Binaural Synthesis based on the Spherical Harmonic Analysis with Compact Microphone Arrays

by

César Daniel SALVADOR CASTAÑEDA

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

Evaluation Committee

Prof. Yôiti SUZUKI

Prof. Yoshifumi KITAMURA

Prof. Akinori ITO

Prof. Shuichi SAKAMOTO

Tohoku University
September 2013

Dedicated to my family,

especially my parents Rosa and Manuel,

and my sisters Noemi and Melissa.

Dedicated also to my comrades in my hometown,

Trujillo.

Preface

Auditory localization cues play a crucial role in everyday life. They increase awareness of surroundings, improve visual attention, and convey complex information about the spatial structure of an acoustic scene. People in music and cinematography have long recognized the aesthetic importance of the spatiality of sound to evoke ambience and focused attention, thus motivating challenging projects of architectural acoustics and also studies on human spatial hearing. It was during the last decades, thanks to the progress on the synthesis of binaural localization cues and the development of multichannel audio techniques, when 3D audio systems started to emerge, covering demanding applications on telecommunications, such as the transmission of auditory scenes and the generation of virtual acoustic spaces. Hence the motivation of this study stems from the growing demand of 3D audio systems for the re-creation of auditory scenes with high levels of realism.

Among 3D audio systems, binaural synthesis aims to reproduce an auditory scene with high levels of realism by endowing the playback stage with spatial hearing information. Basic perceptual cues for a spatial listening experience arise from the scattering, reflections and resonances introduced by the pinnae, head and torso of the listener. These phenomena can be described by the so-called head-related transfer functions (HRTFs). Typical measured sets of HRTFs, though, neither include the motion of the head nor provide the listener with enough spatial accuracy, characteristics that are required, for example, in the accurate synthesis of moving sound sources. Several approaches that sidestep these two limitations have been proposed. They are based on the recordings made with microphones placed on the surface of a rigid sphere, and the angular interpolation of HRTFs from a representative set of sound sources in the distal region (beyond one meter distance from the head). However, the optimal arrangement of the representative sound sources, and the binaural synthesis of sources in the proximal region (less than one meter distance from the head) has not yet been addressed.

Therefore, this thesis provides a novel method to synthesize the binaural signals for sound sources in both the distal and proximal regions, starting from a sound field captured by spherical microphone arrays. The present proposal exploits the directional structure of the captured sound pressure field, analyzed by means of spherical harmonic functions with high directivity (high order). To make it applicable for the practical applications, the improvement and evaluation of the proposed method are carried out by computer simulations.

Chapter 1 introduces the motivation of this study. A survey on related systems for the synthesis of binaural signals by spherical microphone arrays is carried out, highlighting the use of the spherical harmonic functions for the analysis of the captured sound pressure field. Lastly, the objective of this thesis is presented.

Chapter 2 introduces an ideal scenario for the binaural synthesis of sound sources in the distal region, assuming the sound field is captured by an infinite number of microphones, so as to focus the discussion on the required order of the spherical harmonic functions to cover the full audible frequency range.

Chapter 3 continues the evaluation in the ideal scenario, now to discuss about the optimal arrangement of the representative sound sources for an accurate synthesis. A practical scenario for the binaural synthesis is subsequently introduced, where the synthesis accuracy for distal sound sources is evaluated assuming a finite number of microphones.

Chapter 4 proposes an extension of the method described in Chapter 3 for the binaural synthesis of sound sources in the proximal region. The recorded sound field is now analyzed with multipoles, an extension of spherical harmonics that allows the radial propagation of the sound field.

Finally, the main conclusions of this thesis are given on chapter 5.

Acknowledgements

This thesis would not have been possible without the cooperation and understanding of many people. First of all, I would like to express my gratitude to Professor Yôiti Suzuki, my supervisor, who gave me the opportunity of working for a master degree and guided me in the right direction. I am deeply grateful to Prof. Yoshifumi Kitamura, Prof. Akinori Ito and Prof. Shuichi Sakamoto for their participation on my dissertation committee and for their excellent work and teaching skill. Prof. Shuichi Sakamoto offered a lot of helpful advice throughout the completion of this project, I wish therefore to acknowledge his constant support.

I wish to acknowledge Prof. Shiori Satoshi, Prof. Kuniki Ichiro, Prof. Matsumiya Kasumishi, and all members of the Graduate School Seminar for their valuable comments. Likewise, I express my acknowledgment to Prof. Tomoko Ohtani, Eng. Saito Fumitaka, Dr. Cui Zhenglie, and all the members of the Advance Acoustic Information Laboratory (Prof. Suzuki's Laboratory), for their hard work in evaluating the present research and offering valuable feedback. I extend my gratitude to Jorge Treviño, one of our doctoral students, who kept an eye on my work and gave good suggestions and comments, and Yoshiki Sato, my classmate, who gave me helpful comments and had exchange of views for a master degree. I am grateful to Ms. Miki Onodera, our staff, who gave me valuable help for studying and living in Japan.

I wish to express my gratitude to the Ministry of Education, Culture, Sports, Science and Technology of Japan, whose financial support enabled me to continuing my studies, at

a graduate level, in Japan. In the same way, I express my gratitude to the Embassy of Japan in Peru, whose valuable orientation during the scholarship application period in 2010 made it possible to prepare myself for this new challenge of study in Japan.

I extend my gratitude to all my Professors and mates in the Faculty of Electronic Engineering at Pontifical Catholic University of Peru, where I studied the undergraduate and got the inspiration to follow a career on research. I remind with pleasure as well my last job in ISONAR Sound Research Group at University of San Martin de Porres, where I began to appreciate soundscapes in the sense of acoustic ecology, my gratitude therefore to all my colleagues at ISONAR for so enriching experience.

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 comrades and my family in my hometown. It all started back there in Trujillo, where I grew up surrounded by the most honorable people I have ever known. And most important, words can hardly express my deepest appreciation to my inspiration: My father Manuel, my mother Rosa, and my sisters Noemi and Melissa, who provided me with their endless support.

Contents

Chapter 1

Introduction

1.1 Conventional binaural systems

It is quite remarkable how well humans can separate and selectively attend to individual sound sources in a cluttered acoustical environment. The term "cocktail party processing" was coined in an early study on how the perception of sound with both ears enables to selectively attend to individual conversations when many are present, as in a cocktail party [1]. This phenomenon illustrates the importance of binaural hearing to localize and separate sound sources. Although the understanding of such complex phenomena involves cognitive processing in the auditory pathway, the externals facts of the perception of sound sources in the real world can be simplified and interpreted as signals that are acoustically filtered by the pinna, head and torso of the listener. The complex shapes of the human body cause scattering, reflections and resonances on the arriving sound wave. These acoustic interactions modify the sound pressure field at the entrance of the ear canal, giving rise to differences between the signals at both ears: the interaural time difference (ITD) and the interaural level difference (ILD), which are the primary perceptual cues for the localization of sound [2, 3].

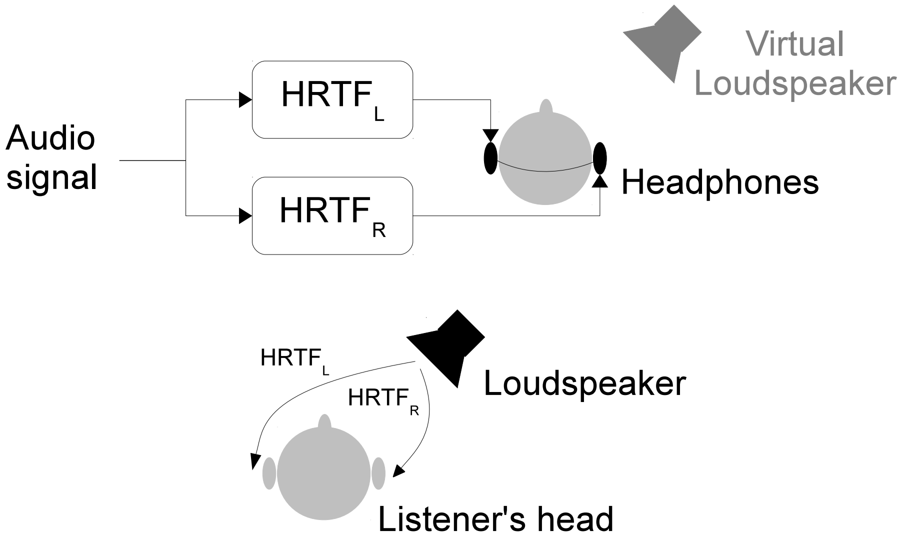


Figure 1.1: A conventional binaural system. An audio signal is filtered with the head-related transfer functions (HRTFs) for the left and right ears. The HRTFs are previously measured with two microphones placed on each listener's ear, for a loudspeaker in a fixed position emitting signals in the full audible frequency range. The reproduction is done over headphones. Ideally, the listener perceives the audio signal as if it was comming from a virtual loudspeaker, at the same position of the loudspeaker used to measure the HRTFs.

The external facts of human spatial hearing have inspired a sound reproduction system for which sound is recorded by using two microphones mounted at the ears of a listener. In an ideal binaural transmission system, both the amplitude and phase of the sound waves incident on the listener's ears are duplicated. A natural extension of this method is to characterize a listener's head beforehand, in the full-audible frequency range, using a loudspeaker at a given position, so as to compute the left and right filters that contains the acoustic transmission path from the given source position to both ears. Such filters conform the so-called head-related transfer functions (HRTFs), which can be used to filter an arbitrary audio signal so as to synthesize the binaural signals. When reproduced over headphones, the binaural signals give the impression of a sound source arriving from a virtual loudspeaker at the given location (see Figure 1.1).

In practice, the synthesis of the binaural signals, produced by an arbitrary sound source at a particular location, requires the HRTFs for a representative set of sound source positions, and the equivalent sound pressure field at the representative positions due to the source signal. The sound pressure signals are filtered with the representative HRTFs at the corresponding positions, and subsequently combined to produce the desired left and right ear signals. The resulting binaural signals can thus be reproduced through headphones (see Figure 1.2) or loudspeakers (see Figure 1.3). When the reproduction is done through real headphones, the sound field is characterized as coming from a set of virtual loudspeakers surrounding the listener. On the other hand, when an array of real loudspeakers are used during the reproduction, the synthesized binaural signals are understood as a pair of virtual headphones. Both approaches rely on the accurate acquisition of both the sound pressure field and the representative set of head-related transfer functions (HRTFs). The scope of the present proposal is limited to the virtual loudspeaker approach.

1.2 Head-related transfer functions

The head-related transfer functions (HRTFs) characterizes how the external ears receive a sound from a point space. They depend on the shape of the pinna, head and torso of each individual human listener, and on the frequency and position of the sound sources surrounding the listener. The geometry for the computation of the HRTFs using an artificial head is shown in Figure 1.4. The left ear HRTFs for sources on the horizontal plane at a 1.5 m distance are shown on Figure 1.5. In general, high intensities appear for sound sources on the same side of the ear (the ipsilateral side), and lower intensities for sound sources on the opposite side of the ear (the contralateral side). A detailed view shows that the HRTFs contain the auditory spectral cues that arise from the shadowing, diffractions and reflections of the sound waves due to the pinna, head and torso. Therefore, the HRTFs characterize the binaural localization of sound sources.

The HRTFs must preferably be measured for each individual listener, but they can also be measured using an artificial head and torso, namely a dummyhead. In both cases, typical measurements only considere a reduced number of sound source positions surrounding the listener's head. More recently, numerical solutions from a 3D model of the torso and

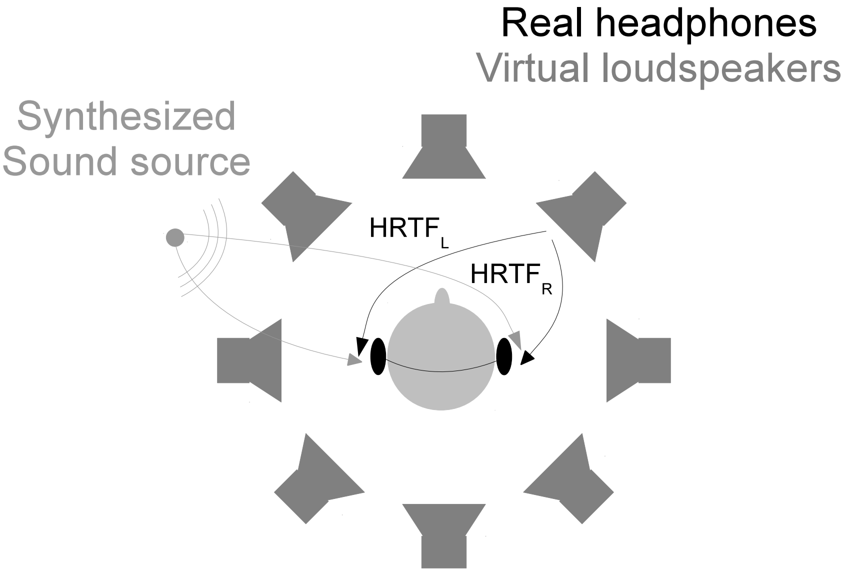


Figure 1.2: Binaural rendering based on weighted sums of filtered versions of an audio signal with a set of HRTFs. The HRTFs are measured for a representative set of sound source positions surrounding the listener, where an array of virtual loudspeakers is assumed to be placed. The binaural signals for an arbitrary position are synthesized by adjusting the complex weights (gains and delays) for each virtual loudspeaker. The reproduction is finally done over headphones. Our proposal follows this scheme: the so-called virtual loudspeaker approach.

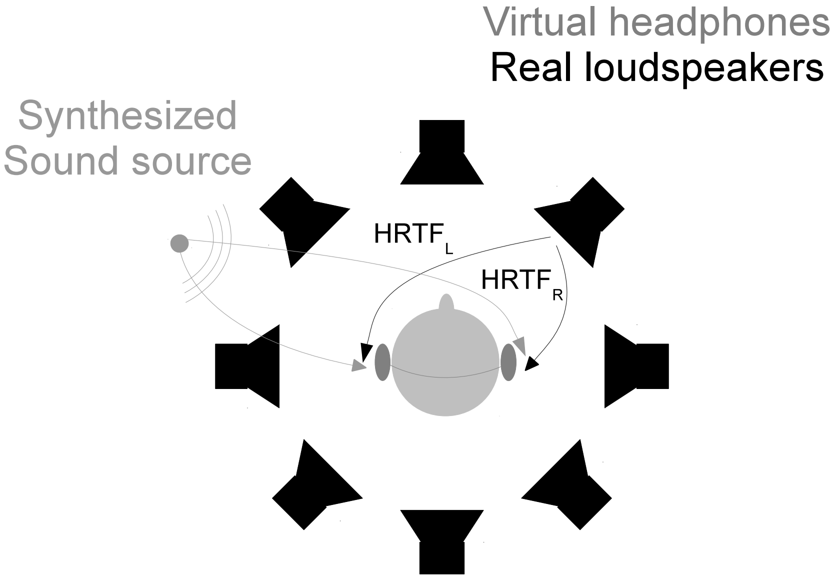


Figure 1.3: Binaural rendering based on weighted sums of filtered versions of an audio signal with a set of HRTFs. The HRTFs are measured for a representative set of sound source positions surrounding the listener, where a loudspeaker array is placed for reproduction. Based on crosstalk canceling techniques, the binaural signals for an arbitrary position are synthesized by adjusting the complex weights (gains and delays) for each loudspeaker.

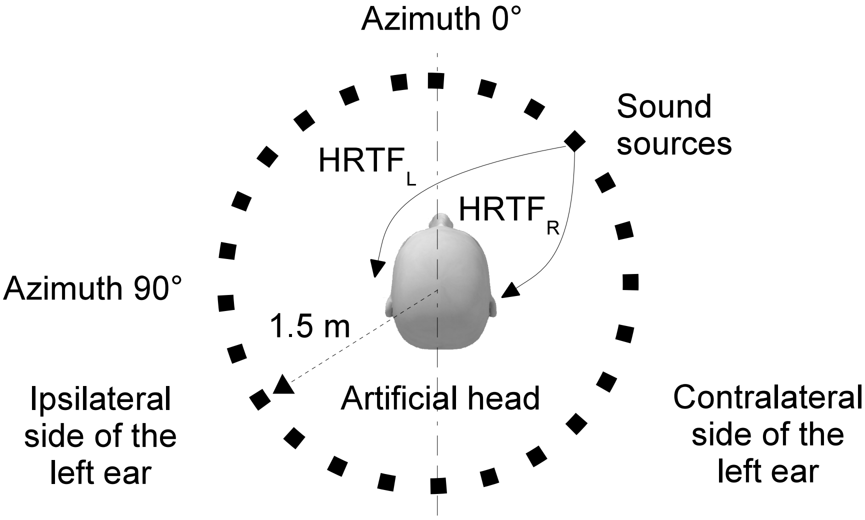


Figure 1.4: Top view of the geometry for the measurement of HRTFs for the left and right ears using an artificial head. The sound sources are equiangularly distributed in the horizontal plane. Assuming high levels of symmetry the discussion is limited to the left ear. Sound sources on azimuths between 0 and 180 degrees are said to lie on the ipsilateral side, and the ones on azimuths between 180 and 360 degrees, on the contralateral side.

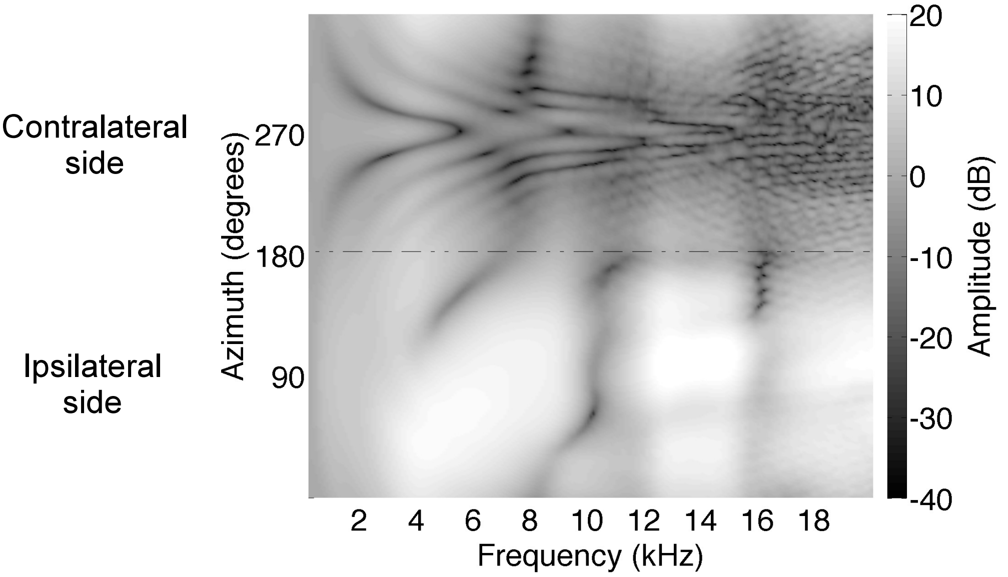


Figure 1.5: HRTFs for the left ear of an artificial head. The HRTFs depend on the shape of the head, and on the frequency and position of the sound sources. The sources are on the horizontal plane and cover the full audible frequency range, from 20 Hz to 20 kHz. In general, ipsilateral sources are received with high intensities at the ear canal, while the contralateral ones are shadowed and scattered by the head. Resonances due to pinna cause a notch pattern around 10 kHz on the ipsilateral side.

head based on the boundary element method (BEM) [4]. However, even numerical methods consider only fixed positions and orientations of the listener's head. Moreover, typical measured sets of HRTFs do not show good accuracy for low-frequency sounds, do not consider the mobility of the head, and only provide information for a limited range of sound source positions. The spatial perception of the synthesized auditory scene is thus significantly affected by these constraints, which generate reversals of sound direction between front and back, excessive elevation, and in head localization of sound sources during the binaural reproduction.

Efforts on these directions have proposed the approximation of the head by simple shapes [5–8]. The most simple model is based on a spherical scatterer of the size of an average human head, in which two microphones are placed at the ends of a horizontal diameter. Experiments on the synchronous movement of the spherical model and listener's head have shown to reduce the reversals of sound direction between front and back, and the excessive elevation, specially for low-frequency sound sources [5]. This approach have inspired sound recording systems based on several microphones placed on the spherical scatterer [9–12], so as to include the mobility of the head. On the other hand, several methods to interpolate HRTFs for not available sound souce positions have also been proposed [13–21], being the spherical harmonic decomposition a promising approach that provides scalable spatial resolution and rotation capabilities [18–21].

Binaural cues for the localization of sound sources

The primary binaural cue for directional perception are the interaural time difference (ITD) and the interaural intensity difference (IID) [2,3]. They determine the lateral angle of a given sound source to the one side or the other. With increasing ITD the perceived sound is located more and more to the side where the stimulation occur first, and with increasing IID, the sound is said to appear further and further toward the side of the stronger stimulation [22, 23]. The ITD and IID are therefore the primary cues for distal sound sources localization. However, for a certain direction, the primary cues determine a circular locus of possible directions, arising the so-called cones of confusion containing the positions with equals ITD and IID. Reversals of sound direction between front and back results from this ambiguity. This effect is called the front/back confusion, referring to the perception of sounds in the back as coming from the front, and the way around.

The primary cue for distance perception is the intensity. In general, intensity decreases as the distance between the sound source and the listener is increased. For a fixed-power point-source in an acoustic free-field, a 6dB loss in sound pressure for every doubling distance. The direct-to-reverberant energy ratio is another cue for environments with sound reflecting surfaces. The ratio of energy reaching a listener via one or more of the reflecting surfaces is inversely related to the distance of the sound source. Recent studies have suggested that listener might also be able to use binaural cues to determine the distances of lateral sound sources by comparing the ITD and IID cues. Although there is some ambiguity about the utility of binaural cues for the distance perception of distal sound sources, there is ample evidence that binaural cues play an important role on the perception of sound sources near the head [24–30].

Therefore, the binaural cues alone can be used to determine both the directions and distances of nearby sound sources, given that listeners could use the distant-invariant ITD to determine the lateral position of the source, and then use the magnitude of the ILD to estimate its distance [24]. Moreover, given that the HRTFs contain the information to compute both ITD and IID, the HRTFs provide a suitable tool to study of the binaural localization of sound.

Models of HRTFs based on orthonormal functions

The limitations that arises from the measurement of HRTFs are motivating the developement of analytical methods to synthesize models for the HRTFs. The expected characteristic of such models are the following:

- The parameters of the model should contain the information related to each indivual.

- The mobility of the head should be considered by means of easy rotations.

- The synthesis of moving sound should be performed by easy interpolations.

Attempts have been made to only measure HRTF sets for a limited range of source positions and to interpolate HRTFs for positions in between. A psychoacoustically motivated approach exploits the limitations of the binaural hearing system to reduce the amount of information to describe HRTFs. The parameters that characterize the HRTFs are extracted by a set of perceptually motivated basis functions. The basis functions form a set of band pass filters that mimic the known spectral limitations of the human auditory system. However, this simple parametric approach has the risk of deteriorating the externalization and cannot be generalized to arbitrary HRTF sets [31].

Alternatively, the HRTFs can be approximated by weighted sums of a finite number I of orthonormal functions f_i as follows

$$H = \sum_{i=1}^{I} c_i f_i + e, \tag{1.1}$$

where the weighting coefficients c_i, computed by minimizing the error e in a least squares sense, are the projection of H on the orthormal basis:

$$c_i = \int H f_i^*. \tag{1.2}$$

A statistical data compression approach aims to perform the approximation in terms of orthonormal eigentransfer functions, which are extracted from the covariance matrix of the HRTFs. Implementations of this approach relies on the principal component analysis [13] and the Karhunen-Loeve expansion [14] have been proposed. A digital filter design approach aims to approximate the HRTFs in terms of rational functions on the z-plane. The use of the pole zero approximation [15] and the common-acoustical pole zero approximation [16, 17] have been proposed to implement this approach. More rencently, the spherical acoustics approach aims to approximate the HRTFs in terms of functions

Table 1.1: Models of HRTFs based on orthonormal functions.

Method	**The orthonormal functions depend on:**	**The weights depend on:**	**Characteristic**
Eigen functions [13, 14].	• Individual. • Frequency.	• Direction.	Reduced dimension.
Ratonal functions [15, 16].	• Frequency.	• Individual. • Direction.	Reduced computational complexity.
Spherical functions [18–20].	• Position.	• Individual. • Frequency.	Scalable spatial resolution.

on the unit sphere. Implementations of this latter approach relies on the use of spherical harmonics [18], multipoles [19], and Fourier spherical Bessel functions [20]. Table 1.1 highlights the characteristics of each approach.

Although there is a risk that the basis functions that are very important in terms of the least-squares error of the fit are not so relevant in terms of human auditory perception, the preservation of all the information contained in the HRTFs is preferable instead of the psycho-acoustically motivated approach that reduces the amount of information. In particular, the spherical harmonics has been selected to be the orthonormal functions for this proposal, because they fully encode the position information with scalable spatial resolution, providing an encoding format compatible with sound field reproduction techniques based on loudspeaker arrays such as wave field synthesis [32, 33] and high-order ambisonics [34].

1.3 Sound recording systems for binaural synthesis

This section briefly describes three spatial sound recording systems that aim to synthesize the binaural signals from the recordings made with a spherical microphone array.

1.3.1 Motion-tracked binaural system

The Motion-tracked binaural system (MTB) [9] is based on a structural model of the human head, for which the size of the head is approximated by a rigid sphere and the shape of the pinna by a notch filter. This simple model allows to consider the mobility of the head by arranging microphones on the surface of the rigid sphere. Three spatial sampling schemes are proposed for the arrangement of microphones depending on the spatial grouping of the original sound sources: equiangularly along the equator for panoramic sound sources, closely spaced along the sides of the sphere for frontal sound sources, and uniformly spaced over the sphere for omnidirectional sound sources. In addition, a head-tracking system allows to select the closest sphere's point to the ear position, where, if necessary, linear interpolation on the microphone signals is performed only on the low frequencies. [1]

A realization of MTB uses 16 microphones for panoramic applications and a spherical scatterer of 8.75 cm radius, which the authors considered to be the traditional radius of the first order approximation of an average human head. The customization to individual listeners is particularly addressed for the effects of different head sizes and pinna shapes. When the radius a of the scatterer differs significantly from the radius b of the listener's head, the apparent locations of the sound sources shift systematically with head motion. For sound sources in front of the listener, the disturbance can be reduced by simply scaling the measured head rotation angle by the scale factor b/a. Moreover, an adjustable simulated pinna notch filter was used to compensate for the front/back confusion and the excessive elevation experienced as consequence of the lack of pinna cues. The limitation of MTB is the degraded perceived realism of the synthesized binaural signals when comparing with systems based on head-related transfer functions.

[1] This interpolation method exploits the Rayleigh's duplex theory of sound localization [2,3], which states that the interaural time difference (ITD) is the dominant cue for frequencies at 1.5 kHz or below and the interaural level difference (ILD) is more dominant for frequencies greater than 1.5 kHZ.

1.3.2 SENZI

SENZI [10, 11] is a system to acquire 3D sound-space information from a rigid spherical microphone array. Its name is an acronym for Symmetrical object with ENchansed ZIllion microphones. In SENZI, the microphone signals are simply weighted and summed to synthesize the listener's binaural signals. The microphone's weights are the solution to a linear least-squares problem. The matrix coefficients of the linear system correspond to spatial samplings of the inverse model of the acoustic scattering from the rigid sphere [6], and its constant terms correspond to HRTFs for a representative of distal sound sources.

An implementation of SENZI uses 252 digital microphones placed on the surface of a rigid sphere of 8.5 cm radius, which the authors considered to be the radius of an average human head, and 2562 representative HRTFs or controlled directions. Both, the positions of the microphones and the representative sound sources, have been decided according to spherical grids based on the subdivisions of the faces of the icosahedron. Audio processing is performed on the frequency domain, and therefore, the microphone's weights needs to be computed by solving an overdetermined linear system for each frequency. Moreover, a new set of weights needs to be computed for each individual listener.

1.3.3 Binaural synthesis by spherical harmonics

The high-order binaural head-tracked system [12] synthesizes binaural signals based on angular interpolations performed by the spherical harmonic decomposition (See Figure 1.6). Starting from the wave nature of sound, the equivalent sound pressure field is derived from the sound pressure field captured by a spherical microphone array [35, 36], and therefore the virtual loudspeakers approach [37, 38] is used to downmix the binaural signals (See Figure 1.7). The sound field is calculated assuming plane-wave like sound sources. The set of recorded signals is first decomposed into its directional components by means of the spherical harmonic functions. Subsequently, the sound field is reconstructed on the directions for which the representative distal HRTFs have been measured.

The representative sound sources on the distal region are interpreted as being rendered by a virtual array of loudspeakers, whose signals are filtered with the HRTFs and add up so as to synthesize the binaural signals.

An existing implementation of this system uses 60 microphones placed on a rigid sphere of 10.1 cm radius. The microphones where distributed according to the Fliege quadrature nodes, which improve the accuracy of the numerical integration over the sphere. The sound arriving at the center of the head consists of a superposition of plane waves weighted by a representative set of HRTFs. Other implementations for the synthesis of distal HRTFs, based on the directional encoding of the sound field and the virtual loudspeaker approach, have also been implemented with techniques such as high-order ambisonics [38], spherical beamforming [39, 40], and wave field synthesis [41]. A benefit of these initiatives is the encoding of the sound field in a directional and scalable format, which can alternatively be reproduced by sound field reproduction techniques such as high-order ambisonics [34] and wave field synthesis [33]. However, there are not enough studies on binaural synthesis for proximal sound sources in this context.

Table 1.2: Existing systems for binaural synthesis based on spherical microphone arrays.

System	**Characteristic**	**Limitation**
Motion-tracked binaural system [9].	A simple scaling operation allows for the customization to individual listeners, particularly for the effects of different head sizes and pinna shapes.	The realism of the synthesized auditory scene is degraded due to the lack of head-related transfer functions during the binaural reproduction.
SENZI [10, 11].	The microphone signals are simply weighted and summed to synthesize binaural signals for distal sound sources.	A new set of microphone's weights needs to be computed for each individual.
High-order binaural head-tracked system [12].	The scalable encoding of the recorded sound field is compatible with other sound field reproduction techniques.	The binaural synthesis for proximal sound sources have not yet been addressed.

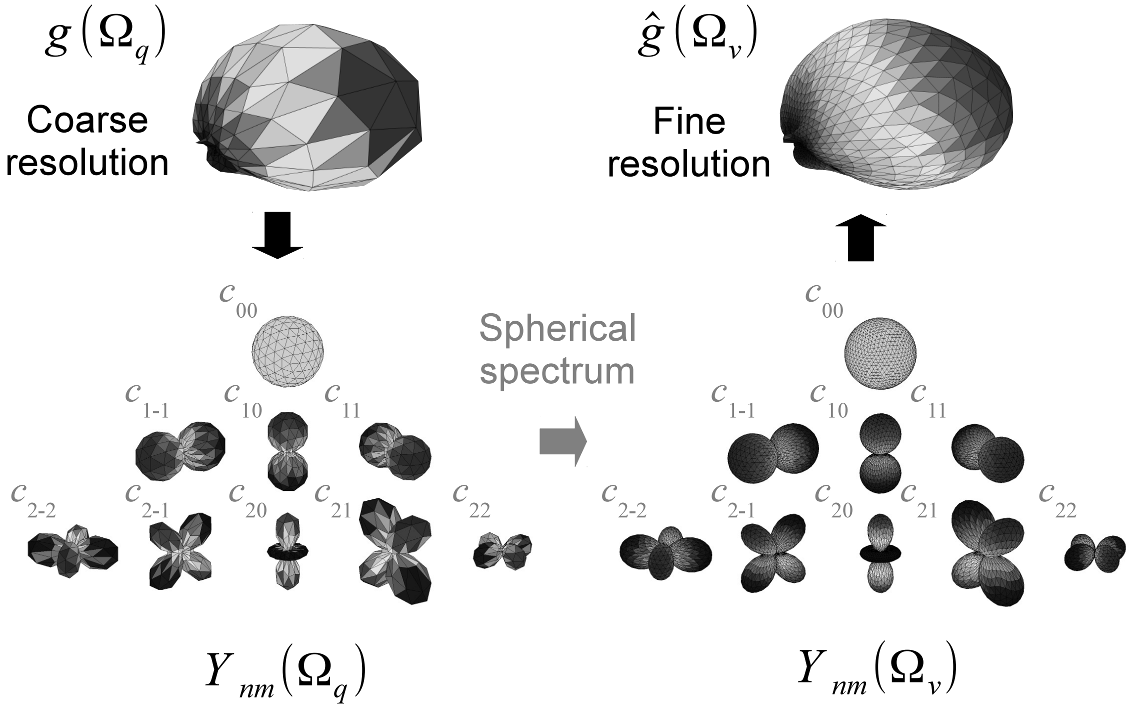


Figure 1.6: The spherical harmonic functions Y_{nm} allow to change the spatial resolution of a function g defined on a coarse sampling Ω_q of the unit sphere. The polar plots of Y_{nm}, with coarse (left) and fine (right) resolutions, are grouped in the bottom triangles. The spherical harmonics conform an orthonormal basis for the set of finite-energy signals on samplings of the unit sphere. Hence, the coarse resolution function $g(\Omega_q)$ can be decomposed into a weighted sum of spherical harmonics sampled with the same resolution. The weights c_{nm} define the spherical spectrum of g. The spectrum can subsequently be used to weight a different set of spherical harmonics defined on a finer sampling Ω_v, so as to reconstruct the function $\hat{g}(\Omega_v)$ with a higher spatial resolution. The converse path is also possible. The independence of Ω_q and Ω_v is exploited by the high-order ambisonics technique [34] to decouple the arrangement of transducers at the recording and reproduction stages.

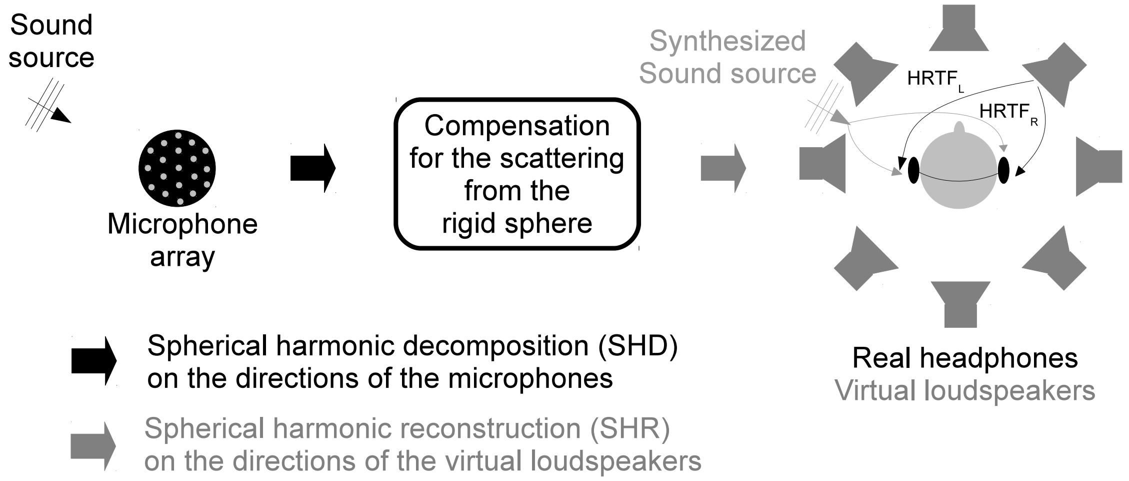

Sound source
Microphone array
Compensation for the scattering from the rigid sphere
Synthesized Sound source
HRTF$_L$
HRTF$_R$
Real headphones
Virtual loudspeakers
Spherical harmonic decomposition (SHD) on the directions of the microphones
Spherical harmonic reconstruction (SHR) on the directions of the virtual loudspeakers

Figure 1.7: Binaural synthesis by spherical harmonics. A sound source in the far field is represented by a plane wave. The sound field due to the plane wave is recorded by the rigid spherical microphone array. The directional structure of the sound field is characterized by its decomposition into spherical harmonics (*cf.* Figure 1.6). A set of filters to compensate for the scattering from the rigid sphere is applied to the spherical spectrum. The compensated spectrum is reconstructed at a set of representative directions, for which a set of HRTFs have been measured. Finally, the binaural rendering is based on the virtual loudspeakers approach (see Figure 1.2), whose weights correspond to the sound field due to the original sound source.

1.4 Objective

The sound field representation in terms of spherical harmonics provide a solid theory for binaural synthesis with the possibility to include the mobility of the head. Several solutions that aims to synthesize the binaural signals based on the this representation have been developed (See Table 1.3). Starting from a representative set of measured HRTFs, the synthesis of distal [18, 38, 41] and proximal [19, 21] sound sources have already been developed. Extending this approach, the binaural synthesis of distal sound sources from the recordings made with a rigid spherical microphone array has also been proposed [12, 41]. However, to full cover the 3D auditory space with a spherical microphone array, it is still necessary to address the binaural synthesis for proximal sound sources.

Therefore, the general objective of the method proposed in this thesis aims to synthesize the binaural signals for the three-dimensional auditory space, starting from the recordings made with a rigid spherical microphone array, and a set of representative HRTFs (see Figure 1.1). The novelty of this proposal is the full covering of the audible space along direction and distance, for which it is necessary to address the synthesis of the binaural signals for distal and proximal sound sources. Starting from the sound field captured by the microphone array, a set of beamformers need to be designed to weight the HRTFs, which had been previously measured or precomputed for a set of representative distal sound source positions. The optimal arrangement of the representative sound sources needs also to be considered. From the overview in Table 1.3, the general goal is subdivided onto the specific objectives organized in Table 1.4.

Table 1.3: Overview of binaural synthesis based on spherical harmonics.

Application	**Distal sources (> 1 m)**	**Proximal sources (< 1 m)**
Binaural synthesis from a set of measured HRTFs	Decomposition of the HRTFs in terms of spherical harmonics [18] and weighted sum of the HRTFs [38, 41].	Acoustic propagation of the HRTF spherical spectrum [19, 21].
Binaural synthesis from 3D recordings and a set of measured HRTFs	Plane wave decomposition via compact spherical microphone arrays [12, 41].	Method proposed on this thesis. Inspired on spherical beamforming [40, 42] and binaural synthesis from computer simulations of acoustic spaces [43].

Table 1.4: Specific objectives.

Application	**Distal sources (> 1 m)**	**Proximal sources (< 1 m)**
Binaural synthesis from a set of measured HRTFs	Chapter 2: To propose an unified theoretical model based on a continuous spherical recording surface.	Chapter 4: References to previous work.
Binaural synthesis from 3D recordings and a set of measured HRTFs	Chapter 3: To evaluate the effect of the spherical microphone array on the accuracy of the synthesis; and to set the minimum number of representative HRTFs for an accurate synthesis.	Chapter 4: To set the minimum distance for proximal sound sources that is possible to synthesize with the extended proposed method.

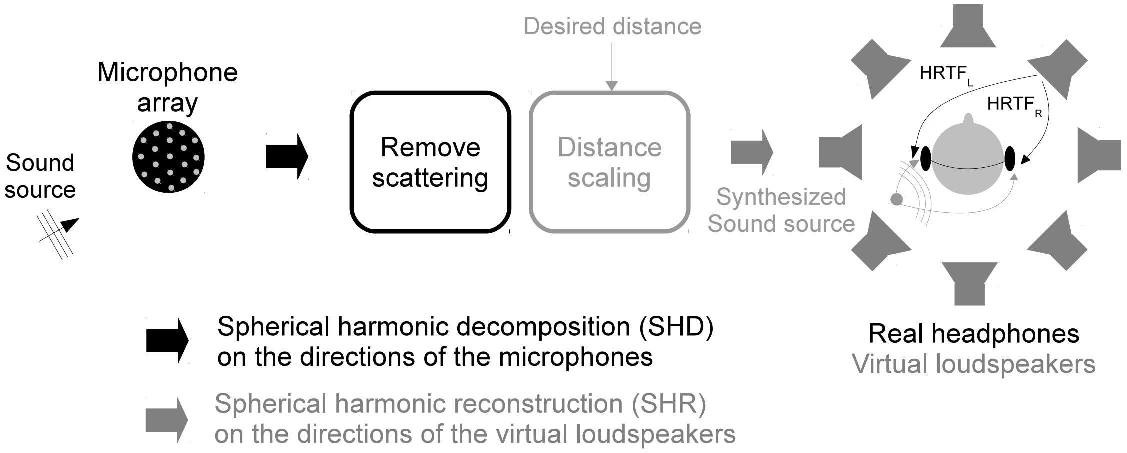

Desired distance
Microphone array
Sound source
Remove scattering
Distance scaling
Synthesized Sound source
HRTF_L
HRTF_R
Spherical harmonic decomposition (SHD) on the directions of the microphones
Spherical harmonic reconstruction (SHR) on the directions of the virtual loudspeakers
Real headphones
Virtual loudspeakers

Figure 1.8: Overview of the proposed method. The sound field due to a sound source in the far field is captured and analyzed with spherical harmonics functions for its binaural rendering based on the virtual loudspeaker approach (see Figure 1.7). We added a filtering stage in the spherical spectrum domain, intended to scale the radius of the virtual loudspeaker array so as to to match an inputted desired distance. Starting from the recordings of a sound source beyond 1 m distance from the center of the listener's head, made with a spherical microphone array, our proposal allows for the binaural synthesis of the sound source at arbitrary distances below 1 m.

Chapter 2

Binaural synthesis for distal sound sources by spherical harmonics

2.1 Introduction

This chapter introduces an ideal scenario to synthesize the binaural signals for sound sources located beyond 1 m distance from the listener head's center, the so-called distal sound sources. The synthesis is based on a weighted sum of HRTFs measured for a representative set of point-like surrounding sound sources. The weighting functions are derived from the spherical harmonic decomposition of the sound pressure field on the surface of a solid sphere. This ideal scenario, where the sound field is assumed to be captured by a rigid spherical array of infinite microphones, allows for the evaluation of spherical harmonics order effects and the arrangement of representative sound sources.

Section 2.2 briefly introduces the preliminary theory to synthesize the binaural signals for sound sources on the distal region. Section 2.3 describes the synthesis method based on the spherical harmonics analysis of the captured sound field. Section 2.4 shows how the decomposition order of the spherical harmonics affects the accuracy of the binaural synthesis for sound sources in the horizontal plane. Lastly, Section 2.5 summarizes the contents of this chapter.

2.2 Preliminaries

This section introduces the theory underlying the binaural synthesis from a rigid spherical recording surface and the superposition of a representative set of HRTFs. The theory of spherical harmonic analysis is then covered, emphasizing its application to characterize the acoustic scattering from the rigid sphere and the analysis of sound pressure fields. In this context, the binaural synthesis based on the superposition of a representative set of HRTFs is lastly described.

2.2.1 Angular solutions to Laplace's equation: Spherical harmonics

A function ψ that is a solution to the Laplace's equation is said to be harmonic. The scalar form of Laplace's equation in the standard spherical coordinate system is the partial differential equation [44]

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

where r is the radial distance, $\theta\in[0,\pi]$ the inclination angle, and $\phi\in[0,2\pi]$ the azimuth angle.

Assuming ψ is constant along the radial distance r, the harmonic angular solutions to Eq. (2.1) can be written $\psi=\Theta_{nm}(\theta)\Phi_m(\phi)$. Thus, separation of variables leads to two ordinary differential equations

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

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

The solution to Eq. (2.2) is found by the transformation of variables $\zeta = \cos\theta$, which leads to the Legendre differential equation. The solution to Eq. (2.3) is simply a complex exponential function of ϕ. Both angular solutions are conveniently combined into a single function $Y_{nm} = \Theta_{nm}\Phi_m$, called the spherical harmonic function of order n and degree m, defined by [44]

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

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

The associated Legendre function is defined by [44]

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

for $\zeta \in [-1, 1]$, and for positive index n and m. The extension to negative index is computed with the following two properties

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

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

The normalization factor 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

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

The spherical harmonics up to oder 2, defined on a continuous sphere of unit radius, are depicted in Figure 2.1. In these polar plots, the radius corresponds to the magnitude of the spherical harmonics. The orthonormality property described by Eq. (2.10) is illustrated by Figure 2.2 for the example in Figure 2.1.

Spherical harmonic transforms

From now in advance, the direction (θ, ϕ) will be abbreviated as $\Omega = (\theta, \phi)$. The integral on the surface of the unit sphere $\mathbb{S}^2$,

$$\int_{\Omega \in \mathbb{S}^2} d\Omega = \int_0^{2\pi} \int_0^{\pi} \sin\theta d\theta d\phi, \tag{2.9}$$

covers the entire surface of $\mathbb{S}^2$.

The spherical harmonics are orthonormal to each other:

$$\int_{\Omega \in \mathbb{S}^2} Y_{nm}(\Omega) Y^*_{n'm'}(\Omega) d\Omega = \delta_{n-n',m-m'}, \tag{2.10}$$

where * is the complex conjugate operator and $\delta_{i,j}$ the Kronecker Delta. They are also complete on $\mathbb{S}^2$:

$$\sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_{nm}(\Omega) Y^*_{nm}(\Omega') = \delta(\theta - \theta')\delta(\phi - \phi'), \tag{2.11}$$

where $\delta(x)$ is the Dirac's delta function. The spherical harmonics thus form a complete set of orthonomal functions, and therefore an orthonormal basis for the set of square-integrable functions on $\mathbb{S}^2$.

The following addition theorem for two different directions Ω_v and Ω_q also holds [45]:

$$\sum_{m=-n}^{n} Y_{nm}(\Omega_v) Y^*_{nm}(\Omega_q) = \frac{2n+1}{4\pi} P_n(\cos\Theta_{vq}), \tag{2.12}$$

where P_n is the Legendre function of Eq. (2.5) for $m = 0$, and Θ_{vq} is the angle between the directions Ω_v and Ω_q. It is useful to deal with relative directions.

The importance of the spherical harmonics rests in the fact that any square-integrable function $f(\Omega)$ on $\mathbb{S}^2$ can be expanded in terms of them as follows [45]:

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

where the constants f_{nm} define the spherical wave spectrum, a projection of $f(\Omega)$ into the

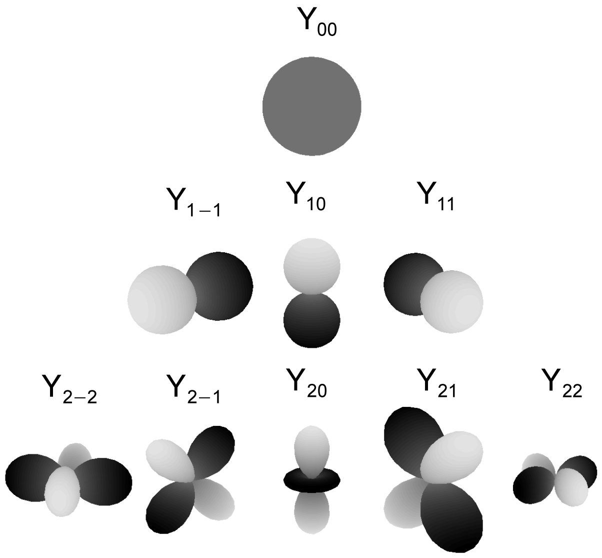


Figure 2.1: Polar plots of the magnitude of the spherical harmonics $|Y_{nm}(\Omega)|$, defined for all directions Ω in the unit sphere $\mathbb{S}^2$, for orders $n = 0, 1, 2$, and degrees $m = -n, ..., n$, respectively. The spherical harmonics form a complete set of orthonormal functions. Any square-integrable function on $\mathbb{S}^2$ can thus be expanded as a linear combination of spherical harmonics.

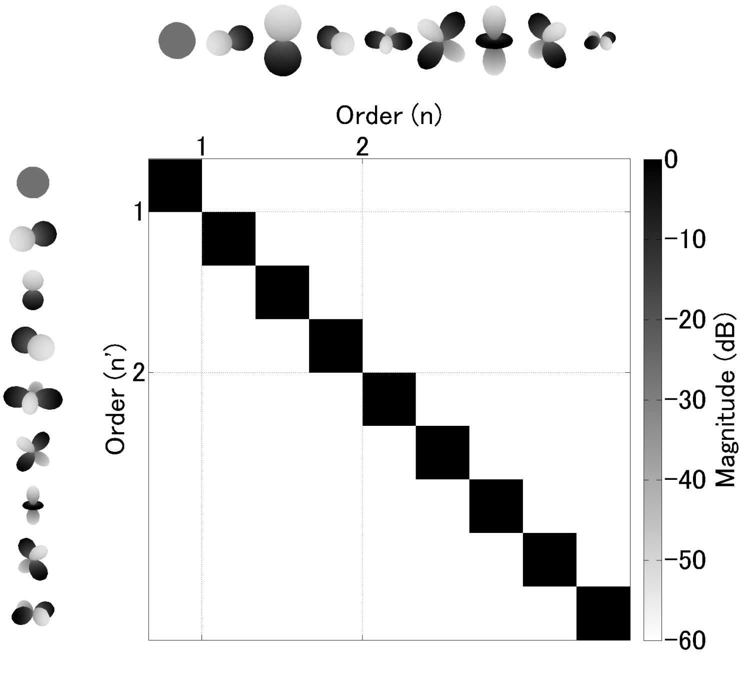


Figure 2.2: The spherical harmonics $Y_{nm}(\Omega)$ of order n and degree m are orthonormal to each other. Their inner product on the unit sphere, evaluated with Eq. (2.10), equals zero except when they coincide. This property does not hold on samplings of the unit sphere.

set of spherical harmonic functions computed as follows [45]:

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

Hence, Eqs. (2.13) and (2.14) are also referred to as the inverse spherical harmonic transform (ISHT) or decomposition (ISHD), and the spherical harmonic transform (SHT) or reconstruction (SHR), respectively.

2.2.2 Acoustic scattering from the rigid sphere

Let $\mathbf{a} = (a, \Omega')$ be a point on a rigid spherical measurement surface of radius a, and $\mathbf{b} = (b, \Omega)$ a point on a spherical radiating surface of radius b (see Figure 2.3). The presence of the rigid sphere of radius a diffracts the sound wave produced by the source at a point $\mathbf{b}$. The total pressure field $S(\mathbf{a}, \mathbf{b}, k)$ on the surface of the rigid sphere, due to a point-like source at $\mathbf{b}$ emiting sound of wave number k, is defined as a sum of the incident pressure in the free field due to the point-like sound source, and the scattered pressure due to the presence of a rigid sphere of radius a.

The interaction of sound with a rigid sphere is therefore characterized by [46]

$$S(\mathbf{a}, \mathbf{b}, k) = -\frac{1}{ka^2} \sum_{n=0}^{\infty} \frac{h_n(kb)}{h_n'(ka)} (2n+1) P_n(\cos\Theta), \quad b > a, \tag{2.15}$$

where Θ is the angle between the measurement point at $\mathbf{a}$ and the source at $\mathbf{b}$, P_n is the Legendre function of Eq. (2.5) for $m = 0$, h_n is the spherical Hankel function, and k is the wave number.

Figure 2.4 illustrates the geometry used to characterize the total pressure field described by Eq. (2.15). The measurement point is placed on the leftmost position of the surface of the rigid sphere. 360 sound sources at a 1.5 m distance from the center of the sphere, equiangularly distributed on the horizontal plane, were used. A microphone is placed on the leftmost position of the equator of a spherical scatterer of 8.5 cm radius. The total

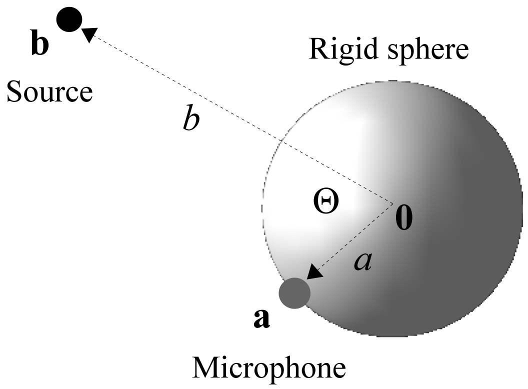


Figure 2.3: The geometry to model the acoustic scattering from the rigid sphere. The total pressure at a point $\mathbf{a} = (a, \Omega')$, on the surface of the rigid sphere centered at $\mathbf{0} = (0, 0, 0)$, equals the superposition of the incident pressure in the free field due to the sound source located at $\mathbf{b} = (b, \Omega)$, and the scattered pressure due to the presence of the rigid sphere.

pressure field characterized for sound sources in the full audible frequency range are shown in Figure 2.5. It has been computed with the algorithm proposed in [6].

The use of microphones placed on opposite points of the equator of a rigid sphere can significantly reduce the front/back confussion [5]. This fact has inspired, directly or indirectly, the use of microphone arrays over the surface of a rigid scatterer to record the sound pressure field for the subsequent binaural synthesis. Rigid spherical microphone arrays provide good spatial samplings of the sound pressure thanks to the use of a rigid scatterer that increases the sensitivity to the arrival direction of sound. The sampled sound field thus recorded can be reproduced over loudspeakers and headphones by characterizing it as a superposition of virtual sound sources surrounding the listener [37, 43].

2.2.3 Superposition of a representative set of HRTFs

Within a spherical volume of radius b, the sound pressure field $\hat{P}$ at a point $\mathbf{r} = (r, \Omega)$, due to a continuous distribution of secondary monopole sound sources on the sphere $\mathbf{b} = (b, \Omega')$, can be calculated using the Kirchhoff–Helmholtz integral theorem. It reads [44]

$$\hat{P}(\mathbf{r}, k) = \int_{\Omega' \in \mathbb{S}^2} P(\mathbf{r}, \mathbf{b}, k) G(\mathbf{r}, \mathbf{b}, k) d\Omega', \tag{2.16}$$

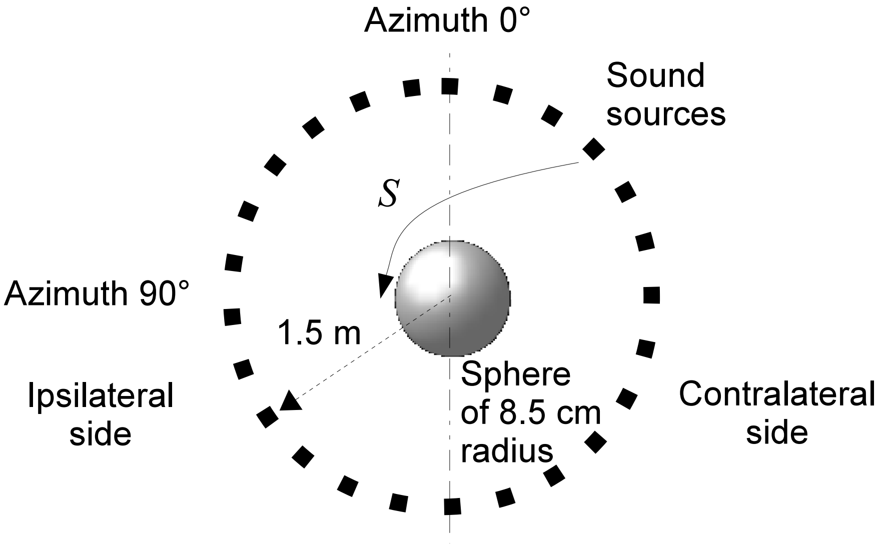


Figure 2.4: Top view of the geometry for the characterization of the total pressure field defined by Eq. (2.15), evaluated at the leftmost point on the surface of the rigid sphere. The sound sources are equiangularly distributed in the horizontal plane. Sound sources on azimuths between 0 and 180 degrees are said to lie on the ipsilateral side, and the ones on azimuths between 180 and 360 degrees, on the contralateral side.

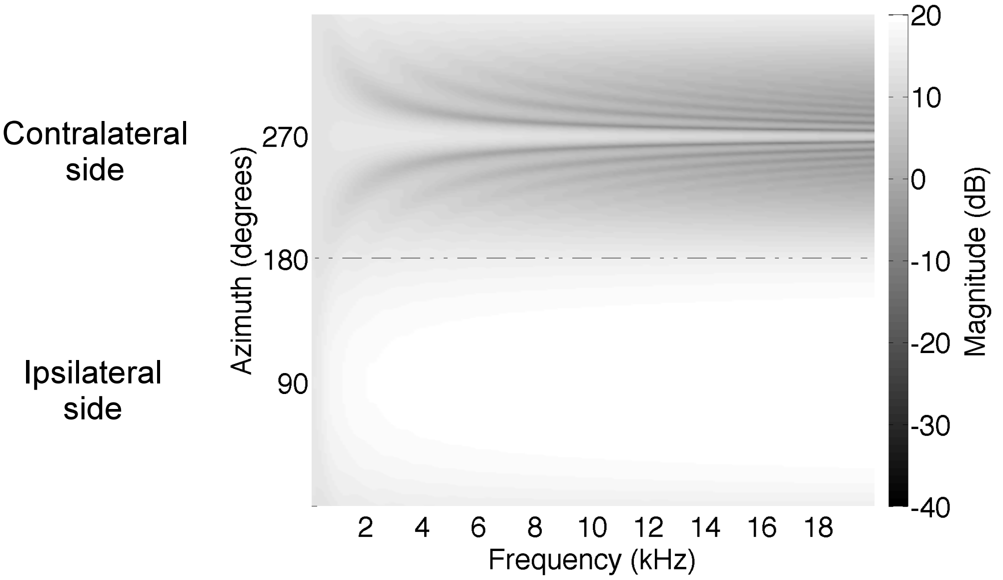


Figure 2.5: Total pressure field at the leftmost point of the rigid sphere described by Eq. (2.15). The total pressure depends on the size of the sphere, and on the frequency and position of the sound sources. The sources are on the horizontal plane and cover the full audible frequency range, from 20 Hz to 20 kHz. In general, ipsilateral sources are received with high intensities at the leftmost point, while the contralateral ones are shadowed and scattered by the rigid sphere. Below 4 kHz, this model approximates the HRTFs fairly good (*cf.* Figure 1.5).

where P denotes the complex amplitude of an incoming wave, with wave number k, arriving from a point in $\mathbf{b}$, and G the Green's function characterizing the sound field at $\mathbf{r}$ due to a monopole located at $\mathbf{b}$.

The synthesis of HRTFs is possible by replacing G in Eq. (2.16) by the HRTFs for sources on $\mathbf{b}$, and P by the density functions ρ at $\mathbf{r}$ for a sound source at $\mathbf{b}$. The HRTFs are interpreted as the complex amplitude of the sound pressure measured at both ears. Therefore, the formula for the synthesis of HRTFs for sources at any point $\mathbf{r}$, starting from a set of HRTFs measured from sound sources on the sphere $\mathbf{b}$, results in the following expression

$$\hat{H}(\mathbf{r}, k) = \int_{\Omega' \in \mathbb{S}^2} \rho(\mathbf{r}, \mathbf{b}, k) H(\mathbf{b}) d\Omega'. \tag{2.17}$$

In practice, the integration on the sphere must be approximated by a weighted sum of a discrete distribution of an integer number V of representative HRTFs, for sound sources on sampled positions $\mathbf{b}_v = (b, \Omega_v)$, and hence, Eq. (2.17) becomes

$$\hat{H}(\mathbf{r}, k) = \sum_{v=1}^{V} \rho(\mathbf{r}, \mathbf{b}_v, k) H_v(\mathbf{b}_v) \alpha_v, \tag{2.18}$$

where α_v are the normalized quadrature weights that approximates the differential $d\Omega'$ at each sampled point.

As for the computation of density functions ρ_v, several sound field techniques have been applied so as to synthesize the HRTFs for distal sound sources ($r = b$), and proximal sound sources ($r < b$). Wave field synthesis (WFS) [41] and higher order ambisonics (HOA) [38] have been proposed to synthesize HRTFs for distal sound sources. The sound field produced by the far field sound sources is typically decomposed into plane-waves, thus neglecting the distance-related effects, which may be important for the rendering with high levels of realism [21]. An extension to include the distance effects of sound sources in the proximal region, based on the modal beamforming technique [40, 42], have also been proposed.

Following the virtual loudspeaker approach, the binaural synthesis for distal sound sources is described in the next section, considering the density functions are derived from the sound field captured by a continuous spherical recording surface. An extension to a discrete recording surface is described in Chapter 3. The binaural synthesis for sound sources in the proximal region is described on Chapter 4.

2.3 Synthesis by a continuous measurement surface

A first appoach to the synthesis of binaural signals is to derive the density functions from the sound field captured by a continuous spherical measurement surface. We interpret $\rho(\mathbf{b}_v, k)$ in Eq. (2.18) as a spherical harmonic decompositon with band-limited spherical spectrum $\rho_{nm}(b)$. Using the spherical harmonic transform in Eq. (2.14), $\rho(\mathbf{r}, \mathbf{b}_v, k)$ can be written

$$\rho(\mathbf{r}, \mathbf{b}_v, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \rho_{nm}(b) Y_{nm}(\Omega_v), \tag{2.19}$$

where the truncated sum over n up to a constant N models the practical limited spatial bandwidth. Here, N must not be confused with the normalization factor N_{nm} of Eq. (2.4).

Let $\mathbf{a} = (a, \Omega')$ represent a rigid spherical measurement surface of radius a (See Figure 2.6), which can be understood as having an infinite number of microphones on the surface of the rigid sphere. The spherical wave spectrum $\rho_{nm}(b)$ can then be computed by back-propagating the sound pressure due to a source at $\mathbf{r}$ measured on the surface of radius a, denoted by $s(\mathbf{a}, \mathbf{r}, k)$, to the radiating surface of radius b. Applying the spherical harmonic transform in Eq. (2.14) to $s(\mathbf{a}, \mathbf{r})$, and replacing the result in Eq. (2.19), $\rho(\mathbf{r}, \mathbf{b}_v, k)$ now reads

$$\rho(\mathbf{r}, \mathbf{b}_v, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} B_n \int_{\Omega'} s(\mathbf{a}, \mathbf{r}, k) Y^*_{nm}(\Omega') d\Omega' Y_{nm}(\Omega_v), \tag{2.20}$$

where the beamformers B_n defined on the spherical harmonics domain must be computed so as to compensate for the measurement effects.

2.3.1 Spherical beamformers

The measured signal $s(\mathbf{a},\mathbf{r})$ can now be modeled with the acoustic scattering from the rigid sphere. Replacing $s(\mathbf{a},\mathbf{r},k)$ in Eq. (2.20) by the result of Eq. (2.15), and using the sum of Eq. (2.12), and the orthonormality property of Eq. (2.10) in the resulting expression yields

$$\rho(\mathbf{r},\mathbf{b}_v,k) = -\sum_{n=0}^{N} B_n(a,b,k)\left[\frac{h_n(kb)}{ka^2 h_n'(ka)}\right](2n+1)P_n(\cos\Theta_v), \tag{2.21}$$

where Θ_v is the angle between the source at $\mathbf{r}$ and the representative source at $\mathbf{b}_v$.

The filters B_n are spherical beamformers to be applied on the spherical spectrum, which can now be chosen in such a way that they compensate the factor in brackets in Eq. (2.21). Therefore,

$$B_n(a,b,k) = -ka^2\frac{h_n'(ka)}{h_n(kb)}, \tag{2.22}$$

whose factors are intended to remove the scattering effects from the rigid spherical measurement surface of radius a [35,36,47] and to backpropagate the pressure field from the center of the array to the radiating surface of radius b [44]. The spherical beamformers up to order $N = 14$ are shown in Figure 2.7, for $a = 8.5$ cm and $b = 1.5$ m.

The replacement of Eq. (2.22) in Eqs. (2.21) and (2.18) results in the following expression for the transfer functions $\hat{H}$ used to synthesize the binaural signals by means of a continuous spherical measurement surface:

$$\hat{H}(\mathbf{r},k) = \sum_{v=1}^{V}\alpha_v H(\mathbf{b}_v,k)\sum_{n=0}^{N}(2n+1)P_n(\cos\Theta_v), \tag{2.23}$$

where the quadrature weights α_v applied to the representative HRTFs are chosen in such way that [36]

$$\sum_{v=1}^{V}\alpha_v Y_{nm}(\Omega_v)Y^*_{n'm'}(\Omega_v) = \delta_{n-n'}\delta_{m-m'}. \tag{2.24}$$

The use of almost regular spherical grids is convenient for computations with spherical harmonics. Their quadrature weights α_v in Eq. (2.24) can be chosen to be equal to the area

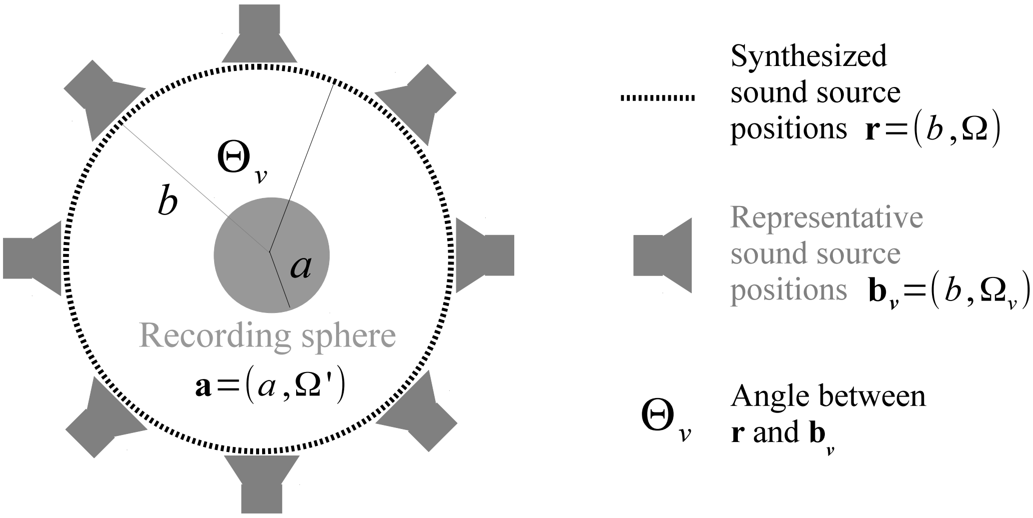

b
Θ_v
a
Recording sphere
$\mathbf{a}=(a,\Omega')$
Synthesized sound source positions $\mathbf{r}=(b,\Omega)$
Representative sound source positions $\mathbf{b}_v=(b,\Omega_v)$
Θ_v
Angle between $\mathbf{r}$ and $\mathbf{b}_v$

Figure 2.6: Geometry for the synthesis of binaural signals for sound sources in the distal region. Sound field due to a sound source is ideally measured by a rigid, spherical and continuous surface of radius a, and then analyzed by spherical harmonics. The rendering assumes a virtual array of loudspeakers placed on the representative positions $\mathbf{b}_v$ (see Figure 1.2). Loudspeaker's driving signals are computed so that they match the captured sound field. The synthesized sound sources lie at the same radial distance of the loudspeakers or farther.

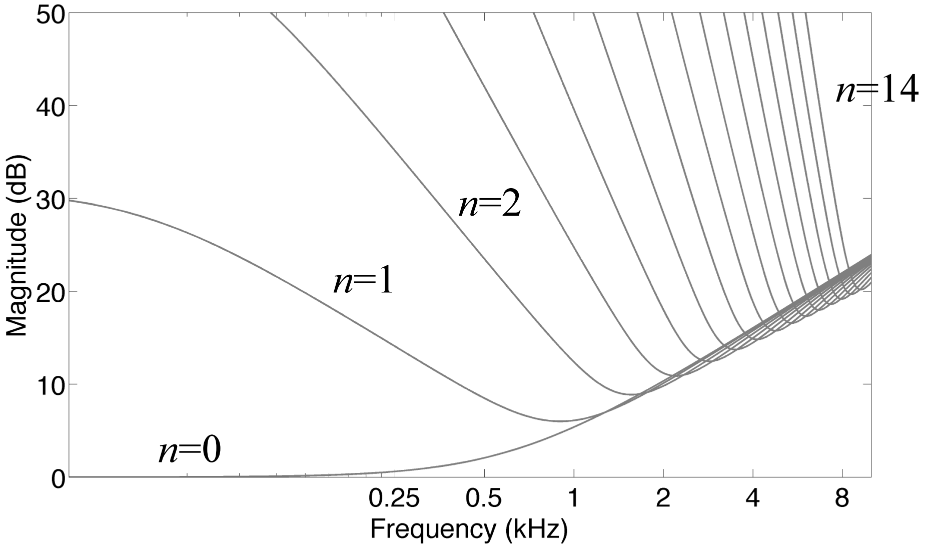


Figure 2.7: Filters to be applied on the spherical spectrum of the recorded signals, along orders $n = 0, ..., 14$. They are intended to remove the scattering from a rigid sphere of 8.5 cm radius due to a sound source at a 1.5 m distance (see Eq. (2.22)). On ideal recordings with infinite microphones, these filters perfectly cancel the scattering up to a given order. On recordings with microphone arrays, the small size of the spherical scatterer will demand very large gains to compensate for the high directivity components at low frequencies.

of the Voronoi cell containing each sampling point, and normalized in such a way that:

$$\sum_{v=1}^{V} \alpha_v = 1. \tag{2.25}$$

Given that Eq. (2.23) is independent of the distribution of microphones, it allows for the evaluation of the isolated effects arising from the arrangement of virtual sources.

2.4 Numerical accuracy of the synthesis

For evaluation, we relied on the computer simulation of a sound field recorded by a rigid, continuous and spherical measurement surface. The representative positions for the virtual loudspeakers were arranged in almost regular samplings of a surrounding sphere. The transfer functions for the whole process were characterized and compared with a set of target HRTFs computed numerically for a dummyhead.

The number $(N+1)^2$ of spherical harmonics for an accurate binaural synthesis is determined by the wave number k of the sound sources. The spatial bandwidth of the captured sound field depends on the radius a of the rigid spherical measurement surface. On the other hand, the spatial bandwidth of the representative set of HRTFs is determined by the radius of the smallest sphere enclosing the listener's head. In both cases, the radial spherical wave spectrum is represented by spherical Bessel functions $j_n(kr)$, which show a rapid decay of orders $n > kr$. Hence, a bound for the required order N of the spherical harmonics analysis to characterize a captured sound field or a set of HRTFs can be approximated by [48]

$$N = \left\lceil \frac{eka}{2} \right\rceil = \left\lceil \frac{e\pi f a}{c} \right\rceil, \tag{2.26}$$

where f is the frequency of the sound sources in Hz, c the velocity of sound in air in m/s, and $e \approx 2.7183$ (see Figure 2.8).

To approximate the entire audible frequency range, from 20 Hz to 20 kHz, using a spherical scatterer of $a = 8.5$ cm radius, which is the size of an average human head, an order $N = 43$ is needed, and therefore, at least $V = (43 + 1)^2 = 1936$ representative sound sources would be required. However, a limited spatial bandwidth is imposed by the finite number Q of microphones in practical arrays. In particular, the recording setup available at the Research Institute of Electrical Communication has $Q = 252$ microphones arranged on a spherical scatterer, which allows for spherical harmonics decompositions up to order $N = 14$, and therefore the accuracy of the synthesis will only be evaluated up to a spatial aliasing frequency of around 8 kHz.

2.4.1 Parameters and conditions for the evaluation

The parameter under evaluation was the number of virtual sources. Its effect on the accuracy of the synthesis is evaluated. The representative sound sources were arranged on spherical grids based on subdividing each face of an icosahedron (see Figure 2.10). These icosahedral grids provide an almost regular covering of the sphere. The HRTFs for sets of representative sound sources at a 1.5 m distance, arranged on icosahedral grids (see Figure 2.10), were computed using the boundary element method (BEM) [4] with a 3D mesh of the SAMRAI dummy head.

The accuracy was evaluated by comparing the transfer functions for the whole synthesis process denoted by $\hat{H}(\theta, f)$ (see Eq. (2.23)) and a set of target HRTFs computed numerically for a dummyhead and denoted by $H(\theta, f)$. Synthesis accuracy along azimuth θ was measured by the spectral distortion (SD) defined by the logarithmic spectral distance between $H(\theta, f)$ and $\hat{H}(\theta, f)$ [49]

$$SD(\theta) = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(20\log_{10}\left|\frac{H(\theta, f_i)}{\hat{H}(\theta, f_i)}\right|\right)^2}. \tag{2.27}$$

Synthesis accuracy along frequency f was measured by the normalized spherical correlation (SC) between $H(\theta, f)$ and $\hat{H}(\theta, f)$ [21]

$$SC(f) = \frac{\sum_{p=1}^{P} H(\theta_p, f)\hat{H}(\theta_p, f)}{\sqrt{\sum_{p=1}^{P} H(\theta_p, f)^2}\sqrt{\sum_{p=1}^{P} \hat{H}(\theta_p, f)^2}}. \tag{2.28}$$

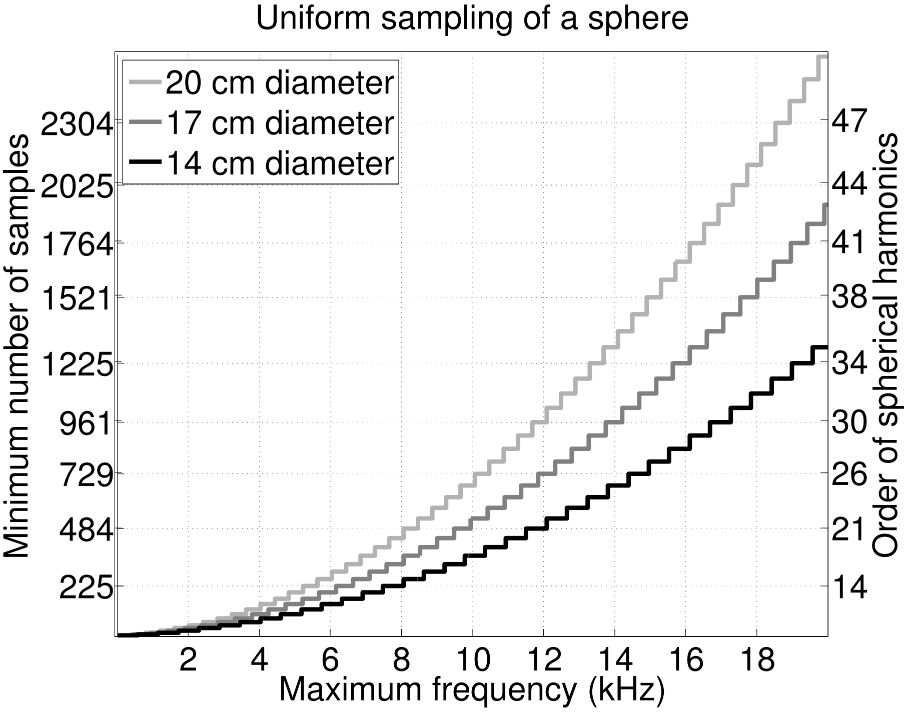


Figure 2.8: The order n of the spherical spectrum provides a measure of the spatial bandwidth for almost uniform samplings of a sphere (see Eq. (2.26)). The spatial bandwidth of a captured sound field depends on the radius a of the rigid spherical measurement surface. On the other hand, the spatial bandwidth of the binaural localization cues, due to the scattering from the listener's head, depends on the radius of the smallest sphere enclosing the head. In both cases, the frequency of sound sources defines the minimum number of microphones to capture a sound field, or the minimum number of virtual loudspeakers associated to the representative HRTFs.

The number of synthesized directions on the horizontal plane were $P = 360$. The number of frequency bins of the target HRTFs was 512 for a sampling frequency of 48 kHz, but the error was computed up to 9 kHz, and therefore up to $I = 97$.

2.4.2 Simulation results

Figure 2.9 shows the target HRTFs on the horizontal plane, which have been computed with a BEM solver. Figure 2.11 shows the HRTFs synthesized with a continuous rigid spherical measurement surface of 8.5 cm radius, spherical harmonic functions up to order 14, and representative sound sources distributed on icosahedral grids. Here, the spatial aliasing effects start to appear around 8 kHz, affecting the shadowed side of the head. Figure 2.12 shows the synthesis accuracy along frequency and direction, measured with the spectral distortion in Eq. (2.27) and the normalized spherical correlation in Eq. (2.28).

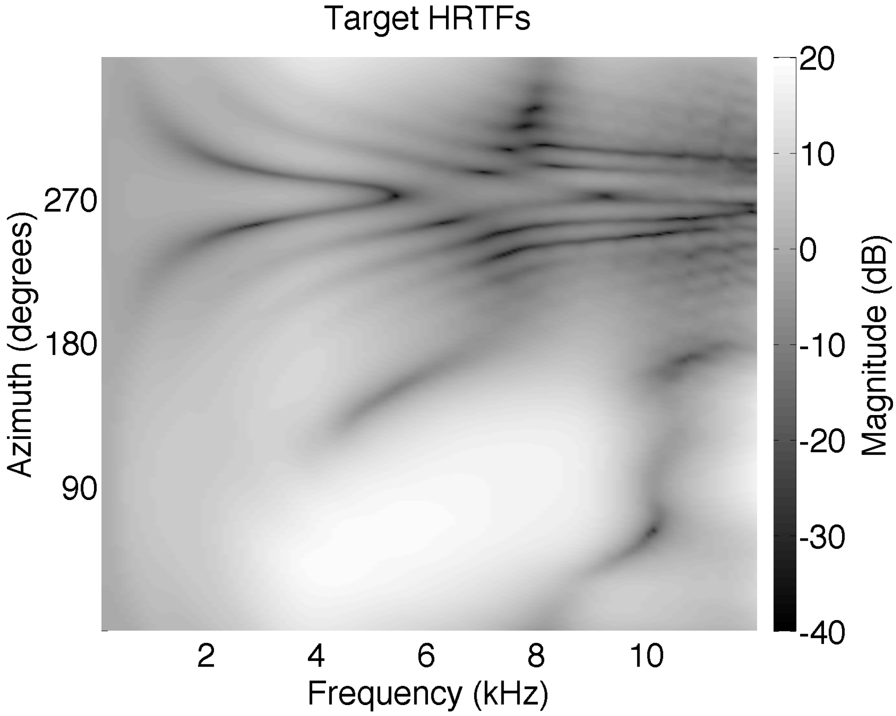
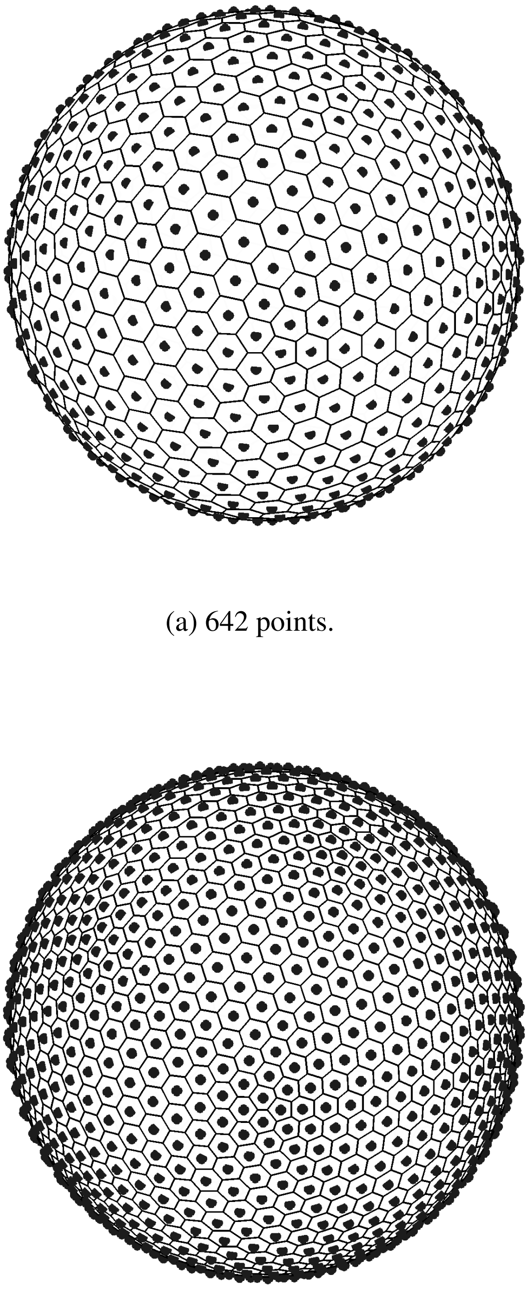


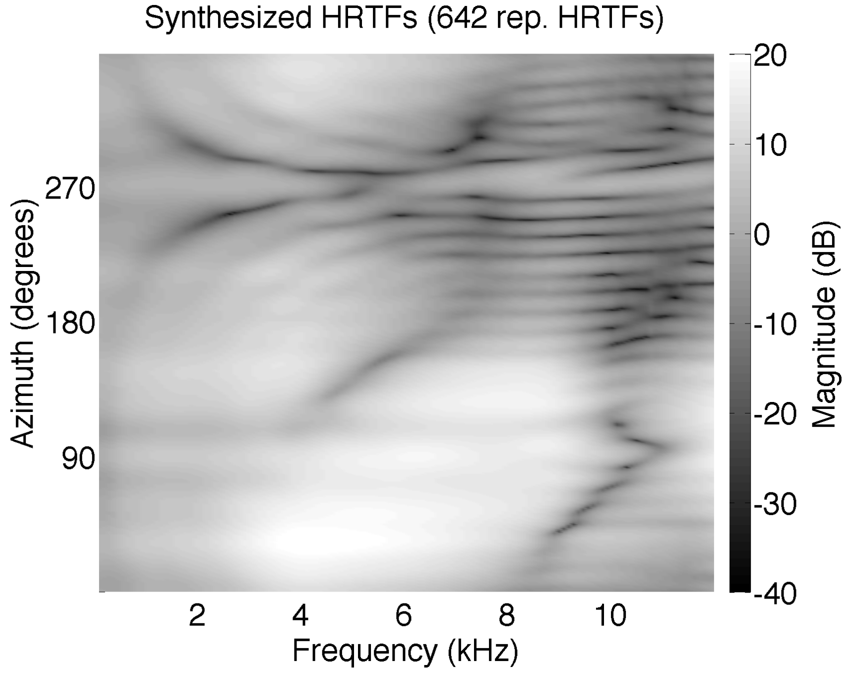
Figure 2.9: Head-related transfer functions for the left ear and sound sources on the horizontal plane. They were computed with the boundary element method (BEM) [4] for a 3D mesh of an artificial head, and 360 sound sources at a 1.5 m distance from the center of the head, equiangularly distributed on the horizontal plane. The transfer functions of the binaural synthesis method, described in Eq. (2.23), are compared with these target HRTFs.



(a) 642 points.

(b) 1212 points.

Figure 2.10: Spherical grids based on subdivisions of the vertex of an icosahedrum. The grid on Panel (a) was obtained by subdividing each vertex into 7 intervals. The grid on Panel (b) was obtained by subdividing each vertex into 10 equal intervals. These icosahedral grids define a set of representative positions used to place the virtual loudspeakers for binaural synthesis (see Figure 2.6).



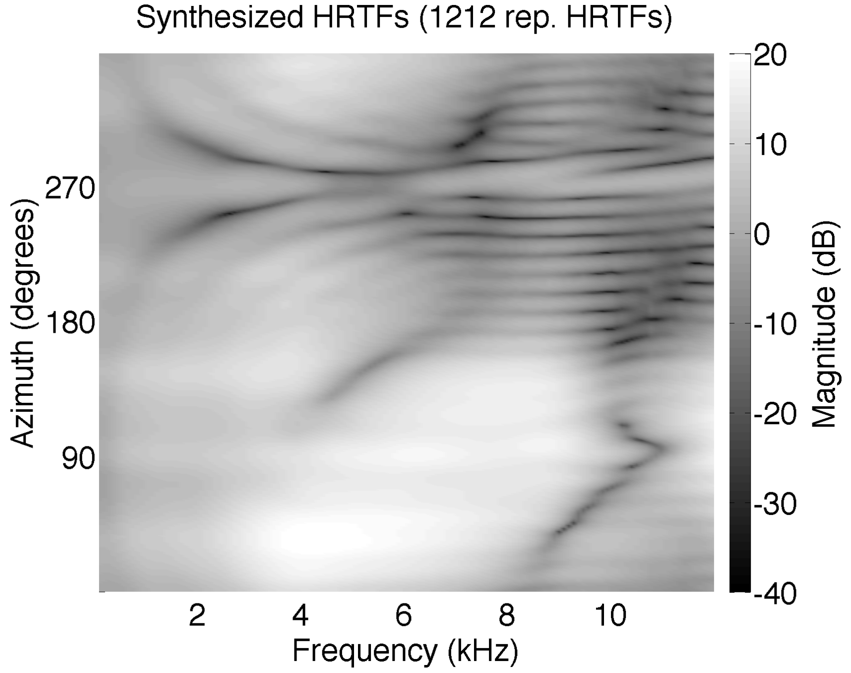
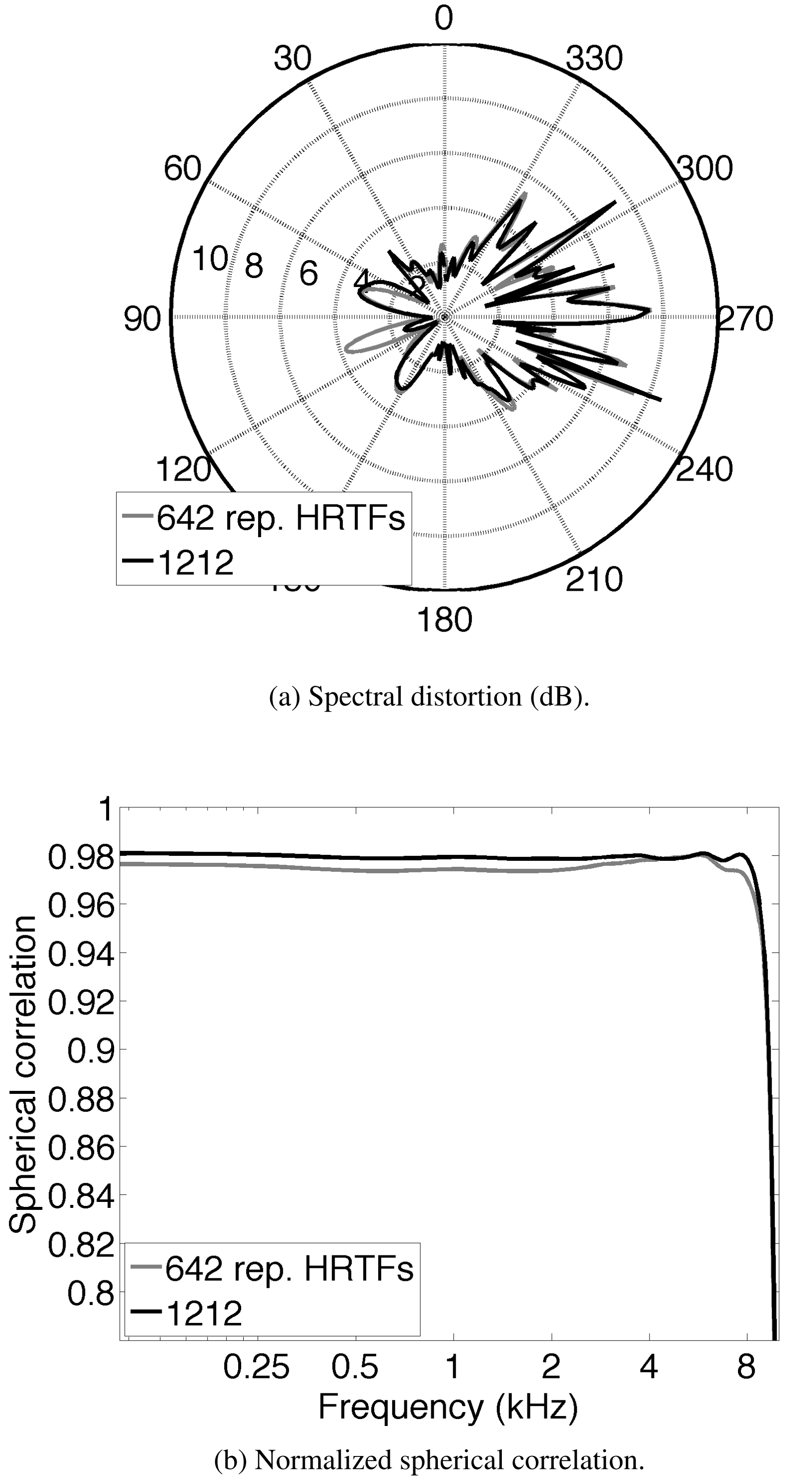


Figure 2.11: Transfer functions of the binaural synthesis method described by Eq. (2.23). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the icosahedral grids shown in Figure 2.10, and spherical harmonics decompositions of order $N = 14$. A visual comparison of the synthesized HRTFs with the target ones in Figure 2.9 shows that the synthesis is fairly good, up a spatial aliasing frequency of around 8 kHz on the ipsilateral side, and around 6 kHz on the ipsilateral side. The lowest bound is in accordance with Figure 2.8.



(a) Spectral distortion (dB).

(b) Normalized spherical correlation.

Figure 2.12: Numerical accuracy of the synthesis with virtual loudspeakers on icosahedral grids. The transfer functions in Figure 2.11 are compared with the target HRTFs in Figure 2.9 using the spectral distortion in Eq. 2.27 and the normalized spherical correlation in Eq. (2.28).

2.5 Summary

With the present approach, a binaural response is synthesized directly from the directional distribution of the incident pressure field on the rigid sphere. Hence, the underlying idea of this approach is closely related to modal beamforming techniques [40, 42]. On the other hand, given the decomposition of the captured sound field in terms of spherical harmonics and subsequent reconstruction in the directions of the representative sound sources, the presented is similar to existing techniques for the angular interpolation of HRTFs based on the spherical harmonic decomposition [18, 19].

Chapter 3

Binaural synthesis by spherical microphone arrays

3.1 Introduction

This chapter continues with the evaluation of the binaural synthesis for distal sound sources introduced in chapter 2. The optimal arrangement of representative sound sources for angular interpolation is discussed in this Section. A practical scenario for the synthesis is subsequently introduced, where accuracy of the transfer functions or the proposed system, compared with a set of target HRTFs, is evaluated assuming a finite number of microphones.

3.2 Sound source distributions for representative HRTFs

Measured or computed sets of HRTFs are usually used for sound sources distributed on spherical grids centered on the listener's head. The following characteristics are desired for the arrangement of the sound sources [48]

- Least number of measurements: For efficiency, the number of points on the proposed sampling grid must at least equal 2000 sample positions to cover the audible frequency range (*cf.* Figure 2.8).

- Equal area division: The proposed sampling grid should have nearly equal area division of the sphere, which makes all the measurements contribute nearly equally when used in the spherical harmonic transform.

- Hierarchical structure of data: The database measured on the proposed sampling grid should be structured such that a low spatial resolution data set, suitable for low frequencies, is imbedded in the high spatial resolution data set.

- Iso-longitude measurement setup: The proposed sampling strategy should measure all elevations at each azimuth in order that the listener and apparatus experience the least rotations.

The three first item are inherent to almost uniform distributions of points on a surrounding sphere. The fourth item is related to practical measurement setups due to the simplicity of mechanical rotations. Given that the scope of our proposal is the almost uniform distribution of virtual loudspeakers, the next section examines two spherical coverings, the icosahedral and the Lebedev grids, which fulfill with the two first items.

3.2.1 Spherical samplings

Although the regular covering of the sphere is still an open mathematical problem [2], several sampling strategies have been proposed for the arrangement of points on a spherical surface. Among them, the icosahedral grids and the Lebedev grids are of special importance for computations based on spherical harmonics.

[2] Uniform distributions of points on a spherical surface are only possible for the special cases of the platonic solids. Many different criteria can be used to distribute points almost uniformly, including minimum energy, covering, packing, Voronoi cells, volume of their convex hull, maximum determinant, cubature weights and norms of the Lagrange polynomials. [50, 51].

The icosahedral grids are based on the subdivision of the icosahedron's edges. A set of icosahedral grids is shown on Figure 3.1. On the other hand, the Lebedev grids are constructed so as to have octahedral rotation and inversion symmetry. The number and location of the grid points together with a corresponding set of integration weights are determined by enforcing the exact integration of spherical harmonics on the unit sphere up to a given order. A set of Lebedev grids appear on Figure 3.2. Both spherical grids fulfill the requirements of least measurements, nearly equal area division, and hierarchical structure.

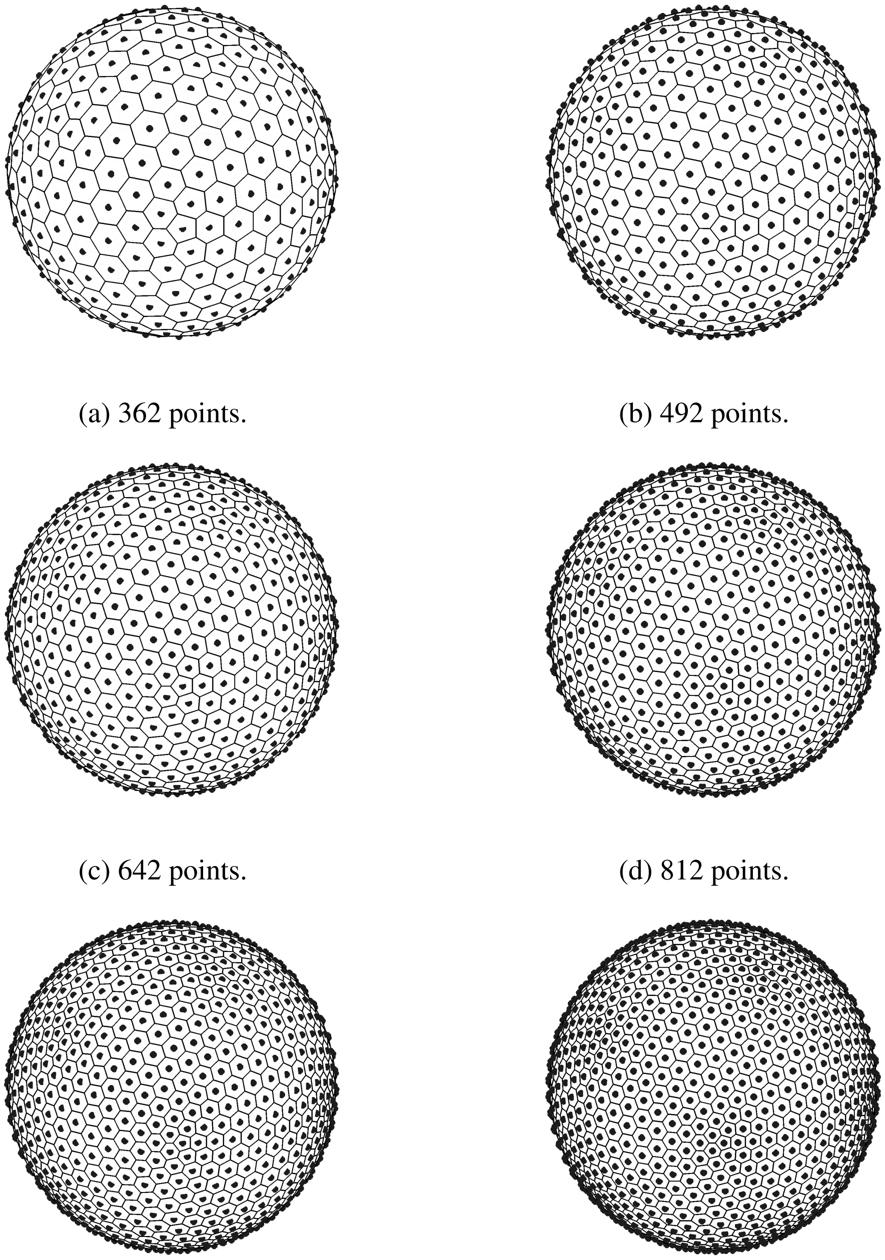
3.2.2 Spatial aliasing

The approximation of functions on the unit sphere in terms of the spherical harmonics is possible thanks to the orthonormality property of Eq. (2.10). However, interference of orders occurs when applying the spherical harmonic transform on samplings of the sphere, specially when the number of spherical harmonics $(N+1)^2$ is lower than the number of samplings V. Therefore, the orthonormality property does not hold anymore, causing the so-called spatial aliasing.

A measure of the interference of orders is given by the orthonormality error, defined by [53]

$$E_{nm} = \sum_{n'=0}^{N} \sum_{m'=-n'}^{n'} \left[\delta(n-n')\delta(m-m') - \sum_{v=1}^{V} \alpha_v Y_{nm}(\Omega_v) Y^{*}_{n',m'}(\Omega_v) \right]^2 . \tag{3.1}$$

The orthonormality error for Icosahedral and Lebedev grids is shown on Figure 3.3. Given the practical limitation of having a finite number of microphones, $Q = 252$, the maximum order is set to $N = 14$. For this order, the orthonormality error graph provides the minimum number of representative sound sources for each type: 362 for Icosahedral grids and 302 for Lebedev grids. Although the error is still noticeable, it remains constant. below this quantity of points and decreasing orders, the interference of orders can be seen on these plots.



(a) 362 points.

(b) 492 points.

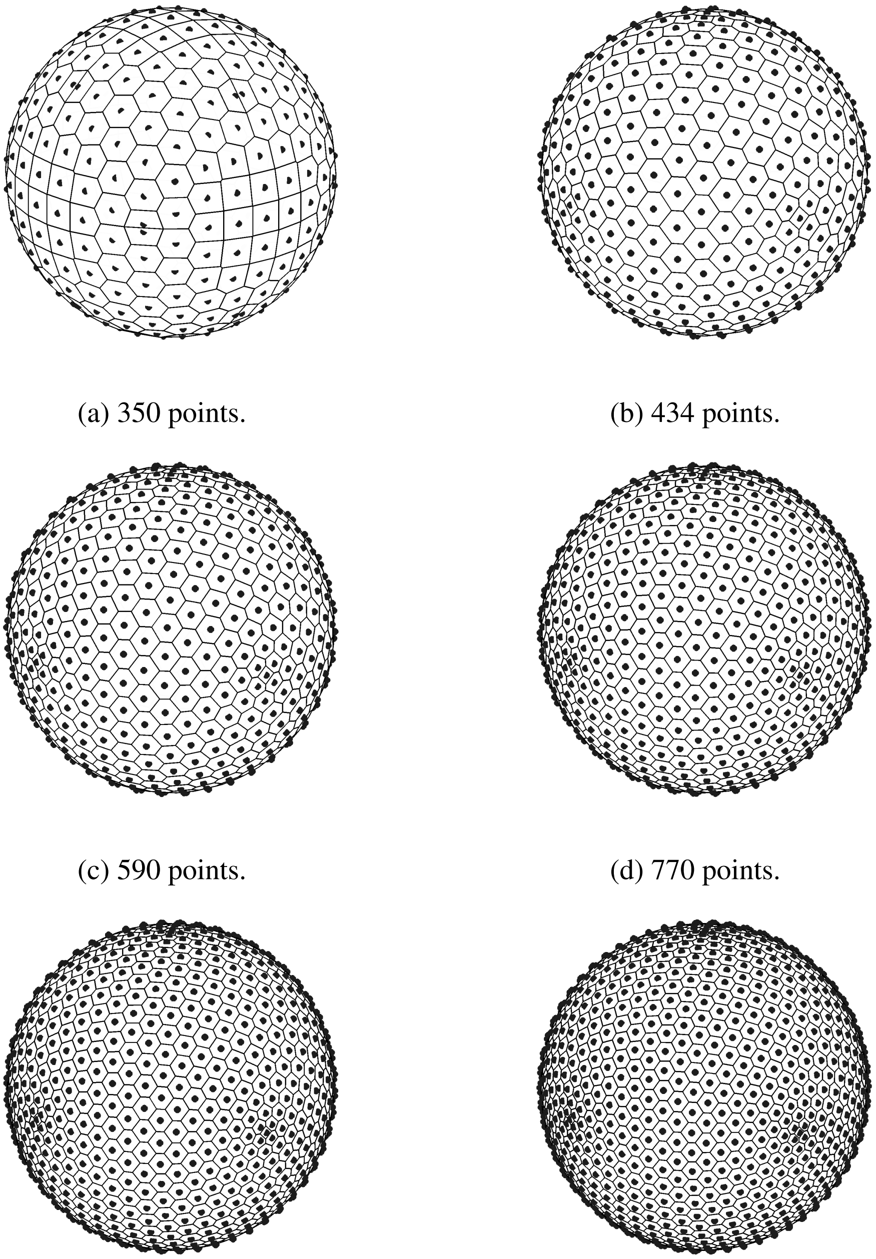
(c) 642 points.

(d) 812 points.

(e) 1002 points.

(f) 1212 points.

Figure 3.1: Spherical grids based on subdivisions of the edges of an icosahedron. The grids were obtained by subdividing each edge into (a) 5 intervals, (b) 6 intervals, (c) 7 intervals, (d) 8 intervals, (e) 9 intervals, and (f) 10 intervals. These icosahedral grids define the representative positions used to place the virtual loudspeakers for binaural synthesis (see Figure 2.6).



(a) 350 points.

(b) 434 points.

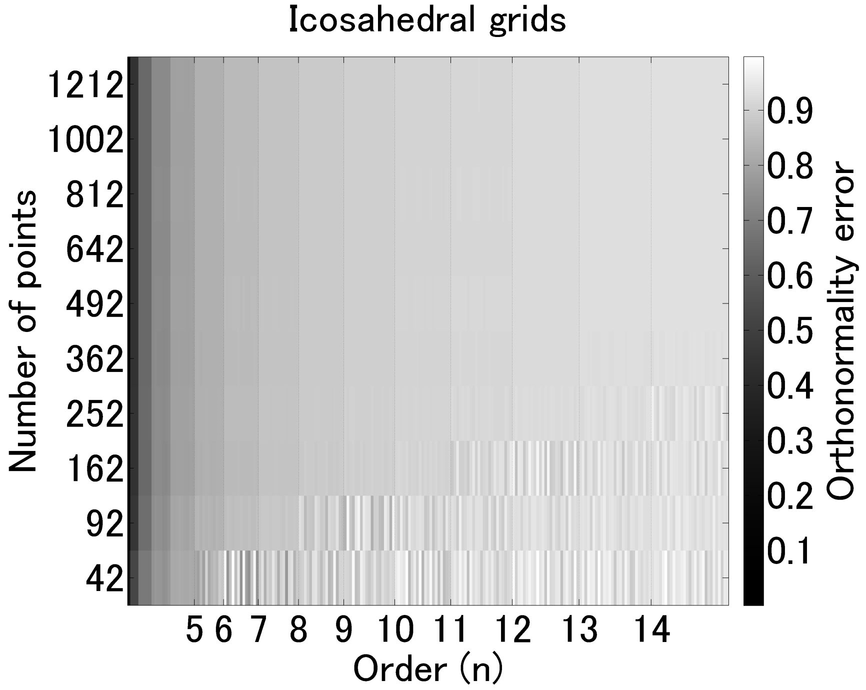
(c) 590 points.

(d) 770 points.

(e) 974 points.

(f) 1202 points.

Figure 3.2: Spherical grids proposed by Lebedev [52]. They are constructed so as to obtain octahedral rotation and inversion symmetry. The number and distribution of points, together with the corresponding quadrature weights (see Eq. (2.24)), are determined by enforcing the exact integration of spherical harmonics on the unit sphere up to a given order. These grids define another set of representative positions used to place the virtual loudspeakers for the synthesis of binaural signals (see Figure 2.6).



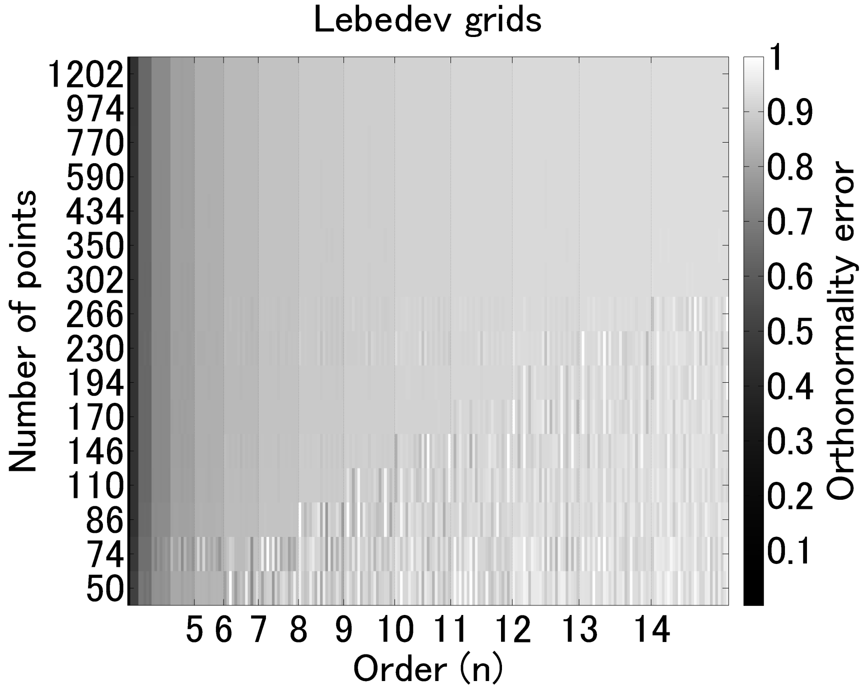


Figure 3.3: The orthonormality error E_{nm} in Eq. (3.1) defines a measure of the spatial aliasing in spherical grids. The measurement is based on the orthonormality property of the spherical harmonics on the continuous unit sphere. This property does not hold on samplings of the sphere. The spatial bandwidth is set by the order n of the spherical spectrum. The wider the spatial bandwidth, the bigger the number of points required to avoid the spatial aliasing.

3.3 Sampling the continuous measurement surface

In practice, a finite number of microphones at points $\mathbf{a}_q$, with $q = 1, ..., Q$, on the surface of the rigid sphere is used (See Figure 3.4). Therefore, the integral in Eq. (2.20) must be approximated with a sum over q. After using the sum in Eq. (2.12) with the resulting sum, Eqs. (2.19) and (2.20) become the formula for the synthesis of the binaural signals using a rigid spherical microphone array and a set of representative HRTFs:

$$\hat{H}(\mathbf{r}, k) = \sum_{v=1}^{V} \alpha_v H(\mathbf{b}_v, k) \sum_{n=0}^{N} (2n+1) B_n \sum_{q=1}^{Q} P_n(\cos \Theta_{vq}) \beta_q S(\mathbf{r}, \mathbf{a}_q, k), \qquad (3.2)$$

where B_n is the filter on the spherical harmonic domain defined by Eq. (2.22), Θ_{vq} is the angle between the virtual source at $\mathbf{b}_v$ and the microphone at $\mathbf{a}_q$, and the quadrature weights β_q applied to the individual microphone signals are chosen in such way that the orthonormality property in Eq. (2.10) is fulfilled, and therefore [36]

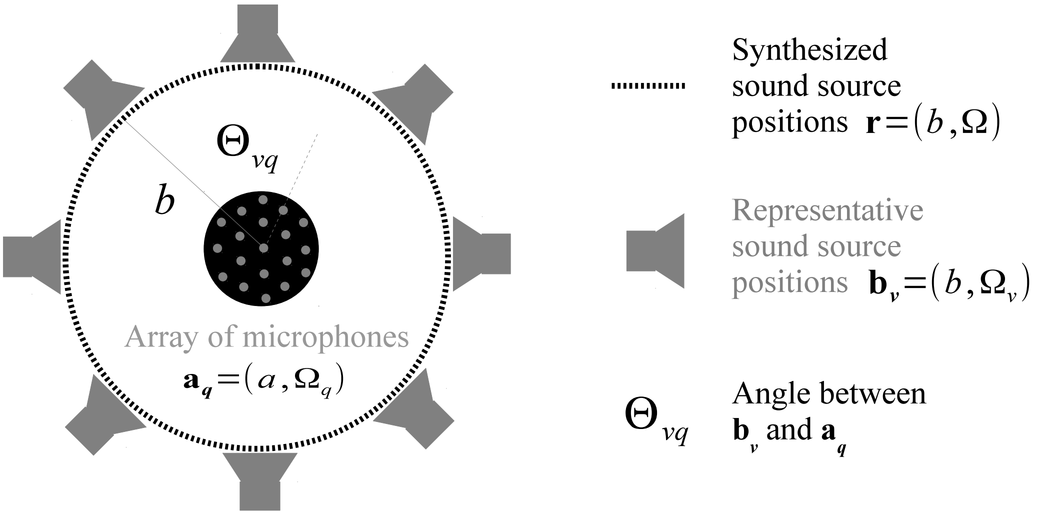
$$\sum_{q=1}^{Q} \beta_q Y_{nm}(\Omega_q) Y^*_{n'm'}(\Omega_q) = \delta_{n-n'} \delta_{m-m'}. \qquad (3.3)$$

3.3.1 Near field order limitation

High gains at low frequencies appeared on the spherical beamformers due to the inversion of higher orders on the model of acoustic scattering from the rigid sphere of small size (See Figure 2.7). In order to avoid such a low frequency distortion, frequency and spatial modes are related so as to select the reconstruction order N according to the wave number k and the size of the scatterer a such that

$$N = \min(\lceil \frac{eka}{2} \rceil, \lfloor \sqrt{Q} - 1 \rfloor), \qquad (3.4)$$

where Q is the number of microphones, which imposes the upper limit to the order. The order selection is illustrated in Figure 3.5.



Θ_{vq}
b
Array of microphones
$\mathbf{a}_q=(a,\Omega_q)$
Synthesized sound source positions $\mathbf{r}=(b,\Omega)$
Representative sound source positions $\mathbf{b}_v=(b,\Omega_v)$
Θ_{vq}
Angle between $\mathbf{b}_v$ and $\mathbf{a}_q$

Figure 3.4: Geometry for the synthesis of binaural signals for sound sources in the distal region. The sound field due to a sound source is measured by the rigid spherical microphone array of radius a, and then analyzed by spherical harmonics up to order N. The rendering assumes a virtual array of loudspeakers placed at the representative positions $\mathbf{b}_v$ (see Figure 1.2). The loudspeaker's driving signals are computed so that they match the captured sound field. The synthesized sound sources lie at the same radial distance of the loudspeakers, or farther.

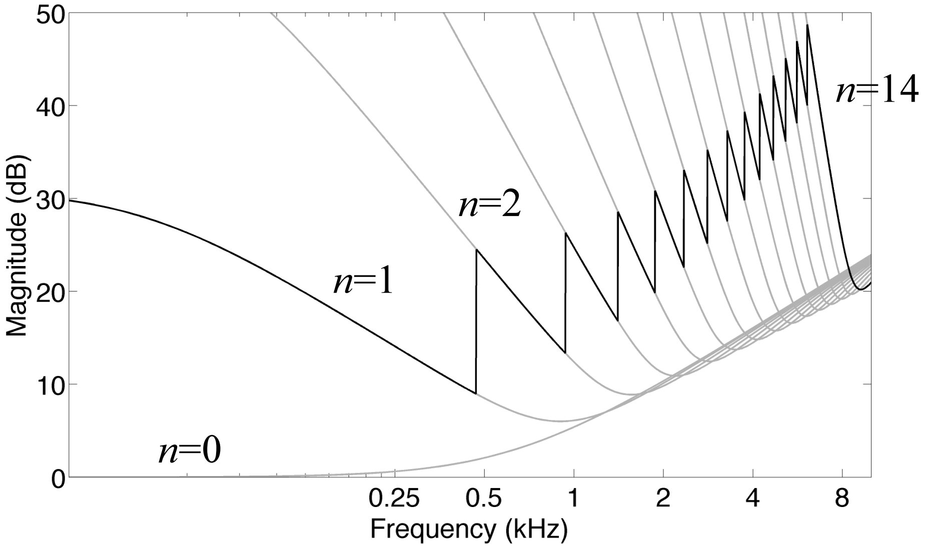


Figure 3.5: Filters to be applied on the spherical spectrum of the recorded signals, along orders $n = 0, ..., 14$. They are intended to remove the scattering from a rigid sphere of 8.5 cm radius due to a sound source at a 1.5 m distance (see Eq. (2.22)). The small size of the spherical scatterer will demand very large gains to compensate for the high directivity components at low frequencies. To sidestep this drawback, the reconstruction order is chosen according to the wave number and the size of the scatterer (see Eq. (3.4)). The dark line sets the upper bound at different frequency bands.

3.3.2 Matrix formulation

The synthesis of the binaural signals for sound sources in the distal region, from the spherical harmonic analysis of compact microphone arrays recordings and a set of representative HRTFs, described by Eq. 3.2, can be expressed in terms of matrix multiplications.

For each ear, each sound sound source position, and each frequency bin, the following expressions represent the transfer function $\hat{H}$ of the whole process required for the proposed binaural synthesis method:

$$\hat{H} = \sum_{v=1}^{V} W_v H_v, \tag{3.5}$$

for H_v the HRTF associated to the v-th virtual loudspeaker, and its corresponding weighting coefficient W_v defined by

$$W_v = \mathbf{F}\mathbf{P}_v\mathbf{S}, \tag{3.6}$$

which depend on the spherical microphone array signals

$$\mathbf{S} = \begin{bmatrix} S_1 \\ \vdots \\ S_Q \end{bmatrix}_{Q\times 1}, \tag{3.7}$$

the directivity patterns matching the arrays of microphones and virtual loudspeakers

$$\mathbf{P}_v = \begin{bmatrix} P_0(\Theta_{v1}) & \cdots & P_0(\Theta_{vQ}) \\ \vdots & \ddots & \vdots \\ P_N(\Theta_{v1}) & \cdots & P_N(\Theta_{vQ}) \end{bmatrix}_{(N+1)\times Q}, \tag{3.8}$$

and the scattering compensation filters

$$\mathbf{F} = -ka^2 \begin{bmatrix} \frac{h'_0(ka)}{h_0(kb)} & \cdots & (2N+1)\frac{h'_N(ka)}{h_N(kb)} \end{bmatrix}_{1\times(N+1)}. \tag{3.9}$$

3.4 Numerical accuracy of the synthesized HRTFs

To approximate the entire audible frequency range using a spherical scatterer of a = 8.5 cm radius, which is the size of an average human head, an order $N = 43$ is needed, and therefore, at least $Q = (43 + 1)^2 = 1936$ microphones over the sphere would be required. However, a limited spatial bandwidth is imposed by the reduced number of microphones in practical arrays, and hence, the accuracy can only be evaluated up to a spatial aliasing frequency.

3.4.1 Parameters and conditions for the evaluation

The parameter under evaluation was the number of representative sound sources. Its effect on the accuracy of the synthesis was evaluated. The representative sound sources were arranged on icosahedral grids and Lebedev grids. The center of the rigid spherical microphone array was set as the reference position, and its radius was chosen to be 8.5 cm. The recorded microphone signals were generated with the model of the acoustic scattering from the rigid sphere in Eq. (2.15), which was computed for 360 sources at a 1.5 m distance from the reference position, equiangularly distributed on the horizontal plane. The target HRTFs from 360 sources at a 1.5 m distance equiangularly distributed on the horizontal plane were computed using the Boundary-Element-Method (BEM) [4] with a 3D mesh of the SAMRAI dummy head. Other HRTFs from sets of virtual sources at a 1.5 m distance arranged on icosahedral and Lebedev grids were also computed with the BEM solver.

The evaluation based on a discrete measurement surface was performed. An array of $Q = 252$ microphones distributed in a icosahedral grid over the surface of the rigid sphere of 8.5 cm radius was assumed for this purpose, which is the available setup at the Research Institute of Electrical Communication, Tohoku University [11]. The number of microphones $Q = 252$ imposed a spatial bandwidth limitation up to the order $N = 14$, and therefore the accuracy could be evaluated up to a spatial aliasing frequency of around 9 kHz. The model of the acoustic scattering from the rigid sphere was decomposed up to order 14 at the positions of the microphones, and reconstructed at the positions of the virtual sources. The resultant signals were downmixed to a binaural signal. This process is formulated in Eq. (3.2).

3.4.2 Simulation results

Figure 3.6 shows the target left ear HRTFs computed with a dummy head model and sources at a 1.5 m distance on the horizontal plane. Figures 3.7 to 3.9 show the transfer functions obtained with the proposed method for representative sound sources arranged on icosahedral grids. Figures 3.10 to 3.12 show the synthesized HRTFs obtained with the proposed method for representative sound sources arranged on Lebedev grids. Figures 3.13 and 3.14 show the accuracy of the binaural synthesis using the spectral distortion in Eq. (2.27) and the spherical correlation in Eq. (2.28), respectively. In both cases, the binaural synthesis were synthesized with 252 microphones arranged on a rigid sphere of 8.5 cm radius and spherical harmonic functions up to order 14. The spatial aliasing effects start to appear around 8 kHz, affecting the shadowed side of the head the most.

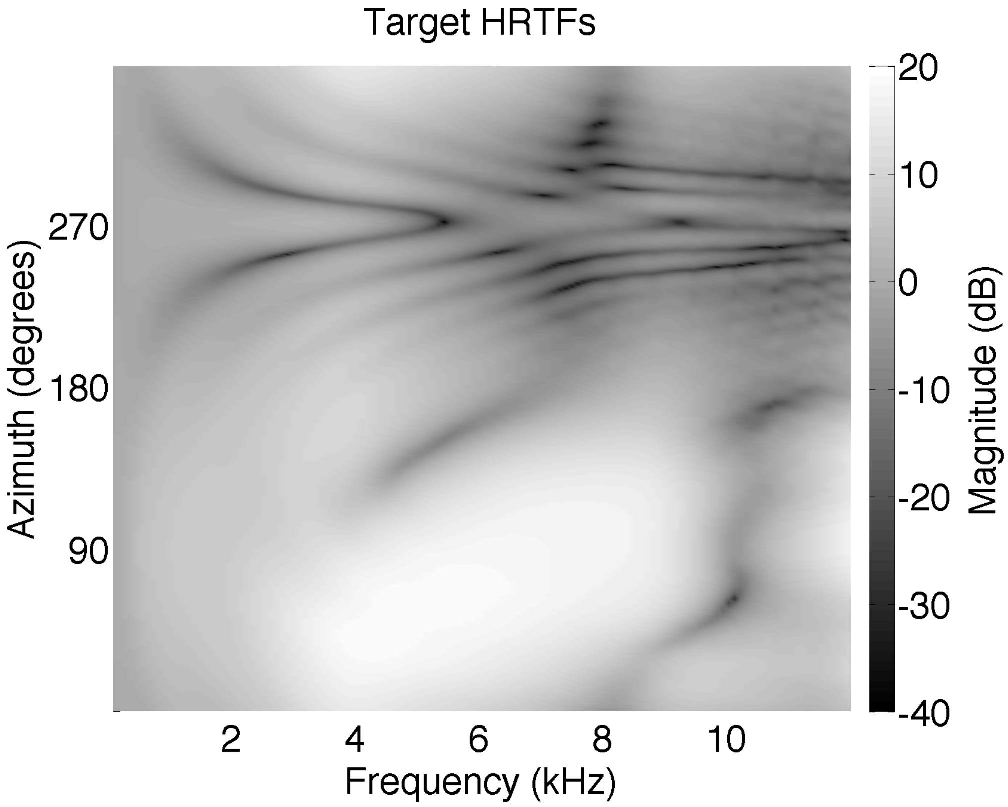


Figure 3.6: Head-related transfer functions for the left ear and sound sources on the horizontal plane. They were computed with the boundary element method (BEM) [4] for a 3D mesh of an artificial head, and 360 sound sources at a 1.5 m distance from the center of the head, equiangularly distributed on the horizontal plane. The transfer functions of the binaural synthesis method, described in Eq. (2.23), are compared with these target HRTFs.

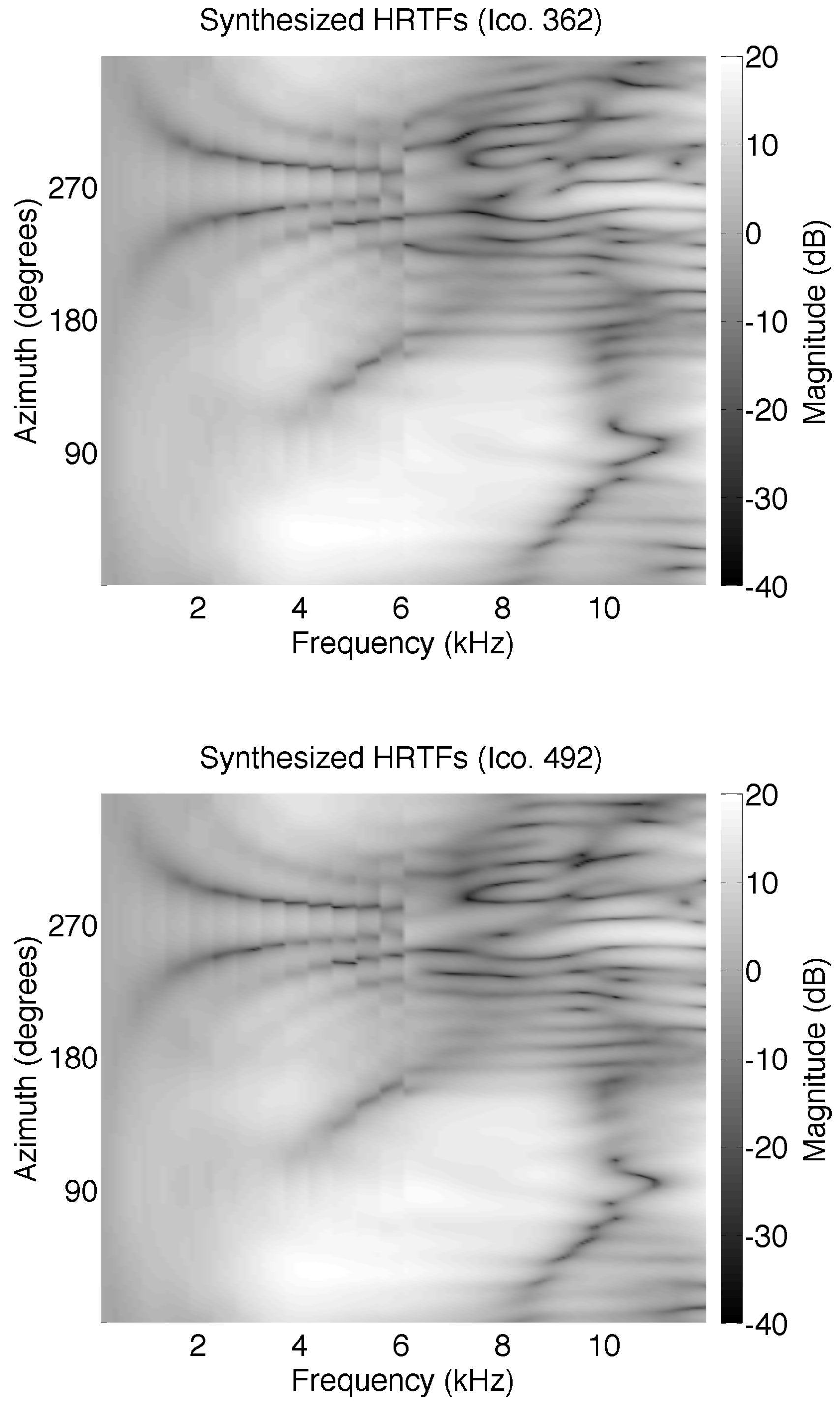


Figure 3.7: Transfer functions of the binaural synthesis method described by Eq. (3.2). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the icosahedral grids shown in Panels (a) and (b) of Figure 3.1, and spherical harmonics decompositions of order $N = 14$. A comparison with the target HRTFs in Figure 3.6 highlights the discontinuities along frequency due to the order limitation set by Eq. (3.4) and illustrated in Figure 3.5.

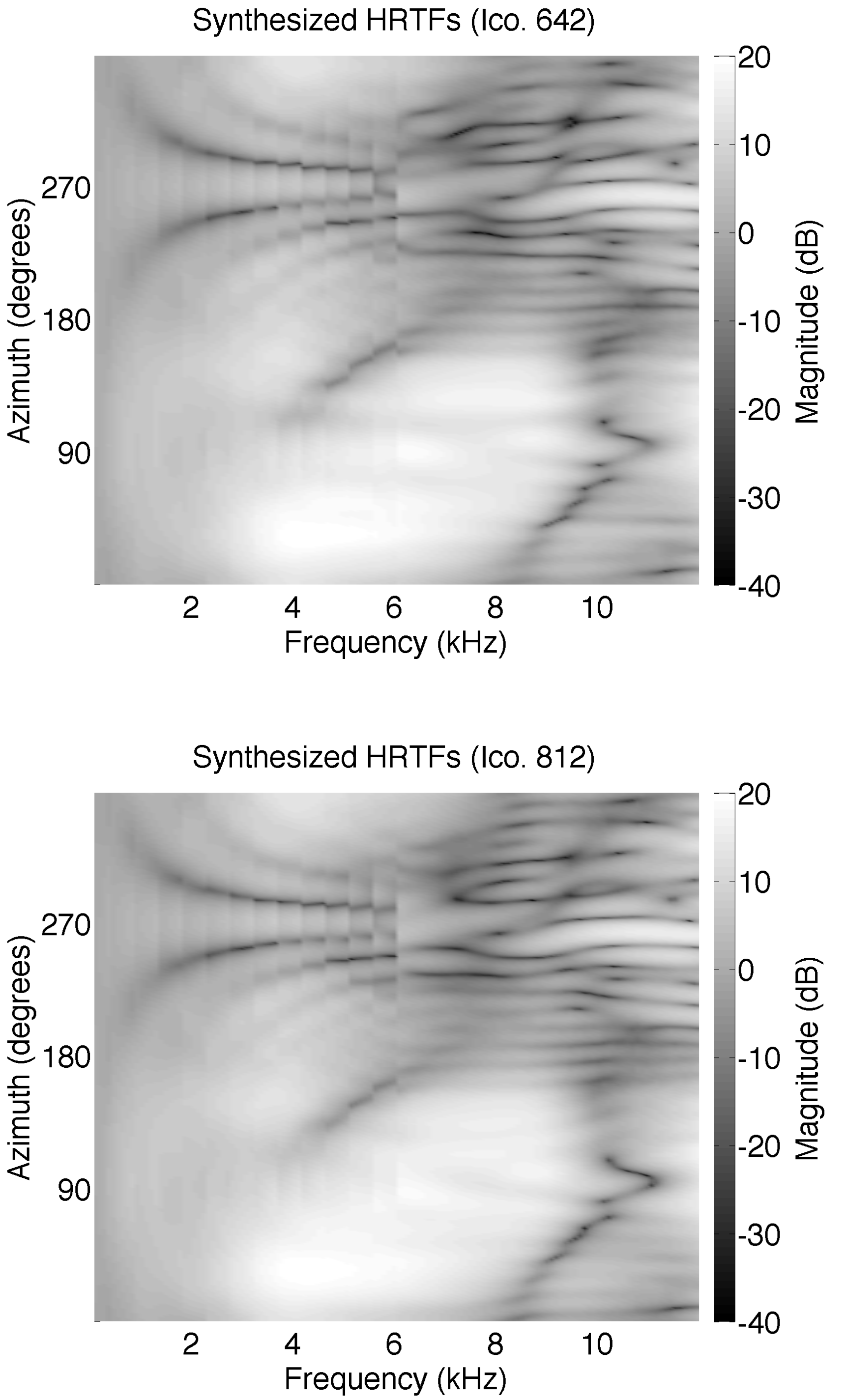


Figure 3.8: Transfer functions of the binaural synthesis method described by Eq. (3.2). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the icosahedral grids shown in Panels (c) and (d) of Figure 3.1, and spherical harmonics decompositions of order $N = 14$. A comparison with the target HRTFs in Figure 3.6 highlights the discontinuities along frequency due to the order limitation set by Eq. (3.4) and illustrated in Figure 3.5.

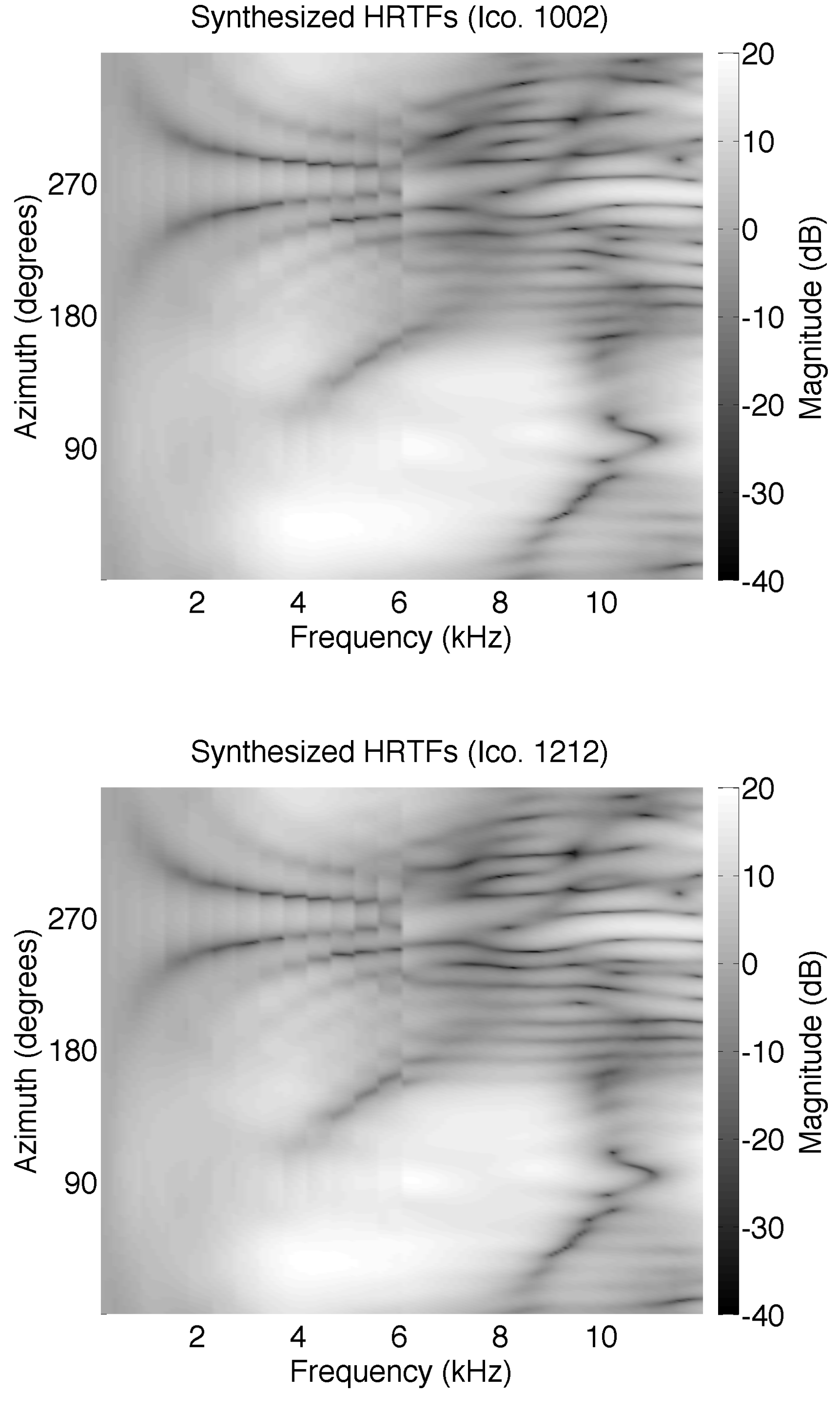
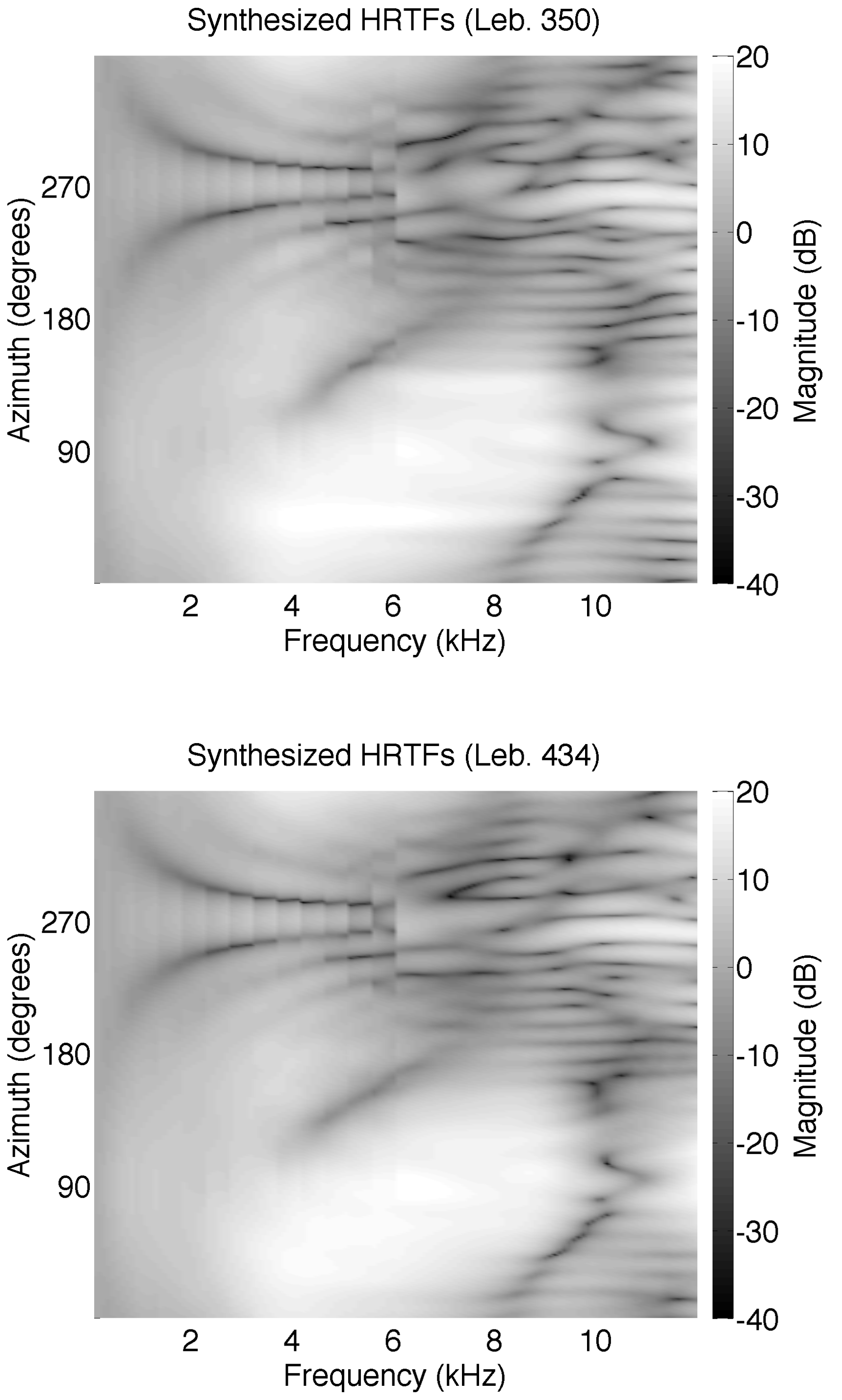


Figure 3.9: Transfer functions of the binaural synthesis method described by Eq. (3.2). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the icosahedral grids shown in Panels (e) and (f) of Figure 3.1, and spherical harmonics decompositions of order $N = 14$. A comparison with the target HRTFs in Figure 3.6 highlights the discontinuities along frequency due to the order limitation set by Eq. (3.4) and illustrated in Figure 3.5.



Synthesized HRTFs (Leb. 350)

Synthesized HRTFs (Leb. 434)

Figure 3.10: Transfer functions of the binaural synthesis method described by Eq. (3.2). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the Lebedev grids shown in Panels (a) and (b) of Figure 3.2, and spherical harmonics decompositions of order $N = 14$. A comparison with the target HRTFs in Figure 3.6 highlights the discontinuities along frequency due to the order limitation set by Eq. (3.4) and illustrated in Figure 3.5.

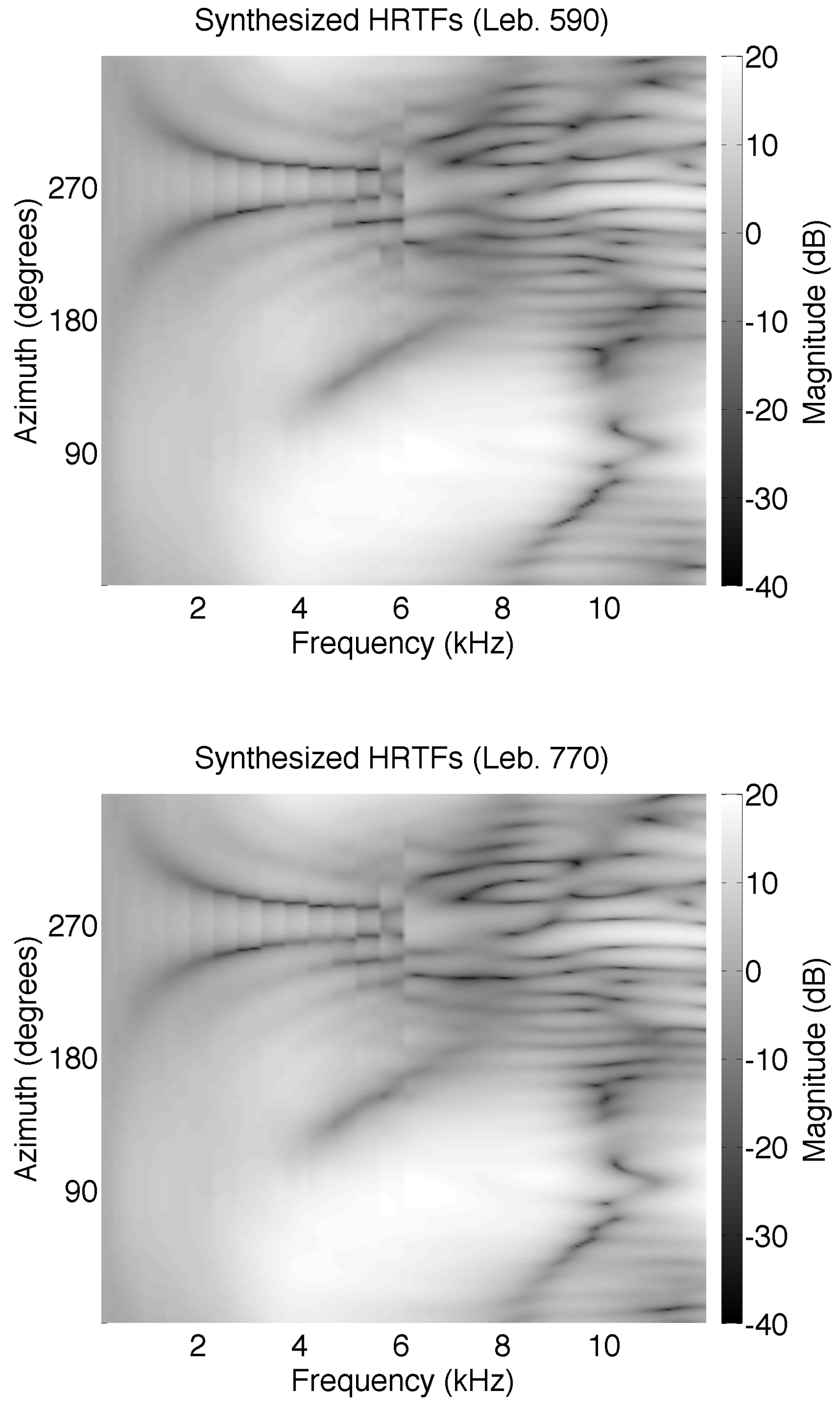
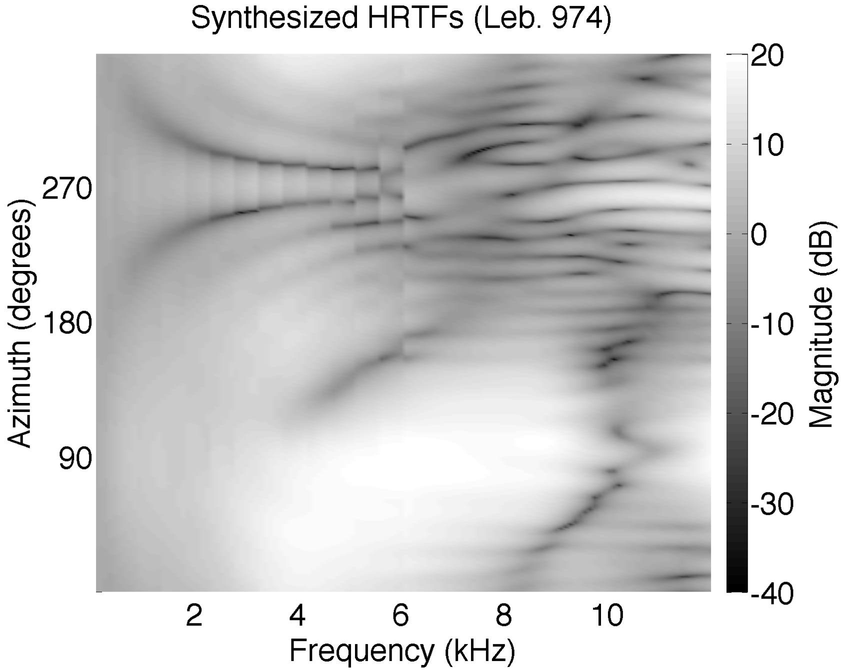


Figure 3.11: Transfer functions of the binaural synthesis method described by Eq. (3.2). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the Lebedev grids shown in Panels (c) and (d) of Figure 3.2, and spherical harmonics decompositions of order $N = 14$. A comparison with the target HRTFs in Figure 3.6 highlights the discontinuities along frequency due to the order limitation set by Eq. (3.4) and illustrated in Figure 3.5.



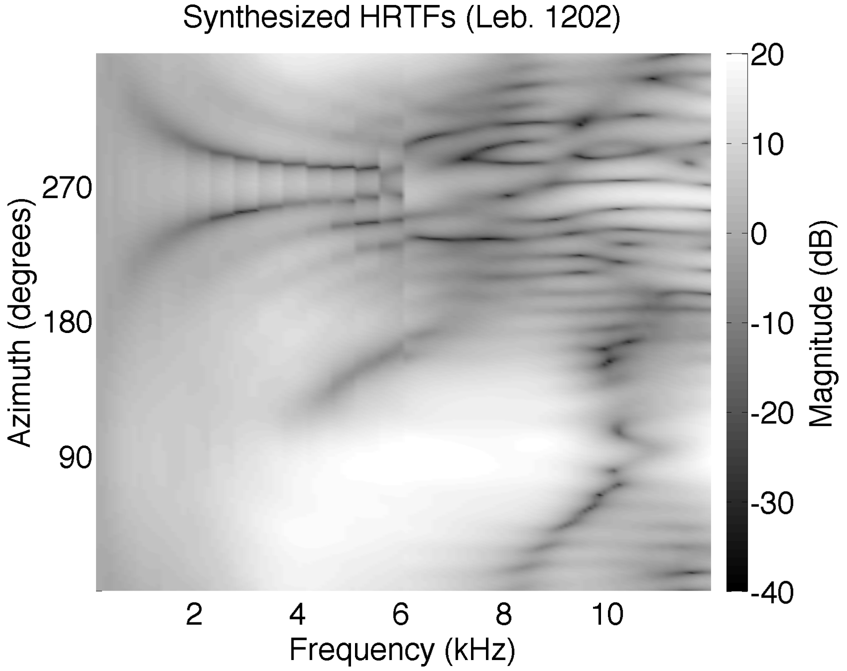
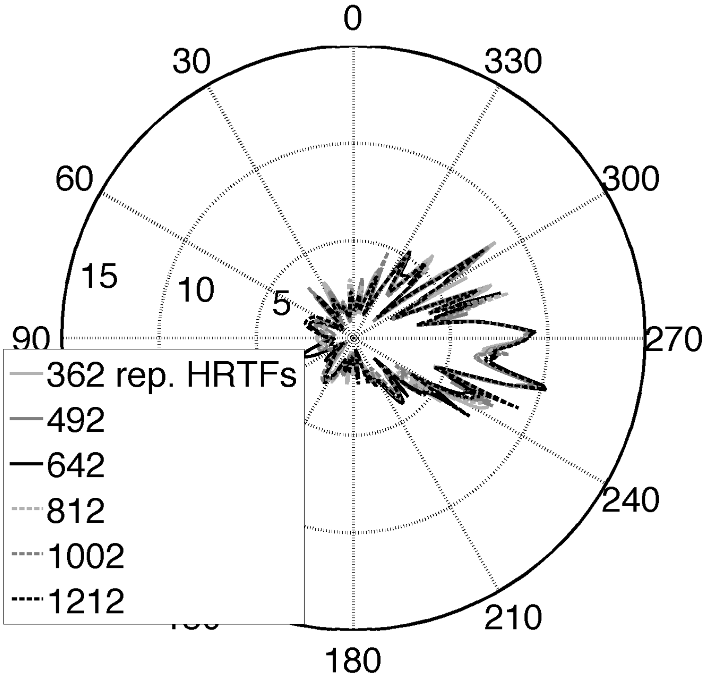
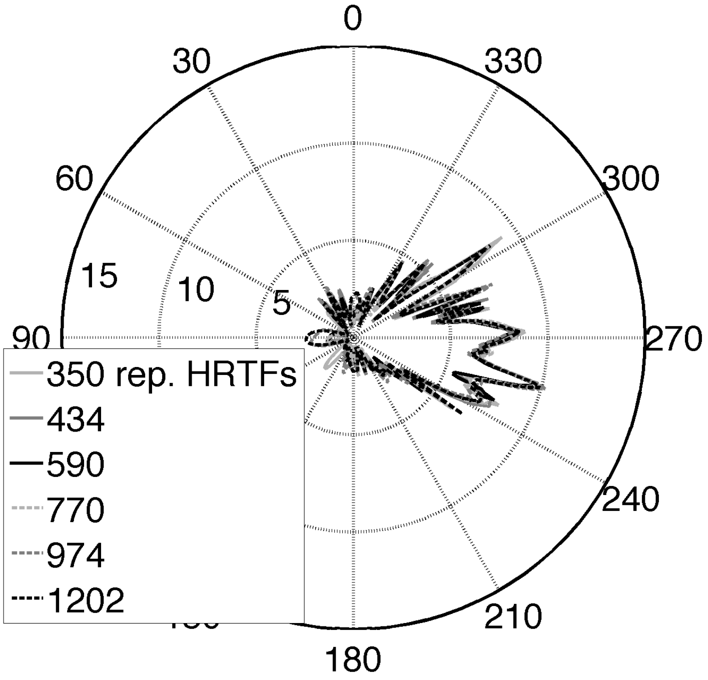


Figure 3.12: Transfer functions of the binaural synthesis method described by Eq. (3.2). We refer to these transfer functions as the synthesized HRTFs. They were computed for virtual loudspeakers arranged on the Lebedev grids shown in Panels (e) and (f) of Figure 3.2, and spherical harmonics decompositions of order $N = 14$. A comparison with the target HRTFs in Figure 3.6 highlights the discontinuities along frequency due to the order limitation set by Eq. (3.4) and illustrated in Figure 3.5.

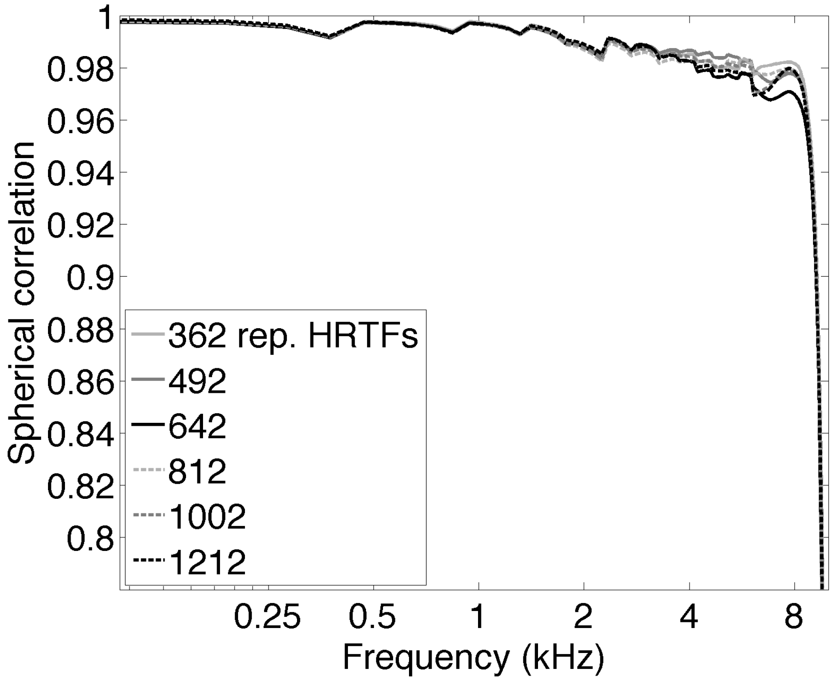


(a) Icosahedral grids.

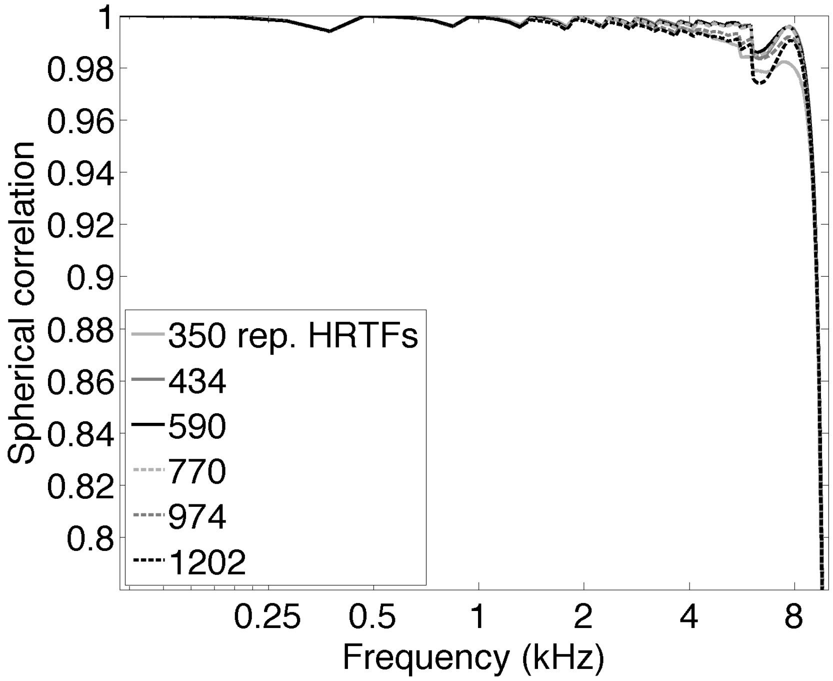


(b) Lebedev grids.

Figure 3.13: Numerical accuracy of the synthesis with virtual loudspeakers on (a) icosahedral and (b) Lebedev grids. The transfer functions in Figures 3.7 to 3.12 are compared with the target HRTFs in Figure 3.6, using the spectral distortion (SD) in Eq. (2.27). The SD is measured in dB.



(a) Icosahedral grids.



(b) Lebedev grids.

Figure 3.14: Numerical accuracy of the synthesis with virtual loudspeakers on (a) icosahedral and (b) Lebedev grids. The transfer functions in Figures 3.7 to 3.12 are compared with the target HRTFs in Figure 3.6 using the normalized spherical correlation in Eq. (2.28).

3.5 Summary

A method for the synthesis of high directional-resolution binaural signals using a limited number of microphones was proposed. The microphones were arranged on the surface of a spherical scatterer the size of an average human head. A limited set of HRTFs computed from virtual sources arranged over icosahedral grids around the listener was used. Spherical harmonics up to order 14 were used for the decomposition of the sound pressure field in the directions of microphones and for its reconstruction in the directions of virtual sources. A set of synthesized HRTFs on the horizontal plane, up to the spatial aliasing frequency around 9 kHz, was obtained with a logarithmic spectral distortion below 5 dB. The benefit of increasing the number of virtual sources was reduction of the low frequency distortion generated by the inversion of high order terms of the acoustic model for scattering from the rigid sphere.

Chapter 4

Extended binaural synthesis method for proximal sound sources

4.1 Introduction

This chapter extends the proposed method for the binaural synthesis of sound sources in the proximal region. The recorded sound field is now analyzed with multipoles, an extension to the spherical harmonics that allows to model the distance propagation of a sound field.

4.2 Preliminaries

4.2.1 General solutions to the wave equation: Multipoles

The wave equation is a physical model of the propagation of a wave in a medium. The medium must be considered as a continuum, that is, the atoms in a particle move at unison. The acoustic propagation of a sound wave in air can be characterized by a change of pressure δp that produces a displacement of particles $\delta\xi$, this displacement also produces a variation of density $\delta\rho$ that produces a change of pressure, and so on. Thus, the wave equation arises from the relation between the spatial variation of a movement equation ($\delta p \rightarrow \delta\xi$) and the temporal variation of a continuity equation ($\delta\xi \rightarrow \delta\rho$) by means of a state equation ($\delta\rho \rightarrow \delta p$). And hence, the behaviour of all these quantities, δp, $\delta\xi$ and $\delta\rho$, obey to the wave equation [54].

When the acoustic pressure field $\psi(r, \theta, \phi, k)$ is represented in the standard spherical coordinate system (r, θ, ϕ) for a wave number k, the wave equation reads [44]

$$\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial^2}{\partial r^2}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi} + k^2\right]\psi(r, \theta, \phi, k) = 0. \tag{4.1}$$

Considering that the spectral solutions are complex exponentials of k, we can focus our attention on the spatial solutions.

The method of separation of variables $\psi = R_{nk}(r)\Theta_{nm}(\theta)\Phi_m(\phi)$ is used to find the spatial solutions to Eq. (4.1). The general solution to the wave equation in spherical coordinates are the so-called multipoles, which are linear combinations of incoming and outgoing sound waves. The incoming waves are represented by the spherical Bessel functions of the first kind $j_n(kr)$ and the spherical Hankel functions of the first kind $h_n^{(1)}(kr)$. The outgoing waves are represented by the spherical Bessel functions of the second kind $y_n(kr)$ and the spherical Hankel functions of the second kind $h_n^{(2)}(kr)$.

The multipole expansion is therefore defined by [44]

$$\psi(r, \theta, \phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(a_{nm}j_n(kr) + b_{nm}y_n(kr)\right)Y_{nm}(\theta, \phi), \tag{4.2}$$

for standing wave type solutions, and by [44]

$$\psi(r, \theta, \phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(c_{nm}h_n^{(1)}(kr) + d_{nm}h_n^{(2)}(kr)\right)Y_{nm}(\theta, \phi), \tag{4.3}$$

for traveling wave solutions. In analogy with Eq. (2.13), the linear combination coefficients a_{nm}, b_{nm}, c_{nm} and d_{nm} are referred to as the multipole spectrum.

4.2.2 Distance propagators based on acoustic holography

Acoustic holography [44] is a method to estimate the sound field on a surface of radius r by measuring acoustic parameters on a different enclosing surface of radius b. Consider the outgoing wave component in Eq. (4.3) of the general solution to the wave equation in Eq. (4.1), evaluated at the measurement surface of radius b

$$\psi(b, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm} h_n(kr) Y_{nm}(\Omega). \tag{4.4}$$

Multiplying by $Y_{nm}(\Omega)$, integrating the product on the unit sphere as in Eq. (2.9), and using the property of orthonormality of the spherical harmonics in Eq. (2.10), the spherical wave spectrum c_{nm} results in

$$c_{nm} = \frac{\psi_{nm}(b)}{h_n(kb)}. \tag{4.5}$$

Evaluating Eq. (4.5) at a different surface of radius r, and inserting the resulting expression into the outgoing component of Eq. (4.3), the sound field at a radius r reads

$$\psi(r, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left(\frac{h_n(kr)}{h_n(kb)} \psi_{nm}(b) \right) Y_{nm}(\Omega). \tag{4.6}$$

The result in Eq. (4.6) can be generalized to work in both the far and near fields by a normalization of the spherical Hankel functions so that they do not tend to zero at infinity. The normalized spherical Hankel functions are defined by [20]

$$R_n(kr) = \frac{i^{n+1} kr}{e^{-ikr}} h_n(kr), \tag{4.7}$$

which tends to 1 when r tends to infinity.

The normalized spherical Hankel functions, together with the term in parentheses in Eq. (4.6), leads to the generalized distance propagator in the spherical harmonic domain

$$\psi_{nm}(r) = \left(\frac{r h_n(kr)}{b h_n(kb)} \right) \psi_{nm}(b). \tag{4.8}$$

4.2.3 Characterization of HRTFs using the principle of reciprocity

When sound propagates from the source to the listener, the received sound at the listener's ears is transformed by the structure and shape of the listener's body. To represent the sound pressure at the ears of the listener, two sources should be taken into account: one is the original acoustic source from the speaker and the other is the secondary source due to the scattering of human body. It is a complicated problem to apply the wave equation in this configuration because the receiver, the listener's ear, is within the scatterer region of human body. The acoustic principle of reciprocity [44] can be used to remove this difficulty and to develop a general representation of the HRTFs.

To apply the principle of reciprocity to the analysis of HRTFs [20], it is assumed that the original acoustic source is located at the listener's ear, and the microphones are assumed to be placed some distance away [3]. Here, all the scattering sources of human body as the secondary level sources with the original sources at the listener's ear together constituting the source field. From Huygens' principle, the sum of the waves from all the sources (including both original and secondary sources) to any point beyond the scatterers (the human body) can be calculated by integration or numerical modeling.

The HRTFs can thus be reformulated as an acoustic radiation problem [21]. When the volume velocity at the entrance of the ear canal is known, the radiated pressure field is completely defined outside a region that encloses all contributing scattering sources. It can be calculated from the Helmholtz equation assuming that the Sommerfeld radiation condition is fulfilled. Therefore, the sound pressure field on the enclosing sphere is completely define only if it large enough to enclose the listener's head, and also the listener's torso in case it was also consider in the measurements.

[3] The principle of reciprocity has also inspired a fast and efficient method to measure HRTFs by using two miniature speakers on the listener's ears and a surrounding array of microphones [55]. However, miniature speakers with a signal to noise ratio of at least 20 dB, at frequencies below 400 Hz, need still to be developed, so as to enhance the measurements for low frequencies [56].

4.3 Synthesis of proximal sound sources

The distance propagators in Eq. (4.8) allow for the embedding of source distance information on the binaural synthesis method derived in Chapter 3. Traditional representations of sound fields based on the spherical harmonics decomposition described by Eqs. (2.13) and (2.14) do not include the sound source distance information. The multipole expansions in Eqs. (4.2) and (4.3), though, can accurately encode the distance of a sound source. However, a spherical spectrum cannot be easily converted into a multipole spectrum, even using the distance compensation described by Eq. (4.8), since this requires establishing the reference distance b during recording. Furthermore, the multipole decomposition seek for accurate reproduction, while some recordings may be enhanced by making sounds appear closer or farther. Therefore, an extended method to synthesize the binaural signals for arbitrary distances is described in this section.

4.3.1 Extended method

We propose a method to synthesize the binaural signals for sound sources in the proximal region from sound fields represented by spherical harmonics (see Figure 1.8). We rely on HRTFs for representative sound source positions on the distal region. The radius of the corresponding virtual loudspeaker array is scaled to match an inputted desired distance using the acoustic holography propagators in Eq. (4.8). The spherical spectrum of the recordings is decoded for the virtual loudspeaker array, and the binaural signals are rendered using the distance-compensated HRTFs. This novel proposal is also an important step towards the re-creation of the binaural signals for sound sources in the proximal region, which would demand to consider both the recording of near field sound sources and the synthesis of HRTFs for sound sources in the proximal region.

Consider a finite number of microphones at points $\mathbf{a}_q$, for $q = 1, ..., Q$, on the surface of the rigid sphere illustrated in Figure 4.1. As in Chapter 3, the integral in Eq. (2.20) must be approximated with a sum over q. After using the sum in Eq. (2.12) with the resulting sum, Eqs. (2.19) and (2.20) become the transfer functions $\hat{H}$ of the proposed method to synthesize the binaural signals from the recordings made with a rigid spherical microphone array and a set of representative HRTFs:

$$\hat{H}(\mathbf{r}, k) = \sum_{v=1}^{V} \alpha_v H(\mathbf{b}_v, k) \sum_{n=0}^{N} (2n+1) B_n \sum_{q=1}^{Q} P_n(\cos\Theta_{vq}) \beta_q S(\mathbf{r}, \mathbf{a}_q, k), \tag{4.9}$$

where Θ_{vq} is the angle between the virtual source at $\mathbf{b}_v$ and the microphone at $\mathbf{a}_q$, and the quadrature weights β_q applied to the individual microphone signals are chosen in such way that [36]

$$\sum_{q=1}^{Q} \beta_q Y_{nm}(\Omega_q) Y^*_{n'm'}(\Omega_q) = \delta_{n-n'} \delta_{m-m'}. \tag{4.10}$$

The spherical beamformers B_n are now defined by

$$B_n(r, a, b, k) = -(ka)^2 h'_n(ka) \times \frac{1}{h_n(kb)} \times \left[\frac{r h_n(kr)}{b h_n(kb)} \right], \tag{4.11}$$

where its three factors, from left to right, are intended to remove the scattering effects from the rigid spherical measurement surface of radius a [35,36,47], to backpropagate the pressure field from the center of the array to the radiating surface of radius b [44], and to propagate the sound pressure field to a closer or farther distance.

4.3.2 Matrix formulation

The extended binaural synthesis method in Eq. (4.9), for sound sources in both the distal and proximal regions, from the spherical harmonic analysis of compact microphone arrays recordings and a set of representative HRTFs, can conveniently be expressed in terms of matrix multiplications.

For each ear, each sound sound source position, and each frequency bin, the following expressions represent the transfer functions $\hat{H}$ of the proposed method:

$$\hat{H} = \sum_{v=1}^{V} W_v H_v, \tag{4.12}$$

for H_v the HRTF associated to the v-th virtual loudspeaker, and its corresponding weighting coefficient W_v defined by

$$W_v = \mathbf{F}\mathbf{P}_v\mathbf{S}, \tag{4.13}$$

which depend on the spherical microphone array signals

$$\mathbf{S} = \begin{bmatrix} S_1 \\ \vdots \\ S_Q \end{bmatrix}_{Q\times 1}, \tag{4.14}$$

the directivity patterns matching the arrays of microphones and virtual loudspeakers

$$\mathbf{P}_v = \begin{bmatrix} P_0(\Theta_{v1}) & \cdots & P_0(\Theta_{vQ}) \\ \vdots & \ddots & \vdots \\ P_N(\Theta_{v1}) & \cdots & P_N(\Theta_{vQ}) \end{bmatrix}_{(N+1)\times Q}, \tag{4.15}$$

and the distance-compensation filters

$$\mathbf{F} = -ka^2 \begin{bmatrix} \frac{h_0'(ka)}{h_0(kb)} \left(\frac{rh_0(kr)}{bh_0(kb)}\right) & \cdots & (2N+1)\frac{h_N'(ka)}{h_N(kb)} \left(\frac{rh_N(kr)}{bh_N(kb)}\right) \end{bmatrix}_{1\times(N+1)}. \tag{4.16}$$

4.4 Numerical accuracy of the synthesized HRTFs

4.4.1 Parameters and conditions for the evaluation

The parameter under evaluation was the number of representative sound sources and the distance of sound sources. Its effect on the accuracy of the synthesis was evaluated. The representative sound sources were arranged on icosahedral and Lebedev grids (see Figures 3.1 and 3.3). The center of the rigid spherical microphone array was set as the reference position, and its radius was chosen to be 8.5 cm. The recorded microphone signals were generated with the model of acoustic scattering from the rigid sphere in Eq. (2.15), which was computed for 360 sources at a 1.5 m distance from the reference position, equiangularly distributed on the horizontal plane. The target HRTFs (See Figures 4.2 to 4.4) for 360 sources at different distances, equiangularly distributed on the horizontal plane, were computed numerically for a dummyhead using a BEM solver [4]. Other HRTFs for sets of representative sound sources at a 1.5 m distance, arranged on icosahedral and Lebedev grids were also computed with the BEM solver.

The evaluation based on a discrete measurement surface was performed. An array of $Q = 252$ microphones distributed on a icosahedral grid over the surface of the rigid sphere of 8.5 cm radius was assumed for this purpose, which is the available setup at the Research Institute of Electrical Communication, Tohoku University [11]. The number of microphones $Q = 252$ imposed a spatial bandwidth or order $N = 14$, and therefore the accuracy could be evaluated up to a spatial aliasing frequency of around 8 kHz. The model of the acoustic scattering from the rigid sphere was decomposed up to order 14 at the positions of the microphones, and reconstructed at the positions of the virtual loudspeakers. The resultant signals were downmixed to a binaural signal. This process is formulated in Eq. (4.9).

4.4.2 Simulation results

Figures 4.2 to 4.4 show the HRTFs for sound sources on the distal (3.0 m, 1.5 m and 1.0 m) and proximal (0.75 m, 0.50 m and 0.25 m) regions. Compared with distal HRTFs, the proximal HRTFs show a slight amplification on intensity on the ipsilateral side, and a slight attenuation on the contralateral side. These HRTFs are the desired output for the evaluation of the accuracy of the proposed method.

Figure 4.5 to 4.10 shows the spectral distortion between the target HRTFs (see Figures 4.2 to 4.4) and the HRTFs synthesized with representative sound sources on icosahedral (top Panels) and Lebedev (bottom Panels) grids. The synthesis of sound sources in the proximal region, on the opposite side of the left ear, shows that the accuracy of the synthesis tends to decrease as the sound sources approaches the listener's head.

Figure 4.11 and 4.16 shows the normalized spherical correlation between the target

HRTFs (see Figures 4.2 to 4.4) and the HRTFs synthesized with representative sound sources on icosahedral (top Panels) and Lebedev (bottom Panels) grids. The synthesis of sound sources in the proximal region, at frequencies up to 1 kHz, shows that the accuracy of the synthesis tends to decrease with both decreasing frequency and distance.

The previous results suggest that the accuracy along azimuth should be based on the mean value of the spectral distortion for sound sources on the ipsilateral and contralateral sides. They also suggest to evaluate the accuracy along frequency by computing the mean value of the spherical correlation over two frequency bands: below 1 kHz (low frequencies) and between 1 kHz and 8 kHz (high frequencies up to the limit imposed by the finite number of microphones).

The synthesis accuracy for sources on the ipsilateral side, shown in Figure 4.19, shows that the accuracy remains almost constant when the synthesis is based on representative sound sources arranged on icosahedral grids. However, when using the Lebedev grids, the use of 770 representative sound sources improves the accuracy at all distances. More important is the evaluation of contralateral sound sources, whose accuracy, in general, clearly decreases as the sound source approaches the listener's head. Moreover, as for the number and arrangement of the representative sound sources, Figure 4.20 shows that the best accuracies correspond to 1002 points on icosahedral grids and 590 points on Lebedev grids.

Therefore, Figure 4.17 shows the mean value of the spherical correlation at low frequencies. This plot shows that the accuracy of the synthesis is not affected by the number of representative sound sources, but it still decreases monotonically as the sound sources approaches the listener's head. On the other hand, at high frequencies, as shown by Figure 4.18, the accuracy does not depend on distance nor frequency, although a slightly improved accuracy is obtained when using 642 sound sources arranged on icosahedral grids.

Finally, Figures 4.21 and 4.22 show the distal and proximal HRTFs, which correspond to the best accuracies obtained with both icosahedral and Lebedev grids.

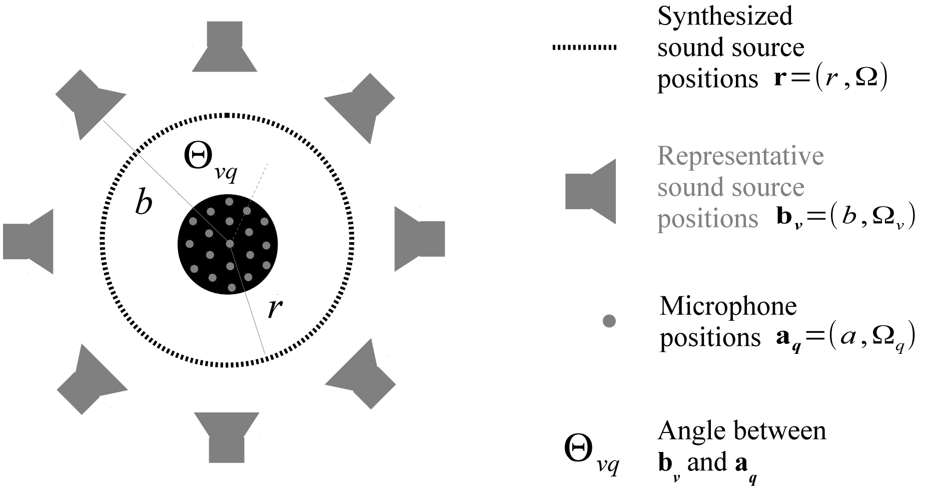
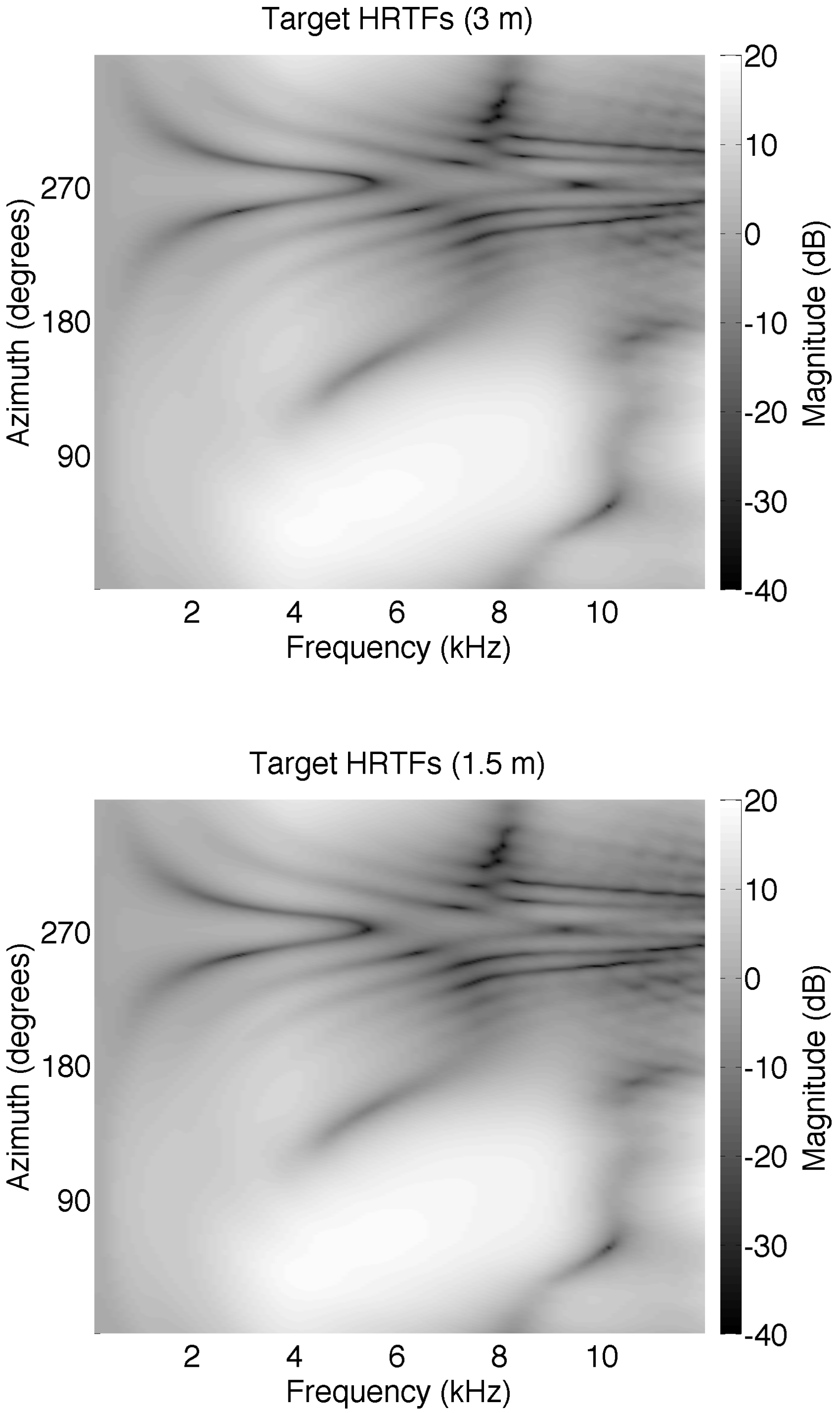

Θ_{vq}
b
r
Synthesized sound source positions $\mathbf{r}=(r,\Omega)$
Representative sound source positions $\mathbf{b}_v=(b,\Omega_v)$
Microphone positions $\mathbf{a}_q=(a,\Omega_q)$
Θ_{vq} Angle between $\mathbf{b}_v$ and $\mathbf{a}_q$

Figure 4.1: Geometry used on the synthesis method of binaural signals for sound sources on the proximal region. The sound field due to a sound source is measured by the rigid spherical microphone array of radius a, and then analyzed with spherical harmonics up to order N. The rendering assumes a virtual array of loudspeakers placed at the representative positions $\mathbf{b}_v$ (see Figure 1.2). The radius of the loudspeaker array is scaled so as to to match an inputted desired distance (see Figure 1.8). The synthesized sound sources can therefore lie at the same radial distance of the loudspeakers, or farther or nearer.



Target HRTFs (3 m)

Target HRTFs (1.5 m)

Figure 4.2: Distance dependence of the left ear HRTFs. They do not vary for sound sources on the distal region.

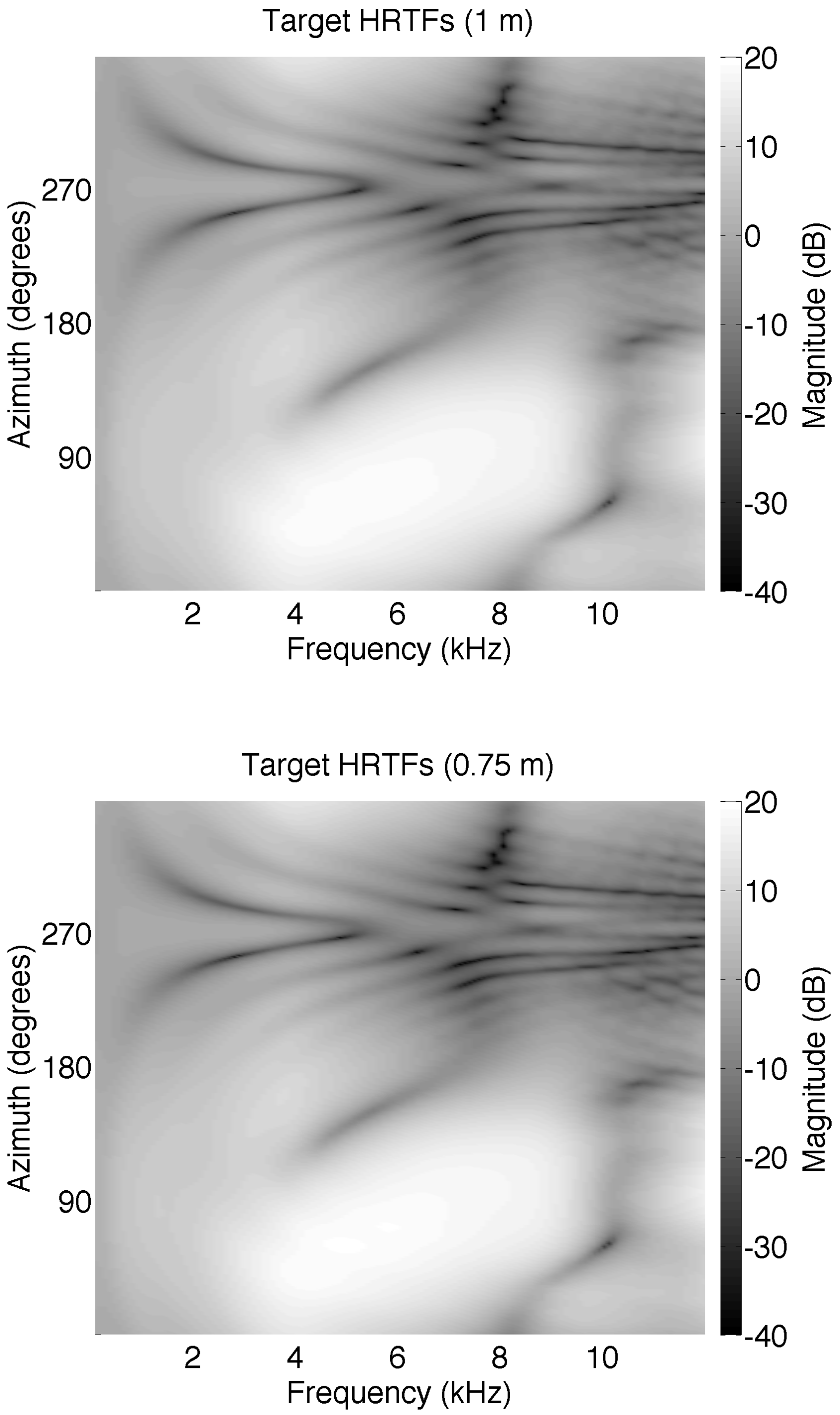
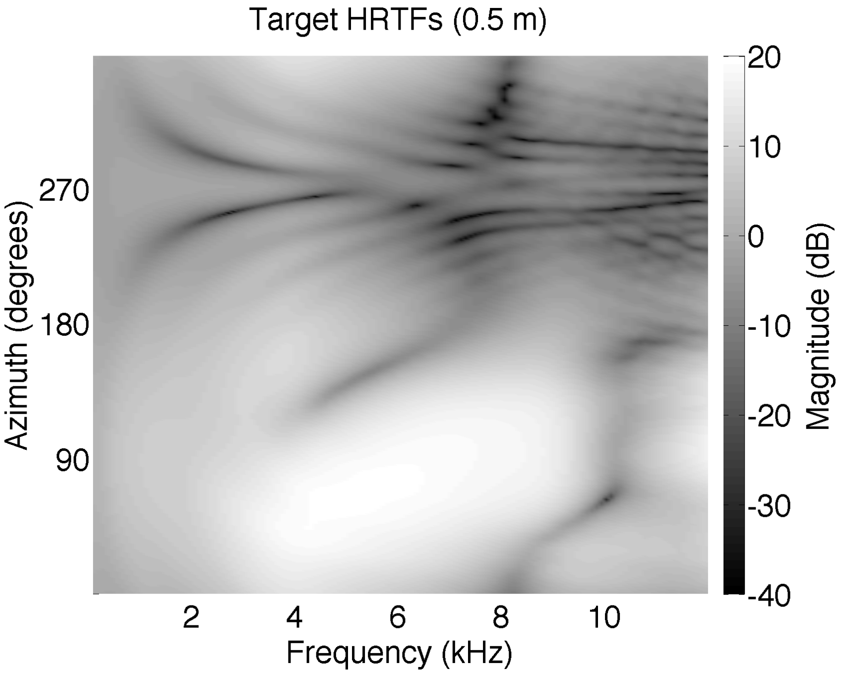


Figure 4.3: Distance dependence of the left ear HRTFs. The intensity slightly starts to increase on the ipsilateral side.



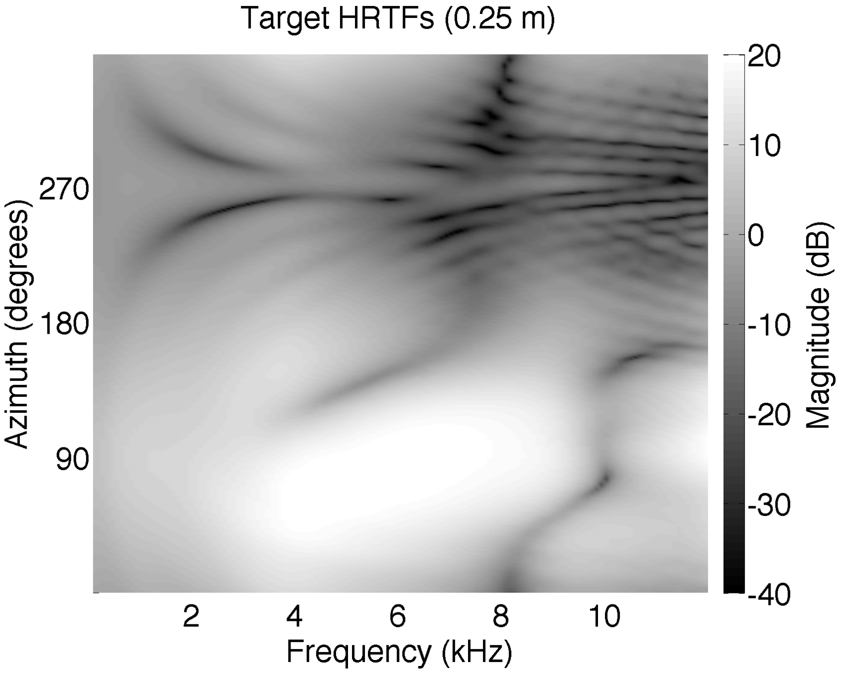
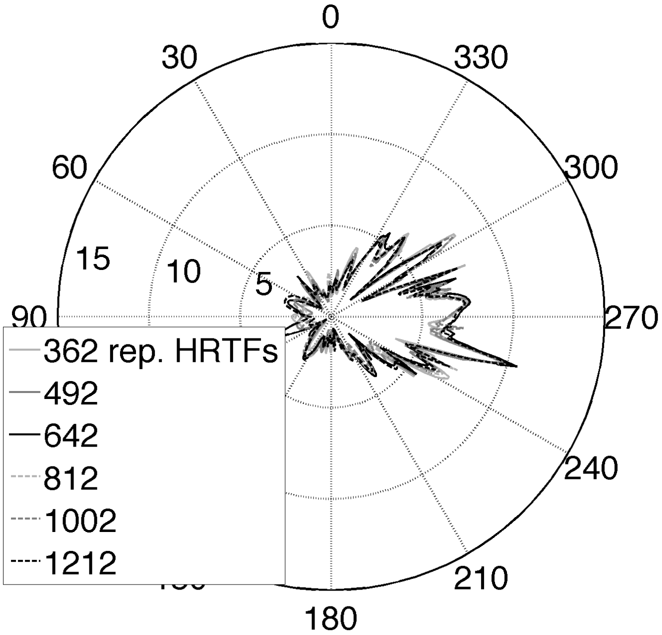
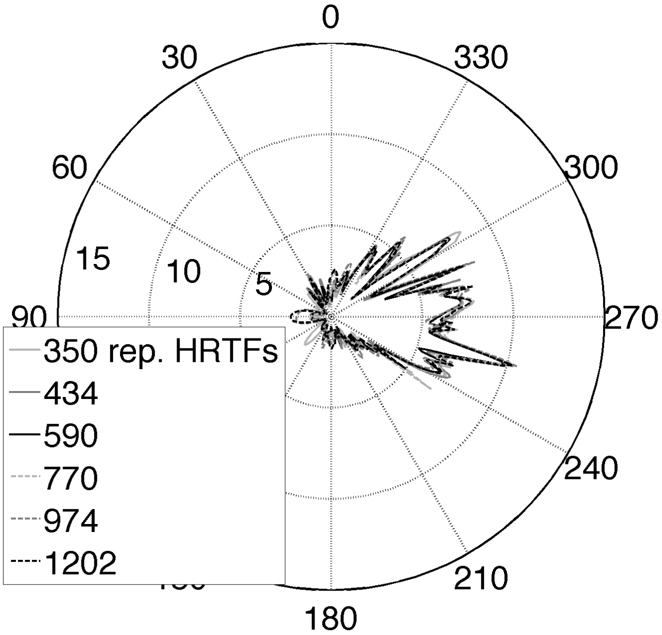


Figure 4.4: Distance dependence of the left ear HRTFs. For sound sources in the proximal region, the intensity increases with decreased distance on the ipsilateral side, and the shadowing effect of the head is attenuated on the contralateral side.

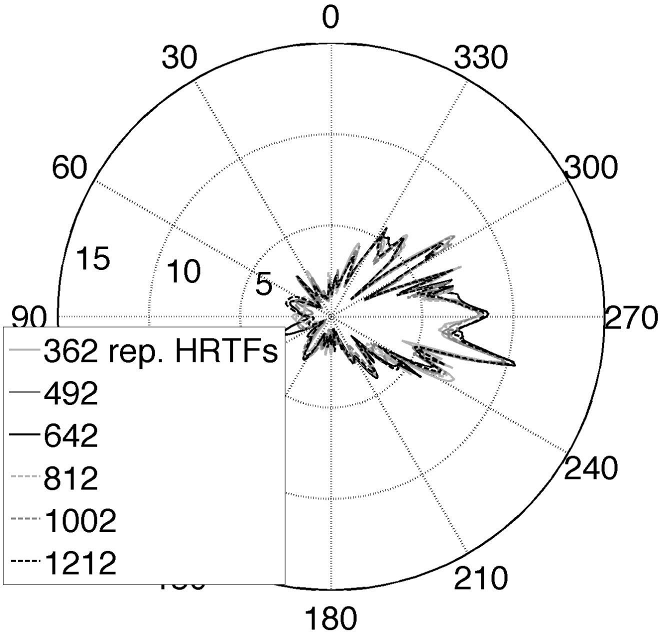


(a) Synthesis at 3 m and virtual loudspeakers on icosahedral grids.

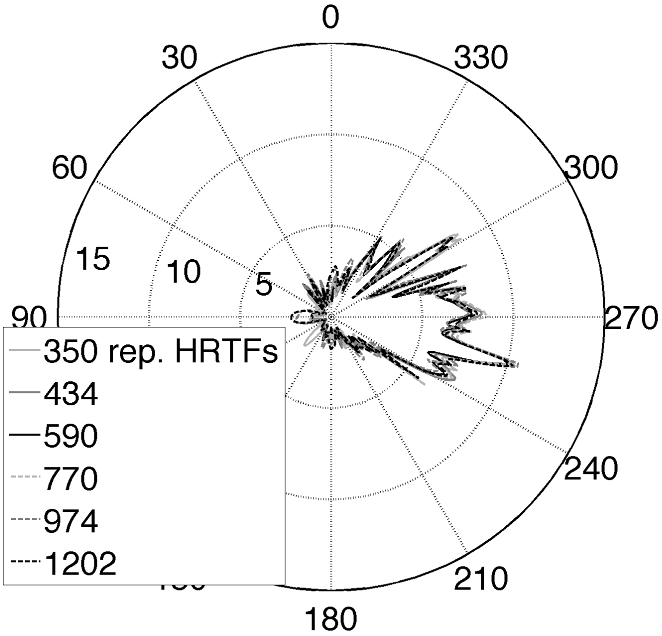


(b) Synthesis at 3 m and virtual loudspeakers on Lebedev grids.

Figure 4.5: Spectral distortion (dB) between the transfer functions of the binaural synthesis method described by Eq. (4.9) and the target HRTFs in the top Panel of Figure 4.2. The synthesis is based on virtual loudspeakers arranged on (a) icosahedral and (b) Lebedev grids. The loudspeakers driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

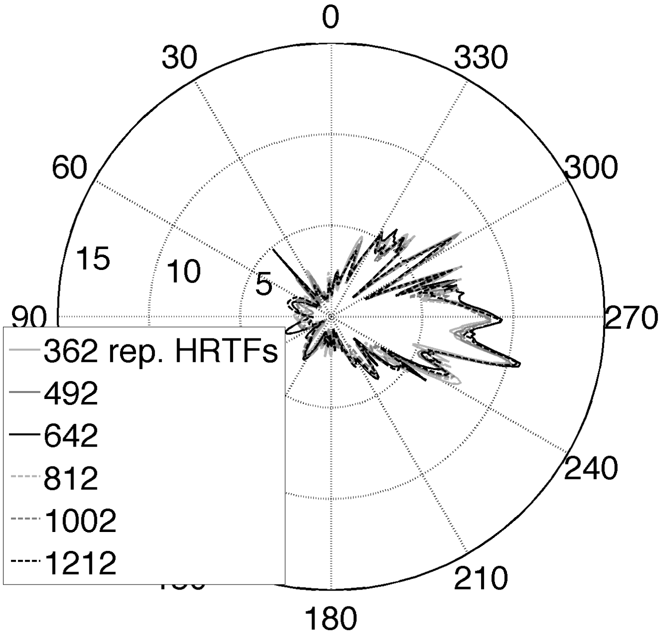


(a) Synthesis at 1.5 m and virtual loudspeakers on icosahedral grids.

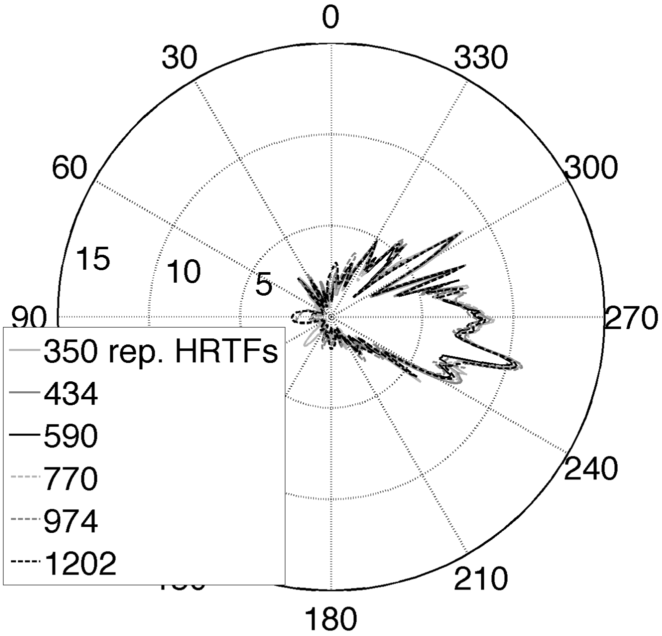


(b) Synthesis at 1.5 m and virtual loudspeakers on Lebedev grids.

Figure 4.6: Spectral distortion (dB) between the transfer functions of the binaural synthesis method described by Eq. (4.9) and the target HRTFs in the bottom Panel of Figure 4.2. The synthesis is based on virtual loudspeakers arranged on (a) icosahedral and (b) Lebedev grids. The loudspeakers driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

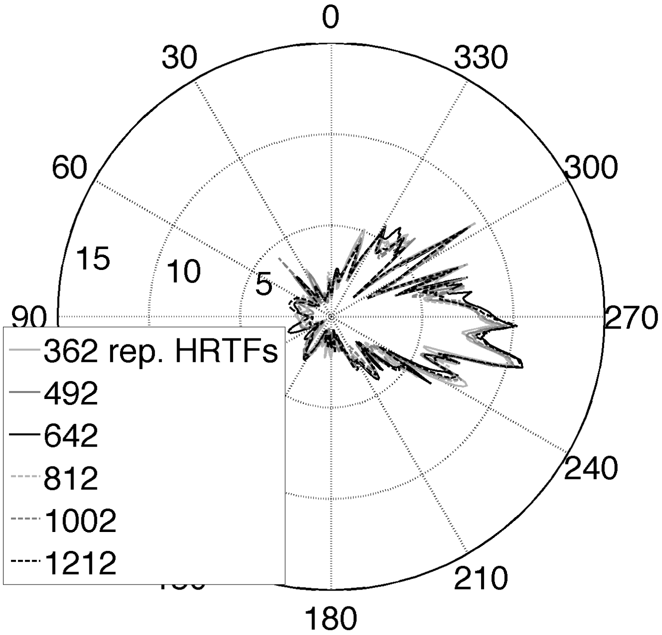


(a) Synthesis at 1 m and virtual loudspeakers on icosahedral grids.

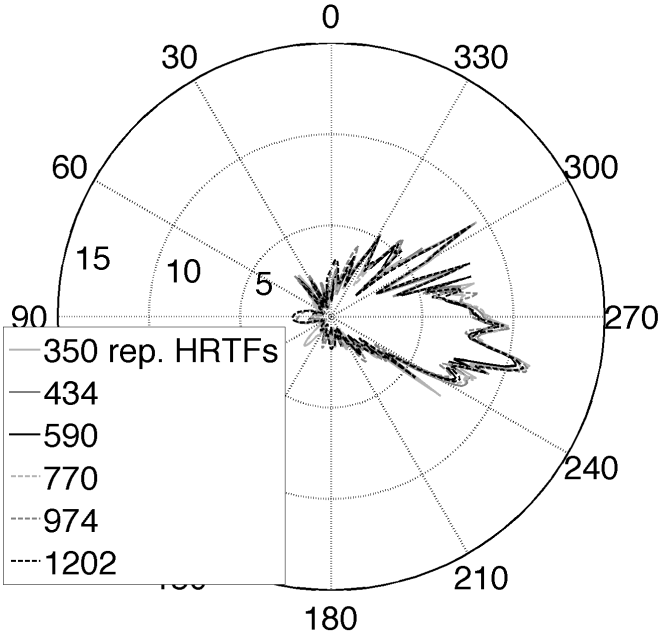


(b) Synthesis at 1 m and virtual loudspeakers on Lebedev grids.

Figure 4.7: Spectral distortion (dB) between the transfer functions of the binaural synthesis method described by Eq. (4.9) and the target HRTFs in the top Panel of Figure 4.3. The synthesis is based on virtual loudspeakers arranged on (a) icosahedral and (b) Lebedev grids. The loudspeakers driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

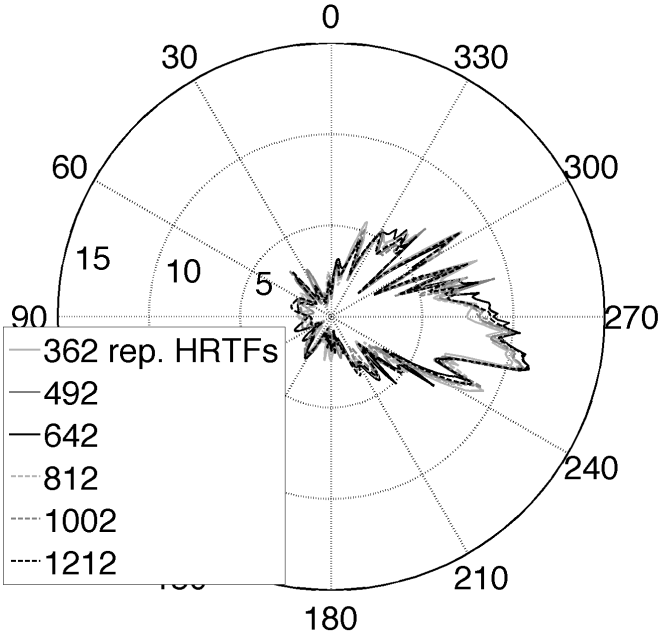


(a) Synthesis at 0.75 m and virtual loudspeakers on icosahedral grids.

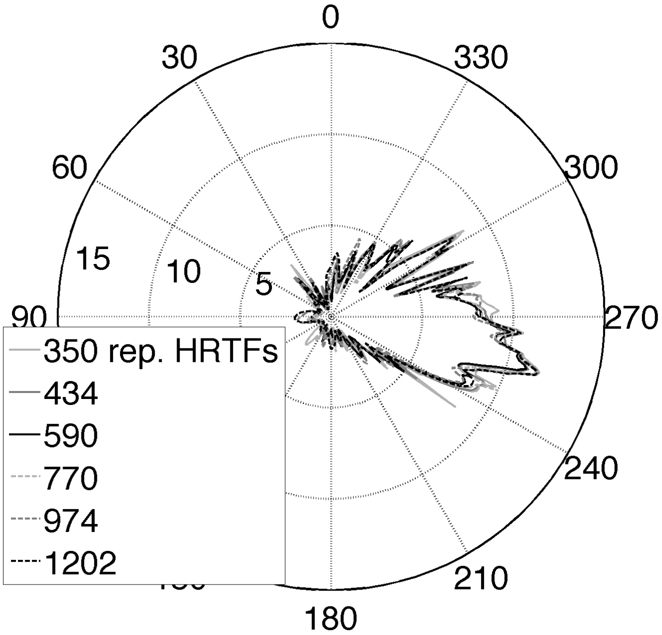


(b) Synthesis at 0.75 m and virtual loudspeakers on Lebedev grids.

Figure 4.8: Spectral distortion (dB) between the transfer functions of the binaural synthesis method described by Eq. (4.9) and the target HRTFs in the bottom Panel of Figure 4.3. The synthesis is based on virtual loudspeakers arranged on (a) icosahedral and (b) Lebedev grids. The loudspeakers driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

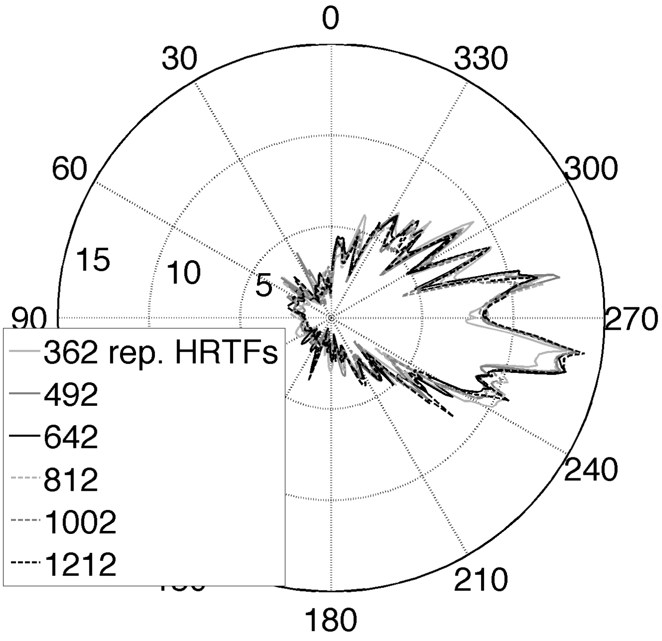


(a) Synthesis at 0.5 m and virtual loudspeakers on icosahedral grids.

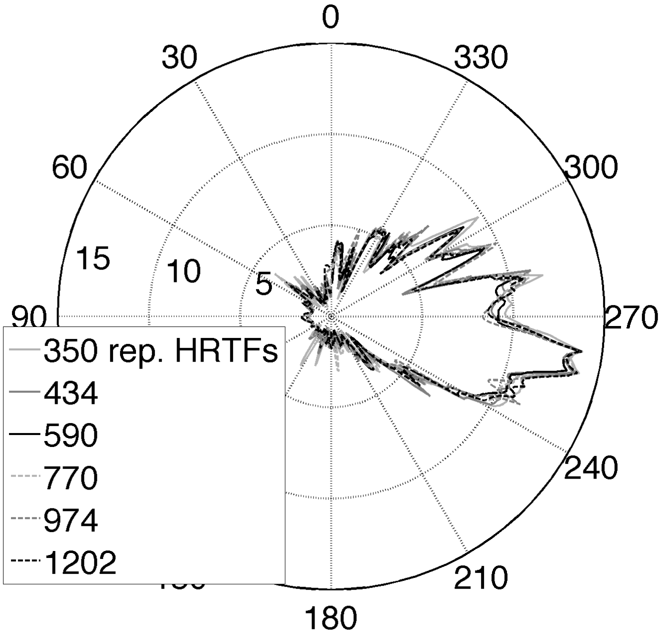


(b) Synthesis at 0.5 m and virtual loudspeakers on Lebedev grids.

Figure 4.9: Spectral distortion (dB) between the transfer functions of the binaural synthesis method described by Eq. (4.9) and the target HRTFs in the top Panel of Figure 4.4. The synthesis is based on virtual loudspeakers arranged on (a) icosahedral and (b) Lebedev grids. The loudspeakers driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

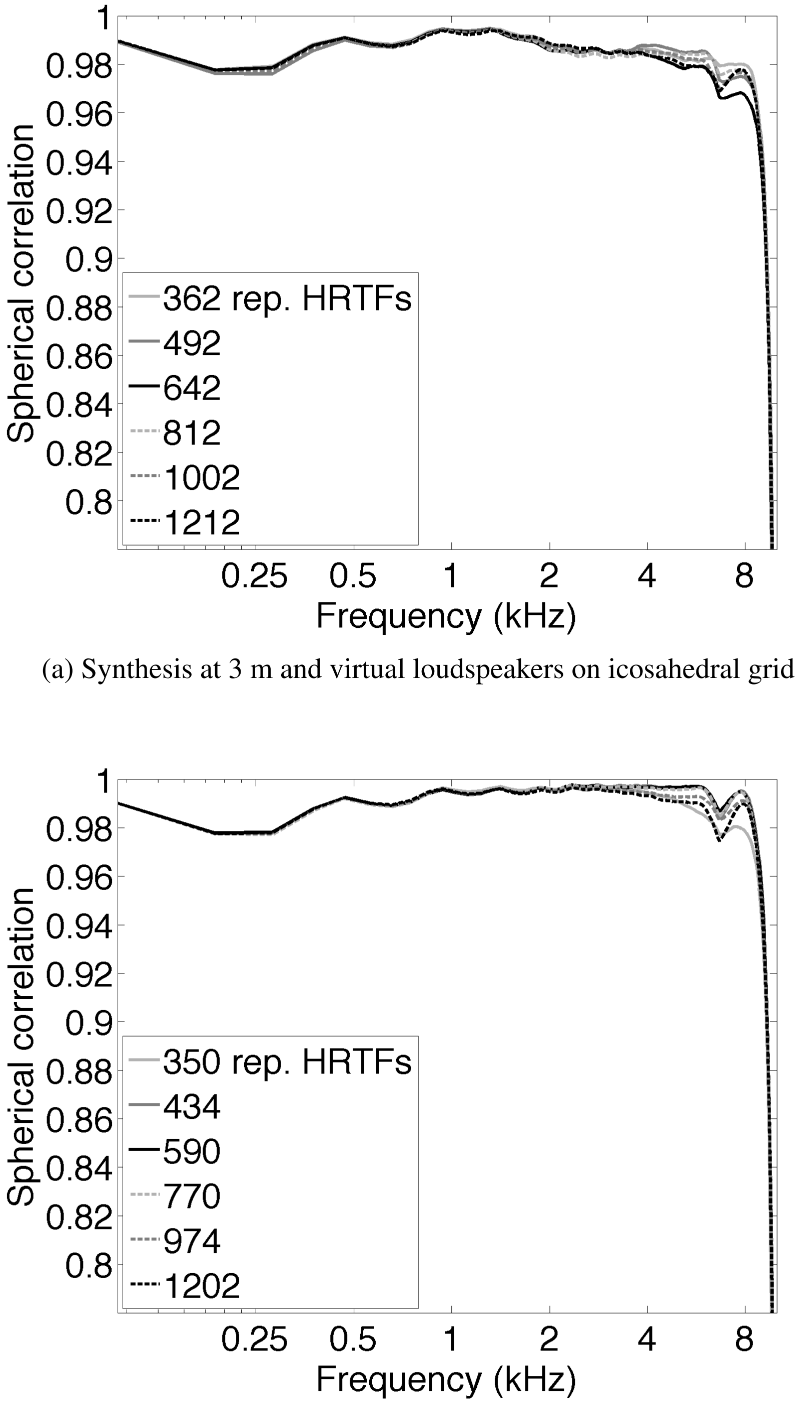


(a) Synthesis at 0.25 m and virtual loudspeakers on icosahedral grids.



(b) Synthesis at 0.25 m and virtual loudspeakers on Lebedev grids.

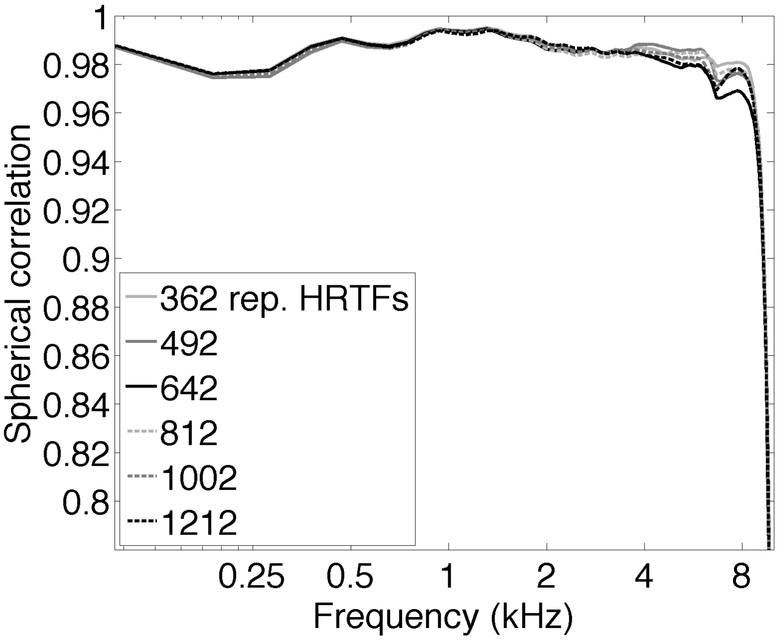
Figure 4.10: Spectral distortion (dB) between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the bottom Panel of Figure 4.4. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. An array of 252 microphones on a rigid sphere of 8.5 cm radius was assumed. A comparisson with Figures 4.5 to 4.9 shows that, for sound sources in the proximal region on the contralateral side, the accuracy of the synthesis tends to decrease with decreased distance.



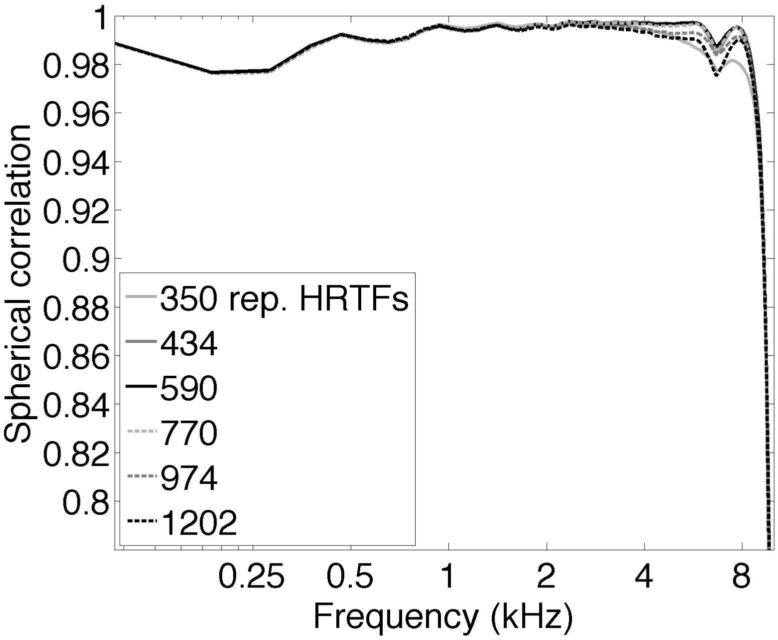
(a) Synthesis at 3 m and virtual loudspeakers on icosahedral grids.

(b) Synthesis at 3 m and virtual loudspeakers on Lebedev grids.

Figure 4.11: Normalized spherical correlation between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the top Panel of Figure 4.2. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. The loudspeaker driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

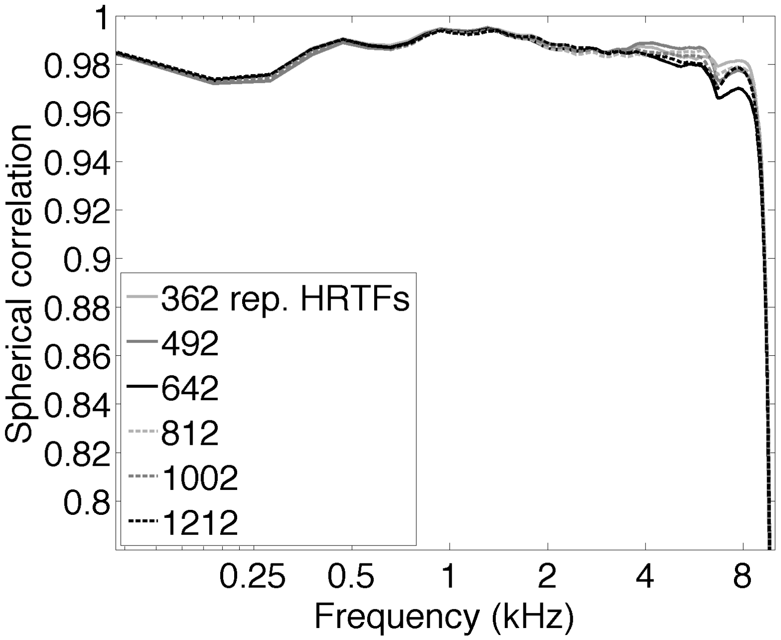


(a) Synthesis at 1.5 m and virtual loudspeakers on icosahedral grids.

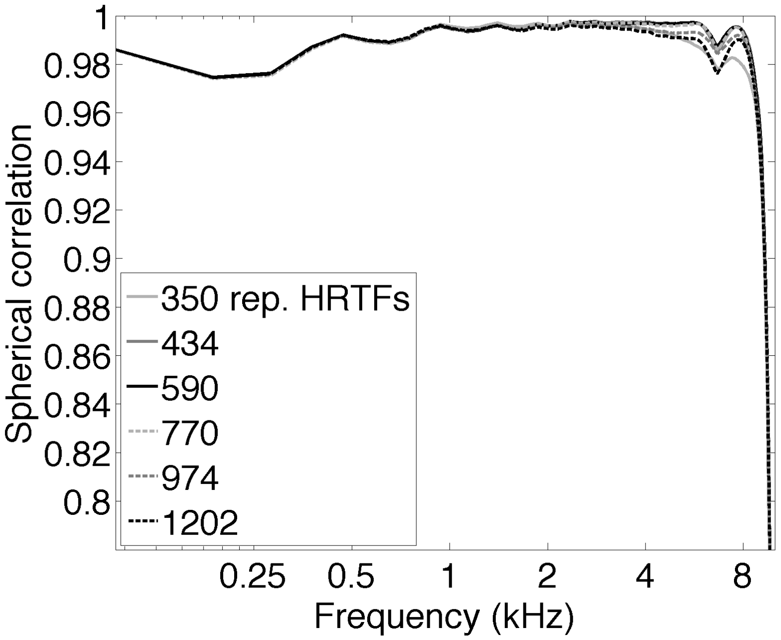


(b) Synthesis at 1.5 m and virtual loudspeakers on Lebedev grids.

Figure 4.12: Normalized spherical correlation between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the bottom Panel of Figure 4.2. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. The loudspeaker driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

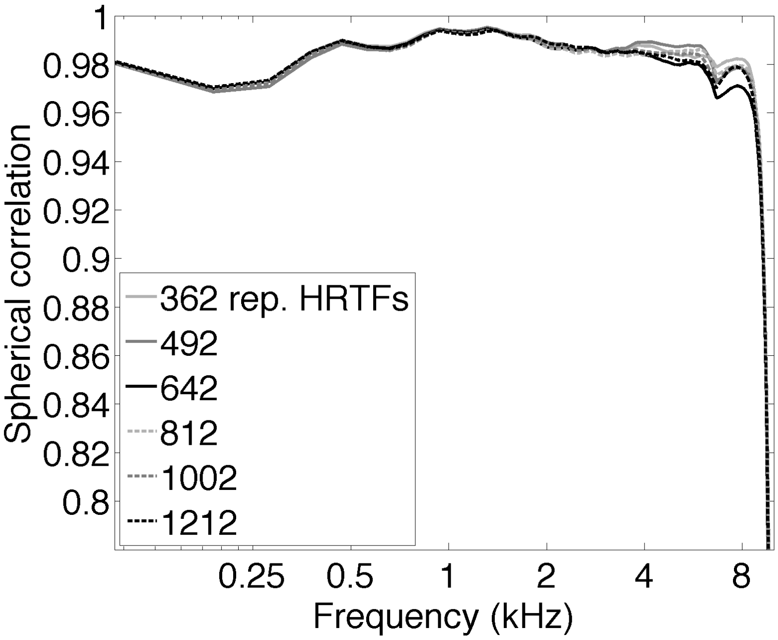


(a) Synthesis at 1 m and virtual loudspeakers on icosahedral grids.

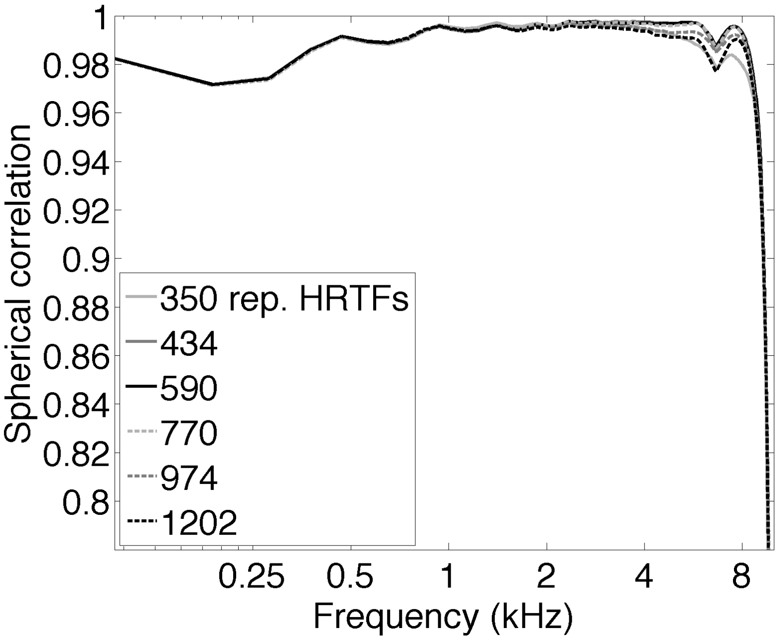


(b) Synthesis at 1 m and virtual loudspeakers on Lebedev grids.

Figure 4.13: Normalized spherical correlation between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the top Panel of Figure 4.3. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. The loudspeaker driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

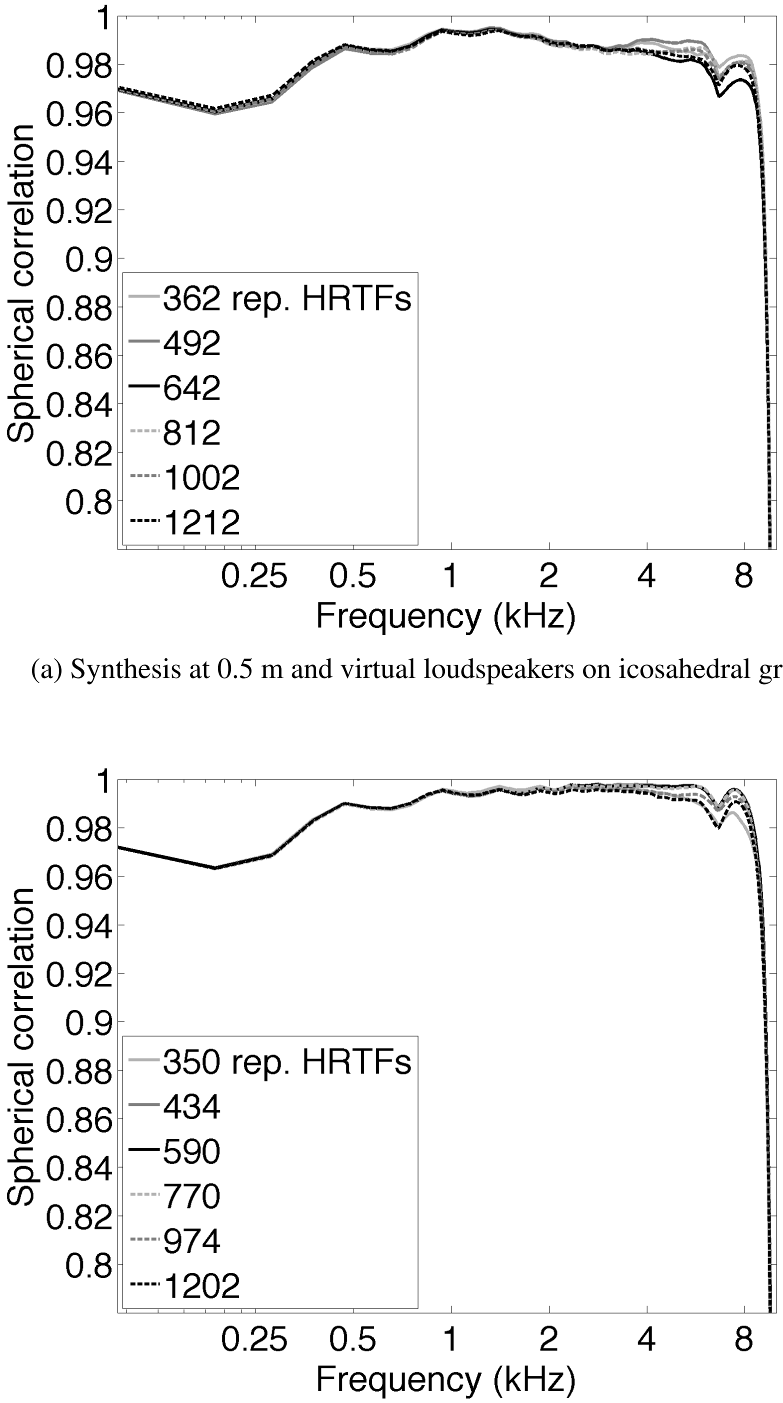


(a) Synthesis at 0.75 m and virtual loudspeakers on icosahedral grids.



(b) Synthesis at 0.75 m and virtual loudspeakers on Lebedev grids.

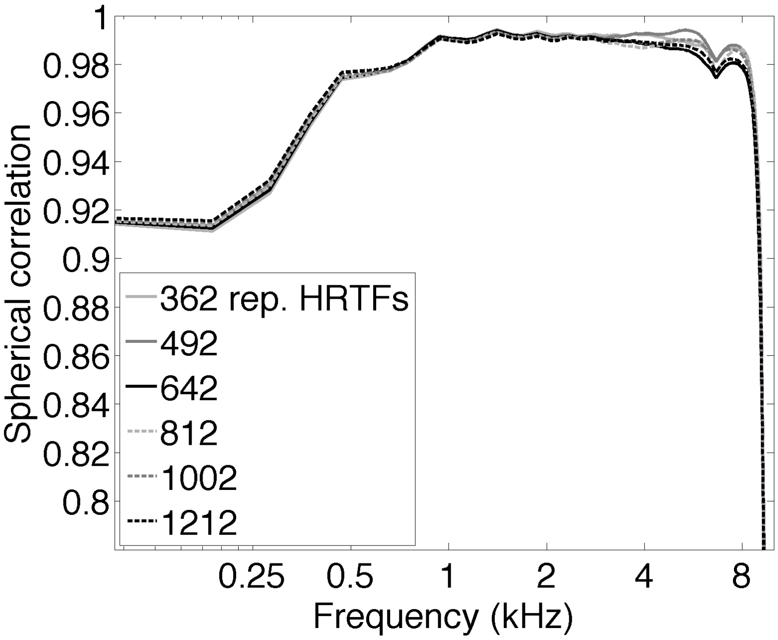
Figure 4.14: Normalized spherical correlation between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the bottom Panel of Figure 4.3. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. The loudspeaker driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.



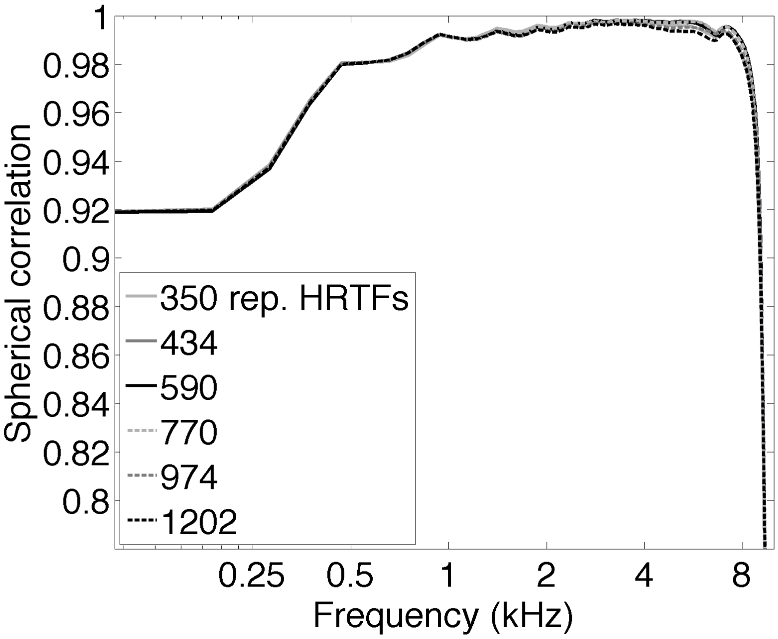
(a) Synthesis at 0.5 m and virtual loudspeakers on icosahedral grids.

(b) Synthesis at 0.5 m and virtual loudspeakers on Lebedev grids.

Figure 4.15: Normalized spherical correlation between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the top Panel of Figure 4.4. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. The loudspeaker driving signals are derived from the recordings made with 252 microphones on a rigid sphere of 8.5 cm radius.

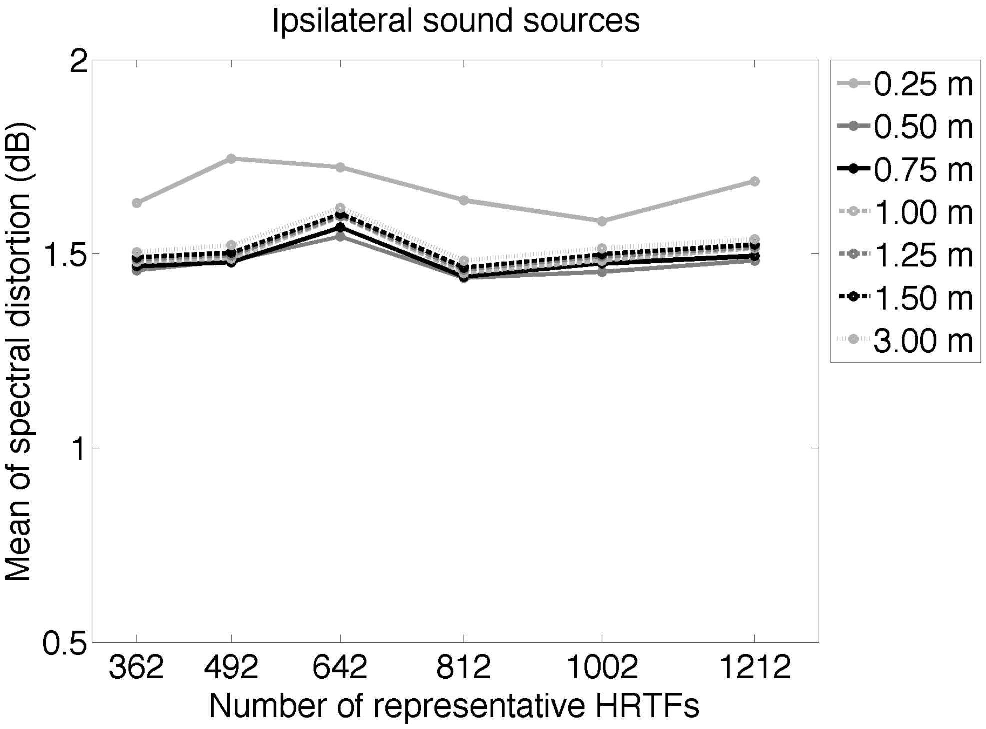


(a) Synthesis at 0.25 m and virtual loudspeakers on icosahedral grids.



(b) Synthesis at 0.25 m and virtual loudspeakers on Lebedev grids.

Figure 4.16: Normalized spherical correlation between the transfer functions synthesized by Eq. (4.9) and the target HRTFs in the bottom Panel of Figure 4.4. Virtual loudspeakers were arranged on (a) icosahedral and (b) Lebedev grids. An array of 252 microphones on a rigid sphere of 8.5 cm radius was assumed. Comparissons with Figures 4.11 to 4.15 show that, for sound sources in proximal region and frequencies below 1 kHz, the synthesis accuracy tends to decrease with decreased frequency and distance.



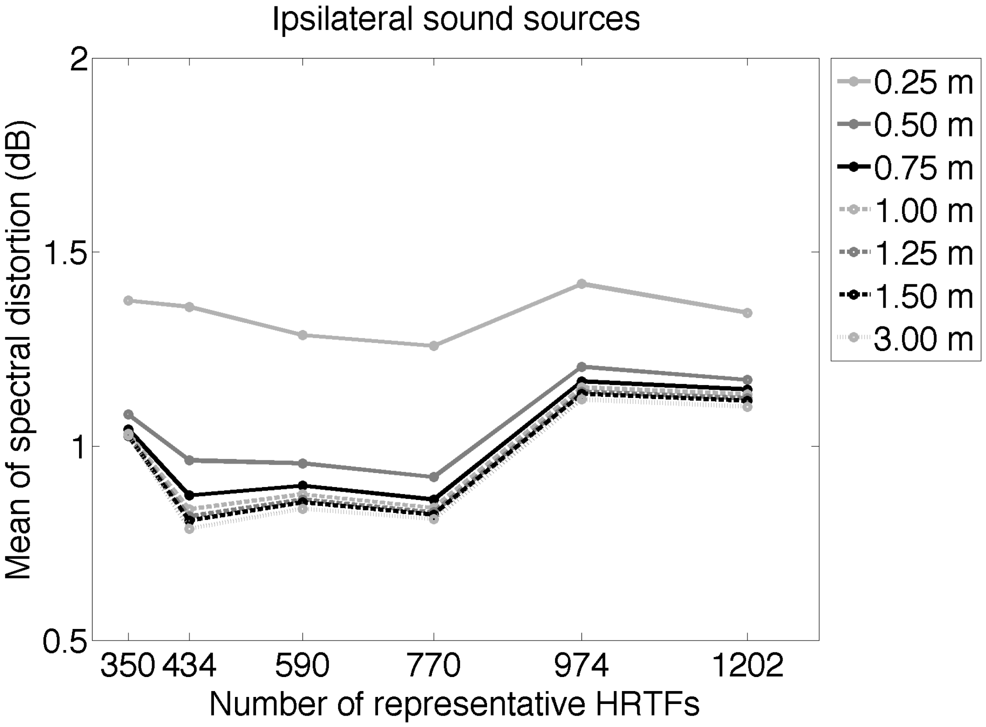


Figure 4.17: Mean value of spectral distortion between the target and synthesized HRTFs for sound sources on the ipsilateral side of the left ear. The synthesis was based on weighted sums of representative HRTFs for sound sources at a 1.5 m distance arranged on icosahedral (top) and Lebedev (bottom) grids. The weights are derived from the compact array signals. Lebedev grids provide slightly better accuracies than icosahedral grids, where the use of 770 representative sound sources improves the accuracy at all distances.

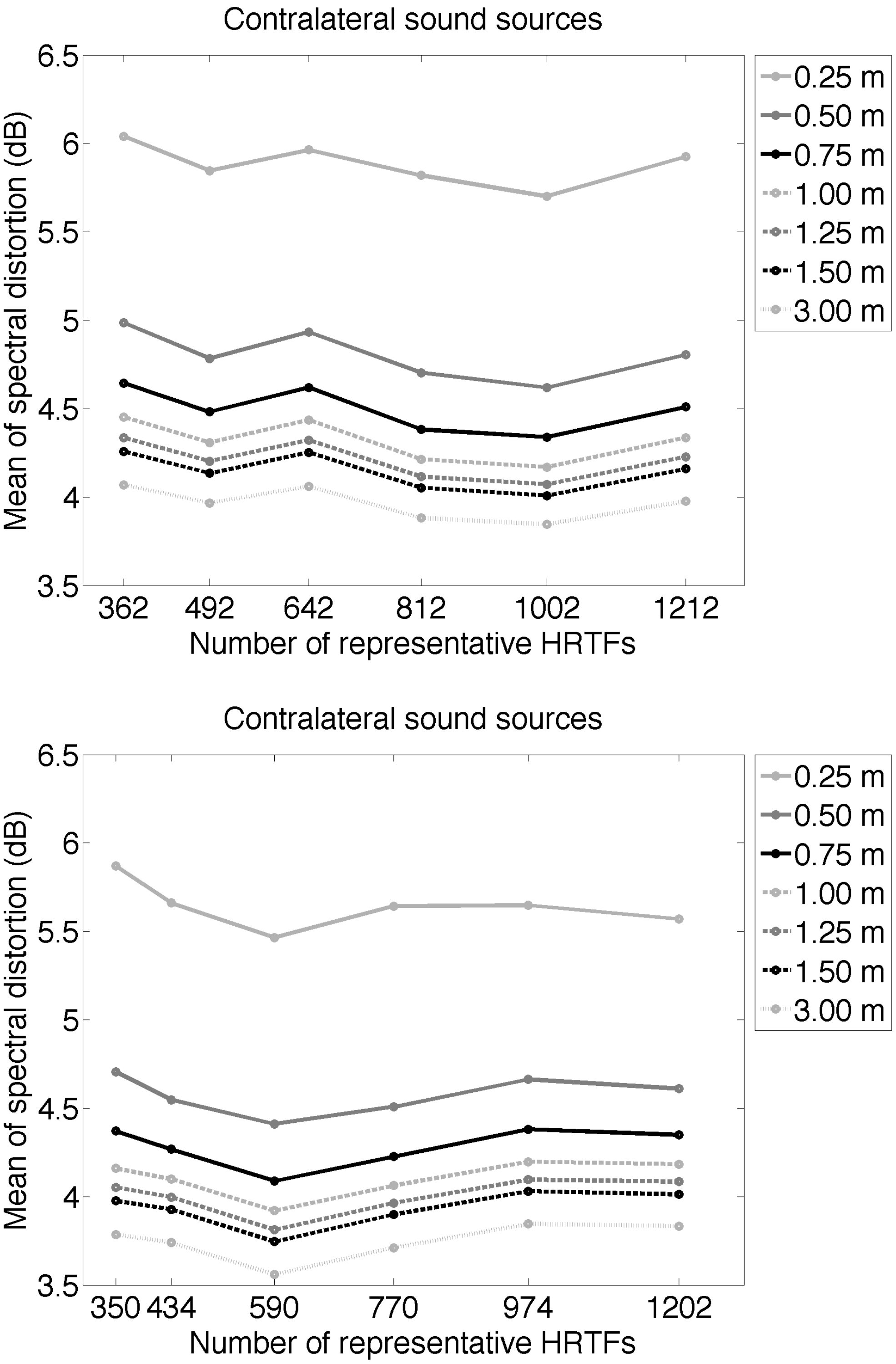
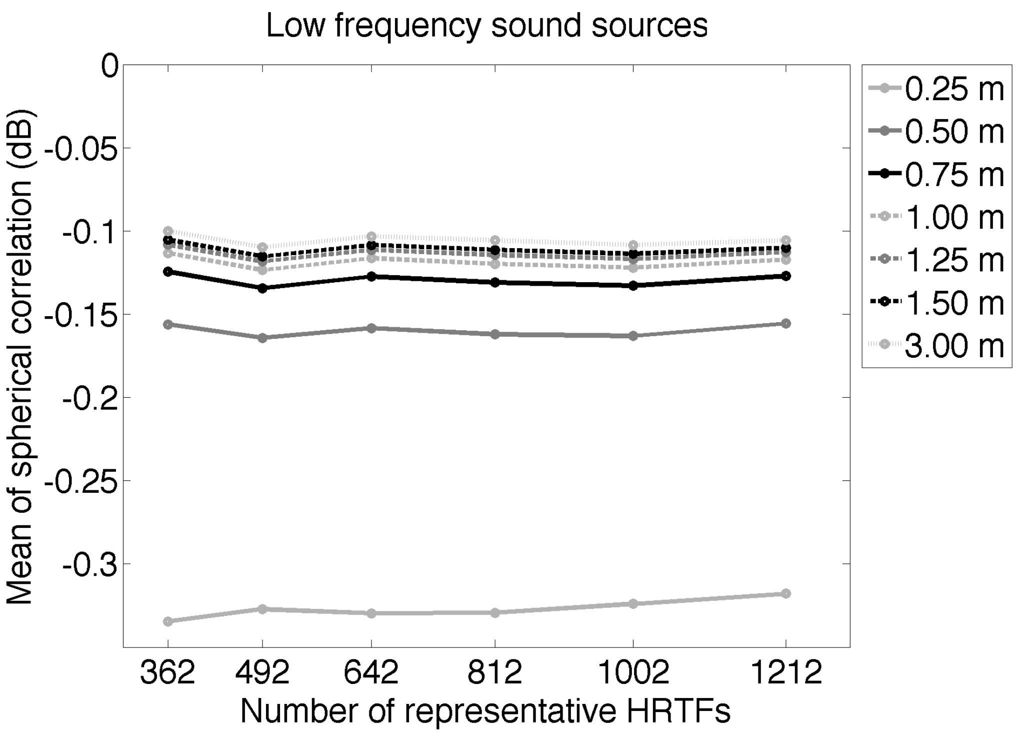


Figure 4.18: Mean value of spectral distortion between the target and synthesized HRTFs for sound sources on the contralateral side of the left ear. The synthesis is based on weighted sums of representative HRTFs for sound sources at a 1.5 m distance arranged on icosahedral (top) and Lebedev (bottom) grids. The weights are derived from the compact array signals. Lebedev grids provide slightly better accuracies than icosahedral grids. In both cases, the accuracy decreases monotonically with decreased distance. The best accuracies correspond to 1002 points on icosahedral grids and 590 points on Lebedev grids.



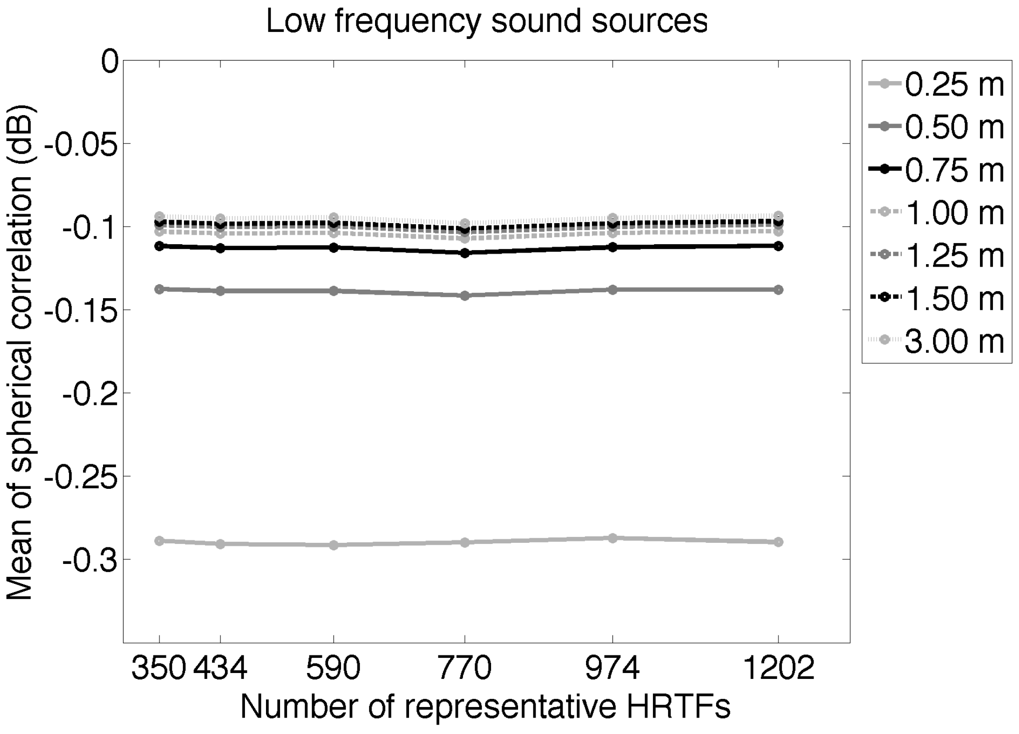
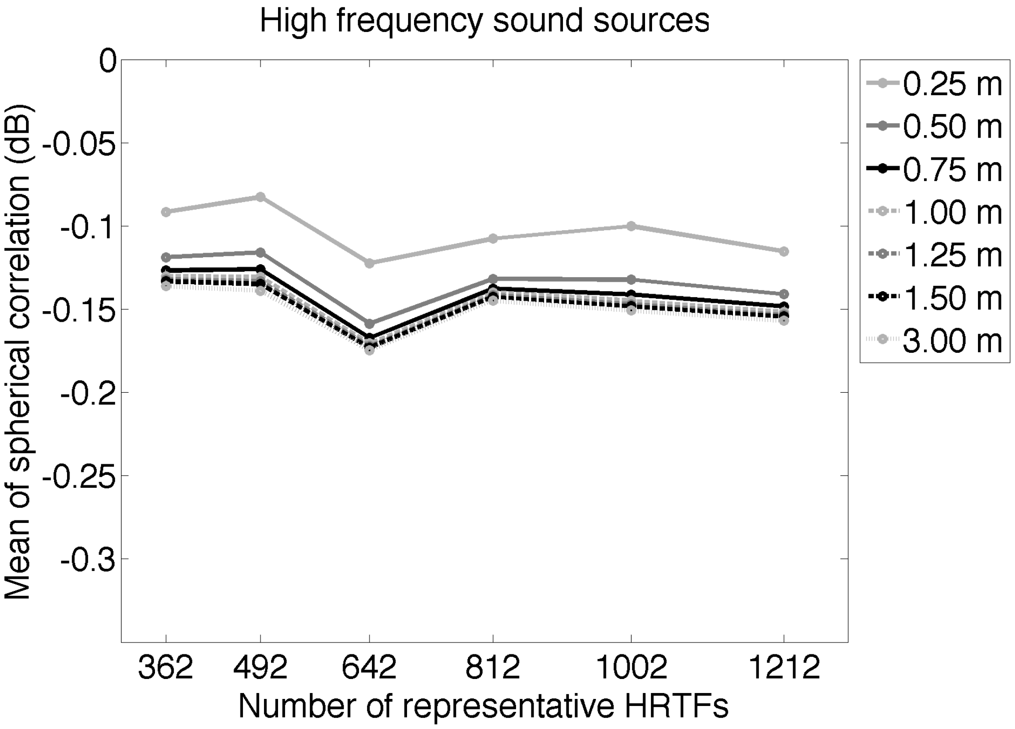


Figure 4.19: Mean value of spherical correlation between the target and synthesized HRTFs for sound sources below 1 kHz. Virtual loudspeakers were arranged on icosahedral (top) and Lebedev (bottom) grids. The weights are derived from the microphone array recordings. Lebedev grids provide slightly better accuracies than icosahedral grids. In both cases, the accuracy of the synthesis is not affected by the number of representative sound sources, but decreases monotonically with decreased distance.



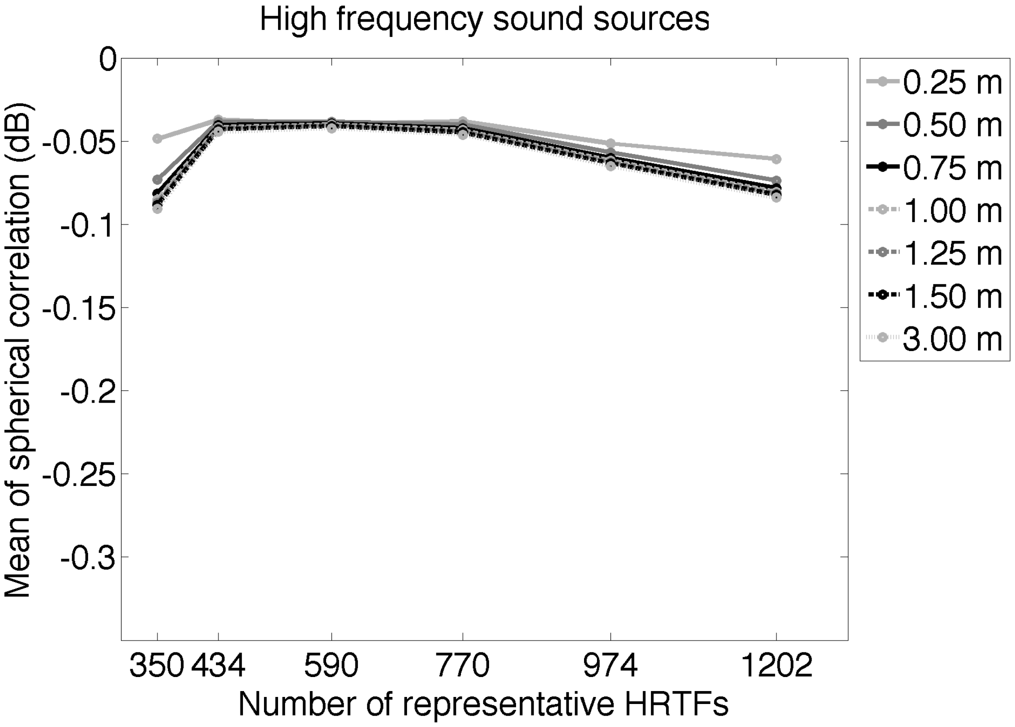


Figure 4.20: Mean value of spherical correlations between 1 kHz and 8 kHz. Virtual loudspeakers were arranged on icosahedral (top) and Lebedev (bottom) grids. The weights are derived from the compact array signals. Lebedev grids provide slightly better accuracies than icosahedral grids. In both cases, the accuracy increases slightly and monotonically with decreased distance. For icosahedral grids, 642 representative sound sources slightly improve the accuracy at all distances.

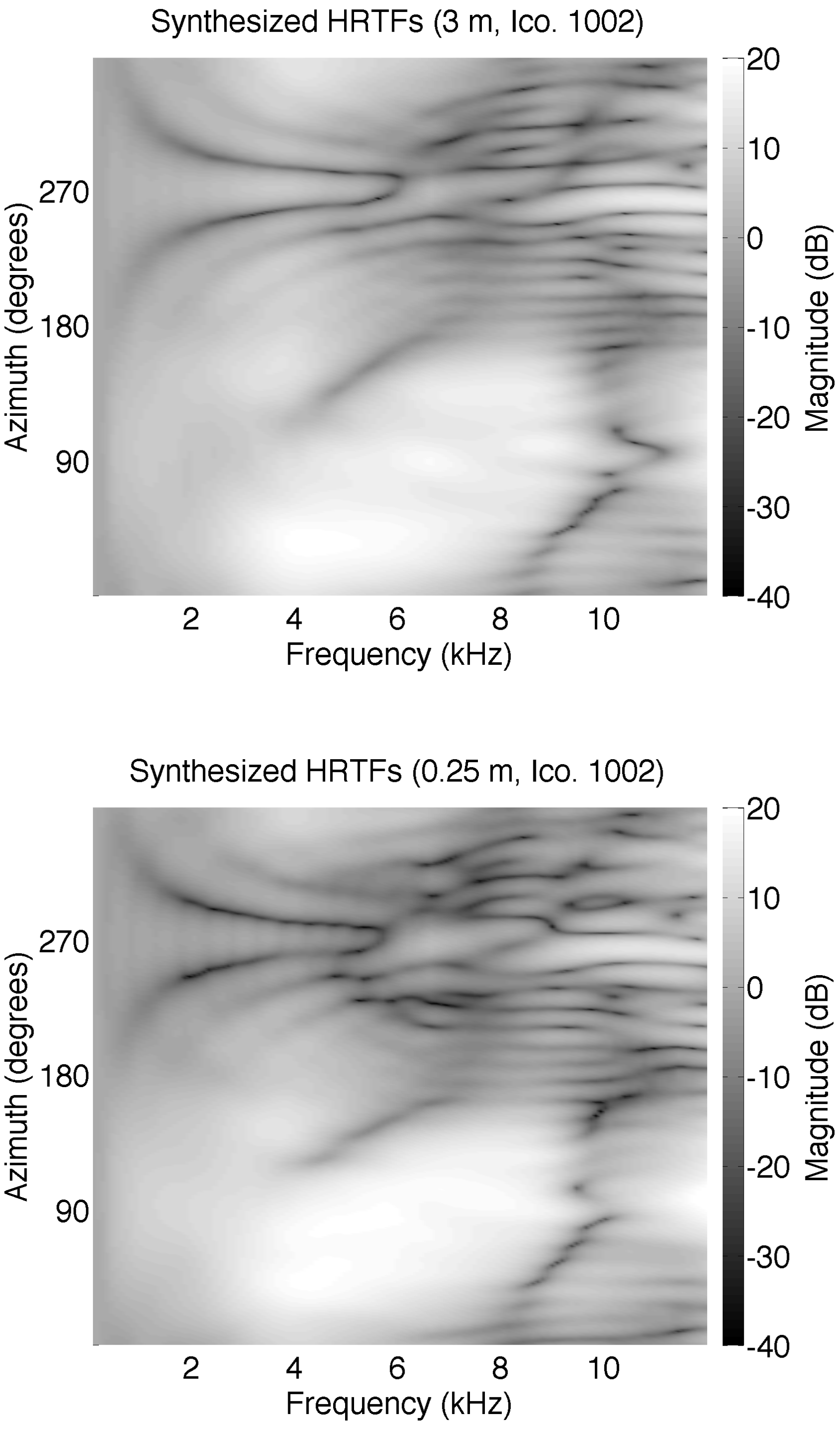


Figure 4.21: HRTFs for (a) distal and (b) proximal sound source synthesized with the extended method. The synthesis is based on weighted sums of 1002 representative HRTFs for sound sources at a 1.5 m distance arranged on an icosahedral grid. The weights are derived from the compact array signals.

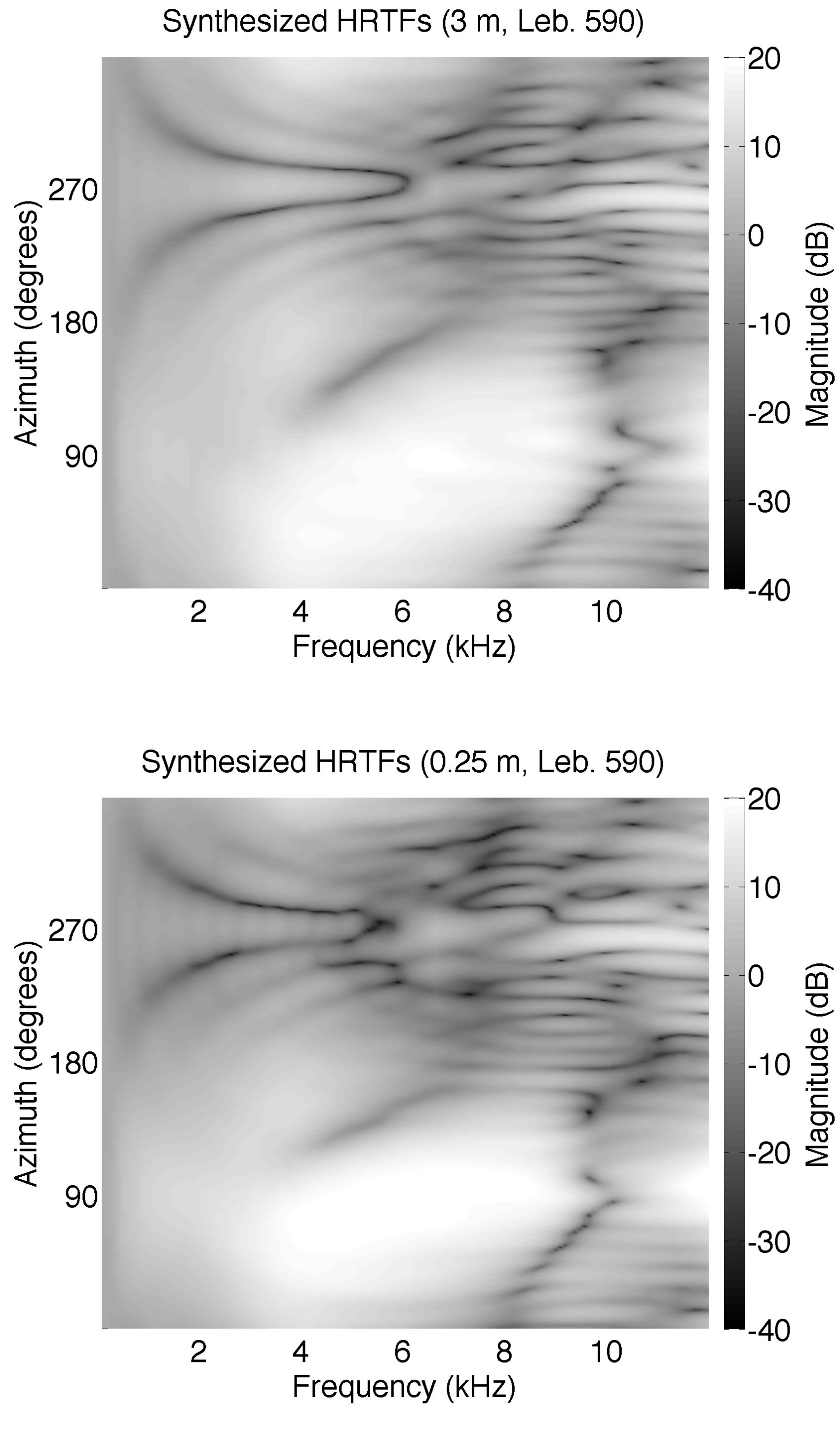


Figure 4.22: HRTFs for (a) distal and (b) proximal sound source synthesized with the extended method. The synthesis is based on weighted sums of 590 representative HRTFs for sound sources at a 1.5 m distance arranged on a Lebedev grid. The weights are derived from the compact array signals.

4.5 Summary

An extended method to synthesize the binaural signals for distal and proximal sound sources has been proposed. The recorded sound field has been analyzed with multipoles, an extension of spherical harmonics that allowed the radial propagation of the sound field. For sound sources at frequencies below 1 kHz, the accuracy of the synthesis tends to decrease with decreased distance, but is not affected by the number of representative sound sources. For sound sources on the contralateral side of the ear, the accuracy also tends to decrease with distance, but it is possible to select an optimum number of repersentative HRTFs that provides the best accuracy.

Chapter 5

Conclusion

A new method to synthesize the binaural signals for the three-dimensional auditory space was proposed in this thesis. The novelty of our proposal is the full covering of the audible space along direction and distance, for which it was necessary to address the binaural synthesis for sources in both the proximal (less than 1 m away) and distal (beyond 1 m) regions. The binaural signals were synthesized by virtual loudspeakers endowed with HRTFs. The driving signals for the virtual loudspekaers were calculated so to match the sound field captured by microphones arranged on the surface of a rigid sphere. The size of the rigid sphere was chosen to equal that of an average human head.

An ideal scenario for the binaural synthesis of sound sources on the distal region, where the sound field is captured assuming an infinite number of microphones, has been described in Chapter 2. The synthesis performed in this way, from the directional distribution of the incident pressure field on a rigid sphere, is closely related to modal beamforming techniques. On the other hand, given the decomposition of the captured sound field in terms of spherical harmonics, and subsequent reconstruction at the virtual loudspeaker array, the reported synthesis accuracies were in agreement with existing techniques for the angular interpolation of HRTFs based on the spherical harmonic decomposition.

A discussion on the optimal arrangement of virtual loudspeakers for an accurate synthesis of sound sources on the distal region have been introduced in Chapter 3. Moreover, the synthesis accuracy evaluation was performed considering a practical scenario where the sound field was captured by a finite number of microphones arranged on the surface of a rigid sphere. It was found that regular samplings of the sphere allowed to set the minimum number of virtual loudspeakers very close to the number of microphones. Accurate synthesis of sources on the horizontal plane was obtained up to the spatial aliasing limit imposed by the finite number of microphones.

An extended method for the binaural synthesis of sound sources in the proximal region has been proposed in Chapter 4. The generalized method was based on the analysis of the sound field in terms of multipoles, an extension of spherical harmonics that allows the radial propagation of the sound field. Theoretically, the closer distance that can be synthesized is equal to the radius of the smallest sphere enclosing the listener's head. This novel proposal is an important step towards the re-creation of the binaural signals for sound sources in the proximal region, which would demand to consider both the recording of near field sound sources and the synthesis of HRTFs for sound sources in the proximal region.

In general, the proposed method shows good accuracy along frequency and direction, as was reported by the spherical correlation and the spectral distortion, up to the limit imposed by the use of a finite number of microphones. The extended method that includes the binaural synthesis of proximal sound sources have been evaluated feeding the proposed system with the model of the acoustic scattering from the rigid sphere, so as to produce the corresponding HRTFs at the output. It was found that at frequencies below 1 kHz, the accuracy of the synthesis tended to decrease with decreased distance, but was not affected by the number of representative HRTFs. However, for sound sources on the contralateral side of the ear, the accuracy also tended to decrease with distance, but it was possible to select an optimum number of repersentative HRTFs that provided the best accuracy.

The computational complexity for a future implementation of the present algorithm can be evaluated from the matrix formulation in Eq. (4.12). Considering the number of products involved in a matrix multiplications, it can be shown that the running time to synthesize a single source, for a number I of frequency bins, is at most of order

$$\mathcal{O}\left(I \times V \times (N+1) \times Q\right), \tag{5.1}$$

where I represents the number of frequency bins, V the number of representative sound sources, N the order of spherical harmonics, and Q the number of microphones. Given that the proposed method allows for the reduction of V nearly close to $Q = (N+1)^2$, the algorithmic complexity for the synthesis of a sound source at each frequency bin can be finally said to be of order

$$\mathcal{O}\left((N+1)^5\right), \tag{5.2}$$

whose lower bound can be set to $\mathcal{O}\left(N^4 \log N\right)$ for large values of N, by using algorithms for efficient matrix multiplications.

Bibliography

[1] C. Cherry, “Some experiments on the recognition of speech, with one and two ears.,” *Journal of the Acoustical Society of America*, vol. 26, pp. 554–559, 1953.

[2] L. Rayleigh, “On our perception of the direction of a source of sound,” in *Proceedings of the Musical Association*, 2nd Session, pp. 75–84, Taylor & Francis, Apr. 1876.

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

[4] 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.

[5] K. de Boer and A. T. van Urk, “Some particulars of directional hearing,” *Phillips Technical Review*, vol. 6, pp. 359–364, Dec. 1941.

[6] 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.

[7] 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.

[8] Y. Tao, A. I. Tew, and S. J. Porter, "A study on head-shape simplification using spherical harmonics for HRTF computation at low frequencies," *Journal of the Audio Engineering Society*, vol. 51, pp. 799–805, Sept. 2003.

[9] V. R. Algazi, R. O. Duda, and D. M. Thompson, "Motion-tracked binaural sound," *Journal of the Audio Engineering Society*, vol. 52, no. 11, pp. 1142–1156, 2004.

[10] S. Sakamoto, S. Hongo, R. Kadoi, and Y. Suzuki, "SENZI and ASURA: new high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," in *Proceedings of the Second International Symposium on Universal Communication*, pp. 429–434, 2008.

[11] S. Sakamoto, J. Kodama, S. Hongo, T. Okamoto, Y. Iwaya, and Y. Suzuki, "A 3D sound-space recording system using spherical microphone array with 252ch microphones," in *Proceedings of 20th International Congress on Acoustics*, (Sydney, Australia), Aug. 2010.

[12] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *119th Convention of the Audio Engineering Society*, (New York, NY, USA), Oct. 2005.

[13] 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.

[14] J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *Journal of the Acoustical Society of America*, vol. 97, pp. 439–452, Jan. 1995.

[15] M. A. Blommer and G. H. Wakefield, "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," *IEEE Transactions on Speech and Audio Processing*, vol. 5, pp. 278–287, May 1997.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] H. Wallach, “On sound localization,” *Journal of the Acoustical Society of America*, vol. 10, no. 4, pp. 270–274, 1939.

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

[24] D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources. hear-related transfer functions,” *Journal of the Acoustical Society of America*, vol. 106, pp. 1465–1479, Sept. 1999.

[25] Y. Suzuki and H.-Y. Kim, “A modelling of distance perception based on auditory parallax model,” in *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America*, (Seattle, USA), Acoustical Society of America, June 1998.

[26] H.-Y. Kim, Y. Suzuki, S. Takane, and T. Sone, “Control of auditory distance perception based on the auditory parallax model,” *Applied Acoustics*, vol. 62, pp. 245–270, Mar. 2001.

[27] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, “Auditory distance perception in humans: A summary of past and present research,” *Acta Acustica united with Acustica*, vol. 91, pp. 409–420, 2005.

[28] A. Kan, C. Jin, and A. van Schaik, “A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function,” *Journal of the Acoustical Society of America*, vol. 125, pp. 2233–2242, Apr. 2009.

[29] N. Kopčo and B. G. Shinn-Cunningham, “Effect of stimulus spectrum on distance perception for nearby sources,” *Journal of the Acoustical Society of America*, vol. 130, pp. 1530–1541, Sept. 2011.

[30] H. Wierstorf, M. Geier, A. Raake, and S. Spors, “A free database of head-related impulse response measurements in the horizontal plane with multiple distances,” in *130th Convention of the Audio Engineering Society*, (London, UK), May 2011.

[31] J. Breebaart and C. Faller, *Spatial audio processing: MPEG surround and other applications*. Chichester, England: John Wiley & Sons, 2007.

[32] A. Berkhout, “A holographic approach to acoustic control,” *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, Dec. 1988.

[33] J. Ahrens and S. Spors, “Wave field synthesis of a sound field described by spherical harmonics expansion coefficients,” *Journal of the Acoustical Society of America*, vol. 131, pp. 2190–2199, Mar. 2012.

[34] J. Daniel, “Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format,” in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, (Denmark), May 2003.

[35] J. Meyer and G. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. II, (Orlando, FL, USA), pp. 1781–1784, May 2002.

[36] B. Rafaely, “Analysis and design of spherical microphone arrays,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 135–143, Jan. 2005.

[37] J.-M. Jot, S. Wardle, and V. Larcher, “Approaches to binaural synthesis,” in *Audio Engineering Society 105th Convention*, (Paris, France), Audio Engineering Society, Sept. 1998.

[38] M. Noisternig, M. Sontacchi, A. Musil, and R. Holdrich, “A 3D ambisonic based binaural sound reproduction system,” in *Audio Engineering Society 24th International Conference: Multichannel Audio, The New Reality*, (Graz, Austria), Audio Engineering Society, June 2003.

[39] D. Menzies and M. Al-Akaidi, “Nearfield binaural synthesis and ambisonics,” *Journal of the Acoustical Society of America*, vol. 121, pp. 1559–1563, Mar. 2007.

[40] W. Song, W. Ellermeier, and J. Hald, “Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise,” *Journal of the Acoustical Society of America*, vol. 123, pp. 910–924, Feb. 2008.

[41] S. Spors and J. Ahrens, “Generation of far-field head-related transfer functions using virtual sound field synthesis,” in *German Annual Conference on Acoustics (DAGA)*, Mar. 2011.

[42] M. A. Poletti and U. P. Svensson, "Beamforming synthesis of binaural responses from computer simulations of acoustic spaces," *Journal of the Acoustical Society of America*, vol. 124, pp. 301–315, July 2008.

[43] B. Støfringsdal and P. Svensson, "Conversion of discretely sampled sound field data to auralization formats," *Journal of the Audio Engineering Society*, vol. 54, pp. 380–400, May 2006.

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

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

[46] J. J. Bowman, T. B. A. Senior, and P. Uslenghi, *Electromagnetic and acoustic scattering by simple shapes*. New York, NY, USA: Hemisphere, 1987.

[47] F. Zotter, *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*. Doctoral thesis, Universitat fur Musik und Darstellende Kunst, Graz, Austria, Sept. 2009.

[48] W. Zhang, M. Zhang, R. A. Kennedy, and T. Abhayapala, "On high-resolution head-related transfer functions measurements: an efficient sampling scheme," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, pp. 575–584, Feb. 2012.

[49] T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, pp. 897–908, Feb. 2007.

[50] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999.

[51] R. S. Womersley and I. H. Sloan, "How good can polynomial interpolation on the sphere be?." Apr. 2001.

[52] V. I. Lebedev, “Quadratures on a sphere,” *{USSR} Computational Mathematics and Mathematical Physics*, vol. 16, no. 2, pp. 10 – 24, 1976.

[53] B. Rafaely, B. Weiss, and E. Bachmat, “Spatial aliasing in spherical microphone arrays,” *IEEE Transactions on Signal Processing*, vol. 55, pp. 1003–1010, Mar. 2007.

[54] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. Princeton University Press, 1987.

[55] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, “Fast head-related transfer function measurement via reciprocity,” *Journal of the Acoustical Society of America*, vol. 120, pp. 2202–2215, Oct. 2006.

[56] N. Matsunaga and T. Hirahara, “Issues of the HRTFs measurement via reciprocal method,” *Technical Report of The Institute of Electronics, Information and Communication Engineers*, vol. 109, pp. 107–112, Oct. 2009.

List of works

International Conferences

1) C. D. Salvador, S. Sakamoto, J. Trevino, J. Li, Y. Yan, and Y. Suzuki, "Accuracy of head-related transfer functions synthesized with spherical microphone arrays," in Proceedings of 21st International Congress on Acoustics, Montreal, Quebec, Canada, 2013.

Domestic Conferences

2) C. D. Salvador, S. Sakamoto, J. Trevino, J. Li, Y. Yan, and Y. Suzuki, "A method to synthesize head-related transfer functions based on the spherical harmonic decomposition," in Proceedings of the 2013 Spring Meeting of the Acoustical Society of Japan, Tokyo, Japan, 2013.