers, T., Cook, J., Famer, W., Ferrè, E. R., Harrison, L., Hemming, R., Ivănescu, A., Reed, L., Roberts, F., Stevens, R., Tatlow, S., & Whittaker, L. (2021). Music and Sound in Virtual/Augmented

Realities—Questions, Challenges and ApproachesA Multidisciplinary Roundtable. *Journal of Sound and Music in Games*, *2*(2), 63–83. https://doi.org/10.1525/jsmg.2021.2.2.63

Svanæs, D. (2019). Phenomenology through Design: A Tale of a Human Tail. *ACM CHI Conference on Human Factors in Computing Systems: Weaving the Threads of CHI*, Glasgow, United Kingdom.

Swann, C., Keegan, R. J., Piggott, D., & Crust, L. (2012). A systematic review of the experience, occurrence, and controllability of flow states in elite sport. *Psychology of Sport and Exercise*, *13*(6), 807–819. https://doi.org/10.1016/j.psychsport.2012.05.006

Swann, G. M. P., & Watts, T. P. (2002). Visualization Needs Vision: The Pre-Paradigmatic Character of Virtual Reality. In S. Woolgar (Ed.), *Virtual society? Technology, cyberbole, reality* (pp. 41–60). Oxford University Press.

Sweet, M. (2015). *Writing interactive music for video games a composer's guide*. Addison-Wesley.

Tabellini, E. (2014). Installation, Technology and Education in the Communication Strategies of the Museo Internazionale e Biblioteca della Musica di Bologna. *Music in Art*, *39*(1–2), 281–291.

Tan, S.-L., Cohen, A. J., Lipscomb, S. D., & Kendall, R. A. (2013). *The Psychology of Music in Multimedia*. Oxford University Press, Incorporated.

Tanenbaum, K., & Tanenbaum, T. J. (2010). Agency as commitment to meaning: Communicative competence in games. *Digital Creativity*, *21*(1), 11–17. https://doi.org/10.1080/14626261003654509

Tariq, M., Trivailo, P. M., & Simic, M. (2018). EEG-Based BCI Control Schemes for Lower-Limb Assistive-Robots. *Frontiers in Human Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnhum.2018.00312

Tarkovsky, A. (1989). *Sculpting in Time: Reflections on the Cinema* (K. Hunter-Blair, Tran.) University of Texas Press.

Taruskin, R. (1995). *Text and Act: Essays on Music and Performance*. Oxford University Press.

Tatlow, S. (2020). Everyone in Space Wants to Hear You Scream: Toward a Framework for Understanding Player Voice in Virtual Worlds. *Journal of Sound and Music in Games*, *1*(3), 15–34. https://doi.org/10.1525/jsmg.2020.1.3.15

Tatlow, S. (2022). Music and Narrative Experience in Final Fantasy XIV: A Realm Reborn. In R. Anatone (Ed.), *The Music of Nobuo Uematsu in the Final Fantasy Series* (pp. 206–235). Intellect.

The 8-Bit Arcade. (2018, November 25). *Disneyland Adventures—Conducting in Mickey's Soundsational Parade (Xbox One Gameplay)*. https://www.youtube.com/watch?v=ZGmZAeolj6I

The Economist. (2017, July 13). Virtual reality gives old fairground rides new purpose. *The Economist*. https://www.economist.com/science-and-technology/2017/07/13/virtual-reality-gives-old-fairground-rides-new-purpose

The Flight's Music Brings Ancient Greece Alive In 'Assassin's Creed Odyssey'. (2018, December 6). https://www.wshu.org/culture/2018-12-06/the-flights-music-brings-ancient-greece-alive-in-assassins-creed-odyssey

Thomas, V. L., & Saenger, C. (2020). Feeling excluded? Join the crowd: How social exclusion affects approach behavior toward consumer-dense retail environments. *Journal of Business Research*, *120*, 520–528. https://doi.org/10.1016/j.jbusres.2018.12.064

Tichon, J., Wallis, G., and Mildred, T. (2006). "Virtual training environments to improve train driver's crisis decision-making," *in Proceedings of SimTecT*, Melbourne. Lindfield, NSW: Simulation Industry Association of Australia.

Tierney, J., Zissman, M. A., & Eddington, D. K. (1994). Digital signal processing applications in cochlear-implant research. *The Lincoln Labatory Journal*, *7*(1), 31–62.

Tom Scott. (2017, June 5). *Blocking People in Real Life: Tom Scott at An Evening of Unnecessary Detail*. https://www.youtube.com/watch?v=fq7oauciBdc

Tom Scott. (2021, September 27). *The world's most useful model railway*. https://www.youtube.com/watch?v=6TLcaJdsRr0

Trapanese, J. (2020, November 14). *Joseph Trapanese: Exclusive Interview and Score – Tales from the Galaxy's Edge* [Interview]. https://www.ilmxlab.com/news/tftge-composer-joseph-trapanese/

Travis, D., Watson, T., & Atyeo, M. (1994). Human Psychology in Virtual Environments. In L. W. MacDonald & J. Vince (Eds.), *Interacting with virtual environments*. J. Wiley.

Triggs, R. (2021, October 29). *Active noise cancelling (ANC) technology types explained*. SoundGuys. https://www.soundguys.com/noise-canceling-anc-explained-28344/

Tylka, J. G. (2019). *Virtual Navigation of Ambisonics-Encoded Sound Fields Containing Near-Field Sources* [Doctoral Thesis, Princeton]. https://www.princeton.edu/3D3A/Publications/Tylka_3D3A_Thesis.pdf

Truax, B. (2001). *Acoustic communication* (2nd ed). Ablex.

Ubisoft. (2018, December 7). *Inside the Studio with Audio Director Lydia Andrew*. https://www.ubisoft.com/no-no/game/assassins-creed/odyssey/news-updates/4vf1lEC7s4dhSKUCP1NA7U/inside-the-studio-with-audio-director-lydia-andrew

United Nations Counter-Terrorism Centre. (2022). *Examining the Intersection Between gaming and Violent Extremism*. Global Programme on Preventing and Countering Violent Extremism and Special Projects and Innovation Branch.

Unreal Engine. (2021, June 4). *Nanite | Inside Unreal*. https://www.youtube.com/watch?v=TMorJX3Nj6U

van Elferen, I. (2016). *Analyzing game musical immersion: The ALI model* (M. Kamp, T. Summers, & M. Sweeney, Eds.; pp. 32–52). Equinox. http://dx.doi.org/10.1558/equinox.23901

Varsányi, A. (2014). Making Museum Audible: Reflections on Implementing a Soundlab within a Museum of Musical Instruments. *Music in Art*, *39*(1–2), 293–309.

Vaske, J. J., & Shelby, L. B. (2008). Crowding as a descriptive indicator and an evaluative standard: Results from 30 years of research. *Leisure Sciences*, *30*, 111–126. https://doi.org/10.1080/01490400701881341

VICE. (2018, January 22). *How to Become TripAdvisor's #1 Fake Restaurant*. https://www.youtube.com/watch?v=bqPARIKHbN8

Vincent, E., Virtanen, T., & Gannot, S. (Eds.). (2018). *Audio source separation and speech enhancement*. John Wiley & Sons.

von Ende, R. C. (1965). Music and Shakespeare. *Music Educators Journal*, *51*(3), 48–50. https://doi.org/10.2307/3390345 von Falkenhausen, L. (1994). *Suspended Music: Chime-Bells in the Culture of Bronze Age China*. California University Press.

Vossen, E. (2018). *On the Cultural Inaccessibility of Gaming: Invading, Creating, and Reclaiming the Cultural Clubhouse* [Doctoral Thesis, University of Waterloo]. https://uwspace.uwaterloo.ca/bitstream/handle/10012/13649/Vossen_Emma.pdf

Vuzix Corporation. (2020, September 1). Netherlands' National Police Unit Using Vuzix Blade Smart Glasses in Police Work Study. *Vuzix Corporation*. https://ir.vuzix.com/news-events/press-releases/detail/1809/netherlands-national-police-unit-using-vuzix-blade

Wagner, K., & Schramm-Klein, H. (2019). *Alexa, Are You Human? Investigating the Anthropomorphism of Digital Voice Assistants – A Qualitative Approach*. Fortieth International Conference on Information Systems, Munich.

Walczak, S. (2021). Predicting Crime and Other Uses of Neural Networks in Police Decision Making. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.587943

Walther-Hansen, M., & Grimshaw-Aagaard, M. (2020). Don't extend! Reduce!: The sound approach to reality. *Proceedings of the 15th International Conference on Audio Mostly*, 8–15. https://doi.org/10.1145/3411109.3411111

Wang, A. L. (2003). An industrial-strength audio search algorithm. *Proceedings of the 4 Th International Conference on Music Information Retrieval*.

Wang, A., Thompson, M., Uz-Bilgin, C., & Klopfer, E. (2021). Authenticity, Interactivity, and Collaboration in Virtual Reality Games: Best Practices and Lessons Learned. *Frontiers in Virtual Reality*, *2*. https://www.frontiersin.org/article/10.3389/frvir.2021.734083

Watkins, R., & Jahankhani, H. (2021). 15—A regulatory investigation into the legal and ethical protections for digital citizens in a holographic and mixed reality world. In H. Jahankhani, L. M. O'Dell, G. Bowen, D. Hagan, & A. Jamal (Eds.), *Strategy, Leadership, and AI in the Cyber Ecosystem* (pp. 321–337). Academic Press. https://doi.org/10.1016/B978-0-12-821442-8.00018-5

Waysdorf, A., & Reijnders, S. (2018). Immersion, authenticity and the theme park as social space: Experiencing the Wizarding World of Harry Potter. *International Journal of Cultural Studies*, *21*(2), 173–188. https://doi.org/10.1177/1367877916674751

Weatherby, A., Henshall, K. R., & McKay, C. M. (2003). The Effects of Frequency Response on Speech Perception for Cochlear Implant Users. *Journal of the American Academy of Audiology*, *14*(10), 582–591. https://doi.org/10.3766/jaaa.14.10.7

Weisz, N., Moratti, S., Meinzer, M., Dohrmann, K., & Elbert, T. (2005). Tinnitus Perception and Distress Is Related to Abnormal Spontaneous Brain Activity as Measured by Magnetoencephalography. *PLOS Medicine*, *2*(6), e153. https://doi.org/10.1371/journal.pmed.0020153

Westerkamp, H. (1995). Listening to the Listening. *Sounding Out Genders: Women Sound Artists/Composers Talk about Gender and Technology*. International Symposium on Electronic Art (ISEA), Montreal.

Westerkamp, H. (2002). Linking soundscape composition and acoustic ecology. *Organised Sound*, *7*(1), 51–56. https://doi.org/10.1017/S1355771802001085

Westerkamp, H. (2019). The Disruptive Nature of Listening: Today, Yesterday, Tomorrow. In M. Droumeva & R. Jordan (Eds.), *Sound, Media, Ecology* (pp. 45–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-16569-7_3

Whalen, Z. (2004). Game Studies—Play Along—An Approach to Videogame Music. *Game Studies*, *4*(1). http://www.gamestudies.org/0401/whalen/

Widge, A. S., Zorowitz, S., Basu, I., Paulk, A. C., Cash, S. S., Eskandar, E. N., Deckersbach, T., Miller, E. K., & Dougherty, D. D. (2019). Deep brain stimulation of the internal capsule enhances human cognitive control and prefrontal cortex function. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09557-4

Wierzbicki, J. (2016). Sound Effects/Sound Affects: 'Meaningful' Noise in the Cinema. In L. Greene & D. Kulezic-Wilson (Eds.), *The Palgrave Handbook of Sound Design and Music in Screen Media: Integrated Soundtracks* (pp. 153–168). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51680-0_11

Willemsen, S., & Kiss, M. (2013). Unsettling Melodies: A Cognitive Approach to Incongruent Film Music. Acta Universitatis Sapientiae, *Film and Media Studies*, *7*(1), 169–183. https://doi.org/10.2478/ausfm-2014-0022

Williams, D. (2021). Psychophysiological Approaches to Sound and Music in Games. In M. Fritsch & T. Summers (Eds.), *The Cambridge Companion to Video Game Music* (pp. 302–318). Cambridge University Press. https://doi.org/10.1017/9781108670289.019

Wiltshier, P., & Clarke, A. (2017). Virtual cultural tourism: Six pillars of VCT using co-creation, value exchange and exchange value. *Tourism and Hospitality Research*, *17*(4), 372–383.

Winston, E., & Saywood, L. (2019). Beats to Relax/Study To: Contradiction and Paradox in Lofi Hip Hop. *IASPM Journal*, *9*(2), Article 2.

Winters, B. (2010). The Non-diegetic Fallacy: Film, Music, and Narrative Space. *Music and Letters*, *91*(2), 224–244. https://doi.org/10.1093/ml/gcq019

Winters, B. (2014). *Music, performance, and the realities of film: Shared concert experiences in screen fiction*. Routledge.

Woolgar, S. (2002). Five Rules of Virtuality. In S. Woolgar (Ed.), *Virtual society? Technology, cyberbole, reality* (pp. 1–22). Oxford University Press.

World XR Forum Crans-Montana. (2020, June 26). *How VR sound drives immersive storytelling ? Martin Rieger - VRTONUNG*. https://www.youtube.com/watch?v=YHbIWDK6dho

Wyatt, S. (2014). Music in the Museum: Intangible Influences. *Music in Art*, *39*(1–2), 267–279.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., & Yu, L.-F. (2021). A Review on Virtual Reality Skill Training Applications. *Frontiers in Virtual Reality*, *2*, 49. https://doi.org/10.3389/frvir.2021.645153

Yamada, M., Fujisawa, N., & Komori, S. (2001). The Effect of Music on the Performance and Impression in a Video Racing Game. *Journal of Music Perception and Cognition*, *7*, 65–76.

Yang, J., & Chan, C. Y. (2019). Audio-augmented museum experiences with gaze tracking. *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, 1–5. https://doi.org/10.1145/3365610.3368415

Young, K. (2017, March). *Assisting VR Gameplay Through the Use of Iconographic Music*. Game Developers Conference (GDC). https://www.gdcvault.com/play/1023945/Assisting-VR-Gameplay-Through-the

Zagal, J. P. (2010). *Ludoliteracy: Defining, Understanding, and Supporting Games Education*. Lulu.

Zeng, F.-G., Rebscher, S., Harrison, W. V., Sun, X., & Feng, H. (2008). Cochlear Implants: System Design, Integration and Evaluation. *IEEE Reviews in Biomedical Engineering*, *1*, 115–142. https://doi.org/10.1109/RBME.2008.2008250

Zhang, W., Samarasinghe, P., Chen, H., & Abhayapala, T. (2017). Surround by Sound: A Review of Spatial Audio Recording and Reproduction. *Applied Sciences*, *7*(5), 532. https://doi.org/10.3390/app7050532

Zoom. (2022, February 9). *Using push-to-talk*. Zoom Support. https://support.zoom.us/hc/en-us/articles/360000510003-Using-push-to-talk

Filmography

Blade Runner, Ridley Scott, Warner Bros: 1982

Frozen, Chris Buck & Jennifer Lee, Walt Disney Studios Motion Pictures: 2013

Inception, Christopher Nolan, Warner Bros. Pictures: 2010

Jurassic Park, Steven Spielberg, Universal Pictures: 1993

Sword Art Online, Tomohiko Itō, A-1 Pictures: 2012, 2014

The Forbidden Planet, Fred M. Wilcox, Metro-Goldwyn-Mayer: 1956

The Matrix, The Wachowskis, Warner Bros: 1999

Tron, Steven Lisberger, Buena Vista Distribution: 1982

Ludography

Citations provided in format: [TITLE], [Developer(s)], [Publisher(s)]: [Date(s)]

APE OUT, Gabe Cuzzilo, Bennett Foddy & Matt Boch, Devolver Digital: 2018

Albion Online, Sandbox Interactive, Sandbox Interactive: 2017

Assassin's Creed Odyssey, Ubisoft Quebec, Ubisoft: 2018

Ballblazer, Lucasfilm Games, Epyx: 1984

Battlefield 4, DICE, Electronic Arts: 2013

Bring to Light, Red Meat Games, Red Meat Games: 2018

Civilization IV, Firaxis Games, 2K Games: 2005

Cuphead, Studio MDHR, Studio MDHR: 2017

Disneyland Adventures, Asobo Studio, Microsoft Studios: 2017[97]

Dragon Front, High Voltage Software, Inc., High Voltage Software, Inc.: 2016

Dune II, Westwood Studios, Virgin Games: 1992

EVE Online, CCP, CCP: 2003

Fail Factory, Armature Studio, Oculus: 2020

Final Fantasy XIV, Square Enix, Square Enix: 2010[98]

Harry Potter: Wizards Unite, Niantic, Niantic: 2019

Half-Life: Alyx, Valve Corporation, Valve Corporation: 2020

Half-Life 2, Valve Corporation, Valve Corporation: 2004

Half-Life 2: Episode 2, Valve Corporation, Valve Corporation: 2007

---

[97] Formerly released as: *Kinect: Disneyland Adventures*, Frontier Developments, Microsoft Studios: 2011
[98] Also including the following expansions developed and published by Square Enix: *A Realm Reborn* (2013); *Heavensward* (2015); *Stormblood* (2017); *Shadowbringers* (2019); *Endwalker* (2021)

Ingress, Niantic, Niantic: 2013

League of Legends, Riot Games, Riot Games: 2009

Leaving Lyndow, Eastshade Studios, Eastshade Studios: 2017

Monkey Island 2: LeChuck's Revenge, LucasArts, LucasArts: 1991

Need for Speed Shift 2: Unleashed, Slightly Mad Studios, Electronic Arts: 2011

No Man's Sky, Hello Games, Hello Games: 2016

Northgard, Shiro Games, Shiro Games: 2017

Pac-Man, Namco, Namco: 1980

PAYDAY 2, Overkill Software, 505 Games: 2013

Pokémon Go, Niantic, Niantic: 2016

Portal 2, Valve Corporation, Valve Corporation: 2011

Rocksmith, Ubisoft, Ubisoft, 2011

Second Life, Linden Lab, Linden Lab: 2003

Super Mario Bros, Nintendo, Nintendo: 1985

Tethered, Secret Sorcery Limited, Secret Sorcery Limited: 2016

Tetris, Alexey Pajitnov, Nintendo: 1984

The Elder Scrolls V: Skyrim, Bethesda Game Studios, Bethesda Softworks: 2011

The Witcher 3: Wild Hunt, CD Projekt RED, CD Projekt: 2015

V Rising, Stunlock Studios, Stunlock Studios: 2022[99]

[99] Early access title

World of Warcraft, Blizzard Entertainment, Blizzard Entertainment: 2004[100]

[100] Also including the following expansions developed and published by Blizzard Entertainment: *The Burning Crusade* (2007); *Wrath of the Lich King* (2008); *Cataclysm* (2010); *Mists of Pandaria* (2012); *Warlords of Draenor* (2014); *Legion* (2016); *Battle for Azeroth* (2018); *Shadowlands* (2020); *Dragonflight* (2022); and *World of Warcraft Classic* (2019)

Appendix A:

An Interrogative Approach to Audio in Virtual Experience

As outlined in section 5.1.4, it is important to note that the construction of virtual experience is inherently inter-dependent: changes to one element of the virtual experience will inevitably affect other elements. That is to say: we could not have a virtual experience that consisted *only* of virtual **externality**, or *only* of virtual **hardware** and **software**; all virtual experiences will have virtual **externality**, virtual **hardware**, and virtual **software**, and the overall virtual experience will be formulated through the combination of all three elements and the relationships between them. This means that in order to understand the potential effect of any given change or design decision, or to analyse any specific element of a virtual experience, it is likely necessary to consider the virtual experience holistically – especially when considering things from an experiential perspective.

This framework is perhaps at its most useful when it is used in this way: an experience is broken down into the constituent elements outlined throughout Chapter 1, with specific considerations for audio highlighted through section 5.1.1–5.1.3. A phenomenological approach to analysing this can be taken by "imagining variations on a phenomenon, and considering at what point those variations would cease to be instances of that phenomenon [to] identify its essential characteristics" (Kamp, 2016, p.172). Within the context of the interrogative approach outlined in this appendix, this means considering how changing the answer to one of these questions might have a cascade effect throughout the answers to other questions.

Virtual Externality

1. What sounds are present in the virtual externality?
    a. What is the real-world location of the externality?
    b. What sounds will there be in this location?
        i. What are the exosonic components of the external sounds (e.g,. reverberation)
    c. Are there any sounds in the externality which we may wish to hear?

        i. Useful sounds (e.g., alarms, doorbells)
        ii. Other participants/users
2. Will these sounds transgress into the virtual experience?
    a. Sensorial immersion: are we sensorially immersed in the experience? Can we hear these sounds?
    b. Narratival immersion: do sounds in the externality conflict with expected sounds in the virtual experience?
    c. Challenge-based immersion: will these sounds confuse the user?
3. How can audio emerging from the virtual externality be controlled?
    a. Can we change the virtual hardware used for the experience?
        i. closed-ear headphones
        ii. open-ear headphones / headband speakers
        iii. external speakers
        iv. multi-modal approaches (e.g., see section 5.3.1)
    b. Can we use virtual software to screen external sound?
        i. Noise-cancellation (including look-ahead/active noise cancellation; see section 3.0.2)
        ii. Raising the auditory floor (e.g., through music; see section 4.1.3)
    c. Can we relocate sound sources beyond the perceptual horizon of the user? (e.g., Disney enforce a no-fly zone over some theme parks; see section 1.3.1)

Virtual Hardware

1. What hardware is used for audio within the virtual experience?
    a. How is sound being broadcast to users?
        i. Speakers (Stereo? Cinematic?)
        ii. Sub-woofer (How many? Where is it located?)
        iii. Headphones (Stereo? Binaural?)
    b. Are any novel broadcasters / haptic devices used?
        i. Rumble pads; see section 5.3.2
        ii. Electrodes (e.g., Cochlear-Implants); see section 6.x.x.
    c. Are any auditory controllers used? i.e., does sound need to be transmitted into the virtual experience?
        i. How many microphones / microphone arrays are used?
        ii. What microphones / microphone arrays are used? (e.g., are any ambisonic microphones used?)

        iii. What are the properties of these microphones / microphone arrays? (e.g., frequency response, polar patterns)
        iv. Where are these microphones / microphone arrays located? (e.g., head-mounted, device-mounted, externally mounted)
    d. Are any novel auditory controllers used?
        i. Digital audio devices (e.g., soundboards)
        ii. Direct audio devices (e.g., musical instruments plugged directly into other hardware)
    e. What platforms are used?
        i. How is audio being processed? (e.g., are devices plugged into a computer?)
        ii. What limitations does this platform have? (e.g., can this platform process multi-channel audio? See section 5.3.2)
        iii. What effects are there on the sound as a result of the audio interface in use?
        iv. What effects are there on the sound as the result of the DSP hardware in use?
2. Is the virtual audio hardware linked to any other hardware in use throughout the experience?
    a. For example: head-tracking from HMDVR is used to affect binaural spatialisation within headphones.
3. What control does the user have over the hardware used for the experience?
    a. How will the experience change to reflect user decisions surrounding audio hardware? (e.g., user decides not to use headphones = binaural audio not possible; what kind of spatialisation will be provided to the user instead?)

Virtual Software

1. What information is intended to be provided to the user through sound?
    a. Signification / explicit information (e.g., birdsong = birds)
    b. Implicit information / exosonic components (e.g., large reverberation = large room)
    c. Affective sound (see Coulthard, 2016)
2. What is intended to be conveyed to the user through music?
    a. What narrative information is conveyed by the music? (e.g., historical, geographical, atmosphere setting; see section 4.1.1)
    b. What emotional information is conveyed by the music? (see Chion, 1995)
    c. Will music increase the embodiment of the user? (see Plank, 2021)
3. How will the user understand this information?
    a. Media literacy (see Aufderheide, 1993; van Elferen, 2016)
    b. Earcons (Blattner et al., 1985) and auditory icons (Gaver, 1986)

    c. Implicit understanding (e.g., spatialisation; see Larsson et al., 2010)
    d. If necessary, how will the user be taught to understand/use the information?
    e. May the intended information be misconstrued by the user?
4. What control does the user have over audio within the experience?
    a. For example: volume control; audio settings (e.g., see figures 18 & 19)
5. Does the user need to transmit audio into the virtual experience?
    a. User-to-user communication (see section 5.4.0–5.4.3)
    b. User-to-experience communication (e.g., voice commands; see section 6.1.1)
    c. External music (e.g., music concerts in Second Life, see Gagen & Cook, 2016)
6. What listening behaviours is the user expected to exhibit?
    a. In what ways might the user be considered auraldiverse? (see Drever & Hugill, 2023)
    b. How attentive will the user be? (consider distracted listening / ubiquitous listening)
    c. How might the design be modified to be aware of these behaviours?