

# A Mobile Augmented Reality Audio System for Interactive Binaural Music Enjoyment.



Steven Thornely

Queensland Conservatorium of Music

Griffith University

A thesis submitted for the degree of

*Bachelor of Music Technology with Honours*

2013

# **Abstract**

This paper details the development and implementation of a Mobile Augmented Reality Audio (MARA) System designed for interactive music playback. MARA technology incorporates position tracking and head tracking to create an immersive auditory environment in headphones that adjusts to user movements in real time. Human hearing uses head movement as a natural method for localising sound. MARA capitalizes on this aspect of sonic perception, allowing it to deliver a convincing augmented audio experience. A Design Research process is employed to investigate and develop each component of a MARA system, including an auralization engine, user tracking unit, and MARA headset. The resulting system is used to create a musical demonstration that showcases the potential for this technology. Broader ideas about music production and consumption with MARA are extracted from this demonstration and discussed, with areas for future research noted.

# Contents

|   |             |
|---|-------------|
| <b>Abstract</b>   | <b>ii</b>   |
| <b>List of Figures</b>                                      | <b>vi</b>   |
| <b>Certification</b>  | <b>viii</b> |
| <b>Acknowledgements</b>                                     | <b>ix</b>   |
| <b>Glossary</b>   | <b>x</b>    |
| <b>1 Introduction</b>                                       | <b>1</b>    |
| <b>2 Literature Review</b>                                  | <b>3</b>    |
| 2.1 The Importance of Interaction . . . . .                 | 3           |
| 2.2 Sonic Immersion . . . . .                               | 5           |
| 2.3 Sonic Immersion and the Human Auditory System . . . . . | 7           |
| 2.4 Previous MARA Development . . . . .                     | 9           |
| <b>3 Methodology</b>  | <b>13</b>   |
| 3.1 Design Research . . . . .                               | 13          |
| 3.2 Theoretical Background . . . . .                        | 14          |
| 3.3 Design . . . . .  | 15          |
| 3.4 Evaluation . . . . .                                    | 16          |
| 3.5 Methodologies in Existing MARA Research . . . . .       | 17          |
| 3.6 Method . . . . .  | 18          |
| 3.7 Summary . . . . .                                       | 18          |

|          |   |           |
|----------|---|-----------|
| <b>4</b> | <b>Designing a MARA System for Music Creation</b>                         | <b>20</b> |
| 4.1      | Introduction . . . . .  | 20        |
| 4.2      | MARA Headset . . . . .  | 20        |
| 4.2.1    | Design Goal . . . . .   | 20        |
| 4.2.2    | Prototyping Stages . . . . .  | 21        |
| 4.2.3    | Current Status . . . . .  | 30        |
| 4.2.4    | Future Improvements . . . . .   | 30        |
| 4.3      | Auralization Engine . . . . .   | 30        |
| 4.3.1    | Design Goal . . . . .   | 30        |
| 4.3.2    | Development Process . . . . .   | 31        |
| 4.3.2.1  | Space and Direction . . . . .   | 31        |
| 4.3.2.2  | Depth . . . . .   | 33        |
| 4.3.2.3  | Automation . . . . .  | 34        |
| 4.3.3    | Critical Evaluation . . . . .   | 35        |
| 4.4      | User Tracking System . . . . .  | 39        |
| 4.4.1    | Design Goal . . . . .   | 39        |
| 4.4.2    | Existing Tracking Systems . . . . .                                       | 40        |
| 4.4.3    | Developing and Implementing a Location Tracking System                    | 40        |
| 4.4.4    | Developing and Implementing an Orientation Tracking Sys-<br>tem . . . . . | 44        |
| 4.4.5    | Critical Evaluation . . . . .   | 50        |
| <b>5</b> | <b>Making Music with MARA</b>   | <b>51</b> |
| 5.1      | Creative Operation of the MARA System . . . . .                           | 52        |
| 5.2      | Musically Creative Ideas with the MARA System . . . . .                   | 54        |
| <b>6</b> | <b>Potential Impact on Music Production and Consumption Paradigms</b>     | <b>56</b> |
| 6.1      | Looking Forward: Advancing MARA Systems for Music . . . . .               | 56        |
| 6.2      | Future Work . . . . .   | 58        |
| <b>7</b> | <b>Concluding Remarks</b>   | <b>60</b> |
|          | <b>Appendices</b>   | <b>62</b> |
| A        | Demonstration of Convolution-Panner and the Blurring Effect . .           | 62        |

## CONTENTS

---

0.0

|                   |  |           |
|-------------------|--|-----------|
| B                 | 'Ping' Response Patch . . . . .        | 63        |
| C                 | List of Required Equipment . . . . .   | 64        |
|                   | C.1 Software: . . . . .                | 64        |
|                   | C.2 Hardware . . . . .                 | 64        |
| D                 | Program Notes . . . . .                | 65        |
| E                 | Software Setup Instructions: . . . . . | 66        |
| <b>References</b> |  | <b>69</b> |

# List of Figures

|      |   |    |
|------|---|----|
| 4.1  | Original circuit of the Philips SHN2500 noise-cancelling earphones  | 22 |
| 4.2  | Headset preamp circuit iteration 1 . . . . .  | 24 |
| 4.3  | Headset preamp circuit iteration 2 . . . . .  | 25 |
| 4.4  | Headset preamp circuit iteration 3 . . . . .  | 27 |
| 4.5  | Iteration 4: MARA mixer circuit . . . . .   | 28 |
| 4.6  | The MARA mixer . . . . .  | 29 |
| 4.7  | The ProTools 7.0 panner (listener’s head is superimposed) . . . .   | 32 |
| 4.8  | The signal routing method for manipulating depth of each sound<br>source . . . . .  | 34 |
| 4.9  | Location error of front/back group . . . . .  | 36 |
| 4.10 | Location error of left/right group . . . . .  | 37 |
| 4.11 | The four soundstages available with the option of depth between<br>sound sources . . . . .  | 38 |
| 4.12 | Minimum Audible Angle for sine waves at varying frequencies and<br>azimuths (Mauro, 2012, p. 15) . . . . .  | 39 |
| 4.13 | The error-correcting cybernetics patch . . . . .  | 43 |
| 4.14 | A patch controlling the reading of the second pan knob in ProTools  | 44 |
| 4.15 | The interface for the compass mapper patch . . . . .  | 45 |
| 4.16 | The theoretical result of convolving a binaural mix with a HRTF<br>panner. In practice, the soundstages become non-uniform and in-<br>distinct. . . . . | 46 |

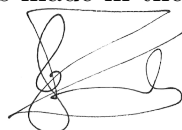
---

|      |   |    |
|------|---|----|
| 4.17 | The user turning their head with the surround-remapper orientation method. In practice, artefacts introduced by the room's rotation are indistinguishable, and a strong binaural image is retained. . . . . | 48 |
| 4.18 | The API SendsPan object in Max for Live . . . . .   | 49 |
| 4.19 | The surround-remapper matrix - send levels from a given SendsPan position . . . . .   | 49 |
| A    | A program to send 'ping' messages to ProTools over the IAC Bus  | 63 |

## Certification

I hereby certify this work is original and has not previously been submitted in whole or part by me or any other person for any qualification or award in any university. I further certify that to the best of my knowledge and belief, this research paper contains no material previously published or written by another person except where due reference is made in the paper itself.

Signed:

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke extending to the right.

Date:

**31/10/2013**



## Acknowledgements

I would like to thank my supervisor Dr. Toby Gifford for his valuable assistance with this project. Toby helped me overcome countless technical hurdles and I am very thankful for his insights.

I'd also like to thank Mr. Matthew Hitchcock for his suggestions that proved crucial to the design of my system. Matt's ideas led me to the convolution-panner and surround-remapper components, which are significant findings from this project that I am very proud of.

Thanks to Brendan and Kai from the A/V department for putting up with me constantly asking for more programs to be installed on the studio computer.

Finally, thank you to all of the Music Technology staff who have taken time to give feedback on my project and help situate my research effectively.

# Glossary

**ADSR** Attack, Decay, Sustain, Release.

**Anechoic** A sound that has no reverberant properties.

**ARA** Augmented Reality Audio.

**Auralization** Applying binaural acoustic cues of a space to a sound source.

**Binaural** Sound that contains specific acoustic information for each ear.

**BRIR** Binaural Room Impulse Response.

**CC** Control Change.

**DAW** Digital Audio Workstation.

**Externalization** Applying acoustic cues to a sound source so that it appears to emanate from outside the head.

**HRTF** Head Related Transfer Function.

**HUI** Human User Interface.

**IAC** Inter-Application Communication.

**ILD** Inter-aural Level Difference.

**Intracranial** A sound that appears to emanate from inside the head when wearing headphones.

**IR** Impulse Response.

**ITD** Inter-aural Time Difference.

**Lateralization** The spatial localization of intracranial sound sources.

**Localization** Deciphering the directional origin of a sound.

**MAA** Minimum Audible Angle.

**MARA** Mobile Augmented Reality Audio.

**MIDI** Musical Instrument Digital Interface.

**PCB** Printed Circuit Board.

**Pseudoacoustic Environment** The representation of the surrounding acoustic environment created by the MARA headset.

**Soundstage** The three-dimensional physical space inferred by the sonic properties of a recording.

**Spatialization** Applying acoustic cues of a space to a sound source.

**SPL** Sound Pressure Level.

**WFS** Wave Field Synthesis.

# Chapter 1

## Introduction

This project investigates the potential for Mobile Augmented Reality Audio to be used as a platform for musical creation. It details the design of a novel Mobile Augmented Reality Audio System for Music, and presents a musical demonstration that showcases the functionality of this system.

Mobile Augmented Reality Audio (MARA) is a technology that allows for a three-dimensional virtual soundstage to be superimposed over a user's natural sound environment. Position tracking and head tracking is incorporated to create an immersive headphone experience that adjusts to user movements in real time (Karjalainen, Tikander, & Härmä, 2004, p. 101). This can create the illusion of a sound object projecting from a fixed location in the room independent of the listener's location or orientation. To ensure perceptual transparency, anechoic sound objects are 'auralized' before being binaurally overlaid into the listener's acoustic space (Mauro, 2012, p. 28). Auralization is the process of adding binaural acoustic cues of the given space to the sound. With MARA technology, it is possible to convincingly 'augment' naturally occurring audio input with sound that appears to originate from a physical location in the listener's surroundings.

MARA is unique to other immersive sound delivery mediums for several reasons. Most notably, MARA allows the listener to *interact* with sound in a three-dimensional environment (Wozniowski, Settler, & Cooperstock, 2012, p. 4).

MARA transparently simulates a sound field, allowing the user to walk around and listen as though sounds were physically emanating from fixed point sources spread about the space. While most spatial audio technologies attempt to recreate a realistic representation of a single perspective, MARA lets the user explore multitudes of perspectives over the temporal course of a piece. MARA differs from surround sound platforms fundamentally, in that the sound in a space is presented directly to the ears as opposed to the sound being projected into the space. This concept is known as binaural sound (Hiipakka, 2012, p. 7).

Binaural recording technology has long existed, but never gained widespread acceptance as a playback medium. This is partly due to several flaws that are inherent in traditional binaural sound: a tendency for front-to-back confusion to occur exists, where it is unclear if a sound is intended to come from in front or behind the listener; and the entire sound stage moves in relation to the listener's head when head movements occur. Both of these issues can negatively impact on the believability of the experience (Alzagi, Dudu, & Thompson, 2004, p. 1). MARA is of particular interest because the inherent issues of binaural sound are resolved when head-tracking is incorporated into the system. Human hearing uses unconscious head movement as a natural method for localising sound (Moustakas, 2012, p. 3). This has been proven to significantly reduce the problem of front-to-back confusion (Hiipakka, 2012, p. 46). The ability to “de-pan” (Gamper, 2010, p. 33) the binaural sound stage also solves the rotation issue.

Prior to this research, a number of implementations of this technology already existed for a range of applications (eg. Lokki et al., 2004), however, the design and use of MARA systems for *music* has seen very limited representation in existing literature. This project collates what is known about MARA systems, and applies this knowledge to the design of a MARA system for a creative musical context. Emphasis is placed on cost effectiveness, ease of execution, and the ability to leverage existing mass-produced low-cost consumer products.

# Chapter 2

## Literature Review

This study straddles the fields of interactive music and immersive sound reproduction. Related literature from these areas is presented in this chapter. The design aspect of this project also calls for a detailed investigation into the mechanics of the human hearing system as well as an overview of the components of established MARA systems.

### 2.1 The Importance of Interaction

This project is rooted within the broader context of human perception, particularly the school of ecological psychology. James J. Gibson pioneered this field with his text, *The Ecological Approach to Visual Perception* (1979), which argues that human visual perception must be discussed in relation to our natural physical environment. Gibson explains that head turning and eye movement (ambient vision) in conjunction with the ability to move around and explore the surroundings (ambulatory vision) are critical components in our perception of sight (p. 303). Importantly, this allows us to continuously update our collection of data on the environment to resolve ambiguities. Andean expresses that Gibson's philosophy can similarly be applied to the context of auditory perception, particularly to counter for blind-spots in semiotic and linguistic approaches (2011, p. 125).

The interaction between listener and environment is a key differentiator between augmented reality audio and traditional sound delivery methods, and is an idea that is emphasised in ecological psychology.

The concept of interacting with music is not new; in fact it has seen considerable development in the digital age due the ability to interface human movement with computer processing at a lower cost (Valbom & Marcos, 2005, p. 871). Dr Garth Paine, a Sydney Music Technologist, has contributed significant research to interactive music in relation to space. Paine explains that interacting with our musical environment is a primal, visceral, and corporeal experience (2007, p. 6). His sentiment, “Sound is fundamentally a temporal medium [that requires] enquiry and patience to unravel the intricate data contained therein” resonates with Gibson’s concept of ‘ambulatory’ perception.

The ability to interact with the sound field holds significant potential for new music listening and creation experiences. In *Sonic Immersion: Interactive Engagement in Realtime Immersive Environments*, Paine discusses his use of real-time sound synthesis controlled by the output of user tracking devices. Here, the physical environment becomes the “performative medium”, and the listener a “creator” and “performative agent” (2007, p. 2). Comparison can be made between this scenario and tribal music, such as that of the Australian aboriginals, in that the space and the people involved form an integral part of the creation of the music (eg. Marett, 2005, p. 88). While interactive music technology may be complex and convoluted at times, its purpose is to recreate a natural, visceral platform to experience music.

Thomas Mitchell identifies that electronic music production commonly involves a breach in fluidity between the performer’s action, the outputted sound, and the audience’s engagement (2011, p. 465). In response to this, he designed a “gestural interface for the performance of live music” titled ‘SoundGrasp’ in conjunction with electronic musician Imogen Heap. SoundGrasp affords the performer a combination of hand gesture recognition and location tracking to control various parameters of a DAW. Not only does this give the performer an intuitive interface with the instrument (a computer), but the aspect of interaction in this example

results in a more meaningful and engaging relationship between the audience, musician, and music.

While many additional examples of interactive music technology exist, the type of interaction that MARA can deliver is quite unique. Whereas the above examples respond to interaction by manipulating musical or textural aspects of the sound, mixed-reality technology has the ability to change the listener's *perspective* of the sound (Wozniowski et al., 2012, p. 3). This can be utilized in an intuitive way so that a listener may interact with the music as if it were being performed by real instruments in the room. On the other hand, there is also potential to employ sounds that do not occur naturally yet make them behave as if they had a physical presence in the space. Andean (2011) explores the relationship between electroacoustic music and the environment of its listeners. The ecological triad of organism/environment/stimulus he discusses (p. 126) becomes particularly interesting in the case of MARA, because these three elements directly influence each other - the stimulus is acoustically situated in the environment; the organism moves within the environment to observe the stimulus.

## 2.2 Sonic Immersion

Sonic Immersion is a form of “sensory absorption within the environment” (Mitchell, 2010, p. 99). The goal of obtaining a sense of immersion within a three-dimensional sound space is in part an extension of our desire to reproduce a live performance as realistically as possible (Baalman, 2010, p. 209). As sound reproduction technology has improved, so has our ability to create immersive sound environments. When monophonic sound was replaced by stereophony, it became possible to achieve a sense of width and depth that was not possible with a single speaker (Moulton, 1990, p. 164). Later, it became common to add more speakers, resulting in quadrophonic, 5.1, and 7.1 speaker systems. Surround sound systems gained popularity on the back of the home theatre industry (Ranjan & Gan, 2013, p. 17). For musical purposes, these systems have some weaknesses: optimal immersion can only be obtained in a small area in the mid-



dle of the speaker array (known as the ‘sweet spot’) (Baalman, 2010, p. 213); and “at least six loudspeakers (not including LFE) are needed to reproduce the spatial impression of a diffuse sound field” (Hiyama, Komiyama, & Hamasaki, 2012, p. 3). Therefore, the most common of these systems - 5.1 surround - has too few full range speakers to facilitate accurate sonic immersion.

A more recent immersive audio delivery method that does not suffer from the limited sweet spot of surround sound technologies is Wave Field Synthesis (WFS). WFS uses the Huygens Principle (Pao & Varatharajulu, 1976, p. 1361) to replicate the propagation of sound waves through a space from a linear array of loudspeakers (Ranjan & Gan, 2013, p. 20). Cost and practicality are a concern with this technology, as many speakers are required to construct a linear array of speakers capable of achieving sufficient realism (Theile, 2004, p. 127).

The third technique for immersive sound reproduction is binaural. This method is of most interest here as it is utilized by MARA technology. Rather than physically produce or synthesize the sound in the space around the listener, binaural techniques present the acoustic properties of the space directly to the ears through headphones (Ranjan & Gan, 2013, p. 18). In addition to the sonic information contained in stereo headphone listening, binaural sound accommodates for the effect of the listener’s ears, head, and torso on the reception of sound (Hiipakka, 2012, p. 8). By sending the same sound information to the ears as what would naturally be received, it is possible to create a highly realistic representation of a three-dimensional sound space.

Tchad Blake is an example of a well-known producer who has utilized binaural recording techniques. Blake’s use of binaural ranges from rock records (such as Pearl Jam’s ‘Binaural’), to electric mandolin and soundscape recordings (Tingen, 1997). Despite the commercial representation of binaural, the format remained largely unknown to most music consumers. This can partly be attributed to several technical flaws that are present in binaural sound. These include potential for front-to-back localization confusion and errors in judging elevation, particularly when non-individualized HRTFs are used (Sung, Hahn, & Lee, 2013, p. 1). Further, listener immersion is compromised when turning the head because the

entire soundscape concurrently rotates.

These limitations are addressed in MARA with the addition of head tracking. Hiipakka explains how head tracking is of importance here:

In natural circumstances people use head movement to achieve more accurate sound source localization. It has been shown that using head tracking in [binaural reproduction] reduces significantly front-back and back-front confusion and increases the probability of externalization of auditory events. (2012, p. 24)

The relationship between head tracking and MARA will be explained further in the sections that follow.

## 2.3 Sonic Immersion and the Human Auditory System

Baalman explains that to achieve a realistic implementation of spatial audio, the “physical properties of sound propagation and/or the psycho-acoustic characteristics of our spatial hearing” need to be acknowledged (2010, p. 209). For listener immersion to occur on the deepest level, each aspect of our auditory system’s localisation methods must be satisfied. This section will provide an overview of how our ears decipher audio input.

On the horizontal plane, localisation is primarily determined by Inter-aural Time Differences (ITDs) and Inter-aural Level Differences (ILDs) (Albrecht, 2011, p. 3). The ITD is the discrepancy in arrival times caused by the separation of the ears. For low frequencies that have wavelengths longer than the radius of the head, our hearing system calculates direction by cross-correlating the phase differences that result from different arrival times (Mauro, 2012, p. 11). While ITDs have some use for higher frequencies, ILDs are most efficient for high frequency localization. Due to the density and makeup of the head, at high frequencies, “the head will be shadowing the one ear more than the other, causing a differ-

ence in sound pressure level between the ears” (Albrecht, 2011, p. 4). The more perpendicular a sound arrives to the head, the more pronounced the ILD will be.

The Head-Related Transfer Function is a model of the spectral filtering that occurs as sounds are diffused and absorbed as they propagate through our body. Albrecht notes that this filtering is primarily caused by the pinna, whose geometry shapes the spectrum of sound in relation to its angle of incidence (2011, p. 6). This shaping can also be caused by reflections off the head, shoulders, and torso to a lesser extent (Hiipakka, 2012, p. 8). Information provided by spectral filtering is used to disambiguate front-to-back location. Without this, sounds on the horizontal azimuth could nearly always be attributed to two possible locations. When considering this scenario over our full sphere of spatial hearing (ie. 3D rather than 2D), the ambiguity of location expands to form what is known as the, “cone of confusion” (Mauro, 2012, p. 10).

In addition to spectral filtering, the hearing system uses small “unconscious” (Moustakas, 2012, p. 3) head movements in order to increase the amount of location data available and improve its integrity. While Härmä et al. say that this changes the HRTF’s effect on the sound (2004, p. 625), the ITD and ILD are similarly effected by head turning, granting us the ability to gather a more accurate perception of an object’s location.

An often overlooked contributor to sound localisation is vision. Tikander, Karjalainen, & Riikonen explain that visual cues can be powerful in our calculation of where a sound is coming from. Tikander et al. note that, “When binaural signals were recorded and listened to later without visual information, frontal externalization was not as good anymore” (2008, p. 3). Similarly, Albrecht states that, “Without cues provided by head movement, the lack of visual information accompanying an auditory event will likely result in localization behind the head, even if acoustical cues suggest a location in front of the head” (2011, p. 6). This phenomenon poses interesting challenges when dealing with an art form restricted to auditory sensory inputs. Practical techniques for compensating for the lack of visual stimulus are discussed in Chapter 5.

Aside from the ability to determine the direction of sound, the human hearing system has the ability to estimate the distance of a sound source, as well as the properties of the space that it inhabits. This is calculated primarily using the reverberant qualities of sound. A sound can typically be divided into three distinct components: direct sound, early reflections, and late reverberation (Baalman, 2010, p. 214). The ratio between direct sound and early reflections gives an indication of distance - the more early reflections relative to direct sound, the further away a sound will be. The decay time and spectral colouration of the late reverberation component provide cues as to the size and composition of the space.

The Sound Pressure Level (SPL), or volume of a sound can also contribute to the perception of distance. As a sound becomes farther away, its volume will decrease due to the dispersion and absorption of the sound through air. As this occurs, the spectral and dynamic components of the sound also change due to air absorption (Huopaniemi, Savioja, & Karjalainen, 1997, p. 1). In particular, transient information becomes less defined, and high frequencies become attenuated more rapidly than low frequencies. Interestingly, the Fletcher Munson Principle is also related to volume. This principle describes the phenomenon where high frequencies and low frequencies become more apparent as volume increases (Göttlinger et al., 2009, p. 3). In synergy, the human auditory system uses these principles to determine the location of a sound source relative to the body.

## 2.4 Previous MARA Development

In the field of MARA technology, research has been conducted in the codes of: location and orientation tracking; auralization and data interpolation; MARA mixers; and applications for MARA. Of note, an overwhelming majority of this research has been funded by the Nokia Research Centre through projects such as KAMARA (Killer Applications for Mobile Augmented Reality Audio) (eg. Gamper & Lokki, 2009). Further, much of this research has been undertaken at the Helsinki University of Technology (now named Aalto University). Lead-

ing academics include Mikka Tikander, Matti Karjalainen, and Aki Härmä, who have published a series of works that detail the implementation and useability of MARA technology. Tikander's, *Development and Evaluation of Augmented Reality Audio Systems* (2009) is a central paper that outlines the author's work in developing the technology since 2002.

The funding supplied by the Nokia Research Centre in this field provides explanation as to why most sources focus on telecommunication applications. Authors external to Nokia such as Lyons et al., Lemordant & Lasorsa and Moustakas et al. have implemented the technology into navigation, gaming, and audio tour applications.

Three major components of a MARA system are the headset, user tracking, and auralization processing. Each of these aspects have been discussed in detail in existing literature. The first required component for a MARA system is the headset. The goal of the headset is to provide the facility for binaural sounds to be played to the listener without the natural sound environment being altered. Tikander et al. (2008) provide an overview of their prototype demonstration in *An Augmented Reality Audio Headset*. In this instance, small microphones capture the sound of the surrounding environment and the output is played through earphones worn by the user (p. 2). The resulting audio effect is known as the 'pseudoacoustic environment' (Rämö & Välimäki, 2012, p. 1). Lokki et al. (2004), Albrecht et al. (2011), and Rämö & Välimäki (2012) present alternate iterations of this headset, although it must be noted that these authors are linked through Aalto University.

The second component of the MARA headset is the ARA mixer. Rämö & Välimäki note that, "An essential part in creating a realistic ARA system is the ARA mixer" (2012, p. 2). A central paper on this topic is provided by Rikonen et al., named *An Augmented Reality Audio Mixer and Equalizer* (2008). Here, the conversion of the ear canal from a quarter-wave resonator to a half-wave resonator as a result of occluding the ear with an earphone is recognised. To mitigate this effect, an equalisation curve should be applied (eg. Gamper & Lokki, 2009, p. 1). In addition to simulating the natural resonance of the ear canal, this

equalisation curve is used to accommodate for frequency build-up caused by spill through the earphones from the outside environment. This is particularly the case for low frequencies that are a greater issue for spill because of their longer wavelength (Riikonen et al., 2008, p. 2).

The next component of a MARA system is user tracking. Various applications to date have integrated head tracking, position tracking, or a combination of these into their MARA systems (see Härmä et al., 2003 and Härmä et al., 2004). For the purposes of this project, it is desirable to incorporate both location and orientation tracking. Many methods for head tracking currently exist. These mainly fall under the categories of acoustic, magnetic, optical, inertial, and mechanical (Peltola, 2012, p. 2).

Karjalainen et al. present a particularly elegant method of both head and position tracking with their 2004 paper. This method makes use of the microphones that are part of the MARA headset. Here, high frequency anchor sources (inaudible) are emitted from three or more speakers around the room. A time code is embedded into these signals, so that the distance between each microphone and each speaker can be calculated given the speed of sound. Since the MARA headset consists of two spaced microphones, both the location *and* orientation of the user is known when the position of each microphone is known. While this method is not suitable for some MARA applications (eg. Peltola, 2012), it is highly applicable to the musical use of MARA. This method does not require additional hardware to be worn, thereby maintaining its transparency to the user. Further, surround sound speaker systems that are common in many listening environments could be used as known anchor sources.

The final component of MARA is auralization. This is the process of adding binaural acoustic cues to an anechoic sound signal so that it becomes externalized when listening in headphones (Albrecht, 2011, p. 9). These cues inform each of the aspects of the human auditory system as discussed in the previous chapter. There are two methods for applying the required auralization properties: through algorithmic emulation, or using convolution (Mauro, 2012, p. 19). When processing an anechoic signal, spatial properties need to be added and the

HRTF needs to be accounted for. The combination of these factors is known as the Binaural Room Impulse Response (BRIR) (Menzer, Faller, & Lissek, 2011, p. 396). It is possible to use algorithmic emulation or convolution to simulate either component of the BRIR, as described by Härmä et al. (2003) and Mauro (2012). The topic of auralization will be expanded upon in detail in the design component of this paper.

# Chapter 3

## Methodology

The goal of this project was to design a Mobile Augmented Reality Audio system for creative musical applications. The following section describes the research methodology and set of methods that were adopted to achieve this goal. The project was broken down into three main design segments: the MARA headset, auralization engine, and user tracking unit. The construction of these modules presented sub-questions to be considered: How could each segment best be designed to accommodate a musically creative experience? The process of research, development, and assessment that occurred here can be described as ‘Design Research’ (Downton, 2003, p. 12).

### 3.1 Design Research

This project adheres to a Design Research methodology as a formal framework for the development of the MARA system. In its most distilled form, Design Research is concerned with the processes involved in accomplishing a predetermined design goal (Zimmerman, Forlizzi, & Evenson, 2007, p. 494). A design process involves constant decision making, and thus involves a complex array of potential variables (Collins, Joseph, & Bielaczyc, 2004, p. 15). The Design Research methodology provides a single context for this multitude of decisions.



It is acknowledged that embracing a meta-level “systems approach” is likely to result in alternate discoveries than if decisions are made out of context to the all-encompassing design goal (Gifford, 2011, p. 50). In addition, it has been observed that unexpected innovations often come to fruition during Design Research due to the exploratory nature of this methodology (Zimmerman, 2003, p. 176).

Design Research consists of two parts: the design component, and the critical assessment - or evaluation - of the design (Collins et al., 2004, p. 21). These parts are mutually reliant, and form a cyclical process called ‘iterative design’. Zimmerman defines iterative design as “prototyping, testing, analysing and refining a work in progress. [...] Test; analyse; refine. And repeat” (2003, p. 176). The ongoing aspect of this type of research allows the designer to engage with the stimuli on a deep level and accumulate a body of knowledge based on these experiences. In contrast to a lab-based, hypothesis-testing paradigm, iterative design allows the researcher to “explore complex product interaction issues in a realistic user context and reflect back on the design process and decisions made” (Keyson & Alonso, 2009, p. 4548). This is particularly relevant in this project, where tactile participation with the system in a real-world setting is a central element.

## 3.2 Theoretical Background

The emphasis on physical interaction is evident in the project itself, the research methodology, and the theoretical framing of knowledge in this project. The following approach to thinking has been embraced as it underpins key concepts that embody this topic and its surrounding fields. One such concept is Gibson’s perspective on ecological psychology. Gibson argues that human perception is intrinsically conjoined with interaction with our environment (1979, p. 2). Grandy (1987, p. 192) identifies that Gibson’s philosophy is analogous to an ‘ecological epistemology’, where knowledge obtained is a holistic construct of layers of perception and interaction in a real-world environment. Grandy notes Gibson’s opposition to the popular definition of information theory provided by Shannon

and Wiener (1987, p. 191), which by contrast focuses solely on the “technical level of communication” (Sveiby, 1996, p. 384). Rather than alienate theoretical knowledge from practice, the ecological epistemology acknowledges the need for a relationship between these two notions.

An additional source of knowledge in this context is derived from *enaction*. “Enactive knowledge is constructed on motor skills” (Luciani, 2009, p. 211). Wessel explains that sensory-motor engagement is a crucial component of the creation of music; further, the auditory experience is similarly engaging (2006, p. 93). The appraisal of a MARA system (the evaluation component of the Design Research methodology), involves extracting knowledge from both interaction and enaction. An ecological epistemology is a theoretical means for acquiring this knowledge.

### 3.3 Design

Now that the methodology and theoretical scaffold have been presented, a closer look at the components of Design Research is appropriate. Design Research is comprised of an iterative process of design and evaluation phases. These phases may take on different forms as appropriate for the project at hand. The term ‘Design’, as Downton states, can be used for “research *for* design, research *into* design, and research *through* design” (as cited in Gifford, 2011, p. 52). This project will focus on the latter of these, explained by Keyson & Alonso: “Research through design focuses on the role of the product prototype as an instrument of design knowledge enquiry” (2009, p. 4548). The cyclical nature of iterative design is highlighted in this sentiment, as it can be observed that the prototype is not exclusive of the reflective data that results from its scrutiny. In fact, this process sees the verb usage of the term ‘design’ become the noun, and vice versa through the course of “progressive refinement” (Collins et al, 2004, p. 18).

This project sees the simultaneous development of MARA headset, auralization engine, and user tracking prototypes. These prototypes are data collection instruments that inform the design of the MARA system. Creativity is core to

the entire process: evolving knowledge is being applied innovatively within each prototype, on a project-level, and within a musical headspace as well.

Knowledge derived from the prototype is not the only source of information. “Designing inevitably employs various kinds of knowledge derived from both sources outside the designing and design as well as sources from within the designer’s own discipline” (Downton, 2003, p. 95). Therefore, the design phase involves a progression from research to internalisation, followed by the application of new knowledge. Research through design begets significant opportunity for continuous research throughout the project. As “subproblems” (Gifford, 2011, p. 51) arise through iterations of design, new research is often necessary to complement the researcher’s existing knowledge. In the case of this project, considerable research into fields including electronics, signal processing, interactive and immersive media, and communication protocols was necessary in response to subproblems presented throughout the design process.

### 3.4 Evaluation

The second component of Design Research is the evaluation stage. This stage involves data collection and synthesis that is used to assess the design. Despite creative practice and music being subjective fields, this project can draw from quantitative data generated from comparison of the augmented reality audio environment with the natural audio environment (as demonstrated in Tikander et al., 2008). In this case, the author’s version of ‘natural hearing’ is a fixed reference point that observations can be based upon. These observations hence form the data collection ‘tool’, and will be compiled under the codes of user immersion and acoustic realism. The analysis of this data can be viewed from a literal perspective, however conclusions should be reflexive, sympathetic to the fact that human hearing is unique for each person. Each component of the system is evaluated in this way in Chapter 4.

Aside from assessing the technical capacity of the system, the emphasis on creativity in this project calls for an evaluation of the creative potential allowed by

the system. Under a similar Design Research methodology, Gifford employs ‘aesthetic evaluation’ for this task, where he puts the system to use before “reflecting on both the musical quality and the intuitiveness of the interaction” (2011, p. 53). In this paper, Chapter 5: Making Music with MARA discusses findings from my aesthetic evaluation of this system. It is important to delineate aesthetic evaluation from creative work evaluation here. The ability for the *system* to be used in a creative context is of primary interest in this case, whereas the musical content itself is of secondary concern.

Technical and aesthetic evaluation of the design can occur on a component level or a systems level (Gifford, 2011, p. 50). In Design Research, each component is tested in isolation initially, and then re-tested in context of the entire system. This poses the questions: ‘How well does it fulfil its intended role by itself?’ and, ‘Is it successfully contributing to the broader system, or is modification required in sympathy with other components in the system?’

### 3.5 Methodologies in Existing MARA Research

The term ‘methodology’ is rarely acknowledged in literature on MARA. When it is used, it could perhaps be better replaced by the term ‘methods’ (for instance, in Moustakas et al., 2012). It appears that the traditional scientific formula of aim, hypothesis, test and analysis is assumed as the basis for writing academic papers in this field. However, in designing novel contributions to MARA technology, these researchers have followed what closely resembles a Design Research methodology. As such, there is an abyss in the written literature that lends itself to a definitive account on the *procedure* that was followed in the development of a MARA system as opposed to the scientific contributions made. A clear account of the design choices made with reference to the project goal would be a valuable addition to the existing body of knowledge.

### 3.6 Method

A creative practice method has been followed that recognises both the technical and creative components of the project. As such the research has been, “initiated in practice, where questions, problems, challenges, are identified and formed by the needs of practice and practitioners ... and the research strategy is carried out through practice” (Gray as cited in Gifford, 2011, p. 4). My practice lies within Music Technology, where I seek to blend music and technology in a cohesive and creative manner. Similarly to the Design Research methodology, a process of critical reflection and evaluation occurs in the creative practice paradigm (Gray, 1996, p. 8). In addition, creative practice can form both the “driver and outcome of the research process” (Hamilton & Jaaniste, 2009, p. 1). This method is hence highly applicable for a creative design project.

The creative practice process has been divided into three modules that form the basis for MARA technology: headset development; location and orientation tracking; and auralization. Each module presented unique sub-questions, which required varied problem-solving approaches. A process of research, experimentation, application and testing was, however, consistent throughout.

### 3.7 Summary

This project followed a Design Research methodology under an ecological epistemology. The method was grounded upon the author’s creative practice of artistic application of musical technologies. This approach was chosen in harmony with the concept of interaction, which is central to this topic.

*A note for the reader:* The next chapter details the design/evaluation cycles that occurred to develop the current MARA for Music System. This process was inherently technical, although informed by creative decisions. As a result, the language in the following chapter is quite technical, drawing upon specific music technology content. This language is a platform to illustrate musical ideas - much

like the way the language of notation is used to communicate musical ideas. This chapter should not, therefore, be considered a departure from musical or creative themes. From Chapter 5, less-technical dialect is used as the implications of the invention become a focus.

## Chapter 4

# Designing a MARA System for Music Creation

### 4.1 Introduction

The following section details the Design Research process throughout this project. Relevant background knowledge is discussed before prototypes are presented and critiqued. These design iterations represent several of the small cycles that occurred during the overall research through design progression.

### 4.2 MARA Headset

#### 4.2.1 Design Goal

The purpose of a MARA headset is to create a transparent representation of the natural acoustic environment while maintaining the functionality of earphones. The headset needs to be discreet and mobile to allow for unrestricted movement. The sound quality must be substantial enough to reproduce subtle sound localisation cues.

### 4.2.2 Prototyping Stages

In their paper, ‘An Augmented Reality Audio Headset’, Tikander et al. outline the components of a MARA headset (2008). A later prototype of this headset is used in ‘A Mobile Augmented Reality Audio System with Binaural Headphones’ by Albrecht et al. (2011). Both papers were partly funded by the Nokia Research Centre, so have a focus toward telecommunication uses. These papers provide an overview of the necessary components in a MARA headset, however do not discuss the choices made in reaching the final design state. Instead of working toward a design goal, Tikander et al. and Albrecht et al. present operational prototypes before discussing their potential applications. The headset development aspect in this project applies findings from these papers in the context of a creative musical design goal.

The first step in this process was purchasing the Philips SHN2500 noise-cancelling earphones. In normal use, the output from the built-in microphones in a noise-cancelling system is phase-inverted and fed into the listener’s ears. In a MARA system, the microphone signals can instead be used to create a pseudoacoustic environment. The Philips SHN2500 model has been used in the majority of documented MARA headsets. These earphones are low-cost, and are simple to disassemble. As they are intended to be used for music listening, they are an appropriate choice for a musical MARA headset. The circuit layout of the unaltered earphones is shown in Figure 4.1. The PCB is responsible for phase inverting and amplifying the recorded ambient sound before mixing this with the earphone audio.

To use the noise cancelling earphones for MARA, the microphone signals need to be amplified and mixed with the earphone audio input. This project saw the design of a number of prototype circuits to accomplish this task. In the first of these, the microphone signals were tapped before they entered the PCB. They were then connected to the preamp stage of an analogue mixer. A circuit diagram is shown in Figure 4.2.

There were several problems with this design. The microphone signal was barely



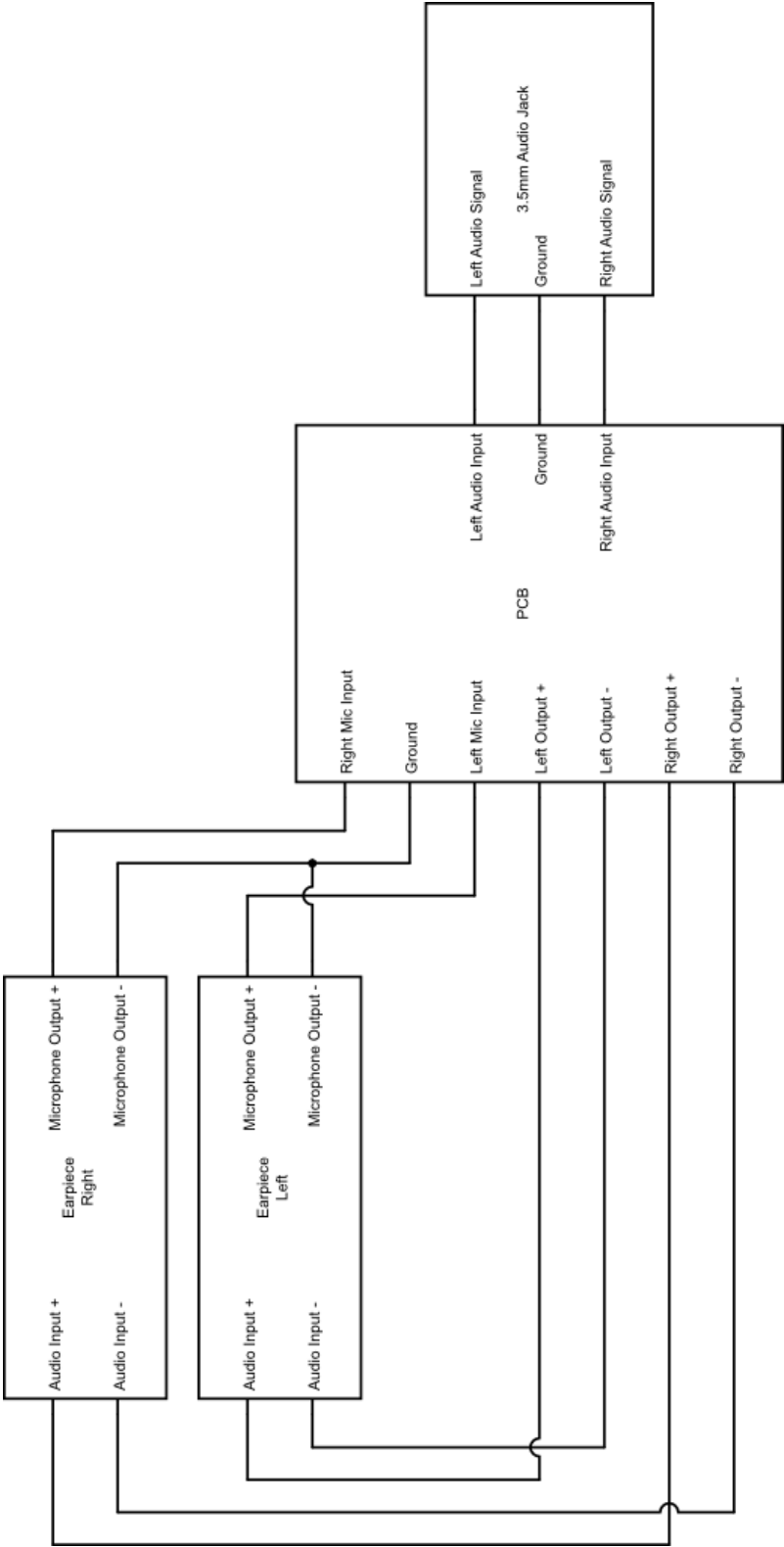


Figure 4.1: Original circuit of the Philips SHN2500 noise-cancelling earphones

audible even when loudly tapping microphones with the preamps at full gain. In addition, there was a very high noise floor. To improve the signal-to-noise ratio, it was necessary to place a preamp closer to the microphones. This would mean that the low-level signal would need to travel less distance through a high-impedance, unbalanced cable.

The second iteration of this circuit utilized the built-in preamp of the PCB. To maintain isolation between the audio signal and microphone signal, the PCB was bypassed between the 3.5mm jack and the earphones. This meant that the output from the PCB would now be the amplified microphone signal only. In this version, the circuit board outputs were wired backwards to reverse the phase. This counters for the phase inversion that occurs in the PCB's noise-cancelling process. This circuit is shown in Figure 4.3.

This design saw the signal-to-noise ratio significantly improve, so the microphone signals could clearly be heard through the earphones. However, early testing found that localizing sounds in the space was impossible. It was found that the PCB was summing the two microphone signals to mono at some point in the circuit. This is likely a result of the manufacturer attempting to reduce costs. Further, noise-cancelling devices are designed to reduce ambient noise, which is essentially omnidirectional. Therefore, the noise-cancelling system does not require a stereo noise signal. This meant that the PCB supplied with the earphones could not be used as a pre-amp for the system.

The third circuit design saw the original PCB replaced with two battery-powered mono preamps. These preamps were small enough to fit inside a pocket, which was in keeping the mobility aspect of the headset. The resulting audio signal was at line level, so required no further boost before being fed into the earphones. This circuit can be viewed in Figure 4.4.

The pseudoacoustic environment resulting from this design was quite satisfactory. When wearing the headset, it was difficult to tell whether the microphones were working or not - the pseudoacoustic environment was often indistinguishable from a natural hearing experience. Localising sound sources was a natural task, and the ability to discern front-to-back location was good, particularly when aided by

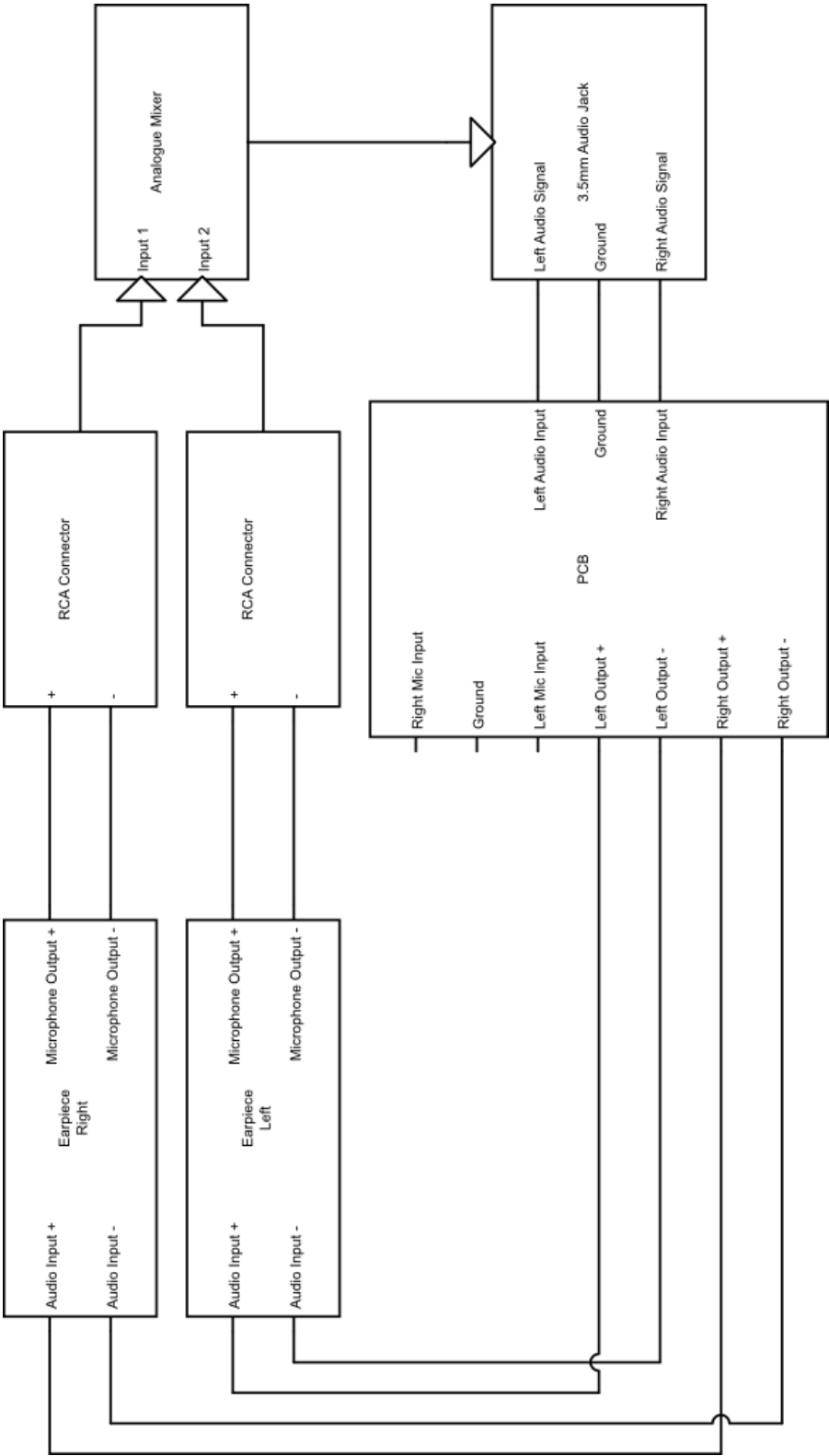


Figure 4.2: Headset preamp circuit iteration 1

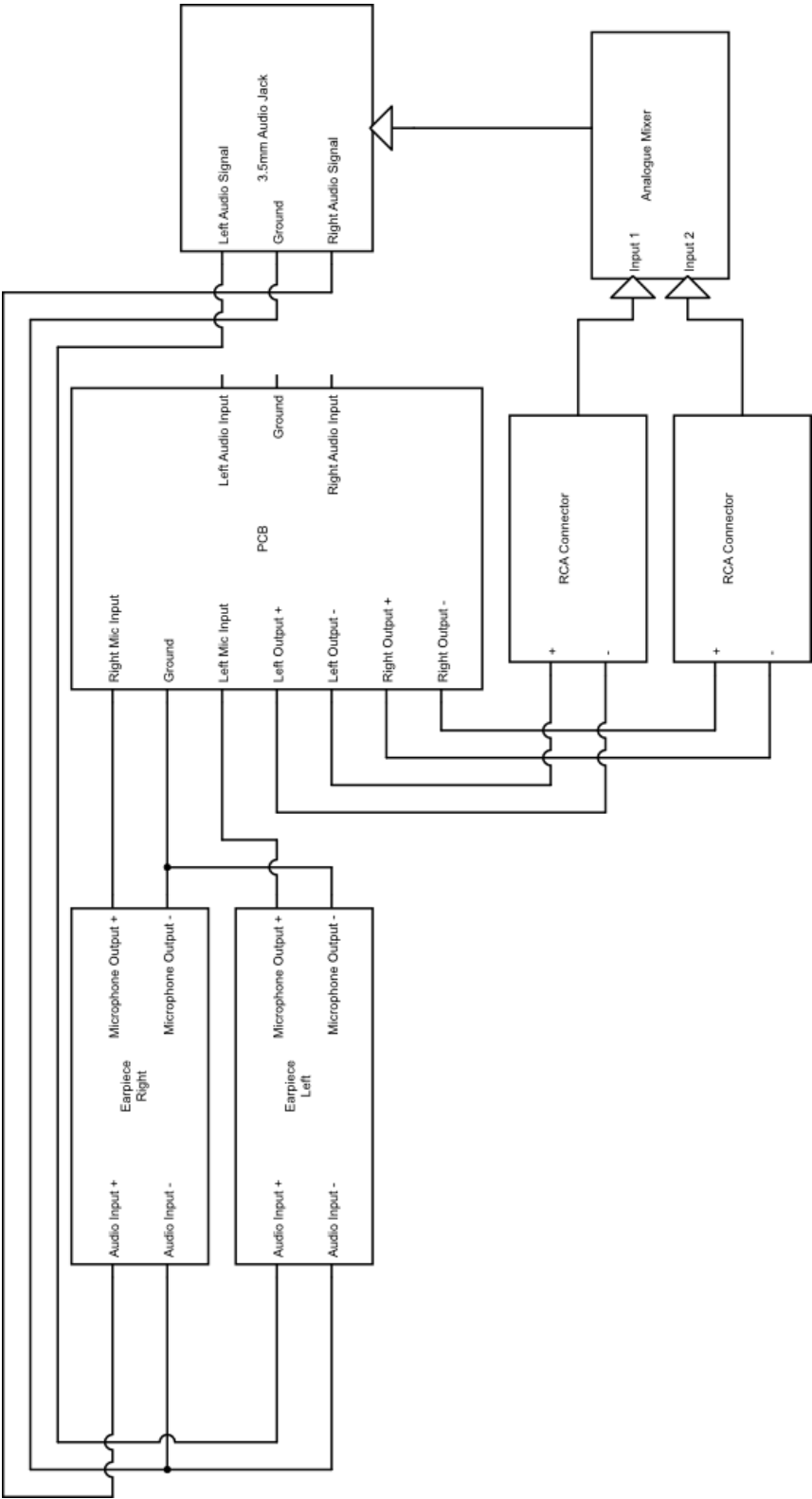


Figure 4.3: Headset preamp circuit iteration 2

visual cues (as found in Tikander et al., 2008, p. 3).

The forth (and current) iteration of the MARA headset improved the usability of the prototype (see Figure 4.5). While Iteration 3 demonstrated a working concept for the headset, it relied upon a wired connection to a mixing console, and was not appropriately contained for practical use. In response to this, a battery powered analogue mixer comprising of two 10 ohm logarithmic double-ganged (stereo) potentiometers and the two microphone pre-amps was built. This circumvented the need for a mixing desk, which meant the unit was now completely mobile.

The circuit was housed in a rugged enclosure that was fitted with a belt clip, a power switch, and the potentiometers positioned for ease of access. Some minor updates to the circuit also occurred, including the addition of a pre-potentiometer auxiliary microphone output. This was incorporated with Tikander's 'Common Modulated Anchor Sources' position tracking method in mind, which requires a split feed of the captured binaural signals. As this tracking method has since been abandoned, these outputs are not in use for this system. Finally, the power source for the circuit was streamlined to a single battery to conserve space in the enclosure.

While the MARA mixer was now mobile, a method of inputting the auralized audio wirelessly was required so that the user's movements were not restricted by an audio cable. To achieve this, real-time WiFi audio streaming between a Mac computer and iPhone has been used via an App named 'Airphones'. A number of similar applications exist to facilitate Mac to iPhone audio streaming, however 'Airphones' is unique as it is designed to be as instantaneous as possible - an appealing feature for MARA, where latency corrupts user-stimuli interaction. A compromise for this near-real-time approach is that 'Airphones' supports a maximum of 16bit/44.1k audio quality, meaning sessions at higher resolutions need to be down-sampled and truncated before they are compatible with this system.

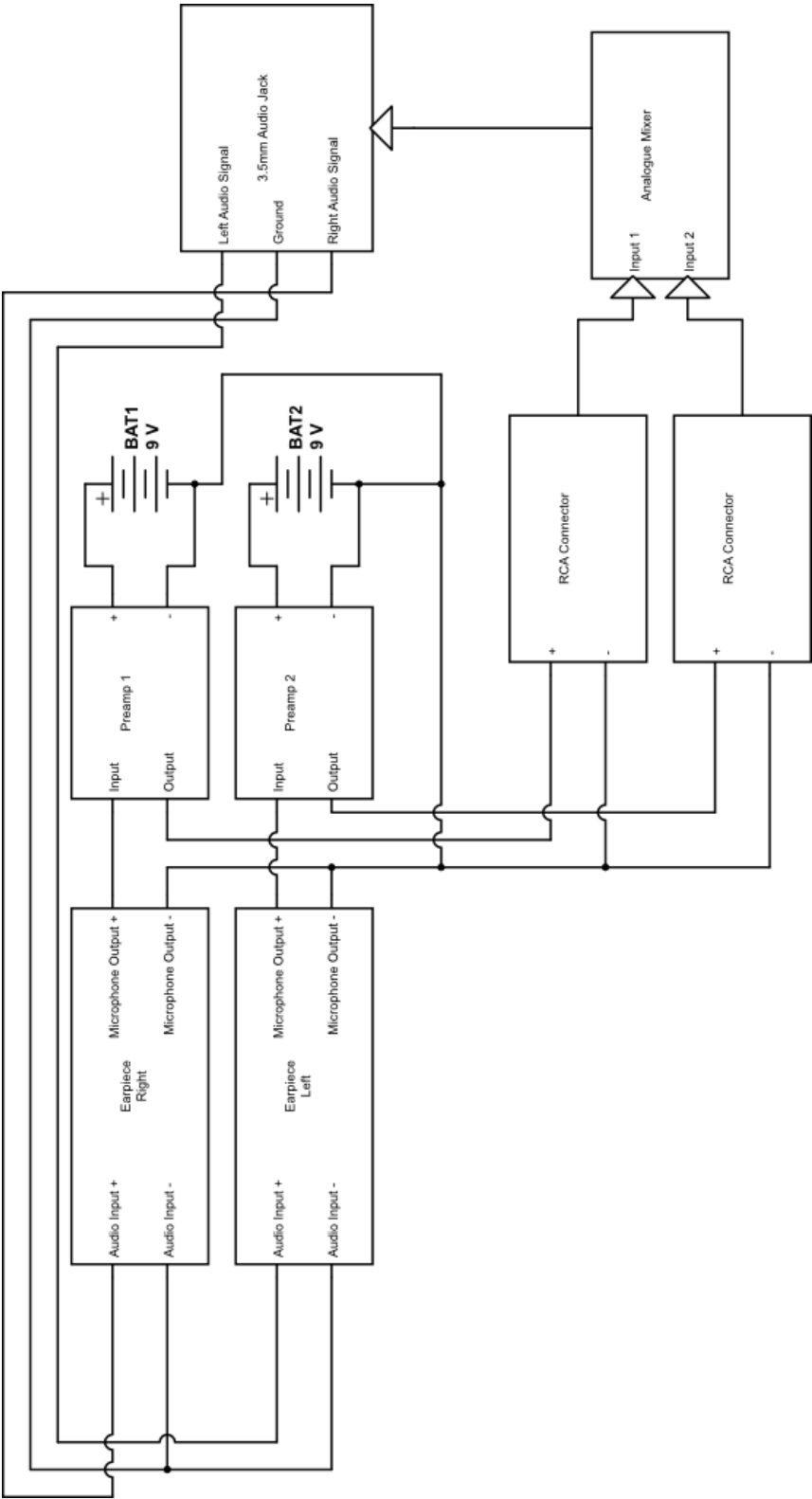


Figure 4.4: Headset preamp circuit iteration 3

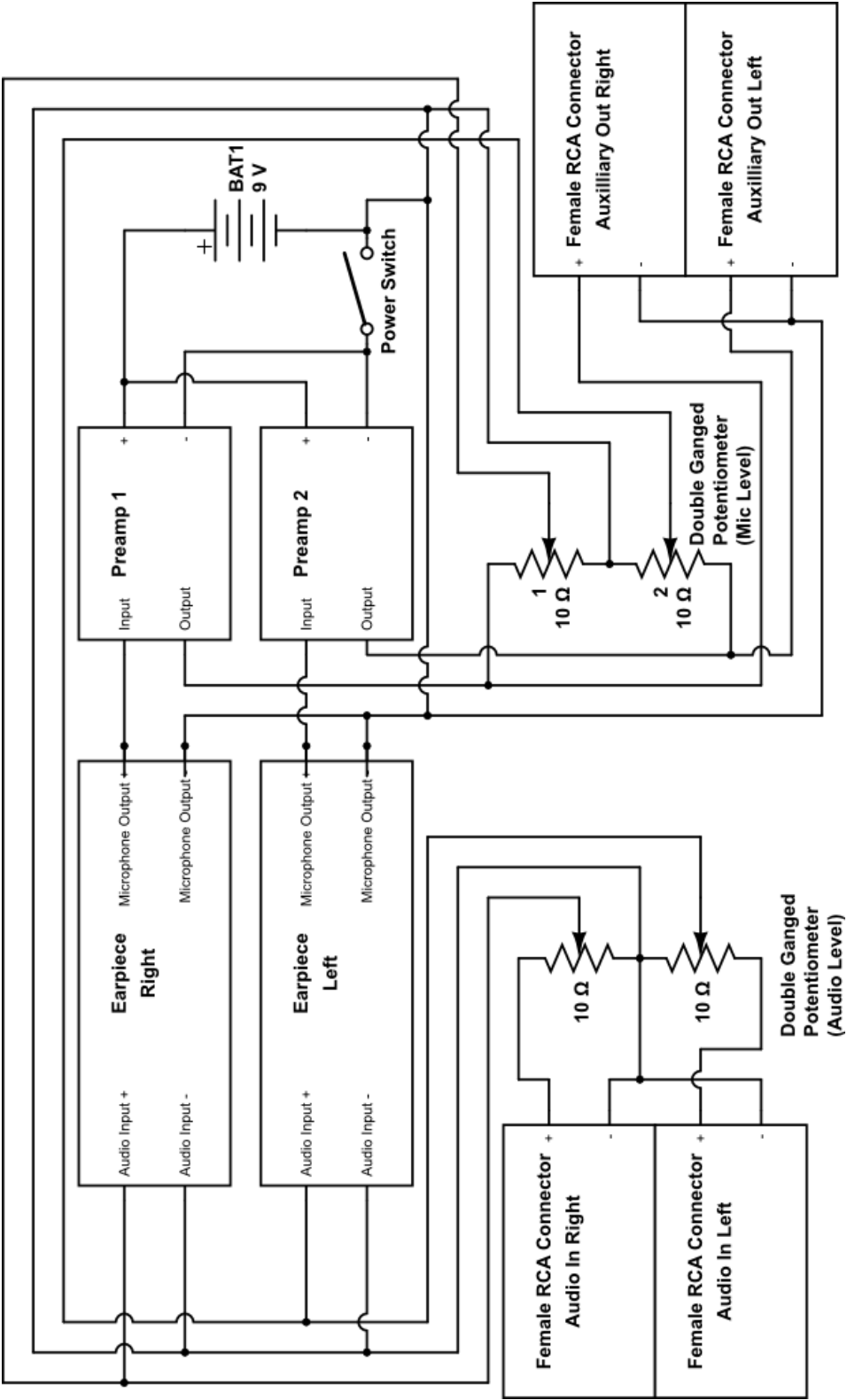


Figure 4.5: Iteration 4: MARA mixer circuit



Figure 4.6: The MARA mixer



### 4.2.3 Current Status

The current iteration of the MARA headset performs its function as intended, however upgrading some components would improve the performance of this unit. Weaknesses of the current build include a slight noise floor (also noted by Albrecht et al., 2011), and somewhat low-quality audio reproduction from the earphone speakers.

### 4.2.4 Future Improvements

Riikonon states that an equalisation curve is usually required in a MARA mixer to compensate for the changed resonant properties of the ear canal caused by obstruction from the earphones (2008, p. 3). Interestingly, the proposed EQ curve appears similar to my observations of the natural frequency response of the Philips SHN2500 earphones (no frequency response graph has been released). This (informal) observation was made by comparing the spectral qualities of the Philips SHN2500's to familiar reference headphones. It has been deemed beyond the scope of this project to incorporate an equalisation curve in the mixer, however this is a viable area for future improvement in conjunction with upgrading the earphone/microphone components.

## 4.3 Auralization Engine

### 4.3.1 Design Goal

The goal of the auralization engine is to provide a software-based platform for applying the three-dimensional sonic information of a space to sound sources. In this case, the auralization engine seeks to recreate the acoustics of the IMERSD live room at the Queensland Conservatorium of Music as realistically as possible. A heavy focus on creativity was placed on the implementation of auralization. The process of placing sounds in the augmented reality environment

needed to be intuitive and rewarding to encourage creative musical use of the system.

## 4.3.2 Development Process

### 4.3.2.1 Space and Direction

The reverberant properties of a room in combination with the HRTF are required to perform auralization. In this case, convolution has been chosen as the method for determining both the reverberation and the HRTF in the chosen room. The convolution method results in a highly realistic representation of the room's acoustic properties, more so than current digital emulation approaches. This is favourable in the context of ARA, where acoustic realism is important because the user has the reference point of the room's *actual* acoustics being fed into their ears through the MARA headset.

By capturing Impulse Responses (IRs) binaurally, the BRIR is obtained. The BRIR holds all of the information required for auralization (Gamper & Lokki, 2009, p. 1). For this project, seven BRIRs were captured that correspond to different positions on the horizontal azimuth around the room. The BRIRs were made by placing the diaphragms of small omnidirectional microphones in the entrance of the author's ears and recording a sine wave sweep (as recommended in Müller & Massarani, 2001, p. 443). This results in a non-individualised HRTF for other users of the system. While non-individualised HRTFs are most common with binaural reproduction, individualised HRTFs are preferred (Hiipakka, 2012, p. 1). Research into the practicality of obtaining individual HRTFs easily for MARA is progressing (eg. Gamper & Lokki, 2009), and is an area for future improvement with this system.

Taking inspiration from Matthew Hitchcock's paper, 'Morphing Aural Spaces in 7.1 Music' (2012), this design process has seen the development of a novel method for BRIR interpolation. This method uses the ProTools 7.0 surround sound panner as an interface to blend a signal between multiple convolution

engines. This is as opposed to the 7.0 panner's normal function, which is to blend signal between multiple speaker outputs. This 'convolution-panner' allows for real-time manipulation of the location properties of a sound within a room. This can be used during the mixing process, and has been integrated with a user tracking system so that it adjusts auralized sounds to remain fixed within the ARA space.



Figure 4.7: The ProTools 7.0 panner (listener's head is superimposed)

Figure 4.7 depicts the ProTools 7.0 panner with a listener's head superimposed in the listening position. The green 'ball' represents the location of a sound source, which, in the surround sound context, is manipulated by blending the relative volumes of each speaker. In the convolution-panner, the green speakers instead represent convolution engines that contain BRIRs taken from the locations of these speakers. Moving the green ball therefore blends the level of signal sent to each convolution engine, which in turn provides the listener with acoustic cues with which to localise the sound.

Several convolution engines were tested to determine the best solution for this context. It was found that Altiverb offered the most control over how the IR was integrated. In particular, the ability to 'colour' the direct sound with the IR meant that externalization remained when increasing the ratio of direct sound to reverberated sound.

The convolution-panner is quite successful at performing auralization and automating sounds to move about the space through time. One limitation of the design is that it is restricted to 7 points of ‘resolution’ as given by the ProTools surround sound panner. When placing sounds in between two IR positions, the location of the sound blurs slightly, due to the differences between each IR. This blurring effect does not occur when panning in stereo, for instance, because the sound coming from each speaker is identical and only the relative volumes change. In theory, this problem makes an argument against the use of binaural convolution here, but in practice, the blurring is almost negligible, and stable soundstages are possible even in worst-case scenario locations. Audio examples and descriptions that demonstrate this are provided in Appendix A.

#### 4.3.2.2 Depth

In addition to being able to manipulate the location of sounds over a 360 degree horizon, the auralization engine also needs to be able to control the perceived distance of the sound source from the listener. This is necessary for both the construction of three-dimensional soundstages within the space, as well as for controlling the relative distance of sounds when the user moves to explore their surroundings. My initial concept for allowing this involved capturing an additional seven BRIRs at a closer proximity to the listener. When experimenting with the Altiverb engine though, I found that increasing the ‘direct’ component of the (coloured) IR successfully changed the perceived distance of the sound. This solution is favourable, as it avoids the aforementioned blurring that occurs if two IRs are dissimilar - blending will occur between two Altiverb settings based on the same IR.

In light of this, each of the seven auxiliaries that contained BRIRs were duplicated, and the duplicates adjusted so that they auralized sounds within a small radius of the listener. The next step was to find a method of interpolating between the ‘near’ and ‘far’ mixes, and therefore variably controlling the perceived distance of the sound. A method of routing the output of each track was devised so that depth could be controlled by a single pan pot in ProTools.



Figure 4.8: The signal routing method for manipulating depth of each sound source

Figure 4.8 shows the audio track titled ‘Snare’ being sent to stereo Bus 31-32. Below, two mono auxiliaries, ‘Far Snare’ and ‘Near Snare’ have been created, with inputs Bus 31 and Bus 32 respectively. With this configuration, the pan dial on the ‘Snare’ track is acting as a blend control between the auxiliaries below. The auxiliaries have been routed to 7.0 busses, where each channel of the bus is mapped to an individual convolution engine. ‘Pan grouping’ has been enabled on the ‘Far Snare’ and ‘Near Snare’ auxiliaries, so the green ball moves synchronously on the panners. Quite simply, the operator can now adjust the location of the sound source by using either of the surround panners, and the distance of the source by panning left or right on the ‘Snare’ track.

#### 4.3.2.3 Automation

A more advanced requirement of the auralization engine is the ability to automate the location and orientation of the entire mix in response to user movements: if the user walks forward, sound sources in front need to move closer, and sources behind become more distant; if the user rotates their head to the right, the mix needs to ‘spin’ an equal amount in the opposite direction. The forward/back,

left/right component of this can be achieved through an extension of the distance manipulation method. The philosophy behind this technique was inspired by the design of the ProTools surround sound panner: sounds can be moved in relation to the listener, but the listener can not be moved in relation to the mix. To create the illusion of sounds remaining fixed when the listener moves, we must therefore move the sounds around the user such that the user's movements are counterbalanced.

To set up location-reactive automation, each sound is first sorted into front/back or left/right groups, according to its location in the mix. Pan grouping is enabled for each group, so moving a single pan control will effect the depth of each instrument in the group. The inputs of the near and far auxiliaries are now set according to which side (front or back, left or right) of the room an instrument is. In Figure 4.8, for example, the snare is placed toward the right hand side of the room. If a guitar track was panned to the left side of the room, it's far/near auxiliary inputs would be reversed (Bus 34 to 'Far Guitar', and Bus 33 to 'Near Guitar', for instance). This means that when the snare moves closer, the guitar will move further away as it would in a real situation. The same concept is used for the front/back group, leaving us with two pan controls - one for the X axis, and one for the Y axis - that control the position of the entire mix.

### 4.3.3 Critical Evaluation

This approach allows for the location of the user to be automated unobtrusively within the ProTools session, although it is not without compromises. The process of grouping each instrument into a front/back or left/right group restricts the flexibility of the mix somewhat - as tracks cannot change groups automatically, a listener cannot walk 'around' a sound source. In fact, the angle of incidence of the sound is fixed. As Figures 4.9 and 4.10 demonstrate, this also causes an error in the location of sound as the listener moves.

Similarly to the blurring effect of the convolution-panner, this error is less damaging in practice than it appears in theory. Sonic cues depicting the distance and

direction of the sound source appear to dominate the mental calculations that suggest the sound has changed location. To reduce the occurrence of this error, sounds that encourage acute localisation can be panned to 0, 90, 180, and 270 - this will mean that when moving perpendicular to a wall, the error will occur on one axis only.

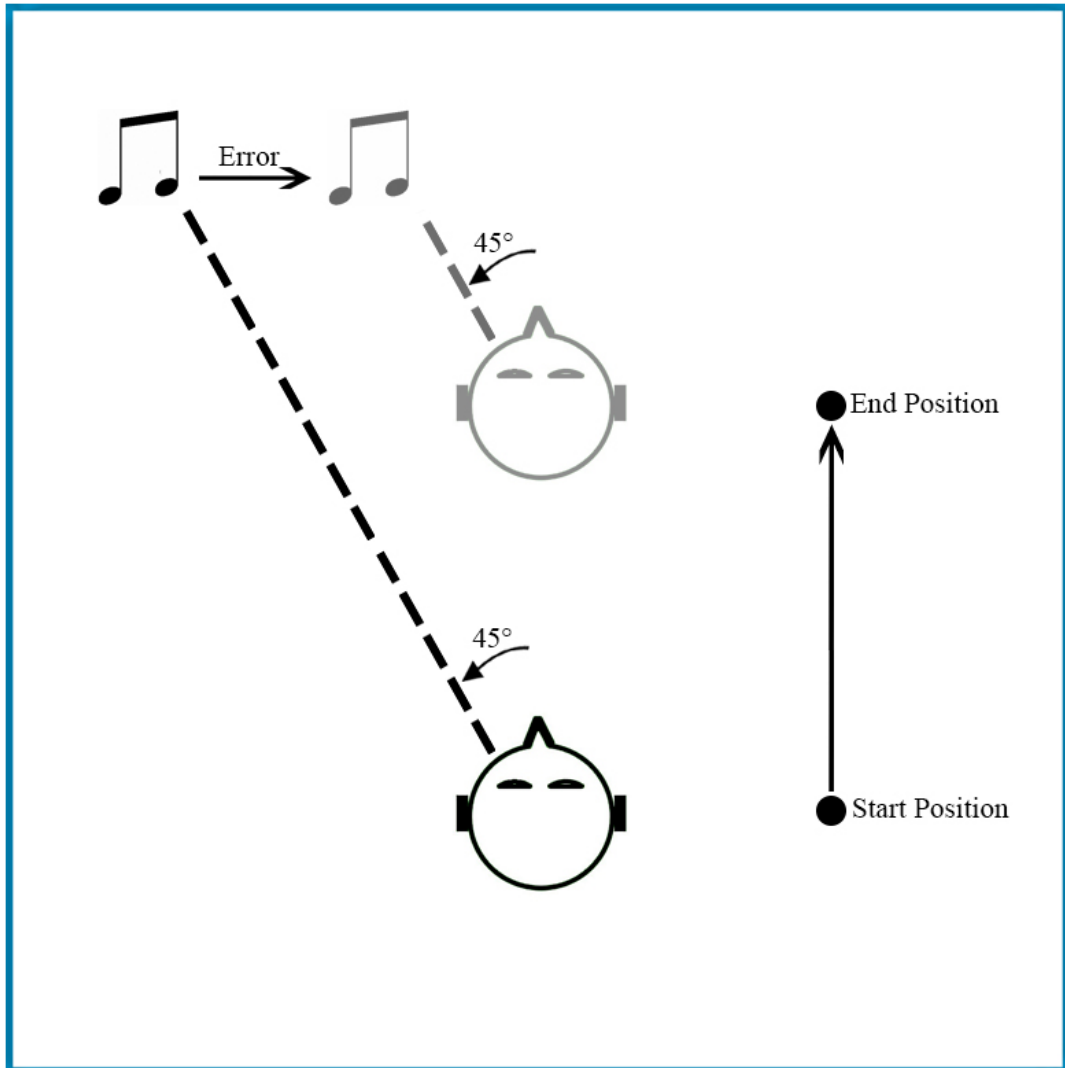


Figure 4.9: Location error of front/back group

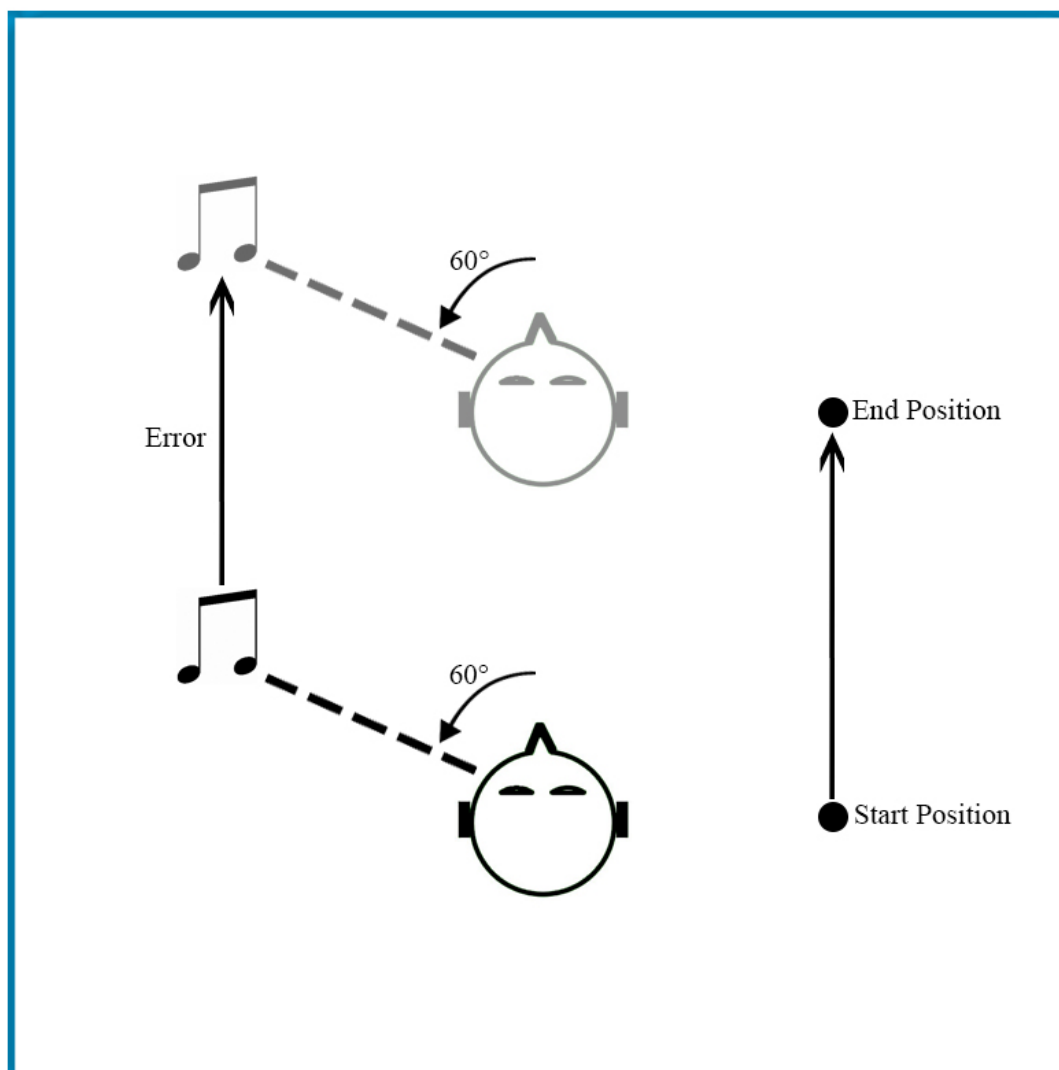


Figure 4.10: Location error of left/right group



The front/back left/right grouping method gives birth to four isolated soundstages upon which a sound artist may construct a mix. Each soundstage is capable of supporting a sense of depth between sound sources. This is achieved by offsetting the depth dials on particular tracks to taste. Figure 4.11 displays the plane for user movement and the four soundstages beyond. A limitation of the current system is that the user cannot breach the inner barrier of this soundstage. Departing from this front/back left/right model would allow the user to interact more fully with more complex mixes, and ultimately become more immersed in the experience. A more sophisticated method for mapping relative distance and orientation of sounds is an area for future investigation.

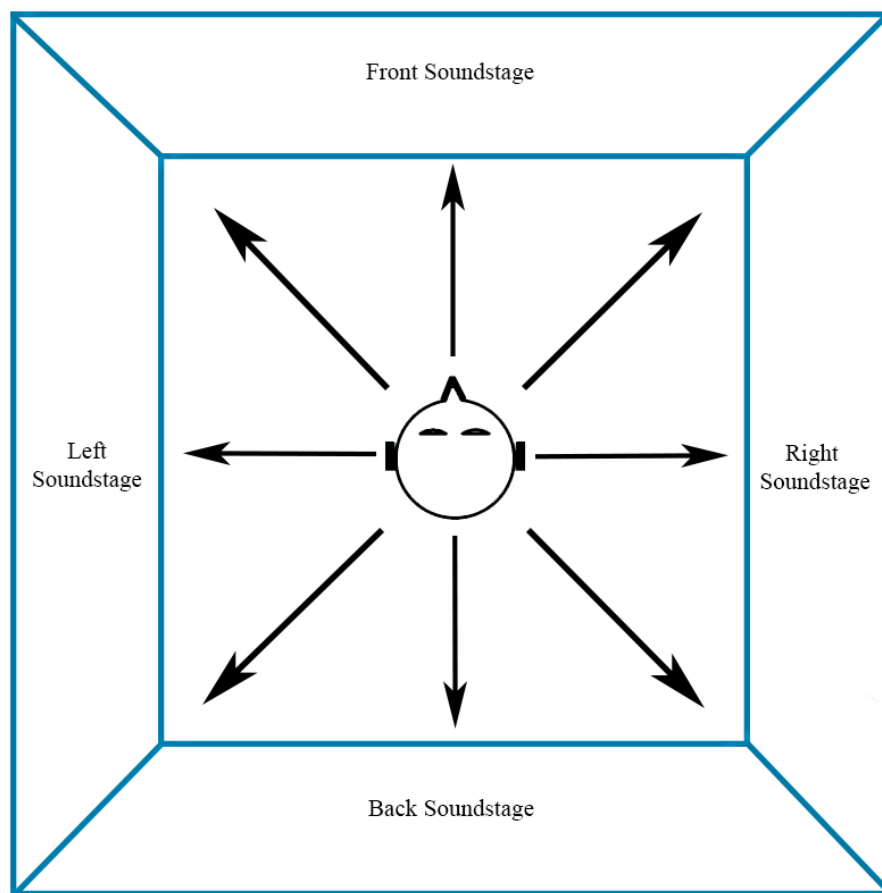


Figure 4.11: The four soundstages available with the option of depth between sound sources

## 4.4 User Tracking System

### 4.4.1 Design Goal

The user tracking system needs to be able to determine the evolving position and orientation of a person in an indoor setting. The system should not restrict the user's mobility or agility. Ideally, the user tracking system should make use of existing technology, thereby making it cost effective and easy to integrate on a large scale.

The accuracy of the tracking system is an important aspect of its functionality. The head tracking component needs to be able to distinguish subtle lateral head movements to make use of this aspect of human sound localization. The smallest change in angle that we are able to perceive is known as the Minimum Audible Angle (MAA) (Mauro, 2012, p. 14). As shown in Figure 4.5, the MAA is dependent on frequency as well as angle towards the side of the head.

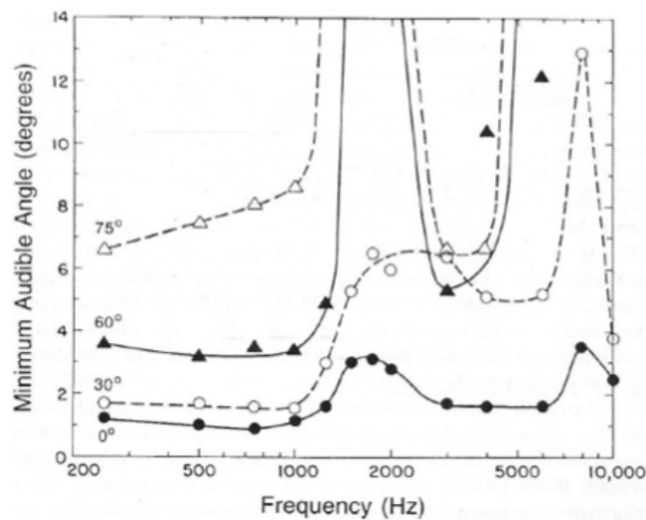


Figure 4.12: Minimum Audible Angle for sine waves at varying frequencies and azimuths (Mauro, 2012, p. 15)

As this figure shows, the MAA is around one degree when sound is in the midrange of human hearing and directly in front of the head. This would therefore be the

ideal resolution of the head-tracking system. However, as this scenario would only occur intermittently, a larger MAA resolution may be acceptable in this context. This is discussed further in the critical evaluation of this component.

#### 4.4.2 Existing Tracking Systems

As noted in Section 2.4, Karjalainen et al. present a working location and orientation tracking system for indoor MARA applications in their 2004 paper. This system is appealing for musical applications of MARA for several reasons. Loudspeakers projecting a reference signal are the only additional hardware required for this design. As loudspeakers are commonplace in most creative musical spaces, this tracking method is highly applicable here. The microphones used to record the reference signal are already part of the MARA headset, giving the headset an extra dimension of functionality. Further, as this method tracks both position and orientation simultaneously, two separate tracking systems are not required. The authors of this paper have demonstrated that the accuracy of this system is comparable to that given with an existing electro-magnetic tracking system.

Unfortunately, this paper offers little information with regards to the practical implementation of the system. The authors were unable to provide the required software templates to recreate this system, so an alternate tracking system needed to be developed.

#### 4.4.3 Developing and Implementing a Location Tracking System

With the ‘Common Modulated Anchor Sources’ tracking method abandoned, I decided to break the user tracking problem into two parts: location tracking and orientation tracking. I would fully develop one of these aspects before considering the other so that a proof-of-concept prototype could be made that had some tracking capability.

The availability of both location and orientation tracking systems for indoor use is quite underwhelming at this time. Commercially available systems do not often have the required accuracy, as they are designed to locate a person in a building - rather than a person's position within a room (eg. Navizon, 2011). This is a rapidly developing field, however, particularly in the smartphone market. In March of 2013, Apple purchased a small indoor location tracking company known as Wi-FiSLAM (Panzarino, 2013). Wi-FiSLAM's location tracking used a combination of Wi-Fi signal mapping and smartphone sensors to generate accurate results (Huang, 2012). With large corporations like Apple interested in this technology, more accurate indoor location tracking may soon be widely available with mobile phones.

An alternative to dedicated location tracking systems is the Xbox Kinect controller. While its intended purpose was for tracking a user's gestures that control a game, the Kinect has also been used for musical purposes including location tracking for music (eg. Heap, 2012). A free program known as 'Synapse' is available which can convert Xbox Kinect gestures into mappable commands that control parameters in Ableton Live. My initial experimentation with this software suggested that the torso mapping function in Synapse would satisfy the accuracy requirements for a MARA system.

As Synapse does not have the ability to control ProTools parameters, a method for transferring the data mapped to Ableton Live controls to ProTools parameters needed to be devised. In this case, the X and Y depth control pots in ProTools needed to be synchronized with two dials in Ableton Live.

ProTools supports several control surface protocols that enable the user to move parameters in the software from physical control surfaces. There is no 'map', or 'MIDI learn' function available, so communication messages need to adhere to one of the supported specifications. The most accessible of these protocols is Mackie's 'HUI' (Human User Interface), which is a MIDI language. As HUI is a proprietary protocol, no official documentation exists to explain its inner workings. All available data on HUI is a result of reverse-engineering of the specification. This information is incomplete at times, meaning parts of the

language needed to be manually decoded for this project.

The HUI protocol uses a ‘digital handshake’ that informs the connected device and the DAW that there is a connection. The device sends a ‘ping’ message of 9 0 0 127 (decimal), or ‘Note on’, C-1 (although the octave is not a standardized reading). The DAW replies with a ‘Note off’ message of the same note. This ‘ping’ procedure occurs just under once per second. This was discovered by setting up a HUI peripheral in ProTools, and using the program ‘MIDI Monitor’ to ‘sniff’ outgoing MIDI messages from ProTools. As the ping message from ProTools was a ‘Note-off’, I concluded that a ‘Note-on’ would likely be the required response, which was the case.

Instead of sending these messages over a physical MIDI cable, the Apple IAC (Inter-Application Communication) driver can be used to send MIDI messages between programs. Using the IAC driver as a ‘virtual MIDI cable’, a patch was made in free programming platform ‘Impromptu’ that repeatedly sent a ping message to ProTools (See Appendix B). After forgetting to enable this program when using the HUI, I discovered that ping messages are not actually necessary for successful HUI communication. While a dialogue appears in ProTools saying that communication cannot be made with the HUI peripheral, the ‘Don’t show this again’ box can be checked, knowing that communication will still work so long as the IAC bus has been correctly set up.

The goal now was to make Ableton Live emulate a HUI control surface by sending appropriate MIDI messages to ProTools. A reverse-engineered version of the HUI specification has been compiled by ‘Theageman’ (2010), who ‘sniffed’ the MIDI messages that were sent and received from a Mackie control surface. This document shows that pan messages are given by delta value Control Change (CC) messages on channels 64-71 (up to 8 pan pots can be controlled simultaneously). Delta values are used because the rotary encoders on Mackie control surfaces (known as V-Pots) and continually variable, so do not have fixed minimum or maximum values. On the contrary, parameters in Ableton Live are given by fixed values; hence, a method for converting delta values to fixed values at a 1:1 ratio needed to be devised. By using trial and error, I found that the minimum negative

delta value for pan automation was given by a message of ‘1’, and the positive equivalent was ‘65’. This meant that sending a CC message of 65 on channel 64 would move the first pan pot in the bank clockwise 2 percent.

A Max for Live patch was created that sent a ‘bang’ every time a Live Dial emitted a new reading, which successfully moved the required pan knob in ProTools - although not at a 1:1 ratio, meaning the controls soon became out-of-sync. This problem occurred due to errors between the number of messages outputted from the live dial and the number of messages received by ProTools. To address this, the concept of cybernetics was incorporated in the Max for Live patch. Ashby states that cybernetics involves, “co-ordination, regulation, and control” of a process of change in a closed loop system (1957, p. 1). In this case, the Max Patch first determines if the delta change of the input control is positive or negative. An increase sends a ‘bang’ message to the positive delta control, and a vise-versa for a decrease. At the same time, the patch is constantly checking if the number of outputted bangs is consistent with the input value. If the input and outputs become misaligned, the patch adjusts to correct for the error. The dial in Ableton could now be moved at a 1:1 ratio to the pan dial in ProTools.

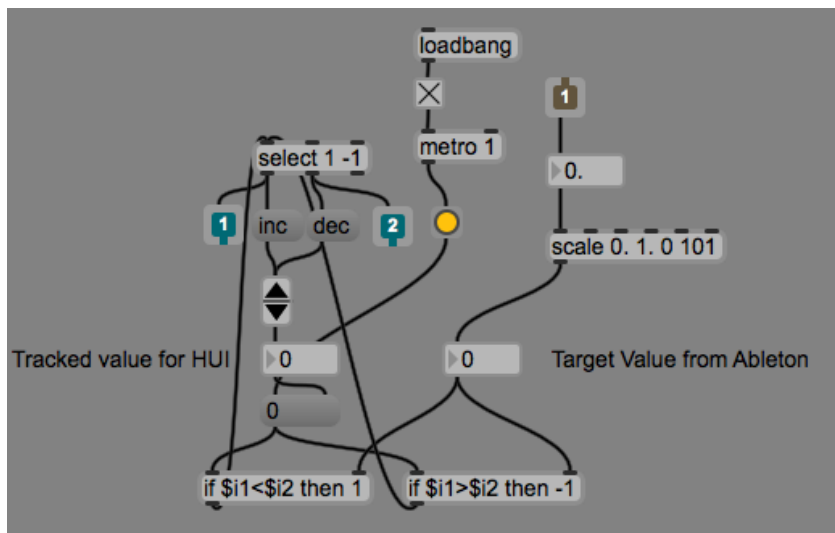


Figure 4.13: The error-correcting cybernetics patch

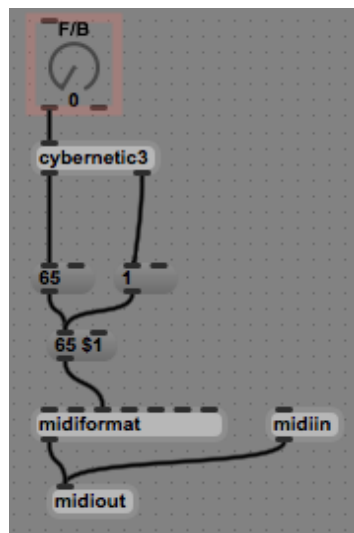


Figure 4.14: A patch controlling the reading of the second pan knob in ProTools

#### 4.4.4 Developing and Implementing an Orientation Tracking System

While the Xbox Kinect has been used for orientation tracking through facial recognition (eg. Mauro, 2012), it would not be suitable in this instance as facial recognition would need to occur in multiple moving locations in the room, and over a 360 degree plane which is not feasible with a single Kinect.

My first point of call for exploring alternate orientation tracking methods was the iPhone's inbuilt compass, as it was to hand. The iPhone was fixed to the top of a hat that the user wears, so the compass reading is consistent with the orientation of the user's ears. Research uncovered an App named c74, which among other features, allows the compass data from an iPhone to be sent to Max for Live via a WiFi connection. Mapping this dial in Ableton is quite simple (see Figure 4.15), and this dial can be further mapped to other parameters as necessary. The accuracy of the iPhone compass appeared reasonable, although faster responsiveness would be preferred. As the iPhone was already in use for wireless playback, this method offered practicality through capitalising upon resources already in the system.

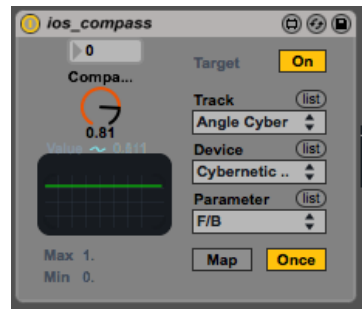


Figure 4.15: The interface for the compass mapper patch

A method for rotating the entire binaural mix in ProTools in response to the compass dial's output in Ableton Live was now required. My initial theory on how to achieve this involved placing a HRTF panner over the entire binaural mix, which would convolve the mix with the location properties given by the HRTF panner, thereby controlling the rotation of the mix. The 'Panorama' plug-in from WaveArts was used to trial this theory. 'Panorama' is a binaural panning engine that accepts stereo inputs (required for a binaural input). The rotation of the soundstage can be controlled by a single 'azimuth' dial.

In practice, the binaural mix was non-uniform in volume and spatial qualities when rotated with this method. This is likely because the input binaural signal already contained HRTF information, so when 'Panorama' convolved this with a second HRTF, abnormal comb-filtering occurred, and vector-based level issues arose. To confirm this, the binaural panner contained in Logic Pro was tested, yielding similar results. The theory behind this method was to move the listener's head in relation to the room and sound sources in a streamlined way (Figure 4.16).

After discussing the flaws of my first method with surround sound expert Matthew Hitchcock, it was recommended that I investigate a plug-in named 'Anymix Pro' by IOSONO. This plug-in acts as a 'surround-remapper', where multiple inputs are remapped to multiple outputs via several GUI techniques. There is a function that can spin each input synchronously to new outputs - a surround sound mix can be rotated as if it was moving around the listener. It occurred to me that placing this plug-in in series before the convolution section of the signal chain



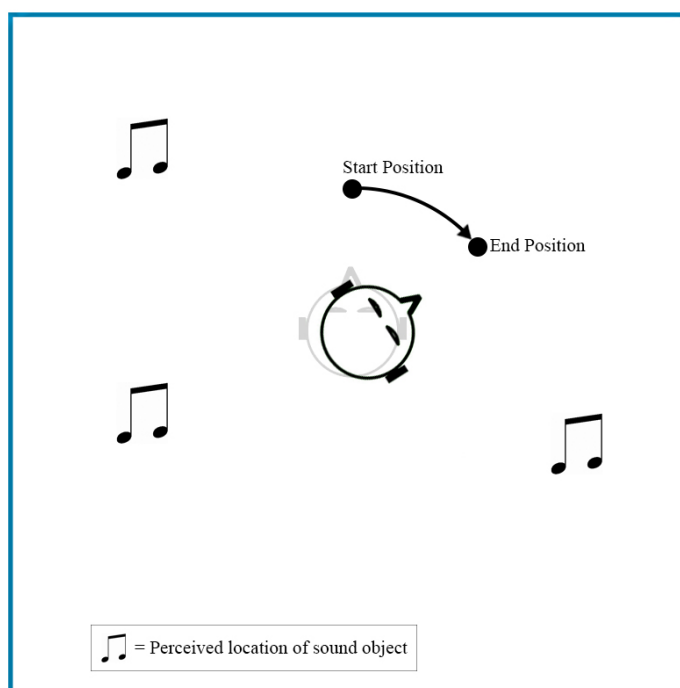


Figure 4.16: The theoretical result of convolving a binaural mix with a HRTF panner. In practice, the soundstages become non-uniform and indistinct.

would enable the entire binaural mix to be spun by a single dial.

The benefit of this method was the strong binaural image that was achieved through the convolution-panner method would be protected. A downside is that directional acoustic properties of the room also rotate with the user. This means that when the user turns their head, the spatial properties of the pseudoacoustic environment will also move, while the sounds in the room remain fixed. Figure 4.17 illustrates this phenomenon.

The next step was to find a method for controlling the angle dial in Anymix Pro from a dial in Ableton Live. The HUI protocol is capable of plug-in automation, so this method was again explored. The necessary messages, however, are not documented in the reverse-engineered specification by Theageman. Through trial and error, I discovered that plug-ins are manipulated using CC delta values from channels 72, 73, 74, and 75 (only four parameters can be controlled simultaneously under the HUI protocol). Of note, a CC message of 15 28 followed by 47 64 may need to be sent to engage the selected plug-in for automation. To scroll across banks of four selected parameters in a plug-in, CC delta value messages are sent on channel 76 (also not documented previously).

After overcoming the challenge of automating plug-ins by emulating a HUI protocol, it was discovered that Anymix Pro does not support control surface automation, so a new method needed to be adopted. Any Anymix parameter can be mapped in Ableton Live, however Ableton does not currently support surround sound routing. While this plug-in could not be used here, it did inspire an alternate solution. With the flexibility offered with Max for Live, I investigated the idea of designing my own ‘surround-remapper’ matrix. To integrate this into a ProTools system, a virtual patch cable program known as ‘JackOSX’ was used.

A default Max for Live object known as ‘API SendsPan’ was used as the basis for my surround-remapper matrix. This object allows inputs to be routed to various sends by way of a single revolving dial. Twelve input tracks were created (one for each send in Ableton), with a SendsPan object inserted on each. Ideally, the surround-remapper would support 14 channels, but this was not possible with

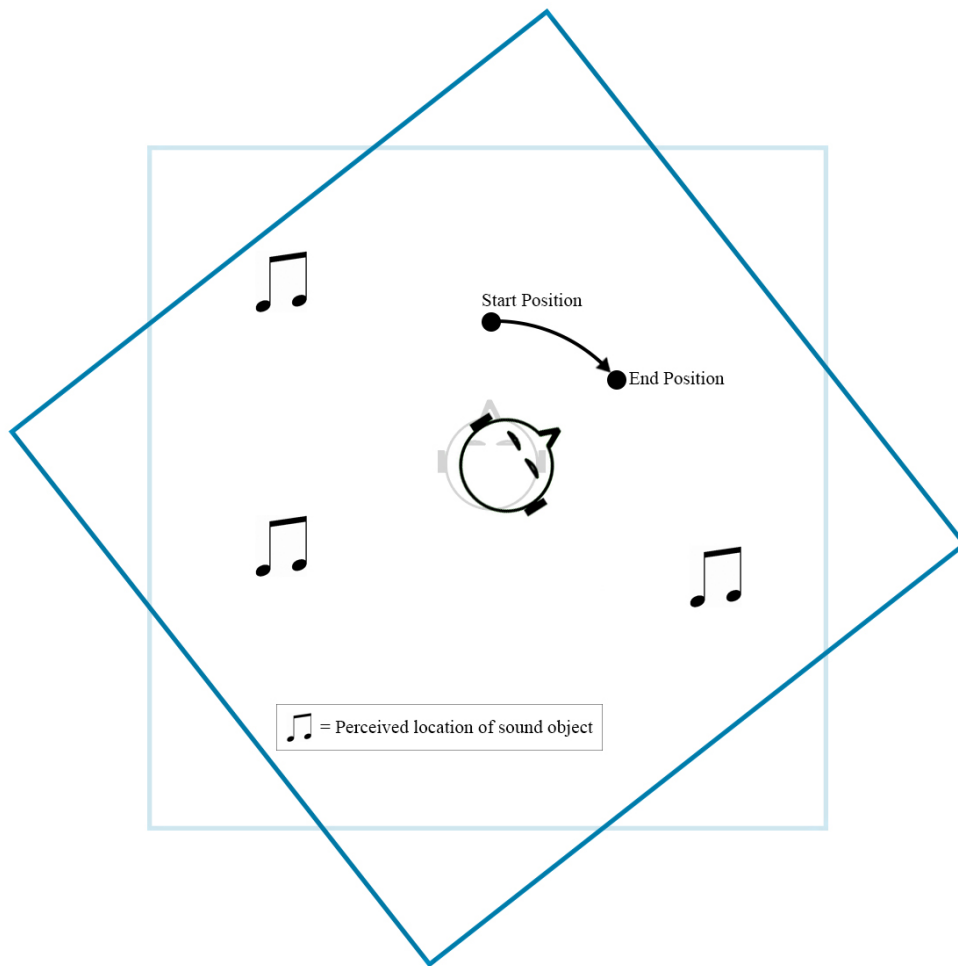


Figure 4.17: The user turning their head with the surround-remapper orientation method. In practice, artefacts introduced by the room's rotation are indistinguishable, and a strong binaural image is retained.

Ableton’s send restrictions. To accommodate this, the rear surround left and rear surround right channels from the ‘Far’ layer were not mapped. These channels were chosen because they are likely to be furthest from the listener’s consciousness - when a sound is behind the listener and in the distance, the listener will likely be focusing on other sounds.

Each SendsPan dial was set to receive data sent from the compass dial, so each displayed the same reading. The routing options enabled the sends to be offset so that when one dial is outputting to Send A, the next dial outputs to Send B etc (see Figure 4.19). Each send in Ableton was then routed through JackOSX and into the inputs of the convolution section in ProTools. With this configuration, when the SendsPan dial is at position 1, each input is being routed unchanged through Ableton. As the SendsPan dial is moved, the inputs to the convolution engines are remapped, thereby rotating the binaural mix.



Figure 4.18: The API SendsPan object in Max for Live

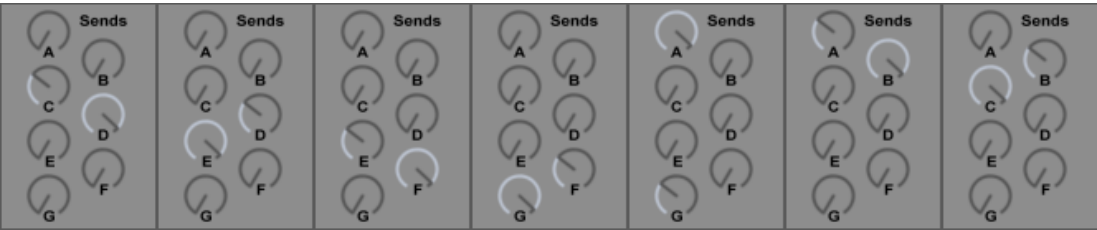


Figure 4.19: The surround-remapper matrix - send levels from a given SendsPan position

An additional object was added before the SendsPan object that inverts the reading given from the compass patch. This way, the mix rotates in the opposite direction to the user, making sounds appear to remain fixed. A patch has also been created that can offset the compass reading by a given angle. This allows the user to calibrate the direction of the room relative to North.

#### 4.4.5 Critical Evaluation

The user tracking system presented here performs its task admirably, particularly considering it was constructed entirely from readily available components. The Kinect location tracking system wirelessly tracks the listener with sufficient resolution, and is easily configurable to suit different spaces.

The orientation tracking system also achieves its goal successfully and its use of the iPhone, an existing component of the system, makes it a particularly practical addition. A weakness of the orientation tracking system is some latency between the listener's movement and the system's response. While every effort has been taken to maximise the efficiency of the compass, calibration, inversion, and surround-remapper patches, the processing required is computationally strenuous and causes some lag. As a result of this, the MAA is not accurately reflected in this system, however, the orientation tracking certainly improves localisation accuracy, and prevents front-to-back confusion (as discussed in Section 2.2). Future versions of this orientation tracking system could make use of a more accurate compass, and use a simpler method implementing compass data into the auralization engine.

## Chapter 5

# Making Music with MARA

A practical demonstration of the MARA system has been developed in response to my observations of the system’s capabilities throughout the design process. Program notes that were provided with this ‘performance’ are attached in Appendix D. The main goal of this demonstration was to showcase the basic functionality of the system. It is expected that the system offers significant creative potential beyond the scope of this example. Comparison can be drawn here with Thomas Edison’s presentation of the phonograph. While Edison recorded ‘Mary Had a Little Lamb’ on his device to illustrate its function (Weber, 2013), others explored the phonograph’s potential much more fully.

Through trialling and evaluating the system, I have gained insights into how MARA might be best used for music creation. I observed that manipulating technical variables in the system had potential to result in inspiring musical results. Further, through testing a diverse range of sounds in the system, concepts about optimising musical elements for the system surfaced. This section will discuss these findings with reference to the ‘Remote Control’ demonstration.

## 5.1 Creative Operation of the MARA System

Facilitating a realistic and immersive sound experience has been an important goal in this project. At several points throughout the design process, the system was set up so that it presented a sound environment slightly peripheral to this goal, but inspired me creatively. This was pleasing to see, as it had been flagged as a potential outcome of the Design Research process. ‘Remote Control’ incorporates several instances of creative configurations of the system. These result in a musical experience that may extend beyond what is possible in a live performance. This is known as *hyper-reality* (Bonanni, 2006, p. 130) and is particularly relevant in augmented reality contexts as the infrastructure for ‘amplifying’ sensory input is already at hand.

In the case of this MARA system, an example of hyper-reality can be achieved by exaggerating the listener’s perception of distance. By adjusting the volume of the ‘Far’ auxiliary relative to ‘Near’, small changes in the listener’s position can create dramatic effects on the perceived proximity of sound sources. For example, moving one meter away from a sound source could make it appear to extend far into the distance, or disappear completely. The absence of visual stimulus makes this possible as there is no conflicting location information arriving through other senses.

The application of this function can be decided according to the producer or user’s taste. When experiencing the system in this configuration, some users have commented that the exaggeration of distance makes it easier to determine how their position effects the sound space. Others have voiced concern that the hyper-realistic environment may impinge on immersion in the performance. The decision to use this feature can be likened to a mix engineer’s decision to use effects - it is done in accordance with the overall production aesthetic. In the case of ‘Remote Control’, I have decided to use this function, although quite subtly. This helps to showcase how the system is responding to the user’s location while maintaining the illusion that a choir of voices is surrounding the room.

Multiple soundstage configurations are another example of creative technical use

of this system. Soundstages can be adjusted to represent the physical dimensions of the space, or represent a space that extends beyond the walls of the room. Further, the area in which the user can move can be adjusted to make for a more or less physically-involved experience. ‘Remote Control’ is designed to maximize the user’s involvement in the performance - the user can explore the extremities of the physical space as if approaching the singers who are positioned along the walls of the room. Each soundstage is two-dimensional, as the singers are placed equidistant from the walls. Inter-instrument depth is an area for future exploration with this system.

The availability of multiple soundstages affords the ability to have multiple focal points throughout the performance. At different times in the demonstration, the lead part occurs in different locations in the room, and the sense of balance is obscured. This encourages the listener to respond by changing their position and orientation.

While it is usually desirable for sounds to embody a fixed position in the space, this system allows the producer to combine these with sounds that ‘follow’ the user. In ‘Remote Control’, some sounds are positioned close to the listener’s left or right ears despite their location or orientation. Sounds can also be made to follow the user, but be independent of their orientation. For instance, an insect could be buzzing around the user’s head and following them wherever they go, but remain in the same place in the sky when head turning occurred. In addition, intracranial sounds can be used to juxtapose the augmented sound space. This gives the producer the ability to place sounds inside the listener’s head, fixed within the space, or moving relative to the listener - a broad creative palette that is unobtainable with other technologies.



## 5.2 Musically Creative Ideas with the MARA System

Design and evaluation iterations of this system have also highlighted several musical concepts that complement MARA well. To illustrate, these concepts have been applied to the ‘Remote Control’ demonstration.

One such finding refers to the horizontal (time-based) versus the vertical (spectral) construction of the arrangement. While a stereo mix may include sounds that play for a short time and then disappear, these may be distracting to a user who is exploring a MARA mix. Sounds that provide continuous location cues are helpful, as the user always has feedback as to where the sound is in the room. If multiple sound sources are emitting streams of location information, the user may decide which to explore at any time. As a result, separation between parts is better achieved spectrally, or through uniqueness in rhythm patterns.

Each vocal part in ‘Remote Control’ has been allocated a unique rhythmic pattern and spectral (pitch and timbre) range. This helps each element to be easily identified at any time. To provide a constant stream of location feedback to the user, small samples of each part have been looped continuously. This use of loops creates an interesting construct of time: the musical elements remain static while the user’s movements are responsible for musical changes through perspective. In most musics, time is a representation of changes in the musical construction of the piece. In this case however, time is replaced as an agent of change by the location of the user. The empowerment given to the user here will be discussed further in Chapter 6. While much of ‘Remote Control’ is loop-based, it also exhibits a morphing of musical and production ideas. This results in a musical experience where the listener is provided with adequate location information to feel comfortable with their augmented reality environment while receiving enough new information to maintain interest in the piece.

In addition to time-based elements, frequency-based findings have emerged that can be used to effectively produce music for the MARA system. As noted in Section 2.3, the human auditory system uses different methods to localise sound

for different pitch ranges. For some frequencies, neither of these methods is particularly effective, meaning the direction of these sounds is very difficult to decipher. Low bass frequencies are omnidirectional to the human ear; hence, they will not translate a strong sense of location in a MARA system. Very high frequencies are also difficult for the human ear to localise. In fact, our ears are most effective at hearing and localising the frequencies present in human speech, which occur well within the boundaries of our hearing range. For this reason, I have chosen to use sounds created by my mouth to showcase the system. This is not to say that the system is limited to human voice sounds, rather, this was a creative decision made cognizant of the system's strengths.

When testing the system, I noticed that I instinctively used the lead vocal part as a point of reference in the mix. Not only do our ears localise this content effectively, but it is instinctual to turn towards a person communicating with us (or in this case, singing lyrics). Knowing this phenomenon, the producer of a MARA experience can attract the user to certain parts of the room at will. In 'Remote Control', the lead vocal enters after some time, giving the listener the opportunity to randomly explore their musical surroundings before their attention is absorbed more fully by the words of the lead vocal.

Section 2.3 discussed how our ears calculate the ratio of direct sound to early reflections to determine the distance of a sound. Transient sounds allow our ears to identify these components most effectively - if a sound has a long ADSR envelope, it is difficult to distinguish between these components. With this in mind, the parts in 'Remote Control' were designed to be quite percussive, regardless of their role in melody, harmony or rhythm. Similarly to the continuous looping of short motifs, this provides the listener with a constant stream of location and distance information to reinforce their immersion in the sound environment.

## Chapter 6

# Potential Impact on Music Production and Consumption Paradigms

*The gramophone changed the way we listen to music; the iPod changed the way we consume it - MARA systems could have similar ramifications on our relationship with music.*

### 6.1 Looking Forward: Advancing MARA Systems for Music

With consumer electronics technology improving rapidly, I envision that a MARA system could soon be operated entirely from a smartphone. BRIRs could be taken using inbuilt microphones in the supplied earphones (as shown by Gamper & Lokki, 2009) and implemented using the device's on-board software. With improved indoor position tracking technology, and using the smartphone compass, the user tracking system would require no further hardware. Processing power in smartphones is inevitably increasing, and it is not difficult to imagine

all auralization processing occurring in a phone. The auralization engine could be programmable by touch screen, or even gesture recognition. For example, aiming the phone in a particular direction could appoint a sound to come from that part of the room. For the consumer, programming would occur ‘behind the scenes’, so that anyone could press play and enjoy the experience as they would a stereo recording.

If a system like this became ubiquitous, it would undoubtedly affect the way we produce and listen to music. An overarching theme in MARA is the power granted to the consumer through interaction. Given the ability to navigate the spatial character of a piece, the recording becomes a fluid entity that is different on each listening. The act of listening to a song would evolve from being passive to active. The listener would go from being subservient to the music presented to them, to being part of it’s creation. This concept is highly topical in the present digital age, where technology is increasingly ‘democratizing’ our interactions with media and art.

This change in the producer/consumer paradigm, for example, could be pertinent to the expanding field of ‘online jamming’. Online jamming makes use of the internet to connect multiple participants in separate places so that they can create music together. Online music creation platforms such as ‘Ohm Studio’ could enjoy the benefits of MARA, where each member of the virtual band could position the other members around their room to mimic jamming in a rehearsal room. Scenarios like this would not only benefit professional producers of music - anybody could enjoy the experience as an observer or contributor to the performance.

MARA also offers significant potential in game development. Using the player as a controller for the game is a strong area of interest presently, and MARA does this and offers layer of immersion in addition. While games are already in development for MARA (eg. Moustakas et al., 2012), there remains significant opportunity to implement music as a central part of new MARA games. This technology could also be adopted in existing game architecture. For instance, Activision’s ‘Guitar Hero’ could incorporate MARA into its multiplayer mode, so

that the sound from fellow band member’s instruments project from their location in the room.

Other applications for MARA-type technology could be found in education, health, and defence, however musical avenues are of particular interest in this paper.

## 6.2 Future Work

The design presented in this paper acts as a proof-of-concept for a MARA system for music. This work can provide a model for the development of more sophisticated system. The next major stage of this project would involve collaborating with others to further develop and refine the findings presented here. While this project has successfully demonstrated that MARA is applicable for musical purposes, the current prototype is complex and therefore not accessible for others to use. Designing a commercially available product would allow others to explore the potential of this technology.

A commercially produced equivalent could improve on several functionality aspects of my system. Areas that present significant opportunity for further research include:

- Adding the dimension of height to the system. This would require roll and elevation tracking of the user’s head.
- The ability to walk ‘past’ a sound object. The auralization engine used in this project required several compromises in order to work with computer-based DAW software. Departing from the reliance on in-built DAW features would make way for improvements in the auralization engine.
- Improving tracking resolution and lowering latency. With smartphone sensor technology improving, accurate indoor tracking systems may soon be ubiquitous. By implementing tracking data in the system more effectively, delay between a user’s movement and the system’s response could be reduced.

## **Potential Impact on Music Production & Consumption Paradigms 6.2**

---

- Streamlining the software components into a single program or application. This would reduce CPU load and make the system more applicable for smartphone use.
- Adding directionality to sound sources. Instruments project sound directionally in an acoustic space. By mapping the three-dimensional sound projection of instruments, it would be possible to replicate these properties with MARA.
- Adding multi-user functionality so that users could enjoy the experience as a group.

# Chapter 7

## Concluding Remarks

This paper has presented a working prototype of a Mobile Augmented Reality Audio System for Music. Using existing research as a basis for design exploration, each component of a MARA system has been purpose-built for music creation. This has led to several innovations. A MARA headset has been developed that can wirelessly receive audio from a host computer. The rationale behind the design process of this component has been documented which is likely to assist in the design of other projects.

A binaural convolution interpolation system called the ‘convolution-panner’ has been developed that allows for real-time manipulation of the horizontal location properties of binaural sound sources using the user-friendly interface of the ProTools surround panner. To my knowledge, this is the first MARA system to be incorporated into a Digital Audio Workstation. Having the system housed within ProTools means that an arsenal of musical tools are available for the producer to use. Existing MARA systems that are not designed for music do not offer such creative control.

A method for adjusting to the user’s position in the room has been developed by automating the distance of each part of the mix relative to the listener. Through innovative application of DAW features, this was achieved using only an X and Y position input. To complement this, the Xbox Kinect was used to track the user’s location. It is likely that this is the first time a Kinect has been implemented in

a MARA system.

The inbuilt iPhone compass has been used to track the listener's orientation. A method for counter-rotating the binaural mix when the listener turns called the 'surround-remapper' has been developed. This has been integrated into the convolution-panner system in such a way as to retain the strong three-dimensional sound field obtained using the convolution method.

Throughout the design/evaluation cycle, a number of technically and musically creative uses for the system have emerged. These have been demonstrated through 'Remote Control', a piece designed especially for use with the MARA system. This demonstration cements MARA as a viable medium for creative music production and consumption, and showcases how each component of this project works in harmony with the design goal.

MARA technology promises to be an exciting addition to music creation and enjoyment. This project has revealed that MARA has great potential to be used for musical purposes; however, little exploration into this has previously occurred. As such, this is a compelling topic that contains considerable opportunity for new discoveries.



# Appendices

## A Demonstration of Convolution-Panner and the Blurring Effect

The following descriptions relate to the three audio tracks provided with the Appendix CD. This CD should be listened to through headphones as it contains binaural audio.

- Track 1 is the reference sample. Here, the audio file was convolved with a single IR that corresponds to a position directly in front of the listener; hence, there is no blurring between multiple IRs.
- Track 2 was convolved with two IRs: one that was captured from the front left, and one from the front right of the room. The centre IR channel (used for Track 1) was muted. This demonstrates the maximum possible level of blurring, as it is exactly in between the two IR positions.
- Track 3 is my improved method of implementing the IRs. Here, I inverted the channels of the front left IR, and replaced the old ‘front right’ channel with the inverted version. This meant that the front right and front left IRs now contain identical information. This method assumes that the room is symmetrical and the listener has symmetrical hearing, although the latter part this assumption is already in operation in this system as a result of

using non-individualised HRTFs. Again, the centre IR channel has been muted, so this shows the maximum possible amount of blurring.

The differences between each of these examples are very subtle. The fact that the worst-case scenario example (Track 2) is so similar to the reference (Track 1) proves that the convolution-panner is a viable method for BRIR interpolation in this context. This demonstration shows that the theoretical flaw of localisation blurring is insignificant in practice.

## B ‘Ping’ Response Patch

```
(io:midi-register-events)
(io:print-midi-destinations)
-----
MIDI Destinations
0: name=IAC Driver Bus 1
1: name=FireWire 410
-----
(define iac (io:midi-destination 0))

(define pinger-on #f)

(define pinger
  (lambda (time delta)
    (io:midi-out time iac 9 0 00 127)
    (if pinger-on
      (callback (+ time delta) pinger (+ time delta) delta))))

(define start-pinger
  (lambda (time delta)
    (set! pinger-on #t)
    (pinger time delta)))

(define stop-pinger
  (lambda ()
    (set! pinger-on #f)))

;; Start ping
(start-pinger (now) *second*)

;; Stop ping
(stop-pinger)
```

Figure A: A program to send ‘ping’ messages to ProTools over the IAC Bus

## C List of Required Equipment

### C.1 Software:

- ProTools 10 HD (required for surround functionality and pan grouping).
- Ableton Live 8 with Max for Live, Max version 5 (c74 App does not currently support 64 bit architecture).
- Synapse with accompanying ‘Live Dial’ Max for Live plug-in.
- JackOSX.
- Airphones App and desktop program.
- c74 App and accompanying Max external.
- Altiverb 7.
- Mac OSX 10.7.
- IAC MIDI driver for HUI communication.
- MIDI Monitor program (optional)
- Impromptu program (optional)

### C.2 Hardware

- 1 x Xbox Kinect controller with power supply.
- 1 x Philips SHN2500 noise-cancelling earphones.
- 2 x Kemo M040N universal preamplifiers.
- 1 x 9V battery.
- 2 x 10 ohm logarithmic double-ganged potentiometers.

- 4 x female RCA sockets.
- 1 x power switch.
- 1 x 3.5mm to stereo male RCA cable.
- 1 x iPhone (tested on 4s).
- 1 x circuit enclosure.
- 1 x belt clip.
- 1 x rigid hat.
- 1 x iPhone case (must not have magnets inside)

## D Program Notes

When you listen to music in earphones, the sound usually seems like it is coming from somewhere inside your head. By manipulating sound based on the way we hear normally, it is possible to make music in earphones sound like it is coming from somewhere in the room. This is called binaural sound.

Until now, binaural sound has suffered a loss of realism whenever the listener turned their head or moved around the space - all of the instruments moved too!

With my Mobile Augmented Reality Audio System for Music - the first of its kind - you can experience a three-dimensional musical environment through earphones that adjusts so that when you move, the sounds stay right where they were before.

Now you can explore. Want to hear more of the lead singer? Simply walk closer to them and you can stand right in front of them. You can investigate each corner of the room, and gain a different perspective of the music everywhere you go. Now you're in control of your own listening experience more than ever.

‘Remote Control’ was written to showcase how the system works. In this piece, you will be immersed inside a choir of voices, with each member standing somewhere along the walls of the room. At times, you may want to simply stand and absorb the sounds around you; as you become accustomed to your augmented reality environment, you can explore the composition by walking around, turning your head, or dancing. The use of loops in this piece creates a fascinating dimension of time - the musical elements remain static while your movements change your perspective of the music. As the piece develops, you’ll find there are focal points in different places at different times. Sometimes, the voices even follow you, whispering in your ears wherever you go.

*Please Note:* The system may encounter glitches that interrupt the listening experience. This system is highly complex and requires a great deal of processing power from both the computer and iPhone. While every precaution has been taken to minimize any crashes, these may be inevitable and beyond our control. I appreciate your understanding that this is a prototype system.

## E Software Setup Instructions:

- Turn on computer.
- Plug Kinect controller in.
- Open Synapse.
- Set up an IAC Bus in Audio MIDI Setup.
- Open JackPilot
  - Set Airphones as the input and output device.
  - Set the audio buffer to 1024 samples.
  - Set 12 virtual inputs and 16 virtual outputs.
- ‘Start’ Jack

- Open ProTools
  - Open demonstration session.
  - Set playback engine to Jackrouter.
  - Set up HUI peripheral to IAC Bus.
  - Load IO settings.
- Open Ableton Live.
  - Set audio engine to Jackrouter.
  - Configure inputs and outputs (12 mono in, 12 mono out).
  - Enable IAC Bus in MIDI preferences
  - Open demonstration session.
- Open Max via a Max for Live Patch.
  - Go to file preferences and set a user defined folder to the project folder.
- Load Jack settings.
- Set up WiFi router.
- Connect computer to network.
- Connect iPhone to network.
- Open Airphones program on computer.
- Open Airphones App on iPhone.
- Open ‘Compass Offset Direct’ Patch in Ableton.
- Open c74 App on iPhone.
- Reopen Airphones and reconnect on iPhone.
- Reopen c74 App.

- Press play in ProTools.

# References

- Albrecht, R. (2011). *Messaging in mobile augmented reality audio*. Doctoral dissertation, Aalto University. Retrieved from <https://aaltodoc.aalto.fi/handle/123456789/3667>
- Albrecht, R., Lokki, T., & Savioja, L. (2011). A mobile augmented reality audio system with binaural microphones. In *Proceedings of interacting with sound workshop on exploring context-aware, local and social audio applications* (pp. 7–11). Espoo: ACM Press. doi: 10.1145/2019335.2019337
- Alzagi, R., Duda, R., Thompson, D. (2004). Motion-tracked binaural sound. In *Proceedings of the 116th aes convention* (pp. 1–14). Berlin: Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=12644>
- Andean, J. (2011). Ecological psychology and the electroacoustic concert context. *Organised Sound*, 16(02), 125–133. doi: 10.1017/S1355771811000070
- Baalman, M. A. (2010). Spatial composition techniques and sound spatialisation technologies. *Organised Sound*, 15(03), 209–218. doi: 10.1017/S1355771810000245



## References

---

- Bonanni, L. (2006). Living with hyper-reality. *Ambient Intelligence in Everyday Life*, 3864, 130–141. Retrieved from [http://link.springer.com/chapter/10.1007/11825890\\_6](http://link.springer.com/chapter/10.1007/11825890_6) doi: 10.1007/11825890\\_6
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, 13(1), 15–42. doi: 10.1207/s15327809jls1301\\_2
- Downton, P. (2003). *Design research*. Melbourne: RMIT Publishing. Retrieved from <http://books.google.com.au/books?id=IVBQAAAAMAAJ>
- Gamper, H. (2010). *Audio augmented reality in telecommunication*. Diploma thesis, Graz University of Technology. Retrieved from [http://www.tml.tkk.fi/~hannes/admin/download.php?download=Audio\\_augmented\\_reality\\_in\\_telecommunication\\_Hannes\\_Gamper\\_2010.pdf](http://www.tml.tkk.fi/~hannes/admin/download.php?download=Audio_augmented_reality_in_telecommunication_Hannes_Gamper_2010.pdf)
- Gamper, H., & Lokki, T. (2009). Instant BRIR acquisition for auditory events in audio augmented reality using finger snaps. In *International workshop on the principles and applications of spatial hearing* (pp. 1–4). Miyagi. Retrieved from [https://mediatech.aalto.fi/~ktlokki/Publs/gamper\\_iwpash.pdf](https://mediatech.aalto.fi/~ktlokki/Publs/gamper_iwpash.pdf)
- Gibson, J. J. (1979). *The ecological approach to visual perception*. London: Lawrence Erlbaum Associates.
- Gifford, T. (2011). *Improvisation in interactive music systems*. Doctoral dissertation, Queensland University of Technology. Retrieved from <http://eprints.qut.edu.au/49795/>

## References

---

- Göttlinger, C., Graz, T., Sontacchi, A., Kunstuniversitat, G., Musialik, C., & Gmbh, A. (2009). Psychoacoustical background of dynamics processing in audio. In *3rd international vdt symposium* (pp. 1–7). Hohenkammer. Retrieved from [http://www.tonmeister.de/symposium/2009/np\\_pdf/C02.pdf](http://www.tonmeister.de/symposium/2009/np_pdf/C02.pdf)
- Grandy, R. (1987). Information-based epistemology, ecological epistemology and epistemology naturalized. *Synthese*, 70(1987), 191–203. Retrieved from <http://www.springerlink.com/index/x6p11116u2482438.pdf>
- Gray, C. (1996). Inquiry through practice: Developing appropriate research strategies. In *Proceedings of the international conference on art and design research*. Helsinki. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Inquiry+through+practice+:+developing+appropriate+research+strategies#0>
- Hamilton, J. G., & Jaaniste, L. O. (2009). Content, structure and orientations of the practice-led exegesis. *Art.Media.Design: Writing Intersections*, 1–10. Retrieved from <http://eprints.qut.edu.au/29703/1/c29703.pdf>
- Härmä, A., Jakka, J., Tikander, M., & Karjalainen, M. (2003). Techniques and applications of wearable augmented reality audio. In *Proceedings of the 114th aes convention* (pp. 1–20). Amsterdam: Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=12495>
- Härmä, A., Jakka, J., Tikander, M., & Karjalainen, M. (2004). Augmented reality audio for mobile and wearable appliances. *Journal of the Audio Engineering Society*, 52(6), 618–639. Retrieved from <http://www.aes.org/>

## References

---

- e-lib/browse.cfm?elib=13010
- Heap, I. (2012). *Imogen heap performance with musical gloves demo: Full wired talk*. Retrieved from <http://www.youtube.com/watch?v=6btFObRRD9k>
- Hiipakka, M. (2012). *Estimating pressure at the eardrum for binaural reproduction*. Doctoral dissertation, Aalto University. Retrieved from <http://urn.fi/URN:ISBN:978-952-60-4865-9><https://aaltodoc.aalto.fi/handle/123456789/6315>
- Hitchcock, M. (2012). Morphing aural spaces in 7.1 music. In *Acmc 2012 conference proceedings* (pp. 17–21). Brisbane.
- Hiyama, K., Komiyama, S., & Hamasaki, K. (2002). The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field. In *Proceedings of the 113th aes convention* (pp. 1–12). Los Angeles: Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=11272>
- Huang, J. (2012). *WiFiSLAM: Modern innovations in indoor positioning*. Retrieved from <http://www.youtube.com/watch?v=OGdvjv1a1Tc&feature=youtu.be#t=1033s>
- Huopaniemi, J., Savioja, L., & Karjalainen, M. (1997). Modeling of reflections and air absorption in acoustical spaces a digital filter design approach. In *Proceedings of 1997 workshop on applications of signal processing to audio and acoustics* (pp. 1–4). Helsinki: IEEE. doi: 10.1109/ASPAA.1997.625594

## References

---

- Karjalainen, M., Tikander, M., & Härmä, A. (2004). Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources. In *Acoustics, speech, and signal processing* (pp. 101–104). Espoo: IEEE. Retrieved from <http://ieeexplore.ieee.org/search/srchabstract.jsp?arnumber=1326773&isnumber=29346&punumber=9248&k2dockey=1326773@ieeecnfs>
- Keyson, D., & Alonso, M. B. (2009). Empirical research through design. In *Proceedings of the international association of societies of design research conference* (pp. 4548–4557). Seoul. Retrieved from <http://www.iasdr2009.org/ap/Papers/SpecialSession/Assessingknowledgegeneratedbyresearchthroughdesign/EmpiricalResearchThroughDesign.pdf>
- Lemordant, J., & Lasorsa, Y. (2010). Augmented reality audio editing. In *Proceedings of the 128th aes convention* (pp. 1–9). London: Audio Engineering Society. Retrieved from <http://hal.inria.fr/hal-00494239/>
- Lokki, T., Nironen, H., Vesa, S., Savioja, L., Härmä, A., & Karjalainen, M. (2004). Application scenarios of wearable and mobile augmented reality audio. In *Proceedings of the 116th aes convention* (pp. 1–9). Berlin: Audio Engineering Society. Retrieved from [http://www.acoustics.hut.fi/~aqi/talks/aes116\\_poster.pdf](http://www.acoustics.hut.fi/~aqi/talks/aes116_poster.pdf)
- Luciani, A. (2009). Enaction and music: Anticipating a new alliance between dynamic instrumental arts. *Journal of New Music Research*, 38(3), 211–214. doi: 10.1080/09298210903325600

## References

---

- Lyons, K., Gandy, M., & Starner, T. (2000). Guided by voices : An audio augmented reality system. In *Proceedings of the international conference on auditory display* (pp. 57–62). Atlanta, USA: International Community for Auditory Display. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.22.2477&rep=rep1&type=pdf>
- Marett, A. (2005). *Songs, dreamings, and ghosts: The Wangga of North Australia*. Middletown: Wesleyan University Press. Retrieved from [http://books.google.com/books?id=oFhUE\\_AIE8QC&pgis=1](http://books.google.com/books?id=oFhUE_AIE8QC&pgis=1)
- Mauro, D. (2012). *On binaural spatialization and the use of GPGPU for audio processing*. Doctoral dissertation, University of Milan. Retrieved from <http://air.unimi.it/handle/2434/172440>
- Menzer, F., Faller, C., & Lissek, H. (2011). Obtaining binaural room impulse responses from b-format impulse responses using frequency-dependent coherence matching. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2), 396–405. doi: 10.1109/TASL.2010.2049410
- Mitchell, B. (2010). The immersive artistic experience and the exploitation of space. In *Ideas before their time: Connecting the past and present in computer art* (pp. 98–107). London: British Computer Society. Retrieved from [http://www.bcs.org/upload/pdf/ewic\\_ca10\\_s3paper2.pdf](http://www.bcs.org/upload/pdf/ewic_ca10_s3paper2.pdf)
- Mitchell, T. (2011). Soundgrasp: A gestural interface for the performance of live music. In *International conference on new interfaces of musical expression* (pp. 465–468). Oslo. Retrieved from <http://eprints.uwe.ac.uk/18266/>
- Moulton, D. (1990). The creation of musical sounds for playback through

## References

---

- loudspeakers. In *8th international conference: The sound of audio* (pp. 161–169). Boston: Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=5420>
- Moustakas, N., Floros, A., & Grigoriou, N. (2012). Interactive audio realities: An augmented/mixed reality audio game prototype. In *Proceedings of the 130th aes convention* (pp. 1–8). London: Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=15826>
- Müller, S., & Massarani, P. (2001). Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society*, 49(6), 443–471. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=10189>
- Navizon. (2011). *Indoor Triangulation System*. Retrieved from <http://www.navizon.com/its/whitepaper.pdf>
- Paine, G. (2007). Sonic immersion: Interactive engagement in real-time immersive environments. *SCAN Journal of Media Arts and Culture*, 1–13. Retrieved from [http://www.scan.net.au/scan/journal/display.php?journal\\_id=90](http://www.scan.net.au/scan/journal/display.php?journal_id=90)
- Panzarino, M. (2013). *What exactly Wi-FiSLAM is, and why Apple acquired it*. Retrieved from <http://thenextweb.com/apple/2013/03/26/what-exactly-wifislam-is-and-why-apple-acquired-it/>
- Pao, Y., & Varatharajulu, V. (1976). Huygens' principle, radiation conditions, and integral formulas for the scattering of elastic waves. *The Journal of the Acoustical Society of America*, 59(6), 1361–1371. Retrieved from <http://link.aip.org/link/?JASMAN/59/1361/1>

## References

---

- Peltola, M., Lokki, T., & Savioja, L. (2012). Augmented reality audio for location-based games. In *Aes 35th international conference* (pp. 1–6). London. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=15176>
- Rämö, J., & Välimäki, V. (2012). Digital augmented reality audio headset. *Journal of Electrical and Computer Engineering, 2012*, 1–13. doi: 10.1155/2012/457374
- Ranjan, R., & Gan, W. (2013). Wave field synthesis: The future of spatial audio. *Potentials, IEEE, 32*(02). doi: 10.1109/MPOT.2012.2212051
- Riikonen, V., Tikander, M., & Karjalainen, M. (2008). An augmented reality audio mixer and equalizer. In *Proceedings of the 124th aes convention* (pp. 1–8). Amsterdam: Audio Engineering Society. Retrieved from [http://www.acoustics.hut.fi/~mak/PUB/AES124\\_ARA.pdf](http://www.acoustics.hut.fi/~mak/PUB/AES124_ARA.pdf)
- Sung, D., Hahn, N., & Lee, K. (2013). Individualized HRTFs simulation using multiple source ray tracing method. In *Aes 49th international conference: Audio for games* (pp. 2–11). London. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=16653>
- Sveiby, K. (1996). Transfer of knowledge and the information processing professions. *European Management Journal, 14*(4), 379–388. doi: 10.1016/0263-2373(96)00025-4
- Theageman. (2010). *Mackie HUI MIDI protocol*. Retrieved from <http://stash.reaper.fm/12332/HUI.pdf>
- Theile, G. (2004). Wave field synthesis: A promising spatial audio rendering

## References

---

- concept. In *Proceedings of the 7th international conference on digital audio effects* (pp. 393–399). Naples. doi: 10.1250/ast.25.393
- Tikander, M. (2009). *Development and evaluation of augmented reality audio systems*. Doctoral dissertation, Helsinki University of Technology. Retrieved from [http://www.usc.edu/dept/civil\\_eng/aerosol/papers/AST37\(3\)271-281.pdf](http://www.usc.edu/dept/civil_eng/aerosol/papers/AST37(3)271-281.pdf)
- Tikander, M., Karjalainen, M., & Riikonen, V. (2008). An augmented reality audio headset. In *Proceedings of the 11th international conference on digital audio effects* (pp. 1–4). Espoo: DAFX. Retrieved from [https://www.acoustics.hut.fi/dafx08/papers/dafx08\\_33.pdf](https://www.acoustics.hut.fi/dafx08/papers/dafx08_33.pdf)
- Tingen, P. (1997). *Tchad Blake: Binaural excitement*. Retrieved from [http://www.soundonsound.com/sos/1997\\_articles/dec97/tchadblake.html](http://www.soundonsound.com/sos/1997_articles/dec97/tchadblake.html)
- Valbom, L., & Marcos, A. (2005). WAVE: Sound and music in an immersive environment. *Computers & Graphics*, 29(6), 871–881. doi: 10.1016/j.cag.2005.09.004
- Weber, J. (2013). *Recorded sound*. Oxford University Press. Retrieved from <http://www.oxfordmusiconline.com/subscriber/article/grove/music/26294>
- Wessel, D. (2006). An enactive approach to computer music performance. *Le Feedback dans la Creation Musical*, 93–98. Retrieved from <http://www.grame.fr/RMPD/RMPD2006/pdf2006/wessel-1.pdf>
- Wozniowski, M., Settel, Z., & Cooperstock, J. (2012). Sonic interaction via



## References

---

- spatial arrangement in mixed-reality environments. *Sonic Interaction Design*, 17(32), 1–12. Retrieved from <http://xplore.staging.ieee.org/ebooks/6504614/6504636.pdf?bkn=6504614>
- Zimmerman, E. (2003). *Design research: Methods and perspectives*. London: The MIT Press. Retrieved from [http://books.google.com/books?hl=en&lr=&id=xVeFdy44qMEC&oi=fnd&pg=PA10&dq=Design+Research:+Methods+and+Perspectives&ots=RS\\_KIn6j35&sig=aQV02PXQ4UZzMu2iNSYWvdWIKDQ](http://books.google.com/books?hl=en&lr=&id=xVeFdy44qMEC&oi=fnd&pg=PA10&dq=Design+Research:+Methods+and+Perspectives&ots=RS_KIn6j35&sig=aQV02PXQ4UZzMu2iNSYWvdWIKDQ)
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 493–502). New York: ACM Press. doi: 10.1145/1240624.1240704