Auralization platform for hearing aid development

Master's Thesis

19gr1077

Aalborg University
Institute of Electronic Systems

This report is compiled in LaTeX. Additionally Mathworks MATLAB® , Inkscape, and Xfig are used to draw figures and charts.

Institute of Electronic Systems
Aalborg University
http://www.aau.dk

Title:
Auralization platform for hearing aid development

Theme:
Sound technology for the hearing impaired

Project Period:
Spring 2019
MSc, 10th Semester

Project Group:
19gr1077

Participants:
Jorge Bravo Díaz

Supervisor:
Flemming Christensen

Number of Pages (including appendix): 101

Date of Completion:
5th June 2019

Abstract:

Hearing aid software testing is costly and time consuming since it requires hardware implementation. Conducting these tests in a virtual environment would make them more accesible and easier to implement. In addition, the placement of Behind The Ear Hearing Aid (BTE-HA) deprives the recorded signal from important spatial cues for sound localization. The main focus of this thesis is to check if it is possible to perform localization tests in virtual environments in comparison with a speaker-based experiment, and to evaluate the performance of different Head-Related Transfer Function (HRTF) sets: measured Blocked Ear Cannal (BEC), measured BTE-HA and processed BTE-HA with a Pinna Reconstruction Filter (PRF). The measured HRTF sets contain 616 positions in a sphere, and 10 subjects took part in the test. The test consisted of 4 stages with 3 rounds each in which sound played from 16 loudspeakers to be identified in different conditions. While the sets' performance is different, the overall results indicate that the virtual environment used is not suitable.

Preface

This report presents the work performed during my master's thesis in the period from February 15th to June 5th 2019. This part is dedicated to thank the people who has shown interest in the project and enriched the thesis with discussion of ideas, knowledge, and/or equipment.

I would like to thank Flemming Christensen for his advice and support during the development of the project. I would also like to thank my fellow student Christoph Kirsch for feedback and discussions throughout the entire project period, and Amalie Damgaard for her repeated participation as pilot subject in the initial trials of the measurement routine and microphone fitting. Claus Vestergaard Skipper is also thanked for being a huge help in setting up the HRTF measurement setup and general laboratory advice.

For citations, the report employs the Harvard method. If citations are not present by figures or tables, these have been made by the authors of the report. Units are indicated according to the SI standard.

Aalborg University, June 5, 2019

Jorge Bravo
<jbravo12@student.aau.dk>

Contents

Glossary

AAU Aalborg University. 12, 28, 36

BEC Blocked Ear Cannal. iii, 12, 17, 23, 26, 32, 33, 38, 39, 52, 53, 55, 57, 59, 63, 64, 65, 66, 69

BTE Behind The Ear. 12

BTE-HA Behind The Ear Hearing Aid. iii, 3, 11, 12, 23, 26, 27, 29, 31, 32, 33, 39, 47, 53, 55, 56, 58, 59, 63, 64, 66, 67, 69

dB Decibel. 9

ESS Exponential Sine Sweep. 30

HA Hearing Aid. 3, 12, 13, 26, 27, 28, 32, 47, 67, 69

HATS Head and Torso Simulator. 19, 27

HRIR Head-Related Impulse Response. 18, 19, 31, 32

HRTF Head-Related Transfer Function. iii, vi, 3, 12, 14, 15, 17, 18, 19, 20, 23, 24, 23, 26, 28, 30, 31, 32, 33, 35, 36, 38, 39, 52, 63, 65, 67, 69, 73, 82

IIR Infinite Impulse Response. 31

ILD Interaural Level Difference. 8, 9, 10, 11, 12

IPD Interaural Phase Difference. 10

IR Impulse Response. 23, 24, 25, 30, 31

ITD Interaural Time Difference. 8, 9, 10, 11

LTI Linear Time-Invariant. 17

PRF Pinna Reconstruction Filter. iii, 32, 33, 39, 53, 54, 55, 56, 58, 60, 63, 64, 66, 69

PRTF Pinna Reconstruction Transfer Function. vi, 33, 32, 33

RIC Receiver-in-Canal. 27, 67

SNR Signal to Noise Ratio. 40

SOFA Spatially Oriented Format for Acoustics. 15, 32, 43, 82

VR Virtual Reality. 68

Chapter 1

Introduction

The development of Hearing Aid (HA) algorithms and signal processing blocks for them is based on detailed knowledge of how hearing works and the hearing pathology to be approached, as well as the possibilities and limitations within the fields of electronics, acoustics, and transducer technologies, to name a few. As such, development of signal processing and algorithms is an expensive, time consuming process for HA manufacturers.

1.1 Problem statement

Testing new features for HA requires their implementation in existing hardware so that they can be tested in specialized laboratory conditions, for compliance with specific requirements. This however, will only tell developers if the algorithm is implemented correctly and not how it sounds or how it affects the listening experience of the users. For this purpose, listening tests can be performed, but they are costly and time consuming. Furthermore, the physical setup needed may be problematic or even impossible to reproduce with the resources at hand.

The motivation of the thesis is to investigate the feasability of performing these tests in a virtual environment with accessible tools. In addition, several studies have shown that the position of Behind The Ear Hearing Aid (BTE-HA) affects the way in which the sound is perceived, leading to localization errors ([Udesen et al., 2013], [Van den Bogaert et al., 2006]). The cause is the lack of spectral cues provided by the shape of the pinna ([Han, 1994], [Gardner and Gardner, 1973]). Therefore, this thesis will also evaluate the performance in terms of localization of different sets of Head-Related Transfer Function (HRTF)s in the selected virtual environment. These sets will represent natural hearing, hearing through the BTE-HA and a correction of the latter set to approach a more natural sound. In summary, the two main aspects that will be investigated during the thesis are:

- Are there differences in performance related to the HRTF set used in a virtual environment?

- Is it possible to conduct localization tests in a non-proprietary virtual environment?

Part I

Problem Analysis

Chapter 2

Spatial sound perception

Correct localization of sound sources, or identification of the direction from which the sound is coming from, is a crucial feature for both animals and humans. In the case of humans, it is not only important for survivability (e.g. environmental awareness), but also for social interactions (e.g. speaker identification). Throughout this chapter, the fundamentals of sound localisation, as well as how the use of hearing aids affects its performance, will be explained.

2.1 Sound source position

The position of a sound source is usually expressed as a point in a spherical coordinate system in which the listener's head is considered the centre of the sphere. The most common notation, and the one that will be used during the rest of the thesis, represents each point in the sphere in terms of:

- Azimuth (θ): represents the angle in the horizontal plane in degrees ($^\circ$). It ranges from 0° to 359° in counterclock-wise direction. Azimuth can also be referred to as a negative angle, meaning that the indicated angle has to be taken into account clock-wise ($\theta = -40^\circ = 320^\circ$).
- Elevation (ϕ): represents the angle in the vertical plane (with respect to the horizontal plane) in degrees ($^\circ$). Usually, elevation ranges from -90° to 90°, with the positive angles referring to the space above the ears and negative angles to the space below.

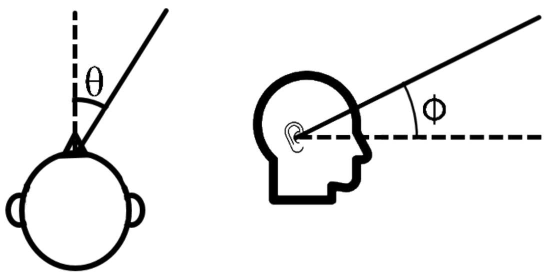


Figure 2.1: Azimuth (left) and elevation (right) representation.

A source located straight in front of the listener's eyes at the height of the ears will be considered the reference point in the spherical coordinates system and corresponds to $\theta = 0°$ and $\phi = 0°$.

2.2 Localization Cues

Sound localization refers to the ability of identifying the origin of a detected sound source in terms of direction and/or distance. Several cues are used by the auditory system in order to pinpoint the location of a sound, and they can be categorized in two groups, binaural and monaural cues. Sound localization can be achieved by the use of monaural cues, but the major contribution to localization is done by the stronger binaural cues.

2.2.1 Binaural Cues

We can categorize as binaural cues the set of cues obtained from both ears. When these cues are aquired through comparison between the signals received in both ears, they are called interaural cues. Due to the separation between the ears, the path followed by the sound to reach each one of them can be different, thus generating Interaural Time Difference (ITD) and Interaural Level Difference (ILD).

- ITD: refers to the difference in time for sound arrival in the ears. It is the result of a different path from the source to each ear and the time needed for the soundwaves to cover it.

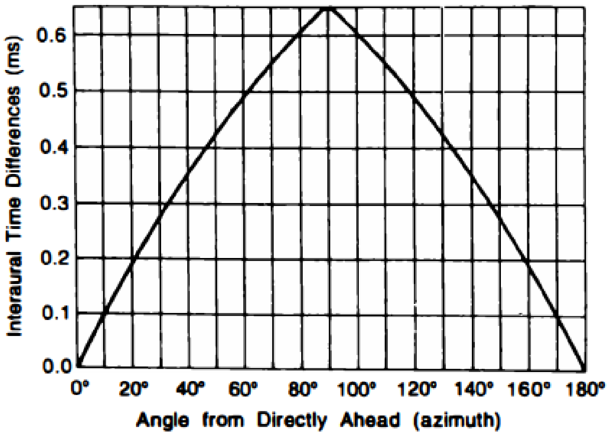


Figure 2.2: ITDs plotted as a function of θ as depicted in [Moore, 2012]. This plot is an approximation since in practice the ITD can vary slightly in frequency for a given θ depending on the method used to calculate it.

- ILD: refers to the difference in intensity between the sound arriving in each ear expressed in Decibel (dB). It is the result of distance attenuation as well as possible shadowing and diffraction generated by the head.

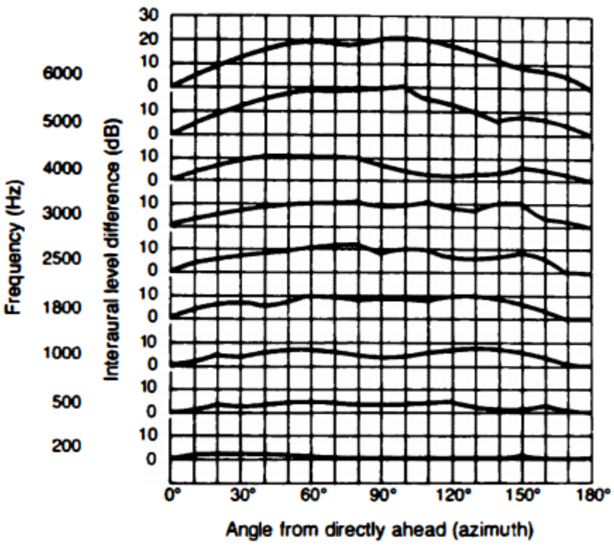


Figure 2.3: ILDs for sinusoidal stimuli as a function of θ, each curve representing a different frequency, as depicted in [Moore, 2012].

The following example(Figure 2.4) illustrates a situation in which the sound source is located in a position such that $\theta \approx -45°$ and $\phi = 0°$. This results in an ITD of approximately 0.38 ms with $t_l > t_r$ and, if the source is generating a 2.5 kHz sine, an ILD of approximately 10 dB with $L_r > L_l$.

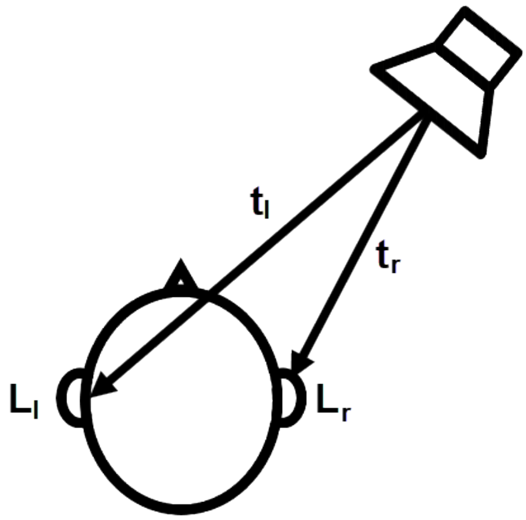


Figure 2.4: Top view of situation in which a sound source is located to the right of the listener. Times are referred as t_r and t_l and levels as L_r and L_l for the right and left ear respectively. The path to the left ear is represented as a direct path for simplicity, since in reality it would go around the head due to diffraction

Due to the physical nature of sound and its wave behaviour, the effectiveness of the binaural cues (ITD and ILD) varies depending on frequency. Low-frequency sounds have a wavelength bigger than the size of the head, and therefore the soundwave "encapsules" the head. This process is known as diffraction and it results in little to no "shadowing" by the head, as the soundwave bends around it. Furthermore, for low enough frequencies, the head could be considered as inexistent, since their wavelength is so big in comparison that it generates an almost undisturbed soundfield. On the other hand, high-frequency sounds have a wavelength smaller than the head, and therefore little diffraction occurs, resulting in more "shadowing". Taking this into account, and for sound sources far from the listener, ILDs can be considered negligible under 500 Hz but have a strong presence for high frequencies. This reflects in the fact that they can be around 20 dB over 5 kHz as shown in Figure 2.3. However, for sound sources located close to the listener considerable ILDs can happen even at low frequencies [Brungart and Rabinowitz, 1999].

Regarding ITDs, they can range from 0 µs for $\theta = 0°$ to 690 µs for $|\theta| = 90°$. For the case of pure tone sinusoidals, the ITD is equivalent to the Interaural Phase Difference (IPD)(phase difference between the two ears). For low frequency, IPDs provide effective information for sound location. However, at high frequency IPDs information can become ambiguous, therefore degrading the value of the cue. As an example, if we take a 4 kHz sinusoid, its period is 250 µs. If this sinusoid is presented from an angle resulting in an ITD of 500 µs ($\theta \approx 62°$), this would mean the ITD corresponds to two whole cycles of IPD, thus leading to ambiguities in the localization of the sound. These ambiguities start to happen when the period of the sound is double the maximum possible ITD. This means a period equal or greater than 1380 µs, which translates into ambiguities starting at 725 Hz. These ambiguities can be resolved by making head movements, but above 1.5 kHz phase differences become highly ambiguous [Moore, 2012]. Head movements can also help localize sources placed in the *cone of confusion*, which happens when several sources placed differently result in the same ITD, and generally improve the overall localization [Hirsh, 1971].

For frequencies above 1.5 kHz, the auditory system uses a method called *phase locking* in order to solve this ambiguities. This method relies on the auditory nerves and the fact that not all nerves send signals for every cycle of the stimulus. The nerves that are startled by the stimulus send signals with a period which is approximately a multiple of the stimulus period. This phenomenon has an upper limit around 5 kHz, but can occur weakly up until approximately 10 kHz [Moore, 2012].

2.2.2 Monaural Cues

We can categorize as monaural cues the set of cues obtained from the analysis of the signal arriving at one ear only, such as the effect of the pinnae shape, distance attenuation and echolocalization. The focus of this section will lay on the effects of the pinnae, since they are highly minimized when hearing through a Behind The Ear Hearing Aid (BTE-HA) due to its location. It is often suggested that the pinna provides information used in the judgement of vertical location ([Moore, 2012]), and the experiments conducted by [Batteau, 1967] showed a degradation in sound localization when the pinnae were absent both for the vertical and horizontal plane.

The experiments performed by [Gardner and Gardner, 1973], in which occlusion of different parts of the pinna with rubber molds was investigated, showed a decrease in localization with the largest effect occuring at the frequency bands with higher central frequencies (8 kHz and 10 kHz).

It is now generally accepted that the spectrum of incoming sounds is modified by the pinna in a way that is dependent of the incidence angle. Further investigation by [Han, 1994] studied how the pinna encodes directional information by occluding different parts of the pinnae of a KEMAR head and torso simulator. The experiments concluded that specific parts of the pinna contribute for an increased localization in different incidence directions (such as the concha for sound sources with $\phi > 0°$).

2.2.3 Summary on localization cues

As a brief summary of the content presented above, the following points can be synthesized [Xie, 2013]:

- The dominant localization cue below 1.5 kHz is the ITD.
- For frequencies above 1.5 kHz, both the ILD and ITD contribute to sound localization. With an increment in frequency, ILD gradually becomes more dominant, starting around 4 to 5 kHz.

- For frequencies above 5 to 6 kHz spectral cues play the most dominant role in sound localization (e.g. spectra cues from the pinna for front/back distinction).
- Dynamic cues introduced by head movements are helpful for vertical and front/back localization.

2.3 Spatial perception with Hearing Aids

While using BTE-HA, spatial perception is altered due to the lack of cues offered by the pinna since the Hearing Aid (HA) is located behind it and the microphones it contains are not affected by the aforementioned phenomena. Modern BTE-HA usually contain two microphones, which are used to perform beamforming in order to estimate the location of the sound source.

Experiments performed by [Van den Bogaert et al., 2006] evaluated localization performance in the frontal horizontal plane for normal hearing subjects and hearing-impaired subjects using commercial HA. The results of this study showed that the performance of hearing-impaired subjects was lower than the normal hearing ones. Furthermore, when testing hearing-impaired subjects without the use of HA, the localization performance was improved. This experiment also showed a negative impact on localization performance introduced by the use of adaptive directional noise reduction.

An analytic comparison of the relative gain difference between measurements taken at the Blocked Ear Cannal (BEC) and BTE-HA was performed by [Møller, 2018] based on a Head-Related Transfer Function (HRTF) dataset obtained by Flemming Christensen at Aalborg University (AAU) [1999, unpublished]. The results showed that for frequencies above 1 kHz the pinna has a stronger contribution, providing up to 30 dB gain relative to the Behind The Ear (BTE) counterpart, varying as a function of incidence direction. The work of [Udesen et al., 2013] showed that microphone placement BTE can lead to ILD changes up to 30 dB, which backs up the obtained results.

Chapter 3

Virtual environment

As stated before, one of the main objectives of this thesis is to validate the feasability of using virtual environments in order to perform analytical acoustics tests, mostly focused on HA algorithms or even testing of different HA models. The use of virtual environments for this purpose could offer an alternative approach for situations in which the conduction of the experiment in real life could become problematic (e.g. setups with an elevated number of speakers). It could also ease some other processes, such as bypassing having to load new HA signal processing modules in the hardware.

The choice of the appropriate virtual environment will be based on the fullfilling of several conditions:

- Free access.
- Non-proprierty coding language.
- Available 3D sound renderers.
- Possibility of including a visual virtual environment alongside the sonic one.

It has been decided that these conditions need to be fullfilled in order to keep the platform as accessible as possible for anyone interested in its use and/or functionality expansion. By doing this, it would be easier to share resources between interested parties and it would ease the further development of the platform.

3.1 Virtual Engine

When referring to the virtual engine through this thesis, it will mean referring to the framework in which the virtual environment is generated. Several engines are available for this purpose at the moment, and the main ones explored for this project are the following:

- **MATLAB**: numerical computing environment with its own proprietary programming language. Although primarily intended for numerical computing, it offers the possibility to generate virtual environments with its 3D World Editor. It allows for a deep level of customization in terms of audio processing, but as mentioned it has its own programming language and it requires a paid license.

- **TwoEars**: computational framework to generate virtual soundscapes. It includes a Binaural simulator and Auditory front-end in order to generate a high-level detailed auditory model. It is Matlab-based and would need to be integrated with the 3D World Editor.
- **Unreal Engine 4**: source-available game engine developed by Epic Games. Its code is written in C++ and it provides the advantages of easily integrating both the audio and graphics renderers, being able to produce a virtual environment without having to integrate separate dependencies. Used by a big number of professional game developers, its results are solid but with a steep learning curve.
- **Unity**: cross-platform game engine developed by Unity Technologies. Although not open-source, its source code is available under a reference-only license. Its code is written in C# and it provides the advantages mentioned in the previous point. It is the most spreaded game engine outside the professional world and provides with a great number of free Assets and plugins.
- **WebVR**: experimental JavaScript API providing support for virtual reality devices. It works in conjunction with other interfaces (such as WebGL) in order to generate web-based applications. It needs integration with more modules in order to produce a fully rendered visual and acoustic virtual environment.

After having investigated the available options, and taking into account the conditions mentioned previously, it was finally decided to use Unity as the virtual engine in which to develop and perform the tests. This decision was supported by the facts that Unity uses an open non-proprietry coding language(C#), it has a free access version, there are plenty of available high-level sound renderers and the learing curve is not as steep as in Unreal Engine 4. Furthermore, due to its extended use in the game development community, it counts with great community support and free-access resources.

3.2 Audio Spatializer

Once the virtual engine has been selected (Unity), it is necessary to find and test suitable audio spatializers that could be used within the virtual framework. The main conditions that need to be met with the audio renderer are a complete integration with Unity, the possibility of using custom HRTF, simulation of spatial audio cues and anatomic and environmental sound interaction. Another key point in the search of the most suitable spatializer is the availability of documentation and community support. After several plugins and libraries were tested, these were the most suitable options:

- **Resonance Audio**: multi-platform spatial audio SDK developed by Google with $C++$ source code available. It is a very powerful audio renderer with a realistic approach to sound interaction, backed by the use of HRTF in order to simulate sound interaction with the ear, and the inclusion of direct sound, early reflections and late reverb in order to simulate sound interaction with the environment. It is also possible to set up source and receiver directivity, as well as to include occlusion effects for objects along the soundpath. It is very well optimized in order to minimize the computational power needed and supports the use of custom HRTF by pre-processing them and generating the corresponding Ambisonics filters. However, as the software is right now, only one set of HRTF can be used at a time and if this set needs to be changed, the plugin needs to be recompiled and built again. It is possible to bypass the one set limitation by tweaking the settings and loading different sets as different ambisonics order filters, but when testing this the plugin was not updated correctly and it was not possible to load the custom HRTF. Furthermore, obtention of subject's HRTF and testing is performed in the same session and therefore the fact of having to recompile the plugin for every subject would have slowed the testing process. There is extensive and detailed documentation available and support both by the community and the original developers. This renderer was discarded due to the reasons mentioned, but it could be a very solid option for tests that do not need continuous changes in the HRTF set.

- **Steam Audio**: multi-platform spatial audio SDK developed by Valve. It has a full *C* API and a Unity plugin with source code in *C#* available. As in the case of Resonance Audio, Steam Audio renders a realistic soundscape by integrating environment and listener simulation with the possibility of using several custom HRTF stored in Spatially Oriented Format for Acoustics (SOFA) files and with physics-based sound propagation (e.g. occlusion and reflections). It offers low-latency HRTF-based binaural rendering for direct sound with the possibility of including Ambisonics recordings. The documentation available is not as extensive and in-depth as the case for Resonance Audio, but there is direct support from the developers as well as a solid community support.
- **SOFAlizer for Unity**: native spatializer plugin developed in $C/C++$ for research purposes [Jenny et al., 2018]. At the time of its development, no other spatializer allowed for dynamic HRTF use was available. It provides a realistic binaural rendering for direct sound (at the moment there is no information about including room acoustics), but lacks some functionalities when comparing it to Steam Audio or Resonance Audio. In addition, almost no documentation or community support are available and only *Windows* is supported.

Based on this analysis and several tests with each one of the renderers mentioned above, it is decided to use Steam Audio as the Audio Spatializer. This decision is supported because of its compliance with the requirements needed. Steam Audio is fully integrated in Unity, offers documentation and support, allows to use different sets of HRTF and change between them during runtime and adds the possibility of including room acoustics in case they are needed. In addition, it offers customizable settings such as the type of interpolation of HRTF to be used (bilinear or nearest), inclusion of air absorption in the acoustical simulation and source directivity adjustment.

Chapter 4

Head-Related Transfer Functions

As stated in section 3.2, one of the key factors when deciding for an audio renderer was the possibility of using custom HRTF. Throughout this chapter, it will be explained what HRTF are and how to obtain and process them.

The importance of being able to use custom HRTFs comes from the fact that they are needed in order to replicate in the virtual environment how a person would listen in reality.

4.1 Definition

As explained through chapter 2, sound generated from a point source interacts with anatomical structures before reaching the ears, generating different pressures containing different types of localization cues. This transmission process from the point source to a single ear can be categorized as a Linear Time-Invariant (LTI) process. HRTFs can be defined as the acoustic transfer functions of the LTI process, describing the filter effect applied to the sound on its way to the ears.

Generally, HRTFs vary as a function of frequency (f), source distance (r), source azimuth (θ) and source elevation (ϕ). However, for $r > 1\,\mathrm{m}$, HRTFs are close to independent of the source distance [Xie, 2013], and therefore can be expressed as follows [Blauert, 2005]:

$$HRTF(\theta, \phi, k) = \frac{P_{\mathrm{BEC}}(\theta, \phi, k)}{P_0(k)} \tag{4.1}$$

Where:

θ	source azimuth
ϕ	source elevation
k	frequency bin index
P_{BEC}	frequency domain complex-valued sound pressure at the BEC
P_0	frequency domain complex-valued sound pressure at the center of the head with the head absent

The reason for obtaining P_{BEC} instead of the pressure at other point (e.g. eardrum) is that by measuring in this way, ear canal resonances are eliminated while still maintaining all the spatial information [Møller, 1992]. If the reference measurement (P_0) is performed with the same microphone as P_{BEC}, both the microphone and speaker frequency response are removed from the HRTF and therefore this method will be used for the obtention of all subject's HRTFs.

4.2 Angular Resolution

The angular resolution of the HRTF set is not limited to the angles from which measurements have been taken. The HRTFs of positions not measured can be estimated by the use of interpolation. The huge advantage of interpolation is that the amount of measured HRTFs can be reduced . However, the higher the number of measured positions is, the more accurate the set of HRTFs will be and the more accurate the interpolated points can be.

There are several interpolation methods, such as time domain interpolation of the neighboring Head-Related Impulse Response (HRIR) horizontally or vertically (used by [Minnaar et al., 2005]), or the interpolation of the closest four available HRIR available (as used by [Sandvad, 1996]). As mentioned in section 3.2, Steam Audio allows the user to select between two interpolation methods [LAKULISH, 2017]:

- **Nearest**: uses the closest measurement position to the source's actual position. If the measurements are regularly spaced, this approach is fast. However, as the closest neighbour will change as the source position moves, it is highly possible that there will be audible artifacts when the change happens.
- **Bilinear**: calculates a weighted average HRTF from the ones available from the four measurement positions closest to the actual source position. This method is slower than Nearest interpolation and requires more computational power. However, this methods presents a smoother transition between positions than Nearest, with less audible changes between the interpolated HRTFs.

In general, noise-like sounds can benefit the most from Bilinear interpolation, as the audible sudden changes can be masked by the noise itself, while speech and music tend work good enough with Nearest interpolation. However, and as computational power is not a problem since the test is not computationally complex, Bilinear interpolation will be used for this thesis. As described in [Sandvad, 1996], the obtention of the estimate HRIR based on the closest four points can be expressed as:

$$\widehat{h} = \sum_{\mathrm{n}=1}^{4} w(n) \cdot h_n \tag{4.2}$$

Where:

$\widehat{h}$	interpolated HRIR
$w(n)$	weighting coefficient of the n^{th} closest HRIR
h_n	n^{th} closest measured HRIR

The weighting coefficients ($w(n)$) are obtained in the following way:

$$w(n) = \frac{V(n)}{\sum_{\mathrm{m}=1}^{4} V(m)} \tag{4.3}$$

where

$$V(n) = \frac{\left|\theta(4) - \theta(1)\right| \cdot \left|\phi(2) - \phi(1)\right|}{\left|\theta(x) - \theta(n)\right| \cdot \left|\phi(x) - \phi(n)\right|} \tag{4.4}$$

Where:

$\theta(x)$	target azimuth
$\phi(x)$	target elevation

By combining this method with regularly spaced position measurements, the transition through non-measured HRTFs can be performed in a smooth way with a low chance of audible artifacts being generated when changing positions.

One of the main limitations in HRTFs measurement is the time constraint, since they need to be measured on individual subjects as for the purpose of this thesis they will not be measured using a Head and Torso Simulator (HATS), and this measurement time needs to be minimized (the more positions measured, the more time needed) in order to reduce the test sessions duration. The works of [Minnaar et al., 2005] showed that the number of measurement points could be reduced from 11975 points in a sphere around the head to 1130 without introducing errors by using linear interpolation. They also showed that a resolution of at least 16° is needed for $|\theta| < 45°$ and $135° < |\theta|$ and 4° to 8° for the rest of the horizontal plane.

Taking into account the results of [Minnaar et al., 2005] experiments, the number of positions measured for the obtention of the HRTFs can be reduced to 616 by using the following horizontal angle resolutions:

- 5° for $45° < |\theta| < 135°$

- 9° for $|\theta| < 45°$
- 9° for $135° < |\theta| < 180°$

In terms of the vertical plane, a resolution of 8° is used for $|\phi| < 40°$. No positions will be measured for $|\phi| > 40°$. With these guidelines, the measurement positions set is defined as represented in Figure 4.1 and Figure 4.2.

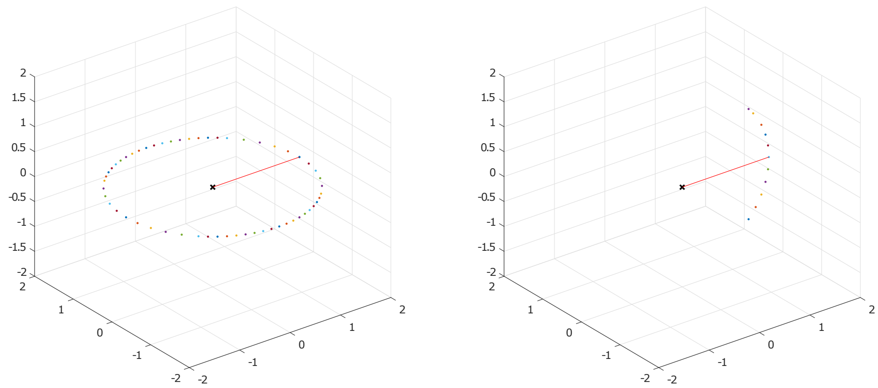


Figure 4.1: Azimuth (left) and elevation (right) positions for the measurements. The black x represents the center of the head, and the red line links it with the position right in front of the eyes of the listener ($\theta = \phi = 0°$).

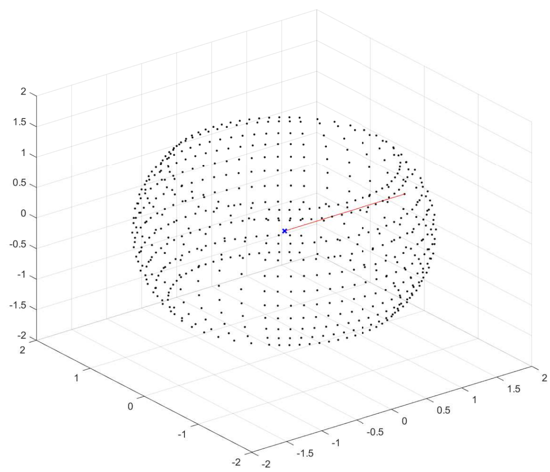


Figure 4.2: All measurement positions for the HRTF set. The blue x represents the center of the head, and the red line links it with the position right in front of the eyes of the listener ($\theta = \phi = 0°$).

Part II

Test Design

Chapter 5

HRTF Measurement Setup

The test routine developed for this thesis will be separated into two main parts:

1. Measurement of personalized HRTFs
2. Localization tests

This chapter will focus on the first part and describe the measurement of the HRTFs, including setup, methods and materials used. During the measurements part, the HRTFs for both BEC and BTE-HA will be simultaneously obtained for both ears.

5.1 Physical setup

Impulse Response (IR) measurements for the acquisition of HRTFs are to be performed inside a large anechoic room [AAU, 2019a] so that only contributions from the direct sound are taken into account. As stated in section 4.2, the number of measurement positions has been reduced from 1130 to 616. Nonetheless, a physical setup with 616 speakers is highly impractical and, in most cases, unlikely to be reproduced.

It is possible however to use a fixed number of speakers and rotate the subject along the longitudinal axis in order to obtain all needed positions. By using this method, the number of needed speakers can be reduced to that of the number of elevations needed. As per the aforementioned requisites, for this thesis $\phi \in [-40..40]$ in 8° increments, which generates 11 possible elevation positions, and thus 11 speakers needed. These speakers need to be positioned within an arc with an arbitrary radius, being the center of the subject's head the center of the arc too. An overview of the measurement setup is depicted in Figure 5.1.

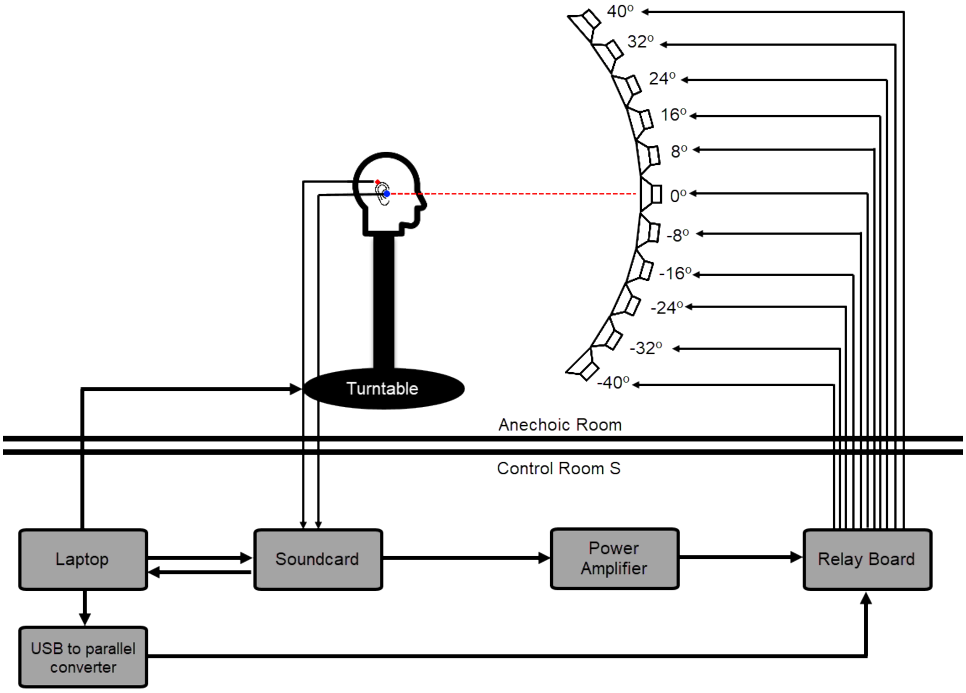


Figure 5.1: HRTF measurement setup diagram. The red dot denotes where the microphone of the BTE-HA is located, while the blue dot denotes the miniature microphone. Connection wires from the microphones located in the contralateral side of the head are not represented in the Anechoic Room section for simplicity.

5.2 Equipment

In this section, some remarks on specific pieces of the equipment will be given. For the complete list of equipment used please refer to Appendix A

Turntable

As explained above, a turntable is necessary in order to rotate the subject. For this purpose it was decided to use an Outline ET250-3D. This decision was supported by the fact that it can withstand and handle the weight of an adult, and it has an angular resolution of 0.5° which will ensure that all of the defined azimuth positions (section 4.2) will be available. Furthermore, this turntable is controllable via UDP protocol, making an integration with MATLAB® easy and allowing full automation of the whole measurement routine.

With this configuration, the subject needs to stand up and still for around 30 minutes. This has proved to be difficult as the subjects get tired as the measurements go on. In order to ease this task, some kind of support is needed. It is not possible to implement support up to the head without the risk of introducing unwanted reflections in the measured IR, and therefore, a support system for the lower back was added to the turntable.

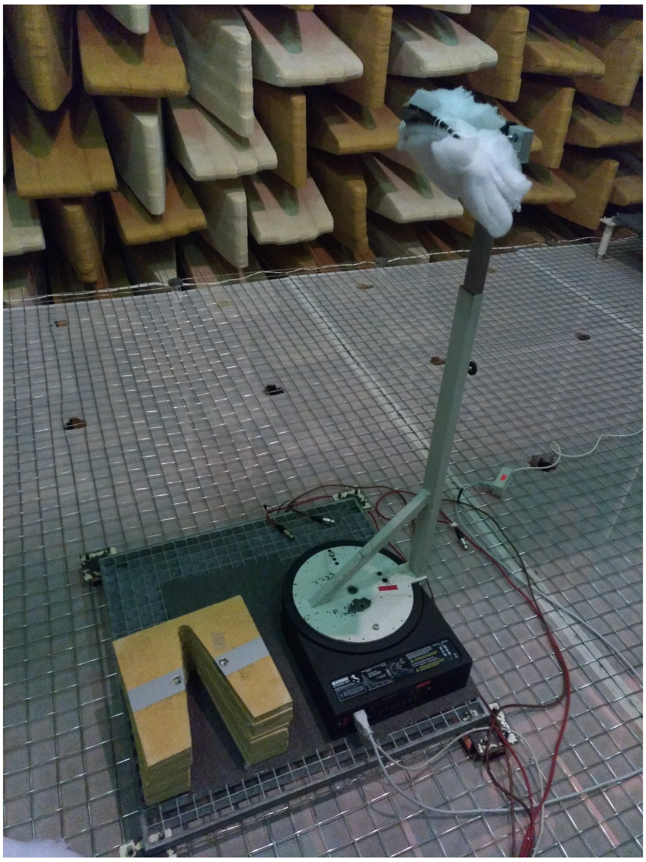


Figure 5.2: Turntable with lower back support and spare wooden boards for height adjustment.

Speaker arc

Since it was decided that the subject would rotate instead of the speakers, it is needed that the speakers are fixed in place in a suitable configuration for the required measurements. In order to do so, 11 custom made ball speakers [AAU, 2019c] (one for each of the defined elevations) were mounted on a 1.72 m radius metal arc. In order to avoid reflections in the higher frequencies, the whole arc has been wrapped in absorbent material (cotton wool). Reflections in the low frequency range will be removed digitally by windowing the measured IRs.

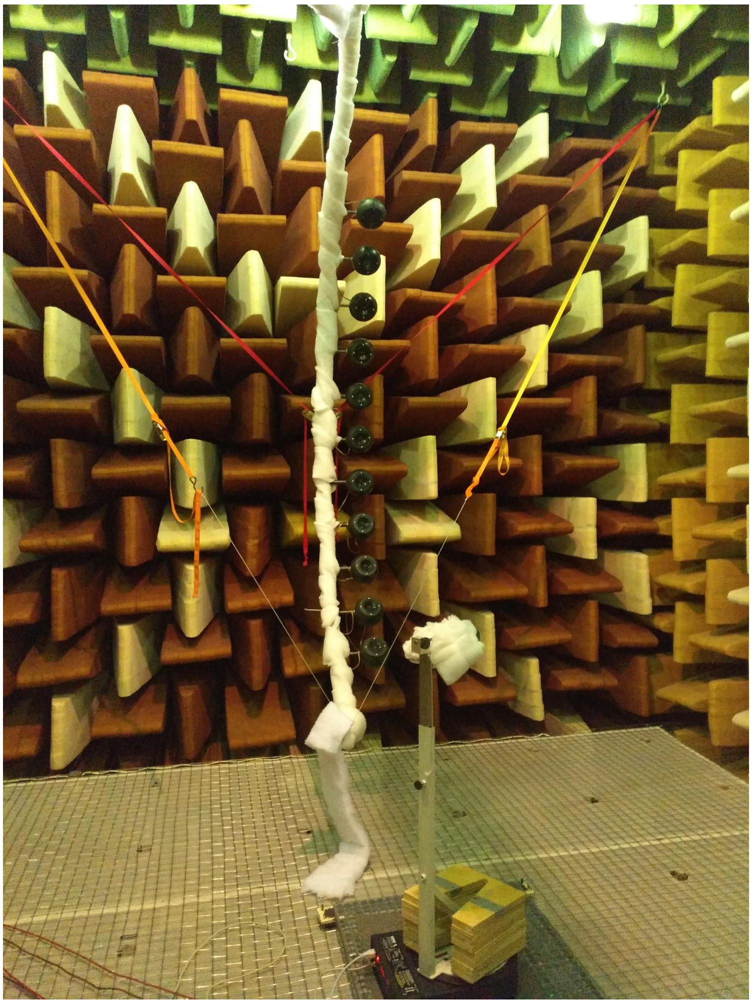


Figure 5.3: Speaker arc configuration as installed in the anechoic room.

BTE-HA

For the purpose of the thesis, two BTE-HA are needed in order to compute the HRTFs to its microphones and later be able to evaluate its localization performance as compared to the HRTFs measured in the BEC in a subject-based listening test. Two *Widex Fusion* HA are used for this purpose. These HA provided by Widex can be considered inert, since no signal processing is implemented in them. However, they provide access to the microphones and the speaker, or receiver.

Figure 5.4: Widex Fusion hanger with two microphones and an in-ear receiver.

As depicted in Figure 5.5, the Widex Fusion model is a Receiver-in-Canal (RIC) BTE-HA. This means that the receivers (or speakers) are placed inside the ear canal, as if they were in-ear earphones, instead of in the HA housing. The receiver is not present during the measurements, as it can be detached and its presence complicated the equipment and fixing of the HA on the subject. This model of BTE-HA contains two Sonion Type 8002 microphones ([Sonion, 2015]) accessible through break-out wires. As the units used are inert, external circuits are needed for microphone biasing using 1.5 V batteries (AAA).

Figure 5.5: Widex Fusion hanger on a HATS.

The aforementioned microphone biasing circuits need to be close to the HA, and therefore placed on the subject during the measurements. A custom harness made for that purpose, modifying a GoPro chest mount harness and adding a plastic box containing the biasing circuitry for the HA microphones (Figure 5.6), was available through AAU and is to be used for this thesis.

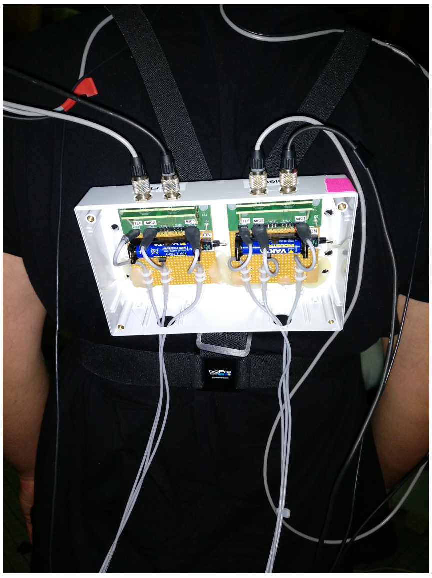


Figure 5.6: Custom-made harness with biasing circuitry and microphone connections.

5.3 Method overview

As stated in section 5.1, in order to minimize the number of speakers used it is necessary to either rotate the subject or the speakers, and in this case rotating the subject is simpler. This can be achieved by having the subject stand on a turntable and rotating it with the desired azimuth angle resolution. In order to align the subject ears and the speaker corresponding to $\phi = 0°$, wooden boards can be stacked on top of the turntable to adjust height. In addition, two laser leveling tools are used to ensure correct placement of the head.

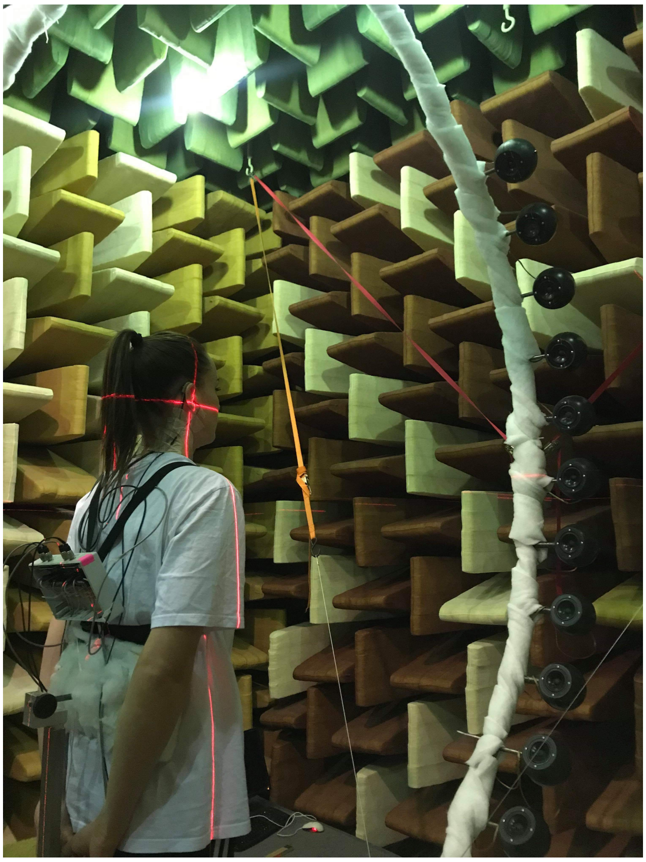


Figure 5.7: Subject in position. The red cross generated by the laser leveller indicates the position in which the ear needs to be.

After the correct height position has been adjusted, the custom harness is mounted, and two miniature microphones and two BTE-HA are fitted (one per ear) and secured with medical tape so they don't move between measurements.

Figure 5.8: Fitting of in-ear miniature microphone and hearing aid. Medical tape is used in order to secure the cables and avoid unwanted microphone movement. The in-ear microphone is fitted into a trimmed E.A.R. Classic earplug with a small indent

Once the harness and microphones are in position, the subject stands on top of the rotating platform, without moving or generating noise during the measurements. The IRs for the four microphones are obtained simultaneously, by playing a 1 s Exponential Sine Sweep (ESS) from 20 Hz to 22 kHz with a 0.2 s silence at the end in order not to lose the tail of the IR. The excitation signal is presented at an approximate level of 70dB in order not to damage the speakers. This signal is played sequentially through the 11 speakers corresponding to the elevation positions (separated by 8°), thus aquiring the IRs of all ϕ for a given θ. This process is then repeated in the 56 selected azimuth positions, generating the needed 616 measurement points. The whole routine is automated through a MATLAB® program that controls the speakers with the use of an Arduino Uno, processes the recorded signals from the four microphones, and controls the platform rotation via UDP.

Without taking into account the preparation and fitting of the microphones, the measurement routine takes between 21 and 27 minutes, making it difficult for the subject to stay completely still for the whole duration. Therefore, a little 1 min break is given every 90° (θ) in which the subject can move while not getting off the platform. These little breaks proved to ease the task for the subjects, while only increasing the duration of the measurement session by 3 minutes.

5.4 HRTF Processing

After the IRs have been measured from all the defined positions, it is necessary to obtain the HRIR as shown in Equation 4.1. However, a small modification will be introduced, since the result will be filtered with 6^{th} order Butterworth bandpass filter, implemented as an Infinite Impulse Response (IIR), with cutoff frequencies 20 Hz and 10.8 kHz. The low-pass filter component is introduced since the microphones used by the BTE-HA hangers used (Sonion Type 8002) have a roll-off at 10 kHz in order to avoid any artifacts or noise that may appear in this frequency area.

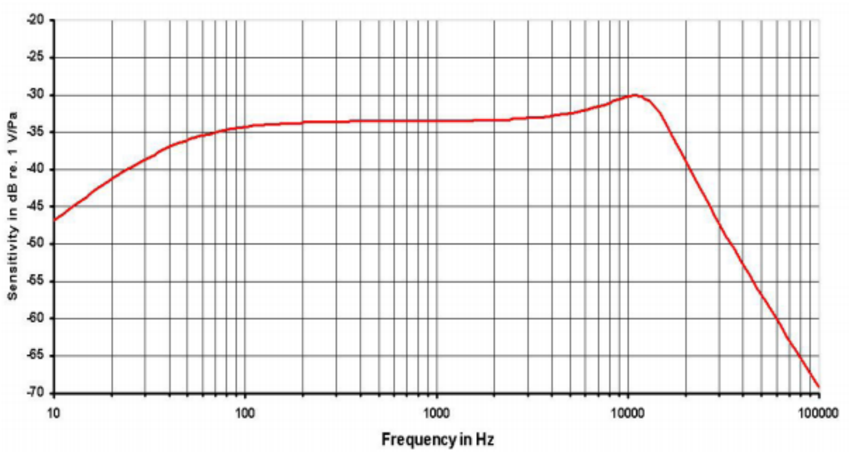


Figure 5.9: Typical response curve of Sonion Type 8002 microphones, as per the datasheet [Sonion, 2015]

Furthermore, as the signal is sampled with the use of a soundcard, nearly half the frequency range will be attenuated due to sharp anti-aliasing filters, which can generate instabilities in the obtained HRIRs. The high-pass filter component is included in order to avoid instabilities in the low-frequency domain of the HRIRs caused by noise. Therefore, the formula used for the acquisition of HRTFs is the one depicted in Equation 5.1. After computing the HRTF, it was needed to bring it back to time domain as HRIR and trim its length to 256 samples in order to ensure that the impulse response had decayed within the included timeframe, meaning that only the impulse response and no reflections are taken into account.

$$HRTF(\theta, \phi, k) = \frac{P_{\mathrm{BEC}}(\theta, \phi, k)}{P_0(k)} \cdot H_{\mathrm{BP}} \tag{5.1}$$

Where:

H_{BP} frequency response of bandpass filter

As stated in section 3.2, SteamAudio allows using custom sets of HRTFs stored in SOFA format, which allows for very flexible ways of defining HRTFs [Institute]. However, it is needed to fullfill several restrictions since SteamAudio only supports a restricted subset [Valve, 2017](directly taken from the source):

- SOFA files must use the *SimpleFreeFieldHRIR* convention.
- The *Data.SamplingRate* variable may be specified only once, and may contain only a single value. Steam Audio will automatically resample the HRTF data to the user's output sampling rate at run-time.
- The *SourcePosition* variable must be specified once for each measurement.
- Each source must have a single emitter, with *EmitterPosition* set to [0 0 0].
- The *ListenerPosition* variable may be specified only once (and not once per measurement). Its value must be [0 0 0].
- The *ListenerView* variable is optional. If specified, its value must be [1 0 0] (in Cartesian coordinates) or [0 0 1] (in spherical coordinates).
- The *ListenerUp* variable is optional. If specified, its value must be [0 0 1] (in Cartesian coordinates) or [0 90 1] (in spherical coordinates).
- The listener must have two receivers. The receiver positions are ignored.
- The *Data.Delay* variable may be specified only once. Its value must be 0.

The fullfillment of this requirements, as well as the creation and storage of the SOFA files is performed with a MATLAB® program using the SOFA API for Matlab and Octave [Majdak and Noisternig, 2016]. Three sets of HRIRs are obtained and stored for each subject:

1. ***subject*_bec**: this set corresponds to the measurements taken at the BEC, representing the usual hearing position for the subject.
2. ***subject*_bteha**: this set corresponds to the measurements taken with one of the microphones of the HA, which represents how the sound would reach the subject in the events of using a BTE-HA.
3. ***subject*_prf**: this set does not correspond to direct measurements, but to the digital processing of the measurements obtained at the BTE-HA and filtered such as it mimicks the behaviour expected in the BEC. More information about the calculation of this set can be found in section 5.5.

5.5 Pinna Reconstruction Transfer Function (PRTF)

As explained in section 2.3, sound perception through BTE-HA varies with respect to the natural perception through the ear. This has become a point of focus for HA manufacturers, since the ultimate goal is to mimick a natural sound perception through the BTE-HA. In order to do so, the sound recorded with the BTE-HA can be filtered through a Pinna Reconstruction Filter (PRF). However, in order to do so, it is necessary to know the direction of the impinging sound. For the purpose of this project, a pre-processing approach has been chosen instead of a real-time live solution. As the HRTFs for both BEC and BTE-HA is obtained for each subject, it is possible to create a new set of *pinna-corrected* HRTFs which are already filtered so that no further live processing is needed.

By doing this, the spectral cues that appear when the pinna is present (BEC) will be recovered for a position in which they were originally not found (BTE-HA). The filter used to modify the measured BTE-HA HRTF set characterizes the relative gain difference between the BTE-HA and the BEC (in frequency domain), which is given by:

$$H_{\text{PRF}}(\theta, \phi, k) = \frac{HRTF_{\text{BEC}}(\theta, \phi, k)}{HRTF_{\text{BTEHA}}(\theta, \phi, k)} \tag{5.2}$$

The response of the PRF, or PRTF, describes the changes in magnitude, and can therefore be expressed as follows for the case of correcting the signal recorded through the BTE-HA:

$$PRTF(\theta, \phi, k) = |H_{\text{PRF}}(\theta, \phi, k)| \tag{5.3}$$

Taking this into account, it can be inferred that it is possible to generate a set of HRTFs corresponding to this correction by filtering the HRTF set obtained from the BTE-HA and filtering it with the corresponding PRF, and this is the procedure used for the processing of this thesis' datasets:

$$HRTF_{\text{PRF}}(\theta, \phi, k) = HRTF_{\text{BTEHA}}(\theta, \phi, k) \cdot PRTF(\theta, \phi, k) \tag{5.4}$$

Chapter 6

Localization test

In order to evaluate the performance of the different sets of HRTFs obtained, a localization test based on [Christensen et al., 1997] has been designed, the main difference being that the current test is to be conducted in a virtual environment instead of using a physical setup. By doing this, the test will provide with information about the performance of both the HRTFs and the virtual platform.

The original test was conducted inside an anechoic room, placing the subjects in the center of a sphere composed by 17 loudspeakers placed in different directions with a distance of 1 m from the center of the head. A 3 s pink noise signal with a 150 ms fade-in and fade-out was used as stimulus, as well as some low/high-pass filtered versions. Each speaker played the stimulus 4 times, using 9 stimulus types. After the sound had finished playing, the subjects had to choose the loudspeaker from which they perceived the sound in a forced choice test with no feedback. The loudspeakers were arranged as follows: In the median plane there were 7 loudspeakers with a 45° angle in between placed in the front and back hemispheres at elevations -45°, 0° and 45° and one placed directly above the subject at elevation 90°. The loudspeaker position right below the subject was omitted. In both the left and right hemisphere there were 3 loudspeakers placed in 45°,90° and 135° azimuth 0° elevation, and 2 in 90° azimuth ±45° elevation. [Christensen et al., 1997]

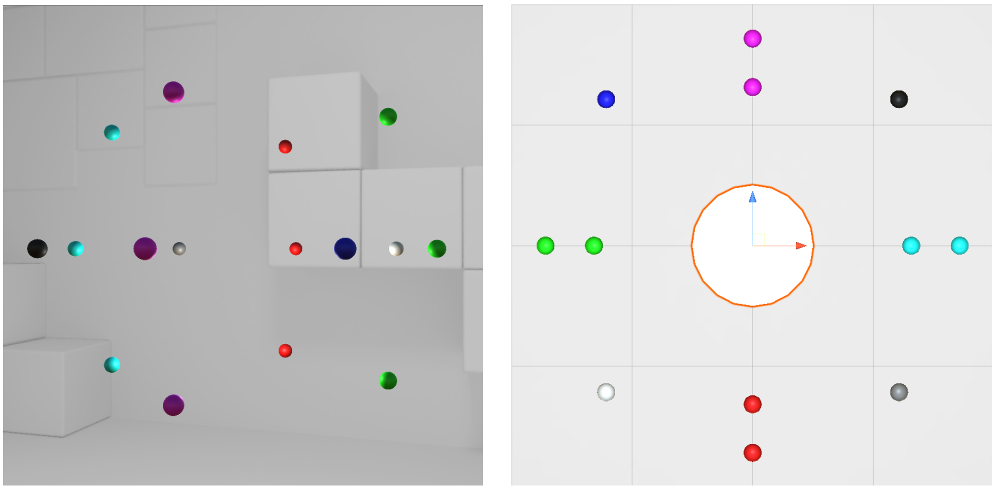
Figure 6.1: Original test setup. Image source [Christensen et al., 1997], edited to maintain subject's anonymity.

However, for the current test, both the procedure and the physical arrangement have been object to modifications. As stated before, the test conducted during this thesis has been designed and implemented in a virtual environment. Taking into account all the considerations evaluated in chapter 3, Unity is set as the virtual environment to use, with the addition of Steam Audio as the audio spatializer. The test is conducted trying to replicate anechoic conditions (no reflections or reverb included) and is conducted in one of AAU's audiometry cabins [AAU, 2019b] using a Lenovo ThinkPad laptop, RME Fireface 802 soundcard and Beyerdynamic DT 990 headphones. The fact that a virtual environment is used allows to bypass most of the limitations that would be present in a physical setup. However, limitations introduced by the measurement routine had to be taken into account. As explained through chapter 5, the maximum and minimum elevation positions measured correspond to $\phi = 40°$ and $\phi = -40°$ respectively. This constraint forced the modification of the maximum/minimum elevations allowed in the setup, since the whole point is to evaluate the performance of the measured HRTFs, and therefore it was decided to limit the elevations in the same way. Even though non-measured positions are interpolated within the tool, it was decided to maintain this constraint in order to use the measured data and not inferred one. Therefore, one of the loudspeaker positions was completely removed from the setup ($\phi = 90°$), while 8 positions were modified in order to comply with the elevation constraint. The new loudspeaker positioning can be found on Table 6.1. It was decided to colour-code the speakers in order to ease their identification.

Table 6.1: Loudspeakers positions for the test

Loudspeaker	Azimuth (θ)	Elevation (ϕ)	Colour
1	0°	-40°	Magenta
2	0°	0°	Magenta
3	0°	40°	Magenta
4	-45°	0°	Black
5	-90°	-40°	Cyan
6	-90°	0°	Cyan
7	-90°	40°	Cyan
8	-135°	0°	Gray
9	180°	-40°	Red
10	180°	0°	Red
11	180°	40°	Red
12	135°	0°	White
13	90°	-40°	Green
14	90°	0°	Green
15	90°	40°	Green
16	45°	0°	Blue

Another physical element changed was the distance from the listener to the speakers. In the original experiment, the speakers were located 1 m from the center of the listener's head. Within the virtual world, some spatializers , such as Resonance Audio, provide with *Near Field Effects* that, according to the documentation, kick off for sources at 1 unit (corresponding to 1 m) distance or below. In the Steam Audio documentation it is not specified if there is any extra processing applied to sounds coming from 1 unit of distance or lower, but just to be completely sure and not risk introducing artifacts due to too close distances, it was decided to use 1.72 units of separation instead. This distance was chosen since it matches the measurements distance from the center of the head to the speakers, but any value greater than 1 unit could have been used.



(a) In-test view of the loudspeaker sphere. Each colored sphere represents one loudspeaker.

(b) Plan of the loudspeakers sphere. The position of the head is located where the blue and red arrow meet.

Figure 6.2: In-test loudspeakers sphere.

The localization test is divided into four stages, each one with three test rounds, one for each HRTF set, and a familiarization round, always rendered using the BEC set for the subject. The familiarization rounds consist of 5 iterations, while the test rounds consist of 32 iterations. Each iteration means a change in the source playing the stimulus, meaning that for the test runs, each speaker plays twice. The order of the sources to be played is randomized for each round within each stage, ensuring that the same source will not play twice in a row, and that every source is not played more than twice. As a first approach, the source order for each round was randomized for every different subject, but this method proved to be prone to require more time than the limit allowed by the frame update rate, crashing the application. This happened due to the fact that for the last position in the order list, there is only a 1/16 probability of randomly drawing the correct source, and in some cases this forced the method to repeat for too long. In order to avoid this uncontrollable crashes, it was decided to pre-define a set of randomized lists (each list containing a random source order). This means that although every subject receives the stimuli from the speakers in the same order, this order was randomly generated, maintaining the intra-subject randomness. Furthermore, as each round is rendered using a different set of HRTFs, and this assignation is done randomly for each subject, not all subjects will coincide in set used and order of sources. As an example, let's imagine that the test consisted on only 3 rounds with 2 iterations each. If our initial randomly generated lists are $R1[16,2]$, $R2[6,1]$ and $R3[0,12]$, this means that every subject would have the stimuli for each round presented as defined by the lists:

Table 6.2: Source order example

Subject	Round	Source order
1	1	[16, 2]
1	2	[6, 1]
1	3	[0, 12]
2	1	[16, 2]
2	2	[6, 1]
2	3	[0, 12]
3	1	[16, 2]
3	2	[6, 1]
3	3	[0, 12]

However, as the HRTF set used during each round is assigned randomly, this would mean that the set to evaluate with this order will potentially be different between subjects (repetitions will occur for sure for more than 6 subjects):

Table 6.3: Source order example with HRTF set

Subject	Round	Source order	HRTF set
1	1	[16, 2]	BEC
1	2	[6, 1]	BTE-HA
1	3	[0, 12]	PRF
2	1	[16, 2]	PRF
2	2	[6, 1]	BEC
2	3	[0, 12]	BTE-HA
3	1	[16, 2]	BTE-HA
3	2	[6, 1]	PRF
3	3	[0, 12]	BEC

During the first and second stages, the subject is in a static mode, with a virtual fixed head position in order to recreate the conditions of Christensen's test as much as possible, as a way to maintain an anchor to an experiment performed outside a virtual environment. For the first stage, a full-band 3 s pink noise signal with a 150 ms fade-in and fade-out is used as stimulus, presented at $\approx 75dB$. Every audio signal used during the experiment was previously filtered with an equalization filter designed specifically for the Beyerdynamic DT990 headphones and provided by Flemming Christensen (Figure 6.3).

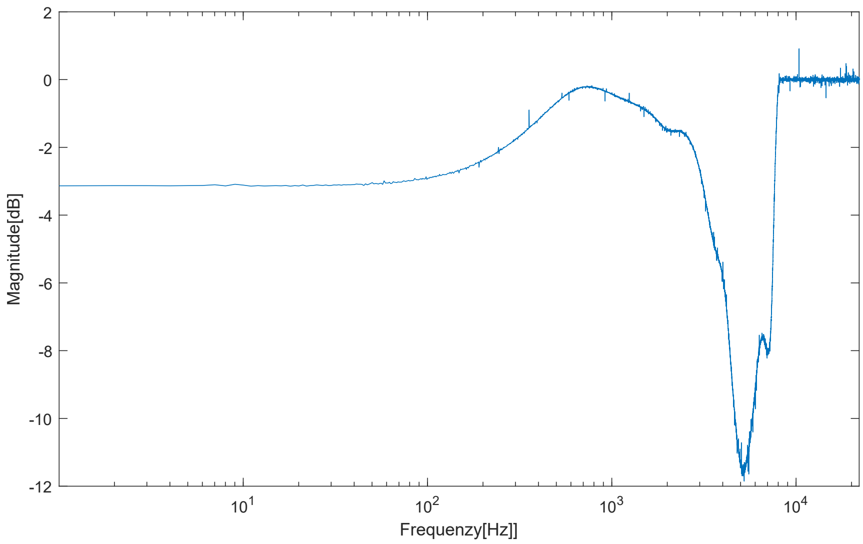


Figure 6.3: Filter response of the headphones equalizer when comparing white noise with its equalized version.

The stimulus level prevails during the whole test, independently of the stage or type of signal played. The subject's answer is not registered until the stimulus has stopped playing. This is done in order to replicate Christensen's test and have a linking point for reference. For the second stage, an excerpt from an audiobook is used as stimulus, while babble noise is played as background noise from four extra speakers located at $\theta = \pm 45°, \pm 135°$, $\phi = 0°$ and at a distance of 2 units from the listener position. The distance for the background noise sources was selected so that they would not block the in-game raycasts from the player to the stimulus sources. The Signal to Noise Ratio (SNR) between the stimulus and the noise is $\approx 6dB$. This SNR value is maintained in all stages that use background noise. From this stage onwards, it is not necessary for the stimulus to stop playing in order to register the subject's answer.

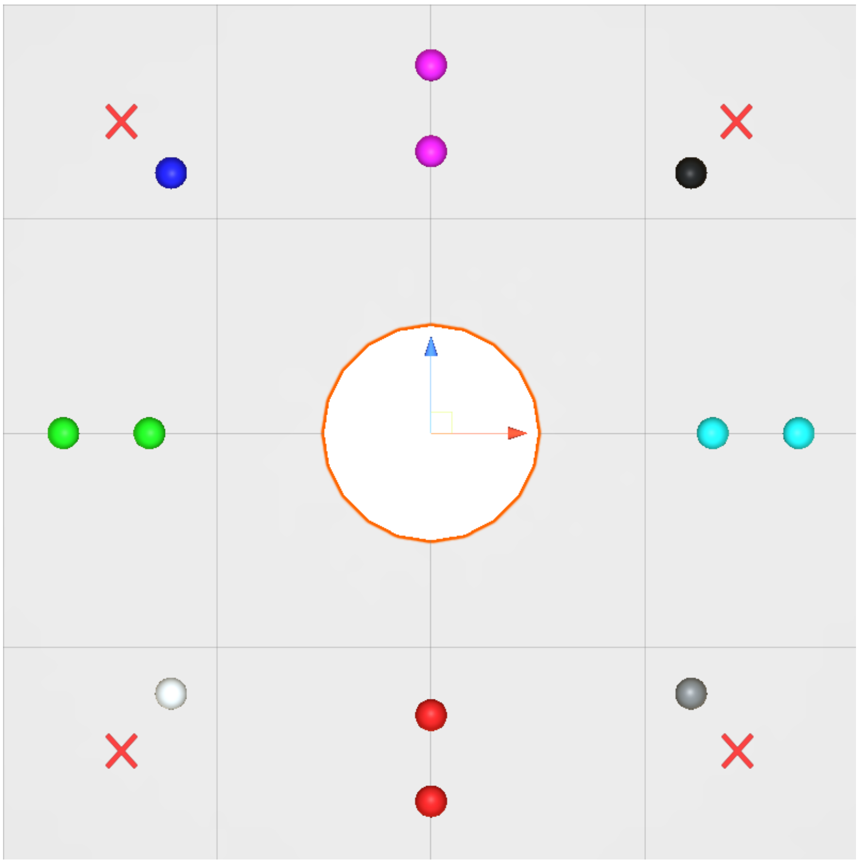


Figure 6.4: Plan view of the test environment with position of the noise sources marked with red X.

Since stages 1 and 2 are bound by a virtual fixed head position, the way in which the subjects input their answers is by the use of a virtual controller displayed during the test. This controller (Figure 6.5) represents the environment in which the subject is located, depicting all loudspeakers positions (excluding the noise loudspeakers) as view from above, and keeping the colour coding used to generate the virtual spheres. For those positions in which several speakers share the same azimuth ($\theta = 0°, \pm 90°, 180°$), three buttons are added, corresponding to:

- **Up**: corresponds to the speaker at $\phi = 40°$.
- **Mid**: corresponds to the speaker at $\phi = 0°$.
- **Down**: corresponds to the speaker at $\phi = -40°$.

For these cases it is necessary to click on the textbox in order to input an answer, while for the rest of the cases the subject has to click on the speaker icon. The controller also displays the Round and Iteration number so that it's easier to follow the development of the test.

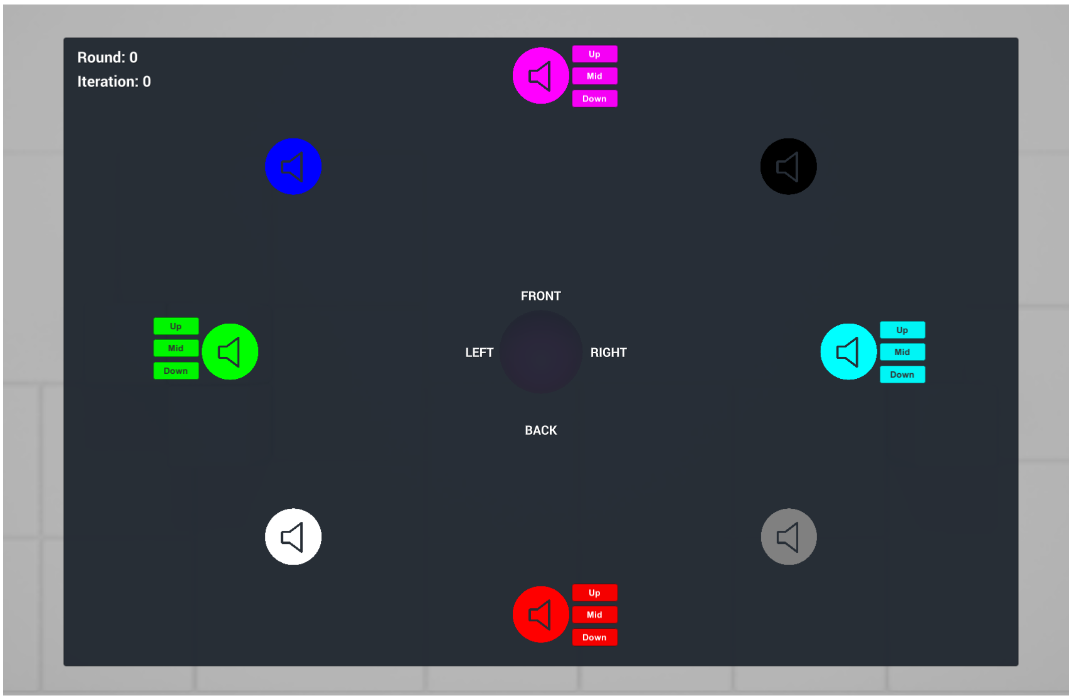
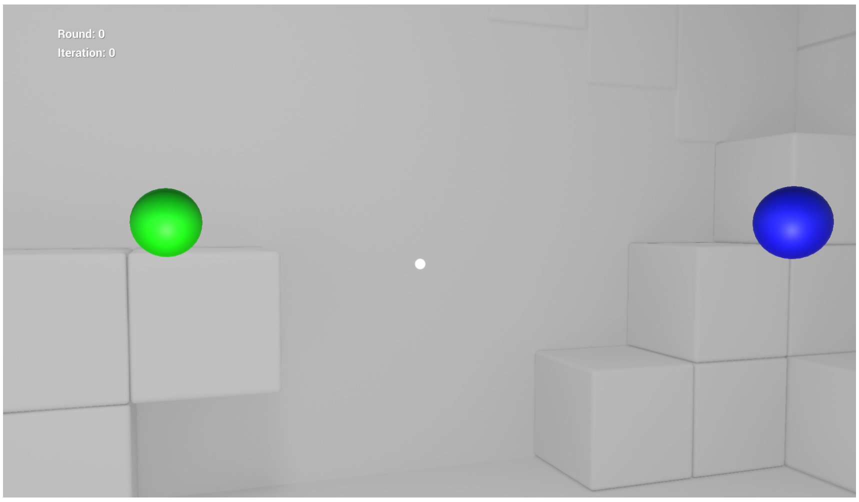
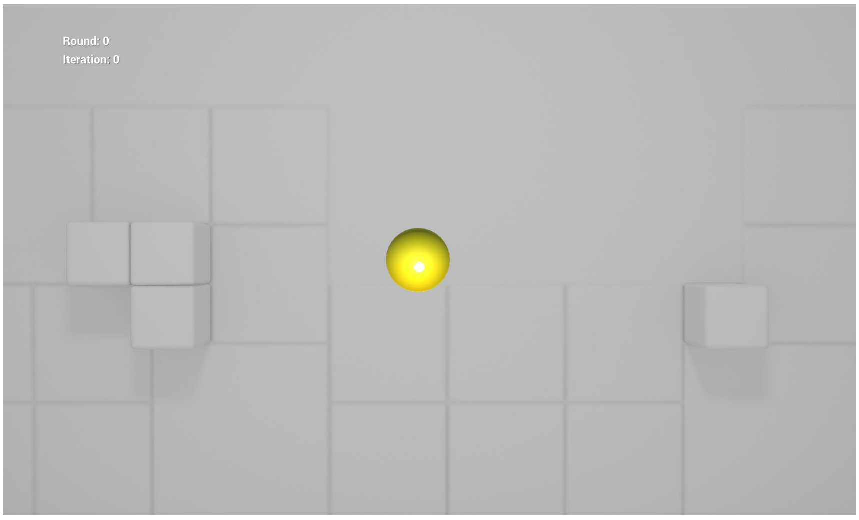


Figure 6.5: Virtual controller used for answer input during stages 1 and 2.

During the third and fourth stages, virtual head movements using the mouse in a first-person perspective are allowed in order to provide further information on the performance in a dynamic situation, that would most likely occur in real life. In terms of stimuli and background noise, stage 3 is similar to stage 1, being the main difference that the stimulus in stage 1 only played once whereas in stage 3 it is looped, and stage 4 is equal to stage 2. For these rounds, the virtual controller is no longer used, and the subjects need to click on the selected loudspeaker after aiming at it. For this purpose, a pointer or crosshead is included indicating the look direction, and the spheres representing the loudspeakers are highlighted in yellow when being looked at. The counters for round and iteration are still displayed on the top left corner of the screen.



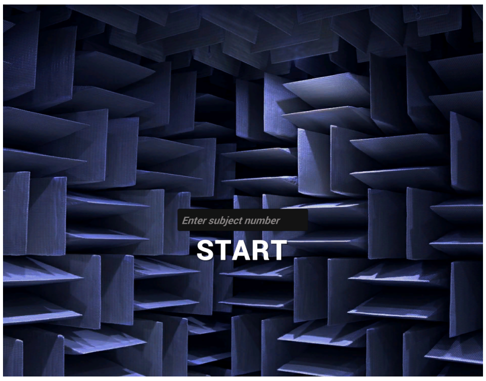
(a) In-test first-person view for stages 3 and 4. Notice the pointer in the middle of the screen indicating look direction.



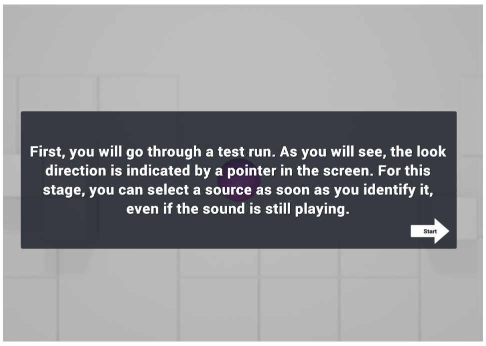
(b) Highlighted sphere when looked over.

Figure 6.6: In-test subject view for stages 3 and 4.

As extra details, the test routine software includes a main menu in which the subject ID needs to be introduced when starting up the program (Figure 6.7a). It also includes in-test instructions describing the mechanics and overall scene before each stage and messages/indicators at the end of the rounds and stages for the subject to be able to take a break (Figure 6.7b). These instructions were orally transmitted to the subjects before the test (while generating the SOFA files), and further instructions or clarifications were also given under subject's demand during the familiarization round of each stage. The complete list of instruction messages can be found in Appendix B.



(a) Main menu for test routine.



(b) Instructions displayed before a new stage.

Figure 6.7: In-test subject view for stages 3 and 4.

The duration of the whole test routine varies from subject to subject, since there is no time limit through the rounds, and depending on the searching strategy of each subject, the main variation in time is noticed during the free head-movement stages. Based on the experience of the subjects tested during this thesis, this part of the routine can take from 20 min to over 1 h. A summary containing the key characteristics for each stage/round of the test can be found in Table 6.4 and more information about the configuration in Unity can be found in Appendix C.

Table 6.4: Test stages summary

Stage	Round	Iterations	Stimulus	Duration of stimulus	Wait for stimulus?	Background noise	Head Movement
1	Familiarization	5	Pink noise	3 s	Yes	N/A	No
1	Test (3x)	32	Pink noise	3 s	Yes	N/A	No
2	Familiarization	5	Audiobook	Looped	No	Babble	No
2	Test (3x)	32	Audiobook	Looped	No	Babble	No
3	Familiarization	5	Pink noise	Looped	No	N/A	Yes
3	Test (3x)	32	Pink noise	Looped	No	N/A	Yes
4	Familiarization	5	Audiobook	Looped	No	Babble	Yes
4	Test (3x)	32	Audiobook	Looped	No	Babble	Yes

Part III

Results

Chapter 7

Results

Although the main purpose of the thesis is to evaluate the use of virtual environments for HA developments testing, the subject pool for this experiment does not include hearing impaired subjects in order to avoid possible additional biasing, introduced by the type of hearing impairement or familiarization with how BTE-HA processed sound is presented to the ear. 10 subjects (3 female and 7 male) in the age range from 18 to 25 years participated in the testing phase. In order to perform a quick evaluation on the subject's possibility of having hearing damage, a questionnaire was sent before the testing session was arranged. This questionnaire (Appendix D) contains 12 questions and is directly taken from [ISO 389-9:2009]. After the evaluation of all questionnaires it was not deemed necessary to drop any of the subjects out of the test.

Subjects were assigned a 6-digit randomly generated ID number that was used through the conduction of the test and measurements in order to maintain anonymity. However, for easier identification of the subjects throughout this thesis, the following re-naming is applied only for results presentation and discussion:

Table 7.1: Subject ID relation with the Subject Nr. that will be used in this report.

Subject Nr.	Subject ID
S1	134420
S2	136624
S3	247433
S4	363189
S5	477865
S6	485756
S7	695095
S8	730012
S9	757555
S10	828039

7.1 Localization test

Prior to present the overall results, it is necessary to check the individual results for each subject in order to detect outliers. The normal probability plots of the Error number for static and dynamic stages are depicted in Figure 7.1 and Figure 7.2 respectively. The continuous lines in the plot correspond to the normal distribution fit that has been derived from estimating mean and variance from the available observations. The '+' markers depict the observations (error number per subject) from which the normal distribution fit is derived. Points being closer to the line indicate that a normal distribution is likely to underlay the observations.

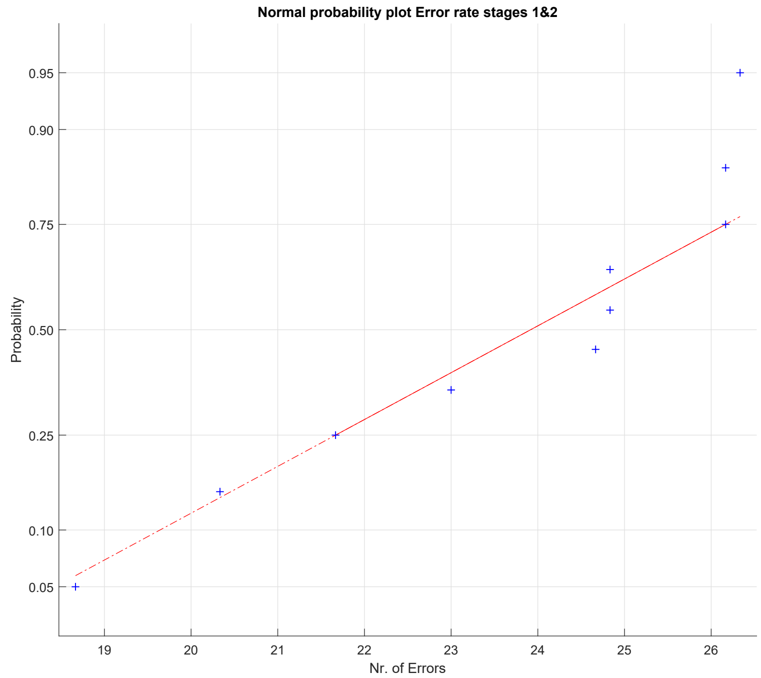


Figure 7.1: Normal probability plot with the average number of errors for each subject during the static stages (1&2).

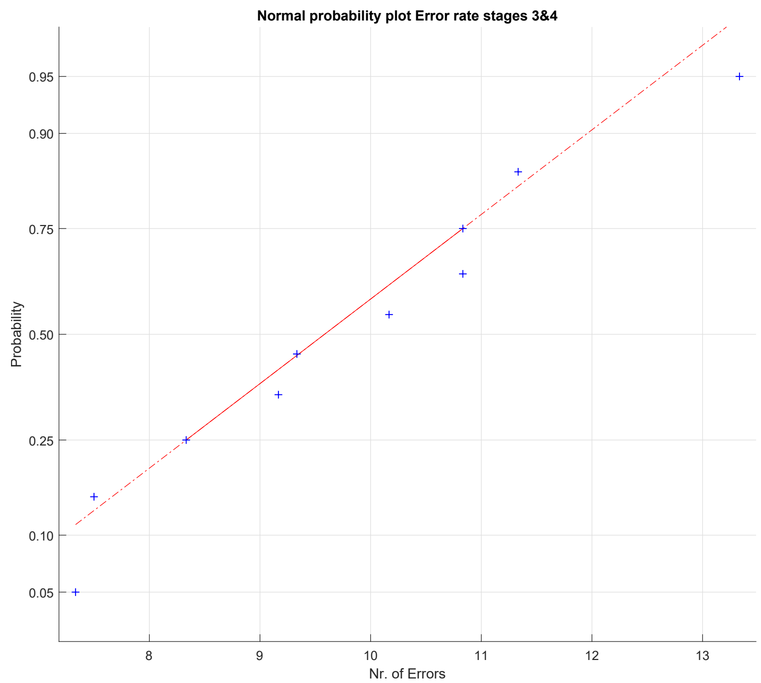


Figure 7.2: Normal probability plot with the average number of errors for each subject during the dynamic stages (3&4).

The datapoints in both figures (Figure 7.1 and Figure 7.2) can be considered reasonably close to their corresponding normal distributions, although the number of observations is small and more samples would be needed for a more accurate characterization of the distribution. Taking this into consideration, the error rate can be characterised by its mean and standard deviation, and with this information it is possible to identify possible outliers. Table 7.2 shows the mean number of errors and standard deviation for all subjects individually and as a group.

Table 7.2: Mean (μ) and standard deviation (σ) of the number of errors per subject. This calculations are done separating fixed head and dynamic stages.

	Stages 1&2		Stages 3&4	
Subject	μ_s	σ_s	μ_d	σ_d
S1	24.83	2.25	7.50	1.50
S2	23	3.50	10.83	3.51
S3	26.16	1.04	7.33	2.84
S4	18.66	6.25	10.83	2.51
S5	24.66	3.54	13.33	4.25
S6	24.83	1.60	8.33	2.08
S7	26.16	4.64	9.16	4.53
S8	21.66	3.01	11.33	2.75
S9	26.33	1.52	10.16	1.15
S10	20.33	5.00	9.33	3.21
Total	23.66	2.66	9.81	1.86

Looking at the information displayed in the table, it can be seen that the average error count for each subject/round is pretty similar, being the extreme values 18.66 and 26.33 errors for the static stages, and 7.33 and 13.33 for the dynamic ones. In the case of the static rounds, the minimum value only differs from the total μ_s by $1.87 \times \sigma_s$, and the maximum value by $\approx \sigma_s$. In the case of the dynamic rounds, the minimum value differs from the total μ_d by $1.33 \times \sigma_d$, and the maximum value by $1.89 \times \sigma_d$. Due to this, it is considered that no subjects need to be removed from the data pool for the results analysis. Therefore, from this point onwards, the results data will be processed as the total data for all subjects. The individual results for each subject can be found in Appendix E.

As explained before, Stage 1 of the test was designed to mimick Christensen's experiment in order to have a comparison point with a test performed in a real environment. The results will be presented in 16by16 matrices, one for each round and stage, in which the abscissa indicates stimulus direction and the ordinate indicates the subject's response. The answers are depicted as blue circles with area proportional to the number of answers for the particular stimulus/response combination. By doing so, right answers will be find in the diagonal of the matrix. The results of Christensen's experiment are displayed in the same way, except for the fact that he used 17by17 matrices since one more speaker was present at $\phi = 90°$.

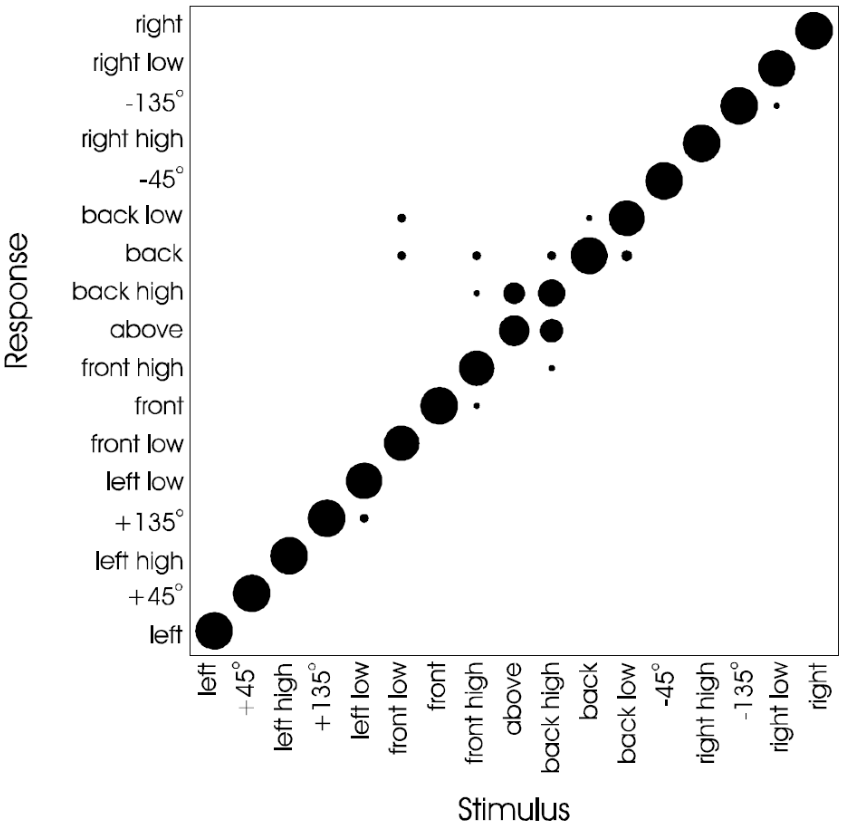


Figure 7.3: Results for Christensen's full bandwidth experiment (12 listeners, 4 repetitions). The largest circle represents 48 answers, and the smallest 1 answer. Image and caption taken from [Christensen et al., 1997].

As it can be seen in Figure 7.3, the full bandwidth experiment in anechoic conditions shows very small error rate, being the largest amount of errors confusions between the *back high* and *above*. For further analysis, Christensen divided the errors into three categories:

1. Median plane: confusions between directions in the median plane.
2. Within cone: confusions between directions in the 45° cone of confusion. Further information about the cone of confusion can be found in [Møller et al., 1996].
3. Out of cone: rest of the errors.

These error categories are maintained for the analysis of the obtained results, although with some modifications to the within cone group definition. Due to the changes in the elevation of the sources (±40° instead of ±45°), the possible cones of confusion for the lateral sides of the subject would only include two sources. However, after a preliminary analysis of the obtained data, it was detected that there was a high error rate between the following positions:

- $\theta = \pm 90°, \phi = \pm 40°$
- $\theta = \pm 45°, \phi = 0°$
- $\theta = \pm 135°, \phi = 0°$

Therefore, it was considered that an extended cone of confusion including all these positions could be use since they are separated only by 5° and the data suggests that it is easy for the subjects to confuse them. Another option would have been to define two cones of confusion, for 40° and 45°, but this would have masked confusions between these two categories as out of cone errors.

As stated before, each stage consists of three rounds. From these rounds, the one which is closer for comparison with a real life test is the one in which the set of HRTFs measured at the BEC is used since is the one that should more accurately describe how the subject perceives sound. The results for this round can be found in Figure 7.4.

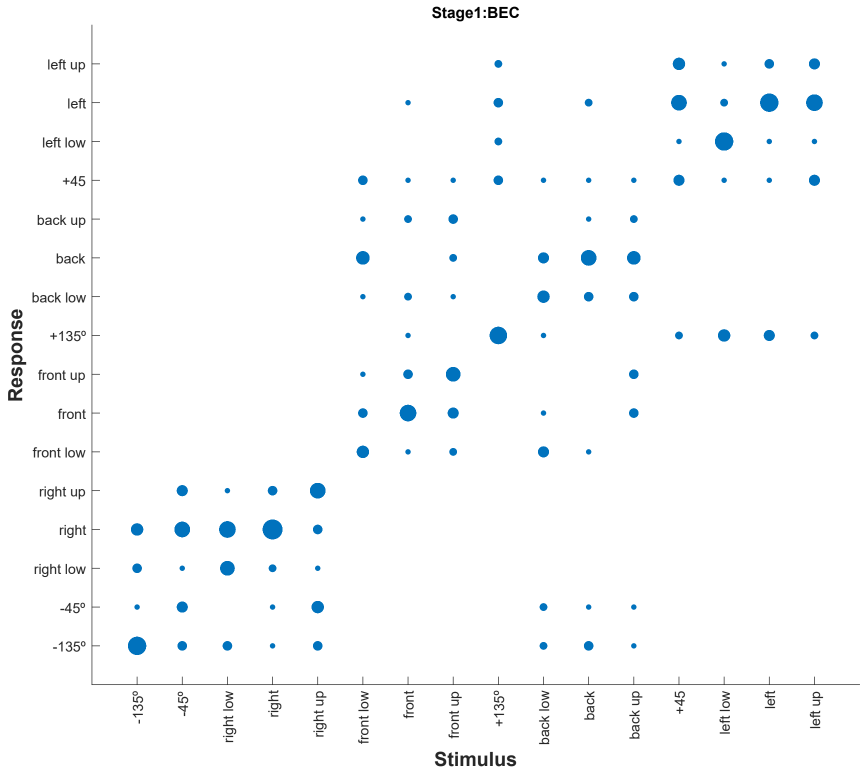


Figure 7.4: Results of the BEC round of Stage 1 (10 subjects, 2 repetitions). The largest circle represents 13 answers and the smallest one, 1 answer.

Looking at the displayed results, it is easy to see that the error rate for this test is considerably higher than the one registered during Christensen's test. Whereas in his test the maximum number of correct answers was achieved for 9 out of the 17 positions (0% error rate) and most of the answers are close to the diagonal, in the test performed in the virtual environment none of the positions registered 0% error rate, and the minimum error rate registered for any position is 35%. The results for the remaining two rounds (BTE-HA and PRF) of the stage can be found in Figure 7.5 and Figure 7.6. As it happens for the BEC round, these two rounds have a high error rate, being the lowest error rate registered for a position 55% for BTE-HA and 35% for PRF.

In terms of general error rate, as opposed to each position individually, the percentage of errors committed in each category can be found in Figure 7.7. The error rate for each category is calculated with the number of errors registered in it over the total number of answers registered during the round. As it can be observed, the error rates are high in general for the three categories, being the highest rates registered in the BTE-HA round.

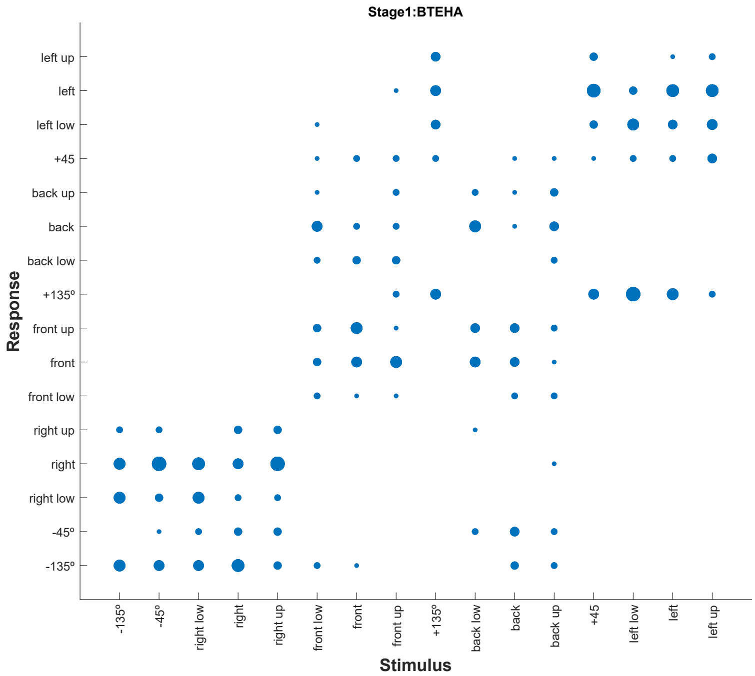


Figure 7.5: Results of the BTE-HA round of Stage 1 (10 subjects, 2 repetitions). The largest circle represents 9 answers and the smallest one, 1 answer.

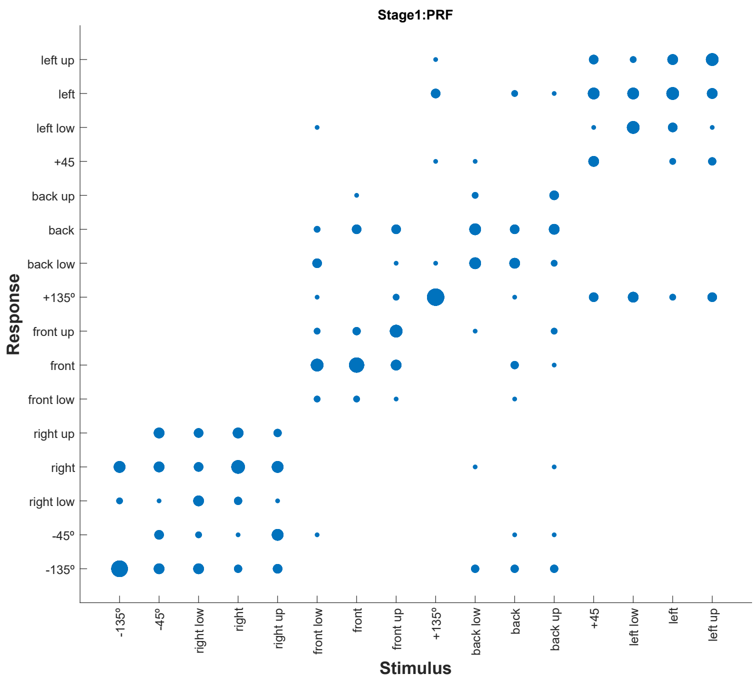


Figure 7.6: Results of the PRF round of Stage 1 (10 subjects, 2 repetitions). The largest circle represents 13 answers and the smallest one, 1 answer.

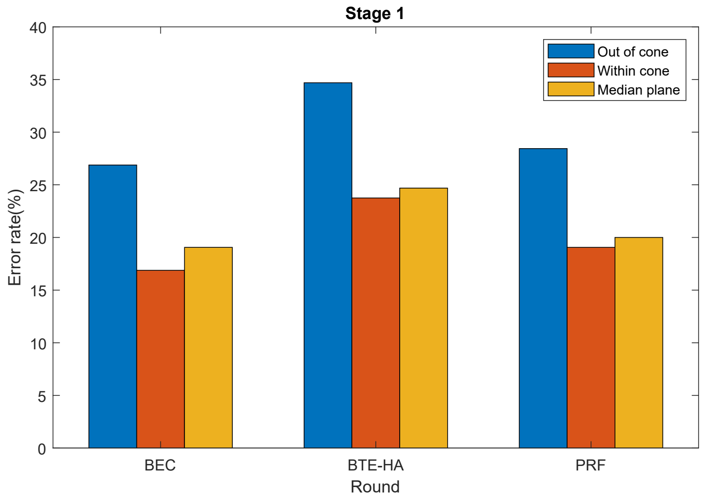


Figure 7.7: Error rate per category and round for Stage 1. Error rate =nr.errors in the category/total answers.

Stage 2 presents a different situation than Stage 1 in terms of soundscape, but they are equal in terms of head movement. Giving a first look at the results obtained during this stage (Figure 7.9, Figure 7.10 and Figure 7.14) it can be seen that the answers are even more spreaded than for Stage 1, meaning a higher error rate. For this stage, the lowest error rate for a position found in the BEC and PRF rounds is 35%, and 50% in the BTE-HA. These values are very similar to the ones obtained in Stage 1, but when looking at the categorized errors(Figure 7.8) it can be observed that the Out of cone error rate in Stage 2 is higher for the three rounds, while the Within cone and Median plane error rates are similar to the ones found in Stage 1.

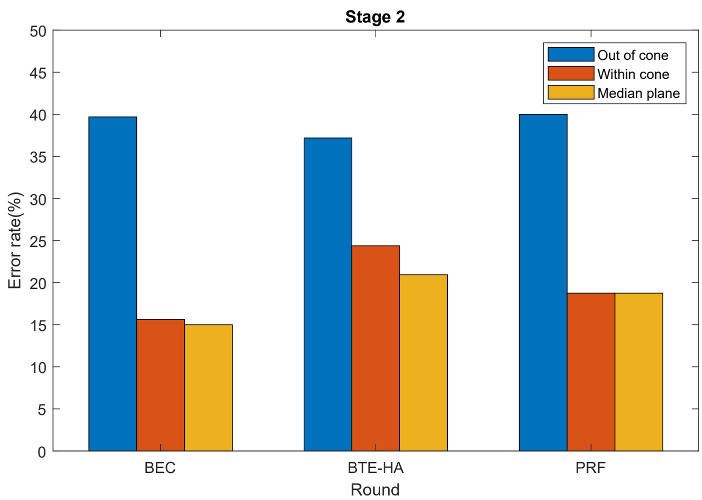


Figure 7.8: Error rate per category and round for Stage 2. Error rate =nr.errors in the category/total answers.

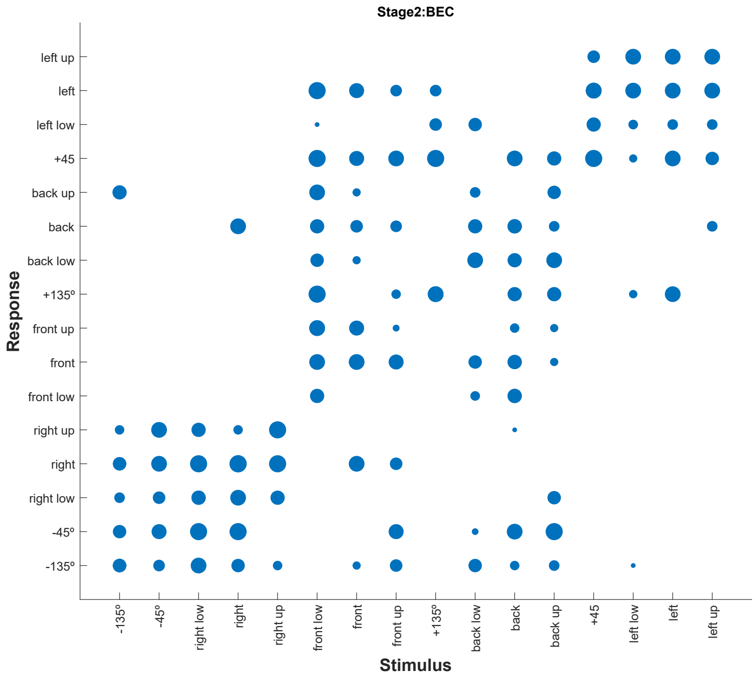


Figure 7.9: Results of the BEC round of Stage 2 (10 subjects, 2 repetitions). The largest circle represents 13 answers and the smallest one, 1 answer.

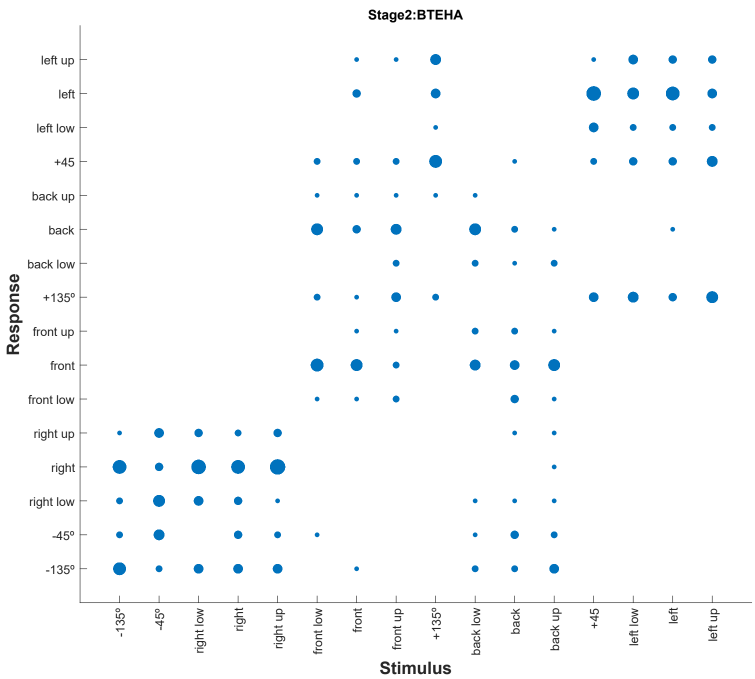


Figure 7.10: Results of the BTE-HA round of Stage 2 (10 subjects, 2 repetitions). The largest circle represents 10 answers and the smallest one, 1 answer.

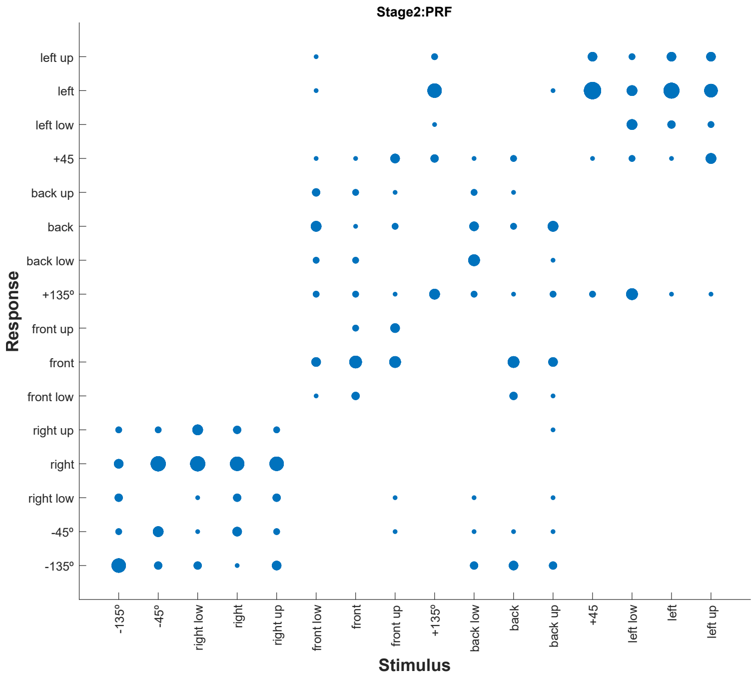


Figure 7.11: Results of the PRF round of Stage 2 (10 subjects, 2 repetitions). The largest circle represents 13 answers and the smallest one, 1 answer.

As explained before, free head movement is introduced in stages 3 and 4, with Stage 3 sharing the soundscape of Stage 1, and Stage 4 the soundscape of Stage 2. The introduction of head movement changes the way in which the errors fall into their corresponding category. For this two stages, it is anticipated not to find Within cone errors, and since the subject will always be looking at the source in evaluation, it can be anticipated that most errors should fall under the Median plane category, with low to no errors in the Out of cone category. This factor should reduce the number of overall errors, and this hypothesis can be confirmed when looking at the results for these stages as it can be observed that most of the registered answeres lay in the diagonal, with some close outliers. Results for Stage 3 are depicted in Figure 7.12, Figure 7.13 and Figure 7.14, while results for Stage 4 are depicted in Figure 7.15, Figure 7.16 and Figure 7.17.

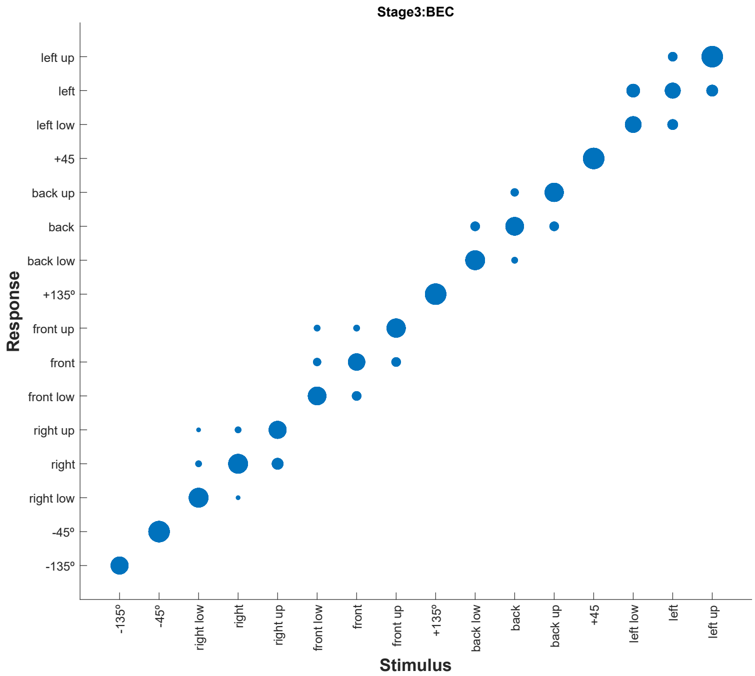


Figure 7.12: Results of the BEC round of Stage 3 (10 subjects, 2 repetitions). The largest circle represents 20 answers and the smallest one, 1 answer.

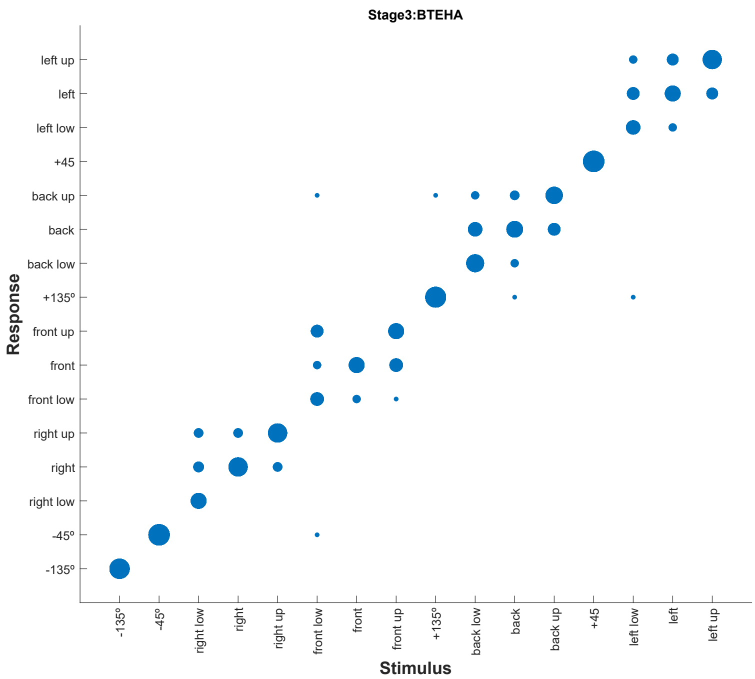


Figure 7.13: Results of the BTE-HA round of Stage 3 (10 subjects, 2 repetitions). The largest circle represents 20 answers and the smallest one, 1 answer.

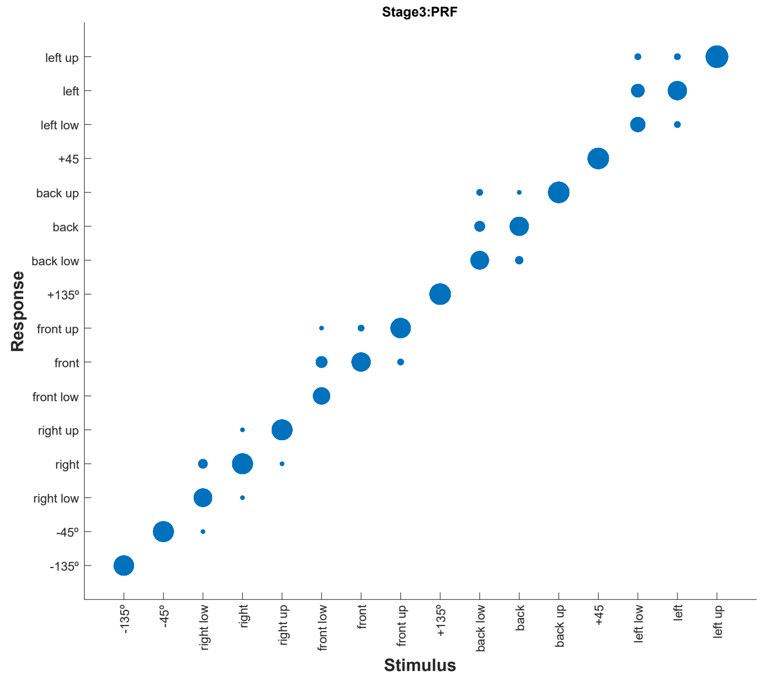


Figure 7.14: Results of the PRF round of Stage 3 (10 subjects, 2 repetitions). The largest circle represents 20 answers and the smallest one, 1 answer.

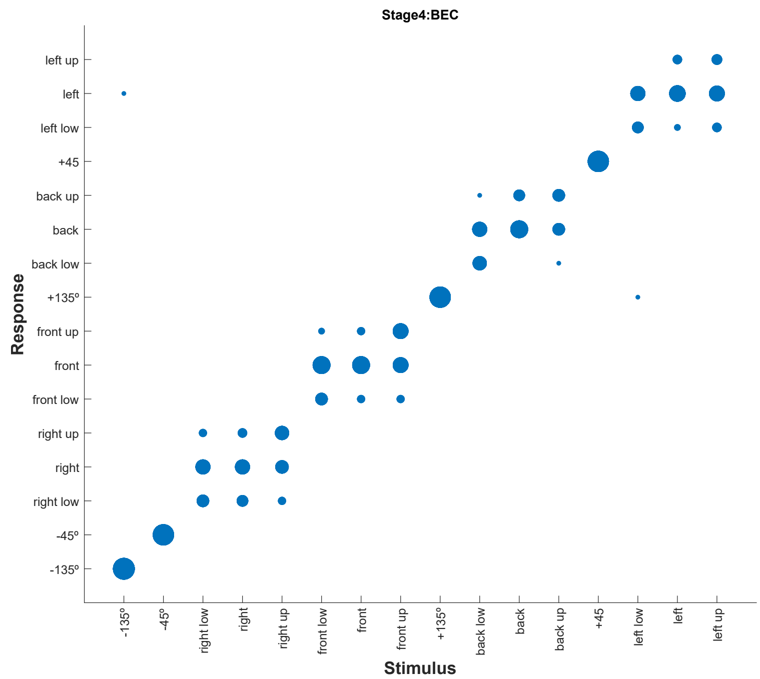


Figure 7.15: Results of the BEC round of Stage 4 (10 subjects, 2 repetitions). The largest circle represents 20 answers and the smallest one, 1 answer.

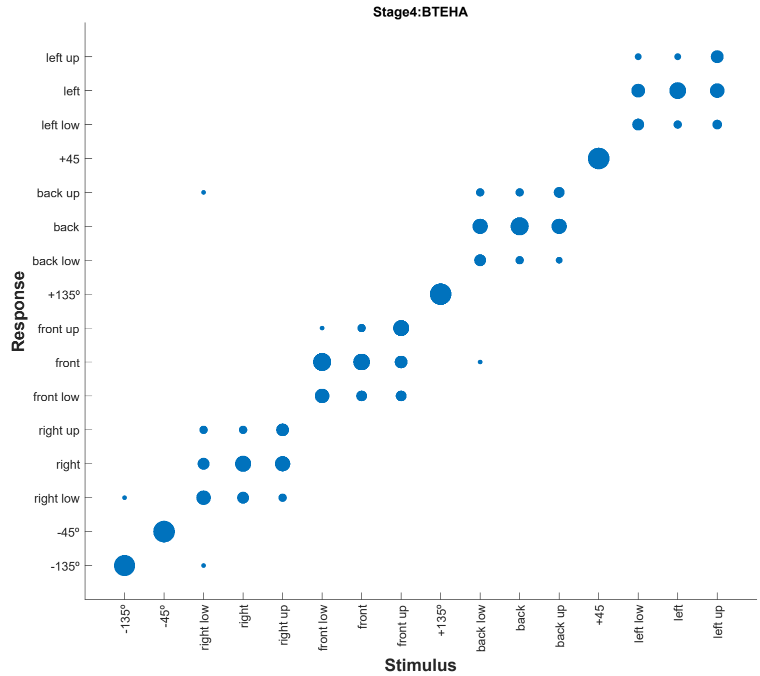


Figure 7.16: Results of the BTE-HA round of Stage 4 (10 subjects, 2 repetitions). The largest circle represents 20 answers and the smallest one, 1 answer.

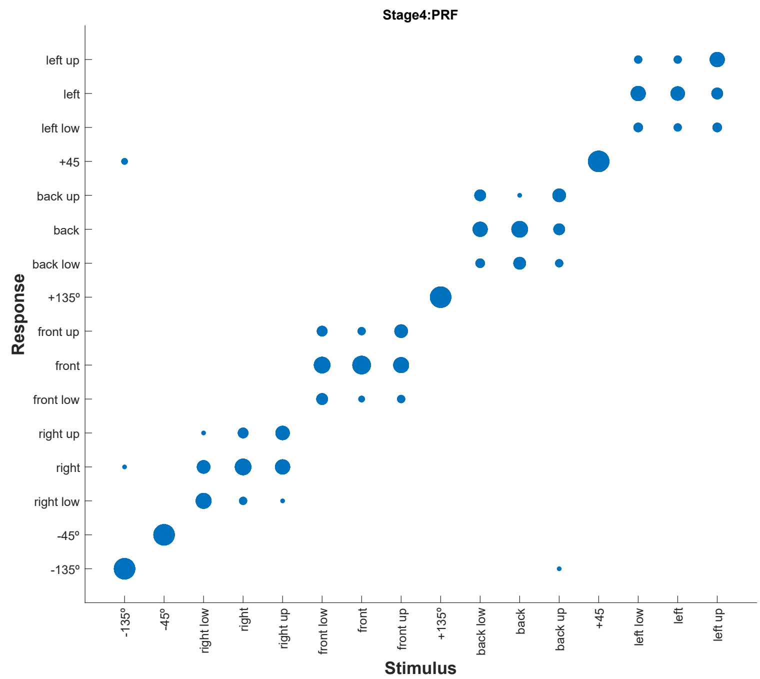


Figure 7.17: Results of the PRF round of Stage 4 (10 subjects, 2 repetitions). The largest circle represents 20 answers and the smallest one, 1 answer.

It can be observed that for all the rounds included in these two stages, there's at least one position in which an error rate of 0% was registered. When looking at the error rate per category for each stage, Figure 7.18 for Stage 3 and Figure 7.19 for Stage 4, the hypothesis that the majority of errors would belong to the Median plane category is confirmed. Since the movement of the head is allowed, it is possible to differentiate between the positions in the cone of confusion, having 0% error rate in the Within cone category. Error rates lower than 1.30% are found in the Out of cone category.

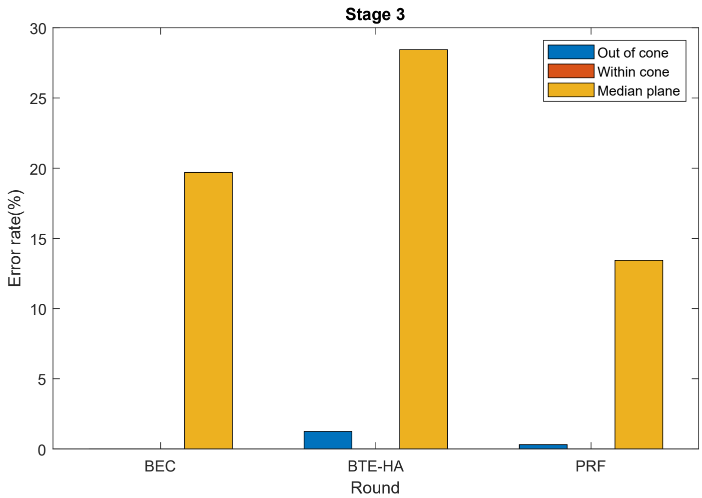


Figure 7.18: Error rate per category and round for Stage 3. Error rate =nr.errors in the category/total answers.

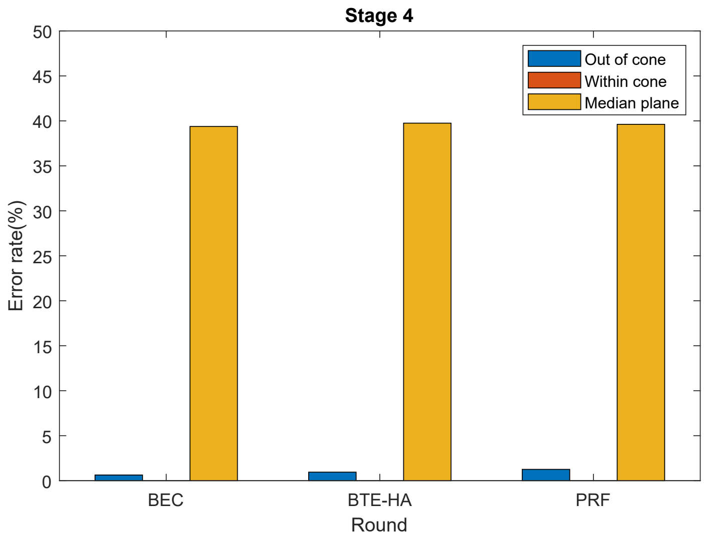


Figure 7.19: Error rate per category and round for Stage 4. Error rate =nr.errors in the category/total answers.

A summary of the test results for all stages/rounds can be found in Table 7.3 and a summary of the Error rate per round and stage can be found in Figure 7.20. With the data extracted from these two sources, it is easily seen that the error rate is higher in the stages in which the head position is fixed. It can also be observed that the error rate is lower for the rounds in which the stimulus is the fullband pink noise with no background noise.

Table 7.3: Results for all subjects. The percentages are calculated with respect to the total number of trials. Notice that for some of the rounds there are less than 320 iterations. This is because 5 of the subjects experienced a previously undetected bug causing the stimulus not to play at Stage 4 Round 3 Iteration 2.

Stage	**Round**	**Iterations**	**Total Errors**	**Out of cone Errors**	**Within cone Errors**	**Median plane Errors**
1	BEC	320	201 (62.81%)	86 (26.88%)	54 (16.88%)	61 (19.06%)
1	BTEHA	320	266 (83.13%)	111 (34.69%)	76 (23.75%)	79 (24.69%)
1	PRF	320	216 (67.50%)	91 (28.44%)	61 (19.06%)	64 (20.00%)
2	BEC	320	225 (70.31%)	127 (39.69%)	50 (15.63%)	48 (15.00%)
2	BTEHA	320	264 (82.50%)	119 (37.19%)	78 (24.38%)	67 (20.94%)
2	PRF	320	248 (77.50%)	128 (40.00%)	60 (18.75%)	60 (18.75%)
3	BEC	320	63 (19.69%)	0 (0.00%)	0 (0.00%)	63 (19.69%)
3	BTEHA	320	95 (29.69%)	4 (1.25%)	0 (0.00%)	91 (28.44%)
3	PRF	320	44 (13.75%)	1 (0.31%)	0 (0.00%)	43 (13.44%)
4	BEC	320	128 (40.00%)	2 (0.63%)	0 (0.00%)	126 (39.38%)
4	BTEHA	317	129 (40.69%)	3 (0.95%)	0 (0.00%)	126 (39.75%)
4	PRF	318	130 (40.88%)	4 (1.26%)	0 (0.00%)	126 (39.62%)

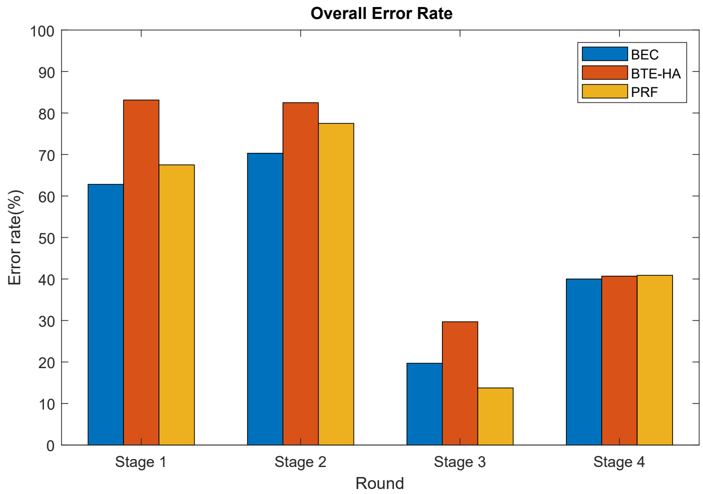


Figure 7.20: Total error rate per round and stage. Error rate = nr.errors in the category/total answers.

7.2 HRTF set performance

By looking at the previously presented data, trends in the performance of the different sets of HRTF used can be observed. In general, it can be seen that the BEC set has the best performance (lowest error rate), being Stage 3 an exception, followed by the PRF set and thus being the BTE-HA set the one displaying the worst performance (highest error rate). Since each of the stages has individual characteristics that make it different from the others, it is not possible to analyze and compare individual results with the outcome being statistically significant. This is due to the fact that even when grouping the results in two groups (locked-head and free-movement), the number of samples for each group and set would be 20. In order to include more data samples, it has been decided to evaluate the performance in terms of error rate difference between the different sets of HRTFs. By doing this, it is possible to group all error rates in three categories, corresponding to the HRTF sets, and not taking into account the stages they belong to. By doing this, each one of the data vectors will have 40 samples instead of 20. Furthermore, this method will allow to evaluate the performance of the sets independently of the soundscape presented or the conditions of the test.

In order to determine if there is a statistically significant difference in performance between the three sets presented, several paired-sample t-tests can be performed. The null hypothesis of the test is that the difference between the observation pairs is normally distributed and zero mean. If this hypothesis can be rejected, it can be concluded that there is a statistically significant difference between the performance of the three HRTF sets. However, for the paired-sample t-test to be viable, it has to be assumed that the pairs of observations are normally distributed. Therefore, the distribution of the difference between them also has to be normally distributed. In order to check this, the same method as in the beginning of the section is used, and the normal probability plots of the different set errors are analyzed.

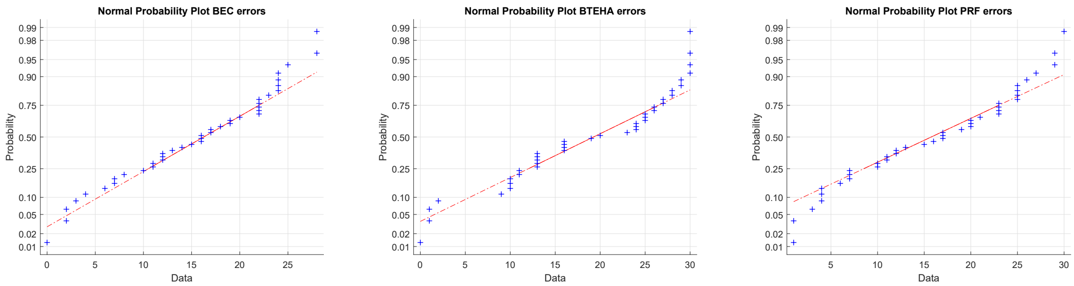


Figure 7.21: Normal probability plot of the raw number of errors of all subjects for each HRTF set, BEC (left), BTE-HA (middle) and PRF (right).

Based on the data observed in Figure 7.21, it can be considered that the observations for each set lay reasonably close to their corresponding normal distribution fits. Therefore, it can be assumed that the observations sets are normally distributed, allowing to perform a paired-sample t-test. Three tests will be conducted, BEC vs BTE-HA, BEC vs PRF and PRF vs BTE-HA. For the case of BEC vs BTE-HA the hypothesis can be rejected with a p-value of 0.0033, for BEC vs PRF, p-value of 0.4195, and for PRF vs BTE-HA, p-value of 0.0056. This means that there is a statistically significant difference in the performance of the BEC and the PRF against the BTE-HA, since their p-values lay below the commonly applied significance level of $\alpha = 5\%$. However, the difference in performance between the PRF set and the BEC set cannot be considered statistically significant.

Chapter 8

Discussion and conclusion

8.1 Discussion: Results

The results obtained during the listening test are meant to help analyse the two main aspects investigated in this project, feasability of conducting specialized listening tests in Unity with the use of Steam Audio, and performance comparison between different sets of HRTFs within the selected framework.

For the first aspect, it is difficult to establish a solid baseline in terms of results obtained, since no literature was found on experiments performed in the same or a similar framework. Furthermore, it is intended to compare the performance of localization tests in this specific virtual environment with the one in a real environment. Thus, in order to evaluate the test's performance, the results of the BEC round from Stage 1 are compared to Christensen's results. As shown in section 7.1, the error rate registered during the round of interest of this thesis' experiment is considerably higher than the one registered by Christensen's test [Christensen et al., 1997]. The lower performance of the virtual test indicates that the feasability of using this system is in question. This is also sustained by comparing the results obtained with results from experiments conducted in different virtual environments, such as [Møller, 1992] and [Mueller et al., 2012]. While the methods of these experiments are not exactly the same as the one used during this test, the results obtained indicated a lower error rate in terms of localizing sources within a virtual environment with a fixed head position. The experiment presented in [Møller, 1992] can be considered the closest in terms of method to the one presented in this thesis, and when looking at its results for the virtual localization test (Figure 8.1) it is easily seen that the response distribution is closer to the one obtained by Christensen (Figure 7.3) than to the one obtained during this thesis' test (Figure 7.4).

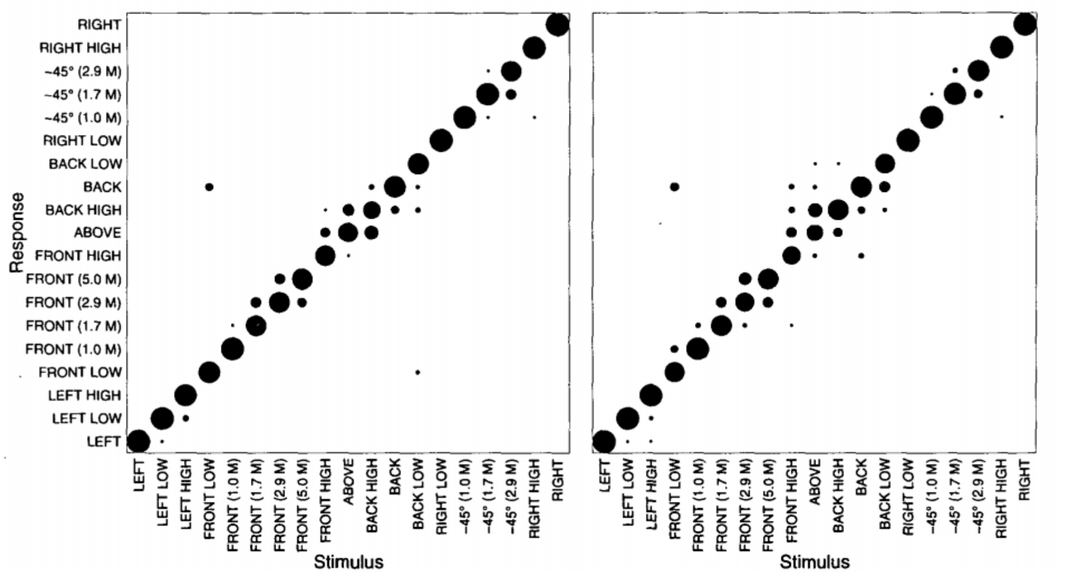


Figure 8.1: Individual binaural recordings. (a) In separate sessions (experiment B, 912 responses). (b) In mixed sessions (part of experiment C, 912 responses). Figure and caption directly taken from [Møller, 1992].

Therefore, it can be concluded that the error rate obtained during the test is considered to be too high for the implementation of localization tests in this specific configuration. Further research would be needed, but the audio spatializer seems to be the limiting element of the framework. While it is usual to encounter front/back confusion situations for binaural rendered sounds, the results for stages 1 and 2 (fixed-head stages) show general difficulties for the subjects to locate the sources in general, and not only in the back/front cases, even registering the *left* position as an answer for a stimulus playing from the *front low*. Furthermore, even in the free head-movement rounds a large amount of errors can be found, being the lowest error rate registered 13.75%, with almost all of them happening in the median plane. In addition to these results, feedback from the subjects was taken at the end of the test, and most of them agreed on high levels of confusion within the first two stages, difficulties telling up from down in stages 3 and 4, and generally a lower sense of direction for the stages with background noise. Further steps in the evaluation of the framework could consist on more focused resolution tests, mainly in the vertical plane. Nevertheless, the audio spatializer used seems not to be fit for the implementation of specialized tests, and the implementation of a custom audio renderer would be advised.

Regarding the performance of the different sets used, the trend of the BTE-HA to score the worst performance of the three sets tested is confirmed, which coincides with the results obtained by [Udesen et al., 2013] and [Van den Bogaert et al., 2006]. The fact that the PRF and BEC sets had a similar performance and that the hypothesis presented by the paired-sample t-test could not be rejected means that the implemented filters fullfilled their functions correctly, being able to closely mimick the behaviour of the BEC.

8.2 Discussion: Methods

HRTF measurement

The measurement of the individual HRTF sets is the backbone of the conducted experiment, since it conditions the sound perception presented to the subject. Several aspects of the measurement process were considered to have room for improvement after observations made during the measurement sessions and input from the subjects.

First of all, in order to include a complete characterisation of the BTE-HA in the measurement of its set of HRTF it would have been necessary to extract the microphone from the plastic housing of the hanger for the reference measurement. This was not possible during the development of this thesis, and all measurements were taken with the microphones inside the hanger. This leads to the effect of the plastic housing to be removed from the HRTF (in the same way as the microphone and speaker responses do). Furthermore, these hangers are meant to be kept in place through the use of the RIC. With the absence of the RIC and the presence of the stiff cables connecting the biasing circuitry to the hangers, deviations from the correct fit and positioning of the BTE-HA could be encountered during the execution of the measurements. The adverse effects of the stiff cables and lack of RIC were countermeasured by the use of the custom harness, medical tape and cable placing, but it was not possible to ensure the correct fitting of the hangers. In addition, no way of controlling position changes of the hanger throughout the measuring session was implemented. It is advised to find a way to control these factors and ensure a correct fitting of the HA for a more acurate measurement of their corresponding HRTFs.

In terms of the in-ear microphones, the fitting and positioning could also be improved. The placement method during this thesis was based on manually trimming and indenting foam earplugs. By trimming the earbuds, it was intended not to have the microphone sticking out of the ear canal, and the indent was made in order to fit the microphone in the earplug. However, the trimming was not done independently for each subject, which can cause the microphones to be in different positions of the ear canal depending on the subject. The indenting was also manually done, and therefore it could not be ensured that all earplugs would have the same depth and width in terms of the indent, meaning that the fit of the microphone could be variable between subjects. In order to ensure fitting and positioning it would have been necessary to provide each subject with personalised earplugs.

Another important factor to take into account is that subjects encountered difficulties in standing completely still and in silence during the whole measurement session ($\approx$28 min in average). Standing still the whole time proved to be difficult since the platform is quite small, meaning that the spread of the legs is minimized, making it harder to stand for a prolongued amount of time. The approach to minimize this problem was to include little breaks every 90° for the subject to move while still standing on the platform. This helped the subjects relax, but most of them still found it difficult to stand in the same position all the time. This problem could be further minimised by including some kind of upholder in the lower back support frame for the subject to slightly lean on and partially sit. If this support is implemented with the correct size, the possible reflections and general interference of the stand in the measurement could be minimized or even removed. In addition, it was noticed during the measurements that it was difficult to ensure the same head position for the whole duration of the measurement since the subjects did not have a reference to look at and small movements could pass by undetected by the test conductor.

Listening test

[Majdak et al., 2010] researched the importance of visual environment, pointing method and training in 3-D localization tests conducted in virtual environments. Their findings show that the localization performance registered did not significantly vary between the different pointing methods tested (head and hand pointing), but in order to avoid any possible bias that could be introduced by the subject being right or left handed, it was decided to use a head-pointing system during the tests. However, it is believed that the pointing system could have been improved by conducting the test in an inmersive Virtual Reality (VR) environment, with the use and headhphones and VR goggles instead of headphones and a computer screen. Furthermore, this method would have allowed for more natural responses in terms of head movement and overall source localization techniques by the subjects. It was observed during the test that subjects with a background in first-person videogames were faster to localize the sources, performing quick movements with the mouse, whereas subjects with low or none first-person videogame experience showed slower searching techniques, usually slow sweeps, and slow reactions. Taking into account the feedback orally given by subjects in the end of the test, the subjects with poor videogame experience expressed difficulties in sound localization for stages 3 and 4, while subjects with more experience did not. This leads to the importance of previous training for tests in VR environments, since previous experience with VR tests and/or videogames could bias the performance of the subjects. Looking at the test in retrospective, one aspect that could have been changed is the length and execution of the familiarization rounds. Several subjects transmitted difficulties in localizing the sources without having a reference in terms of how the sources from each direction sound. Instead of a randomized, limited-sample familiarization round with no feedback, the test could have included a familiarization round in which all sources are presented with a reference to the playing source. [Majdak et al., 2010] showed significant improvements in performance during the training of subjects, so it would be interesting to further investigate this aspect, by comparing trained and untrained subjects' results within the same test.

Another aspect of the test that could be improved is its duration and break distribution. As the test is right now, the average duration lays between 40 min and 1 h, with variable duration breaks in between rounds and stages, with length depending on the subject's assessment. However, it was noticed that the majority of the subjects wanted to minimize the time spent on the test, and therefore chose to have small breaks or no breaks. This could have lead to a decrease in performance during the last rounds, induced by the subjects being tired and more concerned about finishing the test than about the answers given. This aspect could be solved by introducing fixed duration breaks at key points, at the expense of extending the test's duration. Furthermore, subjects also transmitted that during stages 2 and 4, the stimulus was repetitive and became more annoying the more the test advanced. This could be improved by randomizing the starting point of the stimulus track between iterations.

8.3 Conclusion

As stated in chapter 1, the main objective of this thesis is to assess the feasability of performing specialized listening tests (such as localization tests) in virtual environments, as well as to evaluate the performance of different HRTF sets relevant to the development of HA in such environment. It was decided to evaluate a set measured at the BEC (as a characterization of the normal hearing), a set measured with the BTE-HA (as a characterization of listening through HA), and a set consisting on the latter set filtered with PRF generated from the first set. In order to do so, a test routine composed by an HRTF measurement session and a subject-based localization test was designed and implemented. The listening test was based on an experiment conducted in a real environment [Christensen et al., 1997], which was also used as an anchor for comparison with a real environment test. The measurement routine was designed for simultaneous obtention of the three sets of HRTFs in anechoic conditions, with a special focus on minimising measurement time due to the difficulties that these measurements impose on the subjects. One of the core aspects of the listening test presented in this thesis was the selection of a suitable virtual environment for this purpose, with focus on accessibility and possibility of widespread use. As explained in chapter 3, this virtual environment is composed by a virtual engine and an audio spatializer (or audio renderer). After several options and combinations were investigated and tested, it was decided to use Unity with Steam Audio for the development of the test. A localization test with 4 stages presenting 4 different situations was implemented, being the first stage the one comparable with the anchor test.

Analysis of the results reveals that the different sets of HRTFs used registered different performances. In addition, the hypothesis of the BTE-HA set having a worse performance, based on the results obtained by [Udesen et al., 2013] and [Van den Bogaert et al., 2006], was confirmed. However, the general results for the source localization show a considerable amount of errors when comparing it to the anchor test. The registered error rate is also high when comparing it to other tests conducted in different virtual environments, which, although not directly comparable, show trends more similar to the anchor test. Therefore, the current virtual environment used is considered not to be suitable for the desired tests, with the suspicion of the audio spatializer being the main problem. Further investigation in this area would be needed in order to pinpoint what is causing the issue with more accuracy. Another approach for future research in this field would be to implement a different spatializer aimed at specialized listening tests.

Part IV

Appendix

Appendix A

Measurement and test equipment

Table A.1: List of equipment used for HRTF measurements.

Description	Model	Serial Nr.	AAU Nr.
Microphone (x2)	Knowles FG-23329	-	-
HA Hangers (x2)	Widex Fusion	-	-
HA Microphone (x2)	Sonion 8002	-	-
Audio Interface	RME Fireface UFX II	-	108228
Power Amplifier	Pioneer A616	HJ9405069S	08341
USB-parallel converter	Arduino UNO	-	-
DC Power Supply	B&O SN17	3292110	08111
Speaker (x11)	Vifa M10MD39-08 driver in ball mount	-	-
Turntable	Outline ET 250-3D	REIBO012	-
12 Channel Relay Board	AAU custom	-	-
Laptop	Asus GL752	-	-
Screen	Samsung SyncMaster 2493	-	-
USB mouse and keyboard	Logitech	-	-
USB KVM Extender	Proxime CE700A	-	-
Chest mount	GoPro	-	-
Earplugs	E.A.R. Classic	-	-
Laser guides (x2)	Stanley Cubix	-	-
Audio Interface	RME Fireface 802	-	86838
Laptop	Lenovo ThinkPad	-	115512
Headphones	Beyerdynamic DT990	-	-
Screen	Dell U2412M	-	-

Appendix B

Test tutorial messages

These are the messages used in the tutorial panels of the test in order of appearance:

- You are located into a sphere of loudspeakers. Sound will play from the different loudspeakers, one at a time. You will have to identify the source from where the sound is coming from and click it in the controller as soon as it stops.
- You will go through a test run. The controller displays all possible sources. Take into account that there are 4 positions that have 3 sources at them (Up, Mid and Down). For these cases, instead of clicking the speaker icon for the position, you need to press the specific location displayed in text.
- End of the tutorial, please press 'Start Test' when you are ready. [Repeated after each stage's familiarization round]
- Round finished. [Repeated after each round]
- End of the first stage. Please press "Next Stage" when you are ready to continue the test. [Repeated after each stage increasing the stage number]
- For this stage of the test, you are located in the exact same environment as before. However, now there will be background noise as well as the stimulus signal.
- The background noise is babble (mix of different speech signals), and the stimulus to identify is the narrating woman that sounds slightly louder than the noise.
- You will go through a test run first. The controller works as in the previous stage.
- For the next stage, instead of a fixed head position and a controller, you will be able to look around using the mouse, and selecting the source by clicking it. For that purpose, sources now highlight in yellow when the mouse is over them.
- In order to ease source identification, the colour coding from the previous rounds will be maintained.

- First, you will go through a test run. As you will see, the look direction is indicated by a pointer in the screen. For this stage, you can select a source as soon as you identify it, even if the sound is still playing.
- As the previous stage, this one allows for head movements, and the source selection mechanics are unaltered. The difference is that in this case, there will be background noise and you will have to identify the female speaker (as in Stage 2).
- First, you will go through a test run. As in the previous stage, you can select a source before the stimulus has stopped.
- End of the test. Thank you very much for participating!

Appendix C

Unity project settings

The localization test is fully implemented and conducted in Unity 2019.1.2f1 (after updates).

C.1 Scenes

Each of the stages of the test is implemented as a separate Scene in Unity for simplicity.

Menu.unity

This scene contains the main menu and it's the starting point of the test. It has a scene manager script (MainMenu.cs) attached to the MainMenu GameObject. It stores and passes the subject ID to the rest of the scenes. After modifying the randomization processs for the source order for each round, it also initializes the order lists and forwards them to the rest of scenes.

anchor_round.unity

This scene contains Stage 1 of the test. It has a scene manager script (AnchorSceneManager.cs) attached to the SceneManager GameObject. The settings used for initializing this script are displayed in Figure C.1.

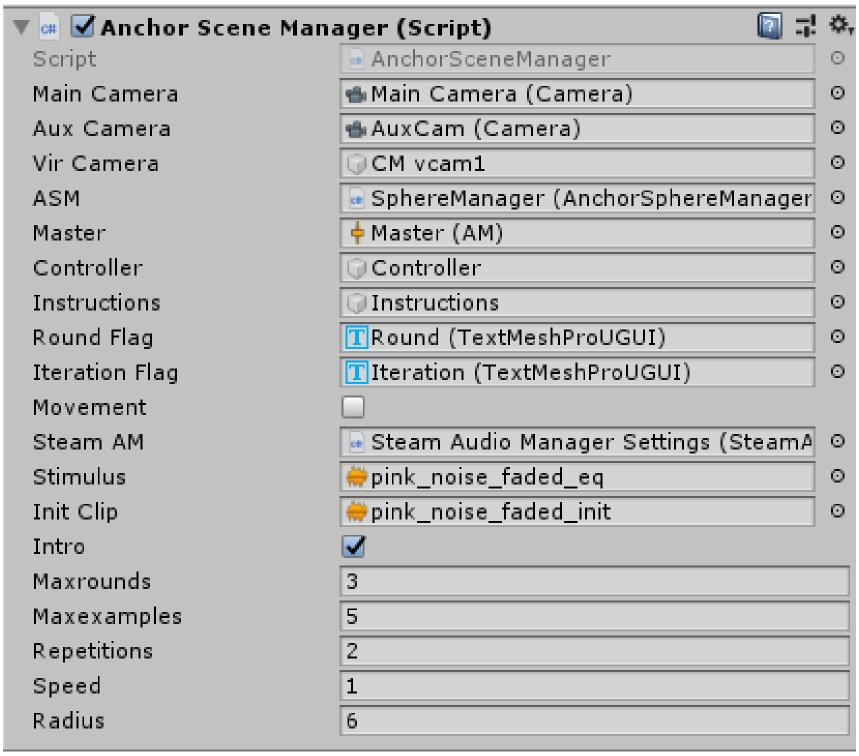


Figure C.1: Settings for the scene manager of Stage 1.

It also contains a sphere generation script (AnchorSphereManager.cs), set up as depicted in Figure C.2.

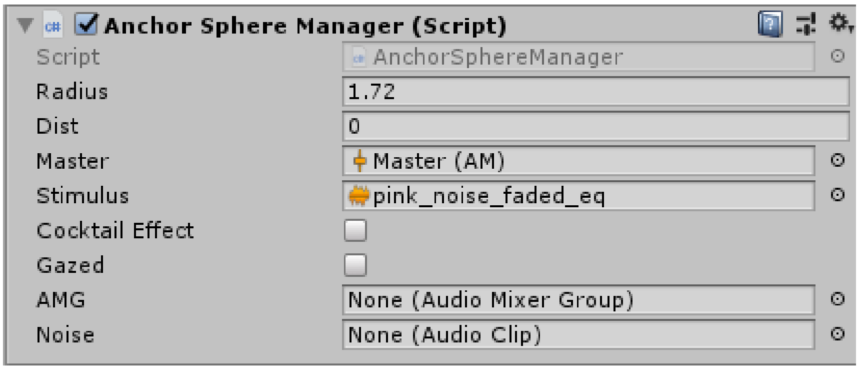


Figure C.2: Settings for the sphere manager of Stage 1.

anchor_round_cocktail.unity

This scene contains Stage 2 of the test. It has a scene manager script (AnchorSceneManager.cs) attached to the SceneManager GameObject. The settings used for initializing this script are displayed in Figure C.3.

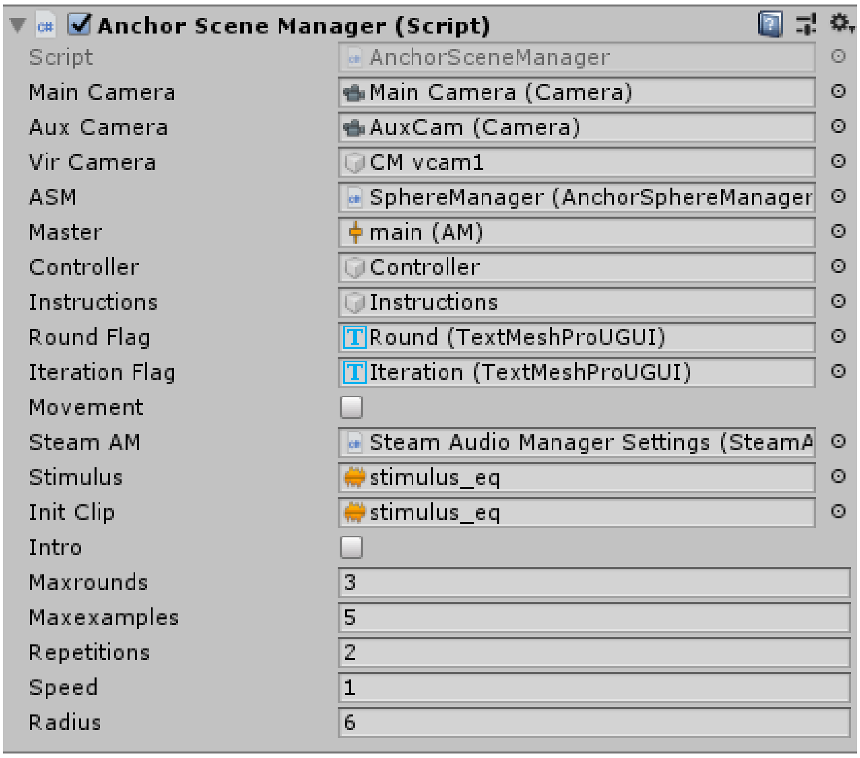


Figure C.3: Settings for the scene manager of Stage 2.

It also contains a sphere generation script (AnchorSphereManager.cs), set up as depicted in Figure C.4.

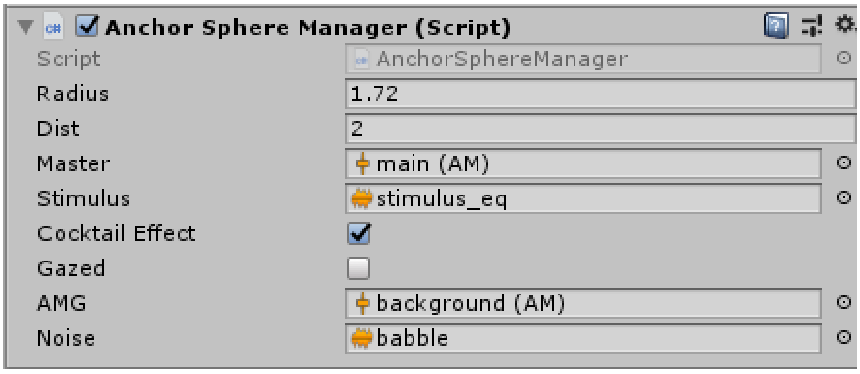


Figure C.4: Settings for the sphere manager of Stage 2.

freemov_round.unity

This scene contains Stage 3 of the test. It has a scene manager script (FreemovSceneManager.cs) attached to the SceneManager GameObject. The settings used for initializing this script are displayed in Figure C.5.

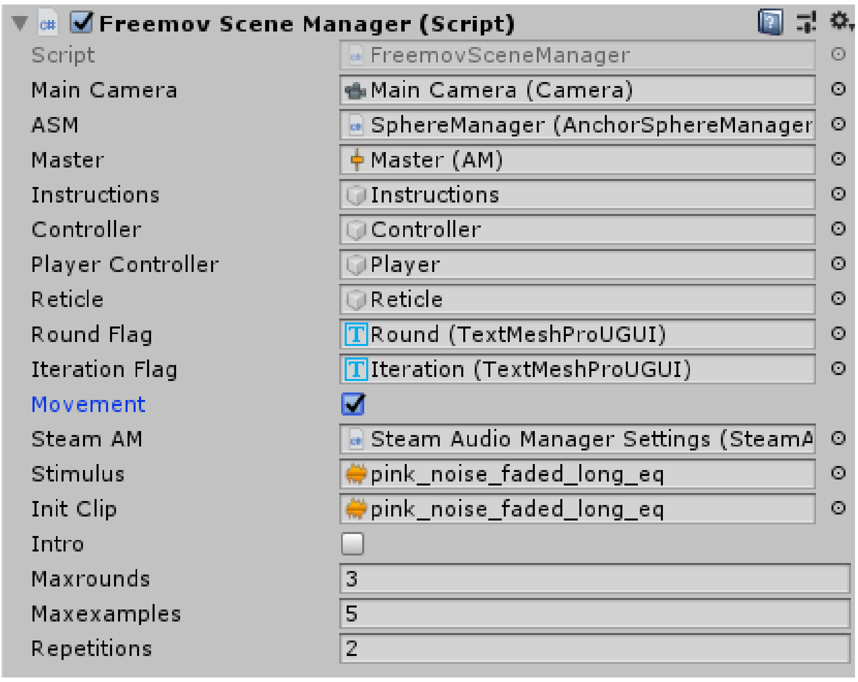


Figure C.5: Settings for the scene manager of Stage 3.

It also contains a sphere generation script (AnchorSphereManager.cs), set up as depicted in Figure C.6.

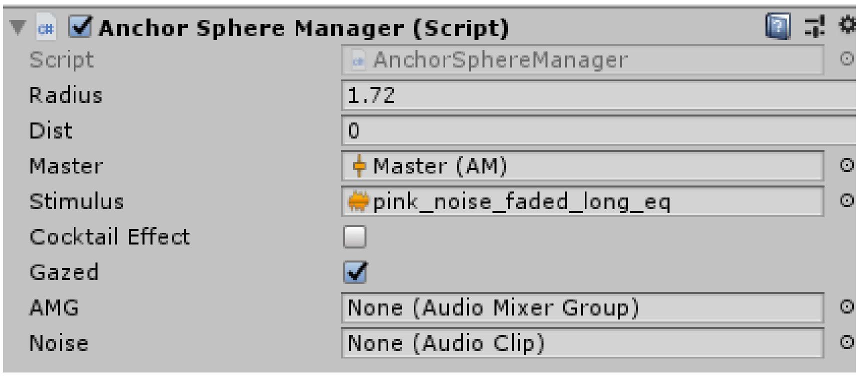


Figure C.6: Settings for the sphere manager of Stage 3.

freemov_round_cocktail.unity

This scene contains Stage 4 of the test. It has a scene manager script (FreemovSceneManager.cs) attached to the SceneManager GameObject. The settings used for initializing this script are displayed in Figure C.7.

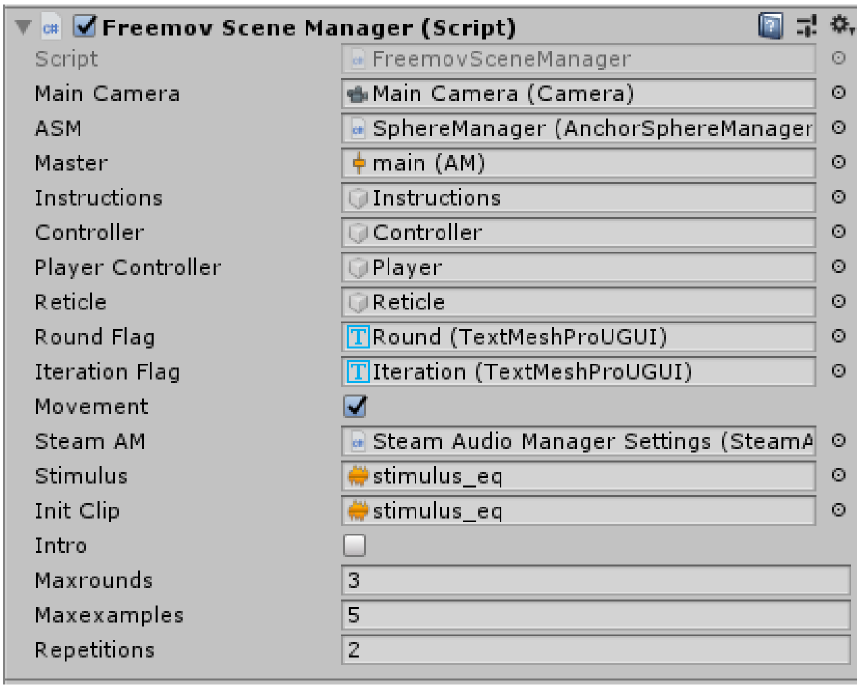


Figure C.7: Settings for the scene manager of Stage 4.

It also contains a sphere generation script (AnchorSphereManager.cs), set up as depicted in Figure C.8.

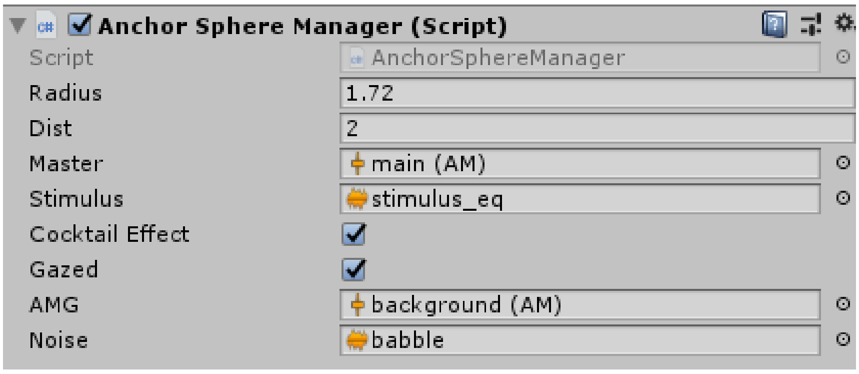


Figure C.8: Settings for the sphere manager of Stage 4.

C.2 General Settings

These settings are maintained the same for the whole project.

Steam Audio Manager Settings

Attached to every scene through Window->Steam Audio Manager, its settings are the following:

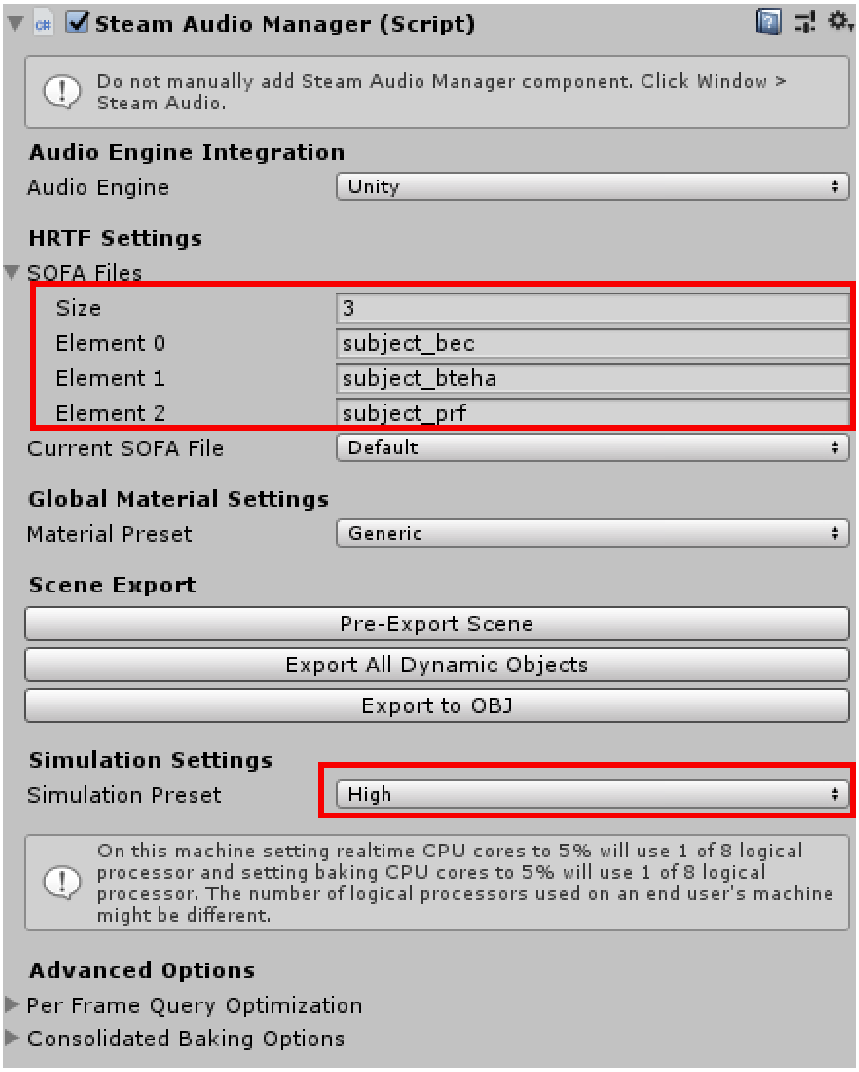


Figure C.9: Settings for the Steam Audio Manager.

It is crucial to define the SOFA filenames correctly. In order to change the set of HRTFs used with each subject, it is necessary to bring the files to the relative path *../listening_test_Data/StreamingAssets* in the build folder and rename them as corresponds.

Audio Settings

Located inside Edit->Project Settings, the Audio settings have to be configured as follows, with special attention to the ambisonics and spatializer plugins:

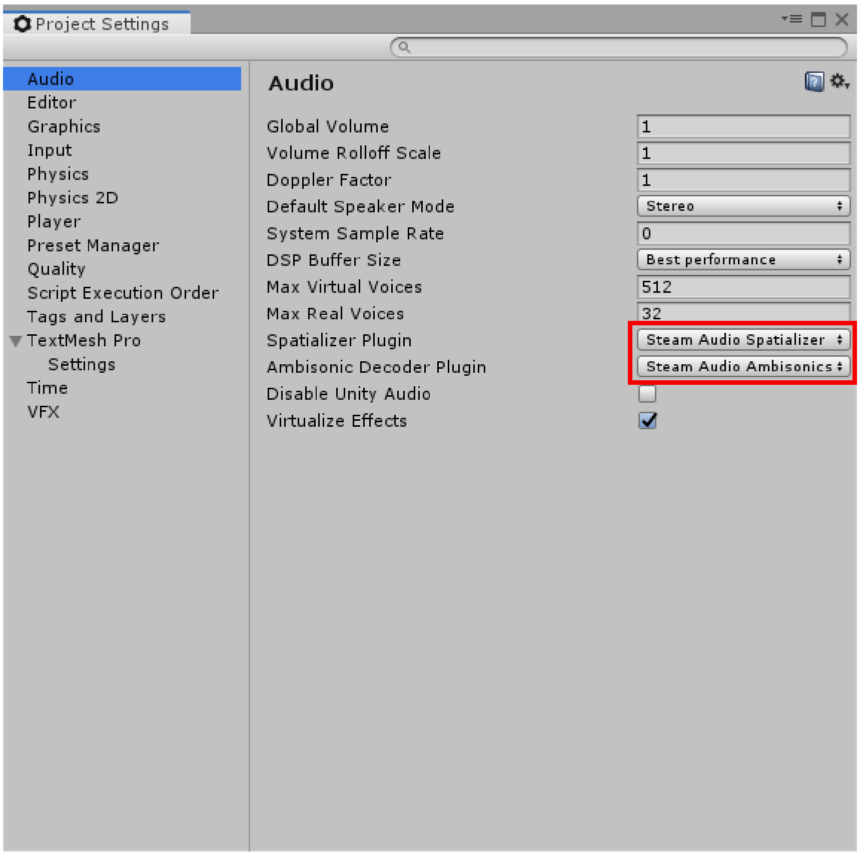


Figure C.10: Audio settings for the project.

Build Settings

Build settings determine the way in which the scenes are ordered when building a version of the software:

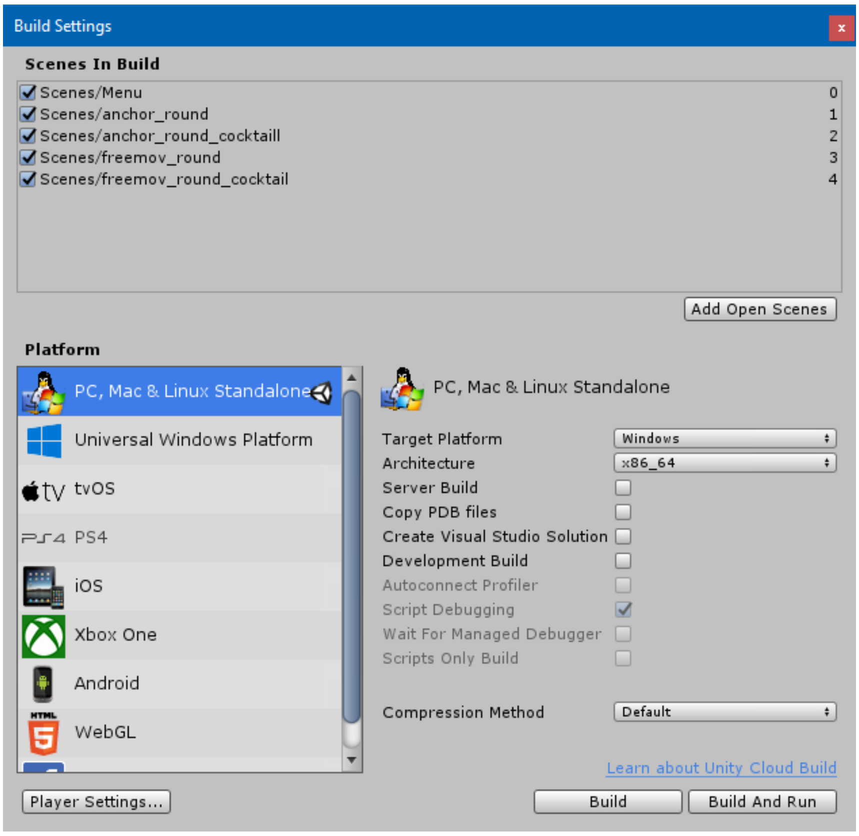


Figure C.11: Build settings for the project.

Results storage

The results obtained during the test are stored in a .csv file located in the relative route *../listening_test_Data* in the build folder, under the name *res_*subjectID*.csv*. A new file is generated for each subject and contains the results for each stage and round in format *setOfHrtf;real_source;selected_source;response_time*.

Appendix D

Hearing test questionnaire

Questionnaire for hearing tests (ISO 389-9:2009)

1.	**Name:**	**Date of birth:**	**Gender:**
2.	**Have you ever had trouble with your hearing (for example, infections, ear noises, drainage etc.)?**		
	☐ Yes ☐ No	If yes, please detail:	
3.	**Have you ever had an operation in your ear?**		
	☐ Yes ☐ No	If yes, please detail:	
4.	**Have your ever taken drugs, tablets or been given injections affected your hearing?**		
	☐ Yes ☐ No	If yes, please detail:	
5.	**Have you worked for several years in a place that was very noisy noisy, i.e. where it was difficult to communicate?**		
	☐ Yes ☐ No	If yes, please detail:	
6.	**Did you wear any hearing protector at that time?**		
	☐ Yes ☐ No	If yes, please detail:	
7.	**Do you attend pop/rock concerts or discotheques?**		
	☐ Never ☐ Once a year	☐ More than once a year	
8.	**Do you play any musical instrument?**		
	☐ Yes ☐ No	If yes, please specify:	
9.	**Do you listen to personal wearable players?**		
	☐ Never ☐ Less than 2 hours per week	☐ More than 2 hours per week	
10.	**Have you been exposed to any loud sounds from, e.g. motorbikes, chain-saws, gunfire, fire-crackers or explosions?**		
	☐ Yes ☐ No	If yes, what kind and how often:	
11.	**Does/did anyone in your immediate family have a hearing disorder?**		
	☐ Yes ☐ No	If yes, please specify:	
12.	**Have you ever had a hearing test before?**		
	☐ Yes ☐ No	If yes, when and where:	
I agree to the storage of my data and their use in connection with the threshold measurements			
Date:		Signature:	

Appendix E

Individual results

Table E.1: Results for subject134420. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	134420	32	22 (68.75%)	11 (34.38%)	5 (15.63%)	6 (18.75%)
1	BTEHA	134420	32	29 (90.63%)	14 (43.75%)	8 (25.00%)	7 (21.88%)
1	PRF	134420	32	23 (71.88%)	6 (18.75%)	8 (25.00%)	9 (28.13%)
1	BEC	134420	32	23 (71.88%)	14 (43.75%)	4 (12.50%)	5 (15.63%)
1	BTEHA	134420	32	25 (78.13%)	11 (34.38%)	6 (18.75%)	8 (25.00%)
1	PRF	134420	32	27 (84.38%)	16 (50.00%)	7 (21.88%)	4 (12.50%)
1	BEC	134420	32	2 (6.25%)	0 (0.00%)	0 (0.00%)	2 (6.25%)
1	BTEHA	134420	32	2 (6.25%)	1 (3.13%)	0 (0.00%)	1 (3.13%)
1	PRF	134420	32	3 (9.38%)	0 (0.00%)	0 (0.00%)	3 (9.38%)
1	BEC	134420	32	16 (50.00%)	0 (0.00%)	0 (0.00%)	16 (50.00%)
1	BTEHA	134420	32	10 (31.25%)	0 (0.00%)	0 (0.00%)	10 (31.25%)
1	PRF	134420	32	12 (37.50%)	0 (0.00%)	0 (0.00%)	12 (37.50%)

Table E.2: Results for subject136624. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	136624	32	17 (53.13%)	4 (12.50%)	7 (21.88%)	6 (18.75%)
1	BTEHA	136624	32	30 (93.75%)	10 (31.25%)	9 (28.13%)	11 (34.38%)
1	PRF	136624	32	20 (62.50%)	6 (18.75%)	6 (18.75%)	8 (25.00%)
1	BEC	136624	32	24 (75.00%)	16 (50.00%)	3 (9.38%)	5 (15.63%)
1	BTEHA	136624	32	24 (75.00%)	12 (37.50%)	7 (21.88%)	5 (15.63%)
1	PRF	136624	32	23 (71.88%)	11 (34.38%)	6 (18.75%)	6 (18.75%)
1	BEC	136624	32	7 (21.88%)	0 (0.00%)	0 (0.00%)	7 (21.88%)
1	BTEHA	136624	32	16 (50.00%)	0 (0.00%)	0 (0.00%)	16 (50.00%)
1	PRF	136624	32	4 (12.50%)	1 (3.13%)	0 (0.00%)	3 (9.38%)
1	BEC	136624	32	14 (43.75%)	0 (0.00%)	0 (0.00%)	14 (43.75%)
1	BTEHA	136624	31	13 (41.94%)	0 (0.00%)	0 (0.00%)	13 (41.94%)
1	PRF	136624	32	11 (34.38%)	0 (0.00%)	0 (0.00%)	11 (34.38%)

Table E.3: Results for subject247433. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	247433	32	25 (78.13%)	8 (25.00%)	9 (28.13%)	8 (25.00%)
1	BTEHA	247433	32	23 (71.88%)	11 (34.38%)	8 (25.00%)	4 (12.50%)
1	PRF	247433	32	25 (78.13%)	11 (34.38%)	8 (25.00%)	6 (18.75%)
1	BEC	247433	32	28 (87.50%)	15 (46.88%)	9 (28.13%)	4 (12.50%)
1	BTEHA	247433	32	27 (84.38%)	15 (46.88%)	9 (28.13%)	3 (9.38%)
1	PRF	247433	32	29 (90.63%)	17 (53.13%)	5 (15.63%)	7 (21.88%)
1	BEC	247433	32	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
1	BTEHA	247433	32	10 (31.25%)	0 (0.00%)	0 (0.00%)	10 (31.25%)
1	PRF	247433	32	1 (3.13%)	0 (0.00%)	0 (0.00%)	1 (3.13%)
1	BEC	247433	32	10 (31.25%)	0 (0.00%)	0 (0.00%)	10 (31.25%)
1	BTEHA	247433	32	11 (34.38%)	0 (0.00%)	0 (0.00%)	11 (34.38%)
1	PRF	247433	32	12 (37.50%)	1 (3.13%)	0 (0.00%)	11 (34.38%)

Table E.4: Results for subject363189. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	363189	32	11 (34.38%)	4 (12.50%)	2 (6.25%)	5 (15.63%)
1	BTEHA	363189	32	26 (81.25%)	8 (25.00%)	8 (25.00%)	10 (31.25%)
1	PRF	363189	32	20 (62.50%)	5 (15.63%)	5 (15.63%)	10 (31.25%)
1	BEC	363189	32	12 (37.50%)	7 (21.88%)	1 (3.13%)	4 (12.50%)
1	BTEHA	363189	32	20 (62.50%)	4 (12.50%)	6 (18.75%)	10 (31.25%)
1	PRF	363189	32	23 (71.88%)	7 (21.88%)	7 (21.88%)	9 (28.13%)
1	BEC	363189	32	4 (12.50%)	0 (0.00%)	0 (0.00%)	4 (12.50%)
1	BTEHA	363189	32	11 (34.38%)	0 (0.00%)	0 (0.00%)	11 (34.38%)
1	PRF	363189	32	4 (12.50%)	0 (0.00%)	0 (0.00%)	4 (12.50%)
1	BEC	363189	32	13 (40.63%)	0 (0.00%)	0 (0.00%)	13 (40.63%)
1	BTEHA	363189	32	16 (50.00%)	0 (0.00%)	0 (0.00%)	16 (50.00%)
1	PRF	363189	31	17 (54.84%)	0 (0.00%)	0 (0.00%)	17 (54.84%)

Table E.5: Results for subject477865. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	477865	32	19 (59.38%)	8 (25.00%)	5 (15.63%)	6 (18.75%)
1	BTEHA	477865	32	28 (87.50%)	13 (40.63%)	6 (18.75%)	9 (28.13%)
1	PRF	477865	32	23 (71.88%)	13 (40.63%)	5 (15.63%)	5 (15.63%)
1	BEC	477865	32	24 (75.00%)	19 (59.38%)	1 (3.13%)	4 (12.50%)
1	BTEHA	477865	32	29 (90.63%)	13 (40.63%)	8 (25.00%)	8 (25.00%)
1	PRF	477865	32	25 (78.13%)	14 (43.75%)	3 (9.38%)	8 (25.00%)
1	BEC	477865	32	15 (46.88%)	0 (0.00%)	0 (0.00%)	15 (46.88%)
1	BTEHA	477865	32	16 (50.00%)	2 (6.25%)	0 (0.00%)	14 (43.75%)
1	PRF	477865	32	7 (21.88%)	0 (0.00%)	0 (0.00%)	7 (21.88%)
1	BEC	477865	32	12 (37.50%)	0 (0.00%)	0 (0.00%)	12 (37.50%)
1	BTEHA	477865	31	19 (61.29%)	1 (3.23%)	0 (0.00%)	18 (58.06%)
1	PRF	477865	32	11 (34.38%)	0 (0.00%)	0 (0.00%)	11 (34.38%)

Table E.6: Results for subject485756. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	485756	32	22 (68.75%)	11 (34.38%)	4 (12.50%)	7 (21.88%)
1	BTEHA	485756	32	24 (75.00%)	10 (31.25%)	4 (12.50%)	10 (31.25%)
1	PRF	485756	32	21 (65.63%)	6 (18.75%)	6 (18.75%)	9 (28.13%)
1	BEC	485756	32	24 (75.00%)	12 (37.50%)	6 (18.75%)	6 (18.75%)
1	BTEHA	485756	32	28 (87.50%)	15 (46.88%)	5 (15.63%)	8 (25.00%)
1	PRF	485756	32	30 (93.75%)	14 (43.75%)	7 (21.88%)	9 (28.13%)
1	BEC	485756	32	2 (6.25%)	0 (0.00%)	0 (0.00%)	2 (6.25%)
1	BTEHA	485756	32	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
1	PRF	485756	32	1 (3.13%)	0 (0.00%)	0 (0.00%)	1 (3.13%)
1	BEC	485756	32	16 (50.00%)	1 (3.13%)	0 (0.00%)	15 (46.88%)
1	BTEHA	485756	32	12 (37.50%)	1 (3.13%)	0 (0.00%)	11 (34.38%)
1	PRF	485756	32	19 (59.38%)	2 (6.25%)	0 (0.00%)	17 (53.13%)

Table E.7: Results for subject695095. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	695095	32	20 (62.50%)	11 (34.38%)	6 (18.75%)	3 (9.38%)
1	BTEHA	695095	32	30 (93.75%)	17 (53.13%)	7 (21.88%)	6 (18.75%)
1	PRF	695095	32	26 (81.25%)	15 (46.88%)	6 (18.75%)	5 (15.63%)
1	BEC	695095	32	22 (68.75%)	13 (40.63%)	7 (21.88%)	2 (6.25%)
1	BTEHA	695095	32	30 (93.75%)	17 (53.13%)	9 (28.13%)	4 (12.50%)
1	PRF	695095	32	29 (90.63%)	20 (62.50%)	8 (25.00%)	1 (3.13%)
1	BEC	695095	32	17 (53.13%)	0 (0.00%)	0 (0.00%)	17 (53.13%)
1	BTEHA	695095	32	1 (3.13%)	0 (0.00%)	0 (0.00%)	1 (3.13%)
1	PRF	695095	32	7 (21.88%)	0 (0.00%)	0 (0.00%)	7 (21.88%)
1	BEC	695095	32	11 (34.38%)	0 (0.00%)	0 (0.00%)	11 (34.38%)
1	BTEHA	695095	31	9 (29.03%)	0 (0.00%)	0 (0.00%)	9 (29.03%)
1	PRF	695095	32	10 (31.25%)	0 (0.00%)	0 (0.00%)	10 (31.25%)

Table E.8: Results for subject730012. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	730012	32	22 (68.75%)	5 (15.63%)	8 (25.00%)	9 (28.13%)
1	BTEHA	730012	32	25 (78.13%)	6 (18.75%)	10 (31.25%)	9 (28.13%)
1	PRF	730012	32	17 (53.13%)	6 (18.75%)	7 (21.88%)	4 (12.50%)
1	BEC	730012	32	22 (68.75%)	8 (25.00%)	6 (18.75%)	8 (25.00%)
1	BTEHA	730012	32	24 (75.00%)	4 (12.50%)	9 (28.13%)	11 (34.38%)
1	PRF	730012	32	20 (62.50%)	3 (9.38%)	10 (31.25%)	7 (21.88%)
1	BEC	730012	32	7 (21.88%)	0 (0.00%)	0 (0.00%)	7 (21.88%)
1	BTEHA	730012	32	16 (50.00%)	0 (0.00%)	0 (0.00%)	16 (50.00%)
1	PRF	730012	32	7 (21.88%)	0 (0.00%)	0 (0.00%)	7 (21.88%)
1	BEC	730012	32	12 (37.50%)	0 (0.00%)	0 (0.00%)	12 (37.50%)
1	BTEHA	730012	32	13 (40.63%)	0 (0.00%)	0 (0.00%)	13 (40.63%)
1	PRF	730012	31	13 (41.94%)	0 (0.00%)	0 (0.00%)	13 (41.94%)

Table E.9: Results for subject757555. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	757555	32	24 (75.00%)	16 (50.00%)	3 (9.38%)	5 (15.63%)
1	BTEHA	757555	32	26 (81.25%)	13 (40.63%)	9 (28.13%)	4 (12.50%)
1	PRF	757555	32	25 (78.13%)	14 (43.75%)	6 (18.75%)	5 (15.63%)
1	BEC	757555	32	28 (87.50%)	17 (53.13%)	7 (21.88%)	4 (12.50%)
1	BTEHA	757555	32	30 (93.75%)	16 (50.00%)	10 (31.25%)	4 (12.50%)
1	PRF	757555	32	25 (78.13%)	19 (59.38%)	3 (9.38%)	3 (9.38%)
1	BEC	757555	32	3 (9.38%)	0 (0.00%)	0 (0.00%)	3 (9.38%)
1	BTEHA	757555	32	10 (31.25%)	0 (0.00%)	0 (0.00%)	10 (31.25%)
1	PRF	757555	32	4 (12.50%)	0 (0.00%)	0 (0.00%)	4 (12.50%)
1	BEC	757555	32	16 (50.00%)	1 (3.13%)	0 (0.00%)	15 (46.88%)
1	BTEHA	757555	32	13 (40.63%)	1 (3.13%)	0 (0.00%)	12 (37.50%)
1	PRF	757555	32	15 (46.88%)	1 (3.13%)	0 (0.00%)	14 (43.75%)

Table E.10: Results for subject828039. The percentages are calculated with respect to the total number of trials.

Stage	Round	Subject	Iterations	Total Errors	Out of cone Errors	Within cone Errors	Median plane Errors
1	BEC	828039	32	19 (59.38%)	8 (25.00%)	5 (15.63%)	6 (18.75%)
1	BTEHA	828039	32	25 (78.13%)	9 (28.13%)	7 (21.88%)	9 (28.13%)
1	PRF	828039	32	16 (50.00%)	9 (28.13%)	4 (12.50%)	3 (9.38%)
1	BEC	828039	32	18 (56.25%)	6 (18.75%)	6 (18.75%)	6 (18.75%)
1	BTEHA	828039	32	27 (84.38%)	12 (37.50%)	9 (28.13%)	6 (18.75%)
1	PRF	828039	32	17 (53.13%)	7 (21.88%)	4 (12.50%)	6 (18.75%)
1	BEC	828039	32	6 (18.75%)	0 (0.00%)	0 (0.00%)	6 (18.75%)
1	BTEHA	828039	32	13 (40.63%)	1 (3.13%)	0 (0.00%)	12 (37.50%)
1	PRF	828039	32	6 (18.75%)	0 (0.00%)	0 (0.00%)	6 (18.75%)
1	BEC	828039	32	8 (25.00%)	0 (0.00%)	0 (0.00%)	8 (25.00%)
1	BTEHA	828039	32	13 (40.63%)	0 (0.00%)	0 (0.00%)	13 (40.63%)
1	PRF	828039	32	10 (31.25%)	0 (0.00%)	0 (0.00%)	10 (31.25%)

Bibliography

AAU, 2019a. *Large Anechoic Room AAU*, Aalborg University, Department of Electronic Systems. [Website]. Available at: <http://doc.es.aau.dk/labs/acoustics/facilities/anechoic_room_large/>. [Accessed 29 April 2019].

AAU, 2019b. *Audiometry cabins AAU*, Aalborg University, Department of Electronic Systems. [Website]. Available at: <http://doc.es.aau.dk/labs/acoustics/facilities/audiometri_cabins/>. [Accessed 23 May 2019].

AAU, 2019c. *Vifa Ball Speaker*, Aalborg University, Department of Electronic Systems. [Website]. Available at: <http://doc.es.aau.dk/labs/acoustics/equipment/vifa_ball/>. [Accessed 29 April 2019].

Batteau, D., 1967. The Role of the Pinna in Human Localization. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 09, 168, pp. 158–80. doi: 10.1098/rspb.1967.0058.

Blauert, J. *Communication Acoustics*. Springer-Verlag Berlin Heidelberg, 1st edition, 2005. ISBN 978-3-540-27437-7.

Brungart, D. S. and Rabinowitz, W. M., 1999. Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America*, 106 (3).

Christensen, F., Lydolf, M., and Sørensen, M., 1997. Human localization of sound signals with reduced bandwidth. *The Journal of the Acoustical Society of America*, 102 (5,Pt.2: Abstract). ISSN 0001-4966. 134th Meeting of Acoustical Society of America, December 1-5, 1997, San Diego, California, USA 134th Meeting of Acoustical Society of America, December 1-5, 1997, San Diego, California, USA pp. 3140 (paper 3aSP7). - Only abstract available.

Gardner, M. B. and Gardner, R. S., 1973. Problem of localization in the median plane: effect of pinnae cavity occlusion. *The Journal of the Acoustical Society of America*, 53(2), pp. 400–408. doi: 10.1121/1.1913336. Available at: <https://doi.org/10.1121/1.1913336>.

Han, H. L., 1994. Measuring a Dummy Head in Search of Pinna Cues. *J. Audio Eng. Soc*, 42(1/2), pp. 15–37. Available at: <http://www.aes.org/e-lib/browse.cfm?elib=6964>.

Hirsh, I. J., 1971. Masking of Speech and Auditory Localization. *Audiology*, 10 (2), pp. 110–114. doi: 10.3109/00206097109072548. Available at: <https://www.tandfonline.com/doi/abs/10.3109/00206097109072548>.

Institute, A. R. *SOFA conventions wiki*, Austrian Academy of Science. [Website]. Available at: <https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics)>. [Accessed 19 May 2019].

ISO 389-9:2009. *Acoustics - Reference zero for the calibration of audiometric equipment - Part 9: Preferred test conditions for the determination of reference hearing threshold levels.*

Jenny, C., Majdak, P., and Reuter, C., 05 2018. *SOFA Native Spatializer Plugin for Unity - Exchangeable HRTFs in Virtual Reality.*

LAKULISH, 2017. *Blog: Bilinear HRTF Interpolation in Steam Audio*, LAKULISH. [Website]. Available at: <https://steamcommunity.com/games/596420/announcements/detail/507059261126291719?>. [Accessed 27 April 2019].

Majdak, P. and Noisternig, M., 2016. *SOFA API for Matlab and Octave.* [Website]. Available at: <https://sourceforge.net/projects/sofacoustics/>. [Accessed 19 May 2019].

Majdak, P., Goupell, M. J., and Laback, B., 2010. 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics*, Feb, 72(2), pp. 454–469. ISSN 1943-393X. doi: 10.3758/APP.72.2.454. Available at: <https://doi.org/10.3758/APP.72.2.454>.

Minnaar, P., Plogsties, J., and Christensen, F., 2005. Directional resolution of head-related transfer functions required in binaural synthesis. *AES: Journal of the Audio Engineering Society*, 10, 53, pp. 919–929.

Møller, H., 1992. Fundamentals of binaural technology. *Applied Acoustics*, 12, 36, pp. 171–218. doi: 10.1016/0003-682X(92)90046-U.

Møller, S., 2018. Pinna reconstruction filter for behind the ear hearing aids. Speciale, Aalborg University, Institute of Electronic Systems. Available at: <https://projekter.aau.dk/projekter/en/studentthesis/pinna-reconstruction-filter-for-behind-the-ear-hearing-aids(8cea1d9d-6632-4113-bc11-25709a7a70ed).html>.

Møller, H., Sørensen, M., Jensen, C., and Hammershøi, D., 1996. Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, 44(6), pp. 451–469. ISSN 1549-4950.

Moore, B. C. *An introduction to the Psychology of Hearing.* Emerald Group Publishing Limited, 6th edition, 2012. ISBN 978-1-78052-038-4.

Mueller, M. F., Kegel, A., Schimmel, S. M., Dillier, N., and Hofbauer, M., 2012. Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes. *The Journal of the Acoustical Society of America*, 131(6), pp. 4732–4742. doi: 10.1121/1.4705292. Available at: <https://doi.org/10.1121/1.4705292>.

Sandvad, J. Dynamic Aspects of Auditory Virtual Environments. In *Audio Engineering Society Convention 100*, May 1996. Available at: <http://www.aes.org/e-lib/browse.cfm?elib=7547>.

Sonion, 2015. *Datasheet microphone 8002*, Sonion. [Website]. Available at: <https://www.sonion.com/wp-content/uploads/8002-2003357.pdf>. [Accessed 9 May 2019].

Udesen, J., Piechowiak, T., Gran, F., and Dittberner, A., 2013. Degradation of spatial sound by the hearing aid. *Proceedings of the International Symposium on Auditory and Audiological Research*, Dec., 4, pp. 271–278. Available at: <https://proceedings.isaar.eu/index.php/isaarproc/article/view/2013-29>.

Valve, 2017. *Steam Audio Unity Plugin 2.0-beta.17*, Valve. [Website]. Available at: <https://valvesoftware.github.io/steam-audio/doc/phonon_unity.html>. [Accessed 19 May 2019].

Van den Bogaert, T., J Klasen, T., Moonen, M., Van Deun, L., and Wouters, J., 2006. Horizontal localization with bilateral hearing aids: Without is better than with. *The Journal of the Acoustical Society of America*, 02, 119, pp. 515–26. doi: 10.1121/1.2139653.

Xie, B. *Head-Related Transfer Function and Virtual Auditory Display.* J.Ross Publishing, 2nd edition, 2013. ISBN 978-1-60427-070-9.