Designing co-located multi-device audio experiences

David Geary

PhD

University of York

School of Physics, Engineering, and Technology

September 2024

Abstract

The widespread distribution and diversification of mobile computing devices presents new opportunities for the development and consumption of new interactive, immersive and accessible media experiences using multiple connected devices. Tools now exist for the creation of these experiences; however, there is still limited understanding of the best design practices and use cases for the technology, especially in the context of audio experiences. Furthermore, the little extant research on multi-device audio experiences is fragmented, leading to problems navigating the research. The work in this thesis approached these issues through a two step programme.

The first step involved improving understanding of these experiences from a conceptual perspective and identifying relevant design concepts. The application space of co-located multi-device audio experiences is explored and documented through a review of the literature and a survey. Through qualitative analysis of the obtained information, an initial set of seven design dimensions, *synchronisation*, *context*, *position*, *relationship*, *interactivity*, *organisation*, and *distribution*, were proposed that capture distinct design considerations and can be used to characterise and compare experiences. The dimensions were evaluated in a series of workshops where they were found to be useful as a co-design tool for experience ideation.

The second step comprised of identifying new experience opportunities and using distinct yet complementary methods to gain new insights. Two case studies were conducted that focused on delivering concurrent personal and shared audio on different devices. The first focused on the design and technical implementation of a novel multi-device audio prototype using a practice-based approach, where a new design dimension, *perception*, was also identified. The second focused on the evaluation of the listening experience of using wireless earbuds with active transparency for augmenting TV audio through a controlled listening test. A novel experimental method is presented and production recommendations were provided for this multi-device audio reproduction medium.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. I also declare that parts of this research have been presented in previous conference and journal publications, which are listed in the contributions in Section 1.2.

Acknowledgements

First and foremost I would like to thank my supervisor Damian Murphy, who's unwavering optimism and support has been invaluable over the course of my programme of study. Furthermore, I would like to thank my past and present industry supervisors Jon Francombe and Kristian Hentschel of Bang & Olufsen and BBC R&D for their guidance and knowledge. Their work and wider experience has been a great source of inspiration throughout my PhD.

I would like to thank colleagues and friends at the AudioLab at the University of York, who have undoubtedly had a positive impact on my PhD experience. In particular, many thanks to Jay Harrison and Joe Williams, who I have shared my PhD journey with through all the ups and downs. Our weekly catch ups were always a welcome interaction, especially during more isolated periods at the beginning and end of our PhDs.

Many thanks to XRStories and BBC R&D for providing the research funding to make this thesis possible, and Bang & Olufsen for being a hospitable host for parts of my research.

I have been fortunate to have spent time with the audio team at BBC R&D across my PhD programme, both online and in-person. I am grateful to have had the opportunity to work alongside and learn from friendly talented researchers, who were always welcoming. I wish them all the best of luck in the future.

I wouldn't of been able to complete this thesis without the support from my family and friends, particularly my parents, Bill and Liz Geary, who have been a constant source of encouragement.

Thank you to Emma, who has been the most caring, understanding and loving partner I could've wished for throughout the last four years, and thanks to my cat, Ivy, who has been the best little companion who has sat by my side (and on my laptop keyboard) for more hours than I can count.

Finally... I would like to thank Frank for his sense of humour and for this thesis template.

Chapter 1

Introduction

Audio impacts our lives in many different ways. At its best, audio can provide pleasure, excitement and joy. Equally, bad audio experiences can be irritating, upsetting or, after long term exposure, can even be harmful to health [1]. The desire for positive audio experiences is especially important as, unlike vision, hearing is a sense that receives input from all directions and cannot easily be blocked or attenuated without external objects. Throughout our lives, we obtain much of our information about our surroundings through sound. Understandably, sound is employed in the same capacity in media entertainment for world-building and storytelling, immersing the listener in the environment envisioned by the creator.

Over the course of the last 50 years, technological developments have transformed many aspects of how we engage and interact with multimedia content. In the 1990s, the advent of the internet enabled new media publishing and consumption avenues combined with social and communication features, ultimately leading to new content delivery mediums such as *on-demand* and *streaming*. Personal mobile computing in the mid 2000s, such as smartphones and tablets, further transformed media engagement, providing now constant access and stimulating new personalised forms of media such as short-form audiovisual content. From the 2010s, computing and hardware advancements led to the eventual commercialisation of virtual reality (VR), building upon research efforts from the 1990s. This facilitated widespread creation of more immersive and interactive visual environments. Finally, over the last 3 years there has been considerable hype around artificial intelligence (AI), particularly generative AI, and how it can disrupt and transform audiovisual media production. However, in the time all this has taken place, everyday consumption of audio has largely been unchanged. Most people still listen to music in stereo and listen to the TV over its inbuilt loudspeakers. Cinema sound systems haven't evolved much aside from adding height loudspeakers. A demonstration of this is that domestic ownership of surround sound systems remains minimal, limited to keen audio enthusiasts rather than the average consumer [2].

A consequence of the aforementioned technological developments is that mobile computing devices are now ubiquitous and it is now common to find ourselves in the presence of many of these devices at once. Thus, there is a recognition that this presents an opportunity to enrich our interactions with multiple co-located devices through connected and complementary usage. There are now dedicated research fields to exploring this possibility known as *multi-device* or *cross-device computing* [3]. However, most of this research is focused on the visual modality. This thesis presents a portfolio of research investigating co-located multi-device audio experiences, focusing on how they can be better understood and designed, and exploring novel multi-device audio concepts through practice and evaluation methods.

1.1 Objectives & Research Questions

The work in this thesis is driven by one overarching research question:

> How can the conceptual understanding and design of co-located multi-device audio experiences be improved and how can new experience opportunities be identified and explored?

To make this objective easier to address, this question is split into a number of smaller, more focused research questions, and are addressed individually across the chapters of this thesis. Questions 1-3 are associated with increasing understanding and design and questions 4-6 are associated with new experience opportunities and how they can be explored.

1. What is the current state of knowledge of co-located multi-device audio experiences?

2. What are the design principles that underpin co-located multi-device audio experiences?
3. How can these design principles be used to explore the range of possible experience design inputs or parameters (design space)?
4. How can these design principles be used to create new experiences?
5. What can we learn of new co-located multi-device audio experiences through prototyping?
6. What can we learn of new co-located multi-device audio experiences through evaluation?

1.2 Contributions

Throughout this thesis, novel contributions are made to the field in the form of knowledge and resources across various areas relevant to co-located multi-device audio experiences. These are outlined below:

- A dataset of co-located multi-device audio experiences and delivery platforms that contains a name, a description of the experience, a description of the audio content and a reference for each entry. The dataset can be used as a point of reference by interested researchers and practitioners in future research related to this area. It is accessible through [4].
- A set of design dimensions that capture the key design principles for co-located multi-device audio experiences which can be used to describe and characterise existing experiences and conceptualise and design new ones. Published as *Design Dimensions for Co-located Multi-device Audio Experiences* in in MDPI Applied Sciences [5].
- A series of workshops dedicated to understanding and ideation of co-located multi-device audio experiences where the structure and exercises can be applied and adapted for future co-design activities involving experiences of this type. Published and presented at *Audio Mostly 2023* [6].
- A novel co-located multi-device audio-visual experience using a creative combination of hearables and loudspeakers, where external loudspeakers are used to augment a pair of

hearables, facilitated through active transparency. A detailed description and explanation of the design and technical implementation can provide useful insights that can be applied in future experiences. Published and presented at *Audio Mostly 2023* [6].

- An original experimental method for approximating the listening experience of augmenting TV with hearables using a simulation technique, that can be utilised in future listening tests investigating concurrent direct and external sound, facilitated through active transparency. Accepted for publication in the *IEEE Communications Magazine: Internet of Sounds* 2024 issue.

- A listening test exploring the application of hearables for audio content 'mirroring', allocating single audio mix elements onto a separate device, for augmented TV programme viewing. Initial recommendations are provided for content provision with this reproduction method, primarily that mirroring of dialogue on the hearables is more preferable than mirroring of music and effects. Accepted for publication in the *IEEE Communications Magazine: Internet of Sounds* 2024 issue.

1.3 Thesis Structure

A breakdown of each chapter in this thesis is provided below. The following chapter (Chapter 2) provides a comprehensive overview of topic areas and relevant literature as a prerequisite to work carried out in the later chapters. This largely covers immersive and spatial sound, multi-device research, experience design and evaluation, devices and other enabling technologies.

Chapter 3 aims to develop a more complete conceptual understanding of co-located multi-device audio experiences through surveying of academic literature and the wider creative arts. A survey and a systematic literature search are employed to collect and synthesise information on tools and applications of co-located multi-device audio experiences and this is used to form a set of design dimensions for mapping the related design space.

Chapter 4 builds on the design dimensions by exploring their utility for conceptualising and designing new co-located multi-device experiences through a series of workshops with audio and design industry experts. A large number of experience ideas are formed and feedback is obtained on the design dimensions.

The next two chapters take the form of case studies to investigate new domestic audio experience opportunities through practice and evaluation perspectives. Chapter 5 adopts a research-by-design method to develop one of the experience ideas from the workshops into a novel working proof-of-concept prototype where the technical and creative processes are described, and design choices and challenges are discussed.

In the second case study in Chapter 6, the underlying prototype concept is adapted to a broadcast TV context and is evaluated in the form of a listening test to investigate novel multi-device audio rendering options for improving TV audio in shared viewing scenarios. Here the challenges of evaluating multi-device audio experiences are reflected upon and recommendations are provided to researchers and creative practitioners for further exploration of this concept.

Finally, Chapter 7 concludes the thesis with a summary of the completed work and how it answers the proposed research question. Reflection and discussion is presented on the contribution of the work to related research areas and opportunities for further work are highlighted.

Chapter 2

Foundations of Audio and Multi-Device Ecosystems

2.1 Introduction

Multi-device audio experiences require knowledge drawn from many different research areas; the main areas are the *Internet of Things* (IoT), *Human-Computer-Interaction* (HCI), *Extended Reality* (XR), spatial and immersive audio, and acoustics. The purpose of this chapter is to form a detailed picture of the current state of knowledge on multi-device audio experiences that motivates much of the subsequent original research reported in this thesis. Like the thesis as a whole, the core focus of this chapter is on audio experiences; therefore, most of the topics discussed pertain to audio. Nevertheless, relevant information can be obtained from audio adjacent research areas, especially where there has been greater activity, for example, on multi-display experiences. Consequently, there is some reporting on these related areas where there is benefit in doing so.

This chapter is broken down into the following sections:

- Section 2.2 details how audio reproduction systems have evolved over time through technological developments, and how these affect audio experiences.

- Section 2.3 then describes how these audio experiences are typically evaluated.
- Section 2.4 explores how multi-device audio experiences have emerged through a diverse range of audio devices with different affordances and contexts of use.
- Section 2.5 covers the methods through which these devices are combined to create experiences including how they are designed, the common use-cases and the technologies used for their implementation, and how they are evaluated.
- Finally, in Section 2.6 a brief summary of the chapter is presented.

2.2 Evolution of Audio Reproduction

2.2.1 Foundations

2.2.1.1 Channel-based Audio

Audio reproduction traditionally occurs through discrete speaker feeds or ‘channels’, and while the enabling technology has been around for a century, channel-based reproduction remains the most common form, with most audio today still consumed through the two-channel stereo format, via headphones or a pair of loudspeakers. Stereo, developed by Bell Laboratories and Alan Blumlein in the 1930s [7, 8], is the most common channel-based audio format, simply consisting of two frontal audio channels in a left/right configuration. It was later introduced commercially in the late 1950s through stereo vinyl records and players.

Stereo has remained popular due to the fact that it is a simple system which offers robust spatial audio experiences within a quadrant in front of the listener, and is therefore suitable for applications such as music, as well as cinema and watching television, where audio is combined with a visual stimulus. A stereo signal can also be easily collapsed into mono without a huge degradation in quality which is beneficial for some broadcasting applications. For an immersive audio experience, however, the limited frontal reproduction area doesn’t provide the level of spatial realism or envelopment that systems with more channels can.

The term multichannel typically refers to systems with more than two audio channels and are referred to collectively as surround sound formats. The primary loudspeaker configuration standard for multichannel audio, and from which many other more recent formats are derived, is 5.1, which comprises of three front channels (Left, Center, Right, or LCR), two surround channels (Left Surround, Right Surround, or LS, RS) and a Low Frequency Effects (LFE) channel for bass frequency content [9]. From 1950 to 1980, prior to 5.1, various multichannel formats had been trialled in cinemas but had limited mainstream adoption due to complexity and cost [10]. It was only with the advent of digital media formats that multichannel audio became more widespread and the concept of home cinema was born, which led to the ITU standardization of the 5.1 surround sound system in 1992 [9]. The purpose of 5.1 is to produce a complete frontal stereo field and complement that with sound information from the rear, usually in the form of ambience and effects, as well as the dedicated LFE channel. Unfortunately, this format fails to meet the demands for full 360° spatial audio as used in more recent virtual and augmented experiences. Furthermore, although a big feature of cinema audio systems, the financial cost and arduous configuration of the system have resulted in home 5.1 systems remaining a niche product [2].

There are a number of more recent surround sound systems that have employed strategies to improve upon the drawbacks of 5.1, typically involving a larger number of loudspeakers and, therefore, channels. These newer systems attempt to increase the spatial audio resolution over a larger listening area. Such systems include 7.1, 10.2 and 22.2, the latter two of the three deploying a bed of elevated loudspeakers to more accurately reproduce sound from above the listener [11, 12]. A large array of loudspeakers is an expensive solution, which is why these systems are not typically consumer oriented, with one exception being Dolby Atmos [13]. Channel-based audio has persisted through its simplicity in that it is inherently linked with the reproduction system. However, now with increasing diversity in audio reproduction devices and environments (including virtual environments), channel-based audio begins to fall short through dependence on static loudspeaker positions, resulting in a lack of necessary flexibility.

2.2.1.2 Binaural

The aim of binaural audio is to reproduce sound signals at the listeners eardrums as accurately as possible. This means replicating spatial cues at both ears which provide the listener with relevant sound localization cues [14]. These spatial cues consist of time and level differences at each ear and frequency filtering due to sound reflections from the head and outer ear. The combination of these effects is known as the Head-Related Transfer Function (HRTF) and this is what forms the basis for binaural sound reproduction.

The simplest approach to achieve binaural sound reproduction is to record a sound environment with two microphones separated by a distance equivalent to the diameter of the head and replay it over headphones. For this purpose, dummy heads with microphones placed at the eardrum, such as the KEMAR [15], have been used to measure sound pressures at these positions. This technique has been utilised for over 50 years and has been applied in hearing aid design, soundscape measurement, and room acoustics, in addition to 3D audio [16].

Knowledge of the HRTF and binaural capture has been available since the 1920s [14], however, a number of problems have limited its application in commercial technologies. The major issue is that HRTFs are specific to the individual as everyone's anatomy is different; therefore, creating an average or generalized HRTF for commercial use is difficult. Moreover, binaural recordings generally aren't suitable for playback over loudspeakers without additional signal processing, and headphone equalization can distort the frequency cues in the HRTF making sound source localization more difficult. As digital signal processing has improved, binaural synthesis has become a possibility where sounds can be rendered binaurally when convolved with an appropriate head-related impulse response which applies the relevant spatial cues. When this is complemented with head tracking technology, realistic, interactive 3D audio environments can be achieved. These technological advancements have no doubt contributed to the development of the thriving *Extended Reality* (XR) sector, where this technology is now commonly implemented [17].

2.2.1.3 Ambisonics

The Ambisonic format is an alternative system for representing 3D sound. Its consists of a total recording, encoding and decoding system which can represent sound independent of the reproduction layout [18]. Sound is captured from all directions using a special microphone array and encoded as velocity and pressure components, known as B-format. The simplest Ambisonic format, known as first-order, can be captured using a microphone array of 4 coincident subcardioid microphones, which record sound in a 360° sphere. The B-format can be represented as three orthogonal figure-of-eight components (X,Y,Z) and an omnidirectional component (W). Higher-order ambisonic formats, a more recent innovation, can be obtained by using larger arrays of microphones which provide greater spatial resolution, albeit at the expense of more specialized equipment [19]. The sound can then be flexibly decoded and rendered for any reproduction configuration, including headphones [20]. As opposed to binaural implementations, head-tracking can be implemented more easily with Ambisonics virtual speaker arrays and using relatively simple time-variant rotation matrices; this avoids the difficult task of interpolation between HRTFs as required with interactive binaural applications [21]. This makes Ambisonics especially applicable for uses in XR, where head-tracking is highly important and head position can be determined from other aspects of the playback system, such as a VR headset. Some Digital Audio Workstations (DAW) now support Ambisonics [22], and suites of audio plugins exist which can apply effects across Ambisonic formats [23]. It has also been included as part of the standardised MPEG-H next-generation audio format [24].

Since the turn of the century, both binaural and Ambisonic audio have started to gain traction in consumer applications, partly driven by the desire for more immersive audiovisual experiences. They are particularly effective when the reproduction medium is headphones, therefore enabling them as suitable candidates for audio in XR. However, challenges remain for both formats. Binaural's efficacy is restricted to the implemented HRTF, where a generalised implementation will not be universally functional for everyone, whereas the lack of hardware decoder support limits wider adoption of Ambisonics. That said, research and industry investment in both mediums is still ongoing, and solutions for such barriers will not be far off [25]. It is reasonable to speculate that both binaural and Ambisonic audio will play an increasing part in audio experiences in the future.

2.2.2 State of the Art

2.2.2.1 Web Audio

Advancements in web technologies, notably HTML5, have presented even greater opportunities for the creation of novel audio experiences. The development of the Web Audio API [26] specifically has provided audio signal processing and synthesis capabilities for web applications. Realising the potential of these features, research in the field of web audio has increased significantly over the last five years, including the formation of new conferences such as the Web Audio Conference [27], which focuses heavily on the area. The benefits of web-based audio include platform independence and ease of prototyping and distribution. The result is that anyone with a network connection can access the full breadth of web audio applications. The Web Audio API has been applied in many areas of audio research including perceptual audio evaluation [28, 29], production tools [30, 31], spatial audio [32, 33] and ear training [34].

A benefit of using the web as a platform for developing audio experiences is that the suite of web Application Programming Interfaces (API) can be utilised to introduce various interactivity features. An example web audio application, WanderCouch [35] provided an enhancement on traditional music festival TV coverage by implementing exploratory 360° video and a live social media feed. Participants could navigate around a festival site using a virtual map, switch between different stages with a range of camera angles and audio streams and read relevant user generated content.

2.2.2.2 Object-based Audio

In addition to the opportunities presented by the Web Audio API and web-based delivery, another promising development in audiovisual media is that of Object-based audio (OBA). Encompassed by the wider paradigm of Object-based media (OBM), OBA is a modern, flexible approach for representing audio content, where audio assets are treated as individual 'objects'. These objects consist of the audio signal coupled with highly descriptive metadata which describe its properties such as spatial position and level. This format enables the audio to be rendered on the listeners' system (also known as the client-side system) and is adapted for their specific audio reproduction environment, unlike channel-based audio. For example, the audio assets will be rendered differently in the case of a 5.1 channel surround system than if headphones were used. Consequently, OBA is, in essence, format-agnostic, meaning the object-oriented content can be rendered for any playback system (e.g. loudspeakers, headphones or mobile devices). Development of web technology, and the proliferation of personal, smart devices has diversified the way in which media is consumed. These outcomes have accelerated the shift from traditional channel-based methods to the more modular object-based approach in broadcast, supported largely by international research efforts which have taken place over the last decade.

One of the significant opportunities of content being rendered at the client side rather than at any point in the signal transmission chain before this is that it enables numerous customisation, interactivity and personalisation options not available previously. Examples of this include switching between different languages, adjusting the volume of dialogue or other sounds and switching between available reproduction formats. The recognition of the importance of these features has led to sizeable research, especially in the area of accessibility in broadcast. In a comprehensive review, Ward & Shirley [36] highlight the importance of the role OBA will play in providing effective accessibility solutions, in particular, how dialogue and background audio objects can be adjusted by a listener to match their preference. Where audio objects are not available, e.g legacy content, source separation technology can be applied to obtain separate dialogue and background stems, which can then be leveraged further through OBA methods [37, 38]. Further audio mix personalisation choices have been explored in sports broadcasting for adjusting crowd noise and other effects [39, 40].

Outside of audio mix adjustment, an object-based system opens the door for new and alternative forms of interactive media content. As sections of content are broadcast as individual elements rather than one item, objects exhibit a degree of modularity where the sequence of objects can change. Consequently, experiences can be designed with variable length and depth, giving listeners different options based on their time availability or subject interest [41, 42]. This format is exemplified by the *Mermaid's Tears*, an object-based pilot production developed by the BBC [43]. This radio drama allowed listeners to select between three main characters, accessing three distinct perspectives of the story. *The Cook Along Kitchen Experience* (CAKE) is another example of a variable object-based experience where the content is changed and scaled depending on the viewer's requirements and the content timeline moves along at the viewer's pace [44].

Audio that is described as an object is independent of the reproduction system. Positional medadata associated with each object can describe its precise location in a sound scene, rather than from which loudspeaker it should be replayed. This enables object-based audio to be a powerful format for 3D audio applications. A clear example of this in practice can be seen in *The Turning Forest*, an audio-focused virtual reality fairytale [45]. In this production, all audio was produced and processed as objects, and then rendered in a game engine via a binaural rendering plugin.

In recognition of the capabilities, but also the new challenges, associated with OBA, efforts have been made to standardise the creation, transmission and delivery of the medium. One major development is MPEG-H Audio [24]. Commissioned by the MPEG standardisation group, this format comprises of a comprehensive audio distribution package which supports channel-based, object-based and Ambisonic sound encoding and reproduction in a vast array of formats ranging from 22.1 surround through to two-channel stereo and binaural. In 2017, MPEG-H Audio was incorporated in the first worldwide deployment of a ultra-high-definition (UHD) broadcasting service in South Korea [46]. MPEG-H also offers personalisation options enabled by OBA, where there is a range of ongoing research in implementing accessibility features such as dialogue enhancement, for both object-based and legacy content [37, 38, 47]. Dolby Laboratories also have their own next-generation audio codec known as AC-4 [48], which offers channel and OBA, personalisation and advanced loudness management. Dolby's newest spatial audio system, *Dolby Atmos*, makes use of additional 'height' speakers to provide an extra dimension for sound designers to accurately position individual sound sources or objects in 3D space, while simultaneously allowing diffuse sound routing through channels or 'beds' [13]. Introduced in 2012, there are now over 2000 Atmos installed cinema systems worldwide [49]. More recently, Dolby have pushed into the home market with a number of commercial systems such as soundbars and televisions now compatible with Atmos [50]. The format is also available on mobile, providing users with binaural audio through headphones.

The development of these standardised formats has been aided greatly by international research collaboration projects such as S3A [51] and Orpheus [52]. The objective of these projects was to address the issues around moving object-based audio from the research laboratory environment into mainstream broadcast production. Primarily, these challenges consisted of workflow difficulties and the lack of suitable tools and frameworks. Thus, considerable research has focused on developing these areas.

The Orpheus Project, running from 2015 to 2018, aimed to develop and validate a comprehensive end-to-end object-based media chain for audio content [52]. Recognising the increasing demands for multi-platform media consumption, the project developed a number of tools and interfaces to ensure OBA is deliverable across all platforms, including a complete internet protocol (IP) radio studio, an *Apple* iOS application and a compatible audio-video receiver for the living room [53]. Additionally, a suite of production tools and an Audio Definition Model (ADM) [54] integrated digital audio workstation (DAW) were also developed [55].

Concurrent with Orpheus, the S3A project's objective was to explore the potential of OBA and other spatial audio technologies to deliver new, immersive audio experiences. One of the major outcomes of the project was the proposal of an end-to-end, object-based workflow pipeline, which implements various components communicating via metadata and using visual interfaces for evaluation [56]. Together, these provide solutions for all parts of the object-based production chain from recording to monitoring. Identifying the absence of object-based audio content resources, part of the S3A project involved creating and publishing a number of object-based audio productions for research use in future audio experiments [57, 58]. This was complemented by the creation of an audio processing software framework called VISR [59], designed purposely for spatial and object-based audio. Furthermore, novel methods for spatial audio reproduction were proposed including an acoustic VR room from 360° video [60] and, most importantly in the context of this thesis, *Audio Device Orchestration* (ADO) [61].

Object-based audio is an emerging next-generation audio format which will bring new experiences and modernise production workflows to be able to cope with the changing landscape of media consumption. The last decade has produced many accomplishments in this area of research, especially in the development of new broadcast chains, audio codecs and production tools. These achievements now provide a solid foundation from which further cohesive research on OBA can proceed. Most importantly, OBA audio is highly relevant to multi-device audio ecosystems and experiences, enabling the necessary flexibility in audio rendering to match the many possible systems of connected audio devices and their possible modes of interaction.

2.3 Perceptual Evaluation of Audio

A part of the research presented in this thesis is concerned with the perceptual evaluation of audio experiences. This section describes the fundamentals of perceptual audio evaluation, covering listening tests and perceptual models.

2.3.1 Listening Tests

Listening tests remain a major method for the sensory evaluation of sound. This is because they provide a systematic approach to obtaining quantitative measurements of how sound is perceived by listeners.

There are two types of measurement that can be obtained through listening tests [62]:

- **Perceptual measurement**: An objective quantification of the sensorial strength of individual auditory attributes of the perceived stimulus.
- **Affective measurement**: An objective quantification of an overall impression of the perceived stimulus.

The type of measurement chosen is dependent on the research question under investigation. Furthermore, there are a number of categories of test methods, each with their own distinct objectives. The two most common methods are described here.

- **Discrimination methods**: These methods are used to observe similarities and differences between stimuli, based on a perceptual or affective attribute. This is usually achieved using pair-wise comparisons, for instance, in A/B testing [63].

- **Integrative methods**: These methods are used to obtain measurements on a single dependent variable (such as overall quality) for many conditions. There are a number of established methods, for example ITU-R BS.1534 (MUSHRA) [64] which is a multiple-stimulus test for assessing audio quality. The listening test in Chapter 6 follows this method.

A further four methods exist, *descriptive*, *mixed*, *temporal*, and *performance* methods; however, these are outside the scope of this review.

2.3.2 Perceptual Modelling

One potential limitation with listening tests is that most are costly and time-consuming. Consequently, efforts have been made of the last 20 years to create computational models that can accurately predict human responses in audio evaluation. These models, known as perceptual or objective measures, can provide quantitative estimates for both perceptual and affective attributes within seconds. As of now, many different models exist for different applications including estimation of audio quality, speech enhancement and source separation quality [65]. While these models can be useful for accurately estimating specific attributes, no one model can accurately estimate every aspect of human audio perception [62]. Therefore, listening tests remain an important method for perceptual audio evaluation, especially where the target application isn't covered by established models.

2.4 Ubiquitous Audio Devices

2.4.1 Personal Audio Devices

2.4.1.1 Mobile Devices

Personal mobile devices such as smartphones, tablets and laptops are everpresent in our daily lives. A 2018 household survey in the UK revealed that 82% of households contain smartphones and 66% contain tablets [2]. In addition, media consumption using these devices is very common, enabled primarily by internet services such as streaming. Mobile phones are no longer just communication devices, they have become powerful personal computers that contain a surprising range of features for data collection and usage. A study in 2010 on smartphone usage reported that participants interacted with their phones between 10-200 times a day, and their phones received up to 1000 MB of data per day [66]. These numbers are likely to be far greater now. A highly influential factor on the rapid uptake of smartphones was the advent of application stores, which enabled third-party developers to create and publish 'apps' which users could download to their own devices. In the first year that *Apple's* App Store was opened, there were over 200,000 apps available on the platform [67]. These software applications provided smartphones with a huge range of new features and services, closing the application gap between them and personal computers. With frequent hardware upgrades and additions of new sensors, research is continually ongoing to find new applications for smartphones. In the field of audio, there is increased focus on incorporating smartphones into many applications such as audio production [68, 69] and environmental sound monitoring [70]. Moreover, developing mobile phones into musical instruments has been another popular avenue of research [71].

Tablets can also access many of these 'apps', and with their larger screen size have been found to be particularly useful as teaching aids for groups in music education [72], and for presentation of user interfaces in audio experiments [73].

2.4.1.2 Wearable audio devices

Wearable computing [74] is another area which has seen significant growth over the last decade and has captured the attention of many consumers. Known more generally as 'wearables' [74], these devices are designed to be worn for extended periods of time and utilise sensors to collect various forms of data, such as biometrics (e.g. heart rate). Thus, a popular consumer application for such devices is health & fitness monitoring, where devices typically take the form of watches and wristbands [74]. Other forms of wearable devices exist such as head-mounted smart glasses which offer lifestyle computing applications, as well as smart clothing where garments are embedded with sensors which can offer physiological signal and ambient environment monitoring [75]. These small computing components have been integrated with personal audio devices, introducing a new category of wearables called 'hearables' [76]. Hearables are devices which are positioned in, on, or around the ear (Figure 2.1), and offer greater functionality than simple audio reproduction. Most consumer hearables have sound control features such as Active Noise Cancellation (ANC) for attenuating unwanted external noise, and 'transparency' for presenting the external sound environment back to the listener. Other features include tactile input, motion sensing, compasses, accelerometers, GPS and biosignal detecting [76, 77]. The functionality of such devices has opened up new possibilities for responsive audio experiences. An example of this is the use of 'acoustically transparent' *Bose Frames* [78] which deliver personal head-tracked audio, while allowing sounds in the real space around you to be heard [79].

Another worthy mention are head-mounted displays (HMDs) which offer immersive visual displays for use in virtual reality (VR). Originally designed as display extensions for computers, they now come as standalone devices with many, such as the *Meta Quest 3*, coming with integrated loudspeakers. The typical role of HMDs is to immerse users in a virtual environment, where applications range from recreational (gaming) to vocational (learning, training). A plausible audio experience is a crucial part of these applications, where audio can provide spatial information about virtual environments and an indication of interaction with virtual objects [81].

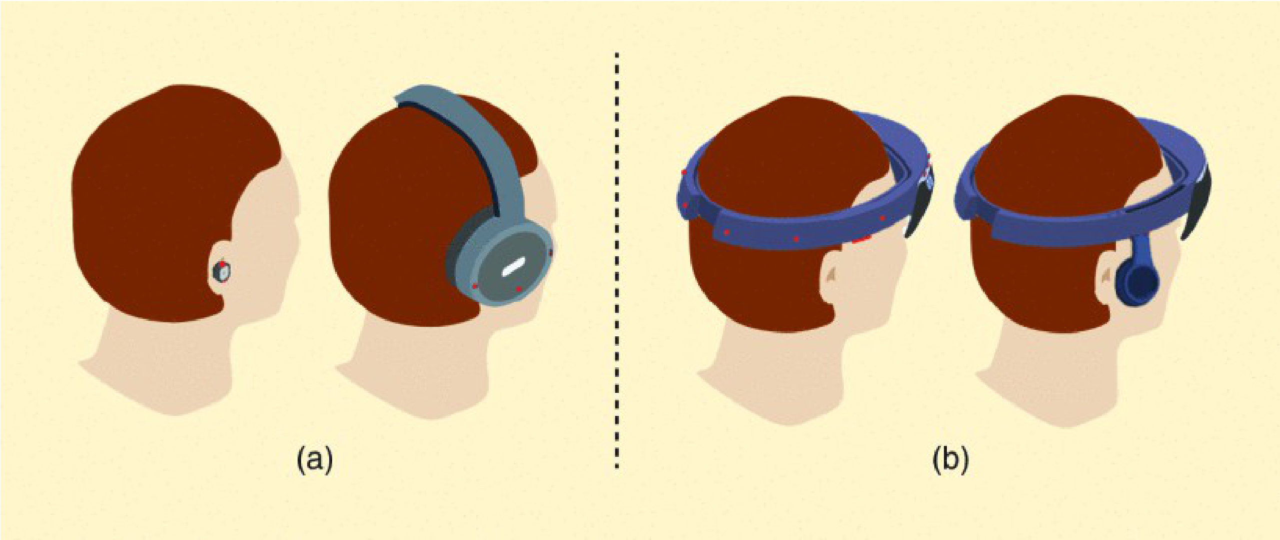


FIGURE 2.1: Two main types of hearable devices: (a) occluding-type hearables, are positioned in or on the ears, and (b) floating-type hearables, which do not touch the ear directly. After [80].

2.4.2 Shared Audio Devices

2.4.2.1 TV

The advent of commercial television was a landmark moment of cultural significance, bringing mass audio-visual media into the home for the first time, and in subsequent decades has been a dominant feature in every living room. However, many aspects of home entertainment have changed significantly since the inception of the first television set. These include the introduction of non-linear, on-demand streaming type services such as those provided by *Netflix* and *Amazon*, providing the consumer with the ability to choose where, when and what they watch or listen to. In the UK, streaming has become one of the most utilised methods of media consumption in the home setting [2]. More recently, short form media content, as hosted by social media platforms such as *YouTube* and *TikTok*, has grown significantly in popularity, especially with younger demographics. This increasing choice of audiovisual content has undermined traditional TV broadcasting. A recent Ofcom report reveals that the proportion of viewers tuning in to traditional broadcast TV is declining, even in the 'core' older demographic (65+) [82]. Average daily watch time for broadcast TV has also decreased from 2h 59m in 2021 to 2h 38m in 2022. This being said, traditional TV broadcasting is still popular and is forecast to remain part of home media consumption in the coming decades [83]. While the delivery context of audiovisual media has changed over time, the primary reproduction method still takes place through a single video screen in a stereo audio format, which has been the standard for decades.

Researchers have recognised the potential of using other common devices to enhance TV experiences. An example of this is 'second-screen' experiences [3, 84], that use a secondary device synchronised to the TV to provide additional visual programme related content.

2.4.2.2 Smart Speakers

Alongside the increasing sophistication of smartphone technology, a number of new devices are now utilising the same advanced computation. These devices are known as 'smart' devices and come in a variety of sizes and form factors, including smart TVs, smart wristbands, smart light bulbs and smart speakers. Smart speakers or voice assistant devices such as the Amazon Echo [85] and the Apple HomePod [86] are becoming increasingly common in households, especially with younger demographics [2]. These devices provide a primary interaction medium through the use of voice commands, which has been enabled by the advancement of speech synthesis and recognition technology. This mode of interaction is especially beneficial for those with visual impairments or poor motor skills. A recent study has reported that these devices are currently under-utilised, with the technological capabilities far exceeding simple playback of music or providing a weather forecast, which are the most common uses [87]. Owners of these devices typically use the same commands on a day-by-day basis and rarely explore new functionalities. Despite this, many interesting applications and services are available for these devices. Amazon's Alexa voice assistant has a library of applications, called 'Skills', which offer an array of services such as ordering food, controlling other smart home devices like lighting and thermostats, booking transport and playing games and quizes [88]. Research with smart speakers specifically has primarily concentrated on privacy and security concerns [89, 90], and consumer usage behaviours [87, 91]. However, the opportunities offered by smart speakers from an audio research perspective are beginning to be explored. An example of this can be seen in [92], where smart speakers have been employed to aid soundscape composition. Through the use of voice commands, sound designers could access a free sound web database and collect and play audio material.

2.4.3 Experimental Audio Devices

Not only has the web improved prototyping of experiences, but the reduction in size and cost of computing components has also contributed too. This has led to reduced barriers to entry, enabling individuals or small teams of researchers, technologists and hobbyists to create their own devices with reduced constraints.

An example of this is DIADs, or Distributed Interactive Audio Devices [93]. These devices consist of small *Raspberry Pi* computers with on board sensors and loudspeakers and are designed for experimenting with interactive multi-device audio experiences. DIADs offer audience interactivity through various motion sensors which can be used to modulate the audio source and can be remote-controlled and live-coded by a central computer to create evolving interactions and experiences. The internal components are enclosed in a spherical casing, chosen as spheres have an association with playful interactions and games. This physical form was put to the test in an art installation, where DIADs were used to create a musical version of lawn bowls [94]. Similar hardware has been implemented in wearable form devices, such as audio bracelets, and used as an augmented-audio toy for outdoor play with children [95]. The possibility of combining wireless and sensor technologies with musical instruments has also been explored. The result is smart-musical instruments that can interface with other connected devices, providing novel musical interactions [96].

Outside of delivering audio experiences, small embedded devices are now employed in environmental sound monitoring applications [97]. These devices not only capture sound through microphones but also process and transmit this data with a wider wireless acoustic sensor network (WASN). The result is a low-cost, real-time, high data volume network that can inform noise mitigation strategies. When combined with visual output hardware such as LEDs, these types of devices can even be used as accessibility aids for hearing-impaired individuals [98]. Projects like these highlight the opportunities that arise from modern lightweight electronics and networked computing, where it is possible to escape from traditional form factors of consumer

FIGURE 2.2: The physical form of DIADS, designed for playful interactions. After [93].

devices, and the associated environments of their use to explore new possibilities for multi-device applications and experiences.

There is now a greater number of devices present worldwide than ever before, and this is only set to increase in the upcoming decades. Additionally, the number of different forms of device available, with different features and use-cases is also increasing at a staggering rate. But it's not only the number of such devices that is remarkable, it is, perhaps more importantly, the ability to wirelessly connect these devices together. Through short-range technologies such as Bluetooth or network connectivity over the Internet, these devices can link together and share data to form complementary networks of devices [67]. As this is an audio focused thesis, the following sections will provide some context and explore multi-device technologies and use-cases relating to audio applications.

2.5 Multi-device Ecosystems

As introduced in Section 2.4.1.1, the proliferation of smartphones and other 'smart' devices has created a digital landscape where many of us, especially since the *COVID-19* pandemic, are increasingly living our lives through interactions with our devices. With many different kinds of devices available, it is now common for one person to own many of these devices, each typically having a target use case. However, as many of these devices contain similar hardware components and software, there is significant overlap in applications between devices. For example, it is possible to watch a live video broadcast on a laptop, tablet or smartphone. This environment has led to changes in human behaviour around media consumption and other practical applications. For instance, people now use more than one device to accomplish a task, such as reading and sending emails, and switch between devices depending on the context. The role of television has changed where television not only provides an activity but serves as a social hub, where the television acts as a background for other household activities, such as the use of other mobile devices [99]. We have moved from a 'one device, one application' environment to one where a single device can have many applications and where multiple devices can be applied to a single application. This section will explore multi-device ecosystems and experiences in more detail.

2.5.1 Existing research domains

The most prominent field of research in multi-device ecosystems is undoubtedly the *Internet of Things* (IoT), which is encapsulated by the idea of many 'things' communicating and collaborating together to achieve a common objective [100]. Behind this idea are a number of key enabling technologies, including sensing and communication technologies, management or network services, and privacy and security systems [100]. Building on the IoT is the recently proposed Internet of Sounds (IoS) which adopts the aforementioned technologies but focuses on their application in audio and sound related use-cases. This has been defined from [101]:

> "*the ensemble of Sound Things, network infrastructures, protocols, and representations of sound-related information that enable services and applications for the communication of sound-related information in physical and/or digital realms*"

Sound Things in this context refer to networked devices that are capable of obtaining, processing, exchanging, or generating sound or sound-related information. Therefore, this research area encapsulates co-located multi-device audio experiences as well as a host of other applications including remote networked music performances [102], environmental sound monitoring [97], and music copyright enforcement [103]. However, like the IoT, the IoS has been largely driven from a computer science and engineering perspective where the predominant focus is on technological challenges such as the design of relevant systems and networks, and there is less of a focus on audience-facing applications and experiences. There are a number of smaller research communities with a more applied science outlook that often report on experiential factors of multi-device audio. The *New Interfaces for Musical Expression* (NIME) community [104], for example, commonly explores novel interaction techniques and how they can be applied in experiences through case studies, many of which involve multiple devices. Another relevant area is the field of *Human-Computer Interaction* (HCI) that covers multi-device or 'cross-device' interactions. Most case-studies, though, focus on the visual modality, such as multi-display applications. Finally, *Audio Augmented Reality* (AAR), that can be described as the addition of virtual or digital audio material to a real auditory environment [105], has seen increased interest due to the development of acoustically transparent 'hearables' [77]. Multi-device applications can overlap with AAR in cases where two or more devices augment the real surroundings or where one or more devices augment the output of another device. In an AAR taxonomy, these examples are encapsulated by *enchanted objects* or *overlay of additional audio information* categories [106].

As seen, multi-device ecosystems span many research disciplines. While having many different perspectives is ultimately positive for the development of these applications and experiences, it introduces a number of challenges. For example, without care it can lead to 'siloing' of research where stakeholders are not aware of others relevant work, increasing the risk of covering the same ground. Additionally, it makes the literature harder to access and digest, especially for newcomers to the space. These challenges are apparent within multi-device audio experience research, meriting the synthesis work in the first chapters of this thesis.

2.5.2 Experience design

Multi-device ecosystems and their applications are complex and encounter many additional challenges compared to their single-device counterparts; many of these stem from differences across the devices such as form factor, computing power, operating system and modes of interaction. To tackle this complexity, and to enable creation of more meaningful multi-device experiences, research has been conducted to increase our understanding of multi-device interactions and user behaviours, and more broadly to identify best design practice.

In Michal Levin's *'Designing Multi-Device Experiences'* [67], she presents the 3C's framework, which outlines three core design approaches for integrated multi-device experiences. These are:

- Consistent
- Continuous
- Complementary

The first is the 'consistent' approach which describes offering the same experience and features across all participatory devices, taking into account differences in device form factor and operating system. The second approach is 'continuous' which outlines how experiences can flow seamlessly between multiple devices, typically facilitated through a user profile system which allows different devices to share data to enable task continuity. The final is the 'complementary' approach which involves multiple devices acting as a collective ensemble to form the experience, typically spreading application features and media content across the devices. The experiences explored in this thesis fall into this category. The framework overall presents a straightforward, high-level starting point for developing new multi-device ecosystems.

As this research area has progressed, and new prototypes have emerged, efforts have been made to consolidate the work through surveys and taxonomies. An excellent example of this is presented by Brudy et al. [107] who report on design characteristics, applications, tracking systems and interaction techniques for multi-device systems. Despite there being no mention of audio throughout the entire review, many of the design approaches are likely to be applicable in the context of multi-device audio. In the HCI community, the concept of artefact ecologies [108] describes how artefacts (or devices) and their use influence the perception of other artefacts and the design of new ones. Tools have been proposed to better understand these relationships; for example, the human-artefact model is a framework for analysing how the needs of users match the affordances of a given artefact [109]. The ecology concept has been applied in the analysis

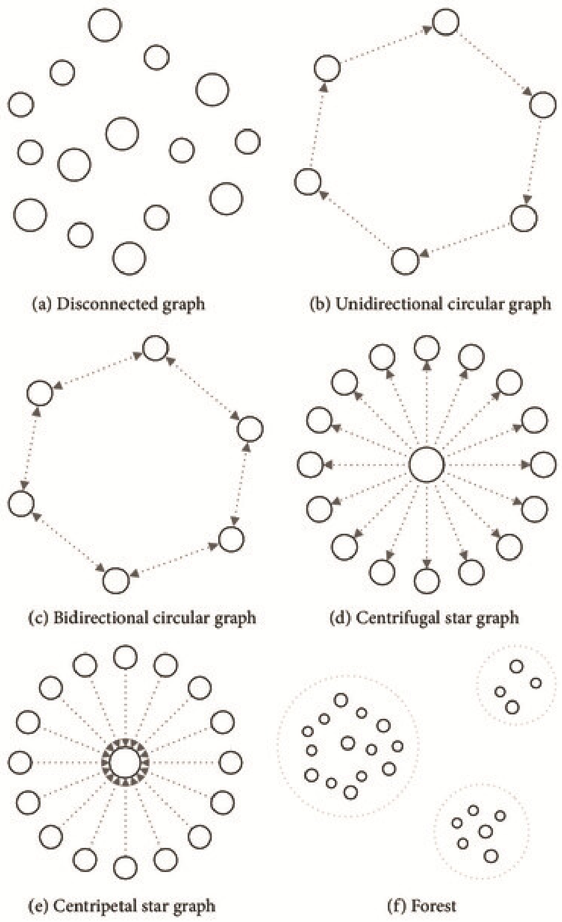


FIGURE 2.3: Abstract graph representation of the six interaction topologies. After [111].

and design of networked music performances to define the role of each involved human actor and audio device within an experience [110]. It connects each actor and artefact through three levels, *role*, *context* and *activities*, and poses appropriate questions related to the actors at each level to clearly define the purpose of each actor and artefact. Complementary to this, Matuszewski et al. [111] have proposed a set of six interaction topologies which describe different possible configurations of entities (humans, artefacts) in networked music experiences (Figure 2.3).

These help the designer to visualise the possible interactions between the entities within the experience. For these types of experiences in general, there is an additional level of complexity for composers where the instrument(s), the music, and the device interactions all need to be considered and composed [112]. In many case studies, applying prototyping and iteration techniques have been important to overcoming these challenges [113].

Bown & Ferguson have made a significant contribution to the multi-device audiovisual media domain in recent years. In [114], they propose the term *'media multiplicities'* to describe aggregate forms of media formed from individual media elements, where the combination of elements leads to new properties and characteristics in the aggregate form. They are described with four key characteristics: *spatial*, where the devices are distributed in physical space; *scatterable*, where devices can move around and be reconfigured; *sensing*, devices can interact and coordinate using sensors and *scalable*, where devices can be added and removed freely. The authors propose that designers of these experiences would benefit from simulation tools that can replicate and model the listening experience, and device interactions and behaviours. However, no tool seemingly exists as of now which enables virtual simulation and monitoring of a flexible multi-device audio environment.

Experience design practice for multi-device audio is developing largely through prototype development, and conceptual frameworks and ideas are beginning to emerge that capture and articulate knowledge gained through these prototypes. However, many of the current concepts are application specific, so it is unclear whether these apply across many different types of experiences.

2.5.3 Tools

The two core requirements for facilitation of modern multi-device audio experiences are modular device systems and flexible audio rendering. Where there is some additional visual stimulus, or the application is music-focused, there is usually a further requirement for control of synchronisation between devices. Researchers recognised that traditional audio production tools and workflows were not sufficient to satisfy these requirements, so efforts have been made to develop tools to fulfill them.

massMobile was one of the first dedicated multi-device technical frameworks which enabled audience device interaction for live music performances [115]. This interaction, however, was control-only, where each client could send control messages to affect parameters on the performer's device. Consequently, audio reproduction only took place over a single device. In the case of dynamic multi-device audio reproduction, there is an increased requirement for more detailed metadata to describe all audio assets and how they should be reproduced. This metadata authoring process was highlighted as one of the significant challenges faced by sound designers and producers. A number of tools were developed during the S3A project including *VISR* [59], a general-purpose framework for audio processing and reproduction suited for OBA, and *Metadapter* [116], a software framework for flexible and adaptable metadata authoring. Early S3A multi-device audio prototypes featured both of these tools where the *VISR* renderer was used to create the loudspeaker feeds and the *Metadapter* to dictate the high-level audio routing [61]. The reproduction setup consisted of a stereo loudspeaker system, with an array of small mobile speakers connected via cables, WIFI or Bluetooth.

Soon after, a decision was made to adapt the system to a web-based delivery format, in order to exploit the device-agnostic and audio rendering benefits of the Web Audio API. The outcome was *Audio Orchestrator*, developed by BBC R&D. The web-system architecture for *Audio Orchestrator* is described in [117]. A central server holds the media assets and the relevant metadata. A 'main' device creates a session and runs a placement algorithm to enable other devices to join, and to assign audio assets. All devices are wall clock synchronised to the main server and communicate using WebSockets and a Cloud-Sync solution developed during the 2-Immerse project [118]. The Web Audio API is used for audio rendering and playback and the user interface was developed using JavaScript. Different sections of the media content (i.e different chapters) are called 'sequences' and are defined by a JavaScript Object Notation (JSON) file which describes the objects and the important timings. Long running rendering objects are delivered using Dynamic Adaptive Streaming over HTTP (DASH) [119]. On top of this system is the production tool which allows producers to upload audio assets, program 'sequences' and adjust audio rendering rules for the allocation algorithm. Once complete, the producer then exports the project where it can be deployed on a server.

There have been other substantial efforts to develop systems and frameworks, based around the Web Audio API, for the creation of such experiences. A good example is *Soundworks* [120], a JavaScript framework which enables artists and developers to easily create distributed real-time music systems. The system architecture shares a few similarities with *Audio Orchestrator* such as server-client configuration and WebSockets for device communication; however, it has a number of features that are distinct for music applications. These are allowances for live monitoring and control of the whole system (e.g for composers/single performers) and plugin support for additional features such as audio synchronisation, file-system management and device position tracking [121]. A framework known as *Nü* [122] extends *Soundworks* by enabling a performer to control web-based audio processes on audience members devices through Open Sound Control (OSC) messages. A final example is *HappyBrackets*, a programming environment for creative coding of multi-device systems, that can also be used as a composition and performance tool for audio experiences [123].

These tools are relevant as their existence illustrates the range of application areas of interest and each contribute to the overall design space of co-located multi-device audio experiences in their own distinct way.

2.5.4 Use Cases

Applications and use-cases for the broader multi-device technology domain are as wide ranging as the number of research disciplines the field intersects. Some of these application areas include education [124], healthcare [125] and collaborative planning [126]. Further discussion of these broader applications is outside the scope of this work.

Loudspeaker arrays, where audio is reproduced over a number of spatially distributed devices, is no doubt the most common form of multi-device audio experience observed in the literature and wider creative industries. Within this category, however, audio experiences can differ significantly depending on the type of devices used, the audio content and the listening environment. Forming arrays of loudspeakers from audience devices is not a new concept. Prior to the internet, one of the first noted examples is that of Jose Maceda's city wide, collaborative soundscape *Ugnayan* which took place in 1974 [127]. The experience consisted of 20 audio tapes that were transmitted over 37 radio channels which were replayed over personal radios on the streets of Manila. While the performance only received mixed reviews at the time due to it's peculiar compositional form, it has a lasting legacy as an early demonstration of audience inclusion as part of an audio performance [128]. Since then, this concept has been explored thoroughly, and has been aided greatly by the uptake of personal mobile devices. A significant example with mobile phones as its focus, called *Dialtones*, was conducted in 2001 [129]. It consisted of the playback of several custom ringtones across audience members' mobile phones to form a dynamic bed of sound throughout a concert hall. Although the audience were passive listeners during the performance, the experience highlighted the capabilities of mobile phone technology in a creative medium which had not been seen or heard before.

With the recent developments in web technology, such as the Web Audio API, prototyping and delivering multi-device audio experiences has become notably easier. The most common use-case is distributed music, involving musical elements spread across many devices. Web-platform experiences began to appear in the early 2010s such as *echobo* [130] and *Fields* [131]. Many more experiences have arisen via development and testing of tools such as *Soundworks*, where numerous interactive music installations have been created and tested, exploring different interaction modes and system configurations [111]. While an abundance of research using the Web Audio API with mobile devices focuses on interactive music applications, the ideas and design principles could be applicable to other forms of audio-visual material.

BBC R&D have explored the use of everyday mobile and smart devices to form speaker arrays for broadcast media content, under the term Audio Device Orchestration (ADO). After promising initial demonstrations and prototypes during the S3A project, the *Vostok-K Incident*, a 13-minute audio drama, was designed and produced specifically for reproduction through an ADO system [132]. Since the *Vostok-K Incident*, BBC R&D have delivered four more ADO experiences through the BBC Taster platform using *Audio Orchestrator*; two audio dramas [133, 134], an orchestral music experience [135], and an augmented sports programme [136]. While these were primarily audio experiences, they also explored synchronised images, animations and video on one or more of the devices. Plenty of scope remains for experimentation in the broadcast content domain including testing of different media genres, longer form content, and introduction of different devices (such as smart speakers) and interaction modes.

Other unconventional forms of devices, such as DIADs mentioned in Section 2.4, enable experiences driven by more unusual forms of interaction, in contrast to common touchscreen interaction of everyday devices. DIADs have been used to in various art installations and performances, with various degrees of audience interactivity. *Bowls* [94], for example, extends a traditional game of lawn bowls through replacing the balls with spherical computing devices with loudspeakers. Another is *Bloom*, a light and sound installation consisting of 1000 distributed audio-devices configured over a wireless network in a public garden [137]. Audience members could walk through the space, interacting with the devices as if they were a natural element of the garden. Traditional sound design approaches were found to be inadequate for such large systems and a statistical perspective to sound reproduction was adopted, similar to granular synthesis [138].

The vast majority of the use-cases and experiences described so far have been designed to be enjoyed as social encounters with multiple listeners, where the output of each audio device can be perceived. Understandably, therefore, headphones are not commonly used as they act primarily to form a personal auditory bubble, deprioritising external sound. There are a few exceptions. A *Soundworks* installation named *ProXoMix* used mobile devices and headphones to create a music remixing experience where participants were assigned a single music track [111]. Using proximity sensing, participants could move closer to each other and hear a neighboring participants' track, allowing for live social composition. Another interesting example is silent discos or films [139], where all participants listen to audio over individual headphones. In this case, only an individuals audio stream can be perceived. Therefore it could be described as both one shared multi-device audio experience, or a collection of individual single-device audio experiences in a single space.

In recent years a collection of multi-device audio experiences have begun to explore the utility of hearables, that allow for personal audio and shared audio through 'acoustic transparency'. These are usually termed 'audio augmented reality' experiences or similar as they allow the surrounding auditory environment to be augmented with a personal audio stream, in the same way that visual augmented reality is the superimposition of digital imagery onto the real visual environment. Research has been conducted with acoustically transparent audio glasses to see how they can complement TV audio, where many benefits have been observed [140]. Another example using audio glasses is *Please Confirm You Are Not A Robot*, an interactive multi-user audio story that consisted as a series of mini-games [141]. Each game involved different forms of interaction with the device and with others, and there was both shared audio and personal audio, as there was a focus on a collaborative experience. Our understanding of how personal and shared audio can be combined is still limited, but with the increasing ubiquity of hearables, new use-cases are bound to be identified.

2.5.5 Research Methods & Evaluation

The improvement of multi-device audio systems and experiences is dependent on thorough evaluation. Due to their relative novelty and complexity, there are no established evaluation methods for multi-device audio systems and experiences. In addition, the multi-faceted nature of these experiences requires a multi-faceted approach, where quantitative and qualitative methods have been employed depending on the focus of the research. This section will cover the research and evaluation methods observed in the literature and highlight any key findings.

As part of their investigations in multi-device audio, BBC R&D have made efforts to evaluate the listening experience of Audio Device Orchestration (ADO) systems. Prior to the *Vostok-K Incident*, two studies were conducted to evaluate an ADO system, from both a quantitative and qualitative perspective. In [142], two ADO systems were compared with mono, stereo and 5.1 surround sound systems with respect to a Quality of Listening Experience (QoLE) rating where participants were asked to base their rating on any aspects of the sound. The results suggested that ADO performed comparably to stereo and the 5.1 surround system in the sweet-spot listening position, and performed better than 5.1 surround away from the sweet-spot position. Despite the experimental ADO setup reducing technical issues, such as device synchronisation, the results clearly highlight the potential of ADO to provide an accessible, high-quality immersive audio experience, especially in scenarios with multiple listeners where sweet-spot listening is not possible for everyone.

Francombe et al. [143] complemented the previous study with a thematic analysis of listener comments and feedback to retrieve positive and negative comments associated with the experience. The authors also wanted to examine the impact of ADO with different types of audio content. For this reason, three items were evaluated, a radio drama, a musical piece and a radio advert. In general, the ADO experience was viewed positively, primarily due to the increased sense of immersion when compared to the reference stereo audio system. The audio-drama in particular received positive reviews, with the music and radio advert receiving more negative comments. In terms of the physical system, some negative comments were received about the sound quality of the smaller speakers and the need for multiple devices; however, these were countered by positive responses such as "using everyday devices", and being, "a great way to have surround sound at home" [143]. A perceptual audio evaluation of ADO alongside four other spatial audio systems was reported in [144]. The authors used a Multiple-Stimulus Ideal Profile Method (MS-IPM) [145] to compare the systems against specific, audio system related attributes. On analysis, the results indicated that the dominant perceptual factor for the chosen systems was 'envelopment'. In terms of this attribute, ADO performed as well as the 5.1 surround sound system, which corroborates the results mentioned previously [142].

For larger public trials, such as the *Vostok-K Incident*, a questionnaire and interaction logging were used for evaluation [146]. The questionnaire provided some valuable insights into the audience's perception of the experience. Overall, the responses were very positive, with over 80% of respondents saying they enjoyed using phones as speakers and would use the technology again. When asked what the worst aspect of the experience was, the most common response was 'unbalanced device volumes'. Furthermore, when asked what they would like to listen/watch the most, the favourite choice was 'drama', with the least favourite being 'sport'. The user interaction data was useful for observing how users engaged with the programme. Francombe et. al. also reported on challenges in the production of the *Vostok-K Incident* and compared practices with traditional production workflows [132]. Unfamiliarity with the technology, altering of the reproduction system, and incompatibility with standard mixing tools (e.g. sidechain compression) were a few of the challenges, and led to an increase in production time.

From the data, the authors identified a number of areas of development including designing future ADO experiences to have maximum impact within the first minute of the program, in order to retain viewership and to maximise the experiences for a low device count. While there are many benefits to a large public-facing trial, such as having a large number of participants, there are some drawbacks which include the likelihood of selection bias and low resolution responses to questionnaire questions. The authors advise that this evaluation method should be used in conjunction with more controlled, standardized tests, like the ones reported previously [142, 143].

Another study was coordinated to find out where listeners would place loudspeakers in a domestic multi-device audio reproduction scenario [147]. Participants were found to typically place the loudspeakers on a functional or aesthetic basis, meaning positions were chosen that were considered convenient (e.g coffee tables or bookshelves). Therefore, it was uncommon to have a spatially balanced arrangement of loudspeakers, especially with larger loudspeaker numbers, therefore possibly reducing its usefulness for spatial audio applications, especially where sound source localisation is important. An additional interesting outcome was that participants tended to re-use the same positions, which would suggest that a system calibration may only be needed on first use.

When developing new products, services or experiences, involving users or similar stakeholders in the process is common to identify problems and to ensure user/stakeholder needs are met. For multi-device audio research, this could involve listeners or producers to identify technology opportunities and challenges, and possible use-cases. BBC R&D organised a series of worksops with audio professionals, based around developing content ideas and showcasing the *Audio Orchestator* [148]. Many positive outcomes were obtained over the workshops including identification of production challenges and ideas for new tools and features, a synopsis of the technical features and opportunities of the system and ideas for new applications for audio content. Shared experiences, interactive or gamified content and personalisation were three use cases that were heavily mentioned. Crucial to the development of new technology is input from target audiences. Due to this and the success of the previous workshops, a set of similar workshops were run with young people [149]. Using a similar structure to the previous workshops, the participants discussed attitudes towards audio technology, important themes for new technology and interesting use cases. While the workshops covered audio technology more generally, a number of constructive outcomes were identified. Most notably, young audiences valued positivity, togetherness and wellbeing as important in the application of new audio technology, which is in line with a number of core human values as identified by Kerlin [150]. Features of personalisation and gamification were also frequently brought up, which is consistent with the workshops with audio professionals [148].

Elsewhere, many of the experiences outlined previously have been developed using practice-based research methods. This largely involves iterative prototyping and evaluation through observation methods, surveys, questionnaires, interviews, and discussions. Early *Soundworks* demonstrations identified challenges using this approach which included audio latency across mobile devices of different models, and uneven output level of devices [120]. *Please Confirm You Are Not A Robot* was evaluated through various questionnaires and guided interviews and found that participants felt engaged and present during the experience and enjoyed the social aspect the most [141].

Some studies have taken a hybrid approach, using a combination of quantitative laboratory experimentation and qualitative methods. As an example, McGill et. al. [140] evaluated different methods for augmenting TV audio with acoustically transparent audio glasses through measured perceptual audio attributes and complemented this with a questionnaire to obtain data on attitudes toward the concept. Overall, the use of the glasses for mirroring the TV audio was rated highly for all media types and significantly improved the viewing experience. Most of all, participants mentioned that the audio glasses were preferred as a means to amplify dialogue or enhance spatialisation of sounds. Most of these features tested here would see usage from the participants given the adoption of this reproduction method.

Research methods used in multi-device audio experience evaluation vary depending on the specific aspect of focus. Controlled listening tests from perceptual audio research have been used to evaluate the listening experiences, whereas interviews and questionnaires have been used to evaluate tools and experiences. All these methods are complementary and are necessary to build a complete picture of how these experiences can benefit users/listeners. Summarising the evaluation outcomes, most are encouraging and demonstrate an interest in multi-device audio technology with a desire for more experiences.

2.6 Summary

This chapter has outlined the important research topics that are necessary to understand the underpinnings of multi-device audio experiences and, consequently, the original research reported in the remainder of this thesis.

A chronological recap of audio reproduction techniques and how consumption of audio has changed over time is presented, ranging from 20th century channel-based reproduction to cutting edge web-based and object-based audio technologies. How the latter facilitate multi-device audio experiences is described and followed by a review of how audio experiences are typically evaluated through listening tests and objective measures.

A large contributor to multi-device audio experiences is the modern day ubiquity and diversity of audio devices with an equally diverse set of features. This multi-device environment is presented, outlining the range of usage contexts and applications. The combination of devices and networking capabilities has enabled the formation of multi-device ecosystems and has catalysed new research, including dedicated research topics of the *Internet of Things* (IoT) and the audio-related subfield of the *Internet of Sounds* (IoS). These research areas are described and their contributions to multi-device audio experiences are highlighted.

Through this research, design principles and considerations are emerging for multi-device audio experiences that aim to aid the future development of these experiences by improving their quality and value to listeners. Many of the current design approaches are summarised. Specific tools such as *Soundworks* and *Audio Orchestrator* now exist that handle much of the technical complexity around creating new multi-device audio experiences, and have resulted in an expanding portfolio of use-cases. These tools are described in detail and how their features cater to different applications is discussed. Current use-cases and applications in the literature are reviewed to highlight how multiple audio devices can be employed in different contexts to achieve a wide range of possible experiences. Understanding and further development of these experiences requires evaluation. As these experiences are multi-faceted, many different approaches are used including workshops, demonstrations, lab-based technical evaluation, and real-world deployment trials. Examples are presented from the literature, and study outcomes are presented, providing an idea of the benefits of multi-device audio experiences.

Multi-device audio experiences have a number of benefits including increased flexibility of reproduction systems, easy provision of spatial audio, and enabling greater audio interaction possibilities. A large proportion of multi-device experience design research has focused on the visual modality, and audio domain work has been limited in scope to specific applications, such as distributed music performances. As highlighted in this chapter, there are many other use-cases that use multiple audio devices to create audio experiences. Therefore, synthesis work is required to understand to what extent these different experiences relate and whether a set of common characteristics or attributes underpin them. The first research questions in this thesis therefore arise as:

1. What is the current state of knowledge of multi-device audio experiences?

2. What are the design principles that underpin co-located multi-device audio experiences?

These are addressed in Chapter 3.

Chapter 3

Design Dimensions of Co-located Multi-device Audio Experiences

3.1 Introduction

Complementary multi-device audio experiences are becoming more common through the proliferation of mobile computing devices and the creation of bespoke multi-device audio production tools. As highlighted in the previous chapter, these experiences are beneficial as they can be an accessible, cost-effective method for the provision of spatial or immersive audio, and can also enable straightforward participatory interaction within experiences, using different modes of interaction. Over the last two decades, the way these devices can be utilised in the design of audio experiences has started to be established through case studies and demonstrations. However, little research has been conducted which reviews and consolidates these collective efforts, unlike in the visual modality [107]. To develop the audio application space further, more work is needed on experience characterisation and design to identify new experience opportunities, and to increase our conceptual understanding of these experiences. Furthermore, similar to many other developing research areas, disjointed use of terminology is also a problem, where many different terms are used to describe similar ideas or concepts. For example, *device orchestration* [132], *media multiplicities* [114] and *mobile multi-speaker audio* [151] can all refer to the use of multiple devices for audio reproduction. Finding some consensus on this issue would help to improve literature accessibility, and may lead to more fruitful collaboration across this discipline.

Consequently, this work addresses the following questions.

1. What is the current state of knowledge of multi-device audio experiences?

2. What are the design principles that underpin co-located multi-device audio experiences?

To achieve this, a survey was carried out to obtain a dataset which included descriptions of co-located multi-device audio experiences, which was then analysed to retrieve information on design aspects and attributes. A set of resulting design dimensions were generated and presented that capture the main distinct design elements of co-located multi-device audio experiences. These provide a semantic framework that aids the understanding and characterisation of such experiences. In addition, the application space was reviewed and mapped with the design dimensions, using a method in which use-cases can be categorised and new instances can be discovered. The survey aspects of the study were published in Geary et al. [152] and presented at the 2022 *AES International Audio for Virtual and Augmented Reality Conference*, and the study in full (Geary et al. [5]) was published as a journal article in *MDPI Applied Sciences Special Issue on Techniques and Applications of Augmented Reality Audio.* This chapter is organised as follows. Within Sections 3.2 and 3.3, the methodology for the survey, and the results are detailed. The design dimensions are presented and discussed in Section 3.4, followed by an analysis and review of the applications in the dataset in Section 3.5. In Section 3.6, the use of the design dimensions, and the overlap between the applications and audio augmented reality (AAR) applications is discussed. Finally, a summary of this chapter is provided in Section 3.7.

3.2 Methodology

3.2.1 Scope of the research

This thesis is concerned only with co-located experiences, where the active devices are within the same physical space or group of adjoining spaces as the listener/audience. This excludes typically remote applications such as teleconferencing and streaming parties. Furthermore, a subset of multi-device audio experiences, where one or more devices control the audio output of a single device (such as the use of an additional controller with a smart instrument [96], or streaming media from one device to another) are also outside the scope of this thesis. The scope of this work and the relationship between these different experiences is displayed in Figure 3.1.

During this period of work, attempts have been made to specifically and contextually define the word ‘device’. However, no satisfactory definition was found; they were either too encompassing or too specific. Nonetheless, the loose criteria for devices in this work is that they are standalone, reproduce audio, and offer some form of interface for interaction. For this reason, multi-channel surround experiences have not been considered as co-located multi-device audio experiences for this work, as they are essentially deemed to be a single device. A single loudspeaker may be used in isolation but they are designed to be used in multiples. A particular use case, somewhere near the definition boundary of these experiences, that was included in this work was silent films or discos. This type of experience differs from multi-channel audio experiences in two ways. Firstly, the loudspeakers in a multi-channel system are specifically designed to be used collectively, such that disruption of the number of speakers or the arrangement would likely diminish the experience. In the silent film/disco case, each set of headphones are independent, where the individual listening experience is approximately the same, regardless of whether there are 50 listeners or 500 listeners. Secondly, multi-channel speakers are purely static output objects with no intentional interaction patterns, whereas in silent films/discos, there is typically some form of device interaction, namely personal volume adjustment or selection of different radio/streaming channels. For other research, there may be a legitimate case for including or adding a different range of experiences to this work.

3.2.2 Survey

To address the research questions, a short survey was conducted to obtain information on both experiences, and production tools and systems. A survey was selected as the chosen methodology in order to obtain as large a sample size as possible, and was deemed more suitable for this purpose versus other qualitative methods such as interviews and focus groups. Researchers, technologists and creative practitioners were contacted through academic networks, mailing lists and social media to contribute their knowledge. The survey was deployed online using the *Qualtrics* platform [153]. To specify the types of experiences desired for this study, a set of inclusion criteria were defined at the beginning of the survey.

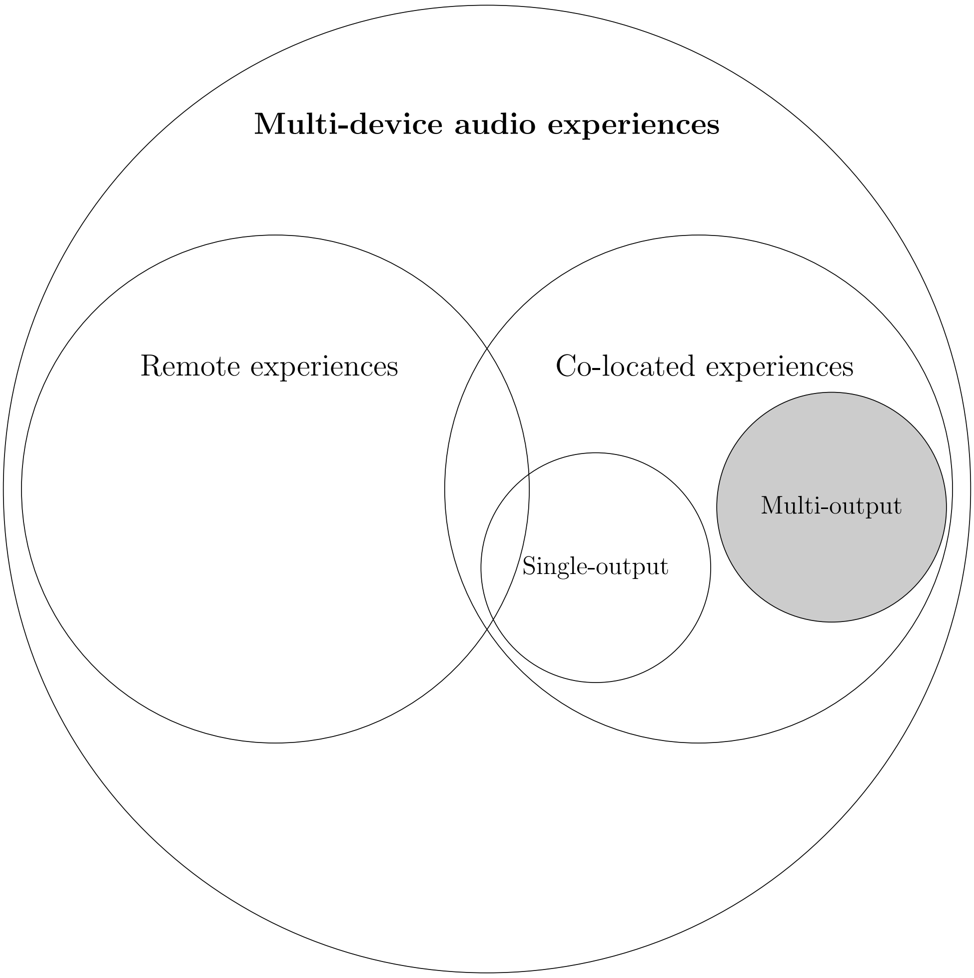


FIGURE 3.1: A Venn-style diagram displaying the types of multi-device audio experiences; the scope of the chapter and the wider thesis is highlighted in grey.

1. *The platform/experience must employ multiple devices with loudspeakers.*

2. *The audio content must be distributed across the devices.*

3. *The devices must be co-located in a single space, or group of adjoining spaces.*

'Platform' refers to the underlying technologies and systems which support and deliver multi-device audio experiences, whereas 'experiences' are the single applications of those technologies which include the audio content. The survey consisted of open questions with free-text fields where participants could describe the experience and its audio content, as well as multiple choice questions capturing the types of devices used, the modes of interaction, whether the number of devices is variable, and the importance of the devices in the ecosystem. Participants were able to respond to the questionnaire multiple times if necessary. The full questionnaire can be viewed in Appendix A.1.

3.2.3 Systematic literature search

To supplement the survey, a systematic literature search was completed to locate relevant papers and capture more instances of co-located multi-device audio platforms and experiences. This was achieved using the *PRISMA* methodology [154]. Both the *Association for Computing Machinery (ACM) Digital Library* [155] and the *Audio Engineering Society (AES) E-Library* [156] were targeted for the search as they were the two databases with the greatest number of relevant publications identified previously. For both databases, the following query string was used to search paper titles, abstracts and keywords. The asterisk is a wildcard character that can denote any number of characters.

(("multi-device" OR "cross-device" OR "distributed" OR "orchestra" OR "multiple devices") AND ("audio" OR "sound" OR "music"))*

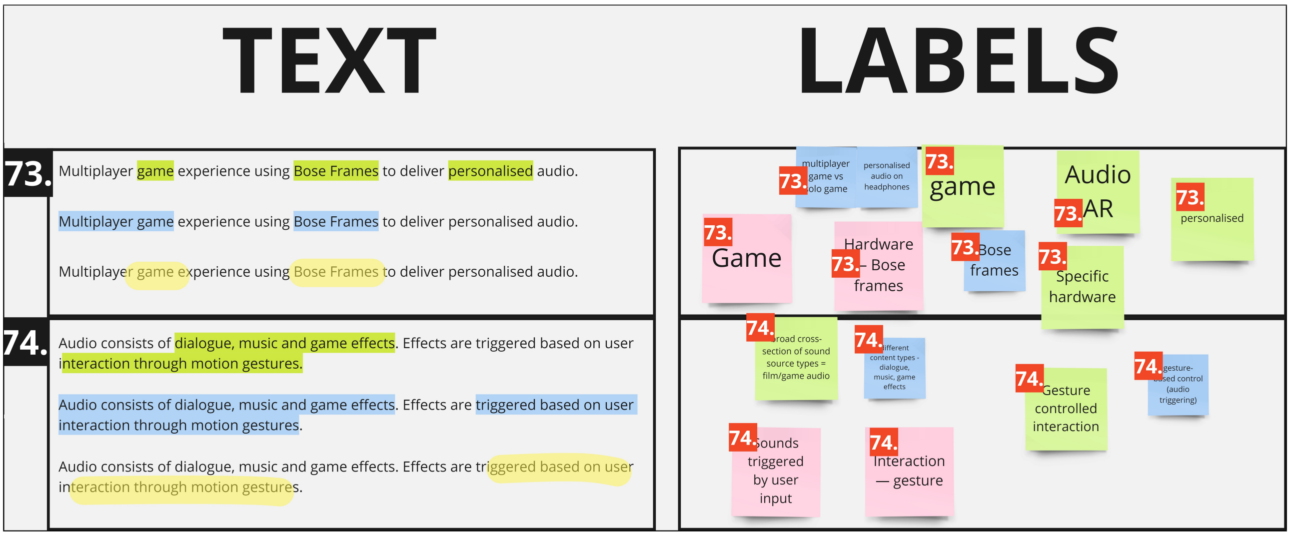
The search was constrained to journal and conference papers published between 2000 and 2021. Using the same experience inclusion criteria as the survey, the initial pool of 361 papers was preliminarily screened through reading the titles and abstracts. This screening stage resulted in 290 paper exclusions and 71 inclusions. These 71 papers were further screened by reading the full texts. Following this process, 11 papers were identified as containing application instances of multi-device audio platforms and experiences that fit into the survey inclusion criteria mentioned above. These instances were then added to the survey by the first author, and the resulting dataset was analysed (see Section 3.3).

3.2.4 Thematic analysis

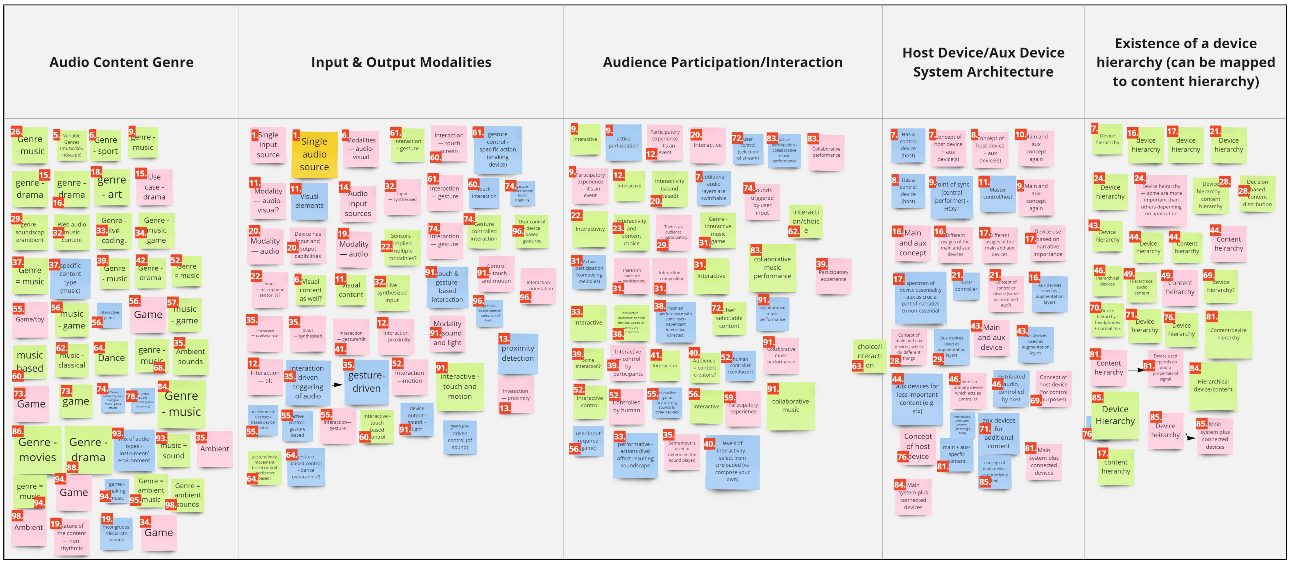
The free-text data was analysed using a thematic analysis approach, based upon Braun and Clarke's methods [157] and qualitative descriptive analysis [158], due to the descriptive nature of the data. The purpose of this analysis was to uncover defining characteristics for these types of experiences, and to use them to develop a framework for understanding and comparing these related experiences. The analysis was conducted in three distinct stages. In the first stage, a panel of three experts individually assigned written codes to each of the text responses (Figure 3.2a). Themes were generated at the second stage, which involved the experts discussing and clustering the codes over two 1-hour group sessions, using a shared *Miro* board (Figure 3.2b). This was an iterative process that started from a large pool of themes in a flat structure that trended over time toward fewer themes in a hierarchical structure. This was achieved largely through merging smaller themes. Over the group sessions there was not much disagreement between experts, however in cases where there was, a majority voting method was used to resolve them. In the final stage, three non-experts were recruited and given the same text responses as given to the experts to assign codes to. The non-expert codes were then compared and placed into the themes generated by the experts, as a form of validation exercise, to investigate whether the expert themes were sufficiently comprehensive.

3.3 Results

The survey received 42 responses in total (including responses from this author and those discovered during the literature search described in Section 3.2). Given the very specific



(A) Assigning written codes to each free text response.

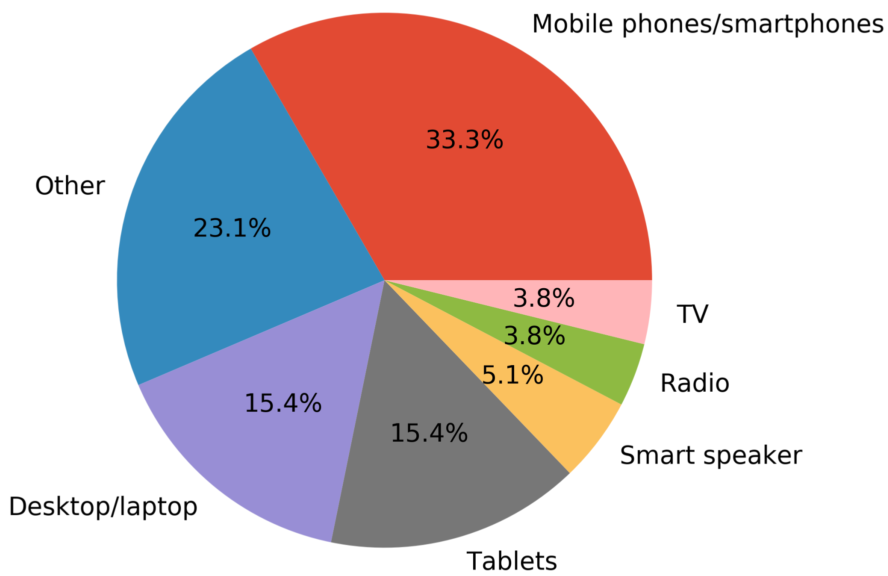


(B) Clustering the codes together into related themes.

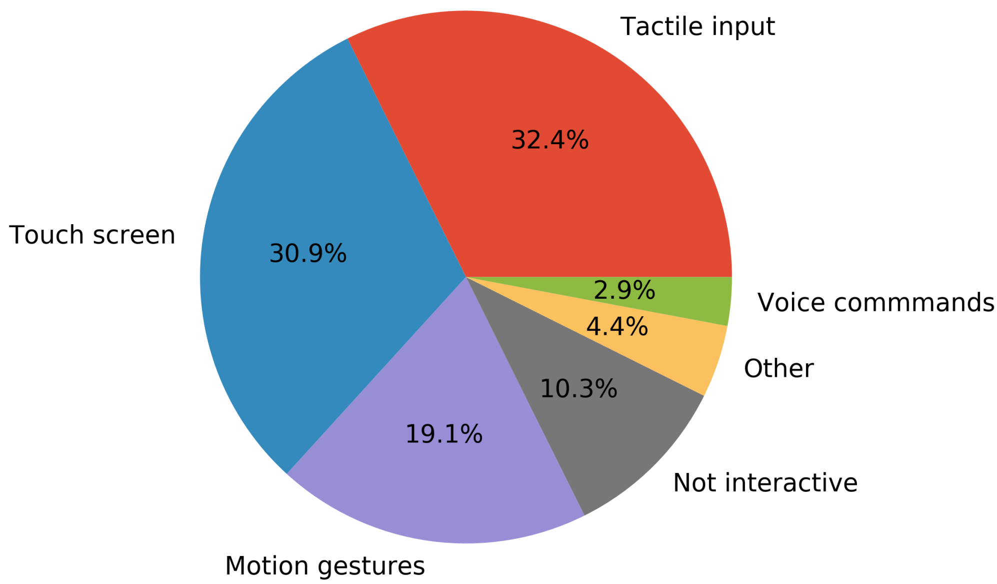
FIGURE 3.2: Assigning and clustering of thematic codes into related themes using *Miro*.

inclusion criteria, and with most examples limited to research, this was an acceptable number. Of these responses, 11 were classed as platforms and 31 were classed as experiences. The full dataset of platforms and experiences is available at [4]. All multiple choice questions and responses can be seen in the Appendices A.1 and A.2, and a full list of experiences in the dataset can be seen in Appendix A.3. Figure 3.3a shows the frequency of device types from the platforms and experiences described in the responses. An observable trend is that the occurrence of each type largely relates to the size and mobility of the devices, with mobile phones being the most common (33%) and TV/radio being the least common (8%). This result roughly reflects the relative market saturation of each type of device (with the exception of TV and radio) [2], but is also due to the increasing number of experiences created through specific multi-device audio production tools such as *Soundworks* [121] and *Audio Orchestrator* [117], which enable easy inclusion of mobile devices. The 'Other' category, receiving 23% of selections, included Raspberry Pi-based embedded audio devices, hearables, and cassette players. Audio-only devices, such as smart speakers, are far less common than those with a visual display, and are only observed within platform ecosystems such as Amazon's *Alexa*[159] and *Sonos*[160]. This can be attributed in part to integration challenges for these devices, where many of the modern experiences implement delivery through a web URL that can be typed or accessed via a QR code.

The frequency of various interaction modes is shown in Figure 3.3b. Tactile input and touch interfaces are most common (63%), followed by motion gesturing interaction (19%), both of which smartphones and tablets commonly possess. It is unclear at this stage whether the most common devices and interaction methods observed are due to a primary desire for these particular interaction methods, or whether the interactions are a consequence of the desired devices. The 'Other' category includes, but is not limited to, proximity-based control through the use of RFID or Bluetooth technologies. Out of all 43 entries, 16% of them are labelled as non-interactive. The number of devices was reported to be variable in 84% of entries. The data collected from the responses to this question are insufficient to determine whether the experience could vary the number of devices *during* or *between* instances of the experiences. Despite this, the results still demonstrate the flexible nature of these experiences and tend to display configuration agnosticism, unlike traditional multi-channel audio experiences. Finally, in answer to the question about device importance, 58% of responses indicated that the devices were perceived to be equally important, whereas 26% reported on some form of hierarchical device structure in the ecosystem. This highlights that the individual roles of each device in an experience may be an important consideration for experience design.



(A) The frequency of types of devices for all survey responses.



(B) The frequency of modes of interaction for all survey responses.

FIGURE 3.3: Pie charts representing the frequency of types of devices and modes of interaction for the multiple choice response data.

Through the thematic analysis process described in Section 3.2, 17 themes were generated (Table 3.1) which were captured under four parent themes: *devices*, *listeners*, *content*, and *other*. These themes were validated by a set of 300 non-expert codes, with 92% of the non-expert codes being represented under the 17 themes defined by the experts.

Parent theme	**Theme**	**Description**
Devices	Device position	The position of the devices
	Synchronisation	The existence of and the accuracy of synchronisation in use
	Device hierarchy	The existence of and the configuration of the device hierarchy
	Types of device	The types of devices in use
	Number of devices	The number of devices in use
Listeners	Number of listeners	The number of listeners present
	Listener position	The position of the listeners
	Listener role	The role of the listener in the experience, including the extent of contribution
	Social interaction	The type and extent of social interaction during the experience
Content	Content interactivity	The extent of interaction with the content, including device input modalities
	Content hierarchy	The existence of and the configuration of the content hierarchy
	Content distribution	How the content is distributed across the devices
	Content genre	What the audio content consists of i.e. music, soundscape
Other	Scale	The overall size and scale of the experience
	Flexible distribution	The ability for reconfiguration of devices and redistribution of audio
	Immersive	Experiences aiming to possess an immersive quality
	Type of experience	The format of the experience, i.e. performance, game etc.

Table 3.1: The 17 themes generated from the thematic analysis of the free-text survey responses, their descriptions and associated parent themes.

3.4 Design dimensions

This section describes a set of seven design dimensions, derived from further analysis and refinement of the themes from the thematic analysis, with additional reference to relevant literature. This process involved 3 steps. Firstly, the themes were analysed and related themes were aggregated into encompassing design aspect groups different from the parent themes. For example, *Listener role* and *Content interactivity* are combined into an *Interactivity* design aspect, despite being under different parent themes. In other cases, larger individual themes were assigned their own design aspect. In the second step, these draft design aspects were compared against existing experience design frameworks in the literature to identify corroboration or conflicts. The design aspects were then revised accordingly to improve the representation of my data, utilising established ideas where appropriate. Finally, once all the high-level design aspects were established, the design dimensions were formed by identifying all the distinct design choices associated with that aspect. These were identified both from my themes and existing frameworks where relevant. Figure 3.4 illustrates the process of generating the design dimensions from the original data. This analysis was conducted by the thesis author with some support from supervisors.

These dimensions aim to represent the key design considerations for co-located multi-device audio experiences. Within each dimension are a small number of categories which constitute the different design choices available for that dimension. For each dimension, examples from the literature are provided and how the dimension affects various aspects of audio experiences is discussed. A few of the following dimensions have been adapted from or are related to Brudy's cross-device taxonomy [107], as many of the important design considerations are applicable to multi-device experiences beyond audio-only. Nevertheless, any adaptations made are intended to make the considerations more relevant to audio experiences. These dimensions relate to:

- the degree of audio *synchronisation*,
- the social and environmental *context* of the experience,

- the physical *positioning* of the devices,
- the interaction *relationship* between audience members and devices,
- the degree of audience *interactivity* with the audio content,
- the roles of the devices and how they are *organised* in the ecosystem, and

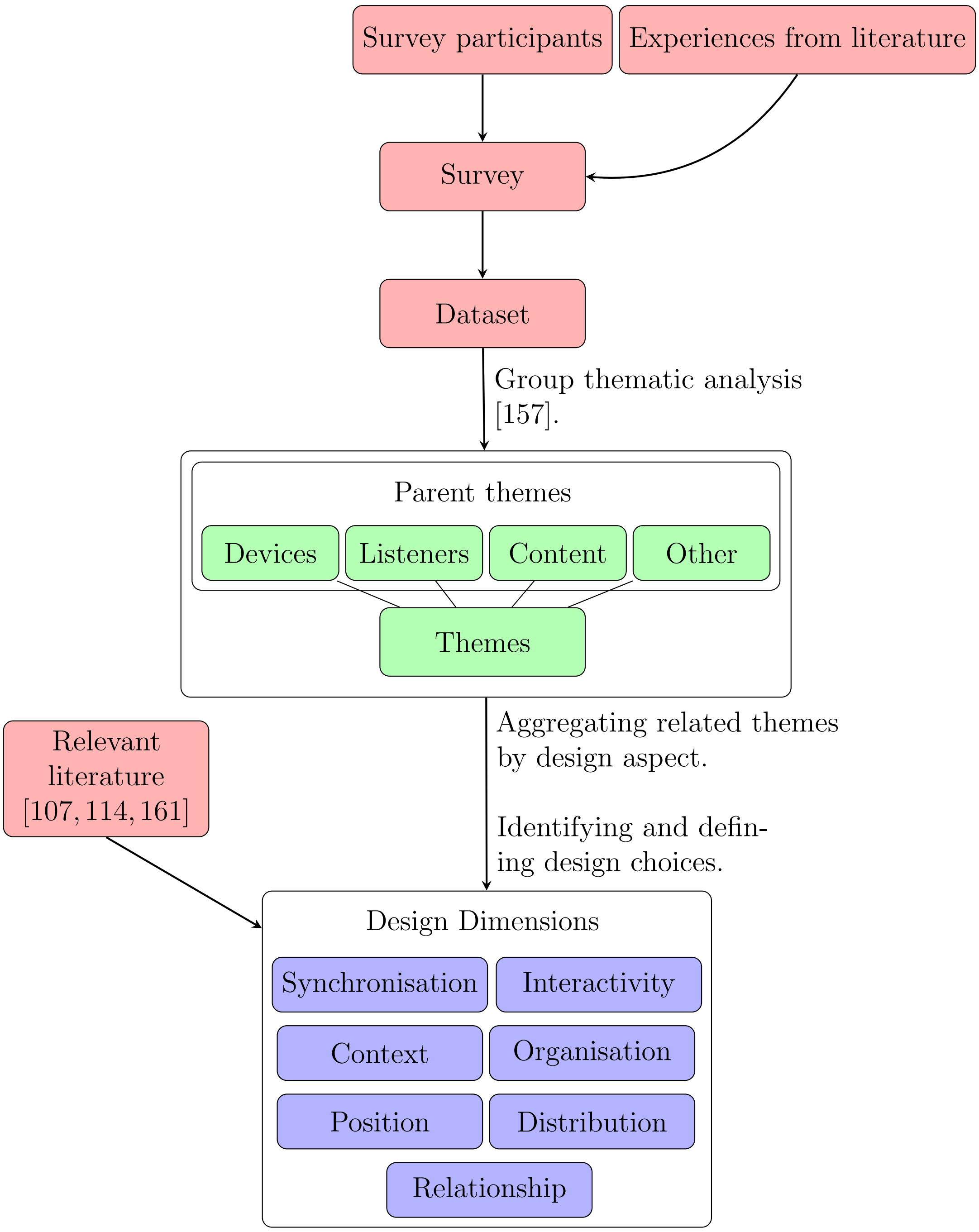


FIGURE 3.4: The methodology of obtaining the design dimensions from the original data.

- the *distribution* of audio content among the devices.

3.4.1 Synchronisation

Synchronisation is defined as the extent to which the devices in the ecosystem are synchronised in their reproduction of audio. This dimension appropriates and extends Brudy's *temporal* dimension [107] for synchronous, multi-device interactions, through three categories of synchronisation for audio content.

1. **Asynchronous** - In this instance, the devices are not synchronised in any form, acting with complete timing independence. The applications are limited, but asynchronous systems can be used for experimental art installations [162] and enchanted object experiences [94].

2. **Loose synchronisation** - Loosely synchronised audio can be achieved through artistic composition, without the use of device networking. The audio material is designed to be played at approximately the same time but without the need for precise timing. Consequently there is variability in timing between devices. Many early multi-device audio experiences were limited to this approach, which typically involved listeners pressing play on devices at the same time [163], or other manual triggering of audio sources [129]. Synchronisation between devices is usually in the region of 100 - 1000 ms.

3. **Tight synchronisation** - This category captures experiences where devices are connected and synchronised using a network. This approach is common in more modern experiences, and enables musical performance applications [164] and narrative-driven experiences such as audio dramas [132]. For such applications, experiences in this category require device synchronisation of $<$100 ms, or from word-level accuracy down to approximately frame-level synchronisation [165]. Variability in timing is much smaller than in *Loose*

synchronisation. Typical synchronisation solutions for these experiences obtain a sync accuracy of around 20 ms for heterogeneous devices after calibration [166].

3.4.2 Context

This dimension captures the social and environmental context of the experience, such as the scope for social interaction and the physical space it takes place in. The three categories generally relate to the number of people and devices present. According to Bown's media multiplicities and their "significance of numbers" [114], the relationship between number of loudspeakers and individual audio elements also affects how the audio is perceived.

1. **Personal** - Personal multi-device audio is an emerging class of experience involving a single listener and a small number of devices. The experiences usually have limited interactivity due to the impracticality of individually controlling multiple devices. Applications therefore tend to consist of overlaying additional audio content to a programme [136] or enhancing the spatial audio image through placing devices in different positions [167].

2. **Social** - Experiences comprised of small to medium sized groups of listeners with multiple devices are classified as *Social*. In this context, rich social interaction is possible and audio scenes with increasing complexity can be created, while still being possible to perceive the individual contribution of the devices in the system [168]. Experiences of this kind can take place in both small domestic environments and larger spaces, depending on the exact number of involved devices and listeners.

3. **Public** - *Public* represents the larger scale experiences with many listeners and devices. Experiences in this category commonly have more than 20 listeners and a large number of devices. In this context, communication or coordination is more difficult and commonly requires a central composer or figure to orchestrate the experience. Individual device contributions are typically displaced by the larger aggregate audio image. Many use-cases

such as art installations [162] and performances [128] fall into this category, and tend to take place in large public venues, including outdoor spaces.

3.4.3 Position

Position is concerned with the physical configuration and positions of the devices in the experience, and is similar to Brudy's *dynamics* dimension [107]. Three categories are identified.

1. **Fixed** - Device positions are typically static and pre-determined, similar to traditional channel-based speaker systems. The number of devices in the system tend to be known and device placements are optimised for the focal aspect of the particular experience, i.e. spatial audio image. For example, Tsui et al. designed a one-person orchestra experience [169], where device positions are carefully calibrated with motion controllers to enable audio control through motion gestures.

2. **Arbitrary** - Many modern tools and frameworks enable flexible, ad hoc device ecosystems where devices can join and leave at will. This ambiguity in the number of devices makes control of device positions difficult. With this flexible positioning, devices can be placed in arbitrary positions. Often, these positions are the same as the listener position if personal devices are used [130, 135]. However, this approach tends to remove control of the position of individual sound sources in the spatial audio image, unless device tracking utilising positional metadata is used.

3. **Dynamic** - In a few cases, experiences are designed with intentional movement of devices. The devices might be carried, worn or even thrown, possibly involving sensors such as accelerometers to affect audio playback parameters. This results in dynamic spatial audio experiences, which evolve over time [95].

3.4.4 Relationship

A number of different people-to-device interaction relationships exist with multi-device ecosystems. These are represented by three categories, similarly adopted from Brudy's *relationship* dimension [107].

1. **One-to-many** - The first category captures use-cases where an individual utilises many devices to expand on or augment a single device listening experience [140]. This is closely aligned with the 'personal' category of the *context* dimension.

2. **Multiple one-to-one** - Many experiences implement a 'use your own device' approach, which is an easy and accessible method for delivering shared audio experiences to a group. Each listener interacts with their own device, therefore known as a multiple one-to-one relationship [129, 131, 170]. This relationship lends itself well to interactive experiences, where each audience member can contribute equally.

3. **Many-to-many** - Other shared or collaborative settings where the number of devices and people are not equal can be associated under this category. In this case there is no link between the number of listeners and the number of devices, where individuals may interact with different devices over the duration of the experience, or with none at all. Examples of this are commonly seen in art installations [162] and music performances [163].

3.4.5 Interactivity

While traditional, passive media experiences, such as TV and film, still have their place, modern technology now allows for experiences with varying degrees of audience interactivity. Consequently, this affects and broadens the possible roles of audience members in an experience, where an audience member can be described as a *bystander*, *spectator*, *customer*, *participant* or *player* [161]. This dimension is split into three categories, a simplification of Striner's [161] spectrum of audience interactivity for entertainment domains.

1. **Passive** - *Passive* experiences capture more traditional media experience formats, where audiences are distinctly passive and have no influence on the media content. Many multi-device audio performances are of this nature [138], but also storytelling experiences typical of traditional radio or TV content [171].

2. **Influence** - This category includes application instances where the audience has some limited form of influence or control over the media content, but the experience creator retains all creative control over the experience. This could be personalisation through selection of audio assets [136], or other influence, perhaps through low impact manipulation of the audio signal [94].

3. **Create** - Highly interactive experiences where audience members could be better described as players or performers, actively creating the audio content of the experience, are situated in this final category [130, 172]. For these experiences, the experience creator passes some of the creative control onto the audience members.

3.4.6 Organisation

Ecosystems of multiple devices introduce new questions around device organisation, from technical, social and content delivery points of view. This dimension is focused on what the roles of each of the devices are in the delivery of audio in the experience. This is linked to the *interaction topologies* presented by Matuszewski et al. for mobile-based music systems [111].

1. **Non-hierarchical** - The simplest approach is to assign all devices to have equal responsibility in the experience. In this configuration, each device contributes to the audio

experience equally, and is regularly observed in social experiences, where personal devices are used [95, 130, 170]. This can be achieved in practice through the disconnected graph or circular interaction topologies [111].

2. **Hierarchical** - In some cases, it may be better to organise devices into two or more tiers, with differing responsibility and affordances. This is effective where the audio content can be organised into an importance hierarchy [132, 134], e.g. dialogue and sound effects in TV and film, or where mobile devices cannot adequately reproduce certain frequencies or amplitudes, and need support from an additional device [120, 131]. A common implementation of this design choice is assigning a single device as the 'main' device, and assigning other devices as 'auxiliary' devices [117]. These device types can be controlled separately and can be assigned different audio assets if necessary. Other configuration examples include star and forest interaction topology graphs [111].

3.4.7 Distribution

The *distribution* dimension captures different approaches for the distribution of audio to the devices in an experience. This can be linked to the previous dimension, where audio distribution can be determined based on the roles of each of the devices in the experience. There are three categories.

1. **Mirrored** - The simplest case is where the same audio content is allocated to every device in the system. This mirrored approach is useful when a simple amplification, or duplication of audio is required, e.g. in a silent disco or silent film, or in commercial multi-room speaker ecosystems such as *Sonos* [160].

2. **Distinct** - Here, the audio assets on each device are unique, where no two devices are playing the same audio. This approach is prevalent in interactive, music-based experiences where each device is allocated a unique instrument or sample, allowing each participant

to have an equal yet distinct contribution to the experience [164].

3. **Hybrid** - In this instance, devices are allocated a mixture of shared audio material as well as unique assets for individual or groups of devices. Audio allocation here is determined through a pre-defined set of rules, which can take into account many factors such as device numbers, types of devices and device locations. Examples of this include larger audio performances with many devices, where there are a number of clusters of devices playing different content [128, 129].

3.4.8 Summary

The original themes presented in Section 3.3 provide an overview of the low-level design aspects of co-located multi-device audio experiences; however, the number of themes and the complexity of relationships between them can be overwhelming, and therefore less practical. These design dimensions aim to provide a simplified framework for the core design considerations for these experiences by abstracting away some of the aforementioned complexity. As a result, the dimensions are likely best utilised in the early stages of experience design. A summary of the design dimensions can be seen in Figure 3.5.

3.5 Application patterns

The dataset obtained from the survey was re-analysed to explore the relationship between various applications of co-located multi-device audio experiences within the framework of the design dimensions. For this exercise, each experience was assessed against each design dimension and given a respective category. The experiences and their categories were then compared to see how the different design choices affect the type of experience. Using this method, four example application categories were identified, where the experiences in each category share a common combination of categories. For each experience in the dataset, their descriptions and corresponding application type, please refer to Appendix A.3. Additionally, in the dimension breakdown table for each of the experience types, the key dimensions are highlighted in bold.

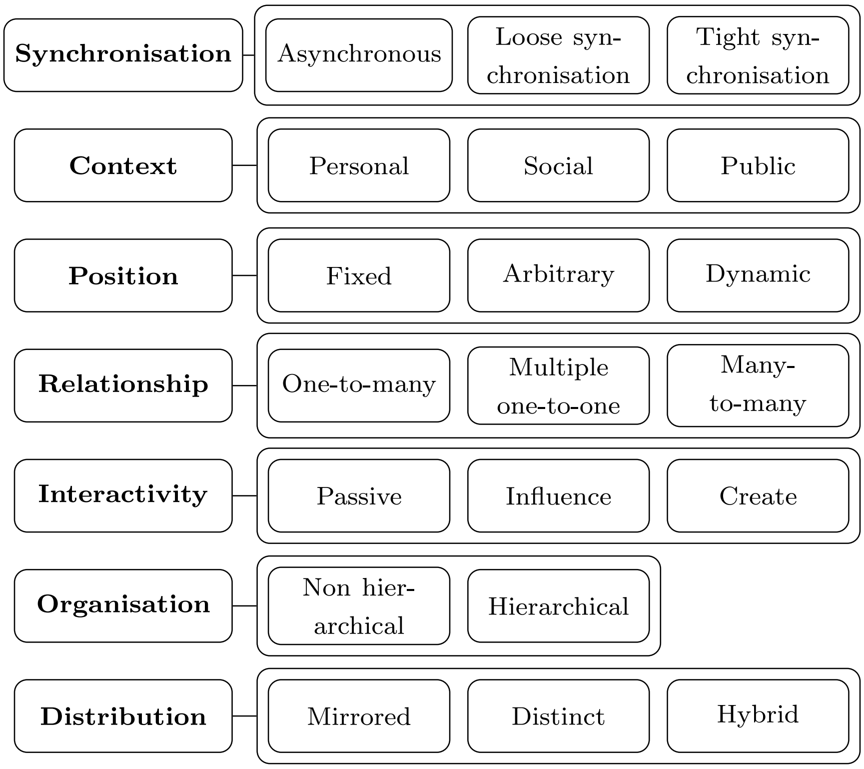


FIGURE 3.5: The seven design dimensions and their associated categories.

3.5.1 Public performances

Key attributes

Synchronisation	Asynchronous/Loose/Tight
Context	Public
Position	Fixed/Arbitrary/Dynamic
Relationship	Many one-to-one/Many-to-many
Interactivity	Passive
Organisation	Non-hierarchical
Distribution	Mirrored/Hybrid/Distinct

The largest category in the dataset consists of large scale performances/installations in public spaces. These experiences are akin to more traditional media experiences with largely passive audiences and they mark the early experimental years of multi-device audio as a medium, many occurring prior to the 21st century, before ‘personal’ devices [128, 163]. However this is not always the case, notably where new forms of audio devices are being explored for their creative potential [138].

3.5.2 Interactive music

Key attributes

Synchronisation	Asynchronous/Loose/Tight
Context	Social
Position	Arbitrary/Dynamic
Relationship	Many one-to-one/Many-to-many
Interactivity	Influence/Create
Organisation	Non-hierarchical
Distribution	Distinct

Interactive music contains interactive experiences that involve audience participation in music composition or performance. Most of the experiences exhibit either loose or tight synchronisation, as expected for the content medium, although a small number do display asynchronicity [172,173], where the musical content has less of an emphasis on rhythm and structure; for example, ambient music or soundscapes. Additionally, most of the experiences employ a non-hierarchical organisation of devices, ensuring that all participating listeners have the opportunity to make an equal contribution to the experience. Where a hierarchical structure is utilised, it is typically in the form of a supporting loudspeaker system to provide better bass frequency reproduction, seen for example in *Drops* [120]. Many of these experiences have been developed using the Soundworks framework [120], which provides tools for developing interactive experiences.

3.5.3 Augmented broadcasting

Key attributes

Synchronisation	Tight Synchronisation
Context	Personal/Social
Position	Arbitrary
Relationship	One-to-many/Many one-to-one
Interactivity	Passive/Influence
Organisation	Hierarchical
Distribution	Hybrid/Distinct

A relatively new set of experiences has emerged over the last decade, that tend to extend a conventional, domestic TV or radio listening experience with additional devices for more immersive spatial audio [132, 134, 136, 171], or to allow personalisation of content [140, 174]. These applications necessitate tight synchronisation, and device positions tend to be arbitrary as they are usually outside of the control of the designer. These experiences leverage the ubiquity of personal devices in the home, such as smartphones, tablets and even newer wearables such as the Bose Frames [78]. Many of these experiences have been created through the *Audio Orchestrator* production tool, which enables device synchronisation and online deployment for domestic consumption.

3.5.4 Social games

Key attributes

Synchronisation	Asynchronous/Loose
Context	Social/Public
Position	Dynamic
Relationship	Many one-to-one
Interactivity	Influence/Create
Organisation	Non-hierarchical
Distribution	Mirrored/Distinct/Hybrid

Finally, the last and smallest application category seen in the dataset contains shared, interactive audio games. These experiences generally involve listeners or players moving around with devices in a physical space and demonstrate audio augmented reality (AAR) characteristics through augmentation of a game with synthesised sound. Audio assets can be triggered and modified through positional and motion-based interactions. These experiences are limited to research studies for now, utilising bespoke, custom made devices, but represent a vision for future possibilities. Examples include Bown & Ferguson's work with DIADS [94], SoundWear [95], a study investigating the effect of non-speech sound augmentation on children's play, and *Please Confirm you are not a Robot* [141], a narrative-focused AAR game.

It is important to emphasise here that these categories are not totally distinct and there is some overlap between them; for example, *Interactive music* and *Social Games* share many of the same attributes. The purpose of this exercise was to highlight the most common design patterns and how they translate into the types of experiences observed within the obtained dataset.

3.6 Discussion

3.6.1 Dimension categories and combinations

When characterising experiences using these dimensions, it is sometimes difficult (and counter-productive) to assign one given category. Many modern experiences are built with technologies that enable them to traverse between categories of given dimensions, even during the experience. For example, some experiences could equally be assigned any of the categories in the *Context* and *Relationship* dimensions, due to the affordance of being able to add and remove devices. It may or may not be better to consider each of these variations separately; however, it is valuable to recognise the possibility that experiences can move along dimensions, and to understand the capabilities of the technologies chosen to build a given experience that may enable this behaviour.

The design dimensions also contain a number of invalid category combinations, of which five have been identified. These are outlined below, where the dimension category is given alongside the dimension name in brackets.

1. One-to-many (Relationship) & Social (Context)
2. One-to-many (Relationship) & Public (Context)
3. Personal (Context) & Multiple one-to-one (Relationship)
4. Personal (Context) & Many-to-many (Relationship)
5. Mirrored (Distribution) & Create (Interactivity)

Combinations one to four exist due to both dimensions containing categories which represent a single listener experience, and are incompatible with the shared experience categories of the comparative dimension. For combination five, if the audio content is dependent on the creative process of an audience member, it is highly unlikely that the audio output of each device will be identical. Knowledge of these invalid combinations can help steer users of the design dimensions toward more useful outcomes.

3.6.2 Relationship with Audio Augmented Reality

Another point of interest is the apparent application overlap between co-located, multi-device audio experiences and audio augmented reality (AAR) experiences, when comparing with Krzyzaniak's six categories of AAR [106]. Specifically, experiences in the *Augmented broadcasting* and *Social games* experience groups illustrate AAR through two forms. *Augmented broadcasting* experiences tend to exhibit AAR through the provision of extra context-relevant audio content. In contrast, *Social games* use an enchanted objects approach to AAR by adding extra purposeful sound to typically silent objects. For example in [95], tiny computing modules transform a bracelet into an interactive musical interface for children. Given the understanding that AAR can be defined as real-world experiences that are accompanied by additional layers of sound under either form mentioned above, there is an inherent requirement for two or more sound sources, where either some or all of those sources could be audio-capable devices. The question then is: when are co-located multi-device audio experiences considered as AAR experiences?

A good starting point is to evaluate the *Organisation* dimension, to consider the roles of the devices in the experience. In most 'non-hierarchical' experiences, where each device contributes to the audio equally, the experience content can be considered as a sum of the output of all the devices. Another perspective is that the experience cannot be 'complete' without the array of devices, so there is not augmentation but equal complementation. Conversely, in many 'hierarchical' experiences, the core experience can be listened to on a single device, and the addition of new devices can introduce enhancements such as spatialised audio and additional audio content, which is consistent with AAR characteristics. An example of this is demonstrated in the BBC trial 'Immersive Six Nations' where the original programme can be viewed on a single main device, but additional devices grant access to extra audio assets such as on-pitch sound effects [136]. This is applicable for *Augmented Broadcasting* use-cases but is less useful for *Social games*, as is illustrated by Soundwear [95] where the device organisation is 'non-hierarchical', yet the purpose of the devices is to add additional audio to an otherwise ordinary outdoor play experience. Thus, when trying to characterise new experiences, it is proposed that the *Organisation* dimension should be used as a guide to determine whether a co-located multi-device audio experience is an AAR experience, but consider the roles of the devices on an individual experience basis, as well as evaluating which form of AAR is exhibited in the experience.

3.6.3 Limitations

It is important to note that the dimensions presented here should not be treated as a fixed model, accounting for all possible design considerations, but should be considered as a helpful starting point for further discussion and refinement through highlighting the most important current design factors. As research in this area continues, it is possible that new categories of experiences will emerge that cannot be sufficiently characterised by the proposed model, and so the design dimensions will have to be revised accordingly. This is in part due to the fact that the dataset of experiences obtained in this study is not an exhaustive list, and it is likely that there are other experiences, academic, commercial or otherwise, that might be considered as multi-device audio experiences under the inclusion criteria used in this research. However, this dataset represents the largest available sample of these specific experiences to date.

3.7 Summary

Research on multi-device experiences is steadily maturing, although, only a small proportion of it is focused on how multiple devices can be used to facilitate more immersive and engaging audio experiences. In addition, research efforts concerning audio tend to be disconnected from each other, which is shown by the lack of consensus on the use of terminology. In this chapter, an attempt has been made to link these research threads and find common ground through an evaluation of current co-located multi-device audio experiences, driven by the following research questions.

1. What is the current state of knowledge of multi-device audio experiences?

2. What are the design principles that underpin co-located multi-device audio experiences?

A survey and structured literature review were used to produce a dataset of known co-located multi-device audio experiences and technologies on which a thematic analysis was performed to determine a set of seven design dimensions. These dimensions, *synchronisation*, *context*, *position*, *relationship*, *interactivity*, *organisation* and *distribution*, capture the most important design considerations of experiences of this type. Furthermore, these dimensions were used to characterise and compare experiences in the dataset, leading to the identification of four example use-case categories of co-located multi-device audio experiences. These are *public performances*, *interactive music*, *augmented broadcasting* and *social games*. More generally, this work provides an overview of research efforts from a practical use case perspective, as well acting as a guide for researchers and creative practitioners in furthering the development of these experiences. More work is necessary to test and validate the design dimensions and is likely to be an ongoing process. This can be achieved through assessing more experiences against the design dimensions, and using the dimensions to explore new edge-case and uncharted applications, by investigating unexplored combinations of the design dimension categories. The following chapter addresses this by exploring the use of the design dimensions as a creative tool for experience designers at the conception stage.

Chapter 4

Evaluating the design dimensions through co-design workshops

4.1 Introduction

In Chapter 3, a set of seven design dimensions were proposed that map out the design space of co-located multi-device audio experiences. These capture the key design considerations for these experiences and highlight the different choices available to a prospective creator or producer. The design dimensions can be useful as a tool for characterising and comparing experiences of this type, which is useful for understanding design patterns and making new associations. The aim of the work reported in this chapter is to investigate the utility of those design dimensions in the practical application of devising new ideas for co-located multi-device audio experiences. This work was completed in collaboration with consumer audio and electronics company Bang & Olufsen (B&O), who are interested and active within cutting-edge audio experiences research.

This work looks to achieve this aim through two research questions:

1. Are the design dimensions useful to professionals for exploring the design space of co-located multi-device audio experiences?

2. Are the design dimensions useful to professionals in the ideation of new co-located multi-device audio experiences?

This chapter is structured into the following sections. Section 4.2 provides a brief introduction to co-design and workshop methods. Section 4.3 describes the co-design workshops with audio professionals, and analysis of the workshop outputs and feedback is presented in Section 4.4. Finally, Section 4.5 summarises the chapter.

4.2 Methods

For the development of new experiences or products, there are a number of established concepts: user-centred design, co-design, and iterative prototyping. User-centred design is the idea that including target users throughout the entire development of an experience or product will achieve the best results by specifically addressing user needs and desires through continual feedback [175]. The related concept of co-design describes a process of collaboration during a design process where stakeholders are brought together to engage in a design task through sharing and combining knowledge [176]. There are a number of benefits associated with co-design; these are improving decision making, idea generation, and long-term user or customer satisfaction with co-design outputs [177]. The final concept of iterative prototyping is the process of repeatedly testing and refining an artefact or experience to quickly learn what does and does not work, and is popular with designers in software design [178] as well as multi-device audio experiences [112]. Workshops are a method that can facilitate the above concepts by bringing groups of people together to work on a design problem. To address the aims of this chapter in assessing the design dimensions applicability in forming ideas for new multi-device audio experiences, the workshop format was judged to be suitable. The workshop format has also been used for exploring use-cases and tools of multi-device audio in the past [148, 149].

Another question was how to incorporate the design dimensions into the workshops. Due to their categorical and modular nature, it seemed sensible to represent them physically as a set of playing cards. Physical cards have been used elsewhere as tools for ideation, such as in *Internet of Things* research [179], as there is evidence supporting this form as effective for idea generation [180]. Each dimension category has its own card and contains a coloured header that identifies the dimension it belongs to together with a description. The body of each card contains a description of the associated category. A picture of the design dimensions is displayed in Figure 4.1.

4.3 Workshops

Three workshops were conducted in total. The first two took place with the Acoustics and Research team at B&O in Struer, Denmark in December 2022 and used a hybrid facilitation format, where one facilitator led the workshop remotely over a video call, and the second facilitator helped the workshop participants locally. The third workshop was run in January 2023 and facilitated in person with the B&O Design team in Lyngby, Denmark. Four participants took part in each of the first and second workshops and nine participants took part in the third workshop.

4.3.1 Structure

Each workshop was 2 hours long and featured a brief introduction to multi-device audio experiences, followed by a warm-up activity and some idea generation exercises. The exercises used were influenced by previous workshops on multi-device audio [148, 149] and *Gamestorming* [181], a resource for co-creation tools. The exercises were tested and refined over two pilot workshops that took place in November 2022 with some volunteers from the B&O Acoustics and Research team. The workshops concluded with a short exit survey to obtain feedback. The workshop agenda is described in more detail below.

Introduction (15 minutes) A short presentation giving a definition of multi-device audio experiences and outlining the workshop structure.

Post up (15 minutes) Participants were given one minute to individually write down as many audio or non-audio devices as they could think of on post-it notes. Once the time had elapsed, all the post-it notes were shuffled on the table and participants were given another minute to repeatedly pick two devices at random and write down a multi-device experience using those two devices. This was repeated for as many combinations as possible until the minute was over. An example output from this exercise is shown in Figure 4.2.

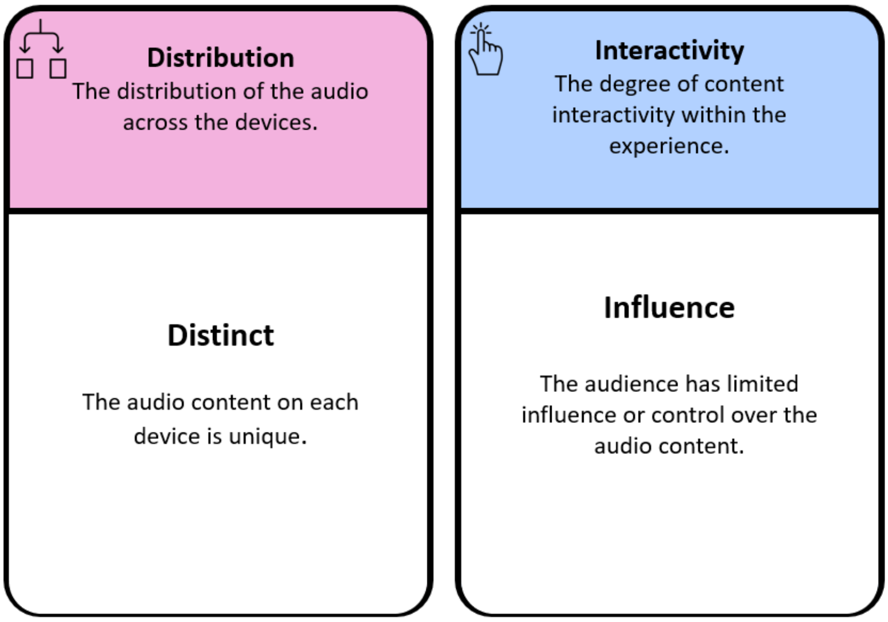


FIGURE 4.1: Two examples of the design dimension cards used in the *round robin* workshop exercise. Each card contains one dimension category, with the parent dimension labelled at the top of the card.

Round robin (20 minutes) This exercise took place in groups of 2 or 3. Each participant was randomly assigned a design dimension card (Figure 4.1). Efforts were made during workshop preparation to ensure that the dimensions were represented equally and that there were no repeated combinations. Participants were given 3 minutes to invent a multi-device audio experience that employed their dimension category and write it down as an annotated diagram on a sheet of paper. After 3 minutes, each participant passed their sheet to the next person in their group. Participants were given the opportunity to briefly explain their ideas to other group members. A second round of 3 minutes started, in which participants were asked to adapt the scenario on their new sheet to comply with their own category. This was repeated until all participants in the group had contributed to all experiences (see Figure 4.3). Each

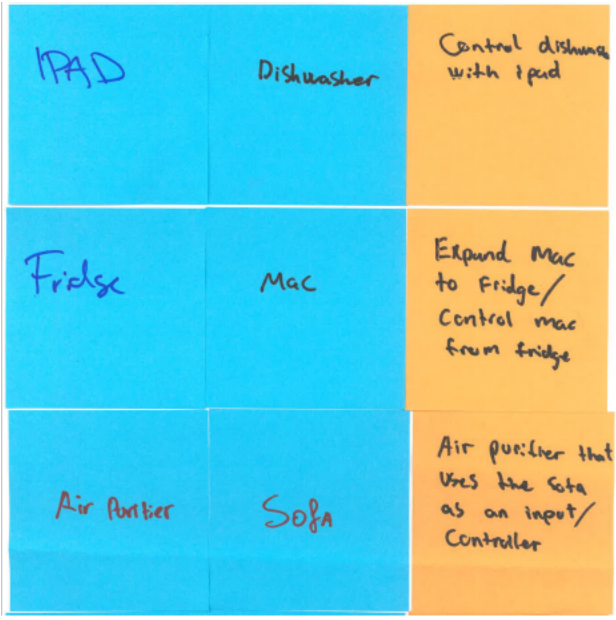


FIGURE 4.2: Example output from the *Post up* activity.

group then shared the resulting ideas with the rest of the workshop participants. Finally, each group was asked to select their favourite idea.

Break (10 minutes)

Mutation grid (25 minutes) Each group was provided with an empty 2D grid on A3 paper. The cells in the header row of the grid consisted of all categories for one design dimension and the cells in the header column were filled with categories from a different design dimension. Both of the dimensions were selected from those assigned to the group in the previous activity. Participants started by entering their chosen idea from the previous exercise into the appropriate cell on the grid. They then populated the rest of the matrix with multi-device audio experience ideas that used each specific combination of the design dimensions provided.

Marketing pitch (25 minutes) Each group was tasked with designing marketing material for a single multi-device audio experience idea of their choice. Participants were asked to think about who the customer is, what the most suitable B&O loudspeaker products are for the experience, and what the unique selling point of the experience is. Once complete, each group

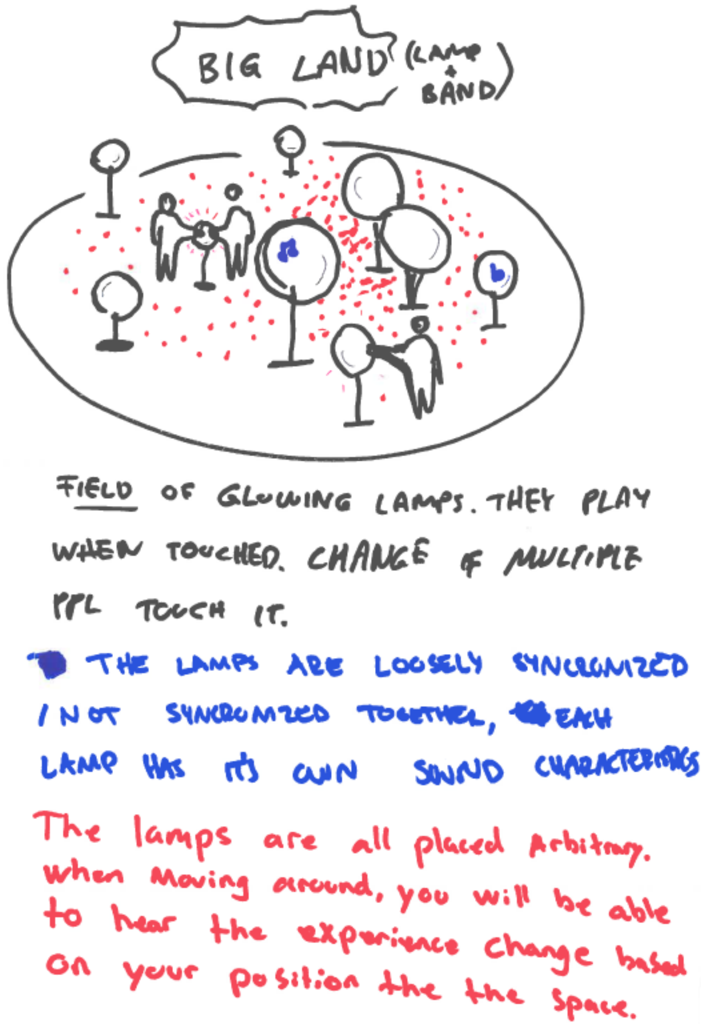


FIGURE 4.3: Example output from the *round robin* activity. The different pen colours indicate each participants contribution in the group. The dimension categories assigned for this idea were *relationship*: many-to-many; *synchronisation*: loose sync; *position*: arbitrary.

presented their materials to the rest of the workshop.

Debrief (10 minutes) A brief summary of the outputs and achievements was given by the facilitators and the participants were thanked for their participation. Finally, each participant was asked to complete a short survey.

Minor modifications were made to the third workshop's structure based on feedback from the previous workshops. These included replacement of the *mutation grid* activity with a second iteration of the *round robin* activity, the inclusion of a template sheet for the *marketing pitch* activity, and replacing the multiple choice survey questions with some free text alternatives. After each workshop, all the worksheets for the exercises were collected, scanned, and stored digitally.

4.4 Results

The results and their analysis are split across two subsections. The first subsection contains description and analysis of the workshop artefacts or 'outputs', and the second subsection contains analysis of the feedback obtained from the exit survey.

4.4.1 Output analysis

The workshop outputs consisted of the main exercise sheets and additional materials such as post-it notes. The data within these outputs was in the form of written text and drawings. A mixed content analysis method was used to analyse the data [182]. This process involved first describing any drawings, transforming the outputs into text form, followed by coding all text data into themes. The themes were then counted to observe their frequency.

4.4.1.1 Post up

The post-up activity resulted in 41 two-device combination ideas, across the three workshops. Twenty-two (54%) of these ideas contained audio and six (14%) of these qualified as co-located multi-device audio under the criteria outlined in Chapter 3. These particular multi-device audio ideas mainly focused on having tightly synchronised audio on both devices. Two common trends were identified in the remaining pool of experiences. The first was the frequency of *control* type experiences, where one device remotely controls some aspect of another device. The second was the presence of information sharing and messaging from one device to another.

4.4.1.2 Round robin and marketing pitch

The outputs from the round robin activity and the marketing pitch were combined and analysed. All outputs from the mutation grid activity were removed due to insufficient detail. Twenty-three experience ideas were obtained in total (one idea was removed as it did not adhere to the multi-device audio criteria). Each experience idea was reviewed and assigned thematic codes using an inductive coding approach. The most prominent codes were *interactive* (16), *shared* (14), and *device tracking* (11). The most common codes relating to audio content were *music* (9) and *soundscapes* (7). Other notable codes (counts ≥ 5) were *personal*, *choice*, *personalisation*, and *live*. Overall, these codes (see Table 4.1) are consistent with outcomes from a previous workshop with audio professionals on multi-device audio and potential use-cases that found that shared experiences, interaction and personalisation are inspiring features for future experiences [148].

Code	**Count**
Interactive	16
Shared	14
Device tracking	11
Music	9
Personal	8
Soundscape	7
Choice	6
Personalisation	5
Live	5
Remote control	4
Listener movement	4
AI	2
Performance	2
Game	2

Table 4.1: The frequency of thematic codes with a count > 1.

To further understand the prominence of *music* and *soundscape* applications, the input design dimensions associated with these ideas were compared to see if the repeated occurrence of particular design dimensions or categories were responsible. For both *music* and *soundscapes*, no patterns in these input dimensions were observed. This suggests that music and soundscape content for multi-device audio experiences may be less constrained by the design dimensions than other forms of content. However, this may also be explained by a simple preference across the workshop participants for music and soundscape experiences. In the round robin exercise, an additional observation was that the experience ideas were more varied within groups in the first two workshops where groups of two were used, as opposed to the third workshop where groups of three were used. This is likely due to the addition of a third design dimension in the groups of three, which caused the outputs to converge towards each other.

Comparing these experience ideas with the experiences collected in Chapter 3, there are many similarities. Of the four use-case patterns reported in that chapter, all workshop ideas match at least three of four key design dimension categories associated with these patterns. This is largely to be expected due to the fact that the ideas were constrained by the design dimensions, and that those experiences in Chapter 3 helped identify the design space represented by the dimensions. In the workshops, the dimensions were assigned randomly for each group. One way to obtain more unique experiences and use-cases would be to explicitly use uncommon combinations of the design dimensions, or to employ more constraints. Despite this, some novel concepts were identified during the workshops, particularly around device interaction and audio allocation. These included audio allocation tied to physical zones; branching-narrative stories where spatially separated devices guide the listeners through a space; synchronised audio and lighting effects; and mixed device systems that facilitate shared audio over loudspeakers and personal audio over headphones.

4.4.2 Workshop feedback

Feedback on the workshop was obtained in both written form through the exit survey and orally during the workshop. Overall, the participants enjoyed the workshop and found the design dimensions useful for understanding multi-device audio experiences and for generating ideas. The free text responses revealed that the design dimensions were good for "*breaking down*" experiences and "*shaping ideas quickly*" in the context of the workshop exercises. Despite this, there was some difficulty at times in understanding some of the dimensions and their categories where, in particular, the *relationship* dimension was difficult to interpret. On improving the workshops, participants requested more information on the design dimensions. Attempts were made to address this for the third workshop through a more detailed introduction of the design dimensions at the beginning of the round robin activity, however, feedback from this workshop highlighted that even more detail was desired.

4.5 Summary

The proposed set of design dimensions for co-located multi-device audio experiences provides a framework for describing and comparing different forms of these experiences through the definition of the key associated design considerations. However, it was unknown if the dimensions could be useful for exploring the design space and generating ideas for new experiences. This work investigated the following two questions:

1. Are the design dimensions useful to professionals for exploring the design space of co-located multi-device audio experiences?
2. Are the design dimensions useful to professionals in the ideation of new co-located multi-device audio experiences?

Three co-design workshops with B&O employees from the Acoustics and Research, and Design teams were conducted. The outputs from the workshops demonstrate that the dimensions are a useful reference for understanding these types of experience and, when presented as ideation cards, can be utilised in a valuable manner to produce a variety of different experience ideas. The modular nature of the dimensions was found to be especially useful for exploring and building new experiences incrementally, which was demonstrated in the round robin exercise — a favourite amongst the workshop participants. The experience ideas from the workshop largely followed conventional design patterns seen across already documented co-located multi-device experiences (Chapter 3). A greater focus on uncommon combinations of the design dimensions is advised to identify new types of experience. Feedback obtained on the dimensions during the workshop revealed that some of the dimensions were harder to understand than others, where particular difficulties were encountered with the *relationship* dimension. Further workshops using the design dimensions should provide more information and explanations prior to their practical usage.

The following two chapters describe two case studies which focus on the unexplored concept of concurrent use of personal audio devices and loudspeakers. The first study focuses on the creation of a novel multi-device audio prototype to demonstrate this concept, where a research-through-design approach is adopted to highlight the design methods and challenges. The second study focuses on the evaluation of the concept and describes a formal listening test to quantify the effects augmenting TV audio with active transparency modes. While the design dimensions are not central to these studies, they have both framed and informed them, and, the opposite is also true. Where the latter is the case, the design dimensions are commented on appropriately.

Chapter 5

Prototyping a novel co-located multi-device audio experience

5.1 Introduction

The workshops in the previous chapter highlighted the utility of co-design for generating new ideas for co-located multi-device audio experiences, where the design dimensions were found to be useful in facilitating this. Prototyping is still a popular design tool for exploring new use-cases and concepts not just in multi-device audio experiences but in wider human-computer interaction and product design. The case study presented in this chapter adopts one of the ideas from the workshops in Chapter 4 and documents the prototyping process undertaken to result in a novel co-located multi-device audio experience. Contributions are made through describing the planning, technical implementation, and audio production of the prototype where such contributions can be considered in future experiences and research.

This work was the outcome of a collaborative research project between the University of York and Bang & Olufsen, and was published in Geary et al. [6] and presented at the 2023 *International Audio Mostly Conference*. This chapter is structured as follows. Section 5.2 discusses the methods used to conduct this type of work, including research-through-design methods and approaches to sound design in multi-device audio experiences. Section 5.3 describes the process through which the prototype was realised, including details of the conceptualisation, technical implementation and sound design. Section 5.4 reports observations from an initial demonstration of the prototype. Section 5.5 presents reflections on the design dimensions including how the design dimensions have influenced the work or vice versa. Section 5.7 highlights the limitations to the work. Finally, Section 5.8 presents a summary of the case study.

5.2 Methods

5.2.1 Research-through-design

The investigation of novel concepts and ideas inevitably involves questions of design. Where creative practice is also involved it can lead to complications, where scientific and artistic processes can seem in conflict. Thankfully, there are well documented tools for these types of design problems. Research-through-design is an established method, and expresses the idea that the design process and the resulting artefact contribute unique insights on complex design oriented problems, resulting in expanding knowledge on those problems [183]. An important aspect of this method is keeping rigorous documentation during the design process which eases synthesis of research contributions. This method has been successfully used in multi-device audio research in the past [111, 131]. Consequently, the prototyping work in this chapter follows this methodology.

5.2.2 Approaches to multi-device sound design

The techniques used for audio production and sound design for multi-device audio experiences are varied and are specific to the type of experience or the type of content, in the same way that techniques differ between stereo and 7.1 surround; and music, film, and radio. Technical constraints further influence these techniques. For example, before networked synchronisation was available, designers used various techniques to embrace or circumvent the lack of synchronisation. For instance, orchestration of manual playback (asking participants to press play at the same time [163]), or using sound textures or soundscapes where tight synchronisation is not important [138]. Another example is using loudspeaker systems to support smartphone-based multi-device audio experiences to make up for the limited frequency response of smartphone speakers, especially in the bass frequency region [131]. Where suitable, using established production methods can provide a good foundation for expansion with multiple devices. Francombe et al. [132] identified that, in the case of broadcast media, starting from a high-quality stereo mix followed by gradual augmentation with the addition of more ad hoc loudspeakers has the benefit of allowing producers to use familiar tools and methods throughout the production process while slowly introducing complexity. However, for very large ecosystems of devices, traditional spatial audio design principles are not sufficient. In these cases, for example large art installations, unconventional techniques are more common. For the experience *Bloom* [138], a statistical approach to sound design was utilised where small variations in timing and pitch applied to each device output led to a granular textured soundscape, similar to the output of digital granular synthesis. The existence of these examples is a compelling argument for research-through-design methods through which many of these examples are reported, and will eventually lead to a consensus on audio production and sound design best practice being formed.

5.3 Creating the prototype

While most of the workshop ideas from Chapter 4 followed similar design patterns as existing experiences, there were some unique concepts that contained interesting combinations of particular devices, or explored new forms of interaction or types of audio content. One idea was selected as the basis for a prototype to explore one of these concepts more deeply. In addition, the design dimensions proposed in Chapter 3 have only been previously used as a descriptive tool to describe and compare experiences, but have yet to be used for the practical application of developing a new experience. Consequently, this prototyping work presents an opportunity to observe how the design dimensions can be of use during experience development. This section details the design and prototyping stage of the project including defining of the core experience features, describing the technical delivery system, and outlining the creative process used for developing the experience.

5.3.1 Idea selection

The idea to prototype was selected through a multi-step process. The first step included creating a shortlist of the best ideas from the workshops. In some cases, ideas were expanded upon or combined to create tangible experiences. Six ideas were proposed.

1. A collaborative soundscape experience, one device per listener, that passed audio objects from one device to another.
2. A multi-device orchestra with instrument allocation dependent on the specific location of the device.
3. An immersive car racing experience with audio from different cars and team voice communications allocated to unique devices.
4. An augmented audio story using earbuds in transparency mode in combination with additional external loudspeakers for story content.
5. A multi-room audio story with branching narrative.
6. An interactive art installation where listeners walk around a set of fixed devices and create sounds by touching them.

For each idea, positive and negative aspects were outlined. These considered a number of factors including access to content, novelty of use-case or interaction pattern, technical complexity, and available tools. After careful consideration, and through discussions with Bang & Olufsen colleagues, the *augmented audio-story* was selected for prototyping. It was selected as it identified a novel use-case, and could be achieved with current tools, already available content, with relatively low technical complexity. This experience stemmed from an idea from the round robin exercise, shown in Figure 5.1. The distinguishing feature is the combined use of headphones and loudspeakers, where speech and a soundscape are played over different devices yet combine into one audio experience. The following sections describe the prototype experience that was designed and explains the processes and techniques used in its creation.

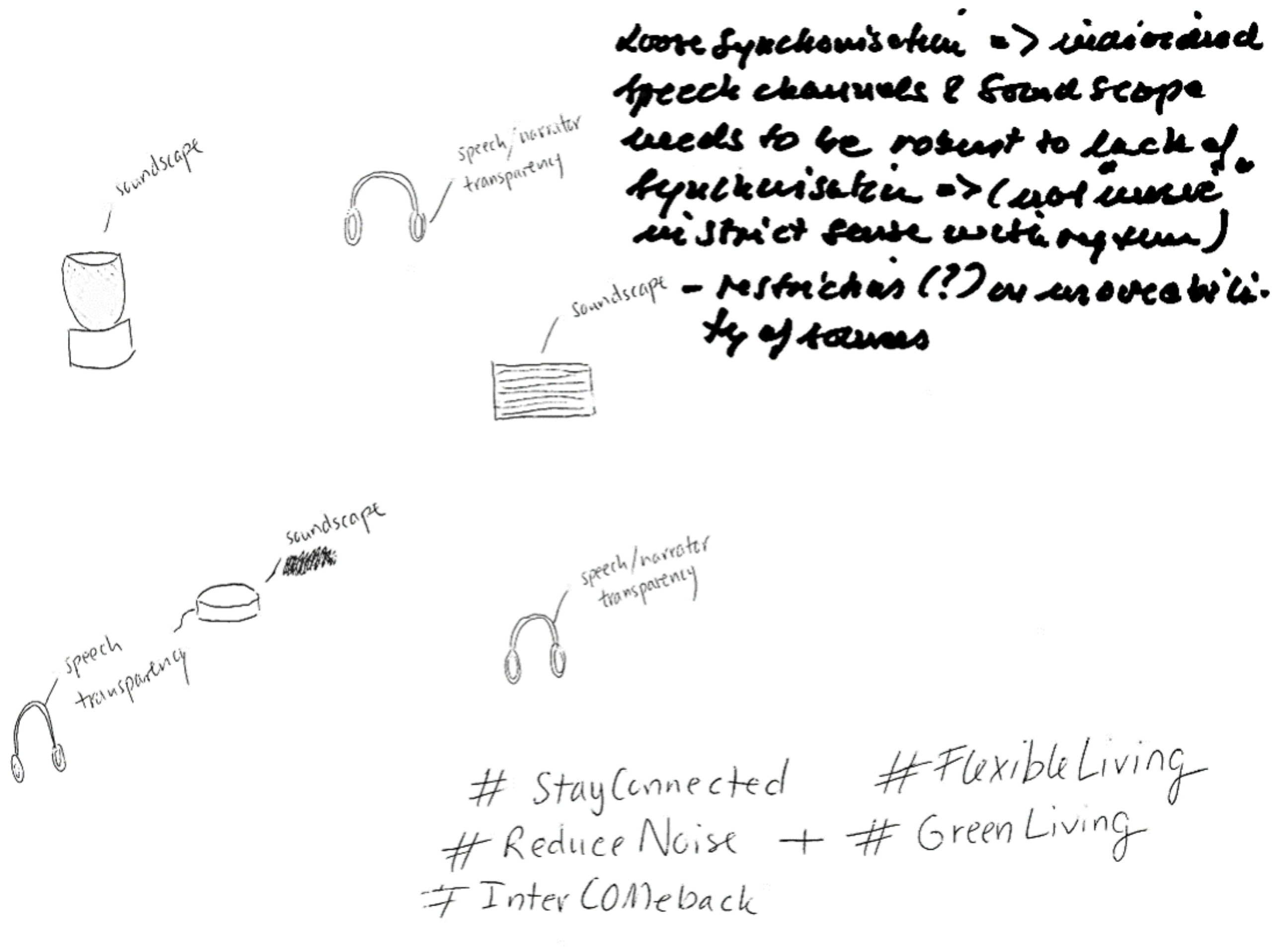


FIGURE 5.1: The main contributing workshop output for the prototype experience, proposing combined headphone and loudspeaker usage.

5.3.2 Concept

The main aim of this prototype is to explore the viability of using headphones or earbuds in transparency mode in combination with loudspeakers. Transparency, also known as 'hear-through', modes allow listeners to hear their external environment through the use of external microphones on the headphone/earbud. A potential benefit of the combination of loudspeakers and headphones/earbuds is that it can supply listeners with both personal audio and shared audio over different devices. The personal audio stream could provide accessibility features such as audio description or other additional content in social settings [140]. In other cases, the personal audio stream could be exploited in creative ways as part of a collaborative audio experience [141]. In addition, the interaction between loudspeaker audio and headphones can lead to interesting spatial effects and increased immersion [140]. What is less understood is how active transparency through hear-through performs for enabling the perception of shared audio content, particularly where it originates from a shared audio device.

The prototype concept comprised of a set of key features, which were used to guide the early development work. These are outlined below.

- Synchronised audio playback between headphones or earbuds and loudspeakers.
- Flexible audio track (or object) routing to any device.
- The content is experienced in a domestic setting.

As our selected idea involved linear narrative content, it was important for the devices to be synchronised to deliver it cohesively. Furthermore, in using consumer B&O products, it seemed sensible to ground the experience within a domestic environment where they are most typically used. Finally, flexible audio track routing was desired so the experience could easily be translated across different device configurations.

5.3.3 Content

The audiovisual content consists of a 5 minute excerpt from the science-fiction fantasy film *Cosmos Laundromat* [184]. This was selected as it contains a variety of scenes within a short time period, providing plenty of creative opportunity in a short demo. In addition, high quality audio and video assets are available under a Creative Commons licence.

The raw audio tracks (120) were firstly mixed down into 20 stems containing separate dialogue, music, and multiple ambience and effects tracks. These were then mixed in stereo as an initial starting point. A choice was made to focus mixing efforts on a core, predetermined configuration of devices first; this consisted of the earbuds and two mono loudspeakers, one placed to the front and right of the listener and another behind and left. The loudspeaker arrangement was somewhat arbitrary, and could be different in other implementations, but was chosen to provide some spatial characteristics with a small number of loudspeakers. This predetermined approach was used to maintain a practical monitoring environment and to create a solid foundation for adaptation. A similar approach was used for the multi-device audio drama *The Vostok-K Incident* [143].

As this was an audiovisual experience, the general mixing philosophy used throughout mostly followed traditional film mixing techniques where important narrative audio content such as dialogue, music, and on-screen Foley tracks were allocated to one or two central loudspeakers, whereas background content (e.g., atmosphere sounds) was allocated to others [185]. In the interest of ensuring audio and visual congruence, the earbuds largely assumed the role of the centre channel in a surround sound format, as the loudspeakers were positioned away from the video-displaying device and/or outside the listener's cone of vision. A number of decisions were made specifically to address particular aspects of the combined earbuds and loudspeaker format. For instance, some of the ambient background objects were mirrored in the earbuds, as employed by McGill et al. [140]. Without any mirroring, the listener (at least with some experience) would be more able to distinguish between the direct earbud output and the indirect loudspeaker output. In this case, by mirroring some of the atmosphere audio tracks on the earbuds, a desired blending effect was achieved between the outputs of the earbuds and the loudspeakers. Perceptually, this increased the presence of the atmosphere sounds while preserving the externalisation of those sounds provided by the loudspeakers. In addition to this, some effects tracks (e.g., flies in a jungle scene) were allocated to a single earbud and the loudspeaker positioned on the opposing side to the earbud selected. In particular instances such as this one, an interesting dynamic spatial effect was achieved where the two instances of the audio signal had different tonal and perceived externalisation qualities. However, this approach was only suitable for off-screen effects where audio and visual congruence was not an important factor.

To compensate for the spectral modification that the earbud and transparency mode applies to external sound sources, some corrective EQ was applied. A 2-3 dB high shelf at 1 kHz was used to boost the high frequency region that was attenuated by the passive occlusion of the earbuds. An additional +3 dB peak-notch filter was applied in the 200-500 Hz region to oppose perceived cuts to that region by the transparency mode. These settings were determined through trial and error while in production.

5.3.4 Technical implementation

A bespoke proof-of-concept demonstrator system was designed to explore the concept (Section 5.3.2). The core functionality of the demonstrator is implemented using a *Max* [186] patch running on a laptop. This patch controls audio and video file playback and audio routing to output devices defined using an aggregate audio device and are connected to the laptop either wirelessly over Bluetooth or WiSA, or a using a wired connection. All the audio tracks are loaded in as a single multichannel WAV file, which is subsequently unpacked into the individual signals before being routed into a 2D matrix. The matrix dimensions are defined where the number of rows corresponds to the number of audio outputs, and the columns correspond to the individual track inputs. Each cell in the matrix can be toggled manually by the user using a *matrixctrl* object to route the column input to the corresponding row output. Within the *Max* patch there is also a configurable delay parameter which can be used to manually synchronise the wired and wireless devices.

Audio allocation can be controlled manually within *Max* using the *matrixctrl* object directly or by defining audio object and device metadata through *Audio Orchestrator*. The latter approach allows allocation 'behaviours' to be assigned to audio objects; these behaviours are then interpreted by an allocation algorithm which decides which device(s) an audio object should be assigned to. A custom build of *Audio Orchestrator* was created that outputs the result of the allocation algorithm as OSC data, which can then transported over UDP to the *Max* patch. This data is then parsed using a *Max* JavaScript object and used to control the *matrixctrl* object and the routing of the audio. A system diagram for the demonstrator is presented in Figure 5.2 and an image of the *Max* patch is displayed in Figure 5.3.

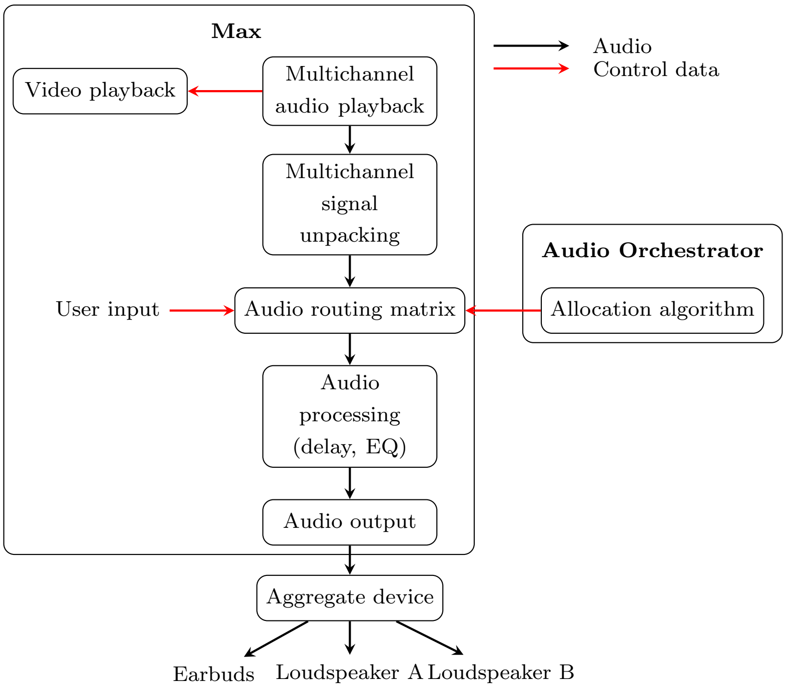


FIGURE 5.2: System diagram. The *Max* patch controls the video and audio playback. Video playback timing is driven by the audio timing. Audio routing can be controlled externally through *Audio Orchestrator* or through user control. The audio is sent to a Mac OSX aggregate device and then on to each device.

5.4 Demonstration

The prototype experience was assembled and demonstrated to a small group of six audio professionals at B&O. In this instance, the devices used were a Beosound A1 (positioned front-right of the listener), a Beolab 20 (positioned rear-left of the listener), and two pairs of Beoplay EX earbuds. All of the devices were connected to a laptop running the *Max* patch. The A1 was connected using a USB connection and the Beolab 20 was connected using a WiSA USB dongle. Both pairs of earbuds were connected directly to the laptop via Bluetooth. A computer monitor was used to display the video. The setup was assembled around a kitchen table in a replica open-plan home environment. The participants were asked to sit down in front of the computer monitor and were each given a pair of earbuds in transparency mode.

Overall the experience was received positively. Listeners especially appreciated having the dialogue routed directly to the earbuds, commenting that it could, for example, improve speech

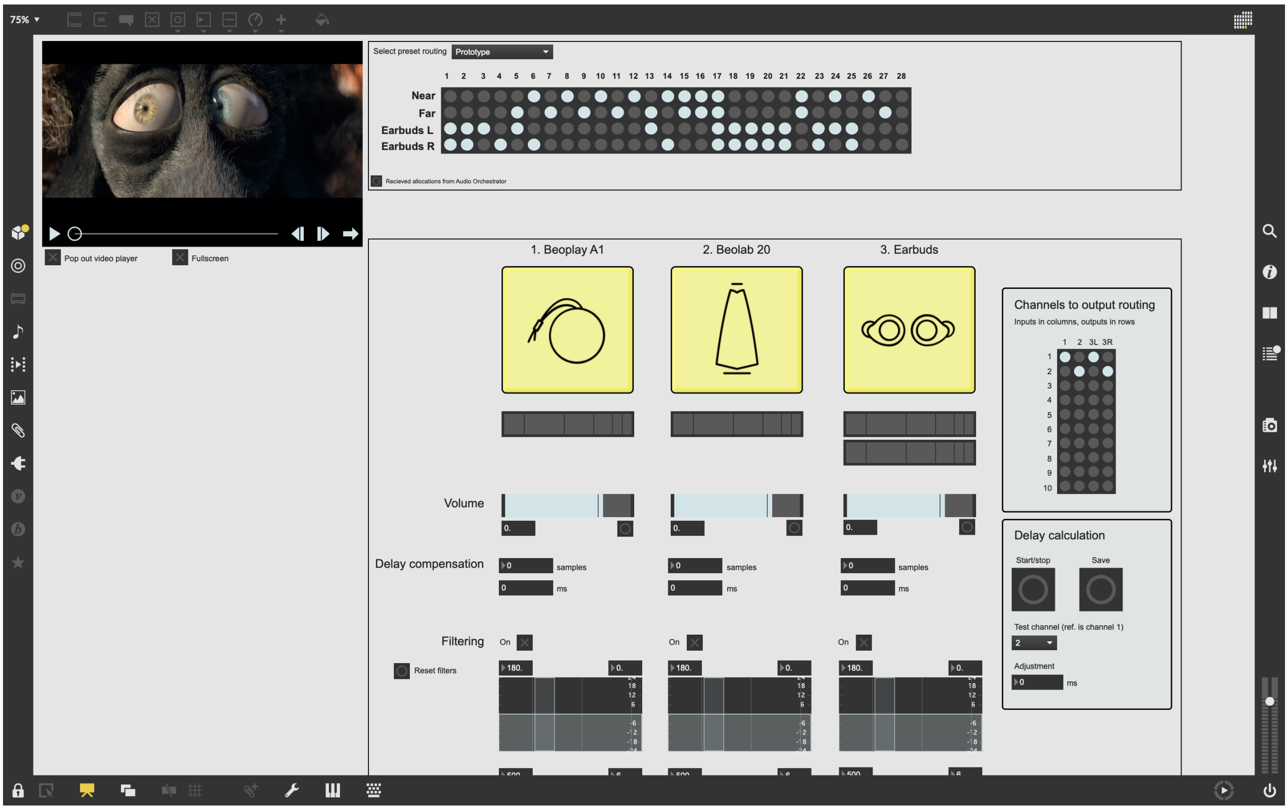


FIGURE 5.3: A screenshot of the main *Max* patch for controlling the prototype experience. The top of the patch contains the video and the audio allocation matrix. In the center is the three device strips with volume, delay and filtering options.

intelligibility compared to if the dialogue were played over loudspeakers in a reverberant space. However, some drawbacks were highlighted, notably the change in sound quality due to listening to loudspeakers through the transparency mode of the earbuds.

5.5 Using the dimensions

The final prototype experience can be characterised using the design dimensions as follows.

- *Synchronisation*: Tight synchronisation
- *Context*: Personal or social
- *Position*: Arbitrary
- *Relationship*: Many-to-many

- *Interactivity*: Passive
- *Organisation*: Hierarchical
- *Distribution*: Hybrid

It became clear over the design process that different dimensions become relevant at different stages of the prototyping work. For instance, the first three dimensions (*synchronisation*, *scale*, and *position*) were defined at the initial concept stage and provided first necessary constraints. The dimensions which have an effect on the audio content (*interactivity*, *organisation* and *distribution*) were considered during the audio production stage of development, and can be defined once the audio content and the creative intent to render the content is known. These latter dimensions are especially useful during this creative phase as they present a number of defined options that the designer can test to see which has the most benefit. In this experience, for example, a *unique* distribution was used initially, but on evaluation of the other options, a *hybrid* approach was ultimately taken as it resulted in the best listening experience for this content.

5.6 New design dimension?

Throughout the prototyping process, a recurring thought was that this idea of shared and personal auditory spaces or 'bubbles' being explored with the prototype was not currently represented within the proposed design dimensions from Chapter 3. It was noted at the end of Chapter 3 that as new technologies and consequent experiences emerge, the dimensions should be reviewed and updated where necessary to retain relevance. Thus, after some consideration it was decided to propose a further dimension to capture this design aspect. It is defined as follows.

Perception: The *perception* dimension expresses the different ways in which audio can be perceived by listeners in the experience through defining the characteristics of the listeners' auditory bubble. It is primarily determined through the types of audio devices in use. Three categories are proposed.

1. **Private** - Audio is delivered and experienced individually where adjacent listeners are not able to perceive the audio allocated to that listener. In this case, audio content is delivered using personal devices such as headphones or other hearables.

2. **Shared** - The audio is delivered from all devices to all listeners where each listener perceives the same aggregate audio scene, facilitated through devices with loudspeakers. In this case the auditory bubble is collective, involving all listeners.

3. **Mixed** - Mediated through acoustic transparency in personal devices, it is possible to have both personal audio and shared audio within the same physical space. This transparency allows a listener to perceive a shared auditory space while retaining a personal auditory bubble, unperceived to others. This is an important aspect of emerging audio augmented reality applications and experiences.

The most familiar example of a co-located multi-device audio experience with the 'private' category would be silent discos, where each listener is largely isolated in their own auditory bubble yet enjoys the collective physical and visual experience of dancing with others. Most of the observed multi-device audio experiences fall into the 'shared' category where loudspeakers are used. Other than the prototype experience developed in this chapter, *Please Confirm you are not a Robot* [141] and augmenting TV audio with audio glasses [140] are examples that use the 'mixed' *perception* category. These three examples all demonstrate *audio augmented reality* but utilise acoustic transparency for different purposes. In *Please Confirm you are not a Robot*, the acoustic transparency is primarily used to enable communication and collaboration between participants. Whereas, in McGill's work with audio glasses, the personal device is used to augment the external (or shared) TV. Finally, the prototype explored in this chapter demonstrates the opposite: using external (or shared) loudspeakers to augment the personal device. Each of these approaches are valid where the choice selected is dependent on the purpose and creative vision of the individual experience. For future co-located multi-device experiences that may consider personal and shared audio devices, it is hoped that this new dimension can help illustrate the design choices available and highlight the interplay between combined personal and shared device use.

5.7 Limitations

Modern transparency modes, while labelled as such, are not completely transparent; many introduce artefacts to the signal including colouration, level bias, and latency [187]. In the demonstrator, equalisation of the loudspeaker signals was used in attempt to counteract this; however, no amount of equalisation could accurately reproduce the real loudspeaker output, as perceived by unobstructed ears. The equalisation settings used were determined through informal listening and trial and error; it would have been preferable to measure the earbud response in order to make better informed decisions about necessary equalisation. More research is necessary to find out how to optimally compensate for imperfect transparency. Secondly, in this specific application there is a problem in which the experience is only accessible using earbuds, and in other shared scenarios where some listeners may have earbuds and others not, those without earbuds receive an incomplete experience. To cater for all listeners, applications should perhaps focus on augmenting or extending loudspeaker content with earbuds in transparency mode; for example, providing access to additional content, or accessibility services.

5.8 Summary

A novel proof-of-concept prototype experience, adapted from an idea from the workshops in Chapter 4, and bespoke demonstrator system were developed to explore the concept of combined personal and shared audio. A research-through-design methodology was used to document and discuss the audio production and sound design approach used in the experience, and the technical implementation. Additionally, the design dimensions were reflected on and were found to be useful for presenting a number of defined options which can each be tested to find the optimal choice. The prototyping process also prompted the proposal of a new design dimension, *perception*, which captures the use of personal and shared audio devices and how audio is perceived by the listeners with these in use. The following chapter continues to investigate combined personal and shared audio experiences, this time using a controlled experimental approach.

Chapter 6

Evaluating hearables for augmenting TV audio in shared viewing environments

6.1 Introduction

In the previous chapter, a novel prototype was developed to explore the concept of concurrent and complementary hearable and loudspeaker audio through the creative use of the hearables' active transparency mode. The benefit of this delivery method is that it provides a more accessible alternative method of presenting both private and shared audio to listeners, a desired trait of emerging audio augmented reality applications, compared with acoustically transparent devices such as audio glasses that are far less commercially widespread.

TV has served as a hub for social activity and shared experiences in the home for many decades and continues to be a present feature in homes today. However, its limited ability to cater to individual preference in these scenarios is a hindrance, especially given the now pervasive nature of personalised devices and media. Hearables with acoustic transparency provide a potential solution through enabling of personalisation within shared TV experiences. In previous work, audio glasses have been used to augment TV audio by allocating (or 'mirroring') parts of the

TV mix on the glasses, where many benefits were reported [140]. The study in this chapter builds on this to observe whether consumer wireless earbud-style hearables with 'hear-through' capabilities have the potential to provide useful personal audio augmentations for improving TV viewing in shared viewing environments. In particular, the work focuses on formally evaluating a simulation of the listening experience under ideal conditions and comparing it against single device TV programme consumption methods.

Hence, the experiment presented in this chapter investigates two research questions.

1. How does mirroring of audio content on an earbud-style hearable affect preference and dialogue clarity versus standard TV viewing?
2. Does the type of programme affect these attributes for this form of reproduction?

The work was completed in collaboration with BBC R&D. This chapter is organised into the following sections. An overview of relevant literature is presented in Section 6.2. The methods used for preparing the audio material and the listening test are described in Section 6.3. The results of the listening test are outlined in Section 6.4. Further discussion to the experiment is presented in Section 6.5. Limitations are considered in Section 6.6, and a chapter summary is provided in Section 6.7.

6.2 Related work

6.2.1 Hearables, hear-through and transparency

Acoustic transparency has become more prevalent with increased use of wearable audio technology outside the home. While outside, we gain lots of information about our surroundings through sound, whether that is to ensure safety (e.g. knowledge of oncoming traffic) or communication (e.g. coffee barista calling out names). Therefore, having some form of acoustic transparency while using hearables enables listeners to more easily take on this information. As described in Chapter 2, hearables are a category of device that not only provide audio playback but have additional mobile computing features that can be utilised for audio processing and other applications (bio-signal monitoring), and can take many different form factors including headphones, earbuds, audio glasses, and hearing aids. Depending on the design of the hearable, acoustic transparency can be achieved passively where the hearable does not occlude the ear canal (e.g. audio glasses) or through signal processing known as 'hear-through', which is used in ear canal occluding devices such as headphones and earbuds. Hear-through processing attempts to compensate for the sound modification imparted by the device, and to replicate the open ear response as closely as possible through the use of equalisation [188]. A technical evaluation of commercial and prototype hearable hear-through modes is presented by Denk et al. [187]. The authors identified three characteristics for achieving good transparency: synchronisation between left and right earpieces, suitable equalisation to the open-ear response and reduction of delay and comb filtering effects. A large variance in achieving the open ear response was observed across the tested devices. A further perceptual evaluation of these devices revealed that most were not able to achieve sound quality comparable to the open ear, and that sound quality is largely determined by the extent to which the open ear transfer function is reproduced [189]. The challenges that need to be overcome for achieving full transparency are minimising the delay between hear-through signal and open ear signal, and finding solutions for personalised hear-through equalisation [188].

6.2.2 Improving TV listening experiences

Given the technological advancements since TVs first became present in living rooms across the world, the audio experience associated with TV hasn't changed all that much in that time. Most broadcast and on-demand TV programmes are still delivered and consumed in channel-based mono or stereo, and are typically reproduced over small in-built stereo loudspeakers within the TV itself. Rather it has been newer devices and experiences that have benefited from innovation in audio. For example, spatial audio formats that are more immersive, such as binaural and Ambisonics, are observed mainly in XR experiences and increasingly in music, but not TV. Additionally, for some time broadcasters have received complaints from consumers about poor speech intelligibility in their TV programmes [190], indicating improvements are still to be made regarding accessibility. Finally, another problem that remains is that the TV, as a facilitator of shared media experiences, cannot itself accommodate the range of individual preferences in a shared viewing environment. This problem is especially apparent now given the ubiquity of personalised forms of media. It also disproportionately affects viewers with specific accessibility requirements who currently may feel guilt or reluctance for imposing their TV sound requirements onto others. In many cases, a compromise between listeners may be achievable. However, where there is no satisfactory resolution, the ability to have personalised audio without diminishing the shared experience would be desirable.

Considering these challenges, there are two general approaches to making improvements to audio in TV experiences. The first is to improve the audio itself through addressing aspects of production, distribution and delivery of content, all of which are the aims of OBA and are covered in Chapter 2. Audio can also be improved through additional signal processing. For example, speech enhancement techniques using neural networks have been found to be effective for reducing listening effort [191]. The second approach is the use of additional connected devices to either enhance or complement TV audio. One method is to use soundzones, being multiple distinct acoustic spaces (or zones) within the same physical space with minimal interference, created using loudspeaker arrays. These zones are generated using signal processing techniques such as acoustic contrast control and pressure control, together with beam-forming technology [192]. Soundzone solutions can be packaged into soundbars for domestic usage; however, there are still significant technical and user interaction challenges that need to be overcome before serious consumer adoption can be considered. Small portable loudspeakers placed close to individuals, amplifying the full mix, not just the components of interest, offer some of the same affordances, albeit still affecting the experience of other listeners within a shared setting [193].

Hearables provide an alternative method where, for instance, accessibility improvements for hearing-impaired listeners can be provided to only those who require them. Bluetooth streaming accessories, for example, that provide a method for sending TV audio directly can yield increased speech quality for hearing aid users [194]. Bone conduction headphones, when used to supplement TV audio through playing the TV mix (known as 'mirroring'), have been demonstrated to provide benefits of increased clarity and spatialisation, especially for elderly listeners [195]. McGill et. al. [140] explored the use of acoustically transparent audio glasses to enhance TV viewing through mirroring various channel combinations of the TV mix on the glasses, and providing additional speech augmentations such as audio description and director's commentary. Benefits of mirroring were observed across all measured attributes including spatial realism, clarity, attention to dialogue and enjoyment, with the largest benefit reported where the full TV mix was mirrored on the audio glasses. In addition, the masking impact on these speech augmentations from the presence of the TV audio signal was minimised. Synchronised smartphones and tablets have been used to increase the loudspeaker count of TV programme experiences to provide a more immersive spatial audio experience [136]. This type of ad-hoc loudspeaker system can perform as well as 5.0 surround sound systems for envelopment [144].

In the previous chapter, a prototype audiovisual experience was presented where personalised audio via hearables was augmented with loudspeakers playing diffuse sounds, perceived by the listener through the hearables' transparency mode. Feedback from demonstrations revealed the potential for using transparency modes to provide personal speech enhancement for loudspeaker delivered audio. Transparency modes enable listeners to perceive their external sound environment while enjoying personal sound content, and are especially useful in outdoor environments. However, their potential to be used as a method for improving audio experiences in domestic settings has yet to be explored. The following sections explore how they might be used to complement TV audio, and how this compares with standard TV audio consumption methods.

6.3 Method

To investigate the proposed research questions, a within-subjects multiple-stimulus listening test was designed and conducted. Additional data on attitudes towards hearables and social listening was collected through an exit survey at the end of the test. This section describes the experimental methods and the test procedure.

6.3.1 Listening conditions

Five different listening conditions were investigated; three involved combining TV and hearables output where the hearables mirror part or all of the full TV mix, which consisted of dialogue (D), and music and effects (ME). The conditions were 'full mix on both devices' (*FM_FM*), 'dialogue on the hearables and full mix on the TV' (*D_FM*) and 'music and effects on the hearables and full mix on the TV' (*ME_FM*). The other two conditions were single device output only. One being the full mix on the TV without the hearables inserted (*TV_ONLY*), and the other where the full mix was played only on the hearables (*H_ONLY*). An overview of the conditions can be seen in Table 6.1. The full mix was preserved on the TV to focus on shared viewing scenarios where not everyone might have hearables. To observe any effects across different types of TV content, three BBC programmes were selected from *Documentary*, *Drama* and *Sports* genres. For each of the three programme types, three 20 second sections were captured to form the test items. Separate dialogue (D) and background stems (ME) were provided by the BBC for each programme. For the documentary and drama programme, the background consisted of music and varying sound effects. The sports programme background contained crowd noise and pitch-side sound effects. The full mixes for each programme were created by summing together the dialogue and background stems. Further explanation of how the test items were prepared is in the next section.

Condition	Hearables audio	TV audio
FM_FM	D, ME	D*, ME*
D_FM	D	D*, ME*
ME_FM	ME	D*, ME*
TV_ONLY	-	D, ME
H_ONLY	D, ME	-

Table 6.1: A summary of each listening condition, highlighting the audio assets that are allocated to each device. *D* corresponds to dialogue and *ME* corresponds to music and effects. The (*) indicates that the signals are perceived through the transparency mode of the hearables.

6.3.2 Stimuli preparation

Wireless, battery-powered devices such as hearables are a challenge to incorporate into controlled listening test environments. Wireless connections, such as Bluetooth, are less reliable than wired connections, and charging requirements can limit how long they might be used for and so

FIGURE 6.1: The hearables fitted to the GRAS KEMAR head and torso for stimuli recording.

disrupt test processes. With earbud-style devices, there is the additional challenge of ensuring the same fit for all participants, where failing to do so can result in drastically different listening experiences. For these reasons, the listening experience of augmenting TV audio using hearables in transparency mode (Figure 6.2) was simulated. This was achieved by splitting each combined device condition into two components: the perceived TV signal and the direct hearables signal. A GRAS KEMAR head and torso was used with the hearables inserted and their transparency mode enabled to create the TV component. This TV component includes the TV audio as captured through the hearables' transparency mode and any passive earbud leakage. The hearables component contains the relevant stereo audio mix (*D, ME, FM*) sent directly to the hearables fitted to KEMAR via Bluetooth and recorded directly with the transparency mode disabled. These two components were then manually synchronised and superimposed to approximate listening to the TV via the transparency mode with additional direct audio from the hearables. Other perceptual evaluation studies of hearables have used similar preparation methods to reproduce hearables signals over headphones[189].

All necessary recordings were recorded three times where the earbuds were refit to KEMAR between each set. The set of recordings with the best fit across both ears were selected for use. This was determined through critical listening of the recordings by the author where the best fit was identified on the best spatial and spectral balance achieved across the left and right earbud. The recordings were made in a quiet recording studio with similar acoustics to an

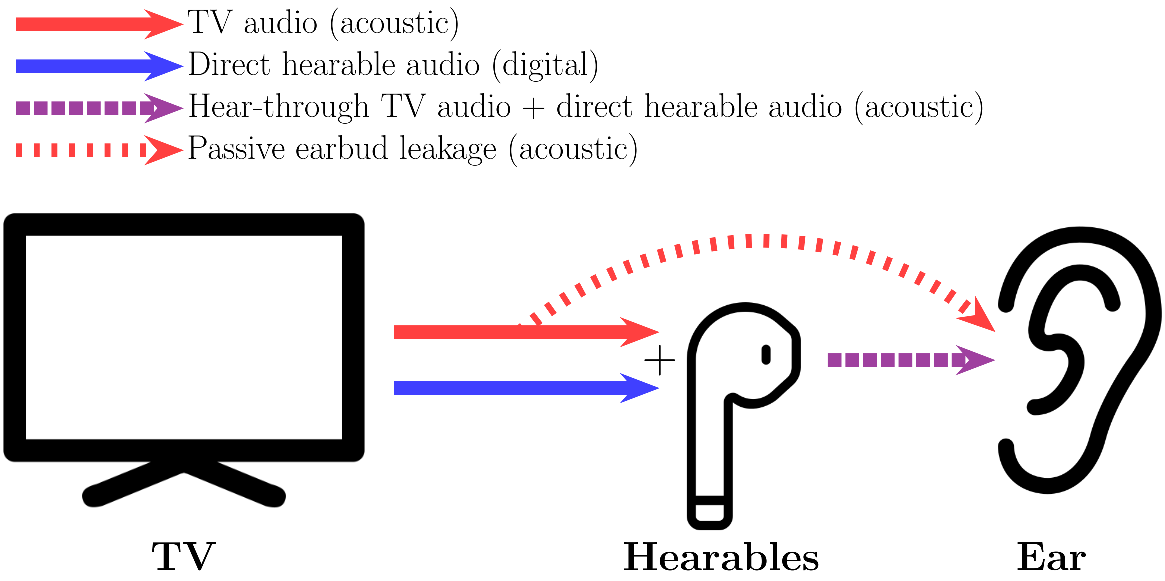


FIGURE 6.2: Concept diagram outlining the listening experience for augmenting TV audio with hearables, showing the relevant audio signal paths.

average living room. For the TV, a Mitsubishi 42" LDT422V monitor with SP-422V speakers was used and the hearables were a pair of Sony Linkbuds S. Measured passive attenuation of the hearables revealed the attenuation shelf began in the 400-800 Hz region and attenuation gradually increased up to attenuation of 30 dB at 20 kHz. The Linkbuds were selected as a typical example of a popular commercially available device available at the time of the study, and in comparison with two other devices from other manufacturers offered a flatter hear-through response. The dummy head was positioned facing the TV at a distance of 2.3m at an angle of 0°. The TV signal was measured at approximately 65 dB SPL at the open ear of the dummy head. All recordings were normalised to -23 Loudness Units Full Scale (LUFS). For creation of the combined device conditions, the TV component and the hearable component were superimposed and the resulting stimuli normalised to -23 LUFS. This method resulted in the combined device conditions containing both device signals in an equal ratio, and the overall loudness matched that of the single device conditions. In this experiment, the user was not given control over individual stimuli levels and so an equal ratio was selected to ensure that both device outputs were able to be perceived by the listener. This also ensured an equal overall loudness of conditions for these listening tests. Keeping all conditions at equal loudness was deemed more important than faithfully reproducing realistic mixing ratios between the two devices, as loudness is known to influence preference ratings.

6.3.3 Listening test

A total of 30 participants took part in a 45 minute-long multiple stimulus listening test, without defined reference and anchor stimuli. The participants, 20 male and 10 female, were between the ages of 22 and 56 and were recruited through university communication channels. All participants self-reported ‘normal hearing’, gave informed consent and were remunerated for their participation. Ethical approval for the experiment (*Geary20231006*) was granted by the University of York ethics committee. Participants were asked to rate each of the five listening conditions on a scale of 1-100 based on their overall preference where 1 represents lowest preference and 100 represents highest preference, for 10 programme items (3×3 programme types plus one random repeat item). Afterwards they listened to the same programme items but this time rating the conditions on *dialogue clarity*. *Dialogue clarity* was selected as this was highlighted as a potential benefit during the prototype demonstration in the previous chapter (Section 5.4). The order of the conditions in each item and the items themselves were randomised for each participant, and participants could freely switch between conditions while making the ratings. The items were presented at the same playback level over a pair of Beyerdynamic DT990 Pro headphones and were equalised for a flat frequency response using Sonarworks SoundID Reference software. The test interface was implemented using the webMUSHRA framework [29] and was displayed on a laptop, connected with an external mouse and keyboard. After the listening test, each participant completed a short exit survey to gather information on hearable ownership and usage, and attitudes around aspects of shared TV viewing experiences. The questions were either presented as statements where participants answered using a 5-point Likert scale or as questions with free text responses. The Likert scale approach was used as it is designed for measuring attitudes and sentiment, and the qualitative data is easy to analyse and interpret. Open questions were included to supplement these questions by capturing any additional information.

6.4 Results

Statistical analysis of these results was conducted using separate two-way repeated measures ANOVAs for *overall preference* and *dialogue clarity* attributes. Two factors were assessed: *listening condition* (*TV_ONLY, H_ONLY, FM_FM, D_FM and ME_FM*) and *programme type* (*documentary, drama, and sports*). Data normality was assessed using a Shapiro-Wilks test on the studentised residuals. Further post-hoc analysis was carried out using paired-samples t-tests with Bonferroni correction. Correlation between the attributes was determined through Pearson's correlation testing. For both attributes, the ANOVA result is reported and subsequent p-value reporting relate to pairwise comparisons from the t-tests.

6.4.1 Overall preference

Figure 6.3 presents the results for the *overall preference* attribute. *Overall preference* was normally distributed ($p>.05$) except for *documentary* × *H_ONLY* ($p=.017$) and *documentary* × *ME_FM* ($p=.024$). A statistically significant two-way interaction was identified between *listening condition* and *programme type* ($F=20.7$, $df=8$, $p<.001$, partial $\eta^2=.417$). Both *listening condition* ($F=96.3$, $df=4$, $p=<.001$, $\eta^2=.769$) and *programme type* ($F=18.4$, $df=2$, $p=<.001$, $\eta^2=.388$) factors were significant. The *H_ONLY* condition across all programme types was significantly preferred versus all other conditions ($p<.001$). The combined device conditions *FM_FM* and *D_FM* were comparable to the *TV_ONLY* listening experience. *ME_FM* was rated significantly worse than all other conditions ($p<.001$). Drama content was significantly preferred over documentary and sports ($p<.001$) when aggregating all conditions. There is a significant preference for *D_FM* ($p<.001$) when listening to the sports programme over the other programmes. Interestingly, there was no significant difference in preference between *D_FM* and *ME_FM* conditions for the drama programme ($p=.101$).

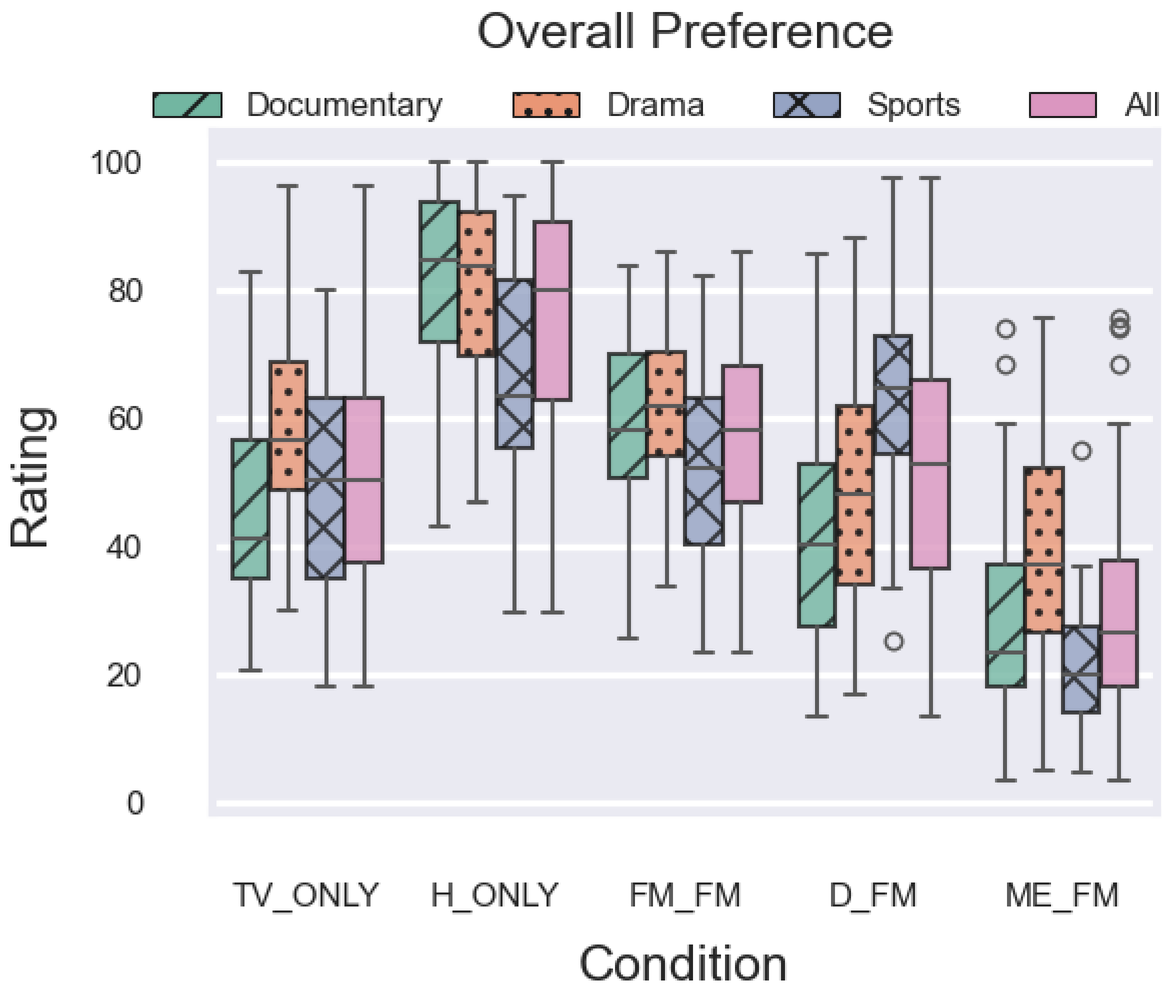


FIGURE 6.3: A boxplot of overall preference ratings for the five listening conditions. Each condition is further grouped by programme type and their combination, represented as the different colour plots. Each boxplot displays the median, the interquartile range (IQR), and outliers. The whiskers include data points distanced 1.5 times the IQR from the nearest quartile.

6.4.2 Dialogue clarity

Dialogue clarity was normally distributed (p>.05) except for *drama* × *FM_FM* (p=.017), *documentary* × *ME_FM* (p<.001), *drama* × *ME_FM* (p=.027) and *sports* × *ME_FM* (p=.008). A statistically significant two-way interaction was identified between *listening condition* and *programme type* (F=20.9, df=8, p<.001, partial η^2=.419). Both *listening condition* (F=121.5, df=4, p=<.001, η^2=.807) and *programme type* (F=82.7, df=2, p=<.001, η^2=740) were significant

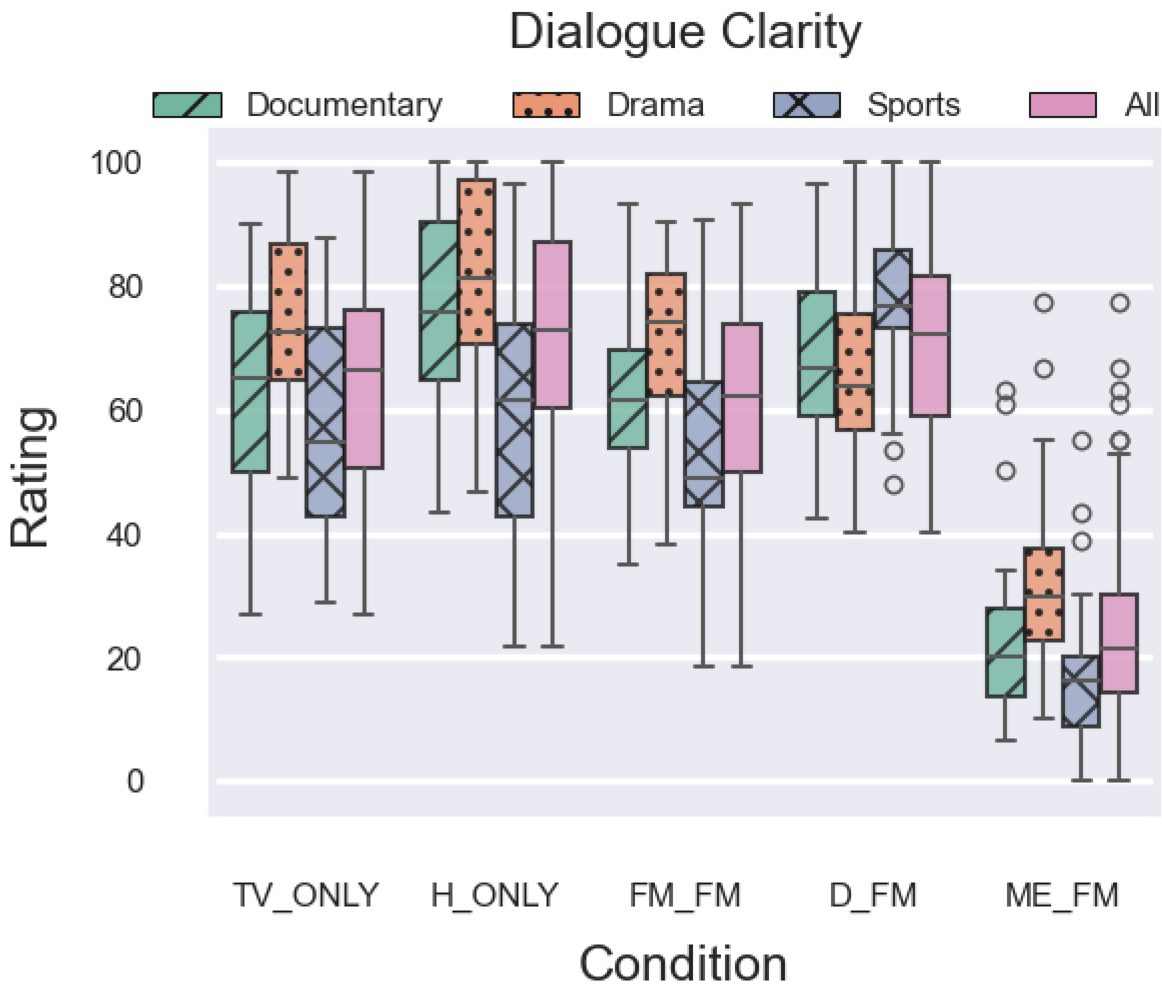


FIGURE 6.4: A boxplot of dialogue clarity ratings for each of the five listening conditions. Each condition is further grouped by programme type and their combination, represented as the different colour plots. Each boxplot displays the median, the interquartile range (IQR), and outliers. The whiskers include data points distanced 1.5 times the IQR from the nearest quartile.

factors. Displayed in Figure 6.4, *dialogue clarity* is broadly similar across all conditions except for the *ME_FM* condition where it is significantly worse than all others ($p<.001$). There is greater dialogue clarity for the *H_ONLY* condition over *TV_ONLY*, *FM_FM* and *D_FM* conditions for documentary and drama programmes, however, this difference is small (mean rating $\delta<15$). No significant difference was observed in *dialogue clarity* for the *D_FM* condition over the others, aside from within the sports programme where there was a significant improvement ($p<.001$). *Dialogue clarity* varies significantly across programme types ($p<.001$) with the drama programme dialogue having the most clarity and the sports programme dialogue having the least. *D_FM* significantly raises the dialogue clarity of the sports programme to above that of the other programmes with the same condition ($p<.001$). In contrast, for the drama programme, the *D_FM* condition reduces the clarity of the dialogue, when compared against the single device conditions such as *H_ONLY* ($p<.001$). Overall preference and dialogue clarity were positively correlated in all combinations of the independent variables ($r>0.5$, $p<.05$) aside from the drama × *D_FM* condition ($r=.21$, $p=.26$).

6.4.3 Survey results

Over half of participants said they owned hearables (16, 53%). For participants whose hearables had a 'transparency mode' (12), when asked how often they use this mode, the most common response was sometimes (4, 33%), with the other responses spread evenly across the scale. The common reasons for using the mode were for the need to pay attention to surroundings whilst outside the home or when communicating with others. Of those that do not own hearables, most either would like to own hearables in the future or would consider it (9, 64%).

Most participants positively valued (agree/strongly agree) socialising with others in shared viewing environments (19, 63%), and would like more control over their personal listening experience (21, 70%). There was a slightly more mixed response on whether participants would feel comfortable using hearables to enhance their listening experience during co-viewing (somewhat agree, 43%; neither agree or disagree, 23%; and somewhat disagree 27%). Participants were similarly split between those who would and would not use hearables to augment the TV. Only limited conclusions can be drawn from these final two responses given the level of abstraction from a real-life implementation of the experience. Free form text feedback from the listening test largely corroborates the observed correlation between the two measured attributes. For example, one participant said, "*during the second half of the test I still found myself rating the conditions that I felt had better sound quality higher than potentially ones that had 'better' clarity...*" which confirms that preference for the condition, whether positive or negative, has influenced subsequent dialogue clarity ratings to some extent.

6.5 Discussion

Previous work identified that mirroring music and other effects on a personal device may provide a more immersive audio experience through louder and improved spatialisation of those assets when compared to just listening on the TV [140]. However, in this experiment, any benefit of mirroring music and effects is negated by quieter dialogue at the 1:1 loudness ratio of the TV and hearables used for the combined device stimuli. Despite no direct comparison between programme types, the drama programme is seemingly preferred over the other programme types. This could be due to the full mix balance being the most satisfactory compared with the other programme types but also could be a simple preference for the drama genre within the participant pool. Concerning the significant preference for *D_FM* when listening to the sports programme over the other programmes, this could be attributed to the perceived importance of the music and effects content to the overall programme. The crowd noise in the sports programme, for example, may be less important or less pleasant to listen to in sports than music and effects in the drama programme. Related to this, the limited difference between the *D_FM* and *ME_FM* conditions for the drama programme suggest that both conditions shift the mix balance away from the ideal seen in *H_ONLY*, and that dialogue and the music and effects might be of nearer equivalence in importance than for the other programme types. Furthermore, the full mix balance of the drama programme is already dialogue focused (Figure **??**), therefore, additional music and effects does not impact too much on the dialogue.

Dialogue clarity varies significantly across programme types. This is most likely the result of the difference in full mix composition between the three programme types, where not only is the content of the *ME* track different across programmes, but also the relative loudness of the *D* and *ME* within the full mix, as illustrated in Figure **??**. Surprisingly, increasing the relative level of the dialogue (*D_FM*) in the documentary and drama programmes provided no additional benefit for dialogue clarity. Given that *overall preference* and *dialogue clarity* are positively correlated, preference for the different conditions is likely influencing the dialogue clarity ratings to some extent.

Separating the influence of the transparency mode specifically from that of changes to the dialogue/background mix in the combined device conditions is a challenge when interpreting the results of this experiment. However, information can be obtained by comparing the single device conditions (*TV_ONLY, H_ONLY*) to the *FM_FM* condition, where the mix balance is unchanged across all three conditions. Given that *FM_FM* was largely comparable to the TV only experience for both attributes, the spectral modification of the transparency mode could be tolerable for many listeners when there is some direct audio signal present. Nonetheless, this will depend on the quality of the transparency mode used, how it was implemented in practice, and the level balance between hear-through and direct audio. An experiment focusing on these details would be useful for confirmation.

6.6 Limitations

There are a number of important caveats and limitations to this work that require reflection. Firstly, the full mix of each programme was created by summing the dialogue (D) and background (ME) stems together. This process, while forming a plausible sounding mix, would have undoubtedly omitted additional mix processing such as side-chain compression of the background to the dialogue which would have presumably improved dialogue clarity and overall preference ratings. In addition, the sports programme consisted of assets used in an OBA production and so no channel-based mix balance appropriate for TV broadcast was available. It is likely that the commentary would have been set at a higher relative level to the background in a broadcast suitable mix, potentially resulting in higher attribute ratings than observed in this experiment. Secondly, as transparency mode characteristics across devices are variable [189], the results of this work may not be generalisable to other devices. Thirdly, the ear canal resonance produced by the KEMAR head and torso was not corrected in our stimuli and may have influenced attribute ratings. However, the effect is likely to be minimal as this was consistent across all stimuli.

The simulation of the combined device listening experience in this experiment has a number of limitations. The first is that the binaural TV component obtained using KEMAR may not be plausibly binaural for all participants, due to differences across personal HRTFs. Another is that an assumption has been made that defining the listening experience as two components and subsequently superimposing these components results in the desired experience. It is arguable that it is likely a good approximate as transparency modes necessitate minimal processing to avoid increasing latency. Regardless, it may not be exactly equivalent. In addition, this experiment assumed near-perfect synchronisation between the hearables and the TV, accounting for a synchronisation solution that would be necessary in a real-life implementation. This circumvents the problem of latency introduced by the transparency mode and Bluetooth transmission. Further work would need to be completed to identify the impact of latency on a listening experience such as this and to consider possible solutions.

Finally, as mentioned in Section 6.3, only a single level balance between the hearable and the TV was tested in this experiment. This is unlikely to be optimal and the preferred balance is likely influenced by factors such as what mix content is reproduced on the hearable, the programme type, and personal preference. Ultimately in the real-life experience, the user would have control over the level of the hearable. Further experimentation is required to identify trends in listener preference for level balance between TV and hearables.

6.7 Summary

The emergence and rise of personal devices and media has diminished the value of TV for many viewers. The TV is designed to facilitate a single shared experience; however, the lack of flexibility to accommodate individual preference in social settings has become a problem for these TV experiences. Consumer earbud-style hearables with transparency modes offer a potential solution for enabling personalised audio within shared TV viewing environments without more complex loudspeaker arrays. In this chapter, a listening test was conducted to identify how best to enhance TV audio using a hearables' transparency mode by transmitting various parts of the TV mix to the hearables, known as content 'mirroring'. This form of reproduction was compared with just listening on the TV and just listening on the hearables, and measured using *overall preference* and *dialogue clarity* attributes.

The research questions under investigation were as follows.

1. How does mirroring of audio content on an earbud-style hearable affect preference and dialogue clarity versus standard TV viewing?
2. Does the type of programme affect these attributes for this form of reproduction?

The results show that where the full mix or dialogue is mirrored on the hearable, it is preferred equally to listening to just the TV output, but less preferred than listening just over the hearables. Music and effects mirrored on the hearables was significantly less preferred than the other conditions, and is likely due to the reduction in dialogue quality and level in that condition. Thus, it is recommend that mirroring music and effects on the hearables should be avoided or done cautiously with TV programme content as to limit negative impact on dialogue comprehension. Benefits of mirroring the dialogue on the hearables were observed for dialogue clarity ratings in the sports programme where there was a low dialogue to background ratio in the full mix. However, no benefit was observed in the other programmes (documentary and drama) with good default dialogue separation.

The type of TV programme was found to significantly affect both measured attributes and is likely to be a consequence of differences in audio mix elements and their respective balance within each programme. However, most of the significant interaction was caused by the difference between the sports programme and the other two. Documentary and drama were rated similarly for both attributes which was expected as they are similar in audio mix constitution. *Overall preference* and *dialogue clarity* were found to be significantly positively correlated in all but one combination of the independent variables. Preference for the mix is likely to have affected listeners in their dialogue clarity ratings. Variance across the results, presumably due to personal preference reinforces the importance of personalization and choice when it comes to TV audio, and emphasises the requirement for personal sound control within shared environments. This sentiment is echoed in the participant responses to the exit survey.

Using hearables to mirror TV content using transparency modes would likely require a number of technical improvements, reducing latency and artefacts, in addition to further research into optimizing delivery systems before being considered for mainstream adoption. Of course, current transparency modes are not designed for this type of application; however, it is hoped that this work raises awareness around the benefits of concurrent personal and shared audio, and can contribute to the discussion around how concurrent use of personal and shared audio devices, and transparency features could improve or evolve in future iterations.

Chapter 7

Conclusion

The broad ambition of this work is to better understand co-located multi-device audio experiences in order to make more immersive, interactive, and accessible audio experiences for listeners in the future. This purpose stems from relevant research focusing heavily on technologies and technical solutions, such as production tools, that answer questions around how such experiences could be created, but lack clarity on the wider issues of why these experiences are important and what the best uses of the technology might be. Furthermore, the disjointed nature of multi-device audio research up to this point forms a barrier to learning and limits communication and collaboration, ultimately slowing progress.

Toward satisfying the stated ambition, in Chapter 1 the following research question was outlined:

> How can the conceptual understanding and design of co-located multi-device audio experiences be improved and how can new experience opportunities be identified and explored?

This overarching question was split into a number of smaller questions that are answered within the chapters of this thesis. Table 7.1 provides an overview of the questions and the chapters in which they are addressed. The following sections of this chapter contain a short summary of each prior chapter (Section 7.1), a list of the original contributions to knowledge (Section 7.2), proposals for future work (Section 7.3), and concluding remarks (Section 7.4).

Question	**Chapter(s)**
What is the current state of knowledge of co-located multi-device audio experiences?	2, 3
What are the design principles that underpin co-located multi-device audio experiences?	3
How can these design principles be used to explore the design space?	4
How can these design principles be used to create new experiences?	4
What can we learn of new co-located multi-device audio experiences through prototyping?	5
What can we learn of new co-located multi-device audio experiences through evaluation?	6

Table 7.1: A table containing each research sub-question and the chapters in which the question is addressed.

7.1 Chapter Summaries

7.1.1 Chapter 2

In Chapter 2, a review of the literature and relevant background was provided to prime the reader in the topics pertinent to the original research presented later in this thesis. This includes a recent history of audio reproduction systems and technologies from stereo, multi-channel, binaural, and Ambisonics, to web-based and object-based audio. From here, a description of listening tests and perceptual modelling provide an understanding of how audio experiences can be evaluated.

The extensive and diverse spectrum of audio devices that provide the opportunity for collaborative multi-device use are explored, providing context into the emergence and development of current multi-device audio experiences. This covers personal audio devices such as smartphones, headphones and headsets; shared audio devices such as television and smart speakers; and finally more bespoke experimental devices. Finally, a comprehensive overview of multi-device ecosystems is presented. This includes the most relevant fields of research, existing multi-device experience design practice, current multi-device audio tools and use-cases, and research and evaluation methodologies. Through this review the fragmented nature of audio related multi-device research became apparent and justified the need for a formal survey.

7.1.2 Chapter 3

Chapter 3 described an exploratory study where the the application space of co-located multi-device audio experiences was explored and documented through a review of the literature and a survey. The survey revealed information on the types of devices used and modes of device interaction observed in these experiences. Through thematic analysis of the obtained text data and further synthesis, a set of seven design dimensions that can be used to characterise and compare co-located multi-device audio experiences were proposed; these were *synchronisation*, *context*, *position*, *relationship*, *interactivity*, *organisation*, and *distribution*. A mapping of the current application space was presented where four categories were identified using the design dimensions, these were *public performances*, *interactive music*, *augmented broadcasting* and *social games*. This work contributes to the wider discussion about the role of multiple devices in audio experiences, and provides a source of reference for the understanding and design of future co-located multi-device audio experiences.

7.1.3 Chapter 4

The design dimensions were evaluated in the context of experience or product conceptualisation where they were used to generate ideas for new co-located multi-device audio experiences. This was achieved through a series of hybrid and in-person co-design workshops conducted with Bang & Olufsen employees. Presented as ideation cards, the design dimensions were found to be useful for understanding the design space of co-located multi-device audio experiences and forming new concepts. A variety of different experience ideas were obtained from the workshops, though most conformed to existing design patterns observed in Chapter 3.

7.1.4 Chapter 5

The case study outlined in this chapter describes the development of a novel multi-device audio experience prototype, selected from the workshops ideas in Chapter 4, exploring the concept of concurrent personal and shared audio delivered on different devices. A research-through-design approach was used to document the development process, where the sound design approach and technical implementation were outlined and discussed. Reflections on the design dimensions revealed that those that influence audio content can be useful to present and test different design choices. Furthermore, these reflections prompted the proposal of a new design dimension, *perception* which exposes the previously unaccounted for design concept of the 'auditory bubble' or shared and personal audio.

7.1.5 Chapter 6

An alternative application of concurrent personal and shared audio, using hearables to augment TV audio, is evaluated in the case study described in Chapter 6. A listening test was conducted to investigate how mirroring audio content on hearables was perceived versus standard TV listening reproduction methods, for different types of TV programme. Thirty participants took part in a multiple stimulus listening test and provided ratings on *overall preference* and *dialogue clarity* attributes for each reproduction condition. In general it was identified that mirroring dialogue on the hearables was similarly preferable to just listening on the TV, but does provide a noticeable benefit with audio content with a low default speech-to-background level ratio, as exhibited by the radio *sports* programme. However, mirroring music and effects on the hearables was mostly detrimental due to the reduction in the level of the dialogue within the audio mix. Both rating attributes were found to be positively correlated, where preference is likely to have influenced *dialogue clarity* ratings. Variance across all ratings, likely due to personal preferences, and survey responses reinforce the need for personal sound control in shared viewing environments.

7.1.6 Summary

Together, the outcomes from these chapters make a significant step toward demystifying co-located multi-device audio experiences as expressed through the original research question. Through the design dimensions, researchers are now better equipped to describe, communicate and design new multi-device audio experiences. The case studies in the latter chapters demonstrate how new experiences can be explored through both practice-based and formal evaluation methods, providing different yet complementary insights. This is necessary to ensure all experience aspects: technological, creative, and end-user receive equal treatment, ultimately yielding better experiences. Key contributions are outlined in the next section.

7.2 Contributions

This thesis contains a number of notable contributions to knowledge across different areas relevant to co-located multi-device audio experiences. These are highlighted below:

- A dataset of co-located multi-device audio experiences and delivery platforms that contains a name, a description of the experience, a description of the audio content and a reference for each entry. The dataset can be used as a point of reference by interested researchers and practitioners in future research related to this area, and can be accessed through [4].
- A set of design dimensions that capture the key design principles for co-located multi-device audio experiences which can be used to describe and characterise existing experiences and conceptualise and design new ones. Published in MDPI Applied Sciences through *Design Dimensions for Co-located Multi-device Audio Experiences* [5].
- A series of workshops dedicated to understanding and ideation of co-located multi-device audio experiences where the structure and exercises can be applied and adapted for future co-design activities involving experiences of this type. Published and presented at *Audio Mostly 2023* [6].

- A novel co-located multi-device audio-visual experience using a creative combination of hearables and loudspeakers, where external loudspeakers are used to augment a pair of hearables, facilitated through active transparency. A detailed description and explanation of the design and technical implementation can provide useful insights that can be applied in future experiences. Published and presented at *Audio Mostly 2023* [6].

- An original experimental method for approximating the listening experience of augmenting TV with hearables using a simulation technique, that can be utilised in future listening tests investigating concurrent direct and external sound, facilitated through active transparency features. Accepted for publication in the *IEEE Communications Magazine: Internet of Sounds* 2024 issue.

- A listening test exploring the application of hearables for audio content 'mirroring', allocating audio mix elements onto a separate device, for augmented TV programme viewing. Initial recommendations were provided for content provision with this reproduction method, primarily that mirroring of dialogue on the hearables was more preferable than mirroring of music and effects. Accepted for publication in the *IEEE Communications Magazine: Internet of Sounds* 2024 issue.

7.3 Future Work

The work in this thesis is interdisciplinary, spanning many different fields of research. Therefore, the scope of any future work is likely to be equally wide-ranging. For the purpose of this thesis, future work is characterised as work that either could be undertaken directly following from the work in this thesis, or as remaining issues in areas relevant to the broad development of co-located multi-device audio experiences.

7.3.1 Continuation of work in this thesis

The design dimensions provide a useful framework for understanding and describing co-located multi-device audio experiences and have been evaluated in the context of experience conceptualisation and design. To remain useful over longer periods of time, the design dimensions need to be further scrutinised; this could be in the form of experimentation with the design dimensions in experience design contexts with a greater number of researchers and creative practitioners. As acknowledged at the end of Chapter 3, the design dimensions represent a starting point where they should be reflected upon and adapted as technology advances and our expectations of audio experiences change.

This thesis has contributed significantly to the idea of concurrent personal and shared audio using multiple devices, and this remains a relatively unexplored area compared with other forms of multi-device audio experiences such as mobile loudspeaker arrays. The prototype outlined in Chapter 5 has not been formally evaluated, therefore a suitable suggestion might be to formally evaluate this or a similar experience to obtain more robust data on the proposed listening experience. Furthermore, an important aspect of personal and shared audio is the ability to communicate with others in the presence of personal sound. This is especially critical where personal ear-occluding devices are in use with active transparency. As listeners tend to use transparency mode for communication outside the home, a reasonable expectation is that this should not be a significant problem. Regardless, future work is needed to investigate different device combinations in concurrent personal and shared audio experiences and how they affect interpersonal communication.

In the listening test results described in Chapter 6 the effect of mirroring dialogue on the hearables on dialogue clarity was mixed, showing contrasting effects depending on the programme. A compounding factor of mix preference was identified to have affected these ratings. Further experiments are needed with a focus on speech intelligibility with hearing-impaired listeners to obtain more conclusive results on the benefits of hearables and dialogue mirroring for increasing speech intelligibility. More generally, only a small number of examples utilising this personal plus shared audio concept exist currently [140, 141]. More experiences need to be developed using this concept to fully understand the benefits and drawbacks of this medium and to identify additional use-cases. For example, subsequent developments of combined TV and hearables usage could see the utilisation of sensors such as digital compasses and accelerometers to provide new interaction possibilities for audiences. Audio mixes could be spatially optimised based on listener location or head orientation. In order for applications like this to be achieved, an inter-disciplinary research approach, as emphasized by the *Internet of Sounds* (IoS) [101] is recommended.

Moreover, social friction on using hearables in shared environments is expected due to hearables currently signalling unavailability to others, which is likely a consequence of long-established headphone usage that has this effect. Hopefully this will improve over time with further proliferation of hearables and normalization of *IoS* & VR/AR applications and experiences.

7.3.2 Wider technical challenges

The flexible and modular nature of device ecosystems in many co-located multi-device audio experiences presents a challenge for testing and monitoring, where it isn't practical to test every possible device configuration. Progress has been made towards solving this with tools such as *Audio Orchestrator*, however further improvements can still be made. A tool that could simulate different device configurations using virtual sound sources could be very beneficial, similar to Dolby's *Dolby Atmos Renderer* [196]. Coupling this with device emulation of common devices such as smartphones, TVs, smart speakers, and a virtual environment with head-tracking to replicate different listening positions could create a complex interactive sandbox environment for multi-device audio experience simulation.

A persistent user experience challenge for heterogeneous loudspeaker array applications is unbalanced device volumes. Currently this is addressed through manual adjustment but can be time-consuming. Finding a technical solution for automatic volume calibration of heterogeneous audio devices would help solve this user pain point and increase experience desirability. Another unmet challenge is how can robust panning be achieved across heterogeneous devices with variable hardware-influenced latency. The prototype and listening test introduced in Chapters 5 and 6 utilised hearables' transparency modes as an interface to enable audio augmentation. As a reminder, this is not the intended use of these modes, but it is hoped that the research presented can highlight this possibility and prompt discussion around improving hear-through processing, particularly reducing inter-device latency and audio artefacts, with greater consideration around audio augmented reality applications.

7.4 Final Remarks

Up to now, commercial success with adaptive co-located multi-device audio experiences has been pretty limited, with the best example being *Sonos* multi-room audio [160]. I believe this is down to a number of reasons. Closed device ecosystems, known as ‘walled gardens’, enacted by large international technology companies remain a problem when it comes to wide scale interoperability and synchronisation between audio devices. It is difficult to see this fact changing soon, however I hope that future standardisation efforts can address this issue, making development of multi-device audio experiences easier than at present. Moreover, the adaptive and flexible nature of many multi-device audio experiences are somewhat bounded by the limitations of channel-based audio and its deeply embedded tools and workflows. I believe that co-located multi-device audio experiences will benefit from increased uptake of object-based audio over time through fully harnessing the flexibility that object-based audio offers.

Speculating on the distant future, and given the steady increase in popularity of hearables, it is possible to imagine a scenario where all audio is consumed over personal devices, and loudspeakers become a relic of the past. Given this scenario, multi-device connectivity will be essential to facilitating shared audio experiences. Regardless of whether this plays out or not, it is hard to imagine a future without co-located multi-device audio experiences, and it is my view that increased interdisciplinary collaboration is necessary to realise the full potential of these experiences.

Appendix A

Multi-device Audio Survey & Results

A.1 Multi-device Audio Survey

Introduction

Welcome, and thank you for taking part in the following survey. We greatly appreciate the time and effort you're taking to complete this survey, and your contribution will significantly aid our research in exploring the landscape of multi-device audio experiences.

Aim

The aim of this survey is to use crowdsourcing to obtain a dataset of multimedia experiences and technologies that utilise multiple devices to reproduce audio, referred to in this study as 'multidevice audio experiences'. The dataset will act as a supplementary resource for further research into the area including supporting characterisation and analysis of multi-device audio experiences, as well as aiding the development of a set of design criteria for multi-device audio experiences. The resulting dataset may be openly published at an unspecified time after the conclusion of the survey.

Instructions

We would like you to tell us about a multi-device audio experience that you are aware of, whether you have experienced it yourself or not. We will ask you some questions about this in a short survey, which should take no longer than 5 minutes to complete. Your task is to answer each of the questions relating to your specific entry. You are encouraged to provide as much detail about your entry as possible using the free text boxes. If you are aware of more than one multi-device audio experience, please feel free to complete the survey for each entry.

Criteria

The term 'multi-device audio experiences' is a broad term which encompasses a wide range of technologies and use-cases. We're collecting experiences that use multiple devices for playing audio, where all playing devices can be perceived by the listener. We would like to obtain information on:

- products, platforms or systems that support and deliver multi-device audio experiences (Example: a SONOS sound system)
- single experiences, or pieces of content which include multiple devices playing audio (Example: A silent disco)

The inclusion criteria in more detail are provided below. Ideally, your entry must fit all three criteria.

1. The Platform/Experience must employ multiple devices with loudspeakers.
   Explanation: Essentially, any platform or experience that includes at least two devices that have loudspeakers (including any type of headphones). This aims to encapsulate any devices capable of playing audio, this includes but is not limited to smartphones, tablets, laptops, TV's, Bluetooth speakers and smart speakers.

2. The audio content must be distributed across the devices.
   Explanation: The experience must be capable of playing audio over multiple devices, whether that's simultaneously or intermittently. An example that does not fit this criterion is streaming audio content from one device to another using AirPlay or Chromecast.

3. The devices must be co-located in a single space, or group of adjoining spaces.
   Explanation: This final specification aims to make a distinction between a loudspeaker array where devices are within the same vicinity, and a group of individual devices remotely connected over the Internet. For example, the devices can be located either in a single room, or across multiple rooms in a building. This criterion excludes teleconferencing platforms such as Zoom or Skype and traditional phone calls.

The criteria will be displayed at the bottom of each page of questions for your reference. This test has been designed to follow the ethical codes and practices laid out by the University of York. This test is anonymised and all data, personal or otherwise, is confidential. This is a confidential and anonymous study, and you can withdraw at any time for any reason. This research is being undertaken by David Geary (drg519@york.ac.uk). If you have any queries, please let them know by email.

I have read and acknowledged all of the above and agree to participate:

- Yes
- No

Questions

1. Would you describe your entry as a platform, or a single experience?

- Platform - a system which supports and delivers experiences
- Experience - a single experience which involves multiple devices playing audio
- Don't know

2. What is the name of your entry?

3. Please provide a brief description of your entry.

4. Describe the audio content of your entry. Specifically, what does the audio content consist of? How is it distributed across the devices?

5. Which devices are a part of the platform/experience ecosystem?

- Mobile Phones/Smartphones
- Tablets
- Desktop/Laptop
- Smart speaker
- Radio
- TV
- Other

6. How do participants interact with the platform/experience? For example, how are the devices interacted with?

- Tactile Input (Keyboard/Mouse/Buttons)
- Touch screen
- Motion gesturing

- Voice commands
- Not interactive
- Other

7. For your platform/experience, is the number of involved devices variable? For example, can devices be added/removed from the platform/experience freely?

- Yes
- No
- Don't know

8. Do some devices in the platform/experience have a greater importance than others? For example, one device may be responsible for playing most of the audio content, while the other devices play a supporting role.

- Yes
- No
- Don't know

9. Can you provide any links to further information about your entry? (E.g. videos, webpages, publications etc.)

10. Is there anything else about your entry that you'd like to tell us?

A.2 Survey responses to multiple choice questions

Question	Option 1	Option 2	Option 3	Option 4	Option 5
Would you describe your entry as a platform, or a single experience? (43)	Platform (14, 33%)	Experience (28, 65%)	Don't know (1, 2%)	-	-
Which devices are a part of the platform/experience ecosystem? (78)	Mobile phones/smartphones (26, 34%)	Tablets (12, 15%)	Desktop/laptop (12, 15%)	Smart speaker (4, 5%)	Radio (3, 4%)
How do participants interact with the platform/experience? (68)	Tactile input (22 ,32%)	Touch screen (21, 31%)	Motion gesturing (13, 19%)	Voice commands (2, 3%)	Not interactive (7 10%)
For your platform/experience, is the number of involved devices variable? (43)	Yes (36, 84%)	No (5, 12%)	Don't know (2, 4%)	-	-
Do some devices in the platform/experience have a greater importance than others? (43)	Yes (11, 26%)	No (25, 58%)	Don't know (7, 16%)	-	-

Table A.1: Multiple choice questions and responses for the multi-device audio experiences and technologies survey. The number and percentage of responses are given in parentheses.

A.3 Full Experiences Table

Name	Description	Experience type	Reference
Fields	Collaborative ambient-style music performance using mobile devices.	Public performances	[131]
Ugnayan	Numerous audio tapes transmitted over different radio channels, replayed on personal radios across an urban space.	Public performances	[?]
Cassettes 100	100 performers walk around with cassette platyers playing different instrumental and environmental sound.	Public performances	[?]
So Predicatable?!	An improvisational dance piece using two Raspberry Pi devices with loudspeakers (DIADs).	Public performances	[?]
Babel	An art installation of a tower of analogue radios all playing simultaneously.	Public performances	[162]
Boombox Experiment	An experiment conducted by the Flaming Lips where 100 audience members, each with their own tape player, played music tapes in synchrony.	Public performances	[163]
Dialtones	Art installation involving mobile phones which played a selection of ringtones.	Public performances	[129]
Bloom	Hundreds of Raspberry Pi devices with loudspeakers (DIADs) in outside installation at Kew Gardens.	Public performances	[138]
Movies on the Meadows	Silent film screening on large screens in fields, with audience members wearing headphones.	Public performances	[139]

Name	Description	Experience type	Reference
Parking Lot Experiments	An experiment by the Flaming Lips where 30 individual tapes where played at the same time on 30 separate car tape deck systems.	Public performances	[197]
Manifolds	Audience in a gallery use their cellphones to project the sound installation's voices into the room. The audience becomes a moving loudspeaker orchestra.	Public performances	[?]
Collective Loops	Participatory music experience where each player controls a segment of a sequencer and can add musical elements to their segment.	Interactive music	[164]
Mesh Garden	Exploratory music-making game for smartphones, creating a piece of distributed ambient music between participants.	Interactive music	[168]
One-Man Orchestra	A human conductor uses motion gestures to control individual smartphones playing music.	Interactive music	[169]
Echobo	Participatory music experience where a central performer and audience members collaborate to create music.	Interactive music	[130]
Pick A Part	An experience built using Audio Orchestrator which enables individual devices to choose different instruments of a classical music piece.	Interactive music	[135]
ProXoMix	Participatory music experience created with Soundworks. Uses bluetooth beaconing so when devices move closer to each other, you can hear other devices music tracks.	Interactive music	[?]
Schminky	Interactive musical game using HP iPaqs and headphones.	Interactive music	[170]
Crowd in c[loud]	Participatory music game where audience members compose melodies on their own devices. These melodies can be shared via profiles and other members can listen and like them, like a social media experience.	Interactive music	[172]
Drops	Participatory music experience created using the Soundworks framework. Triggered sound sources echo across multiple devices.	Interactive music	[111]
GrainField	An individual performer is recorded, every second of recording is sent randomly to audience devices and replayed using a granular synthesizer to create an ambient complimentary backing track.	Interactive music	[173]
The Vostok-K Incident	An immersive 3D audio drama created using Audio Orchestrator.	Augmented broadcasting	[132]
Immersive Six Nations Rugby	Sports programme experience using multiple devices to add additional audio to a highlights show of a six nations rugby match.	Augmented broadcasting	[136]
Spectrum Sounds	Distributed musical experience where short pieces of classical music are spread across a minimum of 3 connected devices.	Augmented broadcasting	[167]

Name	Description	Experience type	Reference
Augmented TV Viewing	A study exploring the application of acoustically transparent Bose Frames to improve TV viewing by intermixing TV and headset audio.	Augmented broadcasting	[140]
BR Audiodeskription App	A mobile app that can deliver audio description to the users connected headphones, while the TV plays the standard broadcast mix.	Augmented broadcasting	[174]
Decameron Nights	Multi-device halloween themed audio story programme.	Augmented broadcasting	[171]
Monster	A halloween themed drama episode over a minimum of three devices.	Augmented broadcasting	[134]
Musical Bowls	Musical game of bowls using DIADs, bespoke Raspberry Pi devices with loudspeakers and sensors.	Social games	[94]
SoundWear	Devices consist of a wearable wrist device which can play audio, designed as a play toy for children. Sounds can be transferred to other devices by aligning them.	Social games	[95]
Please Confirm You Are Not A Robot	Multiplayer AAR game experience using Bose Frames to deliver personalised audio.	Social games	[141]

Table A.2: All the individual experiences in the dataset, with descriptions, the type of experience and the associated reference.

Bibliography

[1] W. H. Organization *et al.*, "Environmental noise guidelines for the european region," World Health Organization. Regional Office for Europe, Tech. Rep., 2018.

[2] C. Cieciura, R. Mason, P. Coleman, and M. Paradis, "Survey of media device ownership, media service usage, and group media consumption in UK households," in *Audio Engineering Society Convention 145*. aes.org, 2018.

[3] T. Neate, M. Jones, and M. Evans, "Cross-device media: a review of second screening and multi-device television," *Pers. Ubiquit. Comput.*, vol. 21, no. 2, pp. 391–405, Apr. 2017.

[4] D. Geary, "Co-located multi-device audio experiences dataset," Jul. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6839250

[5] D. Geary, J. Francombe, K. Hentschel, and D. Murphy, "Design dimensions of Co-Located Multi-Device audio experiences," *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, vol. 12, no. 15, p. 7512, Jul. 2022.

[6] ——, "Using design dimensions to develop a multi-device audio experience through workshops and prototyping," in *Proceedings of the 18th International Audio Mostly Conference*, ser. AM '23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 71–78.

[7] J. C. Steinberg and W. B. Snow, "Auditory perspective - physical factors," *Electrical Engineering*, vol. 53, no. 1, pp. 12–17, 1934.

[8] A. D. Blumlein, "British patent specification 394,325 (improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems)," *Journal of the Audio Engineering Society*, vol. 6, no. 2, pp. 91–130, 1958.

[9] B. ITU-R, "Rec. itu-r bs. 775-2: 2007 multichannel stereophonic sound system with and without accompanying picture," *Int. Telecommun. Union, Geneva, Switzerland*, 2007.

[10] InformIT, "The history of surround sound," https://www.informit.com/articles/article.aspx?p=337317, September 2004, (Accessed on 06/16/2024).

[11] S. Kim, Y. W. Lee, and V. Pulkki, "New 10.2-channel vertical surround system (10.2-VSS); comparison study of perceived audio quality in various multichannel sound systems with height loudspeakers," in *Audio Engineering Society Convention 129*. aes.org, 2010.

[12] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 multichannel sound system and its application," in *Audio Engineering Society Convention 118*. aes.org, 2005.

[13] Dolby, "Dolby atmos - dolby," https://www.dolby.com/technologies/dolby-atmos/, 2020, (Accessed on 01/19/2021).

[14] S. Paul, “Binaural recording technology: A historical review and possible future developments,” *Acta Acustica united with Acustica*, vol. 95, no. 5, pp. 767–788, 2009.

[15] W. G. Gardner and K. D. Martin, “HRTF measurements of a KEMAR,” *J. Acoust. Soc. Am.*, vol. 97, no. 6, pp. 3907–3908, Jun. 1995.

[16] M. Vorländer, “Past, present and future of dummy heads,” *Proc. Acústica, Guimarães, Portugal*, pp. 13–17, 2004.

[17] O. Rummukainen, S. Schlecht, A. Plinge, and E. A. P. Habets, “Evaluating binaural reproduction systems from behavioral patterns in a virtual reality—a case study with impaired binaural cues and tracking latency,” in *Audio Engineering Society Convention 143*. aes.org, 2017.

[18] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

[19] S. Moreau, J. Daniel, and S. Bertet, “3d sound field recording with higher order ambisonics - objective measurements and validation of a 4th order spherical microphone,” in *120th Convention of the AES*, 2006, pp. 20–23.

[20] M. Frank, F. Zotter, and A. Sontacchi, “Producing 3D audio in ambisonics,” in *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet*. aes.org, 2015.

[21] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich, “A 3d real time rendering engine for binaural sound reproduction,” vol. 9, pp. 107–110, Jul. 2003.

[22] REAPER, “Reaper — audio production without limits,” https://www.reaper.fm/, 2021, (Accessed on 01/26/2021).

[23] M. Kronlachner, “Plug-in suite for mastering the production and playback in surround sound and ambisonics,” *Gold-Awarded Contribution to AES Student Design Competition*, 2014.

[24] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, “MPEG-H audio—the new standard for universal spatial/3d audio coding,” *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 821–830, 2015.

[25] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, “Sonic interactions in virtual reality: State of the art, current challenges, and future directions,” *IEEE Comput. Graph. Appl.*, vol. 38, no. 2, pp. 31–43, Mar. 2018.

[26] C. Rogers, “Web audio api,” *W3C Working Draft*, 2012.

[27] J. Varpe and E. Aandahl, “Wac — the web audio conference,” https://webaudioconf.com/, 2021, (Accessed on 03/29/2021).

[28] N. Jillings, B. D. Man, D. Moffat, J. D. Reiss, and Others, “Web audio evaluation tool: A browser-based listening test environment,” 2015.

[29] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, “webMUSHRA—A comprehensive framework for web-based listening tests,” *Journal of Open Research Software*, vol. 6, no. 1, 2018.

[30] M. Buffa and J. Lebrun, “Web audio guitar tube amplifier vs native simulations,” in *Web Audio Conference 2017–Collaborative Audio# WAC2017*. hal.univ-cotedazur.fr, 2017.

[31] M. Buffa, J. Lebrun, J. Kleimola, O. Larkin, and S. Letz, “Towards an open web audio plugin standard,” in *Companion Proceedings of the The Web Conference 2018*, ser. WWW ’18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, Apr. 2018, pp. 759–766.

[32] C. Pike, P. Taylour, and F. Melchior, "Delivering Object-Based 3D audio using the web audio API and the audio definition model," in *Proceedings of the 1st Web Audio Conference.* webaudioconf.com, 2015.

[33] M. Comunità, A. Gerino, V. Lim, and L. Picinali, "Web-based binaural audio and sonic narratives for cultural heritage," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio.* aes.org, 2019.

[34] T. Schaller and I. G. Burleigh, "Using web audio API in web-based tools for ear training of sound engineers," in *Audio Engineering Society Conference: UK 26th Conference: Audio Education.* aes.org, 2015.

[35] M. Wijnants, G. Rovelo, P. Quax, and W. Lamotte, "WanderCouch," *Multimed. Tools Appl.*, vol. 76, no. 4, pp. 5721–5753, Feb. 2017.

[36] L. Ward and B. Shirley, "Personalization in object-based audio for accessibility: A review of advancements for hearing impaired listeners," *J. Audio Eng. Soc.*, vol. 67, pp. 584–597, Aug. 2019.

[37] M. Torcoli, J. Herre, H. Fuchs, J. Paulus, and C. Uhle, "The Adjustment/Satisfaction test (A/ST) for the evaluation of personalization in broadcast services and its application to dialogue enhancement," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 524–538, Jun. 2018.

[38] J. Paulus, M. Torcoli, C. Uhle, J. Herre, S. Disch, and H. Fuchs, "Source separation for enabling dialogue enhancement in object-based broadcast with MPEG-H," *J. Audio Eng. Soc.*, vol. 67, no. 7/8, pp. 510–521, 2019.

[39] H. Stenzel and U. Scuda, "Producing interactive immersive sound for MPEG-H: A field test for sports broadcasting," *Audio Engineering Society Convention 137*, 2014.

[40] M. Paradis, R. Gregory-Clarke, and F. Melchior, "VenueExplorer, object-based interactive audio for live events," *WAC. Paris*, 2015.

[41] M. Armstrong, A. Churnside, M. E. F. Melchior, M. Shotton, and M. Brooks, "Object-based broadcasting - curation, responsiveness and user experience," in *International Broadcasting Convention (IBC) 2014 Conference.* Institution of Engineering and Technology, 2014.

[42] M. Armstrong, S. Bowman, M. Brooks, A. Brown, J. Carter, A. Jones, M. Leonard, and T. Preece, "Taking Object-Based media from the research environment into mainstream production," *SMPTE Motion Imaging J.*, vol. 129, no. 5, pp. 30–38, Jun. 2020.

[43] C. Baume, "The mermaid's tears - bbc r&d," https://www.bbc.co.uk/rd/blog/2017-09-mermaids-tears-object-based-audio, 2019, (Accessed on 04/07/2021).

[44] J. Cox, "Making cake: from object-based production to experience - bbc r&d," https://www.bbc.co.uk/rd/blog/2016-12-making-cake-object-based-production, December 2016, (Accessed on 02/09/2021).

[45] C. Pike and W. Z, "Virtual reality sound in the turning forest - bbc r&d," https://www.bbc.co.uk/rd/blog/2016-05-virtual-reality-sound-in-the-turning-forest, May 2016, (Accessed on 02/09/2021).

[46] S. Meltzer and A. Murtaza, "First experiences with the MPEG-H TV audio system in broadcast," *SET Int. J. Broadcast Eng.*, vol. 4, pp. 47–52, 2018.

[47] C. Simon, M. Torcoli, and J. Paulus, "MPEG-H audio for improving accessibility in broadcasting and streaming," *arXiv [eess.AS]*, Sep. 2019.

[48] K. Kjörling, J. Rödén, M. Wolters, J. Riedmiller, A. Biswas, P. Ekstrand, A. Gröschel, P. Hedelin, T. Hirvonen, H. Hörich, and Others, "AC-4–the next generation audio codec," in *Audio Engineering Society Convention 140*. aes.org, 2016.

[49] C. Sexton, "Immersive audio: Optimizing creative impact without increasing production costs," in *Audio Engineering Society Convention 143*. aes.org, 2017.

[50] B. Roberts, "Dolby atmos: what is it? how can you get it? — what hi-fi?" https://www.whathifi.com/advice/dolby-atmos-what-it-how-can-you-get-it, November 2020, (Accessed on 01/20/2021).

[51] "S3a spatial audio – s3a: Future spatial audio for an immersive listener experience at home," https://www.s3a-spatialaudio.org/, 2020, (Accessed on 04/08/2021).

[52] A. Silzle, R. Sazdov, and M. Weitnauer, "The EU project ORPHEUS: Object-based broadcasting–for next generation audio experiences," *29th Tonmeistertagung*, 2016.

[53] A. Silzle, M. Weitnauer, O. Warusfel, W. Bleisteiner, T. Herberger, N. Epain, B. Duval, N. Bogaards, C. Baume, U. Herzog, and Others, "ORPHEUS audio project: piloting an end-to-end object-based audio broadcasting chain," in *IBC Conference, Amsterdam, The Netherlands, September*. orpheus-audio.eu, 2017, pp. 14–18.

[54] S. Füg, D. Marston, and S. Norcross, "The audio definition model—a flexible standardized representation for next generation audio content in broadcasting and beyond," in *Audio Engineering Society Convention 141*. aes.org, 2016.

[55] M. Noisternig, T. Carpentier, M. Geier, R. T. Herberger, and A. S. Iis, "D3. 2: Implementation and documentation of reverberation for object-based audio broadcasting," 2015.

[56] P. Coleman, A. Franck, J. Francombe, Q. Liu, T. de Campos, R. J. Hughes, D. Menzies, M. F. S. Gálvez, Y. Tang, J. Woodcock, P. J. B. Jackson, F. Melchior, C. Pike, F. M. Fazi, T. J. Cox, and A. Hilton, "An Audio-Visual system for Object-Based audio: From recording to listening," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 1919–1931, Aug. 2018.

[57] J. Woodcock, C. Pike, F. Melchior, P. Coleman, A. Franck, and A. Hilton, "Presenting the S3A object-based audio drama dataset," *Journal of The Audio Engineering Society*, May 2016.

[58] G. Costantini, A. Franck, C. Pike, J. Francombe, and others, "A dataset of High-Quality Object-Based productions," *Audio Engineering Society Convention 147*, 2019.

[59] A. Franck and F. M. Fazi, "Visr—a versatile open software framework for audio signal processing," in *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*. aes.org, 2018.

[60] H. Kim, L. Remaggi, P. Jackson, and A. Hilton, "Spatial audio reproduction system for VR using 360 degree cameras," in *The 12th Asia Pacific Workshop on Mixed and Augmented Reality (APMAR 2019) Proceedings*. epubs.surrey.ac.uk, Mar. 2019.

[61] J. Francombe, R. Mason, P. J. B. Jackson, T. Brookes, R. Hughes, J. Woodcock, A. Franck, F. Melchior, and C. Pike, "Media device orchestration for immersive spatial audio reproduction," in *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, ser. AM '17, no. Article 33. New York, NY, USA: Association for Computing Machinery, Aug. 2017, pp. 1–5.

[62] S. Bech and N. Zacharov, *Perceptual Audio Evaluation - Theory, Method and Application*. John Wiley & Sons, Jan. 2007.

[63] N. Zacharov, *Sensory Evaluation of Sound*. CRC Press, Dec. 2018.

[64] R. ITU-R, "Itu-r bs. 1534-3:"method for the subjective assessment of intermediate quality level of audio systems"," *International Telecommunications Union, Geneva*, 2015.

[65] M. Torcoli, T. Kastner, and J. Herre, "Objective measures of perceptual audio quality reviewed: An evaluation of their application domain dependence," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1530–1541, 2021.

[66] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, ser. MobiSys '10. New York, NY, USA: Association for Computing Machinery, Jun. 2010, pp. 179–194.

[67] M. Levin, *Designing multi-device experiences: An ecosystem approach to user experiences across devices.* " O'Reilly Media, Inc.", 2014.

[68] V. Lazzarini, S. Yi, and J. Timoney, "Digital audio effects on mobile platforms." eprints.nuim.ie, Sep. 2012.

[69] A. M. Cassorla, G. Kearney, A. Hunt, H. Riaz, M. Stiles, and D. T. Murphy, "Augmented reality for DAW-Based spatial audio creation using smartphones," in *Audio Engineering Society Convention 148*, 2020.

[70] M. Green and D. Murphy, "Environmental sound monitoring using machine learning on mobile devices," *Appl. Acoust.*, vol. 159, p. 107041, Feb. 2020.

[71] G. Essl and S. W. Lee, "Mobile devices as musical instruments - state of the art and future prospects," in *Music Technology with Swing.* Springer International Publishing, 2018, pp. 525–539.

[72] A. Hart, "The tablet as a classroom musical instrument," *Athens J. Educ.*, vol. 5, no. 3, pp. 299–324, Aug. 2018.

[73] M. Kock and C. Louven, "The power of sound design in a moving picture: an empirical study with emotouch for ipad," *Empir. Musicol. Rev.*, vol. 13, no. 3-4, pp. 132–148, 2018.

[74] N. Niknejad, W. B. Ismail, A. Mardani, H. Liao, and I. Ghani, "A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges," *Eng. Appl. Artif. Intell.*, vol. 90, p. 103529, Apr. 2020.

[75] S. Seneviratne, Y. Hu, T. Nguyen, G. Lan, S. Khalifa, K. Thilakarathna, M. Hassan, and A. Seneviratne, "A survey of wearable devices and challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2573–2620, 2017.

[76] J. Plazak and M. Kersten-Oertel, "A survey on the affordances of "hearables"," *Inventions*, 2018.

[77] P. Crum, "Hearables: Here come the: Technology tucked inside your ears will augment your daily life," *IEEE Spectrum*, vol. 56, no. 5, pp. 38–43, May 2019.

[78] Bose, "Bluetooth audio sunglasses — bose," https://www.bose.co.uk/en_gb/products/frames.html, 2021, (Accessed on 01/28/2021).

[79] M. McGill, S. Brewster, D. McGookin, and G. Wilson, "Acoustic transparency and the changing soundscape of auditory mixed reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–16.

[80] R. Gupta, J. He, R. Ranjan, W.-S. Gan, F. Klein, C. Schneiderwind, A. Neidhardt, K. Brandenburg, and V. Välimäki, "Augmented/Mixed reality audio for hearables: Sensing, control, and rendering," *IEEE Signal Process. Mag.*, vol. 39, no. 3, pp. 63–89, May 2022.

[81] I. d. V. Bosman, O. o. Buruk, K. Jørgensen, and J. Hamari, "The effect of audio on the experience in virtual reality: a scoping review," *Behav. Inf. Technol.*, vol. 43, no. 1, pp. 165–199, Jan. 2024.

[82] Ofcom, "Media nations 2023: Latest uk viewing and listening trends revealed - ofcom," https://www.ofcom.org.uk/news-centre/2023/media-nations-2023-latest-uk-viewing-and-listening-trends-revealed, August 2023, (Accessed on 03/12/2024).

[83] J. Abreu, J. Nogueira, V. Becker, and B. Cardoso, "Survey of catch-up TV and other time-shift services: a comprehensive analysis and taxonomy of linear and nonlinear television," *Telecommunication Systems*, vol. 64, no. 1, pp. 57–74, Jan. 2017.

[84] S. Basapur, H. Mandalia, S. Chaysinh, Y. Lee, N. Venkitaraman, and C. Metcalf, "FANFEEDS: evaluation of socially generated information feed on second screen as a TV show companion," in *Proceedings of the 10th European Conference on Interactive TV and Video*, ser. EuroITV '12. New York, NY, USA: Association for Computing Machinery, Jul. 2012, pp. 87–96.

[85] Amazon, "Echo & alexa devices - amazon.co.uk," https://www.amazon.co.uk/b?ie=UTF8&node=14100223031, 2021, (Accessed on 02/02/2021).

[86] Apple, "Homepod - apple (uk)," https://www.apple.com/uk/homepod/, 2021, (Accessed on 02/02/2021).

[87] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge, "Understanding the Long-Term use of smart speaker assistants," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–24, Sep. 2018.

[88] T. Martin, "The 50 most useful alexa skills," https://www.cnet.com/how-to/amazon-echo-most-useful-alexa-skills/, 2019, (Accessed on 01/07/2021).

[89] J. Lau, B. Zimmerman, and F. Schaub, "Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, pp. 1–31, Nov. 2018.

[90] A. Alhadlaq, J. Tang, M. Almaymoni, and A. Korolova, "Privacy in the amazon alexa skills ecosystem," *Star*, 2017.

[91] R. Garg and S. Sengupta, "He is just like me: A study of the Long-Term use of smart speakers by parents and children," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–24, Mar. 2020.

[92] L. Turchet and A. Zanetti, "Voice-based interface for accessible soundscape composition: composing soundscapes by vocally querying online sounds repositories," in *Proceedings of the 15th International Conference on Audio Mostly*, ser. AM '20. New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 160–167.

[93] O. Bown, L. Loke, S. J. Ferguson, and others, "Distributed interactive audio devices: Creative strategies and audience responses to novel musical interaction scenarios," *on Electronic Art*, 2015.

[94] O. Bown and S. Ferguson, "A musical game of bowls using the DIADs," in *NIME*. nime.org, 2016, pp. 371–372.

[95] J. Hong, H. Yi, J. Pyun, and W. Lee, "SoundWear: Effect of Non-Speech sound augmentation on the outdoor play experience of children," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, ser. DIS '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 2201–2213.

[96] L. Turchet and M. Barthet, "An ubiquitous smart guitar system for collaborative musical practice," *Journal of New Music Research*, vol. 48, no. 4, pp. 352–365, Aug. 2019.

[97] J. P. Bello, C. Silva, O. Nov, R. L. Dubois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy, "SONYC: a system for monitoring, analyzing, and mitigating urban noise pollution," *Commun. ACM*, vol. 62, no. 2, pp. 68–77, Jan. 2019.

[98] K. P. Yeo, S. Nanayakkara, and S. Ransiri, "StickEar: making everyday objects respond to sound," in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ser. UIST '13. New York, NY, USA: Association for Computing Machinery, Oct. 2013, pp. 221–226.

[99] J. M. Rigby, D. P. Brumby, S. J. J. Gould, and A. L. Cox, “Media multitasking at home: A video observation study of concurrent TV and mobile device usage,” in *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, ser. TVX ’17. New York, NY, USA: Association for Computing Machinery, Jun. 2017, pp. 3–10.

[100] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[101] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, “The internet of sounds: Convergent trends, insights, and future directions,” *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11 264–11 292, Jul. 2023.

[102] P. Cairns, H. Daffern, and G. Kearney, “Immersive network music performance: Design and practical deployment of a system for immersive vocal performance,” *Audio Engineering Society Convention*, 2020.

[103] L. Turchet and C. N. Ngo, “Blockchain-based internet of musical things,” *Blockchain: Research and Applications*, vol. 3, no. 3, p. 100083, Sep. 2022.

[104] “The international conference on new interfaces for musical expression,” https://www.nime.org/, June 2024, (Accessed on 06/15/2024).

[105] M. Lawton, S. Cunningham, and I. Convery, “Nature soundscapes: an audio augmented reality experience,” in *Proceedings of the 15th International Conference on Audio Mostly*. New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 85–92.

[106] M. Krzyzaniak, D. Frohlich, and P. 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*, ser. AM ’19. New York, NY, USA: ACM, Sep. 2019, pp. 160–167.

[107] F. Brudy, C. Holz, R. Rädle, C.-J. Wu, S. Houben, C. N. Klokmose, and N. Marquardt, “Cross-Device taxonomy: Survey, opportunities and challenges of interactions spanning across multiple devices,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–28.

[108] H. Jung, E. Stolterman, W. Ryan, T. Thompson, and M. Siegel, “Toward a framework for ecologies of artifacts: how are digital artifacts interconnected within a personal life?” in *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, ser. NordiCHI ’08. New York, NY, USA: Association for Computing Machinery, Oct. 2008, pp. 201–210.

[109] S. Bødker and C. N. Klokmose, “The Human–Artifact model: An activity theoretical approach to artifact ecologies,” *Human–Computer Interaction*, vol. 26, no. 4, pp. 315–371, Dec. 2011.

[110] R. Masu, M. Bettega, N. N. Correia, T. Romão, and F. Morreale, “ARCAA: a framework to analyse the artefact ecology in computer music performance,” in *Proceedings of the 9th International Conference on Digital and Interactive Arts*, ser. ARTECH 2019, no. Article 25. New York, NY, USA: Association for Computing Machinery, Feb. 2020, pp. 1–9.

[111] B. Matuszewski, N. Schnell, and F. Bevilacqua, “Interaction topologies in Mobile-Based situated networked music systems,” *Proc. Int. Wirel. Commun. Mob. Comput. Conf.*, vol. 2019, Mar. 2019.

[112] L. Dahl, “Wicked problems and design considerations in composing for laptop orchestra,” *NIME*, 2012.

[113] A. Lind, B. Yttergren, and H. Gustafson, “Challenges in the development of an easy-accessed mobile phone orchestra platform,” in *WAC 2021, Web Audio Conference, Online, July 5-7, 2021*, 2021.

[114] O. Bown and S. Ferguson, "Understanding media multiplicities," *Entertain. Comput.*, vol. 25, pp. 62–70, Mar. 2018.

[115] N. Weitzner, J. Freeman, Y.-L. Chen, and S. Garrett, "massmobile: towards a flexible framework for large-scale participatory collaborations in live performances," *Organised Sound*, vol. 18, no. 1, pp. 30–42, Apr. 2013.

[116] A. Franck, J. Francombe, J. Woodcock, R. Hughes, P. Coleman, D. Menzies, T. J. Cox, P. J. B. Jackson, and F. M. Fazi, "A system architecture for semantically informed rendering of object-based audio," *J. Audio Eng. Soc.*, vol. 67, no. 7/8, pp. 498–509, 2019.

[117] K. Hentschel and J. Francombe, "Framework for web delivery of immersive audio experiences using device orchestration," in *Adjunct Proceedings of ACM TVX*, Jun. 2019.

[118] J. Walker, D. L. Williams, I. C. Kegel, A. P. Gower, J. Jansen, M. Lomas, and S. Fjellsten, "2-IMMERSE: A platform for production, delivery, and orchestration of distributed media applications," *SMPTE Motion Imaging J.*, vol. 128, no. 7, pp. 45–51, Aug. 2019.

[119] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[120] S. Robaszkiewicz and N. Schnell, "Soundworks–a playground for artists and developers to create collaborative mobile web performances," in *WAC-1st Web Audio Conference.* wac.ircam.fr, 2015.

[121] B. Matuszewski, "A web-based framework for distributed music system research and creation," *J. Audio Eng. Soc.*, vol. 68, no. 10, pp. 717–726, Dec. 2020.

[122] D. Poirier-Quinot, B. Matuszewski, N. Schnell, and O. Warusfel, "Nü soundworks: Using spectators smartphones as a distributed network of speakers and sensors during live performances," in *Proceedings of 3rd Web Audio Conference.* qmro.qmul.ac.uk, Aug. 2017.

[123] A. Fraietta, O. Bown, S. Ferguson, S. Gillespie, and L. Bray, "Rapid composition for networked devices: HappyBrackets," *Comput. Music J.*, vol. 43, no. 2-3, pp. 89–108, Jun. 2019.

[124] E. de la Guía, M. Lozano, and V. R. Penichet, "Trainab: a solution based on tangible and distributed user interfaces to improve cognitive disabilities," in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 3039–3042.

[125] O. A. Alsos and D. Svanæs, "Interaction techniques for using handhelds and pcs together in a clinical setting," in *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, 2006, pp. 125–134.

[126] A. Chokshi, T. Seyed, F. Marinho Rodrigues, and F. Maurer, "eplan multi-surface: A multi-surface environment for emergency response planning exercises," in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, 2014, pp. 219–228.

[127] B. Taylor, "A history of the audience as a speaker array," in *NIME.* pdfs.semanticscholar.org, 2017, pp. 481–486.

[128] C. Brown and R. Santos, "Ugnayan," *CD notes. Tzadik*, 2010.

[129] G. Levin, "Dialtones-a telesymphony," https://www.fondation-langlois.org/html/e/page.php?NumPage=229, 2001, (Accessed on 12/02/2025).

[130] S. W. Lee and J. Freeman, "echobo: A mobile music instrument designed for audience to play," *Ann Arbor*, vol. 1001, pp. 48 109–42 121, 2013.

[131] T. Shaw, S. Piquemal, and J. Bowers, "Fields: An exploration into the use of mobile devices as a medium for sound diffusion," in *Proceedings of the international conference on New Interfaces for Musical Expression*, ser. NIME 2015. Baton Rouge, Louisiana, USA: The School of Music and the Center for Computation and Technology (CCT), Louisiana State University, May 2015, pp. 281–284.

[132] J. Francombe, J. Woodcock, R. J. Hughes, K. Hentschel, E. Whitmore, and A. Churnside, "Producing audio drama content for an array of orchestrated personal devices," BBC, Tech. Rep., 2019.

[133] E. Young, "Decameron nights: Immersive and interactive with all your audio devices - bbc r&d," https://www.bbc.co.uk/rd/blog/2020-08-audio-drama-surround-sound-spatial, 2020, (Accessed on 02/03/2021).

[134] E. Young and J. Francombe, "Monster: A hallowe'en horror played out through your audio devices - bbc r&d," https://www.bbc.co.uk/rd/blog/2020-10-audio-drama-monster-interactive-sound, 2020, (Accessed on 02/03/2021).

[135] J. Francombe, "Conduct your own orchestra? pick a part with the bbc philharmonic! - bbc r&d," https://www.bbc.co.uk/rd/blog/2020-09-proms-synchronised-audio-music-interactive, 2020, (Accessed on 02/03/2021).

[136] ——, "Get to the centre of the scrum - our immersive six nations rugby trial - bbc r&d," https://www.bbc.co.uk/rd/blog/2021-02-synchronised-audio-devices-sound-immersive, February 2021, (Accessed on 03/02/2021).

[137] S. Ferguson, A. Rowe, O. Bown, L. Birtles, and C. Bennewith, "Networked pixels: Strategies for building visual and auditory images with distributed independent devices," in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, ser. C&C '17. New York, NY, USA: Association for Computing Machinery, Jun. 2017, pp. 299–308.

[138] S. J. Ferguson, A. Rowe, O. Bown, L. Birtles, and C. Bennewith, "Sound design for a system of 1000 distributed independent audio-visual devices," *New Interfaces for Musical Expression 2017*, 2017.

[139] "Movies on the meadows 2019 — cambridge film festival," https://www.cambridgefilmfestival.org.uk/news/movies-meadows-2019, 2019, (Accessed on 04/12/2022).

[140] M. McGill, F. Mathis, M. Khamis, and J. Williamson, "Augmenting TV viewing using acoustically transparent auditory headsets," in *ACM International Conference on Interactive Media Experiences*, ser. IMX '20. New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 34–44.

[141] A. N. Nagele, V. Bauer, P. G. T. Healey, J. D. Reiss, H. Cooke, T. Cowlishaw, C. Baume, and C. Pike, "Interactive audio augmented reality in participatory performance," *Frontiers in Virtual Reality*, vol. 1, 2021.

[142] J. Woodcock, J. Francombe, R. Hughes, R. Mason, W. J. Davies, and T. J. Cox, "A quantitative evaluation of media device orchestration for immersive spatial audio reproduction," in *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*. aes.org, 2018.

[143] J. Francombe, J. Woodcock, R. J. Hughes, R. Mason, A. Franck, C. R. Pike, T. Brookes, W. J. Davies, P. J. B. Jackson, T. Cox, F. Fazi, and A. Hilton, "Qualitative evaluation of media device orchestration for immersive spatial audio reproduction," *J. Audio Eng. Soc.*, vol. 66, no. 6, pp. 414–429, Jun. 2018.

[144] A. Wilson, T. Cox, N. Zacharov, and C. Pike, "Perceptual audio evaluation of media device orchestration using the multi-stimulus ideal profile method," *Journal of The Audio Engineering Society*, Oct. 2018.

[145] N. Zacharov, C. Pike, F. Melchior, and T. Worch, "Next generation audio system assessment using the multiple stimulus ideal profile method," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. ieeexplore.ieee.org, Jun. 2016, pp. 1–6.

[146] K. Hentschel and J. Francombe, “Evaluation of an immersive audio experience using questionnaire and interaction data,” BBC, Tech. Rep., Oct. 2019.

[147] C. Cieciura, R. Mason, P. Coleman, and J. Francombe, “Understanding users’ choices and constraints when positioning loudspeakers in living rooms,” BBC, Tech. Rep., 2020.

[148] J. Francombe and K. Hentschel, “Exploring audio device orchestration in workshops with audio professionals,” BBC, Tech. Rep., 2020.

[149] J. Francombe, K. Hentschel, and S. Clarke, “Exploring audio device orchestration in workshops with young people,” BBC, Tech. Rep., 2020.

[150] L. Kerlin, “WHP 371 - human values: understanding psychological needs in a digital age,” BBC, Tech. Rep., 2020.

[151] H. Kim, S. Lee, J.-W. Choi, H. Bae, J. Lee, J. Song, and I. Shin, “Mobile maestro: enabling immersive multi-speaker audio applications on commodity mobile devices,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’14. New York, NY, USA: Association for Computing Machinery, Sep. 2014, pp. 277–288.

[152] D. Geary, J. Francombe, K. Hentsche, and D. Murphy, “A survey of co-located multi-device audio experiences,” in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*. Audio Engineering Society, Aug. 2022.

[153] Qualtrics, “Qualtrics xm // the leading experience management software,” https://www.qualtrics.com/uk/?rid=ip&prevsite=en&newsite=uk&geo=GB&geomatch=uk, 2022, (Accessed on 04/19/2022).

[154] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group, “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement,” *Int. J. Surg.*, vol. 8, no. 5, pp. 336–341, Feb. 2010.

[155] ACM, “Acm digital library,” https://dl.acm.org/, 2022, (Accessed on 04/19/2022).

[156] A. E. Society, “Aes e-library,” https://www.aes.org/e-lib/, 2022, (Accessed on 04/19/2022).

[157] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[158] H. Kim, J. S. Sefcik, and C. Bradway, “Characteristics of qualitative descriptive studies: A systematic review,” *Res. Nurs. Health*, vol. 40, no. 1, pp. 23–42, Feb. 2017.

[159] Amazon, “Welcome to prime video,” https://www.primevideo.com/, 2020, (Accessed on 01/29/2021).

[160] Sonos, “Sonos — wireless speakers and home sound systems,” https://www.sonos.com/en-gb/home, 2022, (Accessed on 05/05/2022).

[161] A. Striner, S. Azad, and C. Martens, “A spectrum of audience interactivity for entertainment domains,” in *Interactive Storytelling*. Springer International Publishing, 2019, pp. 214–232.

[162] T. Barson, “‘babel’, cildo meireles, 2001 — tate,” https://www.tate.org.uk/art/artworks/meireles-babel-t14041, 2011, (Accessed on 03/24/2022).

[163] G. Kaufman, “Flaming lips’ bizarre ’boombox experiment’ - mtv,” https://www.mtv.com/news/3439/flaming-lips-bizarre-boombox-experiment/, 1998, (Accessed on 03/24/2022).

[164] N. Schnell, B. Matuszewski, J.-P. Lambert, S. Robaszkiewicz, O. Mubarak, D. Cunin, S. Bianchini, X. Boissarie, and G. Cieslik, "Collective loops: Multimodal interactions through co-located mobile devices and synchronized audiovisual rendering based on web standards," in *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '17. New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 217–224.

[165] V. Vinayagamoorthy, R. Ramdhany, and M. Hammond, "Enabling Frame-Accurate synchronised companion screen experiences," in *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, ser. TVX '16. New York, NY, USA: Association for Computing Machinery, Jun. 2016, pp. 83–92.

[166] J.-P. Lambert, S. Robaszkiewicz, and N. Schnell, "Synchronisation for distributed audio rendering over heterogeneous devices, in html5," in *Proceedings of the 2nd Web Audio Conference*. Georgia Institute of Technology, 2016.

[167] BBC, "Spectrum sounds: a lockdown audio composition in seven colours - bbc r&d," https://www.bbc.co.uk/rd/blog/2021-09-hearing-synaesthesia-audio-composition-music, September 2021, (Accessed on 04/12/2022).

[168] T. Carson, "Mesh garden: A creative-based musical game for participatory musical performance," *NIME*, 2019.

[169] C. K. Tsui, C. H. Law, and H. Fu, "One-man orchestra: conducting smartphone orchestra," in *SIGGRAPH Asia 2014 Emerging Technologies*, ser. SA '14, no. Article 11. New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 1–2.

[170] J. Reid, R. Hull, T. Melamed, and D. Speakman, "Schminky: The design of a café based digital experience," *Pers. Ubiquit. Comput.*, vol. 7, no. 3, pp. 197–202, Jul. 2003.

[171] E. Young, "Decameron nights: Immersive and interactive with all your audio devices - bbc r&d," https://www.bbc.co.uk/rd/blog/2020-08-audio-drama-surround-sound-spatial, August 2020, (Accessed on 03/26/2022).

[172] S. W. Lee, A. Deusany, and G. Essl, "Crowd in c[loud] : Audience participation music with online dating metaphor using cloud service," https://smartech.gatech.edu/handle/1853/54594, 2016, (Accessed: 2024/9/23).

[173] B. Matuszewski and N. Schnell, "GrainField," in *Audio Mostly*. hal.science, 2017.

[174] B. Rundfunk, "Br audiodeskription (in german)," https://apps.apple.com/gb/app/br-audiodeskription/id1568128729, (Accessed on 03/25/2022).

[175] D. A. Norman and S. W. Draper, *User Centered System Design: New Perspectives on Human-computer Interaction*. USA: Lawrence Erlbaum Associates, 1986.

[176] M. Steen, "Co-design as a process of joint inquiry and imagination," *Des. Issues*, vol. 29, no. 2, pp. 16–28, Apr. 2013.

[177] M. Steen, M. Manschot, and N. D. Koning, "Benefits of co-design in service design projects," *International Journal of Design*, Aug. 2011.

[178] N. Goldman and K. Narayanaswamy, "Software evolution through iterative prototyping," in *Proceedings of the 14th international conference on Software engineering*, ser. ICSE '92. New York, NY, USA: Association for Computing Machinery, Jun. 1992, pp. 158–172.

[179] S. Mora, F. Gianni, and M. Divitini, "Tiles: A card-based ideation toolkit for the internet of things," in *Proceedings of the 2017 Conference on Designing Interactive Systems*, ser. DIS '17. New York, NY, USA: Association for Computing Machinery, Jun. 2017, pp. 587–598.

[180] K. Vaajakallio and T. Mattelmäki, "Design games in codesign: as a tool, a mindset and a structure," *CoDesign*, vol. 10, no. 1, pp. 63–77, Jan. 2014.

[181] D. Gray, S. Brown, and J. Macanufo, *Gamestorming: A Playbook for Innovators, Rulebreakers, and Changemakers*. "O'Reilly Media, Inc.", Jul. 2010.

[182] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," *Qual. Health Res.*, vol. 15, no. 9, pp. 1277–1288, Nov. 2005.

[183] D. Godin and M. Zahedi, "Aspects of research through design: A literature review," in *DRS Biennial Conference Series*. dl.designresearchsociety.org, 2014.

[184] B. Studio, "Cosmos laundromat: First cycle - blender studio," https://studio.blender.org/films/cosmos-laundromat/, 2023, (Accessed on 04/12/2023).

[185] S. Høier, "Surrounded by ear candy?: The use of surround sound in oscar-nominated movies 2000–2012," *Nordicom Review*, vol. 35, no. s1, pp. 251–262, Aug. 2014.

[186] Cycling74, "What is max? — cycling '74," https://cycling74.com/products/max, August 2024, (Accessed on 08/17/2024).

[187] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, "Acoustic transparency in hearables—technical evaluation," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 508–521, 2020.

[188] R. Gupta, R. Ranjan, J. He, W.-S. Gan, and S. Peksi, "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*. Audio Engineering Society, Aug. 2020.

[189] H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, "Acoustic transparency in hearables—perceptual sound quality evaluations," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 495–507, 2020.

[190] D. Geary, M. Torcoli, J. Paulus, C. Simon, D. Straninger, A. Travaglini, and B. Shirley, "Loudness differences for voice-over-voice audio in TV and streaming," *J. Audio Eng. Soc.*, vol. 68, no. 11, pp. 810–818, Dec. 2020.

[191] N. L. Westhausen, R. Huber, H. Baumgartner, R. Sinha, J. Rennies, and B. T. Meyer, "Reduction of subjective listening effort for TV broadcast signals with recurrent neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3541–3550, 2021.

[192] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.

[193] G. Kbar, A. Bhatia, M. H. Abidi, and I. Alsharawy, "Assistive technologies for hearing, and speaking impaired people: a survey," *Disabil. Rehabil. Assist. Technol.*, vol. 12, no. 1, pp. 3–20, Jan. 2017.

[194] P. Smith and A. Davis, "The benefits of using bluetooth accessories with hearing aids," *Int. J. Audiol.*, vol. 53, no. 10, pp. 770–773, Oct. 2014.

[195] H. Isono, "Television viewing for elderly people using bone-conducting headphones," *Proceedings of the Annual Meeting of Japan Ergonomics Society*, vol. 47spl, pp. 354–357, 2011.

[196] Dolby, "Dolby atmos renderer - dolby professional - dolby professional," https://professional.dolby.com/product/dolby-atmos-content-creation/dolby-atmos-renderer/, July 2024, (Accessed on 07/21/2024).

[197] F. Lips, "1996-1997: The parking lot experiments," https://web.archive.org/web/20081211053043/http://www.flaminglips.com/content/live/shows/s02a.php?sid=8bff976a4c80b0cc13d77e984ade33df, 1996, (Accessed on 04/12/2022).