The University of Derby

College of Engineering and Technology

Department of Engineering

An Investigation into Ambisonic to Binaural Decoding for Virtual Reality.

By

Peter Rowland-Jones

September 2017

Submitted for in part-fulfilment of the requirements for the MSc in Audio Engineering

Acknowledgements

I'd like to thank Bruce Wiggins & Adam Hill for all the help throughout my course & this final piece of work.

I would also like to extend many thanks to Charles Middlicott & Neall Dewsbury for putting up with me during my time at Derby University.

Finally, a thank-you to my family for their continued support.

Abstract

The use of ambisonics for surround sound in virtual reality has grown rapidly. This involving the binauralisation of surround sound audio for use over headphones. When head-tracking is available, user based orientation can be applied to allow a more immersive experience. This project aimed to understand the differences between multiple virtual loudspeaker arrays and orders of ambisonic audio, when including head-tracking. The fundamentals behind implementing such a system are covered, leading into the creation of different ambisonic to binaural decoders in Matlab. These then used with Reaper and a head-tracking system, to allow a listening test to be completed. Results show that due to the lack of psychoacoustic optimisation, differences in localisation accuracy were not significant enough to define an optimum array. Although small differences in terms of lateral localisation can be found. Perceptual attributes were also found to increase in preference with order of array. While the occurrence of front/back confusion can still arise within head-tracked binaural systems.

List of Contents

1 Introduction 10
1.1 Rationale 10
1.2 Aims & Objectives 10
1.3 Structure of report 11
2 Background 12
2.1 Psychoacoustics 12
2.1.1 Interaural Cues 12
2.1.2 Pinnae & Head Movements 15
2.1.3 Rooms & Distance 16
2.2 Fundamentals 17
2.2.1 Binaural Audio 17
2.2.2 Head Related Transfer Functions 18
2.2.3 Ambisonics 19
2.2.4 Ambisonic Decoding 21
2.2.5 Virtual Loudspeaker Methods 23
2.2.6 Velocity & Energy Vector Analysis 24
2.2.7 Recent Investigations into Ambisonic Decoders 25
2.3 Virtual Reality 27
2.3.1 Facebook & Google 28
3 Methodology 30
3.1 Overview 30
3.2 Choice of arrays to test 30
3.3 Development of the Listening Test 32
3.3.1 Matlab Decoder Script 33
3.3.2 Decoder Matrix Generation 33
3.3.3 B-Format HRTF Generation 34
3.4 Head Tracker Implementation 36
3.5 Content creation 39
3.6 Pilot Study 41
3.7 Final Implementation 41
4 Analysis 44
5.1 Energy & Velocity Vector Analysis 44
5.2 Localisation Analysis 45
5.2.1 Multivariate Analysis 45
5.2.2 Individual Analysis 46
5.3 Perceptual Analysis 49
4.4 Stereo to B-Format 50
5 Discussion 51

5.1 Energy & Velocity 51
5.2 Localisation 51
5.3 Perceptual 52
5.4 Stereo to B-Format 52
5.5 Limitations & Observations 53
6 Conclusions 54
6.1 Future Work 55
7 References 56
8 Bibliography 59
9 Appendices 60
9.1 Spherical Harmonic Examples (Up to 3rd Order) 60
9.2 Virtual Arrays 61
9.2.1 Cube 61
9.2.2 Birectangle 61
9.2.3 Dodecahedron 62
9.2.4 26 Point Lebedev Sphere 62
9.3 Matlab Code 63
9.4 Decoder Matrix & Impulse Response Plots 63
9.5 Consent & Debrief Forms 68
9.6 Questionnaire Example 70
9.7 Results Plots 71
9.7.1 Cube/Birectangle Comparison 71
9.7.2 Perceptual Results 72
9.7.3 Stereo to B-Format & Stereo Comparison 73

List of Figures

FIGURE 2.1 - HEAD SHADOWING 12
FIGURE 2.2 - ILD FREQUENCY PLOT 13
FIGURE 2.3 - ITD EXAMPLE 13
FIGURE 2.4 - ITD IMPULSE RESPONSE 14
FIGURE 2.5 - HEIGHT AMBIGUITY 14
FIGURE 2.6 - MAA FREQUENCY & POSITION EXAMPLE 15
FIGURE 2.7 - CONE OF CONFUSION 15
FIGURE 2.8 – PINNAE FILTERING 16
FIGURE 2.9 - FRONT/BACK CONFUSION 16
FIGURE 2.10 - PRECEDENCE EFFECT 16
FIGURE 2.11 - KEMAR DUMMY HEAD 18
FIGURE 2.12 - BINAURAL TRANSMISSION PATH 18
FIGURE 2.13 - SPHERICAL HARMONICS 20
FIGURE 2.14 - AMBISONIC CHANNEL ORDERING 20
FIGURE 2.15 - VIRTUAL REALITY HEADSETS 28
FIGURE 2.16 - SADIE HRTF COMPARISON RESULTS 29
FIGURE 3.1- PLATONIC SOLIDS 31
FIGURE 3.2 - CHOSEN ARRAYS 32
FIGURE 3.3 - CUBE DECODER MATRIX 34
FIGURE 3.4 - CUBE B-FORMAT IMPULSE RESPONSE SET 35
FIGURE 3.5 - DYNAMIC BINAURALISATION 36
FIGURE 3.8 - AMBIX ROTATOR PLUG-IN 38
FIGURE 3.9 - FILTERS FOR PINK NOISE 39
FIGURE 3.10 - FLETCHER MUNSON CURVES 39
FIGURE 3.11 - FINAL TESTING SETUP 42
FIGURE 4.1 - E/V VECTOR ANALYSIS 44
FIGURE 4.10 - N1 C/B 60° LOCALISATION ACCURACY 47
FIGURE 4.11 - N1 C/B 150° LOCALISATION ACCURACY 47
FIGURE 4.12 - N2 C/B 150° LOCALISATION ACCURACY 47
FIGURE 4.13 - N3 C/B LOCALISATION ACCURACY 48
FIGURE 4.14 - SPEECH C/B 60° LOCALISATION ACCURACY 48
FIGURE 4.15 - DEPTH VARIANCE TRACK 1 & 2 49
FIGURE 4.16 - STEREO TO B-FORMAT RESULTS 50

List of Tables

TABLE 1 - LISTENING TEST VARIABLES....42
TABLE 3 - MAX E/V MAGNITUDE RESULTS....44
TABLE 4 – INDIVIDUAL STIMULUS MANOVA RESULTS....45
TABLE 5 - OVERALL MANOVA RESULTS....46
TABLE 6 - DEPTH PERCEPTION RESULTS....50

List of Equations

Equation 1 - Frequency Dependant ILD 12
Equation 2 - Maximum ITD 13
Equation 3 - Sum & Difference Polar Pattern 19
Equation 4 - Basic Ambisonic Encoding Equations 19
Equation 5 - Spatial Aliasing Requirements 21
Equation 6 - Projection Method For Decoding 21
Equation 7 - Spherical Harmonics 22
Equation 8 - Schmit Semi Normalisation 22
Equation 9 - Full 3D Normalisation 22
Equation 10 - Encoding Monophonic Source 22
Equation 11 - Mode Matching 22
Equation 12 - Pseudo Inversion 22
Equation 13 - B-Format HRTF Creation 23
Equation 14 - Binauralisation 23
Equation 15 - Vector Analysis - Pressure 24
Equation 16 - Vector Analysis - Energy 24
Equation 17 - Vector Analysis – Velocity 24
Equation 18 - Vector Analysis - Energy 25
Equation 19 - Euler Characteristic 31

Abbreviations

ACN – Ambisonic Channel Numbering

ANOVA – Analysis of Variance

DAW – Digital Audio Workstation

FFT – Fast Fourier Transform

FuMa – Furse Malham

HMD – Head Mounted Display

HRIR – Head Related Impulse Response

HRTF – Head Related Transfer Function

IFM – Inverse Filtering Method

ILD – Interaural Level Difference

IPD – Interaural Phase Difference

ITD – Interaural Time Difference

MAA – Minimum Audible Angle

MANOVA – Multivariate Analysis of Variance

N3D – Normalised 3D

OBA – Object Based Audio

OSC – Open Sound Control

rE - Energy Vector Magnitude

rV - Velocity Vector Magnitude

SADIE – Spatial Audio for Domestic Interactive Entertainment

SBA – Scene Based Audio

SN3D – Schmit Semi-Normalisation

uE – Energy Vector Angle

uV – Velocity Vector Angle

VBAP – Vector Base Amplitude Panning

VLM – Virtual Loudspeaker Method

VR –Virtual Reality

1 Introduction

This work investigates the decoding of ambisonic material for binaural use. The process of which being a common method of providing surround sound for virtual reality applications over headphones. Overcoming the requirement for large loudspeaker arrays to experience this surround sound format. Ambisonics is favoured due to its separation of encoding & decoding for playback, allowing audio to be prepared for multiple arrays. When used over headphones, virtual loudspeaker methods can be employed to simplify the decoding process. This meaning multiple virtual loudspeaker arrangements can be used.

1.1 Rationale

Thanks to Moore's law (Moores, 2017), current advancements in personal mobile devices have allowed the use of head-tracking with ambisonics for virtual reality media. This to provide highly immersive sound-fields that can be easily rotated. A technique commonly completed using virtual loudspeaker arrays over headphones. These methods have recently been adopted by such companies as Google (Youtube) for 360 media online. The main aspect of this process being the ability to take the same piece of audio and prepare it for multiple loudspeaker arrangements. This process of decoding has been investigated by many researchers (Wiggins et al,2001 & 2004) (Heller, 2008) (Kearney, 2015). A common factor between such testing methods has been the use of static binaural listening. Whereas some investigations (Horbach et al, 1999) have included headtracking, its use with modern technology is still in its infancy. Such studies by Wiggins et al (2001) make key comparisons between real and virtual arrays & their large differences when tested. The rationale for this work is to understand possible implications of using different virtual arrays when employed over virtual loudspeaker arrays and ambisonic orders.

1.2 Aims & Objectives

This research aims to investigate the process of ambisonic to binaural decoding over multiple orders of material. The investigation will allow a better understanding of possible differences found within the virtual arrays. Such as localisation accuracy, & perceptual implications.

The objectives are as follows

- Implement multiple virtual loudspeaker arrays over a binaural system, using different ambisonic orders.
- To build a head tracking system capable of providing and storing location data for use with the decoders.
- Provide insight into favourable decoding methods for binaural use, depending on ambisonic order.

1.3 Structure of report

This work is split up into 3 main sections as follows

1. Background and Fundamentals
    - Psychoacoustics
    - Surround Sound Fundamentals
    - Virtual Reality Applications
2. Methodology
    - Chosen Arrays
    - Development of the Listening Test
    - Head-tracking system Implementation
    - Final Implementation & Pilot Study
3. Discussion & Conclusions
    - Discussion
    - Conclusions
    - Future Work

Section one will introduce main concepts of spatial audio and surround sound fundamentals, while also outlining the basics of decoding methods that will be implemented in further sections. Section 2 will go on utilise this work to develop the chosen methodology for the implemented testing. This section also outlines the use of a head tracking system, capable of providing positional data for the test. The last section will discuss the findings from the chosen tests, while the discussion will then aid conclusions on possible optimum solutions for ambisonic to binaural decoding methods.

[Throughout the text, the following symbols are used to describe each 3D plane, Azimuth = θ & Elevation = ϕ.]

2 Background

2.1 Psychoacoustics

Before being able to create a surround sound system for accurate playback of a sound-field, it is important to understand how the human hearing system completes this process naturally. The following section describes key cues for human hearing and how they are used to aid the auditory system in everyday life. This then leading into a chapter on the fundamentals of re-creating these cues, for use in surround sound systems.

The use of two ears is referred to as a 'binaural' system. This helping humans to distinguish a source direction and distance on a given plane, completed by assessing complex wave-fronts that arrive at each ear. The brain can then use these auditory path differences to make assessments on its surroundings (Mather 2009). That is, that each ear will receive an independent signal that differs in time and level. These differences referred to as interaural cues, used by the auditory system for lateralisation of sound.

2.1.1 Interaural Cues

Interaural Level Difference (ILD) – is the difference in level of a source received at each ear, mainly due to head shadowing. Assuming a human head as a sphere, with a diameter of 20cm and ears placed at 90/270°. Any frequency with a wavelength greater than the given diameter, will be diffracted around the head. While wavelengths above this frequency threshold will be attenuated by the head itself. Therefore, ILD's are frequency dependant & will become ambiguous below a particular frequency, governed by EQ.1 below(Wiggins, 2004) .

$$f = \frac{1}{0.2} * C \sim 1.7kHz$$

EQUATION 1 - FREQUENCY DEPENDANT ILD

C = Approximate speed of sound (343 Metres per sec).

0.2 = Diameter of Sphere (Human Head) in metres.

It must be highlighted that this frequency will be marginally different for everyone, due to having different shape/size head, while also having ears at different locations on the head. An example below is shown in Fig 2.1 & 2.2 of a source placed at 90° θ; here you can see the level differences experienced above ~1 kHz due to head shadowing.

Head Shadowing

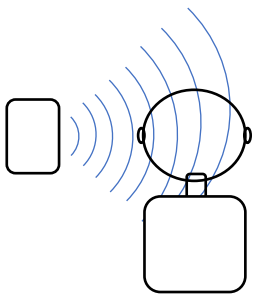


FIGURE 2.1 - HEAD SHADOWING

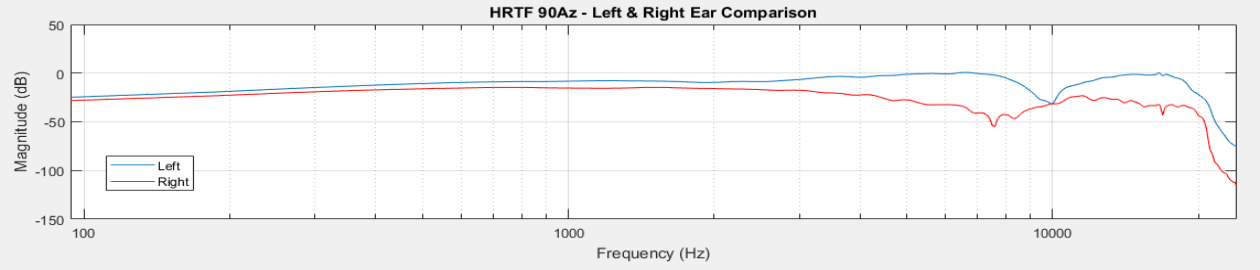


FIGURE 2.2 - ILD FREQUENCY PLOT

Interaural Time Difference (ITD) – is the difference in path length for a soundwave to reach each ear. For example, a source at 330° θ will reach the ipsilateral ear sooner than the contralateral, see Fig 2.3. The distance between the ears will always remain fixed, meaning a maximum time difference can be calculated.

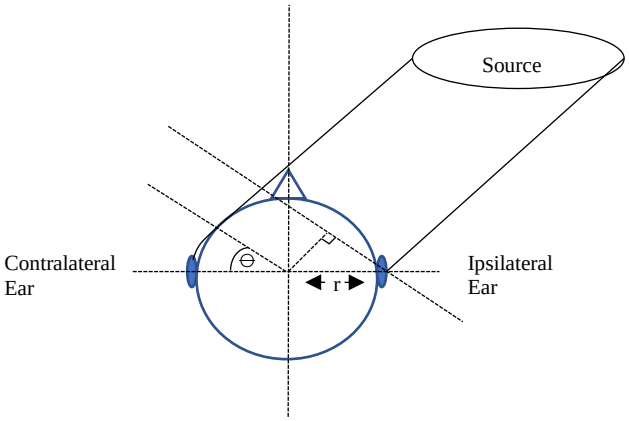


FIGURE 2.3 - ITD EXAMPLE

(Rumsey, 2001)

Taking the previous diameter (d) from EQ.1 as 0.2 metres, and the lateral positions of the ears to be 90/270° θ. The maximum difference in distance, is half the circumference of that sphere, shown in EQ.2 (Wiggins, 2004) .

$$Hc = \pi r = \pi * 0.1 = 0.314m$$

Where r = d/2.

Hc = Half Circumference.

The maximum time difference then resulting in,

$$Max\ ITD = \frac{Hc}{C} = \frac{0.314}{343} = 0.91ms$$

EQUATION 2 - MAXIMUM ITD

Where C = 343m/s

This being the maximum amount of delay that can occur between signals received at the left & right ear.

These ITD's are used to lateralise frequencies <~2kHz, due to higher frequency wavelengths making time differences ambiguous (Mather 2009).

Note the time delay of the impulse for the contralateral ear in Fig.2.4 for a source at 90° θ.

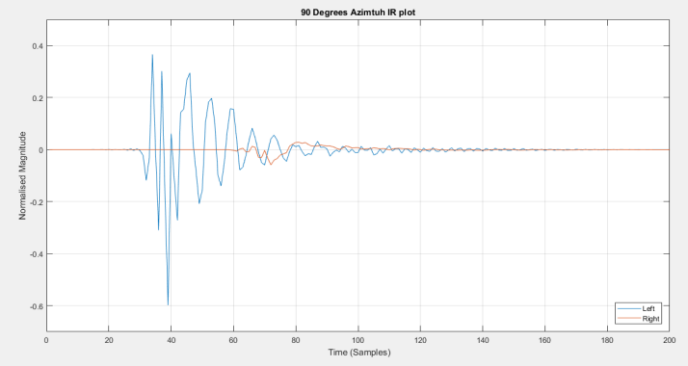


FIGURE 2.4 - ITD IMPULSE RESPONSE

ITD is also used in perception of a sources elevation angle, as shown below in Fig 2.5.

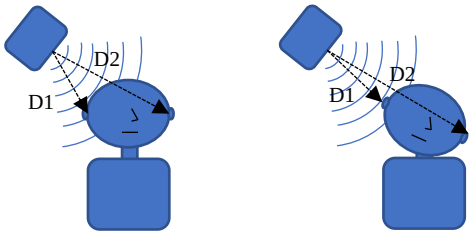


FIGURE 2.5 - HEIGHT AMBIGUITY

Interaural Phase Difference (IPD) – is dependent on the angle of offset from 0° θ in the frontal plane. The distance between the ears being constant, the IPD is location and frequency dependant <~600Hz, due to wavelengths being equal to half the distance between the ears. For example, the path length increment found previously in EQ.2 was ~0.314m; while the wavelength of a 546Hz tone ~0.628cm. Above this frequency, wavelengths become ambiguous; while frequencies below placed at 90 or 270° θ, can cause an IPD of 180°. This causing uncertainty as to whether one ear is lagging or leading, due to the auditory system being unable to distinguish absolute phase (Wiggins, 2004). This also outlined in a recent paper by Collins (2013), in which inaccuracy in these lateral positions is argued to be due to a narrow cone of confusion and pinnae dependencies (These covered in following sections). Collins then goes on to discuss the use of head-tracking to overcome this issue in surround sound systems.

The main difference between time & phase cues is the presence of a time of arrival to the ear. If a continuous sound is used, only phase and level cues can be utilised. Whereas, when using a pulsed source, timing/envelope cues can be used to overcome possible issues found at lateral positions where phase differences lack certainty. This method used in later discussed testing methodologies.

Lastly, a key aspect of lateralisation of a source is the minimum audible angle (MAA); this being the extent to which a source must move from a reference point, for it to be perceived as being in a different location. Mather [2009] states this MAA can be accurate up to 1° for frequencies <~1kHz. With large variances between the 1.5kHz – 3kHz range & a greater resolution of accuracy at 0° & 180° when compared to lateral positions.

This phenomenon commonly termed Duplex theory. An investigation by Wiggins goes some way into displaying this when positioning 3 sources at different positions on the θ plane, shown in Fig 2.6 (Wiggins 2004).

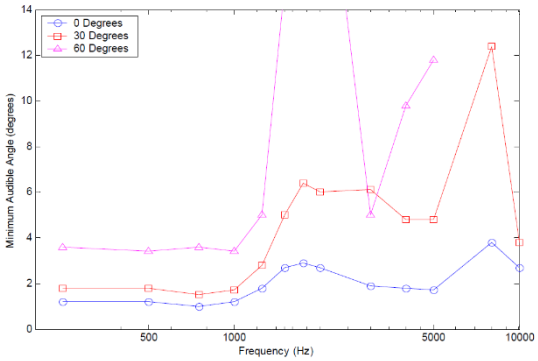


FIGURE 2.6 - MAA FREQUENCY & POSITION EXAMPLE

(Minimum audible angle between successive tones as a function of frequency and position of source.) (Image taken from Wiggins,2004).

It is also apparent from Fig 6 that lower frequency phase cues are more reliable, when assessing the angle of incidence for a sound source. Where as in the crossover region between low and mid frequencies, large variances can be seen due to the crossover between interaural cues.

2.1.2 Pinnae & Head Movements

The previous section discussed the brains ability to lateralise sound sources due to time/level and phase differences between the ears. While these are key, a so called 'cone of confusion' can also occur. In which the path length of a sound source to the ear is the same regardless of its position, see Fig 2.7. Each cross on the circumference of the circle has the same path length to the ear.

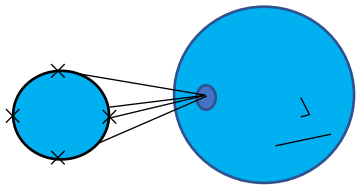


FIGURE 2.7 - CONE OF CONFUSION

For example, a source at 0° θ & a distance of 1.5m, will have the same path length to each ear as 180° θ at 1.5m. To overcome this, the physical shape of the outer ear (Pinnae) helps as a form of acoustic filtering, used to alter frequency and phase components of arriving sounds (Mather, 2009). The upper limit of which being dependant on the physical size of the pinnae. Due to this, the pinnae can create spectral cues that are specific to each angle of incidence, which the auditory system uses to aid its perceived location (Ballou, 2015). Note the similar time of arrival for each position in Fig 2.8, with slight level difference due to pinnae filtering.

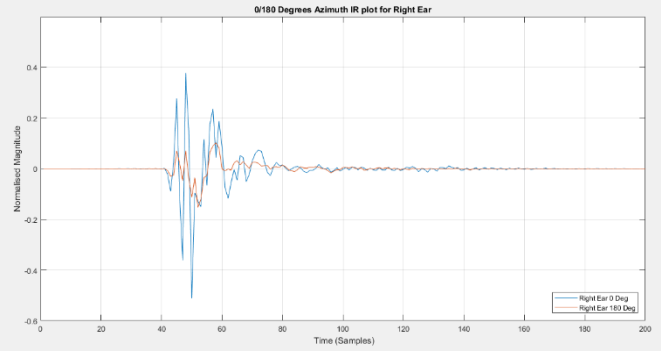


FIGURE 2.8 – PINNAE FILTERING

In conjunction, the addition of head movement can greatly improve localisation. This due to the fixed position and distance between the ears. By rotating the head, interaural cues are altered, to help create an accurate picture of the sources position & reduce such issues as cone of confusion (Ballou,2015) as shown in Fig 2.9. Studies such as Horbach et al (1999) make note of the vast improvements head movements can make to removing front & back confusion while also improving localisation accuracy, a consideration that will be assessed during the following work.

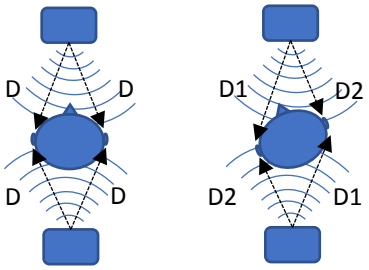


FIGURE 2.9 - FRONT/BACK CONFUSION

2.1.3 Rooms & Distance

A key factor in localisation of sound, is the perception of distance and room size. Where the direct to reverberant ratio is the distance at which a source can be placed before late reflections are perceived at a higher level than the direct sound. Thus, masked by the rooms response & giving a higher overall sense of space and distance of a source (Ballou, 2015). Overall level can also aid the perception of distance, due to the inverse square law. Although, this would require a reference level for this to be used as a distance cue. Distance perception is also frequency dependant, due to air absorption of higher frequencies, but is generally only found over long distances. For example, the clash of thunder when close, compared to a rumble at a distance. It must be noted that interaural level differences play a large part in the perception of distance when within a very close range, whereas the precedence effect will give priority to the first arriving wave front within a 5ms delay (Fig 2.10). Therefore, distance perceptions are also dependant on time and level cues at the ears (Mather, 2009).

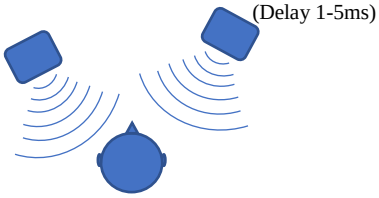


FIGURE 2.10 - PRECEDENCE EFFECT

To summarise, the brains auditory system uses many different cues to aid lateralisation and localisation of sound. The use of time, level & phase differences at the ears can help determine overall position. While the inclusion of reflections and physical movement/filtering, can help discriminate errors between them, & place the source in a perceived space. The combination of these cues helps distinguish a sources location & environment with a high degree of accuracy that should be replicated during playback of convincing sound-fields.

2.2 Fundamentals

The previous chapter outlined the way in which humans perceive sounds sources and their given environment. Allowing the understanding of key concepts for consideration throughout the following work. The next chapter leads onto fundamental areas that will be used in the creation of a system to assess the impact of multiple ambisonic surround sound arrays over headphones.

2.2.1 Binaural Audio

In 1933, Alan Blumlien released a patent outlining differences between stereophonic and binaural audio. This being an extension of the previous work by Bell Labs, in which the capture & reproduction of a wave front was assessed (Blumlein,1933). Findings being that the spacing of microphones/ loudspeakers when recreating a sound stage, could lead to spatial aliasing issues if over the previously discussed maximum interaural time difference. Spatial aliasing being audible distortion due to an insufficient number of loudspeakers representing the sound stage. These holes between loudspeakers causing detent effects when panning a source. When placing a monophonic source between two loudspeakers, two key techniques used are constant power & constant voltage (Sine/Tangent Panning Laws). Theses allowing phantom images to be created dependant on the desire for constant source power or output voltage across the ‘pan pot’ range. The use of vector base amplitude panning (VBAP) can also be employed for source positioning between 2D & 3D arrays (Pulkki, 1997). This technique allowing derivation of loudspeaker vector gains, using inverse identity matrix techniques, similar to those discussed in future sections.

Blumlein went on to present a near coincident spaced omni microphone technique, that he termed ‘Binaural Reproduction’. Due to its similar spacing as a human head with a central baffle, phase differences found at the ears could be re-created (Blumlein, 1933). This providing a simple example of capturing the transfer function from source to the auditory system, that works well when used over headphones. When considering the crosstalk found in loudspeaker playback, Blumlein also presented his ‘Shuffler’ network, which would convert lower frequency phase differences captured at the mic, into amplitude differences at loudspeakers.

Blumlein then went on to make use of coincident mic techniques with figure of 8 microphones. This allowing mono compatibility & spatial equalisation through sum & difference equations, for manipulation of audio post & pre decoding (Wiggins, 2004). These being techniques have been developed for use in periphonic sound, which is discussed later in this section.

2.2.2 Head Related Transfer Functions

The transmission path from a sound source to the ear is termed the 'head related transfer function' (HRTF). The capture of such being very similar to Blumlein's early 'Binaural Audio'. It contains information dependant on position, room, head, torso & pinnae to name just some. To simulate this for binaural playback over headphones, head related impulse response (HRIR) libraries are created. To complete this capture process, either an individual listener can wear small microphones that sit at the entrance or inside of the ear canal, or a dummy head can be used that imitates the human head, torso & pinnae as seen in Fig 2.11 (Aimil, 2016).

FIGURE 2.11 - KEMAR DUMMY HEAD

For acceptable resolution of the HRIR library, many positions are captured, or interpolation between them can be used. Although, this can lead to a lengthy process, so dummy head approaches are favoured; due to the ability to accurately rotate a dummy head if needed & avoid human interference (Rumsey, 2001).

It can be argued that dummy head methods fail to accurately recreate the required cues for all users, such as accurate pinnae filtering. While this may be true, investigations have shown that after a brief training period (~3 weeks), an individual can adapt to the HRTF set and use it to the same accuracy as their own (Ballou, 2015), showing just how adaptive the human auditory system can be.

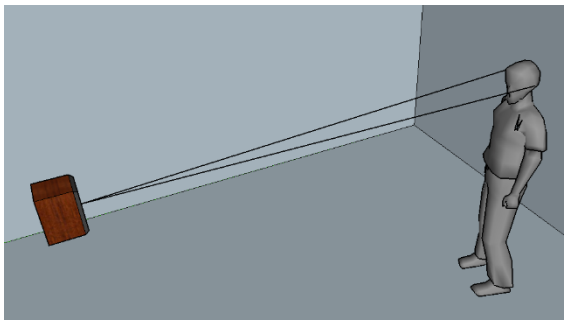


FIGURE 2.12 - BINAURAL TRANSMISSION PATH

When considering audio prepared for loudspeaker playback, the HRTF is a natural process from loudspeaker to each ear, see Fig 2.12. But when listening to this material over headphones, transmission paths are bypassed; such as pinnae filtering. Therefore, listening to music prepared for loudspeakers over headphones results in 'In head' perceptions. To overcome this, a HRTF library can be used. In which, audio feeds for each headphone (Left & Right) are convolved with the HRIR' corresponding to the required location. This concept being expanded in further sections.

2.2.3 Ambisonics

Ambisonics was originally pioneered by Michael Gerzon in the 1970's, as a first attempt to accurately recreate a sound-field in way that overcomes issues found in discrete playback systems (Gerzon, 1974). It initially faced popularity issues due to previous attempts to introduce surround systems to the market that were unsuccessful, such as quadrophony (Hollerweger, 2006). This now changing with companies beginning to understand the advantage of its flexibility as an audio format and its application within virtual reality content.

Ambisonic's allows full sphere isotropic decomposition of a sound-field. It was developed as an extension to Blumlein's original method of deriving multiple polar patterns from coincident microphone techniques. In which sum & difference equations for two microphone feeds can be mixed to allow manipulation of polar patterns. For example, two figure of 8 polar patterns, 90°out of phase on the azimuthal plane, can be used to create any polar pattern in a given direction (0-360° θ).

$$Sum = (L + R) / \sqrt{2}$$

$$Difference = (L - R) / \sqrt{2}$$

$$Figure\ 8 = (\cos(\theta)\ x\ sum) + (\sin(\theta)\ x\ Difference)$$

EQUATION 3 - SUM & DIFFERENCE POLAR PATTERN

(Wiggins, 2004)

In terms of encoding, a sound-field microphone (Soundfield, 2017) or others such as the Eigen mic (mhacoustics(b), 2017) can be used; or encoding software plugins to synthesise ambisonic material. The resulting B-Format channels, can then represent a sound-field in 2 or 3 Dimensions. The basic synthesised encoding equations for each B-Format channel (1st order FuMa) can be found below in Eq.4 (Hollerweger, 2006).

$$W = \sum_{i=1}^{k} si \frac{1}{\sqrt{2}}$$

$$X = \sum_{i=1}^{k} si \cos\theta i \cos\Phi i$$

$$Y = \sum_{i=1}^{k} si \sin\theta i \cos\Phi i$$

$$Z = \sum_{i=1}^{k} si \sin\Phi i$$

Si = mono source

K = number of sound sources

θi /ϕi = Azimuth & elevation direction.

EQUATION 4 - BASIC AMBISONIC ENCODING EQUATIONS

As shown, simple multiplications of sources with sin & cosine terms allow application onto the ambisonic plane to create B-Format audio. Note that the Z channel is only dependant on elevation, and W is an omni-directional channel that requires attenuation of 0.707 (√2) if using Furse Malham conventions.

The increase in ambisonic order is determined by the number of encoded channels, this either being at capture, with a higher order sound-field microphone (Soundfield, 2017), or being encoded with such software as 'ambix' plugins provided by Michael Kronlacher, or 'WigWare' by Dr Bruce Wiggins (Wiggins – Wigware, 2017). An example of resulting spherical harmonics is shown below for 0 & 1st order (Furse Malham) in Fig 2.13, with a full example up to 3rd order in appendix 9.1.

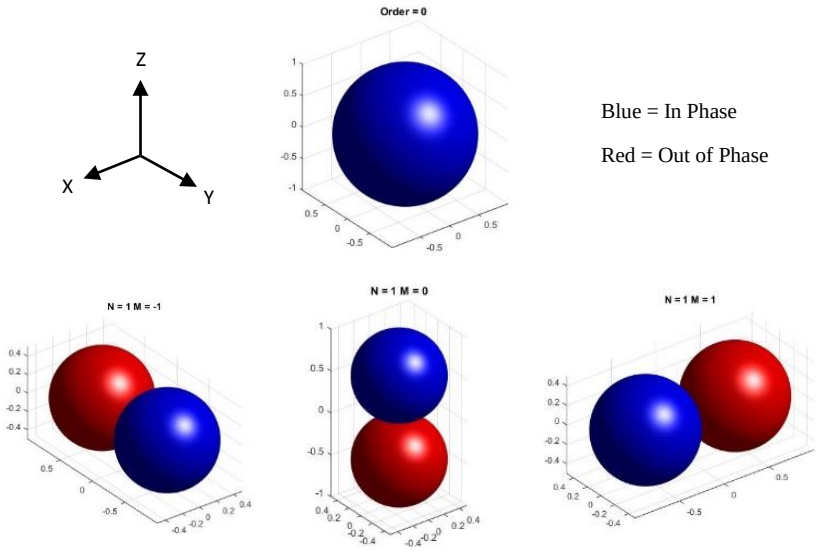


FIGURE 2.13 - SPHERICAL HARMONICS

The ordering of these ambisonic channels must be retained during use. Below in Fig 2.14 is an example of the original 'Furse Malham' (FuMa) channel ordering, while on the right is the more current 'Ambisonic channel ordering' (ACN). The difference being, FuMa referencing from the outer channels inward as they increase in order, while ACN references from left to right. The addition of a 0.707 (1/√2) channel weighting applied to the omni (W) channel is also used in FuMa based ambisonic encoding, which must also be accounted for in conversion between the two formats.

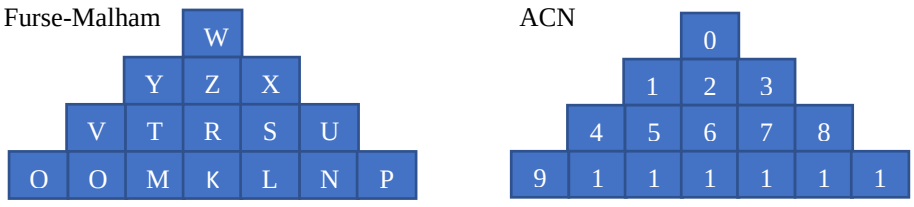


FIGURE 2.14 - AMBISONIC CHANNEL ORDERING

The advantage of ambisonics is that the encoding and decoding of audio streams are separated. Unlike other formats that require the individual preparation of audio for each discrete playback system. This being constrained by the need for more loudspeakers than ambisonics channels. The order of ambisonic playback then being governed by Eq.5,

$$N(2D) = (2n + 1)$$

$$N(3D) = (n + 1)^2$$

(2D & 3D requirements)

N = Number or required loudspeakers

n = Ambisonic order (Kearney et al, 2016)

EQUATION 5 - SPATIAL ALIASING REQUIREMENTS

2.2.4 Ambisonic Decoding

Once the B-Format channels have been created, these can be decoded to a specified loudspeaker array. Projection methods can be used for simple regular shaped arrays. To complete this, the below equation can be used (Eq.6, Wiggins, 2004). Then by varying the 'd' value, a different polar pattern can be achieved for each loudspeaker position.

$$g_w = \sqrt{2}$$
$$g_x = cos(\theta)\cos(\Phi)$$
$$g_y = sin(\theta)\cos(\Phi)$$
$$g_z = \sin(\Phi)$$

$$S = 0.5\ x\ [(2 - d)g_w W + d(g_x X + g_y Y + g_z Z)]$$

Where WXYZ = B Format Channels

S = Speaker Output

θ = Azimuth Location of Loudspeaker

ϕ = Elevation Location of Loudspeaker

d = Directivity Factor (0-2)

EQUATION 6 - PROJECTION METHOD FOR DECODING

When spherically decomposing a sound-field, mode matching methods can be also used. In which, a decoding matrix is derived containing a weighted sum of each channel for each loudspeaker, depending on its position (Hollerweger, 2006). This being a spherical harmonic representation of each loudspeaker location. When using regular arrays, the resulting source positioning is equally wrong in all directions, regardless of whether in-between loudspeakers or positioned directly, overcoming issues with common stereo panning techniques. Allowing an even representation of recreated angles in full periphony, this creating the feeling that the loudspeakers 'disappear', for a natural reproduction (Wiggins, 2004).

To place each loudspeaker location in the spherical harmonic domain, the equation below (Eq.7) can be used (Hughes, et al 2016). This dependant on its position, defined ambisonic order/degree, normalisation scheme and associated Legendre polynomial. The Legendre polynomials being a set of orthogonal solutions to differential equations used in mathematics with spherical symmetry. (McQuarrie, 2003)

$$Y_n^m(\phi,\theta) = N_n^{|m|} P_n^{|m|}(\sin(\theta)) \begin{cases} \cos(|m|\phi) & if\ m \geq 0 \\ \sin(|m|\phi) & if\ m < 0 \end{cases}$$

Where Y = Ambisonic Signal

n = Order

m = Degree

ϕ = Elevation

θ = Azimuth

N = Normalisation scheme

P = Associated Legendre Polynomial

EQUATION 7 - SPHERICAL HARMONICS

The normalisation equations shown below are N3D (Full 3D) (Eq.9) & SN3D (Schmit-Semi Normalisation) (Eq.8) (Kearney et al, 2016). Note that N3D is an extension of the SN3D format. SN3D is most commonly used with ambisonic ACN in current implementations online for virtual reality. These are used to apply an order specific amount of normalisation in terms of gain to each set of ambisonic channels.

$$N_{\ell,m}^{SN3D} = \sqrt{(2-\delta_m)\frac{(\ell-|m|)!}{(\ell+|m|)!}}, \delta_m \begin{cases} 1 & if\ m = 0 \\ 0 & if\ m \neq 0 \end{cases}$$

EQUATION 8 - SCHMIT SEMI NORMALISATION

$$N_{\ell,m}^{N3D} = N_{\ell,m}^{SN3D}\sqrt{2\ell+1}$$

EQUATION 9 - FULL 3D NORMALISATION

The encoded source can then be represented using,

$$b = sY$$

EQUATION 10 - ENCODING MONOPHONIC SOURCE

Where (s) is the source, being multiplied by the spherical harmonic vector value (Y), depending on order & array. (Kearney et al, 2016)

When using mode-matching, a source (b) can be multiplied by the inverse loudspeaker re-encoding matrix (D). A pseudo-inversion (Moore-Penrose) of the re-encoding matrix is completed, to allow inversion of a non-square matrix, shown in Eq.12. This then resulting in the final column vector gains for each channel, Eq.11.

$$g = Db$$

EQUATION 11 - MODE MATCHING

Where

$$D = L^{\dagger} = L^T(LL^T)^{-1}$$

EQUATION 12 - PSEUDO INVERSION

(Kearney et al, 2016) $\quad \dagger = Pseudo\ Inverse \quad T = Transpose$

Completing this pseudo inversion creates a matrix of specific loudspeaker gains required to place a source in each location within the array. Using this method will attempt to account for slightly irregular arrays such as rectangles, by offsetting the spherical harmonic locations, this due to its mathematical derivation. As opposed to using the previously mentioned projection methods that will remain pointing at the loudspeaker location, resulting in an uneven representation of the original sound-field and thus spatial aliasing.

2.2.5 Virtual Loudspeaker Methods

A paper by Wiggins (2001), discusses the use of virtual loudspeaker methods (VLM). This completed by taking a final decoder matrix and its corresponding loudspeaker gains, & multiplying each with its corresponding HRTF location. When convolved with incoming B-Format material of the same order, audio could be perceived as being played through the defined array. This completed using the equations below Eq.13 (Example for 8 loudspeaker array, FuMa) (Wiggins, 2004).

$$W^{hrtf} = \left(\sqrt{2}\right) \cdot \sum_{k=1}^{8} \left(S_k^{hrtf}\right)$$

$$X^{hrtf} = \sum_{k=1}^{8} \left(\cos(\theta_k)\sin(\phi_k) \cdot S_k^{hrtf}\right)$$

$$Y^{hrtf} = \sum_{k=1}^{8} \left(\sin(\theta_k)\sin(\phi_k) \cdot S_k^{hrtf}\right)$$

$$Z^{hrtf} = \sum_{k=1}^{8} \left(\sin(\phi_k) \cdot S_k^{hrtf}\right)$$

$$s_K^{hrtf} = Pair\ of\ speakers\ positional\ HRTF's$$

K = Loudspeaker feed.

EQUATION 13 - B-FORMAT HRTF CREATION

The advantage of this being fewer HRTF's are needed, as they are incorporated into the B-format decoding matrix themselves, as opposed to two per loudspeaker. Once each channels HRTF filter is created, it can be convolved with the given input signals as shown below to allow binauralisation (Eq.14). If the inputted ambisonic channels are rotated prior to convolution, dynamic binauralisation can be applied when used in real-time.

$$L = (HRTF_{Lw} \otimes Ch0) + (HRTF_{Lx} \otimes Ch1) \ldots.$$

$$R = (HRTF_{Rw} \otimes Ch0) + (HRTF_{Rx} \otimes Ch1) \ldots.$$

Where $\otimes = Convolution$

EQUATION 14 - BINAURALISATION

This allowing auditioning of different virtual arrays and orders of ambisonic material over headphones, & the user to explore the sound-field with ease. These are similar techniques to those found in online decoders currently used by Google and Facebook.

2.2.6 Velocity & Energy Vector Analysis

Velocity & energy vectors are a mathematical representation of how a decoder should perform, these proposed by Michael Gerzon in 1992. Each vector has a length which relates to its magnitude, representing overall quality of localisation; & an angle to represent its perceived source location. When used with a virtual source panning around an array, assessment of how the decoder should perform can be made.

Key characteristics of an ambisonic decoder were set out by Michael Gerzon in 1992. (Gerzon, 1992b):

- Velocity & energy vector directions are the same, up to approximately 4kHz. Such that the reproduced azimuth $\phi V = \theta E$ is substantially unchanged.
- At low frequencies, magnitude of the velocity vector is near unity for all reproduced azimuths.
- At mid/high frequencies, the energy vector magnitude (rE), is substantially maximised across as large a part of the 360° sound stage as possible.

(Wiggins, 2004)

The energy a pressure values can be found using the equations given below.

$$P = \sum_{i=1}^{n} gi$$

P = Pressure due to the speakers output

EQUATION 15 - VECTOR ANALYSIS - PRESSURE

$$E = \sum_{i=1}^{n} gi^2$$

E = Energy due to the speakers output

Where gi = ith loudspeaker gain.

EQUATION 16 - VECTOR ANALYSIS - ENERGY

Where the acoustical pressure, is the sum off all loudspeaker gains for that source direction. The lower frequency localisation vector can then be calculated as the overall acoustical velocity vector gain of a sound at the central listening position, by the associated acoustical pressure gain. Resulting in a velocity vector,

$$V_x = \sum_{i=0}^{n} gi \cdot \frac{\cos(\theta_i)}{P}$$

$$V_y = \sum_{i=0}^{n} gi \cdot \frac{\sin(\theta_i)}{P}$$

Where gi = Gain of the I'th Loudspeaker Feed

n = Number of Loudspeakers

θ = Angular position of Loudspeaker

EQUATION 17 - VECTOR ANALYSIS – VELOCITY

The resulting velocity localisation vector lengths (rV) should be maximised to 1 while the vector direction (uV) reflects the perceived source angle.

Optimisation for frequency ranges between 700Hz – 4kHz can be achieved by maximising the energy vector. This being similar to the velocity localisation vector but using 'gi²', therefore the average acoustical energy within the space, see EQ.18. This then used to determine the ratio of acoustical energy to the sound intensity gain of the source (Gerzon, 1992).

$$E_x = \sum_{i=0}^{n} gi^2 \cdot \frac{\cos(\theta_i)}{E}$$

$$E_y = \sum_{i=0}^{n} gi^2 \cdot \frac{\sin(\theta_i)}{E}$$

Where gi = Gain of the I'th Loudspeaker Feed

n = Number of Loudspeakers

θ = Angular position of Loudspeaker

EQUATION 18 VECTOR ANALYSIS - ENERGY

The same process is followed as before, resulting in the energy vectors for mid frequency optimisation. Again, the vector magnitude should be as close to 1 as possible (rE) with vector direction angle (uE) denoting perceived source angle. Due to the energy vector length being an average of 'n' vectors of length 1, it can only be equal to 1 if all sound only originates from a single loudspeaker. This is generally not the desired case, so should be optimised to give the best performance possible (Gerzon, 1992).

When designing an ambisonic decoder, optimisation techniques can be used to provide correct decoding to a wider listening area. To achieve this psychoacoustic optimisation, the decoder matrix is sent to a phase compensated crossover network with a crossover region of ~400Hz, commonly termed dual band decoding. This crossover point would generally be higher (~700hz) but is lowered to compensate for issues encountered by listeners that may not be within the central position. Implementation of this can be completed through applying MaxRe gain values to the higher frequency shelf; a technique also used by [Kearney et al, 2016] with VLM & found to be preferred in many studies such as Heller (2006) to other decoding methods.

Gerzon & Barton [1992] also discuss the use of near field compensation to overcome low frequency addition from curvature of the virtual array, dependant on distance. To overcome this issue, use of a first order high pass filter with a cut-off frequency of 53/d (where d = distance in meters) in the velocity signals path can be implemented. This having a main influence on phase, not amplitude, to overcome issues like that of proximity effect in condenser microphones.

2.2.7 Recent Investigations into Ambisonic Decoders

In 2008, [Heller et al] completed an evaluation of decoders and what makes them ambisonic. The fuel for the study being many methods of ambisonic playback being implemented insufficiently, providing inadequate results and thus poor localisation and spectral qualities. Testing showed that the lack of knowledge in terms of proper

implementation of ambisonic playback has a dramatic effect on the perceived quality. Such as "Phasey" issues commonly mentioned and believed to be due to the use of exact velocity decoding for higher frequencies.

The testing of each decoder was completed in a 'black box' manner to assess individual components of the system rather than just its output. 72 unit impulses were encoded into B-Format that panned around the full azimuth plane. This then played back over each decoder matrix, consisting of 1st order material over 4 loudspeakers (288 IR results per system). From these results, energy and velocity vector analysis was completed; allowing assessment in comparison to the previously mentioned benchmarks set out by Gerzon in 1992.

4 decoders were tested in total, with findings showing that even when using pre-set configurations of software decoders, results were not always optimal for the desired loudspeaker layout. It was noted how the chosen arrays can have an adverse effect on preference, for example a 2D horizontal array may be favoured over a 3D cube array due to containing loudspeakers at ear level.

Further listening tests were carried out to assess the perceptual attributes of the decoders in comparison to the prior analysis, with test material comprising of voice announcements, panned pink noise & musical recordings. With results supporting findings from the initial analysis.

In 2001 [Wiggins et al] conducted a study into the analysis of multichannel surround sound algorithms using HRTF simulation. This completed through a comparison between real source localisation listening tests and HRTF analysis. The systems used were,

- 5 & 8 speaker 1st order ambisonic arrays
- 8 speaker 2nd order (horizontal)
- Amplitude panned over 5 speakers
- Stereo dipole (2 speakers at +/- 5°).

The real listening tests were conducted at the multichannel sound lab at Derby University, with an average ambient noise floor of 43dBa (Peak – 52.1dBa at 100Hz), & RT60 of 0.42 seconds. The room allowed assessment of multiple speaker layouts in quick succession using a multichannel soundcard. In total 28 participants were used, made up of students enrolled on a music technology based course with basic knowledge of the area.

The test material used was band limited pulsed noise, (3 pulses per signal each 2 second's long with 1 second gaps between). The band limited ranges were set to represent approximations of low/mid/high frequencies (0-1kHz, 1kHz – 5kHz, 3kHz – 20kHz) a method also used in Heller (2006). When conducting the listening tests, participants were asked to keep their head as still as possible & then mark perceived source location in ° on a polar plot provided (5° increments).

Virtual loudspeaker methods were used & taken one step further with a symmetrical HRTF dataset and inversion of the B-Format Y channels phase, allowing for accurate horizontal panning. This resulting in just 3 HRTF's needed for any 1st order array.

The HRTF analysis was completed for 3 areas – ITD, ILD & Pinnae filtering. Level differences assessed through mean amplitude differences between the Fast Fourier transform (FFT) of both L & R impulses. Time differences assessed using the group delay between the two ears/impulses. Pinnae filtering then tested through the actual time and level values from the group delay & FFT of the impulses.

Results showed the 2nd order 8 speaker array to perform best overall when compared to the real source results. Interesting results being that the 5 speaker 1st order system seemed to outperform the 8 speaker 1st order system.

As expected group delay and amplitude levels changed vastly as a source was panned around each array, but when compared to the same array being virtualised, results differed massively. This partly being influential on the following work.

In conjunction with this, front and back confusion was commonly reported, mainly being due to the fixed head position used. This also occurred in such arrays as the 5.0 where a point source acting loudspeaker at 0° θ & ϕ still caused approximately 30% of participants to perceive sources from behind them. Another interesting point made, was that front and back confusion was a less reported issue for other systems, apart from the 2nd order 8 speaker array. Implying an increase in ambisonic order could cause more F/B confusion when the head remains in a fixed position.

Lastly a recent study by [Kearney et al] in 2015, assessed height perception in ambisonic based binaural decoding. This also utilising the virtual loudspeaker method for 1st, 3rd & 5th order ambisonic arrays. Comparisons between real world and binaural sources were compared in the vertical plane for spatial resolution/preservation. With mode matching methods used to derive loudspeaker locations in the spherical harmonic domain & psychoacoustic optimisation through dual band decoding.

Within the study, Kearney et al [2015] discusses how the even distribution of loudspeakers over the arrays provide smooth rV in all directions, however not always the case for rE. A comparison is made between a tetrahedron providing perceptual detent effects with non-uniform directions of uV & uE (angles), when in comparison to a cube where both vectors remain harmonious. This supporting Gerzons original findings in 1980 when assessing loudspeaker playback for periphonic sound. Finding the tetrahedron to pull sources towards the loudspeakers (Gerzon, 1980).

This section has gone on to outline the history/application & testing of commonly employed methods for regular surround arrays with ambisonics. These methods are currently being employed with the use of virtual reality content which will be discussed in the next section.

2.3 Virtual Reality

Virtual reality (VR) technology has had a recent rise in popularity due to processing capabilities & online use with 360° content. Audio for virtual reality has always taken a backseat when compared to the requirements of the visual stimulus. Leaving small amounts of room for this audio specific information to be stored/transmitted. Recently the importance of audio as an element of immersion within VR technology has grown. The favoured method for presenting this immersive surround sound the depends on its intended use.

One method being object based audio (OBA), which assigns positional meta-data to each object within the sound scene. This providing a high degree of accuracy for source placement. But when rotating the sound-field, such as that required with head tracking for VR purposes; re-computing positional metadata for each object can become processor intensive. Obviously, this is dependent on the number of objects within the environment. But for dedicated systems is somewhat manageable such as in cinemas & for game engines. These types usually being set up once and maintained (Altman, 2016). OBA is also currently being tested with different compression methods for transfer online, for example with MPEG-H (Shivappa et al, 2016).

In comparison, Scene based audio (SBA) uses Ambisonics; this method recreating a whole scene with more efficiency, making rotation of a sound-field manageable for many devices. When used with VR technology, these lower computational requirements allow higher resolution surround formats to be used, such as higher order ambisonics. The spherical nature also allows the user to always be in the sweet spot when binauralised, obviously lending itself well to the spherical nature of the hearing system.

Online application for this media is growing, with Youtube looking to support 3rd order ambisonic decoding for 360° content online. Although, it seems apparent from currently available forms of VR reality technology, that the presentation of audio with headphones hasn't been standardized. For example, comparing low end mobile phone based headsets (Google VR) that require the user's mobile phone & headphones, to playstation VR & Oculus Rift headsets that come with dedicated head mounted displays (HMD) & headphones (Fig 2.15).

(Google VR, 2017) (Playstation VR, 2017) (Oculus, 2017)

FIGURE 2.15 - VIRTUAL REALITY HEADSETS

Research has indicated that the use of different types of headphone will influence playback within the binaural experiences. For example, frequency response aspects of in-ear monitors providing improved low frequency reproduction, compared to supra-aural headphones (Schulein et al, 2016). Methods such as those presented by Olive et al (2013) propose inverse filtering methods (IFM) for the frequency response of headphones to reduce their influence on HRTF based filter libraries. Issues found within the research did show that participants were not given hearing tests prior to testing, so validity of results cannot be assured. As well as this, this IFM used were only sufficient up to 10 kHz; so, when applied to the full human hearing range is missing a large section of frequency content. Lastly, IFM's would require prior knowledge as to what headphones are being used, so would only apply to dedicated HMD's such as Playstation & Oculus.

It should be mentioned that although the main frontrunners of online VR technology all seem to be using ambisonic based audio processing, decoding methods to binaural for the end user are not well defined. This including the use of different HRTF libraries, orders of material and VLM.

2.3.1 Facebook & Google

Recent developments in the delivery of spatial media online have led to the introduction of software development kits and plug in suites dedicated to making the process accessible to more users. Two key companies currently involved are Google with Youtube Spatial Audio Tools such as Jump (Google - Jump, 2017) & Facebook offering the 360 Workstation (Facebook, 2017).

Current Google (Youtube) platforms allow publication of spherical content using 1st order ambisonic material with the Ambix format (ACN Channel ordering & SN3D normalisation) (Google Youtube 2017). This being decoded using the SADIE (Spatial Audio for Domestic interactive entertainment) Cube HRIR filter set, & dual band decoding. The HRIR filter set currently used was captured using a Neuman KU100 Dummy Head & found to have greater perceptual attributes compared to previously used sets. (UOY(b), 2017) (Fig 2.16).

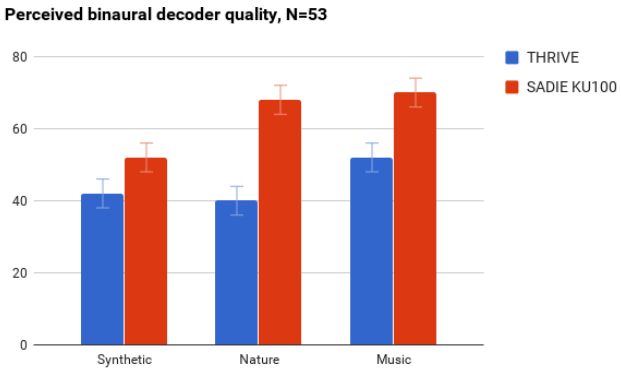


FIGURE 2.16 - SADIE HRTF COMPARISON RESULTS

In comparison, Facebook Spatial Audio Workstation offers a software suite for designing spatial audio. This covering Facebooks .tbe format, Youtube & other ambisonic formats. The supported HMD's for this range from the oculus DK1 & 2 to the HTC vive. Whereas the plugins provided by the workstation offer the option to work with FuMa and Ambix formats in 1st and 2nd order, with visuals such as equirectangular images/videos.

The Facebook .tbe format makes use of 8 channels (Facebook 2017). This due to adhering to popular digital audio workstation (DAW) systems that are restricted to this limit (Pro-Tools). To achieve this, a recent article by Angelo Farina discusses how the .tbe format is essentially 2nd order material with the R (Ambix 6) channel removed, while re-ordering and applying a specific set of gains. This not being as intuitive as the standard ambisonic hierarchal systems used by other software (Farina(a), 2017). Interestingly the virtual loudspeaker array used in decoding for Facebook is not defined in the standard documentation, although deliberation using Eulers characteristic for a spherical array with 8 channels, would mean at least a second order array is being used such as a Dodecahedron or higher.

In summary, the previous chapters have outlined the intricacies of the human hearing system & how ambisonics can be used as a method of attempting to recreate these accurately. This stemming from the early work of Alan Blumlein & Michael Gerzon to the current use of virtual loudspeaker methods, to present virtual reality based audio over the web. The discussed methods and recent work in assessing these decoders will be used in the following chapter to attempt to create a system to allow testing of different virtual arrays while using head tracking.

3 Methodology

3.1 Overview

The research question for this work investigates the use of ambisonic to binaural decoding for VR, and will be completed using VLM. The null hypothesis being, there is no difference between the arrays, in terms of localisation accuracy and perceptual attributes.

As previously discussed, many methods of testing decoders have been implemented in the past for 'real' playback systems. For this testing of virtual arrays, energy and velocity vector analysis will be used initially to gain an understanding of how each array should perform; then a listening test will be completed for each array to assess the real-world implementation. The added benefit of auditioning each over headphones, being that they can be tested in quick succession. The virtual loudspeaker methods outlined in (Wiggins, 2001) & (Kearney et al, 2016) will be used to hard code each loudspeaker HRTF into a B-Format set of impulses for use with head tracking and dynamic binauralisation. This simplifying the processing used to test all arrays.

The following section outlines the creation of each listening test, the content used within them & issues faced during. It begins with an overview of the choice of arrays to test, the Matlab script used to create each set of impulses, & the implementation of a head tracker. The final section outlines the full testing process used and corrections made after a small pilot test was conducted.

3.2 Choice of arrays to test

When considering virtual reality & ambisonics online, the use of decoders takes a slightly different route. That being that because the playback is completed through binaural headphone techniques. Virtual loudspeaker techniques are commonly employed to avoid large HRTF filters sets being required. This meaning that irregular layouts due to geometric constraints do not occur. Whereas Gerzon (1980) makes a point of avoiding certain shapes such as the dodecahedron for this reason, when used in a virtual array these restraints do not apply. In conjunction with this, due to its isotropic nature, the user will always be in the central position. Considering this, the virtual loudspeaker arrays used within this study are regular or diametrically positioned. To which Gerzon made these requirements to satisfy the diametric decoder theorem (Gerzon, 1980):

- All speakers are placed the same distance from the central listening position.
- Speakers are all diametric opposites
- Sum of the feeds to each diametric pair is the same for all pairs.

Although, virtual loudspeaker sampling requirements still stand, and follow that of standard polyhedral shapes. These based on Euler's equations to evenly distribute points over a sphere, resulting in platonic solids. A platonic solid being a 3-dimensional shape in which each face is the same polygon, and the same number of polygons meet at each vertex.

Euler's formula can be used to prove that only 5 platonic solids exist. This being the relationship between the edges, vertices and faces of a shape. When Euler's formula results in an answer of 2, this shape is known as a platonic solid. This formula is not restricted to a result of 2, hence the following equation (Eq.19, MIF, 2017);

$$F + V - E = \chi$$

Where : F = Faces

V = Verticies

E = Edges

χ = Euler Characteristic

EQUATION 19 - EULER CHARACTERISTIC

The 5 basic polyhedra are shown below in Fig 3.1 (Moerman, 2010) :

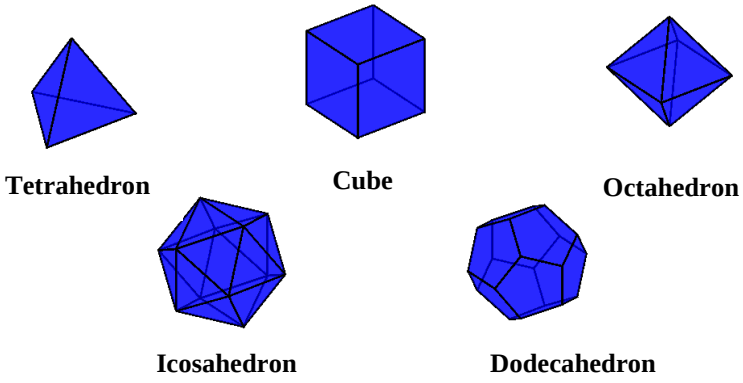


FIGURE 3.1- PLATONIC SOLIDS

For the investigation, 4 arrays were chosen to be tested. These providing a good range of order and sampling schemes.

- 8 Speaker Cube (1st Order)
- 8 Speaker Bi-rectangle (1st Order)
- 12 speaker Dodecahedron (2nd Order)
- 26 point Lebedev grid sphere (3rd Order)

A graphical representation of each is shown below (Fig 3.2), where the central sphere in blue represents the human head position & each smaller sphere is a loudspeaker. A more detailed view of each array can be found in appendix 9.2. These were chosen to assess localisation accuracy over an increase in ambisonic order & different arrangements of loudspeakers. The two 1st order arrays will also test the previously mentioned issue of loudspeakers at ear level (Kearney, 2015). The cube being a representation of an array with no loudspeakers at ear level (0° ϕ), while the bi-rectangle contains 4 loudspeakers at ear level with another four at 90° & 270° (Ear positions) at varying elevations. With the hypothesis that the bi-rectangle will outperform the cube array in terms of localisation accuracy and perceptual preference. This due to the findings in Heller (2006) outlining improved energy vector magnitude at the lateral regions in the rectangular shaped arrays, this being a region of poor localisation for low frequencies as previously discussed.

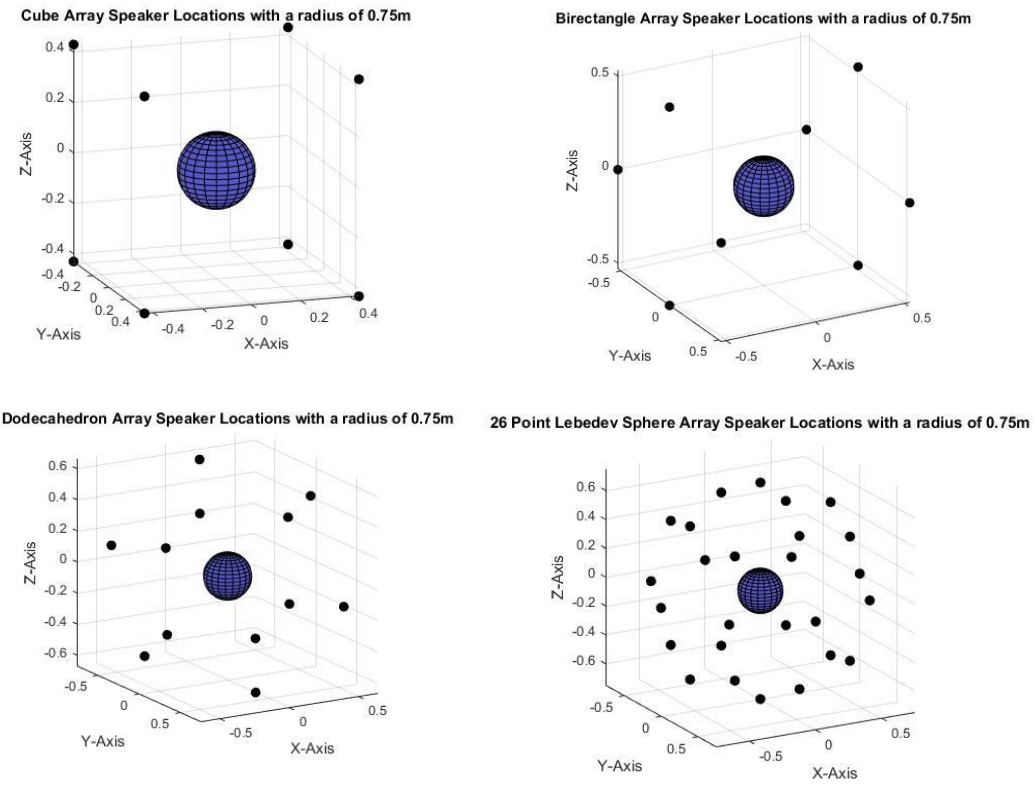


FIGURE 3.2 - CHOSEN ARRAYS

Then finally, the dodecahedron & 26 point lebedev grid arrays will assess the possible increase in localisation accuracy & perceptual attributes with higher order material. The Lebedev grid loudspeaker arrangement derived using quadrature rules to extend spherical harmonic based playback systems for large arrays. (Lecomte et al, 2015).

These arrays were chosen to comply with the loudspeaker positions outlined in the SADIE project (SADIE(a), 2017). This due to the HRTF library being captured in the defined positions for each array at an equal distance (1.5m), rather than having to use interpolation between filters in other sets. The filter set used remained the same throughout testing to ensure validity of results. The chosen set was the diffuse field compensated filters from the Kemar Dummy Head (Subject 001 – 48kHz). Each required position then chosen and used in the Matlab script with its corresponding spherical harmonic.

3.3 Development of the Listening Test

To implement each decoder, Matlab was used, employing the methods outlined in previous chapters such as VLM, & B-Format HRTF simplification. The resulting HRIR filters then being convolved with a given input source (Depending on order & array) in reaper. This allowing the implementation of multiple virtual arrays over headphones for the test. The following section outlines the design of a Matlab script for the HRIR filters, & the build of a headset to allow head tracking during testing.

3.3.1 Matlab Decoder Script

The decoder script generated allowed the creation of B-Format HRTF filters depending on defined array and order of material. This could then be tested within the script using static binauralisation, before use in reaper with head-tracking.

To begin, speaker locations were defined in ° (For simplicity), these then converted to radians for use in the spherical domain. Other functionality variables were set at the top of the script to allow quicker debugging/assessment of multiple arrays.

```
% What Virtual Loudspeaker Array do you want -
% First Ord Cube                  -  1
% First Ord Bi-Rectangle          -  2
% Dodecahedron (2nd ord)          -  3
% 26-Point Lebedev grid (3rd ord) -  4
s.array = 1;
% 1 = SN3D
% 2 = N3D
norm = 1;
% Set Psycho-acoustic Optimisation
psych_opt = 1;
%Playback sound
playback = true;
%Which audio file to play
audiofile = 1; %
% write final file
write_to_file = false;
write_IR_set = false;
%% Loudspeaker Position
if s.array == 1
    s.az = [45 135 225 315 45 135 225 315];
    s.el = [35 35 35 35 -35 -35 -35 -35];
    s.order = 1;
    num_sp = 8;
elseif s.array == 2
    s.az = [90 270 45 135 225 315 90 270];
    s.el = [45 45 0 0 0 0 -45 -45];
    s.order = 1;
    num_sp = 8;
elseif s.array == 3
    s.az = [180 50 310 118 242 0 180 62 298 130 230 0];
    s.el = [63 46 46 16 16 0 0 -16 -16 -46 -46 -63];
    s.order = 2;
    num_sp = 12;
elseif s.array == 4
    s.az = [0 0 90 180 270 45 135 225 315 0 45 90 135 180 225 270 315 45 135 225 315 0 90 180 270 0];
    s.el = [90 45 45 45 45 35 35 35 35 0 0 0 0 0 0 0 0 -35 -35 -35 -35 -45 -45 -45 -45 -90];
    num_sp = 26;
    s.order = 3;
end
% convert the chosen speaker locations to radians
s.az = s.az .*pi/180;
s.el = s.el .*pi/180;
```

3.3.2 Decoder Matrix Generation

A function was then implemented to allow the creation of each spherical harmonic depending on the defined order and its corresponding degree, these cycled through using for loops. Inside the function (Available in appendix 9.3), the inputted order and degree values with speaker locations were used to calculate the speaker positions in the spherical harmonic domain, resulting in a decoder matrix created using mode matching (pseudo inversion) methods. This also providing the option to define normalisation schemes. The chosen implementation for the testing being ACN with SN3D normalisation. An example of the output decoder matrix for the cube array can be found below in Fig 3.3.

```
for n=0:s.order
    for m=-n:n
        tmp = SphHarm(s.az,s.el,n,m,norm);
        s.B(n^2+n+m+1,:) = tmp(:);
    end
end
%Moore-Penrose Pseudo Matrix Inversion
dec = pinv(s.B);
```

Cube Decoder Coefficients =

	0	1	2	3
sp1	0.12500	0.21580	0.21793	0.21580
sp2	0.12500	0.21580	0.21793	-0.21580
sp3	0.12500	-0.21580	0.21793	-0.21580
sp4	0.12500	-0.21580	0.21793	0.21580
sp5	0.12500	0.21580	-0.21793	0.21580
sp6	0.12500	0.21580	-0.21793	-0.21580
sp7	0.12500	-0.21580	-0.21793	-0.21580
sp8	0.12500	-0.21580	-0.21793	0.21580

FIGURE 3.3 - CUBE DECODER MATRIX

3.3.3 B-Format HRTF Generation

To simplify the required number of HRTF filters for each array, a separate function was created to hard-code each loudspeaker position HRTF into the B-Format channels of the decoder matrix. This lowering the required number of filters for each array to a left and right per B-Format channel rather than 2 per loudspeaker. This outputted in a matrix of left and right filters & combined back into stereo impulses for real-time convolution in reaper via the ReaVerb plugin (Reaper, 2017), or binauralisation within the script.

Below is an example of the process for just the cube array. It is understood that this may not be the most efficient way of implementing this process. Although due to the script not being run in real-time, efficiency was not the main priority. Although this could be utilised in future works with possible use of Matlab's audio system toolbox.

To begin, each HRTF filter pair for the array was loaded into Matlab as left and right arrays named 'HRTF_SET_L & HRTF_SET_R'.

```
function [Filters] = BFORM_HRTF(dec,ARRAY_TYPE)
%% This function requires 2 inputs - the generated decoder matrix & the array type
ARRAY = ARRAY_TYPE;
%% Of the 4 choices given at the beginning of the main script choose which
%% HRTF's are being used & define them as L&R arrays.
if ARRAY == 1
    %%%%loadingHRTF's%%%%%%
    [SF1,Fs] = audioread('azi_45_ele_35_DFC.wav');
    [SF2,Fs] = audioread('azi_135_ele_35_DFC.wav');
    [SF3,Fs] = audioread('azi_225_ele_35_DFC.wav');
    [SF4,Fs] = audioread('azi_315_ele_35_DFC.wav');
    [SF5,Fs] = audioread('azi_45_ele_-35_DFC.wav');
    [SF6,Fs] = audioread('azi_135_ele_-35_DFC.wav');
    [SF7,Fs] = audioread('azi_225_ele_-35_DFC.wav');
    [SF8,Fs] = audioread('azi_315_ele_-35_DFC.wav');
    HRTF_SET_L = [SF1(:,1) SF2(:,1) SF3(:,1) SF4(:,1) SF5(:,1) SF6(:,1) SF7(:,1) SF8(:,1)];
    HRTF_SET_R = [SF1(:,2) SF2(:,2) SF3(:,2) SF4(:,2) SF5(:,2) SF6(:,2) SF7(:,2) SF8(:,2)];
    NO_HRTF = 8;
    ambix_channel_no = 4;
end
```

These two arrays were then used in the below function to cycle through each HRTF, with each row of the decoder matrix. Resulting in a final array consisting of left & right filters for each B-Format channel of the decoder matrix.

```
% Loop cycles through the number of B-Format channels,
% For each loop iteration, it multiplies each cell in that channels column    % with its corresponding speaker HRTF (LEFT),
then clears the dec row %variable to do the same for the right channel.
% This resulting matrix is summed so that a 256x2 matrix remains
```

```
% This being the L & R HRTF for that B-Format Channel.
% The process repeats until all channels are completed.
for channel = 1:ambix_channel_no
    for dec_row= 1:NO_HRTF
        Left = dec(dec_row,channel)*HRTF_SET_L(:,dec_row);
        LEFT_out(:,dec_row) = Left;
    end
    clearvars dec_row
    for dec_row= 1:NO_HRTF
        Right = dec(dec_row,channel)*HRTF_SET_R(:,dec_row);
        RIGHT_out(:,dec_row) = Right;
    end
    clearvars dec_row
    Left_Ambix = sum(LEFT_out,2);
    Right_Ambix = sum(RIGHT_out,2);
    FINISHED_LEFT(:,channel) = Left_Ambix;
    FINISHED_RIGHT(:,channel) = Right_Ambix;
end
%% The output of each channel then returned in a final matrix of L&R for all
%% B-Format Channels.
[Filters] = [FINISHED_LEFT,FINISHED_RIGHT];
```

An example of the resulting B-Format filters used for convolution in the test can be found below in Fig 3.4. While the full set for each array can be found in appendix 9.4. This being an example of velocity based decoding, note the inversion of the left impulse for the Y channel (Ambix 1) allowing correct panning with head-tracking from left to right.

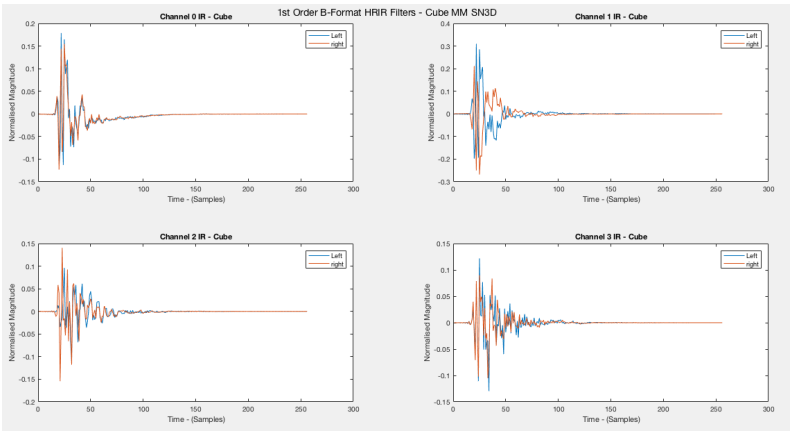


FIGURE 3.4 - CUBE B-FORMAT IMPULSE RESPONSE SET

Once the filters had been created, a static binauralisation function was written to allow the auditioning of each array within Matlab, shown below. This to speed up the debugging process before moving over to Reaper & applying the head rotation element. This allowed test material to be created, such as a source panning around the array from 0-360° to ensure each array was performing properly. Many issues were found at this stage due to bugs in code or incorrect use of channel conventions.

```
function [Binauralize] = BINFUNC(HRTF_AMBIX,ORDER,AUDIO_IN)
%% This function requires 3 inputs - the Resulting B-Format Channel HRTFs from the previous function
%% The order of material being used & the audio input to be used.
ch0 = (AUDIO_IN(:,1));
ch1 = (AUDIO_IN(:,2));
ch2 = (AUDIO_IN(:,3));
ch3 = (AUDIO_IN(:,4));
%% complete convolution for each order (all B-Format channels L conv with channel 1-4, same for right)
if ORDER == 1
    HRTF0 = [HRTF_AMBIX(:,1) HRTF_AMBIX(:,5)];
    HRTF1 = [HRTF_AMBIX(:,2) HRTF_AMBIX(:,6)];
    HRTF2 = [HRTF_AMBIX(:,3) HRTF_AMBIX(:,7)];
    HRTF3 = [HRTF_AMBIX(:,4) HRTF_AMBIX(:,8)];
    MAIN_LEFT = (conv(HRTF0(:,1),ch0)) + (conv(HRTF1(:,1),ch1)) + (conv(HRTF2(:,1),ch2)) + (conv(HRTF3(:,1),ch3));
    MAIN_RIGHT = (conv(HRTF0(:,2),ch0)) + (conv(HRTF1(:,2),ch1)) + (conv(HRTF2(:,2),ch2)) + (conv(HRTF3(:,2),ch3));
end
%% Returns a final L&R HRTF convolved with the chosen Audio file.
```

```
[Binauralize] = [MAIN_LEFT,MAIN_RIGHT];
```

Once all array impulses were generated, these were imputed to reaper using the ReaVerb plugin for real-time convolution within the listening test. The filters were used in this way to avoid having to implement rotational matrixing within Matlab, & the conversion of the script to an object orientated based code. Although as mentioned, future work would allow this. The full code being available on the attached disc, with examples of generated Impulses & Test Material located in Appendix 9.3.

This section has given an overview of the creation of the impulse responses for each virtual loudspeaker array and how they can be used with static binauralisatin. The following section goes on to discuss the use of a headphone based head-tracker to allow dynamic binauralisation. That being the inclusion of a participant's orientation within the virtual array.

3.4 Head Tracker Implementation

As discussed in earlier chapters; the localisation cues used in real life are frequency dependant in terms of time and level differences. One common issue with these functions being the cone of confusion. To overcome this with our binaural system, we mainly use head rotation. Although the influence of monaural cues (4kHz upward) with pinnae filtering and shoulder reflections can also aid dismissing ambiguity in the elevation plane (Stitt et al, 2016). The application of head orientation data has recently become possible using gyroscopic data from either phones or dedicated headsets. But has been assessed using rotating dummy heads since the 90's, Horbach (1999).

The use of this data allows the rotation of the sound field to be directly related to the movement of a user. When used in the ambisonic domain, it is possible to invert this rotation to allow the perception that the sound-field remains motionless and the user rotates within the scene. The application of these methods with binaural technology allows the user to always be in the sweet spot, but requires them to remain relatively stationary. This meaning for continuity purposes the scene should remain fixed. Recent advancements using Microsoft Kinect sensors for optical depth information, gives 6 degrees of freedom (6DOF) during use. Allowing users to explore the scene and move around within it, improving perceived externalisation of sound (Stitt et al, 2016).

The employed tests implemented head-tracking information using gyroscopic based sensors, similar to those found in mobile phones & low budget headsets. This being the most accessible form of VR technology available, but with current advancements in mobile technology it won't be long before 6DOF is available to the wider consumer base (Altman, 2016).

The plug-in used for rotation in the completed listening test was the Ambix Rotator plug in, provided by Matthias Kronlachner. As part of his full plug-in suite designed to be used across multiple platforms (Miscrosof, Mac, Linux) with C++ & JUCE. The formats followed are that of current conventions – (ACN & SN3D). While also providing a conversion plug in to mix between other schemes. (Kronlachner(a), 2017).

Kronlachner outlines the process to be followed with his plug-ins to be (Fig 3.5):

FIGURE 3.5 - DYNAMIC BINAURALISATION

Ensuring that the ACN & SN3D conventions are followed to avoid incorrect playback. These plugins available up to 7th order, in line with reapers 64 track limit and the 3D channel convention $(N+1)^2$ (Kronlachner(b), 2017).

A recent study into the use of head-tracking for binaural (Stitt et al, 2016), made some key findings for its application in terms of latency & externalisation of sound. An assessment was made into the perceivable levels of latency & its effects. Previous work by the authors had noted smearing due to latency causing movement of the source from its intended position. Whereas previous work by Horbach (1999) found latency below 85ms not to be perceived by participants in listening tests. Assessment into source stability found weak correlations between the speed of head movement and accuracy of detection ($r = 0.47$), but interestingly did find a difference between the speed of head movement during trials and the use of either a single source or a complex scene made up of 6 sources. This being positively correlated to an increase in complexity, meaning an increase in speed used for detection. This was independent of the amount of latency used within the tests and proved consistent with an increased number of trials. (Stitt et al, 2016)

A further investigation into the influence of head tracking on externalisation of sound assessed 4 conditions.

- Static HT (No head movement)
- No head movement (Head tracked)
- Head movement (No tracker)
- Head movement & head tracker.

Multiple HRTF sets were used to avoid a learning processes during the tests which could negatively influence results. Participants were then asked to only evaluate the externalisation after a short period of time and when they were still. This allowing (when applicable) head movements to be used but only initially, while evaluations were held under the same static condition. Due to externalisation changing depending on movement. The results showed that as expected, the use of head tracking and movement (HT/HM) performed best. Although the HT/HM condition showed only small improvement compared to static performance. It was also noted that externalisation of sources positioned at 0° θ showed poor results, while the addition of head tracking & small micro movements improved this by approximately 36%. (Stitt et al, 2016)

To allow the use of head tracking with each virtual array for the employed testing, an existing system was implemented called Hedrot. Hedrot is an open source software/hardware solution that has been specifically designed for binaural rendering. This being the main reason it was chosen for use in testing, while also being a fair price. The hardware provides information regarding the orientation of the chip, which when situated on top of a set of headphones, becomes the users head orientation. This used in conjunction with the freely available Hedrot Reciever Max patch, allows transmission of pitch, yaw & roll (Or quaternion) data over open sound control (OSC).

The board itself is made up of widely available parts shown below (Fig 3.6/3.7) Hedrot,2017)) –

- Teensy 3.2 Board
- GY85 (Daughter board) Housing -
  - ADXL345 Accelerometer.
  - Honeywell HMC5883L Magnometer
  - Invensense ITG – 3200 Gyroscope

FIGURE 3.6 - COMBINED BOARDS

Figure 3.7 - Final Headset

To allow the processing of the orientation data, an open-source gradient descent algorithm from Sebastian Madgwick, is used in the Max Patch. The algorithm relies on a beta coefficient to determine the use of the magnometer and gyroscope/accelerometer during movement of the sensors. Optimal values of beta range from between 0.1-0.5 according to Sebastian. Although the receiver application has implemented a variable beta value depending on use. Beta is low when stationary & high when in use. This being defined in the settings as MAX & Variation beta variables, with recommended settings being (Max Beta = 2.5, Beta Variation = 1), these used within the test.

Calibration of the hedrot receiver could take place once the previous steps had been completed. This being done in the hedrot receiver application, allowing individual calibration of each sensor. This was checked before each test, while ensuring the gyroscope remained at 0° when still. Each new participant when wearing the headset was then asked to remain still while quick calibration was completed and the 'centre' function was used to set the users position as a reference for 0° pitch/yaw/roll. Sample rates in the settings of the receiver are set as a basic level of 1kHz but were increased to 2kHz to improve latency during testing.

Once built and calibrated the pitch/yaw/roll information was sent to Reaper via OSC information. The setup was comprised of;

Axes references – X = Front Y = Left Z = Up

Rotation Order – Yaw/Pitch/Roll (Z/Y/X)

Rotation = Inverted

OSC Settings –

Yaw = Inverted (Scale – 0-1)

Pitch = Inverted (Scale 0.25/0.75)

Roll = Not inverted (Scaled – 0-1)

A new control surface was then created in reaper with the same port number as the one transmitted from the Hedrot patch, with binding selected. The learn FX parameter process followed for each of the three planes. This allowed the head tracking to be applied in real time during testing (Fig 3.8).

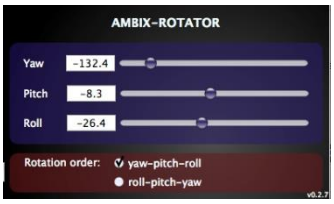


FIGURE 3.8 - AMBIX ROTATOR PLUG-IN

At this point it was possible to allow a user to experience each virtual loudspeaker array in reaper. When wearing the headset, the addition of sound-field rotation provided a more natural experience of the virtual array. This could then be used within the listening test to assess the influence of head movement when testing these decoders. The following section discusses the listening tests used and the creation of content within them. This then leading into findings from a small pilot study & the final implementation of the test.

3.5 Content creation

As previously mentioned, the chosen methodology for the listening test was based on previous work in the assessment of ambisonic decoders. The chosen tests being a localisation test & a perceptual test. The content for both requiring preparation of B-Format audio for each array. These outlined in the following section.

To create the test material, 3rd order 'Ambix' encoders were used in Reaper, to pan a source in different positions on the azimuthal plane. Utilising the hierarchal nature of B-Format material, by removing latter channels when not required for lower order arrays. These files were then saved as 48kHz 16Bit .wav files and numbered according to their position & type.

Band limited pink noise was used in ranges at 0-1kHz, 1kHz-4kHz, 4kHz – 20kHz, similar to methods used in (Wiggins, 2001) & (Heller, 2006). This generated in audacity & inputted to Matlab for filtering. Examples of the filters used can be found below in Fig 3.9.

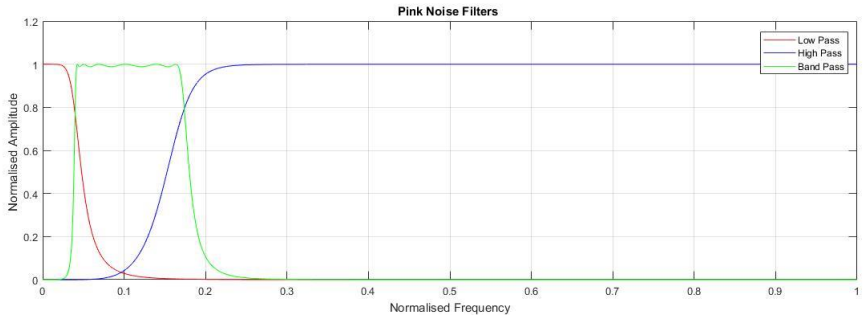


FIGURE 3.9 - FILTERS FOR PINK NOISE

This to cover the frequency ranges discussed in previous chapters concerning interaural cues used for lateralisation & perception of a sound source. Pink noise was used due to having an equal amount of acoustical energy in each octave. Speech was then used, as it is easily localised by the human cognitive system; & sits within the range that humans are most sensitive to (4kHz) according to fletcher munson curves (Ballou, 2015) (Fig 3.10). The clip chosen was an anechoic male voice reading an Archimedes script, (Farina (b), 2017).

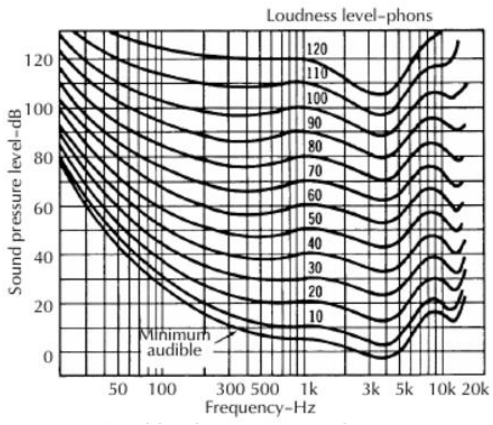


FIGURE 3.10 - FLETCHER MUNSON CURVES

This resulting in 4 stimuli for the localisation test.

- Speech (Archimedes – BBC Anechoic Clip)
- Low Frequency Filtered Pink Noise (N1)
- Mid Frequency Band Limited Pink Noise (N2)
- High Frequency Filtered Pink Noise (N3)

Localisation was assessed over just the azimuthal plane, as including elevation would increase complexity for participant answer capture & assessment of results, but would be of interest for future work. The chosen source locations were as follows:

- 0° - 30° - 60° - 90° - 120° - 150° - 180°

These were kept to one hemisphere due to the use of regular and diametrically paired arrays; so, symmetry between both sides was expected. These locations would allow assessment of the frontal plane, front/back confusion & lateral issues found within binaural listening. This resulted in many conditions when used with multiple arrays, so it was chosen to only use 3 of the overall 4 arrays during the localisation test. These being the Cube/Birectangle & the Lebedev Sphere. Allowing a comparison of 1st order sampling methods for possible detent effects, & possible differences due to an increase in order of material and loudspeaker sampling. Resulting in 89 conditions (Due to repetition of the first 5 clips). These were arranged in a reaper project with markers to allow quick navigation. To randomise the clips, all were assigned values 1-84, inputted to Matlab and randomised with the 'rand' function. This final order was then inputted to reaper for use, then results re-ordered into individual arrays in excel for analysis in Matlab.

Each clip was 5 seconds long; vocals faded in/out & noise 4x1 second pulses with 0.25second gaps. The resulting length of each test was ~7.41 minutes; but as later discussed, in the pilot test participants would regularly want to hear each sample again. So, allowing repetition of each clip resulted in ~15minutes per test. With a break to reduce listening fatigue added, resulting in this part of the test being ~20minutes. A length originally considered to be ok, due to comparisons with work by Lopez et al (2016), but as will be discussed may not have been optimal.

The perceptual listening test stimulus consisted of 2 musical extracts. These provided by MH Acoustics online as demo material (mhacoustics(b), 2017). Two different types were used, the 1st being a large living room with a central guitar player (Simple), the 2nd being a full band scene with the listening position surrounded by each instrument (Complex). Both extracts were Eigenbeam/Ambisonic, 16 Channel .wav files (3rd order, ACN, SN3D). These were chosen to gain an understanding of two types of use for VR technology, these being differences in preference with simple and complex sound-fields as assessed in (Stitt et al, 2016) for speed of localisation. The order of playback being randomised, but the same order for all participants.

1. Cube 2. Dodecahedron 3. Birectangle 4. Lebedev Grid

With participants rating each clip in terms of

1. Width 2. Depth 3. Immersion

These categories based on previous studies by Rumsey & Mason (2000), assessing the perceptual attributes of multiple surround sound systems using a likert scaling system from 0-10 (Lancaster, 2005). Due to the use of the same material, participants heard a 20 second clip of each stimulus for each array. Resulting in ~3-5 minutes to complete this section of the test.

A final piece of content creation involved the use of a separate piece of software. This being a stereo to B-format converter provided by Haydon Cardew (Cardew, 2017). It allowed the up-mixing of stereo content to 3rd order B-Format. A width of 300 was used when completing the up-mixing, as recommended by Haydon himself. The chosen clip being a segment of a popular soundtrack from the film Intersteller – Cornfield Chase (Amazon, 2017). Once the 3rd order file was created, this could be implemented and assessed in the same fashion as the previous perceptual test, to assess participant's preference over stereo or VLM ambisonics with headphones.

Once this content was created and arranged in reaper, a small pilot study was conducted to assess the overall methodology, equipment & findings. These discussed in the following section.

3.6 Pilot Study

A small pilot study was conducted to assess the testing process itself. During which the apparatus used were tested, programs and equipment were calibrated & small errors found. For example, it was noted that the originally proposed method of using a rotating chair for participants to locate each sound source, would only assess their ability to localise 0° in different positions within the virtual arrays. This meaning that their ability to localise sound around them was not being assessed. This was changed to participants being required to keep their back flat against the chair throughout the testing. Small head movements to aid localisation were allowed, but participants were asked to return their head to the centre before answering. It was also noted that the magnometer used in the head rotation software was susceptible to magnetic field influence given off by testing equipment. During the final tests participants were asked to return to 0° to reset. Centring the device every 20 conditions proved acceptable to retain accuracy. This centring possible with just a click of a button so did not impede the testing process itself.

Furthermore, due to the large number of tracks used in reaper to complete assessment, small routing issues were found and rectified before initial tests took place. This possibly not being an issue if the original Matlab script was implemented in a standalone plug in with the head-tracking incorporated.

Originally the localisation test was intended to be used over just two arrays. This was in the interest of keeping the testing time below 30minutes per participant (including consent/debrief), when completing the initial pilot test, it was found that participants were faster at locating sound sources than expected. This being the reason for the introduction of a 3rd array for assessment.

Once this pilot test was completed and changes made to employed methodologies and equipment, the final testing could be completed. An overview of which is given below.

3.7 Final Implementation

To begin, energy and velocity vector analysis of each array was completed, to gain an understanding of how each array should perform. Thereafter the final listening test procedure was as follows:

1. Consent form completed
2. Training Period Given
3. Localisation test (Including Break)
4. Perceptual Tests
5. Debriefing.

The tests were looking to assess 5 Null Hypothesis.

1. There is no difference in localisation accuracy due to different virtual arrays.
2. There is no difference in localisation accuracy due to frequency content.
3. There is no difference in perceptual attributes due to different arrays.
4. There is not difference in perceptual attributes due to sound field complexity.
5. There is no difference in stereo playback over headphones compared to upmixed stereo with VLM & head-tracking.

With some of the key variables during testing being shown below in Table-1:

Independent Variables	Dependant Variables.
The source used – Speech, Noise & Music	Perceived location of source
The position of the source.	Perceptual ratings of arrays
The virtual array used.	Favoured array for stereo and VLM methods
Equipment/calibration	
Room used for assessment.	

TABLE 1 - LISTENING TEST VARIABLES

• Mac mini (Sierra 10.12.6 – 2.3Ghz Intel Quad i7 – 4Gb Ram1600MHz DDR3)
• Focusrite 2i2 A/I
• Phillips SHP2500 Headphones
• HedRot Hardware/Software
• Matlab/Reaper
• USB a/b cable (6m)
• Chair/Desk/Monitor/Mouse/Keyboard
• Consent/Debrief/Questionnaire Forms

TABLE 2 - EQUIPMENT LIST

The equipment used during is outlined above in Table.2, & was completed in MS017 of Markeaton street campus at Derby University, as shown in Fig 3.11.

FIGURE 3.11 - FINAL TESTING SETUP

In terms of choice of participants, constraints were not put in place. Due to the intended use of the technology being open to a large range of the population.

A training period was introduced after the small pilot test was completed. This due to participants possibly having no prior experience with the technology. It consisted of anechoic vocal samples at 0°,90°,180°,270° θ, then a full 360° pan on the azimuth plane. This allowed participants to get used to the process required for the localisation test, prior to capture of results, to improve validity. Once completed, the localisation test was conducted with a break mid-way through at the 40th condition (~10 minutes) to avoid listening fatigue. This time also taken for participants to give feedback about the implementation of the device itself. The first 5 conditions of the test were repeated at the end and results from the original 5 removed, to allow for an unconscious training period within the main test.

Participants were given a polar chart to use as a reference during the test, which was positioned at head height to remove the need for participants to look down to give an answer. This also meant that participants would require their eyesight to complete the test, so blindfolding was not used. Although, it could be argued that in comparison with virtual reality technology, the visual stimulus provided may not always coincide with the audio presented to user.

Once completed, participants took a quick break to remove the head set while preparing for the next test. Then perceptual evaluations were completed with participants seated in the same position, but rotation restraints were not as strict. This due to the desire for participants to explore the sound field before answering. Participants were then given a brief explanation of what each rating based scale was measuring, and played and example as another training period before completing the test.

This was followed up by a final assessment of the stereo & B-Format comparisons. This played over each array, with a hidden anchor of the stereophonic file itself. With participants able to listen to each as many times as they liked and write down their favoured playback type.

This chapter has outlined the creation of B-Format impulse responses in Matlab for each virtual loudspeaker array. It has then gone on to discuss the use of a head-tracking system to allow dynamic binauralisation within the listening test. This listening test design and implementation was then discussed to allow an understanding of its implementation. The following chapter presents the findings from this investigation and goes on to discuss possible evaluations.

4 Analysis

In total, 16 participants were used for the test. Each one taking the same set of conditions in the same order (although randomised for all). Of which the average age was 32.25 (Range – 22-63), with 81.25% male & 18.75% female. The room used for testing had an average noise floor of 31dBa, this was taken as a precaution even though using headphones throughout.

5.1 Energy & Velocity Vector Analysis

Energy and velocity vector analysis was completed for each array, showing the maximum values of each in the table below (Table 3). As shown, the velocity vector values equate to 1 for all arrays, as expected from this method of decoding, while energy vectors slowly increase with a rise in order of array.

Array	Order	Max V	Max E
Cube	1	1	0.50244
Birectangle	1	1	0.56904
Dodecahedron	2	1	0.66398
Lebedev Script	3	1	0.73341

Table 3 - Max E/V Magnitude Results

Below (Fig.4.1) is an example of the plotted E/V analysis (Birdseye view – Mid Right = 0º) of a B-Format source panned around each array on the azimuthal plane. As can be seen, the vector angles are recreated accurately when assessed over 4 positions on the unit circle. When looking at the energy vector magnitude for the birectangle, an increase in the lateral regions can be seen. This also seen in the Lebedev sphere, beginning to warp in each corner.

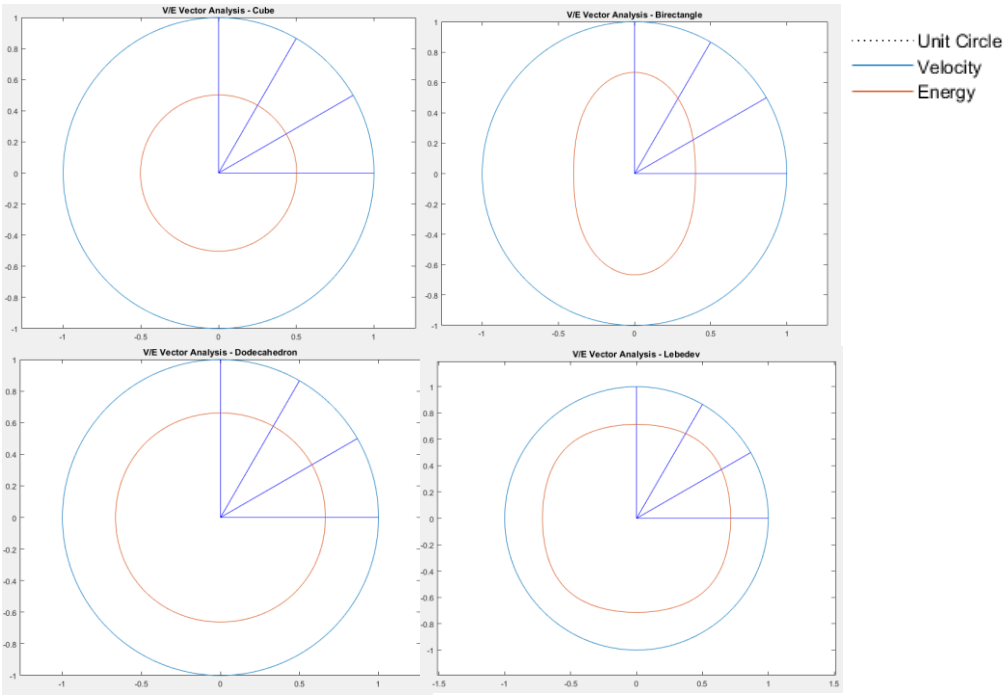


FIGURE 4.1 - E/V VECTOR ANALYSIS

Using these results as a baseline to how each decoder should perform, the listening test was conducted to assess real-world implementation. Localisation tests and perceptual questionnaire results are presented below.

5.2 Localisation Analysis

Once all results were captured from the localisation test, a mixture of MANOVA, ANOVA & T-Tests were completed to assess accuracy of localisation for angle, array & stimulus. This to indicate any differences within the arrays tested. To overcome large variance issues due to the scale used and high amounts of front/back confusion, values that were >180º over the angle being assessed were inverted, essentially taking the smaller value of variance. All tests were conducted in a repeated measures manner, with a confidence level of 95% ($p<0.05$) & sphericity assumed unless otherwise stated. When conducting post hoc testing after ANOVA analysis, Bonferroni corrections were used.

5.2.1 Multivariate Analysis

To being, a multivariate analysis was completed for the accuracy of localised angle for each stimulus & each array; to test the null hypothesis that each array's performance was the same over each frequency range for all tested angles. This chosen to reduce type 1 error from multiple anova tests. No statistical significance was found when assessing each stimulus type, as can be seen in Table 4. Even when using a more conservative form of multivariate test, such as Roy's largest root, significance was still not found. For example (N3) (RLR $p = 0.068$).

N1	**F**	**P**
Pillais Trace	F(12,82) = 0.855	0.59
Wilks' Lambda	F(12,80) = 0.843	0.606
Hotelling's Trace	F(12,78) = 0.830	0.619
Roy's Largest Root	F(6,41) = 1.24	0.306
N2	**F**	**P**
Pillais Trace	F(12,82) = 1.36	0.202
Wilks' Lambda	F(12,80) = 1.35	0.205
Hotelling's Trace	F(12,78) = 1.35	0.208
Roy's Largest Root	F(6,41) = 2.12	0.071
N3	**F**	**P**
Pillais Trace	F(12,82) = 1.056	0.408
Wilks' Lambda	F(12,80) = 1.088	0.382
Hotelling's Trace	F(12,78) = 1.117	0.359
Roy's Largest Root	F(6,41) = 2.148	0.068
Speech	**F**	**P**
Pillais Trace	F(12,82) = 1.031	0.429
Wilks' Lambda	F(12,80) = 1.021	0.438
Hotelling's Trace	F(12,78) = 1.010	0.448
Roy's Largest Root	F(6,41) = 1.548	0.187

a. Design: Intercept + Array

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 4 – Individual Stimulus MANOVA Results

The lack of statistical significance for the per-stimulus Manova, supports the null hypothesis that there was no difference in accuracy of localisation for all 4 arrays at different frequency ranges.

A further manova was completed, to assess the null hypothesis that localisation accuracy will be the same for all conditions over all stimulus types; ultimately looking for the best performing array. When considering Wilks Lamdba ($F(36,145) = 1.372$, $p = 0.09$), no significance was found, see Table 5. Although when using Roy's largest root multivariate analysis ($F(12,51) = 2.597$, $p = 0.009$), some is evident. Suggesting conflicting results for the null hypothesis, that could be due to lack of participants. Further multiple comparisons were conducted (LSD) with results of interest between N1 & N3 ($MD=19.43$, $p = 0.001$), & N1 & Speech ($MD = 11.56$, $p = 0.037$). Suggesting a preference for lower frequency content over higher frequencies and speech, in terms of accuracy of localisation for all arrays. This supporting the use of the velocity decode, optimising only lower frequency content.

Array*Angle*Condition	F	P
Pillais Trace	F(36,153) = 1.363	0.102
Wilks' Lambda	F(36,145) = 1.372	0.099
Hotelling's Trace	F(36,143) = 1.378	0.096
Roy's Largest Root	F(12,51) = 2.597	0.009

Table 5 - Overall MANOVA Results

5.2.2 Individual Analysis

5.2.3 Cube & Birectangle Comparison

The previously discussed research highlighted a key difference between the Cube and Birectangle arrays. That being the arrangement of the array around the central listening position. To test what difference this made to the localisation accuracy of participants, T-tests were completed for just the two arrays. To begin, assessment into front back and lateral positions were assessed over each frequency range, with results showing a positive correlation between improved localisation with an increase of frequency, (refer to Appendix 9.7.1.2).

5.2.3.1 N1

A paired samples t-test was conducted to compare the accuracy at each angle for low frequency content for the two arrays. There was a significant difference in the accuracy of localisation at 60° with cube ($M=78.43$, $SD=53.53$) & the birectangle ($M=99.06$, $SD = 34.74$) conditions; $t(15) = -2.074$, $p = 0.05$). These results indicate rejection of the null hypothesis for no difference between arrays at this angle for low frequency noise. In this case being in favour of the cube, see Fig 4.10.

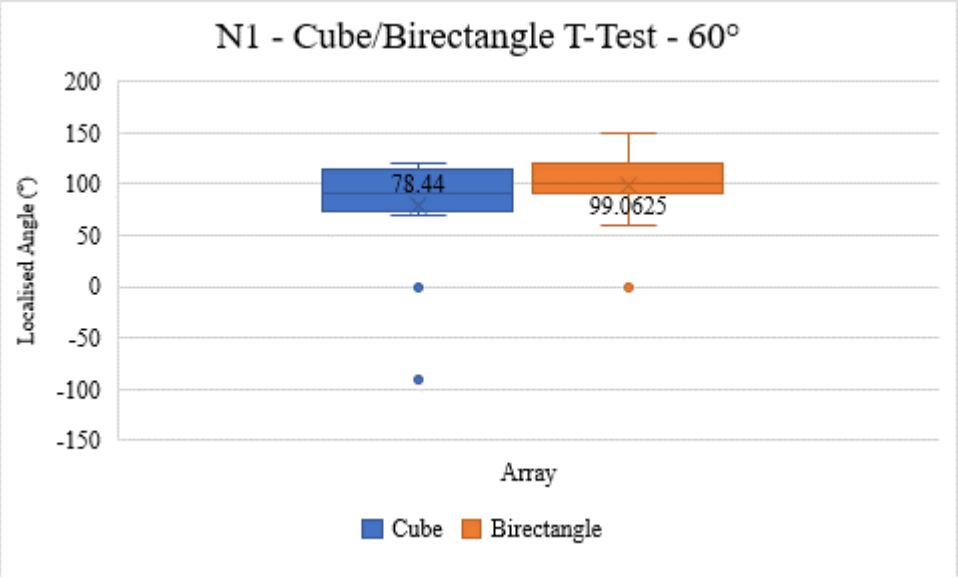


FIGURE 4.10 - N1 C/B 60° LOCALISATION ACCURACY

In conjunction with this, significant differences were found between the arrays for 150°, with cube (M=156.56, SD=37.67) & birectangle (M=130.93, SD = 37.60); t(15) = 2.884, p = 0.01. This once again in favour of the cube, see Fig 4.11.

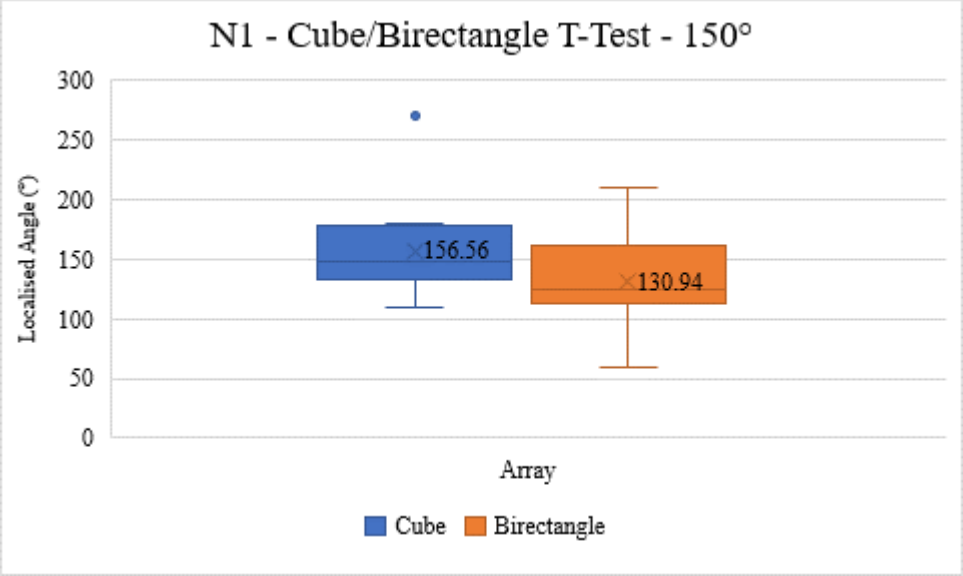


FIGURE 4.11 - N1 C/B 150° LOCALISATION ACCURACY

5.2.3.2 N2

Paired samples t-test were then conducted to compare the localisation accuracy per angle, for mid-frequencies. Results showed one significant difference at 150°, with the cube (M= 126.56, SD = 30.37) & birectangle (M= 140.93, SD= 23.67); t(15) = -2.211, p = 0.043. This rejecting the null hypothesis of localisation accuracy being the same, in favour of the bi-rectangle, see Fig 4.12.

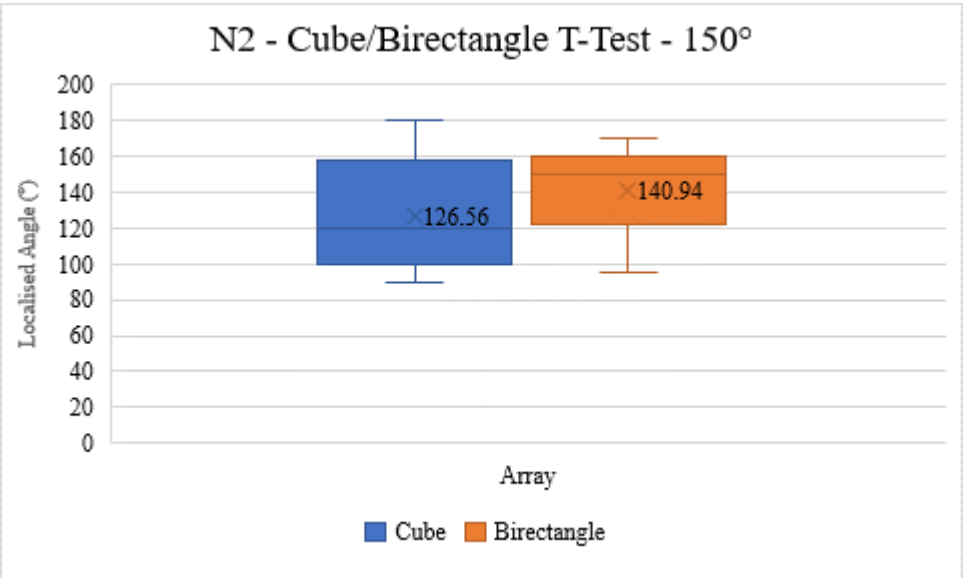


FIGURE 4.12 - N2 C/B 150° LOCALISATION ACCURACY

5.2.3.3 N3

Results for the t-tests at higher frequencies found only one significant difference at 30°, with the cube (M= 65.62, SD= 43.69) & the birectangle (M=44.37, SD= 32.24); t(15) = 2.124, p = 0.05), see Fig 4.13. Again, rejecting the null hypothesis of no difference between localisation accuracy over the two arrays for higher frequency content, in this case supporting the birectangle's accuracy for this condition.

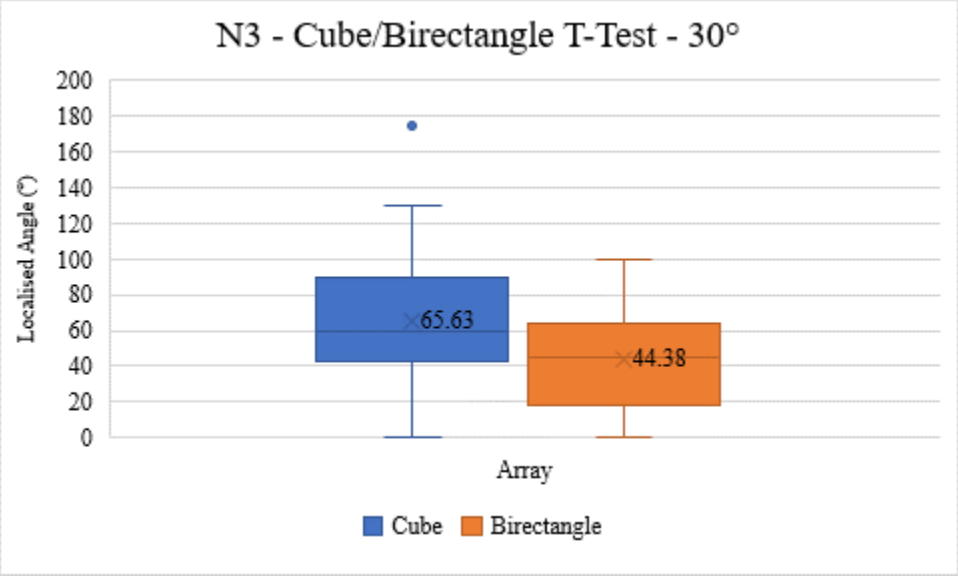


FIGURE 4.13 - N3 C/B LOCALISATION ACCURACY

5.2.3.4 Speech

Lastly, t-test results for the speech condition showed significant difference between the arrays at 60° θ, Cube (M= 107.18, SD=27.68) & Birectangle (M=90.31, SD= 16.17); t(15) = 2.251, p = 0.04). This in favour of the birectangle outperforming the cube for this condition, see Fig 4.14.

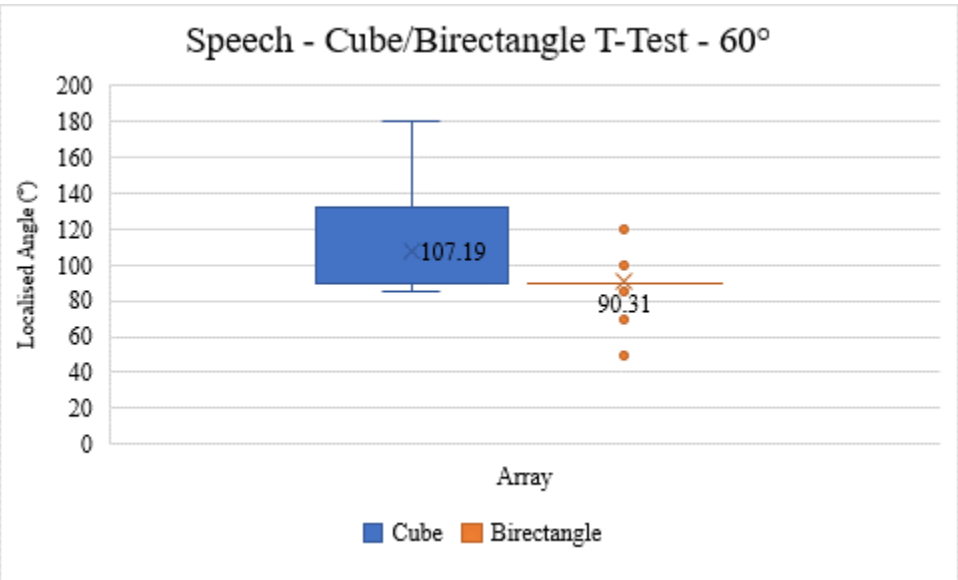


FIGURE 4.14 - SPEECH C/B 60° LOCALISATION ACCURACY

5.3 Perceptual Analysis

For analysis of the questionnaire results, a repeated measures T-Test & ANOVA was completed. The ANOVA to assess the following null hypothesis for each track, (Track 1 – Single Source, Track 2 – Complex).

- The perceived Immersion over all arrays is the same
- The perceived depth over all arrays is the same
- The perceived width over all arrays is the same

Then a further T-Test completed to assess the following null hypothesis for each array, (Cube, Birectangle, Dodecahedron, Lebedev).

- The perceived immersion in each track is the same regardless of sound-field complexity.
- The perceived width in each track is the same regardless of sound-field complexity.
- The perceived depth in each track is the same regardless of sound-field complexity.

The analysis completed provided a within subject's factor analysis for the individual categories in each track. When assessing the results for track 1 (Single Source), no statistically significant results were found, although the Lebedev sphere array was seen to outperform in every category in terms of overall mean, while the dodecahedron was the lowest performing array (Refer to results in appendix 9.7.2), but would require further research.

Following on to the analysis of track 2 (Complex), a statistical significance was found in overall perceived depth between arrays, ($F(3,45)= 2.817$, $P = 0.05$). This indicating rejection of the null hypothesis that the mean perceived depth for all arrays tested was the same. With the Lebedev Sphere scoring highest (M=7.81). This showing that the increase in order of material/array used, can increase perception of depth within a complex sound-field.

Further analysis using repeated measures T-Tests, assessed the difference between the single source & complex sound-fields further. The results showed statistical significance for perception of depth over 3 of the 4 arrays (Shown in Table 6), thus rejection of the null hypothesis of depth perception variance for 3 of the arrays. These being the Bi-rectangle, Dodecahedron & Lebedev, see Fig 4.15.

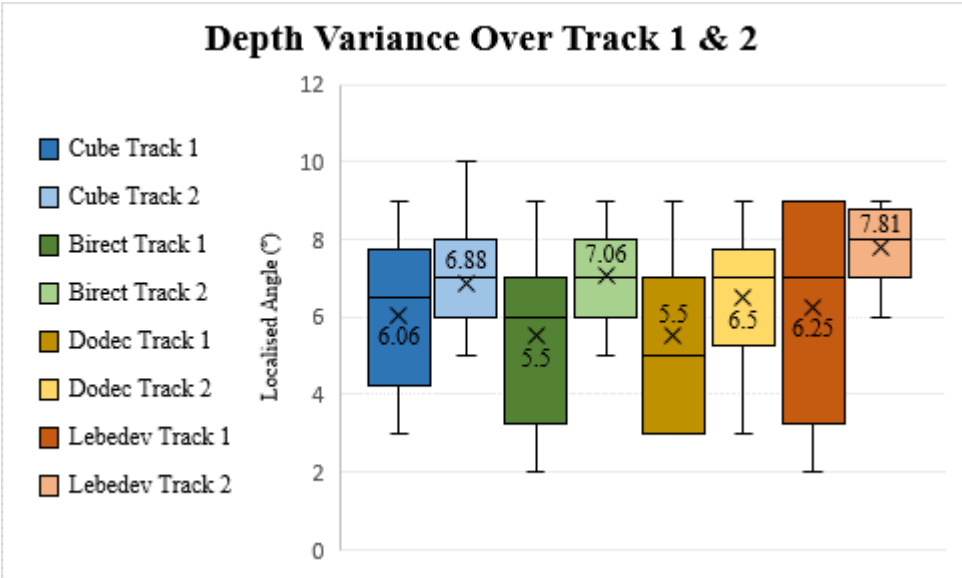


FIGURE 4.15 - DEPTH VARIANCE TRACK 1 & 2

Array	Stats	Track 1	Track 2
Birectangle	(T = -2.581, df =15, P = 0.02)	M = 5.5	M = 7.06
Dodecahedron	(T = -2.070, df = 15, P = 0.05)	M = 5.5	M = 6.5
Lebedev	(T = -2.894, df = 15, P = 0.01)	M = 6.25	M = 7.81

Table 6 - Depth Perception Results

These results indicating that the perceived depth of the sound-field increased significantly due to an increase in complexity. Interestingly, there was a statistically significant improvement in depth perception for the Birectangle 1st order array, as opposed to the Cube which showed none. Even though they both share the same order of ambisonic material and number of loudspeakers. This indicating that the arrangement of virtual loudspeakers around the central sweet spot, can have an influence on perceived depth.

Furthermore, another significant result was also seen in terms of perceived immersion for the Lebedev array. With Track 1(M = 5.37), Track 2 (M = 6.43), (t = -2.221, df = 15, P = 0.042). This suggesting the rejection of the null hypothesis that perceived immersion is the same regardless of complexity of sound-field when using at least 3rd order arrays/material.

Concluding the perceptual analysis, in track 2, the perception of depth results rejected the null hypothesis. This showing that the means across each array were not the same, and that the Lebedev sphere performed best. Although it must be noted that in terms of overall means, the dodecahedron performed worst for both tracks over all perceptual attributes. Indicating that it may not be an increase in order that causes this, but the arrangement of loudspeakers within each order.

4.4 Stereo to B-Format

The results from the stereo to B-Format test are shown in Fig 4.16. This quantifying the number of votes for preference of array. It clearly shows a general preference for listening to stereo up-mixed audio when played back over a virtual 1st order Bi-rectangle array, with 50% of all participants favouring this option. A close second with 25% of results was the original stereo version, while interestingly the 3rd order Lebedev grid array had no votes although seeming to outperform all others in the previous tests.

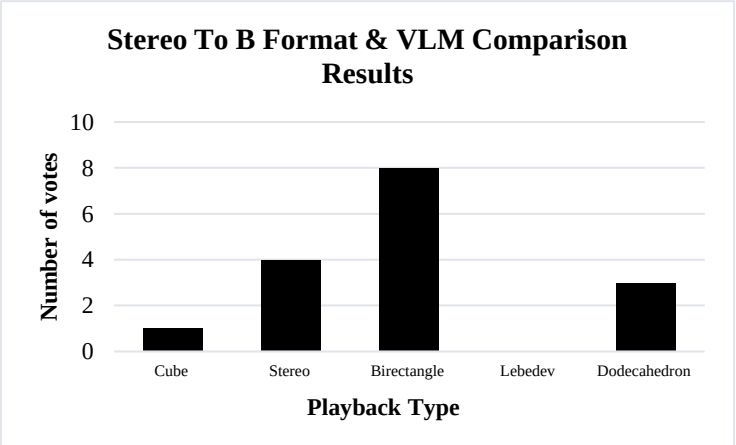


FIGURE 4.16 - STEREO TO B-FORMAT RESULTS

The results from each analysis have been presented in this section showing all significant findings from the tests. The next section will go on to discuss the findings in more detail and possible limitations found.

5 Discussion

This chapter aims to discuss the findings from the employed listening tests results & insight into what these may mean. This using the original E/V analysis findings with reference to previous similar work.

5.1 Energy & Velocity

To begin, the resulting maximum values for rV maintain a value of 1, which is expected of a standard velocity based decoder. Whereas previously mentioned, an rE of 1, would require all sound to be emitting from one transducer in the array. Without using psychoacoustic optimisation, rE can be seen to increase with sampling scheme. This suggesting that localisation accuracy over a full frequency range should improve with an increase in ambisonic order material and array; this even when just using the velocity decode.

When looking at the energy and velocity analysis of each array, a slight improvement in rE is found in the lateral regions of the bi-rectangle, at the sacrifice of rE in the median plane; supporting statements from (Gerzon, 1980) on full periphonic sound-fields. Where 4 channel audio for full periphony could not exceed an average of 0.577 for rE, but a trade-off could be found for specific areas due to its arrangement. This also mentioned in (Heller, 2006) when using rectangle arrays to better suit the geometry of a living room setup in real loudspeaker arrays. With this, it would be expected that the accuracy of localisation in the lateral plane would be higher in the birectangle than in the cube for mid/high frequencies.

It should also be mentioned that due to the approximate nature of deriving loudspeaker locations with a lebedev grid for the 3rd order material, the previously mentioned rE trade off can be found in each corner of the array. Overall these results indicated that the Lebedev should perform best for localisation, while the bi-rectangle should outperform the Cube for lateral localisation accuracy for certain frequency ranges.

5.2 Localisation

The results from the localisation test proved to be somewhat inconclusive. Overall multivariate analysis showing little significant difference between each array, in terms of localisation accuracy. This being the same when considering individual frequency ranges, apart from small possible preference for localisation <1 kHz. Supporting previous comments in terms of dual band decoding, that due to the lack of optimisation techniques frequencies >~ 1 kHz would not be localised well (Heller, 2008).

When considering the front/back confusion that can occur through binaural listening, previous research had commented on how this could be lowered considerably due to the inclusion of head tracking within the sound-field (Collins, 2013). This due to the removal of phase issues at lateral positions encountered in static conditions. Compared to the findings from this study, these confusions still occurred frequently, and participants commented on ‘Phasey’ issues during, once again similar to findings from Heller (2008). This could also be due to overall latency with the employed headset. Recent investigations by Romanov et al (2017) have shown the latency of the GY85 board used to be insufficient when compared to an alternative Arduino board (Arduino, 2017), although not found within this testing, further assessment to the overall latency of the system would be of much help.

The birectangle and cube arrays were compared to assess the differences in loudspeaker arrangement, with initial findings showing less front back confusion for 0º θ with an increase in frequency, these findings contradicting the previously mentioned MAA findings from Wiggins (2004). The results also only indicated small differences, such as the birectangle being the optimal choice for mid/high/speech content at lateral positions, as opposed to the cube for low frequencies. When comparing this to the E/V findings, these indications would be supported in terms of seeing an increase in rE as a source approaches 90º θ.

Looking back to the original hypothesis outlined in the methodology section, these findings support the null hypothesis that no differences exist between all arrays in terms of localisation accuracy. When assessing individual frequency ranges & speech for all arrays, small differences were found at some angles, but would be inconclusive for determining an overall optimum array. With roy's largest root analysis showing significant findings, indicating a preferred use low frequency content, while no significant differences were found over the mid frequencies. Whereas the uncertainty seen in the mid-frequency region, once again supporting lack of optimisation to the decoders. Overall these results have been influenced by certain limitations in the employed test that will be outlined in later sections.

5.3 Perceptual

Results from the perceptual questionnaires were presented as box plots to assess overall variance between preferred arrays for each attribute. Although not statistically significant, in terms of overall means; the Lebedev sphere was favoured for all attributes. Allowing the conclusion that in terms of playback using a velocity decode, an increase in ambisonic order can improve overall perceptions of depth, width, and immersion. Interestingly noted was the performance of the dodecahedron, not performing as expected when compared to lower order arrays. This possibly due to the position of loudspeakers around the listener not being at ear level. Showing that it is not just an increase in order that causes this, but the arrangement of loudspeakers within each order.

Findings went on to show that the increase in complexity of a sound field can improve perceived depth. These findings shown in all arrays apart from the cube. This possibly being another reason to make use of the birectangle over a cube array for VLM. During the complex sound-field test, improvements were also seen in the lebedev in perceived immersion, this not shown in any other array. Concluding that the use of a higher order virtual system can help increase perceptual attributes such as immersion & depth of a sound-field.

5.4 Stereo to B-Format

The final part of the test process assessed participant's preference for arrays in a natural scenario. This allowed a minor finding, that presenting up mixed stereo over virtual loudspeaker arrays for binaural reproduction, may be favoured over general 'in head' listening usually found with stereo & headphones. With the favoured array being the bi-rectangle. So interestingly, when participants were asked to rate a preference of sound-field in terms of perceptual attributes, higher order arrays were preferred. Whereas, when experiencing up-mixed stereo audio over the same arrays, lateral optimisation of rE seems to be a factor of preference. Although, this may be due to the use of recorded material with higher spatial information, when compared to the essentially anechoic loudspeaker conditions for the localisation tests. Furthermore, only 1 track was used for the assessment so further research would be needed to assess other possible influences.

5.5 Limitations & Observations

During testing, observations were made to assess any extra information that may be of importance. The first thing to note, is that every participant involved the rotation of the head to aid localisation of a sound source. During the perceptual testing, it was also noted that majority of participants showed a greater reaction to the second stimulus (Complex). Participants were seen to smile, dance, and spend more time investigating the sound-field & engaging with the stimulus. Promoting the use of surround techniques in virtual reality context to provide a heightened immersive experience. So, the inclusion of this feature in a sound-field is one that would be utilised by consumers.

Also found, was some participants claimed that the higher frequency content was perceived as above them, this supporting previous work stating that ~8kHz content can be perceived above the head [Wiggins, 2004], [Hollerweger, 2006]. Another possible reason for this being the use of generalised HRTF sets skewing higher frequency reproduction as noted in Horbach (1999) due to inaccurate pinnae filtering reproduction.

It must be noted that there were some significant differences in participant's ability to locate sound sources. During the tests, some participants were incredibly accurate and fast at locating sources; whereas some had great difficulty. This was seen over a mixture of younger/older listeners. The fact that these abilities were seen over such a wide variety of participants, would point towards the implementation of the technology itself & not just possible hearing impairment. One possible reason for this being the use of the SADIE HRTF library that is taken from the Kemar Dummy head. As previously mentioned these HRTF's will not apply to a whole population. This may be a reason for some participant's hesitation and prolonged time to localise a source. Another reason for this being the use of anechoic material in the localisation test. Sources panned behind the listener were perceived as in the back of the neck causing much confusion. Further assessment with binaural room impulse responses or artificial ambisonic reverb could be used to rectify this & aid perceived distance.

As can be seen from the results and the lack of statistical significance, some issues were encountered during the testing process. This due to the lack of psychoacoustic optimisation and shelf filters in the employed decoders, causing incorrect reproduction of mid frequencies over each array. This also not accompanied by the use of near-field compensation as advised by Heller (2008). The option to apply max rE in the code found in appendix 9.3 is available, but issues with implementation of phase matched shelf filters caused this to not be used. It was then decided to stay with the mathematically correct velocity decode and use this for the testing. Although the increase in area of correct reconstruction when using the max rE optimisation is not a key aim for the employed techniques, as the user is always in the sweet spot. In terms of accurate localisation of frequency content, this mid frequency optimisation may have narrowed the perceived location of sources and improved results.

Although previous pilot tests showed that the localisation test would take less time than expected, when using multiple participants, it was found not to be the case for all users. Although multiple breaks were given to avoid listening fatigue, based on similar conditions to Lopez et al (2016) localisation testing. It is suggested that if the test was repeated, either multiple groups would be used that assess fewer conditions, thus take less time, causing lower fatigue and higher validity of results. In conjunction with this, a larger sample group would also allow a greater acuity into areas of interest for each array. If this test was to be repeated, participants would be given prior listening tests to assess any inconsistencies between both ears. This being evident during testing when only using 1 hemisphere for source positions, participants with poor hearing could be at a disadvantage during the test, causing invalid results.

6 Conclusions

This work has investigated the use of ambisonic to binaural decoding with head-tracked audio. A Literature review was undertaken discussing the psychoacoustics of human hearing, showing that interaural cues must be recreated over the employed systems, to recreate the way in which the auditory system perceives a source's distance and location. An overview of ambisonic & binaural fundamentals was given, to provide insight into how the employed virtual loudspeaker methods are used during testing & examples of where it is currently being used were given. A review of recent and past investigations into the area were covered, outlining existing testing methodologies and possible things to account for during. This being the use of Gerzon's Energy and Velocity vector analysis to assess decoder performance before implementation; with practical examples of employed listening tests.

An overview of a proposed system was given to allow assessment of multiple virtual ambisonic arrays in quick succession, similar to that used in VR. This employed using Matlab, to create multiple ambisnoinc decoders. These then used in reaper to allow real-time convolution with encoded ambisonic material and rotation of the sound-field. The combination of these techniques allowed the encoded material to be used in a listening test for the assessment of localisation accuracy over multiple virtual loudspeaker arrays. While also used to give perceptual assessment of each array while listening to ambisonic sound-fields. This assessing the favoured virtual loudspeaker array in terms of perceptual attributes rather than just localisation accuracy. A final test allowed participants to experience stereo up-mixed audio over multiple virtual arrays. Results indicate that in terms of perceptual attributes, an increase in ambisonic order and spatial sampling of loudspeakers was favoured for playback with head-tracked audio. Results of the localisation study were somewhat in-conclusive in terms of an overall optimum array to be used. When a closer analysis was completed between birectangle and cube for 1st order comparisons, it was concluded that the bi-rectangle had a better performance for a larger frequency range, supporting findings from the stereo to B-Format test. When looking at the other arrays, significant performance was found over particular angles, this generally being in relation to their overall energy vector magnitude coverage, supporting the use of Gerzon's original analysis techniques.

As discussed in the previous section the employed decoders were not optimised for psychoacoustic performance in terms of shelf filters and NFC. But still allowed the analysis of multiple virtual arrays using velocity optimised decoding methods. The employed listening tests were conducted as planned but showed signs of listening fatigue during many conditions, causing issues with overall validity of some results.

The original aims of this study were to design multiple ambisonic to binaural decoders. These then used to give insight into the differences between multiple virtual arrays with head orientation applied. Although there were issues with creating fully optimised decoders and the employed listening test methods were found to be subject to listening fatigue, conclusions have still been found in terms of favourable traits for use with velocity based decoding. When employed using just a 1st order system for VR, it may be favourable to use a birectangle array with head-tracking to allow for increased localisation in the lateral plane. While also supporting previous research that the use of higher order ambisonic systems can improve perceptual attributes, even when employed over a binaural system where pinnae cues use may be lacking, this being advantageous for immersive VR experiences.

6.1 Future Work

Future work for this area would include the implementation of the phase-matched shelving filters & near field compensation within the decoders. Allowing correct reconstruction in terms of frequency content. This could then be used to assess the localisation accuracy & perceptual attributes of ambisonic audio on the elevation plane. Although methods for capturing participant results may need consideration. Also, the latency of the head-tracking system could be improved for a more natural use of the device, by using and Arduino board as opposed to the GY87, to increase performance in the localisation test. If these changes to the decoder were completed, a plugin could then be designed using Matlab's audio system toolbox; for real-time convolution & rotational matrixing of ambisonic audio without the need for multiple software plugins.

7 References

Aimil Ltd. (2016), Kemar *Dummy Head & Torso* [ONLINE]. Available at: http://www.aimil.com/products/kemar-manikins [Accessed 1 September 2017].

Altman, M. (2016). Immersive Audio for VR. In Audio for Virtual and Augmented Reality. Los Angeles, CA, USA, September 30/08/2016. AES Online Publucation: Audio Engineering Society. pp. 1-8.

Amazon, (2017), *Amazon - Hanz Zimmer - Interstellar - 'Cornfield Chase' Download*. [ONLINE] Available at: https://www.amazon.co.uk/Cornfield-Chase/dp/B00PS8PCFC. [Accessed 1 September 2017].

Arduino. 2017. Arduino. [ONLINE] Available at: https://www.arduino.cc/. [Accessed 8 September 2017].

Ballou, G, 2015. *Handbook for Sound Engineers*. 5th ed. Burlington: Focal Press. (pp 45 – 72)

Blumlein, A. (1933). Improvements in and Relating to Sound Transmission, Sound Recording & Sound Reproduction Systems. Patent, 394,325.

Cardew, H. 2017. *Stereo to B-Format Converter*. [ONLINE] Available at: https://haydoncardew.com/stereo-to-b-format-converter/. [Accessed 1 September 2017].

Collins, T, (2013). Binaural Ambisonic Decoding with Enhanced Lateral Localization. In 134th AES Convention. Rome, Italy, 4-7/05/2013. Online Publication: AES – Audio Engineering Society. 1-10.

Facebook. 2017. *Facebook 360 Spatial Workstation*. [ONLINE] Available at: https://fb360spatialworkstation.zendesk.com/hc/en-us/articles/207887969-Spatial-Workstation-Workflow. [Accessed 31 July 2017].

Farina, A(a). (2017). *Anechoic Directory - Archimedes*. [ONLINE] Available at: http://pcfarina.eng.unipr.it/Public/Anecoic/Archimedes/. [Accessed 1 July 2017].

Farina, A(b). (2017). *Conversion from Ambisonics to TBE format*. [ONLINE] Available at: http://pcfarina.eng.unipr.it/TBE-conversion.htm. [Accessed 16 August 2017].

Gerzon, M (1974) Sound Reproduction Systems. Patent, 1494751.

Gerzon, M, (1980). PRACTICAL PERIPHONY: THE REPRODUCTION OF FULL-SPHERE SOUND. In AES 65th Convention. London, 25-28/02/1980. Online Publication: AES - Audio Engineering Society. 1-11.

Gerzon, M, (1992a). Ambisonic Decoders For HDTV. In Audio Engineering Society 92nd Convention. Vienna, 24-27/03/1992. Vienna: Audio Engineering Society. 1-42

Gerzon, M, (1992b). Psychoacoustic Decoders for Multispeaker Stereo & Surround Sound. In 93rd AES Convention. San Francisco, 1-4/10/1992. Online Publication: Audio Engineering Society (AES). 1-47.

Gerzon, M, 1992(c). General Metatheory of Auditory Localisation. Audio Engineering Society, 92nd Convention.
Google - Youtube. 2017. *Youtube Spatial Audio*. [ONLINE] Available at: https://support.google.com/youtube/answer/6395969?hl=en-GB&ref_topic=2888648. [Accessed 17 August 2017].

Google VR, (2017), *Google Daydream Headset* [ONLINE]. Available at: https://vr.google.com/intl/en_uk/daydream/headset/?utm_source=en-ha-na-sem [Accessed 1 September 2017].

Google.2017.*Jump-Overview*.[ONLINE]Available at: https://support.google.com/jump/answer/6374961?hl=en&ref_topic=6375006. [Accessed 17 August 2017].

Guitierrez-Parera, P, Lopez,J, (2016). On the influence of the frequency response over azimuth localization with consumer headphones. In AES Conference on Headphone Technology. Denmark - Aalborg, 24-26/08/2016. Online Publication: Audio Engineering Society. 1-7.

HedRot Open Source Head Tracker.. 2017. *HedRot User Manual* - PDF Download. [ONLINE] Available at: https://abaskind.github.io/hedrot/. [Accessed 1 September 2017].

Heller, A, Benjamin, A, Lee, R, (2006). Localization in Horizontal-Only Ambisonic Systems. In AES 121st Convention. San Francisco - USA, 5-8/10/2006. Online Publication: AES - Audio Engineering Society. 1-13.

Heller, A, Lee, R, Benjamin, E, (2008). Is My Decoder Ambisonic. *In Audio Engineering Society - 125th Convention*. San Francisco - USA, 2-5/10/2008. Online Publication: Audio Engineering Society. 1 -22.

Hollerweger, F, 2006. Periphonic Sound Spatialization in Multi-User Virtual Environments. PhD Report. Institute of Electronic Music and Acoustics (IEM),:

Horbach, U, Karamustafaoglu, A, Pellegrini, R, Mackensen, P, Theile, G, (1999). Design and Applications of a Data-based Auralisation System for Surround Sound. In Aes 106th Convention. Munich Germany, 8-11/05/1999. Online Publication: Audio Engineering Society (AES). 1-10.

Kearney, G, Doyle, T, (2015). Height Perception In Ambisonic Based Binaural Decoding.. *In 139th Convention*. New York, 1/11/2015. USA: Audio Engineering Society. 1-10.

Kearney, G, Hughes , S, (2016). Moving Virtual Source Perception in 2D Space . In Audio for Virtual & Augmented Reality . Los Angeles , 30/09/2016 - 01/10/2016. USA : Audio Engineering Society . 1-9.

Kronlachner, M,(a) 2017. *Ambisonics plug-in suite for production and performance usage.. Institute of Electronic Music & Acoustics*, Graz, [Online]. ., 1-5. Available at: http://lac.linuxaudio.org/2013/papers/51.pdf [Accessed 1 August 2017].

Kronlachner, M,(b) 2017. *Plug in Suite For Mastering the production and playback in surround sound and Ambisonics*.. Institute of Electronic Music & Acoustics, Graz, [Online]., 1-5. Available at: http://www.matthiaskronlachner.com/wp-content/uploads/2013/01/kronlachner_aes_studentdesigncompetition_2014.pdf [Accessed 1 August 2017].

Lancaster, D., 2005. *Research Methods - A concise introduction to research management and business consultancy*.. 2nd ed. Oxford: Elsevier, Pg – 154-156

Lecomte, P, Gauthier, P, Langrenne, C, Garcie, A, Berry, A, (2015). On the Use of a Lebedev Grid for Ambisonics. In 139th AES Convention. New York - USA, 29/10/2015 - 01/11/2015. Online Publication: AES - Audio Engineering Society. 1-12.

Mather, G, 2009. Foundations of Sensation and Perception: Second Edition. 2. Psychology Press.

Maths Is Fun. 2017. *Euler & Platonic Solids*. [ONLINE] Available at: https://www.mathsisfun.com/geometry/eulers-formula.html. [Accessed 1 September 2017].

McQuarrie, D., 2003. *Mathematical Methods For Scientists and Engineers*. 1st ed. Sausalito, California: University Science Books. Pg – 667,668.

Mhacoustics(a). 2017. *MHAcoustics Audio Demo Page*. [ONLINE] Available at: https://mhacoustics.com/demos. [Accessed 1 September 2017].

Mhacoustics(b). 2017. *mhacoustics - Eigen Microphone*. [ONLINE] Available at: https://mhacoustics.com/products. [Accessed 1 September 2017].

Moerman, K, 2010. *Platonic Solid Function for MatLab*. [ONLINE] Available at: https://ch.mathworks.com/matlabcentral/fileexchange/28213-platonic-solid. [Accessed 1 September 2017].

Moore's Law. 2017. Moore's Law. [ONLINE] Available at: http://www.mooreslaw.org/. [Accessed 15 August 2017].

Oculus, (2017), *Oculus Rift VR Head-Set* [ONLINE]. Available at: https://www.oculus.com/rift/ [Accessed 1 September 2017].

Olive, S, Welti, T, Mcmullin, E, (2013). A Virtual Headphone Listening Test Methodology . In 51st International Conference. Finland , 22-24/10/2013. USA: Audio Engineering Society . 1 -10.

Playstation VR, (2017), *Playstation VR Headset* [ONLINE]. Available at: https://www.playstation.com/en-gb/explore/playstation-vr/ [Accessed 1 September 2017].

Pulkki, V, (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. In AES Volume 45 Issue 6. Online, 01/06/1997. Online Publication: Audio Engineering Society. 456-466.

Reaper - Cockos. 2017. *Reaper Effects Summary Guide*. [ONLINE] Available at: https://www.reaper.fm/guides/ReaEffectsGuide.pdf. [Accessed 25 August 2017].

Romanov, M, Berghold, P, Rudrich, D, Zaunschirm, M, Frank,M, Zotter, F, (2017). Implementation and Evaluation of a Low-cost Head-tracker for Binaural Synthesis. *In Audio Engineering Society - 142nd Convention*. Berlin, 20-23/05/2017. Germany: Audio Engineering Society. 1-6.

Rumsey, F, 2003. *Spatial Audio*. 2nd ed. Oxford: Focal Press - Music Technology Series. (Pp 10-20, 26, 52,53,59,60,62,64,67,68,72)

Rumsey, F, Mason, R, (2000). An Assessment of the Spatial Performance of Virtual Home Theatre Algorithms by Subjective & Objective Methods. In Audio Engineering 108th Convention. Paris, France, 19-22/02/2000. Online Publication: Audio Engineering Society (AES). 1-39.

Schulein , R.B, Fligor, B (2016). In situ Subjective & Objective Acoustic Seal Performance Tests for Insert Earphones. In 141st Convention. Los Angeles, 29/09/2016 - 02/10/2016. USA: Audio Engineering Society. 2,3,7.

Shivappa, S, (2016). Efficient, compelling and immersive VR audio experience using Scene Based Audio / Higher Order Ambisonics. *In Audio for Virtual and Augmented Reality*. Los Angeles, CA, USA, September 30–October 1. AES: Audio Engineering Society. 1-10.

Soundfield . 2017. *Soundfield Mic*. [ONLINE] Available at: http://www.soundfield.com/. [Accessed 1 September 2017].

Stitt, P, Hendrick, E, Katz, B (2016). The Role Of Head Tracking In Binaural Rendering. In VDT International Convention. Tonmeistertagung, November 29th. Online Publication: Research Gate. 2-7.

University of York(a) - SADIE Project.. 2017. *SADIE - Spatial Audio for Domestic Interactive Entertainment.*. [ONLINE] Available at: https://www.york.ac.uk/sadie-project/ambidec.html. [Accessed 1 September 2017].

University of York(b) - Sadie. 2017. *SADIE Google Binaural Filters*.. [ONLINE] Available at: https://www.york.ac.uk/sadie-project/GoogleVRSADIE.html. [Accessed 17 August 2017].

Wiggins, B, 2004. An Investigation Into The Real-Time Manipulation And Contol Of Three Dimensional Sound Fields. PhD Report. Derby: Derby University.Pg, 5 – 100.

Wiggins, B, 2017. *WigWare Plugin Suite.* [ONLINE] Available at: https://www.brucewiggins.co.uk/?page_id=78. [Accessed 1 September 2017].

Wiggins, B, Paterson-Stephens, I, Schillebeeckx, P, (2001). The Analysis of Multi-Channel Sound Reproduction Algorithms using HRTF data.. *In AES -19th International Conference*. Schloss Elmau - Germany, 21 - 24/06/2001. Online Publication: Audio Engineering Society. 1-13.]

8 Bibliography

Ballou, G., 2015. Handbook for Sound Engineers. Taylor & Francis.

G., W., 1998. 3-D Audio Using Loudspeakers. Springer Science & Business Media.

Gilat, A., 2014. MATLAB. Wiley.

M., R., 2008. Auditory Perception. Cambridge University Press.

McClellan,Burrus,Oppenheim,Parks,Schafer,Schuessler, J., 1998. Computer Based Signal Processing Using Matlab 5. 1st ed. New Jersey: Tom Robbins.

Rumsey, F., 2003. Spatial Audio. Taylor & Francis.

Steiglitz, K., 1996. A Digital Signal Processing Primer. 1st ed. California: Addison-Wesley.

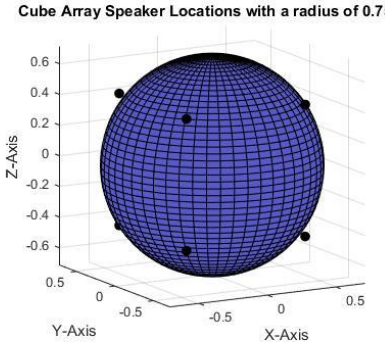
9 Appendices

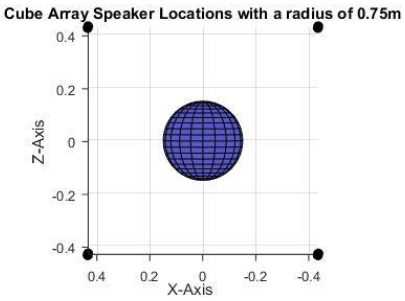
9.1 Spherical Harmonic Examples (Up to 3rd Order)

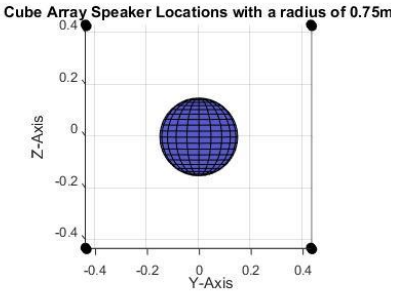
Spherical harmonics up to 3rd order.

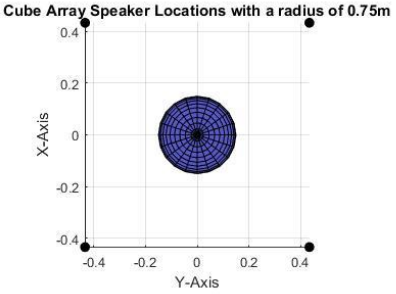
9.2 Virtual Arrays

9.2.1 Cube

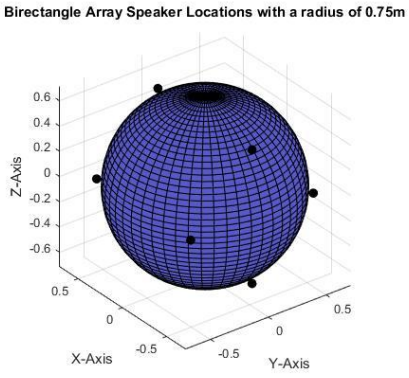


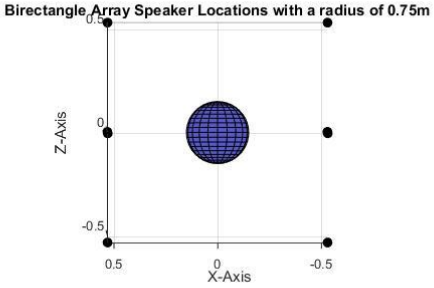


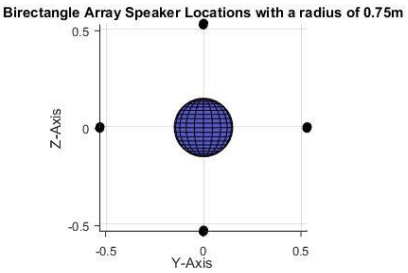


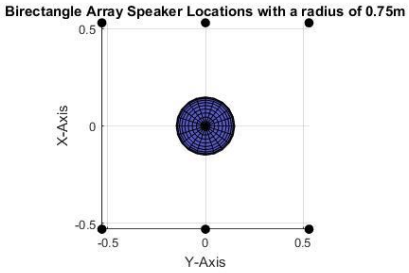


9.2.2 Birectangle

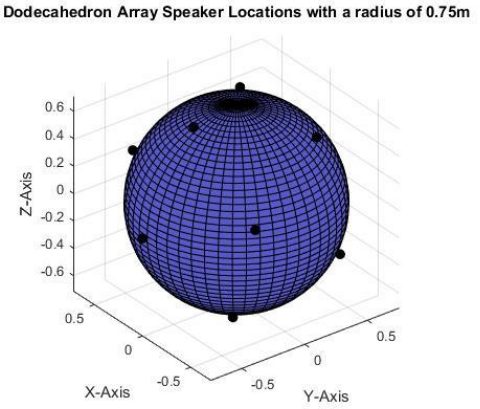


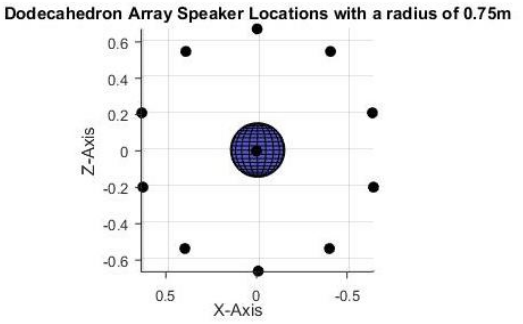


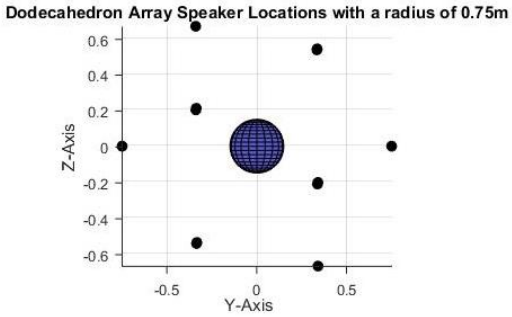


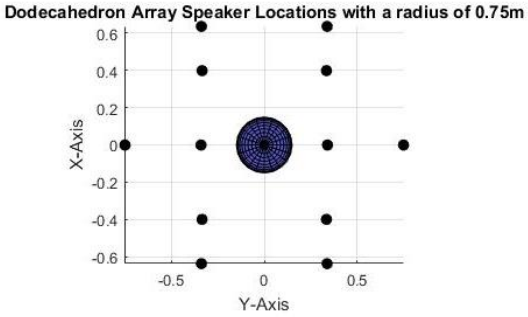


9.2.3 Dodecahedron

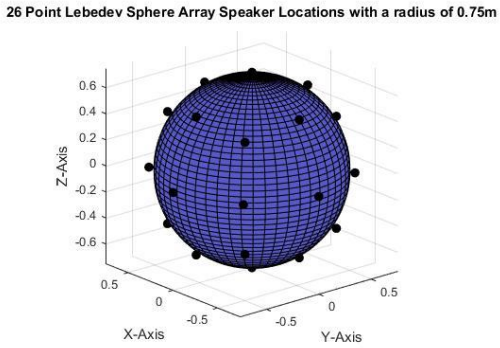


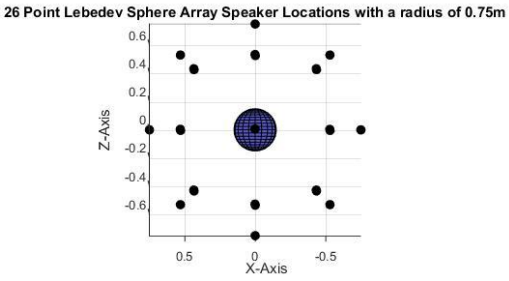


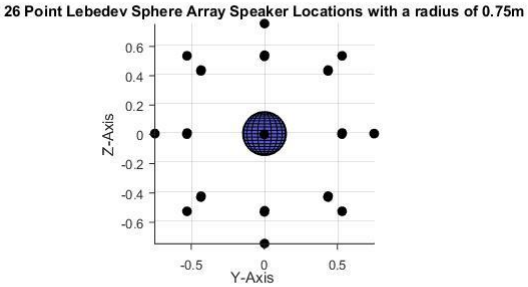


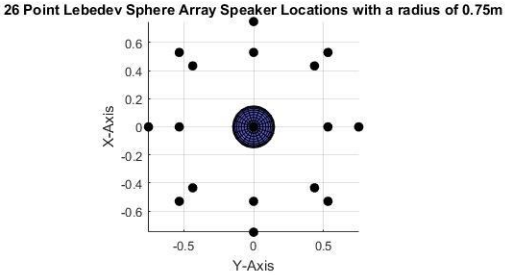


9.2.4 26 Point Lebedev Sphere









9.3 Matlab Code

```
Please Refer to Attached Disc.

Disc Contents
       1 Decoder Script (Main)
       2 Decoder B-Format HRTF Function
       3 Decoder Binaural Function
       4 Folder Containing Impulse Responses Generated With script for test
       5 Folder Containing Excel File Housing Results.
       6 Associated Test Files
```

9.4 Decoder Matrix & Impulse Response Plots

Cube Decoder Coefficients =

	0	1	2	3
sp1	0.12500	0.21580	0.21793	0.21580
sp2	0.12500	0.21580	0.21793	-0.21580
sp3	0.12500	-0.21580	0.21793	-0.21580
sp4	0.12500	-0.21580	0.21793	0.21580
sp5	0.12500	0.21580	-0.21793	0.21580
sp6	0.12500	0.21580	-0.21793	-0.21580
sp7	0.12500	-0.21580	-0.21793	-0.21580
sp8	0.12500	-0.21580	-0.21793	0.21580

Bi-rectangle Decoder Coefficients =

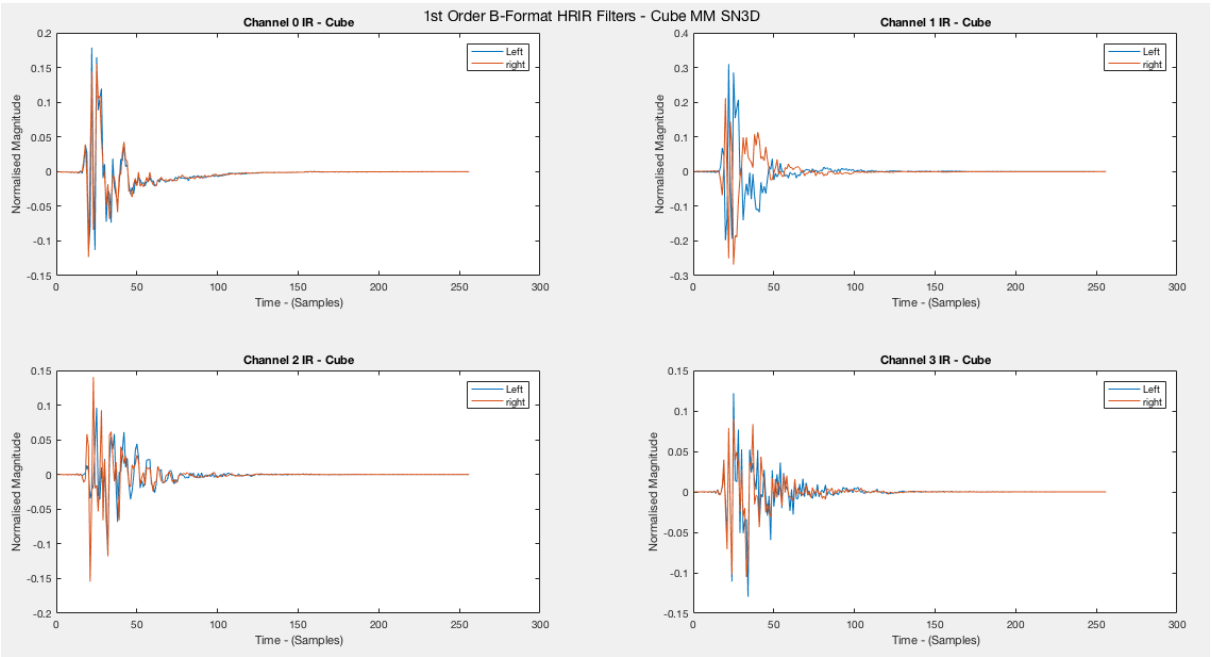
	0	1	2	3
sp1	0.12500	0.17678	0.35355	0
sp2	0.12500	-0.17678	0.35355	0
sp3	0.12500	0.17678	0	0.35355
sp4	0.12500	0.17678	0	-0.35355
sp5	0.12500	-0.17678	0	-0.35355
sp6	0.12500	-0.17678	0	0.35355
sp7	0.12500	0.17678	-0.35355	0
sp8	0.12500	-0.17678	-0.35355	0

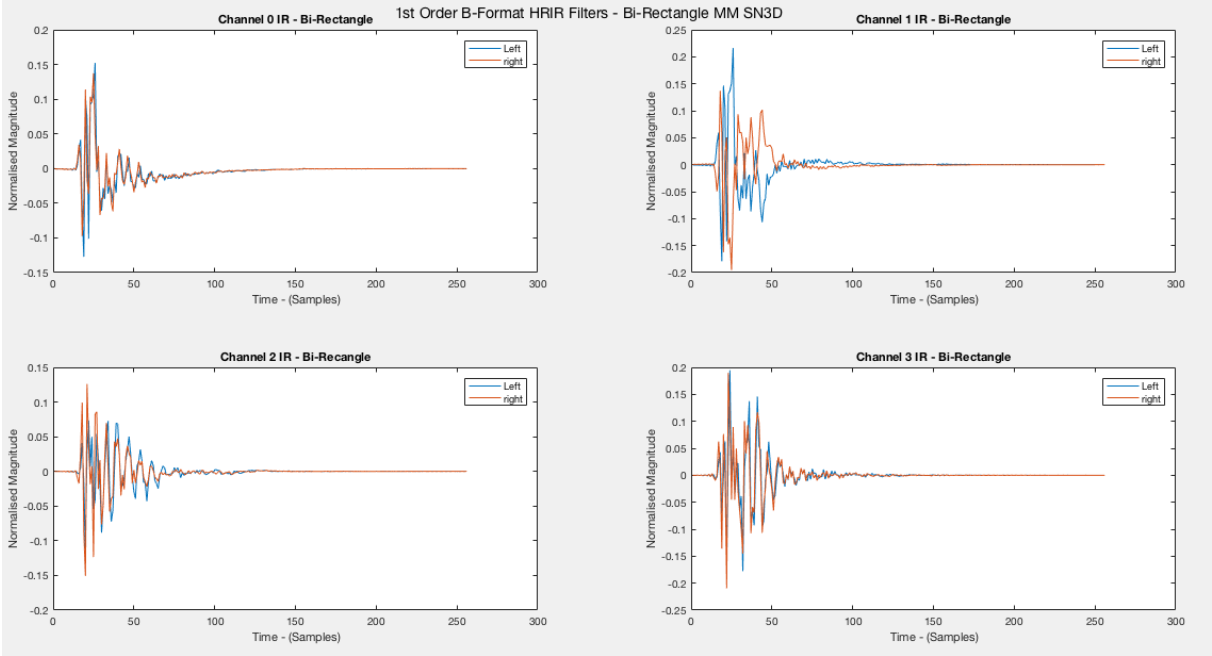
Dodecahedron Decoder Coefficients

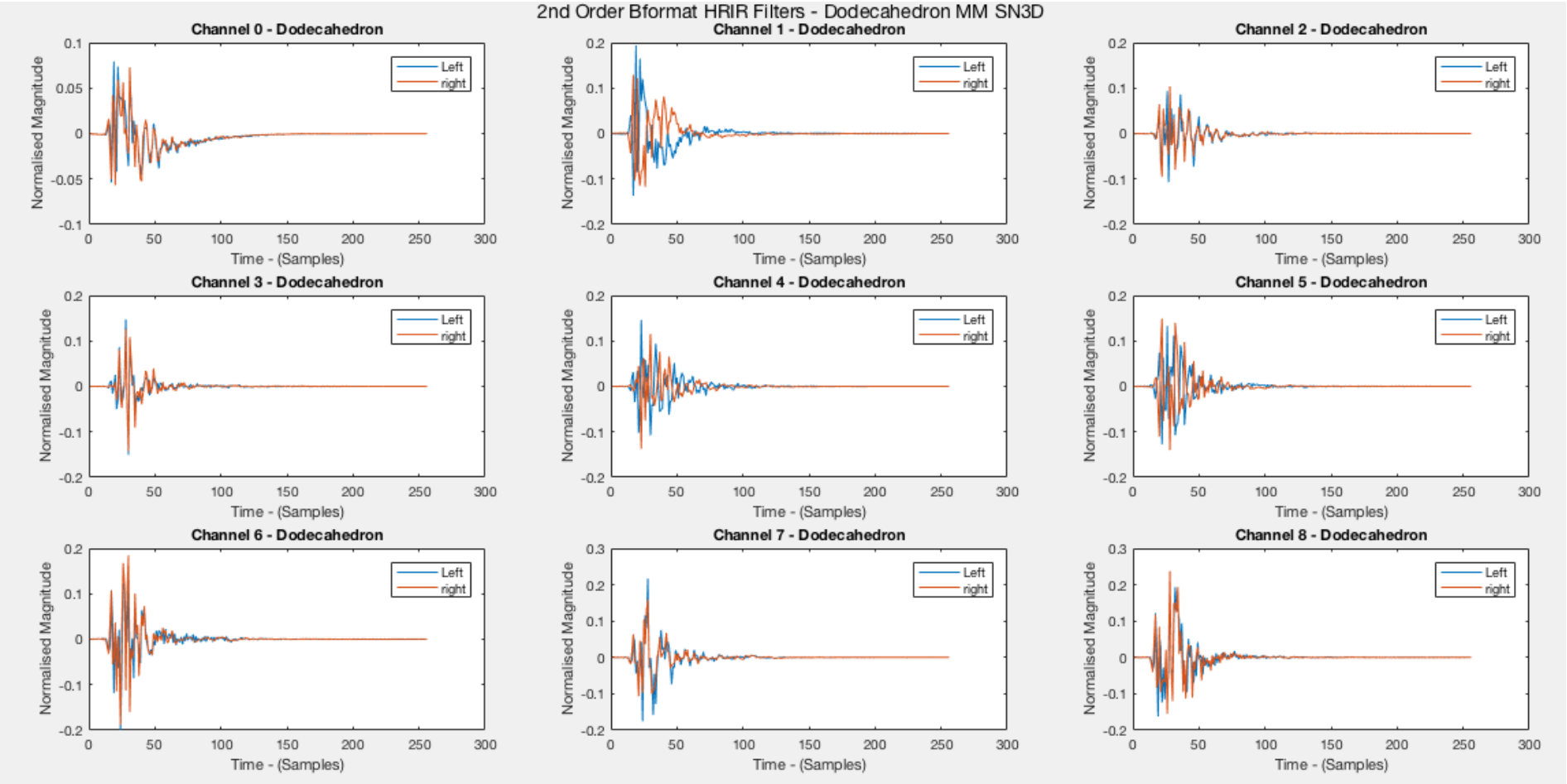
	Channel 0	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6	Channel 7	Channel 8
SP1	0.0837893	0	0.2243033	-0.1115962	0	0	0.2930533	-0.2866388	0.0724429
SP2	0.0846715	0.1325658	0.1821989	0.1119402	0.166994	0.2734096	0.12127	0.2333551	-0.0277549
SP3	0.0846715	-0.1325658	0.1821989	0.1119402	-0.166994	-0.2734096	0.12127	0.2333551	-0.0277549
SP4	0.082398	0.2114379	0.0689645	-0.1117649	-0.2732409	0.1696096	-0.1633526	-0.0917307	-0.1894567
SP5	0.082398	-0.2114379	0.0689645	-0.1117649	0.2732409	-0.1696096	-0.1633526	-0.0917307	-0.1894567
SP6	0.0820717	0	0.0013668	0.2484941	0	0	-0.2088881	0.00339	0.3619804
SP7	0.0820717	0	-0.0013668	-0.2484941	0	0	-0.2088881	0.00339	0.3619804
SP8	0.082398	0.2114379	-0.0689645	0.1117649	0.2732409	-0.1696096	-0.1633526	-0.0917307	-0.1894567
SP9	0.082398	-0.2114379	-0.0689645	0.1117649	-0.2732409	0.1696096	-0.1633526	-0.0917307	-0.1894567
SP10	0.0846715	0.1325658	-0.1821989	-0.1119402	-0.166994	-0.2734096	0.12127	0.2333551	-0.0277549
SP11	0.0846715	-0.1325658	-0.1821989	-0.1119402	0.166994	0.2734096	0.12127	0.2333551	-0.0277549
SP12	0.0837893	0	-0.2243033	0.1115962	0	0	0.2930533	-0.2866388	0.0724429

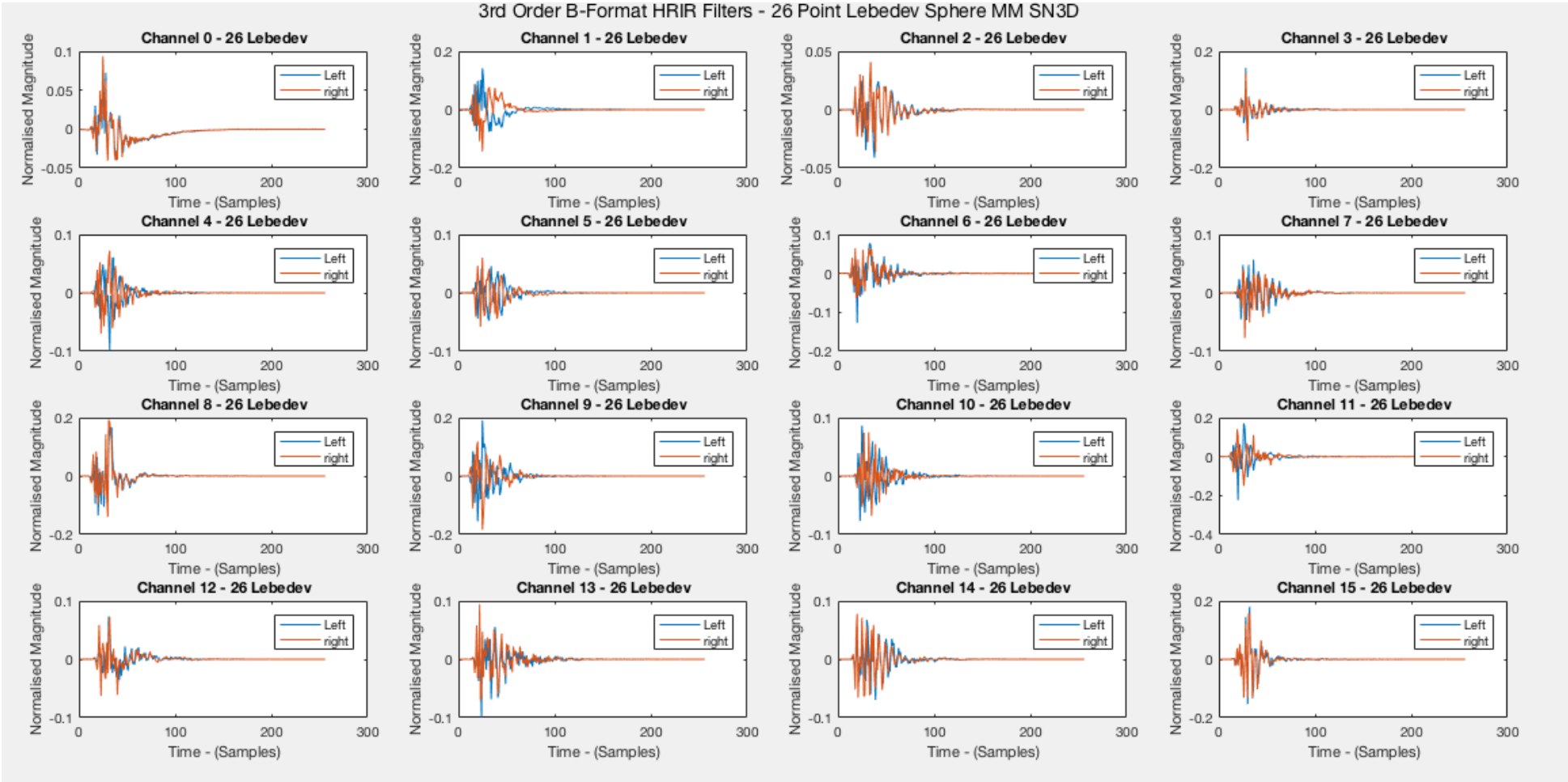
Lebedev Sphere Decoder Coefficients

	Channel 0	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6	Channel 7	Channel 8	Channel 9	Channel 10	Channel 11	Channel 12	Channel 13	Channel 14	Channel 15
SP1	0.0389079	0	0.1450671	0	0	0	0.2226561	0	0	0	0	0	0.3220257	0	0	0
SP2	0.0385738	0	0.0789009	0.0790231	0	0	0.0559981	0.1533042	0.096225	0	0	0	-0.0332496	0.1664536	0.1825742	0.0847237
SP3	0.0385738	0.0790231	0.0789009	0	0	0.1533042	0.0559981	0	-0.096225	-0.0847237	0	0.1664536	-0.0332496	0	-0.1825742	0
SP4	0.0385738	0	0.0789009	-0.0790231	0	0	0.0559981	-0.1533042	0.096225	0	0	0	-0.0332496	-0.1664536	0.1825742	-0.0847237
SP5	0.0385738	-0.0790231	0.0789009	0	0	-0.1533042	0.0559981	0	-0.096225	0.0847237	0	-0.1664536	-0.0332496	0	-0.1825742	0
SP6	0.0384595	0.0578849	0.0574324	0.0578849	0.1019221	0.1018649	-0.0010023	0.1018649	0	0.0785955	0.1677158	0.0598221	-0.0993685	0.0598221	0	-0.0785955
SP7	0.0384595	0.0578849	0.0574324	-0.0578849	-0.1019221	0.1018649	-0.0010023	-0.1018649	0	0.0785955	-0.1677158	0.0598221	-0.0993685	-0.0598221	0	0.0785955
SP8	0.0384595	-0.0578849	0.0574324	-0.0578849	0.1019221	-0.1018649	-0.0010023	-0.1018649	0	-0.0785955	0.1677158	-0.0598221	-0.0993685	-0.0598221	0	0.0785955
SP9	0.0384595	-0.0578849	0.0574324	0.0578849	-0.1019221	-0.1018649	-0.0010023	0.1018649	0	-0.0785955	-0.1677158	-0.0598221	-0.0993685	0.0598221	0	-0.0785955
SP10	0.0382397	0	0	0.1441222	0	0	-0.1106599	0	0.1924501	0	0	0	0	-0.1993501	0	0.2563191
SP11	0.0382397	0.0777874	0	0.0777874	0.1518936	0	-0.1106599	0	0	0.1372053	0	-0.1234986	0	-0.1234986	0	-0.1372053
SP12	0.0382397	0.1441222	0	0	0	0	-0.1106599	0	-0.1924501	-0.2563191	0	-0.1993501	0	0	0	0
SP13	0.0382397	0.0777874	0	-0.0777874	-0.1518936	0	-0.1106599	0	0	0.1372053	0	-0.1234986	0	0.1234986	0	0.1372053
SP14	0.0382397	0	0	-0.1441222	0	0	-0.1106599	0	0.1924501	0	0	0	0	0.1993501	0	-0.2563191
SP15	0.0382397	-0.0777874	0	-0.0777874	0.1518936	0	-0.1106599	0	0	-0.1372053	0	0.1234986	0	0.1234986	0	0.1372053
SP16	0.0382397	-0.1441222	0	0	0	0	-0.1106599	0	-0.1924501	0.2563191	0	0.1993501	0	0	0	0
SP17	0.0382397	-0.0777874	0	0.0777874	-0.1518936	0	-0.1106599	0	0	-0.1372053	0	0.1234986	0	-0.1234986	0	-0.1372053
SP18	0.0384595	0.0578849	-0.0574324	0.0578849	0.1019221	-0.1018649	-0.0010023	-0.1018649	0	0.0785955	-0.1677158	0.0598221	0.0993685	0.0598221	0	-0.0785955
SP19	0.0384595	0.0578849	-0.0574324	-0.0578849	-0.1019221	-0.1018649	-0.0010023	0.1018649	0	0.0785955	0.1677158	0.0598221	0.0993685	-0.0598221	0	0.0785955
SP20	0.0384595	-0.0578849	-0.0574324	-0.0578849	0.1019221	0.1018649	-0.0010023	0.1018649	0	-0.0785955	-0.1677158	-0.0598221	0.0993685	-0.0598221	0	0.0785955
SP21	0.0384595	-0.0578849	-0.0574324	0.0578849	-0.1019221	0.1018649	-0.0010023	-0.1018649	0	-0.0785955	0.1677158	-0.0598221	0.0993685	0.0598221	0	-0.0785955
SP22	0.0385738	0	-0.0789009	0.0790231	0	0	0.0559981	-0.1533042	0.096225	0	0	0	0.0332496	0.1664536	-0.1825742	0.0847237
SP23	0.0385738	0.0790231	-0.0789009	0	0	-0.1533042	0.0559981	0	-0.096225	-0.0847237	0	0.1664536	0.0332496	0	0.1825742	0
SP24	0.0385738	0	-0.0789009	-0.0790231	0	0	0.0559981	0.1533042	0.096225	0	0	0	0.0332496	-0.1664536	-0.1825742	-0.0847237
SP25	0.0385738	-0.0790231	-0.0789009	0	0	0.1533042	0.0559981	0	-0.096225	0.0847237	0	-0.1664536	0.0332496	0	0.1825742	0
SP26	0.0389079	0	-0.1450671	0	0	0	0.2226561	0	0	0	0	0	-0.3220257	0	0	0


1st Order B-Format HRIR Filters - Cube MM SN3D
Channel 0 IR - Cube
Channel 1 IR - Cube
Channel 2 IR - Cube
Channel 3 IR - Cube
Left
right
Normalised Magnitude
Time - (Samples)


1st Order B-Format HRIR Filters - Bi-Rectangle MM SN3D
Channel 0 IR - Bi-Rectangle
Channel 1 IR - Bi-Rectangle
Channel 2 IR - Bi-Recangle
Channel 3 IR - Bi-Rectangle
Left
right
Normalised Magnitude
Time - (Samples)


2nd Order Bformat HRIR Filters - Dodecahedron MM SN3D
Channel 0 - Dodecahedron
Channel 1 - Dodecahedron
Channel 2 - Dodecahedron
Channel 3 - Dodecahedron
Channel 4 - Dodecahedron
Channel 5 - Dodecahedron
Channel 6 - Dodecahedron
Channel 7 - Dodecahedron
Channel 8 - Dodecahedron
Left
right
Normalised Magnitude
Time - (Samples)


3rd Order B-Format HRIR Filters - 26 Point Lebedev Sphere MM SN3D
Channel 0 - 26 Lebedev
Channel 1 - 26 Lebedev
Channel 2 - 26 Lebedev
Channel 3 - 26 Lebedev
Channel 4 - 26 Lebedev
Channel 5 - 26 Lebedev
Channel 6 - 26 Lebedev
Channel 7 - 26 Lebedev
Channel 8 - 26 Lebedev
Channel 9 - 26 Lebedev
Channel 10 - 26 Lebedev
Channel 11 - 26 Lebedev
Channel 12 - 26 Lebedev
Channel 13 - 26 Lebedev
Channel 14 - 26 Lebedev
Channel 15 - 26 Lebedev
Normalised Magnitude
Time - (Samples)
Left
right

9.5 Consent & Debrief Forms

Listening Test Consent Form **Ref No –**

Thank-you for taking the time to read this form. You have been selected to complete a short test to assess ambisonic to binaural dedcoding. To allow more information about participants please answer a few simple questions below.

Gender – Male ☐ **Female** ☐

Age - ☐

Audio Background ? - .....

Any Known Hearing Impairment ? - .....

Please read & initial/sign this form.

You are invitied to participate in this study conducted by myself Peter Rowland-Jones

During the Test you shall be asked to:

- Be seated in a room within Derby Unviersity (Markeaton Street).
- Then listen to audio extracts over headphones, locate sounds, then give feedback.
- The test will be repeated with different conditons and last approximately 30minutes.
- At the end you will be debriefed and have a chance to ask questions concerning the tests to the researcher.

Please **Initial** in each box:

☐ I understand the process to be followed in this listening test and that it will take approximately 30 minutes of my time.

☐ I understand my right to withdraw from the test at any point & my anonymity during observation.

☐ I understand all information will remain confidentail and not shared. This information will be kept on record in a secure place. After any publication of results these will not be able to be removed.

All data gathered will be used towards an MSc in Audio Engineering to assess the implementation of different virtual ambisonic arrays for binaural use.

More information concerning the use of this data and the study itself will be presented after the test with a debriefing form. Any questions you may have now please ask before signing the bottom of this form.

If you have any further questions regarding the study after it is complete please contact Peter Rowland-Jones at p.rowlandjones1@derby.ac.uk & quote the reference number found at the top of this form.

I have read and understood all information presented within this form and consent to participation within this study.

Participant Initials	**Participant Signature**	**Date**
.....	.....	.....
Researcher Initials	**Researcher Signature**	**Date**
.....	.....	.....

Debriefing Form

Firstly, Thank-you for taking part in this study. The results gathered today will be analysed to assess different virtual loudspeaker arrays used with ambisonic material for binaural playback.

More about the invesitgation -

Study Title – An investigation into Ambisonic to binaural decoding using multiple virtual loudspeaker arrays.

Ambisonic audio is a method of providing surround sound experiences. It captures the soundfield in a way that can be seperated from the playback method. This allows the same material to be listened to on many different playback systems, ranging from headphones, to multiple loudspeakers arrays. The flexibility of this method has many advantages, one of these being used with headphones to provide binaural experiences. This being a method of presenting audio to the ears that is closer to how we hear sound in real life. I.E externalising sound to be percieved outside of the head rather than between the ears. There are many ways of binaralising this audio for use with headphones. Todays test has assessed your ability to localise sound within different virtual loudspeaker arrangements & your overall perceptions within them. It will allow a better understanding of best methods for providing these experiences for such use as within virtual/augmented reality formats.

This data will be kept on record & will not be shared with any organisation. It will only be used in the assessment outlined within the consent form. To remove your results please contact the researcher with details provided using your given reference number. Please be aware once any publication of results occours, withdrawal of given data will not be possible.

If you wish to be forwarded any further publication of the data collected today please fill out your current e-mail address below

Participant E-mail - .....

(For Participant)

Contact Details

Researcher – Peter Rowland-Jones E-Mail– p.rowlandjones1@derby.ac.uk

Director Of Studies – Bruce Wiggins E-mail - b.j.wiggins@derby.ac.uk

Phone contact – 01332 593216 [Markeaton Street Campus]

REFERENCE NUMBER -

Thank-you again for taking the time to complete these tests to aid my studies.

Peter Rowland-Jones

9.6 Questionnaire Example

Audio Sample XX

Width = Apparent width of a soundfield.

Depth = Perceived distance/Depth of the soundfield.

Immersion = Overall sense of being enveloped in a different environment.

1

Width = 1---2---3---4---5---6---7---8---9---10

Depth = 1---2---3---4---5---6---7---8---9---10

Immersion = 1---2---3---4---5---6---7---8---9---10

2

Width = 1---2---3---4---5---6---7---8---9---10

Depth = 1---2---3---4---5---6---7---8---9---10

Immersion = 1---2---3---4---5---6---7---8---9---10

3

Width = 1---2---3---4---5---6---7---8---9---10

Depth = 1---2---3---4---5---6---7---8---9---10

Immersion = 1---2---3---4---5---6---7---8---9---10

4

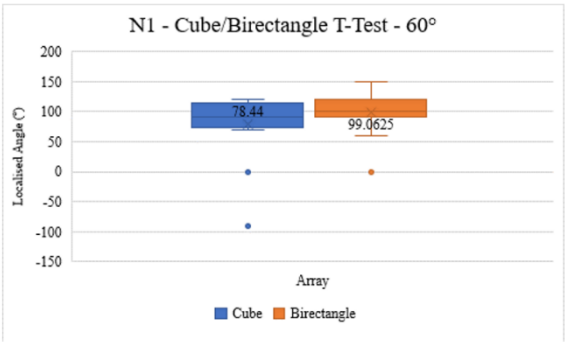
Width = 1---2---3---4---5---6---7---8---9---10

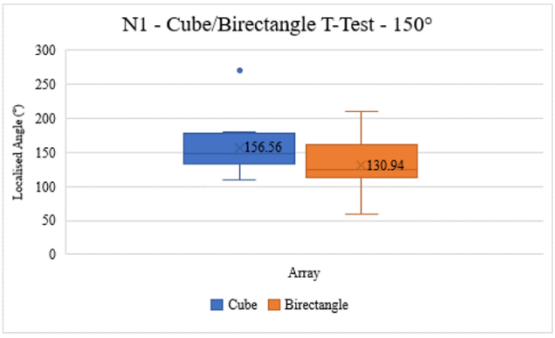
Depth = 1---2---3---4---5---6---7---8---9---10

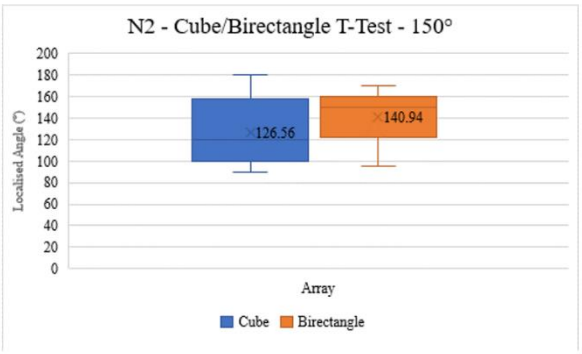
Immersion = 1---2---3---4---5---6---7---8---9---10

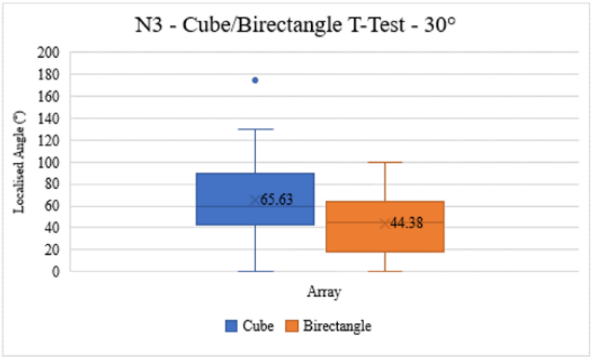
9.7 Results Plots

9.7.1 Cube/Birectangle Comparison

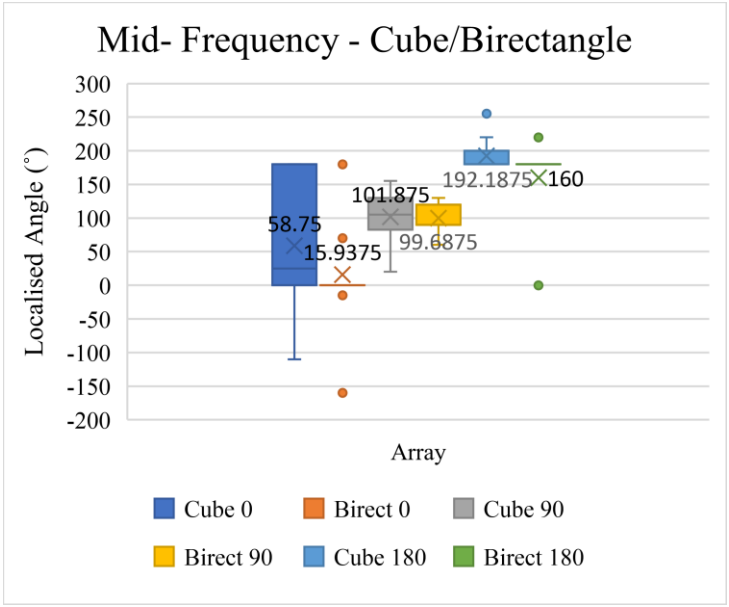


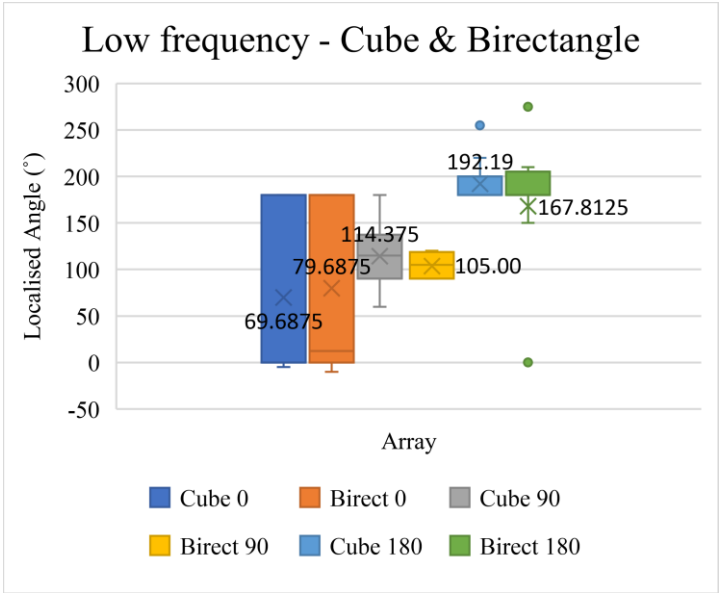


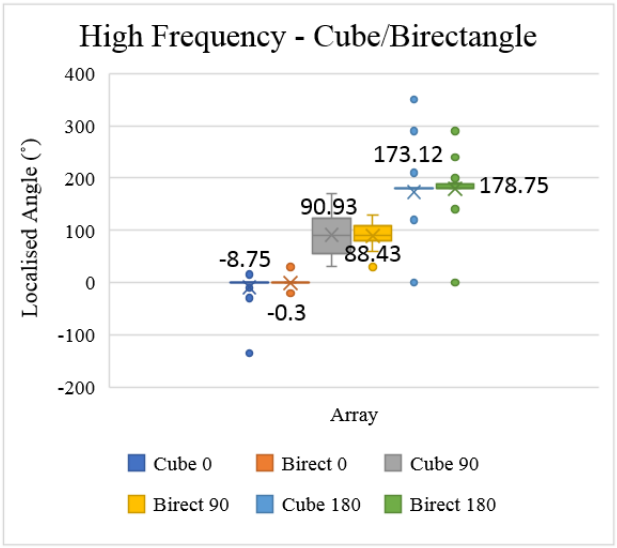


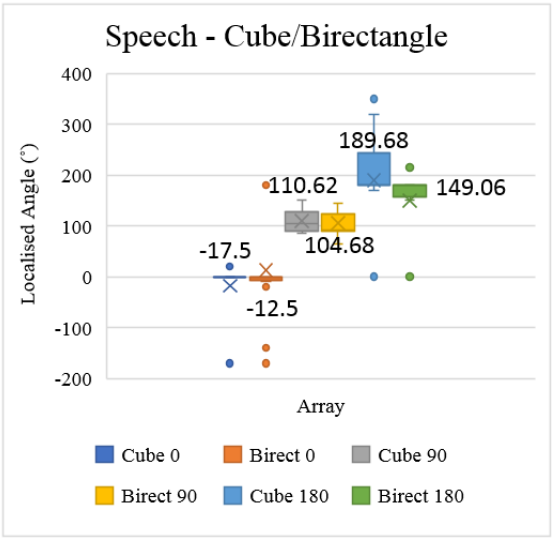


9.7.1.2 – Front/Back/Lateral Assessment

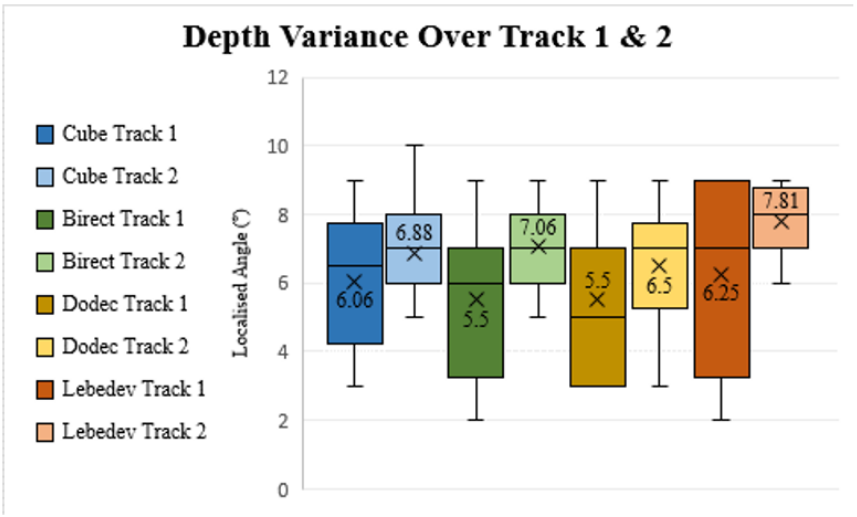


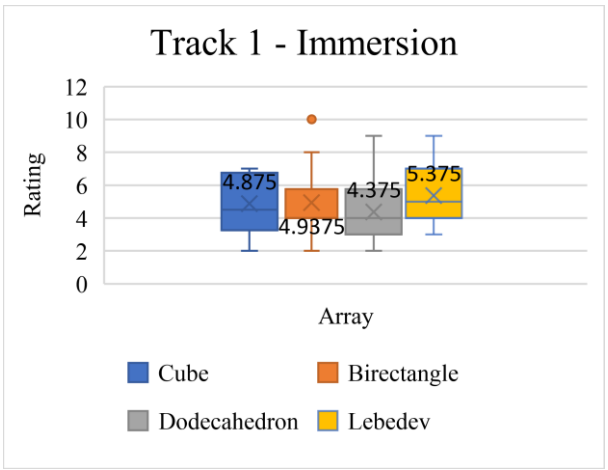


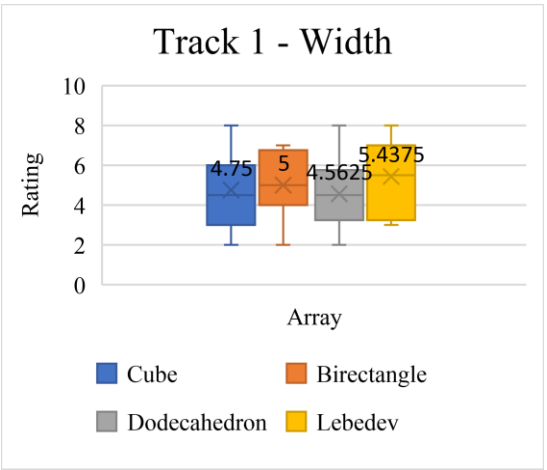


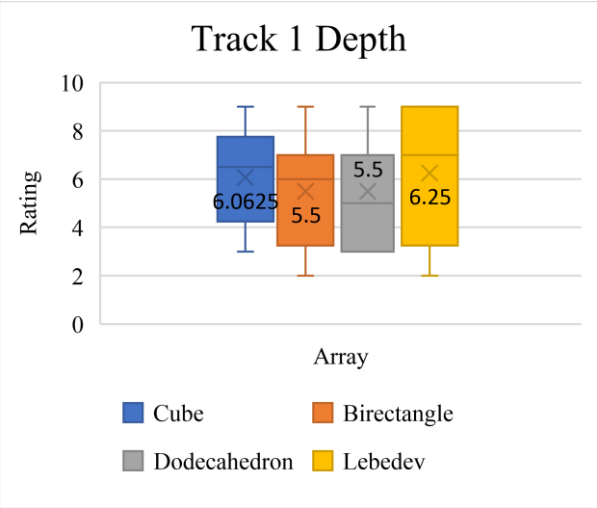


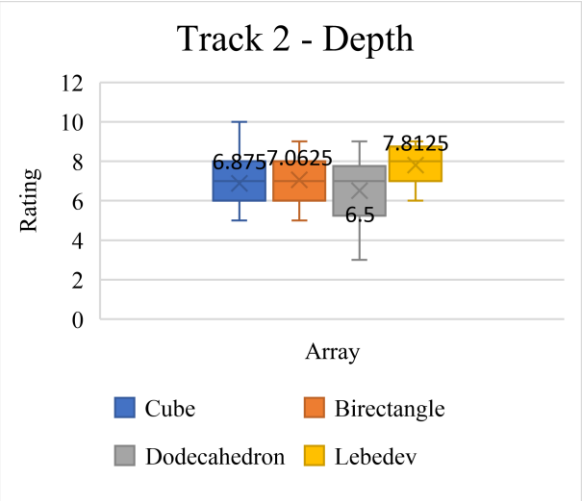
9.7.2 Perceptual Results

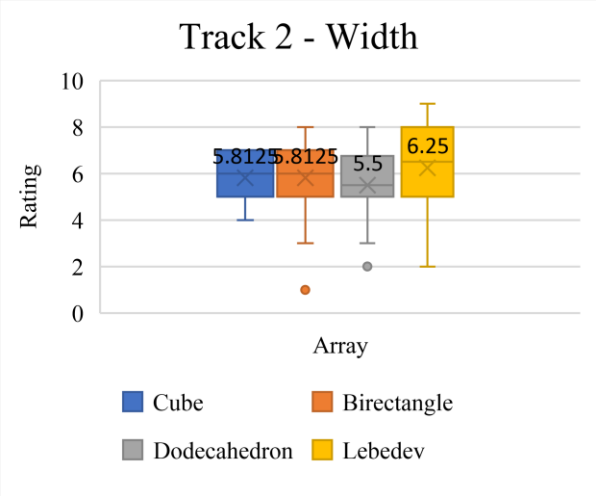


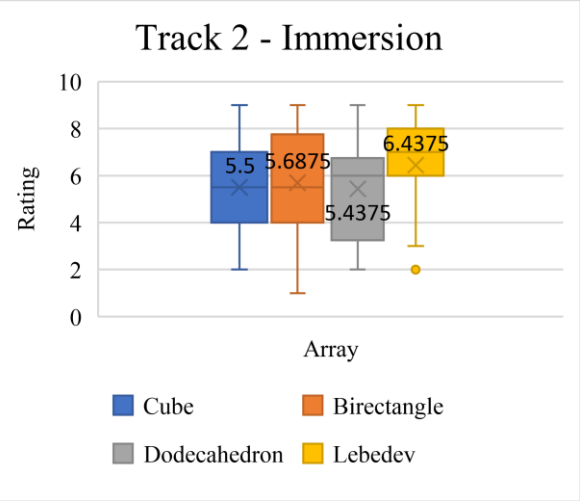












9.7.3 Stereo to B-Format & Stereo Comparison