Fast Individual HRTF Acquisition with Unconstrained Head Movements for 3D Audio

Javier Molina García

Sound and Music Computing, 2019-06

Master's Project

Here you can write something about which tools and software you have used for typesetting the document, running simulations and creating figures. If you do not know what to write, either leave this page blank or have a look at the colophon in some of your books.

Electronics and IT
Aalborg University
http://www.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Fast Individual HRTF Acquisition with Unconstrained Head Movements for 3D Audio

Theme:
Scientific Theme

Project Period:
Spring Semester 2019

Project Group:

Participant(s):
Javier Molina García

Supervisor(s):
Michele Geronazzo
Stefania Serafin

Copies: 1

Page Numbers: 52

Date of Completion:
May 28, 2019

Abstract:

Head-related transfer function (HRTF) is essential to provide a spatialized listening experience over headphones. HRTF-based techniques have been extensively used for 3D audio. A number of studies defend the necessity of individual HRTFs to create convincing immersive audio experience. This project implements a method based on adaptive filtering that allows users to acquire personalised HRTFs based on binaural audio recordings and head tracking information. Simulations were performed to evaluate the influence of the signal to noise ratio, the type of excitation signal, the kind of head movements and the time of the acquisition time on the system's accuracy. The results validate the accuracy of the proposed system, being able to achieve adequate precision even in very low (under 30dB) signal to noise ratio scenarios.

Contents

Preface **vii**

1 Introduction **1**
1.0.1 Head Related Transfer Function (HRTF) 1
1.0.2 Virtual Auditory Displays and Virtual Auditory Environments 3
1.0.3 Personalised HRTF acquisition 4
1.1 Problem Statement 4

2 Related work **7**
2.1 HRTF acquisition based on antropomorphic measurements 7
2.2 HRTF acquisition based on statistical and mathematical methods . . 8
2.3 HRTF acquisition based on spherical harmonics 8
2.4 HRTF acquisition based on adaptive filtering 9
2.5 Excitation signals for HRTF acquisition 10

3 Implementation **11**
3.1 Theoretical Background 11
3.1.1 Adaptive filtering 13
3.1.2 Excitation signal 15
3.2 Software Design 16
3.2.1 Head tracking and sound capturing 16
3.2.2 HRTF/HRIR calculation 16

4 Evaluation **19**
4.1 Simulation scenarios 22
4.2 Simulation results 23
4.2.1 Signal to noise ratio 23
4.2.2 Excitation signal 27
4.2.3 Visitation time 29
4.2.4 Head movements 31

5 Discussion **35**
5.0.1 Excitation signal 35
5.0.2 Head movements 36
5.0.3 Signal to noise ratio 36
5.0.4 Visitation time 36

6 Conclusion **39**

7 Appendix A - Simulation results **43**

Bibliography **47**

Preface

The concept for this research has its origin in my interest for music, spatial auditory displays and new technologies. Since my first experiences with 3D audio as a kid at Hemisferic in Valencia (Spain) to most recent events at sound installations, spatial audio has always been fascinating to me.

Having the opportunity to reproduce these experiences in personal devices, was an important motivation. Pushing the musical perception one step beyond, passing from stereo to a whole three-dimensional experience, is the future of music production. Moreover, with the appearance of new affordable virtual and augmented reality headsets, a whole new world of possibilities opens up.

The applications of 3D audio in a near future are virtually endless, in fields that go from hearing aids to purely artistic purposes. So this project sums up what I was looking for when I started starting my master degree at Sound and Music Computing: having the chance to work with cutting edge technology, while combining it with my passion for music, auditory experiences and art.

Aalborg University, May 28, 2019

Javier Molina García
<jmolin17@student.aau.dk>

Chapter 1

Introduction

Hearing is one of the most empowering of the senses. It is not only the way the pitch, loudness and timbre of the sounds are perceived, but also plays a crucial role in survival [50]. The auditory system is capable of perceiving the distance and direction of a sound source, being a useful resource to warn humans and animals from potential dangers. Human ability to spatially locate sound sources depends on many factors (source's acoustical properties, sound source direction, etc) and varies across individuals [68].

The major mechanisms responsible for the human capability of detecting the direction of the sound have been extensively studied over the years [20][6]. Psychoacoustics research has described most of the main directional localization cues, including Interaural Level Difference (ILD), Interaural Time Difference (ITD)[6].

ITD describes the time difference between the arrival of the sound waves at the left and right ear canals. As soon as the sound source comes from a point that is not placed in the median plane, the distance to each ear will be different, causing a small delay in the arrival time in one of the ears. ILD, on the other hand, describes the attenuation of the sound pressure at the farther ear caused by the shadowing effect of the head [68]. Both descriptors are decisive in sound localization in the horizontal plane.

However ITD and ILD by itself do not explain the source localization system in its entirety. There are other relevant acoustic cues caused by the scattering of the sound by the head, pinna and torso. This filtering caused by the combined effects of the different scatterers can be described by a complex frequency response function called the Head Related Transfer Function (HRTF) [12].

1.0.1 Head Related Transfer Function (HRTF)

A head related transfer function (HRTF) is a function that characterises the sound transmission from a free field to a point in the ear canal of a human subject, for a certain angle of incidence [45] . HRTFs contain temporal, spectral and spatial

Figure 1.1: KEMAR Artificial Dummy Head

information, used by the listener to locate the source of sound.

HRTF measurement is a complex procedure. This task is usually performed in anechoic chambers to avoid undesirable room reflections. There are many different ways to proceed, most of them based on stop-and-go techniques, from running the excitation signal and then changing the position of the speaker or subject [17], to using continuously moving loudspeakers [53] around the measuring subject. These techniques are highly time-consuming and laborious.

Artificial dummy heads have been extensively used for HRTF acquisition. A dummy head is model used to recreate the anatomy of a real human being, stressing in the head, torso and pinna, which are essential in the sound propagation. One of the most widely used dummy head is KEMAR, which was used to create one of the first free HRTF databases available on the internet [17] that has been used in numerous significant works.

The directional cues contained in an HRTF are dependant on the reflection and diffraction of the sound wave at head, torso and pinna, which makes HRTF

highly individual, with a wide variation from person to person due to anatomical differences.

Binaural sound signals recorded at the ear canal of an artificial head or human subject can capture the spatial information of a sound wave [44]. In the same way, the reproduction of these recorded signals at the ear drum can replicate the spatial auditory event, including the environmental acoustic behaviour and the sound source localisation [68]. However there are other methods to obtain binaural signals. Binaural signals can be synthesised using HRTFs [23], replicating the spatial auditory perception of sound events.

When hearing a binaural signal synthesised using a non individualised HRTF, localisation is degraded. This degradation is especially important for the elevation, as well as the front-back perception as these deviations are actively related to the personal geometry of the pinna [52][67]. The use of individual HRTFs entails the perception of the sound source in a more compact, well-defined way, making its localisation more precise [63]. Some studies even suggest that the experience of an accurate and authentic spatial auditory display necessarily imply the use of individual HRTFs [46] [69].

1.0.2 Virtual Auditory Displays and Virtual Auditory Environments

A virtual auditory display (VAD) can be defined as a system for generating sounds with spatial positional characteristics and conveying them to a listener [30]. In complex virtual auditory displays, several sonic events might happen simultaneously. If this sounds are synthesised as if their sources were located in different points in space, the listener can distinguish them in an easier way [9] . The performance in different applications has been demonstrated to be improved by the use of three dimensional virtual auditory displays, from navigation systems [29] to fields as aviation [65].

The spatial sound perception has been underutilised for blind people, being 3D auditory displays a great opportunity to create applications [69] that could be useful for navigation in non-visual environments [38]. With virtual and augmented reality headsets being more popular and affordable, specially after the appearance of headsets as Oculus Rift and Magic Leap, and its use in fields such as communication, military training and entertainment, 3D sound technology will be increasingly researched, used and developed [69] .

Virtual auditory displays and virtual auditory environments (VAE) have become an integral part of user experience in Virtual and Augmented Reality (VR/AR). HRTF plays a crucial role in the rendering of VAD over headphones, so it has been extensively used in interactive VR and AR applications [24] [54]. Spatial audio has not only been proved to increase the sense of presence in virtual environments [34] [37], but some studies suggests that encourages a more exploratory and playful response to this kind of environments [64].

Figure 1.2: Magic Leap headset

1.0.3 Personalised HRTF acquisition

Fast personalised HRTF acquisition methods using head tracking and binaural recordings [22] [57] have been researched in the recent years. This methods give the subject the freedom to move the head freely in order to get the measurements from all the different angles. This is a key improvement for various reasons. It makes it faster and more flexible to obtain the binaural recordings at the ear canal, as the measurements are continuous, avoiding the previously mentioned stop-and-go methods. And, moreover, it makes the whole process more comfortable for the subject that takes part in the measurements, as his head is not fixed during the whole procedure.

The head tracking is key to obtain these measurements accurately, as, when it comes to make this kind of procedures, even small variations in angle can run into major inaccuracies [28]. Simple head tracking devices and more advanced headsets as Oculus Rift [15] have been successfully used to obtain these measurements.

In order to extract the continuous HRTF measurements, Normalized Least Means Squares (NLMS) based approaches have been researched in a number of articles [26] [55] [27]. These methods have been evaluated, mostly through simulations, proving their robustness and accuracy. These results prove this technique's suitability to be used in procedures that allow unconstrained head movement in azimuth and elevation.

1.1 Problem Statement

From the previous research, it is reasonable to extract that virtual auditory displays, and, therefore, 3D audio based on HRTFs is a field in constant evolution that is, and is going to be, widely used in the following years.

Virtual reality and augmented reality (VR/AR) headsets are growing markets that tend to use cutting-edge technologies in order to create more advanced mul-

Figure 1.3: Sennheiser AMBEO headset

timodal experiences. Virtual auditory displays play a main role when it comes to enhance the sense of presence in a virtual environment [34] [37]. Therefore, HRTFs are crucial in the synthesis of real-time virtual auditory environments for VR/AR headsets [68].

The spectral information and directional cues that HTRF contain are highly individual, as they depend on the subject's head, torso and pinna influence on the sound propagation. Specially the perception of front and back, and elevation, is degraded, as they are highly dependant on the anatomy of the subject's pinna [52]. Therefore, the use of individual HRTFs enhances significantly the user's experience [46] [69], making it necessary to find faster and more user-friendly ways to obtain these measurements.

This project aims to find an affordable, portable and easy way to obtain personalised HRTFs. Based on recent articles [22] [57] on HRTF acquisition with unconstrained head movements, this project aims to find a robust algorithm that an obtain these personalised measurements on the fly. This projects seeks to prove that just by using a mobile device, commercially available and affordable headsets (as the Sennheiser's AMBEO headset and headtracker), it is possible to obtain individualised HRTFs.

In order to achieve this goal, the HRTF acquisition algorithm is going to be tested with simulations to evaluate its performance under different conditions. This evaluation tries to demonstrate that this algorithm would be suitable to be

used under non-ideal situations, as it aims to be used as a mobile applications operated by regular users, in conditions far from being ideal. Several different signal to noise ratios, head movements and excitation signals would be tested in order to find out the algorithm's performance in these scenarios.

This leads to the problem statement:

How can we implement a robust algorithm that can extract individualised HRTFs with unconstrained head movements? Would this algorithm be suitable to be used on mobile devices, by regular users, under non-ideal conditions?

Chapter 2

Related work

In this chapter, related studies on the acquisition of personalized Head Related Transfer Functions (HRTF), based in different approaches, are described. The related work regarding individualized HRTF acquisition has been divided in four groups depending the approach taken to obtain these measurements. Moreover, a section describing the research in excitation signals, in order to improve the efficiency and accuracy of these calculations, has been added as it is an important part of the field of interest.

2.1 HRTF acquisition based on antropomorphic measurements

HRTF personalization based on anthropometric measurements has been extensively used. Some of these methods are based on the physical modelling of the receivers body and ear canal, in order to reproduce accurately their body's influence in the sound propagation. Following this approach we can find different techniques. Lopez-Poveda and Meddis [39] describe how HRTFs can be obtained by using the physical modelling of the receiver's concha can be implemented with a diffraction - reflection model. This article focses in the prediction of elevation-dependent spectral features related to the transverse dimensions of the concha.

A similar approach was used by Brown and Duda [8], where the HRTF is obtained implementing a physical model of the receivers head. This implementation is based on Lord Rayleigh's [56] spherical head model, which recreates the head's influence on the sound propagation, combined with Kuhn's [36] study on the phase and group delay for a sphere. This approximation reduces the computational complexity notably. Moreover, a set of physical parameters are available to modify, in order to personalize the HRTF, depending on the user's physiognomy.

These studies on the receiver's physiognomy have also been combined with more advanced physical analysis techniques. For instance, *Tommasini et al* took a

very interesting approach in their 2016 article [63] where a physical model of the pinnae is developed based on the 3D scanning of the receiver's ear.

However, other interesting studies on the necessity of this personalization have been performed in paper's like the one by *Geronazzo et al* [18]. In this article, a different solution in order to personalize HRTFs is described. In this case, based on antrophomorfic measurements, they try to match each user's personal HRTF with the closest HRTF from a large database of non-individualized HRTFs. The results improve the performance significantly withrespect to dummy-head HRTFs and random HRTF selection.

2.2 HRTF acquisition based on statistical and mathematical methods

Mirbagheri and Atlas [42] designed a technique based in statistical methods, named Regression Factor Analysis, that provides a new approach on fast personalized HRTF calculation. This algorithm was tested with a simulation reproducing a sound source which is captured by microphones placed in both ear canals. This algorithm gives promising results, outperforming in time and efficiency most personalized HRTF calculation methods, and providing a full continuous HRTF field. Nonetheless, it has not been tested live.

Moreover, some techniques combining anthropometric measurements with other methods have been implemented. *Hugeng et al*[31] proposes how combining Principal Component Analysis (PCA) and multiple linear regression (MLR) on few anthropometric measurements, it is possible to obtain individualized HRTFs. This method provides individualized HRTFs in the horizontal plane, building on previous studies on the use of multiple regression analysis for individual HRTF acquisition[66], while improving its accuracy.

Other mathematical techniques have been used researching HRTFs individualization in order to get a more accurate experience. *Grindlay et al* [21] combine, anthropometric measurements, as in the previous method, with a multilinear extension of the conventional singular value decomposition (SVD), mapping anatomical data to different parameters for individualized HRTF calculations. This method proves to be able to produce sets of individualized HRTFs based on this anatomical data, outperforming in accuracy other basic Principle Component Analysis based methods [32] .

2.3 HRTF acquisition based on spherical harmonics

Spherical harmonics are a frequency-space basis for representing functions defined over the sphere. They are the spherical analogue of the Fourier series, and Spherical

Harmonic Transforms (SHTs) the equivalent of Fourier transforms on the sphere [5].

Spherical harmonics decomposition and expansion have been extensively used for HRTF acquisition. *Romigh et al* [60] [59] describe how to, using spherical harmonic descomposition, and based on a small a small set of spatial samples, a spatially continuous individualized HRTF can be represented. Individual and non-individual components of the HRTF can also be separated using this technique.

These methods based on spherical harmonics have been extensively studied, with promising results. *Pollow et al* [51] propose a method that, based on spherical harmonic decomposition, can calculate HRTFs on arbitrary points using extrapolation, using measurements from a single radius. On the other hand, *Aussal et al* [3] describe an interpolation method for HRTF measurements, with good accuracy results.

HRTF acquisition usually imply a very long of session of recordings from discrete points in space, so this kind of interpolation and extrapolation techniques could potentially reduce the amount of points needed to provide a countinous HRTF field. Most of the new studies on this field [4] focus on how to reduce the amount of measurement points needed in order to obtain an accurate continuous HRTF field, some of them reducing significantly the amount of sampling points required [61]. However, the computational power required to work with these techniques make it challenging to use these methods in real time applications.

2.4 HRTF acquisition based on adaptive filtering

Some of the most extensively used methods for personalized HRTF acquisition are based on adaptive filtering. From all the adaptive filtering techniques, Normalized Least Means Squares (NLMS) based algorithms have been extensively used, mostly. This kind of adaptive filters tend to mimic a desired filter, which in this case would be the HRIR, by finding the filter coefficients that relate to producing the least mean square of the error signal, which is the difference between the source signal and the signal received in the ear canal.

He, Ranjan et al have been publishing in the last years [26] [55] [27] a series of NLMS based methods for personalized HRTF acquisition. These techniques, in contrast to traditional HRTF acquisition methods, do not require the head of the subject to be in a fixed positions. The movement of the user's head is handled with an activation matrix, which allows the subject to perform unconstrained head movements, making the whole HRTF acquisition process easier and more user-friendly.

These methods have been tested both with real subjects, and more extensively, with simulations. These evaluations proved the robustness and accuracy, while confirming its suitability to use with unconstrained head movement in azimuth

and elevation. This results have encouraged other articles to find ways to combine these techniques with technologies as virtual and augmented reality headsets and head trackers [22] [15] to obtain personalized HRTFs in a fast and accurate way, without needing a lab facilities or expensive setups.

These methods will be explained in detail in section 3, as the implementation that this reports describes is heavily based in this research.

2.5 Excitation signals for HRTF acquisition

There has been extensive research on how to excite HRTF acquisition systems in order to get accurate results and save time in the measurement extraction process.

Gaussian white noise has been been widely used as excitation method for these systems [26] [55] [27] ; however, there are studies proving that other signals could outperform the results of this method.

Maximum length sequences (MLS) [58] have been extensively used in system identification. Moreover, they have been successfully in HRTF measurements [17] in lab facilities. However, these sequences are sensitive to non-linear distortions from the reproduction devices [11][40], which make them not suitable for certain environments.

Non-periodic frequency modulated sweeps [47] have been used for transfer function acquisition, outperforming other excitation signals as the previously mentioned maximum length sequences. Based on this results, *Majdak et al* [40] implemented the Multiple Exponential Sweep Method (MESM). This technique overlaps sweeps, reducing significantly the time required for HRTF acquisition compared with other techniques as Exponential Sweeps and Maximum Length Sequences, while improving its accuracy in noisy conditions.

Finally, *Antweiler et al* [2] proposed the Perfect Sweep as an optimal excitation signal for system identification, giving promising results in NLMS-based systems, and outperforming the previously mentioned signals. This method is described in detail in section 3.1.2, as it was relevant to this project's implementation.

Chapter 3

Implementation

3.1 Theoretical Background

In this section the implementation of this HRTF acquisition system is described. The goal of this implementation is to calculate individual HRTFs while providing a fast, robust and user friendly system to perform these measurements, avoiding the costly, time consuming and unpractical procedures and facilities usually required, as it was explained in section 1.0.1.

In this procedure that we propose, a head tracker is used to record the head position and orientation in every moment. This allows to explore the user's natural head movements of, while permitting to use a way more simple set up with just one speaker, placed in a fixed position, and microphones in both ear canals, as it is described in figure 3.1.

The use of a head tracker makes the process more friendly to the subject, while helping to avoid accuracy problems that could imply major precision errors in the measurements [28]. The user can move the head in a two dimensional space, in azimuth and elevation, freely, while the system captures, synchronously and simultaneously, the head movements and the audio signals received in both ears and the excitation signal. To facilitate this process, a mobile app will explain the user how to move the head to cover the whole space, in order to get fast, continuous, personalised HRTFs, while recording all the sounds and movements necessary to perform these calculations. To ensure the success and accuracy of these measurements, the head must keep a fixed distance to the sound source, as only changes in azimuth and elevation are taken into consideration.

The nature of this system highlights the need of finding an algorithm that can support this freedom of movements. The subject's head movements, in general terms, are going to be quasi-random, with different visitation times in each one of the positions, so regular procedures, as the ones mentioned in section 1.0.1, are not suitable. This report is going to focus in achieving an algorithm with the necessary

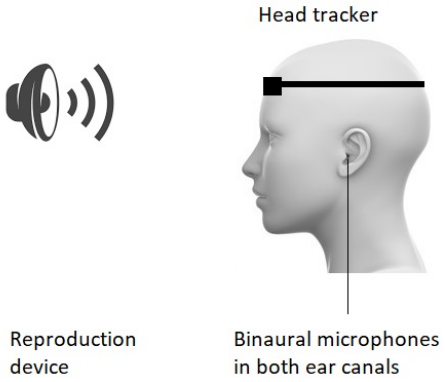

Figure 3.1: System overview

the robustness and accuracy to perform these measurements in such scenarios. In order to confirm these features, different signal to noise ratios, head movement patterns, visitation times on each point in the two dimensional space, and excitation signals will be tested in a batch of simulations that will be thoroughly explained in section 4.

The analysis of the head movements will focus on the pitch and yaw orientations, from where the azimuth and elevation, respectively, can be easily extracted. We will use Sennheiser's AMBEO head-tracker as a reference, as it is a good example of an affordable widely used head tracking device. Also it's easy compatibility with Sennheiser's AMBEO Headset, which facilitates the recording of binaural signals in the ear canal, motivates this decision. AMBEO head-tracker's update rate is 12 hertz, so we will stick to that sampling rate for this implementation, even though other sampling rates could be used for this algorithm. The update rate is a key element in the application, as HRTF calculation algorithm is dependant on this value; the recorded audio signals will be discretized in blocks matching this update rate.

To simplify the notation, the following sections will be focused in obtaining the Head Related Impulse Response (HRIR) in only one of the ear canals. To obtain the HRIR on the other ear canal, exactly the same procedure must be peformed. Considering this system as linear-time invariant, the recorded signal at each one of the ears of the subject can be described by the following equation

$$y(n) = h^T[d(n)]x(n) + e(n), \tag{3.1}$$

where y(n) is the recorded signal vector at the ear canal, e(n) is the measurement noise, h^T[d(n)] represents the HRIR vector at the current direction d(n), and x(n) is the excitation signal vector. As explained in the articles by *He, Ranjan and Woon-Seng* [26] [55] [27], based on the dynamically changing response of the sys-

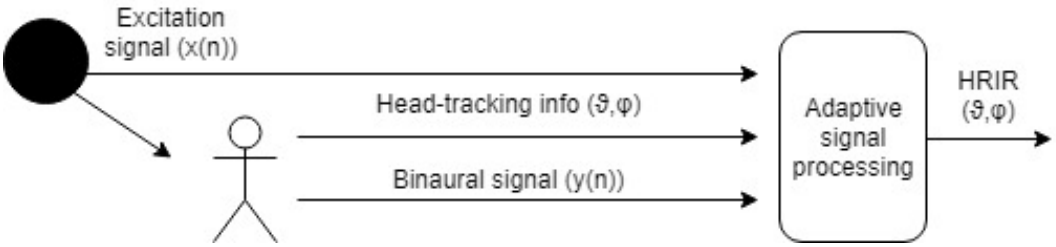

Figure 3.2: System overview

tem, the most suitable way to obtain the HRIRs is by using adaptive filtering methods.

3.1.1 Adaptive filtering

An adaptive filter is a system that aims to model the relationship between two signals in real time with using iterative methods. An adaptive filter can be described by four features: the signals processed by the filter, the structure that defines how the output signal is computed from its input signal, the parameters that can be iteratively changed to alter the its input-output relationship, and the adaptive algorithm that describes how the parameters are adjusted [10] .

According to this last aspect, we can find many different adaptive filtering algorithms. In this paper we are going to focus in Normalized Least Mean Squares (NLMS) adaptive filtering. NLMS adaptive filters have been widely used for different purposes, including echo cancellation [49], noise cancellation [43] and machine learning applications [33] . Some of the reasons of the popularity of the NLMS algorithms is its robustness and low computational cost [25].

The adaptive filtering updating behaviour, in a NLMS algorithm, is described by the following equation

$$\hat{h}(n+1) = \hat{h}(n) + \mu(n)\frac{e(n)}{||x(n)||_2^2}x(n) \tag{3.2}$$

where $\hat{h}$(n) is the time-varying adaptive filter at time n, that corresponds to the HRIR in the respective azimuth and elevation, μ is the step size, x(n) is the excitation signal at time n, and e(n) is the error signal, defined by the following formula

$$e(n) = y(n) - \hat{h}^T(n)x(n) \tag{3.3}$$

where y(n) is the binaural signal at the time n.

In order to handle the random head movements of the subject, the filter coefficients, that in this case corresponds to the HRIR in each specific position, will be stored in a three dimensional matrix, reproducing the behaviour of the activation

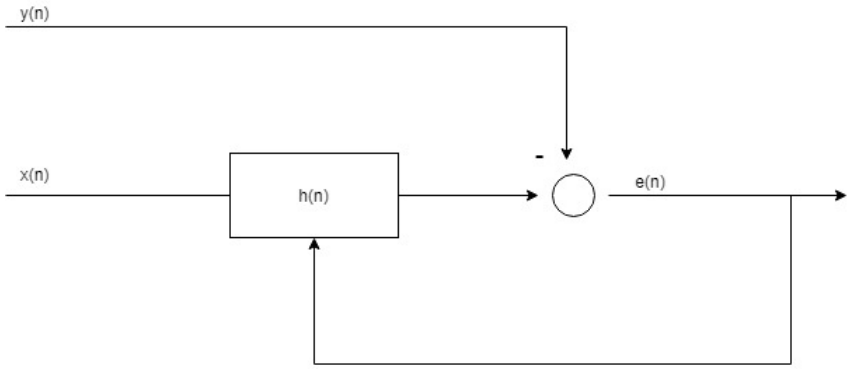

Figure 3.3: Adaptive filtering schema. In our implementation, y(n) represents the binaural signal, x(n) is the excitation signal, h(n) are the filter coefficients (HRIR) and e(n) is the error signal

matrix described in *He, Ranjan and Woon-Seng's* works[26] [55] [27]. This allows to update only HRIR at the time, since for each head position, only one row of the matrix will be active. This implies that the complexity of this algorithm will be the same as in a normal NLMS, but the memory required will be n*m times bigger (being n the number of azimuth positions, and m the number of elevation positions).

Moreover, as *He, Ranjan and Woon-Seng* explained [55] [27], as the HRIRs change gradually with the direction, the adaptive filter should change in a similar way. For this reason, in this implementation, the filter coefficients initial conditions in the unvisited directions are set to the obtained filter coefficients in the contiguous visited directions. For the already visited directions, the filter coefficient initial conditions would be the ones calculated in the previous visitation to that direction.

A variable step size technique is also suggested in the most recent work by *He, Ranjan and Woon-Seng* [27]. This variable step size NLMS method has been successfully used in many occasions [48] [62] [13], with different implementations and purposes, helping to control the learning rate depending on the number of iterations over the same scenario. The behaviour of the implemented variable step size can be described with the following equation

$$\mu(n) = \begin{cases} \mu_{max}, n = 1 \, or \, d(n) \neq d(n-1) \\ max\{\mu(n-1) - \Delta\mu, \mu_{min}\}, otherwise \end{cases} \tag{3.4}$$

In the first visitation on each direction, we use the maximum value for the step size. Then the value of the step size is reduced after each iteration, until it reaches the minimum value. To handle the directions changes and the value of μ for each case, a visitations matrix is used, that counts the number of visitations on each direction. The implementation of this matrix and the selected values will be described in detail in subsection 3.2.2.

Finally, for each direction and each iteration, the normalized mean square error (NMSE) is calculated, using the following formula

$$NMSE = 10log_{10}\left(\frac{||h-\hat{h}||_2^2}{||h||_2^2}\right) \tag{3.5}$$

where $\hat{h}$ is the calculated HRIR and h the HRIR from the CIPIC database.

Since each direction is going to be visited several times, one different HRIR is going to be obtained in each one of this visitations. In order to obtain the best HRIR possible for each direction, the NMSE is calculated. As it is explained in the most recent work by *He, Ranjan and Woon-Seng* [27], there are several strategies to select the most accurate HRIR from the results obtained from the NMSE. These strategies go from the averaging of the best candidates, to just pick the HRIR with the lowest NMSE. As it is explained in the cited article [27], choosing the HRIR with the lowest NMSE provides better results. The implementation of this algorithm of selection will be explained in detail in subsection 3.2.2.

3.1.2 Excitation signal

In all the previously cited examples [26] [55] [27] of personalized HRTF acquisition using NLMS-based algorithms, gaussian white noise is used as the excitation signal. Despite the results have been satisfactory in every case, there are other studies that imply that there other kind of signals that can outperform white noise.

In 2007, *Majdak et al* [40] described how to use a different excitation signal, called multiple exponential sweep method (MESM), for HRTF acquisition. This method is based in the overlapping sweeps in an optimized way. An evaluation, using multiple loudspeakers and fixed position microphones, reduced significantly the measurement duration, outperforming other excitation signals as maximum-length sequence and exponential sweep.

Antweiler, Telle et al [2] propose another approach, where they use a perfect sweep to excite a NLMS-type filter, similar to the one used in this implementation. A perfect sweep can be described with the following equation

$$P(\nu) = \begin{cases} \exp(\frac{-j4m\pi\nu^2}{M^2}); 0 \leq \nu \leq M/2 \\ P^*(M-\nu); M/2 < \nu \leq M \end{cases} \tag{3.6}$$

where ν is the frequency index , M is the length of one period of the sequence, and m is the factor which determines the stretch of the time-stretched pulse. This article implies that this kind of signals increase significantly the convergence speed of the NLMS adaptation algorithm, while provides high robustness against distortions. It outperforms the results of other excitation signals as perfect sequences, ternary perfect sequences and white noise [2].

3.2 Software Design

This project is divided in two main software applications. The first one would capture the head movements and the signal at the ears of the receiver. The second one is a Matlab [41] application that processes both, the audio and head tracking recordings, and calculates the individual HRTF based on them.

3.2.1 Head tracking and sound capturing

In order to facilitate the real time extraction of data for personalised HRTF acquisition, a mobile application was implemented. This application is developed in Swift [19], and can be used in Machintosh mobile devices as iPad.

This application gives instructions to the user on how to orient the head, and the amount of time necessary to get a correct recording. Apart from these instructions, there is also a grid with all the positions successfully visited (marked in green) and all the positions left to be visited (marked in red), as it is described in figure 3.4. The excitation signal will be played through the whole process of the acquisition of the head movements.

The application would make synchronised recordings from both the head movements and the binaural recordings, in order to be processed later in a Matlab application that will calculate individualized HRTFs based on these measurements. This application is built to be used using Sennheiser's AMBEO headset, which provides a simple way of making high quality recordings directly on the ears, and Sennheiser's AMBEO Headtracker, that complements this headset, giving the position of the head, in attitude and heading reference system (AHRS), with an update rate of 12Hz.

At the moment of the completion of this report, this application is still under construction.

3.2.2 HRTF/HRIR calculation

Once we have obtained the binaural audio recording in each ear canal, and the synchronised recording of all the head movements captured with the head tracker, everything is ready to perform the HRTF/HRIR calculations.

The HRIR calculations software is implemented in Matlab [41], a programming platform and language, that facilitates the work with computational mathematics. All the calculations described in the previous section take place in a across several functions.

First of all, the audio recordings are quantized in order to synchronise the different sampling rates of the head tracker and the audio recording. In the case of a head tracker with a sampling rate of 12Hz, and an audio recording with a sampling rate of 44100 samples per second, this would mean that the audio signal should be split in blocks of 3675 samples to be correctly synchronised.

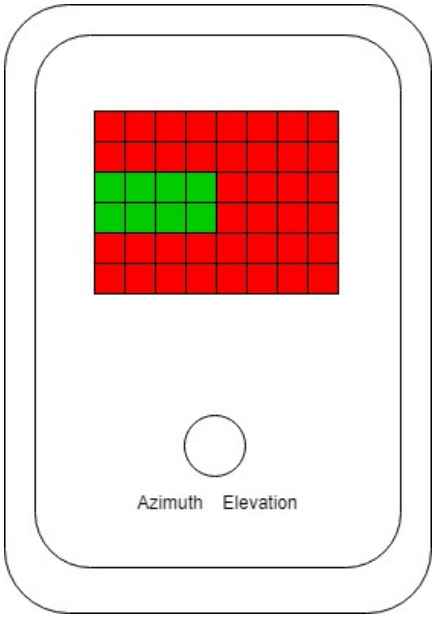

Figure 3.4: Digaram describing the layout of the measurments acquisition application

First of all, the data structures that are going to be used in the calculations are declared. To facilitate the evaluation, that will be explained in detail in chapter 4, we will use a direction grid similar to the one used in the CIPIC database [1]. This direction grid, is a 25x50 matrix where we have 50 elevation points (uniformly sampled in 5.625º steps from -45º to 230.625º), and 25 azimuth points (sampled at -80º, -65º, -55º, from -45º to 45º in steps of 5º, at 55º, 65º and 80º) as it is described in figure 4.1. Any other sampling distribution could have been used, so it is not decisive in the implementation, it is only relevant to the sizes of the data structures.

Following this spatial distribution, we will have the visitation matrix, that will store the number of visitations to each point in the matrix, another matrix to store the value of the minimum NMSE for each direction, a matrix with filter coefficients, that will contain the last HRIR calculated for each direction, and the HRIR matrix, that will contain the final HRIR for each direction.

Once the code is executed, the first thing it does is the update of the visitations matrix. The update of this matrix is strictly related with both the variable step size and the progressive behaviour of this algorithm. If it is the first visitation to an specific point in the matrix, the code will check if any of the neighbours has been previously visited. If so, it will copy the filter coefficients from the neighbour, in order to use it as initial conditions in the filter coefficients calculation for that direction.

According to the number of visitations, the step size (μ) will be updated. The values are set to μ_{max}=0.5, μ_{min}=0.05 with a decrement of 0.0005 in every step.

These conditions are observed to obtain the best overall performance in the cited paper [27].

Once we have updated the step size, and we have the direction and the blocks of both the reference and the binaural signals, the NLMS algorithm explained in subsection 3.1.1 is performed. For each direction, the NMSE is calculated. If that value is smaller than the one stored in the NMSE matrix for that direction, that means that the current HRIR is the most accurate so far. So, in that case, the current NMSE would be stored in the NMSE matrix, and the current HRIR would be stored in the HRIR matrix.

This process would be repeated until the whole recording of audio and head movements are processed.

Chapter 4

Evaluation

The evaluation was performed by implementing a batch of simulations, in which the signal received at the user's ear canals is synthesised by filtering different excitation signals with the HRIRs contained in the CIPIC database [1].

This is a database of high-spatial-resolution head-related transfer functions measured on 45 different subjects, at 25 different azimuths (sampled at -80º, -65º, -55º, from -45º to 45º in steps of 5º, at 55º, 65º and 80º) and 50 elevations (uniformly sampled in 5.625º steps from -45º to 230.625º)[1]. These measurements are obtained from different subjects. HRIRs for subject 003 from CIPIC database are the ones used for this procedure. The decision of choosing this subject was purely random.

In this evaluation we have decided to compare the behaviour of this algorithm using two different excitation signals, gaussian white noise and perfect sweeps. Gaussian white noise has been used extensively used in other cases, like in the studies made by *He, Ranjan and Woon-Seng* [26] [55] [27] in which this algorithm is based on, with successful results.

However, *Antweiler, Telle et al* [2] suggest that perfect sweep outperforms the

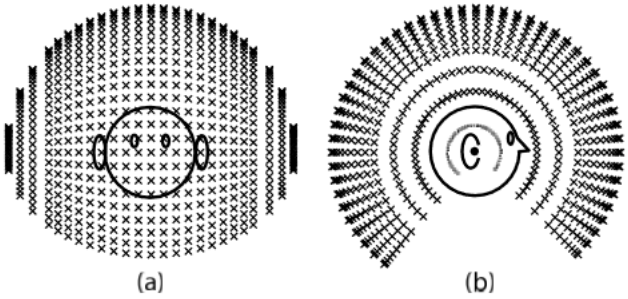

Figure 4.1: Sketch representing the location of the measurement points a) front b) side

results of other excitation signals in this type of NLMS-based algorithms for HRTF acquisition. The perfect sweep used in this simulation was generated using Aulis Telle's code implementing his own article [2] on the use of perfect sweeps on NLMS-based for acoustic system identification. This code was available at Aachen University's website under BSD license.

The chosen length of this excitation signal is half of the block size. As explained in [47], when using this kind of signals, the measurements have to be long enough to extract all the delayed components, so the signal has to be significantly shorter than the capturing period. More information on the decision to use this excitation signal was previously explained in section 3.1.2.

Apart from the excitation signals, the amount of time spent in each point, or visitation time, is analysed to see its influence in the obtained HRTFs. The simulation is run under different visitation time conditions. The combination of the visitation time and the signal to noise ratio, which is also studied in this simulation, seem to be crucial in the accuracy of the obtained HRTFs [26] [55] [27].

Lastly, the influence of the way the head is moved during the process is also studied. Two different kind of movements are simulated in this evaluation procedure. The first one is a basic movement where the head would go across the space in order, by elevation and azimuth, while the second method reproduces random head movements. All these simulation scenarios and their conditions will be explained in section 4.1.

To obtain the accuracy of the measurements, the normalized mean square error (NMSE) is measured, which is described in the following equation

$$NMSE = 10log_{10}\left(\frac{||h - h'||_2^2}{||h||_2^2}\right) \tag{4.1}$$

where $\hat{h}$ is the calculated HRIR and h the HRIR from the CIPIC database.

Moreover, other characteristics of the obtained HRIRs are analysed in order to have a more detailed understanding of the accuracy of the obtained HRIRs. One of this features evaluated is the estimated interaural time difference (ITD) [35], which represents the difference in the reception time of the sound between both ears, and is described by the following formula

$$ITD(\theta) = arg_\tau max IACC(\theta, \tau) \tag{4.2}$$

$$IACC(\theta, \tau) = \frac{\int_{t1}^{t2} \rho_L(\theta, t)\rho_R(\theta, t+\tau)dt}{\sqrt{\int_{t1}^{t2} \rho_L^2(\theta, t)dt \int_{t1}^{t2} \rho_R^2(\theta, t)dt}}$$

(4.3)

where IACC is the interaural cross-correlation, ρL(ϑ, t) and ρR(ϑ, t) are the HRIR for the left and right ear with an incident angle from the source ϑin the moment t.

In the same way, the interaural level difference (ILD) is studied. To analyse the ILD, the power spectra ratio between the left and the right channel is calculated, based on a warped equivalent rectangular bandwidth (ERB) critical bands. The central frequency (f_c(i)) and bandwidth (f_{bw}(i)) of each band are defined by the following equations

$$f_c(i) = QL(\exp\frac{i\gamma}{Q} - 1) \tag{4.4}$$

$$f_{bw}(i) = \gamma L(\exp\frac{i\gamma}{Q}) \tag{4.5}$$

As suggested by several articles on HRTF processing [27] [7], the values selected to obtain both the central frequency and the bandwidth are Q = 9.265, L = 24.7 and γ= 1. From these operations, we obtain 40 bands with central frequencies that go from 26.083 hertz to 16932 hertz. Once we have the central frequencies and bandwiths, the ILD is obtained with the following equation

$$ILD = 10log_{10}\left(\frac{1}{40}\sum_{i=1}^{40}\frac{\sum_{f=f_c(i)-f_{bw}(i)/2}^{f_c(i)+f_{bw}(i)/2} H_R^2(f)}{\sum_{f=f_c(i)-f_{bw}(i)/2}^{f_c(i)+f_{bw}/2} H_L^2(f)}\right)(dB) \tag{4.6}$$

where left and right HRTF magnitude at frequency f is represented by H_L(f), H_R(f), respectively.

Finally, the spectral difference (SD) between the obtained HRTFs and the ones from the CIPIC database is calculated. For this operation, the same 40 ERB frequency bands are used. The spectral difference for the HRTF at each orientation can be defined by the following equation

$$SD = 10log_{10}\left(\frac{1}{40}\sum_{i=1}^{40}\frac{\sum_{f=f_c(i)-f_{bw}(i)/2}^{f_c(i)+f_{bw}(i)/2}[H_S(f) - H_D(f)]^2}{\sum_{f=f_c(i)-f_{bw}(i)/2}^{f_c(i)+f_{bw}/2} H_S^2(f)}\right)(dB) \tag{4.7}$$

One of the goals for this evaluation, is to prove that it this algorithm would be useful for commercial headsets as Sennheiser's AMBEO, as it has been previously mentioned. The head-tracker for this headset, has an update rate of 12Hz, so for this test we are going to use that update rate. However, other head tracking devices can increment their sampling rate until 100Hz, which presumably would increment the accuracy. Nonetheless, in this evaluation we are going to stick to a sampling rate of 12Hz.

For this procedure, the NLMS-based algorithm is executed with a filter length of 600. This filter length is chosen because it has been proved to be large enough to

give good results [27] , while not being computationally too expensive. Also based on *He, Ranjan and Woon-Seng*'s article [26] [55] [27], the values for the variable step size are set to μ_{max}=0.5, μ_{min}=0.05 with a decrement $\Delta\mu$= 0.0005 in every step, as it was explained in section 3.2.2

4.1 Simulation scenarios

Under the previously explained initial conditions, different scenarios where implemented in order to have a deeper understanding of the algorithm's behaviour.

One of the most critical things to test is the algorithm's vulnerability to unwanted noises. This project aims to create an application that can acquire, with accuracy, personalised HRTFs, without depending on lab facilities, as anechoic chambers, or professional loudspeakers. Determining to which extent the algorithm is able to acquire HRTFs with precision under different signal to noise conditions is vital, thus.

The signal to noise ratio conditions evaluated are 50dB, 40dB, 30dB, 20dB and 10dB. In previous tests of similar algorithms [55][27], the simulations where done under a signal to noise ratio of 50dB. It was decided to analyse how the algorithm behaves in noisier conditions, in steps of 10dB, as we expect this to have a major influence in the accuracy of the obtained HRTF, and because the mobile nature of this project's goal suggests that this algorithm could be used in far from ideal situations.

The time visiting each one of the points of the grid is also important in the results. For this simulation, we discretized the space taking as a reference the points used in the CIPIC database [1]. This discretization was convenient for the evaluation, as it made it easier to compare and determine the accuracy of the measurements. Each one of the points in the grid was visited for the same amount of time, testing how the visitation time affects the obtained HRTFs. The visitation times used for each point were 10s, 5s, 1s, 0.5s and 0.25s.

The head movements are also tested in these simulations. For comfort, and to make more friendly the usage of this application by general users, it would be helpful to determine to which extent the nature of the head movements affects the results. In the first scenario the head movements will go across the grid in order, by elevation and azimuth, as it is explained in figure 4.4 and figure 4.5. In the second scenario the head movements would follow a random route across the grid, accessing always contiguous grid locations, replicating how random head movements would be. The path of this random route is described in figure 4.2 and figure 4.3.

Finally, the influence of the excitation signal is also evaluated. The excitation signal can have a relevant role in the behaviour of the NLMS-based algorithm used for the HRTF acquisition. As it was explained more thoroughly in section 3.1.2,

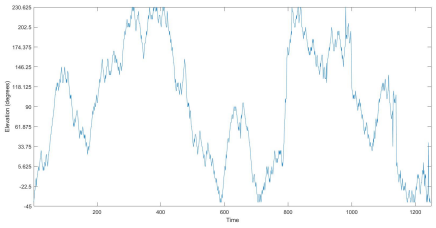

Figure 4.2: Head movement variations in elevation across time (Random head movements)

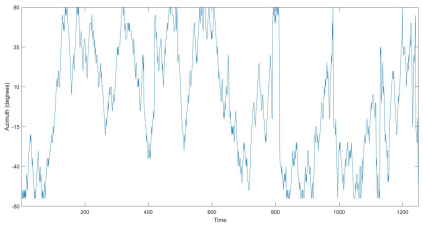

Figure 4.3: Head movement variations in azimuth across time (Random head movements)

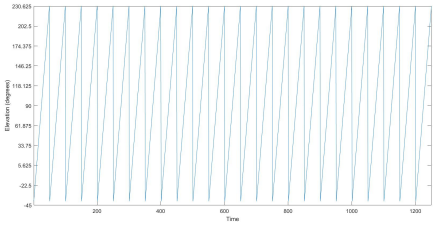

Figure 4.4: Head movement variations in elevation across time (Ordered head movements)

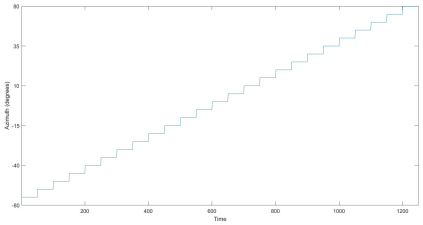

Figure 4.5: Head movement variations in azimuth across time (Ordered head movements)

two excitation signals, gaussian white noise and perfect sweep, are compared to determine its role in the performance of this algorithm.

4.2 Simulation results

Following the previously described conditions, the simulations were performed between the 14th and the 21st of May of 2019. These scenarios, were the following:

- Five different signal to noise (SNR) conditions: 10dB, 20dB, 30dB, 40dB, 50dB.
- Five different visitation times for each point in the grid: 0.25s, 0.5s, 1s, 5s, 10s.
- Two different excitation signals: Gaussian white noise and perfect sweep.
- Two different head movements: following the grid in order, and random head movements.

This makes a total of 5x5x2x2=100 simulation scenarios. The influence of the different variables in the results will be analysed in the following subsections:

4.2.1 Signal to noise ratio

Observing the results we can say the the signal to noise ratio (SNR) is one of the major factors that affect to the accuracy of the measurements, which was somewhat

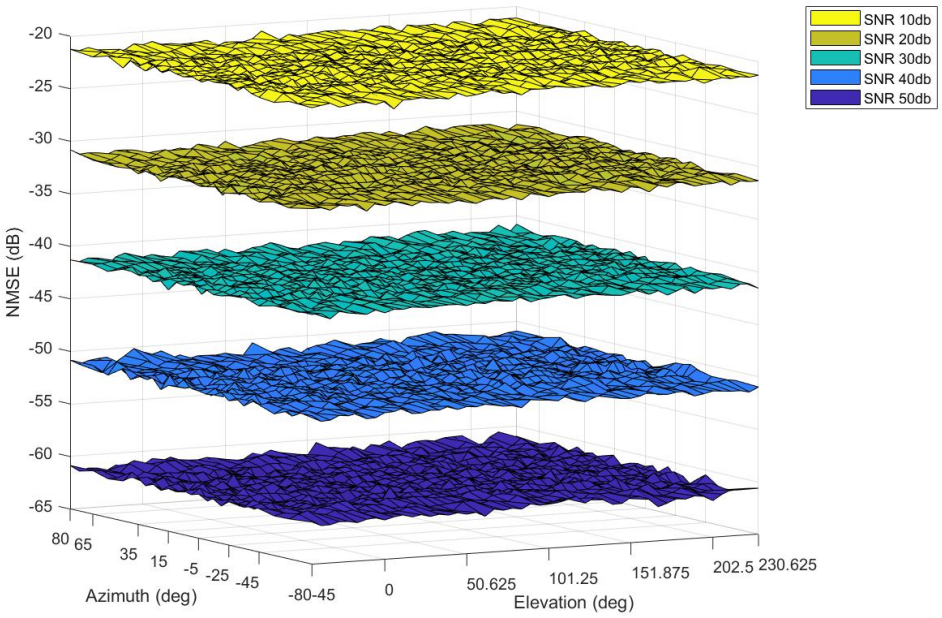

Figure 4.6: NMSE at different SNR conditions. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random.

expected. As in any audio recording or signal, SNR plays a major role when it comes to identify the different sources, and in the quality and definition of the desired signal.

As we can see, the NSME for each one of the SNR levels are very homogeneous across the grid. In the case described in figure 4.6, with a visitation time of 5 seconds, random head movements and using a perfect sweep as excitation signal, the difference within the maximum and the minimum NMSE for signal to noise ratios of 50dB and 40dB is only 2.2541 dB and 2.2543dB respectively.

When the signal to noise ratio is lower, under the same conditions, the results are even more homogeneous, with a difference between the maximum and the minimum NMSE of 1.665 dB, 1.5528 dB and 1.9936 dB for signal to noise ratio scenarios of 30 dB, 20 dB and 10 dB respectively.

However, the difference between the mean NMSE for each one of the scenarios is significant. We can see that the accuracy of the measurements drops as the signal to noise ratio decreases. In figure 4.6 it is easy to perceive how the mean NMSE decreases with the SNR, with almost 40 dB of difference between the results obtained at 50 dB of SNR and the results obtained with a SNR of 10 dB.

Studying the difference between the interaural level difference (ILD) for the reference HRTFs and the one calculated at the simulation, we can also find relevant

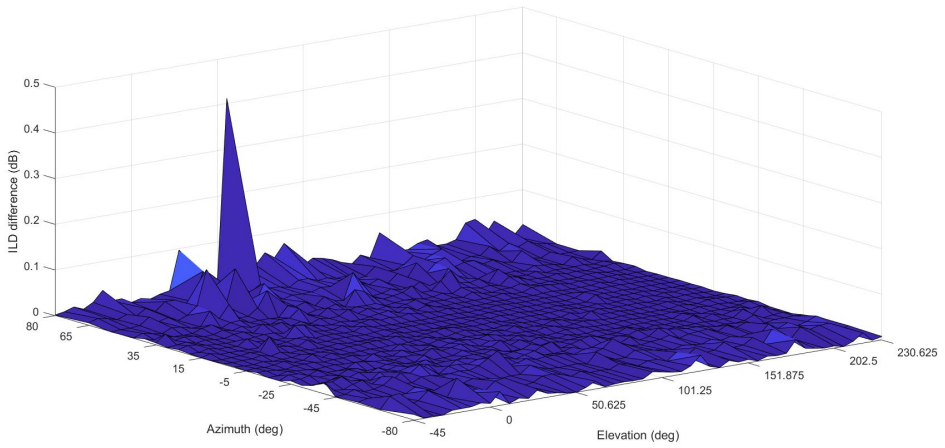

Figure 4.7: ILD difference (dB) between CIPIC's HRTFs and measured HRTFs. SNR = 50dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random.

variations caused by the different signal to noise ratio. As we can see in figure 4.7, with the same conditions as the ones cited before, and a SNR of 50 dB, the values for the difference in decibels between the ILD for CIPIC's HRTF and the ILD for the calculated HRTF is quite homogeneous and low across the whole grid. Some peaks can be observed close to one of the edges, getting to a 0.4 dB difference, but most of the grid presents values under 0.1 dB, with a mean difference of 0.0075 dB.

However, when the SNR decreases, the differences increase significantly. As an example, we can take the results in the same conditions (Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random) but with a SNR of 30 dB, as represented in figure 4.8. The differences across the whole grid still are stable in values under 0.8 dB, but in the areas close to 80^{o} azimuth we can see lots of sharp peaks. The differences get over 1 dB in several cases and one of the peaks overtakes 1.2 dB. Finally, in figure 4.9 we can observe that when the SNR goes as low as 10 dB, the noise affects the ILD across the whole grid, with very high error values, getting over 6 dB at some points. This pattern is repeated in every single case, as the ILD difference always increases as the SNR decreases.

Moreover, once again, we can see that having a good SNR is directly related with having a low spectral difference (SD) value. The values, as we can see in figure 4.11 and figure 4.10, are not as homogeneous in the case of ILD. We can see that the error is lower in the central values of the grid, but it increases when it gets closer to the edges. However, we can see that the values are closely related with the SNR. Under the previously cited conditions, with a visitation time of 5s, random head movements, and using a perfect sweep as excitation signal, the mean SD value is -58.787 dB for a SNR of 50 dB. However, when the SNR gets to 10 dB,

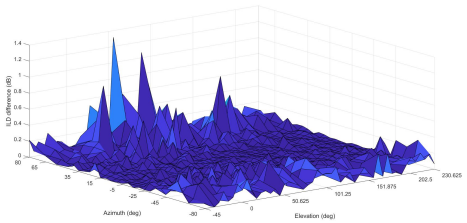

Figure 4.8: ILD difference (dB) between CIPIC's HRTFs and measured HRTFs. SNR = 30dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

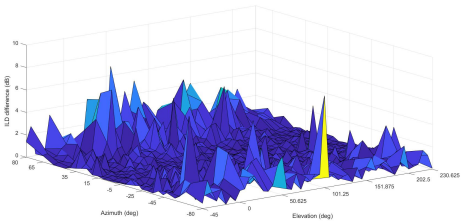

Figure 4.9: ILD difference (dB) between CIPIC's HRTFs and measured HRTFs. SNR = 10dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

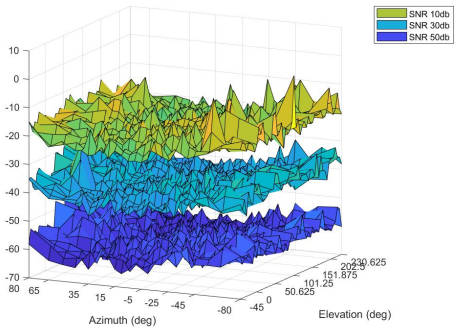

Figure 4.10: Spectral difference (in dB) for the left ear at SNR = 10dB, 30dB and 50dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

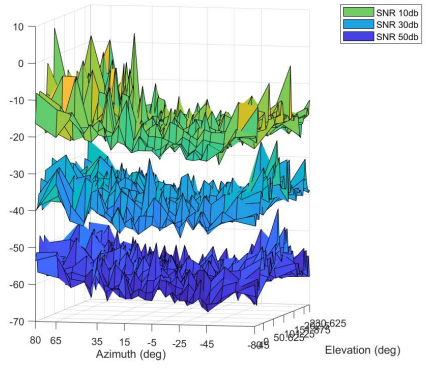

Figure 4.11: Spectral difference (in dB) for the right ear at SNR = 10dB, 30dB and 50dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

this value decreases until only -19.2573 dB.

Finally, the estimated interaural time difference (ITD) is, like all the other descriptors, directly affected by the SNR variations. In general terms, the ITD is very well preserved in most of the cases, as, even in the noisiest conditions, the peaks do not preserve relevant shifts. It is easily observable in figure 4.14 how the ITD differences between the CIPIC's HRIR and the measured HRIRs is almost none for a SNR of 50 dB, only one non zero value can be seen in the plot, and the difference is under 20 microseconds. However, we can see how this almost perfect match in the ITDs is degraded as we decrease the SNR. In figure 4.13 we can see how the results are still good, with only a few non zero values, and all of them under 25 microseconds, but definitely worse than in the previous case. Finally, with a SNR of 10 dB, in figure 4.12 it is easily observable how the differences increases, having exceptional values close to 250 microseconds. However, still, most of the values are

zero.

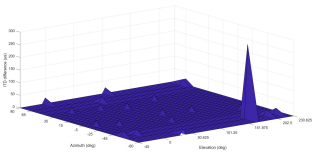

Figure 4.12: ITD difference (us) between CIPIC's HRIRs and measured HRIRs at SNR = 10dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

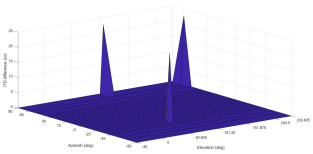

Figure 4.13: ITD difference (us) between CIPIC's HRIRs and measured HRIRs at SNR = 30dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

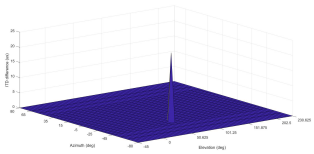

Figure 4.14: ITD difference (us) between CIPIC's HRIRs and measured HRIRs at SNR = 50dB. Simulation scenario : Vistation time = 5s. Excitation signal = Perfect sweep. Head movements = Random

4.2.2 Excitation signal

Two different excitation signals were used in these simulations, gaussian white noise and perfect sweeps. The first thing that was evaluated in this simulation was the differences between the mean NMSE for each one of the excitation signals in every scenario.

As we can observe in figure 4.15, there is no relevant difference in terms of mean NMSE. The difference between the best and the worst excitation signal in every scenario, with random and ordered head movements, and different SNR, for a visitation time of 1 second in each point of the grid, is always under 1 dB. Furthermore it is hard to extract any conclusions, in terms of mean NMSE.

However, when the parameter we change is the visitation time instead of the SNR, we can see major changes between the different excitation signals. The results are pretty similar again, for visitation signals over 1 second. Nonetheless, we can see significant differences for visitation times of 0.5 and 0.25 seconds. As we can see in figure 4.16, there are minor (under 1dB), but obvious, differences when the visitation time is 0.5s. However, when the visitation time drops to 0.25s, the differences are quite significant, with the perfect sweep obtaining a mean NMSE 6dB bigger than the gaussian white noise. Moreover, with this same visitation time and a SNR of 50 dB, the differences are even more remarkable, with a variation of 21.0888 dB between the mean NMSE calculated for the gaussian white noise with random head movements and the perfect sweep under the same conditions.

When it comes to the interaural level difference (ILD) the differences, once again, are very subtle. The plots for the ILD with 1 second of point visitation time, SNR of 30 dB and random head movements using gaussian white noise and perfect sweep as excitation signals can be seen in figure **??** and figure **??** respectively. The behaviour of both signals is pretty similar, however we can see that the gaussian white noise slightly outperforms perfect sweep if we take a look into the mean

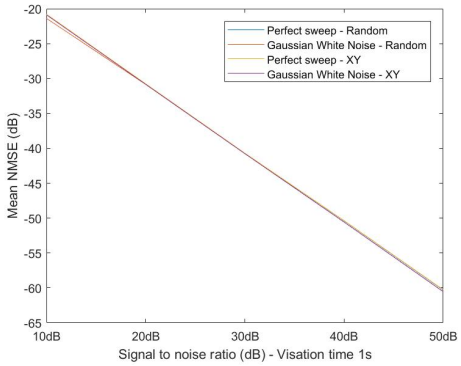

Figure 4.15: Mean NMSE for the different excitation signals with different SNR and same visitation time (1 second)

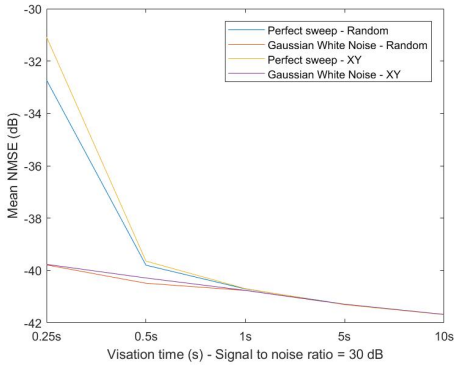

Figure 4.16: Mean NMSE for the different excitation signals with same SNR (30 dB) and different visitation time

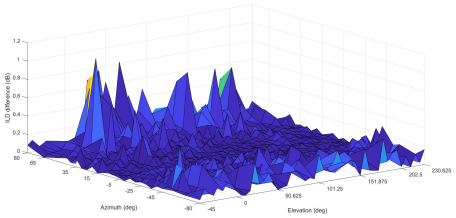

Figure 4.17: ILD difference (in dB) for visitation time = 1s , SNR = 30dB and random head movements using white noise as excitation signal

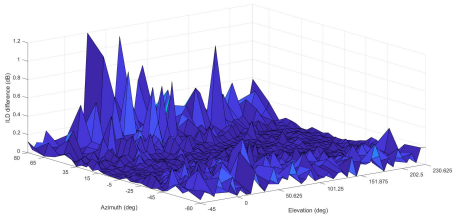

Figure 4.18: ILD difference (in dB) for visitation time = 1s , SNR = 30dB and random head movements using perfect sweep as excitation signal

ILD difference between the measured and the CIPIC's HRTFs in this scenario. The mean ILD difference for gaussian white noise is 0.0701 dB while perfect sweep has a mean difference of 0.074 dB. We can see that the pattern is repeated if the SNR is increased to 40dB, where the mean ILD difference for gaussian white noise is 0.0247 dB and for the perfect sweep is 0.0255 dB. In any of the cases the difference is notable, but it is worth mentioning, as it is consistent for every signal to noise ratio scenario. Only for very short visitation times (under 1 second) and high SNR (over 30 dB), like in the case of the NMSE, it is possible to see bigger differences, with gaussian white noise outperforming perfect sweep.

In terms of spectral difference (SD) once again it is hard to find major differences between both excitation signals. In figure 4.19 we can find the SD for the left ear, with a visitation time of one second, using gaussian white noise as excitation signal, while in 4.20 we have the same plot using a perfect sweep as excitation signal under the same conditions. As we can see the differences are minimal, once again. Taking a look at the mean values, we can say that gaussian white noise outperforms the perfect sweep. The mean SD values, for both ears, using gaussian

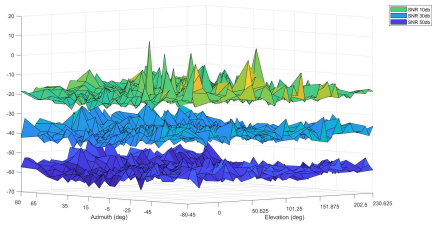

Figure 4.19: SD for the left ear with visitation time = 1s and random head movements using white noise as excitation signal

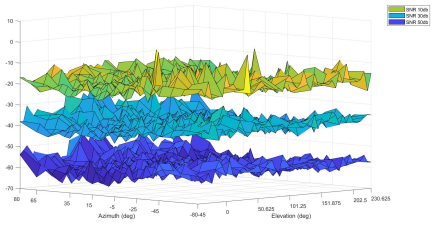

Figure 4.20: SD for the left ear visitation time = 1s and random head movements using perfect sweep as excitation signal

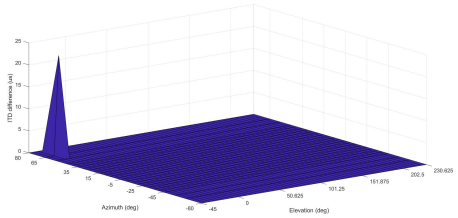

Figure 4.21: ITD difference (us) with visitation time = 1s, SNR = 30 dB and random head movements using white noise as excitation signal

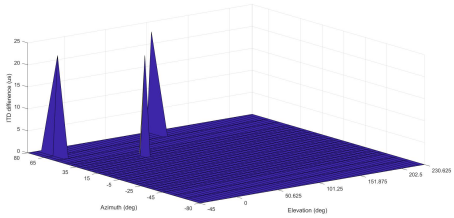

Figure 4.22: ITD difference (us) with visitation time = 1s, SNR = 30 dB and random head movements using perfect sweep as excitation signal

white noise, are -18.7778 dB, -38.6013 dB and -58.3515 dB for 10 dB, 30 dB and 50 dB of signal to noise ratio, respectively. On the other hand, the perfect sweep performs with -18.7093 dB, -38.4954 dB and -58.1070 under the same conditions. The differences are under 0.3 dB in every case.

Finally, the interaural time difference (ITD) is very well preserved in every scenario. However, there are small differences between both excitation signals that are worth remarking. In figure 4.21 and 4.22 there are plots representing the difference, in microseconds, between the estimated ITD for the HRIR's from the CIPIC database and the ones calculated in this simulation. As we can see the error is zero in most of the grid. However if we pay attention to the mean ITD difference, white noise outperforms the perfect sweep, with a mean value of 0.0181 microseconds of difference, while the perfect sweep has 0,0544 microseconds of mean difference. Once again, as we have seen in previous sections, a smaller SNR entails a bigger ITD difference, as we can see in figure 4.23 and figure 4.24. In this case, both signals perform with the seam mean ITD difference, 0,3265 microseconds.

4.2.3 Visitation time

The time spent in each one of the points in the grid, or visitation time, as we are going to see in this section, is directly related with the quality of the obtained

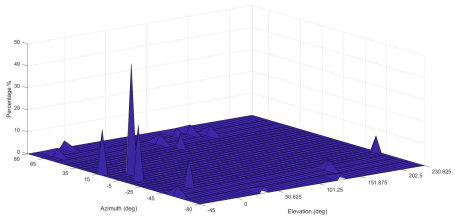

Figure 4.23: ITD with visitation time = 1s, SNR = 10 dB and random head movements using white noise as excitation signal

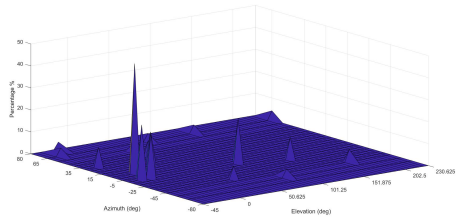

Figure 4.24: ITD with visitation time = 1s, SNR = 10 dB and random head movements using perfect sweep as excitation signal

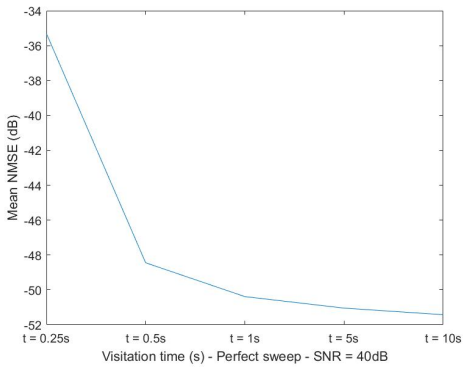

Figure 4.25: Mean NMSE with different visitation times with perfect sweep as excitation signal. SNR = 40 dB and random head movements.

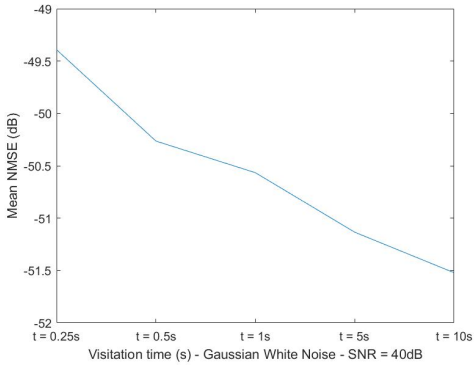

Figure 4.26: Mean NMSE with different visitation times with gaussian white noise as excitation signal. SNR = 40 dB and random head movements.

HRTFs. As we can see in figures 4.25 and 4.26, in every case the mean NMSE decreases when the visitation time increases. Both plots show the performance of the algorithm under a simulation with random head movements and a signal to noise ratio of 40dB. For very short visitation times, the performance of the perfect sweep is significantly worse, as it was explained in section 4.2.2. It is also worth remarking that, the lowest the SNR is, the smallest is the influence of the visitation time. For instance, the difference between 0.25 seconds and 10 seconds of visitation time, using gaussian white noise as excitation signal, random head movements and a SNR of 10 dB, is less than 1.5 dB.

The difference between the ILD for the calculated HRTFs, and the ones from the CIPIC database is also dependant on the visitation times. However, the differences are not excessive. In figure 4.27 we can see the ILD difference using gaussian white noise as excitation signal, a SNR of 40 dB, random head movements and a visitation time of 5 seconds, while in figure 4.28 we can see the same plot under the same conditions but with 0.25 seconds of visitation time. The differences in the plot are

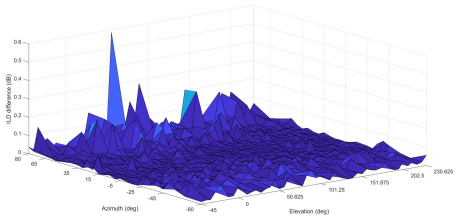

Figure 4.27: ILD difference using gaussian white noise as excitation signal, and random head movements, SNR = 40dB, with 5 seconds as visitation time

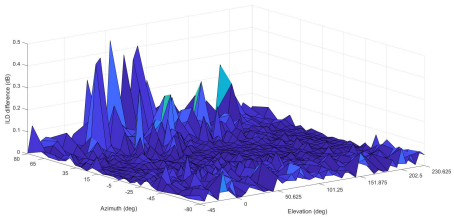

Figure 4.28: ILD difference using gaussian white noise as excitation signal, and random head movements, SNR = 40dB, with 0.25 seconds as visitation time

not remarkable, but if we pay attention to the mean ILD difference in each case, we can see how it decreases when the visitation time increases. When the visitation time is 0.25 seconds, the mean ILD difference is 0.0264 dB while for 5 seconds of visitation time, the mean ILD difference falls to 0.0220 dB. The differences are not significant, but yet worth mentioning.

The spectral difference (SD) is, once again, closely related with the visitation time. In figure 4.29, there is a plot with the mean SD for a simulation at a signal to noise ratio of 40 dB, with random head movements and using gaussian white noise as excitation signal. From this plot we can extract that the visitation time plays a major role in this measurements, as the SD value improves when the visitation time increases. The differences are not dramatic, as it is only 1.8403 dB the difference between the values for 0.25 seconds and 10 seconds, but still we can say that it makes a difference.

Finally, in the case of the interaural time difference (ITD), we can not find obvious differences dependant on the visitation times. For times under 5 seconds, the ITD differences between the estimated ITD for the CIPIC database's HRIRs and the ones calculated by the simulations, remain constant and with very low values. In figure 4.30 we can see the ITD difference with a visitation time of 0.25 seconds. The mean value for the ITD difference is 0.0181 dB, which is exactly the same value for the same test with a visitation time of 5 seconds. Nonetheless, when the visitation time is 10 seconds, the estimated ITD for the CIPIC database's HRIRs matches perfectly (0 dB mean error) with the ones obtained by this simulation.

4.2.4 Head movements

The type of head movements was also analyzed to determine its influence in the HRTF aquisition. As it was shown in figure 4.15 and 4.16, the type of head movements do not affect in a relevant way to the mean NMSE of the computed HRTFs. Only in the case of the simulation using 0.25 seconds as visitation time, SNR of 30 dB, and perfect sweep as excitation signal we can see some relevant difference,

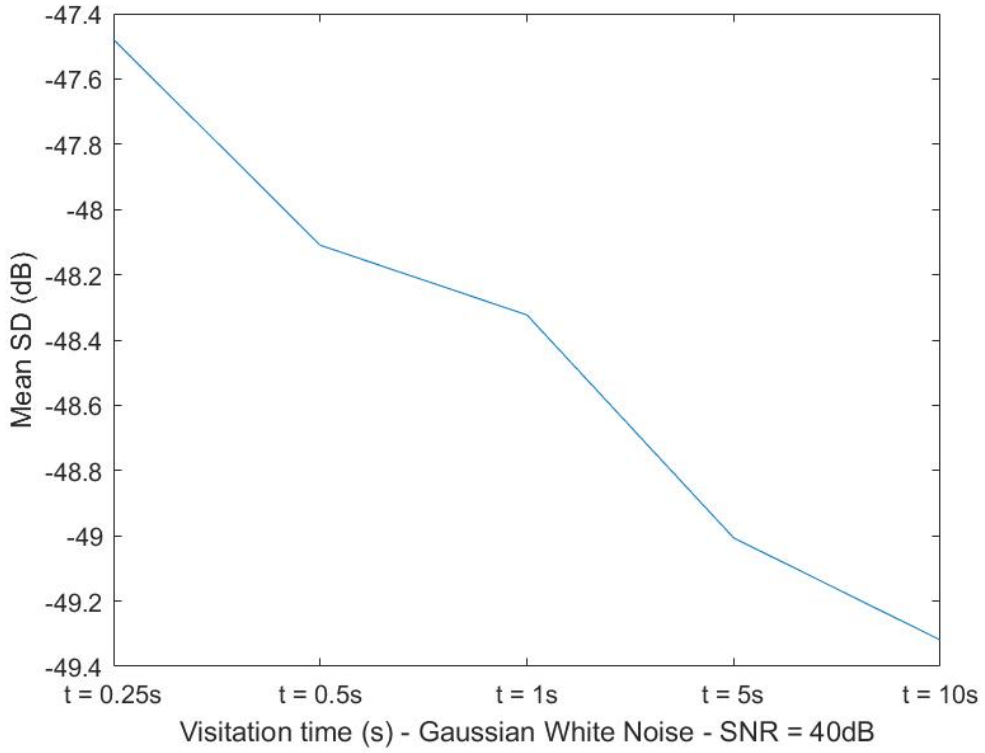

Figure 4.29: Mean SD in dB between CIPIC's HRTFs and measured HRTFs. SNR = 40dB. Excitation signal = Gaussian white noise. Head movements = Random.

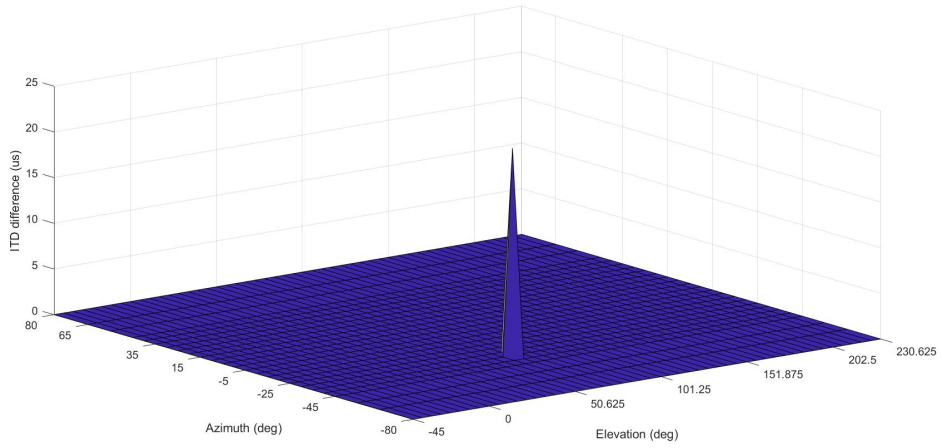

Figure 4.30: ITD difference (us) between CIPIC's HRTFs and measured HRTFs. SNR = 40dB. Simulation scenario : Vistation time = 0.25s. Excitation signal = Gaussian white noise. Head movements = Random.

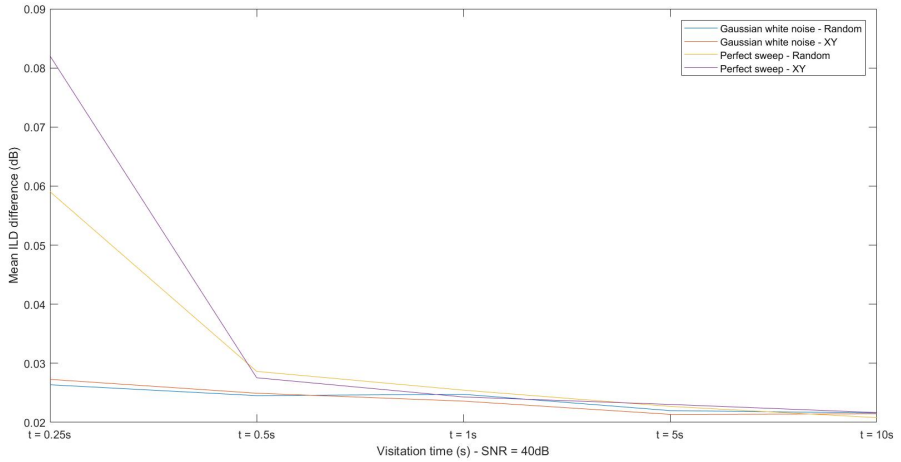

Figure 4.31: ILD difference (dB) between CIPIC's HRTFs and measured HRTFs. SNR = 40dB.

where, surprisingly, the random head movements outperform the ordered head movements by a little more than 1 dB.

In the case of the interaural level difference (ILD) it is hard to understand the influence of the head movements in its results. Taking a look at the mean ILD difference between the CIPIC database's HRTFs and the ones obtained by the simulation, it is still hard to extract any conclusion. In figure 4.31 we can see how the ordered head movements generally outperform the random head movements, but not always, and it is impossible to find a pattern to understand why of this inconsistencies in the performance.

In the case of the spectral difference (SD), once again, the results are not conclusive. If we take a look at the mean SD with the different head movements, we can see that the random head movements outperform the ordered movements. In figure 4.32 we can see the mean SD for each type of signal and head movements, for each visitation time. We can see that the differences are minimal, under 0.2 dB in every case, except for the ones mentioned before, with perfect sweep and short visitation time (0.25s and 0.5s).

Finally, the interaural time difference (ITD) presents some interesting differences between the different head movements. To evaluate the difference we can see figure 4.34 and figure 4.33. The plots display the difference between the estimated ITD for the HRIR's of CIPIC's database and the calculated ones. In figure 4.34 we can see how the ordered head movements present peaks randomly distributed with low values across the grid, while having some specific peaks over 200us. On the other hand, we can see how in the case the random head movements we can see also peaks all over the grid, but always with low values, under 30us. The mean ITD difference also suggests that random head movements perform better, with a

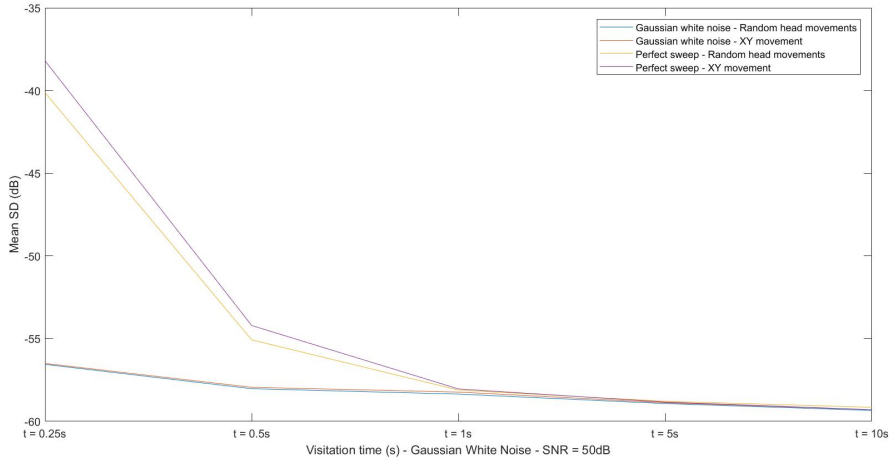

Figure 4.32: Spectral difference. SNR = 40dB.

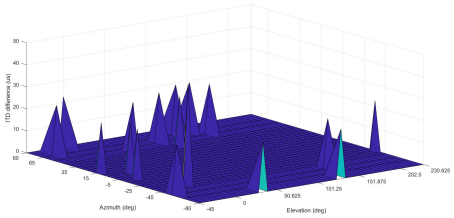

Figure 4.33: ITD difference using gaussian white noise as excitation signal, and random head movements, SNR = 10dB, with 1 seconds as visitation time

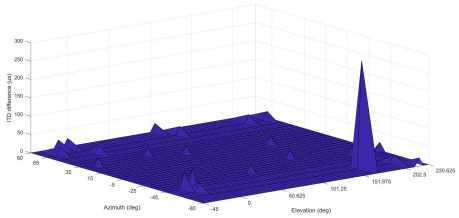

Figure 4.34: ITD difference using gaussian white noise as excitation signal, ordered head movements, SNR = 10dB, with 1 seconds as visitation time

mean value of 0.3265 us, while the ordered movements present a mean value of 0.6531 us.

Chapter 5

Discussion

This project aimed to create a fast HRTF acquisition method, which allowed to obtain personal HRTFs in a faster, portable and more user-friendly way. This project stressed in creating and evaluating an algorithm which allowed to calculate personal HRTFs based on synchronised recordings of binaural sound signals and head tracking information. To evaluate this algorithm, a batch of simulations was performed.

The results of the simulations, which comprehend 100 different scenarios, provide some meaningful results worth discussing. From this results it is possible to extract the influence of different factors that affect the whole HRTF acquisition process.

This results give relevant information that could help to create a robust mobile application that could handle both the extraction of the recordings of both the head tracking and binaural audio signals, and the HRTF calculations based on these recordings.

5.0.1 Excitation signal

From the simulation results we can extract that there are no relevant differences between the excitation signals in most of the cases. Only for very short visitation times, 0.25 seconds and 0.5 seconds, we can see a remarkable difference between gaussian white noise and perfect sweep. In these scenarios gaussian white noise clearly outperforms perfect sweep.

Further investigation regarding the length of the perfect sweep [2], and changing parameters as the head tracking sampling rate, could help to solve this problem with shorter visitation times.

However, it is also worth mentioning that in terms of computational speed, perfect sweep outperforms gaussian white noise. It's rapid convergence speed [2] on NLMS based algorithms was specially relevant with longer visitation times (5 seconds and 10 seconds), increasing the speed of the algorithm over a 20%.

5.0.2 Head movements

The importance of the way the head movements were performed was inconclusive. The results of the simulations performed with ordered head movements and random head movements did not differ significantly.

However, the lack of influence of the head movements on the acquired HRTFs is an interesting and encouraging result for the aim of this project. This conclusion suggests that unconstrained head movements can be suitable for HRTF acquisition. The performance of free movements would make easier and more comfortable for the user the whole HRTF acquisition process,

5.0.3 Signal to noise ratio

The signal to noise ratio plays a crucial role in the quality and accuracy of the calculated HRTFs. Every error measurement performed on the evaluation presented worse results when the signal to noise ratio decreased, as it was predictable.

The normalised mean squared error (NMSE) of the obtained HRTFs increased significantly when the signal to noise ratios decreased. However, it is promising to see that, even though the error increased, the interaural time difference (ITD) and interaural level difference (ILD) still maintained very low values in these conditions.

ITD and ILD play a very important role in the localisation of sounds in the horizontal plane. The preservation of these properties even in poor signal to noise ratio conditions implies that the obtained HRTFs under these conditions would still perform reasonably well on the horizontal plane. On the other hand, localisation in elevation, and in depth (front-back) would still be affected by these interferences.

However, a subjective evaluation would be necessary to determine to which extent this variations affect to the ability to locate sound sources in a virtual auditory display.

5.0.4 Visitation time

The time spent in each one of the points of the grid, or visitation time, also proves to be extremely relevant and directly related with the quality of the acquire HRTFs.

In every case, the quality of the calculated HRTFs increased in every measurement performed over them. It is specially remarkable in terms of normalised mean squared error (NMSE), where the improvement caused by the increase of visitation time could exceed 20 decibels in some cases. Moreover, as it was mentioned in section 5.0.1, the perfect sweep presented some performance problems when the simulation used 0.5 seconds and 0.25 seconds as visitation time.

However, it was very promising to see that when the excitation signal used was gaussian white noise, the differences where not that pronounced. In these kind of scenarios, the differences in NMSE between the simulations using 0.25

seconds as visitation time, and the simulation with 10 seconds as visitation time, never exceeded 4 decibels. This reasonably good performance under very short visitation times give very good perspectives in the goal of shortening the times needed for the HRTF acquisition process.

Chapter 6

Conclusion

An application that calculates personalised HRTFs based on recordings of binaural sound signals and head tracking information has been conceptualised, developed and tested. It uses a normalised least means squared-based algorithm that, using a progressive-activation approach with variable step size, calculates individual HRTFs allowing unconstrained head movements. This algorithm was tested with with a batch of simulations, synthesising sounds using CIPIC's database to reproduce the different angles, in elevation and azimuth, that a sound source would make when a user is moving the head in a two dimensional space. One hundred simulation scenarios were evaluated, to determine the influence of the signal to noise ratio, the type of head movements during the measurements, the time spent in each one of the points in the virtual two dimensional grid, and the type of excitation signal, have on the calculated HRTFs. The normalised mean squared error, interaural time difference, interaural level difference and spectral difference were studied in order to have a better understanding on the way the alteration of this variables affect the HRTF calculation.

The results of the implementation conclude that this algorithm succeeds in computing HRTFs with remarkable accuracy in very different circumstances. Even for very low signal to noise ratios, important directional cues as interaural level difference and interaural time difference, are very well preserved. These directional cues are the main responsible for localisation of sound sources in the horizontal plane.

This feature, combined with the good results for this evaluation, when the simulation performed random head movements, makes it able to use in different possible set-ups. This way, users can obtain their own personalised HRTFs without requiring high end equipment or lab facilities. However, there is a number of interesting ways to continue this project in order to expand and improve these results.

At the moment of the completion of this report, the mobile application designed

to make the synchronised recordings of both the binaural audio signals an the head tracking is not finished. It would be necessary to complete this application in order to confirm if, as the results of the evaluation imply, this system is suitable for real time acquisition of personalised HRTFs.

Moreover, it would be necessary to evaluate this mobile application on test subjects to assure that the instructions are intuitive and easy to use. Also, the time required to obtain all these measurements is crucial in the success of this application. Usability is decisive in order to make customers embrace this system, making all the process clear and comfortable for the user.

Furthermore, testing the whole process in a real scenario would give meaningful information about the behaviour of the system with the different excitation signals. Perfect sweeps have been proved to be more robust against non-linear distortions presented in some reproduction devices as loudspeakers, so it would not be unlikely to assume that it's performance might improve in a real scenario.

Once the mobile application is finished, it would be crucial to test that this setup works with the equipment that we are recommending for this process, Sennheiser's AMBEO headset with Sennheiser AMBEO Headtracker, using a Macintosh mobile device, as an iPad, as a processor and reproduction device.

Using the iPad as reproduction device for the excitation signal would make way easier and portable the process of acquisition of the data, but it would compromise the sound quality. All the articles on HTRF acquisition cited in this paper, use high quality loudspeakers or monitors to reproduce the reference signal, so it is likeable that it would be necessary to use a high quality reproduction device to get the signal with the desired precision and quality.

Furthermore, the quality and location inside the ear of the Sennheiser AMBEO headset are crucial to get accurate HRTF. It would be necessary to test if the microphones are deep enough on the ear canal to obtain an accurate HRTF. If the microphone is placed in the outer part of the ear canal, it might lose some of the characteristics of the user's pinna. These characteristics are mostly related with the perception of the elevation and depth, as well as permitting the distinction between front and back [52][67].

Other approach to improve this project, is finding ways to reduce the time that it takes to get all the measurements. Every HRTF acquisition method using only one speaker requires no less than 30 minutes [27] to perform all the necessary measurements. That is an unacceptable amount of time if we want this method to be embraced by common users with mobile devices. There are different ways to reduce this time.

One of the most obvious ways is to reduce the amount of points where we need to make measurements in order to obtain a complete HRTF field. This could be achieve by different methods, as the one described by *Gamper et al* [14], where they implement tetrahedral interpolation (using Delaunay triangulation) with barycen-

tric weights, in order to interpolate HRTFs in azimuth, elevation, and distance.

Other approach is the use of spherical harmonics-based methods in order to obtain a complete HRTF field based on measurements made on discrete points. There is plenty of literature on this subject, that was explained more thoroughly in section 2.3, and a method based on these techniques was attempted. Due to time constraints, and because of the fact that this algorithms are computationally very expensive, making them unsuitable for a fast HRTF calculation, this improvements were aborted at the time of completion of this report.

However, the literature suggests that these methods provide very promising results, so that it could achive the goal of obtaining a complete HRTF field just performing a reduced amount of measurements. Some of these articles [61] even describe how, in the case of the CIPIC HRTF database, the amount of sampling points could be reduced in more than 400 points, which, therefore, would decrease significantly the amount of time required for this procedure.

Other interesting improvement would be to integrate the measurement acquisition procedure with VR/AR headsets as Magic Leap. In the end, one of the main fields where this techniques can be extensively used is in augmented and/or virtual reality environments.

Moreover, some studies [16] have used VR/AR displays to guide the subjects movements for HRTFs. Nonetheless, the size and nature of this headsets can have an influence on the measurements, so their effect on the recorded signals must be compensated on the HRTF calculation algorithm.

Finally, subjective evaluation on both applications would contribute to improve the project. Subjective evaluation on the mobile application would give us notes on how intuitive it is, and about how easy it is to perform all the actions needed to acquire the necessary information from the head movements and the audio signal.

Also it would be a good way to analyse if the times required for the data acquisition are reasonable and suitable for a commercial product. User feedback can help to find different strategies and detect some possible flaws in the designed procedure, so it would be convenient to perform this kind of tests to polish the whole process and improve the user experience.

Subjective evaluation on the acquired HRTFs would be a necessary input as well. There have been several studies [18] on the actual necessity of the use of personalised HRTFs, some of them suggesting that the quality of the virtual auditory display is highly dependent on this individualisations [69]. However this project would be more meaningful if we can get to prove that there is an actual improvement in terms of accuracy and sense of presence [34] [37] compared with non-individualised HRTFs.

Basic sound localisation tests would give meaningful feedback on the improvement of the HRTF quality for common users. Moreover, VR and AR environments could help to make more in-depth testing on these features. These technologies

could help to determine if the personalization of the HRTFs affects, not only the quality of the sounds, but also features as the sense of presence [34] [37], engaging with the virtual experience, the way the users interact with the virtual environment.

The use of virtual auditory displays has been proved to promote exploratory behaviours in virtual environments [64], so an increase in the quality of the HRTFs could expand these features. A significant improvement in the 3D sound experience would be decisive to introduce this kind of technology in fields as virtual and augmented reality environments or gaming.

Chapter 7

Appendix A - Simulation results

Perfect sweep - Random head movements

Visitation time	0,25s				0,5s				1s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-18,0038	0,9796	0,8047	-19,6928	-18,3990	0,6531	0,7596	-20,4924	-18,7093	0,3265	0,7129	-20,8041
20db	-27,1634	0,4898	0,2692	-27,5828	-28,2170	0,1995	0,2418	-30,2989	-28,6027	0,1088	0,2348	-30,7636
30dB	-34,2657	0,5079	0,1019	-32,7244	-37,7091	0,0726	0,0831	-39,7979	-38,4954	0,0544	0,0740	-40,7037
40dB	-38,3925	0,5261	0,0591	-35,3083	-47,1297	0,0544	0,0286	-48,4435	-48,1562	0,0181	0,0255	-50,3820
50dB	-40,1203	0,5261	0,0485	-36,1968	-55,0741	0,0726	0,0108	-55,0368	-58,1070	0,0000	0,0077	-60,3009

Visitation time	5s				10s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-19,2573	0,5986	0,6583	-21,3290	-19,5771	0,3628	0,6737	-21,7123
20db	-29,0760	0,0726	0,2252	-31,3244	-29,5344	0,1088	0,2049	-31,6972
30dB	-39,0389	0,0544	0,0674	-41,2937	-39,4861	0,0363	0,0668	-41,6826
40dB	-48,8250	0,0181	0,0227	-51,0418	-49,2511	0,0181	0,0208	-51,4204
50dB	-58,7870	0,0181	0,0075	-61,0368	-59,1521	0,0000	0,0069	-61,4325

Figure 7.1: Perfect sweep - Random Head Movements - SD = Mean Spectral Difference (dB) - ITD = mean ITD difference (us) - ILD = mean ILD difference (dB) - NMSE = mean NMSE (dB)

Perfect sweep - XY movements

Visitation time	0,25s				0,5s				1s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-18,0148	1,4331	0,7580	-19,5010	-18,3352	0,6531	0,7373	-20,4736	-18,7222	0,5805	0,7182	-20,7850
20db	-26,9501	1,1066	0,2628	-26,8816	-28,1127	0,2177	0,2468	-30,2465	-28,5888	0,1270	0,2397	-30,7552
30dB	-33,5577	0,8345	0,1227	-31,0812	-37,6602	0,0726	0,0819	-39,6426	-38,4607	0,0544	0,0766	-40,6941
40dB	-36,9720	1,0522	0,0821	-32,8739	-46,9681	0,0544	0,0275	-47,8295	-48,1573	0,0363	0,0243	-50,3556
50dB	-38,1925	1,0159	0,0725	-33,4269	-54,2059	0,0544	0,0120	-53,4570	-58,0405	0,0000	0,0076	-60,2589

Visitation time	5s				10s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-19,3263	0,1995	0,6724	-21,3310	-19,5778	0,4354	0,6382	-21,7152
20db	-29,1082	0,1088	0,2219	-31,3169	-29,5442	0,1088	0,2041	-31,7072
30dB	-39,2152	0,0000	0,0661	-41,3089	-39,4852	0,0544	0,0649	-41,6866
40dB	-48,8532	0,0000	0,0230	-51,0404	-49,5757	0,0000	0,0217	-51,4301
50dB	-58,8348	0,0000	0,0073	-61,0325	-59,2951	0,0000	0,0068	-61,4276

Figure 7.2: Perfect sweep - Ordered movements - SD = Mean Spectral Difference (dB) - ITD = mean ITD difference (us) - ILD = mean ILD difference (dB) - NMSE = mean NMSE (dB)

Gaussian white noise - Random head movements

Visitation time	0,25s				0,5s				1s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-18,1979	0,4898	0,7908	-20,2691	-18,4389	0,5805	0,7680	-20,5715	-18,7778	0,3265	0,7486	-20,8781
20db	-27,8745	0,1451	0,2644	-29,9983	-28,4474	0,3991	0,2330	-30,6037	-28,5948	0,0726	0,2380	-30,8399
30dB	-37,5745	0,0181	0,0835	-39,7884	-38,2456	0,0544	0,0747	-40,4882	-38,6013	0,0181	0,0701	-40,7632
40dB	-47,4791	0,0181	0,0264	-49,3915	-48,1095	0,0000	0,0245	-50,2643	-48,3231	0,0544	0,0247	-50,5661
50dB	-56,5550	0,0000	0,0086	-57,2856	-58,0266	0,0000	0,0082	-60,2139	-58,3515	0,0000	0,0076	-60,5449

Visitation time	5s				10s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-19,2314	0,4717	0,7174	-21,3820	-19,6564	0,2721	0,6654	-21,7691
20db	-29,0847	0,1633	0,2130	-31,3658	-29,6148	0,0907	0,1953	-31,7423
30dB	-39,0446	0,0363	0,0670	-41,2963	-39,4723	0,0907	0,0666	-41,6810
40dB	-49,0063	0,0181	0,0220	-51,1346	-49,3194	0,0000	0,0216	-51,5216
50dB	-58,9242	0,0000	0,0069	-61,1318	-59,3417	0,0000	0,0065	-61,5217

Figure 7.3: Gaussian white noise - Random Head Movements - SD = Mean Spectral Difference (dB) - ITD = mean ITD difference (us) - ILD = mean ILD difference (dB) - NMSE = mean NMSE (dB)

Gaussian white noise - XY movements

Visitation time	0,25s				0,5s				1s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-18,1673	0,4354	0,7314	-20,2500	-18,5098	0,4535	0,7583	-20,5625	-18,7391	0,6531	0,7357	-20,8745
20db	-27,7918	0,1270	0,2489	-29,9947	-28,2175	0,1451	0,2428	-30,3649	-28,6058	0,2902	0,2238	-30,8227
30dB	-37,6771	0,0907	0,0792	-39,7658	-38,1317	0,0181	0,0839	-40,2867	-38,5406	0,0181	0,0752	-40,7593
40dB	-47,3807	0,0181	0,0273	-49,2889	-47,9969	0,0181	0,0249	-50,2540	-48,2902	0,0000	0,0236	-50,5552
50dB	-56,5003	0,0181	0,0092	-56,6706	-57,9427	0,0181	0,0080	-60,1911	-58,2304	0,0000	0,0079	-60,5337

Visitation time	5s				10s			
SNR	SD	ITD	ILD	NMSE	SD	ITD	ILD	NMSE
10db	-19,2397	0,3991	0,6515	-21,3782	-19,6639	0,5079	0,6628	-21,7667
20db	-29,1750	0,1088	0,2153	-31,3596	-29,5540	0,1451	0,2161	-31,7433
30dB	-39,0561	0,0000	0,0677	-41,2890	-39,5199	0,0363	0,0672	-41,6791
40dB	-48,9267	0,0181	0,0214	-51,1367	-49,2813	0,0181	0,0214	-51,5203
50dB	-58,8722	0,0000	0,0069	-61,1249	-59,2952	0,0000	0,0069	-61,5180

Figure 7.4: Gaussian White Noise - Ordered Movements - SD = Mean Spectral Difference (dB) - ITD = mean ITD difference (us) - ILD = mean ILD difference (dB) - NMSE = mean NMSE (dB)

Bibliography

[1] V Ralph Algazi et al. "The cipic hrtf database". In: *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*. IEEE. 2001, pp. 99–102.

[2] Christiane Antweiler et al. "Perfect-Sweep NLMS for Time-Variant Acoustic System Identification". In: Mar. 2012. DOI: 10.1109/ICASSP.2012.6287930.

[3] Matthieu Aussal, François Alouges, and Brian Katz. "A study of spherical harmonics interpolation for HRTF exchange". In: *Proceedings of Meetings on Acoustics ICA2013*. Vol. 19. 1. ASA. 2013, p. 050010.

[4] Alice P Bates, Zubair Khalid, and Rodney A Kennedy. "Novel sampling scheme on the sphere for head-related transfer function measurements". In: *IEEE Transactions on Audio, Speech, and Language Processing* 23.6 (2015), pp. 1068–1081.

[5] JA Rod Blais and MA Soofi. "Spherical harmonic transforms using quadratures and least squares". In: *International Conference on Computational Science*. Springer. 2006, pp. 48–55.

[6] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[7] Jeroen Breebaart, Fabian Nater, and Armin Kohlrausch. "Spectral and spatial parameter resolution requirements for parametric, filter-bank-based HRTF processing". In: *Journal of the Audio Engineering Society* 58.3 (2010), pp. 126–140.

[8] C Phillip Brown and Richard O Duda. "A structural model for binaural sound synthesis". In: *IEEE transactions on speech and audio processing* 6.5 (1998), pp. 476–488.

[9] Martyn Cooper. "Three dimensional auditory display: Issues in applications for visually impaired students". In: Georgia Institute of Technology. 2004.

[10] Scott C Douglas. "Introduction to adaptive filters". In: *Digital signal processing handbook* (1999), pp. 7–12.

[11] Chris Dunn and Malcolm J Hawksford. "Distortion immunity of MLS-derived impulse response measurements". In: *Journal of the Audio Engineering Society* 41.5 (1993), pp. 314–335.

[12] Ramani Duraiswami and Vikas C Raykar. "The manifolds of spatial hearing". In: *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* Vol. 3. IEEE. 2005, pp. iii–285.

[13] Gerald Enzner. "Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses". In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2008, pp. 393–396.

[14] Hannes Gamper. "Head-related transfer function interpolation in azimuth, elevation, and distance". In: *The Journal of the Acoustical Society of America* 134 (Dec. 2013), EL547. DOI: 10.1121/1.4828983.

[15] Woon-Seng Gan et al. "Personalized HRTF measurement and 3D Audio Rendering for AR/VR Headsets". In: May 2017.

[16] Woon-Seng Gan et al. "Personalized HRTF Measurement and 3D Audio Rendering for AR/VR Headsets". In: *Audio Engineering Society Convention 142*. Audio Engineering Society. 2017.

[17] Bill Gardner, Keith Martin, et al. *HRFT Measurements of a KEMAR Dummy-head Microphone*. 1994.

[18] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. "Do We Need Individual Head-Related Transfer Functions for Vertical Localization? The Case Study of a Spectral Notch Distance Metric". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* PP (Apr. 2018), pp. 1–1. DOI: 10.1109/TASLP.2018.2821846.

[19] James Goodwill and Wesley Matlock. "The Swift Programming Language". In: Mar. 2015, pp. 219–244. ISBN: 978-1-4842-0401-6. DOI: 10.1007/978-1-4842-0400-9_17.

[20] D Wesley Grantham. "Spatial Hearing and". In: *Hearing* (1995), p. 297.

[21] Graham Grindlay and M Alex O Vasilescu. "A multilinear approach to HRTF personalization". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2007.

[22] Nguyen Duy Hai et al. "Fast HRFT measurement system with unconstrained head movements for 3D audio in virtual and augmented reality applications". In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2017, pp. 6576–6577.

[23] Dorte Hammershøi and Henrik Møller. "Binaural technique—Basic methods for recording, synthesis, and reproduction". In: *Communication acoustics*. Springer, 2005, pp. 223–254.

[24] Aki Härmä et al. "Augmented reality audio for mobile and wearable appliances". In: *Journal of the Audio Engineering Society* 52.6 (2004), pp. 618–639.

[25] Simon S Haykin, Bernard Widrow, and Bernard Widrow. *Least-mean-square adaptive filters*. Vol. 31. Wiley Online Library, 2003.

[26] Jianjun He, Rishabh Ranjan, and Woon-Seng Gan. "Fast continuous HRTF acquisition with unconstrained movements of human subjects". In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2016, pp. 321–325.

[27] Jianjun He et al. "Fast Continuous Measurement of HRTFs with Unconstrained Head Movements for 3D Audio". In: *Journal of the Audio Engineering Society* (Nov. 2018). DOI: 10.17743/jaes.2018.0050.

[28] Tatsuya Hirahara et al. "Head movement during head-related transfer function measurements". In: *Acoustical Science and Technology* 31.2 (2010), pp. 165–171.

[29] Simon Holland, David R Morse, and Henrik Gedenryd. "AudioGPS: Spatial audio navigation with a minimal attention interface". In: *Personal and Ubiquitous computing* 6.4 (2002), pp. 253–259.

[30] Akio Honda et al. "Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game". In: *Applied Acoustics* 68.8 (2007), pp. 885–896.

[31] Hugeng Hugeng and Dadag Gunawan. "Improved Method for Individualization of Head-Related Transfer Functions on Horizontal Plane Using Reduced Number of Anthropometric Measurements". In: *Journal of Telecommunications* 2 (May 2010), pp. 31–41.

[32] Naoya Inoue et al. "HRTF modeling using physical features". In: *Forum Acusticum 2005*. 2005, pp. 199–202.

[33] Beth Jelfs et al. "Characterisation of signal modality: Exploiting signal nonlinearity in machine learning and signal processing". In: *Journal of Signal Processing Systems* 61.1 (2010), pp. 105–115.

[34] Bill Kapralos, MR Jenkin, and E Milios. "Virtual audio systems". In: *Presence: Teleoperators and Virtual Environments* 17.6 (2008), pp. 527–549.

[35] Brian FG Katz and Markus Noisternig. "A comparative study of interaural time delay estimation methods". In: *The Journal of the Acoustical Society of America* 135.6 (2014), pp. 3530–3540.

[36] George F Kuhn. "Model for the interaural time differences in the azimuthal plane". In: *The Journal of the Acoustical Society of America* 62.1 (1977), pp. 157–167.

[37] Pontus Larsson et al. "Auditory-induced presence in mixed reality environments and related technology". In: *The Engineering of Mixed Reality Systems*. Springer, 2010, pp. 143–163.

[38] Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. "Navigation system for the blind: Auditory display modes and guidance". In: *Presence* 7.2 (1998), pp. 193–203.

[39] Enrique A Lopez-Poveda and Ray Meddis. "A physical model of sound diffraction and reflections in the human concha". In: *The Journal of the Acoustical Society of America* 100.5 (1996), pp. 3248–3259.

[40] Piotr Majdak, Peter Balazs, and Bernhard Laback. "Multiple exponential sweep method for fast measurement of head-related transfer functions". In: *Journal of the Audio Engineering Society* 55.7/8 (2007), pp. 623–637.

[41] MATLAB. *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[42] M. Mirbagheri, L. Atlas, and A. K. C. Lee. "Regression Factor Analysis With an Application to Continuous HRIR Measurement". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.2 (2018), pp. 415–421. ISSN: 2329-9290. DOI: 10.1109/TASLP.2017.2780989.

[43] Jafar Ramadhan Mohammed et al. "An Efficient Adaptive Noise Cancellation Scheme Using ALE and NLMS Filters." In: *International Journal of Electrical & Computer Engineering (2088-8708)* 2.3 (2012).

[44] Henrik Møller et al. "Binaural technique: Do we need individual recordings?" In: *Journal of the Audio Engineering Society* 44.6 (1996), pp. 451–469.

[45] Henrik Møller et al. "Head-related transfer functions of human subjects". In: *Journal of the Audio Engineering Society* 43.5 (1995), pp. 300–321.

[46] Masayuki Morimoto and Yoichi Ando. "On the simulation of sound localization". In: *Journal of the Acoustical Society of Japan (e)* 1.3 (1980), pp. 167–174.

[47] Swen Müller and Paulo Massarani. "Transfer-function measurement with sweeps". In: *Journal of the Audio Engineering Society* 49.6 (2001), pp. 443–471.

[48] Constantin Paleologu, Silviu Ciochina, and Jacob Benesty. "Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation". In: *IEEE signal processing letters* 15 (2008), pp. 5–8.

[49] Constantin Paleologu et al. “An overview on optimized NLMS algorithms for acoustic echo cancellation”. In: *EURASIP Journal on Advances in Signal Processing* 2015.1 (2015), p. 97.

[50] Gerald S Pollack. “Hearing for defense”. In: *Insect Hearing*. Springer, 2016, pp. 81–98.

[51] Martin Pollow et al. “Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition”. In: *Acta acustica united with Acustica* 98.1 (2012), pp. 72–82.

[52] Martin Pollow et al. “Fast measurement system for spatially continuous individual HRTFs”. In: *Audio Engineering Society Conference: UK 25th Conference: Spatial Audio in Today's 3D World*. Audio Engineering Society. 2012.

[53] Ville Pulkki, Mikko-Ville Laitinen, and Ville Sivonen. “HRTF measurements with a continuously moving loudspeaker and swept sines”. In: *Audio Engineering Society Convention 128*. Audio Engineering Society. 2010.

[54] Rishabh Ranjan and Woon-Seng Gan. “Natural listening over headphones in augmented reality using adaptive filtering techniques”. In: *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 23.11 (2015), pp. 1988–2002.

[55] Rishabh Ranjan, Jianjun He, and Woon-Seng Gan. “Fast continuous acquisition of HRTF for human subjects with unconstrained random head movements in azimuth and elevation”. In: *Audio Engineering Society Conference: 2016 AES International Conference on Headphone Technology*. Audio Engineering Society. 2016.

[56] Lord Rayleigh and A Lodge. “Iv. on the acoustic shadow of a sphere”. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 203.359-371 (1904), pp. 87–110.

[57] Jonas Reijniers, Bart Partoens, and Herbert Peremans. “DIY measurement of your personal HRTF at home: low-cost, fast and validated”. In: *Audio Engineering Society Convention 143*. Audio Engineering Society. 2017.

[58] Douglas D Rife and John Vanderkooy. “Transfer-function measurement with maximum-length sequences”. In: *Journal of the Audio Engineering Society* 37.6 (1989), pp. 419–444.

[59] Griffin D Romigh. *Individualized head-related transfer functions: efficient modeling and estimation from small sets of spatial samples*. 2012.

[60] Griffin D Romigh et al. “Efficient real spherical harmonic representation of head-related transfer functions”. In: *IEEE Journal of Selected Topics in Signal Processing* 9.5 (2015), pp. 921–930.

[61] C Sandeep Reddy and Rajesh M Hegde. "On the Conditioning of the Spherical Harmonic Matrix for Spatial Audio Applications". In: *arXiv e-prints*, arXiv:1710.08633 (2017), arXiv:1710.08633. arXiv: 1710.08633 [eess.AS].

[62] Hyun-Chool Shin, Ali H Sayed, and Woo-Jin Song. "Variable step-size NLMS and affine projection algorithms". In: *IEEE signal processing letters* 11.2 (2004), pp. 132–135.

[63] Fabián C Tommasini et al. "Individual head-related impulse response measurement system with 3D scanning of pinnae". In: *Proceedings of Meetings on Acoustics 22ICA*. Vol. 28. 1. ASA. 2016, p. 055007.

[64] Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A Brewster. "Auditory display design for exploration in mobile audio-augmented reality". In: *Personal and Ubiquitous computing* 16.8 (2012), pp. 987–999.

[65] JA Veltman, AB Oving, and Adelbert W Bronkhorst. "3-D audio in the fighter cockpit improves task performance". In: *The International Journal of Aviation Psychology* 14.3 (2004), pp. 239–256.

[66] Wahidin Wahab, Dadang Gunawan, et al. "Enhanced individualization of head-related impulse response model in horizontal plane based on multiple regression analysis". In: *2010 Second International Conference on Computer Engineering and Applications*. Vol. 2. IEEE. 2010, pp. 226–230.

[67] Elizabeth M Wenzel et al. "Localization using nonindividualized head-related transfer functions". In: *The Journal of the Acoustical Society of America* 94.1 (1993), pp. 111–123.

[68] Bosun Xie. *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[69] Song Xu, Zhizhong Li, and Gaviriel Salvendy. "Individualization of head-related transfer function for three-dimensional virtual auditory display: a review". In: *International Conference on Virtual Reality*. Springer. 2007, pp. 397–407.