Sonic and Acoustic Links Between Real and Virtual Worlds in Audio Augmented Reality

Jacob D. Bhattacharyya

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

June 2026

This thesis is dedicated to those that I was fortunate enough to have in my life and unfortunate enough to lose. Firstly, it is dedicated to my grandfather Jim, whose influence on my life and the person I've become is difficult to put to words. Secondly, it is dedicated to my grandfather Debu, whose whole-hearted support for everything I turn my hand to helped give me the confidence to undertake this.

Finally – most importantly – it is dedicated to my brother Casper, whose absence I feel keenly every day. I am certain he would have gone on to do far greater things than this, but nevertheless I hope he would have found it interesting.

Abstract

Audio augmented reality (AAR) applications enhance our real world surroundings with virtual audio. To truly augment our reality, AAR applications must be aware of those surroundings, however AAR applications currently have no understanding of a user's aural environment. This thesis focuses on this fundamental gap in AAR, beginning by establishing a new definition for audio augmented reality, as a system that clearly connects the real world with virtual audio sources. It identifies "acoustic links" and "sonic links" as a way to create this connection and imbue AAR applications with an understanding of a user's real world aural surroundings. Six mixed-methods user studies are presented, exploring the potential of these links.

This thesis first explores acoustic links, where the acoustics of a user's surroundings inform an AAR system. Study 1 explores listener perceptions of different acoustic reproductions in a formal listening test. Study 2 builds on this to explore the plausibility of different acoustic reproductions in real-world spaces and when using AAR applications. The results from these studies suggest that even a simple reproduction of an environment's acoustics is enough to create virtual audio that is plausibly realistic, facilitating an acoustic link between real and virtual.

Study 3 begins exploring sonic links, where the sounds in a real world environment inform an AAR system, by evaluating AAR applications which respond to human-produced sounds as control schemes. It compares existing AAR control schemes with novel sonic controls, finding that speech and "sonic gestures" are viable ways to control AAR applications and by extension create a sonic link. Study 4 explores the use of environmental sounds to create sonic links in three different AAR applications, finding that such environmental sonic links create a more augmented experience and heighten a user's awareness of, and connection to, their real world surroundings. Study 5 builds on this by deploying existing sound classification models to drive these AAR applications, finding that existing detection models are capable of creating sonic links this way and maintaining this more augmented experience. Finally, Study 6 evaluates sonically linked AAR applications in uncontrolled, real-world scenarios over extended periods, finding that these sonic links continue to provide these benefits in the real world, can be facilitated now, and offer a stable experience over time.

The overall conclusion of this thesis is that acoustic and sonic links are viable ways to extend and enhance existing approaches to AAR applications, and are achievable using existing technology.

Contents

List of Tables

List of Figures

Acknowledgements

There are a great number of people who contributed to this work in ways both large and small, but all meaningful.

Firstly, I'd like to thank my supervisors, Steve Brewster and Alessandro Vinciarelli, for giving me the guidance needed to get this work across the finish line, making the past three years such a pleasure, and for our many compelling discussions on what this audio AR business even is. They've supported, encouraged, and challenged me to grow as a researcher and a person, and I couldn't have asked for a better supervisory team.

I'd like also to thank Professors Paul Vickers and Julie Williamson who examined this work, both for their efforts in reading the whole thing in the first place, and for a very comfortable and enjoyable discussion of it in the viva.

I'd like to thank my family – my siblings Robin and Dylan, and my parents, Rana, Oona, and Stephen, for their love and support not just throughout this work but everything that led up to it. This work, and the person I am today, would not be possible without the sacrifices they have all made. Dad, thanks for always believing in and encouraging everything I do. It's a tremendous thing to know there's always at least one person who thinks I've got what it takes and can't be persuaded otherwise. I'm not sure I'll ever understand the positive impact that's had on me. Smit, thanks for teaching me never to back down from a new challenge, for demonstrating to me the value of hard work, and for teaching me that sometimes there's nothing for it but to saddle up and kick it in the *** regardless of whether you feel like it. Mum, I'm not even sure where to begin – a sentence on a page hardly feels enough. Thank you for everything you've done to shape me into the person I am today, your continual love and encouragement, and for quietly serving as a (perhaps dangerous) example of just how much one person can achieve if they set their mind to it. I'm glad that at least one of us managed to finish a PhD. Sorry for messing yours up.

Thanks to my closest friends, Emma, Matt, Finn, Niall, and Patrick, for being there for me these past few years, and for being understanding when the workload meant I couldn't always do the same. Thanks also to Scott, David, Andy, and Billy for providing me with a bit of much-needed escapism and entertainment each week.

I'd like to thank the entirety of the Multimodal Interaction Group, for providing such a positive and supportive environment throughout this work. Thanks firstly to Graham, my F131 desk

neighbour, who has provided me with support, encouragement, good music and good conversation over the past three years. May we never have to see Schiphol Airport again! Thanks to Ammar, who is a continual source of inspiration as both a researcher and a human being. Thanks to Mark, for the advice, wisdom, and mentorship I've been lucky enough to receive from him. Thanks to Shaun, who has always had thoughtful and earnest words of encouragement for me and others. Thanks to Melvin, my very first friend in MIG; to Jesse, my newest; to Iain, Kieran, and Diego, my PhD compatriots and officemates, and to Tom, Joseph, Paddy, Jacqueline, Kat, Laura, and all the others I've had the pleasure of overlapping with, who made this group what it is and ensured the weekly meetings, Wednesday lunches, and Thursday drinks were always highlights of the week.

Beyond MIG I'd like also to specifically thank Eva Fringi, my SONICOM big sister, who has offered me advice and encouragement from my very first day, as well as Rune Jacobsen, who took me under his wing when I started and gave me both a crash course in PhD life and many enjoyable games of pool during the few months he spent in Glasgow.

Finally, I need to thank my partner Izzy. There is nothing I could write here to adequately express the contribution she's made to this work and the person I've become after ten years in her orbit, or how grateful I am to her. She encourages me every day to be the very best version of myself, and supports me in everything I do. While she didn't write any of the words in this thesis, I'm not sure I'd have gotten any of them in here without her or the sacrifices she's made for me over these past three years.

This thesis and the work comprising it was written and conducted largely under the influence of The Smashing Pumpkins, Megadeth, Brian Eno, Steve Reich, Aphex Twin, and Melt-Banana. The dedicated reader may wish to replicate this aural environment for themselves.

Declaration and Contributing Papers

Unless otherwise explicitly stated, the research presented in this thesis is entirely the author's own work.

Study 1 and 2 have been published as a paper at the Audio Engineering Society Audio for Games Conference 2024: Jacob Bhattacharyya et al. "Investigating the Influence of Environmental Acoustics and Playback Device for Audio Augmented Reality Applications". In: *Audio Engineering Society Conference: AES 2024 International Audio for Games Conference*. Tokyo, Japan: Audio Engineering Society, Apr. 2024, p. 10

Study 3 has been published as a paper at the ACM CHI Play 2025 conference: Jacob Bhattacharyya, Alessandro Vinciarelli, and Stephen Anthony Brewster. "Sonomancer: Exploring Sonic Control Schemes for Audio Augmented Reality Games". In: *Proceedings of the ACM on Human-Computer Interaction* 9.6 (Oct. 2025), pp. 976–994. ISSN: 2573-0142. DOI: 10.1145/3748629

Study 5 has been published as a paper at the IEEE ISMAR 2025 conference: Jacob Bhattacharyya, Alessandro Vinciarelli, and Stephen Brewster. "Birds of a Feather Augment Together: Exploring Sonic Links Between Real and Virtual Worlds in Audio Augmented Reality". In: *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, Oct. 2025, pp. 1490–1500. ISBN: 979-8-3315-8761-1. DOI: 10.1109/ISMAR67309.2025.00153

Glossary of Specialist Terms

This section provides a short summary of terms that may be unfamiliar to the reader. These descriptions are designed to provide a basic, accessible understanding, enough to follow the ideas presented in this research, rather than a detailed overview.

- **Acoustics:** Acoustics describe the influence of a space on sounds occurring within it. Echo or reverberation are examples of acoustic effects.

- **Acoustic Transparency:** Headphones and earphones can be acoustically transparent, meaning they do not block the listener's perception of their surroundings. This can be achieved **passively**, where the ear canal is not blocked (such as bone conduction headphones or audio glasses), or **actively**, where a microphone captures the surrounding sounds and plays them back to the listener alongside the virtual audio (such as Apple AirPods' Transparency Mode).

- **Ambisonics:** Ambisonics is a representation of a sound field. An Ambisonics soundfield is 3-dimensional, representing a full sphere, rather than just the horizontal plane as in most conventional speaker formats. An Ambisonics soundfield is speaker-agnostic, and can be decoded to effectively any array of speakers, including binaural playback over headphones. An Ambisonics soundfield has a numerical order n. The higher the order, the higher the resolution of the soundfield, and the higher the number of audio channels used ($(n+1)^2$) [199].

- **Anechoic/Echoic:** The terms anechoic and echoic describe audio recordings which have, or do not have, reverberation of a room included. Most recordings are echoic, including some amount of reverberation from the space where the recording was made. Some recordings are anechoic, and do not include any reverberation, such as some synthesised sounds, or recordings made in anechoic chambers which have no reverberation.

- **Binaural:** Binaural playback is a specific method of playing audio over headphones that incorporates the listener's head-related transfer function. Sounds played back binaurally will reproduce the hearing cues relied on by the human ear to locate sounds, and so will appear to be coming from a position around the listener, rather than from the headphones.

- **Head-Related Transfer Function (HRTF):** A head-related transfer function models the influence of a listener's head on the sounds they hear. The HRTF can be used to reproduce this effect with virtual audio, and the resulting sound will more closely represent an equivalent real sound.

- **Room Impulse Response (RIR):** A room impulse response is a representation of the acoustics of a space. By convolving the room impulse response with an audio file, the resulting audio will sound as if it was being played in that space. This works best with anechoic source audio already free of acoustic effects.

Chapter 1

Introduction

There is no agreement on what defines audio augmented reality[1]. Scholars generally agree that it, like other forms of augmented reality, blends real and virtual elements together [6, 114]. In the case of audio augmented reality (AAR), these virtual elements are auditory – systems which use auditory waypoints for navigation [111, 24], introduce virtual sounds to enhance our physical surroundings [78, 183], or create new entertainment experiences [35, 22]. However, opinions diverge on what makes a system with virtual audio into AAR, or when an auditory experience changes from listening to audio to experiencing AAR. Without a clear understanding of what AAR is, it is difficult to identify its potential, or to identify gaps in our knowledge.

This lack of clarity may explain why AAR applications remain a rarity, despite having a number of strengths compared to visual AR. For one, AAR is much more accessible than visual AR systems, achievable with little more than a smartphone and a pair of earphones, compared to expensive and bulky AR headsets. By relying on our hearing, AAR also offers a 360° experience at all times, and is currently much closer to achieving perceptual plausibility than visually-oriented AR systems. Advances in spatial audio systems allow the believable positioning of sounds in three-dimensional space, while simulations of reverberation can embed such sounds plausibly in a listener's surroundings [132].

Despite AAR's accessibility and unique strengths, it has yet to achieve the same level of impact or success as visual AR systems. This thesis identifies two major problems with the current state of AAR which may contribute to this. The first, as discussed, is the lack of a clear definition for AAR. To fully take advantage of AAR's potential, we must first understand clearly what AAR *is*. The second is a fundamental gap with how AAR systems relate to the real world they seek to augment.

[1] Which makes it difficult to introduce this thesis.

1.1 The Fundamental Gap in AAR

An AR system can augment reality best when it has an awareness of reality, hence why many visual AR applications attempt to identify specific objects in a user's environment, or map the user's surroundings using technologies like Simultaneous Localisation and Mapping (SLAM) tracking [19]. By doing so, visual AR systems can position virtual elements appropriately in the real environment, react to elements of the environment, or simulate light and shadow appropriately for the space to enhance the plausibility of virtual elements. Without such an awareness of the real world, an AR system is limited to simply overlaying virtual elements atop the real world rather than fully integrating the two into a cohesive whole.

Schraffenberger, in her discussions on the nature of AR, argues that augmented reality arises when virtual elements are connected to the real world [163, 161]. Schraffenberger provides two examples of how real and virtual can be connected: spatially, or through their content. Although Schraffenberger did not consider the auditory domain when setting out this definition, it goes some way to defining AAR also. Listening to music does not constitute augmented reality, as the music has no link to the real world. However, if the music emanates from a virtual stage in the listener's living room, this would constitute a spatial connection between real and virtual and, according to Schraffenberger's definition, augmented reality. Likewise, if between songs the virtual musicians complimented the listener on their interior decor, this would constitute augmented reality by way of a content-based connection.

When one considers existing or hypothetical AR applications through this lens, other connections or links between the real and virtual become clear. The most obvious is a geographic link, which many existing AR applications deploy, tying elements of the virtual experience to a specific location in the real-world. Pokémon Go is perhaps the most widely known example of this [135] in AR, while in AAR soundwalks and geolocated music tie virtual audio to real-world locations [183, 78, 23].

One possible link between real and virtual that is both underexplored and highly relevant to AAR is an **auditory link**. Just as audio is an AAR system's output modality, so too could audio be leveraged as a connection between real and virtual elements to facilitate augmented reality. Virtual sounds in AAR are intended to mesh with our real world surroundings, and an AAR system needs to be aware of those sounds to integrate virtual elements alongside them. Our auditory reality cannot be augmented without the system having some understanding of the sounds it contains.

Our aural surroundings are composed of two components – the sounds in our environment (birdsong, cars, human speech, and thousands of other sounds we hear throughout our lives) and the acoustic influence of our environment (the reverberation, echo, absorption and other ways sound is affected by physical space). Either one of these could be leveraged as an auditory link on its own, and both are critical for AAR going forward.

Understanding the specific sounds in our environment (a *sonic link*) is important on a fun-

damental level to make informed decisions about how to present virtual audio. If a system has no awareness of real world sounds, it might site important virtual sounds such that they are masked by other environmental sounds, or *vice versa*. Additionally, real-world sounds could drive AAR applications in other ways, just as real-world objects do in visual AR. Augmented music players could alter music playback based on environmental sounds, for example to maintain audibility of certain elements, or even by acting as an augmented DJ, ensuring that a user's music queue was upbeat in busy urban environments and relaxed in quiet nature environments. Sonically linked AAR systems could provide a level of curation over the real-world soundscape, highlighting key sounds to a user or removing or replacing irritating or meaningless ones. The varied sounds produced by human users could even be leveraged as additional input methods for AAR applications.

The acoustics of a user's environment are also critical for an AAR application to understand (an *acoustic link)*. To seamlessly add virtual audio into a user's environment, a central goal of AAR, it becomes important to simulate the acoustic influence of a space, just as in visual AR it is important to accurately simulate light and shadow. Mapping the acoustics of a space and applying them to virtual audio is fundamental, but acoustics could also drive future AAR applications. AAR audiobooks could heighten user immersion by siting virtual characters within a user's acoustic surroundings, or bring the real world into the narrative by applying the acoustics of the story environment to the user's surroundings.

This thesis asserts that these auditory links are critical to AAR going forward. Without an awareness of a user's real world surroundings, AAR applications can only overlay virtual audio over the real soundscape, rather than integrate real and virtual into something more significant. While the focus of the work presented here is on AAR contexts, the benefits of exploring these auditory links could extend into wider extended reality (XR) also. XR applications encompass the broader categories of virtual, mixed, and augmented reality [182], and are inherently audiovisual. Understanding how the real-world soundscape can be leveraged in AAR could have applications wherever audio is deployed in XR. By exploring the potential of auditory links in AAR, this thesis provides the first step in closing this fundamental gap in AAR.

1.2 Thesis Research Questions

This thesis explores the following overall research questions:

- RQ1: How can an acoustic link between real and virtual elements be created in audio augmented reality?
- RQ2: How can a sonic link between human-produced sounds and virtual elements be created in audio augmented reality?

- RQ3: How can a sonic link between environmental sounds and virtual elements be created in audio augmented reality?

1.3 Thesis Statement

Augmented reality is becoming an increasingly important way for users to interact with the digital world. However, while visual AR systems are seeing continual improvement and growing popularity, audio AR systems are comparatively niche, despite being far closer to achieving perceptual plausibility, offering benefits such as a 360° field of audition, and requiring little more than a pair of earphones to experience. This thesis argues this is for two reasons: AAR itself being poorly defined, and a lack of understanding of how an AAR system can be informed by the user's aural surroundings, just as visual AR systems are informed by visual surroundings. This research makes contributions to both problems, firstly by providing a new definition for AAR that unifies and clarifies existing definitions and discussions, and secondly by exploring what we term 'auditory links' between the real world and virtual AAR elements. This work on auditory links covers both the acoustic domain, exploring the level of fidelity needed to plausibly simulate the real world's acoustics on virtual sounds, and the sonic domain, exploring how AAR systems can be driven by sounds created by the user, or sounds emitted from the user's environment. These contributions are made over six exploratory, mixed-methods laboratory, field, and longitudinal studies, and provide guidance for the development of AAR systems going forward.

1.4 Thesis Structure

The remainder of this thesis is structured as follows:

- Chapter 2 provides a detailed review of existing work relevant to this thesis.
- Chapter 3 discusses the nature of audio augmented reality and proposes a definition for AAR that underpins the remainder of this thesis.
- Chapter 4 explores RQ1 and acoustic linking. It describes two studies on virtual acoustic reproduction, exploring the level of fidelity required to embed plausible virtual sounds in the real world.
- Chapter 5 explores RQ2 – the potential of AAR applications to respond to real world sounds produced by the user. It describes a study evaluating the use of different forms of human-produced sounds to control AAR games.
- Chapters 6 and 7 explore RQ3, how AAR applications can be sonically linked to sounds in the real-world environment. Chapter 6 describes two iterative studies exploring sonic

linking in both a best-case scenario, and with technology available today. Sonic links are assessed with multiple real-world sounds and multiple forms of application. Chapter 7 concludes the experimental work in this thesis by evaluating these applications with real users in a real-world, in-the-wild study over a one week period.

- Chapter 8 provides a discussion on the overall thesis findings and their implications for AAR and wider XR.

Chapter 2

Literature Review

This chapter sets out a literature review exploring the existing work related to this thesis and the contributions it presents. As discussed in Chapter 1, AAR has had limited impact despite offering significant potential in an increasingly augmented world, and this thesis identifies two key gaps in our understanding of AAR.

The first is what AAR actually *is*. Existing work has yet to provide a compelling definition for AAR, with most authors defining it broadly as the introduction of virtual audio into the real world – in doing so implying that existing audio experiences like music, radio, or phone calls are AAR – or as an auditory-focused subcategory of AR, which provides little clarity on when audio experiences become AAR. As the starting point for the original contributions made by this thesis, Chapter 3 discusses this in depth, and sets out the definition of AAR used throughout this work.

The second is a fundamental gap within AAR, the fact that currently AAR applications have no awareness of the real-world soundscape they are designed to augment, no **auditory link** between real and virtual. This chapter provides an overview of this underlying problem – the existing application areas AAR has been deployed in, an overview of how these proposed auditory links have or could be used, and the underlying technology that enables AAR applications, including those presented throughout this thesis.

2.1 Audio Augmented Reality Applications

Cohen *et al.* are considered to have first introduced AAR in 1993, by presenting a system which augmented a silent, real-world telephone with the spatialised sound of a virtual telephone [40, 114]. While Cohen *et al.* presented this as a technical demonstration of the ability to insert virtual sounds into the real world, and primarily envisioned this being applied to telepresence applications for remote robot control, AAR has since been applied to a variety of application areas, though has not yet achieved widespread impact.

2.1.1 Audio Augmented Navigation

One of the most common application areas for AAR has been navigation. Being audio-centric, AAR applications represent an accessible platform for users who are blind or visually impaired (BVI), and a number of AAR applications designed to help these users navigate the world have been proposed. Such approaches to navigation can also be used by sighted users, and provide a potential alternative to other navigational solutions, particularly for scenarios like driving where the user should keep their eyes on the road [24].

Loomis *et al.* presented one of the first such systems – the Personal Guidance System – which presented spatial audio to the user over headphones [111]. The navigation system was contained in a backpack worn by the user, and the system presented users with binaural audio of synthesised speech to demonstrate the direction of a waypoint. More recently, Blum *et al.* developed a smartphone-based AR system for BVI users [24] which sonified nearby points of interest, facilitating exploration rather than navigation to a specific location. Users wore a smartphone in a lanyard around their neck with a pair of earphones, and the system used information from the phone compass and GPS satellites to track the user's position, and presented spatialised audio cues to describe points of interest around the user. These cues consisted of synthesised speech and auditory icons describing nearby locations. While the authors found some difficulties due largely to early smartphone hardware, the system showed promise in tests with both sighted and BVI users.

Albrecht *et al.* [4] explored a musical navigation system, where the user listened to music freely, and the system spatialised the music to indicate directions. The authors evaluated both a bearing-based approach where the music came from the direction of the destination and left navigation to the user, and a turn-by-turn system that re-spatialised the music to provide specific navigation instructions. They tested the system with both pedestrians and cyclists and found it promising with both groups, with users enthusiastic about using such a system in the future.

Heller and Schöning [81] instead manipulated a single component of the music. By using multitrack recordings of music, their NavigaTone system altered the panning of a specific instrument in the music to indicate a navigational bearing. The NavigaTone system was found to enable similar navigational performance as simpler systems that pan all music in the direction of navigation, but enabled a more natural listening experience. The authors also highlighted some potential concerns for real-world usage, like the need for the instrument to be playing whenever a change in direction is needed, for example at an intersection, which is not always guaranteed depending on the musical track.

Although designed to enhance an overall journey rather than offer directions, musical manipulation was also explored for car journeys by Kari *et al.* [97], who developed the SoundsRide AAR system. The SoundsRide system analysed the route to the vehicle's destination, identified 'sound affordances' along the route, and dynamically remixed the user's music such that notable events in the music (e.g. transitions, beat drops) occurred at the same time as notable events on

the route (e.g. exits and entrances of highways and tunnels). The prototype system evaluated in the user study was well-received, with users finding it enhanced immersion and overall experience. In the commercial realm, car manufacturer Kia recently presented a system that runs object detection algorithms on vehicle cameras, detecting natural landscape features and interpreting them alongside vehicle movement into a musical symphony designed to provide BVI passengers with an understanding of their visual surroundings [100].

While many AAR navigation systems are focused on outdoor navigation, some authors have also explored methods for enhancing indoor navigation. Kaul *et al.* developed an AR system which used object detection to identify key objects and obstacles around the user, and presented their locations sonically, as an alternative to traditional white cane exploration [99]. Blessenohl *et al.* developed an indoor navigation system which used a head-mounted depth camera to detect walls, corridors, and obstacles and present them to the user using spatialised audio [21]. While their prototype used a laptop, headphones, and bespoke hat, they note that their system is theoretically achievable using existing consumer technology, and a similar application, waveOut, has been brought to market as a smartphone app by the company Dreamwaves[1]. waveOut presents spatialised auditory beacons that lead the listener to their destination, and relies on the smartphone camera and/or the use of Apple AirPods to track the user's orientation.

While these navigation-centric applications have a strong connection between the virtual audio heard by a user and the real world, this connection is primarily based on visual elements (such as in indoor navigation scenarios) or geographic positions (such as in outdoor navigation). None of these applications deployed an auditory link, despite many of these scenarios potentially benefitting from one – an accurate reproduction of environmental acoustics could site virtual waypoints more believably in a user's real environment, or the volume or tonality of such waypoints could be altered on the fly to ensure audibility based on the nearby soundscape, for example.

2.1.2 Audio Augmented Spaces

Another common theme within AAR has been the augmentation of physical spaces with virtual audio. Vazquez-Alvarez *et al.* [183] developed an AAR 'sound garden', augmenting the Municipal Gardens of Funchal, Madeira with virtual audio sources. The system ran on a Nokia N95 smartphone, and used a GPS receiver to determine the user's position and an IMU sensor pack to report head orientation. Virtual sounds were positioned at five different locations in the gardens, corresponding with key landmarks, and presented over a pair of over-ear headphones. These consisted of synthesised speech providing information on the landmark, and Earcons (non-verbal audio cues that provide information to the user [29]). The authors compared different combinations of these cues, with and without spatial audio, which would be triggered when the user was within a certain distance of the landmark. They found that the use of spatialised

[1] https://www.dreamwaves.io/

audio cues in this way encouraged users to walk more, spend more time in the park, and actively search for the landmarks.

Hazzard *et al.* [78] created a two-part 'sound walk' in Nottingham city centre, where pairs of users would each walk a pre-determined route through the city, experiencing an AAR narrative as they did so. The authors used bone conduction headphones that allowed the users to hear the virtual audio alongside the real world, and found that pairing virtual audio with the existing ambient world sound enhanced the user experience.

There have been many other examples of these sound walks that tie virtual audio to real world locations. One of the most notable in recent years takes place around Notre-Dame Cathedral in Paris, France, developed by Katz *et al.* after the 2019 fire closed the cathedral for renovations [98]. The soundwalk, 'Notre-Dame Whispers', was launched in 2023, and is run on consumer smartphones from a mobile app, presenting the listener with narration tied to different locations around the cathedral. While these soundwalks tie virtual audio to real-world locations, they currently do nothing to tie these into a user's aural environment, overlaying their virtual audio atop the real world rather than truly augmenting it.

The Notre-Dame soundwalk is not the first exploration of AAR for historical and cultural use-cases, with museum tours and heritage applications being another common application area. One of the first such systems was proposed by Bederson [11], who presented a prototype AAR tour guide for museums which relied on IR beacons located at specific exhibits. When using the system, the listener could approach a museum exhibit, and once within a certain proximity the system would present audio descriptions to the user, allowing them to freely explore the museum and still receive additional auditory information.

McGookin *et al.* [123] explored the potential of AAR augmentations for 'un-stewarded' heritage sites – sites of cultural and historical interest which are unstaffed – by developing an AAR application tied to the Antonine Wall in Scotland, an archaeological site with ruins from the Roman Empire. Their application used GPS signals to tie virtual audio, which consisted both of sound effects illustrating different buildings and points of interest at the side, and virtual characters designed to fill the role of re-enactors that stewarded sites often have.

McGookin also explored AAR as a way of blending our digital lives with the real world, with the PULSE system [122]. The PULSE system would connect with a user's Twitter account, and as they moved through the surroundings would present a text-to-speech rendering of a tweet sent from nearby (based on geotagging information). The system was evaluated in a two-week field study, and participants noted the system allowed for a new perspective of their surroundings, and an exposure to different communities within the same space. As the system was audio-only participants found it often receded into the background, and allowed for serendipitous discovery of new aspects of familiar environments.

A system similarly designed for serendipity was 'Audio Aura', proposed by Mynatt *et al.* [129, 130]. The Audio Aura system was designed to provide office workers with background

audio cues that kept them informed of useful information without becoming a distraction, such as waiting email messages, and whether colleagues have been in the office today. While the authors did not conduct a formal evaluation, they noted a positive overall reception with those who tested the system. They also experimented with allowing users to leave auditory messages at their office, for example as a note to say they'll be back soon. A similar system of auditory notes was proposed by Harma *et al.* [74], who referred to it as an 'Auditory Post-It'. Their system allowed the user to leave auditory messages tied to specific coordinates around a user, though their prototype was referenced only to a headtracker and so the prospect of tying messages to different real-world locations was not fully explored.

Much like AAR navigation applications, these systems seeking to augment physical spaces only tie virtual sound into the real world geographically, rather than any other sensory aspect of the space. A real-world environment is more than just a set of coordinates, it contains sights, smells, and crucially for AAR, sounds. Auditory connections between the virtual audio and real world could deepen such experiences, for example by having virtual and real sounds share an acoustic environment, or having the system respond to nearby sounds. While we can easily present a soundwalker with information about the bells of Notre Dame Cathedral as they stand at a specific viewpoint, the experience may be richer when this discussion is prompted by the bells themselves sounding, for instance.

2.1.3 Audio Augmented Objects

Rather than augmenting a location or space with audio, there has also been some work augmenting specific objects with virtual sound. Frohlich and colleagues explored the concept of audio-enabled paper. They developed 'audioprints' – physical printed photographs that had accompanying audio, triggered either through a printed pattern on the back, an embedded chip, or a computer vision system that identified specific images [61, 60]. They have also experimented with audio augmented newspaper [62], where written articles can be paired with supplementary audio using embedded electronics. The interactive newsprint connects wirelessly to an audio device worn by the reader/listener, and the virtual audio can be triggered using capacity buttons on the page. They conducted a user study evaluating these interactive newsprints, and found the system offered new affordances to the user by adding additional content, other perspectives, and even making the newspaper feel more 'local' by adding regional dialects. Grasset *et al.* demonstrated a similar system for a multimodal augmented book [69]. Using a handheld viewing device and a pair of headphones, readers could see additional augmentations when reading a picture book, including three-dimensional audio spatialised to align with elements on the page and thematically linked to the book's content.

Schraffenberger and van der Heide approached the notion of an audio augmented object from an alternative angle – rather than using virtual audio to augment a physical object, they used virtual audio to imply the existence of a physical object [162]. Using a LeapMotion controller,

they created an invisible cube that would float above the controller. When 'touching' the cube, spatial audio cues played from the position where it was touched. While the authors did not conduct a formal evaluation of the cube, they found that it created the sensation of a tangible object, despite both the audio and the cube itself being entirely virtual.

As well as associating additional virtual audio with an object, another approach has been to alter the characteristics of the sound an object already produces. Weger *et al.* proposed a system for 'auditory contrast enhancement', altering aspects of a real sound to enhance important characteristics while maintaining its overall character, for scenarios where sounds contain relevant information such as when using a stethoscope [188, 189]. They found that their system could increase the contrast of a sound without negatively impacting the sound's plausibility. Bovermann *et al.* demonstrated a prototype system for augmenting the sound of a keyboard with digital information. By attaching a contact microphone to the keyboard, the sound of the user typing would alter to reflect weather data, illustrating changes in pressure or cloud movements [27]. Similarly, Bustoni *et al.* developed a system for altering or replacing the sound made when a coffee cup is tapped, with a view to leveraging it as a method for delivering information in the future [33]. They experimented with altering the reverberation on the tapping sounds, and replacing the tapping sounds with sounds of water, wood, metal, and different instruments. By doing so with a very low latency, they were able to do so while maintaining a real-time interaction, but did not explore what forms of information the system could convey to the user. These are some of the few examples of an AAR system where not only the virtual elements are auditory, but the real world elements the system augments are auditory also, though currently represent only early explorations which have not investigated these forms of real-virtual links in depth.

2.1.4 Audio Augmented Music

As one of the most important forms sound can take, music has also been explored in relation to AAR. As well as the examples mentioned earlier, there have also been multiple instances of artists releasing 'geolocated music', for example the Swedish band John Moose whose debut album was released though a mobile app that required the user to be physically located in the woods in order to listen [22]. The musician Bluebrain also released a location-aware album, 'The National Mall', where different tracks were tied to different locations in the park of the same name in Washington D.C. [23]. As a result, each listener's experience of the album becomes unique based on how they choose to explore the space while listening.

Lokki *et al.* also proposed an AAR music player, Acu-Notch, where the system would identify important real-world sounds in the environment (e.g. sirens) and their location, and reduce the volume of the listener's music in the corresponding area of the stereo field to create an auditory notch around the sound [110]. The presented paper was focused on the audio notching, and the sound reactivity part of the system was never implemented or evaluated, though the

concept is a compelling use case which is revisited in Chapters 6 and 7 of this thesis.

2.1.5 Audio Augmented Games

Just as games have been a central application of XR technology, so too have games been a common application area in AAR. One of the first examples was presented by Lyons *et al.*, 'Guided By Voices', which relied on a modified MP3 player, microcontroller, and an RF receiver to enable an AAR roleplaying game [112]. RF beacons are set up at various locations in the play space, which trigger narrative audio as the player approaches, while the system tracks the locations the player has visited to adjust the flow of the narrative.

Ekman *et al.* presented one of the first examples of a multiplayer game with AAR elements, 'The Songs of North', inspired by Finnish mythology [51]. Players took on the role of shamans searching for artifacts located throughout the world, with the ability to hear into an auditory 'spirit world', using a Nokia N-Gage mobile phone. This spirit world was heard alongside visual elements on the device, though audio elements were always available and visuals could only be updated intermittently, by casting spells using the phone interface. In this way, the game could switch between a passive, AAR experience when the player was not casting spells, and a multimodal experience when they were playing more actively. The player's position in the real world was tracked using GSM positioning.

Paterson *et al.* also explored the use of virtual audio as indicator for a supernatural dimension in 'Viking Ghost Hunt' [141]. Players take on the role of a paranormal investigator searching for ghosts at various historical sites around Dublin, Ireland. As they explored these sites the player encountered ghosts, using a smartphone interface to decode messages and information from the ghosts to gather evidence.

Rather than creating an entirely new AAR game, Chatzidimitris *et al.* reimagined an existing game for the medium when they presented 'SoundPacman' [35]. Using street map data and the GPS sensor in a mobile device, the game maps classic game Pacman to real city streets around the user, positioning 'cookies' that need to be collected and ghosts that need to be avoided on the game map. These elements are then presented to the user binaurally. The authors outfitted players with an EEG brainwave monitor, and found that the presence of the ghost sound notably increased players' alertness, suggesting a possible increase in immersion.

Another notable example in recent years is 'Audio Legends', presented by Rovithis *et al.* as a collection of game scenarios used to assess the potential of gesture controls for AAR [153], as most other audio and AAR games rely on simple controls such as physical movement and simple interactions with a controller or mobile device. The minigames were designed to be played at a historical site in Corfu, based on the legends of the local Saint Spyridon, and ran on an iPad Air paired with a Sennheiser Ambeo headset. Players moved physically throughout the space, and completed gestures with the iPad, while game audio was presented over the headset. 29 players gave high ratings for overall experience, usability, and immersion, and the gestural

controls were well-received. This represents one of the only formal assessments of AAR game control schemes, and the potential of other control schemes remains unknown.

While there has been some evaluation of AAR games academically, there are few examples in the commercial space. Most that did exist emerged from the now discontinued Bose AR platform. 'Dead Drop Desperado' was a two-player game built for Bose AR, with one player wearing a Bose AR device such as the Bose Frames audio glasses [142]. The game scenario is a Wild West gun duel, and the other player fires bullets through the mobile app that the AR player must listen for and physically dodge. 'The Clairvoyant' was a narrative experience where the player took on the role of a supernatural medium [190]. Similarly to Viking Ghost Hunt and The Songs of North, the game used the auditory mode as a representation of an alternate world or dimension inhabited by spirits, and players were tasked with communicating with these spirits through gestures (e.g. head shakes and nods).

Outside of Bose AR, another notable example is 'PairPlay', a two-player iOS game / story experience. Each player takes one of a pair of wireless earbuds, and both experience different sides of a co-operative narrative experience overlaid on a normal home environment. As with the Bose AR examples, PairPlay requires specific hardware (Apple AirPods in this case), and there are no significant examples of hardware-agnostic commercial AAR games currently.

Although commercial AAR games are rare, there are some commercial 'audio games', which are audio-only though lack an augmented reality component. Many are designed as an accessible gaming experience for BVI players, while others are designed to put sighted players in a BVI person's shoes, such as the 'blind swordsman' scenarios deployed in games like 'The Vale: Shadow of the Crown' [174] or 'A Blind Legend' [49]. There have also been some games created for smart speakers like Amazon's Alexa, for example 'Skyrim Very Special Edition' [177] or the Alexa version of 'Jeopardy!' [87]. In these smart speaker games, players interact using voice, one of the only examples of an audio or AAR game responding to real-world sound.

There are also a handful of audio exercise games which could be considered an example of AAR. The most widely known is 'Zombies, Run!' [170], designed to help motivate runners by simulating the sound of a pursuing zombie horde. The game also features specific missions tied to running routes that add an overarching narrative to the experience, and through continued use players are able to collect in-game resources. 'Run the Realm' [109] offers a similar premise, providing a fantasy narrative that accompanies the player on their walk, run, or cycle.

Like many of the other forms of application mentioned so far, the vast majority of these AAR games have no awareness of a user's real-world surroundings, and would behave the same were a user playing in a quiet living room or a noisy park environment, despite these spaces having radical sonic differences that could be leveraged to enhance the experience. The acoustics of virtual game sounds could be altered to match the real-world environment (or *vice versa*), or the sounds in a user's environment could be replaced or morphed to become part of the immersive experience, however there has currently been no exploration of the potential of real-world audio

in these experiences.

2.1.6 Summary

While AAR has been deployed in a variety of application areas, it has not achieved the same level of success or impact as traditional AR. Across the application areas outlined here it is also clear that AAR closely follows visual AR in terms of how it integrates with the real world, most often referencing a user's physical or visual surroundings. Despite these applications seeking to blend virtual audio with the real world, there are currently very few AAR applications which respond to the user's aural environment. This provides a strong motivation for the research questions introduced in Chapter 1, and Chapters 4 through 7 of this thesis explore the potential of introducing auditory links to resolve this clear gap in the field.

2.2 Auditory Linking in Audio Augmented Reality

A user's aural surroundings can be considered as two components: the acoustic effects of the space (reverberation, echo, absorption etc.), and the sounds within it. These aspects are independent of one another, and AAR systems could theoretically respond to either component, however little work has explored this potential.

2.2.1 Acoustic Linking

A key goal in AAR is the seamless presentation of virtual elements as being plausibly real, and a crucial part of achieving this is the reproduction of the acoustics of the surrounding environment. With the exception of highly specialised spaces, all spaces exert some influence on the sounds occurring within them, by absorbing acoustic energy, and reflecting sounds to create reverberation or echoes. The human brain is highly attuned to these acoustic cues, and simulating these effects accurately is central to tricking an AAR user's brain into perceiving virtual sounds as real. However, as recently as Yang *et al.*'s review of the field in 2022, a majority of AAR systems neglect to model these acoustic effects [198]. Acoustic modelling in this way can also be computationally expensive, and there is not yet a clear understanding of how precisely these effects must be simulated for AAR scenarios.

In Yang *et al.*'s review, the authors identified 62 AAR systems, of which 45 were designed for indoor use and 17 were designed for outdoor use. The authors found that 33 of the indoor AAR systems did not feature acoustic modelling of the environment (or did not describe it), and none of the 17 outdoor systems featured acoustic modelling either. The 12 systems which did model the acoustic environment did so with varying levels of accuracy, ranging from applying simple artificial reverberation to acoustic simulation through a 3D room model. The results of

the review show that in AAR, acoustics are often an afterthought, and there is limited consensus on how they should be deployed.

The presence of acoustic cues have been shown to provide a number of benefits to the listener. One of the key factors acoustic cues influence is the externalisation of a sound. When listening to virtual audio over headphones, sounds can often be perceived as being internalized – located inside the listener's head instead of outside. The addition of reverberation cues can mitigate this issue, although it is still unclear what cues are most important [15]. Grimaldi *et al.* found that externalisation was significantly improved by the addition of 'early reflection' cues, which arise in the first 80ms of reverberation [71]. This is corroborated by experiments conducted by Begault and colleagues, who found no difference in externalization between a full reverberation simulation and reproducing early reflections only [13]. Leclère *et al.* conducted three experiments on sound externalization and found that the addition of reverberation cues only improved externalization when they resulted in interaural differences [106]. It is also important for these cues to match the listener's environment, as it has been shown that a mismatch between virtual reverberation cues and a listener's physical room can create a 'room divergence effect'. If virtual and real spaces are different, externalization is negatively affected, while if virtual acoustic cues match the listening environment externalization can be improved further [193]. As AAR hinges on siting virtual audio within a 3D, real-world environment, improving the externalization of virtual sounds is key for an AAR system's outcomes.

As well as better externalizing virtual audio, reverberation cues can also affect a listener's ability to localise virtual sounds. The direct-to-reverberant ratio (DRR), the blend of direct and reflected sound waves that we hear, is a key cue that our auditory system uses to estimate the distance to a sound source [32], with more reverberant sounds being perceived as farther away [124, 12]. Without reverberant cues, it can become extremely difficult to accurately gauge distance to a sound [136]. Reverberation cues also affect our ability to pinpoint a sound's location, as additional reflections of a sound arrive at our ears from multiple different angles. The angle of these reflections relative to the sound source can reinforce or muddy the localisation cues we rely on [76], and there has been evidence both of localisation ability degrading [12, 149, 67] and benefitting [13] from the inclusion of reverberation.

In the context of virtual and mixed reality environments, acoustic cues can also enhance the subjective user experience. Potter *et al.* explored the relative importance of visual and auditory cues for immersion in a virtual reality environment [147]. Comparing different combinations of audio features (spatial audio, acoustic simulation) and visual resolution, they found that including acoustic simulation resulted in the same increase in immersion as a five-times increase in video resolution. Other studies have shown that reverberation cues can increase the listener's sensation of envelopment, the feeling of being surrounded by an environment; and increase their overall immersion in an acoustic space [127, 96, 63]. Finally, acoustic cues can contribute to how plausible a virtual sound appears [133, 132]. In this context, plausibility is defined as how

closely a virtual sound matches the listener's expectation of how that sound should be perceived for a given circumstance [107]. Even if the audio source would *actually* sound different, plausibility can still be achieved if it sounds as the listener *believes* it should.

While it is clear that reverberation cues contribute to a sense of plausibility, it is not currently clear how accurate or detailed these reverberation cues have to be in order to achieve plausibility. Brinkmann *et al.* conducted a round robin evaluation of five leading acoustic modelling software packages and conducted a user study to assess the plausibility of the resulting auralisations [30]. Participants compared the model outputs for three different test spaces with a real-world measurement for each test space. While they found many errors in the modellers' outputs, the majority of the auralisations were perceived by study participants as being plausible. Notably however, the participants were only shown a picture of the test spaces rather than experiencing the spaces themselves.

Reverberation cues are commonly rendered in the Ambisonics domain, and multiple studies have been conducted to explore how Ambisonics order (spatial resolution) affects perception of reverberation. Ahrens and Andersson found no perceptual benefits to rendering reverberation above 8th-order Ambisonics [3], while Enge *et al.* compared reverberation reproductions in a VR context, and found no significant change in plausibility between 3rd-order reproductions and 7th-order reproductions [53]. Engel *et al.* compared reverberation cues rendered from 0th to 4th-order with a non-reverberant signal, finding that under their test circumstances listeners could not perceive differences in reverberation above 1st-order Ambisonics [54]. The study was conducted online, with participants using uncontrolled hardware and again shown only an image of the simulated room instead of experiencing it themselves. In a follow-up study, the authors found that differences could be discriminated up to 2nd or 3rd-order [55]. In this study the authors tested one group of participants in a lab environment, and the other group in the physical space that was simulated, but did not report details of this analysis due to paper length constraints. Neidhardt and Zerlik have previously demonstrated that exposure to a real version of a sound makes listeners more critical when judging the plausibility of a virtual counterpart [134], and so determining optimal reverberation simulations will be best done in real-world spaces rather than lab environments.

While most work in this area has focused on accurate reproduction of a given environment, Schneiderwind and Neidhardt also recently explored how far reverberation can be manipulated before plausibility is impacted. They found that late reverberation times could be adjusted within a range of 80–120% of normal without plausibility breaking down [160]. While the most obvious form of acoustic link in AAR is applying real-world acoustics to virtual audio as discussed above, this also implies the potential of future work altering the acoustics of real sounds to match a virtual space, for example in narrative or artistic AAR applications.

Overall, while work has been undertaken to explore the potential of real-world acoustic simulation for AAR elements, there remain a number of unanswered questions. There is still a lim-

ited understanding of how accurately environmental acoustics must be reproduced, particularly in real-world environments and when using an AAR application, rather than conducting lab-based listening exercises. Additionally, reverberation perception has primarily been evaluated using headphones, and there has been little exploration of how acoustic reproduction considerations may change when using audio glasses or other acoustically transparent playback devices which are necessary for AAR.

2.2.2 Sonic Linking

Serafin *et al.* separate a soundscape into 'action sounds' and 'environmental sounds', with action sounds representing sounds made by human actions (such as snapping fingers, clapping, or walking, hitting, or throwing) and environmental sounds representing sounds made by other objects and entities in the environment [166]. Either of these sound categories could be leveraged for AAR experiences, however this has had minimal exploration so far.

Jylhä and Erkut have explored the potential for these action sounds as a general computing system control, under the term 'sonic gesture' [93]. They first explored the potential of sonic gestures through a system that utilised hand claps as control inputs for three different applications – a virtual audience simulation whose clapping speed was mapped to the user's own clapping speed, a system for adjusting music tempo based on the speed of the user's clapping, and a system to control a musical sampler. These applications were only evaluated informally. In a later paper, Jylhä provided a broader overview of the potential of sonic gestures, identifying that sonic gestures could be pitched (e.g. whistling or humming) or unpitched (hand claps, finger snaps), static (a gesture with constant pitch) or dynamic (a gesture with variable pitch), and impulsive or continuous [92]. Jylhä identified hand claps, finger snaps, body taps, whistling, humming, impulsive, fricative, and vowel vocal sounds; breathing, footsteps, scratches, blowing, and knocking or tapping as possible avenues for sonic gesture controls.

Some of these sonic gestures have been demonstrated in other contexts. Sporka *et al.* compared the use of speech input and humming as a means of controlling arcade game Tetris, finding that in that context, humming controls were faster and more accurate than speech controls [173]. Sporka and colleagues have also investigated a computer mouse controlled by whistling [171] and a keyboard operated through humming or whistling [172] as alternative input methods for people with motor impairments. Their whistling mouse was not evaluated in a full user study, but the sonic keyboard was evaluated by five users with varying motor and speech impairments and who responded positively to the tool.

In the realm of auditory applications, Vesa and Lokki proposed a system that could use finger snaps for control input, using the position of the finger snap as another layer to the control [184], allowing for a snap on the user's left, middle, and right to execute different commands. While the paper primarily dealt with the detection and localisation algorithm and wasn't fully evaluated in an application context, the authors proposed using the system as a music player control (skip

backwards, play/pause, skip forward for left/centre/right snaps respectively).

The most common form of action sound-based control is voice input, which has become an increasingly common way for humans to interact with computing systems. Modern smartphones feature voice assistants like Apple's Siri, Amazon Alexa and Google Gemini which can enable hands-free control of our devices [86]. Smart speakers like Amazon Echo devices, Apple HomePods, or Google Home speakers bring this functionality into our homes, allowing control of media systems, smart home devices, and more, and often operating purely within the auditory mode without any visual display. Speech-to-text systems also allow users to utilise their voice for text input, providing an alternative or addition to traditional keyboard inputs [118].

While action sounds have seen exploration in AAR contexts, the use of environmental sounds to drive AAR applications has been very limited until now. Of the 62 AAR systems identified in Yang *et al.*'s review, only one system responded to the presence of environmental sounds [198]. That system, Sawhney and Schmandt's Nomadic Radio was a system for auditory notifications for mobile users [158]. Users wore a wearable speaker which presented notifications, and a microphone which used their sonic context to inform how and when to present these notifications. When presenting a notification, the system considers the notification priority, how recently the user interacted with the system, and the level of nearby conversation to decide how intrusively to present the notification. Less intrusive formats, like a splash sound effect or an abridged reading of a message were used when there was conversation nearby, and when surroundings were quiet a more intrusive presentation like a full readout of a message was used.

The only other example of an AAR system that responded to environmental sound was the Acu-Notch system proposed by Lokki *et al.* [110], discussed earlier. Lokki *et al.* proposed using the auditory notching algorithm in response to key sound events nearby, such as emergency vehicles, however this functionality was never implemented into the system. The Nomadic Radio system was only evaluated with a single user over two days, and the principle of an AAR system responding to real-world sounds has never been evaluated fully, nor have there been any examples of such systems within the last 20 years.

Despite this, there is evidence to suggest consumer interest in AAR systems that interact with environmental sounds. Bustoni *et al.* [34] recently presented results from a survey exploring noise management in AAR. They presented 124 respondents with 21 examples of real-world noise, and surveyed respondents about how they would like an AAR system to alter those noises to reduce annoyance (for example making noise more pleasant, attenuating or removing noises, replacing noise, etc). Participants responded positively to all of their hypothetical augmentation, suggesting a clear consumer interest in AAR systems which can respond to environmental sounds. However, this represents only a first step, and they did not present or evaluate any prototypes for such an AAR system.

Outside of AAR there are also some existing examples of computing systems which respond to environmental sounds. Many smart devices now feature sound detection for accessibility

reasons, detecting important sounds such as doorbells for users who are hearing impaired [38, 28]. Other sound classification systems are also commercially available, for example the Shazam system for music recognition [186] or the Cornell Lab or Ornithology's BirdNET model which can identify different species of bird by their calls [94]. Perhaps the most widespread example of sound-reactive systems is active noise cancellation and mic-through acoustic transparency systems, featured on flagship audio devices like Apple AirPods [2] and Sony's flagship earbuds [3], allowing users to block out external noises.

Despite AAR systems being sonic in their output, there are very few examples of AAR systems that also use sound as an input to inform the application or experience, and the principle overall has never been evaluated. A system cannot augment reality without having an awareness or understanding of reality, and so this represents a fundamental gap in the current state of AAR requiring further research.

2.3 Technological Underpinning of Audio Augmented Reality

There are a number of technologies that enable AAR applications. While this thesis is primarily focused on theoretical and design-based contributions to the field, these areas also underpin the work presented in later chapters, and a primer on each of them is presented here.

2.3.1 Playback Devices and Acoustic Transparency

The playback device used for virtual audio in AAR is a key consideration. To *augment* reality with virtual sound, the user naturally must be able to *hear* reality alongside any virtual elements, and so playback devices which are acoustically transparent and facilitate this are central to an AAR system [91]. As well as facilitating AAR, there has been evidence that these devices can improve the plausibility of virtual sounds and that they can be considered safer than non-transparent playback devices [121]. There are two forms of acoustic transparency: 'active' or 'hear-through' acoustic transparency where the real world is presented to the listener via a microphone, and 'passive' acoustic transparency where the ear is left unblocked, allowing the listener to hear the real world normally.

The most common form of active acoustic transparency comes from microphone-enabled wireless earbuds, often referred to as 'Hearables', such as Apple AirPods. These microphones often capture the user's surroundings to allow for active noise cancellation (ANC) processing[4], and some devices also offer the ability to present this microphone signal to the user instead, similarly to a hearing aid. Devices like Apple's AirPods often offer options in-between, cancelling

[2] https://www.apple.com/airpods-pro/

[3] https://www.sony.com/headphones/products/wf-1000xm5

[4] Where the microphone signal is played back with reversed polarity to cancel out any real-world audio that makes it to the listener's ears, if you're curious.

unwanted noise but allowing the listener to still hear conversations or blending between ANC and transparency. While active devices like these can offer finer-grained control over how a user hears the real world, this additional processing can also bring with it latency and perceptual problems [72]. Denk *et al.* compared seven commercial headsets offering active transparency and found a large variance in latency times, and that some devices could introduce comb filtering effects, colouration, the deterioration of binaural cues, and more [45]. In a later experiment, Schepker *et al.* concluded that currently, actively transparent devices cannot achieve an equivalent quality to open ear listening [159]. However, authors like Stemasov *et al.* [175] and Bustoni *et al.* [34] have envisioned a future where such devices allow the curation and adjustment of our real-world auditory perception, manipulating the volume of specific sounds to better suit our needs.

While passive playback devices do not suffer from these perceptual problems as the ear is left unblocked, although they do not offer the same clear potential for real-world manipulation. Passive devices can take multiple forms, the most common being audio-enabled glasses such as the Huawei Eyewear[5], USound Fauna glasses [181], or Meta's smartglasses[6]. Other approaches include earbuds that clip on the edge of the ear rather than blocking the ear canal, such as Bose's Ultra Open earbuds[7] or Soundcore's C30i[8]. Other devices use bone conduction (often termed 'bonephones') to transmit audio signals directly to the auditory nerve, bypassing the eardrum entirely, however perceptual studies by Barde *et al.* showed that while these devices can produce promisingly externalised audio, localisation performance suffered when using them [9, 8].

As most AAR systems focus on a single user, playback devices are usually limited to personal audio devices like those mentioned above. However, there are some examples of suitable loudspeaker-based systems. Recent advances in speaker technology have enabled 'sound zones', where a speaker system can allow multiple users to experience their own virtual sound experience without overlapping or interfering with others' [90]. Razer have demonstrated a similar product for personal, headphone-free spatial audio in their Leviathan soundbar. The soundbar uses an IR camera to track the listener's head movements, and beamforming approaches to precisely direct soundwaves towards the listener to mimic spatial audio [150]. While such devices have not been applied to AAR scenarios yet, they could allow for in-home or multi-user AAR experiences in the future, blending personal virtual audio with a shared real aural environment.

2.3.2 Spatial and Binaural Audio

As well as the device used to playback virtual audio, another central component of AAR is the use of immersive or 'spatial audio'. Spatial audio is a broad term used to describe audio designed

[5] https://consumer.huawei.com/en/audio/huawei-eyewear-2/

[6] https://www.meta.com/gb/ai-glasses/

[7] https://www.bose.co.uk/en_gb/products/headphones/earbuds/bose-ultra-open-earbuds.html

[8] https://www.soundcore.com/uk/products/c30i-a3330-clip-earbuds

to be 3-dimensional – in spatial audio, virtual sounds can be positioned anywhere in a 360-degree sphere around the user, just as they are in the real world. Spatial audio already has applications in music playback, game audio, cinema and television sound, and artistic installations.

For AAR, the most common form of spatial sound deployed is that of **binaural sound**. Binaural sound is played back over headphones (or similar one-channel-per-ear playback devices), with the left and right signals processed to mimic the effect a listener's ears, head and wider body have on the sound we hear in the real world. As sound travels to our ears, the shape and acoustical properties of our head and body alter the sounds we hear slightly, absorbing and reflecting sound waves in different ways. This results in the sounds received at each ear being subtly different, having different arrival times, frequency content, phase, and more [191]. Our hearing system relies on these interaural differences to localise sounds, and binaural audio reproduces these cues to accurately position virtual audio in three dimensions.

Binaural audio relies on a **head-related transfer function** (HRTF) or **head-related impulse response** (HRIR)[9], which represents the acoustic effect of a person's physiology [44, 32]. HRTFs are measured for a given individual's head, and a sound source at a given position, and when convolved with a given piece of audio, the resulting sound will appear to that individual as if it originates from the corresponding position. As HRTFs are specific to a given person and source position, comprehensive measurement is laborious and complex, requiring highly specific facilities and equipment. More often, binaural sound is processed for a freely available, generic HRTF dataset that provides most of the binaural cues, though not as accurately as individualised measurements. Often these generic HRTFs correspond to head-and-torso-simulators (HATS) or 'dummy heads', which represent an average human.

In an AAR context, headphone-based spatial sound like binaural audio also requires some level of user tracking to present virtual sounds in fixed positions, either relative to the listener's head orientation (egocentric) or relative to the listener's head and body position (exocentric). Rendering sound in a fixed egocentric position allows the listener to gain additional localisation cues from head movements [120], while rendering accurate exocentric spatial audio can benefit plausibility [53]. Approaches for this can include dedicated headtrackers [74, 198], GPS positioning [198], or camera-based inside-out tracking featured on mixed reality headsets like the Quest 3 [37].

Spatial audio is often also achieved using Ambisonics technology. Traditional channel-based approaches to audio use one audio channel for each speaker: one channel for mono, two channels for stereo, and four or more for surround formats. Ambisonics instead uses channels to represent a speaker-agnostic sound field: sound sources are encoded into the Ambisonics domain, where they can be positioned and manipulated, and then are decoded to a target speaker configuration, including to binaural formats [199]. Ambisonics sound fields are described by

[9] HRTFs representing the frequency domain, and HRIRs representing the time domain. Through the wonder and magic of the Fourier transform you can derive one from the other.

their numerical **order**, with higher orders having increased spatial resolution. Higher orders also necessitate more channels, as described in Equation 2.1. Specialist microphones are also capable of recording directly to Ambisonics format, most commonly a 1st-order field, though modern examples are capable of recording at 3rd-order resolution or higher [200, 126, 52], allowing high-resolution, three-dimensional capture of an auditory space.

$$Num\,Channels = (Order + 1)^2 \tag{2.1}$$

In the commercial realm, surround sound formats have employed similar approaches in recent years, moving towards object-based formats like Dolby Atmos and DTS:X. These approaches attach three-dimensional positional metadata to specific sounds, and use a rendering or decoding stage to translate the mix to specific speaker configurations which include height channels. While most notably deployed in cinema contexts, streaming services like Apple Music now offer Dolby Atmos music tracks [47] and many soundbars, TVs, computers, and AV receivers now offer support for these formats, giving many consumers the ability to experience spatial audio at home.

Spatial sound not only enables three-dimensional AAR experiences but offers tangible benefits to the listener. Chiefly, the use of spatial sound has been shown to have beneficial effects on presence, immersion, flow, and emotion [95, 53, 144, 183], which are markers of heightened levels of engagement [31], and may also have beneficial effects on task completion time in virtual environments [155]. Use of HRTF spatialisation also improves the plausibility of virtual sounds, their externalisation, and our abilities to localise them accurately [77, 107, 138, 32], even when using a generic, non-individualised HRTF [14, 192]. Integration of spatial sound processing into an AAR system is therefore highly desirable, and these technologies are deployed in all six of the user studies presented in this thesis.

In recent years these spatial sound systems have become more and more accessible, whether presented over speakers or through personal devices like earphones or headsets. The ability to convincingly position virtual sound sources in three-dimensional space is central to AAR, and the ease with which this can now be done means AAR has never been more accessible. As this technology continues to advance, these spatialised sounds will become increasingly convincing, while the computational requirements to create them will continue to be reduced. It is important to identify and resolve the gaps in our understanding of AAR now, so that future AAR systems can better take advantage of this potential.

2.3.3 Environmental Acoustics

One important component of any AAR system is the modelling of environments for virtual sounds. The primary technical challenge for AAR as a field is seamlessly presenting a virtual sound as real [131, 73], and accurately modelling a real world environment's effect on sound and

applying that to the virtual is key to achieving this. As discussed in subsection 2.2.1, accurate acoustic cues have beneficial effects on sound perception such as improving externalisation or distance estimations.

The two primary terms used to define this seamless presentation of virtual as real are **plausibility** and **authenticity**. Authenticity is the perceptual identity of real and virtual sound – no difference can be detected between the two [20]. Plausibility, meanwhile, is the perceptual identity of virtual sound and a listener's *expectation* of a real sound [107]. Authenticity is far harder to achieve than plausibility as a result [138] – the slightest difference between real and virtual might destroy authenticity, while plausibility might still be maintained [68]. Plausibility, on the other hand has already been achieved under certain conditions [107, 138, 68, 30]. As AAR is concerned with presenting virtual audio as real to the user, integrating environmental acoustics into the system to maximise plausibility is crucial.

The effect an acoustic environment has on sounds within it can be described through its **room impulse response** (RIR), which illustrates how sound evolves in that space over time for a given sound source in a given position, and a given listening position. By convolving the RIR with a given piece of audio, that audio can be made to appear as if it were sounding in the RIR's corresponding acoustic space, in much the same way as HRTFs can be deployed to create spatial sound. RIRs can be created either through direct measurement in a space, or through simulation.

Recording an RIR can be achieved either through recording an impulsive sound in a space, such as a balloon pop, handclap, or gunshot; or by recording an acoustically excited room, and deconvolving that recording with an excitation source such as a sinusoidal sweep [89]. This is a labour-intensive task which requires expensive and specialist equipment when done properly, and many repeat measurements need to be taken at different source/listener positions to be able to truly simulate a room's acoustics.

Simulation approaches avoid this laborious measurement work, but instead usually require a three-dimensional model of the user's surroundings, using techniques like ray-tracing to compute an approximation of the space's RIR [157]. These approaches instead require a 3D model of the user's surroundings, and while often these are available in visual XR contexts, they are rarely otherwise needed for AAR, representing a different challenge for acoustic reproduction. However, there have also been recent examples of smartphone systems capable of mapping a user's surroundings [167, 148], and as SLAM tracking [57], point cloud mapping [117], and Gaussian splatting [197] capabilities become more ingrained in everyday devices, modelling our surroundings for acoustic simulation is likely to become more feasible.

Whether measured or simulated, an RIR can also have differing degrees of spatial resolution, ranging from an omnidirectional RIR which contains no directional information, to a stereo or binaural RIR which models interaural differences, through to RIRs in the Ambisonics domain which offer increasing spatial resolution. In principle, an RIR with higher spatial resolution should offer a closer simulation of a space's acoustics, but as mentioned in §2.2.1, there is cur-

rently no clear understanding of how high spatial resolution should be to maximise plausibility and computational efficiency.

2.3.4 Sound Detection

As discussed in §2.2.2, AAR systems could benefit from the ability to respond to real world sounds. While the AAR applications which have done so are extremely limited, sound classification technology is an area of active research, and deployed in other consumer-facing applications.

The Nomadic Radio system, the only AAR system that has incorporated real-time sound classification, relied on machine learning, using a bespoke Hidden Markov Model to classify nearby speech [39]. Their detection system could identify not only individual sound events, but when a user transitioned from one scene to another, potentially allowing for a more granular understanding of a user's context. Such machine learning models form the basis of most other sound classification systems also. Modern general-purpose sound classification models include Google's YAMNet model and its precursor, VGGish [84], which can classify hundreds of different audio events, with a high level of accuracy [180].

Other, more specialised sound classification models also exist, such as the Cornell Lab of Ornithology's BirdNET model [94]. BirdNET is capable of identifying more than 6000 species of birds from around the world [195], and is available both as a pre-trained model, packaged for Raspberry Pi-based monitoring [194], as a mobile app for on-demand classification,[10] and more. Perhaps the most well-known sound classification system is Shazam, a music classification system that can identify specific musical tracks in an audio stream [186]. Shazam is widely used, often coming pre-integrated into smartphones, and features 'Auto-Shazam' functionality, where the system continuously monitors and identifies nearby music, providing users with both real-time and historical information on unfamiliar music.

While these systems identify the presence of a sound, there is also the matter of identifying the location of the sound, something which is an area of active research. Researchers have been developing machine learning approaches for sound localisation for some time, however these can require specialist microphone arrays [85, 156], be limited to localisation within one plane [187], or cannot handle overlapping sources or real-world environments [85, 46]. While recent advances have begun to solve these problems [46, 2], there are not yet viable pre-trained models for sound localisation like there are for classification tasks.

[10] https://merlin.allaboutbirds.org/

2.4 Conclusion

This chapter has presented a summary of the area of audio augmented reality, the difficulty in defining it, the potential of sonic and acoustic links between the real world and virtual elements for audio augmented reality, and the technology that enables audio augmented reality. A number of key themes and gaps motivate the research questions explored in this thesis, summarised below.

2.4.1 RQ1: How can an acoustic link between real and virtual elements be created in audio augmented reality?

Section 2.2.1 explored the perceptual effects of acoustic cues on an AAR experience, noting their beneficial effects for creating externalised, plausible cues which can heighten immersion and the believability of virtual sound sources. It discussed how there is still no clear understanding of how accurately reverberant cues need to be reproduced to balance these benefits with computing overhead, that most AAR systems do not reproduce acoustics, how these cues are usually evaluated in lab environments rather than the environments AAR will likely be used in, and how they have not been evaluated with transparent playback devices that enable AAR experiences. This leads to the first research question for this thesis.

2.4.2 RQ2: How can a sonic link between human-produced sounds and virtual elements be created in audio augmented reality?

Section 2.2.2 noted that action sounds – sounds produced by human beings – have potential as control inputs for computing systems. It discussed the potential of sonic gestures and voice input, and noted that such control schemes have seen minimal evaluation in AAR. This forms the focus of the second research question, investigating the potential of human-produced sound to inform and control an AAR system as the first evaluation of a sonic link.

2.4.3 RQ3: How can a sonic link between environmental sounds and virtual elements be created in audio augmented reality?

Section 2.2.2 then discussed systems that respond to environmental sounds, noting that there are only two examples of such systems, one of which was conceptual and neither of which was fully evaluated with users. It discussed the fact that sound reactivity has been deployed in non-AAR contexts, and that this represents a fundamental gap in AAR: an AAR system cannot augment one's auditory reality without being aware of it. Section 2.3.4 discussed approaches to sound classification which could be leveraged to enable this, providing both the means and motivation

behind the third research question: exploring how an AAR system can respond and be informed by the sounds composing our aural surroundings.

Now that the research questions underlying the thesis have been established, the following chapters will answer them in turn to create foundations for the field of AAR.

Chapter 3

Defining Audio Augmented Reality

As discussed in Chapter 2, there are a number of competing definitions of AAR, conflicting terms, and no clear understanding of what makes an experience AAR. Synthesising these into one overall model of audio augmented reality will bring clarity to what AAR is, what it can be, and where there are gaps in our understanding. This chapter discusses these existing definitions in greater depth, their themes and merits, offers alternative perspectives, and provides a new definition that unifies these previous discussions and underpins the work presented in the rest of the thesis.

3.1 (Audio) Augmented Reality

Of the common themes in existing literature, one is a view of AAR as a subcategory or more specialised form of augmented reality (AR). This view is shared by authors such as Rovithis *et al.*, Sikora *et al.*, and Lawton *et al.* [153, 168, 105], but also implied simply by the name 'audio augmented reality'. A system could not be an example of *audio* augmented reality without also being an example of augmented reality.

> "[AAR is] a type of AR, in which the virtual component that enriches the real world consists of audio information" – Rovithis *et al.* [153]
>
> "visually augmented reality has its acoustic version–AAR" – Sikora *et al.* [168]
>
> "[AAR is] an instance of AR, whereby experiences and actions in the real world are accompanied by additional layers of sound" – Lawton *et al.* [105]

With that in mind, a viable definition of audio augmented reality must also encompass the criteria for augmented reality, which are more clearly defined. Azuma's definition [6], one of the most commonly referenced, defines augmented reality as a system which:

- Combines real and virtual elements

- Is interactive in real time
- Is registered in three dimensions

Azuma's definition, while popular, is also prescriptive from a technical viewpoint, and the boundaries of AR have shifted in the three decades since it was first proposed. As Nijholt points out, modern AR content can feature non-interactive elements [137], and one can imagine that in a future with ubiquitous AR displays, there would be space for 2-dimensional AR content like heads-up displays.

More recent definitions take a broader view of AR, such as Lindeman *et al.* [108] who define AR only as "the mixing of computer-generated stimuli with real-world stimuli", or Mallem [113] who claims AR allows "spatial and temporal virtual and real worlds [to] co-exist, which aims to enhance user perception in his real environment". Whether or not this use of the term 'worlds' was deliberate, Mallem's notion aligns with Milgram and Kishino's seminal proposal of the reality-virtuality continuum in which all XR applications can be placed [125]. At one end of the continuum exists the real environment, and at the other a fully virtual environment (virtuality). The space between these two points is where mixed reality applications exist, with AR applications existing towards the real end of the continuum, and 'augmented virtuality' existing towards the virtual end, representing a virtual reality environment augmented with real elements. The real and virtual environments defined on the Milgram-Kishino continuum could also be thought of as Mallem's real and virtual worlds, and doing so represents a more flexible and ambitious view of AR.

While both terms can be used interchangeably, 'environment' often implies a smaller scope, and our real world is composed of multiple environments that contribute to a larger whole. We reference natural and built environments, physical and digital environments, visual and aural; social, cultural, political, educational environments, and many more. All of these come together to create the real world we live in, and can often represent only a small part of it. In a future of ubiquitous AR, augmentations could span many of these environments at once, or new environments yet to be imagined. At that point, we are no longer augmenting an environment, but a world, and this should be accounted for in our definitions. While their limitations may only be implied, the words we choose to use when defining something are important, and should be chosen carefully.

Mallem's definition also suggests an AR system must enhance a user's perception of their real environment, however, it is unclear exactly what is meant by this. Is perception enhanced because there are now virtual elements for the user to perceive? This condition would surely be met by the virtual world existing in the first place, and is unnecessary to include. Is an AR system one that improves the user's ability to understand their surroundings? There are many examples of AR applications with no such goal, yet which are clearly AR. AR games and videos seek to entertain, not improve user perception. This thesis argues instead for a level of abstraction, and that an AR system offers a *benefit* to the user. The word 'augmented' implies a positive

alteration, a general improvement[1], and while this could take the form of improving the user's environmental understanding, it could also take any number of other forms.

While many definitions of AR like these focus on how the system relates to the user, Schraffenberger proposes also that we should consider how an AR system relates to the real world. Schraffenberger considers AR to not just be the combination of real and virtual elements, but to arise when the virtual and real have some form of connection between them [161, 163], and gives a particular focus to spatial and content-based relationships. For example, watching the evening news is not AR, until the newsreader is seated at the kitchen table (where they then have a spatial connection to the environment). If the newsreader behaved differently based on the environment or based on the user, this would be augmented reality through a content-based relationship.

A central issue with defining AR and AAR is pinning down when an experience becomes augmented – when it stops being 'listening to music' and becomes AAR, and this notion of the real and virtual worlds being linked in AR provides clarity on this issue. Music listening is not AAR until the virtual music the listener hears is shaped by aspects of the real world. Schraffenberger focuses on spatial- and content-based links between real and virtual, but there is no reason an AR or AAR system needs to be restricted to such relationships. In the AAR context this thesis is focused on, real and virtual worlds could also have auditory links, such as the sounds in one world influencing the other, or the acoustics of one world being shared or altered by aspects of the other.

3.2 Existing Definitions of AAR

While many avoid defining AAR [183, 5, 115, 35, 140], one common view is of AAR merely being the presentation of artificial or virtual sounds to the listener alongside the sounds of the real world [114, 131, 65]:

> "In the most general sense, audio AR is simply the introduction of artificial sound material into the real world" - Mariette [114]
>
> "Extending the real auditory environment with virtual audio entities" – Nagele *et al.* [131]
>
> "A technology that aims to embed virtual auditory content into the real environment of a user" – Gamper [65]

Other authors employ similarly broad definitions, but describe the virtual audio as being 'overlaid' on the real environment For example, Heller *et al.* claim AAR systems "overlay the physical world with a virtual audio space" [80]. Boletsis and Chasanidou consider AAR to

[1] Terrible-Horrible-No-Good-Very-Bad-ified Reality hasn't really taken off in the same way.

allow the "simultaneous perception of the real environment and a virtual audio overlay" [25]. Stecker *et al.* claim that in AAR "synthetic or recorded sound overlays the natural sounds of the physical environment". 'Overlaying' virtual atop real implies a separation between real and virtual, rather than the creation of an augmented gestalt, and broad definitions like these can also be unhelpful – radio, phone calls, and loudspeaker systems can all be thought of as introducing or overlaying virtual sounds into the real world, despite very rarely being argued as examples of AAR. With such broad definitions, it is unclear when experiences like these would become AAR. Some scholars propose additional criteria for an AAR system that can go some way to mitigating this problem.

For instance, Krzyzaniak *et al.* do not strictly define AAR, but claim it is composed of two essential elements: an 'analogue' or real world sound, and a digital system that changes or adds to it [104]. As well as these two essential elements, Krzyzaniak *et al.* also allude to the user benefitting from the augmentation, and to it having relevance to the user, perhaps through a specific "object, activity, place, time, setting, [or] person". This aligns with some of the ideas set forth in the previous section, and, importantly, is one of the only existing definitions to suggest the potential for sound-based links between the real and virtual. However, this view is also unimodal – to meet that criteria an AAR system could only respond to and augment real-world sound, and applications like tying virtual audio to objects or locations would not be AAR, despite the authors including such applications in their taxonomy of AAR in the same paper. While AAR could be unimodal in this way, there is no clear need for it to *only ever* be unimodal, particularly in a ubiquitous AR future, and so AAR should not be defined based on this requirement.

One other stipulation often employed in existing definitions is a requirement for the virtual audio to be presented in a specific manner. The use of spatial audio processing is often listed as a requirement for AAR [119, 70, 103, 82], while other authors suggest AAR requires a 'seamless integration into the real environment' [132], or for users to be 'unable to distinguish' between the real and virtual sound sources [26]. While highly plausible virtual audio is undoubtedly beneficial for many AAR scenarios, and these definitions no longer include traditional audio experiences like music, phone calls, or radio, there is nothing about spatial audio that makes it a defining characteristic of AAR. Consider an audiobook presented in spatial sound and a soundwalk presented in stereo which unfolds differently as the user moves through the physical world. Which would feel the most like it was augmenting a user's reality? While the audiobook might feel more *immersive*, this thesis argues the stereo soundwalk would feel more *augmented*, as it is rooted in and shaped by the real world.

Other authors insert a concept of interactivity into their AAR definitions, aligning with one of Azuma's requirements of AR. Lawton *et al.* claim AAR to be "an instance of AR, whereby experiences and actions in the real world are accompanied by additional layers of sound" [105]. Tikander, although using an alternative term to AAR, describes it as being when "the natural

surrounding sound environment is enriched with interactive virtual sounds" [179]. Similarly, while interactivity may elevate an AAR experience, it is not crucial to it. If one considers the AAR applications discussed in Chapter 2, many do not involve interacting with the virtual audio directly. In many – Vazquez-Alvarez *et al.*'s sound garden [183], Bluebrain's National Mall [23], or McGookin's PULSE system [122] – the augmented nature arises not from the virtual audio, but the fact that the virtual audio responds to the user's interaction with the real world. The connection to the real world is what makes the experience AAR, not the interactivity of the virtual sounds.

McGill *et al.* apply Schraffenberger's notion of real-virtual connections to AAR, which they consider to be "auditory headset experiences intended to co-exist with/supplement reality or exploit spatial congruence with real-world elements, typically rendered on [acoustically transparent] headsets" [121]. However, McGill *et al.*'s definition only concerns itself with spatial connections between real and virtual. As discussed earlier, while the links between real and virtual *could* be spatial, these links could also take other forms.

With so many competing definitions of AAR, many of which partially overlap, there is no clear, accepted definition of the field. To solve this problem, Dam *et al.* used Grounded Theory analysis to arrive at a new, robust definition for AAR. They conducted a workshop with audio experts defining AAR, followed by a literature review, and arrived at a preliminary definition of AAR:

> "Auditory information, customised for the intended user that is capable of sufficiently immersing yet retaining awareness of their environment and designed to provide appropriate assistance in the user's primary task".

This preliminary definition was then followed by further focus groups and expert interviews for further refinement. Their final definition is built on three pillars: an AAR system should be connected to the environment, context-adapted, and goal directed. Dam *et al.* provide a final definition to encompass these pillars:

> "AAR is defined as the augmented auditory stimuli that are goal-directed, immersive, but distinct from the real sounds, and adapted to users' context".

Dam *et al.*'s pillars provide a useful framework for considering AAR, and align well with many themes of previous definitions. For each of their three pillars, they also provide sub-criteria that contribute to that pillar, claiming that as these sub-criteria are met an experience becomes more AAR. They state their environmental connection pillar includes a user being able to perceive an alternate environment (composed of sound) and/or be continuously aware of the real world, but also introduce a criterion that virtual audio sources should "be connected to the physical environment, but at the same time should be perceived as being distinct from the sounds of the physical environment". This notion of a connection to the physical world aligns

with Schraffenberger and McGill *et al.*'s views, that virtual elements should have some link or relationship with the real ones, however Dam *et al.*'s stipulation that the virtual elements should also be distinct is unusual within AAR definitions. Usually, the opposite is argued and, while this could be desirable under certain circumstances, a perfectly plausible presentation could be desirable in others. This stipulation can also only be met in AAR scenarios where virtual sounds are added into the world. Future AAR scenarios, such as the alteration or removal of existing real sounds would be unable to meet this criterion.

Their contextual adaptation pillar, the notion that the experience should change based on the user's immediate reality, also aligns with Schraffenberger. Proposed sub-criteria of this pillar include an AAR system doing this in such a way that allows listeners to maintain social interactions with others, an AAR system manipulating environmental noise such as noise cancellation or transformation, and adapting to "users' temporal and spatial reality to keep users aware of the augmenting audio", such as altering playback to maintain audibility based on environmental noise. Finally, their goal-directed pillar claims AAR systems must help a user accomplish their primary task, requiring AAR cues to be easily recognised, relevant to user goals, and recognisable as being helpful to the user. This concept of an AAR system assisting with a user's goals is perhaps the most flexible way to consider whether a system provides a benefit or not, something often proposed as criteria for AAR.

While Dam *et al.*'s pillars are thought-provoking, their nine sub-criteria result in a very complex model of AAR, and some are at odds with established ideas of the field and potentially restrictive for future AAR applications. The definition provided with these pillars also has flaws, primarily rephrasing the requirement for these pillars to be present (although the environmental connection pillar is absent), and necessitates deeper reading into their taxonomy to understand when a system is goal-directed or adapted to users' context, or when an auditory stimuli is an augmented one. Further work is still needed to distil AAR down to its essential components and provide a clear and accessible definition.

3.3 Competing Terms

As well as differing opinions on what AAR is, there are also differing opinions on whether AAR is even the correct term, and there are other forms of audio-only experience that bear close relation to AAR but fall under different names. Harma and other authors use the term *augmented reality audio* (ARA) [74, 73, 153, 179], as well as subcategories such as *wearable augmented reality audio* [74] and *mobile augmented reality audio* [73]. Harma also uses *augmented audio reality*[2] as a synonym for the same system, which is described as being one "where real and virtual sound scenes are mixed so that virtual sounds are perceived as an extension to the natural ones" [73]. ARA applications are proposed to be unimodal, extending and relating to only the

[2]Presumably someone is also labouring to bring 'Reality: Audio Augmented' to the masses to complete the set

auditory realm, however as noted previously, others have defined AAR in the same way and there is no clear reason as to why such applications would need to be restricted in this way.

Mariette also proposes a number of AAR-adjacent terms. *Spatial audio AR* (SAAR) is proposed as a subcategory of AAR presented using spatial sound processing, and *Personal Location-Aware Spatial Audio* (PLASA) is proposed as a further subcategory of SAAR that is presented to a single user over headphones [114, 116]. Mariette also describes *locative audio* as an umbrella term that covers audio experiences which are tied to a user's position in the real-world, though not necessarily their orientation. SAAR and PLASA applications are already encompassed by AAR and while they have interesting implications – such as the potential for AAR applications to be presented to multiple users at once rather than being purely individual – they are also at risk of adding unnecessary complications. This thesis focuses only on AAR as a whole, and leaves future work to determine whether SAAR and PLASA are useful terms to consider within AAR.

Locative audio, meanwhile, is argued in this thesis to be encompassed by the view of AAR developed in this chapter. As discussed earlier, AR applications rely on virtual elements being linked to the real world. While AAR elements can be linked to real-world locations as in locative audio, they could also be connected to environmental sounds, acoustic spaces, specific objects, actions taken by a user, or any number of other triggers yet to be imagined. As the term most commonly used, and one which encompasses these competing terms, *Audio Augmented Reality* is the focus of this thesis.

3.4 Synthesising a New Definition

Having reviewed a broad range of definitions of AAR, we can create a refined model of the field. This review and discussion identified the following key points and themes:

1. Audio augmented reality is a form of augmented reality.
2. Augmented reality features real and virtual elements, in real and virtual environments, in real and virtual worlds.
3. Augmented reality should offer the user a benefit, as 'augmented' implies an improvement.[3]
4. In augmented reality, the real and virtual elements/environments/worlds should be linked or related to one another.
5. In audio augmented reality, the virtual elements/environments/worlds are sound-based.

[3] 'Expanded' or 'supplemented' reality would have a different connotation.

6. Audio augmented reality encompasses locative audio, SAAR, PLASA, ARA, MARA, and WARA.

The goal with this new definition is to encompass the things that are widely agreed to be AAR, and exclude the things that are widely agreed *not* to be AAR. As AR and AAR are new technologies, it is also important that this definition is not unnecessarily restrictive, to provide space for novel AAR applications in the future. As we have seen already, some prior definitions can be too prescriptive, no longer able to cover modern systems that are broadly considered AAR, and it is important to avoid this. At the same time, if a definition is too broad, such as those which consider AAR to merely be the introduction of virtual audio into the real world, then non-AAR applications can then fall under the same umbrella, making the definition unreliable.

As discussed earlier, this thesis proposes firstly following Mallem's approach and considering AAR in terms of real and virtual 'worlds'. As a more abstract concept than an environment, a world can grow and shrink as required – a world could be composed of a single element or millions, cover a room, a city, or a continent; and characterise elements in physical or abstract terms. A world can encompass existing AAR systems, but also enable novel ones we have yet to imagine.

A world also contains all manner of characteristics which could form the basis of a real-virtual link. While Schraffenberger's concept of real-virtual links are crucial to augmented reality, there is no reason to impose limits on what links can or should be used. Schraffenberger argues for spatial and content-based links, this thesis focuses on sonic and acoustic links, but there is no reason AR and AAR applications of the future could not feature object links, action links, temporal links, emotional links, or any number of other connections between the real and virtual worlds.

This thesis does make an assertion on the benefit an AR system provides, as there is space here to further pin down augmented reality itself. Like Dam *et al.*, this thesis views fulfilling goals as being a valuable way to consider AAR. Achieving a goal is perhaps the broadest way of providing a user with a benefit, however this thesis proposes considering this in terms of the goal itself, and the actions taken to achieve it. Consider that in our day-to-day, non-XR reality, our goals exist within the real world, as do the actions we take to achieve them. In virtual reality, our goals exist within the virtual world, as do the actions we take to achieve them.[4] It follows then, that in augmented reality, the goals we have, and the actions we take to achieve them must be in different worlds, or else we are experiencing reality or virtual reality. When playing Pokemon Go [135], we walk to the park in the real world (real action) to catch Pokémon (virtual goal). We visualise our new IKEA sofa in AR (virtual action) to understand how it will look after purchasing it and placing it in our living room (real goal).

Based on this, this thesis proposes the following definition of AAR:

[4]We may move or manipulate something in the real world as part of that, but only due to imperfect control schemes having not yet figured out The Matrix.

Audio augmented reality is the creation of a virtual world, represented through sound, which is linked to the real world and allows users to achieve goals in one world by interacting with the other.

This definition unifies the key points outlined above, and provides clear criteria for determining whether a system is AAR:

1. Is audio the primary modality?
2. Does the system feature both real and virtual elements?
3. Is there a clear link (spatial / content-based / auditory / ...) between those real and virtual elements?
4. Can the user achieve something in the real world by interacting with the virtual world, or achieve something in the virtual world by interacting with the real world?

3.5 Auditory Links

Having introduced a new definition for AAR which supports the work in this thesis, it is also necessary to introduce the concept of auditory links on which this work is centred. Visual AR systems place a strong emphasis on mapping and understanding a user's physical surroundings such that virtual visual elements can be plausibly integrated: for example, walls are mapped so that virtual displays can be hung on physical walls, or the room is mapped three-dimensionally so that virtual objects can occlude or be occluded by physical objects. Modern hardware like the Apple Vision Pro even simulates real lighting effects for virtual objects, grounding them more plausibly in the user's environment. In AAR, this same potential exists in a user's auditory surroundings, with great potential for elements in the virtual world to respond to real-world sound, or *vice versa*. With reference to the definition of AAR presented here, this thesis defines this concept as an **auditory link**.

As an initial exploration of these auditory links, this thesis considers our aural surroundings to be composed of two parts: the sounds present in a space (speech, birds, cars, etc.), and the acoustic influence of that space (reverberation, echo, etc.). As independent aspects of an aural environment, an auditory link could potentially arise from either. This thesis terms these **acoustic** and **sonic** links, and defines them as:

- **Sonic Link:** The sounds present in one world influence the other through the AR system;
- **Acoustic Link:** The acoustics of one world influence the other through the AR system.

Sonic and acoustic links could be applied to any form of augmented reality experience (hence why they are defined in reference to an AR system rather than an AAR system specifically),

however they have the most obvious applications in AAR. As discussed in Chapter 2, an acoustically linked AAR system could apply the acoustics of the real world to virtual sounds, making them appear more plausibly real. Likewise, the acoustics of the real world could be manipulated to better match an audiobook narrative unfolding in the virtual world, simulating the cathedral the story's protagonist is currently within. A sonically linked AAR system might be as simple as positioning auditory notifications such that they do not clash with existing sounds, or as pervasive as attenuating or replacing the sound of heavy real-world traffic to be less irritating. Originating from the virtual world, enemies in a sonically linked AAR game might morph or distort real-world sounds to heighten the experience.

Just as visual mapping techniques like SLAM tracking and lighting simulation are fundamental to visual AR systems, allowing systems to get ever-closer to the ideal of blending real and virtual elements rather than overlaying virtual atop real, this thesis asserts these acoustic and sonic links as being equally fundamental to AAR, and the remainder of this thesis is devoted to exploring their potential.

3.6 Conclusion

AAR has existed for over 30 years, however throughout that time it has been poorly defined, leading to confusion about what AAR is, and what its potential is. This chapter explored existing definitions of AAR and AR, debated their merits, and presented a robust new definition which synthesises and builds on these prior discussions. It also identified and defined acoustic and sonic links, where an AAR application responds to nearby aural elements in the real world, noting that these have not been explored in great detail before. Exploring the potential of these sonic and acoustic links will provide foundational insights into what AAR applications could be, expand the state of the art for AAR, and possibly lead to AAR and AR applications with deeper integrations between the real and virtual in future. The experimental work presented in the following chapters explores acoustic and sonic links in an effort to reap these benefits.

Chapter 4

Acoustic Linking

As discussed in Chapters 1 and 3, a fundamental gap in AAR at present is the use of acoustic and sonic links to connect virtual sounds with a user's real-world environment. By reproducing the acoustics of a user's real world environment and applying them to virtual sounds, an acoustic link can be created between real and virtual. Chapter 2 demonstrated that reproducing an environment's acoustics for virtual sounds can result in virtual audio that more believably appears part of a listener's real surroundings. Acoustic cues improve the sensation of externalisation, immersion in an acoustic environment, and the plausibility of virtual sounds, all key to the central goal in AAR of creating a seamless blend of real and virtual audio. However, Chapter 2.2.1 also showed there remain unanswered questions with regards to acoustics in AAR.

Firstly, there is currently no clear understanding of how accurate or detailed an acoustic reproduction needs to be to provide these benefits. Work from Enge *et al.* suggested that plausibility was unaffected by increasing spatial resolution above 3rd-order [53], while Engel *et al.* found evidence first that differences above 1st-order Ambisonics were imperceptible, and then that differences could be discriminated up to 3rd-order [54, 55]. At the other end of the spectrum, work from Ahrens and Andersson suggested discrimination was possible as high as 8th-order [3]. This is a key issue as more detailed reproductions require more audio channels and more processing power to deploy – a 1st-order reproduction requires four audio channels where an 8th-order requires 81 – and the sweet spot of perceptual benefit and computational requirement remains unclear.

Secondly, most existing work in this area has focused on lab-based scenarios, studying listener perception in treated environments and under formal listening tests. While this provides valuable insights, these scenarios do not represent the environments where AAR systems are likely to be used, or the tasks AAR users are likely to engage in. Crucially, many of these listening tests ask participants to evaluate the acoustics of a space without being exposed to it, making judgments in an online test environment [54] or based on a picture of the space [30]. In an AAR scenario, users would be judging the acoustic simulation of an environment they are currently located in, and there is evidence to suggest a listener is more critical in this scenario [134].

Additionally, users will engage with AAR under various conditions, such as environments with differing noise levels, when engaged in another task, or using an interactive AAR system, all of which may influence their perception of the virtual sound world. A lab-based formal listening test may not adequately simulate these additional factors and it is important to understand how they might influence the listener's experience.

Thirdly, these existing lab tests usually deploy high-fidelity, professional-level speakers and headphones. While AAR could certainly be experienced using such playback systems, the AAR systems of the future are more likely to be experienced over more accessible and widespread headphones or wireless earbuds which are usually of lower fidelity. Additionally, a central consideration in AAR playback devices is acoustic transparency – as discussed in Chapter 3 we must be able to hear both real and virtual elements to achieve AAR – and devices like audio glasses, bone conduction headsets, or open-ear buds have had minimal evaluation in these contexts. It is important to understand how playback device may influence a listener's perception, and to evaluate acoustic reproductions over representative AAR hardware as well as devices designed for critical listening.

This chapter describes two studies designed to explore these aspects of acoustic linking, and provide insights on RQ1.

> **RQ1: How can an acoustic link between real and virtual elements be created in audio augmented reality?**

The first study explored user perceptions of different acoustic reproductions of a controlled office environment, using both high-end studio headphones and acoustically transparent audio glasses. Study 1 used formal listening tests to assess various aspects of auditory perception. Study 2 focused on the plausibility aspect of user perception, and evaluated different acoustic reproductions in real-world spaces, both outdoor and highly reverberant. Study 2 used an AAR game as a more representative task, again exploring how studio headphones and audio glasses compare.

4.1 Study 1 – Influence of Environmental Acoustics and Playback Device in Lab-Based AAR

As a starting point for acoustic linking in AAR, Study 1 focused on the gaps regarding accuracy and playback device mentioned earlier, exploring how the perception of acoustic reproductions is shaped by the playback device as well as the detail of an RIR.[1] 20 participants assessed six acoustic reproductions of a controlled office environment (Reverberation Time = 450ms), shown in Figure 4.1, both with audio glasses and studio-quality reference headphones, completing localisation tests and assessing perceptual characteristics of the virtual audio. A within-subjects

[1] See Section 2.3.3 in Chapter 2 for more information.

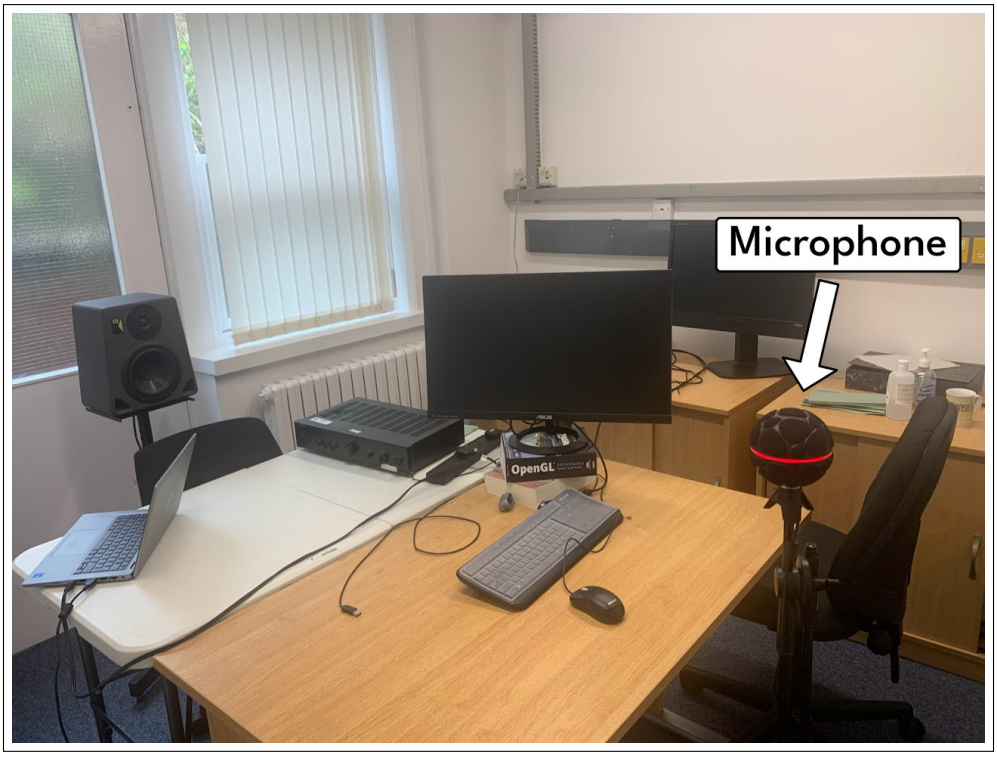

Figure 4.1: Test environment used for Study 1, illustrating the setup for RIR capture. In the user study, the speaker remained in this position and the participant was seated at the microphone position.

design was used, with all participants experiencing all 12 combinations of playback device and acoustic condition. The study lasted approximately one hour.

4.1.1 Experimental Parameters

Study 1's independent variables were acoustic condition, shown in Table 4.1, and playback device, either the Sennheiser HD650 studio headphones[2] or Fauna audio glasses[3]. Neither playback device has a published frequency response graph from the manufacturer, however Sennheiser HD650 headphones are quoted as having a frequency response of 10–41000Hz, while the Fauna glasses are quoted as having a frequency response of 250–20000Hz. Playback devices were chosen to compare auditory perception in critical listening scenarios (headphones) with AAR scenarios (glasses), and are shown in Figure 4.5.

The RIRs under test varied in spatial resolution (the accuracy with which they reproduced directional acoustic cues) and spectral bandwidth (how much of the auditory spectrum they modelled behaviour for), both of which contribute to the 'accuracy' of an RIR. Spatial resolution ranged from single-channel, omnidirectional responses which modelled no directionality but can be captured using widely available microphones or simulated with less computational demand,

[2] https://uk.sennheiser-hearing.com/products/hd-650
[3] https://wearfauna.com/en/

Condition Code	Excitation Source	Spatial Resolution
Dry	None	None
Omni-HC	Handclap	Omni
Stereo-HC	Handclap	Stereo
1O-Sine	Sine Sweep	1st Order Ambisonics
3O-HC	Handclap	3rd Order Ambisonics
3O-Sine	Sine Sweep	3rd Order Ambisonics

Table 4.1: Details of the acoustic conditions used for Study 1.

up to 3rd-order Ambisonics responses which have high spatial resolution but require specialist, often expensive, microphones or computationally demanding simulations to produce. Spectral bandwidth was either limited, using a handclap as an excitation source, or full-spectrum, using a sinusoidal sweep from a loudspeaker. The final six were chosen to cover a range of different accuracy levels, including both complex, specialised RIRs and simpler RIRs which could feasibly be captured by an AAR end-user. Smartphones already feature single or dual-channel microphones as will the AAR devices of the future, so the omnidirectional and stereo handclap RIRs could easily be captured by a novice AAR user without any additional equipment. A dry condition with no RIR was also used as a control.

4.1.2 RIR Capture Process

RIRs of the study test space were captured using a Zylia ZM-1 Ambisonic microphone.[4], loaned from project partners at Imperial College London. The microphone was placed at approximately the ear height of a seated listener, and RIRs were recorded at 3rd-order, the highest resolution the microphone is capable of recording. One 3rd-order RIR was captured of the room being excited by a single handclap, performed 2 metres away from the microphone at a 0° azimuth, chosen to model a lower fidelity, but accessible RIR capture method that AAR users could feasibly carry out themselves. Another 3rd-order RIR was captured of the room being excited by a sinusoidal sweep, covering 20Hz to 20kHz over a period of 5 seconds, to create a high-fidelity RIR modelling behaviour for the full spectrum of human hearing. The sweep was generated using the AURORA plugin suite[5] and played back over a KRK studio monitor speaker, again positioned 2 metres away at a 0° azimuth. The direct sound portion of the RIR was replaced with silence to allow direct sound to be modelled through the spatialisation engine. This ensured that direct sound was rendered at the highest possible level of detail, and was consistent between conditions.

These 3rd-order recordings were then used to produce lower-resolution RIRs:

- 1st-order RIRs were produced by taking the first four (WXYZ) channels of the 3O RIR.

[4] https://www.zylia.co/zylia-zm-1-microphone.html
[5] https://www.aurora-plugins.com/index.htm

- Stereo RIRs were produced by taking the signals from opposing microphone capsules on the ZM-1, recommended by acoustician colleagues at Imperial College London.
- Omni RIRs were produced by taking the 1st (W) channel of the 3O RIR.

Rather than measuring RIRs for each azimuth position a sound was presented at, the same RIR was used for all positions, effectively rotating the reverberant space around the user's head similarly to [55], where this simplification was not found to adversely impact the user experience. Measured RIRs were chosen over simulated ones to isolate the influence of spatial resolution or spectral bandwidth from the accuracy of an acoustic modelling package, while still providing insights that could aid AAR developers when choosing a workflow for simulated RIRs. More detailed reproductions are more computationally demanding to simulate, and so it is still key to determine where the best balance of perceptual benefit and computational complexity lies for acoustic simulations, especially in AAR scenarios which are mobile or battery-constrained.

4.1.3 Experimental Design and Methodology

Participants were seated at a desk in the test environment, directly opposite the KRK loudspeaker used for RIR capture, approximately 2m away. A computer monitor, keyboard and mouse were set up on the desk to allow the participant to see and interact with the experimental interface. Participants evaluated the six different acoustic conditions, once using Sennheiser HD650 open-back headphones, and once using a pair of FAUNA audio glasses. The FAUNA audio glasses are designed to be used wirelessly, but a hardwired development pair were used to remove the influence of wireless latency. The six acoustic conditions are shown in Table 4.1. The order of evaluations was counterbalanced to account for order effects – half of participants used glasses first and half used headphones first, with acoustic conditions for each playback device evaluated in a randomised order.

In each condition, participants completed a localisation test, followed by questions about their perception of the audio. Three sound samples were used as part of the localisation test: a sample of human speech, a sample of acoustic guitar music, and a synthesised 'user interface' (UI) sound designed to evoke notification sounds in existing games or applications. All three sound samples were anechoic, and so free of any reverberation cues of their own which could affect perception. The audio files used in this study are available via Appendix A.2.

Each evaluation procedure was structured as follows:

1. A reference track was played back over the loudspeaker in the room, consisting of the test sounds used in the rest of the experiment. This provided the participant with an understanding of how those sounds 'should' sound, and a real-world reference for making judgments about the virtual sounds.

2. A test sequence was played over the headphones/glasses. The sequence consisted of a white noise burst, played directly in front, to the right, behind, and to the left. The monitor visualised the positions of the noise bursts to give the participant a reference for the audio spatialisation.

3. One of the three sound stimuli was presented to the user to localise, auralised with the appropriate RIR for the current acoustic condition. The user swivelled in their chair to face the sound, and pressed the keyboard spacebar to submit their localisation. The sound was presented for localisation four times, once in each quadrant, in a randomised order. Stimuli varied only in their azimuthal position, and were kept at a 0 degree elevation throughout the study.

4. The participant rated the sound's externalisation, plausibility, realism, and their confidence in their localisations on continuous scales from 0 to 1. These measures are listed in Table 4.2.

5. The participant repeated the localisation and evaluation step for the other two sound stimuli.

6. A questionnaire for the acoustic condition was presented, asking the participant to rate the level of attention they paid to the reverberation, how much of a difference they noticed between the virtual sounds and the reference loudspeaker (both rated on a 7-point Likert scale from "Strongly Disagree" to "Strongly Agree"), and whether they preferred the reference track or virtual sounds (rated on a continuous scale).

7. The participant moved to the next acoustic condition, repeating this process.

The experimental software was developed in the Unity engine. Virtual sounds were spatialised binaurally using the 3DTI Toolkit [44], and the KEMAR dummy head HRTF from the SONICOM dataset [56]. Auralisations for all RIRs were created offline in MATLAB for each anechoic audio file, producing 20-channel audio files corresponding to a dodecahedral loudspeaker layout. Direct sound was rendered using a single instance of the 3DTI Toolkit Unity wrapper, while the reverberant components were decoded to spatialised virtual loudspeakers positioned at the 20 vertices of a dodecahedral speaker layout. Participants were fitted with a Supperware headtracker[6], which tracked 3DOF head movements for localisation trials, and maintained the real-world position of virtual sounds. Both playback devices had output volume levels balanced subjectively to present sounds at an equivalent loudness, and a high-pass filter at 250Hz was applied to the headphones to more closely match the frequency response range of the audio glasses as quoted on the manufacturer website.

[6]https://supperware.co.uk/headtracker-overview

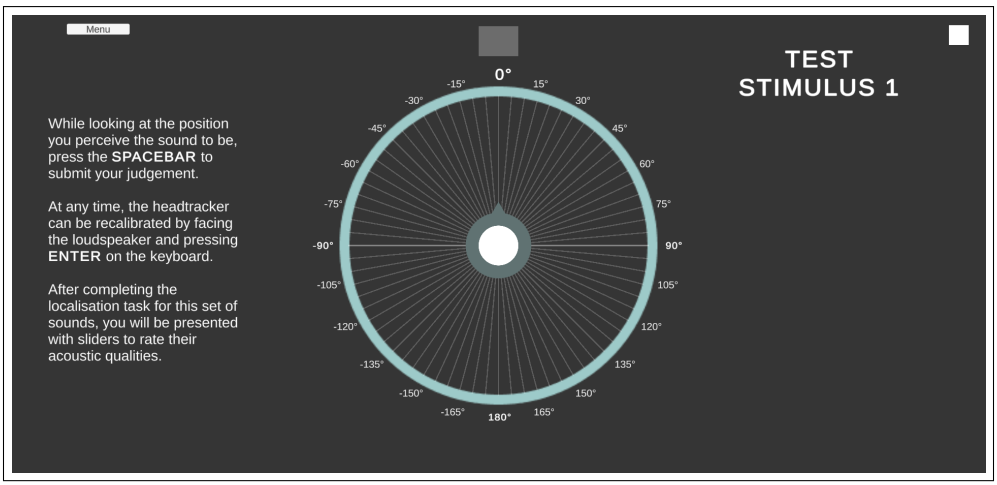

Figure 4.2: Screenshot of the test application used in Study 1, showing the user interface during the localisation test.

Measure	Question
Externalisation	Did these sounds appear to be inside or outside your head?
Plausibility	Do you think these sounds were recorded in this room? (based on [30])
Sound Realism	To what extent did these sounds appear to be part of the real world?
Localisation Confidence	How confident are you that you located the sounds accurately?
Questionnaire - Virtual/Reference Preference	"I preferred how stimuli sounded in the reference track / in this condition."
Questionnaire - Reverberation Attention	"I was conscious of the way the stimuli reverberated."
Questionnaire - Virtual/Reference Discrimination	"I noticed a difference between the sounds in the reference track and the sounds in this condition."

Table 4.2: Questions used in Study 1.

4.1.4 Participants

Twenty people participated in Study 1, recruited from posters and university mailing lists and compensated with a £10 Amazon voucher for their time. To ensure a participant pool that was representative of possible AAR users, the only recruitment criteria were that participants did not have any hearing impairments, and that they could comfortably use a computer screen without wearing glasses to allow for use of the audio glasses. Final participant demographics were:

- **Gender:** 12 men, 8 women
- **Age:** 9 aged 18-24, 6 aged 25-34, 3 aged 35-44, 1 aged 45-54, 1 aged 55-64
- **AR Familiarity:** 1 very unfamiliar, 3 unfamiliar, 4 neutral, 8 familiar, 4 very familiar
- **Spatial Audio Familiarity:** 1 very unfamiliar, 8 unfamiliar, 4 neutral, 6 familiar, 1 very familiar

4.1.5 Results

Localisation error and the measures outlined in Table 4.2 were analysed using ANOVA tests – two-way for questionnaire results (RIR and playback device), and three-way for the other measures (RIR, playback device, and stimulus). Across all measures, data were found not to

Measure	Factor	ANOVA p, F	Sig. Pairwise Comparisons	p	Mean Difference
Plausibility	RIR	**.013, F(5, 684) = 2.932**	Dry - Omni-HC	.033	0.093
			Omni-HC - 3O-HC	.021	-0.099
			Omni-HC - 3O-Sine	.045	-0.096
	Device	.163, F(1, 684) = 1.947			
	Stimulus	**.006, F(2, 684) = 5.116**	Speech - Music	.005	-0.072
Loc. Error	RIR	**< .001, F(5, 2679) = 9.998**	*See interactions*		
	Device	**< .001, F(1, 2679) = 21.742**	*See interactions*		
	Stimulus	.051, F(2, 2679) = 2.973			
	RIR : Device	**.004, F(5, 2679) = 3.452**	1O-Sine*Glasses - Dry*Glasses	< .001	11.6
			1O-Sine*Glasses - Dry*HD650	.002	10.7
			1O-Sine*Glasses - 3O-HC-HD650	.005	11.4
			1O-Sine*HD650 - 3O-HC*HD650	.005	6.3
			Dry*Glasses - Stereo-HC*Glasses	< .001	-13
			Dry*Glasses - 3O-HC*Glasses	.004	-7.7
			Dry*Glasses - 3O-Sine*Glasses	< .001	-16
			Dry*HD650 - Stereo-HC*Glasses	< .001	-12.1
			Dry*HD650 - 3O-Sine*Glasses	< .001	-15.1
			Omni-HC*HD650 - Stereo-HC*Glasses	.009	-11.4
			Omni-HC*HD650 - 3O-Sine*Glasses	.004	-14.4
			Stereo-HC*Glasses - 3O-HC*HD650	< .001	-12.8
			3O-HC*HD650 - 3O-Sine*Glasses	< .001	-15.8
			3O-Sine*Glasses - 3O-Sine*HD650	.029	12.9
	Device : Stimulus	**.038, F(2, 2679) = 21.7422**	Glasses*UI - HD650*Speech	.001	8.6
			Glasses*Music - HD650*Music	< .001	7.9
			Glasses*Music - HD650*Speech	< .001	10.2
			Glasses*Speech - HD650*Speech	.002	8.6
			HD650*UI - HD650*Speech	.005	6.2
Externalisation	RIR	**< .001, F(5, 684) = 5.833**	Dry - Omni-HC	.021	-0.112
			Omni-HC - 3O-HC	< .001	-0.029
			Omni-HC - 1O-Sine	< .001	-0.039
			Omni-HC - 3O-Sine	< .001	-0.018
	Device	**< .014, F(1, 684) = 7.761**	HD650 - Glasses	.014	0.052
	Stimulus	**< .001, F(2, 684) = 9.585**	UI - Speech	< .001	-0.093
			UI - Music	.002	-0.078
Loc. Confidence	RIR	**< .001, F(5, 684) = 14.014**	Dry - Stereo-HC	< .001	0.177
			Dry - 3O-HC	.015	0.082
			Dry - 1O-Sine	< .001	0.132
			Dry - 3O-Sine	< .001	0.158
			Omni-HC - Stereo-HC	< .001	0.158
			Omni-HC - 1O-Sine	.003	0.107
			Omni-HC - 3O-Sine	< .001	0.133
			Stereo-HC - 3O-HC	< .001	-0.095
			3O-HC - 3O-Sine	.048	0.076
	Device	**.004, F(1, 684) = 8.306**	HD650 - Glasses	.004	0.055
	Stimulus	**.014, F(2, 684) = 4.233**	UI - Speech	.018	-0.051
Realism	RIR	.628, F(5, 684) = 0.694			
	Device	.158, F(1, 684) = 1.992			
	Stimulus	**< .001, F(2, 684) = 26.153**	UI - Speech	< .001	-0.126
			UI - Music	< .001	-0.156
[Q] Sound Preference	RIR	.870, F(5, 228) = 0.368			
	Device	.074, F(1, 228) = 3.233			
[Q] Reverb Attention	RIR	.131, F(5, 228) = 1.721			
	Device	.642, F(1, 228) = 0.217			
[Q] Sound Differentiation	RIR	.696, F(5, 228) = 0.605			
	Device	.322, F(1, 228) = 0.985			

Table 4.3: Overall ANOVA and post hoc results for each measure in Study 1.

be normally distributed, and so the Aligned Rank Transform method was used to account for this violation of ANOVA assumptions, with *post hoc* analysis conducted using Tukey HSD tests. Table 4.3 details the full results. Localisation error was calculated using great-circle error, which has been suggested as a more accurate measure for computing localisation error in a three-dimensional field [145], shown in Equation 4.1. While audio was presented only in the azimuthal plane, where great-circle error is equivalent to azimuth localisation error, the great-circle error calculation was still used to allow for the possibility of comparing results with any future studies presenting audio at multiple elevations. As participants' localisations were recorded on their 3DOF head orientation, the great-circle error also accounts for any elevation differences perceived by participants. The full dataset is available via Appendix A.1.

$$great\ circle\ error = \arctan\left(\frac{||xyz_{target} \times xyz_{response}||}{xyz_{target} \cdot xyz_{response}}\right) \tag{4.1}$$

Plausibility

The only significant plausibility findings for the different acoustic conditions were the omnidirectional handclap being less plausible than the 3rd-order RIRs and the dry condition. No significant differences were found when directly comparing the plausibility of 1st- and 3rd-order RIRs, or when directly comparing the influence of impulsive source. The omnidirectional handclap RIR showed a higher degree of externalisation compared to the dry condition, and the three Ambisonic RIRs were also found to have a higher degree of externalisation than the omnidirectional handclap, though the difference between those RIRs and the dry condition was not significant.

Localisation Error

Localisation error was found to have significant main effects of RIR and playback device, as well as interaction effects between RIR and device, and device and stimulus. Notable pairwise comparisons included glasses having a higher localisation error than headphones when using the 3rd-order sine RIR, the 1st-order sine condition having higher localisation error than the 3rd-order handclap condition when using headphones, and glasses exhibiting higher localisation error in the ambisonic conditions and the stereo handclap condition compared to the dry control. No significant differences were found between the 1st- and 3rd-order RIRs when spectral bandwidth was the same.

Significant main effects were also found for all three variables when it came to a user's confidence in their localisations. Participants were more confident in their localisations in the dry and omnidirectional conditions than using RIRs with higher spatial resolution. The stereo handclap RIR also resulted in lower localisation confidence than with the 3rd-order handclap RIR, and participants showed higher localisation confidence using the 3rd-order handclap RIR

than the 3rd-order sine sweep.

Playback Device, Test Stimuli, and Other Measures

Playback device was not found to affect plausibility, though participants showed higher localisation confidence when using the headphones compared to the glasses and a higher level of externalisation. Glasses also exhibited higher localisation error than headphones when assessing the music or speech stimuli, as well us under the 3rd-order sine sweep condition. The speech stimulus was found to be less plausible than the music stimulus, and the UI stimulus resulted in higher localisation error than the speech stimulus when listened to over headphones, with participants less confident in their localisations of the UI stimulus compared to the speech stimulus overall. The UI stimulus was also rated significantly less realistic than the speech and music stimuli. The realism measure was found to be unaffected by acoustic condition or playback device, and similarly all three of the final questionnaire measures showed no significant differences between acoustic conditions, or between playback devices.

4.1.6 Discussion

Overall, the results from Study 1 provided some valuable initial insights on acoustic linking for AAR. The plausibility results align with prior work which finds that the inclusion of acoustic cues improves plausibility. However, the results also showed that the simplest, least accurate RIR actually reduced plausibility compared to the control condition, suggesting that there is a minimum level of accuracy required to hear these benefits. However, the results did not provide enough nuance to understand where this minimum level lies, or where the best balance of plausibility and complexity lies.

The localisation results (both in terms of great-circle error and subjective confidence) align with prior work which finds that reverberation can muddy localisation cues. This is illustrated by the results which show the dry condition outperforming most reverberant conditions. As the results also show that plausibility and externalisation can benefit from reverberant cues, it will be important for future AAR applications to balance these two aspects: maximising the benefits of acoustic cues while minimising or accounting for the impediment to localisation.

Study 1's most useful findings primarily concern playback device. The results clearly show that the audio glasses suffer perceptually compared to the headphones, with higher localisation error, lower localisation confidence, and lower externalisation ratings. This is not surprising, given the glasses feature smaller speakers which may not be as capable as the larger drivers in the HD650s; they represent a novel form factor compared to headphones which many participants will have prior experience with; and the headphones are significantly more expensive, retailing for two or three times more than the glasses which may represent superior design and manufacturing processes. While localisation and externalisation suffered with the glasses, no-

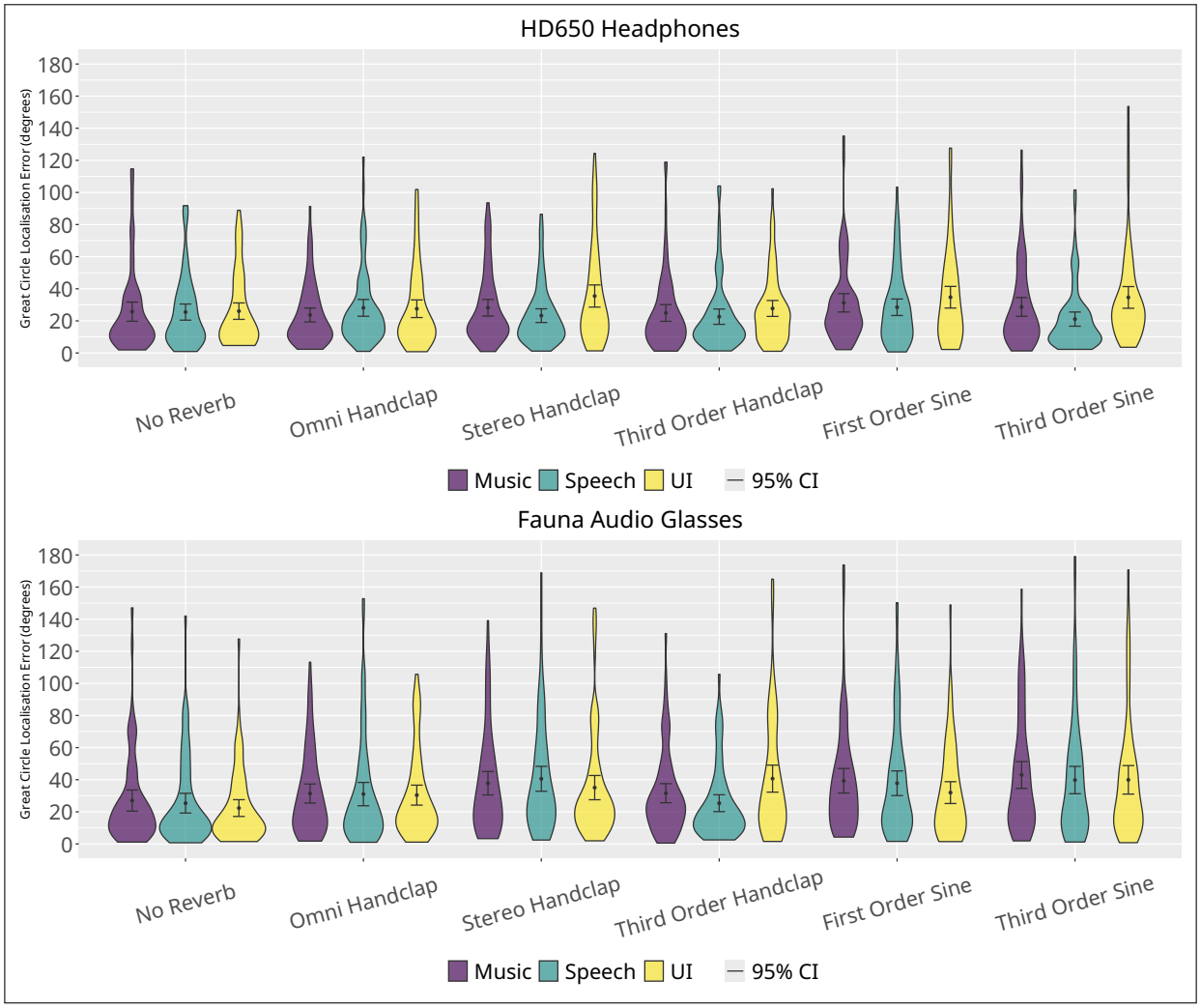

Figure 4.3: Violin plot of localisation error results, separated by playback device, RIR, and stimulus.

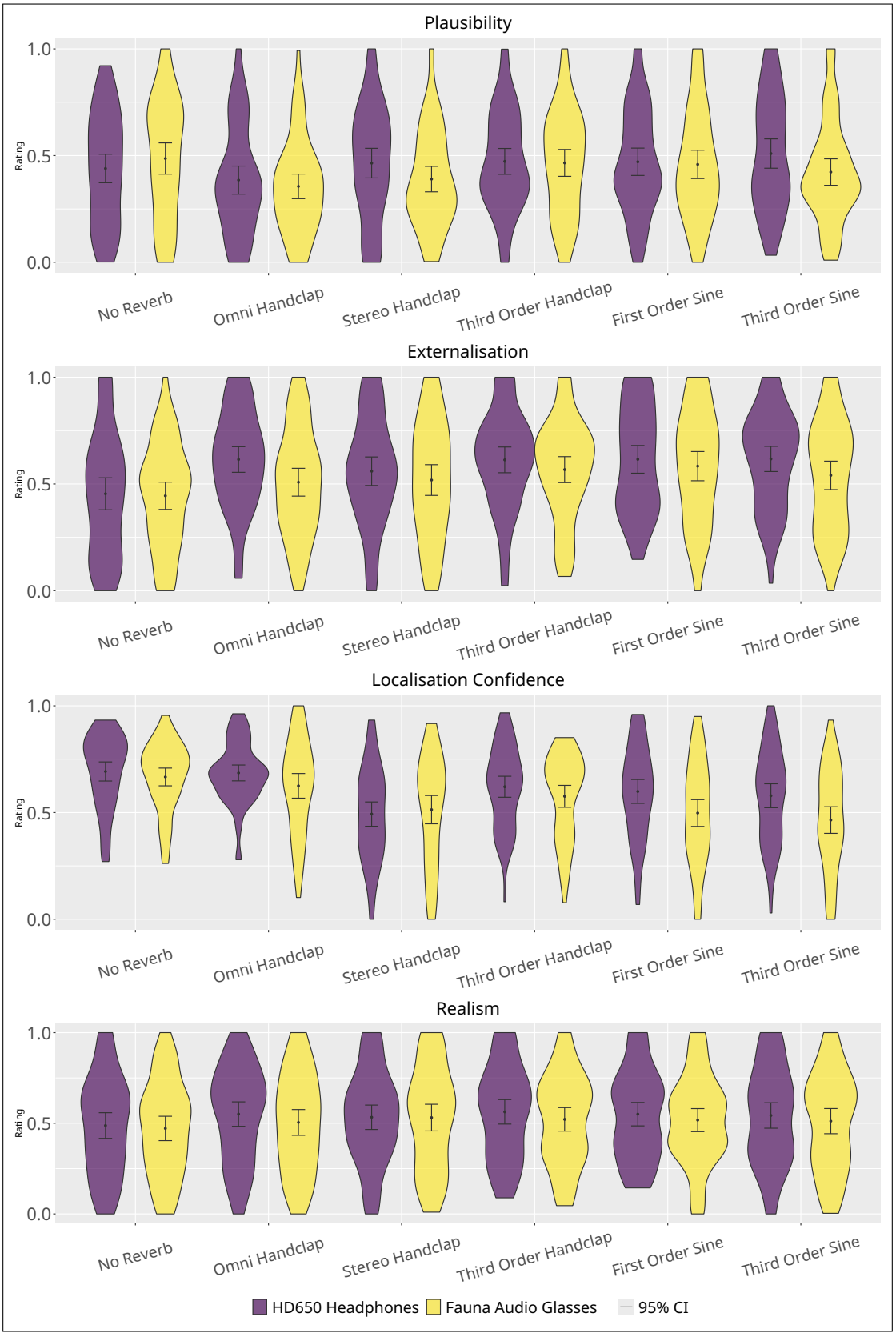

Figure 4.4: Violin plot of main quantitative results, separated by RIR and playback device.

Figure 4.5: The two playback devices used in Studies 1 and 2: Sennheiser HD650 headphones (L) and wired Fauna audio glasses (R).

tably the plausibility and realism measures showed no significant differences between the two, suggesting that for AAR scenarios where localisation is not critical, audio glasses could still represent a viable platform. Even in localisation-centric scenarios, mean localisation errors were inflated by approximately 15°, which could potentially be accounted for by AAR applications.

The results of Study 1 also only represent one space, under formal listening test conditions. A more focused exploration of plausibility was required to identify the ideal level of RIR accuracy for acoustic linking, and explore how perception is affected by real-world AAR scenarios. This formed the motivation for Study 2.

4.2 Study 2 – Influence of Environmental Acoustics and Playback Device in Real-World AAR

Based on the findings from Study 1, Study 2 was designed to focus more closely on the plausibility of acoustic conditions, and assess them in a more realistic AAR scenario than Study 1's lab-based critical listening test. With this in mind, Study 2 investigated multiple real-world environments (designed to both to represent real-world usage locations and to reveal greater differences between RIRs), using a representative AAR application.

A gamified AAR localisation test was developed, where participants took on the role of a sonic wizard or 'sonomancer', and battled sonic 'monsters' by localising them correctly. Par-

ticipants completed these gamified localisation tests in two public spaces, one outdoors and one highly reverberant, using the same headphones and audio glasses as Study 1, and under one of eight acoustic conditions, assessing the plausibility and externalisation of the game sounds afterwards. Between localisation tests, participants also completed a standard MUSHRA critical listening test, directly comparing 8 acoustic reproductions of the current test environment.

4.2.1 Experimental Parameters

Study 2's independent variables were acoustic condition, shown in Table 4.4, playback device (again using the Sennheiser HD650 studio headphones or the Fauna audio glasses), and listening environment (high reverberance space and outdoor space). Listening environments were chosen to represent real-world public environments where AAR may be used, and to cover a range of environments when taken alongside Study 1. The high reverberance environment was also chosen with the aim of exposing more differences between RIR due to its more dramatic acoustic influence. The University of Glasgow Cloisters were used for the high reverberance environment (RT = 2.7s, calculated from captured RIRs), with the University's East Quadrangle used for the outdoor environment (RT = 35ms). Figures 4.6 and 4.7 show these test spaces.

Study 2 utilised the same forms of RIR as Study 1, but introduced two additional 'echoic' RIR conditions. In principle, anechoic source audio should be used with RIRs as it is free of reverberation that may influence the final output. However, anechoic audio like this requires specialist spaces to record, or must be synthesised to be free of reverberation, significantly limiting the audio an AAR developer or AAR sound designer could work with. These conditions were designed to explore how important the echoicity of this source audio is to the listener's perception, and how strictly this principle must be adhered to in an AAR context. Artificial reverberation was added to source audio before RIR processing in these conditions to simulate moderately echoic source audio.

4.2.2 RIR Capture Process

In both test spaces, RIRs were captured using the same process as Study 1. The Zylia ZM-1 microphone was set up at the planned listening position for the user, and then a handclap and sinusoidal sweep were captured from 2m away. The handclap was produced by the same individual, and the sinusoidal sweep reproduced using the same speaker system, as in Study 1. The same MATLAB-based convolution process was also used. The RIR capture setup is illustrated in Figures 4.6 and 4.7. For the echoic conditions introduced in Study 2, a synthetic reverb (IEM FDN-Reverb[7], with Room Size of 20, Reverberation Time of 1s, and Dry/Wet mix of 0.5) was applied to the anechoic source audio files before convolution. The full list of RIR conditions for Study 2 is shown in Table 4.4.

[7]https://plugins.iem.at/

Condition Code	Excitation Source	Spatial Resolution
Omni-HC	Handclap	Omni
Omni-HC-Echoic	*Handclap*	*Omni*
Stereo-HC	Handclap	Stereo
1O-Sine	Sine Sweep	1st Order Ambisonics
3O-HC	Handclap	3rd Order Ambisonics
3O-Sine	Sine Sweep	3rd Order Ambisonics
3O-Sine-Echoic	*Sine Sweep*	*3rd Order Ambisonics*

Table 4.4: Details of the RIRs used in Study 2. Artificial reverberation was added to source audio when using the italicised RIRs.

Figure 4.6: High reverberance test environment used for Study 2, illustrating the setup for RIR capture.

Figure 4.7: Outdoor test environment used for Study 2, illustrating the setup for RIR capture.

4.2.3 Experimental Design and Methodology

Study 2 utilised a broadly similar methodology to Study 1, with some key changes. Firstly, the measures were revised to focus more on plausibility as the results from Study 1 exposed minimal plausibility findings. Two additional plausibility measures were introduced to measure the sensation more rigorously, and other measures were removed, save for externalisation. Study 2 measures are outlined in Table 4.5. Secondly, a MUSHRA listening test (outlined in ITU Recommendation BS.2132-0 [151]) was employed to compare all RIRs directly, with participants making blind comparisons of all acoustic conditions and rating plausibility for each. While the MUSHRA test was initially designed for audio codec comparisons, the overall methodology was suggested by acoustician colleagues in Imperial College London as appropriate for the study's goals. With eight acoustic conditions, it was infeasible to have participants complete a game round for each acoustic condition for each combination of playback device and space, so the localisation game was rendered with a randomly chosen RIR. Weighting was applied to this randomisation so that each RIR was selected a roughly even number of times across the whole dataset.

In contrast to Study 1, no real-world reference for the stimuli used in the game was provided, though as public spaces there were other real-world sounds present. Instead, participants were asked to base their plausibility judgements on their internal reference and expectation for the space's acoustics as this better reflects the user experience of an AAR game or application.

The study had four conditions, one for each of the four combinations of playback device and

test space. In each condition, participants completed a round of the localisation game, then filled in a questionnaire about their perception of the game sounds, and then a MUSHRA listening test evaluating the plausibility of the full selection of RIRs. The order of test environments, and the order of playback devices within each environment, was fully counterbalanced. The study lasted approximately one hour. The full procedure was structured as follows:

1. A test sequence was played to the participant over the current playback device. As in Study 1, the sequence was a spatialised white noise burst, played directly in front, to the right, behind, and then to the left. The laptop screen visualised the noise burst positions to give the participant a reference for the spatialisation.

2. A game narrator gave an introduction to the game scenario: sonic monsters are nearby, and as a sonomancer the participant must localise and destroy them.

3. The participant played the game for two minutes, localising and destroying as many sonic monsters as possible. The sonic monsters were signified using an ominous designed sound loop reminiscent of the Predator or other pop culture monsters, presented at a randomised location in the azimuthal plane. As in Study 1, audio was presented at a 0 degree elevation. Monsters had an angular size of 30°, with localisation anywhere within that angular area considered a successful localisation. 30°was chosen subjectively as an appropriately difficult game experience for non-expert listener participants. A congratulatory or commiseratory voice line was played depending on whether the player successfully localised the monster, before a new monster appeared at a new azimuth position. This process repeated until the end of the game experience.

4. The participant assessed the game sounds on all four measures listed in Table 4.5.

5. The participant was presented with a MUSHRA listening test. Audio stimuli were presented consisting of the Study 1 speech sample rendered under each of the acoustic conditions, each with a 0-100 slider. The participant listened to each, rating it from 0-100 according to one of the plausibility measures. The participant could freely revise their ratings and audition each sound until they were happy with their ratings. The participant then repeated the MUSHRA test for the other two plausibility questions, with plausibility questions presented in a randomised order.

    - In this first set of MUSHRA tests, one of the stimuli was selected at random and represented twice in the test, allowing for a measure of rater reliability. If a participant rated duplicate stimuli significantly differently across the study, they were deemed an unreliable rater and their data were discarded.

6. This process repeated until the participant had completed three game rounds and two MUSHRA rounds.

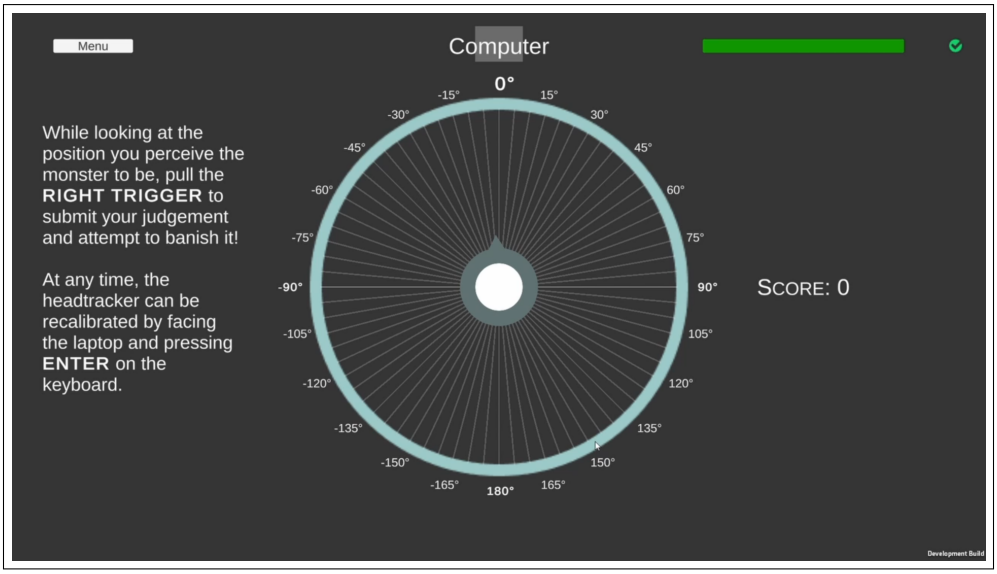

Figure 4.8: Screenshot of the test application used in Study 2, showing the user interface during the localisation game.

Measure	Question
Externalisation	Did the game sounds appear to be inside or outside your head?
Plausibility (Brinkmann)	Do you think these sounds were recorded in this room? (based on [30])
Plausibility (Definition)	Rate the plausibility of the sound[s] [you heard during the game].
Plausibility (Realism)	Did the [game] sound[s] you heard sound as if they could believably be in this real space?

Table 4.5: Questions used in Study 2. Questions varied slightly between the game and the listening test, as shown by the [square brackets].

7. The participant moved to the next playback device, starting the whole process over. When both playback devices had been fully evaluated, the process repeated for the next test space.

As in Study 1, the experimental software was developed in the Unity engine, using the 3DTI Toolkit for audio spatialisation, with the same reverberation decoding system. Participants used the same headtracker, and the same playback devices with the same processing as in Study 1. The user interface of the software is shown in Figures 4.8 and 4.9, and the audio files used in this study are available via Appendix A.2.

4.2.4 Participants

Twenty-four people participated in Study 2 (an additional four people participated but were excluded due to technical glitches or low reliability in the MUSHRA test), again recruited from posters and university mailing lists, who were compensated with a £10 Amazon voucher for their time. As in Study 1, the only recruitment criterion was that participants did not have any hearing impairments and could use a computer screen while wearing the audio glasses. The final participant demographics were:

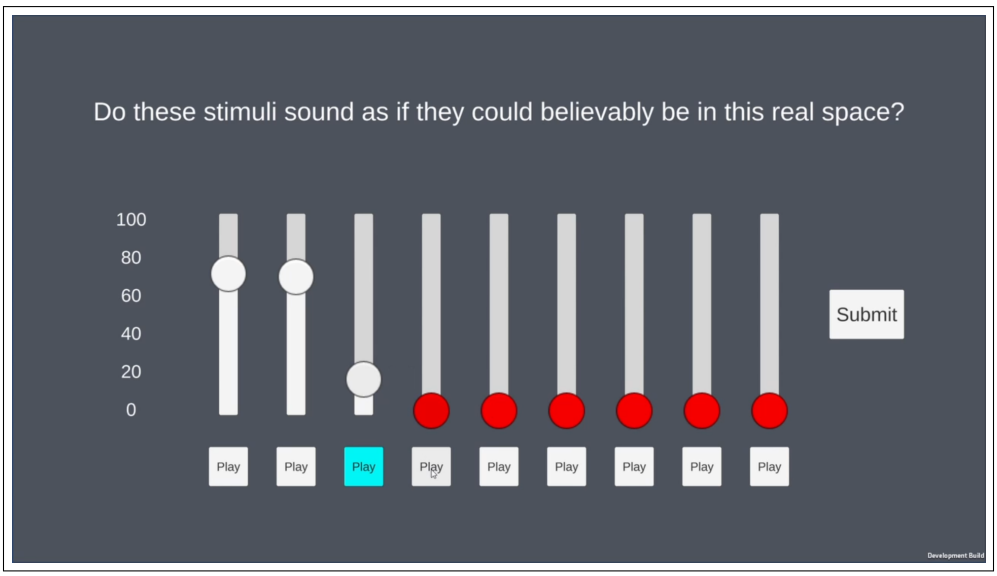

Figure 4.9: Screenshot of the test application used in Study 2, showing the user interface during the MUSHRA listening test.

- **Gender:** 13 men, 10 women, 1 non-binary person.
- **Age:** 7 aged 18-25, 9 aged 25-34, 6 aged 35-44, 2 aged 45-54
- **AR Familiarity:** 2 very unfamiliar, 3 unfamiliar, 4 neutral, 11 familiar, 4 very familiar
- **Spatial Audio Familiarity:** 2 very unfamiliar, 3 unfamiliar, 6 neutral, 9 familiar, 4 very familiar.

4.2.5 Results

For Study 2, the three plausibility questions were aggregated together into one measure, as there were no significant differences between how participants answered them in the listening test, and answers were highly correlated ($p < .001$). Data from both the game evaluations and MUSHRA tests were analysed using a three-way ANOVA (for RIR, playback device, and test space). Once again, data were found not to be normally distributed, and the Aligned Rank Transform method was used to account for this. *Post hoc* analysis was conducted using Tukey HSD tests. Results are shown in Tables 4.6 and 4.7. The full dataset is available via Appendix A.1.

MUSHRA Results

Analysis of the plausibility ratings from the MUSHRA test, shown in Figure 4.11, revealed a complex interaction between RIR and space. Generally, the results showed that RIRs improved plausibility ratings compared to the dry control condition in the high reverberance space, while RIRs decreased plausibility ratings compared to the dry control condition in the outdoor space. In both spaces, the stereo handclap RIR performed more like the dry control condition than any

Measure	Factor	ANOVA p, F	Sig. Pairwise Comparisons	p	Mean Difference
Plausibility	RIR	.563, F(7, 832) = 0.830			
	Device	**.040, F(1, 832) = 4.231**	*See interactions*		
	Space	.051, F(1, 832) = 3.832			
	RIR:Device:Space	**< .001, F(7, 832) = 5.924**	1O-Sine*HD650*Reverberant - Omni-HC*Glasses*Reverberant	.034	0.197
			1O-Sine*HD650*Reverberant - Stereo-HC*Glasses*Outdoors	.046	0.158
			Omni-HC*Glasses*Reverberant - 3O-Sine*Glasses*Reverberant	.044	-0.209
Loc. Error	RIR	**< .001, F(7, 2392) = 4.446**	*See interactions*		
	Device	**.002, F(1, 2392) = 10.098**	*See interactions*		
	Space	.656, F(1, 2392) = 0.198			
	RIR:Device	**.003, F(7, 2392) = 3.040**	*See interactions*		
	RIR:Space	**.007, F(7, 2392) = 2.781**	*See interactions*		
	RIR:Device:Space	**.045, F(7, 2392) = 2.054**	3O-HC*Glasses*Reverberant - Dry*HD650*Outdoors	.031	12.375
			3O-HC*Glasses*Outdoors - Dry*HD650*Outdoors	.020	13.929
			3O-HC*HD650*Reverberant - 3O-Sine-Echoic*Glasses*Outdoors	.030	-11.025
			3O-HC*HD650*Reverberant - Dry*Glasses*Outdoors	.046	-9.365
			3O-HC*HD650*Reverberant - Stereo-HC*Glasses*Outdoors	.045	-5.437
			3O-HC*HD650*Outdoors - Dry*HD650*Outdoors	< .001	8.276
			3O-Sine-Echoic*Glasses*Reverberant - Dry*HD650*Outdoors	< .001	14.147
			3O-Sine-Echoic*Glasses*Reverberant - Omni-HC-Echoic*Glasses*Reverberant	.041	12.067
			3O-Sine-Echoic*HD650*Reverberant - Dry*HD650*Outdoors	< .001	8.276
			3O-Sine-Echoic*HD650*Reverberant - Omni-HC-Echoic*Glasses*Reverberant	.017	13.618
			3O-Sine-Echoic*HD650*Reverberant - Omni-HC*Glasses*Outdoors	.048	12.181
			3O-Sine*Glasses*Outdoors - Dry*HD650*Outdoors	.015	12.767
			Dry*Glasses*Outdoors - Dry*HD650*Outdoors	< .001	14.037
			Dry*Glasses*Outdoors - Omni-HC-Echoic*Glasses*Reverberant	.032	11.957
			Dry*HD650*Outdoors - Omni-HC-Echoic*HD650*Outdoors	< .001	-9.993
			Dry*HD650*Outdoors - Stereo-HC*Glasses*Reverberant	< .001	-10.109
			Omni-HC-Echoic*Glasses*Reverberant - Stereo-HC*Glasses*Reverberant	.032	-8.029
Externalisation	RIR	.926, F(7, 256) = 0.357			
	Device	.923, F(1, 256) = 0.009			
	Space	.336, F(1, 256) = 0.927			
	RIR:Device:Space	**.041, F(7, 256) = 2.126**	*No significant pairwise comparisons*		

Table 4.6: Overall ANOVA and post hoc results for the game experience in Study 2.

Measure	Factor	ANOVA p, F	Sig. Pairwise Comparisons	p	Mean Difference
Plausibility	RIR	**< .001, F(7, 4864) = 9.345**	*See interactions*		
	Device	.0310, F(1, 4864) = 1.031			
	Space	**< .001, F(1, 4864) = 59.951**	*See interactions*		
	RIR:Space	**< .001, F(7, 4864) = 104.509**	1O-Sine*Reverberant - 1O-Sine*Outdoors	< .001	15.223
			1O-Sine*Reverberant - Dry*Reverberant	< .001	27.926
			1O-Sine*Reverberant - Stereo-HC*Reverberant	< .001	23.575
			1O-Sine*Outdoors - Dry*Outdoors	< .001	-18.207
			1O-Sine*Outdoors - Omni-HC-Echoic*Ourdoors	.008	-12.300
			1O-Sine*Outdoors - Stereo-HC*Outdoors	< .001	-14.633
			3O-HC*Reverberant - 3O-HC*Outdoors	< .001	20.270
			3O-HC*Reverberant - Dry*Reverberant	< .001	27.144
			3O-HC*Reverberant - Stereo-HC*Reverberant	< .001	22.793
			3O-HC*Outdoors - 3O-Sine-Echoic*Outdoors	.013	-7.371
			3O-HC*Outdoors - 3O-Sine*Outdoors	< .001	-9.640
			3O-HC*Outdoors - Dry*Outdoors	< .001	-24.029
			3O-HC*Outdoors - Stereo-HC*Outdoors	< .001	-20.456
			3O-Sine-Echoic*Reverberant - 3O-Sine-Echoic*Outdoors	< .001	13.781
			3O-Sine-Echoic*Reverberant - Dry*Reverberant	< .001	28.026
			3O-Sine-Echoic*Reverberant - Stereo-HC*Reverberant	< .001	23.675
			3O-Sine-Echoic*Outdoors - Dry*Outdoors	< .001	-16.658
			3O-Sine-Echoic*Outdoors - Omni-HC-Echoic*Outdoors	< .001	9.149
			3O-Sine-Echoic*Outdoors - Stereo-HC*Outdoors	< .001	-13.084
			3O-Sine*Reverberant - 3O-Sine*Outdoors	< .001	12.040
			3O-Sine*Reverberant - Dry*Reverberant	< .001	28.553
			3O-Sine*Reverberant - Stereo-HC*Reverberant	< .001	24.203
			3O-Sine*Outdoors - Dry*Outdoors	< .001	-14.390
			3O-Sine*Outdoors - Omni-HC-Echoic*Outdoors	< .001	11.417
			3O-Sine*Outdoors - Omni-HC*Outdoors	< .001	10.244
			3O-Sine*Outdoors - Stereo-HC*Outdoors	< .001	-10.816
			Dry*Reverberant - Dry*Outdoors	< .001	-30.903
			Dry*Reverberant - Omni-HC-Echoic*Reverberant	< .001	-24.996
			Dry*Reverberant - Omni-HC*Reverberant	< .001	-26.825
			Dry*Outdoors - Omni-HC-Echoic*Outdoors	< .001	25.807
			Dry*Outdoors - Omni-HC*Outdoors	< .001	24.633
			Omni-HC-Echoic*Reverberant - Omni-HC-Echoic*Outdoors	< .001	19.900
			Omni-HC-Echoic*Reverberant - Stereo-HC*Reverberant	< .001	20.646
			Omni-HC-Echoic*Outdoors - Stereo-HC*Outdoors	< .001	-22.233
			Omni-HC*Reverberant - Omni-HC*Outdoors	< .001	20.555
			Omni-HC*Reverberant - Stereo-HC*Reverberant	< .001	22.474
			Omni-HC*Outdoors - Stereo-HC*Outdoors	< .001	-21.060
			Stereo-HC*Reverberant - Stereo-HC*Outdoors	< .001	-22.979

Table 4.7: Overall ANOVA and *post hoc* results for MUSHRA tests in Study 2. As the *RIR:Space* interaction had 79 significant pairwise comparisons, only comparisons with a common factor are shown here for brevity.

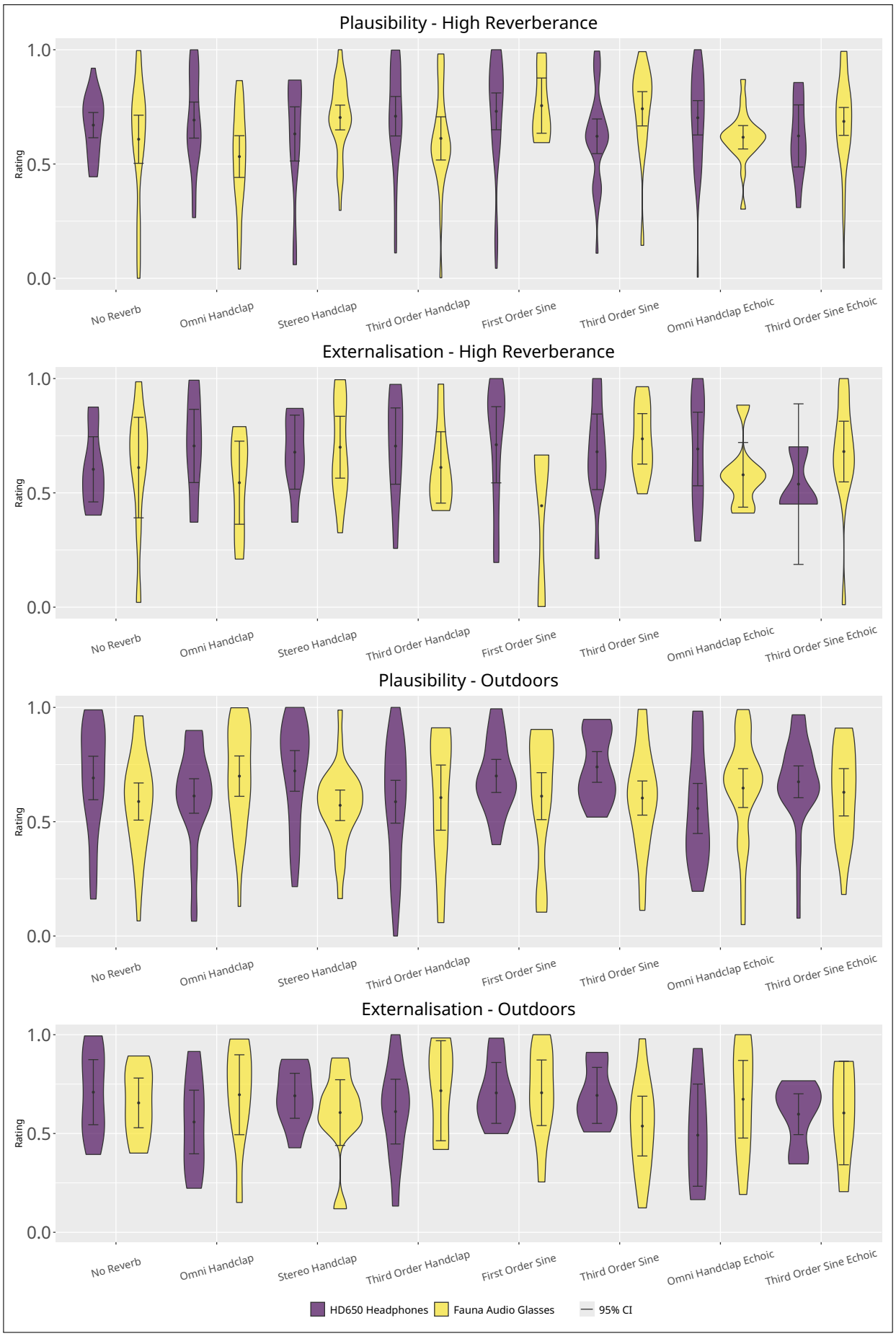

Figure 4.10: Game ratings, separated by playback device, RIR, and test space.

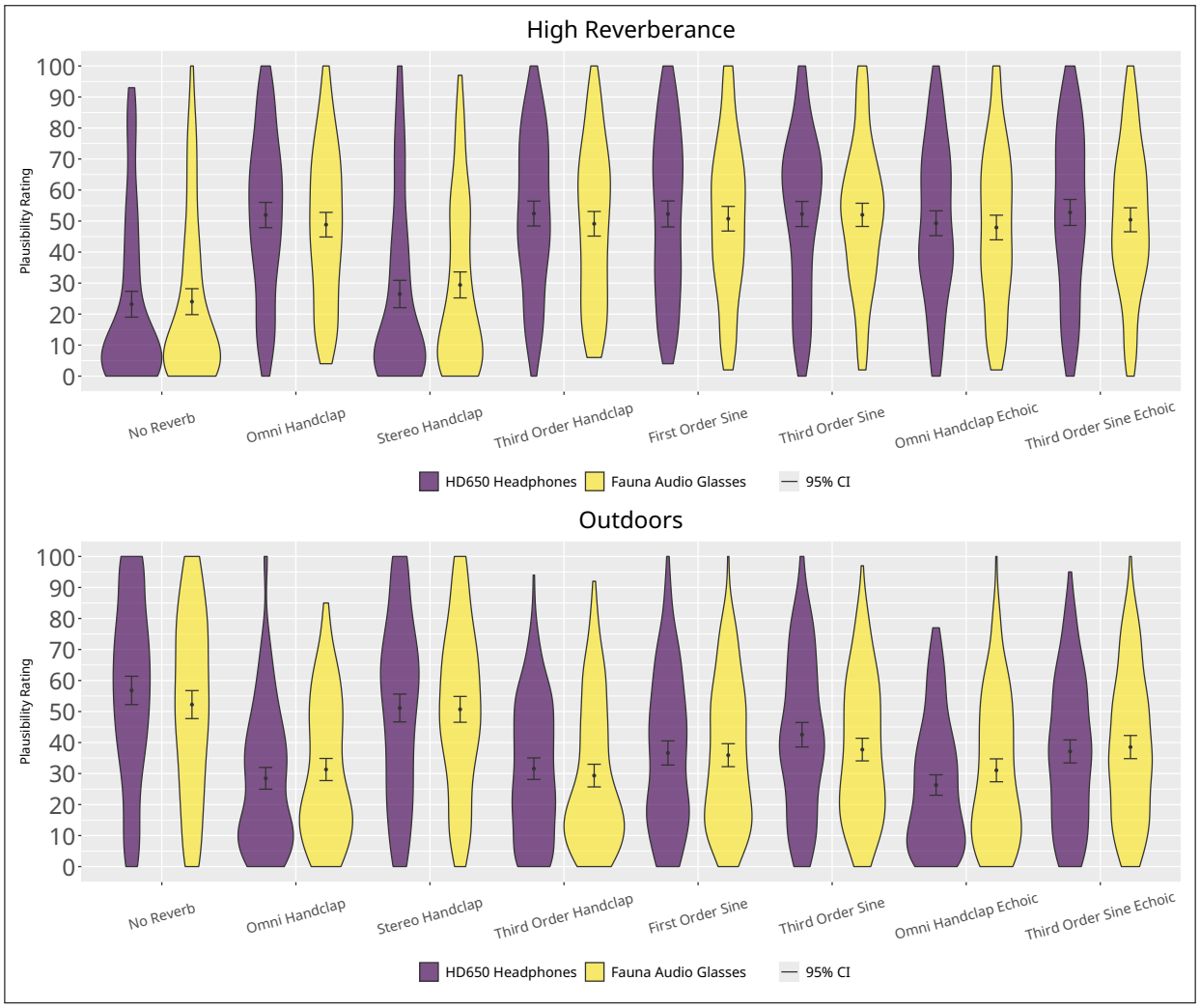

Figure 4.11: MUSHRA plausibility results, separated by RIR, playback device, and test space.

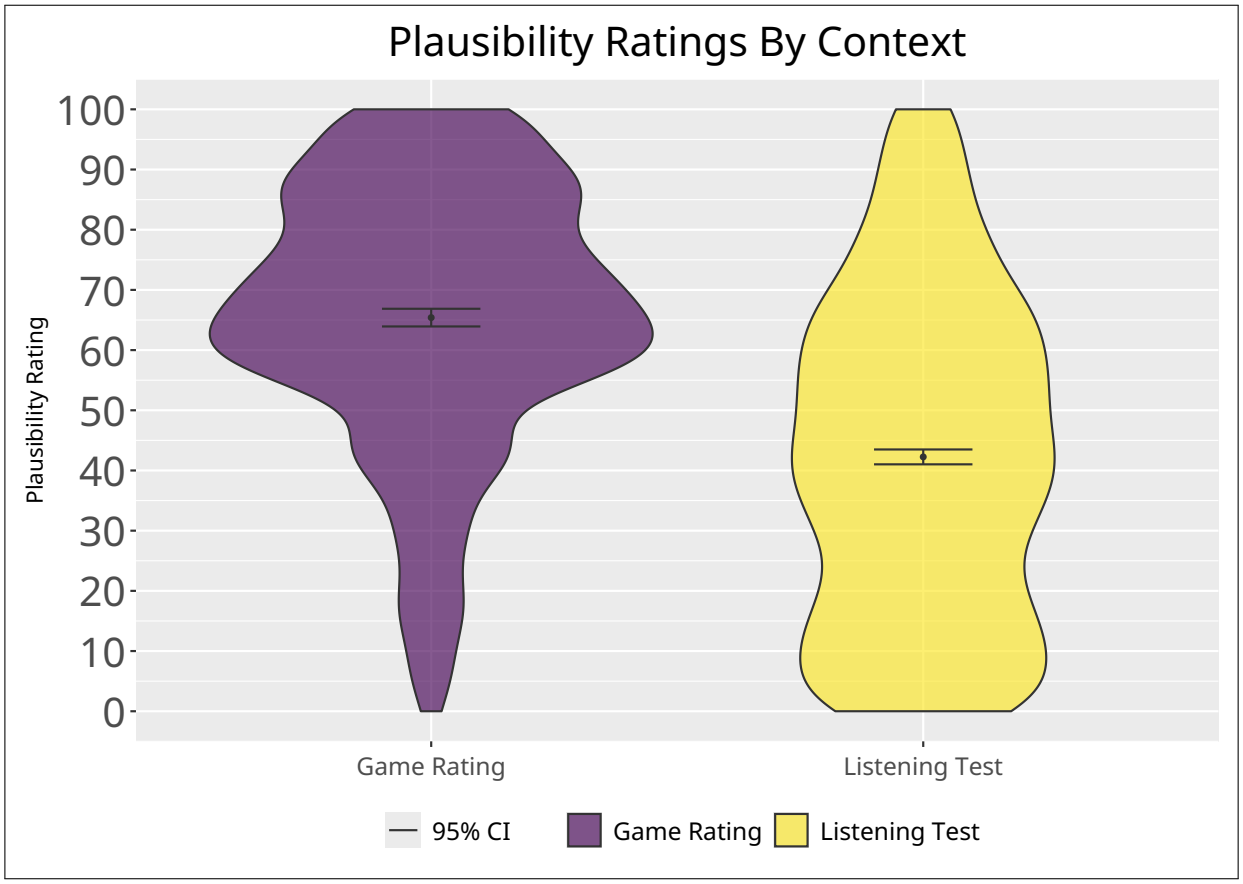

Figure 4.12: Comparison of plausibility ratings given by participants for the same RIR in the game and listening test contexts.

of the other RIRs, rather than performing between the omni handclap and 1st-order RIRs as was expected. A significant comparison between the 3rd-order handclap RIR and the 3rd-order sine sweep RIR in the outdoor space also suggested a lower level of plausibility when a handclap was used as an impulsive source compared to a full-spectrum sweep. No difference in plausibility was detected between the 1st- and 3rd-order sinusoidal sweep RIRs, nor was any plausibility difference detected between the echoic RIRs and their anechoic equivalents.

Game Results

Analysis of the ratings given after the game experience (where participants did not experience all RIRs, and so comparisons are between-subjects), shown in Figure 4.10, revealed a complex, three-way interaction between RIR, playback device, and test space. The omnidirectional handclap RIR was rated as less plausible than the 3rd-order sine sweep RIR when using audio glasses in the high reverberance space, and was deemed less plausible with glasses than the 1st-order sine sweep when using headphones in that space. The only other significant comparison was between the 1st-order sine sweep using headphones in the high reverberance space and the stereo handclap using glasses outdoors. Overall, the results showed no strong trends for plausibility in the game scenario. Localisation error was higher with glasses than headphones in the outdoor condition with no reverberation, and similarly the other significant comparisons did not suggest any strong trends for localisation error. The same three-way interaction was detected for externalisation ratings, though there were no significant pairwise comparisons.

The most notable finding when assessing plausibility ratings in the game was comparing game ratings with MUSHRA ratings. When comparing the plausibility rating given for an RIR after the gameplay task with the rating given to the same RIR in the MUSHRA test, it was found that participants rated sounds as being significantly more plausible during the game than as part of the listening test ($p < .01$, with mean rating of 65.4 in game and 42.25 in listening test), as shown in Figure 4.12.

4.2.6 Discussion

The results from Study 2 provided further insights on plausibility with acoustic reproduction, however, as a broad exploration of acoustic linking in AAR, further work will be required to fully confirm Study 2's implications.. Firstly, Study 2 provides evidence to suggest that spectral bandwidth of an acoustic reproduction contributes to plausibility, with significant differences found between the full-bandwidth sinusoidal sweep RIRs and the limited-bandwidth handclap RIRs. However, this difference was only found in the outdoor environment where RIR plausibility results were unexpected, with the dry control condition deemed the most plausible and the more faithful reproductions being deemed less plausible, despite being a more accurate reproduction of the environment's acoustics. The fact that no significant difference between bandwidth was

detected in the high reverberance space where there is a greater acoustic effect suggests this finding may not hold true in all environments, and further evaluations will be needed to explore the influence of spectral bandwidth on acoustic plausibility in AAR.

The listening test plausibility results also did not demonstrate any significant differences between the 1st and 3rd-order sine sweep RIRs, suggesting there is a possibility that increasing spatial resolution above 1st-order may have diminishing benefits for plausibility. Similarly, no significant difference was detected between plausibility ratings for the two playback devices, suggesting that, at least in terms of plausible virtual sounds, audio glasses may be a viable playback device for AAR. However, as these implications are based on an absence of significance in these results, further work will be required to confirm or disprove them.

The results also suggest that context may be an important factor in virtual sound plausibility, based on the comparison between listening test plausibility ratings and game plausiblity ratings. While this is not a perfectly fair comparison – game ratings were given retroactively and with different material than the MUSHRA, and without a comparison – the large difference may indicate that usage context influences plausibility. As an informal comparison, the results shown here are unable to conclusively show an effect of usage context on plausibility, and this will need investigated in future work. If true, this would be an important finding as the bulk of existing literature on virtual sound plausibility is conducted in formal listening tests, and so performance in these scenarios may not represent real-world performance when listeners are engaged in alternate tasks, as they likely will be in AAR scenarios in the future. Overall, Study 2 suggests that AAR developers may need less accurate acoustic reproductions to leverage their beneficial effects, which are also more achievable and less computationally demanding to render.

Beyond plausibility, Study 2 also corroborated findings from Study 1, again finding evidence that audio glasses result in a higher localisation error than headphones, something important for AAR developers to bear in mind in the future. Finally, the results did not show any significant influence of echoicity, with no significant differences between the two echoic RIR conditions and their anechoic counterparts. While only an initial indication, this may suggest that existing sound design workflows could be applied to AAR scenarios without hampering user experience or perception.

4.3 Qualitative Interview Results

Post-study interviews were carried out for both studies, inviting participants to discuss their experiences of the two playback devices and the different acoustic conditions.

The interviews were transcribed using Microsoft Word, the transcripts corrected by hand, and then analysed line by line to identify common themes, presented below.

Audio Glasses Are Compelling for AAR: In both interviews, participants were asked which of the two playback devices they would prefer to use in a future AAR application, with guided

museum tours, auditory navigation, and AAR games being provided as example scenarios. In both, participants preferred using the audio glasses, praising their form factor and some being pleasantly surprised by their audio quality. In Study 1, 10 participants preferred glasses compared to 7 for headphones. In Study 2 13 preferred glasses, and 8 preferred headphones. Participants often indicated their preference was contextual (3 in Study 1 and 9 in Study 2), particularly based on background noise, noting a degraded experience using audio glasses in the noisier high reverberance space in Study 2.

> *"[I was] very surprised by the glasses, you know, how well they actually worked and how well it resonated." (P19, Study 1)*
>
> *"Because sometimes there were quite a few people walking around, it was easier to ignore with the headphones a bit because you were a bit more separated. But then still you would hear something. But I like the fact that my ears were not covered by the glasses as well, so it's like... I guess it really very much depends on the context of where. So here it depended on how loud the people around me were basically." (P15, Study 2)*

Acoustic Transparency Can Improve AAR Experience: Participants who preferred the headphones often cited superior sound quality, ease of localisation, and their familiarity as being important factors. Many participants also indicated that the glasses helped them feel more 'connected' to their real-world auditory surroundings (3 in Study 1 and 10 in Study 2), or that headphones isolated them from their surroundings (5 in Study 1 and 9 in Study 2). This acoustic transparency was noted as being a positive by 11 participants and a negative aspect by 5 participants. While acoustic transparency is core to facilitating AAR, this was not mentioned directly to participants to avoid biasing them.

> *"I would have thought that before the study, the headphones would have immersed me deeper, but it was the glasses...the glasses sort of put me into the game, whereas the headphones sort of took me out." (P8, Study 2)*
>
> *"I think [I'd use] the head[phones], probably overall because it's 90 percent of the way there in all conditions, whereas the glasses seems like it's awful apart from only in the most echoey thing, then it's amazing" (P1, Study 2)*

Acoustic Cues Were Minimally Different, But Impactful: Participants noted that the inclusion of reverberation improved the plausibility of virtual sounds, and that it affected their ability to localise sounds in both studies, although claims here were inconclusive as 13 mentioned it affected localisation positively and 9 negatively. When asked about differences between conditions, some mentioned certain conditions being clearly worse or less real than others, but participants often noted that any perceived differences were minimal, with one participant even

noting they did not realise different conditions existed. In Study 1, 3 participants mentioned that audio presented over the reference loudspeaker was of higher "quality" or "acuity" than the virtual sounds presented over headphones or glasses. In Study 2, 3 participants also noted that the inclusion of reverberation improved the 'atmosphere' of the game experience in the high reverberance environment.

> *"I really struggled to tell the difference between them, to be honest. I think I could tell what seemed like a...kinda...stronger reverb, I guess? But the subtleties of it were lost on me." (P2, Study 2)*
>
> *"I would be absolutely shocked if my accuracy wasn't massively higher with zero or with lowest reverberation than with everything else. Reverberation made things increasingly difficult. There was a couple reverberation conditions I guess depending on how they were recorded, where I had absolutely almost no idea, like a vague direction, but I was just spinning in place." (P14, Study 1)*
>
> *"I think [the reverberation] contributed quite a lot because you could tell more...like it felt more real. And you could tell more where something is coming from sometimes, as well." (P9, Study 2)*
>
> *"The first moment I was stood here, I think it was headphones on first, and then I heard something and I looked over and then the reverb kind of made it, like, scatter up the ceiling almost and that was a real, like, 'phwoar!' – it really did enhance the atmosphere of the game." (P11, Study 2)*

4.4 Studies 1 and 2 – Discussion and Conclusions

Overall, the results from Studies 1 and 2 provide some useful insights regarding acoustic linking in audio AR. As discussed at the beginning of this chapter, there are three unanswered questions of particular relevance to acoustic reproductions in AAR: how accurate or detailed an acoustic reproduction needs to be to facilitate an acoustic link, how these acoustic reproductions differ perceptually in real-world scenarios compared to lab environments, and how they differ perceptually over lower-fidelity playback devices which better represent consumer end devices. Studies 1 and 2 provide insights on all three aspects.

In terms of RIR accuracy, both studies showed relatively small differences between RIRs, though demonstrate that increasing the spatial resolution or spectral bandwidth of an RIR broadly improves plausibility. Both studies showed that higher resolution Ambisonic RIRs were rated as more plausible than simpler RIR conditions, however neither study found a significant difference between 1st- and 3rd-order RIRs' plausibility ratings. While further work will be needed for confirmation, this provides an initial indication that, at least in the conditions tested in these studies, there may be little need to go above a 1st-order RIR for plausible playback. Study 2

did demonstrate evidence that spectral bandwidth may have an important impact on plausibility, with a significant difference between the full- and limited-bandwidth 3rd-order RIRs in the outdoor environment. As this only applied in one of the Study 2 test spaces, and was not found in Study 1, it remains to be seen how important a concern this is for plausibility. Likewise, no significant differences between 1st- and 3rd-order RIRs were found for localisation error in either study, though both studies demonstrated that deploying acoustic links can also increase localisation error, suggesting that in localisation-focused tasks, AAR developers may need to balance localisation accuracy with virtual sound plausibility. It is important to note that audio was only presented in the azimuthal plane, and these findings may not hold true when audio is presented with varied elevation.

Study 2 explored acoustic linking in real-world use-cases, and as well as providing the plausibility results outlined above, also returned two important insights on real AAR usage. Firstly, the results from Study 2 did not show any significant differences between the echoic conditions and their anechoic counterparts. While significantly echoic audio may have an adverse effect on plausibility, this suggests that AAR developers can deploy similarly echoic sounds without adversely affecting virtual sound plausibility, be they from field or foley recordings, or existing library sounds, aligning with existing sound design practices. Secondly, Study 2 suggests that application context may significantly influence plausibility, finding in-game sounds to be deemed more plausible than those in the listening test. While these game ratings were given retrospectively and for different sound content, this provides evidence that developers can consider simpler, more computationally efficient reproductions, and that findings from critical listening tests may not represent real-world performance.

As one of the first perceptual explorations of audio glasses, the results from these studies provide some interesting findings on their user experience. Firstly, neither study found a notable influence of playback device on plausibility. Study 2 suggested that playback device influenced plausibility in an application scenario, although no significant pairwise comparison was found and so this influence cannot be analysed in greater depth. Both studies' findings also suggest that glasses are perceptually worse than headphones, with a higher localisation error in both studies, and lower levels of externalisation in Study 1. That said, the interview data from both studies suggest that glasses are a promising platform for AAR and one that users are interested in.

4.4.1 Limitations and Future Work

It is important to note a few limitations of these studies, as well as avenues for future work. The first is that as only one set of headphones and audio glasses were tested, these results could be specific to these models. There is also a large price difference between the two (with the HD650 headphones retailing for two to three times more than the FAUNA glasses), which could further exaggerate the perceptual differences found in these studies.

Notably, the Stereo-HC condition performed poorly in both studies. The Stereo-HC con-

dition had been expected to perform partway between the Omni-HC and 3O-HC conditions to reflect its slightly higher spatial resolution, but it actually performed closer to the dry conditions in both studies. As the Stereo-HC condition was rendered by taking signals from opposing capsules on the Zylia microphone, it is possible that this method is not representing a stereo microphone in the intended manner. Future work could explore how other stereo RIRs influence plausibility, as they may still prove a midpoint between the accessibility of an omni RIR and the perceptual benefits of Ambisonics.

Another limitation to acknowledge is that both studies featured egocentric sounds from a static listening position. In an exocentric scenario where users can freely move around a space, the greater spatial accuracy of higher order RIRs may be more influential, and this could be another interesting avenue for future work.

4.4.2 Conclusions

Overall, these studies show that creating an acoustic link provides a tangible step towards the sensation of a virtual sound being located plausibly in a user's surroundings, key for AAR applications. This can be done in a simpler way that reduces processing power and battery consumption, and can be presented over both novel and traditional hardware, be that for audio augmented reality, extended reality experiences, games, or beyond.

As an initial exploration, further work is required to corroborate these findings and explore other AAR contexts, however based on the results of Studies 1 and 2, the following initial recommendations are made for deploying acoustic links in AAR, providing insights on RQ1:

- Full-bandwidth reproductions should be deployed if possible.
- A reproduction with spatial resolution greater than or equivalent to 1st-order Ambisonics should be deployed if possible.
- While anechoic source audio is preferable, source audio featuring mild acoustic cues can also be deployed.
- If localisation is a central part of the AAR task and an acoustic link deployed, the application design should expect localisation errors to be increased by 10-15°, particularly if played back over audio glasses.

Chapter 5

Sonic Linking with Action Sounds

Having established guidelines for achieving an **acoustic** link between real and virtual worlds in Chapter 4, this chapter represents the beginning of an exploration of **sonic** links. As discussed in Chapter 3, this thesis defines a sonic link as when the sounds present in one world influence the other through the AR system, and one form this could take is the use of human-produced action sounds to control or drive an AAR system, as set out in RQ2, which this chapter focuses on.

> **RQ2: How can a sonic link between human-produced sounds and virtual elements be created in audio augmented reality?**

Chapter 2 explored how sonic links could be separated into those originating from environmental sounds in a user's surroundings, and action sounds produced by the user themselves. These action sounds have potential as control inputs for computing systems, with voice input already being widely used in smart speakers and smart device voice assistants. Chapter 2 also explored the concept of 'sonic gestures' – non-speech sounds created by a human such as snapping, clapping, or humming – which also have potential for use in sound-based control schemes.

Neither voice nor sonic input have been evaluated in AAR scenarios, nor has sonic linking (either from environmental or action sounds). Chapter 2 highlighted the importance of an AAR system being able to interpret a user's aural surroundings through a sonic link to be able to augment them. While environmental sounds can include almost any sound, humans can only produce a more limited set. By constraining sonic linking to only action sounds initially, the first steps towards exploring sonic linking can be taken.

To evaluate sonic linking with action sounds, three AAR game scenarios were designed and used as a platform to evaluate three different sonic controls. Games were chosen as test applications as they are both a common use-case for AR, and highly interactive, making them a good candidate for assessing control schemes. A scoping review was carried out of existing AAR games, discussed in Section 5.1, to assess the game scenarios and control schemes currently used in AAR games. The three most common game scenarios, and the three most common control

schemes were identified and formed the basis of the user study's test scenarios. The study then compared these control schemes with speech input, musical input, and sonic gesture input.

5.1 Scoping Review

To evaluate sonic controls in representative AAR games, it was necessary to understand what AAR games already exist, and how they are controlled. A scoping review was carried out to identify the game scenarios and control schemes featured in existing AAR games, allowing for the development of representative AAR games to be assessed in the user study. The review sought both academic and commercial examples of AAR games, with audio-based games considered as examples of AAR if they self-identified as an AAR game or otherwise met the definition of AAR set out in Chapter 3.

5.1.1 Academic Search

The search for academic examples was conducted by searching the Google Scholar, Scopus, and ACM Digital Library databases for relevant papers. Papers were initially identified based on an abstract review, considering the following questions:

- Does the abstract mention presentation of a game? Does it present a system which might be argued to be a game – is the main goal of the system entertainment, play, or fun?
- Does the game or system use audio as its primary output modality?

The databases were searched using the following queries:

- "audio" AND "augmented reality" AND "game"
- "audio augmented reality" AND "game"
- "auditory augmented reality" AND "game"
- "sound-only" AND "game"
- "audio game"
- "auditory game"
- "audio-only game"
- "augmented reality" AND "audio game"
- "augmented reality" AND "auditory game"

As many of these queries returned tens of thousands of results or more, results for a particular search query were no longer reviewed if 50 results in a row were not identified as being relevant. While this introduces a potential selection bias based on search engine ranking algorithms, this threshold still resulted in a thorough literature search. A total of 107 papers from academic literature were identified as relevant, 20 of which were ultimately deemed examples of AAR games for further review.

5.1.2 Commercial Search

To identify commercial AAR games, the iOS App Store, Google Play Store, and the Google and Bing search engines were searched. Mobile storefronts were included to cover mainstream examples of AAR games, with the Web search designed to cover less popular examples. For each, the following queries were used:

- "audio augmented reality game"
- "audio game"
- "augmented reality audio game"
- "audio only game"
- "audio ar game"

Commercial games were judged by similar metrics to the academic search. A result was included for further analysis if it presented a game or game-like system which used audio as its primary output modality, and either self-identified as an AR game or met this thesis's definition of AAR. If a commercial example was mentioned as part of a result, such as a web article, it was searched for directly. Similarly to the academic search, search queries were no longer reviewed once they returned 50 results in a row which were not identified as relevant.

5.1.3 Review Findings

A total of 27 AAR games were found, which were analysed for the control methods they used and the gameplay scenario(s) they featured. The identified games are shown in Table 5.1.3.

Of the 27 AAR games identified:

- The most popular controls were physical movement (16 games), gesture controls (9 games), and a controller or other device (6 games).
- Two games used a sonic control – Rovithis *et al.*proposed a citizen science game where players recorded bird calls around a city [154], though this game was only conceptual. Dimensions [152], the only real example of an AAR game using sound as a control, had

AAR Game / Paper	Control Scheme(s)	Game Scenario(s)	Source
Audio Legends [153]	Physical movement, Gesture	Fight virtual entity, avoid something, progress story	Academic
Bumblebee Jam [50]	Gesture	Make music	Commercial
The Clairvoyant [190]	Gesture	Progress story	Commercial
Collins *et al.*[41]	Physical movement	Solve puzzle	Academic
Dead Drop Desperado [142]	Gesture	Avoid something	Commercial
Dimensions [152]	Real world audio input	Unspecified	Commercial
Eidola [128]	Physical movement	Fight virtual entity	Academic
Fakhour *et al.*[58]	Control device, physical movement	Find something	Academic
Falkland Ghost Hunt [139]	Physical movement, Control device	Navigate virtual world, progress story	Academic
Gampe [64]	Physical movement, Gesture	Progress story	Academic
GenVirtual [43]	Gesture	Abstract - rehabilitation	Academic
Ghast Blasters [83]	Physical movement, control device	Fight virtual entity	Commercial
Guided by Voices [112]	Physical movement	Progress story, fight virtual entity	Academic
Indans *et al.*[88]	Physical movement	Progress story	Academic
Kiriu *et al.*[102]	Physical movement	Gamification - running	Academic
OnTheRun [48]	Physical movement	Progress story, navigate virtual world	Academic
Overherd [196]	Gesture (head)	Fight virtual entity	Commercial
PairPlay [178]	Gesture, Control device	Progress story	Commercial
Please Confirm you are not a Robot [10]	Gesture	Progress story, find something	Academic
Podkosova *et al.*[143]	Physical movement	Find something	Academic
Rovithis *et al.*(concept) [154]	Real world audio input	Abstract	Academic
Siriaraya *et al.*[169]	Physical movement	Gamification - running	Academic
The Songs of North [51]	Physical movement, control device	Find something	Academic
Sonic-Badminton [101]	Physical movement	Abstract/Competitive	Academic
SoundPacman [35]	Physical movement	Find something, avoid something	Academic
Stackable Music [36]	Bespoke control	Make music	Academic
Viking Ghost Hunt [140]	Physical movement, Control device	Navigate virtual world, progress story	Academic

Table 5.1: Table covering all identified AAR games in the Study 3 systematic review.

no clear gameplay scenario or goal and so may be more accurately described as an AAR experience rather than a true AAR game.

- The most common game scenarios were progressing through a story (10 games), fighting virtual entities (5 games), and searching for something (4 games).

5.2 Study 3 – Action Sounds as Input Controls for AAR Applications

As the scoping review showed, the vast majority of AAR games do not use sonic controls, relying instead on physical movement, gesture, or control devices like smartphones for user input. Two games did feature sonic controls, however these were based on environmental sounds rather than action sounds, and were either conceptual or debatable examples of AAR games. In Serafin *et al.*'s framework [166], action sounds were not categorised any further. However, there are multiple distinct categories of sound which humans can produce: speech, musical sound, and non-musical sound, all of which could form user input. As discussed in Chapter 2, speech, music, and sonic gesture (non-musical sound) have not been evaluated for use in AAR before, nor are they in use in any of the AAR games identified through the scoping review. Study 3 was designed to compare these three categories of sound inputs with existing 'traditional' AAR control schemes identified through the scoping review.

5.2.1 Experimental Parameters

Study 3 featured two independent variables: game scenario, and control scheme. Three game scenarios were used, based on the most common gameplay scenarios identified in the scoping review: combat, advancing through a story, and searching for a hidden object. The three sonic control schemes (musical input, speech input, and sonic gesture) were compared with the most common traditional control scheme for each gameplay scenario: physical movement, physical gesture, and use of a game controller. This allowed for a valid comparison between novel sonic control schemes and existing AAR game paradigms. Game scenario and control scheme pairings are shown in Table 5.2.

5.2.2 Experimental Design and Methodology

In all three games, the player took on the role of a sonic wizard or 'sonomancer', as in Study 2, tasked with protecting the real world from the threat of sound-based monsters, providing a cohesive overall narrative that was relevant to the audio-centric applications. In each game scenario, participants evaluated the three sonic control schemes, and the traditional control scheme

Figure 5.1: Test spaces used for the games in Study 3 – a small patio used for the movement condition (L) and a small office space used for the other conditions (R).

for that scenario, resulting in four evaluations per game scenario and twelve evaluation scenarios overall. A within-subjects design was used, where game scenarios were evaluated in a counterbalanced order using a Latin Square design. Within each scenario, participants evaluated the four control schemes in a randomised order to minimise order effects. After each evaluation scenario, the participants were asked to complete a questionnaire, and an overall semi-structured interview was conducted after all twelve evaluations were completed.

The study took place in the same office space as Study 1, where 11 of the 12 evaluations took place. The 12th evaluation, covering a physical movement scenario, took place in an outdoor patio adjoining the office space to allow for more physical movement. The test spaces are shown in Fig 5.1. Building on the findings of Chapter 4, a 1st-order sine sweep RIR of the office space was used for virtual sounds presented in that space. A dedicated RIR was unable to be captured for the patio environment, as the Zylia ZM-1 microphone had been returned to colleagues at Imperial College London, however an appropriate 1st-order sine sweep RIR, corresponding to a similar outdoor environment, was deployed for the outdoor evaluation scenario.

The study was conducted using a Meta Quest 3[1] XR headset, operating in passthrough mode throughout the study. The Meta Quest 3 was chosen as it provided a platform capable of tracking head movement and user position for spatialising egocentric and exocentric sounds accurately, acoustically transparent audio presentation to allow virtual audio elements to coexist alongside real ones, and controllers capable of traditional controller input and gestures. Virtual sounds were spatialised using the 3DTuneIn toolkit [44], as in Studies 1 and 2, and varied only in their azimuthal position.

In the combat scenario, participants had to listen for the sound of a virtual monster which appeared in a random egocentric position around them. They turned on the spot until they felt they had found the monster, then input a command to 'banish' it. If the monster was successfully found (localised within its 30°angular size, as in Study 2), it was banished, causing a success sound to play, and another monster spawned in a new position. If not, the player was "attacked"

[1] `https://www.meta.com/gb/quest/quest-3/`

by the monster, causing a failure sound to play, and the monster moved to a new position, providing a simple motivation to localise monsters successfully.

In the story scenario, participants had to listen for the sound of a virtual ghost, localising it in the same way as the monster in the combat game. If they localised the ghost successfully, a voice line was played providing story information about the sonic monsters and their origin, and a new ghost spawned.

The search scenario featured a similar setup with a magical "ward"[2] that needed to be located and activated, however, unlike the other two scenarios the ward did not constantly advertise its position. Instead, the player had two commands, one which played a $\sim$5 second audio file to reveal the ward's position, and one to reactivate it once they had found it. Auditory feedback informed the player if they were successful or not, and they repeated this process of revealing and activating wards for the course of the condition. The audio files used in this study are available via Appendix A.2.

In each game scenario, participants played until they had completed at least five of these localisation trials, and had played for at least two minutes. This ensured that all participants had the same minimum baseline in each game, both in the number of game loops completed and time spent in the game overall.

When using the novel sonic control schemes, participants said specific words or phrases aloud for the speech conditions, snapped fingers or hummed for the sonic gesture conditions, and played single notes or four-note sequences on a glockenspiel for the musical input conditions. The glockenspiel, shown in Figure 5.2 had colour-coded keys to make note identification easier. The full list of evaluation scenarios and their controls are shown in 5.2. Shorter and longer inputs were chosen in this way to conduct a more robust evaluation of the sonic controls. When using traditional controls, participants held a trigger of the Quest 3 controller, made a sword swing gesture with the Quest 3 controller, or physically walked around the play space. In each of the 12 evaluation scenarios, the current control was demonstrated personally to the participant, and a written tutorial screen was presented with text-based instructions and a button to proceed when the participant understood the upcoming task. This provided space for the participant to familiarise themselves with the control schemes.

Speech recognition, pitch detection, and sound classification systems were integrated into the Unity application for the study, however pilot testing showed that these detection systems had differing levels of reliability from one another, and from the traditional control schemes. A Wizard of Oz methodology was deployed instead to remove this potential confounding effect. Unbeknownst to the participant, a button was pressed on a second Quest controller whenever the participant successfully input the current command, registering the input for the game. While this methodology introduces a level of variability not normally present in the traditional controls, it created a similar level of reliability and input variability for all controls, and produced more

[2] A term for a protective magical spell often seen in fantasy fiction.

Figure 5.2: The glockenspiel used as part of the musical conditions in Study 3.

reliable results for the novel sonic inputs, and the best baseline for comparison. Participants were not informed of the Wizard of Oz methodology until the conclusion of the study, and although not asked directly, no participants suggested they had detected this methodology was in use.

After each evaluation scenario, participants were presented with a short questionnaire about the experience. The questionnaire consisted of the NASA TLX [75], and the Meaning, Immersion, Challenge, and Ease of Control subscales of the Player Experience Inventory (PXI) [1], as well as its Enjoyment questions. The PXI was chosen as a validated instrument focusing specifically on game experience and measures the control scheme's effects on the user's broader game experience. While the PXI is validated as a broader instrument than the subscales used in this study, these were not included in the experimental questionnaire to keep the questionnaire length manageable, and because many of the excluded subscales (e.g. Audiovisual Appeal, Progress Feedback) bear less relevance to assessing control schemes. While the PXI's Ease of Control subscale provides a broad evaluation of the functionality of the control schemes, the NASA TLX task load measurement was also included to explore this further. The NASA TLX is a validated instrument which provides information on the physical, mental, and temporal demands of a test scenario, as well as a user's overall performance, required effort, and frustration with the task, providing deeper insights into the overall functionality of the control schemes being tested. Performance data, such as the number of successful trials, were also recorded for analysis. The full questionnaire used is available in Appendix B.3.2.

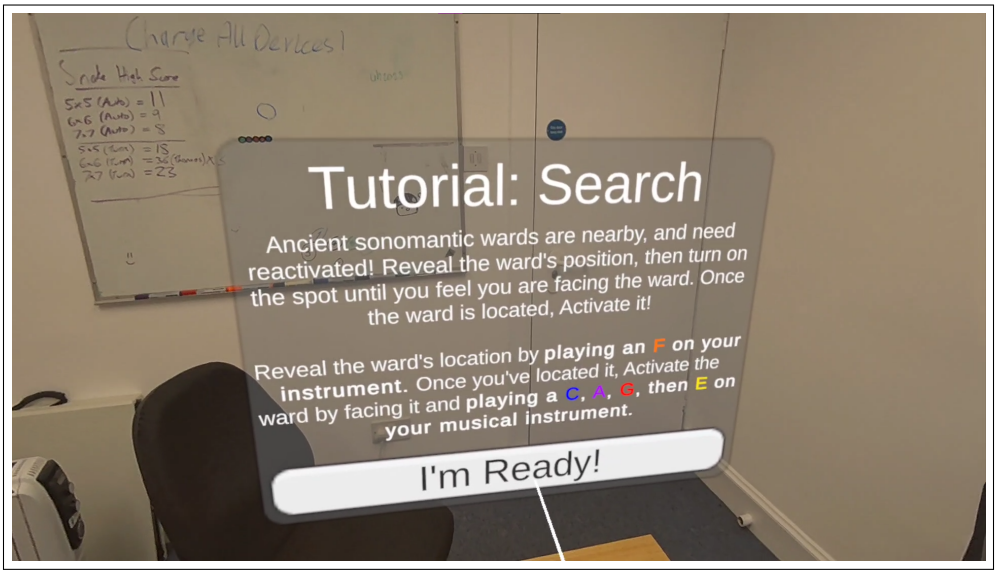

Figure 5.3: Screenshot of the test application used in Study 3, showing an example of the tutorial briefing in passthrough augmented reality.

Game Scenario	Premise / Task	Control Scheme	Control Action
Combat	Destroy sonic monster	Traditional (Gesture) Music Sonic Gesture Speech	Swing sword Single C note Snap fingers "Alakazam!"
Search	Find and reactivate protective ward	Traditional (Movement) Music Sonic Gesture Speech	Stand still to reveal ward. Move to ward and walk in circle to reactivate. Single F note to reveal. CAGE sequence to reactivate Snap fingers to reveal. Hum to reactivate. "Reveal!" to reveal. "I awaken this ward!" to reactivate.
Story	Summon ancient spirit for guidance	Traditional (Controller) Music Sonic Gesture Speech	Hold trigger DEAD sequence Hum "Ancient spirit, I call you forth!"

Table 5.2: Control schemes for each game scenario in Study 3

5.2.3 Participants

24 people participated in Study 3, recruited from posters and university mailing lists and compensated with a £10 Amazon voucher for their time. As with Studies 1 and 2, the only recruitment criteria were that participants did not have any hearing impairments. Final participant demographics were:

- **Gender:** 15 men, 8 women, 1 non-binary person
- **Age:** 9 aged 18-24, 7 aged 25-34, 7 aged 35-44, 1 aged 45-54.
- **AAR Familiarity:** 1 very unfamiliar, 5 unfamiliar, 6 neutral, 10 familiar, 2 very familiar
- **Audio Game Familiarity:** 1 very unfamiliar, 6 unfamiliar, 9 neutral, 7 familiar, 1 very familiar
- **Video Game Familiarity:** 6 unfamiliar, 7 neutral, 7 familiar, 10 very familiar

5.3 Results

Quantitative data were analysed as a single-factor analysis of control scheme within each game scenario. Game scenarios were not directly compared as they were too different to make valid comparisons between, and differences between game scenario were not relevant to RQ2. Friedman tests were used as the PXI and TLX data were ordinal, with the Nemenyi two-sided *post hoc* test used when significant effects were found. Performance data were tested for normality using a Shapiro-Wilk test, with the Aligned Rank Transform method used for analysis in non-normal instances, and a normal ANOVA test deployed when data were normally distributed. Full numeric results are shown in 5.3 and 5.4, and the full dataset is available via Appendix A.1.

Participant completion times for the game portions of the study varied between 53 and 88 minutes (including tutorials and scenario explanations, time spent completing questionnaires, etc.), with a mean completion time of 61 minutes and 34 seconds (SD 7 minutes 56 seconds). The combat scenario took a mean time of 2 minutes 15 seconds, the search scenario a mean time of 2 minutes 59 seconds, and the story scenario a mean time of 3 minutes 23 seconds.

The majority of measures showed no significant difference between control schemes, with only four of the PXI and TLX items showing significant differences between control schemes in at least one game scenario. In the Story scenario, sonic gesture was found to be less meaningful than the speech control (p = 0.037, Mean Difference (MD) -0.5), while the music control had significantly lower ease of control than the traditional controller (p = 0.014, MD 0.56) and sonic gesture (p = 0.05, MD 0.5) control schemes. In the Story scenario, sonic gesture was also rated as requiring more effort than the speech control (p = 0.011, MD 15.5). Significant differences in physical demand were found in all game scenarios, with the speech control being

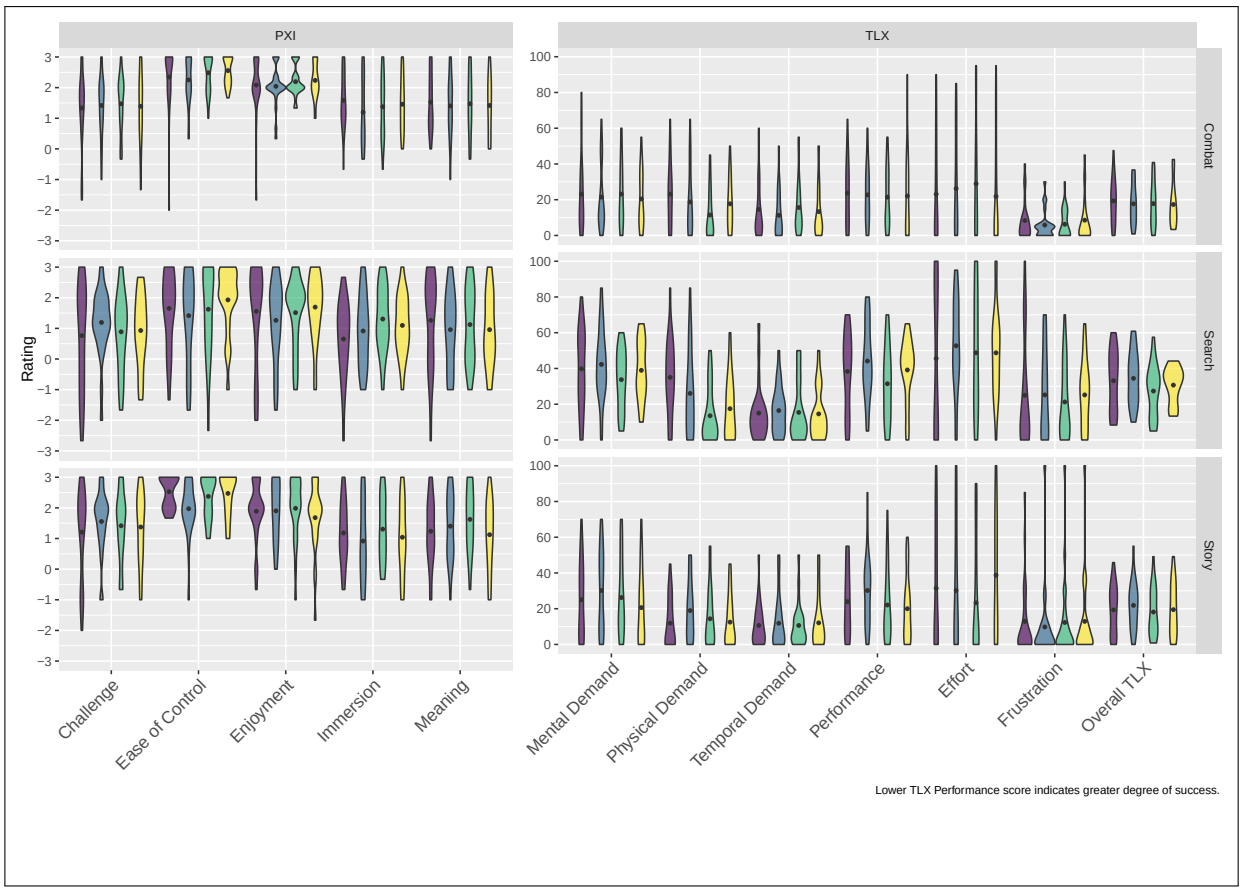

Figure 5.4: Full quantitative results for the PXI and TLX questionnaires in Study 3 across control schemes and game scenarios.

rated as less physically demanding than gesture in the combat scenario (p = 0.023, MD -11.6) and movement in the search scenario (p = 0.003, MD -21,5). Sonic gesture was also rated less physically demanding than movement (p = 0.02, MD -17.5), and the controller was rated less physically demanding than the music control (p = 0.023, MD -7.1). The controller also had a lower overall workload than the music control (p = .011, MD -2.6) in the story scenario.

5.4 Interview Results

Interview results were analysed using inductive, data-driven thematic analysis. Interview results were transcribed using Microsoft Word, then corrected by hand. Transcripts were analysed line by line to develop a coding scheme and overall themes, presented below.

Interest in Sonic Controls: Participants expressed clearer preferences in interviews, with each of the sonic inputs receiving positive sentiment from twelve of the participants, while movement, gesture, and controller saw nine, eight, and two participants expressing positive sentiment respectively. When asked which of the control schemes would be their preferred scheme, eight participants chose sonic gesture, seven chose physical gesture, and controller,

Measure	Scenario	Friedman p (DoF, χ^2)	Sig. Pairwise Comparisons	p	Mean Diff.
PXI - Meaning	Combat	.617 (3, 1.790)			
	Search	.306 (3, 3.617)			
	Story	**.011 (3, 11.128)**	Son. Ges. - Speech	**.037**	-0.5
PXI - Immersion	Combat	.109 (3, 6.048)			
	Search	.396 (3, 2.970)			
	Story	.695 (3, 1.446)			
PXI - Challenge	Combat	.695 (3, 1.444)			
	Search	.66 (3, 1.597)			
	Story	.954 (3, 0.333)			
PXI - Ease of Control	Combat	.168 (3, 5.052)			
	Search	.412 (3, 2.872)			
	Story	**< .001 (3, 16.417)**	Controller - Music	**.014**	0.56
			Son. Ges. - Music	**.05**	0.5
TLX - Mental Demand	Combat	.883 (3, 0.658)			
	Search	.660 (3, 1.596)			
	Story	.055 (3, 7.600)			
TLX - Physical Demand	Combat	**.014 (3, 10.612)**	Speech - Gesture	**.023**	-11.6
	Search	**< .001 (3, 16.827)**	Speech - Movement	**.003**	-21.5
			Son. Ges. - Movement	**.02**	-17.5
	Story	**.008 (3, 11.576)**	Controller - Music	**.023**	-7.1
TLX - Temporal Demand	Combat	.476 (3, 2.494)			
	Search	.672 (3, 1.544)			
	Story	.961 (3, 0.293)			
TLX - Performance	Combat	.73 (3, 1.297)			
	Search	.172 (3, 5.004)			
	Story	.07 (3, 7.068)			
TLX - Effort	Combat	.289 (3, 3.758)			
	Search	.761 (3, 1.167)			
	Story	**.003 (3, 14.042)**	Son. Ges. - Speech	**.011**	15.5
TLX - Frustration	Combat	.509 (3, 2.318)			
	Search	.933 (3, 0.435)			
	Story	.437 (3, 2.717)			
TLX - Overall (Mean)	Combat	.723 (3, 1.327)			
	Search	.463 (3, 0.463)			
	Story	**.018 (3, 10.030)**	Controller - Music	**.011**	-2.6
Enjoyment	Combat	.603 (3, 1.855)			
	Search	.278 (3, 3.845)			
	Story	.132 (3, 5.613)			
Measure	**Scenario**	**ANOVA p, F**	**Sig. Pairwise Comparisons**	p	**Mean Diff.**
Successful Trials	Combat	.836, F (3,92) = 0.285			
	Search*	.725, F (3,92) = 0.440			
	Story*	.358, F (3,92) = 1.089			

Table 5.3: Quantitative analysis results for PXI, TLX, and performance measures in Study 3. Green denotes statistical significance. * denotes non-normal data analysed using Aligned Rank Transform. PXI measures were rated on a seven point scale, TLX from 0-100.

music, and speech were each preferred by five participants. Two participants chose physical movement.

> *"I really, really like the sort of, snapping the fingers and the hum as responses [...] I would probably focus more on the snapping and the hum. Because they're the simplest because they're easiest. So they're sort of more applicable. You can just sort of pick up the game and then you're kind of ready to go".* – P11

> *"In all games, activating by speech, I can imagine if this was part of a longer, sort of, narrative game that that would be a really satisfying, a really good mechanic".* – P21

> *"I thought the integration of the instrument was really fun. [...] you get to play an instrument and that guides you through this quest, I'd be like 'hell yeah!', you know?".* – P20

Sonic Gestures Show Promise, With Some Concerns: While participants were overall very positive about the use of sonic gestures, three participants were concerned about snapping becoming uncomfortable over a longer game session, and three participants said they would not want to use sonic gestures if playing in a public space, particularly the humming. Two participants also disliked sonic gesture overall, with one finding the humming made them feel self-conscious, and the other feeling that in a private gameplay experience they wouldn't want to produce any sounds. Three participants also noted that the finger snapping gave them a sense of power or a feeling of being like a wizard, particularly in the combat scenario.

> *"The snapping of fingers, I think if I was playing longer, the fingers might start hurting and it might be harder to snap and such. So for a shorter time it was ok, but I feel like it would become more burdensome over time".* – P12

> *"I feel like it's a really cool, like I'm actually a wizard because like 'snap' and then I got it correct".* – P14

Speech Controls Can Lead To Self-Consciousness: By contrast, 14 participants stated they would not use speech in a public setting, citing a concern around embarrassment or social unacceptability, and three participants stated they would not use speech controls when playing alone, feeling that they would prefer to use speech in a group scenario, or that they would feel as if they were talking to themselves if playing alone. Five participants expressed an overall negative sentiment towards the speech control, disliking it overall or feeling it was silly.

> *"I don't want to [expletive] shout 'alakazam!' in front of random people".* – P10

> *"I think saying things out loud, that was the silliest one for me. Cause just, yeah, just saying a phrase and like, you either try and make it pompous, but then it's kind of like 'am I making it even more silly?' or you're just trying to say it just with a*

normal voice and at that point it's more of a like [...] [said in monotone] 'open sesame' [end of monotone], I mean no, just, it doesn't sound like I'm having fun, in that sense". – P12

Music Controls Are Polarising: The music control scheme was the most polarising of the three sonic control schemes. While 12 participants were positive on it overall or mentioned it being novel, 11 were negative. Criticisms ranged from it increasing difficulty (seven participants), lowering immersion (three participants), or being unwieldy or undesirable to carry (three participants). However, three participants also cited the physical prop of the glockenspiel as one of the reasons they liked the music scheme. Four participants stated they would not use the music control in a public or group scenario, again feeling it would be socially unacceptable.

"But the xylophone was...made me feel very, just took me out of the game completely because I had to, you know, hold the big xylophone, or maybe not big, but I had to hold the xylophone and I had to remember the notes". – P17

"I like the practical element of the xylophone. I like the note combinations and the fact that you know, I think sort of extrapolating out a little bit, if I was playing a game or if I was doing a task which required a little bit more thinking. In the physical space, holding the xylophone in your hands is obviously a physical connection, as opposed to the largely virtual connection around you." – P8

There Is No 'Best' Control: Of the traditional controls tested, participants were most enthusiastic about physical gesture, with eight noting positive sentiments, and that the control was fun and easy. Only two participants made specific mention of positive aspects of the controller scheme, noting that it felt more accurate. Four participants felt the controller lowered their immersion in the games. Nine participants expressed positive sentiment about the movement control, enjoying being able to move around a space rather than standing still, and six participants expressed desire for a mixture of the different controls.

"Also, the physical gestures with the swipe, because that makes me feel like it's kind of cool as well. Like, I was actually literally imagining, like, slashing the monster's head or something like that". – P23

"I think with the controller, it's obviously not immersive, in the sense that it feels artificial to the world." – P2

"And I don't know about walking around outside, but I think maybe you combine those ones with the moving around and then doing stuff because that was really good, that was really fun, was actually being able to move around the environment kind of looking for things. So I think combining that with some physical thing to hold would I think...that would have been my complete preference, but yeah, I really enjoyed this." – P21

5.5 Discussion

Overall, the results from Study 3 provide a promising first step for sonic linking, suggesting that the use of action sounds as control inputs for AAR applications has promise, at least in game scenarios. The quantitative data suggest that, in terms of overall player experience, the differences between sonic controls and traditional controls are minimal, with the majority of measures showing no significance between controls. At least in the scenarios assessed in this study, this suggests that sonic controls do not perform notably better or worse than traditional controls. Given the traditional controls were chosen based on a scoping review, representing established control schemes already seeing use in AAR games, these minimal differences suggest that action sound-based sonic controls are potentially viable for use in AAR games.

Although the differences between control schemes were minimal, some notable differences were found. Both the quantitative and qualitative results suggest that the musical control can be difficult or unwieldy, with it being rated as more physically demanding than the traditional controller, and rated less easy to control than the physical controller and sonic gesture schemes. The musical control was also the only control not found to be significantly different in physical demand from the outdoor movement scheme, and given the musical control has a much smaller physical component than the movement scheme, this further suggests a level of unwieldiness. As only one musical instrument was tested in this study, it is possible that this could be a function of the glockenspiel rather than musical input overall. In interviews, participants also mentioned the musical control as being harder to use than the other schemes. Some participants mentioned the Quest 3's passthrough cameras did not have the fidelity to easily differentiate between the glockenspiel's notes, which may have increased difficulty. As an initial exploration, musical controls should not be discarded based on these results. Musical input was not rated significantly differently in the Challenge, Effort, or Performance subscales, and just as many participants felt positively about it as felt negatively, suggesting it does have potential.

The speech control performed well, although participants expressed concerns over its use. It offered lowered physical demand than gesture and movement, and required participants to spend less effort for a more meaningful experience in the story scenario than with sonic gesture. However, it was not the most preferred control among participants, and there was no clear consensus from the interview data as to when a speech control is best deployed. Both private and public scenarios were cited as less suitable for speech controls by participants, and so deploying speech controls as the only control scheme may lead to a negative experience for certain users. Many participants noted the speech control felt 'silly', and it is possible that this was influenced by the sonomancer theme used in the study. In other application scenarios with more mundane speech controls, it is possible speech may not have the same concerns.

Of the sonic controls tested, the results suggest that sonic gesture is one of the most promising. While the speech control rated higher for Effort and Meaning in the story scenario, sonic gesture was most preferred in interviews and had the least concerns surrounding public usage

of all the sonic controls tested. In interviews, participants did raise some concerns, such as the finger snapping potentially becoming uncomfortable after an extended time, as well as the same concerns around public use that were common to all sonic controls. Although promising, only two sonic gestures were tested in this study – snapping and humming – and other potential sonic gestures, such as clapping or whistling, could mitigate some of these concerns or provide new layers to a sonic gesture control.

Of the traditional controls tested, physical gesture was best received by the participants, and no significant differences were found between physical gesture and sonic gesture, suggesting that sonic gestures could be deployed as an alternative or supplementary gesture to physical actions. The controller performed adequately overall, however participants did not express strong feelings on it either way, beyond some participants noting a less immersive experience. Physical movement was well-received, however many participants seemed most enthusiastic about using it as a secondary control, being able to move through a virtual space and interact using a different control scheme. A number of participants were enthusiastic about using multiple control schemes at once, suggesting that while sonic linking in this way could be used as the sole control scheme for an AAR application, it could also be used as a supplementary way of interacting in AAR.

5.5.1 Limitations and Future Work

While Study 3 provides a broad overview of sonic linking with action sounds, there are some important limitations to acknowledge, as well as opportunities for future work:

The first is the use of a Wizard of Oz methodology. While this was necessary to ensure an even comparison of input methods, it remains to be seen how these sonic controls function with current detection technology. Secondly, information on participants' musical abilities was not gathered as part of the study. While the musical controls were designed to be simple and accessible for all musical abilities, users with greater musical experience may nevertheless have found less friction when using them, and this could possibly contribute to the polarised opinions found in the qualitative results. Further work should explore the potential of these controls in greater depth, particularly in live detection scenarios.

As a broader overview of sonic controls, the sonic inputs deployed in this study were also deliberately simple, and future work could explore more nuanced sonic controls. For example, future musical controls could explore the use of chords, or musical techniques like sustained notes or glissandi. Singing could also be explored as a hybrid between musical and speech controls, where both the lingusitic and musical elements of a user's voice control a system. Intonation could also be explored as part of speech controls. In this study, only the words spoken were considered as an input, but the way a user speaks those words could be incorporated into future speech controls. For instance in the wizard scenario used in this study, intonation is often

an important part of spellcasting in popular culture[3] and incorporating this into the control could potentially improve the experience in the future.

5.6 Conclusion

RQ2 focuses on how a sonic link can be created between sounds produced by humans and virtual AAR elements. Study 3 represents an initial evaluation of this form of sonic linking and shows both that it can be achieved and that it has promising implications for AAR and beyond. The results suggest that overall, these sonically linked controls perform promisingly and can reasonably be deployed in place of, or addition to, traditional AAR inputs. In particular, sonic gesture is identified as a promising control scheme, offering no disadvantages compared to traditional controls and being well-regarded by participants. The results suggest that sonic gestures could be deployed in scenarios where physical gesture would otherwise be used. Speech and musical controls introduce additional concerns around public usage or unwieldiness, however, participants were enthusiastic about the use of sonic controls in AAR scenarios overall, suggesting that they could have a place in AAR application design going forward.

By evaluating novel sonic control schemes for AAR gameplay, this chapter provides a first step in exploring sonic linking in AAR, and AAR applications being sonic in both their input and output, responding intelligently to real-world sounds around the user. The following chapters explore sonic linking in greater depth, focusing on the potential of sonic links between an AAR application and the environmental sounds around a user.

[3] "It's levi-OH-sa, not levi-oh-SA!" [42]

Chapter 6

Sonic Linking with Environmental Sounds

With Chapter 5 taking the first steps towards understanding sonic links, this chapter continues this exploration by investigating sonic links between a virtual AAR application and the real-world environmental sounds around the user, contributing to RQ3.

> **RQ3: How can a sonic link between environmental sounds and virtual elements be created in audio augmented reality?**

Chapter 5 explored how an AAR system can usefully respond to action sounds created by a user, however the much larger category of real-world sound, and the most important to AAR, is environmental sound. As discussed in Chapters 3 and 2, an AAR system cannot truly augment our auditory reality without being connected to or aware of the sounds within it. Without a sonic link like this, the system is limited to merely overlaying virtual audio atop real rather than combining real and virtual into something greater than either. Sonic links originating from environmental sounds are therefore a crucial aspect of future AAR systems.

By integrating sonic links to environmental sounds, AAR applications of all forms can make informed decisions about how to present virtual audio to a user. In the future, an AAR system could delay an auditory notification until a user is in a quiet location and better able to receive it, reposition virtual audio sources so that they do not overlap with existing real sounds, or even enabling novel applications such as AAR audiobooks which unfold differently depending on the environment where a user listens to it.

Chapter 2 established that sonic linking with environmental sounds is almost entirely unexplored in AAR. Only two AAR papers exist that implement a sonic link like this: the Nomadic Radio system [158], which evaluated the level of conversation noise around a user when determining how to surface an auditory notification, and the Acu-Notch system proposed in [110], where the volume of music would adjust in response to important real-world stimuli. The Acu-Notch system was never fully realised or evaluated, and the Nomadic Radio system was only evaluated by one user over two days. The concept of sonic linking has never been fully developed and evaluated. This chapter marks the first exploration of sonic linking as a concept in AAR.

This chapter presents two studies designed to explore sonic linking with environmental sounds. In Study 4, users evaluated three different sonically linked applications in a real-world space, which were responsive to sounds in the environment. These applications consisted of an AAR music player, an AAR game, and an AAR notification system, which responded to birdsong and vehicle sounds. Study 5 built upon the applications and results from Study 4, taking forward the music player and game applications and using live sound detection models to facilitate sonic links.

6.1 Study 4 – Sonic Linking with Environmental Sounds in a Best-Case Scenario

As the first step in exploring sonic linking in AAR, the focus of Study 4 was a broad exploration of sonically linked AAR applications, focusing on evaluating the overall concept of AAR applications with an environmental sonic link. Participants evaluated six different sonically linked AAR scenarios, consisting of two design variations of three different AAR applications which responded to birdsong and vehicle sounds. The study was run in a loosely controlled outdoor park environment, and used a within-subjects design where all participants experienced all six application variations. The study lasted approximately one hour.

6.1.1 Experimental Parameters

Study 4 featured three AAR applications, each with two design variations, with *application* and *variation* forming the independent variables for the study. The AAR applications consisted of an AAR music player, an AAR game, and an AAR notification system, chosen to cover a range of plausible application areas. Listening to music is one of the most common audio-focused tasks users engage in, while games are a popular form of AR application. The notification scenario represents a novel application of sonic linking, building on related prior work like the PULSE system [122] for social media sonification. Variations were specific to each application, designed to explore nuanced aspects of environmental sonic linking and AAR such as the congruence of a real-world sound and its virtual effect, or sonic linking with different sound sources.

Game Application

The game application was inspired by Pokémon Go [135]. Real-world birdsong generated an auditory 'Sonimon' (sonic Pokémon) positioned randomly around the user. The user rotated on the spot until they were facing the Sonimon and captured it by pressing a button on a controller. The user had three attempts to localise the Sonimon within its 30°angular width (carried forward from prior studies as an appropriate difficulty for non-expert listeners) before it fled. At random intervals of 90–120 seconds an auditory monster would appear, again in a random position

around the user, needing to be destroyed through the same localisation process. The monster was destroyed more easily with more captured Sonimon, incentivising Sonimon collection. If the user could not defeat the monster within four attempts, it destroyed one of their Sonimon and fled.

Sonimon were represented with designed sound – a magical shimmer sound that was common to all Sonimon, layered with bird or dog sounds depending on the Sonimon. The same monster sound as Studies 2 and 3 was used.

Application variations assessed whether the congruence of a virtual event with its real trigger affected the user experience, under the hypothesis that a clearer connection between real sound and virtual effect could positively influence user experience, and improve a user's immersion and sensation of augmentation.

- **Variation A:** *Congruent Scenario.* Birdsong resulted in a bird Sonimon. This had a clear and congruent connection between the real sound and virtual effect.
- **Variation B:** *Incongruent Scenario.* Birdsong resulted in a dog Sonimon. A dog was chosen as a common animal, but with no clear connection to birds.

Music Application

The music player was inspired by the proposed Acu-Notch system [110], and adjusted volume and low-/high-pass filters when specific real-world sounds were detected to improve their audibility. Michael Jackson's 'Billie Jean', the most-streamed song on Spotify from one of the best-selling albums of all time [185], was used as the demonstration track.

Application variations were chosen to explore two different real-world triggers and subsequent use-cases:

- **Variation A:** *Relaxation/Wellbeing Scenario.* The system improved audibility of birdsong. When birdsong was detected, a low-pass filter was applied to the music at 2000 Hz, and volume was lowered to 35% over 2s. This both reduced volume and provided space in the frequency spectrum for the higher frequencies of birdsong. Listening to birdsong has been found to benefit wellbeing and a listener's perceptions of urban environments [79, 59, 176], and so this scenario was designed to maximise these benefits alongside normal music listening. When no birds were detected, music returned to normal over 2s.
- **Variation B:** *Safety Scenario.* The system improved audibility of nearby vehicles. When vehicle sounds were detected, a high-pass filter was applied to the music at 4000 Hz, and volume was lowered to 35% over 2s. This similarly improved audibility of the vehicle in terms of volume and the overall frequency spectrum. Headphones, especially those with noise cancellation, can block out traffic sounds for pedestrians. While this can be desirable, it could also pose a safety risk as users may be unaware of vehicles they cannot

see. By improving the audibility of vehicles, user awareness and safety could potentially increase. When no vehicles were detected, music also returned to normal over 2 seconds.

Notification Application

The notification application was designed for Twitter/X. When real birdsong was detected, it was supplemented with virtual birdsong to represent aspects of the user's Twitter feed. The number of virtual crows represented unread direct messages, the number of virtual pigeons represented the level of activity from nearby accounts, and the number of virtual blackbirds represented the level of activity from followed accounts. Participants were presented with a set of slider controls to vary these three characteristics of a hypothetical Twitter feed and freely experiment with the system. While virtual elements in the game scenario were positioned only in the azimuthal plane, virtual birds in the notification scenario were positioned above the listener, mimicking the effect of birds singing in trees overhead.

Application variations investigated user preference for virtual birds to sound realistic (potentially causing confusion as to whether a bird was part of the system), or distinct from real birdsong:

- **Variation A:** *Realistic Birds.* Virtual birds were presented with natural birdsong samples.
- **Variation B:** *Digital Birds.* Ring modulation and a 'shimmer' effect were added to the birdsong samples to make them appear distinct from real-world birds, while remaining recognisable.

6.1.2 Experimental Design and Methodology

To facilitate a best-case scenario, sounds were detected using a Wizard of Oz methodology where the author identified when specific sound triggers occurred. Sounds can be detected and classified by existing machine learning models, however this introduces a potential liability from missed detections and false positives. By circumventing this and using Wizard of Oz detection, the principle of sonic linking could be evaluated without being influenced by the accuracy of the detection method.

The study took place outdoors, in a small park surrounded by a road, near to a car park, shown in Figure 6.1. This space was chosen as a plausible location for both birdsong and vehicle sounds. Birdsong was selected as the main trigger for the applications as it is a common sound across urban and natural outdoor environments, and represented an interesting case study as birdsong is usually *acousmatic*, heard without being seen. This characteristic helped to facilitate the Wizard of Oz methodology. Vehicle sounds were chosen to assess a safety scenario in the music application.

While the test environment contained birdsong and vehicles, additional speakers were hidden in the undergrowth within the park, playing the sound of birdsong and vehicles to ensure the

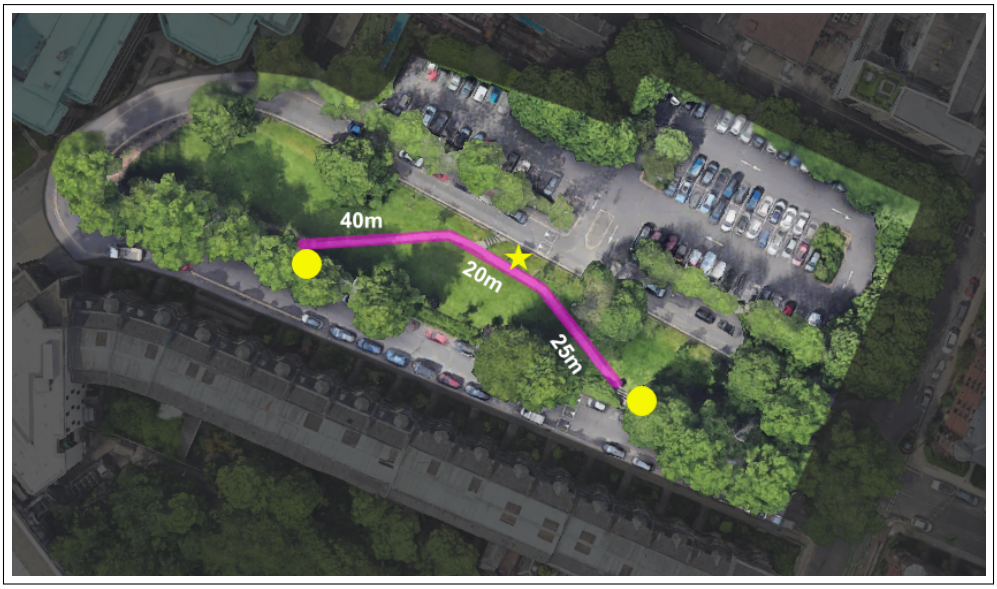

Figure 6.1: Outdoor experimental space used for Studies 4 and 5. Yellow circles indicate positions of birdsong speakers. Yellow star indicates position of car speaker. Pink line indicates walking route for participants.

environment contained a baseline number of the target sounds. This avoided a scenario where there were no birds or cars during an evaluation session, with miminal disruption to ecological validity as the speakers were hidden. Real cars and birds were still considered valid triggers for the application alongside these phantom sounds. Two speakers played native birdsong (one of a common chaffinch and one of a Eurasian blue tit) at intervals, and one played car sounds at intervals. Bird sounds were sourced from xeno-canto.org and car sounds were sourced from freesound.org. Sound files in each speaker were played with a 30 second interval between them, and participants were not informed of the hidden speakers. None of the participants suggested they had noticed them.

Participants evaluated all six of the application variations, using each for five minutes while walking back and forth along an 85 metre path through the park environment. After each evaluation, the participant completed a short questionnaire, containing questions from the validated UEQ-S [164] and Augmented Reality Immersion (ARI) [66] questionnaires, as well as bespoke questions designed to assess the participants' sensation of augmentation. The full questionnaire used is available in Appendix B.4.2. After using both variations of an application, a short interview was conducted to discuss the participant's views on the application and any design changes they would make. After all six evaluations, an overall interview was conducted discussing the applications and sonic linking in AAR generally.

The applications ran on a Quest 3 headset operating in passthrough mode, as in Study 3. This provided a convenient, all-in-one platform which allowed the participant to see their surroundings while providing accurate head and location tracking for virtual sound positioning. Sounds were presented over the Quest 3 headset speakers, allowing both real and virtual sounds

to be audible to the participant at the same time. As in previous studies, the applications were built in Unity and used the 3DTuneIn toolkit for binaural spatialisation alongside the KEMAR HRTF from the SONICOM dataset [56]. Based on the work presented in Chapter 4, a 1st-order RIR corresponding to the test space was used for acoustic simulation in pilot testing, however the computational requirements of this were found to be infeasible for deployment on a Quest 3 when multiple sounds were presented simultaneously, as in the Notification application. As the study was conducted outdoors, where acoustic effects are less significant, acoustic reproductions were omitted in this study.

Music was presented in traditional stereo format, just as in a normal music listening experience, while virtual sounds in the game and notification applications were spatialised binaurally to have specific real-world positions. The sounds were spatialised fully exocentrically, using the Quest 3's native 6DOF tracking to track distance to virtual sound sources. The audio files used in this study are available via Appendix A.2. The applications relied on a Wizard of Oz sound detection method, with the author accompanying the participant with a hidden Quest controller. When a bird or vehicle was heard, the author pressed a button to log the detection.

The post-evaluation questionnaire consisted of the User Experience Questionnaire (short version, UEQ) [164] to assess the applications' overall usability, and the *Engagement* and *Engrossment* subscales of the Augmented Reality Immersion (ARI) questionnaire [66] to assess their immersion in the experience. The *Total Immersion* subscale of the ARI was omitted as pilot testing showed it to be difficult to interpret with regards to the audio-centric applications. Finally, bespoke questions designed to measure participants' sensation of augmentation (how augmented the world/experience feels), as there are no validated instruments for measuring this. Participants were also asked to rate how much they felt the application responded to real sounds around them. These questions were rated on 7-point Likert scales, and are shown in Table 6.1.2

6.1.3 Participants

24 people participated in Study 4, recruited from university mailing lists and compensated with a £10 Amazon voucher for their time. As in Studies 1–3, the only recruitment criterion was that the participants had no hearing impairment. Study 4's participants consisted of:

- **Gender:** 12 men, 11 women, 1 non-binary person.
- **Age:** 8 aged 18-25, 10 aged 25-34, 5 aged 35-44, 1 aged 45-54.
- **AR Familiarity:** 1 very unfamiliar, 4 unfamiliar, 5 neutral, 12 familiar, 2 very familiar.
- **AAR Familiarity:** 5 very unfamiliar, 6 unfamiliar, 7 neutral, 6 familiar.

Measure	Question	Scale
Augmentation	"My world felt augmented."	Strongly Disagree – Strongly Agree
	"The virtual elements felt connected to the real world."	Strongly Disagree – Strongly Agree
	"If I used the application in a different environment, the application would have behaved differently."	Strongly Disagree – Strongly Agree
	"The virtual elements felt like they were part of the same world as the real elements."	Strongly Disagree – Strongly Agree
	"The application was ___ than the real or virtual elements would have been on their own."	Much Worse – Much Better
Real Sound Response	"The application responded to real world sounds around me."	Strongly Disagree – Strongly Agree

Table 6.1: Subjective measures deployed in Study 4 post-round questionnaires.

6.1.4 Results

Quantitative results were analysed as a single factor comparison of application variation within each of the three tested applications. Similarly to Study 3, the overall applications were not directly compared as they were examples of different use cases and not meaningfully comparable. As ordinal quantitative data, the Wilcoxon signed-rank test was used to compare variations. Quantitative results are listed in Table 6.1.4 and shown in Figure 6.2. The full dataset is available via Appendix A.1.

No significant differences were found between variations for the majority of the measures, with the exception of the Augmentation and Real Sound Response questions. The *Congruent* variation of the game resulted in a significantly more augmented experience, and one which felt more responsive to real sounds than the *Incongruent* variation. The car-responsive *Safety* variation of the Music application was also rated as more responsive to real sounds than the bird-responsive *Relaxation* variation.

6.1.5 Interview Results

Interview results were analysed using inductive, data-driven thematic analysis. Interviews were auto-transcribed using WhisperX [7] and reviewed by hand, correcting any transcription errors. As Study 4 placed a larger emphasis on interview data than Studies 1-3, an additional qualitative coder was recruited from the author's research group. Both the author and the second coder independently familiarised themselves with the interview transcripts, then analysed transcripts line by line to develop their own coding schemes. These schemes were then compared, discussed,

Measure		**Variation A (M (SD))**	**Variation B (M (SD))**	**Wilcoxon *p* (W)**	**Effect Size *r***	**A-B Mean Diff.**
UEQ-S	Game	1.92 (0.698)	1.88 (0.755)	.691 (152)	0.094	0.04
	Music	1.36 (0.638)	1.44 (0.766)	.819 (130)	0.056	-0.08
	Notif	0.68 (0.667)	0.57 (0.757)	.563 (158)	0.117*	0.11
ARI *Engagement*	Game	2.09 (0.616)	2.16 (0.552)	.646 (74.5)	0.065	-0.07
	Music	1.86 (0.646)	2.00 (0.564)	.33 (78.5)	0.202*	-0.14
	Notif	0.76 (0.944)	0.63 (0.958	.988 (139)	0.003	0.13
ARI *Engrossment*	Game	1.81 (0.594)	1.82 (0.637)	.888 (91)	0.097	-0.01
	Music	0.72 (0.745)	0.58 (1.04)	.537 (146)	0.137*	0.14
	Notif	0.40 (1.25)	0.40 (1.06)	1 (116)	0.006	0
Augmentation	Game	1.49 (0.667)	1.19 (0.852)	**.020 (168)**	0.460**	0.30
	Music	0.95 (0.557)	1.02 (0.753)	.615 (82)	0.091	-0.07
	Notif	0.89 (0.89)	0.76 (0.82)	.702 (127)	0.076	0.13
Real Sound Response	Game	2.08 (0.717)	1.25 (1.29)	**.007 (95)**	0.566***	0.83
	Music	1.96 (0.806)	2.38 (0.77)	**.028 (19.5)**	0.451**	-0.42
	Notif	1.12 (1.45)	1.17 (1.34)	.863 (56.5)	0.060	-0.05

Table 6.2: Overall results for each measure in Study 4. Green denotes statistical significance. * denotes small effect size, ** denotes moderate effect size, *** denotes large effect size.

Figure 6.2: Study 4 questionnaire responses, separated by application and variation.

and merged to form an overall thematic analysis of the data.

Congruent Sonic Links Are Preferable

Fourteen participants stated a preference for the *Congruent* Game over *Incongruent*. For some, the preference for the *Congruent* variation was based on a wariness or dislike of dogs (two). Seven participants mentioned that the *Congruent* Game felt better connected to reality than the *Incongruent* variation, which often felt 'mismatched' or more artificial by comparison.

> *"I think in the second game, which was the one that had the bird songs as the Sonimons, I think that was really interesting because I could actually feel it kind of reacting to the environment. Like, I could hear the birds chirping in real life, and then I would, you know, then it would appear. So I thought that was really fun". – P4*

> *"I think I significantly preferred the bird one to the dog one because the dog one's like, you could be doing it anywhere and the, sort of, novelty could probably wear off quicker. Whereas with the bird one I really liked how it was connected to the environment". – P24*

> *"The [game] where it was bird sound, making another birds sound, that felt more like it was kind of connecting the real and virtual world in a nicer way and felt more immersed, I suppose." – P15*

Real World Experience

Fifteen participants expressed having a heightened awareness of their real world surroundings while using the applications, with seven also feeling more connected to their surroundings, particularly using the *Wellbeing* Music. Realistic virtual elements were desirable, with 13 participants preferring the natural birdsong in the *Realistic Birds* Notification variation to the *Digital Birds* variation. However, 13 participants also noted that this realism could be problematic, causing confusion in the Notification application as to whether a bird was real or not. Despite participants acknowledging the problem motivating the *Digital Birds* variation in this way, *Realistic Birds* remained preferable.

> *"It made me pay more attention to the bird sounds than ever in my life, I think." – P22*

> *"Particularly, a reason I like the music one is sometimes when you're using digital applications, you feel very disconnected from them. So it was cool to have something that kind of reconnects you to that and responds to the real world. Reminds you to listen to stuff." – P10*

> *"It might be confusing as well is my other point because at the start I wasn't sure if I was hearing real pigeons or not so it could almost become too much noise, like literal noise but also just general, like, noise of information." – P15*

Safety and Utility

Participants were most enthusiastic about the Music application, with 20 expressing interest in using it. Sixteen participants expressed seeing a practical utility and safety benefit, particularly from the *Safety* variation. Twelve participants suggested that it be extended to react to people or speech, and two participants suggested it could have applications for women's safety. One participant disclosed that they were neurodivergent, and expressed interest in adapting the premise to alleviate sound sensitivity issues. While participants were very positive about the app's safety potential, two participants also expressed concern around becoming reliant on such a system.

> *"I really like the [music] scenario with the cars because there's, like, a direct application to that. [...] There were some where it was like 'oh I didn't even realise there was a car.'" – P12*

> *"If someone speaks to me for something, I have to go 'oh wait a minute', and fiddle, faff, and then I've missed what's going on, so if something did that automatically I'd be all over that." – P2*

> *"I like the idea of using it for safety, but one of my concerns would be if we were to have an application where we were using it for safety and people became reliant on it and then it didn't work." – P8*

Interest in Visual Elements

Sixteen participants expressed a desire to incorporate visual elements into future versions of all three applications. Overwhelmingly, participants expressed an interest for this in the Game scenario (16 participants), hoping or expecting to see visuals for the monster or Sonimon when playing. However, some participants also expressed interest in visuals for the Notification (2) and the Music apps (2).

> *"I do quite like the idea where you have to listen for it but like if it was a full game it would be cool to have maybe some visual and audio elements. [...] Maybe they're invisible and you have to hear where they are and then when you look at them then they become visible." – P15*

> *"I feel like I would like a little bit more of a visual component to it, to make it a little bit more engaging. I think it's an interesting concept. I think I could definitely see myself in the future playing something that had that as part of a component, I don't know that I'd play something that had that as the exclusive component." – P4*

Critique of Notification App

Participants were critical of the Notification application, preferring the Music and Game applications. 15/24 participants said they would not use the Notification scenario in its current form, with the most common complaint being that the system felt overwhelming or confusing. Other objections included not using auditory notifications normally (four participants), that they preferred being in control of when they received notifications (four participants), or that they were concerned about blending their digital life with natural soundscapes (three participants). However, some participants did feel the notification scenario was novel (four), or envisaged using it subject to changes (four), or for a different type of data such as reminders or to-do lists (three).

> *"It was just overwhelming. I couldn't really distinguish if it was a message, a nearby activity, or the other option." – P17*
>
> *"So basically, [it] kind of creeps into your offline life in a manner of speaking. [...] You might not want to get notifications from Twitter or X all the time. And this kind of augments Twitter or X into your real world. Now, some people might want that – I personally do not." – P9*
>
> *"A car horn perhaps, maybe it could remind you something like where your keys are maybe – some people are forgetful of their keys – [...] if they can detect a train sound perhaps they can tell you 'have you bought your tickets?'" – P18*

6.1.6 Discussion

Overall, the results of Study 4 provide a promising initial indication that sonic linking has potential for AAR applications. Positive mean UEQ and ARI ratings suggest that participants found the applications usable and immersive overall (though less so for the Notification applications), and participants responded positively to them in interviews. The popularity of the *Safety* Music application also suggests that sonic linking may be viable across multiple use-cases and real-world sounds, and the interest in visual elements from the interview results suggests sonic linking could have potential in visual AR also.

Results from the Game applications also suggest that the type of sonic link deployed can influence user experience. When comparing the *Congruent* and *Incongruent* Game variations, the use of a congruent sonic link resulted in a higher sensation of Augmentation, and a sonic link that felt more responsive to real world sounds (despite being just as responsive in reality as the *Incongruent* variation). This suggests that congruent links should be deployed where possible to provide the most augmented and cohesive experience. The interview findings also suggest that sonic links may improve a user's connection with the real world and their awareness of it – even though an AAR system introduces new elements to our real surroundings, including these appears to elevate rather than dilute real-world experience.

While participants were positive about the Music and Game applications, they were less so around the Notification application. Overall quantitative measures were lower for the Notification application than the other two, and participants commonly reported it could be confusing, often being unsure which birds were virtual and which were real. On one hand, this suggests the applications achieved a 'perfect' augmentation of reality, with the virtual elements being indistinguishable from real ones, something that visual AR applications are still a long way from achieving. However, this also shows that in scenarios where users must be able to differentiate real from virtual, AAR developers must be careful when deploying virtual sounds that can also be found in the real world. Further work is required to explore this problem, as the solution explored in this study (the *Digital Birds* variation) still suffered from this confusion. The participants also raised security concerns surrounding sonic linking, suggesting that developers must be transparent about what data a sonically linked app collects and when.

By exploring a range of sonically linked applications, and doing so in a best-case scenario, this study provided promising initial insights into sonic linking, though only in a broad sense. While the results suggest that sonic linking could offer benefits in a 'perfect' implementation, it is also important to assess sonic linking in more realistic scenarios: evaluating sonically linked applications running on AAR-focused hardware rather than the Quest 3, evaluating them with real sound detection algorithms rather than the Wizard of Oz approach used here, and comparing them with non-sonically-linked applications to explore whether sonic linking represents an improvement to existing approaches, or merely an alternative. This formed the motivation for Study 5.

6.2 Study 5 – Sonic Linking with Environmental Sounds Using Live Classification

Study 4 provided a promising initial exploration of sonic linking, and one motivating further investigation. Study 5 was designed to represent a more realistic sonic linking scenario, implementing a live sound classification model to replace Study 4's Wizard of Oz approach, and moving away from the Quest 3 hardware towards a more representative set of AAR equipment. Additionally, variations of the test applications that did not feature sonic links were evaluated as a control condition, something which was excluded from Study 4 in favour of exploring design variations instead.

Participants evaluated further developed versions of the Game and Music applications from Study 4. The applications relied on the YAMNet sound detection model [84] to detect bird and vehicle sounds, and ran on a laptop and acoustically transparent headset in place of the Quest 3. Additional variations of the Game and Music applications were assessed, where sonic links were not present. The study was conducted in the same park environment as Study 4, and used a within-subjects design, lasting approximately one hour.

6.2.1 Experimental Parameters

Study 5 took a refined approach to that of Study 4, with a largely similar design. The Game and Music applications were carried forward, with an additional 'C' variation introduced which did not feature sonic linking, functioning as a control condition. These control conditions replaced the Notification application from Study 4, which was not carried forward both as the worst performing application and because it did not have a clear unlinked design in the same way that the Game and Music applications did.

The overall application designs of the Game and Music applications were unchanged from Study 4. The design variations used were:

- **Game A:** *Congruent Scenario.* Birdsong resulted in a bird Sonimon. This had a clear and congruent connection between the real sound and virtual effect.
- **Game B:** *Incongruent Scenario.* Birdsong resulted in a dog Sonimon. A dog was chosen as a common animal, but with no clear connection to birds.
- **Game C:** *Unlinked.* After a randomly chosen period of 20-45 seconds, a cat Sonimon was spawned, regardless of environmental sounds, repeated for the course of the game. A cat was chosen as another example of a common animal, but one unlikely to suggest a sonic link. Having no sonic link but otherwise fulfilling the requirements for AAR, this variation represents the existing paradigm for AAR applications. By comparing user experience with the A and B variations, the impact of sonic linking itself on AAR as it has existed so far could be explored.
- **Music A:** *Relaxation/Wellbeing Scenario.* The system improved audibility of birdsong. When birdsong was detected, a low-pass filter was applied to the music at 2000 Hz, and volume was lowered to 35% over 2 seconds. This both reduced volume and provided space in the frequency spectrum for the higher frequencies birdsong usually consists of. Listening to birdsong has been found to benefit wellbeing and a listener's perceptions of urban environments [79, 59, 176], and so this scenario was designed to maximise these benefits alongside normal music listening. When no birds were detected, music returned to normal over 2s.
- **Music B:** *Safety Scenario.* The system improved audibility of nearby vehicles. When vehicle sounds were detected, a high-pass filter was applied to the music at 4000 Hz, and volume was lowered to 35% over 2 seconds. This similarly improved audibility of the vehicle in terms of volume and the overall frequency spectrum. Headphones, especially those with noise cancellation, can block out traffic sounds for pedestrians. While this can be desirable, it could also pose a safety risk as users may be unaware of vehicles they can't see. By improving the audibility of vehicles, user awareness and safety could potentially increase. When no vehicles were detected, music also returned to normal over 2 seconds.

- **Music C:** *Unlinked.* Music was unaltered throughout, representing existing music listening experiences over headphones. As this variation does not meet the definition of AAR, the impact of sonic linking on non-AAR audio applications could be explored by comparing it with the A and B variations.

6.2.2 Experimental Design and Methodology

Study 5 was conducted in the same space as Study 4 (shown in Figure 6.1, using the same sonic triggers. The phantom birdsong speakers from Study 4 were retained, but the phantom vehicle speaker was removed. During Study 4, it was found that participants had no suspicion of the phantom birdsong (likely as birdsong is often *acoustmatic* as discussed earlier), however the phantom vehicle sounds could cause participants to actively look for vehicles which were not there, which risked them discovering the phantom speaker. As the car park had a steady flow of traffic, the phantom vehicle speaker was not needed. Once again, participants were not informed of the hidden speakers and none of the participants suggested they had noticed them.

The qualitative measures from Study 4 were carried forward, and additional bespoke questions were introduced to assess how accurately the system responded to real sounds, and participants' awareness of, and connection to, the real environment, as participants in Study 4 often commented on feeling more aware or connected to their surroundings. The Involvement subscale of the Igroup Presence Questionnaire (IPQ) [165] was used to assess real-world awareness. While only one component of the larger validated instrument, the larger IPQ was less relevant to assessing real-world awareness and omitting other subscales kept the study questionnaire to a manageable length. Additional bespoke questions were used to assess how accurately the system responded to real sounds and connection to the real world, as participants in Study 4 suggested feeling more connected to the environment. The bespoke measures used in Study 5 are shown in Table 6.2.2, and the full questionnaire used is available in Appendix B.5.2. As before, participants assessed all six of the application variations, using the system for five minutes while walking back forth along the path. After each evaluation, participants completed the questionnaire. After all six evaluations, an overall interview was conducted to discuss the participant's experience.

The test applications were once again built using Unity, and used the same spatialisation toolkit as prior studies. The Quest 3 headset was not carried forward, in favour of a hardware setup that more closely resembled AAR equipment. The experiment software was run on a Dell i7 laptop carried in a backpack worn by the participant. Audio was presented over a pair of Soundcore C30i acoustically transparent earbuds, shown in Figure 6.3 which leave the ear canal unblocked, allowing virtual audio to be presented alongside real audio. The participant wore a hat which had a Supperware headtracker secured inside the headband to position virtual sounds correctly, and a Behringer BC lavalier microphone was clipped to the rim of the hat to feed audio to a sound classification model running on the laptop, via a Rode AI-Micro audio

Figure 6.3: The Soundcore C30i acoustically transparent earbuds used in Study 5.

interface. Finally, the participant used a wireless Xbox controller to input commands during the game scenario. As this hardware setup was not able to facilitate 6DOF tracking, virtual sounds were presented egocentrically, having only a fixed orientation relative to the user. The overall equipment setup used by participants is illustrated in Figure 6.4, and the audio files used in this study are available via Appendix A.2.

Sound Classification

Google's YAMNet model was used for classifying sounds in the microphone stream [84]. YAMNet is capable of classifying 521 different audio classes, including birds and cars. Every 480 ms, the model output a frame of classifications for the preceding 960 ms of microphone audio. In the Game scenario, the main "Bird" class was checked, as well as nine other classes covering specific birds or bird sounds on each YAMNet frame. If any of those classes had a confidence ≥ 0.1, this was considered a detection of a bird and a Sonimon was spawned.

In the music scenario, it was found that adjusting music on each YAMNet frame resulted in playback artifacts. Instead, in the Music scenario three frames were averaged together (covering 2 seconds of microphone audio) and the resulting confidence values were used instead. The same bird classes were checked, but a lower threshold of 0.05 was used to compensate for the larger detection window. For the *Safety* variation, the main "Vehicle" class was checked, as well as seven other classes which covered acceleration and engine sounds. If any of these classes had a confidence ≥ 0.25, a vehicle was considered to have been detected.

These detection settings were chosen after comparing YAMNet detections with human ones.

Figure 6.4: The overall equipment setup used in Study 5.

The author walked back and forth through the test space for twenty minutes, manually recording occurrences of birds or vehicles with YAMNet running alongside. These data were then used to set the detection thresholds to provide the best balance of detection accuracy. If a human classification did not have a corresponding YAMNet classification at or above the detection threshold within n seconds, it was recorded as a false negative. If a YAMNet classification at or above the detection threshold did not have a corresponding human classification within n seconds, it was recorded as a false positive. The n second window was implemented to account for delays in human reaction, and the fact that human identifications occurred once, while YAMNet identified every 0.5 seconds.

The final detection thresholds resulted in a false negative rate of 36.84% and false positive rate of 20.0% for birds, using a 3 second window. For vehicles, they resulted in a 1.16% false negative rate, and a 42.3% false positive rate, using a 6 second window (as an individual car would be active in the space for longer than an individual bird). The vehicle detection threshold was deliberately set to minimise false negative rates as in the *Safety* scenario where these were used, a false negative would be more dangerous to the user than a false positive. False positive rates were higher as a result.

YAMNet classifies 521 sound classes, of which 10 bird classes and 8 vehicle classes were used. Chance rate for these classes would be $\frac{18}{521}$ or 3.45%, and the chosen detection settings provided significantly higher detection accuracy. Pilot testing was conducted after setting these thresholds which also showed they provided a reliable experience.

Measure	Question	Scale
Augmentation	"My world felt augmented."	Strongly Disagree – Strongly Agree
	"The virtual elements felt connected to the real world."	Strongly Disagree – Strongly Agree
	"If I used the application in a different environment, the application would have behaved differently."	Strongly Disagree – Strongly Agree
	"The virtual elements felt like they were part of the same world as the real elements."	Strongly Disagree – Strongly Agree
	"The application was ___ than the real or virtual elements would have been on their own."	Much Worse – Much Better
Real Sound Response	"The application responded to real world sounds around me."	Strongly Disagree – Strongly Agree
Real Sound Response Accuracy	"The application accurately detected the real world sounds around me."	Strongly Disagree – Strongly Agree
Real World Connection	"I felt connected to my real world surroundings."	Strongly Disagree – Strongly Agree

Table 6.3: Author-developed measures deployed in Study 5 post-round questionnaires

6.2.3 Participants

20 people participated in Study 5, again recruited from university mailing lists and compensated with a £10 Amazon voucher for their time. As in prior studies, the only recruitment criteria were that participants did not have any hearing impairments. Study 5's participants consisted of:

- **Gender:** 8 men, 11 women, 1 non-binary person.
- **Age:** 10 aged 18-24, 7 aged 25-34, 3 aged 35-44.
- **AR Familiarity:** 1 very unfamiliar, 5 unfamiliar, 1 neutral, 11 familiar, 2 very familiar.
- **AAR Familiarity:** 1 very unfamiliar, 12 unfamiliar, 3 neutral, 3 neutral, 1 very familiar.

6.2.4 Results

Quantitative results were again analysed as a single factor comparison of application variation within each of the two applications, using Friedman tests with Nemenyi *post hoc* tests where relevant to evaluate significant differences. Full quantitative results are listed in Table 6.2.4 and illustrated in Figure 6.5. The full dataset is available via Appendix A.1.

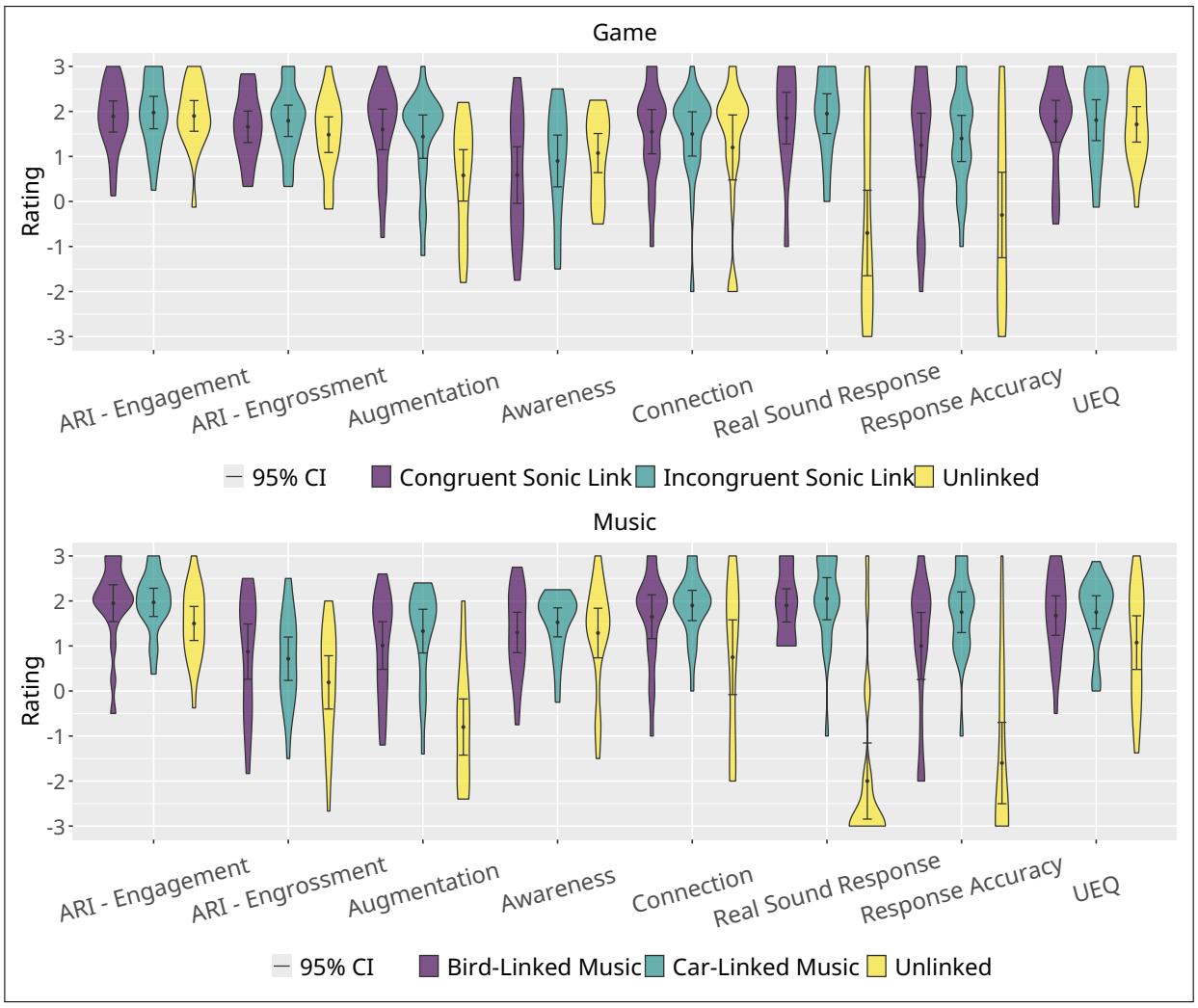

Figure 6.5: Study 5 questionnaire responses, separated by application and variation.

The analysis showed that across applications, sonically linked variations were rated as having a significantly higher Real Sound Response than the unlinked variation. The sonically linked Music variations also rated higher on Real Sound Response Accuracy. While a significant main effect was detected for this measure in the Game application, no significant pairwise comparisons were found. Similarly, an effect on Engrossment was detected in the Music app, but no significant pairwise comparisons were found. Both sonically linked Music variations also rated higher on the Augmentation measure than the unlinked variation, while only the *Congruent* Game had higher Augmentation than the unlinked Game. Finally, participants rated their real world Awareness as being lower with the *Congruent* Game than the unlinked variation. No other significant differences between application variations were found.

Measure		Variation A (M (SD))	Variation B (M (SD))	Variation C (M (SD))	Friedman Test				Sig. Post Hoc Comparisons		
					p	df	χ^2	Kendall's W	Pair	p	Mean Diff.
UEQ-S	Game	1.78 (0.992)	1.81 (0.97)	1.71 (0.841)	0.37	2	1.97	0.049			
	Music	1.68 (0.935)	1.75 (0.783)	1.08 (1.27)	.06	2	5.62	0.140*			
ARI Engagement	Game	1.89 (0.74)	1.98 (0.769)	1.9 (0.728)	.491	2	1.42	0.036			
	Music	1.95 (0.876)	1.97 (0.67)	1.5 (0.807)	.358	2	2.05	0.051			
ARI Engrossment	Game	1.66 (0.746)	1.79 (0.743)	1.48 (0.846)	.449	2	1.6	0.040			
	Music	0.875 (1.31)	0.717 (1.03)	0.192 (1.26)	**.015**	2	8.38	0.209*	*None*		
Augmentation	Game	1.6 (0.962)	1.44 (1.03)	0.58 (1.22)	**.004**	2	10.93	0.273*	A - C	.008	0.86
	Music	1.01 (1.13)	1.33 (1.04)	-0.8 (1.33)	**<.001**	2	26.70	0.668***	A - C	<.001	1.81
									B - C	<.001	2.13
Real Sound Response	Game	1.85 (1.23)	1.95 (0.945)	-0.7 (2.03)	**<.001**)	2	15.97	0.399**	A - C	.01	2.55
									B - C	.003	2.65
	Music	1.9 (0.788)	2.05 (0.999)	-2 (1.81)	**<.001**	2	21.21	0.530***	A - C	<.001	3.9
									B - C	<.001	4.05
Real Sound Response Accuracy	Game	1.25 (1.52)	1.4 (1.1)	-0.3 (2.03)	**.014**	2	8.60	0.215*	*None*		
	Music	1 (1.59)	1.75 (0.967)	-1.6 (1.93)	**<.001**	2	15.53	0.388**	A - C	.012	2.6
									B - C	.001	3.35
Awareness	Game	0.588 (1.35)	0.9 (1.23)	1.08 (0.929)	**.025**	2	7.4	0.185*	A - C	.038	-0.492
	Music	1.3 (0.958)	1.52 (0.688)	1.29 (1.18)	.210	2	3.11	0.078			
Connection	Game	1.55 (1.05)	1.5 (1.05)	1.2 (1.54)	.843	2	0.341	0.009			
	Music	1.65 (1.04)	1.9 (0.718)	0.75 (1.77)	.083	2	4.98	0.124*			

Table 6.4: Overall results for each measure in Study 5. Green denotes statistical significance. * denotes small effect size, ** denotes moderate effect size, *** denotes large effect size.

6.2.5 Interview Results

Interview results were analysed using auto-transcription from WhisperX [7], followed by manual correction. As in Study 4, the author and a colleague analysed the interview transcripts line-by-line to develop independent coding schemes, then discussed and merged these to form an overall thematic analysis of the data, resulting in the following themes:

Game Experience

Sixteen participants expressed a preference for one of the two applications, with 10 of those preferring the Game application the Music player. 17 participants also expressed a preference for one of the Game variations, with 11 preferring the *Congruent* variation, four preferring the *Incongruent* variation, and two preferring the *Unlinked* variation. Participants suggested the *Congruent* variation made more sense to them, and appreciated there being a clear connection between a real-world bird and the virtual bird Sonimon. In the *Congruent* condition, six participants felt the bird Sonimon was harder to localise than in the other variations, and six participants mentioned confusion as to whether they were hearing a real or virtual bird. Preferences for other variations were usually attributed to easier localisation (three participants), or a preference or affinity for dogs or cats (two participants).

> *"I felt that when it was a bird, that I felt more connected to the game and I felt it was more immersive." – P3*
>
> *"I think [my preferred application] was one of the games where there's a bird sound, because it felt more, like, in the reality. It was still kind of hard to distinguish at some points between the actual bird and the bird in the game, but that felt kind of exciting." – P11*

Music Experience

Six of the 16 participants with a preference for one of the applications preferred the Music application to the Game. 18 participants preferred one of the Music variations, with eight preferring the *Wellbeing* variation, seven preferring the *Safety* variation and three preferring the *Unlinked* variation. As in Study 4, participants noted real-world utility for the *Safety* variation: eight participants felt it had potential benefits for road safety, and seven felt it had more general utility. Two participants mentioned moments where the application responded to a car they were unaware of, however three participants also mentioned the *Safety* variation causing the music to be dimmed for the majority of the song. Three participants felt that sonically linked music was interruptive.

> *"I think it's very useful for sort of getting people to be immersed into music and at the same time, not. Sort of. I think it can almost be like a life-saving device if you*

put your headphones on and walk through a busy city with your earphones in" – P4.

"As far as the music applications were concerned, I felt that they were interruptive, because I wanted to listen to music and if there were more birds or more cars around me then my entire experience was getting disrupted and I wasn't feeling the music as I should have" – P3.

"There were times where I didn't hear anything, but it still noticed and then when the music went quiet, then I could hear the car" – P16.

Thoughts on Sonic Linking

Twelve participants felt positively about the idea of sonic linking, with seven participants appreciating how it connected real and virtual elements into a cohesive experience. Some participants again mentioned an impact on their experience of the real environment, with seven participants noticing themselves listening for birds and cars themselves, as if pre-empting the application, and three participants feeling more aware of their surroundings using the sonically linked applications.

Participants felt the sound detection model had mixed accuracy, with 11 participants noticing errors in bird detection, and four noticing errors in vehicle detection. However, this level of accuracy still appeared to be sufficient for the majority of participants.

"It makes you feel that you're present in both sides of the world, I mean the augmented world as well as the physical world." – P13

"I was actively listening to bird noises and being like, 'oh yeah, there's one, they'll probably... and yes, there's the game, it's triggered, it's heard the same thing.' So it was more... I was engaging with the external surroundings in a different way." – P18

"There was maybe some birds [that the system didn't detect], but maybe because it was further away or something like that. But most of the birds I heard, I then got the Pokémon sound." – P15

Future Ideas

Participants had a variety of suggestions for changing, extending, or customising the test applications. The most popular request (from 11 participants) was to introduce speech reactivity, particularly in the Music application. Participants felt this could remove the need to remove earphones when spoken to. Participants also suggested sonic linking in response to nearby people (three participants), doorbells or door buzzers (two), nearby dangers (two), or extending

the Game to respond to other animal triggers (two). Two participants also suggested the music system could operate in reverse to block out annoying sounds like crying babies by raising volume rather than lowering it. Seven participants offered ideas for future applications with sonic links, including adaptive music for cycling, more granular active noise cancellation, or context-sensitive notifications.

> *"Even if it was customisable, if you're in different settings you can turn [music reactivity] on for different noises. Because if you're in the city it's going to be going down for cars all the time, whereas if you're out in the countryside, birds are going to be constant." – P15*

> *"Sometimes people talk to me and I have to take, you know, my music off. And then I wouldn't have to do that if, like, someone's talking to me and then it goes down on its own and then it comes back." – P1*

> *"I listen to the directions to get somewhere on my bike when I'm cycling, and then I hear the directions because the music quietens [(a native Google Maps feature)], but it'll also make more sense for the music to quieten, you know, when there's also those cars." – P17*

6.2.6 Discussion

The principal finding from Study 5 is that the use of sonic linking results in a more augmented AAR experience. In both applications, Augmentation ratings were significantly higher for sonically linked variations than the unlinked variations. In the Music application, both the *Wellbeing* and *Safety* variations had higher Augmentation ratings than the *Unlinked* variation. In the Game application, only the *Congruent* variation had higher Augmentation ratings than the *Unlinked* variation, there was no significant difference between the *Unlinked* and *Incongruent* variations. Mean Augmentation ratings were also negative for the *Unlinked* Music app, and mildly positive for the *Unlinked* Game app. Considering the *Unlinked* Music app represents a non-AAR experience while the *Unlinked* Game variation meets most of the prior definitions of AAR set out in Chapter 3, these findings suggest that the use of sonic links can transform a non-AAR experience into AAR (as shown by the Music findings), while a congruent sonic link can elevate existing AAR approaches (evidenced by the Game results).

The *Congruent* Game was also the majority of participants preferred version of the game application. This preference was most often attributed to the clear real-virtual connection, while those that preferred the other variations usually provided more surface-level reasons such as an affinity for a particular animal or a different level of difficulty. Alongside the quantitative results, this suggests that a congruent real-virtual connection provides a superior experience. While only one example of a congruent sonic link was tested, these benefits could potentially apply in other scenarios also, and future work should explore this. A real bird could just as easily prompt a

virtual assistant to queue up The Beatles' *'Blackbird'* as the next track in a music player with augmented shuffling, or provide an interesting ornithology fact, for example. It will also be important to assess whether congruence affects user experience in other forms of real-virtual link such as the spatial and content-based links proposed by Schraffenberger [163].

In both apps, the sonically linked variations had a significantly higher Real Sound Response rating than the *Unlinked* variations. Mean Real Sound Response ratings were also positive for the sonically linked variations, and negative for the *Unlinked* variations. For the scenarios tested here and the detection settings used, it seems the YAMNet model is capable of creating a sonic link. As no differences between sonically linked variations in *Real Sound Response* or *Real Sound Response Accuracy* were detected, the findings further suggest that YAMNet can achieve such links across multiple applications, and multiple sound triggers. In interviews, some participants did report errors in the detection, which are to be expected. However, the results still show improvements from sonically linked variations compared to *Unlinked* variations, and participants noted an overall preference for sonically linked variations. The results suggest sonic links can be successfully created using existing sound classification models, but future models with improved accuracy will be necessary to improve user experience. Future work should explore improving detection accuracy for sonic linking.

While some participants suggested the sonically linked applications tested were not to their taste, sonically linked variations were more popular than *Unlinked* variations overall, and no negative effects on user experience were detected in the quantitative analysis. The quantitative results also suggest an influence of sonic linking on user immersion as an interaction effect for the *Engrossment* measure in the Music app was detected, though the analysis did not show any significant pairwise comparisons. Future work could explore this further.

Participants also noted changes in their experience of the physical space when using sonically linked applications, much like in Study 4. Many mentioned an increased awareness of their surroundings, either directly or by reporting listening for trigger sounds themselves which suggests an increased awareness or engagement with their surroundings. Interestingly, the quantitative measures introduced to assess *Awareness* and *Connection* with real-world surroundings showed no significant differences, except from between the *Congruent* and *Unlinked* Games, where participants were actually found to have lower *Awareness* in the *Congruent* condition compared to the control. This is a surprising result as it contradicts qualitative findings from both Study 4 and Study 5. This may be a function of question choice, and further work should be done to clarify how sonic linking can affect a user's relationship with their real surroundings.

Once again, participants reported occasional confusion as to whether a virtual bird was real or not, this time in the *Congruent* Game scenario. While this suggests the applications created virtual sounds indistinguishable from real ones, a key aim of augmented reality, this may not always be desirable. Further work is needed to explore how virtual and real audio sources can or should co-exist to minimise user confusion.

6.3 Studies 4 and 5 – Discussion and Conclusions

Until now, sonic linking with environmental sounds has never been properly explored in an AAR context. The work presented in this chapter forms an initial exploration of its potential in AAR, and the results from Studies 4 and 5 suggest sonic linking has strong potential in AAR and perhaps beyond, providing a significant step towards answering RQ3.

Both studies provide compelling evidence that sonic linking in this manner results in an improved AAR experience. Both studies demonstrated that a sonic link creates an AAR experience which feels more augmented – in Study 4 mean augmentation ratings were positive, while Study 5 showed significant improvements to augmentation between sonically linked scenarios and unlinked ones. This is particularly relevant when considering the Music application, where the *Unlinked* variation introduced in Study 5 represents a normal, non-AAR auditory experience. The fact that introducing a sonic link in the Music application can result in a significantly higher, positive mean rating for augmentation compared to the negative mean rating of the unlinked condition suggests that sonic linking can transform non-AAR experiences into AAR, supporting the definition introduced in Chapter 3 and this thesis's position that sonic links will be key to AAR in the future. As the Augmentation measures used in these studies were also developed by the author, it will be important for further work to assess how sonic links influence how augmented an AAR or XR experience feels.

Sonic linking may also improve an AAR user's awareness of their surroundings and increase a sense of connection with those surroundings, however the results from these studies do not clearly demonstrate or disprove this. In qualitative interviews, participants often reported feeling more aware of their environment or noting a feeling of connection. However, the quantitative measures introduced in Study 5 to explore this did not result in any clear conclusions, with the only significant difference actually suggesting that the *Congruent* Game scenario lowered awareness slightly compared to the unlinked control. Given these contradictory findings, future work should explore sonic linking's influence on environmental awareness in greater depth.

A central finding from both studies is that the congruence of a sonic link is an important factor. The results from Study 4 showed that a *Congruent* link resulted in a higher sensation of augmentation than an *Incongruent* link, and also that a congruent link felt more responsive to real sounds, despite responding to the same sounds in the incongruent scenario. Study 5 demonstrated that a congruent link felt more augmented than an unlinked scenario, but that an incongruent link did not. In interviews in both studies, participants also tended to express preference for a congruent scenario over an incongruent one. Study 5's *Unlinked* Game condition represents a current typical AAR experience, where virtual sounds are introduced or overlaid on the real world without a clear link between the two. The *Congruent* condition rating higher for Augmentation, which was not true for the *Incongruent* condition, suggests that congruent sonic links specifically can elevate existing AAR approaches.

6.3.1 Limitations and Future Work

As Studies 4 and 5 represent the first exploration of environmental sonic linking in AAR, there are some limitations to acknowledge, as well as opportunities for future work.

For example, the studies presented here only explored environmental sonic links based on the presence of a sound, the simplest form such a link could take. Future work could integrate aspects of real sound, such as its position or tonality, into a sonic link and explore how this influences the overall experience. Similarly, while Studies 4 and 5 deliberately maximised ecological validity by using a real world environment and introducing plausible, phantom sounds to drive the applications, this still represents a controlled scenario, and sonic linking must also be evaluated in true real-world scenarios.

While sonic links, particularly congruent ones, demonstrate promise for AAR, it is important to acknowledge that only two real-world sounds were used, and only three applications were assessed. These results may not hold true for other real-world sounds or application scenarios and future work should continue exploring the potential of sonic links. It is also important to acknowledge that sonic links may not be suitable for all AAR application designs, at least in the form evaluated here. While a sonic link could be as simple as positioning virtual sounds so as to not overlap with real sounds, not all AAR apps may have a clear congruent link that could be deployed, and designing an AAR application to respond to certain sounds could impose restrictions on the environments where it can be best used. Although these studies represent the first steps in exploring sonic linking with environmental sounds, these are important limitations to acknowledge and for AAR designers to consider.

Participants in both studies offered a number of ideas and suggestions for future sonically linked applications, and participants in Study 4 often mentioned a desire for visual elements in the Game scenario. While this may have been influenced by the use of a Quest 3 as the experiment platform, as this was not commonly suggested in Study 5, it does suggest that users could imagine sonic links being used in multimodal AR scenarios in the future. As sonic links elevated Augmentation, effectively coupling the real and virtual world more tightly together, one can imagine that the AR applications of the future could benefit from sonic links and future work could evaluate sonic linking in a multimodal context also.

6.3.2 Conclusion

With RQs 1 and 2 explored in previous chapters, this chapter took the first step towards answering RQ3, exploring how sonic links can be created between AAR applications and the environmental sounds in a user's real-world surroundings. As the first exploration of this concept and its potential application to AAR, these studies provide compelling evidence that such sonic links can benefit AAR applications. Environmental sonic links result in a more augmented experience, transforming auditory experiences into audio augmented reality. If these links are

congruent, they further elevate AAR experiences compared to existing approaches to AAR applications. Existing sound classification models are capable of creating these links, and the hardware needed to experience sonically linked AAR applications is already available. As a new technology, however, these results could be influenced by a novelty effect, and it remains to be seen how sonic linking functions in the real world, in uncontrolled scenarios, and over a longer period of time. This provided the motivation for Study 6.

Chapter 7

Sonic Linking In The Real World

Chapter 6 began the process of exploring a sonic link between an AAR system and real-world environmental sounds, suggesting that that such sonic links have promise and offer tangible benefits. However, Studies 4 and 5 only evaluated environmental sonic linking in controlled scenarios and short-term usage.

While efforts were taken with the design of Studies 4 and 5 to maximise ecological validity, such as by conducting the studies in a real-world public environment and using hidden speakers to supplement the natural soundscape, real-world environments will present different challenges for sonically linked AAR. For example, while birdsong was chosen for evaluation as it is present in many environments, rural environments or highly urbanised spaces may feature significantly more or less birdsong and either could influence the user experience. Real world environments will also feature many other types of sound that were not featured in Studies 4 and 5, which could influence user experience. For instance, heavy construction sounds may register as vehicle sounds, or target sounds may become masked by crowd noise in busy city streets. It is important to evaluate environmental sonic links in varied real environments to explore potential findings that the prior studies may have missed.

As broad, initial explorations of environmental sonic linking, Studies 4 and 5 were also limited in the length of time participants spent with the applications. Having found strong, promising initial evidence to motivate the use of sonic linking in AAR, it is important to evaluate it also over a longer time scale. As the first explorations of environmental sonic linking, it is possible that the findings from Studies 4 and 5 were influenced by a novelty effect, as this would have been a new experience for participants. By assessing sonic linking over a longer time period, any potential novelty effects can be better identified and therefore improve the interpretation of the results of Studies 4 and 5. Additionally, a longer-term evaluation of sonic linking, particularly in uncontrolled conditions, could potentially reveal more nuanced findings by expanding the scenarios in which the applications are used.

This chapter presents a final, longitudinal study on sonic linking with environmental sounds. The applications from Study 4 and 5 were developed further as representative mobile appli-

Figure 7.1: The Soundcore C30i acoustically transparent earbuds used in Studies 5 and 6.

cations, and an extended, in-the-wild evaluation was conducted, providing further insights on RQ3.

> **RQ3: How can a sonic link between environmental sounds and virtual elements be created in audio augmented reality?**

7.1 Study 6 – Sonic Linking in the Real World

Study 6 was designed to explore environmental sonic linking in uncontrolled, real-world scenarios and over extended use, as well as with more developed test applications that better represented a consumer application. Ten participants were issued with a smartphone pre-installed with the test applications, illustrated in Figure 7.2, and the same transparent earphones used in Study 5, illustrated again in Figure 7.1. The participants used both applications twice per day as part of their everyday routine, returning the device after one week. Experience sampling questionnaires were presented to the participants while using each application, and qualitative data were gathered through daily email interviews and a final exit interview.

7.1.1 Application Design

Game

The Game application from Study 5 was further refined for deployment on an Android smartphone. To reduce the need for specialist hardware, the smartphone compass was used in place

of a headtracker, with participants holding the phone and turning on the spot to localise sounds. While less accurate than a dedicated headtracker, this approach has been found to be sufficient for achieving AAR in prior work [80].

Expanding on the Study 5 implementation, the Game responded to both birds and cars. As Studies 4 and 5 had found congruent links to be beneficial, 10 Sonimon were included which could be caught, five corresponding to birds, and five corresponding to vehicles. To mimic other creature-catching games, some of these Sonimon were made rarer than others. For each set of five, two Sonimon were "common" (75% chance to spawn when target sound detected), two were "rare" (20% chance to spawn), and one was "very rare" (5% chance to spawn). Sonimon also had a randomly determined 'power', which was higher for rarer Sonimon to incentivise collection. Caught Sonimon persisted between game sessions, and a 'Sonidex' view was introduced where players could view their caught Sonimon. The 10 Sonimon in the game were:

- **Crowven** *(Bird):* Common Sonimon representing a crow or corvid. The basic Sonimon loop was overlaid with crow calls.
- **Eaglide** *(Bird):* Rare Sonimon representing an eagle or bird of prey. The basic Sonimon loop was overlaid with bird of prey calls.
- **Phoenos** *(Bird):* Very Rare Sonimon representing a phoenix. The basic Sonimon loop was overlaid with sounds of fire and screeches.
- **Pidgeowary** *(Bird):* Common Sonimon representing a pigeon. The basic Sonimon loop was overlaid with pigeon calls.
- **Twittwoo** *(Bird):* Rare Sonimon representing an owl. The basic Sonimon loop was overlaid with owl sounds.
- **Fumigon** *(Vehicle):* Rare Sonimon representing exhaust fumes from a vehicle. The basic Sonimon loop was overlaid with sounds of an idling engine and fuel pumps.
- **Nitroblast** *(Vehicle):* Very Rare Sonimon representing a nitrous oxide booster system. The basic Sonimon loop was overlaid with idling engine sounds, explosion sounds and compressed air whooshes, similar to how such systems are portrayed in popular culture.
- **Pluggspark** *(Vehicle):* Rare Sonimon representing an internal combustion engine's spark plugs. The basic Sonimon loop was overlaid with idling engine sounds and electrical arcing sounds.
- **Tyreon** *(Vehicle):* Common Sonimon representing a vehicle's tyres. The basic Sonimon loop was overlaid with idling engine sounds, tyre squeals and air compressors.
- **Vroomer** *(Vehicle):* Common Sonimon representing a vehicle's engine. The basic Sonimon loop was overlaid with idling engine sounds, and the sounds of engines being revved.

The random monsters from previous versions of the game were retained to provide a purpose and incentive to Sonimon collection, but these were refined to be sonically linked also. Whenever a triggering sound was detected, there was a 10% chance that a monster would spawn instead of a Sonimon – either a bird or vehicle monster depending on the sound. A player's most powerful Sonimon was deployed to battle the monster, with more powerful Sonimon allowing the monster to be defeated in fewer localisations. If the monster was not defeated, it would eat the player's strongest Sonimon, incentivising proper engagement with the game to maintain a stable of powerful Sonimon. The audio files used in this study are available via Appendix A.2.

Music

Building on the implementation in Study 5, the Music application was expanded to react to both vehicles and birds, as well as speech (a common request from Study 5). When any of these sounds were detected, volume of music would be lowered and the music would be filtered to allow better audibility. As in prior studies, high frequencies were filtered for birds and low frequencies for vehicles, while for speech the frequency band used for telephone calls (300 Hz to 4 kHz) was filtered.

The music selection was also expanded from a single test track to a corpus of 10 albums. This corpus consisted of the top five best-selling albums of all time in the UK Official Albums Chart, and five of the top albums from Von Appen and Doehring's meta-analysis of most influential albums [185]. Only one album per artist was included in the corpus. This created a corpus which included both critical and commercial successes, and which could appeal to a broad range of participants. Audio was imported from legitimate CDs. The final corpus of albums consisted of:

- **Queen** - *Greatest Hits* (UK top seller)
- **ABBA** - *Gold* (UK top seller)
- **Adele** - *21* (UK top seller)
- **Oasis** - *What's The Story Morning Glory* (UK top seller)
- **Michael Jackson** - *Thriller* (UK top seller)
- **The Beatles** - *Sgt. Pepper's Lonely Hearts Club Band*[1] (Critical)
- **Nirvana** - *Nirvana* (Critical)
- **The Beach Boys** - *Pet Sounds* (Critical)

[1] *Revolver* ranked higher on Von Appen and Doehring's list, but *Sgt. Pepper's* was also represented on the UK top seller list and so was included instead.

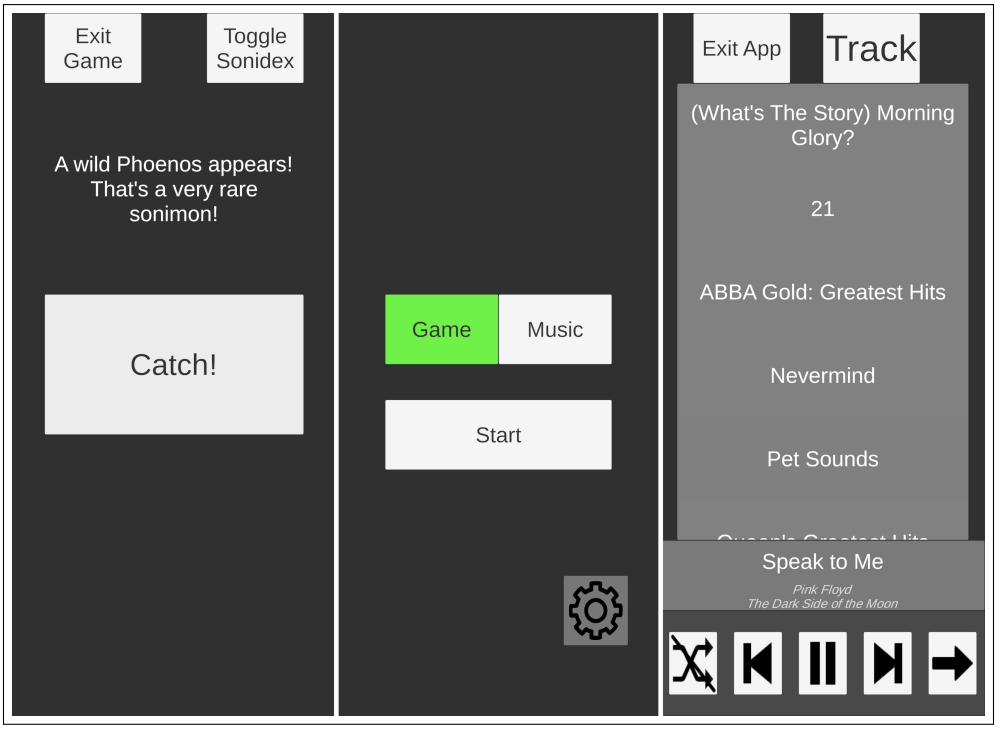

Figure 7.2: The mobile application developed for Study 6, showing the user interface for the game (L), main menu (C), and user interface for the music player (R).

- **Pink Floyd** - *The Dark Side of the Moon* (Critical)
- **Velvet Underground** - *The Velvet Underground and Nico* (Critical)

Additionally, functionality for users to load in custom music files was also included, in case the supplied music was not to participants' tastes.

Figure 7.2 shows the user interface for the different applications.

7.1.2 Experimental Design and Methodology

Participants were supplied with an Android phone (a Google Pixel 4, Pixel 7, or Pixel 9), and a pair of the Soundcore C30i earphones. The test app was pre-installed on the devices, and built using Unity 2022.3. The same spatial audio solution as prior studies was used for game sounds, and music was presented in stereo format. Sound detection was achieved using the same YAMNet system as Study 5, utilising the built-in microphone on the device.

Participants were issued the equipment on a Monday, and had the equipment until the following Monday. They took the device home, and used the application in the real world as part of their normal routine. Participants were required to use both applications at least twice per day for four consecutive days (Tuesday-Friday), using each application until at least one experience sampling questionnaire was completed. Participants were able to use the applications beyond this period if they chose.

After three minutes of using an application, an experience sampling questionnaire was presented to the user. The experience sampling questionnaire contained four bespoke questions designed to assess the detection system's accuracy, how augmented the experience felt, how connected the user felt to their environment, and how aware the user felt of their surroundings, as Studies 4 and 5 suggested sonic linking had an impact on all of these aspects of the experience. All questions were assessed on 7-step Likert scales, from Strongly Disagree to Strongly Agree. The experience sampling questions were:

- **Accuracy:** "The system is responding accurately to sounds."
- **Augmentation:** "My world feels more augmented than normal"
- **Awareness:** "I feel more aware of my surroundings than normal"
- **Connection:** "I feel more connected to my surroundings than normal"

If the participant used the application further, the experience sampling questionnaire was presented again every five minutes. At the end of each day, participants were emailed a short list of interview questions to complete:

- Where did you use the apps today?
- How reliable did you find the applications today?
- Did you have any frustrations with the apps?
- How was the experience compared to a normal music player or game?
- Was your experience of your environment different in any way? Did you behave any differently?
- *How was the experience today compared to yesterday?* (days 2–4 only)

A final exit interview was also conducted at the conclusion of the study, discussing the participant's experience over the week.

7.1.3 Participants

As in prior studies, participants were recruited from University mailing lists. 10 participants took part, with the only criteria again being that they had unimpaired hearing, and that they had not participated in Studies 4 and 5 (to ensure their data were not influenced by those experiences). Study 6's participants consisted of:

- **Gender:** 5 men, 5 women.

- **Age:** 4 aged 18-25, 5 aged 25-34, 1 aged 35-44.
- **AR Familiarity:** 2 very familiar, 4 familiar, 1 neutral, 3 unfamiliar.
- **AAR Familiarity:** 2 familiar, 1 neutral, 6 unfamiliar, 1 very unfamiliar.
- **Walking Frequency:** 5 walked multiple times per day, 3 walked once per day, 1 walked multiple times a week, and 1 walked once a week.
- **Music Listening While Walking:** 5 always listened to music when walking, 2 listened very often, 1 listened often, and 2 listened sometimes.
- **AR Games Experience:** 2 played AR games often, 4 played AR games sometimes, 1 played rarely, and 3 had never played an AR game.

Participants were compensated with a £50 Amazon voucher for taking part.

7.2 Results

Experience sampling questionnaires were analysed using cumulative linked mixed effect models (CLMM), assessing how measures varied over the usage period. CLMMs were used as some participants completed more sampling questionnaires than others, which other tests such as Friedman would not properly account for. Each measure was used as a dependent variable, the day (1, 2, 3, 4) was used as the fixed effect, with a random effect from individual participants. Full quantitative results are detailed in Table 7.2, and shown in Figure 7.3. The full dataset is available via Appendix A.1.

The quantitative results suggest that user experience was largely stable across the four day test period. No evidence was found to suggest that Accuracy, Augmentation, or Connection changed significantly over the course of the week. Awareness was also stable for the Music application. For the Game application, Awareness was found to be significantly higher by the end of the week than at the beginning. For days 2 and 3 of the study, participants also rated the Game system as having higher Accuracy and Connection than day 1, however there was no difference detected for these measures between day 4 and day 1.

7.3 Interview Results

Both the daily interview and exit interview results were analysed using the same approach as Studies 4 and 5 – WhisperX auto-transcription followed by manual correction, then line-by-line analysis by the author and the same second coder as Studies 4 and 5. Analyses were then compared, discussed, and merged to form an overall analysis of the data.

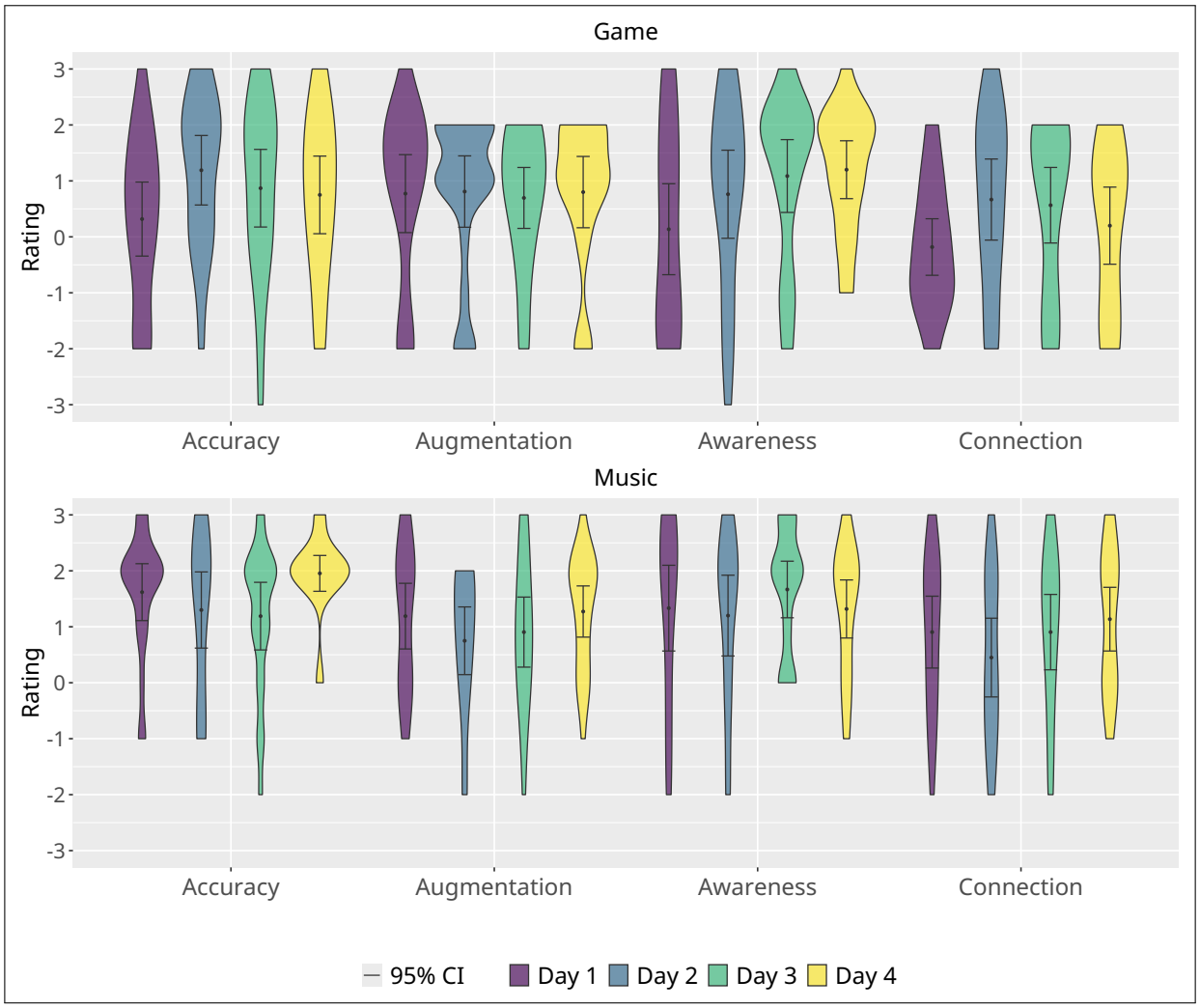

Figure 7.3: Study 6 experience sampling responses, separated by application and measure.

		M (SD) by Day				AIC	Participant Random Effect		Fixed Effects								
							Var.	SD	Day 2			Day 3			Day 4		
		1	2	3	4				Est.	SE	*p*	Est.	SE	*p*	Est.	SE	*p*
Accuracy	Game	0.35 (1.33)	1.22 (1.22)	1.03 (1.22)	0.75 (0.79)	293.8	2.05	1.43	1.395	0.563	**.013**	1.164	0.563	**.039**	0.637	0.578	.270
	Music	1.6 (0.91)	1.3 (1.27)	1.3 (1.08)	1.93 (0.71)	232	0.91	0.95	-0.536	0.648	.408	-0.724	0.599	.227	0.538	0.603	.372
Augment.	Game	0.73 (1.43)	0.83 (1.19)	0.75 (1.16)	0.8 (1.18)	234.1	3.45	1.86	-0.106	0.617	.863	-0.100	0.576	.863	-0.061	0.617	.922
	Music	1.23 (1.15)	0.77 (1.12)	0.92 (1.24)	1.23 (1.06)	251.7	2.46	1.57	-0.846	0.600	.159	-0.590	0.601	.326	0.105	0.583	.857
Awareness	Game	0.3 (1.83)	0.83 (1.6)	1.05 (1.28)	1.2 (1.03)	270.1	3.99	2.00	0.992	0.589	.092	1.429	0.589	**.015**	1.514	0.609	**.013**
	Music	1.45 (1.52)	1.23 (1.54)	1.62 (1.09)	1.38 (1.08)	213.1	7.91	2.81	-0.785	0.655	.231	0.274	0.629	.663	-0.34	0.617	.581
Connection	Game	-0.08 (0.87)	0.7 (1.46)	0.65 (1.26)	0.2 (1.34)	274.2	2.26	1.50	1.406	0.578	**.015**	1.142	0.565	**.043**	0.394	0.567	.488
	Music	0.98 (1.36)	0.45 (1.42)	0.87 (1.52)	1.17 (1.17)	244.5	5.18	2.28	-0.988	0.615	.108	-0.269	0.576	.641	0.524	0.569	.357

Table 7.1: Overall results for each measure in Study 6. Green denotes statistical significance. 86 observations for Game application, 84 observations for Music application.

7.3.1 Daily Interviews

In daily interviews, participants broadly noted a positive experience using the apps, often describing the apps as reliable (21/40 days) or free of frustrations (19/40 days). When participants did mention an issue with the application, it was often a detection error (7/40 days), either missing or misclassifying a sound. Wind was often mentioned as causing problems with the detection. Four participants found it frustrating to not be able to lock the music player without the music stopping (a limitation of Unity). Another common complaint, particularly on day 1 of participation, was that participants found the Sonimon difficult to localise in the game (seven participants). Often they mentioned this becoming easier in later days (5 participants).

> *"The only issue I noticed was in the afternoon, since it was more windy, both apps reacted to that and either generated sonimons when there wasn't a sound in the environment, or lowered the music volume, again with no sound [in] my surroundings." – P6, Day 3*

> *"I was finding it difficult to pinpoint the sound, feeling like it was moving around rather than being in one place, and usually I feel I'm quite good at games!" – P8, Day 1*

> *"I could realistically pinpoint where the sounds were coming from, much better than the previous day!" – P4, Day 2*

Participants primarily used the applications in urban environments (39/40 days), and often as part of their commute (17/40 days). Participants occasionally also used the applications in parks or green space (7/40 days). As in previous studies, some participants also mentioned

noticing changes in their own behaviour, such as listening for the sound triggers themselves (three participants), or noticed an increase in their environmental awareness (seven participants).

> *"I have noticed that especially when I have the game open, now I listen out, especially for birds, without thinking about it." – P6, Day 3*
>
> *"I was more aware of my surroundings and I was noticing if there are any cars crossing by or birds chirping to see what shows up on the app" – P3, Day 2*
>
> *"I definitely felt more aware of my surroundings because I could hear more than I usually can when I am just listening with my headphones on for the music player." – P7, Day 4*

7.3.2 Exit Interviews

Exit interviews were conducted when participants returned the study equipment, discussing the overall experience and how participants felt about the individual applications.

After using the applications for the week, the majority of participants expressed interest in using them, or a refined version, in the future, should similar applications become available or existing applications be expanded to include sonically linked features. Five of the ten participants said they were more interested in sonic linking than they had been before taking part in the study.

> *"It was a lot more fun than I was expecting it to be, it felt just like a game. I could really see it actually taking off in the future. I didn't really have an opinion before, but obviously I'm positive now." – P9*
>
> *"I would actually consider swapping out the music player I have, or at least supplementing it with the alternative, which is a bit more reactive, today." – P10*

As with the daily interviews, five participants reported frustrations with detection errors. Six participants mentioned having a higher level of environmental awareness when using the applications, being better attuned to environmental sounds nearby.

> *"I would say I have mixed feelings on the application after using it. Because I felt like one particular thing that I noticed was that if is really windy, then it doesn't pick up audio as well. I can hear the birds, I can hear the car passing by, but still it didn't hear anything." – P3*
>
> *"I think it worked well in terms of when there was a car going past or someone speaking, like, the volume lowering which was interesting. It did make me more aware of when there was a car because usually I use noise cancelling headphones." – P8*

The Music app was well received overall, and participants offered a number of suggestions for how it could be improved further. The largest frustration was with the player being unable to run in the background (a Unity limitation as mentioned before). Participants were keen to customise how the music responded to stimuli, seeking different behaviour in different contexts like responding to different sounds in certain areas, ignoring cars that have been nearby for an extended period, or having the music dimming being proportional to the volume of the triggering sound. Three participants responded positively to the speech-reactive functionality, though one would prefer it only react to speech directed at the user.

> *"What I also thought about was like even being able to select parts on a map. You know that, okay, if I'm there, if I'm in the park, for example, and I'm listening to music, then I'm not talking to anyone. So then I don't care if there's people talking around." – P1*

> *"But having a seamless conversation with just music in the background, that would very accurately quiet down when I'm talking to someone, it was very helpful, especially with just a very quick interaction at the shop. I didn't have to take my headphones off." – P10*

As with the daily interviews, participants noted that the Game experience began quite difficult, and became easier over the course of the week. Participants often mentioned the experience as being novel, and enjoyed a game experience that did not require a screen. As with the Music app, participants often offered suggestions for improvement, including the addition of visual elements, responding to other animal sounds or having more granular sonic links, such as specific Sonimon for specific car sounds.

> *"What's cool is that, especially in the game, is that you don't have to look at the screen all the time, so you just walk and then you get the audio cue and then you do something." – P1*

> *"It would be fun if it reacted...well, for the game specifically that 'oh, yeah, this specific car sound gives you a specific..."' – P6*

7.4 Discussion

The results from Study 6 broadly serve to reinforce those of Studies 4 and 5, suggesting that those findings hold true both in real-world environments and over extended usage periods. Mean ratings across quantitative measures were positive, indicating that existing smartphone microphones and machine learning models are sufficiently accurate for facilitating an environmental sonic link in the real world. The mean ratings also reinforce the earlier findings that sonic linking can improve a user's sensation of augmentation, environmental awareness, and connection to their surroundings.

The quantitative results do not suggest any evidence of a strong novelty effect for sonically linked AAR applications, with experience sampling ratings for all measures being no lower on day 4 than on day 1, which suggests a stable experience. Had there been a significant novelty effect influencing the results from Studies 4 and 5, there would have been an expectation that ratings would be lower on day 4 than day 1. However, this study only covered a four day period, and longer evaluations could still reveal a longer-term, or mild, novelty effect. Extended testing with a larger sample of users may also reveal more nuanced findings. Overall, however, Study 6 lends further credibility to the findings from Studies 4 and 5: environmental sonic linking results in a more augmented experience, and improves a user's awareness of and connection with their surroundings.

While no measures showed a significant decrease in ratings over the course of the week, a significant increase in Awareness was detected for the Game application on day 4 compared to day 1. There was no improvement in Awareness for the Music app over the same period, and in interviews participants often mentioned the game being difficult at the beginning of the week and becoming easier over time. This suggests then that this change in Awareness reflects the cognitive demands of the Game becoming lower as participants became more familiar and skilled with the game, therefore allowing for greater environmental awareness, rather than this improvement being a function of environmental sonic linking itself.

The interview results also indicate a broadly positive experience, although participants did identify some frustrations. A common complaint was in the reliability of sound detection, with some participants complaining of detection errors, particularly with wind. As an early exploration of sonic linking, and with YAMNet being a general purpose sound classification model, it is to be expected that performance would not be perfect, however it will be important to prioritise detection accuracy with the sonically linked applications of the future. A number of participants also noted difficulties in localising Sonimon in the game application. While this improved over the course of the week, it is likely that this is to do with the use of a phone compass over the headtracker used in prior studies. Many earbuds and headphones now ship with gyroscopes and sensors that can track head orientation, as well as high-quality microphones which could assist with detection accuracy, and as such devices become more widespread and easier to interface with, these problems may be reduced in the future. Importantly, the issues raised by participants were to do with the implementation of environmental sonic linking here, and not the concept itself, and are to be expected as an exploratory study. A majority of participants mentioned an interest in using sonically linked applications in the future, and when taking into account the issues identified by participants, the fact that the majority remained interested in sonic linking, or in many cases more interested in sonic linking than when they began, suggests an appetite for sonically linked AAR applications.

7.4.1 Limitations and Future Work

Study 6 was conducted to address limitations in Studies 4 and 5, however it has some limitations of its own to acknowledge, as well as opportunities for future work to build upon it. While Study 6 did allow participants to use the applications wherever they wished, it is notable that the vast majority of participants used them in urban environments, specifically the city of Glasgow. While these were often supplemented with parks or green spaces, this does mean that the results from Study 6 may not generalise to other environments, for example rural settlements, nature reserves, or metropolitan environments. Exploring environmental sonic linking in such environments could lead to more nuanced findings in the future, and future work should explore this. However, as Glasgow does feature aspects of all of these environments to some degree, the findings from Study 6 provide a good baseline.

As an initial exploration, the findings from this study are also based on 10 participants evaluating the applications for a $\sim$ 4 day period. While no novelty effect was detected in this study, wider, longer-term evaluations could still reveal novelty effects or other valuable insights.

Many participants also offered suggestions for expanding or improving them. For example, many participants suggested the ability to customise the Music app's response to different sounds – something which could help with tailoring to specific environments, or introducing more animal sounds into the Game scenario. Once again, participants also mentioned an interest in visual elements in the Game, suggesting that sonic linking could have applications in multimodal AR. Future work could explore environmental sonic linking in further developed and more varied applications to verify whether these findings hold true in other application contexts.

7.5 Conclusion

RQ3 was posed to assess how sonic links can be created between real-world sounds in a user's environment, and an AAR experience. By evaluating sonically linked applications over long-term, uncontrolled usage, Study 6 concludes the evaluation of environmental sonic linking introduced in Chapter 6, providing additional evidence to support the findings from Studies 4 and 5. The results suggest that even over extended, real-world usage, environmental sonic linking improves a user's sensation of augmentation, as well as their awareness of, and connection to, their real world surroundings. Study 6 finds no evidence to suggest these findings are influenced by a strong novelty effect. Additionally, Study 6 shows that such sonic links can be facilitated with existing tools, using accessible consumer hardware, today, and that users are enthusiastic about their potential in AAR and beyond. Alongside the work presented in Chapters 5 and 6, it provides compelling evidence of the importance and potential benefits of sonic links in AAR asserted in Chapters 1 and 3. The final chapter will discuss these in more detail.

Chapter 8

Conclusion

The opening pages of this thesis asserted that there has existed a fundamental gap in AAR thus far – a lack of auditory links between the real world and the virtual sounds used to augment it. The preceding chapters have set out six studies and additional contributions both to evidence this, and to explore this missing component of AAR.

Chapter 3 reviewed existing definitions of AAR, their strengths and pitfalls, and synthesised a new definition. It then set out a case for auditory links in AAR – AAR systems that linked real and virtual acoustically or sonically. Chapter 3 concluded that "audio augmented reality is the creation of a virtual world, represented through sound, which is linked to the real world and allows users to achieve goals in world by interacting with the other."

Chapter 4 explored the creation of an acoustic link in AAR through Studies 1 and 2. Study 1 explored the perception of various acoustic reproductions in a controlled lab environment with both traditional and AAR-focused playback devices. Study 2 built on these findings by exploring the plausibility of different acoustic reproductions in real-world environments, and using more representative AAR applications. The results from these studies suggest that it may be possible to create plausible acoustic links with less detailed acoustic reproduction than existing literature suggests, particularly in real-world AAR scenarios, although further work will be required to confirm this.

Chapter 5 began exploring the creation of a sonic link in AAR. Study 3 focused on how AAR applications can use human-produced action sounds as control inputs, assessing three different categories of action sound and comparing them to existing AAR control schemes through three AAR minigame experiences. The results suggest that a sonic link between action sounds and an AAR system can be viably created, and particularly identifies sonic gestures as being a potentially promising control input in the future.

Chapter 6 then focused on sonic links originating from environmental sounds through Studies 4 and 5. Study 4 explored the creation of an environmental sonic link in a best-case scenario, evaluating design variations of three different AAR applications. Study 5 then refined these applications and the evaluation protocol further, and introduced representative AAR hardware

and sound classification models to create fully functional prototypes of two of these applications. The results from these studies suggest that environmental sonic links improve a user's awareness of and connection to their surroundings, and that they improve a user's sensation of augmentation, elevating non-AAR experiences to become AAR. These studies also identified congruent links, where there is a clear connection between real-world trigger and virtual-world effect, as being particularly desirable.

Chapter 7 followed up Studies 4 and 5 by evaluating these sonically linked applications over an extended period of uncontrolled, real-world usage. The applications were further refined for use on existing smartphones, and issued to participants for a week of usage in their day-to-day lives. The results suggest that the findings from Studies 4 and 5 hold outside of the controlled test environment, and are not subject to a strong novelty effect, providing a stable experience over the period assessed.

The work presented in these chapters provides insight into the three research questions set out at the beginning of this thesis:

8.1 RQ1: How can an acoustic link between real and virtual elements be created in audio augmented reality?

RQ1 was addressed in Chapter 4, which explored the most obvious form an acoustic link can take – applying the real-world acoustics of a space to virtual AAR sounds to improve their plausibility. Existing work suggests 3rd-order reproductions are the perceptual ceiling for plausibility, with higher spatial resolutions providing no significant improvements to plausibility. However, plausibility is rarely evaluated in real-world or AAR contexts where acoustic perception may be different.

Study 1 addressed RQ1 by evaluating a variety of different acoustic reproductions (varying in spectral bandwidth and resolution) using both professional-grade studio headphones and acoustically transparent audio glasses which represent potential AAR hardware of the future. It provided supporting evidence that the inclusion of acoustic cues improve plausibility of virtual sounds – key to AAR – and that this holds across playback device. However, the Study 1 results did not provide clear insights on how best to balance accuracy and complexity for plausible reproductions.

Study 2 went on to evaluate acoustic reproductions in multiple real-world spaces, again with multiple playback devices, and using both a representative AAR app and a formal listening test to assess both AAR and critical listening scenarios. The results from Study 2 provide an initial suggestion that the plausibility differences between 1st- and 3rd-order acoustic reproductions may be minimal, not demonstrating a significant improvement to plausibility when increasing spatial resolution above 1st-order. As an initial exploration, and an implication that is contrary to existing literature, more work will be required to confirm this. It also provided initial evidence

that the plausibility of acoustic reproductions may also be influenced by application context, however this will similarly need confirmed through further work. For AAR applications, as opposed to the critical listening tests usually used to assess auditory perception, it is possible that simpler acoustic reproductions could be all that is required. Study 2 also investigated the echoicity of source audio, finding no evidence that moderate echoicity is a factor in virtual sound plausibility. This suggests that existing sound design practices can be deployed alongside an acoustic link in AAR without compromising plausibility. Finally, Study 2 found some evidence that spectral bandwidth contributes to plausibility, suggesting that full-bandwidth RIRs should be prioritised. However, this only applied in the outdoor space and not in the highly reverberant test space, meaning that this may not be critical in all AAR scenarios.

Taken together, Studies 1 and 2 suggest that an acoustic link can be created in AAR using acoustic reproductions of a 1st-order or greater resolution. Including such a link improves the plausibility of virtual sounds, a boon for AAR applications. As Studies 1 and 2 used measured RIRs to assess the fundamentals of acoustic reproduction, it is reasonable to expect these findings to hold regardless of whether an AAR app uses a measured RIR, or an equivalently accurate simulation, which AAR applications of the future are likely to deploy. Further work will be needed to explore these findings in greater depth.

8.2 RQ2: How can a sonic link between human-produced sounds and virtual elements be created in audio augmented reality?

RQ2 was addressed in Chapter 5, which explored the use of human-produced sounds – action sounds – as a control input for AAR applications. Study 3 evaluated three different sonic control schemes, comparing them to existing common control schemes and common scenarios identified using a systematic review.

The results from Study 3 suggest that a sonic link can be created using any of the three sonic control schemes – speech, music, and sonic gesture – but that some are better suited than others. Speech control was found to be viable, however it can lead to self-consciousness in users which may make it less appealing in public AAR usage. A musical control was also found to be workable, however it was found to be a polarising control method and likely requires further evaluation before widespread AAR usage. Most promisingly, Study 3 found that a sonic link can be created using sonic gesture, which was well-received by participants and the most popular control, performing competitively with established control schemes. Sonic gesture controls can likely be deployed as a viable alternative, or supplement, to physical gesture in future AAR systems.

As a Wizard of Oz methodology was used, these initial findings are presented under a best-

case scenario, and how these control schemes perform with live detection will need to be assessed. As the first study in this area, more detailed evaluations will also be required, however the initial insights presented here show promise.

8.3 RQ3: How can a sonic link between environmental sounds and virtual elements be created in audio augmented reality?

RQ3 was addressed in Chapters 6 and 7, which explored the use of environmental sounds to drive and inform AAR applications. As this concept has never been evaluated properly before, three studies were conducted to provide a robust exploration and a strong foundation for future work.

Study 4 explored sonic linking in a best-case scenario, evaluating the overall concept of environmental sonic links across three different AAR applications in a loosely controlled real-world environment to maximise ecological validity. The results from Study 4 suggest that the use of environmental sonic links provide a number of benefits in AAR, improving a user's awareness of their surroundings, altering their connection to the real world, and creating an improved sense of augmentation. These findings hold across multiple use-cases, and with multiple environmental sounds. Additionally, Study 4 finds evidence that congruent sonic links, where there is a strong thematic relationship between a real-world trigger and a virtual-world effect, feel more augmented and more responsive to the real world than environmental sonic links without this clear congruence.

Study 5 followed on by exploring environmental sonic linking using currently available technology, providing a more representative evaluation of the current potential of environmental sonic links. It assessed whether the findings from Study 4 can be harnessed with current technology, and whether they hold when compared to an unlinked control condition. The results suggest that the benefits identified in Study 4 are clear improvements to unlinked scenarios. The inclusion of environmental sonic links can elevate a non-AAR application into AAR (reinforcing and supporting the definition of AAR introduced in this thesis, though this should be verified by other authors), and congruent links offer a more augmented experience than existing AAR approaches as well as improving the user's awareness of the real world. Only a small set of real-world sounds and application scenarios were explored, and these findings should be confirmed more broadly. These results were found using existing sound classification models and off the shelf hardware, meaning that these benefits of environmental sonic links can be leveraged today.

Finally, Study 6 explored environmental sonic linking in uncontrolled, real-world scenarios, over extended usage, and with refined, more representative AAR applications. No evidence of a novelty effect was found, and the results suggest that the benefits of environmental sonic linking

found in Studies 4 and 5 hold in real-world usage, however these results are limited to a small sample size and usage period, as well as being specific to usage in the city of Glasgow. Further work will be required to corroborate these findings, however by validating the findings from Studies 4 and 5 in the real world, Study 6 strengthens a set of findings which contribute to a fundamental gap in AAR.

Overall, Studies 4 through 6 suggest that environmental sonic links offer significant benefits to AAR, creating more augmented experiences that can alter the way a user experiences their real-world surroundings. They can be created today, using existing mobile hardware, in uncontrolled real-world environments, and using existing sound classification models. As technology improves, these environmental sonic links will only become more accessible, and more capable.

8.4 Contributions and Recommendations

Across six exploratory, mixed-methods user studies, this thesis has explored both the benefits and techniques of creating acoustic and sonic links in audio augmented reality. It provides the following novel contributions and evidence-backed recommendations to the field:

1. A new definition for audio augmented reality which unifies and clarifies prior discussions. It provides a clear answer as to the difference between 'listening to audio' and AAR, which is supported by the studies presented in this thesis, particularly 4 through 6.

2. Insights on the perceptual differences between headphones and audio glasses: audio glasses inflate localisation error, however have no notable impact on virtual sound plausibility, making them a viable platform for future AAR applications when their limitations are accounted for.

3. Insights on how spectral bandwidth and spatial resolution affect listener perception of different acoustic reproductions in AAR scenarios. This thesis provides some initial evidence that in real-world usage, virtual sound plausibility may be influenced by application context, and that 1st-order reproductions may represent a sweet spot for plausible acoustic reproduction. This thesis recommends the consideration of 1st-order reproductions as a balance between perceptual fidelity and computational complexity, and that spectral bandwidth should be prioritised where possible.

4. The first evaluation of sonic control schemes in AAR contexts. Speech controls have promise and are usable, though can lead to self-consciousness in some users. Sonic gestures are the best-regarded and most usable and this thesis recommends developers consider deploying them in place of or in addition to physical gestures in AAR contexts.

5. The first evaluation of environmental sonic linking. The use of environmental sonic links may transform a non-AAR auditory experience into AAR, and can improve users' aware-

ness of and connection to real world surroundings. This appears to be true across multiple use-cases, multiple triggering environmental sounds, as well as extended usage and both controlled and real-world environments. This thesis recommends the inclusion of environmental sonic links into AAR applications going forward.

6. The congruence of a real-virtual link may have an influence on user experience, particularly the sensation of augmentation. Congruent sonic links offer significant improvements to augmentation compared to incongruent sonic links. This thesis recommends prioritising congruent sonic links where possible.

8.5 Limitations and Open Questions

While this thesis provides a number of key contributions to the field of audio augmented reality, as exploratory work there remain some limitations of the work presented here, and some outstanding questions for future work to explore. Firstly, many of the studies presented in this thesis exposed early insights that were infeasible to fully explore as they were not central to the research questions. Study 2 provided an initial indication that the plausibility of acoustic conditions may be influenced by a user's context, with the same acoustic reproduction being rated as more plausible in an application scenario than a formal listening test. While Study 2 found this difference significant, there were a number of confounding factors that may also have impacted these ratings such as application ratings being given retrospectively and for different sound content, and future work could investigate this potential effect in more detail. Similarly, Study 4 revealed that in an AAR application which uses naturally occurring sounds as virtual elements, users can become confused as to whether a virtual sound is real or not. While this can be desirable and suggests a fully plausible augmentation of reality is possible through sound, when these sounds are informative as they were in Study 4's Notification application, this can result in a reduced user experience. The variations of the Notification application used in Study 4 were designed to explore a potential solution to this problem by using audio effects to make such sounds clearly virtual. However, the results from Study 4 showed that this was not a viable solution and future work could explore alternative methods for resolving this real-virtual confusion.

While Studies 4 through 6 provided compelling evidence of the benefits of environmental sonic linking, they also only explored sonic links as triggers, with applications responding only to the presence of a sound. There are many other characteristics of a sound, such as its position, loudness, pitch, or frequency content, and one can imagine future sonic links drawing upon these aspects of sound to create more nuanced sonic links, something future work could explore. For example, the Game application explored in those studies could be expanded so that Sonimon not only appear when a real bird sings, but also originating from the position of that bird. The Music application could adjust its filtering to more closely match the frequency content of the

triggering sound, or base the level of volume reduction on the loudness of the real sound. As discussed in Chapter 5, the same could be done for sonic controls, incorporating the melody of a spoken phrase or rhythm of a sonic gesture into the control input.

Another key consideration is that the work presented in this thesis has focused only on acoustic and sonic links originating from the real world affecting the virtual one. However, one can also imagine links could originate from the virtual world that could affect the real. An acoustic link originating from the virtual world could see the acoustics of a real space morphed to match the current scene in an AAR audiobook, while a sonic link originating from the virtual could transpose real-world sounds to harmonise with the music a user is listening to, or the real soundscape around a user could be morphed to match a specified 'theme'. As an initial exploration of acoustic and sonic linking such potential applications are outside the scope of this thesis, but are conceptually compelling and seen as a key avenue for future work to build on the findings presented here.

Another key avenue is assessing the potential of acoustic and sonic linking beyond AAR. Throughout the studies presented here, participants expressed interest in audiovisual applications, and there is no reason that acoustic and sonic links could not be deployed as part of multimodal XR applications. The principle findings of this thesis are that acoustic links result in more plausible virtual sounds, and that sonic links result in more augmented experiences, both of which would be a boon to wider XR, though this will need confirmed through future work. Truly multimodal AR could also result in more accessible XR applications. Visual AR is inherently inaccessible for people who are visually impaired or blind, just as audio AR is inherently inaccessible for people who are deaf or hard of hearing. Combining both modalities could lead to more accessible XR applications overall.

Finally, there are some current limitations of acoustic and sonic links themselves which will need overcome in the future. AAR developers cannot reasonably expect users to model their own acoustic environments, and so a robust solution for reproducing environmental acoustics while minimising user effort will need identified to ensure high-quality AAR experiences in the future. Most likely, this will take the form of acoustic modelling simulations which are already the subject of much research [30, 146, 157] – many XR platforms now use LIDAR scanning or similar to map a user's physical surroundings and these models could be used to drive acoustic simulations in the future. Sonic links, particularly congruent ones, may also limit potential AAR applications to environments containing specific sounds. This is the primary reason this thesis suggests prioritising congruent links rather than making them a requirement of future AAR. However, they have also only been explored in three different application scenarios, and specifically as triggers for virtual effects. A sonic link could be as simple as analysing the real soundscape and presenting a virtual sound from the position where it will be most audible, and links like these need not respond to any specific sounds at all. Continued work on sonic linking will likely identify other opportunities to more deeply integrate the virtual sound world with the

real soundscape.

8.6 Concluding Thoughts

Until now, AAR applications have had no awareness or integration with the real-world soundscape they are designed to augment, a fundamental gap in AAR as a field that prevents AAR applications from reaching their full potential. This thesis introduces the concept of acoustic and sonic links, where AAR applications respond to the acoustic or sonic elements in a user's surroundings, as a way to close this gap and elevate the next generation of AAR applications. Through six mixed-methods user studies, novel acoustic and sonic links have been explored in controlled lab environments and real-world evaluation, providing a number of key insights that help to close this fundamental gap. The inclusion of acoustic links elevates the plausibility of the virtual sounds AAR systems introduce into our surroundings, while sonic links create more augmented experiences which alter the way we perceive the world around us. These links can be deployed today, with existing technology, to leverage these benefits. These findings provide a strong case for including acoustic and sonic links in the AAR applications of the future, as well as recommendations for how best to do so.

Appendix A

Data and Audio Files

A.1 Raw Datasets

- The raw datasets for Studies 1 and 2 presented in Chapter 4 are available at DOI: 10.5281/zenodo.10605358.
- The raw dataset for Study 3 presented in Chapter 5 is available at DOI: 10.5281/zenodo.1687852.
- The raw datasets for Studies 4, 5, and 6 presented in Chapters 6 and 7 are available at DOI: 10.5281/zenodo.19224819.

A.2 Audio Files

Relevant audio files for all six studies presented in this thesis are available at DOI: 10.5281/zenodo.19224388.

Appendix B

Experimental Materials

B.1 Study 1

B.1.1 Experimental Instructions

The following script was used to explain the structure of the study to participants:

Thank you for agreeing to participate in this study.

This study is seeking to evaluate how different reproductions and simulations of real-world acoustics affect a user's perception of virtual sounds in an audio augmented reality system. You'll be asked to listen to sounds under different reverberation conditions, and answer questions about your perception of the different acoustic reproductions. At the end of the session, there'll be a short interview about the experience which will be recorded for transcription.

Reverberation, what we're looking at in this study, is the way a sound behaves in a room. If we were having this exact conversation in a church, for example, you can imagine it would sound very different, it'd be a lot more 'echo-y'. That quality is reverberation. If you clap your hands, you'll hear the clap sound lingers for a brief moment, like a very short echo. That's reverberation too. Does that make sense?

Some of the sounds you will hear during the study have been recorded with the direct acoustics of this room. Some have not. Part of the experiment is answering whether you think a given sound was recorded in this room. Before each condition, a reference track will be played over this loudspeaker to allow you to hear the room's true acoustic response to the sounds.

The experiment itself takes place over a number of conditions.

Within each condition, you will listen to 3 blocks of 4 sounds. In each block, you will hear a sound, turn to face the position where you believe the sound to be, and press the spacebar. This is called 'localising' a sound. Once you've localised all the sounds in a block, you'll be asked to rate four acoustic qualities of those sounds, and then the next block will play. Once all three blocks are finished, there'll be a short questionnaire about the overall condition.

You'll be fitted with a head-tracker device which ensures that the sounds have a fixed real-

world position independent of your head movement, and which will track the direction you are facing for submitting your localisations.

The headtracker needs to be calibrated at regular intervals. We'll do this between conditions, and any time you feel the headtracker has drifted. The calibration process is very simple, you just need to look at this black circle here, and press the ENTER key on the keyboard.

This is an information sheet and a consent form. Please read them carefully, feel free to ask me any questions, and if you'd still like to participate, sign the consent form and we can begin.

If you're ready to start, please put on the headtracker, so that the cable hangs to your right. Press the 'Demographics' button on the screen and answer the demographics questions, then we can begin. Your participant number is —.

If A, headphones, if B, glasses

We'll be testing using the — first, please put them on, and press the — button. Then face the loudspeaker, press the ENTER key to calibrate the headtracker, and listen to the reference track

Ok, we're ready to begin. On the next screen, you'll hear a training sequence with its positions marked on the screen, and then the test will begin. You'll hear a sound, all you have to do is turn to face where you think that sound is, and press the SPACEBAR. Sounds may appear at any point around you. After four sounds, you'll rate some qualities of those sounds, and repeat until all three blocks are done.

Both your head and the loudspeaker will also be represented on the screen. If at any point you look at the loudspeaker and the screen doesn't match, look directly at this black circle and press the ENTER key to recalibrate things.

Press the button when you're ready to start.

B.1.2 Qualitative Interview Guide

The following questions were used to structure the qualitative interview:

- How did you find the experience of using these different playback devices? What did you like or dislike about them?
- We're conducting this research in part to explore audio augmented reality applications, those being applications where your real world auditory surroundings are augmented with virtual sounds. For example, when exploring a museum, you might hear a virtual tour guide explaining exhibits, you might receive auditory directions when navigating to somewhere new, or you might play an audio-based game, where the soundscape of a virtual world is overlaid on the real world. If you were using such an audio augmented reality application, which of the playback devices would you prefer to use to experience it? Why?
- Are there any scenarios in which you'd prefer to use the other playback device?

- How much did you feel you noticed the reverberation while playing? Did you notice differences in the reverberation when the conditions changed?
  - Did you notice a difference in ‘quality’? Were some better or worse than others or were they just ’different’?
- How much did you feel you relied on the reverberation to localise the sounds?
- Do you have any other comments on the different conditions you experienced in this study?

B.1.3 Information Sheet and Consent Form

The Influence of Environmental Acoustics on An Audio Augmented Reality System

CONSENT FORM

Before signing this consent form, you will have been given an information sheet explaining the study and what your participation will entail. Do not sign this consent form until you have read and fully understood the information on this sheet, and if you have not received this information sheet, please inform the researcher to be issued with one. The researcher will explain what is required of you throughout the experiment, but if at any point throughout the experiment you have a question, something is wrong, or you would like to end your participation in the experiment, please let the researcher know immediately.

The information provided in this consent form will be kept fully confidential. In order to qualify for participation in this study, you must have unimpaired hearing, and be able to use a computer screen without wearing glasses comfortably.

By signing this consent form, you assert that:

- You are at least 18 years old.
- You have unimpaired hearing.
- You can comfortably use a computer screen without wearing glasses.
- You have been issued with the information sheet for this study, read and understood it, and had an opportunity to ask the researcher any further questions.
- You understand your participation in this study is voluntary, and that your participation can be withdrawn at any time.
- You understand that any personal data recorded over the course of this study will be treated confidentially and kept securely (unless there is a legal requirement to disclose it to a third party), and that your research data will be anonymised.
- The researchers have permission to use any research data generated by your participation in future publications and research, and have permission to store that research data in online repositories, for example Glasgow University's *Enlighten* repository.

Researcher details: Jake Bhattacharyya (j.bhattacharyya.1@research.gla.ac.uk)

Supervisor details: Professor Stephen Brewster (stephen.brewster.glasgow.ac.uk)

This study has been approved by the Ethics Committee (Application Number 300220086).

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (if you would like to be added to the participants mailing list for future experiments):

_____

The Influence of Environmental Acoustics on An Audio Augmented Reality System

PARTICIPANT INFORMATION SHEET

Thank you for your interest in participating in this study. This sheet contains further information about the study, and what your participation will entail. Please read this sheet carefully, and contact the researcher if you have any further questions.

What is the purpose of this study?

This study seeks to investigate the simulation of environmental acoustics, and how differing levels of fidelity affect the user experience of an audio augmented reality application. A test space within the university campus has had its acoustics modelled, to investigate what effect the modelling of environmental acoustics has on a user's perception of virtual sounds, and how different models can be harnessed for the 'augmentation' of a user's auditory surroundings.

Do I have to take part?

No. Your participation in this study is entirely voluntary, and can be withdrawn at any time, and without giving a reason. We will ask you to sign a consent form if you choose to participate, but this will not affect your ability to withdraw your participation.

To qualify for participation in this study, you must have unimpaired hearing, and be able to use a computer screen comfortably without wearing glasses for approximately 30 minutes.

What will happen to me if I take part?

You'll be taken to a test space within the University, and asked to perform a simple listening and sound localisation task (indicating where you perceive the location of a sound to be) with differing types of reverberation applied. Performance metrics for this task will be recorded, and you will be asked to rate other acoustic characteristics of the sounds. You'll be asked to complete this task using headphones and audio glasses.

At the end of the study, you'll be asked to participate in a short interview about your experience. An audio recording of this interview will be made for transcription and future analysis.

What are the possible disadvantages and risks of taking part?

It is possible that the sounds in the listening task could be uncomfortably or dangerously loud, although their loudness has been calibrated to safe standards before the experiment.

You will be afforded the opportunity to set the playback to a comfortable level, and may adjust the playback level throughout the experiment.

What are the possible benefits of taking part?

By providing data on the effects of environmental acoustic modelling, your participation will help the researcher identify possibilities within the field of audio augmented reality, providing a perceptual grounding for further work in that space.

Additionally, you will be compensated for participation with a £10 Amazon voucher.

The Influence of Environmental Acoustics on An Audio Augmented Reality System

Will my taking part in this study be kept confidential?

Yes. A small amount of standard demographic information deemed relevant to the study will be recorded, and the only other personal information recorded in this study will be your signature on your consent form, the audio from your interview (which will be transcribed and the recording deleted), and your contact details (for arranging participation, dispensing of voucher, and any follow-up communications relevant to your data or participation), which will all be kept strictly confidential. Your data within the final results will be identified only with an alphanumeric identifier.

This confidentiality will be strictly maintained according to the GDPR, except under exceptional circumstances, for example the prevention of serious harm.

If you wish, you may opt-in to a mailing list for future experiments conducted within the School of Computing Science. If you choose to opt-in, your chosen email address will be recorded and added to the mailing list.

What will happen to the results of the research study?

The results will be held by the researcher and the University of Glasgow, and they may be used for future publication, or shared with other research partners on the SONICOM project. They may also be uploaded to an online repository, such as the University of Glasgow's "Enlighten" repository.

Your data will be kept anonymous in any publication.

Who is organising and funding the research?

This research is funded by the SONICOM research project.

Who has reviewed the study?

The project has been reviewed by the College Ethics Committee (Application Number: 300220086).

Contact for Further Information:

If you have any further questions, you can use the contact emails below:

Researcher: j.bhattacharyya.1@research.gla.ac.uk

Supervisor: stephen.brewster@glasgow.ac.uk

Thank you for your interest in participating in this study.

B.2 Study 2

B.2.1 Experimental Instructions

The following script was used to explain the structure of the study to participants:

Thank you for agreeing to participate in this study.

This study is seeking to evaluate how different reproductions and simulations of real-world acoustics affect a user's perception of virtual sounds in an audio augmented reality system. Today, you'll be asked to play a short game, answer some questions about the sounds you heard during the game, and then compare some different acoustic reproductions. At the end of the session, there'll be a short interview about the experience, which will be recorded for transcription.

The game is very simple. You are what is called a "sonomancer", a modern day wizard attuned specially to the world of sound. As a sonomancer, you protect our world from the monsters lurking in a parallel universe. These monsters occasionally make their way into our world through a sonic rift, and it just so happens that one of these rifts is here on campus. These monsters exist only in sound, so sonomancers are our only protection against them, and we need you to banish these sonic monsters and save our world.

All you'll have to do is listen carefully for the sounds the monster makes, and turn to face where you think it is. Then, you'll pull the right trigger on this controller. If you found the monster, you'll banish it! Hooray! If you didn't find the monster, it will instead attack you. Your health is shown in the green bar at the top right of the screen. Don't let it hit zero!

Once there are no more monsters left, you'll be asked a couple of quick questions about the sounds you heard during the game. After that, you'll be asked to compare some different test stimuli, which each have different acoustics applied to them.

One of the measures we're looking at in this study is something called **plausibility.** Some questions will ask you to rate the plausibility of a sound. In this study, when we say plausibility, we mean how closely a sound matches your expectation of how it should sound in the environment you're in. A sound which appears exactly as you'd expect it to, if it were really in the space with you, would have a high level of plausibility, and a sound which sounds very different from your expectations would have a low level of plausibility. Does that make sense?

The study takes place over four conditions, two in the Cloisters, and two in the quads. In each space you'll also be evaluating two different audio displays – headphones, and audio enabled sunglasses. In each condition, you'll play three rounds of the game, followed by the ratings.

Do you have any questions?

This is an information sheet and a consent form. Please read them carefully, feel free to ask me any questions, and if you'd still like to participate, sign the consent form and we can begin.

If you're ready to start, please put on the headtracker, so that the cable hangs to your right. Press the 'Demographics' button on the screen and answer the demographics questions, then we can begin. Your participant number is —.

Ok, we're ready to begin. If you press the Tutorial button, we'll give you a short trial run at the game and the evaluation questions before the study itself. You'll hear a noise in front of you, to your right, behind you, and then to your left, to allow you to hear how the spatial position of a sound affects it. Then wait and listen for the monster, turn to face where you think it is, and pull the right trigger.

Listening Test Instructions

This is called a MUSHRA test. Each slider here corresponds to a non-game sound stimulus, which you can listen to by pressing the Play button. Listen to each stimulus, and rate them from 0 to 100 according to the question at the top of the screen. You can switch between stimuli while they're playing by pressing play on another stimuli. Press the Submit button once you've listened to and rated all the stimuli.

B.2.2 Qualitative Interview Guide

The following questions were used to structure the qualitative interview:

- How did you find the experience of using these different playback devices? What did you like or dislike about them?
- If you had to choose, which of the two playback devices would you use for playing an audio-only game like this in the future? Why?
- Over the course of the study you experienced reverbs that ranged in their complexity. Did you feel they also ranged in 'quality'? Do you recall noticing particularly good or bad conditions?
- How much do you feel the reverberation contributed to your game experience?
  - Did that change with different reverberations, or was there not much difference between reverberations?
- Do you have any other comments on the different conditions you experienced in this study?

B.2.3 Information Sheet and Consent Form

SONOMANCER: The Influence of Environmental Acoustics on An Audio Augmented Reality System

CONSENT FORM

Before signing this consent form, you will have been given an information sheet explaining the study and what your participation will entail. Do not sign this consent form until you have read and fully understood the information on this sheet, and if you have not received this information sheet, please inform the researcher to be issued with one. The researcher will explain what is required of you throughout the experiment, but if at any point throughout the experiment you have a question, something is wrong, or you would like to end your participation in the experiment, please let the researcher know immediately.

The information provided in this consent form will be kept fully confidential. In order to qualify for participation in this study, you must have unimpaired hearing, and be able to use a computer screen without wearing glasses comfortably.

By signing this consent form, you assert that:

- You are at least 18 years old.
- You have unimpaired hearing.
- You can comfortably use a computer screen without wearing glasses.
- You have been issued with the information sheet for this study, read and understood it, and had an opportunity to ask the researcher any further questions.
- You understand your participation in this study is voluntary, and that your participation can be withdrawn at any time.
- You understand that any personal data recorded over the course of this study will be treated confidentially and kept securely (unless there is a legal requirement to disclose it to a third party), and that your research data will be anonymised.
- The researchers have permission to use any research data generated by your participation in future publications and research, and have permission to store that research data in online repositories, for example Glasgow University's *Enlighten* repository.

Researcher details: Jake Bhattacharyya (j.bhattacharyya.1@research.gla.ac.uk)

Supervisor details: Professor Stephen Brewster (stephen.brewster.glasgow.ac.uk)

This study has been approved by the Ethics Committee (Application Number 300220086).

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (if you would like to be added to the Computer Science participants mailing list for future experiments):

_____

SONOMANCER: The Influence of Environmental Acoustics on An Audio Augmented Reality System

PARTICIPANT INFORMATION SHEET

Thank you for your interest in participating in this study. This sheet contains further information about the study, and what your participation will entail. Please read this sheet carefully, and contact the researcher if you have any further questions.

What is the purpose of this study?

This study seeks to investigate the simulation of environmental acoustics, and how differing levels of fidelity affect the user experience of an audio augmented reality application. Test spaces within the university campus have had their acoustics modelled, to investigate what effect the modelling of environmental acoustics has on a user's perception of virtual sounds, and how different models can be harnessed for the 'augmentation' of a user's auditory surroundings.

Do I have to take part?

No. Your participation in this study is entirely voluntary, and can be withdrawn at any time, and without giving a reason. We will ask you to sign a consent form if you choose to participate, but this will not affect your ability to withdraw your participation.

To qualify for participation in this study, you must have unimpaired hearing, and be able to use a computer screen comfortably without wearing glasses for approximately 30 minutes.

What will happen to me if I take part?

You'll be taken to two test spaces within the University (the Cloisters and the East Quad), and asked to play a simple sound localisation game (indicating where you perceive the location of a sound to be), with differing types of reverberation applied. Performance metrics for this task will be recorded, and you will be asked to rate acoustic characteristics of the game sounds, and other test stimuli. You'll be asked to complete this task using headphones and audio glasses.

At the end of the study, you'll be asked to participate in a short interview about your experience. An audio recording of this interview will be made for transcription and future analysis.

What are the possible disadvantages and risks of taking part?

It is possible that the sounds in the listening task could be uncomfortably or dangerously loud, although their loudness has been calibrated to safe standards before the experiment.

You will be afforded the opportunity to set the playback to a comfortable level, and may adjust the playback level throughout the experiment.

What are the possible benefits of taking part?

By providing data on the effects of environmental acoustic modelling, your participation will help the researcher identify possibilities within the field of audio augmented reality, providing a perceptual grounding for further work in that space.

Additionally, you will be compensated for participation with a £10 Amazon voucher.

SONOMANCER: The Influence of Environmental Acoustics on An Audio Augmented Reality System

Will my taking part in this study be kept confidential?

Yes. A small amount of standard demographic information deemed relevant to the study will be recorded, and the only other personal information recorded in this study will be your signature on your consent form, the audio from your interview (which will be transcribed and the recording deleted), and your contact details (for arranging participation, dispensing of voucher, and any follow-up communications relevant to your data or participation), which will all be kept strictly confidential. Your data within the final results will be identified only with an alphanumeric identifier.

This confidentiality will be strictly maintained according to the GDPR, except under exceptional circumstances, for example the prevention of serious harm.

If you wish, you may opt-in to a mailing list for future experiments conducted within the School of Computing Science. If you choose to opt-in, your chosen email address will be recorded and added to the mailing list.

What will happen to the results of the research study?

The results will be held by the researcher and the University of Glasgow, and they may be used for future publication, or shared with other research partners on the SONICOM project. They may also be uploaded to an online repository, such as the University of Glasgow's "Enlighten" repository.

Your data will be kept anonymous in any publication.

Who is organising and funding the research?

This research is funded by the SONICOM research project.

Who has reviewed the study?

The project has been reviewed by the College Ethics Committee (Application Number: 300220086).

Contact for Further Information:

If you have any further questions, you can use the contact emails below:

Researcher: j.bhattacharyya.1@research.gla.ac.uk

Supervisor: stephen.brewster@glasgow.ac.uk

Thank you for your interest in participating in this study.

B.3 Study 3

B.3.1 Experimental Instructions

The following script was used to explain the structure of the study to participants:

First of all, thanks very much for agreeing to come down and participate! Basically, what we're looking at today are audio augmented reality games, and specifically, different ways you could control them, and how that affects the experience. The study itself is very simple: you're going to be asked to play three different audio-only games wearing this augmented reality headset, and you'll play four rounds of each game, using a different control mechanism for each. Between rounds, I'll ask you to answer some questions about the experience, and at the end I'll ask you some informal questions in a recorded interview, if that's ok.

The premise of all three games is that you are a sonomancer: a sound wizard, and you need to use your magic to protect our world from the sonic monsters that are threatening it. You'll do that by fighting the monsters in one game, finding and reactivating magical protections in another game, and awakening the ghosts of ancient wizards in the third game. In each game, we'll be evaluating a non-sound control method – either moving around, using a gesture, or using a controller – and three sound-based control methods: playing music on this xylophone, speech commands, and using 'sonic gestures', such as clapping your hands, humming, or clicking your fingers. For all three games we'll put you in this headset, but you'll still be able to see your surroundings through the cameras on the front. Do you have any questions right now?

Great, in that case I'll give you this information sheet, and this consent form. If you can read those both over, feel free to ask me any questions, and then sign on the line for a real good time, we can get going!

Excellent. So as I said we'll be playing three games today, with four rounds for each game. Except for one round which I'll explain when we come to it, your task in each round is to listen for a specific sound, either a monster in the combat game, the sound of a magical ward in the search game, and the sound of a wizardly ghost in the story game. You'll need to turn on the spot to face the sound, and then use the controls to do...whatever it is you need to do with it! I'll give you instructions for each round, though, so don't worry.

Combat Traditional: In this round you'll be fighting a monster. You'll need to listen for where it is, turn to face it, and then swing your controller in a diagonal motion to banish it. If you're accurate, you'll banish it, if not, it'll attack you instead!

Combat Musical: In this round you'll be fighting a monster. You'll need to listen for where it is, turn to face it, and then play a C, one of the blue notes on the xylophone, to banish it. If you're accurate, you'll banish it, if not, it'll attack you instead!

Combat Sonic Gesture: In this round you'll be fighting a monster. You'll need to listen for where it is, turn to face it, and then snap your fingers to banish it. If you're accurate, you'll banish it, if not, it'll attack you instead!

Combat Speech: In this round you'll be fighting a monster. You'll need to listen for where it is, turn to face it, and then say 'Alakazam!' to banish it. If you're accurate, you'll banish it, if not, it'll attack you instead!

Search Traditional: In this round, you'll be trying to find a sonomantic ward. This round is different from the others because we'll be using movement as our control. The ward will be hidden at first, but if you stand still you'll be able to hear it sound briefly. Once you've located it, you'll need to walk in a circle around where you think it is to reactivate it.

Search Music: In this round, you'll be trying to find a sonomantic ward. The ward will be hidden at first, but if you play an F, the orange key, you'll reveal it temporarily. Then turn to face where you think it is, and play four notes: CAGE, or blue, purple, red, yellow. That will reactivate it if you've tracked it down correctly.

Search Speech: In this round, you'll be trying to find a sonomantic ward. The ward will be hidden at first, but if you say 'Reveal!', you'll reveal it temporarily. Then turn to face where you think it is, and say 'I awaken this ward!' That will reactivate it if you've tracked it down correctly.

Search Sonic Gesture: In this round, you'll be trying to find a sonomantic ward. The ward will be hidden at first, but if you snap your fingers, you'll reveal it temporarily. Then turn to face where you think it is, and hum for a few seconds, making sure it's at a decent volume. That will reactivate it if you've tracked it down correctly.

Story Traditional: In this round, you'll be trying to find a ghost of an ancient sonomancer. Listen for where the ghost is, turn to face it, and then hold the trigger on the controller to awaken them. If you've located it correctly, they'll reawaken and speak to you.

Story Music: In this round, you'll be trying to find a ghost of an ancient sonomancer. Listen for where the ghost is, turn to face it, and then play the notes DEAD, or green, yellow, purple, green. If you've located it correctly, they'll reawaken and speak to you.

Story Speech: In this round, you'll be trying to find a ghost of an ancient sonomancer. Listen for where the ghost is, turn to face it, and then say 'Ancient spirit, I call you forth!'. If you've located it correctly, they'll reawaken and speak to you.

Story Sonic Gesture: In this round, you'll be trying to find a ghost of an ancient sonomancer. Listen for where the ghost is, turn to face it, and then hum for a few seconds at a good volume. If you've located it correctly, they'll reawaken and speak to you.

B.3.2 Full Quantitative Measures

The following questionnaire was presented to participants after each evaluation scenario:

Question	Low Anchor	High Anchor	# Scale Steps
PXI			
"Playing the game was meaningful to me."	Strongly Disagree	Strongly Agree	7
"The game felt relevant to me."	Strongly Disagree	Strongly Agree	7
"Playing this game was valuable to me."	Strongly Disagree	Strongly Agree	7
"I was no longer aware of my surroundings while I was playing."	Strongly Disagree	Strongly Agree	7
"I was immersed in the game."	Strongly Disagree	Strongly Agree	7
"I was fully focused on the game."	Strongly Disagree	Strongly Agree	7
"The game was not too easy and not too hard to play."	Strongly Disagree	Strongly Agree	7
"The game was challenging but not too challenging."	Strongly Disagree	Strongly Agree	7
"The challenges in the game were at the right level of difficulty for me."	Strongly Disagree	Strongly Agree	7
"It was easy to know how to perform actions in the game."	Strongly Disagree	Strongly Agree	7
"The actions to control the game were clear to me."	Strongly Disagree	Strongly Agree	7
"I thought the game was easy to control."	Strongly Disagree	Strongly Agree	7
"I liked playing the game."	Strongly Disagree	Strongly Agree	7
"The game was entertaining."	Strongly Disagree	Strongly Agree	7
"I had a good time playing this game."	Strongly Disagree	Strongly Agree	7
NASA TLX			
How mentally demanding was the task?	Very Low	Very High	21
How physically demanding was the task?	Very Low	Very High	21
How hurried or rushed was the pace of the task?	Very Low	Very High	21
How successful were you in accomplishing what you were asked to do?	Perfect	Failure	21
How hard did you have to work to accomplish your level of performance?	Very Low	Very High	21
How insecure, discouraged, irritated, stressed, and annoyed were you?	Very Low	Very High	21

B.3.3 Qualitative Interview Guide

The following questions were used to structure the qualitative interview:

- First of all, do you have any questions for me about the study?
- Overall, what were your impressions of the different games you played? Were there any games that particularly stood out, either for good reasons or bad reasons?
- We're particularly focused on the input methods for controlling and interacting with an audio AR game. You used a few today - movement, controls, physical gesture, music, speech, and sonic gesture. What were your impressions of those different input methods? What did you like or dislike about them? Were any better or worse than others?
- Today we were playing the games in a public space. Do you think your opinions on the games or the controls you used would be different if you were playing in other places, for example at home or in a private space?
- Imagine you were playing an audio augmented reality game in the future, where you're able to hear virtual sounds blended in with your real auditory surroundings. How do you imagine playing it? Would you want to use one of the control schemes you used today, or something else? What does your ideal control scheme look like?

B.3.4 Information Sheet and Consent Form

Evaluating Gameplay Control Methods for Audio Augmented Reality Games

CONSENT FORM

Before signing this consent form, you will have been given an information sheet explaining the study and what your participation will entail. Do not sign this consent form until you have read and fully understood the information on this sheet, and if you have not received this information sheet, please inform the researcher to be issued with one. The researcher will explain what is required of you throughout the experiment, but if at any point throughout the experiment you have a question, something is wrong, or you would like to end your participation in the experiment, please let the researcher know immediately.

The information provided in this consent form will be kept fully confidential. In order to qualify for participation in this study, you must have a normal level of hearing.

By signing this consent form, you assert that:

- You are at least 18 years old.
- You have a normal level of hearing.
- You have been issued with the information sheet for this study, read and understood it, and had an opportunity to ask the researcher any further questions.
- You understand your participation in this study is voluntary, and that your participation can be withdrawn at any time.
- You understand that any personal data recorded over the course of this study will be treated confidentially and kept securely (unless there is a legal requirement to disclose it to a third party), and that your research data will be anonymised.
- The researchers have permission to use any research data generated by your participation in future publications and research, and have permission to store that research data in online repositories, for example Glasgow University's *Enlighten* repository.

Researcher details: Jake Bhattacharyya (j.bhattacharyya.1@research.gla.ac.uk)

Supervisor details: Professor Stephen Brewster (stephen.brewster.glasgow.ac.uk)

This study has been approved by the Ethics Committee (300230087).

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (if you would like to be added to the participants mailing list for future experiments):

_____

Evaluating Gameplay Control Methods for Audio Augmented Reality Games

PARTICIPANT INFORMATION SHEET

Thank you for your interest in participating in this study. This sheet contains further information about the study, and what your participation will entail. Please read this sheet carefully, and contact the researcher if you have any further questions.

What is the purpose of this study?

This study seeks to evaluate different gameplay control methods for audio augmented reality games, how they affect the user experience of an audio augmented reality game, and for what forms of audio augmented reality game they might be best-suited for. In particular, this study investigates control methods which are inherently sonic, such as voice control, musical control, and sonic gesture, and how these compare to control methods which are not sound-based.

Do I have to take part?

No. Your participation in this study is entirely voluntary, and can be withdrawn at any time, and without giving a reason. We will ask you to sign a consent form if you choose to participate, but this will not affect your ability to withdraw your participation.

What will happen to me if I take part?

You'll be asked to play multiple rounds of a simple audio augmented reality game. The games you play during this study will ask you to use different methods to play and control them. Performance metrics for these games will be recorded, and you will be presented with a set of questions to evaluate your experience with the given control method for each game.

What are the possible disadvantages and risks of taking part?

It is possible that the sounds in the game could be uncomfortably or dangerously loud, although their loudness has been calibrated to safe standards before the experiment.

You will be afforded the opportunity to set the playback to a comfortable level ahead of time, and may adjust the playback level throughout the experiment.

Evaluating Gameplay Control Methods for Audio Augmented Reality Games

What are the possible benefits of taking part?

By providing data to evaluate these different control methods, your participation will help provide guidance for development of improved audio augmented reality games.

Additionally, you will be compensated for participation with a £10 Amazon voucher.

Will my taking part in this study be kept confidential?

Yes. A small amount of standard demographic information deemed relevant to the study will be recorded, and the only other personal information recorded in this study will be your signature on your consent form, and your contact details (for arranging participation and dispensing of voucher), which will both be kept strictly confidential. Your data within the final results will be identified only with a numeric identifier.

This confidentiality will be strictly maintained according to the GDPR, except under exceptional circumstances, for example the prevention of serious harm.

If you wish, you may opt in to a mailing list for future experiments conducted within the School of Computing Science. If you choose to opt in, your chosen email address will be recorded and added to the mailing list.

What will happen to the results of the research study?

The results will be held by the researcher and the University of Glasgow, and they may be used for future publication, or shared with other research partners on the SONICOM project. They may also be uploaded to an online repository, such as the University of Glasgow's "Enlighten" repository.

Your data will be kept anonymous in any publication.

Who is organising and funding the research?

This research is funded by the SONICOM research project.

Who has reviewed the study?

The project has been reviewed by the College Ethics Committee (Application Number: 300230087).

Contact for Further Information:

If you have any further questions, you can use the contact emails below:

Researcher: j.bhattacharyya.1@research.gla.ac.uk

Supervisor: stephen.brewster@glasgow.ac.uk

Thank you for your interest in participating in this study.

B.4 Study 4

B.4.1 Experimental Instructions

The following script was used to explain the structure of the study to participants:

First of all, thanks very much for agreeing to come and participate! Today, what we're looking at are audio augmented reality applications, so applications that augment your reality with sound. Specifically, the applications we're looking at today are applications that are also aware of the sounds around you, so the application might behave differently if there are birds nearby, or it hears a car go past, something like that.

The study itself is very simple. We have three different applications – a game, a music player, and a notification system – and each reacts to real sounds around you, most commonly birds. You're going to use each application twice, as we have some variations on each which we're interested in. You'll use the application for five minutes, just walking about this area we're in now and listening to the applications react to any birdsong nearby, and then fill out a questionnaire. I'll also be asking you some informal questions between applications to get your opinion on the design of the applications, and how you might change them or design new applications like these to use in your own life. At the end of the session we'll have an overall interview, during which I'll also ask you what sort of sound-aware audio AR application you might want if you were to design one for your own use, so have a think about that throughout the session.

Does that all make sense so far? Do you have any questions right now?

Great, in that case I'll give you this information sheet, and this consent form. If you can read those both over, feel free to ask me any questions, and then sign on the line for a real good time, we can get going!

Game Instructions

The game scenario is similar to Pokemon Go – when the system detects birdsong, a 'sonimon' will appear nearby. You just have to turn to face where you think it is, and pull the trigger on your controller. If you miss, it might run away. Occasionally, a *monster* will appear too. You'll need to do the same thing, just turn on the spot until you're facing it and pull your trigger. The number of sonimon you've captured will affect how quickly you can destroy the monster.

Music Instructions

In the music scenario, you're going to listen to a Michael Jackson song, and it'll filter some of the music out if the system detects birdsong, or in the second mode if it detects a car nearby. You don't have to do anything but listen.

Notification Instructions

In the notification scenario, the system is going to listen for birdsong and supplement any it detects with additional virtual birds. These virtual birds will indicate information about your hypothetical Twitter feed. You'll see some sliders in front of you which you can adjust to change that hypothetical Twitter feed, to see how the system reacts to the number of direct messages you have waiting, the level of activity from people you follow, and the level of activity in the nearby area, which will each be indicated by different kinds of additional birds.

B.4.2 Full Quantitative Measures

The following questionnaire was presented to participants after each evaluation scenario:

Question	Low Anchor	High Anchor	# Scale Steps
UEQ			
This application was...	Obstructive	Supportive	7
	Complicated	Easy	7
	Inefficient	Efficient	7
	Confusing	Clear	7
	Boring	Exciting	7
	Not Interesting	Interesting	7
	Conventional	Inventive	7
	Usual	Leading Edge	7
Real Sound Response			
The application responded to real world sounds around me.	Strongly Disagree	Strongly Agree	7
ARI - Engagement			
I liked the activity because it was novel.	Strongly Disagree	Strongly Agree	7
I liked the type of activity.	Strongly Disagree	Strongly Agree	7
I wanted to spend the time to complete the activity successfully.	Strongly Disagree	Strongly Agree	7
I wanted to spend time to participate in the activity.	Strongly Disagree	Strongly Agree	7
It was easy for me to use the AR application.	Strongly Disagree	Strongly Agree	7
I found the AR application confusing.	Strongly Disagree	Strongly Agree	7
The AR application was unnecessarily complex.	Strongly Disagree	Strongly Agree	7
I did not have difficulties in controlling the AR application.	Strongly Disagree	Strongly Agree	7
ARI - Engrossment			
I was curious about how the activity would progress.	Strongly Disagree	Strongly Agree	7
I was often excited since I felt as being part of the activity.	Strongly Disagree	Strongly Agree	7
I often felt suspense by the activity.	Strongly Disagree	Strongly Agree	7
If interrupted, I looked forward to returning to the activity.	Strongly Disagree	Strongly Agree	7
Everyday thoughts and concerns faded out during the activity.	Strongly Disagree	Strongly Agree	7
I was more focused on the activity rather than on any external distraction.	Strongly Disagree	Strongly Agree	7
Augmentation			
My world felt augmented.	Strongly Disagree	Strongly Agree	7
The virtual elements felt connected to the real world.	Strongly Disagree	Strongly Agree	7
If I used the application in a different environment, the application would have behaved differently.	Strongly Disagree	Strongly Agree	7
The virtual elements felt like they were part of the same world as the real elements.	Strongly Disagree	Strongly Agree	7
The application was ___ than the real or virtual elements would have been on their own.	Much Worse	Much Better	7

B.4.3 Qualitative Interview Guide

The following questions were used to structure the qualitative interviews:

Post-Application Interviews

- What did you think of the application? What did you like or dislike about it?
- We tried two different variations of the application. Did you have a preference for one of the variations? Why?
- What is the perfect version of this application for you? How would you change it to be something you would use in your own life?

Final Interview

- First of all, do you have any questions for me about the study?
- You used three applications today - a way of receiving notifications, a game, and a music player. Overall, what did you think about them?
- Of the three applications, which was your favourite? Which would you be most inclined to use?
- All the applications you used today were designed to respond to sounds in your real world surroundings, which is what we're focused on today. What are your thoughts on applications doing that? Does anything excite or concern you about that idea?
  - Would there be any circumstances in which you would seek out or avoid an application that did that?
- You experienced three different applications today, all of which were aware of sounds around you, and were audio-only. If you were designing a similar application for using in your own day-to-day life, one that presented audio to you and was aware of the sounds in your environment, what would you want it to do? Take some time to think about it, if you'd like.

B.4.4 Information Sheet and Consent Form

Evaluating Real-Virtual Links for Audio Augmented Reality Applications
CONSENT FORM

Before signing this consent form, you will have been given an information sheet explaining the study and what your participation will entail. **Do not sign this consent form until you have read and fully understood the information on this sheet**, and if you have not received this information sheet, please inform the researcher to be issued with one. The researcher will explain what is required of you throughout the experiment, but if at any point throughout the experiment you have a question, something is wrong, or you would like to end your participation in the experiment, please let the researcher know immediately.

The information provided in this consent form will be kept fully confidential. In order to qualify for participation in this study, **you must have a normal level of hearing.**

By signing this consent form, you assert that:

- You are at least 18 years old.
- You have a normal level of hearing.
- You have been issued with the information sheet for this study, read and understood it, and had an opportunity to ask the researcher any further questions.
- You understand your participation in this study is voluntary, and that your participation can be withdrawn at any time.
- You understand that any personal data recorded over the course of this study will be treated confidentially and kept securely (unless there is a legal requirement to disclose it to a third party), and that your research data will be anonymised.
- The researchers have permission to use any research data generated by your participation in future publications and research, and have permission to store that research data in online repositories, for example Glasgow University's *Enlighten* repository.

Researcher details: Jake Bhattacharyya (j.bhattacharyya.1@research.gla.ac.uk)

Supervisor details: Professor Stephen Brewster (stephen.brewster.glasgow.ac.uk)

This study has been approved by the Ethics Committee (Reference Number Pending).

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (if you would like to be added to the participants mailing list for future experiments):

_____

Evaluating Real-Virtual Links for Audio Augmented Reality Applications

PARTICIPANT INFORMATION SHEET

Thank you for your interest in participating in this study. This sheet contains further information about the study, and what your participation will entail. Please read this sheet carefully, and contact the researcher if you have any further questions.

What is the purpose of this study?

This study seeks to evaluate different ways an audio augmented reality application can be linked to the user's real world surroundings, and how that affects the user experience of the application. In particular, this study investigates sonic links, where the output of the application changes and reacts to the sounds in the user's environment.

This study is investigating these applications in a controlled environment, but we are also planning to run a follow-up study investigating how they work in the wider world. **When that study is running, we will send you another email asking if you'd be interested in participating in that also.** As in this study, participation is not mandatory, but would be very helpful for our data collection.

Do I have to take part?

No. Your participation in this study is entirely voluntary, and can be withdrawn at any time, and without giving a reason. We will ask you to sign a consent form if you choose to participate, but this will not affect your ability to withdraw your participation.

What will happen to me if I take part?

You'll be asked to experience a variety of different audio augmented reality applications, which draw on the sounds in your environment in some way. You'll experience an audio game, a notification system, and a system for listening to augmented music. You'll use these applications in a small park environment on University grounds for approximately one hour.

After experiencing each application, you'll be presented with a set of questions to evaluate your experience. At the conclusion of the study, you'll be asked to participate in a short interview with the researcher about your experience using each application.

The study will last approximately one hour.

What are the possible disadvantages and risks of taking part?

It is possible that the sounds in the game could be uncomfortably or dangerously loud, although their loudness has been calibrated to safe standards before the experiment.

You will be afforded the opportunity to set the playback to a comfortable level ahead of time, and may adjust the playback level throughout the experiment.

Evaluating Real-Virtual Links for Audio Augmented Reality Applications

What are the possible benefits of taking part?

By providing data to evaluate these different applications, your participation will help provide guidance for how these sonic connections can be used to develop improved audio augmented reality applications.

Additionally, you will be compensated for participation with a £10 Amazon voucher.

Will my taking part in this study be kept confidential?

Yes. A small amount of standard demographic information deemed relevant to the study will be recorded, and the only other personal information recorded in this study will be your signature on your consent form, and your contact details (for arranging participation and dispensing of voucher), which will both be kept strictly confidential. Your data within the final results will be identified only with a numeric identifier.

This confidentiality will be strictly maintained according to the GDPR, except under exceptional circumstances, for example the prevention of serious harm.

If you wish, you may opt in to a mailing list for future experiments conducted within the School of Computing Science. If you choose to opt in, your chosen email address will be recorded and added to the mailing list.

What will happen to the results of the research study?

The results will be held by the researcher and the University of Glasgow, and they may be used for future publication, or shared with other research partners on the SONICOM project. They may also be uploaded to an online repository, such as the University of Glasgow's "Enlighten" repository.

Your data will be kept anonymous in any publication.

Who is organising and funding the research?

This research is funded by the SONICOM research project.

Who has reviewed the study?

The project has been reviewed by the College Ethics Committee (Application Number: xxxxxxx).

Contact for Further Information:

If you have any further questions, you can use the contact emails below:

Researcher: j.bhattacharyya.1@research.gla.ac.uk

Supervisor: stephen.brewster@glasgow.ac.uk

Thank you for your interest in participating in this study.

B.5 Study 5

B.5.1 Experimental Instructions

The following script was used to explain the structure of the study to participants:

First of all, thanks very much for agreeing to come and participate! Today, what we're looking at are audio augmented reality applications, so applications that augment your reality with sound. Specifically, the applications we're looking at today are applications that are also aware of the sounds around you, so the application might behave differently if there are birds nearby, or it hears a car go past, something like that.

The study itself is very simple. We have two different applications – a game, and a music player – and each reacts to real sounds around you, most commonly birds. You're going to use each application three times, as we have some variations on each which we're interested in. You'll use each application for five minutes, just walking about this area we're in now and playing the game or listening to the music while the applications react to any sounds nearby, and then I'll ask you to fill out a questionnaire. At the end of the session we'll have an overall interview.

Does that all make sense so far? Do you have any questions right now?

Great, in that case I'll give you this information sheet, and this consent form. If you can read those both over, feel free to ask me any questions, and then sign on the line for a real good time, we can get going!

Game Instructions

The game scenario is similar to Pokemon Go – some of the birds in the area indicate 'sonimon', powerful sonic creatures that you can catch. When the system detects one of these birds, a 'sonimon' will appear nearby. You just have to turn to face where you think it is, and pull the trigger on your controller. If you miss, it might run away. Occasionally, a *monster* will appear too. You'll need to do the same thing, just turn on the spot until you're facing it and pull your trigger. The number of sonimon you've captured will affect how quickly you can destroy the monster, so you want to catch as many as possible.

- A: In this variation, birdsong will result in a bird sonimon.
- B: In this variation, birdsong will result in a dog sonimon.
- C: In this variation, you'll get a cat sonimon, but it won't be tied to any of the sounds in the environment.

Music Instructions

In the music scenario, you're going to listen to a Michael Jackson song, and it'll filter some of the music out if the system detects certain sounds. You don't have to do anything but listen.

- A: In this variation, the music will react to birdsong.
- B: In this variation, the music will react to car sounds.
- C: In this variation, the music won't react to anything.

B.5.2 Full Quantitative Measures

The following questionnaire was presented to participants after each evaluation scenario:

Question	Low Anchor	High Anchor	# Scale Steps
UEQ			
This application was...	Obstructive	Supportive	7
	Complicated	Easy	7
	Inefficient	Efficient	7
	Confusing	Clear	7
	Boring	Exciting	7
	Not Interesting	Interesting	7
	Conventional	Inventive	7
	Usual	Leading Edge	7
Real Sound Response			
The application responded to real world sounds around me.	Strongly Disagree	Strongly Agree	7
Real Sound Response Accuracy			
The application accurately detected the real world sounds around me	Strongly Disagree	Strongly Agree	7
ARI - Engagement			
I liked the activity because it was novel.	Strongly Disagree	Strongly Agree	7
I liked the type of activity.	Strongly Disagree	Strongly Agree	7
I wanted to spend the time to complete the activity successfully.	Strongly Disagree	Strongly Agree	7
I wanted to spend time to participate in the activity.	Strongly Disagree	Strongly Agree	7
It was easy for me to use the AR application.	Strongly Disagree	Strongly Agree	7
I found the AR application confusing.	Strongly Disagree	Strongly Agree	7
The AR application was unnecessarily complex.	Strongly Disagree	Strongly Agree	7
I did not have difficulties in controlling the AR application.	Strongly Disagree	Strongly Agree	7
ARI - Engrossment			
I was curious about how the activity would progress.	Strongly Disagree	Strongly Agree	7
I was often excited since I felt as being part of the activity.	Strongly Disagree	Strongly Agree	7
I often felt suspense by the activity.	Strongly Disagree	Strongly Agree	7
If interrupted, I looked forward to returning to the activity.	Strongly Disagree	Strongly Agree	7
Everyday thoughts and concerns faded out during the activity.	Strongly Disagree	Strongly Agree	7
I was more focused on the activity rather than on any external distraction.	Strongly Disagree	Strongly Agree	7
Augmentation			
My world felt augmented.	Strongly Disagree	Strongly Agree	7
The virtual elements felt connected to the real world.	Strongly Disagree	Strongly Agree	7
If I used the application in a different environment, the application would have behaved differently.	Strongly Disagree	Strongly Agree	7
The virtual elements felt like they were part of the same world as the real elements.	Strongly Disagree	Strongly Agree	7
The application was ___ than the real or virtual elements would have been on their own.	Much Worse	Much Better	7
Awareness			
How aware were you of the real world surroundings while navigating in the virtual world?	Extremely Aware	Not Aware At All	7
I was not aware of my real environment	Fully Disagree	Fully Agree	7
I still paid attention to the real environment	Fully Disagree	Fully Agree	7
I was completely captivated by the virtual world	Fully Disagree	Fully Agree	7
Real World Connection			
I felt connected to my real world surroundings	Strongly Disagree	Strongly Agree	7

B.5.3 Qualitative Interview Guide

The following questions were used to structure the qualitative interview:

- First of all, do you have any questions for me?
- You used two applications today: a game and a music player. Overall, what did you think of the applications you used today? What did you like or dislike about them?
- Of the applications you used, did you have a preferred application?
- The applications you used today often responded to real world sounds around you. How reliable did you feel they were at that?
- What are your thoughts on applications doing that? Do you like that idea or would you prefer they didn't?
- For both the game and the music, we tested some variations of the apps. The music responded to birds or cars, and the game gave you a bird sonimon or a not-bird sonimon. Did you prefer any of those variations?
  - Why?
- Are there any other sounds you'd like these applications to respond to that they don't currently?
- And we also tested variations that didn't respond to sounds. How did you like them compared to the sound-responsive ones?
- Are there any applications you can think of which aren't sound-reactive, but which you'd like to be?
- Any final comments? Any thoughts you'd like to add that we haven't already covered?

B.5.4 Information Sheet and Consent Form

Evaluating Real-Virtual Links for Audio Augmented Reality Applications

CONSENT FORM

Before signing this consent form, you will have been given an information sheet explaining the study and what your participation will entail. **Do not sign this consent form until you have read and fully understood the information on this sheet**, and if you have not received this information sheet, please inform the researcher to be issued with one. The researcher will explain what is required of you throughout the experiment, but if at any point throughout the experiment you have a question, something is wrong, or you would like to end your participation in the experiment, please let the researcher know immediately.

The information provided in this consent form will be kept fully confidential. In order to qualify for participation in this study, **you must have a normal level of hearing.**

By signing this consent form, you assert that:

- You are at least 18 years old.
- You have a normal level of hearing.
- You have been issued with the information sheet for this study, read and understood it, and had an opportunity to ask the researcher any further questions.
- You understand your participation in this study is voluntary, and that your participation can be withdrawn at any time.
- You understand that any personal data recorded over the course of this study will be treated confidentially and kept securely (unless there is a legal requirement to disclose it to a third party), and that your research data will be anonymised.
- The researchers have permission to use any research data generated by your participation in future publications and research, and have permission to store that research data in online repositories, for example Glasgow University's *Enlighten* repository.

Researcher details: Jake Bhattacharyya (j.bhattacharyya.1@research.gla.ac.uk)

Supervisor details: Professor Stephen Brewster (stephen.brewster.glasgow.ac.uk)

This study has been approved by the Ethics Committee (Reference Number Pending).

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (if you would like to be added to the participants mailing list for future experiments):

_____

Evaluating Real-Virtual Links for Audio Augmented Reality Applications

PARTICIPANT INFORMATION SHEET

Thank you for your interest in participating in this study. This sheet contains further information about the study, and what your participation will entail. Please read this sheet carefully, and contact the researcher if you have any further questions.

What is the purpose of this study?

This study seeks to evaluate different ways an audio augmented reality application can be linked to the user's real world surroundings, and how that affects the user experience of the application. In particular, this study investigates sonic links, where the output of the application changes and reacts to the sounds in the user's environment.

This study is investigating these applications in a controlled environment, but we are also planning to run a follow-up study investigating how they work in the wider world. **When that study is running, we will send you another email asking if you'd be interested in participating in that also.** As in this study, participation is not mandatory, but would be very helpful for our data collection.

Do I have to take part?

No. Your participation in this study is entirely voluntary, and can be withdrawn at any time, and without giving a reason. We will ask you to sign a consent form if you choose to participate, but this will not affect your ability to withdraw your participation.

What will happen to me if I take part?

You'll be asked to experience a variety of different audio augmented reality applications, which draw on the sounds in your environment in some way. You'll experience an audio game, and a system for listening to augmented music. You'll experience three variations on each application, for a total of six applications. You'll use these applications in a small park environment on University grounds for approximately one hour.

After experiencing each application, you'll be presented with a set of questions to evaluate your experience. At the conclusion of the study, you'll be asked to participate in a short interview with the researcher about your experience using each application.

The study will last approximately one hour.

What are the possible disadvantages and risks of taking part?

It is possible that the sounds in the game could be uncomfortably or dangerously loud, although their loudness has been calibrated to safe standards before the experiment.

You will be afforded the opportunity to set the playback to a comfortable level ahead of time, and may adjust the playback level throughout the experiment.

Evaluating Real-Virtual Links for Audio Augmented Reality Applications

What are the possible benefits of taking part?

By providing data to evaluate these different applications, your participation will help provide guidance for how these sonic connections can be used to develop improved audio augmented reality applications.

Additionally, you will be compensated for participation with a £10 Amazon voucher.

Will my taking part in this study be kept confidential?

Yes. A small amount of standard demographic information deemed relevant to the study will be recorded, and the only other personal information recorded in this study will be your details on your consent form, and your contact details (for arranging participation and dispensing of voucher), which will both be kept strictly confidential. Your data within the final results will be identified only with a numeric identifier.

This confidentiality will be strictly maintained according to the GDPR, except under exceptional circumstances, for example the prevention of serious harm.

If you wish, you may opt in to a mailing list for future experiments conducted within the School of Computing Science. If you choose to opt in, your chosen email address will be recorded and added to the mailing list.

What will happen to the results of the research study?

The results will be held by the researcher and the University of Glasgow, and they may be used for future publication, or shared with other research partners on the SONICOM project. They may also be uploaded to an online repository, such as the University of Glasgow's "Enlighten" repository.

Your data will be kept anonymous in any publication.

Who is organising and funding the research?

This research is funded by the SONICOM research project.

Who has reviewed the study?

The project has been reviewed by the College Ethics Committee (Application Number: xxxxxxx).

Contact for Further Information:

If you have any further questions, you can use the contact emails below:

Researcher: j.bhattacharyya.1@research.gla.ac.uk

Supervisor: stephen.brewster@glasgow.ac.uk

Thank you for your interest in participating in this study.

B.6 Study 6

B.6.1 Experimental Instructions

AAR Sonic Linking Study – Participant Handout

Thank you for agreeing to participate in our longitudinal AAR study. This document contains an overview of what we're asking of you over the course of the study.

- We are testing two applications: a game and a music player.
- These applications respond to real-world sounds, and so you **must use them outdoors**.
- You will test both applications for a **week**, using both applications on **Tuesday**, **Wednesday**, **Thursday**, and **Friday**.
- On these days, we need you to complete a session with each application in the **morning**, and with each application later in the day **(afternoon/evening).**
- Use each application **at least until it gives you a questionnaire**. You can use the applications for as long as you like, but you must complete **at least one questionnaire** for each application in the morning, and **at least one questionnaire** for each application later in the day.
- If you'd like, you may also use the applications on Saturday and Sunday. You are not required to.
- You'll also be emailed a **short set of questions each day to answer.**

If you don't complete the minimum usage and email interview each day, we won't be able to give you the £50 voucher.

You have been issued with a smartphone and a set of open-ear wireless earbuds. The smartphone comes pre-installed with the test application. The test application opens to a main menu where you can open the music player or Pokemon game.

Music Player

The music player functions like a standard music player. It comes presupplied with a selection of 10 albums. While listening to music through the player, the music will respond to birds, cars and speech, adjusting volume when it detects these sounds so you can hear them better.

Controls along the bottom of the player allow you to play, pause, and skip, as well as setting tracks and albums to loop, or to shuffle the music queue. These should be familiar from other music player applications like Spotify.

At the top right, there is a button to toggle between viewing a list of all tracks in the library, or viewing a list of albums which can be opened.

The player comes with music already, but you may also add your own music if you have digital music files (like MP3s). Connect the supplied device to a computer, and place your custom music into the 'Music' folder on the device. It should then be loaded into the app.

Pokemon Game

The supplied game is a Pokemon-style creature catching game, where you are able to catch audio-only pokemon, or "sonimon". These sonimon will appear when the phone detects the sound of birds or cars, with the type of sonimon that appears depending on the real world sound. There are 10 sonimon to collect, some rarer than others: 5 for birds, and 5 for cars.

Sometimes, these real sounds will result in a monster appearing instead. The monster needs to be defeated, or else it will eat your strongest sonimon!

When a sound is detected and a sonimon appears, it will sound as if it's coming from a specific direction around you. Hold the phone in front of you, and turn on the spot until you feel you are facing the sonimon. Then press the "Catch" button on the screen. If you find it successfully, the sonimon will be caught. If you don't, you get a few more attempts before the sonimon runs away. Remember, the sonimon are **audio-only**, so you won't see them.

The "Sonidex" button at the top of the screen allows you to view the list of sonimon you've caught over the course of the study, and their power. See how strong a sonimon you can catch!

Earbuds

It's very important that you only use the applications with the supplied earbuds. These earbuds clip on the edge of your ear, rather than sit inside it. This allows you to still hear the sounds around you.

The side with the big gold dot should go **behind** your ear.

Checklist

Tuesday

- [ ] **Morning:** Game at least until one questionnaire completed
- [ ] **Morning:** Music at least until one questionnaire completed
- [ ] **Evening:** Game at least until one questionnaire completed
- [ ] **Evening:** Music at least until one questionnaire completed
- [ ] Email interview

Wednesday

- [ ] **Morning:** Game at least until one questionnaire completed
- [ ] **Morning:** Music at least until one questionnaire completed
- [ ] **Evening:** Game at least until one questionnaire completed
- [ ] **Evening:** Music at least until one questionnaire completed
- [ ] Email interview

Thursday

- [ ] **Morning:** Game at least until one questionnaire completed
- [ ] **Morning:** Music at least until one questionnaire completed
- [ ] **Evening:** Game at least until one questionnaire completed
- [ ] **Evening:** Music at least until one questionnaire completed
- [ ] Email interview

Friday

- [ ] **Morning:** Game at least until one questionnaire completed
- [ ] **Morning:** Music at least until one questionnaire completed
- [ ] **Evening:** Game at least until one questionnaire completed
- [ ] **Evening:** Music at least until one questionnaire completed
- [ ] Email interview

Following Monday

- [ ] Return hardware
- [ ] Final interview

B.6.2 Qualitative Interview Guide

The following questions were used to structure the qualitative interviews:

Daily Email Interview

Hi!

You're receiving this email as a daily reminder to try out both the music and game AAR applications. Once you've used each for the day, please reply to this email and jot down a couple of quick reflections on these questions. Thanks!

- Where did you use the apps today?
- How reliable did you find the applications today?
- Did you have any frustrations with the apps?
- How was the experience compared to a normal music player or game?
- Was your experience of your environment different in any way? Did you behave any differently?
- How was the experience today compared to yesterday?

Thanks! You'll receive another reminder email like this tomorrow.

– Jake

Final Interview

- Overall, what did you think of the music app? Anything particularly positive or negative?
- And the game? Anything that stood out as positive or negative?
- How did the experience change over the week?
- Did you use the apps over the weekend? Why / why not?
- Having now used sonic linking apps for an extended period, how do you feel about them? Any change in opinion?
- Are there any other sounds you'd like the apps to react to?
- Is there anything about the music app that you would like to change? Are there any features you would like it to have?
- And what about the game? Anything you would change? Any features you would like it to have?

- If these apps were available on your phone tomorrow, would you use them? Why / why not?

B.6.3 Information Sheet and Consent Form

Evaluating Real-Virtual Links for Audio Augmented Reality Applications
CONSENT FORM

Before signing this consent form, you will have been given an information sheet explaining the study and what your participation will entail. **Do not sign this consent form until you have read and fully understood the information on this sheet**, and if you have not received this information sheet, please inform the researcher to be issued with one. The researcher will explain what is required of you throughout the experiment, but if at any point throughout the experiment you have a question, something is wrong, or you would like to end your participation in the experiment, please let the researcher know immediately.

The information provided in this consent form will be kept fully confidential. In order to qualify for participation in this study, **you must have a normal level of hearing.**

By signing this consent form, you assert that:

- You are at least 18 years old.
- You have a normal level of hearing.
- You have been issued with the information sheet for this study, read and understood it, and had an opportunity to ask the researcher any further questions.
- You understand your participation in this study is voluntary, and that your participation can be withdrawn at any time.
- You understand that any personal data recorded over the course of this study will be treated confidentially and kept securely (unless there is a legal requirement to disclose it to a third party), and that your research data will be anonymised.
- The researchers have permission to use any research data generated by your participation in future publications and research, and have permission to store that research data in online repositories, for example Glasgow University's *Enlighten* repository.

Researcher details: Jake Bhattacharyya (j.bhattacharyya.1@research.gla.ac.uk)

Supervisor details: Professor Stephen Brewster (stephen.brewster.glasgow.ac.uk)

This study has been approved by the Ethics Committee (Reference Number Pending).

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (if you would like to be added to the participants mailing list for future experiments):

_____

Evaluating Real-Virtual Links for Audio Augmented Reality Applications – Long-Term Study

PARTICIPANT INFORMATION SHEET

Thank you for your interest in participating in this study. This sheet contains further information about the study, and what your participation will entail. Please read this sheet carefully, and contact the researcher if you have any further questions.

What is the purpose of this study?

This study seeks to evaluate different ways an audio augmented reality application can be linked to the user's real world surroundings, and how that affects the user experience of the application. In particular, this study investigates sonic links, where the output of the application changes and reacts to the sounds in the user's environment.

This study is investigating these applications "in-the-wild", evaluating how they work in the real world and as part of day-to-day life.

Do I have to take part?

No. Your participation in this study is entirely voluntary, and can be withdrawn at any time, and without giving a reason. We will ask you to sign a consent form if you choose to participate, but this will not affect your ability to withdraw your participation.

What will happen to me if I take part?

You'll be supplied with two audio AR applications – a music player and a game - and the necessary equipment to use them, for one week. Both applications will listen to the sounds in your environment, and respond to specific sounds they detect, such as vehicles, speech and birdsong. You'll be issued the equipment on a **Monday,** and asked to use each application **morning and afternoon** on **Tuesday, Wednesday, Thursday, and Friday**, in the circumstances you would normally consider using them, for example when commuting to work or walking. You'll return the equipment the following **Monday**. There is no requirement to use the applications on Saturday and Sunday, though you are welcome to if you would like.

It is important to only use the application when it is safe to do so. Always pay attention to your surroundings when using the applications.

At regular intervals, the application will present you with a short questionnaire about your experience and how the application is performing. On any given day, you only have to use the applications until you've completed one of these questionnaires for the game, and one for the music player in the morning, and the same again in the afternoon/evening. You are welcome to use either application beyond that if you would like. Each day, you'll be emailed a further short set of questions to answer freely about the experience.

At the conclusion of the study, you'll be asked to participate in an interview with the researcher about your experience using the application.

When in use, the application will also record information about your context when it responds to nearby sounds. This information consists of the **music currently playing**, the **sounds nearby**, and

Evaluating Real-Virtual Links for Audio Augmented Reality Applications – Long-Term Study

your **geographical location**. **The application only records this information when it is running and responding to nearby sounds.**

What are the possible disadvantages and risks of taking part?

It is possible that the sounds you hear could be uncomfortably or dangerously loud, although their loudness has been calibrated to safe standards before the experiment.

You may adjust the playback level freely when using the application.

What are the possible benefits of taking part?

By providing data to evaluate the application your participation will help provide guidance for how these sonic connections can be used to develop improved audio augmented reality applications.

Additionally, you will be compensated for participation with an Amazon voucher.

Will my taking part in this study be kept confidential?

Yes. A small amount of standard demographic information deemed relevant to the study will be recorded, and the only other personal information recorded in this study will be your contact details (for arranging participation and dispensing of voucher), which will be kept strictly confidential. Your data within the final results will be identified only with a numeric identifier.

This confidentiality will be strictly maintained according to the GDPR, except under exceptional circumstances, for example the prevention of serious harm.

If you wish, you may opt in to a mailing list for future experiments conducted within the School of Computing Science. If you choose to opt in, your chosen email address will be recorded and added to the mailing list.

What will happen to the results of the research study?

The results will be held by the researcher and the University of Glasgow, and they may be used for future publication, or shared with other research partners on the SONICOM project. They may also be uploaded to an online repository, such as the University of Glasgow's "Enlighten" repository.

Your data will be kept anonymous in any publication.

Who is organising and funding the research?

This research is funded by the SONICOM research project.

Who has reviewed the study?

The project has been reviewed by the College Ethics Committee (Application Number: xxxxxxx).

Evaluating Real-Virtual Links for Audio Augmented Reality Applications – Long-Term Study

Contact for Further Information:

If you have any further questions, you can use the contact emails below:

Researcher: j.bhattacharyya.1@research.gla.ac.uk

Supervisor: stephen.brewster@glasgow.ac.uk

Thank you for your interest in participating in this study.

Bibliography

[1] Vero Vanden Abeele et al. “Development and Validation of the Player Experience Inventory: A Scale to Measure Player Experiences at the Level of Functional and Psychosocial Consequences”. In: *International Journal of Human-Computer Studies* 135 (Mar. 2020), p. 102370. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2019.102370. (Visited on 02/08/2023).

[2] Sharath Adavanne et al. “Sound Event Localization and Detection of Overlapping Sources Using Convolutional Recurrent Neural Networks”. In: *IEEE Journal of Selected Topics in Signal Processing* 13.1 (Mar. 2019), pp. 34–48. ISSN: 1932-4553, 1941-0484. DOI: 10.1109/JSTSP.2018.2885636. (Visited on 12/01/2023).

[3] Jens Ahrens and Carl Andersson. “Perceptual Evaluation of Headphone Auralization of Rooms Captured with Spherical Microphone Arrays with Respect to Spaciousness and Timbre”. In: *The Journal of the Acoustical Society of America* 145.4 (Apr. 2019), pp. 2783–2794. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.5096164. (Visited on 01/08/2024).

[4] Robert Albrecht, Riitta Väänänen, and Tapio Lokki. “Guided by Music: Pedestrian and Cyclist Navigation with Route and Beacon Guidance”. In: *Personal and Ubiquitous Computing* 20.1 (Feb. 2016), pp. 121–145. ISSN: 1617-4909, 1617-4917. DOI: 10.1007/s00779-016-0906-z. (Visited on 11/17/2025).

[5] Gerasimos Arvanitis, Konstantinos Moustakas, and Nikos Fakotakis. “Real-Time Context Aware Audio Augmented Reality”. In: *Speech and Computer*. Ed. by Andrey Ronzhin, Rodmonga Potapova, and Nikos Fakotakis. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 333–340. ISBN: 978-3-319-23132-7. DOI: 10.1007/978-3-319-23132-7_41.

[6] Ronald T. Azuma. “A Survey of Augmented Reality”. In: *Presence: teleoperators & virtual environments* 6.4 (1997), pp. 355–385.

[7] Max Bain et al. *WhisperX: Time-Accurate Speech Transcription of Long-Form Audio*. July 2023. DOI: 10.48550/arXiv.2303.00747. arXiv: 2303.00747 [cs]. (Visited on 01/20/2025).

[8] Amit Barde et al. “Binaural Spatialisation Over a Bone Conduction Headset: Elevation Perception”. In: *Proceedings of the 22nd International Conference on Auditory Display - ICAD 2016*. Canberra, Australia: The International Community for Auditory Display, July 2016, pp. 173–176. ISBN: 978-0-9670904-3-6. DOI: 10.21785/icad2016.013. HDL: 1853/56562. (Visited on 11/03/2022).

[9] Amit Barde et al. “Binaural Spatialization over a Bone Conduction Headset: Minimum Discernable Angular Difference”. In: *Audio Engineering Society Convention 140*. Audio Engineering Society, 2016.

[10] Valentin Bauer et al. “Designing an Interactive and Collaborative Experience in Audio Augmented Reality”. In: *Virtual Reality and Augmented Reality*. Ed. by Patrick Bourdot et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 305–311. ISBN: 978-3-030-31908-3. DOI: 10.1007/978-3-030-31908-3_20.

[11] Benjamin B. Bederson. “Audio Augmented Reality: A Prototype Automated Tour Guide”. In: *Conference Companion on Human Factors in Computing Systems - CHI ’95*. Denver, Colorado, United States: ACM Press, 1995, pp. 210–211. ISBN: 978-0-89791-755-1. DOI: 10.1145/223355.223526. (Visited on 10/12/2022).

[12] Durand R Begault. “Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems”. In: *Journal of the Audio Engineering Society* 40.11 (1992), pp. 895–904.

[13] Durand R Begault, Elizabeth M Wenzel, and Mark R Anderson. “Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source”. In: *Journal of the Audio Engineering Society* 49.10 (2001), pp. 904–916.

[14] Durand R. Begault and Elizabeth M. Wenzel. “Headphone Localization of Speech”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 35.2 (June 1993), pp. 361–376. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/001872089303500210. (Visited on 01/10/2023).

[15] Virginia Best et al. “Sound Externalization: A Review of Recent Research”. In: *Trends in Hearing* 24 (Jan. 2020). ISSN: 2331-2165, 2331-2165. DOI: 10.1177/2331216520948390. (Visited on 10/31/2022).

[16] Jacob Bhattacharyya, Alessandro Vinciarelli, and Stephen Brewster. “Birds of a Feather Augment Together: Exploring Sonic Links Between Real and Virtual Worlds in Audio Augmented Reality”. In: *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, Oct. 2025, pp. 1490–1500. ISBN: 979-8-3315-8761-1. DOI: 10.1109/ISMAR67309.2025.00153.

[17] Jacob Bhattacharyya, Alessandro Vinciarelli, and Stephen Anthony Brewster. "Sonomancer: Exploring Sonic Control Schemes for Audio Augmented Reality Games". In: *Proceedings of the ACM on Human-Computer Interaction* 9.6 (Oct. 2025), pp. 976–994. ISSN: 2573-0142. DOI: 10.1145/3748629.

[18] Jacob Bhattacharyya et al. "Investigating the Influence of Environmental Acoustics and Playback Device for Audio Augmented Reality Applications". In: *Audio Engineering Society Conference: AES 2024 International Audio for Games Conference*. Tokyo, Japan: Audio Engineering Society, Apr. 2024, p. 10.

[19] Mark Billinghurst, Adrian Clark, and Gun Lee. "A Survey of Augmented Reality". In: *Foundations and Trends® in Human–Computer Interaction* 8.2-3 (2015), pp. 73–272. ISSN: 1551-3955, 1551-3963. DOI: 10.1561/1100000049. (Visited on 09/12/2024).

[20] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press, 1997.

[21] Simon Blessenohl et al. "Improving Indoor Mobility of the Visually Impaired with Depth-Based Spatial Sound". In: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. Santiago, Chile: IEEE, Dec. 2015, pp. 418–426. ISBN: 978-1-4673-9711-7. DOI: 10.1109/ICCVW.2015.62. (Visited on 06/11/2024).

[22] Jon Blistein. *Swedish Band Make Album Available Exclusively in the Woods*. May 2015. (Visited on 11/19/2025).

[23] Bluebrain. *Announcing 'The National Mall' The First Location Aware Album*. https://bluebrainmu mall.html. Mar. 2011. (Visited on 09/10/2024).

[24] Jeffrey R. Blum, Mathieu Bouchard, and Jeremy R. Cooperstock. "Spatialized Audio Environmental Awareness for Blind Users with a Smartphone". In: *Mobile Networks and Applications* 18.3 (June 2013), pp. 295–309. ISSN: 1383-469X, 1572-8153. DOI: 10.1007/s11036-012-0425-8. (Visited on 10/04/2022).

[25] Costas Boletsis and Dimitra Chasanidou. "Audio Augmented Reality in Public Transport for Exploring Tourist Sites". In: *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. NordiCHI '18. New York, NY, USA: Association for Computing Machinery, Sept. 2018, pp. 721–725. ISBN: 978-1-4503-6437-9. DOI: 10.1145/3240167.3240243. (Visited on 10/05/2022).

[26] Riccardo Bona et al. "Automatic Parameters Tuning of Late Reverberation Algorithms for Audio Augmented Reality". In: *AudioMostly 2022*. St. Pölten Austria: ACM, Sept. 2022, pp. 36–43. ISBN: 978-1-4503-9701-8. DOI: 10.1145/3561212.3561236. (Visited on 01/20/2023).

[27] Till Bovermann, René Tünnermann, and Thomas Hermann. "Auditory Augmentation:" in: *International Journal of Ambient Computing and Intelligence* 2.2 (Apr. 2010), pp. 27–41. ISSN: 1941-6237, 1941-6245. DOI: 10.4018/jaci.2010040102. (Visited on 06/07/2026).

[28] Danielle Bragg, Nicholas Huynh, and Richard E. Ladner. "A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users". In: *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. Reno Nevada USA: ACM, Oct. 2016, pp. 3–13. ISBN: 978-1-4503-4124-0. DOI: 10.1145/2982142.2982171. (Visited on 09/12/2024).

[29] Stephen A. Brewster, Peter C. Wright, and Alistair DN Edwards. "An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces". In: *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. 1993, pp. 222–227. (Visited on 11/18/2025).

[30] Fabian Brinkmann et al. "A Round Robin on Room Acoustical Simulation and Auralization". In: *The Journal of the Acoustical Society of America* 145.4 (Apr. 2019), pp. 2746–2760. ISSN: 0001-4966. DOI: 10.1121/1.5096178. (Visited on 11/09/2022).

[31] Jeanne H. Brockmyer et al. "The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game-Playing". In: *Journal of Experimental Social Psychology* 45.4 (July 2009), pp. 624–634. ISSN: 00221031. DOI: 10.1016/j.jesp.2009.02.016. (Visited on 01/10/2023).

[32] Adelbert W. Bronkhorst. "Localization of Real and Virtual Sound Sources". In: *The Journal of the Acoustical Society of America* 98.5 (Nov. 1995), pp. 2542–2553. ISSN: 0001-4966. DOI: 10.1121/1.413219. (Visited on 10/31/2022).

[33] Isna Alfi Bustoni, Mark McGill, and Stephen Brewster. "The Perception of a Tap: Using Auditory Augmented Reality to Alter the Contact Properties of a Physical Object". In: *ACM Trans. Comput.-Hum. Interact.* 32.5 (Oct. 2025), 53:1–53:33. ISSN: 1073-0516. DOI: 10.1145/3745770. (Visited on 11/19/2025).

[34] Isna Alfi Bustoni, Mark McGill, and Stephen Anthony Brewster. "Exploring the Alteration and Masking of Everyday Noise Sounds Using Auditory Augmented Reality". In: *International Conference on Multimodel Interaction*. San Jose Costa Rica: ACM, Nov. 2024, pp. 154–163. ISBN: 979-8-4007-0462-8. DOI: 10.1145/3678957.3685750. (Visited on 11/19/2024).

[35] Thomas Chatzidimitris, Damianos Gavalas, and Despina Michael. "SoundPacman: Audio Augmented Reality in Location-Based Games". In: *2016 18th Mediterranean Electrotechnical Conference (MELECON)*. Lemesos, Cyprus: IEEE, Apr. 2016, pp. 1–6. DOI: 10.1109/MELCON.2016.7495414.

[36] Max Chen, Shano Liang, and Gillian Smith. "Stackable Music: A Marker-Based Augmented Reality Music Synthesis Game". In: *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Stratford ON Canada: ACM, Oct. 2023, pp. 22–28. ISBN: 979-8-4007-0029-3. DOI: 10.1145/3573382.3616071. (Visited on 01/19/2024).

[37] Long Cheng, Michael Schreiner, and Andreas Kunz. "Comparing Tracking Accuracy in Standalone MR-HMDs: Apple Vision Pro, Hololens 2, Meta Quest 3, and Pico 4 Pro". In: *30th ACM Symposium on Virtual Reality Software and Technology*. Trier Germany: ACM, Oct. 2024, pp. 1–2. ISBN: 979-8-4007-0535-9. DOI: 10.1145/3641825.3689518. (Visited on 12/02/2025).

[38] F Wai-ling Ho-Ching, Jennifer Mankoff, and James A Landay. "Can You See What I Hear? The Design and Evaluation of a Peripheral Sound Display for the Deaf". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vol. 5. Ft. Lauderdale: Association for Computing Machinery, 2003, pp. 161–168.

[39] Brian Clarkson et al. "Auditory Context Awareness via Wearable Computing". In: (1998).

[40] M. Cohen, S. Aoki, and N. Koizumi. "Augmented Audio Reality: Telepresence/VR Hybrid Acoustic Environments". In: *Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication*. Nov. 1993, pp. 361–364. DOI: 10.1109/ROMAN.1993.367692.

[41] Karen Collins et al. "An Exploration of Distributed Mobile Audio and Games". In: *Proceedings of the International Academic Conference on the Future of Game Design and Technology*. Vancouver British Columbia Canada: ACM, May 2010, pp. 253–254. ISBN: 978-1-4503-0235-7. DOI: 10.1145/1920778.1920821. (Visited on 01/22/2024).

[42] Chris Columbus. *Harry Potter and the Philosopher's Stone*. Fantasy. 2001.

[43] Ana Grasielle Dionisio Correa et al. "GenVirtual: An Augmented Reality Musical Game for Cognitive and Motor Rehabilitation". In: *2007 Virtual Rehabilitation*. Venice, Italy: IEEE, Sept. 2007, pp. 1–6. ISBN: 978-1-4244-1203-7. DOI: 10.1109/ICVR.2007.4362120. (Visited on 01/19/2024).

[44] María Cuevas-Rodríguez et al. "3D Tune-In Toolkit: An Open-Source Library for Real-Time Binaural Spatialisation". In: *PLOS ONE* 14.3 (Mar. 2019). Ed. by Ifat Yasin, e0211899. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0211899. (Visited on 10/27/2022).

[45] Florian Denk et al. "Acoustic Transparency in Hearables - Technical Evaluation". In: *Journal of the Audio Engineering Society* 68.7/8 (Sept. 2020), pp. 508–521. ISSN: 15494950. DOI: 10.17743/jaes.2020.0042. (Visited on 11/28/2025).

[46] Dhwani Desai and Ninad Mehendale. “A Review on Sound Source Localization Systems”. In: (2022).

[47] Dolby. *Dolby Atmos on Apple Music*. (Visited on 12/02/2025).

[48] Matthew Donahoe. “OnTheRun: A Location-based Exercise Game”. PhD thesis. Massachusetts Institute of Technology, 2011.

[49] Dowino. *A Blind Legend*. Game [Windows, macOS, iOS, Android]. 2015.

[50] Synesthetic Echo. *Bumblebee Jam*. Game [Bose AR]. 2019.

[51] Inger Ekman et al. “Designing Sound for a Pervasive Mobile Game”. In: *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. ACE ’05. New York, NY, USA: Association for Computing Machinery, June 2005, pp. 110–116. ISBN: 978-1-59593-110-8. DOI: 10.1145/1178477.1178492. (Visited on 01/22/2024).

[52] Gary W. Elko, Jens Meyer, and Steven Backer. “A New Sixth-Order Eigenmike® Spherical Microphone Array for Spatial Sound Field Recording”. In: *The Journal of the Acoustical Society of America* 153.3_supplement (Mar. 2023), A143. ISSN: 0001-4966. DOI: 10.1121/10.0018443. (Visited on 12/02/2025).

[53] Kajetan Enge, Matthias Frank, and Robert Holdrich. “Listening Experiment on the Plausibility of Acoustic Modeling in Virtual Reality”. In: *Fortschritte der Akustik–DAGA* (2020), pp. 13–16.

[54] Isaac Engel et al. “Perceptual Comparison of Ambisonics-Based Reverberation Methods in Binaural Listening”. In: (2019), 6 pages. DOI: 10.25836/SASP.2019.11. (Visited on 06/14/2023).

[55] Isaac Engel et al. “Perceptual Implications of Different Ambisonics-based Methods for Binaural Reverberation”. In: *The Journal of the Acoustical Society of America* 149.2 (Feb. 2021), pp. 895–910. ISSN: 0001-4966. DOI: 10.1121/10.0003437. (Visited on 11/16/2022).

[56] Isaac Engel et al. “The SONICOM HRTF Dataset”. In: *Journal of the Audio Engineering Society* 71.5 (May 2023), pp. 241–253. ISSN: 15494950. DOI: 10.17743/jaes.2022.0066. (Visited on 10/11/2023).

[57] Christine Evers and Patrick A. Naylor. “Acoustic SLAM”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.9 (Sept. 2018), pp. 1484–1498. ISSN: 2329-9304. DOI: 10.1109/TASLP.2018.2828321. (Visited on 12/02/2025).

[58] Mohammed Fakhour et al. “A Cultural Scavenger Hunt Serious Game Based on Audio Augmented Reality”. In: *Advanced Intelligent Systems for Sustainable Development (AI2SD’2019)*. Ed. by Mostafa Ezziyyani. Vol. 1102. Cham: Springer International Publishing, 2020, pp. 1–8. ISBN: 978-3-030-36652-0 978-3-030-36653-7. DOI: 10.1007/978-3-030-36653-7_1. (Visited on 01/19/2024).

[59] Danielle M. Ferraro et al. “The Phantom Chorus: Birdsong Boosts Human Well-Being in Protected Areas”. In: *Proceedings of the Royal Society B: Biological Sciences* 287.1941 (Dec. 2020), p. 20201811. ISSN: 0962-8452, 1471-2954. DOI: 10.1098/rspb.2020.1811. (Visited on 06/12/2025).

[60] David M. Frohlich. *Audiophotography: Bringing Photos to Life with Sounds*. Vol. 3. Springer Science & Business Media, 2004. (Visited on 11/19/2025).

[61] David M. Frohlich. “From Audio Paper to Next Generation Paper”. In: *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*. 2017, pp. 9–10.

[62] David M. Frohlich et al. “Designing Interactive Newsprint”. In: *International Journal of Human-Computer Studies* 104 (Aug. 2017), pp. 36–49. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2017.03.002. (Visited on 11/19/2025).

[63] Hiroshi Furuya et al. “The Influence of Total and Directional Energy of Late Sound on Listener Envelopment”. In: *Acoustical Science and Technology* 26.2 (2005), pp. 208–211. ISSN: 1346-3969, 1347-5177. DOI: 10.1250/ast.26.208. (Visited on 11/25/2025).

[64] Johanna Gampe. “Interactive Narration within Audio Augmented Realities”. In: *Interactive Storytelling*. Ed. by Ido A. Iurgel, Nelson Zagalo, and Paolo Petta. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2009, pp. 298–303. ISBN: 978-3-642-10643-9. DOI: 10.1007/978-3-642-10643-9_34.

[65] Hannes Gamper. “Enabling Technologies for Audio Augmented Reality Systems”. In: (2014), p. 114.

[66] Yiannis Georgiou and Eleni A. Kyza. “The Development and Validation of the ARI Questionnaire: An Instrument for Measuring Immersion in Location-Based Augmented Reality Settings”. In: *International Journal of Human-Computer Studies* 98 (Feb. 2017), pp. 24–37. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2016.09.014. (Visited on 02/03/2023).

[67] Christian Giguère and Sharon M. Abel. “Sound Localization: Effects of Reverberation Time, Speaker Array, Stimulus Frequency, and Stimulus Rise/Decay”. In: *The Journal of the Acoustical Society of America* 94.2 (Aug. 1993), pp. 769–776. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.408206. (Visited on 11/25/2025).

[68] Marta Gospodarek et al. “Methodology for Perceptual Evaluation of Plausibility with Self-Translation of the Listener”. In: *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*. 2022.

[69] Raphael Grasset et al. “The Mixed Reality Book: A New Multimedia Reading Experience”. In: *CHI ’07 Extended Abstracts on Human Factors in Computing Systems*. San Jose CA USA: ACM, Apr. 2007, pp. 1953–1958. ISBN: 978-1-59593-642-4. DOI: 10.1145/1240866.1240931. (Visited on 06/11/2024).

[70] Joshua Gregg, Gavin Kearney, and Lauren Ward. “Using Enhanced Audio to Create an Accessible Mobile Augmented Reality App for Visually Impaired Users”. In: (2022), p. 10.

[71] Vincent Grimaldi et al. “Externalization of Virtual Sounds Using Low Computational Cost Algorithms for Hearables”. In: (Dec. 2020). DOI: 10.5167/UZH-198597. (Visited on 11/03/2022).

[72] Rishabh Gupta et al. “Acoustic Transparency in Hearables for Augmented Reality Audio: Hear-through Techniques Review and Challenges”. In: *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*. 2020.

[73] Aki Härmä et al. “Augmented Reality Audio for Mobile and Wearable Appliances”. In: *J. Audio Eng. Soc.* 52.6 (2004), p. 23.

[74] Aki Härmä et al. “Techniques and Applications of Wearable Augmented Reality Audio”. In: *Audio Engineering Society Convention 114*. 2003.

[75] Sandra G. Hart and Lowell E. Staveland. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Advances in Psychology*. Vol. 52. Elsevier, 1988, pp. 139–183. ISBN: 978-0-444-70388-0. DOI: 10.1016/S0166-4115(08)62386-9. (Visited on 02/03/2025).

[76] W. M. Hartmann. “Localization of Sound in Rooms”. In: *The Journal of the Acoustical Society of America* 74.5 (Nov. 1983), pp. 1380–1391. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.390163. (Visited on 11/25/2025).

[77] William M. Hartmann and Andrew Wittenberg. “On the Externalization of Sound Images”. In: *The Journal of the Acoustical Society of America* 99.6 (June 1996), pp. 3678–3688. ISSN: 0001-4966. DOI: 10.1121/1.414965. (Visited on 11/09/2022).

[78] Adrian Hazzard et al. “The Rough Mile: Reframing Location Through Locative Audio”. In: *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*. London United Kingdom: ACM, Aug. 2017, pp. 1–8. ISBN: 978-1-4503-5373-1. DOI: 10.1145/3123514.3123540. (Visited on 10/17/2022).

[79] Marcus Hedblom et al. “Bird Song Diversity Influences Young People’s Appreciation of Urban Landscapes”. In: *Urban Forestry & Urban Greening* 13.3 (Jan. 2014), pp. 469–474. ISSN: 1618-8667. DOI: 10.1016/j.ufug.2014.04.002. (Visited on 06/12/2025).

[80] Florian Heller, Aaron Krämer, and Jan Borchers. “Simplifying Orientation Measurement for Mobile Audio Augmented Reality Applications”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 615–624. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557021. (Visited on 10/12/2022).

[81] Florian Heller and Johannes Schöning. “NavigaTone: Seamlessly Embedding Navigation Cues in Mobile Music Listening”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, Apr. 2018, pp. 1–7. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174211. (Visited on 06/11/2024).

[82] Florian Heller et al. “Where Are We? Evaluating the Current Rendering Fidelity of Mobile Audio Augmented Reality Systems”. In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI ’16. New York, NY, USA: Association for Computing Machinery, Sept. 2016, pp. 278–282. ISBN: 978-1-4503-4408-1. DOI: 10.1145/2935334.2935365. (Visited on 10/12/2022).

[83] Adrian Herbez et al. *Ghast Blasters*. Game [Bose AR]. 2019.

[84] Shawn Hershey et al. “CNN Architectures for Large-Scale Audio Classification”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2017, pp. 131–135. DOI: 10.1109/ICASSP.2017.7952132. (Visited on 11/07/2023).

[85] Kotaro Hoshiba et al. “Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments”. In: *Sensors* 17.11 (Nov. 2017), p. 2535. ISSN: 1424-8220. DOI: 10.3390/s17112535. (Visited on 12/03/2025).

[86] Matthew B. Hoy. “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants”. In: *Medical Reference Services Quarterly* 37.1 (Jan. 2018), pp. 81–88. ISSN: 0276-3869, 1540-9597. DOI: 10.1080/02763869.2018.1404391. (Visited on 11/27/2025).

[87] Volley Inc. *Jeopardy!* Game [Amazon Alexa]. 2015.

[88] Reinis Indans, Eva Hauthal, and Dirk Burghardt. “Towards an Audio-Locative Mobile Application for Immersive Storytelling”. In: *KN - Journal of Cartography and Geographic Information* 69.1 (May 2019), pp. 41–50. ISSN: 2524-4957, 2524-4965. DOI: 10.1007/s42489-019-00007-1. (Visited on 01/19/2024).

[89] *ISO 3382-1: 2009. Measurement of Room Acoustic Parameters*. Geneva, Switzerland, 2009.

[90] Rune Møberg Jacobsen et al. “In the Zone!—Controlling and Visualising Sound Zones”. In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 2022, pp. 1–4.

[91] Giordano Jacuzzi. “"Augmented Audio": An Overview of the Unique Tools and Features Required for Creating AR Audio Experiences”. In: *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[92] Antti Jylhä. “Sonic Gestures as Input in Human-Computer Interaction: Towards a Systematic Approach”. In: *Proceedings of the SMC2011-8th Sound and Music Computing Conference, Padova*. Padova: Sound and Music Computing Network, 2011.

[93] Antti Jylhä and Cumhur Erkut. “A Hand Clap Interface for Sonic Interaction with the Computer”. In: *CHI ’09 Extended Abstracts on Human Factors in Computing Systems*. Boston MA USA: ACM, Apr. 2009, pp. 3175–3180. ISBN: 978-1-60558-247-4. DOI: 10.1145/1520340.1520452. (Visited on 05/31/2024).

[94] Stefan Kahl et al. “BirdNET: A Deep Learning Solution for Avian Diversity Monitoring”. In: *Ecological Informatics* 61 (Mar. 2021), p. 101236. ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2021.101236. (Visited on 05/31/2024).

[95] Ganesh Kailas and Nachiketa Tiwari. “Design for Immersive Experience: Role of Spatial Audio in Extended Reality Applications”. In: *Design for Tomorrow—Volume 2*. Ed. by Amaresh Chakrabarti et al. Smart Innovation, Systems and Technologies. Singapore: Springer, 2021, pp. 853–863. ISBN: 978-981-16-0119-4. DOI: 10.1007/978-981-16-0119-4_69.

[96] Neofytos Kaplanis et al. “Perception of Reverberation in Small Rooms: A Literature Study”. In: (2014).

[97] Mohamed Kari et al. “SoundsRide: Affordance-Synchronized Music Mixing for In-Car Audio Augmented Reality”. In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. Virtual Event USA: ACM, Oct. 2021, pp. 118–133. ISBN: 978-1-4503-8635-7. DOI: 10.1145/3472749.3474739. (Visited on 01/19/2024).

[98] Brian F G Katz et al. “The Past Has Ears at Notre-Dame: Immersive Audio Experiences for Public Engagement”. In: (2024).

[99] Oliver Beren Kaul, Kersten Behrens, and Michael Rohs. "Mobile Recognition and Tracking of Objects in the Environment through Augmented Reality and 3D Audio Cues for People with Visual Impairments". In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, May 2021, pp. 1–7. ISBN: 978-1-4503-8095-9. DOI: 10.1145/3411763.3451611. (Visited on 06/11/2024).

[100] Kia. *Kia Soundscapes: A New Way to Hear Movement*. June 2025. (Visited on 11/24/2025).

[101] Shin Kim, Kun-pyo Lee, and Tek-Jin Nam. "Sonic-Badminton: Audio-Augmented Badminton Game for Blind People". In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. San Jose California USA: ACM, May 2016, pp. 1922–1929. ISBN: 978-1-4503-4082-3. DOI: 10.1145/2851581.2892510. (Visited on 01/19/2024).

[102] Takumi Kiriu et al. "Development of an Acoustic AR Gamification System to Support Physical Exercise". In: *Proceedings of the 27th ACM International Conference on Multimedia*. Nice France: ACM, Oct. 2019, pp. 1056–1058. ISBN: 978-1-4503-6889-6. DOI: 10.1145/3343031.3350589. (Visited on 01/19/2024).

[103] Andreas Kratky. "Walking in the Head: Methods of Sonic Augmented Reality Navigation". In: *Human-Computer Interaction. Recognition and Interaction Technologies*. Ed. by Masaaki Kurosu. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 469–483. ISBN: 978-3-030-22643-5. DOI: 10.1007/978-3-030-22643-5_37.

[104] Michael Krzyzaniak, David Frohlich, and Philip J.B. Jackson. "Six Types of Audio That DEFY Reality!: A Taxonomy of Audio Augmented Reality with Examples". In: *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*. Nottingham United Kingdom: ACM, Sept. 2019, pp. 160–167. ISBN: 978-1-4503-7297-8. DOI: 10.1145/3356590.3356615. (Visited on 10/17/2022).

[105] Mark Lawton, Stuart Cunningham, and Ian Convery. "Nature Soundscapes: An Audio Augmented Reality Experience". In: *Proceedings of the 15th International Audio Mostly Conference*. Graz Austria: ACM, Sept. 2020, pp. 85–92. ISBN: 978-1-4503-7563-4. DOI: 10.1145/3411109.3411142. (Visited on 10/17/2022).

[106] Thibaud Leclère, Mathieu Lavandier, and Fabien Perrin. "On the Externalization of Sound Sources with Headphones without Reference to a Real Source". In: *The Journal of the Acoustical Society of America* 146.4 (Oct. 2019), pp. 2309–2320. ISSN: 0001-4966. DOI: 10.1121/1.5128325. (Visited on 11/24/2025).

[107] Alexander Lindau and Stefan Weinzierl. "Assessing the Plausibility of Virtual Acoustic Environments". In: *Acta Acustica united with Acustica* 98.5 (Sept. 2012), pp. 804–810. ISSN: 16101928. DOI: 10.3813/AAA.918562. (Visited on 11/01/2022).

[108] Robert W. Lindeman, Haruo Noma, and Paulo Goncalves de Barros. “An Empirical Study of Hear-Through Augmented Reality: Using Bone Conduction to Deliver Spatialized Audio”. In: *2008 IEEE Virtual Reality Conference*. Reno Nevada USA: IEEE, Mar. 2008, pp. 35–42. DOI: 10.1109/VR.2008.4480747.

[109] Livia Games. *Run the Realm*. Game [Android, iOS]. 2019.

[110] Tapio Lokki et al. “Application Scenarios of Wearable and Mobile Augmented Reality Audio”. In: *The 116th Convention of the Audio Engineering Society*. Berlin: Audio Engineering Society, 2004, p. 9.

[111] Jack M. Loomis, Reginald G. Golledge, and Roberta L. Klatzky. “Navigation System for the Blind: Auditory Display Modes and Guidance”. In: *Presence: Teleoperators and Virtual Environments* 7.2 (Apr. 1998), pp. 193–203. ISSN: 1054-7460. DOI: 10.1162/105474698565677. (Visited on 09/10/2024).

[112] Kent Lyons, Maribeth Gandy, and Thad Starner. “Guided by Voices: An Audio Augmented Reality System”. In: *Proceedings of the 6th International Conference on Auditory Display*. Atlanta, Georgia, USA: International Community for Auditory Display, July 2000, p. 6.

[113] Malik Mallem. “Augmented Reality: Issues, Trends and Challenges”. In: *Proceedings of the International Conference on Image Processing Theory, Tools and Applications (IPTA 2010)*. Paris France: IEEE, Aug. 2010, pp. 8–8. DOI: 10.1109/IPTA.2010.5586829.

[114] Nicholas Mariette. “Human Factors Research in Audio Augmented Reality”. In: *Human Factors in Augmented Reality Environments*. Ed. by Weidong Huang, Leila Alem, and Mark A. Livingston. New York, NY: Springer, 2013, pp. 11–32. ISBN: 978-1-4614-4205-9. DOI: 10.1007/978-1-4614-4205-9_2. (Visited on 10/06/2022).

[115] Nicholas Mariette and Brian F G Katz. “SOUNDDELTA – LARGE SCALE, MULTI-USER AUDIO AUGMENTED REALITY”. In: (2009), p. 6.

[116] Nick Mariette. “From Backpack to Handheld: The Recent Trajectory of Personal Location Aware Spatial Audio”. In: *PerthDAC 2007: Proceedings of the 7th Digital Arts and Culture Conference*. Perth: Curtin University of Technology, 2007, pp. 233–240.

[117] Milos Markovic, Soren K. Olesen, and Dorte Hammershoi. “Three-Dimensional Point-Cloud Room Model for Room Acoustics Simulations”. In: *ICA 2013 Montreal*. Montreal, Canada, 2013, pp. 015122–015122. DOI: 10.1121/1.4800237. (Visited on 12/02/2025).

[118] Gale L. Martin. “The Utility of Speech Input in User-Computer Interfaces”. In: *International Journal of Man-Machine Studies* 30.4 (Apr. 1989), pp. 355–375. ISSN: 00207373. DOI: 10.1016/S0020-7373(89)80023-9. (Visited on 11/27/2025).

[119] Vincent Martin, Isabelle Viaud-Delmon, and Olivier Warusfel. "Effect of Environment-Related Cues on Auditory Distance Perception in the Context of Audio-Only Augmented Reality". In: *Applied Sciences* 12.1 (Dec. 2021), p. 348. ISSN: 2076-3417. DOI: 10.3390/app12010348. (Visited on 10/04/2022).

[120] Ken I. McAnally and Russell L. Martin. "Sound Localization with Head Movement: Implications for 3-d Audio Displays". In: *Frontiers in Neuroscience* 8 (Aug. 2014). ISSN: 1662-453X. DOI: 10.3389/fnins.2014.00210. (Visited on 12/02/2025).

[121] Mark McGill et al. "Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–16. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376702. (Visited on 10/12/2022).

[122] David McGookin. "Towards Ubiquitous Location-Based Audio: Challenges and Future Directions". In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. Florence Italy: ACM, 2016, pp. 1064–1068.

[123] David McGookin et al. "Shaking the Dead: Multimodal Location Based Experiences for Un-Stewarded Archaeological Sites". In: *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. Copenhagen Denmark: ACM, Oct. 2012, pp. 199–208. ISBN: 978-1-4503-1482-4. DOI: 10.1145/2399016.2399048. (Visited on 06/11/2024).

[124] Donald H Mershon and L Edward King. "Intensity and Reverberation as Factors in the Auditory Perception of Egocentric Distance". In: (1975).

[125] Paul Milgram and Fumio Kishino. "A Taxonomy of Mixed Reality Visual Displays". In: *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), pp. 1321–1329.

[126] Sébastien Moreau, Jérôme Daniel, and Stéphanie Bertet. "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone". In: (2006).

[127] Masayuki Morimoto, Kazuhiro Iida, and Kimihiro Sakagami. "The Role of Reflections from behind the Listener in Spatial Impression$". In: (2000).

[128] Nikolaos Moustakas, Andreas Floros, and Nikolaos Kanellopoulos. "Eidola: An Interactive Augmented Reality Audio-Game Prototype". In: *Audio Engineering Society Convention 127*. New York, NY, USA: Audio Engineering Society, 2009.

[129] Elizabeth D. Mynatt et al. “Audio Aura: Light-Weight Audio Augmented Reality”. In: *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology - UIST ’97*. Banff, Alberta, Canada: ACM Press, 1997, pp. 211–212. ISBN: 978-0-89791-881-7. DOI: 10.1145/263407.264218. (Visited on 10/12/2022).

[130] Elizabeth D. Mynatt et al. “Designing Audio Aura”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’98*. Los Angeles, California, United States: ACM Press, 1998, pp. 566–573. ISBN: 978-0-201-30987-4. DOI: 10.1145/274644.274720. (Visited on 10/17/2022).

[131] Anna N. Nagele et al. “Interactive Audio Augmented Reality in Participatory Performance”. In: *Frontiers in Virtual Reality* 1 (Feb. 2021), p. 610320. ISSN: 2673-4192. DOI: 10.3389/frvir.2020.610320. (Visited on 10/05/2022).

[132] Annika Neidhardt, Christian Schneiderwind, and Florian Klein. “Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms - Literature Review and Theoretical Framework”. In: *Trends in Hearing* 26 (Jan. 2022), p. 233121652210929. ISSN: 2331-2165, 2331-2165. DOI: 10.1177/23312165221092919. (Visited on 11/24/2022).

[133] Annika Neidhardt, Alby Ignatious Tommy, and Anson Davis Pereppadan. “Plausibility of an Interactive Approaching Motion towards a Virtual Sound Source Based on Simplified BRIR Sets”. In: (2018).

[134] Annika Neidhardt and Anna Maria Zerlik. “The Availability of a Hidden Real Reference Affects the Plausibility of Position-Dynamic Auditory AR”. In: *Frontiers in Virtual Reality* 2 (Sept. 2021), p. 678875. ISSN: 2673-4192. DOI: 10.3389/frvir.2021.678875. (Visited on 05/31/2023).

[135] Niantic, Inc. *Pokemon Go*. Game [iOS, Android]. 2016.

[136] Søren H. Nielsen. “Auditory Distance Perception in Different Rooms”. In: *Journal of the Audio Engineering Society* 41.10 (Oct. 1993), pp. 755–770.

[137] Anton Nijholt. “Toward an Ever-present Extended Reality: Distinguishing Between Real and Virtual”. In: *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*. Cancun, Quintana Roo Mexico: ACM, Oct. 2023, pp. 396–399. ISBN: 979-8-4007-0200-6. DOI: 10.1145/3594739.3610726. (Visited on 12/08/2025).

[138] Josefa Oberem, Bruno Masiero, and Janina Fels. “Experiments on Authenticity and Plausibility of Binaural Reproduction via Headphones Employing Different Recording Methods”. In: *Applied Acoustics* 114 (Dec. 2016), pp. 71–78. ISSN: 0003682X. DOI: 10.1016/j.apacoust.2016.07.009. (Visited on 11/09/2022).

[139] Natasa Paterson et al. “Location Aware Interactive Game Audio”. In: *Audio Engineering Society Conference: 41st International Conference: Audio for Games*. London, UK: Audio Engineering Society, 2011.

[140] Natasa Paterson et al. “Spatial Audio and Reverberation in an Augmented Reality Game Sound Design”. In: *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. Audio Engineering Society, 2010, p. 9.

[141] Natasa Paterson et al. “Viking Ghost Hunt: Creating Engaging Sound Design for Location-Aware Applications”. In: *International Journal of Arts and Technology* 6.1 (2013), p. 61. ISSN: 1754-8853, 1754-8861. DOI: 10.1504/IJART.2013.050692. (Visited on 01/22/2024).

[142] Petricore Games, Inc. *Dead Drop Desperado*. Game [Bose AR]. 2019.

[143] Iana Podkosova, Michael Urbanek, and Hannes Kaufmann. “A Hybrid Sound Model for 3D Audio Games with Real Walking”. In: *Proceedings of the 29th International Conference on Computer Animation and Social Agents*. CASA ’16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 189–192. ISBN: 978-1-4503-4745-7. DOI: 10.1145/2915926.2915948. (Visited on 01/22/2024).

[144] Sandra Poeschl, Konstantin Wall, and Nicola Doering. “Integration of Spatial Sound in Immersive Virtual Environments an Experimental Study on Effects of Spatial Sound on Presence”. In: *2013 IEEE Virtual Reality (VR)*. Mar. 2013, pp. 129–130. DOI: 10.1109/VR.2013.6549396.

[145] David Poirier-Quinot et al. “HRTF Performance Evaluation: Methodology and Metrics for Localisation Accuracy and Learning Assessment”. In: *Advances in Fundamental and Applied Research on Spatial Audio*. IntechOpen, Oct. 2022. ISBN: 978-1-83969-005-1 978-1-83969-006-8. DOI: 10.5772/intechopen.104931. (Visited on 06/01/2023).

[146] Barteld N. J. Postma and Brian F. G. Katz. “Creation and Calibration Method of Acoustical Models for Historic Virtual Reality Auralizations”. In: *Virtual Reality* 19.3 (Nov. 2015), pp. 161–180. ISSN: 1434-9957. DOI: 10.1007/s10055-015-0275-3. (Visited on 02/13/2023).

[147] Thomas Potter, Zoran Cvetković, and Enzo De Sena. “On the Relative Importance of Visual and Spatial Audio Rendering on VR Immersion”. In: *Frontiers in Signal Processing* 2 (Sept. 2022). ISSN: 2673-8198. DOI: 10.3389/frsip.2022.904866. (Visited on 11/21/2025).

[148] Swadhin Pradhan et al. “Smartphone-Based Acoustic Indoor Space Mapping”. In: 2.2 (2018).

[149] Brad Rakerd and W. M. Hartmann. "Localization of Sound in Rooms, II: The Effects of a Single Reflecting Surface". In: *The Journal of the Acoustical Society of America* 78.2 (Aug. 1985), pp. 524–533. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.392474. (Visited on 11/25/2025).

[150] Razer. *Beamforming Soundbar with Head-Tracking AI*. 2023. (Visited on 12/01/2025).

[151] *RECOMMENDATION ITU-R BS.2132-0 – Method for the Subjective Quality Assessment of Audible Differences of Sound Systems Using Multiple Stimuli without a given Reference*. Oct. 2019.

[152] RjDj. *Dimensions*. Game [iOS]. 2011.

[153] Emmanouel Rovithis et al. "Audio Legends: Investigating Sonic Interaction in an Augmented Reality Audio Game". In: *Multimodal Technologies and Interaction* 3.4 (Nov. 2019), p. 73. ISSN: 2414-4088. DOI: 10.3390/mti3040073. (Visited on 10/04/2022).

[154] Emmanuel Rovithis et al. *Towards Citizen Science for Smart Cities: A Framework for a Collaborative Game of Bird Call Recognition Based on Internet of Sound Practices*. Mar. 2021. arXiv: 2103.16988 [cs]. (Visited on 01/22/2024).

[155] Dariusz Rumiński. "An Experimental Study of Spatial Sound Usefulness in Searching and Navigating through AR Environments". In: *Virtual Reality* 19.3-4 (Nov. 2015), pp. 223–233. ISSN: 1359-4338, 1434-9957. DOI: 10.1007/s10055-015-0274-4. (Visited on 10/04/2022).

[156] Yoko Sasaki, Ryo Tanabe, and Hiroshi Takernura. "Online Spatial Sound Perception Using Microphone Array on Mobile Robot". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Oct. 2018, pp. 2478–2484. DOI: 10.1109/IROS.2018.8593777. (Visited on 12/03/2025).

[157] Lauri Savioja and U. Peter Svensson. "Overview of Geometrical Room Acoustic Modeling Techniques". In: *The Journal of the Acoustical Society of America* 138.2 (Aug. 2015), pp. 708–730. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.4926438. (Visited on 12/02/2025).

[158] Nitin Sawhney and Chris Schmandt. "Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*. Pittsburgh, Pennsylvania, United States: ACM Press, 1999, pp. 96–103. ISBN: 978-0-201-48559-2. DOI: 10.1145/302979.303005. (Visited on 11/15/2022).

[159] Henning Schepker et al. "Acoustic Transparency in Hearables - Perceptual Sound Quality Evaluations". In: *Journal of the Audio Engineering Society* 68.7/8 (Sept. 2020), pp. 495–507. ISSN: 15494950. DOI: 10.17743/jaes.2020.0045. (Visited on 11/09/2022).

[160] Christian Schneiderwind and Annika Neidhardt. *Modified Late Reverberation in an Audio Augmented Reality Scenario*. 2024. DOI: 10.5281/ZENODO.3457782. (Visited on 10/15/2024).

[161] Hanna Schraffenberger and Edwin Van Der Heide. "Everything Augmented: On the Real in Augmented Reality". In: *Journal of Science and Technology of the Arts* 6.1 (Jan. 2014), pp. 17–29. DOI: 10.7559/CITARJ.V6I1.125. (Visited on 10/20/2022).

[162] Hanna Schraffenberger and Edwin van der Heide. "Sonically Tangible Objects". In: *Skyler and Bliss Original Citation Adkins, Monty and Segretier, Laurent (2015) Skyler and Bliss. In: xCoAx 2015: Proceedings of the Third Conferenc on Computation, Communication, Aesthetics and X. Universidade do Porto, Porto* (2015), p. 233. (Visited on 11/19/2025).

[163] Hanna Kathrin Schraffenberger. "Arguably Augmented Reality: Relationships between the Virtual and the Real". PhD thesis. The Netherlands: Leiden University, 2018.

[164] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. "Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)". In: *International Journal of Interactive Multimedia and Artificial Intelligence* 4.6 (2017), p. 103. ISSN: 1989-1660. DOI: 10.9781/ijimai.2017.09.001. (Visited on 08/29/2024).

[165] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. "The Experience of Presence: Factor Analytic Insights". In: *Presence: Teleoperators and Virtual Environments* 10.3 (June 2001), pp. 266–281. ISSN: 1054-7460. DOI: 10.1162/105474601300343603. (Visited on 02/07/2023).

[166] Stefania Serafin et al. "Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions". In: *IEEE Computer Graphics and Applications* 38.2 (Mar. 2018), pp. 31–43. ISSN: 1558-1756. DOI: 10.1109/MCG.2018.193142628.

[167] Oliver Shih and Anthony Rowe. "Can a Phone Hear the Shape of a Room?" In: *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. Apr. 2019, pp. 277–288. DOI: 10.1145/3302506.3310407.

[168] Marjan Sikora et al. "Soundscape of an Archaeological Site Recreated with Audio Augmented Reality". In: *ACM Transactions on Multimedia Computing, Communications, and Applications* 14.3 (July 2018), 74:1–74:22. ISSN: 1551-6857. DOI: 10.1145/3230652. (Visited on 10/12/2022).

[169] Panote Siriaraya et al. "Investigating the Use of Spatialized Audio Augmented Reality to Enhance the Outdoor Running Experience". In: *Entertainment Computing* 44 (Jan. 2023), p. 100534. ISSN: 1875-9521. DOI: 10.1016/j.entcom.2022.100534. (Visited on 01/19/2024).

[170] Six to Start. *Zombies, Run!* Game [iOS, Android]. 2012.

[171] Adam J. Sporka, Sri H. Kurniawan, and Pavel Slavík. "Acoustic Control of Mouse Pointer". In: *Universal Access in the Information Society* 4.3 (Mar. 2006), pp. 237–245. ISSN: 1615-5289, 1615-5297. DOI: 10.1007/s10209-005-0010-z. (Visited on 11/26/2025).

[172] Adam J. Sporka et al. "CHANTI: Predictive Text Entry Using Non-Verbal Vocal Input". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vancouver BC Canada: ACM, May 2011, pp. 2463–2472. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979302. (Visited on 11/26/2025).

[173] Adam J. Sporka et al. "Non-Speech Input and Speech Recognition for Real-Time Control of Computer Games". In: *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. Portland Oregon USA: ACM, Oct. 2006, pp. 213–220. ISBN: 978-1-59593-290-7. DOI: 10.1145/1168987.1169023. (Visited on 01/28/2025).

[174] Falling Squirrel. *The Vale: Shadow of the Crown*. Game [Xbox, Playstation, Nintendo Switch, PC]. 2021. (Visited on 06/12/2023).

[175] Evgeny Stemasov et al. "Augmenting Human Hearing Through Interactive Auditory Mediated Reality". In: *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*. 2018, pp. 69–71.

[176] E. Stobbe et al. "Birdsongs Alleviate Anxiety and Paranoia in Healthy Participants". In: *Scientific Reports* 12.1 (Oct. 2022), p. 16414. ISSN: 2045-2322. DOI: 10.1038/s41598-022-20841-0. (Visited on 06/12/2025).

[177] Bethesda Game Studios. *The Elder Scrolls V: Skyrim*. Game [Windows, PlayStation 3, Xbox 360, PlayStation 4, Xbox One, Nintendo Switch, PlayStation 5, Xbox Series X/S]. Nov. 2011.

[178] Left Right Studios. *PairPlay*. Game [iOS]. 2021.

[179] Miikka Tikander. "Usability Issues in Listening to Natural Sounds with an Augmented Reality Audio Headset". In: *J. Audio Eng. Soc.* 57.6 (2009), p. 12.

[180] Eleni Tsalera, Andreas Papadakis, and Maria Samarakou. "Comparison of Pre-Trained CNNs for Audio Classification Using Transfer Learning". In: *Journal of Sensor and Actuator Networks* 10.4 (Dec. 2021), p. 72. ISSN: 2224-2708. DOI: 10.3390/jsan10040072. (Visited on 12/03/2025).

[181] USound. *Fauna: Glasses That Let You Listen*. https://usound.com/fauna-glasses-that-let-you-listen/. 2021.

[182] Minna Vasarainen, Sami Paavola, and Liubov Vetoshkina. "A Systematic Literature Review on Extended Reality: Virtual, Augmented and Mixed Reality in Working Life". In: *International Journal of Virtual Reality* 21.2 (Oct. 2021), pp. 1–28. ISSN: 2727-9979. DOI: 10.20870/IJVR.2021.21.2.4620. (Visited on 06/02/2026).

[183] Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A. Brewster. "Auditory Display Design for Exploration in Mobile Audio-Augmented Reality". In: *Personal and Ubiquitous computing* 16.8 (2012), pp. 987–999.

[184] Sampo Vesa and Tapio Lokki. "AN EYES-FREE USER INTERFACE CONTROLLED BY FINGER SNAPS". In: *Proceedings of the 8th International Conference on Digital Audio Effects (DAFx-05)*. Madrid, Spain, 2005, pp. 262–265.

[185] Ralf Von Appen and André Doehring. "Nevermind The Beatles, Here's Exile 61 and Nico: 'The Top 100 Records of All Time' – a Canon of Pop and Rock Albums from a Sociological and an Aesthetic Perspective". In: *Popular Music* 25.1 (Jan. 2006), pp. 21–39. ISSN: 0261-1430, 1474-0095. DOI: 10.1017/s0261143005000693. (Visited on 07/24/2025).

[186] Avery Wang. "The Shazam Music Recognition Service". In: *Communications of the ACM* 49.8 (Aug. 2006), pp. 44–48. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/1145287.1145312. (Visited on 01/14/2025).

[187] Jing Wang et al. "Binaural Sound Localization Based on Deep Neural Network and Affinity Propagation Clustering in Mismatched HRTF Condition". In: *EURASIP Journal on Audio, Speech, and Music Processing* 2020.1 (Dec. 2020), p. 4. ISSN: 1687-4722. DOI: 10.1186/s13636-020-0171-y. (Visited on 12/03/2025).

[188] Marian Weger, Thomas Hermann, and Robert Höldrich. "Real-Time Auditory Contrast Enhancement". In: *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*. Newcastle upon Tyne: Department of Computer and Information Sciences, Northumbria University, June 2019, pp. 254–261. ISBN: 978-0-9670904-6-7. DOI: 10.21785/icad2019.026. HDL: 1853/61505. (Visited on 11/07/2023).

[189] Marian Weger and Robert Höldrich. "A Hear-through System for Plausible Auditory Contrast Enhancement". In: *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*. Nottingham United Kingdom: ACM, Sept. 2019, pp. 1–8. ISBN: 978-1-4503-7297-8. DOI: 10.1145/3356590.3356593. (Visited on 04/11/2025).

[190] Weird Sisters Interactive. *The Clairvoyant*. Game [Bose AR]. 2019.

[191] Elizabeth M Wenzel et al. "Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio". In: Routledge, 2017.

[192] Elizabeth M. Wenzel et al. “Localization Using Nonindividualized Head-related Transfer Functions”. In: *The Journal of the Acoustical Society of America* 94.1 (July 1993), pp. 111–123. ISSN: 0001-4966. DOI: 10.1121/1.407089. (Visited on 10/31/2022).

[193] Stephan Werner et al. “A Summary on Acoustic Room Divergence and Its Effect on Externalization of Auditory Events”. In: *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. June 2016, pp. 1–6. DOI: 10.1109/QoMEX.2016.7498973. (Visited on 11/24/2025).

[194] Ashley Whittaker. *Classify Birds Acoustically with BirdNET-Pi*. https://www.raspberrypi.com/nev birds-acoustically-with-birdnet-pi/. May 2022.

[195] Connor M. Wood and Stefan Kahl. “Guidelines for Appropriate Use of BirdNET Scores and Other Detector Outputs”. In: *Journal of Ornithology* 165.3 (July 2024), pp. 777–782. ISSN: 2193-7192, 2193-7206. DOI: 10.1007/s10336-024-02144-5. (Visited on 12/03/2025).

[196] Worthing and Moncrief. *Overherd*. Game [Bose AR]. 2019.

[197] Tong Wu et al. “Recent Advances in 3D Gaussian Splatting”. In: *Computational Visual Media* 10.4 (Aug. 2024), pp. 613–642. ISSN: 2096-0662. DOI: 10.1007/s41095-024-0436-y. (Visited on 12/02/2025).

[198] Jing Yang, Amit Barde, and Mark Billinghurst. “Audio Augmented Reality: A Systematic Review of Technologies, Applications, and Future Research Directions”. In: *Journal of the Audio Engineering Society* 70.10 (Nov. 2022), pp. 788–809. ISSN: 15494950. DOI: 10.17743/jaes.2022.0048. (Visited on 11/03/2022).

[199] Franz Zotter and Matthias Frank. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Vol. 19. Springer Topics in Signal Processing. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-17206-0 978-3-030-17207-7. DOI: 10.1007/978-3-030-17207-7. (Visited on 10/26/2023).

[200] Zylia. *Zylia Portable Recording Studio - Technology White Paper*. https://www.zylia.co/white-paper.html. Oct. 2017. (Visited on 12/02/2025).