Spatial Domain Modeling of Head-Related Transfer Functions

Spatial Domain Modeling of Head-Related Transfer Functions

by

Shichao Hu

A thesis submitted to the Graduate School of Information Sciences of Tohoku University in partial fulfillment of the requirements for the degree of Doctor of Science

Evaluation Committee

Prof. Yôiti Suzuki

Prof. Akinori Ito

Prof. Takuo Suganuma

Assoc. Prof. Shuichi Sakamoto

Tohoku University

March 2019

Dedicated to my parents, Jianguo Hu and Guihua Liu,

and my wife, Shenjing Liu.

Preface

Human beings can localize sounds extremely well in daily life. This ability is called spatial hearing perception. In contrast to visual perception, which covers only the front region of a person's hearing, spatial hearing provides three-dimensional information of the surrounding environment, even in darkness. This information is useful in our daily life, together with vision, when seeking objects and warning ourselves to evade potential danger. With sound localization capability, we are well able to tell the direction of traffic. Thereby, we can immediately ascertain the positions of speakers and auditory objects that might be obscured from sight. Furthermore, the spatial separation of sounds improves intelligibility in a noisy environment by a phenomenon known as the "cocktail party effect." Therefore, spatial hearing plays a vital role in our daily communications.

Apart from daily communications, sound localization has also been considered increasingly for use in modern audio devices such as hearing aids, game audio, virtual and augmented reality applications, and teleconferencing. Because spatial hearing is an important aspect of our perception, it is necessary to consider it when designing those information systems. Nevertheless, the present telecommunication systems that we are using every day can transmit only limited media information to us without sufficient spatial auditory awareness. The original spatial sound impression of auditory events can be reduced severely by conventional technologies. The exciting atmosphere and enjoyment of a concert and sport event might become tedious and flat when presented to us, even when using the latest information communication systems. Therefore, it is urgent to improve spatial hearing perception for the audio products of today. If technology that can achieve “sense of presence"

can be realized, then we might even stay at home and “sit" at the best place among the audiences to enjoy every great moment brought to use by musicians and athletes. For the reasons presented above, spatial hearing is important for both our daily communications and information systems in the consumer industry.

Over the past century, many scientists and researchers have made numerous attempts to ascertain the mechanisms and principles of spatial hearing perception. Furthermore, many audio applications have been developed in attempts to reproduce original sound fields for listeners such as stereophonic and multi-channel (e.g. 5.1 channels and 22.2 channels) surround sound techniques. No matter how complex such sound systems become, humans can receive sound with only two ears. Therefore, binaural sound pressures should contain all acoustic information related to spatial hearing perception. From a physical perspective, sound waves from different positions interact with a listener’s anatomical structures (e.g., pinnae, head, and torso) with correspondingly different effects. In other words, the sound waves travel from a certain sound source in space and reach a listener’s ears after scattering, diffraction, and reflection from a listener’s pinnae, head, and torso. The processes of scattering, diffraction, and reflection convert the sound field information into the obtained binaural sound pressures. Therefore, the values of these binaural sound signals also vary according to the sound source directions. Human brains can then use these cues to ascertain the directions of sound arrival, which is the basic principle of human spatial hearing. Human spatial hearing ability is mainly attributed to the discrimination of acoustic frequency characteristics arising during sound transmission from source positions to a listener’s ears. These transmission characteristics are described by a head-related transfer function (HRTF), which is conventionally represented as a response along the frequency with spectrum characteristics strongly dependent on the sound source direction: each direction of the sound incidence corresponds to a pair of HRTFs with specific frequency characteristics. A pair of HRTFs enables synthesis of a virtual sound that seems to originate from a certain position in a free space by modifying the frequency components of any sound source. That is the basic method for a type of spatial audio rendering called binaural

synthesis. This notion, which was born in the 1970s, was first validated at Kobe University in 1980. For the past few decades, psychoacoustic experiments and audio techniques based on HRTFs have become popular and conventional modes of investigating spatial hearing and 3D audio application development.

Actually, HRTFs are obtainable by calculation or measurement at different directions. Although HRTF values vary with the source direction as well as the subject because of the different size and shape of human anatomical structures, they show little dependence on distance beyond 1 m. When plotting HRTF values at a certain distance along all directions on a sphere, the spherical surface would have many peaks, notches and spatial variations distributed at different regions on the sphere. Results of earlier psychoacoustic experiments suggest that the spatial characteristics within limited areas on the sphere, which are called HRTF local features, are extremely important for spatial hearing perception. Moreover, the directional resolutions of HRTFs required in binaural synthesis depend on the source direction. For these reasons, methods must be found for analyzing HRTFs locally. Nevertheless, to date no method exists which allows for locally analyzing HRTFs. Methods either evaluate an HRTF direction by direction, or consider all directions simultaneously. To address this difficulty, this study proposes a new method applying the notion of wavelets with the aim of better studying HRTF local features. This proposal is expected to open a new avenue for the study and application of HRTFs. The structure of this study to achieve the objectives above is presented below.

Chapter 1 introduces the motivation and background survey of this study. Based on a survey of earlier research and its limitations, the study objective is proposed. Chapter 2 reviews a conventional method of modeling HRTF spatial patterns, and illustrates that HRTF local features cannot be captured well using this conventional method. In Chapter 3, a method is proposed to model HRTF spatial patterns based on continuous spherical wavelet transform. Using this proposed method, the HRTF local features are modeled and the approximation errors are compared between the conventional method and the proposed method. Chapter 4 further extends the proposed method introduced in chapter 3 using the

discrete spherical wavelet transform for the modeling of HRTF spatial patterns. Because of the better characteristics of the analysis functions, the expansion coefficients are expected to control the spatial resolutions of the HRTF better and enable more efficient representations the HRTF local features. Chapter 5 describes the use of the proposed modeling method to recover the time-domain head-related impulse responses which are necessary for actual signal processing in binaural synthesis. In Chapter 6, an auditory model is introduced to evaluate the proposed method with respect to human hearing perception. Finally, the main conclusions of this thesis are presented in chapter 7.

Acknowledgements

This thesis wouldn't have been possible without the help of many people. First of all, I would like to express my best gratitude to my advisor, Professor SUZUKI Yôiti, who taught me a lot both in and out of the academic research. Prof. SUZUKI Yôiti offered me a lot of valuable advice on both my way of thinking and every detail of my study with his vast patience and support during the past five years. I'm sure that I will benefit constantly from his teaching throughout my life. I also greatly appreciate his kind understanding of my career choice and great support for my study and life, which is even beyond my imagination.

I am deeply grateful to Prof. ITO Akinori, Prof. SUGANUMA, Takuo, and Assoc. Prof. SAKAMOTO, Shuichi for their participation on my dissertation committee. Assoc. Prof. SAKAMOTO, Shuichi also offered me a lot of helpful advice both for my living and research since I came to Japan.

I also would like to express my great gratitude to Asst. Prof. TREVIÑO, Jorge in my laboratory, who is always willing to give his important comments and inspiring discussion at any time of my needs, without which I could not have overcome the difficulties of my research. Furthermore, Asst. Prof. TREVIÑO, Jorge sets a great example for me with his wide knowledge, ways of thinking and doing research, which I will benefit after my graduation.

I wish to acknowledge Prof. SHIOIRI, Satoshi, Prof. KITAMURA, Yoshifumi, Prof. MATSUMIYA, Kazumichi, Assoc. Prof. KURIKI, Ichiro, and all members of the Graduate

School Seminar for their valuable comments. Likewise, I express my acknowledgment to Eng. SAITO, Fumitaka, Asst. Prof. CUI, Zhenglie, and all the members of the Advance Acoustic Information Laboratory, for their hard work in evaluating the present research and offering valuable feedbacks. I extend my gratitude to Asst. Prof. CÉSAR Salvador, who helped me a lot on the beginning of my programming and offered me many advice and discussion. I am grateful to Ms. Miki ONODERA, our staff, who gave me valuable help for studying and living in Japan.

As a Christian, I would like to express my gratitude to the brothers and sisters at Sendai 2SC church, with whom we have been carrying each other constantly in our belief and life since I came to Japan.

I extend my gratitude to Prof. GUO, Jun, Vice President of my undergraduate university (BUPT), who was willing to introduce me to this laboratory. I also would like to express my gratitude to Prof. LI, Junfeng, who offered many valuable advice on my research and my life career.

Finally, I am profoundly thankful to everyone who have contributed to this important step: the references of this study, my colleagues in other laboratories, my friends in Sendai, and of course my family in my hometown.

Contents

Chapter 1

Introduction

1.1 Motivation of this study

Regarding the information perceived from a sound, apart from its loudness, pitch and timbre, the spatial information such as sound source directions and distances, also helps enrich the impression the perceived auditory scene. Although the spatial hearing ability is always unconsciously neglected by the listener compared with speech communication, it plays an important role in our daily life communication. The spatial hearing perception helps us be aware of the surrounding environment that always involves in a series of auditory events, such as the position of the speaker, the direction of the cars in the traffic and so on.

Since spatial hearing plays an important role in our daily life, it should not be missing in the design of current audio-related devices, such as the Virtual Auditory Display (VAD) devices, hearing aids, virtual reality (VR), augmented reality(AR), teleconference and so on. Therefore, it is of great importance for us to understand the mechanism of our spatial auditory system and principle of spatial hearing, in order to develop advanced audio applications.

The investigation of sound localization for human beings dates back to more than one hundred years ago [1]. Thanks to the modern signal processing technology, the knowledge

of sound localization has also been much investigated for the past several decades [2–4].

The perception of spatial hearing with only two ears largely depends on the interaural differences including the interaural time difference (ITD) and interaural level difference (ILD) [3]. These two localization cues provide the dominate cues for the sound localization in the horizontal plane. When a sound is presented from the side, the path from the sound source to the contralateral ear is interrupted by a listener's head. The far ear will be shadowed and a sound attenuation result in that side. ILD is direction dependent in the horizontal plane and can be employed by a person as a important cue for sound localization. The ILD due to head shadow effects depends on the wavelength of sound compared with the size of the head. For a sound source at the side, this interaural level difference can be as much as 35 dB at high frequencies [5]. For a sound source spanning frequencies below 1 kHz, the sound's wavelength can be several times larger than head and, in this case, the ILD between ears can be neglected [1]. For a low frequency sound, listeners are sensitive to the interaural time difference rather than ILD. However, the sensitivity to phase differences declines with increasing frequency up to a limit of 770Hz [1]. Another psychoacoustic experiment also showed that the performance of pure tone sound was worst for sounds of 1500-3000 Hz and better for lower and higher frequencies [6]. In the range around 1500 to 3000, the stimuli are too high in frequency to provide ITD and too low to provide ILD. There has been lots of investigation on the ITD and ILD in horizontal plane sound localization, whose results are consistent with the notion that the spatial information lies in ITD for low frequency and ILD for high frequency.

Apart from the binaural differences, the spectral features caused by the reflection and diffraction by the pinnae, head and torso also provide important cues for vertical sound localization and front-back disambiguity [7]. These spectral cues are important especially for relative higher frequencies. There is also research showing that the first and second notches of HRTFs can be regarded as important spectral cues [8].

The traditional way of investigating the spatial hearing perception is based on the above cues, which can be further explained by the so called head-related transfer function and its

binaural differences introduced in the next. However, the perception of sounds' directions depend on many factors, which also includes the dynamic cues(e.g., head movement or dynamic sound source;) [9] and the properties and number of sound sources [10]. Visual information can also influence the perception of direction of sounds [11]. Therefore, the spatial hearing perception is a type of multimodal perception in which the dominate cues can be regarded as the interaural differences and the spectral cues, as well as the dynamic cues. This study is aimed at investigating the human spatial hearing by means of analyzing the head related transfer functions.

1.2 The head-related transfer function

Here, it is denoted that a sound source direction in space is specified by its azimuthal angle $\theta \in [-180°, 180°]$ and elevation angle $\phi \in [-90°, 90°]$, in which $(0°, 0°)$ and $(90°, 0°)$ represent the front and left direction, respectively. This spherical coordinate is shown as Fig. 1.1

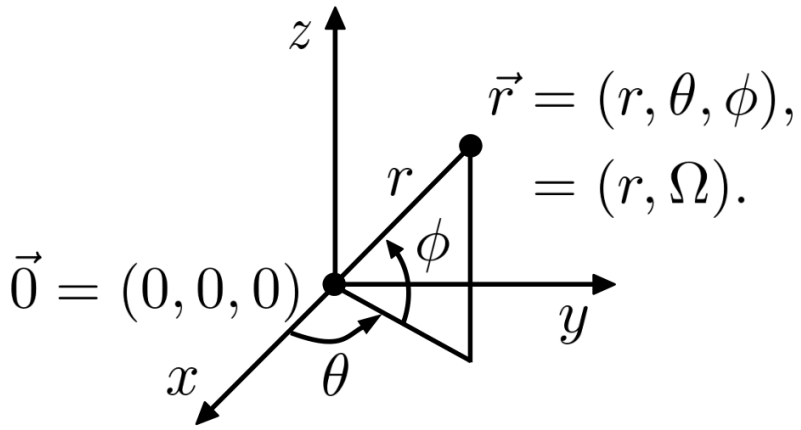


Figure 1.1: Spherical coordinate system.

Head-related transfer functions (HRTF) describe the sound transmission from the free field to the ear drum in space [2, 3, 12, 13]. This function, $H_s(\theta, \phi, r, f)$, is calculated as the complex-valued frequency response of sound pressure at the ear drum $P_e(\theta, \phi, r_s, f)$,

divided by that in the center of the head when it is absent ($P_c(\theta, \phi, r, f)$):

$$H_s(\theta, \phi, r, f) = \frac{P_e(\theta, \phi, r_s, f)}{P_c(\theta, \phi, r, f)}, \tag{1.1}$$

where r_s and r are the distances from the sound source to the evaluated ear and the center of the head, respectively. (θ, ϕ) denotes the azimuth and elevation angles of the sound source whose frequency is f in space. The ITD and ILD can be calculated by the differences of time and magnitudes of the left and right signal, both of which depends on the frequency. As described in Eq. (1.1), the HRTF depends on the direction, frequency and distance to the sound source. Since it is individual variable related to the listener's anthropometric parameters, they vary among different subjects.

Figure 1.2: Loudspeaker array for measuring HRTF in Tohoku University.

HRTF is conventionally obtained through measurement for discrete samplings of directions, using a set of loudspeaker array Fig.(1.2). To measure the HRTF. an ideal impulse

signal is a Dirac delta function which is impossible to be create in practice. An alternative way is to use other excitation signals such as the TSP signal [14] or a sinusoidal signal [3]. The measured values of HRTF for different individuals are usually stored as an HRTF dataset. Nowadays, there are many available HRTF datasets in the world, such as the dataset for KEMAR by the MIT Media Lab, the CIPIC dataset, the Tohoku University dataset, and the ARI HRTF dataset, and so on [15–19].

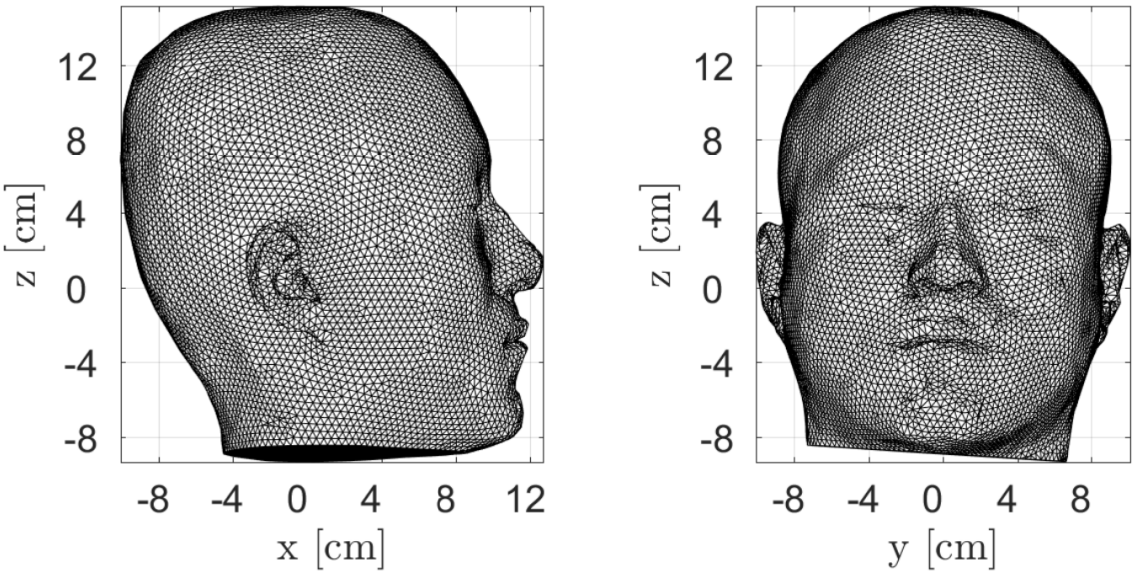


Figure 1.3: An individual head model.

Apart from measurement, HRTF can be calculated as well using a numerical simulation like the boundary element method (BEM) and Finite-Difference Time-Domain method (FDTD) , which is a widely used method for acoustic radiation and scattering problems [20–22]. To calcute the HRTF using the simulation methods, a 3D scanner of a the surfaces of a human or artificial head is needed, as shown in Fig. 1.3. [23–27] The calculation of the HRTF can be time consuming; however, it can cover more positions than the measuring method which only provides data for positions where a loudspeaker is available. Besides, with the BEM approach, HRTFs for the near field (e.g. the distance of the sound source < 0.25m) can also be calculated, while it can be very demanding to measure. Alternative calculations of the transfer functions for very simple models like a rigid sphere or a snowman model, rather than the head model have also been studied [28, 29].

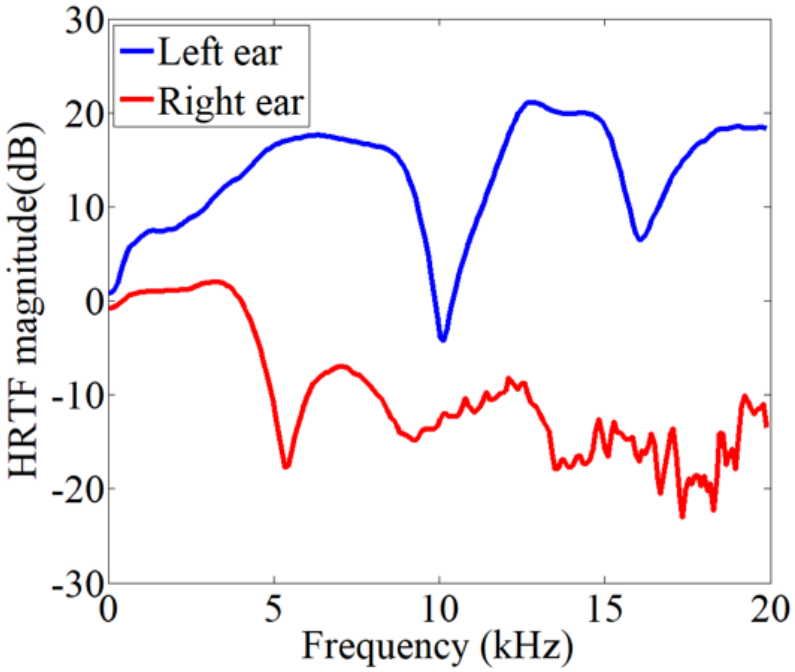


Figure 1.4: HRTFs at azimuth 80 deg. and elevation 0 deg.

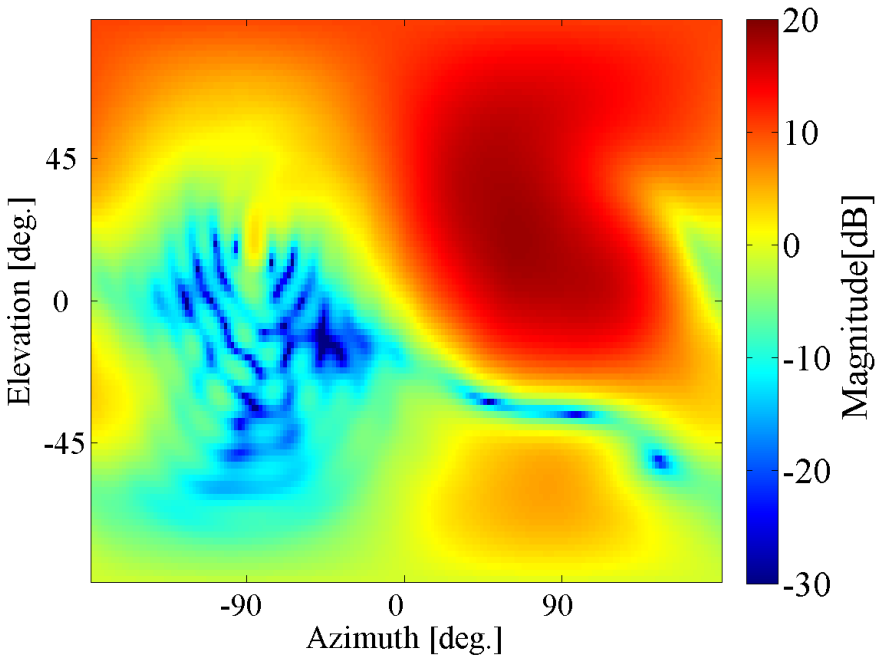


Figure 1.5: HRTF for every two degrees in azimuth (from -180° to 180°) and elevation (from -90° to 90°) angles at a frequency of 7.4 kHz.

As described by Eq. (1.1), the HRTF is a function of several variables including the frequeny and direction of the sound source. Therefore, the values of HRTF vary for different frequencies and directions. Fig. (1.4) shows the HRTF for a sound source in direction

(80°, 0°) over the full audible frequency range. Fig. (1.5) shows the BEM based calculated HRTF at every two degrees for the azimuth (from $-180°$ to 180°) and elevation (from $-90°$ to 90°) angles at the frequency 7.4 kHz. This results in a total of 10,622 directions. The HRTF plotted in the spatial domain (along direction) clearly shows its direction dependency for sound sources at different directions; this is very convenient for the observation and analysis of the data as the spatial variations of the HRTF introduced by the reflections and diffractions can be clearly revealed.

1.3 HRTF modeling

HRTF datasets contain discrete samplings; these raw data is difficult to handle by examining the data point by point. Modeling HRTF has many benefits; it helps to observe the data in different domains and analyze the properties of the HRTFs conveniently. It is also important to model the HRTF for further processing of HRTF such as interpolation, extrapolation, etc. Furthermore, modeling the HRTF with a small number of parameters can greatly reduce the size of the datasets, which is important to store and transmit them. The HRTF is a function of direction and frequency, these functions can be modeled in these two domains [30]. For the past decades, many attempts have been made to model HRTF, both in the frequency and spatial domains.

1.3.1 HRTF modeling in the frequency domain

The HRTF can be regarded as the frequency response of a linear time invariant (LTI) system. Filters models for LTI system and signal processing methods like principle component analysis (PCA) [31] and filter models [32] are widely used to model the HRTF in the frequency domain. The purpose of HRTF modeling is to approximate the original data with a simple formula; therefore, approximation error is an important metric to evaluate different modeling method. A commonly used way of evaluating the approximation error

is based on the spectral distortion which is given by the following equation [33, 34]:

$$\text{Erms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[20\log_{10}\frac{|H_{\text{synth}}(f_i)|}{|H_{\text{target}}(f_i)|}\right]^2}, \tag{1.2}$$

where $H_{\text{synth}}(f_i)$ and $H_{\text{target}}(f_i)$ are the approximated and target HRTF at the frequency f_i, respectively. N is the total number of sampled frequencies.

Filter models for HRTF

According to the theory of signal processing, the general form of describing a differential equation for a discrete-time LTI system is as follows:

$$x(k) = -\sum_{i=1}^{n} a_i x(k-i) + \sum_{i=0}^{m} b_i u(u-i), \tag{1.3}$$

$$H(z) = \frac{\sum_{i=0}^{m} b_i z^{-i}}{\sum_{i=0}^{n} a_i z^{-i}} = \frac{B(z)}{A(z)}, \tag{1.4}$$

where $u(k)$ is the input to the system and $x(k)$ is the corresponding output. The order (m, n) of the numerator (m) and denominator (n) polynomials $B(z)$ and $A(z)$ correspond to the number of zeros and poles of the transfer functions, respectively. This system model is called the autoregressive moving-average model (ARMA) [7, 35]. The HRTF for a certain direction can be represented by these zeros and poles, this reduces the data size effectively [7]. To calculate the coefficients in this model, a common criteria is to minimizes a quadratic expression: the difference in energy between the target and synthetic data. The HRTF can be smoothed by controlling the order of the model [7]. Some studies use other optimization methods to do the calculation, such as the frequency warping method [36]. The HRTF data can also be preprocessed to obtain its directional transfer function (DTF), by subtracting the average of HRTF for all directions from each HRTF.

A similar way to model the HRTF in the frequency domain is based on a finite impulse

response (FIR) model representing the HRTF with only zeros [37]:

$$x(k) = \sum_{i=0}^{m} b_i u(u-i), \tag{1.5}$$

$$H(z) = \sum_{i=0}^{m} b_i z^{-i} = B(z), \tag{1.6}$$

Another method, based on the ARMA model, is the common-acoustical-pole and zero(CAPZ) model for the HRTF [38]. This method assumes that this HRTF set shares the same poles in the approximation. The order for modeling can be effectively reduced with this method.

Principle component analysis for HRTF modeling

Principle Components Analysis (PCA) is a statistical processing method which derives several basis functions from an HRTF dataset and decomposes the target data using these basis functions. Each HRTF can be represented as a weighted sum of these basis functions [39, 40]:

$$\mathbf{d_k} = \mathbf{C}\mathbf{w_k}, \tag{1.7}$$

where $\mathbf{C}$ is a matrix, the colums of which are the selected eigenvectors. $\mathbf{d_k}$ is the k-th HRTF or DTF. $\mathbf{w_k}$ is the vector of the coefficients, representing the contribution of each basis function to the approximation.

In Kisterler's work [40], it was found that the HRTFs can be modeled as a linear combination of five basic spectral shapes (basis functions), which accounts for approximately ninety percent of the variance in the original HRTF magnitude functions.

It should be noticed that the basis component functions are derived from the target HRTF set; this means that the basis functions vary for different target datasets. Although the flexibility of target dependent basis functions allows for a more efficient modeling and compression of HRTF data, it is also a barrier for analyzing and comparing HRTFs from different datasets.

1.3.2 HRTF modeling in the spatial domain

Representing the HRTF in the spatial domain has many advantages: it allows for the observation of the spatial variations introduced by the acoustic scattering of the head for different directions. Besides, looking at the HRTF in the spatial domain allows for a quick comparison of HRTFs between different subjects and for different frequencies. An example of the HRTF magnitudes in the spatial domain are plotted as in Fig. 1.6.

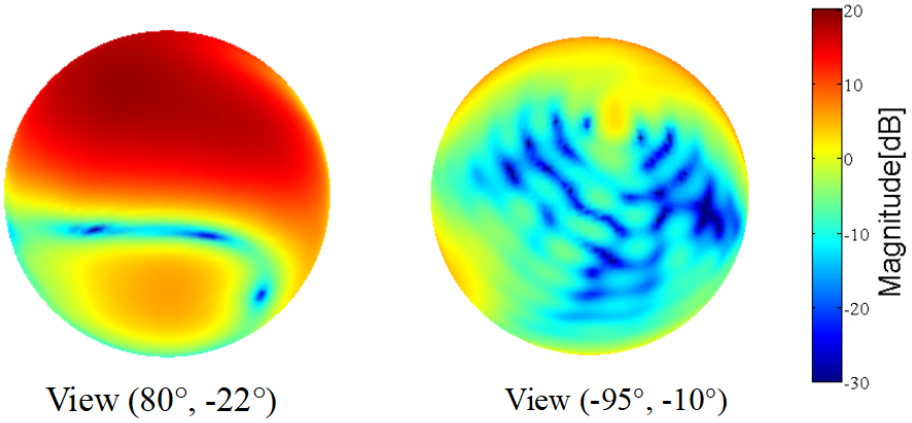


Figure 1.6: Plotting the HRTF magnitudes in the spatial domain.

The HRTF can be decomposed into a weighted sum of spatial basis functions. The HRTF of a given ear can be generally represented as:

$$H(\theta, \phi) = \sum_{i=1}^{N} c_i f_i(\theta, \phi) + \epsilon, \tag{1.8}$$

where $H(\theta, \phi)$ is the target HRTF data set at set frequency at azimuth and elevation angles θ and ϕ, respectively. If given the spatial basis functions $f_i(\theta, \phi)$ with a total number of N, the corresponding coefficients c_i can be obtained in such a way to represent the target HRTF data in the spatial domain.

Various methods are used to select or derive the spatial basis function $w_i(\theta, \phi)$. These basis function are usually chosen as spatial PCA , Legendre polynomials, or the spherical harmonics [41–43]. The spherical harmonic decomposition has been widely applied to HRTF modeling in the past decades due to its several advantages. The spherical harmonics

form an orthogonal basis for functions on the sphere, low orders give an ideal approximation of the coarse structures of in the target dataset [44]. Modeling methods using the spherical harmonics also allow for interpolation since the basis functions are continuous functions over all direction [45, 46].

1.3.3 Local analysis of the HRTFs

Although the HRTF spatial patters have been much studied with spherical harmonics, only the overall spatial characteristics are described with this method. However, as suggested by the previous study, humans localize sounds with different resolutions at different directions as shown in Fig. 1.7 [47]. Besides, the required directional resolutions of the HRTF for convincing spatial audio rendering also vary according to the directions (Fig. 1.8) [48]. Based on these perceptual evidences, HRTFs should be considered to be modeled locally at different directions.

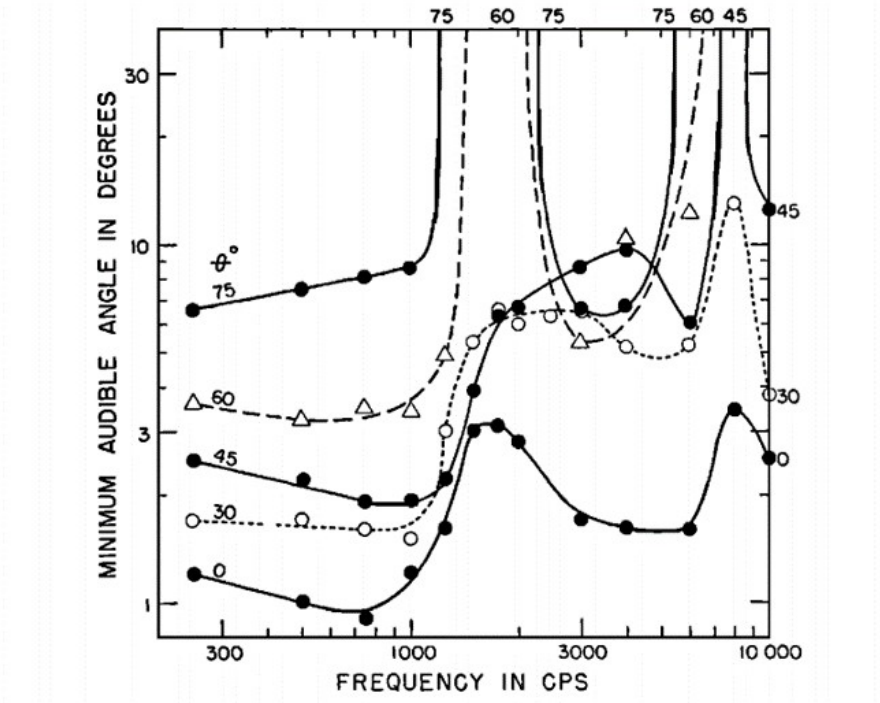


Figure 1.7: Average minimum audible angle as a function of the stimulus frequency (Mills AW,1958). [47]

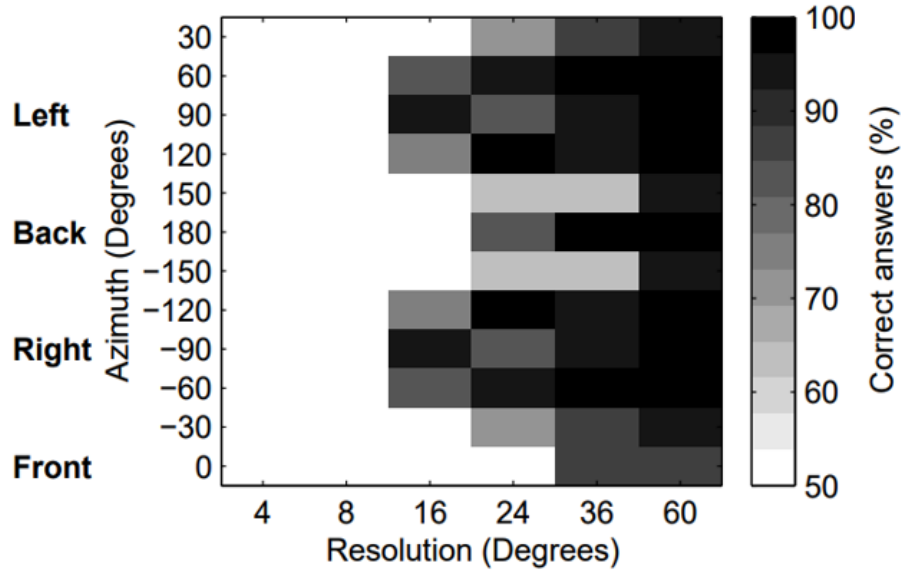


Figure 1.8: Required directional resolutions of HRTFs for binaural audio rendering (Minnaar P, et al.,2000). [48]

1.3.4 Summary of the exsiting modeling methods

Modeling HRTF has many advantages. Basically, HRTF modeling in the frequency domain allows for convenient frequency interpolation while modeling in the spatial domain allows for convenient spatial interpolation. Besides, modeling HRTF have some other benefits as summarized in Table 1.1.

Table 1.1: Benefits of modeling HRTFs

Benefits	HRTF modeling methods	
Smoothing HRTFs for designing psychoacoustic tests	Smoothing in frequency domain	• IIR filter [7] • PCA [39]
	Smoothing in the spatial domain	• Spherical harmonics [44]
Interpolating HRTF for unknown directions	In frequency domain	• Common pole zero model [49]
	In the spatial domain	Interpolating ITD • Spherical harmonics [45] Interpolating HRTF magnitudes • Spherical harmonics [46,50]
Extrapolation of HRTF for unknown distances	Along the distance	• Spherical harmonics [46,50] • Fourier legend transform [51]
Preprocessing for HRTF personalization	HRTF magnitudes	• PCA [52, 53]
	Interaural cues	Sinusoidal models • For ITD [54] • For ILD [55]

Table 1.2: Existing method of HRTF modeling.

Domain	**Method**	**Compression**	**Frequency interpolation**	**Spatial interpolation**	**local analysis**
Frequency domain	Principle component analysis(PCA) [39, 40]	Possible	Possible	Possible but limited	Not possible
Frequency domain	Filter models(FIR, IIR) [7, 37]	Possible	Possible	Possible but limited	Not possible
Spatial domain	Spherical harmonic decomposition [41]	Possible	Not possible	Possible	Not possible

Table 1.2 shows some of the existing methods for HRTF modeling. All of them allows for data compression, either in the frequency domain, or the spatial domain. The spherical harmonic decomposition is a conventional method for modeling of the HRTF in the spatial domain, since it provides good modeling efficiency, especially at low frequencies. Besides, the use of continuous basis functions allows for the spatial interpolation of HRTF. Nevertheless, there are still some limitations to this method:

- The expansion coefficients of the spherical harmonic decomposition carry information for all directions. This is because the basis functions take significant values for all directions. The analysis of those direction dependent local features is difficult with this method.

- Some local features of the HRTF require a large number of harmonics to be approximated; this is not very efficient.

- The target HRTF dataset to be decomposed using the spherical harmonics requires data for all directions. In some cases, the HRTF data is available only for a local region.

- Perceptual studies suggest that the angular resolution needed to characterize human sound localization depends on source's direction [56]. The directional resolution of the HRTF required for binaural synthesis varies for all directions [48]. The spherical harmonic expansion considers all directions simultaneously. Local analysis is difficult using methods based on the spherical harmonics.

Due to the above limitations of spatial modeling based on the spherical harmonic expansion. A way is necessary to capture local features of the HRTF.

1.4 Objective

HRTFs are functions of frequency and direction. As these two variables are independent, the HRTF can be modeled along one or the other. This thesis follows the latter, since qualitative analysis is more easier in the spatial domain than the frequency domain. The variations of HRTF magnitudes in the spatial domain exhibit many direction dependent features introduced by the filtering of pinnae, head and torso; this is an important property of HRTF whose values much rely on the directions.

None of the existing modeling methods allow for local analysis of the HRTF. This thesis seeks to develop a new method to model the HRTF in the spatial domain based on a set of wavelets on the sphere which take significant values only on a small region on the sphere. This method is expected to allows for a local reconstruction of the HRTF. This method is also expected to be capable of representing local features of the HRTF more efficiently than the existing methods. A spatial correspondence between the local variations of the

HRTF and the expansion coefficients is also expected. This method way may open a new approach to the study of the HRTF, whose values largely depend on the sound source's direction. The following figure shows the structure and research map of this thesis.

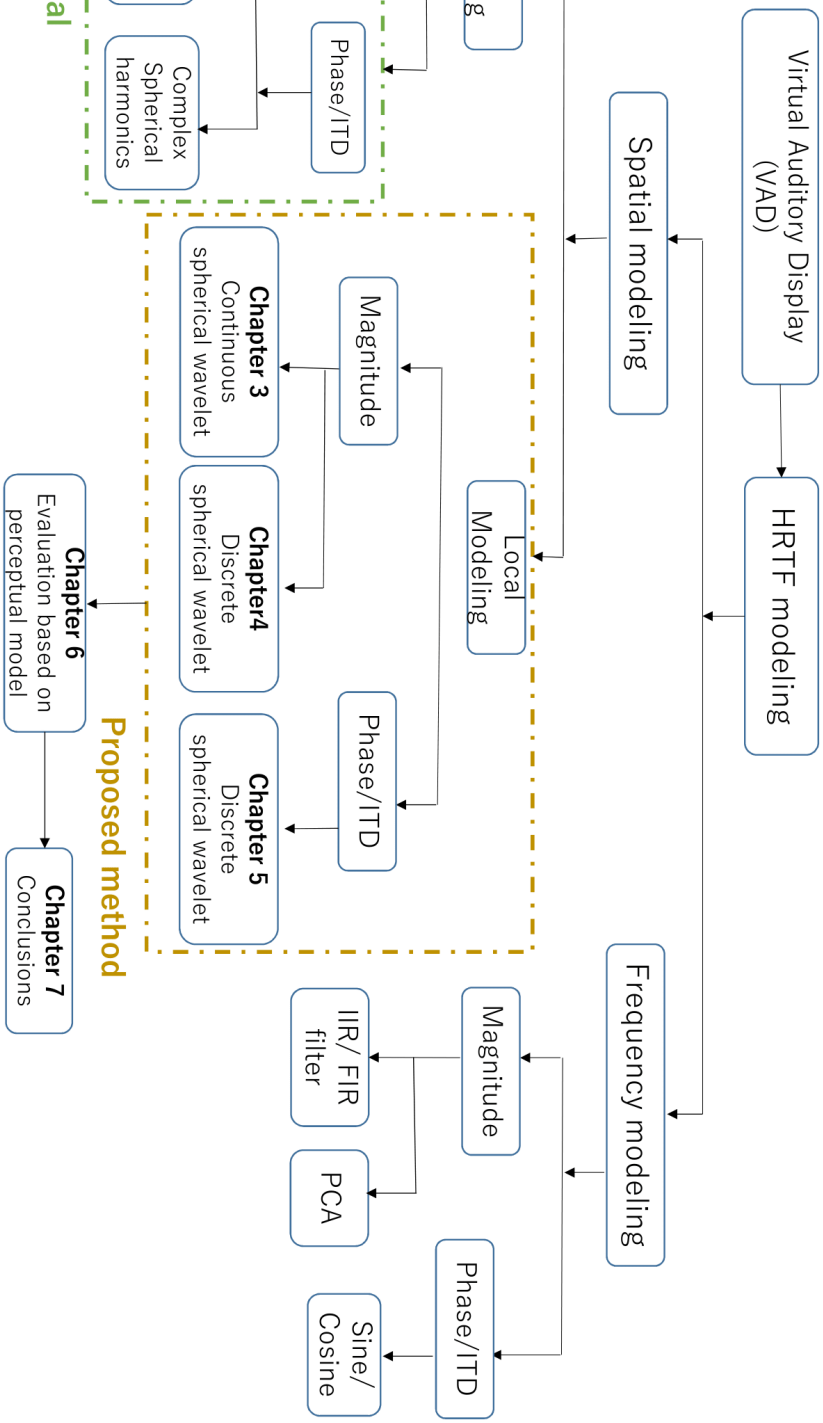


Figure 1.9: Research map of this thesis.

Chapter 2

Modeling of head-related transfer function based on spherical harmonics

2.1 Introduction

This chapter introduces spherical harmonic functions which are a set of orthogonal functions covering different spatial frequencies on the sphere. These functions can be used to represent functions defined on the surface of a sphere. Here, the HRTF dataset for all directions are decomposed by these basis functions into a series of coefficients. The HRTF dataset thus can be represented by these coefficients. Section 2.2 briefly introduces the preliminary theory to represent HRTF dataset with a set of basis functions, the spherical harmonics. Section 2.3 validate this modeling method by simulations. Section 2.4 summarizes the content of this chapter.

2.2 Preliminaries

The spherical harmonic functions are derived as the solution to the directional part of the Laplace's equation. An arbitrary surface function on the sphere can be represented a set of spherical harmonics of different spatial patterns. The spherical harmonic transform basically decompose a target function on the sphere into a set of expansion coefficients. Therefore, the original surface function on the sphere may be described with those decomposed expansion coefficients. Since this study focuses on the HRTF magnitudes in the spatial domain. The HRTF magnitudes for all directions are modeled with the spherical harmonic decomposition.

2.2.1 Spherical harmonics

Assume that the solution function ψ for the Laplace's equation is harmonic. The Laplace's equation in the standard spherical coordinate system is the partial differential equation [57]

$$\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]\psi(r,\theta,\phi)=0. \tag{2.1}$$

Here r denotes the radial distance. θ, and ϕ describe the inclination angle and the azimuth angle, respectively.

If under the assumption that ψ is independent on r, the solutions to Eq. (2.1) can be described as $\psi = \Theta_{nm}(\theta)\Phi_m(\phi)$. Therefore, the original Laplace's equation can be divided into two differential equations

$$\left[\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d}{d\theta}\right)+n(n+1)-\frac{m^2}{\sin^2\theta}\right]\Theta(\theta)=0, \tag{2.2}$$

$$\left[\frac{d^2}{d\phi^2}+m^2\right]\Phi(\phi)=0. \tag{2.3}$$

Eq. (2.2) can be solved by the transformation of variables $\zeta = \cos\theta$, which derives

the Legendre differential equation. Eq. (2.3) can be regarded as a complex exponential function of ϕ. The above two solutions are combined as a joint function $Y_{nm} = \Theta_{nm}\Phi_m$, which is called the spherical harmonic function of order n and degree m, [57]

$$Y_{nm}(\theta, \phi) = N_{nm} P_n^m(\cos\theta) e^{im\phi}, \tag{2.4}$$

where P_n^m is the associated Legendre function and N_{nm} is a normalization coefficient.

The associated Legendre function can be described as [57]

$$P_n^m(\zeta) = (-1)^m (1-\zeta^2)^{m/2} \frac{1}{2^n n!} \frac{d^{n+m}}{d\zeta^{n+m}} (\zeta^2 - 1)^n, \tag{2.5}$$

where $\zeta \in [-1, 1]$, and n is a positive integer defining the order, while $m \in [-n, n]$ denotes the degree of the sollution function. The above sollution can be further extended based on the following descriptions.

$$P_{-n-1}^m(\zeta) = P_n^m(\zeta), \tag{2.6}$$

$$P_n^{-m}(\zeta) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(\zeta). \tag{2.7}$$

The normalization coefficient N_{nm} is chosen such that $\int_0^{2\pi} \int_0^{\pi} |Y_{nm}(\theta, \phi)|^2 \sin\theta d\theta d\phi = 1$, that is

$$N_{nm} = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}}. \tag{2.8}$$

Up to now, the above solutions deal with complex values where the analysis functions, the spherical harmonics, are of complex values, too. The real-valued spherical harmonics read as below.

$$Y_n^m(\theta, \phi) = \begin{cases} (-1)^{m+1} \sqrt{\frac{2n+1}{2\pi} \frac{(n+m)!}{(n-m)!}} P_n^{-m}(\sin\phi) \cos(m\theta) & m > 0 \\ \sqrt{\frac{2n+1}{4\pi}} P_n^0 \sin(\theta) & m = 0 \\ (-1)^m \sqrt{\frac{2n+1}{2\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\sin\phi) \sin(m\theta) & m < 0 \end{cases} \tag{2.9}$$

The plotting of the spherical harmonic Y_n^m is shown as Fig. 2.1. As can be observed,

the spatial variations of spherical harmonics for higher orders change more rapidly on the sphere, which enables to cover higher spatial frequency.

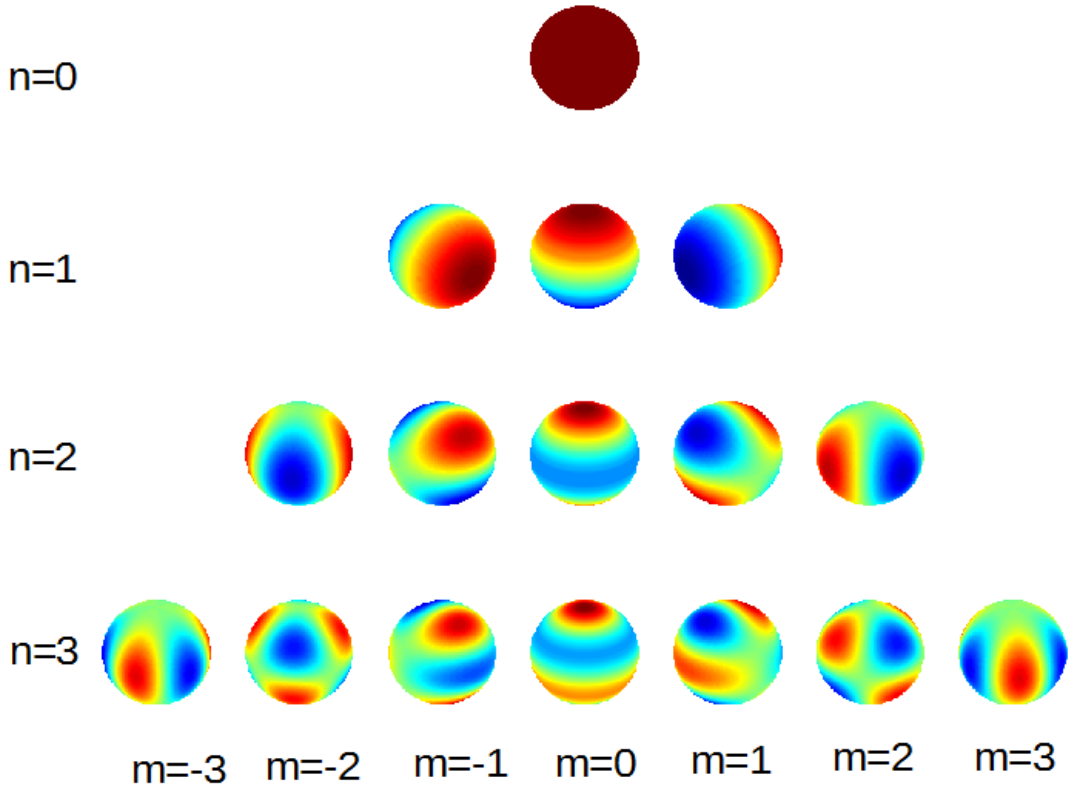


Figure 2.1: Spherical harmonic Y_n^m of degree m and order n ($n = 1, 2, 3$).

Spherical harmonic transforms

The integral calculation along the direction $\Omega = (\theta, \phi)$ on a sphere $\mathbb{S}^2$ with a unit radius can be described as

$$\int_{\Omega \in \mathbb{S}^2} d\Omega = \int_0^{2\pi} \int_0^{\pi} \sin\theta d\theta d\phi. \tag{2.10}$$

An important property of the spherical harmonic functions is that each one is orthonormal to another.

$$\int_{\Omega \in \mathbb{S}^2} Y_{nm}(\Omega) Y^*_{n'm'}(\Omega) d\Omega = \delta_{n-n',m-m'}. \tag{2.11}$$

Here * is the complex conjugate operator and $\delta_{i,j}$ the Kronecker Delta. They are also

complete on $\mathbb{S}^2$:

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_{nm}(\Omega)Y_{nm}^{*}(\Omega') = \delta(\theta,\theta')\delta(\phi,\phi'), \tag{2.12}$$

where $\delta(x)$ denotes the Dirac's delta function.

After defining the spherical harmonics as above, the decomposition of a target function $f(\Omega)$ on $\mathbb{S}^2$ can be represented as a weighted sum of spherical harmonics as follow [58]:

$$f(\Omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_{nm}Y_{nm}(\Omega), \tag{2.13}$$

where the expansion coefficients f_{nm} denotes the spherical wave spectrum, which can be calculated as below [58]:

$$f_{nm} = \int_{\Omega\in\mathbb{S}^2} f(\Omega)Y_{nm}^{*}(\Omega)d\Omega. \tag{2.14}$$

As for the case of real-valued spherical harmonics, the transform can be described as below.

$$f_{nm} = \int_{-\pi}^{\pi}\int_{-\pi/2}^{\pi/2} f(\theta,\phi)Y_{n}^{m}(\theta,\phi)\sin(\phi)d\phi d\theta. \tag{2.15}$$

After the spherical transform, the expansion coefficients can be used to represent the target dataset which is always a point cloud. For the decomposition of discrete target data, it can be computed with least square method in which the analysis functions, spherical harmonics, are discretized in space [59].

2.2.2 Representing HRTF magnitudes with spherical harmonics

As HRTF are functions of direction in space, many attempts have been made to model the spatial part of HRTF using spherical harmonic functions [41]. For each frequency, the HRTF magnitudes can be represented as a weighted sum of spherical harmonics described by the following function:

$$H(\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} B_{nm}\cdot Y_{nm}(\theta,\phi), \tag{2.16}$$

where $H(\theta, \phi)$ is a dataset of HRTF magnitudes at a certain frequency. $Y_{nm}(\theta, \phi)$ is the spherical harmonic function of order n and degree m. B_{nm} is the coefficient of $Y_{nm}(\theta, \phi)$ in the approximation. In practice, the order n is truncated to a maximum order N, in such a way, Eq. (2.17) becomes:

$$H(\theta, \phi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} B_{nm} \cdot Y_{nm}(\theta, \phi) + \epsilon, \tag{2.17}$$

where ϵ is the error in the approximation. To calculate the coefficients B_{nm}, we can not use the Eq. (2.15) directly as it is for a continuous function. However, the HRTF dataset are commonly obtained by measurement or calculation for discrete directions. Here a least square method can be used for the calculation if the dataset and its transform is in matrix form:

$$\mathbf{H} = \mathbf{W}\mathbf{C} + \epsilon, \tag{2.18}$$

Where $\mathbf{H}$ is the vector target HRTF dataset. $\mathbf{W}$ is the matrix of spherical harmonics, in which each column is spherical harmonic for a certain order and degree. $\mathbf{H}$ is the vector of expansion coefficients.

$$\mathbf{C} = \mathbf{W}^{+}\mathbf{H}, \tag{2.19}$$

- Once we can obtain the expansion coefficients, we can use them to represent the HRTF data size which usually has a big data size. An advantage of spherical harmonics is that it allows for convenient interpolation of HRTF as long as the expansion coefficients is known.

2.3 Numerical evaluation

This method is tested by applying it to the target HRTF data set calculated using the Boundary Element Method for the SAMRAI dummy head ([23]). The HRTF dataset for sound sources at 1.5 m were calculated at frequencies between 93.75 to 20, 000 Hz with the interval of 93.75 Hz and samples every two degrees in azimuth and elevation angles for a total of 16, 022 directions. Here the target HRTF is considered to be the magnitudes

values, and the real-valued spherical harmonics are used in the approximation ([41]).

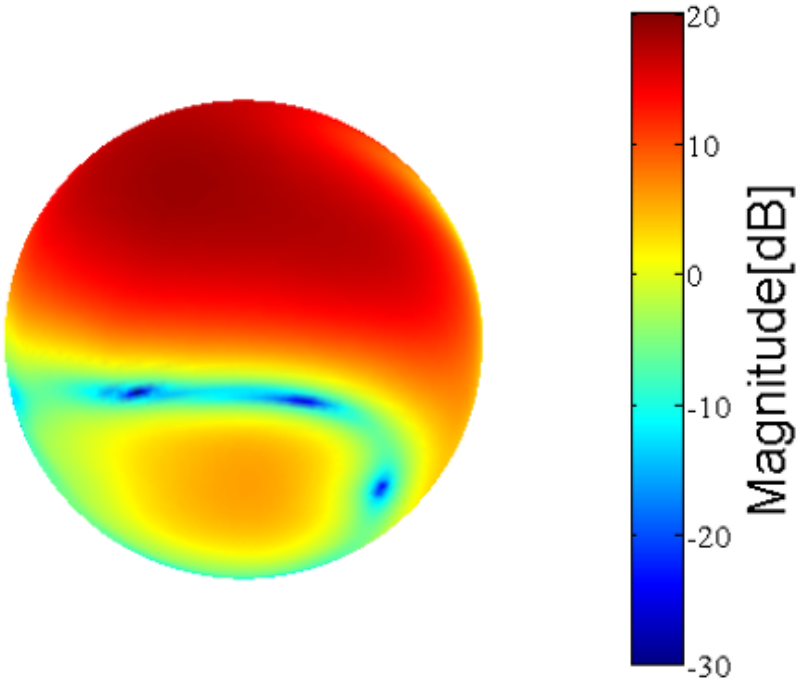
For an objective evaluation, a conventional measure of approximation error in the frequency domain is defined using the Root Mean Squared (RMS) value ([60]), here a relative error for evaluating the approximation accuracy in the spatial domain is defined as:

$$E_{mnl} = \frac{1}{N}\sum_{m=1}^{N}\frac{|H_{synth}(\theta_m,\phi_m) - H_{target}(\theta_m,\phi_m)|}{|H_{target}(\theta_m,\phi_m)|}, \tag{2.20}$$

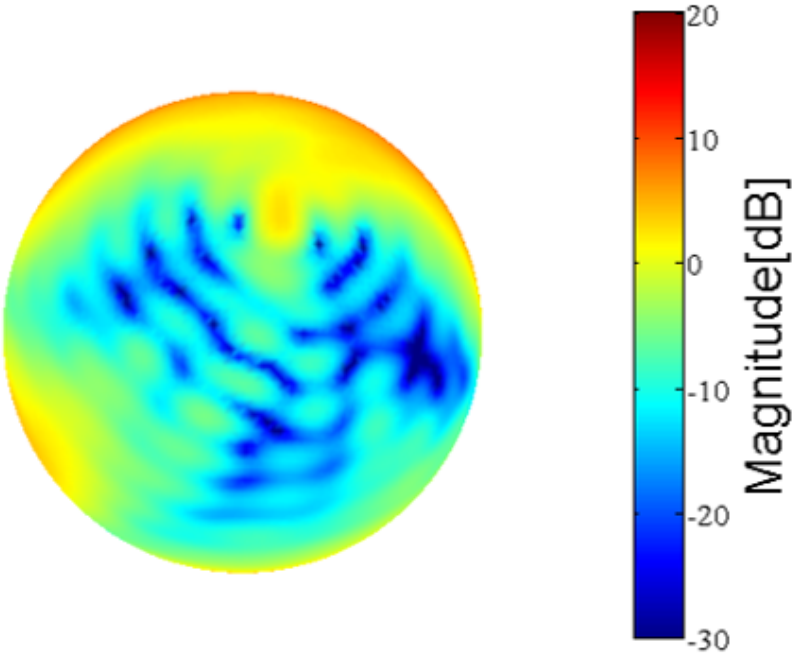
where $H_{synth}(\theta_i,\phi_i)$ and $H_{target}(\theta_i,\phi_i)$ are the reconstructed HRTF and target HRTF at the sampling direction (θ_i,ϕ_i), respectively; N is the total number of the HRTF samples under study.

2.3.1 Parameters and conditions for the evaluation

The evaluated dataset is a calculated HRTF using BEM method [23] for all directions at the frequency of 7.4 kHz at a distance of 1.5 m. The magnitudes of this target HRTF is shown in Fig. (2.4).



(a) View(170°,−22°)

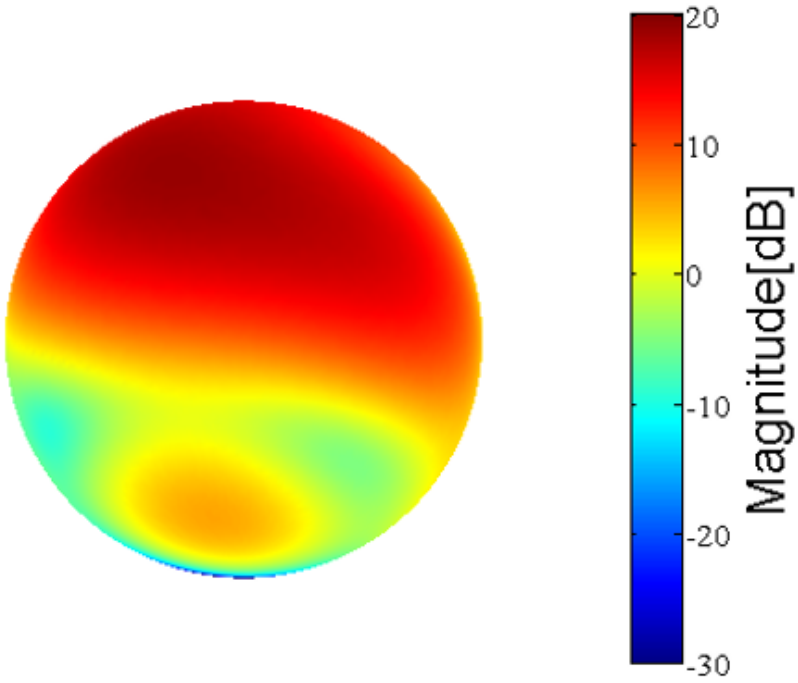


(b) View(−5°,−10°)

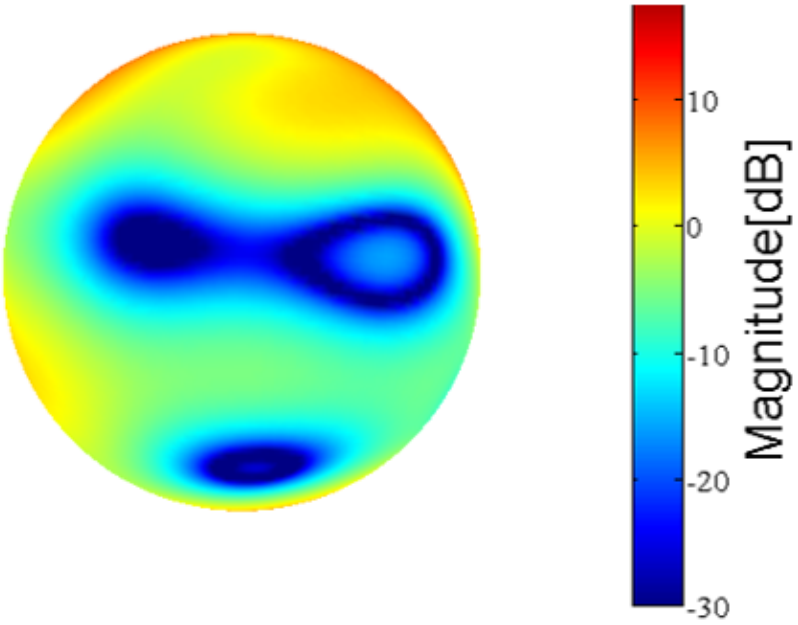
Figure 2.2: Target HRTF to be modeled

2.3.2 Simulation results

To calculate the expansion coefficients, the least squared method is used for the discrete samplings of HRTF on the sphere as shown in Eq. (2.19)

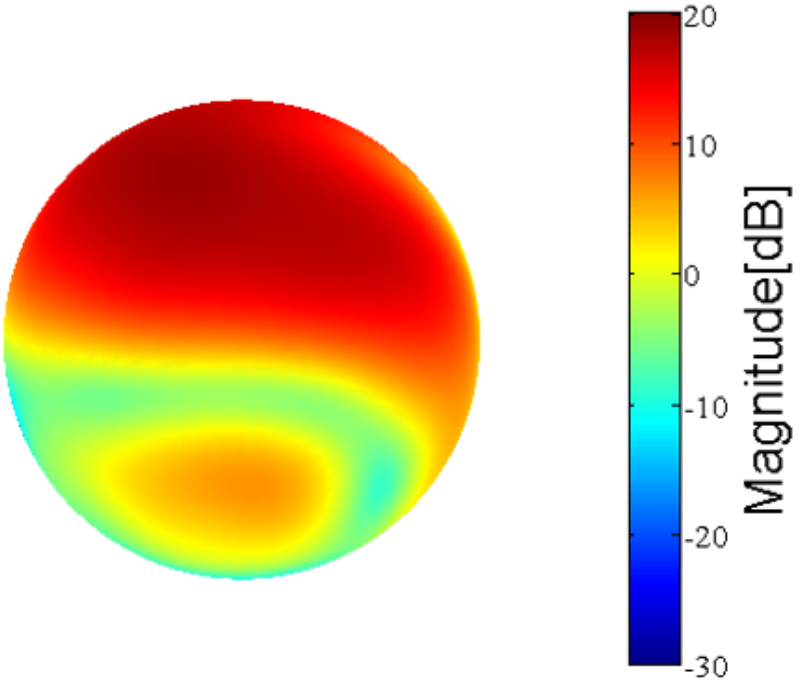


(a) View(170°,−22°)

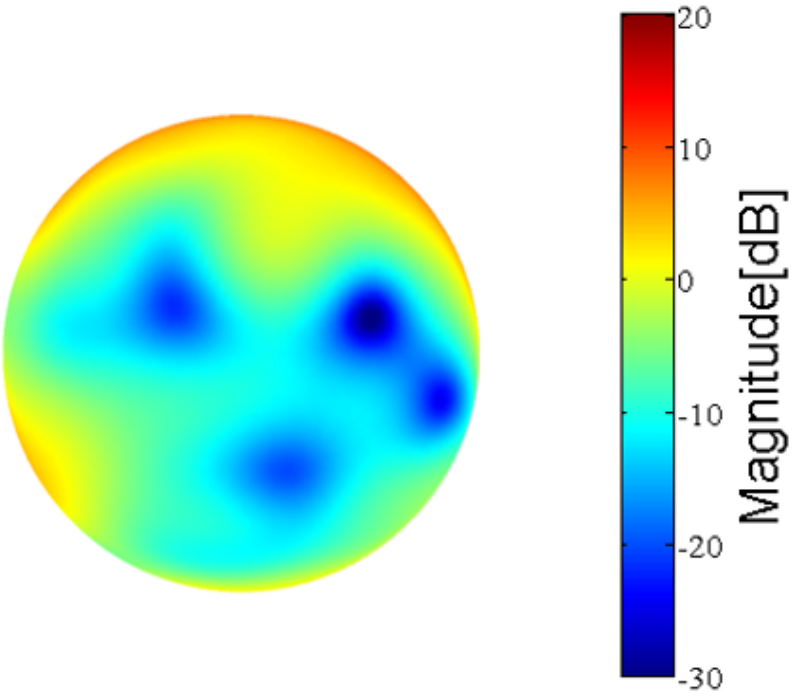


(b) View(−5°,−10°)

Figure 2.3: Reconstructed HRTF using spherical harmonics up to order 5.

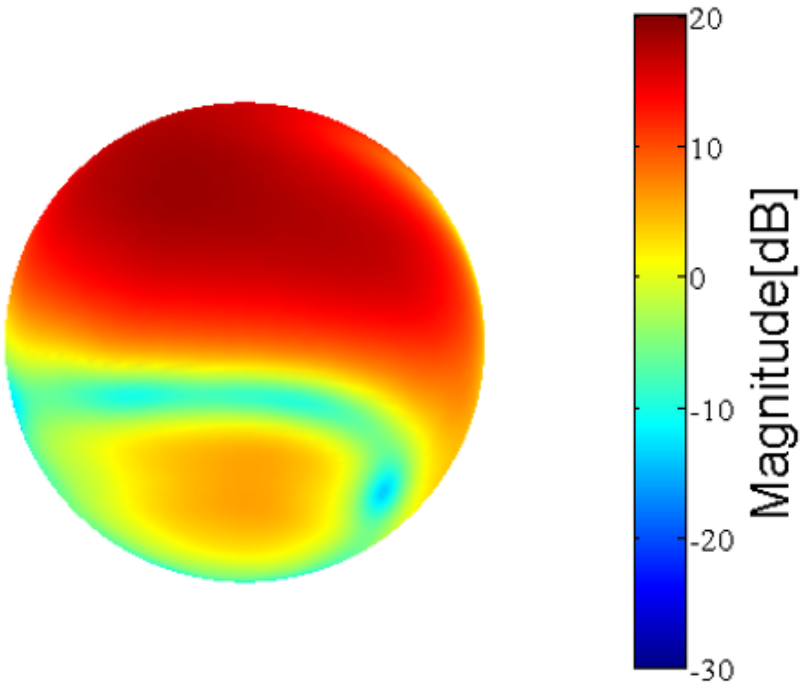


(a) View(170°,−22°)

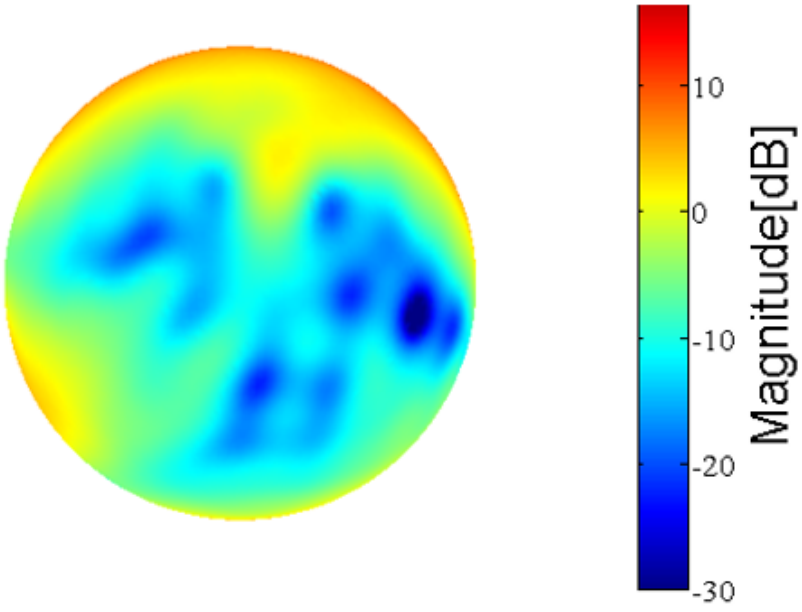


(b) View(−5°,−10°)

Figure 2.4: Reconstructed HRTF using spherical harmonics up to order 10.

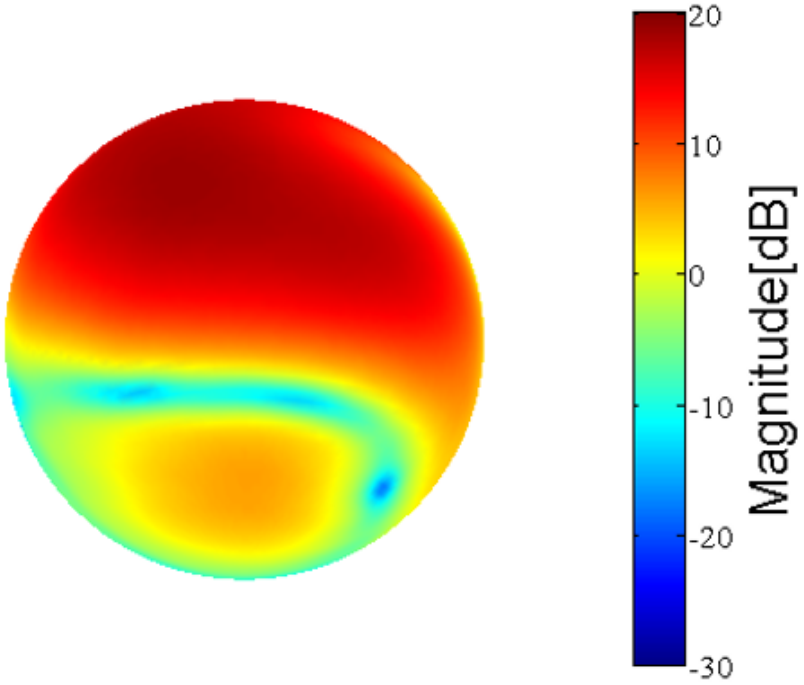


(a) View(170°,−22°)

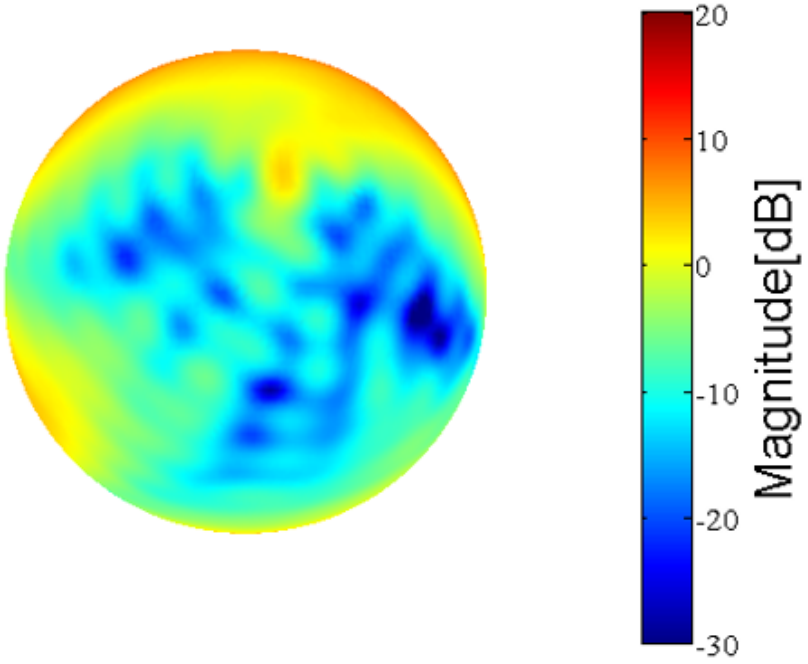


(b) View(−5°,−10°)

Figure 2.5: Reconstructed HRTF using spherical harmonics up to order 20.



(a) View(170°,−22°)



(b) View(−5°,−10°)

Figure 2.6: Reconstructed HRTF using spherical harmonics up to order 30.

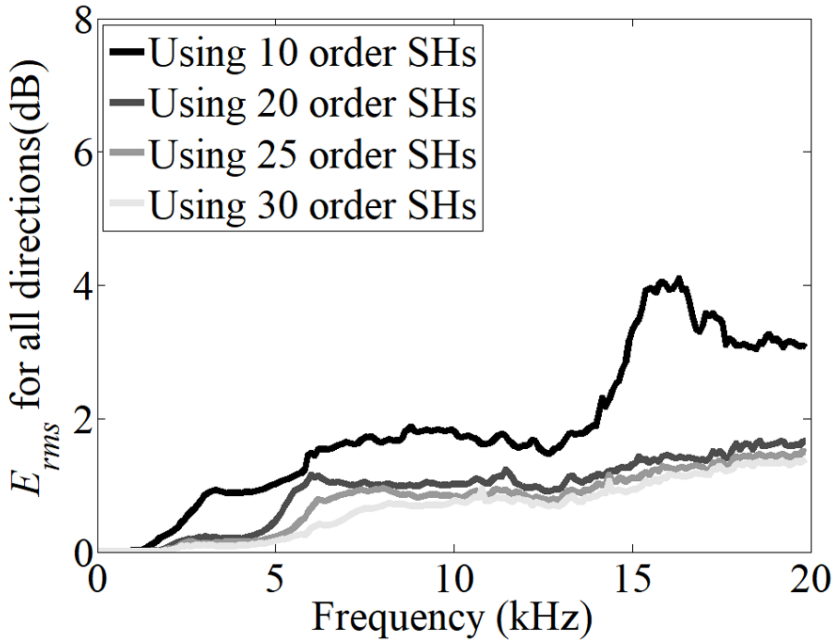


Figure 2.7: Approximation errors based on spherical harmonics at multiple frequencies.

The approximated HRTF up to different orders (order 5, 10, 20 ,30) are shown from Fig. (2.3) to Fig. (2.6). As can be observed in these figures, approximation to a higher orders enable to catch finer details of the HRTF. Some sharp changes of the approximated HRTFs do not appear until the order reaches as a rather high order. The approximation errors (average decibel distance across all directions between the approximated HRTF magnitude and target HRTF magnitude) based on spherical harmonics at multiple frequencies are shown in Fig. 2.7.

2.4 Summary

In this chapter, a common method to analyze HRTFs covering all directions is is introduced based on the spherical harmonic decomposition (e.g., [41, 61]). Rather than looking at the data direction-by-direction, this approach characterizes the HRTFs using a set of expansion coefficients. The spherical harmonic functions take significant values over all directions on spheres, which are called global functions. Therefore, each coefficient includes information from all directions at a particular spatial frequency. This is suited as

a global representation of the target data; however, it requires knowledge of the HRTFs for a sampling covering all directions. In addition, perceptual studies suggest that the minimum audible angle to characterize human sound localization depends on the source's direction ([56]). The directional resolution of HRTF required in binaural synthesis varies for all directions on the sphere around the head ([48]). Since the conventional methods based on spherical harmonics are difficult for local analysis, methods are needed to analyze the HRTFs at different resolutions for different directions.

Chapter 3

Locally modeling head-related transfer functions with continuous spherical wavelets

3.1 Overview

The spherical harmonic decomposition has been a conventional method for modeling the HRTF spatial pattern. However, there still exists some limitations with this method considering the modeling efficiency and the human spatial hearing characteristics. This chapter proposes to use continuous wavelets on the sphere for representing the HRTF spatial patterns.

3.2 Modeling HRTFs using continuous spherical wavelets

In this section, a series of continuous spherical wavelets or local analysis functions constructed by a projection of wavelets in Cartesian plane are proposed, which take signif-

icant values over a small region of directions on the sphere rather than the global functions like spherical harmonics. These local functions still obtained by scaling and translating an initial wavelet function on the sphere basing on the classic wavelet theory.

3.2.1 Continuous spherical wavelet functions on the sphere

Wavelet transform has various applications in audio and image processing [62,63]. The morlet wavelet [64] is a common wavelet which has a simple composition: a oscillation function and a Gaussian window. A basic morlet wavelet has the following description:

$$w(x) = \cos(\alpha x) e^{-\frac{x^2}{2\sigma^2}}, \tag{3.1}$$

where $w(x)$ is a morlet wavelet. α controls the oscillation rate of the cosine function. σ determines the width of the Gaussian window. The classic wavelets are defined in 1D and Cartesian coordinate as the samplings in these cases can be absolutely equal. While for the case on the sphere, as there is no perfect equally distributions of points on the sphere, there are no perfect ways for constructing wavelets on the sphere. Here we try to project the wavelets from the Cartesian plane to the functions on the sphere, so as to construct wavelet functions on the sphere. The projection method is simple which just replace the 'x' with a spherical distance in Eq. (3.1). The proposed local function is thus defined as:

$$W_0(\theta, \phi) = \cos\left[\alpha \cdot D_0(\theta, \phi)\right] \cdot e^{-\frac{D_0^2(\theta,\phi)}{2\sigma^2}}, \tag{3.2}$$

where $W_0(\theta, \phi)$ is a spherical wavelet functions center at (θ_0, ϕ_0). $D_0(\theta, \phi)$ denotes the spherical angle between (θ, ϕ) and (θ_0, ϕ_0) for a unit sphere, and parameters α and σ control the oscillation rate and the width of Gaussian window, respectively. Since multiple wavelets which cover different spatial frequencies are required, a scaling factor S is introduced here. Therefore, a series of scaled wavelet function are generated:

$$W_{0,S}(\theta, \phi) = \sqrt{S} \cos\left[\alpha \cdot S \cdot D_0(\theta, \phi)\right] \cdot e^{-\frac{S^2 \cdot D_0^2(\theta,\phi)}{2\sigma^2}}. \tag{3.3}$$

With the above generation method, continuous spherical wavelets of different spatial frequencies are expected to be constructed with the corresponding different scaling factor S. After defining those local functions on the sphere as in Eq. (3.3) . A target HRTF dataset on the sphere can be approximated as a weighted sum of those wavelets functions, which can be described as follows:

$$H(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{i \in \mathbb{D}(\ell)} c_{\ell,i} \cdot W_{i,\ell}(\theta, \phi), \tag{3.4}$$

Here, $W_{i,\ell}(\theta, \phi)$ is the local function whose scale is ℓ and the center position is (θ_i, ϕ_i); $c_{\ell,i}$ is the corresponding expansion coefficient in the decomposition. The samplings of points corresponding to a certain scale ℓ is denoted as $\mathbb{D}(\ell)$. Due to the reason that it is very difficult to construct local functions with perfect discrete frame. The expansion coefficients are obtained with a least square sense.

3.3 Local represention of HRTF using the propsed local functions

The main difference of the proposed local functions from the conventional spherical harmonics is that the former takes significant value in a local region while the latter consider all directions at one time. Therefore, the proposed wavelet functions may have a good ability to describe the HRTF local features distributed in different local regions.To conduct a local representation, only those wavelets important to a target local region is selected for the purpose. Since the low-scale local functions are constructed with big spatial windows, they have wide influence in space, while those high-scale those have more compact support on the sphere. To implement it, all local functions with the center positions inside the target local region are selected in the reconstruction of the target HRTFs. Besides, the local functions around the local region, however with its amplitude decreased to 10 % of the peak value are selected for the local representation as well.

3.4 Numerical evaluation

3.4.1 Parameters and conditions for the evaluation

This analyzed dataset is the same as the one in Sec. (2.3), a calculated dataset for every two degrees in azimuth and elevation angles at frequency between 93.75 Hz to 20 kHz. The parameters in Eq. (3.3) are empirically set to $\alpha = 0.667$ and $\sigma = 1$. There are many different ways to define the sets $\mathbb{D}(\ell)$ ([65]) of the manifolds on the sphere. Here, an initial sampling is decided as vertexes of an icosahedron [66] whose 12 points are uniformly distributed on the sphere. To derive the directions of samplings for the higher scales, an midpoint is introduced in the projected curve of each edge of the previous scale. To define the scaling factor in the discrete form, wavelets functions on the sphere are generated in a dyadic step as below:

$$S_\ell = 2^{\ell-1}. \tag{3.5}$$

The spherical wavelets constructed using this method and the corresponding located positions thus can be shown in the following figures.

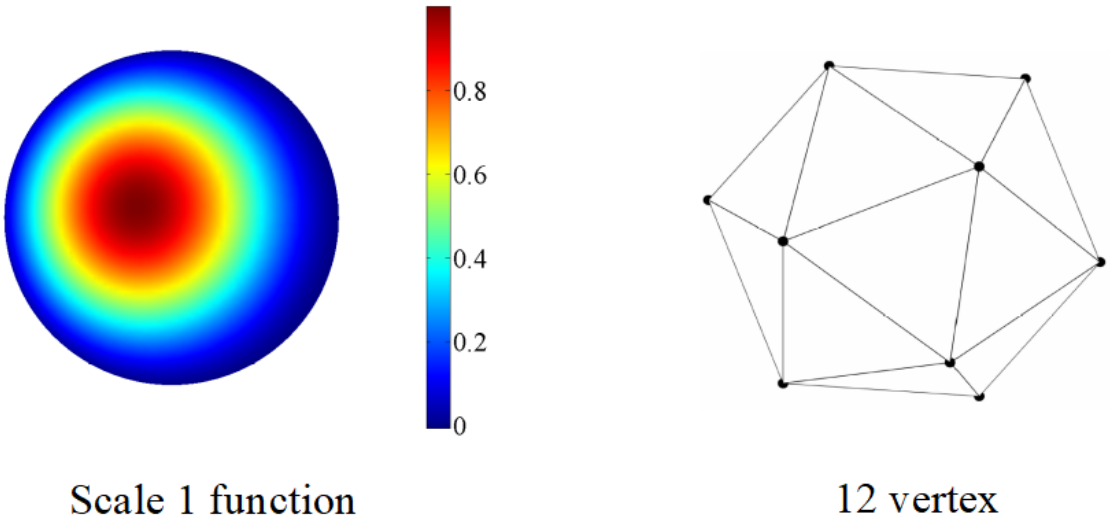


Figure 3.1: Contructed wavelet of scale 1 using the proposed method in chapter4 and its located positions on the sphere.

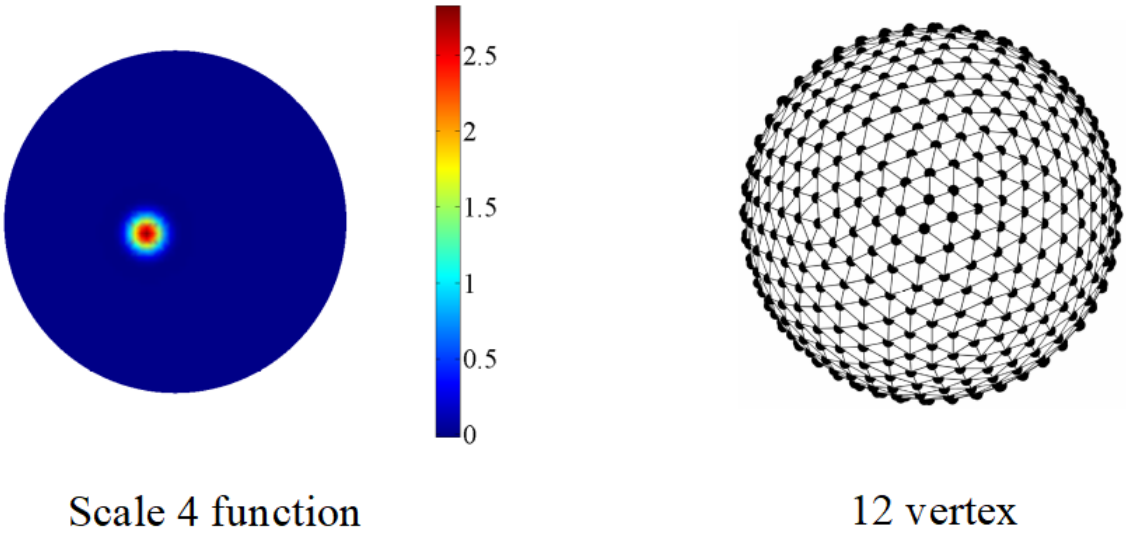


Figure 3.2: Contructed wavelet of scale 3 using the proposed method in chapter4 and its located positions on the sphere.

3.4.2 Simulation results

3.4.3 Objective measurement

The goal of this study is to approximate the original HRTF magnitudes using a set of analysis functions. To measure the approximation error in the spatial domain, the mean normalized error is used, which is defined as

$$E_{mnl} = \frac{1}{N} \sum_{m=1}^{N} \frac{|H_{synth}(\theta_m, \phi_m) - H_{target}(\theta_m, \phi_m)|}{|H_{target}(\theta_m, \phi_m)|}, \tag{3.6}$$

where $H_{synth}(\theta_m, \phi_m)$ and $H_{target}(\theta_m, \phi_m)$ are the reconstructed HRTF magnitude and target HRTF magnitude at direction (θ_m, ϕ_m), respectively; and N is the total number of HRTF samples under study.

Representing for all directions

For modeling the HRTF spatial patterns, as shown in Fig. (3.3), an example of the HRTF magnitudes covering all directions in space at a single frequency 7.4 kHz is given. Then, the target HRTF magnitudes on the sphere are represented based on the modeling proposal introduced in Section 3.2.1 up to a certain scale. The approximated HRTF magnitudes are shown in Fig. (3.4), Fig. (3.5) and Fig. (3.6) which respectively shows the modeling results based on the proposed method up to scale of 3, 4 and 5. The corresponding E_{mnl} is 0.09, 0.04 and 0.01, respectively. This result shows that the approximation up to a higher scale covers more spatial frequencies and therefore yields a smaller approximation error.

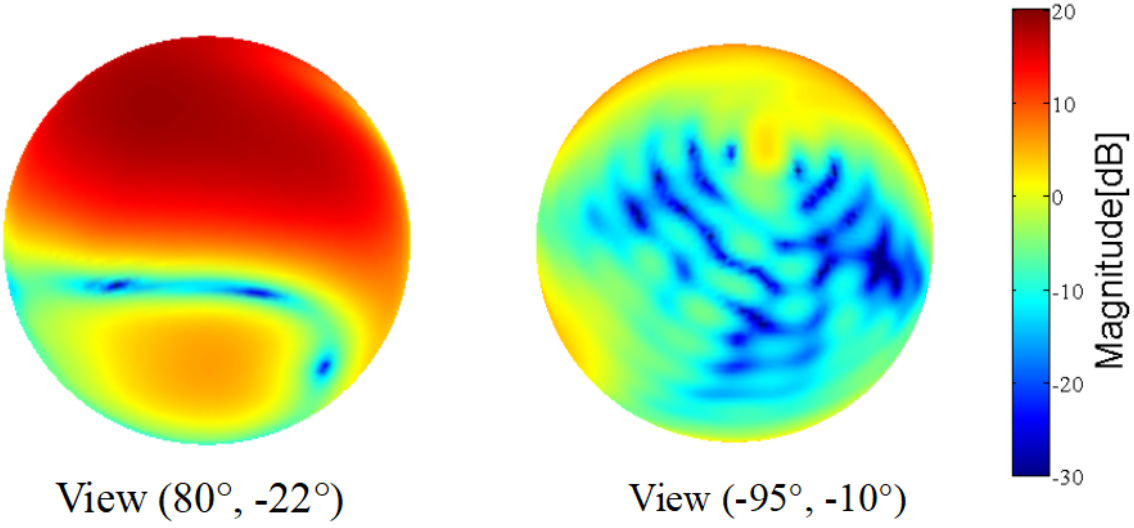
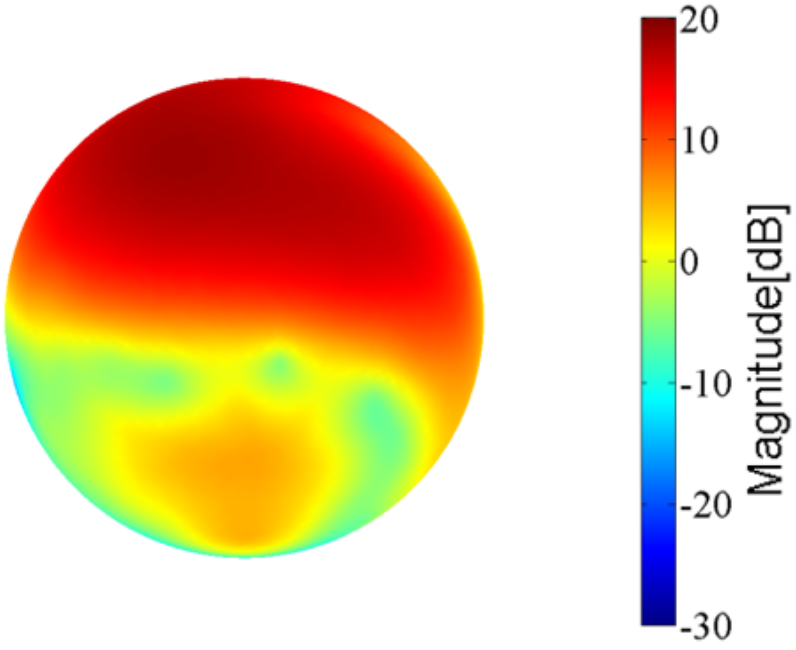
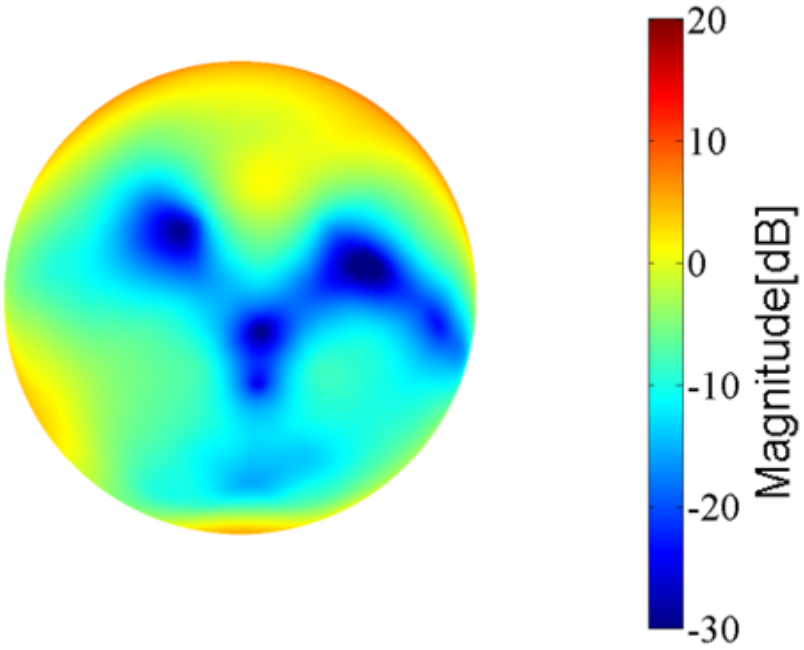


Figure 3.3: Target HRTF.

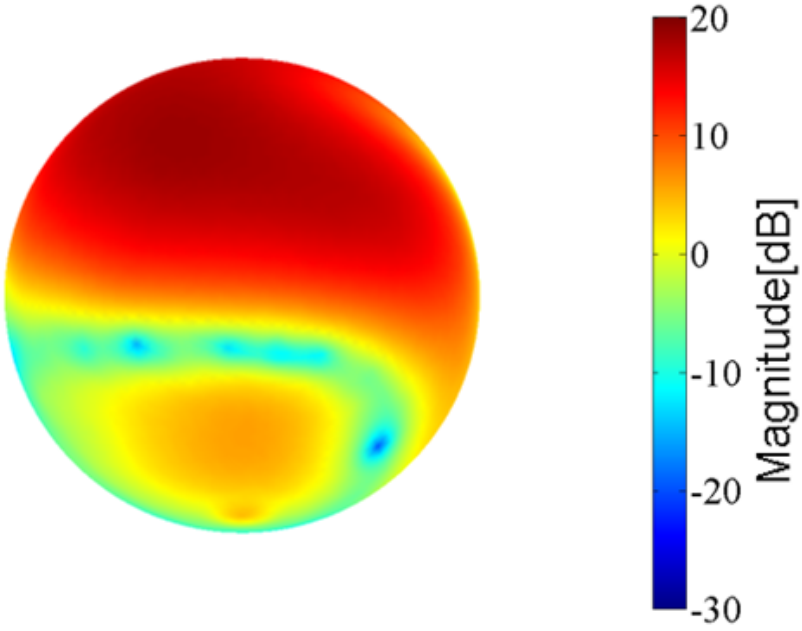


(a) View(170°,−22°)

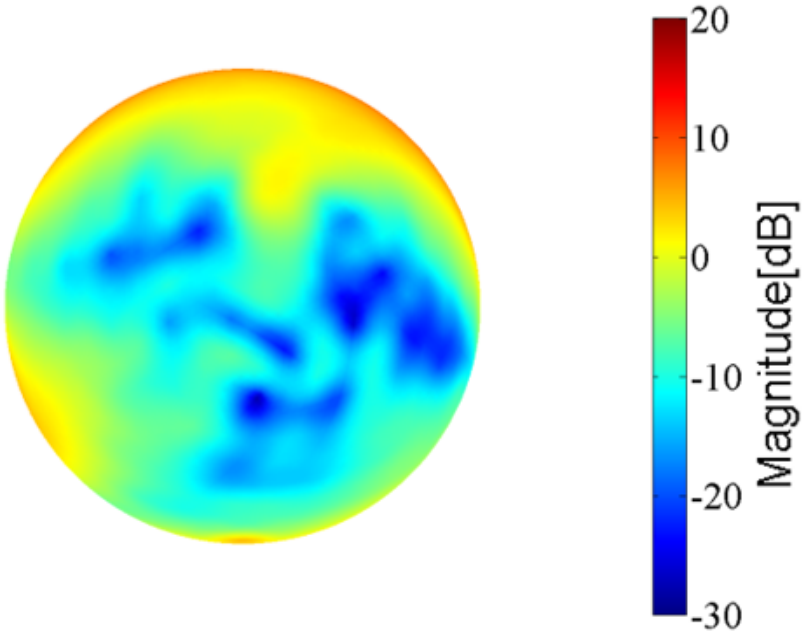


(b) View(−5°,−10°)

Figure 3.4: Reconstructed HRTF using local functions up to scale 3. E_{mnl}=0.09.

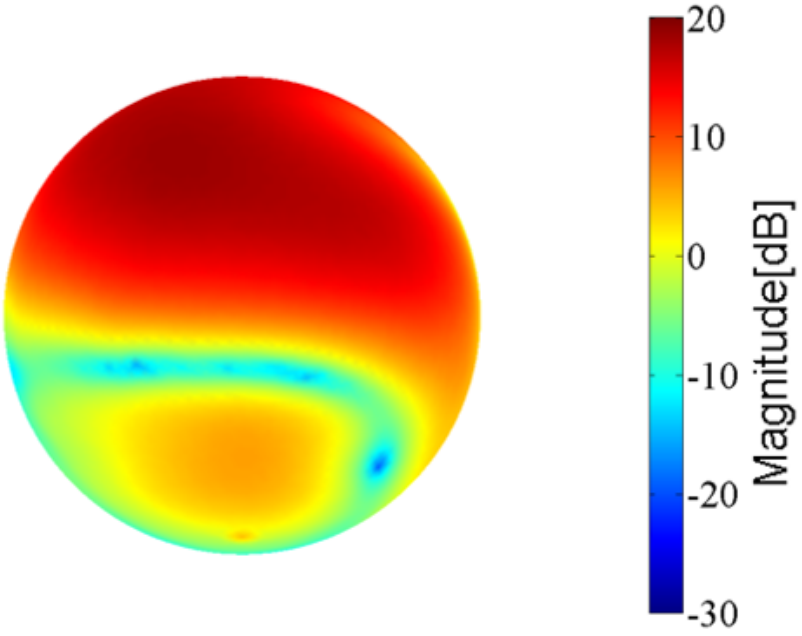


(a) View(170°,−22°)

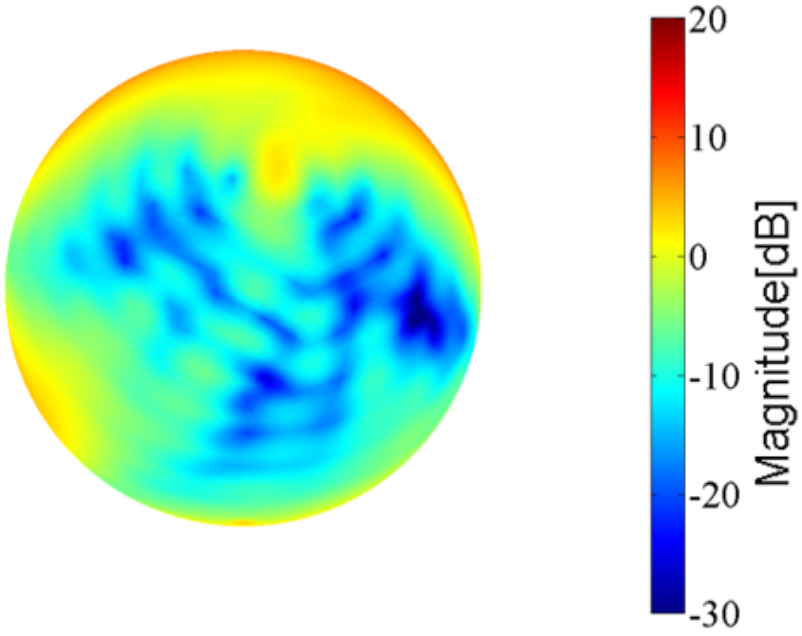


(b) View(−5°,−10°)

Figure 3.5: Reconstructed HRTF using local functions up to scale 4. E_{mnl}=0.04.



(a) View(170°,−22°)



(b) View(−5°,−10°)

Figure 3.6: Reconstructed HRTF using local functions up to scale 5. E_{mnl}=0.01.

Representing for a local region of directions

Since the main difference between the proposed wavelets on the sphere and spherical harmonics is that the former decomposes the target HRTFs with analysis functions with

compact support on the sphere. The proposed method introduced in this chapter is expected to better capture HRTF local features, in such a way that the target HRTF magnitudes over a local region may be more efficiently represented. Same as the method introduced previously in section 3.3, those wavelets that have a great impact on the target local region are selected in the approximation.

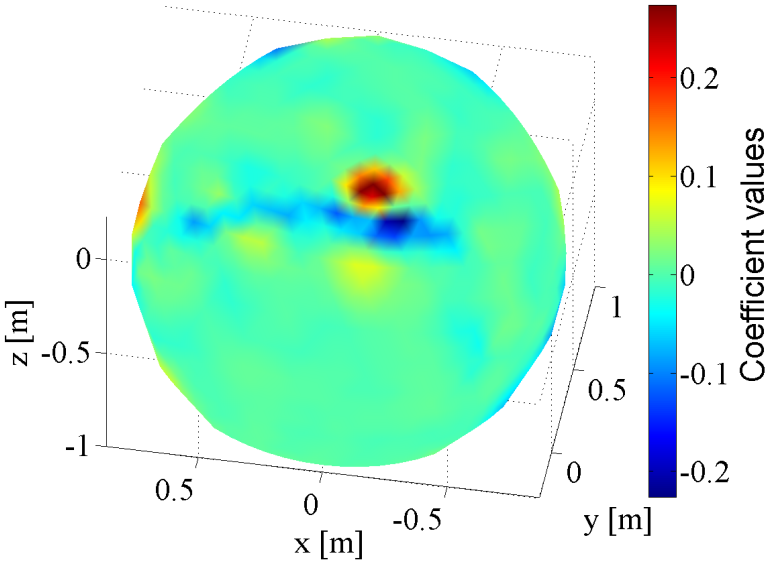
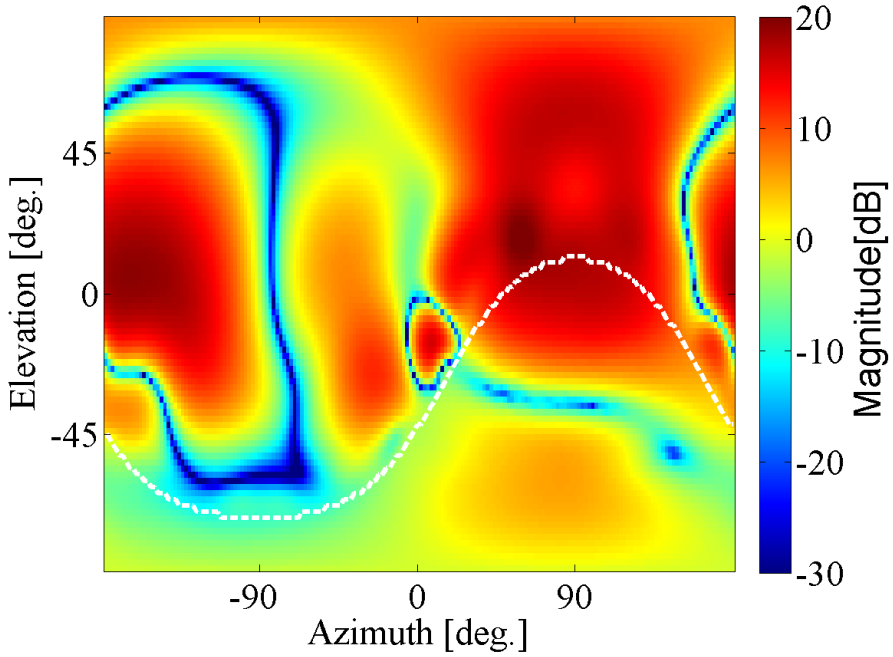
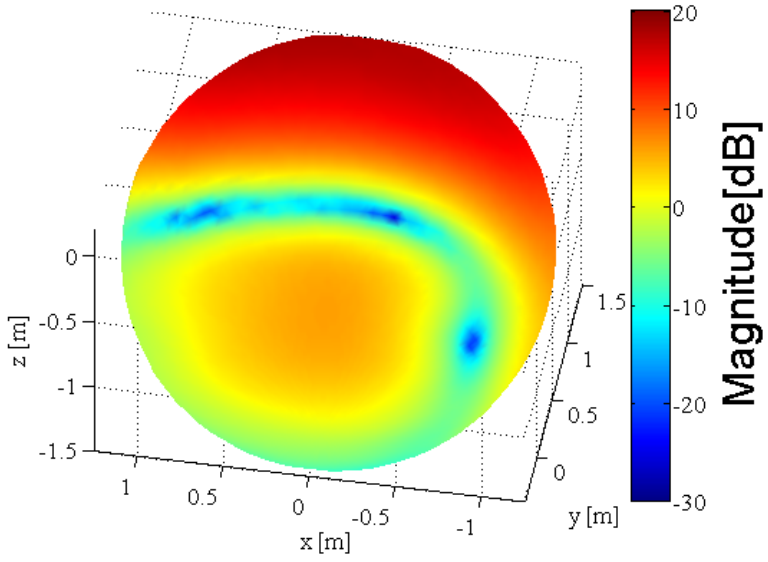


Figure 3.7: Values for the coefficients of scale 5.

Fig. 3.8 givens an example of this local representation, in which the target HRTFs of a local region center at (90°, −48°) on the sphere are approximated up to scale 5. The radius of this local region one the sphere is chosen to be a spherical distance of 1.5 m with a corresponding size of 2.888 steradians. Fig. (3.7) shows the expansion coefficients of scale 5. It can be observed that there seems to be a correspondence between the narrow deep notch of the HRTF and the scale 5 coefficients. This indicates that the expansion coefficients may have a potential to describe the local features of HRTF. The approximation error of this regional representation results in terms of E_{mnl} of 0.006.



(a) The local region below the white doted line.



(b) The local region on the sphere.

Figure 3.8: The local region to be reconstructed is center at (90°, −48°) with a size of 2.888 steradians on the sphere. Panel(a) present this region unfolded to euclidean plane; panel(b) present the same region on the sphere.

A comparison is made between the E_{mnl} values for the local representation of HRTF at 7.4 kHz over the local region introduced above using the proposed modeling method based on the continuous spherical wavelets and the conventional method based on spherical

harmonics (e.g., [41]). When modeling with a comparable number of analysis functions, the proposed continuous spherical wavelets up to scale 4 with 275 wavelet functions and scale 5 with is 937 wavelet functions yield E_{mnl} values of 0.022 and 0.006 respectively, which has smaller values of E_{mnl} values than that of modeling using spherical harmonics up to order 16 with 256 spherical harmonics and order 30 with 961 spherical harmonics, with E_{mnl} of 0.054 and 0.018, respectively.

3.4.4 Approximation error for multiple frequencies

Here, the proposed method is implemented for modeling HRTF magnitudes of multiple frequencies, linearly distributed between 93.75 to 20,000 Hz. In order to compare the local modeling performance of the proposed method and the conventional method [41] based on spherical harmonics, the simulations now are extended for the case of multiple frequencies. Although the conventional modeling method based on spherical harmonics is suitable for global representation of HRTFs, only the local region center at (90°, −48°) is evaluated by the two methods. Besides, these two modeling methods are set to have a comparable numbers of the parameters in the comparison. In detail, the proposed local modeling method up to scales 4 with 275 wavelets and scale 5 with 937 wavelets is compared with the modeling based on spherical harmonics up to, respectively, order 16 with 289 spherical harmonics and order 30 with 961 spherical harmonics. The comparison results are shown in Fig. (3.9) and Fig. (3.10) which suggests that the proposed method using the local functions yields smaller Erms values than that using the spherical harmonics when the number of used analysis functions are comparable in the approximation.

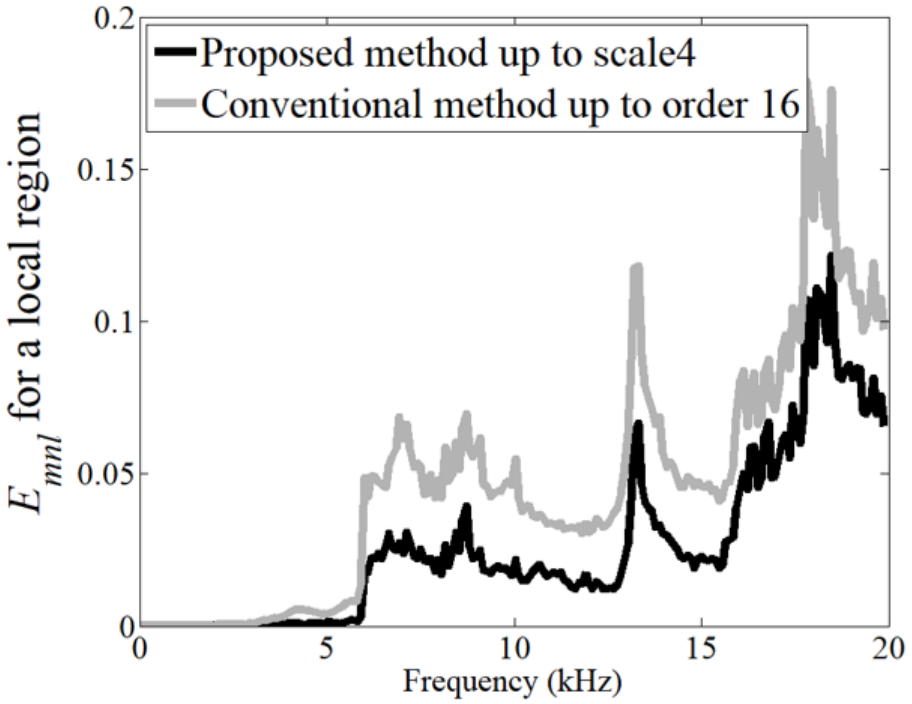


Figure 3.9: Comparison with spherical harmonics up to order 16: Erms with the proposed method up to scale 4 (number of local functions is 275) and that with the conventional method up to order 16 (number of spherical harmonics is 289)

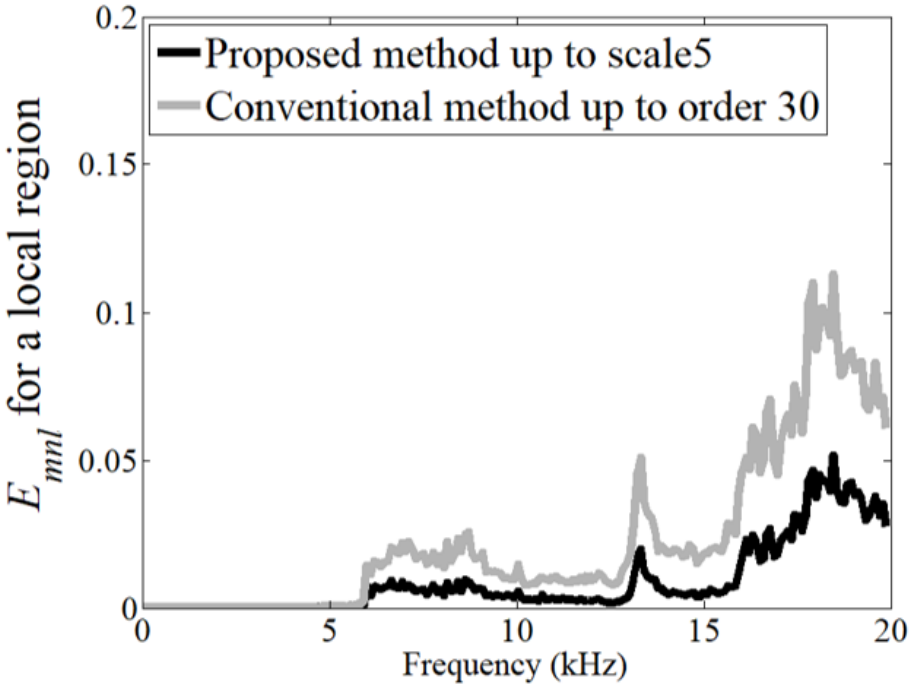


Figure 3.10: Comparison with spherical harmonics up to order 30: Erms values with the proposed method up to scale 5 (number of local functions is 937) and the conventional method up to order 30 (number of spherical harmonics is 961).

3.5 Summary

In this chapter, to locally model the HRTF spatial patterns, a set of continuous spherical wavelets is proposed for the purpose. Different from the spherical harmonics that take significant values in all directions, the proposed wavelets on the sphere have compact support on the sphere with different scaling factors and center positions. Numerical simulation shows that approximation up to a higher scale has corresponding smaller approximation error. Furthermore, the proposed method allows for local representation of the HRTF with appropriately selecting the wavelet functions that intersect the target local region. Therefore, it can achieve the direction dependent spatial resolution needed for accurate sound localization in a certain target region. Furthermore, in the decomposition of HRTF of a local region at 7.4 kHz, there seems to be a correspondence between the values for high-scale coefficients and the spatial details of the HRTFs. It may indicate that the expansion coefficients have a potential to describe the HRTF local features. The correspondence may be improved by defining analysis functions with lower redundancy. Finally, given a comparable number of analysis functions, the proposed method yields smaller values of approximation errors to represent HRTF spatial pattern of a local region than a conventional spherical harmonics method when using a a comparable number of analysis functions.The main content of this chapter is also summarized in a published paper by the author (copyright at the Journal of the Acoustical Society of America, https://asa.scitation.org/doi/10.1121/1.4962805). [67]

Chapter 4

Discrete modeling of head-related transfer functions with spherical wavelet transform

4.1 Introduction

The continuous spherical wavelet proposed in 3 drives the expansion coefficients from the sampled HRTF dataset through the least square optimization. However, the target HRTF dataset is always discrete along the direction and there is a mismatch between the target data and the analysis function. For this reason, it is difficult to control the spatial resolution of the reconstructed HRTF through the expansion coefficients if the analysis functions lack orthogonality. To tackle this problem, this chapter proposes to model the HRTF spatial patterns based on discrete spherical wavelet transform with better characteristics among the analysis functions.

4.2 Review of spherical wavelet transform

For the purpose of modeling HRTF in a certain local region on the sphere, P. Bates et al. recently proposed to use Slepian functions to decompose the target HRTFs inside a certain local region [68]. The Slepian functions are orthogonal analysis functions corresponding to the target local region, which means the construction of the Slepian functions depends on the target region on the sphere. This local modeling method makes sense if HRTFs only inside that region are of interest. However, since the construction of Slepian functions depends on the target local region and do not consider other regions simultaneously. Therefore it is hard to control the spatial resolution of the reconstructed HRTFs locally over the entire sphere. The wavelet analysis on the sphere may be a solution to this problem. According to the classic wavelet theory, a series of wavelets for different scales and positions is generated by dilating and translating a mother wavelet [69–71]. While for the case on the sphere, this way of generating wavelets is not feasible enough. Despite that there exists several methods which construct continuous wavelets or localized functions on a sphere by projecting the wavelets from the euclidean plane to the sphere [72–77], it is still difficult to set a discrete frame for the dilation and translation process on the sphere. For this reason, representing discrete data on a sphere still remains a problem with the continuous spherical wavelet analysis, thus the same for representing discrete HRTF data on the sphere. To generate discrete spherical wavelets for efficiently modeling target functions on a sphere, Schröder and Sweldens proposed the lifting scheme for the purpose [78–80]. A main difference from the classic wavelet analysis which constructs the wavelets through the dilation and translation of a initial wavelet, lifting scheme straightly obtain the expansion coefficients from the predefined discrete manifolds. Although lifting scheme based spherical wavelets have been extensively applied in the field of computer graphics and medical signal processing [81–83], to the best of the authors' knowledge, these spherical wavelets have not yet been applied to the modeling of HRTF. Here in this chapter, the discrete wavelet analysis based on lifting scheme is conducted for modeling the HRTF spatial patterns.

4.2.1 Multiresolution analysis

Although the lifting scheme process does not construct wavelet functions directly, its multi-resolution analysis is very similar to that of the classic wavelet theory. Here, the target function is divided into two parts: the approximation and detail part, as if processed through a low-pass and a high-pass filter, respectively. Fig. 4.1 shows an example, where a high-pass filter ("High") and a low-pass filter ("Low") calculate the wavelet coefficients W_ℓ and scaling coefficients S_ℓ at a certain scale level $\ell = 1, ..., L-1$, which corresponds to the detail and approximation parts, respectively. An example of the multi resolution analysis of function on the sphere using spherical wavelet transform based on lifting scheme is shown in Fig. 4.2.

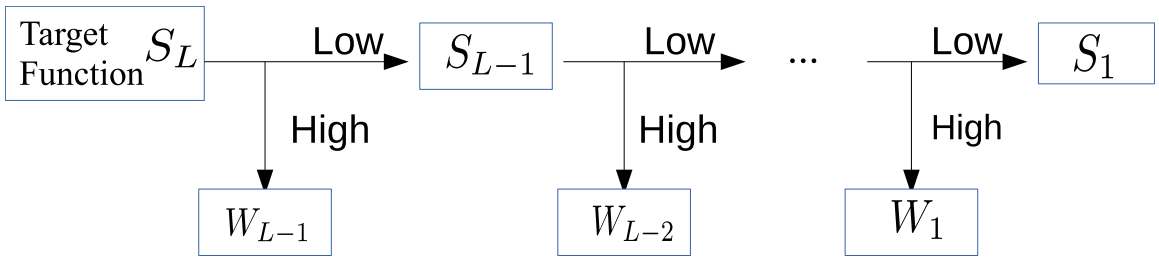


Figure 4.1: Multiresolution analysis based on wavelet analysis.

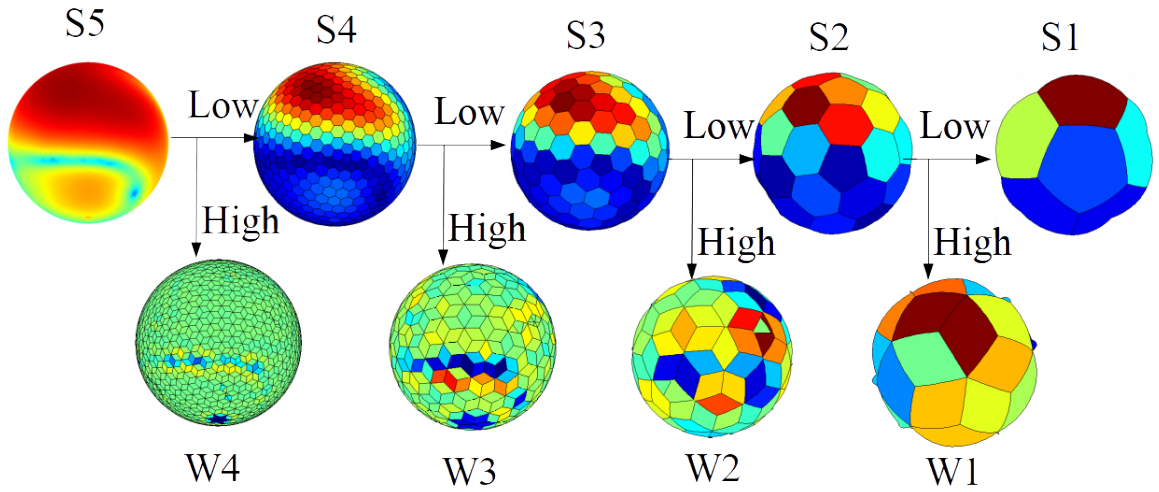


Figure 4.2: Multiresolution analysis of function on the sphere using spherical wavelet transform based on lifting scheme.

4.2.2 Lifting scheme based spherical wavelets

To uniformly define the discrete manifolds on the sphere, the icosahedron and its subdivision samplings with different resolutions are used for the purpose. A vertex set at level ℓ is denoted as $V_\ell = \{\vec{v}_{\ell,i}\}_{i=1,...,I_\ell}$, where $\vec{v}_{\ell,i}$ defines a vertex at level ℓ, and I_ℓ is the total number of vertex at this level. An icosahedron vertex is thus defined as a root level set V_1, which has a total of points $I_1 = 12$. The next level vertex set $V_{\ell+1}$ can be obtained by adding midpoints at every edge of the previous level and then projecting them on the sphere; this midpoint set is denoted as $M_\ell = \left\{\vec{m}_{\ell,j}\right\}_{j=I_\ell+1,...,I_{\ell+1}}$, where $\vec{m}_{\ell,j}$ denotes the midpoint at level ℓ.

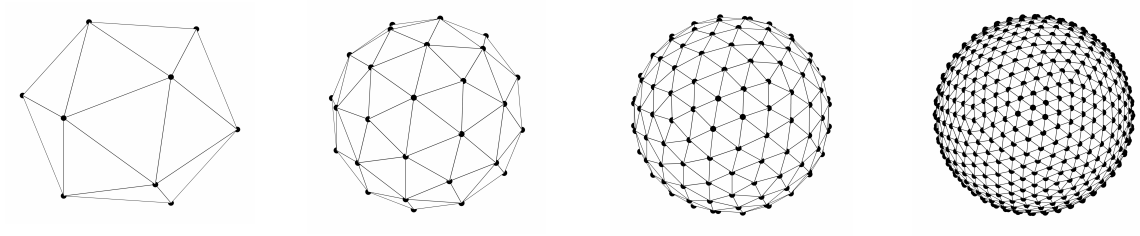


Figure 4.3: Sampling of an icosahedron V_1 and its following subdivisions V_2, V_3, and V_4 from left to right, with the vertex number of 12, 42, 162, and 642, respectively.

A vertex set can be divided into a lower-scale vertex set and the corresponding midpoint set, namely $V_{\ell+1} = V_\ell \cup M_\ell$. For example, Fig. 4.3, shows an icosahedron sampling set V_1, and its subdivision vertex sets V_ℓ ($\ell = 2, 3, 4, ...$).

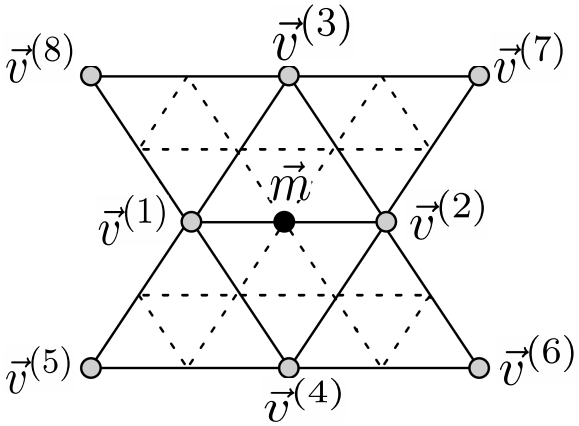


Figure 4.4: Local neighborhoods ($N(\vec{m}) = \vec{v}^{(k)}, k = 1, ..., 8\}$) of a point $\vec{m} \in M_\ell \subset V_{\ell+1}$ using the local naming scheme. The dashed-lines are the edges of next subdivision.

To conduct the spherical wavelet transform, a local naming scheme is applied, where

each $\vec{m}_{\ell,j} \in M_\ell$ is only filtered with points over a small neighborhood. As an example, in Fig. 4.4, a midpoint $\vec{m} \in M_\ell$ is surrounded by the neighborhood set $N(\vec{m})$ which consists of 8 points ($N(\vec{m}) = \vec{v}^{(k)}, k = 1, ..., 8$) [79, 84]. When performing the forward spherical wavelet transform, the vertex level goes from the leaf level to the root level as ℓ decreases. By implementing the steps backwards, the inverse transform is realized. Generally, the process of the forward transform can be shown as below.

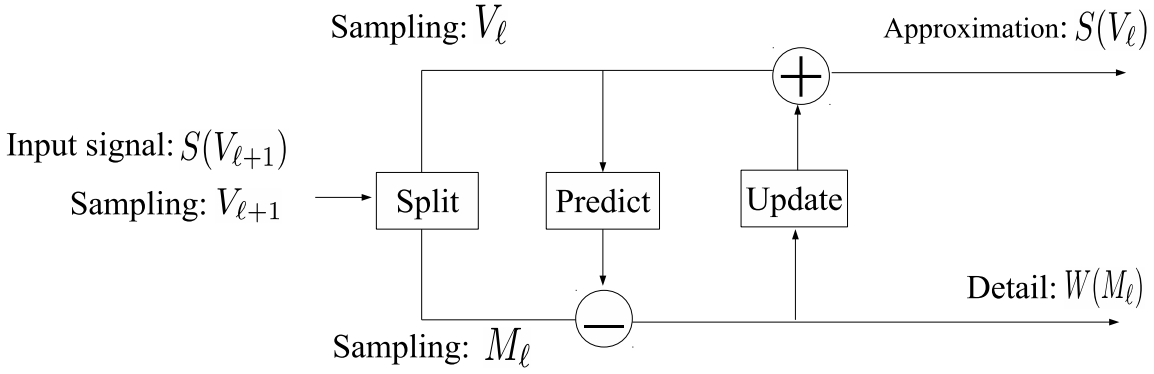


Figure 4.5: Lifting scheme forward wavelet transform.

- Forward step 1:

  $\forall \vec{v}_{\ell,i} \epsilon V_\ell : \quad S(\vec{v}_{\ell,i}) \leftarrow S(\vec{v}_{\ell+1,i})$

  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \quad W(\vec{m}_{\ell,j}) = S(\vec{v}_{\ell+1,j}) - \sum_{\vec{v}_{\ell,i} \epsilon N(\vec{m}_{\ell,j})} z^w_{\ell,i,j} S(\vec{v}_{\ell,i}),$

  where $S(\vec{v}_{\ell,i})$ denotes the scaling coefficient corresponding to a point $\vec{v}_{\ell,i}$ on V_ℓ, and $W(\vec{m}_{\ell,j})$ the wavelet coefficient corresponding to the point $\vec{m}_{\ell,j}$ on M_ℓ. Please notice that the target function H which is set to be the same as the initial scaling coefficients at the finest level L, namely $H(\vec{v}_{L,i}) = S(\vec{v}_{L,i})$; $z^w_{\ell,i,j}$ are the weights for filtering the neighborhood set $N(\vec{m}_{\ell,j})$ around a midpoint $\vec{m}_{\ell,j} \epsilon M_\ell$ in order to predict the value of the midpoint. There are different ways to define the filtering neighborhood and the corresponding weights. Here in this study, the butterfly subdivision scheme [79] is applied for the purpose, which leads to that, $z^w_{\ell,1,j} = z^w_{\ell,2,j} = 1/2$, $z^w_{\ell,3,j} = z^w_{\ell,4,j} = 1/4$, and $z^w_{\ell,5,j} = z^w_{\ell,6,j} = z^w_{\ell,7,j} = z^w_{\ell,8,j} = -1/16$.

- Forward step 2:

With the wavelet coefficients $W(\vec{m}_{\ell,j})$ calculated in step 1, the scaling coefficients $S(\vec{v}_{\ell,i})$ are updated in order to represent the approximation as:

$\forall \vec{m}_{\ell,j} \epsilon M_\ell : \ S(\vec{v}_{\ell,v_1}) = S(\vec{v}_{\ell,v_1}) + z^s_{\ell,1,j} W(\vec{m}_{\ell,j}), \ S(\vec{v}_{\ell,v_2}) = S(\vec{v}_{\ell,v_2}) + z^s_{\ell,2,j} W(\vec{m}_{\ell,j})$.

Here, $\vec{v}_{\ell,v_1}$ and $\vec{v}_{\ell,v_2}$ denote the two endpoints on the parent edge of $\vec{m}_{\ell,j}$. $z^s_{\ell,i,j}$ are weights to satisfy that the resulting wavelet has a vanishing integral: $z^s_{\ell,i,j} = I_{\ell+1,j}/2I_{\ell,i}$, in which $I_{\ell,i}$ is the integral of the scaling function of $\vec{v}_{\ell,i}$.

As shown in Fig. 4.5, when performing the forward transform, the higher level vertex set $V_{\ell+1}$ corresponding to the scaling coefficients $S(V_{\ell+1})$ are split into the lower-level vertex set V_ℓ and the midpoint set M_ℓ, respectively. The coefficients of V_ℓ are used to "Predict" the coefficients of midpoint set M_ℓ and the prediction difference is used as the wavelet coefficients $W(M_\ell)$, which corresponds to the "Forward step 1". In "Forward step 2", the obtained wavelets coefficients $W(M_\ell)$ calculated in step 1 update the coefficients of V_ℓ and the resulted scaling coefficients $S(V_\ell)$ represent the approximation of $S(V_{\ell+1})$. Therefore, the forward step 1 and 2 actually correspond to the high-pass filter and low-pass filter in Fig. 4.1 which calculate the wavelet coefficients and scaling coefficients, respectively.

The inverse transform can be performed by impelementing the above two steps backwards.

- Backward Step 1:

  Calculate $S(\vec{v}_{\ell,i})$:

  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \ S(\vec{v}_{\ell,v_1}) = S(\vec{v}_{\ell,v_1}) - z^s_{\ell,1,j} W(\vec{m}_{\ell,j}), \ S(\vec{v}_{\ell,v_2}) = S(\vec{v}_{\ell,v_2}) - z^s_{\ell,2,j} W(\vec{m}_{\ell,j})$.

- Backward Step 2: $\forall \vec{v}_{\ell,i} \epsilon V_\ell : \ S(\vec{v}_{\ell+1,i}) \leftarrow S(\vec{v}_{\ell,i})$

  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \ S(\vec{v}_{\ell+1,j}) = W(\vec{m}_{\ell,j}) + \sum_{\vec{v}_{\ell,i} \epsilon N(\vec{m}_{\ell,j})} z^w_{\ell,i,j} S(\vec{v}_{\ell,i})$

In the forward transform of the lifting scheme, the scaling coefficients S and wavelet coefficients W of a target function H are calculated. While in the inverse transform, the

target function is synthesized through the calculated expansion coefficients. Here, the forward and inverse transform based on the lifting scheme can be denoted as $\mathbb{W}$ and $\mathbb{W}^{-1}$, respectively. Please notice that the expansion coefficients consist of the scaling and wavelet coefficient S and W, namely, $C = \{S, W\}$ in the spherical wavelet decomposition, and can be obtained through the forward transform as $C = \mathbb{W}\{H\}$. By implementing the inverse spherical wavelet transform, the approximation of the target function can be realized with the decomposed expansion coefficients, namely $H = \mathbb{W}^{-1}\{C\}$.

Different from the classic wavelet theory, the lifting scheme analysis avoids constructing the analysis function directly. However, if the value of coefficient corresponding to a position of interest set as 1 and the rest expansion coefficients set as 0, the generation of the corresponding analysis function can be realized. With this method, the wavelet or scaling function of level ℓ and position $\vec{m}_{\ell,p}$ or $\vec{v}_{\ell,p}$ in the lifting scheme analysis can be generated as below:

$$Y_{\ell,p} = \mathbb{W}^{-1}_{\ell,p}\{C\}. \tag{4.1}$$

Here, $\mathbb{W}^{-1}_{\ell,p}$ denotes the inverse transform $\mathbb{W}^{-1}$ from the root to the leaf level; the expansion coefficient $c_{\ell,p} \in C$ is set as 1, and the rest of the coefficients are set as 0.

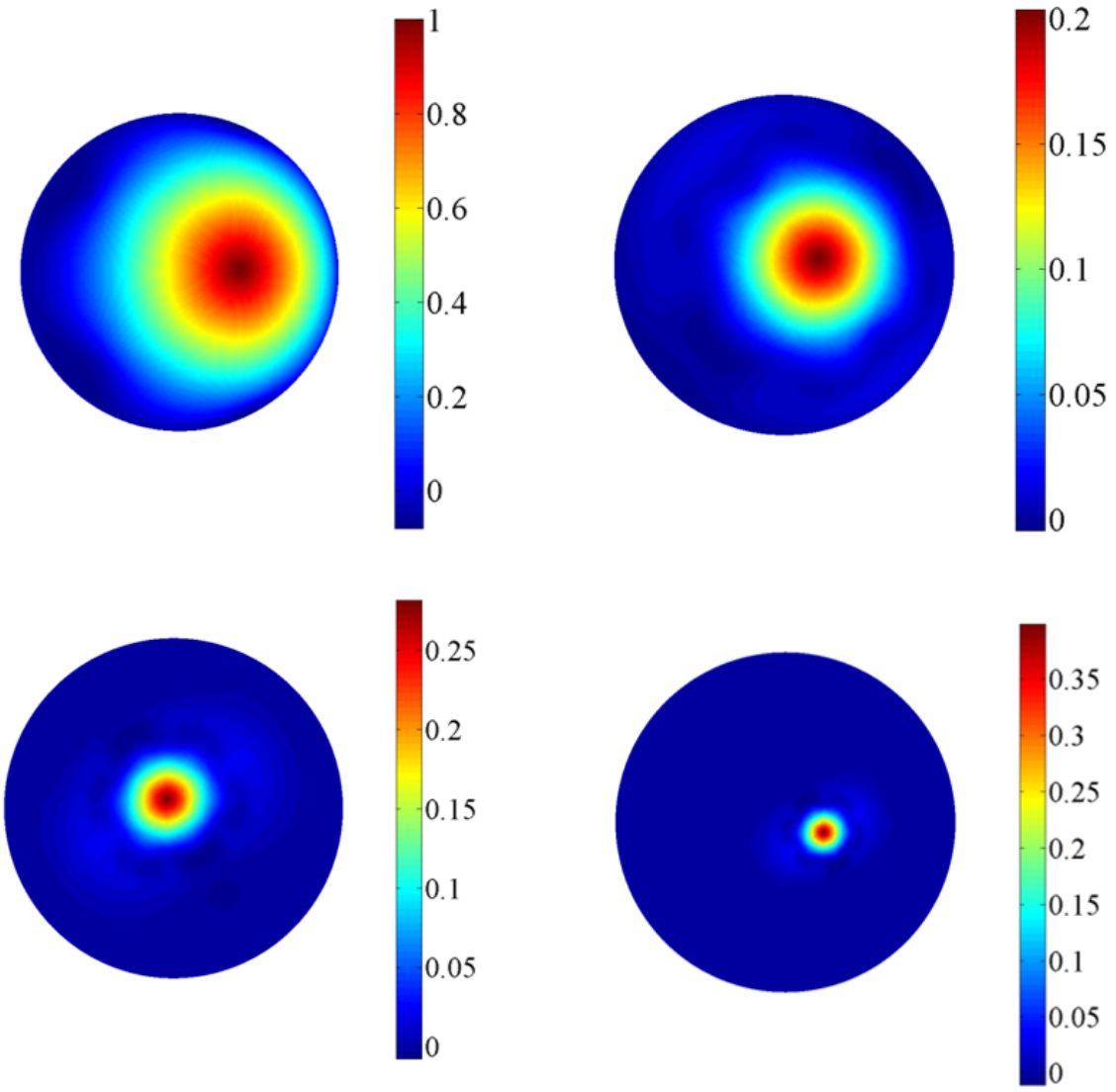


Figure 4.6: Analysis functions in Eq. 4.1 for the spherical wavelet transform (the leaf level is set to scale level 6) based on the lifting scheme with the butterfly subdivision: scaling function of scale level 1 (upper left), and wavelets of scale level 1 (upper right), 2(down left), and 3(down right). The finest resolution corresponds to level 6.

Fig. 4.6 shows the scaling function of scale level $\ell = 1$ and spherical wavelets from scale $\ell = 1$ to 3, which corresponds to the center positions for locating the scaling function and wavelets on the sphere correspond to the mesh vertex in Fig. 4.3. It can be seen that the low-scale spherical wavelets have greater support, while the high-scale wavelets have more compact support, which accounts for the low spatial frequency and high spatial frequency, respectively.

4.3 Modeling HRTF spatial patterns with spherical wavelets

4.3.1 Modeling HRTF magnitudes with spherical wavelets

After defining the analysis function in the previous section, the target HRTF magnitudes $H(\theta, \phi)$ can be decomposed as the function below:

$$H(\theta,\phi) = \sum_{\vec{v}_{1,i}\in V_1} S(\vec{v}_{1,i})\varphi_{1,i}(\theta,\phi) + \sum_{\ell=1}^{L-1}\sum_{\vec{m}_{\ell,j}\in M_\ell} W(\vec{m}_{\ell,j})\cdot\psi_{\ell,j}(\theta,\phi), \tag{4.2}$$

where $\varphi_{1,i}(\theta,\phi)$ and $\psi_{\ell,j}(\theta,\phi)$ denote the scaling function and spherical wavelet corresponding to $\vec{v}_{1,i}$ and $\vec{m}_{\ell,j}$, respectively; $S(\vec{v}_{1,i})$ and $W(\vec{m}_{\ell,j})$ denote the corresponding coefficients in the approximation. Please notice that, in order to implement this method, the sampling positions (θ,ϕ) of the target HRTF magnitudes along the direction should match a certain kind of manifolds on the sphere such as the icosahedron subdivision in the spherical wavelet analysis. The required HRTF target data can be obtained through the interpolation of the HRTF or generation the HRTF magnitudes at the desired directions based on numerical simulation such as the boundary element method (BEM) [23]. An example of a scaling function $\varphi_{1,i}(\theta,\phi)$ of scale level $\ell = 1$ and spherical wavelet $\psi_{\ell,j}(\theta,\phi)$ of scale level $\ell = 1, 2, 3$, and the corresponding meshes for locating these analysis functions are shown in Fig. 4.6 and Fig. 4.3, respectively. For efficient representation of the target HRTFs, ℓ is truncated to a certain scale level. The expansion coefficients are obtained based the decomposition method introduced in Section 4.2.

4.3.2 Modeling interaural level differences with spherical wavelets

It is known that the HRTF spectrum plays an important role in elevation localization, while the interaural level differences contribute to the horizontal sound localization of the higher frequency components and elevation localization as well [13, 85]. Here the lifting

scheme based spherical wavelet transform is applied as well to the HRTF interaural level difference(ILD). Fig. 4.7 shows an example of the ILD at 7.4 kHz.

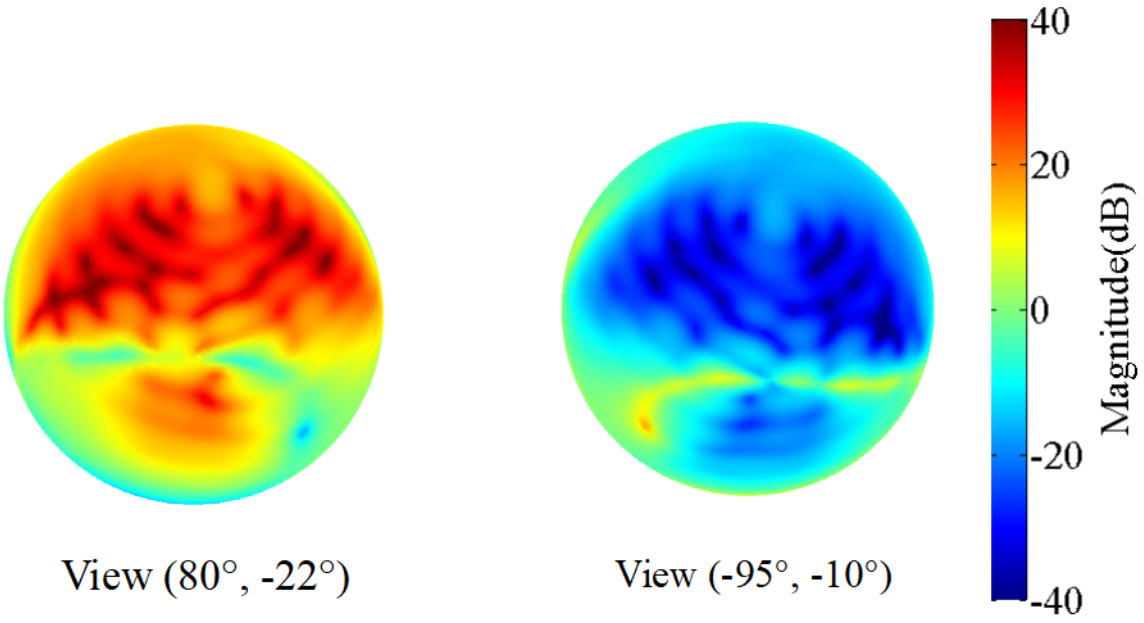


Figure 4.7: ILD of the HRTF in 3.3 at frequency 7.4 kHz.

The ILD is defined as the difference of the level of the left and right HRTF in decible scale and can be described with the following equation:

$$ILD_f(\theta,\phi) = \frac{HL_f(\theta,\phi)}{HR_f(\theta,\phi)}, \tag{4.3}$$

where $HL_f(\theta,\phi)$ and $HR_f(\theta,\phi)$ are respectively the left and right HRTF magitude at frequency f and direction (θ,ϕ). The ILD at a frequency f can be represented based spherical wavelets with the following equation:

$$ILD_f(\theta,\phi) = \sum_{\vec{v}_{1,i}\in V_1} S(\vec{v}_{1,i})\varphi_{1,i}(\theta,\phi) + \sum_{\ell=1}^{L-1}\sum_{\vec{m}_{\ell,j}\in M_\ell} W(\vec{m}_{\ell,j})\cdot\psi_{\ell,j}(\theta,\phi). \tag{4.4}$$

4.4 Numerical evaluation

To evaluate the proposed modeling method in this chapter, numerical simulation is implemented to validate the effectiveness of the method described in Section 4.2.

4.4.1 Objective measurement

To calculate the approximation error in the spatial domain, the same error measurement is used as that in the previous chapters as below. Please notice that this objective criteria may not necessarily reflect the perceptual performance.

$$E_{mnl} = \frac{1}{N}\sum_{m=1}^{N}\frac{|H_{synth}(\theta_m,\phi_m) - H_{target}(\theta_m,\phi_m)|}{|H_{target}(\theta_m,\phi_m)|}. \tag{4.5}$$

Here $H_{synth}(\theta_m,\phi_m)$ and $H_{target}(\theta_m,\phi_m)$ are the approximated HRTF magnitude and the target HRTF magnitude at direction (θ_m,ϕ_m), respectively. N is the total number of HRTF samplings of the target dataset.

4.4.2 HRTF database

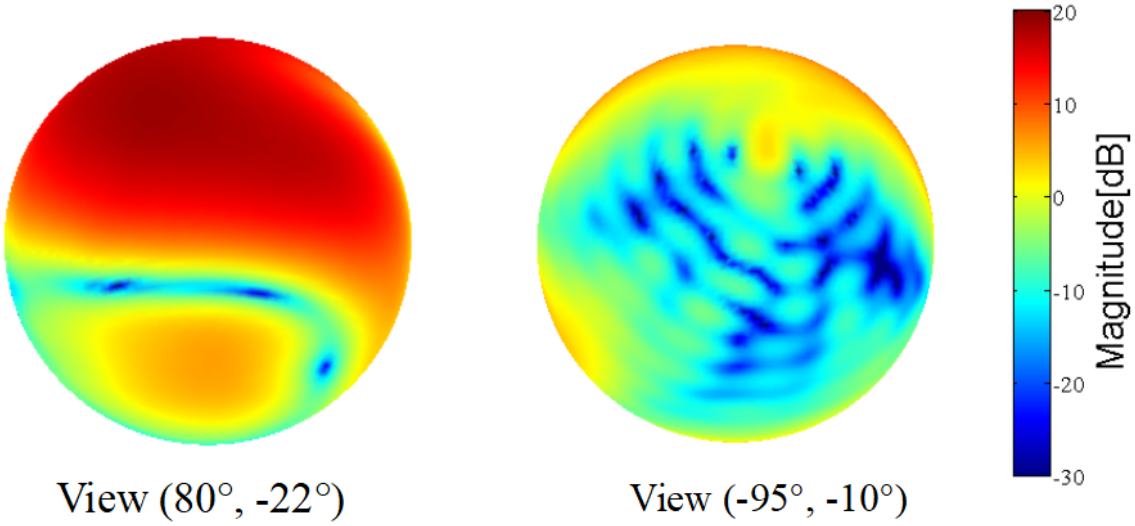


Figure 4.8: Original HRTFs (left ear) at 7.4 kHz. The panels from left to right are the HRTF with views of (80°, −22°) and (−95°, −10°), respectively.

HRTF datasets are always obtained through measuring the impulse responses at predefined grid positions or calculating based on simulation methods at desired positions. The target HRTF applied here is the same as that used in the chapter 3 which covers all directions at multiple frequencies. The number of the HRTF samplings along the direction is 10242 at each frequency bins. This number of samplings is high enough to recover HRTFs

at all directions within the audible frequency range [61], and allows for visualizing the fine details of HRTF spatial patterns. The target HRTF at 7.4 kHz for the left ear is shown in Fig. 4.8.

4.4.3 HRTF representation using the proposed method

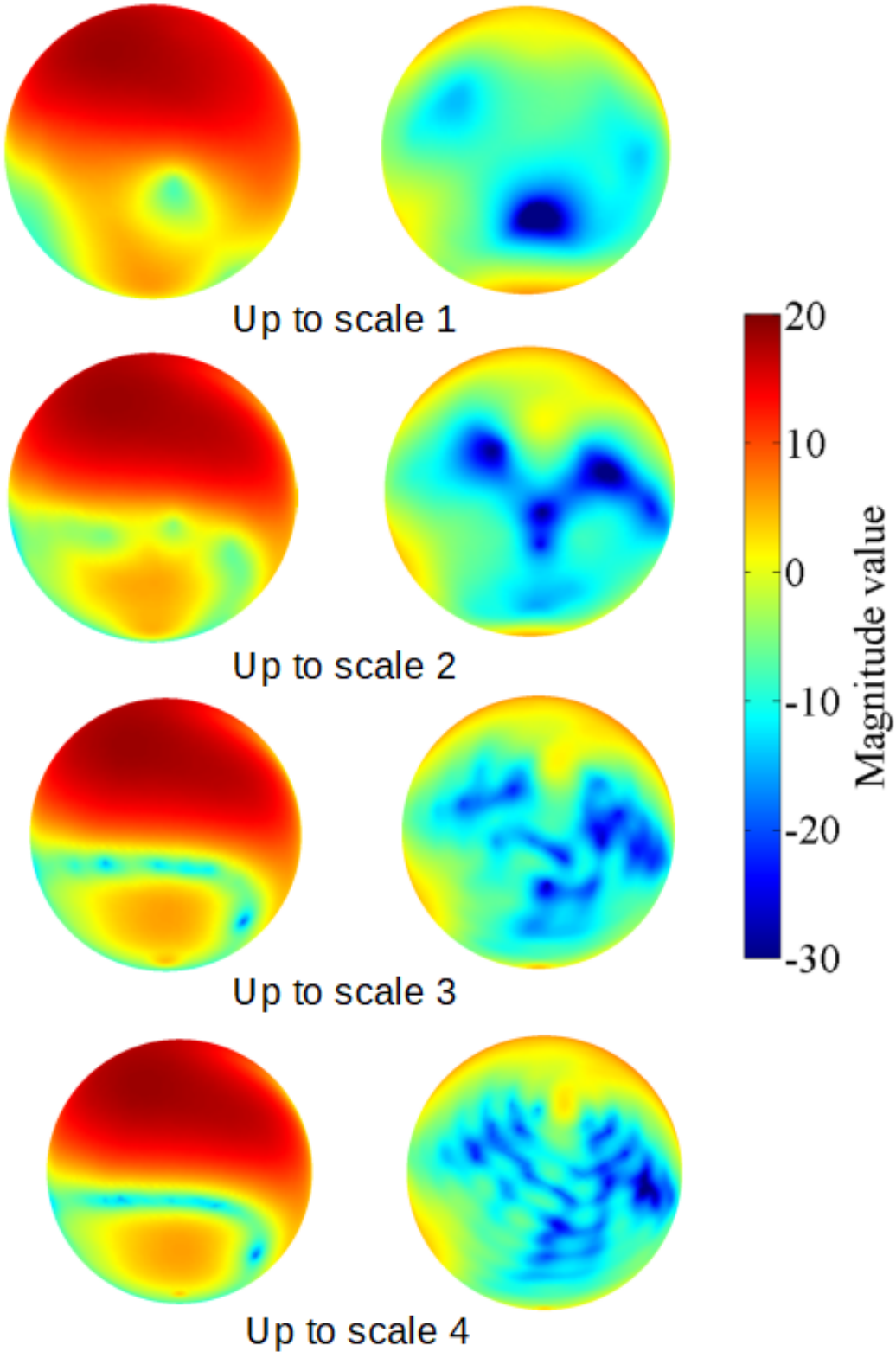


Figure 4.9: Approximated HRTF (left ear) in Eq. 5.4 with spherical wavelets up to different scales.

Based on the lifting scheme based spherical wavelet analysis, the target HRTF magnitudes whose sampling points match the icosahedron subdivision of scale level $\ell = 6$ (10242 points in total) are decomposed to a root level $\ell = 1$. The spatial variations of the HRTF magnitudes can be smoothed by truncating ℓ to a certain level in the reconstruction. Fig. 4.9 shows that the original HRTF with a total sampling number of 10242 is approximated up to scale 1 (number of coefficients = 42), 2 (number of coefficients = 162), 3 (number of coefficients = 642), and 4 (number of coefficients = 2562), which yield E_{nml} = 0.260, 0.105, 0.048 and 0.009 respectively. The expansion coefficients from scale level 1 to 5 are shown in Fig. 4.10. The target HRTF magnitude is equal to the scaling function at scale level 6. This result suggests that the higher scale spherical wavelets catch finer details, while the low scale spherical wavelets approximate of the coarse structure of the HRTF. Thus, the original target HRTF with a high sampling amount can be represented by the expansion coefficients. A perfect reconstruction of the target HRTF can be realized by implementing the backward transform using all 5 scale's expansion coefficients.

One of the motivations of this study is to visualize and describe the HRTF local features using the expansion coefficients. In the previous work of the authors' group, the coefficients of a set of local analysis function could describe the local features to some degree [67].

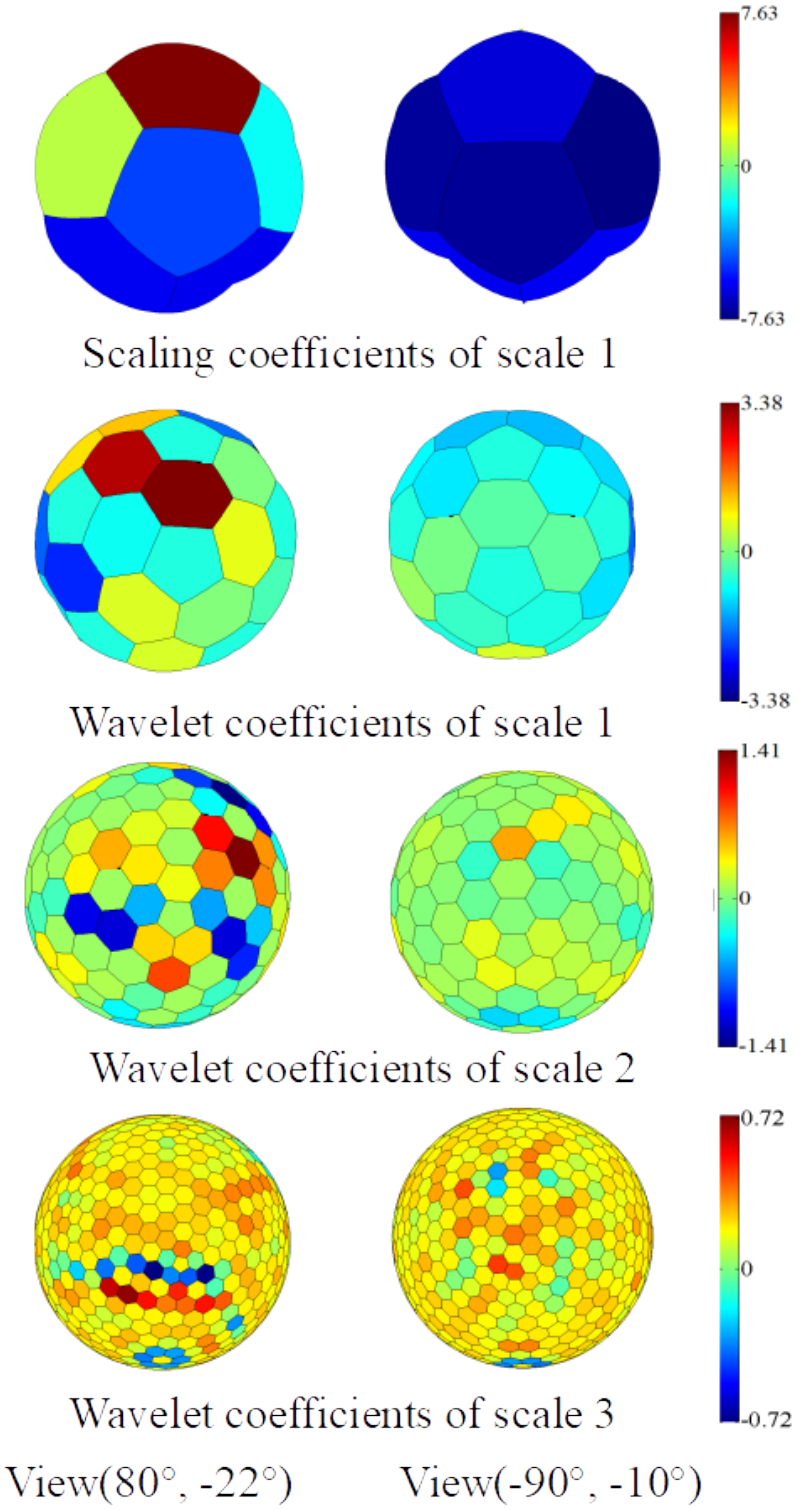


Figure 4.10: Expansion coefficient values of spherical wavelets at each scale in the decomposition of the target HRTF at 7.4kHz (Fig. 4.8) from the scale level $\ell = 1$ to the scale level $\ell = 3$.

By taking look at the expansion coefficients in Fig. 4.10, some interesting assumptions

may be made. The lowest scale (scale 1) coefficients may describe the spatial filtering effects excluding the pinna effects; the significant differences between the coefficient values of the ipsilateral and contralateral sides clearly reveal the head shadow effect. For the expansion coefficients of scale 5 , most of their values are nearly 0. It suggests that the spatial spectrum at this scale plays little role in the composition of the target HRTF. The coefficient values of scale 4 are also close to 0 except at some small local regions that corresponds to distributions of the sharp spatial variations of the original HRTFs. For the coefficients of scales higher than 1, most of their values are small and the significant values are only taken at some local regions. Since the pinna effects are more likely involved in the fast-changing spatial details [86, 87], these higher scale coefficients seem to have a closer relevance to the pinna effects, which highly depend on the direction. The distribution of the expansion coefficients at different scales not only classify the acoustic effects of different anatomical parts, but also exhibits in which direction those filtering effects take place. This is a significant advantage over spherical harmonic decomposition. Therefore, the proposed method provides a tool for visualization and insights of the spatial filtering effects inside the HRTF data. Furthermore, the distribution of the expansion coefficients should also have the individual information across different subject, this may also open a means for HRTF individualization.

The ILD in Fig. fig:ILD is also appximated with simulation, and the result is shown in Fig. 4.11 and Fig. 4.12.

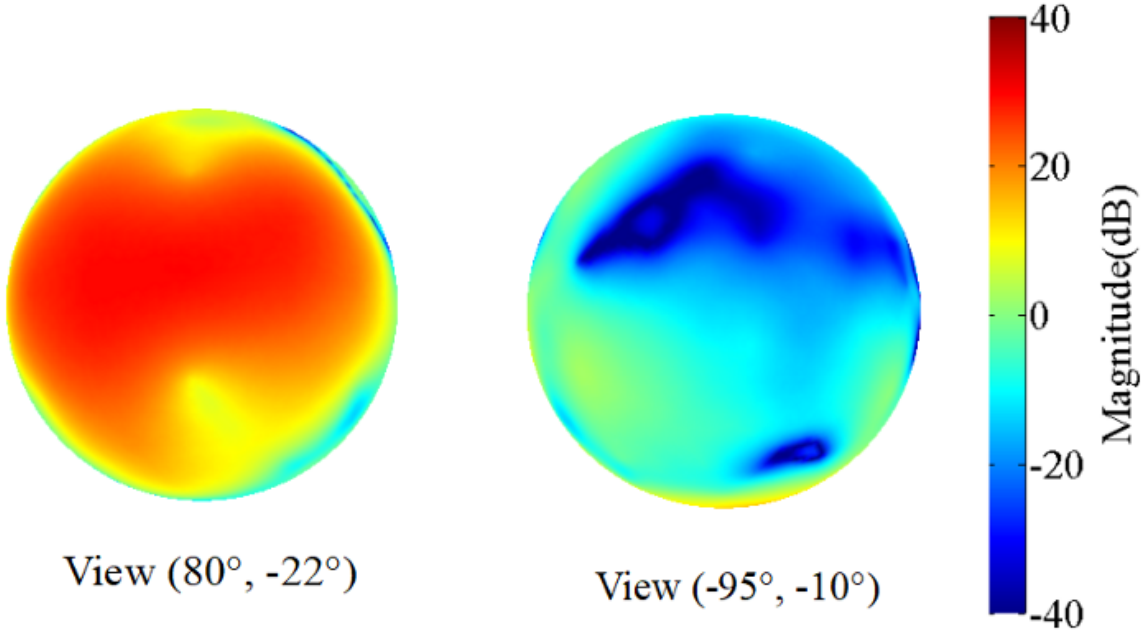


Figure 4.11: The reconstructed ILD at 7.4 kHz with modeling up to level 1.

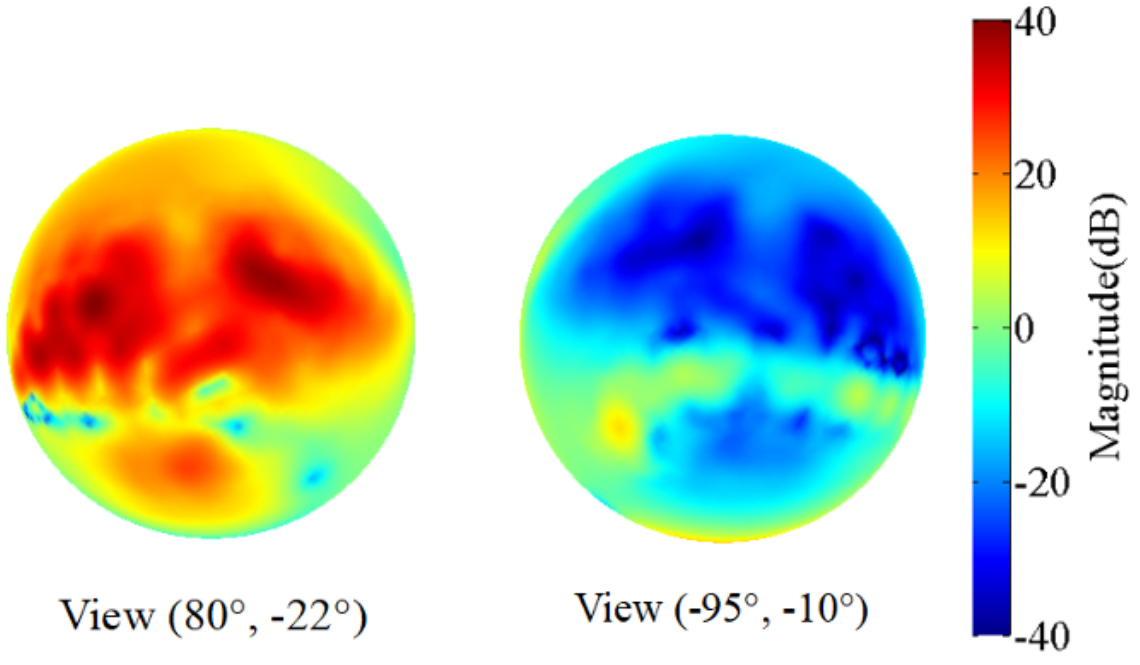


Figure 4.12: The reconstructed ILD at 7.4 kHz with modeling up to level 3.

4.4.4 Efficient representation of the HRTF using the proposal

Fig. 4.10 shows that the expansion coefficients are close to 0 in most directions at the higher levels. This indicates that the rapidly changing spatial variations mainly occur in some small and local regions. An assumption can be made that those low–value coefficients play little role in the reconstruction and could be removed to more efficiently represent the HRTF. To perform the compression, the analysis functions of each scale are first normalized to the same energy. The expansion coefficients with the absolute values above a certain threshold after the normalization are preserved to reconstruct the HRTF. In this manner, the HRTF can be represented by a small number of coefficients.

The following part of this section evaluates this efficient representation of the HRTF in all directions and compares it with the spherical harmonic method when using same number of analysis functions. The comparison is further extended to evaluations of HRTFs in some local regions on the sphere.

Comparison with spherical harmonics in all directions

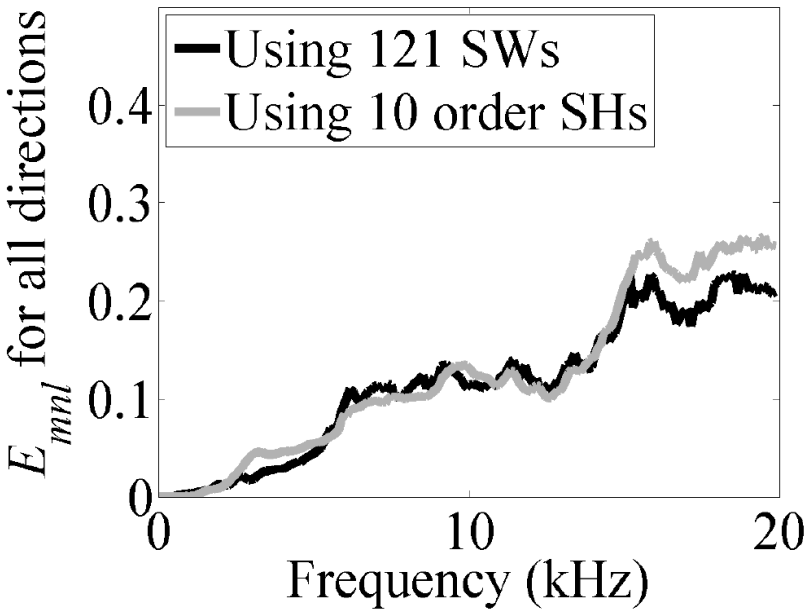


Figure 4.13: Comparisons of E_{mnl} for the target HRTF in all directions betweeen the approximation using 121 spherical wavelets (SWs) and spherical harmonics (SHs) up to order 10.

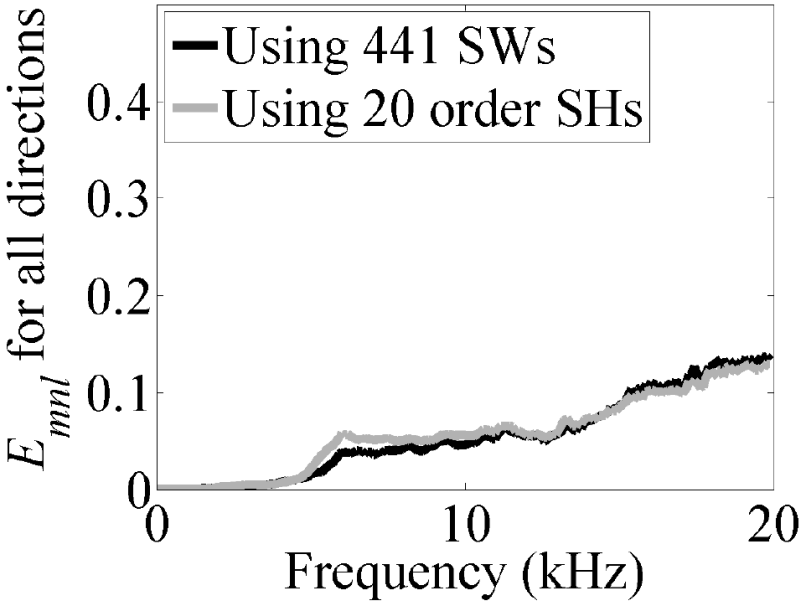


Figure 4.14: Comparisons of E_{mnl} for the target HRTF in all directions betweeen the approximation using 441 spherical wavelets (SWs) and spherical harmonics (SHs) up to order 20.

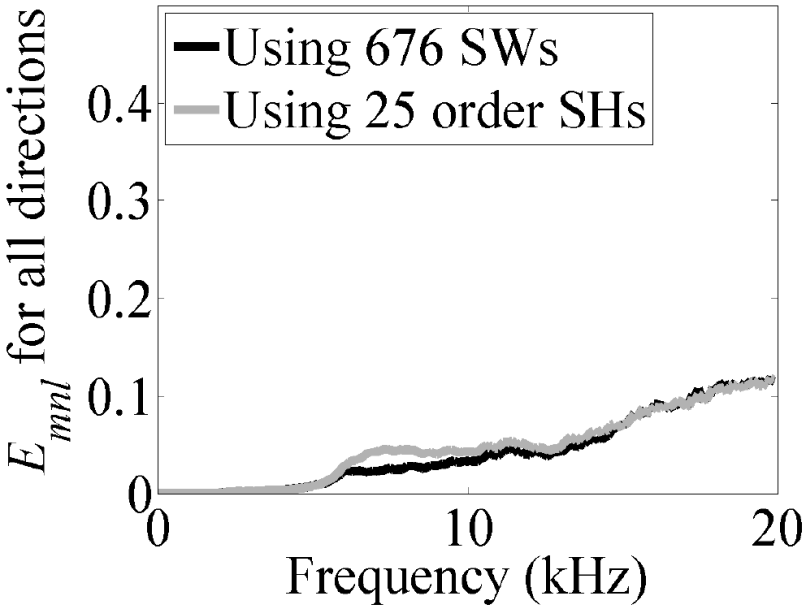


Figure 4.15: Comparisons of E_{mnl} for the target HRTF in all directions betweeen the approximation using 676 spherical wavelets (SWs) and spherical harmonics (SHs) up to order 25.

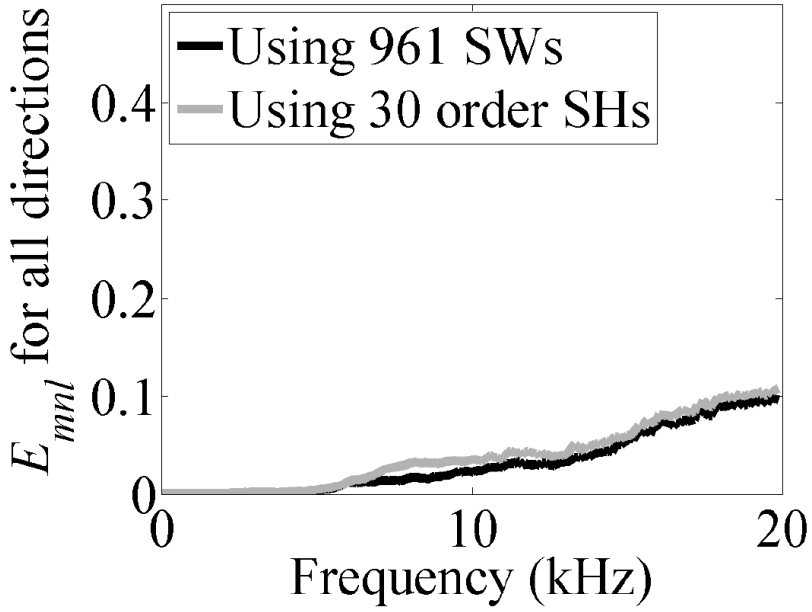


Figure 4.16: Comparisons of E_{mnl} for the target HRTF in all directions betweeen the approximation using 961 spherical wavelets (SWs) and spherical harmonics (SHs) up to order 30.

The spatial representation of the HRTF in all directions is conducted with the same target HRTF as above. To compare the performance with spherical harmonics, the approximation errors are evaluated using the same number of analysis functions between the two methods. Namely, the approximation errors using spherical wavelets (SWs) with the number of coefficients of 121, 441, 676, 961 for reconstruction are compared to the errors when using spherical harmonics (SHs) up to order 10, 20, 25, and 30, respectively, which represent the HRTF with different spatial accuracy. Please note that in this study, we use real spherical harmonic representations with icosahedron samplings, which are well-conditioned for the spherical harmonic decomposition, because the orthonormality error decreases [88]. Fig. 4.13, Fig. 4.14, Fig. 4.15, and Fig. 4.16 plots the comparison results. It shows that our proposed method is slightly better or at least comparable with the conventional method based on real-valued spherical harmonics in terms of the approximation error when representing the HRTF magnitudes in all directions with a same number of analysis functions in the above conditions.

Comparisons with spherical harmonics at local regions

An expected advantage of the proposed method over the spherical harmonic method is to better represent the HRTF local features. Next, comparisons with the spherical harmonic method are conducted with the HRTF represented at some local regions. To perform this local representation, only the significant–value spherical wavelets corresponding to the target local region are selected for the reconstruction; this selection is done in two steps. The first selecting step depends on the local area under evaluation. Because the low–scale spherical wavelets contribute to the coarse structure and have an influence over a larger area on the sphere, all analysis functions of scale 1 are preserved in the reconstruction. For the higher–scale spherical wavelets, only those close to the area under evaluation are used. More specifically, the spherical wavelets whose radius of influence intersects the area under evaluation are selected for this local representation. The radius of influence of a spherical wavelet is defined as the distance between its center position (where it has its maximum value) and the position where its amplitude decreases to the minimum value. Second, among the selected wavelets in the first step, only the spherical wavelets with significant expansion coefficient values are preserved.

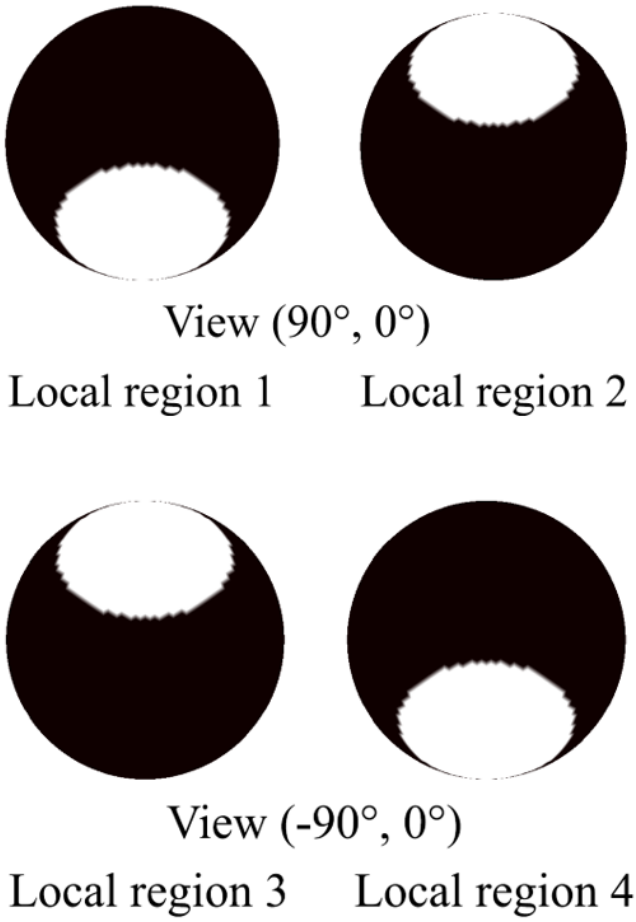
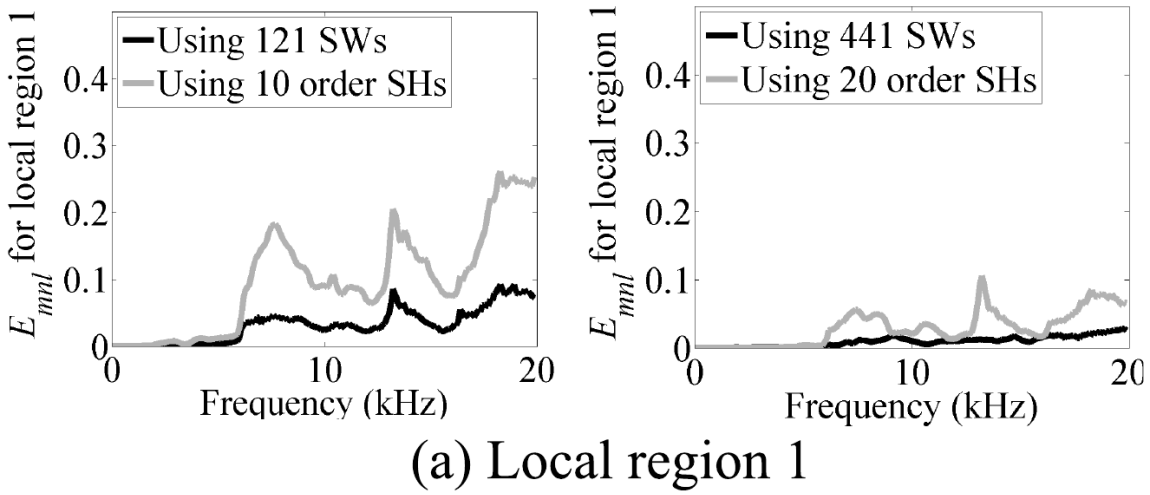


Figure 4.17: Four local regions (in white) selected for the evaluation of the local representation.

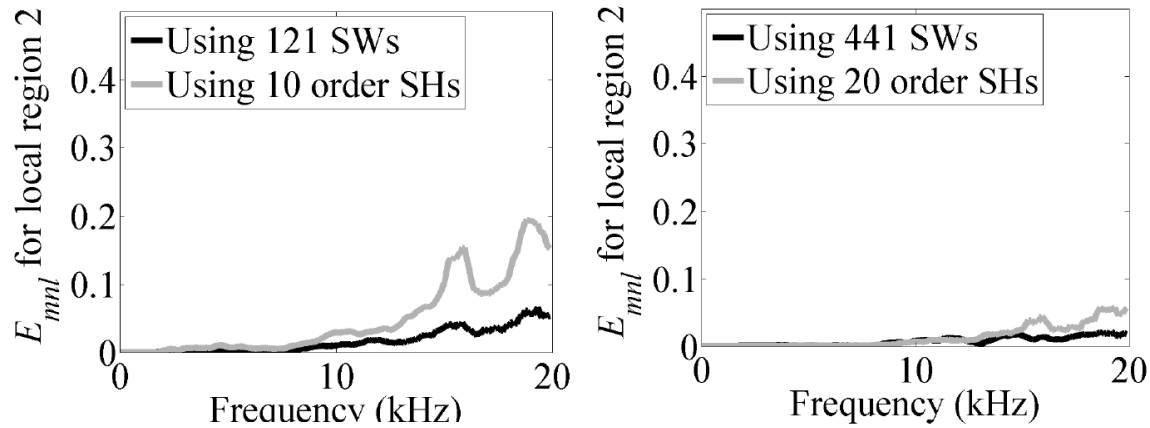
The target HRTFs at the audible frequencies are reconstructed inside four spherical caps that are centered at (90°,−48°), (90°,48°), (−90°,48°), and (−90°,−48°), respectively, all with a size of 1.40 steradians, and denoted as local region 1, 2, 3, and 4, respectively. These four regions together cover most directions in space, as shown in Fig. 4.17. For these four local regions, the E_{mnl} values are compared between the approximation using spherical wavelets with 121 and 441 highest coefficients for reconstruction, and real-valued spherical harmonics up order 10 (number of harmonics=121) and order 20 (number of harmonics=441), respectively. The results shown in Fig. 4.18, Fig. 4.19, Fig. **??**, and Fig. 4.21, suggest that when using these same number of analysis functions, our proposed method yields smaller approximation errors than for the spherical harmonic method. Therefore, the efficiency of modeling HRTF local features using spherical wavelets is val-

idated.



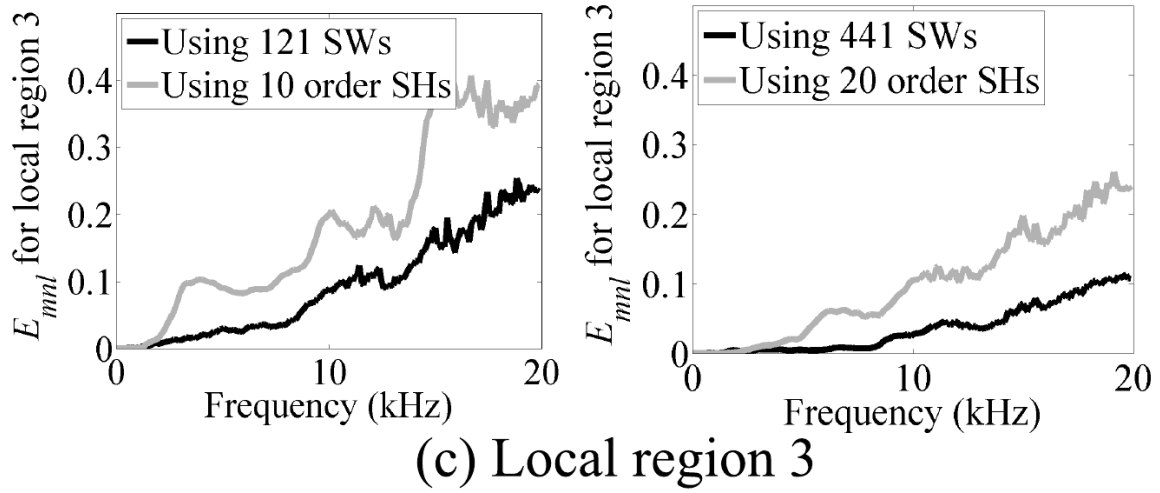
(a) Local region 1

Figure 4.18: Comparisons of E_{mnl} for representing target HRTFs insider local region 1 between the approximation using the same number of spherical wavelets (SWs) and spherical harmonics (SHs).



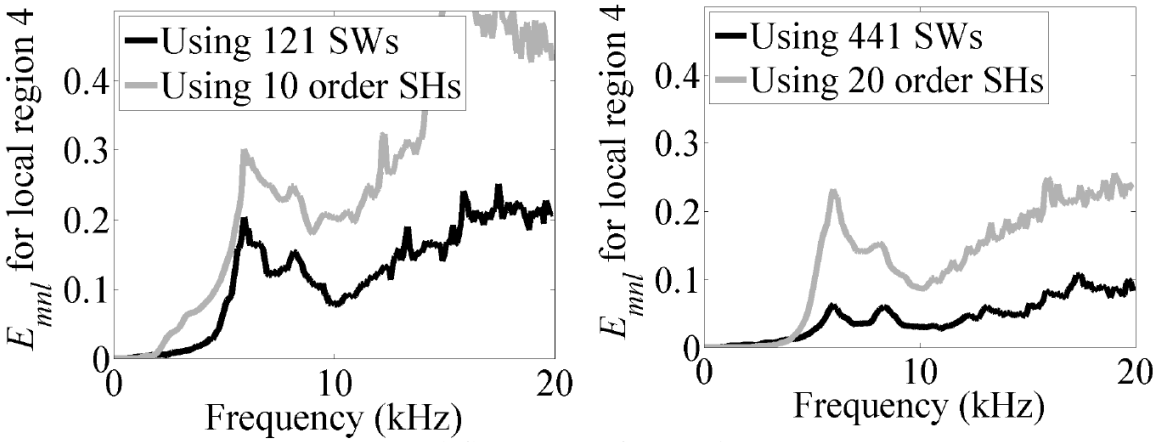
(b) Local region 2

Figure 4.19: Comparisons of E_{mnl} for representing target HRTFs insider local region 2 between the approximation using the same number of spherical wavelets (SWs) and spherical harmonics (SHs).



(c) Local region 3

Figure 4.20: Comparisons of E_{mnl} for representing target HRTFs insider local region 3 between the approximation using the same number of spherical wavelets (SWs) and spherical harmonics (SHs).



(d) Local region 4

Figure 4.21: Comparisons of E_{mnl} for representing target HRTFs insider local region 4 between the approximation using the same number of spherical wavelets (SWs) and spherical harmonics (SHs).

4.5 Summary

In this chapter, the HRTF spatial variations are represented using the lifting scheme based spherical wavelets, which are widely used in the field of computer graphics. Numerical experiments showed that when using the same number (121 and 441) of analysis functions, approximation of the HRTF at the evaluated local regions based on spherical wavelets yields smaller errors than for the spherical harmonic method. In addition, the expansion coefficients of the spherical wavelets could well correspond to the direction dependent HRTF local features. This provides a tool for visualizing and analyzing the acoustic filtering effects inside the target HRTF data. Future work will consider evaluations of several individual HRTFs to study the individual information that benefits from the proposed method. The main content of this chapter has also been presented in a published paper by the author of this thesis (copyright at Applied Acoustics, https://www.sciencedirect.com/science/article/pii/S0003682X18301397). [89]

Chapter 5

Modeling head-related impulse response with spherical wavelets

5.1 Overview

In the previous chapters, the modeling of HRTF is focused on its magnitude in the spatial domain. However, in order to render the binaural spatial audio, the time-domain head-related impulse response (HRIR) or complexed-valued HRTF with both magnitude and phase are required. Therefore this chapter studies the decomposition of the time domain HRIRs and the representation based on the spherical wavelet.

5.2 HRTF model with minimum-phase reconstruction

The complex-valued HRTF contains both magnitudes and phase. The binaural HRTF magnitudes are conventionally used to explain the human elevation localization and the horizontal localization for the higher frequency components based on the interaural level differences. While the HRTF phase, especially the interaural phase cues are believed to

contribute to the horizontal localization for the lower frequency (below 1500 Hz) components of sound sources. The interaural phase cues can be further explained with the interaural time differences (ITDs) which should be frequency dependent. That is, the actual values of ITD vary along the frequency.

Various investigation has been proposed to integrate activity across frequency channels and turn the multiple, one or each frequency channel, ITD estimates into a single perceptual ITD cue [90–92]. An existing perceptual evidence also suggests that a pair of HRIRs of a certain direction can be reconstructed with the magnitude and the interaural time delay (ITD) without introducing no obvious perceptual artifacts [93]. Since the magnitude parts of HRTFs have been represented with spherical wavelets previously, the representation of ITD will be considered in this chapter.

In the minimum-phase reconstruction, the original HRTF $H(w)$ can be represented as product of a minimum phase component $H_{min}(w)$ and an all pass filter $H_{ap}(w)$ as described as below.

$$H(w) = H_{min}(w)_{ap}(w) \tag{5.1}$$

Ideally, the minimum phase component of HRTF is obtained by factoring the transfer function polynomial and identifying zeros and poles which are inside unit circle. However this is not practical and a non parametrical method of obtaining minimum phase component is used here [32].

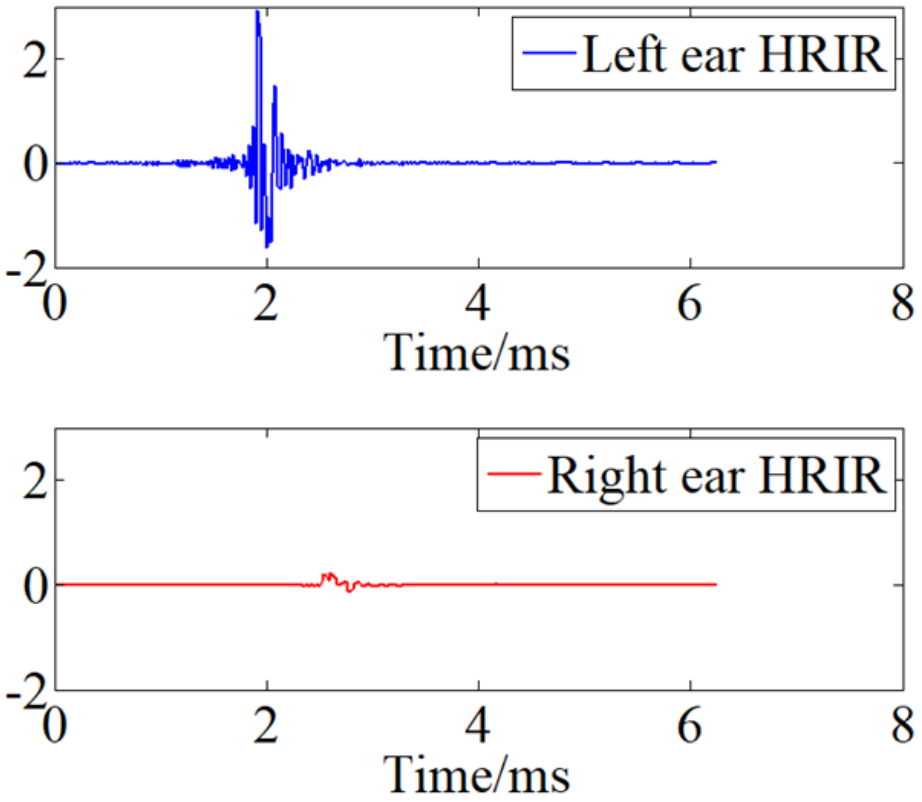


Figure 5.1: Original HRIRs at direction $(80°, 0°)$

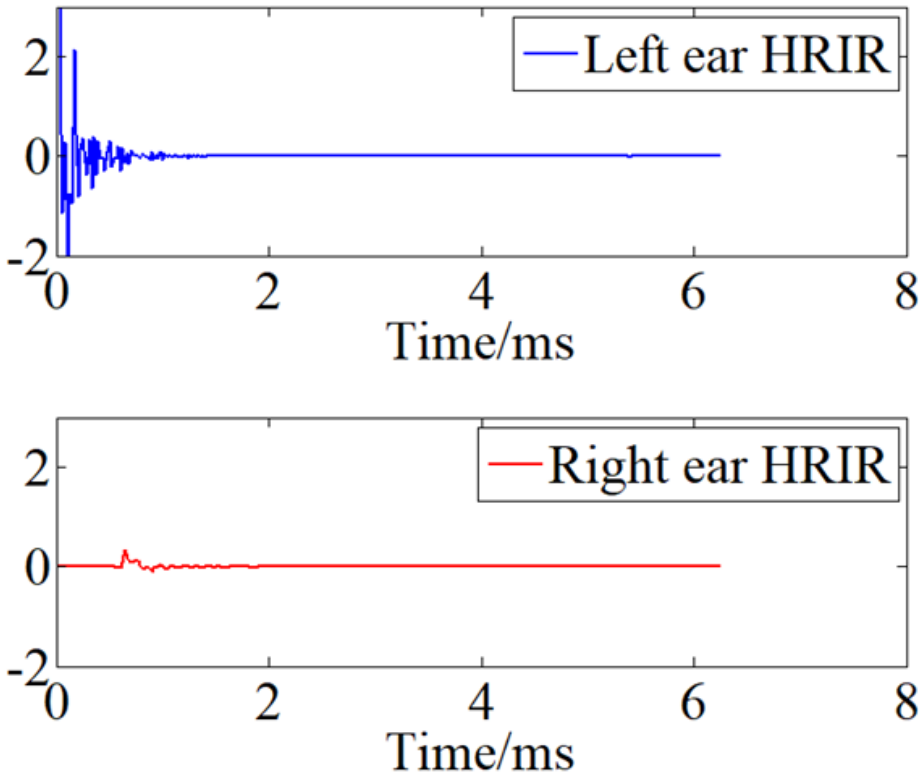


Figure 5.2: HRIRs based on minimum-phase reconstruction.

Here, Fig. 5.1 and Fig. 5.2 present an example of the original HRIRs and the reconstructed HRIRs based the minimum phase model. The original HRIRs in Fig. 5.1 and constructed HRIRs in Fig. 5.2 look quite different, regarding the waveforms, however they should be perceptually identical.

5.3 Estimating interaural time delays using spherical wavelets

5.3.1 Calculation of interaural time delays

According to the literature, there exists three common methods to calculate the ITDs. The first method compares the onset time arrival for the left and right HRIR signal and the difference of the onset time can be regarded as the ITD [94–96]. An example of this onset detection method is shown in Fig. 5.3. The second method considers the interaural cross correlation, and the time delay τ which maximizes the coherence of the signal at one ear with respect to the signal at the other is used to determin the ITD (Eq. 5.2 and Eq. 5.3) [97–99]. While the last method calculates the interual time difference and use the slope of the regression along the frequency as the ITD value [100–102]. All of the above methods should have close results. An excellent review can be found with a paper by Katz and Noisternig [103]. Here in our study, we used the group delay method since it seems to derive ITD smoothly along the direction in space and its effectiveness has been validated in a previous perceptual study [102].

$$ITD(\theta,\phi) = argmax IACC(\theta,\phi,\tau),\ s.t. |\tau| < 1ms, \tag{5.2}$$

$$IACC(\theta,\phi,\tau) = \frac{\int_{t1}^{t2} p_L(\theta,\phi,t) p_R(\theta,\phi,t+\tau) dt}{\sqrt{\int_{t1}^{t2} p_L^2(\theta,\phi,t) dt \int_{t1}^{t2} p_R^2(\theta,\phi,t) dt}} \tag{5.3}$$

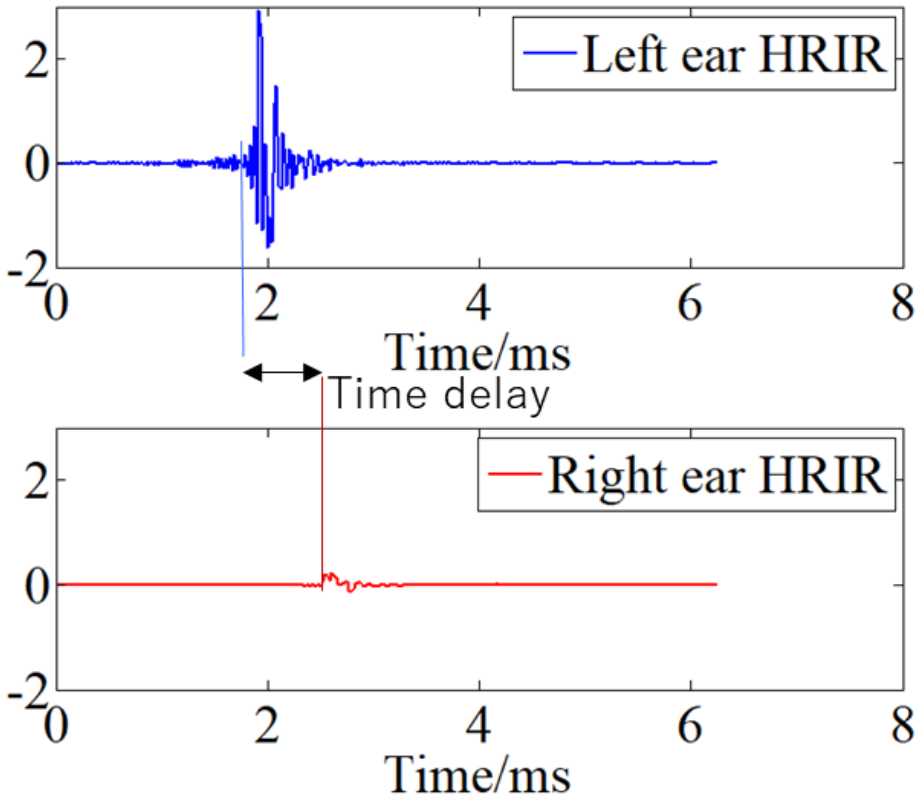


Figure 5.3: ITD calculation based on the onset time difference between the binaural signals for HRTF at (80°, 0°).

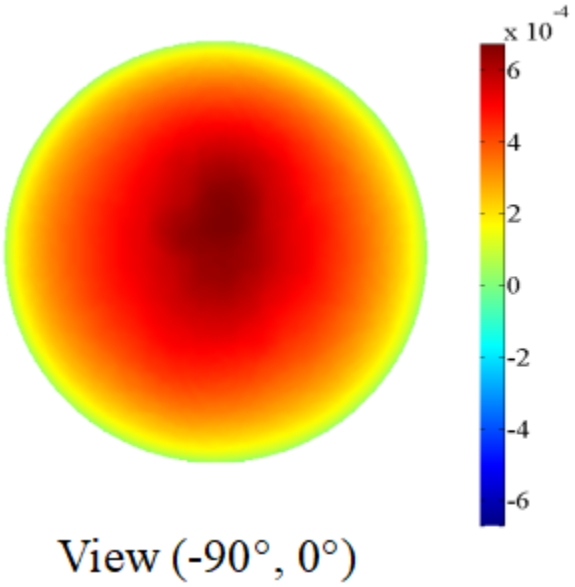


Figure 5.4: ITDs with the right side view.

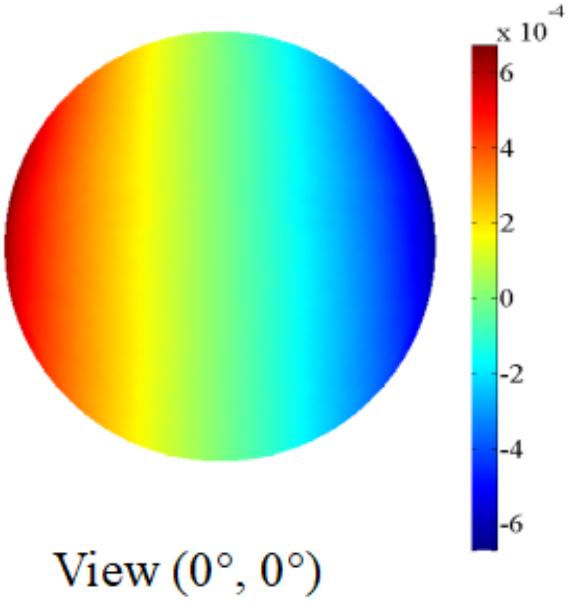


Figure 5.5: ITDs with the front view.

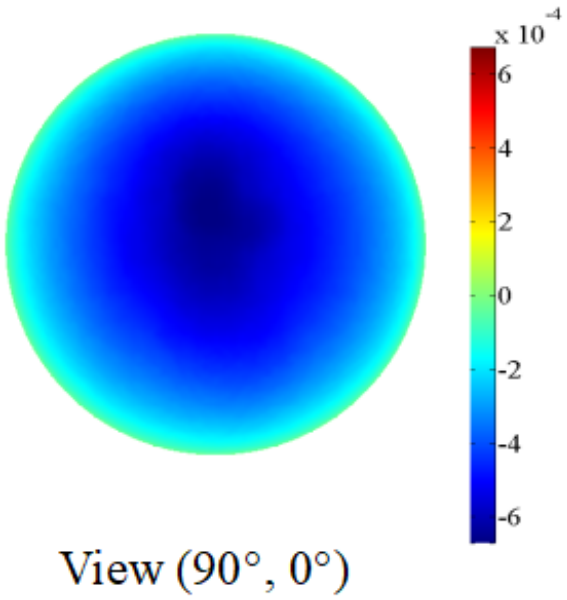


Figure 5.6: ITDs with the left side view.

The calculated ITDs in the spatial domain based on the group delay method are shown in , Fig. 5.4. Fig. 5.5, Fig. 5.6.

5.3.2 Modeling of interaural time delays with spherical wavelets

The decomposition of the target $ITDs(\theta, \phi)$ can be described using the following matrix form:

$$ITDs(\theta, \phi) = \sum_{\vec{v}_{1,i} \in V_1} S(\vec{v}_{1,i}) \varphi_{1,i}(\theta, \phi) + \sum_{\ell=1}^{L-1} \sum_{\vec{m}_{\ell,j} \in M_\ell} W(\vec{m}_{\ell,j}) \cdot \psi_{\ell,j}(\theta, \phi). \tag{5.4}$$

Here, $\varphi_{1,i}(\theta, \phi)$ and $\psi_{\ell,j}(\theta, \phi)$ are the scaling function and spherical wavelet corresponding to $\vec{v}_{1,i}$ and $\vec{m}_{\ell,j}$, respectively; $S(\vec{v}_{1,i})$ and $W(\vec{m}_{\ell,j})$ are the expansion coefficients in the decomposition.

5.4 Evaluation

In this section, numerical experiments are conducted to validate the effectiveness of the method described in Section 5.3.2.

5.4.1 Objective measurement

The goal of this study is to approximate the original HRTF magnitudes using a set of analysis functions. To measure the approximation error in the spatial domain, the root mean squared (RMS) value is used, which is defined as

$$E_{rms} = \sqrt{\frac{1}{N} \sum_{m=1}^{N} \left[ITD_{synth}(\theta_m, \phi_m) - ITD_{target}(\theta_m, \phi_m) \right]^2}, \tag{5.5}$$

where $ITD_{synth}(\theta_m, \phi_m)$ and $ITD_{target}(\theta_m, \phi_m)$ are the reconstructed ITD and target ITD at direction (θ_m, ϕ_m), respectively; and N is the total number of HRTF samples under study.

5.4.2 Results

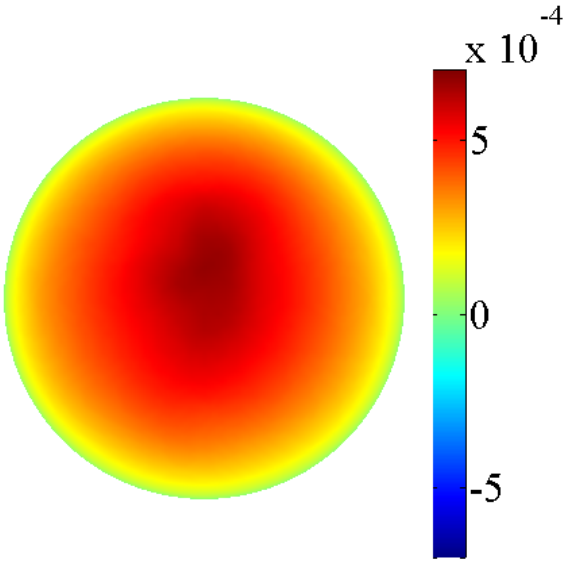


Figure 5.7: ITDs appximated with 121 spherical wavelets (Right side view).

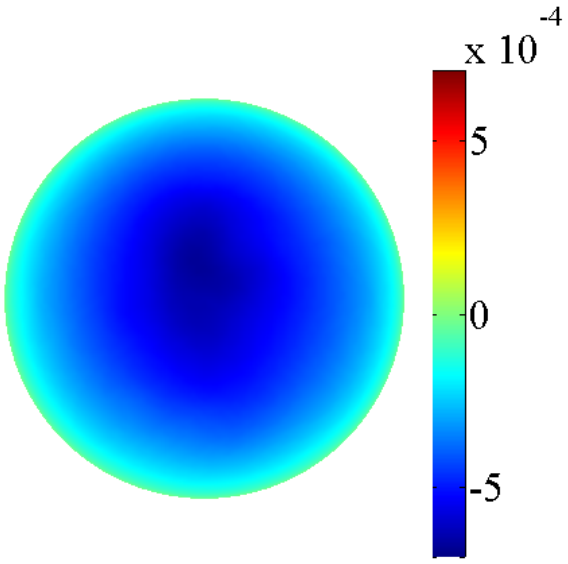


Figure 5.8: ITDs appximated with 121 spherical wavelets (Left side view).

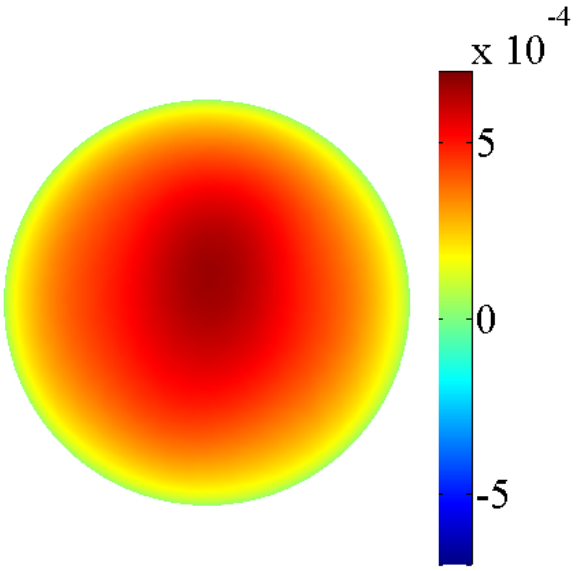


Figure 5.9: ITDs appximated with 10 order spherical harmonics (Right side view).

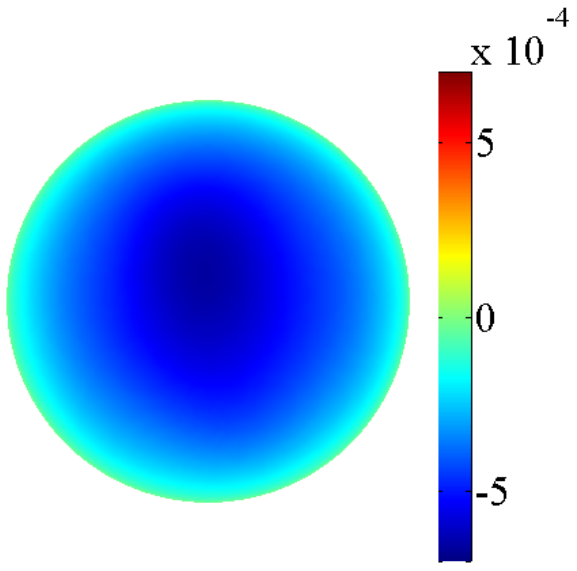


Figure 5.10: ITDs appximated with 10 order spherical harmonics (Left side view).

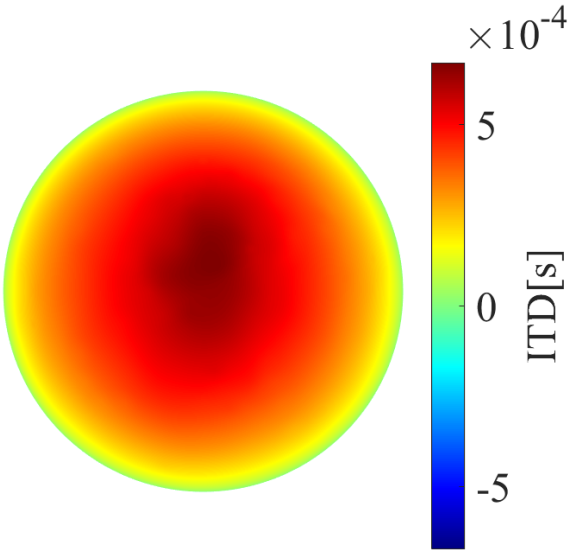


Figure 5.11: ITDs appximated with 441 spherical wavelets (Right side view).

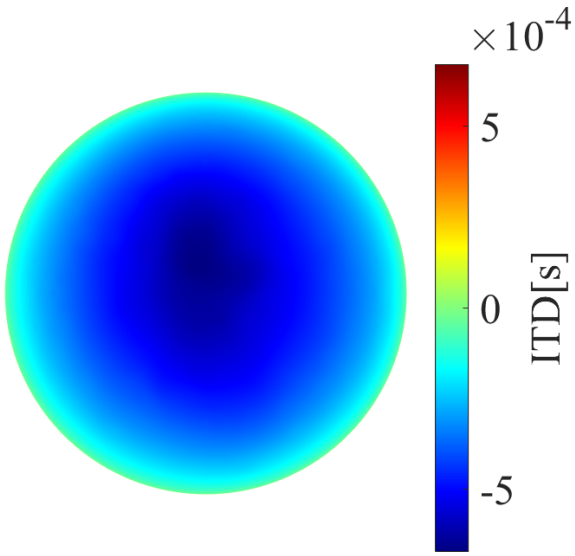


Figure 5.12: ITDs appximated with 441 spherical wavelets (Left side view).

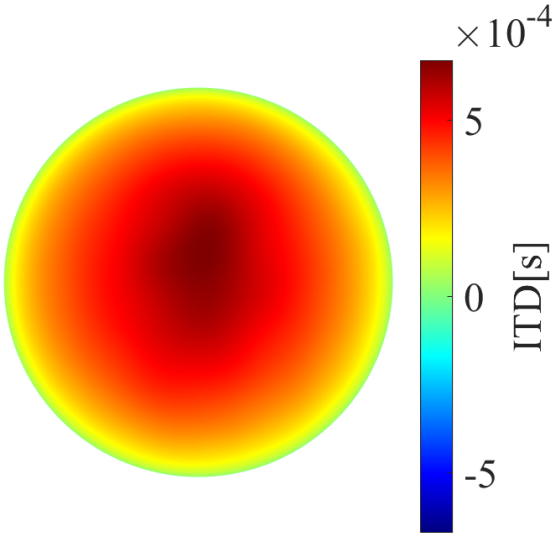


Figure 5.13: ITDs appximated with 20 order spherical harmonics (Right side view).

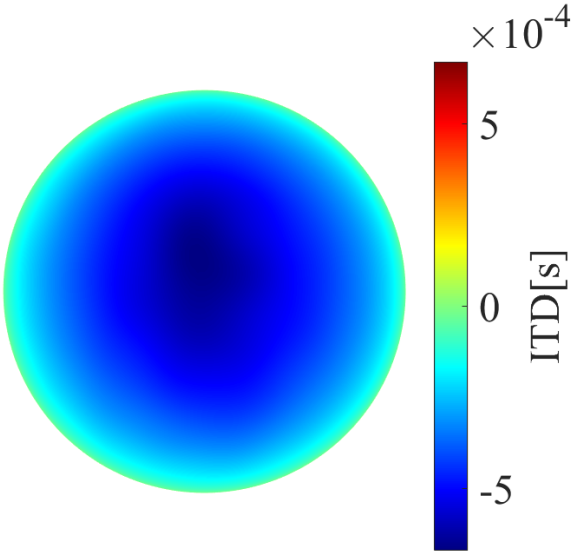


Figure 5.14: ITDs appximated with 20 order spherical harmonics(Left side view).

As shown in Fig. 5.7, Fig. 5.12 and 5.9, 5.10, the ITDs are estimated based on 121 spherical wavelets and 10 order spherical harmonics, respectively. Through visual observation, these two modeling methods do not show significant differences due to the fact that the target ITDs are smooth and not so difficult for modeling. However, in the tables of the calculated approximation errors as shown in Table. 5.1 and Table. 5.2, modeling ITDs with spherical wavelets yield smaller approximation errors than modeling with spherical harmonics when using a same number of analysis functions. Since the approximation errors

are so small, it still remains a question whether these approximated accuracy has significant perceptual necessity or not.

Table 5.1: Comparison of approximated ITDs in all directions using same number of spherical wavelets and spherical harmonics.

Modeling with SWs	Approximation error (μs)	Modeling with SHs	Approximation error (μs)
121 SWs	3.14	10 order SHs	3.46
441 SWs	2.22	20 order SHs	2.37
676 SWs	2.07	25 order SHs	2.18
961 SWs	1.95	30 order SHs	2.09

Table 5.2: Comparison of approximated ITDs in local regions using same number of spherical wavelets and spherical harmonics.

Approximation error (μs)	With 121 SWs	With 10 order SHs	With 441 SWs	With 20 order SHs
Region 1	2.30	2.94	1.34	2.27
Region 2	2.44	4.21	1.34	2.64
Region 3	2.47	3.93	1.33	2.60
Region 4	2.29	2.95	1.33	2.25

5.5 Summary

This chapter reviews the minimum-phase reconstruction of the HRTF, which is a common method to recover the HRIR or complexed valued HRTFs for binaural rendering. Since the HRTF magnitudes has been modeled in the previous chapters, the modeling of ITD is considered here. According to the calculated approximation error (objective evaluation), modeling with spherical wavelets yields smaller approximation errors than modeling with spherical harmonics when using same number of analysis functions in the case of both representing the ITDs in all directions and local regions. The approximated ITDs will help recover the HRIRs for the binaural spatial audio rendering. The reconstructed HRIRs

will also be evaluated in the next chapter based on some perceptual models which are more relevant to the human spatial hearing perception.

Chapter 6

Evaluation based on the perceptual model

6.1 Overview

Until this chapter, the evaluation of the proposed method has been based on objective evaluations with simple criteria. This chapter applies an auditory model which characterizes the human sound localization to the evaluation of the modeling method. In such a manner, this chapter aims to evaluate the proposed method in a way more relevant to human spatial hearing perception.

6.2 Perceptual model of human sound localization

In the past decades, a lot of research has been done trying to figure out the acoustic cues that human listeners utilize to localize the sound direction and how those acoustic cues are processed by the brain. However, most of the proposed human sound localization models in

the past based on simple matching methods with ITD and ILD cues cannot work very well, or cannot predict both the horizontal and elevation directions simultaneously. [104–107]

Here in this thesis, a human sound localization model based on a probabilistic model is applied which is called "An ideal-observer model of human sound localization" proposed by Reijniers et al. recently. [108]

6.2.1 Review on the Ideal-observer model of human sound localization

Ideal-observer model has been widely used for predicting the behavioral performance in the psychological experiments. As with the case of human sound localization, given a pair of signals to the left and right ear, human is able to extract the relevant acoustic cues that are related to sound localization. Here the interaural time difference (ITD) and the spectral information of the left and right ear are considered to be utilized for the acoustic cues for the sound localization. The ITD cue is known to play a vital role for the horizontal sound localization in the lower frequency range while the binaural spectra information may play a more important role in the higher frequency horizontal localization and elevation localization.

Acoustic information

Although the ITD is freuqency dependent, it is considered to be a single delay value which is still believed to have reasonable perceptual significance [40]. As for the spectral information, due to the limitation of human hearing system, it is decomposed into a certain number of frequency channels and represented in log scale. To summarize, the acoustic information used by the listener can be describe as

$$\mathbf{X} = [X_{itd}, \mathbf{X_L}, \mathbf{X_R}], \tag{6.1}$$

where X_{itd} is the ITD of the binaural signal. $\mathbf{X_L}$ and $\mathbf{X_R}$ describe the monaural spectra of the left and right ears, respectively. The ITD is measured by the human brain for processing and can be written as

$$X_{itd} = itd(\theta) + \sigma_{itd}, \tag{6.2}$$

where X_{itd} is the exact ITD corresponding to the directrion θ, and σ_{itd} denotes the error of the measurement due to the limited precision of the hearing apparatus.

Similarly, the logmagnitude spectra in the left and right ear can be represented as below.

$$\mathbf{X_L} = \mathbf{S} + \mathbf{H_L}(\theta) + \sigma_{\mathbf{L}} + \sigma_{\mathbf{S}}; \mathbf{X_R} = \mathbf{S} + \mathbf{H_R}(\theta) + \sigma_{\mathbf{R}} + \sigma_{\mathbf{S}} \tag{6.3}$$

Here $\mathbf{S}$ is the sound source spectrum. $\mathbf{H_L}(\theta)$ and $\mathbf{H_R}(\theta)$ are the logmagnitude of the left and right HRTF. $\sigma_{\mathbf{L}}$, $\sigma_{\mathbf{R}}$ and $\sigma_{\mathbf{S}}$ are the corresponding measurement error of the left, right ear HRTF logmagnitudes and sound source spectrum when utilized as the perception cues by the listeners.

Instead of using the monaural magnitudes directly as above, they can also be transformed into a new basis which consists of the difference and average of the left and right magnitudes as follows:

$$\mathbf{X}_- = \mathbf{X_L}(\theta) - \mathbf{X_R}(\theta) = \mathbf{H_L}(\theta) - \mathbf{H_R}(\theta) + \sigma_- \tag{6.4}$$

$$\mathbf{X}_+ = (\mathbf{X_L}(\theta) + \mathbf{X_R}(\theta))/2 = \mathbf{S} + (\mathbf{H_L}(\theta) + \mathbf{H_R}(\theta))/2 + \sigma_+, \tag{6.5}$$

with

$$\sigma_- = \sigma_{\mathbf{L}} - \sigma_{\mathbf{R}}, \sigma_+ = \sigma_{\mathbf{S}} + (\sigma_{\mathbf{L}} + \sigma_{\mathbf{R}})/2. \tag{6.6}$$

Here $\mathbf{X}_-$ and $\mathbf{X}_+$ are the interaural difference and average of the left and right ear magnitude, respectively; σ_- and σ_+ are the corresponding measurement errors when utilizing these cues. Therefore, the acoustic information can be summarized as below instead.

$$\mathbf{X} = [X_{itd}, \mathbf{X}_-, \mathbf{X}_+], \tag{6.7}$$

It is believed that the just noticed difference (JND) varies along the ITD value [109], which may suggests that the measurement of ITD does not follow a normal distribution since the perception of ITD is not uniform in time unit. The just noticed difference (JND) basically increases as the actual value of ITD and may read as

$$jnd(itd) = a + b \cdot itd. \tag{6.8}$$

Here, $a = 32.5\mu s$, and $b = 0.095$. Further, the formula below transforms the ITD scale from time unit to jnd unit.

$$itd \rightarrow \int_0^{itd} \frac{1}{jnd(x)} dx = \frac{1}{b}[log(a + b \cdot itd) - log(a)]. \tag{6.9}$$

Therefore, the ITD will be represented in this new unit scale which may be more relevant to the actual utilization of the ITD cues. Fig. 6.1 and Fig. 6.2 shows the comparison between the ITD in time and jnd units, which suggests that the ITD in jnd units is not as flat as that in time units and gradually increase as with the ITD value.

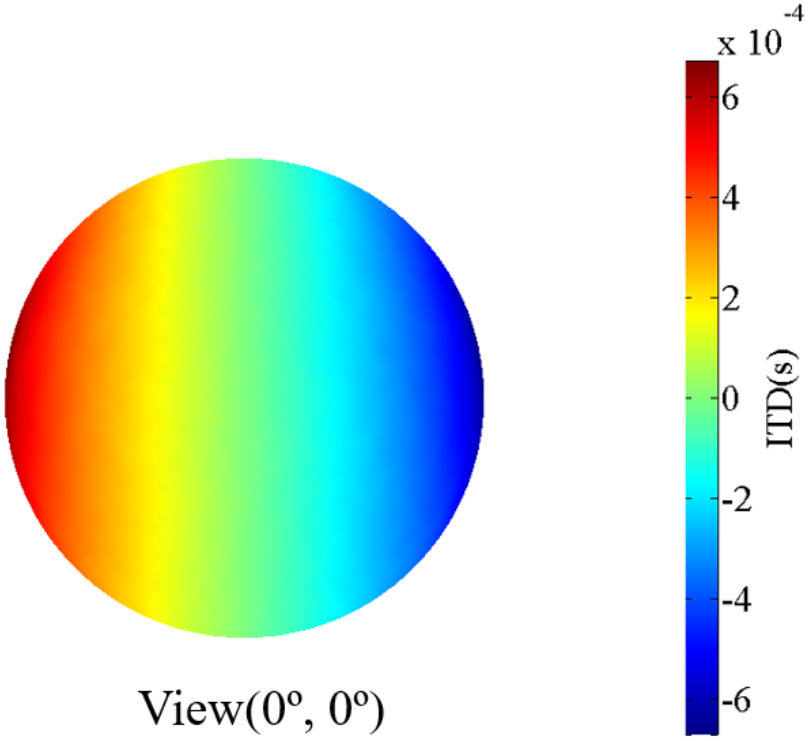


Figure 6.1: ITD in time units.

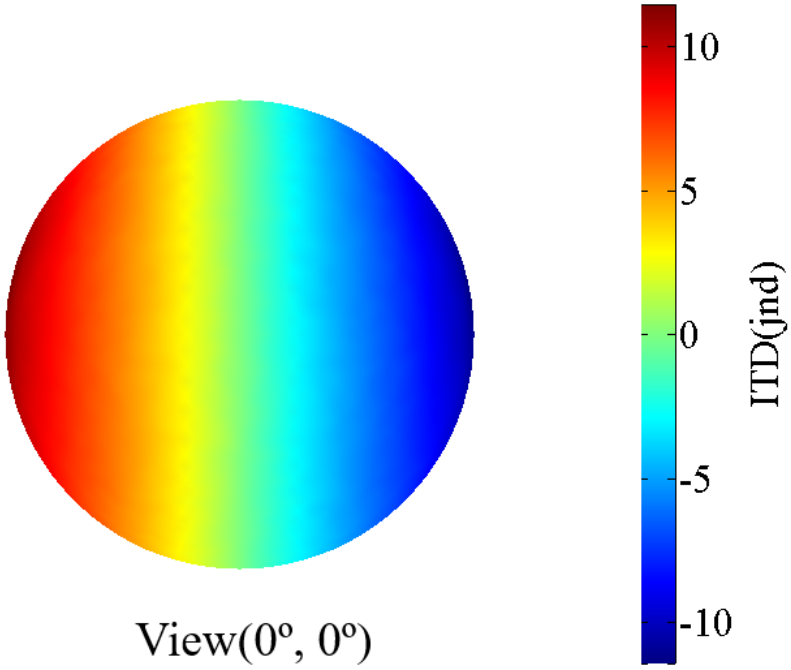


Figure 6.2: ITD in jnd units.

Rather than represent the log-magnitudes along each frequency bin, the received signal is decomposed into a certain number of frequency channels. These frequency channels are chosen based on the equivalent rectangular bandwidth (ERB) of the auditory filter such as the Gammatone filter [110]. There may be different ways to calculate the ERB for each center frequency, while a typical one is used here as shown in the equation as below. [111]

$$ERB(f) = 6.23f^2 + 93.39f + 28.52. \tag{6.10}$$

The center frequencies are chosen so that there are no severe overlaps between the adjacent frequency channels, namely, a center frequency is calculated as the sum of the previous center frequency and its corresponding ERB. Fig. 6.4 shows an example of the Gammatone filters with some different center frequencies. Fig. 6.4 shows 30 four order Gammatone filters covering frequency range from 300 Hz to 15000 Hz.

With the defined frequency channels, the left and right ear signal are processed by those auditory filters one by one, and the energy of each channel's output represents spectra information.

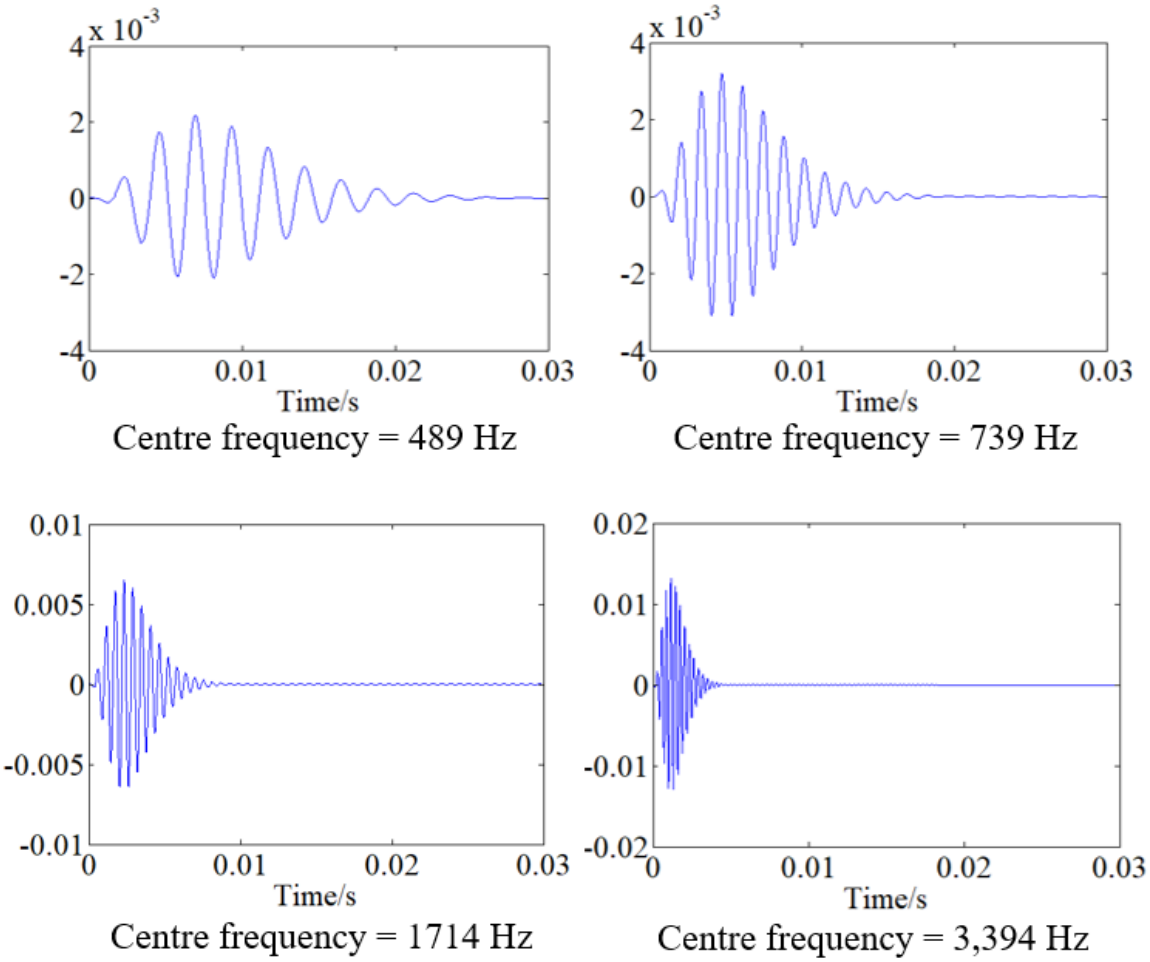


Figure 6.3: Gammatone filter with different center frequencies.

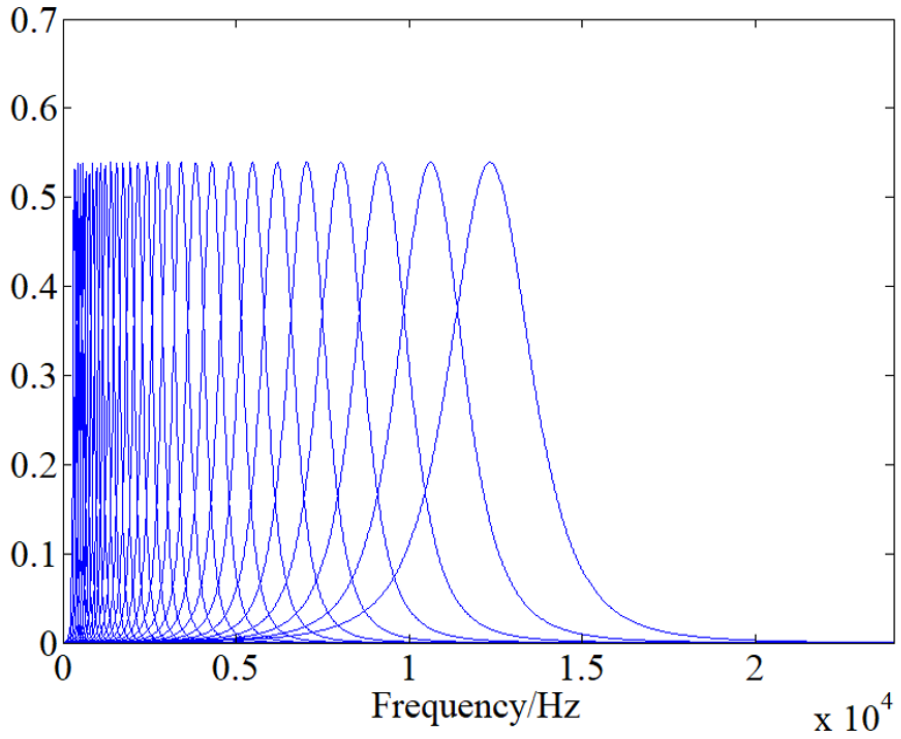


Figure 6.4: Frequency responses of 30 Gammatone filte covering frequency from 300 Hz to 15000 Hz.

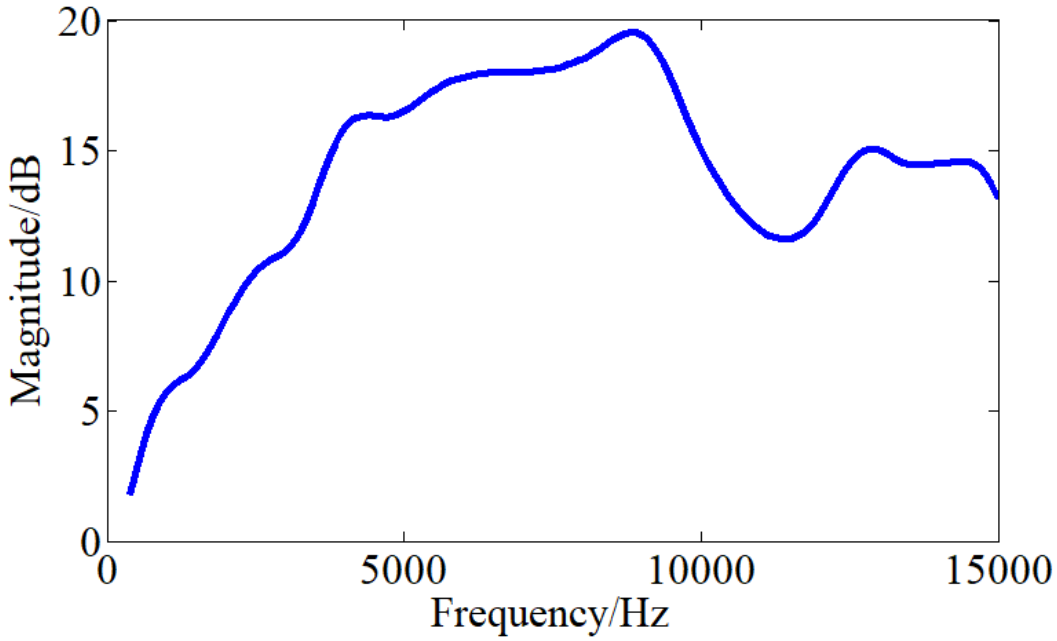


Figure 6.5: HRTF magnitude along the frequency in certain direction.

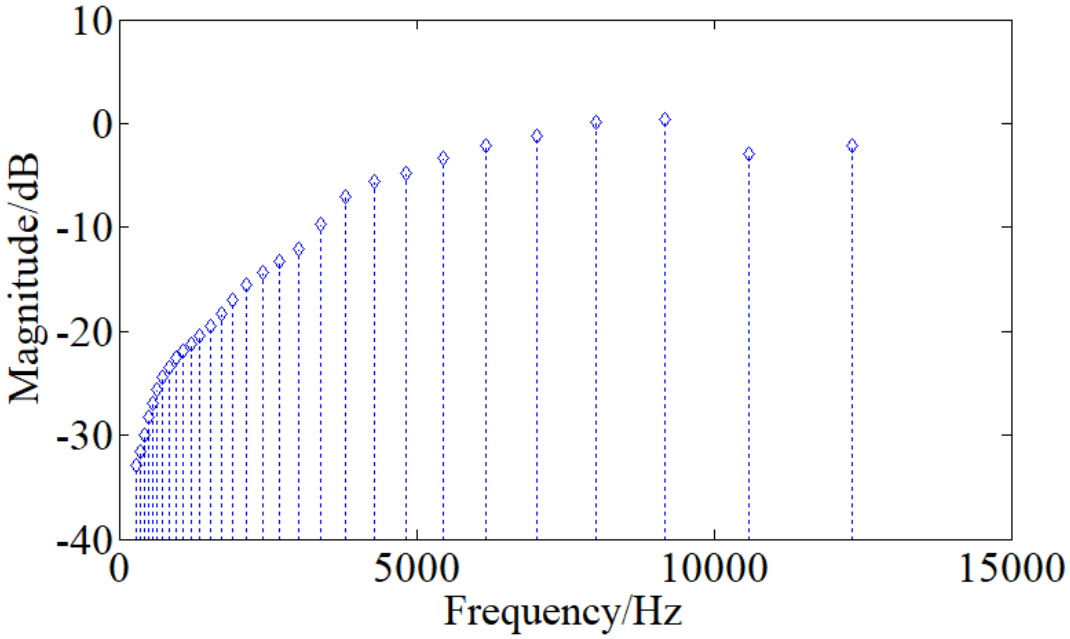


Figure 6.6: Frequency channel responses based on 30 auditory filters.

Estimating sound source position based on Bayes' rule

To predict the sound source direction θ given the acoustic information $\mathbf{X}$, the Bayes' rule is applied for this purpose.

$$P(\theta|\mathbf{X}) = \frac{P(\theta)P(\mathbf{X}|\theta)}{P\mathbf{X}} \tag{6.11}$$

where $P(\theta)$ is the prior knowledge of the perceived sound direction by the listener. This prior knowledge should not be uniformly distributed in space. However, for simplicity, $P(\theta)$ and $P(\mathbf{X})$ are considered uniformly distributed parameters. Therefore, the calculation of the posterior probability $P(\theta|\mathbf{X})$ can be reduced to the calculation of the likelihood P(X|θ).

By including the measurement error $\sigma_{\mathbf{itd}}$, $\sigma_{\mathbf{L}}$, $\sigma_{\mathbf{R}}$, and $\sigma_{\mathbf{S}}$, the likelihood $\mathbf{P}(\mathbf{X}|\theta)$ can be written as

$$\mathbf{X}_\theta[\sigma] = \mathbf{T}_\theta + \sigma \tag{6.12}$$

$$P(X|\theta) = \frac{1}{(2\pi)^{N/2}|\sum|^{1/2}} \cdot exp\{-\frac{1}{2}(X - T_\theta)^T \sum^{-1} (X - T_\theta)\} \tag{6.13}$$

Here, the directional template $\mathbf{T}_\theta = [itd, (\mathbf{H_L} - \mathbf{H_R}), (\mathbf{S} + (\mathbf{H_L} + \mathbf{H_R})/2)](\theta)$, and the covari-

ance matrix is denoted as

$$\sum = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sum_{-} & 0 \\ 0 & 0 & \sum_{+} \end{bmatrix}. \tag{6.14}$$

with $\sum_{-} = 2\sigma_I^2 \cdot \mathbf{I}$ and $\sum_{+} = (\sigma_I^2/2 + \sigma_S^2) \cdot \mathbf{I} + \sigma^2 \cdot \mathbf{1}$. This covariance variable determines how much uncertainty of human hearing when processing the acoustic information. A bigger measurement error indicates that the listener may be not good at extracting the cues precisely for predicting the direction of the sound source.

In order to evaluate the average localization performance, the localization test is done under a certain number of trials with multiple noise realization. The average localization error can be calculated as

$$\epsilon(\theta) = \frac{1}{M} \sum_{i=1}^{M} arccos\{\hat{\theta}(\mathbf{X}_\theta[\sigma_\mathbf{i}]) \cdot \theta\}. \tag{6.15}$$

Here M is the total number of the test trials, and $\mathbf{X}_\theta[\sigma_\mathbf{i}]$ denotes the acoustic information with the noise realization σ_i. Please notice that, this error metric introduced here is the absolute error of the sound localization. The localization error can also be calculated with different ways such as the lateral and polar localization error, which accounts for the azimuth and elevation localization error, respectively.

Fig. 6.8 shows a example of localization estimation of a target sound source at $(75°, 26°)$, with running the test for 500 trials. The simulated result shows that the ideal observer model can estimate the source position with good accuracy. Besides, an interesting finding may also observed that this model also estimate the sound source with front back confusions even with a relatively low probability.Since humans may also localize sounds with possibility of estimating the sound source positions with front back confusions, it suggests that the this model may be a good option for evaluating the human sound localization with perceptual significance.

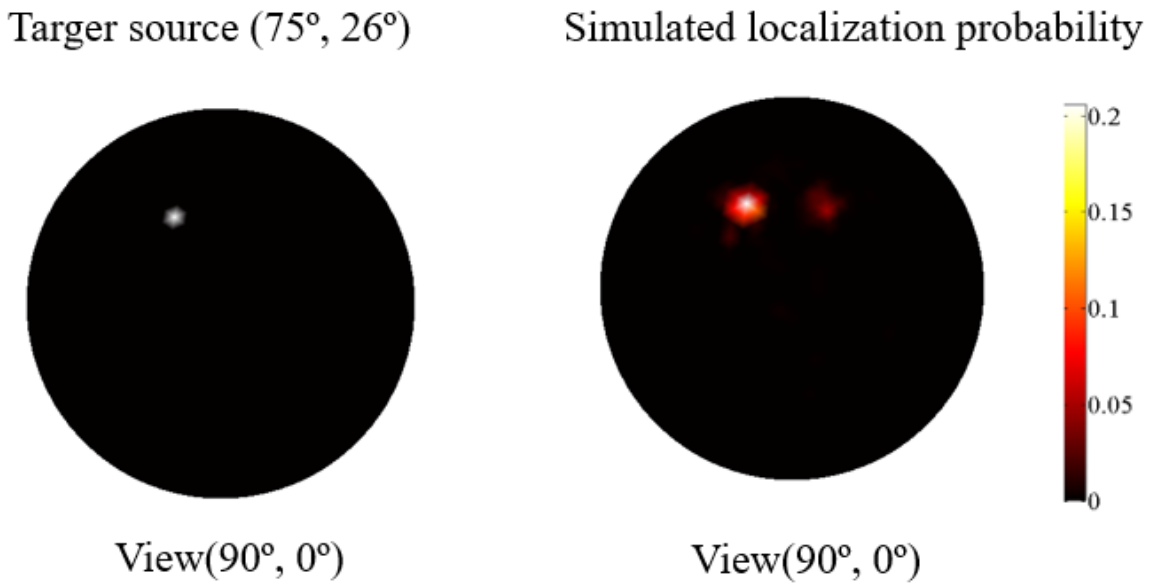


Figure 6.7: Example of the localization test with 500 trials of a sound source (white point in the left pannel) and the simulated localization probability (right pannel).

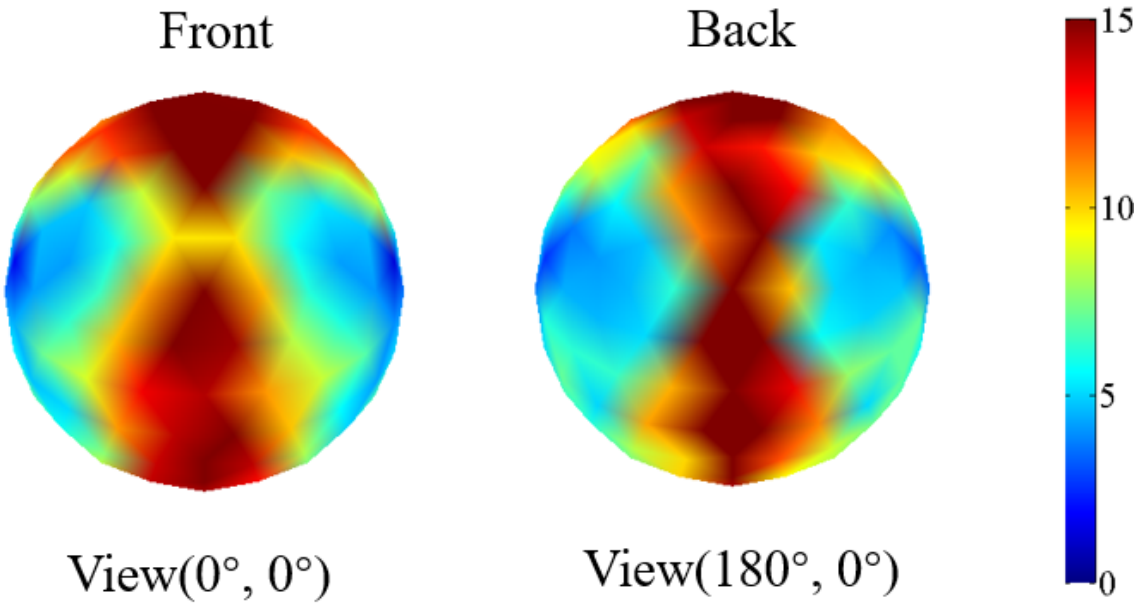


Figure 6.8: Average localization error of 500 trials for 160 source directions on the sphere.

Fig. 6.8 shows the average localization error of 500 trials for each of 162 directions distributed on the sphere. The localization performance is similar to that in the study by Reijniers et al. [108].

6.3 Evaluation of the proposal based on the perceptual model

In the previous chapter, the HRTF magnitude and ITD have been represented with spherical wavelets, which shows better approximation efficiency than the spherical harmonics under the evaluated conditions. However, those evaluations are conducted with very simple objective error metric, which may not reflect the perceptually related significance. Therefore, in this section, the modeled HRTFs based on spherical wavelets and spherical harmonics are evaluated and compared based on the ideal observer model.

To recover the HRIRs for evaluation with the perceptual model, both the ITD and HRTF magnitudes are necessary, as suggested by the minimum-phase reconstruction introduced in Chapter 5. Here, the HRTF magnitudes are approximated with a same method as that introduced in Chapter 4, in which the spherical wavelets are used for representing the HRTF magnitudes. Specifically, a few local regions are selected for recovering the time domain HRIRs based on the HRTF magnitudes approximated with spherical harmonics and spherical wavelets. The recovered HRIRs in each local region are then analyzed and compared based on the ideal observer model.

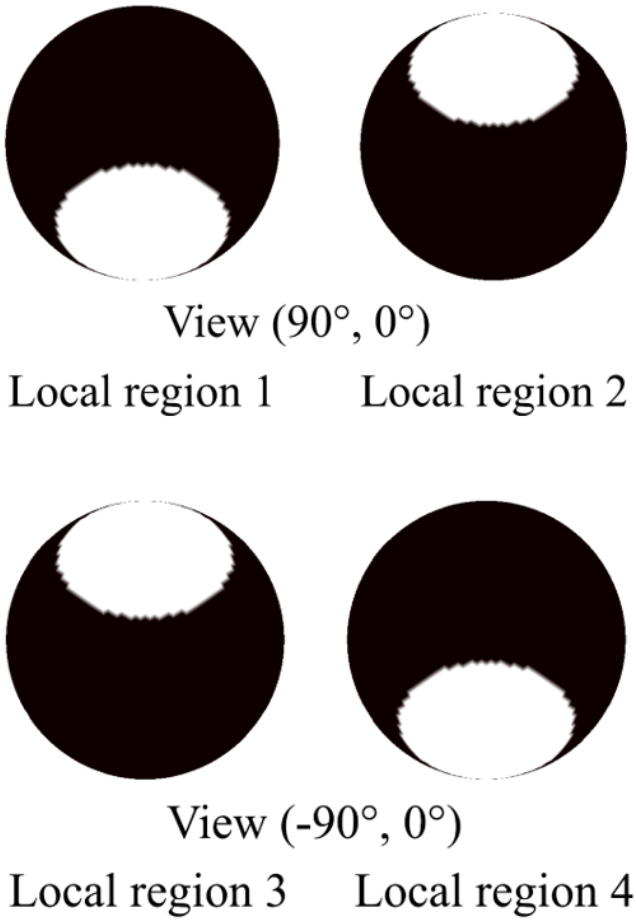
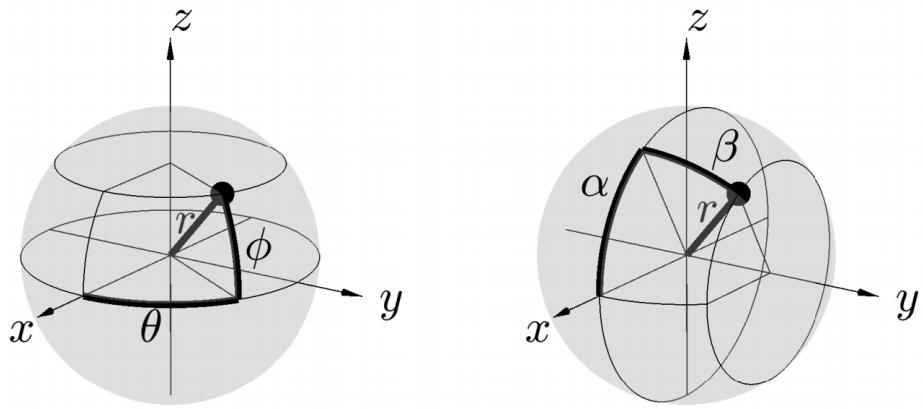


Figure 6.9: Four local regions (in white) selected for the evaluation of the local representation.

For the numerical experiments, the HRTF magnitudes are reconstructed with spherical harmonics and spherical wavelets with a same number of analysis functions in a local region centered at (90°,−48°) with a size of 1.40 steradians (Local region 1 in Fig. 6.9).

6.3.1 Lateral and polar localization error



(a) Vertical polar. **(b)** Interaural polar.

Figure 6.10: Spherical coordinate systems.

In order to evaluate the lateral localization and polar localization separately, the interaural polor coordinate (Fig. 6.10 (b)) is used, in place of vertical coordinate (Fig. 6.10 (a)) which has been used so for in this study. In the polar coordinate, a point (r, α, β) where the lateral angle $\beta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, and the polar angle $\alpha \in [-\pi, \pi]$.

In each local region, HRTFs 17 directions that are nearly uniformly distributed are evaluted with the perceptual model. The Table 6.11 shows result, which suggests that there is no obvious difference between the simulated localization performance based on the spherical wavelet and spherical harmonic methods. It may indicate that the perceptual model used in the simulation may be very tolerant of the loss of the HRTF spatial details. The small differences of the HRTFs synthesized by this two methods may not be reflected by the introduced perceptual model.

Average Lateral Error	*Local region 1*	*Local region 2*	*Local region 3*	*Local region 4*	*Average*
With 121 SWs	0.26	0.29	0.33	0.32	0.30
With 10 order SHs	0.38	0.48	0.42	0.48	0.44
With 441 SWs	0.35	0.31	0.27	0.26	0.30
With 20 order SHs	0.29	0.26	0.25	0.25	0.26
Average Polar Error	***Local region 1***	***Local region 2***	***Local region 3***	***Local region 4***	***Average***
With 121 SWs	3.42	5.44	5.88	4.65	4.85
With 10 order SHs	3.95	6.53	7.42	4.28	5.54
With 441 SWs	3.34	5.11	6.34	4.23	4.76
With 20 order SHs	3.11	5.79	6.29	4.20	4.85

Figure 6.11: Average lateral localization error and polar localization error of the evaluated directions for HRTFs in each local region.

6.3.2 Absolute localization error

Localization Error	(90.0° , 31.7°)	(58.28° , -54°)	(121.7° , -54°)	(74.6° , -26.3°)
With 121 SWs	7.52	7.78	6.90	4.13
With 10 order SHs	8.26	8.65	8.48	6.06
With 441 SWs	6.89	7.51	6.83	4.25
With 20 order SHs	7.26	7.46	6.69	5.09

Localization Error	(105.3° , -26°)	(28.1° , -59.8°)	(78.4° , -42.9°)	Average
With 121 SWs	7.05	13.49	5.0	7.41
With 10 order SHs	9.01	12.5	6.75	8.53
With 441 SWs	6.50	13.7	4.45	7.12
With 20 order SHs	8.10	12.8	4.86	7.46

Figure 6.12: Absolute localization error of 500 trials for 17 source directions in local region 1. The values in red means the error based on spherical wavelets (SWs) is bigger than that based on spherical harmonics (SHs).

The evaluation based on the ideal observer model shown in Fig. 6.12 shows that when modeling HRTF magnitudes with same number of analysis functions in the evalauted conditions, the localization error of the SWs method has slightly better performance than the SHs method except for a small cases. The HRTF magnitudes are evaluated in other Local regions (local region 2, 3, and 4 in Fig. 6.9) as well, the average error shown in Table. 6.13 suggests a consistent result as that in local region 1.

Average Error	Local region 1	Local region 2	Local region 3	Local region 4
With 121 SWs	7.41	11.39	11.51	8.35
With 10 order SHs	8.53	8.65	12.00	8.73
With 441 SWs	7.12	10.66	11.39	7.99
With 20 order SHs	7.46	10.95	11.41	8.20

Figure 6.13: Average absolute localization error of all evaluated directions in each local region 1, 2, 3, and 4 in Fig. 6.9.

Although the above results suggest that HRTFs magnitudes approximated by the spherical wavelets seem have very close performance compared with the spherical harmonics based ones. This result may suggests that some spatial details of HRTFs may be not so perceptually important which was also shown by a previous perceptual study in which the HRTFs magnitudes are represented with spherical harmonics up to a truncated order [44]. However there also exists study suggesting that the spatial details in the higher spherical harmonics also contribute to the perception of timbre. [112]. In any sense, the ideal observer model used here may be a appropriate tool for reflecting the perception related performance with the tested HRIRs. However, the ultimate evaluation of the HRTFs should be relying on the behavioral experiments which however are out of the scope of this study.

6.4 Summary

In this chapter, a auditory model called ideal observer model is introduced and reviewed, which shows a reasonable performance for modeling the human sound localization. The HRTF magnitudes reconstructed by the spherical wavelets and spherical harmonics were compared in terms of the modeling accuracy based on the objective error metric, however those error metric does not consider the human perception. Therefore, those reconstructed HRTFs are evaluated by the auditory model introduced here and the evaluation is compared between the modeling with spherical harmonics and spherical wavelets when using same number of analysis functions. The numerical experiment shows that when representing the HRTF in a evaluated local region, the spherical wavelet method yield slightly smaller localization error than the spherical harmonic method.

Chapter 7

Conclusions

This thesis focuses on the modeling of spatial patterns of the HRTF. Rather than the conventional method based on the spherical harmonics, this thesis mainly proposed modeling methods with considering the HRTF local features. These models are used for:

- Chapter 3: Modeling HRTFs with continuous spherical wavelets
- Chapter 4: Modeling HRTFs with discrete spherical wavelets
- Chapter 5: Modeling HRIRs with spherical wavelets
- Chapter 6: Evaluation based on a perceptual model

The main difference between the proposed method and the conventional method is the analysis function, either local functions or global functions. With modeling based on global functions, some fast changing spatial variations may require modeling up to a high order which is not efficiency. Besides, the fact that human's spatial hearing has different resolutions along the direction, which also suggest that methods are needed to modeling the HRTF locally in space.

In chapter 3, in order to model the HRTF local feature, a set of continuous local functions inspired by the wavelet transform on the sphere is proposed to capture HRTF local features rather than the conventional method based on spherical harmonics that are global

functions on the sphere. Simulation results show that with a same number of analysis functions, the proposed continuous local function yields smaller approximation error than the spherical harmonic method. Besides, there is a correspondence between the higher scale expansion coefficients and some HRTF local features. This correspondence can be improved by improving the orthogonality among the analysis functions.

Since the classic wavelet theory can be rather difficult to be applied for the case on the sphere, the local functions defined in chapter 3 lack orthogonality and therefore is not convenient to control the spatial resolution on the sphere based on the expansion coefficients. In chapter 4, the discrete spherical wavelets are introduced and applied for the purpose of modeling the HRTF spatial patterns. With the discrete spherical wavelet based on the lifting scheme, the HRTF magnitudes are decomposed and represented with a set of expansion coefficients. Thanks to the good orthogonality of the discrete spherical wavelets, the derived expansion coefficients can have a good description of the HRTF spatial features. To evaluate the modeling by this method, the approximation error is compared between the spherical harmonics methods and the spherical wavelet method. When using same number of analysis function for representing the HRTF magnitude in all directions, the proposal performs slightly better or at least comparable with the spherical harmonic method with the evaluated condition. Besides, when comparing the approximation error in the defined local regions, the proposed modeling method based on discrete spherical wavelet yields smaller error than the spherical harmonic method during the audible frequency range.

In chapter 5, the modeling based on spherical wavelets is extended for the HRIRs. Rather than model the HRIRs themselves, they are first decomposed into a ITD value and a pair of magnitudes, and the latter has been modeled in the previous chapter. Then the modeling of the HRIRs is reduced to the modeling of ITDs. Same with the previous comparing method, the target ITDs are represented with SHs and SWs with a same number of functions in the evaluated conditions. The simulation results suggest a slight better performance with the SWs method. However, it still remains a problem whether such a small difference is perceptually significant or not.

In chapter 6, a human sound localization model is reviewed and proposed for evaluating the SWs based HRTF modeling, aiming at better reflecting the perceptual advantage of the proposed method. The numerical experiments show that the HRTF reconstructed with the spherical wavelets method have smaller approximation error than the spherical harmonic method in a evaluated local region. Therefore, the perceptual significance of the proposed method is validated.

Bibliography

[1] L. Rayleigh, "On our perception of sound direction," *Philosophical Magazine*, vol. 13, no. 74, pp. 214–232, 1907.

[2] E. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1848–1861, 1974.

[3] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*. Cambridge, MA, USA; London, England.: The MIT Press, revised ed., 1997.

[4] Y. Suzuki, D. Brungart, H. Kato, K. Iida, D. Cabrera, and Y. Iwaya, *Principles and Applications of Spatial Hearing*. World Scientific, 2011.

[5] J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional sensitivity of sound-pressure levels in the human ear canal," *The Journal of the Acoustical Society of America*, vol. 86, no. 1, pp. 89–108, 1989.

[6] T. Sandel, D. Teas, W. Feddersen, and L. Jeffress, "Localization of sound from single and paired sources," *the Journal of the Acoustical Society of America*, vol. 27, no. 5, pp. 842–852, 1955.

[7] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 159–168, 1990.

[8] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835–850, 2007.

[9] Y. Iwaya, Y. Suzuki, and S. Takane, "Effects of listener's head movement on the accuracy of sound localization in virtual environment," in *Proc. of the 18th International Congress on Acoustics*, pp. 26–29, 2004.

[10] D. S. Brungart, B. M. Sheffield, and L. R. Kubli, "Development of a test battery for evaluating speech perception in complex listening environments," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 777–790, 2014.

[11] G. J. Thomas, "Experimental study of the influence of vision on sound localization.," *Journal of Experimental Psychology*, vol. 28, no. 2, p. 163, 1941.

[12] V. Mellert, K. Siebrasse, and S. Mehrgardt, "Determination of the transfer function of the external ear by an impulse response measurement," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1913–1915, 1974.

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

[14] Y. Suzuki, F. Asano, H.-Y. K. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *Journal of the Acoustical Society of America*, vol. 97, pp. 1119–1123, Feb. 1995.

[15] W. G. Gardner and K. D. Martin, "Hrtf measurements of a kemar," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[16] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 99–102, IEEE, 2001.

[17] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: effects of visual environment, pointing method, and training," *Attention, perception, & psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[18] C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. Van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the sydney york morphological and acoustic record-

ings of ears database," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 37–46, 2014.

[19] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a head-related transfer function database with high spatial resolution," in *7th Forum Acusticum (EAA)*, 2014.

[20] R. D. Ciskowski and C. A. Brebbia, *Boundary element methods in acoustics*. Springer, 1991.

[21] X. Yuan, D. Borup, J. Wiskin, M. Berggren, and S. Johnson, "Simulation of acoustic wave propagation in dispersive media with relaxation losses by using fdtd method with pml absorbing boundary condition," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 46, no. 1, pp. 14–23, 1999.

[22] T. Sakuma, S. Sakamoto, and T. Otsuru, *Computational simulation in architectural and environmental acoustics*. Springer, 2014.

[23] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *Journal of the Acoustical Society of America*, vol. 119, pp. 2589–2598, May 2006.

[24] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Comparison of simulated and measured hrtfs: Fdtd simulation using mri head data," in *Audio Engineering Society Convention 123*, Audio Engineering Society, 2007.

[25] Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, "Boundary element simulation of hrtfs and sound fields produced by virtual acoustic imaging systems," in *Audio Engineering Society Convention 105*, Audio Engineering Society, 1998.

[26] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. i. rigid model calculation," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2440–2448, 2001.

[27] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. ii. impedance effects and comparisons to real measurements," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2449–2455, 2001.

[28] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *Journal of the Acoustical Society of America*, vol. 104, pp. 3048–3058, Nov. 1998.

[29] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The use of head-and-torso models for improved spatial sound synthesis," (Los Angeles, CA, USA), Audio Engineering Society, Oct. 2002.

[30] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space," in *Audio Engineering Society Convention 107*, Audio Engineering Society, 1999.

[31] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[32] J. O. Smith, *Introduction to digital filters: with audio applications*, vol. 2. Julius Smith, 2007.

[33] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolating head related transfer functions in the median plane," in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pp. 167–170, IEEE, 1999.

[34] K.-S. Lee and S.-P. Lee, "A relevant distance criterion for interpolation of head-related transfer functions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1780–1790, 2011.

[35] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response filter models of the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3278–3278, 1995.

[36] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *Journal of the Audio Engineering Society*, vol. 48, no. 11, pp. 1011–1031, 2000.

[37] J. Li, J. Zhang, S. Sakamoto, Y. Yan, *et al.*, "An efficient finite-impulse-response filter model of head-related impulse response," in *Proceedings of Meetings on Acoustics*, vol. 19, p. 050173, Acoustical Society of America, 2013.

[38] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 7, pp. 188–196, Mar. 1999.

[39] W. L. Martens, "Principal component analysis and resynthesis of spectral cues to perceived direction," in *Proceedings of the International Computer Music Conference*, (San Francisco, CA, USA), pp. 274–281, International Computer Musis Association, 1987.

[40] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *Journal of the Acoustical Society of America*, vol. 91, pp. 1637–1647, Mar. 1992.

[41] M. J. Evans, "Analyzing head-related transfer functions measurements using surface spherical harmonics," *Journal of the Acoustical Society of America*, vol. 104, pp. 2400–2411, Oct. 1998.

[42] B.-S. Xie, "Recovery of individual head-related transfer functions from a small set of measurements," *Journal of the Acoustical Society of America*, vol. 132, pp. 282–294, July 2012.

[43] S. Shekarchi, J. Christensen-Dalsgaard, and J. Hallam, "A spatial compression technique for head-related transfer function interpolation and complexity estimation," *The Journal of the Acoustical Society of America*, vol. 137, no. 1, pp. 350–361, 2015.

[44] G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 921–930, 2015.

[45] M. Aussal, F. Alouges, and B. Katz, "Hrtf interpolation and itd personalization for binaural synthesis using spherical harmonics," in *Audio Engineering Society Conference: UK 25th Conference: Spatial Audio in Today's 3D World*, Audio Engineering Society, 2012.

[46] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics," *Acta Acustica united with Acustica*, vol. 98, pp. 72–82, Jan. 2012.

[47] A. W. Mills, "On the minimum audible angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.

[48] P. Minnaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 10, pp. 919–929, 2005.

[49] K. Watanabe, S. Takane, and Y. Suzuki, "A novel interpolation method of HRTFs based on the common-acoustical-pole and zero model," *Acta Acustica united with Acustica*, vol. 91, no. 6, pp. 958–966, 2005.

[50] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 45–48, May 2004.

[51] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values," *Acoustical Science and Technology*, vol. 38, no. 1, pp. 1–13, 2017.

[52] H. Hu, L. Zhou, H. Ma, and Z. Wu, "Hrtf personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, vol. 69, no. 2,

pp. 163–172, 2008.

[53] W. W. Hugeng and D. Gunawan, "Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements," *arXiv preprint arXiv:1005.5137*, 2010.

[54] Y. Suzuki, K. Watanabe, Y. Iwaya, J. Gyoba, and S. Takane, "Adjustment of interaural time difference in head related transfer functions based on listeners' anthropometry and its effect on sound localization," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2485–2485, 2005.

[55] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki, and K. Aso, "Estimation of interaural level difference based on anthropometry and its effect on sound localization," *The Journal of the Acoustical Society of America*, vol. 122, no. 5, pp. 2832–2841, 2007.

[56] A. W. Mills, "On the minimum audible angle," *Journal of the Acoustical Society of America*, vol. 30, pp. 237–246, Apr. 1958.

[57] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, 1999.

[58] C. Müller, *Spherical harmonics*, vol. 17 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, 1966.

[59] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized HRTF fitting using spherical harmonics," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, NY, USA), pp. 257–260, IEEE, Oct. 2009.

[60] K.-S. Lee and S.-P. Lee, "A relevant distance criterion for interpolation of head-related transfer functions," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, pp. 1780–1790, Aug. 2011.

[61] W. Zhang, T. Abhayapala, and R. A. Kennedy, "Insights into head-related transfer function: spatial dimensionality and continuous representation," *Journal of the Acoustical Society of America*, vol. 127, pp. 2347–2357, Apr. 2010.

[62] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE transactions on information theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[63] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on image processing*, vol. 1, no. 2, pp. 205–220, 1992.

[64] J. Lin and L. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of sound and vibration*, vol. 234, no. 1, pp. 135–148, 2000.

[65] E. B. Saff and A. B. Kuijlaars, "Distributing many points on a sphere," *The mathematical intelligencer*, vol. 19, no. 1, pp. 5–11, 1997.

[66] F. Klein, "Lectures on the icosahedron," *London, Kegan Paul*, 1913.

[67] S. Hu, J. Trevino, C. Salvador, S. Sakamoto, J. Li, and Y. Suzuki, "A local representation of the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. EL285–EL290, 2016.

[68] A. P. Bates, Z. Khalid, and R. A. Kennedy, "On the use of slepian functions for the reconstruction of the head-related transfer function on the sphere," in *Signal Processing and Communication Systems (ICSPCS), 2015 9th International Conference on*, pp. 1–7, IEEE, 2015.

[69] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.

[70] G. Kaiser, *A friendly guide to wavelets*. Springer Science & Business Media, 2010.

[71] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[72] I. Bogdanova, P. Vandergheynst, J.-P. Antoine, L. Jacques, and M. Morvidone, "Stereographic wavelet frames on the sphere," *Applied and Computational Harmonic Analysis*, vol. 19, no. 2, pp. 223–252, 2005.

[73] J. D. McEwen, M. P. Hobson, D. J. Mortlock, and A. N. Lasenby, "Fast directional continuous spherical wavelet transform algorithms," *IEEE transactions on Signal Processing*, vol. 55, no. 2, pp. 520–529, 2007.

[74] Z. Khalid, R. A. Kennedy, S. Durrani, P. Sadeghi, Y. Wiaux, and J. D. McEwen, "Fast directional spatially localized spherical harmonic transform," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2192–2203, 2013.

[75] M. Holschneider, "Continuous wavelet transforms on the sphere," *Journal of Mathematical Physics*, vol. 37, no. 8, pp. 4156–4165, 1996.

[76] J.-P. Antoine and P. Vandergheynst, "Wavelets on the 2-sphere: A group-theoretical approach," *Applied and Computational Harmonic Analysis*, vol. 7, no. 3, pp. 262–291, 1999.

[77] M. Holschneider, A. Chambodut, and M. Mandea, "From global to regional analysis of the magnetic field on the sphere using wavelet frames," *Physics of the Earth and Planetary Interiors*, vol. 135, no. 2-3, pp. 107–124, 2003.

[78] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM journal on mathematical analysis*, vol. 29, no. 2, pp. 511–546, 1998.

[79] P. Schröder and W. Sweldens, "Spherical wavelets: Efficiently representing functions on the sphere," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 161–172, ACM, 1995.

[80] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Applied and computational harmonic analysis*, vol. 3, no. 2, pp. 186–200, 1996.

[81] P. Schroder, "Wavelets in computer graphics," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 615–625, 1996.

[82] P. Yu, P. E. Grant, Y. Qi, X. Han, F. Ségonne, R. Pienaar, E. Busa, J. Pacheco, N. Makris, R. L. Buckner, *et al.*, "Cortical surface shape analysis based on spherical wavelets," *IEEE transactions on medical imaging*, vol. 26, no. 4, pp. 582–597, 2007.

[83] H. Laga, H. Takahashi, and M. Nakajima, "Spherical wavelet descriptors for content-based 3d model retrieval," in *Shape Modeling and Applications, 2006. SMI 2006. IEEE International Conference on*, pp. 15–15, IEEE, 2006.

[84] N. Dyn, D. Levine, and J. A. Gregory, "A butterfly subdivision scheme for surface interpolation with tension control," *ACM transactions on Graphics (TOG)*, vol. 9, no. 2, pp. 160–169, 1990.

[85] H. S. Colburn and A. Kulkarni, "Models of sound localization," in *Sound source localization*, pp. 272–316, Springer, 2005.

[86] S. Sekimoto, R. Ogasawara, Y. Iwaya, Y. Suzuki, and S. Takane, "Numerical investigation of effects of head sizes and ear positions on head-related transfer functions," in *Proc. of The Japan-China Joint Conference of Acoustics 2007*, vol. 6, 2007.

[87] M. OTANI, Y. IWAYA, T. MAGARIYACHI, and Y. SUZUKI, "Spatial distribution of the low-frequency head-related transfer function spectral notch and its effect on sound localization," in *Principles And Applications Of Spatial Hearing*, pp. 195–204, World Scientific, 2011.

[88] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Boundary matching filters for spherical microphone and loudspeaker arrays," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 3, pp. 461–474, 2018.

[89] S. Hu, J. Trevino, C. Salvador, S. Sakamoto, and Y. Suzuki, "Modeling head-related transfer functions with spherical wavelets," *Applied Acoustics*, vol. 146, pp. 81–88, 2019.

[90] R. M. Stern, A. S. Zeiberg, and C. Trahiotis, "Lateralization of complex binaural stimuli: A weighted-image model," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 156–165, 1988.

[91] K. E. Hancock and B. Delgutte, "A physiologically based model of interaural time difference discrimination," *Journal of Neuroscience*, vol. 24, no. 32, pp. 7110–7117, 2004.

[92] T. M. Shackleton, R. Meddis, and M. J. Hewitt, "Across frequency integration in a model of lateralization," *The Journal of the Acoustical Society of America*, vol. 91, no. 4, pp. 2276–2279, 1992.

[93] D. J. Kistler and F. 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*, vol. 91, no. 3, pp. 1637–1647, 1992.

[94] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[95] L. A. Abbagnaro, B. B. Bauer, and E. L. Torick, "Measurements of diffraction and interaural delay of a progressive sound wave caused by the human head. ii," *The Journal of the Acoustical Society of America*, vol. 58, no. 3, pp. 693–700, 1975.

[96] D. R. Begault and L. J. Trejo, "3-d sound for virtual reality and multimedia," 2000.

[97] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1480–1492, 1999.

[98] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002.

[99] E. H. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002.

[100] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," in *Audio Engineering Society Convention 98*, Audio Engineering Society, 1995.

[101] J. Huopaniemi and J. O. Smith III, "Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters," in *Audio Engineering*

Society Conference: 16th International Conference: Spatial Sound Reproduction, Audio Engineering Society, 1999.

[102] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller, "The interaural time difference in binaural synthesis," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[103] B. F. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, 2014.

[104] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, no. 5, pp. 592–605, 2011.

[105] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 791–802, 2014.

[106] A. Josupeit and V. Hohmann, "Modeling speech localization, talker identification, and word recognition in a multi-talker setting," *The Journal of the Acoustical Society of America*, vol. 142, no. 1, pp. 35–54, 2017.

[107] M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ild and itd," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 68–77, 2010.

[108] J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, and H. Peremans, "An ideal-observer model of human sound localization," *Biological cybernetics*, vol. 108, no. 2, pp. 169–181, 2014.

[109] J. E. Mossop and J. F. Culling, "Lateralization of large interaural delays," *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1574–1579, 1998.

[110] V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 433–442, 2002.

[111] B. C. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *The journal of the acoustical society of America*, vol. 74, no. 3, pp. 750–753, 1983.

[112] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4087–4096, 2017.

List of works

Journal papers

1) Shichao Hu, Jorge Trevino, César Salvador, Shuichi Sakamoto, and Yôiti Suzuki, Modeling head-related transfer functions with spherical wavelets, Applied Acoustics. 146: 81-88. 2019

2) Shichao Hu, Jorge Trevino, César Salvador, Shuichi Sakamoto, Junfeng Li, and Yôiti Suzuki, A local representation of the head-related transfer function, Journal of Acoustical Society of America, 140(3), pp.EL285-EL290, 2016

International Conferences

3) Jorge Trevino, Shichao Hu, César Salvador, Shuichi Sakamoto, Junfeng Li, and Yôiti Suzuki, A compact representation of the head-related transfer function inspired by the wavelet transform on the sphere, IIH-MSP,IEEE, Sep.,2015, Australia.

Domestic Conferences

4) Shichao Hu, Jorge Trevino, César Salvador, Shuichi Sakamoto, Junfeng Li, and Yôiti Suzuki, A compact representation of head-related transfer functions using wavelets on the sphere. Proceedings of the auditory research meeting, Acoustical Society of Japan, Aug., 2015