Design Considerations for a Wearable, Bi-Modal Interface

A thesis submitted in partial fulfilment of

the requirements for the Degree of Doctor

of Philosophy in Human Interface

Technology in the HIT Lab NZ,

University of Canterbury

by Amit P. Barde

University of Canterbury

2018

Supervisory Team	
Senior Supervisor	Rob Lindeman
Secondary Supervisor	Mark Billinghurst
Associate Supervisor	Gun Lee

Abstract

The design of wearable interfaces that mediate interactions between people and the real world requires us to examine these interactions from the perspective of the users. The quality of these interactions is determined by how designers choose to implement the exchange of information that occurs between the user and the interface. Design considerations that have informed the development of existing interaction techniques in wearable interfaces have been examined in this thesis, and alternate methods proposed based on these investigations. We hypothesized that a bi-modal (auditory and visual) form of information delivery supported by contextual awareness should deliver a more natural form of interaction by leveraging the strengths of the visual and auditory senses.

This thesis tests this hypothesis and presents empirical results from five studies that look at how audio-visual feedback mechanisms can be incorporated into a wearable interface for effective information presentation. Results from these experiments demonstrate that this form of information delivery is a viable alternative to current wearable interfaces. The use of a bone conduction headset in conjunction with a traditional wearable visual interface was shown to be effective at mediating interactions with the real world provided the user is presented with well-designed auditory and visual cues.

The application-based studies in this thesis demonstrate that it is possible to have an unobtrusive interface capable of presenting the user with information which is both informative and intuitive in nature. An audio-visual information delivery design that does not prioritise one faculty over the other was shown to be effective at decreasing time on task, hence improving user efficiency.

This thesis also summarises limitations of the current work and identifies areas for future work, such as contextually aware wearable interfaces that incorporate computer-vision and bio-feedback. An argument for future wearable interfaces to incorporate audition and vision based Augmented Reality features is also made. In addition to this, a case for testing these interfaces in a real-world environment to assess their usability and durability is also presented.

Acknowledgements

I would like to thank my academic supervisor Rob Lindeman for his patience and support. Taking over supervision duties at a critical juncture in the PhD cycle must not have been easy. Huge thanks are also due to Mark Billinghurst. Without him I might never have had the opportunity to pursue research in the field of my liking at HITLab NZ. His guidance from the time I submitted an application to the lab right up to this point has been invaluable. Thanks to William (Deak) Helton for taking me on as a student despite having a significant number of students to look after to begin with. Deak's advice on experimental design and an amazing sense of humour to boot have made it that much easier to deal with the challenges of doctoral studies. A big thanks also to my associate supervisor Gun Lee. Gun's expertise in statistics have been responsible for accurately analysed and reported results in most of my published work. Having him around to advise me on a subject that I was so apprehensive about when I started out has been of great help. Thank you also to Greg O'Beirne of the Communication Disorders department for allowing me to use their sound booths for some of my experiments. I would also like to thank Matt Ward, my collaborator on two experiments. Without his mathematical expertise and programming genius, it would have proved hard to run through the number of experiments I have over the course of my PhD. Thanks also to Eduardo Sandoval, Robin Watson and all the other students and staff at the HITLab NZ. You have made my stay at the lab both fruitful and enjoyable. Thank you also to Simon Hoermann who provided some valuable advice towards the end of PhD, and Robert Sazdov for inculcating in me the habit of reading and summarising academic papers. Finally, thanks also to Sidney Wong for volunteering to read through the thesis and suggesting changes that have made this thesis easier to read.

I am eternally grateful to my maternal grandmother and my parents. Witnessing their work ethic and tireless efforts to provide for and take care of a family has been an inspiration. To my grandmother, without whose unshakable belief in my abilities I would never have been here. And finally, I owe a huge debt of gratitude to my wife Shailee for her unconditional support and patience.

Deputy Vice-Chancellor’s Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 3 *of the thesis is a reproduction of the work undertaken and published in collaboration with William S. Helton, Mark Billinghurst and Gun Lee. Results from this work have been presented at the following conference.*

140th Convention of the Audio Engineering Society

Barde, A., Helton, W. S., Lee, G., & Billinghurst, M. (2016, May). Binaural Spatialization over a Bone Conduction Headset: Minimum Discernible Angular Difference. In *Audio Engineering Society Convention 140*. Audio Engineering Society.

Please detail the nature and extent (%) of contribution by the candidate:

The experiment was planned and executed in consultation with Mark Billinghurst. He provided significant feedback on experiment design and data collection methodology. William (Deak) Helton provided supervisory support and verified the experiment design prior to the commencement of the study. Gun Lee provided a significant amount of support, helping me analyse the data that was obtained. All the data analysis was carried out under his supervision. All the authors helped contribute the editing of the paper with William (Deak) Helton and Gun Lee’s contribution not exceeding 2% of the final paper. Mark Billinghurst’s contribution towards the editing was significantly more but did not exceed 7% of the final paper.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all.

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Mark Billinghurst* Signature: M Billinghurst Date: *14/1/2018*

Deputy Vice-Chancellor’s Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 4 *of the thesis is a reproduction of the work undertaken and published in collaboration with William S. Helton, Mark Billinghurst and Gun Lee. Results from this work have been presented at the following conference.*

22nd International Conference on Auditory Display (ICAD-2016)

Barde, A., Lee, G., Ward, M., Helton, W. S., & Billinghurst, M. (2016). A Bone Conduction Based Spatial Auditory Display as Part of a Wearable Hybrid Interface. International Community on Auditory Display.

Please detail the nature and extent (%) of contribution by the candidate:

Matt Ward assisted with data collection (50%). He was also responsible for all the software development involved with the experiment which included creating the programme that ran the experiment, data collection tools and analysis of the data. William (Deak) Helton provided guidance regarding experimental design and suggested some changes that would make the experiment feasible. Similar guidance was provided by Mark Billinghurst regarding the use of a spatialised auditory cues and choice of the wearable interface.

Gun Lee’s contribution extended to reviewing the data that was analysed. Mark Billinghurst, William (Deak) Helton and Gun Lee helped review the paper. Their suggestions helped organise content in manner that was relevant to the target conference’s submission guidelines. Their editing contributions and suggested changes have contributed to less than 4% of the final papers.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Mark Billinghurst* Signature: M Billinghurst Date: *14/1/2018*

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 4 *of the thesis is a reproduction of the work undertaken and published in collaboration with Matt Ward. Results from this work have been presented at the following conference.*

2016 Human Factors and Ergonomic Society (HFES) Annual Meeting

Barde, A., Ward, M., Helton, W. S., Billinghurst, M., & Lee, G. (2016, September). Attention Redirection Using Binaurally Spatialised Cues Delivered Over a Bone Conduction Headset. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1534-1538). Sage CA: Los Angeles, CA: SAGE Publications.

Please detail the nature and extent (%) of contribution by the candidate:

Matt Ward assisted with data collection (50%). He was also responsible for all the software development involved with the experiment which included creating the programme that ran the experiment, data collection tools and analysis of the data. William (Deak) Helton provided guidance regarding experimental design and suggested some changes that would make the experiment feasible. Similar guidance was provided by Mark Billinghurst regarding the use of a spatialised auditory cues and choice of the wearable interface.

Gun Lee's contribution extended to reviewing the data that was analysed. Mark Billinghurst, William (Deak) Helton and Gun Lee helped review the paper. Their suggestions helped organise content in manner that was relevant to the target conferences submission guidelines. Their editing contributions and suggested changes have contributed to less than 4% of the final papers.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Mark Billinghurst* Signature: M Billinghurst Date: *14/1/2018*

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 4 *of the thesis is a reproduction of the work undertaken and published in collaboration with Matt Ward. Results from this work have been presented at the following conference.*

2016 Human Factors and Ergonomic Society (HFES) Annual Meeting

Ward, M., Barde, A., Russell, P. N., Billinghurst, M., & Helton, W. S. (2016, September). Visual cues to reorient attention from head mounted displays. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1574-1578). Sage CA: Los Angeles, CA: SAGE Publications.

Please detail the nature and extent (%) of contribution by the candidate:

I assisted Matt Ward with approximately 50% data collection. I also reviewed the paper for readability but did not directly contribute to the writing. Mark Billinghurst contributed by means of his supervision and guidance and also assisted with editing the paper. William (Deak) Helton contributed in a similar capacity. He also helped expand the initial idea for the experiment and suggested practical changes in the experiment which made it feasible to run. He also provided editing and assisted in reducing the length of the paper to meet publication guidelines. His editing contributions make up less than 5% of the final paper.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Mark Billinghurst* Signature: M Billinghurst Date: *14/1/2018*

Table of Contents

List of Figures

Sketches on pages 41, 58, 77, 96 and 115 were used as part of posters designed to recruit participants for user studies during the course of the PhD. These have been hand-drawn by Shailee Adke.

List of Tables

Chapter I

1. INTRODUCTION

This thesis covers our examination of auditory and visual cue design concepts and their implementation in a wearable bi-modal interface. Through our research we examine the current state of wearable interfaces and propose a novel design approach incorporating spatialised auditory and visual feedback for an unobtrusive wearable interface.

As an audio engineer and sound designer I have always been concerned with the quality of sound that one hears. My firm belief that sound does more than simply provide 'background noise' to visuals on screens has led me to constantly strive to achieve better ways of implementing it and championing its cause for visual media. My initial forays into research focused heavily on spatialisation of sound over loudspeakers. That was before I encountered the fascinating worlds of Virtual Reality (VR) and Augmented Reality (AR). Naturally, I was drawn to VR since it seemed the most natural medium for a person with my background to move towards. But since my initial encounters with VR, I have gravitated towards AR. I have grown to enjoy the challenges that AR brings to the table. I liken AR and VR to live sound mixing and mixing in the studio respectively. Both require a certain amount of skill and understanding of sound to execute well; but in AR the use of sound must be such that users hear no anomalies in the synthetic audio environment as they move around in a real one. The design and implementation challenges that such a medium brings along with it have been great motivating factors for pursuing research in the field.

1.1 Motivation

The motivating factor for undertaking research in the field of wearable interfaces has been to explore bi-modal interaction methods in this domain. The predominantly visual nature of the technology and its associated pitfalls has led us to examine alternate means of information presentation for wearable interfaces. Mobile phones and tablets have now become a staple in our lives [1] [2]. In a matter of a decade and a half wide spread proliferation of these devices has resulted in an information revolution. Today, access to information via these devices is only an arm’s length away. Our devices are now intrinsically intertwined with our lives, and we now depend on them for everything from mail to finding our lost items. But this has brought with it a unique problem; to get the information we need, we must look at a screen. This form of interaction occupies nearly all our time when using these devices. Besides the social implications of such behaviour, its effects on other aspects of our lives can be profound. Overwhelmingly, and not surprisingly, safety is our major concern when operating such devices. A large number of us are guilty of using cell phones while driving. Accident statistics [3] [4] [5] show that increasingly, mobile devices are responsible for accidents that are taking place on our roads. Distracted drivers are as much of a danger to other road users and pedestrians as drunken ones. Most countries now place severe restrictions or completely ban the usage of mobile devices while operating motor vehicles.

The overtly visual nature of our interactions with devices can often not only prove harmful [3] [5] [4], but limits our interactions with the world around us [6]. With our visual faculty completely occupied by a screen, we tend to completely lose awareness of the environment that we are in. Naturally, the best way to address such an issue is by using sound as an alternate means for presenting information. Initial exploration revealed that there existed a small set of ‘audio only’ interfaces in the consumer market to address the problem that visual interfaces posed. These consisted of Google’s Pixel Buds [7], Amazon’s Eco line of audio devices [8], both powered by their respective voice assistants, the Bragi Dash [9] and Doppler Lab’s Here One [10]. The dearth of such interfaces seemed unusual, given the problems posed by visual only mobile interfaces were well documented and that as a result there existed a gap in the market for such an interface. A more extensive investigation revealed that there were indeed a number of concepts that existed for such interfaces. In the research arena, audio only interfaces such as Audio Aura [11], Nomadic Radio [12] and the Spatialised Progress Bar [13] are some examples of fascinating interfaces that used spatialised audio to provide their users with information in an eyes free manner. However, not all information can be represented aurally and these interfaces make use of headphones, earphones or inconveniently places loudspeakers to deliver the auditory cues.

In this thesis we explore the use of spatialised auditory feedback over a bone conduction headset coupled with minimal visual feedback. The bone conduction headset leaves the ears open to perceive the natural acoustic environment and is able to deliver auditory cues. Being able to hear the environment that surrounds you is important in order to assess what is going on in the regions of the environment that vision cannot cover. Walker et al. [14] [15] and

Stanley [16] have clearly demonstrated that the bone conduction headset can be used as a viable alternative to headphones and loudspeakers as a personal auditory display device to provide spatialised auditory cues.

There are two overarching design outcomes that have provided the impetus for this thesis:

- Bi-modal information delivery.
  A bi-modal approach to information delivery is better than relying on a single faculty. Since our perception of the environment around us is an amalgamation of our senses, it is only logical that an interface that mediates information exchanges between the environment and the user makes use of these senses in as natural a way as possible. We have chosen to use the two most dominant senses of audition and vision since they provide us with the spatial awareness required to accurately perceive the environment around us.
- Contextually relevant and mode appropriate information delivery.
  A contextually relevant delivery of information is critical to the success of a wearable interface. The ability to deliver relevant information based on the real-time needs of the user will make such an interface far more usable in every respect in comparison to a device such as the smart phone [17]. An unobtrusive delivery of this information is also essential to its functioning. This is where a mode appropriate method of information delivery based on contextual sensitivity comes into play i.e. information that can be delivered via the auditory channel need not utilise the visual channel and possibly distract the user during a task such as driving. An example of this would be the delivery of a text message accomplished via a text-to-speech interface versus presenting the text on a screen.

1.2 Research Approach

The focus of this thesis is to explore means of designing and studying interactions mediated by a wearable, bi-modal interface. Our aim is to explore how interactions for a bi-modal interface can be constructed based on the design outcomes listed in the previous section. This entails adopting a two-pronged approach to achieve our design goals. The first involves the study and documentation of the capability of the hardware we plan to use for our experiments, while the second explores the aspects of information delivery design which will govern the mediation of interactions by the interface between the user and his/her surroundings. The key research questions that such an approach raises are:

1. What are the advantages of using sound in wearable interfaces?

2. How can we replicate natural auditory perception to make full use of the auditory faculty?

3. What are the limitations of visual information presentation on the current generation of wearable devices that we believe can be remedied by the use of sound?

4. What audio or hybrid audio-visual solutions have been proposed to address these shortcomings?

5. Do these solutions adequately address the issues raised in Q3?

6. If not, where are they lacking and is there an alternative?

7. Does the alternate solution stand up to rigorous, empirical testing?

8. How well does the proposed solution work in a real-world environment?

The questions listed above lend direction to our research by laying out the fundamental issues that need to be addressed as part of this investigation. It is imperative for us to understand existing issues and proposed solutions to those issues before constructing our own arguments. The use of spatialised auditory and visual cues to deliver information as part of a wearable interface utilising a bone conduction headset is a novel but untested approach. Answers to these questions help us validate equipment choices and design decisions made as part of our research.

To this end an extensive literature review and five user studies were carried out. The first three user studies explored the potential of using the bone conduction headset as part of a wearable bi-modal interface. Two of these experiments looked at how well a binaurally spatialised auditory cue was reproduced over a bone conduction headset. The third explored the relationship between auditory and visual cues via an experiment designed to test the limits of the ventriloquist effect in AR. The fourth and fifth experiments were application based studies designed to test auditory and visual cue designs and presentation combinations in simulated and real-world scenarios.

All data obtained from these experiments was analysed and results obtained from such analyses have been interpreted, presented and discussed in Chapters 3 and 4. Verbal questioning was utilised to gain a more ‘subjective understanding’ and was primarily used for the studies described in Chapter 4. Participants in these studies tended to be more vivid in their descriptions when questioned on their experience of using the interface compared to the questionnaires filled out after the studies. Observations made by myself and my collaborators were mostly limited to how participants reacted to stimuli or a given task during the studies. The author noted any unusual behaviour or perception exhibited or reported by participants.

A secondary, but important, consideration when planning the research approach was to obtain results that were of high ecological validity. To enable this we adopted a two-fold strategy which involved the use of consumer grade, inexpensive, off-the-shelf components and testing our interface prototype in real-world environments or those that mimicked real-world environments. Using inexpensive, consumer grade technology allows for replication without the need for proprietary hardware or software components, and our test environments closely resembled real-world scenarios which greatly enhance the ecological validity of our results. Finally, with the usability of our planned prototype in mind, we have also given consideration to the physical nature of the delivery platform. This is an often overlooked but important aspect, since ergonomic comfort is essential for an interface to be genuinely useful. This will ensure that the interface is unobtrusive and feels less like an intermediary between the user and their environment.

1.3 Structure of the Thesis

The following chapters in this thesis are structured in a fashion such that they answer the research questions presented in Section 1.2 in an ordered manner.

Chapter 2 addresses questions 1 though to 6. This chapter covers literature related to auditory and visual perception. Audio-visual perception is also covered briefly. This chapter also presents a review of the wearable technology over the years. Some important interfaces and proposed concepts are covered to give the reader an overview of how technology in this domain has evolved. A broad range of issues that affect the usability of current AR interfaces is touched upon. The chapter concludes with a recommendation for the use of the bone conduction headset as part of a wearable interface and three main queries resulting from this proposal.

Chapter 3 addresses question 7. Keeping in mind the delivery format proposed for sound in Chapter 2, it is important that its suitability as part of a wearable interface be established. In order to do so, we re-visit some of the studies carried out in auditory perception research and replicate them with our chosen audio delivery format. The studies we have conducted allow us to directly compare results with the only existing results of their kind [16] that we are aware of. Such direct comparisons provide a good barometer of the performance of the selected delivery format and as a result its usefulness as part of a wearable interface. The first two experiments described in the chapter cover psychoacoustic studies related auditory perception of binaurally spatialised stimuli over a bone conduction headset in the horizontal and vertical planes. The third experiment explores the relationship between the auditory and visual factors affecting perception.

Chapter 4 covers application based studies which **addresses question 8**. Once we have validated and established the workings of the bone conduction headset, we move on to the application phase. Here a prototype of our interface, made by combining the Google Glass and a bone conduction headset, is put to the test. This chapter covers two application based studies that are designed to evaluate the usability of such an interface in two different scenarios. The first replicates a scenario which emergency service personnel are likely to encounter in the event of a natural disaster, in this case a blue-sky eruption [18]. The second experiment replicates a more mundane, domestic scenario that we may have encountered; searching for a misplaced set of keys. The results of these studies demonstrate that a wearable interface incorporating bi-modal feedback has the potential to significantly decrease task completion times and improve accuracy. The bi-modal feedback is also shown to reduce the cognitive overload associated with trying to monitor multiple data streams on a visual only interface.

We summarise our work in **Chapter 5** and present some closing arguments for the use of a bi-modal feedback mechanism in wearable interfaces. Some design guidelines for bi-modal interfaces are touched upon, and future directions to further the work we have carried out are discussed as well.

1.4 Research Contribution

The findings resulting from this thesis contributes in two ways to the existing body of knowledge in human-computer interaction studies. Firstly, the outcomes of our experiments will contribute directly to the understanding and advancement of concepts to be implemented when designing a bi-modal, wearable interface. These design concepts stem from studying and understanding the operational parameters of the devices that make up the interface. Secondly, the work in itself is novel in that it addresses the issue of information presentation for wearable interfaces in a unique manner. To our knowledge an interface similar to our prototype does not exist. We have not only proposed a novel method, but also implemented and validated its functionality.

Chapter II

2. BACKGROUND

The previous chapter introduced the motivation behind the research focus of this thesis. It also covered research questions that this thesis has aimed to address in some detail. In order to address these research questions it is first important to look at the work that has already been undertaken by previous researchers. What are the problems that have been identified? Do these problems still persist? If so, how has enquiry into these issues proposed to or addressed the problems? Answers to these fundamental questions provide a measure of the progress made to date, and potentially, a roadmap for the future. Since this thesis deals with the design, development, implementation and evaluation of a wearable interface that provides visual and spatialised auditory feedback to its user; it is only natural that we look at visual and auditory cues in isolation. It is equally important for us to look at how these two types of cues work in conjunction with each other. Aspects related to the cognition of inputs from these cues also require some examination. Lastly, the study of wearable interfaces from the past and present and their evolution over time itself is essential.

To address these aspects of the research study, and provide a comprehensive overview of the parts that make it a whole, we have divided the remainder of this chapter into the following sections:

- Psychoacoustics and Auditory Perception
- Visual Perception
- Audio – Visual Cognition
- Wearable Computing – A History

This chapter will conclude with an overview of some wearable interfaces utilising auditory feedback, visual feedback and a combination of the two followed by a timeline detailing the evolution of wearable interfaces. It will be clear from this overview of the interfaces that a 'holistic solution' to the issue of information presentation on a wearable computer has yet to be addressed. A large part of the work undertaken appears to have presented the use of a single faculty – auditory or visual – as a solution to the problem. What little research has been undertaken in the use of both the faculties together for information presentation often relegates the use of auditory feedback to a secondary, less utilitarian position [19].

2.1 Psychoacoustics and Auditory Perception

The Merriam Webster dictionary defines psychoacoustics as "*a branch of science dealing with the perception of sound, the sensations produced by sounds, and the problems of communication"*[20]. For the purposes of this thesis, the first two parts of the definition dealing with the perception of sound and the sensations it produces are of most relevance. While an all-encompassing study dealing with the "*problems of communication*" is beyond the scope of this thesis, the issue of communication is still of interest to us. After all, the thesis does explore issues regarding information presentation on wearable interfaces. A moderate understanding of psychoacoustics is essential for a person designing an interface that involves the use of sound. This allows a designer to foresee user and interface interactions, and create experiences that are most meaningful and constructive. In this section we will briefly address the perception of sound and sensations produced by sound in order to better understand the mechanisms of the auditory faculty.

2.1.1 The Perception of Sound

'Sound' is mechanical vibrations resulting from the movements of molecules of an elastic medium [21]. These mechanical vibrations (waves) consist of compressions and rarefactions of the medium through which the wave propagates. For the most part, sound propagation is usually assumed to be through air. However, sound can also propagate through liquids, solids and gaseous materials as well. The perception of sound begins with arrival of the wave at the tympanic membrane, commonly referred to as the ear drum [21] [22]. The incident sound wave sets the ear drum into motion which then transfers these vibrations onto the middle ear. These mechanical vibrations are then converted to electrical impulses in inner ear and transmitted to the brain through the auditory nerve. Two major aspects that govern the perception of sound at the ear are:

- **Frequency**: Humans are generally thought to be able to perceive frequencies between 20 Hz and 20 kHz, though the upper limit is closer to 17 kHz in adults [23] [24]. How frequencies affect the sensations resulting from a sound that one hears will be covered in the next section (2.1.2).

- **Loudness, Intensity and Hearing Thresholds**: In the pure psychoacoustic sense loudness is subjective and therefore may seem relatively misleading when used to explain auditory perception. However, for the purpose of this thesis and to provide a brief overview of the processes that make up hearing and auditory perception it is adequate.

  For a sound to be auditorily perceived it must first be heard i.e. the sound waves must reach the tympanic membrane and excite it. An incident sound wave, being a mechanical wave, must apply some pressure on the tympanic membrane to produce the required vibration to generate the perception of sound. This implies that a sound has to be of a certain ‘loudness’ or intensity to be heard. Objective measurements of loudness can be made to detect the levels at which a sound is just noticeable or can cause pain. These limits are commonly referred to as hearing thresholds. The threshold of hearing is commonly taken to be 0dB SPL (Sound Pressure Level), while the threshold of pain is set between 115dB SPL and 140 dB SPL [25]. Therefore, from the above explanation it becomes clear that a sound with a greater intensity generates a higher sound pressure level at the tympanic membrane and consequently sounds ‘louder’. A sound of 0 dB SPL generates a pressure of approximately 2×10^{-5} Pa (Pascal) [26]. It must be noted that hearing thresholds vary depending on the spectral content (frequency distribution) of the sound. The thresholds listed here are those obtained from measurements made of a 1 kHz test tone played at a distance of 1m [27].

2.1.2 Sensations Produced by Sound

In the context of this thesis, the sensations produced by sound refer to the localisation of sound. We have chosen to put forth localisation of sounds as a ‘sensation’ since it is this aspect of auditory perception that makes it tangible. It provides us with a picture of the environment outside the visual field. Localisation of an auditory source depends broadly on three major aspects of auditory perception [28] [29] [30]. These can be categorised as:

- Binaural Cues
- Spectral Cues

- Head Related Transfer Functions (HRTF)

Binaural Cues: Binaural cues, as the name implies, refers to cues that relate to the use of both the ears. Early experiments by Lord Rayleigh lead to establishment of the basic tenets of binaural hearing and auditory localisation known today as the *duplex theory* [31] [32] [33]. The duplex theory states that we use two distinct methods to localise sounds based on the cues the ears receive. These localisation methods are classified based on:

- Interaural Intensity Difference (IID): Also known as the Interaural Level Difference (ILD), IIDs are a measure of the difference in intensity of sound as the two ears. The intensity differences are a result of the ‘head shadow’ effect [34] [35] [36] i.e., the head casts an ‘acoustic shadow’ on the contralateral ear resulting in a lower intensity of the incident sound at the ear [37] [38] (Figure 2.1). Interaural Intensity Differences have been shown to be active mainly for frequencies above 3000 Hz [33] to 4000 Hz [39] [40]. At and above these frequencies, the wavelength of the incident sound wave approaches or is smaller than the diameter of the head resulting in partial absorption and reflection. This results in the head casting an ‘acoustic shadow’ mentioned earlier, leading to a lower level of the sound reaching the contralateral ear.

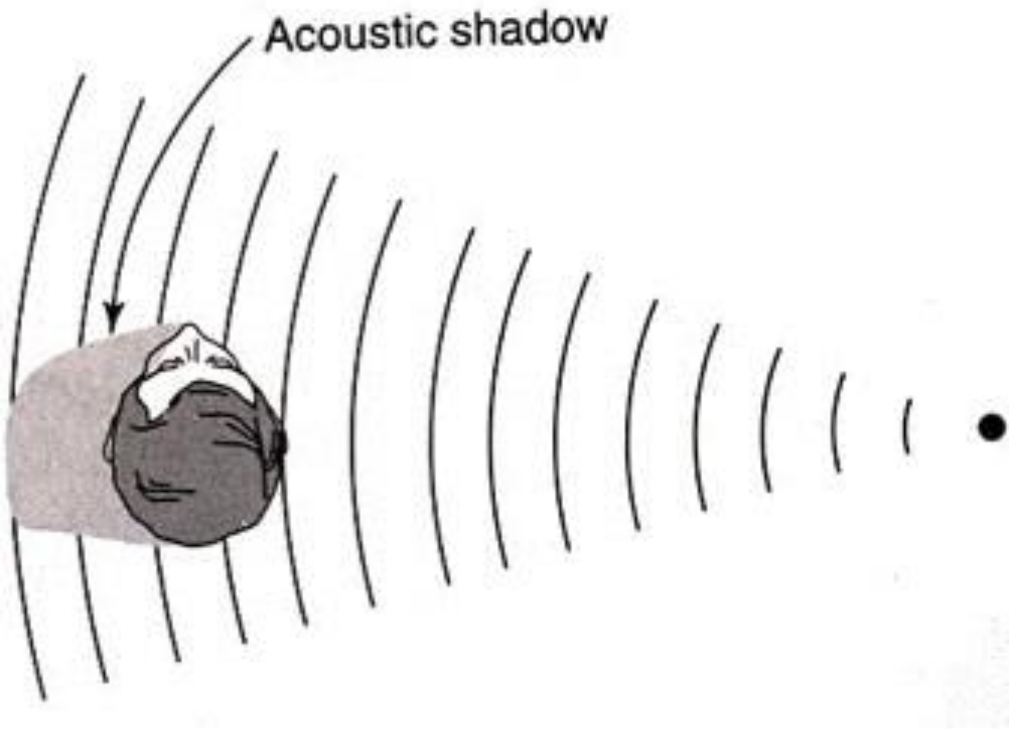


Figure 2.1: Interaural Level Difference (ILD) and the ‘Acoustic Shadow’ at the contralateral ear [41]

- Interaural Time Difference (ITD): Also referred to as Interaural Phase Difference (IPD), ITDs are a measure of the difference in time at which a wave arrives at the two ears [32] (Figure 2.2). Consequently, a difference between the time of arrival at the ears is also assumed to result in a difference in phase that plays a part in auditory localisation [42]. The Interaural Phase Difference is shown be effective up 1300 Hz [33] – 1500 Hz [37] [43] [40].

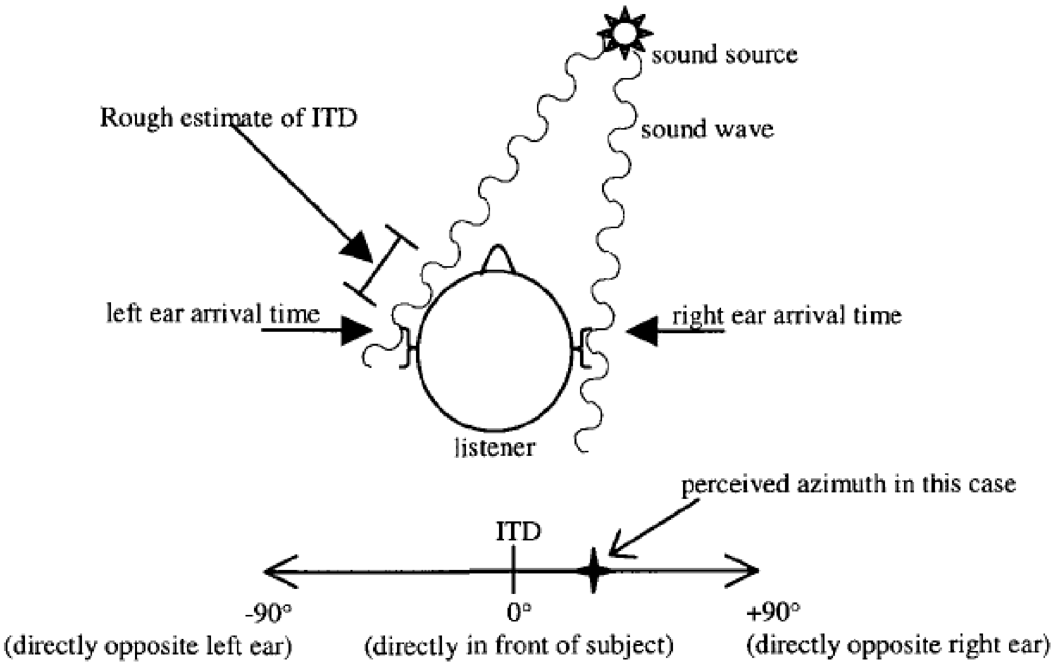

sound source
Rough estimate of ITD
sound wave
left ear arrival time
right ear arrival time
listener
perceived azimuth in this case
ITD
-90°
(directly opposite left ear)
0°
(directly in front of subject)
+90°
(directly opposite right ear)

Figure 2.2: Interaural Time Difference (ITD) [29]

It must be noted that these two mechanisms operate in unison and the localisation of the sound source is a result of a combination of IID and ITD cues i.e. localisation depends both on the time of arrival of the first wave front at the two ears [44] and the intensity of arriving wave front [45]. While these two cues are considered of fundamental to localisation in the horizontal plane, they still fail to explain the lack of localisation accuracy observed for sounds presented in this plane. Two of the most common problems reported by researchers are; a significant lack of localisation accuracy at frequencies between 1300 Hz to 4000 Hz [33] [36] [37] [42], and front-back confusions[1] [38] [42] [46] (Figure 2.3). The most commonly accepted explanation for these inconsistencies in localisation is the use of pure

---

[1] Front-Back confusions refer to the percept of localising a sound source that is actually in front of the head to the back. This phenomenon generally occurs when the ears receive identical binaural cues and therefore cannot distinguish the position of the sound source in the space around the head.

tone stimuli [42] [47]. Since our natural acoustic surroundings do not contain pure tones, it is only natural that our ears respond differently when confronted with such stimuli. Therefore, for auditory localisation to function optimally there must be other mechanisms at work that help us localise auditory sources. To unravel these mechanisms requires a shift from use of tonal stimuli to broadband stimuli. The use of broadband stimuli brings into focus frequency based filtering also known as spectral cues. This next sub-section briefly covers some aspects of spectral filtering and how the ‘colouring of sound’ by our pinna affects localisation.

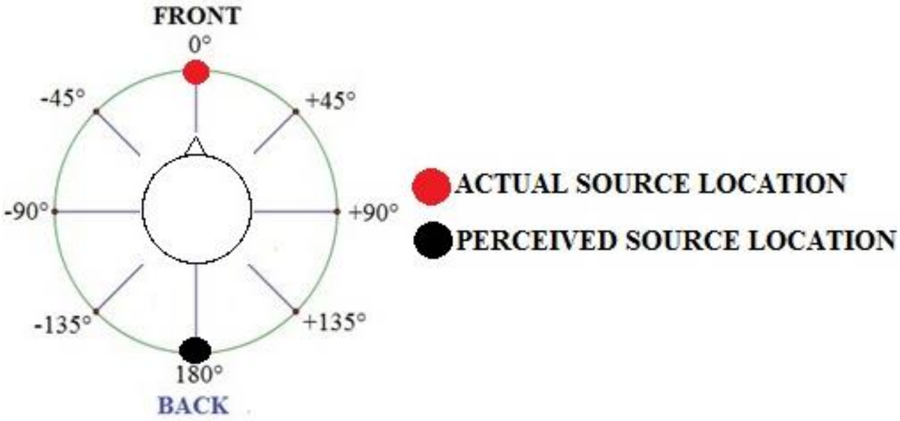


Figure 2.3: Front-Back Confusion. Sound source localised to a position diametrically opposite to its true location

Spectral Cues: An incident sound wave first encounters the pinna or the outer ear on its way to the ear drum. The grooves and notches of the pinna modulate the mid and high frequency content of this incident sound wave in a time and direction dependent manner [28] [37] [38] [40] [48]. It is this frequency based modulation of the incoming sound which allows for ambiguities such as front-back confusions to be resolved [38] [42] mentioned earlier in the section. Spectral cues have also been shown to help with localisation in the vertical plane [49] [28] [50]. It must be noted that the usefulness of spectral cues is directly related to the bandwidth of the signal. Wide bandwidth signals have been shown to generate better, more effective spectral cues than those with a narrow bandwidth [39] [42]. The shape of the pinna and how the incoming sound waves interact with each other in the concha also play an important role in determining how the filtering process takes place [28] [48] [40]. The fundamental nature of spectral filtering has been explained by Rodgers [40] by examining a combination of the incident sound and the first major coherent reflection. As the angle of incidence of the incoming sound wave changes, the distance between the ear canal and the first significant reflecting surface changes (Figure 2.4).

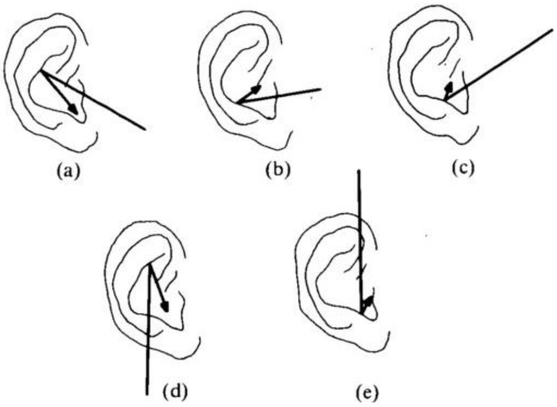


Figure 2.4: Spectral filtering – Interaction of an incoming sound wave with the pinna [40]

To illustrate the point, consider figure 2.4(a). We can see from this figure that the first prominent reflection of the incident sound will be created by the top of the concha. In contrast to this, figure 2.4(c) shows an incident sound wave arriving from the top, with the first prominent reflection resulting from the bottom of the pinna. It can be clearly seen that there exists a definite difference in the time of arrival of both these reflections at the ear canal. However, what is more interesting are the changes that these different angles of incidences cause in the spectral make-up of the sound reaching the ear canal. For figure 2.4(a) notches or 'spectral minima' are seen at two points spanning the bandwidth the sound occupies. These are observed between 6 kHz – 7 kHz, and 18 kHz – 20 kHz. Similarly, for figure 2.4(c) the spectral minimum is recorded at about 13.5 kHz. This aspect of the pinna's effect on the spectral distribution of the incoming sound has also been documented by Zachrov et al. [51]. It is thought that these differences in the spectral characteristics imposed by the pinna on the incoming sound are responsible for resolving front-back and up-down confusions that may arise from localisation based purely on binaural cues.

Head Related Transfer Function (HRTF): HRTFs can, very broadly, be considered an extension of spectral cues. In addition to the pinna, Head Related Transfer Functions (HRTFs) also incorporate the direction dependent frequency filtering due the effects of the head and torso [29]. They are an objective measure of the frequency dependent filtering that takes place within the ears due to effects of the pinna, head and torso. Typically, HRTFs are measured for both ears with a source placed at a fixed distance from the head at several azimuths and elevations [29] [51]. HRTFs obtained by such methods are generally referred to as ‘individualised HRTFs’ since the measurements obtained are unique to the individual whose transfer characteristics are being measured [29] [52]. Figure 2.5 demonstrates how these measurements are obtained in an anechoic chamber.

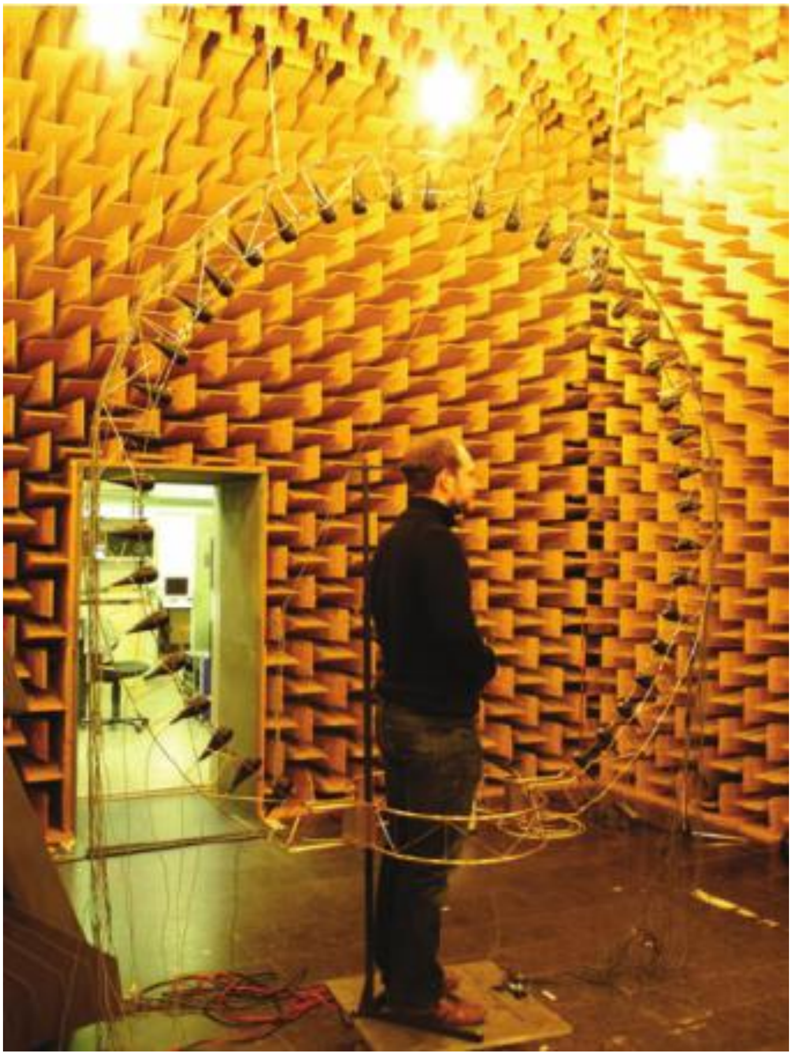


Figure 2.5: Individualised HRTF measurements being made in an anechoic chamber [53]

Similarly a more ‘generalised’ set of transfer characteristics can also be measured using a Head and Torso Simulator (HATS) more commonly referred to as a ‘dummy head’ [54]. The dimensions of the dummy head are based on average measurements of the head, torso and ears obtained from a large number of subjects. Figure 2.6 demonstrates the use of a dummy head in an anechoic chamber to obtain a generalised or ‘non-individualised’ HRTF set.

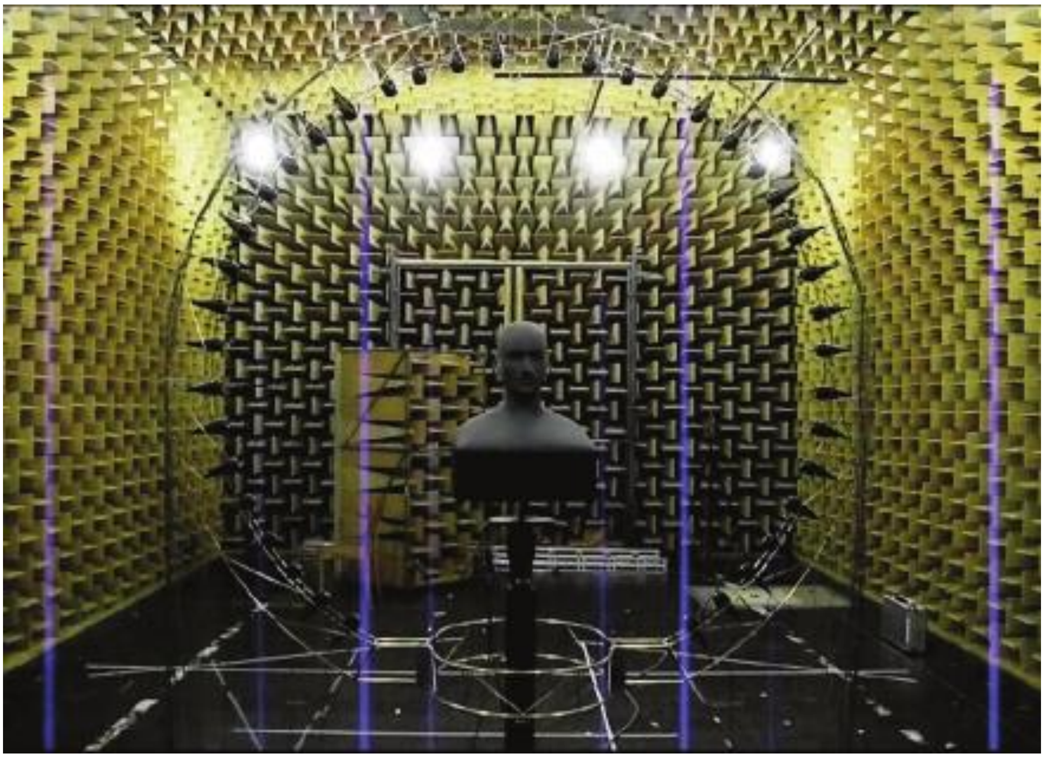


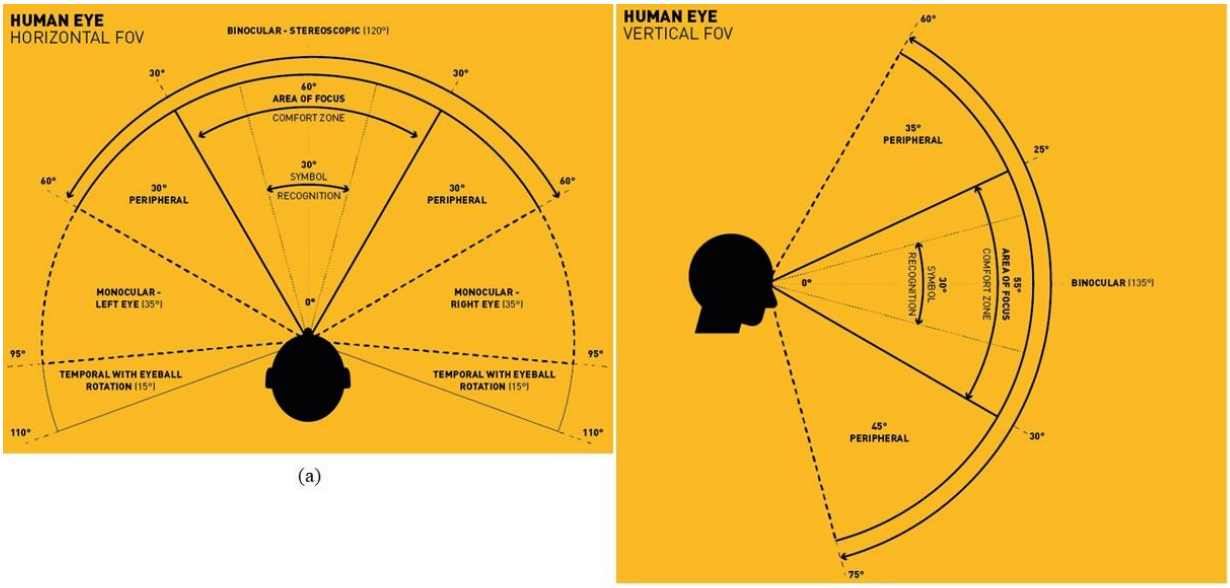
Figure 2.6: Generalised HRTF measurements being made in an anechoic chamber using a ‘dummy head’ [53]

A variation of the use of the non-individualised HRTFs has been demonstrated by Wenzel et al. [55] [56]. Their experiments make use of HRTFs obtained by a ‘good localiser’. These HRTFs are then convolved with the sound source and presented to subjects to make localisation judgements. Their results demonstrate that such a method results in good localisation accuracy even though the HRTFs used in the experiment do not belong to any of the participants. There is some evidence to suggest that over time participants tend to learn and adapt to newly presented HRTFs, leading to good localisation accuracy [57]. With the advent of Virtual Auditory Displays (VADs) the use of HRTFs is seen as a key factor in being able to deliver believable auditory percepts in both real and virtual environments [56] [54].

2.2 Visual Perception

Vision is thought to be our predominant perceptual faculty. It is, quite literally, a ‘window’ to the world we live in. The human eyes perceive approximately 80° to 100° in the horizontal plane and 60° to 70° in the vertical plane in each direction from a central point of view [58]

[59] (Figure 2.7). The area of focus decreases from the central visual field out towards the periphery. The high focus region covers an area spanning approximately 2° on either side of a central view point. Several studies [60], [61] demonstrate that we often use vision to corroborate our perception of the space around us. These studies have also shown us that vision behaves as somewhat of an ‘anchor’ i.e. auditory percepts supported by vision always appear to be closer to the visual stimuli than they physically are [60] [61]. Engaging this faculty seems the most natural manner in which information can be presented to a user.



(a)

(b)

Figure 2.7: (a) Horizontal Field of View (b) Vertical field of View [62]

The relative strength of the visual faculty is also an indicator of the importance it holds to us in a veritable arsenal of perceptual tools we possess. Binocular vision can be seen as an evolutionary step [63] in the dominance of vision over other faculties. None-the-less, the very factors that make vision the dominant perceptual input also limit its uses i.e. inability to focus on several information streams simultaneously. Dominant visual cues take precedence at the expense of weaker ones [64] making it almost impossible to represent everything visually due to the inherent workings of visual perception. The dependence of visual perception on a host of complex cues [64] – depth, colour, motion – make it a complex system to replicate in a wearable computing environment. This means that presentation of such cues, while the visual faculty is already occupied, could distract a user simply because the newly presented cue attracts more attention since it possesses to a greater degree one of the attributes of a visual cue mentioned above. A misrepresentation of any of these cues results in an ineffective and often incorrect display of information. Attempting to convey information via multiple information streams only complicates matters. The ability to circumvent such issues is of paramount importance in wearable computing where only information that requires visual attention is made available via that modality. The use of audio cues makes for an ideal means of supplementing the visual display of information. For the purpose of this discussion, we shall confine ourselves to visual perception from an ego-centric point of view. Wearable devices by nature are designed to provide, in most cases, an ego-centric view of the environment around the user be it augmented in some form or virtual. It is important that we are able to replicate perceived visual cues in a real environment in augmented reality in order to provide a seamless augmented interaction environment to the user. Such perceptual cues must extend to static and dynamic cues and even incorporate our perception of very large objects in the real-world [65]. Besides the factors mentioned above, we must also consider the role vision plays in a multimodal interface, particularly the effect of vision on auditory perception. The next section briefly covers this aspect and how the functions of one faculty affect the other.

2.3 Audio – Visual Cognition

Up until now we have looked at auditory and visual perception in isolation. But from our own experiences we know that these two faculties work in unison. In order to present a strong case for a wearable interface that incorporates these two faculties, it is important for us to understand how these work in tandem. Recognising how one faculty influences the other is imperative in being able to design an interface that is both user-friendly and non-fatiguing. To that end, this section will briefly explore how visual factors affect auditory perception. Witkin et al. [60] have demonstrated that seeing a speaker’s mouth move significantly affects the outcome of auditory localisation. In extreme cases, subjects’ perception appeared to shift from the lateralised position back to the centre when the speaker was seen.

An average deviation of 28° for men and 38° for women was reported when auditory and visual cues were displayed simultaneously[2]. Kyto et al. [66] have demonstrated a similar effect in augmented reality which shows a greater angular deviation, between 32° – 45°,

---

[2] Angular difference between the auditory cue and visual cue at which the subjects reported the two as being spatially separated or ‘disconnected’.

before a disassociation between the auditory and visual cues is reported. Anecdotal evidence suggests that we use vision to corroborate auditory localisation judgements made outside our field of view. This has shown to be the case in a study conducted by Jackson [61] where subjects first appeared to listen to an auditory cue, localise it and then use vision to confirm the initial localisation response evoked by the sound source. Results of this study demonstrate two things. Firstly, vision plays an important role in the perception of our environment, but is limited by what is within our field of view. Secondly, auditory perception does not adhere to similar constraints i.e. the ears do not need be pointed in the general direction of a sound source to be able to hear it.

The visual channel is thought to have a processing ability equivalent to 4.32×10^6 bits/sec [67], equalling approximately a 1024 x 1024 bitmap image with 256 colours. This sort of processing power allows for it to perceive a vast amount of information simultaneously. The auditory channel on the other hand is thought to have a much narrower bandwidth of approximately 9,900 bits/sec [67]. This results in a slower and much smaller volume of information that can be processed through this channel. None-the-less, the complementary nature of their functions [61] can be used to our advantage when designing a wearable interface. Since our sight is fully occupied most times and therefore insensitive to minor changes, it is prudent for us to make use of sound. Its ability to filter out or ignore the prevalent ‘steady state’ of an acoustic environment to rapidly detect any anomalies or impulses [39] in the environment is extremely useful in delivering temporal information. The use of the both these faculties in a context appropriate manner when designing wearable interfaces will ensure that interactions with such an interface are non-fatiguing and useful.

From a philosophical standpoint, there is a need to move away from the *pictorialization of sound* [68], but with the understanding that the two faculties work in harmony to afford us a rich representation of the environment around us. Schafer quotes from McLuhan’s *The Gutenberg Galaxy* to impress up on the reader the importance of sound; “*As our age translates itself back into the oral and auditory modes because of the electronic pressure of simultaneity, we become sharply aware of the uncritical acceptance of visual metaphors and models by many past centuries*”[68]. Subsequent sections in thesis, as you will see, both accept and reject this idea. Acceptance comes in the form of acknowledging that visual presentations dominate today’s wearable interfaces (Section 2.4). Rejection of this idea comes through existing interfaces that are aurally dominant. Our own empirical studies demonstrate that neither of them are fool-proof solutions to the issue of information presentation. A middle ground that entails exploiting the workings of both faculties makes for an ideal solution.

The next section covers the history of wearable computing. This section lays out, chronologically, the evolution of the wearable computers utilising some specific examples and a brief description of each. This section will provide the reader a good overview of how the field has evolved over time and where it currently stands.

2.4 Wearable Computing – A History

Before we dive into the history of wearable computing, it is worth looking at what the term stands for and some of the properties associated with wearable computers. Wearable computing, as the name implies refers to computers that can be worn on one’s self. Wearable computers are designed to be worn as clothing, built into clothes [69], [70] or can even be integrated into glasses like Google Glass [71], [72] or a wristwatch [70] [73]. Wearable computing moves the computer from the desktop to the user’s body [69]. It also means that the computer has evolved from a passive device [70] into one which is, in many ways, an integral part of the user. This has led to a change in the way the user interacts with the device and is what sets it apart from more traditional forms of portable communication [70]. Wearable computing is poised to revolutionise the way we interact with our devices and each other. The ability to access information on-the-go without having to deal with bulky peripherals like keyboards or constantly having to retrieve a portable device to look at are just some of the uses that wearable interfaces offer. Before we move further, let us look at three important goals mentioned by Billinghurst and Starner [70] that wearable computers must satisfy:

- Mobility: Over the last decade and a half, large scale proliferation of mobile phones, laptops, PDAs etc. has made the mobility of a communications device more important and ubiquitous than ever before. It is a sign of the times we live in [74] that a computer must be able to go where the user goes. By default a wearable computer satisfies this requirement since it is worn by the user.
- Augmentation: It must be able to augment reality, not replace it. By this we mean that the wearable interface must be able to provide its users with useful information about

the environment they are in. The augmentation could be in the form of computer generated imagery overlaid on the real-world or spatialised audio cues.

- Context Awareness: This is probably one of the most important aspects of wearable computing. With the wealth of information available these days, it has become near impossible to segregate incoming data. A system that is contextually aware could filter incoming information and present it to the user depending on the environment he/she may be in at a given moment. Such a system could also make decisions regarding how an information stream needs to be presented based on several factors that could it be programmed to take into account.

A wearable device also allows its user to become an information as well as ‘experience gathering’ tool. The experience derived from an object is the reason why an object is acquired in the first place [75]. It is important that wearable computers allow us to experience reality in a chosen manner versus the detached relationship we currently share with our computers [70] and mobile devices. Hull et al. [75] have shown how wearable computers can be used to generate stimulating, immersive experiences in a designated space. Billinghurst et al. [76] have on the other hand successfully demonstrated the use of wearable devices in a collaborative work environment. The goal should be to extend such experiences to information presentation without having a primary mode of information display i.e. audio cues or visual cues. A dynamic form of information representation which allows information to be represented with the use of auditory and/or visual cues must be developed to provide seamless interaction within an environment.

Wearable computing has a long and varied history. Several researchers have proposed different methods of information presentation for wearable devices. Chief among these has been the use of the visual faculty. With an ever increasing access to information, this method resulted in an exponentially rising demand on the visual faculty rendering several of these interfaces ineffective as fixed information display and retrieval systems [13]. The end of the 1980s and beginning of the 1990s saw an increased emphasis being placed on the auditory faculty as an alternate means of delivering information. Spatialised auditory feedback which mimicked natural auditory perception was seen as an answer to the problems that visual only visual interfaces faced. The Virtual Auditory Display (VAD) was championed by several researchers [55] [56] [6] [11] for the purpose of information presentation. Applications for the VAD range from complex information handling in attention critical environments to mundane tasks such as e-mail and message notifications. In the following section we list and explain some of the features of wearable devices that have been developed through the years. Our list contains an assortment of devices that incorporate visual or auditory feedback or the two in tandem. It is important to catalogue these devices and understand their functioning in order to build a case for an interface that addresses some of their shortcomings.

The Sword of Damocles

No discourse on wearable technology is complete without acknowledging the seminal work of Ivan Sutherland. *A Head-Mounted Three Dimensional Display* [77] (Figure 2.8a) is one of the first known demonstration of wearable technology, in both the augmented and virtual reality spheres, as we know it today. The device incorporated miniature cathode ray tubes to display simple information to the user. Half silvered mirrors used in the prisms through which the user looked, allowed them to see both the images from the cathode ray tubes and objects in the room simultaneously. The device also included a head position sensor which was used to communicate the position and orientation of a user's head to the computer powering the system. The system allowed users to turn completely around and tilt their head up or down up to 30° – 40°. Though basic in its design and tethered (Figure 2.8b) to a computer which reproduced basic shapes in real-time, the device is considered a breakthrough in wearable technology.

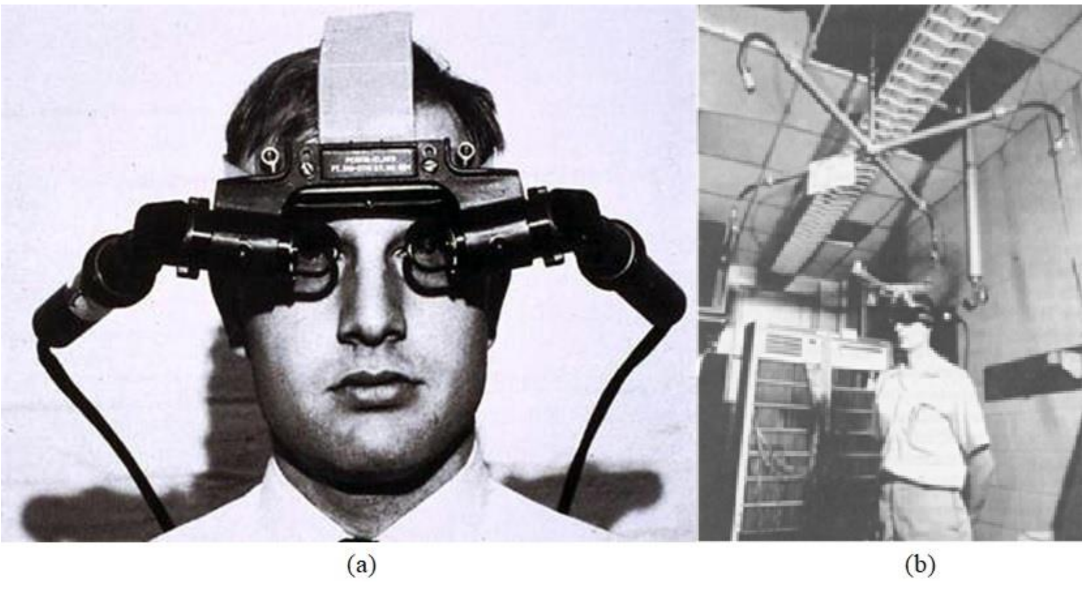

(a) (b)

Figure 2.8: The Sword of Damocles. Ivan Sutherland's head mounted display [77]

Wearable Computer for Three Dimensional Computer Supported Collaborative Work (CSCW)

In their work on computer supported collaborative work environments, Billinghurst et al. [76] have demonstrated the use of use of head mounted displays to accomplish collaborative tasks in a shared space. The study forced two participants to collaborate on a search task. One user was designated the ‘spotter’ and could see all the virtual objects in the shared space. The other user was designated as the ‘picker’. This user’s role was to find the objects made visible by the spotter and drop them over a target. Results from this study demonstrate that a wearable computer mediated virtual collaborative work environment is effective at achieving consistent and relatively fast task completion times. Further explorations of collaboration between two subjects wearing video see-though head mounted displays also showed promise for such a form of interaction. Participants in this study were able to see, via a camera mounted on the head mounted display, what the other person was looking at. This significantly lowered task completion times in comparison to using a head mounted display that fed the image to a monitor, which was then displayed to the user with the head mounted display with the addition of the second user’s graphical notations. While both these displays can be looked upon as exclusively visual interfaces, the use of the instructor’s voice in both experiments is an addition of the auditory element.

Remembrance Agent

The Remembrance Agent [69] developed by Thad Starner and colleagues is a text based wearable interface that attempts to extend the traditional computer interface using wearable technology. It works on the premise that the wearable computer has the ability to display messages unobtrusively or urgently grab the user’s attention. The system was designed to display text and/or graphics in physically meaningful locations in the visual field. The Remembrance Agent was also designed to provide contextually relevant feedback to its user and was envisioned as a constant ‘brain storming system’ capable of providing its user with suggestions that they may never have considered in a certain scenario. Additionally, the system also makes use of video cameras, infrared sensors and bio-sensors in order to learn more about its user and provided a more tailor-made response leading to a seamless interaction between the user and their environment mediated via a smart wearable system.

Nomadic Radio

One of the first audio only wearable interfaces, the Nomadic Radio [6] [12] consisted of shoulder worn speakers and a microphone unit (Figure 2.9). The system was designed as a means to convey messages such as e-mail and voice and provide updates of weather, traffic and any other information the user may specify as being of interest to him/her. A key operational aspect of the device was its ability to present incoming information in a spatialised, contextually aware and preferential manner. It does this by listening to the user's surroundings and learning how the user accesses information presented him/her. The incoming information is scaled in a dynamic manner ranging from ambient auditory cues to voice message summaries based on the user's preferences and the priority accorded to an incoming message. Interaction with the system is achieved via voice and tactile input. The Nomadic Radio makes use of spatialised auditory output to positon cues in a circle around the user's head.

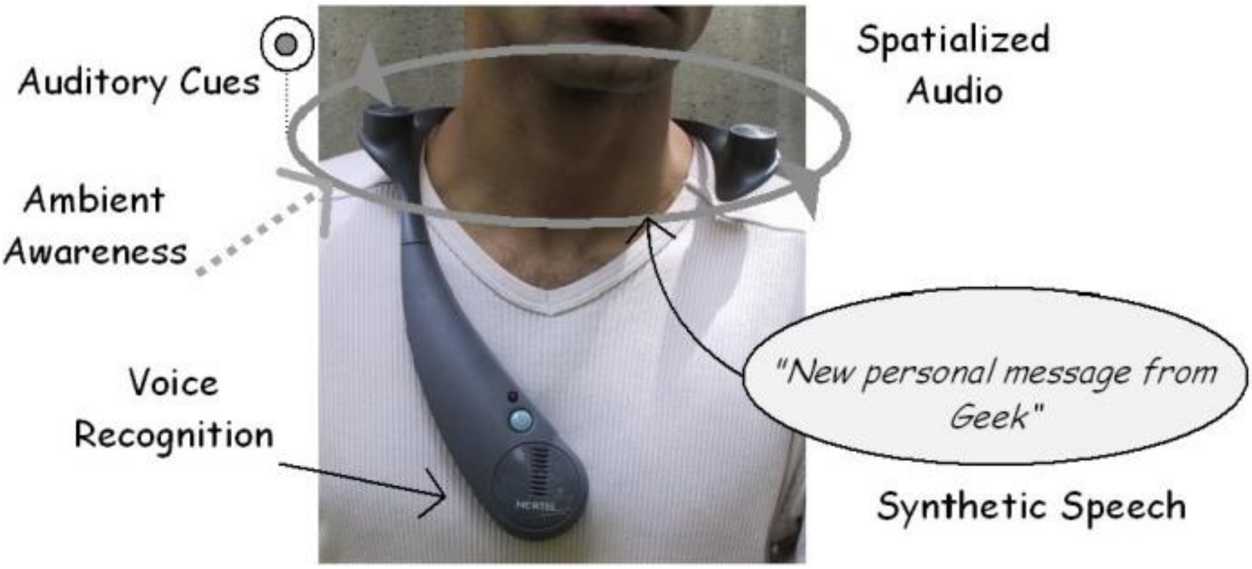


Figure 2.9: Nomadic Radio [17]

Spatial Information Displays on a Wearable Computer

This experiment [78] is one of the first studies that we are aware of that examines the effects of using spatialised auditory and visual cues for information display. The study looks at completion times for a search task using a wearable display. Users were provided with a visual cue, auditory cue or a combination of the two to help them complete the search task. Search times recorded when these cues were presented were compared with those obtained when no cues were provided to the users. Results from this study clearly demonstrated that use of either visual or auditory cues that provide some spatial information was extremely useful. Interestingly, the study also noted that several users found that visual cues ‘overloaded’ their senses. The same appeared to be true of visual and auditory cues presented simultaneously. None-the-less, this study shows us that the judicious use of one of the cues can help in achieving faster task completion times. If the use of these cues can be made context sensitive, such wearable devices have the potential to increase productivity without placing excessive demands on the user’s cognitive abilities.

Audio Aura

Audio Aura [11] was one of the first and highly influential mobile auditory interfaces. The system was designed to work within a building equipped with a network of infrared sensors. Users of the system wore electronic tags which allowed the system to identify and track each one individually. Individualised auditory cues were triggered when a user entered a designated area and were delivered over wireless headphones. The auditory cues provided the user with information such as new e-mails or reminders for meetings. The auditory cues used in the Audio Aura system were not spatialised. The primary focus of the system was to provide ‘serendipitous’ auditory cues to the user. The premise behind this was that the user ‘discovered’ these cues rather than being startled by them. To implement this, the designers of the Audio Aura used environmental sounds that represented an entire ecological system. In the implementation described in their paper, the beach is used as a metaphor to represent several aspects of incoming information. For example, a single seagull cry denoted that there was one unattended or new message that the user had, while several seagull cries signalled that there were multiple messages requiring the user’s attention. Group activity was represented by waves, with louder more close sounding waves signalling heightened group activity. While not context sensitive in the true sense, the system did provide its user with usable cues regarding important activities when he/she entered a specific area.

Situated Documentaries

Situated Documentaries [79] describes an experimental wearable augmented reality interface that enables users to experience media presentations that are actually integrated with an actual outdoor location (Figure 2.10a). The system uses a tracked, see-through head mounted display that overlays 3D graphics and other imagery over the real-world (Figure 2.10b). Integration of GPS based tracking with the displays allows for this wearable system to display or mark points of interest from a distance (Figure 2.10c). As the user approaches a chosen point of interest, such as a building, additional visual markers displayed on the structure inform him/her of various aspects of the building such as the rooms it contains, tours of the structure etc. (Figure 2.10d). The user can then choose from any of these options to access the various bits of information which are then displayed on the head mounted display (Figure 2.10e). The system also provides some non-spatialised auditory output in the form of narrations and non-speech sounds. This system is a great application based example of wearable computing.

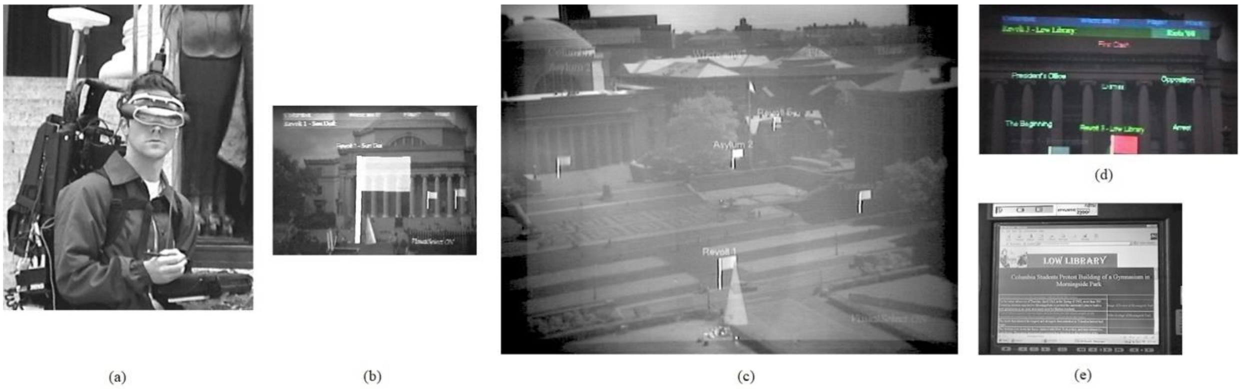


Figure 2.10: Situated Documentaries – A bi-modal, wearable, GPS enabled Augmented Reality (AR) display. (a) The backpack based system with a see-through HMD and a pen-based hand-held computer (b) View through the HMD of points-of-interests in a building (c) View through the HMD of points of interests within a specified area (d) Specific points-of-interest in a building as seen through the HMD (c) Hand-held computer display [79]

Spatialised Audio Progress Bar

The spatialised audio progress bar [13] designed by Walker and Brewster provides an alternative to the mundane task of monitoring the progress of, say a file transfer, on the screen (Figure 2.11). The audio progress bar provides an eyes-free means of monitoring this task via means of spatialised auditory cues placed around the head. The spatial position and presentation rate of two sounds conveys the progress, rate and endpoint of the task. The first sound is played from a fixed position in front of the head and acts as a reference. The second sound communicates the rate of the progress (by way of its angular speed while orbiting the head) and the end of the task (by means of the sound being played at a fixed position around the head). The perception of these two distinct sounds at fixed positions around the head indicates task completion.

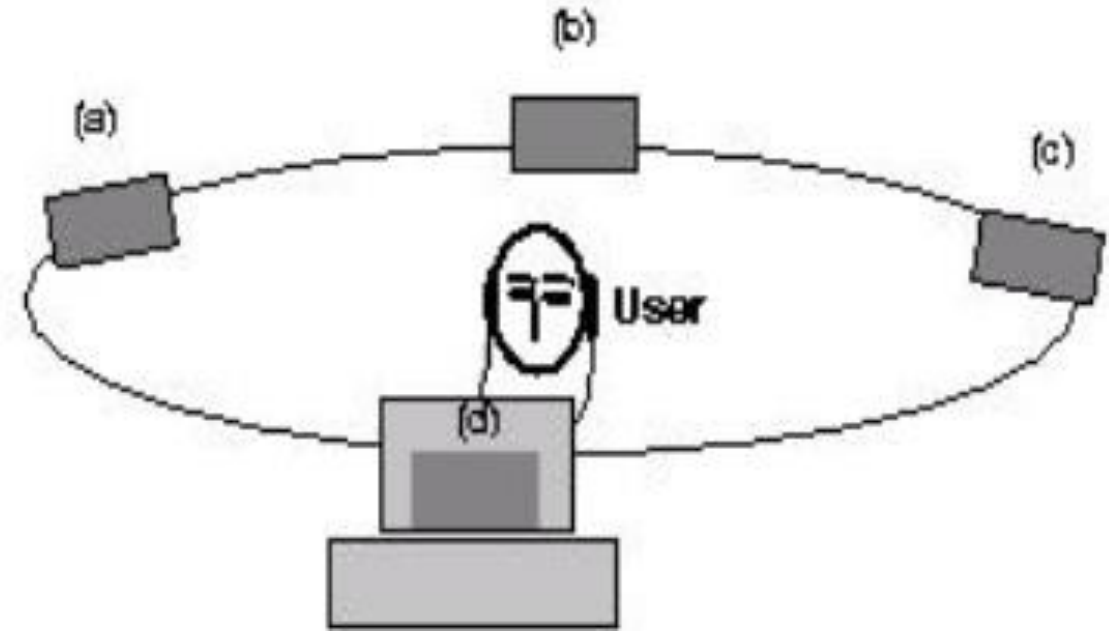


Figure 2.11: Spatialised audio progress bar [13]

Diary in the Sky

The Diary in the Sky [80] explores the technique of using spatialised sound to position sound items according to their semantic content. In the case of the prototype tested here, the sound items represent calendar entries. The entries are ‘time stamped’ and placed on a ‘clock face’ around the user’s head with the position directly in front of the user representing 12 o’clock (Figure 2.12). Such a display is thought to facilitate a better recall of events in comparison to conventional small screen displays.

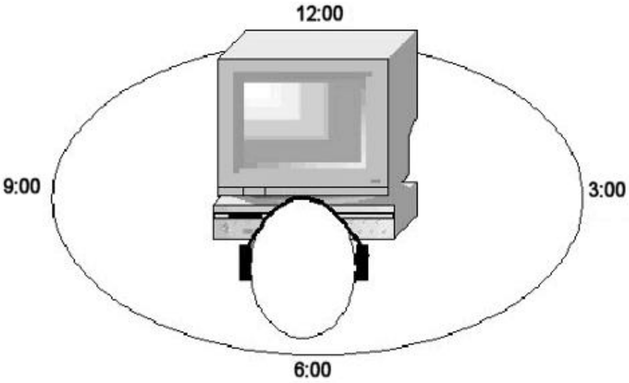


Figure 2.12: Diary in the sky [80]

AudioGPS

The AudioGPS [81] (Figure 2.13) system uses a traditional Global Positioning System (GPS), but with an emphasis on the concept of Minimal Attention User Interface (MAUI). The idea behind this being that the user need only pay some amount of attention to the information being delivered to complete a task, in this case, navigation. This system uses a Geiger counter metaphor to inform users of their distance to the target. Direction mapping is achieved by using equal pitch differences in semitones.

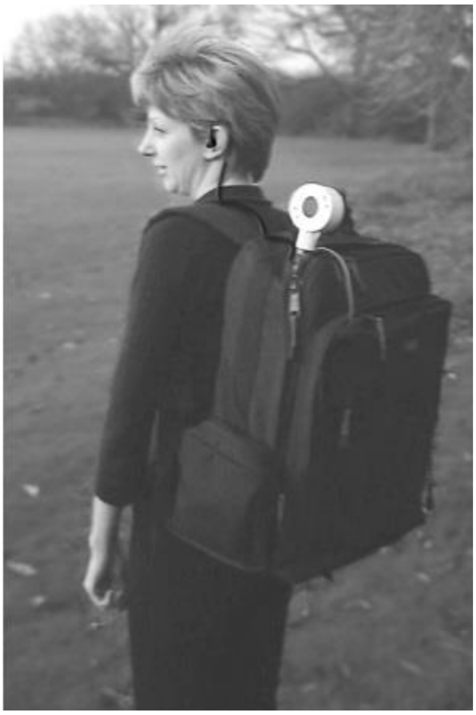


Figure 2.13: Audio GPS – A Minimum Attention User Interface (MAUI) [81]

Spatial Audio Interface for Generating Music Playlists

This interface developed by Hiipakka and Lorho [82] aims to eliminate visual interactions associated with interacting with one's music collection. The interface relies on a spatialised presentation of levels around the user's head. Interaction with the interface is achieved using a tactile input. The system can be used to navigate a large list of musical items organised in a hierarchical structure, and to create personal playlists. The user interface relies on a horizontal and vertical presentation of information to make it easier to distinguish between and access various 'levels' of the organisational structure of the playlist.

Augmented Reality Audio for Wearables

Harma et al. [83] describe an implementation for augmented reality audio that makes use of earphones to deliver both synthetic and real-world auditory cues to the user. Synthetic auditory cues are convolved with generalised HRTFs and presented to the user, while the natural acoustic environment is conveyed via microphones embedded in the earphones. While this system is a good work-around to the problem of acoustic occlusion posed by the use of headphones or earphones, the impractical nature of this solution renders it ineffective. A large amount of calibration and equalisation is necessary to make the surrounding acoustic environment sound 'natural'. Added to that, usability issues such as wire and bone conducted sounds severely limit its usefulness. Similar issues were encountered with an identical system developed by Tikander et al. [84].

System for Wearable Audio Navigation (SWAN)

SWAN [14] [15], developed at Georgia Tech by Bruce Walker and colleagues is a bone conduction based auditory navigation system exclusively for visually impaired users (Figure 2.14). The system uses a bone conduction headset as means to provide spatialised auditory cues to user in order to help them navigate an environment. The use of a bone conduction headset leaves the ears open to perceiving the natural acoustic environment thereby avoiding problems that the use of headphones create as mentioned earlier. The system relies on a Geographical Information System (GIS) to provide its users with wayfinding auditory cues.

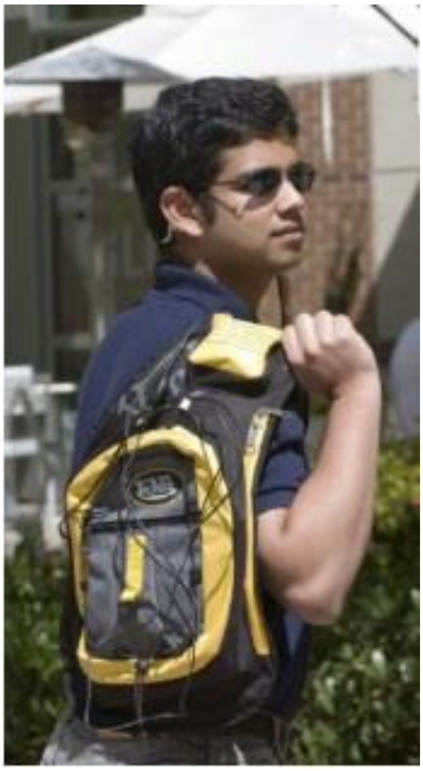


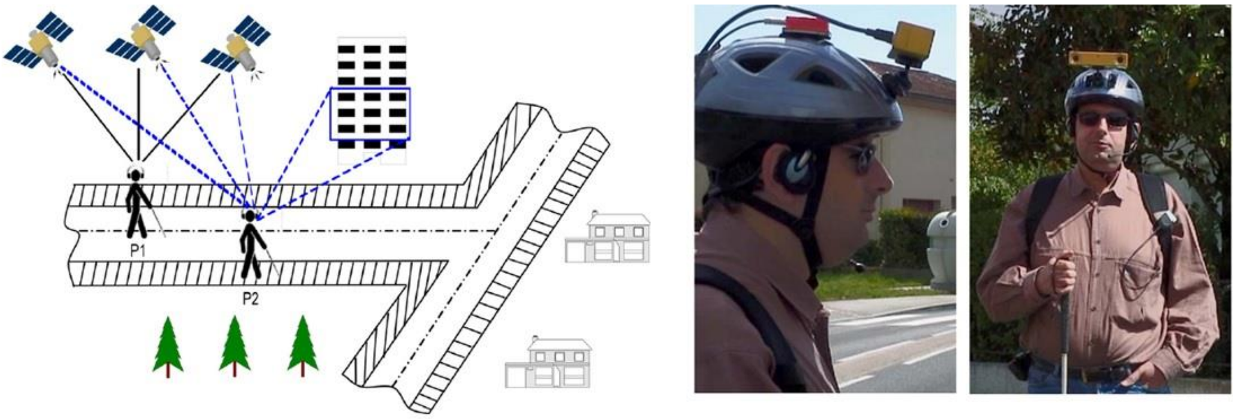
Figure 2.14: System for Wearable Audio Navigation (SWAN). An audio only wearable interface designed to aid navigation tasks in a complex environment for the visually impaired [15]

Geo-Aware Broadcasting for In-Vehicle Entertainment and Localizability (GABRIEL)

The GABRIEL system [85] developed by Villegas and Cohen provides simultaneous spatialised audio feedback to driver of a tourist vehicle and its occupants. The use of bonephones or ‘nearphones’ was proposed as method of audio delivery to the driver in order to provide navigation information. Passengers receive site specific information via headphone. In both cases, information delivery is spatialised. For the driver this helps with precise navigation, while passengers can be provided with an immersive narrative of a point of interest.

Navigation Assisted by Artificial Vision and GNSS (NAVIG)

NAVIG [86] (Figure 2.15) is a system that proposes the use of multiple information streams in order to help guide visually impaired users. Input from satellite data, on-board orientation and acceleration devices and image recognition sensors aims to improve positional precision of its user. Object identification via advanced image recognition algorithms help users understand the environment that surrounds them. The system adopts a macro – micro approach, with the GNSS system dealing with large scale navigational tasks to help the user reach a particular location from where the image recognition side of the system takes over to provide more detailed guidance information. Information regarding user trajectory and the user's position within it is proposed to be delivered via spatialised auditory feedback. NAVIG is designed to provide its user with increased autonomy by collating data from multiple information streams and presenting them in a manner which is easily understood by its user.

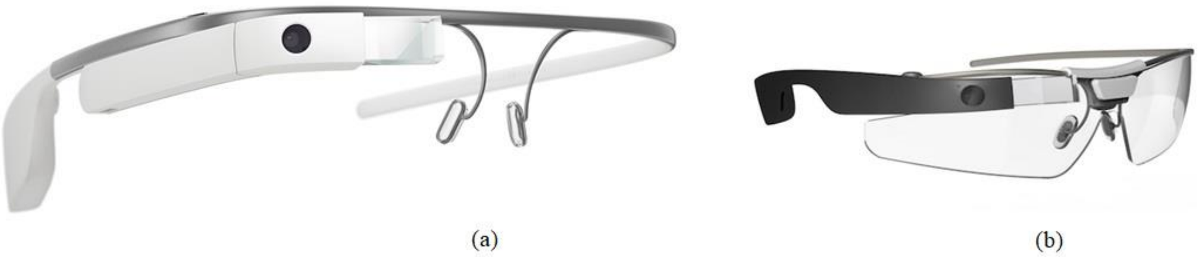


(a) (b)

Figure 2.15: NAVIG – Navigation Assisted by Artificial Vision and GNSS [86]. Figure (a) Block diagram showing how the system provides navigation information by tracking the user via a GNSS system and Figure (b) shows a prototype of the NAVIG system being worn by a visually impaired user.

Google Glass

The Google Glass [72] [71] (Figure 2.16) is an optical see-though, head mounted display device. Users are able to use voice commands and touch based input to interact with the device. The device is also equipped with a bone conduction transducer on side through which the user is able to receive auditory alerts. The bone conduction transducer can also be used as a headset for receiving calls. In its latest incarnation, the Google Glass has been re-born as an enterprise device helping workers in factories assemble equipment, install cabling in aircraft etc. [87].



(a) (b)

Figure 2.16: Google Glass [72]

Microsoft HoloLens

The Microsoft HoloLens [88] (Figure 2.17) is widely regarded as the first fully self-contained, mixed reality interface that allows its users to place and interact with virtual objects in the real-world. The device consists of an array of sensors and cameras that map the environment around the user. The HoloLens is also equipped with a speaker array on either side that sits just above the user's ears. It is capable of delivering fully spatialised auditory experience that is semi-customisable. Interactions with the device are based on fixed set of hand gestures that the HoloLens is programmed to recognise.

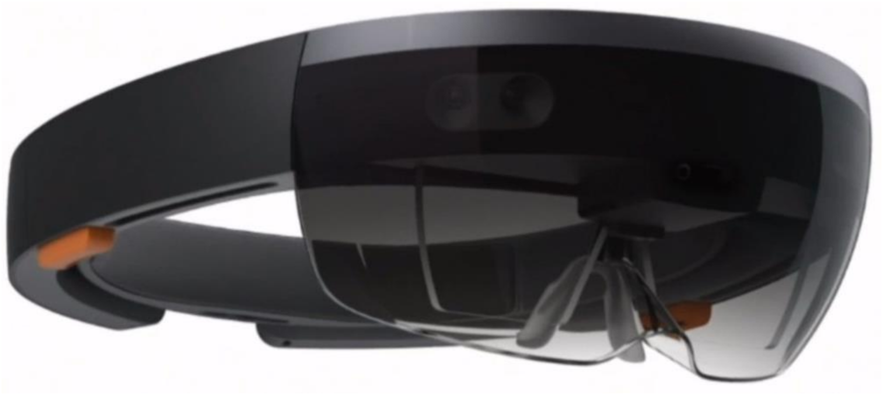


Figure 2.17: Microsoft HoloLens – Self-contained 'mixed reality' interface [88]

2.4.18 Hearables

In the recent past, audio augmented reality devices commonly referred to as 'hearables' have made an appearance in the consumer market. The leaders amongst this new wave of AR devices are the Here One [10] active listening system (Figure 2.18b) and the Bragi Dash Pro

[89] (Figure 2.18a). Both these devices provide the user with audio augmented perspective of the world. They allow the user complete control of the listening environment and are equipped with the latest bio-sensing features. These devices provided the added freedom of wireless connections to most current mobile phones.

Figure 2.18: Hearables (a) Bragi Dash Pro [89] (b) Doppler Labs Here One [90]

In this section we have provided a brief description of a number of wearable interfaces with a wide variety of features. The aim of this section is to provide the reader with an overview of the current state of the technology and how it has evolved over the years. Studying the strengths and weaknesses of these interfaces has shaped the direction this thesis has taken. Primarily, we have identified that interfaces which allow for bi-modal representation of data via the auditory and visual channels make for better information presentation devices. The ability to choose which faculty to engage based on the incoming information makes for a flexible interface which is then able to adapt itself based on the input it receives and the needs of its user at any given moment. In the following section we will present a short summary of the wearable interfaces we have covered in a chronological order and classify them according to their display types i.e. auditory, visual or a combination of the two. This section is designed to provide a quick overview of the interfaces we have reviewed and present an overall picture how the technology has evolved over time.

2.5 The Evolution of Wearable Technology

Wearable technology has been around for almost fifty years in one form or another. From Ivan Sutherland's tethered head mounted display capable of displaying only wireframes [77] to the Microsoft HoloLens [88], wearables have come a long way. In this section we list out some wearable interfaces that have been developed over the years and the modalities which they engage. This gives us good idea of how wearable technology has progressed and allows us to develop a system taxonomy based on these developments. Arguably the second decade of the twenty-first century has seen the most advances made in wearable technology with several commercial systems being launched. To incorporate this spurt in wearable interfaces we have distributed them – research and commercial – over three time lines to highlight their evolution over time. They have been divided into visual, auditory and hybrid or bi-modal interfaces represented by black, blue and green flags respectively.

The first time line (Figure 2.19) depicts the evolution of interfaces in the 1990s. This follows what appears to be a relatively quiet period in the development of wearables after the heads-up display was developed by Sutherland [77] and the demonstration of the Videoplace – which can be considered the first AR collaborative environment – at the Milwaukee Arts Centre in the beginning of October 1975 [91]. We have chosen to refer to this period as a quiet period for development since there does not appear to be a significant amount of literature available on the subject, although we are aware of several patents [92] [93] [94] [95] [96] [97] [98] filed for various forms of visual AR displays mainly in the defence sector. We see that the 90s (Figure 2.19) consisted mainly of visually augmented interfaces with some auditory and bi-modal interfaces being developed towards the middle and end of the decade. Important developments of this decade could be the AR system developed by Caudell and Mizell [99] to aid engineers at Boeing to assess and carry out aircraft maintenance without having to carry aircraft manuals around, the KARMA system developed by Feiner, Macintyre [100] and the wearable conferencing space developed by Billinghurst, Bowskill [101]; the first true hybrid wearable AR system to our knowledge.

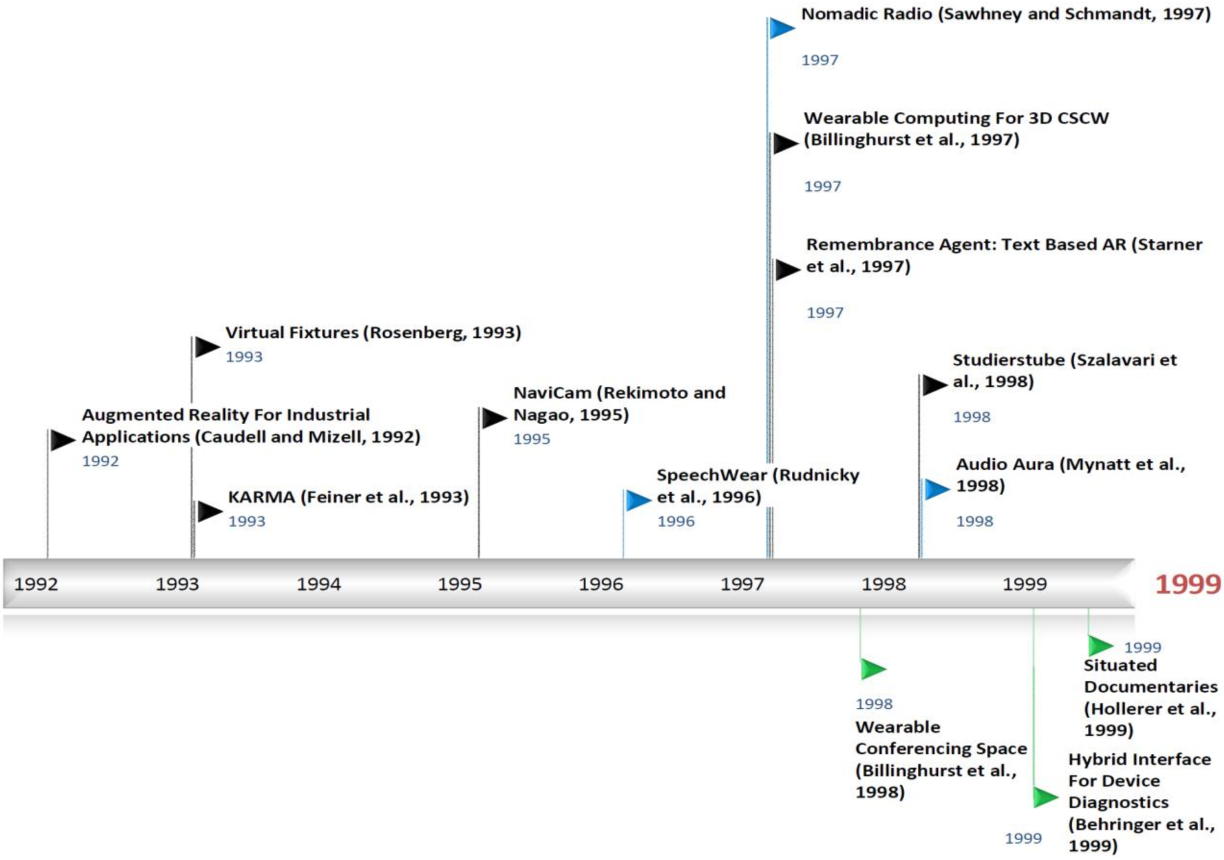

Nomadic Radio (Sawhney and Schmandt, 1997)
1997
Wearable Computing For 3D CSCW (Billinghurst et al., 1997)
1997
Remembrance Agent: Text Based AR (Starner et al., 1997)
1997
Virtual Fixtures (Rosenberg, 1993)
1993
Studierstube (Szalavari et al., 1998)
1998
NaviCam (Rekimoto and Nagao, 1995)
1995
Augmented Reality For Industrial Applications (Caudell and Mizell, 1992)
1992
Audio Aura (Mynatt et al., 1998)
1998
SpeechWear (Rudnicky et al., 1996)
1996
KARMA (Feiner et al., 1993)
1993
1992
1993
1994
1995
1996
1997
1998
1999
1999
1999
Situated Documentaries (Hollerer et al., 1999)
1998
Wearable Conferencing Space (Billinghurst et al., 1998)
Hybrid Interface For Device Diagnostics (Behringer et al., 1999)
1999

Figure 2.19: History of Wearable Interfaces – The 90s. Black flags represent Visual Interfaces, Blue represent Auditory Interfaces and Green represent Bi-Modal/Hybrid Interfaces

Moving on the next decade (Figure 2.20), we see that there is a significant increase in the number of audio augmented reality interfaces that were developed. This follows the realisation that the use of only visual displays for wearable interfaces severely limits information delivery due to limited screen size [13]. The greater cognitive load impressed upon the sensory systems [102] also appears to have spurred this change in direction. On the other hand, interfaces such as SWAN developed by Walker and Lindsay [14] are designed to cater specifically to the visually impaired user. Utilising bone conduction headsets to convey spatialised audio beacons for navigation, it allows the user to remain aware of their surroundings by leaving the ears open. Other augmented reality audio headsets use complex systems to convey natural spatiality of an environment overlaid with synthetic audio cues [83], [84].

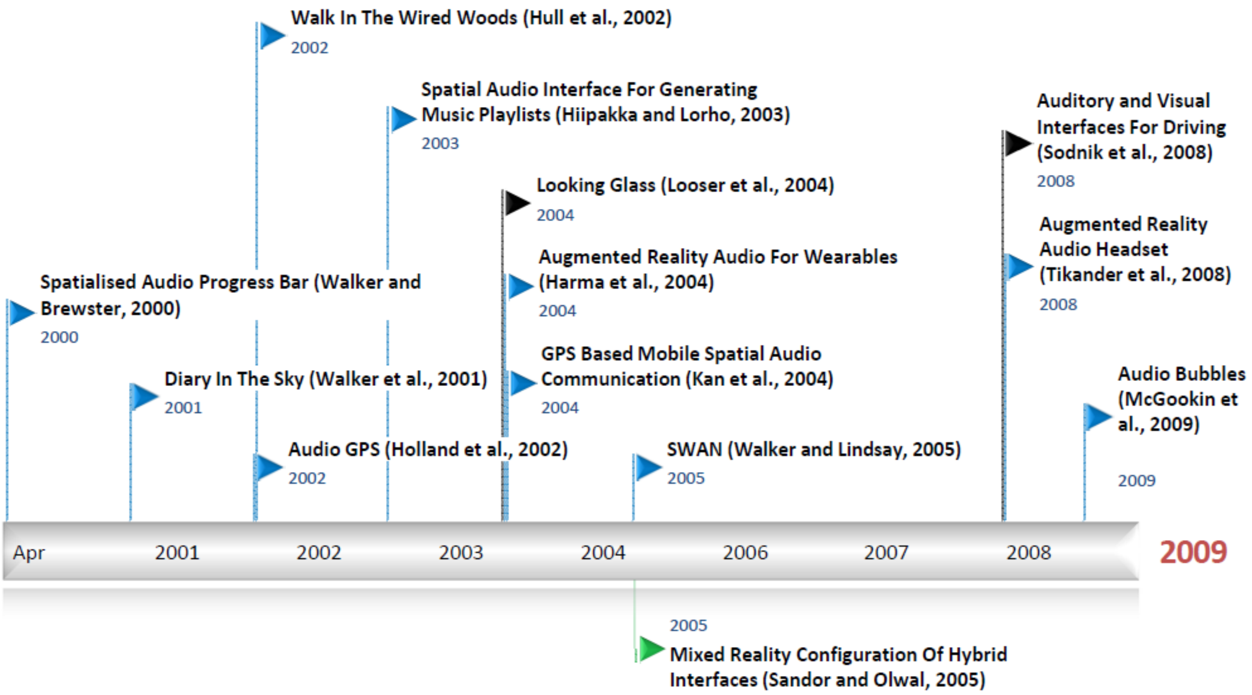

Walk In The Wired Woods (Hull et al., 2002)
2002
Spatial Audio Interface For Generating Music Playlists (Hiipakka and Lorho, 2003)
2003
Auditory and Visual Interfaces For Driving (Sodnik et al., 2008)
2008
Looking Glass (Looser et al., 2004)
2004
Augmented Reality Audio For Wearables (Harma et al., 2004)
2004
Augmented Reality Audio Headset (Tikander et al., 2008)
2008
Spatialised Audio Progress Bar (Walker and Brewster, 2000)
2000
GPS Based Mobile Spatial Audio Communication (Kan et al., 2004)
2004
Diary In The Sky (Walker et al., 2001)
2001
Audio Bubbles (McGookin et al., 2009)
2009
Audio GPS (Holland et al., 2002)
2002
SWAN (Walker and Lindsay, 2005)
2005
Apr
2001
2002
2003
2004
2006
2007
2008
2009
2005
Mixed Reality Configuration Of Hybrid Interfaces (Sandor and Olwal, 2005)

Figure 2.20: History of Wearable Interfaces – The first decade of the new millennium

The current decade (Figure 2.21) has seen a marked increase in wearable interfaces becoming available to the consumer. A significant number of these interfaces now incorporate sound and vision. With the advent of devices like the Google Glass [71], [72] and Microsoft HoloLens [88], multi-modal wearable interface now set to take a step closer to becoming an integral part of our lives – integrated into our clothing, accessories and maybe even our bodies to provide us with seamless real and virtual world interactions.

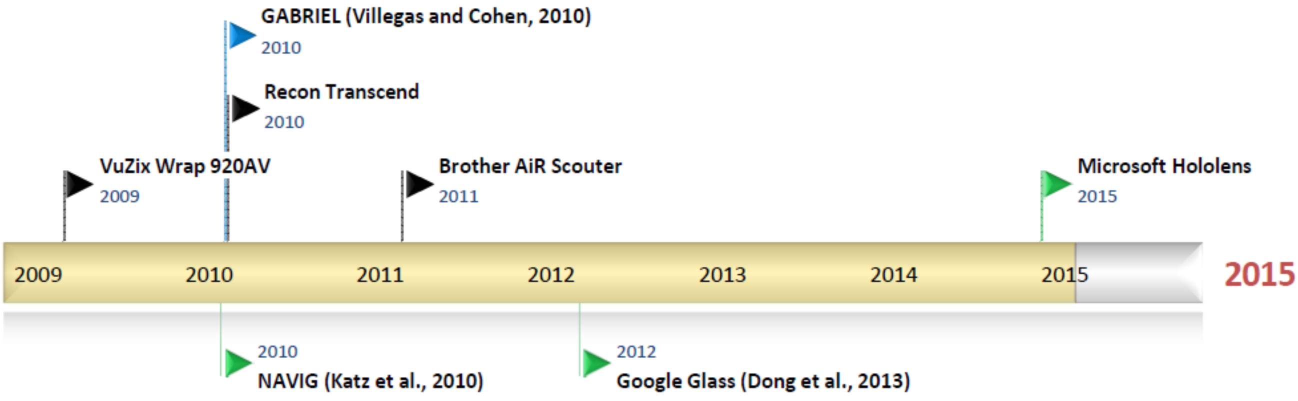


Figure 2.21: History of Wearable Interfaces – The current decade (2010 – present)

2.6 Conclusion

This chapter has presented a review of pre-existing research in the fields relevant to this thesis, namely auditory perception, visual perception and wearable technology. We began this review with an introduction to the perception of sound. With regards to this topic, we have covered aspects of auditory perception; beginning from the earliest theories on the perception of sound to the some of the more recent work being undertaken in psychoacoustics. Following this, we briefly looked at visual perception. We consider this to be an important aspect of our research, since an overview of wearable interfaces has shown us that there seems to be an overwhelming disposition amongst designers of these interfaces to engage the visual faculty. The links between visual and auditory perception and their interdependence has also been addressed.

Typically our perception of the environment consists of cues we receive from a number of our faculties. Chief among these are vision and audition. It is only natural that we explore how these two faculties depend on each other to help us perceive the environment around us. Looking at the research that has been done in this area also provides us with vital design cues which will help construct more holistic design principles for future designers of wearable interfaces. Finally, we have provided a historical overview of the wearable space. This overview has helped us identify existing design trends for wearable interfaces and the drawback associated with them. The dominant themes resulting from this overview are:

- Wearable interfaces tend to primarily engage the visual faculty.
- Those that use auditory cues, relegate their use to providing simple updates and notifications.
- Some interfaces that do use spatialised audio suffer due to improper implementation of the medium.
- Audio only wearable interfaces have been proposed as a ‘solution’ to traditional visual based systems. A recurring theme of ‘the need for eyes free interaction’ and small screen size are presented as a justifications.
- Bi-modal and audio only interfaces, for the most part, tend to use headphones or earphones as delivery platforms for auditory cues.

In addition to this, the review has also made us aware of some exciting but as yet to be validated work. This primarily revolves around the use of the bone conduction headset. Up until now the bone conduction headset has been primarily used as a means to deliver spatialised auditory cues to the visually impaired to aid navigation [14] [15]. There seems to be limited research on the use of the bone conduction headset with an existing wearable visual interface. The handful of studies we have come across [103] [85], recommend the use of a bone conduction headset in some environments. In addition to this, studies involving spatialisation over the bone conduction headset have mostly made use of individualised HRTFs [104] [16]. Research into the use of the headset as a medium for auditory delivery in an AR environment is relatively uncommon. The lack of empirical user studies evaluating the headset in multiple environments as part of a bi-modal, wearable interface has prompted us to look at this area of research. It has helped us formulate an outlook and revisit some existing work with bone conduction headsets. The three preliminary queries resulting from the existing body of research we have looked at are:

- How accurate is localisation in the horizontal plane for binaurally spatialised sources presented over a bone conduction headset?
- How accurate is localisation in the vertical plane for binaurally spatialised sources presented over a bone conduction headset?
- How does a binaurally spatialised source presented over a bone conduction headset affect our visual perception of the real-world?

These questions are dealt with in Chapter 3 as three separate studies that outline the performance of the bone conduction headset. An important aspect of these studies has been our approach to designing and implementing them. In order to maintain high ecological validity, the design and implementation processes make use of freely and/or low cost off the shelf components in order to replicate current development trends in the industry. Chapter 4 maintains the same outlook towards the design and development of the wearable interface prototype as well. The two studies in Chapter 4 provide detailed results regarding the implementation of our prototype in two real-world environments.

Chapter III

3. AUDITORY PERCEPTION OVER A BONE CONDUCTION HEADSET

The previous chapter has covered in some detail the processes involved in auditory perception. These processes can now be replicated to provide a synthetic auditory environment that provides the user with an experience very similar to natural hearing. Costs, monetary and computational, associated with these processes have drastically declined in recent years allowing a wider set of researchers and developers access to tools which allow for the creation of realistic auditory experiences. Some of the wearable interfaces mentioned in the previous chapter make use of these low-cost methods to implement spatial auditory displays as feedback mechanisms. One of our primary goals has been to demonstrate the use of low-cost, off-the-shelf components to design and develop a wearable device that can provide the user with feedback in a contextually relevant manner. To that end we have also used a development pipeline and tools that are currently in use to be able to lend greater ecological validity to our results. This is the first part of a two-part evaluation and implementation strategy adopted for this project. In this chapter we describe three psychoacoustic experiments that we have carried out to determine the feasibility of using a bone conduction headset as part of bi-modal, wearable interface. We have evaluated the use of the bone conduction headset for critical functional aspects such as localisation accuracy and externalisation. This chapter explores three basic but significant research questions:

- How good is localisation of a binaurally spatialised stimulus in the horizontal plane for material reproduced over a bone conduction headset?
- How good is localisation and elevation perception of a binaurally spatialised stimulus in the vertical plane for material reproduced over a bone conduction headset?
- How do visual cues affect the perception of a binaurally spatialised stimulus over a bone conduction headset?

Answers to the questions above are of utmost importance because they lay the foundation for the studies that follow. Examining factors like localisation accuracy and externalisation has aided the design of auditory cues that have been used in the studies described in Chapter 4. Results from the experiments in this chapter have been directly applied to auditory and visual cue designs for our prototype. Answers to these basic questions have helped lay the foundation to address the wider issue of information presentation on wearable interfaces. Consequently, we have been able to formulate principles that are in accordance with the desired design outcomes stated in Chapter 1.

At the outset, it is important to mention that comparisons of results obtained as part of these studies have been made with the few studies that are available on binaural spatialisation over a bone conduction headset. Comparisons to results obtained for similar studies carried out with headphones have also been made to illustrate the efficacy of the bone conduction headset.

Study I

BINAURAL SPATIALISATION OVER A BONE CONDUCTION HEADSET: MINIMUM DISCERNIBLE ANGULAR DIFFERENCE

Our first experiment was set to evaluate binaural spatialisation over a bone conduction headset in the horizontal plane. The study was loosely based on an experiment carried out by Mills [105] which evaluated the *Minimum Audible Angle* between two successively presented test tones at varying angular separations. This study was presented as a full paper at the *140th Convention of the Audio Engineering Society* in Paris, France held between 4th June and 7th June 2016.

3.1 Introduction

As we have already seen in Chapter 2, research exploring the mechanism of auditory perception has a long and varied history. Over a period of close to a century and a half, various aspects of auditory perception have been discovered, dissected and re-interpreted. From the earliest theories put forth by Lord Rayleigh [31] [106] [32], to the cutting edge research currently being undertaken, there exists a wealth of knowledge regarding how humans perceive sound. However, despite all our accumulated understanding of the subject of auditory perception, relatively little is known about the perception of binaurally spatialised sound over a bone conduction headset. It can even be argued that outside of audiology research there remains a comparatively large gap in our understanding of how auditory perception via bone conduction works. The lack of understanding is even starker when we look at binaurally processed sounds being reproduced over a bone conduction headset. With the exception of a few studies [14] [104] [107] [16], there remains a dearth of information on the subject of binaural spatialisation over a bone conduction headset.

As we have already seen in the literature review, spatial audio delivery in a mobile setting is typically implemented using headphones. This approach tends to isolate the user from the surrounding acoustic environment [12] [13] [80] [81]. Such isolation can prove dangerous and even fatal in a real-life context as shown in reports of people involved in accidents caused by headphone use [108] [109]. A bone conduction headset is an ideal alternative to headphones since it leaves the ears open to perceive the surrounding acoustic environment. This ‘natural’ interaction with the acoustic environment afforded by the bone conduction headset [110] makes it a safer alternative to headphones.

Over the last two decades or so, there have been great strides made in wearable computing, including the commercial release of devices such as Google Glass [72] [71] and Recon Jet [111]. However, an ever shrinking screen size in wearable devices has resulted in problems of information presentation [112] [13]. Spatial auditory displays have been suggested as one way to manage information display in such devices [12]. It was hypothesised that such a display will aid with information management and relieve some of the cognitive stress associated with such tasks. Since a large percentage of the wearable interfaces that employ auditory feedback use headphones, it would be prudent to look at alternate form factors for the delivery of auditory material to make these devices safer and more effective. This is where the bone conduction headset stands out.

Our study begins by classifying the limits of binaurally spatialised sound reproduced over a bone conduction headset. This involves codifying the tolerances of critical factors such as localisation accuracy, effect of visual stimuli on the auditory stimuli and the perception of direction and distance. This is important because it will provide designers of future wearable interfaces and their applications a point of references based on which informed design decisions can be made.

3.2 Background

Virtual auditory displays using non-individualised HRTFs have been the subject of research for close to three decades [52] [55] [56] [12] [13]. These displays process the incoming sound so that it appears to derive from specific external spatial locations [52] [55] [56] [16]. Traditionally, these displays have made use of headphones to deliver the processed audio signal. However the use of headphones negatively impacts the perception of the ambient acoustic environment [14] which is potentially hazardous. Not being able to hear ambient acoustic cues that convey the state of the environment to the user and potentially alert them to danger could put the user's safety at risk. Some researchers have proposed the use of 'hear-through' headphones/earphones as a solution to this problem (Figure 3.1) [83] [84].

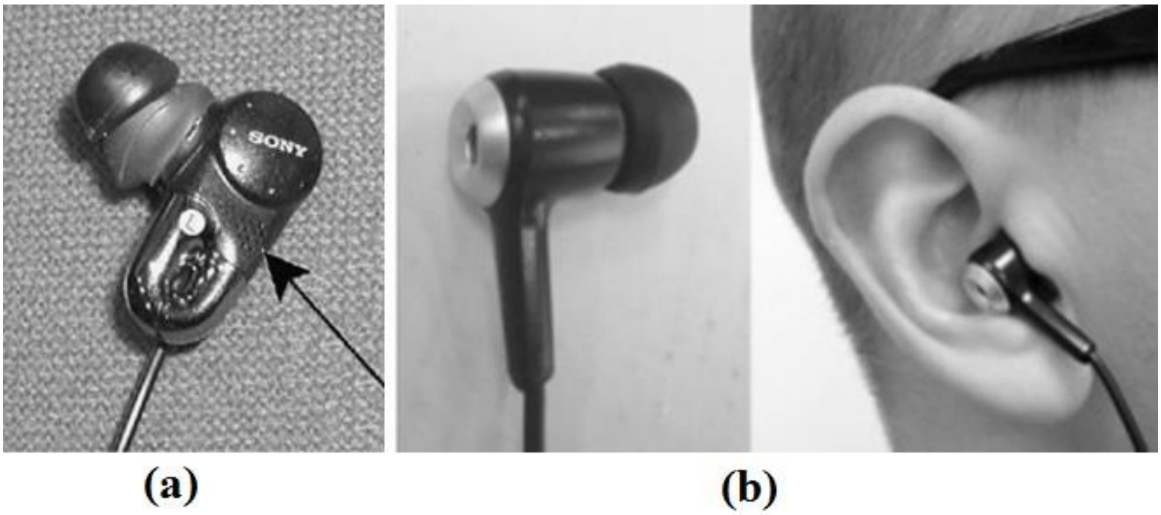


Figure 3.1: Earphone based Augmented Reality Audio (ARA). Figures (a) [83] and (b) [84] show different ways in which a microphone has been incorporated into earphones to capture the ambient acoustic environment.

For such a solution to work, extensive equalisation and calibration must be carried out to provide a ‘near natural’ sounding environment. Tikander et al. [84] report several operational issues that could render such a set-up ineffective, such as distorted audio output resulting from loud sounds overloading the microphone and/or amplifier circuitry. While such a problem can be dealt with to a certain extent using fast acting compressors and gates, sounds of eating, wires brushing clothing etc. can cause users to abandon the use of a such a device due to the ‘irritability’ such noises may introduce. In our opinion these issues can be avoided with the use of a bone conduction headset.

Up until now, binaurally spatialised information presentation over a bone conduction headset has primarily focused on aiding the visually challenged with navigation tasks. This works relatively well in virtual [14] [103] and real-world environments (Figure 3.2) [15].

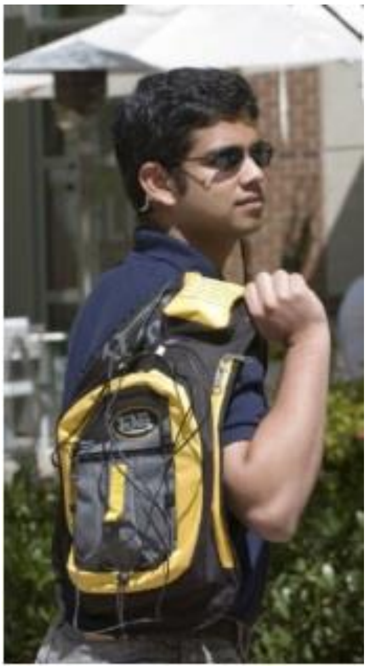


Figure 3.2: System for Wearable Audio Navigation (SWAN). An audio only wearable interface designed to aid navigation tasks in a complex environment for the visually impaired [15]

Villegas and Cohen proposed the use of a bone conduction headset as means of providing spatialised auditory beacon navigation [85] for sighted users. Other studies have briefly explored the possibility of binaurally spatialised sound presented over a bone conduction headset to some degrees of success [104] [107] [103]. MacDonald et al. [104] suggested that binaural reproduction over a bone conduction headset could provide spatial resolution which is comparable to, or better than, headphones when treated with individualised HRTFs.

Our study will explore the use of freely available ‘binaural plugins’ to process the audio signal being sent to the bone conduction headset. Results from some of the exploratory studies mentioned earlier [104] [107] [103], and a significantly more detailed body of work presented by Walker, Lindsay, Stanley and Wilson [14] [113] [114] [16] led us to believe that the bone conduction headset has the potential to be used as part of wearable, bi-modal interface capable of providing spatialised auditory feedback. However, the lack of empirical data on binaural spatialisation over a bone conduction headset makes it crucial to run experiments to gather qualitative and quantitative data regarding their operational parameters.

This study is the first in a series of five that will explore the subject of binaural spatialisation over a bone conduction headset and its real-world applications. In doing so we hope to make significant contributions to the understanding of how spatialisation over a bone conduction headset works and present a design template for future applications of the device. In the following section we outline our study evaluating the extent of the perceived angular discrimination between two successively spatialised sound sources.

3.3 Method

3.3.1 Apparatus

We adopted the following hardware and software components for our experiment:

- Unity3D: A freely available game engine developed by Unity Technologies [115]
- 3Dception: A ‘binaural engine’ plugin for Unity developed by Two Big Ears [116]
- AfterShokz Sportz3: A low cost bone conduction headset [117]

The Unity3D engine was chosen for the flexibility and ease of integration with a number of third party plugins it offered. The binaural engine was chosen after extensive testing and comparisons with other binaural plugins. While the inner workings of the 3Dception spatialiser are not available, we believe the results we have obtained in our studies that have used this plugin can be extended to similar binaural plugins available in the market. Subjective opinions from a range of users, online forums and reviews [118] were also taken into consideration. The AfterShokz bone conduction headset was chosen due to its relatively low cost, and that it provided better sound quality and output level in comparison to the competing Goldendance MGD02 headset. Sound quality ratings were obtained from participants of a pilot study run prior to this experiment.

This combination of software and hardware and its implementation is a step away from traditional practices implemented in psychoacoustic studies which use individualised HRTFs [104], HRTF databases [12] [13] [119] or HRTFs of an empirically verified ‘good localiser’ [55] [56] [120]. The experiment was developed in the Unity3D environment and used the 3Dception binaural plugin to spatialise two sound sources placed equidistant from the main camera position. The main camera also served as the ‘global listener’ i.e. the sound heard at this position was the perspective of the user (Figure 3.3).

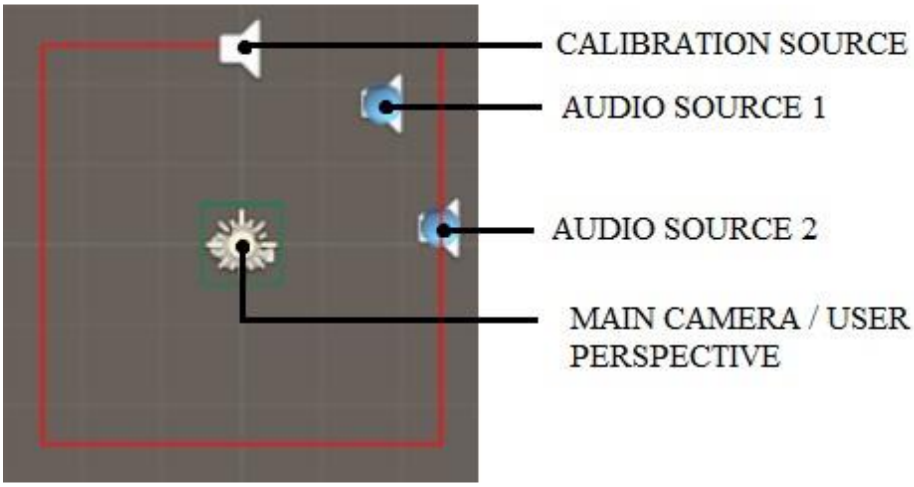


Figure 3.3: Experimental setup in Unity3D [121]

The sources were created using a feature provided in the binaural engine with the ‘global listener’ slaved to the main camera to provide a ‘first person perspective’ of the audio events. Sound was delivered to the participants via the AfterShokz Sportz3 bone conduction headset connected to a Zoom UAC-2 audio interface (Figure 3.4). The experiment was run on a Dell Inspiron laptop (Operating System: Windows 8, Processor: 2.2 GHz Intel Core i7). The study was carried out in a sound proof booth that conformed to the HTML 2045, ISO 8253 and ISO BS EN6189 standards. It had an Rw (Weighted Sound Reduction Index) between 35bB and 60dB (RT60 < 0.3s).

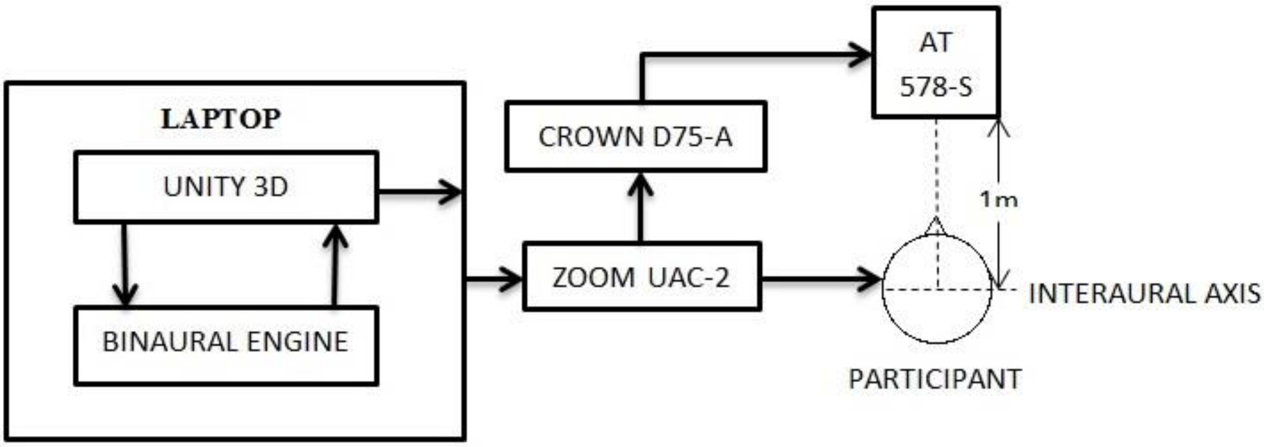


Figure 3.4: Block diagram of experimental setup [121]

3.3.2 Stimuli and Calibration

Two 1000ms pink noise bursts with a 25ms onset and offset time, separated by one second were used for the experiment. Pink noise was used since it has been demonstrated that broad band stimuli were easier to localise than other spectrally deficient stimuli [14] [122] [42] [123]. Prior to beginning the experiment, participants were asked to calibrate the headset. This was achieved by playing a pink noise signal, the calibration source, over a loudspeaker placed 1m away from the participant's head at approximately 70dBA (Figure 3.4). Participants were instructed to position their heads such that they were on the normal to the speaker's cone and adjust the level on the bone conduction headset until they felt that it matched that of the speaker. The loudspeaker was turned off once the calibration was completed. The duration and level of the stimuli were chosen to reproduce those used by previous researchers in psychoacoustic studies [14] [120] [122]. The signal was routed from the audio interface to a Crown D-75A amplifier and reproduced by a single PhonicEar AT578-S loudspeaker. The stimuli were presented at a constant level and no artificial reverberation or any other form of processing carried out on them.

3.3.3 Participants

A total of 26 untrained participants (20 male, 6 female) aged between 19 and 41 (Mean: 25.03 years) volunteered to take part in the study. All, except one of the participants reported normal hearing. This participant's results were excluded from the study. No audiometric testing was carried out on any of the participants to verify their self-reported normal hearing function. All participants were compensated with a NZD $10 voucher for participating in the study.

3.3.4 Experimental Procedure

Subjects were first shown how to put on the bone conduction headset (Figure 3.5). For the experiment participants were presented with ten different angular separations between the two stimuli. These ranged from 0° (no separation) to 45° in 5° increments. This range (0° – 45°) was chosen based on a pilot study[3] carried out with fifteen participants that indicated an average minimum discernible difference of 20° between two binaurally spatialised stimuli presented over a bone conduction headset. Two participants in the pilot study reported spatial separation between the stimuli for an angular difference of 5°. Keeping this in mind, we chose to use this value as the minimum presented angular separation for the main experiment. No participant in the pilot study demonstrated the ability to discriminate between stimuli separated by less than 5°.

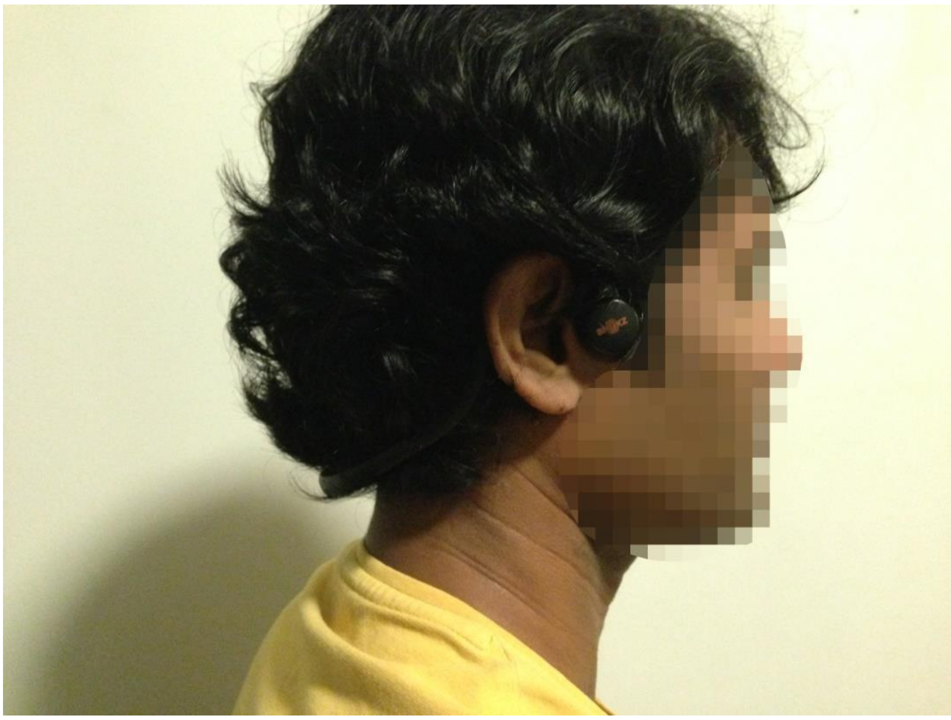


Figure 3.5: Participant wearing the Bone Conduction Headset (BCH) [121]

---

[3] Refer to Chapter 6 – Appendix I (Pilot Study)

Each of the angular separations was presented ten times, resulting in a hundred trials per quadrant (total trials = 400). The angular separations were presented in a random order. The experiment lasted approximately ninety minutes and participants were given a ten minute break once half the trials were completed. Before starting the main block of trials, participants completed a single block of trials not exceeding five minutes. This block exposed participants to all the angular separations they would encounter in the main experimental block. This was done to familiarise the participants with the experiment and how they were expected to provide localisation feedback. Participants were asked to use a response chart (Figure 3.6) given to them to indicate localisation responses. There was no compulsion to look at the chart if participants felt they had understood the signed angle protocol before calling out the response. Localisation responses were called out using the signed angle protocol explained to the participants prior to beginning the experiment[4]. This method of judgement estimation has been validated by previous studies [55] [56] [122] [124] and has been shown to be relatively easy to learn [125].

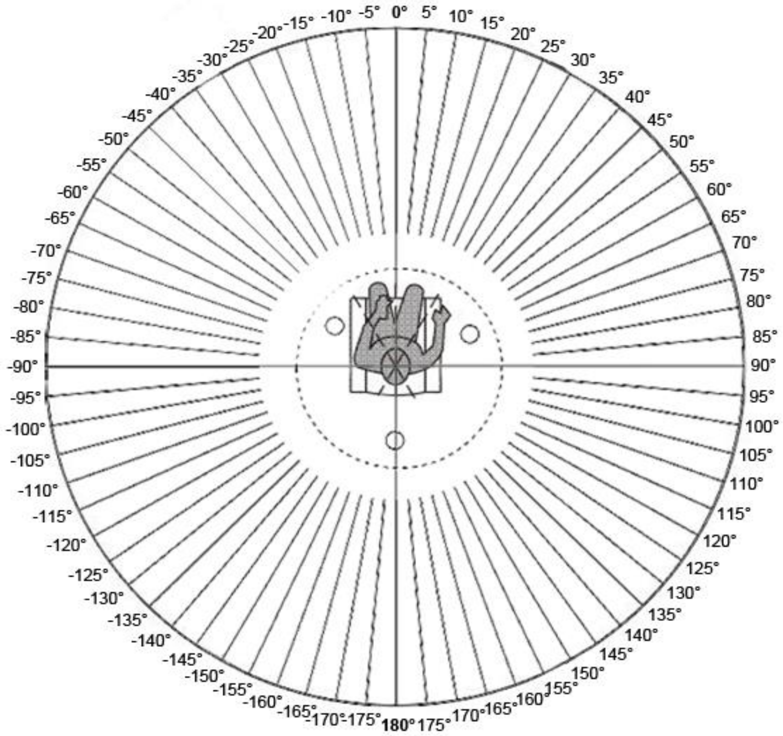


Figure 3.6: Response chart used by participants to localise the two successively spatialised sources of sound [121]

---

[4] Labelling conventions in psychoacoustic research dictates that the point directly in front of the participant's head be designated as 0° with the rest of the area around the head divided in two equal, opposite halves of 180° each signed + (positive) for the right hand side and – (negative) for the left hand side

Participants were asked to look forward throughout the experiment, but this wasn’t strictly enforced. No chin brace was used to keep the head in a fixed position either. The main block of four hundred trials was divided for presentation into eight 45° sectors (0° – 45°, 45° – 90° etc.). Each angular separation between the sound sources (0°, 5°, 10°, 15°…45°) was presented five times per sector in a random order (50 trials per sector). A complete set of presentations, fifty trials, was made for every sector before moving onto the next. A random switching order between sectors was employed to mitigate any learning effects that could arise from an ordered presentation. We chose to present the stimuli sector-wise, rather than for every quadrant, since this allowed us to introduce a larger level of randomness in the presentations and the ability to analyse the data in greater detail.

3.4 Results

Analysis was carried out on the data obtained from all 26 participants who took part in the study (N = 26). We used two methods to divide and evaluate the data: a front-back distribution and a front-back-left-right distribution. In both cases, the non-normal distribution of data required the use of the Wilcoxon’s Signed Rank Test to determine significance. Each angular separation, starting from 0° up to 45°, was presented 5 times per sector for a total of 50 trials per sector. Localisation judgements were recorded and tabulated. Recognition rates for the presented angular differences were then obtained by dividing the number of times an angular presentation was identified as being different by the total number of times it was presented. Figures 3.8 and 3.10 show the mean values for recognition rates obtained for the front-back and front-back-left-right configuration respectively. Paired testing was carried out once these values were tabulated for each participant across both configurations to obtain the minimum discernible angular difference between successively spatialised sound sources.

3.4.1 Front-Back

The front-back division of the auditory space around the head is a traditional approach seen throughout research concerned with auditory perception. The line joining -90° and +90° represents the interaural axis and passes through the ear. The interaural axis is considered the ‘boundary’ which divides the auditory space around the head into the front and back sections (Figure 3.7).

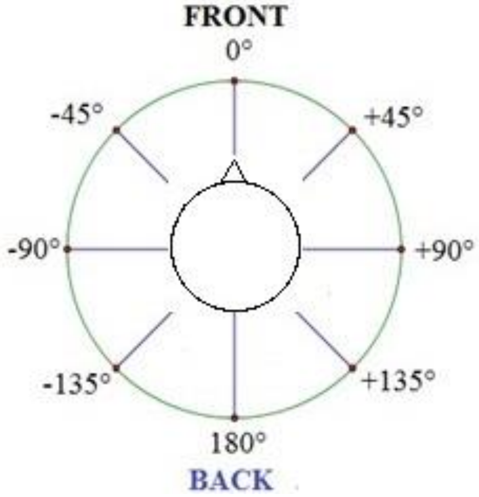


Figure 3.7: A traditional front-back configuration widely seen in psychoacoustic research [121]

Our first approach to analysing the data uses this methodology. The data was analysed in two directions i.e. in pairs starting from a 0° angular difference paired with a 5° angular difference up to a 45° angular difference and back down from a 45° angular difference to a 0° angular difference. This allowed us to check for any variations that may show up when pairs were compared using either ends of the angular differences as anchors. The results of the Wilcoxon's Signed Rank Test conducted for angular differences paired with 0° angular separation for the front indicated a significant difference ($z = -4.02$, $p<.001$) between source presentations with no angular separation and a 10° angular separation. All pairs from 10° onwards showed a significant difference. Similarly, when the test is run on angular separations paired with 45° downwards i.e. towards 0°, a significant difference ($z = -3.54$, $p<.001$) is seen between an angular separation of 35° and 45°. This suggests that the minimum discernible angle between two successively presented, binaurally spatialised sources using a bone conduction headset is 10° in the front.

The same kind of pairing was employed to test angular differences for the rear. The results of the test revealed similar pattern for the 0° and 45° angular separations paired with increasing and decreasing angular separation values (Back 0° vs. Back 10° ($z = -3.51$, $p<.001$) Back 45° vs. Back 35° ($z = -2.86$, $p<.005$)). Results for the rear also show a minimum discernible angular separation of 10°. The results do not show any significant difference between the same angular separations for the front and back (Figure 3.8).

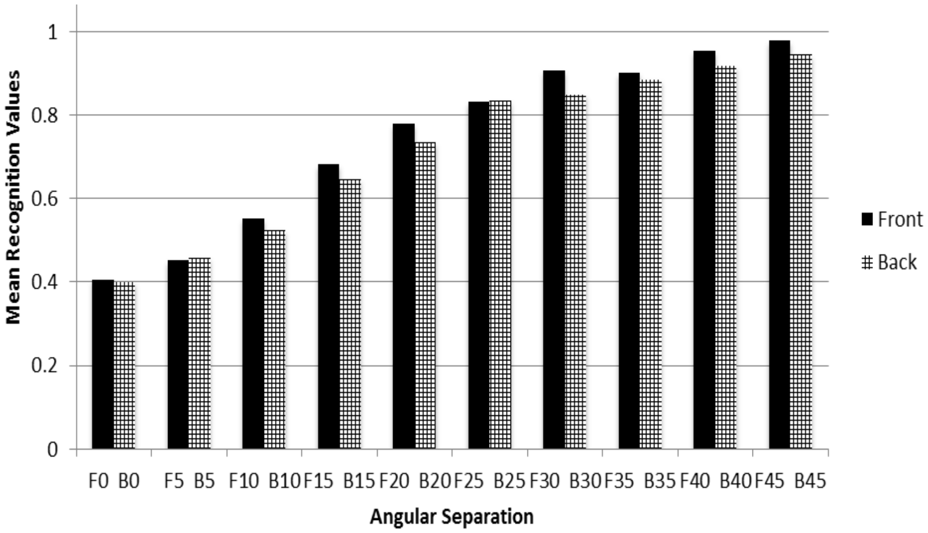


Figure 3.8: Mean recognition values obtained for the Front-Back configuration. The means for no angular difference (0°) indicate 'false positives' i.e. the perception of an angular difference where none exists. F-Front, B-Back [121]

3.4.2 Front-Back-Left-Right

In this scenario, we propose an alternate method to evaluate the data we have obtained. This approach enables the segregation and interpretation of the data in greater detail. The auditory space around the head has been divided into four quadrants resulting in the separation of the space into the front, back, left and right directions. A configuration similar to the one used by Wightman and Kistler [52] was employed, but we have extended the coverage of sides from ±60° to ±120° to ±45° to ±135° (Figure 3.9).

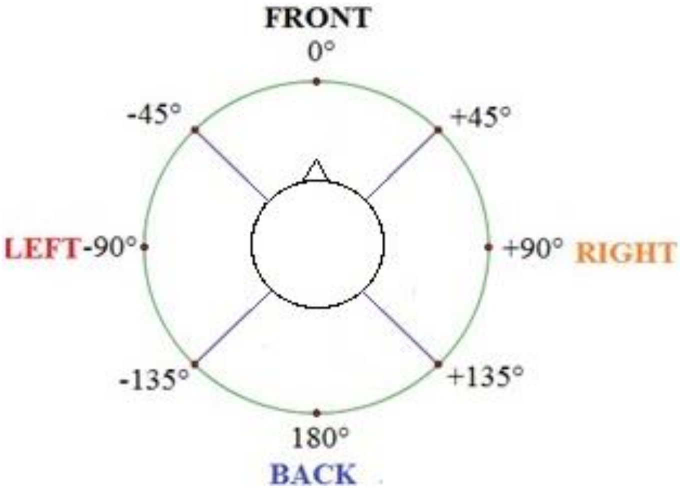


Figure 3.9: A modified approach to evaluating results obtained as part of the perceptual study [121]

In this configuration, a consistent significant difference between the pairs with 0° angular separation and 45° angular separation for the front is observed for an angular separation of 20° between the two sound sources ((0°: $z = -4.08$, $p<.001$), (45°: $z = -2.54$, $p<0.05$)). Similarly, for the back, consistent differences between the pairs are observed for angular differences of 20° and above between the two sound sources ((0°: $z = -4.31$, $p<.001$), (45°: $z = -2.40$, $p <0.05$)).

For angular separations paired with the 0° and 45° separations to the left a consistent significant difference is observed for an angular separation of 15° between the two sound sources ((0°: $z = -3.35$, $p<0.001$), (45°: $z = -2.83$, $p<0.005$)). Identical pairings made for the right quadrant show a significant difference for a 10° angular separation onwards i.e. the first significant difference is seen for the 0° - 10° ($z = -2.32$, $p<0.05$) pair for angular separations paired with 0° separation, and the first significant difference is seen for 45° - 35° ($z = -2.46$, $p<0.05$) pair for all separations paired with the 45° angular separation. While we expected that the front quadrant would demonstrate a greater power of resolution as previously seen in literature [42] [46], the results here appear to suggest that the front and rear quadrants have a similar power of resolution (≥ 20°). The left and right quadrants appear to have a lower threshold of angular discrimination of 15° and 10°.

Pair-wise comparisons made between the same angular differences presented in the front and to the left demonstrate significant differences starting from a 15° angular separation ($z = -3.05$, $p<0.05$). A similar comparison made between the front and right yields significant differences between 15° ($z = -3.19$, $p<0.01$) and 30° ($z = -3.22$, $p<0.01$). All angular separations above and below these thresholds show no significant difference. Comparisons made between the rear, and the left and right quadrants show similar results. Pair-wise comparisons between the rear and left show significant differences between a 15° ($z = -2.55$, $p<0.05$) separation going up to a 40° ($z = -2.54$, $p<0.05$) angular separation. The rear and right pairs differ significantly only between 15° separation ($z = -2.99$, $p<0.05$) and 30° separation ($z = -2.93$, $p<0.01$). No significant differences were observed for pair-wise comparisons between the same angular presentations between the front and back, and right and left quadrants. While these results do not come close to the values obtained by other researchers for the front or the sides [46] [105], it is quite clear from the means that even a 10° angular separation approaches a recognition rate of approximately 60% for the front-back-left-right (Figure 3.10). All angular separations from 25° onwards display a recognition rate in excess of 80% for both configurations (front and back), with 40° and 45° angular separations consistently approaching or exceeding 90% rate of recognition (Figures 3.8 and 3.10).

All participants reported externalization, except for two subjects. A method previously employed by Stanley [16] and Gardner [122] was used to make judgements on externalization. This involved participants indicating whether the sound sources were heard at the centre of their head, on the surface or completely outside at some distance from their head. Almost all participants reported hearing the sound sources on the surface of their heads or 'hovering' close to their neck and face. Only 2 of the 24 participants who externalized the sources localised them further away, from 1m – 5m. The level to which the sources were externalised is to be expected since no reverberation or any other form of signal processing was applied. This was an encouraging result as it suggests the possibility of providing a richer aural experience to the user of a bone conduction headset without compromising perception of the ambient acoustic environment in which the listener is located.

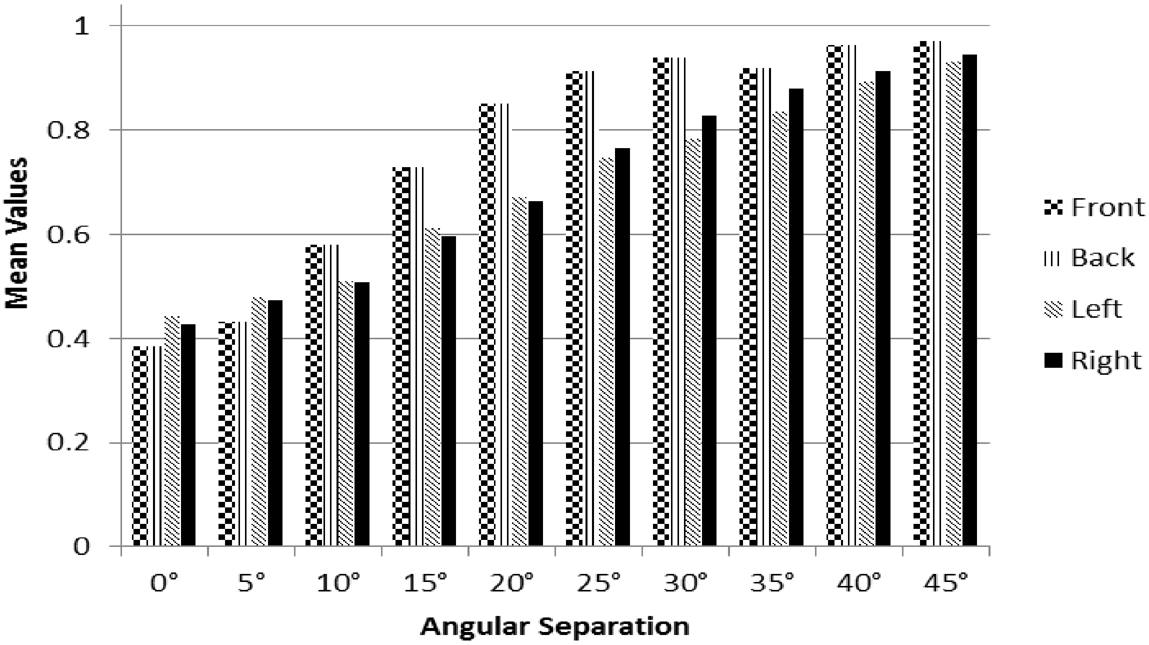


Figure 3.10: Mean recognition values obtained for the Front-Back-Left-Right configuration. The means for no angular difference (0°) indicate 'false positives' i.e. the perception of an angular difference where none exists [121]

The results also demonstrate the well-established principle of reversals (front-back and back-front confusions) and localisation smear or deviation from target cues. A total of 80.6% of the stimuli presented in front were localised to the back. A markedly lower percentage, 14.8%, presented at the back were localised to the front. While both of these types of confusions have been reported in earlier literature [52] [55] [56] [122] [46], the values that we have obtained appear to be quite high in comparison to the approximately 29% (front-back confusions) seen in the study by Wenzel et al. [56]. This could be due to a combination of the non-individualised HRTFs used to treat the stimuli and the lack of any optimisation, as suggested by Walker and Stanley [113], for bone conduction headset based reproduction. The data also shows a tendency of participants to identify co-located sources as spatially separate. These localisation judgements account for approximately 40% of the co-located trials being identified as spatially separate and are treated as false positives (Figures 3.8 and 3.10). While this is definitely concerning when considering information delivery exclusively via an auditory display, we believe that the false positives will not adversely affect the information delivery in a wearable, audio-visual interface.

A mean deviation of 32.4° for the front and 34° for the rear was observed for the traditional front-back division of the auditory space. This is in agreement with existing literature that demonstrates that localisation accuracy in the front is greater than at the back [105] [46]. For the alternate arrangement (Figure 3.9), a similar trend for localisation in the front (Mean: 30.7°) and back (Mean: 32.4°) was observed. The sides, show larger deviations (Left: 35.5°, Right: 34.9°). Errors in localisation judgements calculated here account for front-back and back-front confusions [56]. Another trend that was observed during the course of data analysis was the ‘wrap around’ artefact described by Stanley [16]. This refers to the perception of a target presented in the right hemi-field appearing in the left or vice-versa (Figure 3.11).

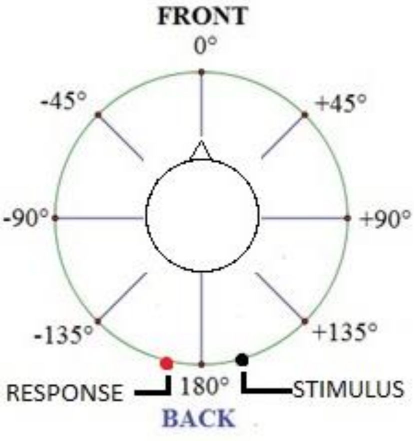


Figure 3.11: Wrap around artefact [121]

Another form of the wrap around artefact that was observed was a ‘reflected wrap around’. In this case a stimulus presented in the frontal quadrant was localised to the diagonally opposite, rear quadrant (Figure 3.12). This was mainly observed for stimuli that were presented within ±30° of the median plane in front of the observer.

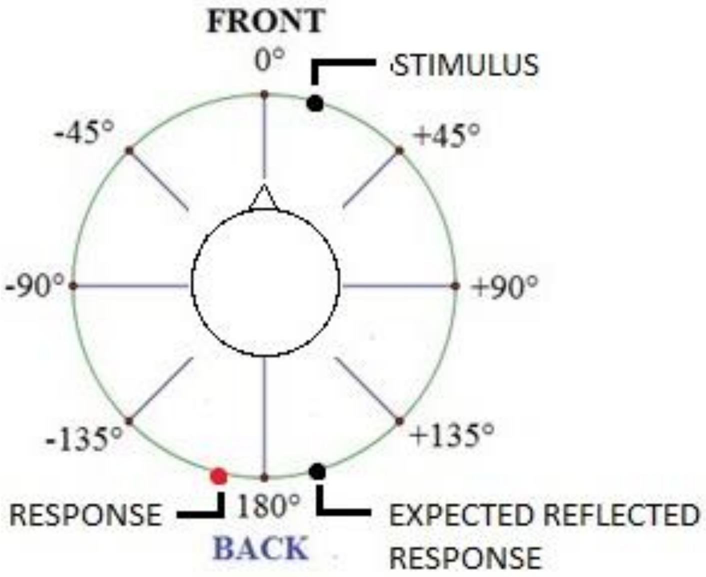


Figure 3.12: Reflected wrap around artefact [121]

3.5 Discussion

The results obtained here demonstrate that discrimination between two, binaurally spatialised sources of sound is possible over a bone conduction headset. The results, not surprisingly, reveal that that ability to discriminate between two successively spatialised sources is significantly diminished in comparison to similar tasks undertaken under free-field or headphone based conditions. None-the-less, we have been able to achieve results similar to what our pilot studies suggested the bone conduction headset was capable of achieving (20° in the front and back for the front-back-left-right configuration and 10° for the front-back configuration). In light of these results, we can safely assume that the spatial separation afforded by binaural reproduction over a bone conduction headset is sufficient for it to be used as a spatial auditory display device. The spatial separation and externalisation afforded by it is of greater importance than the accuracy.

While the deviation values observed here are approximately twice as much as those obtained in earlier studies on auditory perception [52] [55] [56] [122] [46] [120], whether headphone or loudspeaker based, they fall within the range for which the ventriloquist effect is operational as demonstrated by Kyto et al. [66] in an Augmented Reality environment and are marginally higher than the upper limit of 30° as demonstrated by Jackson [61] in the real-world. The Ventriloquist Effect (VE) is a phenomenon wherein spatially separated auditory and visual events appear to be co-located. This is an important aspect since the perception of sound in a bi-modal environment is inextricably linked to visual perception as we will see in the third study covered in Sections 3.13 through 3.18. The issue of false positives encountered in this study also ceases to exist when an auditory and visual cue are simultaneously presented to participants.

These results bode well for future developments since the constraint going forward is the user’s ability to perceive multiple simultaneously or successively presented sound sources, and the inclusion of visual cues. The availability of bone conduction headsets at an affordable price and the increased quality of reproduction mean they are a viable alternative for a range of wearables. It provides an ideal alternative to headphones in a multi-tasking environment where awareness of one’s surroundings is critical.

3.6 Conclusion

In this study binaural spatialisation over a bone conduction headset was explored using inexpensive, off-the-shelf hardware and software components. The results demonstrated the ability of this setup to provide adequate spatial separation between sources. While discrimination between successively spatialised sources, 10° in the front for front-back configuration and 20° for the front-back-left-right configuration, was not as good as that seen in the free-field or headphones based conditions, we feel it is sufficient for the purpose for which this was explored. Further investigation into binaural spatialisation over a bone conduction headset is required to assess localisation performance in the vertical plane in a manner similar to previous studies [52] [55] [56] [122] [120]. The results of this study are encouraging and point toward the potential of bone conduction headsets being used as a spatial information display for wearable hybrid interfaces in the coming years.

Study II

BINAURAL SPATIALISATION OVER A BONE CONDUCTION HEADSET: THE PERCEPTION OF ELEVATION

Our second experiment looked to evaluate the perception of elevation for material spatialised over a bone conduction headset. This seemed like a natural second step to take in order to evaluate the effectiveness of the target device at being able to reproduce spatialised content that would be informative. Early results from this study were presented as a poster at the *22nd International Conference on Auditory Display (ICAD-2016)* in Canberra, Australia between the 2nd and 8th of July 2016. Subsequently, a full paper was submitted to and accepted for presentation at the *AES Conference on Headphone Technology* in Aalborg, Denmark between the 24th and 26th of August 2016. An extended version of the paper has been submitted to the *Journal of the Audio Engineering Society* and is currently under review.

3.7 Introduction

As we have already seen in previous sections (refer to sections 2.4 And 3.2), several researchers have proposed the use of an auditory interface for information display on wearable devices. Besides the obvious benefit of helping de-clutter the screen, the serendipitous nature of sound [11] has been shown to reduce cognitive load in comparison to a visual only interface [80]. This is an important characteristic of sound, given that wearable devices of the future are likely to be ‘always on, always accessible’ computers that could be worn for a significant portion of the day. In such a scenario, distractions brought about by visual only interfaces will be inevitable. Despite auditory interfaces being proposed as a solution to the issues posed by visual interfaces, we believe that the inherent singular nature of information delivery using either of the modalities is a limiting factor. For an interface to truly succeed at information delivery it must be able to provide its user with actionable information in an unobtrusive and context sensitive manner. This can only be achieved if we design the interface to the strengths of our two most prominent faculties – audition and vision.

In this thesis, we aim to explore the use of these two faculties to provide relevant information to the user as and when required. The experiment described in the following sections is a continued attempt to assess and quantify the operational capabilities of the bone conduction headset. In this study we attempt to further our understanding of binaural spatialisation over a bone conduction headset by examining the perception of elevation. As with the previous experiment, we will compare our results with existing headphone and bone conduction based studies to gauge performance and suggest future directions for our research.

3.8 Motivation and Background

We have already established that binaural spatialisation over a bone conduction headset is of a good quality; and that the localisation it affords in the horizontal plane compares well with existing studies in the field [104] [16]. These results have been the primary motivation driving this study. It is only logical that having explored binaural spatialisation and localisation accuracy in the horizontal plane, we sought to do the same in the vertical plane. Results from this experiment, along with those from the previous study have helped us establish the operational parameters that of the bone conduction headset – important information that will help guide design decisions in the future.

Applications of virtual auditory displays extend from complex military applications which help make cockpits of fighter jets less complicated [58] [115] to everyday information retrieval challenges faced by users of mobile devices [12] [13] [126]. Almost all of these displays make use of HRTFs to process the audio signal being delivered to the user. This allows for a more realistic sound stage reproduction that mimics the natural acoustic environment. Such a display also allows for the use of sound in a more efficient manner utilizing its ‘serendipitous’ nature [11]. Alerts, messages and other forms of non-visual communication can easily be delivered without interrupting a primary task that the user may be involved in. For detailed overview of the background information please refer to sections 2.4 and 3.2. Section 2.4 provides a broad overview of the progression of wearable auditory displays over the years, while section 3.2 deals specifically with the literature that is most relevant to this study.

3.9 Method

3.9.1 Apparatus, Stimuli and Calibration

The apparatus, stimuli and their calibration were identical to that used in the previous study. The only difference between the previous study and this one is that instead of a two sound sources, this one used only a single sound source. The duration of the stimulus was identical to the ones used for the previous study. Refer to section 3.3 for detailed list of apparatus and stimuli used. The section also explains in detail how the calibration was carried out and the equipment used for the purpose.

3.9.2 Participants

Fourteen participants (12 male, 2 female) aged between 19 and 43 years (Mean: 27.5, SD: 9.8) volunteered to take part in the study. All participants reported normal hearing. No audiometric screening was undertaken to verify their claims of having normal hearing. All participants were compensated with NZD $10 gift vouchers for taking part in the experiment.

3.9.3 Procedure

Similar to the previous study, participants were first shown how to put on the bone conduction headset and then seated in front of the loudspeaker (Figure 3.13). A complete explanation of the experiment and what was expected of them was given. Following this, the calibration process was completed.

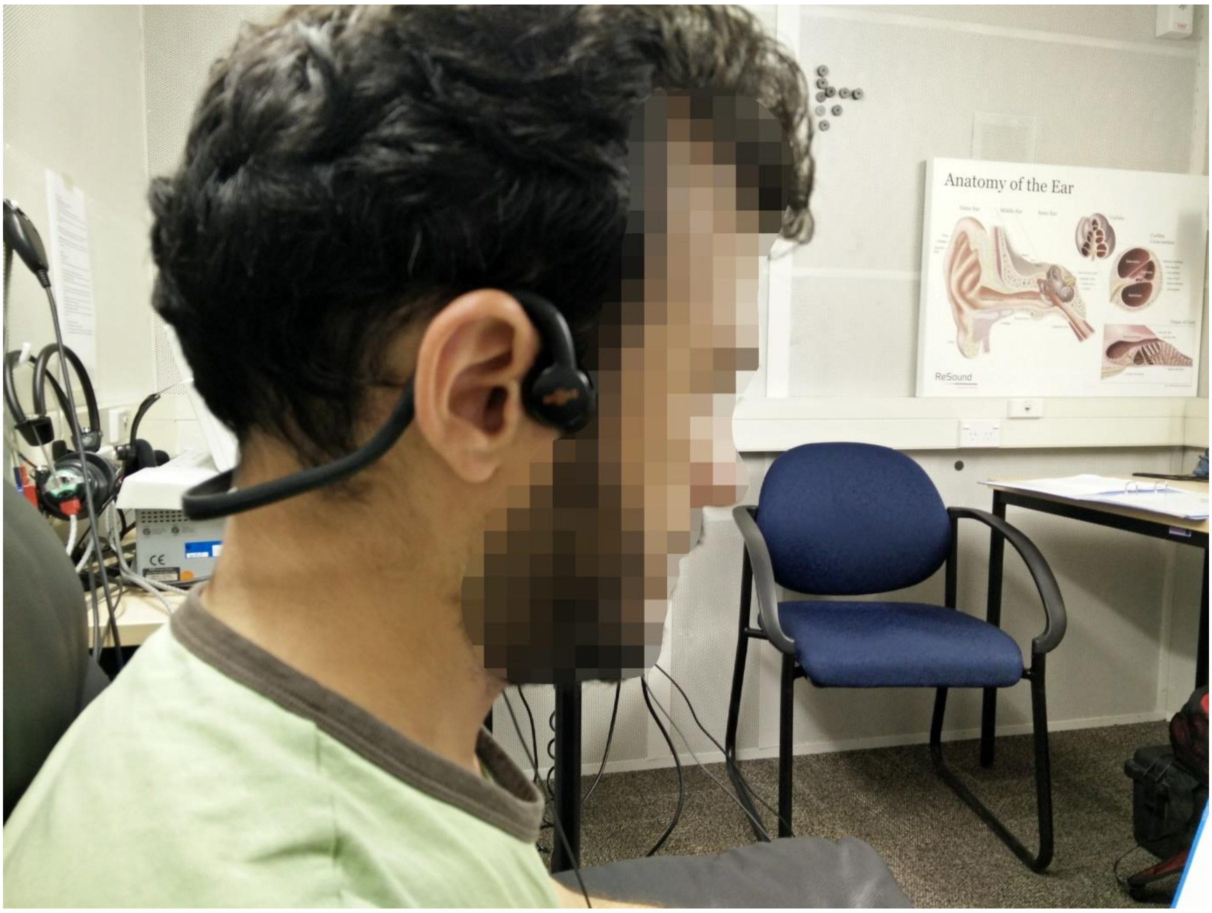
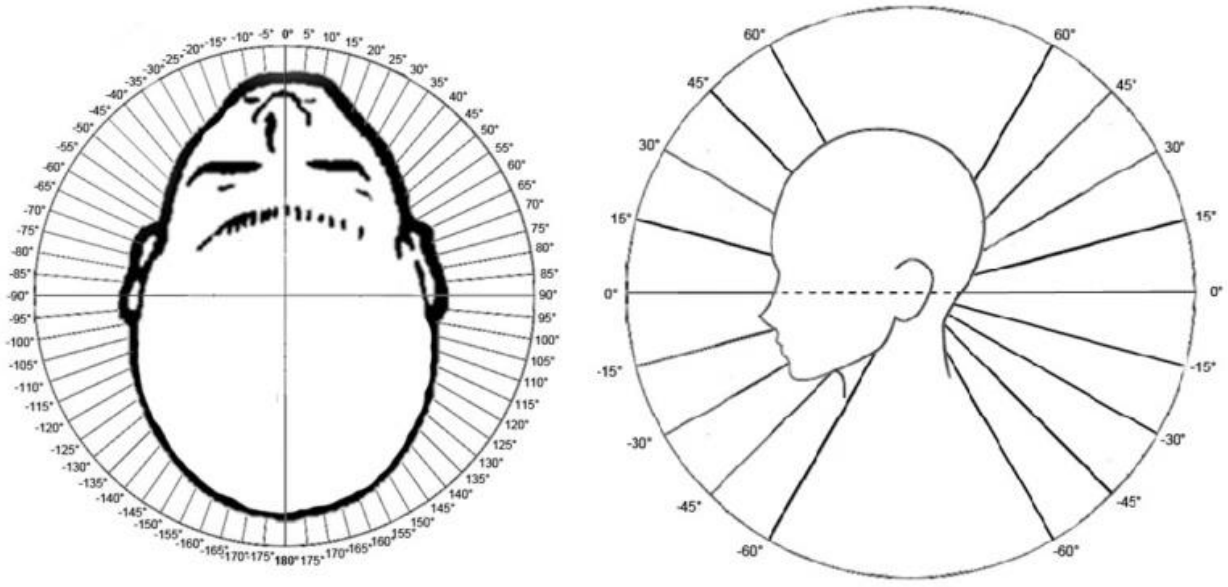


Figure 3.13: Participant wearing the bone conduction headset

Prior to the main experiment, participants undertook a trial block not exceeding five minutes. The purpose of this trials block was to allow participants to gain familiarity with the task of localizing the presented stimulus in the horizontal and vertical planes. Localisation judgements were made using a combination of two response charts (Figure 3.14). The charts were placed in the horizontal (Figure 3.14a) and vertical (Figure 3.14b) positions to help participants ‘visualize’ and interpret what they heard to provide as accurate a localisation as possible of the stimuli.



(a) (b)

Figure 3.14: Response charts used by participants to localise the stimulus in the (a) horizontal plane & (b) vertical plane

For the chart depicting the vertical orientation, participants were asked to call out the nearest signed angle if they felt that a stimulus lay somewhere between the two angles shown on the chart. For the main experiment participants were presented with the stimulus in the vertical plane within a ±45° range spread around a 360° horizontal plane. The elevations were spread over 15° intervals resulting in a total of seven elevations. Spatialisation was limited to within the ±45° range since pilot testing showed severe degradation of azimuth perception as elevation exceeded ±45°. In the azimuth spatialisation was spread over a full 360° plane around the participant’s head. The step size for azimuth was also set to 15°. Each elevation was reproduced twice for every azimuth. This resulted in a total of fourteen trials across all elevations for every azimuth. Ninety-eight trials were conducted per quadrant (14 elevations [2 trials/elevation] x 7 azimuths/quadrant) resulting in a total of 392 trials per participant for all elevations within a 360° space around the head.

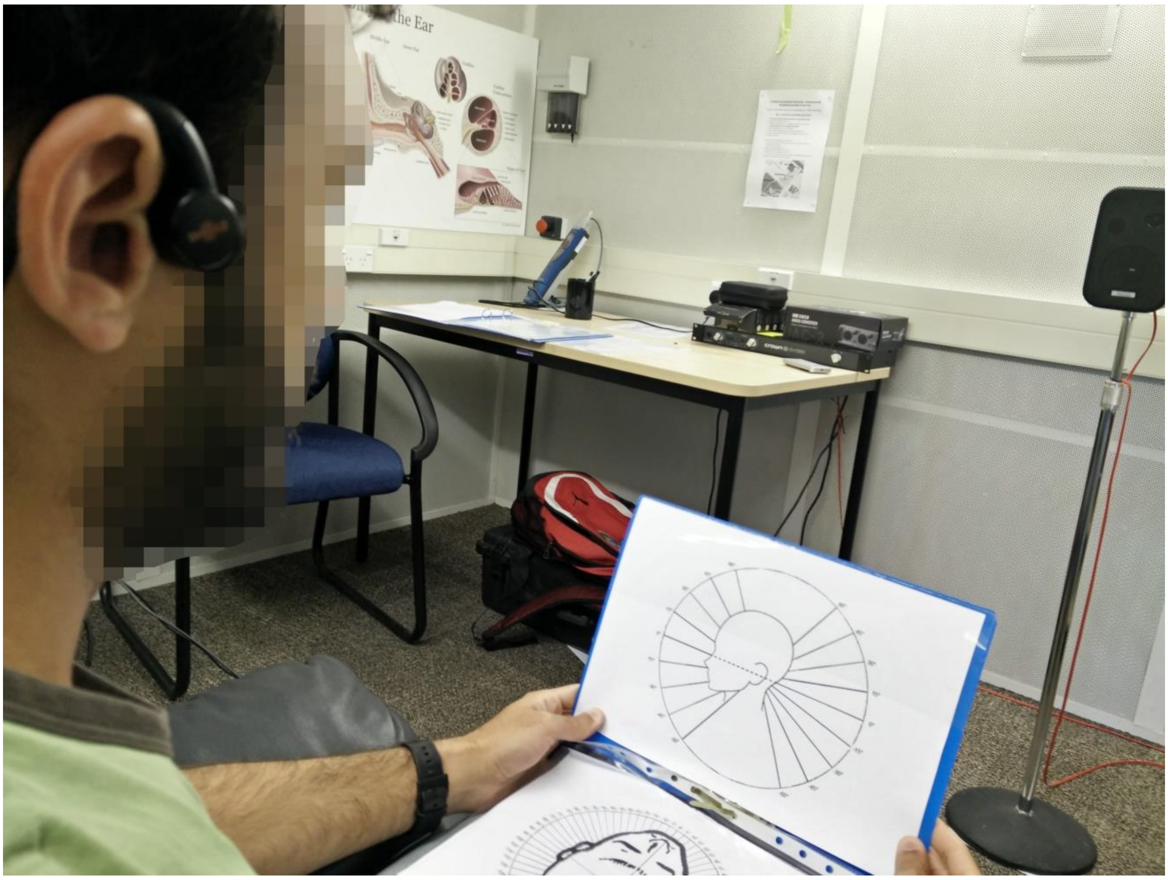


Figure 3.15: Participant using the response charts to indicate the perceived location of the stimulus

The experiment was divided into two blocks with participants getting a ten minute break once half the trials were completed. Unlike the previous experiment [121], this one was conducted quadrant-wise. A random switching order between the quadrants was employed during the experiment. All trials within a given quadrant were completed before switching to the next one. During the experiment participants were asked to call out the location where they perceived the stimulus using the signed angle protocols explained to them prior to the start of the experiment. They could refer to the response charts in order to be able to provide as accurate a judgement as possible (Figure 3.15). All participants were instructed to keep their heads straight and look directly at the calibration speaker for the duration of the experiment, although this was not strictly enforced. Externalisation ratings were also solicited from participants at the end of the experiment. These ratings were based on methods previously employed by Stanley [16] and Gardner [122] in their studies.

3.10 Results

Two of the fourteen participants opted out of the experiment after completing two blocks (N = 12). The results were evaluated in a manner similar to that described in [121]. A traditional front-back (Figure 3.16) and an alternate front-back-left-right configuration (Figure 3.23) were employed. The following section describes the results obtained using the front-back configuration.

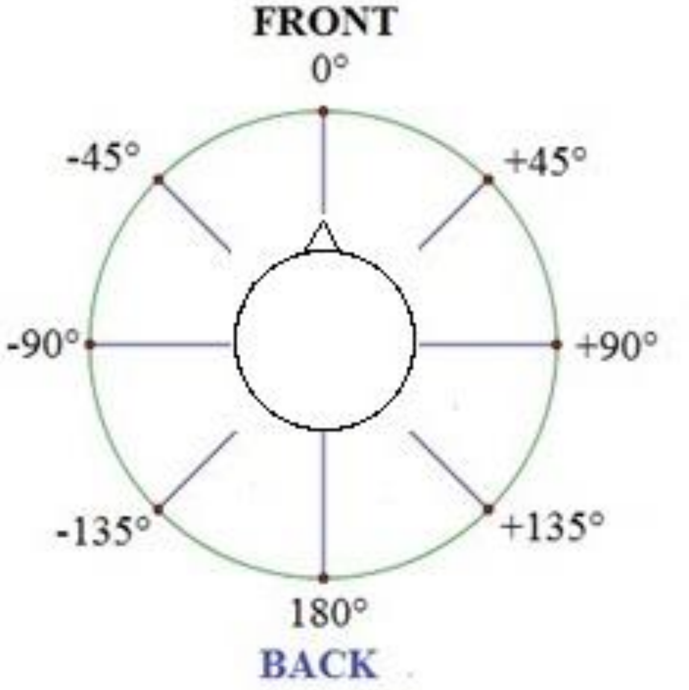


Figure 3.16: A traditional front-back configuration widely seen in psychoacoustic research [121]

Results obtained in this configuration demonstrate behaviours similar to those in headphone-based localisation studies using non-individualised HRTFs. Besides localisation accuracy in the horizontal and vertical planes, we also recorded the absolute errors in vertical localisation observed over the course of the study (Figure 3.17). These errors of judgement range between 17° and 49° for the front and 20° to 44° for the back. A high level overview of the errors demonstrates a significant effect of elevation (F (6, 66) = 14.115, $p < 0.001$) on the magnitude of errors.

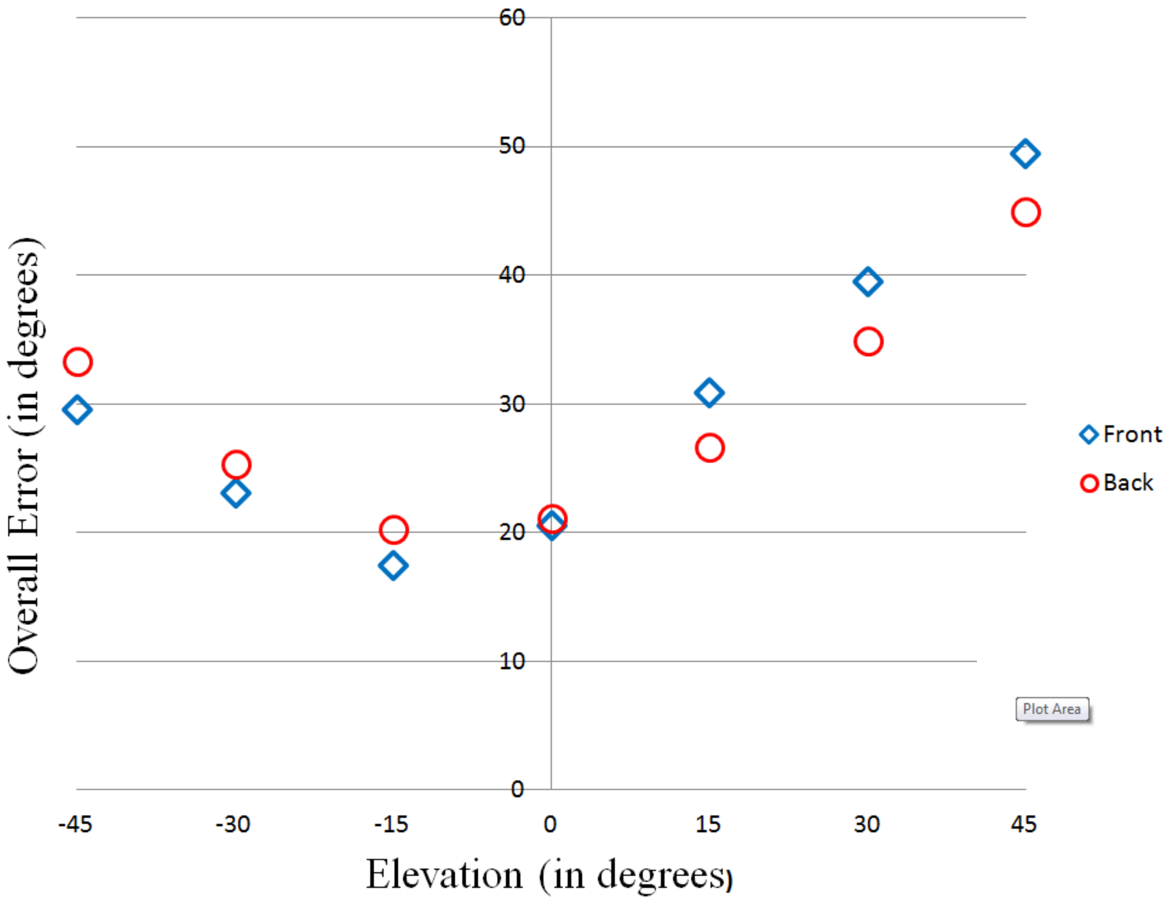


Figure 3.17: Mean errors for front and back across all elevations

We can see a distinct decrease in errors as we move from +45° to -45°. There also appears to be a noticeable pattern separating errors in the front and back above and below the interaural plane. For all elevations below the interaural plane, errors in the rear hemisphere appear to be of a greater magnitude than those in the front, while the opposite is true for elevations above the interaural plane. The front-back hemispheres and the interactions between the hemispheres and the same elevations compared across them appear to have no significant effect on the errors. To obtain a clearer picture of error distribution, underestimation and overestimation errors were calculated. A localisation judgement was considered to be an overestimation when the judged elevation was above the target source (Figure 3.18a). Similarly, a localisation judgement was considered to be an underestimation when it was perceived to be below the source (Figure 3.18b).

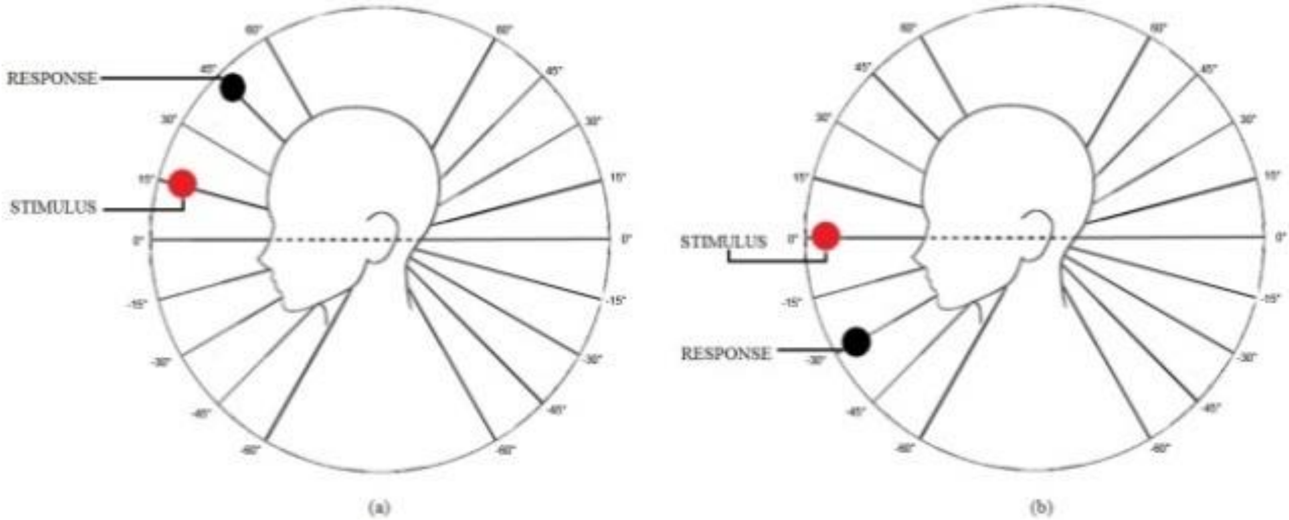


Figure 3.18: Judgement errors of (a) overestimation and (b) underestimation as observed in the experiment

Both, underestimations and overestimations were the absolute errors of judgement made by participants. They were analysed for differences that may exist between the front and back in addition to the effect of the target elevations themselves on the judgements. The data obtained was tested for normal distribution. The resultant non-normal distribution of the data motivated the use of the Aligned Rank Transform (ART) [127] before a two-way repeated measures Analysis of Variance (ANOVA) was carried out. Analysis of the overestimations showed no statistically significant differences between the front and back (F (1, 11) = 2.534, p = 0.140). This indicates that localisation accuracy in both hemispheres follows a similar trend. There was no significant interaction effect between elevation and front-back hemispheres either (F (6, 66) = 1.472, p = 0.202). Elevation, on the other hand, was found to have a significant effect on the errors of overestimation (F (6, 66) = 69.427, $p < 0.001$).

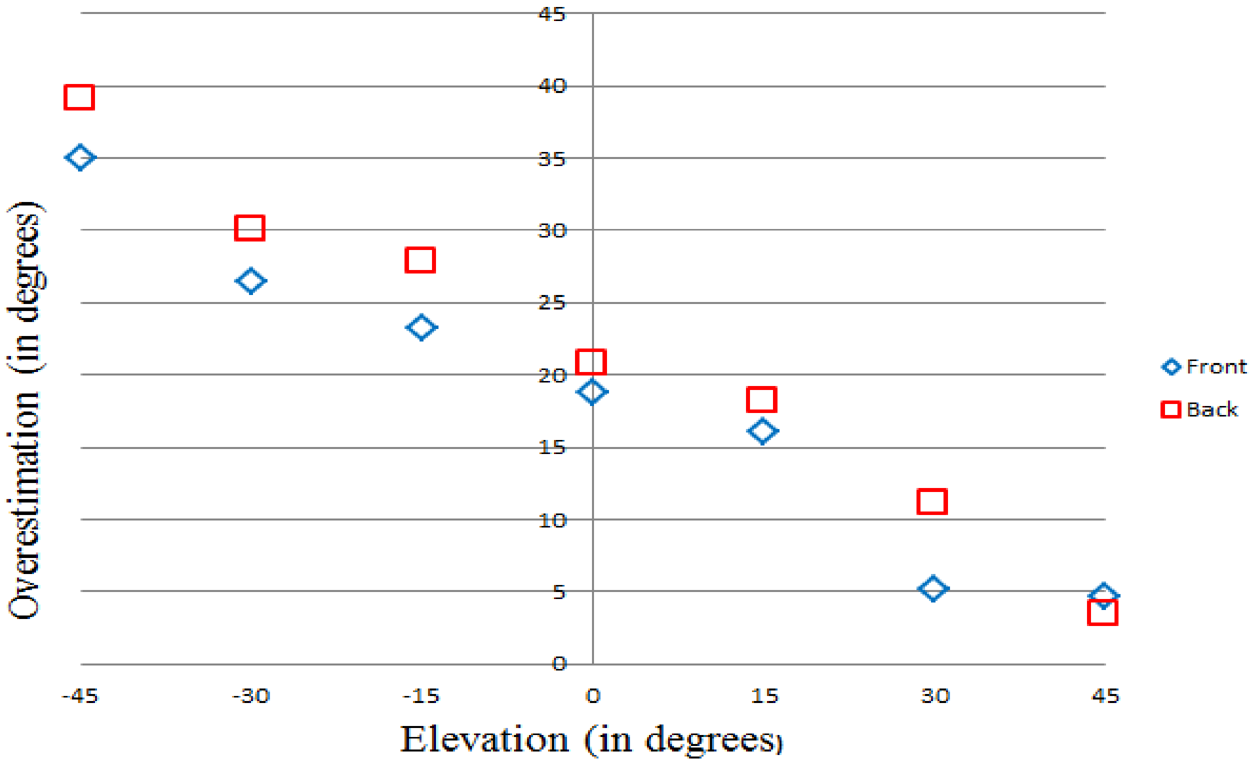


Figure 3.19: Mean errors of overestimation for front and back across all elevations

Figure 3.19 demonstrates the extent to which a change in elevation appears to affect the errors of overestimation. Mean errors range from 4.6° for the front and 3.4° for the back at 45° elevation, to 35.02° for the front and 39.1° for the back at -45° elevation. While no significant differences in overestimation errors were found between the front and back, figure 3.19 shows a trend that these errors are larger in the rear hemisphere in comparison to the front. A comparable trend was observed with the front and back demonstrating no significant effect on the errors of underestimation (F (1, 7) = 0.004, p = 0.953). Elevation, as seen before, significantly affected the magnitude of errors encountered (F (6, 66) = 91.945, p < 0.001) (Figure 3.20).

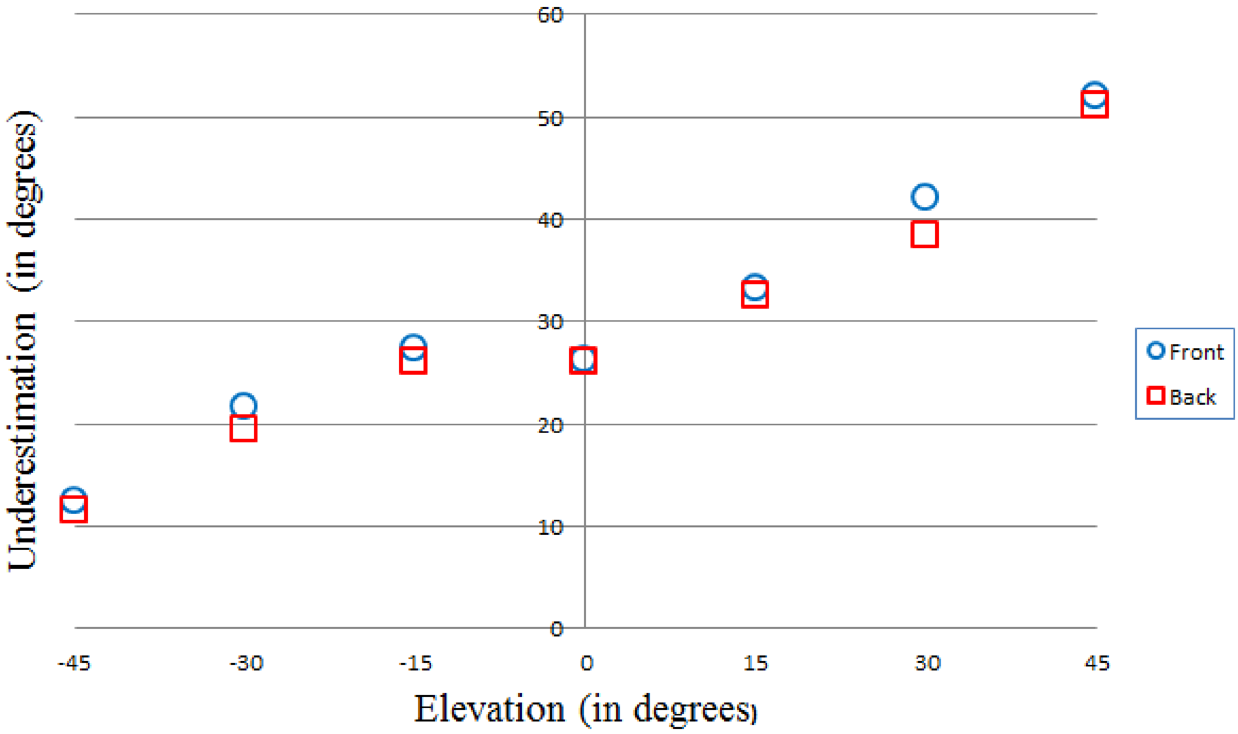


Figure 3.20: Mean errors of underestimation for front and back across all elevations

The interaction between the two independent variables of elevation and hemispheres (front-back) was not significant (F (6, 66) = 0.297, p = 0.936). The manner in which the errors were distributed, and their magnitudes, suggests that errors in underestimation appear to occur to a greater extent for target elevations above the interaural plane. While targets at -45° elevation only exhibit an approximate error of 12° and 11° for the front and back respectively, at a +45° elevation those errors increase to approximately 51° for the front and 50° for the back (Figure 3.20). There appears to be a very distinct pattern to the errors in underestimation and overestimation recorded here. On resolving these errors, we were able to obtain a vertical range within which the judgements are localised (Figure 3.21). Resolution of the errors was achieved utilising a method used by Wenzel et al. [55]. For example, if a sound source spatialised at a +30° elevation was localised at -15°, this judgement was reflected back in to the upper hemisphere i.e. +15°. Following this, the localisation blur was calculated depending on where the reflected judgement appeared. In the example above, while the absolute error of judgement is 45, the error recorded after the judgement is resolved work out to be 15°. This method of resolution is widely used in psychoacoustic studies involving the perception of a sound source in the horizontal plane.

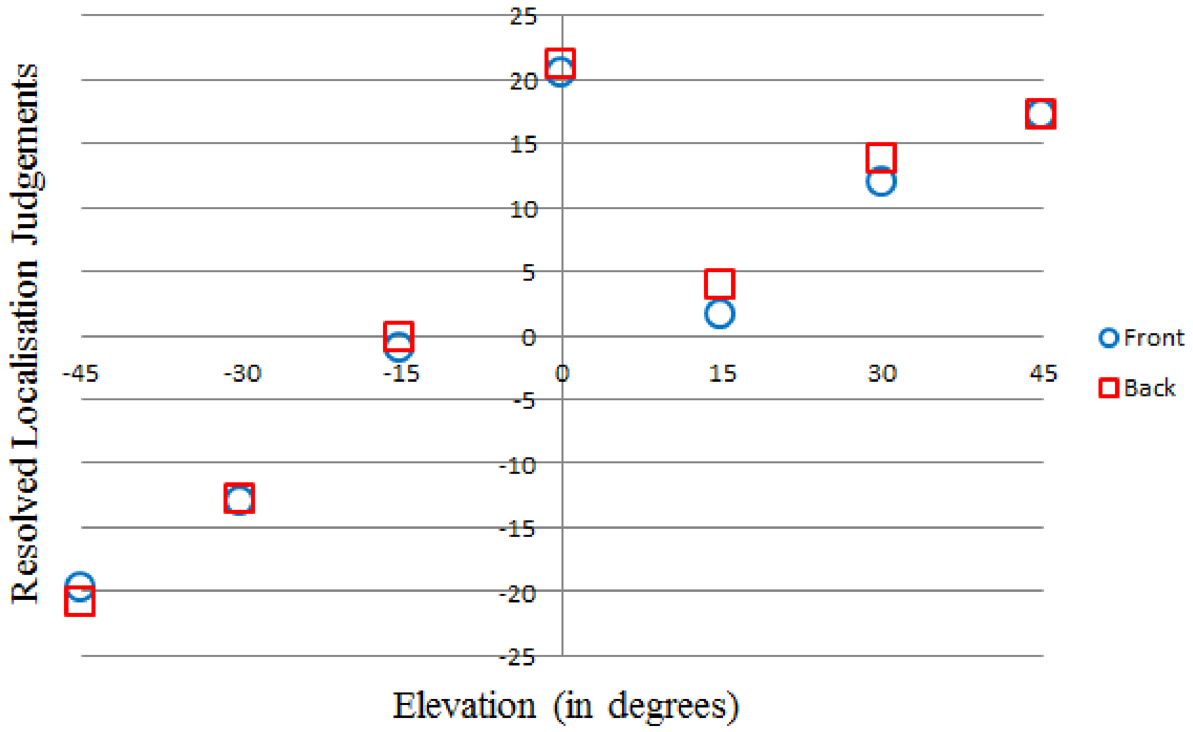


Figure 3.21: Mean resolved localisation judgements

Figure 3.21 demonstrates that the result of resolving the underestimations and overestimations in judgements is a ‘window’ within which all elevation localisation judgements appear to be made. This extends to an approximately equal angular distance of 20° from the interaural plane in either direction.

As expected with use of non-individualised HRTFs, front-back and back-front reversals were observed. Since the experiment also involved spatialisation in the vertical plane, up-down and down-up reversals were also observed. An up-down reversal is where a stimulus presented in the upper hemi-field is localised to the lower hemi-field (Figure 3.22b). Occurrence of an event exactly opposite to the one described above constitutes a down-up confusion.

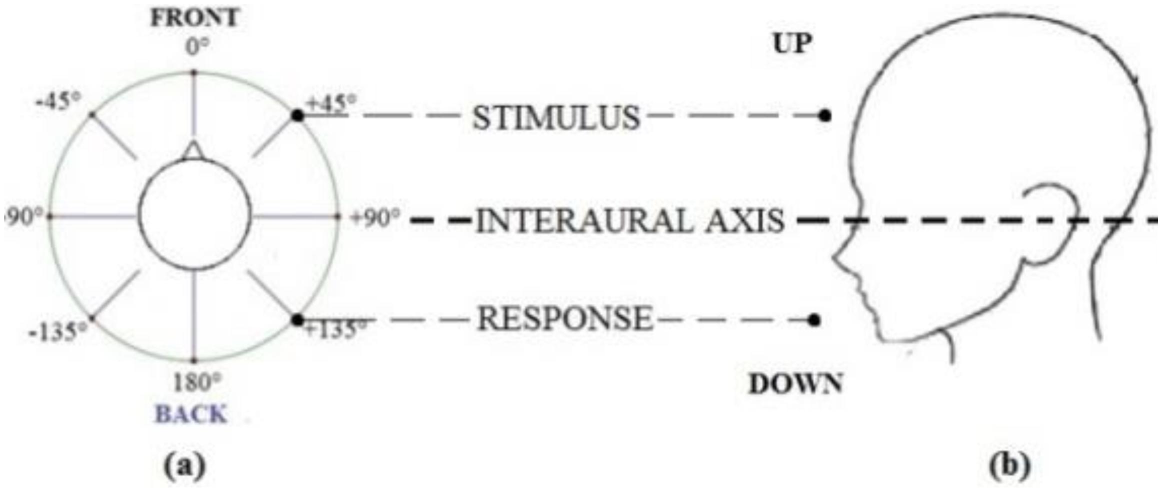


Figure 3.22: (a) Front-Back confusions (b) Up-Down confusions

In this study, 23.8% of the trials in the front resulted in up-down confusions while 21.8% of the trials in the rear hemisphere demonstrated the same result. Down-up reversal rates on the other hand were comparatively low at 4.7% and 6.9% respectively for front and back. Front-back confusions were quite high with 82.03% of the trials in the front being localised to the back, while 4.3% of trials exhibited back-front confusion. Angular deviation for both horizontal and vertical planes appears to be quite high. Localisation blur in the vertical plane was observed to be approximately 21° and 20° respectively for the front and back. A relatively high deviation in the horizontal plane of approximately 44° and 51° is seen for the front and back respectively. These high deviation values appear to be because of two participants who displayed an extremely impaired ability to localise the presented stimulus. If their results are excluded, the deviation values drop to approximately 37° for the front and 41° for the back. These though, are still comparatively higher than the values seen in [121].

In the following section we look at the results obtained from the alternate configuration mentioned earlier in the section. The configuration used here is the same as that in [121]. This extends the sides from ±60° - ±120° used by Wightman and Kistler in [52] to ±45° - ±135° (Figure 3.23).

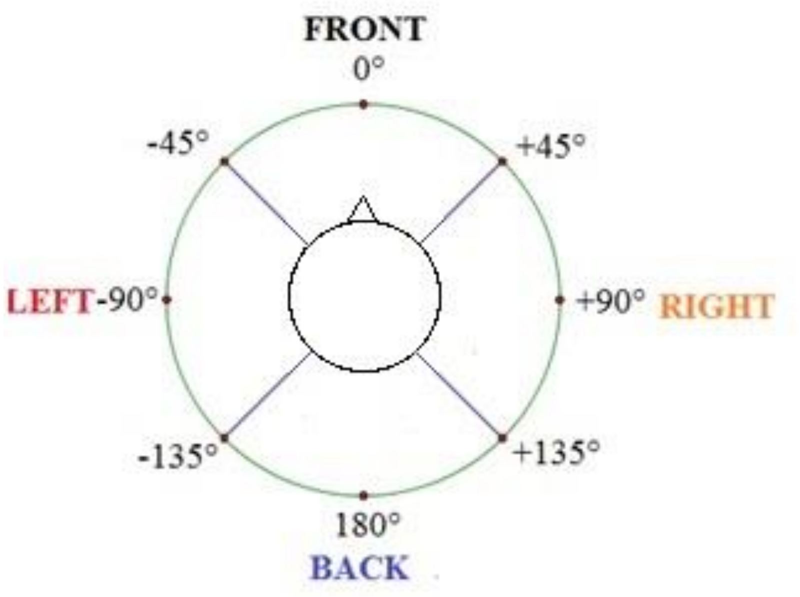


Figure 3.23: Alternate configuration used for evaluating results [121]

The same analysis was run on the error distribution obtained after dividing the area around the head into four equal quadrants. The overall error distribution for the four quadrants can be seen in figure 3.24.

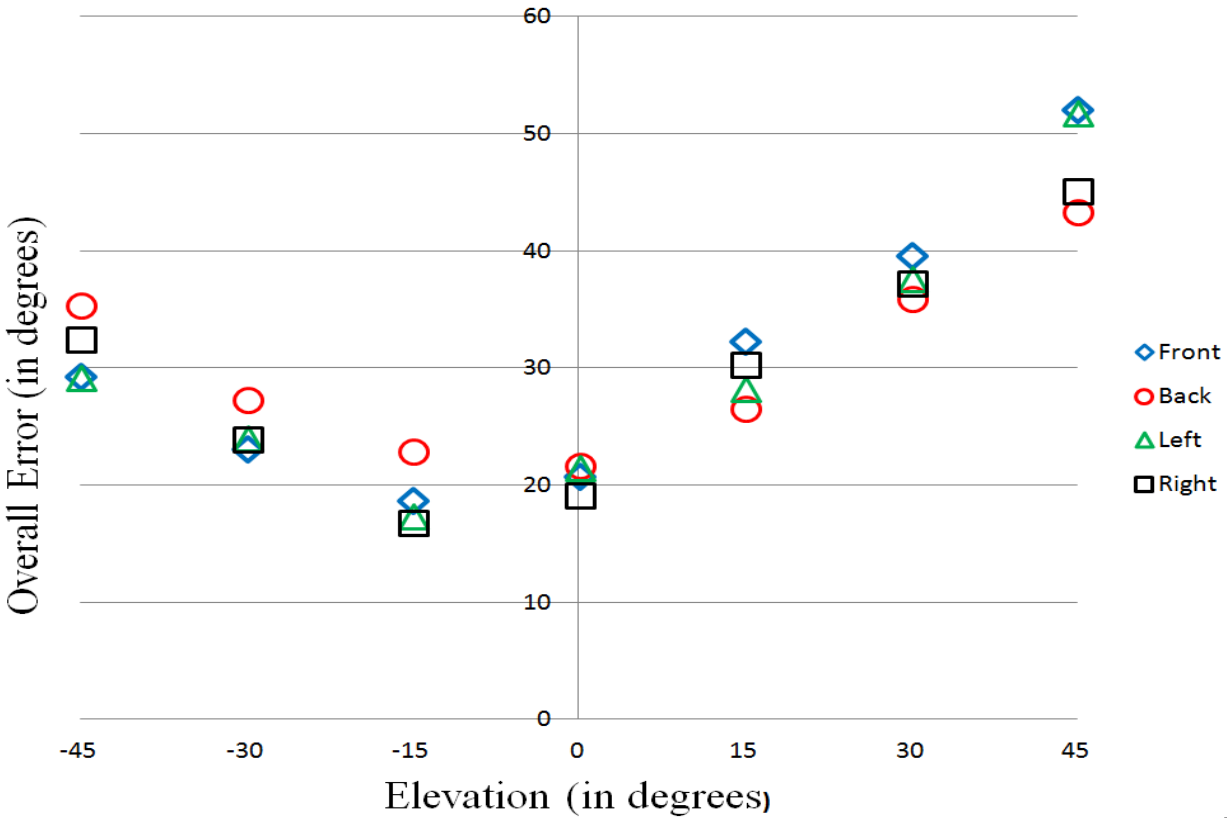


Figure 3.24: Mean errors for front, back, left and right across all elevations

The data follows a pattern similar to the one seen in the previous configuration. Just as with the front-back distribution, only a change in elevation appears to have a significant effect (F (6, 66) = 13.481, p < 0.001) on the magnitude of error. For errors of overestimation, a two-way repeated measures analysis of variance (ANOVA) with ART revealed a significant effect of elevation (F (6, 66) = 93.187, p < 0.001). Similar to the previous configuration (Figure 3.16), changes in elevation significantly affected the magnitude of errors recorded (Figure 3.25).

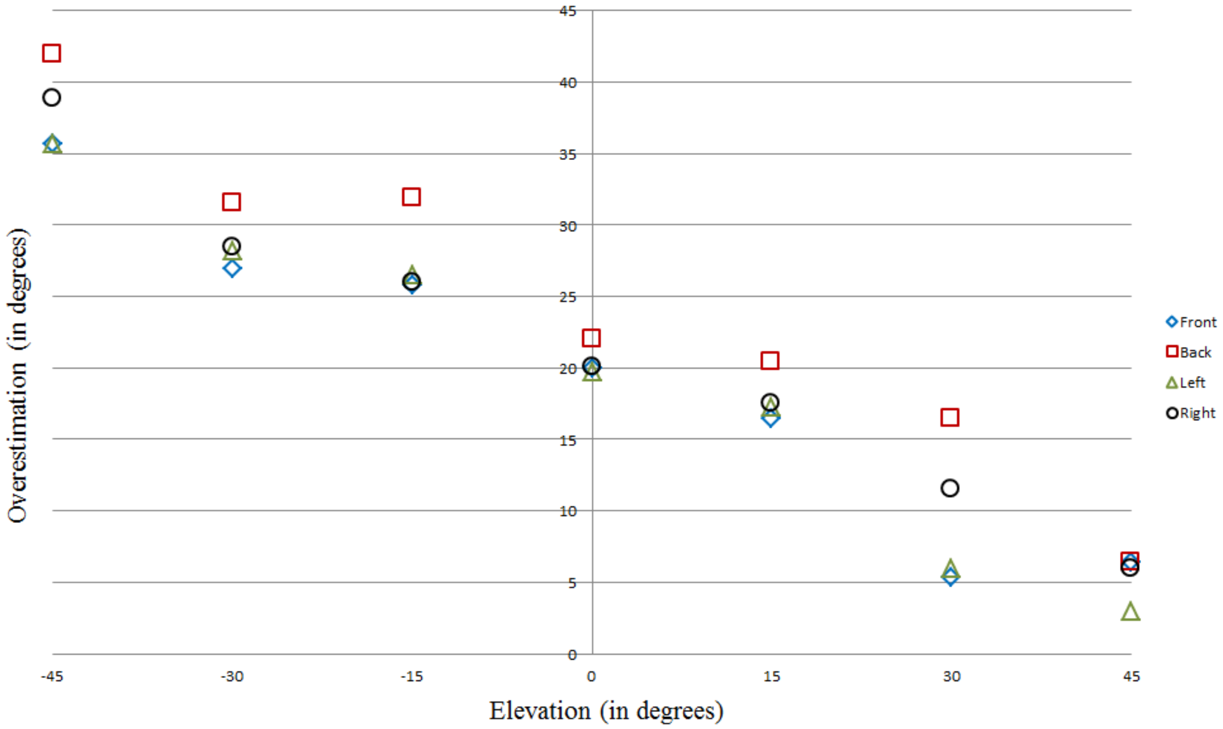


Figure 3.25: Mean errors of overestimation for front, back, left & right across all elevations

Overestimation errors increase as the elevation decreases. An elevation of -45° appears to show the largest errors in overestimation, with this elevation in the rear quadrant resulting in an overestimation of nearly 42°. The same elevations across the four quadrants demonstrated no significant effect on the recorded error rates (F (3, 33) = 2.547, p = 0.073). There was no significant interaction effect between the two variables (F (18, 198) = 1.605, p = 0.061). While not statistically significant, from figure 3.25 we can observe a trend that errors in the rear quadrant are consistently greater across all elevations than the rest of the quadrants.

Similarly, for errors of underestimation a significant effect of elevation was found (F (6, 66) = 124.422, p < 0.001). As with the overestimation errors, error distribution for underestimated targets follows a distinct pattern. Underestimation appears to drastically increase as the elevation increases, with the 45° elevation demonstrating an approximately 54° underestimation (Figure 3.26).

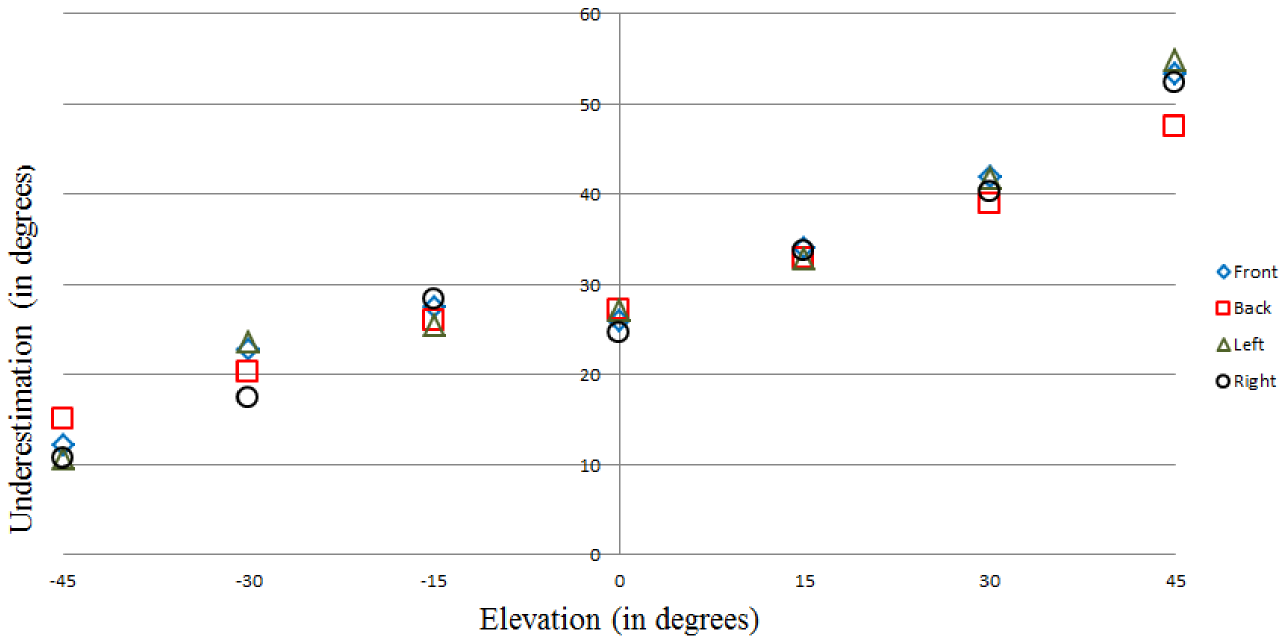


Figure 3.26: Mean errors of underestimation for front, back, left & right across all elevations

No significant effect across the quadrants was found (F (3, 33) = 1.538, p = 0.223). The comparison between elevations themselves across the quadrants demonstrates closely matched errors rates (Figure 3.26). The interaction effects between the quadrants and elevation do not appear to be statistically significant either (F (18, 198) = 0.820, p = 0.675).

Resolution of the underestimations and overestimations gives us a range within which spatialised targets appear to be perceived (Figure 3.27). This range is almost similar to one seen for the front-back configuration. For the front, localisation estimations appear to be concentrated within a range of approximately -22° to 20°. This remains the same for the back with a marginal increase in the upper quadrant from 20° to approximately 21°. The left and right resolutions of estimations also demonstrate similar 'windows' of approximately -20° to

22° and -19° to 19° respective within which localisation judgements appear to be concentrated.

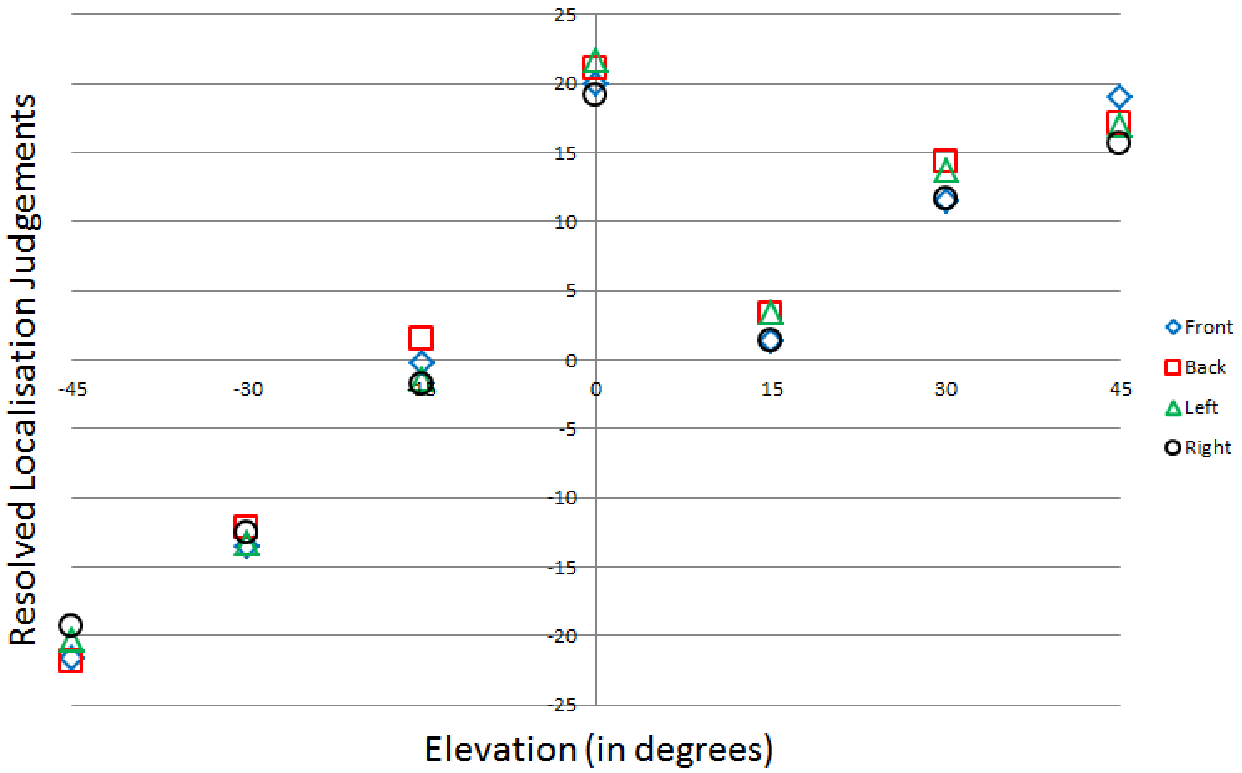


Figure 3.27: Mean resolved localisation judgements for the front, back, left & right distribution of auditory space

Externalisation was reported by eight out of the 12 participants. Of the four participants who did not report any externalisation, three indicated that the stimulus sounded like it was at the centre of the head while the fourth participant indicated that the stimulus felt like it was just under the surface of the head. Of the 8 participants who reported externalisation, five reported the stimulus as being on the surface of the head, two between the surface and one metre away from the head and one participant reported hearing the stimulus over a metre from the surface of the head. The worst localisation appeared to be in front top quadrant. This result can be looked up on as being exactly opposite to established localisation theories and results which have demonstrated that the worst localisation is said to occur in the upper rear quadrant [56] [46].

3.11 Discussion

The results seen here agree with existing literature on binaural spatialisation over headphones using non-individualised HRTFs [56] [55]. Despite the larger variations in localisation results, the trends observed here indicate that binaural spatialisation over a bone conduction headset appears to fundamentally work in the same manner as it does over headphones. Although the impaired perception is a result that we expected considering no optimisation of the stimulus or the bone conduction headset of any sort was carried out. Taking this into consideration, our results here are encouraging and bode well for the use of a bone conduction headset as an auditory display device capable of reproducing binaurally spatialised sound. While our study brought to light a surprising result contrary to existing literature for localisation in the upper quadrants [56] [46], the difference in overestimation recorded for the front quadrant in relation to the rear for both configurations is less than a degree. This could be due to the division of the front and rear space for both the configurations. Oldfield and Parker [46] have stated that the worst localisation accuracy in their study was observed for upper elevations in the rear quadrant between 150° to 180°. For the purpose of our study, we considered the rear to extend from +90° to -90° for the front-back configuration (Figure 3.16) and +135° to -135° for the alternate configuration (Figure 3.23).

Besides this, all are also in agreement with existing literature on auditory perception in the free field as well as over headphones. Thus, we can safely assume that even in this respect the bone conduction headset appears to demonstrate similar working principles but with a greater degree of deviation. We also noticed that the addition of the vertical component caused significant distortion of localisation in the horizontal plane as demonstrated by the difference between the results for deviations in the horizontal plane observed here versus [121].

3.12 Conclusion

We demonstrated the effectiveness of the bone conduction headset in reproducing a binaurally spatialised sound source. While the bone conduction headset appears to be prone to significantly greater inaccuracy that that observed in headphone-based localisation studies, it does offer the advantage of unhindered, natural auditory perception. Since we propose the use of a bone conduction headset as part of a wearable hybrid interface incorporating auditory and visual feedback, we believe that a significant number of the localisation errors seen here will be accounted for or be rectified by the addition of visual cues.

Upcoming studies such as the one carried out in [128] will explore how vision and audition can be combined in order to provide an unobtrusive, yet useful stream of information. Experiments with greater ecological validity will be able to provide us with more data regarding the performance of the bone conduction headset as part of a wearable hybrid interface across a broad range of use cases. The results obtained from [121] and this study are a clear indication of the use of bone conduction headsets as an alternative form of auditory display.

Study III

THE VENTRILOQUIST EFFECT IN AUGMENTED REALITY

While the first two studies examined the issue of binaural spatialisation over a bone conduction headset exclusively using audio, our third and final psychoacoustic study looks at the how visual cues can affect the perception of a sound source. We do this by looking at the ventriloquist effect and how it pertains to auditory and visual perception in augmented reality. This is an especially important topic since the aim of this thesis is to develop novel ways of interaction using both the visual and auditory modalities.

3.13 Introduction

This experiment is a continuation in our ongoing efforts to evaluate the use of the bone conduction headset in bi-modal, wearable interfaces [121] [129]. The experiment explores the extent to which the ventriloquist effect (VE) is operational in a real-world environment augmented with spatialised auditory cues delivered via the bone conduction headset. In the following section we provide an overview of existing literature related to spatialised auditory feedback over a bone conduction headset and audio-visual perception. We then outline the methodology and the results we have obtained. Finally, these results are compared to existing studies and future directions for research and application based studies are presented.

3.14 Background

The virtual auditory display (VAD) has existed for close to three decades. From making information delivery in the cockpit of a fighter jet less complicated [58] [115] to helping the visually impaired navigate the real-world using spatialised auditory beacons [130] [131], VADs have been used for across a broad spectrum of applications to provide information to a user. Most of these applications have made use of Head Related Transfer Functions (HRTFs) to process the audio signal being delivered to the user. Using HRTFs to process auditory information allows the VAD to generate a realistic sound stage that mimics our perception of the acoustic environment. The HRTFs used though can vary; encompassing a range of available non-individualised libraries [55] [56], to the much harder to measure, individualised ones [83]. Existing studies demonstrate that the use of individualised HRTFs in VAD greatly enhance the localisation accuracy. Traditionally, VADs used for audio augmented reality applications have made use of headphones or earphones as display devices [83] [17] [132] [84]. While these maintain the privacy of the information delivered, they also tend to isolate the user from the surrounding acoustic environment [14]. Such isolation, while necessary in some cases, may prove to be a hindrance in an AR environment. Attention critical tasks that also require a user to maintain awareness of their surroundings may be affected negatively by the use of headphones in such a scenario. The only faculty capable of ‘keeping an eye out’ for environmental changes when the visual faculty is engaged is audition. The loss or deterioration of this natural ability could prove dangerous in environments where a change in the acoustic environment could be the only indicator of impending danger such. Such environments are routinely encountered by firefighters, urban search and rescue teams and even drivers. Several researchers have attempted to circumvent this drawback of the traditional headphone/earphone based VADs by attaching microphones to them [83] [84] . Their placement on either side, on top of the headphone/earphone unit is thought to be sufficient to provide not just an acceptable reproduction of the ambient acoustic environment, but also good localisation ability of any acoustic event occurring in the environment. While studies suggest that such a system works well, issues related to calibration and deployment in the real-world appear to be major hurdles [83] [84]. Careful calibration taking into account the frequency response of the headphones/earphones must be carried out in order to reproduce an accurate rendering of the acoustic environment. In addition to this, practical operational limitations such as distortions caused by loud sounds resulting in an overload of the circuitry, sounds of eating and wire conducted sounds are just some of the problems encountered by such an augmented reality audio system [84]. We believe that these problems can be addressed by using the bone conduction headset as a VAD leaving the ears free to do what they do best – to perceive the acoustic environment [14]. Studies demonstrate how perception of a space and localisation results can be affected by the disturbances induced in the frequency spectrum of an audio signal by the electroacoustic transducer reproducing it. While research on the subject of binaural spatialisation over a bone conduction headset is relatively limited, there are studies that demonstrate the potential of the bone conduction headset as a VAD [121] [129] [14, 85, 103, 104, 133, 134]. Villegas and Cohen, propose the use of a bone conduction headset as an alternative to the headphones in several virtual and real-world navigation tasks [85]. Walker et al. have demonstrated the usefulness of the bone conduction headset in a real-world navigation task for visually impaired users [14] [114]. Based on these studies and some of our own research [121] [129], we believe that the bone conduction headset has immense potential to be integrated into a bi-modal, wearable interface. The following sub-section covers some literature on the phenomenon of VE and how it relates to AR.

Bi-modal presentation, auditory and visual, naturally lends itself to the AR environment. The addition or overlaying of information on an environment that is occupied by the user is more often than not perceived favourably [135]. Negative perceptions of bimodal AR systems often arise due to ineffective information management. Commonly, such systems appear to be heavily geared towards visual information presentation with the auditory channel being used to communicate some basic system states and possibly as a communication pathway [13]. Research combining the two modalities for effective information delivery in mobile AR interfaces is sparse. Of the studies do exist, most are focused on congruent audio – visual cues in the environment [136] [19].

The Ventriloquist Effect (VE) is a phenomenon wherein spatially separated auditory and visual events appear to be co-located [137]. Despite the spatial separation that exists between the two events, the visual event appears to 'capture' the auditory event [137] [136] [138]. The phenomenon has been studied in some detail in the real-world environment. It is known that there is a certain angle or a range beyond which angular separation between the two events causes them to be perceived as two individual events occurring at different points in space in the horizontal plane. Jackson [61] demonstrated the "visuo-auditory" threshold to be between 20° - 30°. A study by Witkin et al. [60] demonstrated that the angle of deviation of the sound source from the visual cue was 17° for men and 18° for women in the eyes closed condition. This rose dramatically to 28° for men and 38° for women in the eyes open condition. In a video see through AR environment, Kyto et al. have demonstrated VE to be between 32° - 45° [66]. In this experiment we attempt to combine the real-world and augmented approach in order to investigate how and to what extent the ventriloquist effect is operational in an audio augmented real-world environment. Since perception often involves integrating information from different modalities to generate a unified experience [139], it is of utmost importance that the limits of such cross-modal interactions be studied to aid in better interface design.

Two factors set this study apart from previous work in the area.

- Firstly, the ventriloquist effect in the AR domain has not been explored using the bone conduction headset. This makes the study unique since we have obtained results that, to the best of our knowledge, do not exist.
- Secondly, the study has explored the perceptual relationship between a synthetic auditory cue and a visual cue in the real-world. In the literature that we reviewed prior to this study we were only able to find the ventriloquist effect being studied in a real-world environment [60] [61] or a video see-through environment with synthetically generated auditory and visual cues [66].

3.15 Method

3.15.1 Apparatus

The apparatus for this study consisted of a ‘real-world’ analogue that was used to deliver the visual cues, a bone conduction headset to deliver the auditory cues and a tracked ‘flystick’ to record participants’ responses (Figure 3.28).

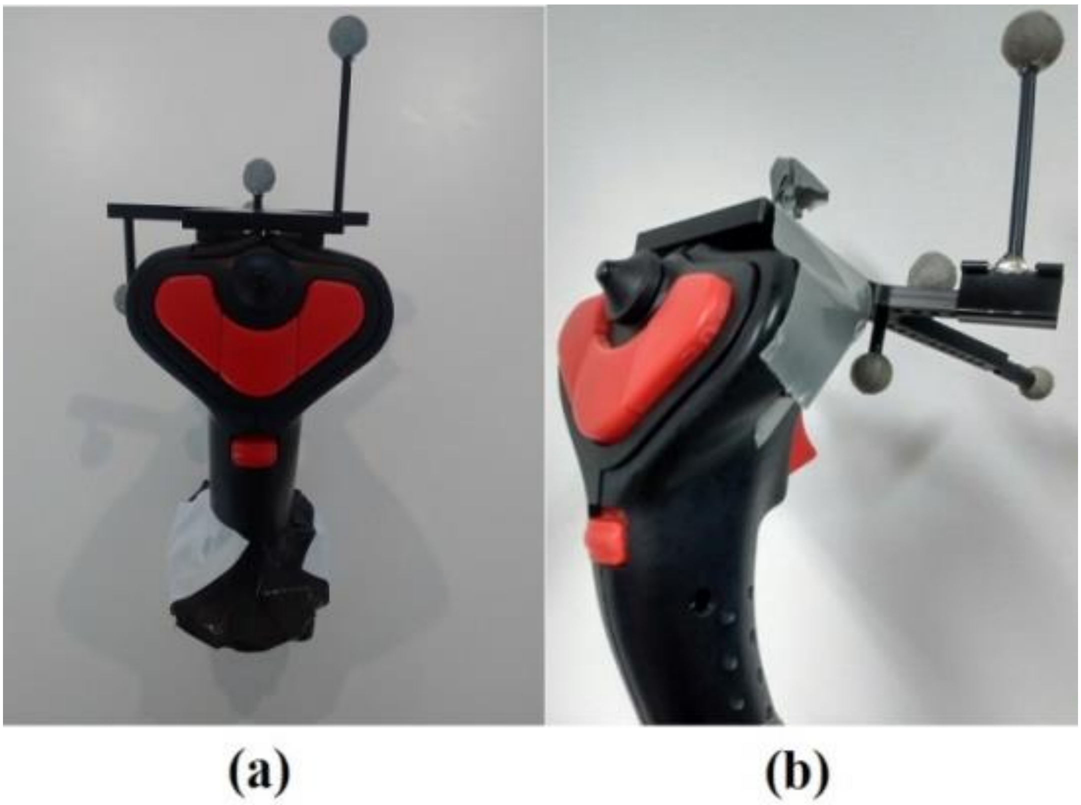


Figure 3.28: Tracked flystick used for the experiment. (a) Front View (b) Side View

The real-world analogue used for this experiment was a set of three screens connected to each other at 60°. The screens measured 2400mm x 1830mm and were mounted 600 mm above the floor. The visual cue was projected on to these screens by three NEC LT265 projectors. A four camera tracking system, ARTTRACK2 [140], mounted on top of the screens (Figure 3.29) was used to acquire positional data from the tracked flystick.

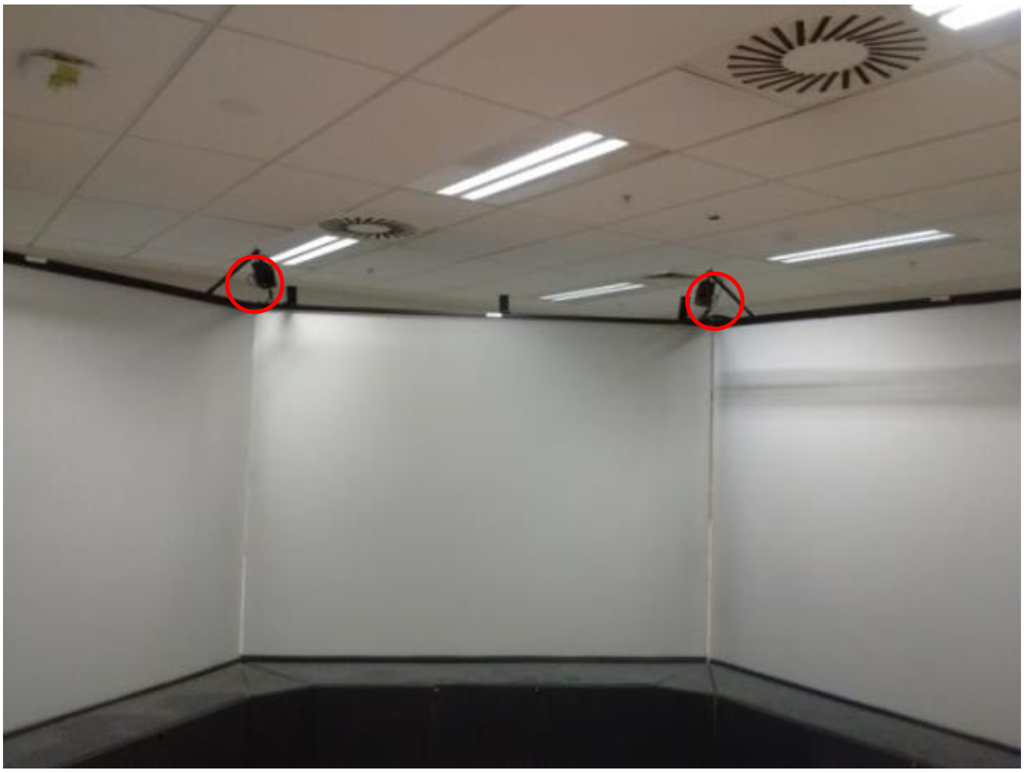


Figure 3.29: The 'Vision Space' system that served as a real-world analogue. Tracking cameras have been circled

Participants wore an Aftershokz Sportz3 bone conduction headset [117] (Figure 3.30). Auditory stimulus used for the experiment was delivered to the BCH via the PC's on-board sound card. The experiment was developed using the Unity3D game engine [115]. Spatialisation of the auditory stimulus was achieved using the 3Dception plugin for Unity3D developed by Two Big Ears [116]. All participant input data was recorded using the buttons on the flystick.

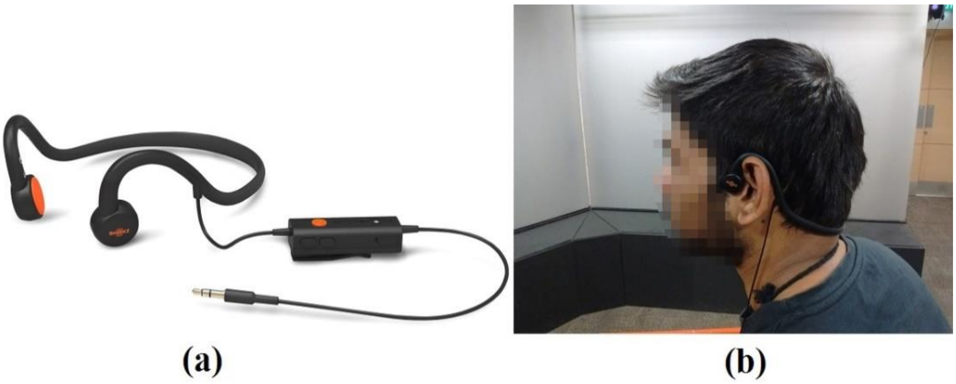


Figure 3.30: (a) Aftershokz Sportz3 [117] (b) Participant wearing the bone conduction headset

3.15.2 Stimuli

Auditory and visual stimuli were presented in this experiment. The auditory stimulus was a one second burst of white noise with a 25ms on-set and off-set ramp. The visual stimulus consisted of a yellow disc approximately 58mm in radius (Figure 3.31). The stimuli were placed at pre-determined positions separated by 5° in a 180° arc around the participant.

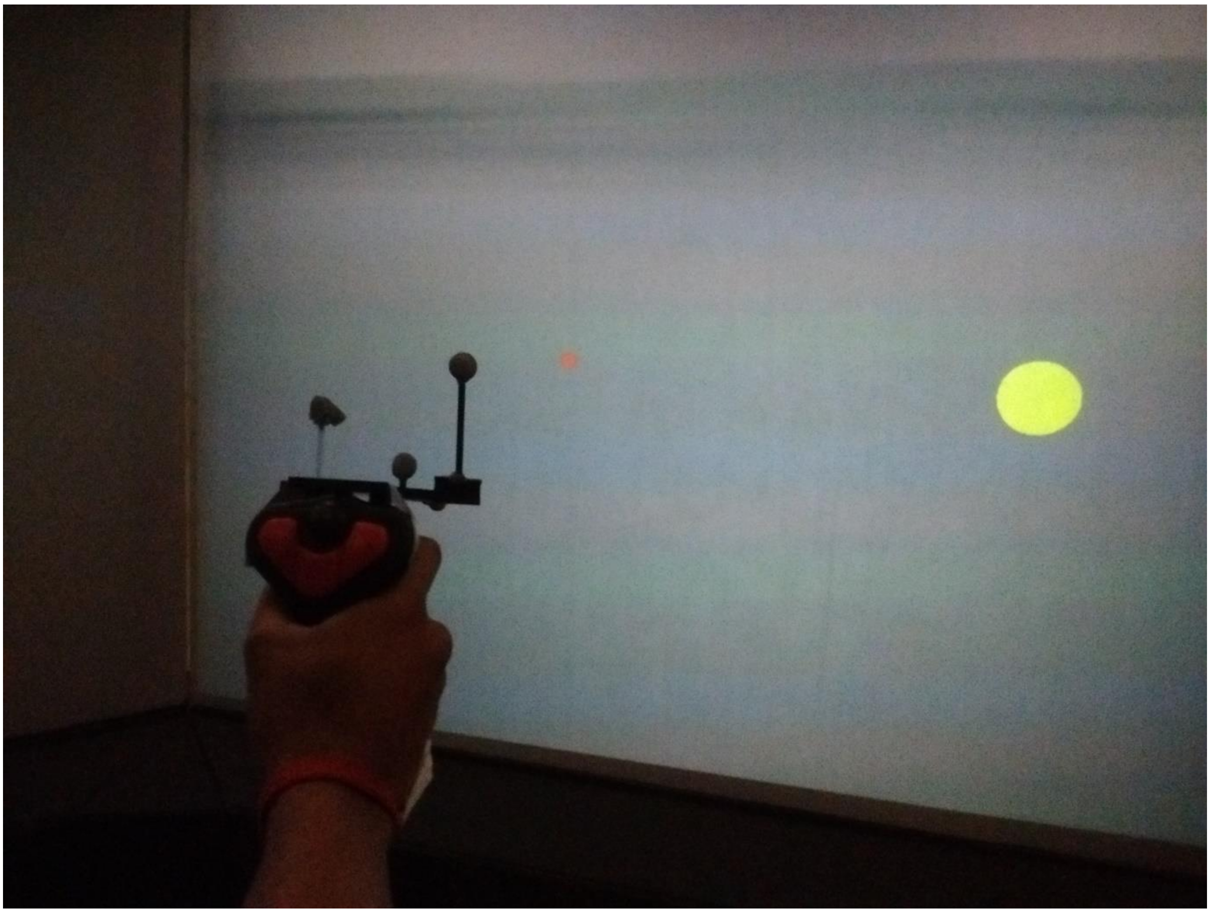


Figure 3.31: Visual stimulus represented by the yellow disc. The red dot indicates where the flystick is being pointed in the real-world

3.15.3 Participants

A total of 15 participants (8 male, 7 female) between the ages of 21 and 37 (Mean: 27.06 years, Std. Dev: 5.4) volunteered to take part in the study. All participants reported normal hearing in both ears. No testing was carried out to verify this since there appears to be no known relation between localisation performance and audiogram results unless the hearing loss is profound [58]. All participants received a $10 shopping voucher as compensation for their efforts.

3.15.4 Procedure

Participants were seated on a rotating chair approximately three metres from the central screen. The position was chosen such that the visual stimuli could be presented anywhere within a 180° arc around the participant. Participants could adjust the height of the chair for comfort. The noise floor of the space in which the experiment was conducted was approximately 50 dBA.

A brief explanation of the experiment and what was expected of the participants was provided. They were then handed the bone conduction headset to put on, followed by the flystick. A short explanation on how to use and indicate responses using the flystick followed. Participants were able to track the flystick's position in the real-world environment using a red dot that was projected onto the screens (Figure 3.13). This allowed the participants to see where the flystick was being pointed in the real-world environment. Participants were asked to press the button on the left if they thought the two stimuli were co-located or 'together'; and the button on the right if they were perceived in different locations (Figure 3.32).

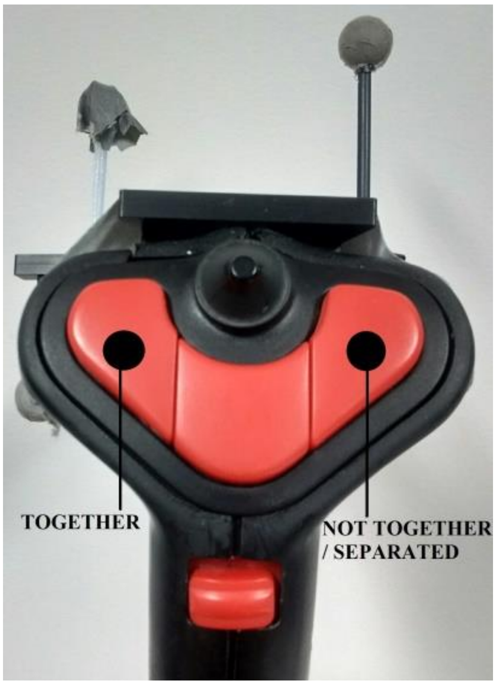


Figure 3.32: Left Button: Co-located stimuli. Right Button: Spatially separated stimuli

Once participants indicated they could begin, 13 practice trials encompassing all the angular separations between the auditory and visual cues were run. This was done to familiarise participants with the response protocol. This period was also used by the participants to adjust the volume on the bone conduction headset if necessary. The main block of trials consisting of 156 repetitions were then run. Each of the 13 angular separations was presented 12 times ensuring that these presentations were equally distributed about the median plane i.e. six trials per separation either side of the median plane. For each trial, the auditory and visual cues were presented simultaneously. The first eleven angular separations used ranged from 0° to 50° in 5° increments. The remaining two were a 60° and 180° separation between the auditory and visual stimuli. All separations were presented in a randomized manner to prevent any learning effects.

The experiment was self-paced and ran for an average of 30 minutes. Participants were required to indicate whether they perceived the auditory and visual cues to be co-located or as separate events. Perception of co-located and separate events was conveyed using the flystick as shown in figure 3.32. If the participants indicated that the auditory and visual events were spatially separated, they were asked to use the flystick to indicate where they perceived the auditory stimuli. Responses were recorded by depressing the trigger on the flystick which logged the angular positon of the flystick in the horizontal plane. All results were tabulated and saved for analysis at a later stage. Externalisation ratings, irrespective of the response to the stimuli were sought at the end of each trial as well. An image (Figure 3.33) was displayed at the end of each trial. Participants were required to point at one of the choices and press the trigger to indicate the level of externalisation they experienced.

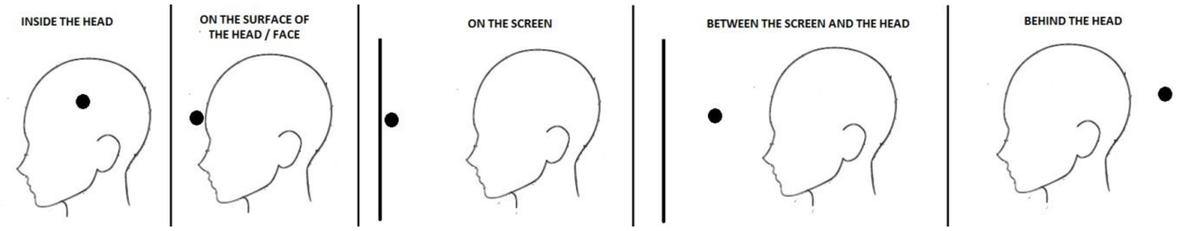

INSIDE THE HEAD
ON THE SURFACE OF THE HEAD / FACE
ON THE SCREEN
BETWEEN THE SCREEN AND THE HEAD
BEHIND THE HEAD

Figure 3.33: Image displayed on screen asking participants to indicate the level of externalisation

3.16 Results

Following the experiment, data from all participants who completed the study (N = 15) was analysed. Recognition rates i.e. identification of the presented stimuli as spatially separated were obtained for all the angular separation presentations (Figure 3.34). This was achieved by dividing the number of times a separation was identified as being spatially separate by the total number of times the separation was presented. A basic analysis of the data demonstrates, unsurprisingly, that the 180° separation between the stimuli was recognised the maximum number of times.

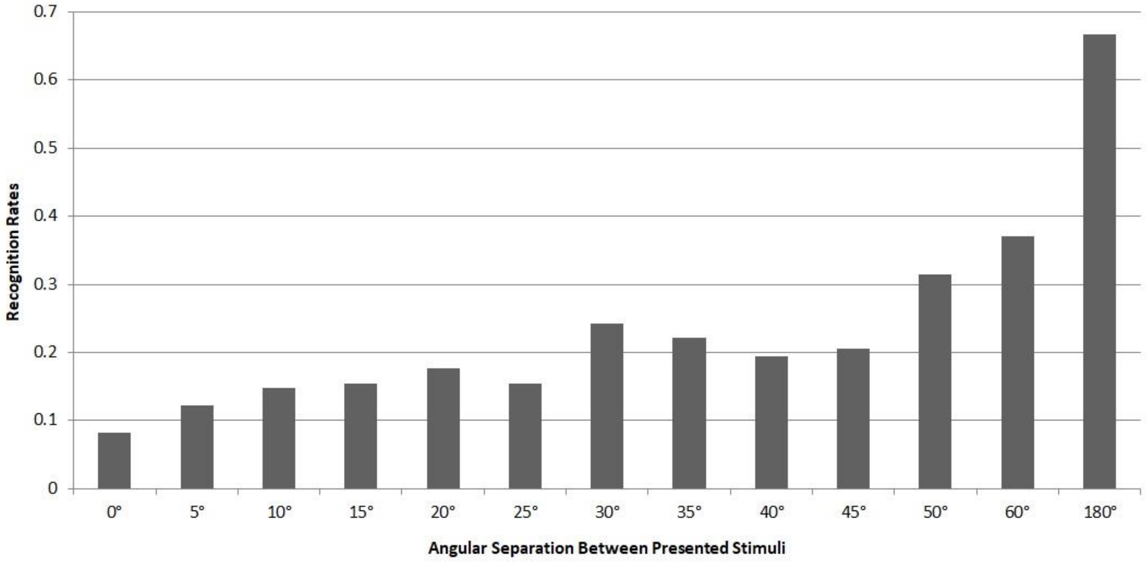


Figure 3.34: Recognition rates for stimuli identified as spatially separate across all presented angular separations between the two stimuli

Further, a paired t-test was run to compute the minimum angle at which the ventriloquist effect appears to breakdown for spatialisation over a bone conduction headset. To enable this, pairwise comparisons between all angular separations with 0° (no separation) were made. This approach was adopted to give us a clear understanding as to where we begin to see the ventriloquist effect breakdown. For the purpose of this study, the failure of the effect was judged to be at the point where the t-test demonstrated consistent significant differences between successive pairs. This approach was also used in the first experiment [121] described in this chapter to determine the minimum discernible angular difference between two successively spatialised sound sources.

The first significant difference between 0° and a chosen angular separation is seen for the 0° – 20° pair (t (14) = -2.466, p = 0.027). This is followed by a pair that does not demonstrate any significant difference. As per the criteria laid out earlier in this section, a consistent significant difference between successive pairs is what is required for an angular difference to be considered as the point at which the ventriloquist effect breaks down. This is seen for the 0° - 30° (t (14) = -3.384, p = 0.004) onwards. An increase in the mean from 0.0827 ± 0.095 to 0.2433 ± 0.21 (p = 0.004) demonstrates that an increased angular difference between the two stimuli, and a consequent percept of two spatially separated sources, leads to the breakdown of the ventriloquist effect. All pairs following 0° – 30° show a significant difference. Based on these results, it appears that at least a 30° angular separation between a visual and auditory stimulus, irrespective of their position in a 180° arc around the participant, is required for material spatialised over a bone conduction headset for them to register as spatially separated.

A further investigation to look at the issue of the ventriloquist effect based on the orientation of the stimuli around the head was also carried out. To facilitate this, the 180° arc in front of the participants was divided into three sections; one 90° (-45° to +45°) arc in front of the participant and two 45° sections on either side (-45° to -90° and +45° to +90°). For the purpose of the analysis, we divided the sections and stimuli locations in the following manner:

- Both cues between -45° and 45°.
- Visual cue between -45° and 45°.
- Auditory cue between -45° and 45°.
- Both cues outside the -45° to +45° range.
- 180° difference i.e. diametrically opposite at -90° and +90°.

Following this, section-wise recognition rates were tabulated. A repeated measures ANOVA was carried out according to the target cues distribution criteria laid out above. Since the data violated Mauchly’s test of sphericity, the following analysis utilised the Greenhouse-Geisser correction. A statistically significant difference (F (1.386, 19.405) = 37.03, p < 0.001) between the specified sections was observed. Post-hoc testing using the Bonferroni correction demonstrated a significant decrease in the ability of participants to recognise the stimuli as separated when at least one of the cues was outside the ±45° range. Statistically significant differences (p < 0.001) were observed between the data obtained for both cues between ±45°

versus that when only the visual cue is between ±45°. No significant differences (p= 1) are observed between conditions where at least one of the cues are outside the ±45° range.

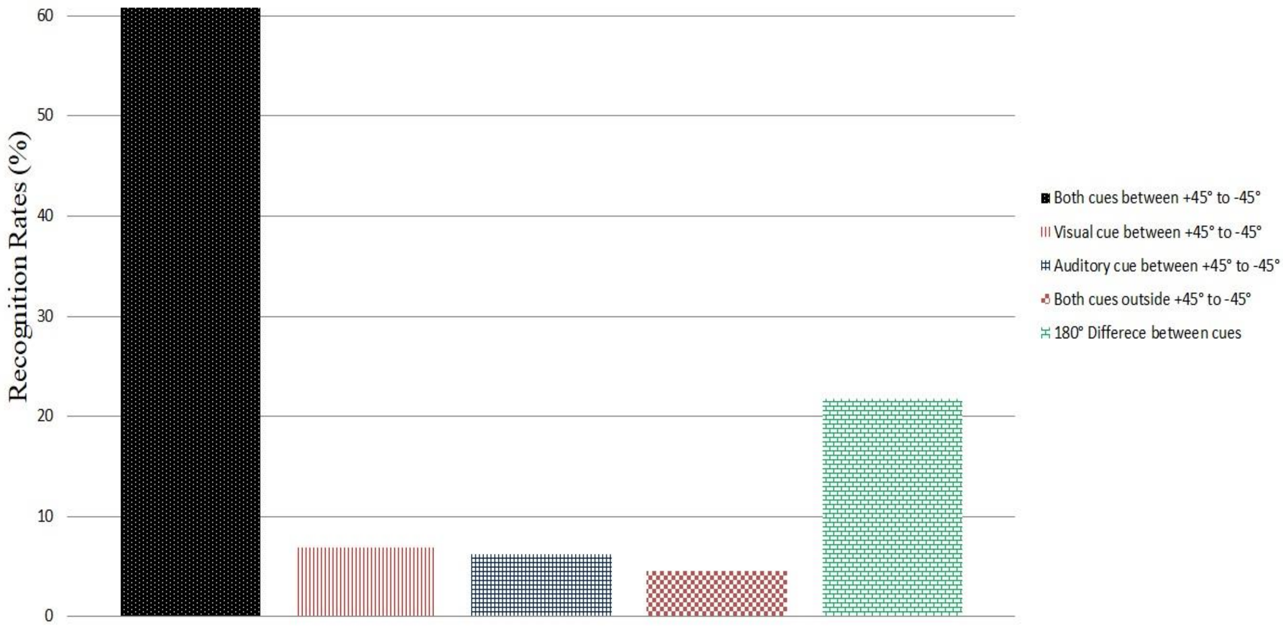


Figure 3.35: Recognition rates (stimuli identified as being spatially separated) for various cue distribution types

From figure 3.35 we can see that the recognition of the presented cues as two separate events occurs most frequently directly in front of the observer. Participants appeared to be best at discriminating the two cues as separate events when both were presented within the 90° arc in front of them. This ability appeared to diminish drastically when at least one of the cues was presented outside the arc. Participants appeared to localise the auditory cue at the position of visual cue. Discrimination between the two cues was significantly lower than expected even in the condition in which they are presented in diametrically opposite locations. It is possible that the presence of the visual cue in the participants' peripheral vision, as a result of not fixing the position of the head via clamps or a bite bar, attracted their attention leading them to localise the auditory cue to the same point.

In addition to these results, we have also obtained externalisation ratings for each trial. Figure 3.36 shows the percentages accrued for each rating type. Percentages were obtained by dividing the total number of trials across all participants by the total number of externalisation ratings for each rating type. Percentages for the ratings have been calculated for cues across the following distributions:

- Overall ratings: These take into account the ratings collapsed across all trials.
- Both cues within +45° and -45°: These ratings only account for all trials that fall within the specified range. It must be noted that this range falls within an arc subtending 90° directly in front of the participant. Percentages for these ratings were obtained by dividing the total number of trials within this range by the total number of externalisation ratings for each rating type. The same system for calculating percentages was applied for the following two distributions.
- Both cues outside +45° and -45°.
- Both cues separated by 180°.

Figure 3.36 clearly demonstrates that participants displayed a tendency to localise the presented auditory cue on the screen. Approximately 30% of the total trials appear to have elicited this response. This percentage falls marginally to just over 27.5% when only analysing cues that were both either within a +45° to -45° range or outside of it.

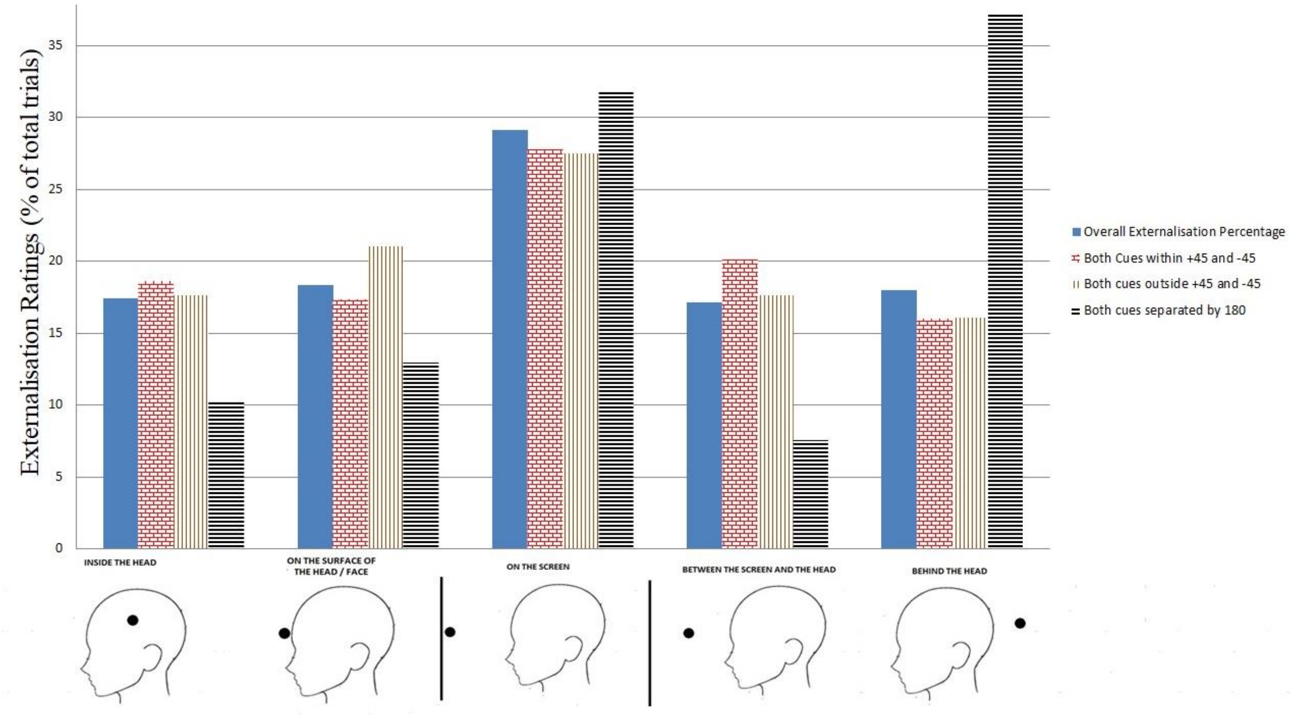


Figure 3.36: Externalisation rating percentages collapsed across all trials, when both cues were presented within and outside the +45° to -45° area and when both cues were presented diametrically opposite (180°) each other

A one-way ANOVA of the overall results shows no significant difference between the externalisation ratings (F (4, 30) = 0.73, p = 0.58). This result appears to suggest that the participants were unable to categorically decide on the level of externalisation based on the auditory cue. A similar analysis between the three areas listed earlier demonstrates that a significant difference does exist between them (F (2, 12) = 19.48, p < 0.001). Further pairwise comparisons made using the Bonferroni correction reveals significant differences between cues separated by 180° and those that are presented within +45° and -45° (p < 0.001) and outside this range (p < 0.001). Figure 3.36 also demonstrates that in close to 38% of the trials where cues were presented at diametrically opposite directions, participants reported the auditory cues as being located behind the head. This could be explained by reversals in auditory perception that are experienced in VADs that employ non-individualised HRTFs [55] [56] [141].

In addition to this, both commonly reported externalisation ratings can be explained considering the effect of the visual cue. We have already seen how a visual cue can 'capture' a sound source and make it seem like the auditory and visual sources are a singular entity. It appears that the same phenomenon is taking place here, but with the perspective of auditory distance. In our opinion, the visual cue appears to be dictating the level of externalisation reported by participants. Additionally, we also feel that not randomising the positions of externalisation ratings for each trial may have contributed to the externalisation reported by participants. In the case of the cues that were separated by 180°, it appears that the failure to see a visual cue within the frontal visual field results in the auditory system assuming complete responsibility for localisation. This results in the auditory cue being perceived as being at the back of head due to the use of non-individualised HRTFs as described earlier in this section.

3.17 Discussion

We have been able to demonstrate that the ventriloquist effect is operational even for binaurally spatialised auditory cues over a bone conduction headset. While we expected the audio AR system demonstrated here to exhibit higher 'tolerances' prior to the breakdown of the effect, the results seem to indicate otherwise. The phenomenon of the ventriloquist effect is seen to persist up to angular difference of 20° between the auditory and visual cues, with a recurring difference only occurring from an angular separation of 30° onwards. It is for this reason we can suggest a range between 20° - 30° to be the angles within which the ventriloquist effect breaks down. By this we mean that the specified range is the minimum angular difference between a visual and auditory cue required for them to be perceived as spatially separate events. The 20° limit for VE appears to exist for the area directly in front of the observer; an arc of about 10° in front of the observer. However, in this case we have defined the area in front as the 90° arc directly in front of the participant. Therefore, looking at figure 3.35 we can see that the percentage of recognition rates was quite high (approximately 60%) when both cues were presented within this arc. There is a significant fall-off in the recognition rates (less than 10%) when either the auditory or visual cue was presented outside the 90° arc. A likely explanation for these results is the manner in which visual and auditory perception function. Both visual and auditory perception have shown to function best within in a few degrees on either side of the median plane [63] [105]. This increased resolution in auditory and visual perception directly in front of the head could account for the higher recognition rates when both cues were presented within the 90° arc in front of the user. Conversely, the significant decrease in recognition rates outside this region could be attributed to the visual cue ‘capturing’ the auditory cue. Another possible explanation for the reduction in recognition rates could be a conflict between the actual perceptions of the two stimuli versus ‘learned perceptions’. This means that when presented with conflicting audio-visual stimuli, a participant was likely to co-locate both stimuli in absence of a clear distinction between their spatial locations. While this form of perceptual conflict has been covered in psychoacoustic studies [141], it appears little is known about the phenomenon when comparing audio-visual stimuli. For the purpose of this experiment, it appears that participants defaulted to the location of the visual stimulus when such a conflict was encountered.

Our results are in line with the real-world outcome as demonstrated by Jackson [61]. They also fall well within the range (32° - 45°) demonstrated by Kyto et al. [66] They are particularly encouraging when compared with the study run by Kyto et al., since their experiment utilised auditory cues reproduced over headphones. The use of non-individualised HRTFs in both studies lends a greater degree of ecological validity to the results. This is especially true with our study since we have used commercially available, ‘off-the-shelf’ non-individualised HRTFs to process the auditory cues. Our augmented audio reality system more accurately represents what audio based wearable systems of the future are likely to resemble.

3.18 Conclusion

We have been able to study the limits of the ventriloquist effect in audio augmented reality utilising a bone conduction headset and a real-world analogue. Results obtained here demonstrate a good level of consistency in comparison with similar studies. These are encouraging and need to be investigated further. Such investigations should include the use of vision based wearable devices such as Google Glass in collaboration with a bone conduction headset [129]. In addition to these, it is also worth looking at how visual cues affect distance perception and externalisation of an auditory source. As we have mentioned in the previous section, our skewed externalisation ratings appear to demonstrate that there is a significant effect of vision on the perception of distance. Perhaps, there could be a range within which this proposed effect of distance is effective; similar to the ventriloquist effect, but affecting the perception of distance. Future studies must also look at the relationship between auditory and visual cues in order to study how learned perception versus actual perception of stimuli can affect the design of wearable devices. Such studies will help establish the usability of such devices in the real-world. In the next chapter we take a look at two experiments conducted based on the results obtained here. These provide some indication of how a bi-modal wearable interface would perform in real-world scenarios.

3.19 Summary

This chapter has presented three psychoacoustic studies we have run to assess the feasibility of using a bone conduction headset as part of a wearable, bi-modal interface. Important aspects of auditory perception such as localisation accuracy and externalisation were explored. Results from these experiments demonstrate that a bone conduction headset affords good externalisation and localisation accuracy in the horizontal plane. Its performance in the vertical plane is not nearly as good. The results from this study lead us to conclude that outside a ‘window’ of about 20° - 25° either side of the horizontal plane passing through the ears spatialisation, localisation and externalisation break down leading to severe inaccuracies in auditory perception. Accordingly, a recommendation stemming from these results would be to avoid using auditory cues to draw attention to points of interest in the environment that occupy a significant elevation.

Never-the-less, our third study demonstrates that audio-visual congruency for a simultaneously presented synthetic auditory cue and real-world visual cue is robust.

Statistically, the breakdown of the VE has been demonstrated to occur for approximately a 30° difference between an auditory and visual cue. Taking this into account, we believe that an auditory cue in the horizontal plane could be used to re-direct attention to the point of visual interest in the vertical plane. This relies on the premise that the visual cue will 'capture' the auditory cue provided the visual cue falls within the 40° - 50° window described earlier in Section 3.10. This window, not surprisingly, corresponds to the approximate vertical field of view for human beings as shown in Chapter 2.

In the following chapter we describe two application based studies that we have conducted using a wearable, bi-modal prototype. These studies apply the findings from our psychoacoustic experiments to simulated real-world scenarios. We have used the results from our studies in Chapter 3 to design and implement the auditory and visual cues to assess how these can be delivered via a wearable, bi-modal interface in the real-world.

Chapter IV

4. APPLICATION BASED STUDIES

The previous chapter looked at the psychoacoustic studies that were conducted to evaluate the operational characteristics of the bone conduction headset. We examined the localisation accuracy and externalisation afforded by the bone conduction headset for binaurally spatialised sound reproduced over it. The interaction between, and effects of auditory, and visual cues on the perception of objects in the real-world environment were also explored. Based on the results of these experiments we have been able to demonstrate that the bone conduction headset holds an enormous potential to be utilised as an information display device as part of a bi-modal wearable interface.

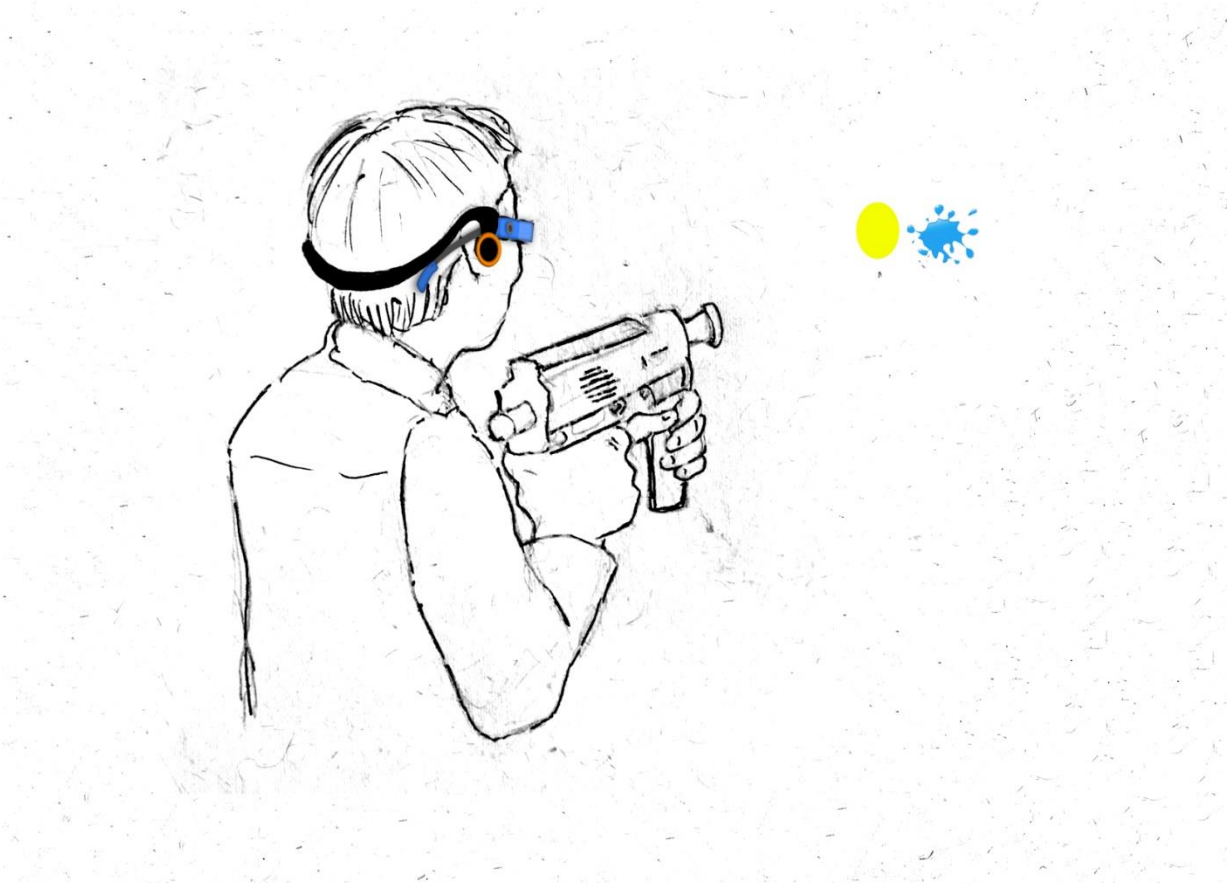
In order to explore this possibility and evaluate how it may perform in real-world or close to real-world conditions, experiments which mimic such conditions must be carried out. Testing out a prototype device in some of the potential scenarios it can be used in is vital to establishing its usefulness and practicality. To this end, the following studies in this chapter lay out two potential scenarios in which our prototype device can be used in. The first one mimics a scenario that first responders such as fire service personnel, police and disaster management crews may find themselves in. We have achieved this by designing a divided attention task that forces the user to switch between a primary task, in the far field, and a secondary task, in the near field. Auditory or visual cues or a combination of the two were delivered at random points during this task to let the users to let them know that a point of interest required their attention in the far field. Reaction times to these cues were measured to determine how efficiently auditory and/or visual cues or a combination of the two were at attracting and then redirecting the user's attention to these points of interest.

For the second study, we have chosen to implement a seemingly mundane domestic scenario. We tend to leave objects of importance such as keys, mp3 players and other small items around the house without much thought. Finding these items when the need arises can sometimes prove difficult and frustrating. This experiment approaches the problem in an innovative manner utilising visual cues and spatialised auditory beacons. These cues were delivered to the user via the Google Glass and a bone conduction headset. Results from this study were very promising and demonstrated that a wearable device incorporating auditory and visual feedback can be used to simplify such tasks, at least within the home environment. While similar technologies such as the TrackR Bravo [142] are currently available in the market, we believe our prototype adopts an innovative and novel approach that a user may find more intuitive.

The remainder of this chapter describes in detail the implementation and results obtained from the studies mentioned above. Our results demonstrate a good level of usability and localisation accuracy afforded by the prototype device.

Study IV

USING BONE CONDUCTION BASED SPATIAL AUDITORY DISPLAY AS PART OF A WEARABLE INTERFACE



Papers resulting from this study have been published at the *22nd International Conference on Auditory Display (ICAD-2016)* held in Canberra, Australia between 2nd and 8th July 2017 and the *2016 Human Factors and Ergonomic Society (HFES) Annual Meeting* held in Washington D.C., U.S.A. between 9th and 13th October 2016. Both papers were co-authored with Matt Ward.

The first experiment evaluating the wearable bi-modal interface prototype explored how such a device would perform in a dangerous and attention critical environment. The aim of this study was to analyse how efficiently such a device was at redirecting attention. There are many scenarios where it might be important to direct the user's attention. For example, first responders in a natural disaster scenario are required to tend to the injured as well as maintain awareness of the environment around them. In addition to these tasks, monitoring and responding to communications between various parties is also critical. In such a scenario, the user's attention will be divided between many tasks, and so it is important to direct their attention to possibly dangerous events occurring in their vicinity. This study attempts to mimic such a scenario and evaluate users' performance when equipped with a wearable, bi-modal interface.

4.1 Introduction

One of the most common ways of interacting with mobile devices is via the visual interface. The inherent nature of current mobile devices makes visual interaction a necessity to retrieve almost any information. In the recent past, wearable mobile devices such as Google Glass [72] [71] and the Recon Jet [111] have attempted to simplify this interaction by moving the screens from our pockets and onto our faces. Unfortunately, this has failed to address the issues that visual interaction itself poses. Due to the inordinate number of data streams that compete for our visual attention and screen space on our devices, it becomes almost impossible to sift through the vast amount of information being presented to us simultaneously. As users we find it extremely stressful to divide our attention between these multiple streams, causing sensory and cognitive overload [13].

Addressing this problem of wearable displays has been one of the key focus areas of this thesis. Wearable displays with their severely limited screen space and a constant demand on the user's visual attention make for potentially unsafe devices. Attention critical tasks such as driving, search and rescue etc. may be adversely affected by the use of such wearable devices [3] [4] [5]. This presents us with a set of unique challenges; (1) information presentation without overloading the user and (2) unobtrusive information delivery requiring minimum attention from a user perspective. Following up on the psychoacoustic studies described in Chapter 3, this study expands on the results obtained from the experiments to explore practical means of utilising the bone conduction headset as part of a wearable, bi-modal interface. We analyse the hypothesise that presenting information using two modalities results in a lower cognitive load and increased task efficiency.

4.2 Background

Wearable spatial auditory interfaces have been covered in detail in Chapters 2 and 3. We have also looked at the problems with interfaces that employ headphones or earphones as means to deliver auditory information. Isolation from the surrounding acoustic environment [143] [144] and other operational factors [83] [84] limit the usability of such auditory displays as practical alternatives to a visual display. Recent developments in digital signal processing have seen the introduction of more sophisticated Audio Augmented Reality (AAR) devices such as the Here Active Listening system [90] [10] (Figure 4.1c) and Bragi Dash Pro [89] (Figure 4.1a). Both these devices need to be inserted into the ear canal and rely on an external, on-board microphone input to reproduce the ambient environment around the user. Their sizable form factor occludes the pinna (Figure 4.1b); meaning signals reaching the microphone are not ‘naturally’ filtered in the frequency domain. From existing research we know that this can cause a large number of front-back confusions and a significantly diminished ability to localise in the vertical plane [50].

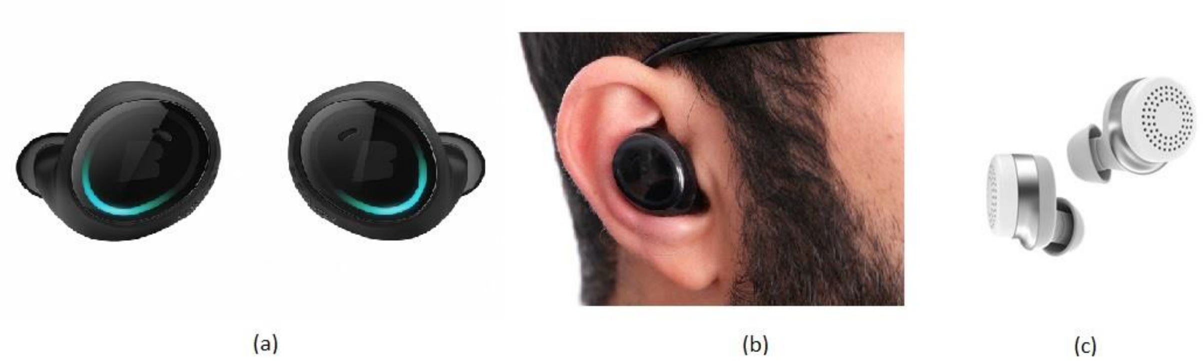

(a) (b) (c)

Figure 4.1: (a) Bragi Dash Pro [89] (b) Form factor of a hearable (c) Doppler Labs Here Active Listening System [90]

Such issues with auditory perception are unacceptable, especially when viewed in light of attention critical environments where errors in auditory perception can prove deadly. Leaving the ears open to the natural acoustic environment is the safest and most prudent option in such cases. It is for this reason that we have championed the cause of the bone conduction headset through this thesis. Previous studies have demonstrated that bone conduction headsets have an enormous potential to be used as an auditory display [14] [104] [114] [15] [107] [16].

While the use of the bone conduction headset has been primarily restricted to its implementation as an auditory display for the visually impaired [14] [114] [15], some studies demonstrate its effectiveness even for sighted users [103] [85]. However, besides the work done by Valjamae et al. [103] and Villegas and Cohen [85], we are unaware of any other studies that incorporate the use of a bone conduction headset as part of a wearable interface for users with normal vision. Considering this lack of research in the area, we hope to demonstrate the practical utility of incorporating the bone conduction headset into a wearable interface. In the following sections we describe our use of the bone conduction headset as a spatial auditory display incorporated into a wearable, bi-modal interface. We then present a user study conducted to evaluate the use of audio-visual cues in a visual search task. The ability to reorient attention with the use of these cues is explored. Existing studies demonstrate significant performance improvements for a visual search task when auditory cues are used [58] [115] [145] [146]. Being able to exploit the mechanism of auditory and visual perception is likely to result in a more efficient and usable interface. Our study tests this premise and presents clear performance indicators for the various cue types such as the time taken for the on-set of head motion in direction of the target once the cues have been presented.

4.3 Method

4.3.1 Apparatus

The apparatus used for the study can be divided into three categories; ‘real-world’ analogue (for far domain stimuli presentation) and tracking equipment, handheld and worn tracked devices used by the participants and a bone conduction headset used to deliver auditory cues.

The real-world analogue used for this experiment was a set of three screens connected to each other at 60°. The screens measured 2400mm x 1830mm (74.65° x 59.53° visual angle) and were mounted 600 mm above the floor. Images were projected on to these screens by three NEC LT265 projectors. The tracking system used for the study comprised of four

ARTTRACK2 cameras mounted on top of the screens, paired with the DTrack software [147] (Figure 4.2). The cameras are capable of tracking objects up to 4.5m. For detailed specifications of the camera see [140]. Equipment used by participants consisted of a Recon Jet [111], a wearable ‘smart glass’ and the Steradian S-7X laser tag gun [148] (Figure 4.3). Both devices had retro-reflective markers affixed to them to allow their positions to be tracked during the experiment. The laser tag gun was modified such that depressing the trigger on the gun allowed the participant to ‘shoot’ targets that were displayed on the screens. This was achieved by connecting a pair of leads attached to the trigger inside the gun to the circuit board of a mouse. The Recon Jet has a widescreen 16:9 WQVGA display with images on it set to appear as they would on 30-inch HD display at 7 feet. The field of view afforded by the display is approximately 16° [149] For more detailed technical specification see [111]. The Recon Jet was connected wirelessly through a router. Tracking data was transferred to the PC using the VRPN software [150] [151].

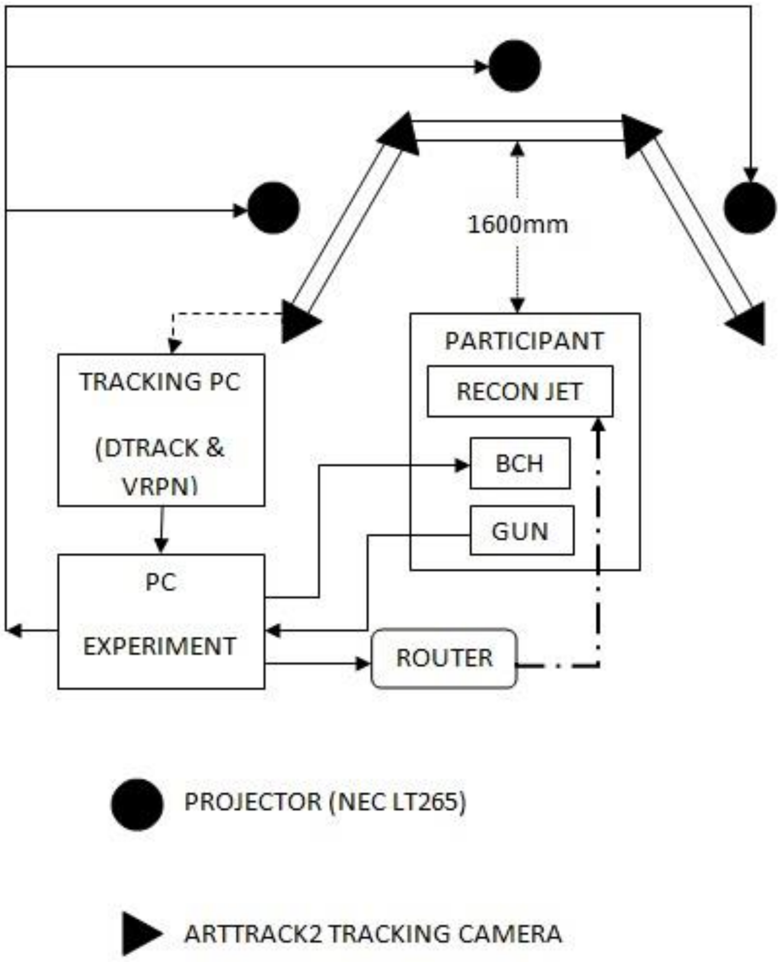


Figure 4.2: Block diagram of the experimental setup

Participants also wore a pair of bone conduction headsets (Aftershokz Sportz3) [117]. Auditory stimuli for the experiment were reproduced over the bone conduction headset. The auditory stimuli were delivered to the BCH via the PC’s on-board sound card.

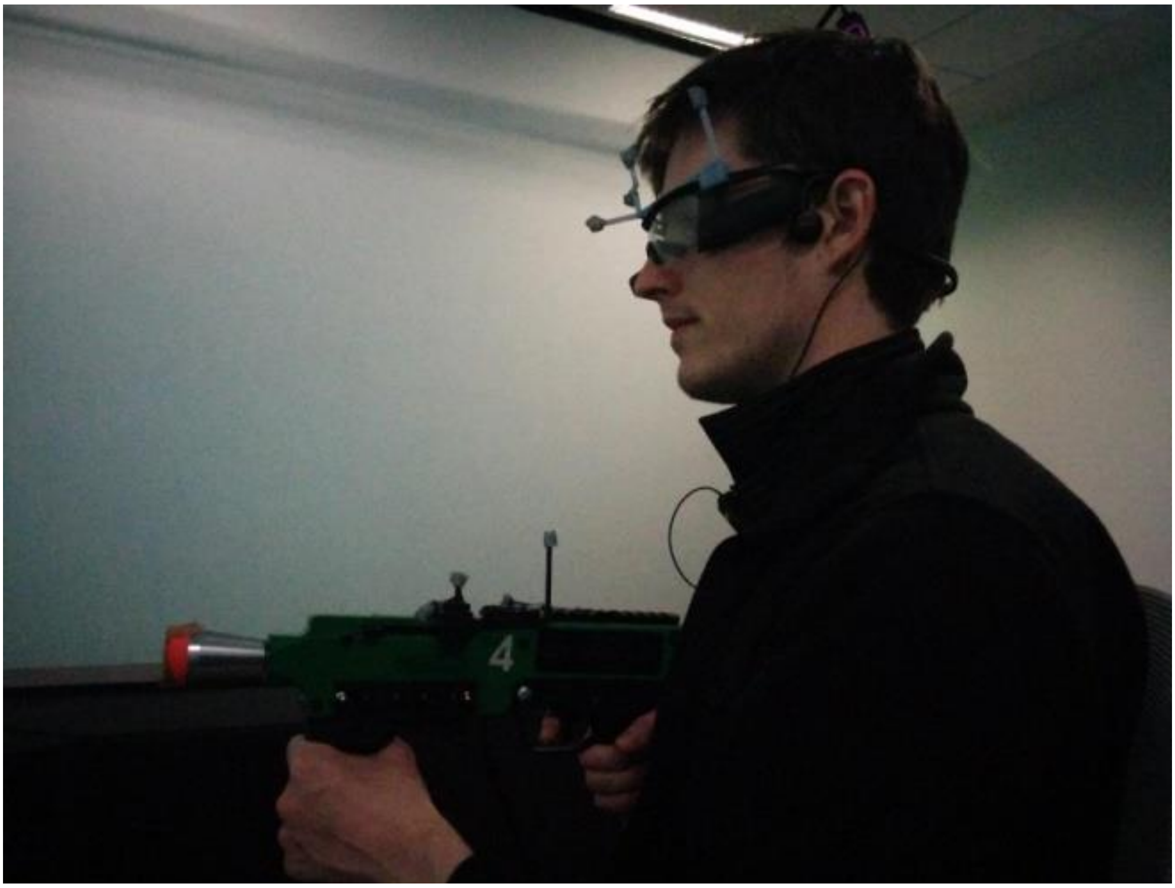


Figure 4.3: Participant holding the laser tag gun with markers affixed on top to allow the position of the gun to be tracked. Also seen in the picture are the Recon Jet with markers for tracking, and the Bone Conduction Headset

4.3.2. Stimuli

Participants were presented auditory and visual stimuli for the experiment. A detailed explanation of the stimuli is given in the following sections.

- Visual Stimuli

  Visual stimuli were delivered on the projection screens representing the far field, and the Recon Jet display. Stimuli displayed on the projection screens were targets that appeared at random intervals during the experiment, and a string of numbers at the bottom of the centre screen. Targets consisted of yellow discs of approximately 58mm radius that turned blue when shot and appeared at predefined positions of ±50° and ±100° (Figure 4.4). The targets were positioned at the centre of the screens. The numerical string used a black Arial typeface of 65mm size positioned in the horizontal centre and approximately 690mm below the vertical midpoint of the centre screen.

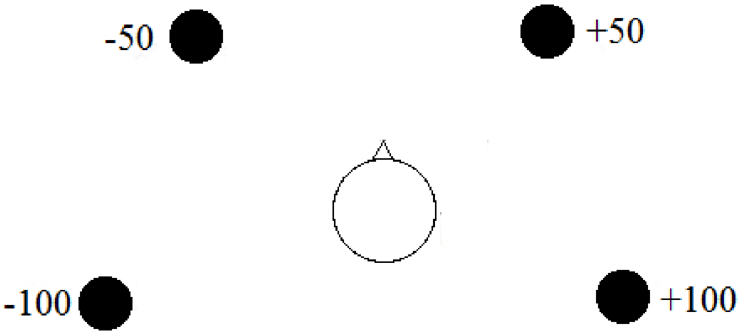
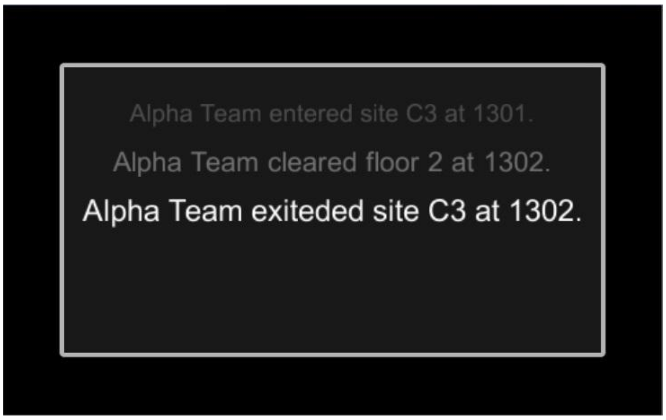


Figure 4.4: Target Positions

Text displayed on the Recon Jet used a white Arial type face and was positioned in the centre of the screen (Figure 4.5). Preceding messages were listed above in grey. Visual interrupt signals delivered on the Recon Jet consisted of static cues, pursuit visual cues and a blank screen. The static visual cues consisted of white arrows 1.3° in width and 6.5° in length (see figure 4.6). The arrows were angled at 40° for targets appearing at ±50° and 80° for targets at ±100°. Pursuit visual stimuli caused all objects on the screen to move in the direction of the target at 16.2° per second.



Alpha Team entered site C3 at 1301.

Alpha Team cleared floor 2 at 1302.

Alpha Team exiteded site C3 at 1302.

Figure 4.5: Messages displayed on the Recon Jet screen

In addition to the messages on the Recon Jet and targets projected on the screens, the participant also saw two smaller ‘dots’ on the screens. These dots represented the position of the participant’s head (yellow dot) and the position of the gun (blue dot). Both moved around on the screen as the participant rotated along the horizontal arc on which the targets appeared.

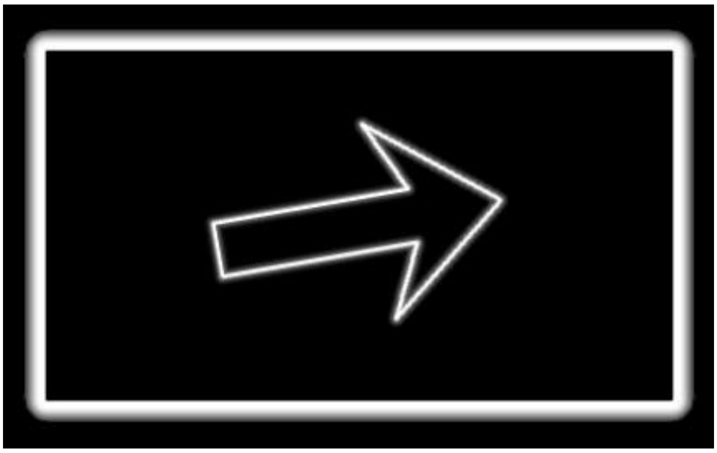


Figure 4.6: Visual Cue – White Arrow

- Auditory Stimuli

  Auditory stimuli consisted of a 1 second alarm sound or ping (25ms on set and offset rate). The same sound was used for two of the three types of auditory cues that were delivered. The alarm tone was presented either as a static sound or a binaurally spatialised dynamic audio cue moving in the direction of the target. The binaurally spatialised, dynamic cue simulated the motion of the alarm from the participant’s position towards the target on the screen. The cue was designed in accordance with alarm design guidelines prescribed by Walker and Kramer in [152]. The duration and level of the auditory cue were chosen to represent those used by previous researchers [14] [113] [114] [122] [120]. Despite studies demonstrating that wideband noise is easier to localise [14] [114] [122] [42] [123] than most other forms of stimuli, we chose to use a ping for its aesthetic appeal [153]. The third auditory cue consisted only of silence. The static auditory cue was delivered at approximately 70dBA. The dynamic cue on initiation had approximately the same loudness level, but decreased as the cue moved towards the target. A logarithmic fall off with the addition of the Doppler Effect was modelled to replicate real-world auditory percepts.

The visual and auditory cues used here are analogous in that they encompass similar perceptual characteristics, but in different domains (Table 1).

Table 1: A list of auditory and visual cues used in the experiment. A total of 9 cues encompassing a combination of all the cues above were presented to the participants

AUDITORY CUES		VISUAL CUES	
S0	No sound (Silence)	V0	Blank Screen
S1	Static alarm	V1	Static (Arrows pointing in the direction of the target)
S2	Binaurally spatialised, dynamic alarm	V2	Pursuit visual cue

The experiment was designed and built within the Unity3D [115] environment. Binaural spatialisation of the auditory cue was achieved using the 3Dception Binaural Engine plug-in for Unity developed by Two Big Ears [116]. We have chosen to adopt the use of a plug-in versus the traditional approach of using individualised HRTFs or HRTF libraries since we believe this lends a greater degree of ecological validity to the study. The plug-in was chosen after an extensive phase of testing and comparisons with existing binaural engines.

4.3.3 Participants

Thirty participants (20 male, 10 female) between the ages of 18 and 34 (Mean: 24, Std. Dev: 4.3) volunteered to take part in the study. Participants reported normal hearing in both ears. No testing was carried out to verify their claims of having normal hearing since there appears to be no known relation between localisation performance and audiogram results unless the hearing loss is profound [58]. Five of the participants had prior experience with binaural spatialisation over a bone conduction headset, having taken part in one or all of the experiments described in chapter 3. All participants were compensated with NZD $20 shopping vouchers for their efforts.

4.3.4 Procedure

Participants were seated on a rotating chair 1600mm from the central screen (Figure 4.2). The position was situated approximately on the normal from the central screen such that targets could be presented anywhere on a 200° horizontal arc. Participants could adjust the height of the chair for comfort.

Participants were then handed the Recon Jet and bone conduction headset to put on. If required, they were helped positioning the screen of the Recon Jet so that the text displayed on the screen appeared clear. The bone conduction headset was put on such that the drivers of the headset sat in front of the ears. Participants were also handed the laser tag gun. Following this, a short calibration process was run. For the calibration process, participants were asked to look at a red dot that appeared directly in front of them on the central screen. Following this participants were required to rotate their heads through approximately 180°. This was done to ensure that participants had a full range of motion that allowed them to reach targets at ±100°, and to verify if tracking information was being gathered in the right manner. Following this, several messages were displayed on the Recon Jet and the bottom of the central screen. This ensured that participants were able to read text appearing on the Recon Jet and the main screen with a minimal movement of the head. Following the calibration process, six practice trials were conducted. These trials allowed participants to see the different audio-visual cues that could be presented to them over the course of the experiment.

The main experiment was divided into three blocks. Each block was followed by a five-minute break. During each block, participants were instructed to read aloud number strings appearing on the central screen and messages appearing on the Recon Jet. Messages displayed on the Recon Jet were chosen at random from one of three structure types; [Alpha] team entered site [C][37] at time [1407], [Alpha] cleared floor [2] at time [1407], or [Alpha] team exited site [C][37] at time [1407] (Figure 4.5).

During a third of the trials for which messages were displayed on the Recon Jet, a cue would interrupt the participant one second after the message's onset. Simultaneously, a target would appear at ±50° or ±100°. Participants were required to 'shoot' or mark the target using the modified laser tag gun as quickly as possible (see Figure 4.7). Once the target had been shot, participants returned their gaze to the central screen, and the alternating display of number stings on the central screen and messages on the Recon Jet resumed. Within each block there was one target event trial for each combination of visual and audio cues for each position for a total of 36 target events (9 event types x 4 target locations) and 72 non-event messages. The experiment took on average 65 minutes to complete.

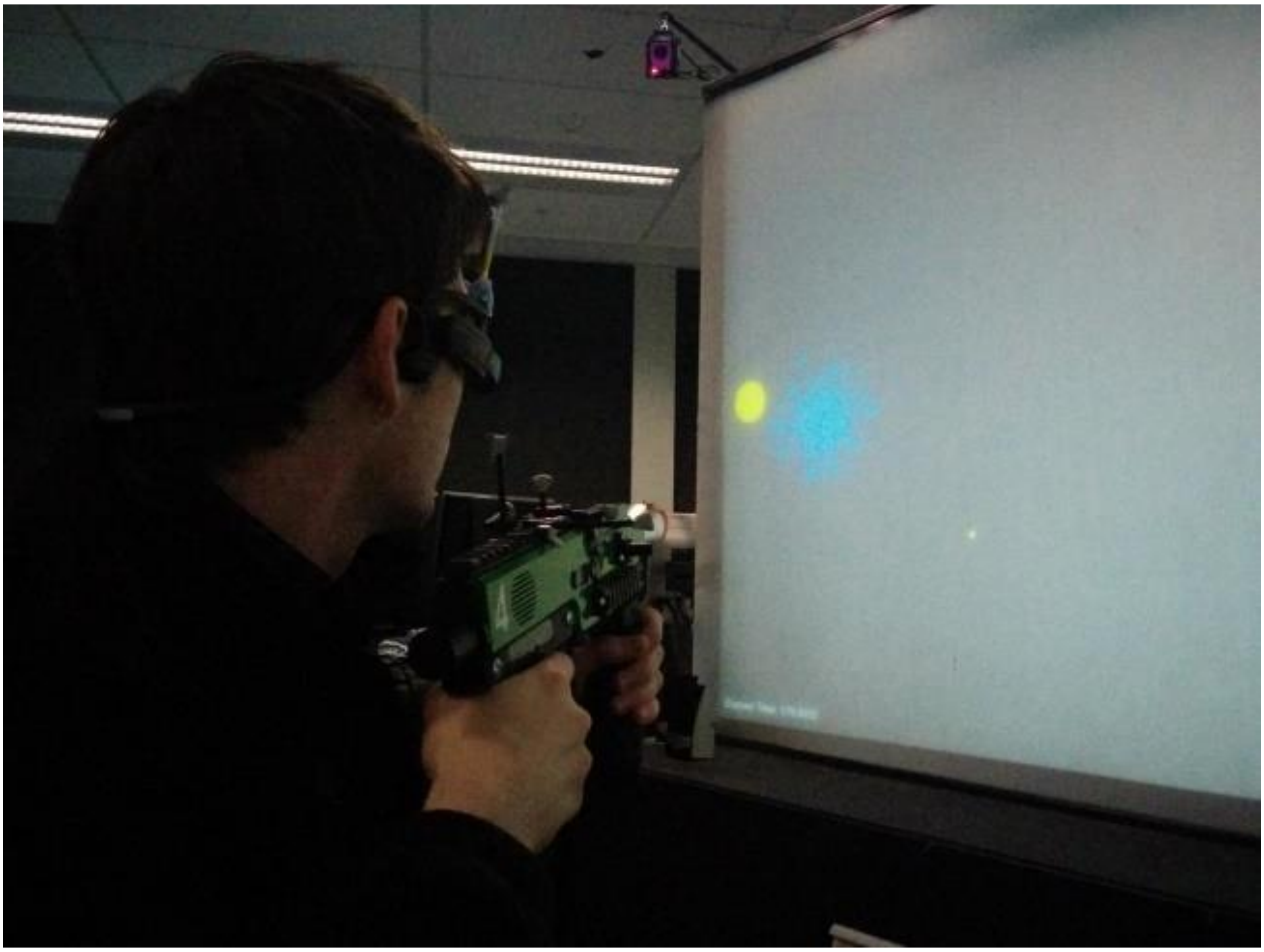


Figure 4.7: Participant attempting to shoot a target

4.4 Results

Two participants were excluded from the analysis (N = 28) – one for not following instructions, and the other due to failure of the on-board sound card to deliver audio signals. Additional technical difficulties with the Recon Jet, primarily associated with power management, meant data from the third block of trials for an additional six participants was recorded incompletely or lost. Data gathered from the third block was excluded from the analysis for all participants to maintain uniformity. Since this study investigated how attention can be redirected in the real-world using a combination of cues, the onset of head motion was accorded a greater importance than the rest of the measures such as time for the gun to reach target and time until target was hit. Here we report results for the auditory and visual cues presented in the nine different combinations specified earlier (Table 1). For an over view of the results obtained for the visual cues only, see Ward et al. [154].

A 3x3x4 (visual cues: 3, auditory cues: 3 and target positions: 4) repeated measures analysis of variance (ANOVA) was carried out to test for the main and interaction effects between the factors. Since the data violated Mauchly's test of sphericity, values determined by the Greenhouse-Geisser correction were used. The results demonstrate significant main effects of all three independent variables; audio cues (F (1.677, 80.513) = 104.671, $p < 0.001$), visual cues (F (1.767, 84.822) = 60.736, $p < 0.001$) and target positions (F (2.244, 107.732) = 54.592, $p < 0.001$). The results also demonstrate significant two-way interactions between all pairs (visual x audio: F (3.038, 145.835) = 16.041, $p < 0.001$; visual x position: F (3.432, 164.742) = 12.754, $p < 0.001$; audio x position: F (3.928, 188.535) = 8.248, $p < 0.001$)). In addition to this, a significant three-way interaction between the three independent variables was also observed (F (5.528, 265.30) = 4.054, $p = 0.01$).

Interaction effects between the auditory and visual cues were explored by fixing the levels of the target position. Statistically significant interactions (F (2.942, 158.852) = 11.414, $p < 0.001$) were observed between the auditory and visual cues at the -100° target position. Main effects for auditory (F (1.784, 96.32) = 47.764, $p < 0.001$) and visual cues (F (1.806, 97.504 = 44.114, $p < 0.001$) demonstrated statistically significant results. Similar results were observed for the target at +100° with interaction effects between the two cues demonstrating statistical significance (F (3.258, 169.395) = 5.852, $p < 0.001$). Main effects for the cues also demonstrate statistically significant results (Audio: F (1.562, 81.214) = 20.299, $p < 0.001$; Visual: F (1.927, 100.179) = 30.847, $p < 0.001$). No significant interactions between the audio and visual cues were observed for the targets at -50° (F (2.209, 119.266) = 1.839, p = 0.159) and +50° (F (3.1, 161.204) = 0.971, p = 0.410). The -50° position demonstrated significant main effects for both auditory (F (1.558, 84.145) = 37.285, p < 0.001) and visual cues (F (1.802, 97.331) = 5.982, p = 0.005). At the +50° position a significant main effects were observed for only the auditory cues (F (1.785, 92.81) = 30.743, p < 0.001). The lack of a significant effect for the visual cues suggests that the peripheral vision over-rides any of the visual cues when the targets appeared in these regions. Participants appear to lock onto these targets because of the auditory cues and/or peripheral vision, rendering the visual cues ineffective.

The preceding analysis was then followed up with an analysis of variance (ANOVA) for each target position to compare performance between the different cues and their combinations. All positions displayed a significant difference in performance between the different cue types and their combinations (-100°: F (4.494, 242.66) = 31.008, p < 0.001; +100°: F (3.656, 190.112) = 17.182, p < 0.001; -50°: F (3.851, 207.944) = 12.063, p < 0.001; +50°: F (5.018, 260.93) = 9.672, p < 0.001). Post hoc tests using the Bonferroni correction of pair wise comparisons between the cueing conditions give a detailed picture of the effectiveness of the cues for each of the four positions. For the auditory cueing conditions (V0S0, V0S1 and V0S2) only, the static (V0S1) and dynamic auditory (V0S2) cues outperform the no cue condition, V0S0, at all target positions (-100°: p < 0.001; +100° (V0S2): p < 0.001; -50° (V0S1): p = 0.006; -50° (V0S1): p < 0.001; +50 (V0S1): p < 0.001; +50 (V0S2): p = 0.04) except +100° V0S1 (p = 1). No significant difference was observed between the static (V0S1) and dynamic (V0S2) cueing conditions at -100° (p = 0.194), -50° (p = 1) and +50° (p = 1). A significant difference, though, was observed at +100° (p = 0.008). While further investigation is required, the binaurally spatialised auditory cue consistently demonstrates a faster onset of head motion time across all targets (Figure 4.8).

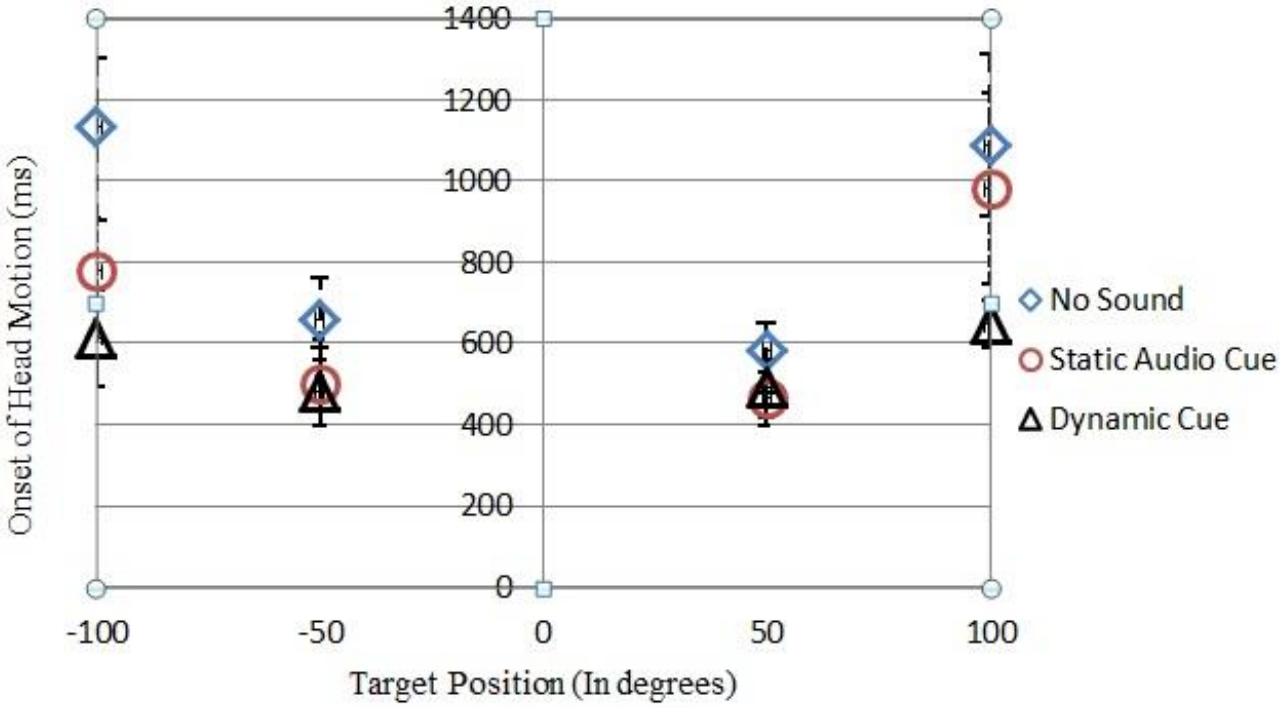


Figure 4.8: Average time for onset of head motion measured across all auditory cueing conditions

For the cueing conditions using V1 paired with the auditory cues (V1S0, V1S1 and V1S2), a significant difference is seen between V1S0 and V1S2 at all positions (-100°: $p = 0.02$; -50°: $p < 0.001$; +50°: $p < 0.001$; +100°: $p = 0.001$). Significant differences were also seen between V1S0 and V1S1 at +100° ($p = 0.012$), -50° ($p < 0.001$) and +50° ($p = 0.001$), while -100° showed no significant difference between the cues ($p = 0.06$). This result is similar to the one obtained with only auditory cues earlier. No significant differences were observed between V1S1 and V1S2 at any of the positions ($p = 1$). Both these conditions show closely matched onset times for head motion, with V1S2 displaying a marginally quicker onset (Figure 4.9).

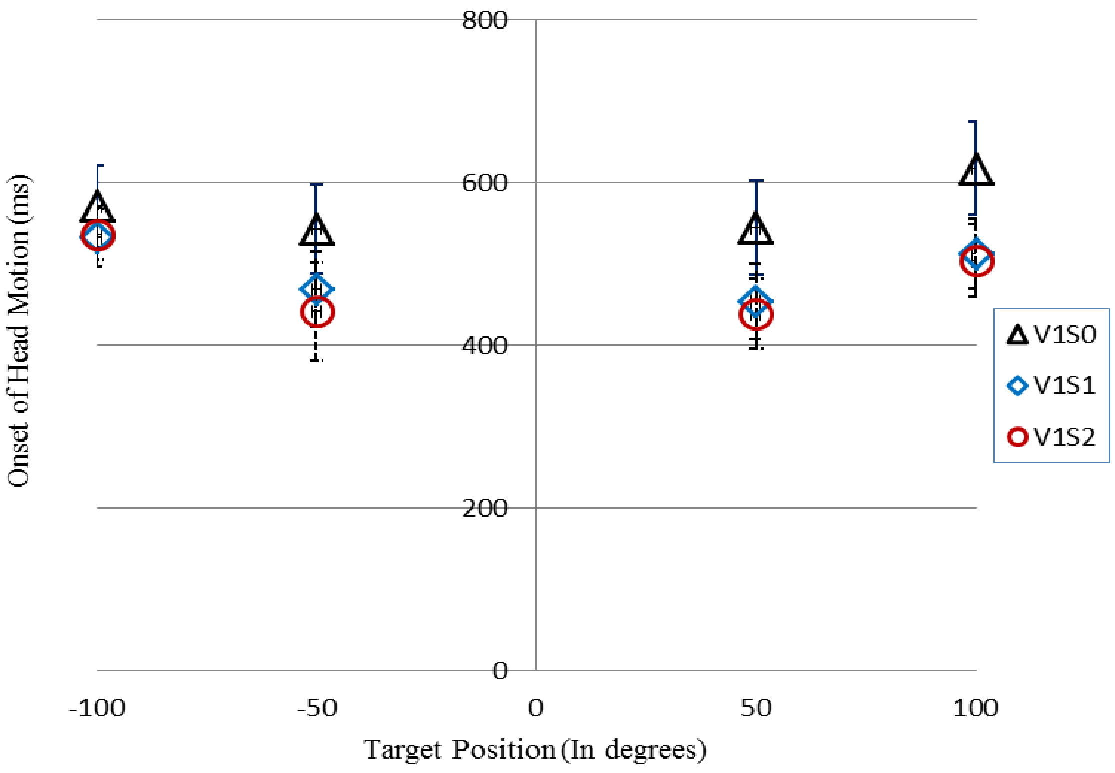


Figure 4.9: Comparisons between on-set of head motion times for the static visual cue (V1) paired with the auditory cues. A significant difference exists between on-set of head motion times for no auditory cue vs. auditory cueing conditions. No significant difference is seen between the static (S1) and dynamic (S2) auditory cueing conditions when paired with the static visual cue (V1)

For the cueing conditions using V2 paired with the auditory cues (V2S0, V2S1 and V2S2), a significant difference is observed between conditions V2S0 and V2S1 at -100° ($p < 0.001$), -50° ($p = 0.01$) and +50° ($p = 0.003$). No significant difference between the cueing conditions is seen at +100° ($p = 0.261$). Comparisons between V2S0 and V2S2 display significant differences at -100° ($p = 0.001$), -50° ($p = 0.044$), +100° ($p = 0.017$) and +50° ($p < 0.001$). Comparisons between the V2S1 and V2S2 pair does not show any significant difference at -100°, +100°, -50° and +50° ($p = 1$). These cueing conditions (V2S1 and V2S2) appear to have similar onset times across all targets (see figure 4.10).

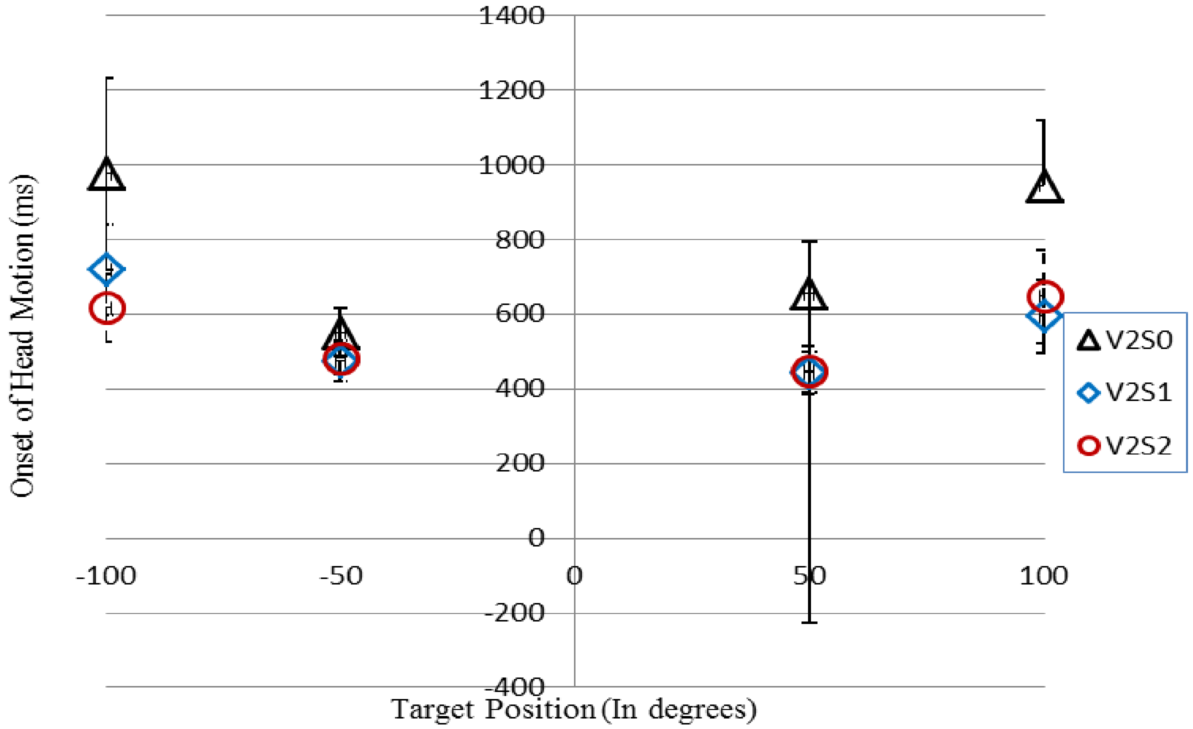


Figure 4.10: Comparisons between on-set of head motion times for the pursuit visual cue (V2) paired with the auditory cues

From the analysis that has been carried out, it is clear that the presence of a visual or auditory cue definitely elicits a quicker onset of head motion from the time the target appears at any one of the positions. The lack of significant differences (-100°: p = 1, -50°: p = 0.753, +50°: p = 1 and +100°: p = 1) between the spatialised auditory cue (S2) and the static visual cue (V1), points to the fact that both these cues are nearly equally good at redirecting attention. Although a combination of these two cues consistently registered the quickest time for the onset of head motion across all positions, in some cases these differences do not appear statistically significant. This lack of statistical significance appears mainly when this cue (V1S2) is compared with other cues that include either a binaurally spatialised auditory cue (S2) or a static visual cue (V1). While V1S2 appears to be best suited for attention directions tasks, a comparison of the onset of head motion times between +100° and -100° for this cue shows that the cue performs marginally better for the right side i.e. 100° (506.61 ms vs. 529.53 ms) (see figure 4.11).

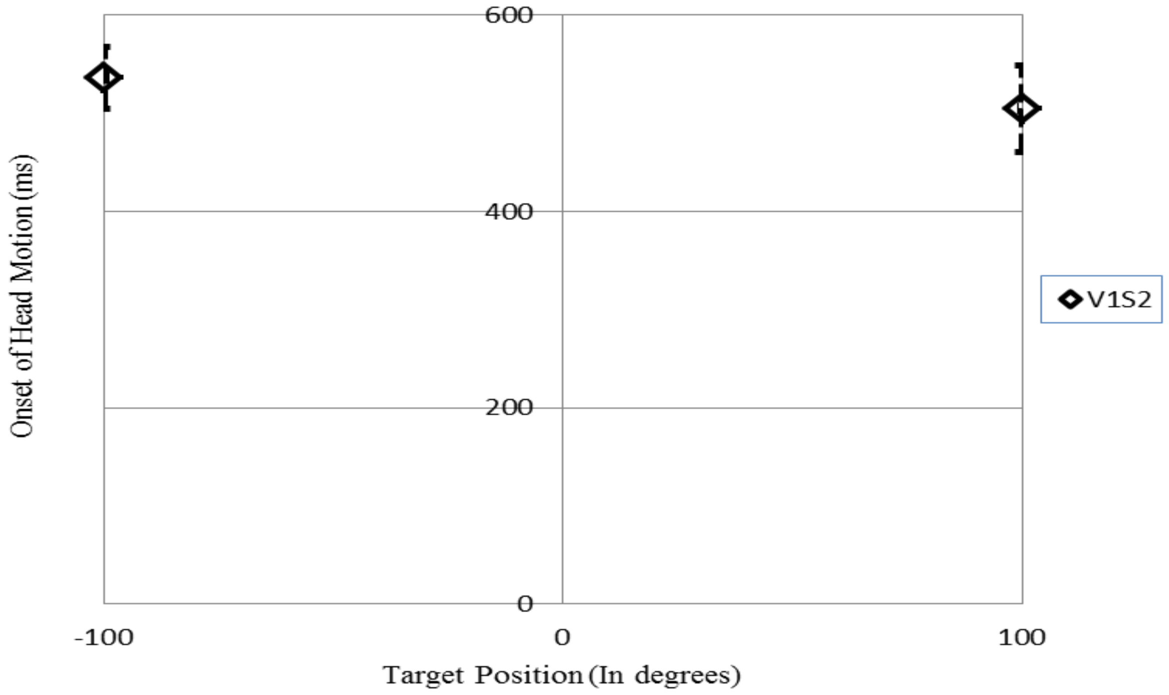


Figure 4.11: Onset of head motion times for the cueing condition V1S2 (static visual cue and dynamic auditory cue) for ±100°

4.5 Discussion

We have been able to demonstrate the benefits of using auditory cues in an attention redirection task via this study. The results in this case can be categorized into two distinct types: (1) auditory cues only and (2) audio-visual cues. The first part of the results section falls under the auditory cues category. The results obtained for these cueing conditions suggest that the binaurally spatialised, dynamic auditory cue is effective for redirecting attention to targets that do not occupy the user's field of view i.e. ±100°. The absence of a significant difference between the static and dynamic cueing conditions for targets at ±50° is likely because the targets fall within the user's peripheral vision [155]. The appearance of the target in the peripheral vision could be responsible for over-riding both the auditory cues, negating their effect. This effect extends across the two auditory cueing conditions S1 and S2 paired with the two visual conditions V1 and V2 for targets at ±50°. Conversely, when either of the auditory conditions paired with a visual cue was compared to the performance without an auditory cue, a clear difference in the onset of head motion times and target acquisition times is observed. This is indicative of the fact that even in the presence of a visual cue delivered during a visually demanding task, an auditory cue is more likely to attract attention and help reorient user attention in the space around them. Another observation that points to the effectiveness of the binaurally spatialised dynamic cue is the absence of a significant difference between targets on the same side i.e. ±100° and ±50°. This result effectively demonstrates that the binaurally spatialised auditory cue is as good at redirecting attention to targets outside the visual field as the visual percept is at acquiring targets at ±50° in this this experiment.

In the case of combinations of the visual cues, V1 and V2, with auditory cues S1 and S2, the pairing of the static visual cue, V1, with the dynamic auditory cue, S2 appears to provide the best results. As we have demonstrated with the auditory cueing condition only, these results show superior performance in comparison to other cueing condition pairings when compared with onset of head motion and target acquisition times for targets outside the visual range. This study clearly indicates that the use of auditory cues in conjunction with visual cues to reorient attention is possible. The results from our study compare favourably with those of Perrott et al. [58], Nelson et al. [115] and Rudmann & Strybel [145].

Despite these positive results, it is important to note that there are some limitations that must be taken into consideration when viewing these findings. The most obvious one is the use of three screens in the far field as substitutes for a real-world environment. While this set-up has served well to demonstrate the effectiveness of both auditory and visual cues at redirecting the user's attention to a point in the environment, the lack of complexity in the environment is a key drawback. The use of a simple far-field visual cue can also be limiting factor in this study. Additionally, we have used only a single 'alarm ping' in this experiment as the solitary auditory cue. The same auditory cue was presented to a participant for the duration of the study – either as a binaurally spatialised or non-spatialised cue. Taking this into account, further studies could look at projecting a more complex scene in the same environment. The use of associative sound cues versus alarm sounds can also be explored.

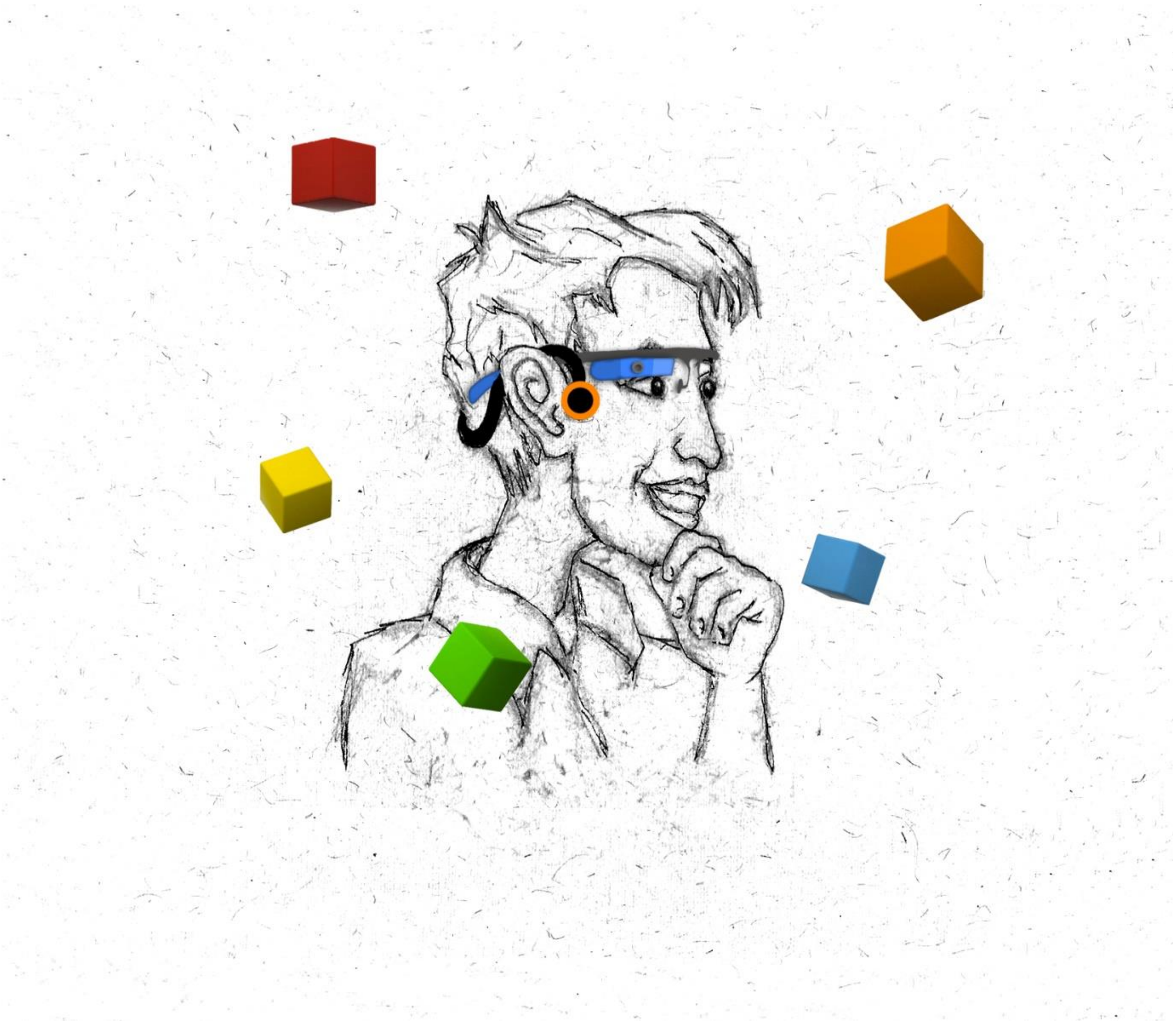
4.6 Conclusion

We have demonstrated the use of a binaurally spatialised, dynamic auditory cue in conjunction with a visual cue to redirect user attention. These reorientation cues appear to be most effective for targets outside the visual field, but have also shown to be of use within the peripheral vision in comparison to having no auditory cue at all. The use of an auditory cue or alarm in a visually demanding task cannot be underestimated. The dynamic auditory cue appears to be able to redirect the user’s attention without inducing a frantic search of the visual field, a behavior that was seen with the static auditory cues. Similar to a ‘3D’ auditory cue delivering azimuth, elevation and distance information used by Nelson el al. [115], our dynamic auditory cue exhibits superior performance compared to the static cue. These results also demonstrate that the binaurally spatialised, dynamic auditory cue will be useful in the event that a user does not latch on to a visual cue that may be presented simultaneously. The outcomes from this study also appear to suggest that a ‘dual delivery’ of cues across two different modalities appears to ensure that the system is somewhat fail safe.

For the purpose of this experiment only four specific targets were used. In the future, it will be worthwhile exploring how both visual and auditory cues will perform in the presence of visual and auditory distractors. In addition to this, both auditory and visual cues can be tested in more complex and even real-world environments. It would be prudent to explore the use of associative auditory cues versus alarm sounds to determine how effective they are at redirecting a user’s attention. Results from such experiments could provide further evidence for the use of auditory and visual cues in a context sensitive manner. Such results would also provide further evidence of the utility of a BCH as an auditory display incorporated into a wearable interface.

Study V

A REAL-WORLD SEARCH TASK EXECUTED USING A BI-MODAL, WEARABLE AUGMENTED REALITY INTERFACE



In this second application based study, we take another real-world approach to the use of our prototype device. This experiment explores the possibilities of using the wearable, bi-modal interface for an indoor search task. Participants were required to navigate a relatively complex environment to find targets in a room. Target search was aided by visual and auditory cues provided to users via the Google Glass and a bone conduction headset.

This study builds on the findings of the previous study by comparing similar visual and auditory cues, albeit in a different environment. A direct comparison of similar cue types between two different environments and search strategies can be made in order to effectively predict the kind of cues that will be most effective based on the environment and use case.

4.7 Hypothesis

The experiment carried out as part of this study was to evaluate the effectiveness of visual and auditory cues at being able to guide a user to a designated target. The experiment builds on the findings of a previous study of a similar nature [129], but with the added requirement on the user to navigate a complex environment.

We hypothesised that a map would be the best at guiding participants to the target, with the arrow performing marginally worse based on the search strategy that we hoped the participants would employ. This hypothesis was proposed since the orientation of the map matched that of the participants when they entered the search area. A map with the location of the target indicated on it (Figure 4.17a) was, therefore, expected to guide participants to the target faster than any other cue. In the case of the arrow (Figure 4.17b), the expected search strategy was for participants to look at the direction in which the arrow was pointing along with the distance-to-target readout and immediately proceed in the indicated direction. The assumption in this case was that while the arrow did not provide a specific location for the target, the direction in which it was pointing along with a distance-to-target readout would provide sufficient information for participants to locate the target in a relative short time span.

Similarly, amongst the auditory cues, a static cue was expected to perform better than a dynamic cue. The ability to ‘home in’ on the static cue was seen as being more intuitive in comparison to the dynamic cue. The use of these two auditory cues allowed us to compare their effectiveness across different task types in two environments; namely attention redirection (Study IV) versus a search task in a complex environment. A simple measure of time taken to find the target from the moment the cue(s) were delivered was calculated to evaluate the effectiveness of each of the cue combinations. Detailed analyses and the treatment of data to negate the influence of biases and effects are provided in the results section.

4.8 Method

4.8.1 Apparatus

The apparatus for the experiment can be divided into three categories: (1) tracking cameras with an allied software package used to track the movements of participants through the designated experimental area, (2) hardware worn by the user though which auditory and visual cues were displayed; hardware worn by participants that allowed their movements to be tracked through the experimental space and (3) handheld and desktop based equipment used by the experimenters to initiate, record and terminate trials during the experiment.

The tracking equipment used for the experiment consisted of six OptiTrack Flex 13 cameras [156] paired with OptiTrack's Motive optical motion capture software [157] developed by NaturalPoint. The cameras were spread across the room in a manner which allowed maximum coverage of the experimental area without compromising tracking accuracy. The Motive software communicated with the experiment developed in Unity3D [115] via the Motive plugin for unity. Data was streamed to the experiment build via a 'connector' from Motive. The participants wore a cap with eight markers affixed to it to allow their heads to be tracked. Participants were also required to wear the Google Glass [72] and an Aftershokz Sportz3 [117] bone conduction headset as part of the experiment (Figure 4.12). Visual cues were delivered via the Google Glass while spatialised auditory cues were delivered via the bone conduction headset.

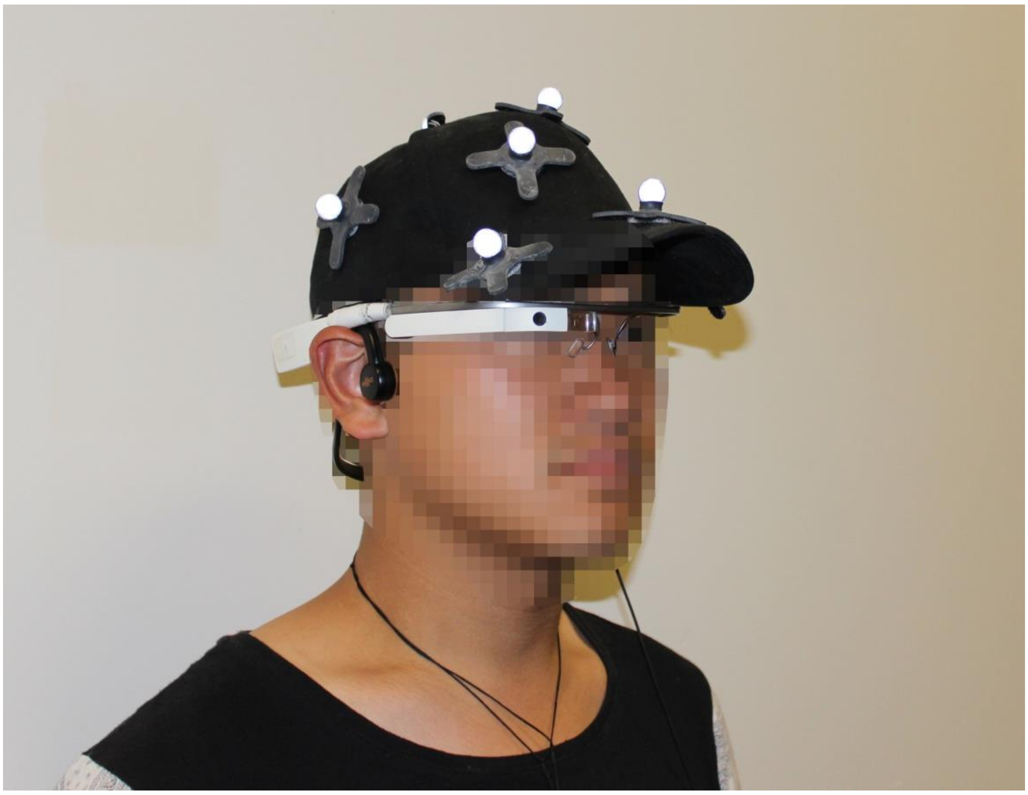


Figure 4.12: Participant wearing a cap with trackers along with the Google Glass and a bone conduction headset

To allow participants to move freely through the designated search area, the auditory cues were delivered wirelessly to the BCH via Wi Digital System's AudioStream ProAV [158] stereo wireless transmitter and receiver package (Figure 4.13). Audio was streamed to the transmitter via the PC's on-board sound card.

Figure 4.13: Wi Digital AudioStream ProAV wireless system

The experiment itself was initiated by the experimenters and certain aspects of it recorded in real time during the trials. To facilitate this, a desktop PC and mobile phone were used. The desktop PC that ran the experiment was also used to manually initiate and terminate the experiment, while the mobile phone was used to record the number of incorrect judgements participants made during a trial. A block diagram of the experimental setup shows how the elements of this study interfaced with each other (Figure 4.14)

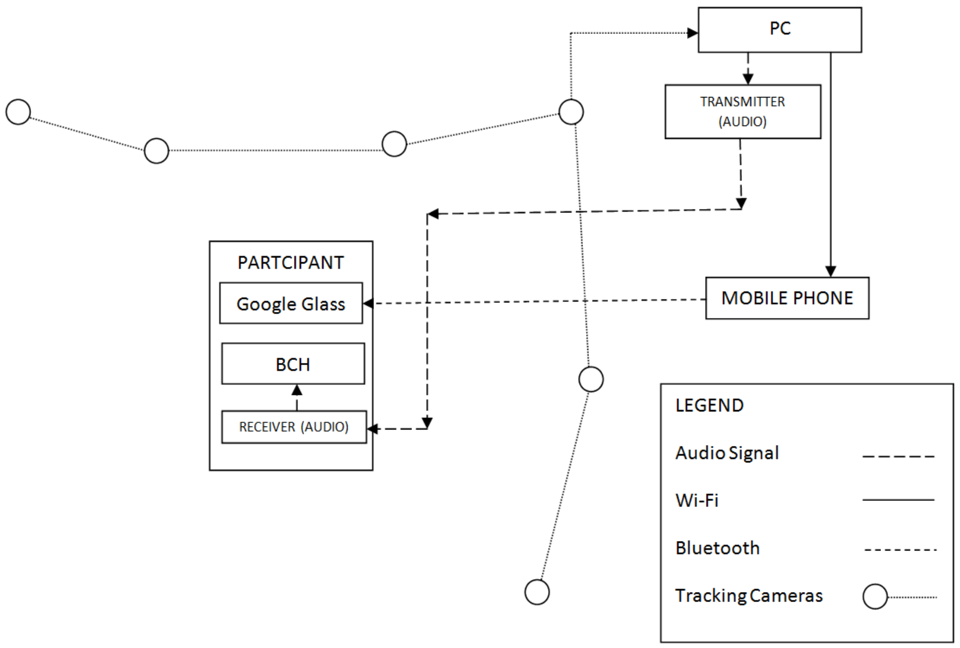


Figure 4.14: Block diagram denoting signal flow between equipment used for the experiment

4.8.2 Environment

The experiment was carried out in a large room with a test area of 4.2m x 6.2m. Six tables of various sizes were positioned within this space with between four and fifteen laminated cards placed on them (Figure 4.15).

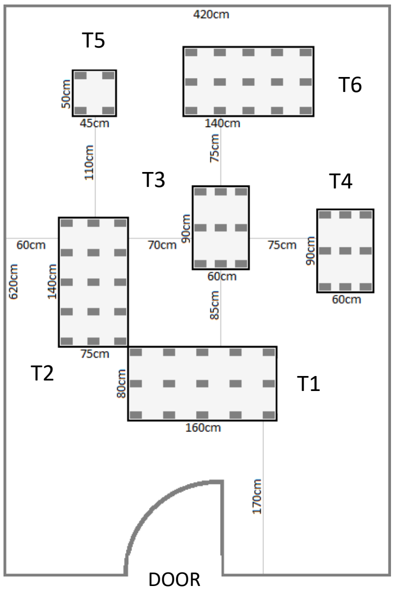


Figure 4.15: Environment layout (to scale). Test area includes an entry (door), 6 tables and 67 cards

Each card was 12.7cm by 7.6cm. Cards were blank on both sides excluding a single target card. The target card had a green dot on a single face measuring 6cm (Figure 4.16). A total of 67 cards were placed on the tables in the experimental space during each trial.

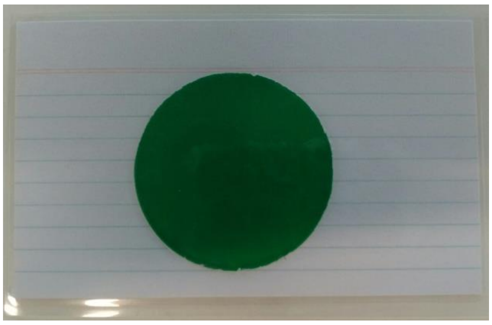


Figure 4.16: Target Card

4.8.3 Stimuli

Participants were presented with auditory (static and dynamic) and visual stimuli (map and arrow) for the experiment. A detailed explanation of the stimuli is provided in the following sub-section.

Table 2: Cue Types

AUDITORY CUES	**VISUAL CUES**
No sound	Blank Screen
Static Sound	Map
Dynamic Sound	Arrow / Heading

- Auditory Stimuli
  Auditory stimuli consisted of a one second ping. The same sound was used for the two auditory stimuli that were presented to the participants. Both stimuli were binaurally spatialised using the 3Dception plug-in for Unity developed by Two Big Ears [116]. While one was a static sound source that was positioned at the target location and constantly pinged, the other was a dynamic stimulus that shot out

towards the target from the participant's position. This auditory cue traced a parabolic trajectory to the target and was identical to the one used in [129] The dynamic stimulus was played only once when the participant entered the room. An inverse square law fall off of the auditory cue was modelled in order to facilitate a better direction and distance judgement. The static auditory cue increased in level as the participant approached the target. This behaviour reflects a common occurrence observed among sound sources in everyday life i.e. while the level of the source itself remains the same, the distance from it determines how loud the sound is at the observer's position. We have chosen to do this based on anecdotal evidence and our own experiences looking for a misplaced cell phone by ringing it and homing in on the ringtone.

For the dynamic cue, a longer fall off indicated that the target was further away. This cue was used as it has been shown to be effective at redirecting a user's attention in space [129]. Its use allows us to directly compare efficacy with the static cue for a given task in a more complex environment. The auditory cue was designed in accordance with alarm sound design guidelines prescribed by Walker and Kramer [152]. While wideband noise has been shown to be localised the best [114] [120] [42] [123], the alarm sound was chosen for its aesthetic appeal [153] and greater ecological validity.

- Visual Stimuli

  Two visual stimuli were used to provide navigation information. The first one consisted of a 'map' (Figure 4.17a) of the search area with tables marked out as white rectangles. A red dot superimposed on any of these rectangles indicated the area on the table where the target was located. The second visual stimulus consisted of an arrow (Figure 4.17b) that pointed in the direction of the target relative to the user's current heading. The distance of the participant from the target, in metres, was also displayed above the arrow. The arrow updated the distance and direction in real time based on the tracking information it received from the programme. The length of the arrow changed depending on the horizontal angular distance between the participant's head and the target i.e. the arrow grew in the length if the rotation required to orient oneself towards the target was large. Tracking for this cue was actively only in horizontal plane. Moving one's head down towards a target did not alter the heading or distance.

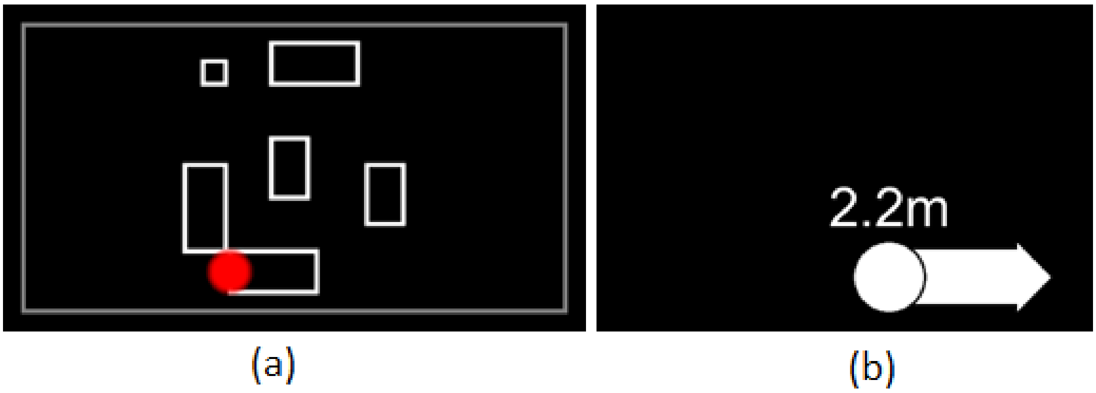


Figure 4.17: Visual Cues as displayed on the Google Glass. (a) Map Cue (b) Arrow/Heading Cue. Black background appears completely transparent when displayed on the Glass

4.8.4 Participants

Thirty-six participants (20 Male, 16 Female) aged between 21 and 38 years (Mean: 25.7 years, Std. Dev: 4.3 years) took part in the study. All participants reported normal hearing. No audiometric screening was carried out to verify this since there appears to be no know relation between auditory localisation performance and audiogram results unless the hearing loss is significant [58]. A total of 17 of the 36 participants had taken part in at least one of the author’s previous auditory experiments and were familiar with the bone conduction headset. Approximately 88% of the participants reported using headphones or earphones with their mobile device at least few times a week. Of the 36 participants, 17 reported being familiar with a wearable computing device. While not a large population, these numbers reflect the increased usage of headphone/earphone with mobile devices over the last few years. Familiarity with wearable devices also indicates that the wider population is gradually beginning to gain awareness of such devices, with more and more becoming available in the consumer space in the last three – five years. All participants were compensated with NZD $10 gift vouchers

4.8.5 Procedure

As part of the search task, participants were required to find a card with a large green dot on it placed amongst other cards spread out evenly on the tables in the room (Figure 4.18). The card was placed face down to prevent participants from being able to recognise the target as soon as they entered the room. They were guided to the target using any one of the cue combinations shown in the table 2.

All participants were provided with an explanation of the experiment and what was expected of them prior to beginning the trials. They were instructed to find the target using the auditory or visual cues presented to them over the Glass and/or BCH. The entire experiment was divided into four blocks of nine trials each giving us a total of 36 trials per subject. Each block had one trial where participants got no cues at all (see table 2). The first block of trials was set aside as practice trials and participants were permitted to ask the experimenters questions or express doubts they may have during this period. On completing this block, no more interaction between the participant and experimenters was permitted. The entire experiment took about 30 minutes to complete.

The experiment was designed such that the cues would be delivered to the Glass HMD and/or bone conduction headset as soon as the participants entered the search space and the tracking cameras had detected the head tracker. For the experiment itself, participants were asked to leave the room so that one of the experimenters could place the target card at a designated position on one of the tables. Following this, the other experimenter ‘primed’ the experiment. The priming allowed the experimenters to avoid accidental starts. Once the experiment was primed, the participant was asked to enter the room. The cues were delivered to the participants as soon as they entered the room and a timer began counting the number of seconds it took the participant to find the target. During the trial, one of the experimenters used a mobile phone connected wirelessly to the PC to log the number of incorrect flips made by the participant while attempting to locate the target. On finding the target the trial was ended by one of the experimenters and the same cycle repeated for the next set of cues. All trials were capped at 40 seconds for the sake of brevity i.e. if the target was not located in 40 seconds, the trial was terminated. On completion of the experiment the participants were presented with a post-study questionnaire. The questionnaire asked participants to rate cue preference and speed of target acquisition. Additionally, the participants were also asked to indicate which of the auditory and visual cues they preferred.

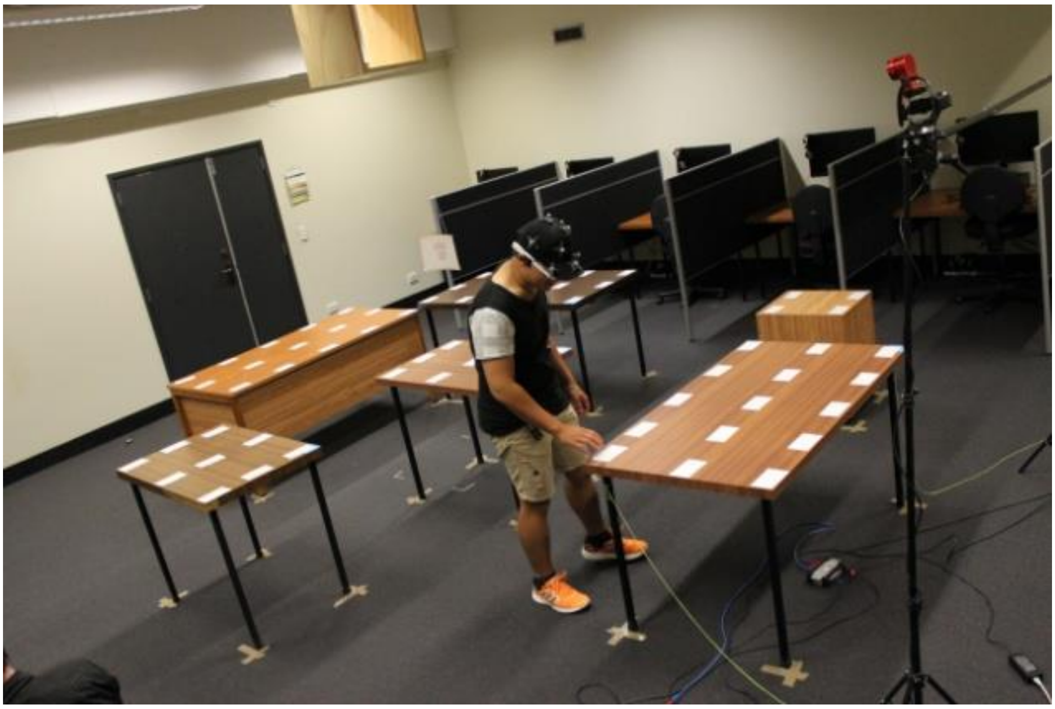


Figure 4.18: A participant executing the search task

4.9 Results

On completing the first block of practice trials participants proceeded to undertake the three main trial blocks. As we have already mentioned in the previous section, each block consisted of nine trials resulting in a total of 27 trials that were used for the analysis. To negate the random effect of the distance to the target card from the starting point on response times between participants, each participant was presented with the same sequence of target locations. A random order of cue types was delivered to the participants within the set sequence of target locations. This ensured that any order effects were accounted for. The process also safeguarded against the dependence of cue type and distance to target effects on each other. One participant’s data was excluded for the analysis for failure to complete the trials as a result of an equipment malfunction (N = 35).

A preliminary ANOVA predicting response time for the target table, current participant, and the number of times a particular cue had been presented was performed. When not accounting for the cue effects, there were no significant differences in the average search times between participants (F (34, 474) = 0.90, p = 0.630, η^2=0.06). Performance over time appeared stable (F (7, 474) = 1.12, p = 0.346, η^2=0.02) with no evidence of a linear (p = 0.168) or quadratic (0.278) increase or decrease in performance as participants progressed though the trial blocks. An effect of distance to the target remained. As expected, the central table and those towards the far end of the search space took longer to reach than the one directly at the entryway (F (5, 475) = 6.66, p < 0.001, η^2=0.07). The effect of the target's table did not differ significantly between participants (F (175, 485) = 0.86, p = 0.877, η^2=0.24) or across blocks (F (14, 474) = 0.99, p = 0.465, η^2=0.03). To eliminate the effect of the target's location (table) on the cues, the average time to find a target for each trial's table was subtracted from the trial response time. The following analyses use and report the adjustment from the average time for a table caused by each cue. The average time to find a target on each table and the average adjustment for each cue and for each participant are shown in table 3 below.

Table 3: Average time to find targets on all tables, average adjustments for each participant and cue type

Tables	T1	T2	T3	T4	T5	T6			
Mean	13.42	16.22	17.76	15.03	15.83	19.28			
	Cue Types								
Auditory Cues	No Sound			Spatialised Static Audio Cue			Spatialised Dynamic Audio Cue		
Visual Cues	None	Map	Arrow	None	Map	Arrow	None	Map	Arrow
Participants									
1	40.0	6.8	15.9	18.6	6.3	17.9	31.1	9.4	13.9
2	31.0	6.2	12.2	10.2	7.4	23.5	17.7	6.1	15.8
3	29.4	7.1	10.9	16.7	7.1	11.3	36.8	6.5	11.3
4	39.0	9.0	21.9	35.9	13.6	19.0	23.1	12.5	23.0
5	34.8	8.2	12.9	14.0	8.5	10.7	24.1	5.2	18.3
6	33.1	5.7	10.6	16.3	6.8	15.0	20.0	6.3	12.3
7	40.0	7.0	15.2	26.8	5.0	14.4	33.7	10.4	14.3
8	31.5	7.9	16.5	20.8	8.5	20.3	34.3	7.9	21.8
9	35.9	7.6	14.6	19.6	8.7	22.9	40.0	8.1	17.4
10	29.3	6.2	14.0	10.9	8.8	11.7	31.7	5.6	23.0
11	40.0	9.4	15.8	32.2	9.4	14.9	34.3	7.6	24.9
12	33.2	9.5	28.7	22.6	10.4	25.9	30.8	10.5	17.0
13	28.6	8.9	15.1	25.4	7.2	14.8	34.6	7.7	14.7
14	18.4	7.4	11.3	17.1	5.9	12.0	22.7	4.8	13.1
15	28.1	4.6	23.1	25.5	5.3	11.0	25.5	4.4	13.4
16	40.0	8.7	14.5	23.7	8.3	12.7	26.6	9.4	12.8
17	36.5	7.7	12.6	17.5	6.3	13.8	30.3	6.7	9.1
18	23.5	7.6	15.6	16.7	4.8	21.3	35.4	6.6	10.2
19	13.8	4.7	9.4	9.3	3.7	9.7	27.3	5.5	7.5

20	27.5	6.8	17.4	29.8	10.8	20.1	25.6	9.7	22.9
21	31.5	8.6	18.1	14.2	9.0	18.5	29.1	10.6	18.6
22	40.0	7.5	17.5	13.4	9.6	16.6	38.9	12.3	12.4
23	34.3	5.5	8.2	18.1	8.1	11.0	35.6	6.2	9.3
24	28.7	6.6	17.4	17.3	9.9	26.3	26.5	10.7	16.1
25	21.7	8.6	18.7	11.7	8.0	18.8	11.1	10.5	21.7
26	40.0	7.9	21.2	31.1	6.0	11.9	32.3	7.4	14.4
27	40.0	9.2	11.4	16.3	6.7	17.3	26.1	6.5	14.5
28	25.1	5.2	15.7	18.9	5.2	13.4	30.4	4.9	11.6
29	20.0	7.0	9.2	16.5	5.8	15.4	40.0	4.8	14.0
30	35.5	4.0	6.6	12.0	5.7	8.8	31.3	5.1	8.0
31	32.7	9.0	14.6	22.1	6.4	15.6	36.4	6.0	16.3
32	24.5	6.6	10.2	16.3	7.1	11.1	27.2	4.6	8.2
33	38.2	5.5	10.9	14.9	9.2	8.1	20.6	6.1	20.1
34	37.6	6.0	9.6	11.0	5.9	12.0	28.6	4.8	12.6
35	34.4	6.5	12.4	13.2	8.0	9.2	20.2	5.4	11.6
Average adjustments for all cue types (Figure 4.19)									
	14.9	-9.4	-1.9	1.9	-9.5	-1.1	12.5	-9.0	-1.3

Next, a repeated measures ANOVA of the visual and auditory cue types was performed predicting average search time adjustment from the table average for each participant. A Mauchly's Test of Sphericity found no significant deviations from the homogenous difference variance between the auditory cueing conditions (W = 0.970, $X^2(2) = 1.013$, p = 0.603). A similar analysis of the visual cueing conditions (W = 0.655, $X^2(2) = 13.96$, p = 0.001) and the interaction between visual and auditory cueing conditions (W = 0.318, $X^2(9) = 37.13$, $p < 0.001$) reveal significant deviations indicating a violation of the assumption of sphericity. To account for this a Greenhouse-Geisser correction is applied to the following results.

There exist significant differences in search times between the audio cueing conditions (F (1.94, 66.01) = 38.05, $p < 0.001$, $\eta^2 = 0.582$). Between the two auditory cues, in the absence of a visual cue, the static sound cue performed significantly better than the dynamic cue (t(34) = 7.9, $p < 0.001$, Cohen's d = 1.6). No significant differences were observed between the dynamic and no auditory cueing conditions (t(34) = 1.646, p = 0.109, Cohen's d = 0.35). As figure 4.19 shows, the presentation of visual cues significantly decreased search times (F (1.49, 50.56 = 348.34, $p < 0.001$, $\eta^2 = 0.911$). In particular, in the no audio cue trials, presentation of both the arrow cue (t(34) = 12.75, $p <0.001$, Cohen's d = 2.15) and map cue (t(34) = 23.56, $p <0.001$, Cohen's d = 3.98) lead to significantly shorter search times than the no cue control condition. Additionally, between the two visual cue types the map lead participants to the targets significantly faster than the arrow cue in the absence of the auditory cues (t(34) = 10.32, $p < 0.001$, Cohen's d = 1.75). Interaction effects between the cues, while seemingly significant (F (2.85, 96.80) = 35.06, $p < 0.001$, $\eta^2 = 0.508$), appear to be greatly affected by the visual cues. These effects are shown in figure 4.19. This is especially true for the map cue when paired with any of the auditory cueing conditions. The cause of the main effects for the auditory cues and the audio-visual cue interaction is the decrease in search time caused by the static audio cue when no visual cue is present (F (1, 34) = 35.06, $p < 0.001$, $\eta^2 = 0.763$). When the no visual cue conditions are excluded, the effect of the audio cues (F (1.98, 67.19) = 0.64, p = 0.530, $\eta^2 = 0.018$) and the interaction between audio and visual cues (F (1.90, 64.75) = 0.44, p = 0.637, $\eta^2 = 0.013$) on search speed disappears.

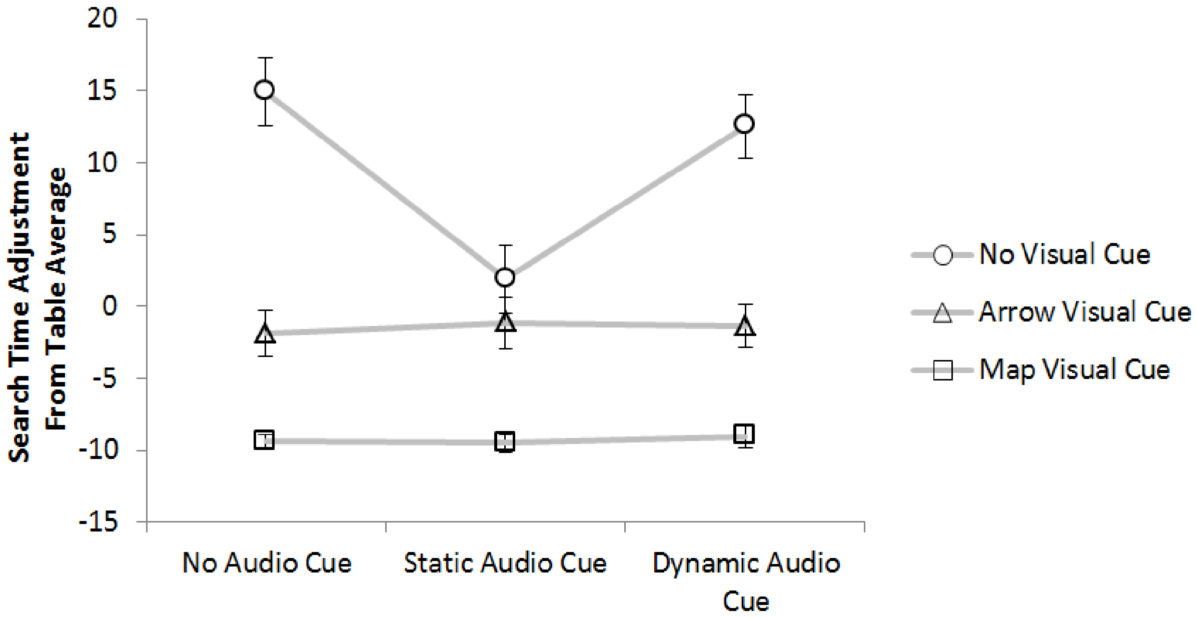


Figure 4.19: Search time adjustments from table average for each combination of visual and auditory cues. Error bars represent 95% confidence limits

Analysis of the number of incorrect flips demonstrates that the count was the least for the visual cue displaying the map as expected (Table 4). Table 4 shows the average incorrect flips across all participants for all cue combinations. Since the difference in variance between the conditions is too extreme to support analyses utilising ANOVA, comparisons between the in the incorrect flip rates were made using paired sample t-test. The test was adjusted for family-wise error rate using the Sidak correction. The corrected significance threshold for the comparisons is p = 0.0014.

Table 4: Average incorrect flips for all cue combinations

	No cue	Map	Arrow
No Cue	15.9	0.3	1.8
Static Sound	3.6	0.2	2.4
Dynamic Sound	13.3	0.2	1.9

Participants in the no audio/no visual cues condition made more incorrect flips than in any other condition (t's(35) > 7.87, p's < 0.001) except for the dynamic sound/no visual cue conditions (t(35) = 1.88, p = 0.069, Cohen's d =0.31). There were no differences between the map cue condition across all audio cueing conditions (t's(35) < 0.87, p's > 0.388), nor between the arrow cueing condition analysed across all audio cueing conditions (t's(35) < 1.66, p's > 0.106). the map cue alone resulted in fewer incorrect flips in comparison to the arrow cue alone (t(35) = 5.20, p < 0.001, Cohen's d = 0.87). The static auditory cue on its own caused participants to make a similar number of incorrect flips in comparison to the arrow cue (t(35) = 3.16, p = 0.003, Cohen's d = 0.53).

Analyses of the participant ratings for cue type preference[5] demonstrate an inclination towards the combined cueing condition (Figure 4.20). A multiple measures ANOVA run on the data obtained from the question demonstrates a significant difference exists between the cues (F (2, 68) = 41.865, p < 0.001). Post hoc tests using the Bonferroni correction exhibited participants' preference for the visual cue over the auditory cue (1.68 ±0.68 vs. 2.74 ±0.57, p < 0.001). No such significant difference (p = 1) was observed between the visual and combined cue types (1.68 ±0.68 vs. 1.59 ±0.66). A possible explanation for such a result could be a natural preference for the visual cue over the auditory cues [159] [160]. This holds true especially for the combined cueing condition where participants are more likely engaged visually versus aurally.

---

[5] Refer to Appendices: Experiment 5 Post Study Questionnaire, page 172

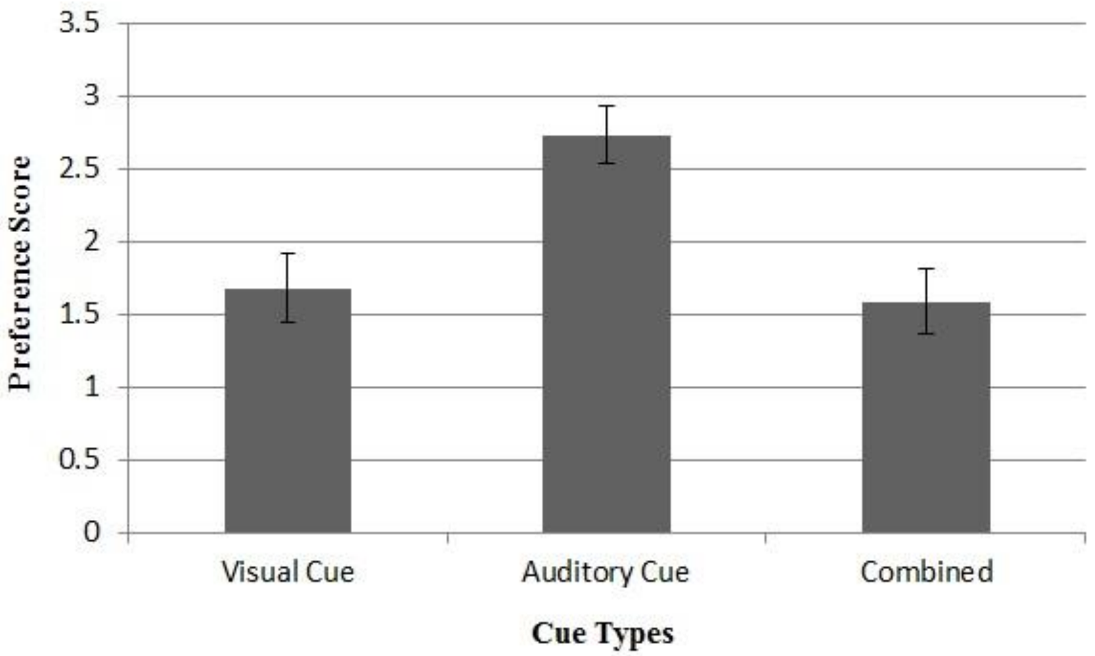


Figure 4.20: Preference ratings for the different cue types. A higher score denotes a lower preference

Amongst the visual cues, participants indicated an overwhelming preference (97.2 %) for the map cue. Similarly, for the auditory cues, participants showed a preference for the static sound cue (88.88%). Pairwise t-tests conducted among both groups of cues, auditory and visual, demonstrate a similar significant difference ($t(35) = 2.03$, $p < 0.001$) between the cue types within those groups i.e. participants preferred the map amongst the visual cues and the static sound among the auditory cues.

As part of the questionnaire, participants were also asked to rate cues based on their perception of target acquisition times. The best (fastest) cue was rated 1, while the slowest was rated 3. Self-rated speed scores for the different cues align with results discussed earlier in this section. Most participants felt that they performed best with the visual cues, followed by a combination of the visual and auditory cues. The auditory cue was rated as the slowest for target acquisition. Figure 4.21 shows the means of the self-rated target acquisition scores. Data from 2 participants was excluded for analysis of this question for failure to follow instructions on how the question must be answered. A multiple measures ANOVA performed on the data acquired from this questionnaire demonstrates that a significant three-way interaction exists between the participant ratings for the three cue types ($F(2, 66) = 22.691$, $p < 0.001$).

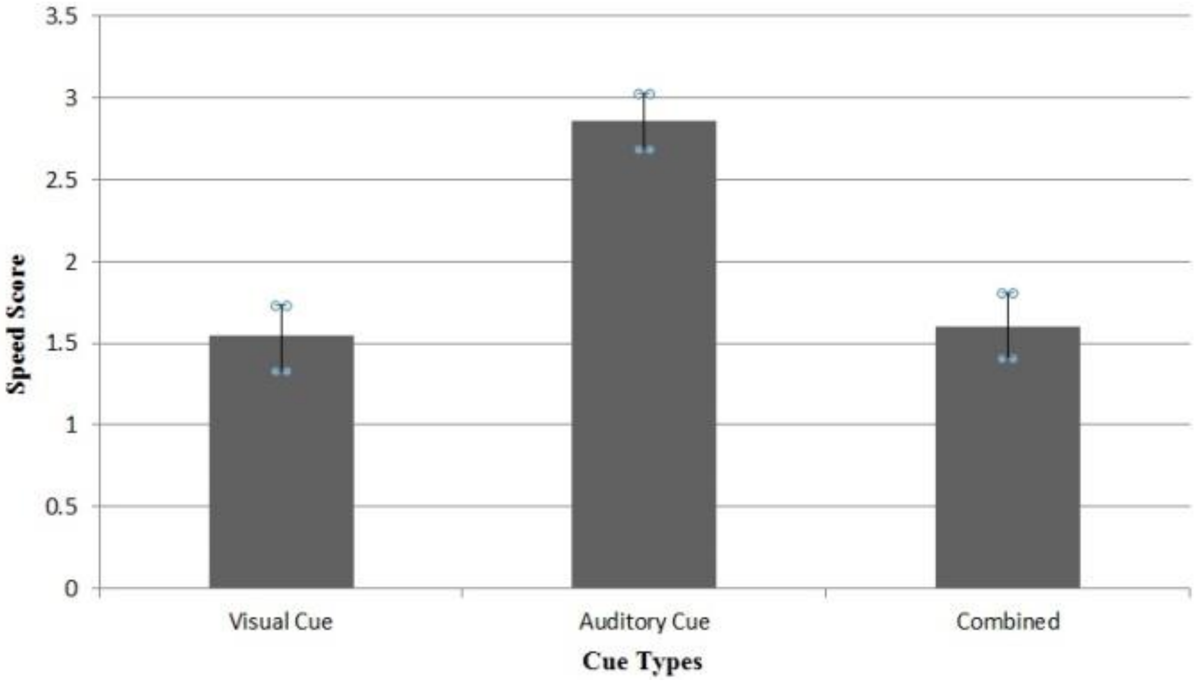


Figure 4.21: Self rated speed scores. A higher score indicates a greater target acquisition time

Post hoc tests using the Bonferroni correction revealed that participants felt they performed significantly faster (1.68 ± 0.68 vs. 2.73 ±0.57) with the visual cues in comparison to the auditory cues ($p< 0.001$). They did not seem to think that the combined cues had any significant effect (1.68 ±0.68 vs. 1.59 ±0.66) on the search times ($p = 1$). As seen in the earlier part of this section, these observations made by the participants are backed up by the data that has been obtained. The lack of a significant difference between the visual and combined cueing conditions was due to the visual cue as we have already mentioned earlier. We hypothesise that in the combined cueing conditions participants have relied almost exclusively on the visual cue versus the auditory cue leading to both the empirical data and subjective data demonstrating similar results.

4.10 Discussion

We have been able to demonstrate the effectiveness of using both auditory and visual cues in a simulated real-world search task. Unsurprisingly, the results obtained here demonstrate that visual cues are significantly better at aiding a search task than any other form of cueing in the absence of a prominent environmental marker. Providing the users with the location of the target superimposed on a rudimentary map of the area is sufficient to guide them to the target in the least amount of time. The heading indicator on the other hand proved to be more of a challenge. We attribute this to a moderately demanding learning curve; rotating and not moving in the direction the arrow was pointing to did not seem apparent to most participants from the outset. We also noticed that participants constantly monitored the arrow in an attempt to be as precise as possible as they moved through the environment resulting in a longer time required to reach the target. The optimal strategy of using the heading and distance to quickly reach a table and execute a search task was rarely employed by any of the participants.

Amongst the auditory cues, the static sound source performed significantly better than the dynamic source. Participants appeared to be able to home in on the target based on where they heard the sound and how loud it was. In the case of the dynamic cue, search times approached those seen in the no cue condition. While every effort to model a realistic fall off in the level of the auditory cue was made, it appears this did not translate well to the bone conduction headset. This is an important point to note since, despite being successful at directing participant attention in the right direction, the lack of any distance information made it relatively hard for participants to zero in on an area to begin their search. Based on these results we could conclude that while binaurally spatialised auditory cues are good at redirecting attention to a particular point in space [129], they lack the accuracy to aid in a search task in a complex environment. The binaural spatialisation afforded by the bone conduction headset appears to be good enough to allow for an aurally aided search task to be carried out with relatively accuracy and speed in comparison to no cues.

4.11 Conclusion

A visual search task was conducted using auditory and visual cues. While the visual cues performed significantly better than the auditory cues, it must be noted that their use may not be suited to all situations. In scenarios where the visual faculty is engaged actively in monitoring the environment, auditory cues provide an ideal alternative for information delivery. We have managed to build on previous studies [129] by showing that both auditory and visual cues can be used in a context sensitive manner to execute a task with greater accuracy and efficiency. While confirming some findings from previous studies, our experiment has also been able to counter some of them; namely that, a ‘dual delivery’ of cues [129] does not make for a better system. While the system may provide redundancy in some respects, there appears to be no advantages to providing both auditory and visual cues simultaneously as demonstrated by the experiment. This is especially true for the static auditory cue delivered with the map or arrow visual cue.

While the results from this study are encouraging, a major limitation has been our inability to use real objects for the search task. Finding an inconspicuous method to ‘tag’ items in a manner that would not make them pop out of the environment proved difficult. In addition to this, it remains to be seen how effective the perception of binaurally spatialised stimuli reproduced over the bone conduction headset is in a noisier environment. Obtaining real-time and accurate location and head orientation information at the ‘eye-level’ for a user is also a challenging proposition. This is an important consideration for the usability of such an interface since features like spatialised auditory beacons and waypoint based navigation depend almost exclusively on orientation of a user to provide accurate information. Future investigations could involve more complex, life-like environments possibly simulating certain scenarios.

We envision a range of uses for the wearable device that has been tested here. Augmenting the visual and auditory channels without interfering with their natural functions is likely to aid in critical tasks such as urban search and rescue, driving and firefighting. These are likely to benefit significantly from the use of non-visual interaction mediated by the BCH, with the visual faculty being engaged in a more context sensitive manner. Mundane and time consuming activities such as searching for lost keys, as we have demonstrated, can also be made easier with the aid of a wearable device incorporating auditory and visual feedback.

Chapter V

5. CONCLUSIONS AND FUTURE WORK

This thesis has investigated the possibility of bi-modal information delivery for a wearable interface. The key questions addressed in this thesis were outlined in Chapter 1, and include the advantages of using sound in wearable interfaces (Q1), replication of natural auditory perception (Q2), limitations of visual information displays (Q3), proposed solutions to address the limitations of visual information displays (Q4) and whether these solutions adequately address the issues raised in Question 3 (Q5). These questions were dealt with in a methodical manner beginning with a thorough review of auditory and visual perception in Chapter 2. Additionally, have also looked at how individual transfer characteristics of subjects' ears can be measured.

A comprehensive review of wearable interfaces has been provided with timelines (Figure 2.19 – 2.21) to clearly lay out the progress of this technology. These timelines present not only the evolution of wearable interfaces through the 20th century and beyond, but also show us how the approaches to designing these interfaces have evolved over time. The most important finding from this review, and the one that this thesis hinges on, has been the approach used for information delivery on wearable devices. We see that most interfaces employ an 'either/or' approach to information delivery. By this we mean that the interfaces are, for the most part, designed to use either visual feedback or auditory feedback as the primary mechanism to deliver information. Those that make use of both mechanisms tend to relegate the use of the auditory faculty to provide simple alerts. It seems presumptuous of designers of such interfaces that only the visual faculty can be fully engaged to deliver information, or that the auditory faculty somehow addresses all the problems encountered with visual interfaces. Some of the work covered in Chapter 2 clearly demonstrates that spatialised audio can be used in a bi-modal interface to deliver information [78] [101].

Based on the reviews in Chapter 2 and the research approach laid out in Chapter 1, we have been able to answer Question 6 and identify where these interfaces are lacking and propose an alternate solution; the use of a bone conduction headset (BCH) as a potential wearable information delivery device as part of a wearable interface. The BCH meets the key requirements of leaving an unobstructed air-conduction pathway and being relatively unobtrusive. Our choice of this information delivery approach is informed by studies that examined its function in typical auditory perception studies [16], virtual environments [14] [103] and real environments [15].

Chapter 3 addressed the question of testing and evaluating binaural spatialisation using a bone conduction headset – Question 7. This was accomplished by running psychoacoustic studies to establish that binaural spatialisation over a BCH is possible, and to evaluate the accuracy of localisation afforded by the BCH. Our studies differ to those carried out earlier in two ways. Firstly, we have tested the headset in ‘regular room’ environments. Secondly, our application based studies have demonstrated good usability of the wearable interface in at least one real-world scenario. These approaches were adopted keeping in mind the ecologically valid methodology laid out in Chapter 1. Our psychoacoustic studies demonstrate that perception of binaurally spatialised sound sources is, indeed, possible over a bone conduction headset. Studies 1 and 2 show strong alignment in localisation accuracy and externalisation with one of the few known, comprehensive studies on binaural spatialisation over a bone conduction headset [16]. Similarly, results from the third experiment compare favourably with those obtained in similar real-world [61] and video see-through AR based studies [66].

Chapter 4 addressed Question 8 – how the proposed prototype functioned in real-world environments. This chapter presents two studies that looked to evaluate the use of the proposed bi-modal wearable interface. The interface prototype was put together by pairing a bone conduction headset with two existing wearable computers with small screen visual displays – the Recon Jet and Google Glass. Results from these studies show that the bi-modal form of delivery works well. In the first study [129], a dynamic, spatialised auditory cue was successfully used to redirect the user’s attention to a point in the real-world. Reaction time comparisons to the visual and auditory cues were closely matched. The participants also appeared to take roughly the same time to reach the target for both cueing conditions. These results demonstrates that auditory and visual information appears to work equally well when information with a similar intended outcome is delivered, i.e., in the case of this experiment, redirecting attention to a point in the real-world.

In the second study we explored the use of a spatialised auditory beacon which participants could home in on as part of a search task. During the dual-display condition (dynamic arrow and stationary spatialised auditory beacon) we noticed that some participants, who appeared to primarily use the auditory beacon, periodically glanced at the visual display to confirm that they were moving in the right direction. This study also shows us that there is merit to the idea of using an auditory beacon in a complex environment. This strategy leaves the visual faculty free to help the participant navigate the environment. Conversely, visually displaying the location of the target within a complex environment prior to entering it ensures that the participants are able to navigate the environment using the shortest possible path without having to look at it again.

Both of these experiments demonstrate that if well designed, a wearable, bi-modal interface can effectively exploit the visual short term memory (VSTM) capacity for three to four objects [161] or the temporal nature of an auditory cue to provide the desired information.

5.1 Contributions

The key contributions resulting from the work presented in this thesis are:

- A strong case for bi-modal (audio-visual) information presentation in wearable interfaces.
- Demonstrated viability of using spatialised auditory feedback in a real-world scenario where vision is already occupied.
- A strong case for a contextually aware wearable interface that is capable of choosing a display channel – auditory or visual – based on the user’s activity and environment.
- The design of a wearable interface that is unobtrusive in both form and function.

The work presented in this thesis has demonstrated the potential for bi-modal information presentation in wearable interfaces. In particular, the use of spatialised auditory feedback was shown to provide better localisability, whether for objects in the real-world outside the field of view, or those hidden in plain view. Our studies exploring binaural spatialisation over a bone conduction headset demonstrate that a good level of accuracy and externalisation can be obtained using off-the-shelf components. While the phenomenon of binaural spatialisation over a bone conduction headset itself is not novel, our approach of using existing technology is. Unlike previous studies, we have used a freely available binaural plugin encompassing ‘drag and drop’ features for use with an existing freely available game development platform.

This approach, along with the environments in which the experiments have been carried out, lends great ecological validity to our findings.

Through our studies we have been able to demonstrate that a bi-modal system that engages the visual and auditory faculties appropriately works significantly better than an interface that engages only a single faculty. Our contribution to the field in this respect is novel in that it has addressed the issue of information presentation in a unique manner. The use of bilaterally applied bone conduction transducers for a wearable interface, to our knowledge is unheard of outside the fields of audiology and some studies exploring navigation aids for visually impaired persons [14] [15] [114]. Despite not being able to able to implement the functionality of contextual awareness in our interface, we have demonstrated in Study 4 the usefulness of implementing the auditory faculty for information delivery in the event that the visual faculty is occupied. We envision future iterations of wearable interfaces to be equipped with the capability to select an information delivery channel based on a number of environmental parameters. More importantly, we have been able to design and test a fully functioning prototype of our wearable interface that meets two out of the three criteria laid out in Chapters 1 and 2, and revisited in the next section.

5.2 Limitations

In the beginning of this thesis, we specified three important criteria for effective wearable interfaces.

- Bi-modal feedback (visual and auditory) capabilities.
- Mobile and Unobtrusive
- Contextually Aware

Of these three criteria, our prototype meets the first two. It incorporates a bi-modal feedback mechanism and is relatively unobtrusive. However, we have been unable to address the third criterion, that of a contextually aware interface. We feel that this is an extremely important area to address in the future. Contextual awareness is a powerful tool that has the capability to greatly increase interface usability. Being able to sense the environment users are in and their responses to the environment – heart rate, pupil dilation etc. – to make use of the appropriate information display apparatus may help increase user task efficiency. In addition to this, a contextually aware interface is also capable of addressing safety issues. A combination of inadequate technical resources and programming expertise, combined with a lack of time did not allow for the implementation of this functionality in the interface tested as part of this thesis.

Some of the other limitations of the work described in this thesis are:

- Limited hardware testing: Only two bone conduction headsets could be tested prior to commencing the studies. The qualitative difference between two BCHs was the primary consideration for having selected one over the other. The generational differences between the headsets resulted in a noticeable difference between the sound qualities. A few other BCHs became commercially available during the course of this thesis, but for the sake of consistency, using these in any of the studies was not feasible.
- Limited number of wearable interfaces: Only two wearable interfaces with visual displays – Recon Jet and Google Glass – were available for testing during the course of the study. Using both these interfaces, one for each application based experiment, could be looked upon as a drawback. We would have liked to run both the studies listed in Chapter 4 using the two devices for the sake of uniformity. This would have also allowed us to directly compare how the auditory and visual cues functioned across both interfaces for the same environmental conditions.
- Participants: Since all the experiments were run at the University of Canterbury, a sizable population of participants who took part in the experiments were young students with an average age of 26 years. This is not representative of the general population outside the academic environment. There also appeared to be skewed gender ratio, with male participants outnumbering female participants 2:1. While concerted efforts were made to involve a larger number of female participants to help balance the gender ratio among participants, this did not always prove successful.
- Varying degrees of auditory and visual acuity amongst the sample population: Carrying on from the previous point, the sample population that participated in these studies was more likely to have better visual and auditory acuity. This could lead to a skewed perception of the effectiveness of the wearable interface tested here.
- Testing environments: Considering the envisioned uses for the wearable interface tested here, not being able to trial it in outdoor environments has not allowed us to fully validate its usability across a range of scenarios. This is an important point, since the use of the BCH has been proposed on the premise that it allows for the natural

perception of the acoustic environment. Not being able to examine its functioning in a real-world outdoor environment limits the extent to which the results obtained in these studies can be applied across different scenarios.

- Lack of optimisation: In keeping with the ecologically valid approach, no optimisation was performed on any of the hardware components used in the experiments. This is especially true of the BCH. No attempt was made to optimise the auditory cues for presentation over the BCH keeping its frequency response in mind.

In addition to the limitations described above, another important factor that limits the scope of the results obtained here is the availability and familiarity with wearable interfaces. Considering that wearable interfaces are in their infancy, their relative inaccessibility limited the familiarity participants had with such devices. The ability to truly evaluate interactions mediated by an interface lie in comparisons between those that users expect, and those that are designed based on an existing body of work in the field. With wearable interfaces forecast to become ubiquitous over the coming years, this will be a premise that can be tested in the near future.

5.3 Future Work

There are many directions for future work to continue the research undertaken in this thesis. One important aspect of our work that we would like address in the future, is more testing and validation. While the two application-based studies have demonstrated good results, we would like to test our prototype in more real-world environments. Such testing will help validate the design and implementation approaches we have utilised in conception and implementation of the prototype. There is also a need for testing the prototype in an outdoor environment. To validate the choice of a BCH as an auditory display device in a bi-modal wearable interface, its usability must be tested in a range of outdoor and indoor settings. Since some of the intended users of this interface involve emergency service personnel, it would be prudent to test the interface in environments that they regularly encounter; such as urban disaster response, search and rescue, urban firefighting etc. Auditory and visual cueing conditions, similar to those seen in Chapter 4, can be re-tested in such scenarios to evaluate their effectiveness. Objective measures such as task completion time, target acquisition time and path selection can be combined with subjective feedback from users to refine the design of information delivery on such wearable interfaces.

We would also like to develop future wearable prototypes with the addition of computer vision and bio-feedback mechanisms. Contextual awareness could be implemented using computer vision technology, while the addition of bio-feedback could have varying applications from military to domestic search and rescue. Computer vision could potentially be used for areas outside of the visual field. Redirecting attention using auditory cues to visual points of interests outside the users FOV identified by a computer vision system could greatly enhance safety in a range of scenarios. Wearable bio-feedback mechanism could be trained to take into account the emotional state of the user before presenting them with information. Such systems also have potential uses in the health and palliative care industries. A wearable system that monitors its user's physical and emotional state could potentially be used to provide them with visual and auditory stimulation at critical periods of physical or emotional instability.

Future wearable interfaces could also incorporate features that exploit the potential of Augmented Reality (AR). A review of some wearable interfaces in Chapter 2 and the second study in Chapter 4 have shown that AR is a powerful medium that holds immense promise for real-world applications. Being able to effectively combine auditory and visual representations to present an augmented perspective of the world has innumerable applications. The research into such bi-modal interfaces and the degree to which they are usable will depend almost exclusively on evaluating their functionality in the real-world. Building on existing interaction paradigms, research into wearable AR interfaces hinges on demonstrating their viability in an ever-changing, digital world.

Besides the obvious objective conclusions resulting from studies, it is also important to analyse findings in a subjective and social context. Far too often, designers fail to gauge the social implications of interfaces they have designed. Besides just 'usability testing', it would be worthwhile to look at how social interactions are influenced when these interfaces are used. I say this, because having been at the intersection of technology and art, like Walter Benjamin I believe that our issue lies not in development of technology, but its integration into our social fabric.

"*...society has not been mature enough to incorporate technology as its organ, that technology has not been sufficiently developed to cope with the elemental forces of society.*" [162]

- *Walter Benjamin, The Work of Art In the Age of Mechanical Reproduction*

VI

6. REFERENCES

[1] Center, P.R. *Mobile Fact Sheet*. 2017 [cited 2018 8 January]; Available from: **http://www.pewinternet.org/fact-sheet/mobile/**.

[2] Zealand, R.N., *A Report on a Survey of New Zealanders' Use of Smartphones and other Mobile Communication Devices 2015*, R.N. Zealand, Editor 2015: New Zealand.

[3] WHO, *Mobile Phone Use: A growing Problem of Driver Distraction*, in *Deacade of Action for Road Safety* 2011, World Health Organization.

[4] DOT. *Distraction: Facts and Statictics*. 2015 [cited 2016 3rd Feb]; Available from: **http://www.distraction.gov/stats-research-laws/facts-and-statistics.html**.

[5] CDC. *Injury Prevention & Control: Motor Vehicle Safety*. 2015 [cited 2016 3rd Feb]; Available from: **http://www.cdc.gov/motorvehiclesafety/distracted_driving/**.

[6] Sawhney, N. and C. Schmandt. *Design of Spatialized Audio In Nomadic Environments*. in *International Conference on Auditory Display*. 1997. Palo Alto, California.

[7] Gibbs, S. *Google Pixel Buds Review: Bluetooth earbuds are a missed opportunity*. 2017 [cited 2017 27 December]; Available from: https://**www.theguardian.com/technology/2017/dec/04/google-pixel-buds-review-bluetooth-earbuds-headphone-apple-airpods-translation**.

[8] Amazon. *Echo & Alexa Devices*. 2015 [cited 2017 7 - 10]; Available from: https://**www.amazon.com/Amazon-Echo-And-Alexa-Devices/b/ref=sv_devicesubnav_0?ie=UTF8&node=9818047011**.

[9] Bragi. *The Dash*. 2015 [cited 2016 28 December]; Available from: https://**www.bragi.com/thedash/**.

[10] Labs, D. *Here One*. 2016 [cited 2016 28 December]; Available from: https://hereplus.me/about#.

[11] Mynatt, E.D., et al. *Designing Audio Aura*. in *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems*. 1998. ACM Press/Addison-Wesley Publishing Co.

[12] Sawhney, N. and C. Schmandt. *Nomadic Radio: Scaleable and Contextual Notification For Wearable Audio Messaging*. in *SIGCHI conference on Human Factors in Computing Systems*. 1999. ACM.

[13] Walker, A. and S. Brewster, *Spatial Audio In Small Screen Device Displays.* Personal Technologies, 2000. **4**(2-3): p. 144-154.

[14] Walker, B.N. and J. Lindsay, *Navigation Performance In A Virtual Environment With Bonephones*, in *International Conference On Auditory Displays*2005: Limerick, Ireland.

[15] Wilson, J., et al. *SWAN: System for Wearable Audio Navigation*. in *1th IEEE International Symposium on Wearable Computers*. 2007. IEEE.

[16] Stanley, R.M., *Measurement And Validation Of Bone Conduction Adjustment Functions In Virtual 3D Audio Displays*, in *School of Psychology*2009, Georgia Institute of Technology.

[17] Sawhney, N. and C. Schmandt, *Nomadic Radio: Speech and Audio Interactions For Contextual Messaging In Nomadic Environments.* ACM Transactions on Computer-Human Interaction (TOCHI), 2000. **7**(3): p. 353–383.

[18] Doherty, A.L., *Blue-sky Eruptions, Do They Exist? Implications for Monitoring New Zealand's Volcanoes.*, in *Geology*2009, University of Canterbury: Christchurch. p. 181.

[19] Barfield, W., M. Cohen, and C. Rosenberg, *Visual and Auditory Localization as a Function of Azimuth and Elevation.* The International Journal of Aviation Psychology, 1997. **7**(2): p. 123 - 138.

[20] Webster, M.-. *Definition of Psychoacoustics*, 1948.

[21] Lin, Y. and W.H. Abdulla, *Principles of Psychoacoustics*, in *Audio Watermark: A Comprehensive Foundation Using MATLAB*. 2015, Springer. p. 15 - 49.

[22] Yost, W.A., *Fundamentals of Hearing*, in *Fundamentals of Hearing: An Introduction*. 1994.

[23] Heffner, H.E. and R.S. Heffner, *Hearing Ranges of Laboratory Animals.* Journal of the American Association for Laboratory Animal Science, 2007. **46**(1): p. 20 – 22.

[24] Jackson, L.L., R.S. Heffner, and H.E. Heffner, *Free-field Audiogram of the Japanese macaque (Macaca fuscata).* Journal of the Acoustical Society of America, 1999. **106**(5): p. 3017 - 3023.

[25] University, S.F. *Threshold of Pain*. Handbook For Acoustic Ecology 1999 [cited 2017 26 - 08]; Second:[Available from: **http://www.sfu.ca/sonic-studio/handbook/Threshold_of_Pain.html**.

[26] Laboratory, N.P. *Acoustics*. 2006 [cited 2017 10 December]; Available from: **http://www.npl.co.uk/educate-explore/factsheets/acoustics/**.

[27] Communications, A., *Introduction to Audio: Acoustics, Speakers and Audio Terminology*, A. Communications, Editor 2017, Axis Communications.

[28] Bloom, P.J., *Creating Source Elevation Illusions by Spectral Manipulation.* Journal of the AES, 1977. **25**(9): p. 560 - 565.

[29] Cheng, C.I. and G.H. Wakefield, *Introduction to Head Related Transfer Functions: Representations of HRTFs in Time, Frequency and Space.* Journal of the AES, 2001. **49**(4): p. 231 - 249.

[30] Langendijk, E.H.A. and A.W. Bronkhorst, *Contribution Of Spectral Cues To Human Sound Localization.* Journal of the Acoustical Society of America, 2002. **112**(4): p. 1583 - 1596.

[31] Rayleigh, L. *On Our Perception of the Direction of a Source of Sound*. in *Proceedings of the Musical Association*. 1875. Taylor & Francis, Ltd.

[32] Rayleigh, L. *On the Perception of the Direction of Sound*. in *Proceedings of the Royal Society of London.* 1909. The Royal Society.

[33] Middlebrooks, J.C. and D.M. Green, *Sound Localization by Human Listeners.* Annual Review of Psychology, 1991. **42**(1): p. 135 - 159.

[34] Wightman, F.L. and D.J. Kistler, *Hearing in Three Dimensions: Sound Localization*, in *AES 8th International Conference: The Sound of Audio*1990.

[35] Hartmann, W.M., *Localization of a Sound Source in a Room*, in *AES 8th International Conference: The Sound of Audio*1990. p. 27 - 32.

[36] Litovsky, R.Y., *Binaural Hearing.* 2008.

[37] Kapralos, B., M.R. Jenkin, and E. Milios, *Auditory Perception and Spatial Auditory Systems.*

[38] Wang, D. and G.J. Brown, *Binaural Sound Localisation*, in *Computational Auditory Scene Analysis*. 2005, John Wiley & Sons, Inc. p. 1 - 27.

[39] Wagenaars, W.M., *Localization of Sound in a Room with Reflecting Walls.* Journal of the AES, 1990. **38**(3): p. 99 - 110.

[40] Rodgers, C.A.P., *Pinna Transformations and Sound Reproduction.* Journal of the AES, 1981. **29**(4): p. 226 - 234.

[41] Shanna, B., *Interaural Level Difference,* 2010, Scottish Sensory Centre.

[42] Stevens, S.S. and E.B. Newman, *The Localization of Actual Sources of Sound.* The American Journal of Psychology, 1936. **48**(2): p. 297 - 306.

[43] Moore, B.C.J., *Controversies and Mysteries in Spatial Hearing*, in *AES 16th International Conference: Spatial Sound Reproduction*1999.

[44] Wallach, H., E.B. Newman, and M.R. Rosenzweig, *The Precedence Effect in Sound Localization.* The American Journal of Psychology, 1949. **62**(3): p. 315 - 336.

[45] Hass, H., *The Influence of a Single Echo on the Audibility of Speech.* Journal of the AES, 1972. **20**(2): p. 146 - 159.

[46] Oldfield, S.R. and S.P.A. Parker, *Acuity of Sound Localization: A Topography of Audiotory Space. I. Normal Hearing Conditions.* Perception, 1984. **13**: p. 581 - 600.

[47] Shinn-Cunningham, B. *Localizing Sound in Rooms*. in *ACM SIGGRAPH*. 2001. Snowbird, Utah, USA.

[48] Campbell, R., *Monaural and Binaural Level Cues: A Behavioural and Physiological Investigation*, 2006, University of Oxford. p. 293.

[49] Blauert, J., *Sound Localization In The Median Plane.* Acustica, 1969. **22**: p. 205 - 213.

[50] Oldfield, S.R. and S.P.A. Parker, *Acuity of Sound Localization: A Topography of Audiotory Space. II. Pinna Cues Absent.* Perception, 1984. **13**: p. 601 - 617.

[51] Zacharov, N., O. Tuomi, and G. Lorho, *Auditory periphery, HRTF's and directional loudness perception*, in *110th Convention of the AES*2001: Amsterdam, The Netherlands.

[52] Wightman, F.L. and D.J. Kistler, *Headphone Simulation Of Free Field Listening. II: Psychophysical Validation.* The Journal of the Acoustical Society of America, 1989. **85**(2): p. 868 - 878.

[53] Masiero, B., M. Pollow, and J. Fels. *Design of a fast broadband individual head-related transfer function measurement system*. in *Forum Acusticum*. 2011.

[54] Jo, H., Y. Park, and W.L. Martens, *Evaluating Candidate Sets Of Head Related Transfer Fucntions For Control Of Virtual Source Elevation*, in *AES 40th International Conference*2010, AES: Tokyo, Japan.

[55] Wenzel, E.M., F.L. Wightman, and D.J. Kistler, *Localization with non-individualized virtual acoustic display cues*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1991, ACM: New Orleans, Louisiana, USA. p. 351-359.

[56] Wenzel, E.M., et al., *Localization using nonindividualized head-related transfer functions.* The Journal of the Acoustical Society of America, 1993. **94**(1): p. 111-123.

[57] Futoshi Asano, Y.S., Toshio Sone, *Role Of Spectral Cues In Median Plane Localization* The Journal of the Acoustical Society of America, 1990. **88**(1): p. 159 - 168.

[58] Perrott, D.R., et al., *Aurally Aided Visual Search in the Central Visual Field: Effects of Visual Load and Visual Enhancement of the Target.* Human Factors, 1991. **33**(4): p. 389 - 400.

[59] Gibson, J.J., *The Ecological Approach To Visual Perception*. 1979.

[60] Witkin, H.A., S. Wapner, and T. Levanthal, *Sound Localization With Conflicting Visual and Auditory Cues.* Journal of Experimental Psychology, 1952. **43**(1): p. 58 - 67.

[61] Jackson, C.V., *Visual Factors in Auditory Localization.* Quarterly Journal of Experimental Psychology, 1953. **5**(2): p. 52 - 65.

[62] Noodles, N.C.-A., *Human eye—Horizontal Field of View & Human eye —Vertical Field of View*, 2016, Medium.

[63] Heesy, C.P., *Seeing In Stereo: The Ecology and Evolution Of Primate Binocular Vision And Stereopsis.* Evolutionary Anthropology, 2009. **18**(1): p. 21-35.

[64] Drascic, D. and P. Milgram. *Perceptual Issues in Augmented Reality*. in *Stereoscopic Displays and Virtual Reality Systems III*. 1996. San Jose, California, USA.

[65] Loomis, J.M. and J.M. Knapp, *Visual Perception Of Egocentric Distance In Real And Cirtual Environments*, in *Virtual And Adaptive Environments: Applications, Implications and Human Performance Issues*, L.J. Hettinger and M. Haas, Editors. 2003, Lawrence Erlbaum Associates. p. 21-45.

[66] Kyto, M., K. Kusumoto, and P. Oittinen, *The Ventriloquist Effect In Augmented Reality*, in *ISMAR*2015: Fukuoka, Japan.

[67] Dicke, C., *A Holistic Design Concept For Eyes-Free Mobile*, in *Computer Science and Software Engineering*2012, University of Canterbury: Christchurch, New Zealand.

[68] Schafer, R.M., *Notation*, in *The Soundscape: Our Sonic Environment And The Tuning Of The World*. 1994, Destiny Books: United States of America. p. 128.

[69] Starner, T., et al., *Augmented Reality Through Wearable Computing.* Presence: Teleoperators and Virtual Environments, 1997. **6**(4): p. 386-398.

[70] Billinghurst, M. and T. Starner, *Wearable Devices: New Ways To Manage Information.* Computer, 1999. **32**(1): p. 57-64.

[71] Dong, J., et al., *Wearable Computing Device with Indirect Bone-Conduction Speaker* 2013.

[72] Heinrich, M.J., et al., *Wearable Computing Device*, in *Google Patents*, USPTO, Editor 2013, Google Inc.: U.S.A.

[73] Motorola, I. *moto 360: A Watch For Our Times*. 2014 [cited 2015 25th February].

[74] Dicke, C., J. Sodnik, and M. Billinghurst, *Spatial Auditory Interfaces Compared to Visual Interfaces for Mobile Use in a Driving Task*, in *Ninth International Conference on Enterprise Information Systems (ICEIS)*2007: Funchal, Madeira, Portugal. p. 282 - 285.

[75] Hull, R., J. Reid, and E. Geelhoed, *Creating Experiences With Wearable Computing.* IEEE Pervasive Computing, 2002. **1**(4): p. 56-61.

[76] Billinghurst, M., S. Weghorst, and T.F. III. *Wearable Computers For Three Dimensional CSCW*. in *International Symposium on Wearable Computers*. 1997. IEEE Computer Society.

[77] Sutherland, I.E. *A Head-Mounted Three Dimensional Display*. in *Fall Joint Computer Conference*. 1968. ACM.

[78] Billinghurst, M., et al., *Spatial Information Displays On A Wearable Computer*. Computer Graphics and Applications, IEEE, 1998. **18**(6): p. 24-31.

[79] Hollerer, T., S. Feiner, and J. Pavlik. *Situated Documentaries: Embedding Multimedia Presentations in The Real World*. in *International Symposium On Wearable Computers*. 1999. San Francisco, CA, USA: IEEE.

[80] Walker, A., et al. *Diary In The Sky: A Spatial Audio Display For A Mobile Calendar*. in *People and Computers XV—Interaction without Frontiers*. 2001. Springer London.

[81] Holland, S., D.R. Morse, and H. Gedenryd, *AudioGPS: Spatial Audio In A Minimal Attention Interface*. Personal and Ubiquitous Computing, 2002. **6**(4): p. 253-259.

[82] Hiipakka, J. and G. Lorho. *A Spatial Audio User Interface For Generating Music Playlists*. in *Proceedings of the 2003 International Conference on Auditory Display*. 2003. Boston, MA, USA.

[83] Harma, A., et al., *Augmented Reality Audio For Mobile And Wearable Appliances*. Journal of the Audio Engineering Society, 2004. **52**(6): p. 618-139.

[84] Tikander, M., M. Karjalainen, and V. Riikonen. *An Augmented Reality Audio Headset*. in *11th Int. Conference on Digital Audio Effects (DAFx-08)*. 2008. Espoo, Finland.

[85] Villegas, J. and M. Cohen, *GABRIEL: Geo-Aware Broadcasting For In-Vehicle Entertainment And Localizability*, in *AES 40th International Conference*2010, AES: Tokyo, Japan.

[86] Katz, B.F.G., et al., *NAVIG: Navigation Assisted by Artificial Vision and GNSS*, in *Workshop on multimodal location based techniques for extreme navigation (Pervasive 2010)*2010: Helsinki, Finland.

[87] Levy, S. *Google Glass 2.0 Is A Startling Second Act*. 2017 [cited 2017 31 - 08]; Available from: https://**www.wired.com/story/google-glass-2-is-here/**.

[88] Microsoft. *HoloLens*. 2015 [cited 2015 April 08]; Available from: **http://www.microsoft.com/microsoft-hololens/en-us**.

[89] Bragi. *Bragi Dash Pro*. 2017 [cited 2017 30 - 08]; Available from: https://**www.bragi.com/thedashpro/**.

[90] Labs, D. *Here Active Listening*. 2015 [cited 2016 6 Feb]; Available from: https://**www.hereplus.me/**.

[91] Krueger, M.W., *Responsive Environments*, in *National Computer Conference*1977.

[92] Uyeda, L.R. and J.A. Emery, *Information Display For Diver's Face Mask*, 1973.

[93] Crost, M.E., et al., *Apparatus For Adding Electronic Display Information To A Night Vision Goggle Display*, 1976, The United States of America as represented by the Secretary of the Army.

[94] Evans, C.D., et al., *Compact Helmet Mounted Display*, U.S.P.a.T. Office, Editor 1988, Kaiser Aerospace and Electronics Corporation: United States of America.

[95] Bettinger, D.S., *Spectacle-Mounted Ocular Display Apparatus*, 1989.

[96] Cheysson, F. and J.B. Migozzi, *Wide Field High Optical Efficiency Display Device*, 1989, Thomson CSF.

[97] Kubik, J.C., *Headwear-Mounted Periscopic Display Device*, 1989, Iota Instrumentation Company: United States of America.

[98] Becker, A., *Miniature Video Display System*, 1990.

[99] Caudell, T.P. and D.W. Mizell. *Augmented reality: An Application of Heads-Up Display Technology To Manual Manufacturing Processes*. 1992. IEEE.

[100] Feiner, S., B. Macintyre, and D. Seligmann, *Knowledge-Based Augmented Reality*. Communications Of The ACM, 1993. **36**(7): p. 53-62.

[101] Billinghurst, M., et al. *A Wearable Spatial Conferencing Space*. in *Second International Symposium On Wearable Computers*. 1998. IEEE.

[102] Vazquez-Alvarez, Y. and S.A. Brewster. *Designing Spatial Audio Interfaces To Support Multiple Audio Streams*. in *12th International Conference On Human Computer Interaction With Mobile Devices And Services*. 2010. Lisboa, Portugal.

[103] Valjamae, A., et al., *Binaural Bone Conducted Sound In Virtual Environments: Evaluation Of A Portable, Multimodal Motion Simulator Prototype.* Acosutical Science And Technology, 2008. **29**(2): p. 149 - 155.

[104] MacDonald, J.A., P.P. Henry, and T.R. Letowski, *Spatial Audio Through A Bone Conduction Interface.* International Journal Of Audiology, 2006. **45**(10): p. 595 - 599.

[105] Mills, A.W., *On The Minimum Audible Angle.* Journal of the Acoustical Society of America, 1958. **30**(4): p. 237 - 246.

[106] Rayleigh, L. *On the Amplitude of Sound Waves*. in *Proceedings of the Royal Society of London*. 1877.

[107] Lindeman, R.W., H. Noma, and P.G.d. Barros, *An Empirical Study of Hear-Through Augmented Reality - Using Bone Conduction To Deliver Spatialized Audio*, in *IEEE Virtual Reality*2008, IEEE: Reno, Nevada, USA.

[108] Topping, A. *Listen up: wearing headphones can endanger life, study finds*. 2012 [cited 2017 28 - 08]; Available from: https://**www.theguardian.com/technology/2012/jan/16/headphones-can-endanger-life-study**.

[109] Staff, L.S. *Pedestrians & Headphones Don't Mix (Infographic)*. 2012 [cited 2017 28 - 08]; Available from: https://**www.livescience.com/17995-accidents-pedestrians-headphones-infographic.html**.

[110] Kondo, K., N. Anazawa, and Y. Kobayashi, *Comparison of Output Devices for Augmented Audio Reality*. IEICE TRANSACTIONS on Information and Systems, 2014. **97**(8): p. 2114-2123.

[111] ReconInstruments. *Recon Jet*. 2015 [cited 2015 01 August 2015]; Available from: **http://www.reconinstruments.com/products/jet/**.

[112] Brewster, S. and A. Walker, *Non-Visual Interfaces For Wearable Computers.* COLLOQUIUM DIGEST-IEE, 1999.

[113] Walker, B.N. and R.M. Stanley, *Thresholds Of Audibility For Bone Conduction Headsets*, in *International Conference On Auditory Display*2005, ICAD: Limerick, Ireland.

[114] Walker, B.N. and J. Lindsay, *Navigation Performance With A Virtual Auditory Display: Effects Of Beacon Sound, Capture Radius and Practice.* Human Factors: The Journal of the Human Factors and Ergonomics Society, 2006. **48**(2): p. 265 - 278.

[115] Nelson, W.T., et al., *Effects of Localized Auditory Information on Visual Target Detection Performance Using a Helmet-Mounted Display*. Human Factors, 1998. **40**(3): p. 452 - 460.

[116] Thakur, A. and V. Nair, *3Dception*, 2015, Two Big Ears: Edinburg, Scotland.

[117] AfterShokz. *Sportz3*. [cited 2015 2nd June]; Available from: **http://aftershokz.com/collections/wired/products/sportz-3**.

[118] Devana, A. *3D Audio: Weighing The Options*. 2015 [cited 2015 20 - 04]; Available from: **http://designingsound.org/2015/04/3d-audio-weighing-the-options/**.

[119] Zhao, H., et al. *Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study*. in *Tenth Meeting of the International Conference on Auditory Display*. 2004. Sydney, Australia.

[120] Wersenyi, G., *Localization In A HRTF Based Minimum Audible Angle Listening Test On a 2D Sound Screen For GUIB Applications*, in *115th Convention of the Audio Engineering Society* 2003.

[121] Barde, A., et al., *Binaural Spatialisation over a Bone Conduction Headset: Minimum Discernable Angular Difference*, in *140th Convention of the AES* 2016, AES: Paris, France.

[122] Gardner, W.G., *3D Audio Using Loudspeakers*, in *School of Architecture and Planning*1997, Massachusetts Institute of Technology.

[123] Weinrich, S.G., *Horizontal Plane Localization Ability and Response Time as a Function of Signal Bandwidth*, in *98th Convention of the AES*1995: Paris, France.

[124] Wightman, F.L. and D.J. Kistler, *Headphone Simulation Of Free Field Listening. I: Stimulus Synthesis.* Journal of the Acoustical Society of America, 1989. **85**(2): p. 858 - 867.

[125] Evans, M.J., *Obtaining Accurate Responses in Directional Listening Tests*, in *104th Convention of the AES*1998, AES: Amsterdam, The Netherlands.

[126] III, J.J.A. and A.M. Chiu, *An Effectiveness Study Of A CAD System Augmented By Audio Feedback.* Computer and Graphics, 1977. **2**(4): p. 231-233.

[127] Wobbrock, J.O., et al. *The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures*. in *ACM Conference on Human Factors in Computing Systems (CHI '11)*. 2011. Vancouver, British Columbia: ACM Press.

[128] Barde, A., et al., *Attention Redirection Using Binurally Spatialised Cues Over a Bone Conduction Headset*, in *2016 HFES Annual Meeting*2016: Washington D.C., USA.

[129] Barde, A., et al., *A Bone Conduction Based Spatial Auditory Display As Part of a Wearable Hybrid Interface*, in *22nd International Conference on Auditory Display (ICAD 2016)*2016: Canberra, Australia.

[130] Walker, B.N. and J. Lindsay, *Development And Evaluation Of A System For Wearable Audio Navigation*, in *Human Factors And Ergonomics Society 49th Annual Meeting*2005.

[131] Parseihian, G., A. Brilhault, and F. Dramas, *NAVIG: An Object Localization System For The Blind.*, in *8th International Conference on Pervasive Computing* 2010: Helsinki, Finland

[132] Lokki, T., et al., *Application Scenarios of Wearable And Mobile Augmented Reality Audio*, in *Audio Engineering Society Convention 116*2004, AES: Berlin, Germany.

[133] Walker, B.N. and R.M. Stanley. *Evaluation of Bone-Conduction Headsets for Use in Multitalker Communication Environments*. in *49th Meeting of the Human Factors and Ergonomics Society*. 2005.

[134] Lindeman, R.W., H. Noma, and P.G.d. Barros, *Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction To Display Spatialized Audio.* 2007.

[135] Zhou, Z., et al., *The Role of 3-D Sound in Human Reaction and Performance in Augmented Reality Environments.* IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2007. **37**(2): p. 262 - 272.

[136] Kanaya, S. and K. Yokosawa, *Perceptual Congruency of Audio-Visual Speech Affects Ventroloquism with Bilateral Visual Stimuli.* Psychonomic bulletin & review, 2011. **18**(1): p. 123 - 128.

[137] Alais, D. and D. Burr, *The Ventriloquist Effect Results From Near-Optimal Bimodal Integration.* Current Biology, 2004. **14**(3): p. 257 - 262.

[138] Roffler, S.K. and R.A. Butler, *Localization Of Tonal Stimuli In The Vertical Plane.* Journal of the Acoustical Society of America, 1968. **43**(6): p. 1260 - 1266.

[139] Hairston, W.D., et al., *Visual Localization Ability Influences Cross-Modal Bias.* Journal of Cognitive Neuroscience, 2003. **15**(1): p. 20 - 29.

[140] ART. *ARTTRACK2*. 1999 [cited 2015 1st Dec]; Available from: **http://www.ar-tracking.com/products/discontinued/arttrack2/**.

[141] Martens, W.L., A. Guru, and D. Lee, *Effects of individualised headphone response equalization on front-back hemifield discrimination for virtual sources displayed on the horizontal plane*, in *20th International Congress on Acoustics (ICA)*2010.

[142] TrackR. *TrackR Bravo*. 2015 [cited 2017 30 - 08]; Available from: https://**www.thetrackr.com/bravo**.

[143] Wilkins, P.A. and W.I. Acton, *Noise and accidents—a review.* Annals of Occupational Hygiene, 1982. **25**(3): p. 249 - 260.

[144] Lichenstein, R., et al., *Headphone Use and Pedestrian Injury and Death in the United States: 2004 - 2011.* Injury Prevention, 2012.

[145] Rudmann, D.S. and T.Z. Strybel, *Auditory Spatial Facilitation of Visual Search Performance: Effects of Cue Precision and Distractor Density.* Human Factors, 1999. **41**(1): p. 146 - 160.

[146] Bolia, R. and W.T. Nelson, *Spatial Audio Displays for Target Acquisition and Speech Communications*, in *Virtual And Adaptive Environments: Applications, Implications and Human Performance Issues*, L.J. Hettinger and M. Haas, Editors. 2003. p. 187 - 197.

[147] ART, *DTrack*, 1999, Advanced Realtime Tracking: Munich, Germany.

[148] Technologies, S. *Steradian S7-X Laser Tag Gun*. 2003 [cited 2015 12 Sept]; Available from: **http://www.steradiantech.com/xseries/s7x/**.

[149] SocialCompare. *Recon Jwt* 2014; Available from: **http://socialcompare.com/en/review/recon-jet**.

[150] II, R.M.T., *Virtual Reality Peripheral Network (VRPN)*, 1998.

[151] II, R.M.T., et al. *VRPN: A Device-Independent, Network-Transparent VR Peripheral System*. in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. 2001. ACM.

[152] Walker, B.N. and G. Kramer, *Auditory Displays, Alarms and Auditory Interfaces*, in *International Encyclopedia of Ergonomics and Human Factors*, W.K. Informa Healthcare, Editor. 2006, CRC Press. p. 1021 - 1025.

[153] Walker, B.N., R.M. Stanley, and J. Lindsay. *Task, User Characteristics, and Environment Interact to Affect Mobile Audio Design*. in *PERVASIVE*. 2005.

[154] Ward, M., et al., *Visual Cues to Reorient Attention from Head Mounted Displays*, 2016.

[155] Posner, M.I., *Orienting of Attention.* Quarterly Journal of Experimental Psychology, 1980. **32**(1): p. 3 - 25.

[156] NaturalPoint. *OptiTrack Flex 13 Camera*. 2015 [cited 2016 28 December]; Available from: **http://optitrack.com/products/flex-13/**.

[157] NaturalPoint. *Motive*. 2015 [cited 2016 28 December]; Available from: **http://optitrack.com/software/**.

[158] Systems, W.D. *AudioStream Pro AV*. [cited 2016 28 December]; Available from: **http://www.widigitalsystems.com/index.php?route=product/product&path=110_134&product_id=89**.

[159] Shams, L., Y. Kamitani, and S. Shimojo, *Illusions: What You See Is What You Hear*. Nature, 2000. **408**(6814): p. 788.

[160] Posner, M., M.J. Nissen, and R.M. Klien, *Visual Dominance: An Information-Processing Account Of Its Origins And Significance*. Psychological Review, 1976. **83**(2): p. 157 - 171.

[161] Vogel, E.K. and M.G. Machizawa, *Neural Activity Predicts Individual Difference in Visual Working Memory Capacity*. Nature, 2004. **428**(6984): p. 748 - 750.

[162] Benjamin, W., *The Work of Art In the Age of Mechanical Reproduction*, in *Literary Theory: An Anthology*, J. Rivkin and M. Ryan, Editors. 2004, Blackwell Publishing Ltd. p. 1235 - 1241.

Appendices VII

APPENDIX I

PILOT STUDY

This section provides details of a pilot study carried out prior to beginning our main experiment. This was run to establish the feasibility of using a bone conduction headset as part of a bi-modal, wearable interface. Following up on the study a similar experiment was carried out using the bone conduction headset listed in Chapters 3 and 4, and the one listed here. The same scene in Unity was used to gauge their performance. Based on the result of this second pilot study, we made the decision to use the Aftershokz Sportz3 bone conduction headset for our main experiment. Unfortunately, this second pilot study and the accompanying results have been lost due to data corruption on an external drive. None-the-less, localisation results for this study demonstrated no significant differences between the headsets. The primary reason for choosing the Aftershokz Sportz3 for our experiments was its on board amplifier that did a remarkable job in providing a good auditory output level. Subjectively, the Aftershokz headset also appeared to have a better frequency response in comparison to the Golden Dance headset.

7.1 Aim

The aim of the pilot study was to assess the difference between the accuracy of spatialised audio sources when presented over two different audio reproduction media – BCH and headphones – for a given set of conditions.

7.2 Hypothesis

A bone conduction headset will allow for sufficiently good externalisation, comparable to headphones, to afford an acceptable level of accuracy for spatialised sound sources.

7.3 Assets

- Golden Dance Audio Bone MGD02 bone conduction headset (Frequency Response: 50 Hz – 12 kHz; as specified by the manufacturer)
- Beyer Dynamic DT770Pro Headphones (Frequency Response: 20 Hz – 20 kHz; as specified by the manufacturer)

- Four sound source – music loop, beeps, woosh (reversed cymbal hit) and male speech
- A scene constructed in Unity 5
- Oculus Spatializer Plugin (OSP) for Unity

The following images indicate the frequency spectrum (y-axis) and duration (x-axis) of the sound sources used for this study

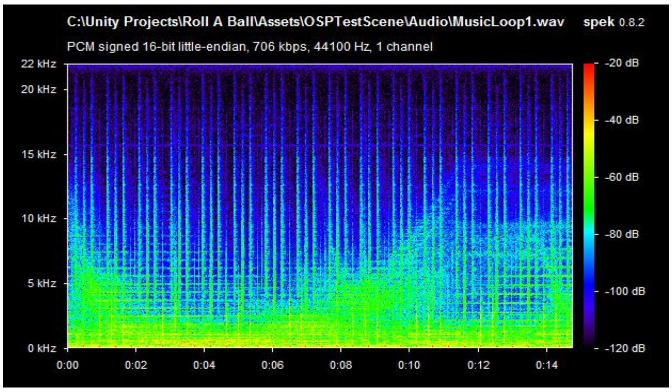


Figure 7.1: Frequency spectrum of the music loop

The music loop used was supplied along with the OSP Test scene package. As we can see from the frequency spectrum alongside, the sample's energy appears to be concentrated mainly in the lower end i.e. lower frequencies. There appears to be no significant energy in the frequencies above 7 kHz – 8 kHz.

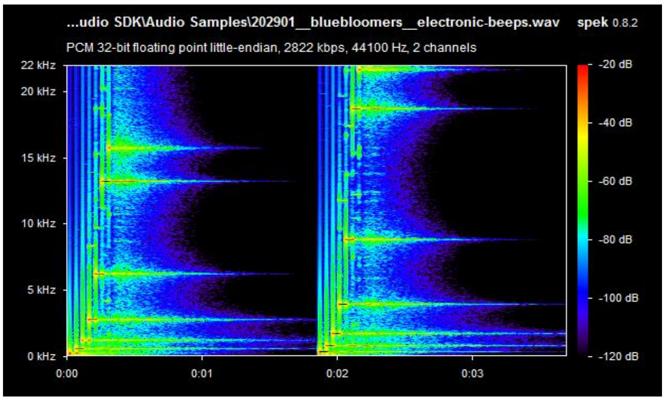


Figure 7.2: Frequency spectrum of the beeps

The beeps sample used here shows relatively high energy distribution over the whole frequency spectrum. The image alongside shows the distribution of frequencies of the individual 'beeps'. Their relatively short duration and impulse like quality should make them easy to localise.

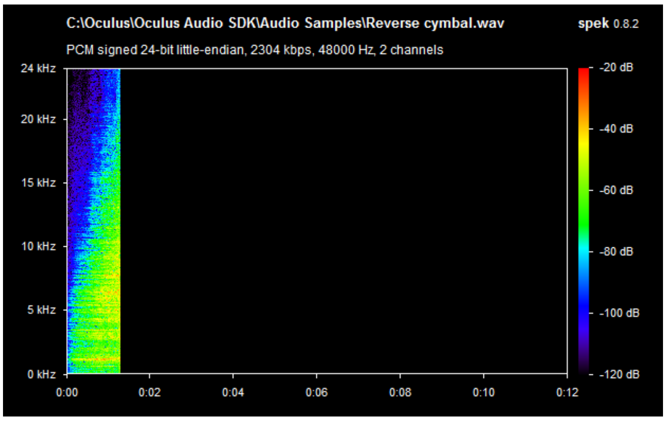


Figure 7.3: Frequency spectrum of the ‘woosh’. The ‘woosh’ was obtained by reversing the sound of a cymbal hit

The ‘woosh’ appears to have a relatively high distribution of energy in the lower frequencies. The energy content in the higher frequencies rapidly declines as the sound progresses, although there does appear to be some strength in the higher frequencies between 8 kHz – 11 kHz. The one second sound is followed by eleven seconds of silence following which it repeats. This is the only sound among the four sources that does not play constantly for the duration of the test

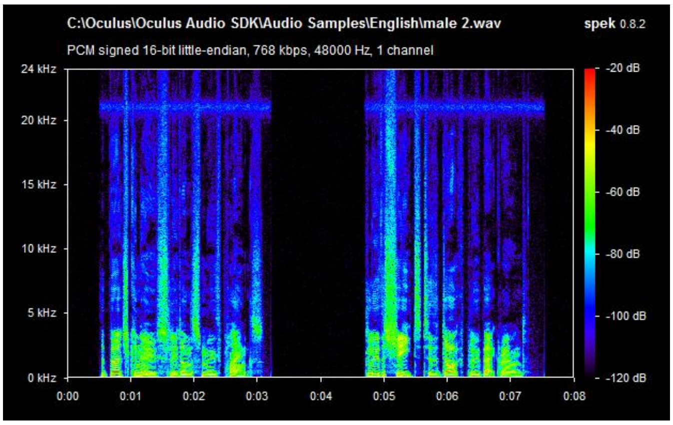


Figure 7.4: Frequency spectrum of male speech

The male speech sound source consists of two sentences separated by little over a second. In this case too, the energy of the sound source is concentrated in the lower end of the frequency spectrum with hardly any content with significant energy in it exceeding 5 kHz.

As is evident from the images, no attempt has been made to ‘book match’ these sources with regards to sampling frequency, bit depth or number of channels. The “force to mono” function in Unity was applied on all sources which had more than one channel. This was done to allow the OSP to spatialise the sources as accurately as possible. Specialization works best when the sources to be spatialised are in the ‘mono’ (single channel) format.

The four sources were chosen to loosely represent a mix of music, notification sounds and speech elements in use in today’s mobile devices.

7.4 The Unity Scene

As has been mentioned before, the test was conducted by constructing a scene in Unity and then playing this back to the participants (Figure 6.5). It allowed for a relatively accurate placement of the sound sources in relation to the listener. The camera in Unity was tied to the listener to afford the person a ‘first person’ view of the sound field. The screenshot below shows how the scene was setup in Unity. The white dot in the centre represents the camera and the player. The participants’ perception of the sound sources was tied to the camera object’s orientation as mentioned above. The four red spheres represent the sound sources – music loop, beeps, woosh and male speech.

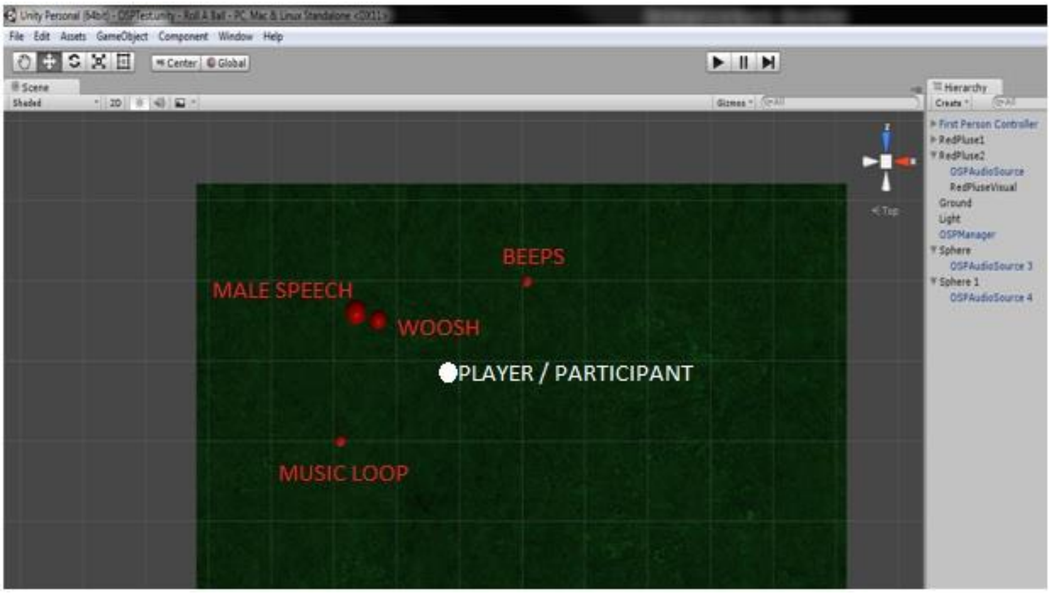


Figure 7.5: A view of the distribution of sound sources around the listener in Unity3D

Using the screenshot above as a reference, the distribution of sound sources was as follows:

- Music Loop – Azimuth: -90°, Elevation: 0° (in line with left ear)
- Beeps – Azimuth: +90°, Elevation: 0° (in line with the right ear)
- Woosh – Azimuth: 0°, Elevation: 0° (directly in front, level with the ears)
- Male Speech – Azimuth: 0°, Elevation: +45° (directly in front, at an angle of 45° from the interaural axis)

A simpler representation of sound source distribution is shown in the figure below

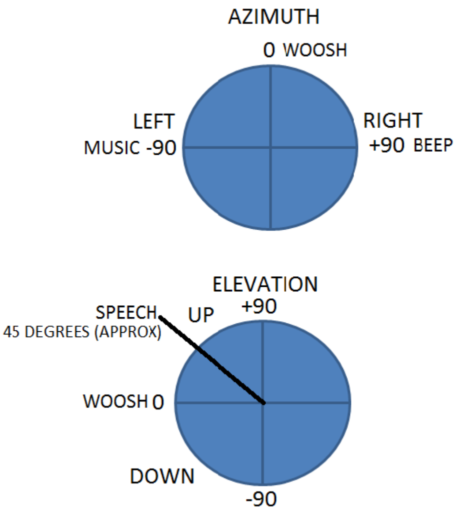


Figure 7.6: Representation of sound source distribution by azimuth and elevation

The screen shot below (Figure 6.7) shows us what the scene looked like when it was run. While none of the participants were allowed to look at the screen during the listening test, this represents the participants' orientation in the sound field during the listening test. The red sphere in front represents the 'woosh' which was placed directly in front (azimuth: 0°) of the participants and is assumed to be at ear level (elevation: 0°).

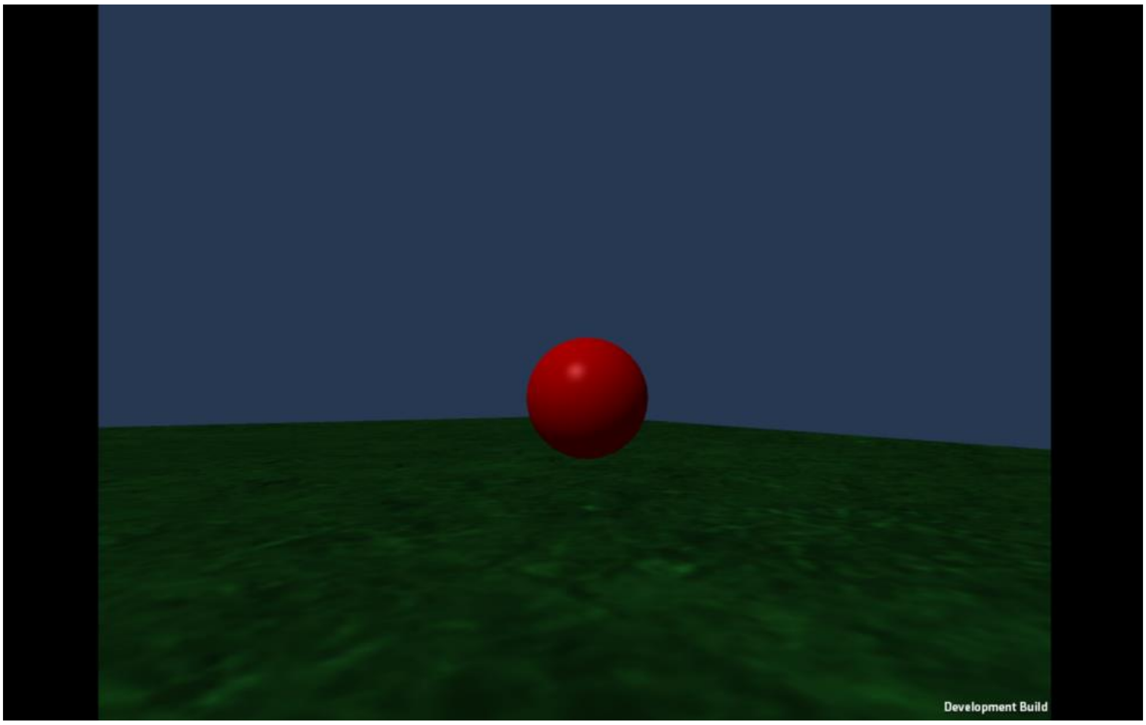
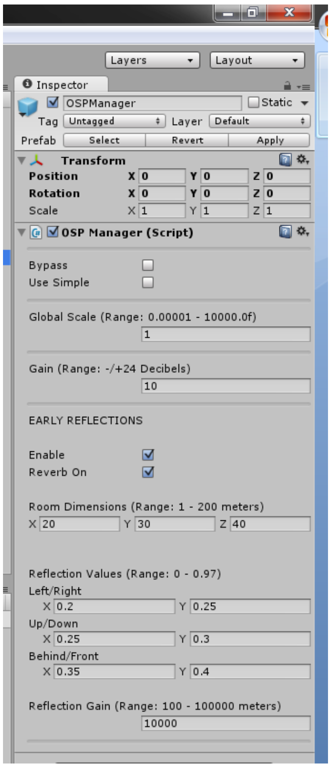


Figure 7.7: Representation of a sound source as seen from the user's perspective

7.5 Running the Pilot Study

For both reproduction conditions, over BCH and headphones, participants were asked to put on the apparatus before the scene was started. The experiment was conducted in the lab with no attempt to provide the participants with a silent space, darkened room etc. Participants were also not asked to close their eyes. Such a setting, to a certain extent, is representative of a typical ‘office environment’. Participants were asked to report where they heard the sound source in a spherical sound field around them after explaining the co-ordinate system being used for the purpose of this study. They were also asked whether the sound sources appeared to be located inside or outside their head. If outside, they were asked to indicate an approximate at which the sources were located. Participants were also required to identify the number of sources that were being presented to them.

No attempt was made to spatialise the scene at the same sound pressure level for reproduction over BCH and headphones. Although, the BCH did require the volume on the on board sound card to be turned up all the way. The output on the OSP Manager too was increased by 10dB since spatialisation appears to cause a drop in the overall level of the sound sources. The settings used for the OSP Manager can be seen in the screen shot below (Figure 6.8).



☑ OSPManager ☐ Static
Tag Untagged
Layer Default
Prefab

Transform
Position X 0 Y 0 Z 0
Rotation X 0 Y 0 Z 0
Scale X 1 Y 1 Z 1

☑ **OSP Manager (Script)**
Bypass ☐
Use Simple ☐

Global Scale (Range: 0.00001 - 10000.0f)
1

Gain (Range: -/+24 Decibels)
10

EARLY REFLECTIONS

Enable ☑
Reverb On ☑

Room Dimensions (Range: 1 - 200 meters)
X 20 Y 30 Z 40

Reflection Values (Range: 0 - 0.97)
Left/Right
X 0.2 Y 0.25
Up/Down
X 0.25 Y 0.3
Behind/Front
X 0.35 Y 0.4

Reflection Gain (Range: 100 - 100000 meters)
10000

Figure 7.8: The OSP Manager in Unity3D

7.6 Results

The following table shows us user performance with the bone conduction headset

Table 5: Tabulated localisation judgements made with the bone conduction headset

PARTICIPANT	SOURCES IDENTIFIED	MUSIC (A -90°, E 0°)	BEEPS (A +90°, E 0°)	WOOSH (A&E 0°)	SPEECH (A 0°, E +45°)
P1	4	-90°	+100°, marginally behind the head	+100°, sounded close to the head	Behind, slightly to the right, no elevation
P2	3	-100°, distance: approximately 10 ft, below interaural axis	+90°, close (less than 1m), 45 elevation	Exactly behind, no elevation	Exactly behind, no elevation, about 30 ft away
P3	3 (Definitely) + 1 indistinguishable source (woosh). Unsure	-80° to – 85°, 1m away, no elevation	+90°, no elevation, 20 cms away	Could hear source but presumed to be part of music / not distinguishable	0° (in front), -10° elevation (reported as being "under the chin")
P4	4	-150°, 100m away, -40° elevation	+120°, 5m – 10m away, no elevation	Exactly behind, 2m away, no elevation	Exactly behind, -50° elevation
P5	4	-125°, 2m away, +45° elevation	+100°, 0.5m, no elevation	Exactly behind, 1m away, +50° elevation	Exactly behind, 5m away, no elevation
P6	4	-135°, 1m away, +30° elevation	+90°, 0.5m away, no elevation	+135°, +30 elevation	+160°, 2m away, no elevation
P7	4	-90°, 2 – 3m away, +25° elevation	+90°, <1m away, no elevation	-10°, 2m away, +15 elevation	0, <1m away, no elevation. Reported as appearing "sometimes behind" during the play cycle
P8 Unsure about externalization	4	-45°, 15 feet away, +15° – +20° elevation	+90°, <1m away, no elevation	-110°, 15 feet away, +15° - +20° elevation	-110°, 15 feet away, +15° - +20° elevation. Unsure about

					distance
P9 Headset not worn properly	5	-90°, 100m away, +45° - +60° elevation	Identified as two sources. +90°, <1m away, +2° elevation	-160°, 0.5m away, +10° elevation	Exactly behind, 2m away, no elevation
P10	4	-60°, 2-3m away, +30° - +45° elevation	+90°, 0.5m away, no elevation	+170°, 10m away, 10° elevation	+170°, 5m away, -5° elevation
P11 Headset worn OVER the headscarf. Appeared mostly unsure about localisation.	3	-70°, 2m away, no elevation	+90°, 1m, +70° elevation	Identified as part of "music" source	0, 1m away, no elevation
P12 Externalisation reported for all sources except the 'woosh'. Some confusion regarding how to wear	4	-90°, 2 -3m away, no elevation	+75° - +80°, close (no distance), reported as "drops falling from top to bottom"	Exactly to the back and inside the head	Heard on both sides at a distance of 1m, no elevation reported
P13 Noisy environment in the lab. Some uncertainty regarding externalization	5	-90°, 3-4m away, +17° elevation	+100°, 2m away, no elevation	Exactly behind, <1m, +10° - +20° elevation	Exactly behind, 6-7m away, +30° elevation

From the table above we can see that lateral localisation (left and right) is relatively good for both sound sources – music loop and beeps. Localisation of the music loop though shows a greater variation than the beeps.

The BCH also appears to be relatively good at reproducing spatialised sound sources with few participants unsure about or not reporting any externalisation.

Eight participants suffered from reversals in the case of the 'woosh' and male speech sound sources. This could be due the absence of strong high frequency cues in the sound sources, particularly the speech. Participants also reported some movement of the 'woosh' and that it was perceived at an elevation. This perception of elevation and movement could be attributed to the relatively short, but fairly prominent high frequency cues present in the 'woosh'. These high frequency cues last for less than a second but appear to have an impact on the elevation perception and impart dynamism to the sound source. Almost all participants reported the source (woosh) moving towards them.

The table below shows the results for the same scene spatialised over headphones for the same sound source positions. Ten participants from the previous study participated in this.

Table 6: Tabulated localisation judgements made with headphones

	MUSIC		BEEPS		WOOSH		SPEECH	
P1	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	5	90	0	180 R	45	180 R	45
P2	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	70	120	0	180 R	0	180 R	0
P3	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-100	10	135	0	180 R	45	180 R	30
P4	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-80	0	90	0	15	45	180 R	0
P5	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-100	0	80	0	180 R	30	18- R	0
P6	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-135	20 - 30	120	35	175	30	160	-20
P7	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-100	20	90	0	120 TO 90	180 TO 90	-160	50
P8	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	25 - 30	90	0-010	180 R	0	180	30 - 45
P9	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	2 TO 3	90	0	180 R	5 TO 10	180 R	45
P10	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	-20	90	0	N135 TO P150	20	180 R	0
P11	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	N90 - N95	5 TO 10	85	5 TO 10	-5	10	-10	-15
P12	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-70	15 - 20	90	5	180 R	0	175 - 180	60

There appear to be no major differences between the ten participants who were common to both the studies. The table below shows they fared across the two studies. The green highlight indicates the correct target acquisition for the headphone condition and the orange for the BCH condition.

Table 7: Comparisons of localisation judgements made with headphone vs. bone conduction headset. All highlighted blocks indicate correct localisation.

	MUSIC		BEEPS		WOOSH		SPEECH	
	A = -90	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
P1 – HP	-90	5	90	0	180	45	180	45
P5 – BCH	-125	45	100	0	180	50	180	0
P2 – HP	-90	70	120	0	180	0	180	0
P13 – BCH	-90	17	100	0	180	10	180	30
P3 – HP	-100	10	130	0	180	45	180	30
P1 – BCH	-90	NR	100	NR	100	NR	170	NR
P5 – HP	-100	0	80	0	180	30	180	0
P8 – BCH	-45	15	90	0	-110	15	-110	15
P7 – HP	-100	20	90	0	120 - 90	180 - 90	-160	50
P6 – BCH	-135	30	90	0	135	30	160	0
P8 – HP	-90	25 - 30	90	0 - 10	180	0	180	30 - 45
P2 – BCH	-100	NR	90	45	180	0	180	0
P9 – HP	-90		90	0	180	5 '- 10	180	45
P9 – BCH	-90	50	90	2	-160	10	180	0
P10 – HP	-90	-20	90	0	-135 TO 150	20	180	0
P4 – BCH	-150	-40	120	0	180	0	180	-50
P11 – HP	-90 TO '-95	5 '- 10	85	5 '- 10	-5	10	-10	-15
P3 – BCH	-85	0	90	0			0	-10
P12 – HP	-70	15 - 20	90	5	180	0	175 - 180	60
P10 - BCH	-60	30	90	0	170	10	170	-5

From the table above we can see that there appears to be a fairly even distribution in terms of target localisation for both the conditions – headphones and BCH. Besides the 'Beeps' no other sound source shows a pair of correct acquisitions for a given target. For the 'Beeps" sound source, four correct target acquisitions (azimuth and elevation) were made while only three acquisitions were made for the headphone condition.

A paired sample t-test was carried out on the data to analyse the differences between binaural spatialisation over a headphone and bone conduction headset. One participant was excluded from the analysis for the music, beeps and speech pairs while two were excluded for the pair representing the whoosh. Analysis shows a significant difference in localisation between headphone and bone conduction headset based reproduction (t (8) = 2.848, $p < 0.05$). No such differences were observed for the rest of the pairs; Beeps (t (8) = 0.121, p = 0.907), Speech (t (8) = 1.179, p = 0.272) and Whoosh (t (7) = 1.958, p = 0.091).

Nearly all azimuthal source acquisitions for the ‘woosh’ and ‘speech’ show reversals in both conditions and only some participants have indicated the right elevations for them. This appears to be the case for the other two sources too, as seen in the table below (headphone condition only). Interestingly, the music and beeps too appear to be localised behind the head (classified as front-back reversal) in both target conditions (azimuth: 0° or 180°). The highlighted numbers show reversals. In this case we have assumed reversals to be any source being ≥ ± 160°. For the purpose of this study we can also look at reversals as any source, placed in the front quadrants, which is localised in the rear quadrants.

Table 8: Reversals in localisation judgements (Headphones)

	MUSIC		BEEPS		WOOSH		SPEECH	
P1	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	5	90	0	180 R	45	180 R	45
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	180 R	45	-10	10	90	45	135	45
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
			180 R	-5	-90	30	-170	45
P2	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	70	120	0	180 R	0	180 R	0
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	180 R	0	160 TO 170	0	120	20 - 30	170	30
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	180	0	180 R	0	-120	60	n160 - 170	80
P3	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-100	10	135	0	180 R	45	180 R	30
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	90	60	150 - 160	10	100	0	120	20 - 30
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-170	30	170	45	-120	0	-135	45
P4	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-80	0	90	0	15	45	180 R	0
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	-150	-45	135	25	135	45	135	75
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**

	-135	-35	110	20	-135	20	-135	25
P5	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-100	0	80	0	180 R	30	18- R	0
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	-110	0	180 R	0	110	15	110	0
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-150	0	-160	0	-80	15	-100	20
P6	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-135	20 - 30	120	35	175	30	160	-20
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	180 R	10	135	0 - 15	20	5	100	0
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-160	-50	-170	-30	-90	0	-110	-40
P7	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-100	20	90	0	120 TO 90	180 TO 90	-160	50
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	180 TO 170	0	160	-40	160 - 60	n45 - p30	120	40
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-170	10	180 R	0	n120 - n60	0 - 30	-135	60
P8	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	25 - 30	90	0-010	180 R	0	180	30 - 45
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	-170	0	130 - 145	0	70	0	135 - 140	30 - 45
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-135	0	N170 TO N160	0 - 10	N80 TO N90	10	-135	45
P9	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	2 TO 3	90	0	180 R	5 TO 10	180 R	45
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	180 - 135	45	180	10	90	0	90	45 - 60
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-125	60	180	45 - 60	-90	5	-135	45
P10	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-90	-20	90	0	N135 TO P150	20	180 R	0
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	-150	20	-170	10	90	0	160	0
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-130	20	180 R	-20	-100	40	-120	0
P11	**A = -90**	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	N90 - N95	5 TO 10	85	5 TO 10	-5	10	-10	-15
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	-165	15	75	10	95	0	110	5 TO 10

P12	**A = 180**	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
	-150	-30	25	0	-100	5	-125	-5
	A = -90	**E = 0**	**A = 90**	**E = 0**	**A = 0**	**E = 0**	**A = 0**	**E = 45**
	-70	15 - 20	90	5	180 R	0	175 - 180	60
	A = 0	**E = 0**	**A = 180**	**E = 0**	**A = 90**	**E = 0**	**A = 90**	**E = 45**
	N175 - 180	5	175	20	100	10	110	60
	A = 180	**E = 0**	**A = 0**	**E = 0**	**A = -90**	**E = 0**	**A = -90**	**E = 45**
		95	90	5	-120	45	-80	70

The beeps, compared to the other sources, show good elevation perception. This, though, could be attributed to the fact that this source stood out from the rest as being the loudest and was consistently placed the closest to the head. This lack of distance due to its level appears to have contributed to a large number of participants localising the source at the right elevation. Some participants also reported the 'beeps' to be heard as two distinct sounds, one participant reporting them to sound like "drops falling from top to bottom".

7.7 Redressals

Having looked at the data we can clearly see that the bone conduction headset is a good alternative to the headphones as part of a hybrid wearable interface. It definitely appears to have the ability to reproduce spatialised audio in a relatively faithful manner. The reversals we see are, in my opinion, due to the nature of the sound sources chosen. The lack of high or low frequency energy appears to be causing significant front-back confusion. This indicates that a broad band sound with a relatively even distribution of energy across the frequency spectrum will perform better. To address this I propose creating and/or sourcing such sources. Keeping in mind the uses of this proposed hybrid interface, the sources will resemble audio beacons and/or alerts similar to the ones in use on current mobile devices. Speech, male and female, will also be tested in upcoming experiments.

The use of four different stimuli, in addition to speech, covering the whole frequency spectrum can be tested to determine the minimum audible angle for each of the roughly dived bandwidths over the spectrum that encompasses human hearing i.e. 20 Hz to 20 kHz. The bandwidths can be divided into lows (20 Hz – 250 Hz), low-mids (250 Hz – 1000 Hz), high-mids (1000 Hz – 7000 Hz) and highs (7000 Hz – 20,000 Hz).

APPENDIX II

INFORMATION SHEETS, CONSENT FORMS & QUESTIONNAIRES

This section contains the information sheets, consent forms and questionnaires used for the all our experiments. University regulations state that all participants must be provided with information sheets and consent form prior to taking part in any experiment. In this section we present an information sheet and a consent form used in one our studies as samples. Further, pre and post study questionnaires that participants were asked to fill out as part of the experimental process are also presented.

INFORMATION SHEET FOR PARTCIPANTS

Welcome! You are invited to take part in this study which will form a part of the investigators, Amit Barde & Matt Ward's, doctoral theses. The purpose of this study is to understand and evaluate the use of auditory and visual cues in a real-world search task. This study will form a part of an ongoing investigation into the design and use of hybrid wearable interfaces.

As part of this study you will be required to perform a simple search task. The sequence of the experimental procedure will be as follows:

- You will be presented with a bone conduction headset and the Google Glass, both of which you will be required to wear.
- Visual cues will be presented to you via the Google Glass, while auditory cues will be delivered over the bone conduction headset.
- You are to navigate within the test space using these cues and find a target object.
- To help familiarise you with the experiment, a trial run with all the cues presented once will be carried out.
- The main experiment will run for 50 minutes.
- You will receive a $10 Westfield Voucher as compensation.
- Your participation in this study is voluntary.
- You have the right to withdraw at any stage without penalty up until the point the data is entered into the computer.
- If you choose to withdraw, all data relating to your participation in the study will be discarded.
- Your participation in this study and the data generated as a result will be treated with the utmost regard to anonymity and confidentiality.
- No personal information will be collected. Only population demographics such as age, sex and frequency of mobile and/or wearable device usage will be collected.
- You are entitled to receive a copy of the results by contacting the researchers at the conclusion of this study. The results of this study may be published. Any publication of the results will not involve divulging participant details.
- Only demographic data that has been collected will be made public as required for publication. All data generated as part of this study will be kept in a secure location i.e. in the research's personal locker and/or a password protected computer located in a secure, access controlled facility. Only the researcher and this supervisor(s) will have access to this data.

- The data will be archived and stored for 10 years before being destroyed as university regulation governing the storage and destruction of data accumulated as part of a doctoral thesis stipulates. The results of this study will be used in part or full, as part of the researcher, Amit Barde's, doctoral thesis.

Please indicate to the researcher on the consent form if you would like to receive a copy of the summary of the results of this project.

This project is being carried out by the researchers, Amit Barde & Matt Ward, as a requirement for PhD degree under the supervision of Dr. Rob Lindeman, Dr. Gun Lee and Dr. Mark Billinghurst. They can all be contacted via their respective email addresses given below.

Dr. Rob Lindeman: **gogo@hitlabnz.org**

Dr. Gun Lee: **gun.lee@canterbury.ac.nz**

Dr. Mark Billinghurst: **mark.billinghurst@canterbury.ac.nz**

They will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ehtics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (**human-ehtics@canterbury.ac.nz**).

If you agree to participate in the study, you are asked to complete the consent form and return it to the experimenters.

Amit Barde.

Pre and post task questionnaires for the first two experiments described in Chapter 3 were the same. Experiment 3 also utilised the same pre-task questionnaire as experiments 1 and 2 but did not have a post-study questionnaire since externalisation ratings were sought at the end of each trial.

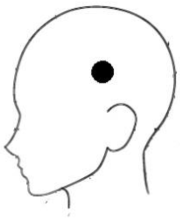
Experiment 1, 2 and 3 Pre-task questionnaire

1. Gender: Male ☐ Female ☐ Other ☐
2. Age: ☐☐ years
3. Do you have any previous experience with auditory experiments? Conducting, participating or assisting.
   Yes ☐ No ☐
4. Do you own a mobile device? I.e. cell phone, tablet, mp3 player etc.
   Yes ☐ No ☐
5. If you've ticked **Yes** in Q4, how often do you use one or more of these devices?
   Everyday ☐ Few times a week ☐ Few times a month ☐ Few times a year ☐
6. Do you use headphones/earphones to listen to music and/or other material?
   Yes ☐ No ☐
7. If you've ticked **Yes** in Q6, how often do you use headphones/earphones?
   Everyday ☐ Few times a week ☐ Few times a month ☐ Few times a year ☐
8. How often do you use headphones/earphones to listen to material on any of your mobile devices?
   Everyday ☐ Few times a week ☐ Few times a month ☐ Few times a year ☐
9. Do you have normal hearing in both ears?
   Yes ☐ No ☐

Experiment 1 and 2 post-study questionnaire

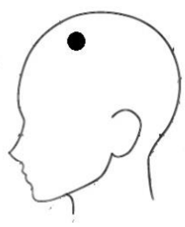
Where did you hear a majority of the stimuli? (CIRCLE A NUMBER)

Centre of the head



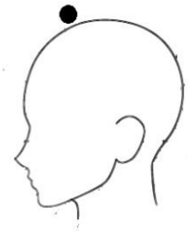
1

Between the centre and surface of the head



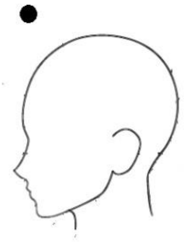
2

On the surface of the head



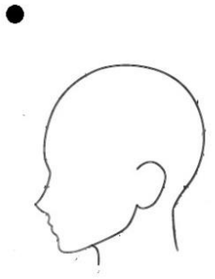
3

Less than or equal to 1m from the surface of the head



4

More than 1m from the surface of the head



5

Experiment 5 pre-study questionnaire

1. Gender: Male ☐ Female ☐ Other ☐

2. Age: ☐☐ years

3. Do you have any previous experience with auditory experiments? Conducting, participating or assisting.
   Yes ☐ No ☐

4. Do you own a mobile device? I.e. cell phone, tablet, mp3 player etc.
   Yes ☐ No ☐

5. If you've ticked **Yes** in Q4, how often do you use one or more of these devices?
   Everyday ☐ Few times a week ☐ Few times a month ☐ Few times a year ☐

6. Do you use headphones/earphones to listen to music and/or other material?
   Yes ☐ No ☐

7. If you've ticked **Yes** in Q6, how often do you use headphones/earphones?
   Everyday ☐ Few times a week ☐ Few times a month ☐ Few times a year ☐

8. How often do you use headphones/earphones to listen to material on any of your mobile devices?
   Everyday ☐ Few times a week ☐ Few times a month ☐ Few times a year ☐

9. Do you have normal hearing in both ears?
   Yes ☐ No ☐

10. Are you familiar with a wearable computing device? E.g. Google Glass.
    Yes ☐ **(Go to Q.11)** No ☐ **(Return the questionnaire)**

11. How often have you used such a device?
    Once ☐
    Everyday ☐
    Few times a month ☐
    ☐

Few times a year

12. How would you describe your experience using such a device?

Excellent, I found it useful ☐

Good, with potential to be a lot more useful ☐

Neutral, the novelty of using such a device wore off quickly ☐

Bad, I do not see how such a device could be helpful ☐

Other (Please explain the experience in your own words) ☐

Experiment 5 post study questionnaire

1. Which visual cue did you prefer?

   Environment Mapped ☐ Dynamic Arrow ☐

2. Which auditory cue did you prefer?

   Static (Constant Ping) ☐ Dynamic ☐

3. Which cueing condition did you most prefer? (**1 – Best, 3 – Worst**)

Visual Cue	Auditory Cue	A combination of the two
☐	☐	☐

4. You were fastest at finding the target using the (**1 – Fastest, 3 – Slowest**)

Visual Cue	Auditory Cue	A combination of the two
☐	☐	☐

5. Would you use any of these cues in a real-world scenario? E.g. To find an address or while driving etc.

   Yes ☐ **Go to Question 6** No ☐ **Go to Question 7**

6. Which one?

   Visual Cue ☐ Auditory Cue ☐ A combination of the two ☐

   Depends on the scenario (driving, walking etc.) ☐

   Go to Question 7

7. Why? (Provide a short explanation for your choice)
