# Audio augmented objects and the audio augmented reality experience

by Laurence Cliffe

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

March 2022

# Abstract

This thesis explores the characteristics, experiential qualities and functional attributes of audio augmented objects within the context of museums and the home. Within these contexts, audio augmented objects are realised by attaching binaurally rendered and spatially positioned virtual audio content to real-world objects, museum artefacts, physical locations, architectural features, fixtures and fittings.

The potential of these audio augmented objects is explored through a combination of practice-based research and ethnographically framed studies. The practical research takes the form of four sound installation environments delivered through the use of an augmented reality mobile phone application that are deployed within a museum environment and in participants' homes.

Within these experiences, audio augmented objects are capable of being perceived as the actual source of virtual audio content. The findings also demonstrate how the perceived characteristics of real-world objects and physical space can be altered and manipulated through their audio augmentation.

In addition, audio augmented museum objects present themselves as providing effective interfaces to digital audio archival content, and digital audio archival content presents itself as an effective *re-animator* of *silenced* museum objects. How audio augmented objects can function as catalysts for the exploration of physical space and virtual audio space within both the home and museum is presented. This is achieved by the uncovering of a sequence of interactional phases along with the uncovering of the functional properties of different types of audio content and physical objects within audio augmented object realities.

By way of conclusion, it is proposed that the audio augmented object reality alters the current, popular experience of acoustic virtual reality from an experience of *you* being *there*, to one of *it* being *here.* This change in the perception of the acoustic virtual reality has applications across an array of audio experiences, not just within cultural institutions, but also within various domestic listening experiences including the consumption and delivery of recorded music and audio-based drama.

# Acknowledgements

Firstly, I would like to thank my supervisors Chris Greenhalgh, Adrian Hazzard and James Mansell. I would also like to thank Steve Benford, Alan Chamberlain, Michel Valstar, Joel Fischer, Paul Tennent and the rest of the staff and students at the Mixed Reality Laboratory, and at the Horizon Centre for Doctoral Training, for their time, advice, support and encouragement. Many thanks also to Annie Jamieson at the National Science and Media Museum, and also to Aleks Kolkowski and Nick Redfern.

I would also like to give a special thank you to Lee Hazeldine, Adam Paliwala, Beverley Cliffe, Jack Cliffe and Ben Cliffe for all their support and encouragement.

I gratefully acknowledge the funding and support from the Horizon Centre for Doctoral Training at the University of Nottingham and the Engineering and Physical Sciences Research Council (EPSRC) (RCUK Grant No. EP/L015463/1).

# Table of Contents

# Figures

# Tables

## Publications

Cliffe, Laurence, Mansell, James, Greenhalgh, Chris & Hazzard, Adrian, 2020. *Materialising contexts: virtual soundscapes for real-world exploration: Personal and Ubiquitous Computing.* Personal and Ubiquitous Computing. 25(4), 623-636. Springer, New York.

Cliffe, Laurence, Cormac, Joanne, Mansell, James, Greenhalgh, Chris & Hazzard, Adrian, 2019. *The Audible Artefact: Promoting Cultural Exploration and Engagement with Audio Augmented Reality In: AM '19: Proceedings of the 14$^{th}$ International Audio Mostly Conference: A Journey in Sound.* 176-182. ACM Press, New York.

# 1 Introduction

This chapter introduces three of the key terms within this thesis: *Augmented Reality, Audio Augmented Reality* and *Audio Augmented Objects.* Following a brief statement on the motivation behind the research, the scope of the presented research is described, along with its anticipated and accomplished contributions to date. Lastly, the structure of the thesis is outlined, providing details of the contents of each chapter.

Within this thesis, and line with Schraffenberger & Heide (2014) and Hugues et al. (2011), *Augmented Reality* (AR) is considered as the virtual augmentation of physical space, as well as physical objects. A slightly more comprehensive view is taken to that of Azuma et al. (2001), where it is suggested an AR experience should include '*a combination of real and virtual objects*, *should run interactively in real-time* and *should align real and virtual objects with each other*'. More broadly speaking, a similar view is taken to Schraffenberger & Heide (2014) in that a real component and a virtual component is required, and it is the relationship between the two that constitutes the augmentation and its nature.

*Audio Augmented Reality* (AAR) is therefore considered as a mixed reality experience realised through a combination of virtual audio and the physical reality. The term AAR, having one of its earliest references in a 1995 paper by Benjamin Bederson (Bederson, 1995), is considered here as a physical reality that has been augmented with exclusively virtual audio content. Here the term is used to distinguish from AR, which, though often assumed to refer to visual augmentation, suggests the augmentation of reality with any variety of virtual content.

An *Audio Augmented Object* is a physical object, residing in reality that has had virtual audio content attached to it, or attached to the location at which it is situated. Just as virtual graphical content within an AR experience can be perceived as being situated in the real-world through a device's camera, within an AAR experience virtual audio content can be perceived as being situated in the real-world through a device's speaker or headphones. By placing a perceptibly realistic virtual audio source at the same location as a physical object it is possible to create the perception that the audio is emanating from the same location as the object, or even the object itself.

It is the potential of a real-world object to be perceived as the actual source of virtual audio content that realises an *audio augmented object,* and it is the construction of experiences that contain such objects that offer new and exciting listening experiences.

## 1.1 Motivation

Through my creative practice I have developed a way of working across the arts, communication and marketing that embraces experimental and exploratory creative processes. This artistic exploration and enquiry generates and informs the use of innovative aesthetics for the purposes of commercial communications. Such practice has included working with sound, open-source hardware and software, radio transmissions, live data and custom digital processing techniques, and has often sought to exploit both emerging and near-obsolete technologies in order to create innovative visual and sonic artwork. Influenced by fields including cybernetics, artificial intelligence and science fiction, the resultant works are often realised by exploiting technology's perceived failings in order to expose otherwise hidden content, or to generate initially indeterminate output, that can be used as a foundation for research and exploration.

As part of this practice, I have become increasingly interested, not just with the aesthetic or functional possibilities of malfunction, but also the creation of contexts or systems within which these possibilities can happen. For it appears that it is the unmediated processes of a system that has been determined capable of generating creative output within which such possibilities are unearthed.

> *"The system is the work of art; the visual work of art is the proof of the system."* - Sol LeWitt (1999)

It is possible to think about the system, or indeed the creation of a context within which creative content can be produced, as being a creative practice in itself. While LeWitt's comments suggest that a system can be a work of art, within the work of Kit Galloway & Sherrie Rabinowitz we can see an emphasis on the production of context, rather than the production of artefact (Durland, 1987). Within these contexts, we also see the creation of spaces within which the audience are provided the means to create and compose their own creative, non-artefactual, temporal content.

Cuartielles (2012) demonstrates an interesting link between the fields of sound art and ubiquitous computing. In his paper *Embodied Sound* (Cuartielles, 2012), Cuartielles suggests that there is a '*clear link'* between the assertions of Weiser (Weiser, 1999), regarding the importance a precise positioning system has to the advancement of ubiquitous computing, and the needs of soundscape composers.

Like Cuartielles' *Embodied Sound* installation (Cuartielles, 2012) it is amongst a framework of music technology, sound art, ubiquitous computing and the virtual augmentation of reality that the work within this thesis is positioned.

## 1.2 Scope of thesis

The research presented within this thesis seeks to gain a greater understanding of the contribution of audio augmented objects to the audio augmented reality experience. This includes both how they are experienced by listeners, as well as how the listener experience can be authored by interested parties. It is envisaged, given the scope and breadth of existing audio augmented reality experiences, that these interested parties would include museum and gallery curatorial staff, artists, installation designers, audio game designers, musicians, composers and producers of audio-based content seeking to create innovative experiences. Furthermore, the work in this thesis intends to inform the design and realisation of audio augmented object realities via a thorough examination of their attributes and potential functions within the acoustic virtual realities they create.

As such, the questions that I am attempting to answer by undertaking the research presented here firstly address the character and user experience of an augmented reality that contains audio augmented objects. Secondly, they explore the potential functionality that audio augmented objects add to an audio augmented reality. Thirdly, and finally, they attempt to determine what the best practices are for realising functional audio augmented objects.

The practice on which this thesis is based was undertaken between March 2019 and October 2020. Chapter 4 details, and reflects upon, the design, development and deployment of four AAR installations. The first installation, *The Mingus Demonstration,* was demonstrated to curatorial staff at the *National Science and Media Museum* (NSMM) in Bradford, UK. The second installation, *The McMichael Experiment,* was deployed during a workshop attended by researchers, museum professionals, museum visitors and members of the public, again at NSMM. *Alien Encounters*, the third installation, was deployed for members of the public to experience during an evening open event at NSMM, and the final installation, *Horror-Fi Me,* was deployed for use in user's homes.

Whilst the first three installations were deployed prior to the COVID-19 restrictions, the final project; *Horror-Fi Me,* was deployed during the COVID-19 lockdown restrictions. During this period, due to the closure of public galleries and museums, this final installation was designed to be remotely deployed for participants to author and experience within their own homes. Whilst this change of context was unanticipated, its impact served to broaden both the scope of the practice and therefore also the scope of the research. Whilst the change in context from public gallery to domestic environment served to broaden thinking around the potential applications of audio augmented objects and offer insights into what their more general characteristics are, it also focused attention on how they could be realised.

In line with both the context and the content of *Horror-Fi Me,* this broadening of scope also provided insights into applications for various domestic listening experiences including: the composition, delivery and consumption of domestic musical listening experiences and audio dramas.

Furthermore, given the contexts within which we see how the *acoustic virtual reality* has evolved (as outlined in section 2.5.1), it appears, with hindsight, that this unexpected digression is not far detached from the original concerns of the research. What also unifies this broadening of scope is *the cultural artefact*. Whether virtual or real, audible or silent, we remain concerned with technical interventions which create new experiences through audio with existing *cultural artefacts*.

## 1.3 Contributions

The contributions of this thesis arise from the design, development and deployment of four audio augmented reality sound installations and the subsequent study of two of these installations. The findings from the research presented within this thesis have contributed to an understanding of how audio augmented objects can be used to promote exploration and engagement with both physical space and virtual audio content.
This includes how audio augmented objects can be used to engage listeners with digital audio archival content, how they can effectively communicate their presence beyond line-of-sight and how they can reframe themselves as objects as well as the space around them.

Contributions are made to the field of HCI, primarily through a developed understanding of the affordances of audio augmented objects, which in turn has the potential to inform the specific use of audio within the design of mixed reality experiences more generally. For example, their ability to effectively localise attention within the physical reality, reframe the perception of reality and to realise intuitive and contextually relevant interfaces to digital audio assets can form and inform parts of larger mixed reality experiences. Furthermore, the understanding of such affordances has the potential to inform the creation of audio-centric interfaces beyond the cultural.

As well as contributions to curatorial and artistic practice, the research also offers contributions to the field of audio entertainment. By considering the change in the perception of the acoustic virtual reality from an *acousmatic* to a perceptively *direct* listening experience, it is possible to imagine the creation of new, mixed reality domestic listening experiences.

It is also hoped that the initially presented theoretical understanding of what defines and constitutes an audio augmented reality, along with its historical and contextual positioning within the development of audio and listening technologies, can be a useful contribution to the field of sound studies.

In addition, this research has directly informed and contributed to the development of the practical components of two subsequent AHRC funded research projects, *Sonic Futures* and *The Media of Mediumship.*

## 1.4 Thesis structure

Within this final section of this chapter I will now outline the structure of the thesis and the contents of the subsequent chapters.

In chapter 2 I review the key terms and the theoretical and technical assets involved within the construction and realisation of audio augmented reality experiences and audio augmented objects. This is achieved through a review of relevant studies, projects and example experiences, and through an examination of augmented reality more generally, prior to specific focus on audio augmented reality. The review of literature presented within this chapter also serves to identify and explain the initial trajectory and context of the practice-based research presented within chapter 4. The three key research questions which form the basis of initial enquiry within the practice are subsequently outlined. These questions address how audio augmented objects change the audio augmented reality experience, what their functional properties are, and how experiences containing audio augmented objects can be best realised.

Chapter 3 details the practice-based and research-through-design methodology I have developed and engaged with in order to conduct the research described in the previous chapter. This chapter also outlines the relevance and importance of the ethnographic framing of the studies and interviews conducted within chapters 5 and 6, and the specific relevance of such an approach within the context of research into the experiential qualities of sound technologies.

The fourth chapter provides some background to my creative practice and recounts the story of the practice undertaken as research for this thesis. It details, in chronological order, the design, development, authoring and deployment of four audio augmented reality sound installation projects. Each project is reflected upon in turn, with each subsequent project building upon the findings of the last, reflecting an iterative design approach. A section of conclusions is included at the end of this chapter which draws together all the findings from this creative practice.

Chapters 5 and 6 describe, and present the results of, the ethnographically-framed studies of two of the deployed audio augmented reality sound installations detailed within the previous chapter. Each study is followed by an interview with a professional practitioner who's practice directly relates to the context and content of the installation. In chapter 5 the findings from the deployment and study of an AAR installation within the setting of a museum are presented. This is followed by interview with the museum's curator of sound technologies, who experienced sound installation, and also an earlier prototype experience. In chapter 6 the findings from the remote deployment and study of an AAR installation within participant's homes are presented. These are followed by the findings from an interview with the composer who created the musical soundtrack for the installation experience.

The interviews with these professional practitioners are undertaken in addition to the study of user interactions in order to obtain insights into the user experience, and to inform the potential application and design of such experiences. Conclusions are drawn at the end of each of these chapters in light of each study's findings, the accompanying practitioner interviews, and in relation to the previously discussed literature and related contemporary discourse.

In chapter 7 the key recurring themes evident within the findings from the literature, practice, studies and interviews are discussed whilst drawing on discourse from a variety of related sources which include media theory, human-computer-interaction, sound studies and the arts.

Finally, in chapter 8, conclusions are drawn explicitly addressing the research questions outlined within section 2.9. These conclusions are considered in light of the discussion within the previous chapter and attempts to align and address the findings from all aspects of the research conducted in relation to excepted theories around the subject.

# 2 Literature review and related work

Within this chapter, a review of literature around the subject of audio augmented objects deals with the question of what is, and what constitutes an audio augmented reality. This is considered, not from a technological perspective, but rather an experiential perspective, in an attempt to highlight both the commonalities and the differences between the user experiences of different technologies, and in order to ascertain what different technological innovations and approaches have brought to the table.

Following a subsequent review of what constitutes an *audio augmented reality* (see section 2.1), and an explanation of what defines it from an *acoustic virtual reality* (see section 2.1.1), a review of different *audio augmented realities and acoustic virtual realities* is undertaken, which includes: *mobile audio* (see section 2.2), *locative audio* (see section 2.3) and differing types of *binaural audio* experiences (see section 2.4).

Based on the review of these different types of audio experiences, along with theoretical observations relating to them, the components that define and comprise an *audio augmented reality* are identified. These component parts of an audio augmented reality, when considered in relation to literature regarding the history of sound and listening technology, enable a timeline to be presented that considers a trajectory towards the creation of increasingly sophisticated *acoustic virtual realities*. When combined with listening technologies that have an increasing ability to augment physical space, these *acoustic virtual* realities, in turn, realise increasingly capable *audio augmented realities.*

Example projects, experiences and studies that have previously realised audio augmented objects are presented and discussed. The function, character and experience of these objects are also discussed, with such attributes as their ability to promote engagement with physical space, and to remotely engage users with their presence, offering glimpses into their potential.

In a similar way the component parts of an audio augmented reality are presented, the components that combine to realise an *audio augmented object*, or an *audio augmented object reality* are also presented. Additionally, the functional requirements of the enabling technologies involved in the creation of these components are also identified, along with a context that will facilitate the deployment of audio augmented object-based experiences and their study.

Last, but not least, the research questions that I will seek to answer within this thesis are presented and described.

Overall, the following review of literature around audio augmented reality and audio augmented objects serves as a record of the point of departure into the practice-based research presented within chapter 4. My intention is that it serves to explain an initial position that evolves and develops over the course of the practice, as well as providing and explaining initial subjects of enquiry.

## 2.1 Audio augmented Reality

Prior to looking at examples of previously created audio augmented objects, it seems appropriate to first define the scope of AR, and specifically AAR, that we are considering within the context of this thesis. This effort seems of particular relevance when dealing with audio augmented reality as opposed to visually orientated augmented reality due to the physical properties of sound itself. Unlike visual AR, sound envelopes the user, or in this case the listener, it fills the physical space of the experience and becomes part of it, having the ability to reframe and re-organise it. In comparison, visually orientated AR is retained, if not within the bounds of a screen, then within the limits of line-of-sight, lacking both the permeable and pervasive qualities of audio.

AR, and therefore AAR by association, is generally perceived to involve a high-technical intervention that overlays digitally rendered virtual content onto the real-world. Attempting and working towards a definition for AAR that includes all types of listening experiences that overlay virtual audio content on to the physical environment, regardless of what that technology may be, allows us to place AAR within an appropriate technological linage, effectively distinguishing it from other technologies and better understanding what makes it different and special.

Rather than thinking about a reality augmented by that which is determined to be an *augmented reality technology* we can begin to include all sorts of technologies, even those which are so embedded within our everyday lives that they are barely considered *technologies*, let alone *augmented reality technologies*.

In relation to visually applied AR, Wang (2018) uses the helpful example of the measuring cup to illustrate this point rather well. He suggests that the measurement units printed onto the outside of the cup, along with the transparent material from which it is made, constitute an *augmented reality,* as additional information pertaining to exactly how much liquid it has inside is overlaid onto our reality. This wide-reaching interpretation of augmented reality is also taken up by Schraffenberger & Heide (2014), who propose an interpretation of AR that includes all modalities, not just those that require computational technologies. Additionally, Kee & Compeau (2019) along with Kee et al. (2019) use plaques, monuments, physical signposts and place markers as examples of ways in which historians have traditionally augmented the present with information about the past. Krzyzaniak et al. (2019), specifically regarding AAR, include such technologies as Sat Navs, telephones and Karaoke within their taxonomy of audio augmented reality and suggest that there are technologies that pre-date the term *augmented reality* but could be considered as such.

Such far-reaching and over-arching interpretations quickly become embroiled within and around the philosophical question of what the exact nature of reality is within an augmented reality experience, but this question, on an immediate level, is worth some investigation. As pointed out by Hugues et al. (2011) and Schraffenberger & Heide (2014), if reality is defined as everything, then it cannot be augmented as it *is* already reality, and therefore the question remains as to exactly what is it that is augmented? One possible answer is that it is our perception of reality as experienced through our senses that is augmented, rather than reality itself (Hugues, Fuchs and Nannipieri, 2011; Schraffenberger and Heide, 2014). If we apply a more cybernetic interpretation to the way in which humans use and interact with technologies, thinking of them as extensions of the human body, or as McLuhan (1994) puts it: *"…as an extension of the human nervous system",* then the proposition that AR is in fact the augmentation of our senses and thus an augmentation of our perception of reality, rather than the augmentation of reality itself, holds true.

The discussion around the exact nature of reality within an AR environment is as multi-faceted as it is fascinating and therefore more than capable of distracting us from the subject in hand. Needless to say that it illustrates the importance of exploring and studying non-ocular augmenting technologies and suggests that we should consider all technologies that alter our perception of reality as technologies of *augmented reality*.

If we consider AAR within the scope of technologies that alter our perception of reality through our sense of hearing we immediately have a very broad church that would include possibly all and every musical and audio playback or listening device since the invention of audio recording. If, on the other hand, we consider AAR in relation to other mobile audio listening technologies we can perhaps begin to examine AAR within a manageable and relevant technical sphere. For it is the technical development of mobile audio listening that allows recorded audio to exist within almost any location or space. It is this technical development that enables, perhaps to the greatest extent, that which is amongst our primary concerns in relation to AAR; the association between physical space and virtual sound. The importance of this association is exemplified in the following quote from science-fiction author William Gibson, and is used within the preface to Michael Bull's seminal work on mobile audio listening - *Sounding Out the City: Personal stereos and the management of everyday life.*

> *'The Sony Walkman has done more to change human perception than any virtual reality gadget. I can't remember any technological experience that was quite so wonderful as being able to take music and move it through landscapes and architecture'* - Gibson *in* Bull (2000)

This quote from Gibson also alerts us to another important concern, that of the listener's freedom of movement through space, the feature that distinguishes the personal stereo experience from the home hi-fi experience. Just as we expect a VR experience to contain an explorable environment with some degree of freedom of movement, even if it's just the freedom to orientate the head and look around, should such agency be a requirement for an audio augmented reality experience? If so, we can swiftly make a distinction between static technologies that augment our perception of our audible reality, such as a home-situated smart speaker or a radio, and those which create a spatially interactive and explorable environment.

Within visually orientated VR and AR environments the freedom of spatial exploration is realised through the spatial positioning of virtual three-dimensional vector graphics, within an AAR environment the freedom of spatial exploration is realised through the inclusion of spatially positioned virtual audio sources. We can therefore conclude that if we are to afford AAR listeners with a comparable level of immersion to that which can be experienced within a VR, or a visual based AR environment, then the realisation of an explorable three-dimensional environment, to some degree at least, should be a pre-requisite. Along with thinking about degrees of exploration, the difference between a VR environment that you can look around from a stationary position, and one that you can move through are potentially good examples here, we can also think of such experiences as having differing degrees of freedom or agency to explore.

Krzyzaniak et al. (2019) provide a useful, though by no means exhaustive, taxonomy of audio augmented reality examples and categorise these into six main types. Although included within these types are those that *'overlay extra audio information onto the real-world*', personal or mobile listening technologies are not explicitly included within this type. Furthermore, there are other suggested types, that include: '*Deliberate blending of acoustic and digital sound' and 'Realtime digital modification of acoustic sounds'* that include the examples: '*Karaoke'* and '*Electric guitar distortion pedals'* respectively. Within the definition proposed here, such examples could be dismissed as being examples of AAR as they do not realise an explorable spatial environment, even though they augment our perception of reality with sound.

So, in summary, our scope of interest in technologies that could be usefully described as being *audio augmented reality technologies* are those which are capable of realising a freely explorable virtual and, or, physical environment the perceived reality of which is augmented through our sense of hearing by the inclusion of virtual audio content.

Whilst attempting to position AAR within the technical scope of mobile listening it is worth looking at the different approaches to realising such an AAR experience. It is perhaps convenient and somewhat straight forward to think of an AAR experience delivered through headphones and a smart phone as a development of mobile listening technology, but what about, for example, speaker-based AAR experiences?

If we look at two speaker-based AAR examples, the outdoor AAR installation presented by Lawton et al. (2020) and the indoor AAR installation presented by Gimenes et al. (2016), we can still maintain that a freely explorable environment is realised and augmented by virtual audio sources. In both cases the virtual audio is included via loudspeakers and the perceived realities of the installation environments (in these cases the realities of the physical environment) are augmented by the inclusion of this virtual audio content. Within speaker-based AAR experiences the ability for the listener to have a personalised experience is dependent on their ability to freely move around the installation environment, the same as with a headphone-based experience, though in these cases this is realised by their proximity to the installation's speaker arrays.

Thinking about the presence of physical speakers placed and used to playback audio content within a physical environment as constituting an AAR installation raises several further questions. Firstly, does this mean that sitting down and listening to some music at home creates an AAR environment? Secondly, can we begin to associate the type of augmentation our perception of reality is experiencing in relation to the type of AAR technology deployed?

In relation to the first question, we can make a clear distinction between experiencing the playback of a piece of recorded music emanating from the speakers of our hi-fi system at home and an AAR experience by thinking of the former as being acousmatic and the latter being visualised or direct sound. Within AAR there the seems to be an attempt to project the acousmatic into the realm of visuality by allowing it to be perceived as believably originating from a source within reality. Within the nature installation presented by Lawton et al. (2020) we see how this is attempted with the remotely recorded sound of bird song, with Gimenes et al. (2016) we see how this is attempted with the placement of a recorded musical composition at specific coordinates within physical space and in Thielen (2018) we see how this is attempted with the music of a remotely recorded string quartet. Additionally, if we think again about freedom and the agency to explore, such experiences appear distinct from the passive *'armchair'* listening experience.

Our second question essentially deals with identifying a link between the type of AAR technology deployed and how exactly this transformation from acousmatic to visualised sound is attempted. Krzyzaniak et al. (2019) tentatively suggest that one of the essential elements of AAR is some form of *'physical analogue sound'* combined with *'some digitally mediated addition or modification to the sound that provides some benefit'.* Whilst I would disagree with the first point (which I assume to mean the requirement of an element of the acoustic reality), the second point shines a light on how AAR can be achieved in practice. If we do consider the inclusion of the ambient acoustic reality (perhaps enabled via a hear-through delivery system such as bone-conducting headphones, active noise cancellation technology or indeed a speaker-based AAR experience) then we can identify such practice as an attempt to realise AAR through *modification* (Krzyzaniak, Frohlich and Jackson, 2019), but also through *addition* or *reinforcement* of the acoustic reality. If we consider an AAR environment deployed via closed-cup headphones then we could say that an AAR is achieved through the *replacement* of the acoustic reality; reality exists as a visual element of the experience, the virtual audio as the perceived accompanying soundscape.

Within the discipline of sound studies, specifically the history of sound technologies, it could be suggested that rather than being unique, original or at least highly distinguishable from other audio experiences, what we think of as AAR is in fact just the next step in the evolution of augmented listening. This audio augmented reality journey is perhaps, most recently, best illustrated by a lineage that includes the *Walkman*, *Discman*, the *MiniDisc* player, iPods and iPhones (Bull, 2000), though is evident through the history and development of listening technologies (Sterne, 2003; Ouzounian, 2020). This could be described as an ability to enable a greater sense of *acoustic virtual reality*, or greater proficiency to augmentation the listener's environment with virtual audio content, then was previously possible, a trait that appears to distinguish each subsequent and significant development of our listening technologies.

We see, emerging here, two different realisations or constructions of an audio augmented reality, one by design and one through inference, the former relying on associations made between virtual and real content by the author of the experience, the latter relying on associations made by the listener.

Firstly, let's consider Azuma et al's (2001) broadly acknowledged (Carmigniani and Furht, 2011) definition of AR where, it is suggested, an AR experience should include: *a combination of real and virtual objects*, *should run interactively in real-time* and *should align real and virtual objects with each other*. We can quite easily confirm that an experience within which the associations between virtual audio content and real-world objects have been pre-determined, or designed, by the author would adhere to Azuma et al's (2001) definition of AR, and would therefore constitute the realisation of an audio augmented reality.

On the other hand, an experience such as a linear mobile listening experience, where no explicit associations between the virtual acoustic reality of the audio content and reality of the listener's physical environment have been authored, presents as a little more problematic. Within such an experience, in order for it to qualify as an AR or, for that matter, an AAR experience, according to Azuma et al's (2001) definition, we'd require an explicit association between the virtual content and the physical reality. For this we need to rely on chance and the ability of the listener to find, make and imagine these associations, which we know, according to Bull (2000) and Aceti (2013) are quite possible, thus making the experience of mobile listening, even a linear stereophonic one, an experience that is capable of realising an audio augmented reality.

It seems important, at least at this stage, to pursue and attempt to nail down what we are dealing with when we talk about audio augmented reality, and this is for a number of reasons. The strength of the meaning and the experience created via serendipitous association reported by Bull (2000) through the use personal stereos needs to be understood if we are to attempt to transcend it with new mobile listening technologies to create new and different experiences. We can then attempt to answer such questions as: *What is different about this latest innovation in AAR? What characteristics does it share with previous experiences? As* Hugues et al. (2011) suggest, if we resist the temptation of '*reducing AR to a set of technical characteristics'* we will be better positioned to understand its full potential.

### 2.1.1 Acoustic virtual reality

Iida (2019) describes *acoustic virtual reality* as being the '*reproduction of an existing sound field'* or '*the creation of a virtual sound field*.' Whereas researcher and acoustician Rozenn Nicol (2020) goes a little further and provide us with a technical description of how you might go about its creation, suggesting that one such approach would be a *'binaural reproduction of the spatial scene'.* Indeed, the primary concern of binaural audio, via the use of HRTF[1] profiles (Iida, 2019), is the creation of an *acoustic virtual reality;* the ability, with varying degrees of effectiveness, to transport the listener to an alternate acoustic reality, whether virtual or a reproduction.

The important point here is the way binaural audio can do this with *varying degrees of effectiveness.* For example, a binaural audio rendering can either utilise a standardised HRTF profile, or a personalised HRTF profile that matches the physical dimensions of the listener's head, the latter resulting in a more reliable three-dimensional acoustical sensation (Iida, 2019). The composition of the binaural sound field could be a binaural field recording of the interior of a supermarket; with the beeps of the tills, people shuffling and coughing, children crying, an announcement over the *tannoy* in the background, all presented to the listener with the acoustic characteristics of a high ceiling and the hard tiled floors that the listener would recognise from such a location. Or, it could be a simple, virtually created spatial scene, with a single bird song and a gentle breeze. The point is, they will both take you

[1] Binaurally rendered audio uses HRTF (Head-Related Transfer Function) profiles that can be standardised or specifically modelled for the listener. These emulate both the interaural time difference (ITD) and interaural level difference (ILD) responses of spatialised sound sources (the difference in time and frequency a sound source is perceived by the listener's two ears) in order to create a realistic perception of a sound positioned in three-dimensional space.

somewhere, just to *varying degrees.* This phenomenon also extends beyond the binaural audio format as it is possible for one to listen to a recording of a live concert in stereo, or even in mono for that matter, and still feel transported to the concert hall or festival, it is possible to imagine being there, just perhaps not with the same degree of realism afforded by a binaural audio recording.

## 2.2 Mobile audio

The predominately stereophonic mobile audio listening experience currently represents one of the most ubiquitous and freely explorative acoustic virtual reality experiences with the ability to augment almost all varieties of physical space. The mobile audio experience is capable of creating explorable physical environments, the listener's perception of which has been augmented with audio.

Probably the most influential of all mobile listening technologies is the *Walkman*, a mobile musical delivery system that enabled people to '*take music and move it through landscapes and architecture'* (Gibson in Bull, 2000). Even within this now seemingly rudimentary linear listening experience, listeners where able to realise profound moments of an audio augmented reality through a combination of place and music (Bull, 2000). For Bull and sound artist Janett Cardiff these moments are inextricably linked to chance, or serendipitous encounters, where a meeting of the listener's real-world environment combines with the virtual audio source to present them with a meaningful and personal experience, an experience that is often described as being '*cinematic'* or '*filmic'* (Aceti, 2013; Bull, 2013).

Even within the scope of the limited controls and storage capacity of the *Walkman*, choices are made and interactions with the technology happen. The choice of what music to select, where to go, or which route to take become ingredients in the creation of the serendipitous. This choice of musical selection is increasingly facilitated by the increased musical storage capacity and musical selection controls of the portable Compact Disc player, the *MiniDisc* player and the *iPod* or portable mp3 player respectively, the control of which is increasingly facilitated by its playback technology; rewind, playback, fast forward; skip, shuffle, loop; *'play my top-rated songs'*. With increased personal control over an increased library of personally selected material one would surmise that the chance of the serendipitous and cinematic encounter becomes greater, along with its ability to effectively augment physical space with, for the main part, a stereophonic *acoustic virtual reality*.

Interestingly, this relationship between AR and the cinematic experience has a historical connection, with Heilig's *Sensorama*, a cinematic experience that attempted to take the viewer into the screen by stimulating all the senses, being cited as one of the first examples of AR (Carmigniani and Furht, 2011)

Behrendt's taxonomy of Mobile Sound (Behrendt, 2015) includes four categories: *Musical Instruments, Sonified Mobility, Sound Platforms* and *Placed Sound*, and it is, perhaps, Behrendt's category of *Placed Sound* that remains of most relevance here. Behrendt describes *Placed Sound* as a mobile listening experience where the listener does not contribute their own sounds to the experience, rather they create their own mix, or their own personal audio experience, as a result of their trajectory through space.

Behrendt defines *Placed Sound* within the context of a GPS enabled mobile, or locative listening experience, where audio content is located at specific geographic locations, where it can be triggered for playback. But this definition, if we consider either the serendipitous encounter, or the purposely constructed experience of meaning by the listener, as referred to by both Aceti (2013) and Bull (2013), also appears to hold true for the linear, manual mobile listening experience that the likes of the *Walkman* and the mp3 player are capable of realising. To choose to play a specific piece, or a specific selection of music, for a specific journey, or at a specific point within that journey, is to *place sound* within a location with which a personalised audio experience is encountered. Similarly, to experience a meaningful encounter between location and audio by chance whilst engaged with such listening technology is to have a personal experience with *placed sound*. There seems, though, one immediate distinction that can be defined by authorship. Behrendt specifically refers to *Placed Sound* as sounds placed by someone other than the listener; sounds that have been placed by someone to be found and listened to by someone else. Whilst the mobile, mp3 audio listener has agency over what audio content to play and where, they do not have the agency to either create or manipulate the nature of the audio content itself.

We therefore begin to see, with the move from a linear mobile experience to that of what Behrendt defines as *Placed Sound* or locative audio, a change in the way in which meaningful encounters between sound and locations are constructed and experienced. Within the linear mobile listening experience, the listener has the agency to construct meaning between audio and location, though with little, if any, degree of manipulation over the delivery of the authored content. Within the locative audio experience, construction of meaning between audio and location is within the hands of the author, though the listener is afforded a greater degree of control of the delivery of that authored content and the composition of their audio experience within the limits of what has been provided to them.

## 2.3 Locative audio

Janett Cardiff's soundwalks, along with other locative listening experiences, extend this *filmic* experience beyond the aestheticizing potential of linear listening's serendipitous encounters. This is typically achieved by purposely

authoring interactive audio experiences which trigger specific segments of pre-recorded audio content based on a listener's GPS coordinates.
The move to located listening constitutes a change in the agency of the construction of meaning, but one that also marks an acceleration in the potential construction of meaning.

Through the direct authorship and association of sound and place, the meaningful experience has been deliberately designed to happen and for connections between content and place to happen (Hazzard *et al.*, 2017). Although perhaps not always as intended by the authors, the direct association of a piece of virtual audio content with a specific real-world location is a product of curatorial, musical or narrative intention and construction, the how or if, a meaningful encounter is experienced is with the ears of the listener. But these types of experiences are site-specific, for the most part non-spatial and most often non-nomadic, meaning that exploration of space is authored and pre-determined rather than freely explorable. Within these experiences we see the user's agency regarding the determination and association of both context and content relinquished in favour of a directly authored association between content and location.

In *The Rough Mile* (Hazzard *et al.*, 2017), where pre-recorded audio is used to augment a specific outdoor location, and Sikora et al's archaeological AAR experience (Sikora *et al.*, 2018), where pre-recorded audio is used to augment locations in and around an archaeological site, could both be categorised as examples of *transformative soundscaping*. In both of these examples, virtual audio sources are used to reframe, rather than to directly compliment, the physical reality. In the case of Sikora et al's AAR experience, this change of context is from rural to urban; in *The Rough Mile*, this change of context is from city centre to fictional narrative. Being outdoor experiences, both rely on GPS technology for determining the position of the user within the physical landscape. In Sikora et al's AAR experience, the listener's GPS coordinate values are plotted on a virtually authored representation of the landscape based on satellite imagery, onto which are placed virtual sound sources for the user to encounter in the real-world.

Hazzard et al. (2017) provide compelling evidence, and testimony from their participants, for the inclusion of situated ambient noise within such an experience, which, in this particular case, was realised through the use of bone-conducting headphones. Participants of *The Rough Mile* generally noted that their ability to hear both the real and virtual sound sources within the experience added to their feeling of immersion, and that it better situated the virtual sound within the physical location. Although there is also evidence to suggest that occasional loud ambient sounds masked the ability to hear the virtual audio through the bone-conducting headphones.

Nevertheless, both of these examples illustrate the ability of a virtual soundscape to alter how a physical environment is perceived and engaged with by the listener.

## 2.4 Binaural audio

As previously outlined, binaural audio presents a sophisticated way of either constructing or reproducing the experience of a *virtual sound field* (Iida, 2019; Nicol, 2020). Binaural audio's ability to achieve this is a consequence of the listener being able to perceive different sound sources within a binaural audio mix as emanating from different and specific positions in space around them, just as they would in reality (Iida, 2019).

In the following section I look at three different types of binaural audio experiences; *static binaural audio*, *dynamic binaural audio* and *six-degrees-of-freedom dynamic binaural audio*. Through the provision of examples and an assessment of their differing characteristics, it is proposed that each of these different types of the binaural audio experience are capable of realising *virtual sound fields* with increasing sophistication and an increasing perception of reality.

### 2.4.1 Static binaural audio

Recent experimental programming in binaural offerings of *Audio horror*, such as Peter Strickland's radio dramatisation of Nigel Kneale's 1972 television drama *The Stone Tape* (Strickland, 2015) *and The Paragon Collective's* podcast *The Darkest Night* (Shudder, 2021), have demonstrated the potential to offer an '*unique', 'alternative'* and '*highly effective'* form of horror (Hancock, 2018). Hancock attributes this to binaural audio's ability to *'afford a uniquely acoustic virtual reality'* that can realise the physicality of sound and offer a more intense experience of uncertainty, shock and immersion. Hancock also suggests that a combination of three-dimensional technology, mobile listening and private headphone use has the potential to realise audio experiences that will rival those of video gaming and cinema (Barrios-O'Neill, 2018; Hancock, 2018).

Cultural historian Johnathan Sterne traces binaural listening technology back to the invention of the binaural stethoscope along with other binaural instruments dating back to as far as 1829. Interestingly, the binaural stethoscope's superior quality and clarity of sound was attributed to, amongst other factors, its ability to isolate the physician from external sound sources and intensify and concentrate the listening experience. The increased mobility that the binaural stethoscope facilitated is also cited as a beneficial feature, and a feature which, along with its ability to isolate and concentrate, it shares with modern headphones which are seen as a prerequisite for an effective binaural listening experience (Hancock, 2018; Karathanasopoulou, 2019; Kiss, Mayer and Schwind, 2020).

Probably the best known, pre-digital example of binaural radio drama is Andrew Sachs' *The Revenge* (1979) (Sachs, 1978). *The Revenge* is an audio play that contains no dialogue and is comprised entirely of binaurally recorded content that situates the listener alongside a man being chased by the police. Seen very much as a piece of experimental programming at the time (Sachs, 1978). In *The Revenge*, as Sachs himself notes in the interview which precedes the broadcast of the radio play, he comments:

> *"…the action is all around. You, the listener, are always at the centre of proceedings and the sounds and the noises, or whatever it is, go right around you. And it's a system whereby if somebody's behind you as the listener, you actually want to turn around and you think he's there, but of course he's not."* - Andrew Sachs on *The Revenge* (Sachs, 1978)

Hancock reflects that this effectively situates the listener within the play's '*acoustic space'* (Hancock, 2018) and this placing of the listener within the space of the play or drama is nicely illustrated by the actual practice of binaural recording. In the case of *The Revenge*, in-ear microphones are worn by the play's principal character to achieve a simulated first-person auditory account of the narrative's unfolding soundscapes. A similar approach was applied with *The Stone Tape* (Strickland, 2015), though in this case the '*dummy head'* recording approach was used, which enables the generation of binaural recordings through the use of a dummy head with microphones embedded within the ears (Jahn, 2015). In either case, the head of the play's central character, and the head of the dummy, become the head of the listener, and therefore determine both the position and orientation, or the *sound perspective*[1], from which the listener will perceive their audio experience.

> *"…the whole narrative thread is carried by Andrew as the principal character in the play, and it's important that we realise that the microphone is with him, and with his adventure from the beginning of the play to its climax."* - Ronald Mason on *The Revenge* (Sachs, 1978)

Although binaural audio offers an increased sense of reality and belonging within the space of the narrative, the experience for the listener remains a passive one. Owl Field's *3D Escape Room: Frequency* (OwlField, 2020) offers an element of interaction within a binaural audio podcast in the form of a non-linear narrative which is made navigable by selecting different podcast episodes titled with different radio frequencies. Although the listener can

---

[1] Crook (1999) and Ouzounian (2020) define *auditory, sound or acoustic perspective as being* the listening *point-of-view* or, rather *point-of-listening,* within an acoustic space specifically in relation to the *point-of-listening* within a simulated or remotely experienced acoustic environment. Additionally, Crook (1999), draws a comparison between *perspective* as the *visual* representation of three-dimensional objects and volumes of space, and *sound perspective's* ability to achieve the same.

determine different outcomes for their adventure by choosing which *frequency* to play at the end of each podcast episode, interaction does not extend to the delivery or manipulation of the binaural audio content itself.

Janet Cardiff's soundwalks, specifically *Her Long Black Hair*, affords the static binaural audio experience not just a sense of mobility, but a greater physicality and locational belonging by incorporating this type of audio content within the context of a location-based experience. As such, this particular work appears to bridge both the *static binaural* and the *locative audio* experience. Set, and designed to be listened to within New York's Central Park, *Her Long Black Hair* consists of six audio tracks that the listener is instructed to play at specific locations on a map depicting a pre-determined route through the park. In *Her Long Black Hair,* the listener is linked to their physical surroundings by the pre-recorded audio content, the narrative and through the use of physical photographs that the listener is instructed to look at in specific locations, and that directly reflect their physical location. Cardiff herself makes this interesting observation:

> *"Audio affects our perception of the physical world. We understand three-dimensional space by using our vision but also by the character of sounds we hear. If these sounds are manipulated and changed, then our perception of reality can be drastically affected."* - Janet Cardiff (KQED-Arts, 2015)

What is of particular interest here is how *Her Long Black Hair* provides just one example of how artists involved in using located, or *'placed sound'* (Behrendt, 2010, 2012)*,* make connections between physical space and virtual sound. Additionally, we also see how Cardiff recognises the ability of sound to change our perception of reality and, within *Her Long Black Hair*, uses physical props to give the virtual audio content a greater sense of reality, or authenticity, within the physical environment.

Although *Her Long Black Hair* undoubtedly pushes the boundaries of the *static binaural audio* experience in that various tools are used in order to attempt to place the listener directly within the soundwalk's *auditory* or *sound perspective*, the listener remains, as Sachs and Mason (Sachs, 1978) point out, at the centre of *somebody else's* proceedings; the *auditory* or *sound perspective* is fixed and not their own, they remain largely a passive listener within the narrative experience, even though this is less distinctive than in the previous examples.

Figure 1. *Her Long Black Hair* (2004) by Janet Cardiff.

### 2.4.2 Dynamic binaural audio

*Static binaural audio* remains the current standard for the delivery of binaural audio-only content, with *dynamic* binaural audio largely remaining a component part of a visual-based AR or VR experience (Hancock, 2018; Bauer *et al.*, 2019). A combination of recent technical developments and changes in listening behaviour, similar to those referred to by Hancock (Hancock, 2018) and Barrios-O'Neill (Barrios-O'Neill, 2018) have, though, rendered possible the opportunity for such audio-only experiences to explore the use of *dynamic* binaural audio.

*Dynamic* binaural audio differs from *static* binaural audio in that it takes into account the bodily movement of the listener in real-time. We can further differentiate between types of dynamic binaural audio here and divide dynamic binaural audio into that which just tracks the rotation of the head of the listener, or having three-degrees-of freedom (3DoF)[1], and that which tracks both the rotation of the listener's head and their bodily position in three-dimensional space, or having six-degrees-of freedom (6DoF)[2]. The absence of embodied interaction with three-dimensional sound is considered a significant obstacle in realising its immersive potential by Begault (2000),

[1] *Three-degrees-of-freedom* (*3DoF)* refers to the freedom of movement in three-dimensional space and includes the three degrees of rotational movement (*pitch*, *yaw* and *roll*).
[2] *Six-degrees-of-freedom* (*6DoF)* refers to the freedom of movement in three-dimensional space and includes the three degrees of rotational movement (*pitch*, *yaw* and *roll*) and the three degrees of translational movement (*surge*, *heave* and *sway*).

who also points out that head movement allows the listener to better localise sound sources by comparing interaural clues.

Most commonly this is achieved through headtracking (Bauer *et al.*, 2019) and delivering spatially rendered audio to the listener through headphones. Using these technologies it is possible to place virtual audio content that the listener can perceive as existing within a specific location in their real-world environment (Bauer *et al.*, 2019).

Audio-only headtracking technologies, such as *Bose AR* enabled frames and headphones (Bose, 2018) and more recently *Apple's Airpods Pro* (Apple, 2021a), have enabled some initial experimental studies and development of *dynamic* binaural audio-only experiences (Bauer *et al.*, 2019). One example of which is *Sonic Samurai* (Bose, 2021), a 3DoF audio game enabled by Bose's AR framework and hardware where the listener has to dodge the incoming attacks from virtual ninjas made evident by encroaching sounds from different directions. The listener then has to turn and orientate themselves to face the incoming enemy attack in order to fend them off. *Sonic Samurai* utilises sensors within the headband of the Bose AR device to determine the listener's orientation, and the user's phone to detect the swings of their virtual Samurai sword. Other examples include *Spatial Scenes* (Apple, 2021d), a mindfulness, meditative audio mobile app that creates virtual audio soundscapes intended to immerse the listener within a dynamic binaural virtual audio scene with 3DoF. Enabled by the movement sensors embedded within *Apple's AirPod Pro* headphones and described as a '*positionally aware audio'* experience, the premise is that, when virtual audio sources that comprise a virtual binaural soundscape stay located in virtual space, the listener has a more immersive audio experience.

Though undoubtedly the more intuitive approach, headtracking remains difficult to deploy within an audio-only experience due to the lack of prevalence, or take-up amongst consumers, of these still early and developing technologies. As a more accessible, or fall-back option, device tracking can be used as an alternative to headtracking (Heller and Borchers, 2014; Cliffe *et al.*, 2020) to create a dynamic binaural audio experience. This is achieved by using both the positional coordinates and orientation of the listener's handheld device, as determined by a combination of the device's internal sensors, as a reference for the listener's position and orientation within a 6DoF experience, or just the orientation of the device within a 3DoF experience. A *dynamic* binaural audio production has the ability to place the listener within their *own* proceedings. Essentially, the listener is no longer the *dummy head*, or the microphones in somebody else's ears, rather they have their own *sound perspective* on the experience.

Tim Crook, in his book *Radio Drama* (Crook, 1999), refers to *sound perspective* as being composed through a process of *choreophonic design*[1], and suggests that, just like perspective in the visual arts, the *sound perspective* is also the representation of three-dimensional objects and volumes of space. He also suggests that the sound artist has the means to create *sound perspective* not just from the point of view of the listener's imagination, but also from their physical experience of the binaural audio. In keeping with Crook's visual analogy, we could perhaps think of the difference between static and dynamic binaural audio as being similar to the difference between a painted scene and a three-dimensional sculpture. In the former the scene's view point has been pre-determined by the artist for the viewer. In the latter the artist has provided a work that can be viewed from many view points and many different angles. We could therefore conclude that within a dynamic binaural listening experience, the listener, within the context of the work at least, can determine their own *picture space* as a result of having involvement in the *choreophonic design* of their experience.

Th previous point also illustrates the difference between dynamic binaural audio experiences with 6DoF and 3DoF. Whilst the 3DoF experience remains a predominately static activity (albeit you can rotate around a single location, or *sound perspective*), the 6DoF experience realises a fully mobile and explorable experience with no fixed *sound perspective.* There is no *sweet-spot* in the 6DoF experience, no privileged listening position, every location has the potential to be a *sweet-spot*, determined as a consequence of authorship, exploration and listener discretion. In relation to the composition of early spatial music, Ouzounian (2020) suggests that audiences were not thought of as being *active* or *productive*, rather as *recipients*. Furthermore, the composer Henry Brant commented that spatial music should consider *'the premise that there is no one optimum position'* (Ouzounian, 2020)

Due to the increase in exploratory movement that better matches natural interactions with sound, along with the greater potential control over the *choreophonic design* of their experience and a multitude of *sound perspectives* to explore and appreciate, we can safely conclude that the *dynamic binaural audio experience* has the ability to realise a more immersive acoustic virtual reality than *static binaural audio*. Similarly, one can determine that a *dynamic binaural audio* experience with 6DoF has the ability to create a more convincing *acoustic virtual reality* than one with 3DoF.

### 2.4.3 Six-degrees-of-freedom (6DoF) dynamic binaural audio

Enabled by the persistent tracking capabilities of modern smartphones (a subject detailed in sections 2.6.3 and 2.6.4), there have emerged a small number of experimental prototype mobile applications and demonstrations

[1] Crook (1999) defines *choreophonic design* as the composition of the listener's *acoustic perspective* or, rather poignantly, as the *'picture space'* for the imagination.

that have managed to realise virtual audio experiences with 6DoF outside of the visually orientated VR experience. Two notable examples of which are *Fields* by *Planeta* (Planeta, 2021) and a prototype virtual audio application created by artist Zach Lieberman (Lieberman, 2017; Stinson, 2017).

Figure 2. Three-dimensional visualisations of virtual sound sources in physical space. In *Fields* (left) and in Lieberman's prototype app (right).

The *Fields* mobile application is described as an augmented reality spatial sound creation tool. It allows users to position a series of pre-defined audio samples, supplied by various sound artists, in three-dimensional space and create dynamic binaural sound installations with 6DoF. These audio samples are placed by a process of selection and then tapping the camera view of the physical environment to instruct the application where to place the audio sample in the real-world. Once placed, the audio is visualised as floating, blossom-like orbs; augmented reality visualisations that pulsate and change colour in time with their associated audio content.

Lieberman's virtual audio prototype creates similar visualisations of the virtual audio content in real space, though these remain as static, three-dimensional representations of the virtual sound wave. In Lieberman's app this virtual audio content is given an even greater sense of space and time in reality by enabling the user to walk around and view this visual, three-dimensional model of the virtual audio content from different angles, and the ability to play it by walking through these audio visualisations. Ingeniously, this virtual audio content can be mixed live, or scratched, by walking back and forth through the visualised audio in a similar way to moving a record back and forth against the needle of a record player.

Both *Fields* and Lieberman's demonstration utilise the SLAM-based tracking and persistent capabilities of *Apple's* AR framework *ARKit.* These specific features of the audio framework ensure that the virtual audio content that has been placed in reality stays in its specified location and can continue to be perceived as emanating from its location whilst the user explores their authored sound environment with 6DoF. The importance of persistence and SLAM-based tracking in the realisation of explorable virtual audio realities is discussed in detail in sections 2.6.3 and 2.6.4 respectively.

The move from a binaural audio experience with 3DoF to one with 6DoF constitutes a move from an experience within a fixed location to a mobile one. Within a binaural audio experience with 6DoF the fixed rotational position is exchanged for any rotational position in three-dimensional space. This difference makes the fully dynamic binaural listening experience (one with 6DoF) intrinsically mobile and more realistic, as the delivery of audio content to the listener responds to both their orientation and their position within the three-dimensional reality.

The dynamic binaural listening experience with a full *six-degrees-of-freedom* within the physical reality constitutes a new type of listening. This active and mobile form of listening entices the listener away from the fixed position of static binaural audio listening, and the fixed position in three-dimensional space of the dynamic binaural audio listening experience with *three-degrees-of-freedom*. With the additional *three-degrees-of-freedom* representing the listener's position in three-dimensional space, the *six-degrees-of-freedom* dynamic binaural listening experience promotes the exploration of the physical three-dimensional reality. The *six-degrees-of-freedom* dynamic binaural audio listening experience remains, for the most part, an experience which is undertaken within virtual reality and the virtual worlds of video gaming (Hancock, 2018; Bauer *et al.*, 2019). Therefore, it is the experience of this type of listening within reality, rather than within a virtual world, that constitutes a new listening experience. However, we should remember that, in reality, we experience dynamic binaural audio with six-degrees-of-freedom every day, this is how we naturally experience the sounds that envelop us. So, it is therefore the natural affordances of the listening experience applied within the context of an acoustic virtual reality that presents as new and full of opportunity.

## 2.5 Constructing audio augmented realities

By looking at these different types of audio experiences and considering audio augmented reality as an experience where virtual or acousmatic sound is experienced within a physical reality, one can see that the AAR experience is one that comes about through a combination of an *acoustic virtual reality* and the *audio augmentation of physical space*. The rendering of an *acoustic virtual reality* is realised by the sound reproduction method's ability to render a realistic virtual audio experience. This is combined with the audio augmentation of physical space which is facilitated by the listening technology, both these factors combine to realise the creation of an audio augmented reality.

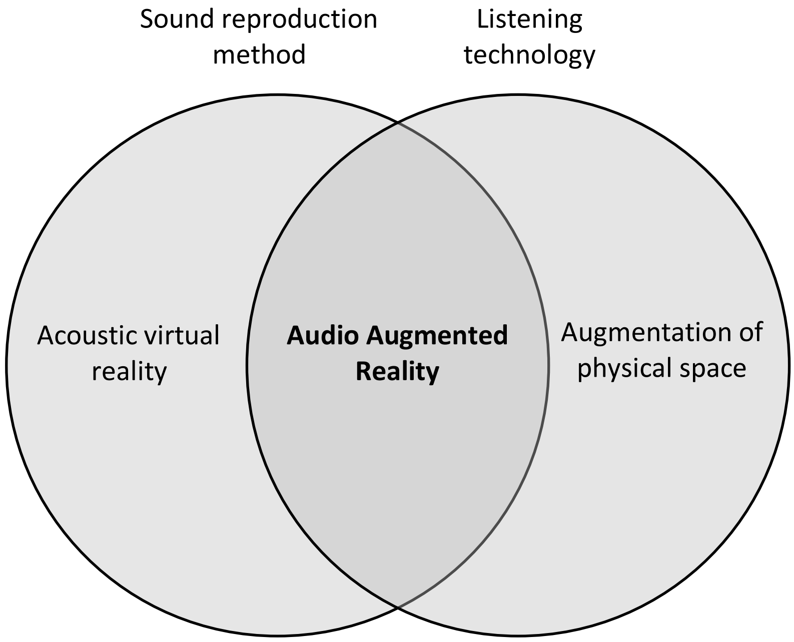


Figure 3: The construction of an audio augmented reality. The combination of the sound reproduction method's ability to create a acoustic virtual reality and the listening technology's ability to augment physical space combine to create an audio augmented reality.

The proficiency to render an acoustic virtual reality can be seen as relative to the innovations within each of these technologies. With the method of sound reproduction, this is illustrated by the experience of monophonic, through to stereophonic, quadrophonic, surround sound, binaural and dynamic binaural audio's increasing proficiency to realistically render a three-dimensional acoustic virtual reality. Within mobile listening technology, this is illustrated by an increasing proficiency to facilitate the augmentation of physical spaces with audio, as illustrated by the Walkman, through to the iPod, iPhone and GPS enabled location-based audio experiences.

So, for example, early domestic radio receivers would have enabled audio augmented realities via a monophonic acoustic virtual reality in people's living rooms, whereas *The Walkman* enabled an audio augmented reality via a stereophonic acoustic virtual reality in practically any location.

This model presented in Figure 3 positions itself with Bull's (2013) suggestion that the mobile audio listening experience is an augmented reality experience. When considering the experience of listening to music on an iPod, Bull (2013) suggests that the sense of place is heightened and that the listener feels a greater sense of *connection* with the environment; the listener's perception of reality has been augmented with audio.

Furthermore, Figure 3 extends this suggestion to include stationary listening experiences as well. The trajectory of the combination of these technologies, from a monophonic static environment to a fully mobile three-dimensional audio environment is illustrated within Figure 4.

Audio technologies that maintain the ability to augment our perception of reality can be thought of as audio augmented reality technologies. Whether by design or by chance association, this is not always guaranteed, but nevertheless the potential for it to happen remains. Additionally, as Hugues et al. (2011) maintain in relation to AR in general, AR technologies can be considered as either those that augment our perception of reality and those that intend to immerse us within an artificial reality.

Furthermore, we can also determine that, depending on the listening technology, the two contributing components to an Audio Augmented Reality are brought together by either design or chance association. For example, within a locative audio experience, the virtual audio content has been explicitly authored to be present and to be experienced within a specific physical space; the association between audio and physical space has been designed. Within a linear mobile audio experience, association is mainly through chance or the self-mediated activities of the listener; the audio content has not been designed for a specific location in space, its association is coincidental.

### 2.5.1 The pursuit of acoustic virtual reality

We can see from Figure 3 that the realisation of an audio augmented reality is a combination of a method of sound reproduction and an audio playback, or listening technology. With an increasing ability to render a realistic, three-dimensional acoustic virtual reality the method of reproducing sound is provided the ability to augment physical space by the audio playback technology, again with an increasing level of ability and sophistication. We can see how this combination has enabled a static monophonic audio augmented reality through to a stereophonic mobile audio augmented reality, and the realisation of what is considered to be the state-of-the-art in audio augmented reality (Bauer *et al.*, 2019); a fully dynamic binaural listening experience with translational as well as rotational freedom, or *six-degrees-of-freedom* (6DoF), made possible by the sensors in a contemporary smartphone and it’s processing capability.

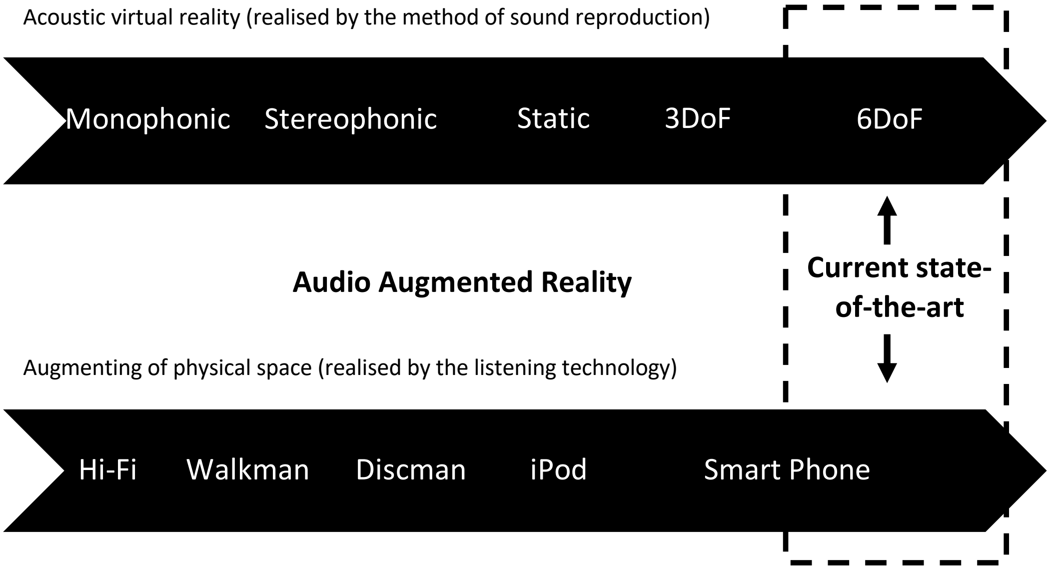


Figure 4. Example sound reproduction methods and listening technologies combining to create increasingly sophisticated audio augmented realities. On the right we can see the current, state-of-the-art, smart phone enabled, binaural listening experience with six-degrees of freedom (6DoF).

It is also worth noting that the example acoustic virtual realities, along with the listening technologies, presented in Figure 4 are largely contemporary, consumer-level technologies. The endeavour to realise acoustic virtual realities are also present within other, what could perhaps be considered as chronological out-lying technologies, experimental demonstrative technologies, or indeed other perhaps less ubiquitous consumer-level technical solutions. These could include: *Ambisonics*[1], surround-sound cinema systems, 5.1 and quadrophonic speaker arrays, and experimental scientific and industry demonstrations, a good example of which would be Bell Laboratory's *Oscar* (Ouzounian, 2020).

Although positioned below what is identified here as *the current state-of-the-art*, the trajectory towards realism is also illustrated by *Dolby Atmos*[2] and *Apple's* recent application of it within their spatialised video and musical content (Apple, 2021b). *Apple's* declaration that *'First there was mono, then stereo, and now there's Spatial Audio'* not only illustrates the increased desire to represent virtual space, but it also infers that there is a desire to consume it.

What is also important to consider within the trajectory towards realism presented in Figure 4 is that the advancement of the perception of acoustic

[1] Ambisonics is a method for recording and playing back three-dimensional audio. It was invented in the 1970s but never commercially adopted. It is currently used as a solution for embedding 360° ambient audio into VR experiences and video games.
[2] Dolby Atmos represents a development in surround sound technology by adding audio channels that deal with elevation, allowing sounds to be positioned three-dimensionally. Initially developed for the cinema market it is currently being marketed commercially.

virtual reality can happen within a method of sound reproduction and listening technology and not just as a consequence of the evolution of one of these technological factors. Typically, this is achieved through the use of reverb (convolution reverb is a particularly useful reference point here) and the desire to place sound within virtual sonic environments and the creation of virtual sonic environments within the sound production process.

Sterne (2003) makes this interesting point when he considers the role of the audio engineer within the pursuit of acoustic virtual reality:

> *'To understand the aural dimensions of virtual reality, we need to consider audio engineers' century-long obsession with creating what we would now call virtual acoustic spaces in recordings.'* (Sterne, 2003, p. 338)

But, as suggested by Apple (Apple, 2021b) and other consumer orientated technological innovations of the last few years, such as *Bose's* AR frames and headsets (Kiss, Mayer and Schwind, 2020) and *Dolby Atmos,* one would conclude that there is also a desire demonstrated by consumers to pursue and explore the evolving acoustic virtual reality. The desire to go beyond the serendipitous cinematic offerings of the stereophonic mobile audio experience, as described by Bull (2000) and Aceti (2013), appear evident.

The construction of meaning in terms of both authorship and design, and listener association and inference afforded by the ability to audibly augment the physical reality with near centimetre accuracy would appear to offer new experiences in acoustic virtual reality.

The introduction of individual objects directly augmented with virtual audio content that this technology enables constitutes an innovation in the current state-of-the-art AAR experience. In other words, the refined ability of the listening technology (in this case AR capable smartphones) to augment the physical environment, combined with a 6DoF binaural audio acoustic virtual reality, has the potential to evolve the current, state-of-the-art audio augmented reality experience through the realisation of audio augmented objects.

## 2.6 Audio augmented objects

We have seen, amongst the previously discussed literature and projects, the invocation of imaginary sound sources, rendered in audio for the aural imagination; the encroaching ninjas in *Sonic Samurai*, the spatially positioned voices of visually absent actors and props in the binaural audio dramas and the imaginary sources in the soundscapes of *Spatial Scenes.* Whether static, dynamic, 3DoF or 6DoF binaural audio, all of the sound sources in these listening experiences remain in the mind's eye.

An audio-focused study presented by Sodnik et al. (2006), where spatial sound is added to visually rendered objects in an augmented reality scene in order to determine the users' ability to localise the virtual objects, gives us some important glimpses into the potential of audio augmented real-world objects.

Firstly Sodnik et al. confirm that sound sources can be effectively located by listeners, with what they describe as a high degree of accuracy, even without personalised HRTF profiles. The ability of users to successfully localise the sound sources without personalised HRTF profiles indicates the effectiveness of binaural audio within an AR application in this regard. It also suggests that its effective inclusion within an AR experience can be achieved without the need for any technical calibration by the user, which also makes it an accessible addition to the AR experience.

Additionally, the authors note that, in the first instance, users utilise sound as their method for localizing the object and sound source, and then confirm this initial judgement with the provided visual cue of the virtual object itself. When applied to audio augmented real-world objects, rather than virtual objects, such findings indicate that virtual audio can be used to create an awareness of the presence of physical objects. But there is also an interesting relationship between the virtual audio and the real-world object in promoting the exploration of both the virtual soundscape and the physical environment. As presence can seemingly be suggested with sound and confirmed with vision, it seems that perhaps through expectation, curiosity or, indeed, considered design, a user's attention can be engaged enough to encourage movement towards confirmatory visual engagement with audio augmented objects that are perhaps, initially, beyond the user's visual periphery.

Whilst Sodnik et al's study hints at the immergence of the audio augmented physical object and its potential, the audio augmented physical object is something that has, it appears, come to fruition within other projects, largely as a result of such objects being combined with dynamic binaural audio.

Furthermore, the previously discussed transformative soundscaping of Sikora et al's audio augmented archaeological project could easily give way to the realisation of an audio augmented physical object, the change from location of an imaginary object to the location of a real-world object is of little significance to the applied technology, other than the granularity of augmentation which is afforded to us by a change from GPS enabled localisation to SLAM-based localisation. The change from imaginary object to real object in such examples appears as more of a design decision. Though, as Sodnik et al's study suggests, this difference between a sound in a location in space and a sound in a location in space that is occupied by a physical object has potential wide-reaching implications for designers and users alike.

What's interesting is that the direct augmentation of the object is not necessarily what is of immediate interest within the scope of the projects' that have realised audio augmented objects, rather what seems of importance within the context of such projects and studies, is the overlaying or triggering of audio content based on some mode of interaction with a physical object. In other words, focus has been on the association of virtual audio content with a physical object, rather than positioning the physical object as a potential or probable source of the virtual content.

In Montan's thesis regarding the implementation of an early AAR system within the context of a gallery environment (Montan, 2002), binaural audio content is positioned relative to a photographic print mounted on the gallery wall. In addition to the descriptive audio content that the photograph was augmented with enhancing engagement with the picture, it was also observed that it effectively advertised the position of the photograph to the listener prior to its visual discovery. Montan also recounts how the authoring of what they describe as *acoustical zones*, or zones of audio within the space of the gallery, rather than attached to specific objects, gave the listener an indication of what they could expect within the physical space of the exhibition as the explored it through bodily movement. These observations that provide interesting initial insights into the possible function of audio augmented objects and the relationship between physical object and binaural audio are further enhanced by Montan's use of *sound zones*. *Sound zones* describe zones of space that changed the acoustic treatment of the audio content delivered to the listener based on their proximity to the photographs. One example describes how moving towards a picture changed the acoustics of the binaural audio content that was positioned in the same location as the picture to one that was perceived to be within a small room to one perceived as being outside. This, it is suggested, gave the listener a sense of entering into the picture, which, incidentally, depicted an outside scene.

Zimmermann and Lorenz's LISTEN system (Zimmermann and Lorenz, 2008) provides an excellent example of the capabilities of AAR within the context of a cultural institution. The LISTEN project, which they describe as *'an attempt to make use of the inherent everyday integration of aural and visual perception'*, delivers a personalised and interactive location-based audio experience based on an adaptive system model. It does this by tracking aspects of the visitor's behaviour (which artworks have been visited, how long were they visited for) to assign the visitor a behavioural model and adjust the delivery of audio content accordingly. The LISTEN system relies on a substantial technical background infrastructure to realise this personalised and invisible technical front-end experience for the visitor, who can wander freely through the exhibition space with just a set of customised headphones. Like Montan (2002), Zimmermann and Lorenz also suggest the concept of the attractor sound, which, based on the visitor's personalised profile model, suggests other nearby artworks to the visitor that may be of interest to them via spatially located audio prompts. Furthermore, LISTEN characterises many of the key differences between the usual audio guide experience and an interactive, adaptive and immersive approach. This includes binaurally rendered, three-dimensional surround sound based on the listener's movement and the delivery of related audio content based on the listener's proximity to an exhibit. The authors report that two-thirds of participants rated their experience with the LISTEN system as being 'enriching', and clear positive feedback was gathered in relation to the combination of artwork and auditory information realised through the system.

Finally, in two exploratory studies where physical, everyday objects have been augmented with three-dimensional, virtual audio sources (Yang and Mattern, 2019; Yang, Frank and Sörös, 2019), we are provided with evidence that suggests users are able to accurately localise virtual audio sources, and therefore the location of audio augmented objects, via off-the-shelf headphones. In these studies, this is achieved through the appropriation of video game authoring and audio spatialisation software, and also indicates that effective and accurate localisation of audio augmented objects can be achieved while walking.

### 2.6.1 Function, character and experience

Within the field of HCI, in the work of Sodnik et al. (2006), Montan (2002) and Zimmermann & Lorenz (2008), we are provided with some tantalising glimpses into the potential function, character and experiential qualities of augmented objects that deserve further exploration.

Within all these projects and studies, we see how a combination of spatially positioned binaural audio and a physical object can, and does, create an augmented reality environment within which the presence of a physical object can be perceived prior to being seen.

Furthermore, we can see the potential for using audio augmentation of space, as well as object, to create a sense of expectation. This expectation, which can be resolved or affirmed by visual confirmation of the presence of the physical object itself, can also be used to lead listeners through, or to suggest potential trajectories around, physical space. As such, through the audio augmentation of object and space, there appears the potential for these audio augmented reality environments to function as interfaces with which to explore and engage with both the real-world environment and virtual audio content.

Specifically, within the work of Montan (2002) and Zimmermann & Lorenz (2008), we see how audio augmentation has the potential to increase engagement with real-world objects, with Montan's work indicating that this engagement can be both multi-layered and a potentially intimate experience.

Lastly, it also appears such environments can be authored in an accessible and intuitive manner, forgoing the requirement for any technical setup, calibration or training on the part of the user.

### 2.6.2 Constructing an audio augmented object reality

Given the model for the construction of an AAR environment outlined in Figure 3, it would seem possible to realise an audio augmented object and construct an audio augmented reality with them. Within such an experience, physical objects would represent points in physical space and combine with a dynamic binaural acoustic virtual reality with 6DoF to realise an audio augmented object reality.

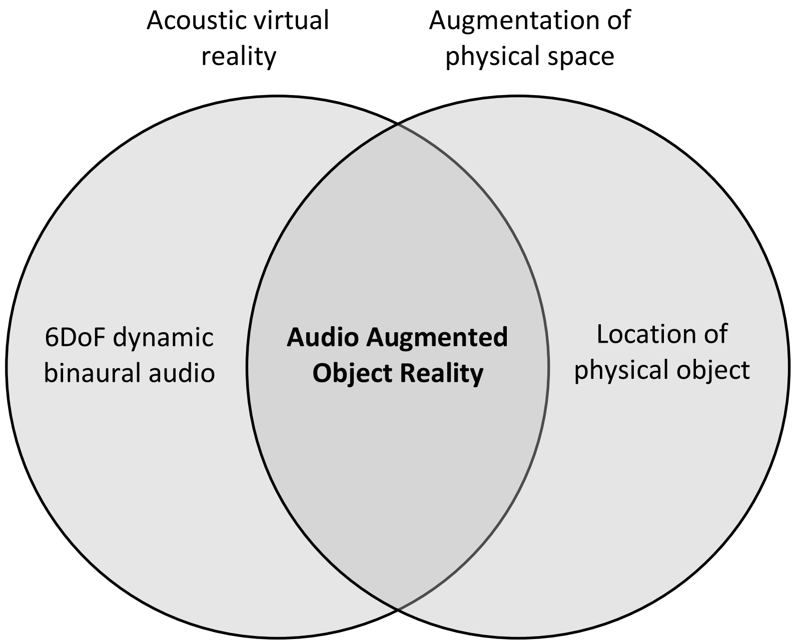


Figure 5. The construction of an audio augmented object reality. A product of a carefully authored association between the binaurally rendered acoustic virtual reality and the location of the physical object.

Within this combination we can see the important role that the robustly persistent virtual model plays in ensuring that the object is perceived as the location of the source of the virtual audio content, and can continue to be perceived as such within the three-dimensionally rendered binaural audio environment, or the *acoustic virtual reality.*

Unlike the creation of an audio augmented reality (see Figure 3), an audio augmented object reality is a product of design, not chance. The two component parts, the *acoustic virtual reality* and the *physical space*, are associated through careful authorship involving the specific location of the physical object within real-world space, and the relational position of the virtual audio source within the binaurally rendered acoustic virtual reality. It should be noted that this model assumes, and relies upon, a stationary real-world object in order to effectively author and maintain the illusion of a physical object as a source of the virtual audio content.

In order to realise individual audio augmented objects there appears to be some common requirements of the applied technology. These requirements can be identified within the previously discussed examples as being the robust and effective mapping of the physical environment, and the effective localisation of the user and the virtual content within this mapped environment. Within the following section I outline why reliability and persistence are important features and why they seem specifically important to the realisation of audio augmented realities and audio augmented objects, over more traditional, or visually orientated, augmented reality experiences.

### 2.6.3 The importance of persistence

Within the context of AR the concept of persistence can describe two differing, though equally important, functionalities. Firstly, we can prescribe the AR experience's ability to allow virtual content to remain at its assigned physical location, beyond the camera's field of vision, as being *persistent.* This functionality allows users to return to the location of virtual content and find it in situ after it having left the camera's field-of-view. Secondly, the AR experience's ability to re-render a previously authored AR environment; allowing the author of the augmented reality to *save* the virtual environment, and the user to *load* the authored virtual environment at another time, is referred to as a *persistent* AR experience.

This second definition of persistence, which describes the concept of persistent AR environments, provides both AR and AAR experiences with the highly desirable and powerful potential of scalability and is discussed further in chapter 8.4. The first concept of *persistence*, which is sometimes referred to as *extended tracking*[1] and which describes the ability of individual virtual objects to remain in situ within the current AR environment, is, for the AAR experience, crucial rather than desirable.

In visually orientated experiences this can be illustrated by the difference between an AR game, where characters and game objects are required to remain in their current positions to allow for a large scale interactive and collaborative virtual environment, and a virtual animated character that is triggered only when the user scans a printed advert in a magazine. Within an AAR virtual model, where we are attempting to create a acoustic virtual reality though the use of spatialised and binaurally rendered audio sources, this notion of persistence remains integral. This issue delves directly into the

[1] *Extended tracking* is a term sometimes used to differentiate persistent virtual objects from persistent virtual experiences, the idea being that the location of the virtual content is tracked, or extended, beyond the camera's field-of-view, or persists when it is no longer within the camera frame. See: https://www.wikitude.com/augmented-reality-extended-tracking/

different way in which we perceive sound and vision and their differing physical properties. Just like the device's camera, we can only see what we are looking at, though we are able to hear all the sound sources within our locality (within the natural physical properties of acoustics) irrespective of our orientation. Therefore, within the construction of an audio augmented reality, where one is concerned with the construction of a acoustic virtual reality, the question of virtual object persistence is crucial. In essence, in order to perceive spatially rendered virtual audio sources as spatial, we need them to persist within their desired locations regardless of the camera's field of vision, this way it is possible to construct an explorable acoustic virtual reality where virtual sound sources can be perceived as emanating from specific real-world locations regardless of the orientation of the user, just as we do in reality.

The functional importance of this within the AAR experience is illustrated well by Zimmermann & Lorenz's concept of the *attractor sound* (Zimmermann and Lorenz, 2008) as well as by Behrendt's category of *placed sound* which is described within her taxonomy of mobile sound art (Behrendt, 2015). In the former we see how a virtual sound's persistent location in reality has the potential to *attract* a listener's attention towards its location and the location of the object that has been augmented with the virtual audio content. With Behrendt's *placed sound* this persistence is more broadly associated with the realisation of a mobile and explorable model for interaction design, specifically a persistent virtuality that is explorable through walking.

### 2.6.4 Simultaneous Localisation and Mapping (SLAM)

Simultaneous Localisation and Mapping (SLAM) describes the process of constructing and updating a map of an arbitrary location while tracking the location of the virtual objects within it. Within the scope of AR it refers explicitly to a technology's ability to create a virtual map of the AR experience's physical environment and realise *extended tracking*, or *SLAM-based tracking*. This is primarily achieved by obtaining and collecting an array of feature points of the given setting that can be used to localise and re-localise the position of both the user and the virtual content within the physical environment.

The focus on the augmentation of individual objects within an indoor environment of the research presented here denotes the need for a highly precise Indoor Positioning System (IPS). Whilst outdoor positioning systems rely largely on GPS, this technology becomes unreliable in an indoor environment (Lymberopoulos *et al.*, 2015) thus prompting alternative solutions for the reliable tracking and positioning of users and the location of virtual content within indoor spaces.

In the most part, IPS's use a fusion of different technologies in order to achieve differing levels of satisfactory positioning results, and what technologies are used within this fusion is largely dependent on the type of environment in which such a system is being deployed and the device through which the person will be detected. He & Shin (2018) provide us with study results which illustrate that a smartphone-based IPS that uses camera-based tracking has the ability to realise a low-cost and highly accurate IPS over other options such as Wi-Fi, Bluetooth, FM and geomagnetic sensing. Furthermore, we can see in the work of Cervenak & Masek (2019) that centimetre accuracy is achieved through the use of SLAM, camera-based smartphone sensing as an IPS.

Due to its ability to determine the location of the camera (as well as the user if deployed in a handheld or body mounted situation) SLAM technology provides a way in which virtual sound sources can persist and be spatially determined beyond the view of the camera frame, and an indoor positioning system (IPS) within which the location of the user and the virtual content can be determined, potentially with a high degree of accuracy, within both the virtual and physical environment.

## 2.7 A context for practice and study

The previously outlined work by Montan (2002) and Zimmermann & Lorenz (2008) speak to the potential of the gallery and museum environment for the practical application and study of audio augmented objects. There are also other early audio augmented reality projects in which such environments have been identified as fruitful contexts within which to study the combination of virtual audio content and physical object, Bederson (1995) and Bijsterveld (2015) being two such additional examples. Whilst previous AR interventions within these environments highlight the potential curatorial and institutional benefits for AAR interventions within these contexts (Thiel, 2014).

Within such environments we also find, for the main part, stationary objects; objects with a fixed position within indoor architectural space. Such an environment enables us to realise audio augmented objects through the previously identified combination of a 6DoF dynamic binaural audio acoustic virtual reality and a fixed position in physical space (see Figure 5). It also allows for the concentration of focus on the mobile listener, space and object, without the possible distraction of mobile objects. In addition to this an indoor gallery environment provides a place where a robust IPS can be achieved through SLAM-based tracking (as outlined in section 2.6.4), in order to realise a system within which both the locations of the listener and the augmented objects can be persistently localized and tracked.

Curator and author Caleb Kelly (2017) reminds us that gallery and museum spaces are largely organised around the understanding of visual primacy, even to the extent that the architectural design of such spaces are conceived to present visual exhibits. Within such contexts, where visual objects are organised and confined by line-of-sight, considering Montan's (2002) and

Zimmermann & Lorenz's (2008) suggestion that audio augmented objects have the ability to advertise their presence prior to being seen illustrates the appropriateness of such contexts for finding out how audio augmentation changes things. The museum or gallery enables this at a spatial and architectural level, as well as at an object level.

Findings within such a context could have direct curatorial consideration; introducing methods and practices for curating sonically as well as visually, and for designing visitor experiences with objects and space based on the affordances of sound.

The recognised potential of audio augmented objects and space to engage audiences with audio content via their movement through these augmented environments (Montan, 2002; Zimmermann and Lorenz, 2008) provides interesting opportunities in light of some recent projects involving the digitization of existing analogue audio archives.

The *Unlocking Our Sound Heritage* (British-Library, 2019) and the *LARM Audio Research Archive* (Mortensen and Vestergaard, 2014) projects present two examples where existing and significant quantities of analog audio archival material have, and continue to be, digitised for the purposes of preservation and improving access. It would seem that, within museums and galleries, opportunities await for marrying up this digitised archival audio content with appropriate and relevant objects and artefacts. Through further investigation of this apparent mutually beneficial relationship between object and audio, possible approaches to interfacing with both objects and audio archival content could be determined.

Lastly, audio-only augmentation, within the context of a gallery or museum where we are concerned chiefly with the display of visual artefacts, presents a possible opportunity to augment without visual distraction or interference.

## 2.8 Summary

So, it would seem that binaural audio represents just one way of creating an *acoustic virtual reality*, albeit a sophisticated one, and *acoustic virtual reality* is an experience that we have been familiar with certainly since the advent of recorded sound, though potentially it precedes this too.

*Acoustic virtual reality* is therefore distinct from *audio augmented reality* in that *audio augmented reality* is the merging of virtual audio content with reality, and *acoustic virtual reality* is the ability of that virtual audio content to take you to another acoustic reality, or transport you to *another place.*
The creation of an *audio augmented reality* can happen in two possible ways; by explicitly authored association, or by chance-based listener association. Either way, it is possible that meaningful, mixed-reality experiences can occur through the combination of virtual audio and the physical reality.

Additionally, we can see that these types of experiences can take place with a variety of different listening technologies and, over time, technological innovation has enabled ever more realistic and vivid experiences in virtual acoustic reality. As such, it would appear that the realisation of an acoustic virtual reality with 6DoF represents the latest innovation in audio augmented reality, and the realisation of an *audio augmented object reality* has the potential to further contribute to and extend this experience.

We can also identify the technological components required to construct an *audio augmented object reality* along with how and why these are specific to the creation of an *audio augmented reality*, or indeed, an *audio augmented object reality* as opposed to a visually orientated AR experience.

Through some of the early explorations into AAR that have realised the experience of an audio augmented object, we can begin to identify some of the seemingly unique characteristics and functions of these objects and their relationship to the listener, the virtual audio content and the physical environment.

Finally, by identifying the common context within which much of the previous and innovative work with audio augmented objects has been conducted, and by recognising many of its unique characteristics and potentials, an appropriate context for both the practice and study of audio augmented objects can be determined.

## 2.9 Research questions

An acoustic virtual reality with 6DoF binaural audio combined with the mobility of the smartphone represents the latest state-of-the-art audio augmented reality experience. In addition to being the latest innovation in the pursuit of an acoustic virtual reality (as outlined in Figure 4), the addition of audio augmented objects, that is real-world objects that are augmented with virtual audio content, offers potentially new experiences within audio-centric mixed reality. As such, my initial research question is:

1. *What are the characteristics and experiential qualities of dynamic binaural audio augmented realities that contain audio augmented real-world objects?*

Secondly, it seems prudent to gain a greater understanding of the function of audio as a tool for promoting engagement with physical objects through their audio augmentation, as identified by Sodnik et al. (2006), Montan (2002) and Zimmermann & Lorenz (2008). Whilst this first point seeks to explicitly address the apparent extension of presence afforded real-world objects via their audio augmentation, we should also seek to explore the function the real-world object can take in promoting engagement with the virtual audio content, as, again, suggested in the work of Sodnik et al. (2006), Montan (2002) and Zimmermann & Lorenz (2008). Furthermore, given this formerly identified, and potentially substantial functionality, it appears sensible to ask more broadly:

2. *What are the potential functions of audio augmented objects within dynamic binaural audio augmented realities?*

Finally, although the chosen context of the gallery and museum focuses attention chiefly upon cultural engagement within these environments, the main concern of generating new forms of engagement with existing cultural assets, real or virtual, would seem transferable across other cultural contexts.

Whilst it appears straight forward that practice and study within this context has the potential to generate theory that can inform best practices for curators and artists alike, investigation into the object and audio relationship could also inform the creation of other listening experiences that seek to culturally engage listeners in new ways. As such, we can attempt to determine:

3. *What are the best practices for realising culturally engaging dynamic binaural audio augmented realities that contain audio augmented objects?*

In summary, the research presented within this thesis seeks to gain a greater understanding of the contribution of audio augmented objects to the audio augmented reality experience. This includes both how they are experienced by listeners, as well as how the listener experience can be authored by interested parties. It is envisaged, given the scope and breadth of existing audio augmented reality experiences, that these interested parties would include museum and gallery curatorial staff, artists, installation designers, audio game designers, musicians, composers and producers of audio-based content seeking to create innovative experiences. Furthermore, the work in this thesis intends to inform the design and realisation of audio augmented object realities via a thorough examination of their attributes and potential functions within the acoustic virtual realities they create.

# 3 Methodology

A practice-based approach puts sound at the centre of the research; the writing is not just *about* sound; it is also a product *of* sound. The studies, findings and conclusions presented here are a direct consequence of sounds heard and interacted with and their relationship to the physical environment within which they are experienced. For Lingold et al. (2018) such an approach (alongside other examples of practically applied digital enquiry into sound) presents as part of a long-awaited departure from text-centric enquiry within the humanities and challenges us to attend to *how* we hear, as well as *what* we hear.

Attempting to find out what the experiential qualities and characteristics of audio augmented objects are within the setting of an audio augmented reality experience, required methods and approaches that would, in the first instance, facilitate the realisation of these types of experiences to study, and then provide a means through which theoretical findings could be identified. The following sections describe the roles of the *practice-based* and *research-through-design* approach, along with the *ethnomethodologically-informed design ethnography,* that was engaged with in the delivery of the research presented within this thesis. The suitability of the artist as ethnographer is also outlined, as is the suitability of an ethnographical perspective in relation to the study of sound technologies. Finally, I describe how my approach and methods needed to be adapted due to the COVID-19 lockdown.

## 3.1 Practice-based research-through-design

It is intended that the developed methodology is also an appropriate companion to the creative, interdisciplinary and exploratory nature of the practice outlined in section 4.1.

In the first instance, Benford et al's assertion that creative practice is capable of generating generlisable theory, due to the experimental nature of the practice which leads to concepts that may otherwise have not been considered, would appear to fit well (Benford *et al.*, 2013).

Whilst this approach advocates the use of experimental, or perhaps more appropriately *exploratory,* practice, it is Benford et al's observation that such practice can lead to *generalisable theory* that seems of upmost importance in considering a practice-based approach in this instance. It is this observation that confirms the trajectory of the output of exploratory creative practice (as outlined in section 1.2). It is also this observation that indicates the interdisciplinarity of such creative practice; where exploratory creative practice is attributed with the potential to pave the way for the development of thinking and the application of technology across disciplines and professions.

Such an approach very much echoes LeWitt's declaration that artists are able to arrive at conclusions that logical progression cannot (LeWitt, 2003), as well as Youngblood's observation that it is essential to progress that new technology is put into the hands of artists (Youngblood, 1970).

Within this methodology, the creative practice has designed, developed and realised settings within which the audience experience and the characteristics of audio augmented objects can be studied, and these settings have taken the form of interactive sound experiences containing one or more audio augmented objects. As described by Gaver & Bowers (2012) in relation to the *research-through-design* approach, the research is in the writing about the artefac*t* (in this case the *designed experience*) and not the artefact itself. Though the writing requires the artefact, in the first instance, to come about, it also requires it in order to communicate, make sense and meaning. The artefact, and the practice that created it, can therefore be considered as the catalyst for the research.

Qualitative data has been gathered from participant interactions with these interactive sound experiences which has taken the form of video recordings of participants' direct interactions with the installations, audio recordings of structured interviews with participants regarding their experience, audio recordings of freely-offered verbal feedback from participants along with their written feedback. Observational field notes taken by myself, the researcher who was present during these audio augmented reality experience deployments, were also collected.

This approach can be aligned with what could be described as *performance-led research in the wild* (see Figure 6); a cyclical process where practice contributes to theory, and theory, in turn, contributes to practice (Benford and Giannachi, 2012; Benford *et al.*, 2013). Similarly, we see this research feedback loop present within the practice-led approach described by Haseman (2007) as *action research*, who identifies this cyclical process of enquiry as one of the approach's most '*serviceable features'*.

Within the *performance-led* research approach described by Benford et al. (2013) potential theories and frameworks can be formulated and applied to future public-facing deployments which are grounded in the effective analysis of the results of '*recognised research methodologies'*, the end-goal of such an approach being the formulation of broader and generalisable theory. Benford et al. (2013) also suggest that many of the techniques that artists adopt within the creation and deployment of interactive artworks can prove useful within the development of other types of experiences as the '*craft knowledge'* artists develop over years of practice can provoke different approaches to thinking about interaction design. Within such an approach, what is of ultimate concern is the generation of theory as a product of the creative process and application of the research-related project. This approach is also recognised by Gaver & Bowers (2012) as a potential route to the discovery of generalizable theory within a field.

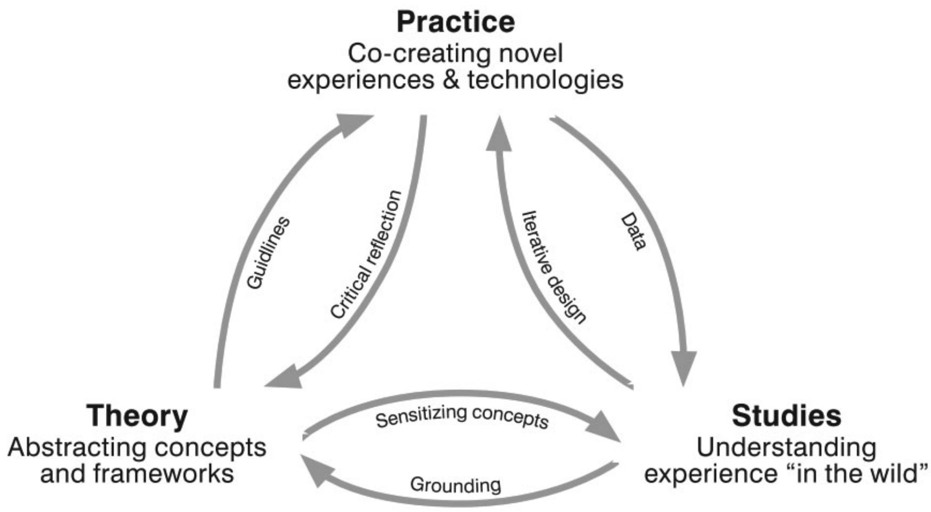


Figure 6: From practice to theory in an artist-led research approach (Benford and Giannachi, 2012).

Additionally, as the practice of developing prototypes and realising deployable experiences required working across the arts and computer science, it has meant that learning can take place not just from participants, but also from my own repeated experiences with the projects, and via a process of self-evaluation, informed by the dual nature of the practice. A similar complimentary approach is identified by Taylor et al. (2011) who suggests that the personal experiences of the researcher as participant play an important part within this cyclical and iterative design process as it can help, in combination with the analysis of each design iteration, to identify emerging and embedded trends that could potentially compromise either artistic integrity or effective HCI (Human-Computer Interaction) practice. Additionally, Taylor et al. (2011) point out that this dual design and participatory role can lead to dialogue and exchange with users that would ordinarily be absent, and can gain insights that may be unobtainable in a traditional design process, both of which having the potential to drive development forward. Indeed, this approach is described by Benford et al. (2013) as having three distinct perspectives: *practicing, studying* and *theorizing*, through which the individual will need to navigate and adopt appropriate positions, specifically in the case of the *artist-researcher,* and *'shifts in perspectives'* to suit relevant activities in order to maintain the integrity of both the practice and the research.

It should be mentioned that this seemingly fruitful and mutually beneficial approach does not come without compromise. As Benford et al. (2013) suggest, developing installation based, and mobile artistic experiences in the real-world, for real audiences, requires working to real-world schedules, venue and festival timetables, and the access and time constraints and deadlines associated with the possible platforms that can be utilised to publicly deploy such projects. In relation to this, Taylor et al. (2011) suggest that this may, at times, require compromise, and the complexity of a specific design iteration may need to be adapted to facilitate an effective deployment, which constitutes a trade-off between the ease and importance of the deployment and the speed of the project's development.

Whilst Benford et al. (2013) and Gaver & Bowers (2012) both outline the importance of maintaining artistic integrity within such a practice-based approach, they additionally warn of the dangers of the forced application of theory on the creative process. This point is made in relation to the fear of losing that which makes a practice-based approach an important contributing element to the research. Whilst all efforts have been made to maintain the artistic integrity of the creative works presented within this thesis, I can confirm that the constraints and pressures of real-world deployment have, at times, led to compromises regarding the inclusion of additional and desirable functions that have been recognised as a consequence of the development process, though I'd maintain that this has not compromised the creative element of the experiences. As such, these identified, though yet to be deployed and studied, functionalities are presented and discussed in chapter 8.4.

## 3.2 Ethnomethodologically-informed design ethnography

Examples of this artist-led and practice-based approach can be found within some of the related works presented in section 2 such as: *Sonic City: The Urban Environment as a Musical Interface* (Gaye, Mazé and Holmquist, 2003), *Designing from within: humanaquarium* (Taylor *et al.*, 2011) and *Psychogeographical Sound-drift* (Chaparro and Duenas, 2015). Explicitly within the work of Gaye et al. (2003), we see this iterative and developmental creative process combined with ethnomethodologically-informed ethnographic studies to provide insights into the user experience of the deployed creative and technological work, fulfilling Benford & Giannachi's (2012) assertion that the development of theory should be grounded in the effective analysis of the results of '*recognised research methodologies*.

The two studies presented within this thesis constitute ethnomethodologically informed *design ethnographies* (Crabtree, Rouncefield and Tolmie, 2012) in that they concern themselves directly with the interactional work undertaken by participants within the settings of the designed sound installation experiences. This is approached from the perspective that understanding the setting of the work is arrived at from the study of the setting itself, rather than the application of external theories or processes (Crabtree, Rouncefield and Tolmie, 2012).

All the presented qualitative data was obtained from the recorded observations of participants' interactions within these settings and then used to inform the design of subsequent experiences, or presented with the intention of informing the design of future experiences.

Within the research presented here, the *artistic practice* has created the opportunities to collect data and the *approach* has ensured that this data is relevant to the users and settings of primary concern and provided a framework within which theory can be identified and put into practice. In addition, the analytical perspective of ethnography has provided a means by which this data can be analysed in order to ground the formulation of theory, as well ground the design of the experiences within a *design ethnography* (Crabtree, Rouncefield and Tolmie, 2012). Therefore, as suggested by Benford et al. (2013), this methodology attempts to accommodate ethnomethodological study and theorising within the *performance-led* research approach.

Therefore, within this thesis you will find the mainstays of ethnographic analysis applied to the collected qualitative data. These include ethnographer *competence* of both the designed experience and the setting, *thick descriptions* based on *vivid exhibits*, *praxiological accounts* of documented interactions, and an outline of *the machinery of interaction* in order to inform the ongoing and future design of such experiences within the studied settings. All this intends to ground the theoretical findings within the analytical perspective of a *design ethnography*.

Potential recipients of this research were engaged in the early stages of the development of each of the deployed experiences and these individuals participated within the studies, feedback their thoughts and offered reflections on how the designed experiences may be useful within their practice. Whilst this data proved valuable for the development of both design and theory, the early involvement of these expert users also enabled a good degree of understanding of the practical requirements and considerations of the potential deployment settings. This allowed for, at a minimum, a *vulgar competence* (Garfinkel, 1967; Crabtree, Rouncefield and Tolmie, 2012) in the understanding of the work of the setting on the part of myself, the ethnographer. This *competence*, or understanding, of the interactional setting is seen integral to the subsequent analytical process involved in making sense of what is being done (Crabtree, Rouncefield and Tolmie, 2012). Additionally, the involvement of the recipients of the research (those that the research is for) within the developmental process is reflected upon within the *research-through-design* approach described by Gaver & Bowers (2012) and performs the function of *sensitising* the initial design concepts as outlined by Benford et al. (2013) within their *'performance-led research in the wild'* approach (see Figure 6).

In summary, the research presented within this thesis is grounded within an ethnomethodologically-informed design ethnography in that the studied interactional settings are the museum and the home, and the participants and users are museum visitors, curators, composers and domestic occupants.

### 3.2.1 The artist-ethnographer

The importance of the ethnographer's understanding of the studied interactional setting is attributed with significant weight by Crabtree et al. (2012) who assert:

> *"It is the ethnographer's understanding of the setting from within which the exhibits are extracted which is crucial."* (Crabtree, Rouncefield and Tolmie, 2012, p. 112)

The above point positions the *artist-ethnographer* well for the ethnographic interpretation of the qualitative data as they undoubtedly have a detailed understanding of the artwork that they themselves have created. Indeed, one can see the additional significance this assertion has in relation to the construction of an interactive installation setting, where the artist has gained intimate knowledge of the experiences and character of the interactive setting through the detailed and involved process of making and testing. In short, one could argue that the *artist-ethnographer's* competence in understanding the experience they have created, when combined with competence within the deployed setting, goes beyond the *vulgar* and sets them in good stead for uncovering what Crabtree et al. (2012) describe as the *prize* of ethnography; *the machinery of interaction.*

Whilst within the *performance-led* research approach described by Benford et al. (2013) the differing perspectives of *practicing, studying* and *theorising* are evident for the *artist-researcher*, I would argue that the perspectives of the *artist-ethnographer* are less clear-cut*.* The ethnographical perspective described by Crabtree et al. (2012) as one of observation, documentation and explication of the setting's work fits well with the artist's perspective of observation, interpretation, reflection and presentation; just as the ethnographer documents and interprets the world around them, so does the artist. Although the purposes are distinct (perhaps one of provocation on the part of the artist and the development of understanding on the part of the ethnographer) the skills that are applied are less so. It should perhaps be noted that, rather than *'the artist as ethnographer'* as identified by Foster (1996) within what is termed the *'ethnographic turn*' within contemporary arts practice, the suggestion here concerns the suitability and application of the skills involved in artistic practice within the realm of ethnomethodologically informed research, whilst maintaining that the output of artistic practice can, indeed, be ethnographic. This suggestion is illustrated well by what Richardson & St. Pierre (2008) term as CAP (creative analytical processes) ethnography, where the creative practice, in their case writing, is identified as a form of inquiry that attempts to understand the world due to the interpretative nature of the creative work undertaken.

### 3.2.2 Sound technologies and the ethnographic perspective

Johnathan Sterne's assertion that sound technologies are *social artefacts* positions ethnography well for engaging them in study, reflecting its focus on *work* as a social enterprise, and its '*fundamentally social and accountable nature'* (Crabtree, Rouncefield and Tolmie, 2012):

> *'sound technologies are social artefacts… The secret to a hard-disk recorder may, Indeed, lie somewhere inside the hard drive and the microprocessor, but only if we consider those technologies as social artefacts that in turn lead us beyond themselves into other fields of practice.'* (Sterne, 2003, p. 338)

Additionally, within sound studies, Sterne (2003) suggests that the analysis of mobile listening technologies (with specific reference to *The Walkman*) provides a rich feeding-ground for study, where multiple expected and unexpected connections await, but only if we consider such technology as *social artefacts.* This viewpoint is exemplified by Michael Bull's seminal study into personal stereo listening in: *Sounding Out the City: Personal Stereos and the Management of Everyday Life* (Bull, 2000), who also notes the application and generalisation of the ethnographically analysed data beyond individual scenarios into the *'wider social and historical characteristics of society'* (Bull, 2000).

Lastly, Sterne (2003) also advocates the study of the seemingly obvious or the banal aspects of sound technologies, suggesting that it is the study of aspects that are less likely to draw critical attention that may well tell us the most. Again, this fits nicely with the ethnomethodological approach, specifically relating to the importance of capturing *vivid exhibits* of mundane everyday practices in an effort to uncover the *machinery of interaction* (Garfinkel, 1967; Crabtree, Rouncefield and Tolmie, 2012).

## 3.3 Adapting to circumstances

It should be noted that the absence of any first-hand observational data in study 2 (see section 6) was a consequence of the COVID-19 lockdown that required the design of a remotely deployable audio augmented reality experience within participant's homes. However, post-participatory interviews with users were conducted and recorded as a means of understanding what they *did* and what their experiences *were*.

In lieu of these direct, first-hand recorded observations and *vivid exhibits* of participants interactions that are the mainstay of ethnomethodological practice, it was intended that by getting the participants to *tell me* what they did, I could go some way towards getting them to *show me* what they did. As Crabtree et al. (2012) suggest, useful data pertaining to participants' interactions can be obtained by making observations that relate to the answers to questions given by participants regarding their experience.

Whilst it is understood that these post-participatory interviews are, from a design ethnographical point-of-view, decontextualized from the flow of the work with which we are interested in studying, it is also understood that they are not without their ethnographical merit. As Crabtree et al. (2012) propose, such data, when used in conjunction with other direct observational data, remains useful and of interest. As such, this data obtained via interview is also complimented by reflective and direct observational data obtained from the practice of creating, developing and deploying the studied audio augmented reality experiences. It was also anticipated that this study data would contribute to the findings based on the first-hand recorded observations of participants' interactions within the previous study of this audio augmented reality technology. These observations regarding participants' remote interactions with the audio augmented reality experience documented in chapter 6 have been made by *drilling-down* into the methods of the participants' practice with the remotely deployed experience; a result of the answers to direct questions concerned with teasing out the work undertaken by the participants within the experience setting.

By way of summary, within this chapter, a methodology has been described within which the experiential qualities and characteristics of audio augmented objects can be determined. This practice-based approach provides both the experiences to study, and the contexts within which to study these experiences. Whilst these two factors provide insights into the realisation and deployment of audio augmented objects, the ethnographic perspective provides insights into the user experience of audio augmented objects. The particular suitability of the described methodology to the subject of sound technology has also been identified, and how the described methodology has been adapted and applied to accommodate for the events of the COVID-19 lockdown have also been outlined.

# 4 Practice

Initially, in this chapter, I provide an introductory background to the type of creative practice I have engaged in this research, after which I detail the specific practice involved in designing, developing and deploying a series of AAR experiences comprised of audio augmented objects. Each of the experiences, which range from sound installations deployed within the public spaces of museums and galleries, to publicly accessible smartphone-based applications devised for use within the home environment, are reflected upon and conclusions are drawn. The user interactions with two deployed AAR experiences, *The McMichael Experiment* (see section 4.3) and *Horror-Fi Me* (see section 4.5), are subsequently studied in detail within chapters 5 and 6 respectively.

In keeping with the engaged methodology (see chapter 3), an iterative design approach is employed which attempts to ground the design of each subsequent installation based on the reflections of the practice and deployment of the previous installation. By way of summary, the research trajectory across the presented installations can be broadly defined as follows. Firstly, the realisation of multiple single audio augmented objects (in that each object has its own individual virtual sound source). This initial investigation is followed by the realisation of one physical object augmented with multiple virtual audio sources, which in turn is followed by an environment populated with two single audio augmented objects situated within audio augmented space. Finally, an installation environment is realised that contains multiple physical objects augmented with multiple virtual audio sources situated within audio augmented space.

## 4.1 Background

The practice presented within this thesis has deployed what could be considered *artistic interventions* within gallery and museum spaces comprising small scale interactive sound installations. This approach has drawn much from early artistic interventions with visual orientated augmented reality experiences, such as the AR(t) collective *Manifest.AR's* intervention *'We AR in MoMA'* (Rhodes, 2014), and also from an understanding of the intrinsic interventionist and disruptive qualities of the medium of AR (Baker, 2014).

In 2010 the artist group *Manifest.AR* curated an unofficial art exhibition where virtual artworks were placed within MoMA's gallery spaces which visitors could view through their smartphones, leading Thiel (2014) to remark: *'The institutional walls of the white cube are no longer solid...'.* In the case of Audio AR, Thiel's remarks take on a more literal meaning, both in relation to the internal walls of the gallery and its external boundaries, and present opportunity as well as challenge. If we can hear the visual before we see it, then the dividing walls of gallery rooms are no longer obstacles to our exploration, if we can hear the contents of the institution before we arrive, then these objects are no longer confined by external architecture. According to media theorist Sabine Breitsameter (Breitsameter in Behrendt, 2012) this fluid and borderless design approach stems from *'a sonic understanding of space'* which allows for a space which is more permeable and one that *'doesn't suggest the same kind of hard and fast boundaries of a visual construction of space'*.

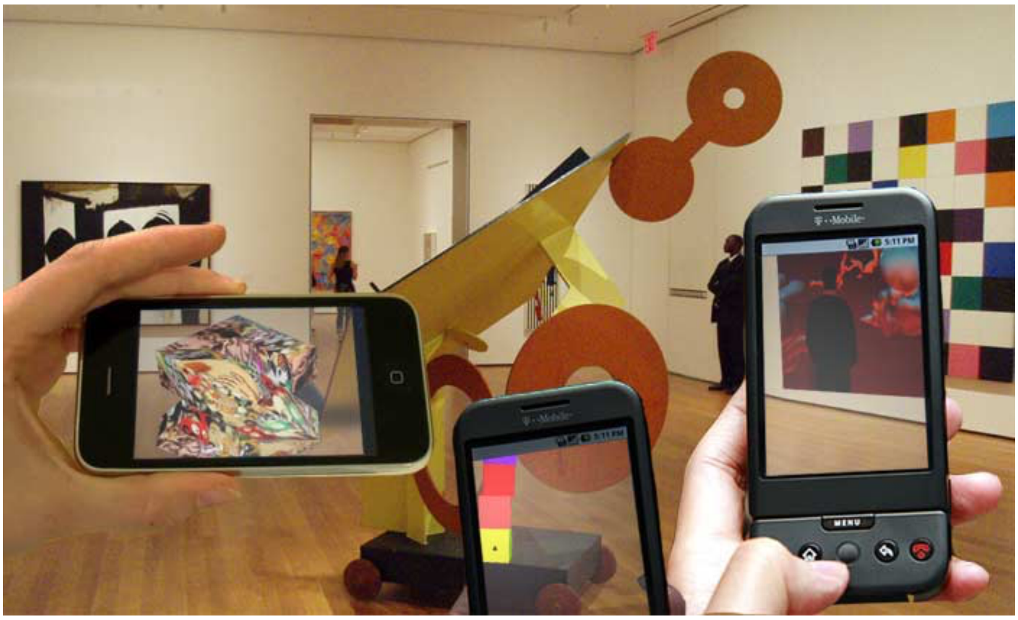

Figure 7. *We AR in MoMA* (2010) by *Manifest.AR* at the Museum of Modern Art, New York.

It is entirely possible to imagine the potential curatorial and institutional possibilities such an Audio AR intervention could have in terms of extending the boundaries of an exhibition. Additionally, in relation to the institutional appropriation of such an intervention, it is perhaps worth noting Zimmerman & Lorenz's positive curatorial feedback on their LISTEN system (Zimmermann and Lorenz, 2008) which acknowledges the curatorial potential of innovative, less descriptive and enriched audio content.

The audible augmentation of the art or museum object also leads us to think about how, in relation to contemporary curatorial and artistic practice, these objects could advertise their presence, and potentially the presence of other related objects around them, beyond the traditional confines of line-of-sight within such contexts (Kelly, 2017).

This extension of the object's, or the location's, communicable and cultural footprint could even extend beyond the confines of the architecture of the cultural venue or institution itself, with audio augmented objects or exhibitions advertising their presence and inciting interaction with them through related experiences beyond the walls of the gallery.

Another driving force has been the lack of a connection between collections of *silenced* sound making objects in museums and the audio archives that contain their related recordings. The recent digitization of analogue sound libraries has both highlighted this problem, and provided potential opportunities for solving it. This subject it discussed further in section 7.3.5.

It is my intention that the previous description of the type of exploratory creative practice that has been engaged with (see section 1.1) illustrates the interdisciplinary nature of the research undertaken, especially in relation to the notion that the creation of a technical system, the output of which is intended to be creative, constitutes an artistic practice in itself.

## *4.2 The Mingus Demonstration*

*The Mingus Demonstration* was devised as a SLAM-based, image recognition audio augmented reality prototype experience. Intended to be authored and deployed as a smartphone application, with audio content delivered to the user via headphones, this prototype was demonstrated to curatorial staff at the *National Science and Media Museum* (NSMM) in Bradford, UK. The intention was to author a prototype AAR application capable of augmenting multiple physical objects with 6DoF dynamic binaural audio and with a view to creating a virtual soundscape comprised of these objects for a listener to explore.

### 4.2.1 Design and development

The *Vuforia Engine*[1] was adopted as a means to realise a SLAM-based image recognition and tracking feature within a system that was useable from both an authoring and curatorial perspective in a variety of locations. This decision was informed and inspired by the artwork detection project presented by Seidenari et al. (2017). Along with artwork recognition, the use of image recognition and tracking technology presented opportunities for the development of an Indoor Positioning System (IPS). The *Vuforia* SDK enables the development of mobile augmented reality applications that use computer vision technology to recognise and track image targets and three-dimensional objects in real-time, and is compatible with both the iOS and Android mobile application platforms. The *Vuforia Engine's* camera-based object recognition and tracking capabilities not only facilitate the recognition of the artwork and artefacts to which virtual audio sources can be associated, but also additionally enables the implementation of an IPS where the listener's angle and distance can be determined in relation to tracked, stationary two or three-dimensional objects.

---

[1] *Vuforia Engine* is a software development kit (SDK) for creating AR applications with computer vision functionality that can recognise images and objects in the real-world. The *Vuforia Engine Package* can be used in combination with the *Unity Game Engine* for authoring cross-platform mobile applications.

Through an authoring approach similar to the one presented in the *LISTEN* system by Zimmerman & Lorenz (2008), where a world model is combined with a locative model, it is possible to determine the listener's position both in the physical and virtual environment. Within the *LISTEN* system, the world model contains geometric information relating to the physical real-world environment and the objects within it, which it describes as the visitor moves and interacts with the system. On the other hand, the location model defines areas of interaction within the world model and enables the system to determine the visitor's location and head orientation by mapping their position to predetermined virtual zones within the space and their position in relation to object identifiers.

Within *The Mingus Demonstration*, the locative model is authored within *Unity*[1] as zones of space of a specified size and shape, situated at specific coordinates in three-dimensional space in relation to a unique and recognisable image target. The *Vuforia SDK* acts as our world model, which it creates on-the-fly, recognising and tracking the location of the image target in the physical environment. Because our listener is holding the camera, the system knows the listener's position and bodily orientation in relation to the tracked image and therefore can determine the listener's position and orientation in relation to our authored zones of space.

Additionally, the system is capable of determining the listener's current focus by returning the angle and distance of the listener in relation to the tracked object. An additional and important feature of this camera-based IPS is made possible through Vuforia's Extended Tracking or SLAM capability, delivered through either *Apple's ARKit*[2] or *Google's ARCore*[3], when compiled for delivery as either an iOS or Android mobile application respectively. *Vuforia's* extended tracking enables the continued recognition and estimated location of a tracked object outside of the camera's field of view. This fusion-based sensing technology extends our ability to determine the location of our physical objects and their associated virtual audio sources in relation to the listener's position in space. By being able to estimate both the angle and distance of the virtual audio sources around the listener, we can deliver a virtual and interactive three-dimensional soundscape based on the listener's physical, real-world environment.

### 4.2.2 Authoring and deployment

*The Mingus Demonstration* was authored and deployed as an iOS app for use with an iPhone. The experience comprised of three printed photographic

[1] *Unity* is a cross-platform game authoring application. See: https://unity.com
[2] *ARKit* is *Apple's* SLAM-based AR software development kit (SDK) that enables third-party developers to build AR applications that utilise an iOS device's camera and motion sensors.
[3] *ARCore* is *Google's* AR SDK that enables third-party developers to build AR applications that utilise an Android device's camera and motion sensors.

display images each associated with a single virtual audio source. The images and audio content were chosen in order to trial disparate types of audio content within an experience such as this, though at the same time could be appreciated as belonging to a singularly themed installation or exhibition. Centered around the 1960's New York jazz musician *Charles Mingus*, the photography comprised of an image of one of his albums covers, a black and white portrait photograph of the musician and a black and white photograph of a 1960's New York Street scene. The audio content with which these three photographs were augmented with were a piece of solo jazz piano music, an excerpt from a recorded interview with the musician and an archive field recording of a 1960's New York Street.


Figure 8. *The Mingus Demonstration* image targets. These images were augmented with the following audio content (from left to right): a piece of solo jazz piano music from the depicted album, an excerpt from a recorded interview with the musician and an archive field recording of a 1960's New York Street.

The audio middleware application *FMOD*[1] was used to create audio events that could be triggered within *Unity* when the AR application successfully recognised the tracked images. From within FMOD the audio events were given a specified range of 3 meters, this determined the distance from the centre of the sound source from which the sound would be audible. Additionally, rather than being omni-directional sources, each of the FMOD audio events were given an angle within which the audio signal would be fully audible, and outside which the audio signal would begin to attenuate, in essence they were given directional characteristics. This focal angle was set at 60°, meaning that if the AR camera was in range and within this viewing angle the sound source's signal would not be attenuated, beyond this angle the signal would be attenuated until the angle reached 180°, at which point the signal would be fully attenuated. Both the range and the focal angle were programmed using *FMOD's distance* and *direction event parameters*. These ranges and angles are represented graphically in Figure 9.

[1] *FMOD* is described as an audio middleware program for authoring sound effects and interactive music sequences for video games. *FMOD* provides an application programming environment (API) that allows the parameters of an authored *FMOD audio event* to be manipulated via components and scripts within the *Unity Game Engine.*

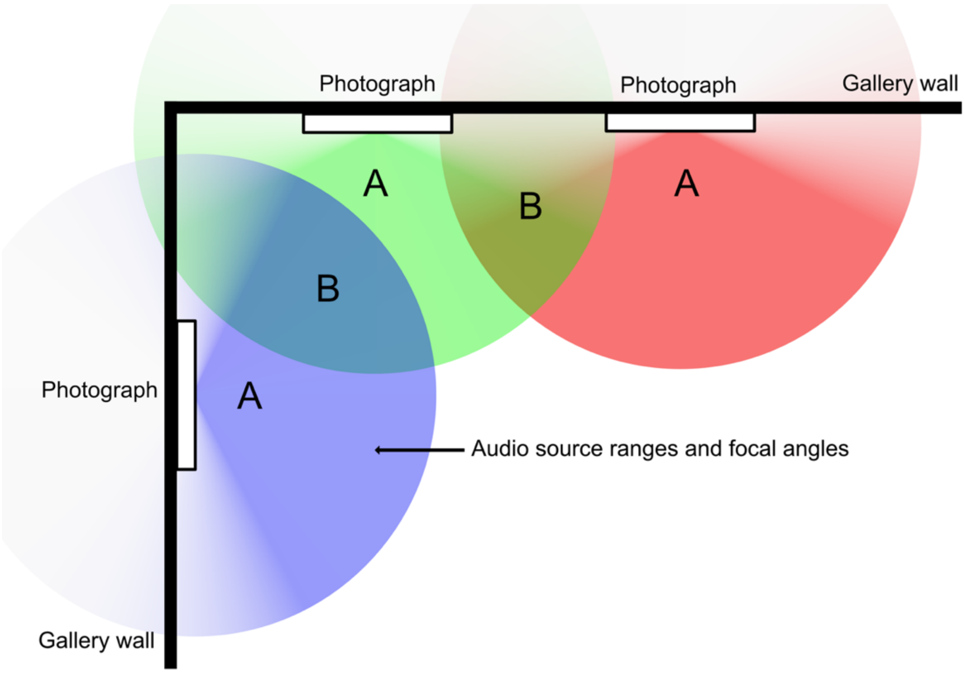


Figure 9: Interactional design and layout for *The Mingus Demonstration.* Showing the sound source positions, ranges and angles in relation to the audio augmented photographs. Areas (A) indicate locations where only individual audio sources were audible, areas (B) indicate locations where more than one audio source was audible.

The demonstration was setup in the corner of a gallery room, and the photographs were positioned so as to create locations (taking into account the source's ranges and directivity) within which both individual and multiple sound sources could be heard. These locations are indicated (A) and (B) respectively within Figure 9. In practice the physical positioning of the photographs in combination with the range and angles of the sources created locations within which listeners could focus their attention on the individual audio content attached to the individual photographic images, along with locations where solo piano music could be heard alongside the recorded interview or the sound of a busy New York Street. By moving around, it was possible for listeners to find locations where one audio source could provide an accompanying background to another, or they could mix, or were afforded the opportunity to compose to some degree, the virtually created soundscape for themselves, albeit with the limited resources of three differing audio sources.

The binaural rendering of the spatialised sound sources was performed by the *Resonance Audio*[1] plugin. *Resonance Audio* is an open source, cross-platform audio spatialisation plugin that uses a generic HRTF profile to simulate how real sound waves interact with human ears. Binaural audio processed with a HRTF profile allows listeners to perceive sounds as having a specific location in physical space when heard through headphones. Although primarily designed

[1] https://resonance-audio.github.io/resonance-audio/

as a game engine plugin for the spatial positioning of virtual audio sources within virtual worlds, within the AAR system described here the virtual sounds are given a specific presence within real physical space due to being attached to, or located at the same position as, real physical objects. Figure 10 shows the overall system architecture of *The Mingus Demonstration*, along with the responsibilities of the individual hardware and software components.

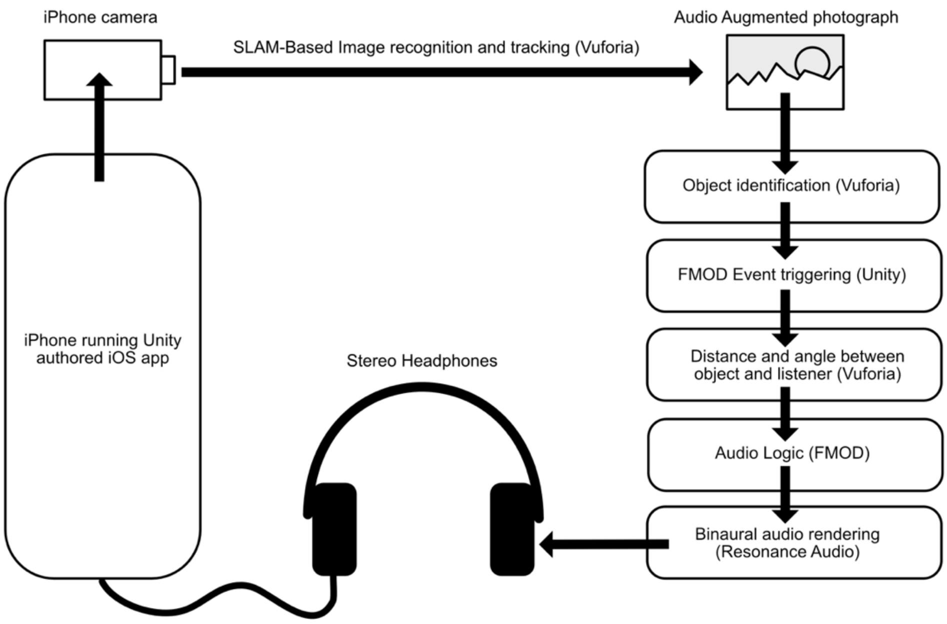


Figure 10. System architecture for *The Mingus Demonstration*.

Naphtali and Rodkin define the core set of components required to construct an AAR system; sensors, control methods, rules and conditions, and a delivery mechanism (Naphtali and Rodkin, 2020). Within the AAR system used to realise *The Mingus Demonstration*, we can define our primary sensor component as being the phone's camera, which provided real time tracking of our listener's position and recognised environmental elements. Our control methods were virtual colliders, authored zones of space in the virtual environment, the position of which in the real-world physical environment were determined by our sensor component. These colliders acted as triggers for our rules and conditions, which was the audio logic that determined the audio content delivery. The delivery mechanism, the device with which our listener will interface with system, comprised of a smartphone and headphones, capable of realising our core set of system components either via an installed application or intrinsically via its hardware and software, capable of delivering personalised, high-fidelity, dynamic binaural audio.

### 4.2.3 Reflections on practice

This initial demonstration at NSMM was received favourably by curatorial and collections staff in terms of its potential application within a museum context, and in terms of its immersive qualities. The capability of the system to robustly and effectively augment the photographic objects with dynamic binaural audio content was also evident during the development process and during various lab-based tests, as well as in the demonstration. Of specific interest, from a curatorial perspective, was how the system enabled the addition of an extra layer of interpretive content in an economical and cost-effective manner, given that no internal infra-structure was required, although concerns were aired regarding the suitability of headphone-based audio experiences within public spaces (a subject considered in more detail within section 7.3.3). Furthermore, one member of the curatorial staff expressed an interest in finding out more about the demonstration's subject matter (something that they were not familiar with), suggesting the potential of the experience to provoke interest and engage users with an exhibition's subject matter.

Also of interest was how multiple sound sources could be interacted with through movement, and how these could be combined to compose an over-arching context or atmosphere for an exhibition around a specific subject. It was therefore determined that the described approach to AAR warranted further exploration, and demonstrated a potential for deployment within galleries and museums. A more detailed discussion on *The Mingus Demonstration* from a curatorial perspective is included in section 5.4.

It was determined that the *range* and *angle* parameters of the sound sources could be included as fully adjustable parameters within an authoring tool which could facilitate the creation of AAR soundscapes as required, allowing for as much, or as little, overlap of sources as desired. It was also noted that these factors work in direct combination with the positions of the augmented physical objects and the architecture of the indoor space. These reflections on possible authoring techniques for audio augmented reality soundscapes comprised of audio augmented objects are discussed in greater detail in section 7.5.

It's also proposed that this approach demonstrates a contribution to indoor positioning within gallery and museum environments through the application of camera-based image detection for determining visitor location and focus. Which, in turn, enables the system to expand upon the concept of the *attractor sound* (Zimmermann and Lorenz, 2008) through a system reliant on little background infrastructure. Also concluded was that this appropriation and application of Simultaneous Localisation and Mapping (SLAM) for solely audio augmentation purposes works well with this particular technology's current shortcomings. The small and gradual movement of overlaid graphics placed on real-world objects that can sometimes be observed with visual based AR applications (sometimes referred to as *drift*) was not so acute or obvious when translated to the spatial position of audio sources, a shortcoming which is the result of the mapping technologies adjusting their placement of virtual augmentations as they build up, or gain additional information about their environment (Apple, 2021c; Google, 2021).

Initial prototype designs centred around tracking the objects to which the virtual sound sources were going to be attached to, and using these as reference points to determine the listener's position and orientation, an approach that seemed natural given that these were the objects that I wanted to detect. But through the prototype development stages, once a system had been developed that demonstrated a useable degree of accuracy and reliability, and through the trials and manipulations involved in sculpting the positions and dimensions of the virtual audio sources in physical space, a feature-based tracking approach emerged. This approach is discussed in section 4.3.1.

Finally, also observed within this initial demonstration, specifically in relation to sound content of differing categories, was the interest and engagement expressed by museum staff in both singularly delivered audio sources, and simultaneous multi-source content delivery. The former with which the individual source could be afforded greater focus and attention, and the latter with which participants could begin to claim agency over the composition of their own soundscape to accompany their physical reality.

## *4.3 The McMichael Experiment*

This section outlines the development and deployment of an AAR installation at the National Science and Media Museum (NSMM) in Bradford, UK. The intention was to build upon *The Mingus Demonstration* (detailed in section 4.2) by augmenting a physical museum artefact with virtual audio content. It was also intended that the possible feature-based tracking approach that emerged as a result of the previous deployment should be explored further within the context of this project. The user interactions with this sound installation are the subject of study in chapter 5.

Whilst the development of this installation drew inspiration from the successes and findings of the previous project, the design of the installation was largely motivated by an investigation into the potential for a *silenced* museum object to act as an interface for digital audio archival content. As discussed in section 2.6, this potential was identified, with somewhat limited success, in the project presented by sound scholar and cultural historian Karen Bijsterveld (2015) and discussed in detail by Mortensen & Vestergaard (2014) within what they termed a *listening exhibition* curated at the Media Museum in Odense, Denmark in 2012 titled 'You are what you hear'. Through the implementation of their *Exaudimus* system, Mortensen & Vestergaard proposed a way of exhibiting and interfacing with radio heritage which has been enabled by the digitisation of analogue audio archive content by the Danish Broadcasting Corporation. Within this approach, we see how, through authorship and embodied visitor interaction, the exhibition demonstrates potential as an accessible and immersive interface to the sound archive itself.

We can imagine the audible output of this project and the *Exaudimus* system to be of a similar nature, given the similar context and type of physical and virtual audio artefacts used. But the application of different technological solutions within each AAR system and the apparent absence of three-dimensional audio spatialisation within the *Exaudimus* system, along with differences in the material contextualisation of the audio content (*listening situation* verses direct augmentation of the sound artefact), denote the issues around both authorship and user experience being very much different.

Although Mortensen & Vestergaard (2014) reported some success in generating engagement with the audio archive content contained within the exhibition, significant issues arose around initiating interaction with, and triggering the playback of the audio archive content. The authors attribute this to what they termed *cultural constraints*, the reluctance of visitors to touch, pick up and directly interact with physical objects within a gallery environment, something which goes against normal behaviour within such a context. Unfortunately, the triggering of archive audio playback was largely dependent on such direct interactions with the constructed listening situations within the exhibition.

This installation also presented the opportunity to explore the differences between augmenting *silent* objects and augmenting *silenced* objects with audio content; the difference between a photograph that is expected to be silent, and a radio that is expected to generate sound. This was considered in addition to creating an installation experience that would help to provide deeper insights into the relationship between the museum artefact, sound and engagement.

### 4.3.1 Design and development

Further exploration of the *feature-based* approach outlined in the previous section (4.2.3) involved providing the object tracking software (*Vuforia*) with isolated images of unique and static physical features within the experience environment. The approach meant that the listener's position and orientation in relation to these physical features could be determined and, in turn, so could the position of the user in relation to the object to be augmented with sound. This concept was devised as a result of two key observations made as a result of the practice thus far. Firstly, within *Unity* it is possible to position virtual content relatively to the tracked image, meaning that the tracked image does not necessarily need to be the object that is augmented. Secondly, how can an audio augmented object attract listeners' attention if it cannot be initially recognised, for example, it is obscured or beyond the view of the camera and therefore cannot be recognised, localised and subsequently tracked.

Primarily designed as an image, fiducial or QR code recognition and tracking function, *Vufuria's* image recognition database was supplied with photographs of unique features present within the physical environment. Experiments where then conducted tracking unique and stationary building features with a view to determining how they could be used to either provide trajectory way-markers for an experience, or to provide additional information to the system as to the position of the user and provide reference locations from which virtual audio content could be virtually placed.

An experimental app was authored using the same tools as *The Mingus Demonstration* outlined in section 4.2.2, though, as this remained solely an experiment in image and location tracking, no audio content was used.


Figure 11. Using image recognition to locate and track the unique features of a physical environment.

Figure 11 shows the photographic images of unique features present within the physical environment provided for *Vuforia* to track (top row), and the successful recognition and *SLAM-based* tracking of these features within the physical environment, indicated by the red dot (bottom row).

The ability of the image recognition and tracking software to successfully identify and track unique features within an indoor space presented exciting opportunities for the authoring of experiences comprising of audio augmented objects in museum spaces. For this project it opened up the possibility of tracking unique features in the installation location rather than the augmented object itself, thus opening up the possibility of objects communicating their presence to visitors beyond line-of-sight. Within the construction of an AAR experience this could be realised through the positioning of, (potentially multiple) virtual audio sources relative to a trackable environmental feature.

Figure 12 shows a preliminary outline for an installation based around the audio augmentation of a 1950's radio and television set (see Figure 13), which was selected after a tour of NSMM's collection of electronic instruments, radio equipment and photography archives and consultation with the museum's collections and curatorial staff.

In Figure 12 you can see how the feature-based tracking technique was intended to be implemented within the experience by measuring the location of the radio in relation to a uniquely identifiable feature within the museum's physical environment. It was intended that a virtual world model would then be constructed that would place the virtual sound source directly in line with the physical radio's speaker, relative to the position of the tracked feature.

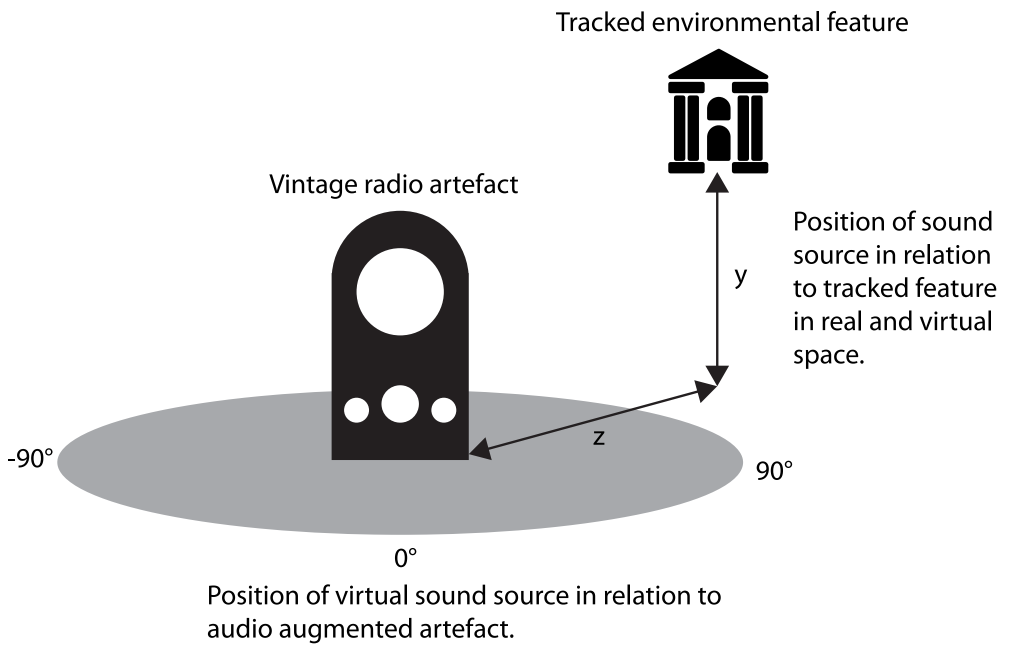


Figure 12. Outline for realising an audio augmented museum object.

Unfortunately, due to the exact location of the installation within the museum space, and for that matter the radio object, being undetermined during the final development of the project's mobile application, this feature-based tracking approach had to be abandoned in place of a more traditional marker-based tracking approach. This problem, which highlights the intimate relationship between architectural space and indoor AAR environments of this type and design, is further reflected upon in section 4.3.3.


Figure 13. McMichael model 512R 12 inch television and radio receiver. Made in Slough, England, c. 1951.

### 4.3.2 Authoring and deployment

As before, an image target, though this time in the form of a QR code, was uploaded to Vuforia where the image feature points are extracted and stored in a database. This image target was included as a game object within the Unity scene, with another game object added as a child of this image target object, to represent the virtual audio source. This child object was positioned virtually in relation to its parent image target to reflect the actual required position of the virtual sound source in our real-world environment (see Fig. 2).

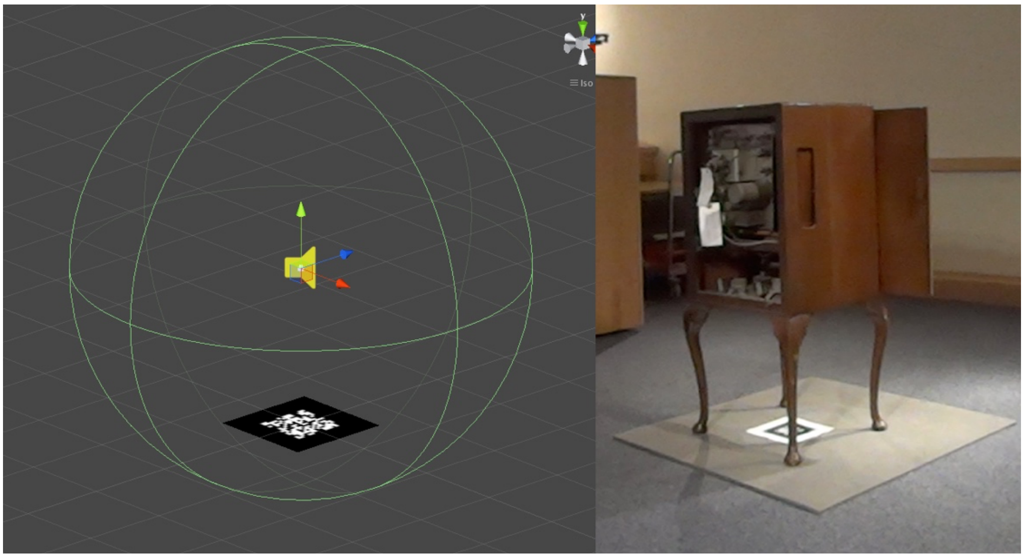

Figure 14. Positioning a virtual audio source in relation to a real-world object.

On the left in Figure 14, we see the virtual environment during development, showing the position of the virtual audio source and its collider component in relation to the position of the tracked image. On the right, we see the position of the tracked image in relation to our radio object in our real-world installation environment. The speakers of the radio were situated in the bottom of the main body of the radio unit.

The authored FMOD audio event was attached to this child game object, along with a collider object for triggering it. Key to this authoring approach working in relation to the designed model of spatial interaction (Figure 15) is the use of collider components on both the virtual audio event triggers and on Vuforia's *ARCamera* object. The addition of a rigid body component on the latter, combined with these collider components, renders our user's mobile camera position within both the virtual and physical world of our AAR application much the same as a first-person perspective player within a video game, and, as such, other similar game-orientated authoring approaches can be adopted within FMOD. The approach of commandeering game authoring techniques, specifically collision detection, for spatial augmented experiences, is utilised and reflected upon within the field of HCI by Greenhalgh & Benford (1997) in their model of spatial interaction for a remote teleconferencing application.

The appropriation of the VR authoring technique of collision detection through the placing of collider components around the virtual sound sources and the *ARCamera* object begin to realise a model of spatial interaction with similarities to Greenhalgh & Benford's (1997) Auras; spatial zones around objects that define their region of interaction with other objects. Similarly, this approach enables an awareness of these objects to each other, indicated by their position and orientation. This awareness could be used to design a model and author a subsequent experience that can take advantage of this information to determine a user's current focus within the system, and to allow an object to determine if it is the current point of focus.

The design of focal length and width for individual virtual sound sources within the model can be achieved through the dimensions of both its range and its associated collider, the shape of its directivity pattern and through the attenuation of its signal based on the parameters of distance and angle between it and a listener. Again, this echoes Greenhalgh & Benford (1997) and the concept of the *nimbus* feature of an object as both a focal and advertising determiner.

By adjusting the audible presence of virtual sound sources based on listener's proximity and orientation, we can design an element of focus into the experience where individual sound sources and objects can be identified and coherent and curatorially useful soundscapes can be composed. Additionally, this audible presence could be manipulated, or focal range extended, in order to give specific sources priority, or to enable them to advertise their presence more vocally than other sources within the experience. It is these points that are of particular interest as they constitute a manipulation of the usual, or expected, attributes of a physical sound source. According to the normal physics of sound, these sound sources would continue to emanate through the soundscape, with only the altered characteristics virtually attributed to them by the game engine's audio spatialisation effects, perhaps through means of occlusion, change of environment, volume and position within the game or experience. We may see here the emergence of a model for spatial audio interaction for use within applied AAR systems, where a considered compromise is brokered between the creation of a realistic acoustic virtual reality and a functional and coherent model of spatial interaction.

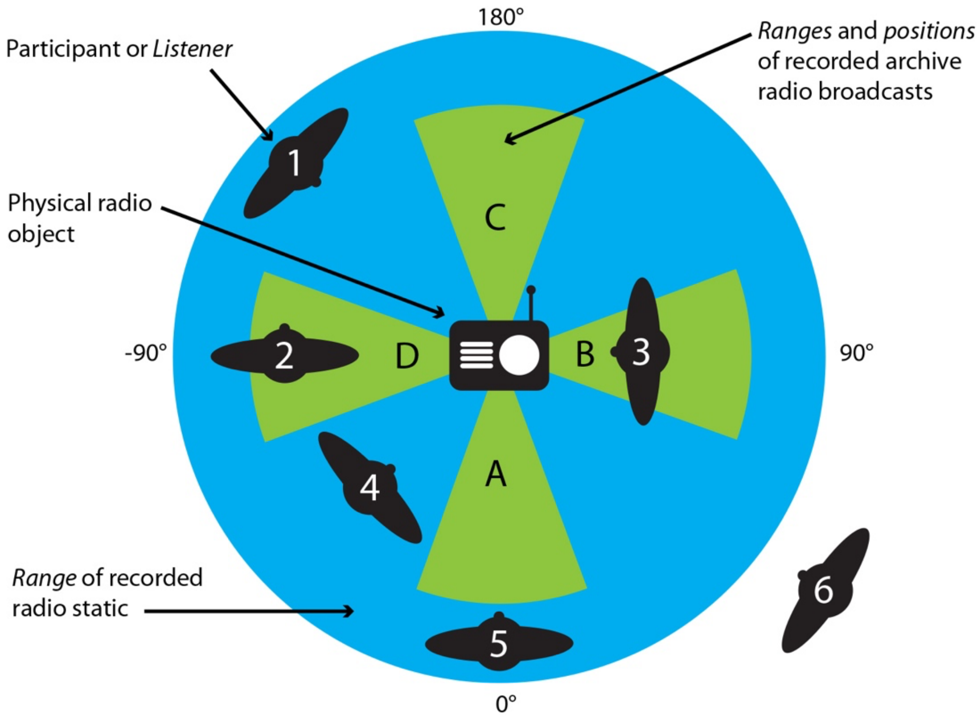


Figure 15. Spatial audio interaction design for *The McMichael Experiment*.

In Figure 15, we see the spatial audio interaction design for *The McMichael Experiment*. In the centre is the physical vintage radio artefact, which is represented in the virtual space by the QR code on the floor below it (as shown in Figure 14). There are four looped virtual archive radio broadcasts positioned around the QR code image target; these are positioned at 0°, 90°, 180°, and − 90° and are indicated by areas A, B, C and D on the diagram respectively.

As previously mentioned, for the purposes of this installation, a 1950s television and radio receiver were selected from the museum's collection. Also utilised was contemporaneous archival radio broadcast material obtained from an online Internet archive resource. This material included a science-fiction radio drama, a live concert hall musical performance recording, the narrated introduction to a religious music programme and an episode from a detective drama serial. All the chosen audio content was historically and geographically accurate in relation to the chosen radio receiver from the museum's collection. In addition to the recorded archival radio broadcast audio content, various recordings of radio static were obtained by recording the output from an out-of-tune contemporary radio receiver. Table 1 shows the included audio content and details of their attributes and functions.

The real-world positions of these virtual archive radio broadcast transmissions are achieved within the FMOD event authoring environment by cross-fading from the background radio static sound to the appropriate archive recording when the listener is in the relevant position in relation to the tracked QR code. The cross-fading between these two audio sources is extended by 10° in each direction from its centre position, with a further 10° transitionary non-linear cross-fade to allow for a degree of comfortable, and smooth transitional listening, so small body movements do not result in sudden loses of the perceived broadcast signal. The audio sources were positioned around the radio in this fashion to promote 360° exploration of the physical artefact, and so that multiple audio sources could be tuned in to through the embodied interactions of the listener around one augmented source. The fine tuning of the crossfade angles were a result of trial and error in the authoring process in an attempt to create smooth and seamless auditory experience, and to try and emulate the tuning of an analogue radio dial with bodily movement.

| **File name** | **Description** | **Type** | **Function** | **Position** | **Range** | **Loop length** |
|---|---|---|---|---|---|---|
| Paul-temple.aif | Crime drama | Spoken word | Archive content | 0° | 2 m | 02.05 |
| Chapel-in-the-valley.aif | Religious music programme | Spoken word and music | Archive content | 90° | 2 m | 00.42 |
| Variety-bandbox.aif | Live recorded musical concert | music | Archive content | 180° | 2 m | 01.56 |
| Red-planet.aif | Science-fiction drama | Spoken word | Archive content | − 90° | 2 m | 02.17 |
| Static.aif | Untuned radio static | Sound effect | Transitional ambience | n/a | 3 m | 00.06 |

Table 1. Audio content included in *The McMichael Experiment*.

It is the listener's focus, along with their position in relation to the virtual sound source, which is situated in the same physical location as audio augmented object, that additionally determines the delivery of the audio content to the user. Within the context of this installation, and the associated interaction model, the listener's focus is determined by the angle of their hand-held iPhone in relation to the tracked image target. It is this, in addition to their bodily position in relation to the tracked image target, that provides a spatial interactional model that encompasses degrees of listener position, proximity and focus.

These three spatial interactional variables (position, proximity and focus) and their associated outcomes in terms of audio content delivery for the respective listener can be illustrated through a closer inspection, and a comparison of the positions of listener 2 and listener 3 in the interaction design diagram (Figure 15).

In Figure 15, we see listener 2 at a position of−90° in relation to the radio object and the tracked target, and therefore currently at a position where they can hear broadcast D, in contrast to listener 3, who is at a position of 90°in relation to the radio and therefore can currently hear to broadcast B.

It is the listener's proximity to the radio object that also determines if they are currently within hearing range of the broadcasts located at their current positions, and the degree to which the broadcast's signal is attenuated and mixed with background static. We can see that both listener 2 and listener3 are within range of broadcasts D and B respectively, and therefore are able to hear these broadcasts, though listener 3's closer proximity to the object means that its signal will be less attenuated than listener 2, who is further away.

Our last interactional variable, that of focus, is illustrated by both listener 2 and listener 3 within Figure 15. The focus variable is determined by the angle of the listener's device in space (in this case their handheld iPhone) in relation to the position of the tracked object. We can see that listener 2 is facing away from the radio, with it situated on their immediate right-hand side, and as a result will perceive the spatialised audio content as being emitted from their right-hand side (the direction of the radio). In contrast, we see listener 3 directly facing the radio, who, as a result, will perceive the virtual audio sources as emanating from directly in front of them.

In light of this explanation of the spatial interactional variables of position, proximity and focus, we can determine the differences in the delivery of audio content for all our listeners' locations in Figure 15. Perhaps notable here is listener 6, who, although directly facing the radio, will hear nothing as they are well out-side the range of both static and broadcast. Similarly, we see that the location of listener 5 determines that, although they are within range of the static, with the radio directly in from of them, they are beyond the range of the broadcast.

Furthermore, the delivery of audio associated to the radio object to listener 2 in in Figure 15 has the potential to encourage engagement with the object by tempting their focus, but additionally leaves them open to impressions of other potential virtual sound sources within the context of an experience with multiple audio augmented objects.

Both the real-world distance and angle between these virtual sound sources and the user can be accessed as parameters within FMOD in order to author adaptive transitions in the delivery of the audio content. This is achieved in the same way a player's character may experience virtual sound sources when exploring the virtual domain of a video game, or the way in which instrumentation within an adaptive soundtrack may be manipulated in relation to the player's health, or the proximity of enemy characters.

### 4.3.3 Reflections on practice

In practice the feature-tracking approach was not possible due to the unknown nature of the deployment environment, although the initial experiments that formed part of this project's development show promise for this approach solving the problem of augmenting unseen objects with audio content and thus making them audible beyond a visitor's line-of-sight. The QR code solution did, though, demonstrate the ability to successfully position sound sources relatively to a recognisable and trackable image, therefore I would maintain that within a fixed environment, to which sufficient access is available to carefully author an AAR experience like this, a feature-based tracking approach still remains viable. This viability appears to hinge upon the ability to carefully survey and author the virtual world model to reflect the precise positioning of tracked features and objects that require augmenting. This also presupposes that the objects to which virtual audio content will be attached are in a fixed location within the installation environment.

As (Behrendt, 2010, 2012) points out, the interactional modality of *placed sound* is walking, and within an indoor environment the direction and space within which one can walk is determined by the shape and confines of architectural space and the obstacles and objects situated within it. Therefore, when considering the interactional design of indoor *placed sound* experiences, the architectural setting of the experience needs to be carefully considered and built into the interactional model.

By associating virtual and spatialised audio sources to objects, or features, that may not be directly related to the experience, one can begin to think about how artworks or artefacts within gallery and museums could advertise their presence beyond the traditional confines of line-of-sight. For example, a unique architectural feature within a gallery could be tracked in order to position and trigger an audio source attached to an object beyond a partition wall.

Additionally, one can see how this could also be used to curate and design visitor trajectories through an exhibition or a collection of artefacts by triggering sound sources at certain times in certain locations, or in relation to other objects. For example; If you're listening to '*x*' then you can also hear '*y*', guiding you through the exhibition regardless of whether the camera has seen the actual audio augmented object.

Also possible would be advertising the location of other objects in relation to the one you are currently viewing; *associated objects* that work well together sonically as well was contextually, guiding and suggesting trajectories to the listener. Of particular interest would be how the reliability of such targets is affected by changes in lighting conditions and human traffic.

Furthermore, and in relation to my previous observations on the appropriation of video game sound design techniques within an AR experience, such an approach as the one outlined above could have the potential to create rich narrative experiences through the use of both *augmented diegetic* and *augmented non-diegetic* sounds. Converting a seemingly warm and welcoming environment into a potential scene from a horror movie is perhaps a convenient illustrative example. This approach could include banks of audio content that could relate to user preferences for type or genre of content. It also provides an opportunity to explore in more detail how additional sound effects, in addition to the primary audio content, can be used to promote, incite and effect engagement. This would perhaps constitute a study on how more filmic and game-like sound design approaches can be used in this context, outcomes from which could be used to inform trajectory and narrative based approaches to AAR experiences.

In relation to this project and *The Mingus Demonstration* (see section 4.2) we see a possible and interesting distinction emerging between the audio augmentation of *silent* objects and the audio augmentation of *silenced* objects. The audio augmentation of *silenced* objects seems to go beyond, or project further into reality, the virtual audio content due to the augmented object being a possible and believable source of real sound. This appears to be a different experience from the usual super imposition of audio content on the expectedly *silent* object, or indeed the usual visual AR experience where we see, for example, computer graphics super imposed over the cameras view of the real-world.

Finally, the development of a non-tactile interface to digital audio archival content presents a plausible solution to the problem encountered by Mortensen & Vestergaard's (2014) *Exaudimus* system (where interaction and access to the archival audio was hindered by the requirement for to physically pick items up). Whilst this presents a possible way of using silenced museum artefacts as interfaces to digital audio archival content, it also presents this digital audio archival content within its original context (in this case a period appropriate domestic radio set) as well as presenting the material radio object within the context of its original sound and function.

## *4.4 Alien Encounters*

The third audio augmented reality experience *Alien Encounters* was presented to a visiting public at a National Science and Media Museum (NSMM) *Lates* event, a free, quarterly, adults-only open night at the museum where hands-on experiments, music, comedy, art installations and talks are organised and arranged around a specific theme. An attending public of diverse gender, age and cultural background are invited to explore and participate in these additional activities at the museum, which are presented alongside the existing collections and current exhibitions. *Alien Encounters* was also demonstrated at the *Audio Mostly 2019* and *Halfway to the Future* conferences.

At the National Science and Media Museum *Alien Encounters* was exhibited alongside other technical and experimental demonstrations and installations within the NSMM's *Wonderlab*, a gallery of interactive, science themed exhibits. Visitors were free to approach the installation and request to participate. This particular *Lates* event coincided with the museum's launch of a new temporary *space-themed* exhibition, as such artists were encouraged to explore the sights and sounds of space within their presented work.

The busy, freely open and public nature of the deployment conditions of *Alien Encounters* within NSMM's *Lates* event negated the possibility of the type of ethnographical study which the two subsequent installations were subjected to (see chapters 5 and 6). This was due, in the main part, to the inability to obtain ethical consent from the public participants prior to their participation in the installation, eliminating the possibility of obtaining any video, audio or photographic *vivid exhibits*. Having said this, observational and reflective notes were taken during and immediately after the event by myself, the sole researcher present during this specific deployment.

This particular installation comprised of two *silenced* audio augmented objects in the form of two transistor radios each augmented with their own individual virtual audio broadcast (see Figure 16). Whilst within *The McMichael Experiment* (see section 4.3) a single radio object was augmented with multiple audio broadcasts, within *Alien Encounters* each of the two radios were augmented with their own single audio broadcast. In essence this different approach was taken in order to obtain different observations from both practice and participation and answer different questions which could then be reflected upon. These included: What are the implications for authoring multiple audio augmented objects in close proximity? What is the user experience of a soundscape comprised of multiple audio augmented objects and what are the characteristics of such a soundscape?

The first radio was a 1950's portable Bush radio receiver that was paired with an episode of the 1950's science-fiction radio drama *The Red Planet*, the second was a 1970's portable Panasonic radio receiver that was paired with a recording of David Bowie's *Starman.* The content of the radio broadcasts was chosen to reflect the 'space' theme of the *Lates* event as well as the historical period to which the radio receivers belonged.


Figure 16. The two audio augmented radios used for *Alien Encounters.* Left: The Bush TR 82C (Bush Radio Ltd. 1959-1963). Right: The Panasonic R-1449 (Matsushita Electric Industrial Co. Ltd. 1970).

### 4.4.1 Design and development

*Alien Encounters* worked in a similar way to the previous *McMichael Experiment* installation (see section 4.3) in that the virtual audio broadcast content was made available for listeners to *tune into* amongst a background sound of radio static via the position and orientation of their body in physical space. However, there were several significant differences between this and the previous installation.

Firstly, this installation would be comprised of more than one audio augmented *silenced* object, making it possible to explore the authoring strategies involved in the construction of an audio augmented reality comprised of multiple audio augmented objects, as well as observing listener interactions with them.

Secondly, a hardware interface was devised that would allow the iPhone to be mounted on the side of the listener's headphones to enable the position and orientation of the listener's head to determine the output of the spatial audio, see Figure 17. This essentially enabled the exploration of a *head-tracking* approach, rather than the handheld *device-tracking* approach formally implemented. Whilst the *device-based* tracking approach for the delivery of dynamic binaural audio presents as an effective and accessible solution (Heller and Borchers, 2014; Cliffe *et al.*, 2019, 2020), the use of a generic HRTF profile by the binaural rendering software suggest that a more compellingly realistic result would be achieved through the implementation of a *head-based* approach to tracking.

Figure 17. Hardware interface for *Alien Encounters.*

Additionally, for this installation, *Vuforia* was exchanged for *Unity's ARFoundation*[1] plugin, which represented a more flexible and scalable option

[1] *ARFoundation* is a cross-platform AR SDK for use with the *Unity* game engine. It makes available the key features of *ARCore* (for *Android*) and *ARKit* (for *iOS*) to enable the authoring of cross-platform mobile AR applications.

over *Vuforia* largely due to licensing considerations, but also because authoring assets such as image targets could be stored locally, rather than having to be uploaded in advance to a third-party server space.

Another important change within the design of *Alien Encounters* compared to the previous installation is the way in which the radio static audio content is used in relation to the listener's embodied interactions within the installation. This time the static sound was not attached to either of our physical radio objects so as it creates the effect that they were its source, rather it was the sound that was experienced when the listener was beyond the *range* or *focal angle* of our virtual broadcasts, and it did not have any virtual or physical location within the virtual three-dimensional soundscape. In effect, this meant that its role changed from being a component of the virtual audio reality, to being a contextual framing device.

### 4.4.2 Authoring and deployment

With this view, and perhaps to better illustrate this point, let's look at the spatial interaction involved with this specific installation and its implications for the delivery of the audio content to listeners.

In Figure 18 both *broadcast A* and *broadcast B* have a *range* of 2 meters. There is an *overlap* of these broadcasts as they, and the real-world counterparts (the radios) are positioned approximately 1 meter away from each other. This overlap area represents a space where listeners can potentially hear both *broadcast A* and *broadcast B* at the same time.

The audio logic has been authored such that broadcasts are only audible to listeners when they are within its *range* and have a *focal angle* of the virtual sound source (or physical radio) of equal to or less then 40°. The broadcast signal attenuates from its centre, and eventually cross-fades to radio static on the periphery of its range.

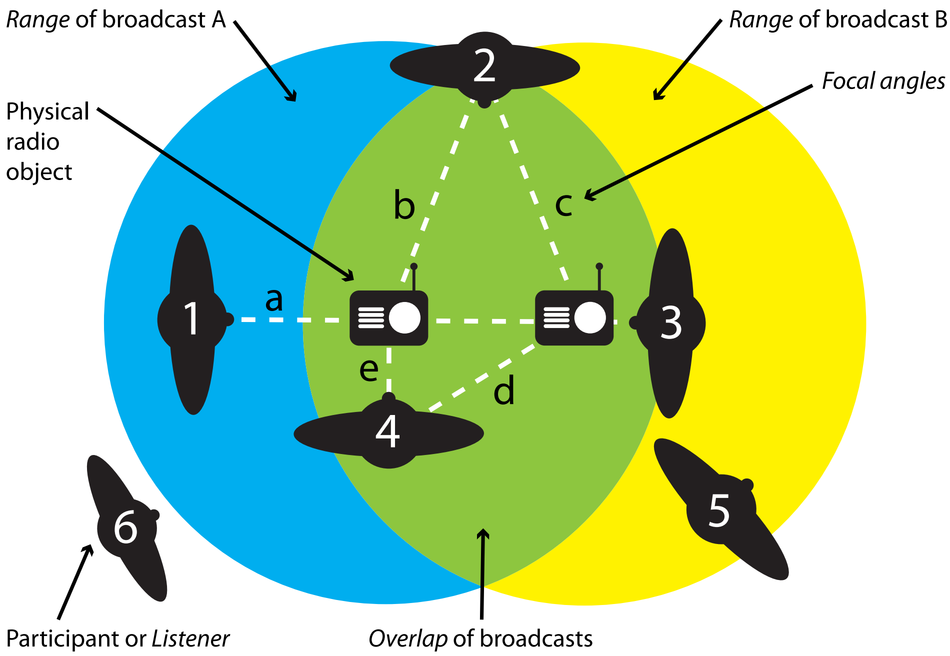


Figure 18. Spatial and audio interaction design for *Alien Encounters*.

We can look at each of the listener's positions and angles in relation to the radios illustrated in Figure 18, and estimate the content of the audio being delivered to them at their current locations:

- *Listener 1:* Will hear broadcast A only. Although they have a *focal angle (a)* of less than 40° of broadcast B, they are out of range of broadcast B.
- *Listener 2*: Will hear both broadcast A and broadcast B, as they are within both broadcast *ranges* and have *focal angles (b and c)* of both broadcasts of less than 40°. They may also hear some static as they are on the periphery of both broadcast *ranges.*
- *Listener 3:* Will also hear both broadcasts as they are within both *ranges* and *focal angles (a),* although broadcast A, for them, will be heavily attenuated.
- *Listener 4.* Is within both broadcast ranges, though they only have a *focal angle (e)* of less than 40° of broadcast A, therefore this is the only broadcast they will hear.
- *Listener 5.* Although within range of broadcast B, they have no focal angle less than 40°. This listener will hear only radio static.
- *Listener 6.* Although having *focal angles* less than 40° of both broadcasts, they are outside both broadcast ranges and so will hear only radio static.

A specific advantage of this model is that as the user's proximity to the location of a sound source shortens, or as they become increasingly focused on one particular source, their angle in relation to the other source becomes increasingly acute, and therefore becomes increasing attenuated until it is removed from their soundscape, allowing them to focus on their source or object of interest. This is clearly demonstrated by the position of listener 4 in Figure 18.

### 4.4.3 Reflections on practice

Prior to visitors taking part, instructions for use were seldom sought, if they were required this was usually due to visitors asking what they should do, and involved instructing the visitor to *'put on the headphones and explore the space'.* Many visitors wanted to know what it was, and sometimes how it worked, prior to having a go, after initially being attracted to the installation by the displayed vintage radios, the interactions of another participating visitor, or the spectacle of the interface. This observation reminds us that, although efforts are often focussed upon interactions with the virtual audio content, the physical object, along with the spectacle of physical interaction, can be the catalyst for interaction with the virtual audio content and therefore the physical and virtual models both require consideration.

The nature of the experience became apparent to the majority of visitors once they put on the headphones and heard the sounds. In these cases, the association between bodily movement and the delivery of the audio content became very quickly apparent, and often resulted in an enjoyable and engaging experience of exploration. At this point it was also apparent that visitors were making a quick connection between the virtual audio source and the physical object, this was made observable by their bodily movement towards the objects. This connection between the virtual audio and the physical artefact was often accompanied by confirmatory lateral sweeps of the head to evaluate the changes in the audio signal, similar to those described by Heller & Borchers (2014).

These lateral movements, observed during the initial stages of the participants interactions, may prove useful as an orientation device for the system as well as the user. These subtle and exploratory type of movements help the *SLAM* technology build and maintain a more accurate model of the physical environment within which it is operating (Apple, 2021c; Google, 2021). Again, if these could be established early on within the environment, perhaps whilst receiving initial instructions, they would help build a more robust world model from the outset. This procedure could perhaps be encapsulated within a system design that provides initial verbal instructions, along with an initial object to *find*, that both trains the user in what to expect from the experience, and serves to effectively *localise* the virtual map.

Through the use of the *angle* and *range* parameters of the individual sound sources, the ability for a listener to both focus on individual sound sources and compose a soundscape from the output of multiple sound sources seems entirely possible. Though, as mentioned previously, this is intrinsically linked to the shape and area of the physical architectural space and the positions of the audio augmented objects within it. As with the first spatial model (Figure 15) we can begin to see how this approach could be used in a larger space with multiple audio augmented objects, and how the peripheries of *ranges* can be used to advertise the location of objects within the soundscape. Additionally, it was also possible to identify the potential role of the *focus* parameter to help achieve this with more precision, especially when dealing with an area with multiple audio augmented objects.

Also observed was how the addition of audible radio static within the experience acted as a *sonic boundary* which defined the setting of the experience, in a similar fashion to Hazzard et al's *global* and *musical boundaries* (Hazzard, Benford and Burnett, 2015), and promoted focus on the physical artefacts. The listener's audio would cross-fade to static once they were beyond the range of a broadcast, it was at this distance that visitors were observed being physically repelled back towards the radio objects and their associated broadcast audio.

The use of static as a *non-spatialised* audio component may have contributed to this effect allowing the spatialised radio broadcasts to punctuate this ambience and advertise the position of the physical objects more effectively, promoting a sense of *spatial focus.*

Silence, the result of the full attenuation of the source signal due to increased distance between the user and the source, was occasionally perceived as a system failure. This may have also been due to the archive broadcast material reaching the end of its looped playback, and participants experiencing a short break prior playback starting again. Therefore, on reflection, it may be more appropriate to ensure that the ranges of transitional or background ambience better fit the architectural dimensions of the space.

The noisy environment, due to the large quantity of visitors, the close proximity of other demonstrations, and groups of visitors talking amongst themselves, led to some participants having difficultly hearing the virtual sound sources, even with the over-ear headphones and after turning up the iPhone's volume to maximum. This led to some visitors kneeling down directly in front of the radios to shorten the distance between the virtual sound source and headphone-mounted iPhone, thus decreasing the distance between the source and the user, and therefore decreasing the attenuation of the audio signal in an effort to render it audible. Although this was obviously problematic in terms of visitor's experience of the audio archive elements of the installation, it re-enforced and amplified, with significant clarity at times, the relationship and engagement between the visitor, the virtual audio and the physical artefact, with visitors interacting with the 'silenced' radio, perhaps just as they would with one that was working with limited volume in a noisy environment.

The headphone-mounted iPhone interface, with the iPhone's camera facing forward and the its screen facing towards the rear of the visitor, presented itself as an advert to prospective, spectating participants. The screen showed the cameras view of the radio objects along with the superimposed virtual objects representing the position of the sound sources.

Whilst the particularly low lighting level in the gallery facilitated this ability of the experience to advertise itself via its backwards facing and glowing screen, it also compromised the reliability of the system that had been previously experienced in a different setting. Whilst reliability issues with AR SDK's tracking and mapping had been anticipated to some extent due to the low lighting conditions (Apple, 2021c; Google, 2021), and a table light was used to light the radio objects themselves in an attempt to overcome any problems, the virtual sound sources required relocating after every third or fourth experience. The extent to which these virtual objects *drifted* went well beyond a threshold useable for the purposes of directly associating a virtual sound source with a physical object, without regular repositioning.

Another observation of specific interest involved a visitor, who had previously taken part, attempting to attract the attention and direct their co-visitor, who was currently interacting with the experience, towards a specific location they themselves had previously encountered. This point, which illustrates a desire amongst listeners to share content they have found with others, highlights the opportunity for facilitating sharing and social interaction amongst participating listeners within the system. At the same time this also highlights the difficulties in squaring such a desire to communicate with the inherently isolating experience of a personal audio experience delivered via headphones. This subject is discussed further in section 7.3.3.

Many visitors exchanged post-experience comments with the invigilating researcher which often reflected upon the *'magical'* and *'real'* nature of their experience along with an occasional disbelief that the radios were not actually switched on and the source of the audio content themselves. This suggested that the *head-based* tracking approach did indeed give a greater perceived sense of reality over the previously deployed *device-based* approach. Perhaps of even greater interest here is the inability of some listeners to accept that the radios were not the actual source of the audio broadcast content that they were hearing, suggesting that the experience created a situation where virtuality was, at times, indistinguishable from reality. These experiences with this installation highlight the distinction between the *willing* and *unwilling* suspension of disbelief (Elferen, 2012); the difference between a *conscious* suspension (in the literary sense) in order to pursue the authors intent, and an *unconscious* suspension, reflecting an immersion in virtuality as reality. This observation is discussed in detail within section 7.1.

This is, perhaps, made all the more fascinating due to how the *acoustic virtual reality* is manipulated through the rendering of the radio static audio content as an ambient feature, rather than as a spatially positioned sonic artefact of the radio (as it would persist in a real acoustic reality) in that this factor did not seem to distract from the perception of an acoustic reality.

As with the *device-based* approach in the previous installation, there appears to be a moment of delight that accompanies the association between bodily movement and the interactive audio content delivery. Perhaps even a short and subsequent moment of playful experimentation and interaction with the positional qualities of the sound source. Additionally, positive comments were also exchanged regarding the free, nomadic nature of the interaction in comparison to a tethered VR experience and its somewhat cumbersome interface.

Regarding this installation's specific spatial interaction design, it appears that a successful model for the effective deployment of audio augmented objects within galleries could exist in a compromise between the coherent exploration of the environment, and a *suspension of disbelief* rendered plausible through the accurate replication of the real physical qualities of sound. In other words, what would appear to both facilitate exploration and engagement with museum objects, and render a compelling *acoustic virtual reality*, is a compromise between adjusting the physical properties of the virtual acoustic space in order to maintain functional coherency, and adhering to them for a sense of *acoustic virtual reality.* This point is discussed further in section 7.2.3.

By way of summary, this installation has demonstrated how AAR can attach virtual sound sources to real-world objects, and how it has the potential to enable the creation and curation of perceptually realistic acoustic virtual realities comprised of audio augmented objects that can help to promote the exploration of museum and gallery artefacts and provide a physical interface to digitised audio archival material. It has also demonstrated a workable object detection and nomadic indoor positioning solution by way of SLAM-based AR that extends the capabilities of art and artefacts to advertise their presence to visitors through audio augmentation.

## *4.5 Horror-Fi Me*

Originally proposed and accepted as a site-specific sound installation for the *Live Cinema III: Festival of Research and Innovation 2020*[1], *Horror-Fi Me* was subsequently adapted to allow users to author an AAR experience within their own homes, and using their own equipment, in order that it could be deployed under the circumstances of COVID-19 *lockdown.*

[1] *Live Cinema III: Festival of Research and Innovation* presents cutting edge academic research, master classes, workshops and a programme of screenings. The 2020 event took place online: http://livecinema.org.uk/live-cinema-iii-festival-of-research-and-innovation-2020/

*Horror-Fi Me* was designed with the intention of utilising the facets of the existing relationship between audio and horror as outlined by Barrios-O'Neill (2018). It was envisaged that previous and recent experiments with static binaural audio within this sub-genre, such as those demonstrated in peter Strickland's adaptation of *The Stone Tape* (Strickland, 2015), could be further built upon by realising an audio horror experience that utilised dynamic binaural audio with a full *six-degrees-of-freedom*[1].

Of course, given *Audio horror's* interest in the realism and effectiveness afforded by the static binaural audio experience (Barrios-O'Neill, 2018; Hancock, 2018), it is anticipated that there would be interest in a technological advancement that has the potential to create an even greater sense of three-dimensional audio realism within this specific genre. But it was also anticipated that the following project would provide insights into the application of dynamic binaural audio across other contexts, such as within museums and art galleries, as well within domestic settings.

Although dynamic binaural audio is seen as a key component part of an audio augmented reality (AAR) experience (Bauer *et al.*, 2019), this project further extends such audio content into the realm of AAR by incorporating *audio augmented objects* and features from the listener's environment within the experience in a similar way to the previously described installation experiences (see sections 4.2 to 4.4). Not only does the listener have their own *sound, acoustic* or *auditory perspective*[2] on the experience and has a degree of agency over *choreophonical design*[3] afforded to them by the freedom of spatial interaction, but the virtual audio content is provided with physical anchors in the shape of objects and features within the listener's reality. The importance of the freedom of spatial interaction within the audio augmented object reality is discussed further in section 7.2.4, the implications of the audio augmented object on the perception of reality are discussed further in section 7.1.

By association, it seeks to obtain insights into how an *audio horror* may be realised that could rival the immersive qualities of video games and television using a similar combination of technologies as those suggested by Hancock (2018). Again, it was intended that this association would be by no means exclusive, as any such findings would surely be transferable across multiple

---

[1] *Six-degrees-of-freedom* or *6DoF* refers to the freedom of movement in three-dimensional space and includes the three degrees of rotational movement (*pitch*, *yaw* and *roll*) and the three degrees of translational movement (*surge*, *heave* and *sway*).

[2] Crook (1999) and Ouzounian (2020) define *auditory, sound or acoustic perspective as being* the listening *point-of-view* or, rather *point-of-listening,* within an acoustic space specifically in relation to the *point-of-listening* within a simulated or remotely experienced acoustic environment. Additionally, Crook (1999), draws a comparison between *perspective* as the *visual* representation of three-dimensional objects and volumes of space, and *sound perspective's* ability to achieve the same.

[3] Crook (1999) defines *choreophonic design* as the composition of the listener's *acoustic perspective* or, rather poignantly, as the *'picture space'* for the imagination.

genres and contexts. This concept is very much echoed by Esslin *in* Crook (1999) in relation to radio drama, where he suggests that the listening audience become the visual directors of their own experience as the visual imagery conjured up is a matter of personal choice and preferred consciousness. Being a consequence of *'performance in time and space'*, this, he suggests, has the ability to create a visual quality arguably higher than that of television, film and theatre.

*Horror-fi Me* also aimed to elaborate on Hancock's vision of a highly immersive screenless, audio-only experience by investigating the role *audio augmented objects* have to play in providing a surrogate screen for such experiences. The idea of exploring the possibility of a *screenless cinematic experience* is also a response to the work of artist Janet Cardiff (Aceti, 2013) and Michael Bull (2000). Artist Janet Cardiff, who is perhaps best known for her soundwalks, along with sound studies theorist Michael Bull, both demonstrate how cinematic experiences can be realised through the augmenting of everyday life with mobile audio content delivery technologies. The personal stereo listener, and indeed more recently, and with greater efficiency, the iPod and iPhone listener, rely on the possibility of serendipitous collisions between linear, stereophonic audio content and their experiences of everyday life, such as those analysed in Bull's *Sounding Out the City*, for the realisation of what is described as a '*filmic experience'.*

Cardiff's soundwalks, along with other locative listening experiences, extend this filmic experience beyond the aestheticising potential of linear listening's serendipitous encounters. This is typically achieved by purposely authoring interactive audio experiences which trigger specific segments of pre-recorded audio content based on a listener's GPS coordinates. AAR enables the finer-grained audio augmentation of specific parts of a location, or an architectural structure, as well as the objects or artefacts within. It also enables, through the determination of the listener's orientation, the authoring of interactive and spatialised, binaural soundscapes for the listener to explore. The augmenting precision this technology affords the locative and mobile listening experience leads one to hypothesise that an AAR experience could go beyond the capabilities of linear stereophonic mobile audio listening, and beyond the capabilities of GPS enabled soundwalks and location-based listening experiences, in terms of reframing a listener's perception of their location. Furthermore, one could hypothesise that an immersive, live, cinematic and fictional encounter with a real-world location could be experienced.

*Horror-Fi Me* therefore invited listeners to enter into, and become composers of, their own horror movie soundscape, a live cinematic experience that sought to promote exploration, intrigue and suspense through the creation of a live cinematic experience where reality is the screen.

Interest in the creation of a cinematic experience and the creation of virtual space extended to the use of music to provide an accompanying musical score to the listener's experience. In much the same way the soundtrack for a video game provides a contextual musical accompaniment to the personal adventures of the player (Phillips, 2014, p. 70), it was intended that the inclusion of a musical soundtrack for *Horror-Fi Me* would help to ground the experience within the genre of horror cinema and provide a soundtrack for the listeners' adventures around their homes. Furthermore, the association between the contextualisation of the mode of interaction and the memory of real and imagined experience established within the preceding installations suggested that the contextualisation of *Horror-Fi Me* through the use of background music would help to associate the experience with participants' memories and experience of horror cinema.

As such, Nick Redfern, an award winning contemporary classical composer whose style fitted the style of music traditionally used for scoring film and video games within the horror genre, was approached to write a score for the experience.

Finally, following on from the *Alien Encounters* installation (see section 4.4), it seemed prudent to create an experience comprised of multiple audio augmented objects across a larger interactive setting, within which the listener's experience of audible objects that are both beyond and on the periphery of their vision could be explored. Such an experience would have the additional potential to gain further insights into the functional properties of different types of sounds and objects, and would facilitate an investigation into the creation of a sonic atmosphere as a result of multiple audio augmented objects. It was also envisaged that the creation of a contextually rich virtual sonic atmosphere and its deployment within an ordinary setting would enable the combined ability of virtual audio and physical reality to evoke memory, imagined experiences and to generate affect and a sense of telepresence to be investigated more generally.

### 4.5.1 Design and development

In the previous installations (see sections 4.2 to 4.4), the positions of the sound sources have either been pre-authored and built-in as part of the application's design and the interactive settings were pre-determined (a specific gallery space or dedicated area within a museum). With *Horror-Fi Me* the arbitrary nature of the deployment environment (beyond the assumption that it will be a domestic space) necessitated that a generic and accessible authoring environment be built into the mobile application that would facilitate the attaching of virtual audio content onto available domestic features, locations and objects. Although deploying *Horror-Fi Me* in this manner was not without its challenges, namely incorporating a generic and accessible frontend authoring process into the application, it did unexpectedly present an opportunity to study the remote authoring process, though this did come at the expense of being able to finely craft a site-specific AAR installation experience that could more carefully consider the architectural confines of an experience's setting.

As originally proposed, the application allowed real-world objects and locations to be identified and have virtual audio content attached to them. Within the lockdown adapted version the user performed part of the experience's authoring process themselves by using the mobile application to *tag* key objects, locations and features within their domestic environment that the application then used to construct a cinematic reframing (see Figure 20). Users were also able to save and reload their authored experience for others, or for themselves, to revisit again at a later time.

### 4.5.2 Authoring and deployment

Due to circumstances requiring participants to utilise their own hardware, efforts were made to make what were considered to be relatively new technological developments as accessible as possible. In some cases, this has meant only using elements embedded within the employed software development kits that were capable of performing with the lowest specification of user devices possible, rather than those that are known to produce better results, though only on more recent models.

*Horror-Fi Me* was deployed as an iOS application and made available for users to download through *Apple's TestFlight*[1] application by publishing and promoting a download link. The application's minimum required hardware specification was an iPhone SE 1st generation, 2016 model or higher, with the only additional equipment required being a standard set of stereo headphones.

It was originally intended that the *Bose AR SDK* (Bose, 2018) would be included within the application to provide owners of *Bose AR* enabled frames and headphones with the option of connecting to the application and realising a dynamic binaural audio experience via headtracking. Unfortunately, this was not possible due to *Bose* discontinuing their *AR SDK* (Roettgers, 2020). This was perhaps illustrative of the developmental nature of this technology, at a consumer level at least. The only other presently viable option for realising a dynamic binaural headtracking experience that can be remotely deployed for use on users' own equipment appears to be *Apple's AirPod Pro* (Apple, 2021a). Unfortunately, these were released when development on the

[1] *TestFlight* is an online service operated by *Apple* for testing iOS mobile applications. It allows for the remote distribution of development versions of new iOS application to beta testers, and for the distribution of a publicly available download and installation link for the app without the need for commercial distribution via *AppStore:* https://developer.apple.com/testflight/

application was almost complete, as such, *Horror-Fi Me* continued to rely on device tracking to realise dynamic binaural audio[1].

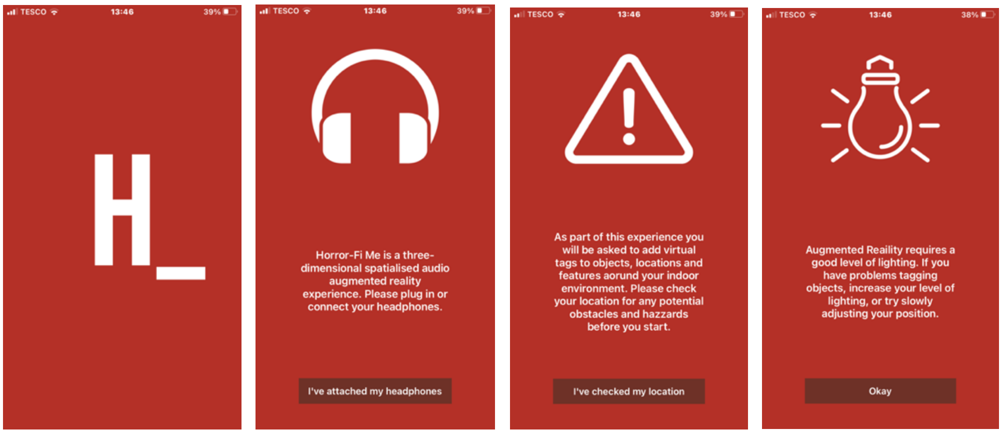


Figure 19. *Horror-Fi Me* mobile application introductory screens.

The launch and introductory screens (see Figure 19), along with the application's name, provide some context and indicate to the user that this is an audio augmented reality experience. They also ensure that the user has attached their headphones, undertaken some basic safety checks for potential trip hazards and obstacles, and that they are undertaking the experience within an environment with a good level of lighting before they can proceed. It also provides some basic information on how to overcome problems they may encounter within a poorly lit environment. A well-lit environment remains, for the time being, a key prerequisite for robust and reliable mobile AR experiences of all kinds (Cervenak and Masek, 2019).

On proceeding beyond the introductory screens, users are presented with the main authoring interface of the application which allows them to attach a series of audio samples with pre-defined attributes to the locations of objects and features within their real-world environment (see Figure 20). This is achieved via a process of *tagging*, where users are instructed through the interface to place a series of virtual *tags* within their real-world environment at the locations of the real-world objects and features. The specified locations of these *tags* represent the coordinates in three-dimensional space where the appropriate audio samples will be placed to form the virtual soundscape of the experience. With the positions of these *tags* being in approximately the same locations as the objects and features that the user has been instructed to identify to the application, the virtual sound sources can then be perceived by the listener to emit from these locations, and potentially from the objects themselves, and *audio augmented objects* are able to be realised within the rendered dynamic binaural soundscape.

[1] A build of *Horror-Fi Me* was subsequently released via *Apple's TestFlight* that included support for head-based tracking for users with *Apple AirPod Pro* earbuds as well as device-based tracking for users with standard stereo headphones.

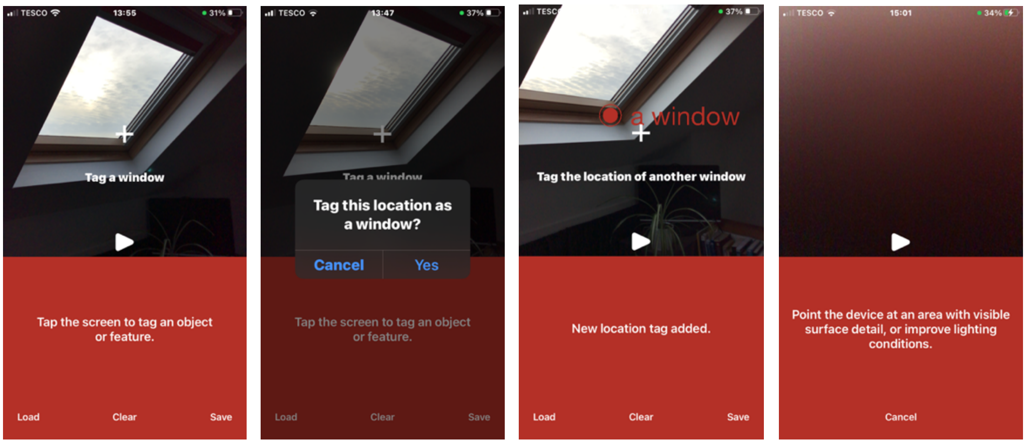


Figure 20. *Horror-Fi Me* mobile application screenshots. Showing *tagging* process and user feedback.

The tagging process closely resembles visually orientated AR mobile applications where the coordinates of real-world objects need to be identified in order to place virtual graphical content at the same spatial coordinates. One such example of this is *Placenote's StickyNotes* (Placenote, 2021), a *virtual sticky notes* mobile application that allows users to write and attach virtual notes to real-world objects.

As shown in Figure 20, the user is initially presented with a target crosshair, a message indicating what feature or object to tag, and a play button superimposed over the camera view of their environment. This is presented to the user above an interface panel that includes a message area used for feeding back to the user information about the current state of the application. Additionally, we can see the *load*, *clear* and *save* buttons at the bottom of the screen, enabling the user to load a previously authored soundscape, clear the currently authored soundscape or save the currently authored soundscape.

The user is instructed to tap the screen to add a tag for the current object or feature (as indicated in the message panel underneath the crosshairs) at the current location. After being asked to confirm their selection, we can see the position of the virtual tag along with the object or feature's name within the real-world environment. The message panel below the crosshairs is updated to instruct the user which object to tag next (in this case the user is being instructed to tag another window), feedback is also provided to the user within the bottom panel that a new tag has been successfully added.

The screen on the far right of Figure 20 shows a message being relayed to the user to point the camera at an area with more detail or increase lighting, a consequence of the SLAM-based tracking and mapping being interrupted by an obscured camera. Other interruptions to tracking can be detected and information relayed to the user on how to remedy them, including when tracking is still being initiated, and when movement is detected that is too fast to allow for reliable mapping and tracking of the environment.

| Description | Sound File | Range | Rendering | Reverb | Loop | Virtual Tag | Type | Position | Max. Volume |
|---|---|---|---|---|---|---|---|---|---|
| musical score | horror-fi-music-stereo.aiff | 0 | stereo | 0 | false | Window 1 | contextual | none | 0.9 |
| wind | wind.aif | 4 | binaural | 0.2 | true | Window 1 | object | none | 0.4 |
| rain on window | rain.aif | 6 | binaural | 0.2 | true | Window 1 | object | none | 0.4 |
| wind 2 | wind2.aif | 4 | binaural | 0.2 | true | Window 2 | object | none | 0.4 |
| rain on window 2 | rain2.aif | 6 | binaural | 0.2 | true | Window 2 | object | none | 0.4 |
| wind 3 | wind3.aif | 4 | binaural | 0.2 | true | Window 3 | object | none | 0.4 |
| rain on window 3 | rain3.aif | 6 | binaural | 0.2 | true | Window 3 | object | none | 0.4 |
| wind 4 | wind4.aif | 4 | binaural | 0.2 | true | Window 4 | object | none | 0.4 |
| rain on window 4 | rain4.aif | 6 | binaural | 0.2 | true | Window 4 | object | none | 0.4 |
| thunder | thunder.aif | 40 | binaural | 0.3 | false | Window 1 | object | x: 0.0, y: 20.0, z: 0.0 | 0.9 |
| crows | crows.aif | 8 | binaural | 0.3 | false | Window 2 | object | x: 0.0, y: 2.0, z: -1.0 | 0.5 |
| drips | drip.aif | 5 | binaural | 0.2 | true | Sink/Tap | object | none | 0.3 |
| door knocking 1 | door-knock01.aif | 5 | binaural | 0.2 | false | Front Door | object | none | 0.7 |
| door knocking 2 | door-knock02.aif | 5 | binaural | 0.2 | false | Front Door | object | none | 0.7 |
| door knocking 3 | door-knock03.aif | 6 | binaural | 0.2 | false | Front Door | object | none | 0.7 |
| door knocking 4 | door-knock04.aif | 6 | binaural | 0.2 | false | Front Door | object | none | 0.7 |
| door knocking 5 | door-knock05.aif | 8 | binaural | 0.3 | false | Front Door | object | none | 0.8 |
| smashing cup | smashing-cup.aif | 8 | binaural | 0.1 | false | Kitchen Cupboard | object | x: 0.0, y: -1, z: 0.0 | 1.0 |
| smashing plate | smashing-plate.aif | 12 | binaural | 0.2 | false | Kitchen Cupboard | object | x: 0.0, y: -1, z: 0.0 | 1.0 |
| smashing teapot | smashing-teapot.aif | 8 | binaural | 0.1 | false | Kitchen Cupboard | object | x: 0.0, y: -1, z: 0.0 | 1.0 |
| ticking clock | ticking-clock.aif | 8 | binaural | 0.1 | true | Clock | object | none | 0.6 |
| chimes | chiming-clock.aif | 8 | binaural | 0.2 | false | Clock | object | none | 0.7 |
| ghostly breath 1 | ghostly-breath-1.aif | 5 | binaural | 0 | false | Window 1 | personal | x: -0.50, y: 0.5, z: 0.0 | 0.7 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ghostly breath 2 | ghostly-breath-2.aif | 5 | binaural | 0 | false | Window 1 | personal | x: 0.50, y: 0.5, z: 0.0 | 0.7 |
| creak 2 | creak02.aif | 6 | binaural | 0 | false | Window 1 | personal | x: 0.0, y: -1.0, z: 2.0 | 0.4 |
| creak 1 | creak01.aif | 6 | binaural | 0 | false | Window 1 | personal | x: 0.0, y: -1.0, z: 2.0 | 0.4 |

Table 2. *Horror-Fi Me's soundLibrary*.

Table 2 represents the application's *soundLibrary* (an array of *Sound* objects) which contains all the *Sounds* available for the construction of the virtual soundscape within *Horror-fi Me*. All the audio for the experience, with the exception of the musical score, was obtained from the online BBC sound effects library (BBC, 2022) and, if required, was converted into mono for use within the binaural soundscape. *Apple's AVFoundation*[1] (the audio-visual framework upon which *Horror-Fi Me* is built) requires audio files to be monophonic in order for them to be spatialised and binaurally rendered. The sounds that were required to play on a loop were carefully cropped so as to achieve seamless audio playback and, in the case of the *dripping tap* and *ticking clock* samples, in order to preserve the sample's consistent timing.

Each *Sound* has a set of attributes that determine its behaviour within the soundscape. We can see in Table 2 that each *Sound* has, along with other attributes a *Type* of either *object, personal or contextual*. A *Sound* which is designated a *Type* of *object* will be positioned at the same location as its specified real-world object, unless the *Sound* has a position specified, in which case the its position will be relative to the location of the *tagged* real-world object.

[1] *AVFoundation* is an audiovisual framework for developing on MacOS and iOS devices. See: https://developer.apple.com/av-foundation/

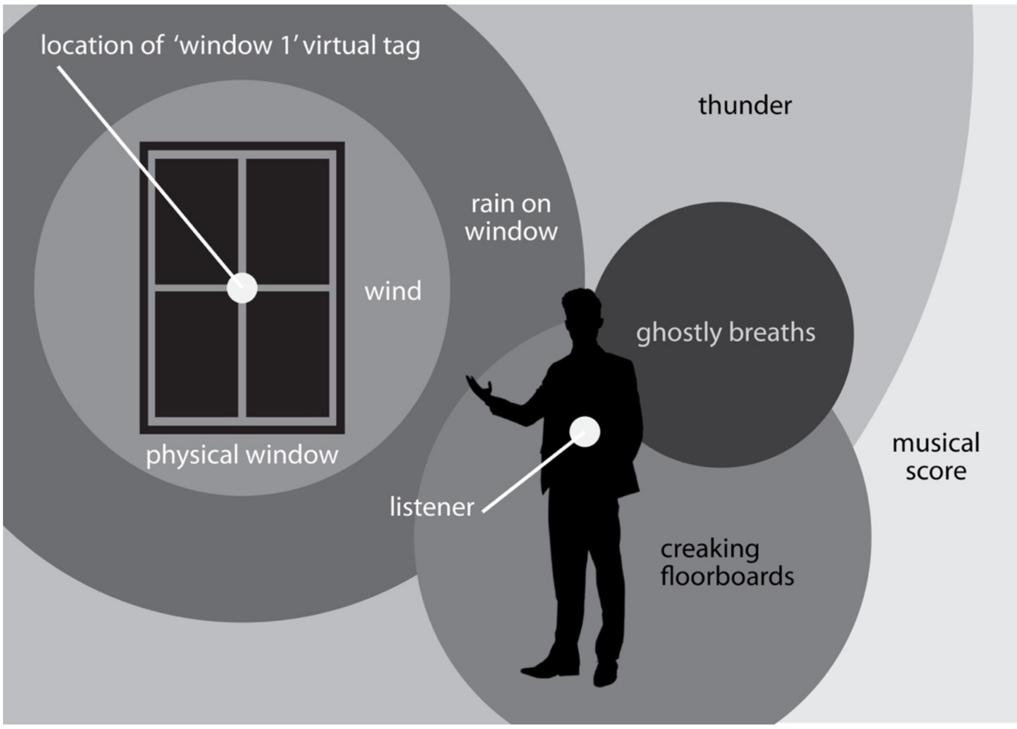


Figure 21. The result of tagging a window in *Horror-Fi Me*. The placement of the '*window 1'* tag and the resultant rendered *Sounds* within *Horror-Fi Me's* virtual soundscape.

The example in Figure 21 illustrates how one virtual tag can specify the position of multiple *Sounds*, each with their own attributes. We can see from Table 1 that the *wind, rain on window* and *thunder Sounds* all have the same virtual tag. When this tag is placed by the user, all three of these *Sounds* are included within the soundscape and their position within it is specified as being the same as the tag (if the *Sound* has a position of none) or relative to the position of the tag (if the *Sound* has a position specified). In the case of *window 1* (see Figure 21) we can see that the centre of both the *wind* and *rain on window Sounds* are positioned on the location of the tag and the *thunder* is positioned 20 meters above the location of the *window 2* virtual tag, having a three-dimensional coordinate value of*: x:0, y:20, x:0*. This effectively places the wind and rain on the window and the sound of the thunder in the sky. In addition to this we can see in Figure 3 that these three *Sound*s all have different *ranges* as indicated by their radial size. A *Sound*'s *range* specifies the maximum distance in meters from the centre of the source's location from which it will be audible by the listener. The approximate audible ranges of the sources are indicated by the diameter of the circles in Figure 21.

Figure 21 also indicates the approximate position of two *personal* type *Sounds* (*ghostly breaths* and *creaking floorboards*) that are included within the soundscape when the *window 1* tag is added by the user. While *personal* type *Sounds* work in much the same way as the *object Sounds*, their position always remains relevant to the current position of the *listener* in real time. This makes it possible to have *Sounds* that follow the listener, are positioned on the listener's shoulder, or will always be positioned at a certain distance above, below, in front, to the rear or to the side of the listener's current and real time position within the experience. Within the *Horror-Fi Me* experience *personal* type *Sounds* are used to ensure that a ghostly breath sound is always positioned just to the side of the listener and a creaky floorboard sound is always positioned behind the listener.

We can also see the stereo rendered musical score that is always audible, having no spatial location or defined range and specified as a *contextual Sound type* within the sound library. The *contextual Sound type* is reserved for non-binaurally rendered, non-positional audio content that can play alongside the other content, in this case this is used to provide the stereo musical score for the experience. Both the *contextual* and *personal* type *Sounds* are always rendered with the first virtual tag even though they have no direct, or relative, real-world stationary location. This is to ensure that they are rendered within the soundscape if the experience is played with just one virtually tagged location, which is the minimum number of *tags* required in order to *play* the experience.

By rendering multiple *Sounds* with varying attributes at a variety of locations within the soundscape as a result of placing just one virtual *tag* it is possible to quickly build a rich and complex virtual soundscape and audio augmented reality.

Within *Horror-Fi Me*, all non-looping *Sounds* are triggered randomly at a point in time over the duration of the experience, the length of which is determined by the musical score. The musical score also determines the opportunities at which these randomly occurring, non-looping *Sounds* are triggered. The time (in seconds) within the musical score that contains periods of low dynamic levels were identified and placed within an array, and non-looping, spatialised *Sounds* are chosen at random to play during one of these moments. The looping, spatialised *Sounds*, such as the *rain* and *wind*, provide additional ongoing spatial and contextualised content for the experience.

The objects and features that the users were asked to *tag* within their homes were selected in an attempt to try and realise a reasonable sized space within which the experience would be set and also to try and ensure that a good separational distance between *Sounds* would be achieved. By trying to ensure the best possible separation between the Sounds, within an environment of indeterminate proportions, I was also attempting to ensure that a reasonable level of spatial exploration of the both the physical and audible environment would take place. Such an approach also sought to ensure that fundamental experiences associated with the embodied interaction with spatial audio could take place, such as the discovery of new *Sounds* (*Sounds* becoming audible) and the disappearance of *Sounds* (*Sounds* becoming inaudible) through physical movement. To account for the indeterminate shapes and sizes of the deployment environment, all the wind and rain *Sounds files* were adjusted to be slightly different so as to avoid *phasing*[1] issues between identical sources that could potentially be audible at the same time from different locations within the rendered 360-degree virtual soundscape. This said, care was taken to ensure that these samples still sounded as though they were the result of the same rain storm happening outside of the same house in order to maintain contextual continuity throughout the experience. In practice this was achieved by cropping the same audio loop at different points, as all rendered audio content begins playing simultaneously, this meant that no loop of the same audio content would be playing from the same point at the same time.

A sense of contextual continuity was also provided by the musical score, a *stereo rendered Sound* that takes place within the listener's head, in the same way a usual stereo headphone-based listening experience would, though independent of the external, spatialised sounds of the augmented audio reality. It was conceived that this would provide an appropriate musical accompaniment to the experience that would reflect and reference the cinematic score of a horror movie. A contemporary classic musical composer, whose style, technique and compositional style reflects that of the genre of horror cinema, was commissioned to write a score for the experience. The approach taken was very much in line with that described by composer Winifred Phillips (2014) in reference to the commissioning of video game music, where, rather than directly composing for the experience, an appropriate piece of music is commissioned that stands on its own merit, and then is repurposed within the context of the experience. Additionally, through the popular association of contemporary classical music with the horror movie genre, the musical element of the experience becomes the contextualised soundtrack to the personal adventures of the listener, again, in a similar way to that described by Phillips (2014) in relation to video games.

It should be noted that there is no explicitly authored or intended narrative to the experience other than the musical score which, being composed in the sonata form, naturally comprises a beginning, middle and end in the form of an exposition, development and recapitulation. At this stage, the experience, for the main part, intends to realise a virtual cinematic soundtrack within the listener's reality beyond that which is achievable through a linear, stereophonic listening experience, rather than overtly imposing a narrative structure upon it.

In summary, there are three different *Sound types* each having their own positional qualities:

---

[1] *Phasing* describes the timing differences when combining identical (or near identical) signals. This can result in unwanted increases or decreases in amplitude of the signal and even complete signal cancellation.

- *objectRelated:* related to the position of a stationary real-world object or feature
- *personal*: related to the listener's current position
- *contextual*: not related to the position of either the listener or a real-world location

In addition to its *position*, each *Sound* data type has a number of other definable members, or parameters, (as indicated in Table 2) that facilitate the authoring of the AAR experience, these are as follows:

- *file:* The reference to the *Sound's* audio file.
- *range:* The distance in meters from which the *Sound* is audible. This determines the distance at which the *Sound* will be fully attenuated.
- *rendering:* The spatial rendering algorithm that will be applied to the *Sound*. (For example, HRTF is used for binaurally rendered and spatialised *Sounds*, stereo and *Ambisonic* being other possible options.)
- *reverb:* The amount of reverb to be applied to the *Sound*. For example, this can be adjusted to match the probable acoustic qualities of the location where the *Sound* will be positioned.
- *loops:* A Boolean value to indicate if the *Sound* will loop or will be a *one-shot.*
- *object:* A reference to the object that the *Sound* will be associated with (if any).
- *descriptiveName:* A human readable name for the *Sound* (mainly for displaying within the user-interface).
- *type*: A *Sound*'s *type* determines how it will be positioned within the soundscape.
- *position:* The position of the *Sound*, depending on its *type.* A *Sound* with no specified position will have the same position as it's associated real-world object (if it has one). If a *Sound* does have a specified position, the specified position will be relative to the position of its associated real-world object. This is useful for creating a cluster of different *Sounds*, or 'child *Sound*s' with different attributes and positions around a single, real-world object or feature (as illustrated in Figure 21).
- *maxVolume:* The maximum volume that the *Sound* can achieve.

The *soundLibrary* (Table 2) represents a library of possible *Sounds* (along with their characteristics) that can be used within a soundscape. Once they are used they become instances of a *Source* within a soundscape. This is so we can keep track of which *Sounds* are currently in use within a soundscape, and control their playback and other values within the audio experience.

The members that make up a *Source* data type are as follows:

- *audioPlayer: The audioPlayer object that will play the Source's sound file.*
- *buffer: The audio buffer to load the audio file into.*
- *sourceNode: The AR node to which the source's audioPlayer object will be attached (this provides the audioPlayer object with its position in virtual and physical space.*
- *id: A unique ID so it can be identified if required.*
- *object:* A reference to the object that the *Source* will be associated with (if any).
- *descriptiveName:* A human readable name for the *Source* (mainly for displaying within the user-interface).
- *range:* The distance in meters from which the *Source* is audible. This determines the distance at which the *Sound* will be fully attenuated.
- *distance:* Variable to hold the current distance in meters between the Source and the listener.
- *loops:* A Boolean value to indicate if the *Source* will loop or will be a *one-shot.*
- *reverb:* The amount of reverb to be applied to the *Source*. For example, this can be adjusted to match the acoustic qualities of the location where the *Source* will be positioned.
- *rendering* The spatial rendering algorithm that will be applied to the *Source*. (For example, HRTF is used for binaurally rendered and spatialised *Source's*, stereo and *Ambisonic* being other possible options.)
- *position:* The position of the *Source*.
- *type:* A *Source*'s *type* determines how it will be positioned within the soundscape.
- *volume:* Variable to hold the Source's current volume.
- *file:* The name of the *Source's* audio file.
- *maxVolume:* the maximum volume that the *Source* can achieve.

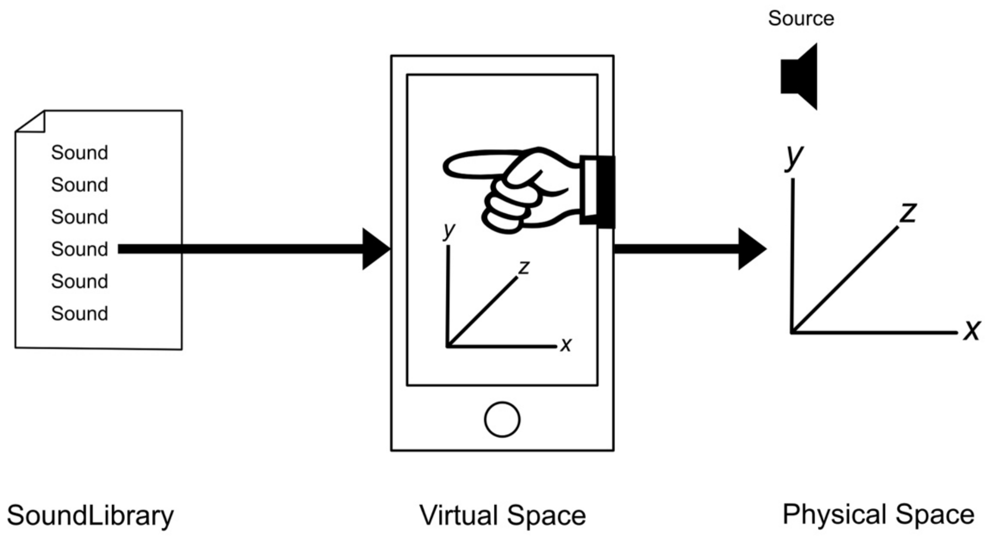


Figure 22. Diagram showing the conversion of a *Sound* to a *Source*. From the *SoundLibrary*, through virtual space, to a perceived location in physical space.

Whilst this conversion from a *Sound* to a *Source* uses many of the same values, it adds some important *objects* and *variables* that enable it to become an interactive audio source within our soundscape. These *objects* include an *audio buffer*, an *audio player* and the *AR node* to which the audio player will be attached, and which in turn will be attached to the *root node* of the soundscape. It also adds important *variables* such as *volume* (the current volume of playback of the audio player) and *distance* (the current distance between the source and the listener).

By keeping track of the distance between the sound source and the *listener,* and using the sources *range* and *maxVolume* values, the level of attenuation of the *Source's* volume can be calculated based on the *listener's* current position. These variables ensure that the correct volume of the *Source* is perceived by the listener both in relation to the *listener's* current position, and in relation to the individual characteristics of the *Source* itself, namely it's *position, range, reverb, volume and maxVolume* parameters.

### 4.5.3 Reflections on practice

Whilst the *AVFoundation* framework allows for custom values to be set for a distance attenuation model across the whole audio environment, determined by the parameters of the *AVAudioEnvironmentNode*[1] object, it does not

[1] The *AVAudioEnvironmentNode* within the *AVFoundation* framework is responsible for the simulation of the virtual 3D audio environment, and defines properties for the distance attenuation and reverberation of all the virtual audio sources attached to it.

provide a method through which distance attenuation characteristics can be determined for individual sound sources. In essence the framework assumes, quite rightly, that the perceived acoustic characteristics of the sound within the virtual environment should be determined by the acoustic characteristics of the environment that it is within. This assumption makes perfect sense when we want sound sources to be perceived within the virtual environment just as they would be within a real physical environment, such as sound sources situated within virtual spaces with differing acoustic properties of a video game. But this model creates problems for the construction and deployment of audio augmented reality experiences, as well as for realising audio augmented objects.

Firstly, *Horror-Fi Me* was designed for use within any indoor domestic environment, the dimensions and acoustic properties of which were largely unknown. For this reason, rather than adopting a specific audio environment with specific characteristics, a more generic approach was required that would render the sources as *possible* or *probable* within a variety of real-world spaces. A default audio environment for the space was therefore adopted and the *Sources* were attributed with their own characteristics (*range, reverb, rendering* and *type*) so that they could be individually adjusted for the best possible perception of their reality in general space. This approach also proved useful for the rendering of sources that were intended to be perceived as existing outside of the *acoustic space*, such as the sound of crows outside the window, or the sound of thunder. This meant that the range, reverb and attenuation of these virtual sound sources could be individually adjusted to better reflect their behaviour within their intended reality, rather than being determined by the audio environment of the audio engine.

Secondly, not having the ability to control distance attenuation at an individual sound level resulted in incoherent and indiscernible sound sources, or rather, created a soundscape of sound sources that became part of an over-arching ambience, rather than being individually discernible and easily perceptible as coming from a specific location. By individually controlling these characteristics of the sounds much more control could be leveraged over the authorship of the audio experience; specific sounds could be authored to travel further than others and attract a listener’s attention, controlled zones of ambience could be created, and, perhaps most importantly, limiting the ambient range of a source made it easier to localise and isolate the source (when in range) than within a soundscape comprised of sound sources with largely equal volume levels. It therefore appeared that this approach facilitated the perception of a more precise augmentation and the perception that individual physical objects were augmented, not just an approximate location in three-dimensional space.

Finally, this point remains of interest, not just regarding the listener’s perception of the object-audio relationship, but also for the crafting of three-dimensional, audio-based interactivity in general. For example, it would not have been possible to contextualise the mode of interaction through the use of radio static in *The McMichael Experiment* (see section 4.3) or *Alien Encounters* (see section 4.4) without applying a distance attenuation model at a source level for both the broadcast content and the sound of radio static.

Distance attenuation characteristics were determined for individual sound sources within *Apple's AVFoundation* framework through the inclusion of three key features within the software. Firstly, the inclusion of *Range* parameters and *Distance* variables for the individual *soundSources.* Secondly, through the inclusion of a real-time, exponential decay calculation which updated the volume of all the *soundSources* present within the experience based on the *soundSources' Range* parameter and the current value of the *soundSources' Distance* variable. The value of a *soundSource's Distance* variable was determined by the distance between the soundSource and the listener which, in turn, was calculated from the *position* of the *soundSource* and the current *position* of the *listener.*

Through *Horror-Fi Me* we can begin to see the development of a parent application, with the potential for remotely authoring some of these previously devised and deployed AAR experiences. Both the *The McMichael Experiment* (see section 4.3) and *Alien Encounters* (see section 4.4) were custom applications with pre-determined settings for all the sound sources. As *Horror-Fi Me* allows users to tag virtually any object or location within any setting and position a virtual sound source relative to the position of the tag and playback the resultant soundscape, its versatility and usefulness as an authoring tool for such experiences is evident. Although the *Sources* used within the authored AAR environment are currently accessed from a pre-determined list of *Sounds* (from the application's *SoundLibrary* file, which also contains many of the attributes of the *Sounds*) it is entirely feasible that this file could be loaded from a remote location for use within the authoring process and the attributes of each *Sound* could be assigned within the authoring process.

In relation to its potential use within other contexts, we can see from the outdoor situated AAR installation described and evaluated by Lawton el al. (2020) that the approach taken with *Horror-Fi Me* could solve many of the problems that their paper highlights, specifically in relation to its speaker-based approach to AAR, and the need for a headphone-based binaural audio solution, but also due to the increased portability and flexibility of installation that the *Horror-Fi Me* system presents.

Within HCI, though Lawton el al. (2020) recognise the advantages of a headphone-based approach in masking unwanted ambient noise from the user experience, they also recognise the downsides of such an approach when it comes to wanting to include potentially useful *natural* ambient content within the experience. I would suggest that recent advances in *Active Noise*

*Cancellation*[1] (ANC) technologies could overcome this; a feature which is becoming increasingly prevalent in new, consumer-level headphones (Kiss, Mayer and Schwind, 2020). This point is discussed further in section 7.3.3.

It should be noted that any potential deployments are presented with the usual caveats resulting from the current shortcomings of this technology. These include inconsistent spatial rendering in poorly lit environments and an inability to correctly map, or re-localise an existing map, due to a lack of, or an excess of, user movement. Whilst some of these issues can be managed by clear user instruction within the *user interface* (UI) it is also worth noting that the recent inclusion of LiDAR[2] technology within smartphones can overcome some of these limitations, specifically issues pertaining to poor lighting conditions (vGIS, 2021).

Returning briefly to the subject of including instructional information within the UI, the personal *Sound type* (which positions virtual audio content relative to the current position of the user, rather than positioned relative to the location of a physical and stationary object) presents as a possible way of including audible UI instructions. Using this *Sound type*, an audible instructional guide could be introduced who positions themselves next to the listener and provides guidance, instructions or hints and tips for a successful experience.

Finally, a couple of points regarding the optimisation of the system are perhaps worth a mention. Firstly, rather than playing all the sound sources once the soundscape is rendered, regardless of whether they are audible from the listener's current position, sounds could only be played when they are within range. This could significantly reduce the load on user's devices, and improve performance for older phones. Additionally, multiple audio sources of the same type could be identified and forced to play asynchronously, in order to optimise memory and performance, rather than having multiple samples cropped at different points to avoid *phasing*.

## 4.6 Discussion and conclusions

The following section discusses and draws conclusions on the key points reflected upon with regards to the design, development and deployment of all of the installations. These include: realising the experiences, the perception of

---

[1] *Active Noise Cancellation* (ANC) technology uses microphones within the headphones to invert and cancel exterior soundwaves. These microphones can also be used to create a *hear-through*, or *transparent*, listening experience by not inverting the exterior soundwaves and instead amplifying them and delivering them through the headphone's speakers along with other audio content.

[2] LiDAR stands for Light Detection and Ranging and is a method for sensing and determining distances by emitting a laser and measuring the time for the light to return to a receiver.

reality within the experiences, deploying across the different contexts and the intuitive and functional nature of the interactional experience.

### 4.6.1 Realising the experiences

SLAM-based tracking provides a way of realising robust indoor positioning systems and audio augmented reality experiences. Whilst the use of feature point detection provides a useful way of augmenting arbitrary environments with virtual audio content, image tracking remains more reliable with significantly less drift of virtual content over time and distance. This observation, discerned from initial prototyping experiments and the evaluation of these projects during development, echo the conclusions drawn from a study into the reliability of different AR tracking technologies, which found that marker-based AR is significantly more accurate than *markerless* AR (Cheng, Chen and Chen, 2017).

The image recognition and tracking capabilities of current AR authoring tools and SDKs can be used to successfully track unique architectural features, signage and other uniquely identifiable visual elements within an indoor environment in order to provide reference locations from which virtual content can be relatively positioned. This can be achieved by providing photographs of these features and locations as reference images for the AR authoring tools to track, and realises an AR environment with the added reliability of one that utilises image tracking over feature point detection when the direct augmentation of actual recognised and tracked images, or features, is not of concern.

Of course, such a solution relies on the AR experience being authored and deployed within a known physical environment, where the locations of virtual content can be carefully located in relation to these stationary, physical and visually unique trackable features. The reliance on site-specificity for this approach also extends to access to the site to allow for the careful authoring of the virtual environment. With either solution, the short-comings of the AR technology's ability to successfully and accurately localise virtual content, which most often manifests as *drift* (the gradual movement of physically anchored virtual content away from its specified location over time and distance) remain less of a concern for AAR than within visually orientated AR experiences.

Although applied within a different situation, it is easy to identify how the authoring process developed and deployed as part of *Horror-Fi Me* could be applied to curatorial practice. Just as the home user can attach virtual audio content to objects within their home using the process of tagging, the curator could apply the same system to the tagging of artefacts within a gallery or museum environment. Furthermore, the parameters of the virtual sound sources could be used to provide relatively fine curatorial control over the authorship of the virtual soundscape. For example, the *range* and *angle* parameters could be used to generate overlapping and non-overlapping sources and to prioritise specific sound sources over others.

### 4.6.2 The perception of reality

Whilst the AAR experiences and installations outlined within this chapter illustrate an ability to render realistic acoustic virtual realities that rival the object-based AAR experiences discussed in section 2.6, there remains much that can be included that has the potential to further this perception of reality, especially within the creation of site-specific experiences. These additional features are identified and discussed within chapter 8.4. Furthermore, we see how, through the modality of spatial interaction, the problems encountered by Mortensen & Vestergaard (2014) through a reliance on physical interaction with audio augmented objects are overcome. Also, again in relation to the object-based AAR experiences discussed in section 2.6, we also see how the concept *attractor sound* (Zimmermann and Lorenz, 2008) can be developed and deployed.

Whilst headtracking presents itself as a desirable asset, it is by no means a necessity. Although there is some evidence to suggest that it led to a greater perceived reality of the virtual audio content, the ability to localise and perceive virtual audio content as emitting from a physical object via device-based tracking remains evident.

The strength of the audio augmented object's ability to reframe reality and generate context is further illustrated by its capacity to sustain this function even when the usual physical properties of acoustics are altered. This point relates to the use of custom attenuation parameters for the virtual sound sources, attenuation parameters that often reside well beyond what would be experienced within the acoustic reality. These virtual characteristics, with little foundation in reality, appear to have little, if any, effect upon the virtual sound source's ability to directly augment the physical object, prevent it from functioning as previously described, or indeed preventing it from being perceived as *real*. The ability of the experience to achieve this seems dependent on a number of components, and the striking of a balance between *function* and *acoustic virtual reality* in order to create a *functional acoustic virtual reality*. This subject is discussed further in section 7.2.

### 4.6.3 Deploying across contexts

The positive feedback obtained from curatorial staff and museum visitors, regarding *The Mingus Demonstration* (see section 4.2) and *Alien Encounters* (see section 4.4) respectively, illustrate the potential of audio augmented objects within the gallery environment. Similarly, the deployment and remote authoring of audio augmented objects within *Horror-Fi Me* (see section 4.5) illustrate their potential within a domestic setting. We should, however, acknowledge concerns over the deployment of headphone-based AAR experiences within public gallery environments, therefore this issue is discussed in detail in section 7.3.3.

Within the gallery or museum environment we can observe the audio augmented object's potential to engage audiences with both physical objects and virtual audio content, and more specifically digital audio archival content. Additionally, within the domestic space, we can observe the audio augmented object's potential in creating new musical and audio narrative listening experiences.

The deployment of these indoor audio installations has highlighted the importance of the relationship between the virtual audio experience and the physical architectural space. Within sound studies, as Frauke Behrendt (2010, 2012) points out, the interactional modality of *placed sound* is walking, and walking within an indoor environment is determined by the dimensions of interior architectural space. One can therefore conclude that the deployment of an indoor AAR experience needs to carefully consider, and integrate, the physical space within which it will be experienced.

In order to achieve this, the installation designer, or virtual soundscape author, requires sufficient and timely access to the deployment space in order to effectively author such experiences, primarily by means of measuring and surveying the space, but also to ensure that the space is appropriate for the effective use of AR technology. Indeed, the process of authoring the AAR experience can take place within the deployment environment itself. Furthermore, intimate access to the deployment environment for site-specific AAR experiences not only opens up the possibility of utilising a *feature-based* tracking approach, but also allows the precise authoring of *attractor sounds* (Zimmermann and Lorenz, 2008), *associated objects*, *sonic boundaries*, visitor trajectories and narrative experiences through the virtual and physical space. These points are considered further, along with the possibility of authoring a more detailed and precise virtual model, in section 7.5.

Of course, as has been illustrated by the practice documented within this chapter, this is not always possible, and, as we are interested in the provision of authorship to third-party *curators of sound*, alternative solutions are required. This design problem, again, highlights the requirement for a general AAR authoring platform that provides the functional authoring elements of *Horror-Fi Me's* virtual tagging process, but enhances it with the ability to specify a virtual sound sources *range, angle, reverb* and *maxVolume* parameters. Further details on how this could be achieved are also discussed in section 7.5.

Within both the gallery and domestic context we can see important functions for both the individual audio augmented object, and the overall virtual soundscape that they collectively contribute to. Within the museum or gallery this can be identified as an individual audio augmented object's ability to communicate additional information about itself, beyond that which it can achieve by its visual or physical presence. Collectively, such objects within these spaces have the potential to generate an over-arching sonic atmosphere that can help communicate an exhibition's theme and provide context for the individual objects themselves. Within the domestic space, we see evidence of its ability to do the same, though perhaps with different outcomes; collectively they present the potential to reframe the overall reality, whilst individually these objects display the potential to reframe their own physical and visual reality.

The distinction between these two contexts, and the differences in the immediate experiences and characteristics of the audio augmented objects within them, seem to be determined by the nature of the objects themselves, rather than by the type of space within which they are situated. Within *The Mingus Demonstration* (see section 4.2) we are augmenting physical photographic objects, objects that are expected to be *silent.* Within *The McMichael Experiment* (see section 4.3), *Alien Encounters* (see section 4.4) and *Horror-Fi Me* (see section 4.5) we are augmenting *sound-making* or *silenced* objects.

Through this distinction, we can perhaps begin to determine that audio augmented *silent* objects can be attributed with the ability to communicate additional layers of information about themselves and play a part in generating a functional sonic atmosphere that can help to contextualise them, individually and collectively. Whereas audio augmented *silenced* objects can be attributed with the ability to reframe the reality of their context and their individual visual and physical realities. The specific functionality of the components of an audio augmented object reality are discussed further in section 7.2.

### 4.6.4 Intuitive and functional interactions

Largely as a result of the *Alien Encounters* deployment, we can begin to gain insights into the intuitive nature of the listener's experience within the AAR installation, and how the nomadic nature of the experience compliments the applied AR technology. This is made apparent by users' quick familiarisation and interaction, and how their exploratory movements around the space facilitate the mapping and localisation of the SLAM-based tracking.

In the case of *The McMichael Experiment* (see section 4.3) and *Alien Encounters* (see section 4.4) we can see how the inclusion of background audio in the form of radio static provided context for the spatial interaction. This *contextualisation of the mode of interaction* also provided a sonic boundary to the interactive setting and acted to promote a sense of *spatial focus* for the binaurally rendered audio content, which punctuated the non-spatialised sound of the radio static, thus making it easier to determine the location of the audio augmented object and its virtual sound source. Again, the acoustic properties of this contextualising audio content, like the attenuation properties of the spatialised audio sources, has little resemblance to the acoustic properties of reality, though this remains of no consequence to the overall perception of reality of the virtual acoustic model. In essence, in the case of *Alien Encounters* (see section 4.4), we see embodied spatial interaction replacing the tactile machine interaction involved in tuning a radio dial. To this end, this associative interaction appears to convincingly replace the user's previous experience with no observed or recorded concern or consequence.

# 5 Studying The McMichael Experiment

The chapter presents the findings from the deployment and study of *The McMichael Experiment* sound installation (detailed in section 4.3) at the *National Science and Media Museum* in Bradford, UK. It also presents the findings from an interview with the *National Science and Media Museum's* curator of sound technologies, who experienced *The McMichael Experiment* sound installation, and who also experienced the early related prototype experience *The Mingus Demonstration* which is detailed in section 4.2.

## 5.1 Practicalities

*The McMichael Experiment* was deployed as part of the *National Science and Media Museum's Gallery Listening Sessions*; a set of workshops exploring the question of what 'sonic engagement' should mean, and how it should be achieved in the context of museums of science and technology. A small group of interested participants were invited to take a guided tour of the museum's collection stores and take part in a small number of workshops. After the museum tour, attendees were invited to participate in the AAR study. A total of 10 attendees participated in the study, and these participants were reflective of the *Gallery Listening Session* attendees in general; researchers, museum professionals, museum visitors and members of the public, of mixed age and gender.

As detailed in section 4.3, *The McMichael Experiment* used augmented reality technology to attach virtual sound sources to a vintage radio receiver from the museum's collection. The audio content consisted of archival radio broadcast material, contemporaneous with the chosen radio receiver, along with a contemporary recording of radio static. The installation was designed in a way that enabled participants to virtually *tune-in* to the archival radio broadcasts amongst the sound of radio static by positioning themselves at different points around the radio receiver.

Participants were provided with an iPhone, pre-installed with a copy of the AAR application and a pair of stereo headphones and were instructed to explore the radio object and the space around it. No additional details regarding what would happen, how the technology worked or what they could expect was provided.

So that the social interactions between users could be observed and recorded, participants were instructed to experience the installation in pairs. Participants either self-organised themselves into pairs, or the pairings were the result of their availability to participate having completed other unrelated workshop activities. The decision to study the social interactions between participants stemmed from both an understanding of audio technologies as social technologies (Bull, 2000; Sterne, 2003) and from an understanding of museums as social spaces. Within such spaces not only do visitors have to negotiate and interact with other visitors, but they are often visited by groups, sets of friends and families[1].

Ethical approval was obtained prior to the deployment of the installation (see appendix A), and informed consent relating to the gathering and usage of the recorded data was obtained from all workshop attendees prior to their participation (see appendix B).

Both video and audio recordings were captured of the participants just prior to, during and after their engagement with the study. Participants were given the opportunity to provide both verbal and written feedback via printed feedback forms relating to their experiences with the installation subsequent to their participation. Verbal feedback was captured on the video camera and took the form of an open-ended discussion. Written feedback was collected on feedback forms; these were completed anonymously by participants as free text in an effort to encourage the collection of honest thoughts and descriptions from participants relating to their experience that they may have felt less willing to disclose during discussion.

The participants' written feedback was prompted by the question: *How would you describe your experience with the augmented radio?* Verbal feedback was captured on the video camera's microphone, with participants being asked, if they were not initially forthcoming on their own accord, what they thought about the experience they had just undertaken. The bodily interactions between all pairs of participants and the radio installation were recorded on a single, wide-angle video camera that covered the interactional setting of the installation. From this view, participants were recorded entering, interacting with and leaving the setting of the installation.

The video recordings of participants' interactions within the installation setting act as a resource to facilitate and communicate an understanding of what is taking place within the participants' interactional *work* (Crabtree, Rouncefield and Tolmie, 2012). A process of interaction analysis was employed (Jordan and Henderson, 1995) to uncover the recurring themes within the participants' activities. This analytical process was undertaken inductively and reflects the six phases of thematic analysis identified by Braun & Clarke (2006). This process entailed familiarising myself with the video footage, generating labels for interesting activities, gathering the identified labels under themes, reviewing these themes, defining and naming the themes and reporting on them.

---

[1] Considering the museum as a diverse social space within the development and deployment of audio augmented reality technology was reflected upon by curatorial staff during *The Mingus Demonstration* (see section 4.2).

## 5.2 Participant feedback

In the written feedback, all but one of the ten participants described their experience as being either *'interesting'* or *'fascinating'*. Two participants commented on the authentic *'valve warm sound'* and the *'period appropriate programming'*, one commenting that *'It was interesting to have new technology used to interpret a story about an older object'* and that they would like to see this technology used throughout the museum.

Two participants made direct references to how their bodily movements were tuning the radio into the different broadcasts, and likening this to their practical experiences and memories of tuning a traditional radio receiver. There were comments made about being able to listen to individual broadcast material, as well as being able to construct or compose an individual soundscape experience from the different elements available; *'picking up and losing the sounds'.*

Additional positive references were made to the exploratory nature of the experience and its potential for being adapted as a maze, puzzle or mystery solving experience. One participant mentioned that they would have liked additional visual or textual information displayed on the phone's screen to complement and provide information about the audio they were currently listening to. Furthermore, this feature was suggested as an additional means of navigation within the experience, to visually indicate the whereabouts of specific sounds or, if you miss something, provide a means by which it could be easily found again.

In relation to the verbal feedback, participants identified with the experience of using their proximity and their position in relation to the radio to find the broadcast material amongst the sound of static as being a metaphor for what it may have been like, or what it was like, to originally tune this type of analogue radio receiver, as one participant commented:

> *"It reminded me of how difficult and frustrating it used to be to tune a radio, because walking around the object was like tuning it."*

Mentioned again in relation to the evoking of memory was the '*Faithful reproduction of the warm valve sound'* indicating the potential importance of historical accuracy in the sonic delivery of the audio augmented object. Participants also expressed an interest in further levels of sonic engagement with the object, for example one participant mentioned that they almost expected to hear 'more stations when pointing the phone at the tuning dial on the radio'. Two participants made reference to the *'abstract'* nature of the experience and expressed interest in having a more literal and faithful relationship between the object and the delivery of the audio content. One participant commented on how the combination of the real object and the virtual audio triggered their imagination, much like listening to music being a catalyst for the mind's eye, but suggesting that having a physical object in front of them which directly related to the content on their headphones in some way amplified this experience:

> *"It just brings the sound out more, so you're kind of just looking at the object, imagining things, the object's actual sounds but without touching it."*

## 5.3 A sequence of interactional phases

This section outlines eight distinct themes, or phases, of interaction that were defined via the analysis of the video recordings of participants interactions the installation.

We see how, through a process of familiarisation, our participants quickly associate their bodily movements to the receipt of the spatialised audio sources, and then begin to explore the interactional setting to see what they can find. Subsequent to this, we witness our participants returning to investigate the location of some of these sources and engage in listening to them. This phase of focussed listening can sometimes result in a more attentive and engaged listening activity, observable by participants attempting to achieve a very close proximity to the location of the virtual sound source. We see how personal space and acceptable social proximities affect the process of virtual sound exploration and investigation, and how these social constraints become more flexible during phases of engaged listening and manufacture an apparent disassociation with the physical environment. We will now look at each identified interactional phase in a bit more detail.

### 5.3.1 Preparation

It is envisaged that the application will eventually be made available for listeners to download onto their own devices, enabling institutions to economically deploy experiences like this, as such, familiarity and access to an appropriate device would be assumed. Although all participants automatically put on their headphones when they were ready to start, four participants needed to be reminded to put their headphones on the correct way around (essential for the correct orientation of the binaural audio content). Observable from the recorded video of the participants' interactions with the installation, I noted that two out of the ten participants required instructional prompts to engage in an exploration of the space.

### 5.3.2 Familiarisation

This phase of familiarisation is distinguishable within the video recordings of the participants' interactions by the various lateral movements the participants made. This seems to indicate an initial process of familiarisation with the association between bodily movement and the interactive positioning of the surround sound. These movements are often terminated by an acknowledging sign of appreciation, perhaps confirmation that the association has been recognised and understood. These lateral movements were observed being performed in a variety of different ways. Some participants swayed from side-to-side with their device held in alignment with their body and head. One participant waved their device in a lateral motion within a few moments of starting the experience and kept their body stationary whilst doing so. Another participant rotated their upper body in a lateral motion, and therefore also the device they were holding.

Figure 23. Lateral movements of the device during *The McMichael Experiment*.

Figure 24. Lateral movements of the upper body and device during The McMichael Experiment

During this phase of familiarisation, a detachment of the focal gaze from the screen of the device was often observed. In other words, the participant, through their particular process of positional familiarisation, was observing the physical object directly, rather than secondarily through the screen of the device.

### 5.3.3 Exploration

After the brief familiarisation phase described above, all our participants can be observed within the video recordings of their interactions walking around the radio a full 360°, often pausing briefly at the locations of the archive audio broadcast content. The direction of exploration, clockwise or counter-clockwise, most often determined by the first participant to start moving around the object, equally the length of the participants' pauses at the locations of the audio signals were often determined by one participant resuming their exploration around the radio and prompting the other to resume theirs. This behaviour leads to each member of our pair of participants exploring adjacent locations of the sound source, as one member begins to travel to the location of the next broadcast, so does the other member.

This type of exploratory behaviour is observed amongst all our participant pairs, though there are some occasional exceptions. These exceptions appear to take place either when one of the participants has become engaged in the next phase of investigatory interaction, or if the participants appear to have a greater degree of social familiarity with each other. The latter is indicated by an observed acknowledgment of each other, and an observable indication of appreciation for the content they have found and a willingness to share that appreciation with another participant.

Figure 25. Visual acknowledgment and content sharing during *The McMichael Experiment*.

### 5.3.4 Investigation

Within this phase I observed members returning to the locations of the audio broadcasts that they had initially identified during their exploratory phase. We also begin to see exploratory interpretations of the smartphone device as an interface to the audio content. These interpretations take on a variety of styles, with one participant holding their device aloft in an antennae-type fashion, directly reflecting the subject of both the virtual and the physical, another uses their device as a virtual microphone, moving it towards points of interest around the artefact. Others listen through the window of the screen, or rather, observe the radio through the screen of the device whilst listening through their headphones. During this phase of interactional activity, we also observe participants sharing the same audio sources and interacting with the installation in much closer proximity to each other.

### 5.3.5 Focussed listening

The investigation phase, where the participants revisit the virtual audio broadcasts they identified within their exploratory phase, quickly develops into focussed listening. This is discernible within the video recordings of participants remaining stationary for prolonged periods for the first time since beginning their interactions with the installation. Evident within this interactional phase is an apparent disassociation with the physical object itself, with participants being observed closing their eyes or seemingly focussing on other more distant objects whilst they concentrate on the audio content. This behaviour is also documented in the following participant's written feedback, though it is interesting that despite the visual disassociation with the radio object, a strong sonic and physical attachment to it remains:

> *"It was a fascinating experience. The object came alive, I entered a new sonic dimension where I was totally immersed. (I also closed my eyes repeatedly). I was trying to understand the context of sound content, the words of the man speaking."*

Again, despite this visual disassociation with the object whilst engaged in these periods of focussed listening, these events initially take place at either the front or the back of the object, areas of distinct visual interest compared with the two rather plane wooden sides, with the exposed electronic and mechanical insides at the rear, and the TV screen and radio dials at the front. This behaviour is observed despite the location of the two audio broadcasts at the sides of the object.

Figure 26. Focussed listening during *The McMichael Experiment*.

### 5.3.6 Second-level focussed listening

Throughout the recordings of all the pairs of participants, it is possible to witness moments when at least one of the participants engages in listening in much closer proximity to the object, often crouching down in order to obtain a physical position very close to the centre of the virtual sound source. Again, this happens exclusively at the front or to the rear of the object where the object's mechanical and electrical interfaces and inner workings can be seen respectively. This observed activity suggests that the visual component of an audio augmented object can form an important part of a user's interactions with the augmented object as well as the audible component.

Figure 27. Second-level of focussed listening during *The McMichael Experiment*.

### 5.3.7 Interruption and finishing

Interruption to a participant's interactions with the installation, which often resulted in the termination of their participation, stemmed from one of the pair of participants deciding they had finished. Evident throughout all the recorded interactions, in all but one of the 5 pairs of participants, the end of participation is initiated by one participant removing their headphones, which prompts the other to do the same, even though the participants never started at exactly the same time. In the one event in which this did not happen, the other participant was engaged in second- level focussed listening.

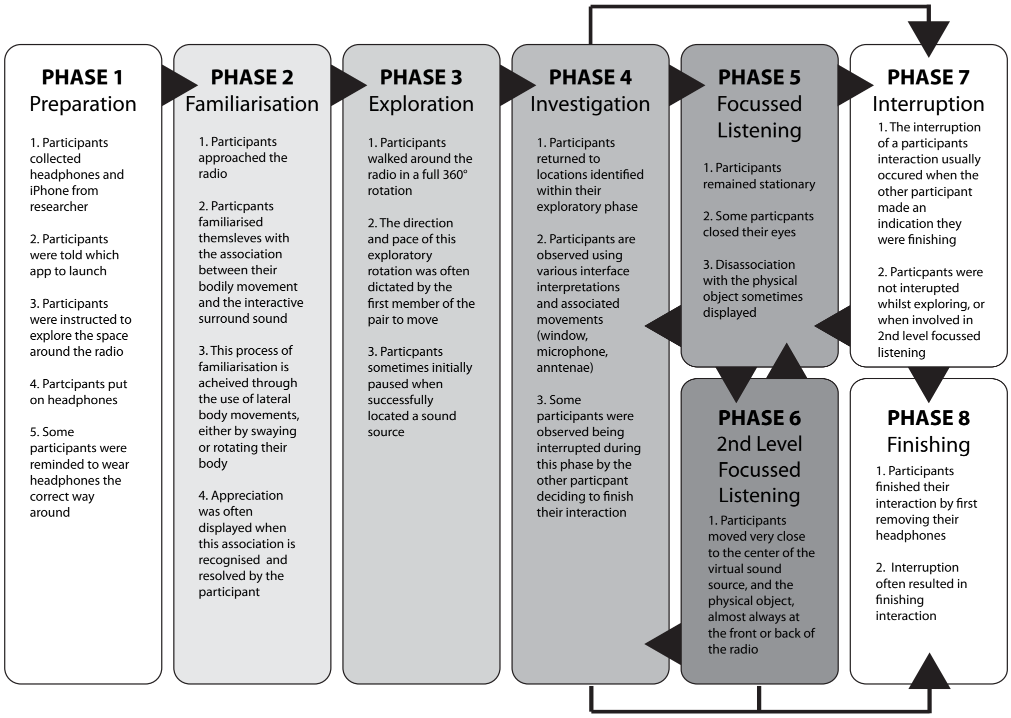


Figure 28: An identified sequence of interactional phases.

Figure 28 shows the eight distinct phases of interaction and their sequence defined as: *preparation, familiarisation, exploration, investigation, focussed listening, second-level focussed listening, interruption* and *finishing*.

## 5.4 A curatorial perspective

Subsequent to the presentation of *The Mingus Demonstration* (see section 4.2) and deployment of *The McMichael Experiment* (see section 4.3) at the *National Science and Media Museum* an interview was undertaken with the museum's *Curator of Sound Technologies*, who participated in both of these experiences. The interview was conducted with a view to gaining insights into the suitability of this technology for deployment within a museum and gallery, its potential and the implications involved. It was also envisaged that the interview would reveal the curatorial perspective on this technology.

The audio recording of this structured interview was transcribed and thematically analysed using an inductive approach, undertaken using the six phases of thematic analysis identified by Braun & Clarke (2006). Several themes were defined and are presented within the following sections as: *Transparency and intuitiveness*, *The shared and social experience*, *The curatorial potential of the audio augmented reality experience*, *Audiences for sound and sound for audiences* and *Objects and experience*.

After initially refreshing the interviewee's memory of the two AAR experiences they participated in, these themes were identified through the curator's responses to the following set of interview questions:

1. *What are your general thoughts regarding the two prototype AAR experiences?*
2. *Can you outline what you think the specific strengths and potentials are?*
3. *Can you outline what you think the specific weaknesses or challenges are?*
4. *What are your main concerns about this technology?*
5. *Do you think the audio content encouraged engagement in the museum object? If so, how do you think it achieved this?*
6. *Do you think the presence of the museum object encouraged engagement in the audio content? If so, how do you think it achieved this?*
7. *Do you see opportunities for ARR as a way of augmenting existing collections with audio?*
8. *Do you see opportunities for AAR as a way of exhibiting sound as a central feature of an exhibition? (Not just as a way of augmenting existing collections of physical objects)*
9. *Do you recognize any potential in the generation or composition of more experimental and ambient soundscapes, as well as the direct audio augmentation of individual objects? (For example: the overlaying of multiple and diverse audio sources) If so, how do you see this working?*
10. *What are your thoughts on audience interfaces relating to this technology? Thus far, both demonstrations have used handheld smartphones. Do you have any comments relating to the deployment and interactional and social issues regarding this approach?*

Within the following sections each of the defined themes resulting from the thematic analysis of the interview with the curator are presented in detail.

### 5.4.1 Transparency and intuitiveness

In addition to commenting on the installation's *'intuitive'* nature, the curator reflects on the some of the overall attributes of *The Mingus Prototype*, attributing its accessibility to an appropriate object/audio relationship which is both intuitive and unusual:

> Curator: *"…because it combined recognizing what's happening... so I'm looking at a photo of a New York Street and I'm hearing the sound of a New York Street… so, at one time it's easy to grasp what's going on with it… but at the same time it's potentially quite surprising because you weren't expecting that to happen perhaps."*

In the next quotation, the curator goes a little further and suggests that it is not only the quick association that can be made between object and audio that make it intuitive, but it is this which quickly promotes further exploration of physical space:

> Curator: *"It's quite intuitive so you can kind of work out what's happening once you've looked at the different objects… that's the music and that's the thing then you can explore that space through the sound and I think that's quite an easy thing for people to realise what's happening then play with it."*

The curator suggests that the immediate and instant association of *The Mingus Prototype's* image and audio association, compared to the more abstracted tuning experience of *The McMichael Experiment*, make it more likely to initially engage visitors:

> Curator: "*It's [The Mingus Prototype] almost like it's less prescriptive… it just kind of happens to you, if you just kind of walk the space, you experience the soundscape"*

Specifically in response to *The McMichael Experiment, the curator reflects:*

> Curator: *"I can imagine instances when you might probably get bored before you've worked out what's happening because, for our galleries, we need a pretty much instant hit – do a thing, then this happens."*

This is not to say that the approach demonstrated by *The McMichael Experiment* was not without potential, and the curator suggests that, if an experience similar to this were to be pursued, *"instructions or guidance would be needed".*

The curator suggests that the agency that both experiences afford to visitors in relation to them having control over the composition of their own soundscape, as demonstrated perhaps to the largest degree by the ability to mix multiple sound sources together, is a desirable feature for the museum:

> Curator: *"it's [being able to compose as a visitor] interesting from a curatorial point-of-view because its… we're always interested in ways visitors can create their own experience out of what we present to them, and it's not just a black box – here's the thing, here's another thing, you do it in the order we say. So anything that allows visitors to play around with that… and find how it works for them is really good."*

Whilst the above excerpt reflects upon this functional element from a curatorial perspective, the following excerpt considers this attribute from the point-of-view of the visiting public, and suggests that its success, particularly within the context of a museum, is reliant upon the type of audio content, as well as a clear and obvious relationship between the object and the audio:

> Curator: *"I can imagine that working really well if the content's right, and that's probably the really difficult thing of working out what the suitable content would be. I think that idea of hearing things from different objects and mixing them up could work really well. But it's got to be the right sort of content… I think for our visitors it's got to be quite clear and obvious, and on-gallery it's almost like everything has to be really exaggerated. Because our audience is coming in to just have a bit of fun with the kids, so they're not critical listeners, they're not trained, they're not like an art gallery audience. If you're doing some sound-based work on display in an art gallery, you almost have like a trained audience to some extent…"*

This latter point of accessibility is further reflected upon regarding visitors having their own compatible device, and a will to utilise it:

> Curator: *"not everyone will have a suitable device, and even if they do, they might not want to do that."*

Therefore, the curator maintains that an exhibition has to work for people who can't, or don't wish to participate – for them the headphone and device-based experience needs to be an add-on, an additional to, and not a conditional feature in experiencing an exhibition.

The previous observation is reflected upon further, and the lending out of compatible devices to the visiting public presents itself as a possible solution to this problem. Though, again, the curator suggests that this solution poses its own issues, such as the security of those devices, their robustness and durability. The curator suggests that 'listening stations', a stationary point at which a tethered pair of headphones is available for visitors to use, may be a viable option.

Furthermore, this 'listening station' approach also provides an opportunity to physically contextualise the device hosting the application for this type of experience, with a view to perhaps engaging younger audience members:

> Curator: *"A lot of our visitors are families with children, and how well does this thing work with for 7-year-old child… and sometimes just delivering it in a more fun way so, if it's like a ray-gun-scanner thing and the kids can point it at stuff… that could help make it a bit more appealing to children as well."*

### 5.4.2 The shared and social experience

The problems of using a headphone-based audio experience within family and social groups are also considered further in terms of the social and shared experience of visiting a gallery, specifically in relation to the difficulties associated with talking about audio-based content retrospectively:

Curator: *"It makes it hard for them to talk about what they are experiencing. You want people to talk about what they are experiencing, and that's quite hard to do for a lot of people with sound after the event, because it's like: 'that thing that it did near the start where it did that thing and…' it's really hard for most people to describe that and share it. Whereas if you're hearing it all together and it's like 'what's that?' and you're all instantly switched on and everyone knows what the other people are talking about."*

Also mentioned, in relation to how the isolating effects attributed to the use of headphones could be overcome, is how the experience could be *"transferred into a more shared experience"* potentially by utilising a *"carefully designed distribution of speakers…".*

The above statement hints at this later confirmed view that, whilst the experience poses much promise, it is how it is delivered within the gallery environment that remains unresolved, though the intuitive and transparent nature of the experience we saw described earlier is referred to again and attributed with much potential:

Curator: *"The idea that it's less interactive, it's just there doing its thing, might be a better solution for us.. there's huge promise in this for a museum like this, it's the practical consideration and budgetary too".*

Additionally, the proposed use of a speaker-based system is again later reflected upon in terms of its reliability on time-based, proximity or manual triggering, each with their own considerations and implications within a gallery and exhibition environment:

Curator: *"you have to ask if that's going to work for all of our visitors, or are most people going to miss most of it because there are not going to be in the right place at the right time, and triggering sound and all that is great, but if you've got a really busy gallery, with load of kids running around… what happens to your carefully designed triggering as people approach this or that."*

Somewhat of a distraction, though perhaps worthy of consideration in terms of the practical application of this type of experience within a cultural institution, the additional 'budgetary' remark made in the previous excerpt refers to the curators feeling this experience provides a cost-effective and economical deployment option. This is determined by comparing this visitor-hosted, mobile application approach to an institution-hosted and deployed audio experience and the additional hardware and infrastructure expenses incurred with the latter. Specifically related to this matter is the idea that existing objects, collections and gallery spaces can be re-framed and re-contextualised virtually, without costly refurbishments to physical space, this is also seen as a practical and economical solution to keeping things fresh for repeat visitors:

> Curator: *"…it could be really effective, particularly for permanent galleries because, and for us we have a lot of repeat visitors, so if you can do things that do change the way people are presented with things… it's also a relatively cheap way to change things out, compared to changing interpretation panels and having new video content made…"*

### 5.4.3 The curatorial potential of the audio augmented reality experience

In the following interview excerpts the curator gives thought to what they perceive would be some of the desirable functions of audio augmented reality experiences. In the first case this relates to the construction of an atmosphere that can help promote thought about the subject matter of an exhibition:

> Curator: *"if it can help to evoke an atmosphere and a kind of feeling that makes the visitor think more about what they're looking at and listening to here – I think that could be really useful."*

Secondly, the curator provides the following description of a potential functional element that very closely resembles that of the *attractor sound,* as referred to by Zimmerman & Lorenz (2008), adding that such a function should perhaps have an element of surprise or be of an unanticipated nature:

> Curator: *"I can imagine that if this could be developed in some way so that it can actually help to draw attention to the object or display… if you walk past something and it makes a really interesting sound that you maybe weren't expecting then that would be a good way to grab peoples' attention and then deliver more, if that's through the sound or in other ways once you've got them there."*

Again, the use of surprising or *unexpected* sound is mentioned in relation to the function AAR experiences can perform within museums and galleries in terms of providing access to, and exhibiting, audio archival content in new and more engaging ways:

> Curator: *"I think that [audio archive content] is really important for us… it's important that we connect with that content because that's the only way, on the whole, our visitors will get what these objects are about. Their purpose is to show you a film or play you music… but again that can become very clichéd – press the button and hear the Kings speech through this microphone or whatever. So if there are ways of actually making that a bit more sophisticated and engaging, and bringing in interesting and maybe unexpected, unknown content…"*

The curator also sees potential in the system being able to relate or connect different objects together, perhaps by historical period, type, or another curatorial association of interest, perhaps using different focal levels of content based on the listener's proximity. In the following excerpt they provide a nice musical example of how this might be achieved:

> Curator: *"…it ties into an idea of a very dynamic exhibition, where it's not just things stood in isolation and move from one to another but the exhibition itself is kind of moving and dynamic and pulling you in different ways and you experience something different just by walking across the room… and I can see that working particularly well as a way to kind of connect things… if it was musical instruments, the ambient sound is 'the band' with all those instruments playing, but then as you get closer to each instrument that changes and it starts to pull out of the ensemble… then as you get really close to it, maybe you hear the player of that instrument telling you something about that instrument…"*

Within the following two interview excerpts, *The Mingus Prototype's* and *The McMichael Experiment's* potential are recognised in terms of creating an effective and innovative way of providing an additional layer of interpretative content. Additionally, this is seen as a positive improvement upon existing solutions to exhibiting sound and the sounds of objects:

> Curator: *"You could build a soundscape around a set of objects that… in ways that don't kind of burden them, there are not expected to read a big text panel… it's just there."*

> Curator: *"I think it gets away from that push a button to get the sound of 'this' – 'this' thing makes 'this' sound. But builds it into a sort of environment… if you have a themed gallery and having a whole continuous soundscape that blends together to be an ambient soundscape that blends together and that's not too noisy and offensive, but then spotlights parts of the soundscape as you get nearer to the relevant object."*

Further reflecting upon the concept of multiple sound objects forming a blended and coherent soundscape, the curator references upon their experience of *The Future Starts Here*[1] exhibition, specifically its ability to provide discernable content around individual objects, as well as an over-arching soundscape:

> Curator: *"…*they [the exhibition's audio content] *all worked individually as you approached the exhibit and you heard whatever you were supposed to hear about that particular part of it, but the whole thing in*

[1] *The Future Starts Here* was an exhibition at London's V&A museum in 2018 that bought together new, emerging and potential future technologies as an enquiry around possible future society: https://www.vam.ac.uk/exhibitions/the-future-starts-here

> *the gallery as a whole blended into this almost stereotypical hubbub of modernity... and it was really, really effective."*

The curator then compares the approach taken to the curation of the audio content in *The Future Starts Here* to how individual sound sources are often exhibited and managed, suggesting that an *inclusive approach* that embraces multiple, bleeding audio sources has the potential to provide a legitimate alternative to the dominant *isolationist approach*:

> Curator: *"…instead of insisting on isolating everything, we actually managed to blend it together in a way that overcomes the problems of too much noise on gallery and things bleeding and over-lapping too much, but kind of works, so you get something out of it wherever you are... I think it brought a lot to that exhibition* [*The Future Starts Here*] *and made it a lot more interesting than walking through exhibitions that are sort of fundamentally quiet, except you hear the burst of sound when someone pushes the button or something, or the repeating video… "*

### 5.4.4 Audiences for sound and sound for audiences

The curator also suggests that this technology presents itself as an opportunity to deliver an additional layer of content for those who wish to engage with further interpretive content, and that this presents itself as a particularly attractive feature for an institution that is largely geared towards family audiences:

> Curator: *"Although our audience is largely families, and most of our exhibitions are focused on that audience, we do have quite a wide range of people… and we're always trying to think of ways we can sort of layer interpretation and give options for people who want something else who don't just want the seven to eleven family experience, but want to know more about the objects, want a more in-depth interpretation… so this might be one way of delivering that…"*

Additionally, the curator also considers that the audience who would most likely engage with, and have the access to the devices that would be capable of running, an experience such as this are also those who would most likely want to engage with an additional layer of interpretive content:

> Curator: *"…that* [additional interpretative content] *would have the advantage that that audience probably are more likely to engage with using their own devices for things."*

It is worth noting at this point that if this additional interpretative content were to be personally selectable by the audiences, for example different types of content for children and adults, then this would seem to be something that would be difficult to realise with a communal, distributed speaker-based audio content delivery system. Rather, such a feature would seem only to be possible with a personalised audio content delivery interface, such as headphones.

Staying on the subject of audiences for sound, the curator suggests that most audiences have little no expectation about sound in museums, though if you are explicitly exhibiting sound, or indeed curating an exhibition about sound, then an expectation and a readiness to engage with it is generated:

> Curator: *"If it's an exhibition about sound you sort of have the advantage of that's what people are expecting, to be honest a lot of people don't… have no expectations about sound in museums… but if you have an exhibition that is about sound then maybe you've got over that first hurdle, people are expecting to hear things, they're expecting that there will be things to listen to."*

Interestingly, in the following excerpt the curator suggests an *object-first* approach to exhibiting sound, proposing that exploration or access to the sound content within the virtual soundscape, could be initiated by the physical object. This appears in contrast to a *sound-first* approach (using sound to initiate engagement with the object), considered from the position of augmenting physical objects with virtual sound, rather than augmenting a virtual soundscape with physical objects. Additionally, they indicate that the chosen approach (whether *object-first* or *sound-first*) would be a matter of curatorial priority:

> Curator: *"…maybe within that you kind of build your virtual soundscape with the things [sounds] you want to display in that and use visual triggers of objects to pull people around that soundscape, rather than: 'Ooh, what's that sound? Look at that thing.' it's 'That's a weird thing let's go and look at that.' …then you hear the sound, which, curatorially, is what you want them to attend to."*

### 5.4.5 Objects and experience

Furthermore, the curator reflects upon a subject of specific interest to the scope of this research; the relationship between the object and the visitor and proposes that the visitor is more concerned with the experience rather than the physical object itself:

> *"In a way it's not about the object, the objects are what we have… from the visitor's point-of-view, it's really about the experience."*

Figure 29: A *Gaumont-Kalee* GK21 35mm cinema projector. From the collection of The Science Museum Group on display in the café of the PictureVille Cinema at The National Science and Media Museum, Bradford, UK.

In the café of the *PictureVille Cinema* at the *National Science and Media Museum*, where this interview with the museum's curator of sound technologies took place, there stands a large and imposing mid-twentieth century cinema projector (see Figure 29). Considering and referring to this object in light of our discussion, the curator continues to reflect upon the object and experience relationship, indicating that it is not only about the personal experience of the visitor with the object, but about the experiences of the object itself, and communicating those to the visitor. Additionally, they propose that the operational and mechanical sounds of objects within a soundscape, for example, could play a part in achieving this:

> Curator: *"I mean that projector there… what matters is the experience of everyone who was involved in that, the projectionist that used it, the people who built it, the audiences in the cinema that watched the films on it… it's about trying to get at that… sometimes it could be relatively simple, but really effective if it's done well… projectors make sound,*

*they're not silent, some of them make quite a lot of noise… and you can imagine some kind of soundscape that uses that."*

## 5.5 Discussion and conclusions

The following section discusses and draws conclusions on the key points resulting from the study of *The McMichael Experiment* and the curatorial interview presented in the previous section. These include: the sequence of interactional phases, the observed handheld interactions of participants, the potentials and implications of audio augmented museum objects, the realisation of an intuitive and transparent interface, and the importance and function of exploratory freedom. Also discussed are the problems associated with the use of headphones in public and social spaces, and the identification of different types of audio content and its function within the audio augmented object reality.

### 5.5.1 A sequence for engagement

Within the study of *The McMichael Experiment* sound installation, we see phases of interaction that resonate with the findings from some of the previously mentioned related works in this area. This includes the use of virtually attached sounds as advertisers that draw users towards the audio augmented object for closer investigation. This is identified, though not specifically exploited, by Zimmerman and Lorenz (2008) and could be said to be evident within our participants' trajectories from exploration through to investigation and focussed listening. Furthermore, we also see evidence of a second level of focussed listening within the work of Montan (2002) where differently treated zones of reverb are triggered upon a user's close proximity to the audio augmented object, generating a soundscape within a soundscape and the feeling amongst participants of entering into a different space from outside. Based on these commonalities, we can perhaps begin to generalise more widely across various applications for audio augmented objects and begin to provide some foundations of a theoretical model for attraction and engagement with them. This subject is discussed further in section 7.2.1.

The curator also suggests that it can be the surprising nature of the realisation of this association between object and audio that can grab a visitor's attention and initiate engagement, again echoing Zimmerman and Lorenz's (2008) concept of the *attractor sound*, and providing an insight into how this functionality could be best authored. The effective use of unfamiliar and unusual audio content as a means of engaging visitors within a museum is also reflected upon by cultural theorist Bijsterveld (2015). A discussion on how these sounds can be used to engage users with audio augmented objects in practice can be found in section 7.2.1.

### 5.5.2 Handheld interactions

Additionally, we see within this study the successful association between hand and ear, as documented by Helller and Borchers in their *AudioTorch* project (Heller and Borchers, 2014). Logic would dictate that a headtracking-based approach to delivering an AAR experience with binaural audio would result in a more intuitively realistic audio augmented reality, given binaural audio's reliance on HRTF models for the creation of three-dimensional acoustic virtual realities (Iida, 2019). Though we can determine that a handheld-based tracking approach for both translational and rotational movement is capable of delivering an experience within which the spatial perception and localisation of virtually spatialised sound sources can be successfully achieved by listeners. Furthermore, the initial *familiarisation* phase, observable via repeated lateral movements and their association with the effect on the delivery of the spatialised audio content, is also observed within the *AudioTorc*h project.

The observed actions of participants pointing the device's camera at points of visual interest around the object would lend itself well to scanning or magnification orientated experiences where, for example, sub or super-sonic audio content could be rendered audible upon closer inspection with the handheld device. The rendering audible of ordinarily inaudible subjects such as insect or other biological sounds would provide an enticing subject matter for the further exploration of this approach within exhibitive contexts, and the interactional approaches described by Montan (2002) and observed by Kelly in the work of sound artist Vicky Browne (Kelly, 2019) lend additional validity to such an interactional approach. These examples and approaches also fit well with the focussed listening activity identified within this study, and with the inclination expressed by some participants to have more audio detail presented to them upon closer inspection of physical details of the object. This is discussed further in section 7.2.1.

### 5.5.3 Audio augmented museum objects

*The McMichael Experiment*, along with the curator's suggestion of the validity of both an *object-first* and an *audio-first* approach to the authorship and curation of such experiences also demonstrated how the presence of the physical object can be used to engage users with virtual audio content. This is perhaps most evident in the participants' exploration of the more visually interesting areas of the radio object and their ensuing engagement with the virtual audio content available at those locations. In short, we see how the object's material characteristics can help promote engagement with the immaterial virtual audio content and vice versa, the latter demonstrated by the audio's ability to draw in users and entice them to consider the physical object from specified locations. In summary, the audio augmented object has the ability to engage listeners with its physical and visual presence via sound, and engage listeners with sound via its physical and visual presence. These *object-first* and an *audio-first* approaches are discussed further in section 7.2.1.

Understanding a little more about the dual nature of the interactional experience and character of the audio augmented object leads to some interesting considerations regarding their specific and potential function. It is worth reconsidering the use of digital audio archival material within this installation in relation to this last point and, as such, one should consider the potential of audio augmentation as a means of creating interfaces for engagement with both digital audio archives and museum objects. Additionally, what also appears to be of primary significance here is that, by augmenting a historical museum artefact with contemporaneous audio archival content, a mutual and functional contextualisation of both these components occurs; the artefact functions as an interface for the audio and provides it with physical presence and historical context, and the audio functions as an interface for the object and provides it with additional physical presence and historical context. Furthermore, this factor presents itself as a way in which technically and mechanically silenced museum objects can once again become *experiential objects.* Furthermore, this study shows that participants engaged with the provided audio archival content, with some expressing the desire to be able to quickly and efficiently revisit content they had previously encountered in order to engage with it further. The subject of using audio augmented objects as interfaces to audio archival content is explored further in section 7.3.5.

We can also conclude that the deployment of an audio augmented reality experience of this type within a gallery or museum environment presents itself as an effective way of adding an additional interpretive layer of content and a practical and economical way of re-framing or re-interpreting existing collections of objects. An in-depth discussion of the implications of the audio augmented object reality on galleries and museums is included in section 7.3.

### 5.5.4 A transparent and intuitive interface

The curator's comments about the experience being *'less interactive'* and *'just there doing its thing'* could be considered as an interpretation of an intuitive and transparent interface. The curator suggests that this is a result of a strong object and audio association, an association which is a direct consequence of augmenting a real-world object with dynamic binaural audio. We can therefore conclude that the creation of an audio augmented reality experience comprised of, or containing, audio augmented objects contribute to the realisation of a highly intuitive user interface. Furthermore, this assertion is supported by participants' swift ability to familiarise, explore, investigate and engage with an audio augmented object without prior instruction. It should also be noted that the transparency of the experience is anchored in the naturalistic and, what the user perceives to be, the near instantaneous feedback of their embodied interactions with the dynamic binaural audio content. In summary, an audio augmented reality experience containing audio augmented objects utilises the natural affordances of our experience with sound and sound making objects, and therefore has the ability to realise a highly intuitive interface. The realisation of intuitive and naturalistic interfaces within the audio augmented reality is discussed further in section 7.3.3.

It should be noted that the observed intuitive and naturalistic characteristics of the AAR interface are evident once the nomadic and explorative nature of the experience is understood and excepted by the participants. With nothing to do other than to spatially explore, there can be uncertainty, in the first instance, around what participants should do. This specific issue is explored further in section 7.2.1.

### 5.5.5 Exploratory freedom

The freedom to explore the three-dimensional reality of the interactional setting also seems to be an integral component in the creation of the naturalistic experience. Albeit hindered by the confines of gravity and indoor architectural space, we are naturally free to explore the three-dimensions of an interior acoustic reality and the resultant changing soundscape, and these AAR experiences reflected this freedom.

The nomadic nature of the experience also enables users to have a degree of compositional control and ownership over their audio experience. From a curatorial perspective, not only is this described as being desirable, but it is seen as a factor in the creation of a more *meaningful* visitor experience, which is explicitly stated by the curator as being a fundamental goal of the museum experience. This is also supported by the participant feedback within the first installation study, where the physical act of tuning into the radio broadcast audio evoked personal memories and imagined experiences. In short, it seems that the freedom of exploration facilitates a freedom to compose, to generate a personalised audio experience which, in turn, has the potential to create a more meaningful experience for the listener. This topic is discussed further in section 7.2.4.

An uninhibited freedom to explore the dynamic binaural soundscape also creates a functional fluidity which the curator recognises as being an attractive contribution to the creation of a '*dynamic exhibition'.* This is suggested in terms of individual sounds within an over-arching exploratory soundscape having the ability to '*move'* and '*pull'* visitors in different directions around an exhibition environment. Whilst this suggestion again confirms the concept of the *'attractor sound'* (Zimmermann and Lorenz, 2008) it is also suggestive of a soundscape comprised of multiple audio augmented objects where each object acts as a component sound source beckoning for attention within an over-arching exploratory *sonic atmosphere*.

Whilst we can begin to determine some of the curatorial functions of a nomadic audio augmented reality experience and the audio augmented objects it contains, we can also find function within the resulting *sonic atmosphere* created by such an experience. The example exhibition soundscape described by the curator in section 5.4.3 indicates that the collective audio output of multiple audio augmented objects could be constructed in such a way to promote thought and engagement with an exhibition's context, and therefore function as a tool to help facilitate the exchange of knowledge within such an environment. Again, we see this playing out within *The McMichael Experiment* with multiple sound sources (the radio broadcast content and the sound of radio static) combining to create a sonic atmosphere reminiscent for participants of a historical period and their interactions with its entertainment technology. This topic is discussed further in section 7.3.2.

### 5.5.6 The headphone conundrum

Whilst there are reservations about a headphone-based rather than a speaker-based delivery system, for the most part relating to the shared and social experience, we should be aware that a headphone-based approach solves the problem of trying to isolate disparate audio sources within the gallery. It also enables both the careful authorship of an overarching sonic atmosphere and the careful isolation of audio sources as desired. Whilst you can embrace the *'hubub'* with a speaker-based system, control over the isolation of sources presents as more problematic, and the personalisation of delivery either highly-restrictive (turn-based) or impossible to realise, which is something that the curator identifies as being very desirable, not just in terms of delivering one audio experience that can be personally and individually consumed, but offering multiple layers of audio content that can both be personally selected and personally delivered. This (the headphone-based approach and the nomadic experience) also has the added benefit of enabling the listener to have compositional control and ownership over their own audio experience, which in turn facilitates the benefits previously outlined.

We can also observe a desire to share with other participants audio content that has been found and enjoyed and, more generally, how initial social constraints become less ridged during the investigative and listening phases. The implications of the headphone-based approach, its impact on the shared and social experience and how they may be overcome is the subject of discussion in section 7.3.3.

### 5.5.7 Audio content

From recordings of the types of media they played, to recordings of their electrical and mechanical workings, operation and production it seems that a wide variety of sounds and noises can be considered. Whilst this is great news for sound artists, curators and audio archivists alike, these different sounds also have different functions, from attracting a listener's attention (with more unusual content) to prompting personal memories and helping to communicate a myriad of historical and social knowledge and information, both individually and as a part of a *functional sonic atmosphere.* Furthermore, it appears that the perceived *historical authenticity* of the audio content and the perceived *sonic authenticity* of its audio production are also of importance, not just for the realisation of an authentic or realistic virtuality, in that it can be perceived as being like *'the real thing'*, but also in terms of provoking a listener's imagination and memory. We should also acknowledge the successful role the radio static component played in terms of its function as a contextualised transitional element between the radio broadcast content, but also in terms of its contribution to the evocation of listeners' memories and imaginations. Both these factors illustrate the important function the *contextualisation of the mode of interaction* can play within the audio augmented reality experience, specifically if we make the connection between *embodied interaction* and *embodied cognition*, in that direct and embodied interaction with a subject can help facilitate an understanding and meaningful engagement with it (Truax, 2012; Kirsh, 2013).

In summary, the identified sequence of interactions provides a foundation for the design of audio augmented reality experiences containing audio augmented objects. Within this we see that the combination of dynamic binaural audio and a freely explorative, nomadic experience can realise a highly transparent, naturalistic and intuitive interface. This, in turn, enables an interactive audio experience that has the ability to attract and engage audiences with both physical objects and virtual audio content. We can also see a connection between the freedom of exploration, agency over creation, or composition of the sound experience, and the generation of meaningful experiences that can evoke an audiences' memory and imagination. We also learn how collections of audio augmented objects can create over-arching sonic atmospheres for framing exhibitions, re-framing existing collections of objects and creating additional layers of interpretive content that can facilitate the creation and communication of experience. Whilst we begin to learn about the function of different types of audio content within these experiences, we also gain important insights into the potential issues regarding the deployment of these types of experiences and the requirements of specific audiences and institutions which can better inform their design.

# 6 Studying Horror-Fi Me

This chapter presents the findings from the deployment and study of *Horror-Fi Me*, a remotely deployed iPhone application which enabled users to author their own dynamic binaural audio augmented reality experience by attaching virtual audio content to objects and features within their own homes (see section 4.5 for details). The study is followed by the findings from an interview with the composer of *Horror-Fi Me's* specifically commissioned soundtrack.

## 6.1 Practicalities

*Horror-Fi Me* was launched and presented as part of the *Live Cinema III: Festival of Research and Innovation 2020*[1]*.* A download link for the application was circulated as part of the festival's online marketing material, and attendees were encouraged to download and try out the experience prior to its presentation at an online version of the festival.

The analytical data provided within the mobile application's *TestFlight*[2] account logged a total of 102 installations of the *Horror-Fi Me* mobile application and a total of 97 sessions. The *TestFlight* application defines an *Installation* as the number of times the app has been installed on a compatible device, and it defines a *Session* as: *'the number of times the app has been used for at least two seconds'* (Apple, 2020). Both these totals are based on users who have agreed to share this data with myself, the developer.

Subsequent to the launch of *Horror-Fi Me, TestFlight* reported that a total of 29 copies of the application had been downloaded and installed on to a compatible device with at least one session logged against it. Of these 29 installations 17 had more than one session logged, with 10 installations logging 5 sessions or more. All 29 participants associated with these installations were approached for an interview, with 10 participants responding and agreeing to this request. The 10 participants interviewed were aged between 27-65 years old and of mixed gender and all had successfully downloaded and installed the app, authored an AAR experience and undertaken that AAR experience within their own homes with their own iPhones. These 10 participants used a variety of stereo headphone types to

---

[1] *Live Cinema III: Festival of Research and Innovation* presents cutting edge academic research, master classes, workshops and a programme of screenings. The 2020 event took place online: http://livecinema.org.uk/live-cinema-iii-festival-of-research-and-innovation-2020/

[2] *TestFlight* is an online service operated by *Apple* for testing iOS mobile applications. It allows for the remote distribution of development versions of new iOS application to beta testers, and for the distribution of a publicly available download and installation link for the app without the need for commercial distribution via *AppStore:* https://developer.apple.com/testflight/

listen to the resulting audio content, including on-ear, over-ear, ear-buds and *AirPod*[1] headphones.

Ethical approval was obtained prior to the launch of the application (see appendix C), and informed consent relating to the gathering and usage of the recorded data was obtained from all application users who participated in the study.

The interviews with these participants took a structured form in an attempt to specifically address and gather additional insights into the areas of interest identified within the conclusions of the previous study detailed in section 5.5. This included the study of multiple audio augmented objects and the virtual soundscape produced as a result of having multiple and simultaneously audible sound sources across a larger interactive setting that could facilitate the locating of objects that could be heard, though are yet to be seen. Additionally, through the use of a variety of audio augmented objects, audio sources and the creation of an explicit context, it was envisaged that further insights could be gained into the characteristics, experience and possible functions of audio augmented objects within the audio augmented reality experience.

The specific questions put to the participants of the study of *Horror-Fi Me* were as follows:

1. *How did you find the experience?*
2. *Did you undertake the experience more than once? If so, why and how many times did you undertake the experience?*
3. *What, if anything, did you think worked especially well?*
4. *What, if anything, do you think didn't work well?*
5. *Did the sounds sound like they were coming from the locations and objects you tagged? If so, which sounds in particular?*
6. *Did you tag a clock? If so, what type of clock did you tag, and did the clock sound fit with your clock?*
7. *How would you normally describe the environment in which you undertook the experience?*
8. *Did the experience make your environment feel different?*
9. *Do you think the experience encouraged you to move around and explore your environment? If so, how do you think it achieved this?*
10. *Were you encouraged to explore the locations of any sounds? If so, any sounds in particular?*

---

[1] *Airpods* are earbud-type headphones manufactured by *Apple*. The Pro model of *Airpod* contain accelerometer, gyroscope and magnetometer sensors within the ear piece as well as externally positioned microphones. This feature means that they are capable of tracking the orientation of a user's head as well as the elimination of external sound via *Active Noise Cancelation* technology (ANC) or the inclusion of external sounds via a *transparent* listening mode.

11. *When you were moving around, did you notice if you were guided by the locations of sounds, or by the locations of the objects you tagged?*
12. *How did you find the process of tagging objects and locations?*
13. *When you were playing the experience, did you find the presence of your phone or the phone's screen distracting in any way?*
14. *If this type of experience was available within a gallery or museum space do you think you would engage with it?*
15. *Is there anything else you'd like to add relating to your experience with Horror-Fi Me?*

In lieu of video recordings of participants' interactions with the experiences they authored, interviews were conducted with participants after they had completed their experiences. The reasons and justification for this change in approach are detailed in section 3.3.

Here, it is the quotes from participants' relating to their interactions within the installation setting that act as a resource to facilitate and communicate an understanding of what is taking place within the participants' interactional *work* (Crabtree, Rouncefield and Tolmie, 2012). These are included with an understanding that by getting participants to tell us what they did, we can gain information about what they actually did. Again, this point, and its relevance to the ethnographical perspective, is detailed in section 3.3.

A process of interaction analysis was employed (Jordan and Henderson, 1995) to uncover the recurring themes within participants' activities. This analytical process was undertaken inductively and reflects the six phases of thematic analysis identified by Braun & Clarke (2006).

Several themes were defined and are presented within the following sections as: *Functionality*, *Reframing reality*, *Expectation and presence*, *Enjoyable exploration* and *Memory, mood and atmosphere*.

## 6.2 Findings from participant interviews

In this section each of the themes defined from the analysis of the participant interviews are presented in detail.

### 6.2.1 Functionality

As a part of successfully authoring their remote AAR experiences, all participants managed to attach virtual audio sources to specific locations and objects within their real-world environment via the aforementioned process of *tagging,* with no participants reporting any *perceived* audible changes in the locations of their *anchored* spatialised audio sources:

> Participant 7*: "It never felt like it was… I was never questioning that* [the sounds staying anchored in their location]*. I assumed it stayed where it was. I didn't think, oh, that sounds a bit odd."*

Furthermore, we can hear how a participant managed to successfully re-listen to virtual audio content that they had previously heard by returning back to the physical spaces within which they had previously heard it:

> Participant 3: *"I was very much going back and forth to listen back to the sounds I had heard there before…"*

In addition to this, we can observe how, when explicitly asked, all the participants agreed that the virtual sounds sounded like they emanated from the locations and objects that they had tagged, and that this was perceived as being realistic:

> Participant 2: *"I think all of the sounds felt realistic, yes definitely, I think I was a bit surprised…"*

Other participants, one triggering a spatialised audio sample of a knocking sound attached to their front door, the other triggering a large mechanical ticking clock sound attached to a clock in their house, commented on the perceived realism of the virtual sound sources, and how they effected their bodily exploration through physical space :

> Participant 3: *"it made me want to move away from the door, it was a bit scary I guess. So it's like, well it's horror, so you know not to open the front door when it's knocking."*

In the next excerpt one participant comments on how they didn't even consider that the virtual clock sound wasn't emanating from the clock in their house which they had tagged, even though they knew their clock didn't make that kind of sound, citing their immersion within the experience as the reason why they perhaps didn't question this:

> Participant 2: *"Yeah, I didn't really think about that* [if the clock sounded realistic] *I think your mind does… make it so because you are hearing something and even if it's I think… because you're hearing something, because you become involved in the app and that experience it sort of, it doesn't clash. It didn't make me think oh yeah, well my clock doesn't make that sound. You know what I mean? It's like that virtual reality stuff… it's so real."*

What's also interesting in the above quote is that the participant draws a similarity with virtual reality (VR) in relation to the *realness* of this experience. This is, perhaps, somewhat intriguing in itself as we are dealing with virtual audio content and not the primarily visual-based content of a VR experience.

As such, the virtual audio's ability to suspend disbelief on a level similar to primarily visual-based VR experiences speaks to the potential effectiveness of AAR in this regard.

### 6.2.2 Reframing reality

The previous participant's quote was made in relation to them realising that their small, battery powered, wall-mounted analogue clock bore little resemblance to the large, mechanical ticking grandfather clock audio sample that they had attached to it. The disassociation between this virtual audio content and the physical object to which it has been attached appears to have remained unnoticed by the participant up until their post-participatory interview. We can witness this same phenomenon more explicitly within the following quote from another participant, who notices this disassociation, again seemingly retrospectively, in relation to both the audio augmented clock and tap. In addition to this, we also hear how the audio augmentation has rendered a familiar object with new characteristics:

> Participant 1: *"That's really interesting, it's a really good question* [Did you feel that the sounds you were hearing we're coming from these actual objects?] *because the clock that I er… tagged is quite a contemporary clock – it doesn't tick, you know. And suddenly it becomes an old clock and, er the dripping tap is in the utility room which is only three years old, it's still brand spanking new, and of course it doesn't drip. So, It's er… well different, it's a different tap!"*

This rendering of the familiar as *different* also extends to the overall setting of the experience, not just the individual objects within it, as illustrated by the following quotations:

> Participant 1: *"No one ever knocks on our door as we live down the bottom of a lane, or very, very rarely does anyone knock on our door, so it's then, again, you're in a different house."*

> Participant 2: *"It's sort of like it's not your house… while you're doing it, it does feel like it's not your… space. It wasn't my house anymore, I was a stranger…"*

Whilst we have accounts from 9 out of our 10 participants that either make reference to, or confirm when explicitly asked, that the experience '*reframed'*, rendering '*different', or* indeed rendered *'unfamiliar',* their environment, for one participant these transformations are far less compelling:

> Participant 5: *"It's really cool and effective, but it's still your house."*

Interestingly, this participant only managed to *tag* two windows prior to *playing* the experience, and no other physical objects or features (such as the door, tap, clock or kitchen cupboard). Therefore, this participant's audio content would have comprised of the musical soundtrack and the audio samples related to the window tags, such as the wind, rain, thunder and possibly the sound of cawing crows.

What also stands out is that Participant 5 appears relatively indifferent in their response to the interviewer's question: *Do you think the audio content made your environment feel different?* This observation can be made by comparing their response with some of the other participant responses to this same question:

> Participant 5: *"Obviously the creepy piano in the background makes it a little bit different. So yeah, it was almost like an extra experience where it was interesting to see how the room you're normally in changes as you're walking backwards and forwards, seeing how the sort of the direction of the volume of the rain and the sounds and that kind of thing."*
>
> Participant 2*: "Well yes, it actually does because you're just looking at it from a different point-of-view, and the fact that my door doesn't creak like that or, you know the bangs or whatever."*
>
> Participant 4*: "Yeah, like I said, I did it last night and I was glad I didn't do it in the dark, it was spooky, cos' it was a spooky clock, and I don't know whether the wind was blowing? You know sometimes when you're concentrating on one thing but you get the atmospheric noises in the past, but all of a sudden you've got this heavy breathing somewhere which did made me look around…"*

In comparison to the other participants, we can see that not only does Participant 5's response seem somewhat uninvolved, but we can see how they revert to their experience with the functional and interactive qualities of the technology, rather than providing any experiential or affectual feedback like our other participants.

Perhaps equally as interesting is how, in response to this question about the feel of their environment, participants 2 and 4 quickly draw upon their experiences with individual audio augmented objects.

### 6.2.3 Expectation and presence

Although I am beginning to gain insights into the function the visual presence of an audio augmented object has in the creation of an alternate object, atmosphere and an immersive audio augmented reality experience overall, it seems that it is not a requirement. Or, at least, not a requirement when it comes to generating the possible existence of such an object within the experience. This observation can be made by looking at some of the participants' experiences with virtual sounds on the periphery of their hearing which were attached to objects beyond the periphery of their vision:

> Participant 1: *"The clock fascinated me, I don't know whether it was me or the app, but you can hear when you get close to the clock it gets louder. You can sort of hear it when you go into the kitchen, then it's gone… again the dripping tap, when you head towards it, it's getting louder… lovely sensation of when you go towards something it gets louder"*
>
> Participant 2: *"Yes, definitely, the sounds were [guiding me]. I think it was easy to get involved in it… I was waiting for the next one... when you're sort of moving around and wondering what the next sound is going to be."*
>
> Participant 3: *"There was definitely moments when I was just kind of free-roaming and just kind of like not really thinking about where I was in the room and just letting the sound sort of, I guess, take me to those places."*

Seemingly, such findings only go to demonstrate the natural affordances of sound, after all we are often in the business of hearing things before we see them, but within an AAR experience the implications of this appear to be of significant importance. Whilst we can see that the probable existence of an object can be suggested to the listener, the specific qualities and characteristics of that object can also be communicated (a *dripping* tap, a *mechanical* clock etc.) both of which can then be reaffirmed, or confirmed, visually once the user achieves line-of-sight with the object; the listener *expects*, then the listener *sees*.

Whilst dealing with the perceived existence of audio augmented objects, it may also be noteworthy how one participant recounts: *'the dripping tap is in the utility room which is only three years old*" subsequent to undertaking the experience. Does this excerpt suggest that, not only does the tap still exist in this location for the participant (which, of course, it does), but the qualities attributed to it by the audio augmentation do as well.

### 6.2.4 Enjoyable exploration

There are a considerable amount of observations that can be made that point to *Horror-Fi Me* being an *'enjoyable'* and *'fun'* experience for the participants, and that support the idea that these types of AAR experiences have a playful quality about them.

> Participant 1*: "Lovely sensation of when you go towards something it gets louder"*

> Participant 2: *"Made me smile and jump with the sounds when I heard it for the first time… Fun, enjoyed it."*
>
> Participant 3: *"I thought it was really good. It was wicked... I guess like, it felt a little bit like playing a game, like a 'LA Noire' type game on PlayStation or something."*
>
> Participant 5: *"Spinning around to understand and experience the panning…"*
>
> Participant 5: *"I was more focussing on me walking to and fro… seeing how like, where I was effecting the sound".*
>
> Participant 5: *"…and like, walked to each window. Um, so I had like a big pair of headphones on. Um, yeah and just like spinning around next to the window and seeing how the sound changed as I was facing like different directions which was really responsive. Then seeing how it changed when I was walking from one side of the house to another."*
>
> Participant 6: *"I thought it worked really, really well as you moved around, I liked the fact that the sounds seamlessly gelled into each other, rather than it being a… you can hear a glitch when it switches from one sound file to the next, as one faded into the other… you had the layered sound, it was really good."*
>
> Participant 7: *"I was probably trying to figure out what I could influence, and what I was influencing…"*

We can also see from the previous interview excerpts that these *playful* interactions develop a familiarisation and understanding of the system and the encompassing effects that the participants' embodied interactions have on the delivery of their audio content. Furthermore, we see how this familiarisation and understanding manifests as exploration of the physical and audible space, and how these initial explorations lead to further exploratory forays.

During some participant interviews it sometimes proved difficult to delve deeper into the nature of the participants' experiences with individual augmented objects due to an apparent fascination with the novelty of the experience of manipulating interactive binaural audio with the movement of the body. We can also see how the perceived immediate responsiveness of their embodied interactions with the system is referenced in relation to this.

Additionally, there's some evidence of exploration being initiated by the increasing absence of sound, the fading away of currently playing sources as the user moves away from them. This appears to encourage the user to explore further, or in this case to re-explore, through an appreciation of the effect of their direct interaction with the spatial characteristics of the audio content:

> Participant 3: *"because of that* [the audio fading away as I moved] *I was wanting to track back round, and sort of hang about… to see if I could hear them again"*

### 6.2.5 Memory, mood and atmosphere

> Participant 3*: "I guess like, it felt a little bit like playing a game, like a 'LA Noire' type game on PlayStation or something."*

Many of the previously discussed excerpts demonstrate that participants felt a change in the *feel* or *mood* of their physical environment. The participants' descriptions of this include feelings of unfamiliarity, spookiness and a reference to the creation of a video game-like feel. In addition to Participant 3's reference to the video game title *LA Noire* [1] concerning their description of what the experience felt like, in the following excerpt we can hear Participant 7 reference the video game title *Silent Hill* [2] explicitly in relation to describing the experience's atmosphere:

> Participant 7: *"…specifically Silent Hill… I played it on a GameCube, probably about 15 years ago… it was a ghost adventurer type of thing… the atmosphere took me straight back to that... the atmosphere and being able to kind of… I'm not a massive gamer but it just reminded me of the atmosphere and being able to kind of take myself to a completely different time very quickly... and to get space as well... there wasn't just one fixed point, there was layers of that as well. So that helped with that transformation really"*

Within this excerpt we can also hear how this atmosphere is closely associated with the memory of playing this this video game around 15 years ago. The stimulation of memory through interaction with the experience and the effect it has on the atmosphere of the physical space is also mentioned in the following excerpt:

> Participant 6: *"…there was an element of wanting to know what sound is going to occur and what, how that was going to provoke my imagination and memory. Which was good. I quite liked that. I liked the fact it provokes memories, and er… it changed the space in that regard."*

---

[1] *LA Noire* is described as a *neo-noir* detective action-adventure video game inspired and influenced by the 1940s and 1950s cinematic *film-noir* style. It features an original musical score along with licensed recordings by artists from the *film-noir* period.
[2] *Silent Hill* is described as a survival horror video game influenced by the psychological horror literary genre. It features an original musical score along with licensed musical recordings from a variety of genres including industrial and rock.

We can also see how Participant 6 alludes to feelings of anticipation, curiosity, and even reward, in association with the exploration and delivery of audio content that could '*provoke'* the *'imagination and memory'*.

There are more components involved in the creation of *atmosphere* or *mood* within the experience though, and it is not solely the consequence of the personal interpretation of audio content, or indeed the association of personal memories that the embodied interactions with spatially interactive binaural audio appear to prompt. The following interview excerpts suggest that the construction of *atmosphere*, in part, is a product of participants' pre-conceptions of what they think, or indeed what they expect, the experience should be like:

> Participant 2: *"Definitely created, I think the mood, well I thought it was the mood that you were trying to create."*
>
> Participant 3: *"…well it's horror, so you know not to open the front door when it's knocking."*
>
> Participant 6: *"I suppose the fact that the title of the interface had the word 'horror' attached to it, so you're already… there's a connotation, an expectation."*

Finally, we can see that a pre-conception and expectation is created by the contextualised framing of the experience, suggesting that the name, user interface design and other visual and non-audible elements play an integral part in the creation of a different framing of the listener's environment and the creation of both *mood* and *atmosphere.*

## 6.3 A compositional perspective

Upon completion of the application, and after its subsequent launch, a structured interview was conducted with the composer of *Horror-Fi Me's* soundtrack. The audio recording of this structured interview was transcribed and thematically analysed using an inductive approach, analysis was undertaken using the six phases of thematic analysis identified by Braun & Clarke (2006). On this occasion, the main themes were defined as: *Adaptability*, *Re-enforcing contexts*, *Musical movement*, *Complimentary content* and *New possibilities*.

After briefly refreshing the interviewee's memory of the *Horror-Fi Me* AAR experience and the role they played as the composer for its musical score, the interviewee was asked the following questions:

1. *What are your general thoughts regarding the Horror-Fi Me experience?*
2. *What approach did you take in writing the music for Horror-Fi Me?*

3. *How (if anything) would you approach differently next time?*
4. *Do you have any specific thoughts regarding how the spatialised audio effects worked in conjunction with the music?*
5. *Can you outline what you think the specific strengths and potentials of Horror-Fi Me are?*
6. *Can you outline what you think the specific weaknesses or challenges of Horror-Fi Me are?*
7. *Do you see any further compositional opportunities within this type of experience?*

Within the following sections each of the defined themes resulting from the thematic analysis of the interview with the composer are presented in detail.

### 6.3.1 Adaptability

The approach taken very much reflects a way of working outlined by Phillips (Phillips, 2014), where an original piece of music is commissioned of a specific style or genre for use within a video game. Rather than specifically scoring for sequences or events within the game, the piece remains a musical work that stands on its own merit with, for example, any variations in theme, dynamics and instrumentation being a consequence of the process of musical composition. Although the piece of music is applied within the context of the new media experience of audio augmented reality, it was written as a piece that could be performed in concert, if so desired:

> Composer: *"I always work in a monothematic way, I might have sub-themes, or sub-gestures, but generally every strand of the piece of music will be directly related to one theme… I was hugely influenced by Schoenberg, and I've adapted his serial techniques… every piece I write, the process that I go through is the same. When I write, and it doesn't matter for what… all the music I write is absolute, it's about the music itself. I don't work in a programmatic way, there's no plot in the music, it's just the music itself."*

Specifically reflecting on the linear use of the composition within the experience, in the following interview excerpt the composer reflects on how this non-programmatical composition in the sonata form would lend itself well to being applied in a non-linear way within the experience:

> Composer: *"…they* [the musical sections] *always relate very carefully to each other, and in a way that's what these individual movements are, like I say there isn't a programmatical element in the music – it's not about anything. So there is no reason why it couldn't be broken up."*

### 6.3.2 Re-enforcing contexts

The composer goes on to reflect upon the relevance of their musical style to cinema and, more specifically, horror cinema:

> Composer: *"What a lot of the horror film composers from the 1950s… they borrowed a lot of Schoenberg's ideas because they found that expressionism in his music suited the horror genre. …you know, most people's reference to contemporary art music is through cinema."*

We can see these comments verified in the interviews with participants' in the previous study, perhaps most explicitly in the reference to *'the creepy piano'* in section 6.2.2.

### 6.3.3 Musical movement

In addition to this, as a participant in the experience, the composer also reflects on how their music was afforded a sense of movement within architectural space:

> Composer: *"My world's a very static world… What was nice about the app was that it allowed the music to move through the house, which was interesting for me…"*

### 6.3.4 Complimentary content

*Horror-Fi Me's* composer also reflected upon a feeling of *'wholeness'* regarding how the two spatially distinct components (the binaurally rendered sound effects and the stereo rendered musical content) functioned and worked together within the experience:

> Composer: *"I was really pleased by it, again because I had no preconceptions about what was going to happen… I felt a sense of wholeness with it."*

### 6.3.5 New possibilities

Through experiencing their music being used in this way, the composer goes on to reflect on other possibilities and discusses some potential ways in which the music could be adapted and used to further effect within an experience like this. The composer also reflects on how the technology behind *Horror-Fi Me* could be used to create new musical artforms and new ways of creating and delivering domestic musical experiences, hinting that although the way we materially interact with it has evolved, the way in which we experience it within the domestic environment has changed little:

> Composer: *"…because the nature of music has changed so much, particularly over the last 30 years, there's the potential for a new artform, a domestic artform as well. The thing about the iPhone you know, is that I have hundreds and hundreds of albums on here, but in a way I don't. Because they're not even there are they, they're somewhere else. Because music has changed so much in the way we purchase music and we interact with it, that maybe now is the time to think about our audio experience, and have it as this malleable thing as you're walking around the house."*

Referencing *generative music*, in the following excerpt the composer reflects upon how the system could be utilised for the delivery of new musical content and experiences:

> Composer: *"…where it's generative music, it invents itself… the potential of having an artform which changes the melody and harmony changes itself as you go around would be amazing, you'd never hear the same piece twice."*

In addition to this, they reflect on how the composition is playable and all theoretically crafted, and could, if desired, be played by performing musicians. Further reflecting upon this they suggest:

> Composer: *"One could write for the environment within which one is walking through."*

Adding to this previous point, this time from a historical perspective, and with reference to the use of classical music in video games, they suggest that this new listening experience provides an opportunity for a new way to experience and write music:

> Composer: *"Even with the Tomb Raider score it was done with a symphony orchestra, playing slow classical music. Maybe the art of writing for this app is to not actually even consider the concert hall, but to consider the environment which somebody is walking through… What this app development should be able to do is get us away from that, and get us away from narrative, and allow us to experience sound and music in a completely different way, which isn't concert hall, which isn't Tomb Raider, but it's actually something that's a bit more domestic and a bit more unusual - a new artform."*

## 6.4 Discussion and conclusions

The following section discusses and draws conclusions on the key points resulting from the study of *Horror-Fi Me* and the interview with the composer presented in the previous section. These include: the authoring of the experience, how realism is constructed, creating and exploiting expectation, cultural authenticity and the use of non-spatialised audio. Also discussed are the new musical opportunities arising from these studies.

### 6.4.1 Authoring the experience

Firstly, we can conclude from the participants' successful remote authoring and undertaking of their own audio augmented reality experiences that the deployed mobile application is a capable tool for the creation and delivery of AAR experiences within which virtual sound sources have the potential to be perceived as emanating from within the physical environment. Furthermore, one can determine from this successful domestic deployment by non-professional participants that a successful gallery or museum deployment by curators and creative professionals could also be realised using a similar approach and mobile application authoring tool. This subject is discussed further in section 7.5.

### 6.4.2 The construction of realism

The somewhat underwhelming experience regarding the creation of a contextualising sonic atmosphere by the participant who only tagged two windows (and no individual physical objects) suggests that the physical audio augmented object plays an important part in the reframing of reality, and that the listener experience of interactions with architectural features augmented with contextualising ambient audio is not enough to achieve a believable reframing on its own. This suggests that components within the experience with a more direct connection between the visual and audible are required to achieve this; associations that go further beyond the acousmatic listening experience where the physical source of the sound is visually present and not just its spatialised audio source. It should perhaps be noted that it remains entirely possible that the situated, though imaginary audio sources (such as the rain against the windows and the crows outside, unless of course it is actually raining and there are crows outside) take on a greater level of realism when directly augmented objects are present. This can be largely determined by the feeling of telepresence reported by many of the participants and that the sources of the audio and their location in reality, for the main part, remained unquestioned. This further indicates that the acousmatic listening experience, through the introduction of the audio augmented object, has been replaced by a direct listening experience and that the presence of audio augmented objects provides a stimulus for the *aural imagination* regarding the sources of sounds that cannot be visually determined or verified. A discussion on how audio augmented objects alter our listening experience is included in section 7.1.1.

### 6.4.3 Creating expectation

Through the process of audio augmentation, a physical object's virtual, or indeed real, characteristics can be communicated to the listener prior to it being visually apparent. These characteristics place an expectation in the mind of the listener, an expectation that can be affirmed by the visual and physical presence of the object itself. Furthermore, these projected characteristics need not directly relate to the exact nature of the physical object itself. The association between the object and the characteristics communicated by the audio with which it has been augmented can be fluid, and at times tenuous, whilst still maintaining a sense of realism. How far these associations can be stretched is, perhaps, the subject of further examination. Though it does appear that a listener's perception of these objects can be challenged and, in some cases, altered in a way that contradicts their visual appearance.

It would make sense that this process operates best when an audio augmented object is first heard and then seen, and it would also appear that this process provides an authoring strategy for the creation, management and manipulation of object-audio associations and the creation of *sonic atmospheres*. Additionally, once encountered, these characteristics can persist throughout, and beyond, the duration of the experience without continual visual or audible presence. This last point seems quite revealing in that it suggests that every audible encounter during the experience can continue to contribute to both the sense of presence of the object and its characteristics, and the presence of the *sonic atmosphere.* This finding correlates with those from the previous study (see section 5.3.5) where, despite a visual disassociation with the audio augmented object, a strong sonic and physical attachment to it can remain. Many of these points are explored further in section 7.2.

A listener's perception of objects can be altered via the associative relationship between the virtual audio content and the physical object, the *dripping tap* and the *creaking door* being two such examples. Whilst we can probably accept that most doors could *creak* and most taps could *drip*, it seems significantly further beyond the realms of possibility, and imagination, that a battery powered clock, or small digital clock, could tick and chime like a large mechanical grandfather clock. It therefore appears that the boundaries of convincing associations can be stretched, significantly. When a participant's experience of an object is dramatically altered via audio augmentation, the visual reference of the physical object combined with the virtual audio propels the virtual audio content into reality. These instances remind us that this is not in fact an audio only experience, but an AAR, or mixed-reality experience, where, in the spirit of the project's original design (see section 4.5), reality becomes the screen, or the visual stimuli, that compliments the spatialised virtual audio content. This topic is discussed further in section 7.1.2.

On the subject of the presence of character, one can think of this much in the same way as a film may use a wide-angle and recognisable view of specific city, and then cut to an interior of a restaurant, by association the restaurant is perceived to be present within this city, though it could, in fact, be anywhere. Of course, additional references can be assigned to the restaurant in order to bolster this assumption, but the point is, these signifiers need not be present for the duration; context has been assigned, remains, and can be reasserted as required. A discussion regarding how context can be assigned within an audio augmented reality can be found in section 7.2.3.

Additionally, we see how an experience which is comprised of the ingredients of self-determined exploration, such as anticipation, curiosity, fulfilment and reward, stimulate the participants' imagination and memory and help contribute to the creation of a meaningful experience. Again, not only does this clarify the importance of the nomadic experience, but the way in which this unfolds through a process of familiarisation, exploration and investigation further supports the identified sequence of interactional phases revealed as a result of the previous study (see section 5.3). The importance of self-determined exploration in the creation of a meaningful experience is discussion further in section 7.2.4.

### 6.4.4 Cultural authenticity

The strong association between the style of music used and horror cinema, along with the participant testimonials that compare their memory and experience of the application to that of playing horror-style video games, suggests that the inclusion of music within the *Horror-Fi Me* experience made a significant contribution to the stimulation of associated memories. Though, interestingly enough, these associations focus on the experience of horror-style video gaming, rather than horror cinema itself, suggesting that the closest experiential reference point for users is, in fact, the video game experience, rather than a cinematic or musical one as suggested by the soundtrack's composer. Furthermore, we see how the similarity between this type of audio augmented reality experience and video games extends through to the authoring process, namely the adaptability and successful appropriation of contemporary compositional arrangements for the purposes of supporting the creation of context, or atmosphere, within such experiences. It is also shown that the perceived atmosphere of such an experience relies on a combination of factors that include how it is initially framed and presented to the participant, the type of audio content delivered, how this audio content is applied and the considered contextualisation of the mode of embodied interaction.

### 6.4.5 Non-spatialised audio

Whilst the music's successful function in relation to the creation of an appropriate *sonic atmosphere* for the experience is evident, we should also acknowledge what is referred to by the composer as the *'wholeness'* of the musical experience. The composer's personal observation seems of additional consequence as they are commenting on the experience of their own stereophonically rendered musical composition being combined with dynamic binaural audio content. Although the type of functional compatibility observed by Strickland relating to non-binaural audio content's ability to allow the spatialised audio content to punctuate through the overall audio production (Strickland in Jahn, 2015) is not explicitly referred to, the continued ability for users to successfully localise the spatial audio sources within the audio mix, along with the composer's appreciation of the *'wholeness'* of the musical experience suggests that dynamic binaural audio and stereo rendered audio content can effectively be combined within the context of a single audio experience. The function of this type of audio content within the audio augmented object reality is outlined in section 7.2.1.

We also see evidence of an additional functional relationship between these two differently rendered audio sources. Some of the composer's comments regarding the movement of the music through space very much echo science-fiction author William Gibson's observation regarding the personal stereo experience's ability to *'take music and move it through landscapes and architecture'* (Gibson in Bull, 2000) This physical spatialisation of the stereophonic musical score is thought of not just in a way similar to the stereophonic mobility afforded by *The Walkman* experience, but with the added spatial interactivity of the translational and rotational positioning of the listener in relation to the physically anchored and spatialised audio content, distinctively separating it from other mobile audio listening experiences. The potential for the creation of new mobile audio listening experiences as a result of an audio augmented object reality is discussed in section 7.4.

Furthermore, we see how, in a similar way to video games, contemporary compositional techniques can be appropriated within the audio augmented reality experience and how previous insights into the contextualising and complimentary function of contemporary and popular cultural references within the audio augmented reality experience can be further re-enforced through the inclusion of a musical component.

### 6.4.6 The musical experience

What also presents itself of particular interest here is the correlation between the musical AAR example described by the curator and the compositional possibilities relating to this technology outlined here by the composer. Within the interview with the curator (see section 5.4.3) a possible example application for this AAR experience is described involving a group of audio augmented musical instruments that collectively play a piece of music, though upon closer inspection their parts become individually discernible and additional information or layers of audio content can be triggered. Although this initial concept is provided from the curatorial perspective of a functioning epistemic sonic atmosphere, by combining this idea with that of the composer's regarding the creation of a *'generative'* and '*malleable'* composition with changes of '*melody and* harmony' based on user movement, one can begin to envisage exciting compositional potentials and musical listening experiences that further advance sound reproduction's fascination with the creation of acoustic virtual realities, as outlined in section 2.5.1. The possibilities for new musical listening experiences afforded by an audio augmented object reality are discussed in section 7.4.

Additionally of interest, is the potential re-materialisation of the musical listening experience that is indicated by the composer's comments about the intangible nature of the consumption and delivery of contemporary music and the suggestion that the musical listening experience demonstrated by *Horror-Fi Me* could alter this. If one considers, for example, the comparison between the vinyl album, its artwork, packaging and the consumption of mp3 formatted and distributed music, then one could perhaps consider the projection of music into the physicality of real-world space as a re-materialisation of the currently intangible nature of the domestic musical listening experience. This materialisation of the virtual music listening experience proposes that, via the audio augmented object, the acousmatic musical listening experience can be transformed into a direct listening experience which could also include such genres as audio drama. This point is discussed in more detail in section 7.4.

In summary, we see how the developed mobile application is capable of enabling lay-users to remotely author and play audio augmented reality experiences comprised of audio augmented objects. Within these authored experiences, we also see how the physical environment, along with the individual objects within it, can be reframed and afforded new characteristics. Additionally, we see how these virtual characteristics can persist during and beyond the immediate experience, how they function to create and maintain context and to uphold *the suspension of disbelief*. Along with gaining further insights into strategies for authoring and curating audio augmented reality environments, we see how the presence and character of these audible objects can create a sense of expectation in the listener which can be verified, or re-enforced, by the presence of the physical object itself. Furthermore, we see how the perceived presence of the physical sound making object turns what would otherwise be perceived as an *acousmatic listening* experience into a *direct listening* experience. Again, as in the last study, we see how the listener's embodied interactions with the dynamic binaural audio content evoke memories and realise a highly intuitive interactive experience. Here, memories, context and cultural references are evoked, created and assisted through the inclusion of a musical soundtrack to the experience. Whilst demonstrating some of the functional properties of music within an audio augmented reality experience, this also provides insights into the possibilities for a *dynamic binaural music.*

# 7 Discussion

Within the following chapter some of the key themes evident within the literature, practice, studies and interviews are discussed in further detail. The themes that have been chosen for discussion not only reflect those which recur within the practice, studies and related work, but they have been chosen for discussion with consideration to how they may help to shed light upon the research questions. The chosen themes are considered alongside discourse from a variety of related interdisciplinary sources, including media theory, human-computer-interaction, sound studies and the arts. The themes discussed are: the perception of reality in the audio augmented object reality, how functional audio augmented object realities can be designed, and the potentials and implications for employing audio augmented objects within the studied contexts of the museum and home. Also discussed is how audio augmented objects can be effectively realised.

## 7.1 The perception of reality

The creation of audio augmented objects realises new experiences in acoustic virtual reality. For example, we have seen evidence within the findings of both studies how augmenting objects with audio can reframe the character of physical space as well as the character of the object itself and other objects around it. In addition, we also have seen how audio augmented objects can help affirm the presence of virtual sound sources in the physical reality (see section 6.2.3), and how the virtual characteristics of both space and object, and the implied presence of virtual content, can persist and provide an ongoing contribution to the user experience (also see section 6.2.3).

### 7.1.1 Breaking and remaking the *Acousmatic Field*

Through the various projects described within chapter 4 we see a variety of different types of physical objects being augmented with virtual audio content, from photographs to radios, architectural features through to everyday household items.

By placing virtual sound sources within physical space and allowing them to be perceived as emanating from real-world objects the listener experience can be *transformed* from an *acousmatic* one, to a *direct* one. Equally, the same practice can result in the *construction* of an acousmatic listening experience, this transformation, in either case, is dependent on the type of object and the virtual audio content with which it is augmented.

Composer Pierre Schaeffer's description of the *Acousmatic Field* (Schaeffer, 1966) alerted us to new modes of listening afforded by the advent of new technology, in Schaeffer's case this was the magnetic tape recorder. Here we consider the changes in the experience of listening afforded us by audio augmented objects and the way in which this alters our perception of reality.

Whilst Schaeffer's *Acousmatic Field* deals with the removal of the sound's source from the listening experience, with the audio augmented object, we are, in many ways, dealing with its reintroduction. This is perhaps most evident when we are audio augmenting *sound making objects*; objects that are directly responsible for the creation of their own sound, or to use Schaeffer's Pythagorean distinction, those objects with which we have a *visible, touchable* or *measurable relationship* with the direct source of the sound.

An audio recording of rain pattering against the glass of a window is perhaps a good example here. If heard through loudspeakers or headphones with no association to its original, physical source in reality, then we are involved in a *acousmatic* listening experience; the Pythagorean veil remains intact. On the other hand, by directly associating this audio recording with a plausible and probable physical source in reality, we are bypassing the *acousmatic* experience; the sound source has visibly and physically returned to the listening experience and it ceases to be perceived as *acousmatic*.

Placing sound in three-dimensional space does not make it non-acousmatic (at least from Schaeffer's Pythagorean perspective) as the source of the sound is still not visible. This, on the other hand, changes as soon as we place that sound at the same location as a plausible, physical sound making object. As the source can perceived to be present and visible. Therefore, this change only takes effect if the physical object is indeed perceived to be the source of the virtual content, but (as identified in section 4.4.3) this is entirely possible, and in these cases a perceivable transformation has taken place from what Schaeffer terms a *non-live listening* experience into a *live listening* experience.

This transformation not only relies on the effective positioning and rendering of realistic virtual sound sources, but also on the plausibility of the augmented object to emit the virtual sound with which it has been associated. It seems that the nature of this transformation is also dependent on the type of object that is being audibly augmented. For example, to reunite a radio broadcast from a sound archive with a radio set would be to reconstruct a lost acousmatic experience. To attach a sound recording of rain against a window to a physical window, replaces a potentially acousmatic experience with a perceptively direct listening experience. Therefore, the effect of either breaking or remaking the *Acousmatic Field* appears to be a consequence of the *type* of object and its relationship with the audio content with which it has been augmented.

In either case, whether recreating an acousmatic listening experience, or creating a direct listening experience (where the augmented object is perceived as the direct source of the sound itself), a more connected and intimate relationship between object and listener is created. Whilst the audio augmentation of a radio with archival radio broadcast material is not a direct listening experience, it is more familiar and truer to its original context and therefore a *more* direct listening experience than, for example, listening to the same content via a web-based audio player.

This familiar, contextualised and more direct experience with the source of the sound (whether perceived as direct or contextually acousmatic) constitutes a more haptic type of listening experience, one that communicates the physicality of the sound, its source and their relationship to the listener.

Within the discipline of sound studies haptics, in this sense, is not thought of something that you can necessarily physically touch, or that can touch you, rather the creation of *haptic space* (Kassabian, 2013), the creation of a more intimate and close space than which can be achieved via the acousmatic or *detached* acousmatic experience.

The less direct the listening experience, the greater the degree of detachment from the sound source which negates the intimacy of the experience and which reflects Deleuze & Guattari's philosophical concept of *Smooth* and *Striated space* (Deleuze and Guattari, 2005). Within Marks' interpretation of Deleuze & Guattari's concept of *Smooth* and *Striated space*, we can equate the perceived experience of being present with the source of the audio as a *close*, or *smooth*, experience, and the act of having to imagine that source as a more *distant, or striated,* experience. The former, albeit not tactile, resulting in a more intimate and physically close experience.

This point is largely reflected upon in relation to Marks' distinction between *haptic visuality* and *optic visuality*, haptic visuality being a close experience that draws on other sensory experience (such as kinesthetics) and *optic visuality,* relying solely on the visual experience (Marks, 2002).

Here we can perhaps propose a distinction between a *haptic listening* experience and an *auditory listening* experience. *Haptic listening* being a hearing experience that draws on other senses as well as the auditory, such as vision, touch and kinesthetics, and *auditory listening*, being one which relies solely on the auditory experience.

We have seen, specifically in relation to *Alien Encounters* (see section 4.4) and *The McMichael Experiment* (see section 4.3), this Intimacy of haptic space reflected in participant's interactions with both the audio augmented object and the virtual audio content. In the former, this spatial intimacy is reflected in participants' close inspection and exploration of the audio augmented objects, in the latter, this is reflected in participants' focussed listening and secondary-focussed listening phases of interaction (see section 5.3).

By way of a summary, the inclusion of audio augmented objects within the virtual acoustic reality realises a listening experience that draws on other sensory experiences. This realisation of a more *haptic listening* experience is capable of creating a *smoother*, closer, more direct and intimate relationship between the audio augmented object, the virtual audio content and the listener.

The breaking or remaking of the acousmatic field relies on the plausibility of the physical object and virtual audio relationship, in other words, an expectation that the object is a viable source of such a sound is required on behalf of the listener. This therefore begs the question: What happens when these physical objects are augmented with different types of sounds? What is the effect of using sounds inspired by the object or associated with different types of objects?

### 7.1.2 A third reality

Firstly, let's consider again participants' experiences of the grandfather clock audio sample described in chapter 6. Within their accounts, participants experience this audio content emanating from a variety of modern, non-mechanical clock types with little concern or recognition of its effect on the reality of their experience.

Within the discipline of computer science Azuma et al. (Azuma *et al.*, 2001) consider AR experiences that remove real-world objects, in addition to adding virtual objects, as *mediated* or *diminished reality,* Carmigniani & Furht (2011) further observe that this removing of objects from the real-world corresponds to the covering over of the object with virtual content that matches the background in order to create the impression that it is no longer there. Although Carmigniani & Furht's (2011) observation can, perhaps, be best illustrated from the perspective of visually applied AR, for AAR their observation sits well with Schaeffer's definition of the acousmatic listening experience, as well as many of the recorded experiences of participants presented within this thesis. We can, though Schaffer's definition, go some way to explaining how an object's character can be perceived as being different and at odds with its physical and visual presence due to it being augmented with an audio source if indeed perception moves from object to content. We can also, perhaps, begin to understand, and apply a little of the logic surrounding the visual augmentation of reality to the audio augmentation of reality, if we consider how, through the use of appropriate and corresponding contextual background audio, by covering over the object with virtual audio content, the physical presence can be altered, if not entirely removed. A method of audio blending, where the augmented object's audio corresponds to, or is in some way complimented by, the background audio content, thus making its newly acquired character blend in with the physical environment, in a similar way to visually masking it with virtual content that matches the background within a visually orientated AR application.

In relation to the participants' experiences with the grandfather clock audio augmented object, it may be worth considering the *The McGurk Effect. The McGurk Effect* describes an auditory-visual illusion that takes place when visual and auditory stimuli are merged and perceived as being something different. Perhaps one of the best illustrative examples of *The McGurk Effect* is *when hearing the sound 'ba', while seeing the face of a person articulate the word 'ga', many perceive the sound 'da'.*

Although within our example we cannot explicitly identify the perception of an alternative *third,* as outlined within the cross-modal illusion of *The McGurk Effect*, we do see the perceived character of the physical object changed into something different due to the interaction between the object's visual presence and the virtual audio components of the experience.

We can therefore begin to determine that an audio augmented object's perceptible character within the context of an AAR experience is a consequence of its physical appearance, the virtual audio content with which it has been augmented and the sonic context within which it is presented. By creatively combining these three components it appears that the perceived reality and character of audio augmented objects can be manipulated to generate a perceptible third reality that is a consequence of the combination of the visual reality and virtual acoustic reality.

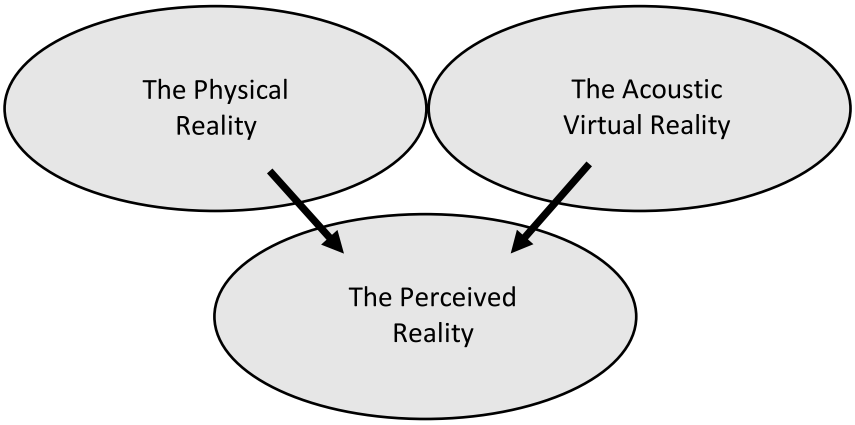


Figure 30. The construction of a third reality. The perceived reality is a product of the physical reality and the acoustic virtual reality.

### 7.1.3 Environmental authenticity

Described by composer Winifred Phillips (2014) in relation to the musical content of video games, *environmental authenticity*, along with *cultural authenticity*, are seen as contributing factors in the creation of an atmosphere that helps realise a *suspension of disbelief.* The concept of *environmental authenticity* within *Horror-Fi Me* appears to extend across the musical content, the sound effects and samples, through to the individual audio augmented object.

Whilst the musical content and many of the audio samples ground *Horror-Fi Me* within the genre of the cinematic horror experience, providing it with a sense of *cultural authenticity*, Phillips (2014) also suggests that the musical content of a video game experience, provides a source of 'emotionally engaging content' that can also contribute to the *suspension of disbelief.*

Phillips here is referencing Coleridge's concept of the *suspension of disbelief* which implies a willing on behalf of the audience to want to believe in what they know to be the fantastical within a fictional narrative experience. Of course, this willingness is pressed upon the audience, and their complicity is manipulated by such literary and cinematic trappings as grounding the fantastical within the believable, or as Phillip's puts it: *'environmental authenticity'*. Or, indeed the creation of a compelling narrative that, in order to immerse oneself within it, the reader or listener most forego any misgivings regarding it's probability. Either way, ultimately the virtual is still distinguishable from reality, and disbelief's suspension, as Phillips (2014) points out, is a result of a conscious contract between viewer and author.

In her book *Gothic Music*, musicologist Isabella van Elferen (2012) cites Schopenhauer's conviction that *'music is an unconscious exercise in metaphysics in which the mind does not know it is philosophizing'* to make a distinction between the *willing* and the *unwilling suspension of disbelief*. Though interesting with regards to the inclusion of musical content within an AAR experience (which is discussed further in section 7.2.3), it is Elferen's unveiling of the two distinct ways that belief can be suspended that seems of specific interest here:

> *"Musical experience is an unwilling suspension of disbelief that can mould listeners into believing in Gothic ghosts: the haunting voices of film music, the domestic spectres of television, the interactive supernatural of computer games."*
> (Elferen, 2012, pp. 180–181)

Within the *Horror-Fi Me* experience we see, through *environmental authenticity*, *cultural authenticity* and the *'unconscious philosophising of the musical listening experience'*, the creation of both *willing* and *unwilling suspensions of disbelief,* the latter manifesting itself with participants seemingly unable to distinguish reality from virtuality.

Just as Phillips (2014) and Elferen (2012) assert the importance of music in generating a sense of cultural authenticity and a belief in the fantastical, Within the discipline of sound studies Bijsterveld (2015) suggests that sounds that have a degree of imperfection, a historical quality, or those that display a degree of unfamiliarity can be attributed with a greater degree of acoustical correctness by listeners. Bijsterveld's suggestion is considered within the context of museums where such audio content can be attributed to generating a better connection to the past. This perhaps goes someway to explaining why, despite the unfamiliar interactions, the act of tuning into period recorded radio broadcasts through the sound of static remained so evocative and *real* for participants in *The McMichael Experiment* and *Alien Encounters*. Indeed, given Bijsterveld's observations, one could argue that the addition of the sound of *out-of-tune* radio static amongst the period radio broadcast material attributed the broadcast content, and the experience itself, with a greater historical authenticity. Additionally, Bijsterveld's observations may also explain the effectiveness of some of the historical and unfamiliar audio content in *Horror-Fi Me*, namely the ticks and chimes of the grandfather clock, in suspending disbelief in the virtual acoustic reality.

Bijsterveld's observations also emphasis the additional importance of the authentic production quality and rendering of virtual audio sources in order to provide an overall sense of environmental authenticity, achieved through close attention to historical and cultural authenticity.

Within the field of HCI, like the experience outlined by (Pausch, 1996), the experiences presented here indicate that believability is different from realism, in that they contain experiences that could not be normally perceived in reality, though can still be perceived as being believable. Instead, this believability, or *unwilling suspension of disbelief*, hinges on a number of different components of the experience that individually add *environmental authenticity* and contribute to a *willing suspension of disbelief*, and that collectively have the potential to contribute to the realisation of an un*willing suspension of disbelief,* or an experience that has the potential to be perceived as being indistinguishable from reality.

## 7.2 Designing audio augmented object realities

In this section I present a framework for attracting engagement within audio augmented object realities. This framework extends the previously presented phases of interaction by providing examples of the type of audio content that can facilitate each phase and help to engage the listener further. Also discussed is how it is possible to design an audio augmented reality that can be perceived as being highly realistic even though it may not strictly adhere to the expected properties of natural acoustics. This observation allows functionality to be designed within the virtual acoustic reality that would not normally be possible in reality, with little consequence to the perception of the reality of the virtual content itself. The component parts of an audio augmented object reality are also described, and the functional attributes of each of these components are detailed along with the importance of the freedom of exploration regarding the creation of meaningful, serendipitous encounters within these experiences.

### 7.2.1 A framework for attraction and engagement

The identified sequence of interactional phases described in sections 5.3 through 5.3.7 and illustrated in Figure 28, along with the identification of both an *object-first* and an *audio-first* approach to soliciting the attention of users (see section 5.4.4 and Figure 31) allows one to think about a general model for attraction and engagement within audio augmented object realities.

Whilst the identified interactional phases help us to understand and design a model that will facilitate a user's experience from awareness through to participation, familiarisation, discovery and engagement, the consideration of both physical objects and virtual audio content as attractors within these experiences allows us to understand how engagement with both object and audio can be promoted.

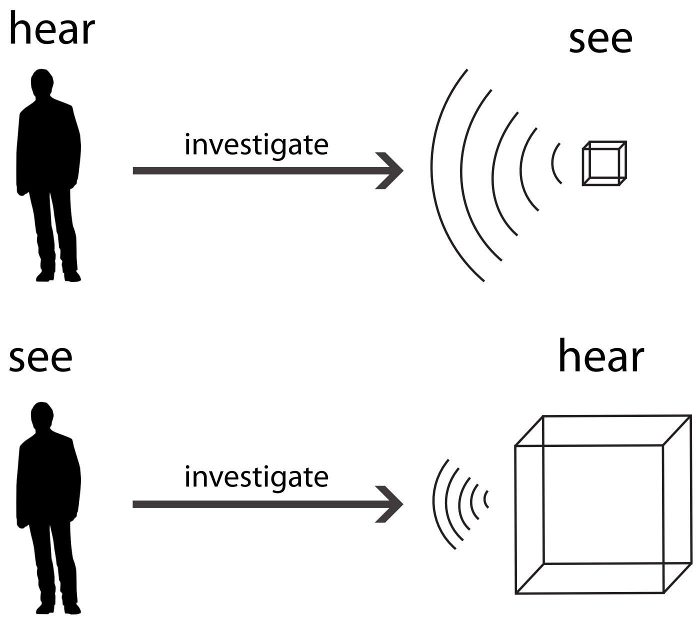


Figure 31. Promoting investigation. Promoting investigation of the physical object through virtual sound, and investigation of virtual sound through the physical object; the respective *audio-first* and *object-first* design approaches.

The possibility of both an *object-first* and an *audio-first* approach brings into focus the function that both sound and object can perform within the context of a single experience. By thinking about how engagement can be initiated and promoted with both virtual sound content and physical objects, we could determine that the virtual sound content could be used to promote awareness and interest in objects, and the physical presence of the object can be used to promote engagement with the virtual audio content. Those that have a small physical or visual presence (though are nonetheless of equal importance within the context of the experience) can be promoted with *attractor sounds*, whilst those that are perhaps physically or visually imposing can be used to draw visitors towards associated virtually exhibited audio (see

Figure 31). Additionally, whilst Figure 31 emphasises the potential use of an audio-first approach for objects that are physically small in size, or with little immediate or long-range visual presence, this approach could similarly be taken for objects that are perhaps obscured by the presence of other physical objects or architectural features.

Furthermore, the use of unfamiliar and unusual audio content as a means of engaging users, as reflected upon by Bijsterveld (2015) and suggested within section 6.3, allow us to begin to consider what type of audio content is best suited to specific stages of the model. This is considered alongside a potential relationship between the identified interactional phase of focussed listening and the micro-investigative nature of the device-based interactions and the recognised desire by participants to enter into and engage with secondary levels of audio content. This latter point is illustrated within the work of Montan (2002) and within participants' feedback of their interactions with *The McMichael Experiment* in chapter 5.

| **Interactional phases** | **Type of Audio Content** |
|---|---|
| Participation | The use of exciting, entertaining and hidden audio content within the experience can help promote participation by others via the spectacle of exploratory enjoyment. |
| Preparation | Initial non-spatialised audio instructions can be used to help explain the experience to users in a familiar way. |
| Familiarisation | A transition to dynamic binaural spatialised instructions can help familiarise and explain to users how their interactions effect the delivery of the audio content. |
| Exploration | Exploration can be facilitated by sonically contextualised space and investigative sounds (see below). |
| Investigation | Unusual, unfamiliar and transient spatialised sounds can be used to promote investigation of objects. Visually imposing objects can be used to promote the discovery of more subtle audio content. |
| Focussed Listening | Investigative sounds can give way to more involved and prolonged audio content. |
| Second-level Focussed Listening | Micro sounds, along with sub and super-sonic audio material can be delivered to users upon closer inspection of objects. |

Table 3. Types of audio content. Promoting, facilitating and delivering the previously identified interactional phases of an audio augmented object reality with different types of audio content.

Table 3 indicates what variety of audio content would function best within each interactional phase, but also how this audio content can function to guide the listener through to the next interactional phase and promote further engagement with the experience and its physical and virtual content. This therefore outlines a model for attraction and engagement within the audio augmented object reality, a model that considers both engagement with virtual audio content, and engagement with physical objects within these experiences. The following sections describe in greater detail how specific types of audio content can aid the user's journey through these interactional phases:

> *Participation:* In section 4.4.3 we saw how the spectacle of currently participating users can help to promote interest in the experience to other, prospective users. By ensuring that the experience contains the entertaining facets of exploration, such as self-determined investigation, curiosity and reward, the enjoyment of current participants can be used promote the participation of others. This point should be considered within the construction of the virtual audio model and the physical environment, as the reflections in section 4.4.3 also describe how the display of *silenced-objects* (those that are recognisably noise-making though currently silent) can help to indicate a possible hidden layer of explorable content.
>
> *Preparation:* The creation of a highly intuitive user experience through the use of audio augmented objects is illustrated by participants' consistent and swift ability to familiarise and undertake the experiences presented in chapter 4. This is also illustrated by the curator's comments in section 5.4.1 that echo the experiences' naturalistic affordances, in that the constructed virtual acoustic reality remains explorable and perceptible in same way as an acoustic reality. Whilst the overall intuitive nature of the experiences is well documented, there remained occasions when participants required some explanation of what to expect, how it worked, or what they should do (see section 5.3.1). It is proposed that, prior to their engagement with the potential unfamiliarity of dynamic binaural audio, initial instructions in standard stereo audio could be provided.
>
> *Familiarisation:* By way of introduction to the interactional concept of virtual dynamic binaural audio, the above instructions can give way to instructive and demonstrational content rendered in dynamic binaural audio. Such an approach could fulfil the objectives of preparation, and provide familiarisation with dynamic binaural audio prior to the delivery of any experience-critical audio content. The audio presented within this phase can make use of the recorded lateral movements of participants as they familiarise themselves with the interactional properties of dynamic binaural audio as observed and described in section 5.3.2.
>
> *Exploration:* A key point in the process of engaging the user with object and audio, exploration can be facilitated through the inclusion of the previously described experiences of self-determined investigation, curiosity and reward. Whilst self-determined investigation can be ensured by maintaining the nomadic nature of the experience, curiosity

and reward can be achieved through the use of sonically contextualised space and *investigative sounds* and *investigative objects*. As outlined in section 4.4.3, sonically contextualised space serves not just to contextualise users' embodied interactions within the experience, but also enable the user to localise objects and audio content. This can be achieved by allowing spatialised binaural audio content to punctuate a layer of stereo rendered audio, providing an obvious indicator to the location of physical objects of interest and audio content that requires further investigation. Although exploratory freedom remains an integral part of these types of experiences (see sections 4.5 and 7.2.4) practically, for reasons of safety, or the provision of space for the interactional setting of the experience, exploratory boundaries may, and most probably will, need to be determined. As we have also learnt in section 4.4.3 , by setting maximum audible ranges for audio augmented objects (see section 7.5) and sonically contextualising space in manner that suggests the user is out-of-range, sonic boundaries that define the boundaries of exploration can be successfully determined.

*Investigation:* Part of the aforementioned phase of *exploration*, *investigation* can be encouraged within this phase in a number of ways. Firstly, there appears to be an agreement that the use of unusual, unfamiliar and transient sounds would be useful here for the purpose of attracting a listener's attention (see sections 5.5.1 and 7.2.1). As spatialised sound sources attached to physical locations and objects, such sound then become attractors or localisers to, and for, specific coordinates within three-dimensional space. Additionally, to aid this interactional phase of investigation, both the object-first and audio-first methods can be used to both draw attention to physical objects, locations and virtual audio content within the interactional setting (see Figure 31).

*Focussed Listening:* Within the interactional phase the initial *investigative*, or *attractor*, sounds can give way to more involved and prolonged audio content. Such content can be triggered based on a listener's proximity to the audio augmented object and a calculation of their current focus. By doing so, it may be possible to determine, to a certain extent, that the listener is indeed interested and ready to be engaged with such content based on the fact that they have been attracted to engage further due to being attracted by the visual presence of the object or it's *attractor sounds*. In short, we can, to some degree, calculate that there is some level of interest prior to dispensing with more involved and attentive content.

*Second-level Focussed Listening:* The interactional phase of *Second-level Focussed Listening* remains primarily a consequence of the physical intricacy of the audio augmented object and the magnification-type of interactions facilitated by a device-based tracking approach. For example, in relation to *The McMichael Experiment* (see section 4.3) one participant mentioned that they almost expected to hear *'more stations when pointing the phone at the tuning dial on the radio'*. These reports seem reflective of the findings of Montan (2002) in relation to the design of different 'acoustical zones' within the context of a single AAR subject for increasing immersion and engagement and a reported impression of entering into the subject, when moving from one zone to another. Such an approach is also consistent with the work of the artist Vicky Browne, where the elements of Browne's sound installation *Cosmic Noise* are described by curator and author Caleb Kelly (2019) as having *'micro-ecologies'*, where the work can be listened to as a whole, or attention can be focused on certain elements to reveal *'specific and often minute sounds'*. Providing a listener's attention to the subject can still be determined (see section 7.5) more subtle and intricate audio content can be delivered upon a listener's closer inspection of certain areas of the physical audio augmented object. Good examples of this would be sub or super-sonic audio content, or amplified audio subtleties of which would normally be barely audible to the human ear. Example content for these *microsonic* interactions could be biological or mechanical sounds that form a small part of a much larger audio augmented assemblage.

The sequence of interactions presented in Figure 28, along with the functional qualities of different types of audio content outlined in Table 3, provide a framework with which attraction and engagement can be designed within an audio augmented object reality. Using this framework one could design an experience that specifically exploits the concept of the *attractor sound*, as described by Zimmermann & Lorenz (2008) and harnesses the potential of *second-level of focused listening*, as described in the AAR project presented by Montan (2002) and Kelly (2017), in relation to the work of sound artist Vicky Browne.

### 7.2.2 Designing the alternate reality

Perhaps one of the most interesting considerations when designing an audio augmented reality is that the constructed acoustic virtual reality doesn't have to strictly adhere to the rules regarding the physical properties of an acoustic reality in order for it to be perceived as being real by the listener.

For the designer this is rather convenient, as it enables the construction of functional coherence within the experience whilst maintaining *the suspension of disbelief*. Within the work presented here, this is best illustrated by two factors. Firstly, the use of custom attenuation characteristics, and secondly the use of non-spatialised audio content within the *acoustic virtual reality*. Within *Alien Encounters* (see section 4.4) and *The McMichael Experiment* (see section 4.3) we see the former utilised to determine the range of the spatialised radio broadcasts, and the latter utilised to frame the audio experience in non-spatialised radio static. Within both of these installation experiences these factors are implemented with seemingly little consequence to the perception of reality, or rather, despite these unnatural acoustic properties, the audio augmented radios remain perceived as the source of the virtual audio content.

By presenting the audio content with these *unnatural* characteristics it is possible to author and control coherency within the experience by deciding exactly what can be heard and where. For example, if we return to the augmented radios once more, within a normal acoustic environment both radios would be audible within the acoustic space simultaneously. Within the audio augmented reality presented in *Alien Encounters* the range and angle of the virtual audio sources have been authored to include areas where only the individual radios can be heard, as well as areas where both radios can be heard.

Whilst this describes the useful separational function of an audio augmented reality for the presentation of multiple audio sources and audio augmented objects, it also highlights another interesting finding.

The recorded comments and observed reactions of participants within the studies indicate that this bending of the rules appears to have little, if any, detrimental effect on the perception of the reality of the audio augmented environment. This seems somewhat odd and remarkable having spent much time and effort constructing a convincing acoustic virtual reality through the use of responsive and dynamic binaural audio and the augmentation of sound making objects with authentic sounding content.

But it is, perhaps, precisely through this endeavour that a point has been reached where the physical laws of acoustics can be tweaked in the interests of functionality whilst still maintained a perception of reality. In other words, there is enough *environmental, cultural* and *interactional authenticity* present within the acoustic virtual reality to allow certain rules to be broken whilst still maintaining an *unwilling suspension of disbelief.*

If we return briefly to the concepts of *environmental* and *cultural authenticity* discussed in section 7.1.3, we can see how both these components of the experience don’t just ground the fantastical as being perceived as real, but they enable additional functionality to be authored into the experience. We can therefore propose a link between *environmental* and *cultural authenticity* present within an audio augmented reality experience and the potential within that experience to manipulate the acoustic virtual reality without detriment to the perceived reality for the listener.

This allows for the creation of an acoustic hyper-reality, an acoustic virtual reality capable of being indistinguishable from reality but imbued with additional functionality beyond that which is possible within the scope of the physical properties of acoustics. To what extent the audio augmented object reality can be manipulated and still maintain the perception of reality remains a topic for further study.

### 7.2.3 A functional atmosphere

Within an experience containing multiple audio augmented objects we can begin to map the components within both the virtual and physical realities that help create a functional atmosphere; an atmosphere that can help to convey a sense of time and place, evoke memory and provide a conducive environment within which a more detailed and attentive form of engagement can happen. Furthermore, we can determine which elements can help us assign context and environmental and cultural authenticity, thus allowing us introduce desired functionality into the experience.

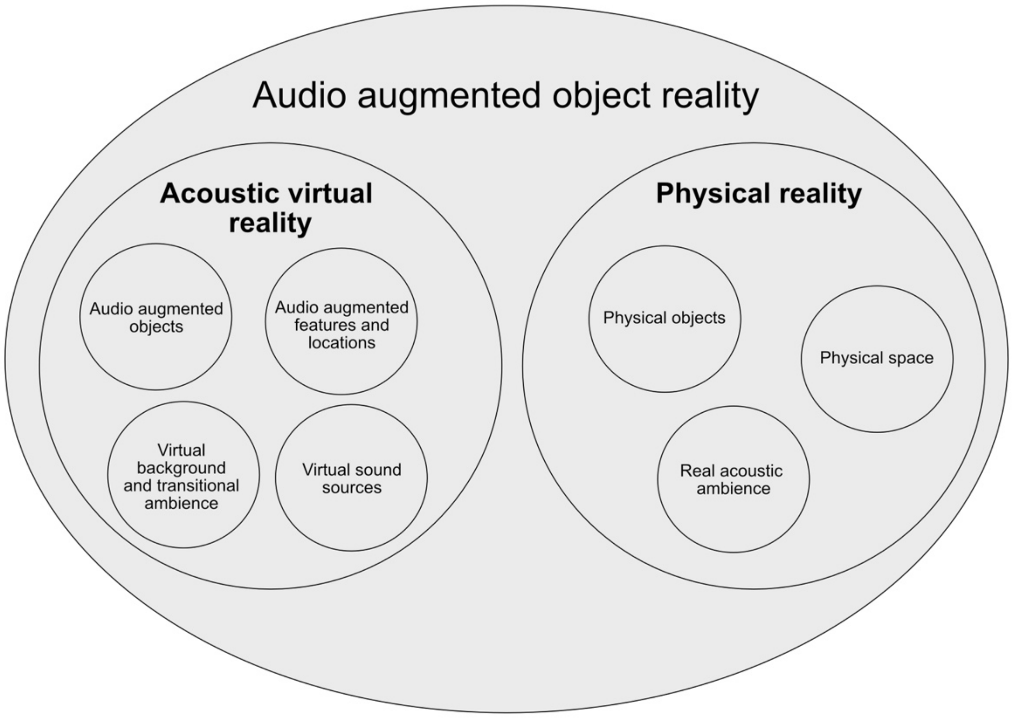


Figure 32. An audio augmented object reality and its potential components.

Figure 32 shows the component parts which can make up an audio augmented object reality. Whilst all the following components of an audio augmented object-based reality are responsible for the creation of an experience's atmosphere and context, and provide potential contributions to *environmental* and *cultural authenticity* they are also capable of providing the following additional functions within an audio augmented object-based experience:

The functional qualities of the acoustic virtual reality:

*Audio augmented objects:* Collections of audio augmented objects add to the sonic atmosphere though also function as attractors or locators to the real physical object within an *audio-first* approach (as illustrated in Figure 31) and are capable of contextualising and changing the perception of the physical object itself. Finally, audio augmented objects can provide environmental authenticity for *virtual sound sources.*

*Audio augmented features and locations:* Audio augmented features, locations and objects that are of not of specific interest can contextualise the physical reality, provide initial and ongoing familiarity with the interface and function as ongoing *contextual anchors*. *Contextual anchors* can be thought of as consistent contextual reference points amongst other content (such as the directly augmented objects) that may change. A good example is the audio augmented windows in the *Horror-Fi Me* experience (see section 4.5), whilst other content may change, or come and go, the wind and the rain sounds remain consistent, providing an adhering and continual context for the experience. These augmented features and locations can also act as waypoints, guiding listeners towards specific points of interest, or along specific trajectories. Finally, audio augmented features and locations can also provide environmental authenticity for *virtual sound sources.*

*Virtual background or transitional ambience:* Virtual background or transitional ambience can take the form of music or audio artefacts that help to contextually frame the experience. The stereophonic or monophonic rendering of this content can assist in the localisation of the spatialised audio content, and therefore also the audio augmented physical objects themselves. The spatial interactivity involved with this type of content and the individual audio augmented objects, creates a contextualisation of the mode of interaction, which can assist in the creation of an affectual and immersive experience (see section 6.2.5). Furthermore, it's possible that this direct and embodied interaction with contextually appropriate audio content can facilitate a greater level of engagement and meaning with the experience via embodied cognition (see section 7.2.4).

*Virtual sound sources:* Virtual sound sources that are not attached to physical objects or features but have a perceived location in physical space can be used to provide additional context and cultural and environmental authenticity. Such sound sources are spatially positioned beyond the boundary of the experience so that they can be heard but their potential source can never be visually verified by the listener.

The functional qualities of the physical reality:

*Physical objects:* Physical objects can act as attractor objects for virtual audio content within an *object-first* approach (as illustrated in Figure 31). They can also provide interfaces to, and context for, audio archival content. Importantly, physical objects within the experience the potential to provide environmental authenticity by virtue of their relationship with the virtual audio content with which they are augmented, and grounding this virtual content in reality. In section 5.2 we also see how the physical object can assist in the creation of an experience with personal meaning through the evocation of memory.

*Physical space:* In addition to providing context, the physical space of the audio augmented object reality determines the size of the physically and virtually explorative space.

*Real acoustic ambience*: The real acoustic ambience of the experience's location can be utilised as a part of the experience to enable verbal social interaction between participants (see section 7.3.3) as well as to provide additional environmental authenticity.

In summary, all components of an audio augmented object reality are capable of contributing to a sense of *cultural* and *environmental authenticity*, though they each have some specific functionality of their own. Whilst both *cultural* and *environmental authenticity* can help suspend disbelief, other functional attributes can help to manipulate exploration of the interactional setting of the experience, reframe the perception of objects and space and maintain a sense of context. Other functional attributes include the ability to provide coherent and effective localistion of spatialised content (both physical and virtual) and to provide opportunities for deeper levels of engagement with subject matter. The audio augmentation of physical objects can help provide interfaces to engagement with physical objects and provide them with context. Furthermore, audio augmentation can provide context for inanimate objects and an interface to audio archival material, whilst the audio augmentation of space by means of virtual background and transitional ambience can define boundaries to the interactive setting.

### 7.2.4 Manufacturing serendipity

Throughout the description and analysis of their deployment of the *Exaudimus* system, Mortensen and Vestergaard (2014) iterate that their interest lies in the creation of serendipitous moments of engagement, rather than assisting the listener in the collection of declarative knowledge on the subject matter. As maintained by composer Barry Truax (2012) and Mortensen and Vestergaard (2014), such serendipitous encounters have the ability to realise engaging cultural experiences and have the potential to extend interest in the installation's subject matter beyond the duration of the experience. Such serendipitous and explorative expeditions could be likened to Debord's philosophical theory of the dérive (Debord, 1958); a détournement, where one is concerned with potential points of departure, rather than a specific destination. Mortensen and Vestergaard (2014) make reference to this type of take-away chance encounter or, recontextualisation of the familiar or seemingly mundane, reflecting that it acts as a catalyst for extended engagement.

Within the quotes of participants involved in the study of *The McMichael Experiment* (see section 4.3), as within the study conducted by Mortensen and Vestergaard (2014), we see evidence of the role personal memory plays in realising these moments of serendipity. Additionally, we see evidence that suggests how virtual sound sources, when combined with physical artefacts, have an ability to stimulate the imagination and realise these moments of engaged exploration.

Truax (2012) explicitly attributes this phenomenon to the ability of sound to create relationships between listeners and their environment, combined with a relationship between embodied interaction and embodied cognition, the idea that bodily movement influences our process of acquiring knowledge and understanding. Furthermore, with regard to this audio-object relationship, we can perhaps look towards Schafer's work on the soundscape and its composite elements (Schafer, 1977), where sounds are prescribed with the ability to indicate age and reflect the state of society within which they were conceived. Such attribution gives the object the power to speak to, and engage with the user beyond the immediate scope of the audio content with which it has been augmented.

Truax's suggestion (Truax, 2012) invokes Bull's (2000) observations within the field of sound studies on the use of personal portable audio systems. Users have been augmenting their environments for decades with such devices, and Bull's observations point towards the importance of nomadic agency within the system, where listeners remain free to explore their own relationships between virtual sound, the physical environment and its contents. Though, evidently, we should not dismiss the ability of serendipitous experiences to increase engagement, awareness and understanding of subject matter on their own, our identified phases of focused listening, also observed by Montan (2002), offer opportunities to create moments within the experience when declarative knowledge could be imparted.

Mortensen and Vestergaard (2014) suggest that within immersive environments such as these learning outcomes are not facts, rather experiences, feelings and memories. But we can, perhaps, have our cake, and eat it. By initially engaging listeners with chance serendipitous encounters through exploratory freedom and association, we can draw them into phases of focused listening during which declarative knowledge can be imparted.

## 7.3 Within the gallery

The practice detailed in sections 4.3 and 4.4 and study detailed in chapter 5 underscore the appropriateness of the museum or gallery as context for the deployment of audio augmented object-based experiences. Whilst the previously discussed design implications and functional attributes of audio augmented objects can be seen as largely transferable across various contexts and types of experience, it is worth considering the function and implications for these experiences specifically within the context of the gallery.

Within the gallery, a system that relies on the modality of non-tactile, embodied spatial interaction overcomes the problems encountered by Mortensen & Vestergaard relating to their reliance on physical interaction with objects within such an environment (Mortensen and Vestergaard, 2014). Within the gallery such an approach also enables the explicit authoring and design of *attractor sounds* (Zimmermann and Lorenz, 2008) as objects need not be physically, or even visually, encountered prior to their presence being made aware to users.

The suitability of audio augmented object-based experiences for deployment within galleries is reflected by the positive feedback attributed to *The Mingus Demonstration* (see section 4.2) and *Alien Encounters* (see section 4.4) by curatorial professionals. This suitability is also reflected by the observed potential of audio augmented objects to engage audiences with both physical objects and virtual audio content, including digital audio archival content (see section 5.5).

The thoughts participants shared regarding the potential uses and applications for this type of technology within a museum, gallery or heritage site, along with their interest in undertaking such an experience, suggests that an experience like Horror-Fi Me would be well received as part of an exhibition. The findings also indicate that the convenience and agency offered museum and gallery attendees through the utilization of their own technology to undertake such an experience is desirable. Specifically, we find participants welcoming the exploratory freedom afforded to them by the model of spatial interaction inherent within the experience as a refreshing change to the restrictive nature of the more widely available audio guide. The enjoyable and performative interactional experiences that this exploratory and spatial freedom facilitates is also worth considering within such a context.

Of particular curatorial and artistic interest is how the participants' experiences demonstrate the ability of an audio augmented object to communicate its presence, location and character beyond line-of-sight. An audio augmented object's ability to do this can be used to affect the listener's exploration of physical and audible space, by both inviting and deterring exploration. Once seen, both the object's existence and characteristics can remain present for the listener throughout the entirety of the experience regardless of the object's audibility within the soundscape. Furthermore, this sense of presence, along with the characteristics with which it is associated, could possibly extend beyond the scope of the experience. Whilst we can see that the presence of something can be suggested through audio, it's the confirmatory visual association of the virtual audio's source in the guise of a physical object that makes the suggestion believable within the context of the experience, and that this association has a fluidity and malleability that can be creatively exploited.

### 7.3.1 An absence of experience

Whilst much emphasis is placed upon the visitor experience by the curator in section 5.4, *silenced* objects present as objects with which the experience has been lost; the radio that can no longer be tuned, or the musical instrument that can no longer be played. Within the gallery or museum, such objects present themselves as *silenced* for a number of reasons, these include: health and safety, preservation, technical obsolescence and the cost and practicalities of the ongoing maintenance of antiquated technologies. Additionally, as the earlier movie projector example indicates (see Figure 29), such objects should not be limited to the media playing or musical type, nor are the sounds of media playing or musical objects limited only to the media content or music that they are capable of playing. Indeed, it is perhaps through a thorough investigation of all possible sounds associated with all sorts of objects that the unusual and the unexpected content can be uncovered that the curator also suggests can help entice initial engagement with objects.

Virtual audio, as realised through a museum-based audio augmented reality experience, presents itself as an antidote to this dilemma. It provides a way for visitors to experience objects that cannot be picked up and played, turned on, tuned in and turned up. Of course, whilst not being able to interact with these objects in the usual tactile way, new ways of remotely interacting with them need to be devised but, by thinking carefully about how this can be achieved, not only can an experience with a silenced or silent object be realised, but an experience with the object's associated audio content can also be attained. Additionally, such interactions have the capacity to help tell the stories and experiences behind the objects, and connect them to relevant and appropriate audio archival content. Again, the importance of both these factors are clearly expressed in the excerpts from the curatorial interview in section; in essence, that which matters is *the experience*, and it is *the experience* that can again be made possible, albeit in a new way.

By including, or at least considering, all sounds associated with an object as worthy contenders for an object's audio augmentation, it should be possible to curate virtual soundscapes around collections of audio augmented objects that go some way to capturing the experiences and atmospheres associated with specific objects, places and moments in time.

### 7.3.2 Atmosphere as communicative gestalt

The creation of such a holistic and interpretive experience would speak strongly to the practices of soundscape studies, acoustemology and sonic ethnography (Droumeva, 2016); where the study of a setting's soundscape is attributed with a potential to communicate an in-depth understanding of time and place, more so than the sum of its component sound sources and, at times, to greater effect than perhaps more traditional methods of documentation and enquiry (Lingold, Mueller and Trettien, 2018). Though, perhaps a more considered definition is provided by Gershon (Gershon, 2019), who suggests the epistemological potential of sound, as realised within the practical application of sonic ethnography, lies in its ability to promote an understanding of things from a different, non-ocular point-of-view.

The creation of a thought-provoking sonic atmosphere for exhibiting collections of objects was, to a small degree, realised by *The Mingus demonstration* (see section 4.2), and seems entirely possible to achieve on a larger scale. The curatorial interest expressed in this functionality, along with that of an intuitive and easy-to-use virtual environment, has the potential to be achieved with collections of audio augmented objects which can contribute to an over-arching and communicative sonic atmosphere, as well as affording more detailed investigation of their individual audio content. Such an intuitively explorable sonic atmosphere also creates an interface through which visitors can, to quote the curator: '*create their own experience out of what* [is presented] *to them'.*

The audio content of the V&A's *The Future Starts Here* exhibition also provides us with a great example of an ambient soundscape working as an anthropological or sonic ethnographical document. Just as the curator anticipated, the '*stereotypical hubbub of modernity'* was intentionally crafted in an attempt to place listeners within futures which they had the *conceptual tools* to imagine for themselves amongst the complexity of possible connections between the exhibited technologies. Superflux, the team responsible for the video content that provided much of the audible content for *The Future Starts Here* exhibition, remark:

> *"...the future is given a near-ethnographic and anthropological quality, allowing people to easily place themselves within these futures… One of the key objectives of our work was to comprehend the complex connections between the objects and technologies in the show, and their wider societal, cultural and environmental implications across multiple futures… We wanted visitors to consider the future not as a destination where technologies will suddenly empower them, but as a complex, messy and plural process, a space of diverse potential, and an encounter with otherness… we wanted to hand the audience the conceptual tools to form unique perspectives, engage with diverse futures, and*

*understand that they have agency to influence the future(s)."* (Superflux, 2017)

These remarks regarding the curatorial intentions and decision making involved in the creation of *The Future Starts Here's* atmosphere echo the curator's interest in the ethnographical quality of the exhibition, they also speak to the holistic and interpretive qualities of sonic-ethnography present within a soundscape-based approach to exhibiting sound and the sound of objects. Additionally, they highlight the importance of visitor agency, which again drives home the appropriateness and the possibilities associated with a freely explorable, nomadic soundscape-based approach to the sonic curation of audio augmented objects.


Figure 33. *The Future Starts Here (2018).* An exhibition at the Victoria and Albert Museum, London.

> Curator: *"…they all worked individually as you approached the exhibit and you heard whatever you were supposed to hear about that particular part of it, but the whole thing in the gallery as a whole blended into this almost stereotypical hubbub of modernity.. and it was really, really effective."*

On the point of visitor agency, we can return to Bull's (2013) suggestion that the connectivity to the environment afforded to the listener via the ability to freely associate virtual audio and physical reality differs from that which can be achieved via directed association (narrative, historical or otherwise). Although Bull (2013) insists that this experience is a product of the nature of the technology, the environment and '*the user's cognitive orientation*', the *differing* that the agency of association represents could be one that has the potential to facilitate a greater level of connection with the subject.

Comparing the speaker-based listening experience of *The Future Starts Here* with an imagined headphone-based binaural audio version of the same exhibition is an interesting exercise. Being aware of the functions of specific curatorial interest in exhibiting sound and using sound as a means to exhibit, brings into sharp focus exactly what a headphone-based binaural audio solution has to offer.

Indeed, with a technological solution that presents itself as being such a close simulation of reality, it suggests that one should remain further focused upon what exactly is has to offer. Exactly what puts the hyper into the reality in this case? What can the virtual, binaural headphone-based soundscape experience offer over the unmediated natural exploration of a physical soundscape comprised of multiple sound sources, speaker-based or otherwise?

Whilst the listening experience within *The Future Starts Here* was a speaker-based one, a headphone-based binaural listening experience provides an opportunity for further sonic curatorial control and a level of personalisation of the sound experience not possible with a distributed speaker-based solution. With personal headphone audio delivery, the problem of timed and looped audio events is overcome by ensuring that, for example, narrative-based content always begins when the visitor arrives, rather than being half way through its timed or looped cycle. Additionally, different layers of sonic engagement can be delivered simultaneously to different visitors, whilst one visitor roams the ambient soundscape, another can be engaged with detailed narrative content around a specific exhibit. Although, as the curator points out, this latter function is possible within a distributed, and carefully curated, speaker-based system, the greater degree to which the soundscape can be curated in audio augmented reality eliminates the problems of any unwanted bleeding and over-lapping of sources.

These advantages should be considered not just in conjunction with the practical and economic advantages of deploying a mobile-based audio augmented reality solution, but also in relation to the creation of a sonic atmosphere that can help communicate understanding of a subject, connect the listener to the subject and gallery environment and function as a catalyst for knowledge exchange.

### 7.3.3 Headphones, Hearables and hear-throughs

Although the use of headphones in public spaces is widely accepted, within the public space of the gallery or museum environment it forms a barrier between visitors, hindering any potential shared and social experience of attending an exhibition. As is detailed within section 5.5.6, groups of visitors wanting to share or describe audio content they have found to other members of their group may find this difficult to do with whilst using headphones.

These issues present as problematic for the realisation of audio augmented objects with such contexts as the effective delivery of virtual binaural audio content, and therefore the effective realisation of audio augmented objects, relies on the delivery of this binaurally rendered audio content to users through headphones (Bauer *et al.*, 2019; Lawton, Cunningham and Convery, 2020).

We should consider the fact that the binaural audio experience is a natural phenonium, we hear binaurally, so we therefore need to consider what the headphone-based experience offers over a distributed speaker-based experience within the gallery environment in order to both justify its deployment and consider how these issues could be resolved.

Just we have heard within the curatorial feedback that the deployment of a headphone-based audio experience within galleries presents problems, we have also heard how it also resolves some existing problems related to distributed speaker-based solutions and also provides new and exciting opportunities for curatorial practice.

We could, as illustrated by Lawton et al. (2020) realise audio augmented objects through a distributed network of loud speakers embedded, for example, underneath or within the objects that we want to augment, and then rely on our natural sense of binaural hearing to navigate, find and explore such objects in a similar way we have been doing with the headphone-based approach presented within this thesis.

But to do so would compromise many of the desirable characters that the headphone-based approach facilitates. As the curator in section 5.4.2 states that with the headphone-based approach the problems associated with timed and looped audio events are overcome; the triggering of audio content is personalised, meaning the other users are not encountering critical content halfway through that has been activated by other visitors. Additionally, the greater level of curation of the sonic curation is afforded by the headphone-based approach, meaning that the natural characteristics and physical properties of sound in space can be manipulated (such as the attenuation of a sound source's volume over distance) allowing for the precise authoring of the sonic experience.

In addition to these positive factors, we have also seen how, within the studies presented here, how the headphone-based experience has enabled the personalised and self-composed user experience that has resulted in the realised of a more meaningful experience for users.

Given the recognised affordances, benefits and curatorial opportunities specific to the headphone-based delivery, it would seem only appropriate to spare some thought on how these few, though not unimportant, issues could be potentially resolved.

Firstly, it is perhaps worth considering the social problems associated with the use of headphones. Though it is impossible to deny that the use of over-ear headphones that cancel out the exterior acoustic ambience in order to concentrate the user's attention on the delivery of a virtual acoustic reality isolate the user from the acoustic reality itself, it is arguable that this type of listening experience can lead to a more general sense of *belonging*.

Bull (2000) refers to the mobile audio listening experience as being, on the one hand, socially isolating, and on the other, socially adhering. In other words, the headphone-based mobile audio listening at once *accompanies* listeners as well as *isolates* them. Bull (2000) draws upon Adorno and Eisler's cultural theoretical concept of the construction of a state of *'we-ness'* within the musical listening experience, achieved through a constructed sound world, a type of mediated company (Adorno and Eisler, 1994). This, as Bull (2000) reflects in relation to the personal-stereo experience, results in an experience that is both *isolated* from the immediate reality, though *accompanied* through the realisation of a state of *being with* and a *part of* a shared cultural sphere of identity and production.

In relation to this shared experience of production and consumption it is perhaps worth reflecting again upon the observed user interactions with *The McMichael Experiment* (see section 4.3) where we saw, despite the presence of traditional headphones, participants demonstrating physical evidence of a mutual, enjoyable and shared sense of both being within the installation together, along with a mutual acknowledgement and acceptance of this state within the shared experience.

Figure 34. Participants of the *The McMichael Experiment.* Acknowledging each other, communicating visually and sharing enjoyment of specific parts of the installation whilst using traditional headphones.

Moreover, by considering the *shared* and *accompanied* experience as being a consequence of both *mutual production* and *mutual consumption,* underlines the importance of the free exploratory nature of the experience. As previously discussed, it is the freedom of exploration that enables the user's agency over the composition of their audio experience.

Although participants may feel that they are *in it together* and part of a communal experience, the fact that the listener is isolated from their immediate reality remains and effects their ability to communicate with others around them. As the curator observes in section 5.4, sound remains a particularly difficult medium to describe after the event, and the ability to truly communicate and share excitement and enjoyment over specific pieces of sonic content, beyond a smile and a nod, remains encumbered by the use of traditional headphones.

The recent development of *Active Noise Cancellation* (ANC) technologies, along with the consideration of headphones as *hearable* interfaces (Plazak and Kersten-Oertel, 2018), rather than passive listening devices, raises an opportunity to resolve this issue. An increasingly prevalent feature in new, consumer-level headphones (Kiss, Mayer and Schwind, 2020), ANC enabled headphones have embedded microphones that enable them to listen to and cancel out external sound sources. These microphones can also be used to enable a transparent mode of listening that, instead of cancelling out external audio sources, can be used to mix these external audio sources with the virtual audio content delivered through the headphones themselves. Through the inclusion of a *hear-through* listening technology into the audio augmented object reality, one can envisage an experience were one can both appreciate the virtual acoustic reality and retain the ability to verbally communicate with friends and family.

Additionally, Plazak & Kersten-Oertel (2018) suggest that the capabilities of *hearables* extend well beyond two-way audio transparency. Citing the inclusion of sensors, including the use of bio-sensors, within these products, they suggest that additional features could include interfacing with AI assistants to negotiate stressful situations and connect users to appropriate content based on bio-feedback, as well as other real-time feedback of the user condition to improve human-computer interfacing.

This observation by Plazak & Kersten-Oertel (2018) is somewhat confirmed by both *Bose's* AR enabled headphones and *Apple's AirPod Pro* and *AirPod Max* model headphones that, through the sensors embedded within these models, enable head-tracking, in addition to transparent listening modes.

Plazak & Kersten-Oertel (2018) suggest that one of the key affordances of the *hearable* is its ability to allow the user to attend to unimpeded visual interaction with the real-world and blend into everyday environments in a uniquely seamless manner that reflects ubiquitous computing's goal of a perceptively transparent human-computer interface (Weiser, 1999).

Furthermore, experiencing the dynamic binaural acoustic virtual reality through an effective *hear-through* headphone-based interface could add to a heightened sense of *environmental authenticity*. A consideration that draws upon the previously discussed idea that to embed the virtual into recognised elements of reality blurs the distinction between them and enables a greater grounding of the fantastical as being indistinguishable from reality.

I would therefore propose that an effective hear-through, headphone-based interface could deliver the recognised benefits of a traditional headphone delivered dynamic binaural acoustic virtual reality with audio augmented objects, and resolve the problems relating to sharing and sociability within public space, and also further develop the sophistication of the audio augmented object-based reality.

It should be noted that there may be instances when a speaker-based AAR system remains referable to a headphone-based one. Based on the findings from the curatorial interview in section 5.4 and the subsequent discussion presented here, we could determine that if fine-grained curatorial control over the audio content is not a priority and the creation of an over-arching sonic atmosphere is of primary concern, then a speaker-based option may be preferable. Such an option should perhaps also be given further consideration if the exhibited content is targeted at families with small children, or if blending the virtual content seamlessly into reality without the requirement of any physical interface is a desirable option. Such a decision should though carefully consider that the audio augmentation would be one in which an AAR is realised through *modification*, *addition* or *reinforcement*, rather than through *replacement* (Krzyzaniak, Frohlich and Jackson, 2019).

### 7.3.4 Curating with audio augmented objects

Within the gallery both the object-first and audio-first approaches illustrated in Figure 31 can be put to work. One can see how this could also be used to curate and design visitor journeys through an exhibition, or a collection of artefacts, by triggering sound sources at certain times in certain locations, or in relation to other objects. Also possible would be advertising the location of other objects in relation to the one you are currently viewing, associated objects that work well together sonically as well was contextually, guiding and suggesting potential trajectories to the listener. Such approaches build upon Zimmerman & Lorenz's concept of the attractor sound (Zimmermann and Lorenz, 2008), a feature of the LISTEN system that recommends additional artworks, via emitted and localized virtual sound sources, to users of the system based on adaptive and personalised recommendations. Giving objects within these cultural spaces the ability to communicate to visitors beyond line-of-sight has the ability to provide great potential, and significant challenges, for the designers and curators of such spaces. Spaces where the visual has maintained primacy from architectural design through to curatorial decision making for centuries (Connor, 2011), and constitutes a new design space and way of seeing within such environments.

Additionally, regarding the authorship and curation of trajectories through an exhibition space, such an approach places the object in the role of both waypoint and content, making an object potentially a functional, and a thematic, part of the system. This is considered in relation to Hazzard et al's definition of an audio event that *'tells, or supports the telling of the narrative'* as being thematic, and an audio event that *'supports participants in their navigation or comprehension of the experience'* as being a functional one (Hazzard *et al.*, 2017). In addition to being functional instruments that guide and suggest exploratory trajectories through exhibition space, the object-first and audio-first approaches can be interchanged within the context of a single exhibition depending on what curators want people to attend to; either sound or object within different parts of exhibition.

### 7.3.5 Interfacing with history

The recent digitisation of existing analogue audio archive content has both highlighted the need for new methods to promote exploration and engagement with this type of audio content, as well as providing the potential opportunity to do so (Mortensen and Vestergaard, 2014; British-Library, 2019).

With millions of physical recordings now digitised, indexed and tagged with descriptive meta data (BBC, 2022; The-British-Library-Board, 2022) opportunities for extending these recordings' ability to engage with a wider public through the use of innovative and creative digital-based solutions, with a few exceptions (Cliffe *et al.*, 2019, 2020) remain, for the most part, unexplored.

With the radio-based installations presented here, this is achieved by using the body like a tuning dial on an analogue radio set, allowing the visitor to find clear signals of archival content amongst the sound of static. Bijsterveld (2015), in relation to Mortensen and Vestergaard's similar approach (Mortensen and Vestergaard, 2014) describes this as a *'highly original framing of the exhibition sounds'* and one where '*the exhibition space itself mimicked the technology behind the sounds that were the topic of the exhibition'*. One could also argue that this act of embodied, interactive tuning constitutes a physical contextualisation of the virtual digital archive content.

Within Mortensen and Vestergaard's installation this physical contextualisation is extended through the construction of listening situations, where the settings of the original physical listening environment associated with the different pieces of audio content (an armchair for content programmed in the evening, a car seat for drivetime content and a bedroom for teenage content) are reconstructed within the gallery space. Again, we see an exploration into how the material can be used to promote and focus engagement with the immaterial, a mixed reality exercise in the contextualisation of the virtual with the physical.

Mortensen and Vestergaard's approach is largely justified by an understanding that learning associated with immersion is experience driven (Mortensen and Vestergaard, 2014). As such, the authors anticipated visitor learning outcomes to include experiencing situations, feelings and memories, not hard facts. As with the installations presented here, their approach bears fruit in the form of positive participant feedback in relation to awareness, interest, engagement and the evocation of memories associated with the audio archive content used within the exhibition.

Understanding a little more about the dual nature of the interactional experience and character of the audio augmented object leads to some interesting considerations regarding their specific and potential function as interfaces to audio archival content. What appears to be of primary significance here is that, by augmenting a historical museum artefact with contemporaneous audio archival content, a mutual and functional contextualisation of both these components occurs; the artefact functions as an interface for the audio and provides it with physical presence and historical context, and the audio functions as an interface for the object and provides it with additional physical presence and historical context. Furthermore, this factor presents itself as a way in which obsolete and inanimate museum objects can once again become *experiential objects,* as well as interfaces to audio archival content.

## 7.4 Beyond the gallery

The successful authoring of an audio augmented object-based experience and the successful undertaking of this experience by participants within the study of *Horror-Fi Me* (see chapter 6) illustrate the experience's suitability beyond the gallery. Furthermore, we have heard how audio augmented objects have the potential to realise new musical and audio listening experiences (see section 6.3.5). In this section the application of audio augmented object-based experiences for the purposes of creating new domestic listening experiences is considered.

### 7.4.1 Towards a dynamic binaural music

The composer's comments in section 6.3 support dynamic binaural audio augmented reality's consideration as a new listening experience, and as the next spatial iteration of audio delivery and consumption that does much to vindicate the AAR continuum presented in section 2.1.1.

Evidence suggests that the dynamic binaural audio augmented reality represents the realisation of a new domestic listening experience that transcends the previous incarnations of the audio augmented realities outlined within the presented AAR continuum in section 2.1.1 and that offers a new way to associate music with, and move music through, space.

Such a *Dynamic Binaural Music* has the potential to further advance the site-specificity of GPS enabled location-based musical listening experiences. Enabled by the increased positional precision of SLAM-based augmented reality, it is possible that music could be scored for a specific location that responds to the listeners subtle interactions with architectural space, its features, fixtures and fittings and the objects contained within it.

Within the interview with *Horror-Fi Me*'s composer we learn how the domestic musical listening experience can be afforded a new sense of movement and spatial belonging through the introduction of audio augmented objects and features within the physical environment. Furthermore, we see how this is achieved, within this instance, using recognised classical contemporary compositional techniques, and how this could be harnessed in future compositional practice and the realisation of a *dynamic binaural music*.

### 7.4.2 Towards a new musical acoustic virtual reality

The musical exhibition example described in section 5.4.3 offers a different, though equally interesting, binaural musical listening experience. In the given example, a musical listening experience is described where the different instruments present within the musical piece exist at specific locations in physical space. Within this, a listening experience is described where, depending on the listener's proximity to these physical locations, they may be able to hear the whole ensemble or pick-out the individual instruments within it. Whilst the previous example of a *Dynamic Binaural Music* described a music that could be specifically scored to take advantage of the spatial interactivity present within the AAR experience, this example describes a way in which existing recorded music could be afforded a new sense of virtual acoustic reality.

Applied to the delivery and consumption of recorded music within the domestic setting, by positioning the individual instruments from a single piece of music within the physical reality, one can envisage a listening experience where you could wonder between *The Beatles* during a famous Abbey Road recording session, or place the members of John Coltrane's famous jazz quartet in different corners of your living room. Appropriately, *MagicBeans,* a company who specialise in the design of interactive three-dimensional soundscapes, describe this as *6DoF Music*; a re-envisaging of the musical listening experience as something that you can walk through and even superimpose onto real physical objects (MagicBeans, 2022).

Additionally, it would seem that the realisation a coherent and responsive musical listening experience within such a context would rely on having existing recordings with good separation between instrument tracks, whilst new recordings could be produced, and possibly composed, with this application in mind.

The placing of recorded music into the physical reality has the potential to a realise a greater sense of *presence*. Whilst the experience of an *acoustic virtual* reality is often described as a feeling of *being there* (Nicol, 2020), this particular manifestation of an *acoustic virtual* reality would seem to be more about it *being here;* the bringing into the physical reality of the music, rather than the taking of the listener into the music, a change in perception, researcher and acoustician Rozenn Nicol (2020) suggests, is a much harder illusion to perform.

Presently, and over recent years, we have seen an emergence of interest in the development of the perception of space within the musical acoustic virtual reality (Apple, 2021b; Dolby, 2022). Indeed, if we return to Figure 4, in which I outline the pursuit of acoustic virtual reality within the evolution of audio listening and recording technologies, we see that this trajectory remains chiefly concerned with the creation of *space in sound*, rather than *sound in space*. Whilst recent developments in *acoustic virtual reality* include and appreciate the importance of the binaural listening experience in relation to a heightened sense of *acoustic virtual reality* (Apple, 2021b; Dolby, 2022), they remain chiefly concerned with the perception of *being there*.

The departure from '*space in music'* to '*music in space'* represents a break from the current trajectory of *acoustic virtual reality* previously outlined in Figure 4. Such an approach appears to have more in common with *Musique Concrète* and early pioneering sound arts, such as the spatially dispersed speaker arrays of Edgard Varèse's *Poème électronique* within Le Corbusier's 1958 *Philips* World Fair Pavilion (Ouzounian, 2020).

The difference the inclusion of audio augmented space or objects creates for the recorded musical listening experience is the difference between '*space in music'* and '*music in space'*, the latter realising a perceived experience, as Nicol (2020) puts it, of *it being here.*

### 7.4.3 New audio dramas

Just as the acousmatic musical listening experience can be transformed to a more direct experience through the inclusion of audio augmented objects, so can the narrated audio listening experience.

By including audio augmented objects within the context of an audio-based drama it would be possible to bring the story into the listener's reality; the imagined *there* becomes the experienced *here.*

Including audio augmented objects within this specific type of listening experience further transforms the *dummy head* approach utilised by

Strickland (2015) and Sachs (1978) (as discussed in section 2.4.1) by both making the audio experience the listener's own, and placing it within their physical reality.

Furthermore, just as static binaural audio has provided *Audio horror* listeners with a heightened experience over stereophonic audio (Jahn, 2015; Hancock, 2018), dynamic binaural audio with an audio augment physicality offers a potential route to Hancock's vision of an audio experience that can rival that of video gaming and cinema (Hancock, 2018).

Even within the stereophonic listening experience we are familiar with the concept of being transported to another place but, as Nicol (2020) points out, the illusion of *'they are there'* is stronger and consequently more complicated to achieve than the illusion of *'you are there'*. With audio augmented objects situated within a dynamic binaural audio experience narrative events and situations can be experienced within a convincing *virtual acoustic reality.*

We can perhaps imagine the resultant experience of this particular application of an audio augmented object reality as being a combination of the *Horror-Fi Me* experience (see section 4.5) and *The Darkest Night* (Shudder, 2021). For it is through *dynamic binaural audio* and *audio augmented objects* that the virtual knock at the door has the potential to be perceived as being indistinguishable from reality, and can be embedded into the fabric of an interactive and immersive fictional narrative.

## 7.5 Realising audio augmented object realities

Within the final section of this chapter, the practical ways in which an audio augmented object reality can be created are discussed. This is largely considered from the perspective of enabling the provision of authorship to third-party *curators of sound,* that is, enabling those who wish to make such experiences make them.

These discussions are the result of the practical efforts of attempting to bolt together various frameworks and software development kits in order to achieve an audio augmented object reality. They are also the result of the practical deployment, study, curatorial feedback, interviews and the involvement in early-stage exhibition development workshops where practical needs, opportunities and problems have provided interest and potential platforms for such an approach.

### 7.5.1 Placing virtual sound sources

The borrowed process of tagging real-world objects in order to specify the required position of virtual audio content to the system appears to transfer well from visually orientated virtual sticky note AR applications such as *Placenote* (Placenote, 2021). Furthermore, the project has demonstrated that the developed system is capable of enabling novice users to author AAR environments in arbitrary locations (see section 6.4).

Additionally, whilst there appears to be a significantly greater margin of error with audio AR as opposed to visually orientated AR in relation to the user's perception of the system's ability to keep virtual content in position, the initial and prolonged accuracy of the tagged locations remains proportional to the user's perception of a convincing virtual reality. As such, any further development should always consider ways in which this can be, as a minimum, maintained, and, if at all possible, enhanced. Although the system presented in Figure 10 performed well overall, there were times when virtual content drifted from its authored position. This was usually the result of low-lighting conditions, and virtual audio content that is not correctly positioned at the same location as the physical object cannot be perceived as being emitted by that object. The exact allowable margin of positional error, and its relationship to the other factors we have discussed, such as other forms of *environmental authenticity* within the experience, provides an opportunity for future study.

### 7.5.2 Creating realistic (and functional) audio augmented objects

The ranges and focal angles of placed audio sources can be made available as adjustable authoring parameters for maintaining the coherent audibility of exhibited sounds and for the construction of user composable soundscapes, as well as for the fine tuning of *attractor sounds* and *attractor objects,* within an *audio-first* and *object-first* approach (see Figure 31). These adjustable authoring parameters can also work in conjunction with the positions, shapes and types of physical objects and the size and dimensions of the architectural space within which they are contained.

In Figure 35, we can see how a sound source's parameters can be adjusted to fit the position of a radio's speaker, determine from what distance the object will be audible within the resultant soundscape and at what angle. We can also see from Figure 35, how visual representations of these attributes, along with the sound sources position in three-dimensional space, can be communicated visually within the authoring process. Such a feature would allow the author to easily detect, determine and create overlaps of audio sources, create attractor sounds for objects and effectively position virtual audio content onto real-world objects.

This is especially apparent in relation to the attempted realistic audio augmentation of historical sound making artefacts as presented in the study of The McMichael Experiment (see chapter 5), where objects may have been designed to project their sound from a specific direction, such as the speaker situated at the bottom or the front of a radio artefact. But this issue also translates to other exhibition environments where one can imagine both free standing artefacts which can be approached from potentially any direction and wall mounted artefacts which may be required to project their sound in a specific direction.

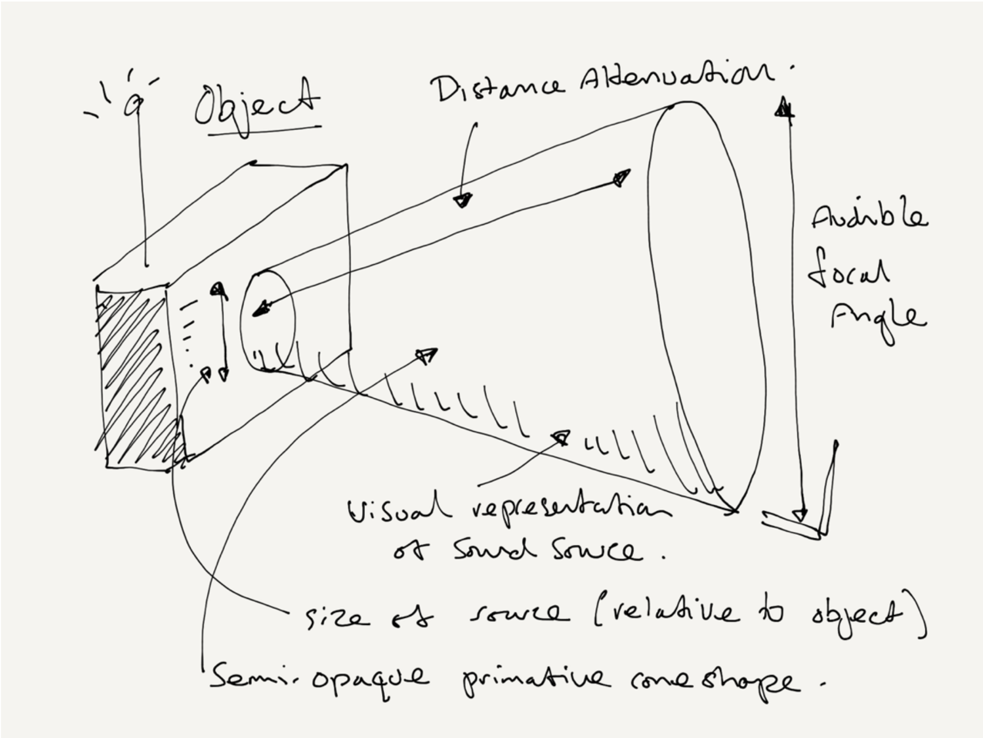


Figure 35. Aiding the authoring process with visualisations. Visualising sound sources and their parameters and authoring a sound source's parameters to emulate the audio output of a silenced-object.

### 7.5.3 Authoring an explorable virtual model

In Figure 36, we can see how a transition to a map-based view can provide the author with information about how the soundscape will be perceived within the confines of the architectural space. Again, be visualising the positions and characteristics of the virtual audio content, the authorship of *attractor sounds* and *attractor objects* can be facilitated. Within larger and more dimensionally complex indoor spaces, listener trajectories through space can be effectively authored, with sounds designed to entice listeners down corridors, or indicating the presence of virtual and physical content behind partition walls or adjacent rooms and gallery spaces.

Perhaps most applicable to gallery spaces, though by no means exclusively, sound sources can be assigned and illustrated as being either directional or omni-directional. A good example of which would be audio augmented wall-mounted paintings within a gallery space being assigned as directional sources, and free-standing sculptural exhibits assigned as omni-directional sources (see Figure 36). Of course, within such a context, it would be up to the author of the audio augmented object reality to determine if they indeed wanted their free-standing exhibits as being audible from all directions, or if they wanted to attract attention to it from a specific direction. These decisions, along with others, are largely determined by two factors. Firstly, what the author would like the listener to curatorially attend to, and secondly, the dimensions of the interior architectural space which, as previously discussed, ultimately determines the confines of the virtual world model.

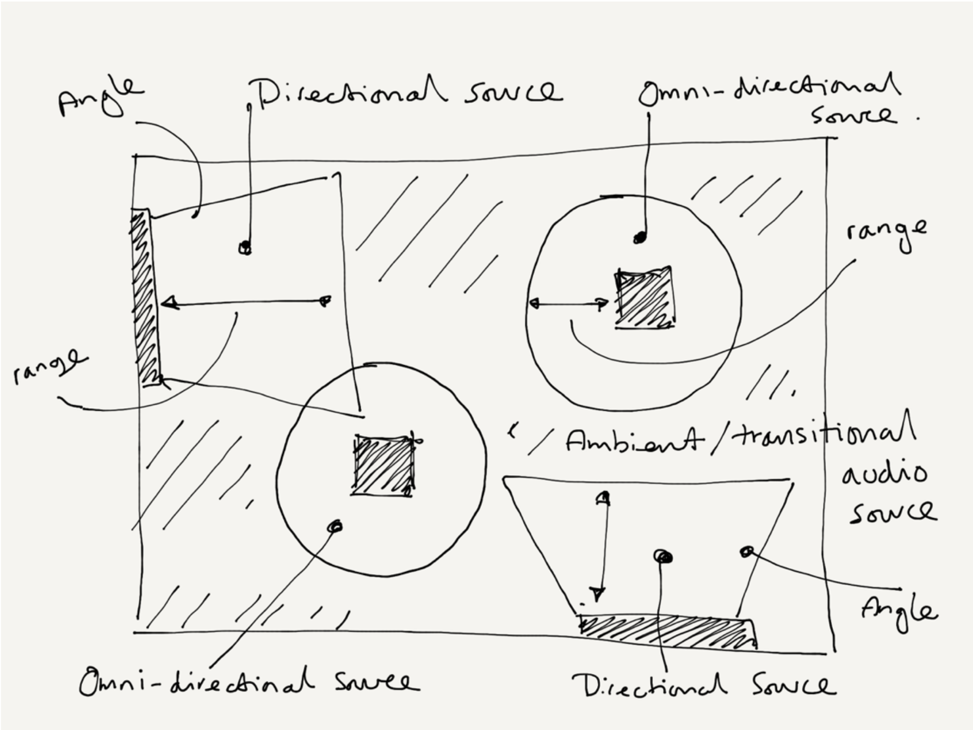


Figure 36. Illustrative map-view of a soundscape. Indicating sound sources of differing positions, directional types, angles and also the inclusion of a *background ambient* or *transitional* audio source.

Having said this, a map-based view of the virtual soundscape also aids the effective authoring of the *sonic atmosphere*. The way in which the different authored components of an audio augmented object reality (see section 7.2.3) interact with each other determines the structure of the explorable sonic atmosphere. For example, how individual sound sources overlap and sit within any included ambient or transitional audio content, and how, and to what extent dividual audio sources may overlap amongst themselves. Both these factors determine the sonic perspectives available to the listener to explore, or rather, what audio content they are able to encounter within different parts of the experience and where. By creating, and carefully crafting, overlapping audio content, the identified functional properties of sonic atmospheres can be exploited as discussed in section 7.2.3.

### 7.5.4 Manipulating the acoustic reality

In order to author an acoustic virtual reality that doesn't necessarily follow the rules of acoustics requires a flexible audio framework.

Currently, within *Apple's AVFoundation* framework, distance attenuation parameters, including both roll-off factor and reference volume, are controlled globally by the *AVAudioEnvironmentNode,* and are, as such, considered as characteristics of the audio environment as a whole, rather than individual characteristics of the sound sources therein.

In order to realise the authorship of a coherent and navigable soundscape within which individual sound sources can be positioned and given different curatorial priority, and within which individual sounds can be easily discerned by a listener, a requirement for individual control over a source's distance attenuation parameters is required. The manipulation of these characteristics at a source level, rather than at an environmental level also enable the precise and exact authorship of *attractor sounds* and *sonic atmosphere.*

Within the *Horror-Fi Me* application (see section 4.5), a custom distance attenuation model had to be authored which exponentially controlled the volume of the virtual source based on its specified range and the distance of the listener from it. Whilst the native *audioSource* component within *Unity*[1] enables individual control over a sound source's distance attenuation characteristics, it currently does not have the ability to render these binaurally. Thankfully, due to a resurgence of interest in binaural audio, there are an increasing amount of software plugins and audio frameworks available that enable the individual manipulation of a source's characteristics and have the ability to these sources binaurally. Within *The McMichael Experiment* (see section 4.3) this was achieved through a combination of *FMOD*[2], the *Resonance Audio*[3] spatialisation plugin and *Unity*. One more recent, and straight forward, solution is provided by *DearVR*[4] which combines the binaural functionality of *Resonance Audio* with the precise control over a source's characteristics in a single *Unity* plugin.

### 7.5.5 Scalability through persistence

The ability to create persistent AR environments means that more experiences can be authored and more users can explore these experiences. Within a system that incorporates a persistent virtual model, in that it can be freely created, saved, loaded and experienced at any time, enables the creation of a single application that allows both the authorship and exploration of such an experience. A persistent model could also facilitate a double-application approach, meaning that a parent *authoring* application could be created, and a child *player* application could be created which plays the experiences created with the parent application. The single application approach was deployed for *Horror-Fi Me* (see section 4.5), allowing users to both author and explore the experience they had created.

---

[1] *Unity* is a cross-platform game authoring application. See: https://unity.com
[2] *FMOD* is described as an audio middleware program for authoring sound effects and interactive music sequences for video games. *FMOD* provides an application programming environment (API) that allows the parameters of an authored *FMOD audio event* to be manipulated via components and scripts within the *Unity Game Engine.*
[3] https://resonance-audio.github.io/resonance-audio/
[4] https://www.dear-reality.com/products/dearvr-unity

By combining *persistence* with a remote storage service, it would be possible to create *authoring* and *playing* applications that are capable of creating and playing multiple experiences, rather than just a single experience, as in the case of *Horror-Fi Me*. Within such a model, the coordinates of virtual sound sources, along with their authored attributes, could be uploaded and saved remotely. The *point-cloud*, or *world-map* data, as assembled by the *ARKit*[1] or *ARCore*[2] frameworks could also be saved so that virtual content can be repositioned correctly within the physical environment. Current options for achieving such a model include *Microsoft's Azure Spatial Anchors*[3] and *Google's Cloud Anchors*[4].

This approach also has powerful implications for singular and site-specific experiences as singular player applications can be configured to retrieve the latest virtual content and world map data. Within the context of the museum or gallery, new experiences can be created, edited and added to with the knowledge that he visitor's player application will retrieve the latest, updated version of the experience.

### 7.5.6 Access and deployment

The importance of the relationship between architectural space, the position of objects and the virtual model in realising an audio augmented object reality underlines the importance of access to the deployment setting for the authorship of the experience. Whilst it is possible, as demonstrated by the installations described in chapter 4, to augment objects with audio via an immediate *tag and explore* approach, it is only through extended access and surveying of the installation setting that the precise authorship of *attractor sounds, associated objects*, *sonic boundaries*, *visitor trajectories* and *sonic atmospheres* can be effectively authored.

With *Unity's ARFoundation*[5] framework supporting building AR applications to the *Android* mobile operating system (via *ARCore*) and the *iOS* mobile operating system (via *ARKit*), both authoring and playback applications can be made available across both these platforms. Furthermore, applications made for the purpose of authoring and playing audio augmented reality experiences are capable of running on devices that are up to five years old[6].

[1] *ARKit* is *Apple's* SLAM-based AR software development kit (SDK) that enables third-party developers to build AR applications that utilise an iOS device's camera and motion sensors.
[2] *ARCore* is *Google's* AR SDK that enables third-party developers to build AR applications that utilise an Android device's camera and motion sensors.
[3] https://azure.microsoft.com/en-gb/services/spatial-anchors/#overview
[4] https://developers.google.com/ar/develop/cloud-anchors
[5] *ARFoundation* is a cross-platform AR SDK for use with the *Unity* game engine. It makes available the key features of *ARCore* (for *Android*) and *ARKit* (for *iOS*) to enable the authoring of cross-platform mobile AR applications.
[6] The minimum hardware specification for *ARKit* is an iPhone SE 2016 , the minimum hardware specification for *ARCore* is an Android device running Android 7.0 or later.

Although deployment remains extremely accessible, due to the relatively low-specification and ubiquity of the consumer level technology required to run an audio augmented reality experience, there remain factors that can help deploy the virtual content onto a user's devices once they have installed the application.
With both AR frameworks reliant upon computer vision for the recognition, tracking and mapping of physical space, a well-lit physical environment remains a pre-requisite for the deployment of a robust AR experience. Additionally, a featureful, rather than a featureless, deployment environment also remains essential. The plain white walls of a contemporary gallery space provide little to no feature-points to populate the virtual world map and enable the correct positioning of both user and virtual content. If, on the other hand, these walls a populated by uniquely identifiable pictures, or the space between them is populated by uniquely identifiable objects, then an featureful world map can be created that can successfully localise and re-localise both the user and the virtual content within the physical environment.

Using the previously discussed persistent virtual world model possible to deploy a virtual soundscape based around the identification of a *reference feature* - a static, uniquely identifiable and trackable image, painting, photo or architectural feature. This *reference feature* can then be set as the world origin (coordinates 0,0,0) and all other virtual content can be positioned relative to this origin. An XML or JSON file can be created detailing the name, parameters and position of each virtual audio source to render within the authored soundscape. On playback, successful recognition of the *reference feature* loads the virtual content from file. The advantages of this approach is that re-localisation occurs on successful image recognition, and virtual anchors are created around the *reference feature* with the rendered and correctly positioned virtual content in relation to the *reference feature*. This approach could be part of the authoring process with the *reference feature* image being uploaded with the JSON data at authoring runtime, and retrievable during playback.

Creating an experience that encourages the user to move around the interactional setting will aid in the localisation of the user's, and the virtual content's, position within the virtual model. Whilst the freely explorative nature of the approach to the creation of audio augmented object-based experiences outlined within this thesis facilitates this, instructional narrative could be included within the experience to aid localisation. For example, *'examine the portrait above the fireplace with your phone's camera'* - the portrait being the *reference feature* that, upon recognition, loads and renders the virtual soundscape.

# 8 Conclusions

This thesis has researched the characteristics, experiential qualities, functional attributes and creation of audio augmented objects within the context of the museum and the home. This research has been undertaken through a practice-based approach, ethnographically framed studies and thematically analysed interviews. By way of a general conclusion, it is proposed that the audio augmented reality object alters the current, popular perception of acoustic virtual reality from an *indirect* to *direct* listening experience. It is also proposed that this change in perception has the potential to offer new audio experiences across a variety of listening contexts. The following sections consider the original research questions outlined in section 2.9 and provide conclusions based on the presented research findings.

## 8.1 The audio augmented object reality

This section considers the first research question: *What are the characteristics and experiential qualities of dynamic binaural audio augmented realities that contain audio augmented real-world objects?*

Firstly, it should be noted that the audio augmented object reality can be a highly intuitive and transparent virtual reality.

Audio augmented objects signal an experiential change in the common consumption of audio-only acoustic virtual realities; they have the potential to transform the acousmatic acoustic virtual reality into a more direct, or *live listening*, experience, one within which the sound sources of the acoustic virtual reality can be perceived as being present. As such, they seize to be audio-only experiences, or indeed augmented reality or augmented virtuality experiences, and become true mixed reality experiences comprised of virtual audio content and components of the listener's physical reality that have the potential to be indistinguishable from each other.

Whilst this vivid perception of reality within an acoustic virtual reality signifies a new type of experience enabled by the associated technologies, the association between virtual audio and physical location is nothing new. The difference here is that this association is explicitly authored, the connection between the virtual audio content and the physical reality is predetermined, finely tuned and designed. Without a reliance on serendipitous connections between the virtual and the real, the audio augmented object reality, like the locative audio experience, can place sound within a specific reality. The difference between it having the potential to happen, and it being designed to happen.

Though, whilst the locative audio experience has the potential to effectively reframe a geographic location to a practitioner's specifications, the audio augmented object reality has the potential to effectively reframe both location and object to a practitioner's specifications. As with the locative audio experience, one of the key aspects is the specification of function through authorship that, in the case of the audio augmented object reality, can be realised at the level of individual objects.

The augmented object reality, where the perception of the virtual as being indistinguishable from the real has been achieved, suggests that an experience has being created that is neither augmented reality or augmented virtuality, it is at once both, and is therefore a mixed reality experience. Sitting in the centre of *Milgram's reality-virtuality continuum*, such an experience is perceptively at once a reality that has been augmented with virtual content, and a virtuality that has been augmented with real content, the virtual and the real are combined to form one perceptively coherent experience in mixed reality.

Ultimately, and more generally, the key characteristic of the audio augmented object reality is a product of its physicality and its perceived reality. This combination enables, across contexts, a virtual acoustic environment that can be perceived as being within the listener's reality, rather than a virtual acoustic environment that the listener perceives as being somewhere else. This creates an acoustic virtual reality that is *'here and now'*, rather than *'there and then'.*

## 8.2 The function of audio augmented object realities

The following section considers the second research question: *What are the potential functions of audio augmented objects within dynamic binaural audio augmented realities?*

The ability of the audio augmented object reality to create an alternative or artificial reality speaks to its potential to reframe not just the acoustic reality, but the user's perception of the visual reality as well. With this view we can firstly conclude that perhaps one of the most vivid functions of audio augmented objects is their ability to reframe, or recontextualise, physical environments. This ability to reframe includes both their collective ability to create alternative atmospheres, as well as their individual ability to reframe the perception of themselves as physical objects. When both these methods are combined, the sense of expectation created by the collective sonic atmosphere can help to exaggerate the listener's perception of individual objects. Using these properties, designers of audio augmented object experiences can harness this functionality to, for example, reframe physical environments, collections of objects or, indeed, both. Such functionality, it would seem, is highly desirable for museum and gallery curators wishing to re-curate existing collections of art and artefacts while foregoing extensive physical alterations to existing displays and gallery layouts. Furthermore, by manipulating the acoustic reality *attractor sounds* that guide users towards physical objects, the creation of *sonic boundaries* that determine the interactional setting of the experience and the creation of *sonic atmospheres* that can facilitate exploration, engagement and learning and be authored.

## 8.3 Creating audio augmented object realities

This next section considers the final research question: *What are the best practices for realising culturally engaging dynamic binaural audio augmented realities that contain audio augmented objects?*

Current AR authoring frameworks and game engines that are inherently visually centric, and remain principally interested in the creation of visually centric experiences, can be effectively appropriated for the authorship and playback of audio augmented object realities. The methods used within these AR applications and frameworks for placing virtual graphical content within the physical reality translate well for placing virtual audio content within the physical reality. Whilst, in many cases, virtual graphical objects can be replaced with virtual audio objects, the visually orientated approaches and assumptions within these tools can be used to visualise the position and character of virtual audio content within the authoring process. In short, It could be said that AR technology works remarkably well, if not better, for the purposes of AAR.

The creation of *environmental authenticity* remains of crucial importance within the audio augmented object reality, just as with a visual AR experience, where this is achieved by embedding virtual graphical content within the camera's view of the physical reality.

Within the audio augmented object reality *environmental authenticity* is achieved, in the main part, by the delivery of dynamic binaural audio attached to physical objects. The perception of *environmental authenticity* within an audio augmented object reality can be increased in a number of ways. These include: the addition of audio augmented features and locations, virtual sound sources and an appropriate physical context.

By increasing *environmental authenticity*, and *cultural authenticity*, within the experience, it is possible to manipulate the realism of the acoustic virtual reality. That is, the expected behaviour of the acoustic virtual reality as a simulation of reality can be altered to provide functionality within the experience.

By continuing to consider *environmental* and *cultural authenticity* during the design process, a robust and reliable audio augmented object reality has the potential to be indistinguishable from reality, yet provide functionality beyond that which is possible in acoustic reality.

## 8.4 Limitations and future work

It should be noted that the work presented in this thesis has its limitations. Firstly, the limited number of participants within the two studies should be acknowledged. Secondly, though representative across the adult age range, the studies did not involve children. This seems important, especially when considering the museum environment as a place that family groups visit. In this respect, it would seem prudent that further research within this context includes children and family groups as participants.

Additionally, it is envisaged that there are opportunities for focused, laboratory-based studies arising from the work presented here. Such studies could focus upon users' abilities to localise and engage with different types of audio content and physical objects used to construct audio augmented objects. Such studies could also make use of application logs and heat maps to better determine the types of audio content and physical object combinations that best aid exploration, attraction and engagement. Such studies could also be used to determine exactly how far the realism of the acoustic virtual reality can be manipulated while still maintaining the perception of reality.

While discussion regarding the generalisation of developed theory has attempted to stay within the context of the research conducted, recognising the importance of *the setting* within the ethnographic perspective, enticing opportunities are apparent across other environments. For example, these include audio-based navigation and informational aids across a variety of different contexts. As such, studies of audio augmented objects across broader contexts would undoubtedly lead to the broader generalisation of theory, as well as their broader application.

Much technical work also remains within the developing field of audio augmented reality, and in particular, the audio augmented object reality. The following observations are a result of the practical development and deployment of these types of experiences over the course of the research presented within this thesis. They represent opportunities observed and noted that couldn't necessarily be implemented due to time constraints, access constraints or the availability of consumer-level technology needed to deploy them remotely.

### 8.4.1 Increasing environmental authenticity

By utilising the vertical plane detection feature of *ARCore* or *ARKit* within the authoring process it would be possible to create an accurate model of the physical setting of the audio augmented reality experience. Such a model could be used to author, audition and deploy a highly realistic and site-specific virtual audio environment. Exterior and interior walls, materials and even furnishings could all be included and defined in order to replicate the reflective, occlusive and absorptive properties of the physical environment.


Figure 37. Plotting vertical planes to define real-world geometry. (left). Using the defined geometry to enable virtual content to seemingly occlude and interact with the real-world in a virtual ball-shooting AR game (Lundgren, 2017).

Figure 37 shows two screen shots from a virtual ball shooting game. Within the game virtual soft balls are occluded by real-world objects, they also interact with the real-world geometry by seemingly bouncing off it. In the first instance the geometry of the real-world is defined by using *ARKit's* plane detection feature and plotting the corners of the vertical planes within the real-world space to create a virtual model of the architectural space. This model is then used by the AR framework's physics engine to allow the virtual content to interact with it in a realistic way, including bouncing the virtual balls off of real-world surfaces. Within an audio AR application, a similar approach could be used to allow the audio engine to interact more realistically with the geometry of the real-world environment.

Within visually orientated AR applications the practice of *environmental probing* allows for adjustments of a three-dimensional model's shadows and reflections in order to increase its *environmental authenticity*. It would seem that this practice could extend to the *probing* of the acoustic environment in order to increase the environmental authenticity of virtual audio. This practice could be facilitated by the recording of impulse responses in order to determine the acoustic characteristics of the interactional setting.

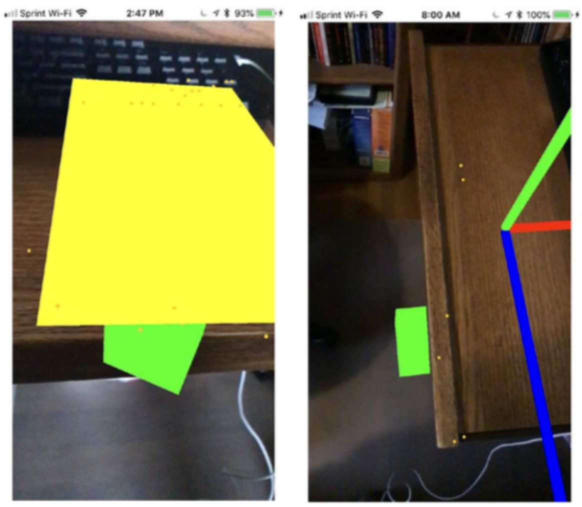


Figure 38. A method for occluding virtual objects behind real-world objects (Wang, 2018).

The object-based occlusion method illustrated in Figure 38, being visually based, uses transparent virtual objects placed in the same location as real-world objects to create the illusion of correct visual occlusion of virtual objects by real-world objects. This method could also be applied to the occlusion of sound sources, where virtual objects can also be placed in the locations of real-world objects to create an audio based equivalent, resulting in muffled sounds, the result of reflected and obstructed, rather than direct sound.

### 8.4.2 Optimising functionality

Having determined that function within an audio augmented reality can be created by manipulating the virtual acoustic reality, there remains a question around to what extent it can be manipulated whilst maintaining the illusion of reality. Furthermore, as there appears to be a connection between this relationship and the inclusion of elements that provide *cultural* and *environmental authenticity* to an AAR experience, the question remains: How much do you need of each component? Or, perhaps more specifically: What elements are required, and to what extent, in order to maintain the perception of reality within a dynamic binaural acoustic virtual reality consisting of audio augmented objects?

This question also relates to the exact allowable margin of positional error of virtual audio content, and its relationship to other forms of *environmental authenticity* within the experience.

Such questions seem important here as, having identified a connection, they could help to uncover best practices for realising desired functionality within an audio augmented object reality.

### 8.4.3 Extending persistent virtual models

Much potential remains within persistent virtual models. By utilisng third-party cloud-based services, such as *Google's AR Cloud Anchors* or *Microsoft's Azure Spatial Anchors* (as outlined in section 7.5.5) the possibility of user-authored audio augmented objects that can be shared and experienced amongst different users becomes a possibility.

Within such a model, audio content, locational mapping data and the coordinates of sound sources in the form of *AR anchors* could be uploaded to a remote server by users for others to find and explore. This creates the potential for large scale, evolving and live experiences to be authored by multiple users, with new audio augmented objects and locations capable of being created, uploaded and updated within experiences in real-time.

Particular topics for further study within such a model would be the social interactions that take place between users, how and what sort of content is created and what types of objects are augmented and who for.

### 8.4.4 GPS and AR hybrid experiences

Audio augmented objects could be used to extend the functionality of GPS-enabled locative listening experiences. By using geo-located sound sources combined with AR located sound sources, audio augmented objects could let their whereabouts be known over much larger distances, well beyond that achievable by AR alone. Once located, and when the listener is within close proximity to the object, AR functionality could take over to allow closer inspection, or more fine-grained audio augmentation at an object level.

Perhaps a good example of how this could be applied would be to audio augmented statues situated around a city centre. Within such an experience GPS-enabled audio augmentation of geographical coordinates could facilitate long-range localisation. Once the user is within a defined proximity to the statue, AR-enabled audio augmentation could directly augment the statue with audio archival content, creating the impression the statue is delivering a speech.

Similarly, through a combination of these two positioning technologies, an experience could be authored that would allow for a transition from an indoor environment to an outdoor environment, or vice versa. By using GPS to determine the position of the listener and a sound source outdoors, an experience could be authored that allows the listener to find the whereabouts of an indoor location, within which there is a transition to an AR indoor positioning system.

### 8.4.5 Head tracking

The projects presented here have determined that a handheld tracking approach is capable of realising an immersive audio augmented reality within which spatialised sound sources can be successfully localised by listeners. Though this has proved practical within the scope of the work presented here, that fact that binaural audio renderers use HRFT profiles to virtually simulate how we hear in reality indicates that a head-based tracking approach would significantly contribute to the sophistication of the rendered acoustic virtual reality. Indeed, within the one project which did use a head-based tracking approach (see section 4.4), an acute sense of reality was observed being experienced by participants (see section 4.4.3). The handheld approach was pursued over a head-based approach in the subsequent projects in interests of accessibility and remote deployment, given that the handheld approach had proven usable and could be implemented with a phone and a standard pair of stereo headphones.

The increasing availability of sensor-embedded headphones, or hearables (as discussed in section 7.3.3) at a consumer level, indicate that a head tracking approach to audio augmented reality could soon be an accessible and practical option for the remote deployment of these experiences.

Recent and fast-moving developments in this technology enabled the latest version of the *Horror-Fi Me* mobile app to be authored with this in mind. Within this latest version users who have access to *Apple's AirPod Pro* or *Airpod Max* model headphones can make use of the sensors embedded within these headphones to enable rotational tracking of the head. Accessibility is maintained by continual support for standard stereo headphones. In both cases translational tracking is determined by the location of the handheld device.

Going forward, the benefits of the head tracking approach on the perception of reality and the construction of authentic and functional acoustic virtual realities would be another useful subject for further enquiry.

### 8.4.6 Final words

It is proposed that the work presented within this thesis provides a foundation for future work with audio augmented objects, has highlighted the types of new and exciting audio-based experiences that can be created with them, and has provided insights into how to realise such experiences.

As a creative practitioner working with sound, the future possibilities associated with making physical objects audible and enhancing our reality them seem numerous. From telling their stories and histories, enabling us to find them, interacting with them, composing with them, entertaining and communicating ideas with them, audio augmented objects offer many new ways to explore our reality.

# 9 Bibliography


Aceti, L. (2013). “Not Here, Not There: An Analysis Of An International Collaboration To Survey Augmented Reality Art.” *Leonardo*, 19 (1), pp. 1–16.

Adorno, T. and Eisler, H. (1994). *Composing for the Films*. New York: Continuum.

Apple. (2020). *App analytics metrics and dimensions*. Available at: https://help.apple.com/app-store-connect/#/itc21781223f (Accessed: December 21, 2020).

Apple. (2021a). *AirPods (2nd generation) - Technical Specifications - Apple (UK)*. Available at: https://www.apple.com/uk/airpods-2nd-generation/specs/ (Accessed: December 21, 2020).

Apple. (2021b). *Apple Music’s Zane Lowe explains how Spatial Audio will transform music - Apple (UK)*. Available at: https://www.apple.com/uk/newsroom/2021/06/apple-musics-zane-lowe-explains-how-spatial-audio-will-transform-music/ (Accessed: July 28, 2021).

Apple. (2021c). *ARKit | Apple Developer Documentation*. Available at: https://developer.apple.com/documentation/arkit (Accessed: October 20, 2021).

Apple. (2021d). *Spatial Scenes on the App Store*. Available at: https://apps.apple.com/us/app/spatial-scenes/id1532491431 (Accessed: January 6, 2022).

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B. (2001). “Recent advances in augmented reality.” *IEEE Computer Graphics and Applications*, 21 (6), pp. 34–47. doi: 10.1109/38.963459.

Baker, D. L. (2014). “Wearable Apocalypses: Enabling Technologies for Aspiring Destroyers of Worlds.” in Geroimenko, V. (ed.) *Augmented Reality Art: From an Emerging Technology to a Novel Creative Medium*. Switzerland: Springer, pp. 305–312.

Barrios-O’Neill, D. (2018). “Fearful Sounds: Cross-Platform Studies of Sonic Audio and Horror.” *Revenant*, (3), pp. 1–4.

Bauer, V., Nagele, A., Baume, C., Cowlishaw, T., Cooke, H., Pike, C. and Healey, P. G. T. (2019). “Designing an Interactive and Collaborative Experience in Audio Augmented Reality.” in *EuroVR: International Conference on Virtual Reality and Augmented Reality*. doi: https://doi.org/10.1007/978-3-030-31908-3_20.

BBC. (2022). *BBC Sound Effects*. Available at: https://sound-effects.bbcrewind.co.uk/ (Accessed: February 16, 2022).

Bederson, B. B. (1995). "Audio Augmented Reality: A Prototype Automated Tour Guide." in *CHI95: Conference on Human Factor in Computing Systems*. New York, NY, United States: ACM Press, pp. 1–2.

Begault, D. R. (2000). *3-D Sound for Virtual Reality and Multimedia*. California: NASA.

Behrendt, F. (2010). *Mobile Sound: Media Art in Hybrid Spaces*.

Behrendt, F. (2012). "The sound of locative media." *Convergence: The International Journal of Research into New Media Technologies*, 18 (3), pp. 283–295. doi: 10.1177/1354856512441150.

Behrendt, F. (2015). "Locative Media as Sonic Interaction Design: Walking through Placed Sounds." *Wi: Journal of Mobile Media*, 9 (2).

Benford, S., Adams, M., Tandavanitj, N., Farr, J. R., Greenhalgh, C., Crabtree, A., Flintham, M., Walker, B., Marshall, J., Koleva, B., Egglestone, S. R. and Giannachi, G. (2013). "Performance-Led Research in the Wild." *ACM Transactions on Computer-Human Interaction*, 20 (3), pp. 1–22. doi: 10.1145/2491500.2491502.

Benford, S. and Giannachi, G. (2012). "Interaction as Performance." *Interactions*, 3 (19), pp. 38–43.

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

Bose. (2018). "Bose AR Design Guidelines - 12-4-18." Available at: https://developer.bose.com/sites/default/files/Bose%20AR%20Design%20Guidelines%20v1.0.pdf.

Bose. (2021). *Bose Developer Portal | Sonic Samurai Creates Revolutionary Interactive Gaming Experience Using Bose AR*. Available at: https://developer.bose.com/bose-ar/get-inspired/sonic-samurai-creates-revolutionary-interactive-gaming-experience-using-bose-ar (Accessed: January 6, 2022).

Braun, V. and Clarke, V. (2006). "Using thematic analysis in psychology." *Qualitative Research in Psychology*, 3 (2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

British-Library. (2019). *Unlocking Our Sound Heritage - The British Library*. Available at: https://www.bl.uk/projects/unlocking-our-sound-heritage# (Accessed: January 10, 2022).

Bull, M. (2000). *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Oxford: Berg.

Bull, M. (2013). “The End of Flânerie: iPods, Aesthetics, and Urban Experience.” in Ekman, U. (ed.) *Throughout: Art and culture emerging with ubiquitous computing*. MIT Press.

Carmigniani, J. and Furht, B. (2011). “Augmented Reality: An Overview.” in Furht, B. (ed.) *Handbook of Augmented Reality*. New York: Springer. doi: 10.1007/978-1-4614-0064-6.

Cervenak, R. and Masek, P. (2019). “ARKit as indoor positioning system.” in. (2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)), pp. 1–5. doi: 10.1109/icumt48472.2019.8970761.

Chaparro, I. and Duenas, R. (2015). “Psychogeographical Sound-drift.” in. (the 2015 ACM SIGCHI Conference), pp. 187–188. doi: 10.1145/2757226.2764559.

Cheng, J. C. P., Chen, K. and Chen, W. (2017). “Comparison of marker-based AR and markerless AR: A case study on indoor decoration system.” in *Lean & Computing in Construction Congress*.

Cliffe, L., Mansell, J., Cormac, J., Greenhalgh, C. and Hazzard, A. (2019). “The Audible Artefact: Promoting Cultural Exploration and Engagement with Audio Augmented Reality.” in *AM’19: Audio Mostly 2019*. (AM’19: Audio Mostly 2019), pp. 1–7. doi: 10.1145/3356590.3356617.

Cliffe, L., Mansell, J., Greenhalgh, C. and Hazzard, A. (2020). “Materialising contexts: virtual soundscapes for real-world exploration.” *Personal and Ubiquitous Computing*, 20 (3), pp. 1–16. doi: 10.1007/s00779-020-01405-3.

Connor, S. (2011). “Ears Have Walls, On Hearing Art.” in Kelly, C. (ed.) *Sound*. Whitechapel Gallery/MIT Press.

Crabtree, A., Rouncefield, M. and Tolmie, P. (2012). *Doing Design Ethnography*. London: Springer-Verlag. doi: 10.1007/978-1-4471-2726-0.

Crook, T. (1999). *Radio Drama; Theory and practice*. Routledge. doi: 10.4324/9780203006276.

Cuartielles, D. (2012). “Embodied Sound.” in *Proceedings of the Participatory Design Conference,*. Toronto, Canada, pp. 1–4.

Debord, G. (1958). *Theory of the dérive*. Oakland: Bureau of Public Secrets.

Deleuze, G. and Guattari, F. (2005). *A Thousand Plateaus Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press.

Dolby. (2022). *Dolby Atmos - Official Site - Dolby*. Available at: https://www.dolby.com/technologies/dolby-atmos/ (Accessed: March 7, 2022).

Droumeva, M. (2016). "Curating Aural Experience: A Sonic Ethnography of Everyday Media Practices." *Interference Journal*, (5), pp. 72–88.

Durland, S. (1987). "Defining the image as place." *HIGH PERFORMANCE: A Quarterly Magazine for the New Art Audience*, pp. 1–15. Available at: http://www.ecafe.com/museum/hp_gy_1987/hp_gy_1987.html.

Elferen, I. van. (2012). *Gothic Music The Sounds of the Uncanny*. Cardiff: University of Wales Press.

Foster, H. (1996). *The Return of the Real: The Avant-Garde at the End of the Century*. Cambridge, Massachusetts: MIT Press.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. London, United Kingdom: Polity Press.

Gaver, B. and Bowers, J. (2012). "Annotated Portfolios." *Interactions*, pp. 1–10. Available at: https://dl.acm.org/doi/pdf/10.1145/2212877.2212889?download=true.

Gaye, L., Mazé, R. and Holmquist, L. E. (2003). "Sonic City: The Urban Environment as a Musical Interface." in *2003 Conference on New Interfaces for Musical Expression (NIME-03)*, pp. 109–115.

Gershon, W. S. (2019). "Sonic Ethnography in Theory and Practice." doi: 10.1093/acrefore/9780190264093.013.547.

Gimenes, M., Largeron, P.-E. and Miranda, E. R. (2016). "Frontiers: Expanding Musical Imagination With Audience Participation." in *NIME '16*. Available at: http://www.nime.org/proceedings/2016/nime2016_paper0068.pdf.

Google. (2021). *Build new augmented reality experiences that seamlessly blend the digital and physical worlds*. Available at: https://developers.google.com/ar/ (Accessed: October 20, 2021).

Greenhalgh, C. and Benford, S. (1997). "MASSIVE:A Collaborative Virtual Environment for Teleconferencing." *ACM Transactions on Computer-Human Interaction*, 2 (3), pp. 1–23.

Hancock, D. (2018). "'Put on your headphones and turn out the lights': Exploring Immersive Auditory Horror in 3D-sound Podcasting." *Revenant*. Edited by D. Barrios-O'Neill. (Fearful Sounds: Cross-platform studies of sonic audio and horror), (3), pp. 55–71.

Haseman, B. (2007). “Rupture and Recognition: Identifying the performative research paradigm.” in Barrett, E. and Bolt, B. (eds) *Practice as Research: Approaches to Creative Arts Enquiry*. London: I.B.Tauris & Co Ltd, pp. 1–218.

Hazzard, A., Benford, S. and Burnett, G. (2015). “Sculpting a Mobile Musical Soundtrack.” in. (the 33rd Annual ACM Conference), pp. 387–396. doi: 10.1145/2702123.2702236.

Hazzard, A., Spence, J., Greenhalgh, C. and McGrath, S. (2017). “The Rough Mile.” in. (the 12th International Audio Mostly Conference), pp. 1–8. doi: 10.1145/3123514.3123540.

He, S. and Shin, K. G. (2018). “Geomagnetism for Smartphone-Based Indoor Localization.” *ACM Computing Surveys*, 50 (6), pp. 1–37. doi: 10.1145/3139222.

Heller, F. and Borchers, J. (2014). “AudioTorch.” in. (the 16th International Conference), pp. 483–488. doi: 10.1145/2628363.2634220.

Hugues, O., Fuchs, P. and Nannipieri, O. (2011). “New augmented reality taxonomy: Technologies and features of an augmented environment.” in Furht, B. (ed.) *The Handbook of Augmented Reality*. New York: Springer, pp. 47–63.

Iida, K. (2019). “Head-Related Transfer Function and Acoustic Virtual Reality.” doi: 10.1007/978-981-13-9745-5.

Jahn, P. (2015). *Remixing The Stone Tape: Interview with Peter Strickland*. *Electric Sheep: A deviant view of cinema*. Available at: http://www.electricsheepmagazine.co.uk/2015/11/28/remixing-the-stone-tape-interview-with-peter-strickland/ (Accessed: January 22, 2021).

Jordan, B. and Henderson, A. (1995). “Interaction Analysis: Foundations and Practice.” *Journal of the Learning Sciences*, 4 (1), pp. 39–103. doi: 10.1207/s15327809jls0401_2.

Karathanasopoulou, E. (2019). “Atmosphere in radio and architecture:” *Radio Journal: International Studies in Broadcast and Audio Media*, 17 (1), pp. 113–129.

Kassabian, A. (2013). *Ubiquitous Listening Affect, Attention, and Distributed Subjectivity*. Berkeley and Los Angeles, California: University of California Press.

Kee, K. and Compeau, T. J. (2019). “Seeing the Past with Computers, Experiments with Augmented Reality and Computer Vision for History.” doi: 10.3998/mpub.9964786.

Kee, K., Poitras, E. and Compeau, T. J. (2019). “History All Around Us: Toward Best Practices for Augmented Reality for History.” in Kee, K. and Compeau, T. J. (eds) *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*. Ann Arbor: University of Michigan Press.

Kelly, C. (2017). *Gallery Sound*. London: Bloomsbury. doi: 10.5040/9781501305948.

Kelly, C. (2019). “Caleb Kelly, Material Sound: sound as more than sound.” *Material Sound Catalogue Essay*, pp. 1–15.

Kirsh, D. (2013). “Embodied cognition and the magical future of interaction design.” *ACM Transactions on Computer-Human Interaction*, 20 (1), pp. 1–30. doi: 10.1145/2442106.2442109.

Kiss, F., Mayer, S. and Schwind, V. (2020). “Audio VR— Did Video Kill the Radio Star?” *Interactions*, pp. 1–6. Available at: https://dl.acm.org/doi/pdf/10.1145/3386385.

KQED-Arts. (2015). *One Collective Breath: Janet Cardiff’s “The Forty Part Motet” | KQED Arts - YouTube*. Available at: https://www.youtube.com/watch?v=rZXBia5kuqY (Accessed: December 22, 2021).

Krzyzaniak, M., Frohlich, D. and Jackson, P. J. B. (2019). “Six types of audio that DEFY reality!: A taxonomy of audio augmented reality with examples.” in *Audio Mostly (AM’19)*. New York: ACM Press, pp. 1–8. doi: 10.1145/3356590.3356615.

Lawton, M., Cunningham, S. and Convery, I. (2020). “Nature Soundscapes: An Audio Augmented Reality Experience.” in *In Proceedings of the 15th International Audio Mostly Conference (AM’20), September 15–17, 2020, Graz, Austria.* New York: ACM Press (AM’20: Audio Mostly), pp. 1–8. doi: 10.1145/3411109.

LeWitt, S. (1999). “Paragraphs on conceptual art.” in Alberro, A. and Stimson, B. (eds) *Conceptual art: A critical anthology*. Cambridge, Massachusetts: The MIT Press.

LeWitt, S. (2003). “Sentences on conceptual art.” in Harrison, C. and Wood, P. (eds) *Art in theory 1900 -2000*. Oxford: Blackwell.

Lieberman, Z. (2017). *zach lieberman on Instagram: “Quick test recording audio in space and playing back -- (video has audio !) #openframeworks.”* Available at: https://www.instagram.com/p/BYs0iI3g8kZ/ (Accessed: January 5, 2022).

Lingold, M. C., Mueller, D. and Trettien, W. (2018). *Digital Sound Studies*. Durham and London: Duke University Press.

Lundgren, B. (2017). *GitHub - bjarnel/arkit-occlusion: A demonstration of vertical planes "tracking" and occlusions with ARKit+Scenekit*. Available at: https://github.com/bjarnel/arkit-occlusion (Accessed: March 18, 2022).

Lymberopoulos, D., Liu, J., Yang, X., Choudhury, R. R., Handziski, V. and Sen, S. (2015). "A realistic evaluation and comparison of indoor location technologies." in. (the 14th International Conference), pp. 178–189. doi: 10.1145/2737095.2737726.

MagicBeans. (2022). *MagicBeans Website*. Available at: https://www.magicbeans.xyz/ (Accessed: March 7, 2022).

Marks, L. (2002). *Touch: Sensuous Theory and Multisensory Media*. Minneapolis: University of Minnesota Press.

McLuhan, M. (1994). *Understanding Media The Extensions of Man*. The MIT Press.

Montan, N. (2002). *AAR: An Audio Augmented Reality System*.

Mortensen, C. H. and Vestergaard, V. (2014). "Embodied Tuning: Interfacing Danish Radio Heritage." *Journal of Interactive Humanities*, 1 (1), pp. 23–36. doi: 10.14448/jih.01.0003.

Naphtali, D. and Rodkin, R. (2020). "Audio augmented reality for interactive soundwalks, sound art and music delivery." in Filimowicz, M. (ed.) *Foundation in sound design for interactive media: a multidisciplinary approach*. New York: Routledge.

Nicol, R. (2020). "Creating Auditory Illusions with Spatial-Audio Technologies." in Blauret, J. and Braasch, J. (eds) *The Technology of Binaural Understanding*. Springer, pp. 581–663.

Ouzounian, G. (2020). *Stereophonica: Sound and Space in Science, Technology, and the Arts*. Cambridge, Massachusetts: The MIT Press.

OwlField. (2020). *3D Escape Room: Frequency · Owl Field*. Available at: https://www.owlfield.com/pf/escape-room-frequency/ (Accessed: January 23, 2021).

Pausch, R. (1996). "Disney's Aladdin:  First Steps Toward Storytelling  in Virtual Reality." in Jnr., C. D. (ed.) *Digital Illusion: Entertaining the future with high technology*. ACM Press.

Phillips, W. (2014). *A Composer's Guide to Game Music*. Cambridge, Massachusetts: MIT Press.

Placenote. (2021). *Placenote: Spatial sticky notes*. Available at: https://placenote.com/annotations (Accessed: November 1, 2021).

Planeta. (2021). *Fields – Spatial Sound in AR*. Available at: https://fields.planeta.cc/ (Accessed: January 5, 2022).

Plazak, J. and Kersten-Oertel, M. (2018). "A Survey on the Affordances of 'Hearables.'" *Inventions*, 3 (3), p. 48. doi: 10.3390/inventions3030048.

Rhodes, G. A. (2014). "Augmented Reality in Art: Aesthetics and Material for Expression." in Geroimenko, V. (ed.) *Augmented Reality Art: From an Emerging Technology to a Novel Creative Medium*. Switzerland: Springer.

Richardson, L. and St. Pierre, E. A. (2008). "Writing: A Method of Inquiry." in Denzin, N. K. and Lincoln, Y. S. (eds) *Collecting and Interpreting Qualitative Materials*. Los Angeles: SAGE, pp. 473–500.

Roettgers, J. (2020). "Another company is giving up on AR. This time, it's Bose.," 10 July. Available at: https://www.protocol.com/enterprise/collab-software.

Sachs, A. (1978). "Andrew Sachs - The Revenge." United Kingdom: BBC Radio 3.

Schaeffer, P. (1966). *Treatise on Musical Objects: An Essay Across Disciplines*. Translated by C. North and J. Dack. Oakland, California: University of California Press.

Schafer, R. M. (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, Vermont: Destiny Books.

Schraffenberger, H. and Heide, E. van der. (2014). "The real in augmented reality." in. *xCoAx 2014: Proceedings of the Second Conference on Computation, Communication, Aesthetics and X.*

Seidenari, L., Baecchi, C., Uricchio, T., Ferracani, A., Bertini, M. and Bimbo, A. D. (2017). "Deep Artwork Detection and Retrieval for Automatic Context-Aware Audio Guides." *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13 (3s), pp. 1–21. doi: 10.1145/3092832.

Shudder. (2021). *Darkest Night: A Podcast Experience*. Available at: https://www.shudder.com/series/watch/darkest-night-a-podcast-experience/396498fa03a0b2be?season=1 (Accessed: September 30, 2021).

Sikora, M., Russo, M., Derek, J. and Jurčević, A. (2018). "Soundscape of an Archaeological Site Recreated with Audio Augmented Reality." *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14 (3), pp. 1–22. doi: 10.1145/3230652.

Sodnik, J., Tomazic, S., Grasset, R., Duenser, A. and Billinghurst, M. (2006). "Spatial Sound Localization in an Augmented Reality Environment." in *OZCHI 2006*. ACM Press, pp. 1–8.

Sterne, J. (2003). *The Audible Past, Cultural Origins of Sound Reproduction*. doi: 10.1354/books/unregistered/9780822384250.

Stinson, L. (2017). "An Artist Uses an iPhone to Visualize Sounds in AR." *Wired*. Available at: https://www.wired.com/story/an-artist-uses-an-iphone-to-visualize-sounds-in-ar/.

Strickland, P. (2015). "The Stone Tape." *Fright Night BBC Radio 4*. BBC Radio 4.

Superflux. (2017). *The Future Starts Here*. *Superflux. Translating future uncertainty into present day choices.* Available at: https://superflux.in/index.php/work/the-future-starts-here/# (Accessed: March 19, 2021).

Taylor, R., Schofield, G., Shearer, J., Wallace, J., Wright, P., Boulanger, P. and Olivier, P. (2011). "Designing from within: humanaquarium." in *Annual Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, pp. 1–10.

The-British-Library-Board. (2022). *British Library - Sounds*. Available at: https://sounds.bl.uk/ (Accessed: February 16, 2022).

Thiel, T. (2014). "Critical Interventions into Canonical Spaces:Augmented Reality at the 2011 Venice and Istanbul Biennials." in Geroimenko, V. (ed.) *Augmented Reality Art: From an Emerging Technology to a Novel Creative Medium*. Switzerland: Springer.

Thielen, E., Letellier, J., Sieck, J. and Thoma, A. (2018). "Bringing a virtual string quartet to life." in. (the Second African Conference for Human Computer Interaction), pp. 1–4. doi: 10.1145/3283458.3283477.

Truax, B. (2012). "Sound, Listening and Place: The aesthetic dilemma." *Organised Sound*, 17 (3), pp. 193–201. doi: 10.1017/s1355771811000380.

vGIS. (2021). *2020 iPad Pro: Does the LiDAR sensor improve spatial tracking?* Available at: https://www.vgis.io/2020/04/23/2020-ipad-pro-does-the-lidar-sensor-improve-spatial-tracking/ (Accessed: November 20, 2021).

Wang, W. (2018). "Beginning ARKit for iPhone and iPad, Augmented Reality App Development for iOS." doi: 10.1007/978-1-4842-4102-8.

Weiser, M. (1999). "The Computer for the 21st Century." *Mobile Computing and Communications Review*, 3, pp. 3–11. doi: https://doi.org/10.1145/329124.329126.

Yang, J., Frank, Y. and Sörös, G. (2019). "Hearing Is Believing." in. (the 10th Augmented Human International Conference 2019), pp. 1–9. doi: 10.1145/3311823.3311872.

Yang, J. and Mattern, F. (2019). "Audio augmented reality for human-object interactions." in. (the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 ACM International Symposium), pp. 408–412. doi: 10.1145/3341162.3349302.

Youngblood, G. (1970). *Expanded Cinema*. Toronto and Vancouver: Clarke, Irwin & Company Limited.

Zimmermann, A. and Lorenz, A. (2008). "LISTEN: a user-adaptive audio-augmented museum guide." *User Modeling and User-Adapted Interaction*, 18 (5), pp. 389–416. doi: 10.1007/s11257-008-9049-x.

# 10 Appendices

A. *The McMichael Experiment* project information sheet.

B. *The McMichael Experiment* consent form

C. *Horror-Fi Me* project information sheet.

D. *Horror-Fi Me* consent form

Appendix A.

# PROJECT INFORMATION

**Date:** 30/01/2019

**Project:** Understanding the deployment of interactive audio systems at cultural institutions, cultural venues and heritage sites

**School of Computer Science Ethics Reference:**

**Funded by:** UK Engineering and Physical Science Research Council (FAST & HORIZON projects)

**Purpose of the research.** The project aims to develop an understanding of how three-dimensional interactive audio might be deployed at cultural institutions, cultural venues and heritage sites in order to enhance visitor experiences and curatorial potential.

**Nature of participation.** Your participation in the research is voluntary and involves taking part in one or more engagement activities (see below), which seek to understand how interactive audio might be relevant to your own work.

**Participant engagement.** Your participation may include: being interviewed, which will be recorded on an audio recorder; workshop activities that include prototype testing and/or group discussions that will be recorded on video, audio and photographically. Where relevant, it may also involve being observed by the researcher, and documented on video and audio, while you are using and testing prototypes and undertaking some of your everyday workplace activities. You will be able to decide, in discussion with the researcher, exactly which activities you will be involved in.

**Benefits and risks of the research.** The data collected will contribute to the researcher's PhD work and will help to shape future technologies and contribute to research publications and presentations. The study involves the gathering of personal data and there is a risk that use of your data may identify you in research reports and publications. Data collected will be held in a secure and safe manner in accordance with the Data Protection Act (2018), including the General Data Protection Regulation (GDPR 2018).

**Use of your data.** Data gathered during the research will be used in supervision sessions, project meetings, project presentations and project reports. The results of the research may be disseminated via conference workshops, presentations, publications and the researcher's thesis. All published data will be anonymous unless you specifically ask to be identified in publications and presentations.

**Future use of your data.** Your data may be archived and reused in future for purposes that are in the public interest, or for historical, scientific or statistical purposes. For example shaping the development of new technologies and contributing to research publications and presentations. Data will be stored on the researchers password protected personal computer and a secure digital environment provided by the University of Nottingham.

**Mixed personal data** The research will gather 'mixed' personal data, i.e., data that simultaneously involves multiple participants and/or is irreducibly social in nature. In this case, mixed personal data includes multi-party conversation recorded on audio or video, group photographs and video. We can only delete mixed personal data if all parties to it withdraw their consent. However, we will redact any data that identifies you in public presentations and reports of this research insofar as this is practicable and the data has not already been made public by yourself (e.g., posted on social media).

**Procedure for withdrawal from the research** You may withdraw from the study at any time and do not have to give reasons for why you no longer want to take part. If you wish to withdraw please contact the researcher who gathered the data. If you receive no response from the researcher please contact the School of Computer Science's Ethics Committee. If you withdraw then your data will be excluded from subsequent use and deleted where possible (note that data – which is normally anonymous – cannot be deleted from publications).

**Contact details of the ethics committee.** If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

**Researcher contact details**

Appendix B.

# CONSENT FORM

University of Nottingham
UK | CHINA | MALAYSIA

**Date:** 30/01/2019

**Project:** Understanding the deployment of interactive audio systems at cultural institutions, cultural venues and heritage sites

**School of Computer Science Ethics Reference:**

**Funded by:** UK Engineering and Physical Science Research Council (FAST & HORIZON projects)

| **Please tick the appropriate boxes** | **Yes** | **No** |
|---|---|---|
| **1. Taking part in the study** | | |
| a) I have read and understood the project information sheet dated 30/01/2019, or it has been read to me. I have been able to ask questions about the study and my questions have been answered satisfactorily. | ☐ | ☐ |
| b) I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. | ☐ | ☐ |
| c) I understand that taking part in the study requires the collection of personal data including both audio and video recordings. | ☐ | ☐ |
| d) I understand that taking part in the study will involve being interviewed, participating in workshops and/or being observed at work. | ☐ | ☐ |
| e) I understand that taking part in the study may involve bodily movement, gesture and proximity to other study participants. | ☐ | ☐ |
| **2. Use of my data in the study** | | |
| a) I understand that data which can identify me will not be shared beyond the project team. | ☐ | ☐ |
| b) I agree that the data provided by me may be used for the following purposes: | | |
| – Presentation and discussion of the project and its results in research activities (e.g., in supervision sessions, project meetings, conferences). | ☐ | ☐ |
| – Publications and reports describing the project and its results. | ☐ | ☐ |
| – Dissemination of the project and its results, including publication of data on web pages and databases. | ☐ | ☐ |
| c) I give permission for my words to be quoted for the purposes described above. | ☐ | ☐ |
| d) I give permission for my visual image contained in photos or video gathered during the research to be used for the purposes described above. | ☐ | ☐ |
| **3. Reuse of my data** | | |
| a) I give permission for the data that I provide to be reused for the sole purposes of | ☐ | ☐ |

1

future research and learning.

b) I understand and agree that this may involve depositing my data in a data repository, which may be accessed by other researchers ☐ ☐

**4. Security of my data**

a) I understand that safeguards will be put in place to protect my identity and my data during the research, and if my data is kept for future use. ☐ ☐

b) I confirm that a written copy of these safeguards has been given to me in the University's privacy notice, and that they have been described to me and are acceptable to me. ☐ ☐

c) I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data. ☐ ☐

**5. Copyright**

a) I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented. ☐ ☐

b) I wish to be publicly identified as the creator of any audio or visual materials that I produce or co-produce as a part of this project. ☐ ☐

**6. Signatures (sign as appropriate)**

| **Name of participant** (IN CAPITALS) | Signature | Date |
|---|---|---|

If applicable:

For participants unable to sign their name, mark the box instead of signing ☐

I have witnessed the accurate reading of the consent form with the participant and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

| **Name of witness** (IN CAPITALS) | Signature | Date |
|---|---|---|

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

| **Name of researcher** (IN CAPITALS) | Signature | Date |
|---|---|---|

**7. Researcher's contact details**

Provide the participant with a copy of the completed form either by email or hard copy as they prefer.

Appendix C.

# PROJECT INFORMATION

University of Nottingham
UK | CHINA | MALAYSIA

**Date:** 02/06/2020
**Project:** Applying Audio Augmented Reality (AAR)
**School of Computer Science Ethics Reference:**
**Funded by:** UK Engineering and Physical Science Research Council

**Purpose of the research.** The project aims to develop an understanding of how Audio Augmented Reality (AAR) might be deployed within and around cultural institutions, cultural venues, heritage sites and domestic settings.

**Nature of participation.** Your participation in the research is voluntary and involves taking part in one or more engagement activities (see below), which seek to understand how interactive audio might be relevant to your own work.

**Participant engagement.** Your participation may include: (1) being interviewed, either in person (which will be recorded on an audio recorder) or via a teleconferencing application (which will be recorded from within the software application itself); (2) workshop activities, that include testing AAR prototypes and/or group discussions, that will be recorded on video, audio and photographically; (3) . being observed by the researcher (or recording yourself), documented on video and audio, while you are using and testing prototypes or (where relevant) undertaking everyday activities that are relevant to AAR (e.g. if you are a curator or sound designer). Data logs of your interactions with the technical prototypes may also be recorded. You will be able to decide, in discussion with the researcher, exactly which activities you wish to be involved in. Any in-person activities will be contingent on and follow current legislation and Government guidance on adapting activities to the COVID-19 situation.

**Benefits and risks of the research.** The data collected will contribute to the researcher's PhD work and will help to shape future technologies and contribute to research publications and presentations. The studies involve the gathering of personal data and there is a risk that use of your data may identify you in research reports and publications. Data collected will be held in a secure and safe manner in accordance with the Data Protection Act (2018), including the General Data Protection Regulation (GDPR 2018).

**Use of your data.** Data gathered during the research will be used in supervision sessions, project meetings, project presentations and project reports. The results of the research may be disseminated via conference workshops, presentations, publications and the researcher's thesis. All published data will be anonymous unless you specifically ask to be identified in publications and presentations.

**Future use of your data.** Your data may be archived and reused in future for purposes that are in the public interest, or for historical, scientific or statistical purposes. For example shaping the development of new technologies and contributing to research publications and presentations. Data will be stored on the researchers password protected personal computer and a secure digital environment provided by the University of Nottingham.

**Mixed personal data** The research will gather 'mixed' personal data, i.e., data that simultaneously involves multiple participants and/or is irreducibly social in nature. In this case, mixed personal data includes multi-party conversation recorded on audio or video, group photographs and video. We can only delete mixed personal data if all parties to it withdraw their consent. However, we will redact any data that identifies you in public presentations and reports of this research insofar as this is practicable and the data has not already been made public by yourself (e.g., posted on social media).

**Procedure for withdrawal from the research** You may withdraw from the study at any time and do not have to give reasons for why you no longer want to take part. If you wish to withdraw please contact the researcher who gathered the data. If you receive no response from the researcher please contact the School of Computer Science's Ethics Committee. If you withdraw then your data will be excluded from subsequent use and deleted where possible (note that data – which is normally anonymous – cannot be deleted from publications).

**Contact details of the ethics committee.** If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

**Researcher contact details**
Name: Laurence Cliffe
Email: laurence.cliffe@nottingham.ac.uk

Appendix D.

# CONSENT FORM

**Date:** 02/06/2020

**Project:** Applying Audio Augmented Reality (AAR)

**School of Computer Science Ethics Reference:**

**Funded by:** UK Engineering and Physical Science Research Council (FAST & HORIZON projects)

| **Please tick the appropriate boxes** | **Yes** | **No** |
|---|---|---|
| **1. Taking part in the study** | | |
| a) I have read and understood the project information sheet dated 02/06/2020, or it has been read to me. I have been able to ask questions about the study and my questions have been answered satisfactorily. | ☐ | ☐ |
| b) I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. | ☐ | ☐ |
| c) I understand that taking part in the study requires the collection of personal data potentially including both audio and video recordings. | ☐ | ☐ |
| d) I understand that taking part in the study will involve being interviewed, participating in workshops and/or (with my specific agreement) being observed or recorded in other relevant activities. | ☐ | ☐ |
| e) I understand that taking part in the study may involve bodily movement, gesture and (for in-person workshops only) proximity to other study participants. | ☐ | ☐ |
| Before trying an AAR prototype at home (only): | | |
| f1) I agree that I have undertaken an appropriate safety audit prior to the session, ensuring that I have sufficient space free of any potential trip hazards. | ☐ | ☐ |
| f2) I agree that I have arranged a chaperone for the duration the study. | ☐ | ☐ |
| f3) I agree that I have provided the researcher with appropriate contact information, and agreed appropriate contact channels, for both myself and my chaperone. | ☐ | ☐ |
| f4) I understand that I will be contacted by the researcher prior to the agreed start time of the session, and that I will be contacted again after the agreed duration of the session. | ☐ | ☐ |
| f5) I understand that in the event that I cannot be contacted by the researcher after the agreed duration of the session my chaperone will be contacted. | ☐ | ☐ |
| **2. Use of my data in the study** | | |
| a) I understand that data which can identify me will not be shared beyond the project team. | ☐ | ☐ |

1

b) I agree that the data provided by me may be used for the following purposes:

- Presentation and discussion of the project and its results in research activities (e.g., in supervision sessions, project meetings, conferences). ☐ ☐
- Publications and reports describing the project and its results. ☐ ☐
- Dissemination of the project and its results, including publication of data on web pages and databases. ☐ ☐

c) I give permission for my words to be quoted for the purposes described above. ☐ ☐

d) I give permission for my visual image contained in photos or video gathered during the research to be used for the purposes described above. ☐ ☐

**3. Reuse of my data**

a) I give permission for the data that I provide to be reused for the sole purposes of future research and learning. ☐ ☐

b) I understand and agree that this may involve depositing my data in a data repository, which may be accessed by other researchers ☐ ☐

**4. Security of my data**

a) I understand that safeguards will be put in place to protect my identity and my data during the research, and if my data is kept for future use. ☐ ☐

b) I confirm that a written copy of these safeguards has been given to me in the University's privacy notice, and that they have been described to me and are acceptable to me. ☐ ☐

c) I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data. ☐ ☐

**5. Copyright**

a) I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented. ☐ ☐

b) I wish to be publicly identified as the creator of any audio or visual materials that I produce or co-produce as a part of this project. ☐ ☐

**6. Signatures (sign as appropriate)**

________________________ ________________________ __________

**Name of participant** (IN CAPITALS) Signature Date

If applicable:

For participants unable to sign their name, mark the box instead of signing ☐

I have witnessed the accurate reading of the consent form with the participant and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

________________________ ________________________ __________

**Name of witness** (IN CAPITALS) Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

**Name of researcher** (IN CAPITALS) Signature Date

**7. Researcher's contact details**

Provide the participant with a copy of the completed form either by email or hard copy as they prefer.