Head related transfer function experiment and modeling

Qiao, L.

Award date:
2016

Head Related Transfer Function Experiment and Modeling

Liang Qiao

Eindhoven, August 2013

Technische Universiteit
Eindhoven
University of Technology

Department of Mathematics and Computer Science
Industrial and Applied Mathematics

Head Related Transfer Function Experiment and Modeling

Master Thesis

Liang Qiao

Eindhoven, August 2013

Summary

A sound signal generated at a point in space, travels through the medium with a certain velocity as a pressure wave. The wave reaches the ear canal and then arrives at the eardrum. Before the sound wave arrives at the eardrum, the wave is filtered because of the diffraction and reflection properties of head, pinna, and torso, that are described by anthropometric characteristic. Head Related Transfer Function (HRTF) is used to describe this filtering phenomenon. A pair of HRTFs, one for each ear, is used to synthesize a binaural sound that seems to come from a particular point in space. Thus, with HRTF, virtual surround can be created. Virtual surround is an audio representation technique that attempts to create the sensation of more sound sources than actually present at arbitrary positions in a room. Currently, HRTF is widely used in home entertainment products such as headphone and surround sound speaker in home theater.

Ear anthropometries are individual specific. Thus each individual has its own HRTF. In nowadays practice, one uses a non-personalized HRTF to create virtual surround. Such non-personalized HRTF may lead to a misinterpretation of the location of the sound source. The goal of the project, started up at Philips Research, is to exploit the possibilities of obtaining a personal HRTF from the ear anthropometry. Intent of Philips Research is to obtain the best personalized HRTF that optimizes individual virtual surround. Best practice is to measure the HRTF in an anechoic room. But to obtain the HRTF of an individual is a customer-unfriendly procedure. For optimum performance, a set of HRTFs need to be measured by playing test-tones on the loudspeakers and record them with in-ear microphones. Instead of doing that, transform of an existing HRTF or construction of an HRTF based on the individual anthropometric data would be an effective method to get an approximate personalized HRTF, according to acoustic theory. HRTF is determined by anthropometries. For sure, individual anthropometric data is easier, quicker and friendlier to obtain. Thus, to find the relationships between anthropometric and HRTF data should be a rewarding enterprise.

The strategy is to discover the characteristics of the anthropometric data and the features of HRTF. In order to do so, we apply statistical method such as correlation analysis and Principal Component Analysis (PCA) on the HRTF database. Since the scale of the database is small, no strong conclusion can be drawn. As a second strategy, we consider first principal from acoustic theory that explains the relationship as certain aspect of the anthropometry of the ear and head. For that, we conduct acoustic experimentations. These experimentations are designed to obtain different HRTFs by changing one or several parts of a dummys head and ear. The results show some features, as notches and peaks, in HRTF determined by certain parts of head and ear. Combined with acoustic theory, we use geometry to fit the wave scattering paths. Models are developed to predict these features from anthropometric data. The CIPIC database is used to validate the models.

The results are evaluated under the measurement precision and equivalent rectangular band (ERB). The model we has more than 80% accuracy to predict the features of HRTF with its anthropometric data. As a second objective, the HRTF evaluation results also provide the model are very effective.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Human anthropometric properties play a role in human listening experience. When a sound is generated at a point in space, it propagates, reaches the ears, and vibrates the eardrums. Before it arrives at the eardrum, the torso, head and ear influence its direction of propagation, energy level, and other features. The ears collect a sound and send it into the ear canals, while the head and torso reduce the energy of the incident sound wave. Meanwhile, the impact from the ears, head and torso help to locate the position, where the sound is generated.

Head Related Transfer Function (HRTF) describes how the sound waves are affected by the anthropometric properties of the head, ear, and torso. HRTF is used to simulate binaural hearing of sounds from different directions. Each direction is characterized by two impulse responses, one for each ear. Theoretically, HRTF is a transfer function with parameters as the sound source location and frequency; HRTF is highly correlated with morphemic properties of the human head.

By knowing HRTF, sound without any information of the sound location can be transferred into sound that seems to be played at a specific position. HRTF is typically used to simulate multiple loudspeakers of a surround sound set-up through the headphones or stereo speakers. The perception of the sound sources, more than actually present at arbitrary positions, is created by a virtual surround sensation, largely reproduced in home entertainment systems nowadays.

Since the HRTF reflects the impact of the anthropometric properties, the differences from person to person lead to a highly personalized HRTF. A non-personalized HRTF can lead to misinterpretation of the sound source. Front to back or left to right confusions happen frequently, when applying non-personalized HRTFs to the listeners. Thus, the HRTF a personalization is prerequisite for a listening experience of high quality.

The difficulty with individual HRTFs is that they are highly personalized and laborious to get. For optimum performance, a set of HRTFs need to be measured by playing test-tones on the loudspeakers and record them with in-ear microphones. From the view point of a consumer, this way of measuring the HRTF is not friendly and convenient. Although there are some theoretical methods that can partly solve the inconvenience, they are complicated and time-consuming. Besides, theoretical studies have not revealed the high frequency range HRTF. In comparison to HRTF, personal anthropometric data is easier to obtain. A photo of the head is taken and used to analyze the useful anthropometries. Even an actual and precise measurement procedure is more acceptable to an individual than measuring the HRTF with an in-ear microphone. Although the accuracy of the personalized HRTF from the anthropometric data may be less than the ones provided by actual HRTF measurements, it is simple and effective to apply.

1.2 Objective

Human ear has a complicated structure; there are various ways to describe the ear. Since in this project the sound sources are located at the same height as the ears, the main characteristics of the ear that influence the perception of sound in the horizontal plane at that height need to be exploited.

HRTFs are available in open databases. They can be represented in several ways. We usually combine different HRTFs from the horizontal plane into an azimuth-and-frequency-dependent HRTF. The HRTF contains redundant information. Instead of using the whole HRTF set, we distinguish meaningful characteristics from the HRTF that are influenced by the anthropometries. Also these characteristics should contain enough information to regenerate a personal HRTF.

The methods that are used to analyze the relationship are divided into statistical and non-statistical methods. From the statistical point of view, it requires a large range of samples to draw a reliable conclusion. A large number of anthropometric and HRTF data sets need to be obtained. To apply a non-statistical method, such as a first principle method, a physical or numerical model is built according to acoustic theory. To build the model, specific experiments are designed and carried out. The relationships are expected to be observed from these experiments. Also a validation of the model is required.

Since the measured HRTF contains redundant information for sound perception, the vital characteristics of HRTF need to be investigated. To accomplish this, researchers perform listening experiments to see if reducing a characteristic might influence the sound interpretation.

1.3 Strategy

In order to achieve a good virtual surrounding experience for individuals, HRTFs need to be personalized. We tackle this problem by transforming existing un-personalized HRTFs (reference) to the target person's HRTFs by comparing their anthropometric data. The complete personalization process is presented in Figure 1.1. To distinguish the relationships is the focus of this

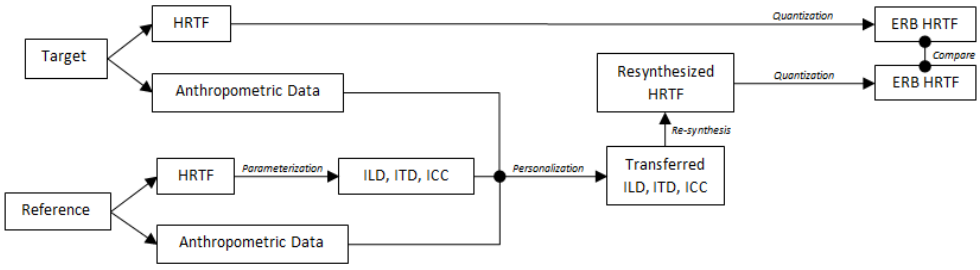


Figure 1.1: HRTF personalizatoin procedure

report. Moreover, other techniques are involved to personalize an HRTF. Signal processing is used when we measure an individual's HRTF in the anechoic room and transform it into usable data. Image analysis is applied when we extract the features, which are used to identify HRTFs. We use transformation algorithms to transform a reference HRTF to a target HRTF with frequency scaling following the information provided by features. A reference HRTF is suggested based on a 3D head model that we build with anthropometric data. The transformed HRTFs are passed through an Equivalent Rectangular Band (ERB) filter to make them correspond to the human listening system. Also we evaluate the accuracy of our model under ERB.

1.4 Content Outline

In this chapter, we introduce the background of HRTF and the objective we intend to solve in this project. Also, the whole personalization process and the involved techniques are presented. The following chapters are organized as: Chapter 2 gives a more specific description of our data and techniques; Chapter 3 presents the method and results in the statistic model; Chapter 4 introduces the experiments we designed and conducted; Chapter 5 describes the model that we build based on the experimental results; Chapter 6 shows the results of the model with statistical method and first principal method; Chapter 7 concludes the advantage and disadvantage of our model and proposes further development.

Chapter 2

Preliminaries

As Algazi metioned [2], although the exact HRTF is complicated, its general behavior can be estimated from fairly simple geometric models of the torso, head and pinnae. These models can be individualized to particular listeners if appropriate anthropometric measurements are available. Thus, using anthropometric data to personalize an HRTF is an effective way to get an individualized HRTF than measuring it in a special listening room. However, the relationships between the anthropometric data and the corresponding HRTFs are unclear. The project aims to find these relationships. Before applying different modeling methods, several relevant data sets have to be obtained, like sets of anthropometric data and of HRTF data.

2.1 Auditory System and Anthropometric Data

2.1.1 Auditory System

The auditory periphery, starting with the ear, is the first stage of the transduction of sound in a hearing organism. While not part of the nervous system, its components feed directly into the nervous system, performing mechanoeletrical transduction of sound pressure-waves into neural action potentials.

The folds of cartilage surrounding the ear canal are called the pinnae. Sound waves are reflected and attenuated when they hit the pinnae, and these changes provide additional information that helps the brain determine the direction from which the sounds come. The sound waves enter the auditory canal, a deceptively simple tube. The ear canal amplifies sounds that are between 3 and 12 kHz. At the far end of the ear canal is the eardrum (or tympanic membrane).

Figure 2.1 shows some components of a pinna, such as tragus, helix, cymba and concha. The tragus is a small pointed eminence of the external ear, situated in front of the concha, and projecting backward over the meatus. Concha is the bowl-shaped part of the pinna (the external) nearest the ear canal. Cymba conchae is the narrowest end of the concha. Helix is the folded over outside edge of the ear.

2.1.2 Anthropometric Data

Consumer Lifestyle Louvain (CL Louvain) of Philips research measured ear and head data on 150 subjects [11]. The data set includes information on head and ear, and also on gender, age and nationality. Age can be a factor influencing the ability of the sound capture. The measurements are slightly different from those in the CIPIC database. There are 9 head-related properties and 8 ear-related properties. The ear data are from the right ear, which are the average values of the same properties during two measurement sessions. An anatomical reference is chosen in order to carry out the measurements uniformly. In the database, the tragion and the Frankfurter Horizontal are used. The tragion is defined as the point on the upper edge of the tragus. The Frankfurt plane is a plane passing through the inferior margin of the left orbit (the point called

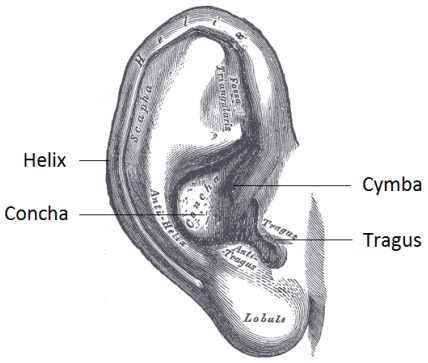


Figure 2.1: Pinna

the left orbital) and the upper margin of each ear canal or external auditory meatus, was most nearly parallel to the surface of the earth, and also close to the position the head is normally carried by the individual. Table 2.1 shows that anthropoemtric measurements in the CL Louvain

Table 2.1: Anthropometric properties in the CL Louvain database

Circ. Top Head FF	Circ. Top Head Leuven	Circ. Rear Head FF	Circ. Rear Head Leuven
Dist. Top Head FF	Dist. Top Head Leuven	Dist. Rear Head FF	Dist. Rear Head Leuven
Head Width			
Ear Length	Ear Width	Ear Ridge	Ear Protrusion
Ear Gage	Tilt Outward Skull	Tilt Outward Ear	Tilt Backward Leuven

database. Some abbreviations are used. For instance, the Circ. Top Head FF means the measured circumference over the top of the head while the head is held in the Frankfurter horizontal. The Dist. Rear Head Leuven denotes the distance from the ear tragus point to the rear of the head based on the Leuven plane. Ear length refers to the greatest distance from the tragus point to a point on the ear helix.

Algazi et al [2] defined a set of 27 anthropometric characteristics, including 17 for the head and torso and 10 for the ear. They are measured from 43 human subjects with 27 men dn 16 women. His CIPIC database describes these anthropometric data. In the CIPIC database, there are 31 subjects whose ear photos are available. This allows other researchers to derive more anthropometric data than he 17 included. The anthropometric properties which is linked to the CL Louvain database are listed in Table 2.2.

Table 2.2: Part of anthropoemtric properties in the CIPIC database

Head width	Head height	Head depth	Head circumference
Pinna width	Pinna height	Pinna flare angle	

2.2 Definition, Measurement and Feature of HRTF

2.2.1 Definition of HRTF

HRTF describes how a sound wave played at the sound source is filtered. We assume that the sound source is placed at a position of (r, θ, ϕ), where r represents the distance from the sound

source to the center of the head O. θ represents the azimuth in the horizontal plane, ϕ represents the elevation in frontal plane, see Figure 2.2. We also assume that the radius of the head is α. Let $p_0(r, f, t)$ denote the responses caused by the sound source at O without the existence of head, $p_L(r, \theta, \phi, t, a)$ and $p_R(r, \theta, \phi, t, a)$ denote the response caused by the sound source in the left and right ear canal. We assume that the acoustic system in the neighborhood of the head is a linear system. Then the relation between output p_L, p_R and input p_0 in time domain satisfies

$$p_L(r, \theta, \phi, t, a) = h_L(r, \theta, \phi, t, a) \star p_0(r, t)$$
$$p_R(r, \theta, \phi, t, a) = h_R(r, \theta, \phi, t, a) \star p_0(r, t)$$

where $h(t)$ is called Head Related Impulse Response (HRIR). The equivalent relation between output P_L, P_R and input P_0 in the frequency domain is:

$$P_L(r, \theta, \phi, f, a) = H_L(r, \theta, \phi, f, a) \times P_0(r, f)$$
$$P_R(r, \theta, \phi, f, a) = H_R(r, \theta, \phi, f, a) \times P_0(r, f)$$

where H_L and H_R represent the complex amplitude generated by the sound source in the left and right ear canal. P_0 is the complex amplitude measured at the point where the center of the head is supposed to be in the free-field. H is then called Head Related Transfer Function, which can be transferred from h using the Fourier transformation:

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt.$$

In our project, we focus on the HRTF for the horizontal plane and fix the sound source distance and head radius (see Figure 2.2), so the formula simplifies to:

$$P_L(\theta, f) = H_L(\theta, f) \times P_0(r, f)$$
$$P_R(\theta, f) = H_R(\theta, f) \times P_0(r, f)$$

Thus, if the HRTF is known, the complex amplitude caused by the sound source in the left and

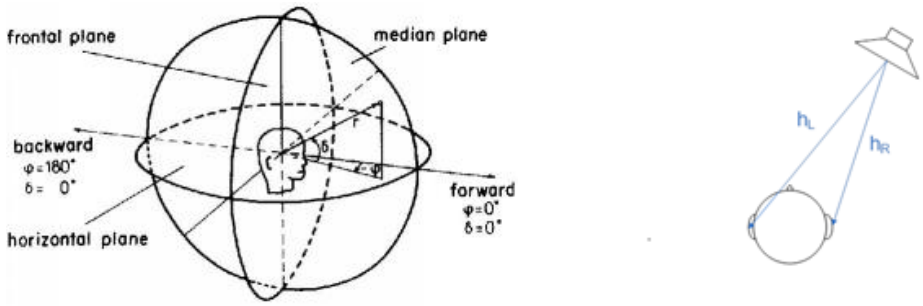


Figure 2.2: Spatial coordinate for HRTF

right ear canal can be determined if P_0 is measured. So if a signal is filtered by the HRTF, then

$$S_L = H_L S_0$$
$$S_R = H_R S_0$$

It is called the binaural recording and playback, as well as the method we used in the experiment to measure the HRTF. A stimulus $s_0(t, \theta)$ is played and the impulse response $s_l(t, \theta)$ and $s_r(t, \theta)$ at the ear canal are measured. HRTF can then be calculated from the formulas above.

2.2.2 Measurement of HRTF

HRTF data sets in the project needs to be measured, although they are laborious and time consuming to get. In order to minimize the influence of early reflections and reverberations on the measured response, HRTFs are typically measured in an anechoic chamber. The walls of the anechoic chamber are made or covered by sound-absorbing materials. Usually, the experiment is simplified imposing a limitation on the equipment and time. Although HRTF is a continuous function, only limit points can be measured. A further interpolation is applied to simulate HRTF in the whole space. The chair for the listeners is in the center of the room; it is surrounded by 19 speakers. The 19 speakers are set up in a semicircle, with an angular distance of 10 degrees, from 0 degree to 180 degrees. The assumption is made that there is no much difference from ipsi-lateral to contra-lateral (In the experiment, the ipsi-lateral means the right side, since the equipment is on the right side of the listener). The semicircle formed by the 19 speakers is used to cover the ipsi-lateral. Two additional speakers are placed at an elevation of 15 degrees at azimuths 0 and 60 degrees. Using logarithmic sine-sweeps the HRTFs are measured by microphone capsules connected to a custom-made microphone amplifier. In an experiment, the subject is positioned in

Figure 2.3: HRTF measurement setup

the middle of the straight connection between the 0 degree speaker and 180 degree speaker. Two Philips webcams, fitted with reference indicators and their feeds displayed on a computer monitor, the subject is able to hold its head at a fixed position. An adjustable chair is used to accommodate for the different heights of the subjects. This way, using the ear-canal as a (front-back) reference every subject has its head centered at the same position. The tilt of the head is referenced in the Frankfurt plane. In the experiment that we refer to, the HRTF of 51 individuals are measured. They are from different gender, age, and nationality; 13 of them are subjects from the CL Louvain database [11].

In the CIPIC database, HRTFs are measured in a similar way. Sound source location is specilified by the azimuth angle θ and elevation angle ϕ in interaural-polar coordinates. In order to obtain approximately uniform density on the sphere, azimuths are sampled from 0 to 45 in steps of 5, at 55, 65, 80, 100, 115, 125, and from 135 to 180 in steps of 5.

Figure 2.4 shows ipsi- and contra-lateral of subject 003 from the CIPIC database. From top to bottom, HRTF with different azimuths in the horizontal plane are plotted.

2.2.3 Feature of HRTF

In Figure 2.4, we can see the ipsi- and contra lateral HRTF, they are different. But still they have features in common. For instance, in ipsi-lateral HRTF, all of them have a notch around 9 kHz and a peak around 5 kHz. Moreover, in order to get the whole spectrum of HRTFs in the

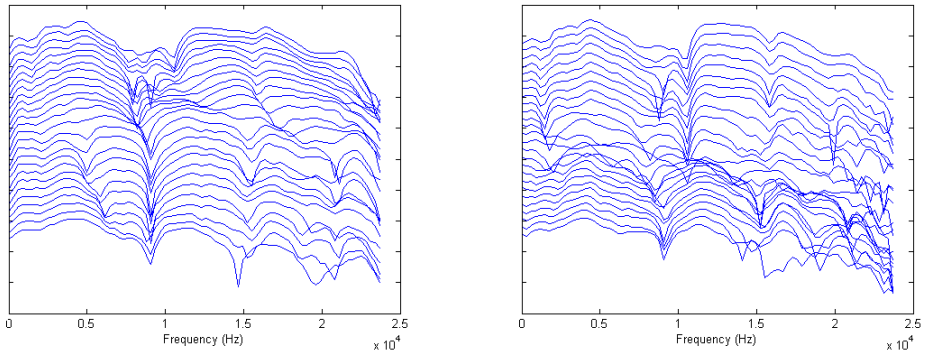


Figure 2.4: Ipsi- and contra-lateral HRTF, plotted for different azimuth

horizontal plane, the HRTFs from the same lateral are considered as one set. Figure 2.5 shows a 3D image of the ipsi-lateral HRTF and the contra-lateral HRTF. Structures, like notches, peaks, peak areas can be observed. Instead of using the whole spectrum, a simple representation of the

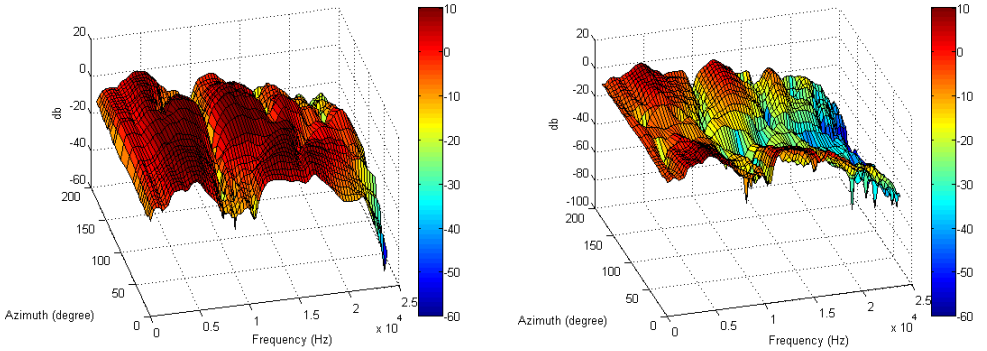


Figure 2.5: Ipsi- and contra-lateral HRTF

HRTF set is wanted. Such representation can be suggested by a principal component analysis [5], or by features annotated artificially. The features are used to annotate the common qualitative structure of an HRTF. Based on the distinguished features, relationships to the anthropometric properties can be built. By assigning the appropriate value to the features, statistical analysis can be applied in subsequent procedures. A number of features are extracted and studied [3]. Raykar

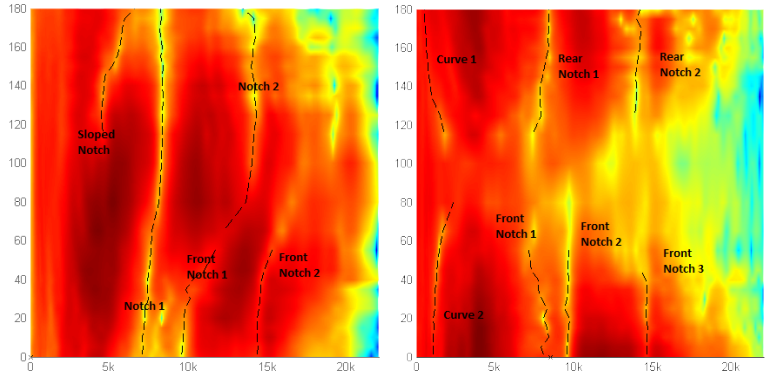


Figure 2.6: Notches annotated in the ipsi- and contra-lateral HRTF

[7] summarizes that the prominent features contributed by the pinna are the sharp notches in the spectrum, commonly called the pinna spectral notches. There is substantial psychoacoustical, behavioral, and neurophysiological evidence to support the hypothesis that the pinna spectral notches are important cues for localization. Table 2.3 lists the notches we annotate in HRTF set.

Table 2.3: Features

Lateral	Name	Abbreviation	Position	
			Frequency (kHz)	Azimuth
Ipsi-	Sloped Notch	iSN	3 – 6	120 – 180
	Notch 1	iN1	5 – 8	0 – 180
	Notch 2	iN2	10 – 15	60 – 180
	Front Notch 1	iFN1	9 – 12	0 – 60
	Front Notch 2	iFN2	14 – 18	0 – 60
Contra-	Curved Notch 1	cCN1	0.5 – 1	120 – 180
	Curved Notch 2	cCN2	0.5 – 1	0 – 60
	Front Notch 1	cFN1	5 – 8	0 – 60
	Front Notch 2	cFN2	9 – 12	0 – 60
	Front Notch 3	cFN3	14 – 18	0 – 60
	Rear Notch 1	cRN1	5 – 8	120 – 180
	Rear Notch 2	cRN2	10 – 15	120 – 180

2.3 Technique

This section describes the techniques we use in this project. The statistical method and first principle method are techniques to discover the characteristics of the anthropometric data and the characteristic of HRTF. Besides these two techniques, techniques from signal processing, 3D registration, transformation and numerical methods with partial differential equation are performed to complete the whole personalization procedure.

2.3.1 Signal Processing

Stimulus

During the measurement, a stimulus is needed. The stimulus has the form

$$x(t) = A \sin \phi(t)$$

with

$$\omega = \frac{d\phi}{dt}.$$

Since the human auditory system processes frequencies in a logarithmic scale that is approximated by the Bark scale, an exponential function for the angular frequency as function of time is preferred.

$$\omega = \frac{d\phi}{dt} = Be^{Ct}.$$

If the start and the end frequency, ω_1 and ω_2 at $t = 0$ and $t = T$, respectively, are given, the parameters B and C can be calculated. They are given by

$$B = \omega_1$$
$$C = ln(\frac{\omega_2}{\omega_1})/T$$

Table 2.4: Sweep signal

f (Hz)	20	500	1000	2000	4000	8000	16000	20000
t (s)	0	0.46	0.56	0.66	0.76	0.86	0.95	1

So

$$\omega(t) = \omega_1 e^{ln(\frac{\omega_2}{\omega_1})t/T} = \omega_1 e^{ct/T}$$

where

$$c = ln(\frac{\omega_2}{\omega_1})$$

Then

$$\phi(t) = \frac{\omega_1 T}{c}(e^{ct/T} - D)$$

We take $\phi(0) = 0$, so that $D = 1$ and

$$x(t) = A\sin(\frac{\omega_1 T}{c}(e^{tc/T} - 1))$$

In order to measure HRTF in the laboratory, ω_1, ω_2 and T can be chosen to generate different sweeps. In this experiment, the frequency of the sweep varies from 20 Hz to 20 kHz (see Figure 2.7), which is the range of perceivable frequencies for the human auditory system, the sample rate (f_s) we use is 48 kHz, larger than twice of 20 kHz, according to the Nyquist Shannon sampling theorem. We set the total time T equal to 1 second. With these choices,

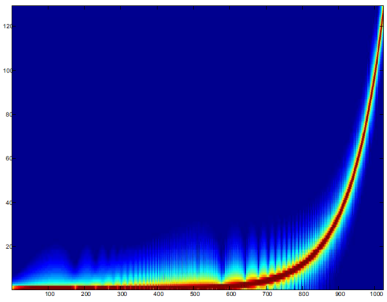


Figure 2.7: Sweep signal

$$c = ln(\frac{\omega_2}{\omega_1}) = ln(\frac{20000}{20}) = 3ln10 \approx 7$$
$$\omega(t) = \omega_1 e^{ct/T} \approx \omega_1 e^{7t}$$

and

$$t \approx \frac{1}{7}\ln\frac{F}{20}.$$

when frequency of the stimulus increases to F Hz. The table shows the corresponding time that the stimulus generates a certain frequency.

Sampling

The signals need to be sampled since the computer can not deal with the continuous-time signal. Since $T = 1$ sec, we take $\Delta t = 1/(4.8 \times 10^3)$ sec. Then time sample $N = T/\Delta t = 48000$. The test signal is discretized:

$$x(n) = A\sin(\frac{\omega_1 T}{c}(e^{n\Delta tc/T} - 1)), \quad n = 0, \ldots, N$$

Discrete Fourier Transformation (DFT), which transforms the discrete signal from the time domain to the frequency domain, is defined as:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}, \quad k = 0, \ldots, N-1.$$

In the experiment, we measure the response of the sweep $y(n\Delta t)$ from the microphone in time domain. We use the same rate to record, and in the A/D convertor, high frequencies above $f_s/2$ are filtered. There is still some noise that requires some post-processing steps to reduce.

Similar, we have the DFT of $y(n\Delta t)$:

$$Y(k) = \sum_{n=0}^{N-1} y(n)e^{-j\frac{2\pi}{N}kn}, \quad k = 0, \ldots, N-1.$$

The transfer function H of any linear time-invariant system at frequency domain is:

$$H(k) = Y(k)/X(k).$$

So if we consider the test sweep as input, the measured sweep as output, the space between the speakers and the listener as the system, its frequency response $H(e^{j\omega n\Delta t})$ can be calculated. And the spectrum of HRTF is defined as:

$$A(f) = 20\log_{10}(|H(k)|)$$

The frequency resolution of the human auditory system is modeled by Equivalent Rectangular Bands (ERB, see Chapter 2.6) that have bandwidths equivalent to the bands induced by the basilar membrane in the cochlea.

2.3.2 Statistical methods

Principal Component Analysis

Principal component analysis is a powerful statistical technique that is used to reduce the dimensionality in a correlated dataset. Thus the database structure is simplified and its main information is kept. Generally, PCA is an analysis of variance. Directional information is highlighted whereas redundant information is neglected. The resulting data is a set of orthogonal principal components that are sorted according to their variance in the original data.

Martens [5] presented a model that is based on PCA and minimum-phase reconstruction. The data set included critical-band-filtered HRTFs of 36 source positions in the horizontal plane of two subjects. He used only 4 of 24 components for reconstruction to explain 90% of the data set.

Kistler and Wightman [4] analyzed HRTFs of 10 subjects using PCA and showed that more than 90% of HRTF variance can be approximated with only 5 of 150 basis spectral functions. Judgments on the horizontal plane were accurate even when only the first PC was used for reconstruction. The first principal component therefore contains most of the interaural intensity information necessary for lateral discrimination. Front-back and up-down performance dramatically decreased when using less than five PCs. Basis functions 2-5 turned out to be crucial parameters for front-back and up-down discrimination. So when using less than five components,

the fine spectral details are not accurately represented by the model, therefore subjects are not able to distinguish properly front-back and up-down.

Middlebrooks [6] verified if PCA of different data sets give almost the same results. Because in theory, differences in measurement setup and subjects should have only little effect on the resulting components, he compared own measured HRTFs with the database by Kistler and Wightman. A high correlation between the two different sets of basis vectors was found (PC1: 96%, PC2: 88%, PC3: 70%). This reveals that PCA is relatively robust to different measurement techniques.

In our project, we also use PCA to reduce the dimension of data. We assume that the data matrix X has n rows and p columns. N rows mean n different cases and p columns represent p samples in one case. Matrix X is denoted by:

$$X = [x_1, x_2, \ldots, x_n],$$

where $x_j = [x_{1j}, x_{2j}, \ldots, x_{pj}]'$, $j = 1, 2, \ldots, n$. The sample mean of each case is defined by:

$$\bar{x}_j = \frac{1}{p}\sum_{r=1}^{p} x_{rj}, \quad j = 1, 2, \ldots, n.$$

and the sample variance of each case is defined by:

$$s_{jj} = \sigma_j^2 = \frac{1}{p}\sum_{r=1}^{p}(x_{rj} - \bar{x}_j)^2, \quad j = 1, 2, \ldots, n,$$

where σ_j denotes the standard variance. The sample covariance between ith and jth cases is defined as:

$$s_{ij} = \frac{1}{p}\sum_{r=1}^{p}(x_{rj} - \bar{x}_j)(x_{ri} - \bar{x}_i), \quad i, j = 1, 2, \ldots, n.$$

Now the correlation coefficient between ith and jth can be defined:

$$r_{ij} = \frac{s_{ij}}{\sigma_i \sigma_j} \quad i, j = 1, 2, \ldots, n.$$

The sample mean vector $\bar{x} = [\bar{x}_1, \ldots, \bar{x}_n]$ can be viewed as the center of gravity of the sample points in R^p. If a unit mass is given to every point, the mean vector denotes the center of mass. The matrix of covariance and the matrix of correlation are defined as:

$$S = s_{(ij)} \quad i, j = 1, 2, \ldots, n,$$
$$R = r_{(ij)} \quad i, j = 1, 2, \ldots, n.$$

The covariance matrix S measures the spread of the data points about the mean.

The purpose of principal component analysis (PCA) is to transform a high dimensional dataset to a much lower dimensional one. To do that, a rotation of the standardized variables $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$, say $y_1, y_2, \ldots, y_n$ need to be found. $y_1, y_2, \ldots, y_n$ are called principal components if they are uncorrelated and satisfy $\mathrm{Var}(y_1) \geq \mathrm{Var}(y_2) \geq \cdots \geq \mathrm{Var}(y_n)$. The standardized matrix $[\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n]$, denoted as $\hat{X}$, is defined as:

$$\hat{x}_j = \frac{x_j - \bar{x}}{\sigma_j}, \quad j = 1, 2, \ldots, n.$$

Let Y denote $[y_1, y_2, \ldots, y_n]$ and W_D denote the rotation matrix, then

$$Y = W_D \hat{X}.$$

The solution of PCA is given by the following equation:

$$C_{\hat{X}} = W_D \Lambda W_D^T,$$

where the rotation matrix W_D happens to be the eigenvector of the covariance matrix of $\hat{X}$.. Thus the eigenvectors are called weights for each principle components, which are the new vector $y_1, y_n, \ldots, y_n$. They form the same space, but are uncorrelated to each other. y_1 has the largest variance, which means that it contains the greater part of the information because of its largest contribution to the variance spreading out the population as much as possible.

The percentage p_i which the principle component accounts for is evaluated by its respective eigenvalue λ_i:

$$p_i = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_N} \times 100\%$$

In applications, only the principal components that account for a high percentage are used. The new space that they form is of lower dimension than the original one without losing important information. Besides, the W_D gives a way to weight each old vector and help to investigate the meaning of the new linear combinations. PCA can be submitted to the anthropometric data and HRTF data sets to discover the information they contain. For instance, we represent the HRTF data set as a matrix $A \in N \times M$:

$$A = (x_{kj})$$

where k refers to speakers $k = 1, 2, \ldots, M$ and j to frequency points $j = 1, 2, \ldots, N$. So two different direction PCA can be applied to the data matrix. The vectors u_k, v_j are defined as:

$$\begin{aligned} u_k &= (x_{k1}, x_{k2}, \ldots, x_{kN}), \quad k = 1, \ldots, M \\ v_j &= (x_{1j}, x_{2j}, \ldots, x_{Mj}), \quad j = 1, \ldots, N \end{aligned}$$

Thus the corresponding centers are given by

$$c = \frac{1}{M} \sum_{k=1}^{M} u_k$$

and

$$d = \frac{1}{N} \sum_{j=1}^{M} v_j$$

and the covariance matrices by

$$S_c = \frac{1}{M} \sum_{k=1}^{M} (u_k - c)(u_k - c)^T$$

and

$$S_d = \frac{1}{N} \sum_{j=1}^{N} (v_j - d)(v_j - d)^T$$

Then the eigenvectors and eigenvalues can be obtained according to

$$\begin{aligned} S_c &= W_c \Lambda_c W_c^{-1} \\ S_d &= W_d \Lambda_d W_d^{-1} \end{aligned}$$

Two matrices are generated from the rotation matrices Wc and Wd:

$$\begin{aligned} Y_c &= W_c A \\ Y_d &= W_d A^T \end{aligned}$$

A simple example is given with M = 2, N = 8913. The HRTF data is from the ipsi-lateral transfer function of 0363's profile in the experiment. M = 2 means that only the speaker 1 and speaker 2 are taken. The PCA is submitted to the dimension of speaker. Following the previous procedures, the eigenvalues and the eigenvectors can be obtained:

$$\lambda_1 = 1.971, \quad \lambda_2 = 0.029$$

and the respective eigenvectors are

$$w_1 = (0.7071, 0.7071)^T, \quad w_2 = (-0.7071, 0.7071)^T$$

So the new matrix Y is

$$Y = WX = (w_1, w_2)^T (u_1, u_2)^T$$

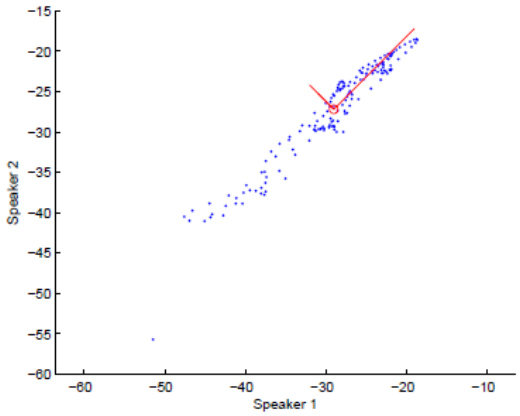
Here

$$y_1 = w_1^T (u_1, u_2)^T = 0.7071u_1 + 0.7071u_2$$
$$y_2 = w_2^T (u_1, u_2)^T = 0.7071u_1 - 0.7071u_2$$

The percentages of these two principle components are:

$$p_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} 100\% = 99\%$$
$$p_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2} 100\% = 1\%$$

The y_1 and y_2 are called the principal components (PCs) and they are independent of each other. y_1 accounts for 99% and explains almost the whole variance. The two dimensional space can be replaced by a one dimensional space spanned by y_1 without losing important information.



Correlation Analysis

Correlation analysis may help to discover the mutual relationship between one anthropometric data feature and one HRTF data feature by setting a threshold value. The strategy of the correlation analysis is as follows. If the Pearson correlation is higher than a threshold, a potential relationship can be concluded. Considering the error of the measurement and annotation and the small scale of the database, correlation values higher than 0.6 or smaller than -0.6 are viewed as describing potentially important and useful information.

For convenience, we give a description of the correlation method used. Let z_i denote the i-th property and define the representation measurement vector (containing M samples) as

$$z_i = [z_{i1}, z_{i2}, \ldots, z_{iM}], \quad i = 1, 2, \ldots, N_A$$

Also, we use x_l to denote the l-th feature ($l = 1, 2, \ldots, N_F$) and define the corresponding data vector

$$x_l = [x_{l1}, x_{l2}, \ldots, x_{lM}], \quad l = 1, 2, \ldots, N_F$$

In order to investigate the relation between different anthropometric properties and features, the Pearson correlation coefficient is calculated according to the formula

$$\rho_{x_l,z_i} = \frac{\sum_{j=1}^{N}(x_{lj} - \mu_{x_l})(z_{ij} - \mu_{z_i})}{\sqrt{\sum_{j=1}^{N}(x_{lj} - \mu_{x_l})^2}\sqrt{\sum_{j=1}^{N}(z_{ij} - \mu_{z_i})^2}}$$

μ_{x_l}, μ_{z_i} denote the mean of x_l and z_i. σ_{x_l},σ_{z_i} denote the standard deviation of x_l and z_i. The Pearson correlation matrix consists of correlation coefficients ρ_{x_l,z_i}.

2.3.3 First principle method

Generally, researchers believe that the features like notches and peaks in the HRTF are caused by the interference of the direct wave and its reflection and diffraction by the head and ear or by the resonance of the ear. It is important information for localization. With the knowledge of the physics of sound, models can be built to predict the features in the HRTF from the anthropometric data. Also, relationships derived from statistical methods can be verified.

Ear

Batteau [3] suggested that the external ears or pinnae have a role in human localization. By attaching artificial pinnae to microphones, Batteau showed experimentally that the pinnae provide major elevation cues, and he argued that they created azimuth cues as well. Moreover, Batteau proposed a simple pinna model, with the wall of the concha and the rim of the pinna acting as reflectors.

According to Batteau, the sound waves with high frequency are reflected by the outer ear, when their wavelength is small enough to compare with the pinna dimensions. The sound waves are divided into the direct waves and the reflected waves (especially reflected by the ear pinna). Interference between the direct and reflected waves causes sharp notches to appear in the high frequency side of the HRTF with a periodicity that is inversely proportional to the time delay of each reflection. This theory gives insight in the function of the pinna. In fact, the size of the pinna is around or smaller than 80 mm, which means that only waves with frequencies in the range of 3-4 kHz are comparable and affected. And only waves with frequencies in the range of 5-6 kHz or above are largely influenced. Further, theory indicates that the interference between the direct waves, the reflected waves, and the diffracted waves form a frequency filter and causes sharp notches and peaks in the HRTF.

As mentioned in Simone [9], reflection models usually assume all reflection coefficients to be positive. If this were the case, the extra distance travelled by the reflected wave with respect to the direct wave must be equal to half a wavelength in order for destructive interference to occur, which translates into spectral notches in the frequency domain. Let c denote the speed of sound in air, l the wavelength and d the half length of l. Then frequency f satisfies the relation:

$$f = \frac{c}{l} = \frac{c}{2 \times d}$$

This is also the assumption when tracing reflection points over pinna images based on extracted notch frequencies. In Simone's article, it is stated that the notch around 10 kHz to 12 kHz is caused by the length of helix and length of antitragus.

Furthermore, Shaw [8] worked on the pinna frequency response and developed Batteau's model. Shaw investigated a variety of simplified shapes for physical pinna models, seeking the simplest shape that could account for the spectral features found in real PRTFs. This led to the flange-and-cavity model. The relative simplicity of this model suggests that a fairly small set of anthropometric parameters should be sufficient to characterize the pinna, but does not specify them explicitly. Through patient acoustic experimentation, Shaw identified a number of longitudinal and transverse resonant modes of the pinna, and the preferred directions for exciting these modes. Shaw's resonance model showed that the ear ridges and valleys and the ear canal form a resonance system. The obvious resonances are around 3 kHz, 5 kHz, 9 kHz, 11 kHz and 13 kHz, which is well explaining the positions of the peaks. He also argued that the first peak around 3 kHz is caused by the resonance of a quarter-wavelength of the ear canal. Since the ear canal is around 2.5 cm, we obtain:

$$f = \frac{c}{4d} = \frac{340}{4 * 0.025} = 3400Hz$$

The measured resonance around 2 kHz is caused by the compliance of the canal and the tympanic membrane [11]. The paper also states that two peaks around 7 and 9 kHz are caused by vertical modes of resonance involving the concha, cymba and triangular fossa, and two higher peaks around 12 and 14 kHz are caused by horizontal transverse modes of resonance involving the front and back parts of the concha and cymba.

Head and Torso

The so-called 'snowman model', given by Algazi [1], provides a simple example of a structural extension of the spherical-head model. A spherical torso is placed directly beneath a spherical head to account for shoulder reflections. According to Algazi, the Snow Man model provides the major low-frequency localization cues. When the source spectrum is limited to 3 kHz to prevent the pinnae from providing elevation cues, essentially the same localization performance is obtained.

Xie [12] mentioned that different clothe materials affect the reflection of the torso. But the frequency under 3 kHz almost will not be affected by the clothes according to his experiment. Only above 5.0 kHz, the clothe material affects the HRTF in the experiment without pinna. But the pinna largely affects the high frequency waves.

2.3.4 3D Head-Ear Model

A 3D head-ear grid is required when applying a numerical method. Researchers use 3D scanners or simply use basic geometry models such as spherical heads to obtain the grids. The 3D scanner is limited in use and unfriendly to the tester. The basic geometry model is too brief to represent the details of individuals. In our project, we develop a 3D model based on the anthropometric data. It will be used in a further personalization step. The 3D point set is constructed with ear

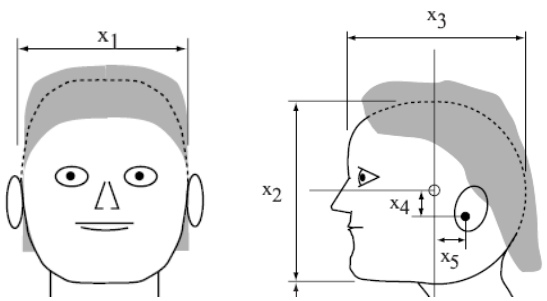


Figure 2.8: CIPIC head annotation

and head measurements from the CIPIC database, as shown in Figure 2.8 and 2.9. First we build

the point set of the head. The coordinate system of the head consists of three planes. They are front, lateral and horizontal planes. In the front plane, we let the y axis be the midline of the

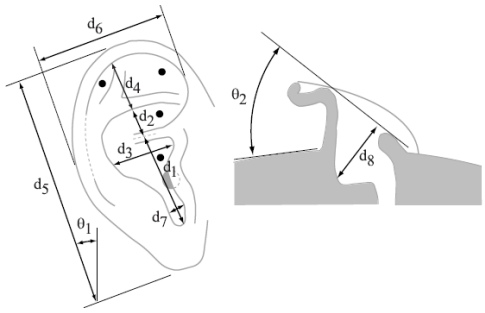


Figure 2.9: CIPIC ear annotation

head and the x axis cross the ear tragus point. We assume that the widest length (denoted by x_1 in Figure 2.8, same below) of front head is just under the eyes, which has a vertical distance of x_4 to the ear tragus point. We approximate the upper and lower head shapes with two half ellipses, as shown in Figure 2.10.

In the lateral plane, we let the y axis parallel to the midline of the head top and cross the ear tragus point and the x axis be the line that also crosses the ear tragus point. The tragus has a vertical distance of x_5 to the midline. In consistency of the front plane, the widest length (x_2) of lateral head is also just under the eyes, which has a vertical distance of x_4 to the tragus. We approximate the upper head shape with a half ellipse. In the horizontal plane, we let the x axis

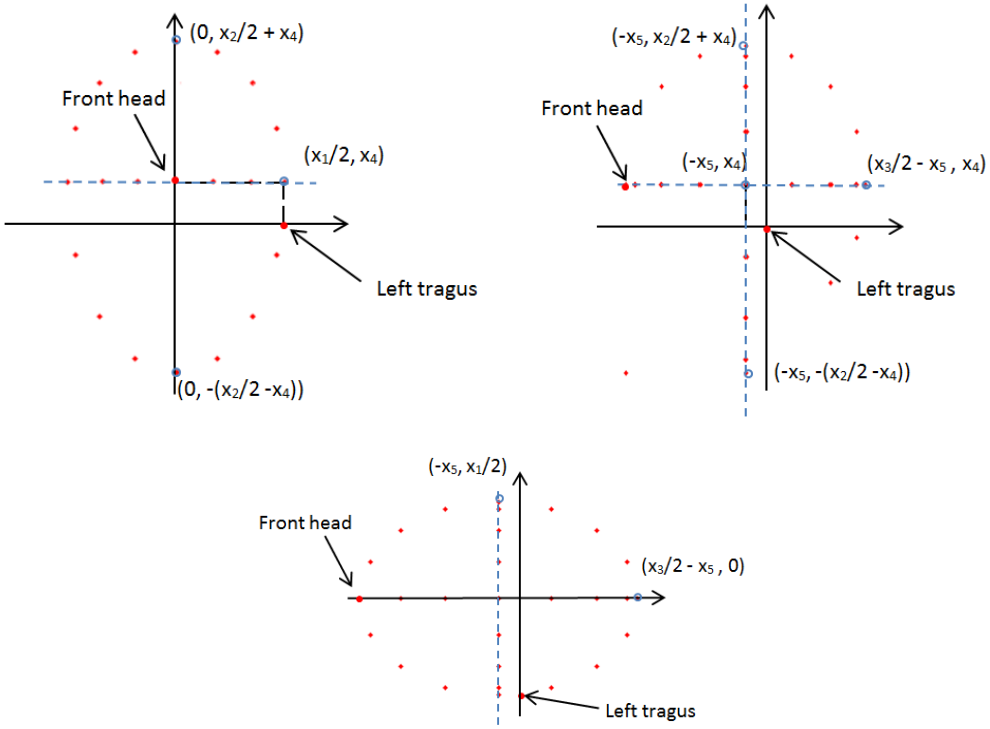


Figure 2.10: 3D head point cloud

be the midline of the head top facing front and let y axis be parallel to the midline of the head top facing lateral and cross the tragus point. The second midline has a vertical distance of x_5 to

the tragus point. We approximate the top head as a complete ellipse. Furthermore, in order to

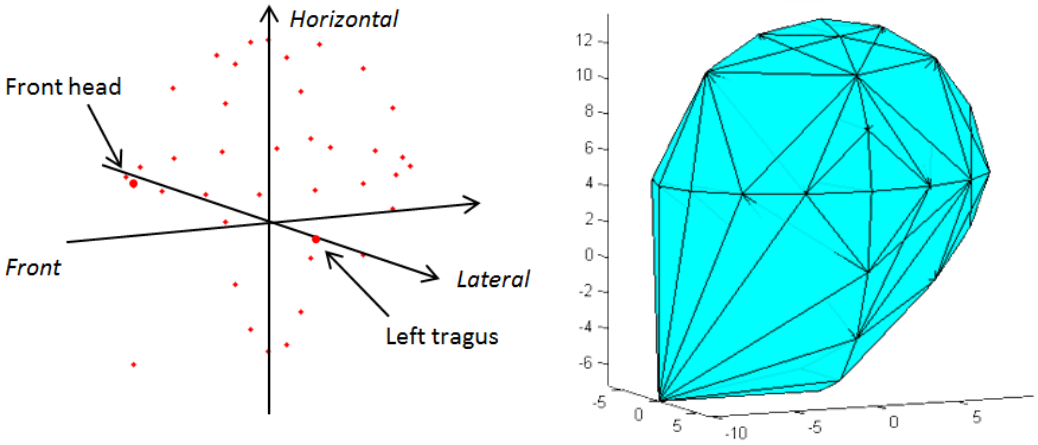


Figure 2.11: 3D head point cloud

link these three planes, we make the front head point and the tragus point as mark points. Then a 3D coordinate system can be derived, as shown in Figure 2.11.

In contrast to the way we construct the head coordinate system, we first approximate the ear shape in the plane parallel to the ear surface, called the surface plane. We approximate the shape of the ear helix and concha with two Fibonacci sequences, as shown in Figure 2.12. Using the measurements denoted in Figure 2.9, we calculate the points of the ear in its surface plane. Since there is no specification for the location of the ear tragus point, we take the assumption that it is located at the middle of the edge of the ear concha. In order to validate the approximation result,

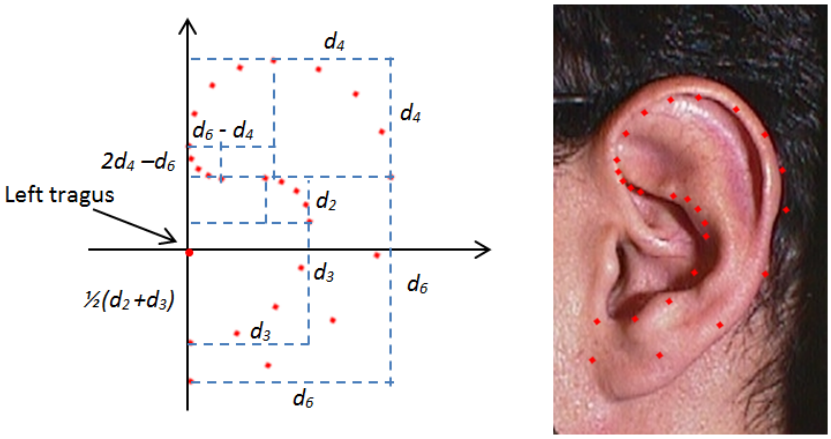


Figure 2.12: 3D head point cloud

we use the photos in the CIPIC database to do benchmarking. The two Fibonacci spirals fit the ear curves quite well.

After that we define the coordinate system (x, y) in the surface plane, we project the points into the three planes that we build for the head. The corresponding angles are $(0, \sigma_1, \sigma_2)$ for the front, lateral and horizontal planes. The corresponding points (x,y) in 3D is $((x\cos\theta_1 + y\sin\theta_1)\cos\theta_2, (x\cos\theta_1 + y\sin\theta_1)\sin\theta_2, (-x\sin\theta_1 + y\cos\theta_1))$. Then we can correlate the ear coordinate system to the head coordinate system by a translation of $(x_2/2, 0, 0)$ for the left ear. The right ear is a reflection of the left ear by lateral plane.

2.3.5 Transformation

After we discover general relationships between anthropometric data and features, the personalized features can be predicted by knowing the anthropometric data. The features are the marks and we use them to shift a reference H_l to a target one. We apply the shift in the linear frequency domain. For each azimuth, the features in H_l give certain frequencies. For instance, features in a reference H_l at 0 azimuth give frequencies $[f_1, f_2, f_3]$ and in a target H_l give $[\tilde{f}_1, \tilde{f}_2, \tilde{f}_3]$. We fix boundary points (0 and $fs/2$) and shift f_1 to $\tilde{f}_1$, f_2 to $\tilde{f}_2$ and f_3 to $\tilde{f}_3$.

$$\tilde{H}_l(\tilde{f}, \theta) = H_l(f, \theta)$$

With conditions:

$$\begin{cases} \tilde{H}_l(0, \theta) = H_l(0, \theta) \\ \tilde{H}_l(f/2, \theta) = H_l(f/2, \theta) \\ \tilde{H}_l(\tilde{f}_i, \theta) = H_l(f_i, \theta), \quad i = 1, \ldots, N_f \end{cases}$$

N_f is determined by the features and azimuth angles. One of the methods to tackle this problem is to apply linear interpolation. First we stretch or compress the interval:

$$\tilde{f} = \frac{\tilde{f_{k+1}} - \tilde{f_k}}{f_{k+1} - f_k}(f - f_k) + \tilde{f_k}, f \in [f_k, f_{k+1}]$$

Then:

$$\tilde{H}_l(\tilde{f}, \theta) = \tilde{H}_l(\frac{\tilde{f_{k+1}} - \tilde{f_k}}{f_{k+1} - f_k}(f - f_k) + \tilde{f_k}, \theta), f \in [f_k, f_{k+1}]$$

2.3.6 Equivalent rectangular bandwidth

The equivalent rectangular bandwidth or ERB is a measure used in psychoacoustics, which gives an approximation to the bandwidths of the filters in human hearing, using the unrealistic but convenient simplification of modeling the filters as rectangular band-pass filters. The center frequencies of ERB are defined by:

$$f(b) = QL(\exp^{b\gamma/Q} - 1)$$

With b the parameter band number, γ the warp factor, Q the filter Q-factor, L the minimum bandwidth. The bandwidth of the auditory filters can be described with $\gamma = 1$, $Q = 9.265$ and $L = 24.7$. Then before 24 kHz, we have 22 center frequencies, see Table 2.5. We define the

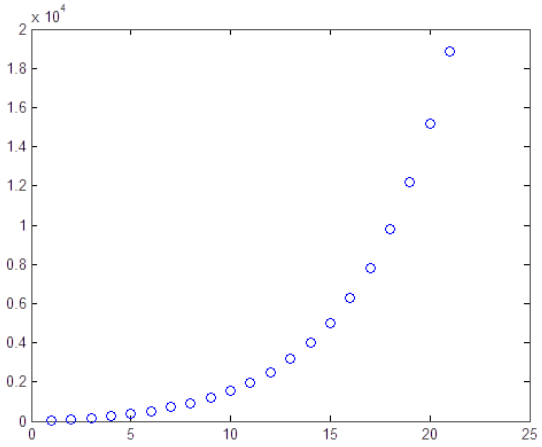


Figure 2.13: Equivalent Rectangular Band

Table 2.5: Equivalent rectangular band

b	center frequency	bandwidth
1	26	55
2	88	68
3	164	85
4	258	105
5	376	131
6	521	162
7	702	201
8	926	250
9	1205	310
10	1550	385
11	1979	477
12	2511	593
13	3171	735
14	3990	912
15	5006	1132
16	6267	1405
17	7833	1744
18	9775	2164
19	12185	2685
20	15176	3332
21	18888	4135
22	23493	2934

bandwidth that is corresponding to f(b) as:

$$w(b) = f(b + 0.5) - f(b - 0.5)$$

And the subbands B are

$$B(b) = \begin{cases} [0, f(1 + 0.5)] & b = 1 \\ [f(b - 0.5), f(b + 0.5)] & b = 2, \ldots, 21 \\ [f(22 - 0.5), 24000] & b = 22 \end{cases}$$

In the evaluation state, we assume that humans cannot distinguish sounds with same intensity and frequency that are located in the same ERB band.

Chapter 3

Statistical Method

As we mention in Chapter 2, data isS obtained from two different sources. CL Louvain of Philips provides anthropometric data and Applied Sensor Technology of Philips measured HRTFs from subjects. The CIPIC database is the other data source which contains both anthropometric and HRTFs of 45 subjects. The internal correlations for one kind of dataset are investigated first. It is necessary and important to inspect whether the samples from different databases provide a good representation of the whole dataset. Therefore, no matter the statistical method or the first principle method we apply, the variation within the sample set needs to be investigated. The differences between anthropometries of individuals contain valuable information for HRTF personalization. Compared to the wavelength, the torso, head and ear are supposed to play different roles in affecting the sound waves. The relationships between these anthropometries can provide first conclusions that are determinative in procedures followed.

3.1 Data

3.1.1 Anthropometric Data

From Philips Research, CL Louvain measured anthropometric data from 150 subjects with 17 different head and ear properties. Applied Sensor Technology measured HRTFs from 51 subjects, in which there are 13 subjects whose anthropometric data are contained in the CL Louvain database. The advantage of the CL Louvain database is that all the ear properties are related to ear canal (ear tragus point), thus these measurements can be real influence factors reflected in the HRTF. The disadvantage is that the number of subjects, whose anthropometric data and HRTF are both available, is small.

In the CIPIC database, anthropometric measurements and corresponding HRTFs from 45 subjects are contained. The advantage of the CIPIC database is that the number of usable data is enough. The disadvantage is that the ear related anthropometric data is not measured with the ear tragus point. Fortunately, there are 31 subjects whose ear photos are included. From these photos, usable anthropometric data can be derived. Thus we analyze these 31 subjects in the CIPIC database.

3.1.2 HRTF

The HRTF measurements, either in the CL Louvain database or in the CIPIC database, are obtained from laboratory. In Chapter 2, we describe the head related impulse response data. By using Fourier transformation, HRIR is transformed into HRTF with frequencies from 0 Hz to 24000 Hz.

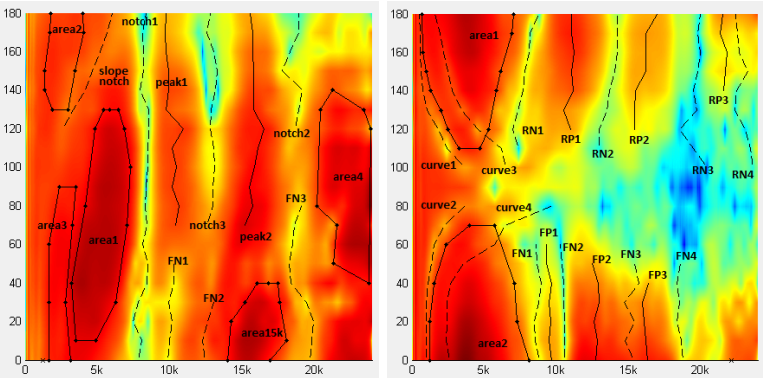


Figure 3.1: Annotated features in HRTF graph

3.1.3 Feature of HRTF

Features are extracted from HRTF, see Figure 3.1. We define feature value as the value of a mathematical feature, such as frequency, area or slope. For instance, the average frequency of notch 1 in ipsi-lateral HRTF is a feature value. We annotate 34 different features and calculate 143 feature values from them.

3.2 Result from Statistical Method

3.2.1 Anthropometric data

Mean, Standard Variance and Coefficient of Variation

The mean values, variance, and the coefficients of variation (CV) are calculated to investigate whether the 13 samples form a good representation of the whole group of 51 subjects in the CL Louvain database. Table 3.1 gives the results for each anthropometric property. The mean of most properties of 13 samples has a difference of less than 6% from the whole group. Tilt Outward Skull decreases 28%, which is considered as a bad representation. Tilt Backward LV, which has a 18 % difference, should be ruled out. For the measurement of distance of rear head, the result measured with the Frankfurt (FF) plane is better than the Leuven (LV) one.

The variance of most properties of 13 samples is smaller than the whole group. This means that the samples, that are far away from the center, are not covered.

The distribution of a property can be suggested by its CV and then fitted. For instance, Circ. Top Head FF is likely to have a normal distribution, and then is fitted with its mean value and variance. With a small CV, the distribution is more like normal. With a CV close to 1, the distribution is more like an exponential one.

We also investigate the mean and variance of several properties in the CIPIC database. These properties include head width, height, depth and circumference and pinna width, height and flare angle. They are the same as or related to the properties we investigate in the CL Louvain database. The mean of head width stays almost the same in the two databases. The mean of head depth and head circumference are approximately twice of Dist. Rear Head FF and Circ. Rear Head FF. For ear length (pinna height), the difference is smaller than 5%. The consistency means that these properties are not largely affected by the different measurement methods and conditions. However, the mean of ear width is not close to the pinna width. This difference may be caused by different measurement methods since the former one is measured at the beginning of the tragus point and the latter one not.

Table 3.1: Mean, variance and coefficient of variance (CV) of anthropometric data in the CL Louvain database

Properties	μ_{13}	σ_{13}	CV_{13}	μ_{all}	σ_{all}	CV_{all}	d_μ	d_σ
Circ. Top Head FF	35.63	1.36	3.8%	35.28	1.71	4.8%	1%	20%
Circ. Top Head LV	35.07	1.27	3.6%	34.56	1.71	4.9%	1%	25%
Circ. Rear Head FF	28.12	1.15	4.1%	27.38	1.79	6.6%	3%	36%
Circ. Rear Head LV	26.63	1.20	4.5%	26.13	1.81	6.9%	2%	34%
Dist. Top Head FF	13.11	0.73	5.5%	12.71	0.97	7.6%	3%	25%
Dist. Top Head LV	12.81	0.79	6.2%	12.34	0.93	7.5%	4%	14%
Dist. Rear Head FF	9.23	1.27	13.8%	8.70	1.22	14.0%	6%	5%
Dist. Rear Head LV	8.57	1.23	14.4%	7.74	1.16	14.9%	11%	6%
Head Width	14.52	0.60	4.1%	14.70	1.08	7.4%	1%	45%
Ear Length	6.67	0.40	5.9%	6.35	0.48	7.6%	5%	18%
Ear Width	3.50	0.26	7.6%	3.53	0.40	11.2%	1%	33%
Ear Ridge	2.18	0.22	10.2%	2.10	0.24	11.6%	4%	9%
Ear Protrusion	2.31	0.36	15.7%	2.18	0.38	17.4%	6%	4%
Ear Gage	8.62	1.45	16.8%	8.43	1.18	14.0%	2%	23%
Tilt Outward Skull	3.67	4.23	115.1%	5.06	3.71	73.2%	28%	14%
Tilt Outward Ear	11.88	4.53	38.1%	12.66	4.47	35.3%	6%	1%
Tilt Backward LV	4.64	4.44	95.7%	5.66	5.82	102.9%	18%	24%

Table 3.2: Mean and variance in the CIPIC database

Properties	μ	σ
Head Width	14.53	0.88
Head Height	21.34	1.35
Head Depth	19.94	1.98
Head Circumference	57.44	6.01
Pinna height	6.45	0.26
Pinna width	2.92	0.08
Pinna flare angle	0.46	0.01

Table 3.3: Correlations between the properties higher than 0.6

Property 1	Property 2	Correlation
Circ. Top Head FF	Circ. Top Head Leuven	0.96
Circ. Top Head FF	Dist. Top Head FF	0.62
Circ. Top Head FF	Dist. Top Head Leuven	0.60
Circ. Top Head FF	Head Width	0.64
Circ. Top Head Leuven	Dist. Top Head FF	0.64
Circ. Top Head Leuven	Dist. Top Head Leuven	0.62
Circ. Top Head Leuven	Head Width	0.68
Circ. Rear Head FF	Circ. Rear Head Leuven	0.90
Dist. Top Head FF	Dist. Top Head Leuven	0.92
Dist. Rear Head FF	Dist. Rear Head Leuven	0.85

Correlation within Anthropometric Data

A correlation analysis is performed for all the anthropometric properties in the CL Louvain database to find out the potential internal relationship. As shown in Table 3.3, we analyze the pairs that have a correlation larger than 0.6. The measurements based on two different reference planes (FF and LV) are highly correlated. The circumference over the top of the head is correlated (0.6) with the distance from the ear tragus point to the top of the head. This high correlation is expected, because usually a longer top distance leads to a longer top circumference. However, no high correlations are found between head related properties and ear related properties.

A similar correlation analysis is conducted with the properties in the CIPIC database. Between different ear related properties, the correlation is high. This is also true for the the head related properties. But no high correlation is found between ear and head properties.

The statement 'big head leads to big ear' can not be verifies, since there are no clear relationships between the ear-related properties and the head-related properties. Thus, they both can be used as personal information. Although there are correlations between some head-related properties, they should not be replaced and may influence the sound wave in different ways. The measurements based on two different planes show high correlation and thus we can only take the measurements under one plane, say FF, because the definition of FF is clearer and measurements of 13 samples with FF plane is a better representative from mean value analysis. A single parameter for head size is likely sufficient for more detailed personalization schemes. The conclusion is that the dimension of the anthropometric properties in the CL Louvain database can be reduced to 13.

Principal Component Analysis

In the anthropometric database, data of 150 individuals is measured. As said, the measurements can be divided into two main parts, one for the ear and one for the head. They are analyzed separately.

The measurements for the ear include 7 items: ear length, ear width, ear ridge length, ear protrusion length, ear gage length, the angle of tilt outward of the ear, the angle of tilt backward of the ear. The first five are measured in centimeters and the last two in degrees. Applying PCA on the ear data, 7 principal components are obtained. The first four PCs contribute 0.432, 0.195, 0.148 and 0.143, respectively, to the total variation. They account for 92% of the information of the ear data. The first PC contains 43% of the information, and is contributed by all of the ear items. The first 5 measurements contribute almost equally and significantly more than the last two measurements. We conclude that the first PC is the component describing the measurements of the length and the shape of the ear and it contains the largest variance of all ear related data. The angle measurement of the ear is the second important component.

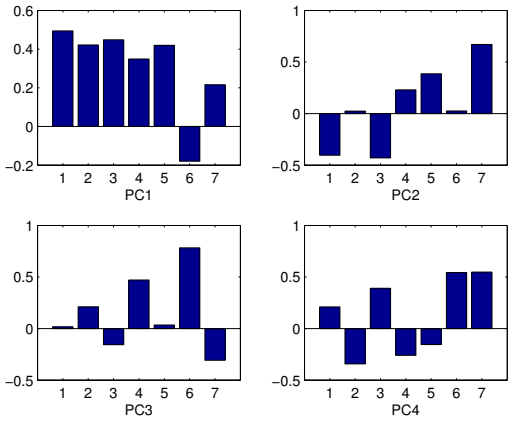


Figure 3.2: The first 4 principal components of the ear properties

The measurements for the head include 10 items: circumference over the top head , rear head, and the distance between the top head and the rear head, measured by two different methods, head width, and the angle of tilt outward of the skull. The first nine are measured in centimeters and the last one in degrees. Of course, correlation analysis reveals that each pair of measured data from the two different methods shows high correlation. We delete the repeated so that six measured items remain. The first four PCs contribute 0.411, 0.223, 0.151 and 0.090 respectively,

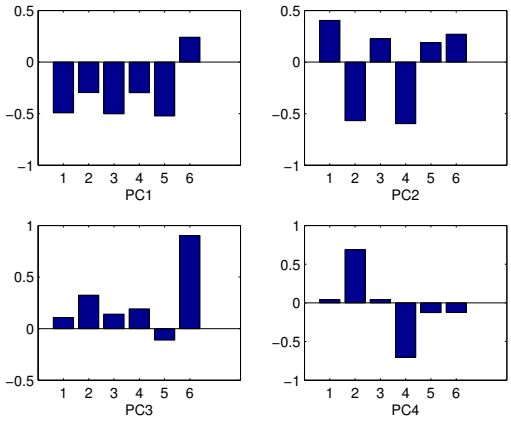


Figure 3.3: The first 4 principal components of the head properties

to the total variation. They contain 87% of the information of the former head data. The first PC contains 41%. It relates mainly to the lengths of the different parts of the head. Besides, the measurements of the front vertical plane contribute more than the measurements of the horizontal plane. The second PC containing 22% relates to the measurements in the horizontal plane. The third PC containing 15%, emphases the angle of tilt outward of the skull. In conclusion, the length measurement of head in the front horizontal plane contains most variance information, and is followed by the length measurement of head in the horizontal plane. The angle measurement is the third influence factor.

3.2.2 HRTF

Principal Component Analysis

The HRTF data for one individual at 19 different azimuth angle can be also analyzed with this method. The idea is to find out the largest contribution w.r.t degree range of the speakers or the largest contribution w.r.t frequency range of the HRTF.

PCA is applied on the data set H, which is formed by the HRTFs from 19 different azimuth angles. The first four principle components of the ipsi-lateral HRTF distinguish the directionality.

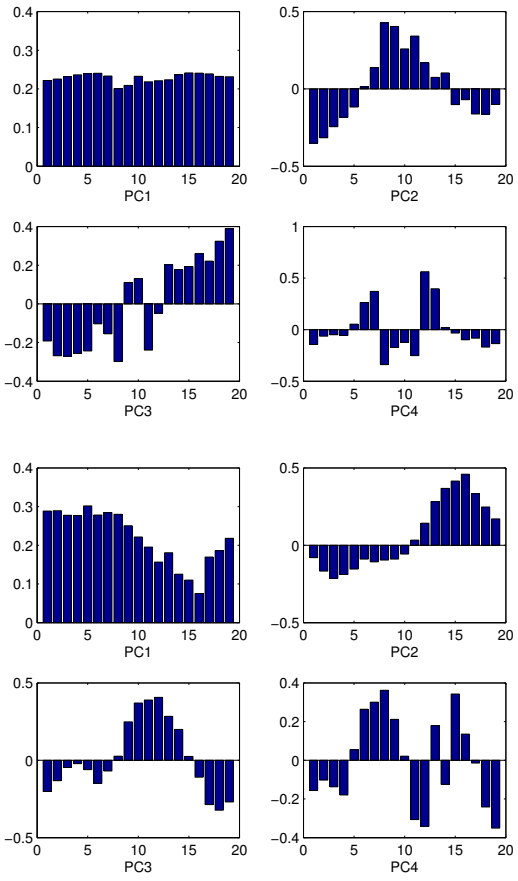


Figure 3.4: The first 4 principal components of HRTF respect to azimuth

They account for 64.5%, 19.9%, 5.1% and 3.4%, respectively. The interpretation is that the second principle components' weights (PCW) distinguish the front and rear part and the third PCWs distinguish the central from the lateral part. In contrast, the PCWs of the contra-lateral HRTF distinguish the variation of the central-lateral (second PC) more than the one of the front-back (third PC). This observation is consistence with the result in [11].

3.2.3 Features

Coefficient of variation

We calculate the coefficient of variation (CV) of the features. Generally, the frequency related features (as the average frequency, min frequency, etc.) as well as the slopes of notches and peaks show a small variation (mostly smaller than 10%). The CVs of the energy levels of most of the notches and peaks are larger than 50%, which shows that the energy level is not as stable as the

frequency related features. The areas of these spatio-peaks also vary a lot since the annotations are more subjective.

Specially, the CV of the intersection of the two curved notches in the contra-lateral HRTF is 5%, and its mean value is 92 degree. This can be considered as supporting evidence of the similarity between the frontal part and rear part of the contra-lateral HRTF.

The distributions of the features are firstly suggested by their CV. For most of them, normal distributions are suggested. For instance, the distribution of the average frequency of Notch 1 in the ipsi-lateral HRTF is fitted with a normal distribution. With a mean value μ_x and standard variance σ_x,

$$p(x) = \frac{1}{\sqrt{2}\pi\sigma_x} e^{-\frac{(x-\mu_x)^2}{\sigma_x^2}}$$

The energy level of these notches and peaks show a different distribution. They seem to follow exponential distributions or gamma distributions. The non-normal distribution gives an explanation that these features have a long tail distribution, since in the exponential distribution, the mean value equals the standard variance, which leads a CV of 1. For example, the energy level of Notch 1 in the ipsi-lateral HRTF, the distribution fitting is:

$$p(x) = \lambda e^{-\lambda}$$

where $1/\lambda$ is its mean value.

Correlation Analysis

By performing a correlation analysis, some internal feature relationships are found. For example, in the ipsi-lateral HRTF, the average frequencies of Notch 1 and Rear Peak 1 have high correlation (0.63) and in the contra-lateral HRTF, the notches and peaks that appear in frontal part also have high correlation with each. The same goes to the notches and peaks which are in rear part of the conra-lateral HRTF. The energy levels of these spatio-peaks are highly positively correlated to their area and their maximum amplitude points. Clearly, a larger area or higher amplitude leads to a higher energy level. The minimum and maximum frequencies of a spatio-peak are also correlated. Occurrence of highly correlated features may be caused by one and the same anthropometric property or by different correlated properties.

Principal Component Analysis

Before we apply PCA on the features, the features whose CV is close to 1 or larger than 1 are ruled out, since as analyzed, they may be not annotated well. There remain 53 frequency or slope related features.

No principal component with large variance of feature exist. The first principal component contains 16% of the feature information and the second contains 13%. 17 PC are needed to obtain 85% information.

3.2.4 Correlation between Anthropometric Data and Features

The correlation analysis is applied to find high correlation between anthropometric data and HRTF features. We analyze the result with an absolute correlation value higher than 0.6.

Head-related anthropometric data

Although the circumference over the top of the head is measured by two different methods, both are highly relevant. This leads to the same group of highly related features that they share. In this case, only one of these two anthropometric data sets is analyzed namely the first one, measured according to FF plane. The same strategy is taken for the following three groups. The correlation analysis shows that the top head circumference has positive correlation with the energy levels of

Table 3.4: Part of features

Property	Feature	mu	sigma	CV
Average Frenquency	iN1	8408.87	612.71	7%
	iN2	13362.94	903.33	7%
	cRN2	13302.75	934.67	7%
	cRP2	15494.43	1075.84	7%
	cFN1	7667.70	617.31	8%
Max Point (Amplitude, db)	iPA3	-21.89	1.56	7%
	cPA1	-21.85	1.55	7%
Max Point (Azimuth, deg)	iPA3	156.67	11.43	7%
	cPA1	178.63	4.91	3%
Max Point (Frequency, Hz)	iPA15	15276.25	822.73	5%
Min Frequency (Hz)	iPA15	14022.58	835.26	6%
Max Frequency (Hz)	iPA1	7517.98	673.37	9%
	iPA15	16845.19	1026.39	6%
	cPA2	7048.37	662.19	9%
Min Azimuth (deg)	iPA3	130.59	9.88	8%
	cPA1	109.41	6.45	6%
Max Azimuth (deg)	iPA3	179.41	2.38	1%
	cPA2	69.61	5.99	9%
Slope (deg)	iPA1	84.43	4.19	5%
Aver. Slope	sn	70.67	6.03	9%
Frequency to edge	c1	530.50	18.41	3%
Max Curvature FT	c1	7.15	0.29	4%

Notch 1 and Peak 1 in the ipsi-lateral HRTF. It means that if the circumference over the top head increases, the energy levels of Notch 1 and Peak 1 of the ipsi-lateral increase as well. By checking the correlations of the average frequencies of these notches and peaks appearing in the contra-lateral, although values are small, we found negative signs. Also the rear head circumference has clear influence on the rear part of the contra-lateral in the low frequency range.

In the previous section we conclude that the correlation between the top distance of the head and the top circumference of the head is high (more than 0.6) and that they share some common features, especially, in the low frequency range.

The frequencies at 0 azimuth of Spatio-Peak 2 and Curve 2 in the contra-lateral HRTF have high negative correlation with the rear distance of the head. We conclude that the frequencies of the features in the frontal part decrease if the value of Circ. Rear Head or Dist. Rear Head increases.

Ear-related anthropometric data

Ear width is correlated with the frequencies of notches and peaks in the medium frequency range (at about 10 kHz) in the front part. Ear Ridge affects the high frequency range in the rear part. Our correlation analysis shows that the slope of Sloped Notch has high correlation with the protrusion of the ear. This seems reasonable because it happens only in the rear part of the ipsi-lateral HRTF.

The tilt outward skull has a positive influence on the energy level of Spatio-peak 1 of the ipsi-lateral HRTF. It seems that the angle of the tilt outward ear will affect the sound capture of the contra-ear negatively. The low and medium frequency part of the contra-lateral HRTF decreases. The high frequency part of the ipsi-lateral HRTF increases.

We also apply a correlation analysis to the new features suggested from PCA and anthropometric data. Unfortunately, no high correlation is found for the first several principle components. Also, the correlation calculation between the new anthropometric data (first PC) and features

does not provide high values.

3.3 Conclusion

From the results, the ear-related properties and the head-related properties are independent from a statistical view point. They both should be used as personal information. Some head-related properties show correlation with each other.

The 13 samples follow the same statistics for most of the anthropometric properties, such as ear width and ear length. They do not follow the statistics for these properties related to the angles (Tilt Outward Skull, Tilt Outward Ear, Tilt Backward Leuven). They show prominent variations and are considered as unreliable representations.

For the features, the 13 samples show variation for the frequency related features, the slopes of notches and peaks. The coefficients of variation of the energy level of the notches and peaks are large, which shows that the energy level has high variation.

The statistical investigation also show the consistency of the anthropometric and HRTF data in different database.

3.3.1 Lower dimensional representation

The result of PCA applied to different objects can provide lots of information. The anthropometric data application gives the insight how to combine these different anthropometric items and form independent ones. A linear combination of the anthropometric data forms a new physical or nonphysical part, which could not be known intuitively, to influence some features. This is applied to do a correlation analysis to the features to discover the mutual relationships.

Also, PCA helps to discover some interesting phenomenon, like the results when applying PCA to the speaker azimuths with the frequency domain. The principle components are interpretable. The weights of the first PC show the difference between ipsi-lateral and contra-lateral. The second ones describes the difference between the front part and the rear back. The third ones likely emphasize the central-lateral differences. This shows how the different azimuth ranges of the speakers combine to influence the spectral cues on HRTF.

3.3.2 Mutual relationships

The correlation analysis reveals relationships that physics can explain. The acoustic physics suggests that the ear anthropometric properties have more influence on the features above 3 kHz than on the features in the lower frequency part, which is influenced mainly by head anthropometric properties. The rear head circle has clear influence on the rear part of the contra-lateral in the low frequency range. When the circumference of the rear head increases, the average frequency of the notch and the Spatio-Peak 1 shift decrease. Since the circumference of the rear head is comparable to the sound wave with a frequency of about 1 kHz, this is reasonable. A larger ear width makes the frequency of notches and peaks that appear at about 10 kHz in the front part decrease. It has more apparent influences on the medium frequency range. This makes sense because the wavelength related to a frequency of 10 kHz is 3.4 cm and the ear width (3 cm) can affect it.

In accordance with physics, we find that the frequency of the notches and peaks is negatively correlated with the anthropometric data. By checking the correlations of the average frequencies of these notches and peaks appearing in the both laterals around 10 kHz, we find negative value with the ear properties.

Unfortunately, the lower dimensional features, derived from PCA, do not show high correlations with the anthropometric data. From acoustic theory, the relationship between the original anthropometric data and the original HRTF features seem more reasonable. Several correlated anthropometric properties and features are listed in Table 7.1.

Table 3.5: Correlated anthropometric data and features

Lateral	Anthropometry	Feature
Ipsi	Ear Protrusion	Sloped Notch
Ipsi	Ear Ridge	Notch 2
Contra	Ear Width	Front Notch 1
Contra	Dist. Rear	Curved Notch 2

Their relationships cannot be confirmed by the theory in a statistical way because of the limitation set by the small database. Theory confirms that these anthropometric data and corresponding feature behave rightly and interpretably. Then the relationships derived from the statistical method can be applied to build the models, which can predict the features from the personal anthropometric data.

Chapter 4

First Principle Method

4.1 Experiment setup

Lack of data is a problem we encounter in this research. Although the original data sets in the CL Louvain database are not small, the data set is too small to analyze the relationships. Therefore, we go deeper into the physics behind. To figure out the theoretical relationships between the features and the anthropometric properties, experiments are designed. We can modify the ear properties and observe the change of the features in the HRTF sets. Also, we can construct the ear step by step to validate of the corresponding relationships.

4.1.1 Hypothetical Relationship

In order to perform the analysis, the data should be extracted from the original database first. The anthropometric data included in the CIPIC database is mostly not suitable for examining effects on HRTFs, as it is not measured with relation to the ear canal. Therefore, more suitable data is measured from the subjects' ear photos. All these original pinna images have as reference either a ruler or a black strip of 5 cm long. Since the research focuses on the change of azimuths in the horizontal plane, the horizontal measurements, like the distances from the tragus point (used to indicate the location of the ear canal) to ear helix, to concha posterior wall, are expected more relevant to our research. These measurements are extracted from these 2D images, in this way the influence of the ear protrusion angle is not included in our measurements. Luckily, the information of ear protrusion is available in the CIPIC database. It can be used to adjust the result later.

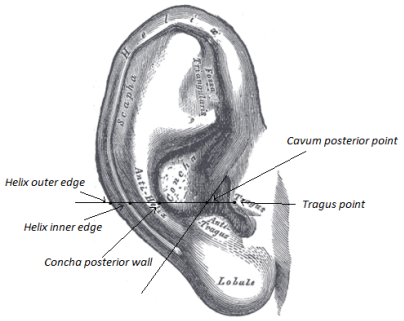


Figure 4.1: Measured points of the pinna

Table 4.1: Means and variances of m_1, m_2, m_3 and m_4

Feature	m_1	m_2	m_3	m_4
Mean	1.05	1.58	2.23	2.81
Variance	0.18	0.30	0.30	0.30

Table 4.2: Means and variances of path length differences derived from notch frequencies

Feature	iN1	iN2	iFN1	iFN2	iSN	cRN1	cRN2	cFN1	cFN2	cFN3
Mean	2.25	1.15	1.53	1.02	2.85	1.89	1.09	2.12	1.51	1.01
Variance	0.25	0.1	0.16	0.09	0.39	0.23	0.12	0.28	0.17	0.09

There are 31 pinna images available in the CIPIC database. For each subject, we measure 4 different horizontal distances, which start from the tragus point and end at cavum posterior point, concha posterior point, helix inner edge and helix outer edge, respectively, as showed in 4.1. They are denoted as m_1, m_2, m_3, and m_4, respectively. Generally we have a relationship as

$$m_1 \leq m_2 \leq m_3 \leq m_4$$

Since only the left pinna images are available (excluding subject 003, whose right pinna image is available only), we assume that the left and right ears are similar.

The mean values and variances of some anthropometric data and features in HRTFs are analyzed. Then hypothetical relationships are formulated.

We use Simone's model [9]. According to acoustic theory, when the path difference from the direct wave and reflected wave equals half of a wavelength, the destructive interference occurs. A notch may appear on the HRTF. At the 0 azimuth, the path difference d equals to twice of the measurement m, so the formula is:

$$f = \frac{c}{2 \times d} = \frac{c}{4 \times m}$$

Where c is the sound speed in air (340 m/s) and d is the path difference. However, according to [10], since the impedance of the pinna is larger than the air impandence, there may be a boundary created because of the impedance discontinuity, which could produce its own reflection and ultimately reverse the phase of the wave. In this case, a delay of half a wavelength would not produce notches in the spectrum any longer. Instead, destructive interference would appear for full-wavelength delays.

$$f = \frac{c}{d} = \frac{c}{2 \times m}$$

Here d is the path difference, which could double some measurements of the ear properties m like m_1, m_2, m_3, m_4. We extract the average frequency of the main notches from the HRTF sets and calculate the corresponding half path difference. The results are listed in Table 4.2.

When comparing the values of Table 4.1 and Table 4.2 in Figure 4.2, several features which have similar mean values can be grouped based on their mean values, as shown in Table 4.3.

We assume that the features which have similar mean values of the distance are formed by the corresponding anthropometric property. For example, the Notch 1 in the ipsi-lateral HRTF is considered formed by the destructive interference of the path difference m_3. Front Notch 1 may be formed by the path difference m_2. However, due to the large confidence intervals, these attributions are likely to contain errors. We design an experiment to check these hypotheses.

Table 4.3: Means and variances of measurements and features

Feature	m1	iFN2	cFN3	cRN2	m2	iFN1	cFN2	m3	iN1	m4	iSN
Mean	1.05	1.02	1.01	1.09	1.58	1.53	1.51	2.23	2.25	2.81	2.85
Variance	0.18	0.09	0.09	0.12	0.3	0.16	0.17	0.3	0.25	0.3	0.39

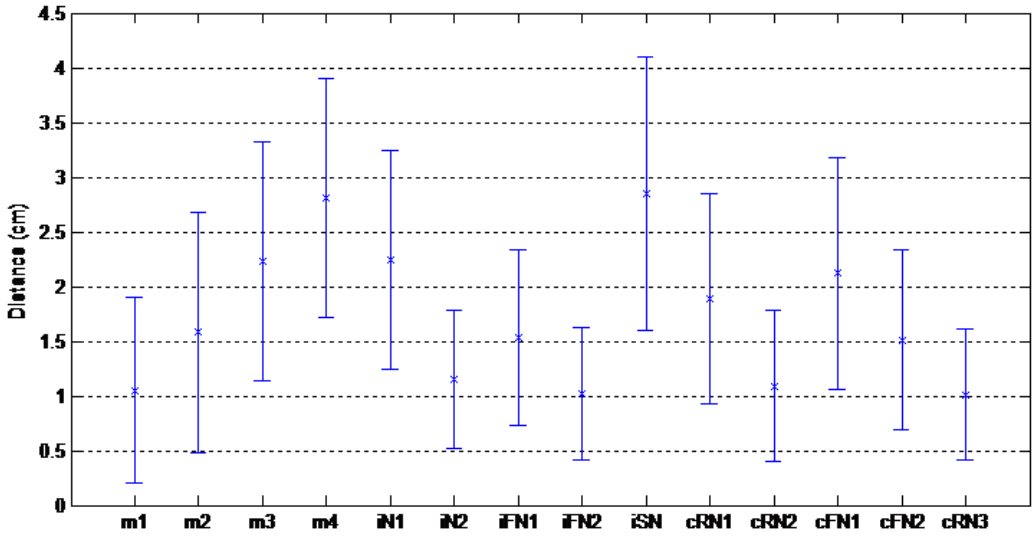


Figure 4.2: Mean distance (measured or calculated from frequency) and 95 % confidence intervals

4.1.2 Strategy

The experiments are conducted in an anechoic room. We use the $B\&K$ Head and Torso Simulator as the experiment subject. The original right ear from the dummy head is considered the control group. The experimental groups are the ears that we change. In order to investigate the effects of different ear structures, we modify only one part at a time.

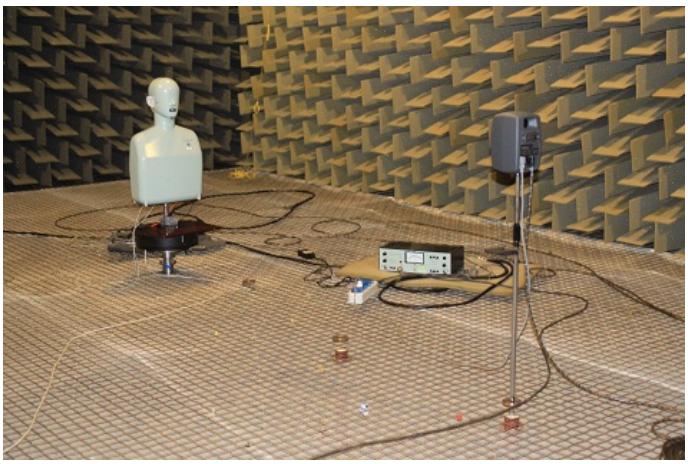


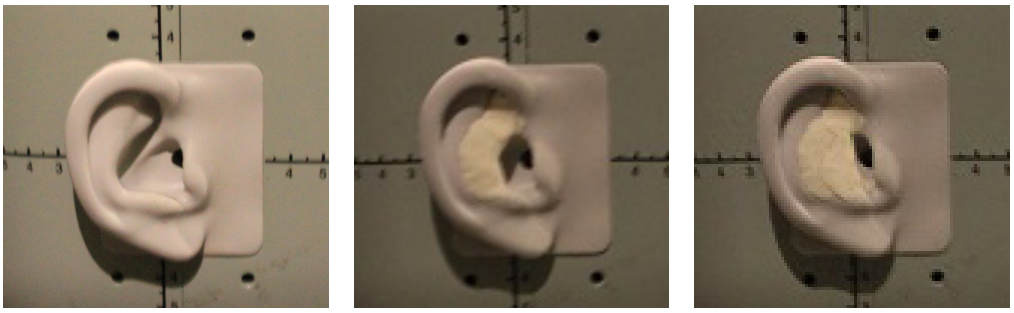
Figure 4.3: Experiment setup

Modification Experiment

Since we expect that the notches are formed by the ear helix, concha and other details in the outer ear, we can use clay to reshape the ear details or block some parts of the ear and observe the difference between their HRTFs.

Experiment 1 is focused on the ear concha. The change of the cavity of the concha affects the distances m_1 and m_2. It is designed to know the effect of the ear concha. The change of the

ear concha may cause different changes of the physical phenomenon. By decreasing the cavity, see Figure 4.4, which also decreases the distance from the ear canal to the concha posterior wall, some changes can be observed. Experiment 2 is focused on the ear helix. Making the helix



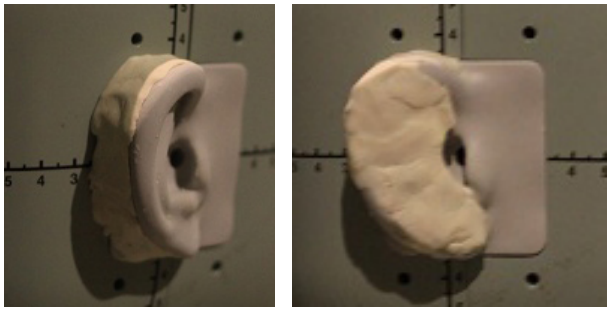
(a) (b) (c)

Figure 4.4: (a) Control group, (b) experimental group with small concha, (c) with smaller concha

thicker or blocking its valley changes the distance m_3 and m_4, see Figure 4.5. Experiment 3 is

Figure 4.5: ear with upper helix blocked

conducted to investigate the function of the ear back part. The space of the ear back part may influence the sloped notch, which is connected to the distance m_4, see Figure 4.6. Experiment 4



(a) (b)

Figure 4.6: (a) Back ear blocked, (b) ear only with concha

is designed to test the influence of the shoulder. The control experiment is the original dummy without any cover on its shoulder. The experiment group is a dummy with a sound absorbing material on its shoulder. Experiment 5 is designed to figure out the influence of the ear canal and its corresponding physical phenomenon. When the ear canal is blocked, only the sound wave with

Table 4.4: Construction procedure

Step	1	2	3	4	5	6
Ear component	Canal	Tragus	Concha	Helix	Cymba wall	Lobe

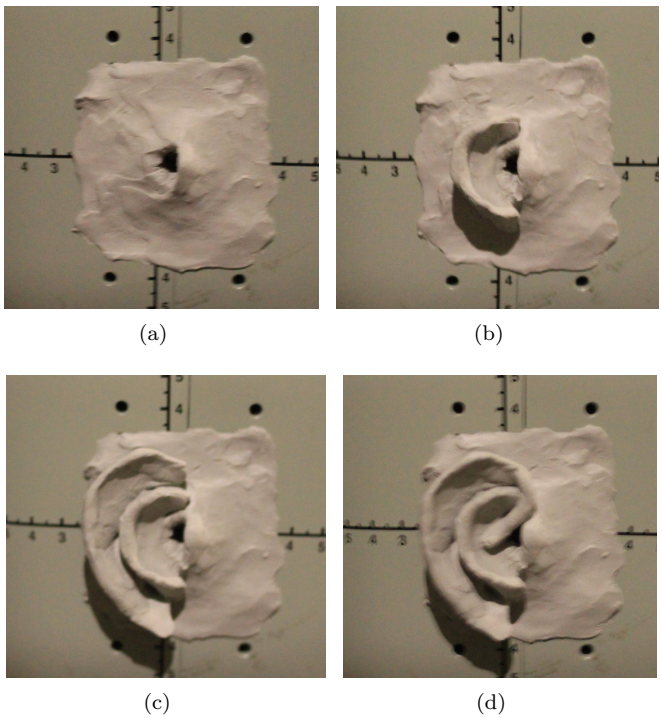
Table 4.5: Construction procedure

Control	Experimental 6	
Concha	Larger concha	Larger concha

large wavelength can cross the clay and get into the ear canal. The resonance is expected to be obvious.

Construction Experiment

By using the clay, an artificial ear can be constructed from the very beginning. First, the ear canal can be built. The length of it also can be modified to observe the change of the phenomenon it causes. This is not easy to be implemented using the dummy ear. Furthermore, the details of the ear can be added step by step. The ear concha can be added. Also the length of it is modified to test the change of the feature in HRTF. At last, a similar ear can be constructed. With at each addition step, a corresponding feature is expected to appear. It helps us to understand the HRTF structure. Part of the procedure is shown in Figure 4.7. Experiment 6 is focused on the effect



(a) (b)

(c) (d)

Figure 4.7: (a) Canal, (b) concha (c) helix (d) cymba wall

of the ear concha in the construction procedure. It only contains concha and the corresponding HRTFs are expected to have a simple structure and obvious change during the experiment.

4.2 Experiment Results

The features in the HRTFs from the experiments show continuity. We assume that they are caused by the same phenomena. This property allows us to focus only on the features in ipsi-lateral HRTF if the similar features appear in both laterals. Because comparing to the ipsi-lateral HRTF, the contra-lateral HRTF is less influenced by the change of the azimuth. The sound waves from ipsi-lateral azimuthal positions travel along the head. It only distinguishes the sound that is coming from the frontal part or the rear part.

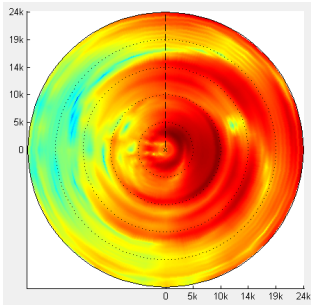
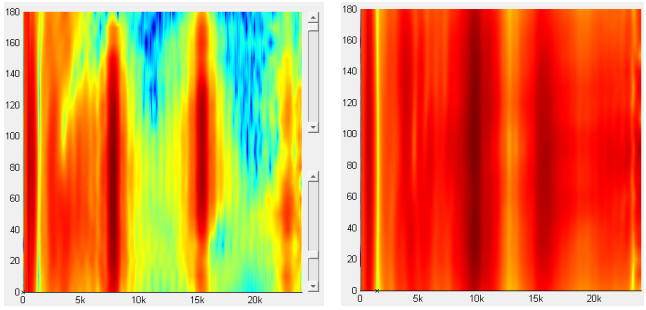


Figure 4.8: HRTF in coordinates

4.2.1 Ear canal

The result shows that a peak around 7 kHz is formed by the ear canal. Normally, the resonance caused by the ear canal appears around 5 - 8 kHz. The dummy ear canal, with a length of 1.1 cm, causes a resonance around 7.8 kHz. In the construction experiment, the artificial ear canal has a length of 0.9 cm and forms a peak around 10 kHz.

When the entrance of the ear canal is blocked in Experiment 5, the destructive interferences is reduced, see Figure 4.9. In contrast, the resonance phenomenon is displayed clearly. However, the interference below 7.5 kHz, especially the sloped notch, has not been affected much. The figure can be viewed as a basic resonance structure in the whole HRTF set. Some of the differences are caused by the interferences.



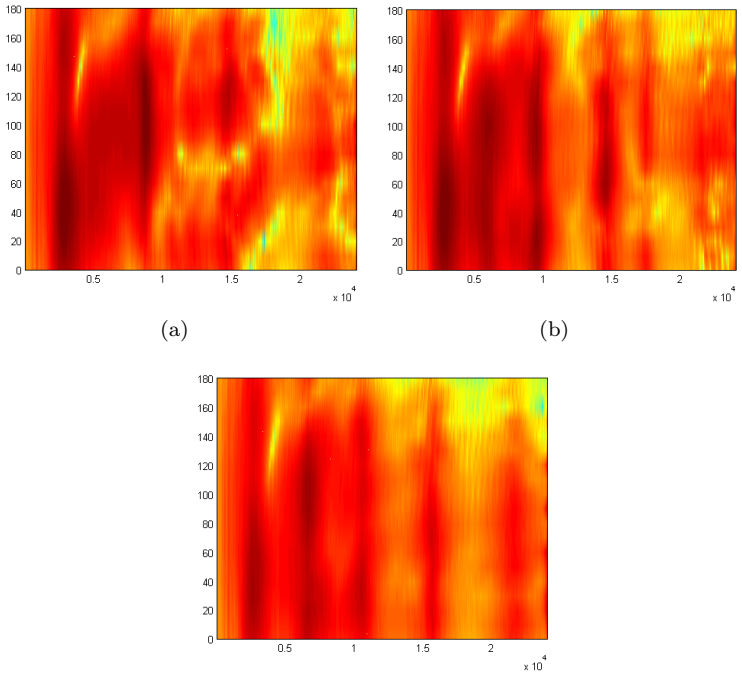
(a) (b)

Figure 4.9: (a) Ear canal being blocked, (b) containing ear canal only in construction experiment

4.2.2 Ear Concha

The result shows that the ear concha is responsible for a notch appearing around 10 kHz. In Experiment 1, we performed two consecutive experiments. In the first one, we used clay to increase the thickness of the concha posterior wall, which means the distance from the tragus

point to the posterior wall decreases. In the second one, we continued to increase the thickness. The obtained HRTFs are showed in Figure 4.10.



(a) (b)

(c)

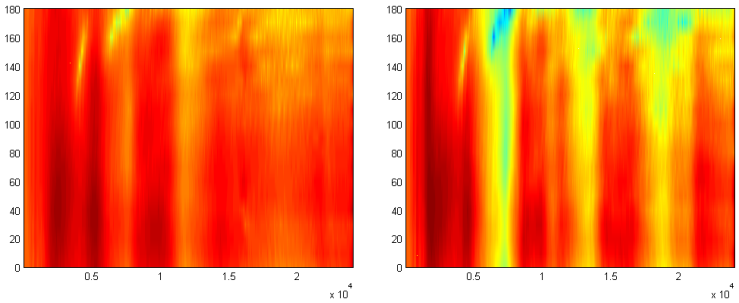
Figure 4.10: HRTFs from (a) original concha, (b) small concha, (c) smaller concha

Compared to the original right ear, the reshaped ears have smaller conchae. The notch that appears around 10 kHz shifts to high frequency range in Figure 4.10 a (12.5 kHz) and b (14 kHz). Also the notches that appear below 10 kHz shift to a higher frequency range. As we expected, the 10 kHz notch is corresponding to m_2, which is the distance from the tragus point to the concha posterior wall. The decrease of the concha leads the decrease of m_2, and causes the increase of frequency of the notch around 10 kHz (as pointed by the arrows).

However, the change of concha does not affect the frequency range below 5 kHz a lot. If we consider the notch around 5 kHz is unchanged, it could be formed by the ear helix, which is unchanged in this experimental group. Also the reshaped conchae have straighter posterior walls, which may cause the disappearance of the sloped notch between 10 - 15 kHz in the original HRTF. The ear only has a small narrow concha, and the HRTF shows periodical notches and peaks which are almost straight and remain the same in both frontal and rear azimuthal parts. Figure 4.11 shows the HRTFs corresponding to a small concha and an even smaller concha without other outer ear details. In Figure 3 12 b, the notches even show an equal distance in frequency domain. In order to have a further validation, Experiment 7 is conducted. In the construction experiment, the ear only contains the ear canal. By adding the concha, we observed the appearance of a curved notch around 10 - 20 kHz in the frontal part. By making it larger, the notch shifts to left.

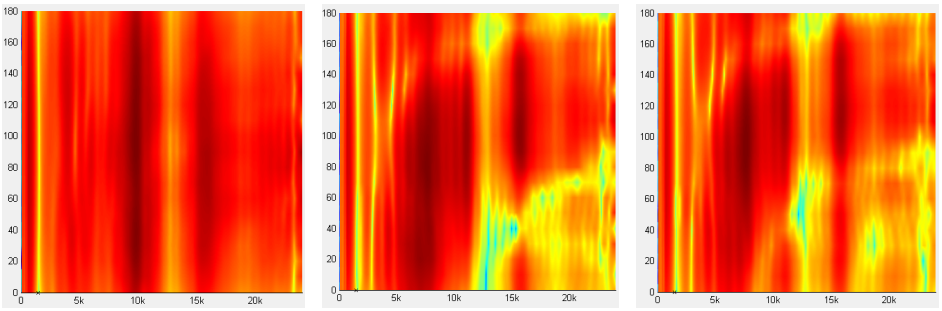
4.2.3 Ear Helix

The construction experiment shows that the helix forms a notch around 8 kHz, which is different from the notch formed by the ear concha.



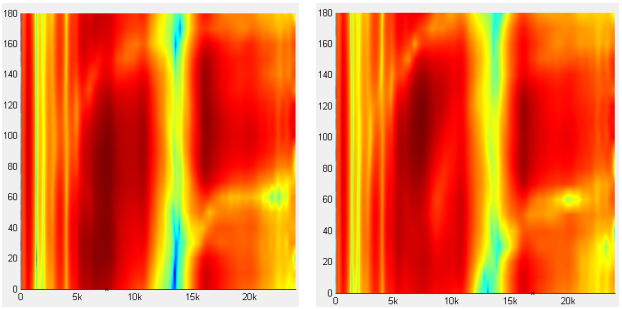
(a) (b)

Figure 4.11: HRTFs from (a) small concha, (b) smaller concha



(a) (b) (c)

Figure 4.12: HRTFs from (a) only canal, (b) adding small concha, (c) larger concha in construction experiment



(a) (b)

Figure 4.13: (a) Without helix, (b) with helix in construction experiment

4.2.4 Ear Cymba

The result shows that the cymba part plays an important role in forming the rear azimuthal parts of the notches. We use clay to block the whole cymba part. The result shows that almost the whole notch 1 and the rear part of the notch 2 disappear. In order to verify that the cymba part

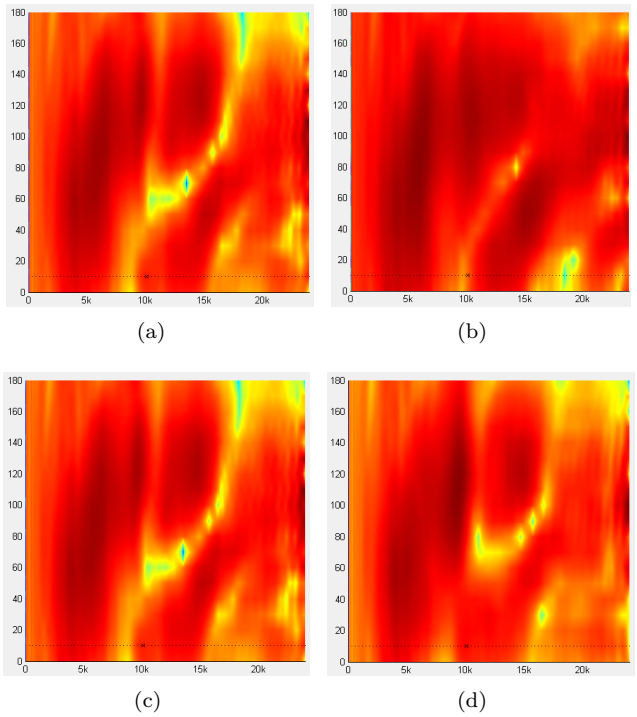


Figure 4.14: (a) Original, (b) cymba blocked, (c) original, (d) upper ear blocked

indeed causes the notch disappearance, we carry out the second experiment. We leave the cymba open but close the upper ear. The result shows that although there is a small energy change for the notch, the main structure remains.

4.2.5 Ear back

From the result, the sloped notch is affected by the ear back. Experiment 4 is applied, which uses the clay to reshape the ear back. By filling in clays, the distance from the ear back to the ear

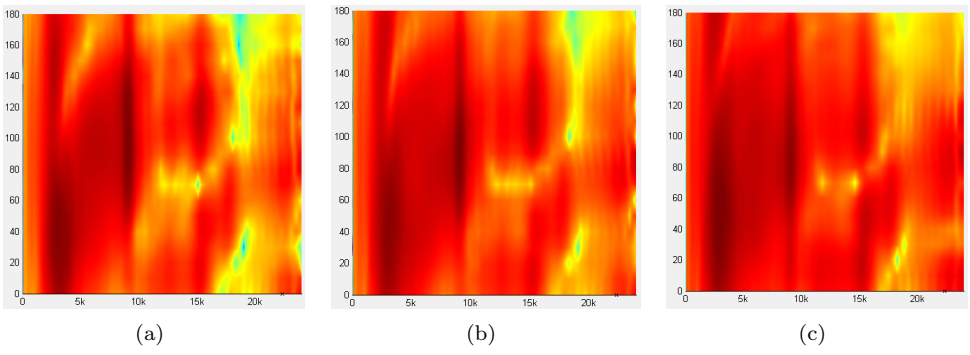


Figure 4.15: (a) Original, (b) thicken the ear back, (c) thicken more

canal becomes larger. Thus, the corresponding frequency gets smaller.

4.3 Conclusion

From the experiment and analysis, a relationship between the basic structure and its corresponding anthropometric items is described.

Table 4.6: Feature

	HRTF structure	Anthropometric items
Ipsi-lateral HRTF	Resonance around 5 kHz	Ear canal
	Sloped notch in rear part	Ear back
	Notch around 6-9 kHz in rear part	Ear cymba
	Notch around 6-9 kHz in frontal part	Ear helix
	Notch around 10-12 kHz in frontal part	Ear concha
	Notch around 12-18 kHz in frontal part	Ear helix harmonic
Contra-lateral HRTF	Curved Notch in frontal part	Rear head
	Curved Notch in rear part	Frontal head
	Notch around 6-9 kHz in rear part	Ear canal
	Notch around 6-9 kHz in frontal part	Ear helix
	Notch around 10-12 kHz in frontal part	Ear concha
	Notch around 12-18 kHz in frontal part	Ear helix

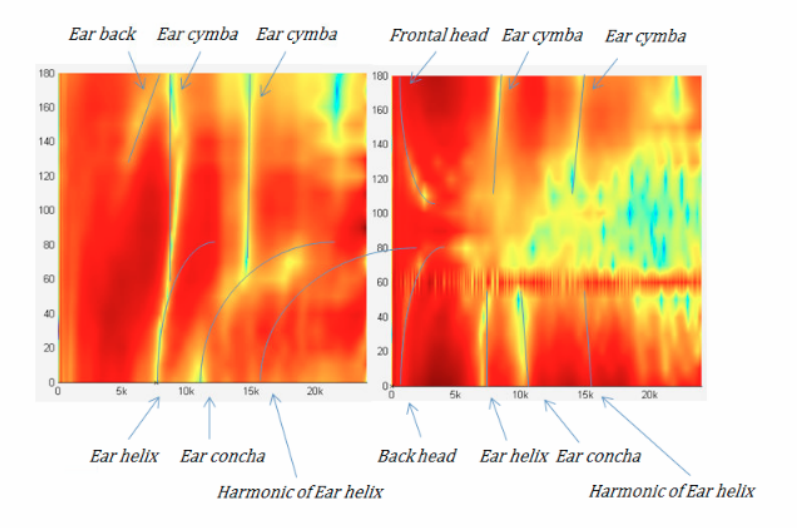


Figure 4.16: Features and corresponding anthropometric properties

Chapter 5

Model

5.1 Model Construction

The model is built to predict the notches in HRTFs. From the experimentations, we already obtain some relationship between the anthropometric data and the features. The model is constructed based on this finding. In the ipsi-lateral HRTF, we approximate Sloped Notch, Notch 1 and Notch 2. The resonance peak is excluded because the measurement of the ear canal is unavailable. The results from Chapter 4 show that Front Notch 1 and Rear Notch 1 in the contra-lateral HRTF are continuous with Notch 1 in the ipsi-lateral HRTF. This is also true for Front Notch 2 and Rear Notch 2 in the contra-lateral HRTF and Notch 2 in the ipsi-lateral HRTF. So for the contra-lateral HRTF, we only approximate Curved Notch 1 and Curved Notch 2.

The model is to predict the position of notches in HRTF, which means that for a given azimuth, the frequency of the notch in this azimuth is predicted. As we know, the notches are formed by the destructive interference of the incident wave and reflected wave. So the path differences are essential factors to influence the prediction. The path difference d is determined by the anthropometric data $\mathbf{a}$ and the angle θ of the incident wave. The anthropometric data $\mathbf{a}$ could contain both length measurement and angle measurement, such as ear width and ear protrusion angle. So for a notch N, its frequency f_N at an azimuth θ is described by

$$f_N = G(d(\mathbf{a}, \theta)), \quad \theta \in [0, \pi]$$

G is the model we build. Furthermore, according to previous research,

$$f_N = \frac{c}{2 \times d(\mathbf{a}, \theta)}, \quad \theta \in [0, \pi]$$

where c is the speed of sound. Now we use what we obtain from the statistical method and first principal method to determine d for different notches.

5.1.1 Ipsi Notch 1

According to the result in Chapter 4, Ipsi Notch 1 is divided into two parts, the frontal part and the rear part. Taking the ear protrusion angle θ_p into consideration, the frontal part is from azimuth 0 to $\pi/2 - \theta_p$ and the rear part is from azimuth $\pi/2 - \theta_p$ to π. The rear part is not only formed by the ear helix, but also influenced by the ear cymba. From the experiment we conclude that, when the cymba is blocked, the rear parts of Ipsi Notch 1 disappears. Based on this fact, we assume that when the sound wave arrives at the cymba, it gets diffracted. The diffracted wave hits the ear helix and goes back. Thus, the path difference between the incident wave and the diffracted and reflected wave cause the destructive interference. In order to reduce the anthropometric properties involved, we assume that the ear canal point, the diffraction point in the cymba and the reflection point at the helix are on the same straight line. Also the incident

wave is far enough from the source and acts as a plane wave when it arrives at the ear. At 0 azimuth, the wave and ear has an intersection angle θ_p, which is the ear protrusion angle, in the horizontal plane. Figure 5.1 shows the path of the direct wave and reflected wave when the wave comes from the rear part. Let d_c denote the distance from the ear canal to the ear cymba. The

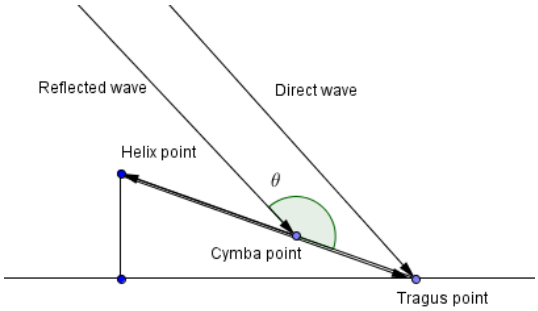


Figure 5.1: Wave path of the ear helix

path of the reflected wave equals ear width l plus the distance from the ear cymba to the ear helix $l - d_c$. The path difference d is dependent on the angle θ of the incident wave. The incident wave travels a distance $d_c \cos(\pi - \theta - \theta_p)$ more to hit the ear. The path difference is then given by

$$d = 2l - d_c - d_c \cos(\pi - \theta - \theta_p), \quad \frac{\pi}{2} - \theta_p \leq \theta \leq \pi - \theta_p$$

When the azimuth θ is beyond $\pi - \theta_p$,

$$d = 2l - 2d_c, \quad \pi - \theta_p \leq \theta \leq \pi$$

In some research only the notch at 0 azimuth is approximated and the influence of the ear protrusion angle is not considered. In this case, the frontal part is mainly affected by the ear helix. However, from our result, the frontal part of Ipsi Notch 1 is also affected by the ear cymba similarly as the rear part, see Figure 5.2. So there are two possibilities. 1, the path difference d with the function of ear cymba is modeled as:

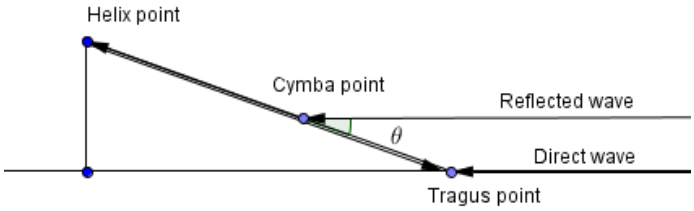


Figure 5.2: Wave path of the ear helix

$$d = 2l - d_c + d_c \cos(\theta + \theta_p), \quad 0 \leq \theta \leq \frac{\pi}{2} - \theta_p$$

2. the path difference without the function of ear cymba is modeled as:

$$d = l + l \cos(\theta + \theta_p), \quad 0 \leq \theta \leq \frac{\pi}{2} - \theta_p$$

Based on all Ipsi Notch 1 data that we find in the database, it should be continuous through all azimuth from 0 to 180. By checking this continuity condition, the model without the function of ear cymba is ruled out.

5.1.2 Ipsi Notch 2

The form of Ipsi Notch 2 is similar to Ipsi Notch 1. The only difference is that the reflection point is at the ear concha wall. So the distance l is from ear canal to ear concha.

5.1.3 Ipsi Sloped Notch

Ipsi Sloped Notch is affected by the ear back. In Chapter 4, we claim that when the upper part of the ear back is filled with clay, the Ipsi Sloped Notch shifts to left in the HRTF graph. This means that its frequency becomes smaller. The wave travel path is along the upper part of the ear back.

We can use the ear measurements like concha length and helix length to simulate this case. The statistical results show that Ipsi Sloped Notch starts around an azimuth of $(\pi - \theta_p)/\pi \times 180$, and ends at 180. Thus we assume that the incident wave travels along the outer ear in the horizontal plane and the affected wave travels along the upper ear and both end at the ear canal. Figure 5.3 shows the wave paths when they come from rear part of the horizontal plane and being affected by the ear back. We model the path along the upper ear as a half circle. The formula of the path

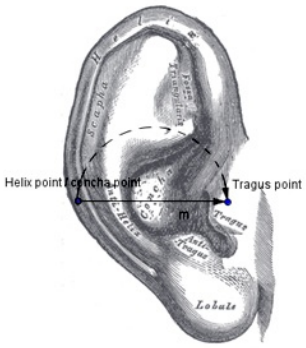


Figure 5.3: Wave path of the ear back

difference d is:

$$d = \frac{\pi m(\theta)}{2} - m(\theta), \quad \pi - \theta_p \leq \theta \leq \pi$$

m is dependent on azimuth because as the azimuth increases, the start point shifts from helix point to concha point linearly.

$$m(\theta) = d_c + \frac{l - d_c}{\theta_p}(\pi - \theta), \quad \pi - \theta_p \leq \theta \leq \pi$$

5.1.4 Contra Curved Notch

The curved notches in the low frequency range of the contra-lateral HRTF are the most unique features in HRTF. Because of their low frequency, it is assumed that they are caused by the effect of the head. We built a head model to explain how the destructive interference happens. In order to take account of the difference between the frontal head and the rear head, we use two half ellipses to model the shape of the head instead of using a sphere. Let $2a$, $2b$ and $2c$ denote the head width, frontal head distance and rear head distance respectively. Let the intersection point

of the head width and the head depth be the origin of our carterian coordinate system. So the function of the front circumference of the head is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad y \geq 0.$$

The function of the rear circumference of the head is

$$\frac{x^2}{a^2} + \frac{y^2}{c^2} = 1, \quad y \leq 0.$$

When a wave comes from the front part, one sound wave travels along the frontal head and arrives at the contra-ear canal, another sound wave travels along the rear head and also arrives at the contra-ear canal. The difference between these two paths causes a notch in the HRTF graph. We

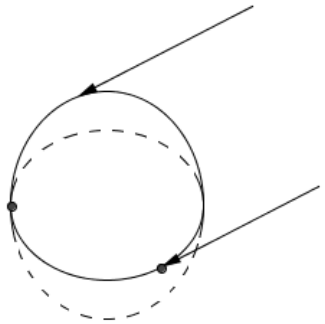


Figure 5.4: Wave path affected by the head

assume that the waves arrive at the frontal head and the rear head simultaneously and that they are tangent to the head curve, see Figure 5.4. Let C_1 denote the upper and C_2 denote the lower ellipse curves.

$$C_1 = b \int_{\frac{\pi}{2}-\theta}^{\frac{\pi}{2}} \sqrt{(1 - \frac{b^2 - a^2}{b^2} \sin^2 \phi)} d\phi, \quad 0 \leq \theta \leq \pi$$

$$C_2 = c \int_{\theta}^{\pi} \sqrt{(1 - \frac{c^2 - a^2}{c^2} \sin^2 \phi)} d\phi, \quad 0 \leq \theta \leq \pi$$

Thus the path difference is

$$d = |C_1 - C_2|.$$

5.1.5 Other Notches in the Contra-lateral HRTF

As the results from Chapter 4 show, Front Notch 1 and Rear Notch 1 in the contra-lateral HRTF are continuous with Notch 1 in the ipsi-lateral HRTF. So the frequency of Ipsi Notch 1 at 0 and 180 azimuth are the frequency of Contra Front Notch 1 at 0 azimuth and the frequency of Contra Rear Notch 1 at 180 azimuth. From this observation, these notches usually from azimuth 0 to 60 or 120 to 180 are vertically straight in the HRTF graph, which means that each notch has the same frequency over different azimuths. So for Contra Front Notch 1

$$d = l + l \cos \theta_p, \quad 0 \leq \theta \leq \frac{\pi}{3}$$

For Contra Rear Notch 1

$$d = 2l - d_c, \quad \frac{2\pi}{3} \leq \theta \leq \pi$$

The same analysis is also applied to Front Notch 2 and Rear Notch 2 in the contra-lateral HRTF and Notch 2 in the ipsi-lateral HRTF.

5.2 Result

With the previous analysis, $d(\mathbf{a}, \theta)$ is modeled for different cases. The frequency of each notch is calculated, on basis of that model.

5.2.1 Ipsi Notch 1

We use the ear width and ear cymba width to reconstruct Ipsi Notch 1. The average ear width and ear cymba width is 2.23 cm and 0.5 cm. Figure 5.7 shows an approximation for Ipsi Notch 1 by using its anthropometric data and actual HRTF graph.

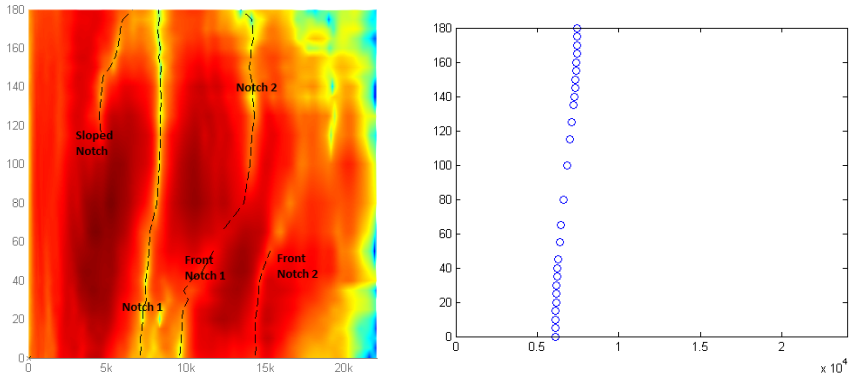


Figure 5.5: Ipsi Notch 1 approximation

5.2.2 Ipsi Notch 2

The average ear concha wdith is 1.58. Similar as Ipsi Notch 1, we approximate Ipsi Notch 2 with the ear concha width and the ear cymba width. Figure 5.5 shows an approximation for Ipsi Notch 2 by using its anthropometric data.

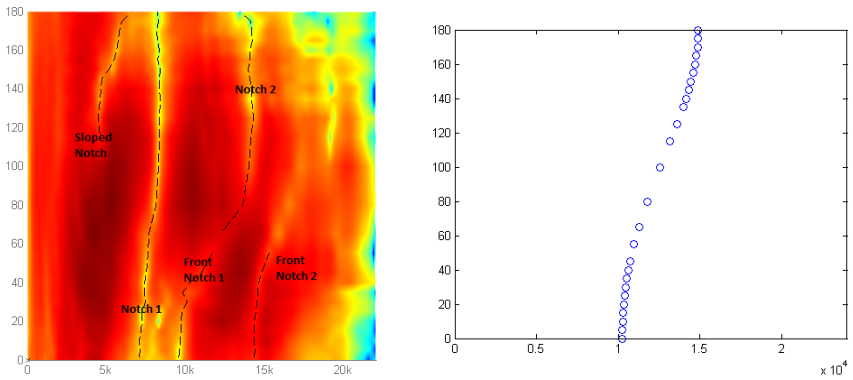


Figure 5.6: Ipsi Notch 2 approximation

5.2.3 Ipsi Sloped Notch

We use the ear width and ear concha width to reconstruct Ipsi Sloped Notch. Figure 5.6 shows an approximation for Ipsi Sloped Notch by using its anthropometric data.

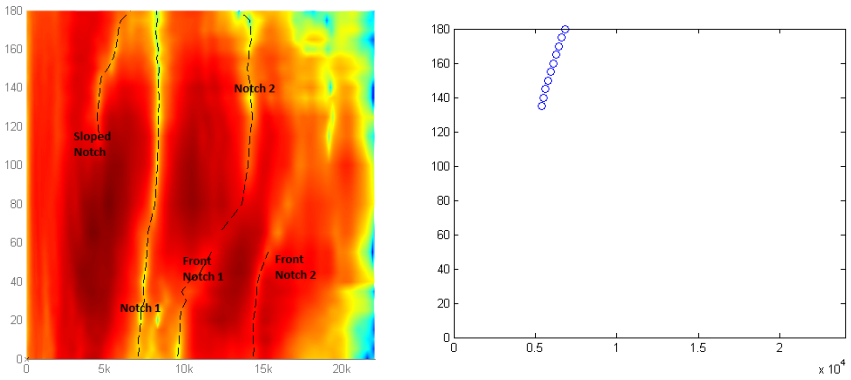


Figure 5.7: Ipsi Sloped Notch approximation

5.2.4 Contra Curved Notch

The average head width is around 14 cm from the CIPIC database, thus the curve generated from the formula is showed in Figure 5.8. Since the circumference of the rear head is smaller than the circumference of the front head, the frequency to 180 azimuthal position, which is 500 Hz, is smaller than 0 azimuthal position.

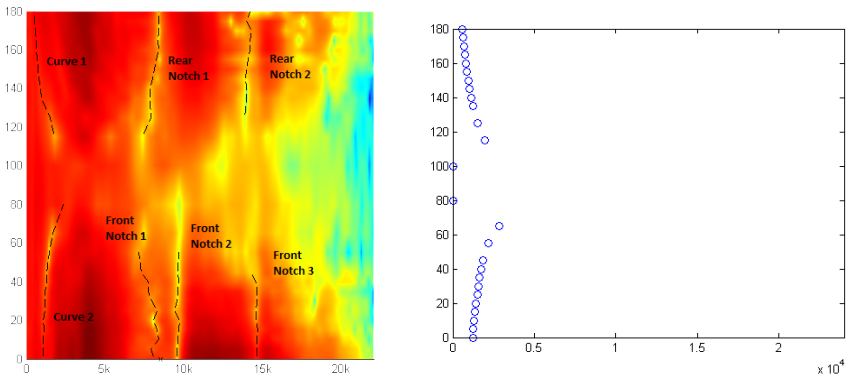


Figure 5.8: Contra Curved Notch approximation

5.2.5 Conclusion

All notches in the HRTF graph are modeled. The 7 anthropometric properties that we use are the ear width, the concha width, the cymba width, the ear protrusion, the head width, the distance of the frontal head and the distance of the head distance. Figure 5.9 shows all approximated notches in the HRTF graph. In this figure, Contra Front Notch 1 and 2, Contra Rear Notch 1 and 2 are also approximated. From the result of the statistical method. The harmonics of Ipsi Notch 1 and Contra Front Notch 1, which is Ipsi Front Notch 2 and Contra Front Notch 3, are also calculated and presented. The precision of the model is evaluated in the next chapter.

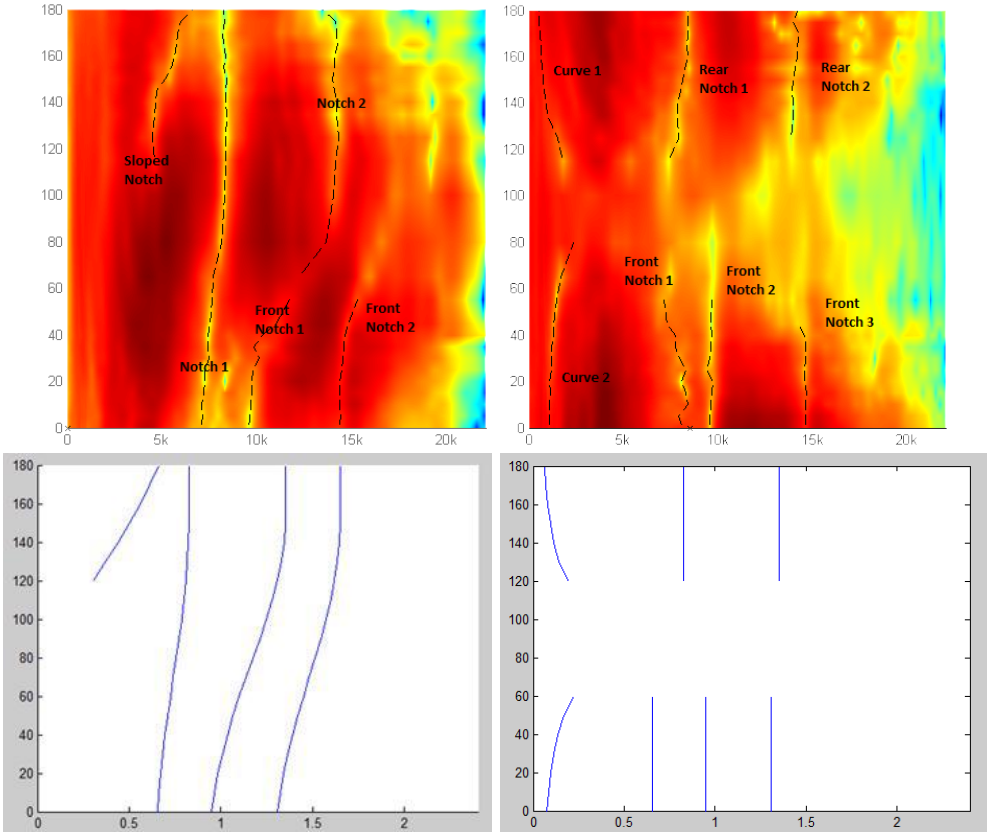


Figure 5.9: Reconstruction of all annotated notches

Chapter 6

Result

In order to validate the model we build, we use the data in the CIPIC database to do benchmarking. As we conclude in the last chapter, the anthropometric properties required in the model are the ear width, the concha width, the cymba width, the ear protrusion, the head width, the distance of the frontal head and the distance of the rear head. Except the ear cymba width, the data of these anthropometric properties are contained in the CIPIC database. However, some of the measurements, such as the ear width and the concha width, are not related to the ear canal. They are not the correct measurements in our model. Fortunately, the CIPIC database also contains the pinna photos of 31 subjects with a length reference. Then the cymba width, the concha width and the ear width can be measured from the photos. So these 31 subjects in the CIPIC database are the samples of benchmarking.

On one hand, from these 7 anthropometric properties, the notches can be approximated. On the other hand, we can annotate the actual notches in HRTF graph. By comparing the approximated notches and the actual notches, the effectiveness of the model is evaluated. Furthermore, we apply Equivalent Rectangle Bandwidth to evaluate the result in the perspective of the human hearing.

6.1 Benchmarking

As introduced, HRTF in the CIPIC database has 26 difference azimuths in the ipsi-lateral of the horizontal plane. For each measured notch N_m, at each azimuth θ it has a frequency $f_{N_m}(\theta)$. The approximated notch N_a is obtained from the model. To compare these two notches, we consider the measurement precision. We assume the d with error bound is $d \pm cs$, where s stands for the standard deviation and c for constant. So we construct a tolerance band for the approximated notch, see Figure 6.1.

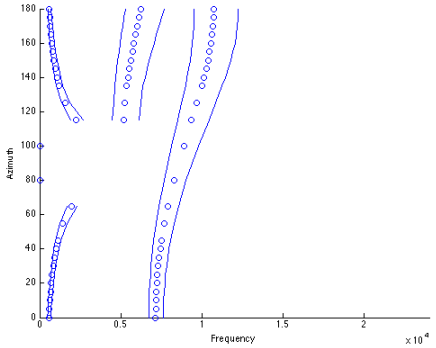
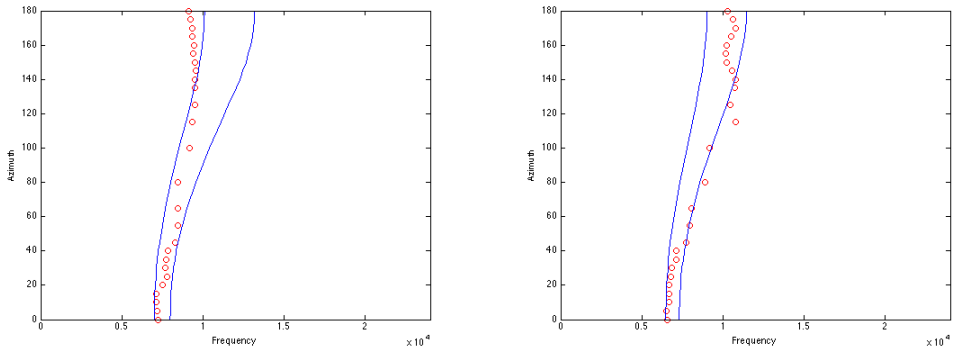


Figure 6.1: Approximated notches and their tolerance band

If the approximated notch at a certain azimuth is located in the tolerance band, we conclude that the corresponding azimuth of this notch is well approximated. For each notch of each subject, we determine the number of points inside the tolerance band, and calculate the percentage of all points of one notch. Then we take the assumption that a good approximation is over 60%. So we count the number of subjects who have a precision approximation over 60% and compute the percentage of them in the whole 31 group. Also we determine the azimuths of each notch inside the band to evaluate the precision of this model at each azimuth.

6.1.1 Ipsi Notch 1

The standard variance s of the ear width is 0.3 cm and we construct a tolerance band with $c = 1.2$. Figure 6.2 shows that the measured (red dots) Ipsi Notch 1 of subject (a) 21, (b) 26 in the CIPIC database and its tolerance band (within blue line). For the red points inside the blue band, they are well approximated.



(a) (b)

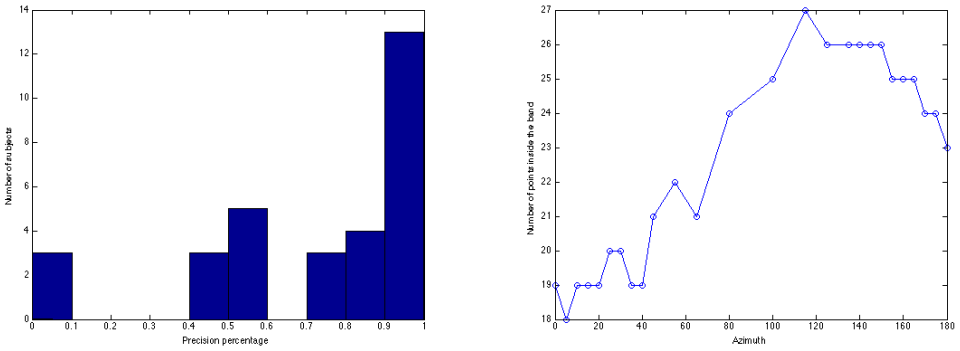
Figure 6.2: The measured (red dots) Ipsi Notch 1 of subject (a) 21, (b) 26 in the CIPIC database and its tolerate band (within blue line)

In order to evaluate how well each Ipsi Notch 1 is approximated, we count the number of the points inside the tolerance band and calculate its percentage for each subject. The histogram is made to show the result, see Figure 6.3. The x-axis denotes the percentage of the inside-band points of one Ipsi Notch 1. The y-axis denotes the number of subjects who have the corresponding percentage. In total we have 31 subjects. There are 13 subjects who have over 90% precision, which is 42% of all the subjects. If we consider over 60% presicion is a good approximation, there are 20 subjects who satisfy this condition, which is account for 65%.

Over all the azimuths, we also investigate the azimuths which are well approximated. For each azimuth, we count the times it is inside the tolerance band over 31 subjects. In Figure 6.3 (b), the x-axis denotes the index of azimuth from 0 to 180 and the y-axis denotes the times. It shows that the rear azimuths are approximated better than the frontal azimuth.

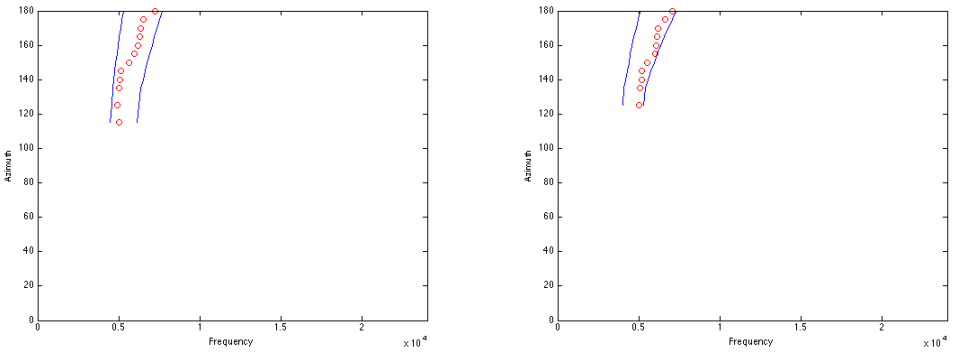
6.1.2 Ipsi Sloped Notch

We apply the same evaluation strategy on Ipsi Sloped Notch. The standard variation s of the concha with is also 0.3 cm. We take $c = 1.2$. Figure 6.4 shows the measured (red dots) Ipsi Slope Notch of subject (a) 1, (b) 30 in the CIPIC database and its tolerance band (within blue line). These two Ipsi Sloped Notches are well approximated. We investigate the whole group of 31 subjects and Figure 6.5 shows that the well approximated is account for 55% if we consider an approximation over 60% precision percentage is good, see Figure 6.5.



(a) (b)

Figure 6.3: (a) Histogram of precision percentage of Ipsi Notch 1; (b) Number of inside band points for each azimuth



(a) (b)

Figure 6.4: The measured (red dots) Ipsi Slope Notch of subject (a) 1, (b) 30 in the CIPIC database and its tolerance band (within blue line)

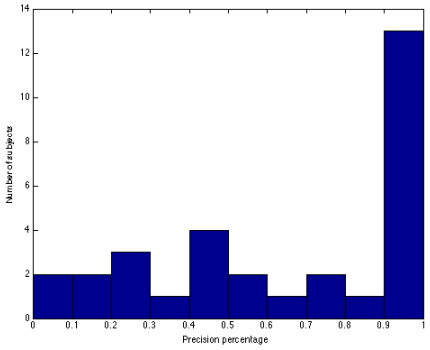
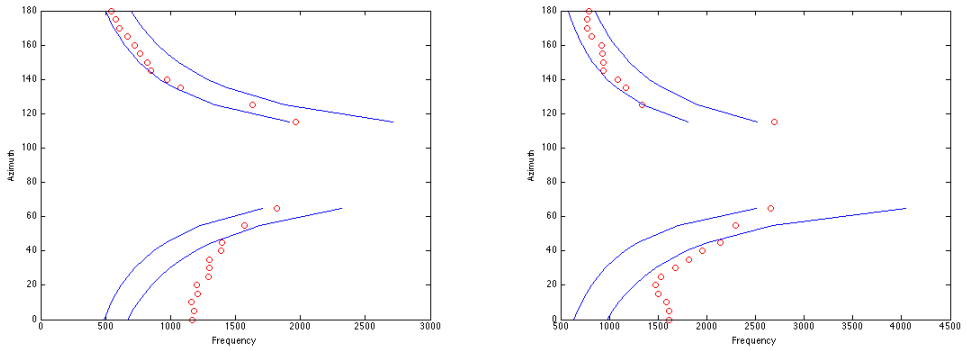


Figure 6.5: Histogram of precision percentage of Ipsi Sloped Notch

6.1.3 Contra Curved Notch

Contra Curved Notch is modeled with the head related properties. The standard variation s of the distance of the top head is 0.9 cm, and the standard variation s of the distance of the rear head is 1.1 cm. We take $c = 1.5$.

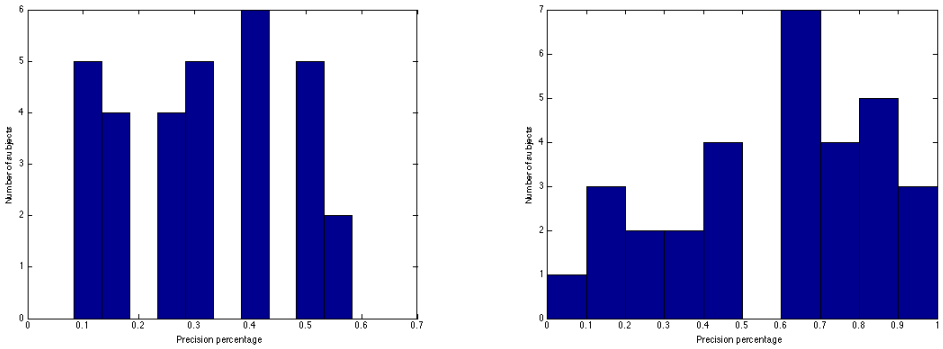
Figure 6.6 shows the measured (red dots) Contra Curved Notch of subject (a) 1, (b) 3 in the CIPIC database and its tolerance band (within blue line). It shows that Curved Notch in the rear



(a) (b)

Figure 6.6: The measured (red dots) Contra Curved Notch of subject (a) 1, (b) 3 in the CIPIC database and its tolerance band (within blue line)

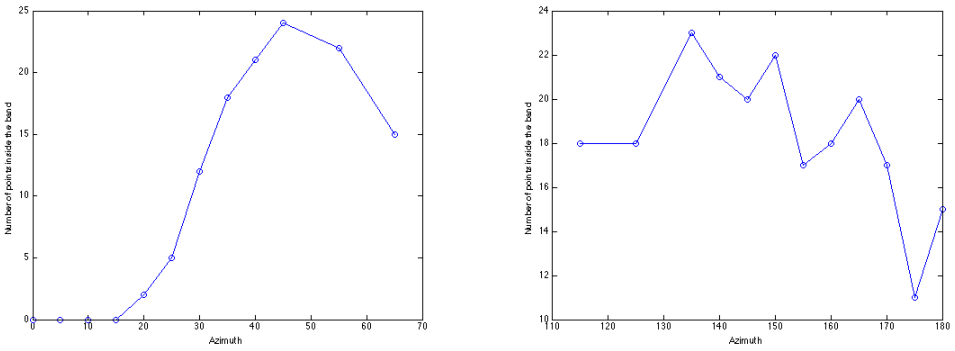
part of the graph (Curved Notch 2) is better approximated than Curved Notch in the frontal part of the graph (Curved Notch 1). We confirm this statement by calculating the precision percentage for Curved Notch 1 and Curved Notch 2 of each subject, see Figure 6.7. There is 19 subjects who get approximation over 60% for Curved Notch 2. But for Curved Notch 1, all approximation is under 60%, which means that the model for Curved Notch 1 is not good. For each Curved Notch,



(a) (b)

Figure 6.7: Histogram of precision percentage of Curved Notch (a) 1 and (b) 2

the azimuths from 35 to 165 degree are better approximated than the lateral azimuths such as 0 and 180, see Figure 6.8.



(a) (b)

Figure 6.8: Number of inside band points for each azimuth of Curved Notch (a) 1 and (b) 2

6.2 Evaluation by Equivalent Rectangular Bandwidth

For each approximated notch, we also apply Equivalent Rectangular Bandwidth (ERB) (see Chapter 2) to evaluate its precision. We calculate the central frequency of the approximated notch and actual notch at every azimuth. A band related to this central frequency is computed according to ERB. For each azimuth, if the two notches are both located in the band, there is no difference in listening system and we consider the approximation at this azimuth is good. Let f_{N_a} denote the frequency of the approximated notch and let f_{N_m} denote the frequency of the measured notch at azimuth θ, then its central frequency is

$$f_a(\theta) = \frac{1}{2}(f_{N_a}(\theta) + f_{N_m}(\theta))$$

The corresponding band b equals to

$$b = \frac{Q}{\gamma}\log(\frac{f_a}{QL} + 1);$$

where $\gamma = 1$, $Q = 9.265$ and $L = 24.7$. So the lower frequency bound is

$$f_l(b) = QL(\exp^{(b-0.5)\gamma/Q} - 1)$$

the upper frequency bound is

$$f_u(b) = QL(\exp^{(b+0.5)\gamma/Q} - 1)$$

The band width is

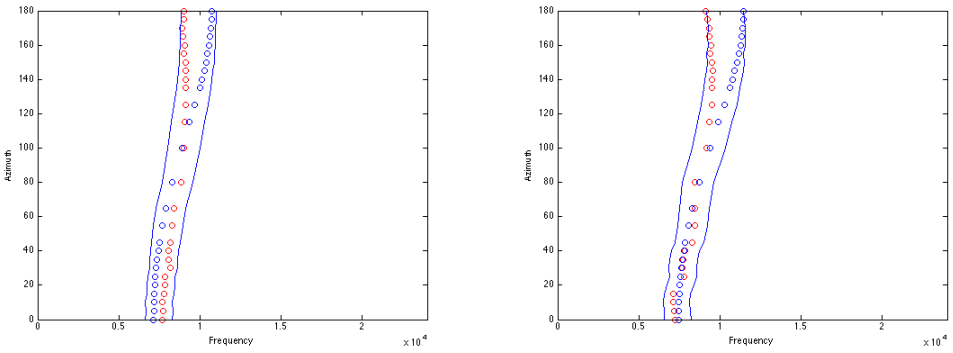
$$w(b) = f_u(b) - f_l(b).$$

6.2.1 Ipsi Notch 1

For Ipsi Notch 1, the ERB evaluation shows that the model is approximated well. Notch 1 has an average frequncy of 8 kHz and the corresponding bandwidth is 1780 Hz, which is from 7.15 kHz to 8.93 kHz. So the ear width can vary from 1.93 cm to 2.40 cm. This difference is larger than the standard variation 0.3 cm.

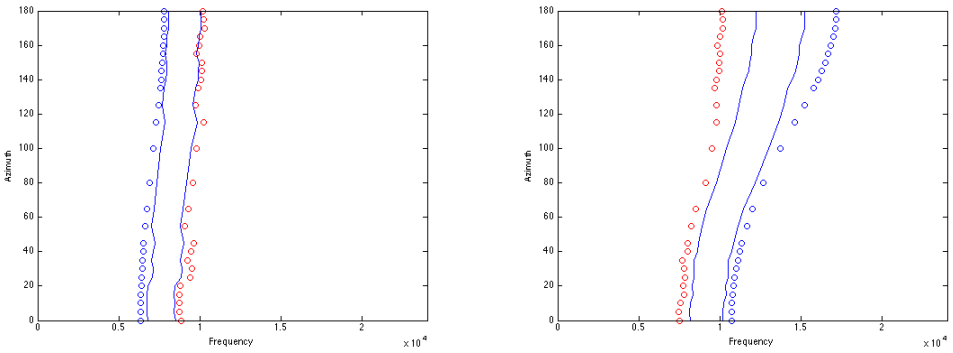
There are some subjects whose approximated notch and measured notch are not inside their ERB band, see Figure 6.10.

The histogram 6.11 shows that over 60% precision there are 25 subjects, which is account for 80% of the whole group. Over 70% precision there are 21 subjects, which is account for 68% of the whole group.



(a) (b)

Figure 6.9: The measured (red dots) Ipsi Notch 1 of subject (a) 1, (b) 21 in the CIPIC database and its ERB band (within blue line)



(a) (b)

Figure 6.10: The measured (red dots) Ipsi Notch 1 of subject (a) 18, (b) 24 in the CIPIC database and its ERB band (within blue line)

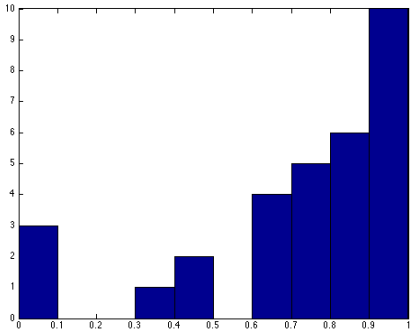
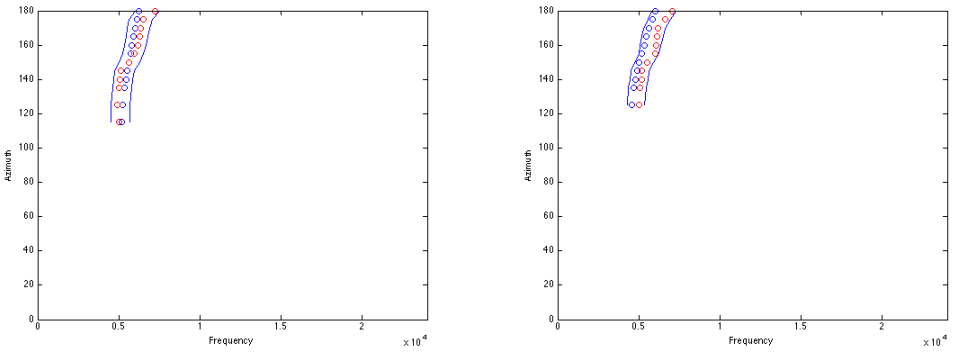


Figure 6.11: Histogram of precision percentage of Ipsi Notch 1

6.2.2 Ipsi Sloped Notch

Ipsi Sloped Notch has an average frequency of 6 kHz at azimuth of 180 degree. The corresponding bandwidth is 1.35 kHz, which is from 5.36 kHz to 6.71 kHz. The corresponding measurement length is from 2.6 cm to 3.2 cm. The difference is larger than the standard variation 0.3 cm.

The histogram 6.13 shows that over 60% precision there are 20 subjects, which is account for 65% of the whole group. Over 70% precision there are 18 subjects, which is account for 58% of the whole group.



(a) (b)

Figure 6.12: The measured (red dots) Ipsi Sloped Notch of subject (a) 1, (b) 30 in the CIPIC database and its ERB band (within blue line)

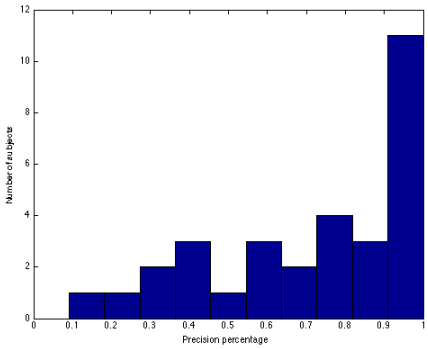
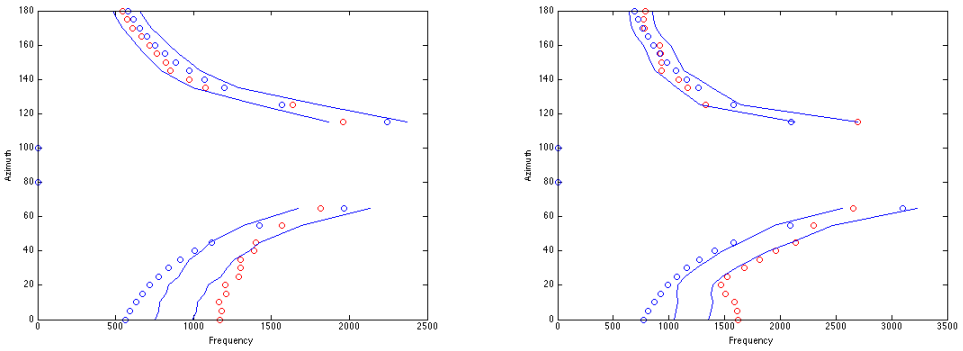


Figure 6.13: Histogram of precision percentage of Ipsi Sloped Notch

6.2.3 Contra Curved Notch

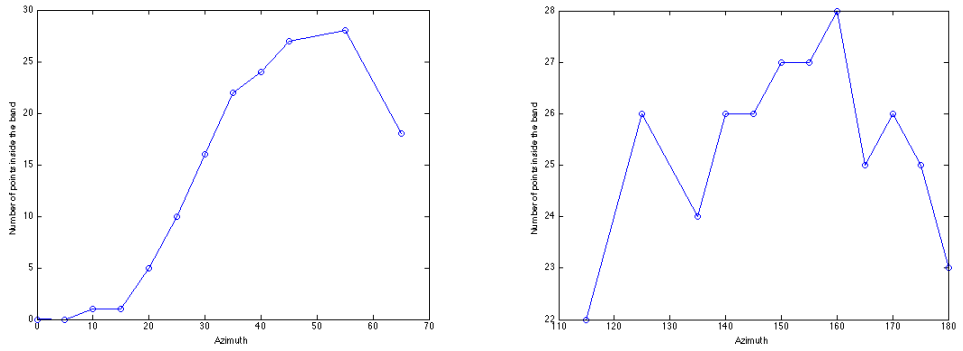
Curved Notch 2 has a frequency of 500 Hz at azimuth of 180 degree. The corresponding bandwidth is 158 Hz, which is from 425 Hz to 583 Hz. Although the bandwidth is relatively small compared to the bandwidth of Ipsi Notch 1, Curved Notch 2 is well approximated. Figure 6.15 (a) and (b) show the percision percentages for each azimuth of Curved Notch 1 and 2. The latter figure shows that the least approximated azimuth of Curved Notch 2 is azimuth of 115 degree, but it still has percision percentage of $22/31 = 71\%$. The model for Curved Notch 1 is relatively weak, see Figure 6.15 (a). The precision percentage is under 50% for azimuth from 0 to 30.

The histogram 6.16 (b) shows that for Curved Notch 2 over 60% precision there are 27 subjects, which is account for 87% of the whole group. Over 70% precision there are 25 subjects, which



(a) (b)

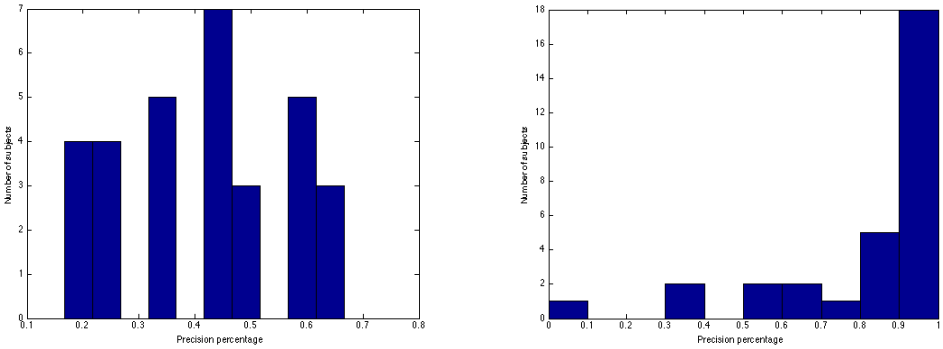
Figure 6.14: The measured (red dots) Ipsi Sloped Notch of subject (a) 1, (b) 3 in the CIPIC database and its ERB band (within blue line)



(a) (b)

Figure 6.15: Number of inside band points for each azimuth of Curved Notch (a) 1 and (b) 2

is account for 80% of the whole group. But as shown in 6.16 (a), there are only 10% of all the subjects who have over 70% percentage for Curved Notch 1.



(a) (b)

Figure 6.16: Histogram of precision percentage of Curved Notch (a) 1 and (b) 2

6.3 Analysis

The approximated notches are azimuth-dependent, which is the same with the measured notch. This is because although we take the assumption that the reflected point of the wave is unchanged, the difference between the incident wave and reflected wave changes when the angle of the incident wave changes. In our model, the shape of each notch is well approximated. For instance, most of measured Ipsi Notch 1 is tilted to the higher frequency part in HRTF graph. Sloped Notch is not a full azimuth notch and also tilted to the higher frequency. Curved Notches are relatively symmetric and both bended from lateral to middle. The qualitative relationship between azimuth and frequency is well modeled.

Table 6.1 shows the precision percentage of the different notches under different evaluation method.

Table 6.1: Precision percentage of the different notches under different evaluation method

Feature	Measurement precision	ERB
Ipsi Notch 1	65%	80%
Ipsi Sloped Notch	55%	65%
Curved Notch 1	0%	10%
Curved Notch 2	62%	87%

Curved Notch 1 is not well approximated. The measured notches have a larger frequency at 0 azimuth and a larger curvature, comparing to the approximated notches. The model of Curved Notch 1 needs to improve. No matter the shape and the location on the HRTF graph, Curved Notch 2 and Ipsi Notch 1 are well approximated. The approximation of Ipsi Sloped Notch is also not bad.

The results are better with the ERB evaluation than with the measurement precision.

The error may come from different aspects. The annotation of the notches in the HRTF graph is not absolutely objective. Distractive notches appear in some graphs and they can be mis-annotated. Also the measurements from the photo can cause error. The exact reflection points on the ear helix and the ear concha is hard to determine.

The model only considers the effect on the horizontal plane. So some assumptions are taken to simplify the situation. In fact the outer ear affect the wave in three dimension and this influence

is not considered in our model.

Chapter 7

Conclusions

7.1 Relationship

There are clear relations between the HRTF features and the anthropometric data. Because the range of the database is limited, the relationship between the anthropometric data and corresponding HRTF data obtained from statistical methods are not solid. In spite of that, conducting the acoustic experiments, we have confirmed them. Table 7.1 shows the relations suggested by statistical methods and confirmed by acoustic experiments.

Table 7.1: Correlated anthropometric data and features

Lateral	Feature	Anthropometry
Ipsi	Sloped Notch	Ear Protrusion
Contra	Front Notch 1	Ear Width
Contra	Curved Notch 2	Dist. Rear

Besides the relationships obtained from statistical methods, other relationships, which are not clear in the statistical results, are revealed through the experiments, see Table 7.2.

Table 7.2: Relationship between features and anthropometry

Lateral	Feature	Anthropometry
Ipsi	Resonance around 5 kHz	Ear canal
Ipsi	Sloped notch in rear part (Sloped Notch)	Ear back
Ipsi	Notch around 6-9 kHz in rear part (Notch 1)	Ear cymba
Ipsi	Notch around 6-9 kHz in frontal part (Notch 1)	Ear helix
Ipsi	Notch around 10-12 kHz in frontal part (Notch 2)	Ear concha
Ipsi	Notch around 12-18 kHz in frontal part	Ear helix harmonic
Contra	Curved Notch in frontal part (Curved Notch 1)	Rear head
Contra	Curved Notch in rear part (Curved Notch 2)	Frontal head
Contra	Notch around 6-9 kHz in rear part (Rear Notch 1)	Ear canal
Contra	Notch around 6-9 kHz in frontal part (Front Notch 1)	Ear helix
Contra	Notch around 10-12 kHz in frontal part (Front Notch 2)	Ear concha
Contra	Notch around 12-18 kHz in frontal part (Front Notch 3)	Ear helix

7.2 Model

The relationships we established can be used to predict individual HRTF features from the anthropometric data by building relatively simple models. We build three different kinds of model. They

simulate the change of the path difference between the incident wave and the reflected wave from the ear helix, the ear back, and the head circumference. In our model, only seven anthropometric properties are used, see Table 7.3.

Table 7.3: Anthropometries used in our model

Ear	Pinna width	Concha width	Cymba width	Ear protrusion
Head	Width	Frontal distance	Rear distance	

We conclude in Chapter 3 that there is no obvious relationship between the ear related properties and the head related properties, so it is reasonable to use both of them in the model. Besides, when we apply PCA in the ear related properties, variations are spread differently on the length measurements and the angle measurements. So the use of the ear width and protrusion is maintaining the variety.

The models are verified by the measurement precision and the ERB evaluation. Ipsi Notch 1, Ipsi Sloped Notch, and Contra Curved Notch can be approximated well. The shape of the notches are well modeled, which shows that the qualitative relationship between the anthropometric data and the HRTF feature are correct.

From the results of benchmarking, we conclude that the rear parts of the features are modeled better than their frontal parts. This is because the tolerance bands are wider in the rear parts then in the frontal parts. It is also correct with the acoustic precision since human can determine the coming direction of a sound in the front better than in the rear.

The model do not show all the features. For instance one of the peaks in the HRTF graph are caused by the ear canal. Since we do not have the relative data, we can not model the peak.

The error of the model may come from different aspects. The annotation of the features bring the error at the first place. The features are annotated manually. For instance, the annotation of Notch 1 can be different in some azimuths, see Figure 7.1 There is a small notch near Notch 1 and it could be mis-annotated in some HRTF. In the second place, the measurements of the ear and the head also bring the error, especially when we measure the ear width and the concha width from the photos. In the third place, the models are relatively simple and we take some assumptions when we build the models.

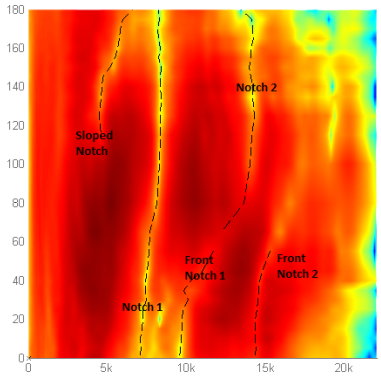


Figure 7.1: Notch 1 may be mis-annotated

7.3 Recommendation

7.3.1 Development

If more HRTF data are available, more solid relationships may be obtained from statistical methods. Then the model can be built on pure statistical methods and the laborious experiments can

be skipped. In this project, we already show that the anthropometric and HRTF data in different database are similar in the sense of statistics. It is a promising way on collecting data from different database and perform different statistical methods.

Notches are not the only kinds of the HRTF feature. As annotated in the HRTF graph, the peaks and the energy level of a Spatio-area are also features. The peaks is related to the length of the ear canal, which could be measured in the future and contained into our model.

The model only focus on the horizontal plane and has it limitation. The similar research can conducted on the vertical plane and its HRTF. Further information can be obtained from this complementary research and a more accurate model could be built.

To eliminate the error, we can develop the whole procedures in three different ways. We can re-annotate the features according to the results of our models or leave the controversial features aside and re-calculate the percision percentage. Secondly, we collect our database with correct anthropometries instead of using the measurements from the photos. Thirdly, we can develop more complicated models if we consider more anthropometies and physical phonomena such as resonance.

7.3.2 Further Steps towards Personalization

After the model is built and the features are predicted by the anthropometric data, personalization needs to be applied. The features are the marks and we use them to shift a reference HRTF graph to a target HRTF graph, see Chapter 2.

The transformation is only applied on the frequency domain, the amplitude of the HRTF is untouched. The energy level related to the amplitude is also needed personalization. The relationship is under discovery.

About evaluation, listening test could be performed. Informal listening tests can be done by several subjects indicated that the apparent location of sounds synthesized by personalized HRTF. Listening tests are performed at a set of random angles from 0 to 360 with a 10 degree difference which covers the horizontal plane. In-ear phones were used to avoid the need for headphone compensation. The listening test can tell us the validations of the personalization at different azimuth. It is also an indirect evalution method of our model.

Bibliography

[1] V Ralph Algazi, Richard O Duda, and Dennis M Thompson. The use of head-and-torso models for improved spatial sound synthesis. *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 2002. 17

[2] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE, 2001. 5, 6

[3] Dwight W Batteau. The role of the pinna in human localization. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 168(1011):158–180, 1967. 9, 16

[4] Doris J Kistler and Frederic L Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91:1637, 1992. 12

[5] William L Martens. *Principal components analysis and resynthesis of spectral cues to perceived direction*. Ann Arbor, MI: MPublishing, University of Michigan Library, 1987. 9, 12

[6] John C Middlebrooks and David M Green. Observations on a principal components analysis of head-related transfer functions. *The Journal of the Acoustical Society of America*, 92:597, 1992. 13

[7] Vikas C Raykar, Ramani Duraiswami, and B Yegnanarayana. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *The Journal of the Acoustical Society of America*, 118:364, 2005. 10

[8] EAG Shaw. The acoustics of the external ear. *Acoustical factors affecting hearing aid performance*, pages 109–125, 1980. 17

[9] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. Fitting pinna-related transfer functions to anthropometry for binaural sound rendering. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 194–199. IEEE, 2010. 16, 34

[10] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. Structural modeling of pinna-related transfer functions. In *In Proc. Int. Conf. on Sound and Music Computing (SMC 2010)*, 2010. 34

[11] S Verbruggen. Anthropometry for earphones and headphones measurement. Technical report, 2011. 5, 8, 17, 28

[12] Bosun Xie. *Head-related transfer function and virtual auditory display*. 2013. 17