Generating Personalized HRTF Using Scanned Mesh from iPhone FaceID

WENKANG LIU

Division of Applied Acoustic

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

www.chalmers.se

Master's thesis 2023

Generating Personalized HRTF Using Scanned Mesh from iPhone FaceID

Supervisor: Sergejs Dombrovskis, China Euro Vehicle Technology AB
Examiner: Jens Arhens, Chalmers University of Technology

Cover: 3D scanning for Kemar via the HRTF plotting calculated by Kemar's 3D mesh

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustic
Chalmers University of Technology
Gothenburg, Sweden 2023

Generating Personalized Head-Related Transfer Function (HRTF)
using Scanned Mesh from iPhone FaceID
Wenkang Liu
Division of Applied Acoustic
Chalmers University of Technology

Abstract

In recent years, the advancements in virtual reality (VR) and augmented reality (AR) technologies have been impressive. Binaural audio rendering plays a vital role in these technologies and is used in various applications such as gaming, video conferencing, and hearing aids. Providing a high-quality immersive experience in a virtual environment heavily relies on the spatial audio quality.

The head-related transfer function (HRTF) describes how sound is filtered by the head, torso, and ears as it travels from the sound source to the listener's eardrum. To achieve spatial audio that better matches auditory perception, researchers have proposed several HRTF personalization methods, including measurement methods, database matching methods, modeling simulation methods, and anthropometric parameter regression methods.

This paper proposes a new modeling simulation method for personalized HRTF workflow that consists of three parts. Firstly, the participant's face and torso are scanned in 3D using the iPhone Face ID component. Secondly, the scanned mesh is optimized and cleaned using MeshLab and Blender. Finally, the personalized HRTF is generated using Mesh2hrtf and COMSOL. The effectiveness of the personalized HRTF is evaluated by comparing the simulated HRTF with the measured HRTF. Moreover, a test is designed using an adjustable equalizer-based headphone-speaker control to evaluate the performance of the generated personalized HRTF.

The results demonstrate that the HRTF generated using the FaceID scan grid is highly comparable to the measured HRTF and produces predictable outcomes in the listening test. This method shows promise as a low-cost alternative for customizing HRTFs.

Keywords: spacial audio, auditory perception, psychoacoustics, head-related transfer functions(HRTF), 3D scanning, mesh optimization, boundary element method(BEM), sound quality

Acknowledgements

I would like to sincerely thank all those who contributed to the successful completion of this project. First and foremost, I would like to thank my supervisors, Jens A. and Sergejs S., for their invaluable guidance and support throughout the research process. Their insights and feedbacks were instrumental in helping us achieve our goals.

I would also like to thank the faculty and staff of Divison of Applied Acoustic for providing me with support and resources. In addition, I would like to thank my classmates and friends who provided helpful discussions and feedback during the research process.

Finally, I would like to thank all the participants involved in the data collection process. Their contributions were critical to the success of this project. We hope that our research will have a meaningful impact in the field of HRTF and audio signal processing.

Wenkang Liu, Gothenburg, Dec. 2023

Contents

1

Introduction

Lots of theoretical studies show that about 60% of people perceive the objective world by visual perception, 30% by hearing, and the other 10% by touch and smell. Therefore, besides vision, voice is the main means of getting information about the objective world, which is very important in our daily life. Human hearing system plays an important role in understanding the message conveyed by language and understanding the dialogue in TV. In order to avoid danger, the auditory system is able to pick up the sound of danger and react in time. In contrast to the visual system, the auditory system "Listens" in all directions, and sounds can be heard from the back, up, or down. Furthermore, because of the physical nature of sound, for objects that are invisible because of obstacles, the auditory system can determine the basic characteristics and status of an object by means of voice. The human auditory system is very powerful. Using only two ears, the multi-dimension information can be easily distinguished, such as the vertical direction, the horizontal direction, the sound distance and the environment information. The research on the auditory system is one of the most advanced research fields in the field of audio signal processing.
In recent years, spacial audio technology has attracted the attention of domestic and international research institutes and manufacturers. How to realize spacial audio with high fidelity is one of the hot topics in the field of multimedia. Using a spacial audio technique to realistically display the spatial orientation of a sound source, using a binaural playback technique, for example, Head Related Transfer Function (HRTF). The Head Related Transfer Function Technique takes into account the transmission of sound waves to the ear as a filter, taking into account only the free field (with the exception of the listener itself and no other reflection object). More in details, HRTF reconstructs the spatial direction signal which is generated by the convolution of the original audio signal and the HRTF, so as to reproduce the real spacial sound in the two ears.

1.1 Motivation

The human auditory system is a complex system that is vital to our daily lives. Sound waves are collected by the external ear canal and transmitted to the inner ear, where they are amplified and processed by the auditory chain in the tympanic membrane. However, the sound that reaches the eardrum is not the same for all listeners because it is affected by the listener's head, torso and body (Møller, 1992). The brain analyzes these sound signals and calibrates our hearing with the help of

vision.

Interaural level difference (ILD) and interaural time difference (ITD) help the brain to discriminate the direction of sound in the horizontal plane. However, these cues are not sufficient for vertical localization. The complex diffusion, resonance and reflection of sound in the ear provide subtle differences in the frequency spectrum that can help the brain perceive vertical differences. In this case, the spectral and temporal properties of sound provide spatial cues that are hidden within fixed spectral cues.

Therefore, the use of HRTF rendering to produce realistic spatial audio effects can greatly improve our auditory immersion in virtual environments (Blauert, 1997). HRTF describes the reflections from the head, shoulders and neck and their effects on the sound reaching the eardrum. Since sounds from different locations have different frequency characteristics after passing through the HRTF, humans can distinguish between sounds from different locations.

Based on the structure modeling approach, the parts of the head, shoulders, and ears that influence the HRTF are created separately. These modules are then combined to get personalized HRTF. The method can be used to deal with every part of the system, and it is easy to realize in real time. Brown et al. (1998) proposed a structural model for the synthesis of binaural sound. The input is mono sound, followed by head shadow and shoulder echo. Then, the signals of both models are combined, and spacial audio is output with Pinna model. The model is based on the sound wave propagation and diffraction and considers the influence of human structure on sound waves. It has a practical physical meaning to model the shoulders, head, and ears of a person separately. According to the characteristics of the listener, the parameters of the model can be adjusted to get the personalized HRTF. The structure model is easy to implement in DSP (Digital Signal Processor). Geronazo et al. (2010) developed a personalized approach to Pinna Related Transfer Function (PRTF). Their approach only takes into account the effect of ear reflection on sound waves. The PRTF is divided into resonant and notch components. Later, they presented a personalized HRTF approach based on auricle parameters and a structural model based on auricle model. Through the use of human measuring parameters, the HRTF can be customized. Compared with the non-personalized structure model, the personalized HRTF is more effective, and the computation is low, so it can be realized in real time. Geronazo et al. (2013) developed a Mixed StructuralMode1 model for HRTF. Every module in the model can be chosen from the integration module, the measuring module and the data base. The synthesized HRTF has an exponential level, and the optimum combination is chosen as the HRTF. The hybrid structure model is flexible and has better location performance, but it needs a lot of calculation.
The HRTF personalization algorithm based on Principal Component Analysis (PCA) is generally used to analyze the HRTF database, and the Principal Components (PCs) with high correlation and relatively few are selected for further processing, so as to reduce dimension. Finally, high dimensional reconstruction is carried out to

obtain personalized HRTF. Kistler et al. (1992) measured the HRTF of 10 individuals at 265 azimuth angles, and then analyzed these 5,300 ($2 \times 10 \times 265$) HRTF with PCA. The results indicated that 90% of the original HRTF data set could be included in the original data set. The performance of the reconstructed HRTF with the five components is similar to that of HRTF. However, when the number of PC is reduced, the quality of the reconstructed HRTF is correspondingly reduced. Afterwards, Shin et al. (2008) used PCA for personalization of HRTF. Their approach first extracted Pinna Related Impulse Response (PRIR) from all individuals' temporal HRIR in the middle vertical plane (with 0 horizontal angle) in the HRTF database, and analyzed the ear response. Then, the first 5 main components are selected, and the measurer uses a graphical interface to adjust the weights of the 5 main components, and then re-synthesize the PRIR with the adjusted weights. The experiment results indicate that the HRIR with the proposed method has less location error and lower confusion rate than non-personalized HRIR. Hwang et al. (2008) proposed a similar approach to that of Shin, who also performed PCA analysis on HRTF database vertical data. However, they extended their choice of primary components to 12, and the difference between the reconstructed HRIR and the original HRIR was below 4.8%. The tester only has to make adjustments to the first three critical PCWs, which saves the adjustment time compared to the 5 PCWs.
The HRTF individualization approach based on database matching is used to map the individual HRTF to the database, and the HRTF is used as the personalized result. The shortcoming of this method is that it needs a lot of data base and representative data to get good effect. Apart from the structure model and the personalized approach based on the PCA, the other HRTF individualization approach is based on the measurement parameter. This approach is based on the assumption that some measuring parameters of the human body are similar, and that the HRTF and the individual HRTF are similar. Zotkin et al. In 2002, a personalized HRTF scheme was proposed on the basis of measuring parameter matching. Seven ear measurement parameters were chosen, and HRTF database was used to find the most suitable individuals for these 7 parameters. The HRTF of an individual is then used as a personalized HRTF. Experimental results indicate that the HRTF of the proposed method is superior to the conventional HRTF in terms of precision. Subsequently, Zotkin et al. (2003) proposed a Head and Torso Model to compensate for the low frequency loss, which further improved the experimental results. Algazi et al. (2007) proposed a new approach to modeling the ear based on measurement parameters. The Pinna Related Transfer Function (PRTF) was broken down into a few small pieces, and then a low order filter was used to describe them. Based on the measurement parameters of the human body, they set up the relationship between the parameters of the body and the filter coefficients, and finally, the personalized PRTF was obtained. The experiment results indicate that only a few parameters can be used to get the coefficients of the filter, and the PRTF can be approximated well. Iida et al. (2014) estimated a trough central frequency in the PRTF on the basis of measurement parameters of a person's ear, and then searched the HRRF database for HRTF nearest to those trough center frequencies, resulting in the best matching HRTF being deemed personalized. The HRTF localization performance of this method is similar to that of HRTF. Meshram et al. (2014) proposed an image

modeling approach for personalized HRTF. First, they take pictures of the head, shoulders, and other parts of the body with a camera. Then, they use advanced imaging techniques to estimate the 3D model of the head. Then, the acoustic equation can be solved by simulation to get personalized HRTF. The HRTF obtained by this method is superior to the HRTF of KEMAR, but the computation is too much and the calculation time is large. Torres et al. (2015) used Active Shape Models (ASM) to obtain the parameter characteristics of the subject subject through computer vision, and then select suitable parameters to search for the HRTF with the nearest parameter from the HRTF library.
Various methods have been proposed to personalize HRTFs, but they all have their limitations. In this paper, a new method to personalize HRTFs is tested by using 3D grid scanning of an iPhone. The performance of the method will be evaluated by comparative analysis with existing techniques and by conducting auditory experiments. By developing and testing an accurate and cost-effective method for personalizing HRTFs, this study can contribute to improving auditory immersion in virtual environments and suggest directions for further research.

1.1.1 Classical HRTF Measurement Method

HRTF measurements are usually performed in an anechoic chamber so that the measured HRTF does not record information about a specific space that should not be present. In the early stages of HRTF research, impulse response signals at different locations were recorded by microphones placed inside the ear. The basic principles and methods of HRTF measurements today are the same as in the early days. However, the early HRTF measurement process was more complex and the results were worse than the digital measurement techniques used today. The drawbacks of the analog measurement methods used in the early studies were mainly due to the fact that the hardware and software used were not sufficient to support such fine measurements.

Figure 1.1: Dummy head HRTF measurements in anechoic chamber, RWTH Aachen University

The team from RWTH Aachen University published HRTF measurement data for dummy in 2017. The article describes how HRTF measurements can be carried out using classical methods. Compared to earlier measurement methods, the authors used a more miniaturized device and a tighter measurement process. The measurement goal of the authors' team was to achieve reliable individual HRTF measurements in an anechoic chamber in as short a time as possible with minimal impact on the measurement itself. A new loudspeaker array was used in the experiment, and this new design allowed the device to be significantly reduced in size. As can be seen in Figure 1, the setup is a circular array of loudspeakers, which are placed along the zenith direction. The dummy head stands on a turntable and rotates with the center point of the head as the center point of the arc. During the test, each speaker in turn emits an impulse response, and the speakers built into the ear canal of the dummy head record these sounds. For each recording of the speaker array, the turntable changed the horizontal angle of the dummy's head until all horizontal angles were measured. For the collected data, some post-processing was performed, including the measurement pulse loudness was cropped to the same length, the reference transfer function was regularized, etc.
Although this method has greatly simplified the traditional HRTF measurement process, there are still many inconveniences in the actual measurement: if a real person participates in the test, the subject has to maintain a stable sitting position for a long period of time, otherwise the microphone in the ear canal will be displaced, and the body deformation will also make the sound not experience correct reflection and diffusion; and the test conditions such as anechoic chamber, microphone array and back-end control system are also demanding.

1.1.2 Numerical HRTF Simulation

In recent years, numerical calculation methods have emerged as an alternative approach to obtaining Head-Related Transfer Functions (HRTFs). This technique involves modeling the linear transformation of sound before it reaches the listener's ear canal, including spatial cues. The Boundary Element Method (BEM) is the most commonly used numerical method for HRTF calculation, which uses the Helmholtz equation to describe sound waves in a domain and transforms it into a boundary integral equation. However, this approach is based on the assumption that only surface features of the ear, head, and shoulders are relevant, and the propagation through other body parts is ignored(Katz, 2001a). Additionally, human skin has been shown to have acoustic rigidity, while hair does not.

Accurate three-dimensional geometrical shapes of the pinna, head, and torso are required for personalized HRTF calculation, and the accuracy of the calculation depends largely on the accuracy of the 3D geometrical measurements, especially in the high-frequency range. However, to ensure accurate acquisition, such measurements also require considerable costs. To reduce the cost of personalization, several approximate acquisition methods have been developed, such as physiologically-based personalized methods and subjective experiments based on a small number of measurements. Nonetheless, Zhong and Xie (2012) have pointed out that the accuracy of HRTF, especially in the high-frequency range, needs to be improved, and there is still a significant gap.

As artificial intelligence continues to advance in the field of acoustics, AI-based methods have shown great potential in improving the efficiency of obtaining personalized HRTFs.

Gebru et al [10] designed a HRTF prediction system based on deep learning, and the input parameters of this system can be measured without a professional listening room, which also reduces the cost and obtains good results. Good results were obtained.

1.2 Outline

This report is bifurcated into two main sections. The first section, i.e., Chapter 2, comprehensively explicates the process of generating personalized HRTF from Kemar scanning to deriving it in the SOFA format. Furthermore, this chapter also presents a comparative analysis of the results obtained from the proposed method with those of other conventional software models.
The second chapter encompasses a basic listening test designed to evaluate the effectiveness of the HRTF simulated through mesh scanning of the test participants. This chapter delves into the design of the test software, outlines the experimental procedure, and presents the results of the listening test.

2

Individual HRTF modeling

This chapter provides an introduction to the main focus of this study, which is the simulation workflow for personalized head-related transfer functions (HRTFs) based on individual head models. The chapter is divided into two parts: the first part describes the methodology for obtaining the head model using Heges 3D software and an iPhone XR, including pre-processing steps such as cleaning, repairing, and simplifying the scanned mesh. The second part compares different simulation methods and presents post-processing results.

It is worth noting that RWTH Aachen University has conducted professional high-resolution scanning and HRTF practical tests on the same Kemar model, enabling a comprehensive comparison between the Kemar model scanned with an iPhone and with professional 3D equipment. The actual test results can serve as the most accurate reference for comparison.

The conclusion of this chapter provides a basis for the auditory tests in the next chapter and suggests corrections. Therefore, the present study aims to explore the simulation method for personalized HRTFs based on individual head models and provide valuable insights and references for future auditory research.

2.1 Theory

In 1974, Jens Blauert first proposed the concept of head dependent transfer function. He pointed out that when the head is fixed and stationary, the sound waves emitted by the sound source reach the ears through scattering and reflection from the head, auricle, trunk, etc., and can be regarded as a linear time invariant (LTI) filter. Its characteristics can be fully described by the frequency domain transfer function of the filter. This filtering process is represented by a head related transfer function, and its corresponding time-domain form is called Head Related Impulse Response (HRIR). Specifically, the head related transfer function HRTF describes the filtering effect of the head, auricle, and torso when receiving sound from an acoustic point source at a specific location in the listener's ear canal under free field acoustic conditions. Its definition is as follows:

$$HRTF_L = HRTF_L(r, \theta, \phi, \omega, a) = \frac{P_L(r, \theta, \phi, \omega, a)}{P_0(r, \omega)} \tag{2.1}$$

$$HRTF_R = HRTF_R(r, \theta, \phi, \omega, a) = \frac{P_R(r, \theta, \phi, \omega, a)}{P_0(r, \omega)} \tag{2.2}$$

Among them, P_L and P_R are the sound pressure of the sound source at the left and right ears of the human body, P_0 is the sound pressure of the sound source at the center of the line connecting the two ear canals of the human body in the absence of the human body, and r is the distance from the sound source to the center of the head, θ is the horizontal azimuth angle of the sound source, ϕ Is the height angle of the sound source, ω Is the frequency of sound waves, α It is a morphological parameter of the human body.
Usually, there are two main ways to obtain HRTF: one is through experimental measurements in an anechoic chamber, and the other is through theoretical calculations. The methods obtained through measurement are mainly divided into two types: one is to use linear time invariance, and the other is to use deconvolution. For linear time invariant systems, when the input signal is a unit impulse response $\delta(t)$. The output $h(t)$ of the system is the impact response of the linear time invariant system, which is the transfer function of the system.

$$x(t) * h(t) = y(t) \tag{2.3}$$

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ 1, & t = 0 \end{cases} \tag{2.4}$$

Deconvolution refers to inputting any input signal $x(t)$ into a system with a system function $h(t)$ to obtain an output signal of $y(t)$. When $x(t)$ is known and $y(t)$ is measured experimentally, the system function $h(t)$ only needs to be calculated using the following formula:

$$x(t) * h(t) = y(t) \tag{2.5}$$

$$h(t) = IFFT\{FFT(y(t))/FFT(x(t))\} \tag{2.6}$$

The CIPIC database is obtained through linear time invariance measurement, using a random signal (Golay signal) as the excitation, and the autocorrelation function of the signal is a strict unit impulse response $\delta(t)$. The databases such as Listen HRTF are measured and calculated through deconvolution methods.
The essence of the method of obtaining HRTF through theoretical calculation is to physically solve the scattering and diffraction processes of sound by the head, ear, and other objects.

2.1.1 Head-related Transfer Function

The head-related transfer function (HRTF) is an acoustic transfer function that describes the distance between a point source in a free field and a specified location in the listener's ear canal, and plays an important role in creating an immersive virtual acoustic environment (VAE) for headphone or speaker playback. HRTF is highly personalised and depends on the direction and distance (near-field HRTF) Head-related impulse response (HRIR) is the time-domain representation of HRTF. All relevant acoustic information for localising real sound sources is contained in

the HRTF, i.e. ITD and ILD, and the monaural spectral factors. As each person's anatomy is different, the HRTF is unique for each individual. VAEs created using a non-personalised HRTF may have a poor listening experience, such as reduced accuracy of sound image localisation and a confusing sense of distance. For far-field VAEs, it is usually possible to adjust the sound pressure to vary with the distance of the sound, according to the inverse square law. In the near field, however, the HRTF varies significantly with distance, and this is when a separate HRTF is needed to accurately describe it.
In practice, early individualised HRTFs were obtained from acoustic measurements. This was done by placing a microphone in the subject's ear canal so that the microphone recorded the sweep signal from the different directions of the acoustics. Eventually the signals from all the different vertical and horizontal angles were collected to create a frequency response function with the input signal of the sound. Testing a high density HRTF data set is time consuming, especially for real subjects - it often means sitting motionless in an anechoic chamber for several hours. The use of sparse HRTF datasets interpolated or extrapolated with distance or direction to obtain high-density HRTF datasets is effective in reducing the number of measurement points, but still requires a large number of measurements.

2.1.2 Burton-Miller Boundary Element Method (BM-BEM)

The Burton-Miller Boundary Element Method (BM-BEM) is a numerical approach that is employed in the field of computational acoustics to solve problems concerning wave propagation and scattering. This variant of the Boundary Element Method (BEM) was introduced by Burton and Miller in 1971 and is specifically advantageous for solving exterior acoustic problems, such as scattering from objects in a free field or radiation from a vibrating surface.

BM-BEM transforms the governing equation of the acoustic problem into an integral equation on the boundary of the domain, which is then discretized to obtain a linear system of equations that can be solved numerically. The numerical solution provides the values of the acoustic pressure and/or velocity at all points in the domain.

Compared to traditional BEM, BM-BEM offers various advantages. Firstly, it employs an alternate representation of the fundamental solution which leads to a computationally efficient symmetric coefficient matrix instead of a non-symmetric one used in traditional BEM. Secondly, BM-BEM provides a stable numerical solution even for highly oscillatory kernels, which can be challenging to handle in traditional BEM.

To overcome these issues, the Burton-Miller equation combines the Helmholtz equation and its normal phase derivative equation. It can uniquely solve the full frequency band in the external region.

Mesh2HRTF uses a 3-dimensional Burton-Miller with a BEM implementation with the Multilevel Fast Multipole Method (ML-FMM) and provides add-ons for existing

cross-platform applications for pre-processing of geometric data and visualisation of results.

2.1.3 Personalized HRTF via Mesh2HRTF and COMSOL

For simulations using BEM via different software, appropriate scan meshes are required. However, the processing workflow for calculating BEM may differ.
For Mesh2HRTF simulations, an ideal 3D mesh should have a high-resolution ear shape and a relatively low-resolution head and torso part to reduce computation time and improve accuracy.
Mesh2HRTF workflow requirements:
1. A computer with at least 16GB RAM.
2. An accurate 3D mesh of the individual ear and head shape.
3. Mesh correction and simplification are performed in Blender, which is the final step in mesh processing.
4. Simulating in "NumCalc" of Mesh2HRTF under the Python environment.

Some additional free or open source software for cleaning up 3D meshes and listening to the generated SOFA files.
Sergejs D., the senior system engineer from CEVT and the supervisor of this project, has created a starter guide to Mesh2HRTF. For more detailed steps, please refer to the citation.[4]

COMSOL is a wide-used simulation software that can also be used to simulate the acoustic properties of an individual's head and torso.
COMSOL workflow requirements:
1. A computer with at least 16GB RAM.
2. An accurate 3D mesh of the individual ear and head shape.
3. Simulating in the Acoustics Module under COMSOL Multiphysics.
The detailed software operation steps (excluding hardware) are demonstrated in an article titled "Head and Torso HRTF Computation" in the Application Gallery page of COMSOL.[5]

2.2 Methods

Mesh2HRTF is an open-source project available on GitHub that provides a user-friendly package for the numerical computation of head-related transfer functions (HRTFs) for researchers and enthusiasts in the field of binaural spatial audio. The software reads the 3D human body mesh, calculates the corresponding sound fields, and produces HRTFs using NumCalc. To accommodate multiple computational platforms, Mesh2HRTF is primarily a command-line tool focused on the numerical core, which includes the 3D Burton-Miller alignment BEM and the Multilevel Fast Multipole Method (ML-FMM) implementation. It also offers add-ons for existing cross-platform applications to pre-process geometric data and visualize results.
For the Pressure Acoustics Boundary Element interface in Acoustics Module, COM-

SOL, no more specific boundary element method is described.
Before establishing a 3D grid, the original HRTF data will be pre-processed as follows:
(1) Perform minimum phase processing on the original HRIRs data in the database to remove delay information;
(2) Transform the removed delayed HRIRs data obtained in (1) into HRTFs through Fast Fourier Transform (FFT);
(3) Calculate the logarithmic domain form of HRTFs, log-HRTFs;
(4) Perform a mean removal operation on HRTFs in each logarithmic domain;

2.2.1 More Things Need to Know

SOFA file: The spatially oriented format for acoustics (SOFA) aims at representing spatial data in a general way, allowing to store not only HRTFs but also more complex data, e.g., directional room impulse responses (DRIRs) measured with a multichannel microphone array excited by a loudspeaker array. In order to simplify the adaption of SOFA for various applications, examples of implementation of the format specifications are provided together with a collection of exemplary data sets converted to SOFA. In this project, the personalised HRTF obtained from the Mesh2HRTF simulation will be recorded in a file in SOFA format.

2.3 Acquisition of 3D Meshes.

Research by Mesh2HRTF developers suggests that an ideal 3D mesh should have approximately 40,000 elements with lengths between 0.5 mm and 10 mm.[3] COMSOL requires a 3D mesh of the head and torso. Typically, the mesh around the ear is detailed, while the head and torso areas are simplified to speed up calculations and improve accuracy. Mesh2HRTF and COMSOL have similar requirements in this regard. [5]
The goal of scanning is to obtain a total mesh that includes a high-resolution ear and low-resolution head and torso. The total mesh needs to be scanned carefully with the highest resolution around the ear's contour and relatively quickly with mediocre resolution around the head and torso. In this study, the Heges 3D v1.6 software and a standard iPhone XR using iOS 15 were used to perform all scans.

2.3.1 Selection of Software and Hardware

Various methods can be employed to obtain a precise 3D mesh of the individual ear and head shape. One approach is to use professional 3D scanners, including portable laser scanners commonly used for traditional HRTF simulation. Alternatively, mobile devices with laser scanning or structured light scanning capabilities can also perform 3D scanning, with flagship smartphones in the Android and IOS camps offering such features[1]. While laser scanning can be used for larger objects, structured light scanning, such as the Face ID on the iPhone, is considered

to have sufficient resolution (0.5mm) for the boundary element method required for customized HRTF[2].

Before scanning, some software and hardware settings can also be adjusted. The Heges app can support sharing the user interface to another iOS device, which is very helpful for operators. This is particularly important when scanning the cochlea, as some imaging angles cannot be observed well through the posture of taking a selfie. It is also important to set the size and coordinate axis of the mesh in advance, as some mesh processing software cannot easily modify the size. Additionally, using the finest resolution when scanning the head and torso may cause the device to crash due to insufficient cache or memory. It is reasonable to lower the resolution to 1 mm or 2 mm, which has been proven to be a reasonable and effective operation. Although the mesh will be further down-sampled to even lower resolutions during processing, starting with a lower resolution that meets the actual requirements (approximately 2 mm to 5 mm) will result in a smoother surface when merging with high-resolution cochlear meshes. Cleaning up rough surfaces can be very time-consuming and laborious.

2.3.2 Preparation and Strategy for Scanning

To optimize the conditions for 3D scanning, several steps can be taken. First, it is important to expose the ear and skin surface as much as possible by covering or removing the hair. Hair can distort the interaural time difference (ITD) and other aspects of the head-related transfer function (HRTF) that are important for accurate scanning. It is recommended to use a tight-fitting swim cap or wig cap to compress and tidy the hair, and to clean any beard to minimize the amount of 3D data needed to clean up the remaining hair. This will improve the accuracy of head boundaries. It is also important to avoid wearing additional items such as glasses and earrings since they reflect light and have small details that cannot be accurately scanned.

Reflective make-up should also be avoided as matte surfaces are preferred for most 3D scanners. Clean skin is required to ensure accurate scanning. It should be noted that the actual simulation calculations do not require such high resolution, but slightly higher resolution than the simulation has better fault tolerance in the subsequent mesh processing.

The scanning strategy should begin with scanning the left and right ears in detail, followed by the face, and finally the neck and shoulders. Since it is common to obtain inadequate meshes (discontinuous surfaces or spiky features), each part should be scanned multiple times, with at least two qualified meshes obtained for each part. The scanning time is approximately 30 minutes for those familiar with the process, which is considered a significant advantage over traditional methods using professional laser scanners.

The targets of the scanning are also listed in Table 2.1:

Body part	Resolution(mm)	Scanning times	Scanning object
Ear	0.5	Multiple	Kemar
Head	1	Multiple	Kemar
Torso	2	Single	Kemar

Table 2.1: Desired mesh quality for different body parts

2.3.3 Scanning Process

Using an iPhone or even a professional scanner to scan the ear can pose challenges. The scanning method is based on the principle that Face ID performs best when the object is about 30-50 cm away. Therefore, capturing an image of a person in a selfie form is the ideal choice. This ensures that the person being scanned is in the best position for resolution, while allowing the operator to maintain an appropriate distance and avoid being captured in the scan.

Starting from the back of the head can be a good option because any accumulated error in this area is less noticeable.

To help the 3D scanner locate the ear accurately, it is recommended to add a geometric reference in the scanning scene. The reference object should be simple and have a regular shape, which helps the scanning software locate it and facilitates future quality checks. It should also be large enough to be viewed from multiple angles to avoid losing track. In this study, we chose a frame made of a toy building kit as a reference object, which can be seen in the figure2.1.

If the entire scanning process includes the reference object, it may not be possible to achieve excellent mesh cleaning at the connection between the head and the reference object (although this seems to have little effect from the simulation results). Therefore, assuming no reference object is included is also considered feasible. As

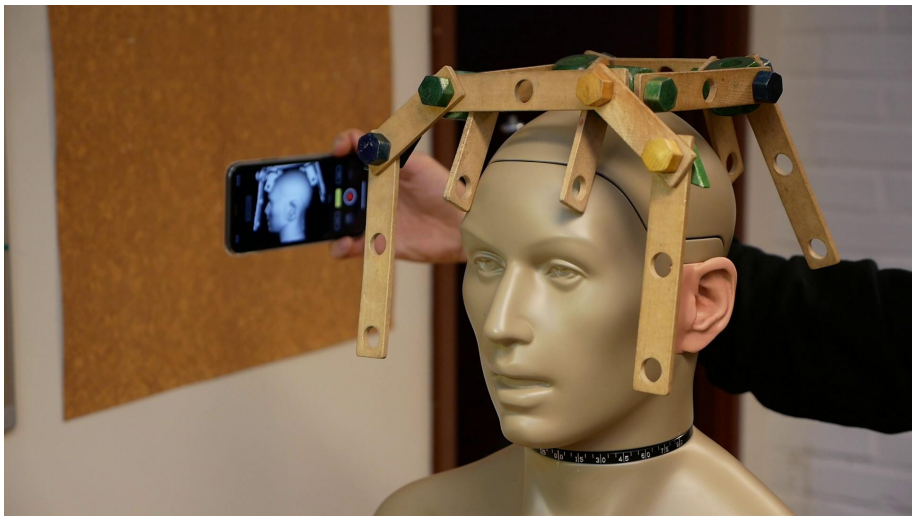

Figure 2.1: Scanning of Kemar in this thesis

can be seen from the figure above, the operator is not recorded in the scanner, while Kemar is clearly recorded.

To better simulate real-world scenarios, Kemar was placed on another tester's lap to mimic tiny unintentional vibrations of the human body. In practice, having an additional person between the scanner operator and Kemar made it helpful to plan the scanning route in advance. Also, due to the increased distance, it became somewhat difficult to scan the entire head in one go. As a result, many scans were taken before the desired head scan was obtained.Without a backup, many defects like the one shown in the Figure 2.2 are likely to be overlooked.

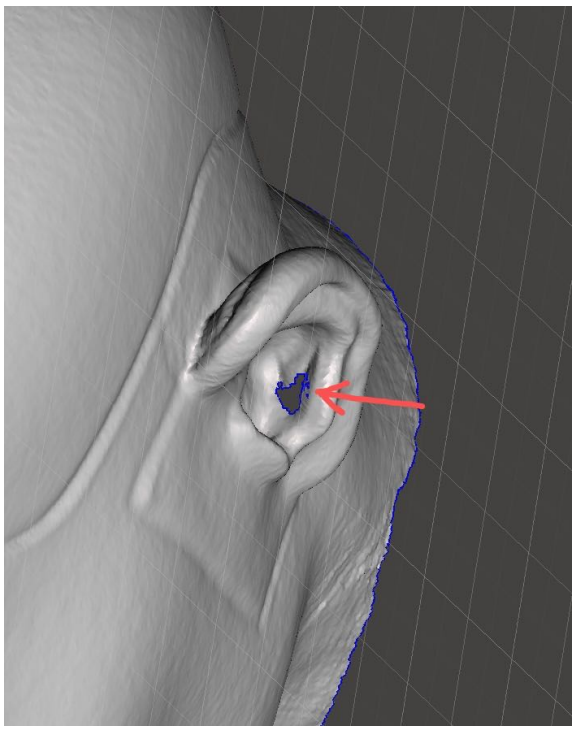

Figure 2.2: Bad scan of left ear (hole appears)

During the scanning process, it is also necessary to check and compare with the actual shape in a timely manner. In addition to obvious scanning errors such as grids containing holes and wrinkles, as shown in Figure 2.3, when scanning the ear shape in the direction of the arrow, the resulting mesh structure is often distorted.

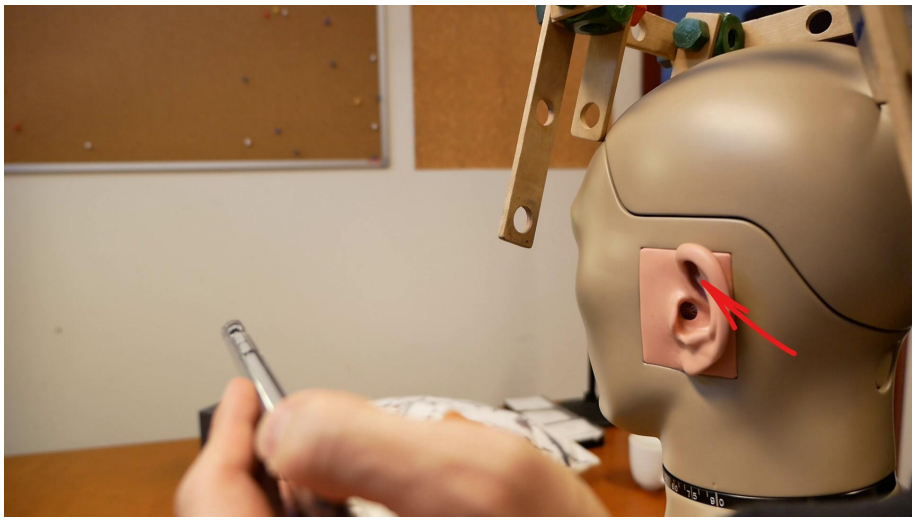

Figure 2.3: Kemar dummy difficult to scan area

A long scan of the Figure 2.3 area is needed to ensure that the ear crease is sufficiently deep and narrow.

This is a three-part scanning process, with the head scan performed first, followed by detailed scans of the right and left ear and finally the torso. Attempting to capture all the details in one scan is challenging due to the high demand it places on the phone's processing and storage capacity. Additionally, it would result in a low fault tolerance rate and costly re-scans. It is important to note that conducting multiple scans will improve the chances of a successful scan in the later stages. It is also necessary to clear the storage space on phone in advance. A single scan of Kemar's torso for this project can be as large as 700 MB, which means that a full scan process can take up to 10 GB of storage

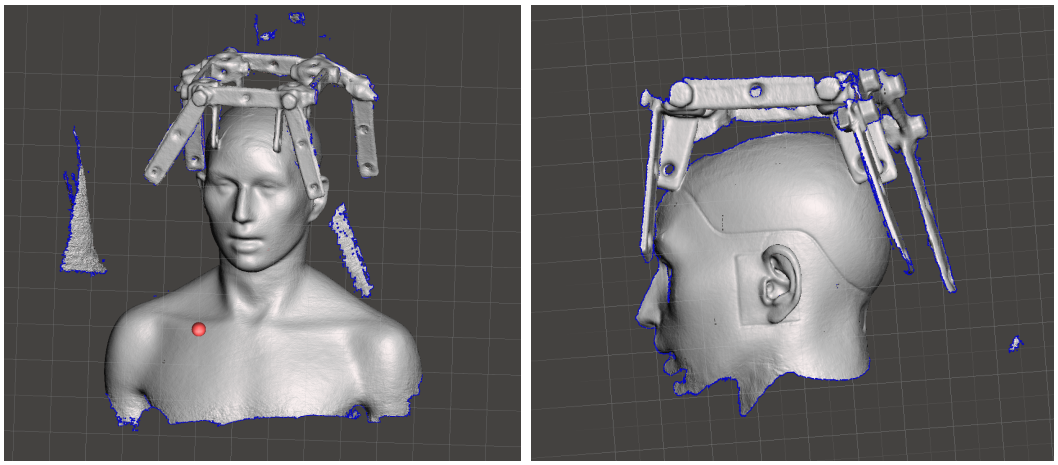

Figure 2.4: Original scan of Kemar's body and left ear

Some of the results of the scan are shown in the Figure 2.4.As can be seen, the scan results often contain many isolated surfaces, as well as some uneven surfaces and

even holes. Therefore, further mesh cleaning is very necessary.

2.3.4 Cleaning and Merging of Meshes

After the initial scan, a high-resolution ear mesh and a relatively low-resolution head and torso mesh are applied to specific software for cleaning and merging. The software used in this project includes Blender, Meshmixer, and Meshlab. It should be noted that these are not the only available software, but free and open-source software was prioritized to make the experiment more universal and valuable.

First, some basic automatic cleaning is necessary. For simulation, the 3D mesh must be an airtight fluid shell without isolated, overlapping geometric shapes. In the scanned mesh, there will always be sharp protrusions and disconnected free bodies, which will cause simulation errors. Therefore, some basic automatic cleaning is necessary before cutting and merging meshes.

It should be noted that for surfaces that are difficult to completely remove or paint, such as hair and eyebrows, down-sampling should be performed in advance and then the mesh should be cleaned. This is to prevent mesh collapse during painting or the formation of sharp corners between meshes, which is a common problem when dealing with dense meshes.

In cleaning operations, taking Meshlab 2021.10, Blender 3.0.0 and Meshmixer 3.5.474 used in this peoject:

Step 1. Delete unnecessary geometries such as reference objects: use 'Edit/Select Faces and Vertices inside polyline area' to select the objects that need to be saved. Then press 'I' to invert the selection and press 'DELETE' to delete. Note that the selection under this method is perspective-based, which may cause some unexpected consequences: when selecting isolated surfaces near the ears, some useful surfaces of the neck or back of the head may also be automatically selected and deleted due to perspective. So be extra careful.
Step 2. Delete all isolated surfaces: use 'Edit/Select Connected Components in a region' and drag it to the main mesh for selection. Then press 'I' to invert the selection and press 'DELETE' to delete. Step 3. Save and 'Reload all layers'. Save the preliminary modified mesh.
Step 4. Use 'Filters/Remeshing, Simplification and Reconstruction/Simplification: Quadric Edge Collapse Decimation' to simplify the mesh.
Step 5. Use 'Filters/Remeshing, Simplification and Reconstruction/Surface Reconstruction: Screened Poisson' to further down-sample the head and torso. Set 'Reconstruction Depth = 12'. This operation will reduce the accuracy of the mesh, so it cannot be performed on the ears.The other function of this step is to close the grid below the shoulders, as you can see from Figure 2.4, the torso mesh is not closed.The picture on the right shows the placement of the microphone during the actual HRTF measurement.
Step 6. Please pay special attention to the mesh at the entrance of the ear canal.

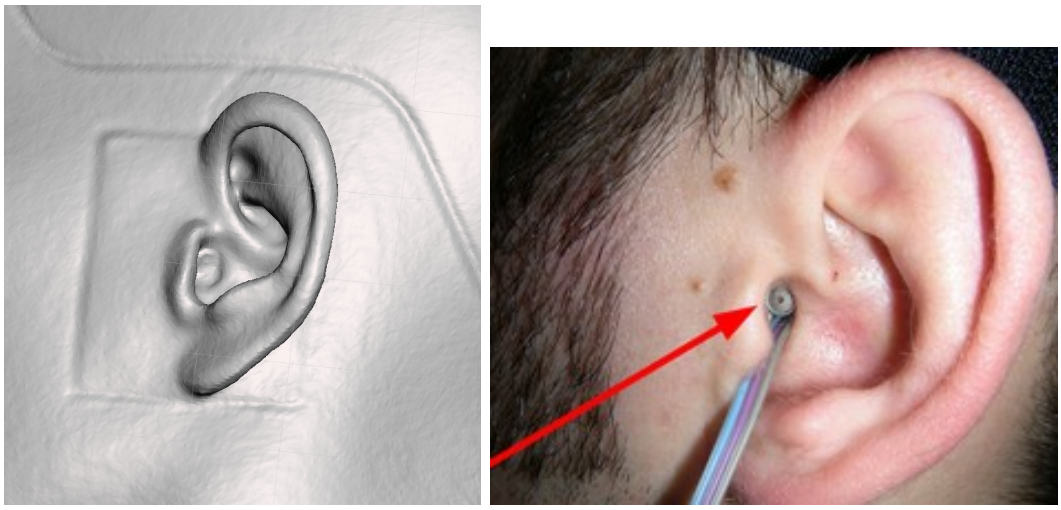

Figure 2.5: Ideal external ear canal surface and ear canal in reality[20].

Sometimes, many sharp geometries or holes may form in the ear canal because the structure deep inside cannot be detected. At this point, some automatic repair algorithms are needed to fill in. This is a tricky detail to consider because the parameter points for Mesh2HRTF calculation are at the entrance of the outer ear canal (which is realistic because the microphone is almost impossible to be placed at the eardrum). Ideally, 'filling the ear canal entrance' should be the default. In this experiment, the appropriate processing in Meshlab was not found, so 'Smooth' in Blender's 'Sculpt Mode' was selected to smooth the geometry of this part.The ideal external auditory canal surface is shown in Figure 2.5.

Please note that for the head and torso, the priority of downsampling can be increased appropriately. In practice, the torso and head files are too large, causing the software to often lag or even crash when working with such meshes. Simplifying these meshes will significantly improve the efficiency and success of the process. Saving frequently is also a forced habit, as Meshlab is very prone to jamming or crashing when performing some automated algorithms (especially noise reduction). If not saved in time, this means that the job needs to be started all over again.

The next step is the mesh merging. The mesh of the left and right ears that have been clipped, and the mesh of the head that has been subtracted from the ear part will be used.This is followed by the merging of the meshes which can be done in a variety of ways, with different software offering many options. This stage of the process varies from person to person. Finally it is necessary to check that the merged mesh joints are smooth.

The merging process in this project is as follows: Step 7: Down-sampling the mesh. Use 'Filters/Remeshing, Simplification and Reconstruction/Surface Reconstruction: Screened Poisson' to down-sample the torso, which will effectively reduce the size of the file. In this project, the number of meshes for the torso and head is controlled to be no more than 15,000.

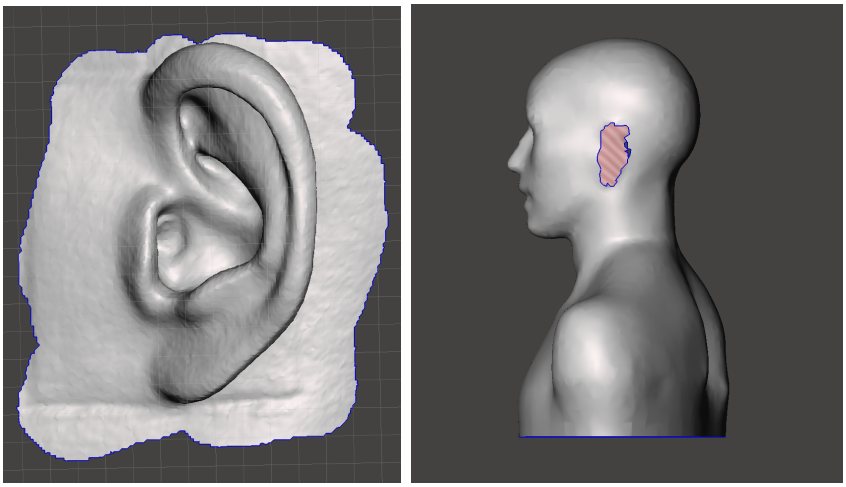

Figure 2.6: Cropped left ear (left) and torso (right) meshes

Step 8: Cut the head and ear meshes to the size shown in Figure 2.6. Check that all meshes should contain partially overlapping surfaces. Cutting too much surface will cause the mesh to be misaligned.
Step 9: Merge the ears and torso using the 'Point-based gluing' method in the 'Align' tool.As can be seen from Figure 2.7, the additional ear grid can greatly assist in the positioning of the merge.

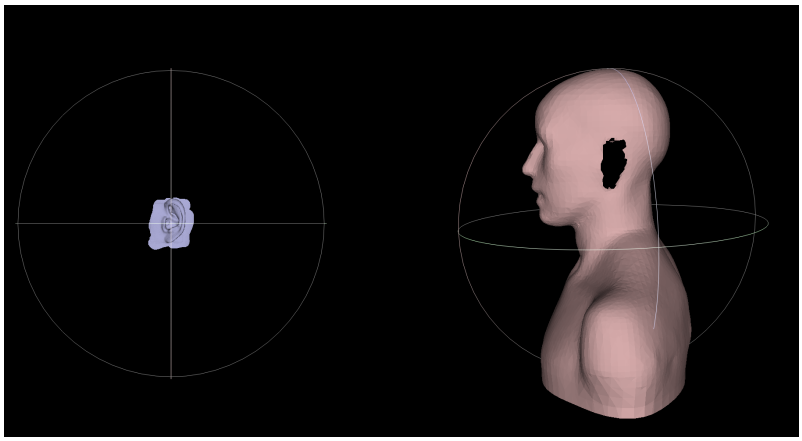

Figure 2.7: Merging of meshes using Point-based gluing

Step 10: Save the merged mesh and reload all layers. Step 11: Use 'Surface Reconstruction: Screened Poisson' to create a new mesh around the scanned data. Set 'Pre-Clean' to 'Yes', set "Merge all visible layers" to Yes and adjust 'Reconstruction Depth = 12'.
Step 12: Use 'Smooth' in 'Script Mode' under Blender to smooth the joints of merged meshes. Radius' and 'Strength' in the Smooth function can be adjusted to suit your needs. The settings for this project are 'Radius = 45 px' and 'Strength = 0.4'. As can be seen in Figure 2.8, the smoothed mesh is much flatter and more consistent with the actual Kemar surface in the area where the ear meets the head.

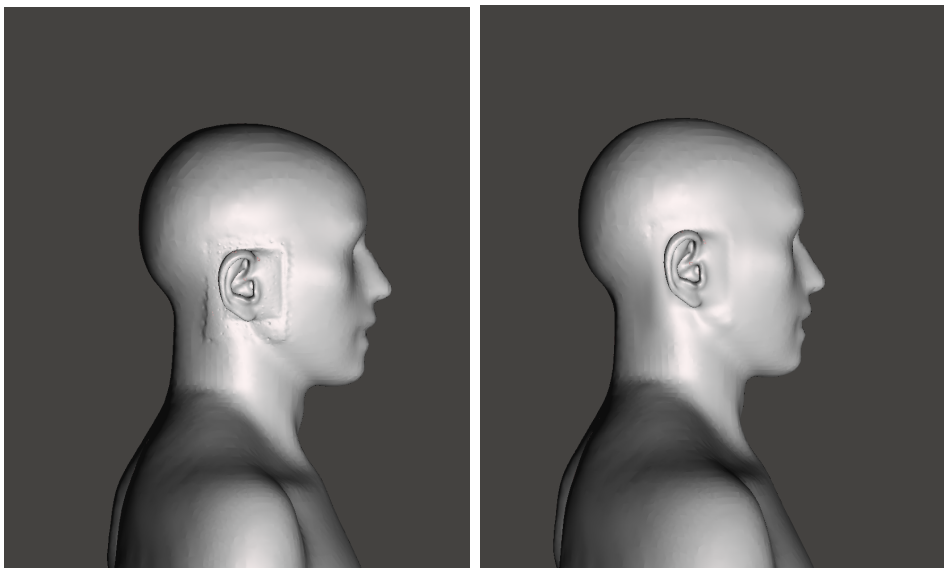

Figure 2.8: Merged Kemar mesh (Left) and Kemar mesh after smoothing (Right)

Step 13: Use 'Make Solid' in MeshMixer to ensure that the overall mesh is airtight. This is an option because in the later stages of the simulation process, the airtightness needs to be ensured, otherwise it will indicate that the 'computational data does not regress'.

The above process is how the mesh is processed and the prepared mesh can be used in the Mesh2HRTF and COMSOL simulations. Please save most of the mesh files and make a note of them, as this will greatly facilitate later modifications as required by the simulation phase.

For the overall meshing process, the time consumed is dependent on the quality of the scanned mesh. In practice, poor meshes can often take up to a day to fix, compared to a good mesh that can be fully processed in just one hour.

2.4 Simulation of Individual HRTF in Mesh2HRTF

This section describes the procedure for placing the prepared meshes into the different workflows.

2.4.1 Pre-processing in Mesh2HRTF

Pre-processing phase is carried out to complete all the parameter settings in Mesh2HRTF. The rest of the simulation process will either complete automatically or stop with an error. The HRTFs for the left and right ear are generated separately and these HRTFs can then be combined to obtain the final personalised HRTF.

Timon et al. suggest gradually reducing the resolution of the mesh as the distance to the ear close to the HRTF increases to reduce the amount of operations. This

approach is also optimised by considering the curvature of the geometry. The resulting graded meshes allow for faster simulations in the HRTF with equal or better accuracy than previous work.

In Mesh2HRTF there is a component called 'hrtf mesh grading', which is used to optimise the 3D model simulated by Mesh2HRTF. In order to obtain maximum efficiency, the mesh used to simulate the left ear contains only fine details on the left side, while the right side is significantly simplified and vice versa.

The pre-processing of Mesh2HRTF includes the optimisation of the mesh and the setting of other parameters.
Step 1: The optimisation of the mesh is first carried out in blender by importing the full resolution mesh into the '3d Model uniform.blend' example Blender file. This example file will be used twice, this time to determine the spatial position of the 3D model. This is done so that the centre of the head is at the origin, the face is in the positive direction of the X-axis and the left and right ear canals are crossed by the Y-axis.
Step 2: The 3D model processed in Step 1 is placed into the 'hrtf mesh grading Windows Exe' folder and run. The output results in two meshes. They are '3Dmesh graded left.ply' and '3Dmesh graded right.ply'.' The results of the '3D mesh graded left' run are shown in Figure 2.9. It can be seen that the left ear of the optimised model is complete, while the otherwise dense mesh of the right ear has been extensively simplified. The other file runs with the opposite result.

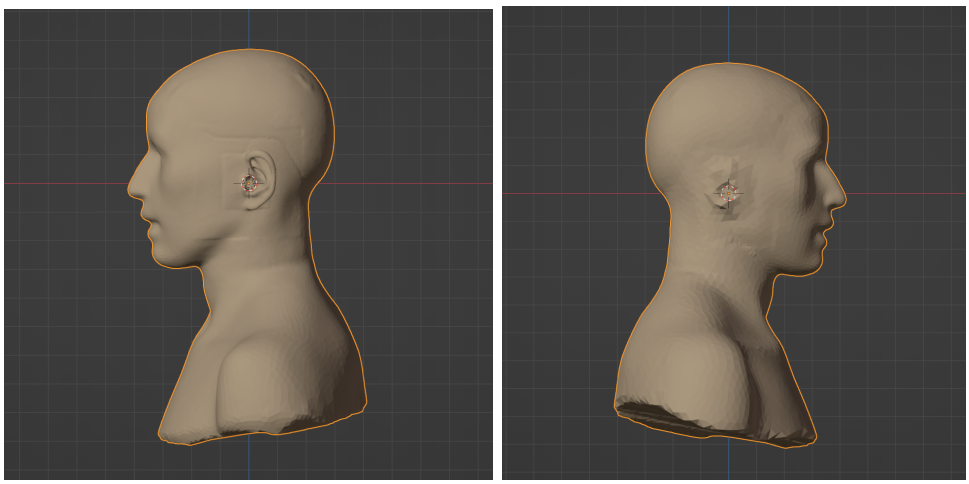

Figure 2.9: Comparison of left (Left) and right ears (Right) after optimisation of mesh

Step 3: The two exported meshes are imported again into the '3d Model uniform.blend' file. This will set different material properties for all surfaces. There are three material properties, 'Skin', 'Left ear' and 'Right ear'. The 'Skin' contains the properties of human skin, while the 'Left ear' and 'Right ear' materials represent the properties of the blocking ear canal microphone. 'Skin' was set to the vast majority of the surface, with the 'Left ear' and 'Right ear' finding only a triangle selected

as the most representative. In this project, the 'Left ear' and 'Right ear' materials were selected in the same triangle as shown in Figure **??** for the Y-axis shot into the ear.

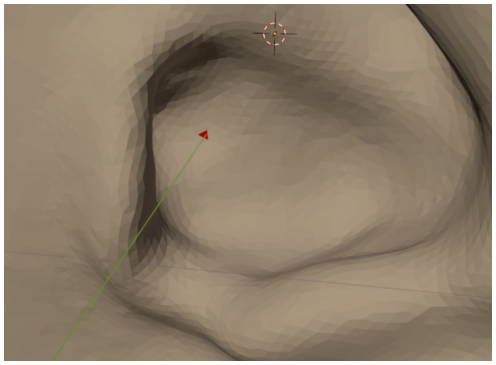

Figure 2.10: Default Vibrating Element

Step 4: Use the Python console in Blender to export the final mesh.The Python console can be found in Sergejs' Beginner's Tutorial. The code for this project will be shown in Appendix B.

Some other settings:
Step 5: Adjusting the parameters of 'EvaluationGrids'. The HRTF data set is assumed to be one in a spherical space, with different points on the surface of the sphere corresponding to the amplitude-frequency characteristics of the sound emitted from that direction to the ear canal. Each point contains a set of impulse responses for both ears. In Mesh2Input's EvaluationGrids, the location and density of the points sampled in this sphere can be set. In this paper, two formats of Evaluation Grids are set, one following the ARI HRTF database and the other optimised according to the Kemar HRTF measurements provided by the RWTH Aachen University. Points containing the same results as the measurements are set in the second Evaluation Grid. ARI data includes full azimuthal space (0° to 360°) and elevation angles from -30° to +80°,the resolution of the frontal space in the horizontal plane is 2.5° 1550 points in total. The customised data set contains full azimuthal space (0° to 360°) and elevation angles from -90° to +90°,the resolution of the frontal space in the horizontal plane is 1° and 6220 points were selected for the surface.

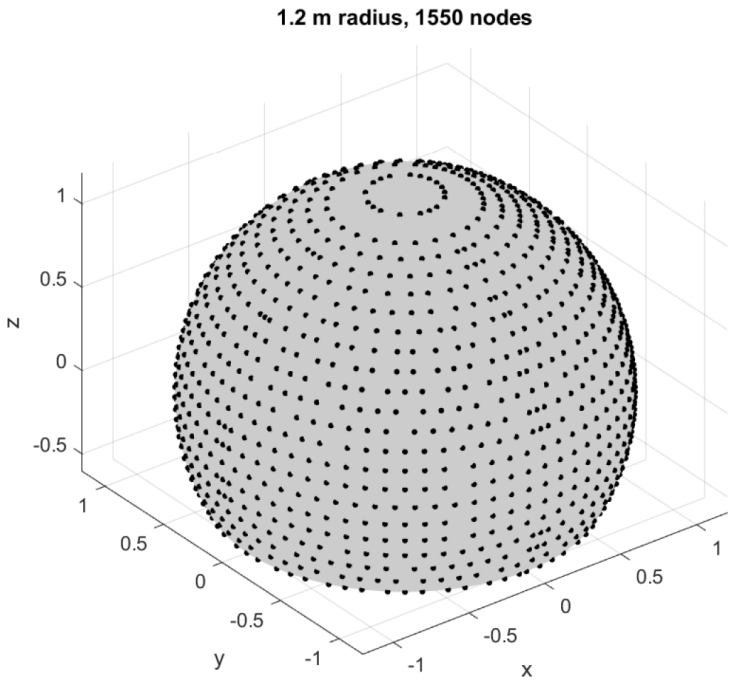

Figure 2.11: Schematic diagram of the spatial distribution of HRTF collection points under the ARI standard

2.4.2 NumCalc Simulation

After the Blender project has been exported, 2 project folders (for the left and right HRTF side) will be found, containing the "info.txt" file and other files and folders. Move this file to the mesh2hrtf-tools folder and run 'NumCalcManager.py'.
The result of the operation is a personalised HRTF SOFA file for the left and right ear respectively. In this project, this operation usually takes between 8 and 10 hours, depending on the frequency range of the HRTF previously set and the density of the Evaluation Grids.

2.4.3 HRTF SOFA File Generation

After the simulations in the previous section, the HRTFs for the left and right ears have been generated. The only step left is to merge the HRTFs of the left and right ears to obtain the final Kemar HRTF.

It is really easy to complete this step, Sergejs provides the 'finalize hrtf simulation.py' script in the beginner's tutorial. This script can automatically synthesize two HRTFs and get the final PDF.

In addition, some images containing HRTF are also provided as shown in Figure 2.12. This is a plot of the HRTF amplitude and frequency characteristics at ear level, which can help to check the performance more visually.

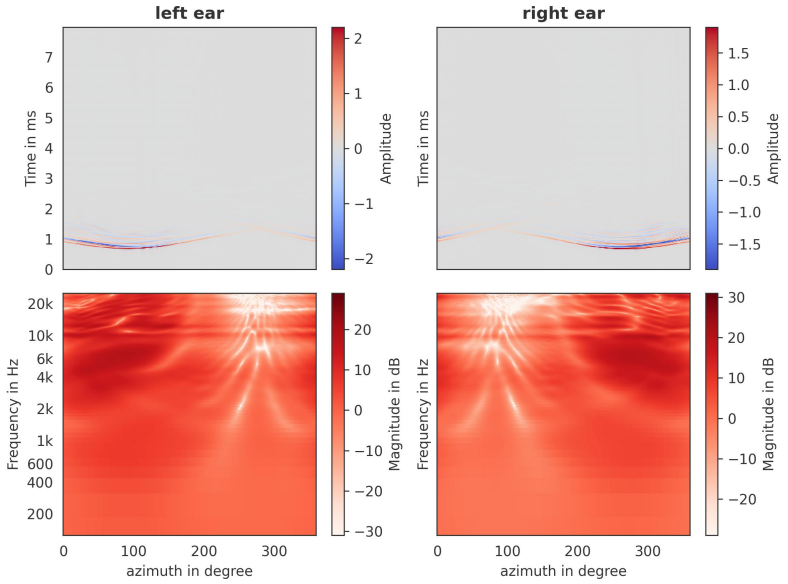

Figure 2.12: Kemar Mesh2HRTF Results

2.5 Simulation of Individual HRTF in Comsol

In COMSOL Multiphysics, the boundary element interface uses the boundary element method (BEM) to model acoustic problems via pressure acoustics. This interface is quite accurate for HRTF analysis, as the HRTF model represents a purely radiative problem in a free field. This simulation also does not require additional parameter settings and computing performance.

Please note that COMSOL's HRTF simulation has some limitations. For example, in the frequency band above 8000 Hz, the BEM module will show that the simulation results do not converge. Therefore, the COMSOL simulation data is used in this paper only as a reference for performance testing of 3D meshes and comparison of different simulation software, and all calculation results will not be used in further experiments.

2.6 Results

In the data comparison session, all the meshes that were used for testing are listed in Table 2.2.

Model from Aachen is downloaded from ITA HRTF-database which scanned via a high resolution laser scanner.
The original Aachen model included the entire surface of Kemar, but in this project, to maintain consistency of the model, the part of the Aachen model below the shoulders was cut off as shown the right picture in Figure 2.13. All low-resolution models and high-resolution model based on iPhone scans have a same out-looking, and the

scanner

Source of model	Scanning devices	Mesh quality
Model from Aachen	Structure light 3D scanner	70k surfaces
High-resolution model	iPhone XR	900k surfaces
Low-resolution model (COMSOL)	iPhone XR	20k surfaces
Low-resolution model (Mesh2hrtf)	iPhone XR	15k surfaces

Table 2.2: The model list for all simulations used in this Chapter 2

only difference is the number of meshes. If the full potential of iPhone's performance is unleashed, the highest resolution model can reach 900,000 triangles, which is higher than the resolution of professional scanning software. As described earlier in the mesh processing method, these iPhone-based meshes were cropped and stitched in post-processing.

As can be seen in Figure 2.13, the surface scanned with a cell phone still has some unevenness, while the surface from a professional scanner is much smoother. This means that in practical operation, there is always a potential risk at the joints of the mesh.

In the context of COMSOL and Mesh2HRTF, the low-resolution meshes are handled differently. As mentioned earlier, the Mesh2HRTF component includes an optimization mechanism that simplifies the head and shoulder meshes to a greater extent, while retaining the ear meshes. In contrast, COMSOL's low-resolution bar mesh is globally simplified by software from MeshLab, which uniformly simplifies all surfaces to a single resolution.

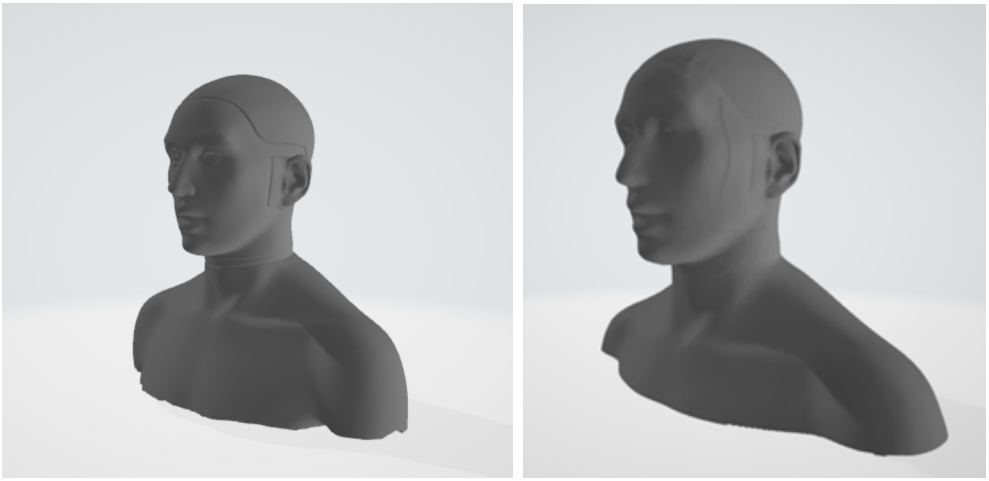

Figure 2.13: The different models used in comparison: Model from Aachen(Left) and High resolution model (Right)

2.6.1 COMSOL Simulation Results

In this project, the upper frequency limit in COMSOL simulations was set to 8250 Hz for low-resolution model. When this limit is exceeded, COMSOL reports that the calculation will not conform to linear regression. However, the frequency limit is increased to 8700 Hz and 8950 Hz, respectively, when using high-resolution model and the Aachen model. This suggests that the quality of the model has a significant impact on the performance of the simulation for the acoustic boundary element module in COMSOL.

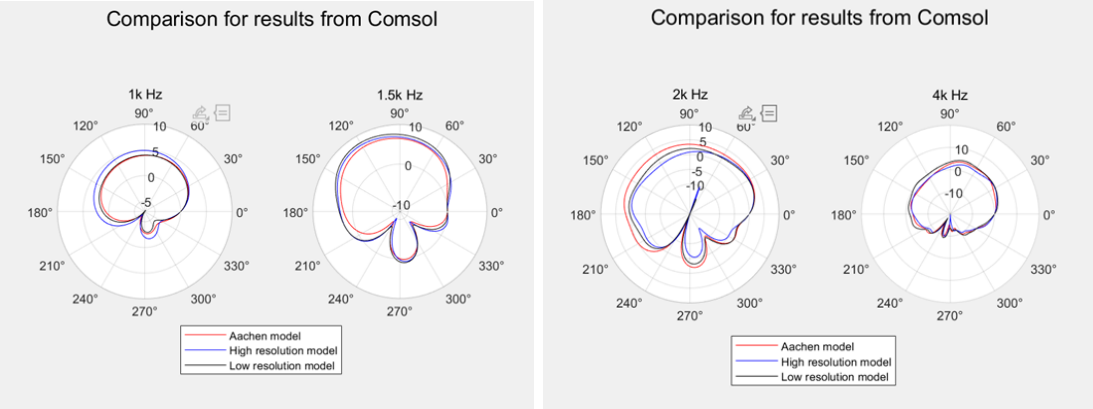

Figure 2.14: The low frequency range radiation pattern for different models in COMSOL

Figure 2.14 illustrates the HRTF values calculated for various models positioned in the horizontal plane (xy-plane). The HRTF has been normalized to 0 dB in the frontal direction (polar angle $\theta = 0$). The figure consists of four subplots showing the HRTF performance calculated for each model at different frequencies. By analyzing the presented plots, it can be inferred that the differences between the HRTF values of each model below 4K Hz are relatively small, usually within 5 dB. This error is similar to the computational results presented in the COMSOL example.

However, the outcomes are not entirely satisfactory for the higher frequency ranges that have not been addressed in the example paper. The presented Figure 2.15 illustrates the radiation patterns of three models at 6k and 8k Hz. When compared at 6k Hz, the amplitude distributions of the models appear similar in various directions, but begin to exhibit errors of up to 10 dB in some locations. Notably, this difference becomes noticeable at the position directly opposite the weaker side of the ear (around 270 degrees). At 8k Hz, the errors start to become significant. For the vicinity of the 270 degree direction, there is no similarity in the amplitude characteristics, and the errors are generally above 20 dB. For other positions that were expected to perform better, the errors are around 10 dB.Furthermore, at 8k Hz, the high-resolution models exhibit greater consistency with the results of the Aachen scanning model.

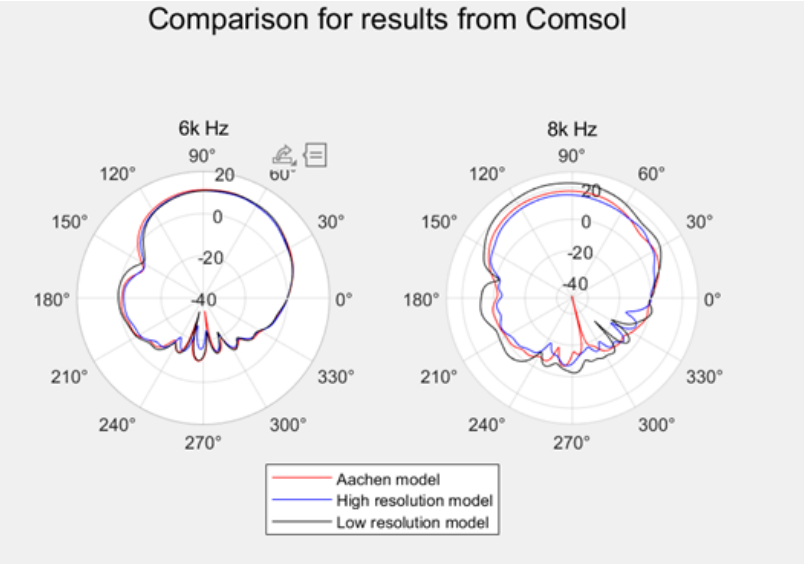

Figure 2.15: The higher frequency range radiation pattern for different models in COMSOL

Assuming the model obtained from the laser scanner used by Aachen RWTH as the ideal model, the simulation results and the amplitude distribution of the low-resolution model obtained from the iPhone scanning are similar to those of the ideal model up to 6K Hz. However, at 8K Hz, the amplitude distribution of the low-resolution model is dissimilar to that of the ideal model. On the other hand, the high-resolution model obtained from the iPhone scanning is still similar to the amplitude-frequency distribution of the ideal model at 8K Hz, and the error performance in comparison at the same frequency is lower than that of the low-resolution model. The errors between all models increase with an increase in frequency.

The above simulations demonstrate that, in BEM simulations using COMSOL, a mesh generated from iPhone scanning can replicate the personalized HRTF simulations mentioned in the example file. It was found that 8000 Hz is the highest frequency range that can be computed with COMSOL. This implies that personalized HRTF obtained from COMSOL may be difficult to apply in auditory experiments due to the limitation of frequency range. From the simulation results, higher resolution models will maintain more accurate results in relatively higher frequency ranges.

2.6.2 Mesh2hrtf Simulation Results

For the Mesh2hrtf results, a special MATLAB plugin called SOFA API for Matlab and Octave version 1.1.3 is required to process the SOFA files. This plugin loads the analysis and renders the HRTFed audio using SOFA files. This module is also used in the listening test later on.

The sources, resolutions and scanning devices of all HRTFs in this comparison are shown in Table 2.3.All HRTFs used for comparison were horizontal with respect to

the ear-nose horizontal plane.

HRTF	Approach	Resolution	3D mesh sources
ITA HRTF-database	Measurement	-	-
Simulation	Mesh2HRTF	-	Laser scanner
Simulation	Mesh2HRTF	High	iPhone
Simulation	Mesh2HRTF	Low	iPhone

Table 2.3: Acquisition source, quality and scanning method of the HRTF used

Figure 2.16 shows the left HRTF results of different models simulated at 30-16000Hz. It can be observed that they are generally consistent from a global perspective, and their amplitude-frequency distributions are very similar. For frequencies below 1kHz, the addition of HRTF does not bring any significant changes to the sound at all angles. However, around 2kHz, some reflections can be observed at around 50 degrees, which can be attributed to the shoulders. Around 8kHz, higher sound pressure levels are received at 0-180 degrees, while many dips are observed at 180-360 degrees, indicating that the sound from specific directions at these frequencies reaching the ear canal will be significantly lower.

Around 10kHz, differences between the simulation and measurement values can be noticed: the amplitude-frequency distributions of all simulated values are somewhat fragmented. This could be due to insufficient smoothness of the fine surface transitions in the mesh preparation.

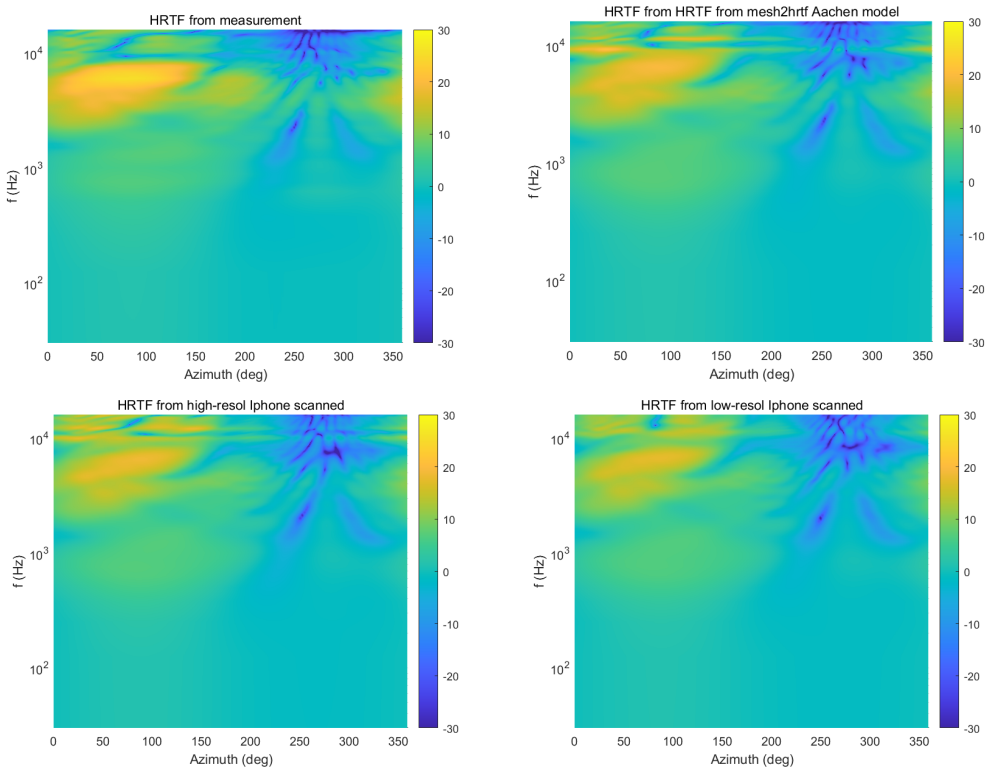

Figure 2.16: Simulated HRTF results from different models by Mesh2hrtf

The Error Plotting is a graphical representation obtained by subtracting the simulated HRTF from the measured HRTF. Specific points on the plot are labeled to facilitate observation. The results show that, in the frequency range up to 2 kHz, the differences between the simulated and measured values are negligible for all models. However, between 4 kHz and 6 kHz, the simulated values for all models are consistently lower than the measured values.

Differences between the simulated and measured values become increasingly evident above 6 kHz. This can be attributed to several factors. Firstly, misalignment along the horizontal axis, possibly due to initial differences in the horizontal plane of the Kemar mannequin or displacement of surfaces during the scanning process, results in significant vertical errors in the error plotting. Secondly, errors in the scanning grid are amplified at high frequencies, which is difficult to avoid.It is worth mentioning that, in order to simulate the actual scanning process, the iPhone was placed on the subject's lap to introduce some jitter, which undoubtedly adds some uncertainty to the mesh.

This can be observed from the error plotting, where the HRTF generated based on the high-resolution model has smaller errors compared to the other two models in the frequency range up to 4 kHz. In the frequency range of 4 kHz to 8 kHz, the HRTF generated based on the professional laser scanner has more correlated results. However, in the frequency range above 10 kHz, all simulation results have significant differences compared to the measured values.

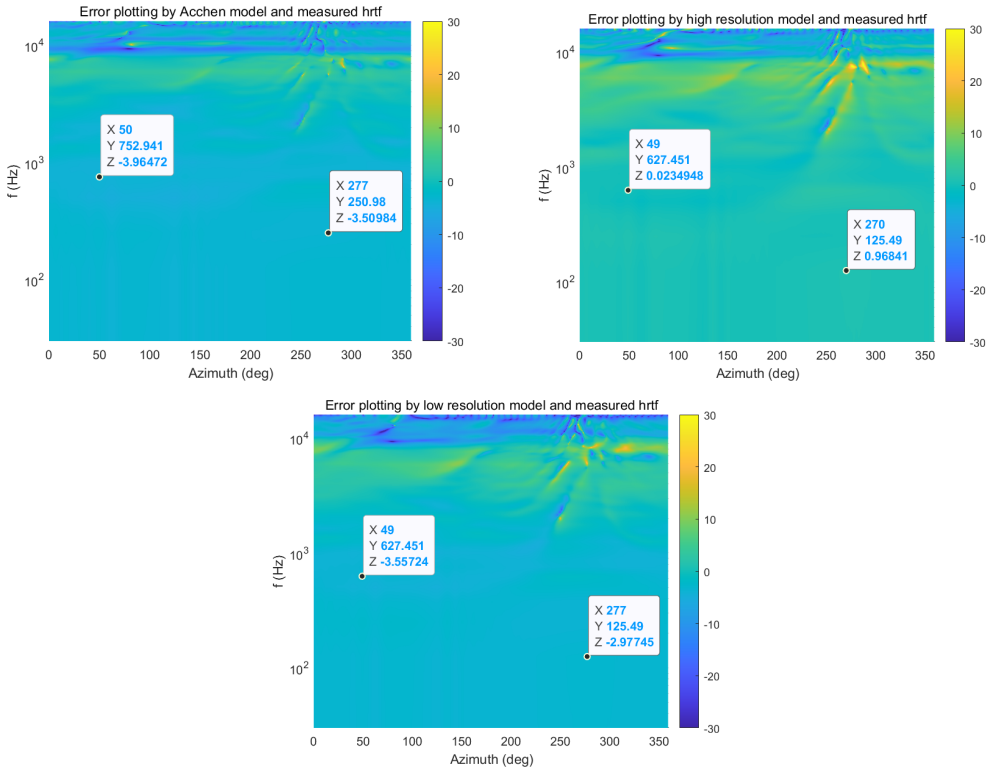

Figure 2.17: Error plotting for different models

2.6.3 Comparison of Simulation Results

For the comparison between COMSOL and HRTF, both were compared on the HRTF horizontal plane, and the radiation pattern was used in this section.

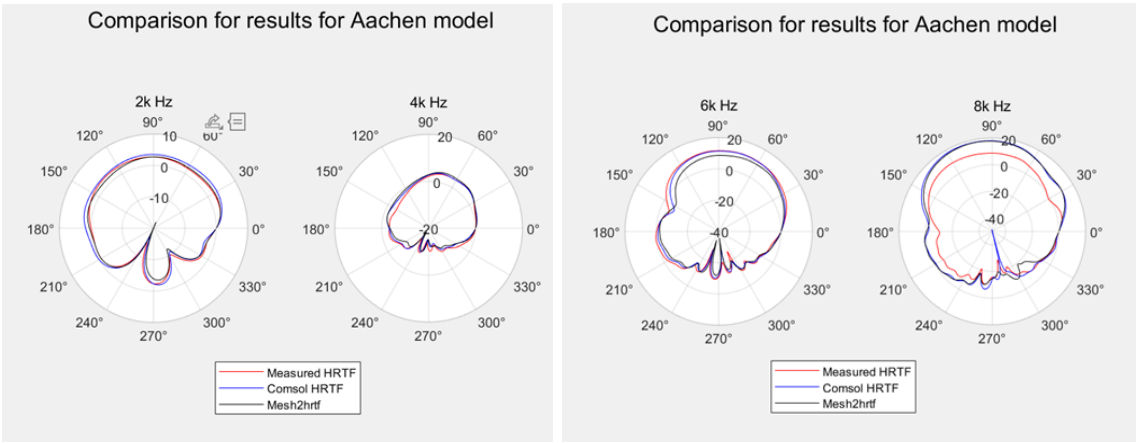

Figure 2.18: The radiation pattern for different simulation tools

Figure 2.18 shows that for different simulation software, the simulated values are very similar across all frequency ranges. Any minor differences are likely due to

different optimization algorithms used. Within the frequency range of up to 6kHz, the simulation results match the scanning results very well, with an average error not exceeding 5dB. However, at 8kHz frequency range, the simulated values are about 15dB higher than the measured values. This is a very interesting result, and the potential reason for this discrepancy could be attributed to some uncertainties in the actual measurement process, such as microphone performance or environmental conditions in the measurement room.

2.7 Discussion

This chapter outlines the process of preparing HRTFs based on iPhone scanning, which can be divided into three main parts: obtaining the 3D mesh, cleaning and optimizing the 3D mesh, and comparing simulated HRTF results. However, this project did not conduct in-depth research on the algorithmic aspects of HRTFs.

For 3D mesh acquisition, Faceid structured light components are applied to acquire 3D meshes with the software Heges 3D to obtain high precision ear meshes and relatively coarse torso meshes. These scanned meshes are saved in .stl format for further processing.

Regarding the cleanup and optimization of the mesh, all operations are performed in several open-source software described in the previous section. These steps include mesh cleanup, simplification of the head and torso meshes, merging of the ear and torso meshes, and finally overall mesh correction and error correction. There are no very standardized steps here, just the functions that should be performed step by step, and the results will be more desirable.

For the simulated HRTF, a comparison with the measured values shows that the simulation results of Mesh2hrtf and COMSOL are in high agreement with the measured values up to 8k Hz, with an average error of less than or approximately 5 dB. above 8k Hz, they are more different. Specifically, almost all simulation results above 8 k Hz will show an error of about 15 dB compared to the measured results.It is worth mentioning that COMSOL cannot consistently obtain results above 8kHz, however the same mesh can be obtained in Mesh2hrtf at more than 16k Hz.

The following questions were sent down during the process.

1. The impact of hair is particularly severe and exists in the scanning process of real people. In this project, wigs and swimming caps were used to cover the hair. Although the hair on the top of the head can be effectively covered, surfaces that cannot be covered, such as those on the temples and neck, still produce unsatisfactory results. Especially for people with long hair, wearing a hat cannot accurately restore the shape of the head. Secondly, although the effect of hair on ear scanning is not obvious, the merging of meshes requires facial meshes around the ears, which are almost impossible to avoid the impact of hair in actual scanning. So far, more effort is still needed for 3D scanning of people with long hair.

2. Frequent 3D operations often lead to the problem of the overall mesh not being closed. Usually, this problem occurs after merging the meshes, and automatic errors occur when loading the mesh for simulation and processing. Unfortunately, Meshlab cannot solve this problem well, but the 3Dbuilder software in the Windows environment can automatically repair it. Generally, the mesh repaired by 3Dbuilder is sufficient for the next preprocessing step. However, please note that the automatically repaired surface often means that the mesh is not so accurate, and these areas usually appear in the area where the trunk and ear meshes are merged. Subsequent research may help solve this problem.

The 3D scanning part is now a relatively mature workflow, however, the ideal HRTF preparation should unify the cleaning and merging of meshes in one software or application, and more effort is undoubtedly needed to achieve a fast and easy collation of qualified meshes.

3

Listening Test of Individual HRTF Performance

Listening tests are a widely recognized and reliable method for evaluating the acoustic performance of sound systems. In contrast to the technical measurements described in the previous section, experiential testing through listening can provide a more comprehensive and nuanced assessment of performance, including various subjective factors such as perception and preference. In this context, the present chapter reports on a listening test involving a cohort of six participants, which was conducted to evaluate the efficacy of an individualized HRTF processing approach.The personalized HRTFs were acquired using the same methodology as in the previous chapter.

3.1 Theory

This section describes the theories used in listening tests.

3.1.1 Two-device Test

Two-device Test is a commonly employed methodology in audio quality evaluation and comparison studies. This test involves the simultaneous use of two audio devices to assess and compare their performance. It serves as a controlled experiment to measure various aspects of audio reproduction and gauge any perceived differences between the devices being tested.

In a typical Two-device Test, a specific audio source is played in parallel on both devices under investigation. Listeners are presented with the audio output and are tasked with evaluating and differentiating the sound quality, clarity, or any other relevant perceptual attributes. These attributes may include but are not limited to frequency response, stereo imaging, dynamic range, distortion, tonal balance, spatial characteristics, or other factors that contribute to the overall auditory experience.

The evaluation process often involves the use of subjective rating scales or methodologies, such as pairwise comparison, ranking, or rating scales, to collect listener preferences or judgments. Listeners may provide ratings or rankings based on their perception of audio quality, preference for one device over the other, or the perceived differences in audio reproduction.

3.1.2 Digital Equalization

Digital equalization refers to the process of modifying the frequency response of an audio signal using digital signal processing techniques. The goal of digital equalization is to correct or enhance the frequency balance of an audio signal, typically by adjusting the amplitude of specific frequency bands. This can be done using various types of digital filters, such as parametric, graphic, or shelving filters, which are designed to alter the amplitude response of the signal over a specified range of frequencies.

Parametric equalizer is a type of digital equalizer that allows the user to adjust the frequency response of an audio signal by manipulating the parameters of one or more digital filters. Unlike graphic equalizers, which provide fixed frequency bands with fixed levels of gain or attenuation, parametric equalizers offer more precise and flexible control over the frequency response of an audio signal.

A typical parametric equalizer consists of one or more filters, each of which can be adjusted to target a specific frequency range, or band. Each filter is defined by several parameters, including center frequency, bandwidth, and gain. The center frequency determines the center point of the band affected by the filter, while the bandwidth controls the range of frequencies affected. The gain parameter determines the amount of boost or cut applied to the selected frequency range.

Parametric equalizers are commonly used in professional audio production, live sound reinforcement, and home audio systems to correct for frequency imbalances or to enhance the tonal quality of an audio signal. They offer a high degree of precision and flexibility, allowing users to tailor the frequency response to match the specific requirements of a given sound system or recording.

3.1.2.1 Shelving Filters

Shelving filter is one type of equalizer filter used in this listening test. It is commonly used to adjust the frequency response of an audio signal by boosting or attenuating specific frequency ranges.

Shelving filters work by gradually increasing or decreasing the amplitude of frequencies above or below a specific cutoff point, known as the "Shelf Frequency". The filter's slope determines the rate at which the amplitude changes beyond the shelf frequency.

There are two types of shelving filters: high shelf filters and low shelf filters. A high shelf filter attenuates or boosts frequencies above the shelf frequency, while a low shelf filter attenuates or boosts frequencies below the shelf frequency.

3.1.2.2 Peak Filter

Peak filter works by increasing or decreasing the amplitude of a narrow range of frequencies around a specific center frequency, known as the "Central Frequency." The width of the range of frequencies affected by the filter is determined by the "Q Factor", which is a measure of the filter's bandwidth.

Peak filters are useful for adjusting the tonal balance of an audio signal by selectively boosting or cutting specific frequency ranges. They are often used to remove resonances or other frequency-specific issues in recordings, or to enhance certain aspects of a sound, such as the "punch" of a kick drum or the "presence" of a vocal.

3.2 Method

This section describes the design and implementation of the listening test.

3.2.1 Set up

In order to validate whether the Kemar results from the previous chapter could also be applied to spatial audio reproduction based on personalized HRTFs generated from iPhone scans of human ears, a two device test-based auditory experiment was conducted.

The same stimuli were convolved with the HRIRs corresponding to specific speaker positions, eg. directly in front and directly to the left(0 and 90 degrees). And played back using Sennheiser HD 560S headphones, Focusrite Scarlett 4i4 audio interface, Lake People G109 headphone preamplifier, and Genelec 8030C speakers. The sound EQ and comparison functions were implemented in Matlab, which will be detailed in a later section.

Equipment	Device model	Application
Laptop	Macbook Pro	Control of all systems
Headphone	Sennheiser HD560S	HRTF rendered audio
Loudspeaker	Genelec 8030C	Initial audio
Audio interface	Focusrite Scarlett4i4	4-channel Audio Generation
Amplifier	Lake People G109	Headphone volume control
Cables	6.5 mm Audio Cable	Devices connection

Table 3.1: List of equipment used in the listening test

In this experiment, a 15-inch Macbook Pro was used as the control device, and participants were instructed to perform the two device test EQ operation on the device. Participants were instructed to keep their head facing the Macbook Pro

at all times, while the horizontal direction of the speaker was adjusted as needed. It is worth mentioning that in the auditory test, the pitch angle of the speaker's geometric center relative to the participant's ear canal was always 0 degrees, and the distance was always 1.2 meters.

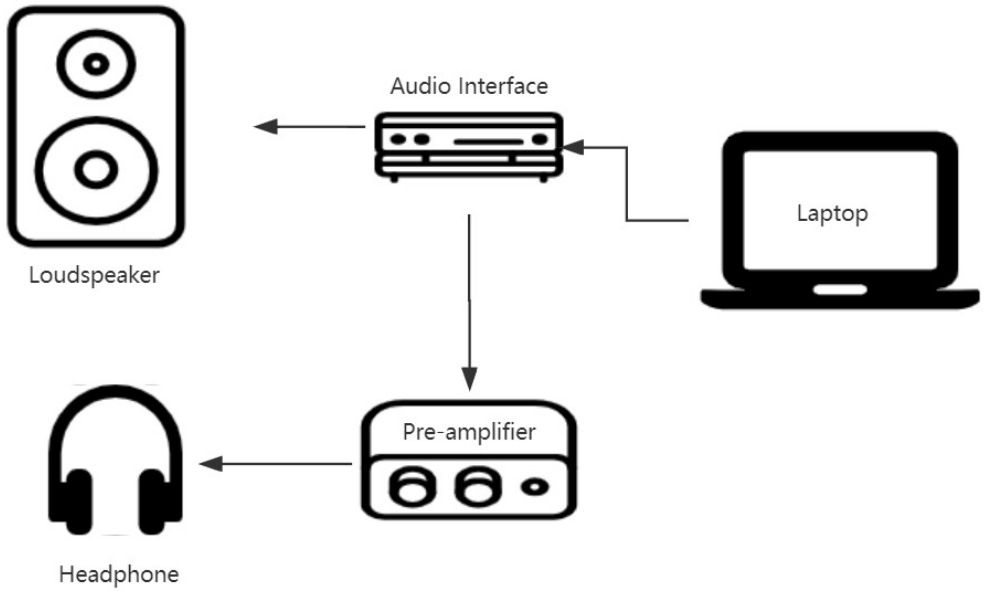

Figure 3.1: Schematic of experimental set-up

The experiment was conducted in an acoustically treated auditory laboratory, and participants were allowed to adjust the playback volume to a comfortable level during the auditory experiment. This means that different participants may choose slightly different playback volumes based on their personal preferences, and these volumes may not be consistent with the volume convolved with the initial HRIRs.

3.2.2 GUI Design in MATLAB

The design of the Graphical User Interface (GUI) is based on real-time audio processing using MATLAB, utilizing plugins for ultra-low latency real-time audio processing. This module serves as the basic architecture for all GUI designs. In addition, the SOFA API plugin is utilized, which in the previous Mesh2HRTF data analysis, was used to read HRIR data sets from SOFA files. In the experiment, it was necessary to align the actual positions of the loudspeakers with specific positions of the impulse response of the HRTF. Therefore, position parameters, including horizontal and vertical angle information, need to be reflected in the design.

As shown in Figure 3.2, the entire software operation process is as follows: First, mono audio is copied, producing three identical output channels. Then, for Channel 1 and Channel 2, the signals are respectively added with spatial cues of the same

spatial position, that is, convolved with the HRIR (the time domain representation of the HRTF) at that position. This produces a personalized HRTF-rendered audio output. The output of the third channel is directly connected to the Genelec 8030C, which serves as the reference loudspeaker mentioned in the two device test. The audio of this channel is not rendered. Next, a module consisting of four second-order IIR filters is added to channels 1 and 2, allowing listeners to adjust the parameters of these filters in real-time to achieve equalization (EQ) goals. Participants adjust this EQ module to obtain headphone output consistent with the reference sound source. All EQ filter parameters will be saved and used to analyze the performance of personalized HRTF.

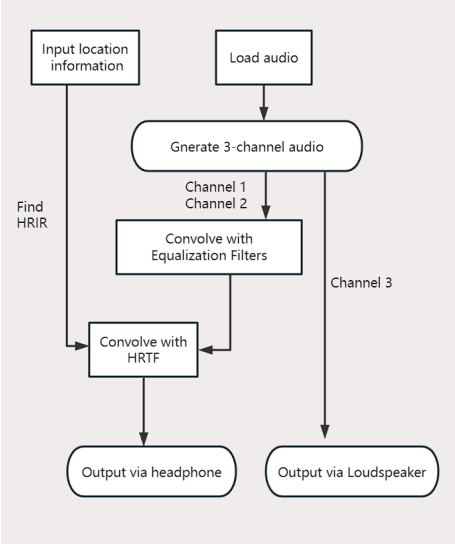

Figure 3.2: GUI flow chart design

MainFunction is one part of the code that reads audio data, applies equalization filters, and plays back the audio through the output device. The audio data is read in blocks, and a filter is generated before each block is processed. If the audio player is on, the function reads filter coefficients from GUI and applies filters on audio data. The filter coefficients are generated based on the user's input. The audio is then convolved with the filter to create a reverberation effect, and the resulting audio is played back through the output device in Channel 1 and Channel 2. Finally, the function checks if the end of the audio file has been reached or if the audio player has been turned off, and releases the audio device writer and the audio file reader if necessary.

Another component of the code is the app file which determine the GUI. This GUI automatically loads the desired audio file and sets specific HRIRs, applies four equalization filters to the audio data, and plays back the resulting audio through the audio output device. As shown in Figure 3.3, the GUI includes switches and options for loading, playing, and pausing audio files on the left, a switch for switching output hardware, sliders and options for adjusting filter coefficients in the middle, and options for saving all current filter coefficients on the right.

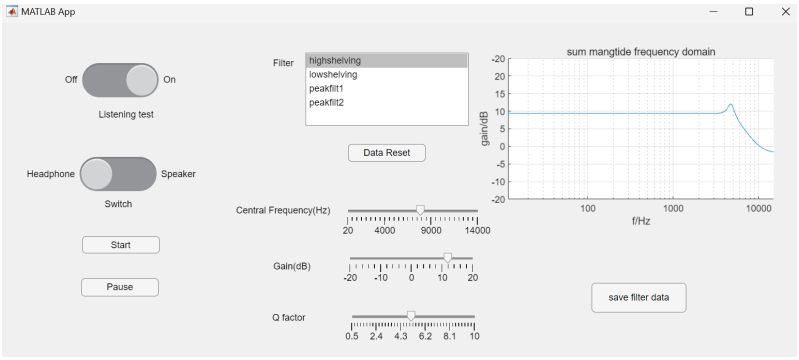

Figure 3.3: GUI in MATLAB including a simple equalizer with four filters

It also features a real-time information display of the audio data being processed, the current frequency of the EQ operation, and the gain of the audio signal are graphically displayed in the upper right corner of the interface. In addition, the GUI automatically stores information about the filter parameters, so that the center frequency, gain and Q-factor of the original filter are automatically displayed on the slider after switching to another filter. If the sound is tested for longer than the length of the audio, the audio is automatically repeated.

3.2.3 Listening Test Protocol

3.2.3.1 Participants' HRTF Generation

Prior to the auditory testing, the 3D mesh of the participants' ears, head, and upper torso were scanned, cleaned, merged, and simulated using mesh2hrtf to calculate personalized HRTFs, which were saved in .sofa format.

Sampling Frequency	Database	Number of positions
48000	ARI	1550
44100	ARI	1550
48000	Default	1850
44100	Default	1850

Table 3.2: List of individual HRTF format

Due to the varying physical features of the participants, the degree of mesh cleaning and modification differed. The main factor causing this difference is the presence of hair on the skin surface. In the previous Kemar comparison, the mesh cleaning could be clearly confirmed because Kemar does not have hair. However, for participants with long hair, it was difficult to accurately generate the head contour, and the resulting mesh may be influenced by various aspects, such as the negative impact on surface optimization, which lack repeatability. This may affect the accuracy and repeatability of personalized HRTFs to some extent. Generally, personalized HRTFs

of participants with short hair are considered to have higher accuracy.

3.2.3.2 Participants

Five participants (4 males and 1 female) were recruited for the study. The average age of the participants was 33 years old. All participants stated to have normal hearing and they were students or staff members of Chalmers University of Technology. Among them, three participants were considered experienced listeners and participated in the subjective listening test, while the remaining two participants lacked expertise in tuning but provided subjective evaluations of the sound quality.

3.2.3.3 Stimuli

For this test, a pop rock music piece was chosen as the stimulus. The piece is approximately 5 minutes in length, featuring various instrumental sounds and vocals that cover the entire frequency spectrum. The piece also includes numerous repetitive melodies, which can facilitate tuning for the listener. The music will loop automatically upon completion of the playback.

3.2.3.4 Procedure

At first, the purpose of the study and the two device test were briefly introduced to the participants, followed by a detailed instruction on how to switch between the devices and perform the equalization task.

Participants were given the opportunity to practice using the devices and perform the equalization task on a set of predetermined stimuli. The stimuli used for training were played repeatedly, allowing participants to become more comfortable with the test.

Figure 3.4: Listening test implementation: in a room with thick curtains and carpets

The experimenter provided guidance to the participants to ensure that they understood the equalization task and were able to operate the devices correctly. Once the participants demonstrated a clear understanding of the equalization task and were able to operate the devices effectively, they were considered ready to begin the actual two device test.

After completing the equalization task and saving the filter coefficients, participants provided feedback on the performance of the two devices and expressed their subjective preferences. Considering the difficulty of the equalization task for non-expert listeners, participants were allowed to voluntarily abandon the equalization results and only record their subjective feelings.

3.2.3.5 Questionnaire

The following table is the questionnaire of listening test table. The goal of this questionnaire is to find out if the listener could notice the difference of localization via different audio processing method including individual HRTF generated and non-individual HRTF generated sound.

Question	Processing method	Evaluation
Sound Quality	individual HRTF	
Sound Quality	non-individual HRTF	
Localization	individual HRTF	
Localization	non-individual HRTF	

Table 3.3: Questionaire of Listening test

For evaluation, the participants should use good or bad for sound quality while use accurate and inaccurate to describe the localization. In general subjective listening tests, there is often a scoring system for the results. The scoring system describes the expression of a gradient for the evaluation, e.g. excellent, very good, good, fair, bad, very bad, not at all. However, due to the lack of sufficient reference standards and clear calibrations in this experiment, especially for localization, it is not possible to obtain concise and accurate results without using gradient evaluations.

Another thing worth mentioning is that interviews were conducted for each participant after the test. More details based on the information provided by the participants will be mentioned later.

3.3 Results

3.3.1 Graphic Representation of the Results

Participants selected the filters provided by each GUI for testing under different relative positions of headphones and impressions. Figure 3.5 shows the result of their EQ.

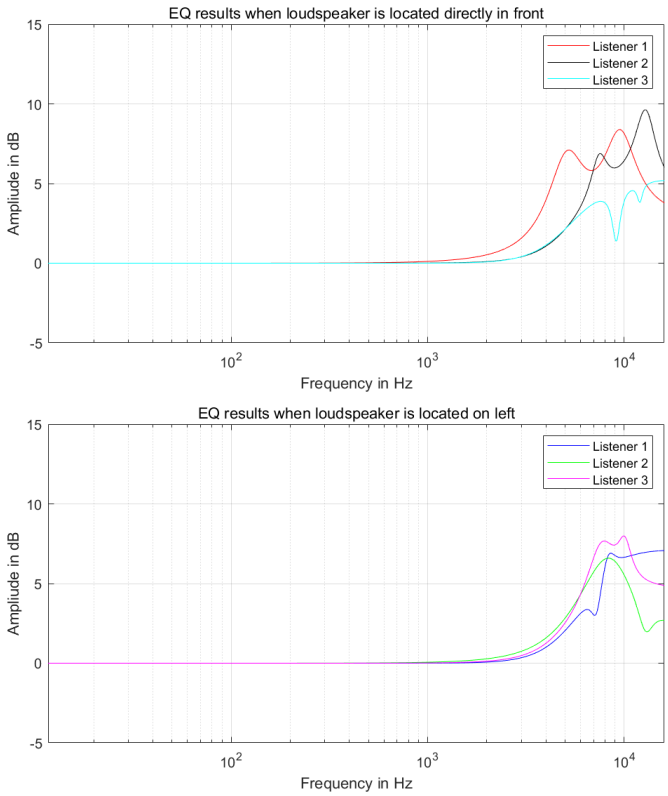

Figure 3.5: EQ results when the audio is placed in different positions. Top picture is the audio in th e front and bottom picture is the audio in the left side.

The EQ experiment conducted in this study mainly focused on the frontal (0 degrees) and left (90 degrees) directions. The participants were asked to EQ the sound in these two directions for approximately half an hour. The distance between the speakers and the participants was approximately 1.5 meters. As shown in the figure, the participants used a computer to adjust the sound. The sound was repeatedly switched back and forth, and the participants adjusted the center frequency, gain, and bandwidth of the four filters to achieve the most consistent sound.

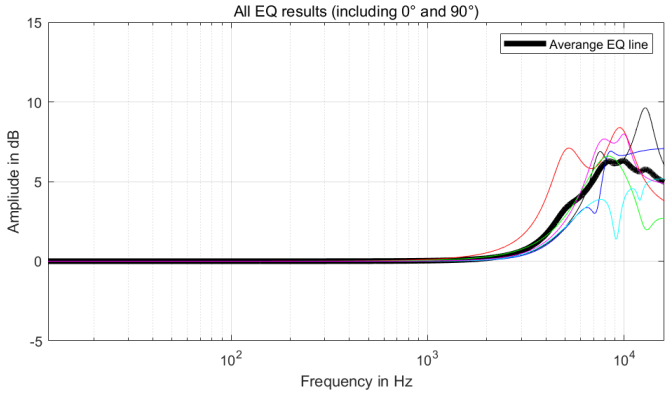

Figure 3.6: Overall EQ results including data in all directions (black bold line is the average of all EQs)

The quanstionare results shown in Table 3.4. It can be found that 80% participants think individual HRTF have a better sound quality compared with the non-individual sound while all of them believe individual HRTF provide a better localization.

Property	Processing Method	Preference
Sound Quality	individual HRTF	80%
Sound Quality	non-individual HRTF	20%
Localization	individual HRTF	100%
Localization	non-individual HRTF	0%

Table 3.4: Results of Questionaire

Issues identified in the post-test interview session:

1. The tuning of loudness is very crucial in listening tests. If the loudness is different, the whole listening experience will be very different.

2. The EQ only targets two individuals, as EQ is usually tailored to experienced listeners. Adding a head tracker may introduce high latency in the EQ (in fact, EQ already has some latency in the SOFA API and Matlab). To achieve the ideal EQ, I may need help from others.

3. It is difficult for candidates who had no listening experiecnce before in this field to evaluate the performance of stacking HRTFs on multi-channel audio, as the experiments did not involve a speaker array as a control (most people think that adding HRTFs to monaural audio is less intuitive than multi-channel audio, as multi-channel seems to increase details, but if you simply stack the left and right HRTFs on two-channel audio, the sound image will drift. This is much worse than direct two-channel audio). EQ cannot directly solve this problem, and a more professional way is needed to record the sound and then find ways to combine spatial information with HRTFs (BRIR). From a psychoacoustic perspective, people are theoretically less sensitive to non-low-frequency room information, and they will automatically compensate for subtle amplitude changes to distinguish what is a sound source. However, after multiple layers of filtering (rendering), amplitude changes can lead to unpleasant results.

3.3.2 Comparision with Headphone Transfer Function

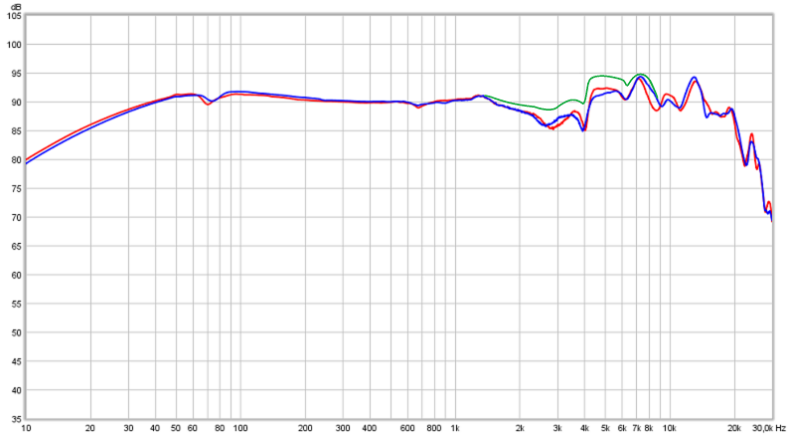

Figure 3.7: Transfer function of Sennheiser HD560S

The experiment was not headphone calibrated, so the calibration curve given by the listener also contains information about the headphone transfer function. A comparison of these two functions shows that the EQ curve can approximately compensate for the differences due to the frequency response curve. This means that the individual HRTF is reliable for sound reproduction in terms of localization and sound quality.

3.4 Discussion

In the listening experiment again, the two sections were given different subjective listening tests. Inexperienced listeners were required to complete questionnaires and interviews containing questions, while experienced listeners were required to complete additional tuning tests.

The results of the test showed that the personalized HRTF-rendered sources produced by this method were preferred over the non-personalized HRTF-rendered sources in terms of both sound quality and localization. For the EQ test, the integrated EQ curve can be approximated as a compensation for the frequency response curve of the headphone itself, which means that the individual HRTFs produced are highly usable.

In this study, several issues during the generation of personalized HRTFs need to be addressed. One of the major challenges is the impact of hair on the accuracy of HRTF generation. Covering the hair with a wig or swim cap is not sufficient as the hair on the surface of the temples and neck cannot be covered, leading to unsatisfactory results. Individuals with long hair are also challenging to scan as wearing a cap cannot restore the shape of the head accurately. Moreover, the merging of the mesh requires facial meshes around the ear, which is almost impossible to avoid

the impact of hair during the actual scanning process. Therefore, further efforts are needed to achieve accurate HRTF generation for individuals with long hair.

4

Conclusion

This chapter summarises the findings from the preceding chapters of this thesis.

4.1 Review

In this paper, a new inividual HRTF processing method based on the 3D mesh scanned by FaceID via iPhone 11 is mentioned. The results show that the generated individual HRTF produced by this method has a lighter preparation process compared to the traditional personalized HRTF method, and participants gave high ratings in the listening test.

Despite the challenges faced in the HRTF generation process, this study demonstrates a relatively mature 3D scanning workflow in Chapter 2. However, there is a need to uniformly clean and merge meshes in one software or application to enable fast and easy collation of qualified meshes for ideal HRTF preparation. In general, further efforts are needed to refine the HRTF generation process and improve the accuracy of the results.

A simple listening test was conducted in Chapter 3 for the generated individual HRTF. Although the cost of HRTF is much reduced compared to the traditional way of HRTF, the production process of personalized HRTF proposed in this paper is still difficult to be tested on a large scale in a population. Therefore, this listening test was conducted with only five participants. From the results, all participants were satisfied with the personalized HRTF produced.

4.2 Discussion of Contributions

Despite the challenges faced during HRTF generation, this study demonstrates a relatively mature workflow for 3D scanning. However, there is a need to unify the cleaning and merging of meshes in one software or application to enable quick and easy collation of qualified meshes for the desired HRTF preparation. Overall, further efforts are needed to refine the HRTF generation process and improve the accuracy of the results.

Chapter 2 describes the workflow of the new method for HRTF preparation and the comparison of results with other methods. 3D meshes generated using Face ID

require more cleaning and optimization steps than 3D meshes generated based on high performance laser scanners. However, the results of individual HRTF parameters simulated from meshes acquired by different means are relatively similar and within 6 dB SPL for the frequency bands that mainly affect localization.

Chapter 3 describes a listening test to test the individual HRTF based on the new method in Chapter 2. This listening test is based on the Two Device Test and focuses on comparing the sound quality and localization of the individual HRTF with that of the non-individual HRTF.The results show that most of the participants have a higher preference for generated individual HRTF.

4.3 Future Work

There are a number of open source spatial audio production tools available. For example, the Spatial Audio Real-Time Applications (SPARTA) suite, developed and open-sourced by the Aalto University Acoustics Laboratory, can process a variety of Ambisonics-based effects and render them to loudspeakers or headphones as needed. These tools can all be used with headphone equipment for perspective tracking, and SPARTA and BST can input personalized HRIR or BRIR files for binaural rendering. Using these tools allows for more extensive personalized HRTF testing and exploration. and loudspeaker synthesis for virtual acoustic environments, and enables data conversion from sound objects to Ambisonics. All of the above mentioned tools can be used with headset devices for perspective tracking, and SPARTA and BST can input personalized HRIR or BRIR files for binaural rendering.

In addition, the human visual perception of space is equally important. As Salmon [**?**]said, numerous studies on 3D acoustic-visual interaction have shown that visual factors play an obvious guiding role in sound source localization, distance perception, and sound externalization, and also have a certain influence on the perception of spatial sense of the acoustic environment. Therefore, in practice, the goal of approximate restoration of the spatial sound field is often achieved through visual factors, for example, in the synchronization of sound and picture, content correspondence, the sound immersion is further enhanced. Therefore, we can further improve the quality of spatial audio perception with the help of head-up display devices.

In the end of this paper, some problems about HRTF's low dimension representation and personalized modeling in HRTF are pointed out, and some improvements have been made. But because of the limited time and the experiment condition, the work is still insufficient. In the study of the artificial neural network personalized HRTF modeling, the measuring procedure of the main shape of human body is adopted the traditional measuring method, which is used to measure the body directly. This method is affected by measuring instruments and operation methods, and it has the disadvantages of low measuring precision and bad stability. In order to get more accurate data of human morphology, to improve the accuracy of the prediction of HRTF and to obtain better location performance, it is necessary to try to do 3D

scanning with depth camera.

Bibliography

[1] Brown C P, Duda R O. A structural model for binaural sound synthesis[J]. IEEE Transactions on Speech & Audio Processing, 1998, 6(5):476-488.

[2] Geronazzo M, Spagnol S, Avanzini F. Estimation and modeling of pinna-related transfer functions[C]// Digital Audio Effects (DAFx- 10), Graz, Austria, 2010:431–438.

[3] Geronazzo M, Spagnol S, Avanzini F. A head-related transfer function model for real-time customized 3-D sound rendering[C]// Signal Image Technol. and Internet-Based Syst. (SITIS '11), Dijon, France, 2013:174–179.

[4] Geronazzo M, Spagnol S, Avanzini F. Mixed structural modeling of head-related transfer functions for customized binaural audio delivery[C]// International Conference on Digital Signal Processing, Fira, Greece. IEEE, 2013:1-8.

[5] Kistler D J, Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.[J]. Journal of the Acoustical Society of America, 1992, 91(3):1637-47.

[6] ShinKH,YoungjinP.Enhancedverticalperceptionthroughhead-relatedimpulseresponse customization based on pinna response tuning in the median plane[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2008, 91(1): 345-3 56.

[7] Hwang S, Park Y, Park Y. Modeling and customization of head-related transfer functions using principal component analysis[C]// International Conference on Control, Automation and Systems (ICCAS2008),Seoul, Korea, 2008: 227-231.

[8] Hugeng, Wahab W, Gunawan D. Effective preprocessing in modeling head-related impulse responses based on principal components analysis [J]. Signal Processing, 2010, 4(4): 201-212.

[9] Zotkin D N, Duraiswami R, Davis L S, et al. Virtual audio system customization using visual matching of ear parameters[J]. 2002, 3(3):1003-1006.

[10] Zotkin D N,Hwang J,Duraiswaini R,et al. HRTF personalization using anthropometric measurements[C]// IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, NY, USA, 2003: 157-160.

[11] Algazi V R, Duda R O,Satarzadeh P. Physical and filter pinna models based on anthropometry[C]// 122th Audio Engineering Society Convention, Vienna, Austria, 2007.

[12] Iida K, et al. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae [J]. The Journal of the Acoustical Society of America, 2014, 136(1): 317-333.

[13] Meshram A, Mehra R, Yang H, et al. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatia sound[C]// IEEE International Symposium on Mixed and Augmented Reality. 2014:53-61.

[14] Torres-Gallegos E A, et al. Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database[J]. Applied Acoustics, 2015, 97: 84-95.

[15] Møller, H., Hammershøi, D., Hundebøll, J. V., & Jensen, C. B. (1992). Transfer characteristics of headphones: Measurements on 40 human subjects. Presented at the 92nd Convention of the Audio Engineering Society, Vienna, Austria.

[16] Vogt, Rips, Emmelmann.(2021) Comparison of iPad Pro®'s LiDAR and TrueDepth Capabilities with an Industrial 3D Scanning Solution. Technologies https://doi.org/10.3390/technologies9020025

[17] Urban, S., Lindemeier, T., Dobbelstein, D., & Haenel, M.W. (2022). On the Issues of TrueDepth Sensor Data for Computer Vision Tasks Across Different iPad Generations. ArXiv, abs/2201.10865.

[18] Ziegelwanger, H., Kreuzer, W., & Majdak, P. (2016). A priori mesh grading for the numerical calculation of the head-related transfer functions. Applied acoustics. Acoustique applique. Angewandte Akustik, 114, 99-110 .

[19] Sergejs D.(2022). Basic_HRTF_tutorial. Retrieved fromhttps://sourceforge.net/p/mesh2hrtf/wiki/Basic_HRTF_tutorial/

[20] Details In-The-Ear HRTFs. (n.d.). Retrieved from https://www.oeaw.ac.at/isf/das-institut/software/hrtf-database/details-in-the-ear-hrtfs

[21] Head and Torso HRTF Computation(2022).Retrieved from https://www.comsol.com/model/head-and-torso-hrtf-computation-75011

[22] Salmon,François H,Etienne E,Nicolas P. (2020). The Influence of Vision on the Perceived Differences Between Sound Spaces [J].Journal of the Audio Engineering Society.68.522-531.10.17743/jaes.2020.0046.

A

Why is iPhone XR

Some suggestions are given by Heges, the iOS 3D Scanner app:

It seems that the iPhone 13 lineup has inferior depth data quality compared to older devices, such as the iPhone X, which produces better scans with richer details. Screenshots of two scans, one made using the iPhone X and the other made using the iPhone 13 Pro, are attached as evidence of this difference.
It appears that the iPhone 13 lineup has lower quality depth data than previous iPhones, potentially due to inferior hardware intentionally put in place to save on manufacturing costs. The creator of Heges, an iOS 3D Scanner app, has attempted to inquire about this with Apple but has not received a response.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY