Proposing Factors Towards a Standardised Testing Environment for Binaural and 3D Sound Systems

By

Sebastian Chandler Črnigoj

A thesis submitted to Liverpool John Moores University in partial fulfilment of the requirements for the degree of Doctor of Philosophy

June 2020

Abstract

Binaural sound systems are a growing industry in the upcoming age of three-dimensional (3-D) technology. While many commercial and home sound systems are entering the market, there is no clear method of determining their suitability for different applications. Thus, a standardised methodology of testing such systems is proposed in order to evaluate and compare new and existing binaural microphone array systems.

This thesis presents a thorough literature review into the current techniques and methods for the capture, and playback, of binaural audio. Furthermore, the literature defines some of the broad range of current evaluation methods for a given binaural system and analysis of head-related transfer functions. Current development challenges in binaural audio are identified; elevation (height) informational cues, individual user hearing 'signature', playback devices on multiple platforms and a set (standardised) testing environment, for such systems.

For the first time, a method of codifying the accuracy of binaural localisation cues in humans using binaural systems is investigated. This provides an indication of how people interpret binaural audio within a 3D soundscape. Humans, using head-related transfer functions (HRTF), are capable of determining the direction of arrival (DOA) of a sound relative to their position in space. A system capable of capturing the informational cues contained in HRTFs demands a 'fool-proof' testing methodology, owing to the complex nature of human hearing, or more specifically sound localisation.

The implicating factors which determine the location of a sound, and methods of capturing such sounds, have been determined. Data suggests there are common localisation issues relating to given areas around a subject as well as the unique characteristics of the sound.

A testing, and comparison, methodology is proposed based on the data collected mentioned above. Subjects were positioned within a circular loudspeaker array and instructed to communicate the perceived location of a sound from a series of 24 possible locations. The accuracy of a subject's result was calculated based on precision and an overall score was assigned to each participant. Validation methods were created through the mathematical probability of conducting the experiment through guesswork, and simulations were run to compare theoretical versus actual. Further validation methods were employed, and subject sample size was investigated.

This proposed methodology provides quantitative and qualitative comparison methods to determine the function and suggested application of any given binaural sound system. The proposed testing procedure aims to determine issues pertaining to localisation abilities and unify manufacturers' method of validating such binaural systems.

Results indicate a direct correlation between higher-scoring locations and subjects. Certain locations were more difficult across all participants, whilst other high-scoring locations were easier to approximate. The simulation provided results matching those of the theoretical calculation of the mathematical probability, and subject sample sizes were speculated to a certain minimum requirement.

Declaration

The author declares that no portion of this thesis has been submitted in support for another degree, qualification or other institute of learning. The work presented in, and contribution to, this thesis is in its entirety the work of the author. There was no collaborative group work undertaken in this research.

Acknowledgements

Firstly, I would like to express my deep gratitude to my supervision team, in particular Dr Karl Jones for his endless and unwavering support throughout the course of my study. Further thanks go to my other supervisors, Dr Paul Otterson, Dr David Ellis and Dr Stephen Wylie.

My thanks also go to my friends and family who have supported and encouraged me during these times.

It is my privilege to extend my gratefulness to the Audio and Music Production staff for their continued support and endless theoretical discussions regarding my research.

Lastly, I thank all students, and participants alike, who assisted me in my relentless quest for results during my experimentation process.

Table of Contents

Abstract ii
Declaration iv
Acknowledgements v
Table of Contents vi
List of Tables & Figures x
Nomenclature and Abbreviations xiv
Definitions xv
Chapter 1: Introduction – Aim and Thesis Outline 1
1.1 Background 1
1.1.1 History 1
1.1.2 Immersive, Spatial and '3D' sound 2
1.1.3 Non-binaural 'Spatial' Audio 3
1.1.3.a Stereo 3
1.1.3.b Surround Sound 5
1.2 Aim and Objectives 6
1.3 Thesis Outline 7
1.4 Summary 8
Chapter 2: Literature Review 9
2.1 Head-Related Transfer Functions 9
2.1.1 Distance/Time 10
2.1.2 Loudness/Intensity 12
2.1.3 Timbre and Frequency 13
2.1.4 Summation and Crossover/Trading of Interaural Cues 13
2.1.5 Head-and-Torso Related Transfer Functions 14
2.1.6 Database Storage and Formatting of HRTF (and HpTF) Datasets 14
2.1.7 Binaural Localisation in the Vertical Domain 15
2.2 Binaural Audio Capture 15
2.3 Binaural Audio Playback/Reproduction 18
2.3.1 Headphone Methods 18
2.3.1.a Closed Headphones 19
2.3.1.b Open and Semi-open Headphones 19
2.3.1.c In-ear Phones 20
2.3.1.d Multi-driver 21

2.3.2 Binaural Audio Playback with Loudspeakers 23
2.4 Psychological Testing Factors 24
2.4.1 Psychoacoustic Factors 24
2.4.1.a Haas Effect 25
2.4.1.b Inside-the-head 'Locatedness' 26
2.4.2 Front-and-Back Confusion 27
2.5 Current Testing Regimes and Procedures 27
2.5.1 Mathematical Measurements ('Quantitative') 28
2.5.2 Psychoacoustic Measurements ('Qualitative') 28
2.5.3 Stimulus/-i 29
2.6 Limitations & Challenges 30
2.6.1 Height Information 30
2.6.2 Adaptability to Individual Users 31
2.7 Literature Overview 32
2.8 Chapter Summary 32
Chapter 3: Theory 33
3.1 Qualitative or Quantitative Measurement 33
3.2 Evaluating Negatively Influencing Factors 34
3.3 Stimulus Selection 36
3.4 'Pseudo-Randomised' Locations 37
3.5 Subjects and Subject Validation 38
3.6 Hypothesis 40
3.7 Summary 40
Chapter 4: Design 42
4.1 Assessment System of Results 42
4.1.1 Point-based System 44
4.1.2 Calculating the Probability of Guesswork 46
4.2 Creating a Loudspeaker Array 47
4.3 Location and dB Levels 49
4.4 Binaural Recording, Playback and Testing Procedure of Experimentation 49
4.5 Summary/Chapter End 50
Chapter 5: Method – Standardising Testing Environments 51
5.1 Order of Locations 51
5.2 Preparation Procedure for Experimentation 52
5.3 Experimentation Procedure 53
5.4 Evaluation of a Binaural System's Efficiency 54

5.4.1 Experiment 1 using Stimulus 1 54
5.4.2 Experiment 2 using Stimulus 2 55
5.4.3 Validation Process 55
5.6 Creating a Scoring and Point System 56
5.7 Simulation of the Experimental Procedure 56
5.7.1 Simulation 1 using Dataset 1 57
5.7.2 Simulation 2 using Dataset 2 57
5.7.3 Simulation 3 using Dataset 3 58
5.7.4 Simulation 4 using Dataset 4 58
5.7.5 Simulation 5 using Dataset 5 58
5.8 Summary 58
Chapter 6: Results and Analysis 60
6.1 Experiment results 61
6.1.1 Experiment 1 61
6.1.2 Experiment 2 65
6.1.3 Further Analysis and Comparison of Experiments 1 and 2 68
6.1.3.a Experiment 1 – Headphone 70
6.1.3.b Experiment 1 – Open-field 72
6.1.3.c Experiment 2 – Headphone 74
6.1.3.d Experiment 2 – Open-field 76
6.1.3.e Comparison of Consistency Locations from Headphone versus Open-field 77
6.2 Verification Check – Simulation 78
6.2.1 Result for Dataset 1 78
6.2.2 Result for Dataset 2 79
6.3.3 Results for Datasets 3, 4 and 5 80
6.3.4 Simulation Analysis 82
6.4 Summary 84
Chapter 7: Discussion 85
7.1 Initial Observation of Results and Stimuli 85
7.1.1 Experiment 1 86
7.1.2 Experiment 2 88
7.1.3 Crossover and Summary of Experiments 1 and 2 91
7.1.4 Point system, Simulation and Validation Review 92
7.2 Subjects and Subject Population Size 93
7.3 Superimposing Factors and Challenges 94
7.3.1 Front-and-Back Confusion 94

7.3.2 Inside-the-head 'Locatedness' and Characterising & Personalising HRTFs 96
7.3.3 Haas Effect 97
7.4 Summary 97
Chapter 8: Conclusions 98
8.1 Main Contributions 98
8.2 Recommendations for Further Work 100
8.3 Summary 101
References 103
Bibliography 111
Appendix A 112
Appendix B 114
Appendix C 115
Appendix D 116

List of Tables & Figures

List of Figures:

Figure 1 ‘The listening experience’ (Taken from Roginska, 2018). 3
Figure 2 Typical loudspeaker positioning for a stereo listening environment 4
Figure 3 Example of a 5.1 loudspeaker positioning for a surround sound listening environment 5
Figure 4 The difference in arrival times between the left and right ear 11
Figure 5 Industry-standard Binaural recording ‘dummy-head’, the Neumann KU-100, (taken from; https://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description) 16
Figure 6 Bar-and-pinnae binaural microphone, the 3Dio Free Space Binaural Microphone (taken from; https://3diosound.com/collections/microphones/products/free-space-binaural-microphone) 17
Figure 7 Closed-back headphones, Sennheiser Momentum 2.0 (taken from: https://www.rtings.com/headphones/learn/open-vs-closed-back) 19
Figure 8 Open-back headphones, Sennheiser HD 650 (taken from: https://www.rtings.com/headphones/learn/open-vs-closed-back) 20
Figure 9 Diagram of in-ear phones 21
Figure 10 Multi-driver 7.1 surround headphones with driver positions (Taken from Roginska, 2018) 22
Figure 11 The precedence effect as described by W. M. Hartmann, the left channel dominates thus the sum is perceived as arriving from only one location. (Hartmann, 1999) 26
Figure 12 The lateral view of an example loudspeaker array with a 10 degree of separation between each driver 36
Figure 13 Accuracy versus Precision of a point-based system, 0/0 versus 100/100 43
Figure 14 Accuracy versus Precision of a point-based system, 50/100 versus 100/50 43
Figure 15 Accuracy versus Precision of a point-based system, 100/50 versus 100/75 44
Figure 16 An example of the point system for the different degrees of error 45
Figure 17 The circular array of loudspeakers located in the isolation chamber 48
Figure 18 A switch box outside the isolation chamber, designed to individual control each of the 24 loudspeakers 49
Figure 19 Acceptable versus Unacceptable examples for orders of locations 52
Figure 20 Position of subject (or dummy-head) and loudspeakers in the loudspeaker array (aerial perspective) 53
Figure 21 Stimulus 1 spectrogram (left) and waveform (right) 54
Figure 22 Stimulus 2 Spectrogram (top) and waveform (bottom) 55
Figure 23 An example of an answer set for the experimentation procedure 56
Figure 24 Subject ‘X’ results from open-field test (left) and headphone test (right) 61
Figure 25 Graph of overall results from all Experiment 1 subjects (binaural) 62
Figure 26 Graph of overall results from all Experiment 1 subjects (open-field) 63
Figure 27 Graph of the differences between both tests in Experiment 1 (open-field – headphone) .. 64
Figure 28 Graph of overall results from validated-only subjects in Experiment 1 (headphone), subjects with an overall score below 45% in the open-field test are presented as 0, and do not contribute towards the mean 65
Figure 29 Graph of overall results from all Experiment 2 subjects (headphone) 66
Figure 30 Graph of overall results from all experiment 2 subjects (open-field) 66
Figure 31 Graph of differences between both tests in Experiment 2 (open-field – headphone) 67
Figure 32 Graph of overall result from validated-only subjects in Experiment 2 (headphone) 68
Figure 33 Headphone results of Experiment 1 for each degree of acceptable error 71

Figure 34 Open-field results of Experiment 1 for each degree of acceptable error............................73
Figure 35 Headphone results of Experiment 2 for each degree of acceptable error..........................75
Figure 36 Open-field results of Experiment 2 for each degree of acceptable error............................77
Figure 37 Dataset 1 simulation results 79
Figure 38 Dataset 2 simulation results 80
Figure 39 Dataset 3 simulation results 81
Figure 40 Dataset 4 simulation results 81
Figure 41 Dataset 5 simulation results 82
Figure 42 Position of subject and loudspeakers in the loudspeaker array (aerial perspective) (copy of Figure 20) 86
Figure 43 Comparison of results in headphone vs open-field tests for any degree of error in Experiment 2..91

List of Tables:

Table 1 Statistical results of Experiment 1 (binaural) 61
Table 2 Statistical results of Experiment 1 (open-field) 62
Table 3 The percentage error of each statistical overall result ((Open-field – Headphone)/Openfield * 100)) 63
Table 4 The average decrease in overall results from open-field to headphone in Experiment 1 63
Table 5 Overall results from validated-only subject in Experiment 1 (headphone) 64
Table 6 Statistical results from Experiment 2 (binaural) 65
Table 7 Statistical results from Experiment 2 (open-field) 67
Table 8 The percentage error of each statistical overall result ((Open-field – Headphone)/Open-field * 100) 67
Table 9 The average decrease in overall results from open-field to headphone in Experiment 2 67
Table 10 Overall results from validated-only subjects for Experiment 2 (headphone) 68
Table 11 A more accurate representation of successful subjects from both validation experiments 69
Table 12 Summary of all standard deviations across both experiments, where verified results include only the results above a pass mark of 45% 69
Table 13 The order of locations and results of Experiment 1 along with the various degrees of acceptable error (headphone) 70
Table 14 The error difference between consistency location (#5) for Experiment 1 (headphone) 71
Table 15 The order of locations and results of Experiment 1 along with the various degrees of acceptable error (open-field) 72
Table 16 The error difference between consistency location (5) for Experiment 1 (open-field) 73
Table 17 The order of locations and results of Experiment 2 along with the various degrees of acceptable error (headphone) 74
Table 18 The error difference between consistency location (14) for Experiment 2 (headphone) 75
Table 19 The order of locations and results of Experiment 2 along with the various degrees of acceptable error (open-field) 76
Table 20 The error difference between consistency location (14) for Experiment 2 (open-field) 77
Table 21 The percentage error difference of consistency locations in Experiment 1 ((open-field – headphone)/open-field * 100) 78
Table 22 The percentage error difference of consistency locations in Experiment 2 ((Open-field – Headphone)/Open-field * 100) 78
Table 23 Statistical results of Dataset 1 (simulation) 79
Table 24 Statistical results of Dataset 2 (simulation) 79
Table 25 Statistical results of Dataset 3 (simulation) 80
Table 26 Statistical results of Dataset 4 (simulation) 80
Table 27 Statistical results of Dataset 5 (simulation) 81
Table 28 Statistical results of repetitions of all five datasets (simulation) 83
Table 29 New averages for the results of all five datasets (simulation) 83
Table 30 Highest and lowest scoring locations in Experiment 1 87
Table 31 Highest and lowest scoring locations for Experiment 2 89
Table 32 The locations, by descending order of highest correct estimated locations for the headphone, and open-field, tests in Experiment 2 90
Table 33 Differences in results of shared locations in Stimulus 1 and Stimulus 2 92
Table 34 Results of confusion between front and rear locations for Experiment 1 (binaural and open-field) 95

Table 35 Results of confusion between front and rear locations for Experiment 2 (binaural and open-field) 96

Nomenclature and Abbreviations

r	Radius of a circle (metres)
θ	Degree of azimuth relating to the direction of arrival of a sound (radians)
c	Speed of sound (metres per second^{-1})
F	Frequency (hertz)
Λ	Wavelength (metres)
T	Time (seconds)
3D	Three-dimensional
HRTF	Head-related transfer function
HRiR	Head-related impulse response
DOA	Direction of arrival
LFE	Low-frequency effects
ITD	Interaural time difference
IID	Interaural intensity difference
ILD	Interaural level difference
dB	Decibel
IHL	Inside-the-head locatedness
DSP	Digital signal processing
HAT	Head-and-torso (binaural 'dummy' head)
HNT	Head, neck and torso (binaural 'dummy' head)
STFT	Short-time Fourier transform
HpTF	Headphone-to-ear transform function
ADA	Acceptable degree of accuracy

Definitions

Ambisonic (format)	*High-fidelity, multi-channel audio signal that contains some form of localisation perception, a spherical surround sound system*
Anechoic	*Free from echo (or reverberation)*
Auricle	*The outer structure of the ear, the lobe*
Automation	*The editing of a sound signals volume, spatial information, etc.*
Azimuth	*The angle of a receiving sound relative to the horizontal plane of a given receiver*
Binaural	*A digital (stereo) signal containing localisation information in the form of interaural differences*
Closed-back (headphones), Circum-aural	*Headphones which enclose the pinnae in an isolated chamber*
Colouration	*The spectral properties or response of any given sound/equipment (i.e. headphones)*
Diotic	*Pertinent to the use of both ears, binaural hearing*
Direction of arrival (DOA)	*Location of a sound source relative to the user's position*
Far-field	*Sound sources at a distance of 1 m or greater*
Headphones	*A playback device consisting of a pair of transducers for transposing electrical signals into sound, worn on the head*
Head-related transfer functions (HRTF)	*Spatial information and characteristics contained in audio received by a pair of sensors (receiver, 'ears')*
Immersive	*Generating or reproducing a life-like experience through digital audio (or video)*
Interaural	*The differences and variation of a single audio signal between two ears, particularly timing and intensity*
Near-field	*A sound source within a distance of 1 m (opposite of far-field)*
Open-field	*Un-obstructed listening environment, 'free-form' hearing, without the use of headphones or another alternative device*
Pan, panning	*The editing of an audio signal in terms of its audio image, the distribution of a sound signal into a multi-channel field*
Pink noise	*An artificially generated noise, whereby energy per frequency interval is inversely proportional to the frequency*

Pinna, Pinnae	*Another term for auricle, more often used when relating to animals*
Plane	*A level (component) of a 3D spectrum (e.g. horizontal, median, etc.)*
Playback, Played	*The delivery of an existing (recorded) audio signal through a delivery device, i.e. headphones*
Priming	*Preparation of test subjects for psychoacoustic testing. (e.g. providing instructions or information prior to testing)*
Psychoacoustic	*The psychological perception of sound*
Quadraphonic	*Sound reproduction through the use of four channels, often used for '3D' or surround sound*
Receiver	*Subject or capture device*
Spatial	*Pertaining informational cues which depict, or portray, a sense of life-like, immersive or '3D' environment*
Stereo, Stereophonics	*Two-channel delivery of an audio signal which gives the user (listener) an impression of multiple sound sources*
Supra-aural ('over-ear')	*With relation to headphones, sitting on the ear (as opposed to over or in)*
Surround sound	*A system based on stereophonics involving multiple audio signals to create a sense of realistic immersion in a sound environment*
Timbre	*The distinct tonal characteristics of any given sound which allows it to be identified*
Tonal	*A description for the spectral property of a sound*
Transmitter	*The perceived location of a sound source relative to each respective listener, the object of the DOA*
White noise	*An artificially generated noise with equal intensity at varying frequencies (thus a constant power spectral density)*

Chapter 1: Introduction – Aim and Thesis Outline

1.1 Background

Binaural audio is a form of audio reproduction that attempts to recreate the spatial awareness ability found in humans and many animals. Binaural audio, or spatial recording array/diotic (Nikunen et al., 2016), as it was known previously, was initially introduced in the late 19th century. This was demonstrated in 1884 when Luis of Portugal, was unable to attend an opera in person but experienced it through an early version of binaural audio played through the use of a French invention belonging to Clément Ader, namely the théâtrophone (Bertho-Lavenir, 1989).

This invention was initially created using a simple form of stereophonics, the delivery of two separate audio channels from two locations. This would later become one of the most common delivery methods of audio to date. It provided a sense of immersion through positioning audio cues within the frontal, horizontal plane. A certain instrument, or any individual sound, could be placed (panned) to appear as though it were arriving from the left, or right, or anywhere in between depending on the amount of automation used.

There has been a recent reignited interest in binaural systems, owing to the increase of 3D technology such as virtual reality systems, which demand higher quality and clarity from binaural recording technology. Thus, many areas of the audio technology industry are attempting to reproduce the immersive localisation abilities found naturally in humans.

The growth of many new and existing audio systems that seek to capture and reproduce near-fully immersive audio suggests that there is a need for a set, standardised testing procedure to validate and compare said systems, thus helping the end user to determine the most suitable system for their needs.

Binaural audio is currently widely confused as a generic spatial audio concept, rather than its specific function and properties. In order to distinguish and define the core principles of binaural audio, first the broader picture should be considered. The sub-sections below outline the position and definition of binaural audio for the purpose of the work presented in this thesis.

1.1.1 History

Clément Ader's 'Théâtrophone' (The Theatrophone, 1895) was furthered in the United Kingdom by the Electrophone Company Ltd in 1894 (Van Drie, 2015) but was ultimately unsuccessful owing to the requirement of specialised and personal headsets (what we now call headphones). This two-channel audio was not broadly recognised until the work of Alan

Blumlein in 1931 who filed for patents of stereophonics of records and film, now commonly named just stereo (Birkinshaw, 1968). Blumlein's work on acoustics and binaural audio capture in 1933 provides the foundation and fundamental characteristics of the process used in modern binaural systems.

1.1.2 Immersive, Spatial and '3D' sound

To understand the importance and workings of binaural audio to accurately determine a standardised testing procedure, first the broader spectrum of immersive audio needs to be considered.

Immersive audio, otherwise known as spatial or '3D' audio, is a multichannel audio format capable of reproducing the life-like sound localisation abilities naturally found in humans (and certain animals) (Mayfield, 2016). Like immersive systems for other senses such as vision, a certain set of characteristics such as perception of a sources' location are required. Fewer of these qualities will deter the accuracy and ability to localise or perceive the direction of arrival (DOA) of any given sound (Seo & Jeon, 2019). Such features range from various frequencies, loudness intensities, unique tonal characteristics referred to as timbre, etc. (See section 2.1).

All forms of audio rely on an individual's unique experience, and the cues that are detected. Even a single-channel audio, or mono, source such as a loudspeaker, can be interpreted as arriving from a certain elevation, distance and azimuth relative to the individual's facing direction. Additional audio channels provide further information and can therefore create the 'virtual' audio space surrounding any listener. Roginska (2018) describes this as the 'listener's perspective' and illustrates the process, as reproduced in Figure 1 (Roginska, 2018).

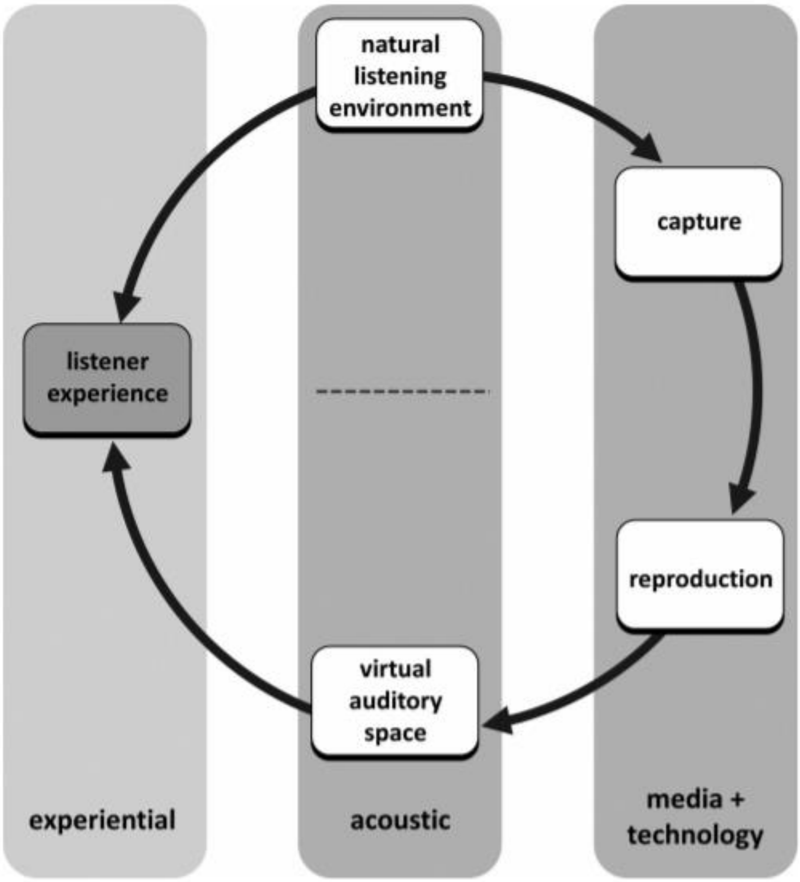


Figure 1 'The listening experience' (Taken from Roginska, 2018).

1.1.3 Non-binaural 'Spatial' Audio

This section looks at the technology and methods of immersive audio that are not strictly defined as binaural. For the purpose of this thesis, binaural audio will be defined as "the method of reproducing audio as heard by humans", or in technical terminology "*the capture and reproduction of a stereo channel with particular informational cues that portray a spatial perception*" (Zhang et al., 2017).

<u>1.1.3.a Stereo</u>

Stereophonic (two channel) sound is the first mainstream hi-fidelity audio to introduce a listening environment and with spatial awareness. This perception of sound directionality dates to work by Alan Blumlein in his 1930s patent and to date stereo technology largely relies on concepts developed following his work (Birkinshaw, 1968). Through loudspeaker or headphone reproduction, the listener is positioned at an angle between two (or in the case of surround sound, multiple) sound sources, and in the case of loudspeakers, often in-front, angled towards the listener to create a perception of a sound image (Whitaker & Benson, 2002).

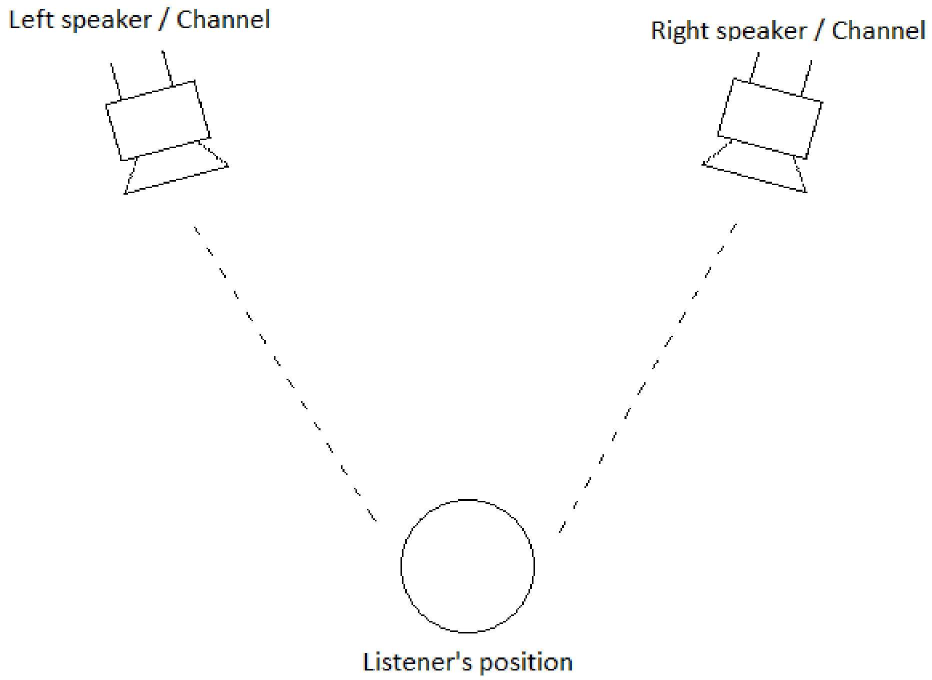


Figure 2 Typical loudspeaker positioning for a stereo listening environment

The perceived efficiency of a sound's direction is greatly impacted by the physical space (i.e. room) owing to its acoustic characteristics. A minute shift in position can negatively affect the experience due to factors such as room treatment (acoustics) to reflections and reverberation. Many professional consumers of stereo introduce counteractive actions to improve a stereo field in a particular "idealistic" listening location. This calibration of reverberation, equalisation and loudspeaker positioning can create a very enhanced and life-like listening experience. These acoustic localisation cues can be referred to as inter-channel time and level differences (ICTD and ICLD respectively). As such, there are a significant number of studies that suggest guidelines and findings (Rory & Hyunkook, 2017). Rory & Hyunkook (2017) devised an experiment to determine the impact of amplitude across a vertically positioned phantom image on localisation. They concluded that "*The results of the study showed that the localisation thresholds obtained were not significantly affected by sound source or presentation method. Instead, the only variable whose effect was significant was interchannel time difference (ICTD)*".

Stereo channels are commonly stored in a matched (paired) format and are widely accepted by most of today's technology.

1.1.3.b Surround Sound

Primarily referring to the reproduction of sound through loudspeakers, surround sound is the reproduction of audio through a set of multiple loudspeakers which attempt to create a sense of an audio field using positioning and frequency-dependent filtering. Within this work, to differentiate from other immersive or '3D' sound systems, surround sound will refer to modern sound systems following the '*n-m*' format described in the 2012 international standards ITU-R BS.775-3 (ITU, 2012). This denomination defines the number of front channels versus rear and/or side channels. A more commercial and domestic nomenclature for defining the number of channels is a decimal point configuration, e.g. 5.1. This defines the number of channels between main (5), such as left and right, and low-frequency channels (commonly noted as low-frequency effects, LFE) (1) often in the form of a woofer or sub-woofer loudspeaker. All cases of surround sound have suggested guidelines for loudspeaker placement, and a basic positioning for a 5.1 surround channel setup is shown in Figure 3.

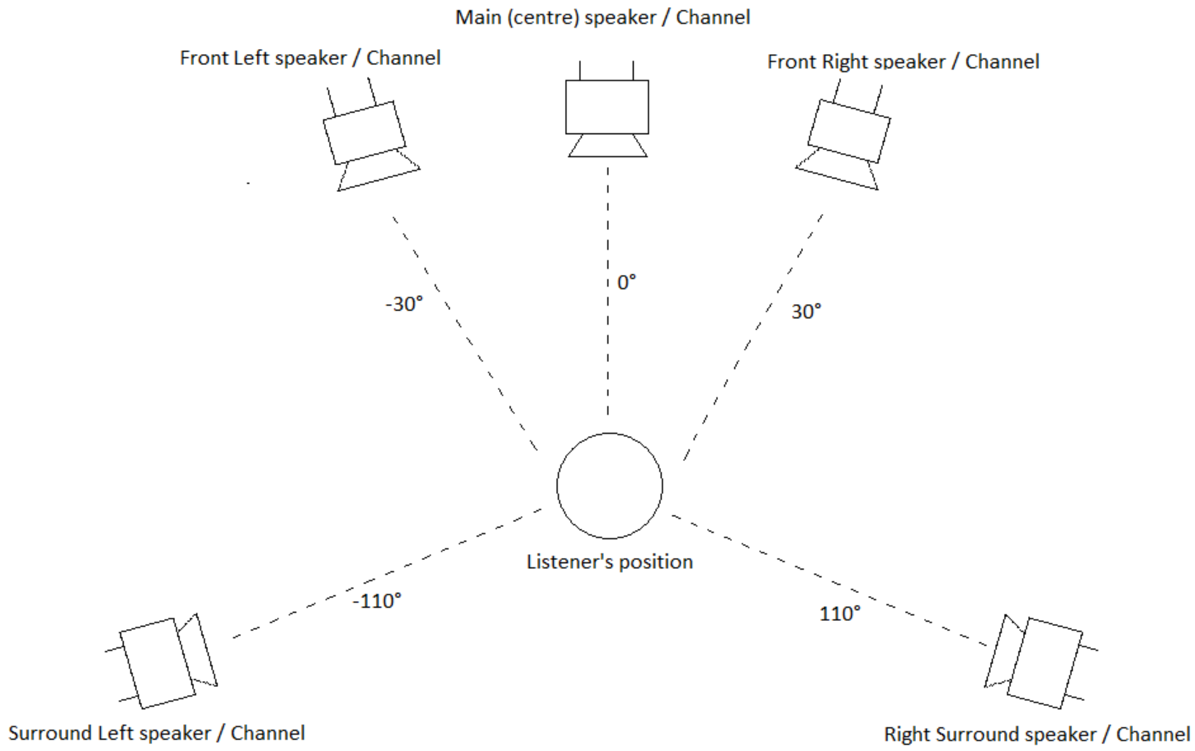


Figure 3 Example of a 5.1 loudspeaker positioning for a surround sound listening environment

This widely distributed 5.1 channel surround system, or in '*n-m*' format described as 3-2 stereo, is available on most Blu-ray and on-demand media. The extended ITU-R BS.775-3 international standard represents the channels as 3-2-1, the latter number representing the LFE channel mentioned above. This system is considered as the optimal configuration of sound based on availability and fidelity for the widely distributed audio formats (Griesinger, 2001). By this definition, the system consists of the following components; (i) two loudspeakers (left and right) at approximately 30° either side of the central listening direction which also double up for stereo compatibility, (ii) two 'surround' loudspeakers at 110° commonly dubbed LS and RS for left-surround and right-surround respectively, (iii) a centred loudspeaker at the sound source at 0° and finally (iv) a low-frequency effects channel with little suggested positioning owing to the low level of spatial information present. This channel is generally used for frequencies below 120 Hz such as rumbling, owing to its efficiency in the reproduction of low frequencies.

1.1.3.c Synthesised/processed Binaural Audio and Modelling

The increase in digital technology has brought new methods of synthesising and replicating immersive and binaural audio. Binaural modelling, commonly named binaural beats in entertainment and media, is one of many digital signal processing (DSP) methods that synthesises binaural audio (Wahbeh et al. 2007). In relation to human-end-user consumption, these models aim to apply processing techniques to recreate the spatial information and characteristics contained in 'real-world' sound, or even to create the impression of a false direction-of-arrival of a sounds' source and thus provide a sense of immersion in space to the listener. Alternative studies have investigated the manipulation of sound metadata in sound localisation technology, for robotics, and the evaluation of auditory scenes in military uses (Keyrouz, 2014), (Abouchacra et al., 2001).

There are extensive studies that aim to apply such binaural processing to a range of applications (Lim et al., 2018), (Kokkinakis, 2018), (Gantt, 2017) such as, but not limited to, speech technology, hearing aids, research tools and audiology (Blauert, 2013).

1.2 Aim and Objectives

The Aim of this research is to propose a methodology towards developing a standardised testing procedure for use with any given binaural system which seeks to improve binaural audio, for human application. This standardised methodology for testing binaural systems will allow consideration of a system's viability, and comparison with other existing systems. The Objectives to achieve this are:

- Determine the viability of evaluating psychoacoustic testing for spatial awareness to support further work in understanding localisation abilities in humans and furthermore provide a basis for a standard testing methodology

- Investigate and determine acoustic factors that could negatively influence results in the testing of a binaural system
- Investigate and determine human factors that could negatively influence results in the testing of a binaural system
- Develop a standardised testing environment that evaluates binaural systems
- Develop a method of validating test subjects for consideration of binaural systems, supporting the testing environment above

1.3 Thesis Outline

The following contents, in order of appearance, present the work undertaken within this thesis:

(i) The preface provides the terminology, relevant acronyms and nomenclature used, a list of tables found in this thesis and a table of contents of other chapters and sections.

(ii) Chapter 1 introduces the concept of binaural audio and its background and sets out the Aim and Objectives of this thesis.

(iii) Chapter 2 presents a current literature review of binaural audio, along with possible testing and measurement methods for binaural systems. Latterly, it defines the novelty of this work and summarises the necessary requirements for proposing towards a standardised testing environment for binaural systems.

(iv) Chapter 3 suggests and speculates some of the potential methodology, findings, observations, and outcomes of this work. Furthermore, it provides a hypothesis on the key factors required towards proposing a standardised testing environment.

(v) Chapter 4 presents a measurement model and a blueprint for a testing environment for evaluating binaural systems.

(vi) Chapter 5 outlines the process of creating a testing environment along with the methodology undertaken during the experimentation process.

(vii) Chapter 6 presents the raw results, and a preliminary analysis, following the experimentation procedure.

(viii) Chapter 7 further analyses the results from Chapter 6 and considers the reliability or justification of various results. Furthermore, Chapter 7 discusses the aforementioned results in conjecture with the theory and hypothesis presented in Chapter 3.

(ix) Chapter 8 summaries and concludes the key findings and contributions of the work presented in this thesis. Chapter 8 also suggests some directions for further work based on the findings, or challenges faced, during this research.

(x) Lastly, the relevant references and appendices are provided at the end of this thesis.

1.4 Summary

This chapter introduced the initial concept of binaural audio in broad terms and its origin for the purpose of this work within its rightful area of study. Furthermore, the chapter highlighted the Aim and Objectives of the research and defined the scope of the project undertaken. Additionally, it defined the contents of this thesis and their respective running order. The next chapter focuses on a more in-depth literature review of the current work and research undertaken in the field of binaural audio.

Chapter 2: Literature Review

The previous chapter outlined the scope of the project and the novelty of the research in its field. This chapter reviews the current research in binaural audio as well as the necessary prerequisites towards proposing a standardised testing methodology for current, and future, binaural systems.

As live broadcast has international standards for broadcast loudness levels or medical hearing tests have universal procedures of measuring frequency response, current research indicates there are no such set standards for evaluating and defining the efficiency of binaural systems. As such, the broader subject of binaural audio is investigated.

Binaural audio, a subgroup of immersive and spatial audio as seen in Chapter 1, is first and foremost defined by the ability to receive audio cues from a single sound source and approximating the direction-of-arrival (DOA) using such cues based on a minimum of two sensors (ears/pinnae).

2.1 Head-Related Transfer Functions

The interaural sensory reactions contain information that is analysed by the brain using the minute differences between both ears which can be mathematically modelled. In the frequency domain these are referred to as head-related transfer functions (HRTF) and head-related impulse responses (HRIR) in the time domain (Xie, 2013). Many proposed and revised mathematical functions have been used to define these relationships as discussed in the following sections (Hao, 2007); (Zhong, 2013); (Blauert, 2013). Xie 2013 defined the formula for calculating HRTFs based on the pressure at the two ears, omitting the effect of the torso and assuming a spherical head:

$$H_L(\theta, f) = -\frac{1}{(ka)^2}\sum_{l=0}^{\infty}\frac{(2l+1)j^{l+1}(-1)^l P_l(\sin\theta)}{dh_l(ka)/\,d(ka)} \qquad (2.1)$$

$$H_R(\theta, f) = -\frac{1}{(ka)^2}\sum_{l=0}^{\infty}\frac{(2l+1)j^{l+1}(-1)^l P_l(\sin\theta)}{dh_l(ka)/\,d(ka)} \qquad (2.2)$$

Where H_L and H_R are the left and right HRTFs respectively, θ is the azimuth, f the frequency, $P_{L/R}$ the frequency in free-field sound pressure, k is the wave number (2πf/c), a is the head radius (m) and h_l (ka) is the lth-order 2nd kind spherical Hankel function (Xie, 2013).

Roginska defines a more modern formula (Equations 2.3 and 2.4) for the relationship of head-related transfer functions in her 2018 study (Roginska, 2018) where θ is the azimuth, ϕ is the elevation, d is the distance from source, ω is the angular frequency, Y_L & Y_R are the spectra of acoustic signals at the listener's ears, H_L & H_R are the HRTFs (Left & right respectively) and X is the spectrum of the sound source:

$$Y_L(\theta, \emptyset, d, \omega) = H_L\ (\theta, \emptyset, d, \omega)X(\omega) \qquad (2.3)$$

$$Y_R(\theta, \emptyset, d, \omega) = H_R\ (\theta, \emptyset, d, \omega)X(\omega) \qquad (2.4)$$

Roginska (2018) then describes that HRTF is extracted through the cross-correlation of the input with the output, resulting in:

$$H_L\ (\theta, \emptyset, d, \omega) = Y_L\ /\ X(\omega) \qquad (2.5)$$

$$H_R\ (\theta, \emptyset, d, \omega) = Y_R\ /\ X(\omega) \qquad (2.6)$$

Equations 2.5 and 2.6, "*where the process of localizing a sound source can thus be described as the extraction of (θ, ϕ, d) based on the information contained in Y_L (θ, ϕ, d, ω) and Y_R (θ, ϕ, d, ω)*" (Roginska, 2018).

Yu (2018) describes the following HRTF observations and trends, adapted from an earlier experiment conducted by Brungarts (1999). Overall magnitudes of HRTFs present on the same lateral hemisphere increase as the source distance decreases, particularly sub-1kHz, whereas the magnitudes of HRTF at opposite hemispheres decrease with a closer source distance. Therefore, HRTF levels are expressed as distance-dependent ILD/IID cues (Yu, 2018), (Brungarts,1999).

The human hearing system is able to differentiate the phase difference in audio signals detected in each ear. This inter-aural time difference (ITD) for humans, varies minutely depending on the threshold of each individual and be can be as low as 10 microseconds, as discovered by Helmut Haas (Haas, 1951). Sound localisation features are a key component when attempting to reproduce the effectiveness of human hearing and the ability to detect a sound's location, to a minimum 15° azimuth (Perrott & Saberi, 1990); (Plack, 2005); (Mills, 1958). The following interaural differences are largely based on the initial research conducted by Lord Rayleigh in 1907 (Rayleigh, 1907). A leading use of this, is to emphasize hearing cues that are compiled to detect the source of any given sound.

2.1.1 Distance/Time

The interaural time difference (ITD), defines the azimuthal degree of any given sound source along the horizontal plane. ITD is the time difference between a sound arriving at one ear, and the other ear, owing to the separation of ears by the head (Figure 4) (Gelfand, 2010). Therefore, the maximum possible difference is at 90° azimuth. Fedderson et al. (1957) describe the time scale to be approximately 650 microseconds, giving leeway for various head dimensions. The precedence effect is determined by the strength of the delay between the signals arriving at the two ears and is easier to locate if it is in the range of 2 – 40 milliseconds (Wallach et al., 1949).

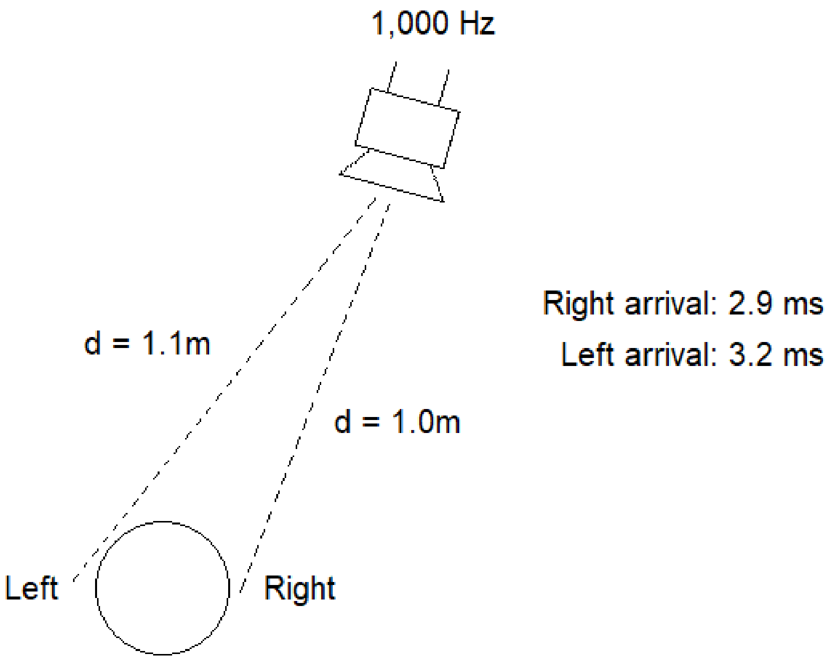


Figure 4 The difference in arrival times between the left and right ear

Any shorter delay between the two sounds (e.g. source directly in-front) greatly diminishes a person's ability to locate the sound in space beyond the rough estimate by a process of elimination, or through other cues such as vision. Blauert (1983) suggested that any difference above 50 ms is perceived by the brain as two different sounds, thus removing the ability to localise the source. The ITD is one of three factors that enable us to locate any sound source, thus allowing us to fundamentally understand and reproduce binaural audio.

$$ITD_{n,p}(w) = \frac{1}{w}(\angle\frac{X_n^r(w)}{X_n^l(w)} + 2\pi p) \qquad (2.7)$$

Equation 2.7 shows the proposed calculation for ITD (in seconds) (Zhou et al., 2011). Based on the right and left spectra of the n-th frame, the integer p is the phase unwrapping factor, which is a priori unknown, *w* is the angular frequency, X_n^l and X_n^r are the short-time fourier transforms (STFTs) of the left and right channel of the binaural signal respectively.

A more accurate equation for calculating ITD proposed by Howard & Angus (2009), taking into consideration the travel delay around a subject’s head, is given in Equation 2.8.

$$ITD = \frac{r(\theta + \sin(\theta))}{c} \qquad (2.8)$$

Equation 2.8 for calculating interaural time differences takes into consideration the path around an assumed spherical head where r (metres) is half the distance between the pinnae, c is the speed of sound (metres per second) and θ (radians) is the angle of arrival of the sound from the median (Howard & Angus, 2009).

Howard & Angus (2009) therefore determined that the maximum level of ITD (occurring at +/- 90° radians azimuth) can be calculated as:

$$ITD_{max} = \frac{0.09\,m \times(\pi/2 + \sin(\pi/2))}{344\,ms^{-1}} = 6.73\, \times 10^{-4}\, s\,(673\,\mu s) \qquad (2.9)$$

2.1.2 Loudness/Intensity

The interaural intensity difference (IID), occasionally referred to as interaural level difference (ILD), defines the location of a sound based on the level/amplitude of the arriving signal and its difference between each ear. Humans are able to locate the physical distance of a sound based on the directivity and/or reflections of a signal based on the arrival ratio at the ear. This allows us to locate sounds even in enclosed environments owing to the theory proposed by Helmus Haas in his doctoral thesis (Haas, 1951). The extreme differences in loudness in an ear based on proximity, such as a whisper, can also be determined through the comparison of the sound to a relative sound from further away. This factor is very limiting and often has drawbacks particularly for a moving sound source. A listener perceives a closer sound to move faster than a distant signal, thus creating the acoustic counterpart to the visual concept of motion parallax (Schwartz & McDermott, 2012).

$$ILD_n(w) = 20\log_{10}\left|\frac{X_n^r(w)}{X_n^l(w)}\right| \qquad (2.10)$$

Equation 2.10 shows the proposed calculation for ILD (in dB) (Zhou et al., 2011); where *w* is the angular frequency, X_n^l and X_n^r are the STFTs of the left and right channel of the binaural signal and *n* is the *n*-th frame of ILD.

As described in the previous section, Howard & Angus (2009) also revised the IID/ILD calculation to include our relevant head circumference and its respective additional travel time. Howard & Angus determined that there is a minimum frequency below which the effect of interaural intensity difference is useful for localisation, where the head diameter is approximately 1/3 wavelength in size. As such, a head diameter (width) of 18 cm corresponds to a minimum of frequency of:

$$f_{\min\,(0=\pi/2)} = \frac{1}{3}\left(\frac{c}{d}\right) = \frac{1}{3} \times \left(\frac{344\,ms^{-1}}{0.18m}\right) = 637\,Hz \qquad (2.11)$$

Equation 2.11 (Howard & Angus, 2009) shows the minimum IID frequency for localisation, where f is the frequency (Hertz), c is the speed of sound (ms^{-1}) and d is the distance between ears (m).

Howard & Angus (2009) concluded that IID is a cue for direction at high frequencies, whereas ITD is a cue for direction at low frequencies.

2.1.3 Timbre and Frequency

The frequency of the wave determines whether we can process and evaluate the directionality of a sound owing to the phase difference between our receivers (pinnae/ears). This additional spectral information adds to the perceptual information when attempting to localise any given direction of sound.

$$c = f \times \lambda \qquad (2.12)$$

Equation 2.12 (Beranek & Mellow, 2012) shows the calculation for the speed of sound; where the frequency is f (Hertz), λ (metres) is the wavelength and c is the acoustic velocity, (m/s). At an atmospheric pressure of 1 atm (1013.25 mbar), and a temperature of 20°C, c will be approximately 343 m/s.

The human auditory system effectiveness is dependent on frequency, owing to the size of the average human head and its related wavelength, from one pinna to the other. Humans can only discern a distinct change in phase when the wavelength is up to double the subject's head width. The distance between each pinna, approximately 18-22 cm for adults, allows us to accurately locate a source under approximately 770 Hertz (Wang & Brown, 2006). Many musical instruments fall below this frequency, along with the musical pitch standard tuning note A at 440 Hz.

2.1.4 Summation and Crossover/Trading of Interaural Cues

Understanding the combination of these cues is crucial in recreating near-perfect sound localisation. The time, or phase delay, only works during the low range of frequencies and transitions into an interaural intensity difference over the higher frequency range. This crossover begins at around 700-800 Hz where both cues function contrastingly until 2800 Hz where interaural intensity differences predominantly take over, owing to the wavelength corresponding to our head dimension, allowing us to differentiate the signal level drop from one pinna to the other. Furthermore, these two functions create a crossover range of frequencies that reduce the localisation effectiveness (Howard & Angus, 2017). However, both of these functions still restrict our ability to differentiate between the front and rear of the listener, assuming other external cues are not present (such as vision), sometimes referred to as the ‘cone of confusion’ (Plack, 2005) (This is discussed in more detail in section 2.4.2). At further, extreme, or constant frequency tones such as a pure sine wave, the human auditory system struggles to locate sound as effectively owing to the undetectable change in phase (Blauert, 1983).

2.1.5 Head-and-Torso Related Transfer Functions

Any sound in the free-field domain is subject to the acoustic environment, i.e. reflection and refraction. This includes the physical space we occupy in this environment, thus affecting the characteristics of a sound wave and the way we perceive it using inter-aural functions. Normally this only affects localisation in extreme cases of any given acoustic interference such as reverberation. The nature of our forever-changing environment therefore rarely impacts the ability to localise a given sound source. There is, however, evidence to that suggest that the torso, and to some extent the rest of our body, strongly impacts on the resulting audio heard by our ears, and hence the HRTF calculations with included head-and-torso (HAT) models (Gumerov, 2002).

Chen et al. proposed measurements for calculating extended HRTFs included with a head-neck-torso (HNT) model (Chen et al., 2012). Chen et al. state that there are discrepancies in results from standard HAT and HNT responses and concludes that the influence and function of the neck should be included in the calculation of near-field HRTFs.

$$D(r_0, f) = 20\ log_{10}\left|\frac{H_{HNT}(r_0,f)}{H_{HAT}(r_0,f)}\right|\ (dB) \qquad \textit{(2.13)}$$

Equation 2.13 shows the evaluation of discrepancies in HRTFs of HAT and HNT magnitudes (Chen et al. 2012), where r is the position vector and f is the frequency (Hertz).

2.1.6 Database Storage and Formatting of HRTF (and HpTF) Datasets

The demand for information on how a human locates a source of sound to improve immersive audio, particularly binaural, has led to extensive data gathered from studies into HRTFs. The tedious process of gathering HRTFs through physical/mathematical measurements and inaccessibility or transportability of a system has produced many public libraries and open-access databases;

AUDIS – A European Union-funded project *Auditory Displays*, provides data taken from collecting HRTFs using binaural recording and human responses of 20 subjects. These were conducted at a distance of 2.4m, following 10° and 15° azimuth vertical and horizontal spacing respectively (Blauert et al., 1998).

ARI – The database of HRTFs gathered by the *Acoustics-Research Institute* constitutes samples from over 70 subjects in 1550 positions ranging from 5° azimuth in both vertical and horizontal planes. The full database and other documentation are available online (Majdak et al., 2017).

KEMAR – One of the earliest databases, the *Knowles-Electronics Mannequin for Acoustic Research* contains over 800 samples of HRTF measurements available online based on 10° increments of horizontal azimuth and 5° increments along the vertical azimuth at a distance of 1.4m (Gardner & Martin, 1995).

PHOnA – A publicly available archive of headphone-transfer functions (HpTFs) provided by the *Princeton Headphone Open Archive* contains measurements from extensive studies compiled by PHOnA to provide optimisation of immersive audio headphone reproduction (Boren et al., 2014).

AUDIS have released a suggested set of recommendations for measuring HRTFs (dubbed 'Golden Rules'), based on their research, available at (see footnote)[1].

There are many other databases and libraries of HRTF measurements being developed or expanded, many of which opt to store them in varying formats, commonly in *.wav or relevant Matlab® file extensions. This distribution of HRTFs has brought forward a standard of storage and file exchange, published by the Audio Engineering Society (AES) in 2015. AES69-2015 describes the procedure of storing spatial audio information for head-related transfer functions (HRTFs), directional room impulse response (DRIR) and other more demanding systems in terms of response complexity (AES69-2015, 2015).

2.1.7 Binaural Localisation in the Vertical Domain

The literature above refers exclusively to localisation abilities within the lateral, horizontal domain owing to the complexity, limitation and lack of technology for binaural audio in the vertical domain at present. Therefore, this thesis will be based on the theory and literature in respect to localisation strictly in the horizontal domain/plane. The exact details and issues concerning localisation with elevation, as well as attempts at overcoming these, are discussed in section 2.6.1.

2.2 Binaural Audio Capture

Modern methods and techniques for capturing binaural audio aim to simulate the way our brain differentiates between the two interaural varying signals, at each sensor (ear). This is

[1] http://dx.doi.org/10.5278/VBN/MISC/AUDIS, Accessed: 31 August 2019

most commonly achieved through two omni-directional microphones placed in a life-size dummy head. This is undertaken to reproduce the time difference between an audio signal arriving at each ear, which are described by head-related transfer functions (discussed in section 2.1).

Figure 5 shows an industry standard binaural microphone dummy-head, the Neumann KU-100. These are positioned in a hypothetical subject's location within an environment, thus capturing the different arrival times and intensities of a signal at each ear, or more accurately each channel (left/right). Various other manufacturers design similar systems that seek near-full immersion or the capture of binaural audio.

Figure 5 Industry-standard Binaural recording 'dummy-head', the Neumann KU-100, (taken from; https://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description)

The 3Dio system omits the dummy-head-like features and only uses the pinnae reproduction and functions to capture superimposing HRTFs (3Dio, 2019).

The KEMAR (Knowles-Electronis Mannequin for Acoustic Research) dummy-head developed in the early 1970s is the first and still the most prominent head-and-torso (HAT) simulator (KEMAR, 1972). Designed with the intent of acoustic and audiology research, the KEMAR dummy-head has been used in a plethora of research studies of binaural audio and its relevant HRTFs.

Other manufacturers look at capturing the essence of binaural audio more directly at the source and do so through the process of embedding capsule microphones in the ear canal of a human subject, suggesting that personalised HRTFs are recreated to their exact physical dimensions and characteristics (i.e. pinna structure) (Roland CS-10EM, 2017).

Some researchers and manufacturers believe that extended HRTFs in systems such as the head-and-torso (HAT) are superior to their torso-free counterparts and have designed their systems accordingly. The effectiveness of different systems such as HAT versus the bar-and-pinnae (BAP) are discussed and considered in further detail in section 2.6.1.

Figure 6 Bar-and-pinnae binaural microphone, the 3Dio Free Space Binaural Microphone (taken from; https://3diosound.com/collections/microphones/products/free-space-binaural-microphone)

Other, more unconventional, binaural methods aim to capture the sense of audio immersion using other techniques such as microphone arrays and simulation. With the increasing demand for immersive, and therefore '3D', media and given that one of the leading applications of binaural technology is entertainment and media, specifically virtual reality (VR), it is important to be aware how such systems may complement VR technology in creating full-immersion of "3D" audio. Examples of 3D audio capturing systems include the *Ambeo VR Mic* (Ambeo, 2017), by *Sennheiser* and the H3-VR by Zoom (Zoom, 2018). These systems capture audio using four identical microphone capsules, allowing the user full/multi-channel ambisonics[2]. Furthermore, they also have binaural simulation recording modes which engage two of the microphones to act as the receivers thus recreating binaural recording (albeit at a lower efficiency). It is important to note that this is still, only a form of surround sound, thus it is only capable of creating a perceived location of a sound based on approximate fields around the listener. This does not provide the directional and detailed localisation information

[2] 'Full-sphere' surround sound, including both vertical and horizontal planes

contained in HRTFs to specific areas or degrees of azimuth or elevation, unlike more accurate binaural systems.

2.3 Binaural Audio Playback/Reproduction

Currently, the most effective approach to binaural reproduction is through the conventional use of headphones. In stereo, by definition, the left channel is transmitted to the left ear, and the right channel to the right ear, ideally in perfect separation from not only each other, but also any external noise. It could be argued that any stereo signal is effectively binaural, albeit to a low degree. For the purpose of audiophiles, and of this research, binaural audio will refer to the distinct characteristics audibly present which convey some information of spatial awareness, physically referred to as head-related transfer functions (see section 2.1). Assuming the ability to perfectly capture these HRTFs, a reproduction of a binaural signal needs to be achieved in a way which will retain all the minute physical differences of the audio received at each receiver (ear). These differences (HRTFs), are a relation of time, amplitude (intensity) and timbre (colouration/ unique spectral characteristic) between one receiver and the other. The design theory of headphones (and earphones for that matter) works on the principle of isolating the left channel from the right, and vice-versa. This bi-phonic reproduction, originating from the music theory term for two distinct lines, of an audio signal allows for a more accurate representation of the basic human hearing.

2.3.1 Headphone Methods

Roginska (2018) states the most effective method of reproducing binaural accurately is through the use of in-/over-ear monitoring and playback devices, such as headphone, (Roginska, 2018). The ability to reliably isolate the right and left channels to their respective receivers (ears) delivers a certain level of clarity which reproduces the natural hearing functions of human audio-localisation. This procedure ensures that the right channel is delivered to the right ear and the left channel to the left ear whilst also avoiding cross-talk cancellation, an acoustic effect whereby two signals in precise phase relationship cancel each other, which in turn would distort the auditory image (Elliott et al., 2016). Theoretically, perfect isolation of channels to their respective receivers should provide life-like reproduction of the HRTFs present in human hearing. Most headphones provide a controlled listening environment allowing us to isolate other possible interferences such as background noise. This design characteristic of headphones proves an advantage in being able to relay each channel directly to its intended ear. The following sub-chapters are the current types of headphones, as summarised by Roginska (2018).

2.3.1.a Closed Headphones

Closed headphones, whether of circum-aural or supra-aural structure, are designed to enclose the ears fully in a chamber in an attempt to isolate the listeners' ears from the environment. This acoustic isolation aims to reduce the environmental noise to improve the users' attention to the intended audio. These sealed headphones provide approximately 10 dB level of isolation, with improved efficiency in the higher frequencies (Roginska, 2018). Further efficiency improvements can be achieved with in-ear headphones which are currently the most effective noise-cancelling option, at 23 dB of isolation. It is, however, important to note that some studies suggest extreme isolation in headphones can negatively impact the users' listening experience through the complete lack of ambience and acoustic environment (Roginska, 2018).

Figure 7 Closed-back headphones, Sennheiser Momentum 2.0 (taken from: https://www.rtings.com/headphones/learn/open-vs-closed-back)

2.3.1.b Open and Semi-open Headphones

Similar in design to closed (sealed) headphones, open and semi-open headphones differ in the seal of the ear-cups. Open-back headphones aim to allow for some environmental noise-spill in order to provide a sense of natural or spatial perception.

Figure 8 Open-back headphones, Sennheiser HD 650 (taken from: https://www.rtings.com/headphones/learn/open-vs-closed-back)

2.3.1.c In-ear Phones

The most accessible and distributed form of personal-audio reproduction in the current age are earphones, also known as earbuds. Unlike headphones, these are positioned right at the entrance of the canal, sitting in the pinnae, with the in-ear monitor variations being sealed into the ear canal anywhere from the entrance to halfway-in, referred to as the blocked-meatus method (Santos et al., 2014). Earphones are small in size and provide compact and portable accessibility of audio reproduction and can range significantly in quality. Although they theoretically re-create the most human-like hearing listening system, listener response varies largely owing to the difference in shape and size and proximity to the eardrum.

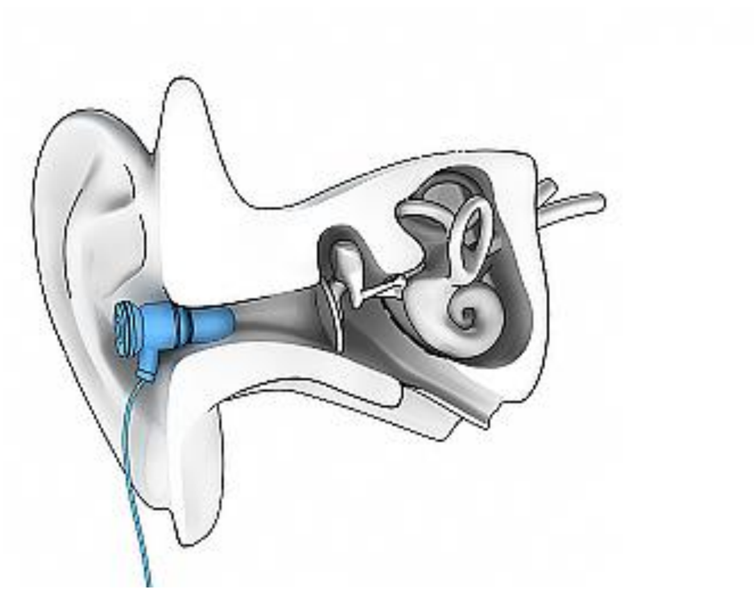


Figure 9 Diagram of in-ear phones

2.3.1.d Multi-driver

The surge in virtual reality, ‘3D’ gaming and media has brought forward a new application of headphones and therefore expanded its technology. The constant demand to improve binaural audio reproduction has sparked attempts at multi-driver headphone utilization. These headphones function in a surround sound configuration, aiming to dedicate different drivers to different respective frequencies. A 7.1 headphone surround system will normally consist of 10 drivers divided equally between each ear. In this example, each ear is comprise a subwoofer (low-frequency effects/LFE), centre, left, left surround and left surround-back (See 1.1.3.b).

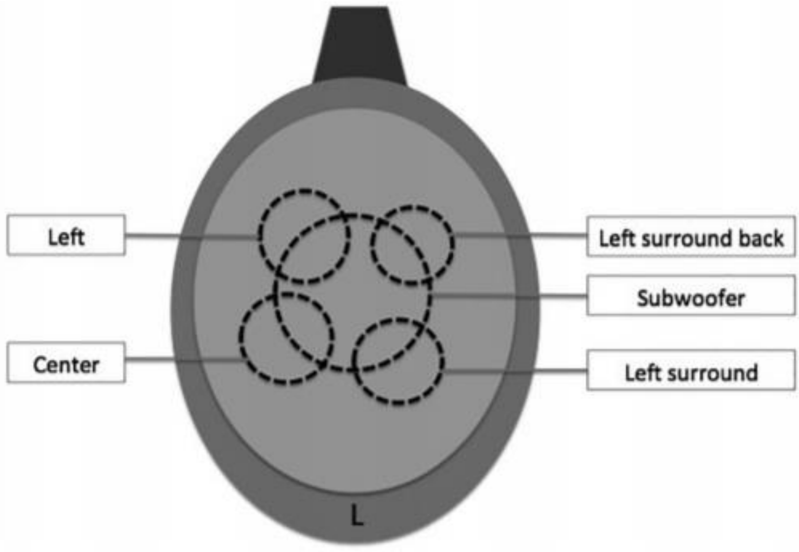


Figure 10 Multi-driver 7.1 surround headphones with driver positions (Taken from Roginska, 2018)

Regardless of headphone design, the HRTFs vary in each individuals' physical complexity (ear response) and therefore greatly impact the perceptual localisation of any given sound (Rumsey, 2011) (further information is given in section 2.6.2). The extent of these binaural cues depicts the quality and range of binaural audio through headphones. With the large variation of headphone types, listener experience will vary based on functionality and design.

All types of reproduction methods, as well as their different designs, can be calibrated and thus improved in terms of maximising localisation ability. These calibrations are a combination of frequency response efficiency, physical design (with stronger presence in circum-aural headphones) and resonance. These characteristics of headphones (known as colourations) are considered and minimised in an attempt to improve the spatial audio image (Rumsey, 2016). Calibration also extends to a listener's individual morphology, which defines personalised HRTFs of our structural characteristics (Roginska, 2018). The summation of these factors is referred to as headphone transfer functions (HpTF). Lindau & Brinkmann (2012) proposed using regularisation methods in order to compensate for some of these HpTF colourations in order to calibrate headphones (Lindau & Brinkmann, 2012). Even a slight re-positioning of headphones on a listener can greatly impact response functions and thus negate the calibration of headphone colourations.

In summary, the use of headphones provides a viable solution to current binaural reproduction and playback, however their very advantage also provides a disadvantage to certain applications of binaural audio. Furthermore, it could be argued that binaural audio through headphones is 'static' in respect to the fact that it is user-centric and does not allow for freedom of head-movement. If the user were to turn right (i.e. 90° azimuths relative to facing forward) to face a certain perceived location of an audio cue, the location would once again appear to be coming from the right (180° relative to initial forward-facing direction). This results in a loss of perception in space especially in virtual-reality applications. Inaccurate or improper calibration of binaural audio through headphones can result in deterioration of the perceptual audio image and results in a psychoacoustic phenomenon referred to as inside-the-head location (IHL) (Boren & Roginska, 2011), see section 2.4.1.b.

2.3.2 Binaural Audio Playback with Loudspeakers

The current viability of binaural audio is arguably only available on stereo headphones, even then the binaural element is to a very limited extent, owing to the alienation of each signal to its respective, matching ear. Listening to binaural audio on a medium other than headphones, introduces further challenges through acoustic influences which are not present in headphone reproduction, i.e. reverberation and deflection. Following the assumption that the audio was recorded using the standard method of dummy-head recording (containing the relevant superimposing interaural cues), the restricting factor is the automatic function of the brain attempting to locate the source of the playback in a new acoustic environment, thus dismantling the characteristics embedded in the two audio signals. This occurs owing because both ears are listening to both signals along with any possible acoustic reverb and deflection occurring in the listening space not strongly present in headphone reproduction, and therefore interpreting them as a joint signal rather than a single track intended for each ear, creating a form of crosstalk cancellation (XTC). This is the interference that contains data cues for the right ear being heard equally by both left and right ears. To reduce this effect, a form of "barrier" would have to be established to separate the two signals to each respective ear. A way of establishing this could be to create a filter system that prevents the crossing of either signal. This brings its own issues owing to the distortion of sound content being heavily filtered. This has yet to be perfected and is being investigated by industry-leading expert Dr. Edgar Choueiri of Princeton University (Choueiri, 2011). Choueiri aims to alleviate this issue through crosstalk cancelation as a means of reducing unintended levels of degradation in loudspeaker playback of binaural interaural cues. The study investigates the ability to apply filters to audio signals in order to direct them to each respective ear in a standard two-loudspeaker (stereo) setup. These optimal crosstalk cancellation filters (BACCH filters) assume that sound dispersion travels in a free-field environment, free of deflection, diffraction and absorption, thus emitting sound as a natural source. Here, the ideal XTC-filter is expressed in Equation 2.12.

$$H^{[P]} = C^{-1} = \frac{1}{1-g^2e^{-2i\omega t_c}}\begin{bmatrix} 1 & -ge^{-i\omega t_c} \\ -ge^{-i\omega t} & 1 \end{bmatrix} \qquad (2.12)$$

The research presented in this thesis shall not attempt to propose a standardised testing regime for binaural loudspeaker reproduction but instead look to propose a testing environment with the possibility of flexibility to expand to loudspeaker techniques. For the purpose of this thesis, the focus for investigating binaural systems will be chosen on current reproducible binaural headphone methods.

2.4 Psychological Testing Factors

Creating and researching factors towards proposing a standardised testing environment for a human end-user requires an unbiased protocol of experimentation. Various psychological factors can influence the biasing of results (negatively impacting them, such as the precedence effect) when exposing human subjects to testing, particularly of the senses.

2.4.1 Psychoacoustic Factors

Owing to the complex diversity of the human brain, and hence its hearing system, its function to interpret and convert audio signals to cues can lead to various psychoacoustic factors. These psychoacoustic factors are present in everyday hearing, as a result of the acoustic properties of the environment in combination with the human hearing function. As such, results can negatively impact the human ability to localise the direction-of-arrival of a particular cue.

Psychoacoustic measurements are generally non-HRTF experiments for binaural (diotic) hearing, often conducted on human subjects owing to the nature of subjective results. These results are a product of a chosen stimulus and thus an analysis method is created for their respective experiments. The stimulus is played to the participants through a chosen delivery method, most commonly as headphones or through an array of loudspeakers, and observation of the perceived location of the transmitter is noted. This is occasionally done through verbal communication of perceived azimuthal degree or in other cases with the use of head-movement trackers, with the participants being asked to face the perceived location (Choueiri, 2011), this is addressed further in section 2.5.2. When conducting either form of measurements, verbal or tracking, it is essential to eliminate any biasing effects which could impact the outcome. Many external effects which would influence psychoacoustic testing in humans, or to some extent animals, and create incorrect results are considered and approached in section 3.2.

Some psychoacoustic studies on the localisation abilities of humans based on spectral frequency and colouration have observed certain patterns in localisation perception. Iida et al. (2007) state that many studies have shown spectral information is the cue for localisation on the median (horizontal) plane. Furthermore, they state that previous studies found that spectrum changes, i.e. distortions, exceeding 5 kHz function as cues for localisation in the median planes (Blauert, 1969), (Hebrank & Wright, 1974), (Gardner & Gardner, 1973), (Middlebrooks, 1992). Iida et al. 2007 further defined the peak of HRTF to occur at approximately 4 kHz and suggests that humans interpret, or analyse, other notches/peaks based on this reference point (Iida et al., 2007).

2.4.1.a Haas Effect

The Haas effect, or precedence, of the first wave-front is a psychoacoustic hearing effect. This occurs when two sounds of the same perceived loudness, but varying distance from the receiver (ear), appear to be arriving from a single location, that of the shorter-distanced source (Haas, 1949, 1972). Haas discovered that any reverberation or reflected sound within an acoustic space is perceived as the same location with greater amplification with the pretence that it occurs within 35ms of each ear. This only applies if two arriving sounds are within the range of 20-40ms, above the minimum human threshold. Time variations beyond 50-80ms result in distinct echoing effects (Everest & Pohlmann, 2015). This is a natural phenomenon often undesired in the capture of binaural audio, however it is crucial to understanding why some results in a real-world scenario may vary, particularly in the reproduction of binaural audio (or any immersive audio for the matter) through loudspeakers.

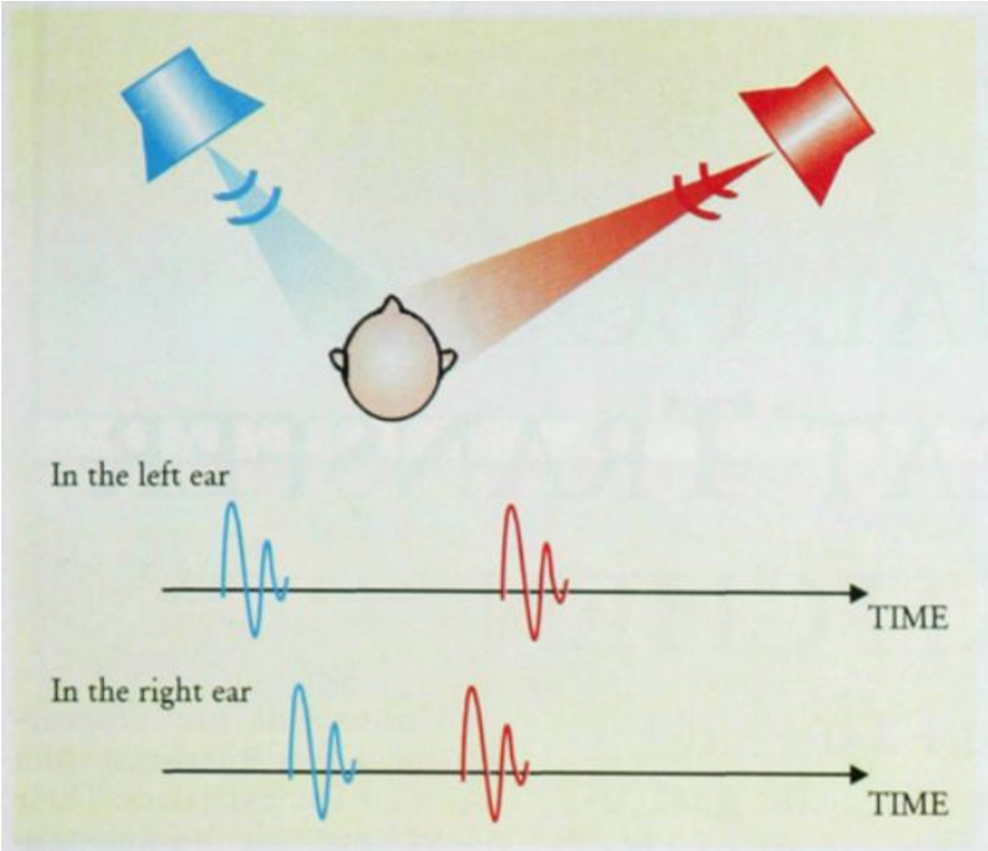


Figure 11 The precedence effect as described by W. M. Hartmann, the left channel dominates thus the sum is perceived as arriving from only one location. (Hartmann, 1999)

Howard & Angus (2009) conclude the Haas effect as follows.

- "*The ear will attend to the direction of the sound that arrives first and will not attend to the reflections provided they arrive within 30ms of the first sound.*
- *The reflections arriving before 30ms are fused into the perception of the first arrival. However, if they arrive after 30ms they will be perceived as echoes.*" (Howard & Angus, 2009)

Some studies such as (Madsen, 1990), (Howard & Angus, 2009) refer to the Haas effect as one of two 'ITD and IID trading effects'. The other effect describes the relationship of time versus loudness trading. It defines the effectiveness of both ITD and IID when both are present, stating that the former exclusively functions only within the maximum time delay of 673 µs (as seen in section 2.1.4).

2.4.1.b Inside-the-head 'Locatedness'

A well-documented audio phenomenon dubbed 'inside-the-head locatedness' (IHL) or internalisation, is a psychoacoustic defect present in human hearing, particularly present in the reproduction of audio through headphones. The most natural form of this sensation is heard from a subject’s own speaking voice. This of course seems natural as it is present throughout a subjects’ life. However, with other sounds that are normally external, it causes a strange sensation and disturbance when listening to audio, owing to the inversion of localisation. This perception, or psychological illusion, occurs when the localisation cues in a diotic simulation are nearly identical to each other. As such, this effect is dominantly present in headphone reproduction owing to the lack of distance and acoustic field found in loudspeaker reproduction. Other studies speculate and suggest that IHL is also caused by bone conduction and resonance created by wearing headphones (Ebata et al., 1968).

Naturally, loudspeaker reproduction alleviates the issue of IHL but can still be achieved under certain circumstances (Hanson & Kock, 1957). The lack of inter-aural differences is therefore in conclusion perceived as originating from within the head. This eliminates a sense of perception, or to some degree a subjects’ ability to localise the direction of arrival of any sounds’ source when creating and/or reproducing immersive ‘spatial’ audio. The result of this occurrence is a loss of a sense of externalisation of sound. The use of artificial pinnae in ‘dummy-head’ recordings (binaural capture) is thought to currently be the most effective way to eliminate some levels of IHL (Durlach & Colburn, 1978). This replicates the precise acoustic properties of human ears.

2.4.2 Front-and-Back Confusion

A regular factor of sound localisation in humans is the inability to accurately distinguish between front and back sources of any particular sound. This ‘cone of confusion’ as it is often referred to, is a reversal error whereby locations in the front hemisphere and in the rear hemisphere are seemingly interchangeable in a subjective measure owing to the low levels of ITD and/or IID trading. Such confusion is also said to be present in cases of back-to-front as well as in the elevation/vertical domain (Oldfield & Parker, 1984), (Wenzel, 1991).

2.5 Current Testing Regimes and Procedures

There are numerous publications that investigate the need for a standardised method of testing binaural systems including Nicol et al., (2014) and Le Bagousse et al., (2011). Nonetheless, it would appear that there are currently no set standards or guidelines to follow when conducting experiments for sound localisation in humans or any given animals. Work by Salvador et al. (2017) considers a design theory for binaural synthesis and the evaluation of head-related transfer function datasets. Many publications devise their own methodology of assessing results, thus making it difficult to ensure their applicability to other situations or to compare them with rival systems (Moravec et al., 2018).

There are two main approaches that are taken when determining the ability of a binaural system to recreate binaural techniques. Firstly, a mathematical or physical analysis of the head-related transfer functions, and secondly, a psychoacoustic testing procedure based on the localisation abilities of any given subject through a perceived localisation ability.

2.5.1 Mathematical Measurements (‘Quantitative’)

Frequently researchers look to determine the viability of a given microphone system, or the ability to localise a sound, through the physical response rate of brain functions (Moravec et al., 2017). This practice is most commonly achieved through the measurement and calculation of HRTFs conducted through binaural microphones embedded in the pinnae, or alternatively through binaural dummy systems (and in some cases bone-conduction). Small flat-frequency-response condenser microphones are positioned in the entrance of a subjects’ ear canal and sealed in before being captured and analysed. Others embed the microphones deeper into the canal through a probe tube, with results suggesting the canal improves individualised HRTF features (see 2.6.2) (Hiipakka et al., 2012). The datasets sent by the transmitter determine whether spatial cues are contained within the information measured by the receiver in the form of interaural differences between the two microphone signals. However, this does not necessarily prove whether or not a human (or animal) is able to accurately locate a given sound on the respective system in question, but rather whether superimposed interaural cues are present in such a signal. Thus, the physical measurement of binaural cues can be defined as objective. Such binaural systems are often used for applications not explicitly pertaining to human hearing, but for localisation technology, audiology, further experimentation, etc. (Keyrouz, 2014). Owing to the vast variety of methods used to measure HRTFs, careful consideration is required to compare the range of existing measurement techniques. It is vital to be aware that many other methods have been conducted for a range of applications. Publications that look into determining localisation abilities and response rates in mammals have conducted tests through the vibrations and responses of the ears on a series of anesthetized animals based on their frequency response of HRTFs (Xu et al., 1999); (Rice et al., 1992); (Grana et al., 2017).

2.5.2 Psychoacoustic Measurements (‘Qualitative’)

Psychoacoustic measurement techniques for localising binaural audio (diotic hearing) are considered a subjective judgement of where any given sound is arriving from, to a human listener’s perception of it. Unlike mathematical quantitative measurements, these cannot be divided into physical attributes such as frequency or interaural differences. These are driven by emotion and reflex response instead. This perceptual judgement of a sounds’ location is determined based on the colouration of a sound (frequency/etc. characteristics) and the accuracy of a sounds' source (Letowski, 1989). The perceived location of a sound (based on either elevation, distance and/or azimuth) is noted through verbal communication or through the use of head-tracking devices whereby the subjects are instructed to face the direction of arrival. (Choueiri, 2011). Lewald et al. (2000) used a visual laser pointer positioned on the vertical axis of the nose to indicate the accuracy of a location.

Psychoacoustic measurements for binaural (diotic) hearing are often conducted on human subjects owing to the communication of results. These results are a product of a chosen stimulus and thus an analysis method is created for their respective experiments. The stimulus is played to the participants through a chosen delivery method, most commonly headphones, and the perceived location of the transmitter is noted. This is occasionally done through verbal communication of perceived azimuthal degree or through the use of head-movement trackers, with the participants being asked to face the perceived location (Choueiri, 2011).

It could be argued that either of these methods, qualitative or quantitative, are suitable for varying applications, thus selection of which method to use (or both) is subjective to the requirement of the experiment and/or system. When conducting either form of measurements, qualitative or quantitative, it is essential to eliminate any biasing effects which could negatively impact the accuracy of a locating a sound. Many external effects, which could influence psychoacoustic testing in humans and create incorrect results, are considered in section 3.2.

2.5.3 Stimulus/-i

To the author's knowledge, there is little information available on the use of a given stimulus in an experiment of HRTFs or general auditory localisation abilities, nor are there many scientific justifications behind the use of a particular stimulus. Furthermore, there also appears to be no set standards, beyond some recommended/suggested theories, on the properties of a stimulus. As the perception of a sound is based on the relevant HRTFs present, it is therefore naturally easier to estimate the direction of arrival of a stimulus with larger interaural times and intensity differences. However, it is important to note that this refers solely to the ability to locate the DOA of the stimulus based on its angle of arrival and not its physical acoustical properties. The ability to localise the source of a stimulus is additionally based on its frequency properties, and to a minor degree, the subjective assessment of it, that is psychological association (recognition) (Blauert, 1983). Many experiments conducted to this date choose to use a quantitative method of measuring binaural localisation abilities, namely HRTFs, thus instead this section will highlight some case studies of the chosen excitations in the aforementioned research studies. Some case study examples of the stimuli used in research on the ability to localise sound are presented below.

Jiang et al. (2018) investigated human HRTF responses for vertical localization using pink noise. The experiment was conducted using stimuli with lengths of 5s each, on a total of 8 subjects, between the ages of 22 and 30 with reported normal-hearing, in a sound-proofed room with background noise levels of 30 dB. The stimuli were delivered at 75 dB SPL and the subjects were asked to give the perceived location of arrival. The location was noted using two laser head-trackers, the former determined the subjects' head-position and orientation, and the latter was a hand-held tool for the subjects to aim and point to the estimated location. Additionally, in the case that the stimuli appeared to be located within the head, subjects were asked to communicate results orally. The authors conclude and determine that front-back confusion is reduced significantly in the static reproduction of a full-bandwidth stimulus and is furthermore alleviated almost entirely in the case of dynamic reproduction (where subjects are encouraged to reposition themselves). Additionally, the authors draw similar conclusions from the results of up-down confusion, but to a lesser extent.

Yao & Chen (2013) conducted subjective listening tests on 15 subjects which were required to locate the perceived DOA of four different stimuli to investigate the relation between non-adjusted and adjusted HRTFs. The stimuli, a 2-second burst of white noise arriving from different directions in a random order, was played for a total of 24 samples, 18 of which were in unique locations. This procedure was repeated for all four of the HRTF settings (non-modified, modified to +/-5 dB, modified to +/-10 dB and lastly modified HRTFs to +/- 15dB). The results for the unmodified HRTF stimulus produced a mean of 60% with results deviating between 33% and 88%. Each of the four tests lasted approximately 5 minutes and the increasingly modified HRTFs produced more accurate results.

Yu's experiment investigating human HRTFs, was conducted on 56 Chinese subjects using a 24-bit quantised and 96 kHz sampled frequency logarithmic sweep signal (Yu, 2018).

2.6 Limitations & Challenges

Binaural audio, and its relevant technology, is constantly expanding and improving to meet the demands of a near-realistic immersion of audio in the 21st century. Some of the limitations and problematic areas of expansion are investigated.

2.6.1 Height Information

Localisation in the vertical (elevation) plane is naturally inefficient owing to the lack of interaural cue differences. The way the human brain can detect vertical changes of a sound source is through the shape of the pinna or auricle. The structure filters and provides a modified frequency response based on the direction of the incoming signal. The ability to interpret and localise sound in the vertical domain is dependent on the interaction of a sound with the structure of the pinnae, thus resulting in a colourisation of the monaural spectrum (Jiang et al., 2018). This seems to be limited (to the ability of the pinna shape) and provides six main directionality points, three directional and three refracted (Grothe et al., 2010), (Batteau, 1967). These superimposing monaural cues transpose and change the characteristics of sound that determine the location of a sound into the domain of the horizontal plane as opposed to the vertical (Klein & Werner, 2016), (Gardner & Gardner, 1973). Kim (2018) describes the ability to perceive an elevated sound source as dependent on the spectral modification produced from various reflections bounced off the head, shoulders and pinnae, thus conforming some of the conclusions found in head-and-torso (HAT) or head-neck-torso (HNT) experiments (Kim, 2018). Furthermore, the ability to localise sound through the vertical plane in a natural free-field domain is influenced, and in some cases improved, by the acoustic factors in the surrounding environment. The ability to adapt to the environment (e.g. physical repositioning) is limited in humans, particularly when attempting the playback of audio since the playback device will be static relative to the user, thus removing the precision when sound is localised on a vertical plane/elevation. Some animals (and to some extent humans) have the ability to utilise this, through the movement of their ears (Pena et al., 2001).

Immersive systems are generally categorised as binaural, synthesised (audio objects) or discrete channel based. Currently the attempt of height reproduction relies largely on discrete channel-based systems that aim to use microphone and loudspeaker arrays to accomplish a sense of elevation in audio.

2.6.2 Adaptability to Individual Users

Binaural audio is greatly limited owing to the differences created in an individual users' hearing range, head dimensions and auricle structure (Alberti, 2006). This creates a challenge for binaural audio, to expand for a mass audience and thus commercial viability. These individual user characteristics would change the brain functions which determine the location of a sound based on the informational cues received. For example, a stimulus recorded on a Neumann KU-100 dummy-head (Figure 5), with a corresponding ear-to-ear distance of 18 cm would function more accurately for users of a lower average head size. This would change the perceived location of a sound, by a varying amount of degrees' azimuth. Furthermore, the shape and structure of the auricle and pinna and ear in general, would distort localisation features based on directionality of the arriving sound source at the dummy-head.

Various studies have adopted calibration and adjustment methods to characterised HRTFs in order to determine an improved accuracy in user-individual based binaural audio (Rumsey, 2001), (Xie, 2013), (Orduña-Bustamante et al., 2018). Rumsey (2001) explains the process as time-consuming and impractical owing to the required controlled conditions and investigates whether subjects are capable of adapting and developing to the ‘foreign’ HRTFs on their own. Nonetheless, the use of individualised HRTFs in binaural reproduction for humans is proven to positively impact the ability to accurately localise the direction of arrival of a given sound. Unfortunately, the adaptability of individualised/personalised binaural audio, whether synthesised or physical, is yet to become available for the commercial industry owing to the practical limitations of wide-spread delivery.

2.7 Literature Overview

The broader concept of binaural audio can be widely interpreted and investigated in numerous ways. To outline the necessary prerequisites of working towards proposing a standardised testing procedure for systems which capture and/or playback binaural audio, the various, different, fields are condensed in this subchapter for the readers convenience. As such, the following elements are necessary for proposing a move towards creating a standardised testing environment:

(i) Defining and justifying the type of measurement which should be incorporated (qualitative, quantitative or both),

(ii) Defining the stimulus, or stimuli,

(iii) Defining the locations, and/or an order of locations, from which the stimuli should be delivered from,

(iv) Defining the subjects and environment the test should be conducted in/on along with any potentially negatively influencing factors,

(v) Defining a method of analysing, and verifying, results of such a test (i.e. scoring and/or point system for qualitative response).

2.8 Chapter Summary

This chapter provides an in-depth review of the current literature and the origins of binaural, both in terms of the basic analogue principle behind the process of localising and estimating the source of any given sound, as well as the technology to capture, recreate and playback binaural audio as an electrical signal. Furthermore, it defined the current characteristics and procedures of previously conducted experimentations on head-related transfer functions and general localisation abilities, particularly in humans. The next chapter will discuss the relevant theory behind establishing and proposing a standardised testing environment.

Chapter 3: Theory

Chapter 2 outlined the inner workings of binaural audio, its current evaluation procedures in terms of HRTFs, along with the potential influencing factors on binaural systems and their testing procedures. This chapter suggests and justifies the reasons for the methodology that will be outlined in Chapter 4, furthermore, it provides an estimation of the expected outcome against which the experimental results can be compared for seen in this work.

In an attempt to create an unbiased testing procedure, the following is proposed. Acoustic and psychological factors are raised, and measures are suggested to counteract issues raised in Chapters 1 and 2. The stimuli are considered and chosen based on their frequency properties and impact on human localisation abilities. The locations from where the stimuli are triggered from and their successive running order is suggested. A hypothetical method of validating a subject's ability to localise any given sound, as well as a subject's set of results, is proposed. A scoring system is investigated which allows for some degree of subjective error with a scaling factor for 'near-correct' answers, and finally, the theoretical probability of 'guessing' such a test is calculated, utilising a simulation to prove such figures.

3.1 Qualitative or Quantitative Measurement

The scientific definition of binaural audio is commonly presented in the form of mathematical measurements, namely HRTF. As described in the literature review, the majority of publications to date investigate the ability to localise the DOA of a source through recording the time and intensity differences between two ears and thus devise mathematical formulae to present such work. Although creating a testing environment for binaural systems does not strictly require quantifying a users' HRTFs, it can be considered vital information to be aware of which particular stimuli, or locations, are hypothetically easier or harder to localise. Ultimately, it is the subjective, qualitative, judgement of the end user that defines the efficiency of a binaural system.

"*Although quantitative methods are useful in measuring a signal's physical attributes reaching a listener's ears (e.g. spectral content, time attributes, etc.), it is a listener's perceived judgement of the quality of a spatial auditory image that is more relevant to the actual listening experience."* (Roginska, 2018)

A more psychological approach to investigating binaural responses, against which a binaural system should be tested, is to quantify a qualitative response to particular stimuli, or locations, through subject participation on the perceived judgement of a sound's location.

Following the two techniques mentioned above, it can be concluded that mathematical HRTFs are necessary, and thus recommended, when investigating how humans and animals localise various stimuli and DOAs and to what extent. Alternatively, when devising a system that is intended for human use, particularly entertainment, it is ultimately the qualitative perception of whether a user is capable of localising sounds and their respective DOAs, that is required. As such, the remainder of this thesis will primarily focus on the perceptual judgement of being able to localise a sound, rather than the measurement of physical HRTFs.

3.2 Evaluating Negatively Influencing Factors

The potential, superimposing, factors that could negatively impact the outcome or result of the testing environment are split into two categories (i) psychological impact of a testing environment and human procedural errors ("exam pressure"), and (ii) acoustic properties of the subject, space and stimuli, relating to the free-field space (reverberation, etc.). As such, it is suggested that the following choices are implemented.

Firstly, the testing of binaural systems (and hence therefore human localisation abilities), should be conducted in anechoic, or near-anechoic, conditions, i.e. in an isolation booth or anechoic chamber. This aims to remove the possibility of external influences, such as noise, owing to the isolation and simultaneously reduce reflection and reverberation. This is a common procedure not only in experiments on hearing (intensity versus frequency tests), but also in the testing of frequency responses in microphones, loudspeakers, and so on (Silman & Emmer, 2011), (Brittain, 1951), (Floyd, 2008). Alternatively, Algazi et al. (2001) state that HRTF testing (and therefore localisation abilities) can be carried out in untreated rooms and environments, providing the aforementioned space is free of many reflections or interferences.

The stimulus/-i (discussed separately in section 2.5.3), should ideally be triggered through sealed in-ear phones and/or high-fidelity loudspeakers depending on the testing stage. This ensures the direct delivery of the signals to each ear without any cross-interference or cancellation. Owing to the cost limitations, requirements and complications of in-ear phones, it is acceptable to use closed-back headphones and certain loudspeaker drivers providing they have a flat frequency response in the range of the excitations. The minor acoustic variations the headphones create between the pinnae and the driver are argued to be minimal for the approximation of a sound to the desired azimuthal degree in the work presented in this thesis. A system that demands higher accuracy (i.e. binaural hearing aids) requires the use of personalised HRTFs and is discussed in the further work section Chapter 8. Given the requirement for high fidelity (definition) in audio to maintain all the spatial characteristics captured in binaural audio, a binaural stimuli, and playback resolutions, should ideally exceed a minimum of 48-96 kHz sampling rate in computer file formats such as .wav or .flac, to ensure that the relevant transfer functions are present. There is currently no record of a set standard or recommendation for higher resolution samples. The sample rate and file format are based on the cases presented in section 2.5.3. Recordings should be captured using a sampling rate that is sufficient to prevent the loss of informational cues. The stimulus should be captured through the binaural system in question and played back through headphones as mentioned above. This stimulus should be triggered through a series of loudspeakers (an array) positioned around the system.

The loudspeakers should be positioned in a free-field environment, equidistant from the subjects at a range of 1m to 1.5m. This positions the subjects in the far-field environment (relative to the source) and enables the use of localisation abilities, particularly the approximation of distance. This far-field environment contributes to localising the source of a given sound through IID/ILD and frequency spectrum as opposed to exclusively through ITD as in the near-field environment. The increments of the loudspeakers degree of separation (per azimuth and/or elevation) is relevant to the intended use of the system in question. A system designed for entertainment, such as virtual-reality, may only require an accuracy of 10°- 15° whereas mission critical systems require a precision of 1°- 2°.

The results of a subject's own perception and interpretation of the direction of arrival (DOA) of any given sound should ideally be done by instructing them to face the perceived DOA, which can be identified by a head-tracker/laser pointer that records the degrees of azimuth (and/or elevation where necessary). Alternatively, the results should be communicated to, and noted by, an external observer whereby the loudspeaker locations are numbered and/or lettered, and the subject defines the location based on the perceived number/letter. These methods ensure that the test is 'blind' where the subject is unable to see their previous results to avoid the psychological tendency of being reluctant to answer the same number/answer multiple times, which is known as response bias of a Likert scale (Likert, 1932). In the work presented in this thesis, this measure is further taken into consideration through the randomisation of locations as detailed in section 3.4.

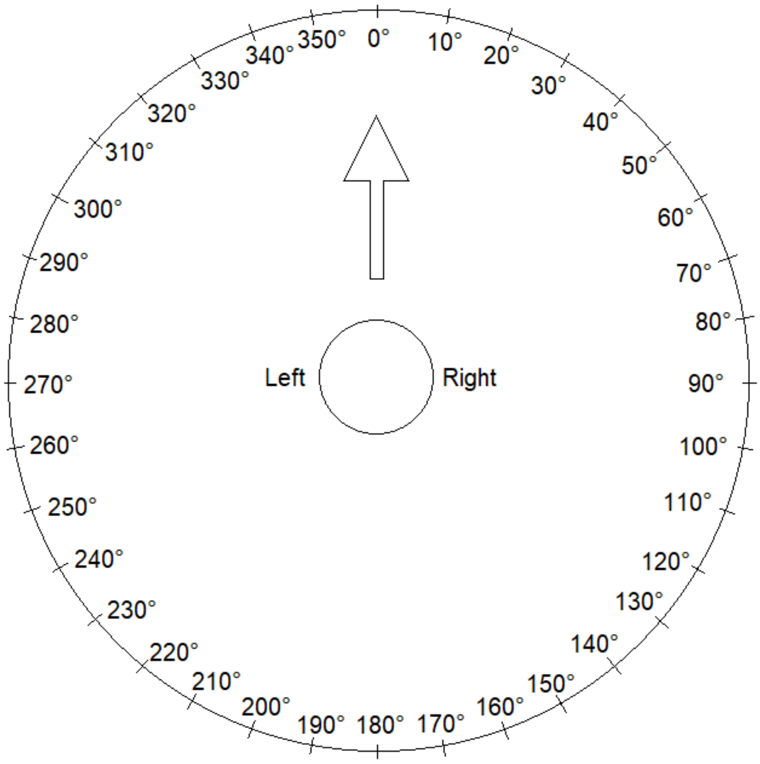


Figure 12 The lateral view of an example loudspeaker array with a 10 degree of separation between each driver

3.3 Stimulus Selection

The stimulus, or stimuli, should be capable of determining a system's binaural accuracy for a wide variety of applications. As such, the literature review presented in Chapter 2 suggests the necessary pre-requisites of an excitation are as follows:

(i) The stimulus should not be a pure tone (i.e. sinewave) as it contains too few HRTFs and interaural differences to be efficiently localised by a subject.

(ii) The stimulus should refrain from being a constant and prolonged level owing to a loss of interaural intensity differences.

(iii) The stimulus should be played at a constant decibel level of integrated loudness between 70-80 dB throughout the experiment and should not exceed a peak of 90 dB to prevent ear fatigue and hearing impairment.

It is important to note that the above recommendations may vary depending on the desired outcome of the experiment. For example, an investigation on the ability to localise 'harder-to-locate' sounds on a particular binaural system, will require the use of pure tones. These only provide a fundamental, standardised, procedure for testing the basic responses of binaural systems, particularly for the use in the entertainment sector (such as 3D film, VR and so on).

Very little speculation can be done on the timbral properties of a sound owing to a distinct lack of current research conducted on the correspondence between familiarisation of a sound and the ability to accurately locate a familiar sound compared to an unknown sound.

3.4 'Pseudo-Randomised' Locations

The randomisation procedure outlined in this section servers as a qualitative measurement of validation in ensuring that the subject is attempting the experiment and not 'guessing' (or psychologically predicting) the location.

To determine the accuracy of a binaural system, multiple locations need to be tested to investigate whether the system is capable of capturing the immersive nature of sound as perceived by humans. The randomisation (or sequence) of locations, from where the stimulus is emitted, is crucial to prevent the subject from discovering patterns and/or determining the location through any means other than their perceptive judgement of 'binaural' hearing. As such, particular patterns and repetitions should be avoided. More specifically, it is suggested to avoid using the same amount of samples as there are unique loudspeakers owing to (i) the ability to remove previous locations as the experiment progresses and guess the latter locations, through the process of elimination and (ii) being unable to establish a control location to check for consistency in the results.

A system establishing its qualitative response should measure localisation response for all possible locations (i.e. -135°, -150°, -175° azimuth, etc.) but should not attempt to localise every possibility precisely once within the same testing procedure. This will prevent the ability to distinguish patterns by the test subject and more importantly reduce the possibility of process-by-elimination. A system with 12 lateral locations (a location every 30° azimuth) and 12 samples, could by chance coincide with each possible answer (even with randomisation). As such, it is necessary to over-sample, or under-sample, to increase the difficulty of establishing patterns and arrangements. Additionally, it is important to note that oversampling, particularly in complex systems with large numbers of locations, results in a longer procedural time and thus affects ear fatigue and loss of attentive brain functionality (Roginska, 2018), (Hood, 1968), (Gelfand, 1981).

Ideally, subjects should be able to approximate a singular location, or general area, of a sound consistently and therefore score similarly (ideally matching, even if incorrect) on the two same locations (assuming localisation is not based on prior excitations).

It is suggested that the locations of the stimulus are generated randomly through an unbiased digital computation process (e.g. *RANDBETWEEN(1,30)* in the example of Microsoft Excel) for a sample total of approximately 80-85% of possible locations (26 for the example shown below). This under-sampling provides difficulty in attempting to establish patterns and recognitions. Furthermore, where possible, ensure that there are 1 or 2 locations which repeat to check for consistency, where these single locations do not repeat more than twice (to enable as many different locations as possible), or in a successive arrangement. A suitable randomisation of locations, for a testing procedure with 30 possible answers/locations, could be:

30, 4, 12, 22, 20, 8, 5, 29, 15, 13, 11, 1, 24, 19, 8, 18, 10, 20, 23, 6, 17, 10, 2, 21, 3, 9

The literature outlined in section 2.1 suggests that any accuracy within 15° azimuth or elevation in human hearing is perceived as the same location, thus suggesting the highest number of required locations is 24 (360° divided by 15° gap). However, this is merely a recommendation and pertains more specifically to systems intended purely for media and entertainment purposes. Systems demanding a higher accuracy of localisations are discussed in Chapter 8.2.1. Subjects should therefore be positioned within a loudspeaker array (i.e. Figure 12) and listen to a stimulus with an unbiased randomisation procedure to prevent pattern recognition.

3.5 Subjects and Subject Validation

Owing to the nature of any subjective experimentation, it is crucial to ensure that test subjects are capable of successfully conducting the experiment. In order to find out whether binaural audio is being captured, or played back, accurately through a new and/or existing system, a pre-test has been devised to determine whether the subjects are capable of conducting such a test within a natural hearing environment. This enables a subject's results and perceived judgement of a stimulus's location to qualify as 'trust-worthy' or more precisely, 'validated'. As such, a method of validating subjects is proposed.

A subject's hearing in the 'open-field' ("free-form"[3]) domain determines whether the subject in question is capable of conducting a test of such nature. Therefore, the following open-field test is proposed to validate subjects, whereby the localisation abilities of a subject are determined using the loudspeaker locations mentioned previously in this chapter to trigger the stimulant, as opposed to the use of a pre-recorded audio signal delivered through headphones (or other appropriate alternatives). The subject is positioned in the centre of an array or system

[3] Without the use of headphones or similar device

used to create the binaural recording and follow the identical procedure used for determining the accuracy of a binaural system. This enables the observer to assign a qualitative hearing response in terms of localisation abilities to any subject, and furthermore allows them to compare the results to those of the intended (binaural) test. Following the assumption that the binaural system is working perfectly, the subject should be able to score identically on all locations throughout both tests and ultimately score a matching overall grade or percentage. In practice however, it is expected that the headphones (binaurally reproduced[4]) test will have a reduced accuracy based on variance in the type of system, stimulus, and other factors detailed in sections 2.1 through 2.4. The rate of decreased efficiency is to be determined upon the average results and comparison of each test (validation versus binaural), ideally within a certain range of agreement to define consistency between localising the same set of locations (i.e. 20% difference in an overall binaural score of 60% or above would be acceptable for a validation score of 80%). A subject achieving an overall score of 65% on the validation test and 45% on the binaurally-captured test for any given system can still define the system as efficient and accurate as it is a reflection of how consistent the binaural system is in capturing and reproducing the sound signal. An overall result ranging from 20-30% lower than the open-field (non-headphone) test may suggest that the system, or experimental procedure, is considered faulty and further investigation is necessary.

To familiarise the subjects with the procedure of the experiment and the stimulus, a set of locations should be introduced prior to the experiment. This allows the subjects to be aware of the process and more importantly the stimulus, preventing subjects from focusing on the sound contents rather than the objective of the experiment. The results of the locations shared by these ‘guideline’ locations could serve towards showing whether a subject is capable of learning to localise. An increase of accuracy in these locations could overall be speculated to be a result of ‘learning-to-localise’.

This procedure will not test subjects for general hearing (frequency response) test as subjects will be based on their ability to localise sounds in terms of direction of arrival (DOA) and not whether they are able to physically hear it as the stimulus is consistent through the experiment. It is currently believed that a subject’s ability to hear a given frequency, or sound, better than other sounds is not a direct correlation to whether the subject is capable of localising the DOA of a sounds’ source accurately. It could be argued that a hearing impairment in one ear could affect localisation abilities as HRTFs are partly based on inter-aural loudness/intensity differences, however, additional validation procedures are suggested to create a theoretical

[4] Headphone test refers to the method of audio delivery through headphones, as opposed to the open-field (non-headphone) test of the validation procedure

‘pass’ mark which excludes any subject who is unable to achieve a minimum score defined by the average result and hypothetical probability of guessing the experiment (see 4.2.1). Hence, to avoid the possibility of a large population size of hearing-impaired subjects, the experiment will attempt to only use subjects below the age of 60 where hearing abilities decrease rapidly.

Studies and research conducted on subjective abilities or perceptions, with a given set amount of possible answers, along with the central limit theorem of probability, suggest a minimum population size of 30 to ensure an accurate average and representation. This eliminates the inflation of numbers and possibilities (Ruggieri, 2016), (Canals & Canals, 2019). Given the multiple verification, and validation, procedures outlined in this thesis, a smaller population size may be acceptable as the results could be considered more ‘trustworthy’. The author speculates that a population sample size in upwards of 25 may be acceptable given the methods presented in this thesis.

3.6 Hypothesis

The measures outlined above suggest a blueprint procedural guide to testing a binaural system. The outcome of this procedure defines a qualitative response figure that is attributed to an overall score of a binaural system for the broad use of entertainment and media. A methodology following the principles recommended in this chapter should also serve as a process of validation with regards to acceptable subjects and results. Furthermore, it takes into consideration the different factors which could negatively impact results and hence discredit the overall outcome. A manufacturer testing a new, or existing, binaural system, based on the requirements outlined in this chapter, should be able to determine a qualitative level of efficiency for its intended use. Additionally, manufacturers and developers of binaural systems should be able to compare it to other existing systems as per standardisation and unionisation of the test. This allows them to determine a more particular intended use of the system and position it within its respective field of technology, or alternatively, to investigate potential improvements in accordance with their desired aims and objective for the system. The pilot study undertaken in this work should, in accordance with the hypothesis, determine whether an appropriate set of results from a binaural system is possible and whether a subject’s perceived quality of an auditory image is capable of being quantified consistently in order to propose a standardised testing environment for binaural systems.

3.7 Summary

This chapter outlined some of the theory behind proposing, and creating, a standardised testing environment and hence suggested some testing procedures for binaural systems.

Furthermore, it provided a hypothesis on the outcome of proposing a standardised testing environment for binaural systems. The next chapter will present the design and modelling of the testing environment to be created to test the qualitative responses of binaural systems.

Chapter 4: Design

Chapter 3 determined a postulation on the outcomes of this research and on its experimental work conducted. Furthermore, it provided the specifics of the hypothesis to determine whether this research was successful in proving, or disproving, the viability of a standardised testing environment for binaural systems. This chapter outlines the blueprints for creating such an environment.

4.1 Assessment System of Results

In order to determine and assign a qualitative, numerical response for testing binaural systems, a method of representing its efficiency is necessary. The ability to localise a particular stimulus (based on frequency, intensity, etc.), or to distinguish a particular location based on prior knowledge (learning to localise) defines how localisation abilities function. However, it is ultimately the overall experience and perceived judgement of any stimulus from any given location which defines how efficient and successful a binaural system may be within its relative industry.

As with the standard procedure of assessing any test with a given factual answer set, there is only ever a singular correct answer for each 'question'. In mathematical terms, a given equation with any number of integers can only ever produce one correct result. Therefore, logically, if a subject estimates any amount of loudspeaker locations of a stimulus to the correct matching locations for all repetitions, they (or the system) could be said to have an accuracy of 100% (and therefore 100% precision, as there are no partial-correct answers). Owing to the complexity of human hearing, and therefore its localisation abilities, it may be appropriate to include a certain degree of error when evaluating response rates. In mathematical terms, an answer rounded up or down to certain decimal points may still be accepted (unless stated otherwise). However, this does not differentiate the subject's ability to answer accurately. As such, rather than simply extending the area of correct locations, a scaling factor should be considered to include and discern the 'near-correct' locations from the precisely correct ones. For example, a correctly estimated location would award 1 point, whilst an estimated location adjacent (either side) of the correct location is awarded 0.5 of a point respectively. Figure 13, Figure 14, and Figure 15 show the various possibilities of accuracy versus precision through a scoring system. The results are presented from a hypothetical experiment to illustrate the approach.

Correct Answer:		Subject Answer:		Points Awarded:		Correct Answer:		Subject Answer:		Points Awarded:	
1		10		0.0		1		1		1.0	
2		9		0.0		2		2		1.0	
3		8		0.0		3		3		1.0	
4		7		0.0		4		4		1.0	
5		1		0.0		5		5		1.0	
6		2		0.0		6		6		1.0	
7		3		0.0		7		7		1.0	
8		4		0.0		8		8		1.0	
9		5		0.0		9		9		1.0	
10		6		0.0		10		10		1.0	
		Total Points:		0/10				Total Points:		10/10	
		Accuracy:		0%				Accuracy:		100%	
		Precision:		0%				Precision:		100%	

Figure 13 Accuracy versus Precision of a point-based system, 0/0 versus 100/100

Figure 13 shows an example of two different sets of results from a subject's experiment, through the point system outlined above. The experiment has 10 different locations, with 10 samples, a correct estimation of the location, awards 1 point, with a total of 10 points available. Therefore, a subject scoring 0 points (left) has an accuracy of 0%, and a precision of 0%. As such, a subject who scored the maximum of 10 points (right), has an accuracy of 100% and a precision of 100%.

Correct Answer:		Subject Answer:		Points Awarded:		Correct Answer:		Subject Answer:		Points Awarded:	
1		1		1.0		1		2		0.5	
2		2		1.0		2		3		0.5	
3		3		1.0		3		4		0.5	
4		4		1.0		4		5		0.5	
5		5		1.0		5		6		0.5	
6		1		0.0		6		7		0.5	
7		2		0.0		7		8		0.5	
8		3		0.0		8		9		0.5	
9		4		0.0		9		8		0.5	
10		5		0.0		10		9		0.5	
		Total Points:		5/10				Total Points:		5/10	
		Accuracy:		50%				Accuracy:		100%	
		Precision:		100%				Precision:		50%	

Figure 14 Accuracy versus Precision of a point-based system, 50/100 versus 100/50

Figure 14 shows the effect of the experiment once a scaling point factor is introduced, with adjacent locations (to the correct) awarding 0.5 points. Both subjects scored 5 points, with different accuracies and precisions. The first subject (left) achieved an accuracy of 50%, with 100% precision. The latter subject (right) had an accuracy of 100% owing to the subject estimating the approximate location correct, but with a reduced precision, of 50%.

Correct Answer:	Subject Answer:	Points Awarded:	Correct Answer:	Subject Answer:	Points Awarded:
1	1	1.0	1	1	1.0
2	2	1.0	2	2	1.0
3	3	1.0	3	3	1.0
4	4	1.0	4	4	1.0
5	5	1.0	5	5	1.0
6	5	0.0	6	5	0.5
7	6	0.0	7	6	0.5
8	7	0.0	8	7	0.5
9	8	0.0	9	8	0.5
10	9	0.0	10	9	0.5
	Total Points:	5/10		Total Points:	7.5/10
	Accuracy:	100%		Accuracy:	100%
	Precision:	50%		Precision:	75%

Figure 15 Accuracy versus Precision of a point-based system, 100/50 versus 100/75

Figure 15 shows the effect of the introduction of a scaling-point factor. The same set of results allow for a different overall score when given a wider acceptable degree of error (right).

Given that the overall score of a system is a combination of precision and accuracy, and that overall results can be identical for various levels of precision and accuracy, it is ultimately up to the binaural systems' manufacturer to determine whether a systems accuracy or precision is more vital, based on the intended application of the system in question.

4.1.1 Point-based System

Given a dataset of answers following the theory highlighted above, a qualitative figure can be awarded to any given subject's result set, and therefore also to the binaural system undergoing the testing procedure. However, this procedure works on the principle of a narrow depth of field in terms of possible accuracy (to 15° azimuth) and a singular possible answer. Instead, a multiple-possible answer set is proposed to add a broader range of acceptance, with a reduced scaling factor for the locations +/-1, +/-2, etc. adjacent to the correct locality. This allows for some degree of error, whilst still awarding the correct locations with a higher score. In the case of a system with 24 possible locations divided equally along the lateral field, and a test with 20 samples (triggers), the maximum available points would be 20. A test which includes an acceptable error rate of +/-1 (relating to +/- 15° azimuth in this particular example), and awards 2 points and 1 point for the correct and 'semi-correct' answers respectively, would have a maximum possible score of 40 points. In both examples, assuming a perfect set of results, this relates to 100% accuracy. However, if we assume a different set of results whereby 10 out of 20 results are correct and the remaining 10 results are either of the adjacent locations, the first scoring system would therefore award an overall score of 50% and the latter scoring system 75%. The acceptable degree of accuracy (ADA) is directly proportional to the intended use of the system and can therefore provide a blueprint for proposing recommendations on a system's ADA.

The work undertaken in this research should be able to propose guidelines and suggestions towards creating a standardised testing procedure. Given that the most common use of binaural audio currently is in the entertainment and media sector, the scoring system will be based as such (for guidelines and suggestions on systems created for other uses, refer to section 8.2 further work, and section 8.1 main contributions):

A correct location should award 3 points, with each adjacent location (+/-15°) awarding 2 points and each adjacent location to a degree of +/-2 locations (+/-30°) awarding 1 point respectively. Therefore, the maximum possible for 20 samples of correct locations is 60 out of 60 (100%). A subject estimating (or perceiving) the location with a 15° azimuth error (locations directly adjacent to the correct one) for all 20 samples would be awarded an overall result of 66.67% (40 out of 60) and a subject estimating the locations adjacent to the correct one by 30° azimuth would be awarded 33.33% (20 out of 60). In summary, a result estimated within an area of 60° azimuth of the correct location (30° either side) will award a varying amount of points between 1-3, with 3 points being the highest/exact location and 0 points for any estimation within the remaining 300° azimuths.

	A	B	C	D	E	F	G	H	I	J	K
4							#	Pts.		#	Pts.
5							6	3		6	3
6							18	3		21	0
7							17	3		19	1
8							8	3		7	2
9							9	3		8	2
10							5	3		6	2
11							16	3		16	3
12							24	3		24	3
13							7	3		8	2
14							11	3		10	2
15							20	3		17	0
16							10	3		7	0
17							12	3		12	3
18							19	3		17	1
19	**Total points gained;**			40			2	3		2	3
20							13	3		11	1
21	**Percentile accuracy**			66.66667			4	3		4	3
22							1	3		1	3
23							23	3		23	3
24							5	3		5	3
25						Total;		60			40

Figure 16 An example of the point system for the different degrees of error

Figure 16 shows an example of different points awarded for various acceptable degrees of error with the correct locations in column G, versus a subject answer set in column J.

Following the principles of the scoring system outlined above, the likelihood of scoring highly is improbable should the subject approximate the locations to any degree further than +/-15°, showing that 20 semi-correct locations which still award points can still only achieve an overall total of 33.33% assuming all 20 samples are estimated at 30° adjacent to the correct location. However, this does not take into account the various 'partially-correct' locations. Instead, the theoretical probability of any given overall score is investigated.

4.1.2 Calculating the Probability of Guesswork

With any ability or skill-based examination, there is a certain likelihood that the test in question could be 'guessed' and a subject be awarded a significantly high result, thus demanding a form of validation or exclusion procedure for such anomalies. The first proposed validation is seen above, in section 4.1.1. The second method is most commonly done through the use of a large population size (enough subjects), or more accurately, through quantitative reassurance as mentioned in section 3.1. This defines a minimum required sample size of subjects in order to reduce the probability of a skewed result, i.e. large irregularities or deviations in numbers. However, owing to the tedious and time-consuming nature of testing binaural hearing responses, and therefore binaural systems, an alternative method is proposed.

To validate or determine whether a subjects' answer is of their own, and not a product of some external factor or guesswork, the likelihood of scoring certain results through pure random guessing is investigated. In a simple multiple-choice test with 4 possible answers (A, B, C and D), the probability of answering correctly, independent of previous randomisations, would be ¼, or more specifically, 25%. In an experiment with a total of 24 locations (loudspeakers) the theoretical probability would therefore be 1/24 (4.16%). However, as seen above, this only takes into account a singular correct answer/location, excluding the additional four locations, two each side adjacent to the correct location. Therefore, strictly speaking, this equates to a probability of 5/24 (20.83%) that a subject will be able to guess any possible correct location. This does not take into account that the correct location may only be semi-correct and not award full points based on the system described above. Taking into account the scaling factor of the remaining locations, two locations at 2 points and two more locations at 1 point respectively, the five correct answers would award 1.8 points each (the mean of 1, 2, 3, 2, 1 points). Following this theory, it can then be expressed that through pure randomisation, 20.83% of attempts will award 1.8 points each, and the remaining 79.17% will award 0 points. Repeating this procedure for a total of 20 attempts, or samples, it would hypothetically equate to on average, correctly answering 4.167 times. In other terms, using pure and independent randomisation, on average any subject will only ever achieve an overall result of 7.5/60, or more accurately represented as 12.5% (4.167 by 1.8 points each). This is based on the hypothetical likelihood that each location is guessed randomly but does not however, take into consideration the likelihood that a result could deviate, although through very low probability, score an overall results in the range of 40% or higher.

To investigate the possibility of a subject's overall result in either of the two procedures, open-field or headphone, a simulation of the experiment is required. In theory, the simulation (permitting occasional 'strays') should match an average of 12-13% overall score as detailed above. The outcome of the simulation, in combination with the theoretical calculation, should suggest and propose a minimum 'pass' mark (or benchmark), which eliminates a subject should their results fall below the threshold of 13%. However, in order to allow for anomalies, such as the occasional randomly-generated high result or a combination of correct approximation and guesswork, some headroom is suggested in order to avoid the higher averages. A simulation of the experiment and the likelihood of guessing it is conducted in Chapter 5.

4.2 Creating a Loudspeaker Array

The loudspeaker array was created by positioning a series of 24-identical Visaton 3-inch (FR10 HM) drivers equally spaced on the azimuthal plane at 0° elevation. The loudspeaker array was rotated by 7.5° to position the audio drivers off-axis. This was done in order to avoid 0°, 90°, 180° and 270°, owing to the maximum and minimum interaural differences ('easiest' and most difficult locations to estimate) (See Figure 17). Furthermore, this also avoided the correlation and assimilation of the recording to that of a stereo signal (stereo can be considered as 90° and 270° for the left and right channel respectively). Each audio driver was mounted on a custom-made, 7mm medium-density fibreboard stand which was laser-cut to create 24 identical copies (See Appendix A). The mounted audio drivers were then locked in place on a circular rig also created out of fibreboard. The circle, which held the drivers in place, had a radius of 1m and raised the loudspeakers by approximately 5cm to avoid strong absorption or reflection from the floor and to allow for natural dispersion of the sound.

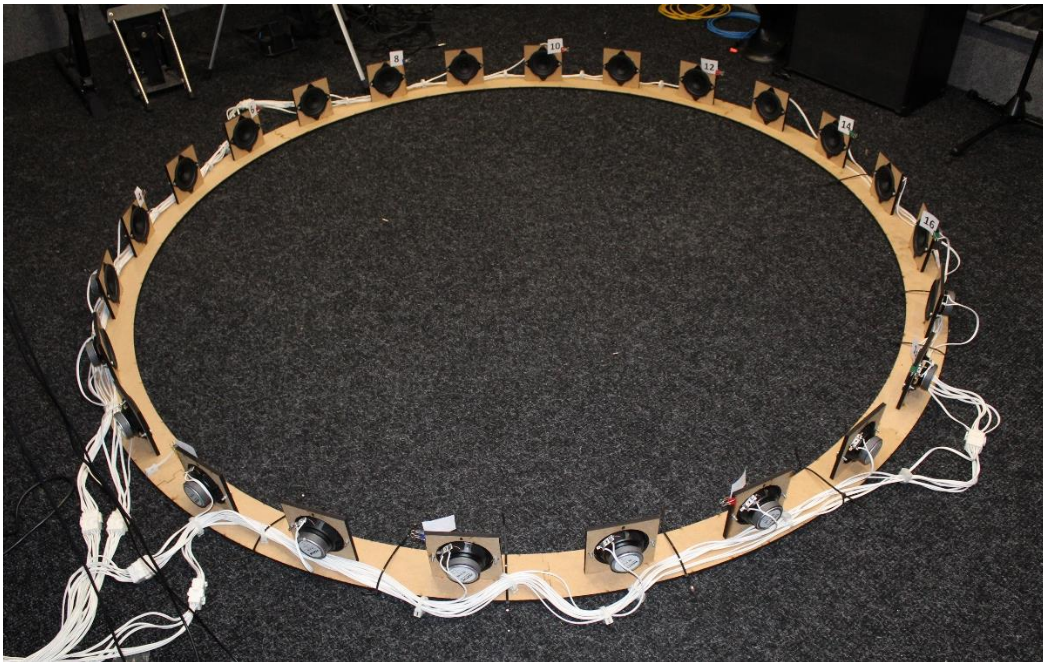


Figure 17 The circular array of loudspeakers located in the isolation chamber

The audio drivers were visibly numbered in ascending order from 1-24 for the human test subjects to easily communicate the perceived location of the DOA of the test sound. These were then connected using 2-core, 13 strand, 12-gauge loudspeaker audio cable (ProPower 14512) which ran through to the outside of the isolation booth where they were connected to a switch box and a *NAD* power amplifier (NAD Stereo Integrated Amplifier 310). The switch box, shown in Figure 18,was designed using a project box and 24 independent switches. It allowed for the control of any given single, or combination, of loudspeaker connectivity through the use of toggle switches.

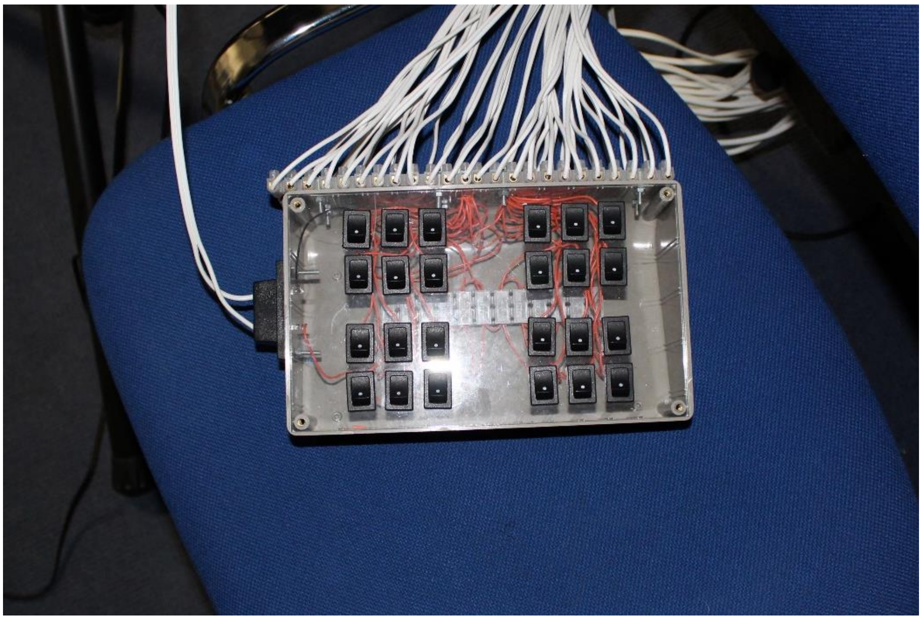


Figure 18 A switch box outside the isolation chamber, designed to individual control each of the 24 loudspeakers

4.3 Location and dB Levels

The experiments were conducted in an isolation booth (DV1560) produced by *DEMVOX*™ (DemVox, 2019) with internal measurements of 2.560m height, 3.572m width and 4.712m length. The isolation booth has an average resting (background) level of 34 dB A weighting and an average reverberation time (RT60) of 0.1s – 0.25s at a frequency of 1 kHz. The resting room level was measured using a Bruel and Kjaer decibel metering device (2239A). This decibel meter was also used to ensure that the levels of the excitation (the stimulus/signal output of the loudspeakers) was kept between 70-80 dBA and did not exceed 90 dBA. This was done to prevent ear fatigue, hearing loss and to furthermore comply with ethical approval provided by Liverpool John Moores University. See Appendix B for attached ethical approval.

4.4 Binaural Recording, Playback and Testing Procedure of Experimentation

The binaural audio was recorded using a Binaural Enthusiast B1-E dummy-head (Dobosz, 2019)). The dummy-head was positioned within the equilateral centre of the loudspeaker array (Figure 20) facing toward 0° and a stimulus was triggered through each loudspeaker with the correct corresponding location according to the sequence determined in section 3.4. The excitations were recorded onto a portable recorder (TASCAM DR60), at a sampling rate of 96 kHz, using a 32-bit .WAV digital file format. The dummy-head was removed, and the subjects were asked to position themselves in place of the dummy-head, once again, facing 0° azimuth. The recordings were played back using the DR60 and using on-ear closed-back headphones (Audio Technica ATH-60x). The play-back followed the dB levels specified in section 4.3 and were kept constant for each subsequent participant. The full methodology and experimental procedure can be found in Chapter 5, particularly section 5.3.

4.5 Summary/Chapter End

This chapter presented the initial prototype environment for the work undertaken in this research. Chapter 5 will outline the methodology undertaken to determine and generate results which should in turn clarify the confidence of the hypothesis.

Chapter 5: Method – Standardising Testing Environments

Chapter 4 discussed the design theory behind proposing a blueprint for creating a standardised testing procedure. Furthermore, it highlighted additional measures required to ensure a validation procedure not seen in other work within the field of testing binaural systems to date. This chapter defines the work undertaken to determine the viability of proposing guidelines towards creating a standardised testing methodology for binaural systems.

To simplify and define the material work undertaken in this research for the requirements of proposing a standardised testing procedure for binaural systems, the following is summarised:

(i) The procedure of creating the loudspeaker array used in the experiment, namely the materials, the selection, and positioning, of audio drivers, the wiring, and lastly the system to control said audio drivers.

(ii) The physical environment where the experiment was conducted, and the audio (relative decibel) levels used for the stimulus.

(iii) The chosen stimuli used in the experiments, Stimulus A and Stimulus B, and their spectral physical properties (i.e. frequency and duration).

(iv) A running order of locations from where a stimulus was triggered from (which loudspeaker position played first, either through the dummy-head recording or directly to a subject in the validation test).

(v) The procedural instructions given to subjects as well as the responsibilities of the observer during and after the experiment. A definition of the process for testing a binaural system along with the validation of a subject's localisation abilities.

(vi) The subjects, and the number of subjects, participating in each relative experimental stage.

(vii) The process for creating a second measure of validation, through a simulation of the experiments.

5.1 Order of Locations

To test the binaural system, a random sequence of loudspeaker locations was required. This was created through a randomisation procedure using Microsoft Excel and the *RANDBETWEEN,(1,24)* function for a total of 20 samples and ensuring that any given number only ever appeared a maximum of twice and in a non-successive order as explained in section 3.4. The running order of locations was manually checked to ensure at least one repetition of a location to check for consistency in results and a maximum of two locations repeating, thus ensuring that a maximum amount of different locations was present during the experiment. An acceptable order of locations versus an unacceptable order is shown in Figure 19. The green column (left) is considered an acceptable order of locations owing to a lack of successive repeats, 80% of total samples available, and a singular repetition of a number matching the modal average (a frequency of 2). The red column (right) is considered unacceptable owing to the increased frequency of the mode (a frequency of 3), a repeating location (location 11), as well as a second repeating number, thus reducing the sample size to 71%.

Order of locations examples				
Acceptable:			Unacceptable:	
	19			5
	6			9
	23			18
	17			4
	10			7
	13			11
	8			11
	12			3
	15			4
	2			9
	16			1
	20			24
	24			19
	22			2
	4			21
	7			17
	21			14
	11			6
	3			5
	4			11
Mode:	4			11
Freq:	2			3
Successive Repeat:	0			1 (11)

Figure 19 Acceptable versus Unacceptable examples for orders of locations

5.2 Preparation Procedure for Experimentation

The subjects undertaking the test were given verbal instructions (Appendix C) on the experimental procedure and informed about the protection and storage of their personal data acquired through their participation in the experiment. The subjects were further instructed to listen to a total of four familiarisation (guideline) locations prior to the test to acquaint themselves with the stimulus, as well as the concept of the experiment. These locations were 1, 7, 13 and 19 and were made known to the subjects. As orientation tests, these were not recorded.

5.3 Experimentation Procedure

Subjects were asked to position themselves in the circular array, facing 0° azimuth (see Figure 20). Subjects were then further instructed to communicate the perceived DOA of a stimulus based on the relative loudspeaker number attached to the mount. The perceived number was communicated to the observer located outside of the isolation booth using a simple talk-back system to ensure maximum possible sound isolation. This was repeated for all 20 samples following the order of locations determined in Chapter 3. Upon completion of the experiment, the subjects were assigned a unique identification number, to prevent any breach of data protection and furthermore to allow comparison of results to a subject’s validation result. Subjects were asked not to discuss potential results with any other participant. Each experiment lasted approximately 20 minutes, which included time for instructions, the binaural and validation tests and the rotation to the next subject.

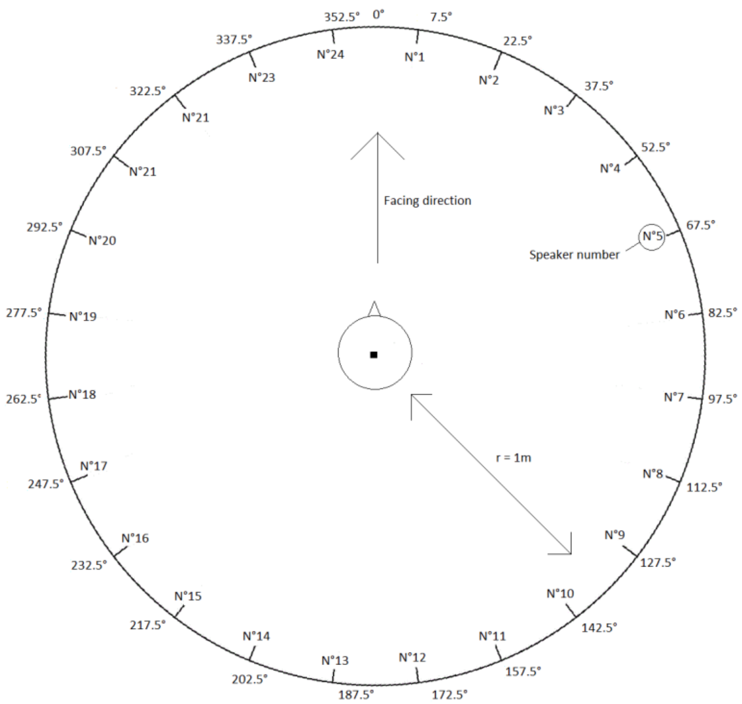


Figure 20 Position of subject (or dummy-head) and loudspeakers in the loudspeaker array (aerial perspective)

5.4 Evaluation of a Binaural System’s Efficiency

Experiments were conducted on two different stimuli Stimulus 1 and 2 with a different order of locations for each. As such, this thesis shall refer to the experimentation of stimulus 1 as Experiment 1, and Experiment 2 for stimulus 2. Each experiment has an additional validation test to compare and evaluate results. Therefore, a total of four tests were conducted: (i) Experiment 1 – Headphone test (binaurally delivered audio), (ii) Experiment 1 – Open-field test (natural free-hearing delivered audio), (iii) Experiment 2 – Headphone test and (iv) Experiment 2 – Open-field. Subjects in all experiments were given the same instructions as detailed in section 5.3 and all experiments followed the same testing procedure, except for the order of triggered locations and stimuli.

5.4.1 Experiment 1 using Stimulus 1

The first experiment was conducted using a single-click tone with a burst of 0.006s and a total duration of 0.04s (Figure 21). The stimulus was artificial and created in Audacity[5] (Audacity 2.3.1). This was done to follow similar stimuli used in previous experiments in the field (see section 2.5.3).

Figure 21 Stimulus 1 spectrogram (left) and waveform (right)

The stimulus was played through the following loudspeaker locations, determined using the randomisation methodology seen in section 5.1:

6, 18, 17, 8, 9, 5, 16, 24, 7, 11, 20, 10, 12, 19, 2, 13, 4, 1, 23, 5

Experiment 1 was conducted on a total of 34 subjects.

[5] https://www.audacityteam.org/

5.4.2 Experiment 2 using Stimulus 2

The second experiment used a 0.3s long sample of pink noise also generated in Audacity[1]. As with Stimulus 1, a pink-noise stimulus was used in the experimentation of binaural localisation abilities.

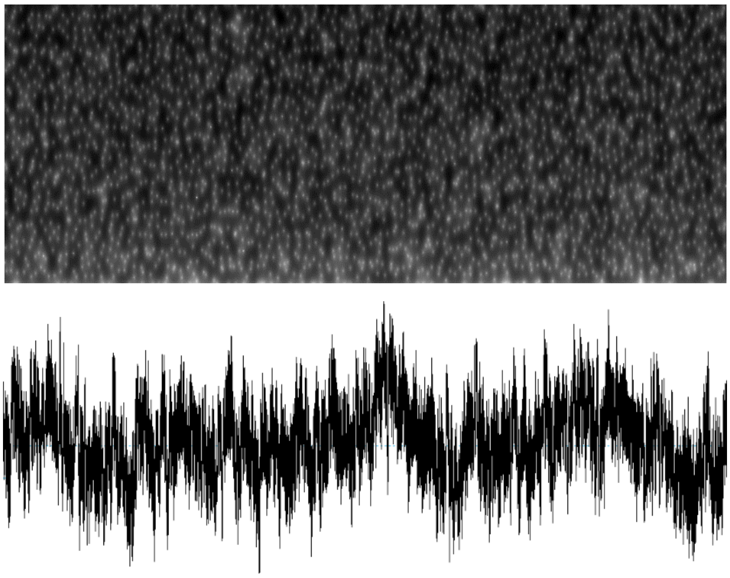


Figure 22 Stimulus 2 Spectrogram (top) and waveform (bottom)

The stimulus was played using the following loudspeaker locations:

19, 1, 4, 17, 24, 10, 9, 22, 11, 12, 2, 4, 5, 23, 20, 14, 21, 1, 15, 3

A total of 11 subjects undertook Experiment 2.

5.4.3 Validation Process

As an additional method of validating any given subject's results and answer set, all subjects were asked to repeat the experimentation procedure, both for Experiment 1 and Experiment 2. This time, the subjects were asked to listen to the excitations triggered through the loudspeakers, without the use of a binaural recording or headphones. This validation test (open-field test), conducted in an open, free-hearing, situation, would determine whether a given subject was capable of localising the stimulus in its natural form, thus verifying the validity of their binaural test. The subjects were then asked for their unique identification number, provided previously, to allow for comparison in the result sets (binaural test versus validation test).

Each subject was given an overall score of efficiency on each stimulus, including the binaural and validation tests.

5.6 Creating a Scoring and Point System

The scoring system was also designed in Microsoft Excel (Excel 2013) using the respective correct answer set for that particular stimulus. A subject estimating the apparent or perceived location to the exact correct loudspeaker number (e.g. 6) was awarded 3 points; 2 points were awarded for the locations adjacent to the correct loudspeaker (5 and 7 in the given example) and 1 point for the locations two places adjacent from the correct loudspeaker (4 and 8 based on the example once again). All estimations of a location outside of these 5 ‘correct’ loudspeakers/numbers (1-3, 9-24 in the example stated previously) were awarded 0 points. Therefore, the maximum possible points a subject could be awarded is 60 (3 points for each of the 20 samples).

A subject’s perceived answers were entered into the Microsoft Excel spreadsheet and the scores were determined using IF statements (see Figure 23). The totals were added together and transposed into a percentage score, thus enabling the identification of a subject’s, or system’s, overall efficiency as a percentage, qualitative, measurement whereby 60/60 points are equivalent to a score of 100%.

	A	B	C	D	E	F	G	H	I	J	K
4							#	Pts.		#	Pts.
5	**Test type (Pre-/3Dio/etc)**			Pre-test			6	3		6	3
6							18	3		20	1
7	**Stimulus (tone/bird noise/etc)**			Tone			17	3		18	2
8							8	3		7	2
9	**Date:**			DD/MM/YY			9	3		9	3
10							5	3		5	3
11	**RNG Set:**			A/B/C/etc			16	3		19	0
12							24	3		24	3
13	**Max. pts available (20 x 3)**			60			7	3		6	2
14							11	3		10	2
15							20	3		20	3
16							10	3		10	3
17							12	3		15	0
18							19	3		20	2
19	**Total points gained;**			46			2	3		2	3
20							13	3		14	2
21	**Percentile accuracy**			76.66667			4	3		4	3
22							1	3		1	3
23							23	3		23	3
24							5	3		5	3
25						Total;		60			46

Figure 23 An example of an answer set for the experimentation procedure

5.7 Simulation of the Experimental Procedure

As discussed in Chapter 3 and Chapter 4, the probability of a subject scoring highly, or a certain overall score, is theoretically calculated. The results from this simulation should coincide with the hypothesis in section 4.1.2 which calculated the overall average result of a subject conducting the experiment through pure guesswork to 12.49%. The combination of results from the simulation and the calculation should suggest, or even determine, the viability of setting a 'pass mark', and where the threshold of acceptable results should lie (i.e. above 45% overall score, allowing for an additional 5% of headroom).

The simulation, designed in Microsoft Excel, was created to follow the principle provided in Chapter 3. A set of "correct" loudspeaker locations were generated following the methodology (Chapter 4), and then a set of "answers" were randomised using the same RANDBETWEEN(1,24) function. An overall score was determined using the point system detailed in section 4.1. This simulation was repeated for a population (theoretical subjects) size of 2000, with all theoretical subjects aiming to answer the same correct order of locations. A second simulation was created, following the order of locations used in Experiment 2 (see section 5.4.2) in order to determine the success rate of guesswork in that particular experiment.

To eliminate any possibility of biased numbering generation within the Excel coding and the randomisation procedure, three additional simulations were run using the same procedure with a different correct location set. The third simulation used an ascending order of numbers from 3 to 22. The fourth simulation used the same speaker/location (12) as the correct answer for each possible sample. The fifth simulation generated a random answer set for each of 2000 different theoretical subjects (no set/same answer set). All five simulations were created with a sample size (number of answers) of 20 in an attempt to reproduce the experiment as accurately as possible.

5.7.1 Simulation 1 using Dataset 1

The first simulation test, based on a random set of 20 samples/numbers following the randomisation procedure suggested in the chapter above, used the following "correct" locations:

17, 16, 14, 12, 20, 4, 6, 11, 3, 15, 24, 13, 7, 6, 23, 8, 23, 10, 9, 15

5.7.2 Simulation 2 using Dataset 2

The second simulation was calculated using the following randomly generated set of numbers, repeating the procedure seen in the first simulation:

19, 1, 14, 17, 24, 10, 9, 22, 11, 12, 2, 4, 5, 23, 20, 14, 21, 1, 15, 3

This set of numbers was used to check for consistency in theoretical/simulated results versus the human results seen in the second experiment. This should allow for the comparison of a theoretical versus practical response and discern whether any given subject conducting the test was likely to have achieved the score obtained simply by guessing the locations.

5.7.3 Simulation 3 using Dataset 3

Simulation 3 was undertaken using the following correct answer set:

3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22

This set of numbers, as aforementioned, was created to check for consistency and/or irregularities in the randomisation algorithm within Microsoft Excel. This would determine whether the programme created patterns or favouritism towards certain arrangements of numbers.

5.7.4 Simulation 4 using Dataset 4

Simulation 4 was conducted with the following correct answer set:

12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12

This simulation was created using the same location (12) to determine whether the randomisation algorithm created a biasing of numbers towards a particular, singular, location/answer. Dataset 4, along with Dataset 3, should be sufficient in defining whether there are any programming biases towards any given numbers. The results from these datasets can only indicate the presence of a biased number, however, are not capable of identifying which location/number. In the event of a biasing, further investigation is required.

5.7.5 Simulation 5 using Dataset 5

Lastly, simulation 5, created with a correct answer set of:

X, X, X, X, X, X, X, X, X, X, X, X, X, X, X, X, X, X, X, X

(whereby X = any number between 1-24)

This double randomisation procedure, where every correct answer is randomised along with each possible answer-attempt (using multiple possible answer sets), was created to simulate 2000 different experiments (or stimuli tests). This set of numbers defined whether guesswork above 40-50% was probable in any possible variant of correct answer set in an experiment.

5.8 Summary

This chapter defined the procedural process of designing and conducting an experiment to determine the viability of creating a standardised testing environment for binaural systems. It particularly outlined the various materials and technology used to create such a test. Furthermore, the validation procedures for defining a human subject's localisation abilities, and the probability of guesswork through a simulation, was conducted and presented. Chapter 6 presents the preliminary findings and results from the experiments shown in this chapter.

Chapter 6: Results and Analysis

Chapter 5 outlined and defined, the precise process and methodology undertaken in this research to determine the viability of proposing a standardised testing environment for binaural systems. This chapter presents the findings and results of the experimental procedure highlighted in the methodology. In addition, this chapter also analyses some of the results and outcomes to be discussed in Chapter 7. This chapter presents the results of the two experiments (1 and 2) as well as the results of the simulated experiment in order to determine the likelihood of guesswork, in conjunction with the objectives set (section 1.2). For the readers convenience, the following results are described and presented in this chapter: (i) Results of Experiment 1 (binaural “headphone” test) and Experiment 1 (validation “natural-hearing” test), (ii) Results of Experiment 2 (binaural test) and Experiment 2 (validation test), (iii) Comparison of consistency locations within each experiment and the comparison of Experiment 1 against Experiment 2, and (iv) Results of the Simulated Experiments (Dataset 1, 2, 3, 4 and 5).

To investigate and determine the validity of results, various statistical observations were made. The overall scores gathered from all subjects for the binaural, validation and simulation tests were compared and analysed, namely:

(i) The mean (MEAN), the average of all results within one dataset,

(ii) The median (MEDIAN), the midpoint of the frequency distribution (the mid-point of all numbers arranged in ascending order),

(iii) The mode (MODE), the most common repeating value from all participating subjects,

(iv) The max (MAX), the maximum and highest occurring value, and

(v) The minimum (MIN), the lowest occurring value in the entire population of results for the relevant experiment.

Each overall result was determined based on the scoring system defined in section 5.6. The computation of all overall results produced the outcomes shown in the sections below. An example of a singular subject’s set of results is shown in Figure 24, whereby the subject’s perceived answers are compared to the correct answers and the overall score is calculated below. Again, the validation results refer to those conducted through the natural (free form) hearing, whereas the binaural results refer to the results obtained through the use of headphones. The consistency locations are the locations which repeat within a given experiment (see section 5.1).

	Correct location;		Subject estimation;	
	#	Pts.	#	Pts.
	6	3	6	3
	18	3	19	2
	17	3	20	0
	8	3	6	1
	9	3	8	2
	5	3	5	3
	16	3	19	0
	24	3	24	3
	7	3	6	2
	11	3	11	3
	20	3	21	2
	10	3	9	2
	12	3	12	3
	19	3	19	3
	2	3	1	2
	13	3	14	2
	4	3	5	2
	1	3	1	3
	23	3	23	3
	5	3	5	3
Total;		60		44
	Total points gained;		44	
	Percentile accuracy		73.3333	

	Correct location;		Subject estimation;	
	#	Pts.	#	Pts.
	6	3	7	2
	18	3	18	3
	17	3	15	1
	8	3	7	2
	9	3	7	1
	5	3	8	0
	16	3	14	1
	24	3	16	0
	7	3	8	2
	11	3	12	2
	20	3	17	0
	10	3	8	1
	12	3	10	1
	19	3	18	2
	2	3	3	2
	13	3	13	3
	4	3	6	1
	1	3	24	2
	23	3	21	1
	5	3	6	2
Total;		60		29
	Total points gained;		29	
	Percentile accuracy		48.33333	

Figure 24 Subject 'X' results from open-field test (left) and headphone test (right)

6.1 Experiment results

Each experiment was carried out on various subjects, some of which participated in both Experiment 1, and Experiment 2 (sections 5.4.1 and 5.4.2 respectively). Each subject followed the same procedure highlighted in the methodology chapter and was awarded an overall qualitative score for both tests as defined by the scoring system designed in section 5.6. All subjects participating in the experiment were between the ages of 18 and 60 covering a variety of genders and ethnic backgrounds.

6.1.1 Experiment 1

Experiment 1 was conducted with Stimulus 1 (see Figure 21 in section 5.4.1) and undertaken on a total of 34 subjects participating in both the binaural and the validation tests, whereby each subject contributed to 2.94% of the overall statistics.

Table 1 Statistical results of Experiment 1 (binaural)

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
48.55%	47.50%	40.00%	68.33%	6.70%

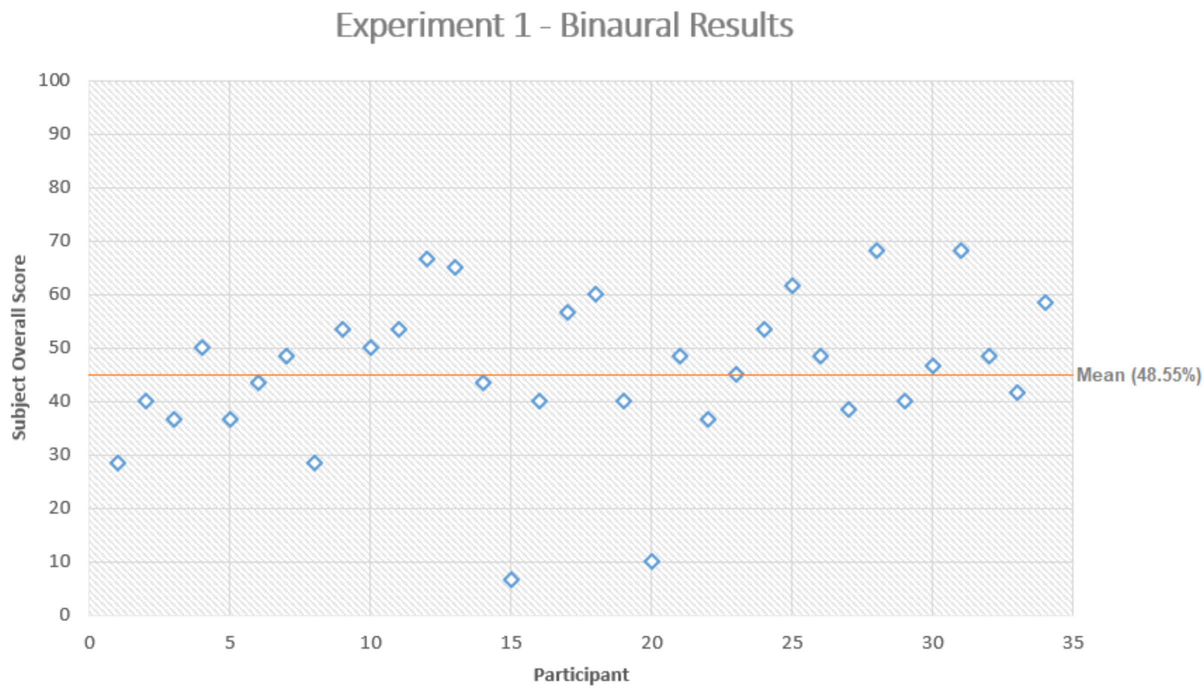


Figure 25 Graph of overall results from all Experiment 1 subjects (binaural)

Figure 25 presents the initial results of the binaural test including the un-validated subjects. To determine which subjects are attempting the experiment, beyond the limits of guesswork, the results of the validation test are presented (Figure 26). The statistics observed are provided in Table 2:

Table 2 Statistical results of Experiment 1 (open-field)

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
59.06%	60.00%	61.67%	83.33%	33.33%

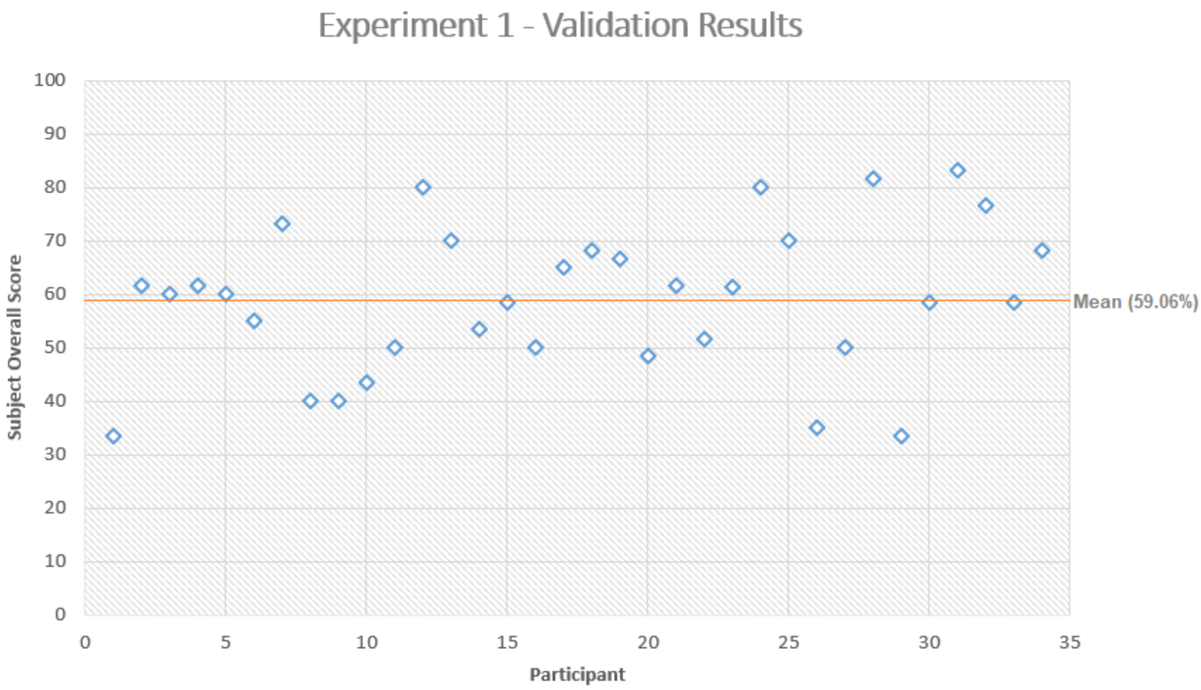


Figure 26 Graph of overall results from all Experiment 1 subjects (open-field)

The percentage error (comparison) of the validation results versus the results of the binaural for each statistical category were observed as detailed in Table 3:

*Table 3 The percentage error of each statistical overall result ((Open-field – Headphone)/Openfield * 100))*

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
-22.31%	-20.83%	-35.14%	-18.00%	-79.99%

The differences and percentages of each subject's overall result error from all 34 subjects between the headphone and the open-field tests were compared (Figure 27). These differences produced results as shown in Table 4:

Table 4 The average decrease in overall results from open-field to headphone in Experiment 1

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
13.18%	11.67%	11.67%	51.66%	5.00%

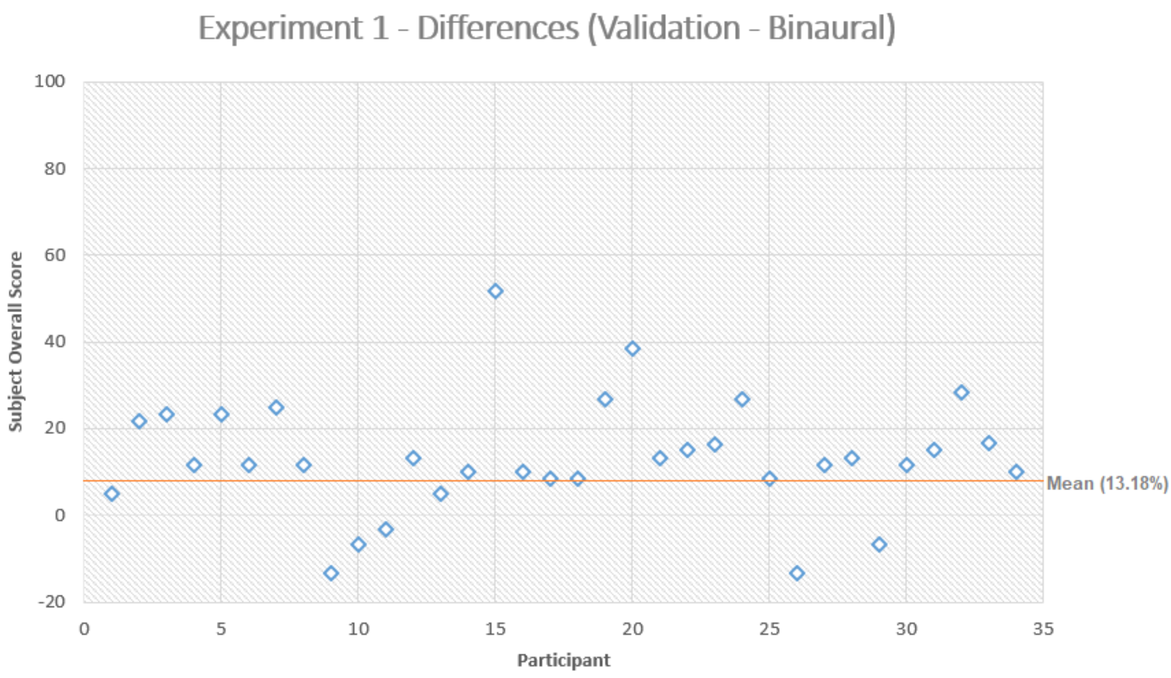


Figure 27 Graph of the differences between both tests in Experiment 1 (open-field – headphone)

To exclude the subjects' results which were considered unsuccessful for the experiment, the headphone scores below an overall of 45% were omitted, as determined in the theory chapter (see Chapter 3). Out of the total 34 subjects, 28 were able to achieve an open-field score above 45%, the remaining 6 were unable to achieve results above 45% and were therefore excluded, resulting in a pass rate of 82.35% (28 out of 34). Therefore, each subject represented a total of 2.95% towards the pass rate. The new results and statistics were calculated as shown below.

Table 5 Overall results from validated-only subject in Experiment 1 (headphone)

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
46.84%	47.50%	40.00%	68.33%	6.67%

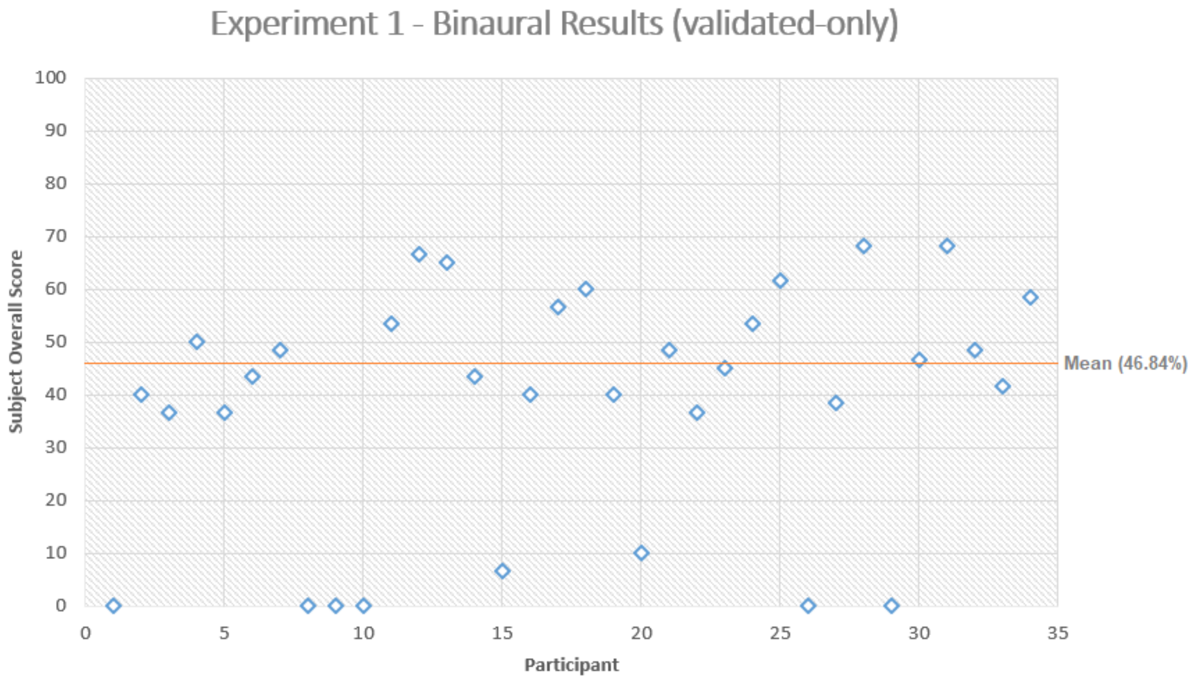


Figure 28 Graph of overall results from validated-only subjects in Experiment 1 (headphone), subjects with an overall score below 45% in the open-field test are presented as 0, and do not contribute towards the mean

6.1.2 Experiment 2

Experiment 2 was carried out on a total of 11 subjects, some of whom had also carried out Experiment 1 prior to this, with each of the 11 subjects contributing 9.09% towards the overall statistics result. This experiment was conducted using Stimulus 2 (see Figure 22 in section 5.4.2), once again following the same principle created earlier within this research and used during Experiment 1. This experiment was carried out using the order of locations shown in section 5.4.2. Overall results for the headphone test can be seen in Figure 29. The MEAN, MEDIAN, MODE, MAX and MIN of the overall results are detailed in Table 6.

Table 6 Statistical results from Experiment 2 (binaural)

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
55.15%	53.33%	46.67%	83.33%	33.33%

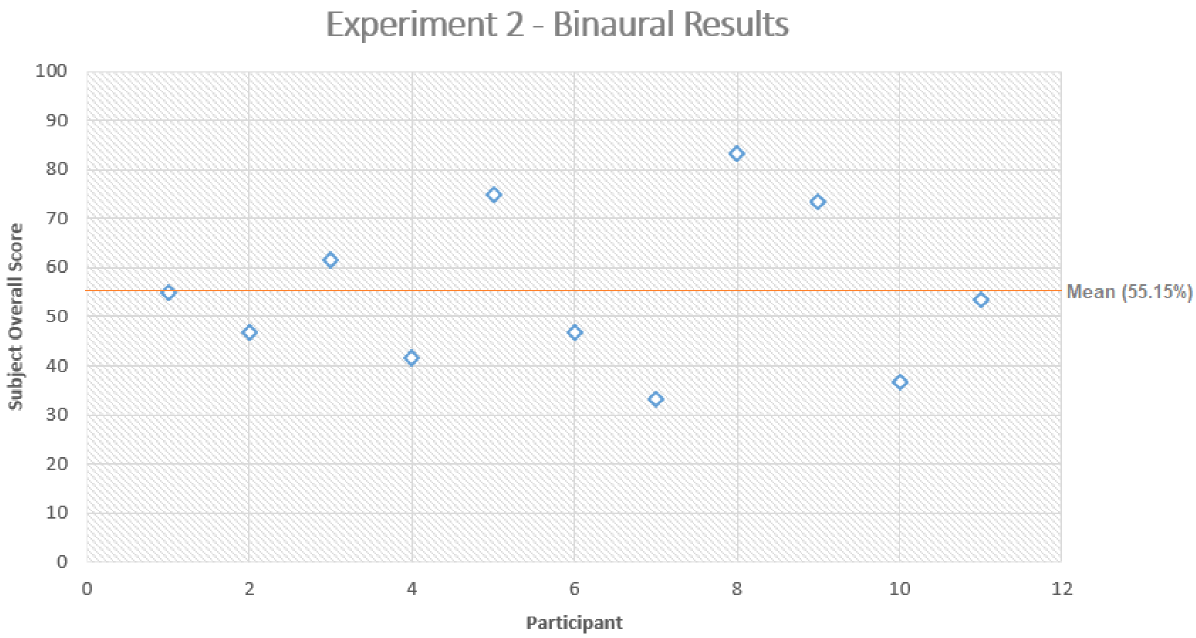


Figure 29 Graph of overall results from all Experiment 2 subjects (headphone)

The open-field test (conducted using the same procedure but through loudspeaker playback as opposed to headphones) provided results seen in Figure 30 and statistical observations were made in Table 7.

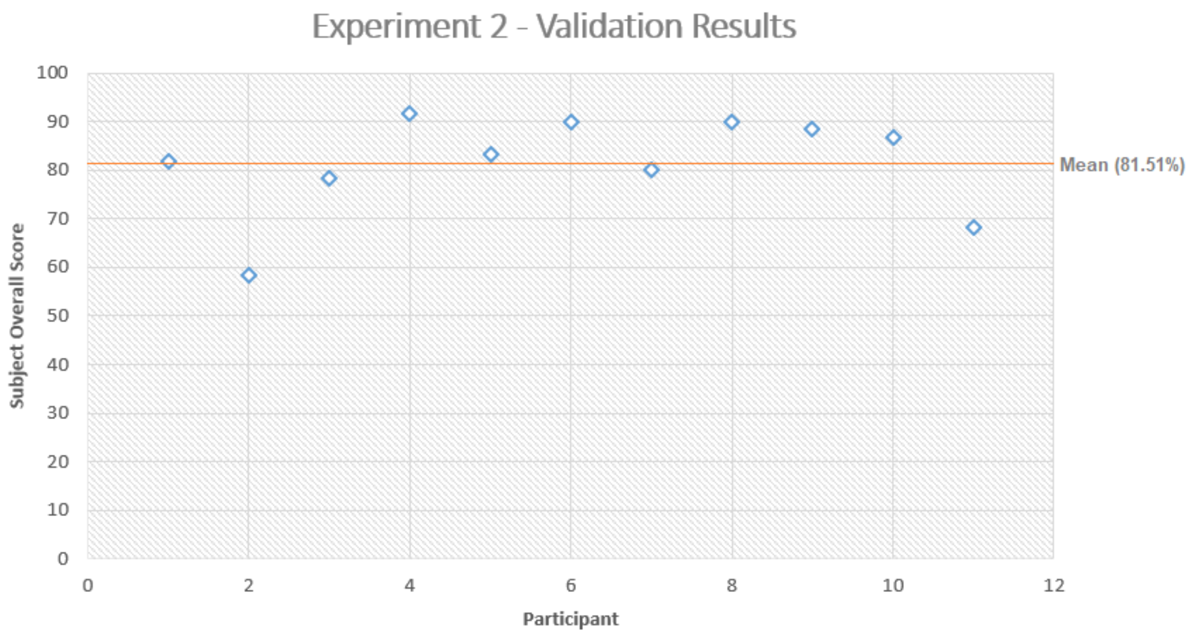


Figure 30 Graph of overall results from all experiment 2 subjects (open-field)

Table 7 Statistical results from Experiment 2 (open-field)

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
81.51%	83.33%	90.00%	91.67%	58.33%

The decreased efficiency from the open-field test (natural-hearing test) to the headphone test (binaural test) was observed and the error is presented as a percentage function (Table 8).

*Table 8 The percentage error of each statistical overall result ((Open-field – Headphone)/Open-field * 100)*

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
-32.34%	-36.00%	-48.14%	-9.09%	-42.86%

All statistical differences from subjects were also observed (Figure 31) and averages were once again calculated. This produced differences as shown in Table 9.

Table 9 The average decrease in overall results from open-field to headphone in Experiment 2

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
26.36%	16.66%	50.00%	50.00%	6.67%

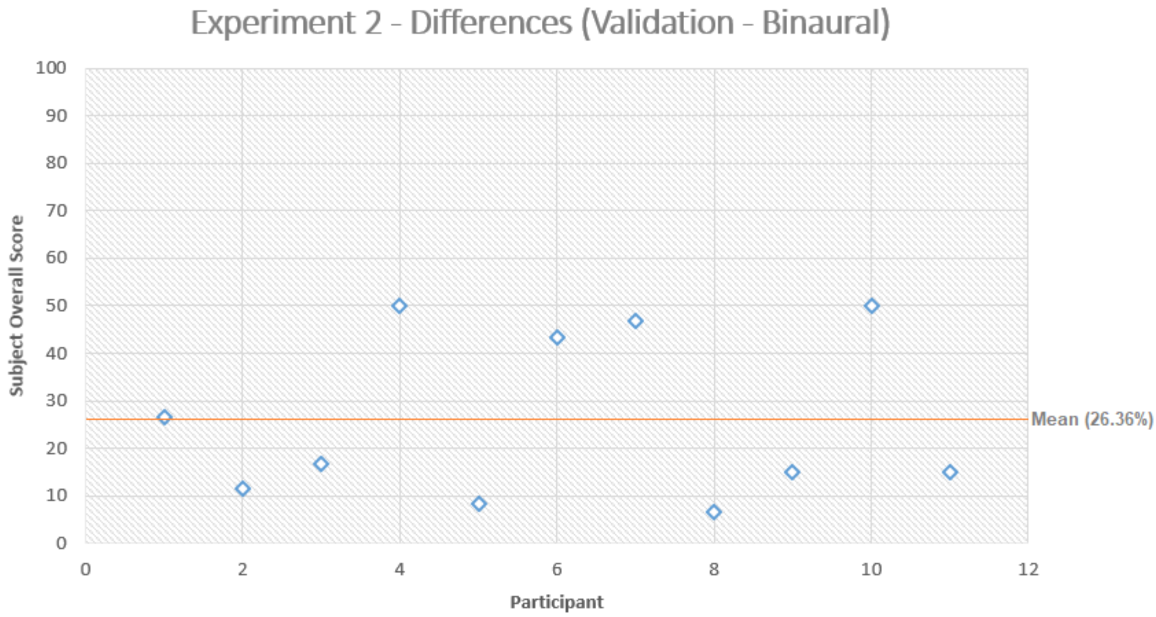


Figure 31 Graph of differences between both tests in Experiment 2 (open-field – headphone)

Finally, employing the use of the validation method presented in the methodology chapter, the headphone/binaural results which achieved a higher score than their validated counterparts were excluded along with the subjects which were awarded a score below the pass threshold of 45.00%. Out of 11 subjects, no subjects scored higher on the headphone test compared to the open-field test, and all 11 subjects were able to achieve an overall validation score of 45.00% or above, resulting in a pass rate of 100.0% (11 out of 11). As such, the results presented in Table 10 and Figure 32 match those of the raw binaural results presented at the start of this sub-section. The lower population size meant that each subject contributed as 9.09% towards the overall pass rate. Using only the validated subjects, the new overall binaural results are presented in Table 10.

Table 10 Overall results from validated-only subjects for Experiment 2 (headphone)

MEAN:	MEDIAN:	MODE:	MAX:	MIN:
55.15%	53.33%	46.67%	83.33%	33.33%

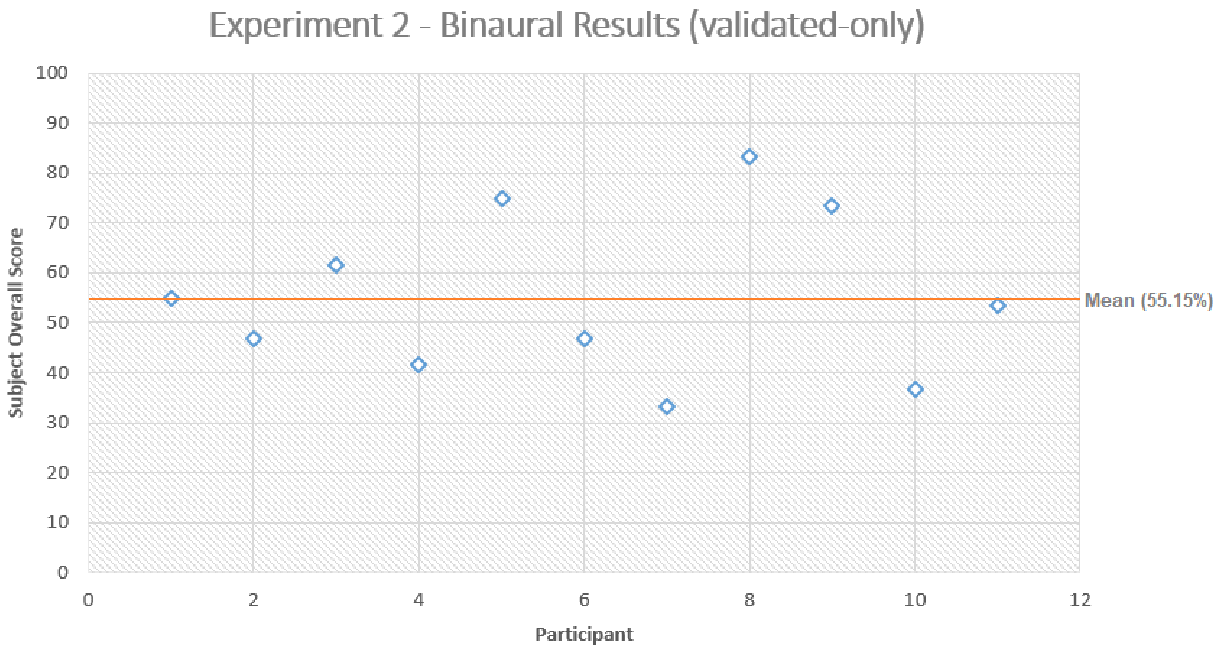


Figure 32 Graph of overall result from validated-only subjects in Experiment 2 (headphone)

6.1.3 Further Analysis and Comparison of Experiments 1 and 2

In an attempt to ensure that both sound stimuli were viable in terms of localisation ability, the open-field/validation tests were observed independently of the headphone/binaural test to investigate their pass rates. Stimulus 1 was successfully completed by 79.41% of subjects (27 out of 34, based once again on a 45.00% pass mark). Stimulus 2 was successfully completed by all 11 subjects, thus having a 100.00% pass rate. As an additional note, the minimum achieved overall score of Stimulus 2 was 58.33%, 13.33% points above the pass mark. The new statistics of successful-subjects-only for both stimuli and open-field experiments are provided in Table 11.

Table 11 A more accurate representation of successful subjects from both validation experiments

Stimulus/Experiment	MEAN:	MEDIAN:	MODE:	MAX:
1	63.68%	61.67%	61.67%	83.33%
2	81.51%	83.33%	90.00%	91.67%

To additionally investigate the validity of subjects and their results, the comparison between individual locations and the consistency of answers was undertaken. Firstly, the standard deviation across all overall results was calculated to check for variation/dispersion of scores. Table 12 summarises the standard deviations for all experiments.

Table 12 Summary of all standard deviations across both experiments, where verified results include only the results above a pass mark of 45%

	Unverified		Verified	
Experiment:	**MEAN:**	**ST.DEV (plain):**	**MEAN:**	**ST.DEV (plain):**
1 – Headphone	45.88%	14.00%	56.90%	8.06%
1 – Open-field	59.06%	13.68%	63.68%	10.16%
2 – Headphone	55.15%	15.69%	61.88%	12.97%
2 – Open-field	81.51%	9.76%	81.51%	9.76%

The results above indicate a similar overall result spread of answers within an area above or below the average to a difference between 0% and 8.06-15.69%. To investigate the consistency of individual answers, the percentage results of attempted answers for each correct location are presented in Table 13 and Table 15 for Experiment 1, along with Table 17 and Table 19 for Experiment 2. The total percentage of subjects awarded 3, 2, 1 or 0 pts. respectively for each location, are also shown. Specifically, the percentages indicate how many subjects were able to answer each given location to each degree of error or separation. The results are also presented visually in bar charts in Figure 33, Figure 34, Figure 35 and Figure 36.

6.1.3.a Experiment 1 – Headphone

Table 13 The order of locations and results of Experiment 1 along with the various degrees of acceptable error (headphone)

Correct Loc.	Correct Ans.	Correct Ans.+/-1	Correct Ans.+/-2	Incorrect Ans.
6	32.4%	44.1%	11.8%	11.8%
18	20.6%	44.1%	14.7%	20.6%
17	5.9%	11.8%	32.4%	50.0%
8	8.8%	32.4%	35.3%	23.5%
9	5.9%	11.8%	38.2%	44.1%
5	17.6%	55.9%	11.8%	14.7%
16	8.8%	11.8%	23.5%	55.9%
24	8.8%	52.9%	2.9%	35.3%
7	2.9%	44.1%	26.5%	26.5%
11	14.7%	35.3%	11.8%	38.2%
20	8.8%	26.5%	26.5%	38.2%
10	8.8%	17.6%	20.6%	52.9%
12	20.6%	20.6%	14.7%	44.1%
19	17.6%	41.2%	17.6%	23.5%
2	47.1%	23.5%	2.9%	26.5%
13	14.7%	26.5%	17.6%	41.2%
4	23.5%	29.4%	23.5%	23.5%
1	50.0%	29.4%	5.9%	14.7%
23	11.8%	29.4%	14.7%	44.1%
5	20.6%	41.2%	11.8%	26.5%

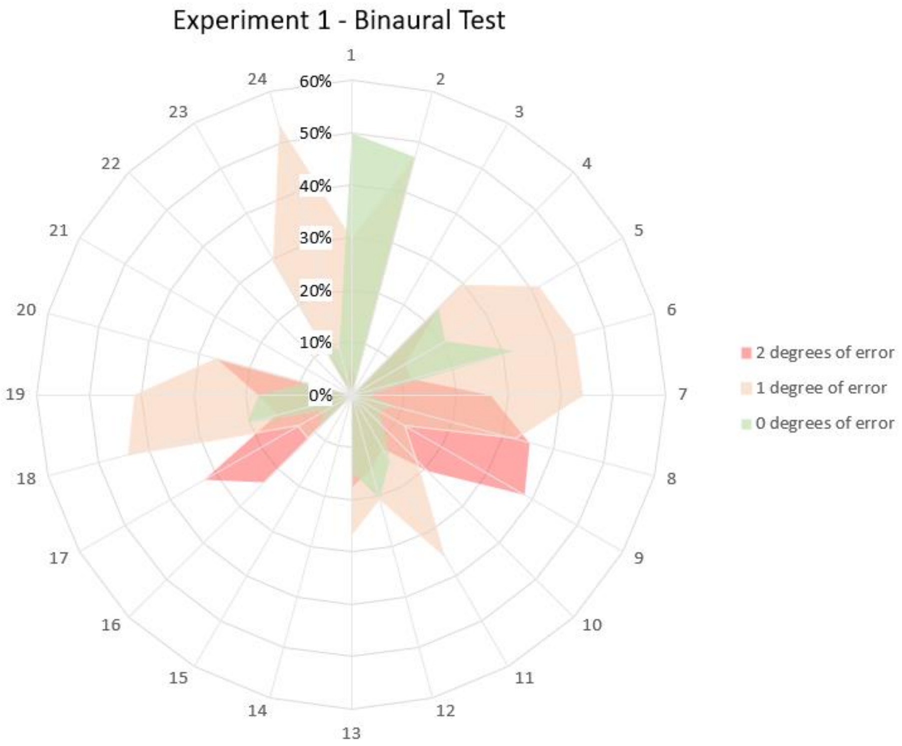


Figure 33 Headphone results of Experiment 1 for each degree of acceptable error

To determine the partial success of the experiment, the repeating (consistency) number within this experiment (5) was checked for a percentage error between the higher and the lower values. Location 5 was answered correctly to the exact location by a total of 17.6% subjects, while the latter location 5 was correctly answered 20.6% of the time. This produced an error difference of 14.56% ((20.6-17.6)/20.6 * 100). The error difference was also observed for the other two degrees of acceptable error (separation) (see Table 14).

Table 14 The error difference between consistency location (#5) for Experiment 1 (headphone)

Correct Ans.	Correct Ans.+/-1	Correct Ans. +/-2	Incorrect Ans.
-14.56%	-26.30%	0%	-44.53%

6.1.3.b Experiment 1 – Open-field

Table 15 The order of locations and results of Experiment 1 along with the various degrees of acceptable error (open-field)

Correct Loc.	Correct Ans.	Correct Ans.+/-1	Correct Ans.+/-2	Incorrect Ans.
6	41.2%	44.1%	8.8%	5.9%
18	20.6%	38.2%	23.5%	17.6%
17	8.8%	8.8%	29.4%	52.9%
8	5.9%	35.3%	32.4%	26.5%
9	17.5%	23.5%	20.6%	38.2%
5	41.2%	35.3%	11.8%	11.8%
16	14.7%	11.8%	14.7%	58.8%
24	58.8%	26.5%	0.0%	14.7%
7	11.8%	44.1%	26.5%	17.6%
11	11.8%	32.4%	11.8%	44.1%
20	41.2%	32.4%	14.7%	11.8%
10	17.7%	32.4%	11.8%	38.2%
12	32.4%	32.4%	2.9%	32.4%
19	38.2%	38.2%	8.8%	14.7%
2	67.6%	26.5%	0.0%	5.9%
13	32.4%	23.5%	8.8%	35.3%
4	38.2%	35.3%	11.8%	14.7%
1	79.4%	14.7%	0.0%	5.9%
23	38.2%	41.2%	5.9%	14.7%
5	41.2%	32.4%	20.6%	5.9%

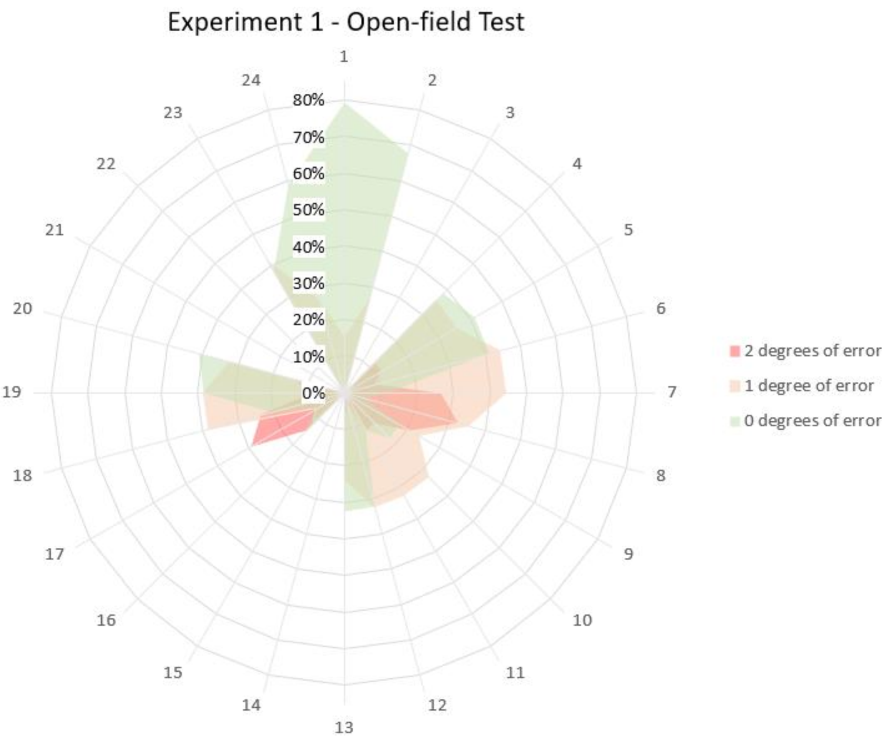


Figure 34 Open-field results of Experiment 1 for each degree of acceptable error

Again, the consistency location (5) is observed for its error rate in Table 16.

Table 16 The error difference between consistency location (5) for Experiment 1 (open-field)

Correct Ans.	Correct Ans.+/-1	Correct Ans. +/-2	Incorrect Ans.
0%	-8.30%	-42.86%	-50.00%

6.1.3.c Experiment 2 – Headphone

Table 17 The order of locations and results of Experiment 2 along with the various degrees of acceptable error (headphone)

Correct Loc.	Correct Ans.	Correct Ans.+/-1	Correct Ans.+/-2	Incorrect Ans.
19	63.6%	18.2%	9.1%	9.1%
1	54.5%	9.1%	0.0%	36.4%
14	36.4%	27.3%	18.2%	18.2%
17	36.4%	54.5%	9.1%	0.0%
24	63.6%	0.0%	9.1%	27.3%
10	0.0%	45.5%	18.2%	36.4%
9	18.2%	36.4%	0.0%	45.5%
22	45.5%	9.1%	0.0%	45.5%
11	9.1%	36.4%	9.1%	45.5%
12	45.5%	18.2%	9.1%	27.3%
2	36.4%	9.1%	9.1%	45.5%
4	27.3%	36.4%	9.1%	27.3%
5	18.2%	45.5%	36.4%	0.0%
23	45.5%	9.1%	0.0%	45.5%
20	54.5%	9.1%	18.2%	18.2%
14	36.4%	9.1%	45.5%	9.1%
21	27.3%	9.1%	18.2%	45.5%
1	36.4%	36.4%	0.0%	27.3%
15	18.2%	45.5%	0.0%	36.4%
3	18.2%	45.5%	9.1%	27.3%

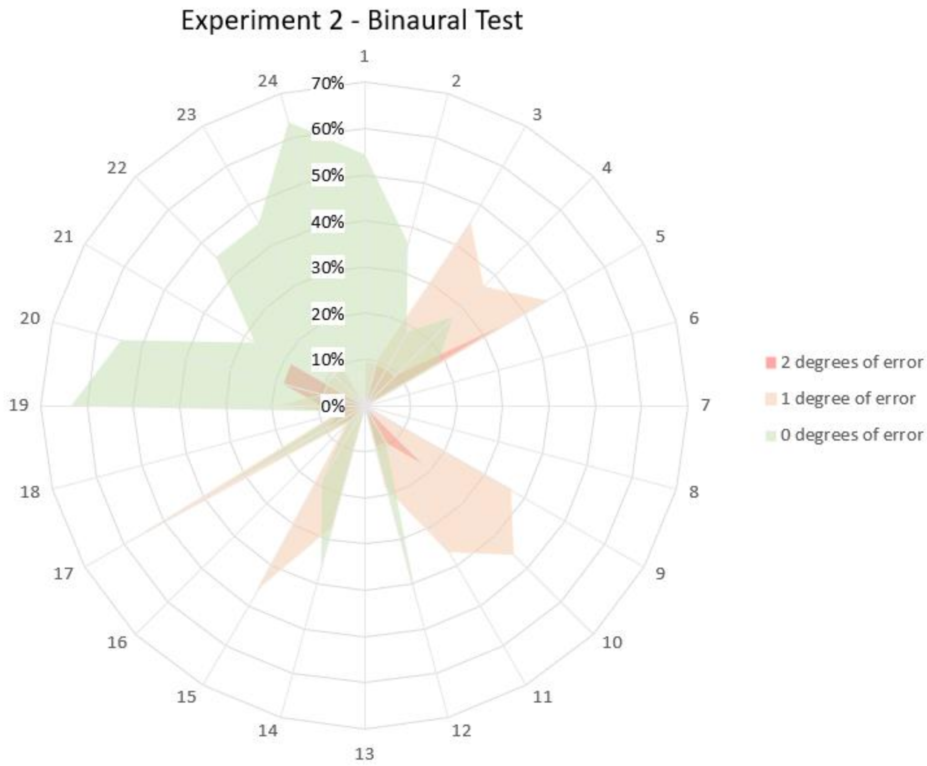


Figure 35 Headphone results of Experiment 2 for each degree of acceptable error

Table 18 The error difference between consistency location (14) for Experiment 2 (headphone)

Correct Ans.	Correct Ans.+/-1	Correct Ans. +/-2	Incorrect Ans.
0%	-66.67%	-60.00%	-50.00%

6.1.3.d Experiment 2 – Open-field

Table 19 The order of locations and results of Experiment 2 along with the various degrees of acceptable error (open-field)

Correct Loc.	Correct Ans.	Correct Ans.+/-1	Correct Ans.+/-2	Incorrect Ans.
19	36.4%	63.6%	0.0%	0.0%
1	90.9%	0.0%	0.0%	9.1%
14	54.4%	36.4%	0.0%	9.1%
17	45.5%	27.3%	9.1%	18.2%
24	81.8%	0.0%	9.1%	9.1%
10	72.7%	27.3%	0.0%	0.0%
9	54.5%	45.5%	0.0%	0.0%
22	45.5%	0.0%	9.1%	45.5%
11	72.7%	18.2%	0.0%	9.1%
12	54.5%	36.4%	9.1%	0.0%
2	81.8%	18.2%	0.0%	0.0%
4	54.5%	27.3%	9.1%	9.1%
5	36.4%	63.3%	0.0%	0.0%
23	72.7%	9.1%	0.0%	18.2%
20	45.5%	36.4%	0.0%	18.2%
14	45.5%	36.4%	9.1%	9.1%
21	54.5%	36.4%	9.1%	0.0%
1	90.9%	9.1%	0.0%	0.0%
15	45.5%	27.3%	9.1%	18.2%
3	63.6%	36.4%	0.0%	0.0%

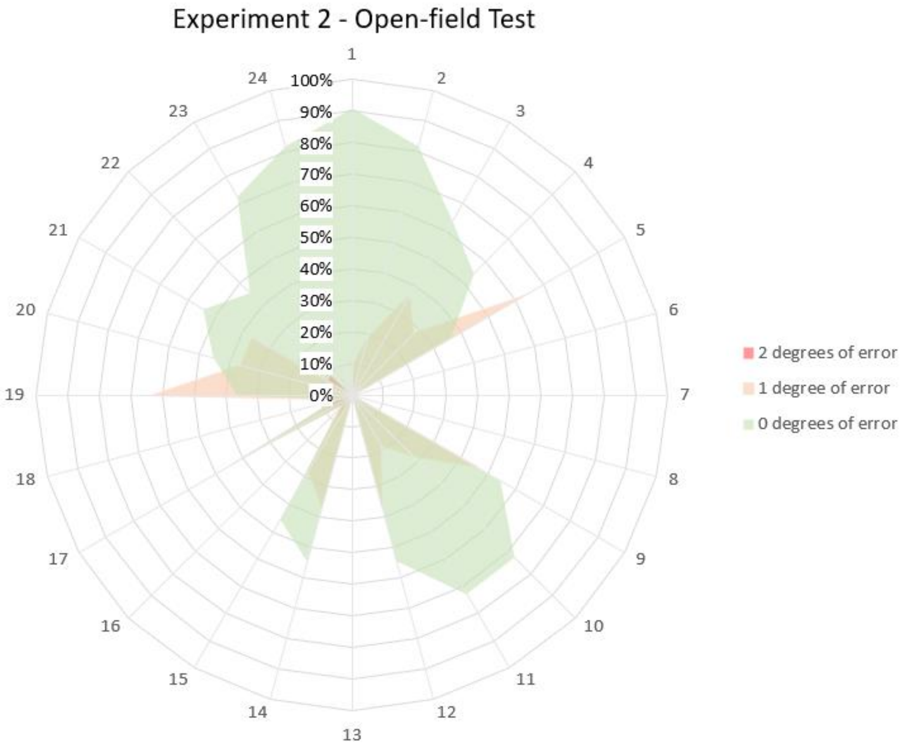


Figure 36 Open-field results of Experiment 2 for each degree of acceptable error

Table 20 The error difference between consistency location (14) for Experiment 2 (open-field)

Correct Ans.	Correct Ans.+/-1	Correct Ans. +/-2	Incorrect Ans.
-16.67%	0.00%	-100.00%	0.00%

6.1.3.e Comparison of Consistency Locations from Headphone versus Open-field

The precisely correct answers for consistency locations (5 and 14 for Experiments 1 and 2 respectively) were observed for similarities and were further checked for average decreases from validation test to binaural test in their respective experiments, as previously undertaken for overall results seen in sections 6.1.1 through 6.1.3.d. To distinguish between the two instances of location 5 and the two instances of location 14, they are referred to as 5A, 5B and 14A, 14B respectively in this section, with 5A and 14A being the former of both relative experiments (1 and 2). These numbers show 2 of the 20 locations that contributed to the average decrease in localisation efficacy from the open-field test to the headphone test, calculated in section 6.1.1 and 6.1.2. Note that these refer solely to the exact correct answers, and additionally, take into account all subjects and do not exclude any based on the validation procedure seen previously. Table 21 shows the values related to Experiment 1 while

Table 22 provide them for Experiment 2.

*Table 21 The percentage error difference of consistency locations in Experiment 1 ((open-field – headphone)/open-field * 100)*

5A (Headphone)	17.6%
5A(Open-field)	41.2%
5A (Error Difference)	-57.3%
5B (Headphone)	20.6%
5B (Open-field)	41.2%
5B (Error Difference)	-50.0%

*Table 22 The percentage error difference of consistency locations in Experiment 2 ((Open-field – Headphone)/Open-field * 100)*

14A (Headphone)	36.4%
14A (Open-field)	54.4%
14A (Error Difference)	-33.1%
14B (Headphone)	36.4%
14B (Open-field)	45.5%
14B (Error Difference)	-20.0%

6.2 Verification Check – Simulation

All five simulations were executed using a sample (theoretical subject) size of 2000, with varying correct locations/answers for each of the five different tests. These simulations were conducted an additional four times to create an average of averages, as seen in section 6.3.4. Sections 6.2.1, 6.2.2 and 6.2.3 show one of these five simulations for each of the five different datasets. Minimum scores were omitted from these results as each simulation produced a significant number of overall scores with a result of 0. These can instead be seen in the graphs for each respective dataset. The graphs in each of these subsections represent each hypothetical subject's overall result, with the result (in percentage) along the x-axis and the participant (20-21 per subject to allow distance between each point). The larger cluster of points represents the average overall score, as indicated by a mean trend line. The respective mean, median, mode and max values are presented in tabular form above each graph.

6.2.1 Result for Dataset 1

The first simulation, based on the order of numbers (virtual locations) given in section 5.7.1, provided the results given in Table 23.

Table 23 Statistical results of Dataset 1 (simulation)

MEAN:	MEDIAN:	MODE:	MAX:
12.56%	11.67%	11.67%	33.33%

The average distribution of overall result scores can be seen in Figure 37.

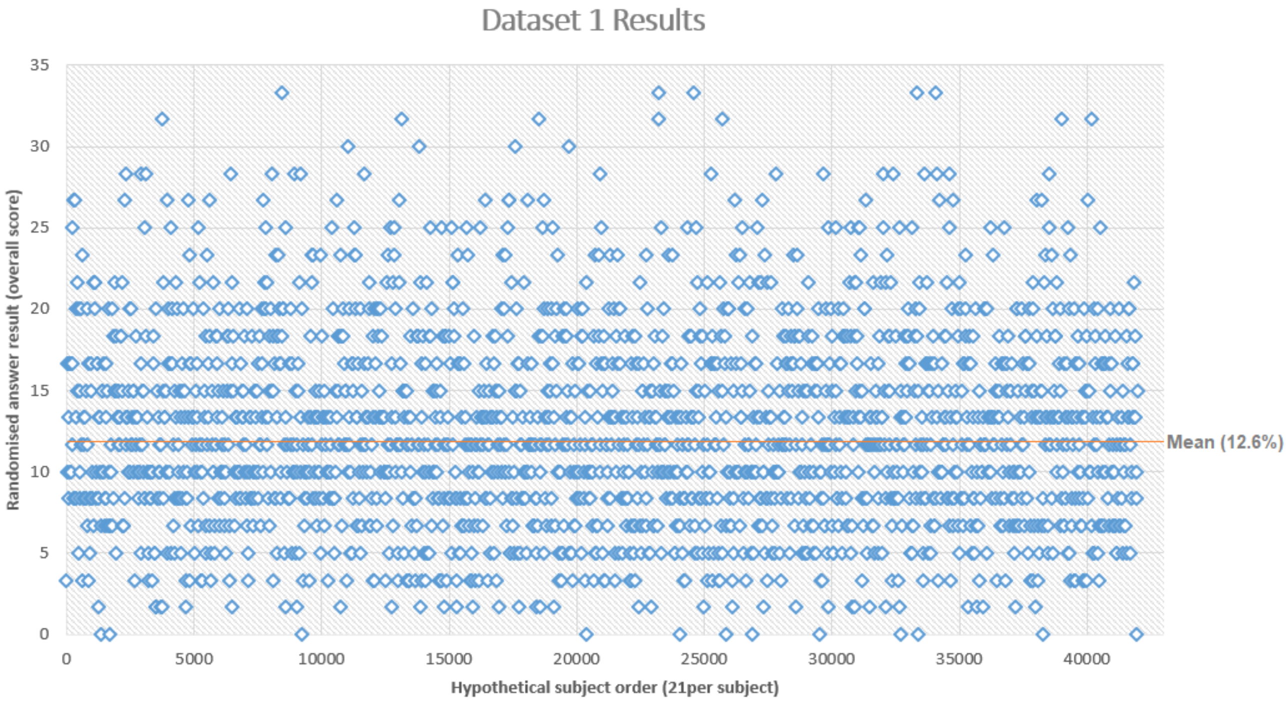


Figure 37 Dataset 1 simulation results

6.2.2 Result for Dataset 2

The numbers used for the second simulation (see section 5.7.2) resulted in the figures provided in Table 24:

Table 24 Statistical results of Dataset 2 (simulation)

MEAN:	MEDIAN:	MODE:	MAX:
12.40%	11.67%	10.00%	38.33%

Again, the full distribution of numbers is shown in Figure 38;

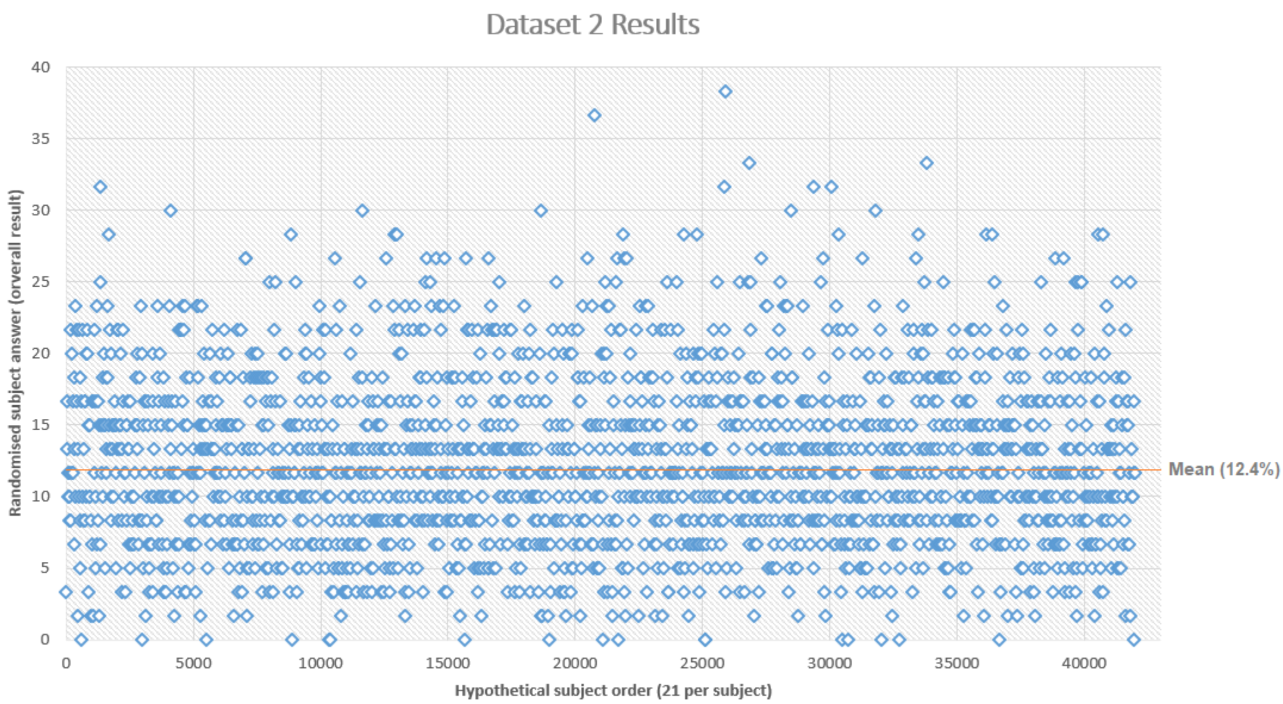


Figure 38 Dataset 2 simulation results

6.3.3 Results for Datasets 3, 4 and 5

Simulation 3, the first of three verification checks for number consistency and true randomisation, was an ascending order of numbers from 3-22 and produced statistical results seen in Table 25.

Table 25 Statistical results of Dataset 3 (simulation)

MEAN:	MEDIAN:	MODE:	MAX:
12.55%	11.67%	11.67%	38.33%

Simulation 4, simulated using the location 12 as the correct number for each possible answer/location produced the results shown in Table 26.

Table 26 Statistical results of Dataset 4 (simulation)

MEAN:	MEDIAN:	MODE:	MAX:
12.39%	11.67%	10.00%	35.00%

The last simulation, simulation 5 (the latter of the three verification simulations), was conducted using a double-randomised set of numbers where each theoretical subject had their own unique correct set of answers. The statistical results of this simulation can be seen in Table 27.

Table 27 Statistical results of Dataset 5 (simulation)

MEAN:	MEDIAN:	MODE:	MAX:
12.19%	11.67%	11.67%	36.67%

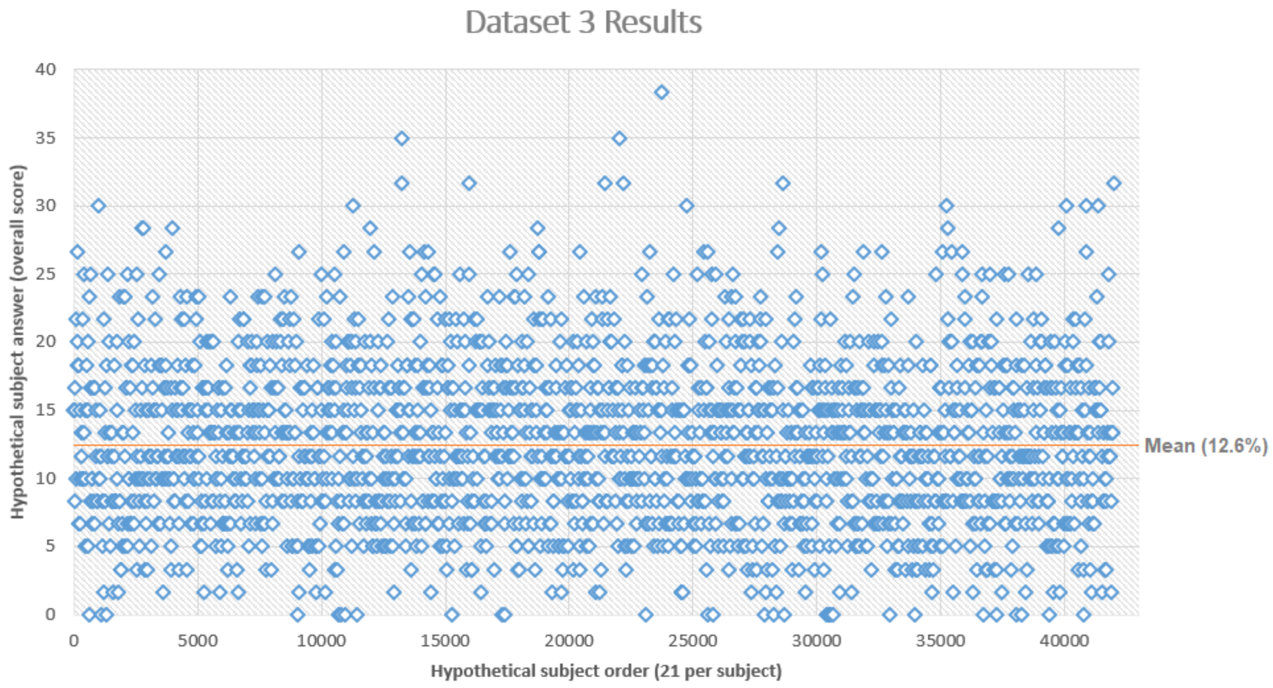


Figure 39 Dataset 3 simulation results

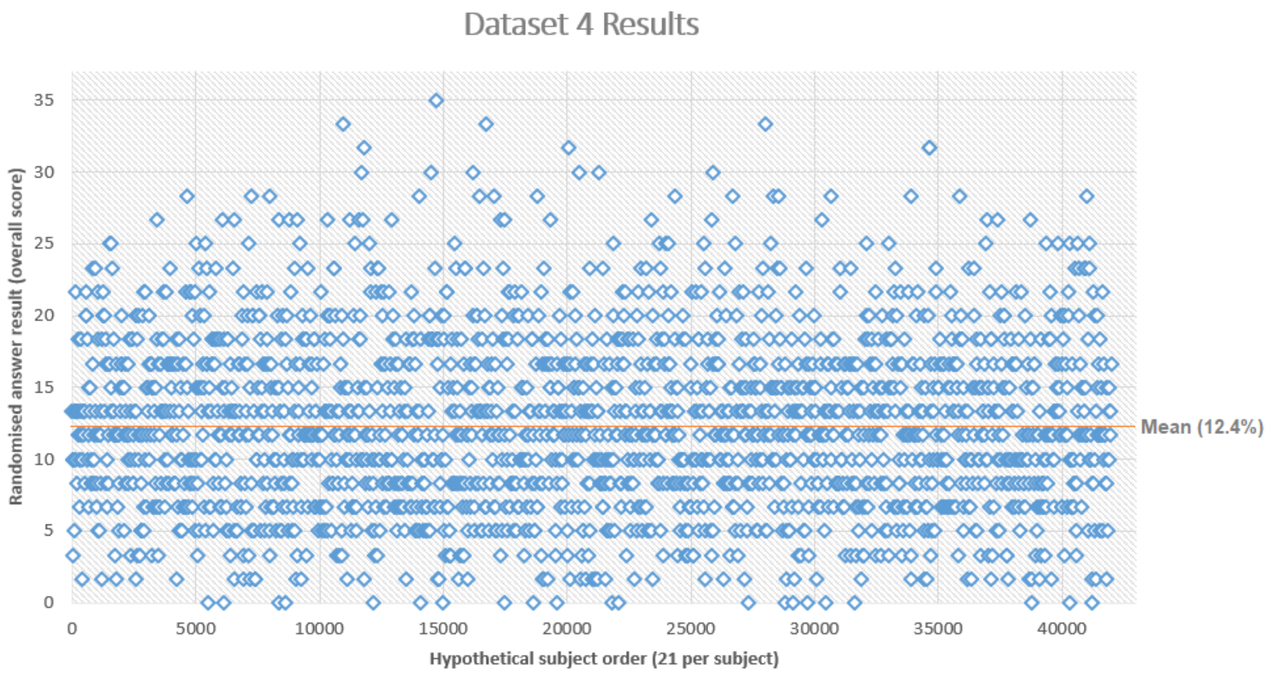


Figure 40 Dataset 4 simulation results

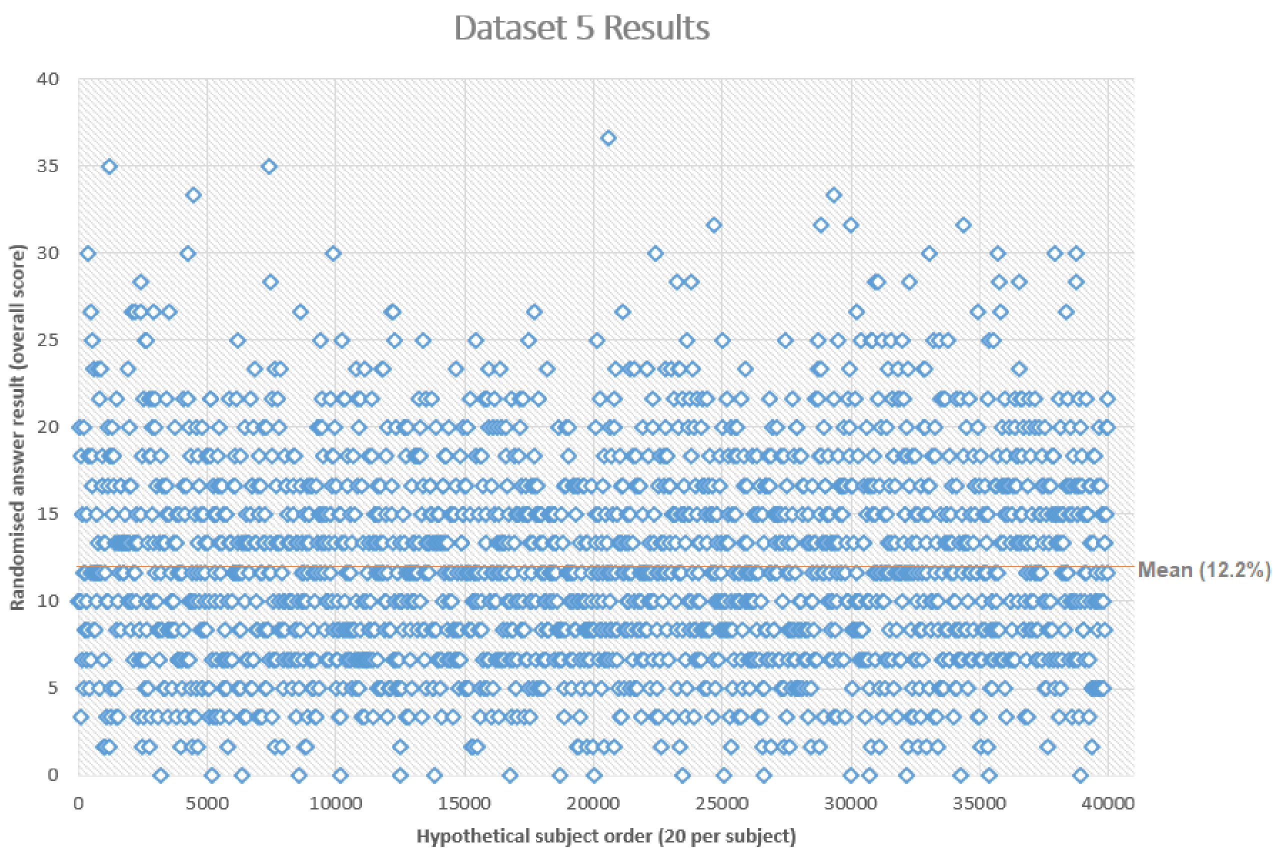


Figure 41 Dataset 5 simulation results

6.3.4 Simulation Analysis

All five simulations produced statistical results within 5% of each other, e.g. the mode of simulations 1 through 5 is 11.67%, 10.00%, 11.67%, 10.00% and 11.67% respectively, and the maximum of each simulation is 33.33%, 38.33%, 38.33%, 35.00% and 36.67%. The larger gap between the maximum values of each simulation could be owing to occasional high and low result (stray). To investigate this further, the simulations were undertaken an additional four times for each simulation, totalling 10,000 theoretical subject results for each dataset. The following averages and highest values of each simulation observed are shown in Table 28 and Table 29 respectively.

Table 28 Statistical results of repetitions of all five datasets (simulation)

Dataset	MEAN (1-5)	MAX (1-5)
1	12.56, 12.57, 12.93, 12.50, 12.58	41.66, 36.67, 38.33, 40.00, 33.33
2	12.50, 12.32, 12.44, 12.53, 12.53	35.00, 35.00, 36.67, 35.00, 35.00
3	12.55, 12.48, 12.41, 12.26, 12.46	43.33, 35.00, 36.67, 36.67, 35.00
4	12.29, 12.40, 12.50, 12.45, 12.40	35.00, 40.00, 36.67, 36.67, 33.33
5	11.97, 12.19, 11.79, 11.81, 11.97	40.00, 35.00, 33.33, 33.33, 36.67

This created the following averages for each respective category:

Table 29 New averages for the results of all five datasets (simulation)

Dataset	Average (MEAN)	Average (MAX)
1	12.63	38.00
2	12.46	35.33
3	12.43	37.33
4	12.41	36.33
5	11.95	35.67

The results shown in Table 29 suggest that any subject attempting an experiment, through pure ‘guesswork’, using the generational procedure for order of locations seen in section 5.1 can only be expected to achieve between 12-13% overall result. Furthermore, the highest overall result achieved ranges between 33.33% to 43.33%, with an average of 35.33% to 38.00% depending on the dataset used. It is important to note that datasets 1 and 2 followed a more rigorous selection procedure (order of locations, again, seen in section 5.1) and explain the higher and lower values of the MEAN and MAX respectively. Dataset 5 achieved a lower result of averages owing to the randomisation procedure of both the theoretically correct answer set as well as the simulated answer attempt, thus representing the potential of guesswork across 2000 different experiments instead. Furthermore, the results indicate that in any given dataset a subject can only achieve a maximum score of 43.33% in the likelihood of 1/10,000 (e.g. Dataset 1). The statistical distribution in terms of standard deviation, is discussed in section 7.1.4.

6.4 Summary

This chapter presented the raw results and data gathered from Experiment 1 and Experiment 2 along with the results obtained during the simulation of the experiments. Additionally, it compared and analysed some of these results for various observations. Chapter 7 will review and discuss these findings to determine the success of the initial aim and objectives.

Chapter 7: Discussion

Chapter 6 presented the findings and data gathered from the experiments conducted during this research. Additionally, it compared and analysed this data for various correlations and/or similarities. This chapter discusses the initial theory against the outcome and findings of the results gathered. Furthermore, it critically evaluates the possibility of creating a standardised testing procedure, and any possible challenges or further requirements, that determines the possibility, and thus viability, of creating a qualitative measurement of a binaural systems' efficiency.

7.1 Initial Observation of Results and Stimuli

In order to discuss the observations of all the experiments conducted, along with their respective findings, the following is summarised for convenience:

Two experiments (Experiment 1 and 2) were conducted using the same procedure with the exceptions of the stimulus and the order of locations. Each experiment consisted of two tests: (a) Headphone test - a binaural test conducted through headphones with a pre-recorded stimulus played through the respective order of locations and, (b) Open-field test - a validation test which tested subjects' localisation abilities in the free-hearing domain through the same stimulus and order of locations used during the headphone test of that experiment.

A simulation of the experimental procedure was executed with a population size of 2,000 which was repeated five times to achieve a running average. Five different orders of locations were investigated. Datasets 1 and 2 determined the various statistics and probability of guesswork in the experiment, whilst datasets 3 through 5 were used as verification methods to investigate potential biasing in the number generation.

The position of each correct location is shown again in Figure 42.

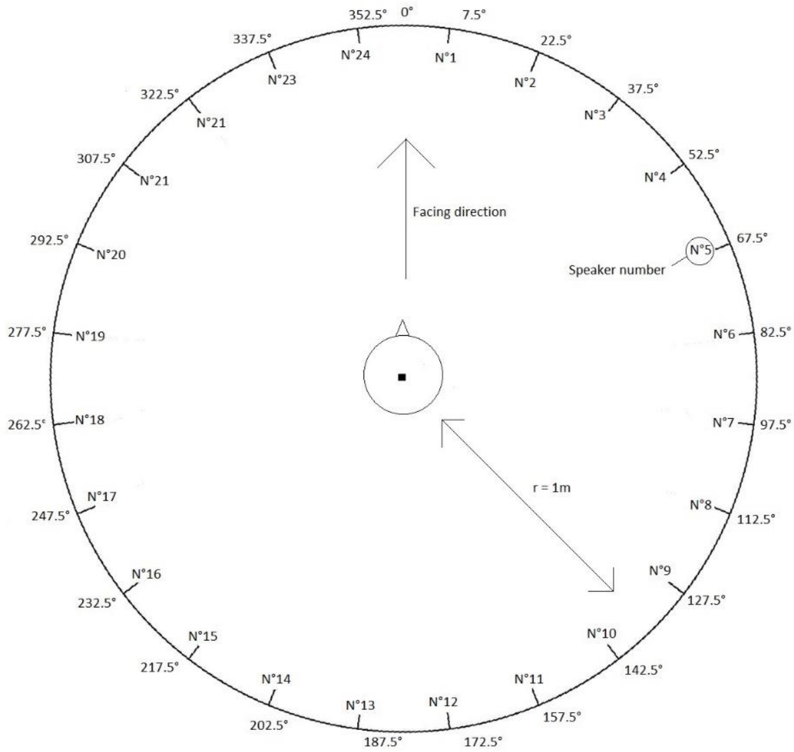


Figure 42 Position of subject and loudspeakers in the loudspeaker array (aerial perspective) (copy of Figure 20)

7.1.1 Experiment 1

The procedure for Experiment 1 was given in section 5.4.1, while the results were presented in section 6.1.1. This experiment showed a variation of results in terms of overall percentages, with particular locations being approximated consistently correct or incorrect, across different subjects. The 'difficult' locations (individual locations with a low-scoring average, i.e. below 45%) were mostly harder to locate for all participating subjects. This could be owing to the location being in a harder position to locate, such as locations 11 through 13 owing to front-back confusion, or due to the particular stimuli used. Figures varied for the validation procedure of Experiment 1, with the mean result achieving a higher value per location as well as per overall result of subjects. The lower-achieving results of the headphone test are speculated to be related to the limitations of the headphones compared to the free-hearing listening environment conducted in the open-field test. The use of in-ear monitors (blocked-meatus method, section 2.3.1.c) along with individualised and characterised HRTFs could possibly improve the results of the headphone test. Some higher overall results within the headphone test could be owing to the likelihood that the HRTFs of the headphones and system used, match the subject. Naturally, this also functions as a disadvantage to certain subjects/users.

The majority of subjects were able to achieve an overall score between 30% and 60% for the headphone test, whilst the simulation results suggest that with any given order of locations a subject attempting to guess the experiment can only be expected to achieve, a score of 12.60%, with maximum averages of 38% for a population size of 10,000. This suggests that subjects participating in the experiment employed some level and use of estimation of localising abilities within the experiment. The majority of subjects participating in the open-field test of this experiment were able to achieve overall scores between 46% of 72%. Following the open-field test, subjects who may have impaired hearing, or a distinct lack of localisation ability (subjects with overall scores of <45% on the validation test), were excluded from the initial results of the headphone test, producing a new set of averages and validated results as seen in section 6.1.1, with the new majority of scores being between 49% and 75% on the headphone test. The increase in the mean from 48.5% to 59%, suggests the necessity of an open-field test to exclude subjects who are less capable of localising a sounds' source or direction even in the natural free-hearing domain.

Additional observations are made on the 'easy' and 'difficult' locations to localise, that is, the highest and lowest scoring locations respectively. The lowest and highest scoring locations, and percentage results, for each degree of accuracy are presented in Table 30.

Table 30 Highest and lowest scoring locations in Experiment 1

		3 points	2 points	1 point	0 points
Headphone	**Highest**	1 (50%)	5 (55.9%)	9 (38.2%)	16 (55.9%)
	Lowest	7 (2.9%)	9, 16, 17 (11.8%)	2, 24 (2.9%)	6 (11.8%)
Open-field	**Highest**	1 (79.4%)	6, 7 (44.1%)	8 (32.4%)	17 (52.9%)
	Lowest	17 (8.8%)	17 (8.8%)	1, 24 (0%)	1, 2, 5, 6 (5.9%)

Both tests of the experiment show that the highest point of accuracy is location 1, with more than half of the subjects being able to estimate the location to the exact, correct, loudspeaker. Logically, there are therefore very few subjects who were not able to estimate location 1 to any degree at all, with only 14.7% and 5.9% of subjects for the headphone and open-field tests respectively, being unable to localise location 1. This is most likely owing to the natural tendency to localise sources directly in-front.

The lowest scoring location is split between location 7 and 17, which are located at 112.5°, and 262.5° degrees azimuth respectively. These locations are located just beyond the frontal hemisphere, where localisation abilities reduce significantly. This is additionally seen in the 0-points column of the table, whereby location 17 achieved the highest percentage of incorrect guesses, with 52.9% of subjects being unable to approximate it to a range of 60° azimuth.

Table 30 shows similar correlations and localisation abilities within both tests, with locations being within the same approximate, relative, area of each other (i.e. within 30° azimuth left or right). Location 17 occurred regularly in the lowest values whilst location 6 often appeared within the highest point values. As such, it can be speculated that localisation abilities are generally consistent between the natural hearing abilities and that of the reproduction of binaural hearing through a system, with a relative decrease in overall efficiency, in this particular binaural system.

7.1.2 Experiment 2

Experiment 2 was conducted using the procedure shown in section 5.4.2 and the results of this experiment presented in section 6.2.2. Similar observations were made for Experiment 2 as with Experiment 1. Most of the subjects who participated in Experiment 2 were able to achieve an overall score of 40% to 70% on the headphone test and 71% to 91% for the open-field test. The higher results, compared to Experiment 1, could be owing to a lower population size and thus equate to an inflated percentage (each subject contributing 9.09% towards the overall average, compared to that of Experiment 1 which is 2.94%), but it could also be that the stimulus used in Experiment 2 was easier to localise. This could be owing to the wider range of frequencies within the stimulus, or the length of exposure, or a combination of both.

Again, the highest and lowest scoring locations for each degree of error have been extracted and are presented in Table 31.

Table 31 Highest and lowest scoring locations for Experiment 2

		3 points	2 points	1 point	0 points
Headphone	**Highest**	19, 24 (63.6%)	17 (54.5%)	14 (45.5%)	2, 9, 11, 21, 22, 23 (45.5%)
	Lowest	10 (0.0%)	24 (0.0%)	1, 1, 9, 15, 22 (0.0%)	5, 17 (0.0%)
Open-field	**Highest**	1, 1 (90.9%)	19 (63.6%)	1, 1, 2, 3, 5, 9, 10, 11, 14, 19, 20, 23 (9.1%)	22 (45.5%)
	Lowest	5 (36.4%)	1, 24 (0.0%)	4, 12, 14, 15, 17, 21, 22, 24 (0.0%)	1, 2, 3, 5, 9, 10, 12, 19, 21 (0.0%)

The highest scoring ('easiest') locations in the headphone test were tied between 19 and 24, whilst the open-field test was tied between both repetitions of location 1, at 90.9%. It may be worthwhile to note that location 1 is adjacent to location 24, by 15° azimuth to the right, and that location 1 scored 54.5%, making it the second highest scoring location.

The most difficult locations to localise were 10 and 5 for the headphone and open-field tests respectively. However, location 5 also scored 0.0% in the 0 points column for both tests, meaning all subjects were able to locate it to 15° or 30° azimuth left or right, but not necessarily to the exact location. This could be owing to the lower population size and the results are therefore divided between correctly, and incorrectly estimated locations. Table 32 shows the results of each location for the headphone and open-field tests, arranged in descending order of accuracy.

Table 32 The locations, by descending order of highest correct estimated locations for the headphone, and open-field, tests in Experiment 2

Headphone	Any points	0 points		Open-field	Any points	0 points
5	100.0%	0.0%		1	100.0%	0.0%
17	100.0%	0.0%		2	100.0%	0.0%
14	90.9%	9.1%		3	100.0%	0.0%
19	90.0%	9.1%		5	100.0%	0.0%
14	81.8%	18.2%		9	100.0%	0.0%
20	81.8%	18.2%		10	100.0%	0.0%
1	72.7%	27.3%		12	100.0%	0.0%
3	72.7%	27.3%		19	100.0%	0.0%
4	72.7%	27.3%		21	100.0%	0.0%
12	72.7%	27.3%		1	90.9%	9.1%
24	72.7%	27.3%		4	90.9%	9.1%
1	63.6%	36.4%		11	90.9%	9.1%
10	63.6%	36.4%		14	90.9%	9.1%
15	63.6%	36.4%		14	90.9%	9.1%
2	54.5%	45.5%		24	90.9%	9.1%
9	54.5%	45.5%		15	81.8%	18.2%
11	54.5%	45.5%		17	81.8%	18.2%
21	54.5%	45.5%		20	81.8%	18.2%
22	54.5%	45.5%		23	81.8%	18.2%
23	54.5%	45.5%		22	54.5%	45.5%

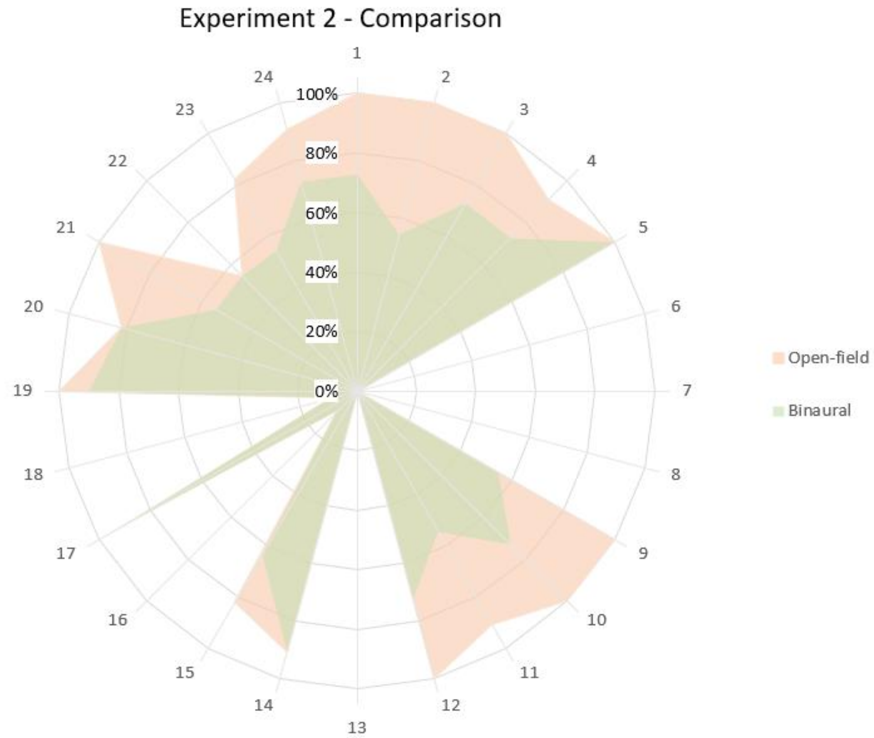


Figure 43 Comparison of results in headphone vs open-field tests for any degree of error in Experiment 2

The results shown in Table 32 and Figure 43 illustrate some similarities and patterns between accuracies of locations within a range of 60° azimuth (30° either side of the correct location), however owing to the low population size difficult locations are still hard to distinguish, soh, the number of subjects participating in the experiment should be carefully considered.

7.1.3 Crossover and Summary of Experiments 1 and 2

Both experiments, and stimuli, performed differently under the same testing conditions, with some exceptions, to the order of locations, stimuli and subject population sizes. The locations, not considering the lower population size in Experiment 2, do however, show some overlapping areas. Table 33 presents the locations used in both binaural experiments, along with their respective scores for any degree of accuracy (within 30° azimuth left or right).

Table 33 Differences in results of shared locations in Stimulus 1 and Stimulus 2

Location	Experiment/Stimulus 1	Experiment/Stimulus 2
1 (twice in Experiment 2)	85.3%	63.6%, 72.7%
2	73.5%	55.5%
4	76.5%	72.7%
5 (twice in Experiment 1)	85.3%, 73.5%	100.0%
9	55.9%	55.5%
10	47.1%	63.6%
11	61.8%	55.5%
12	55.9%	72.7%
17	50.0%	100.0%
19	76.5%	90.9%
20	61.8%	81.8%
23	55.9%	55.5%
24	64.7%	72.7%

Even with the change of stimulus between experiments, the results show the significant patterns of localisation abilities and favourable directions of arrival. Some of the highest scoring locations in both experiments were located directly in-front of the subjects (23, 24, 1, 2) along with locations due left and right (18, 19 and 6, 7) owing to high levels of HRTFs present, with the frontal-hemisphere locations being more accurate (19 and 6 respectively) and thus supporting the literature and theory reported previously in this thesis.

7.1.4 Point system, Simulation and Validation Review

The point system devised within this work defined a qualitative assessment by enabling a quantitative analysis of a binaural system to an overall result of each subject, as well as a percentage of subjects which were able to pass the test using the validation procedure. The raw data can be analysed by the manufacturer of the system in order to improve and locate the problematic areas, and furthermore, a database of results from the validation test can provide the foundation for further investigation on human hearing localisation abilities. The degree of accuracy, or acceptable error, within the point system can be modified to adapt to the requirement of a system. The work presented in this thesis is primarily intended for the use of binaural systems in entertainment and media, however, a system demanding a higher accuracy (i.e. mission-critical robotics and artificial intelligence) can take a stricter approach to not only the degree of acceptable accuracy, but also the overall validation pass threshold.

The results of the simulation work determine the average overall result of a subject, assuming the subject were to attempt the test randomly and independent of the perceived DOA. Furthermore, the results determine the likelihood of a high-average guesswork set of results within a certain population size. Chapter 3 suggested the hypothetical average overall result of guesswork lies between 12% and 13%. The results from the simulation define the probability and average overall result between 12.4% and 12.6% for the first four datasets, with a standard deviation of 5.9-6.1, depending on the order of locations used while the double-randomisation (dataset 5) procedure scoring as low as 11.95%. Chapter 3 also suggested, and recommended, a pass threshold of 45% on the binaural and validation tests to allow for headroom, for the probability of a higher overall score of guesswork and additionally to eliminate/exclude the subjects which may have guessed some locations and approximated the 'easier' locations. The pass rate of experiment 1 suggests that a threshold of 45% could be harsh owing to the large percentage of eliminated subjects, however, it is speculated that the average overall results are lower in experiment 1 owing to the stimulus rather than the pass mark.

7.2 Subjects and Subject Population Size

Throughout the experiments, subjects expressed no discomfort with the procedure of the experiment, nor any misunderstanding of the objective of the experiment. The stimuli, both through the headphones and loudspeakers, were at audible and comfortable sound pressure levels for the subjects and no amount of ear fatigue is speculated owing to the short exposure time of the experiment. All 8 subjects who participated in both experiments verbally expressed a firm belief that Stimulus 2 was 'easier' to locate, and results seem to support these claims. This suggests that pink noise is a more appropriate selection of stimuli than the former, single-click stimulus. Furthermore, the increase of efficiency in localisation abilities could be owing to subjects becoming adjusted to the test, and therefore imply the possibility of training subjects on how to localise the direction of arrival of a sounds' source. Further investigation is required to determine whether either, or both, of these are the basis of an increase in localisation abilities for the latter stimulus.

The open-field test shows a clear correlation with the headphone test. As such, the procedure of validating subjects is considered effective enough to justify the lack of a hearing test. Subjects with either favourable or biased hearing (i.e. loss of hearing in a particular ear) develop coping mechanisms to localise a sounds' source whilst subjects with more severe hearing impairments result below the pass threshold of validation and thus do not contribute to the final qualitative score of a binaural system. In the case that results across many subjects show similar patterns in efficacy for a specific area of locations, further investigation is required to determine the source of the biasing.

Experiment 2 showed problematic percentage results owing to the low population size, and it is therefore suggested that an experiment should be conducted on a minimum of 20 subjects to allow for certain anomalies (i.e. hearing impairments). Given the correct experimental conditions, it is expected that approximately 70% of subjects will succeed in passing the headphone test, based on the validation procedure, allowing for a dataset of 14-15 validated subjects. Ultimately, a higher population size is recommended, however, owing to the time-consuming procedure of experimentation, the author suggests a population size of 20 initial subjects is adequate, given the additional validation procedures.

7.3 Superimposing Factors and Challenges

Acoustics, psychoacoustics and psychology influence the way humans perceive the location of a sound. In order to create an unbiased testing procedure, many of these factors, or effects, need to be considered and counteracted. This proves difficult whilst still attempting to maintain a realistically reproducible testing environment for manufacturers in terms of time and monetary requirement.

7.3.1 Front-and-Back Confusion

Front-and-back confusion is a common challenge in the reproduction of immersive, namely binaural, audio. The accuracy of the front and rear locations provides an idea of the efficiency of a binaural system and furthermore provides an insight into the possible intended use of the system under test. The results of the binaural dummy-head used in this work are investigated for locations 23, 24, 1, 2 (front) and 11, 12, 13, 14 (rear). Each location, where possible (based on the order of locations), is shown with its corresponding incorrect (0 points) estimations along with the number of subjects which estimated the location as one of the locations on the opposing side. Table 34 and Table 35 allow for comparison of front-and-back confusion.

Table 34 Results of confusion between front and rear locations for Experiment 1 (binaural and open-field)

Experiment 1 - Headphone		
Location	**Incorrect**	**% of Incorrect opposites**
2	26.5%	33.3% (3/9)
1	14.7%	20.0% (1/5)
24	35.3%	58.3% (7/12)
23	44.1%	6.7% (1/15)
11	38.2%	46.2% (6/13)
12	44.1%	53.3% (8/15)
13	41.2%	57.1% (8/14)
Experiment 1 – Open-field		
Location	**Incorrect**	**% of incorrect opposites**
2	5.9%	50.0% (1/2)
1	5.9%	50.0% (1/2)
24	14.7%	80.0% (4/5)
23	14.7%	80.0% (4/5)
11	44.1%	33.3% (5/15)
12	32.4%	27.3% (3/11)
13	35.3%	58.3% (7/12)

Table 35 Results of confusion between front and rear locations for Experiment 2 (binaural and open-field)

Experiment 2 - Headphone		
Location	**Incorrect**	**% of Incorrect opposites**
2	45.5%	40% (2/5)
1, 1	36.4%, 27.3%	75% (3/4), 66.7% (2/3)
24	27.3%	66.7% (2/3)
23	45.5%	20% (1/5)
11	45.5%	80% (4/5)
12	27.3%	66.7% (2/3)
14, 14	18.2%, 9.1%	0.0% (0/2), 0.0% (0/1)
Experiment 2 – Open-field		
Location	**Incorrect**	**% of incorrect opposites**
2	0.0%	0.0% (0/0)
1, 1	0.0%, 0.0%	0.0% (0/0), 0.0% (0/0)
24	9.1%	100.0% (1/1)
23	18.2%	50.0% (1/2)
11	9.1%	0.0% (0/1)
12	0.0%	0.0% (0/0)
14, 14	9.1%, 9.1%	100.0% (1/1), 0% (0/1)

Observations of Experiment 1, the more trustworthy of the two experiments in terms of amount of data gathered, show relatively large percentages of incorrectly estimated locations which were estimated to be approximately opposite of the correct location (i.e. subject guessed/perceived the location as number 12 where correct answer was 1). In most of the examples, more than half of the incorrect locations were perceived as almost exactly opposite, suggesting that front-and-back confusion is moderately present, with increasing levels and difficulty distinguishing between front and back during the headphone test as opposed to the free-field, natural hearing environment.

7.3.2 Inside-the-head 'Locatedness' and Characterising & Personalising HRTFs

Observing and tracking inside-the-head locatedness (IHL) is a relatively difficult task when attempting to localise the source of a sound. As described in Chapter 2, this occurs in the reproduction of binaural audio through headphones, owing to the mismatch of a users' HRTFs against the HRTFs of the binaural system. This phenomenon is difficult to quantify, or describe, and the only way of observing such results is through the communication of subjects conducting the experiment. As such, potential cases of IHL were not observed given it reflects a singular subject's localisation ability rather than the overall efficiency and broad application of the binaural system.

The work presented in this thesis focuses primarily on the broad delivery of binaural audio to a human audience for the purpose of entertainment and media. Therefore, individualised results were not observed, except for reduction in localisation accuracy between headphone and open-field tests. A binaural system aiming to achieve a greater accuracy, to suit the systems' intended application such as binaural hearing aids, should consider not only using a stricter grading and point system as aforementioned, but also employ characterisation of HRTFs, thus greatly reducing the occurrence of IHL. The procedure of characterising and personalising HRTFs is a further time-consuming and tedious process and as such is not considered necessary for entertainment systems.

7.3.3 Haas Effect

The precedence effect is minimally present when conducting experiments in anechoic or near-anechoic conditions. Given that the experiments were solely conducted in an isolation booth with minimal reflections and external sound sources, and that binaural audio is typically captured in a similar studio environment, it is expected that the Haas effect will not impact such experiments or testing. Binaural audio captured in untreated environments (i.e. field recordings) requires further investigation and observation.

7.4 Summary

This chapter discussed and analysed some further findings and results in correlation with the initial theory presented in Chapter 3. Furthermore, this chapter considered the results and investigated which components of the experimental procedure were considered successful and/or necessary. This chapter also proposed certain factors required towards proposing a standardised testing environment for binaural systems as per the aim and objectives in Chapter 1. Chapter 8 will provide the final conclusions drawn from all the observations and findings conducted throughout the experimentation procedure.

Chapter 8: Conclusions

Chapter 1 set an aim and a series of objectives for the work presented in this thesis, furthered by, and based on, the current literature and work undertaken by other researchers in the subject of binaural audio which is presented in Chapter 2. Chapters 3, 4 and 5 put forward a theory and an experimentational methodology for undertaking the investigation into the required parameters for proposing a standardised testing environment of binaural systems. Chapter 6 outlined all the results produced from said experiments and their procedure. Chapter 7 observed and further analysed results in detail against the speculated theory in Chapter 3.

This chapter summarises the fundamental results from Chapter 7. Furthermore, this chapter proposes several recommendations for further work in standardising the testing environment of any future binaural audio systems. Lastly, this chapter presents the final conclusions of the work presented in this thesis.

8.1 Main Contributions

The underlining literature behind binaural systems and the testing procedures, or measurements of efficacy, of binaural systems has been investigated. With little-to-no pre-existing standards, or in-depth testing methodology, the research to determine whether such a standardised testing procedure is possible, has been undertaken.

The work presented in this thesis has determined the viability of creating a standardised testing procedure and environment for binaural systems as well as the viability of creating such a standardised testing regime (Objective 4). These binaural systems are evaluated through a subjective, qualitative, measurement and are assigned a percentage figure of overall efficiency in terms of human localisation abilities through the system in question. The experiments, conducted on the Binaural Enthusiast dummy-head, show positive levels of response rates, using the methodology provided in Chapter 5. These results indicate certain levels of consistency of localisation in subjects, with variations owing to the difference in HRTFs, localisation and/or hearing abilities, and stimulus selection.

Many validation techniques, both statistical and physical, have been developed and employed to ensure the removal of biasing datasets or subjects (Objective 5). A method of validating subjects which may not be able to complete the test successfully, owing to hearing impairments or other factors, has been demonstrated. The likelihood, and potential, for guesswork has been investigated and determined, with simulations showing the majority of scores between 6-18%. This provides and suggests a benchmark, a pass criterion, for which subjects need to achieve to contribute to the binaural systems' efficiency. The theory, in conjecture with the results, suggests the pass threshold to be approximately 45% to allow for additional headroom in high-scoring guesswork or a combination of guesswork and localisation abilities, particularly for those locations with high levels of HRTFs (easier locations to estimate).

Experiment 1 showed a lower pass-rate of subjects at 41.18%, compared to that of the latter, Experiment 2 at 72.72%. This could be owing to the stimulus, order of locations or population size of subjects. Further investigation is required to determine the exact cause. Both experiments showed similar results of overall scores, particularly in results of the same individual location. Additionally, the results of the open-field test were also observed for consistency in locations, both between its respective headphone (binaural) test, and its counterpart in Experiment 2. Once again, results indicate certain levels of consistency, given some decrease in efficiency between the open-field (validation) and headphone test. This states that the system in question has significant HRTFs present and can reproduce binaural audio to an efficiency of 56.9% and 61.9% for Stimulus 1 and Stimulus 2 respectively. The consistency of results implies that irrespective of the two stimuli, subjects (humans) have certain favourable locations and directions-of-arrival. As such, a definitive conclusion can be drawn that a standardised testing procedure for evaluating the efficiency of any binaural system is possible (Objective 1). Further investigation into subject population size is recommended, however, subject sizes of 30-35 are deemed appropriate so long as the pass-rate is no lower than 40%.

Many of the external, negatively influencing, effects have been investigated and some countermeasures have been deployed (Objectives 2 and 3). The use of near-anechoic, or sound-proofed isolation chambers should be sufficient to counteracting reflections, or more specifically, reverberation. This is under the assumption of resting dB sound pressure levels of below 35 dB, and a reverberation time of 0.1 to 0.5 ms. The occurrence of the natural phenomena referred to as inside-the-head 'locatedness' is not considered, and further work is required to determine the full possibility or likelihood of this occurring. This IHL is generally a result of mismatching HRTFs relative to the system and audio delivery and customisation of HRTFs would have to be introduced. It is likely that the average success rate would increase significantly and could therefore undergo stricter testing conditions to adapt to a systems' requirements.

8.2 Recommendations for Further Work

Given the complex nature of human localisation abilities and the reproduction of perceived HRTF cues, binaural audio, and therefore the testing of binaural systems, has much room for improvement. The work presented in this thesis attempts to propose factors towards creating a standardised testing environment and is not able to scientifically determine many other causes or phenomena that occur during the reproduction of binaural audio. The following paragraphs are the authors proposed fields of work that require further research, in order to create, or improve, a blueprint for such a standardised testing environment of binaural systems. Specifically, the (i) HRTF Personalisation, (ii) further investigation into external effects (physical and psychological) and (iii) creating the standardisation of binaural systems.

The mismatch in HRTFs between a subject and the binaural system produces lower levels of accuracy and therefore a personalisation method is required to improve accuracy. For a binaural system that aims to achieve higher accuracy demands the application of personalisation and customisation of HRTFs. This could be achieved through a process whereby the audio is captured through different physical dimensions of the dummy-head (i.e. adjustable width of the distance between the pinnae) and the subjects would have the option to choose the most suitable, or closest, to their respective pinna-to-pinna width. Other possibilities include signal processing and filtering of cues present in the audio, or a calibration procedure which adjusts signals to the perceived direction of arrival (DOA) of a subject. Every human has their own unique set of HRTFs and therefore their own perception of a sounds' DOA. The personalisation of localisation cues is therefore crucial in achieving near-perfect recreation of localisation abilities.

The work presented in this thesis considered some of the external effects on testing regimes, particularly those with negative impact. A deeper understanding of these effects would allow for a more rigorous and trustworthy testing environment. As such, the psychoacoustic and psychological effects which require further work are condensed for the readers convenience:

(i) Inside-the-head 'locatedness' (IHL), the internalisation of audio, appears to be a by-product of binaural audio reproduction and methods of eliminating IHL are desired. Implementation of such IHL-elimination is required to be included in the testing environment.

(ii) Front-to-back confusion is the inability to distinguish the locations from the front to back, and vice-versa. This occurrence is generally countered through the physical repositioning or interaction with the environment, however, given that this is not possible with an audio-only environment, other cues would have to be developed.

The research, and experiments, presented in this thesis merely suggest the possibility and viability of a standardised testing environment for binaural system. A testing regime is to be developed to be able to quantify the quality of a system and compare it to other industry systems, and therefore ultimately to be able to define an intended application, or use, of the binaural system.

The criterium for proposing towards the standardisation of a testing environment presented in this thesis provides a foundation and blueprint for binaural systems intended in the entertainment sector (i.e. virtual reality). A more rigorous testing procedure is required for binaural systems demanding higher accuracy (e.g. hearing apparatus).

8.3 Summary

Designing a blueprint for evaluating binaural systems is a time-consuming and strenuous procedure owing to the delicate nature of human hearing, and thus localisation abilities. Subjective measurements require the participation of subjects in a testing environment and experimentation, whilst objective measurements only determine the levels of HRTFs present and merely suggest whether a listener is hypothetically capable of approximating the location of the sound. Furthermore, there are many external effects, both psychological and acoustic, which need to be taken into consideration. The work presented in this thesis investigated many of these factors and attempted to propose some rough guidelines towards developing a standardised testing environment for such binaural systems. Additionally, experiments were conducted which determined the possibility of awarding binaural systems with a qualitative measurement.

To summarise the final, individual, contributions based on the initial requirements presented at the end of Chapter 2, the following is reviewed:

(i) Qualitative measurement system is considered successful and an overall percentage score is attributed to each binaural system. This is defined by the average overall score of test subjects who are proficient at localising the direction of arrival of a given stimulus through a process of validating subjects.

(ii) Two stimuli are created and experimented on. The overall observation is that localisation abilities vary to a smaller degree, however problematic locations are prevalent and consistent between either stimulus.

(iii) A procedure for generating the order of locations has been developed. The results of the simulation experiments show a low likelihood of achieving a high overall score on the experiment through guesswork. More locations and/or samples are possible, given a longer testing procedure.

(iv) The location and procedure of the experimentation process is considered necessary to avoid various acoustical interference (i.e. reverberation). It is expected that further improvements can be made given the correct provisions such as in-ear phones, etc.

(v) The scoring system developed shows a range of possible scores, with the validation (open-field) test excluding subjects with insufficient localisation abilities. The point-based system is considered adequate for binaural systems intended for the capture and reproduction of 3D film and audio, immersive music, virtual reality, etc. A more rigorous point system is required for more demanding systems, and the current system of measurement provides a blueprint and baseline for such a system.

(vi) Experiment 1 showed a large enough population size of subjects whilst Experiment 2 showed difficulties in reliable results. As such, the work of this thesis suggests an acceptable number of subjects, or participants, of 30. Subject population sizes below 25 may struggle to achieve dependable and viable results.

References

Abouchacra K. S., Breitenbach J., Mermagen T., Letowski T. (2001), Binaural Helmet: Improving Speech Recognition in Noise with Spatialized Sound, *The Journal of Human Factors and Ergonomics Society*, Vol. 43 (3), pp. 584-594

Alberti P. W. (2006), The Anatomy and Physiology of the Ear and Hearing, *University of Toronto Press*

Algazi V. R., Duda R. O., Morrisson R. P., Thompson D. M. (2001), Structural composition and decomposition of HRTFs, *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 103-106

Batteau D. W. (1967), The role of the pinna in human localization, *Proceedings of the Royal Society of London, Series B: Biological Sciences*, Vol. 168, pp. 158-180

Beranek L. L. and Mellow T. (2012), *Acoustics: Sound Fields and Transducers*, Elsevier Science and Technology

Bertho-Lavenir, C. (1989), Innovation Technique Et Société Du Spectacle: Le Théâtrophone à L'Exposition De 1889, *Le Mouvement Social*, Vol. 149, pp. 59-69 [In French]

Birkinshaw D. C. (1968), The work of Alan Blumlein, *British Kinematograph, Sound & Television Society*, Vol. 50 (7), pp. 206-219

Blauert J. (1969), Sound Localization in the Median Plane, *Acta Acustica united with Acustica*, Vol. 22 (4), pp. 205-213

Blauert J. (1983), *Spatial Hearing – The Psychophysics of human sound localization*, The MIT Press, Cambridge, Massachusetts pp. 203-237

Blauert J., Bruggen M., Hartung K., Bronkhorst A., Drullman R., Reynaud G., Pellieux L., Krebber W., Sottek R. (1998), The AUDIS Catalog of Human HRTFs, *16th International Congress on Acoustics and 135th Meeting Acoustical Society of America*, pp. 2901-2902

Blauert J. (2013), *The Technology of Binaural Listening*, Springer

Boren B. and Roginska A. (2011), The Effects of Headphones on Listener HRTF Preference, *Audio Engineering Society Convention 131 (AES)*

Boren B., Geronazzo M., Majdak P., Choueiri E. (2014), PHOnA: a public dataset of measured headphone transfer functions, *Audio Engineering Society Convention 137 (AES)*, pp. 834-840

Brittain F. H. (1951), A non-reflection room and its uses for Acoustical Measurement, *British Kinematography*, Vol. 19 (5), pp. 148-152

Brungart D.S and Rabinowitz W. M. (1999), Auditory localization of nearby sources. Head-related transfer functions, *The Journal Of The Acoustical Society of America*, Vol. 106, pp. 1465

Canals C. and Canals A. (2019), When is n Large Enough? Looking for the Right Sample Size to Estimate Proportions, *Journal of Statistical Computation & Simulation*, Vol. 89 (10), pp. 1887-1899

Chen Z. W., Yu G. Z., Xie B., Guan S. Q. (2012), Calculation and Analysis of Near-field Head-Related Transfer Functions from a Simplified Head-Neck-Torso Model, *Chinese Physics Letters*, Vol. 29 (3), pp. 034302-1-034302-4

Choueiri E. Y. (2011), Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers, *Princeton University*

Dobosz P. (2019), *Binaural Enthusiast B1-E Dummy Head with BE-1P Binaural Microphones and Battery Box*, Binaural Enthusiast [Available at: https://binauralenthusiast.com/product/etiam-ullamcorper-dollor-3/] [Accessed on: 05-12-2019]

Durlach N. I. and Colburn, H. S. (1978), Binaural Phenomena, In Carterette E. C. and Friedman M. P. *'Handbook of Perception'*, New York Academic Press, pp. 365-466

Ebata M., Sone T., Tadamoto N. (1968), Improvement of Hearing Ability by Directional Information, *The Journal of the Acoustical Society of America*, Vol. 43 (2), pp. 289-297

Elliott S., House C., Cheer J., Simon-Galvez M. (2016), Cross-talk Cancellation for Headrest Sound Reproduction, *Audio Engineering Society*, (Conference Paper) [Presented at: The Conference on Sound Field Control, July 18-20, 2016, Guildford, United Kingdom], Paper no. 4-2

Everest F. A. and Pohlmann K. C. (2015), *Master Handbook of Acoustics*, 6th Edition, McGraw-Hill Education

Feddersen W. E., Sandel T. T., Teas D. C. and Jeffers L. A. (1957), Localization of High-Frequency Tones, *The Journal of the Acoustical Society of America*, vol. 29, pp. 988

Floyd E. T. (2008), *Sound reproduction loudspeaker and rooms*, Focal Press

Gantt M. (2017), Binaural Beat Technology: A Complimentary Path to Post Deployment Wellness, *Bethesda Marlyand TriServicing Nursing Research Program*

Gardner M.B. and Gardner R. S. (1973), Problem of localization in the median plane: effect of pinnae cavity occlusion, *The Journal of the Acoustical Society of America*, Vol. 53 (2), pp. 400-408

Gardner W. and Martin K. (1995), HRTF Measurements of KEMAR, *The Journal of the Acoustical Society of America*, Vol. 97 (3), pp. 907-908

Gelfand S. A. (1981), *Hearing: An introduction to Psychological and Physiological Acoustics*, Marcel Dekker Inc.: Butterworths

Gelfand S. A. (2010), *Hearing: An introduction to Psychological and Physiological Acoustics*, 5th Edition, Informa Healthcare

Grana G. D., Huston K. A., Badea A., Pappa A., Scott W., Fitzpatrick D.C. (2017), The organization of frequency and binaural cues in the gerbil inferior colliculus, *The Journal of Comparative Neurology*, Vol. 525 (9), pp. 2050-2074

Griesinger D. (2001), Surround: The Current Technological Situation, *British Kinematograph, Sound & Television Society*, Image Technology, Vol. 83 (2), pp. 22-31

Grothe B., Pecka M., Mcalpine D. (2010), Mechanics of Sound Localization in Mammals, *Physiological Reviews*, Vol. 90 (3), pp. 983-1012

Gumerov N. A., Duraiswami R., Tang Z. (2002), Numerical study of the influence of the torso on the HRTF, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 1965-1968

Hao, M., Zhou L., Hongmei H., Zhenyang W. (2007), A Novel Sound Localization Model Based on Head Related Transfer Function, 2007 *8th International Conference on Electronic Measurement and Instruments*, pp. 428-432

Haas H. (1949, 1972), "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache", translated into English in 1972, The Influence of a Single Echo on the Audibility of Speech, *Journal of the Audio Engineering Society*, Vol. 20 (1972), pp. 145-159

Hanson R. L. and Kock W. E. (1957), Interesting Effect Produced by Two Loudspeakers under Free Space Environment, *The Journal of the Acoustical Society of America*, Vol. 29 (5), pp. 655-663

Hartmann W. M. (1999), How we localize sound, *Physics Today*, Vol. 52 (11), pp. 24-29

Hebrank J. and Wright D. (1974), Spectral Cues used in Localization of Sound Sources on the Median Plane, *The Journal of the Acoustical Society of America*, Vol. 56 (6), pp. 1829-1834

Hiipakka M., Kinnari T., Pulkki V. (2012), Estimated head-related transfer functions of human subjects from pressure-velocity microphones, *The Journal of the Acoustical Society of America*, Vol. 131 (5), pp. 4051-4061

Hood J. D. (1968), Observations Upon the Relationship of Loudness Discomfort Level and Auditory Fatigue to Sound-Pressure-Level and Sensation Level, *The Journal of the Acoustical Society of America*, Vol. 44, pp. 959-964

Howard D. M. and Angus J. (2009), *Acoustics and Psychoacoustics*, 4th Edition, Focal Press, pp. 109-113

Howard D. M. and Angus J. (2017), *Acoustics and Psychoacoustics*, 5th Edition, Focal Press, pp. 108-122

Iida K., Motokuni I., Itagaki A., Morimoto M. (2007), Median plane localization using a parametric model of the head-related transfer function based on spectral cues, *The Journal of Applied Acoustics*, Vol. 68, pp. 835-850

ITU (2012), Multichannel stereophonic sound system with and without accompanying picture, *International Telecommuncation Union*¸ Recommendation ITU-R BS. 775-3 (08/2012)

Jiang J., Xie. B., Mai H., Liu L., Yi. K., Zhang C. (2018), The Effects of Dynamic Cue and Spectral Cue on Auditory Vertical Localization, *Audio Engineering Society* (Conference paper), [presented at: International Conference on Spatial Reproduction – Aesthetics and Science, August 7-9, 2018, Tokyo, Japan]

Keyrouz F. (2014), Advanced Binaural Sound Localization in 3-D for Humanoid Robots, *IEEE Transactions on Instrumentation and Measurement*, Vol. 63 (9), pp. 2098-2107

Kim S. (2018), Auditory Localization training using generalized head-related transfer functions in augmented reality, *Acoustical Science and Technology*, Vol. 39 (4), pp. 312-315

Klein F. and Werner S. (2016), Auditory Adaptation to Non-Individual HRTF Cues in Binaural Audio Reproduction, *Journal of the Audio Engineering Society*, Vol. 64 (1), pp. 45-54

Kokkinakis K. (2018), Binaural Speech Understanding with Bilteral Cochlear Implants in Reverberation, *American Journal of Audiology*, Vol. 27 (1), pp. 85-95

Le Bagousse S., Paquier M., Colomes C. (2011), Sound Quality Evaluation Based on Attributes – Application to Binaural Contents, *Audio Engineering Society Convention 131 (AES)*

Letowski T. (1989), Sound Quality Assessment: Concept and Criteria, *Audio Engineering Society Convention 87 (AES)*

Lewald J., Dorrscheidt G. J., Ehrenstein W. H. (2000), Sound localization with eccentric head position, *Behavioural Brain Research*, Vol. 108 (2), pp. 105-125

Likert R. (1932), Technique for the Measurement of Attitudes, *Archives of Psychology*, Vol. 22, pp. 140-195

Lim J. H., Kim H., Jeon C., Cho S. (2018), The effects on mental fatigue and the cognitive function of mechanical massage and binaural beats (brain massage) provided by massage chairs, *Complementary Therapies in Clinical Practice*, Churchill Livingstone Inc., Vol. 32, pp. 32-39

Lindau A. and Brinkmann F. (2012), Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings, *Journal of the Audio Engineering Society*, Vol. 60 (1), pp. 54-62

Majdak P., Balazs P., Laback B. (2007), Multiple exponential sweep method for fast measurement of head-related transfer functions, *The Journal of the Audio Engineering Society*, Vol. 55, pp. 623-637

Mayfield M. (2016), The Practical Realities of Immersive Audio, *Boxoffice*, Vol. 152 (7), pp. 22-26

Middlebrooks J.C. (1992), Narrow-band sound localization related to external ear acoustics, *The Journal of the Acoustical Society of America*, Vol. 92, pp. 2607-2624

Mills J. (1956), On the Minimum Audible Angle, *The Journal of The Acoustical Society of America*, Vol. 30, pp. 237

Moravec M., Liptai P., Zbojovsky J., Badidova A. (2017), Assessment of the Washing Machines Sound Quality by the Binaural Measurement System, *Technical Sciences and Technologies*, Vol. 1 (7), pp. 161-165

Nicol R., Gros L. Colomes G., Warusfel O., Noisternig M., Bahu H. (2014), A Roadmap for Assessing the Quality of Experience of 3D Audio Binaural Rendering, Proceedings of the EEA Joint Symposium on Auralization and Ambisonics 2014, Universitätsverlag der TU Berlin, pp. 100-106

Nikunen J., Diment A., Virtanen T., Vilermo M. (2016), Binaural rendering of microphone array captures based on source separation, *Speech Communication*, Vol. 76, pp. 157-169

Oldfield S. R. and Parker S. P. (1984), Acuity of sound localisation: a topography of auditory space 1: Normal hearing conditions*,* *Perception*, Vol. 13 (5), pp. 581-600

Orduña-Bustamante F., Padilla-Ortiz A. L., Torres-Gallegos E. A. (2018), Binaural speech intelligibility through personal and non-personal HRTF via headphones, with added artificial noise and reverberation, *Speech Communication*, Vol. 105, pp. 53-61

Pena J. L., Viete S., Funabiki K., Saberi K., Konishi M. (2001), Cochlear and neural delays for coincidence detection in owls, *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, Vol. 21 (23), pp. 9455-9459

Perrott D. R. and Saberi K. (1990), Minimum audible angle thresholds for sources varying in both elevation and azimuth, *The Journal of The Acoustical Society of America*, Vol. 87, p.1728

Plack C.J. (2005), *The Sense of Hearing*, Psychology Press

Rayleigh L. (1907), On our perception of sound direction, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Philos. Mag. 13, pp. 214-232

Rice J.J., May B., Spirou, Young E. (1992), Pinna-based Spectral Cues for Sound Localization in Cat, *Hearing Research*, Vol. 58 (2), pp. 132-152

Roginska A. (2018), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*, Audio Engineering Society (AES), p. 88

Rory W. and Hyunkook L. (2017), The Reduction of Vertical Interchannel crosstalk: The Analysis of Localization Thresholds for Natural Sound Sources, *Applied Sciences*, Vol. 7 (3), p. 278

Ruggieri E. (2016), Visualizing the Central Limit Theorem through Simulation, *PRIMUS*, Vol. 26 (3), pp. 229-241

Rumsey F. (2001), *Spatial Audio*, Focal Press, Oxford, United Kingdom

Rumsey F. (2011), Whose head is it anyway? Optimizing binaural audio, *Journal of the Audio Engineering Society*, Vol. 59 (9), pp. 672-677

Rumsey F. (2016), Headphone Technology: Personalization, Perception, Preference, *Journal of the Audio Engineering Society*, Vol. 64 (11), pp. 940-944

Salvador C. D., Sakamoto S., Trevino J., Suzuki Y. (2017), Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets, *Acoustical Science and Technology*, Vol. 38 (2), pp. 51-62

Santos I., Colella-Santos M. F., Marques do Couto C. (2014), Sound pressure level generated by individual portable sound equipment, *Brazilian Journal of Otorhinolaryngology*, Vol. 80 (1), pp. 41-47

Schwartz A., McDermott J. H. (2012), Spatial cues alone produce inaccurate sound segregation: the effect of interaural time differences, *The Journal of the Acoustical Society of America*, Vol.132, pp. 357-368

Seo J. G. and Jeon J. Y. (2019), Impact of early reflection on real-time estimation of direction of arrival in different building environments, *Building & Environment*, Vol. 147, pp. 434-444

Silman S., Emmer M. B. (2011), *Instrumentation for Audiology and Hearing Science: Theory and Practice*, Plural Publishing

Van Drie, M. (2015), Hearing through the theatrophone, *Sound Effects*, Vol. 5 (1), pp. 73-90

Wahbeh H., Calabrese C., Zwickey H., Zajdel D. (2007), Binaural beat technology in humans: A pilot study to asses neuropsychologic, physiologic and electroencephalographic effects, *The Journal of Alternative and Complementary Medicine*, Vol. 13 (2), pp. 199-206

Wallach H., Newman E. B., Rosenzweig M. R (1949), The Precedence Effect in Sound Localization, *The American Journal of Psychology*, Vol. 62, No. 3, pp. 315-336

Wang D. L., Brown G. J. (2006), *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, Wiley-IEEE Press

Wenzel E. M. (1991), Three-dimensional Virtual Acoustic Displays, *NASA Ames Research Centre Technical Memorandum*

Whitaker J. and Benson B. (2002), *Standard Handbook of Audio Engineering*, The McGraw-Hill Companies, 2nd Edition

Xie B. (2013), *Head-Related Transfer Function and Virtual Auditory Display,* 2nd edition, J. Ross Publishing, p. 22

Xu L., Middlebrooks J. C. (1999), Encoding of sound-source elevation by the spike patterns of cortical neurons, *ProQuest Dissertations Publishing*, (thesis)

Yao S. N. and Chen L. J. (2013), HRTF Adjustments with Audio Quality Assessments, *Archives of Acoustics*, Vol. 38 (1), pp. 55-62

Yu G., Wu R., Liu Y., Xie B. (2018), Near-field head-related transfer-function measurement and database of human subjects, *The Acoustical Society of America*, Vol. 143 (3), pp. 194-198

Zhang W., Samarasinghe P. N., Chen H., Abhayapala T. D. (2017), Surround by Sound: A Review of Spatial Audio Recording and Reproduction, *Applied Sciences*, Vol. 7, pp. 532-551

Zhong X. (2013), *Influence of Microphone Placement on the Measurement of Head-Related Transfer Functions*, Proceedings - 2013 International Conference on Computational and Information Sciences, pp. 1831-1834

Zhou C., Hu R., Tu W., Wang X., Gao L. (2011), Binaural Moving Sound Source Localization by Joint Estimation of ITD and ILD, *Audio Engineering Society*, 130th Convention Paper (8830)

Bibliography

3Dio (2019), *FS Binaural Microphone*, 3Dio, [Available at: https://3diosound.com/] [Accessed: 31 August 2019]

Ambeo 3D Audio (2017), Ambeo VR Mic, Sennheiser electronic GmbH & Co. KG, [Available at: https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic] [Accessed on: 31-08-2019]

Audio Engineering Society (2015), AES69-2015: AES standard file exchange – Spatial acoustic data file format, *Audio Engineering Society*, [Available at: http://www.aes.org/publications/standards/search.cfm?docID=99] [Accessed: 01-12-2019]

Chandler-Crnigoj S. L., Jones K. O., Ellis D. L., Wylie S. R., Investigating Binaural Localization Abilities for Proposing a Standardized Testing Environment for Binaural Systems, *2018 International Conference on Information Technologies (InfoTech-2018)*, IEEE Conference Rec. No. 46116, 20-21 September 2018, St. St. Constantine and Elena, Bulgaria, pp. 1-4 (Available in Appendix D)

DemVox (2019), Liverpool John Moores University, *DEMVOX Sound Isolation Booth DV1560* [Available at: https://en.demvox.com/testimonials-and-pictures/testimonies/17-england/207-liverpool-john-Moores-university] [Accessed on: 05-12-2019]

KEMAR (2015), *Knowles-Electronics Mannequin for Acoustic Research*, GRAS Sound and Vibration, [Available at: http://kemar.us/] [Accessed on: 31-08-2019]

Moravec M., Izarikova G., Liptai P., Badida M. (2018), Development of psychoacoustic model based on the correlation of the subjective and objective sound quality assessment of automatic washing machines, *Journal of Applied Acoustics*, Vol. 140, pp. 178-182

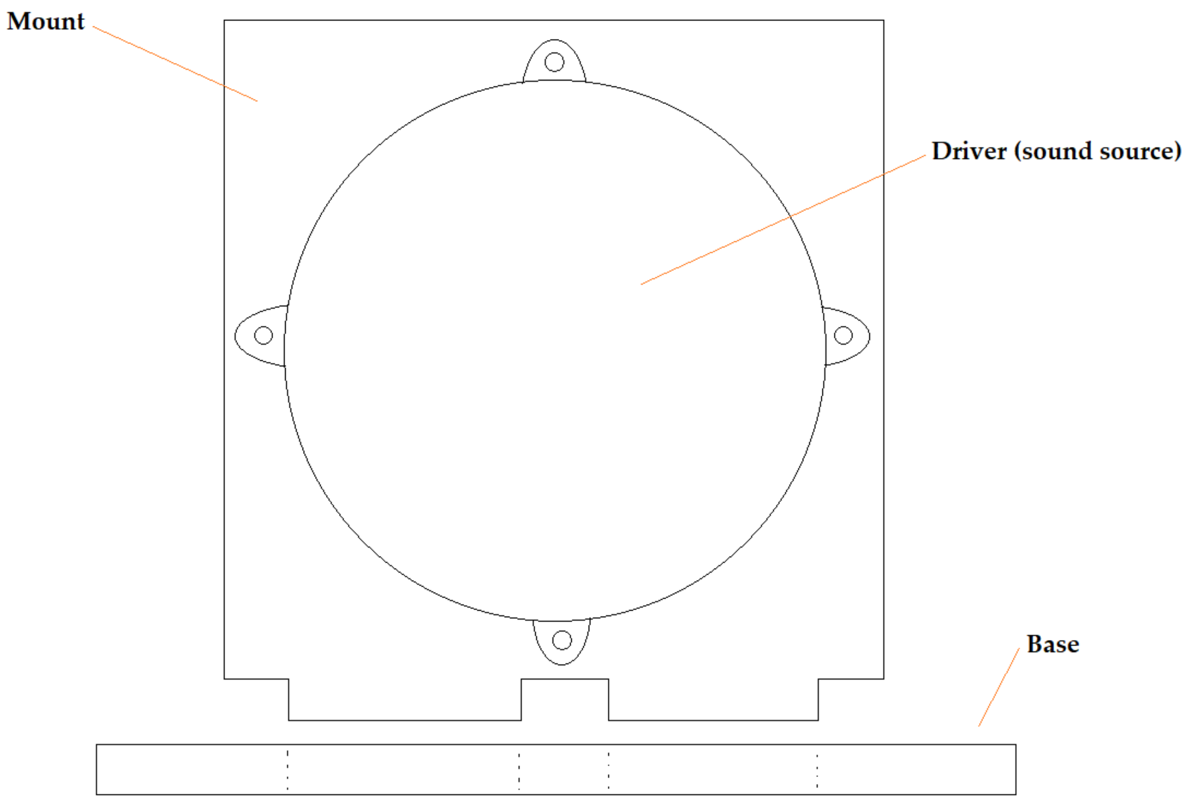
Roland (2017), CS-10EM Binaural Microphones/Earphones, Roland Corporation, [Available at: https://www.roland.com/us/products/cs-10em/] [Accessed on: 31 August 2019]

The Theatrophone (1895), The Theatrophone, *Scientific American*, Vol. 39, pp. 1600-1609

Zoom (2018), H3-VR Handy Recorder, Zoom North America, [Available at: https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder] [Accessed on: 31 August 2019]

Appendix A

Diagram of mounted drivers (top) and image of a mounted driver (bottom)




Speaker location number (#24)
24
Mount
Driver (sound source)
Base
Connections to adjacent room and control panel

Appendix B

Ethical Approval provided by LJMU University Research Committee (UREC)

Dear Sebastian/Karl

With reference to your application for Ethical Approval:

16/EEE/001 – Sebastian Chandler Crnigoj, PGR - Investigation into Binaural Sound Applications (Karl Jones)

The University Research Ethics Committee (UREC) has considered the above application by proportionate review. I am pleased to inform you that ethical approval has been granted and the study can now commence.

Approval is given on the understanding that:

- any adverse reactions/events which take place during the course of the project are reported to the Committee immediately;
- any unforeseen ethical issues arising during the course of the project will be reported to the Committee immediately;
- the LJMU logo is used for all documentation relating to participant recruitment and participation e.g. poster, information sheets, consent forms, questionnaires. The LJMU logo can be accessed at http://www.ljmu.ac.uk/corporatecommunications/60486.htm

Where any substantive amendments are proposed to the protocol or study procedures further ethical approval must be sought.

Applicants should note that where relevant appropriate gatekeeper / management permission must be obtained prior to the study commencing at the study site concerned.

For details on how to report adverse events or request ethical approval of major amendments please refer to the information provided at http://www.ljmu.ac.uk/RGSO/93205.htm

Please note that ethical approval is given for a period of five years from the date granted and therefore the expiry date for this project will be December 2021. An application for extension of approval must be submitted if the project continues after this date.

LIVERPOOL JOHN MOORES UNIVERSITY

Mandy Williams, Research Support Officer
(Research Ethics and Governance)
Research and Innovation Services
Kingsway House, Hatton Garden, Liverpool L3 2AJ
t: 01519046467 e: a.f.williams@ljmu.ac.uk

Appendix C

Instructions for experimental procedure – Open-field and Headphone Tests

Observer (to participant): Welcome. Please position yourself within the circular loudspeaker array, on the provided seating area, facing forward, towards 0° between location 1 and 24 and pay attention to the following instructions.

You will listen to a series of excitations from any of the possible 24 loudspeakers around you, as numbered above each driver. After each trigger of any given sound, please communicate the perceived result, numbered location, through the talk-back microphone provided to your right. You may turn to identify the numbered driver and you may take your time with the response of your perceived location. After you have communicated your perceived location (result), please return yourself to the 0° orientation, once again, facing forward. The observer will note the result and continue with the procedure for a given number of samples and you will be informed when the procedure is over. The procedure is expected to take approximately 10-20 minutes.

*Applicable for binaural test only: Please put the headphones on [*provided by the observer*], with the cabled headphone, marked with an 'L', on your left ear.

To familiarise you with the procedure and to get you accustomed to the stimulus, four locations will be demonstrated first. These locations are - 1, - 7, - 13 and - 19. [- *Triggered stimulus between each demonstration location*]. Please re-position yourself once again, and signal when you are ready to begin with the procedure.

[*Experimental procedure*]

You will be assigned an individual unique identification number to allow your results to be paired with your headphone/open-field results from this experiments counterpart. These are to ensure your anonymity with regards to general data-protection. The observer asks that you do not communicate your perceived results with other potential participants. Thank you for participating in the experiment.

Appendix D

2018 International Conference on Information Technologies (InfoTech-2018), IEEE Conference Rec. No. 46116
20-21 September 2018, St. St. Constantine and Elena, Bulgaria

INVESTIGATING BINAURAL LOCALIZATION ABILITIES FOR PROPOSING A STANDARDIZED TESTING ENVIRONMENT FOR BINAURAL SYSTEMS

1st Sebastian L. Chandler-Crnigoj
Electronics and Electrical Engineering
Liverpool John Moores University
Liverpool, United Kingdom
s.l.chandlercrnigoj@2013.ljmu.ac.uk

2nd Karl O. Jones
Electronics and Eletrical Engineering
Liverpool John Moores University
line 4: Liverpool, United Kingdom
k.o.jones@ljmu.ac.uk

3rd David L. Ellis
Electronics and Electrical Engineering
Liverpoool John Moores University
Liverpool,. United Kingdom
d.l.ellis@ljmu.ac.uk

4th Stephen R. Wylie
Built Environment
Liverpool John Moores University
Liverpool, United Kingdom
s.r.wylie@ljmu.ac.uk

***Abstract*— Binaural sound systems are a growing industry in the upcoming age of three-dimensional (3-D) technology. While many commercial and home systems are entering the market, there is no clear method of determining their suitability for different applications, such as gaming, movies and so on. Thus, a standardized methodology for testing such systems is proposed which evaluates and compares new and existing binaural microphone array systems. The implicating factors which determine the location of a sound, and methods of capturing such sounds, have been identified. A testing and comparison methodology is proposed based on data collected. The proposed methodology provides quantitative and qualitative comparisons to determine the function and suggested application of any given binaural sound system.**

Keywords*— *Audio technology, microphone arrays, psychoacoustics, binaural, sound localization.

Introduction

The ability to localize a sound's source in space is the fundamental characteristic in creating a perception of three-dimensional audio. The increasing demand for hyper-realistic technology that can capture, or simulate, such environments calls for a standardised procedure of testing and validating such systems. There are no current set standards for testing binaural systems.

Traditional binaural capture systems predominantly work by recreating the hearing characteristics of humans. This is done through a binaural microphone array which aims to replicate many of the human head-related transfer functions (HRTFs), which can be seen further below. These HRTFs contain spatial characteristics that inform the human brain of a perceived location of a sound, relative to the position of the listener. Binaural arrays often come in the form of a 'dummy-head' which replicates the human head and its reverberation characteristics. This can be seen in the current leading binaural microphone, the Neumann KU-100 [1]. Many audio technology companies seek to improve and design their own binaural systems, with no current unified method of testing, or comparing, such systems to competing products. This paper works towards proposing a standardised testing environment and procedure for testing new and existing binaural systems based on datasets collected in the experiment outlined below.

I. LITERATURE

Binaural hearing is defined as the act of listening with two sensors a short distance apart. The human (or animal) brain response system determines the location of a sound based on the variance of a sound at each ear. For the purpose of this paper, binaural audio has been defined in two categories; (1) The physical properties of human hearing and localization abilities and, (2) psychological response to various stimuli and testing regimes.

Head-related transfer functions are the cues and physical properties of any sound arriving at two sensors (ears), more specifically, they are the brain processors that distinguish the minute differences created between two sensors. These binaural cues are categorized under the following:

a) Loudness/intensity difference between two sensors (ears), commonly known as interaural intensity difference (IID),

b) Time differentiation between two sensors (ears), interaural time difference (ITD),

c) Timbre, the unique frequency of each given familiar sound.

The combination of these three main localization cues, are what create a sense of direction of arrival (DOA) for any given sound. A binaural system capable of accurately reproducing these cues should in theory achieve near-perfect sound localization through recordings for the purpose of immersive or 3D audio.

II. METHODOLOGY

A. Binaural factors

For a binaural microphone array to work appropriately, it must be able to capture sound from a 3D environment and then reproduce it to a human through appropriate headphones. In this work, a human subject's ability to locate a sound's position is first tested, and then a second test is carried out using binaural headphones playing a similar set of sounds. Both test procedures have a great deal a similarity. A comparison of the two sets of results provides an indication of the "quality" of the binaural microphones array in capturing 3D sound, assuming that the human can determine a sound's location.

Firstly, a subject's ability to localize a sound's source is considered. This localization (hearing) ability first needs to be tested in the natural domain. This ensures the credibility of the individual's results following the binaural test. The inter-aural difference varies in any human subject (based on head dimensions, brain response time, etc.), thus demanding certain pre-test subject conditions. These conditions will be evaluated through a pre-test designed to determine the ability of a subjects' localization ability to a certain percentile accuracy. The given accuracy will dictate whether the subjects' results from a binaural recording are reliable, excluding the potential for 'guesswork'. A consistency of results from both tests also determines the accuracy of a binaural system under test.

Owing to the nature of any psychological testing conducted on humans, it is vital to exclude any, and all, circumstances that could negatively bias the testing procedure, or the validity of the data collected. Any form of pattern recognition would implement an advantage to the subjects' estimation of a speakers' location. For example, playing sounds in a cyclic nature around the test subject. Hence, an unsystematic sequence of sounds needs to be utilised. To create a set of random number generated (RNG) locations from which subjects are to locate the DOA, a mathematical function is required. This pseudo randomisation algorithm feature attempts to exclude certain biased patterns and favouritism.

B. Pre-test and stimuli

Experiments were conducted in a DEMVOX sound isolation booth [2]. The participants were asked to position themselves at the centre of a loudspeaker array ring, with a radius of 1 metre. The array of loudspeakers contained 24 identical drivers mounted on laser-cut MDF, where each loudspeaker was a Visaton FR 10 HM [3]. In relation to positions on a circle, the speaker No. 1 was positioned at 7.5° while the participants faced 0° (See Figure 1). This was done to intentionally avoid degrees of 0, 90, 180 and 270, owing to pre-existing literature of sound localization at these regions (i.e. stereo recording) as well as to avoid front-and-back confusion [4]. The loudspeakers were positioned at every 15° azimuth, facing the subject.

The audio stimulus was chosen for its frequency properties relating to the efficiency of human hearing at certain frequency bands [5]. These equal loudness contours depict the optimal sound pressure level (SPL) of hearing at the target stimulus level. This ensured that accuracy of a subjects' hearing was owing to their ability to do so, rather than ability to hear intensity.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

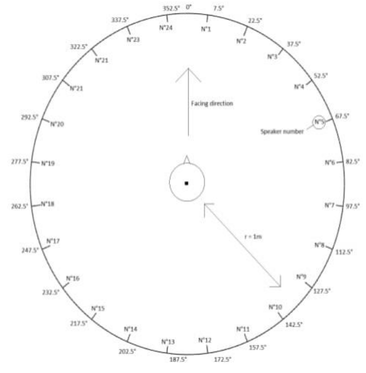


Fig. 1. Position of loudspeakers (Subject faces 0°)

The stimulus, shown in Figure 2, was a single 'click' tone, that was played through one loudspeaker at a time for a total of twenty samples with each sample coming from a different loudspeaker. After each sample, the subjects were asked to identify the direction of arrival, relative to the 24 available locations, starting at 1 (front, 7.5° azimuth).
Sets of RNG locations were created using a template spreadsheet, for various desired findings. A group set of random numbers between 1 and 24 (for each loudspeaker/location) which included the possibility of repeating locations, another set for 1 – 24 without the possibility of repetitions and lastly a set with biased weightings which intentionally focused certain problematic (or favourable) locations (e.g. directly left and right). These were done to further investigate the potential application of a specific binaural system. For example, an application of a system capable of accurately reproducing complex waveforms in the frequency range of 2-5 kHz would be recommended for capturing dialogue (human speech).

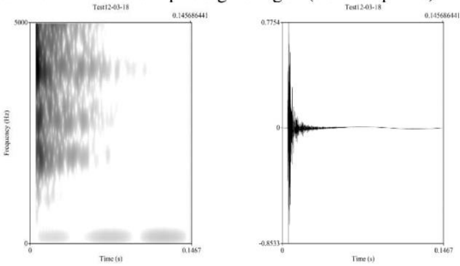


Fig. 2. Waveform (left) and spectrogram (right) of stimulus

The participants were given starting reference points at locations 1, 7, 13 and 19 to familiarize them with the stimulus and the objective as well as the procedure of the test. The results were communicated verbally by the participant to the observer, who noted them independently to prevent subjects from seeing previous answers [6]. This was done to counter the psychological effect of answering multiple choice style questions where the answer was a repetition or pattern (e.g. 3, 3, 3).

The subjects were given a short break between the two tests in an attempt to prevent listening (ear) fatigue and to comply with ethical testing procedures.

B. Binaural Capture and Test

The process seen in section 3.1. was repeated by replacing the subject with the binaural microphone system under test (e.g. 3Dio [7]), with the sound being recorded. The stimulus was played back to participants using headphones [8]. and the participant will once again be asked to attempt to localize the approximate location of the sound.

C. Data Capture, Point System and Analysis

Results for all the testing procedures were communicated to the observer for independent note taking and examination. The data was recorded in a customised Microsoft Excel spreadsheet which compared observed results versus their respective, correct loudspeaker locations. The subjects were given an anonymous identification number to match their natural-hearing test with the binaural system under test. Any further tests on other binaural systems with the same subject eliminated the requirement of the initial experiment and pre-test.

A correct location of a sound awarded the subject with 3 points. Therefore, the total of twenty samples awarded the maximum possible of 60 points. Furthermore, 2 points were awarded for the identification of the sound coming from a loudspeaker immediately adjacent to the 'true' loudspeaker (N°4 and N°6, if sound is coming from loudspeaker location N°5, and finally 1 point was given for the locations two positions away (N°3 and N°7, relative to previous example given). Therefore, an estimation of a loudspeakers' position within 30° azimuths either side of the correct location was still awarded points. Figure 3 gives an illustrative example.

	A	B	C	D	E	F	G	H	I	J	K
1	Localisation experiment										
2											
3	Subject ID:			xxxx			Correct location;			Subject estimation;	
4							#	Pts.		#	Pts.
5	Test type (Pre-/3Dio/etc)			Pre-test			6	3		6	3
6							18	3		18	3
7	Stimulus (tone/bird noise/etc)			Tone			17	3		20	0
8							8	3		5	0
9	Date:			11/04/2018			9	3		4	0
10							5	3		4	2
11	RNG Set:			B			16	3		19	0
12							24	3		24	3
13	Max. pts available (20 x 3)			60			7	3		6	2
14							11	3		9	1
15							20	3		20	3
16							10	3		10	3
17							12	3		12	3
18							19	3		18	2
19	Total points gained;			41			2	3		2	3
20							13	3		12	2
21	Percentile accuracy			68.3333333			4	3		4	3
22							1	3		1	3
23							23	3		22	2
24							5	3		5	3
25						Total;		60			41

Fig. 3. Example of point-based evalutation system

The accuracy of these results reflected a subjects' ability to localize sound as a percentile figure, more specifically, a qualitative dataset. A large enough sample size of subjects meant that an overall efficiency of a binaural system was estimated. This estimation was the average of the results observed during the experiment.

D. Result Analysis

For this section, subjects that met the eligible criteria were selected. Results were observed and compared through individual answers as well as percentile accuracy of the overall score. Results ranged from 48.3%, to 80% (29/60 and 48/60 pts. respectively). The total results observed amounted to a mean of 60.25% and a median of 59.95%.

Problematic locations for binaural systems have been investigated and identified to certain regions. Figure 4 shows collated data from a set of randomly number generated samples with correlating results plotted on the bar chart. The numbers on the left show the order of loudspeaker locations used to play the test sample for a total of twenty samples. The bars represent the percentage of answers which awarded zero points, thus constituting for an answer of at least 30° azimuth incorrect for each respective location. Loudspeaker location five (N°5) was repeated in this particular randomisation of numbers, to investigate a consistency of results from a particular location.

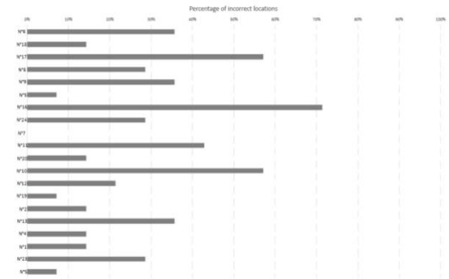


Fig. 4. Common problematic areas (Percentage of answers with a result of a minimum 30 degree azimuth error)

III. Concluding Comments and Further Work

In this paper, a two-step process for evaluating and comparing new and existing binaural microphone systems was proposed. Furthermore, problematic locations relating to the human localization abilities have been identified. An advantage of this procedure is to standardise methods of testing binaural systems in their respective industry.

In addition, we wish to investigate and determine further application for such systems by experimenting with various other stimuli and sample patterns. From the data collected, this would allow the categorisation of systems into various mediums and industries, (e.g. virtual reality). This data would also aid in further understanding the ability of human hearing and localization, as well as its psychological impact.

Acknowledgment

The authors would like to thank John Moores University for providing the funds and necessary equipment for the experimental procedures conducted during the creation of this paper. Furthermore, S. L. Chandler-Crnigoj would like to thank the remaining authors for their support and technical knowledge throughout the course of the papers production, particularly to K. O. Jones for unwavering support and management of this project.

REFERENCES

[1] Georg Neumann GmbH (**2018**). *Dummy Head KU*-100 (*Available at:* https://en-de.neumann.com/ku-100)

[2] DEMVOX™ (**2018**). *Sound Isolation Booth* (*Available at*: en.demvox.com)

[3] Visaton GmbH & Co. KG (**2017**). *Visaton FR HM10 – 8 Ohm*, (*Available at*: heimkino.visaton.de/en/products/fullrange-systems/fr-10-hm-8-ohm)

[4] Fletcher, H. et al (**1933**). Loudness, Its Definition, Measurement and Calculation, *Bell Telephone Laboratories*, The Journal of the Acoustical Society of America, vol. 5, pp. 82

[5] Hofman, P. M. et al. (**2003**). Binaural weighting of pinna cues in human sound localization. *Experimental Brain Research*, Vol. 148, Issue 4, pp. 458-470.

[6] Adrian, F. (**1986**). Response bias, social desirability and dissimulation, *Personality and Individual Differences*, Vol. 7, Issue 3, pp. 385-400

[7] 3Dio (**2018**). *Free Space XLR Binaural Microphone (available at:* https://3diosound.com/products/free-space-xlr-binaural-microphone)

[8] Audio Technica U.S., Inc. (**2018**) *ATH-M30X Professional Monitor Headphones (available at: https://www.audio-technica.com/cms/headphones/f6e3988012a67cd1/index.html)*