Individualisation and reverberation factors in the subjective assessment of plausibility in a binaural auditory display

Thesis · July 2014

DOI: 10.13140/RG.2.1.1876.1202

1 author:

Andrea Genovese
New York University
15 PUBLICATIONS **4** CITATIONS

SEE PROFILE

THE UNIVERSITY OF YORK
DEPARTMENT OF ELECTRONICS

MENG INDIVIDUAL PROJECT REPORT

Individualisation and Reverberation Factors in the Subjective Assessment of Plausibility in a Binaural Auditory Display

---

submitted for the degree of Master in Engineering in:

ELECTRONIC ENGINEERING
WITH MUSIC TECHONOLOGY SYSTEMS

ANDREA FELICE GENOVESE
Y4803772

SUPERVISORS:
MR ANTHONY I. TEW & DR HELENA DAFFERN

May 2014

Individualisation and Reverberation Factors in the Subjective Assessment of Plausibility in a Binaural Auditory Display

ANDREA FELICE GENOVESE

May 2014

UNIVERSITY of York

Abstract

This report forms part of the assessment for the final year of the four-year MEng course in Electronics Engineering with Music Technology Systems at the University of York, UK. This project also stands as a collaboration with the Ph.D. research of Chris Pike from BBC R&D, Media City UK, Salford.

The project portrayed by this report is a research-oriented experiment aimed to investigate the influence of the difference between different reverberation conditions in the subjective assessment of plausibility in a binaural auditory display. The project follows the work of recent experiments that used plausibility as an alternative assessment criteria for binary quality judgement tasks, and used Signal Detection Theory as the appropriate analysis methodology. For this experiment, the recording and processing of individual Head-Related Transfer Functions was introduced as an ulterior factor to compare to previous studies. It was found that the introduction of individualisation brought no real improvements compared to previous studies. Reverberant conditions proved to increase the level of plausibility in comparison to the level reached in anechoic conditions. A number of problems encountered and procedural mistakes decreased the level of confidence in the validity of the results, nevertheless similar results to previous studies were achieved. The report covers the needed theoretical background on the topic and relevant past experiments. The whole project was subdivided into four main stages fur-

ther sub-divided into sub-stages. Each stage is fully documented and described in every step of its procedure. The final chapter illustrates the conclusions drawn from the analysis and a discussion about the problems and limitations encountered during the project period.

Acknowledgements

I would like to express my gratitude to the University of York and the Department of Electronics, for the resources and facilities provided which made this project possible.

A big personal thanks goes to the following people who played a vital role in my MEng project:

First of all I would like to thank my supervisor Tony Tew, for his great advice, patience and help demonstrated during the project period, and Chris Pike from BBC R&D in Salford, who was the main inspirator behind this project and also provided invaluable help by supporting me with advice, suggestions and resources such as the headphone equalisation code and the SDT analysis code.

A very special thanks goes to Andrew Chadwick, which patiently helped me to set-up the experiment environment in the anechoic chamber, lent me the equipment for the measurements, put up with my constantly changing timetables and taught me how to saw wood. Without his help, things would have been much harder.

Thanks to Anton Brunyee for helping me build the phantom power supply unit and for his considerable help in this project. Likewise, thanks to all the other experiment participants, for the patience demonstrated in the lengthy and uncomfortable measurement sessions and for having

been available to last-minute calls. Thanks also to all the people who helped me proofread this report.

Two special thanks go to Andres Calabbresi and Andrea Fassina for having helped me understand electronics better and the moral support given to me throughout the years.

Thanks to my second supervisor Helena Daffern, thanks to my past and present academic supervisors Adar Pelah, Andy Hunt and Dave Pearce. Finally thanks to Fraunhofer IIS for having introduced me to spatial audio, and to everyone that worked there with me.

A final thanks goes to my family because without their support I would not have come to study in England to pursue this degree.

Glossary

This section provides a reference to all the acronyms and abbreviations used throughout this report. More detailed explanations are covered in the introductory chapters.

- **HRIR** - Head Related Impulse Response
- **HRTF** - Head Related Transfer Function
- **BRIR** - Binaural Room Impulse Response
- **BRTF** - Binaural Room Transfer Function
- **HpIR** - Headphones Impulse Response
- **HpTF** - Headphones Transfer Function
- **ITD** - Interaural Time Differences
- **ILD** - Interaural Level Differences
- **CC** - Cross Correlation
- **SDT** - Signal Detection Theory
- **QoE** - Quality of Experience
- **FIR** - Finite Impulse Response
- **DAW** - Digital Audio Workstation

Contents

Abstract **3**

Acknowledgements **5**

Glossary **7**

1 Introduction **17**

1.1 Overview 19

1.1.1 Role in industry 21

1.1.2 Applications 23

1.2 Issues of binaural audio 24

1.2.1 Individualisation and Reverberation Factors 25

1.2.2 Quality Assessment scales 26

1.3 Project Motivation 30

1.3.1 Objectives 31

1.3.2 Research Hypothesis 32

1.4 Report Structure 33

2 Literature and Design **35**

2.1 Binaural audio production: Technical background 36

2.1.1 Binaural Cues 37
2.1.2 Binaural recording 40
2.1.3 Binaural synthesis 40
2.1.4 Other 3D sound technologies 45
2.2 Signal Detection Theory 49
2.3 Previous experiments 52
2.3.1 Interaction of factors experiment 53
2.3.2 Externalisation in dark environments 55
2.3.3 Assessing plausibility using SDT 57
2.3.4 Plausibility in small room environments 61
2.4 Experiment Design 65
2.4.1 Requirements 66
2.4.2 Breakdown of stages 67
2.4.3 Specifications 68
2.4.4 Project Management 72
2.4.5 Listening Test Design 72
2.4.6 Differences with initial plan 77

3 Equipment and Facilities 80

3.1 Facilities 81
3.1.1 Environment Set-up 83
3.2 List of Equipment 89
3.2.1 Experiment Chair 92
3.2.2 Power supply unit 93
3.3 Problems Encountered 102

4 Individualisation Measurements 105

4.1 HRIR Measurements 106

4.1.1 Sine-sweep Technique 107

4.1.2 Free-Field measurements 108

4.1.3 Subject Recruitment 111

4.1.4 Measurement procedure 113

4.1.5 DAW workspace 116

4.1.6 Individual HpIR measurements 116

4.2 Signal Processing 118

4.2.1 HRTFs processing 119

4.2.2 Free-field transfer functions 119

4.2.3 ITD and ILD correction 124

4.2.4 Headphones compensation filters 126

4.3 Problems Encountered 132

5 Listening Test 137

5.1 Stimuli preparation 138

5.1.1 Item List 140

5.1.2 Individualisation of stimuli 140

5.2 Listening Test 142

5.2.1 Preparation 144

5.2.2 Familiarisation process 144

5.2.3 Routine 145

5.2.4 Informal feedback 150

6 Analysis and Conclusions 153

6.1 Analysis 154

6.1.1 Results 156

6.1.2 Speaker positions 160

6.1.3 Correction groups 161

6.2 Conclusions 163

6.2.1 Comparison with previous studies 165

6.3 Discussion 166

6.3.1 Limitations 167

6.3.2 Further work 169

A Instructions for participants 178

B Supporting pictures 182

C Electret Microphone Capsules Datasheet 185

D Supporting Material 190

List of Figures

1.1 Head-related planes 18

1.2 Auditory Scene 20

1.3 Morphological head parameters 27

1.4 Morphological ear parameters 27

2.1 Localisation Cues 38

2.2 Inter-aural cues 38

2.3 Reversal error 39

2.4 Standard Artificial Head 41

2.5 Non-standard Artificial Head 41

2.6 HRTF pairs 42

2.7 HRTF-HRIR 43

2.8 Recording HRTFs 43

2.9 Convolution process 44

2.10 BRIR 45

2.11 Crosstalk Cancellation 46

2.12 3D audio with speakers 48

2.13 Plausibility experiment 50

2.14 Externalisation experiment 57

2.15 SDT in plausibility 60
2.16 Plausibility Analysis - Lindau 61
2.17 Plausibility Analysis - Pike 63
2.18 SDT sensitivities and biases - Pike 63
2.19 Project Flow Chart 73
2.20 Conceptual Set Up 75
2.21 Routine design flow chart 76

3.1 Anechoic Chamber 81
3.2 Anechoic Chamber Dimensions 82
3.3 Reverberant Room 83
3.4 Reverberant Room Dimensions 84
3.5 Hanging Speaker 86
3.6 Anechoic Chamber set-up 87
3.7 Environment Set-up 88
3.8 Listening Room set-up 88
3.9 Loudspeaker 90
3.10 Soundcard 90
3.11 Headphones 91
3.12 amplifier 91
3.13 Ear Sponges 92
3.14 Experiment Chair 94
3.15 Chair back view 94
3.16 Wooden support 95
3.17 Adjustable headrest 95

3.18 Head-strap and Blindfold 96

3.19 Wooden Platform 96

3.20 Microphone Capsules 97

3.21 Electret Microphone pins 98

3.22 Pre-amp configuration 98

3.23 Circuit Schematic 100

3.24 Testing the circuit 101

3.25 Phantom Power Circuit 102

4.1 Sine-sweep technique 109

4.2 Bent Capsules 110

4.3 Capsule connection 110

4.4 Free-field measurements 111

4.5 Measurement scene 115

4.6 DAW interface 117

4.7 Export settings 117

4.8 Signal Processing - Front speaker 120

4.9 Signal Processing - Rear speaker 121

4.10 Signal Processing - Free field measurement 123

4.11 ITD and ILD correction 127

4.12 HpTF measurements 128

4.13 Filter Wrapping 130

4.14 Equalised filters 131

4.15 Non-ideal conditions 135

5.1 Experiment scene 143

5.2 Preparation of Subject - 1 146

5.3 Preparation of Subject - 2 146

5.4 Preparation of Subject - 3 147

5.5 Listening test code 149

6.1 Anechoic Chamber results 157

6.2 Listening Room results 158

6.3 SDT distributions 159

6.4 Position-dependency 162

6.5 Correction groups 164

B.1 Anechoic Chamber - distant view 182

B.2 Listening Room - distant view 183

B.3 1st chair version 183

B.4 2nd chair version 184

B.5 1st circuit version 184

List of Tables

3.1 Speaker Positions 86

3.2 Equipment Used 89

3.3 List of components 99

4.1 Details of participants 113

5.1 List of Items 141

6.1 Overall sensitivities and biases 156

6.2 T-test for d'_{min} 160

6.3 T-test for $d'_{p_c 60\%}$ 160

6.4 T-test for $\beta = 1$ 160

6.5 Position-related sensitivities 161

6.6 Correction Groups sensitivities 163

Chapter 1

Introduction

Contents

1.1 Overview 19
1.1.1 Role in industry 21
1.1.2 Applications 23
1.2 Issues of binaural audio 24
1.2.1 Individualisation and Reverberation Factors . . 25
1.2.2 Quality Assessment scales 26
1.3 Project Motivation 30
1.3.1 Objectives 31
1.3.2 Research Hypothesis 32
1.4 Report Structure 33

Spatial audio research is a branch of acoustics and audio technology that aims to exploit the sound localisation abilities of our brain. One of the most famous examples of spatial audio is the two-channels stereo sound format which allow us to distinguish left/right directions in sound. Developing the next step from stereo is the main purpose of spatial audio: this field looks into ways to develop a full multichannel 3D surround sound experience where sound can come from all directions, elevations and distances. The resulting 3D experience is referred to as an auditory scene.

It is not necessary to have multiple sound sources in order to obtain a multichannel auditory scene. Virtual acoustics is a sub-field of spatial audio that aims to virtualise the multichannel set-up into headphones, creating *virtual sound sources* that can potentially make the virtual experience undistinguishable from the real experience. Any position in the horizontal and median planes can theoretically be recreated by virtual acoustics; the horizontal plane delines the source azimuth angle and the median plane the elevation angle (figure 1.1). One of the main tools used by virtual acoustics to achieve the spatialisation effect is *binaural audio* technology, probably the most well-known technique for obtaining virtual 3D sound. Binaural audio attempts to make the eardrums vibrate in the same way that a real acoustical source does.

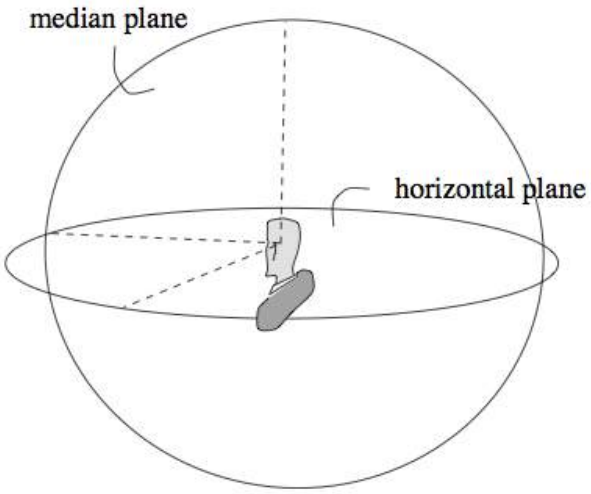


Figure 1.1: Horizontal and median planes of perception [1]

A perfect virtual experience is hard to recreate outside the laboratory and authentic virtual scenes are unpractical to deliver in large scale. More lenient assessment parameters, like plausibility [2], are now reported as sufficient to indicate realism of the binaural spatial audio experience. Understanding binaural audio, its principles and the issues related to it, is vital to understand this project and the research motivations behind it.

The experiment portrayed in this report is a research-based project aimed

at explorng and investigating the influence of certain factors, such as reverberation and individualisation, in the subjective assessment of binaural audio. The experiment is focused on the assessment of plausibility of binaural virtual sound sources in different reverberation environments using individualisation of spatial cues (section 1.2.1).

This chapter introduces the reader to the concept of binaural audio and the essential background, further explained in chapter 2, necessary to understand the project. An overview of the project's aims and the break down of the project stages is then illustrated.

1.1 Overview

> "Binaural: of, relating to, or used with both ears."
>
> - The Oxford Dictionary.

Binaural Audio is an emerging audio format which aims to reproduce the experience of multi-channel surround sound on headphones, or more generally, a two-channel playback source. This format is informally referred to *"3D Audio"* or *"Virtual Surround Sound"* to reflect its main feature: an 'out of head' localisation of the virtual sound sources in the surrounding environment.

In fact, the purpose of binaural audio is to provide the user with the same sensation experienced, for example in a surround sound set-up (e.g. the 5.1 six channels home cinema format) through, headphones or a simple stereo speaker set-up (cross-talk cancellation [3], see next section). The spatialisation effect is obtained by exploiting the psycho-acoustical functions of the human brain which automatically detects, combines and processes the acoustical spatial cues present in real sound sources[4]. These cues, referred to as binaural cues allow us to perceive and localise sound

in the surrounding space (see section 2.1.1).

The spatial cues recreated in the binaural signals allow the listener to perceive the location of the virtual sound sources somewhere in the surrounding environment as opposed to standard stereo which delivers the sensation that the sound sources are within the head[5]. The consequential effect of binaural audio is the sensation of "being there" in the soundscape rather than the sound "being here" in our heads[6]. Figure 1.2 illustrates a graphical example of what is meant by auditory scene. In this case a 5.1 surround set-up is virtualised (excluding the .1 low-frequency channel), the arrows indicating the concept that the virtual sound sources can potentially be moved anywhere in angle and distance.

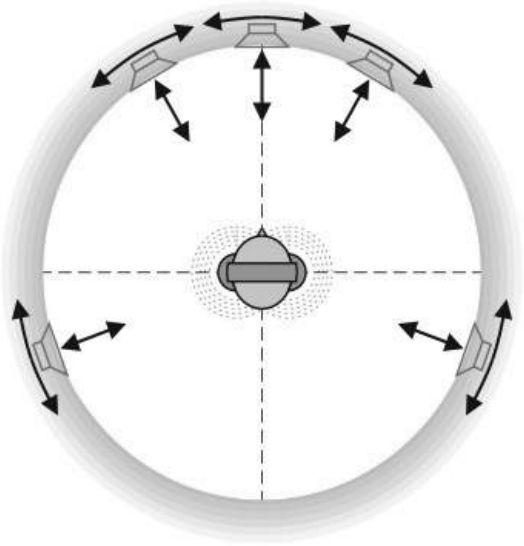


Figure 1.2: A 5.1 virtual auditory scene [7]

Binaural audio formats aim to virtually reproduce the spatial perception of a real sound-field by recreating the same binaural cues of the real sound sources. This can be pragmatically achieved either from direct recording with dummy heads or various synthesis techniques (see chapter 2). In [5] Pike expresses the concept that binaural synthesis "can be used to create auditory events at locations outside of the head with well defined direction and distance. [...] Binaural processing therefore has the potential to create realistic three-dimensional auditory scenes,

which could give a more immersive and engaging listening experience."

1.1.1 Role in industry

In today's modern technology, the increasing popularity of mobile portable devices such as smart-phones and tablets have increased the everyday number of users that listen to audio through headphones and earphones [8] [9]. This trend forms the perfect environment for the production, development and distribution of binaural audio formats. An ever increasing number of software applications dedicated to media entertainment delivery make use of binaural technology to improve the spatial impression for the listener. Furthermore, binaural offers a surround sound experience that could only otherwise be achieved by an expensive and non-portable surround system, thus making it an affordable attractive option for many users.

However, due to the issues that are mentioned in the next section (1.2), big companies involved in audio production and technology are not yet entirely confident about investing in binaural technology. In fact, despite its increase in popularity, binaural is still in an experimental stage. Currently binaural audio is well-known within the audiophile niche but still has to reach the degree of popularity needed to attract the masses. In any case, binaural over headphones is considered the best way to make immersive 3D audio mass consumable [2].

From a technical point of view, one of the issues with real-time converting of non-binaural audio formats into binaural audio, has been that of designing an efficient conversion method compatible with low-power devices such as mobile portable devices [10]. New rendering techniques combined with more powerful processors in modern devices aim to minimise latency and solve this issue. In [11] a hybrid conversion technique that makes use of a combination of frequency-domain sparse convolution

and a stochastic reverberator algorithms, it is illustrated how substantial improvements in computational speed can be obtained. Applications such as teleconferencing, that require a fast real-time rendering process, would considerably benefit in implementing these improvements.

Codecs like MPEG-Surround are specifically tailored to deliver multichannel sound, including binaural, to mobile user listeners in a flexible way for both real-time and on-demand audio content[10], while maintaining a legacy function that would allow mono/stereo sound to be played from devices not compatible with multichannel formats. MPEG-Surround is currently developed by Fraunhofer IIS, Philips, Dolby and LSI [12]. This codec aims, among other functions, to optimise the real-time rendering process of mono/stero/multichannel sound into binaural by making efficient use of the limited computational processing power of mobile devices. Furthermore MPEG-Surround is able to transmit multichannel data sound at stereo bit-rates which are much lower [12]. Thanks to this codec, the issue of how to deliver binaural sound to the masses in an efficient way could be solved and a bigger audience target could be reached.

Considering the perceptual improvements brought by individualisation, it is highly impractical to measure the individual acoustical parameters of the entire consumer target and potential users in the world. The drawback of producing individual binaural sound lies in the lengthy individual acoustical response recording sessions to be performed on the subject himself and the expensive equipment needed. This makes individualisation a non-ideal solution, but still important in research and worthy of further investigation. Future research might make individualisation more attractive for the industry. An Italian research [13] has investigated a method to extract morphological parameters from photos of the user's head and ears, and assign, from a database, the acoustic binaural cues recorded for a morphologically similar subject. Other research ([14], [15]) uses mathematical models to simulate personalisation of bin-

aural audio by controlling the morphological parameters, thus avoiding the standard lengthy measurement sessions.

1.1.2 Applications

Binaural audio has been studied and explored in applications and media production for a considerable amount of time. Artists like Lou Reed (*Street Hassle*, 1978) and Pearl Jam (*Binaural*, 2000), have experimented with music and produced albums recorded in binaural. Most of the binaural music produced has come from independent artists or studio recording teams like *Kall Binaural* [16]. Broadcasters like *BBC* [17] or German radio *BR Klassik* [18], already offer the possibility of listening to some of their content in a binaural format as well as standard stereo format.

The numerous possibilities enabled by the application of virtual acoustics could open the doors to new features for audio/video broadcasting not only on portable devices such as tablets and smart-phones, but also television and gaming platforms. Some examples are teleconferencing, video-gaming and creations of virtual auditory scenes. A good example is a study in Chile [19] that experimented the use of 3D virtual auditory scenes (*AudioDoom*) on blind children. In this case, binaural audio was used to create a navigable entertaining experience based on spatial sound that could allow the visually impaired children to "explore" the virtual sound-scape environment.

Entertainment is not the only purpose of binaural audio; various research fields have benefited from the innovative tools brought by binaural technology. Zahorik in [20] describes how "binaural technology has enabled realistic virtual listening simulation of a variety of room environments from anechoic rooms to concert halls". Not only these techniques allow the listeners to "evaluate the acoustics of different environments without being physically present in the environments, but also afford architec-

tural acousticians and sound engineers to reach a level of control of the acoustic stimulus captured by the listener's ear that would be impractical or impossible in real acoustic listening spaces".

1.2 Issues of binaural audio

It is beyond the scope of this project to explore in detail the efficiency aspect of binaural audio; more focus is instead dedicated to the psychoacoustical-related issues of this format. While computer systems improve year-by-year in processing power and computational speed, binaural audio finds most of its problems to be related with subjective perception and accuracy of the virtual auditory scene.

Pike in [5] mentions that "despite huge research and development effort, binaural technology has not yet reached widespread success in media entertainment. There is currently limited evidence of binaural processing creating an improvement in the quality of the listening experience in media entertainment applications when compared with stereo signals". It is therefore necessary to develop further understanding of the factors that influence the subjective assessment of binaural audio quality, how to better tailor it for the application and in which instances it actually improves the listening experience.

Subjective assessment of binaural audio produced using generic non-individualised parameters (i.e. binaural material produced using standardised spatial cues parameters recorded on artificial heads) has demonstrated in several occasion, that many listeners experience a poor 'externalisation' effect ([21] [22]). This issue leads to errors and inaccuracies in terms of localisation of virtual sound sources, their directionality, perception of distance and perception of timbre. *Lateralisation errors* occur when the listener perceives the location of a virtual source as $\pm 90^{\circ}$ on

either side of the head (full lateral) rather than a partial lateral position (e.g. at $\pm 45^{\circ}$). These errors can even result in listeners perceiving the virtual sound sources "in-the-head" rather than "out-of-head", practically nullifying the purpose of binaural audio for that particular listener.

1.2.1 Individualisation and Reverberation Factors

In recent years, audio technology research dedicated to binaural audio has been focusing on how to ameliorate the externalisation of the sound localisation for individuals. The accuracy of the 3D localisation of virtual sound sources is still very dependant from individual-to-individual. It has been proven that subjective morphological characteristics of the head, pinna (outer ear) and ear canals can modify the binaural spatial cues at the ears' entrance [15], especially spectral cues. The use of binaural audio prepared for a specific head/pinna shape, on an individidual whose morphological characteristics are very different, can lead to localisation error[23].

Thus, the individual binaural cues, caused by head measures and pinna ear shape, can be taken into account when producing binaural material using an "individualisation" process. In this way, the resulting material would be specifically tailored for the subject whose morphological characteristics have been taken into account. In literature, it is generally agreed that the individualisation of binaural audio processing can significantly improve the spatial perception and localisation accuracy. This thesis is further supported by the research conducted by Møller's [24] and Minaar's [25] research, which focused on the extent of the importance of individualisation. However, individualisation would only successfully work for the one subject chosen for the individualisation; any other person would experience a degree of localisation accuracy highly dependent on the degree to which his head/pinna morphology matches

with the person for which the binaural content was created [24].

The influence of reverberation factors in the subjective perception and assessment of binaural is also topic of debate. The research of Begault [22] and Völk [26] established that the presence of room reverberation in binaural sound can reduce localisation errors on the azimuth plane (at the cost of an increase of elevation error) and can generally improve the perceived sense of "externalisation", defined as the perceived distance of an auditory event from the centre of the head. Kim et al. [21] point out that "externalisation" errors that occur when using non-individual binaural audio, can be reduced by applying room reverberation to the stimuli. In these papers, it is agreed that the best combination to reach an optimal binaural 3D experience is to apply reverberation to individualised binaural audio. Some studies like the one of Zahorik [20] and Lindau [27] suggest that the effects of reverberation have more impact on the listeners' perception of quality, compared to the effects of individualisation.

The two figures below illustrate the concept of individual parameters. Figure 1.3 shows a MRI scan of a head showing the location of the auditory channels and the distance between them. The scans were are used to inspect the distance between the subject's eardrums and inspect what impact this distance would have on the perception on the perception of non-individual binaural. Figure 1.4 shows 3D models of a particular ear shapes operating at different acoustic modes. The red shade shows the region's high degree of influence on perception of spectral cues.

1.2.2 Quality Assessment scales

The mission of virtual acoustics, and therefore binaural audio, is to virtually recreate a real sound scene to the listener, providing a full and accurate aural immersion. Listening tests are the main source of quality assessment input for binaural audio production; they are used to

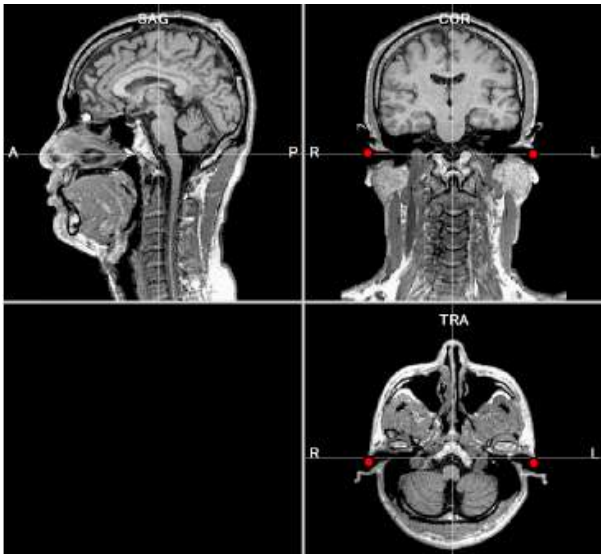


Figure 1.3: Example of a head MRI highlighting the auditory channels locations [14]

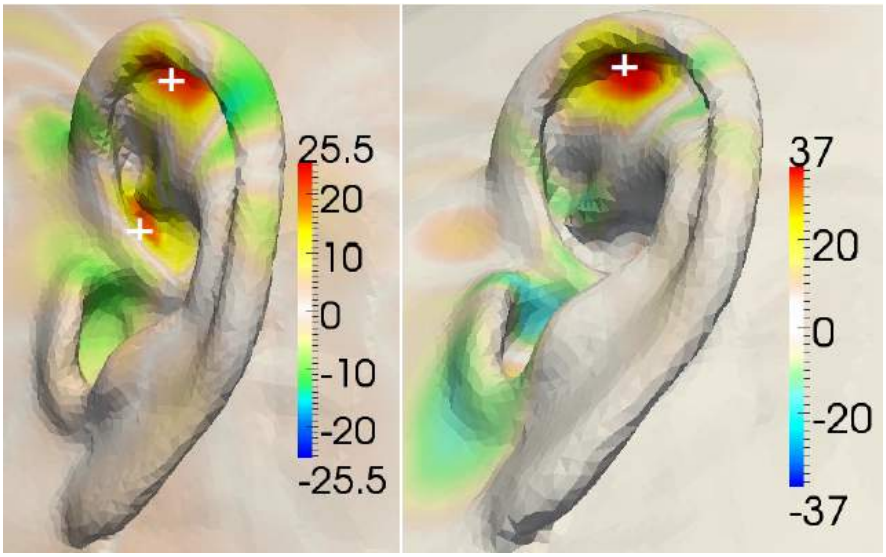


Figure 1.4: Example of two different ear models [15]

judge the quality of the virtual audio by using an assessment criterion given by the experimenter. Aside from spatial impression, signal colouration and timbre degradations can be affected by the binaural processing. When validating binaural systems, unidimensional scales or evaluations of listener's preferences cannot "measure the degree to which the desired enhancement of spatial impression is achieved or how each aspect affects a listener's liking of experience" [5]. Spatial quality assessment is a first step towards the validation of binaural systems and it is the aspect of binaural which is relevant for this report.

Assessable spatial audio attributes can be divided into physical attributes and psychoacoustical attributes. Physical-related attributes are perceptual attributes that can be "directly linked to a physical or mathematical property of either the sound source, the acoustic space or the sound reproduction system" [2]. Some examples are *source location*, *sound depth* and, more importantly, *reverberation* (room-related attributes). The second kind of attributes are more related to the listening experience than to the physical properties of the sound sources or the room. These attributes can be defined as *naturalness, readability, emotion*, and also *plausibility*.

It is of trivial importance to find the best quality assessment scale that can be used to objectively judge the Quality of Experience (QoE) of the virtual binaural auditory scene but this is still subject of debate [2] in the audio engineering community. Some of the criteria used in the past look for degree of *authenticity*, *externalisation* or generally, *sense of being there* [2]. The problem with most of these scales is that they are *multidimensional* and prone to being misinterpreted; in other words not clear enough to focus on one particular aspect and making it unclear what aspects are actually evaluated by the listeners [5].

As our brains make judgements by combining the inputs of all our senses [28], Blauert and Jekosch in [29] argue that pure real auditory sound scenes do not exist in real life as the influence of the visual senses in loc-

alising sound is too vast. It is therefore impossible to recreate an authentic sound scene using virtual audio simulation without complementary stimuli for the other senses that add to the aural perception. For these reasons, a direct immediate comparison between real audio and simulated audio cannot be realised experimentally, but is also not necessary.

In a recent experiment, Lindau [27] proposed a judgement binary variable based on a subjective assessment of *plausibility* as opposed to *authenticity*. In fact, the latter its considered an overly strict criterion [30], as trying to assess objective reality of the virtual scene would demand a degree of similar perceptual identity with the relative real acoustic event that simply cannot be achieved. Therefore a *plausibility* subjective assessment binary decision variable seems more appropriate when judging the quality of realism of a binaural event.

The plausibility of an acoustic environment, is defined as:

> "A simulation in agreement with the listener's expectation towards an equivalent real acoustic event"
>
> - Lindau and Weinzierl [27].

Plausibility can be assessed in a simple binary way. Lindau [27] and Pike [5] had experiment subjects sit in a room with speakers, wearing headphones. Signals would then be presented in a random way either from one of the loudspeakers or from the headphones in binaural (representing the same positions of the loudspeakers). The assessment was based on a YES/NO paradigm which asked the listeners whether the source was for them real (coming from loudspeakers) or non-real (virtual, coming from the loudspeaker). Plausible virtual sound sources would then be perceived as coming from the loudspeakers instead of the headphones. To ensure that the listeners were not using an immediate comparable reference, each item was played only from either loudspeaker or headphones

in a random way, hence making subjects base their judgement only on their inner expectation.

This way, the task of assessing how far binaural synthesis is able to provide substitutes for real sound fields can be based on each listener's own experience and expectations that result in different inner references of 'reality'. The referral to this inner reference very often corresponds to the scenario in which most users evaluate the quality of a simulation. Plausibility is also a less ambiguous variable to judge than the other ones mentioned due to the simplicity of the question.

Using signal detection theory analysis (see section 2.2) with the plausibility assessment, the individual response bias of each participant can be accounted for and separated from the averaged sensorial difference, making it possible to find out whether subjects actually distinguish reality from the simulation or if their decisions are close to guessing.

1.3 Project Motivation

The problems and the issues illustrated so far in this report raise the need for further understanding of what factors can influence the perceived QoE. It has been established in section 1.2.2 that multi-dimensional assessments can result in unclear interpretations of listening test experiments, therefore a uni-dimensional assessment would serve a more suited investigation procedure for assessing a particular aspect or attribute.

Given that plausibility is now a widely accepted scale for judging spatial audio, it would be interesting to investigate the interaction of external physical factors with this assessment methodology. *Plausibility* represents a realistic subjective scale for which users of 3D audio would typically assess the quality of the material presented; however, external conditions

may affect the context over which this decision is made.

Although previous experiments have been conducted in this direction (see chapter 2, section 2.3) and explored some attributes of spatial perception, no research has yet investigated the interaction of different reverberation conditions and individualised binaural audio using a binary yes/no plausibility assessment methodology.

The dependent variable involved would be that of reverberation; an anechoic chamber could then be compared with a typical small reverberant room, and the difference between experiment results for the two rooms can be analysed. Individualisation is here a fixed variable. The choice of using individualisation derives from the fact that, as mentioned before, it is widely accepted that creation of binaural spatial audio using individual response parameters can significantly improve the spatial perception, and therefore improve plausibility. Also as mentioned, current research aims to make individualisation a remote process and therefore more practical.

1.3.1 Objectives

This project aims to pursue the following objectives:

- Design an experiment that allows to investigate the influence of the interaction between reverberation and individualisation factors on the assessment of plausibility, in a binaural auditory scene.
- Measure, for a number of subjects, the individual acoustical binaural cues that can be used to produce spatial audio tailored for the subject.
- Conduct the experiment on the subjects and assess subjective perception of spatial audio quality based on a plausibility decision,

in two rooms differing for reverberation characteristics, using individualised binaural audio

- Analyse the experiment results for the two different environments using signal detection theory, and draw conclusions about the influence of reverberation in the plausibility assessment
- Compare the conclusions drawn in this experiment with previous findings
- Discuss whether it would be worth to conduct further research in this direction and what are the limitations in this experiment
- Analyse whether the objectives were accomplished and assess whether the project was carefully planned

1.3.2 Research Hypothesis

Although this project is research-based and the questions are open-ended, it is necessary to establish research hypotheses in order to set a direction for the experiment performed.

As a result of literature research and previous experiments, some introduced in this chapter and some in the next chapter, the following hypotheses have been made:

- Different reverberation conditions will yield different plausibility response for each individual subject.
- The combination of room reverberation conditions and individualised binaural audio will achieve the highest degree of plausibility.

1.4 Report Structure

This report is broken down into separate chapters, each of which represents a different stage of the project development in chronological order. Each stage can be divided into sub-stages which generally consist of requirements analysis, design, implementation, testing and correction.

Chapter 2 is dedicated to the preparatory work and literature review done in the first preparatory stage of the project ,which can be referred to as "*Stage 0*". The chapter covers a more technically detailed theoretical background on how binaural audio is produced and why it is preferred to other 3D sound technologies. The Signal Detection Theory model is covered as the chosen approach to the problem. Subsequently, past experiments related to this project are portrayed and discussed. After the discussion on the background and literature, the project requirements, specification analysis and listening test design are illustrated along with a flow chart diagram that served to organise the managment aspect.

Chapter 3 examines the hardware and the equipment prepared for the experiment and the measurement stages. The environment set-up is fully documented along with a list of everything that was done in preparation for the measuring stage. It is explained how the two environments, an anechoic chamber and a listening room, were carefully prepared with the same set-up, making sure all distances where reproduced identically in both environments. The chapter generally focuses on design, implementation and testing of equipment specifically built for the project such as a phantom power supply and a special experiment chair.

Chapter 4 describes the measurement procedure used to obtain the individual head-related-transfer-functions (HRTFs) which were used to create individual binaural material for each of the participants in the experiment. A testing stage highlighted some unexpected problems which

were eventually solved, although the perfect ideal conditions that were initially aimed for, could not be reached. Headphones equalisation measurements were also part of this stage. Finally, the measurements were processed in MATLAB, to produce the filters used in the following stage to create the binaural virtual sound material.

Chapter 5 presents the experiment implementation stage, the creation of binaural content and the test procedure. All the steps taken throughout the execution, starting from the choice of material and the creation of the stimuli, to the calibration of the system, are fully described. This part also comprises details on the instructions given to participants as the way they were prepared for the plausibility assessment formed a vital part in affecting their subjective judgement.

Chapter 6 finally portrays the processing of the results of the experiment using Signal Detection Theory, which are compared with the results of previous related studies, and the conclusions derived from it. This chapter includes a discussion on the limitations that affected the project and a portrayal of the problems encountered during the work period. The central topic of the discussion is the extent to which certain procedural mistakes might have affected the subjective assessment of the individuals, and, consequentially, the results. A final section on future possible work illustrates some change proposals for equipment and procedure adjustments for a more ideal repetition of this experiment.

The Appendix part that concludes the report includes the data-sheet of the microphone capsules, relevant MATLAB®(Mathworks, Inc.) and PYTHON code used, pictures of the equipment and project environments, and copies of the experiment instructions given to the participants.

Chapter 2

Literature and Design

Contents

2.1 Binaural audio production: Technical background . . 36
2.1.1 Binaural Cues 37
2.1.2 Binaural recording 40
2.1.3 Binaural synthesis 40
2.1.4 Other 3D sound technologies 45
2.2 Signal Detection Theory 49
2.3 Previous experiments 52
2.3.1 Interaction of factors experiment 53
2.3.2 Externalisation in dark environments 55
2.3.3 Assessing plausibility using SDT 57
2.3.4 Plausibility in small room environments 61
2.4 Experiment Design 65
2.4.1 Requirements 66
2.4.2 Breakdown of stages 67
2.4.3 Specifications 68
2.4.4 Project Management 72
2.4.5 Listening Test Design 72
2.4.6 Differences with initial plan 77

This chapter represents the introductory work done in the initial months of the project. A deeper understanding of the project needs to be backed

up by further understanding of binaural audio, hence the technical aspects of binaural audio production are covered. A more detailed explanation is given on the use of HRTFs as a method to achieve binaural synthesis. The literature section summarises the findings of previous relevant studies of Begault [22], Völk [26] and Lindau [27] and the research done in preparation for the project. An experiment conducted recently at BBC R& D [5] is also included for its relevance with this project. The last section in this chapter goes through the project requirements, specifications and design stages.

2.1 Binaural audio production: Technical background

The two main pathways for the creation of binaural are recording and synthesis. In the past decades several improvements in both methodologies have been achieved as a result of the exploration of virtual 3D audio. The most direct form of producing binaural audio is direct recording through a dummy head (artificial head), however, the drawback of direct recording is that audio recorded in binaural would not be suited for playback over mono/stereo systems.

Binaural synthesis techniques simulate the response of a dummy head and render non-binaural formats such as mono, stereo or multichannel into spatial binaural audio. This process is made possible by a filtering process between the audio and location-specific Head Related Transfer Functions pairs (HRTFs). HRTF pairs represent the properties of the spatial cues that are related to a specific point in space. When filtering a signal through an HRTF pair (Left ear and Right ear) these properties get applied to the signal, which becomes spatialised. For example, a multichannel format, conceived to be reproduced over a specific multichan-

nel speaker set-up, can be virtualised using the HRTFs related to the same spatial positions used in that specific format (see image 1.2 in chapter 1).

2.1.1 Binaural Cues

The human ears can detect binaural spatial cues which are received, combined and deciphered by our brains in order to localise the origin of a sound source. Localising sound is not only a matter of processing binaural cues: the influence of the other human senses, mainly vision, is integrated with the aural perception to enhance and complete the awareness and detection of the origin of sound sources.

In the early 20th century, Lord Rayleigh developed the so-called *duplex theory* of sound localisation[31]. Through his early acoustic experiments, he theorised the influence of *Inter-aural Time Differences* (ITD) and *Inter-aural Level Differences* (ILD) in the localisation of sound at different frequencies. He concluded that low frequencies were affected more by ITD, which is related to the head size, while high frequencies were influenced by the amplitude modulation of the pinna and thus by the ILD [28]. It is now accepted by the audio engineering community that ITDs are more influential than ILDs, and dominant when the two are in conflict. These cues are usually dominant for detecting sources' azimuth angle. In the case of elevation angle, ITD and ILD cues sound quite similar regardless of the angle: a source placed, for example, at a $+10^{\circ}$ elevation would have similar ITD and ILD to a source placed at an elevation of 0°, therefore not little information is provided by these cues. This is the situation where spectral cues further refine our perception.

Spectral cues in signals such as *spectrum shape*, *energy* and *width*, or more generally the distribution of frequencies, also have a strong influence on sound localisation perception, particularly for monaural perception [32]. Elevation differences between signals often present variances in notches

and peaks in the perceived frequency response. Figure 2.1 graphically shows the binaural cues described in these paragraphs, elevation is perceived mainly by spectral cues (A) while angle is detected by ITD (B) and ILD cues (C). Figure 2.2 is a further illustration of how ITD and ILD respond to the angle of a source.

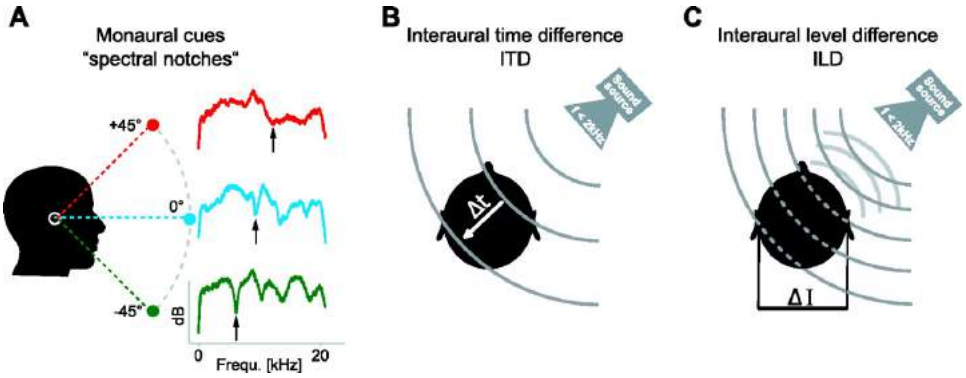


Figure 2.1: Cues for sound localisation [33]

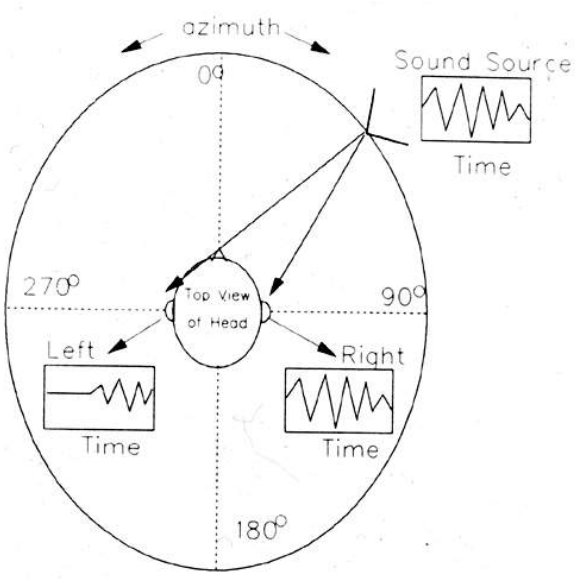


Figure 2.2: Representation of ITD and ILD [34]

A front/back confusion effect called *reversal error* often arises when reproducing binaural through headphones. Listeners sometimes perceive the sound source as behind the head rather than the front. However it is debated that this effect might be caused by the influence of the visual senses on our perception abilities as there are no differences in ITD or ILD

cues between a source located at an azimuth angle of $0°$ (front) or $180°$ (back) [35]. In other words, our eyes tell our brains there are no possible real sound sources in front of us, which is what we 'expect' to see, hence the virtual sound source 'must be' at the back, where we cant see. The reversal error, can be partially solved by using dark environment conditions [26] (easily reproducible for an experiment or listening test but quite unrealistic in a real-life situation) or head-tracking techniques [22]. Head tracking is performed using sensors or optical systems that can track the head direction of the listener in a virtual acoustic space and consequently fix the directionality of the virtual sources and pinpoint its location in space, instead of having them revolve along with the listener. This technique can reduce the reversal error rates in headphones binaural audio: as seen in figure 2.3, head movements would create an ITD difference which in turn provides the brain the necessary binaural cues for correct source localisation detection [22].

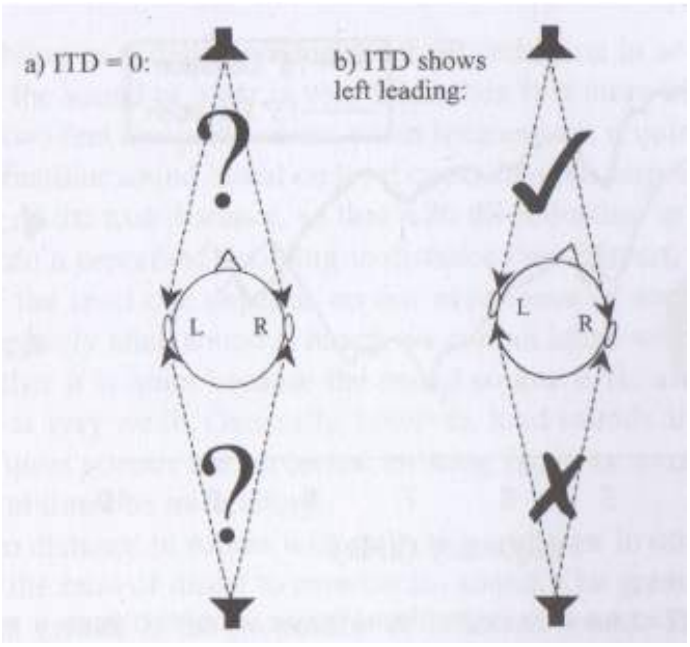


Figure 2.3: Front/Back perception [35]

An experiment conducted by Han [36] investigated the influence of memory effect as a source localisation cue. He showed that under certain conditions source movements do not account for resolving front/back inver-

sions once the position and the sound source were identified and memorised.

2.1.2 Binaural recording

Direct binaural recording is achieved by recording the sound, in a studio or *in-loco*, around a dummy head. Dummy heads are plastic mannequin reproductions of real human heads, complete with detailed plastic ears and ear canals. The idea behind a dummy head is that tiny microphones are inserted in the canals in the position where the human eardrum would normally be. The position of the microphone capsules record the sound as the eardrums would normally hear it in a human being, and, as a result, the spatial cues of the recorded sounds are preserved and caught by the listener when playing back the audio over headphones. Futhermore, the outer ear plastic piece, the exact position of the microphone and the depth of the canals would add further spatial characteristics to the sound which are unique for the dummy head used. In fact, the sound recorded through dummy heads is non-individualised as the morphological characteristics of the dummy, which are based on average ear shapes chosen by the constructors, may not match those of the end-user and lead to non-optimal spatial impression of the content [25].

Figure 2.4 shows a typical dummy head used for binaural recordings made by German company, Neumann. Figure 2.5 shows an unconventional dummy head constructed by an independent recording group for simultaneous binaural field recording of multiple listening orientations.

2.1.3 Binaural synthesis

Binaural audio can be simulated by applying special Finite Impulse Response (FIR) filters representative of the acoustical response of our ears

Figure 2.4: The KU-100 by Neumann [37]

Figure 2.5: Non-standard artificial head used for binaural recordings [38]

to a specific location in space. These filters are commonly called Head-Related-Transfer-Functions, in short HRTFs, and are always matched in pairs. Each pair contains the spatial binaural cues for left and right ear related to the position for which they are recorded [39]. Filtering a mono item with a HRTF pair transfers these cues to the item and assigns it a virtual position and distance associated to the pair. Figure 2.6 shows how each virtual source position is represented by a HRTF pair, one filter for the left ear and one for the right ear.

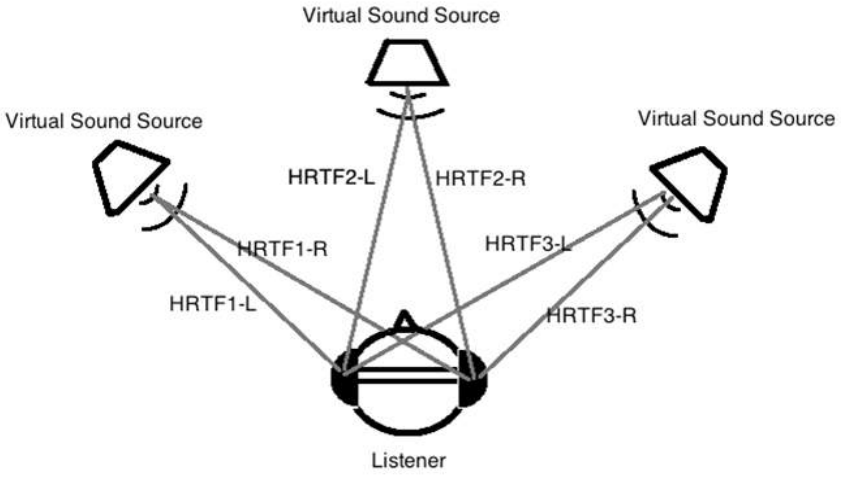


Figure 2.6: HRTF pairs for three virtual sources

HRTF pairs can be obtained from Head Related Impulse Responses (HRIRs) by computing a Fourier Transform; in fact, HRIRs are the time-domain version of HRTFs, and are easy to record. Figure 2.7 shows the relation between a frequency-domain HRTF pair and its time-domain version HRIR pair. The red and blue colours indicate the left and right channel. As the filtering process can be more efficiently and quickly performed in frequency domain using fast convolution, rather than time domain, it is very usual to transform HRIRs into HRTFs before synthesis.

To record HRIRs, a dummy head is used to record click-sound impulse responses played from loudspeakers placed at the desired positions (figure 2.8). The resulting HRIR pair will then be associated with the exact position of the loudspeaker. Alternatively, a sine-sweep technique con-

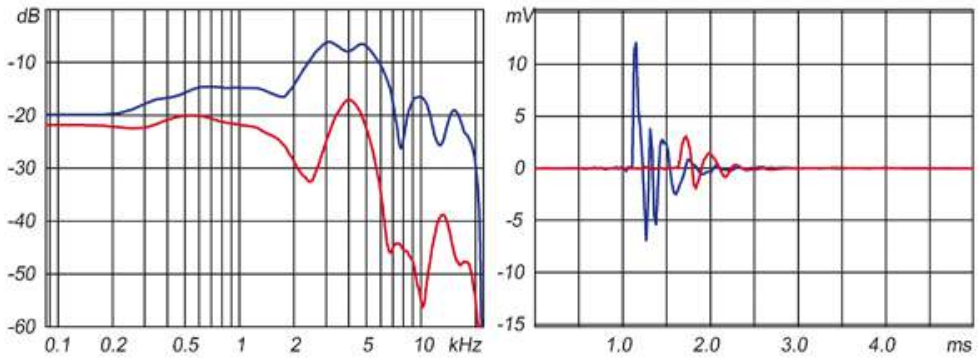


Figure 2.7: An example of a HRTF in frequency domain (left) and HRIR in time domain (right), for two channels (left and right ear) [40]

sists in a binaural recording of a sine-wave signal that sweeps through all audible frequencies and a deconvolution process can extract the HRIR pair from the recording [41].

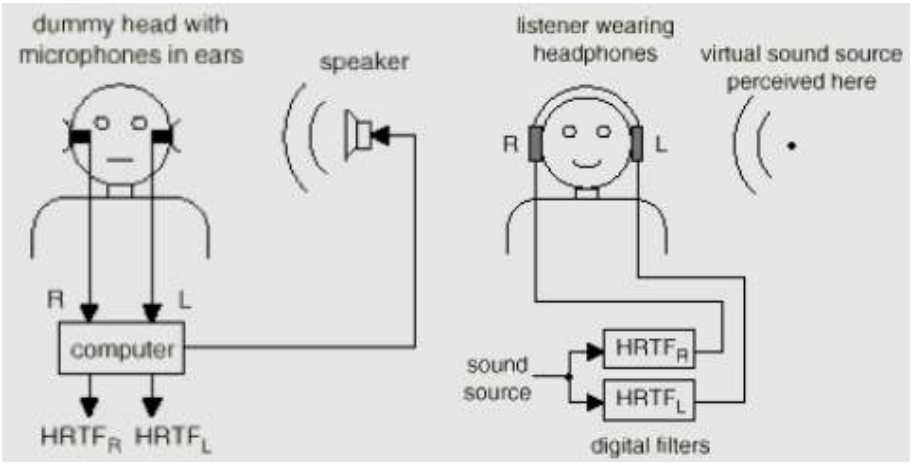


Figure 2.8: Concept of binaural audio simulation using non-individualised HRTFs [42]

The process of rendering mono items into binaural happens through either time-domain convolution or frequency-domain fast convolution, between the item and a chosen HRTF pair, creating two filtered signals, one for each ear. The spatial attributes information contained in the HRTF are transferred in this way to mono non-spatialised sound sources. In its final version, the resulting binaural signal is a sum of multiple HRTF convolutions (see figure 2.9): all the signals filtered for the left ear gets summed in the left channel, and the same happens for the right channel.

Using this methodology, it is possible to synthesise a binaural auditory

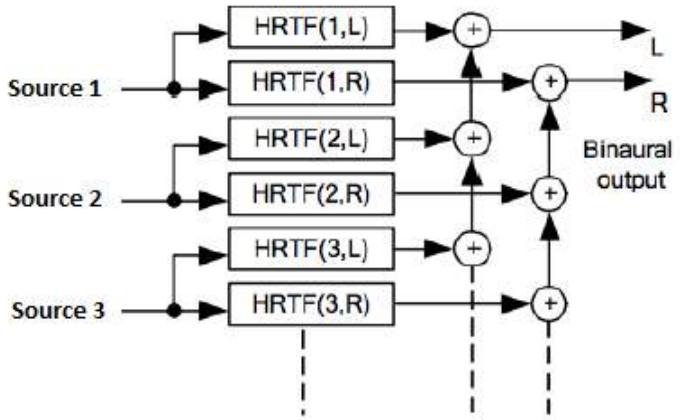


Figure 2.9: HRTF pairs are convolved with sources to create the binaural output [10]

scene without the need of a direct recording of the field using a dummy head. Yet, using dummy heads does not take into account the morphology of the individual which can modify the sound ulteriorly, to the point that the virtual positions lose all their spatial impression [43].

HRTFs can in fact be influenced by the already-mentioned external physical factors like room reverberation or by the ear/pinna shape over which they are recorded. Individual HRTF recordings are performed in the same procedure as non-individual recordings, with the microphone capsules inserted inside the subject's ear canals instead of the dummy's canals. As explained in chapter 1, this procedure can lead to creation of binaural content specifically tailored for the individual and thus improve the accuracy of the localisation of virtual sound sources in the 3D space [24], [25].

HRTFs are usually recorded in anechoic conditions. When the procedure is performed in a reverberant room environment, the echoic HRIRs will contain reverberation information. Reverberant HRIRs and HRTFs are called BRIRs and BRTF, respectively Binaural Room Impulse Response and Binaural Room Transfer Function. According to [22], the presence of reverberation in BRIRs, significantly influence the perceived externalisation ('sense of distance') and accuracy of directional localisation by the

end- user, meaning a better preservation of the spatial quality properties. The drawback of BRIRs is the higher computational power required to process the audio into binaural. See figure 2.10 for a representation of a BRIR FIR filter: the impulse response is divided into three parts, direct sound, early reflections and late room reverberation. Anechoic impulse responses lack the presence of any reflections as the materials used in anechoic rooms are able to absorb them, leaving only a direct sound component in the time-domain response.

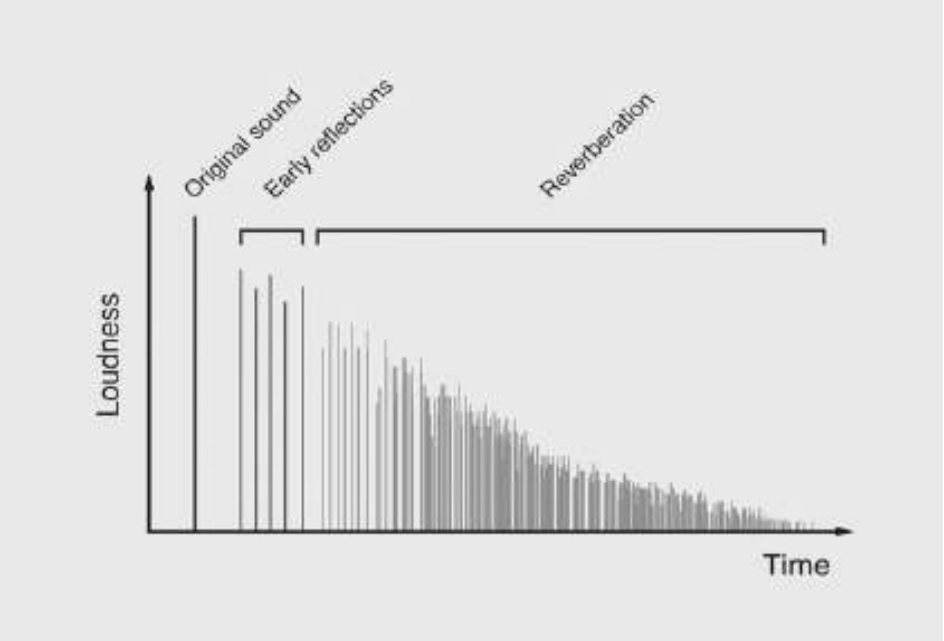


Figure 2.10: BRIR - Binaural Room Impulse Response: reverberant time-domain version of HRTF [44]

2.1.4 Other 3D sound technologies

3D audio is not limited to the use of binaural over headphones. Other technologies like crosstalk cancellation filters, higher order Ambisonics, vector based amplitude panning and wave-field synthesis, permit to experience 3D audio using speaker set-ups although some limitations are present. Although these other 3D sound reproduction techniques are not essential for the understanding of this project, they are still related to virtual acoustics and their limitations justify why binaural over headphones

is now the most virtual acoustics medium where more effort is put for development. Mainly, all the following techniques need loudspeaker set ups, limiting any mobility that the listener might want during reproduction.

Crosstalk Cancellation

Binaural audio can be played back by a simple stereo speaker configuration. Usually, binaural audio played from loudspeakers loses its spatialisation effect. This is due to an effect called crosstalk where the audio channel meant for the left ear gets picked up from the right ear and vice-versa. Special filters, in addition to the HRTF filters, can be applied to each speaker in order to avoid this effect, thus this technique is called crosstalk cancellation [3]; figure 2.11 illustrates the concept. The drawback of this methodology is the necessity for the listener to stand in a specific sweet-spot, otherwise the spatial impression would be lost. Furthermore, expensive head tracking technologies would be needed to account for head rotations. For these reasons crosstalk cancellation is not a popular reproduction technique.

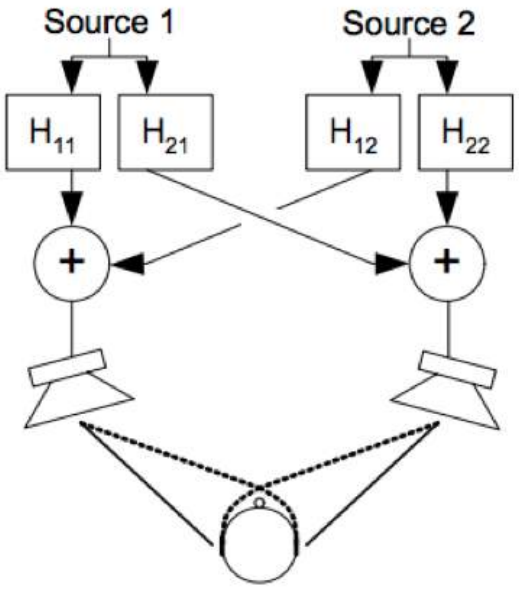


Figure 2.11: Crosstalk cancellation concept [10]

Ambisonics

Ambisonics is a recording technique used to capture a full sphere 3D sound field by adding a height dimension to the recording which turns the number of channels from 2 to 4. The decoding of the signal drives a number of loudspeakers related to the recording channels. This first-order ambisonics can be further increased to several dimensions of higher order (hence, the term highest-order ambisonics) reaching for example 16 or 32 channels and consequently a better sound field [45]. In simple terms, as explained in [45], the principle is "based on the idea of capturing the sound impinging on a single point in space from any and every direction [...] the incident sound can be measured by a combination of microphone capsules mounted in coincidence – or as near coincidence as is physically possible". The advantage of ambisonics is a very flexible playback ability, which, thanks to decoders, is able to recompute the format for a variable number of loudspeakers. However ambisonics is prone to phasing artefacts the moment the listener moves or turns since any one virtual source will be reproduced by several speakers with strong correlations.

Vector-based amplitude panning

Vector Based Amplitude Panning (VBAP) works on the principle that panning the volume between a number of speakers which are set up in specific configurations, for example triplets, can recreate any virtual position between the speakers [1]. This is controlled by adjusting the gains of the triplet with the position specific parameters. The figure below (2.12) illustrates the concept of using triplets of speakers.

The drawback of this solution is that, not only is one required to stand in a sweet-spot, but it is also highly expensive and impractical to set up this configuration in small areas for home-cinema purposes. On a perceptual side, a displacement of the virtual sound sources towards the

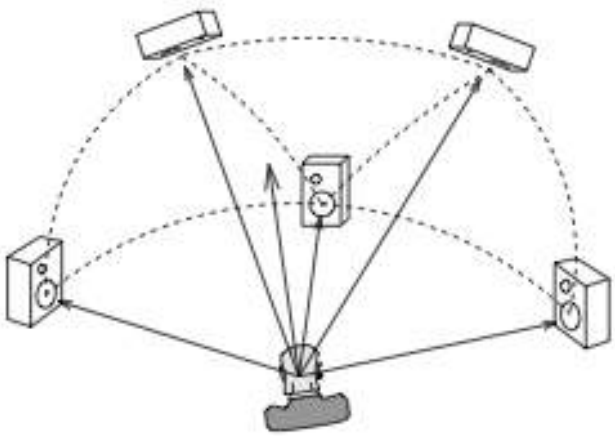


Figure 2.12: Triplet-wise amplitude panning with five loudspeakers [1]

median plane has been recorded when the loudspeakers are not placed symmetrically with the median plane, requiring compensation [46].

Wave field synthesis

Wave field synthesis is used to recreate a whole sound field using a very large number of speakers [1]. The advantage of this technique is that the localisation of the virtual sources does not depend on the listener's position as the representation of a virtual source and a virtual room is achieved by rendering an acoustically correct sound field. "The binaural ear input signals that are active for the auditory event thus arise in a natural way within the sound field, contrary to dummy head stereophony" [47]. This allows more mobility within the sound field and would make this technique suitable for creating areas of sound-field to multiple people simultaneously (e.g. *Tresor* club in Berlin, Germany, is a venue for concerts that use wave field synthesis technology [48]).

Unfortunately, like the VBAP technique, it is a very impractical and high-cost technique in most situations, as "the most restricting boundary condition is that the system produces the sound field accurately only if the loudspeakers are at a distance of maximally a half wavelength from each other. The centroids of loudspeakers should thus be a few centimetres

from each other to be able to produce high frequencies correctly also, which cannot be achieved without a very large number of loudspeakers" [1].

2.2 Signal Detection Theory

Signal Detection Theory (SDT) provides a model for the perceptual processes when detecting weak signals in the presence of internal noise [27]. SDT can apply to any area of psychology, and psychoacoustics, where two types of stimuli must be discriminated [49]. This particular method of analysis is appropriate for a binary judgement, such as that of plausibility, and already explored in similar experiments. Here, an overview of SDT given by [49] is portrayed and its application in yes/no tasks is justified.

SDT is specifically suited to yes/no tasks. Yes/no tasks involve *signal trials*, that imply *yes* as the correct answer, and *noise trials* that imply *no* as the correct answer. The response of the participants is based on a *decision variable* that is subjectively evaluated during each trial. If the decision variable is sufficiently high during a given trial, the answer would be a *yes* (signal is perceived). In the other case, a subject would respond *no* (noise was perceived). In [49] it is explained that "the value that specifies *sufficiently high* is called the *criterion*".

If a subject is capable of distinguishing between signal and noise, the decision variable is affected by the stimuli presented; other external factors such as fluctuations in attention may also affect the decision. Whenever a *yes* answer is associated correctly with a *signal trial*, a *hit* response is recorded. On the contrary, whenever a *yes* answer is given in a *noise trial*, it is labelled *false alarm* response. Likewise, a perception of a signal as noise is called a *miss*, and a correct perception of a noise signal is called *correct*

detection. *Hit rate* and *false-alarm* rate only can describe the individual's or group's performance on the yes/no task involved.

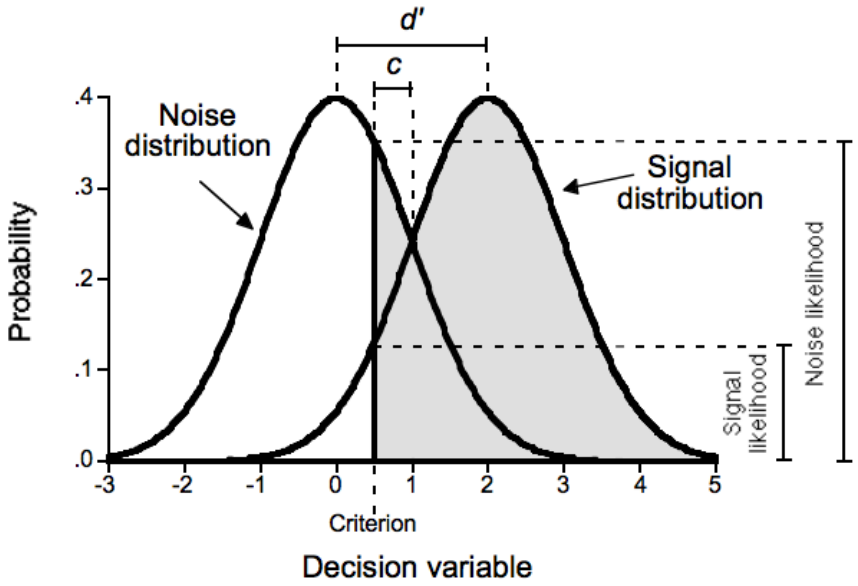


Figure 2.13: Distribution of the decision variable across noise and signal trials [49]

In figure 2.13 the *signal distribution* represents the distribution of values realised by the decision variable across signal trials, the hit rate is represented by the shaded region in the distribution that exceeds the criterion value. Similarly, the *noise distribution* corresponds to the distribution of decisions in noise trials and the false-alarm rate is represented by the shaded region exceeding the criterion value. Setting the value for the criterion affects the *bias*, for example it could be set to a very low level (*liberal*) to bias answers towards a *yes*, or to a very high level (*conservative*) to bias answers towards a *no*. The response bias (general individual tendency towards a *yes* or *no* response, determined by the location of the criterion) and the overlap of the distributions are reflected on the *hit rate* and *false-alarm rate*; it can be deduced that an increased false-alarm rate would be caused by a higher overlap of the two distributions.

Individual *sensitivity* to the decision variable, which is subjective, also add another layer of influence to test responses, making it harder to discern the true reasons for a particular answer, in a specific trial. If, for ex-

ample, different test conditions result in different *hit* rates, it would not be clear whether the conditions differ in sensitivity or response bias. Stanislaw and Todorov [49] explain how "SDT separates the response bias and sensitivity" and the two can be calculated separately. It is in fact pointed out that sensitivity is related to the overlap of the two distributions, which is in turn related to the distance of the means of the distributions (*d'*) and the standard deviation of the distributions. The overlap of the two curves will decrease if *d'* increases or if the standard deviation decreases. Hence, sensitivity can be quantified using the *hit* and *false-alarm* rates to determine the distance between the means, relative to their standard deviation. The distance value can be calculated as follows ([49], [5]):

$$\hat{d}'_i = Z(p_{Hit_i}) - Z(p_{FA_i})$$

Where the symbol ˆ indicates an estimated variable and *Z(p)* is the inverse cumulative normal distribution. A complete overlap of the two distributions, meaning a value of $d' = 0$, indicates a complete inability to distinguish signal from noise. Greater values correspond to greater ability and a value of $d' = \infty$ imply perfect discrimination ability. Negative values of *d'* can occur as a consequence of response confusion (confusing *yes* for *no*) or when the false alarm rate is greater than the hit rate.

Regarding the response bias, there are two accepted ways to calculate this. The most straightforward measure is calculating the distance *c* between the criterion point λ and the *neutral point* where the two distributions have equal value and neither is favoured. If *c* is negative, it reveals a bias toward responding *yes*, if *c* is positive, the bias goes toward *no*. The other way to quantify response bias is through the variable β, based on a likelihood ratio. The numerator of the ratio would be the height of the signal distribution at *x* and the denominator would be the height of the noise distribution at *x*. A $\beta < 1$ value indicates a bias towards *yes*, $\beta > 1$ is a

bias towards *no*. A completely fair observer would have a *c* value of 0, and a β value of 1.

To measure bias β, it is first needed to measure the individual response criterion, which is estimated from the false alarm rate and already provides a first indication of bias. The formula, taken from [49] and [5], is:

$$\hat{\lambda}_i = Z(1 - p_{FA_i})$$

Hence, the bias can be hence be calculated by the ratio of likelihood at the position of the response criterion [5]. φ represents the normalised probability density:

$$\hat{\beta}_i = \frac{\varphi(\hat{\lambda}_i - \hat{d}'_i)}{\varphi(\hat{\lambda}_i)}$$

In conclusion, Signal Detection Theory is a model that allows for the response bias to be interpreted separately and independently from the sensory difference. Having independent measures of these two subjective variables can lead to better interpretation of individual's response to a binary task test such as the one of plausibility where *noise* and *signal* can be applied to *real* and *simulated* stimuli.

2.3 Previous experiments

Past experiments have investigated the extent of synthesis quality using HRTFs measured in different reverberation environments on artificial or human heads. The following experiments were taken as a starting point on how to prepare and design the project experiment. The findings and experimental conclusions were used to determine what other aspect could be explored in order to motivate this project and contribute to the research in this field by bringing further knowledge. This section covers

in detail those experiments closely related to this project which were researched prior to the design stage including a discussion on the validity of the methodologies in relation to the assessment of *plausibility*.

2.3.1 Interaction of factors experiment

This experiment, "*Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualised Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source*", was conducted by Begault et al. [22] in 2001.

The main objective of this research was to investigate the interaction of several factors that could influence perception by focusing on the subjective judgement of a virtual binaural speech signal. Listening test subjects were asked to assess, through a software interface, the directionality of the virtual sources, the "externalisation" perception and the amount of "realism" of the experience using an arbitrary scale. This study also aimed to establish which listening conditions would bring the least amount of *localisation errors* in the median and horizontal plane, and *reversal errors* between front and back. Although the influencing factors in QoE perception were already identified, it was previously not clear if all of those factors "contributed equally to the accuracy and overall quality of auditory localisation in a virtual acoustic display, or if instead these factors contributed only to specific aspects of localisation" [22].

The listening test conditions that were used to determine the influence of the factors involved, in different combinations, were the following:

- Individual HRTFs vs. Artificial HRTFs
- Anechoic vs. Low reverb vs. Full reverb
- Head-tracking vs. No head-tracking

A software interface was used during the experiment by the subject to evaluate source direction and elevation angle on a graph, and perceived *realism* with a slider. *Externalisation* was determined by judging whether the sound seemed inside or outside the head and rating this sensation on a scale 0 to 4.

The conclusions drawn from this study served as a further validation of commonly held assumptions: the paper concludes that the optimal combination of factors, for minimal perceptual errors, consists in a head-tracked, reverberant, individualised test condition. According to the conclusions, head-tracking is the main factor that considerably reduces the number of front/back confusion rates whereas reverberation is the main factor that improves the azimuth localisation error in the horizontal plane with no difference between small and big reverberation amounts. Nonetheless, a degradation in elevation error in the median plane was documented. Externalisation error is also mainly affected by reverberation, this is verified by other studies [43] that also confirmed how an 80ms reflection time is sufficient to improve the perceptual sense of distance. An interaction between individual HRTFs and head tracking was found to slightly improve azimuth localisation error as well. It has to be noted that the reverberation used in this study was artificially added to the dry anechoic items, hence not directly related to the room reverberation where the subject was presented the audio material.

As demonstrated by Møller et al., the problem with focusing only on speech signals is that individual HRTFs gave no advantage in localisation accuracy for speech. This might be due to the fact that "most spectral energy of speech is in a frequency region where ITD cues are more significant than spectral cues" [22]. A necessary expansion of the experiment would be to inspect the interaction of individualisation and other factors in the assessment of other kind of signal items, such as music and sound effects, which cover different regions of the audible frequencies.

It was also concluded that the methodology used to assess the "realism" amount was far too vague and did not lead to any useful conclusion as no variability between conditions was recorded. This was explained by the fact that perhaps each listener had a different understanding of what "realism" meant due to the lack of an acoustical reference of what was "real" and what was not. *Plausibility* as an assessment parameter aims to address this problem by providing a reference of what can be perceived as real, while at the same time, an immediate comparison of the signal versions like in the assessment of *authenticity* can be avoided.

2.3.2 Externalisation in dark environments

A similar experiment was conducted by Völk et al. in 2008 and published in a conference publication: "*Externalisation in binaural synthesis: effects of recording environment and measurement procedure*" [26].

The listening test described in [26] focused on a qualitative assessment of *externalisation* in terms of subjectively perceived distance from the centre of the head, using similar combinations of variables as Begault [22]:

- Human-Head HRTFs vs. Artificial-Head HRTFs
- Room reverberation vs. Free field

It was emphasised that in this test the Human Head HRTF did not represent individual HRTF recordings for every single subject but represented the HRTFs recorded on the ears' canals of a particular "good listener". The reverberation conditions in this test were natural part of the recording as opposed to [22] where the reverberation was synthesised. This made it a non-ideal scenario for judging the effect of individual HRTF recordings as the situation was not very different from that of an artificial head.

Another limitation of this experiment concerned the stimulus chosen for the experiment. Listeners were presented with 200ms long burst noises revolving twice in a circle around the head starting at a random chosen direction. Listeners had to judge perceived distance rather than directionality. Although suited for the experiment in [26], noise bursts were non-ideal for judging *realism* as they did not reflect a real-life listening situation (i.e. not an ecological stimulus); hence would not be an item suited to judge *plausibility*. On the other hand, noise bursts contained the same energy in each critical frequency band causing all the spectral cues in HRTFs available with the same perceptual weight [26].

An interesting factor in the present experiment was the use of dark-room test conditions as an attempt to cut-off the influence of the visual senses in the subjective judgement. This decision well reflected the fact that appropriate visual stimuli could not be synthesised for the audio presented in the test. It was hypothesised that the complete inhibition of the visual senses could improve the perceived *realism* of the experience and ideally clear front/back confusion; however, it was not possible to completely block the inputs to the non-auditory senses: "it is only applicable to give as little input as possible to them and to keep the conditions for the comparison as constant as possible" [26]. Darkness hence did not mean there was no visual stimulus at all, but that darkness was the only visual stimulus present and that the influence of visually perceptible sound sources could be avoided. Völk redefined the goal of virtual acoustics as "the creation of virtual events that arise in the corresponding real scene in complete darkness" [26].

The results indeed showed that a good degree of front/back externalisation can be achieved (although the experiment did not assess reversal error) by the combination of human-measured HRTFs and reverberation. It was noted that this experiment did not use head tracking technology that was normally used to clear reversal errors. In general, the results

and conclusions of this experiment mostly agreed with [22] shown in section 2.3.1. In fact, the highest degree of externalisation was found to be obtainable with human-measured reverberant HRTFs whereas the worst condition was that of artificial non-reverberant HRTFs. Figure 2.14 illustrates the results of Völk's research. Overall, anechoic conditions presented the worst results in terms of *externalisation* perception. This could be explained by the fact that anechoic conditions were quite unrealistic conditions for a real life situation. "Our hearing system acts like being in a comparative real-life situation, although listening to a virtual auditory display; [...] the little amount of diffuse energy in anechoic HRIRs occur most likely when a sound source is very close to the head" ([26]).

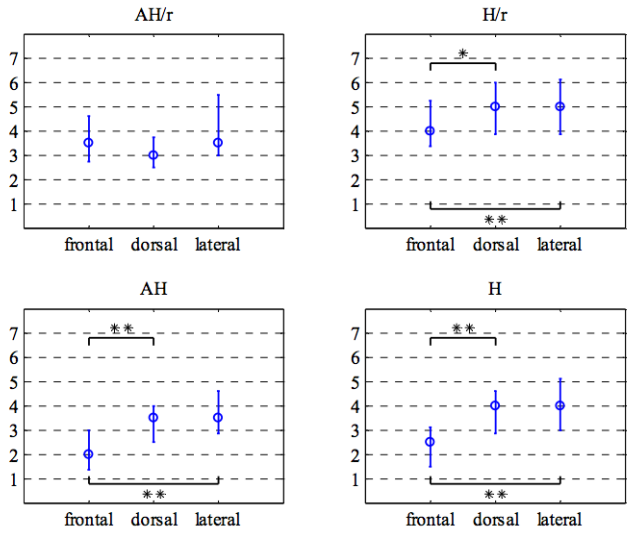


Figure 2.14: Externalisation amount in the assessment of different HRTF sets. Left column shows artificial head (AH) results while the right column shows results for the Human head (H). The index */r* indicates the presence of reverberation. [26]

2.3.3 Assessing plausibility using SDT

One of the key studies relevant for this project was conducted by Lindau and Weinzierl in 2011 ("*Assessing the plausibility of virtual acoustic environments*", [27]); the concept of *plausibility* as an assessment parameter was introduced and explored in a listening test which served as main inspiration and model for the experiment depicted in this report.

It was already explained in section 1.2.2 how *plausibility* was proposed as a more appropriate parameter than *authenticity* for assessing the realism of a virtual auditory scene. In [27], a new methodology for assessing plausibility using a YES/NO binary test paradigm was proposed. The simplicity of the assessment minimised the ambiguities of multi-facets assessments by focusing the judgement on one simple question. Easily interpretable results can hence be obtained and discussed.

The listening test placed the subjects in an auditorium room where real and simulated stimuli could be presented in the same acoustic setting. A dummy head (with headphones on) was used to record BRIRs on the horizontal plane in a reverberant room. The recorded positions corresponded to five randomly specified speaker positions in the room. The same room was used for the listening test: subjects were placed in the same spot where the dummy was used in the recording stage and five loudspeakers were placed around the listener in the same set-up used for the recording. Subjects were instructed to keep headphones on at all times and were presented with randomized real or simulated stimuli. The task was to evaluate, after each acoustic presentation, whether the presentation was real (coming from the speakers) or simulated (from headphones) using the mentioned binary Yes/No test paradigm. To prevent memory effects which could bias individuals towards one direction, stimuli were varied in content and source direction, every combination of content and source direction was presented only once in the test.

Lindau's experiment [27] made use of state-of-art head-tracking technology with low latency and time aligned BRIR interpolation that allowed free head-movement and avoided cross-fading artefacts. Spectral colouration from headphones response was flattened by applying non-individual equalisation filters called Headphones Transfer Functions (HpTF). The headphones used were electrostatic STAX SR-2050II, renowned to be acoustically relatively transparent headphones. Furthermore, al-

though no individual BRIRs were used, individuals' head width size were taken into account for individualising ITD levels. In order to achieve a statistically significant sample population, 11 participants took part at the experiment, each of them assessing 100 signals.

"A strong and inter-individually different response bias was expected due to personal theories about the credibility of virtual realities and the performance of media systems in general" [27]. A standard Signal Detection Theory approach was used to take account of the expected inter-individual response bias of the individual listeners; as a result, discriminability from an inner reference could be tested without using an objective criterion and with high confidence. The *signal trials* in this case were the simulated stimuli while the *signal plus noise trials* were represented by the real stimuli (from the loudspeakers)(see figure 2.15). Consequently, a *hit* would be represented by a correct detection of a simulation, while a *false-alarm* would be represented by a wrong perception of a real stimulus as simulated. Once again, the sensory difference of the two kinds of stimuli is represented by the distance of the two distributions' maxima and the individual response bias is reflected by individually differing response criteria λ. The *decision variable* (measured in arbitrary units) involved was the subjectively perceived level of sensation that would determine whether a simulated stimulus was "plausible" or not. A complete overlap of the two trials distributions would indicate a complete degree of plausibility.

Lindau clarified that "In terms of the SDT observer model, proving 'perfect' plausibility would require proving a sensitivity d of zero. From the view of inferential statistics, however, a direct proof of the null hypothesis is impossible. Thus, one has to draw back to rejecting a directional and specific alternative hypothesis by negating an effect that is small enough to be regarded as perceptively irrelevant (a *minimum-effect hypothesis*)" [27]. Using a two-alternative forced choice test paradigm,

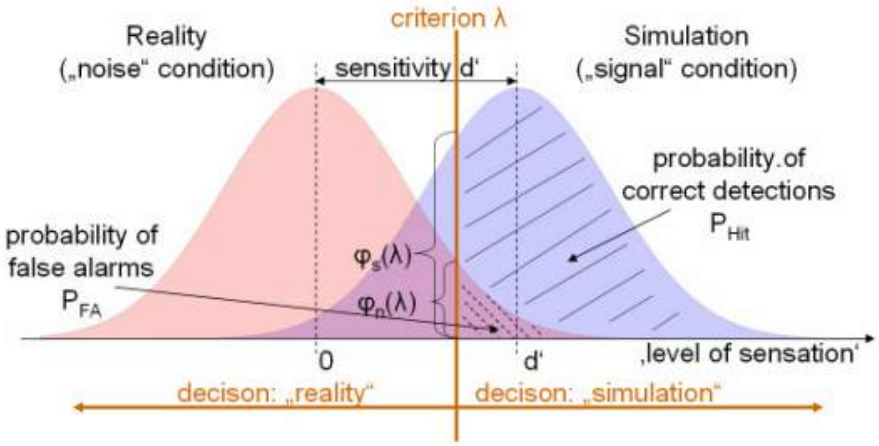


Figure 2.15: The SDT model applied to the assessment of plausibility, the *noise condition* is represented by reality, *signal condition* by simulation [27]

it was possible for Lindau to find the values that would determine the minimum effect hypothesis. For this purpose, he assumed that plausibility would be reached if $P_{Hit} = 0.5255$ which is less than 3% exceeding the pure guessing rate (which is at $P_{Hit} = 0.5$); this value was stressed to be a far stricter criterion than the ones commonly used in similar cases. A critical value of $d'_{min} = 0.84$ was calculated, as the minimum effect hypothesis to reject was $d' \leq d'_{min}$, a group sensitivity d_{avg} below this value would prove the system to be plausible.

The experiment succeeded in satisfying a strict test of plausibility and it was revealed that subjects were almost perfectly guessing, despite the use of non-individual BRIRs. Figure 2.16 shows an almost exact overlap of the probabilities of the *signal* and *noise* distributions meaning that a very high degree of plausibility was reached. Results were calculated by averaging the individual sensitivities d'_i for the whole group, d_{avg}. It was pointed out in [27] that the latest-state of art technology in head-tracking allowed the shown degree of plausibility to be reached. A previous experiment ran by Hohn & Lindau few years before ,in 2007 [50], was conducted with a less advanced system; the technical improvements of the more recent experiment were vital in improving plausibility as the previous experiment failed to perform within the critical parameters.

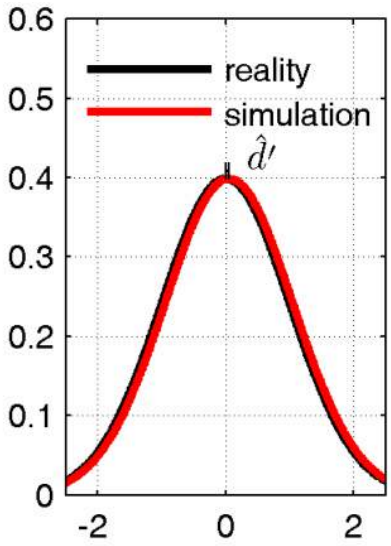


Figure 2.16: SDT analysis result in the assessment of plausibility over the group average, y axis is probability, x axis is the magnitude rate of the decision variable [27]

Lindau and Weinzierl also conclude that "spectral differences resulting in the use of non-individual acoustic simulations did not interfere with perceived plausibility. On the other hand, plausibility was shown to be sensitive to excessive latency and cross fading artefacts [caused by head-tracking real-time interpolation]" [27]. Room reverberation factors were not explored in this experiment and the question of how would different reverberation conditions affect the plausibility was not examined. It could be established whether a similar test would bring different results by involving these factors, and if a comparable degree of plausibility could be reached without the use of expensive head-tracking technology.

2.3.4 Plausibility in small room environments

A recent experiment, very similar to [27], was run by Pike et al. in 2014 [5]. The study, "*Assessing the Plausibility of Non-Individualised Dynamic Binaural Synthesis in a Small Room*", aimed to replicate Lindau's experiment in a smaller room environment with different loudspeaker positions, and analyse the results using similar parameters. This study was not yeat published at the time of the literature review stage, but it was

included due to its relevance and close relationship with the present project.

The use of a smaller room ($V = 99m^3$) was hypothesised to be more challenging to use for plausible simulation due to the proximity of the sources to the listener. The consequential smaller reverberation amount could make spatial and timbral characteristics more easy to detect [5]. Similar equipment as [27] was used, including a state-of-art head-tracking system. A set of BRIRs was recorded with a KU100 dummy head (figure 2.4) with headphones put on top of its ears. Like in [27] this was done to ensure that the headphone effects on the external sounds were simulated, since in the plausibility assessment the headphones are worn by the participant when real loudspeakers are presented. A headphone equalisation filter measured on the same dummy head was applied to the audio. It was denoted that, according to [51], when non-individual BRIRs were used, non-individual headphones compensation measured on the same head used for measuring BRIRs, enabled a more realistic binaural simulation than individual headphones compensation. The headphones equalisation filters were calculated using 10 measurements done on a dummy head, with the headphones removed and replaced each time, according to the procedure described in [52].

A very similar minimum effect hypothesis to [27] was proposed with $P_{Hit} = 0.55$ and $d'_{min} = 0.1777$ in order to compare the effects of the study. It was calculated that a minimum of 1071 samples had to be collected in order for the results to be meaningful (in [27] was 1077). This was realised by presenting 100 stimuli each to 11 subjects. The same routine as in [27] was also used, using a variety of monophonic items that included speech, orchestral ensemble recordings and individual instruments, over five randomly chosen positions. A familiarisation stage was included where listeners where presented each of the items used from a random real loudspeaker positions and subsequently from binaural virtual posi-

tions.

The results (see figure 2.17 and 2.18) demonstrated that, differently than [27] a meaningful sensory difference between the real and simulated cases was observed. Thus, plausibility had, in this case, not been reached despite individual sensitivity values were very close to d'_{min} [5] and equivalent to a detection rate $P_{Hit} = 0.6$. No individual bias was observed as possible to notice in figure 2.18.

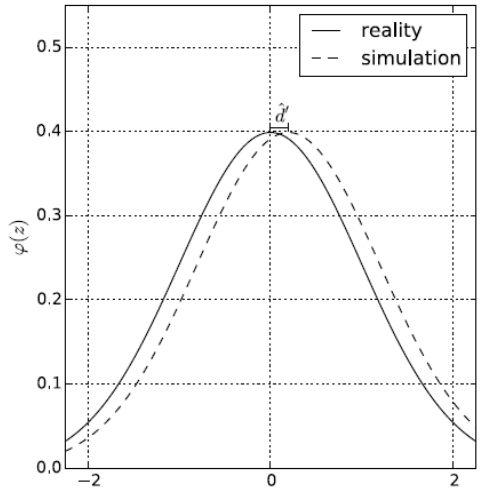


Figure 2.17: SDT analysis result in the assessment of plausibility over the group average sensitivity [5]

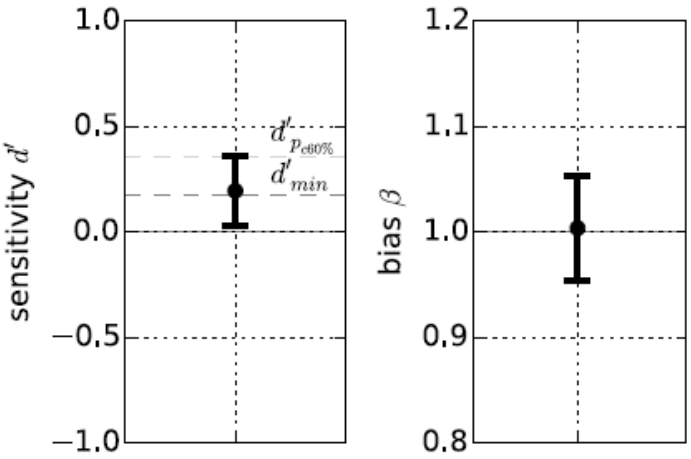


Figure 2.18: Average of individual sensitivities d'_i and biases β_i with 90% confidence intervals [5]

Pike theorised that the reasons for these results might be due to a non-perfect head tracking system which had to individualise ITD values for each participant in a way which some found quite challenging [5]. The use of a smaller reverberation environment compared to [27], different dummy head and different, closer, speaker positions, might also have had an effect. Pike stated that "It was noted by a few participants that the simulated stimuli seemed to have less precisely defined position, which could be described as a larger apparent source width or increased localisation blur. Also, for the elevated source positions, it was noted that in some instances the source location appeared lower than the visible loudspeaker. This might explain the greater sensitivity observed for elevated loudspeakers. Two participants also commented that the source direction was sometimes initially ambiguous, before moving their head, and that the perceived direction sometimes reversed after head movement" [5]. It was debated, in the conclusions, whether the familiarisation process, which was not performed by [27] might have had an influence in the results. The inclusion of the process was intended to avoid participants to wrongly interpret the acoustic effects of the headphones over external sounds to be artefacts, increasing the plausibility artificially.

Although a strict test of plausibility was not passed, informal participants' feedback portrayed the experiment as challenging, and many stated they were simply guessing [5]. As it had not been established what level of plausibility was required for what application, it might be possible to relax the minimum effect hypothesis to more lenient parameters and let this system configuration to pass the test. "Realism is often not the creative aim" [5], in those cases it might be more appropriate to simulate un-realistic sound-fields.

The environmental conditions used for this experiment, excluding the use of head tracking, were very similar to the conditions of one of the environments used in the project experiment. It would therefore be appro-

priate to directly compare the project's results with the results in [5] and assess the changes brought by individual BRIR measurements. The other project's environment yielded different reverberation conditions which also could be assessed.

2.4 Experiment Design

Initial decisions about the direction of the project were agreed with Chris Pike from BBC R&D department in Media City UK, Salford. This project did benefit from the collaboration in terms of advice and resources, in turn it was intended to help the research conducted by Pike ([5]) with ulterior insights to analyse.

Following the study of the past research done in the field, and considering the findings of each of them, it was decided that a similar procedure to relevant past experiments for the objective of a plausibility assessment could be reached. In particular, the procedure of assessment defined by Lindau [27], was taken as a model for expanding the test of plausibility to different reverberation environments, still unexplored in the context of a plausibility assessment with a SDT analysis approach. An interesting addition to the experiment was to consider the use of individual HRTFs measurements, so far not considered to be determinant for a plausible environment. Several initial design constraints were taken into account, mainly, the impractical difficulty to set-up a head-tracking system from scratch and the non-availability of acoustically transparent STAX headphones.

Previous experiments suggested that reverberation was more influential than individualisation [22], and that anechoic conditions [26] brought less "externalisation" than reverberant conditions. Both these assumptions could be further verified in the context of plausibility assessment.

The initial specifications, described in the Literature Review report handed in December/2013, were subject to several changes after discussion with the project supervisor about the feasibility of some of the aspects of the experiment. Here, it is depicted the final approved design before moving on to the next stage; section 2.4.6 recalls the initial design and the major differences between the initial and final versions of the experiment. Further minor changes had to be applied in the later stages as a result of problems encountered and unexpected limitations, those changes would be discussed later on in the report.

The project was broken down in stages and substages, each main stage is represented by a dedicated chapter that covers the design process for that particular stage with more details.

2.4.1 Requirements

In order to fulfil the objectives stated in section 1.3.1 the following requirements were taken as essential for the project:

1. The sampe experiment procedures of [27] and [5] will have to be adapted and reproduced
2. The experiment routine will be run in two environments that differ in reverberation characteristics, using the same participants
3. The same identical set-up will have to be reproduced in both rooms for direct comparison
4. The number of participants must be chosen to be high enough to provide a meaningful amount of data for the analysis
5. Binaural virtual content tailored for each individual participant will have to be produced for the listening test

6. Due to the absence of head-tracking technology, participants will need to keep the head still in a fix position throughout the experiment. Failure to do so may result in an immediate detection of simulated stimuli as soon as the head rotates

7. Real and simulated stimuli will be assessed in terms of plausiblity

8. Signal Detection Theory will have to be used to analyse the results and measure the degree of plausibility evaluated through a Yes/No task

9. Results will have to be compared to previous study in order to derive meaningful conclusions

2.4.2 Breakdown of stages

After having established the requirements four main stages, dividable into sub-stages were identified and planned:

- **Stage 1** - Preparation of equipment
- **Stage 2** - HRTF measurements
- **Stage 3** - Listening Test
- **Stage 4** - Analysis of results

The breakdown of the project into four stages helped to define and clarify the exact order of tasks necessary to complete the project, and focus the attention and effort in an organised way.

2.4.3 Specifications

More detailed specifications were defined for each stage of the project. This process took into account the equipment available for the experiment:

- Stage 1

- The two environments have to be placed in the University of York for practicality:
  - Anechoic environment
  - Echoic reverberant environment
- The set-up will use 2 speaker positions to present the real stimuli:
  - One loudspeaker placed at the front, at zero elevation at convenient distance
  - The other placed in a rear point, at a convenient position
  - The chosen positions and distances have to be replicated and maintained in both environments
  - A listening point has to be set in a convenient position and maintained in both rooms
- A special experiment chair needs to be constructed:
  - Comfortable headrest to avoid tiredness
  - Possibility to fix participants' head in position
  - Avoid any block of direct sound path
  - Minimise the reflections emitted
  - Possibility to transport it between the two environments
- Measurement equipment:

  - 2x Electret Microphone Capsules
  - Microphones sponge encapsulation for each participant

- Measurements microphones need to be powered by a phantom power supply:
  - Circuit needs to be designed and constructed
  - Powered by 9V battery
  - 2 channels
  - XLR input and output sockets

- Other equipment, chosen based on availability, will have to cover the following:
  - 4x loudspeakers of same model/brand
  - 2x two-channels amplifier
  - Soundcard with customisable sample rate
  - DAW
  - Headphones
  - Loudspeakers' mounting equipment suitable for both rooms

- Stage 2

- Number of participants has to be high enough to obtain meaningful amount of data
  - Balance between males and females

- Free-Field recordings have to be performed for all the speakers
  - Channels have to be calibrated for equal gain
  - Free-field recordings must be inspected to find out eventual inconsistencies in channels gain parameters

- HRIR Measurements have to be performed for both environments
  - HRIRs recorded for the two speaker positions in each environment
  - Sine-sweep technique [41]
  - Participants need to keep the head still during recording
  - Headphones have to be worn by the participant, on top of the microphones, during the recording
  - Recordings have to be processed to obtain the individual HRTF pairs for both positions, in both rooms
- Headphones Impulse Response (HpIR) need to be recorded individually for each participants
  - Sine-sweep technique [41]
  - For left and right headphone output
  - Recordings have to be processed to obtain the equalisation filters in the way described by [52]
  - HpTFs compensation filters have to be applied to the related HRTF pairs
- Filters must be inspected and screened for anomalities
  - Central position ITD must be zero
  - Central position ILD has to be corrected using the free-field ILD
  - Eventual corrections must apply

- Stage 3

- Original items used to create stimuli have to be dry and monophonic

- Stimuli chosen will have to be varied in content and be ecologically viable
  - Male and Female speech
  - Orchestra Ensembles
  - Ambient sounds
  - Single Instruments
- Each participant will have the virtual stimuli created with his own HRTFs
  - Same environments and set-up of stage 2 has to be kept
  - Same listening point of stage 2
  - Head position has to be kept fix
  - Stimuli has to be created independently for both room environments
- Loudspeakers volume gains will have to be adjusted to have the same perceptual volume as the headphones' virtual audio
- Listening Test routine as in [27]:
  - Familiarisation process
  - One presentation at the time, avoid repetition of signal items
  - Randomised content between real/simulated
  - Avoid of direct comparison
  - Yes/No binary decision
- Listening Test has to be repeated by each participant in both rooms

Stage 4

- Results have to be formatted for the SDT analysis code (Provided by Chris Pike, BBC R & D):

  - Signal Item
  - Speaker Position
  - Real/Simulated
  - Yes/No answer

- The analysis has to be run separately for the two listening environments
  - Further investigation has to be directed for position-dependency and signal-dependency

- Conclusions have to be compared with past experiments
  - SDT analysis result have to be compared with that of [27] and [5]
  - Observations have to be compared with the findings of [22] and [26]

2.4.4 Project Management

Having a set of specifications ready allowed for the further breakdown of the main stages into substages for project management purposes. Image 2.19 shows the resulting flow chart representative of this project.

2.4.5 Listening Test Design

The absence of head-tracking technology meant that subjects were required to keep the head fixed in a steady position. It was decided that a specific chair had to be built in order to have the listeners still, a comfortable headrest could serve for the purpose. It was also determined to use a headstrap that would encompass the listener's head and the headrest. At the same time it was necessary to minimise the amount of reflections

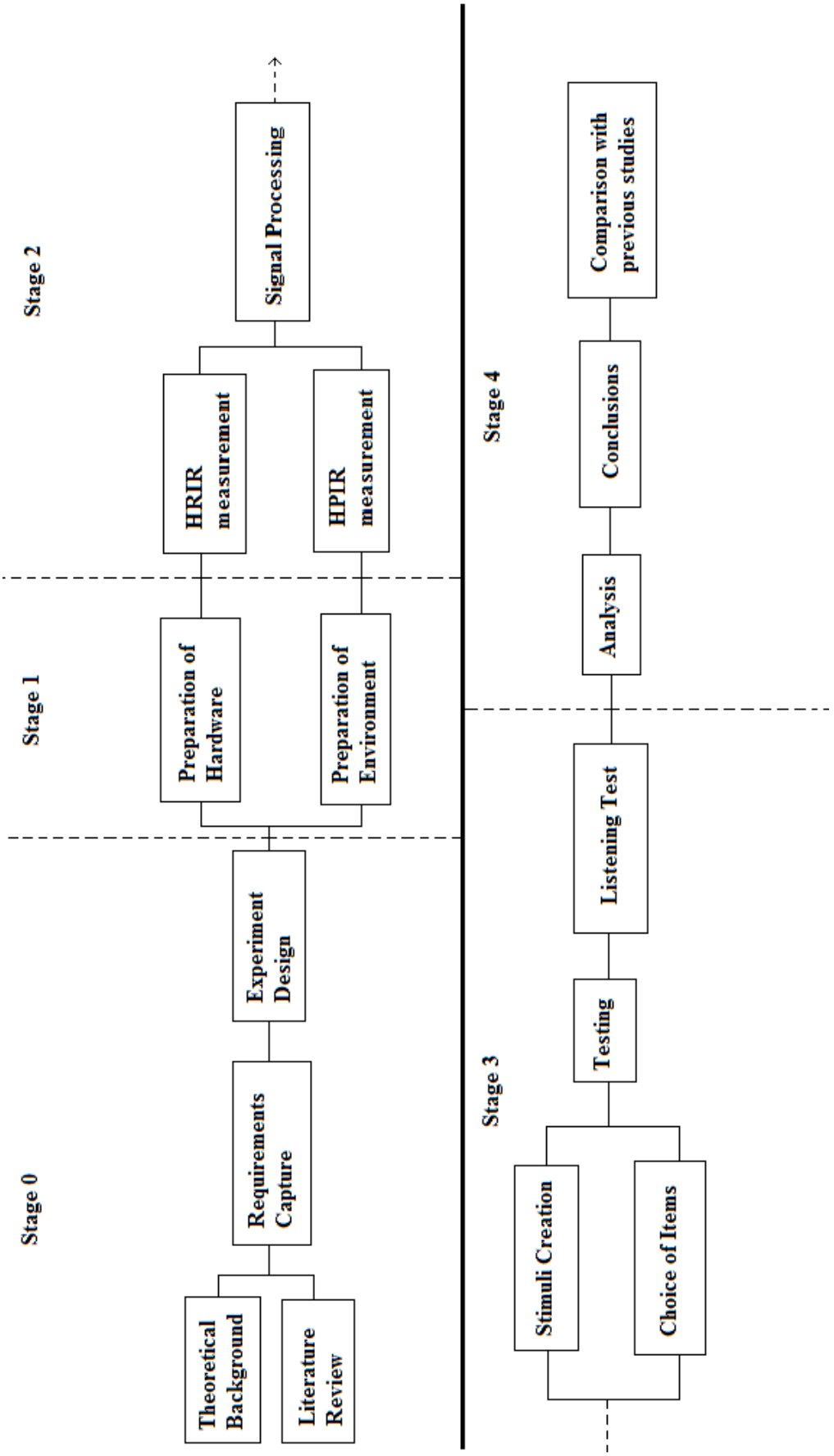


Figure 2.19: Project Flow Chart

produced in the anechoic chamber environment; hence the surface area of the modifications would have to be kept as small as possible. Placing the chair in the anechoic chamber's grid required the use of a support where the chair could stand; it was found that a wooden board would have been appropriate for this task although non-ideal due to the reflective surface. A later change from this initial design was to include a blindfold due to the reasons explained in chapter 3.

To satisfy the specifications, the main important concern was to design a set-up which could have been reproducible in both room environments. Anechoic chambers often have the costraint of having a metallic suspended grid instead of a floor, to minimise reflections. This aspect influenced the choice of where to place the speakers; no more than 2 positions were necessary for the purposes of this project. The most critical position for virtual sound localisation is the front [22], hence one of the positions was decided to be at 0° azimuth angle and 0° elevation. The other position did not have to be chosen before inspecting the rooms used for the test. The important thing at that stage was to choose a position which would have been practical to reproduce in both rooms and easy to maintain in place. It was established that the ideal case would have been to find a lateral rear position at a non-zero elevation, for the sake of variability.

An early concept of the set-up is shown in figure 2.20, distances *d1* and *d2* are not defined at this stage such as the position of speaker #2:

The listening test routine was agreed with the supervisor to be adapted from the routine of [27]. Listeners would have been sitting at the centre of the room on the special chair, with their head-position fixed and wearing headphones. A training process would have been necessary to clear any ambiguities with the assessment and avoids participants to wrongly interpret the acoustic effects of the headphones over external sounds to be artefacts. Contrary to [5], the process would have to be conducted with a signal item not present in the actual listening test, in order to

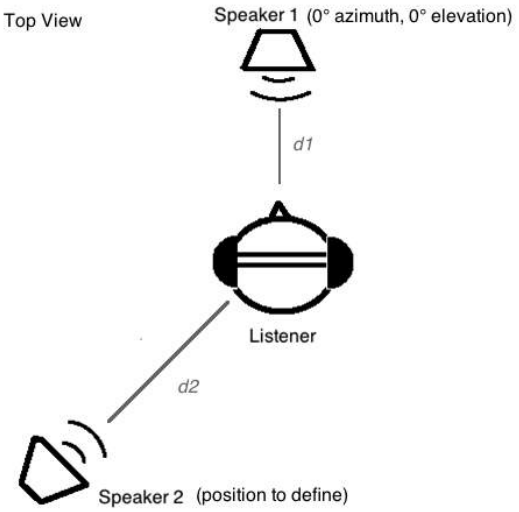


Figure 2.20: Conceptual experiment set-up

avoid a memory effect. The signals would have to be fed in a random way between speakers and headphones for the two positions. To further reduce memory effects and immediate comparison of stimuli (suitable for an assessment of *authenticity*) each presented signal item would not have been repeated, furthermore it would not have been re-presented for the same spatial position whether real or simulated although it could have been re-presented for the other position. After each presentation, a yes/no decision had to be made by the subject. To simplify this decision, the *yes* or *no* answers were transmuted to *simulated* and *real* in respective order. This change was meant to reduce the chances of misunderstandings and confusions, by making the decision paradigm more straightforward for the participants. Each answer would then have to be manually recorded by the experimenter on a laptop before going on to the next presentation. The same identical procedure would have to be repeated in the other listening environment using the binaural material created specifically for that subject and that environment. Figure 2.21 shows a high-level flowchart design for the listening test used as starting point in the development of the routine.

Initially it was thought that having the possibility to see the loudspeakers

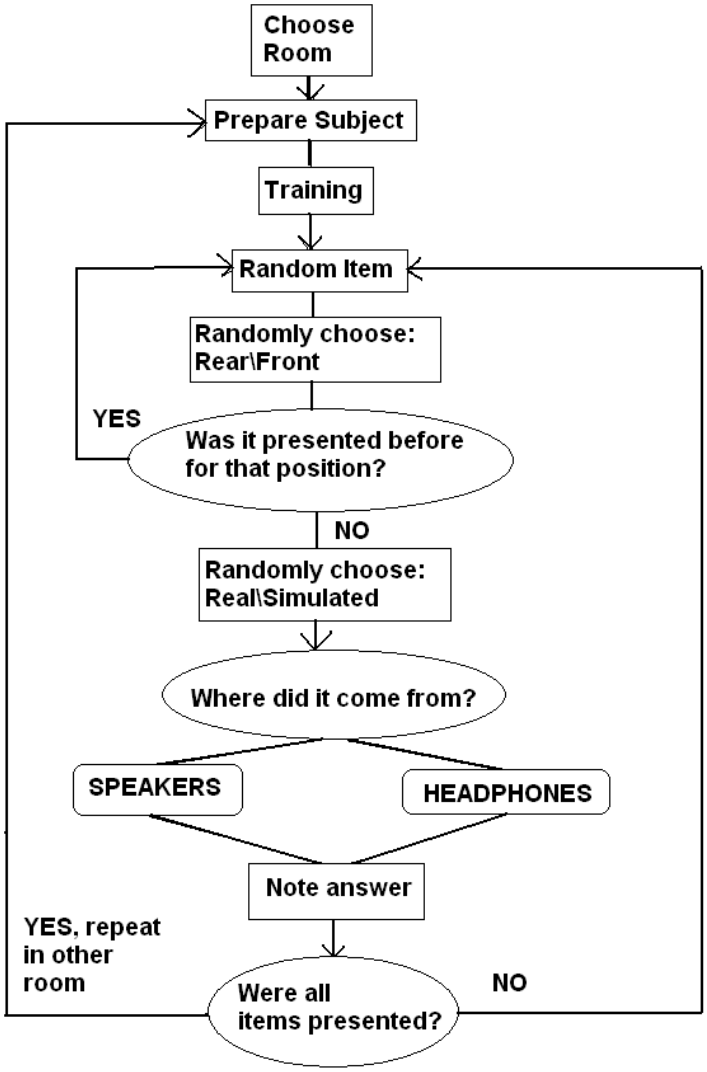


Figure 2.21: Routine design flow chart

during the test would have improved the chances of having a plausible auditory scene as this would have "tricked" the brain into processing the presence of a plausible sound source in front of it. It was later considered that slight inaccuracies in the spatial localisation of frontal binaural audio might have benefited from darkness conditions, hence, it was decided that a blindfold would have been applied after the training session where the participants could see the speaker positions and roughly memorise them.

Lindau [27] and Pike [5] found a minimum required number of samples to be around 1077. They both used 11 subjects, each of one made 100 decision, leading to a total number of samples of 1100. At this point of the experiment it could not be predicted how many subjects would have been possible to recruit for the experiment but was considered possible to tailor the number of stimuli presentations for the amount of participants and therefore reach the amount of 1100 data samples.

In summary, the variable between experiment sessions was the reverberation profile of the room environment. Individualisation was a fix variable. Secondary dimensions to be inspected were source position and stimuli-type.

2.4.6 Differences with initial plan

A number of changes from the initial design [53] were made. Main reasons for the changes were: impracticality of the proposal, unavailability of equipment and difficulty of the operation.

An initial number of speakers was suggested to be five, as that is a common number for home-cinema systems. It was established that the number of speakers was not, in fact, essential and the restraints imposed by the difficulty of setting up the environment in the anechoic chamber led

to a decision to downscale the number of speakers to two. Analysis of the influence of speaker position became a secondary objective for this project and a focus on reverberation factors reduced the number of dimensions explored for the sake of practicality. A further reduction of dimensions was achieved by rejecting the initial consideration of using non-individual HRTFs to be compared against individual HRTFs; individualisation was changed to be a fix variable in order to keep the experiment simple and more focused on one single aspect. It was also suggested that source localisation and externalisation could be assessed; however, this option would have also increased the complexity of the experiment to a point where the feasibility in the given time could be compromised.

The use of Headphones equalisation was previously stated as "optional". The equalisation process was upgraded to the status of "essential" as the improvements brought by applying equalisation could have a big impact on the perception of plausibility. The headphones used were initially intended to be STAX headphones, as used in [27] and [5]. The unavailability of this equipment item and its high cost (impossible to cover with the given project budget) led to a second choice of headphones based on what was available in the department; the choice fell on the DT-990-PRO by Beyerdynamic®.

An initial proposition of using darkness conditions for the experiment was at a certain point considered unnecessary and scraped off the design, but later reconsidered in the implementation stage for the reasons explained in chapter 5. The decision to remove the darkness condition derived from the fact that having the subject see a source in front of himself could provide a better visual stimuli than darkness; thus, improving plausibility.

At the time, Analysis of Variance (ANOVA) was proposed as analysis approach. However, for the reasons previously stated in this chapter, SDT was considered the best choice for this experiment as its appropriate for

binary tasks and for discriminating sensorial difference from individual bias.

Chapter 3

Equipment and Facilities

Contents

3.1 Facilities 81
3.1.1 Environment Set-up 83
3.2 List of Equipment 89
3.2.1 Experiment Chair 92
3.2.2 Power supply unit 93
3.3 Problems Encountered 102

This section covers the first stage of the project which involved the research for suitable facilities, the environmental set-up and the preparation of the hardware equipment used for the project. Some parts of this stage, the experiment chair and especially the phantom power supply unit, proved to be particularly difficult and lengthy tasks. Many versions had to be created and tested before the specifications could be met.

This chapter includes detailed description of all the sub-stages, including set-up schematics, a list of equipment and pictures of the equipment and the listening environments.

3.1 Facilities

The experiment environments were chosen between those available in the *Department of Electronics* at the *University of York*. Two different reverberant conditions had to be assessed, a small reverberant room and, ideally, an anechoic room. The *Audiolab* building, *Genesis 6*, is provided with a full anechoic chamber available to students for booking. This environment was also practical due to proximity to the equipment storage room. Figure 3.1 shows the external view of the chamber, figure 3.2 shows a diagram of the room dimensions: the chamber volume was $V = 3.5^3 = 42.875m^3$; a suspended floor grid (used to allow absorption wedges to block floor reflections) is placed at 0.7 meters from the bottom. Anechoic conditions (RT60 = 0) have to be assumed although some external low-frequency noise, produced from a nearby facility of the chemistry department could not be fully blocked and reflections would have been produced from the bouncing of sound on the speakers and the amplification equipment. Nevertheless these aspects were judged to be small enough to be negligible for this experiment, which did not require fully optimal anechoic conditions.

Figure 3.1: External of the anechoic chamber, Genesis 6, *University of York*

The choice of the other room had to fall on a reverberant small room en-

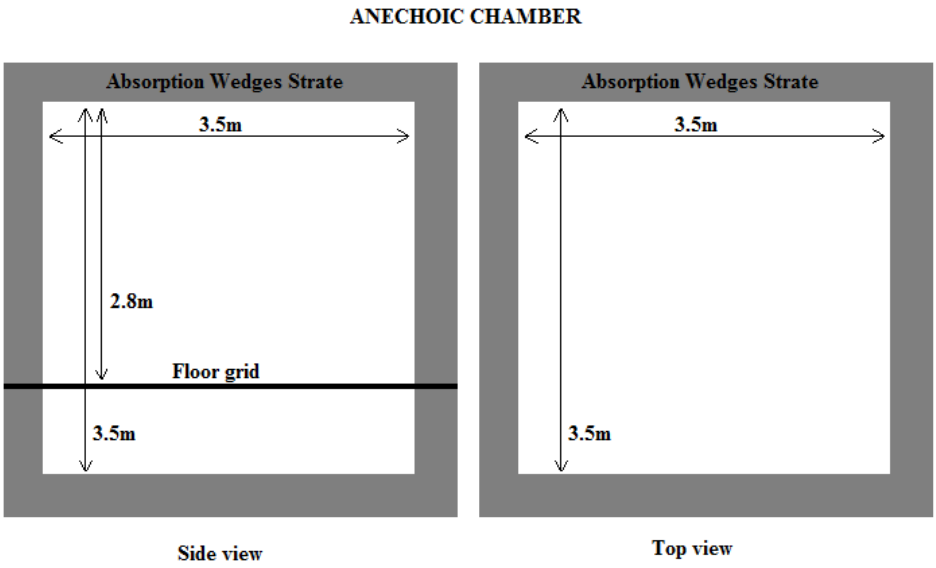


Figure 3.2: Anechoic chamber dimensions

vironment, similar to that of [5]. The dimensions had to be big enough to allow the set-up to be first implemented in the anechoic chamber (which has more logistical constraints), and subsequently reproduced in the room. Other criteria used to choose the room was proximity to the chamber, proximity to equipment and availability to book it without risk of clashing with other projects or lectures. The choice fell on the listening room of the *Genesis 6* building (figure 3.3), used for research purpose, listening test and PhD work. This room is very near the anechoic chamber and usually not very requested for bookings, making it available for planning sessions last-minute in emergency cases. A *Google Calendar* system managed by the facilities technician, Andrew Chadwick, was integrated with an email address so that available times could be easily checked and bookings could be made online for both the chamber and the listening room. As possible to see from the schematic in figure 3.4, the listening space in the chosen room occupies only a portion of the total room, the listening space could be isolated by curtains for a total listening space volume of $V = 5.05 * 4.7 * 2.6 = 61.711m^3$. The walls and floor were made of different reflective materials and some reflections also were pro-

duced by objects in the room that could not be removed (the surround system speakers and piano). It was revealed, by the room design plans, that immediately above the ceiling there was a void space of approximately 0.7 meters. It is not clear whether this void space would have affected reverberation and at which frequency bands, whatever the effect, it did not matter in this experiment as long as the room used remained the same throughout the project as it was indeed the case. Due to these factors, it was decided that the reverberation time RT60 of the room had to be calculated using an empirical approach: an impulse response was measured using the methods of section 4.1.2, using a dedicated software (*AURORA*), the RT60 was calculated across frequency bands and found to be $150ms$ at $1kHz$.

Figure 3.3: Chosen reverberant room environment: Listening Test Room, Genesis 6, *University of York*

3.1.1 Environment Set-up

The next logical step in the first stage of the project was to define the speaker positions to be used in the set-up, measure and calculate the positions and the angles and then replicate the set-up in both the environ-

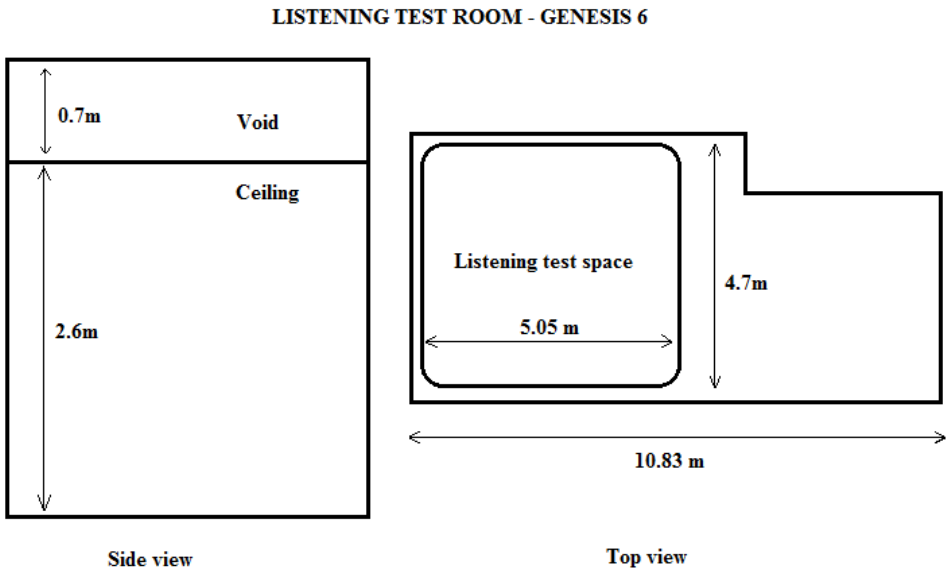


Figure 3.4: Listening room dimensions

ments. The chosen approach was to find a suitable set-up in the anechoic chamber first due to the difficult logistical aspect of setting up equipment on the suspended grid. Afterwards it would have been possible to measure the distances between the listener's position and the speakers, calculate the angles, and replicate the set-up in the bigger listening room. It was suggested by the project supervisor to avoid placing the loudspeaker stand-poles in the chamber and instead set up metallic wires across the chamber where the loudspeakers could be hanged to. As it was considered important to have a central loudspeaker position in the front, the first wire was set up, with the help of the technician, to run across the room parallel to the back wall. A second wire was found to be convenient to place on the left side of the chamber, perpendicular to the first wire. Both wires were set to a height of approximately 2 meters from the grid.

The loudspeakers to be used were chosen according to availability. Four *KEF-HTS3001* speakers were made available for this project (figure 3.9), two for each room, this specific model proved to be particularly well suited and flexible enough to be hung by the wires. Two adjustable wire supports were constructed for the two speakers designated to be set-up

in the chamber, figure 3.5 shows how the hanging procedure was carried out. The first speaker was placed centrally by the first wire to point at the centre of the room. The height was adjusted to be approximately as tall as the position of the listeners' head when sitting on a chair. The second speaker was put along the wire on a random position at the rear left in comparison to the listening point, the elevation also was chosen arbitrarily according to what was convenient. Once this position had been set, both loudspeakers placements on the wires were marked with tape so that the same set-up could always be reconstructed. To calculate the distances, the experiment chair described in section 3.2.1, was placed in the middle of the room where the listening point was intended to be, facing the first loudspeaker.

Distances from the chair and the floor grid where measured using a measure tape, the same was then done for the other loudspeaker. To calculate the angles, simple geometric calculations based on trigonometry theorems [54] were applied: the perpendicular distance from the chair to the wall and from the chair to the speaker allowed to calculate the azimuth angle, the height difference between the loudspeaker and the chair's headrest together with the direct measured distance of the chair to the speaker allowed to calculate the elevation angle. These operations are represented by the superimposed geometrical figures in the picture of image 3.6. The angles were calculated in the following way [54] using the distances between objects:

$$\theta = tan^{-1}(\frac{opposite}{adjacent})$$

Table 3.1 summarises the measured distances (in centimetres) and calculated angles for both speaker positions, figure 3.7 shows a sketch of the final set-up. The accuracy of these measures, especially elevation angle, depend on the height of the participant, however this concern had no influence on the experiment as long as the configuration was kept intact

and identical for all the participants.

Figure 3.5: Details of one of the loudspeakers hanging from the wire

Table 3.1: Speaker Positions

Speaker Number	Distance from chair	Distance from floor	Azimuth	Elevation
1	141cm	120cm	$0°$	$0°$
2	175cm	185cm	$-126°$	$15°$

After having established the details of the loudspeaker placements, the other room environment could be constructed. It was agreed with the technician that it would have been possible to leave the set-up in the chamber untouched for the totality of the project period. However the same thing could not have been promised for the listening room. For this reason the set-up was prepared to be easily mounted and dismounted. A listener position was defined and marked on the floor, then the speakers were placed on top of adjustable stands at the defined distances and elevations. Figure 3.8 shows the complete set-up for the listening room.

Appendix B presents further pictures of the listening environments and other equipment prepared at this stage.

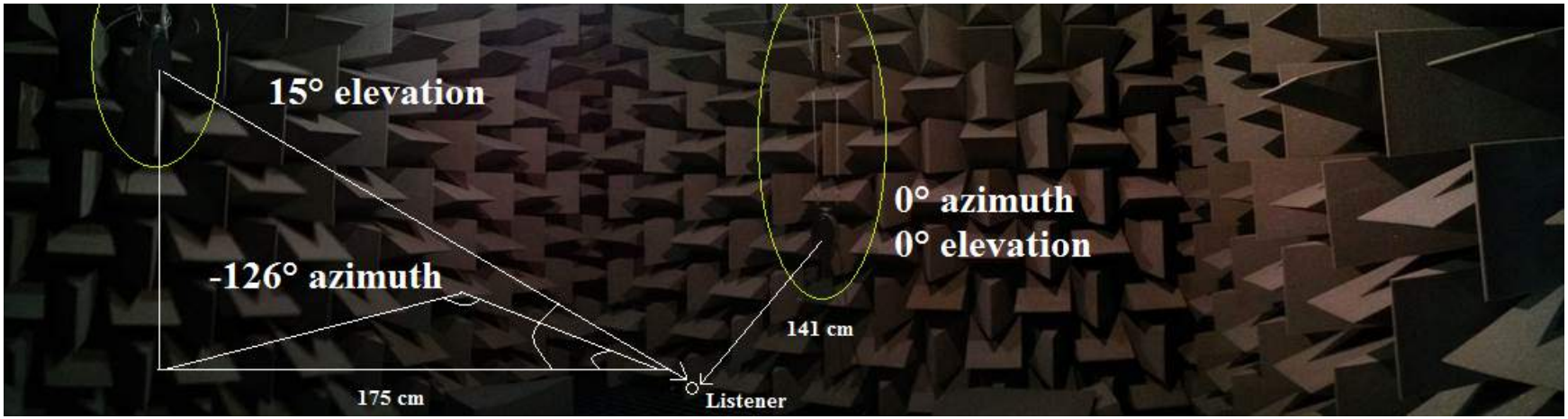


Figure 3.6: Picture of the set-up in the anechoic room, visualisation of geometric calculations of angles and distances are superimposed

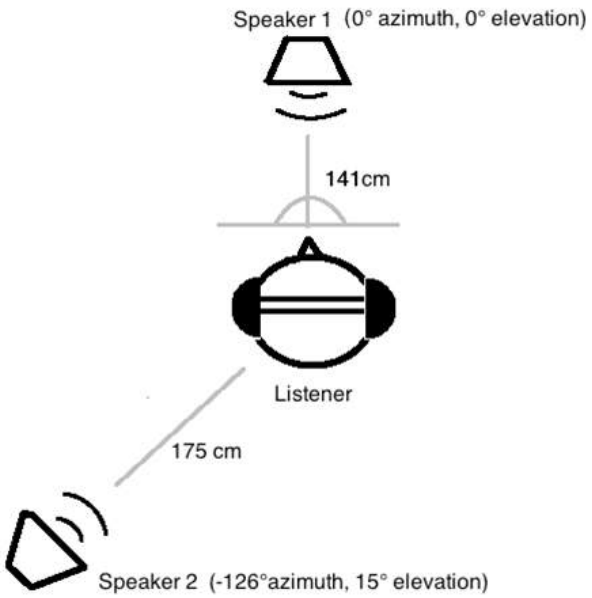


Figure 3.7: Details of the experiment set-up chosen for both environments

Figure 3.8: Picture of the set-up in the listening room

3.2 List of Equipment

In order to be able to quickly switch each participant between measurement session, some of the equipment used for the project was doubled in number as to ensure that the two environments were ready for simultaneous use.Time efficiency was thus improved for both measurement and listening test stages. Table 3.2 shows the list of all the hardware and software used for the project, indicating the stage for which they have been used. Furthermore, this section includes pictures of some of the equipment used.

Table 3.2: Equipment Used

Item	Quantity	Brand	Model	Stage
Loudspeakers	4	KEF	HTS3001	2,3
USB Soundcard	2	Focusrite	Scarlett 2i4	2
USB Soundcard	1	MOTU	Ultralite MK3	2,3
Amplifier	2	Behringer	A500	2,3
Headphones	1	Beyerdynamic	DT-990-PRO	2,3
3-Pin Electret Capsules	4	Senheiser	KE4	2
Capsules' Ear Sponges	100	-	-	2
Phantom Power Supply	1	-	-	2
XLR-LEMO cable	2	-	-	2
XLR cable	4	-	-	2
Jack cable	4	-	-	2,3
Reference Speaker Cable	4	-	-	2,3
Experiment Chair	1	-	-	2,3
Head-strap	1	Velcro	VEL60327	2,3
Blindfold	1	TFY	Eye Mask	3
Surgical Tape	1	Micropore	25mmx5m	2
Measure Tape	1	Silverline	Chunky Tape	1,2,3
Laptop	1	Apple	MacbookPro	2,3,4
DAW	1	Cockos	Reaper	2
Signal Proc. Software	1	Mathworks	MATLAB	2,3,4

Most of the equipment used for this project, such as the amplifiers (figure 3.12), were chosen based on its subject to availability in the department, the most acoustically transparent model of headphones available was chosen (figure 3.11). The microphones capsules used in this project were also chosen according to availability. Sponge covers for the electret mi-

crophones were prepared in abundance as they had to be changed every time the capsules were inserted in the ear canals of a participant for hygiene purposes. The ear-sponges were obtained from standard sponge ear-plugs by the department technicians (figure 3.13). The initial soundcard model used for the HRTF measurements (*Focusrite - Scarlett 2i4*), was later changed to another model (*MOTU Ultralite mk3*, figure 3.10) for the HPTF measurements and the listening test, due to the reasons explained in chapter 4.

Figure 3.9: Loudspeaker model - HTS3001 by KEF

Figure 3.10: Ultralite MK3 soundcard by MOTU

Figure 3.11: DT-990-Pro headphones by Beyerdynamics

Figure 3.12: Reference Amplifier A500 by Behringer

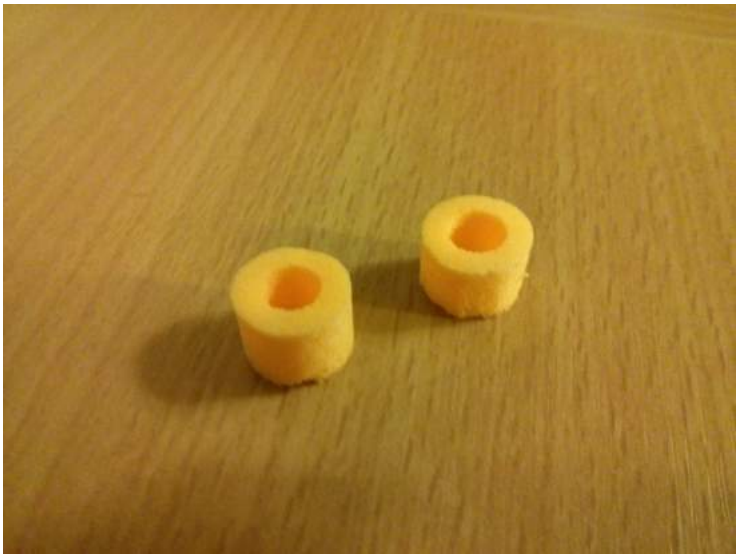


Figure 3.13: Sponge covers for the microphone capsules

3.2.1 Experiment Chair

Due to the absence of head-tracking technology (too complex to implement for the purpose of this project), it was determined that participants would be required to have their heads kept still in a central fixed position in both the measurement stage and experiment stage. It was decided that the most appropriate way to do that would have been to use a special modified chair that would allow participants to rest their head on a headrest, where it could be fixed with the use of a velcro strap. The chair had to minimise the amount of reflections and, more importantly avoid any blockage of the direct path between real sound sources and listener's ears. For this reason it was more practical to modify an existing chair instead of choosing one already provided with a headrest.

The chair specifications where the following:

- No blockage of direct sound path from real sound sources to the ear
- Headrest to fix the listener's head with a strap
- Comfortable position to prevent participants feeling tired during the listening test

- Miminised amount of reflections
- Easily transportable between the two environments

A standard non-mobile chair from the *Genesis 6* labs was selected to be modified. The chair (figure 3.14) was designed to have a central pole, fixed to the back of the chair, with the headrest placed at its end. To fix the pole, a wooden support (figure 3.16) was constructed with the help of Andrew Chadwick and drilled to the legs of the chair. The support was designed to be distant enough to allow the pole to run up straight without having to follow the oblique profile of the chair backrest (figure 3.15). Another reason to use the wooden support was its possibility to be taken off the chair easily and remounted on another chair, in case the first one broke. The headrest (figure 3.17) was designed to encourage listeners to adopt a straight position and also to be adjustable so that the height of the participants could be taken into account, thus avoiding slouching or uncomfortable sitting position. A velcro strap was selected to serve as head-strap in order to fasten the head to the headrest (figure 3.18). The chair was light enough to be transported between measurements and experiment sessions. To allow the chair to stand on the anechoic chamber grid, a wood support platform was placed beneath the chair (figure 3.19). The exact position of the wooden board on the grid, and the position of the chair on the board, were marked with adhesive tape in order to be able to reproduce the positions accurately. In the listening room, adhesive was put at measured distances on the floor to mark the chair-legs position and have a prompt reference for rebuilding the environment every time it was necessary.

3.2.2 Power supply unit

Most condenser microphones need a phantom power supply of around 48V, these units are widely available on the market for purchase. Electret

Figure 3.14: Finished version of the experiment chair - Front view

Figure 3.15: Finished version of the experiment chair - Back view

Figure 3.16: Details of the chair's wooden pole support

Figure 3.17: Details of the adjustable headrest

Figure 3.18: Head-strap lace and blindfold used for the project

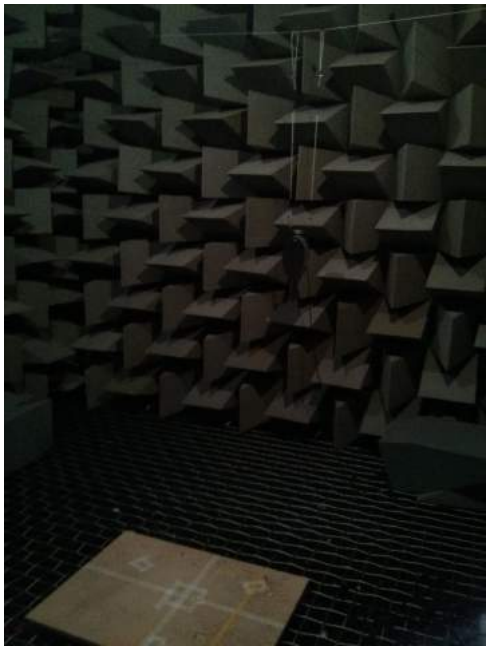


Figure 3.19: Details of the wooden platform suppor

Figure 3.20: 3-Pin Electret microphone capsules used for testing

condenser microphones, however, need a smaller amount of operating voltage (0.9V to 15 V according to the data sheet of appendix C), making the unit very rare in the market due to the non-commercial use of these type of microphones. It was necessary to design and build a phantom power supply unit from scratch, in order to be able to perform the individual HRTF measurements. The circuit then had to be encased to ensure stability and protection, while making it easy to transport.

The electret microphone capsules available were 3-pin back-electret models produced by Senheiser. The 3 pins (figure 3.21) represent the output AC signal, the DC current input, and the ground reference connection respectively. An electret microphone is made of a Field-effect Transistor (FET) transistor which has the advantage of being low-noise at the cost of being highly susceptible to overload charges. The main feature of electret microphones is the elimination of the need for a polarised power supply due to the use of stable dielectric material *electret* which is permanently embedded with static electric charge. The data-sheet included in appendix C highlights the high signal-to-noise ratio and the wide frequency response of these particular capsules. Back-electret capsules are suitable for transducer configurations, the pre-amp circuit suitable for these capsules is shown in figure 3.22, the signal is taken between the output pin (pin 1) and the ground (pin 3), the power is fed to via pin 2.

The resistor sets the gain and ouput impedance and the capacitor blocks the DC current.

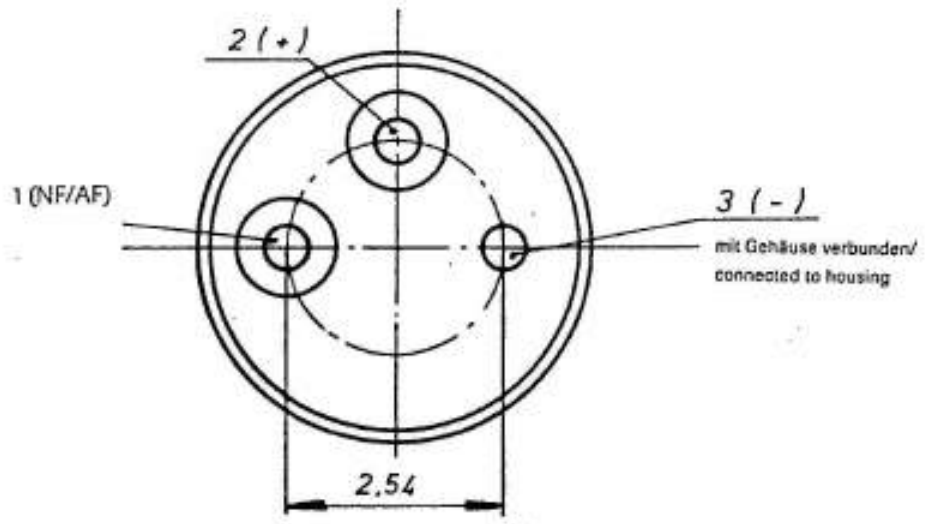


Figure 3.21: Detail of the pin connections of the electret capsules used [55]. 1 = output signal, 2 = VCC+, 3 = Ground

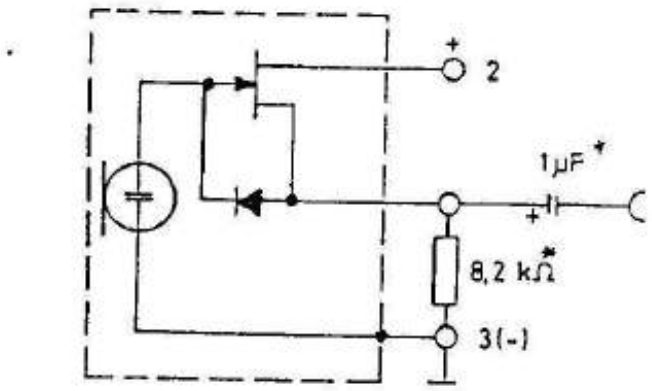


Figure 3.22: Typical pre-amp configuration [55]

The unit specifications were the following:

- Use of a battery
- 2 independent channels
- Feed 0.9 to 15 Volts and 150 uAmps to each channel

- Block the DC current from the output signal
- XLR inputs for the soundcard
- XLR outputs for the microphones cables

The final design was built to meet the specifications using the lowest amount of circuit complexity possible. The capsules specifications allow a voltage as little as 0.9 V to be enough to drive the current trough the capsules. The most practical option was to use a 9V battery and a potential divider to split the voltage between the two channels; $10k\Omega$ resistors (R1, and R2) split the voltage equally to 4.5 Volts per channel. C1 and C2 were used to isolate the channels from shared connections and avoid crosstalk effects. Values for R3,R4,C3,C4 were chosen for the nearest available values from the circuit shown in figure 3.22, once again, the resistor was used to match the impedances and the capacitor blocked the DC current from the 9V battery. R5 and R6 were used to keep the capacitors correctly polarised; the value was suggested by the supervisor to be high enough to allow all frequencies pass through. Table 3.3 enlists the components needed for implementing the design shown in the schematic of figure 3.23.

Table 3.3: List of components

Label	Type	Value
MIC1, MIC2	Electret Condenser Microphones	-
R1, R2, R3, R4	Resistor	$10k\Omega$
R5, R6	Resistor	$1M\Omega$
C1, C2	Electrolytic Capacitor	$10\mu F$
C3, C4	Tantalum Capacitor	$6\mu 8F$
-	Prototype Board	-
-	2x XLR Female Sockets	-
-	2x XLR Male Sockets	-
-	Battery	$9V$

Two specific electret microphone capsules were used for testing the measurement equipment (figure 3.20) while two different capsules were used for the actual measurement stage described in chapter 4. The reason for

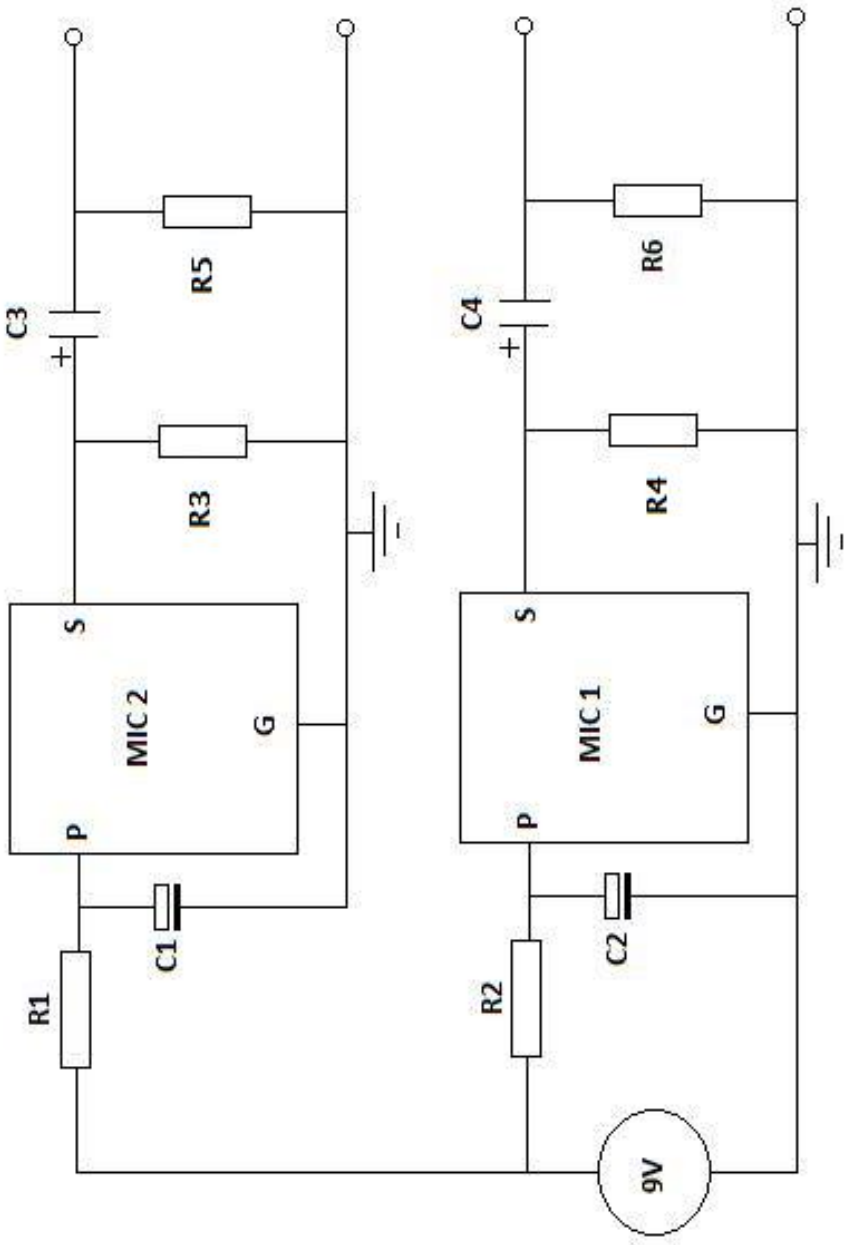


Figure 3.23: Schematic for the phantom power supply circuit

using different capsules was that the capsules used for the measurements were bent in a specific way which made them suitable for being inserted in the ear canals. To avoid the risk of possibly overcharging the only available pair of bent capsules and blowing them up by feeding too much voltage, the other un-bent capsules (figure 3.20) had to be used when testing the unit. The circuit was first built on a breadboard and tested before being soldered to a prototype board. The testing process involved the connection of the microphones to the phantom power supply circuit with a 9V battery connected. An oscilloscope was used to monitor the two channels simultaneously and verify the correct pickup of sound in both microphones. As possible to see in figure 3.24: the oscilloscope correctly detected the impulse on one of the channels, while the other remained silent (that meant there was no unwanted crosstalk). The testing procedure also checked for correct filtering of DC current, high signal-to-noise ration and finally stability of soldered components (to avoid loose connections).

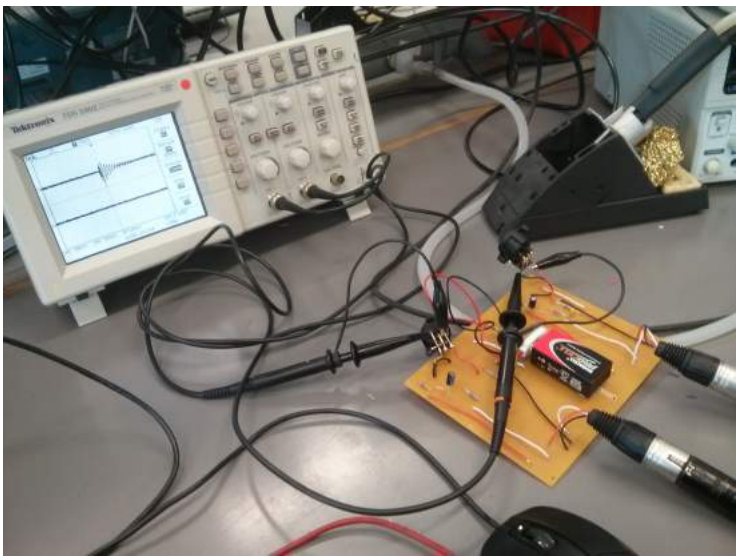


Figure 3.24: Testing process for the phantom power supply using an oscilloscope to check the signal output

Once the circuit passed the testing criteria and was judged to meet the specifications, a plastic enclosing case was ordered and drilled to encapsulate the unit. The case served to provide a higher degree of physical

protection to the unit, allowing easy and safe transportation between environments. The case was provided with holes for input/output connections, drilled with professional laboratory equipment, and equipped with a removable lid, which could be fixed with screws. Figure 3.25 shows the finalised unit.

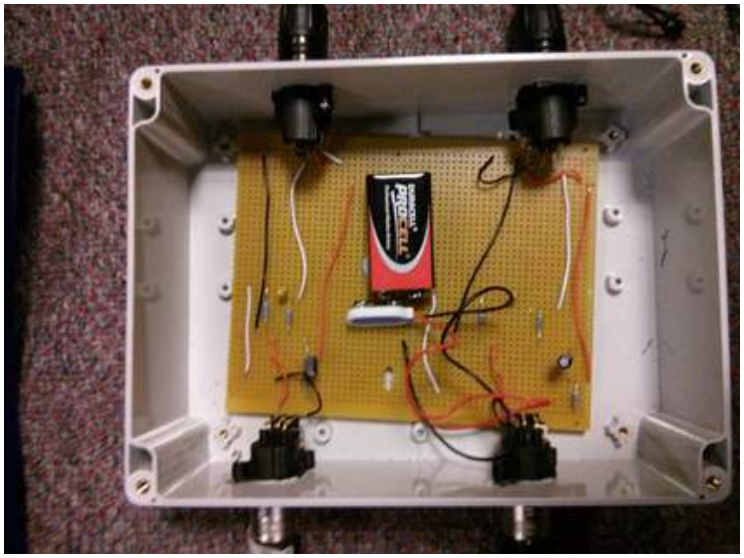


Figure 3.25: Finished version of the phantom power supply circuit

The input impedance of the soundcard would act as a High Pass Filter on the signal, being the value of that impedance around $10k\Omega$ the resulting break frequency was calculated to be $2.34Hz$ according to the formula:

$$f_{cutoff} = \frac{1}{2\pi RC}$$

3.3 Problems Encountered

A compromise had to be made when choosing the headphones for the experiment. Ideally, as done in [27] and [5], STAX headphones should have been used due to their reputation of being acoustically transparent. The unavailability of this item and its high cost led to the second non-optimal choice of the DT-990-Pro Open headphones model. It is not clear

whether this compromise would have affected the results of the experiment. Either way, the second choice of headphones still represented a high-quality option.

Two previous versions of the chair were discarded due to their failure to meet the specifications. The first chair was constructed by the *University of York, Electronics Department* technicians, however the reflective area behind the head was too big and would have covered the direct sound from the rear speaker. The first re-adaptation reduced the reflective area but still presented a possible path-block for the sound of the rear speaker to the ear. The third and final version (the version shown in this section) solved this problem by substituting the lateral support poles with a central pole. This process slowed down the progress of the project at its very initial stage as time was first spent waiting for the technicians to complete the first version and then looking for the construction material for the later versions.

It also needs to be mentioned that the successful completion of the power supply unit took much more time that it was initially allocated. Two failed versions of the circuit board were designed before the final illustrated version, also, many unexpected problem were encountered. Initial designs were implemented on PCB boards, however the lengthy waiting times to obtain the printing of a PCB board eventually led to usage of a prototype board instead. The first design was very similar to the final design (figure 3.23) but did not include capacitors C1 and C2. Its first implementation completely failed the testing stage due to bad soldering of components which created a short circuit in the board and blocked any signal from going through. The design was re-implemented with more care and was initially thought to be working as only one channel at the time was tested, but a more attentive test procedure which involved the two channels to be tested simultaneously exposed the presence of unwanted crosstalk between channels. A review of the design led to the

conclusion that a ground loop between the channels was shared and allowed the signal picked from one capsule to be sent to the other. After having included capacitors C1 and C2, it looked like the design would have passed the test, as it initially did. However a repetition of the testing stage showed an exceedingly high level of noise picked up by both channels. Several days were spent trying to identify the problem, the circuit was again re-implemented but the problem persisted. Eventually, it was discovered that both the XLR-to-LEMO cables provided by the department to connect the capsules to the unit had both broken in the same point and at the same time. Being this event very improbable and unlikely to happen, it took some time before it was taken into consideration. Fixing the cable connections confirmed that the design was actually working and the specifications were met.

Due to the reduced opening times of the laboratory facilities, component ordering times, and the unexpected problems, three weeks instead of one were spent on the phantom power unit and the chair. To these three weeks, another one has to be added for the time spent gathering the equipment and setting-up the environments, making a total of more or less 4 weeks of work spent on this stage.

For more information about the previous versions of the chair and the phantom power unit circuit, see appendix B.

Chapter 4

Individualisation Measurements

Contents

4.1 HRIR Measurements 106
4.1.1 Sine-sweep Technique 107
4.1.2 Free-Field measurements 108
4.1.3 Subject Recruitment 111
4.1.4 Measurement procedure 113
4.1.5 DAW workspace 116
4.1.6 Individual HpIR measurements 116
4.2 Signal Processing 118
4.2.1 HRTFs processing 119
4.2.2 Free-field transfer functions 119
4.2.3 ITD and ILD correction 124
4.2.4 Headphones compensation filters 126
4.3 Problems Encountered 132

The second major stage of the project involved the measurement of individual HRTF pairs related to the speaker positions set-up in the previous chapter. Participants were recruited and subjected to the measurements which involved the recording of a sine-sweep signal. Headphone Impulse Responses (HpIRs) were measured in a similar way and then processed in order to calculate the headphone equalisation filters. The signal processing phase served to combine the individual headphones com-

pensation filters with the individual HRTFs thus creating the appropriate binaural filters needed to create individualised spatial content. This chapter covers the whole measurement and signal processing procedure for creating and equalising those filters.

A procedural mistake occurred during the measurement stage and led to some problems which had to be compensated in software, making the procedure non-ideal. The mistake was caused by an initial wrong choice of soundcard which did not allow stepped control of the microphone gains. For later stages the soundcard was changed to another model with stepped controls. Other problems that could not be avoided, as they depended on the subjects' behaviour and ear morphology, affected the quality of the measurement for some of the subjects producing wrong relations between channel levels. Full details of these problems are given in section 4.3

For reference purposes, it is reminded to the reader that the terms BRIR and BRTF indicate reverberant HRIRs and HRTFs.

4.1 HRIR Measurements

The HRTFs related to the chosen speaker positions had to be recorded for a number of subjects in both listening environments. This was achieved by inserting the microphone capsules into each listener's ear canals and recording a sine-sweep signal emitted by the loudspeakers. The recordings would then have to be processed in order to obtain the HRIRs. A safe and short procedure had to be prepared in order to ensure that subjects would not feel too uncomfortable during the measurements due to the invasive presence of the microphone capsules in the ears.

Rooms were prepared as shown in figure 4.5. To improve efficiency between

measurement sessions, 2 amplifiers and 2 soundcard units of the same brand were available and split between room environments. The amplifier levels across rooms were set to be the same; the loudspeaker levels were further checked by using a sound level meter and making sure that all the speakers' volume levels would match at a distance of 1 metre. The microphone gains had to be set up using the knob controls of the *Focusrite Scarlett 2i4* soundcard. Unfortunately these controls were not stepped and had to be visually matched, making the gain-matching process non-ideal and subject to approximations. A laptop and a Digital Audio Workstation (DAW) software was used to control the sending of the signals to the loudspeakers.

4.1.1 Sine-sweep Technique

A popular technique used to record HRIRs is the one described by Farina in [41]. This technique involves the use of a *sine-sweep*, a sine-sweep consists in a sinewave signal that 'sweeps' across a selected range of frequencies throughout a specified time-frame. Recording this signal, and then applying a deconvolution process, allows us to retrieve the room's linear impulse response. The deconvolution process is operated using the sweep recording and the *inverse filter*, which is the time-reversal version of the sine-sweep with a logarithmic decaying amplitude curve. The resulting data from the deconvolution process is the room impulse response (RIR). Figure 4.1 shows three signals: a 48kHz ramped sinesweep (used for this stage), its inverse decaying filter and the resulting IR which is in this case a Dirac delta function (this proves the correct functionality). Using the recording of the sinesweep done by the two capsules inserted in the ear canals, the RIR would instead be a HRIR. The recorded HRIR pair is then transformable into a frequency-domain HRTF pair using fast Fourier transform.

The sweep signal was created in MATLAB using a sample rate of $48kHz$, sweeping across the audible frequencies range up to half the sample rate (11 Hz to 24kHz), with a total duration of $5,5$ seconds meaning 11 octaves covered, each one $0,5$ seconds long. The code script was adapted from a third party version provided by the *Genesis 6* technician. A copy of the code used to produce the sinesweep and the inverse filter can be found in the supporting material CD in the folder *"Code/Stage2/"*.

4.1.2 Free-Field measurements

In order to check that the microphone capsules' gains matched in level and were not subjected to any fabrication defect, a free-field recording was taken. The capsules used for the measurement stage (figure 4.2) were first wired to their respective XLR cable (figure 4.3) and then joined together,as close as possible, on a microphone stand (figure 4.4). The same connections shown in figure 4.5 were used (with the exclusion of the human participant and the chair).

The stand was placed to be in the location where the listener's centre of the head would have supposed to be in each room, pointing to the first speaker. The sine-sweep signal was then played by the speaker using the DAW (section 4.1.5) and recorded. After the recording, the stand was pointed to the other speaker in the same way; the sine-sweep was then recorded for that position. The same measurement procedure was then repeated for the anechoic chamber. These measurements served to find out how accurate the gain matching of the soundcard was, and take account of the gain difference in order to apply it to the individual HRTFs recorded. For details of the processing of the Free-field measurements see section 4.2.2.

This procedure should have ideally been performed at the very beginning, before starting the measurement on the subjects. However, due

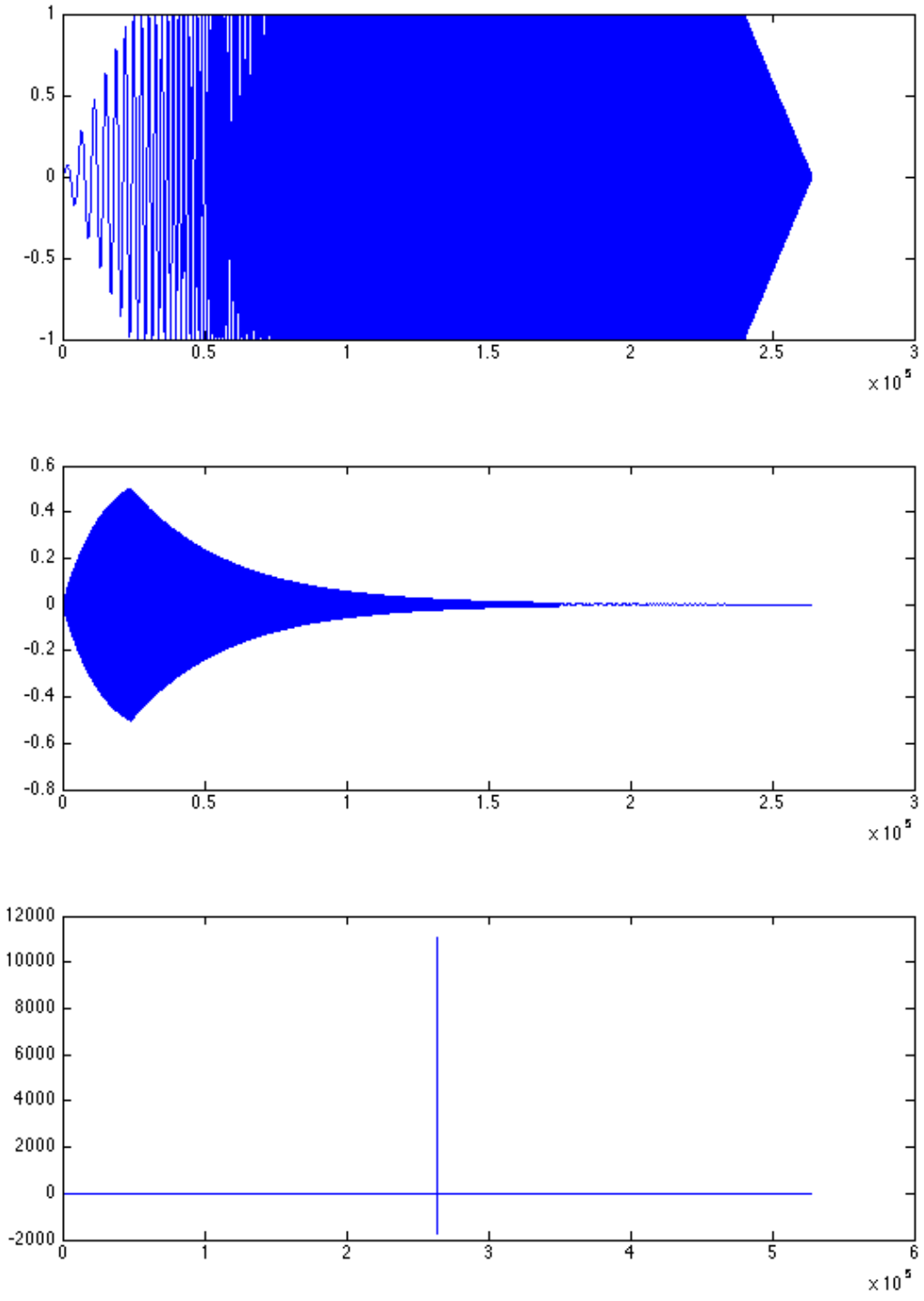


Figure 4.1: The original sine-sweep signal, the decaying inverse filter, and the IR obtained from the deconvolution of the two. X axis is length in samples (sample rate is 48kHz) and Y axis is amplitude

Figure 4.2: Microphone capsules used for recording, same model as the ones used for testing the phantom power supply

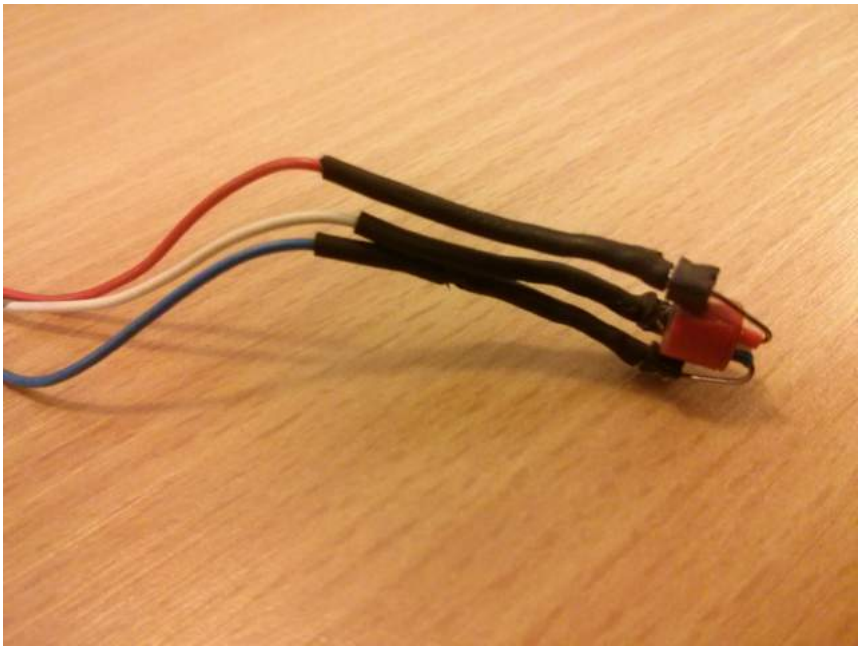


Figure 4.3: Detail of the cable wires connected to the capsule

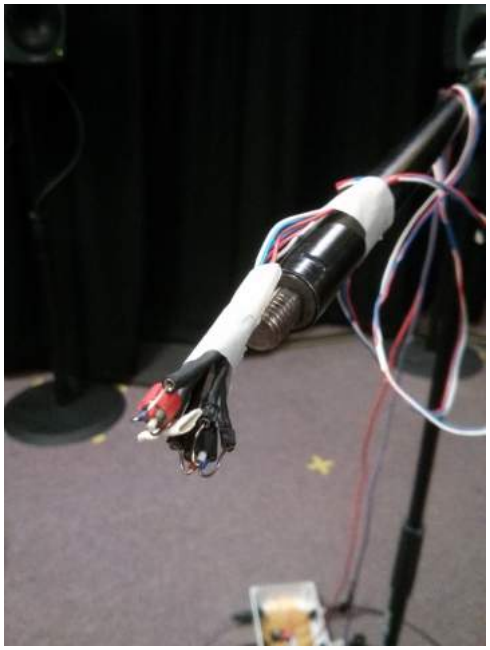


Figure 4.4: Detail of the microphones set-up for the freefield measurements

to an underestimation of needed precautions, these measurements were delayed and performed half way through the individual measurements stage. Unexpected changes in the gain settings, for which no precaution was taken, made the earlier individual measurements on participants unrelated to the calibration measures taken with the free-field technique. More details of this procedural mistake are given in section 4.3.

4.1.3 Subject Recruitment

The number of participants used was aimed to be similar to the one used by [27] and [5] so that the same minimum optimal size of data samples could be maintained. Both studies used 11 subjects which had to make 100 decisions each. These figures were calculated according to this formula:

$$N_{opt} = (z_\alpha + z_\beta)^2 \frac{2\pi}{\hat{d}^2_{min}}$$

Here z_α and z_β are the z values for type I and type II error respectively. This formula assumes perfectly unbiased participants and equal variance between noise and signal conditions [5]. In the case of less participants being recruited, the number of presentations would have been increased to match the minimum sample size required. It was decided that the best approach was to get as many participants as possible up to the number of 11, with the number dynamically increasing along the way as the measurements proceeded, rather than establishing a fix number from the beginning. This way, it was possible to adapt the time-management schedule and allow the number of participants to be dependent on the remaining time available, in case unexpected problems would have increased the length of the measurements to an extent where the schedule had to be revised.

Subjects were recruited among friends and course colleagues who volunteered to participate to the project. The required number of participants was eventually met, although attempts to balance gender as much as possible as typically done in standard listening test procedures, the numbers were unbalanced (7 males, 4 females) but this was considered to be a negligible aspect. Subjects were asked to sign a safety form before moving on with the next steps and, upon confirmed agreement, shown the procedural instructions for the measurements. Both documents are available in appendix A for reference. A brief questionnaire given to participants revealed that the range of participants' age was between 20 to 24. Only two of them (both course-colleagues) had experience in listening to binaural audio before. Table 4.1 below shows the details of each participants. Real names are hidden due to privacy agreement; for reference purposes, each participant is assigned a label from the letter of the alphabet. A total of 11 subjects was eventually reached.

Table 4.1: Details of participants

Name	Age	Gender	Experienced	Hearing disabilities
A	24	M	Yes	Mild Tinnitus
B	22	M	No	-
C	22	M	No	-
D	20	F	No	-
E	21	M	No	-
F	20	M	No	-
G	22	F	No	-
H	22	M	Yes	Mild Tinnitus
I	20	M	No	-
J	21	F	No	-
K	21	F	No	-

4.1.4 Measurement procedure

Once the instructions and the agreement form were presented to the subjects, the measurement sessions could begin. Each subject was firstly seated on the experiment chair, following that, the head-strap was put around the head and tightened (gently but firmly) in order to secure the head to the headrest. The insertion of the microphone capsules was carefully planned: a new pair of sponges (figure 3.13) was used to cover the capsules, the capsules where then rested on the participants' ear canals. The subjects were asked whether he/she preferred to allow the experimenter to push them in or if they wanted do the operation by themselves. Extra care was taken in making sure that the capsules were not over-inserted. Although the size of the ear sponges practically extinguished the risk of over-insertion, the procedure had to be monitored attentively. Once the microphones were in place, surgical tape was used to secure the capsules' XLR cables on the sides of the participants' neck in order to avoid the weight of the cables pulling the microphone away from the ear. Finally the headphones had to be placed over the ears carefully as to prevent pulling out the cables and capsules. The reason for having headphones was that it was necessary to include in the simulated stimuli the effect that wearing headphones would make on external sound sources [27] [5]. To verify that the microphones were still held in place,

the headphones were removed and then placed again upon confirmation of correct capsules' placement. Throughout the whole process participants were asked if they felt any discomfort as to ensure that they were as comfortable as possible.

Before recording, participants were instructed to keep their head straight and still (the head-strap alone was not enough to ensure immobility). Using a DAW, the sine-sweep signal was played from the frontal speaker and then the rear speaker. The headphones were then removed and replaced after having checked the position of the capsules; a second recording was then taken in the same way. Before ending the session, the recordings were quickly listened to on headphones to ensure no unwanted artefacts or external were present. If no problem was found, the second session in the other listening environment was implemented.

Figure 4.5 shows a sketch of a measurement scene. The subject is shown to be sitting on the chair with the head secured to the chair's headrest, the microphone capsules are inserted in the ear canals, the cables are secured with tape to the subject's neck and finally headphones are placed on top. The headphones did not need to be connected as their function was only to provide the acoustic filter. The two speakers were connected to the amplifier which was in turn connected to the soundcard, the microphones were connected to the phantom power supply which was also connected to the soundcard. A USB connection between the soundcard and a laptop allowed the latter to be able to send the sine-sweep signals to the loudpspeakers, and record the microphones signals into a DAW.

All the 11 participants were successfully subjected to the measurements, the two environment sessions were done on the same day for each participant in order to make efficient use of time. Average times for the two measurement sessions together was 20 to 40 minutes depending on the complexity of the capsules insertion phase. There was no requirement for establishing an order of which room had to be recorded first, hence, for

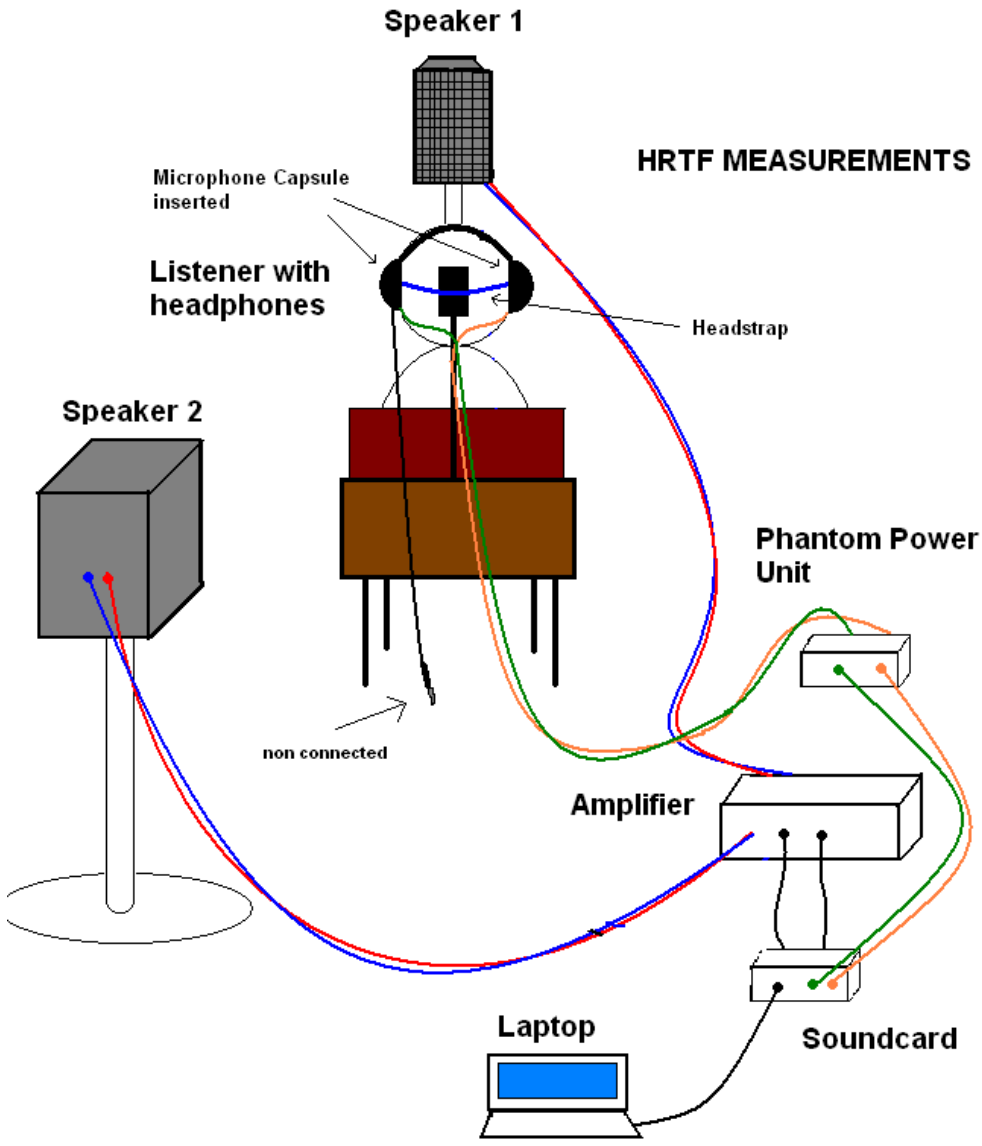


Figure 4.5: A sketch of the measurement procedure, showing the equipment and the connections

every subject, the first room recorded was the one ready to be used, the equipment (chair and laptop) would have then been moved to the other room for the second session.

The author's own HRIR were also recorded, with the help of one of the participants. This allowed the author to directly experience the further manipulations done in the signal processing stage on himself and make sensible decisions, with the guarantee that the individual recordings would maximise the quality of the binaural rendering and the accuracy of the spatial perception.

4.1.5 DAW workspace

The REAPER® DAW was chosen due to the availability of a free trial version, available for MAC OSX, which included all the needed functionalities. The interface shown in figure 4.6 was set to include two new tracks for each participant (one for the left microphone and one for the right microphone). Recordings were time aligned with the original sine-sweep signals and then exported separately for each channel using the settings of figure 4.7. Export sample rate was $48kHz$ as the recording sample rate set on the soundcard; export bit-depth was set to 24 PCM. Exported .wav files were labelled according to the following format in order to keep the files organised: *NAME_ROOM_POSITION_CHANNEL.wav*.

4.1.6 Individual HpIR measurements

Headphones Impulse Responses (HpIRs) had to be recorded individually for each subject in order to take account of the modulating effect of the pinna on the spectral cues of the sound emitted by the headphones and create a compensating equalisation filter that would improve the quality of the timbre by preserving those cues. This process was based on the

Figure 4.6: REAPER DAW interface

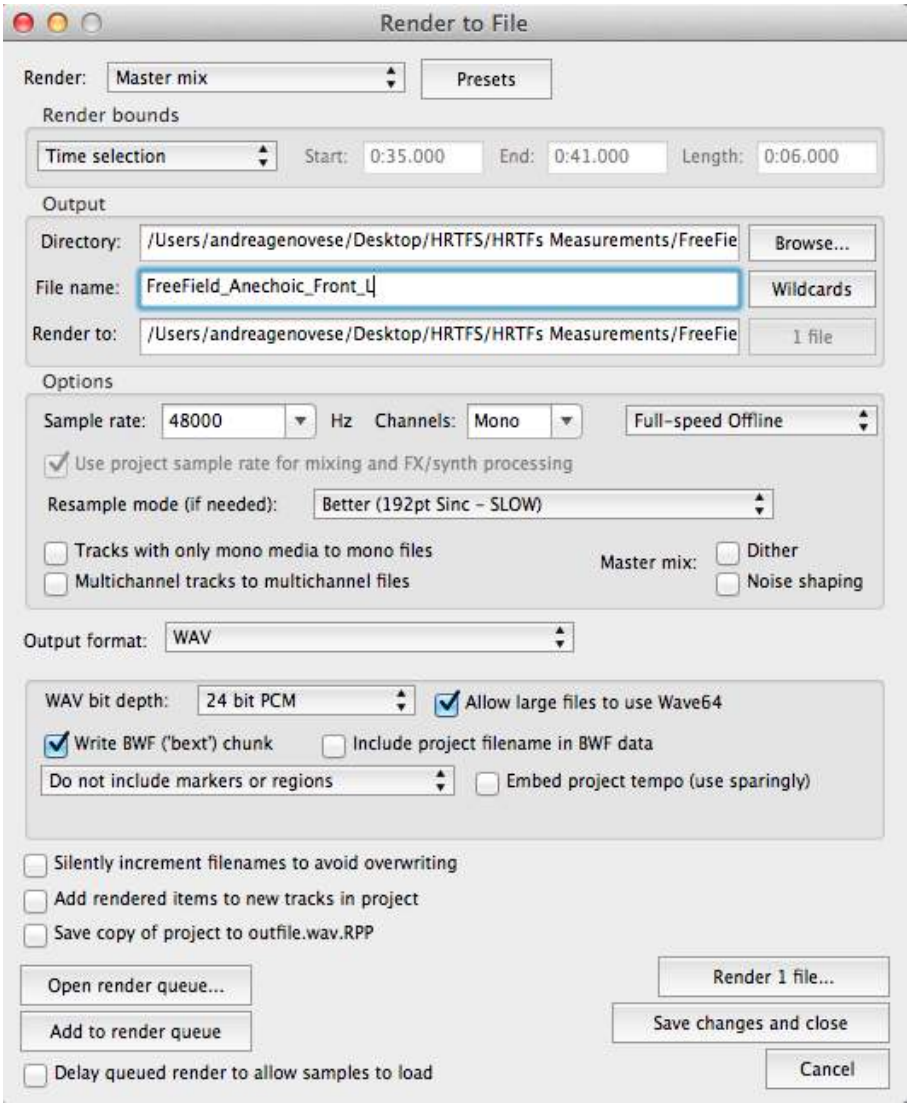


Figure 4.7: Detail of the export settings used in the DAW

process previously done by [27] and [5] on an artificial head. Due to the reasons explained in section 4.3 these measurements were performed in different sessions and in different environments than the one used as the external room dimensions would not matter for the sound emitted by the headphone speakers.

The same procedure as in the HRIR measurement was used, without the need of loudspeakers or the listening chair. The sinesweep signal was first recorded for the left ear and subsequently for the right ear. As done in [5] the measurement was repeated 10 times on each subject, each time removing and replacing the headphones. This accounted for the fact that different headphones placement on top of the head would slightly change the resulting spectral content recorded, the 10 measurements would then be averaged in the signal processing stage.

The recordings were exported from the DAW using the same settings as in figure 4.7. This time the format was *NAME_NUMBER_CHANNEL.wav* where *NUMBER* indicates the number of the measurement out of the 10 taken.

4.2 Signal Processing

Once exported, the recordings had to be processed in MATLAB in order to extract the HRIR and the BRIRs. A student version of MATLAB (R2013a) was installed and used for the processing. The exported recordings were processed in the way described in 4.1.1, using a fast convolution algorithm that operated in frequency domain (a code script of this function is available in the supporting CD).

4.2.1 HRTFs processing

The HRIRs obtained from the deconvolution with the inverse filter were truncated starting from few milliseconds before the direct-sound peak up to $2^{13} = 8192$ samples of length, in order to allow the late reflections enough time to decay to zero. The indexes of the truncation points were found to be the same for every recording thanks to the time-alignment with the sweep signal in the DAW which allowed to precisely export the raw recordings with a consistent time duration. The following step was to normalise the amplitude of the HRIR pair to max the highest peak of the two-channels pair at the value of 1. Using a fast Fourier transform function (also included in the CD) the HRTFs were obtained from the truncated HRIRs. No smoothing algorithm was applied.

Figure 4.8 shows the resulting HRIR pairs related to the author's own recordings for the frontal position. The top row represents the HRIR and HRTF for the anechoic room and the bottom row shows the BRIR and the BRTF for the reverberant room plotted in decibels over a logarithmic scale. The left channel is plotted in blue while the right channels is superimposed in green. The figure shows, as expected, that the presence of reverberation increased the HRIR decay length and created more variety between the spectrum content of the two channels. Figure 4.9 shows the same plots for the recordings related to the rear speaker.

4.2.2 Free-field transfer functions

In order to account for the gain differences between the capsules caused by un-equal level settings and fabrication imperfections, the freefield measurements had to be processed. The same deconvolution and truncation procedure described was implemented on the freefield measurements to obtain the RIR. It was assumed that, in the freefield case, the energy at

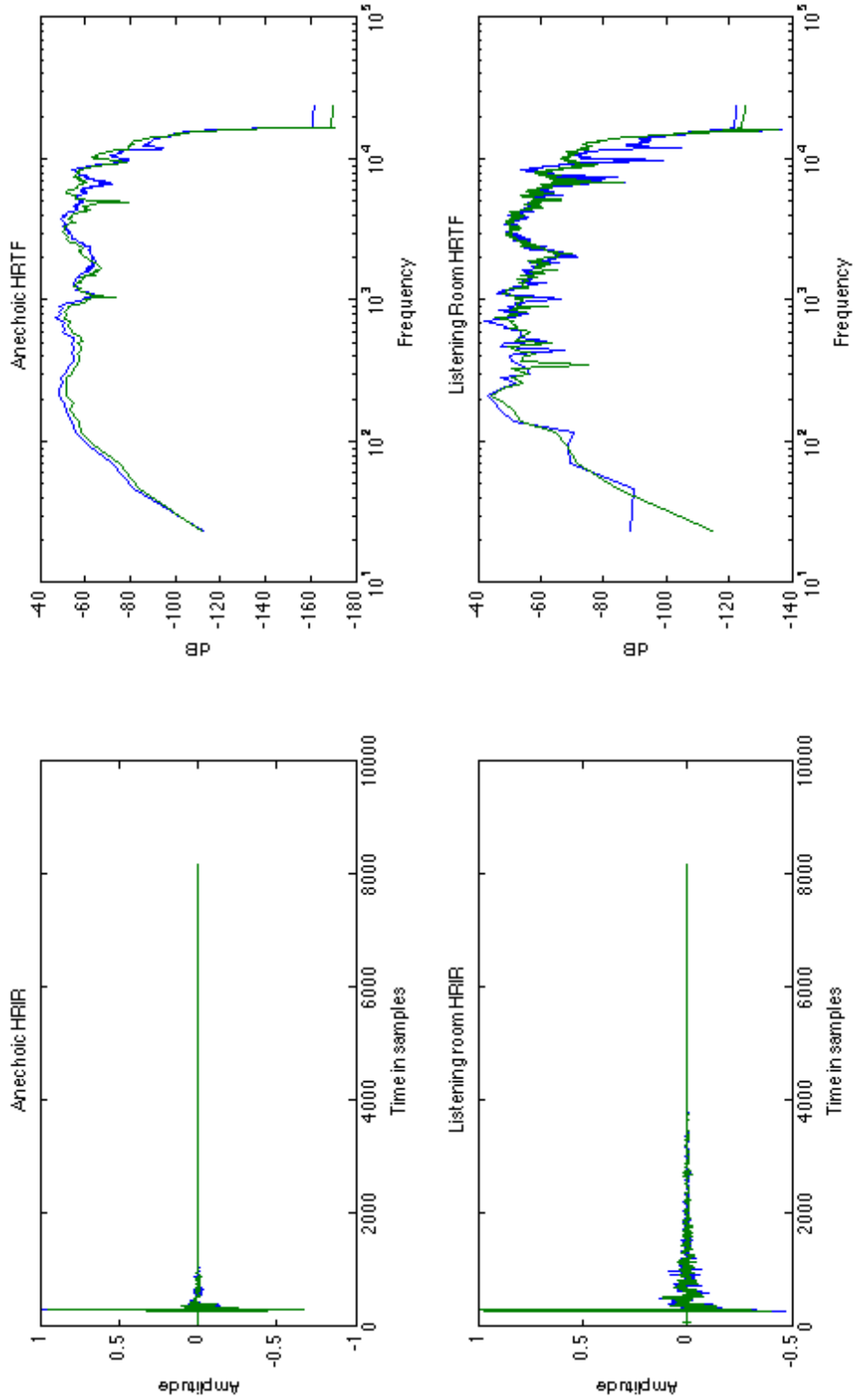


Figure 4.8: Anechoic and echoc HRIRs and HRTFs processed from the recordings related to the **front** speaker

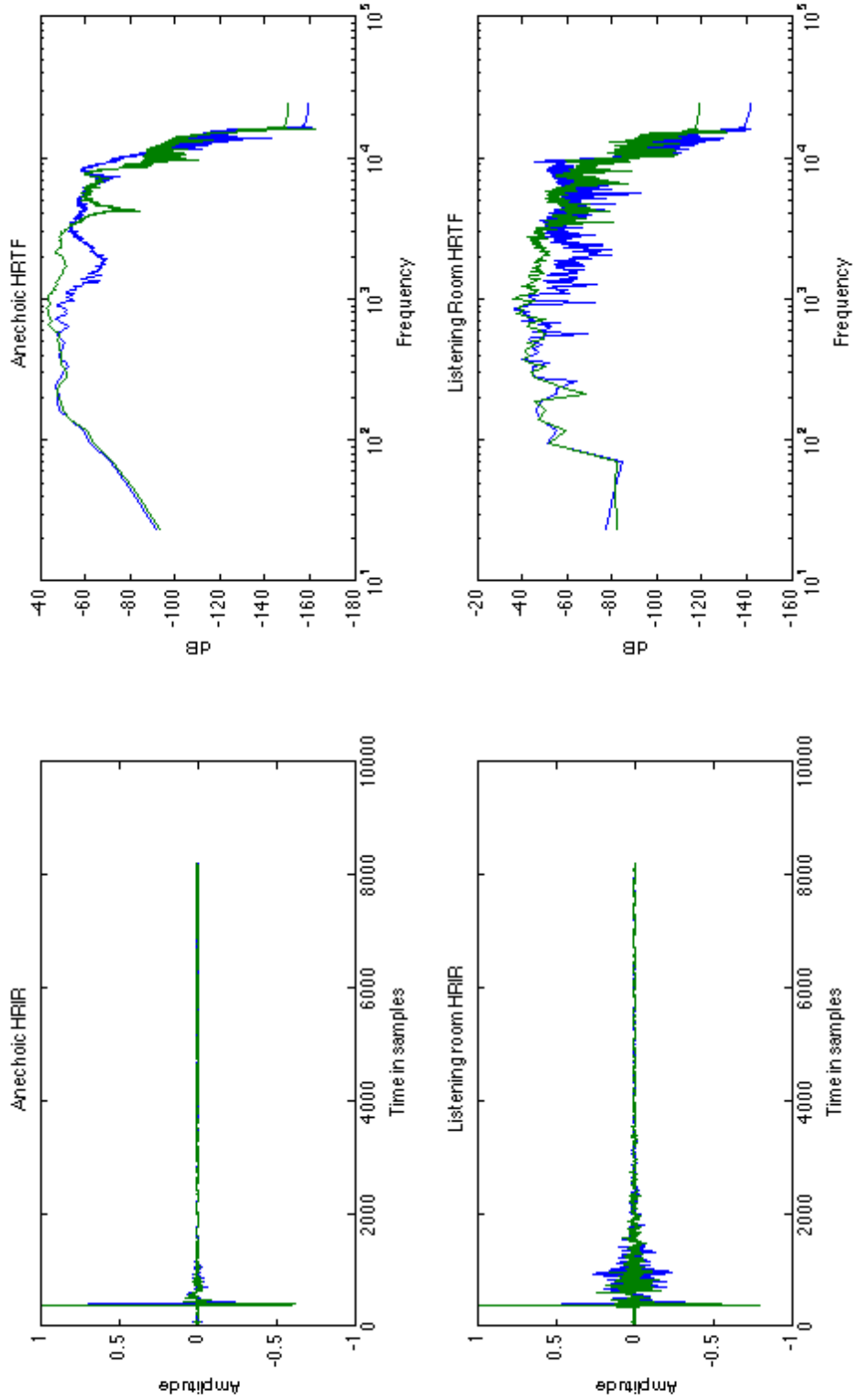


Figure 4.9: Anechoic and echoc HRIRs and HRTFs processed from the recordings related to the **rear** speaker

both microphones should have been approximately the same. In order to check the gain offset between the channels, the average rms energy of both channels was calculated separately and coefficient was derived from the ratio of the two energy measures. The energy was calculated according to the formula:

$$E_{RMS} = \sqrt{\overline{HRTF^2}}$$

The ratio of the energies, served as digital gain coefficient to apply to one of the two channels in order to match the average energy, some differences would still be present but a slight improvement can be obtained. Figure4.10 shows the processing executed on the freefield measurement done in the listening room for the front speaker position. The image shows the following:

1. The RIR of the freefield measurement
2. The transform in frequency domain plotted in dBs, logarithmic scale
3. The superimposed RMS energy levels (light blue for left channel, light green for right channel
4. The corrected frequency response after applying the gain coefficient factor

This procedure was repeated for every speaker involved in the experiment, gain correction coefficients were thus calculated individually for every speaker. The same correction coefficients measured in the freefield case had to be applied to the individual HRTF measurements, each HRTF was processed using the coefficient associated to its particular loudspeaker. Due to the reasons explained in section 4.3 the free-field correction could only applied to some of the subjects measured. As a consequence of those problems, subjects B, C, D, E and F, which were measured before

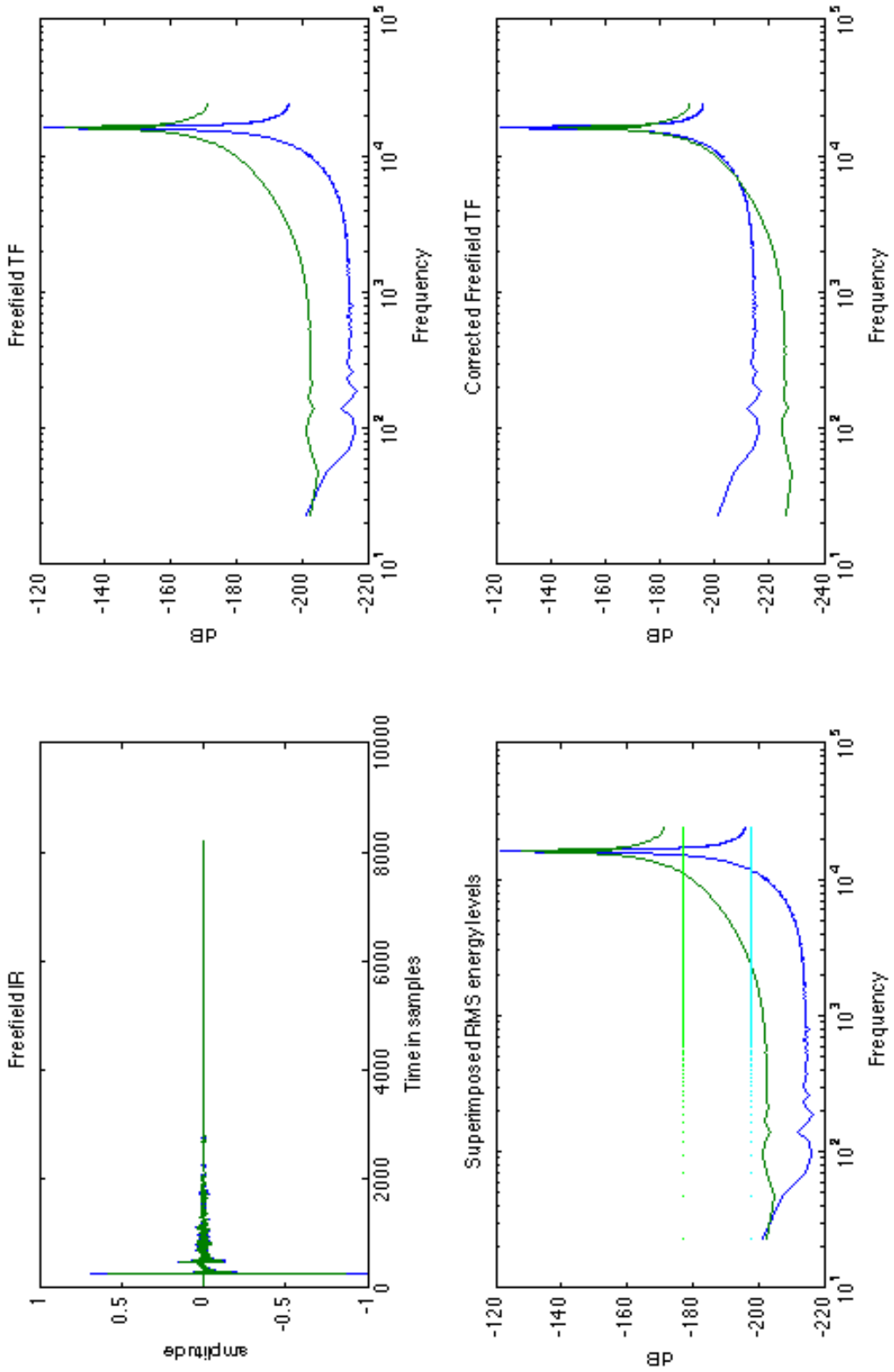


Figure 4.10: Freefield measurement taken in the listening room, front speaker. The rms energy ratio calculated in the frequency domain served as gain correction factor

the free-field measurements were recorded, with different soundcard settings, and also needing a different kind of gain correction.

4.2.3 ITD and ILD correction

To test that the measurements were carried out correctly, and to check the presence of eventual problems, a noise train signal was created and convolved with all the HRTF measurements of each subject. A careful listening stage of every HRTF pair highlighted some problems in the accuracy of the localisation which exposed the need of futher inspection of the HRIRs and HRTFs.

It was found upon inspections that in some cases the direct-sound peaks in the time-domain HRIRs, were not aligned for the front speaker, causing an ITD of few samples (range of ITDs was from 1 to 3 samples), way above the detection threshold of $10\mu s$ [28]. This could have possibly been caused by a wrong positioning of the subjects in the listening environments. To create an ITD, it was enough to have the head rotated few degrees off-shift from the centre of the speaker, meaning that either the subject had moved or that it was wrongly positioned in the first place (see section 4.3 for more details). This error would have caused an offshit of the virtual image a few degrees to either side of the centre angle. This off-shift was verified and confirmed by preliminary tests. All the HRIRs and BRIRs measured for the frontal positions, which exihibited a non-zero ITD, were subjected to a correction script where the delayed channel was shifted earlier in time to have its amplitude peak matched, on the x-axis, with the other channel. The amount of correction calculated for the front position of each environment was then applied to the rear position of the same environment (it would have been inappropriate to calculated the ITD error on the rear position as an ITD was expected).

Even after the correction of the ITDs, preliminary listening tests showed

that in some cases, a substantial location error was still present. Even though the HRTFs inspected did not belong to the author, the amount of error present had objectively a deep impact on the spatial localisation of sound sources. The extent of this error was such that simulated sources meant to be perceived in the front, were perceived instead from an angle somewhere around 60° on the left. A participant (Participant D) was "spoiled" and asked to participate to the preliminary tests where he was presented a signal convolved with his own HRTFs. The subject agreed with the author's opinion that the virtual position was considerably off-shift compared to what it should have been, raising the need of further investigation. Inspection of HRTFs related to the front speaker revealed a visually-substantial ILD difference which could not be ignored, this was identified as the cause of the persistence of localisation errors. It was again assumed that the ILD for the front position should have been approximately zero. Not-surprisingly these errors were present only in subjects B, C, D, E and F for which a gain correction based on the freefield measurements was not applied. For these subjects, it was decided to use the same approach used to compute the free-field gain difference (section 4.2.2) to compensate for ILD differences. For each individual subject out of those mentioned (B to F), the average rms energy levels were calculated for both channels of the frontal HRTF pair and a gain correction coefficient was obtained by the ratio of the two levels; the process was repeated for both listening environments. The coefficient was used to scale the quieter channel to match the average energy of the other channel. The same associated coefficients were applied to the rear HRTF pairs. As with ITDs, an ILD was expected for the rear sound sources meaning that the ILD could only be extracted from the front direction. A second session of preliminary tests, again involving participant D, demonstrated that the gain correction brought significant improvements to the spatial perception of the sound sources. Both the author and the participant agreed that the ILD correction was a crucial factor in improving the directionality of the simulation and simulated sources were now sounded "as they were

meant to be" . It is discussed in section 4.3 how this decision meant that the measurements lost their "pureness" as a manipulation process was involved.

Figure 4.11 shows the application of the corrections on subject D which was an example of a measurement involving both ITD and ILD errors. The first column shows the application of ITD correction in a front position BRIR (with a zoomed view on the time domain peak), the top image is the original BRIR and the bottom image is the ITD-corrected version. The second column shows the related BRTF, original on top and ILD corrected at the bottom. The red markers represent the values for the left channels, the magenta markers represent the right channel.

4.2.4 Headphones compensation filters

The HRTF filters had to be merged with the headphone filters before being ready for usage on the experiment stimuli. The first step was to extract the HpIRs for each subjects from the recordings described in section 4.1.6, 10 recordings per subject were taken, each recording was first deconvolved and truncated (this time up to 2048 samples) for each channel and then normalised. Channels were normalised separately as only one headphone speaker per time was active during the recording. The reason for having 10 recordings per subject was to be able to average the effect of repositioning the headphones over the ear, some slight differences between recordings were therefore expected. Nonetheless, repositioning the headphones on the current participant occasionally caused the microphone capsules to fall out of place, the reinsertion of the capsules would then have included re-insertion differences. This happened particularly often for those subject which possessed ear morphologies that made the insertion process quite "problematic". A first inspection of the recordings, was based on a comparison of the 10 measurements, in frequency

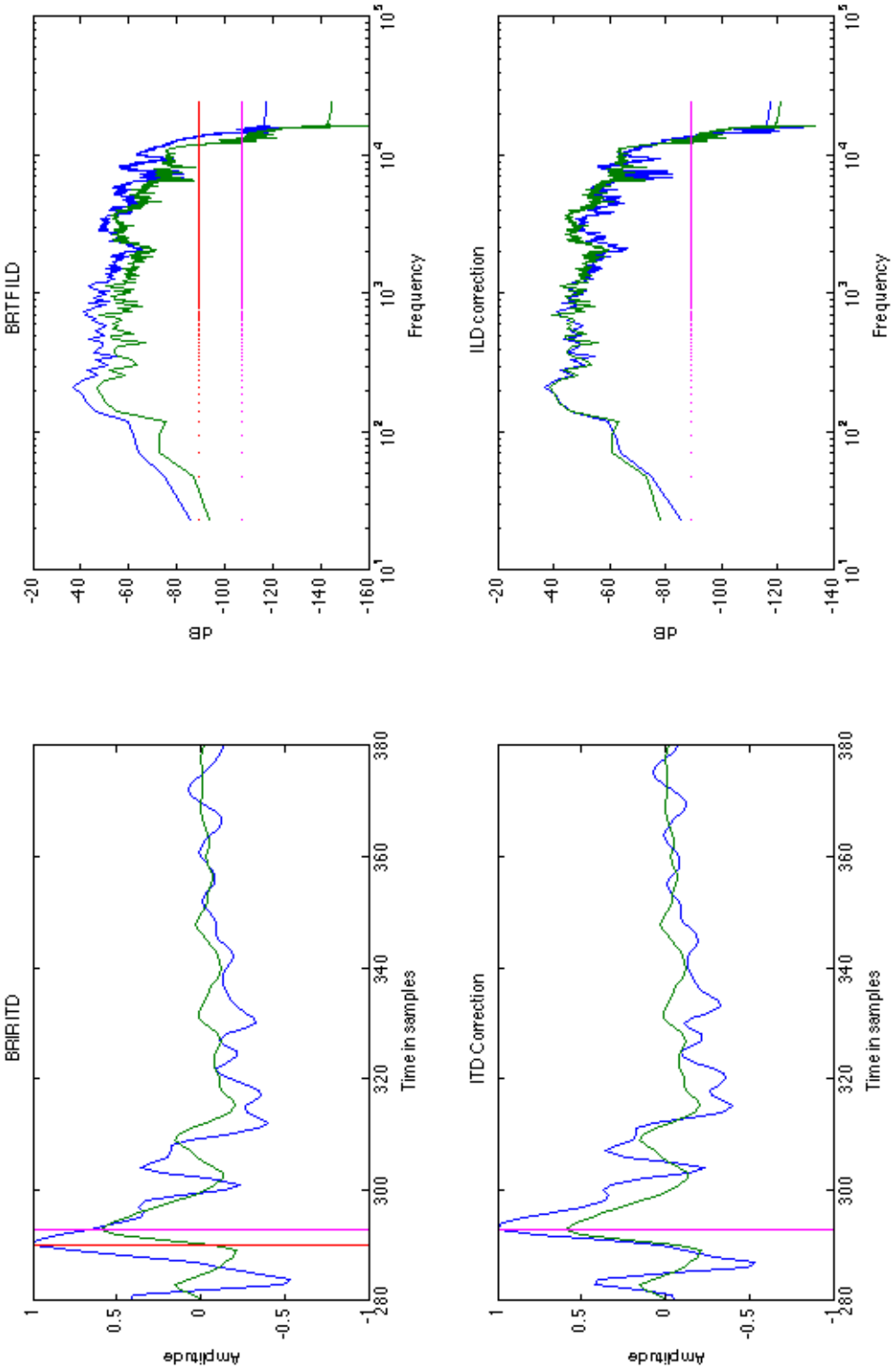


Figure 4.11: Effects of ITD and ILD corrections on subject D, Listening room / front position

domain. Figure 4.12 shows, in different colors, the spectra of all the 10 recordings for subject D, for the right ear. It is shown that most measurements' spectra roughly follow the same shape, yet some of the spectra plotted have a very different shape from the average.

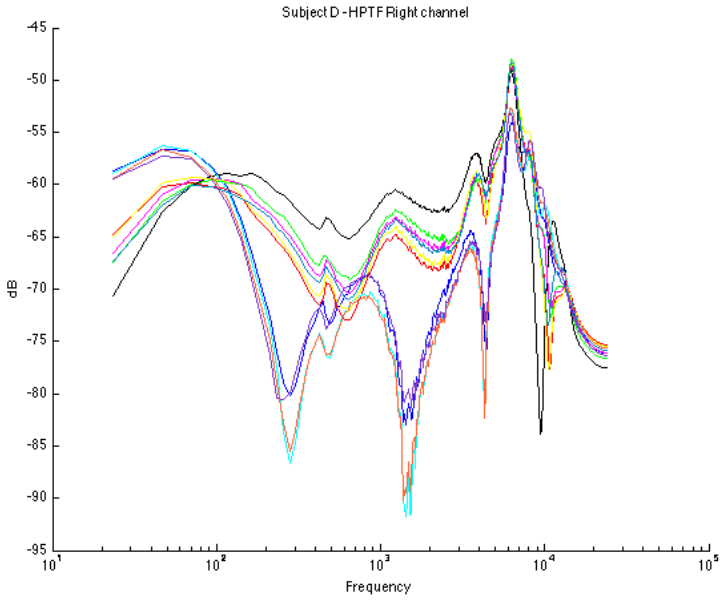


Figure 4.12: Plot of the 10 HpTFs for subject D, right channel

It was theorised that the measurements were not consistent in the example shown due to the reinsertion of the capsules every time they fell off position, raising the possibility that the microphone capsules were not properly inserted. Although some differences were expected, the degree of these were sometimes feared to be too severe in magnitude. It was decided to run a selection process between measurements. For each set of 10 measurements, an average for each channel had to be calculated, each measurement was then cross-correlated to its related average. If a measurement showed a degree of correlation less than an established threshold, the measurement would then be discarded. The threshold was set to be 0.7 as an inspection of the correlation coefficients across all measurements showed that 0.7 was the appropriate threshold in order to rule out only those measurements judged to be "too different" from the average, in the most extreme case, half of the measurements were discarded

(5 out of 10). The cross-correlation formula used was the following (Pearsons product-moment correlation coefficient):

$$CrossCorrelation = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}}$$

The resulting HpTF sets for left and right ear were fed to the compensation filter function, which included equalisation and normalisation of the created filters. This function was provided by Chris Pike, BBC R&D and coded following the methods described in [52]. It is explained in [52] that a perceptually robust equalisation filter for headphones can be obtained by a process of inversion of the upper variance limit of many measured HpTFs. The resulting output was therefore the headphone equalisation filter relative to the specific listener. A final processing phase needed to be applied: inspecting the filters in the time-domain displayed the need for wrapping the compensation filters to avoid the creation of artefacts. A preliminary test with a noise train signal confirmed that the wrapping was indeed needed to avoid artefacts in the filtered signal. Figure 4.13 shows the procedure of wrapping the EQ filters.

The wrapped EQ filters could then be safely applied to the previously-corrected HRTFs, figure 4.14 shows the anechoic HRTF filter related to the front position before and after being merged with the headphone compensation filter. An optimal binaural rendering could then be achieved for the next stage where the individual stimuli content was created. Informal screening of different type of signals with the HRTFs, before and after the application of equalisation, showed that the audio did indeed sounded “brighter” and more similar to the original item if the EQ was present.

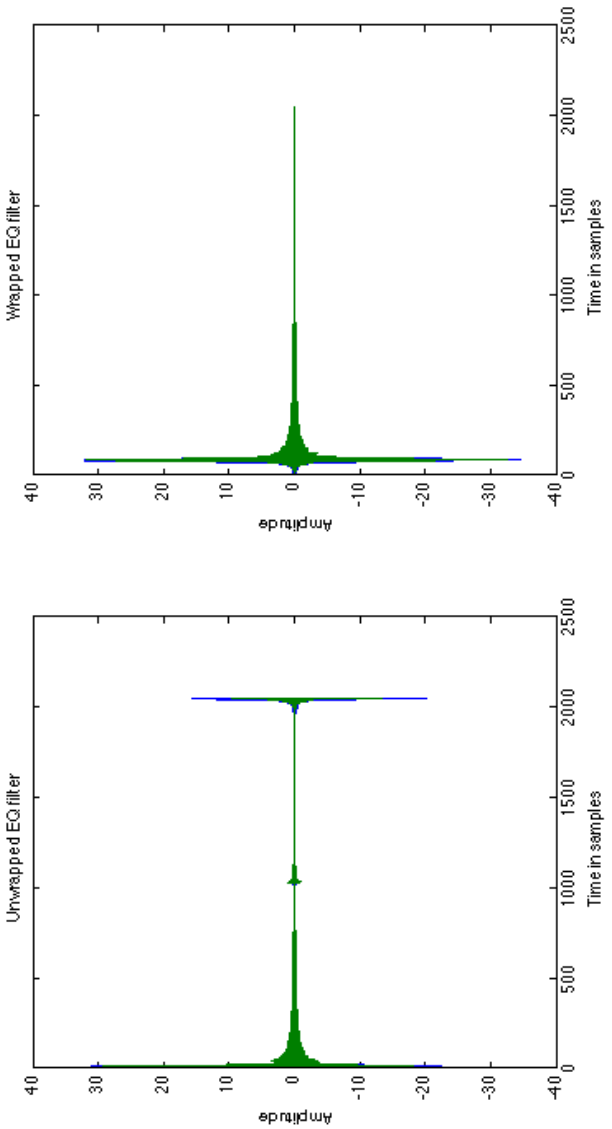


Figure 4.13: EQ FIR filter for subject D, before and after wrapping

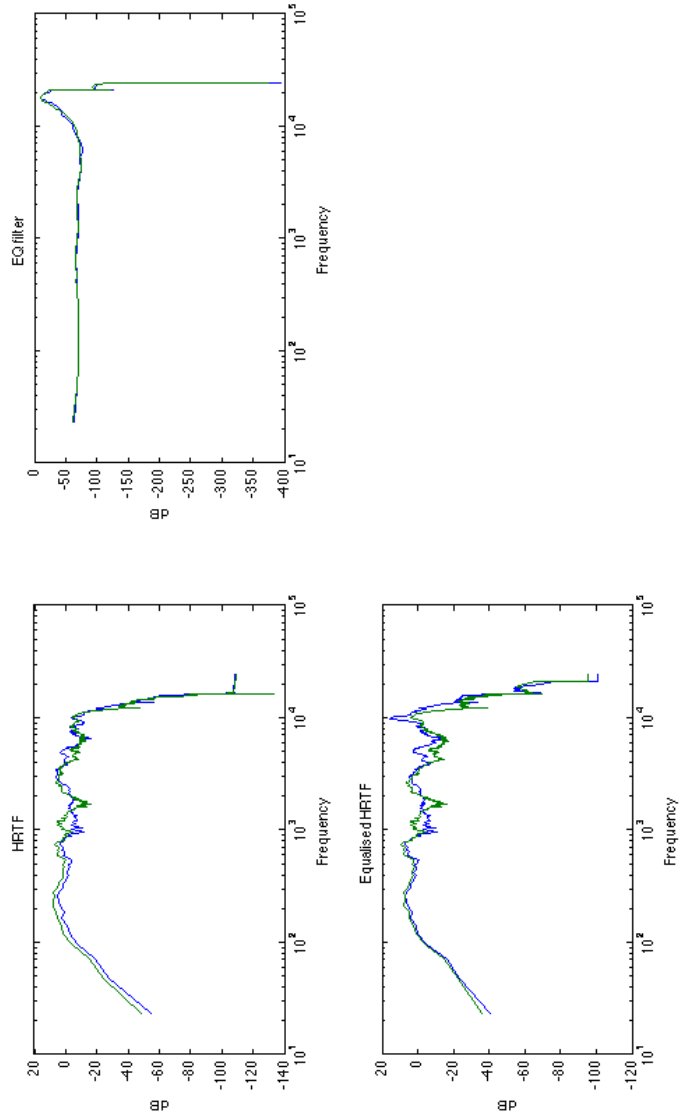


Figure 4.14: Anechoic / Front HRTF for subject A, before and after the merging with the compensation filter

4.3 Problems Encountered

The first initial recordings for the first subject were screened on the DAW. A substantial amount of noise was found in the recording to the point that it was beyond acceptable. After a discussion with the technician, it was suggested that the problem might have been caused by the proximity of the soundcard to the amplifier which would create interference. Up to that point the soundcard was in fact placed directly on the amplifier, the subject was re-called the following day and re-measured, this time having the soundcard properly distanced from the amplifier. The new recordings were screened and it was found that the noise was now absent.

The first soundcard used, *Focusrite Scarlett 2i4*, did not have stepped gain knob controls nor visual feedback for the levels of the two microphone inputs as the *MOTU Ultralite mk3* soundcard did. This meant that the gain between channels had to be visually matched. To do this, the two *Scarlett* soundcards used for the two rooms were stacked on top of each other and the gains were set to be approximately similar. The idea was that even if the levels were not entirely matched, a free field recording would have shown the difference between the channel levels, therefore a gain correction factor could have been derived and applied to the recordings. However, the freefield procedure was delayed as it was not predicted that the levels would have ever been changed or manipulated. Having underestimated the possible scenarios, proper precautions were not taken at the beginning and the exact levels of the input channels were not recorded nor an acoustical calibration process on the microphone levels had been performed. In fact, after the Easter break period, the soundcards had been used by third parties and the channels level settings were found to be changed without the possibility of exactly reproducing them as they were before the manipulation. It was decided at that point to run the freefield measurement procedure and record the exact posi-

tion of the gain knobs in case of further manipulation of the levels. The consequence of this problem was that the calibration would have been valid only for the subjects being measured after the free field procedure, and all the earlier measurements would not have been related to the new settings. As time constraints did not allow for the measurements to be retaken, and the number of measurements taken up to that point was considerably high (5 subjects), the correction procedures illustrated in section 4.2.3 was considered as a better alternative to the rejection of part of the work done so far. The corrected material was screened and it was concluded that the problem could have been overcome by using RMS level matching. Per contra, applying this kind of level correction involved a risky manipulation of the recordings, implying the compromise of their validity into accurately representing the spatial position of the sound source. This non-ideal solution was regarded to be a necessary compromise in order to maintain a good sample population, it was decided to go forward as planned with all the measured subjects, and later on run a separate analysis which would isolate the pre-calibration subjects with the post-calibration subjects. This compromise also led to the re-introduction of the blindfold in the experiment, as it was thought that a potential inaccuracy of frontal spatial position could have been made less substantial by blocking the visual senses of the subject from accurately locating the real sound source in front of them, thus improving the chances of plausibility in case the virtual stimuli for the frontal position would have been perceived as slightly off-centred. To summarise, 5 out of 11 subjects had their levels corrected through the RMS matching technique, all the rest were corrected using the freefield gain difference factor. It was decided to formally devide these groups of subjects into *pre-freefield* and *post-freefield*. Preliminary listening tests for all the resulting HRTFs convolved with a noise train signal showed that no perceptual difference (from the perception of the author and a "spoiled" participant) in directionality was present between the two groups; running a side-analysis procedure would reveal more accurately the consequences of using dif-

ferent gain correction on the plausibility assessment (see chapter 6).

Since the start of the headphones equalisation recording process, the soundcard was changed to the *MOTU Ultralite mk3* which allowed for digitally scalable microphone input levels, thus ensuring the possibility to accurately recreate the levels in case a manipulation occurred. The drawback of this choice was the fact that only one model was available and it was often requested by other students; as a result the scheduling of HpIR measurement sessions and, later, listening test sessions were constrained and more lengthy set-up times were required when switching session from one environment to the other.

Another source of inaccuracy during the measurements was due to particular pinna morphology of some subjects that made the insertion of the microphones very difficult. It was not safe to insert the microphone without the sponges so it was necessary to patiently try to make them fit using a trial-and-error method, until a suitable positioning was found. After a lengthy attempt phase, a twelfth subject could not be measured due to the fact that his ear canals were "too small" to fit the sponges. It is believed that the subjects with the most "difficult" ears presented some inconsistencies in the measurements' ILD levels. This is supported by the fact that people with "easy" ears, where the capsules could be inserted very easily, show less incongruence between channels. Furthermore, once the headphones were placed on top of the ears, the headphone exercised pressure on the cable. This fact enabled the risk that the microphones capsules could be "budged-out" of place when the headphones are put on and the measurements taken, and return to its original position once the headphones are removed and the pressure is released, making the displacement impossible to detect in case of occurrence.

The ITD error in the front position was caused by non-optimal behaviour of some subjects during the measurement. It is believed that in few cases, the subject was not pointed correctly to the centre of the front speaker, or

the head-strap was not tight enough to discourage any head movement. The ITD differences between subjects range from 0 to 3 samples which at 48 kHz sampling rate meant approximately $63\mu s$ of delay between channels. This delay was too high to ignore and so it had to be corrected using zero-padding in one of the channels. The measurements for all the 11 participants were subjected to an ITD correction script.

An ulterior source of inaccuracies was the presence of reflective material (figure 4.15) in the anechoic chamber. Perfect anechoic conditions could not be ensured due to the presence in the chamber of the recording equipment and the experimenter himself. Also the laptop was a mild source of noise due to the cooling ventilator incorporated. Putting extra layers of absorption wedges partially stopped some of the reflections.

Figure 4.15: Reflective material in the anechoic chamber

A procedural inefficiency was due to the fact that the HpIRs measurement sessions for each participant did not occur on the same day of the HRIRs measurement sessions, which would have saved time. The reason lies is due to the fact that it was initially misunderstood that the headphone measurements had to be done individually for each subject rather than be recorded on a dummy head. Due to this reason, an extra week of

work (that could be avoided) had to be allocated for the HpIR measurements. Luckily it was not required to do these measurements in the same listening environments so, instead, the equipment was transported each time to the most convenient place for the participant and the recordings were done *in-loco*, in order to save time.

The total time spent on this stage lasted around two weeks for the HRIR measurements and 2 extra weeks for the signal processing. An extra week was used for the HpIR measurement, making a total of 5 weeks spent on this stage between the end of the university's spring term and the start of the summer term.

A further discussion on the impact of these problems on the rest of the project is given in the conclusion chapter.

Chapter 5

Listening Test

Contents

5.1 Stimuli preparation 138
5.1.1 Item List 140
5.1.2 Individualisation of stimuli 140
5.2 Listening Test 142
5.2.1 Preparation 144
5.2.2 Familiarisation process 144
5.2.3 Routine 145
5.2.4 Informal feedback 150

This chapter covers the third stage of the project. Once the individual HRTF filters were ready, the binaural audio material had to be prepared by convolving the filter pairs to monophonic material. The following pages illustrate the procedure followed for selecting appropriate test items and the implementation of the listening test. The listening test procedure is hereby described in all of its stages: preparation, subject briefing, execution and collection of data.

No particular problem was met during this stage. Some minor limitations on the choice of stimuli were faced (section 5.1) but ultimately had no influence on the implementation of the experiment; all the same, some

secondary analysis possibilities were precluded as a result of the lack of extensive variability of items. The only problem encountered that slowed down the project schedule was related to the lack of MATLAB toolboxes needed to control the soundcard through the laptop. After a thorough research into alternative methods, it was decided to change laptop with a MICROSOFT Windows 7 operating system which allowed the possibility to map and control the output port on the *MOTU Ultralite mk3* soundcard using a simple built-in function. The commented code use for mapping the soundcard in the training and experiment session is included in the supporting CD.

Approximately a week was dedicated to complete this stage, three days were needed to find the stimuli, test the code and prepare the environments and three days were used to run the test on the participants. 9 out of 11 participants, successfully completed the experiment in both sessions at the rate of three a day. At the moment of writing this report, subjects H and E have not yet taken part to the listening test stage due to unavailability to schedule a session in the month of May. If time permits, they will be recalled in a future moment. The lack of data for two of the participants should have been compensated by a higher number of sample data taken from the other subjects (using the formula in section 4.1.3). However, this problem was not expected and therefore less data than the optimal threshold was collected.

5.1 Stimuli preparation

To find material for the experiment stimuli, online resources were researched. It was required that the stimuli were dry (recorded in anechoic conditions) in order to allow for natural reverberation only. The limited availability of anechoic recordings suited to be used as experiment stimuli narrowed down the options. The initial difficulty was to find a good

variety of stimuli material to be used for the test; unfortunately the most interesting resources were only provided with echoic material. Secondary resources were found but the variety of stimuli types was not as varied as was hoped for. More specifically, the majority of the material found was related to individual classical instruments. Not many voice items or sound effects items ere found as well as no ensemble items. It was hoped that some "pop music material" could be included, but apart from item 47 (see table 5.1) none was judged to be suited for the test. A proper analysis of signal-dependency, in terms of the assessment of plausibility, would then not be very reliable in the absence of balanced sample sizes of different types of items.

The list of the items selected for the test is presented in section 5.1.1; the main source for the items was the OpenAIR library [56] supported by the *University of York*, although other sources of material, including online databases, were found. Other criteria used in choosing the material was variability of sounds, sound quality, and ecological viability (sounds that would exist in real life, i.e. no noise train signal). Instrument signals were chosen between strings, woodwind, percussion and brass in order to cover as much frequency range as possible. Voice speech was included in English and German for male and female versions. Some suitable sound effects were found across online resources and chosen for their "peculiarity".

Once the stimuli material was collected it was a simple task to adapt it for the experiment. Each item was loaded into a MATLAB workspace, rendered to a monophonic version (playable by single loudspeakers), resampled at 48kHz and finally normalised to max the amplitude at the value of 1. Each item was then audited and truncated to a shorter duration length as it was pointed out that longer signals would have required the subjects to keep the head still for longer time during the listening test routine. The truncation points were chosen with the intent of creat-

ing acoustically interesting phrases out of the available material. When a truncation would not have been possible without creating artefacts, a ramped linearly decaying amplitude envelope was applied on the signal. All the items taken from [57] were found to be noisy in the low frequency region. The material was loaded into a DAW (AUDITION) and the SNR was improved by using a noise-floor automatic reduction plug-in.

5.1.1 Item List

A copy of all the dry items can be found in the supporting CD. Sources for the audio material were the following:

1. OpenAirLib [56]
2. EBU Sound Quality Assessment Material CD [58]
3. Aalto Mediatech University [57]
4. University of Central London [59]
5. FreeSounds.org [60]

Table 5.1 illustrates a list of all the dry items used for the experiment.

5.1.2 Individualisation of stimuli

For each subject, all items were rendered into individually tailored binaural audio, ready to be presented in the listening test. Each item was convolved, using a fast frequency-domain convolution algorithm, with each processed and equalised HRTF pair. The audio was stored in a session folder prepared individually for each participant. The total number of convolutions was 200 per subject: 50 simulations for the chamber's

Table 5.1: List of Items

Number	Type	Source	Duration (s)
1	Voice (Female, Opera)	3	00:06
2	Voice (Female, Opera)	3	00:04
3	Voice (Female, Opera)	3	00:06
4	Voice (Female, Opera)	3	00:03
5	Instrument (Clarinet)	3	00:08
6	Instrument (Clarinet)	3	00:03
7	Instrument (Clarinet)	3	00:03
8	Instrument (Bassoon)	3	00:02
9	Instrument (Bassoon)	3	00:04
10	Voice (Male, German speech)	2	00:04
11	Instrument (Bassoon)	1	00:05
12	Instrument (Bassoon)	1	00:05
13	Instrument (Bassoon)	1	00:03
14	Instrument (Bagpipe)	1	00:04
15	Sound Effect (Rifle loading)	5	00:02
16	Sound Effect (Female Laughter)	4	00:03
17	Instrument (Flute)	1	00:05
18	Instrument (Clarinet)	1	00:06
19	Instrument (Clarinet)	1	00:07
20	Instrument (Bassoon)	1	00:07
21	Instrument (Clarinet)	1	00:06
22	Instrument (Flute)	1	00:08
23	Instrument (Drums)	1	00:02
24	Instrument (Drums)	1	00:02
25	Intrument (Clarinet)	1	00:07
26	Intrument (Cello)	1	00:05
27	Voice (Female, Opera)	1	00:04
28	Instrument (Trumpet)	1	00:02
29	Instrument (Trumpet)	1	00:03
30	Instrument (Trumpet)	1	00:06
31	Instrument (Trumpet)	1	00:08
32	Instrument (Trumpet)	1	00:10
33	Instrument (Trumpet)	1	00:08
34	Instrument (Trumpet)	1	00:05
35	Instrument (Trumpet)	1	00:06
36	Instrument (Trumpet)	1	00:05
37	Instrument (Bassoon)	1	00:04
38	Instrument (Bassoon)	1	00:04
39	Voice (Female, English speech)	1	00:06
40	Voice (Female, English speech)	1	00:03
41	Instrument (Bassoon)	1	00:03
42	Instrument (Viola)	1	00:08
43	Instrument (Viola)	1	00:05
44	Sound Effect (Keyring)	5	00:05
45	Sound Effect (Sneezing)	5	00:03
46	Instrument (Drums)	5	00:05
47	Instrument (Synthesiser)	5	00:05
48	Voice (Male, Scream)	5	00:05
49	Voice (Female, English speech)	2	00:04
50	Voice (Male, English speech)	2	00:08
51	Voice (Female, German speech)	2	00:04

front speaker, 50 simulations for the chamber's rear speaker, 50 simulation for the room's front speaker and finally 50 simulations for the room's rear speaker. The spatialised audio was labelled using the following format: *"NAME_ROOM_POSITION_ITEM.wav"*,

The supporting CD contains the stimuli prepared for a specific subject (subject A) who was considered to be a particularly "good listener" due to the fact that his ears where the most apt for inserting the microphone capsules. This had an impact on the subject's own HRIR recordings as they where the most consistent across subjects (zero ITD, and negligible ILD).

5.2 Listening Test

The listening test environments had to be set-up in a similar way done in the measurements. Having changed the soundcard to a different model, meant that only one unit was available. Furthermore it was not an easy task to control the soundcard through MATLAB. The only possibility to do so on a MAC OSX machine, was to use the Data Communications Toolbox which had to be purchased. To tackle this problem, a laptop using WINDOWS 7 operating system was made available by the department. To connect the soundcard, an output mapping function allowed to define IDs for the output ports and therefore control whether the sound output had to be played by the speakers or by the headphones. The other equipment used was largely identical to the one used in stage 2 and shown in figure 4.5, with the exclusion of the recording equipment. A typical set-up for this stage can be seen in figure 5.1.

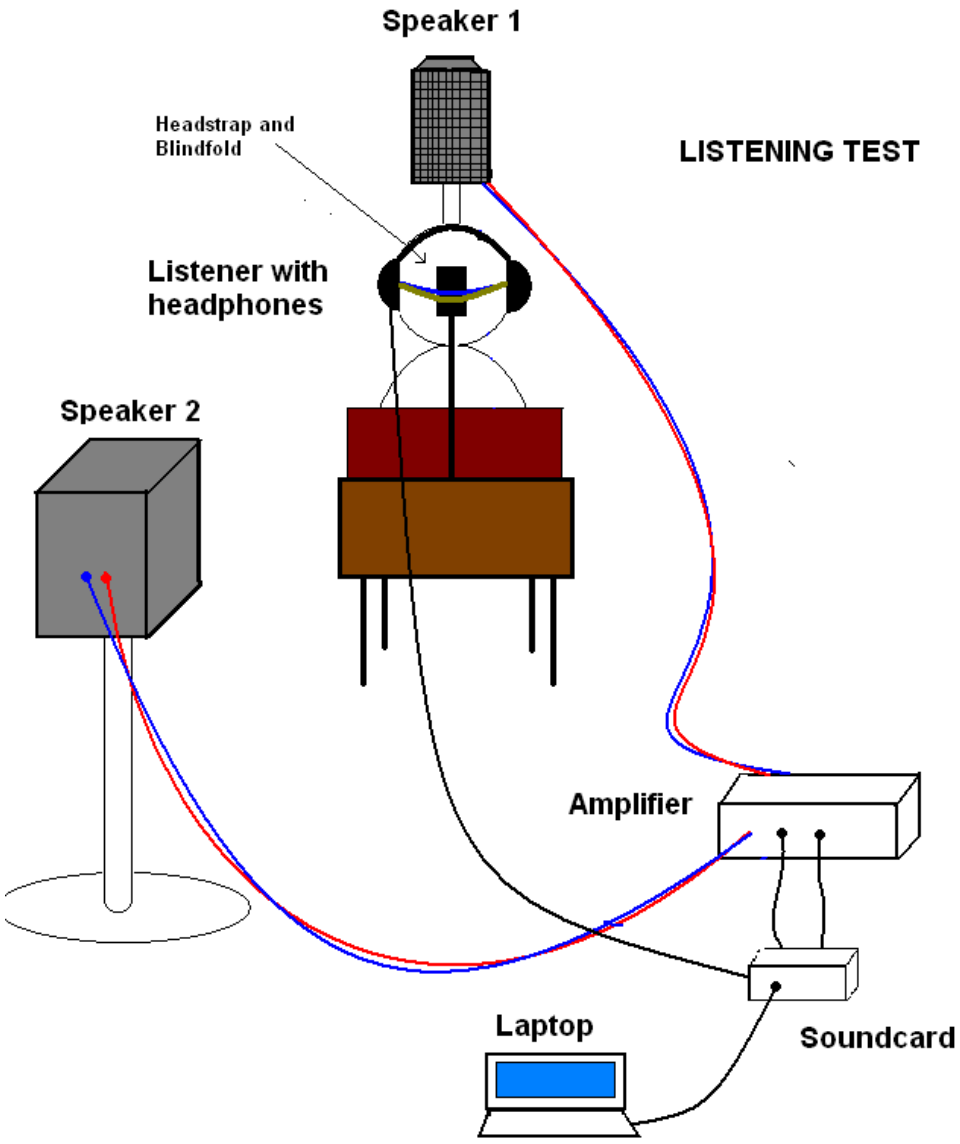


Figure 5.1: A sketch of the listening test, showing the equipment and the connections

5.2.1 Preparation

In order to avoid the possibility of listeners being able to discriminate the provenience of the sound sources by the volume difference, it was necessary to equalise the perceptual loudness between speakers and headphones. With the help of an external collaborator, the author adjusted the volume of the loudspeakers until an equal perceptual loudness was experienced between the real and simulated stimuli. This was verified over a number of different kinds of items selected from table 5.1. Once an optimal setting was found, the external collaborator was asked to repeat the procedure using his own perception. The collaborator agreed on the level settings chosen by the author, de facto rejecting the need for further verification. Preliminary tests ensured that the volumes were consistent throughout all stimuli and signals were correctly routed by the soundcard to the output possibilities. The training routine code (section 5.2.2) and the listening test routing code (section 5.2.3) were confirmed to behave as expected. It was also checked that no audible artefacts in the stimuli were present by using the binaural stimuli produced with the author's own equalised HRTFs but non was reported. The same testing procedure was applied for both environments. For the sake of ensuring consistency, the volume levels for all the loudspeakers were recorded as well as the volume level for the headphones and kept untouched for the whole stage.

5.2.2 Familiarisation process

Subjects were contacted and a schedule of sessions was established. For the reasons previously explained, a familiarisation session was needed in order to avoid people mistaking the effect of wearing headphone as artefacts, thus artificially increasing plausibility. Subjects were firstly briefed about the experiment and instructed on what their task was. The exact

instructions given to participants can be found in the appendix.

The training session started by leading the subject to the environment chosen for the first session, the subject was shown the exact position of the loudspeakers and subsequently seated on the chair, no strap nor blindfold was applied at this stage. The subject was presented with a signal example played from the loudspeakers (Item 51), this item would not have been presented again in the actual tests as its difference between real/simulation was now spoilt. The participant was instructed to put the headphones on and off as the signals were played from the real sources, in order to let him observe the acoustic effect of wearing headphones. The same item was then played through headphones for both the virtual positions. The participant was allowed to re-listen to all the real and simulated position as much as he liked until he felt ready to start the test. When changing session to the other environment the training was repeated in the same way. A training MATLAB script was used to control the output source and the position of the item presented, the code can be found in the supporting material CD.

5.2.3 Routine

After the training, the subject was blindfolded and his head strapped to the chair's headrest by the experimenter, making sure that neither the blindfold or the head-strap would rest on the ears, but above them. Participants were finally reminded to keep their head still and avoid any head movement as much as they could. Figures 5.2, 5.3 and 5.4 show different view angles of how the subject (who agreed to be shown in this report) was prepared for the listening test. The configuration shown in figure 5.1 was then achieved.

The experiment routine described in the previous section was coded in MATLAB and run after the training session was completed. The first

Figure 5.2: Front view of the subject with blindfold and strap in the anechoic chamber

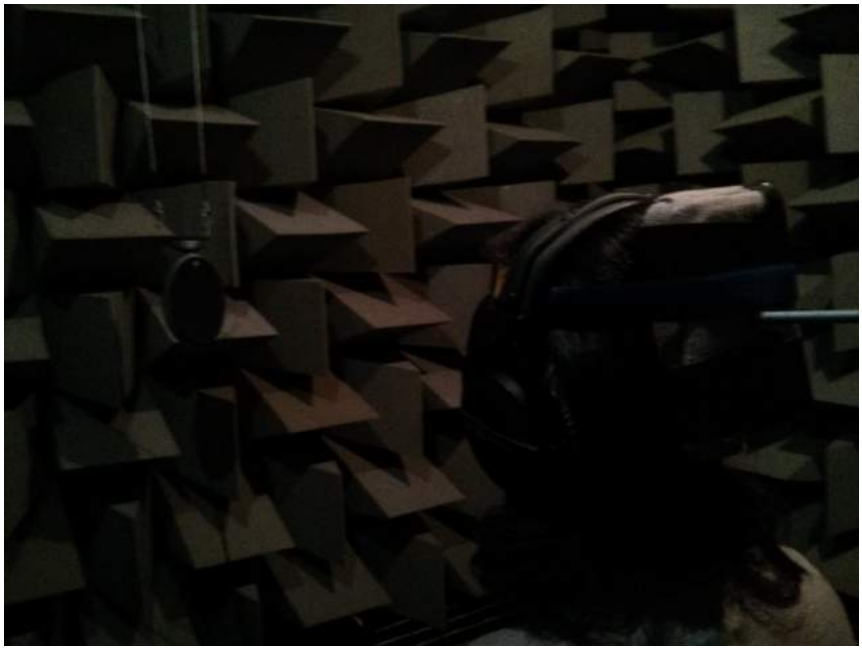


Figure 5.3: Back view of the subject with blindfold and strap, front speaker is visible in the background

Figure 5.4: Lateral view of the subject with blindfold and strap, rear speaker is visible in the background

step was for the experimenter to parse to the code the subject ID and the room session (anechoic or echoic). A total of 200 (50 items x4 possibilities) items per session was available for each subject:

1. Front position - real
2. Rear position - real
3. Front position - simulated
4. Rear position - simulated

The 200 items were ordered in an array list, a random number between 0 and 200 would then choose an item out of the list and present it to the participant. The presented item was then removed from the list in order to avoid repetition. Only 100 items per session were presented as it was decided that if an item was played for a specific position, its related alternative (between real and simulated) had to be taken off the list as well in order to avoid direct comparison of material. After each presentation, the subject was asked to judge plausibility by answering the question

"Where did the sound come from?" for which the answer had to be either *"Speakers"* or *"Headphones"*. Figure 5.5 shows a flow-diagram of the code prepared for the listening test routine.

Results were stored in the form of *"1"* and *"0"* answers(*1* for *Speakers*, *0* for*Headphones*) by the experimenter. Once the answer was parsed to the code, the next presentation could start. This whole process was repeated until the 100^{th} presentation was played. The session results were stored in a dedicated array of structures which was formatted to contain the following information (at a later stage these results were reformatted for the analysis code):

- Subject (A to K)
- Environment (1 for Anechoic, 2 for Room)
- Results:
  - Item (1 to 50)
  - Presentation order (1 to 100)
  - Position (1 for Front, 2 for Rear)
  - Source (TRUE for Real, FALSE for Simulated)
  - Answer (TRUE for Real, FALSE for Simulated)

The approximate running time for each session was 20 to 30 minutes according to the participant behaviour during the familiarisation procedure, for the experiment the subjects were instructed to answer quite instinctively and warned that presented signals could not be repeated. In order to allow participants to have a fresh mindset during the experiment, a resting time of arbitrary length during which the experimenter moved the equipment to the other test environment, was mandatory before the start of the next session. All together, the whole listening test lasted on average around 1 hour and 20 minutes per subject. Details of

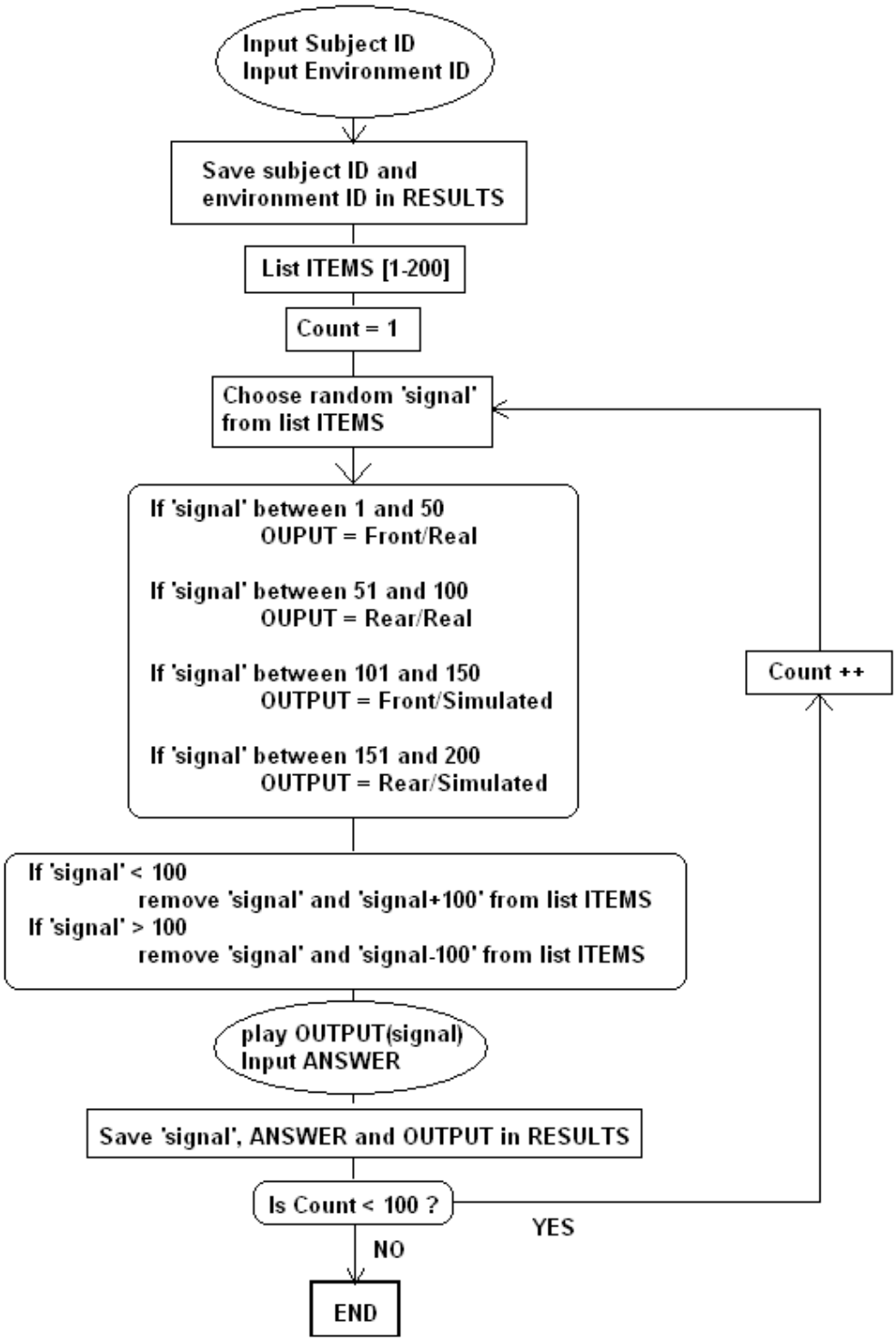


Figure 5.5: Flow diagram for the listening test routine

the instructions given to participants are found in the appendix. A copy of the code used for the experiment routine is on the attached CD with the supporting material.

5.2.4 Informal feedback

At the end of the session, participants were informally interviewed about their experience; not many direct questions were asked as instead the participant was mostly allowed to spontaneously describe his experience. The informal answers received exhibited a high degree of variability. A substantial amount of confusion between real and simulated was often experienced; some subjects even stated that they were almost completely guessing. The high variety of feedback responses was shown by the fact that 5 out of 9 subjects stated that the rear position was for them easier to judge than the front position, 3 out of 9 said the opposite and one subject said neither of them was easier. In fact, Subject A perceived almost the totality of the signals presented during the test as "real", which indicated the possibility of having reached a very high plausibility for his case. When asked for the reasons that made a position more easy to decide than the other, subjects' answers often mentioned spaciousness of sound; they could clearly localise one direction but they could not for the other. One subject stated that for the rear position he felt that "*the sound coming from the back loudspeaker was not able to overcome the sense of noise I feel when the ears are closed in an encapsulation, therefore I knew that it was not from the headphone*".

In regards of reverberation conditions most subjects reported the fact that the test was easier in the anechoic room than the listening room; two subjects proclaimed that neither was easier and none expressed the opposite. It was pointed out that in the anechoic chamber "everything sounded quite close to the ears".

People often mentioned timbre when asked about which aspect was important for them to base their decisions on. It was noted that a couple of subjects made their guesses almost perfectly wrong (i.e. answering "headphones" for every real stimuli and "speaker" for every simulated stimuli) in the anechoic chamber environment. They explained that they could hear a timbral difference between real/simulated. A possibility would be that the training session was misunderstood and the YES/NO answers were inverted; however, in both cases the anechoic chamber session happened after the listening room session where this consistent mistake was not reported. It is therefore believed that the perceived difference did not give away whether the source was real or simulated, as subjects could not correctly associate which was which, and instead judged simulated stimuli to be "more plausible" than real stimuli.

Mismatches in directionality were reported especially for the rear position. It was often stated that simulated sound sounded closer to the ear than real sound sources, but in three other cases the opposite was reported. An elevation increase with simulated sound was reported by two subjects, again, for the rear position. Once again, these factors were not enough for the participants to reach a high rate of correct answers.

It was not clear whether these answers were in any way associated with the problems described in section 4.3 as no relationships in the feedback between the *pre-freefield* and *post-freefield* groups were found in the informal interviews. More accurate conclusions were drawn by looking at the analysis results (see chapter 6). It was established that in case the differences was substantial, only the post-freefield group had to be considered reliable for the answers given despite the non-optimal sample population.

Some slight signal-dependency was noted by the experimenter. On average, speech signals were guessed correctly more frequently than music signals, probably due to the high level of sensitivity demonstrated to ma-

nipulation of speech signals [22]. One specific item (Item 44 - Keyring) was almost always correctly guessed by all the participants. It is not yet clear for what reason would this specific item be more easy to judge; the only thing that can be noted at this stage is the elevated presence of high-frequency content in the item. It was pointed out by one of the participants (subject B) that truncation artefacts were sometimes audible in the convolved material. According to this participant, these artefacts influenced his decisions and helped him identify the sound when it was originated from headphones. This was not reported by any other subject, even if they were directly asked if they experienced the same artefact. It was hypothesised that the problem could have been located in the subject's own HRTFs. A second hypothesis was that some truncation points used in the creation of stimuli might not have been precisely chosen and a very sensitive ear could have caught the clues; however ,further listening of the material by the author did not exhibit truncation artefacts to be significantly different enough between real and simulated stimuli.

Chapter 6

Analysis and Conclusions

Contents

6.1 Analysis 154
6.1.1 Results 156
6.1.2 Speaker positions 160
6.1.3 Correction groups 161
6.2 Conclusions 163
6.2.1 Comparison with previous studies 165
6.3 Discussion 166
6.3.1 Limitations 167
6.3.2 Further work 169

The final stage of the project was the Signal Detection Theory analysis of the collected data. The code for the analysis was programmed in PYTHON by Chris Pike and it was the same code used for the research described in [5]. The parameters were left unchanged as the same kind of analysis had to be computed.

This chapter illustrates the analysis results computed for different groups of variable. Results were grouped for separate analysis, the main focus in on the influence of different reverberation conditions in the plausibility assessment results. Secondary analysis looked at the influence of speaker

position and signal dependency. It is also inspected whether the types of level correction used had an effect

The conclusions provide an interpretation of the data in relation to the similar past studies conducted in the field. The plausibility results are directly compared to those achieved by Pike [5] and Lindau [27]. The impact of reverberation factors with individual HRTF measurements is compared with the results obtained by [26] and [22] who explored the factors that influence "externalisation" and localisation accuracy. A discussion on the results in relation to the problems encountered and the technical limitations that had an impact on the project illustrates why the conditions were not optimal and what it could have been done to improve validity of the results and the organisational aspects of the project. A 'further work' final section illustrates what could be done with additional work on the data collected.

6.1 Analysis

To run the SDT code written in PYTHON, a MAC machine with an inbuilt PYTHON compiler was used. The code read the results form *comma separated values* files (*.csv*) and produced the output analysis figures presented in this chapter.

The results data matrix was reformatted for the code to agree with the following specifications:

- Signal
- Position
- Was simulated?
- Answer

Example:

25_48kHz.wav,1,False,True

The reformatted results were stored in *.csv* files and placed in a specific folder based on the common variable. More specifically, a folder was created for each room environment and the associated results for all the subjects were put in the related folder. The analysis parameters used were left untouched from the ones given in the code. The parameters (explained in section **??**) were set in the code as $P_c = 0.55$ and $d'_{min} = 0.1777$ (P_c is the minimum value of P_{hit} for rejecting the minimum effect hypothesis)

The auditory scene can be judged to be plausible if the minimum effect hypothesis is rejected. More specifically, if the average sensitivity d'_{avg} is less than $d'_m in$, the auditory scene can be safely regarded as plausible. This happens when the probability of correct detection P_{hit} is within the value of $Pc = 0.55$ and 0.5 (which indicates that real/stimulated are completely indistinguishable). A bigger sensitivity value, would indicate the correct detection rate was high to the extent that the simulated signals have to be judged as not plausible. In other words, non-plausibility means that the signal sources were easy to discriminate correctly for the listeners. In [5] Pike also puts a secondary less strict threshold of plausibility at $P_c = 0.6$ which in a 2AFC test (binary choice test) means that $d'_{p_c 60\%} = 0.3583$. Sensitivity results within these values would indicate a semi-plausible scene, which could be accepted depending on the strictness parameters of the application for which the stimulated stimuli is created for. The secondary threshold values are also used in the code provided.

6.1.1 Results

The first analysis was conducted on the whole group of participants that attended the listening test (9 participants, E and H did not attend) separately for the two rooms. Figure 6.1 and figure 6.2 show the averages of individual sensitivities d'_i and biases β_i with 90% confidence intervals for the anechoic chamber and the listening room respectively. Exact values of d'_{avg} and β_{avg} for both rooms are shown in table 6.1. It is shown that the minimum effect hypothesis could not be disproved in both cases although it could have been if the $Pc = 0.6$ thresholds are used.

Table 6.1: Overall sensitivities and biases

Condition	d'	β
Minimum Effect	0.1777	
$P_{c60\%}$	0.3583	
Anechoic Chamber	-0.2777	0.9154
Listening Room	0.2229	1.1457

In the case of the anechoic chamber, the bias is smaller than 1 meaning that on average the subjects were biased towards reporting the stimuli as "simulated". This is reflected by the average sensitivity value being below zero suggesting that participants tended to judge "real" stimuli as "simulated" more often than the other way round. The average sensitivity is between $-d'_{min}$ and $-d'_{p_c60\%}$ meaning a close probability of participants "guessing". It is not clear how to interpret a negative sensitivity average, the bias indicates that people tended to judge sounds from the speaker as coming from the headphones, this might be the result of the absence of reflections in the chamber that reduces the "externalisation" sensation ([26]) to the point that external sounds are perceived as very close to the ears. In the context of spatial audio this situation can't be defined as "plausible" as the objective is to create a plausible 3D audio scene; however, the close proximity of the average sensitivity to d'_{min} shows that listeners were close to be guessing rather than identifying.

In the listening room the bias is reversed towards reporting stimuli as "real". Nevertheless the sensitivity is very close to d'_{min} showing that a plausible scene is almost reached. This indicates that a small but meaningful sensory difference was observed with "simulated" stimuli being recognised frequently enough to not pass the strict parameters set by Lindau [27]. Yet, the sensitivity value is quite below the secondary threshold $d'_{pc60\%}$ meaning that the small sensory difference exposed might not be very substantial.

Figure 6.2 shows the probability distribution of the two rooms. The average estimated sensitivity distribution is represented by the distance of the dotted distribution curve from the central curve. The bigger the overlap between the two distributions, the higher the plausibility. The figure shows that $\hat{d}'$ has a negative value for the anechoic chamber as illustrated in figure 6.1.

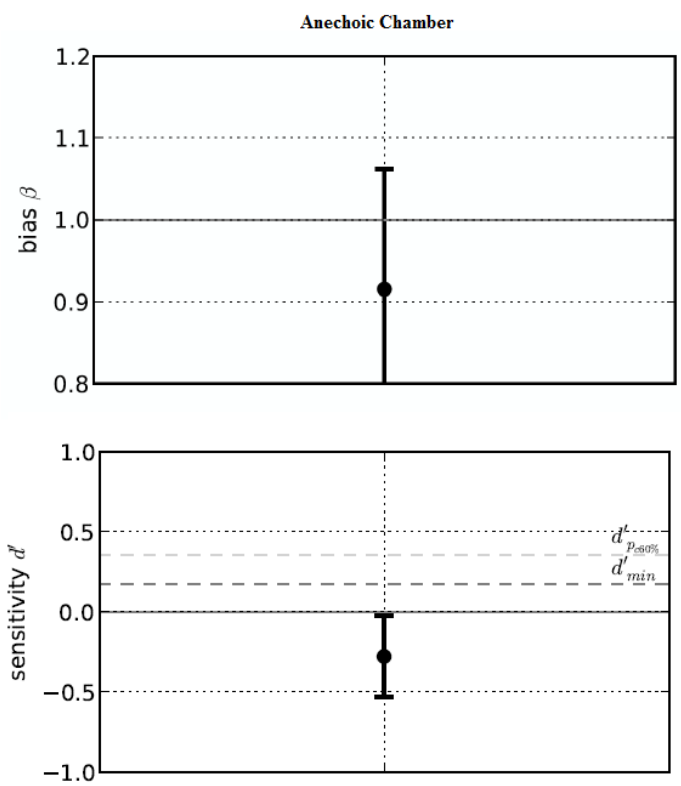


Figure 6.1: Anechoic chamber average bias β_{avg} and sensibilities d'_{avg}

As in [5] a one sided *t-test* was performed (using the same PYTHON

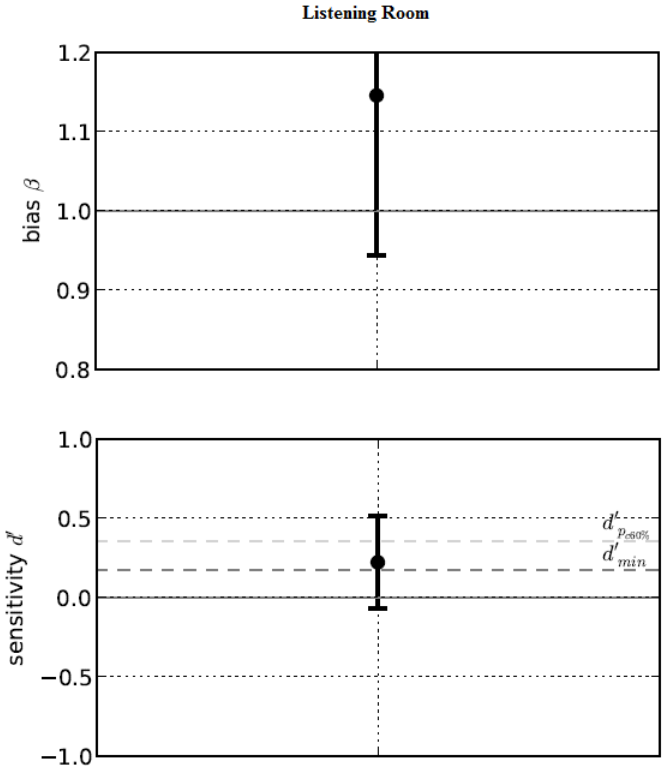


Figure 6.2: Listening room average bias β_{avg} and sensibilities d'_{avg}

code) to assess whether the individual sensitivity values were significantly greater than $d'_{min} = 0.1777$ and $d'_{p_c60\%=0.3583}$ using a 95% significance level. The test was also conducted to check if a significant difference of the test biases from the neutral bias $\beta = 1$ was present. Tables 6.2 and table 6.4 show that in the case of the listening room the values were not significantly greater than $d'min$. On the contrary, in the case of the anechoic chamber it is shown that the differences were significant (i.e. not due by chance). This is explained by the fact that the bias recorded was such to make the sensitivity d' negative, therefore quite far from d'_{min}. It would be appropriate to run another test using instead $-d'_{min} = -0.1777$. Due to time constraint this test has not been yet performed. No significant difference was recorded in either environment when looking at the bias level compared to the neutral bias.

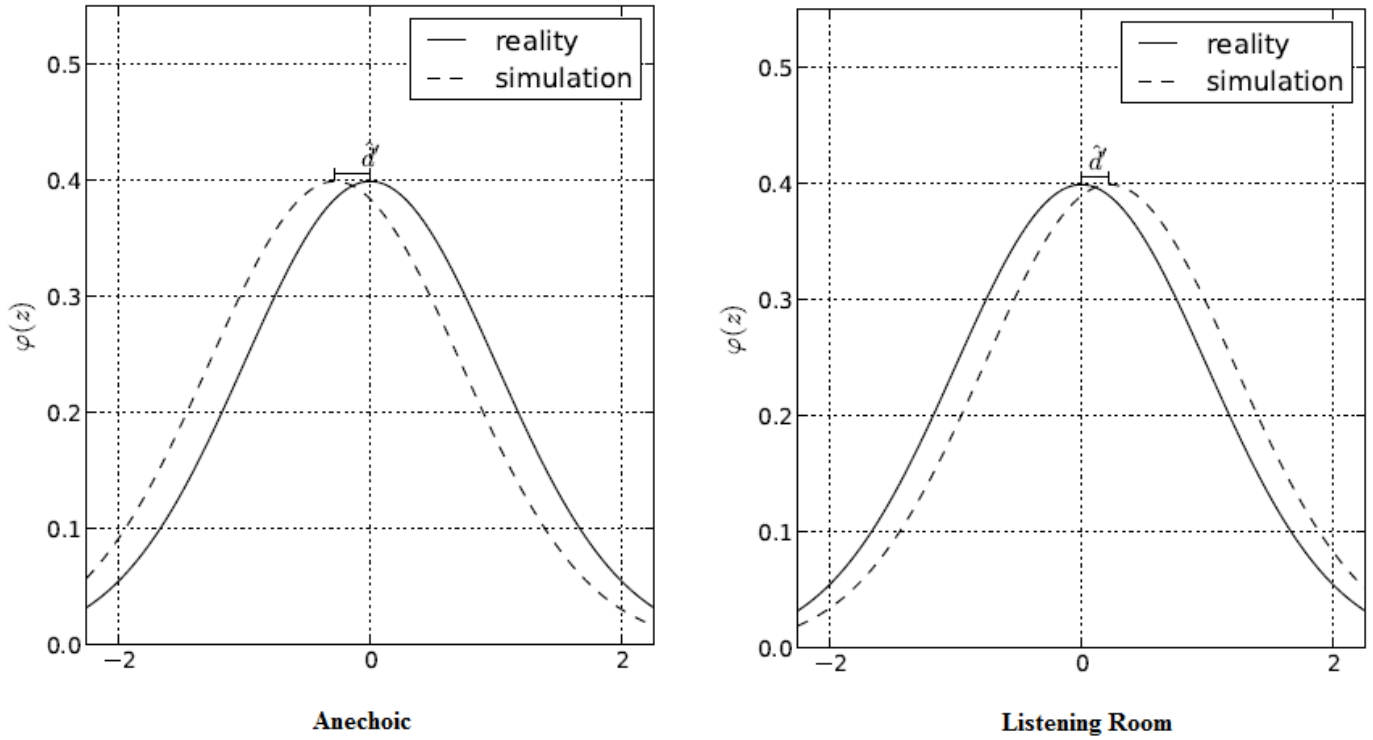


Figure 6.3: Probability density distributions of the equal-variance Gaussian SDT model, left for the anechoic chamber, right for listening room

Table 6.2: T-test for d'_{min}

Condition	d'	t	p
Anechoic Chamber	-0.2777	-3.3208	0.0053
Listening Room	0.2229	0.2885	0.3901

Table 6.3: T-test for $d'_{p_c 60\%}$

Condition	d'	t	p
Anechoic Chamber	-0.2777	-46379	0.0008
Listening Room	0.2229	-0.8653	0.2060

6.1.2 Speaker positions

Further SDT analysis was performed to inspect the effect of different speaker positions on the plausibility assessment in the two different environments. The analysis was run on all the subjects, first an overall average was calculated and then a distinct average for each speaker in each room. Table 6.5 and figure 6.4 show the SDT results for the 4 speakers (2 positions x 2 rooms).

It is shown that the levels of sensitivity differ quite consistently for the anechoic environment. While both positions exhibit a negative sensitivity value, the front position would actually pass a plausibility test if $d'_{min} = -0.1777$ is used as a reversed minimum effect hypothesis. For the reasons described in the previous session it is not clear how to interpret this result. Generally, a tendency towards an overlap of the distributions means the rate of guessing is greater (probability of error and probability of correct identification are the same). A shift of the distribution towards the left means that the probability of both type I (false positive) and type II (false negative) errors is greater that the probability of correct identific-

Table 6.4: T-test for $\beta = 1$

Condition	β	t	p
Anechoic Chamber	0.9154	-1.0711	0.1577
Listening Room	1.1457	1.3465	0.1075

ation.

For the listening room it can be safely stated that the front position passed the test of plausibility compared to the rear position given that $d'_{room/front} = 0.1456$ is considerably smaller than d'_{min}, in this case the minimum effect hypothesis can be rejected and the position judged as plausible.

Table 6.5: Position-related sensitivities

Position	Anechoic d'	Room d'	Anechoic β	Room β
FRONT	-0.1701	0.1456	1.0266	1.1571
REAR	-0.3607	0.1855	0.9378	1.1553

6.1.3 Correction groups

In order to verify the extent of using different correction methods on the individual HRTFs, the subjects were subdivided into a *pre-freefield* group (rms gain adjustment) and (post-freefield) group (calibration through freefield measurements). Subjects B,C,D and F were part of the first group; subjects A,G,I,J,K were part of the second group (subjects E and H did not participate to the listening test). It is shown in table 6.6 and in figure 6.5 that a noticeable degree of difference can be seen between the two groups in the case of the listening room environment. This raises the question of whether the measurements corrected using the rms energy energy method were valid or if that process might have degraded the spatial quality of the binaural material. However, this is not the case for the anechoic chamber where the pre-freefield results are slightly better than the post-freefield, thus making a conclusion on the impact of correction methods is impossible at this stage. Preliminary listening test showed no perceptual difference between the two correction methods so this difference might be the product of chance. A t-test might indicate whether the difference is significant but for the t-test to be reliable, a higher sample population is required.

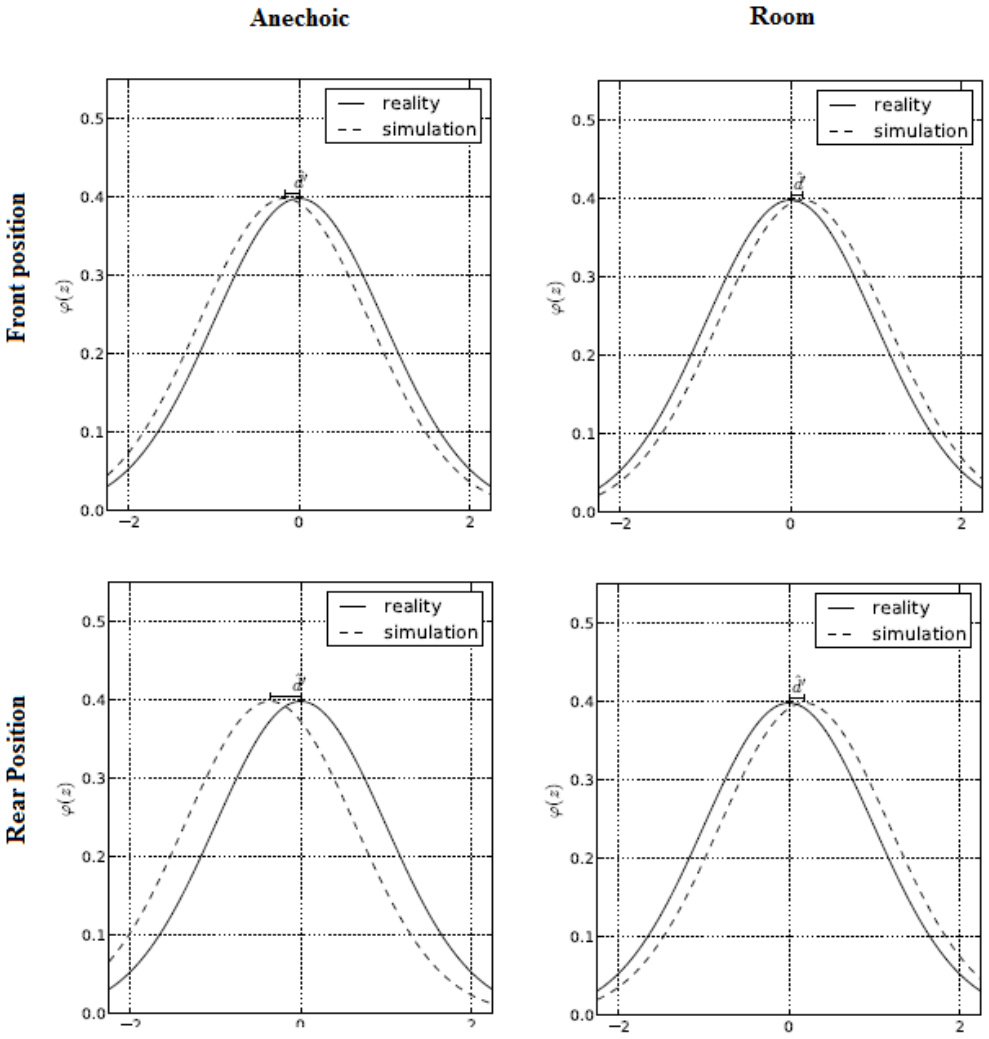


Figure 6.4: Average sensitivities distributions for the two different positions in each room

If we hypothesise that the post-freefield group hold more "legitimate" results than the other group, then the average sensitivity is smaller than d'_{min} meaning that plausibility might be reached in the listening room. The anechoic chamber does however exhibit a value between than $-d'_{min}$ and $-d'_{p_c 60\%}$ meaning that a meaningful sensory difference is present.

Table 6.6: Correction Groups sensitivities

Group	Anechoic d'	Room d'	Anechoic β	Room β
Pre-freefield	-0.2598	0.3209	0.9837	1.2230
Post-freefield	-0.2902	0.1445	0.8608	1.0838

6.2 Conclusions

The main focus of the analysis of results is put on the impact of different reverberant conditions on the plausibility assessment. The listening room environment shows that a strict test of plausibility was nearly passed and it would have been if only the front position is considered. As stated in [5] the level of required strictness has not yet been defined by the audio engineering community; different applications might, in fact, require different levels of plausibility, or even don't require it at all if realism is not the aim. The bias shown, indicates the average tendency of judging stimuli as "real". Lindau expresses in [27] that this might happen due to the unexpected realism of the simulation that is perceived.

Regarding the results associated with the anechoic chamber, the outcome is question of debate. Being the average probability of error higher than the probability of correct detection, it could be said that subjects were confused between presentations being real or simulated rather than convinced of the realism of the simulations. This might be explained by the fact that, as reported in the informal feedback, the sources sounded closer to the ears. As a result, despite of the anechoic chamber being describes ad an "easier" environment, real stimuli was more often judged as com-

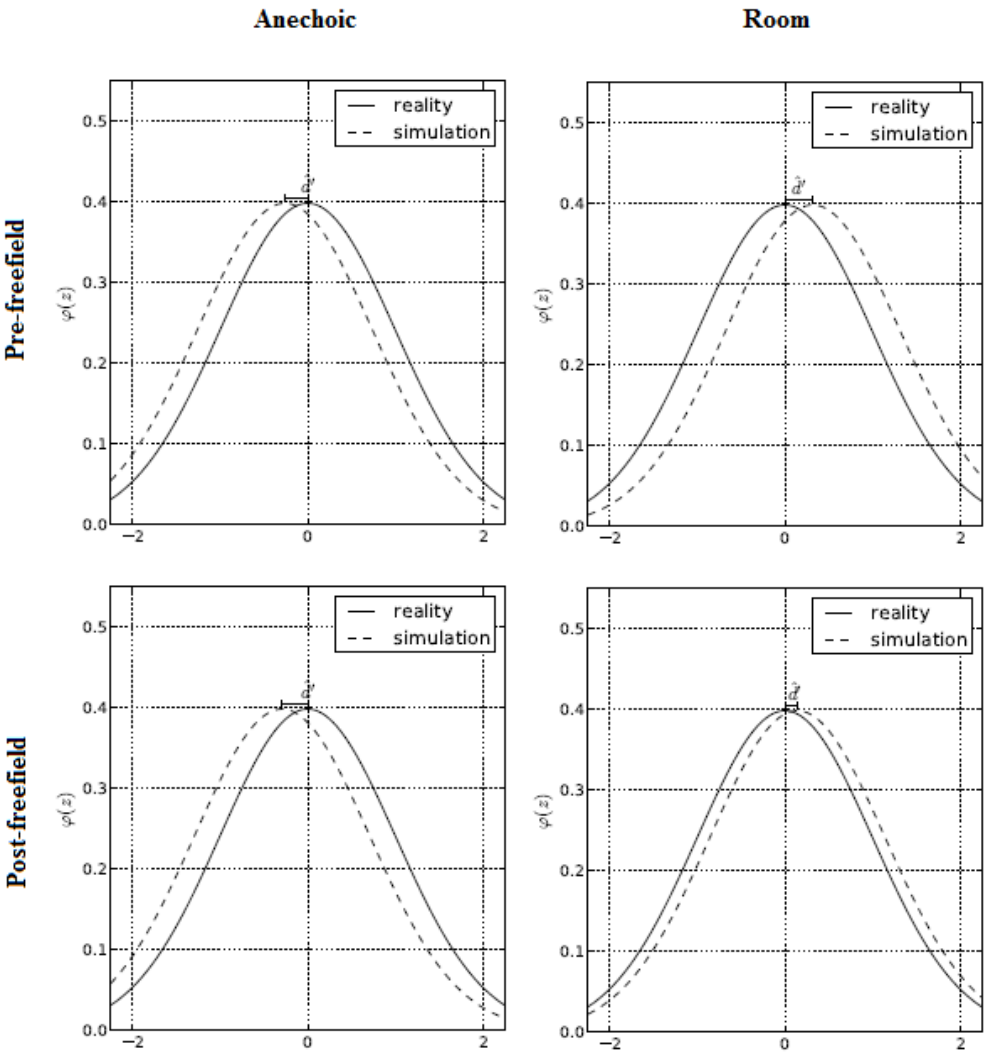


Figure 6.5: Average sensitivities distributions for the two different correction groups

ing from the headphones than coming from the loudspeakers. It seems that rather than simulated stimuli being "plausible", the real stimuli tended to be perceived as "implausible", perhaps the peculiarity of the anechoic conditions raised a certain perceptual confusion.

It is concluded that the level of externalisation in the anechoic conditions, despite the use of individual equalised HRTFs, is quite poor in comparison to reverberant conditions. On the other hand, reverberant conditions and individualisation bring a near-plausible level of realism which is further confirmed by the rejection of the minimum effect hypothesis in the case of the front speaker. The two hypothesis described in the first chapter then are both revealed to be true: a substantial difference between the two environments is reported and the echoic conditions influenced the binaural stimuli to be "more plausible".

Section 6.3 opens a discussion upon the validity of these conclusions and the level of confidence that can be put on these results.

6.2.1 Comparison with previous studies

The experiment portrayed in this project differed in some key aspects compared to the experiments of [27] and [5] that served as a model. The first difference was the lack of head-tracking and related technology. Secondarily the subjects used in those experiments were experienced in the audio technology field, while in this experiment, only 1 out of 9 participants had previous experience with spatial audio. This might be the reason of why a sensitivity bias has been found in this experiment while none has been found by Pike and Lindau.

This experiment was more similar to that of Pike [5] than that of Lindau [27]. Pike used a smaller room environment, rather than an auditorium, similar to the reverberant environment used in this experiment. In fact,

results do agree as the echoic room almost achieves plausibility with a d'_{avg} of 0.2229, while in [5] $d'_{avg} = 0.1954$. It would appear that the use of individual HRTF recordings and individual headphone equalisation filters did not bring an improvement in the assessment. As found in [5] the speaker position had an effect on sensitivity, the rear speaker was in both rooms, easier to judge than the front speaker. The rear speaker also had a non-zero elevation which was defined as another source of increase in sensitivity in [5].

It appears that, as found by Völk [26], the use of darkness conditions helped to increase the plausibility for the front position as shown in figure 6.4. It is also agreed with [26] and [22] that reverberant conditions yielded better externalisation effects on the perception of listeners.

6.3 Discussion

This section opens a discussion upon the validity of these conclusions and the level of confidence that can be put on these results.

The results obtained by the analysis on the different correction groups question the validity of the results and decrease the level of confidence of the drawn conclusions. It is not evident whether the problems in the measurement stage were influential enough to have compromised the results on the whole group. It is safe to assume that the results extracted from the *post-freefield* subjects are confidently reliable due to the certainty of having performed a correct measurement procedure for those participants. However the optimal sample population defined by [27] was not reached and it would be further decreased by looking only at the post-freefield group. Section 6.3.2 discusses what future work could be done to further explore this aspect.

Looking only at the *post freefield* results, it would seem that the test in the listening room environment did achieve the required plausibility. If these results had to be confirmed, the introduction of individual stimuli combined with a reverberant environment would bring an improvement from the test conducted in [5]. It is debatable whether this improvement would be significant enough to actually make a difference; a plausibility test between individual versus non-individual stimuli might help establish this possibility.

The lack of data for two of the subjects should have been compensated by a higher sample population (i.e. more stimuli presentations per subject). Considering that 9 subjects out of 11 showed up for the listening test, an optimal minimum sample population of $N_{opt} = 1077$ could have been reached by having each subject assess 120 signals in each session (120*9 = 1080). The non-availability of the two subjects was not predicted as it was communicated last-minute in spite of having previously agreed on the test schedule with all the participants.

6.3.1 Limitations

The project was limited by several factors. The main source of non-ideal conditions for making this project less reliable was the lack of professional equipment more suited for this kind of test. Head-tracking technology could not be used due to the high cost and complexity of such a system which would have exponentially increased the amount of work needed for this project. Also, the headphones chosen for this test were a second choice; the STAX headphones used in [27] and [5], renowned to be acoustically relatively transparent, were not available in the department and could not be purchased with the project budget only.

It was not predicted in the very early stages that a phantom power supply should have been built from scratch for powering up the electret micro-

phones. Despite the simplicity of this unit, the technical problems encountered in the implementation of this unit (section 3.3), considerably slowed down the project during the spring term, taking away time that could have been use to better prepare the subsequent stages. In fact, by the time that the problems with the soundcard gain settings were found (section 4.3), it would have been appropriate to re-measure some of the subjects but the tight schedule and the little available remaining time did not permit to organise further re-measurement sessions.

Should this experiment be reimplemented, special care will have to be taken for the delicate measurements procedure in order to get high-quality recordings. Many encountered problems could be avoided by making immediate use of a stepped soundcard where the channel gains can be easily matched, yet, this might not be enough to ensure that the gains are equal. In fact, manufacturing defects or components' value inaccuracies could influence the signal gain at the microphone itself or in the power supply unit. A correct gain calibration procedure could be achieved by using an acoustical signal as reference. A simple way to do that involves the use of a single-frequency sinewave signal to be played by a loudspeaker; the channel levels of the microphones should then monitored using a software mixer and finally, the knob-controls of the soundcard should be adjusted until the signal levels in the mixer match in decibels.

Furthermore, the different ear morphology of participants, often asymmetric, could not guarantee a correct insertion of the microphones. In fact, the insertion of the capsules had to be attempted several times and in most cases could not go very deep, also to avoid physically hurting the subject. This would also inequalities and asymmetries in the recordings, further influencing the levels between channels. On the other hand, it is believed that the directionality of the microphone in the ear should not have mattered due to the omnidirectional characteristics of the electret capsules. To avoid these asymmetries a possible solution would be to

select participants based on the ear canal's sizes using a lengthy procedure but this would mean to conduct a discriminative test which does not reflect the response of an average consumer group. Another solution would be to prepare specifically tailored ear plugs for each subject which could fit their ear canals optimally or more practically produce different sizes of sponges from a variety of ear-plugs sizes.

6.3.2 Further work

More exploration of the results could be achieved by analysing the presence of significant differences using further *t-tests*. Due to time constraints, it was not possible for the author to learn the Python code provided by BBC R&D and change it for running different tests. Further work in this direction might reveal if the differences between the *pre-freefield* and the *post-freefield* groups are significant enough to justify the repetition of the experiment with more care or if the difference is caused by chance.

Signal-dependency analysis has not been performed due to the unbalance of signal types in the test and time constraints. Performing this analysis could confirm what has been gathered informally from the listening test: stimuli such as voice or sound effects were more easily detectable than music stimuli, therefore less plausible. It would be interesting to further explore whether the signal-dependency, if confirmed, presents a different impact on the plausibility assessment according to the reverberation environment.

The anechoic conditions represent some kind of contradiction to the assessment of plausibility as they do not represent a situation where most listeners would find themselves into while experiencing binaural audio. Being this kind of environment unrealistic in a real-life scenario, it is hypothesised that participants could not relate to an inner reference related to a past experience of a real auditory scene in anechoic conditions. This

hypothesis would also explain the bias towards judging stimuli as "simulated". The SDT analysis further showed as "implausibility" of real sound sources was the effect achieved. For this reason, this experiment would not suggest exploring anechoic conditions further. On the other hand, it would be worth, especially due to the results seen in figure 6.5, to further investigate and compare the use of individual binaural stimuli compared to artificially-created binaural stimuli in the context of the assessment of plausibility. An interesting test would be that of comparing, always in the context of plausibility, real multichannel sound-scenes with virtual binaural sound-scenes as this could be the situation where most applications will aim for in the future.

Bibliography

[1] Ville Pulkki. *Spatial sound generation and perception by amplitude panning techniques*. PhD thesis, Helsinky University of Technology Laboratory of Acoustics and Audio Signal Processing, 2001.

[2] Rozenn Nicol, Laetitia Gros, Cathy Colomes, Markus Noistering, Olivier Warusfel, Helene Bahu, Brian FG Katz, and Laurent SR Simon. A roadmap for assessing the quality of experience of 3d audio binaural rendering. *Proc. of the EEA Joint Symposium on Auralisation and Ambisonics, Berlin*, pages 100–106, March 2014. Orange Labs, LIMSI-CNRS, UMR STMS IRCAM-CNRS-UPMC.

[3] Edgar Y. Choueiri. Optimal crosstalk cancellation for binaural audio with two loudspeakers. *Princenton University*, 2008. http://www.princeton.edu/3D3A/Publications/BACCHPaperV4d.pdf.

[4] Robert Campbell. *Monaural and Binaural Level Cues: A behavioural and Physiological Investigation*. PhD thesis, University of Oxford, 2006.

[5] Chris Pike, Frank Melchior, and Tony Tew. Assessing the plausibility of non-individualised dynamic binaural synthesis in a small room. *AES 55th International Conference*, 2014. BBC Reseach and development, Salford, UK.

[6] Bo Gehring. Why 3d sound through headphones? , 1997. http://www.fp3d.com/papers/WhyHeadphones.pdf.

[7] Beyerdynamics. *Headzone - The virtual control room*.

2014. http://europe.beyerdynamic.com/headphones-headsets/headphones-headsets/headzone/headzone-technology.html.

[8] J. Heggestuen (Business Insider). One in every 5 people in the world owns a smartphone, one in 17 owns a tablet, 2013. http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10.

[9] eMarketer. Smartphones users worldwide will total 1.75 billion in 2014, 2014. http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536.

[10] Jeroen Breebaart, Jürgen Herre, Lars Villemoes, Craig Jin, Kristofer Kjörling, Jan Plogsties, and Jeroen Koppens. *Multi-channel goes Mobile: MPEG Surround Binaural Rendering*. AES 29th International Conference, Seoul, Korea, 2006.

[11] Juha Vilkamo, Bernhard Neugebauer, and Jan Plogsties. Sparse frequency-domain reverberator. *J. Audio Eng. Soc*, 59(12):936–943, 2012.

[12] MPEG Surround. *MPEG Surround - Inventors*. 2013. http://www.mpegsurround.com/inventors.html.

[13] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno. Reconstructing head models from photographs fir individualised 3d-audio processing. *Computer Graphics Forum*, 27(7):1719–1727, October 2008.

[14] Parham Mokhtari, Hironori Takemoto, Ryouichi Nishimura, and Hiroaki Kato. Computer simulation of hrtfs for personalization of 3d audio. In *Proceedings of the 2008 Second International Symposium on Universal Communication*, ISUC '08, pages 435–440, Washington, DC, USA, 2008. IEEE Computer Society.

[15] Anthony I. Tew, Carl T. Hetherington, and Jonathan Thorpe. Morphoacoustic perturbation analysis. pages 867–872, 2012.

[16] Kall Binaural Audio. Kall binaural audio. http://www.kallbinauralaudio.com/.

[17] BBC. *BBC - Binaural Broadcasting*. 2013. http://www.bbc.co.uk/rd/projects/binaural-broadcasting.

[18] BR Klassik. *BR Klassik - Surround*. 2013. http://www.br.de/radio/br-klassik/service/empfang-und-technik/surround100.html.

[19] Jaime Sanchez and Mauricio Lumbreras. *Virtual Environment Interaction through 3D Audio by Blind Children*. PhD thesis, University of Chile, Department of Computer Science, 1999.

[20] Pavel Zahorik. Perceptually relevant parameters for virtual listening simulation of small room acoustics. *J. Acoust. Soc. am*, 126(2):776–791, August 2009.

[21] S. M. Kim and W. Choi. On the externalisation of virtual sound images in headphone reproduction: a wiener filter approach. *J.acoust. Soc. Am.*, 6(177):3657–65, 2005.

[22] Durand R. Begault, Elizabeth M. Wenzel, and Mark R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc*, 49(10):904–916, 2001.

[23] Jorgos Estrella. *Real time individualisation of interaural time differerences for dynamic binaural synthesis*. PhD thesis, TU Berlin, 2011.

[24] Henrik Mller, Michael Friis Srensen, Clemen Boje Jensen, and Dorte Hammershi. Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc*, 44(6):451–469, 1996.

[25] P. Minaar, S. Olesen, F. Christensen, and H. Møller. Localisation with binaural recordings from artificial and human heads. *J. Audio Eng. Soc.*, 2001.

[26] Florian Vlk, Fabian Heinemann, and Hugo Fastl. Externalization in binaural synthesis: effects of recording environment and measurement procedure. *The Journal of the Acoustical Society of America*, 123(5):3935–3935, 2008.

[27] Alexander Lindau and Stefan Weinzierl. Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica*, 98(5):804–810, 2012.

[28] Philip Mackensen. *Auditive localisation. Head movements, an additional cue in localisation*. PhD thesis, TU Berlin, 2004.

[29] Jens Blauert and Ute Jekosch. Sound-quality evaluation; a multi-layered problem. *Acta Acustica united with Acustica*, 83(5):747–753, 1997.

[30] E. H. A. Langendijk and A. W. Bronkhorst. *Fidelity of three-dimensional-sound reproduction using a virtual auditory display*, volume 107, pp 528-537. J. Acoust. Soc. Am, 1999.

[31] J. W. Strutt (Lord Rayleigh). On our perception of sound direction. *Philosophical Magazine*, 13:214–232, 1907.

[32] J. Angell and W. Fite. The monaural localization of sound. *Psychol.Rev*, 8:225246, 1901.

[33] David Heeger. Lecture notes. *New York University*, -. http://www.cns.nyu.edu/ david/courses/perception/.

[34] Benedikt Grothe, Michael Pecka, and David McAlpine. Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3):983–1012, 2010.

[35] C. J. Plack. *The sense of Hearing*. New York Psychology Press, 2005.

[36] H. Han. On the relation between directional bands and head movements. *92. AES Convention*, 1992.

[37] Neumann. *KU-100 Artificial Head*. 2014. http://www.neumann.com.

[38] Beck Binaural Blog. *Legacy Effects Blog*. 2014. http://legacyeffectsblog.com/beck-binaural-head/l.

[39] Jens Blauert and Robert A. Butler. *Spatial Hearing: The Psychophysics of Human Sound Localization by Jens Blauert*, volume 77. 1985.

[40] Tobias Lentz, Ingo Assenmacher, Michael Vörlander, and Torsten Kuhlen. *Precise Near-to-Head Acoustics with Binaural Synthesis*. 2006. https://www.jvrb.org/past-issues/3.2006/589/.

[41] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*, Feb 2000.

[42] Multiphonie. *Son 3D*. http://multiphonie.free.fr/son_3d.htm.

[43] F. E. Toole. Inhead localization of acoustic images. *The Journal of the Acoustical Society of America*, 48(4B):943–949, 1970.

[44] G. Reid. *Synth Secrets, Part 22: From Springs, Plates & Buckets To Physical Modelling.* 2001. Sound on Sound - http://www.soundonsound.com/sos/feb01/articles/synthsecrets.asp.

[45] Sound on Sound. *YOU ARE SURROUNDED - Surround Sound Explained - Part 3.* 2013. http://www.soundonsound.com/sos/oct01/articles/surroundsound3.asp.

[46] Ville Pulkki. Compensating displacement of amplitude-panned virtual sources. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, Jun 2002.

[47] Gunther Theile and Helmut Wittek. Wave field synthesis : A promising spatial audio rendering concept. *Acoustical science and technology*, 25(6):393–399, nov 2004.

[48] Robert Henke. *Monolake WFS at Tresor, Berlin*. 2009. http://www.monolake.de/concerts/wfs.html.

[49] Harold Stanislaw and Natasha Todorov. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1):137–149, 1999.

[50] Torben Hohn, Alexander Lindau, and Stefan Weinzierl. Binaural resynthesis for comparative studies of acoustical environments. In *Audio Engineering Society Convention 122*, May 2007.

[51] F. Brinkmann and A. Lindau. On the effect of individual headphone compensation in binaural synthesis. *36th German Annual Conference on Acoustics, Berlin, Germany, 2010*, pages 15–18, 2010.

[52] Bruno Masiero and Janina Fels. Perceptually robust headphone equalization for binaural reproduction. In *Audio Engineering Society Convention 130*, May 2011.

[53] Andrea F. Genovese. Meng project literature and preparation. 2013. University of York.

[54] Eric W. Weisstein. Tangent. In *Wolfram Mathworld*.

[55] Senheiser. Ke4 elektret-mikrofonkapsel - datasheet.

[56] University of York. Openairlib anechoic recordings.

[57] Jukka Ptynen, Ville Pulkki, and Tapio Lokki. Anechoic recording system for symphony orchestra. *Acta Acustica united with Acustica*, 94(6):856–865, 2008-11-01T00:00:00.

[58] European Broadcasting Union. Users handbook for the ebu sqam cd. *EBU Tech 3253*.

[59] UCL. Sounds of ucl: anechoic chamber laughter.

[60] FreeSound.

Appendix A

Instructions for participants

This section of the appendix includes a copy of the exact instructions given to participants for the measurement stage (stage 2) and experiment stage (stage 3). An example of the safety agreement document signed by one of the participants, who agreed to have her name published, its included.

Instructions for listening test experiment

Part I of II

During this session you will have some acoustic measurements made on your ears in two different listening test environments (one listening room and one anechoic chamber).

You will be guided into the first room and seated in the experiment chair by the experimenter. The microphones used for the experiments are tiny capsules that are wrapped in clean sponge and they have to be inserted in your ear canals like a pair of earplugs. The experimenter will seat the capsules in your ears and let you push them inside; if you wish you can allow the experimenter to push them in for you. You should not experience any discomfort, but if you do, please tell the experimenter and they will stop immediately.

Surgical tape will be used to secure the connecting leads to your neck and to prevent the capsules from falling out. A headstrap will be used to secure your head to the chair's headrest in order to encourage a fixed head position. Finally headphones will be placed carefully over your ears in a way that won't dislodge the capsules. This final preparation step will have to be repeated until it is completed successfully.

The measurement consists of playing two tones, one from each loudspeaker. It is of the utmost importance that you **keep your head as still as you can** while these signals are played. The measurement itself will take about 10 seconds. A second measurement may be taken in order to have some backup material in case something is wrong with the first measurement.

The exact same procedure will be repeated in a second listening test environment. In both environments, the experimenter will be nearby in the room all the time ready to deal with any troubles or questions you may have.

After the measurements you will asked a few questions for statistical analysis purposes.

Please don't hesitate to tell to the experimenter if anything makes you uncomfortable or if you have any questions at all. At any point during the experiment you are free to withdraw without having to explain the reason, in which case your data will be destroyed and not used for further analysis.

Thank you very much for taking part in this experiment, we really appreciate it.

Instructions for listening test experiment

Part II of II

During this session you will be brought in the same environments where the recording sessions happened, an anechoic chamber and a listening test room. In each room there are two speakers placed around you.

You will be showed the room and the setup and then be seated on the experiment chair. The experimenter will put headphones on you and run a test trial to show you what kind of signals you will have to expect.

After the trial you will have a head-strap put around your head, this will be used to secure your head on the headrest and make sure you don't move your head during the experiment. Finally you will be blindfolded and the headphones will be put back on.

The experimenter will be present with you in the room during the entire experiment session.

You will hear a sequence of signals and for each signal you hear, you will have to answer a simple question:

"Where did the sound come from?"

Please answer either **HEADPHONES** or **SPEAKERS** as you find appropriate

After your answer is noted, the next signal will be presented.

After the session is over, the whole procedure is repeated in the second room.

You will be played each trial once only. You may find it hard to decide whether the sound came from the loudspeaker or from the headphones but do not be concerned. If that happens just try to guess the answer in an instinctive way.

At any time during the test you are allowed to ask for a coffee break or a pause in order to keep yourself concentrated. If you want to stop the experiment, you can tell the experimenter. In case you do, you are not required to give any reason and your data will be eliminated.

Thanks for your participation!

Risk Assessment

Hazard	Severity (1 to 9)	Likeliness (1 to 9)	Prevention Action
Over-insertion of microphone capsules into ear canals	2	1	The subjects will be asked to insert the capsules in the ears by themselves, the experimenter will help only if allowed to and will stop at the first sign of pain or complaint. The ear sponges that wrap the capsules will be big enough to avoid falling into the canals.
Trip hazards over cables or chamber floor grid	1	1	Experimenter will be present all the time to guide the subjects in and out the experiment rooms

Health and Safety declaration

By signing this form I hereby understand the risks involved in the experiment and agree to participate. It will be possible for me at any time to withdraw from the experiment without having to explain a reason, in which case all measured data will be destroyed and not used for further analysis.

Name: Hikaru Takeuchi

Date: 17/04/14

Signature: Hikaru Takeuchi

Further Questions

Gender: female

Age: 22

Do you consider yourself an experienced listener?

No

Do you have any hearing impairment?

No

If yes, what type?

Appendix B

Supporting pictures

This section of the appendix included ulterior pictures of the listening environments and previous versions of the hardware that failed to meet the specifications and had to be reconstructed.

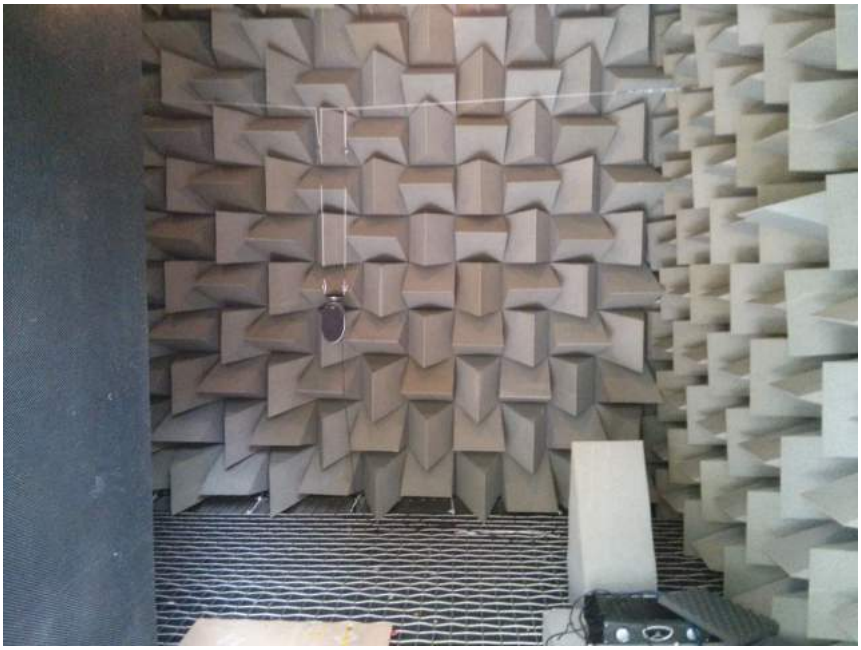


Figure B.1: A view of the anechioc chamber environment front-speaker

Figure B.2: A view of the listening room environment set-up

Figure B.3: First version of the expetiment chair

Figure B.4: Second version of the experiment chair when testing the anechoic chamber wood platform support

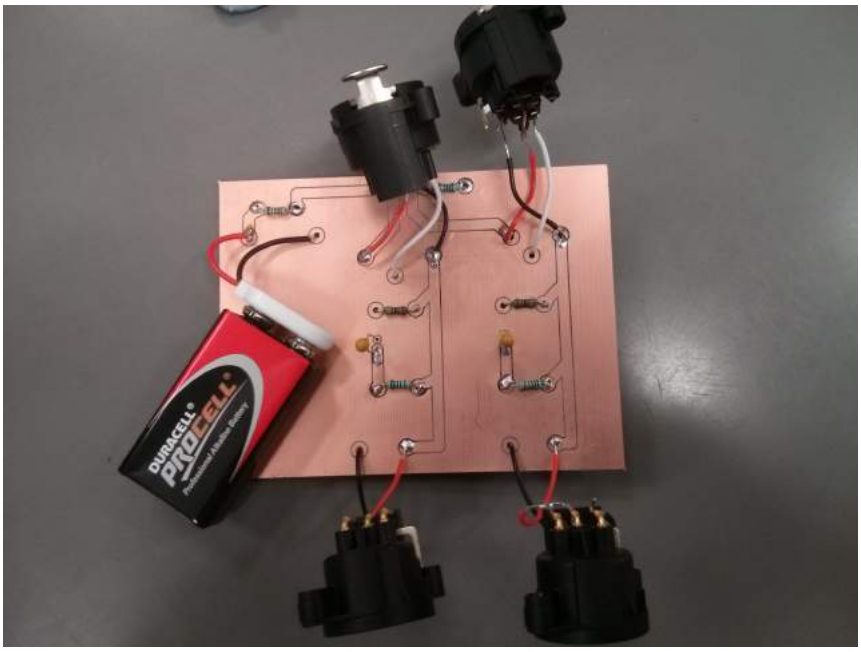


Figure B.5: First version of the phantom power supply circuit

Appendix C

Electret Microphone Capsules Datasheet

Datasheet for the KE4 Electret Microphone capsules by Senheiser.

From [55]

SENNHEISER

Industrie · Industry Information

Elektret-Mikrofonkapsel
Electret microphone capsule
KE 4

23631

Kurzbeschreibung
Elektret-Mikrofonkapsel (Druckempfänger) in Back-Elektrettechnik mit integriertem Impedanzwandler

Eigenschaften
Äußerst geringe Abmessungen
Weiter Übertragungsbereich
Hoher Geräuschspannungsabstand
Körperschallunempfindlich durch Back-Elektrettechnik
Niedrige Betriebsspannung

Besonderheiten
TO 18 – Transistor-Gehäuse

Ausführungen
KE 4-211-1 (Art.-Nr. 02014)
Anhebung im Bereich um 10 kHz. Daher besonders geeignet für Sprachübertragungen.

KE 4-211-2 (Art.-Nr. 02280)
Kapsel mit äußerst linearem Frequenzgang

KE 4-241-1 (Art.-Nr. 02419)
Geringe Stromaufnahme (< 40 µA). Besonders geeignet für Hörgeräte.

Weitere Ausführungen auf Anfrage

Brief description
Electret-microphone capsule (pressure receiver) in back-electret technique with integrated impedance transformer

Features
Very small dimensions
Wide frequency response
High signal-to-noise ratio
Insensitive to structure-borne vibration due to back electret
Low operating voltage

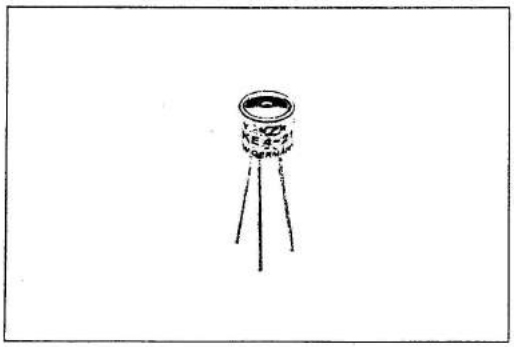
Special feature
TO 18 – transistor housing

Types
KE 4-211-1 (Art.-No. 02014)
Emphasis at approx. 10 kHz. Therefore, especially suitable for speech.

KE 4-211-2 (Art.-No. 02280)
Capsule with extremely flat frequency response.

KE 4-241-1 (Art.-No. 02419)
Very low current consumption (< 40 µA). Especially suitable or hearing aids.

Further types on request



SENNHEISER

Industrie · Industry Information

Elektret-Mikrofonkapsel
Electret microphone capsule
KE 4

Technische Daten KE 4-211-1, KE 4-211-2

(gemessen in Impedanzwandlerschaltung)

Wandlerprinzip	Kondensator-Druckmikrofon
Übertragungsbereich	20 bis 20 000 Hz
Feldleerlauf-Übertragungsfaktor im ebenen Schallfeld bei 1000 Hz (Empfindlichkeit)	10 mV/Pa ± 2,5 dB
Elektrische Impedanz bei 1000 Hz	ca. 1,5 kΩ bei $U_B = 5$ V
Minimale Abschlußimpedanz	4,7 kΩ
Geräuschspannungsabstand nach DIN 45590	ca. 58 dB
Speisespannung	+0,9 ... 15 V
Stromaufnahme	ca. 150 µA bei $U_B = 5$ V
Aussteuerbarkeit	je nach Betriebsspannung
Temperaturbereich: Lagerung	−20° C ... +70° C
Betrieb	−10° C ... +50° C
Klimafestigkeit (Lagerung)	bis +40° C und 90 % rel. Feuchte (SNP 51)

Abweichungen für KE 4-241-1

Elektrische Impedanz bei 1000 Hz	ca. 2 kΩ
Minimale Abschlußimpedanz	ca. 10 kΩ
Stromaufnahme	< 40 µA bis $U_B = 9$ V

Technical Data KE 4-211-1, KE 4-211-2

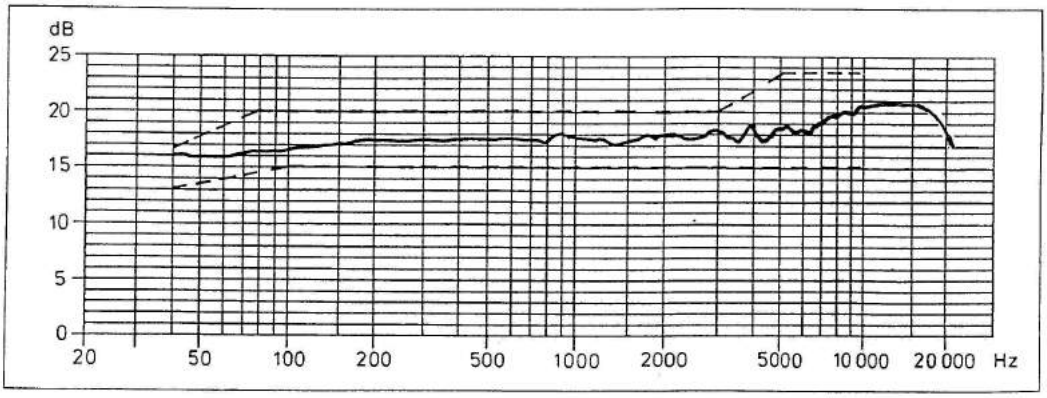
(measured in impedance transformer configuration)

Transducer principle	condenser pressure microphone
Frequency response	20 bis 20 000 Hz
Sensitivity in plane sound wave at 1000 Hz	10 mV/Pa ± 2.5 dB
Electrical impedance at 1000 Hz	approx. 1.5 kΩ at $U_B = 5$ V
Min. load impedance	4.7 kΩ
Weighted S/N ratio according to DIN 45590	approx. 58 dB
Operating voltage	+0.9 ... 15 V
Current consumption	approx. 150 µA at $U_B = 5$ V
Modulation range	depends on supply voltage
Temperature range: Storage	−20° C to +70° C
Operation	−10° C to +50° C
Climatic resistivity (storage)	up to 40° C and 90 % rel. humidity (SNP 51)

Specific data for KE 4-241-1

Electrical impedance at 1000 Hz	approx. 2 kΩ
Min. load impedance	10 kΩ
Current consumption	< 40 µA to $U_B = 9$ V

Frequenzgang KE 4-211-1/KE 4-241-1 — **Frequency response KE 4-211-1/KE 4-241-1**



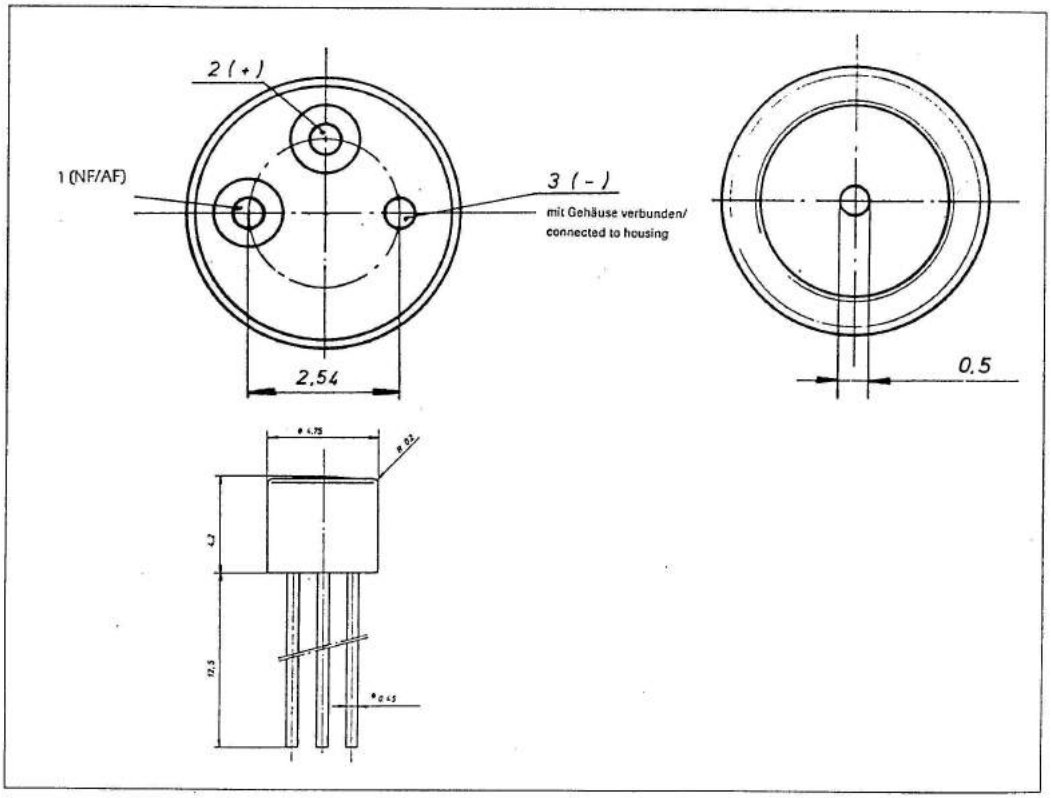
z.H. Herr Dr. Heinz

SENNHEISER

Industrie · Industry Information

Elektret-Mikrofonkapsel Electret microphone capsule KE 4

Abmessungen · Dimensions (in mm)

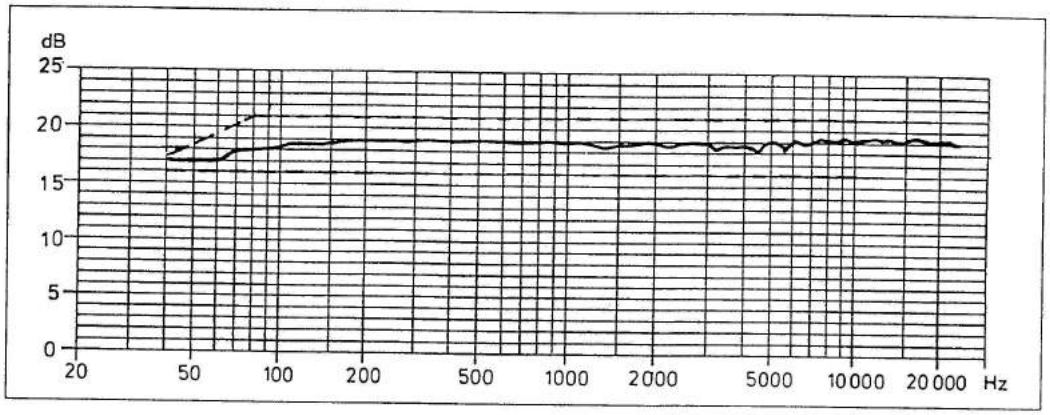


Industrie · Industry
Information

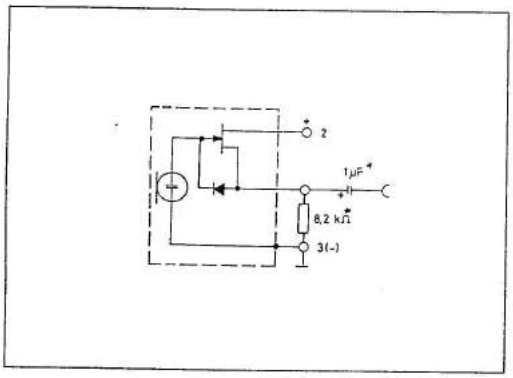
Elektret-Mikrofonkapsel
Electret microphone capsule
KE 4

Frequenzgang KE 4-211-2

Frequency response KE 4-211-2



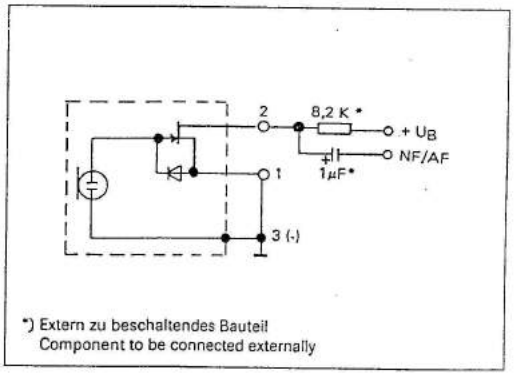
Impedanzwandlerschaltung
Impedance transformer configuration



Hinweis:
In dieser Konfiguration ist die Phasenlage positiv, d. h. ein positiver Druckimpuls erzeugt ein positives gerichtetes Ausgangssignal.

Note:
In this configuration the phasing is positive, i. e. a positive pressure impulse generates a positive output signal.

Verstärkerschaltung
Amplifier configuration



Hinweis:
1. In dieser Konfiguration ist die Phasenlage negativ, d. h. ein positiver Druckimpuls erzeugt ein negativ gerichtetes Ausgangssignal.
2. Der Feldleerlauf-Übertragungsfaktor erhöht sich um 10 bis 14 dB.

Note:
1. In this configuration the phasing is negative, i. e. a positive pressure impulse generates a negative output signal.
2. The sensivity increases by 10 to 14 dB.

Appendix D

Supporting Material

A CD is included in this submission in order to provide supporting material such as the stimuli audio and relevant MATLAB code.

The CD includes the following:

- Dry stimuli used for the listening test
- Rendered binaural stimuli of a particular "good listener" for both room environments (listener A)
- Significant MATLAB code used for all the project stages
- Headphone Equalisation MATLAB code and PYTHON Analysis code provided by Chris Pike, BBC R&D, *Media City UK, Salford*