Binaural Synthesis Based on Spherical Acoustics

by

César Daniel Salvador Castañeda

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

Evaluation Committee

Prof. Yôiti Suzuki

Prof. Akinori Ito

Prof. Kazuyuki Tanaka

Assoc. Prof. Shuichi Sakamoto

Tohoku University

July, 2016

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

Future media, information and communication technology will require to create the perceptual illusion of being immersed in a remote environment, whether real or imaginary, so as to enable more realistic telepresence applications. The creation of different blends between real and virtual environments is being recognized as a central stage of media tools that aim to support the synchronous collaboration among people in different locations [1], the safe spatial navigation and training in hazardous work environments [2], and the affective therapeutic experience in metal health treatments [3], among many other applications.

Different modes of perception, such as vision, touch and hearing, are involved when experiencing the world around us. Modern immersive systems need therefore to include displays for all the sensory modalities, so as to achieve high levels of sense of realism (the perceived correspondence between a technology-mediated experience and a similar experience not mediated by technology, in experimental conditions [4]), sense of presence (the feeling of being immersed in a distinct environment [4]), and sense of naturalness (the degree of consistency between the technology-mediated experience and a similar experience in real life [5]).

Hearing plays an important role in enhancing the sense of spatial presence, fundamentally because the spatial features of sound increase the awareness of the physical surroundings and their localized events [6, 7]. Spatial sound also enhances the sense of social presence because it contributes to the perception of the semantic and emotional components of daily experience [8,9]. The sense of naturalness, on the other hand, can also be enhanced by the spatial features of sound since they contribute to the perception of self-motion [10–12],

a key phenomenon that needs to be considered when simulating the experience of being able to move around freely while interacting with a technology-mediated environment [13].

The spatial nature of hearing likewise allows paying attention to events outside the current visual focus and beyond the distances that can be reached with our hands. This fact is being exploited to develop virtual auditory displays for spatial navigation in situations of little or no visual attention, without reducing the necessary amount of interaction [14]. Such kind of systems particularly seek to enable visually impaired individuals to travel through familiar and unfamiliar environments without the assistance of guides [15–17].

The development of technologies for synthesizing the spatial features of sound is therefore becoming increasingly important, not only for conveying the illusion of being immersed in a distinct environment, but also for displaying and distributing spatial information to the users when their other modes of perception are already highly loaded. Such technologies are in general referred to as spatial sound systems, and they typically involve the use of transducer arrays for capturing and presenting the acoustic variables associated with the auditory scenes.

The current study revolves around a class of spatial sound systems that seek to convey high-definition and personal listening experiences by synthesizing the sound pressure at the ears of the listeners, namely, the binaural signals. In particular, this thesis concerns the design of a binaural system for presenting sounds not only at far distances, but also in the region of space immediately surrounding our bodies where objects can be grasped and manipulated, that is, in the peripersonal space. The main purpose of this investigation is the formulation of methods of mathematical physics that will enable the future development of binaural systems for presenting sounds at far and near distances.

Spherical acoustics is the name adopted in this thesis for referring to a body of physical and mathematical work, one of the aims of which is to allow for the stable representation or encoding of spatial sound information at scalable directional resolutions. This kind of representation in turn facilitates the spatial analysis and signal processing required for instance in binaural systems.

Binaural systems based on spherical acoustics have so far been focused on presenting sounds at far distances. However, there is an increasing interest on developing virtual auditory displays for presenting sounds in the peripersonal space.

The objective of this thesis is to extend or generalize the existing spherical acoustic models to enable the development of stable spatial sound recording and optimal binaural synthesis techniques, not only for sounds at far distances, but also at near distances from the listeners' ears.

For practical purposes, the synthesis of sound on a horizontal plane at the height of the listeners' ears is also addressed in this study due to its importance in many applications. In connection with the kind of transducer array used to synthesize the spatial features of sound, three-dimensional synthesis is associated with spherical geometries, while the synthesis of sound restricted to the horizontal-plane is associated with circular geometries.

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 specific objectives of this thesis are presented.

Chapter 2 overviews the theory of spherical acoustics, which comprises the methods of mathematical physics in spherical geometries that are of interest in binaural systems.

Chapter 3 presents a model for the synthesis of the spatial sound information that is contained in a external model of spatial hearing. This model is called the head-related transfer function. Synthesis is considered at far and near distances, and for spherical and circular geometries.

Chapter 4 presents a model for the capture of spatial sound from far and near distances by using spherical and circular microphone arrays.

Chapter 5 presents a general model for binaural synthesis at far and near distances.

Finally, the concluding remarks of this thesis are presented on chapter 6.

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 Prof. 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. Akinori Ito, Prof. Kazuyuki Tanaka, and Assoc. Prof. Shuichi Sakamoto for their participation on my dissertation committee and for their excellent work and teaching skills. Assoc. 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. Satoshi Shiori, Assoc. Prof. Ichiro Kuriki, Assoc. Prof. Kazumichi Matsumiya, Assoc. Prof. Takashi Nose, and all members of the Graduate School Seminar for their valuable comments. Likewise, I express my acknowledgment to Assist. Prof. Zhenglie Cui, Assist. Prof. Jorge Treviño, Mr. Fumitaka Saito, and all the members of the Advance Acoustic Information Laboratory (Suzuki-Sakamoto Laboratory), for their hard work in evaluating the present research and offering valuable feedback. I extend my gratitude to Jorge Treviño, who kept an eye on my work and gave good suggestions and comments, and in general to all the participants of the research meetings on spatial sound recording (Senzi meetings), who gave me helpful comments and had exchange of views for a Doctor degree. I am grateful to Ms. Miki Onodera, the secretary of the Laboratory, 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 academic staff 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 Alejandro Cornejo and all my colleagues at ISONAR for so enriching experience.

Finally, I am profoundly thankful to everyone who have directly or indirectly contributed to this important step in my life: the authors of the literature used in this study, my colleagues in other laboratories in Japan and around the world, all my friends, and specially my relatives and comrades in Peru. It all started back there in my hometown, 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 Spatial sound systems

Spatial sound systems seek to synthesize the spatial features of sound that are required to convey a listening experience with high levels of presence and naturalness. Research towards this goal has led to two prominent schemes. One is to re-create the entire sound pressure field across an extended listening region, whereas the other attempts to reproduce only the sound pressure at both of the listener's ears.

The first scheme leads to systems based on physical models for sound propagation. A common assumption used to simplify their design is to consider the listening region to be empty; that is, to be free of sound sources or obstacles to sound propagation [18, 19]. Other methods attempt to deal with non-empty or reverberant spaces by calculating inverse dynamic models for use with multichannel sound systems [20, 21]. Some proposals lie in between, combining both wave propagation models and the theory of inverse systems [22, 23]. In all cases, this approach to spatial audio typically requires a complex system with a large number of microphones and loudspeakers. An insufficient number of channels results in poor spatial accuracy; this is particularly true for high frequency sound [24].

The second scheme leads to what are known as binaural systems [25–28]. Binaural systems [25–28] attempt to convey the localization cues that naturally arise from the interactions of sound with the listener's pinna, head and body [29]. Binaural sound can be presented through personal auditory displays based on loudspeaker arrays and cross-talk canceling techniques [30, 31], or simply through headphones. Headphone-based reproduction is becoming particularly attractive because it makes possible to isolate the listener from his surroundings. Binaural reproduction is also becoming popular for immersive personal home systems and mobile communication systems [32]. For the above reasons, this study is focused on the binaural approach to spatial audio. An overview of binaural systems is presented in the next section.

1.2 Binaural systems

Binaural systems [26–28] are a promising class of auditory displays, specially for personal high-definition audio devices, because of their small size, simple signal processing, and good performance in terms of the achieved sense of presence and naturalness. In principle, a binaural systems aim to synthesize the sound pressure signals that would be generated at the ears of each listener, namely, binaural signals, as if the listeners were present in a specific acoustic environment.

1.2.1 Conventional binaural systems

The most direct way to record binaural sounds is by placing two small microphones at the entrance of a listener's ear canals. For convenience, though, it is common to use a mannequin or mock-up of the human head, referred to as a *dummy head* [28]. In this way, the microphones will pick up sound after it has interacted with the pinna, head and body of the listener/dummy head. A natural extension of this method is to characterize a listener/dummy head beforehand, in the full-audible frequency range, using a loudspeaker at a given position, so as to compute the transfer functions for the left and right ears that

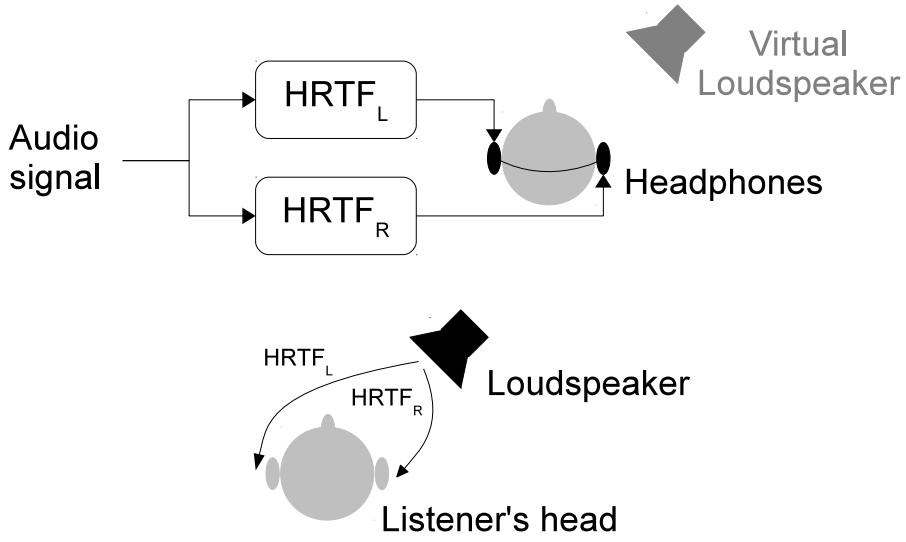


Figure 1.1: A conventional binaural system. An audio signal is filtered with the head-related transfer functions (HRTFs) for the left and right ears. The HRTFs describes the transmission of sound from the loudspeaker position to the listener's ears. The reproduction is done over headphones. Ideally, 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.

describe the acoustic transmission path from the given source position to both ears. The binaural signals can subsequently be generated by multiplying such acoustic transfer functions with corresponding sound source signals in free field.

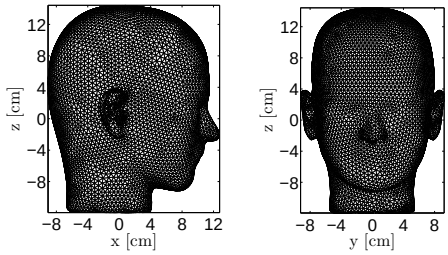
The acoustic transfer functions can be considered to be composed of a representation of the environment (e.g. room acoustics) and a representation of the coloration caused by the interactions of incident sound with the listener's pinna, head and body [29]. The latter component is called the head-related transfer function (HRTF), which comprehensively involve perceptual sound localization cues when listening is done in anechoic conditions. The binaural signals generated by multiplying a dry audio signal with the HRTFs give the impression of a sound source arriving from a virtual loudspeaker at the given location, as illustrated in Fig. 1.1.

1.2.2 Head-related transfer functions

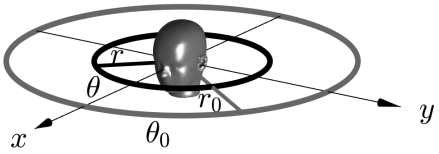
The head-related transfer functions (HRTFs) are a central tool for the binaural presentation of three-dimensional auditory spaces [26, 28]. Their aim is to characterize the scattering of sound due to the external anatomy of a listener, mainly determined by the shape of their pinna, head, and body. HRTFs are represented by linear filters relating the position of a sound source and the sound pressure generated by that source at the ears of the listener [29].

HRTFs contain the auditory cues that are essential to perceive the direction of sound sources in anechoic conditions. Auditory cues for the perception of distance are also present in the HRTFs for sources within 1 m of the listener's head [33–35]. Examples of numerical simulations of circular datasets for left-ear HRTFs at far and near distances are shown in Figs. 1.2, 1.3, and 1.4. The trend of ipsilateral spectral features at the far distance to concentrate around the ear's direction ($\theta = 90°$) in the near distance illustrates the so-called acoustic parallax effect [35, 36]. The decrease in energy of contralateral spectral features at the near distance illustrates the prominence of the head shadowing effect [34]. Such distance cues are particularly useful when presenting auditory scenes comprised of lateral sounds near the heads of the listeners [37].

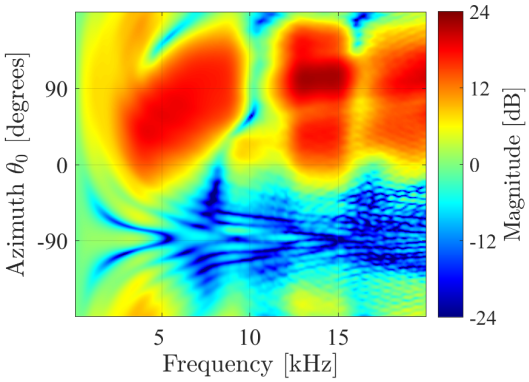
Binaural systems require sets of individual HRTFs characterized for densely distributed sound sources to enable the auditory localization at any specified position. Such HRTF datasets are typically obtained for a spherical array of sound sources at a single distance from the listener's head [38]. Because HRTFs hardly depend on distance when sources are beyond 1 to 1.5 m [33, 34, 39], the radius of the array is typically taken within or beyond this interval. HRTF datasets, however, do not reflect the movements of the listener's head, which are important to provide the dynamic localization cues for enhancing the realism during reproduction [26, 40–42]. This study is focused on binaural systems that are able to capture and present sounds by taking into account the listener's motion.



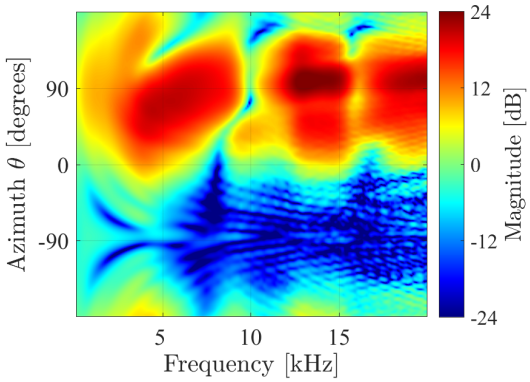
(a) Lateral and front views of a symmetric head model.



(b) Source positions in circular boundaries.



(c) HRTFs at r_0 =150 cm.



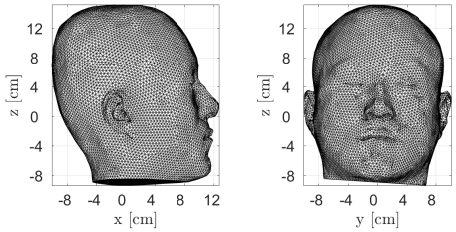
(d) HRTFs at r =25 cm.

Figure 1.2: Two circular HRTF datasets for the left ear ($\theta_0 = \theta = 90°$) calculated with the boundary element method (BEM) [43]. Azimuthal angles θ_0 and θ are measured from the positive $x-$axis, which indicates the front position ($\theta_0 = \theta = 0°$). Magnitudes indicate the difference in sound pressure level at the left ear compared to that which would be observed at the position of the head's center in free-field conditions.

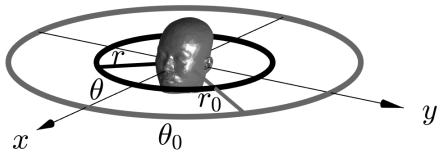
1.2.3 Dynamic binaural systems

Updating the binaural signals to match the listener's motion is not easy. It requires the real-time tracking of the listener's motion and the proper recording of sound reflecting the motion. A recording method is based on the use of a head and torso simulator (HATS) that closely resembles the listener and simultaneously moves as the listener does in a remote place. This kind of system is known as a TeleHead [44, 45]. TeleHeads, however, can match the motion of only one listener at a time.

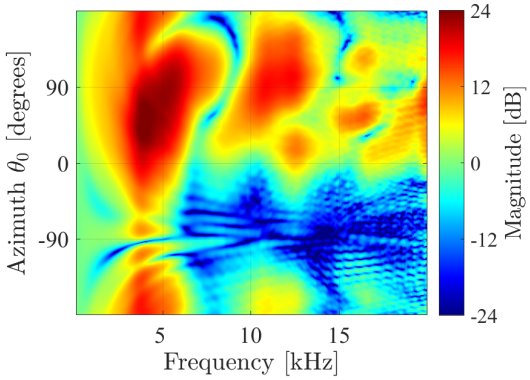
The simultaneous tracking of multiple moving listeners is possible electronically by recording spatial sound information with a spherical microphone array. Its use allows for matching movements through digital computation. A system based on this approach is the motion-tracked binaural (MTB) system [46].



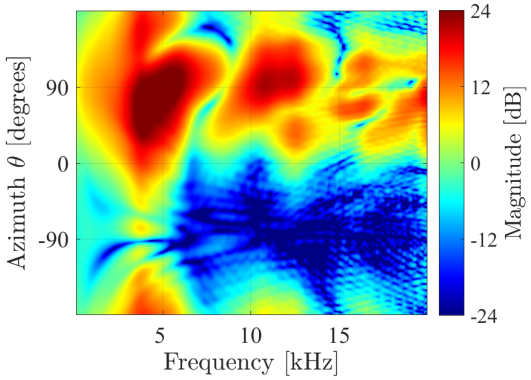
(a) Lateral and front views of the individual head model A.



(b) Source positions in circular boundaries.



(c) HRTFs at r_0 =150 cm.



(d) HRTFs at r =25 cm.

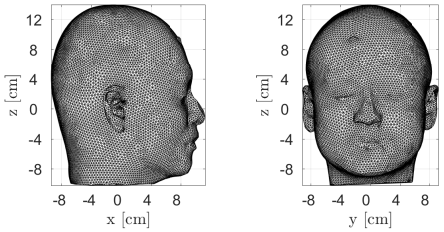
Figure 1.3: Examples of individual HRTF datasets for the left ear ($\theta_0 = \theta = 90°$) calculated with the boundary element method (BEM) [43].

1.2.4 Sound field recording with spherical microphone arrays

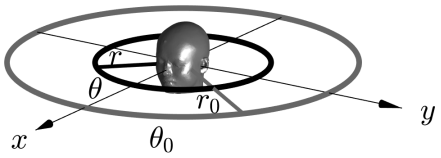
Sound field recording with a microphone array mounted on an acoustically rigid spherical baffle has become an important research issue in areas such as beamforming [47–58], binaural synthesis [59–66], and room acoustics [67–71].

Characterizing a sound field within a region in space fundamentally requires the knowledge of the pressure and its normal derivative in a bounding surface. The consideration of only pressure produces non-unique characterizations at frequencies associated with resonances of the recording boundary exclusively [72]. Since the use of the rigid baffle causes the normal derivative to be zero on the rigid boundary, its use allows for sound field characterizations that are free of resonances. Examples of rigid spherical microphone arrays are shown in Fig. 1.5. This is the kind of arrays considered throughout this study.

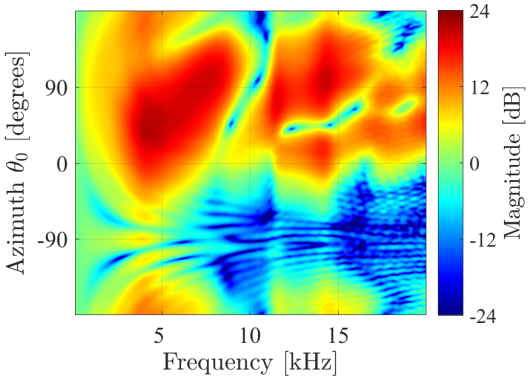
The use of a spherical microphone array in binaural synthesis provides the following advantages: 1) acoustical characteristics of the array can be modeled mathemati-



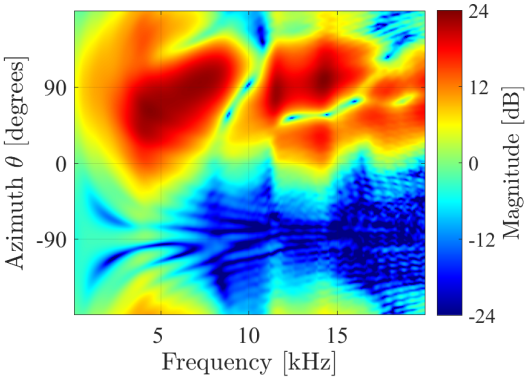
(a) Lateral and front views of the individual head model B.



(b) Source positions in circular boundaries.



(c) HRTFs at r_0 =150 cm.



(d) HRTFs at r =25 cm.

Figure 1.4: Examples of individual HRTF datasets for the left ear ($\theta_0 = \theta = 90°$) calculated with the boundary element method (BEM) [43].

cally, 2) spatial sound is captured from several directions with almost uniform resolution, and 3) simultaneous tracking of multiple moving listeners is possible through digital computations. To enable binaural reproduction, the array recordings need to be adapted in real-time for its rendering with HRTF datasets.

1.2.5 Combining HRTFs and sound field recordings

During the last decade, there has been an increasing interest on the development of binaural systems that are based on combining the spatial information contained in HRTF datasets with recordings made with spherical arrays [46,58–60,63,64,66,73–80]. Such approach is useful to enhance the binaural presentation of localized sounds because the HRTF datasets can provide the individual auditory cues, while the spherical array recordings can provide the dynamic auditory cues to multiple listeners.

Combining HRTF datasets with array recordings leads to systems such as SENZI [81,

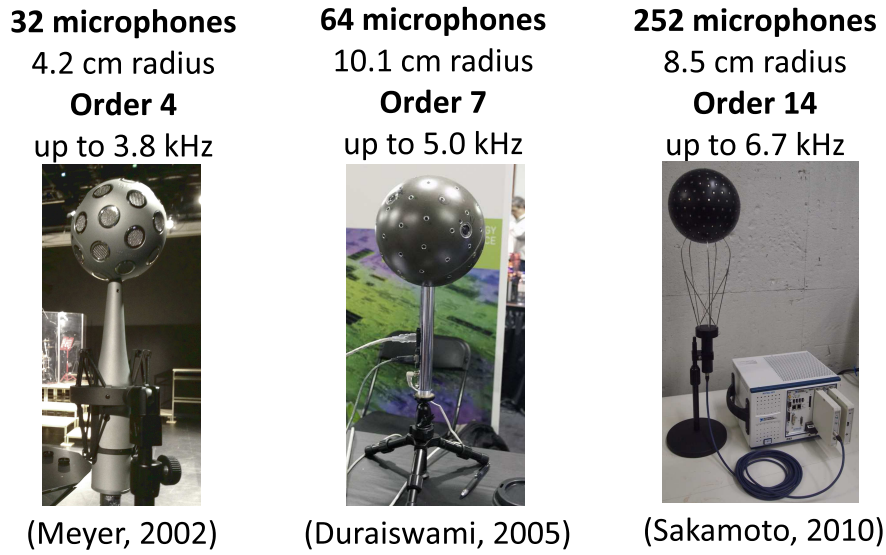


Figure 1.5: Examples of spherical microphone arrays. The order is a measure of directional resolution and increases as the number of microphones increases. Together with the radius of the array, the order also determines the maximum frequency of operation.

82] and the virtual artificial head (VAH) [73]. In these systems, HRTFs constitute a desired spatial pattern to be synthesized by a set of microphone filters. Computing these filters requires inversion of a linear system involving acoustic transfer functions from the source position of each HRTF to each microphone position. Inverse methods used in [82] and [73] consider that the HRTFs and recording grids are coupled, which can limit their flexibility in applications requiring intermediate stages for spatial editing.

In this study, we follow the alternative approach that offers the possibility of decoupling the HRTFs and recording grids. This approach is based on the use of representations of HRTF datasets and spherical array recordings in terms of solutions to the acoustic wave equation at different spatial resolutions (orders). This approach enjoy popularity because they enable scalable encoding and simplify multichannel processing [59, 60, 63–66, 75].

1.3 Spherical acoustics

Spherical acoustics is the name adopted in this thesis for referring to a body of physical and mathematical work [72, 83–88] for use in spatial sound systems that exhibit the

following characteristics:

- The spatial sound information is defined on spherical boundaries.

- The vicinity of a single view point is emphasized.

- The directions are covered with almost uniform radial symmetry.

Since the binaural systems investigated in this thesis share the characteristics mentioned above, spherical acoustics provide a plethora of suitable methods for their study.

The results of spherical acoustic theory rely on assuming that the spatial information is a continuous function defined on a continuous spherical domain. Yet in practice, spatial information on the sphere necessarily requires to be represented by means of a finite number of transducers, for instance a microphone array for recording or a loudspeaker array for reproduction.

In analogy with sampling theory on the time domain, obtaining a discrete version of a continuous function defined on the sphere can be regarded as a process composed of two stages: angular band limitation and angular sampling, as shown in Fig. 1.6. Angular bandwidth is defined as a measure of the variation of a function along directions. Two adjacent microphones, for example, could sample a continuous sound field at their corresponding positions with sufficient accuracy only when the amount of variation is comparable with the distance between the two microphones. Angular band limitation has therefore the role of ensuring a limited angular variation so that the spatial information on the sphere can be described with sufficient accuracy by using a finite number of samplings.

The mathematical details of the spherical acoustic methods used in this thesis are given in chapter 2.

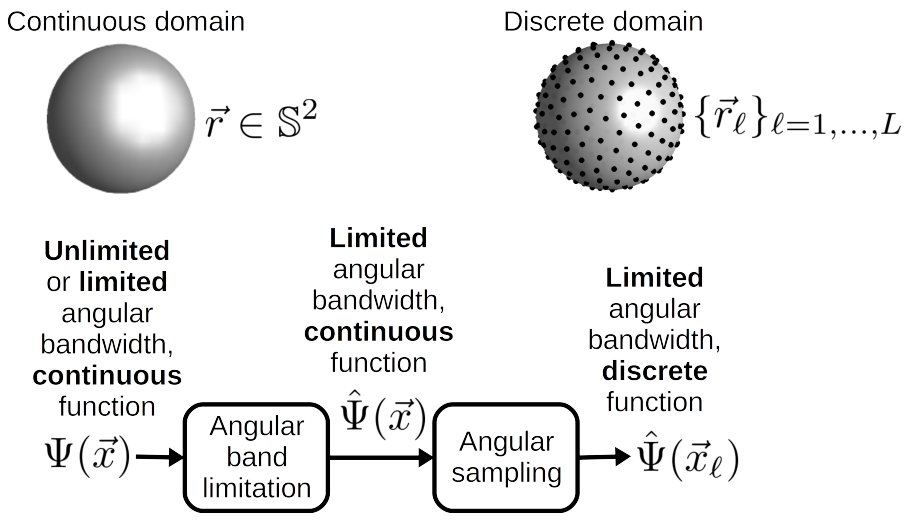


Figure 1.6: Sampling on the sphere. The angular bandwidth is a measure of the variation of a function along directions

1.4 Necessity of spherical acoustic models for binaural synthesis at near distances

Spherical acoustic models have so far been focused on presenting sounds at far distances. However, there is an increasing interest on binaural systems for presenting sounds in the region of space immediately surrounding our bodies, in which objects can be grasped and manipulated, that is, in the peripersonal space.

Spherical acoustic models for near distances would benefit, for example, the development of applications requiring a high degree of consistency between vibration and sound of close objects, as would be the case of training systems in the mechanical industry. Another field of applications can be found in virtual environments for design or entertainment that involve grabbing and moving virtual objects.

Figure 1.7 illustrates the concept proposal used throughout this thesis for realizing a general binaural system at far and near distances, with the possibility of manipulating the sounds along direction and distance.

An overview of spherical acoustic models for binaural synthesis at far distances is presented in Table 1.1. The research areas requiring significant improvement to enable models

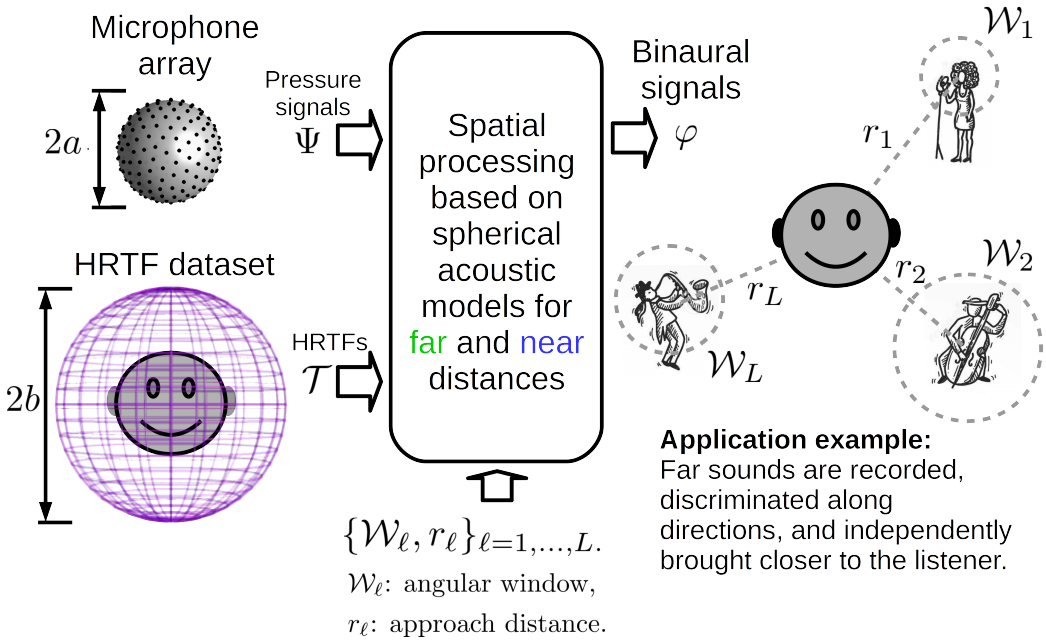


Figure 1.7: Concept proposal for a general binaural system. In the application example, an angular window $\mathcal{W}_\ell$ defines the angular region in which far sounds are approached to a desired distance r_ℓ.

at near distances are emphasized in italics. Because modern binaural systems are based on combining spatial sound information available in microphone array recordings and HRTF datasets, existing models for binaural synthesis are first classified into two groups according to the kind of spatial information. Although binaural synthesis would ideally require to take into consideration information on the sphere, in practice it is sometimes enough to consider circular data limited to a horizontal plane at the height of the ears. This give raise to the second level in the classification. Finally, models are also distinguished according to the nature of the spatial domain, whether continuous or discrete.

It is highlighted in Table 1.1 that existing models for synthesis of HRTFs from circular datasets, formulated in the continuous domain, do not ensure accurate prediction of the pressure decay along distance. They also require splitting the plane into two half planes, which produces discontinuities when synthesis is performed along the axis connecting the ears.

Another research field highlighted in Table 1.1 is related to recording with spherical microphone arrays. In this field, continuous space models still require excessive dynamic

ranges at lower and middle frequencies, which can not be achieved with existing transducers.

In general, discrete models can be derived from continuous models by taking into consideration the discretization scheme shown in Fig. 1.6. This requires proper methods for angular band limitation. Existing methods for this task, however, yield spectral discontinuities and overemphasize the lower and middle frequencies.

In addition, existing models for combining array recordings and HRTF datasets have so far considered the case of far distances only [65, 66, 89, 90].

1.5 Objectives of the thesis

The purpose of this research is to cover the gap in existing theoretical models, so as to enable the development of binaural systems for presenting sounds at far and near distances.

In connection with the present state of spherical acoustic models displayed in Table 1.1, The objectives are therefore stated as follows:

- An stable model for the synthesis of HRTFs.
- An stable model for sound field recording with spherical microphone arrays.
- A general model for binaural synthesis by combining the models for HRTF synthesis and sound field recording.

Modeling throughout this study relies on the following assumptions:

- Monochromatic and omnidirectional acoustic sources
- sound radiation take place in an anechoic environment
- sound field recording is done by a continuous distribution of measurement points
- sound field reproduction is provided by a continuous distributions of secondary sources.

Table 1.1: Overview of spherical acoustic models for binaural synthesis at far distances, and *research areas requiring significant improvement to enable models at near distances*.

Model for		Continuous space	Discrete space
Synthesis of head-related transfer functions for arbitrary source positions	From a spherical dataset of HRTFs	Band-unlimited distance-varying filters for synthesis on three-dimensional space [91–93]	Band-limited distance-varying filters for synthesis on three-dimensional space: • *Existing models that use frequency-dependent band limitation yields spectral discontinuities and overemphasizes the lower frequencies* [89, 94–96]
	From a circular dataset of HRTFs	Band-unlimited distance-varying filters for synthesis on the horizontal plane: • *Existing models do not accurately predict the distance decay due to point sources, or require splitting the plane into two half-planes* [97]	Band-limited distance-varying filters for synthesis on the horizontal plane: • *Existing models that use frequency-dependent band limitation yields spectral discontinuities and overemphasizes the lower frequencies* [89, 94–96]
Sound field recording	With a spherical microphone array mounted on a rigid spherical baffle	Band-unlimited boundary-matching filters for microphone-to-loudspeaker signal conversion: • *Continuous space models requires excessive dynamic ranges at lower and middle frequencies* [47, 49–53]	Band-limited boundary-matching filters for microphone-to-loudspeaker signal conversion: • *Angular band limitation based on Tikhonov regularization is demanding and depends on several parameters* [49, 50, 56, 57]
	With a circular microphone array mounted on a rigid spherical baffle	Band-unlimited boundary-matching filters for microphone-to-loudspeaker signal conversion [98]	Band-limited boundary-matching filters for microphone-to-loudspeaker signal conversion: • *Existing models that use frequency-dependent band limitation yields spectral discontinuities and overemphasizes the lower frequencies* [89, 94–96]

The strong assumptions above are made in order to concentrate the attention on the essential facts of the underlying physics, simplify the mathematical presentation of the problem, and to clearly deliver the scientific message of this work.

An overview of the contributions of the thesis is presented in Table 1.2. In addition, models for combining array recordings and HRTF datasets at near distances are also presented as contributions chapter 5.

Table 1.2: Overview of spherical acoustic models for binaural synthesis, and *summary of the scientific contributions detailed throughout this thesis*.

Model for		Continuous space	Discrete space
Synthesis of head-related transfer functions for arbitrary source positions	From a spherical dataset of HRTFs	Band-unlimited distance-varying filters for synthesis on three-dimensional space [91–93]	Band-limited distance-varying filters for synthesis on three-dimensional space: • *A method for angular band limitation that yields a smooth and bounded spectral response* (in Sec. 3.4.3)
	From a circular dataset of HRTFs	Band-unlimited distance-varying filters for synthesis on the horizontal plane: • *A model that predicts the distance decay due to point sources, and does not require splitting the plane* (in Sec. 3.4.1.3)	Band-limited distance-varying filters for synthesis on the horizontal plane: • *A method for angular band limitation that yields a smooth and bounded spectral response* (in Sec. 3.4.3)
Sound field recording	With a spherical microphone array mounted on a rigid spherical baffle	Band-unlimited boundary-matching filters for microphone-to-loudspeaker signal conversion: • *A model with limited dynamic range at lower frequencies* (in Sec. 4.4.1.3)	Band-limited boundary-matching filters for microphone-to-loudspeaker signal conversion: • *A method for angular band limitation that yields a smooth and bounded spectral response by simply setting a threshold for magnitude* (in Sec. 4.4.2.3)
	With a circular microphone array mounted on a rigid spherical baffle	Band-unlimited boundary-matching filters for microphone-to-loudspeaker signal conversion [98]	Band-limited boundary-matching filters for microphone-to-loudspeaker signal conversion [89].

Chapter 2

Spherical acoustics

2.1 Overview

Spherical acoustics is the name adopted in this thesis for referring to a body of physical and mathematical work that is used in spherical geometries [72, 83–88]. One of the aims of spherical acoustics is to allow for the stable representation or encoding of spatial sound information at scalable directional resolutions (orders). This in turn facilitates the spatial signal processing and analysis that is required for instance in binaural systems.

This chapter presents an overview of the theory of spherical acoustics from three equivalent approaches. The first approach emphasizes the physical nature of this theory and lies on the solutions to the acoustic wave equation in spherical coordinates. The second approach highlights the mathematical aspect of the theory, which arises from the fact that the angular portion of the solutions to the wave equation defines an orthonormal basis of functions on the sphere that is used to represent spatial sound information in general. Finally, the third approach brings up a general scheme for processing of spatial information in the represented or encoded domain, which constitutes a convenient and intuitive framework for the design and implementation of binaural systems in spherical geometries.

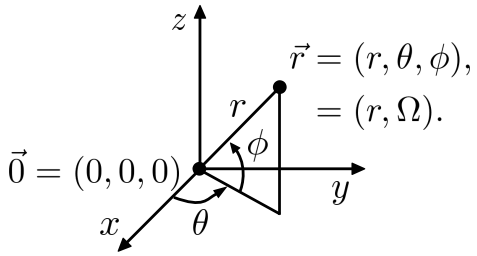


Figure 2.1: Spherical coordinate system.

The spherical coordinate system used throughout this chapter is shown in Fig. 2.1. In these coordinates, a point in space $\vec{r} = (r, \theta, \phi)$ is specified by its radial distance r, azimuth angle $\theta \in [-\pi, \pi]$ and elevation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Angles are merged into the variable $\Omega = (\theta, \phi)$ in such a way that a point in space is also represented by $\vec{r} = (r, \Omega)$.

2.2 Physical approach: the acoustic wave equation

The propagation of sound in space is modeled by the acoustic wave equation, which is a partial differential equation involving acoustic pressure fields $\Psi(\vec{r}, t)$ and their partial derivatives along space $\vec{r}$ and time t. The wave equation in a homogeneous fluid is mathematically formulated as follows [83, ch. 2, pp. 15]:

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\Psi(\vec{r}, t) = 0, \tag{2.1}$$

where ∇ is the gradient operator and c denotes the speed of sound in the homogeneous fluid. Air is considered as a fluid that comes from a homogeneous mixture of the gaseous substances nitrogen, oxygen, and smaller amounts of other substances. The variable c will therefore denote the speed of sound in air in all of what follows.

In the spherical coordinate system shown in Fig. 2.1, the homogeneous wave equation in (2.1) can be explicitly formulated as follows [83, ch. 6, pp. 183]:

$$\left[\underbrace{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial}{\partial\phi}\right) + \frac{1}{r^2\cos^2\phi}\frac{\partial^2}{\partial\theta^2}}_{\nabla^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\Psi(r, \theta, \phi, t) = 0. \tag{2.2}$$

The solution to (2.2) can be obtained by a method called separation of variables. It is based on assuming sound pressure fields that behave independently along time and the spatial coordinates:

$$\Psi(r, \theta, \phi, t) = \Re(r)\Theta(\theta)\Phi(\phi)T(t). \tag{2.3}$$

Sound fields of this form decouple the wave equation into the following sum of ordinary differential equations [83, ch. 6, pp. 185]:

$$\frac{1}{\Re}\frac{d}{dr}\left[r^2\frac{d}{dr}\right]\Re + \underbrace{\frac{1}{\cos^2\phi}\underbrace{\frac{1}{\Theta}\frac{d^2}{d\theta^2}\Theta}_{=-m^2} + \frac{1}{\cos\phi}\frac{1}{\Phi}\frac{d}{d\phi}\left[\cos\phi\frac{d}{d\phi}\right]\Phi}_{=-n(n+1)} - r^2\underbrace{\frac{1}{c^2}\frac{\partial^2}{\partial t^2}T}_{=-k^2} = 0. \tag{2.4}$$

The particular solution is obtained by using separation constants k, n and m. These quantities define the wave number k, the order n, and the degree m of the solution. In particular, the wave number is related to the angular frequency ω, or to the temporal frequency f, according to the following expression:

$$k = \frac{\omega}{c} = \frac{2\pi f}{c}. \tag{2.5}$$

The temporal part of the solution is given by a harmonic oscillator

$$T(t) = T_1 \exp(jkct) + T_2 \exp(-jkct), \tag{2.6}$$

where $j = \sqrt{-1}$. The selection of one of the two terms in (2.6) defines a sign convention in time. In this thesis, the second term is chosen so that the combination constant $T_1 = 0$.

The radial part of the solution is defined by

$$\Re(r) = R_1\mathcal{J}_n(kr) + R_2\mathcal{N}_n(kr), \tag{2.7}$$

where $\mathcal{J}_n$ and $\mathcal{N}_n$ are spherical Bessel functions of the first and second kind, respectively.

Alternatively, the solutions can be written as follows:

$$\Re(r) = R_3 \underbrace{\left[\mathcal{J}_n(kr) + j\mathcal{N}_n(kr)\right]}_{h_n^{(1)}} + R_4 \underbrace{\left[\mathcal{J}_n(kr) - j\mathcal{N}_n(kr)\right]}_{h_n^{(2)}}, \tag{2.8}$$

where the terms in brackets define the spherical Hankel functions of the first and second kind, respectively denoted by $h_n^{(1)}$ and $h_n^{(2)}$. The selection of one or both of the terms in (2.7) or (2.8) will depend upon the locations of the sources and the specific physical conditions of the problem under consideration.

The angular solutions Φ and Θ are combined into a single function $Y_n^m = \Phi\Theta$ called a spherical harmonic of order n and degree m. The spherical harmonic is defined by

$$Y_n^m(\Omega) = Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\sin\phi)\exp(jm\theta), \tag{2.9}$$

where P_n^m is the associated Legendre function. Examples of spherical harmonic functions are depicted in Figure 2.2.

In general, a standing wave solution to (2.2), with $\exp(-jkct)$ implicit, is represented by the following linear combination:

$$\Psi(r, \Omega, k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left[A_{nm}\mathcal{J}_n(kr) + B_{nm}\mathcal{N}_n(kr)\right] Y_n^m(\Omega). \tag{2.10}$$

Alternatively, a traveling wave solution is represented by

$$\Psi(r, \Omega, k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left[C_{nm}h_n^{(1)}(kr) + D_{nm}h_n^{(2)}(kr)\right] Y_n^m(\Omega). \tag{2.11}$$

2.3 Mathematical approach: representation theory

The spherical harmonic functions are orthonormal on the unit sphere $\mathbb{S}^2$ [99, 100]:

$$\int_{\Omega\in\mathbb{S}^2} Y_n^m(\Omega)\overline{Y_{n'}^{m'}}(\Omega)d\Omega = \delta_{nmn'm'}, \tag{2.12}$$

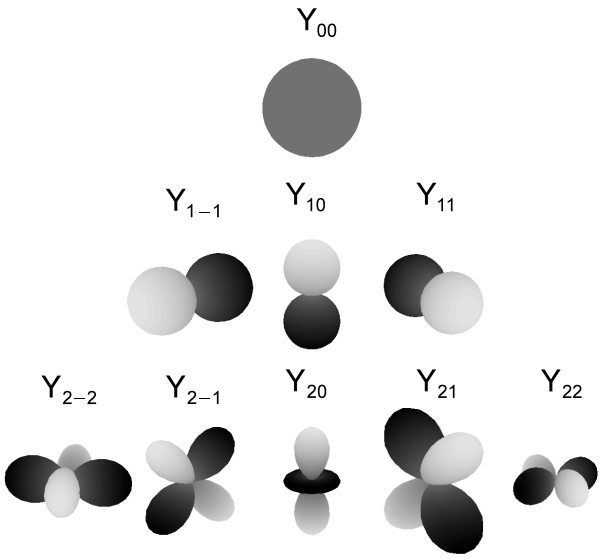


Figure 2.2: Polar plots of the magnitude of the spherical harmonics $|Y_{nm}(\Omega)|$, defined for all directions Ω in the unit sphere $\mathbb{S}^2$, for orders $n = 0, 1, 2$, and degrees $m = -n, ..., n$, respectively. The spherical harmonics form a complete set of orthonormal functions. Any square-integrable function on $\mathbb{S}^2$ can thus be expanded as a linear combination of spherical harmonics.

where $d\Omega = \cos\phi d\phi d\theta$, the overline denotes complex conjugate, and δ is the Kronecker Delta. The spherical harmonics also form a complete set of functions when indexed by their order n and degree m. This is expressed by the following completeness relation [99, 100]:

$$\sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\Omega)\overline{Y_n^m}(\Omega') = \delta(\Omega - \Omega'), \tag{2.13}$$

where the Dirac delta on $\mathbb{S}^2$ is given by $\delta(\Omega - \Omega') = \delta(\theta - \theta')\delta(\sin\phi - \sin\phi')$. Because the set of spherical harmonics is orthonormal and complete, it constitutes an orthonormal basis for the set of square-integrable function on $\mathbb{S}^2$.

Let now consider a sound pressure field Ψ with finite energy along all directions. Since this kind of sound pressure field can be modeled as a square-integrable function on $\mathbb{S}^2$, it can be represented on the orthonormal basis provided by the set of spherical harmonic functions. This originates the spherical Fourier transform (SFT) [99, 100]:

$$\mathcal{S}\{\Psi\} = \Psi_{nm} = \int_{\Omega\in\mathbb{S}^2} \Psi(\Omega)\overline{Y_n^m}(\Omega)d\Omega, \tag{2.14}$$

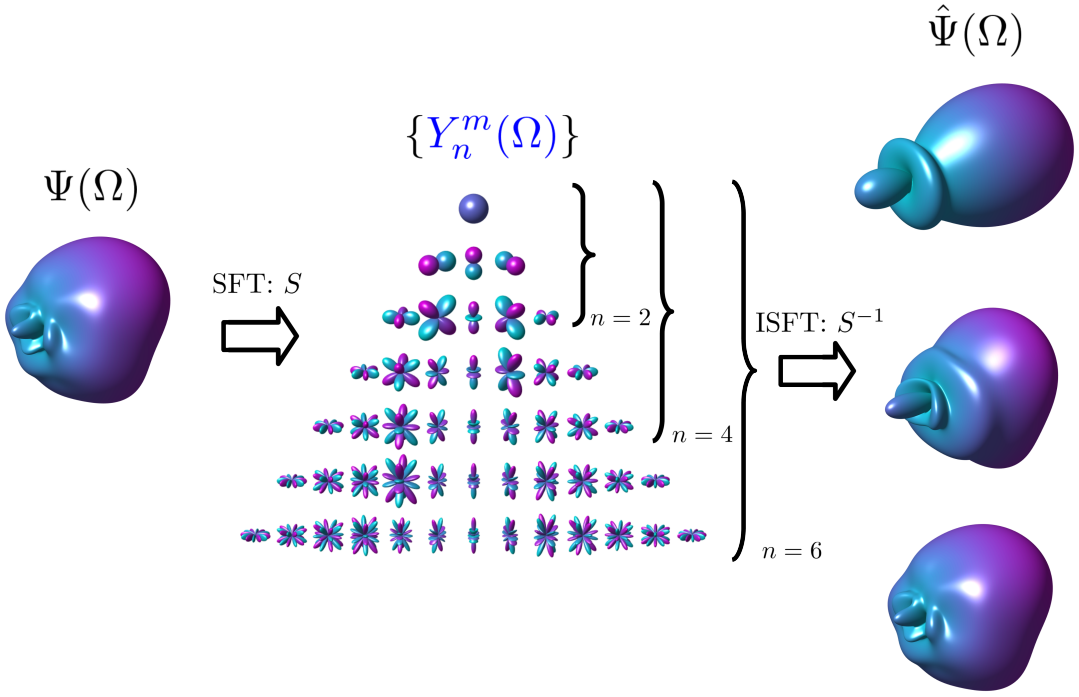


Figure 2.3: Example of using the spherical Fourier transform (SFT) and the inverse spherical Fourier transform (ISFT) for representing a function Ψ on the sphere. The higher the order n, the better the angular variations are represented.

and the inverse spherical Fourier transform (ISFT) [99, 100]:

$$\mathcal{S}^{-1}\{\Psi_{nm}\} = \Psi(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \Psi_{nm} Y_n^m(\Omega). \tag{2.15}$$

In this context, it is important to mention that Ψ is not limited to denote sound pressure fields. In general, Ψ can denote a distribution of information on the sphere, as long as it can still be modeled as a square-integrable function on $\mathbb{S}^2$.

An example that illustrates the action of representing a function Ψ on the sphere based on the spherical Fourier transform, and its recovery at distinct resolutions based on the inverse spherical Fourier transform, is shown in Fig. 2.3. It can be observed that the order of the representation provides a measure of the angular bandwidth of the functions. In fact, the higher the order, the better the rapid angular variations are represented.

The Legendre addition theorem for the spherical harmonics is widely used to simplify

mathematical expressions. This theorem is formulated as follows [99, 100]:

$$\sum_{m=-n}^{n} Y_n^m(\Omega_1)\overline{Y_n^m}(\Omega_2) = \frac{2n+1}{4\pi} P_n(\cos\Theta_{12}), \tag{2.16}$$

where Θ_{12} denotes the angle between the directions Ω_1 and Ω_2.

Performing a transform $\mathcal{S}$ followed by an inverse transform $\mathcal{S}^{-1}$, up to a finite order N, can for instance be interpreted in the light of the addition theorem in (2.16) as an integral equation on the sphere [101–104]. From a proper combination of (2.14), (2.15), and (2.16), it can indeed be shown that

$$\hat{\Psi}(\Omega_2) = \int_{\Omega_1 \in \mathbb{S}^2} K_N(\Omega_1, \Omega_2)\Psi(\Omega_1) d\Omega_1, \tag{2.17}$$

where K_N is referred to as the integral kernel and is defined by

$$K_N(\Omega_1, \Omega_2) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\cos\Theta_{12}). \tag{2.18}$$

This compact form is particularly useful for investigating the combined action $\mathcal{S}^{-1}\mathcal{S}$, since it only requires to understand the properties of the integral kernel K_N.

The integral equation (2.17) also constitutes the theoretical basis for an interpolation method on the sphere. It is used to approximate a function at an arbitrary direction Ω_2 from their values in all directions $\Omega_1 \in \mathbb{S}^2$. In practice, the integral is calculated numerically from a finite number of directions. For this reason, K_N is also known as the interpolating kernel.

Finally, an important consequence of the orthonormality property in (2.12) is the possibility of integrating the product of two functions on the sphere, $\Psi^{(1)}$ and the complex conjugate of $\Psi^{(2)}$, by performing an equivalent sum of products in the transform domain. This is mathematically formulated as follows [99, 100]:

$$\int_{\Omega \in \mathbb{S}^2} \Psi^{(1)}(\Omega)\overline{\Psi^{(2)}}(\Omega) d\Omega = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \Psi_{nm}^{(1)} \overline{\Psi^{(2)}}_{nm}. \tag{2.19}$$

Here, $\Psi_{nm}^{(1)}$ and $\Psi_{nm}^{(2)}$ denote the spherical Fourier transforms of $\Psi^{(1)}$ and $\Psi^{(2)}$, respectively.

2.4 Spatial signal processing approach

The results of spherical acoustic theory rely on assuming that the spatial information is a square-integrable function defined on the unit sphere, that is, a functions on a continuous spherical domain. Yet in practice, spatial information on the sphere necessarily requires to be represented by means of a finite number of transducers, for instance, a microphone array for recording or a loudspeaker array for reproduction.

In analogy with sampling theory on the time domain, obtaining a discrete version of a continuous function defined on the sphere can be regarded as a process composed of two stages: 1) limitation of the angular bandwidth, and 2) angular sampling, as shown in Fig. 2.4. Angular bandwidth is defined as a measure of the variation of a function along directions. Two adjacent microphones, for example, will be able to sample a continuous sound pressure field at their corresponding positions with sufficient accuracy only in the case when the amount of variation is comparable with the distance between the two microphones. Angular band limitation has therefore the role of ensuring a limited angular variation so that the spatial information on the sphere can be described with sufficient accuracy by using a finite number of samplings.

A measure of the angular bandwidth is the order n used in spherical Fourier transforms defined in (2.14) and (2.15), as was exemplified in Fig. 2.3. Since a finite number of data points on the sphere is associated with a finite angular bandwidth, finite-order spherical Fourier representations are also associated with finite angular resolutions. The spherical Fourier transform therefore provide a suitable domain for the directional encoding and decoding of spatial information at scalable angular resolutions. The main benefit of this approach is that the spatial signal processing can be simplified when it is performed in the transform domain. Moreover, spatial signal processing is always in consistency with acoustic propagation phenomena. This originates the general scheme for spatial signal

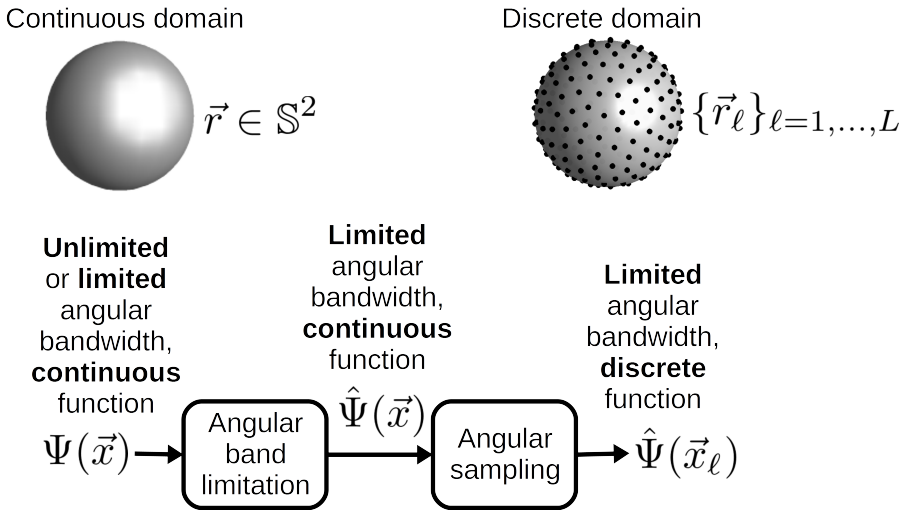


Figure 2.4: Sampling on the sphere. The angular bandwidth is a measure of the variation of a function along directions

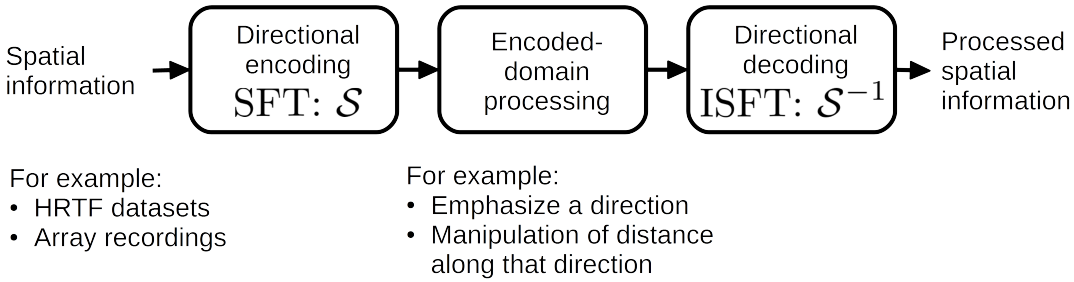


Figure 2.5: Spatial signal processing.

processing shown in Fig. 2.5.

2.5 Summary

The theory of spherical acoustics allows for the stable representation or encoding of spatial sound information at scalable directional resolutions. This in turn facilitates the spatial signal processing and analysis, required for instance in binaural systems.

Chapter 3

Synthesis of HRTFs for arbitrary source positions from boundary values

3.1 Overview

Binaural presentation would ideally require knowing the head-related transfer functions (HRTFs) for all points in three-dimensional space. However, measuring or calculating these HRTF datasets is a complex and time consuming task. This chapter presents models for the synthesis of HRTFs for arbitrary sound source positions. The input to these models are the values of the HRTFs on a boundary at a given initial distance. These models constitute a useful result, since spherical datasets are typically obtained for only one far distance [38]. Moreover, prior knowledge about the HRTFs at the target positions is not required when applying these models.

An existing model for synthesis on the three-dimensional space from values on a spherical boundary is first reviewed. For practical purposes, two existing models for synthesis on the horizontal plane from values on a circular boundary are also reviewed. The first contribution of this chapter is an additional horizontal-plane model that outperforms the

existing ones. The second contribution is a general method for limiting the angular bandwidth of the continuous models, which can be applied to all existing models, disregarding whether they are based on spherical or circular boundaries.

3.2 Distance-varying filters (DVFs) for synthesis of HRTFs

Distance varying filters (DVFs) are used to synthesize the HRTFs for arbitrary sound source positions from initial HRTFs obtained for positions on a boundary at a single distance from the head of a listener.

Synthesis with DVFs relies on the principle of acoustic reciprocity [105] to consider HRTF datasets as sound pressure fields generated by a point source at the ear. This approach originates from the solutions to the acoustic wave equation, where solutions can be separated into a radial and an angular part [83, ch. 6]. The angular part defines an orthonormal basis of functions along angles, which is used to represent the directional information in the HRTFs that will remain unchanged. The radial part is then used to formulate the DVFs in the represented domain. DVFs therefore operate on a directionally encoded or transformed domain. Prior knowledge about the HRTFs at the target positions is not required when applying this model-based approach.

3.3 Spherical DVFs for synthesis on the three-dimensional space

Spherical DVFs [91–93] make it possible to calculate the HRTFs for arbitrary positions in space, once an initial HRTF dataset for discrete positions on a sphere surrounding the listener is given. The fundamentals of the methods presented in [91–93] are reviewed in this Section.

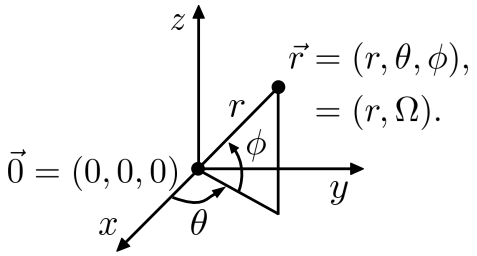


Figure 3.1: Spherical coordinate system.

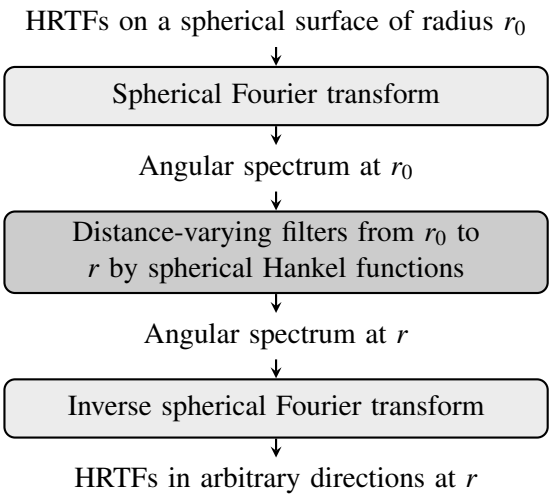


Figure 3.2: Block diagram of the three-dimensional synthesis method.

3.3.1 Continuous spherical boundary

In this section, we discuss the fundamentals of the synthesis method used in [91–93] when continuous distributions of sources are assumed. This method is formulated on the coordinate system shown in Fig. 3.1, a point in space $\vec{r} = (r, \theta, \phi)$ is specified by its radial distance r, azimuth angle $\theta \in [-\pi, \pi]$ and elevation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Angles are merged into the variable $\Omega = (\theta, \phi)$ in such a way that a point in space is also represented by $\vec{r} = (r, \Omega)$. This method is formulated by solving the acoustic wave equation in spherical coordinates. An overview of this method is illustrated in Fig. 3.2.

By $\Psi^{\mathrm{SPH}}(r, \theta, \phi, t) \propto \Re(r)\Theta(\theta)\Phi(\phi)\exp(-jkct)$, we denote a sound field varying independently along each one of its coordinates. In all of what follows k will denote the wave number as defined in (2.5). The upper index refers to the spherical coordinate system. As previously described in (2.3) and (2.4), sound fields of this form decouple the wave

equation in (2.2) into a sum of ordinary differential equations [83, pp. 183]:

$$\frac{1}{\Re}\frac{d}{dr}\left[r^2\frac{d}{dr}\right]\Re + \underbrace{\frac{1}{\cos^2\phi}\underbrace{\frac{1}{\Theta}\frac{d^2}{d\theta^2}\Theta}_{=-m^2} + \frac{1}{\cos\phi}\frac{1}{\Phi}\frac{d}{d\phi}\left[\cos\phi\frac{d}{d\phi}\right]\Phi}_{=-n(n+1)} = -r^2k^2. \tag{3.1}$$

Three-dimensional solutions to (3.1) are obtained by using separation constants m and n. The product of the angular solutions $\Theta(\theta)\Phi(\phi) = Y_n^m(\theta,\phi) = Y_n^m(\Omega)$ produces the spherical harmonics of order n and degree m as defined in (2.9). The radial term produces a Bessel differential equation; one of its solutions is given in terms of the spherical Hankel function of the second kind, which we simply denote by $h_n(kr)$.

The sound pressure field in the frequency domain is thus defined by the solution

$$\Psi^{\mathrm{SPH}}(r,\Omega,k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} C_{nm}^{\mathrm{SPH}}(k)h_n(kr)Y_n^m(\Omega). \tag{3.2}$$

In connection with the general solution for traveling waves given in (2.11), this solution is only considering the waves that travel inwards and converge to the origin of coordinates. This is enough when synthesizing HRTFs, since the radiation of sound towards the center of the listener's head is of main interest.

Because coefficients C_m^{SPH} are invariant with respect to distance, the sound pressure fields at two different distances r and r_0 are linked as follows:

$$\Psi^{\mathrm{SPH}}(r,\Omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\Omega) \underbrace{\left[\frac{h_n(kr)}{h_n(kr_0)}\right]}_{\text{Filters: } D_n^{\mathrm{SPH}}(r,r_0,k)} \times \underbrace{\int_{\Omega_0\in\mathbb{S}^2}\Psi^{\mathrm{SPH}}(r_0,\Omega_0)\overline{Y_n^m}(\Omega_0)d\Omega_0}_{\text{Spherical Fourier transform: } \mathcal{S}\{\Psi^{\mathrm{SPH}}\}}. \tag{3.3}$$

Here, the term in the second line represents a spherical Fourier transform along Ω_0. The term in brackets defines the spherical DVFs. Finally, the sum of spherical harmonics describes an inverse spherical Fourier transform.

Applying (3.3) to spherical HRTF datasets defines the synthesis method shown in Fig. 3.2. Equation (3.3) first decomposes the dataset using an integral over Ω_0 and then

reconstructs it at a target angle Ω. Note that these two variables are independent.

3.3.2 Limitation of the bandwidth of angular spectra

Methods for the limitation of the angular bandwidth are necessary to ensure the stable discretization of the continuous model. Limited angular bandwidths are modeled by truncating the sums along n in (3.3) up to a maximum order. In [94], a frequency-dependent threshold to decide maximum orders was deduced by using the large-order asymptotic form of Hankel functions. Bandwidth limits yielded by this threshold, however, are overestimated and present the risk of emphasizing high-order components that typically have a low signal-to-noise ratio [83].

We have observed that the DVFs in (3.3) have magnitude responses that are monotonically increasing functions of order and relative distance. Based on this observation, we propose an alternative approach for restricting the angular bandwidth by clipping the magnitude responses of ideal DVFs according to the ratio between the surface area of two hypothetical concentric spheres at r_0 and r.

In summary, we formulate both the magnitude-dependent band-limiting threshold using the following expression:

$$\hat{D}^{\mathrm{SPH}}(r, r_0, k) = \begin{cases} s \cdot D^{\mathrm{SPH}}(r, r_0, k) & \text{if } |D^{\mathrm{SPH}}| \leq s^{-2}, \\ 0 & \text{else.} \end{cases} \tag{3.4}$$

Here, $\hat{D}^{\mathrm{SPH}}$ defines a band-limited distance-varying filter, D^{SPH} is the DVF defined in (3.3), and s is a scaling factor defined as follows:

$$s = \frac{r}{r_0}. \tag{3.5}$$

3.4 Circular DVFs for synthesis on the horizontal-plane

The method presented in Sec. 3.3 was a useful result, since spherical datasets are typically obtained for only one far distance [38]. Nevertheless, optimally distributing sound sources on the sphere is a non-trivial problem [106, 107], and obtaining the HRTFs for all required initial positions might not be feasible for many applications.

In many applications, it is enough to present sounds from positions on a horizontal plane at the height of the listener's ears, even though this may not be enough to present complex scenes such as those including reverberations. The reason behind this lies in the fact that real life sound sources are often located on or close to this horizontal plane. Furthermore, the human auditory system can resolve sounds more accurately in this plane [40]. Consequently, horizontal-plane virtual auditory displays would be an important use case. Obtaining HRTFs for a dense set of positions throughout the plane, however, can still be very demanding and time consuming. It would be useful, then, if a method similar to the one presented in [91–93] could be applied to synthesize the HRTFs for arbitrary positions on the horizontal plane, once an initial, circular HRTF dataset is known at a single (typically far) distance.

Ambiguous results are obtained when deriving methods for the horizontal plane by simply restricting the spherical representation to a circle [90, 108]. This is because, even when considering only a plane, sound still propagates along three dimensions.

Two approaches for horizontal-plane synthesis of HRTFs that take into consideration the spatial propagation of sound have been explored in [97]. Both rely on finding horizontal-plane solutions to the three-dimensional acoustic wave equation by assuming sound fields that are invariant along a spatial coordinate [83, chs. 4 and 6].

The first approach assumes invariance along height in cylindrical coordinates. This assumption is equivalent to considering vertical, infinite-length line sources of sound. Approximate solutions for the case of finite-length line sources have been proposed in [109] based on the stationary phase method [83, pp. 137-140]. However, these solutions are only

accurate at a specified distance from the origin. In either way, the main limitation of modeling in cylindrical coordinates is the impossibility of considering the spherical symmetry of sound field propagation due to external sound sources. As a result, the free-field pressure decay along radial distance towards the head's center can not be accurately predicted. The accurate prediction of such distance decay pattern is an important requirement because it matches the distance decay due to point sources of sound, which are the kind of sources used in the definition of HRTFs.

The second approach attempts to account for the spherical, radial convergency towards the head's center by assuming invariance along vertical angles measured from the horizontal plane, which are described by the polar angle β in the interaural coordinate system shown in Fig. 3.3(b). For a given distance, this assumption is equivalent to considering vertical, circular sources of sound that are parallel to the median plane ($xz-$plane), and whose radii decrease as they move away from this plane. This leads to a method that requires the division of the horizontal plane into two regions, in front of and behind the listener, delimited by an axis connecting the ears' canals ($y-$axis). In practice, this can be understood as considering a semicircular array of circular sources (in front or behind), as illustrated in Fig. 3.4(b) for the frontal case. Nevertheless, this method yields results on the front and back regions that take different values along the interaural axis. Discontinuities on the lateral sides are hence inevitable, precisely where the distance cues are more prominent [37].

In this section, we properly account for the spherical, radial convergency towards the head's center by assuming invariance along the elevation angle ϕ defined in the spherical coordinate system shown in Fig. 3.3(c). This assumption is equivalent to considering vertical, semicircular sources of sound that connect points of equal azimuthal angle θ and are terminated by the north and south poles. The vertical semicircles have equal radii; they are also continuously and uniformly distributed around the $z-$axis. In practice, this can be interpreted as considering a circular array of semicircular sources, as illustrated in Fig. 3.4(c). The underlying assumption therefore leads to solutions to the acoustic wave

equation that uniformly consider the horizontal plane as a whole. Consequently, there are no singularities or discontinuities in the results. The angular part of the solutions results in a suitable basis for representing directional information on full horizontal circles. The radial part is used to formulate a set of distance-varying filters to be applied in the represented domain. Our solutions, hence, define a method to synthesize HRTFs at arbitrary distances on the horizontal plane, once a circular HRTF dataset is known.

3.4.1 Continuous circular boundary

In this Section, the methods in cylindrical and interaural coordinates are first reviewed. Then, the proposal in spherical coordinates is presented.

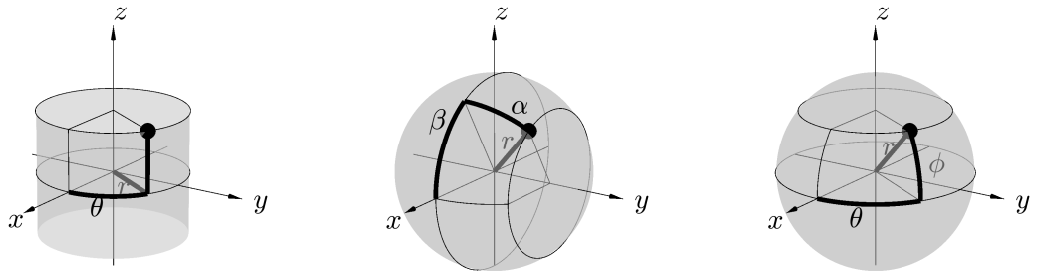
3.4.1.1 Cylindrical coordinates

In this section, we discuss the fundamentals of the synthesis method used in [97] when continuous distributions of sources are assumed. This method is formulated on the coordinate system shown in Fig. 3.3(a), where a point in space is represented by its radial distance r, azimuthal angle $\theta \in [-\pi, \pi]$, and height z. The use of these coordinates is equivalent to considering vertical line sources of infinite length. This method can hence be formulated by solving the acoustic wave equation in cylindrical coordinates, where sound fields are assumed to be invariant along height. An overview of this method is illustrated in Fig. 3.5.

By $\Psi^{\text{cyl}}(r, \theta, z, t) \propto \mathrm{R}(r)\Theta(\theta)\mathrm{Z}(z)\exp(-jkct)$, we denote a sound field varying independently along each one of its coordinates. The upper index refers to the cylindrical coordinate system. Sound fields of this form decouple the wave equation into a sum of ordinary differential equations [83, pp. 116]:

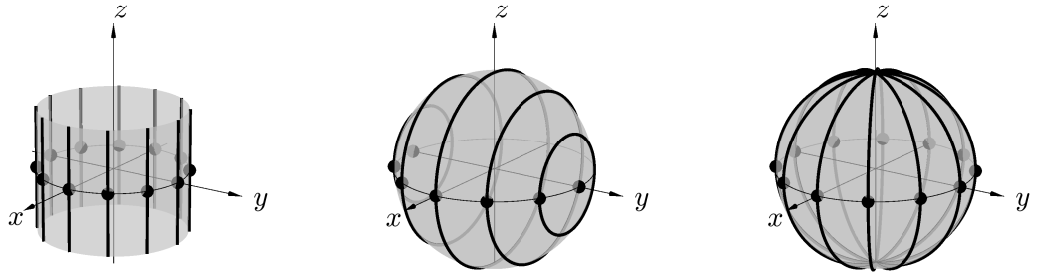
$$\frac{1}{\mathrm{R}}\frac{d}{dr}\left[\frac{d}{dr}+\frac{1}{r}\right]\mathrm{R}+\frac{1}{r^2}\underbrace{\frac{1}{\Theta}\frac{d^2}{d\theta^2}\Theta}_{=-m^2}+\frac{1}{\mathrm{Z}}\frac{d^2}{dz^2}\mathrm{Z}=-k^2. \tag{3.6}$$

Solutions to (3.6) on a horizontal plane are obtained using separation constant m and



(a) Cylindrical coordinates. (b) Interaural coordinates. (c) Spherical coordinates.

Figure 3.3: Coordinate systems for describing the sound source positions used to obtain HRTF datasets on the horizontal plane. The front position is along the positive *x*-axis. The ears of the listener are along the *y*-axis.



(a) A circular array of data points is replicated along vertical lines of infinite length.

(b) A semicircular array of data points is replicated along vertical circles.

(c) A circular array of data points is replicated along vertical semicircles.

Figure 3.4: Replicating an array of data points.

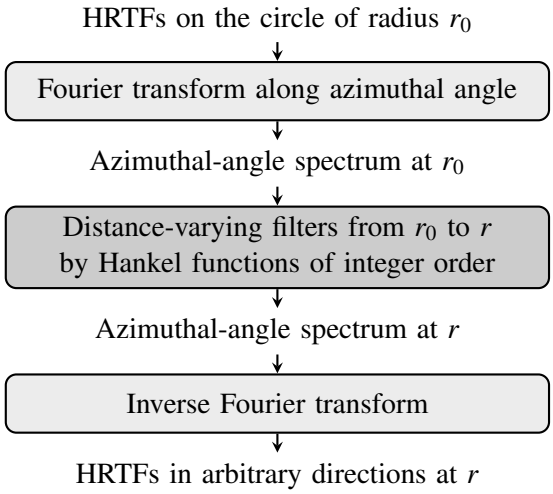


Figure 3.5: Block diagram of the synthesis method in cylindrical coordinates.

assuming that $\mathrm{Z}(z)$ is constant. The term involving $\mathrm{Z}(z)$ thus vanishes and the azimuthal solution oscillates according to $\Theta(\theta) = \exp(jm\theta)$, where $m \in \mathbb{Z}$. The radial term thereof produces a Bessel differential equation; one of its solutions is given in terms of the Hankel function of the second kind and integer order $\mathrm{R}(r) = H_n(kr)$, also known as the cylindrical Hankel function.

Sound pressure fields can be defined by the general solution to (3.6) in the frequency domain according to

$$\Psi^{\mathrm{cyl}}(r, \theta, k) = \sum_{m=-\infty}^{\infty} C_m^{\mathrm{cyl}}(k) H_n(kr) \exp(jm\theta). \tag{3.7}$$

By virtue of the orthonormality of complex exponentials, the coefficients C_m^{cyl} can be calculated using the following expression:

$$C_m^{\mathrm{cyl}}(k) = \frac{1}{H_m(kr)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi^{\mathrm{cyl}}(r, \theta, k) \exp(-jm\theta) d\theta. \tag{3.8}$$

Note that distance and angle variations in (3.7) are fully represented by the Hankel functions and the complex exponentials, respectively. Coefficients C_m^{cyl} hence exclusively depend on k and (3.8) holds for any other distance.

Given the independence of C_m^{cyl} with respect to distance, the relation between the sound

pressure fields at two different distances r and r_0 can be formulated as follows:

$$\Psi^{\mathrm{cyl}}(r,\theta,k) = \sum_{m=-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp(jm\theta) \underbrace{\left[\frac{H_m(kr)}{H_m(kr_0)}\right]}_{\text{Filters: } D_m^{\mathrm{cyl}}(r,r_0,k)} \times \underbrace{\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Psi^{\mathrm{cyl}}(r_0,\theta_0,k)\exp(-jm\theta_0)d\theta_0.}_{\text{Fourier transform along azimuthal angle: } \mathcal{F}\{\Psi^{\mathrm{cyl}}\}} \tag{3.9}$$

Here, the term on the second line is a Fourier transform along θ_0, the term above in brackets defines distance-varying filters in the horizontal plane, and the sum of complex exponentials is an inverse Fourier transform.

The principle of reciprocity [105] implies that HRTF datasets can also be characterized by the sound pressure around the listener due to a point source at each of their ears. Applying (3.9) to these datasets then defines the synthesis method shown in Fig. 3.5. Equation (3.9) first decomposes the dataset using an integral over θ_0 and then reconstructs it at a target angle θ. It should be noted that these two variables are independent. Because this method was derived by assuming infinite-length line sources, the pressure decay along radial distance towards the head's center can not be accurately predicted.

3.4.1.2 Interaural coordinates

In this section, we discuss the fundamentals of the synthesis method used in [97] when continuous distributions of sources are assumed. This method is formulated on the coordinate system shown in Fig. 3.3(b), where a point in space is represented by its radial distance r, lateral angle $\alpha \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, and polar angle $\beta \in [-\pi, \pi]$. These coordinates are a suitable choice for representing the underlying assumption that the interaural HRTF (the left ear HRTF divided by the right ear HRTF) is constant along polar angle β for the hypothetical case of a spherical head. This method can hence be formulated by solving the acoustic wave equation in interaural coordinates, where sound fields are assumed to be invariant along vertical circles parallel to the median plane (xz-plane) and centered on the interaural

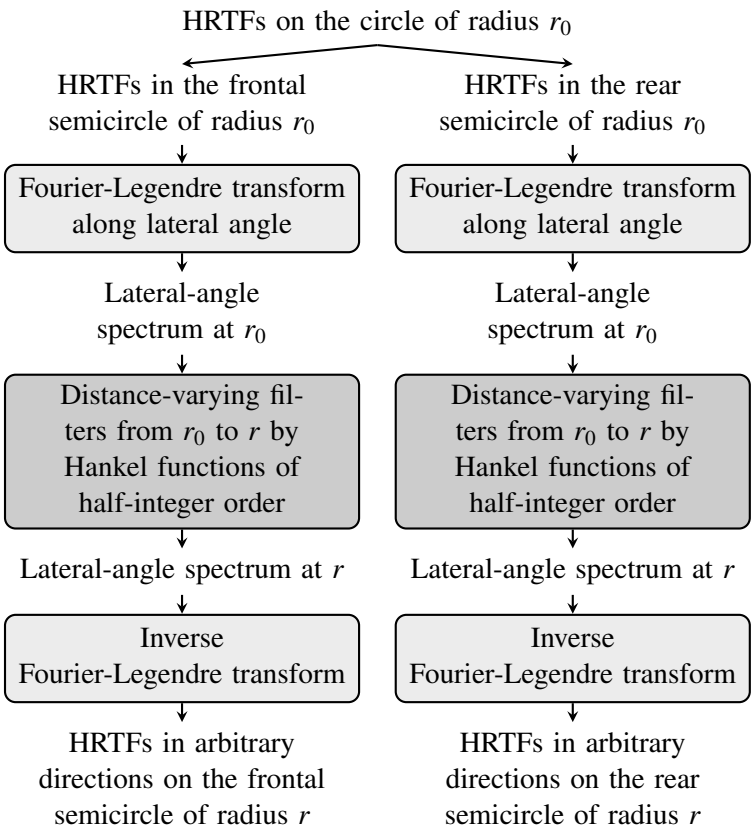


Figure 3.6: Block diagram of the synthesis method in interaural coordinates [97].

axis (y-axis). An overview of this method is illustrated in Fig. 3.6.

By $\Psi^{\text{int}}(r, \alpha, \beta, t) \propto \mathrm{R}(r)\mathrm{A}(\alpha)\mathrm{B}(\beta)\exp(-jkct)$, we denote a time-harmonic sound pressure field with independent spatial variables, where $j = \sqrt{-1}$. The upper index refers to the interaural coordinate system. Sound fields of this form simplify the three-dimensional wave equation to a sum of three ordinary differential equations such as the ones shown in the following expression [83, pp. 183]:

$$\frac{1}{\mathrm{R}}\frac{d}{dr}\left[r^2\frac{d}{dr}\right]\mathrm{R} + \underbrace{\frac{1}{\cos\alpha}\frac{1}{\mathrm{A}}\frac{d}{d\alpha}\left[\cos\alpha\frac{d}{d\alpha}\right]\mathrm{A} + \frac{1}{\cos^2\alpha}\frac{1}{\mathrm{B}}\frac{d^2}{d\beta^2}\mathrm{B}}_{=-n(n+1)} = -r^2k^2, \tag{3.10}$$

where c is the speed of sound in air.

Solutions to (3.10) on the horizontal plane (xy-plane) are obtained by setting a separation constant n and assuming $\mathrm{B}(\beta)$ is constant. The term involving $\mathrm{B}(\beta)$ thus vanishes. The term involving $\mathrm{A}(\alpha)$ results in a Legendre differential equation, whose so-

lution $\mathrm{A}(\alpha) = P_n(\sin\alpha)$ is the Legendre polynomial of order $n \in \mathbb{N}$. The radial term thereof produces a Bessel differential equation; one of its solutions is given in terms of the Hankel function of the second kind and half-integer order $H_{n+\frac{1}{2}}(kr)$. The radial term $\mathrm{R}(r) = r^{-\frac{1}{2}} H_{n+\frac{1}{2}}(kr)$ is known as the spherical Hankel function.

Sound pressure fields can be defined by the general solution to (3.10) in the frequency domain according to

$$\Psi^{\mathrm{int}}(r, \alpha, k) = \sum_{n=0}^{\infty} C_n^{\mathrm{int}}(k) r^{-\frac{1}{2}} H_{n+\frac{1}{2}}(kr)\, P_n(\sin\alpha). \tag{3.11}$$

By virtue of the orthonormality property for P_n, the coefficients C_n^{int} can be calculated using the following expression:

$$C_n^{\mathrm{int}}(k) = \frac{n+\frac{1}{2}}{r^{-\frac{1}{2}} H_{n+\frac{1}{2}}(kr)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Psi^{\mathrm{int}}(r, \alpha, k) P_n(\sin\alpha) \cos\alpha d\alpha. \tag{3.12}$$

Note that distance and angle variations in (3.11) are fully represented by $\mathrm{R}(r)$ and P_n, respectively. Coefficients C_n^{int} hence exclusively depend on k and (3.12) holds for any other distance.

Given the independence of C_n^{int} with respect to distance, the relation between the sound pressure fields at two different distances r and r_0 can be formulated as follows:

$$\begin{aligned}\Psi^{\mathrm{int}}(r, \alpha, k) = \sum_{n=0}^{\infty} \left(n+\frac{1}{2}\right)^{\frac{1}{2}} P_n(\sin\alpha) \underbrace{\left[\frac{r^{-\frac{1}{2}} H_{n+\frac{1}{2}}(kr)}{r_0^{-\frac{1}{2}} H_{n+\frac{1}{2}}(kr_0)}\right]}_{\text{Filters: } D_n^{\mathrm{int}}(r, r_0, k)} \times \\ \underbrace{\left(n+\frac{1}{2}\right)^{\frac{1}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Psi^{\mathrm{int}}(r_0, \alpha_0, k) P_n(\sin\alpha_0) \cos\alpha_0 d\alpha_0}_{\text{Fourier-Legendre transform along lateral angle: } \mathcal{L}\{\Psi^{\mathrm{int}}\}}.\end{aligned} \tag{3.13}$$

Here, the term on the second line is a Fourier-Legendre transform along α_0, the term above in brackets defines distance-varying filters in a horizontal half-plane, and the sum of Legendre polynomials on α is an inverse Fourier-Legendre transform.

The principle of reciprocity [105] implies that HRTF datasets can also be characterized by the sound pressure around the listener due to a point source at each of their ears. Applying (3.13) to these datasets then defines the synthesis method shown in Fig. 3.6. Given that lateral angles are in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, Fourier-Legendre transforms need to act on separate datasets defined over two semicircles in front and behind the listener. Unfortunately, this leads to discontinuous results along the interaural axis. Equation (3.13) first decomposes the dataset using an integral over α_0 and then reconstructs it at a target angle α. It should be noted that these two variables are independent.

3.4.1.3 Spherical coordinates

In this section, we detail our method to synthesize the HRTFs at arbitrary points on the horizontal plane from HRTF datasets given on a single circular boundary. Continuous distributions of sources are assumed throughout this section. The proposal is formulated on the coordinate system shown in Fig. 3.3(c), where a point in space is represented by its radial distance r, azimuthal angle $\theta \in [-\pi, \pi]$, and elevation angle $\phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The underlying idea is to generate spherical datasets by replicating circular datasets along semicircles connecting points of equal azimuth θ and terminated by the north and south poles. Semicircles of this type define the meridians. Our method thus originates from an equivalent problem consisting of solving the acoustic wave equation on these coordinates, under the assumption that sound fields are invariant along meridians. Although this assumption can appear one step removed from physical reality when considering the three-dimensional space, it is accepted here for the practical purpose of having a uniform distribution of semicircular sources on a circle, as shown in Fig. 3.4(c). This assumption allows us to derive a set of distance-varying filters that takes into consideration the radial symmetry of acoustic propagation towards the head center. An overview of our method is illustrated in Fig. 3.7.

By $\Psi^{\text{sph}}(r, \theta, \phi, t) \propto \Re(r)\Theta(\theta)\Phi(\phi)\exp(-jkct)$, we denote a sound field varying independently along each one of its coordinates. The upper index refers to the spherical coordinate system. Sound fields of this form also decouple the wave equation into a sum of ordinary

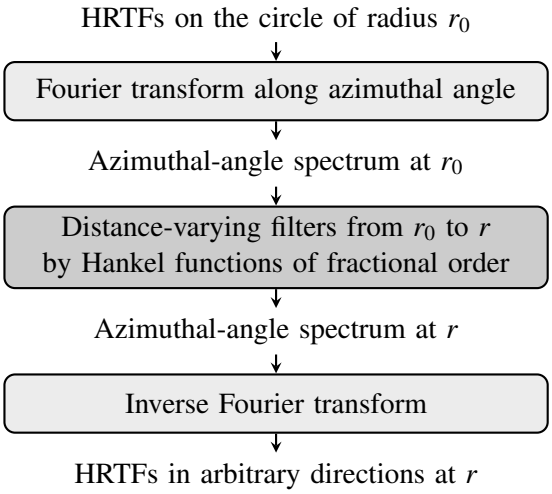


Figure 3.7: Block diagram of the synthesis method in spherical coordinates.

differential equations [83, pp. 183]:

$$\frac{1}{\Re}\frac{d}{dr}\left[r^2\frac{d}{dr}\right]\Re + \frac{1}{\cos^2\phi}\underbrace{\frac{1}{\Theta}\frac{d^2}{d\theta^2}\Theta}_{=-m^2} + \frac{1}{\cos\phi}\frac{1}{\Phi}\frac{d}{d\phi}\left[\cos\phi\frac{d}{d\phi}\right]\Phi = -r^2k^2. \tag{3.14}$$

We look for solutions to (3.14) on a horizontal plane using separation constant m. We assume that $\Phi(\phi)$, the elevation-dependent portion of the solution, is constant. The term involving $\Phi(\phi)$ thus vanishes and the azimuthal solution oscillates according to $\Theta(\theta) = \exp(jm\theta)$, where $m \in \mathbb{Z}$.

We deal with the remaining radial term in (3.14) using auxiliary function $\mathcal{R}(r) = r^{\frac{1}{2}}\Re(r)$ to obtain the following Bessel differential equation for it:

$$\frac{d^2}{dr^2}\mathcal{R} + \frac{1}{r}\frac{d}{dr}\mathcal{R} + \left[k^2 - \frac{\left(\frac{m^2}{\cos^2\phi} + \frac{1}{4}\right)}{r^2}\right]\mathcal{R} = 0. \tag{3.15}$$

For the sake of simplicity, we denote the term inside the parentheses by a parameter μ as follows:

$$\mu^2 = \frac{m^2}{\cos^2\phi} + \frac{1}{4}, \tag{3.16}$$

where $\phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Solutions to (3.15) can now be obtained in terms of the well-known Bessel functions, among which we choose $\mathcal{R}(r) = H_\mu(kr)$, the Hankel function of the

second kind, and fractional order μ. The radial solution finally results in $\Re(r) = r^{-\frac{1}{2}} H_\mu(kr)$.

We can thus define the sound pressure field in a horizontal plane in terms of our general solution to (3.14) in the frequency domain in such a way that

$$\Psi^{\text{sph}}(r, \theta, k) = \sum_{m=-\infty}^{\infty} C_m^{\text{sph}}(k) r^{-\frac{1}{2}} H_\mu(kr) \exp(jm\theta). \tag{3.17}$$

We interpret this result as an incoming converging wave due to a radiating semicircular arc along a meridian. As a consequence of the orthonormality of complex exponentials, coefficients C_m^{sph} remain defined as follows:

$$C_m^{\text{sph}}(k) = \frac{1}{2\pi r^{-\frac{1}{2}} H_\mu(kr)} \int_{-\pi}^{\pi} \Psi^{\text{sph}}(r, \theta, k) \exp(-jm\theta) d\theta. \tag{3.18}$$

Note that distance and angle variations in (3.17) are fully described by $\Re(r)$ and $\exp(-jm\theta)$, respectively. Coefficients C_m^{sph} hence depend on k only and (3.18) holds for any distance.

Because coefficients C_m^{sph} are invariant with respect to distance, the sound pressure fields at two different distances r and r_0 are finally linked according to the following expression:

$$\begin{aligned} \Psi^{\text{sph}}(r, \theta, k) = \sum_{m=-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp(jm\theta) \underbrace{\left[\frac{r^{-\frac{1}{2}} H_\mu(kr)}{r_0^{-\frac{1}{2}} H_\mu(kr_0)}\right]}_{\text{Filters: } D_m^{\text{sph}}(r, r_0, k)} \times \\ \underbrace{\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \int_{-\pi}^{\pi} \Psi^{\text{sph}}(r_0, \theta_0, k) \exp(-jm\theta_0) d\theta_0}_{\text{Fourier transform along azimuthal angle: } \mathcal{F}\{\Psi^{\text{sph}}\}}. \end{aligned} \tag{3.19}$$

Here, the term in the second line represents a Fourier transform along θ_0. The term in brackets defines the proposed set of horizontal-plane distance-varying filters with parameter μ defined in (3.16). Note that μ can correspond to horizontal planes at different elevation angles $\phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. For the goals of this study, we delimit our discussion hereafter to the case $\phi = 0$. Finally, the sum of complex exponentials on θ describes an inverse Fourier transform.

Applying (3.19) to HRTF datasets defines the synthesis method shown in Fig. 3.7.

Because azimuthal angles are in $[-\pi, \pi]$, Fourier transforms act on datasets defined on complete circular boundaries around the listener. Equation (3.19) first decomposes the dataset using an integral over θ_0 and then reconstructs it at a target angle θ. Note that these two variables are independent.

The synthesis model in spherical coordinates presented in this sections can properly predict the distance decay due to point sources, as opposed to the method in cylindrical coordinates. Moreover, the model in spherical coordinates yields continuous results along the interaural axis because it does not require splitting the plane, as opposed to the method in interaural coordinates.

3.4.2 Analysis of distance variation in transformed domains

The angular part of the models in cylindrical, interaural and spherical coordinates are used in this section to analyze the distance variation in circular HRTF datasets.

3.4.2.1 Fourier-Legendre domain

To investigate the relative variation over distance of circular HRTF datasets, we calculate the ratio between Fourier-Legendre transforms. We analyzed the datasets in Figs. 1.2, 1.3, and 1.4 to obtain the corresponding illustrative results shown in Figs. 3.8, 3.10, and 3.12. General remarks are as follows.

i) In Figs. 3.8(a), 3.8(b), 3.10(a), 3.10(b), 3.12(a), and 3.12(b), it can be appreciated that most of the energy in the Fourier-Legendre transform of the datasets is concentrated in the lower lateral-angle modes (α-modes). The transition between α-modes of high and low energy increases as frequency increases.

ii) In the distance variation patterns shown in Figs. 3.8(c), 3.10(c), and 3.12(c), it is observed that the distance variation patterns corresponding to high-energy α-modes are relatively simple as opposed to the ones corresponding to low-energy modes.

iii) The nature of the Fourier-Legendre transform in (3.13) implies that frontal and rear regions must be analyzed separately because $\alpha_0 \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. These analyses are respectively shown in the top and bottom panels of Figs. 3.8, 3.10, and 3.12. Furthermore, distance variations follow different patterns in each of these two regions.

3.4.2.2 Fourier domain

Similarly to the examples in section 3.4.2.1, those shown in Figs. 3.9, 3.11, and 3.13 are based on the ratio between the Fourier transforms of the datasets shown in Figs. 1.2, 1.3, and 1.4. General remarks are as follows.

i) In Figs. 3.9(a), 3.9(b), 3.11(a), 3.11(b), 3.13(a), and 3.13(b), it can be observed that most of the energy in the Fourier transform of the datasets is concentrated in the lower azimuthal-angle modes (θ-modes). The transition between θ-modes of high and low energy increases as frequency increases. The decay of energy along the transition occurs faster than in the Fourier-Legendre domain, as can be observed when comparing with the corresponding examples shown in Figs. 3.8(a), 3.8(b), 3.10(a), 3.10(b), 3.12(a), and 3.12(b).

ii) In the distance variation patterns shown in Figs. 3.9(c), 3.11(c), and 3.13(c), which were calculated based on the ratio between the Fourier transforms illustrated in Figs. 3.9(a), 3.9(b), 3.11(a), 3.11(b), 3.13(a), 3.13(b), it can be appreciated that the distance variation patterns corresponding to high-energy θ-modes of the datasets are relatively simple, in contrast with the ones corresponding to low-energy modes.

iii) Because Fourier transforms are applied on full circular boundaries at once, the distance variations between two circular datasets are described by a single set of patterns.

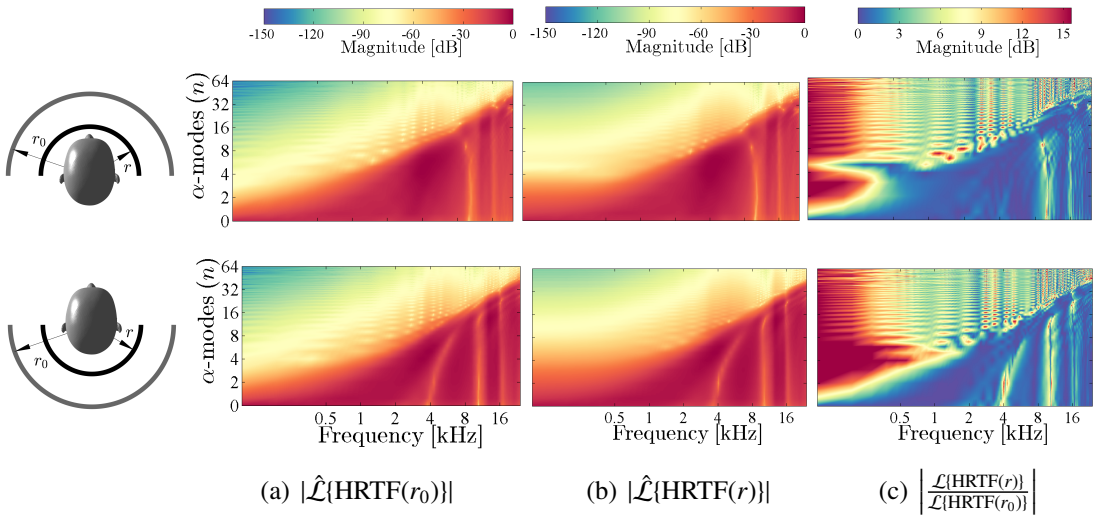


Figure 3.8: Analysis of distance variation between the datasets shown in Figs. 1.2(c) and 1.2(d), from $r_0 = 150$ cm to $r = 25$ cm, using the Fourier-Legendre transform $\mathcal{L}$ defined in (3.13). Frequencies and lateral-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier-Legendre transforms $\hat{\mathcal{L}}$, while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier-Legendre transforms $\mathcal{L}$ at r and r_0.

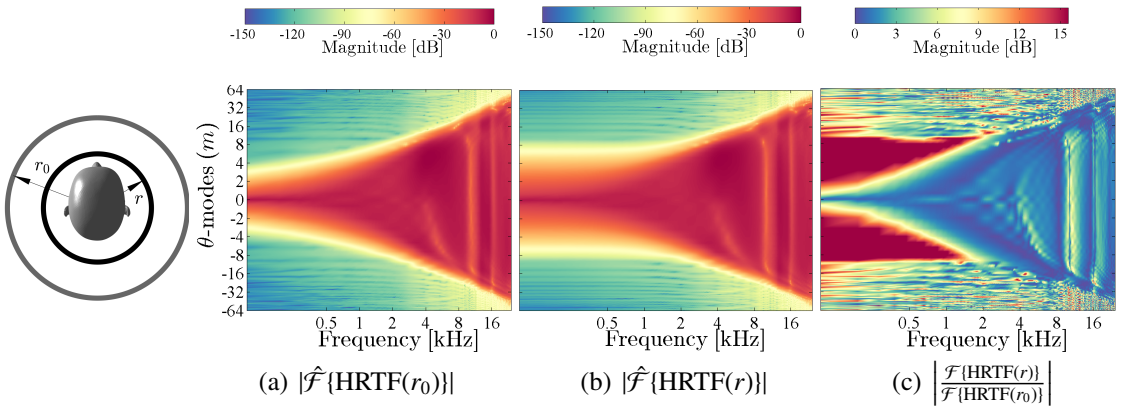
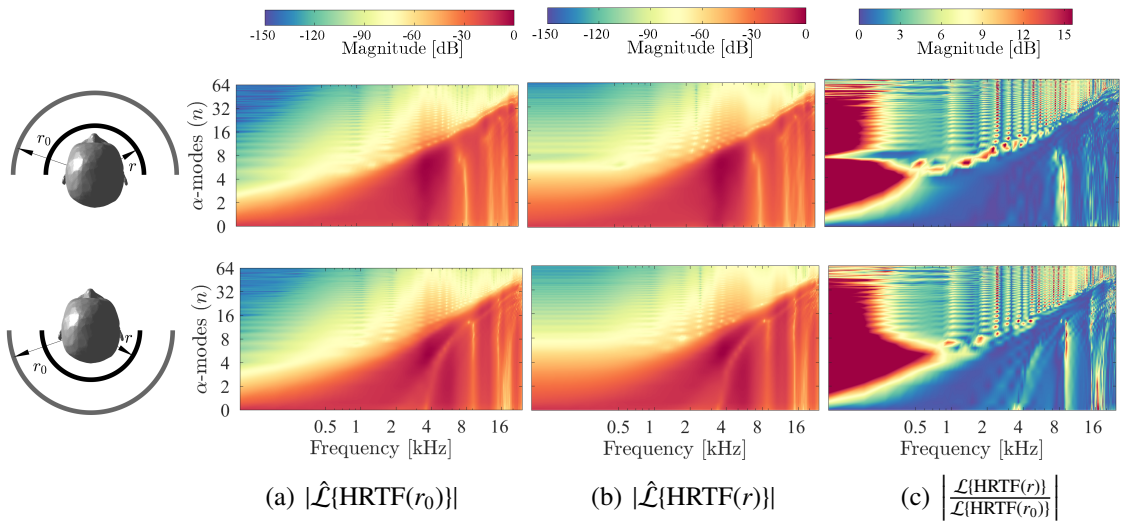
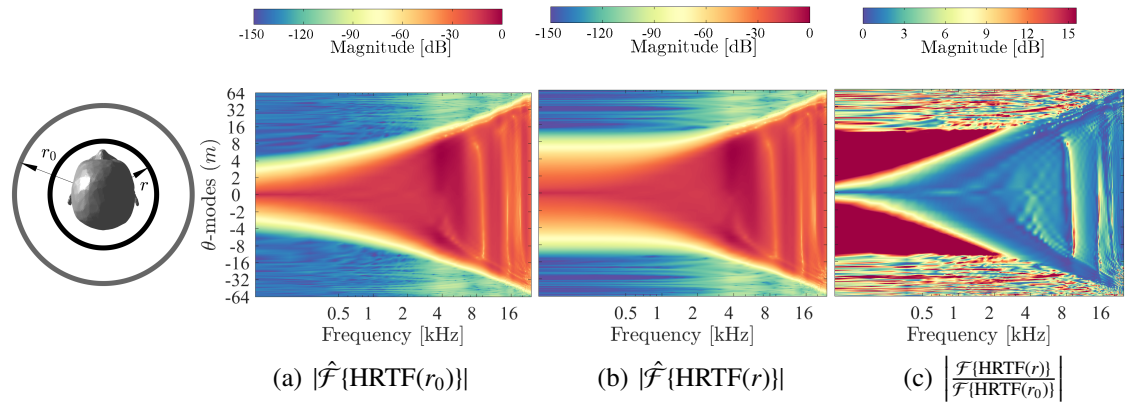


Figure 3.9: Analysis of distance variation between the datasets shown in Figs. 1.2(c) and 1.2(d), from $r_0 = 150$ cm to $r = 25$ cm, using the Fourier transform $\mathcal{F}$ defined in (3.19). Frequencies and azimuthal-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier transforms $\hat{\mathcal{F}}$, while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier transforms $\mathcal{F}$ at r and r_0.



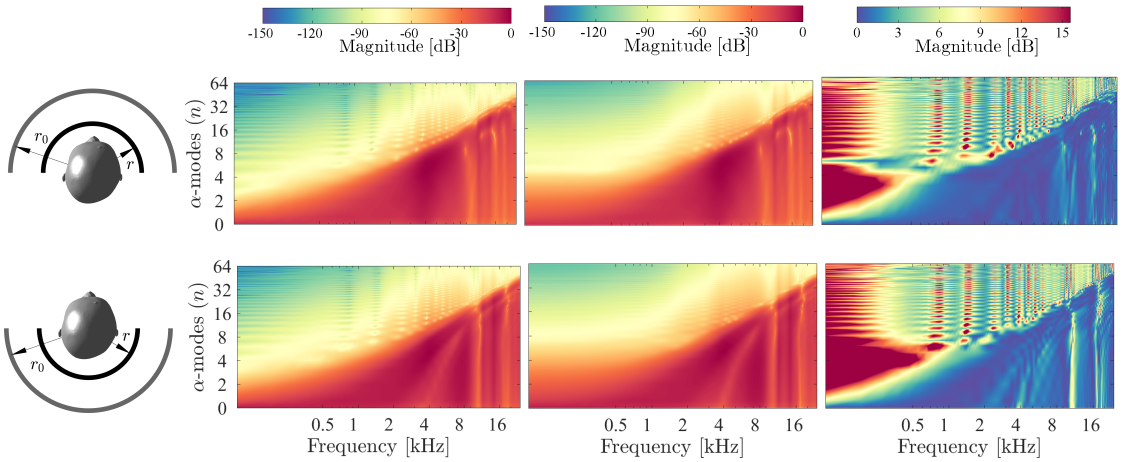
(a) $|\hat{\mathcal{L}}\{\text{HRTF}(r_0)\}|$ (b) $|\hat{\mathcal{L}}\{\text{HRTF}(r)\}|$ (c) $\left|\frac{\mathcal{L}\{\text{HRTF}(r)\}}{\mathcal{L}\{\text{HRTF}(r_0)\}}\right|$

Figure 3.10: Individual A. Analysis of distance variation between the datasets shown in Figs. 1.3(c) and 1.3(d), from $r_0 = 150$ cm to $r = 25$ cm, using the Fourier-Legendre transform $\mathcal{L}$ defined in (3.13). Frequencies and lateral-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier-Legendre transforms $\hat{\mathcal{L}}$, while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier-Legendre transforms $\mathcal{L}$ at r and r_0.



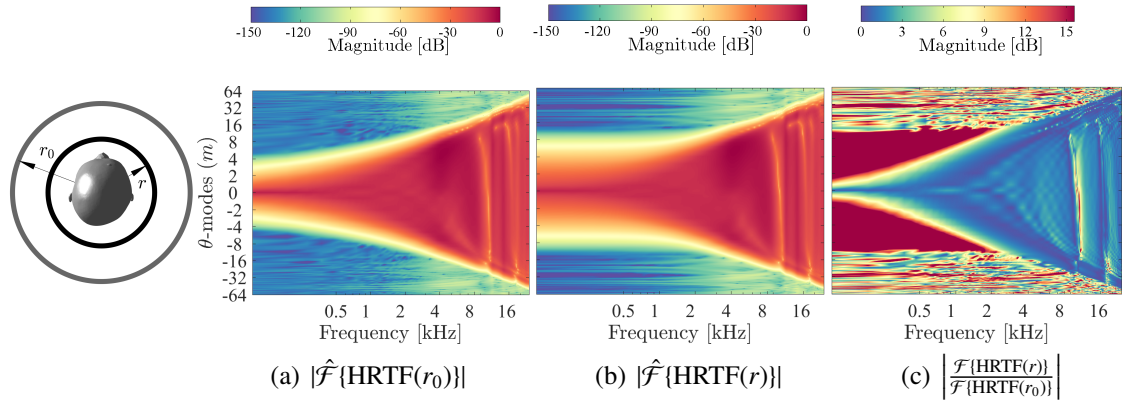
(a) $|\hat{\mathcal{F}}\{\text{HRTF}(r_0)\}|$ (b) $|\hat{\mathcal{F}}\{\text{HRTF}(r)\}|$ (c) $\left|\frac{\mathcal{F}\{\text{HRTF}(r)\}}{\mathcal{F}\{\text{HRTF}(r_0)\}}\right|$

Figure 3.11: Individual A. Analysis of distance variation between the datasets shown in Figs. 1.3(c) and 1.3(d), from $r_0 = 150$ cm to $r = 25$ cm, using the Fourier transform $\mathcal{F}$ defined in (3.19). Frequencies and azimuthal-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier transforms $\hat{\mathcal{F}}$, while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier transforms $\mathcal{F}$ at r and r_0.



(a) $|\hat{\mathcal{L}}\{\text{HRTF}(r_0)\}|$ (b) $|\hat{\mathcal{L}}\{\text{HRTF}(r)\}|$ (c) $\left|\frac{\mathcal{L}\{\text{HRTF}(r)\}}{\mathcal{L}\{\text{HRTF}(r_0)\}}\right|$

Figure 3.12: Individual B. Analysis of distance variation between the datasets shown in Figs. 1.4(c) and 1.4(d), from $r_0 = 150$ cm to $r = 25$ cm, using the Fourier-Legendre transform $\mathcal{L}$ defined in (3.13). Frequencies and lateral-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier-Legendre transforms $\hat{\mathcal{L}}$, while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier-Legendre transforms $\mathcal{L}$ at r and r_0.



(a) $|\hat{\mathcal{F}}\{\text{HRTF}(r_0)\}|$ (b) $|\hat{\mathcal{F}}\{\text{HRTF}(r)\}|$ (c) $\left|\frac{\mathcal{F}\{\text{HRTF}(r)\}}{\mathcal{F}\{\text{HRTF}(r_0)\}}\right|$

Figure 3.13: Individual B. Analysis of distance variation between the datasets shown in Figs. 1.4(c) and 1.4(d), from $r_0 = 150$ cm to $r = 25$ cm, using the Fourier transform $\mathcal{F}$ defined in (3.19). Frequencies and azimuthal-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier transforms $\hat{\mathcal{F}}$, while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier transforms $\mathcal{F}$ at r and r_0.

3.4.3 Limitation of the bandwidth of angular spectra

Another issue addressed in this investigation deals with angular resolution in practical filter implementation. Continuous, circular distributions of sources are necessarily considered when deriving distance-varying filters from analytic solutions to the acoustic wave equation. Yet HRTF datasets are obtained in practice for discrete distributions of sources. Relating discrete and continuous distributions requires additional assumptions. In particular, following the Nyquist sampling theorem on the circle [89], we consider that continuous distributions have finite angular bandwidths. Implementations therefore require the establishment of thresholds to restrict the action of distance-varying filters to the corresponding angular bandwidth. We examine an existing threshold [95] to limit the angular bandwidth according to frequency and find that this traditional threshold does not ensure accurate results, especially at low frequencies and positions close to the head. To cope with this problem, we propose a new threshold to limit the angular bandwidth in accordance with the magnitude decay of spherical wavefronts along distance in free-field.

In this section, we address an issue regarding practical implementation of distance-varying filters. Continuous distributions of sources were considered in sections 3.4.1.1, 3.4.1.2 and 3.4.1.3 when formulating distance-varying filters. Yet HRTF datasets are obtained for discrete distributions of finite angular resolution and, hence, representations of datasets on transform domains are accurate in a limited angular bandwidth. Implementations thus require thresholds to restrict the action of distance-varying filters to a limited angular bandwidth.

Figures 3.8, 3.9, 3.10, 3.11, 3.12, 3.13 show that most of the energy in angular spectra of datasets is concentrated in low angular modes up to a frequency-dependent high-to-low energy transition. Small values taken at higher modes would, in practice, lead to a low signal-to-noise ratio that can negatively impact the results; this is in agreement with the ill-conditioned nature of the acoustic propagation problem [83, pp. 291-293]. Furthermore, the behavior of these low-energy components at different distances follows a complex pattern that may be difficult to capture in a general model. For these reasons, we focus our

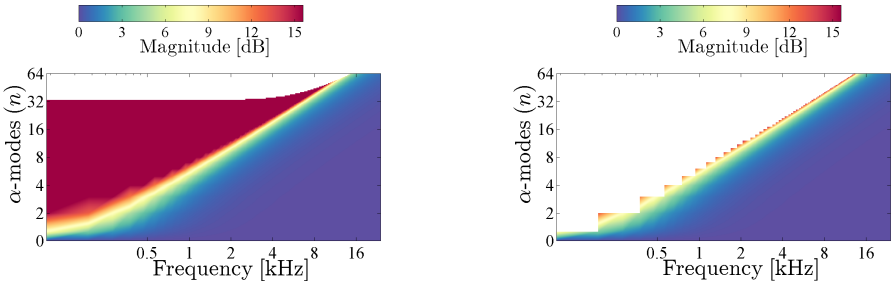
analysis on limiting the angular bandwidth so as to preserve the high-energy modes only.

Limited angular bandwidths are modeled by truncating the sums along modal indices n in (3.13) and m in (3.19) to a finite number of terms. In [95], a frequency-dependent threshold to decide bounds for modal indices was deduced by using the asymptotic expansion of spherical Hankel functions for large orders. In general, bounds L for $n \leq \lfloor L \rfloor$ or $|m| \leq \lfloor L \rfloor$, beyond which the distance-varying filters are set to zero, can be defined as follows [95]:

$$L = \left\{ \left[\frac{1}{\ln\left(\frac{r}{r_\mathrm{h}}\right)} \ln\left(\frac{\left(\frac{r}{r_\mathrm{h}}\right)^{\frac{3}{2}}}{\left(\frac{r}{r_\mathrm{h}} - 1\right)^{\frac{3}{2}} \epsilon} \right) + 1 \right]^4 + \left[kr_\mathrm{h} + \frac{1}{2} \left(\ln \frac{1}{\epsilon^3} + \ln \left(kr_\mathrm{h}\right)^{\frac{1}{2}} \right)^{\frac{2}{3}} \left(kr_\mathrm{h}\right)^{\frac{1}{3}} \right]^4 \right\}^{\frac{1}{4}}. \quad (3.20)$$

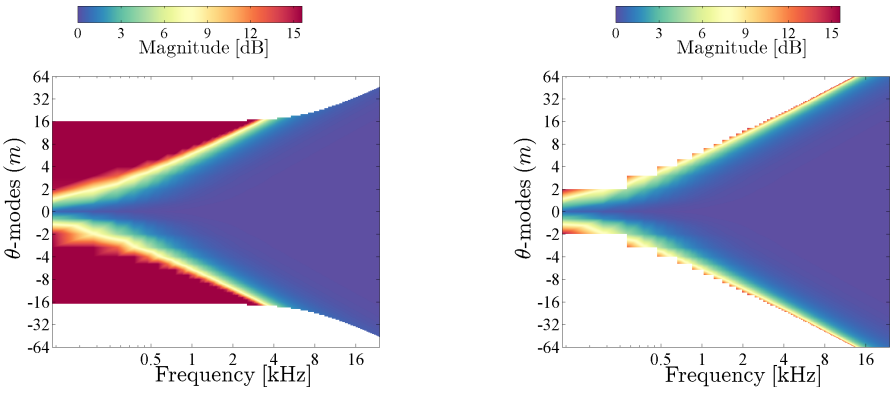
Here, ϵ is a specified approximation error to represent spatial information inside the smallest radius r_h containing the head, due to a source at a distance r. Examples of filters designed using this threshold, for $\epsilon = 10^{-5}$, $r_\mathrm{h} = 15$ cm and $r = 25$ cm, are shown in Figs. 3.14(a) and 3.15(a). Filters designed with this threshold attempt to fully approximate distance-varying patterns, such as the ones shown in Figs. 3.8(c), 3.9(c), 3.10(c), 3.11(c), 3.12(c), 3.13(c) by further including information on spectral regions comprising high modes at low frequencies. Nevertheless, resulting filters have the risk of overemphasizing components of low signal-to-noise ratio, as angular spectra of initial datasets at far distances contain low-energy components in these spectral regions.

To cope with this problem, we first observe that the distance-varying filters in (3.9), (3.13) and (3.19) have magnitude responses that are monotonically increasing functions of relative distance, and of n or m. Based on this observation, we propose to restrict the angular bandwidths of distance-varying filters by clipping their magnitude responses according to the ratio between the surface area of two hypothetical concentric spheres (or cylinders) at r_0 and r. In addition, to generalize the distance-varying filters to work in both the far and near fields, we scale the filters by a factor $\left(\frac{r}{r_0}\right)^{\frac{1}{2}}$ in cylindrical coordinates, or by a factor $\frac{r}{r_0}$ in interaural and spherical coordinates. Such scalings are equivalent to the use of Hankel functions normalized by the magnitude of their far-field asymptotic forms [92] or by the magnitude of the near-field radiation of a point source [110]. We formulate a magnitude-



(a) Frequency-dependent threshold. (b) Magnitude-dependent threshold.

Figure 3.14: Distance-varying filters in interaural coordinates $\hat{D}_n^{\text{int}}(r_0, r)$ from $r_0 = 150$ cm to $r = 25$ cm, whose directional bandwidths have been limited according to the frequency-dependent threshold in (3.20) and proposed magnitude-dependent threshold in (3.21). Colorbars indicate magnitude gains.



(a) Frequency-dependent threshold. (b) Magnitude-dependent threshold.

Figure 3.15: Proposed distance-varying filters in spherical coordinates $\hat{D}_m^{\text{sph}}(r_0, r)$ from $r_0 = 150$ cm to $r = 25$ cm, whose directional bandwidths have been limited according to the frequency-dependent threshold in (3.20) and proposed magnitude-dependent threshold in (3.21). Colorbars indicate magnitude gains.

dependent band-limiting threshold (MBT) according to the following expression:

$$\hat{D}(r, r_0, k) = \begin{cases} s \cdot D(r, r_0, k) & \text{if } |D(r, r_0, k)| \leq s^{-2}, \\ 0 & \text{else,} \end{cases} \tag{3.21}$$

were

$$s = \begin{cases} \left(\frac{r}{r_0}\right)^{\frac{1}{2}} & \text{cylindrical coordinates,} \\ \frac{r}{r_0} & \text{interaural or spherical coordinates.} \end{cases} \tag{3.22}$$

In (3.21), $\hat{D}$ is a distance-varying filter with limited angular bandwidth and D is one of the filters defined in (3.9), (3.13) or (3.19). Examples of filters designed with this threshold, for the methods in interaural coordinates and spherical coordinates, are respectively shown in Figs. 3.14(b) and 3.15(b), where it is observed that the action of the filters is properly restricted to the spectral regions of high-energy components. Magnitude gains at the edges correspond to thresholds $\frac{r_0^2}{r^2}$ in (3.21), while staircase patterns are due to the integer nature of n and m. Note that no prior knowledge on the target HRTFs is required when using (3.20) or (3.21).

3.4.4 Numerical experiments

In this section, we numerically compare the performances of the method in the interaural coordinates defined in (3.13) and the proposal in the spherical coordinates defined in (3.19). Evaluations were performed by comparing synthesized datasets denoted by $\hat{\mathcal{T}}$ and target datasets denoted by $\mathcal{T}$.

Overall accuracy along frequency was calculated by a logarithmic spectral distance that has been shown to be suitable for predicting audible differences between measured and synthesized HRTFs [111]. We refer to this distance as the spectral distortion (SD) in

decibels and write it as follows [111]:

$$\mathrm{SD}(\theta) = \left[\frac{1}{f_2 - f_1} \int_{f_1}^{f_2} \left[20 \log_{10} \left| \frac{\hat{\mathcal{T}}(\theta, f)}{\mathcal{T}(\theta, f)} \right| \right]^2 df \right]^{\frac{1}{2}}. \tag{3.23}$$

Overall accuracy along angles, described for convenience by azimuthal angle θ, was calculated based on circular correlations (CC) that provide a good measure for the similarities of the directional patterns of $\hat{\mathcal{T}}$ and $\mathcal{T}$, independently of a possible gain mismatch [93]. Normalized CC is defined as follows [93]:

$$\mathrm{CC}(f) = \frac{\int_{-\pi}^{\pi} \hat{\mathcal{T}}(\theta, f) \overline{\mathcal{T}(\theta, f)} d\theta}{\left[\int_{-\pi}^{\pi} \left| \hat{\mathcal{T}}(\theta, f) \right|^2 d\theta \times \int_{-\pi}^{\pi} \left| \mathcal{T}(\theta, f) \right|^2 d\theta \right]^{\frac{1}{2}}}, \tag{3.24}$$

with the overbar denoting the complex conjugate.

Using the head models described in Figs. 1.2(a), 1.3(a) and 1.4(a), and the boundary element method (BEM) in [43], we calculated an initial HRTF dataset at a radius r_0 = 150 cm and target datasets at radii r ranging from 15 to 149 cm in regular spacing intervals of 1 cm. Every dataset was calculated for 360 sound sources equiangularly distributed on the horizontal plane. Such a high angular resolution allowed us to focus the study on the distance effects. We considered frequency bins ranging from f_1 = 93.75 Hz to f_2 = 19, 875 Hz, with a sampling frequency of 48 kHz. Because the head model is symmetric, only the left ear was examined.

3.4.4.1 Synthesis of artificial-head HRTFs

To illustrate the local performance first, we present examples of HRTF datasets synthesized in interaural and spherical coordinates in Figs. 3.16 and 3.17, respectively. Discontinuous results along $\theta = \pm 90^\circ$ can be observed in Figs. 3.16(a) and 3.16(b). This is because Fourier-Legendre transforms are applied along two hemi-circumferences separated by the interaural axis and, unfortunately, different values are obtained when approaching this axis. Such a drawback is not present in the results obtained with the proposal in spherical coor-

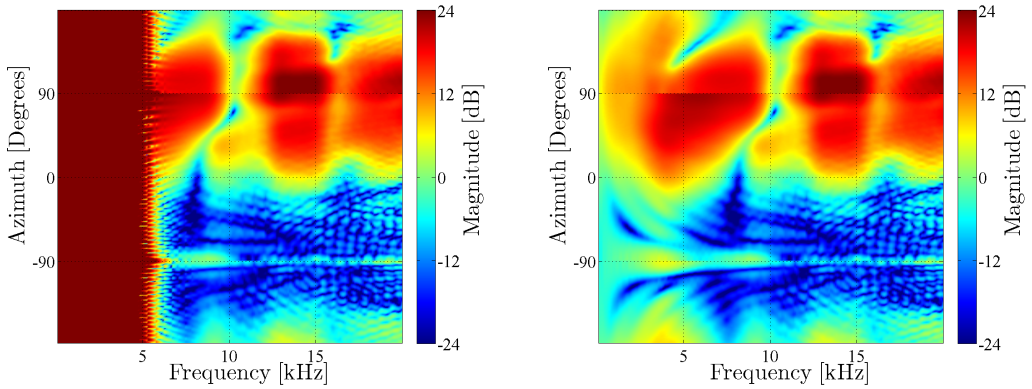
dinates, as can be observed in Figs. 3.17(a) and 3.17(b), as Fourier transforms are applied along all azimuth angles at once.

Overall accuracies obtained with the method in interaural and spherical coordinates, calculated with (3.23) and (3.24), are displayed in Fig. 3.18. It can be observed that the filters designed with the frequency-dependent threshold yield poor overall accuracies at low frequencies and close distances. These limitations are overcome when using filters designed with the proposed magnitude-dependent threshold.

Resulting overall accuracies of the proposal in spherical coordinates, calculated with (3.23) and (3.24), are displayed in Fig. 3.19. Notwithstanding the poor overall accuracies at low frequencies and close distances yielded by the frequency-dependent threshold, the overall performance in spherical coordinates under this condition outperforms the method in interaural coordinates. Performance is further improved when using filters designed with the proposed magnitude-dependent threshold nonetheless. In this case, our proposal performed better than the method in interaural coordinates.

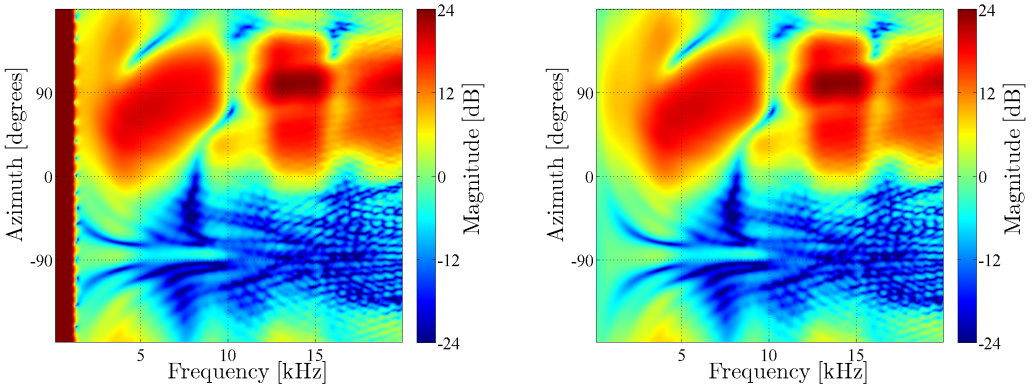
In Fig. 3.19(b), high SD values are still observed for distance variations around $\theta = -90°$. These values would ultimately not have corresponded to significant errors in the synthesized HRTFs, as they result from low levels in the initial and target datasets for the contralateral side. In the same panel, low CC values appear around 10 kHz. The reason behind these inaccuracies is that the initial HRTFs have an energy content that is relatively low in this frequency region, as can be observed in Fig. 1.2(c). The distance-varying filters rely on a simple model that struggles to handle the head-model features in this frequency region, such as the displacement of salient features towards the interaural axis. In general, the human head model used in this study can be considered to be valid up to an average frequency of 16.6 kHz only. In effect, according to the spatial sampling criterion of four elements per wavelength λ [43], the average cell length $\frac{\lambda}{4} = 5.1$ mm limited the evaluations up to an average frequency of $\frac{c}{\lambda} = 16.6$ kHz. Measures of overall accuracy, therefore, provide reliable information for frequencies below this limit only.

Finally, Fig. 3.20 and 3.21 present synthesis examples from lower angular resolutions.



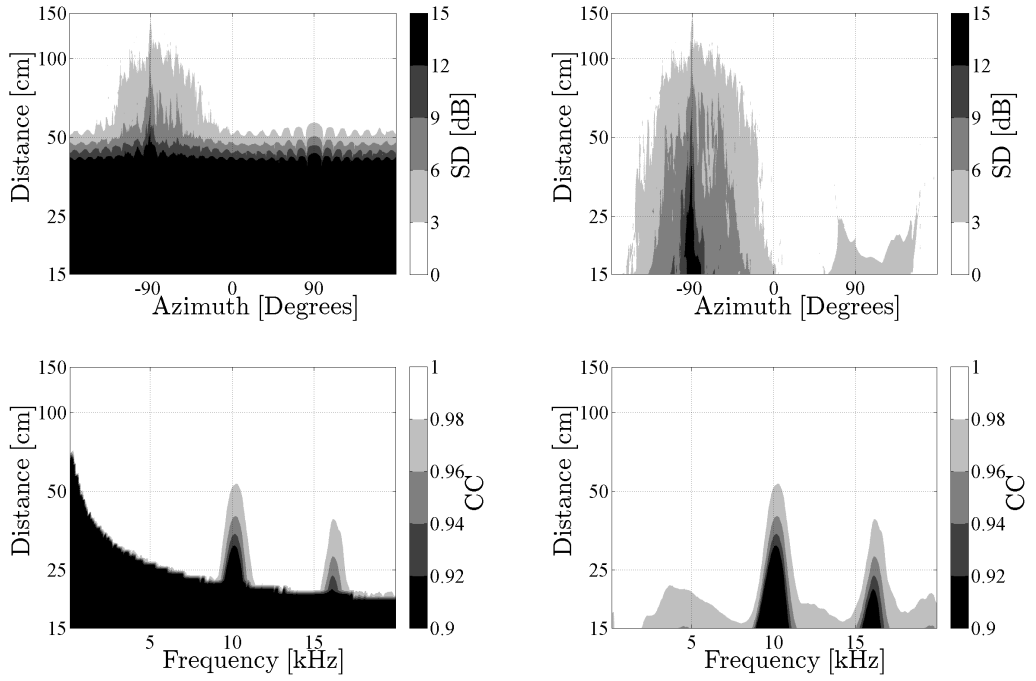
(a) Frequency-dependent threshold. (b) Magnitude-dependent threshold.

Figure 3.16: HRTF datasets at $r = 25$ cm synthesized with the distance varying filters in interaural coordinates. Directional bandwidths were limited according to the frequency-dependent threshold in (3.20) and the proposed magnitude-dependent threshold in (3.21).



(a) Frequency-dependent threshold. (b) Magnitude-dependent threshold.

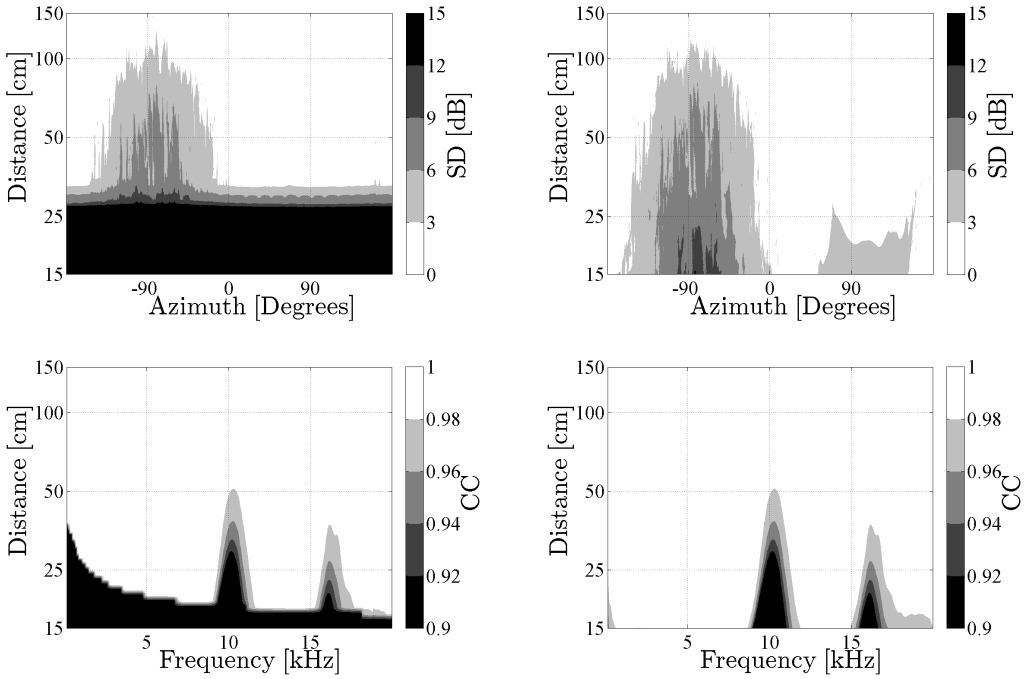
Figure 3.17: HRTF datasets at $r = 25$ cm synthesized with the proposed distance varying filters in spherical coordinates. Directional bandwidths were limited according to the frequency-dependent threshold in (3.20) and the proposed magnitude-dependent threshold in (3.21).



(a) Frequency-dependent threshold.

(b) Magnitude-dependent threshold.

Figure 3.18: Overall accuracies achieved with the distance-varying filters in interaural coordinates. The angular bandwidths of filters were limited according to the frequency-dependent threshold in (3.20) and proposed magnitude-dependent threshold in (3.21). Overall accuracies were calculated using the spectral distortion (SD) in (3.23) and normalized circular correlation (CC) in (3.24).



(a) Frequency-dependent threshold.

(b) Magnitude-dependent threshold.

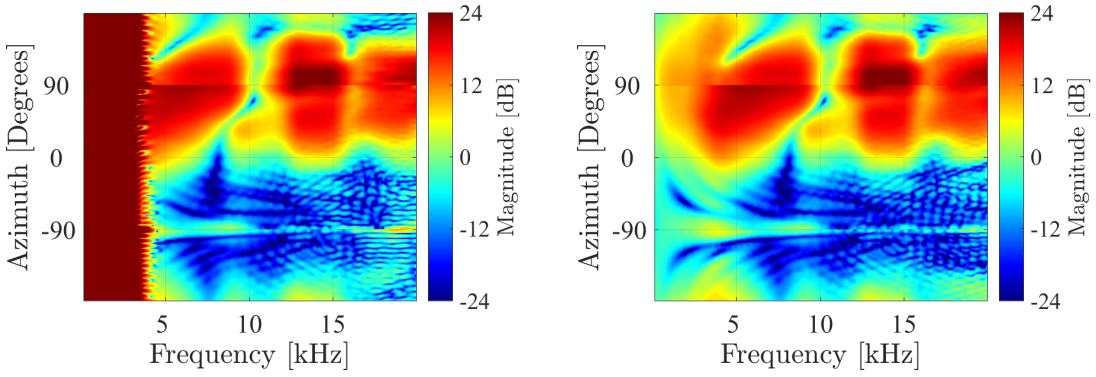
Figure 3.19: Overall accuracies achieved with the proposed distance-varying filters in spherical coordinates. The angular bandwidths of filters were limited according to the frequency-dependent threshold in (3.20) and proposed magnitude-dependent threshold in (3.21). Overall accuracies were calculated using the spectral distortion (SD) in (3.23) and normalized circular correlation (CC) in (3.24).

Two initial HRTF datasets at a radius $r_0 = 150$, for angular resolutions of 3° (120 sources) and 10° (36 sources), were considered in these examples. A resolution of 3° allows for modal representations up to $n = |m| = \frac{120}{2} = 60$, which correspond to an angular bandwidth still comprising the required initial information for an acceptable synthesis, as can be verified in Figs. 3.8(a) and 3.9(a). As a consequence, the results shown in Fig. 3.20 only differ slightly from those shown in Figs. 3.16 and 3.17. A resolution of 10°, on the other hand, allows for modal representations up to $n = |m| = \frac{36}{2} = 18$, and correspondingly an angular bandwidth comprising initial information up to around 8 kHz, as can also be verified in Figs. 3.8(a) and 3.9(a). Results shown in Fig. 3.21 are therefore expected to be accurate only up to around 8 kHz. Interestingly, it is verified therein that synthesis in spherical coordinates using the MBT outperformed the other methods.

3.4.4.2 Synthesis of individual HRTFs

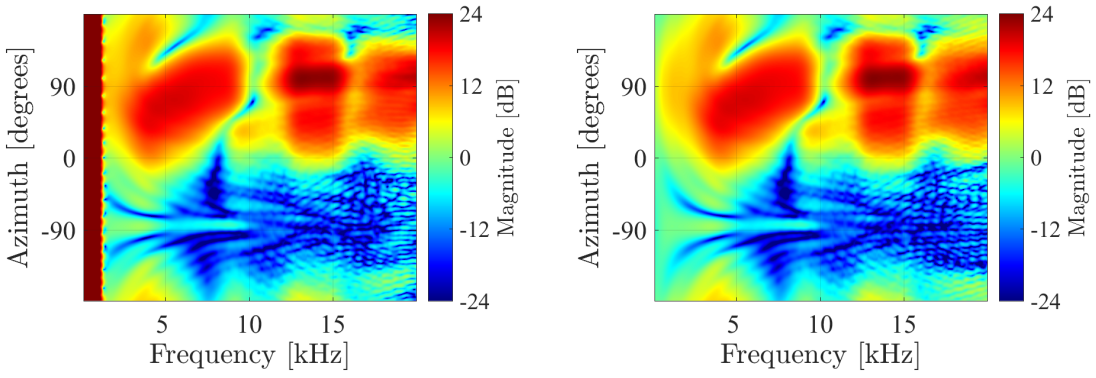
Synthesis examples considering HRTFs obtained for real individual listeners are presented in this section. To illustrate the local performance, examples of HRTF datasets synthesized in interaural and spherical coordinates are shown in Fig. 3.22 and 3.23. This time, the proposed magnitude-dependent band-limiting threshold was used in both synthesis methods. Discontinuous results along $\theta = \pm 90°$ can be observed in Figs. 3.22(a) and 3.23(a) as a result of using the method in interaural coordinates. Such a drawback is not present in the results obtained with the proposal in spherical coordinates, as can be observed in Figs. 3.22(b) and 3.23(b).

Overall accuracies obtained with the method in interaural coordinates, calculated with (3.23) and (3.24), are displayed in Figs. 3.24 and 3.25. For both individual models, high SD values are still observed in the top panels for distance variations around $\theta = -90°$. These values result from low levels in the initial and target datasets for the contralateral side. In the bottom panels, low CC values appear around 10 kHz. The reason behind these inaccuracies is that the initial HRTFs have an energy content that is relatively low in this frequency region, as can be observed correspondingly in Figs. 1.3(c) and 1.4(c).



(a) Int. coord., FBT.

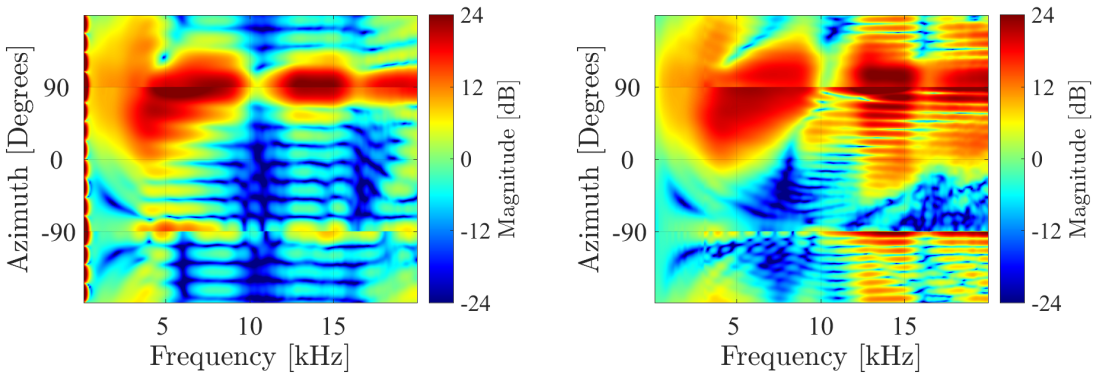
(b) Int. coord., MBT.



(c) Sph. coord., FBT.

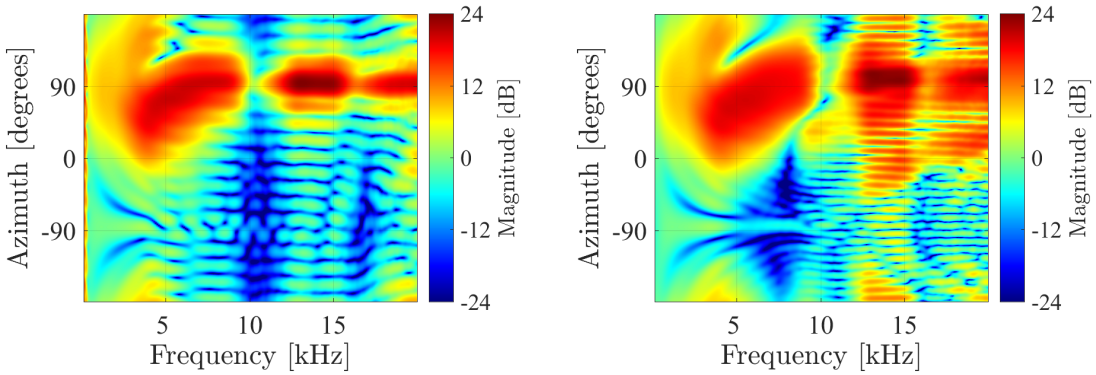
(d) Sph. coord., MBT.

Figure 3.20: HRTF datasets at $r = 25$ cm synthesized from initial datasets at $r_0 = 150$ cm with initial angular resolution of 3°. Synthesis was performed using the methods in interaural and spherical coordinates, with the frequency-dependent band-limiting threshold (FBT) in (3.20) and the proposed magnitude-dependent band-limiting threshold (MBT) in (3.21).



(a) Int. coord., FBT.

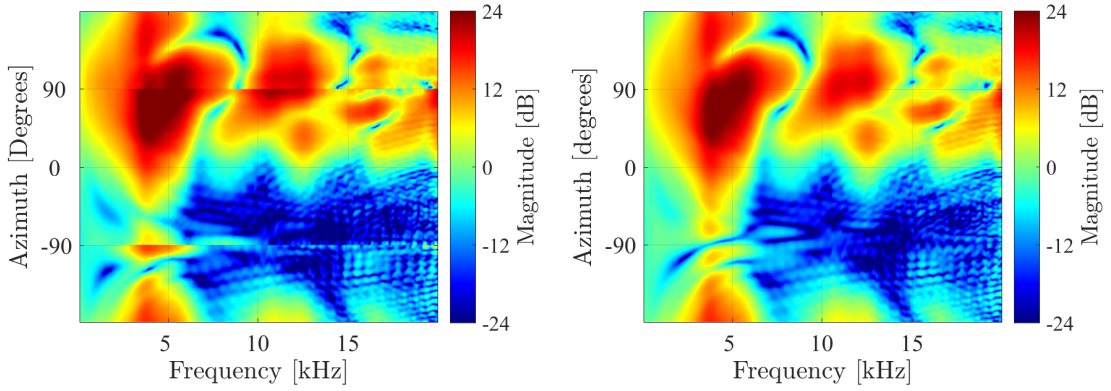
(b) Int. coord., MBT.



(c) Sph. coord., FBT.

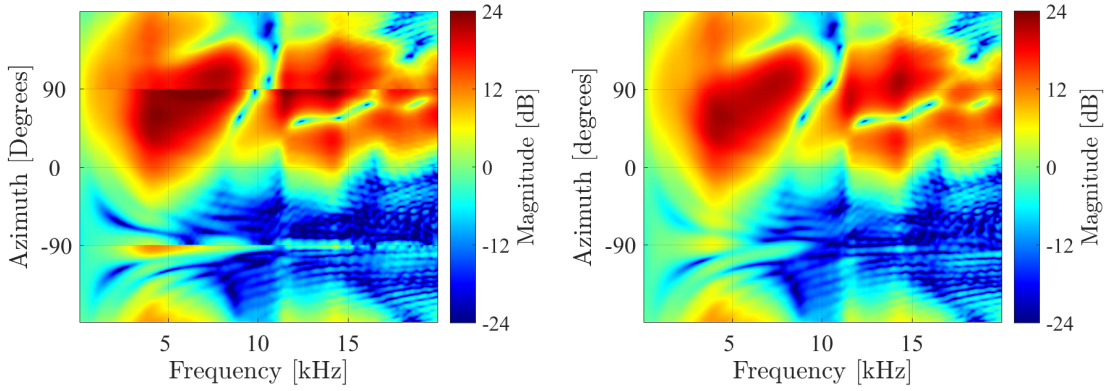
(d) Sph. coord., MBT.

Figure 3.21: HRTF datasets at $r = 25$ cm synthesized from initial datasets at $r_0 = 150$ cm with initial angular resolution of 10°. Synthesis was performed using the methods in interaural and spherical coordinates, with the frequency-dependent band-limiting threshold (FBT) in (3.20) and the proposed magnitude-dependent band-limiting threshold (MBT) in (3.21).



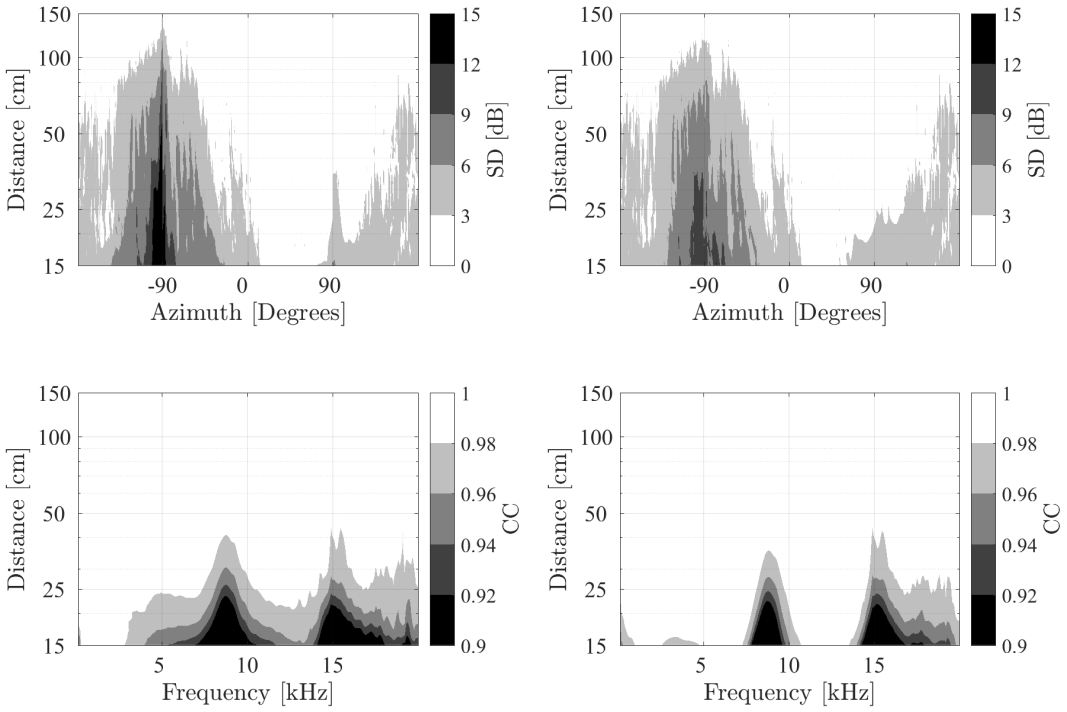
(a) Int. coord., MBT. (b) Sph. coord., MBT.

Figure 3.22: HRTF datasets for the left ear of individual A in Fig 1.3. Datasets were synthesized at $r = 25$ cm with the distance varying filters in interaural and spherical coordinates. Angular bandwidths were limited by using the proposed magnitude-dependent band-limiting threshold (MBT) in (3.21).



(a) Int. coord., MBT. (b) Sph. coord., MBT.

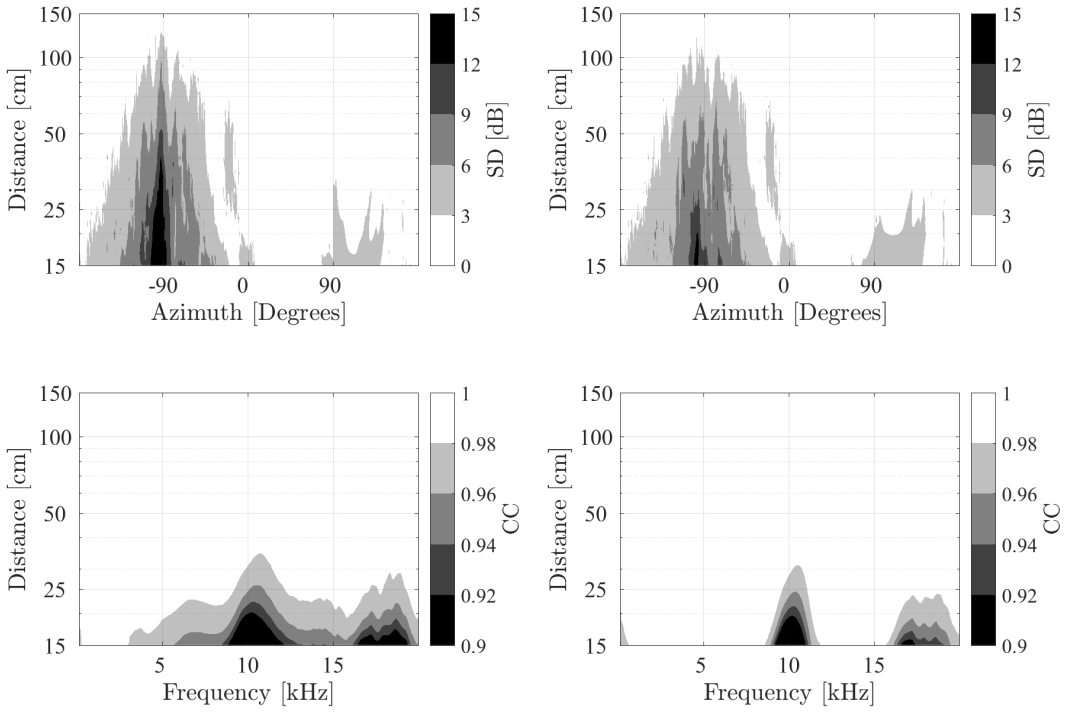
Figure 3.23: HRTF datasets for the left ear of individual B in Fig 1.4. Datasets were synthesized at $r = 25$ cm with the distance varying filters in interaural and spherical coordinates. Angular bandwidths were limited by using the proposed magnitude-dependent band-limiting threshold (MBT) in (3.21).



(a) Int. coord., MBT.

(b) Sph. coord., MBT.

Figure 3.24: Overall accuracies for the left-ear HRTFs of individual A in Fig 1.3. Datasets were synthesized with the distance-varying filters in interaural and spherical coordinates. The angular bandwidths of filters were limited by using the proposed magnitude-dependent threshold in (3.21). Overall accuracies were calculated using the spectral distortion (SD) in (3.23) and normalized circular correlation (CC) in (3.24).



(a) Int. coord., MBT.

(b) Sph. coord., MBT.

Figure 3.25: Overall accuracies for the left-ear HRTFs of individual B in Fig 1.4. Datasets were synthesized with the distance-varying filters in interaural and spherical coordinates. The angular bandwidths of filters were limited by using the proposed magnitude-dependent threshold in (3.21). Overall accuracies were calculated using the spectral distortion (SD) in (3.23) and normalized circular correlation (CC) in (3.24).

3.5 Summary

We have herein presented and evaluated distance-varying filters to synthesize HRTFs at arbitrary distances from a dataset of HRTFs at a single distance around the listener.

When synthesis was limited to the horizontal plane, a new set of distance-varying filters was derived from the solution of the acoustic wave equation for sound fields that are assumed to be invariant with respect to elevation angles in spherical coordinates. Results were free of the lateral side discontinuities that appear when sound fields are assumed to be invariant along polar angles in interaural coordinates. Moreover, to properly account for HRTF datasets obtained for discrete distributions that limit the angular bandwidth, we introduced a new magnitude-dependent band-limiting threshold (MBT) to restrict the action of distance-varying filters to limited angular bandwidths.

We finally evaluated the performance numerically using a model of a human head valid up to 16.6 kHz. Results show that overall accuracies obtained with the proposed distance-varying filters in spherical coordinates outperform the existing distance-varying filters in interaural coordinates. Furthermore, overall accuracies obtained with the proposed MBT outperformed overall accuracies achieved with the traditional frequency-dependent threshold, especially at low frequencies and distances close to the head.

A perceptual evaluation of the distance-varying filters by means of detectability of differences, and localization tests along azimuth and distances, could provide more insight into the validity of the suggested approach.

Chapter 4

Sound field recording using an acoustically rigid boundary

4.1 Overview

This chapter addresses the problem of estimating the sound pressure signals outside a spherical microphone array mounted on the surface of an acoustically rigid spherical baffle. The pressure signals outside the array are estimated from the microphone array signals. For this purpose, a continuous-space model is first formulated for estimating the sound pressure field outside a rigid sphere, once the measurements of sound pressure on its whole surface are available. The presence of the spherical baffle constitutes an acoustically rigid boundary; it provides higher stability to the estimation as opposed to the use of open spherical arrays that does not use a baffle [47]. The continuous-space model is used to derive a discrete-space model by angular-band limitation and sampling. The resulting discrete-space model constitutes a theoretical basis for estimating the sound field outside the spherical microphone array.

Estimating the free-field pressure signals outside the spherical microphone array, as if

its rigid baffle was not present, is an important step that is required in many applications such as beamforming [47–58], binaural synthesis [59–66], and room acoustics [67–71]. In general, the estimation comprises two steps: 1) the removal of the acoustic scattering effects introduced by the interaction of sound with the rigid spherical baffle, and 2) the estimation of the sound pressure in free-field at a specific radius outside the array. It is accepted from the literature [47–71] that existing filters involving these two steps require excessive dynamic ranges at lower and middle frequencies. The model introduced in this chapter is particularly intended to enable the design of filters that properly sidestep the drawbacks found in existing filters.

In spatial sound applications, the spherical array recordings are intended to be presented to the listeners in the form of a sound pressure field through a real loudspeaker array, or alternatively in the form of binaural signals through a virtual loudspeaker array that represents a HRTF dataset. This idea is illustrated in Fig. 4.1. In general, stable sound field estimation is required at the reproduction radius, that is, at the radius of the array of real or virtual loudspeakers.

To formulate the continuous-space model for estimating the pressure outside the continuous recording sphere, it will be assumed in general that a continuous and hypothetical reproduction boundary is placed at an arbitrary reproduction radius. The concept of boundary matching filtering will be introduced in this context. Boundary matching filters aim to convert the physical boundary conditions used during recording into the ones required for reproduction. This concept is formulated using a general theoretical framework for recording and reproduction that is based on the Kirchhoff-Helmholtz integral equation.

4.2 The Kirchhoff-Helmholtz integral equation

Sound fields satisfying the linearized acoustic wave equation are fully characterized within a source-free bounded volume by a set of physical boundary conditions, i.e., the values of sound pressure and its normal derivative on a bounding surface. The mathe-

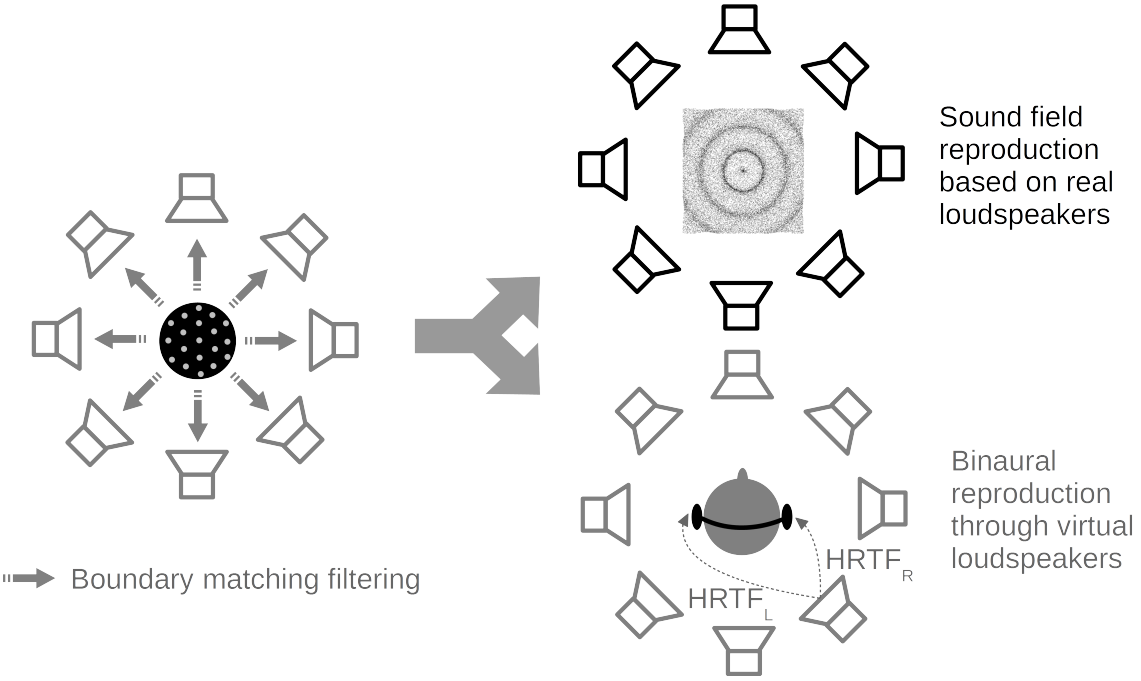


Figure 4.1: Given the pressure signals of a rigid spherical microphone array, the purpose of boundary matching filters is to obtain the pressure signals at a reproduction radius outside the array. These signals can be used for sound field reproduction through a loudspeaker array, or for binaural synthesis based on virtual loudspeakers or HRTF datasets.

matical formulation of such characterization is known as the Kirchhoff-Helmholtz integral equation (KHIE) [83].

The KHIE is used to reconstruct a time-harmonic sound pressure field Ψ that satisfies the acoustic wave equation in (2.1) at any point $\vec{r}$ within a source-free volume $\mathcal{V}$. The sound field is reconstructed from its values of pressure and normal derivative on the bounding surface Γ that delimits $\mathcal{V}$. Reconstruction is performed by means of transfer functions $\mathfrak{g}$ characterizing the transmission of sound in free space from any point $\vec{r}_\gamma$ in Γ to any point $\vec{r}$ in $\mathcal{V}$. The transfer function $\mathfrak{g}$ is known as the Green function in free space. The KHIE is mathematically formulated as follows [83]:

$$\Psi(\vec{r},k) = \int_{\vec{r}_\gamma \in \Gamma} \left[\mathfrak{g}(\vec{r},\vec{r}_\gamma,k) \frac{\partial}{\partial \vec{n}_\gamma} \Psi(\vec{r}_\gamma,k) - \Psi(\vec{r}_\gamma,k) \frac{\partial}{\partial \vec{n}_\gamma} \mathfrak{g}(\vec{r},\vec{r}_\gamma,k) \right] d\vec{r}_\gamma, \tag{4.1}$$

where $\vec{n}_\gamma$ is the normal vector to Γ at $\vec{r}_\gamma$ pointing outwards $\mathcal{V}$, and $\frac{\partial}{\partial \vec{n}_\gamma}$ denotes the derivative

at the point $\vec{r}_\gamma$ in the direction of $\vec{n}_\gamma$.

Recently, the KHIE is enabling a generalized theory for physically motivated sound field techniques that use microphone and loudspeaker arrays, respectively distributed on two bounding surfaces for recording and reproduction (see [72, 84, 85, 87, 88] for a present-day account of this theory). In this theoretical framework, although not always stated explicitly, the problem of stably recording and optimally reproducing a sound field can also be interpreted as finding a way to linearly convert the boundary conditions used during recording into the ones required for reproduction. Such linear conversion can be represented by a set of what is referred to in this study as boundary matching filters (BMFs). Obtaining the BMFs therefore involves the examination of the KHIE in both the recording and reproduction boundaries.

4.3 Boundary matching filters (BMFs) for sound field recording

Boundary matching filters (BMFs) are presented in this section based on the framework provided by the Kirchhoff-Helmholtz integral equation in (4.1). The aim of BMFs is to convert the physical boundary conditions used during recording into the ones required for reproduction. BMFs are then applied to the problem of stably recording a sound field with a spherical microphone array on a rigid boundary, which is important for a wide range of applications including beamforming [47–58], binaural synthesis [59–66], and room acoustics [67–71].

Boundaries of arbitrary geometry fundamentally require the consideration of both pressure and its normal derivative [72, 83–85, 87, 88, 112]. The recording boundary plays a crucial role in this respect. The consideration of only pressure produces non-unique characterizations at frequencies associated with resonances of the recording boundary exclusively[1] [72]. Recording the normal derivative, either using pressure-gradient probes or a

[1] Historically, resonant frequencies has long been identified in the theory of active control of sound [22,

double layer array of microphones, would likely be a limitation in real-world implementations. To enable pressure-based characterizations free of resonances, some physical topologies have been explored for determining the complete boundary conditions from pressure recordings only [72]. Because these topologies will potentially disturb the sound field to be recorded, their invertible characterization is also necessary to reconstruct the free-field or unperturbed sound field. Among such topologies, the use of an acoustically rigid boundary, the so-called Neumann boundary condition [83], is an effective choice because it causes the normal derivative to be zero. A useful boundary condition for reproduction, on the other hand, is the consideration of an open or acoustically transparent boundary, since the matched sound field in this case would exclusively convey the acoustics of the recording environment, providing a characterization whose acoustic propagation during reproduction would only be affected by the reproduction environment. Under these considerations, BMFs aim to compensate for the recording topology so as to enable an acoustically transparent recording system, and they further aim to extrapolate the resulting free-field sound field towards the reproduction boundary.

4.4 Spherical BMFs

Boundaries of spherical geometries are particularly useful in recording scenarios that require to emphasize the vicinity of a single view point with an uniform covering of directions. They conveniently allow for the mathematical formulation of BMFs. In spherical coordinates, sound fields can be expanded in terms of time-harmonic solutions to the acoustic wave equation in (2.2); the number of terms is referred to as the order of the expansion and is a measure of angular resolution. The angular portion of the solutions results in an orthonormal basis of functions that is used to represent or encode the directional components of sound fields with scalable order. This is done by means of the spherical Fourier transforms in (2.14) and (2.15), which in turn facilitates the spatial signal processing. The radial

113], where inverse systems are derived from actual measurements of transfer functions relating the pressure on two boundaries. Such inverse systems can also be interpreted as BMFs.

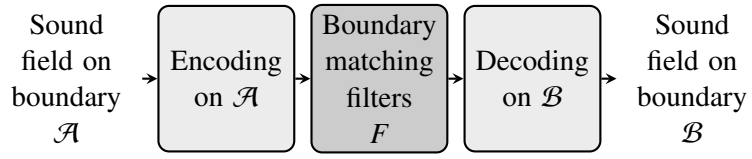


Figure 4.2: General scheme used in model-based sound field techniques that consider a recording boundary $\mathcal{A}$ and a reproduction boundary $\mathcal{B}$.

portion of the solutions is used to derive BMFs that operate on the encoded domain (transform domain). This approach originates the model-based scheme for sound field recording and reproduction shown in Fig. 4.2. It is important to mention that a reference radius for reproduction can be chosen a priori without loss of generality, because the shift to a new radius is possible by means of well-known acoustic propagators [83], as long as the new reproduction boundary is also assumed with no sources in its interior. The present study concerns the formulation of BMFs for stable recording with a rigid spherical boundary.

The coordinate system used throughout this chapter is shown in Fig. 4.3. The geometry used to model sound field recording is shown in Fig. 4.4(a), where the bounding surface $\mathcal{A}$ is composed of recording points $\vec{a} = (a, \Omega_a)$. The geometry used to model sound field reproduction is shown in Fig. 4.4(b), where the bounding surface $\mathcal{B}$ is composed of secondary source points $\vec{b} = (b, \Omega_b)$ and divides the space into an external and an internal volume. All of what follows considers time-harmonic sound pressure fields satisfying the acoustic wave equation in (2.2) with time-harmonic dependence $\exp(-jkct)$, where k denotes the wave number, c the speed of sound in air, and $j = \sqrt{-1}$. The sound pressure fields under consideration also satisfy the Sommerfeld radiation condition at infinity [114].

4.4.1 Recording with a rigid spherical continuous surface

The use of a rigid spherical baffle allow for pressure-based characterizations free of resonances. When examining the KHIE in (4.1) on the rigid boundary, under the consideration of an incident plane wave, or an incident spherical wave, the two classical for the acoustic scattering from a rigid sphere are obtained, respectively for the far field or the near field [83]. These models are illustrated in Fig. 4.5. The angular part of the models results

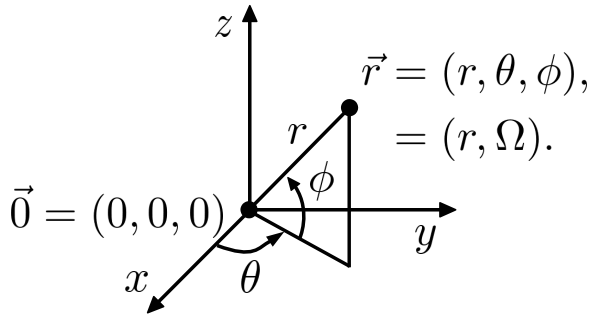
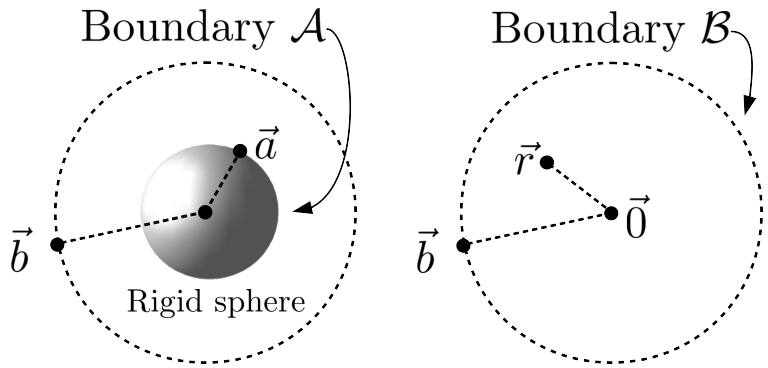


Figure 4.3: Spherical coordinate system. A point in space $\vec{r} = (r, \theta, \phi)$ is specified by its radial distance r, azimuth angle $\theta \in [-\pi, \pi]$ and elevation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Angles are merged into the variable $\Omega = (\theta, \phi)$ in such a way that a point in space is also represented by $\vec{r} = (r, \Omega)$.



(a) Encoding. (b) Decoding.

Figure 4.4: Geometry used to formulate the boundary matching filters. The centers of $\mathcal{A}$ and $\mathcal{B}$ coincide with the origin of the coordinate system in Fig. 4.3.

in a directional encoding scheme. The reciprocal of the radial part directly originates the transform domain filters that are used to compensate for the rigid baffle; the spherical-wave filters further allows for sound field estimation outside the array (sound field extrapolation along the reproduction radius). In the literature, these filters can be recognized as a term inversely proportional to the Green-Neumann functions [63, 83, 115], as the reciprocal of the source mode strengths [47–49, 51, 54, 55, 58–62, 64–71, 115], or included in what is commonly called a radial filter [52, 53, 56, 57].

An important assumption implicit in the derivation of the classical models that underlie plane-wave and spherical-wave filters is that of keeping a constant angular resolution over the full frequency range. Nevertheless, it is accepted that this achieved at the expense of excessive gains at lower frequencies. The gains grow even more with increasing en-

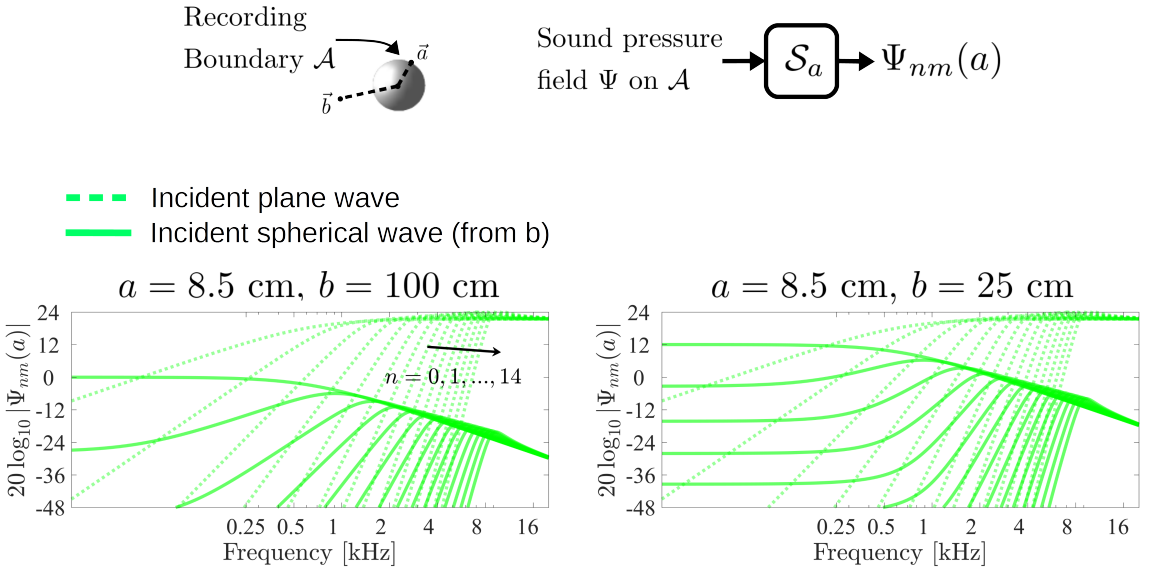


Figure 4.5: Model of pressure over a rigid sphere of radius a due to an incident plane wave ($b \to \infty$) or an incident spherical wave ($b < \infty$). The order n can be interpreted as the spatial resolution of the measurement over the rigid sphere. Existing filters are based on the direct inversion of this model, which leads to high amplitudes for higher orders.

coding order and quickly exceed the dynamic range of existing arrays. Despite the gain of spherical-wave filters decreases as the desired reproduction radius decreases, they still remain high and this behavior systematically extends to the middle frequencies.

Excessive amplifications of classic plane-wave and spherical-wave filters have been addressed by establishing limits to the encoding order. Order limitation is equivalent to restricting the angular bandwidth of spherical harmonic representations and is not an exclusive issue of spherical array encoding. Among the frequency-dependent limits formulated in [94, chapter 9] and [95], two are of particular importance in encoding. The first limit corresponds to the transition between the far and near fields [52, 53, 83], and is proportional to the frequency and reproduction radius. It also sets the theoretical restrictions for sound field extrapolation indicating that, given an reproduction radius, higher frequencies demand higher orders and, therefore, greater dynamic ranges. Current technology has confined most applications to use this limit for encoding on the vicinity of a low-order array. Regarding the future of recording technology, however, high-order arrays are not unrealistic; it has been reported in [74] for instance a setup composed of 252 microphones. The

second limit, although not popular for array encoding, includes a prescribed gain bound to allow for high-order encodings at the expense of decreasing the reproduction radius with increasing frequency. Nevertheless, this limit tends to overemphasize the low-frequencies. In either way, the main shortcomings of filters designed with frequency-dependent limits are the loss of frequency content in the low and middle range, and the discontinuous behavior along frequency due to the integer nature of orders.

The artificial generation of frequency content and its smoothing to sidestep the shortcomings of order limitation have been proposed based on, for example, timbre compensation [116], spherical-cap windowing [75], regularization [56], and band-pass filter banks [57]. The performance of each method depends on tuning parameters according to the targeted application. Of particular interest is the understanding of the effects of regularization, since it is widely used in array recording. There still remain a more general question, however, on whether the results achieved with this kind of methods are in agreement with the underlying physical phenomenon. Because excessive amplification at low frequencies is not in un-encoded sound field recordings, it is likely a result of the sound field encoding model. It is therefore expected that a physics-based refinement of the model will be in consistency with existing evidence that low-frequency content of multipole sources decays systematically with increasing order [117].

This study postulates that the origin of unnatural gains in existing plane-wave and spherical-wave filters might rely on the lack of knowledge of the boundary conditions required for reproduction. In fact, most implementations are mathematically equivalent to considering linear systems that arise from the superposition of the single-source scattering models to describe pressures on the recording boundary due to an array of secondary sources. This consideration results on the widely explored plane-wave decomposition [47–49, 51, 54, 55, 58, 59, 64–68, 70, 71] and the more recently used point-source expansion [60, 61, 63, 115], whose common physical foundations can be traced back to the principle of wave superposition (PWS) [118]. Although an equivalence between the PWS and the KHIE exist [118], no consideration about a reproduction boundary was considered.

Despite the necessity of examining the KHIE in both boundaries being recognized and the mathematical tools exist [72,84,85,87,88], to the best of the authors' knowledge, no formal explanation of the phenomenon of excessive amplifications has been provided in terms of testing a hypothetical reproduction boundary. As a consequence, no analytical solution to this problem has been provided.

This study assumes a hypothetical spherical boundary for reproduction where the KHIE is examined, so as to formulate the corresponding BMFs from a rigid boundary to the hypothetical boundary. This approach allows to identify the reproduction boundary conditions that underlie to plane-wave and spherical-wave filters used in sound field encoding. The plane-wave filters correspond to considering BMFs for a reproduction sphere of infinite radius, since the radiation outwards of sources at infinity is zero according to the Sommerfeld's radiation condition [114]. The spherical-wave filters correspond to considering BMFs for a reproduction sphere of finite radius [2]. It is then proposed the consideration of a spherical boundary for reproduction that has a finite radius and further constitutes an acoustically transparent boundary, that is, an open boundary. This results in a new class of BMFs that provide a smooth analytical solution to the problem of high gains at low frequencies for encoding with low-order arrays. The critic points of the BMFs also define smooth transitions between the near and far fields. For applications requiring high-order encodings, it is proposed a method to establish a magnitude bound without affecting the continuity of BMFs. In general, smooth and bounded results can be ensured using BMFs in the full range of frequencies and distances.

The considerations stated above will be formally treated in a general theoretical framework provided by the KHIE for recording and reproduction in spherical geometries.

[2] In connection with the underlying assumption used in wave field synthesis [84], this kind of reproduction boundary can also be understood as having a rigid spherical shell because only pressure information is available. The evaluation of the KHIE on a rigid spherical shell, however, would produce a different analytic expression.

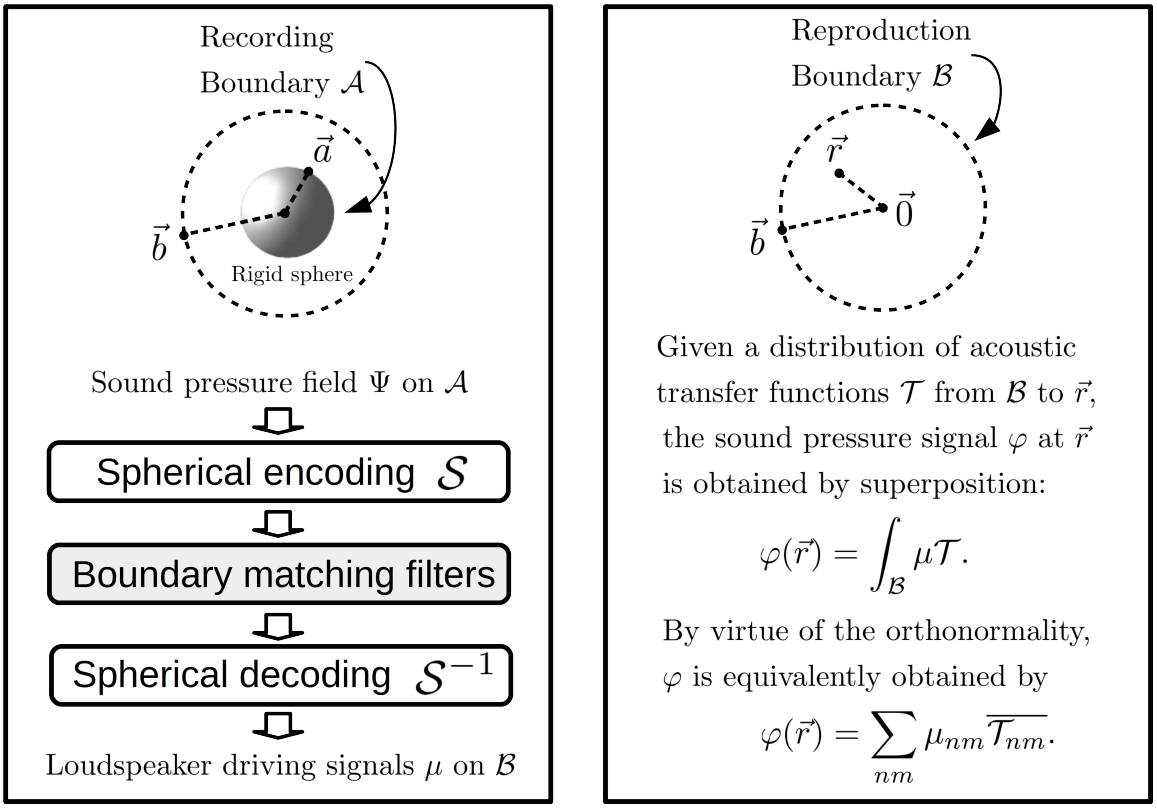


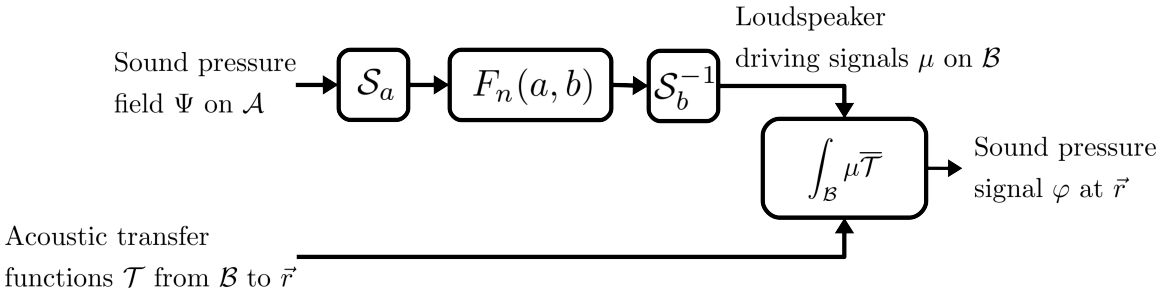
Figure 4.6: Formulation of the model to synthesize the pressure signals φ at an arbitrary point $\vec{r}$ from pressure recordings Ψ. Encoding is based on the spherical Fourier transform denoted by $\mathcal{S}$ and defined in (2.14), while decoding is based on the inverse spherical Fourier transform denoted by $\mathcal{S}^{-1}$ and defined in (2.15).

4.4.1.1 General framework for recording and reproduction

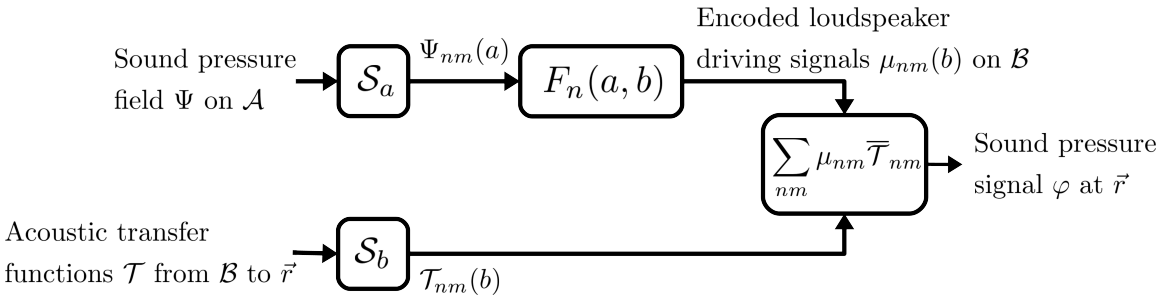
The spherical geometry considered in this study allows for an encoding-decoding scheme based on the spherical Fourier transforms defined in (2.14) and (2.15). The synthesis of sound pressure outside the recording boundary relies on the principle of wave superposition [118], which leads to two dual processing schemes that can be equivalently formulated in the spatial or transform domains, as illustrated in Figs. 4.6 and 4.7.

The sound pressure signal φ at an arbitrary point $\vec{r}$ is calculated based on the diagram shown in Fig. 4.7(b) by using the following linear combination in the transform domain:

$$\varphi(\vec{r}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \overline{\mathcal{T}_{nm}}(\vec{r}, b, k) F_n(a, b, k) \Psi_{nm}(a, k), \tag{4.2}$$



(a) Linear combination in spatial domain.



(b) Linear combination in encoded domain.

Figure 4.7: Dual schemes for the synthesis of pressure signals φ at an arbitrary point $\vec{r}$ from pressure recordings Ψ. Synthesis is based on linear combinations of pressure distributions μ and acoustic transfer functions $\mathcal{T}$. The spherical Fourier transform in (2.14) and its inverse in (2.15) are respectively denoted by $\mathcal{S}_r$ and $\mathcal{S}_r^{-1}$, where r indicates the radius of the corresponding spherical boundary. The boundary matching filters are denoted by F_n.

where the overbar denotes complex conjugate, Ψ_{nm} represents encoded pressure Ψ over $\mathcal{A}$,

$$\Psi_{nm}(a,k) = \int_{\Omega_a \in \mathcal{A}} \Psi(\vec{a},k)\overline{Y_n^m}(\Omega_a)d\Omega_a, \tag{4.3}$$

$\mathcal{T}_{nm}$ represents encoded acoustic transfer functions $\mathcal{T}$ over $\mathcal{B}$,

$$\mathcal{T}_{nm}(\vec{r},b,k) = \int_{\Omega_b \in \mathcal{B}} \mathcal{T}(\vec{r},\vec{b},k)\overline{Y_n^m}(\Omega_b)d\Omega_b, \tag{4.4}$$

and F_n represent encode-domain BMFs that convert Ψ_{nm} on $\mathcal{A}$ onto an encoded distribution of pressure μ_{nm} on $\mathcal{B}$. In practice, μ on the spatial domain would correspond, for example, to the driving signals for an array of loudspeakers distributed on $\mathcal{B}$, while $\mathcal{T}$ can represent a dataset of room impulse responses or head-related impulse responses.

4.4.1.2 Reproduction with infinite and finite spherical boundaries

Discussion on BMFs is limited hereafter to the case when $\mathcal{A}$ is rigid spherical boundary. As for the reproduction boundary $\mathcal{B}$, two conditions are considered. A spherical reproduction boundary of infinite radius corresponds to the classic far-field or plane-wave model

$$F_n^{\mathrm{RI}}(a,b,k) = \frac{(ka)^2 h_n'^{\mathrm{out}}(ka)}{4\pi j^{n+1}}, \text{ Rigid } \mathcal{A} \text{ to Infinite } \mathcal{B}, \tag{4.5}$$

whereas a reproduction boundary of finite radius corresponds to the classic near-field or spherical-wave model

$$F_n^{\mathrm{RF}}(a,b,k) = \frac{-ka^2 h_n'^{\mathrm{out}}(ka)}{4\pi h_n^{\mathrm{out}}(kb)}, \text{ Rigid } \mathcal{A} \text{ to Finite } \mathcal{B}. \tag{4.6}$$

These transform-domain BMFs compensate for the presence of the rigid sphere and extrapolates the resulting free-field representation from $\mathcal{A}$ to $\mathcal{B}$ in such a way that

$$\mathring{\Psi}_{nm}^{\mathrm{ff}}(b,k) = F_n(a,b,k)\Psi_{nm}(a,k). \tag{4.7}$$

BMFs in (4.5) and (4.6) are recognized in the literature as a term inversely proportional to the Green-Neumann functions [63, 83, 115], or as the reciprocal of the source mode strengths [47–51, 54, 55, 58–62, 64–71, 115]. BMFs are also often included in what is commonly called a radial filter [52, 53, 56, 57].

The reproduction conditions are verified by examination of the KHIE on $\mathcal{B}$. BMFs in (4.5) are equivalent to considering a boundary $\mathcal{B}$ of infinite radius, because the radiation outwards of sources at infinity is zero in virtue of the Sommerfeld's radiation condition at infinity [114]. As for the BMFs in (4.6), they are equivalent to using a spherical reproduction boundary of finite radius. On $\mathcal{B}$, the radial derivative of a free-field sound pressure $\frac{\partial}{\partial r}\Psi^{\mathrm{ff}}$ on $\mathcal{B}$ (when the rigid sphere is absent) can also be described by its encodings

$\mathring{\Psi}^{\text{ff}}_{nm}(b,k)$. In general, examining the KHIE on $\mathcal{B}$ reveals that

$$\mu_{nm} = \mathring{\Psi}^{\text{ff}}_{nm}(b,k), \tag{4.8}$$

where F_n represents one of the BMFs defined in (4.5) or (4.6).

4.4.1.3 Reproduction with an open spherical boundary

A useful boundary condition for reproduction is the consideration of an open or acoustically transparent boundary, since the matched sound field in this case would exclusively convey the acoustics of the recording environment, providing a characterization whose acoustic propagation during reproduction would only be affected by the reproduction environment. This condition can be modeled by using the simple source formulation of the KHIE, which considers an interior and an exterior problem separately [119]. The aim here is to obtain the BMFs that relate Ψ on a rigid $\mathcal{A}$ and μ on an open $\mathcal{B}$. For simplicity, the derivation is done in the encoded domain.

An acoustic measurement φ at an arbitrary point $\vec{r}$ outside or inside $\mathcal{B}$ is given by [119]

$$\varphi(\vec{r},k) = \int_{\Omega_b \in \mathcal{B}} \mu(\vec{b},k)\mathfrak{g}(\vec{r},\vec{b},k)d\Omega_b. \tag{4.9}$$

Here, μ is the density of the monopole or simple source distribution, and $\mathfrak{g}$ is the Green's function in free space modeling the transmission of sound from $\vec{b}$ to $\vec{r}$. The density is defined in terms of the radial derivatives of the sound field radiated outside and inside $\mathcal{B}$, in such a way that [119]

$$\mu(\vec{b},k) = \frac{\partial}{\partial r}\varphi^{\text{ext}}(\vec{b},k) - \frac{\partial}{\partial r}\varphi^{\text{int}}(\vec{b},k), \tag{4.10}$$

where superscripts indicate that the radial derivatives on $\mathcal{B}$ need to be computed by considering the external and internal volumes, respectively.

We calculate μ based on its spherical harmonic expansion

$$\mu(\vec{b},k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \mu_{nm}(b,k) Y_n^m(\Omega_b), \tag{4.11}$$

where the encoding μ_{nm} are respectively associated with the terms in (4.10) in such a way that

$$\mu_{nm}(b,k) = \mathring{\varphi}_{nm}^{\text{ext}}(b,k) - \mathring{\varphi}_{nm}^{\text{int}}(b,k). \tag{4.12}$$

Let φ^{ext} and φ^{int} be two general sound pressure fields on the vicinity of $\mathcal{B}$, respectively due to the radiation of μ inwards and outwards. Their representations in terms of time-harmonic solutions to the acoustic wave equation at an arbitrary point $\vec{r}$ are defined as follows:

$$\begin{aligned}
\varphi^{\text{ext}}(\vec{r},k) &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \underbrace{\alpha_{nm}^{\text{out}}(k) h_n^{\text{out}}(kr)}_{\varphi_{nm}^{\text{ext}}(r,k)} Y_n^m(\Omega), \\
\varphi^{\text{int}}(\vec{r},k) &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \underbrace{\alpha_{nm}^{\text{in}}(k) h_n^{\text{in}}(kr)}_{\varphi_{nm}^{\text{int}}(r,k)} Y_n^m(\Omega),
\end{aligned} \tag{4.13}$$

and their corresponding radial derivatives result in

$$\begin{aligned}
\frac{\partial}{\partial r}\varphi^{\text{ext}}(\vec{r},k) &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \underbrace{\alpha_{nm}^{\text{out}}(k) k h_n'^{\text{out}}(kr)}_{\mathring{\varphi}_{nm}^{\text{ext}}(r,k)} Y_n^m(\Omega), \\
\frac{\partial}{\partial r}\varphi^{\text{int}}(\vec{r},k) &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \underbrace{\alpha_{nm}^{\text{in}}(k) k h_n'^{\text{in}}(kr)}_{\mathring{\varphi}_{nm}^{\text{int}}(r,k)} Y_n^m(\Omega).
\end{aligned} \tag{4.14}$$

Because α_{nm}^{out} and α_{nm}^{in} remain invariant along radial distance, the relation between the harmonic representations of pressures an their radial derivatives is given by the following expression:

$$\begin{aligned}
\mathring{\varphi}_{nm}^{\text{ext}}(r,k) &= \frac{k h_n'^{\text{out}}(kr)}{h_n^{\text{out}}(kr)} \varphi_{nm}^{\text{ext}}(r,k), \\
\mathring{\varphi}_{nm}^{\text{int}}(r,k) &= \frac{k h_n'^{\text{in}}(kr)}{h_n^{\text{in}}(kr)} \varphi_{nm}^{\text{int}}(r,k).
\end{aligned} \tag{4.15}$$

To relate the results in (4.15) with the extrapolated sound field $\mathring{\Psi}^{\mathrm{ff}}_{nm}$ defined in (4.8), for the case of an open or acoustically transparent $\mathcal{B}$, the following continuity condition on $\mathcal{B}$ is assumed:

$$\begin{aligned}\varphi^{\mathrm{ext}}_{nm}(b,k) = \varphi^{\mathrm{int}}_{nm}(b,k) &= \frac{1}{2}\Psi^{\mathrm{ff}}_{nm}(b,k) \\ &= \frac{1}{2}\frac{h^{\mathrm{in}}_n(kb)}{kh'^{\mathrm{in}}_n(kb)}\mathring{\Psi}^{\mathrm{ff}}_{nm}(b,k).\end{aligned} \tag{4.16}$$

Combining (4.16) with (4.15) leads to

$$\begin{aligned}\mathring{\varphi}^{\mathrm{ext}}_{nm}(b,k) &= \frac{1}{2}\frac{h'^{\mathrm{out}}_n(kb)}{h^{\mathrm{out}}_n(kb)}\frac{h^{\mathrm{in}}_n(kb)}{h'^{\mathrm{in}}_n(kb)}\mathring{\Psi}^{\mathrm{ff}}_{nm}(b,k), \\ \mathring{\varphi}^{\mathrm{int}}_{nm}(b,k) &= \frac{1}{2}\mathring{\Psi}^{\mathrm{ff}}_{nm}(b,k).\end{aligned} \tag{4.17}$$

Substituting (4.17) in (4.12), and using the Wronskian

$$h'^{\mathrm{out}}_n(kb)h^{\mathrm{in}}_n(kb) - h^{\mathrm{out}}_n(kb)h'^{\mathrm{in}}_n(kb) = -\frac{2j}{(kb)^2}, \tag{4.18}$$

results in

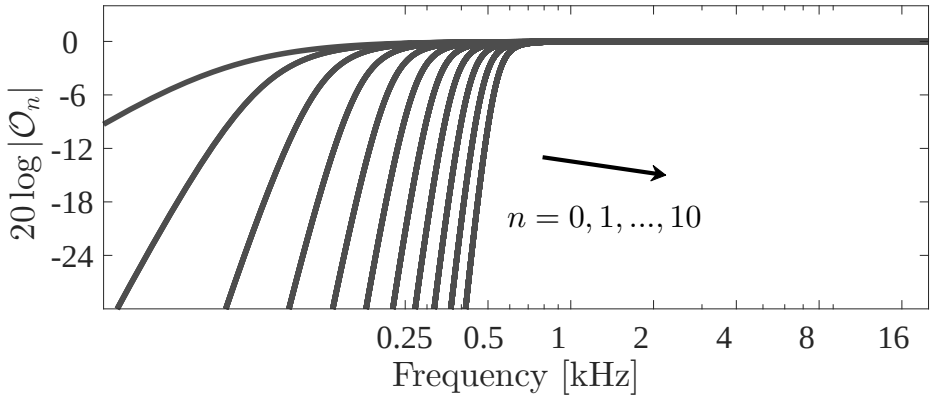
$$\mu_{nm}(b,k) = O_n(b,k)\mathring{\Psi}^{\mathrm{ff}}_{nm}(b,k), \tag{4.19}$$

where O_n are the open-boundary filters in the encoded domain defined by the following expression:

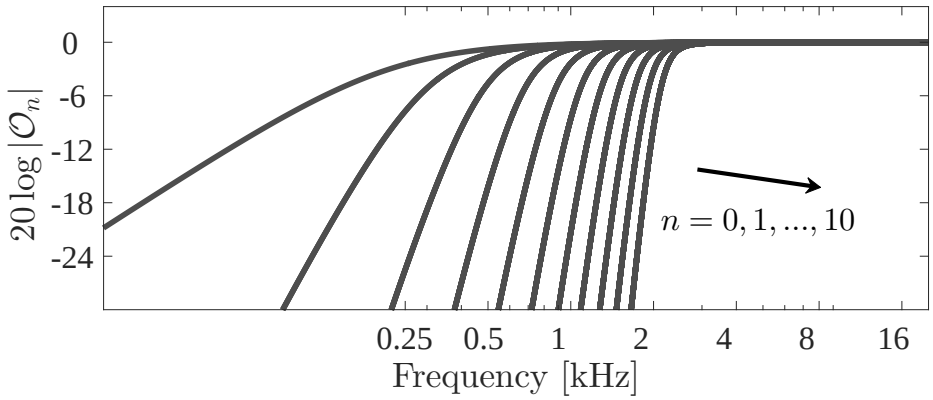
$$O_n(b,k) = \frac{j}{(kb)^2 h^{\mathrm{in}}_n(kb)h'^{\mathrm{out}}_n(kb)}. \tag{4.20}$$

Examples of O_n are illustrated in Fig. 4.8. It is observed that propagation across an open boundary, in which radiation outwards and inwards is considered, has the effect of attenuating the high-order components at low frequency. This is in agreement with empirical evidence that low-frequency content of multipole sources decays systematically with increasing order [117].

Finally, substituting (4.19) in (4.8), results in a new set of BMFs that relates a rigid



(a) b = 100 cm.



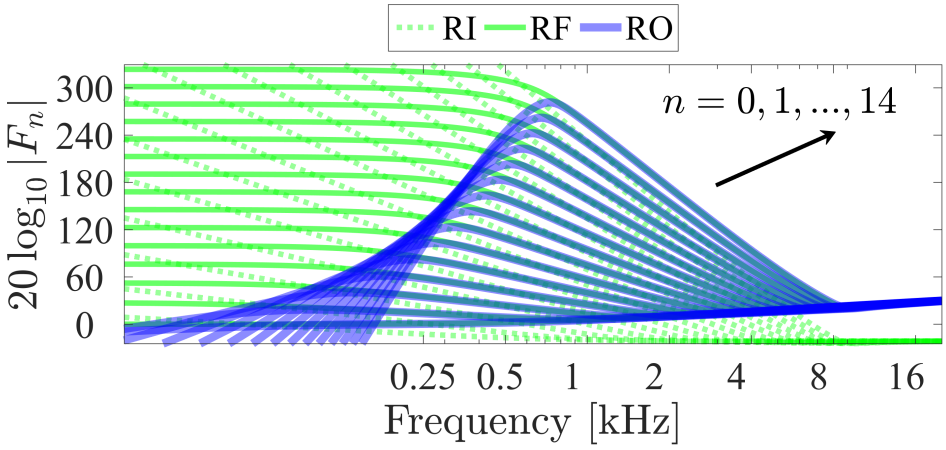
(b) b = 25 cm.

Figure 4.8: Open boundary filters in (4.20) for different orders n.

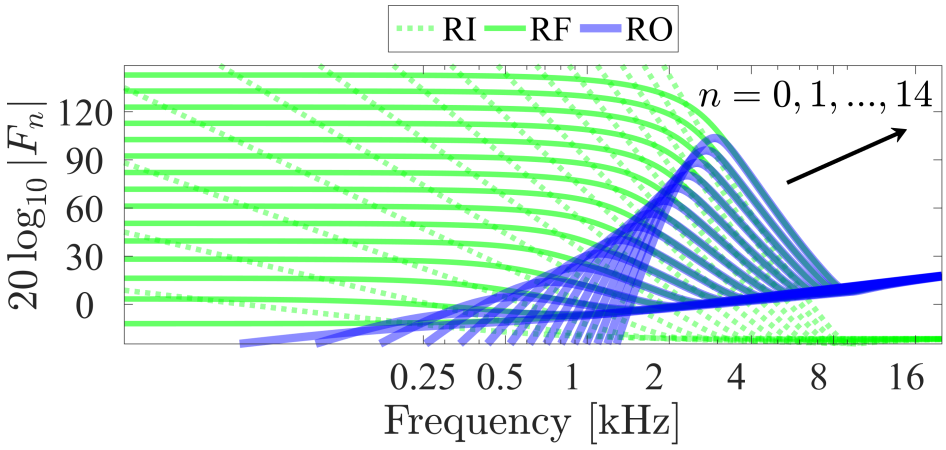
boundary $\mathcal{A}$ to an open boundary $\mathcal{B}$:

$$\begin{aligned} F_n^{\mathrm{RO}}(a,b,k) &= \mathcal{O}_n(b,k) F_n^{\mathrm{RF}}(a,b,k) \\ &= \frac{-ja^2 h_n'^{\mathrm{out}}(ka)}{4\pi k b^2 |h_n(kb)|^2 h_n'^{\mathrm{out}}(kb)}, \text{ rigid } \mathcal{A} \text{ open } \mathcal{B}. \end{aligned} \tag{4.21}$$

For large values of n, such that $n \approx n+1$, local maxima of $|F_n|$ occurs when $n \approx kb$, while local minima occurs when $n \approx ka$. This can be verified by using the first derivative test. Note also that no assumption on $\mathcal{A}$ was made when deriving $\mathcal{O}_n$ and, therefore, these filters can be applied for any other boundary condition on $\mathcal{A}$. Examples of the BMFs in (4.5), (4.6) and (4.21) for different orders n are shown in Fig. 4.9.



(a) a = 8.5 cm, b = 100 cm.



(b) a = 8.5 cm, b = 25 cm.

Figure 4.9: Boundary matching filters F_n^{RI} in (4.5), F_n^{RF} in (4.6) and F_n^{RO} in (4.21) for different orders n.

4.4.2 Limitation of the bandwidth of angular spectra

An angular spectra of limited bandwidth is modeled by truncating the sum (4.2) along n up to a maximum order N.

4.4.2.1 Frequency-dependent band-limited RI- and RF-BMFs

To cope with excessive gains in existing plane-wave and spherical-wave filters, the transition between the far and near fields is used to assign lower maximum orders to lower frequencies. This is equivalent to using the following frequency-dependent band-limited

BMFs:

$$\hat{F}_n = \begin{cases} F_n, & n < \lfloor N \rfloor, \\ 0, & \text{else}, \end{cases} \tag{4.22}$$

where F_n is one of the filters in (4.5) or (4.6).

Keeping a constant reproduction radius b is equivalent to using the following frequency-dependent limit [94]:

$$N = kb. \tag{4.23}$$

This theoretical limit, however, states that high orders and therefore high gains are required to cover large reproduction radii and high frequencies.

To keep the gains bounded, a constant truncation error ε can be introduced, which results in the following limit [95]:

$$N = \left\{ \left[\frac{1}{\ln\left(\frac{b}{a}\right)} \ln\left(\frac{\left(\frac{b}{a}\right)^{\frac{3}{2}}}{\left(\frac{b}{a} - 1\right)^{\frac{3}{2}} \varepsilon} \right) + 1 \right]^4 + \left[ka + \frac{1}{2} \left(\ln \frac{1}{\varepsilon^3} + \ln (ka)^{\frac{1}{2}} \right)^{\frac{2}{3}} (ka)^{\frac{1}{3}} \right]^4 \right\}^{\frac{1}{4}}. \tag{4.24}$$

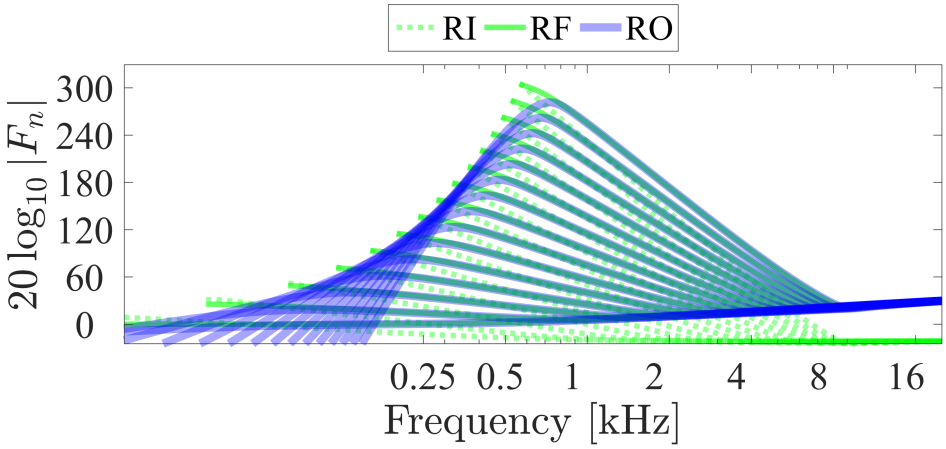
Examples of BMFs designed with (4.24) are shown in Fig. 4.10, where it is observed that this limit tends to overemphasize the low frequencies.

4.4.2.2 Regularized RI- and RF-BMFs

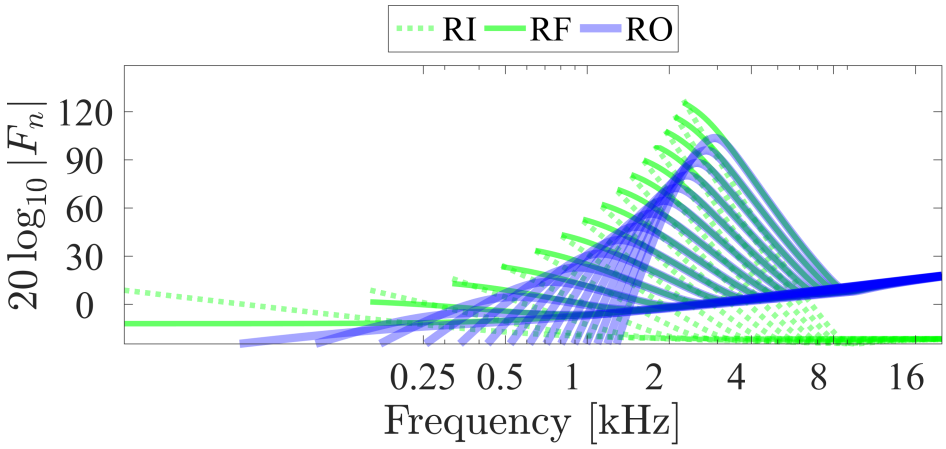
Several methods for spherical array encoding rely on Tikhonov regularization to keep the results bounded. This is equivalent to apply the following BMFs:

$$\hat{F}_n^{\text{reg}} = \frac{F_n}{1 + \lambda^2 |F_n|^2}, \tag{4.25}$$

where F_n is one of the filters in (4.5) or (4.6), and λ is the regularization parameter. A similar expression has been reported in [56]. Examples of BMFs designed using (4.25) are shown in Fig. 4.11.



(a) $a = 8.5$ cm, $b = 100$ cm.



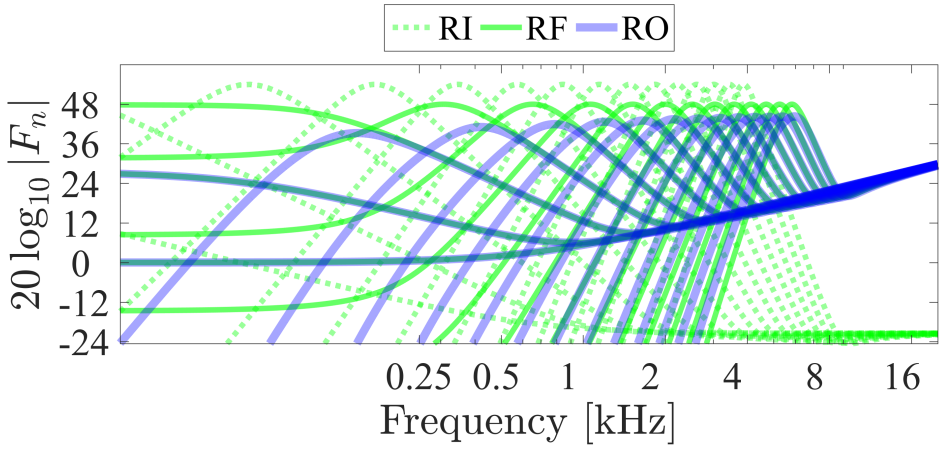
(b) $a = 8.5$ cm, $b = 25$ cm.

Figure 4.10: Examples of discontinuous band-limited RI-BMFs and RF-BMFs based on (4.22) and (4.23).

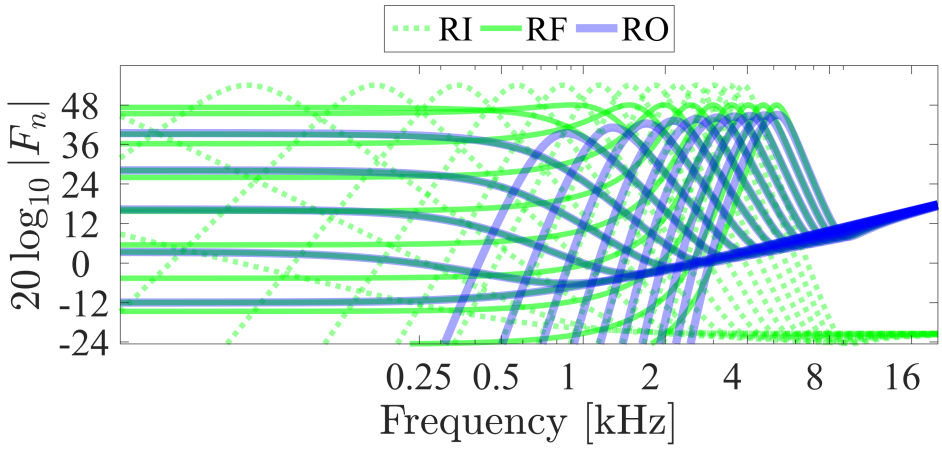
4.4.2.3 Magnitude-bounded RO-BMFs

In (4.21), relative maximum values occurs at frequencies that depend exclusively b. Relative maximum values can therefore attenuated and shifted towards the high frequencies by introducing a new reproduction radius ϱ. This procedure defines the following magnitude-bounded RO-BMFs:

$$\hat{F}_n^{\mathrm{RO}}(a, b, \varrho, k) = |\mathcal{O}_n(\varrho, k)| \times F_n^{\mathrm{RF}}(a, b, k). \tag{4.26}$$



(a) $a = 8.5$ cm, $b = 100$ cm.



(b) $a = 8.5$ cm, $b = 25$ cm.

Figure 4.11: Regularized RI-BMFs and RF-BMFs based on (4.25), for $\lambda^{\mathrm{RI}} = 10^{-3}$ and $\lambda^{\mathrm{RF}} = 2 \times 10^{-3}$, and RO-BMFs with magnitude bounded using (4.26) for $20\log|\gamma| = 40$.

Modifying the transition between the far and near fields in (4.23), ϱ can be calculated as follows:

$$\varrho = \frac{n}{k_\gamma}, \tag{4.27}$$

where k_γ is the wave number at which a prescribed maximum magnitude γ occurs, and is calculated in such a way that

$$k_\gamma : |F_n^{\mathrm{RF}}(a, b, k_\gamma)| = \gamma. \tag{4.28}$$

Examples of BMFs designed using (4.25) are shown in Figs. 4.10 and 4.11, where it can be observed that smooth and bounded BMFs are obtained using this proposed method.

4.4.2.4 Analysis of band limitation

A convenient way to examine the combined effects of F_n up to a finite order N in the spatial domain is derived. Substituting (4.3) and (4.4) in the finite sum version of (4.2) up to N, and swapping the integrals and sums by virtue of their absolute convergence, results in the integral operator form

$$\hat{\varphi}_N(\vec{r}, k) = \int_{\Omega_a \in \mathcal{A}} \int_{\Omega_b \in \mathcal{B}} \overline{\mathcal{T}}(\vec{r}, \vec{b}, k) K_N(\vec{a}, \vec{b}, k) \Psi(\vec{a}, k) d\Omega_b d\Omega_a, \tag{4.29}$$

where

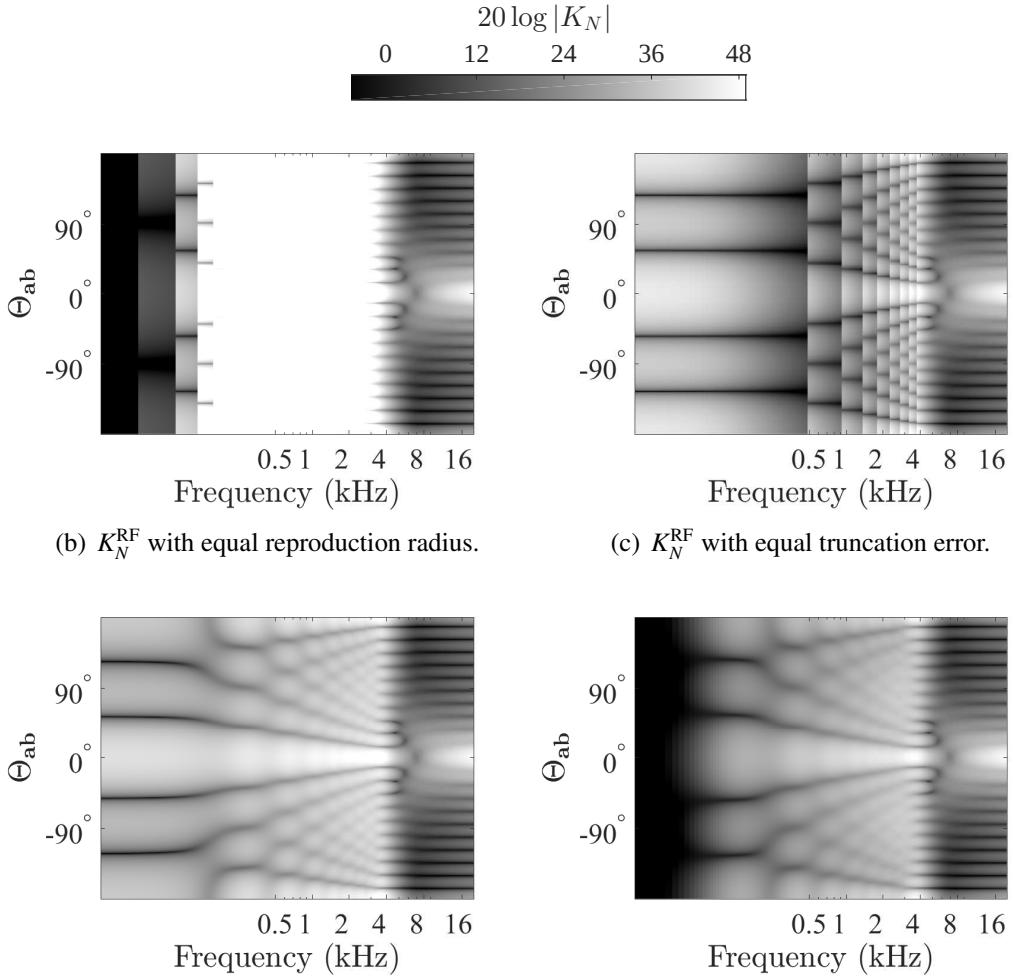
$$K_N(\vec{a}, \vec{b}, k) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} \hat{F}_n(a, b, k) P_n(\cos \Theta_{ab}). \tag{4.30}$$

The function K_N represents the matching between the boundaries $\mathcal{A}$ and $\mathcal{B}$ along angle and distance. Owing to the addition theorem of spherical harmonics in (2.16), the angular part of (4.30) is represented by a Legendre polynomial P_n evaluated at the angle between $\vec{a}$ and $\vec{b}$ denoted by Θ_{ab}. The radial part of (4.30) is one of the BMFs in (4.5), (4.6) or (4.21).

Examples of the angular and radial part of K_N are respectively illustrated in Figs. 4.12 and 4.13. It can be observed that the gains of bounded-magnitude RO-BMFs behave within the theoretical limits for sound field extrapolation established by (4.23). The bounded-magnitude RO-BMFs also provide smooth transition between the far and near fields, in contrast with the discontinuous transitions yielded by RF-BMFs designed with the limit for equal reproduction radius in (4.23). Furthermore, the bounded magnitude RO-BMFs ensure bounded and smooth results over the full range of frequencies and reproduction radii.

4.4.3 Analysis of error due to discretizing the continuous boundaries

In practice, computing (4.3) requires a finite number Q of pressure measurements over $\mathcal{A}$, while computing (4.4) requires a finite number L of secondary sources over $\mathcal{B}$. Given a sampling scheme of $\mathcal{A}$, represented by $\{\Omega_q, \alpha_q\}_{q=1,\ldots,Q}$, where Ω_q are measurement nodes



(b) K_N^{RF} with equal reproduction radius.

(c) K_N^{RF} with equal truncation error.

(d) K_N^{RF} with regularization.

(e) K_N^{RO} with bounded magnitude.

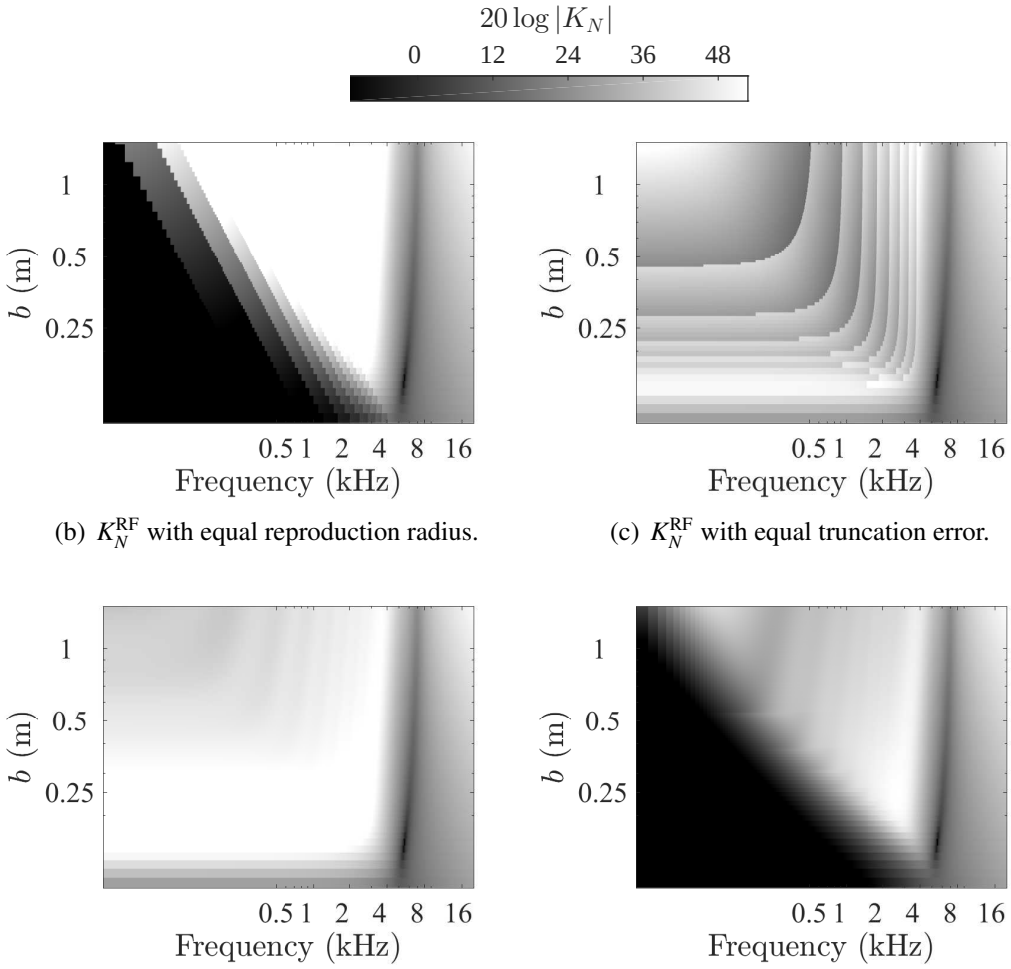
Figure 4.12: Example of K_N in (4.30) for $b = 1$ m and $N = 10$.

and α_q are quadrature weights, the corresponding orthonormality error $\epsilon^{(Q)}_{nmn'm'}$ is given by

$$\underbrace{\int_{\Omega_a \in \mathcal{A}} Y_n^m(\Omega_a)\overline{Y_{n'}^{m'}}(\Omega_a)d\Omega_a}_{\delta^{(\mathcal{A})}_{nmn'm'}} = \underbrace{\sum_{q=1}^{Q} Y_n^m(\Omega_q)\overline{Y_{n'}^{m'}}(\Omega_q)\alpha_q}_{\Delta^{(Q)}_{nmn'm'}} + \epsilon^{(Q)}_{nmn'm'}. \tag{4.31}$$

$$\underbrace{\int_{\Omega_b \in \mathcal{B}} Y_n^m(\Omega_b)\overline{Y_{n''}^{m''}}(\Omega_b)d\Omega_b}_{\delta^{(\mathcal{B})}_{nmn''m''}} = \underbrace{\sum_{\ell=1}^{L} Y_n^m(\Omega_\ell)\overline{Y_{n''}^{m''}}(\Omega_\ell)\beta_\ell}_{\Delta^{(L)}_{nmn''m''}} + \epsilon^{(L)}_{nmn''m''}. \tag{4.32}$$

Components of signals with finite directional bandwidth can be modeled with the Kronecker delta sifting property. Subsequently, the effects of spherical sampling within this



(b) K_N^{RF} with equal reproduction radius.

(c) K_N^{RF} with equal truncation error.

(d) K_N^{RF} with regularization.

(e) K_N^{RO} with bounded magnitude.

Figure 4.13: Example of K_N in (4.30) for $\Theta_{ab} = 0°$ and $N = 10$.

limited bandwidth is modeled by means of integration quadratures. The mathematical formulation is as follows:

$$\Psi_{nm} = \sum_{n',m'}^{N'} \Psi_{n'm'} \delta^{(\mathcal{A})}_{nmn'm'} \approx \sum_{n',m'}^{N'} \Psi_{n'm'} \Delta^{(Q)}_{nmn'm'}, \tag{4.33}$$

$$\mathcal{T}_{nm} = \sum_{n'',m''}^{N''} \mathcal{T}_{n''m''} \delta^{(\mathcal{B})}_{nmn''m''} \approx \sum_{n'',m''}^{N''} \mathcal{T}_{n''m''} \Delta^{(L)}_{nmn''m''}. \tag{4.34}$$

Several sampling theorems on the sphere exist in the literature to relate a finite band-

width and a sampling scheme. In general, we assume that

$$\begin{aligned} (N'+1)^2 &\leq Q, \\ (N''+1)^2 &\leq L, \\ N &= \min(N', N''). \end{aligned} \tag{4.35}$$

It can be shown that

$$\hat{\varphi} = \sum_{n,m}^{N} \sum_{n',m'}^{N'} \sum_{n'',m''}^{N''} \overline{\mathcal{T}_{n''m''}} F_n \Psi_{n'm'} \Delta^{(Q)}_{nmn'm'} \Delta^{(L)}_{nmn''m''}. \tag{4.36}$$

Since

$$\Delta^{(Q)}_{nmn'm'} \Delta^{(L)}_{nmn''m''} = \delta^{(\mathcal{A})}_{nmn'm'} \delta^{(\mathcal{B})}_{nmn''m''} - \delta^{(\mathcal{A})}_{nmn'm'} \epsilon^{(L)}_{nmn''m''} - \delta^{(\mathcal{B})}_{nmn''m''} \epsilon^{(Q)}_{nmn'm'} + \epsilon^{(Q)}_{nmn'm'} \epsilon^{(L)}_{nmn''m''}, \tag{4.37}$$

reproduced pressure signal can also be written as follows

$$\hat{\varphi} = \hat{\varphi}_N + \xi_Q + \xi_L + \xi_{Q,L}, \tag{4.38}$$

where

$$\hat{\varphi}_N = \sum_{n,m}^{N} \overline{\mathcal{T}_{nm}} F_n \Psi_{nm}, \tag{4.39}$$

$$\xi_Q = -\sum_{n,m}^{N} \sum_{n',m'}^{N'} \overline{\mathcal{T}_{nm}} F_n \Psi_{n'm'} \epsilon^{(Q)}_{nmn'm'}, \tag{4.40}$$

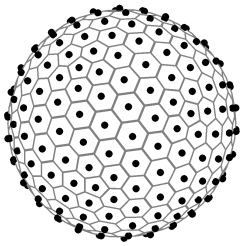
$$\xi_L = -\sum_{n,m}^{N} \sum_{n'',m''}^{N''} \overline{\mathcal{T}_{n''m''}} F_n \Psi_{nm} \epsilon^{(L)}_{nmn''m''}, \tag{4.41}$$

and

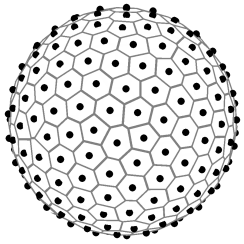
$$\xi_{Q,L} = \sum_{n,m}^{N} \sum_{n',m'}^{N'} \sum_{n'',m''}^{N''} \overline{\mathcal{T}_{n''m''}} F_n \Psi_{n'm'} \epsilon^{(Q)}_{nmn'm'} \epsilon^{(L)}_{nmn''m''}. \tag{4.42}$$

The following errors are therefore defined:

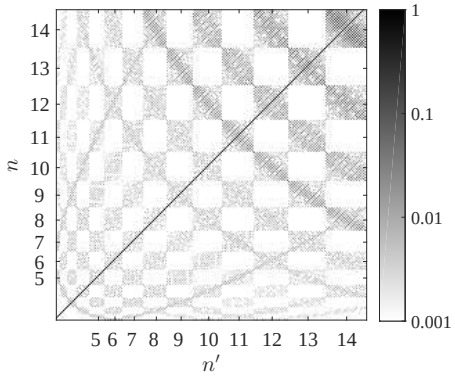
$$E_Q = \frac{|\xi_Q|^2}{|\hat{\varphi}_N|^2}, \quad E_L = \frac{|\xi_L|^2}{|\hat{\varphi}_N|^2}, \quad E_{Q,L} = \frac{|\xi_{Q,L}|^2}{|\hat{\varphi}_N|^2}. \tag{4.43}$$



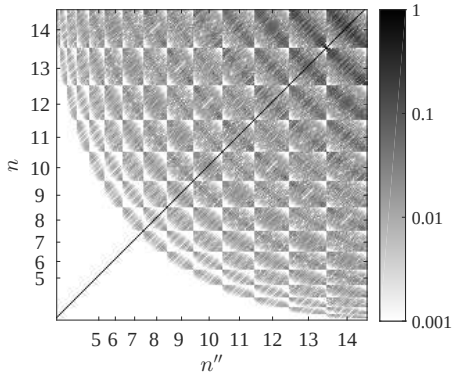
(a) Icosahedral grid with 252 points.



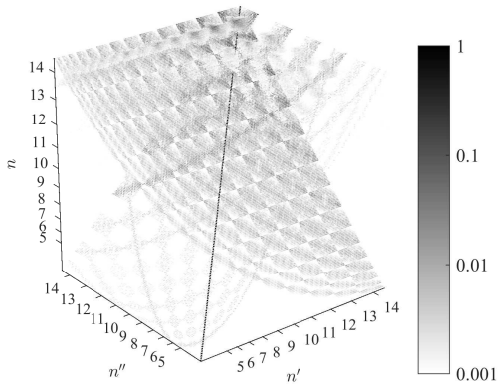
(b) Spherical n-design grid with 225 points.



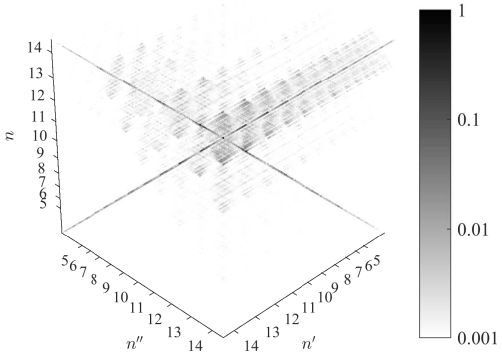
(c) $\Delta^{(Q)}_{nmn'm'}$.



(d) $\Delta^{(L)}_{nmn''m''}$.



(e) $\Delta^{(Q)}_{nmn'm'} \times \Delta^{(L)}_{nmn''m''}$.



(f) $\Delta^{(Q)}_{nmn'm'} \times \Delta^{(L)}_{nmn''m''}$.

Figure 4.14: Perturbation patterns in the spherical harmonic domain due to spatial discretization for two spherical samplings.

4.5 Numerical experiment for binaural synthesis

An application to binaural synthesis is exemplified in this section. In binaural synthesis based on virtual loudspeakers, existing filters for microphone-to-loudspeaker signal conversion do not take into account the boundary conditions for reproduction. However, they

can be classified in the theoretical framework provided in this chapter.

Existing plane-wave filters cannot take into consideration the distance information of the sound source. Furthermore, no information regarding the boundary conditions used during reproduction. These filters can therefore simply be classified as having a reproduction boundary of infinite radius. Existing spherical wave filters, on the other hand, can take into consideration the distance information of the sound source. They can be classified as having a reproduction boundary of finite radius. However, these filters produces high amplification at lower. To apply plane-wave and spherical wave filters to real cases, the amplitudes are typically limited by simply discarding the spectral information at lower frequencies. However, filters for arrays with a large number of microphones, as required in binaural synthesis, still lacks of high amplifications at middle frequencies.

The proposed boundary matching filters for an open reproduction boundary consitute a proper tool for dealing with the excessive amplifications at lower and middle frequencies. This is bases on displacing the critic points of the filter that define their maximum values. Displacement is performed toward the high frequencies.

An application of this filters to binaural synthesis is exemplified in Fig. 4.15. Synthesis was performed by assuming a rigid array for recording that consists of 252 microphones. A dataset of HRTFs for 252 virtual loudspeaker arrays constitute the acoustic transfer functions required by the model. The positioning of both microphones and virtual loudspeakers was decided according to spherical grids that are based on the geometry of the icosahedron. It can be observed that synthesis with the spherical wave filters produces discontinuities due to the gain limitation procedure, which simply discards spectral information at low frequencies. Synthesis with the proposed open boundary filters, on the contrary, yields smooth results at lower and middle frequencies thanks to the fact that the procedure for gain limitation in this case do not affect the continuity of the boundary matching filters. In general, accurate synthesis is always possible up to a maximum frequency determined by the number of microphones in the array.

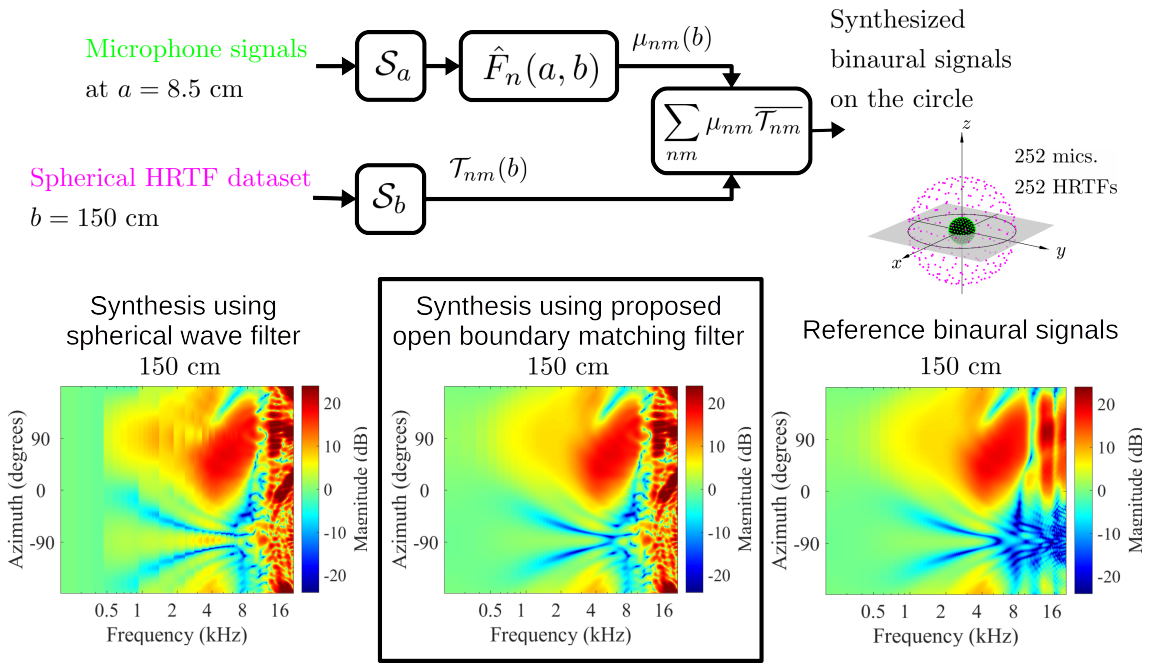


Figure 4.15: Example of binaural synthesis.

4.6 Summary

Boundary matching filters (BMFs) for obtaining free-field representations of sound fields in the vicinity or outside a rigid spherical microphone array were presented.

By assuming a hypothetical reproduction boundary at a reference radius, the classic plane-wave and spherical-wave filters used for recording in the far and near fields were found to be equivalent to using BMFs for a reproduction boundary of infinite radius (spherical-wave filters), or a reproduction boundary of finite radius (plane-wave filters). To cope with the excessive gain at lower and middle frequencies yielded by plane-wave and spherical-wave filters, an open or acoustically transparent reproduction boundary is considered. This results in analytic BMFs that constitute a proper tool for handling the excessive amplifications. Bounded and smooth encoding in the full range of frequencies and distances is possible with BMFs for an open reproduction boundary, without the need of frequency-dependent angular band limitation and regularization. The three classes of BMFs considered in this section were analyzed in a mathematical framework to identify

the effects of angular band limitation and spatial discretization.

The principle of boundary matching filtering might also be applied to other simple geometries.

Chapter 5

Editing and combining spatial sound information for binaural synthesis

5.1 Overview

A general model for binaural synthesis is detailed in this Chapter. An overview of the model and the assumed geometry are respectively shown in Figs 5.1. This model allows to edit and combine distinct types of spatial sound information defined on the sphere. For instance, microphone array recordings and datasets of head-related transfer functions. The model therefore rely on the models formulated in previous chapters. More specifically, the model combines distance-varying filters and boundary-matching filters.

As an application example, this model can be used to implement a binaural system capable of recording far sounds in all directions. Sound in the recorded auditory scene can subsequently be discriminated along directions, and independently brought closer to the listener. Proceeding in this way, sounds can be heard as if they were closer than their original distances.

It is important to mention that each angular window in this model can be interpreted as

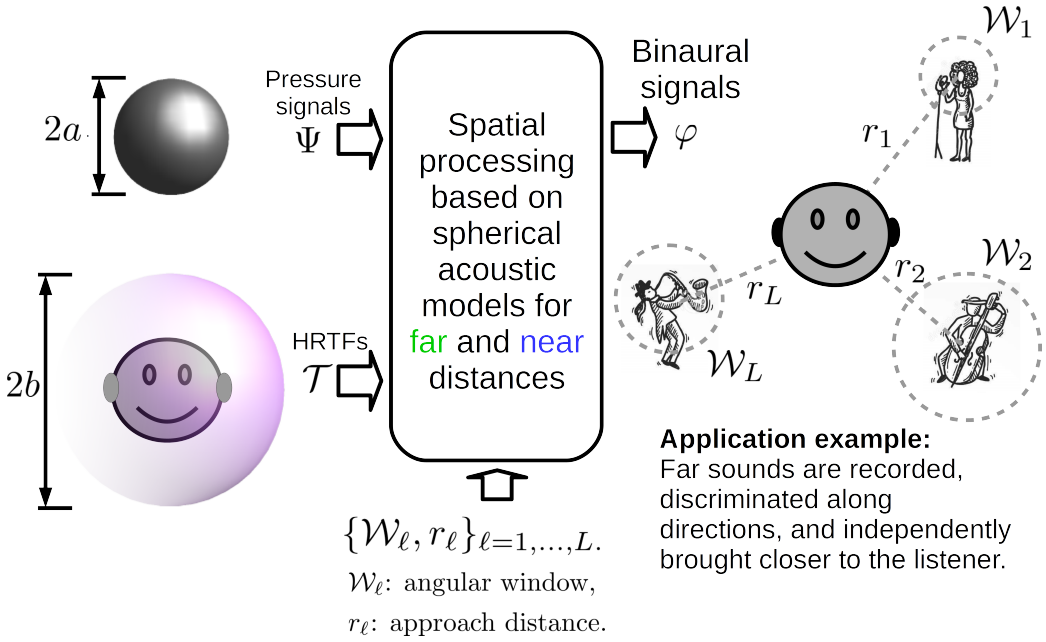


Figure 5.1: General binaural synthesis model. In the application example, an angular window $\mathcal{W}_\ell$ defines the angular region in which far sounds are approached to a desired distance r_ℓ.

an angular region along which distance is edited or manipulated. Before formulating the model in detail, the preliminary mathematical considerations are given in the next section.

5.2 Mathematical considerations

Combining spatial information on the sphere relying on the principle of wave superposition [115, 118] involves the calculation of integral of products.

The integral of the product of two functions on the sphere can be equivalently calculated in the transformed domain. As a consequence of the orthonormality property in (2.12), the integral becomes a sum in the transformed domain, in such a way that

$$\int_{\Omega\in\mathbb{S}^2} \Psi^{(1)}(\Omega)\overline{\Psi^{(2)}}(\Omega)d\Omega = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \Psi^{(1)}_{nm}\overline{\Psi^{(2)}}_{nm}. \tag{5.1}$$

Here, $\Psi^{(1)}_{nm}$ and $\Psi^{(2)}_{nm}$ respectively denote the spherical Fourier transforms of $\Psi^{(1)}$ and $\Psi^{(2)}$.

The integral of the product of three (or more) functions on the sphere, however, requires

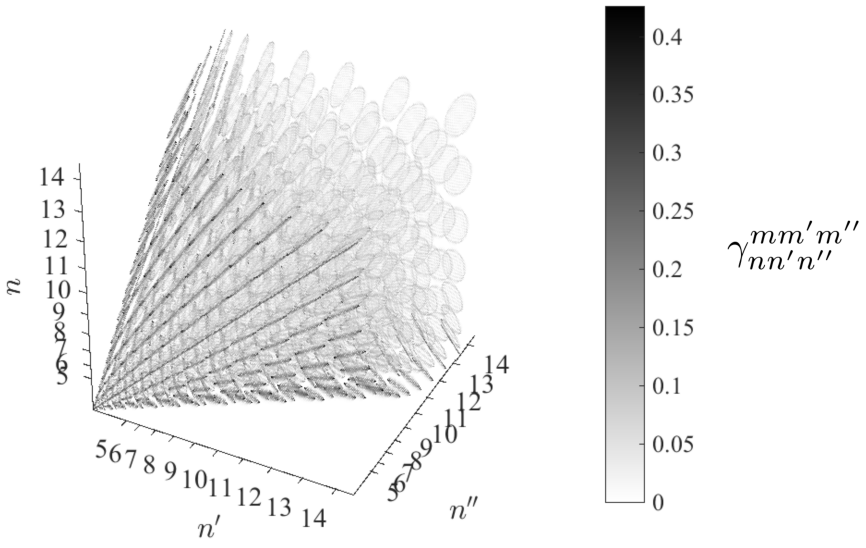


Figure 5.2: Example of coefficients $\gamma_{nn'n''}^{mm'm''}$ in (5.3).

a more elaborated formulation that can be interpreted as the extension of the concept of orhotnormality to consider three or more functions. The mathematical formulation reads

$$\int_{\Omega\in\mathbb{S}^2} \Psi^{(1)}(\Omega)\Psi^{(2)}(\Omega)\Psi^{(3)}(\Omega)d\Omega = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'}\sum_{n''=0}^{\infty}\sum_{m=-n''}^{n''} \Psi_{nm}^{(1)}\Psi_{n'm'}^{(2)}\Psi_{n''m''}^{(3)}\gamma_{nn'n''}^{mm'm''}, \tag{5.2}$$

where $\Psi_{nm}^{(1)}$, $\Psi_{nm}^{(2)}$ and $\Psi_{nm}^{(3)}$ denote the spherical Fourier transforms of $\Psi^{(1)}$, $\Psi^{(2)}$ and $\Psi^{(3)}$, respectively, and

$$\gamma_{nn'n''}^{mm'm''} = \int_{\Omega\in\mathbb{S}^2} Y_n^m(\Omega)Y_{n'}^{m'}(\Omega)Y_{n''}^{m''}(\Omega)d\Omega. \tag{5.3}$$

An analytic solution to (5.3) exist, which defines the so-called Gaunt coefficients. An example of this coefficients is illustrated in Fig. 5.2.

5.3 General model

The mathematical considerations presented above are used to model the general binaural system in Fig. 5.1 by considering the underlying geometry shown in Fig. 5.3.

At a first stage, the loudspeaker signals μ at each approach distance r_ℓ, where ℓ =

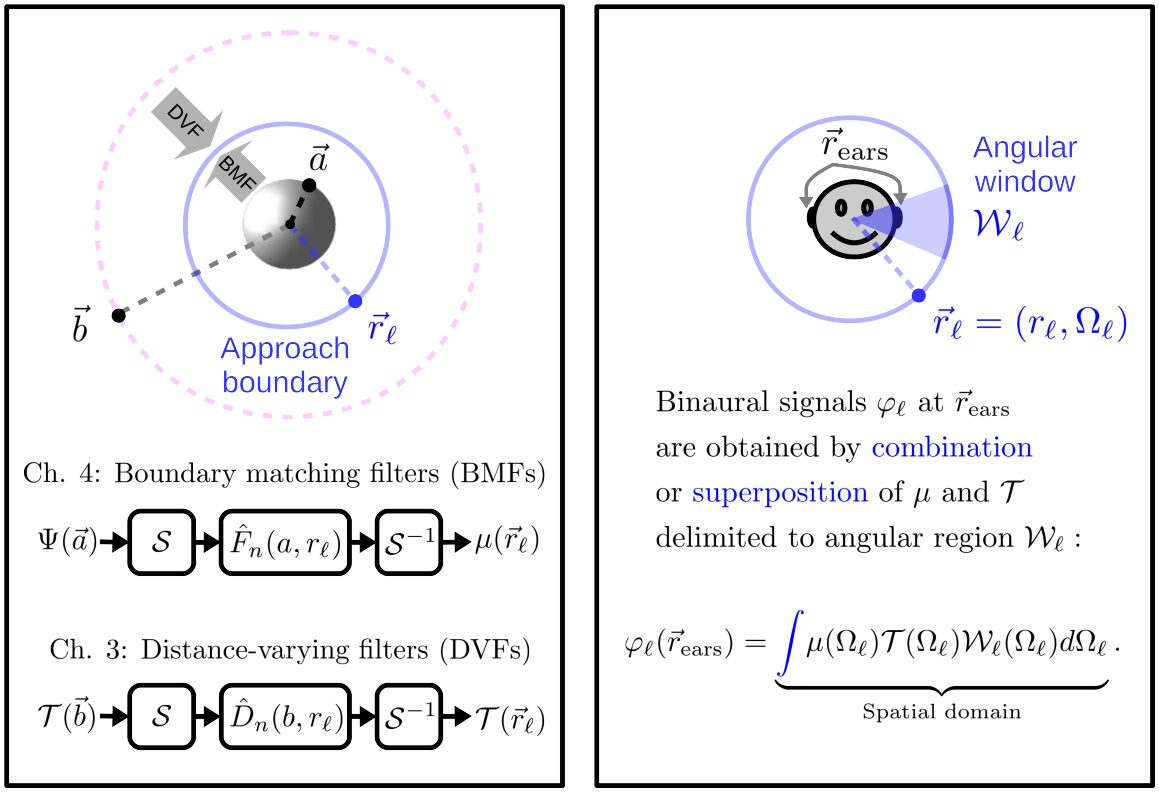


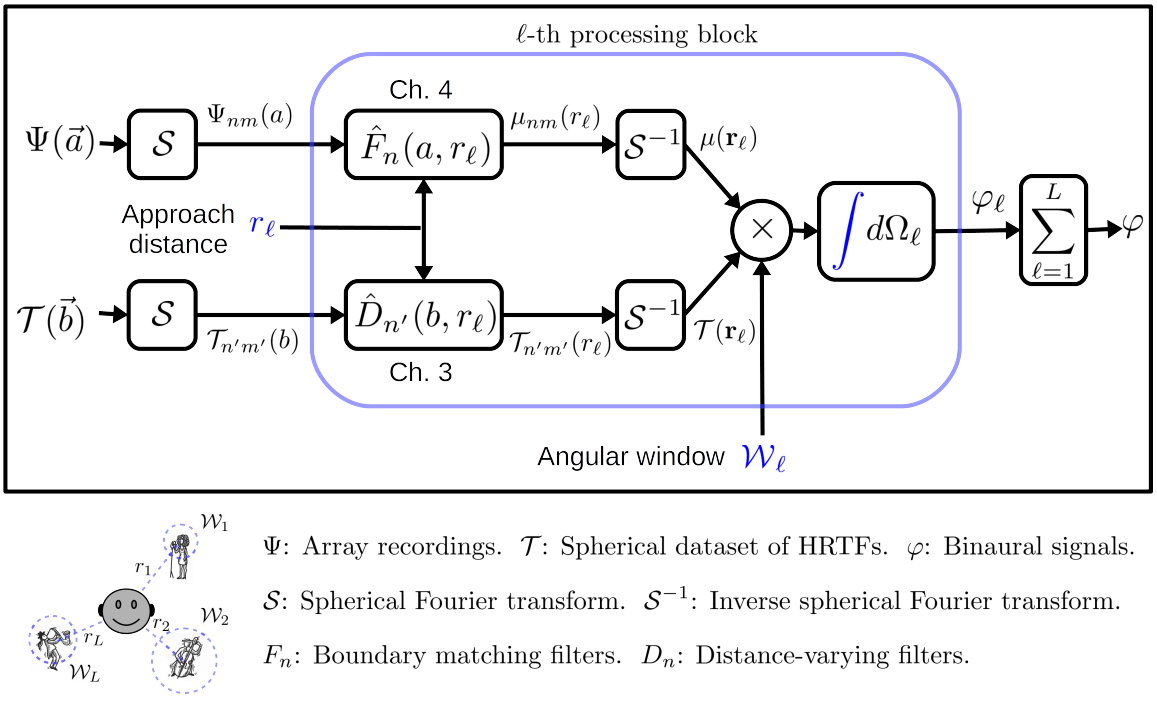
Figure 5.3: Geometry for the general binaural synthesis model. The angular window $\mathcal{W}_\ell$ defines the angular region in which far sounds are approached to a desired distance r_ℓ.

$1, 2, ..., L$, are calculated from the microphone array recordings Ψ by means of boundary matching filter F_n. An dataset $\mathcal{T}$ of HRTFs at the approach distance is also obtained from the initial dataset a far distance b by means of distance-varying filters D_n.

The second stage performs the combination of μ and $\mathcal{T}$ previously obtained for each of the approach boundaries at r_ℓ. Combination is done based on the principle of wave superposition. Superposition in this case, however, has an additional information given by each of the angular windows $\mathcal{W}_\ell$ that allows for the discrimination of sounds along directions. Superposition is formulated in the spatial domain as follows:

$$\varphi_\ell(\vec{r}_{\text{ears}}) = \int \mu(\Omega_\ell)\mathcal{T}(\Omega_\ell)\mathcal{W}_\ell(\Omega_\ell)d\Omega_\ell. \tag{5.4}$$

The direct implementation the general model based on the integral above leads to the spatial processing diagram shown in Fig. 5.4.



Ψ: Array recordings. $\mathcal{T}$: Spherical dataset of HRTFs. φ: Binaural signals.
$\mathcal{S}$: Spherical Fourier transform. $\mathcal{S}^{-1}$: Inverse spherical Fourier transform.
F_n: Boundary matching filters. D_n: Distance-varying filters.

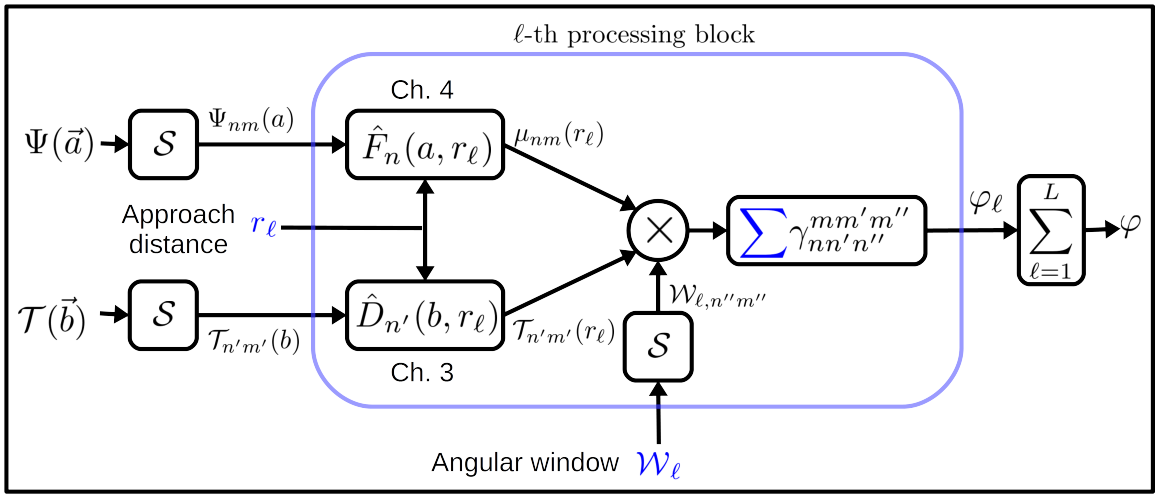
Figure 5.4: General model for binaural synthesis by using acoustic superposition in the spatial domain.

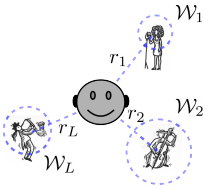
Performing the combination or superposition of spatial information by integrating in the spatial domain, however, necessarily involves the use of samplings of the sphere. This procedure introduces an additional parameters that is the sampling scheme and produces a model that highly depends on this additional sampling scheme.

The distribution of pressure μ and the HRTF datasets $\mathcal{T}$, however, are already available in the transform domain as μ_{nm} and $\mathcal{T}_{n'm'}$, respectively, as can be observed in the diagram in Fig. 5.4. It would be useful, then, if a similar procedure for performing the superposition be available directly in the transform domain.

This is possible by means of the mathematical consideration presented in the previous section. In fact, superposition can also be formulated as follows:

$$\varphi_\ell(\vec{r}_{\text{ears}}) = \int \mu(\Omega_\ell)\mathcal{T}(\Omega_\ell)\mathcal{W}(\Omega_\ell)d\Omega_\ell = \sum_{nm}^{\infty}\sum_{n'm'}^{\infty}\sum_{n''m''}^{\infty} \mu_{nm}\mathcal{T}_{n'm'}\mathcal{W}_{\ell,n''m''}\gamma_{nn'n''}^{mm'm''}, \qquad (5.5)$$





Ψ: Array recordings. $\mathcal{T}$: Spherical dataset of HRTFs. φ: Binaural signals.
$\mathcal{S}$: Spherical Fourier transform. $\gamma_{nn'n''}^{mm'm''} = \int Y_n^m(\Omega) Y_{n'}^{m'}(\Omega) Y_{n''}^{m''}(\Omega) d\Omega$.
F_n: Boundary matching filters. D_n: Distance-varying filters.

Figure 5.5: General model for binaural synthesis by using acoustic superposition in the transform domain.

where

$$\gamma_{nn'n''}^{mm'm''} = \int Y_n^m(\Omega) Y_{n'}^{m'}(\Omega) Y_{n''}^{m''}(\Omega) d\Omega. \tag{5.6}$$

The direct implementation the general model based on the sum above leads to the spatial processing diagram shown in Fig.5.5

5.3.1 Analysis of band limitation

Similarly to the analysis presented in section 4.4.2.4, a convenient way to examine the combined effects of the synthesis process up to finite orders N, N' and N'', such that $N = \max(N', N'')$, is derived in this section. Substituting the results of chapter 3 and 4 in the finite sum version of (5.5) up to $N = \max(N', N'')$, and swapping the integrals and sums by virtue of their absolute convergence, results in the integral operator form

$$\hat{\varphi}_{\ell,N}(\vec{r}_{\text{ears}}, k) = \int_{\Omega_a \in \mathcal{A}} \int_{\Omega_b \in \mathcal{B}} \overline{\mathcal{T}}(\vec{r}_{\text{ears}}, \vec{b}, k) K_{\ell,N}(\vec{a}, \vec{b}, k) \Psi(\vec{a}, k) d\Omega_b d\Omega_a, \tag{5.7}$$

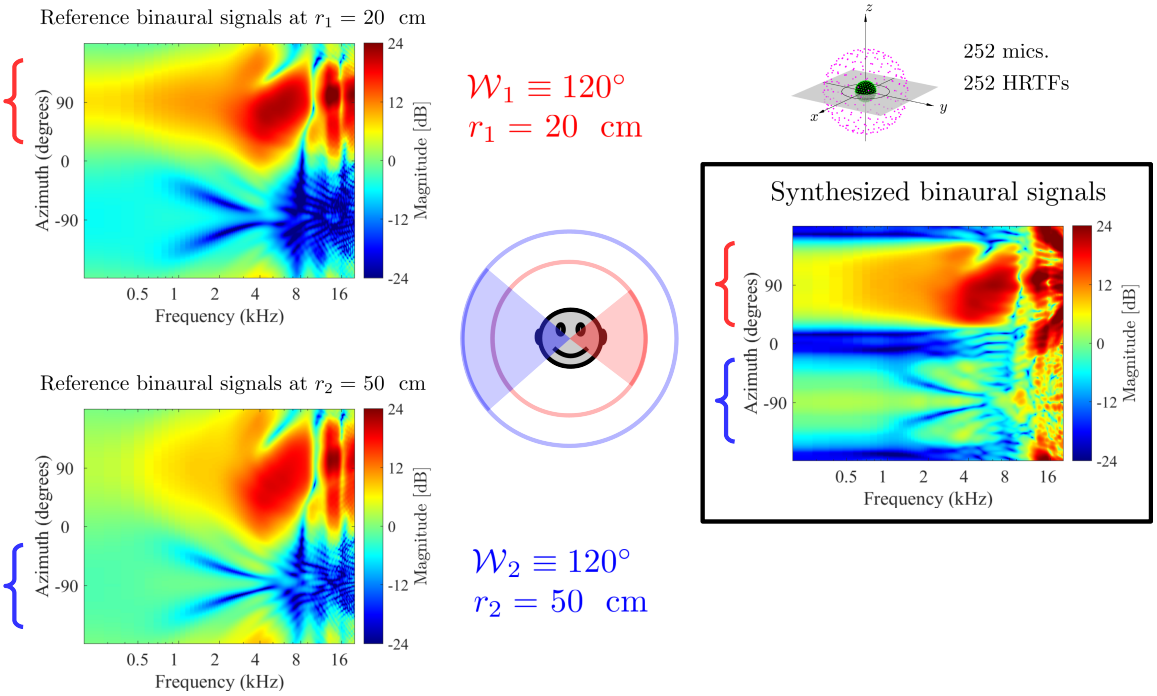


Figure 5.6: Binaural synthesis using a spherical box window.

where

$$K_{\ell,N}(\vec{a},\vec{b},k) = \sum_{nm}^{N}\sum_{n'm'}^{N'}\sum_{n''m''}^{N''} \hat{F}_n(a,r_\ell,k)\overline{Y_n^m}(\Omega_a)\overline{\hat{D}_{n'}}(b,r_\ell,k)Y_{n'}^{m'}(\Omega_b)\mathcal{W}_{\ell,n''m''}\gamma_{nn'n''}^{mm'm''}. \quad (5.8)$$

5.3.2 Synthesis of binaural signals

An application of the general model for binaural synthesis is exemplified in Figs. 5.6 and 5.7. Synthesis in both cases was performed by assuming a rigid array for recording that consists of 252 microphones. A dataset of HRTFs for 252 virtual loudspeaker arrays constitute the acoustic transfer functions required by the model. The positioning of both microphones and virtual loudspeakers was decided according to spherical grids that are based on the geometry of the icosahedron.

Two angular regions discriminated by using the spherical version of a box window (a spherical cap window) were used when generating the synthesized binaural signals shown in Fig. 5.6. The approach distances were also different along each angular window. It

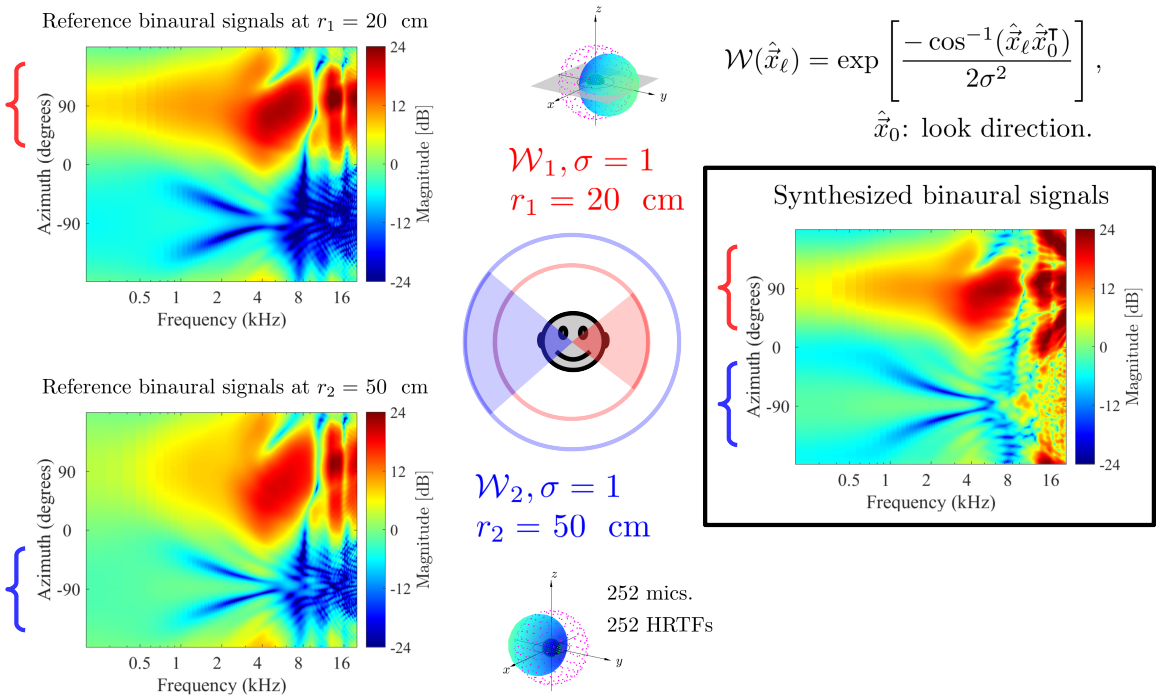


Figure 5.7: Binaural synthesis using a spherical Gaussian window.

can be observed that angular regions can be discriminated and the distance patterns in the binaural signals are also properly synthesized. However, some distortions are also observed in these patterns.

A similar procedure was performed when synthesizing the results shown in Fig. 5.7 but using a spherical Gaussian window in this case. Synthesis results were smother than the case of a box window. However, in both cases, acceptable synthesis was only possible up to a maximum frequency determined by the number of microphones in the array.

5.4 Binaural beamforming approach

Binaural synthesis can also be formulated in terms of the filter-and-sum structure for beamforming shown in Fig. 5.8. When applied to binaural synthesis, the filters w_q aim to synthesize spatial patterns given by HRTF datasets. This structure is convenient for investigating the effects of the number of microphones on binaural synthesis, and also for

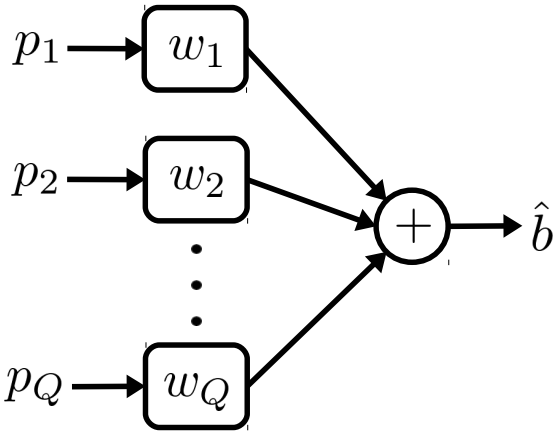


Figure 5.8: Binaural beamforming.

the study of the propagation of noise through the synthesis model. Another benefit of the beamforming approach is that it is not limited to spherical geometries. In this section, a general model for binaural synthesis is formulated in the light of the theory of filter-and-sum beamforming. The formulation starts directly in discrete geometries.

5.4.1 Filter-and-sum beamforming in arbitrary geometries

For a single frequency, the microphone array recordings are organized in the vector

$$\mathbf{p} = \begin{bmatrix} p_1 & p_2 & \cdots & p_Q \end{bmatrix}^\intercal . \tag{5.9}$$

The symbol $^\intercal$ indicates transpose. Each entry p_q of $\mathbf{p}$, where $q = 1, 2, ..., Q$, represents a sample in the frequency domain of a sound pressure signal recorded at an arbitrary position $\vec{a}_q$.

The dataset of HRTFs is organized in the matrix

$$\mathbf{h} = \begin{bmatrix} h_1^{\text{left}} & h_2^{\text{left}} & \cdots & h_L^{\text{left}} \\ h_1^{\text{right}} & h_2^{\text{right}} & \cdots & h_L^{\text{right}} \end{bmatrix}^\intercal . \tag{5.10}$$

Each entry h_ℓ^{left} or h_ℓ^{right} of $\mathbf{h}$, where $\ell = 1, 2, ..., L$, represents a sample in frequency of a free-field HRTF for the left or right ear, respectively. Each entry is characterized for a

single arbitrary sound source position $\vec{b}_\ell$. These kind of positions are referred to as the *virtual loudspeaker positions* throughout this section.

The synthesized binaural signals for the left and right ears are organized in the pair

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{b}^{\text{left}} & \hat{b}^{\text{right}} \end{bmatrix}^{\mathsf{T}}. \tag{5.11}$$

Binaural synthesis can be summarized as the following linear combination of $\mathbf{p}$ and $\mathbf{h}$:

$$\hat{\mathbf{b}} = \mathbf{h}^{\dagger}\mathbf{K}\mathbf{p}, \tag{5.12}$$

where $\mathbf{K} = [\kappa_{\ell q}]$ is a combination matrix of size $L \times Q$ and † indicates conjugate transpose. The matrix elements $\kappa_{\ell q}$ should not be confused with the wave number k defined in (2.5).

Characterizing $\mathbf{K}$ requires to take into consideration the geometry and physical topology of the microphone array, as well as the distribution of virtual loudspeakers used to obtain the HRTF dataset. Most of the existing methods for obtaining $\mathbf{K}$ can be gathered into two dual approaches depending on weather the products $\mathbf{h}^{\dagger}\mathbf{K}$ or $\mathbf{K}\mathbf{p}$ are optimized.

In this study, the first approach is followed because it is closely related with the theory of filter-and-sum beamforming, which provide a convenient framework for investigating the effects of noise. This leads to the diagram for binaural beamforming shown in Fig. 5.8. In this context, the binaural synthesis equation in (5.12) is written as follows:

$$\hat{\mathbf{b}} = \mathbf{w}^{\dagger}\mathbf{p}, \tag{5.13}$$

where the beamformer matrix

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}^{\text{left}} \\ \mathbf{w}^{\text{right}} \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} w_1^{\text{left}} & w_2^{\text{left}} & \cdots & w_Q^{\text{left}} \\ w_1^{\text{right}} & w_2^{\text{right}} & \cdots & w_Q^{\text{right}} \end{bmatrix}^{\mathsf{T}} \tag{5.14}$$

contains the weighting coefficients that are applied to $\mathbf{p}$ so as to synthesize the spatial

patterns defined by **h**. The beamformer matrix is defined from (5.12) and (5.13) as follows:

$$\mathbf{w} = \mathbf{K}^{\dagger}\mathbf{h}, \tag{5.15}$$

where a suitable choice of **K** is assumed[3].

5.4.2 Filter-and-sum beamforming in spherical geometries

In spherical geometries, the binaural synthesis equation (5.12) can be regarded as a discrete-space version of the reproduction equation (4.2) when it is expressed as the integral operator form in (4.29). The entries of the recording vector **p** are spatial samplings of the sound pressure field Ψ, while the entries of the HRTF dataset **h** are spatial samplings of the acoustic transfer functions $\mathcal{T}$ when the listener is present and the reproduction point $\vec{r}$ corresponds to the positions of the ears $\vec{r}_{\text{ears}}$. A clear relation is therefore expected between the combination matrix **K** and the function K_N. In effect, the entries of $\kappa_{\ell q}$ of **K** are defined by the spatial sampling of K_N along the recording and reproduction boundaries.

Although not treated in this thesis, it is important to mention that a similar relation can be established between the binaural synthesis equation (5.12) and the general reproduction equation presented in (5.7). Furthermore, the same reasoning can be applied in other systems of coordinates, as long as it exist an integral operator form to describe the matching between the recording and reproduction boundaries. We delimit the discussion to spherical geometries in the remainder of this thesis.

When a rigid spherical microphone array of radius a is used for recording, each entry p_q of **p** in (5.9) correspond to a microphone signal recorded at $\vec{a}_q = (a, \Omega_q)$. Similarly, when a spherical virtual loudspeaker array of radius b is used to obtain the HRTF dataset **h** in (5.10), each entry h_ℓ^{left} or h_ℓ^{right} corresponds to a virtual loudspeaker position $\vec{b}_\ell = (b, \Omega_\ell)$.

In such spherical geometries, the combination matrix of size $L \times Q$ in (5.12) is calculated

[3]More precisely, **w** is obtained by solving a linear system $\mathbf{Bw} = \mathbf{h}$, were matrix **B** of size $Q \times L$ contains all the transfer functions describing the transmission of sound from each virtual loudspeaker position to each microphone position. Assuming that the pseudoinverse $\mathbf{B}^{+}$ exists, the choice $\mathbf{K}^{\dagger} = \mathbf{B}^{+}$ defines (5.14).

according to the following expression:

$$\mathbf{K} = \mathbf{D}\mathbf{E}^{+}, \tag{5.16}$$

where $\mathbf{E}^{+}$ and $\mathbf{D}$ denote the encoding and decoding matrices, respectively. In general terms, the matrix $\mathbf{K}$ can also be interpreted as a spatial resampling from the resolution of the spherical microphone array to the resolution of the HRTF dataset.

The matrix $\mathbf{E}^{+}$ calculates a representation of $\mathbf{p}$ in terms of harmonic solutions to the acoustic wave equation up to order N. It further compensates for the presence of the spherical baffle, and extrapolates the resulting free-field representation from a to b. The matrix $\mathbf{E}^{+}$ is obtained by calculating the pseudo-inverse of a matrix $\mathbf{E}$ by using Tikhonov regularization [120]. The entries of $\mathbf{E}$ represent acoustic transfer functions from an arbitrary position at a distance b to each microphone position $\vec{a}_q$. The size of $\mathbf{E}$ is $Q \times (N+1)^2$ and its entries, for the assumption of virtual loudspeakers radiating spherical waves, are

$$e_{q,n^2+n+m+1} = \frac{-h_n(kb)Y_n^m(\Omega_q)}{ka^2 h'_n(ka)}. \tag{5.17}$$

Here, h_n denotes the spherical Hankel function of the second kind and order n, while Y_n^m denote the complex spherical harmonic functions of order n and degree m, where $n = 0, 1, ..., N$, and $m = -n, -n+1, ..., n$. The functions h_n and Y_n^m are defined in [83], respectively as the radial and angular portions of the solutions to the acoustic wave equation. They are also functions of the wave number k. The symbol $'$ denotes the derivative of a function with respect to its argument. The benefit of performing regularization by including the radial portion is the attenuation of high-order components at low frequencies.

The matrix $\mathbf{D}$ takes the encodings $\mathbf{E}^{+}\mathbf{p}$ at a radius b and decodes them for each virtual loudspeaker directions Ω_ℓ. The size of $\mathbf{D}$ is $L \times (N+1)^2$ and its entries are

$$d_{\ell,n^2+n+m+1} = \frac{\exp(jkb)}{b} Y_n^m(\Omega_\ell). \tag{5.18}$$

The fraction represents a free-field transfer function from b to the head center. Its purpose is to set the head center as the observation point, in consistency with the definition of free-field HRTFs.

Equating (5.17) and (5.18), and using the analytic Tikhonov regularized solution [120], it can be shown that the entries of $\mathbf{K} = [\kappa_{\ell q}]$ are given by

$$\kappa_{\ell q} = \frac{1}{4\pi \lfloor \sqrt{Q} \rfloor^2} \cdot \frac{\exp(jkb)}{b} \cdot \sum_{n=0}^{\lfloor \sqrt{Q} \rfloor - 1} (2n+1) F_n^{\mathrm{reg}}(a, b, k) P_n(\cos \Theta_{\ell q}). \tag{5.19}$$

The angular part of the sum in (5.19) is defined by the Legendre polynomial P_n of order n evaluated at the cosine of the angle $\Theta_{\ell q}$ between $\vec{b}_\ell$ and $\vec{a}_q$. The radial part is defined by the regularized radial filter

$$F_n^{\mathrm{reg}} = \frac{F_n}{1 + \lambda^2 |F_n|^2}, \quad F_n = -\frac{ka^2 h_n'(ka)}{h_n(kb)}, \tag{5.20}$$

where λ is the regularization parameter.

5.5 Effects of the number of microphones

HRTF datasets can be obtained for high-resolution source distributions using numerical methods [43]. Recently, a perceptual study [121] has reported that low-order HRTF representations might be sufficient to approximate an individual space, whereas an objective study [107] has identified HRTF features that would require high-order representations. Because the required resolution for characterizing an individual space is still an open question, the performance of binaural systems should be evaluated by considering HRTF datasets with the higher resolution that can be achieved.

Existing spherical microphone array recordings, on the other hand, typically contain low-order information only, since high-resolution spherical microphone arrays are still hard to construct using actual technology. This has motivated a recent study [75] on adapting the resolutions of HRTF datasets to spherical microphone array recordings by spatially down-

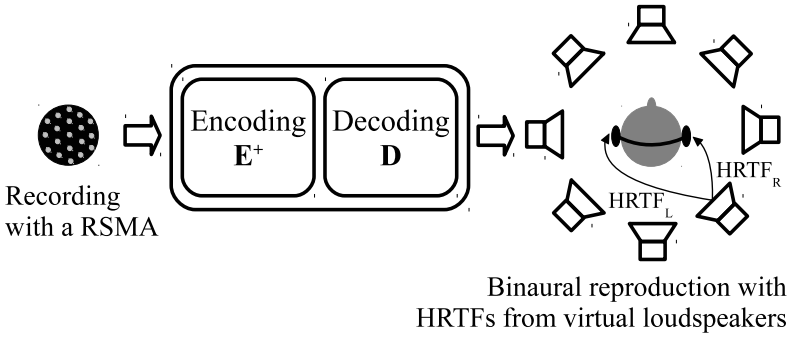


Figure 5.9: Overview of a binaural system composed of the following stages: recording with a rigid spherical microphone array (RSMA), spatial encoding and decoding, and binaural reproduction.

sampling the HRTF datasets. Nevertheless, it has also recently been reported in [64] that high-order information is important to synthesize more directionally sharpened and more externalized (outside-head) sounds. Regarding the future of recording technology, spherical microphone arrays of hundreds of microphones are not unrealistic. For instance, it has been reported in [74, 81, 82] a setup composed of 252 microphones distributed according to an icosahedral symmetry. Plans to construct higher resolution arrays in the near future also exist.

Bearing these considerations in mind, the present section seek to identify the number of microphones that are necessary to synthesize the binaural signals with a specified spatial accuracy. We therefore present an extensive numerical evaluation using spherical microphone array recordings and HRTF datasets of different resolutions, up to the amount required to cover all typical audible frequencies objectively. To focus the analysis on the number of microphones, spatial encodings of spherical microphone array recordings are calculated only. In connection with [75], this is equivalent to a spatial resampling of the spherical microphone array recordings so as to match the resolution of the HRTF datasets. Furthermore, to cope with low-frequency high amplifications typically observed when assuming HRTF datasets characterized for plane-wave sources, we consider the case of point sources, which is more consistent with existing HRTF datasets. An overview of the binaural system under consideration is shown in Fig. 5.9.

The effect of using different numbers of microphones (Q) and virtual loudspeakers

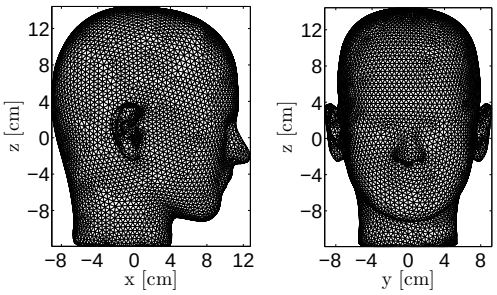
(L) on the synthesis accuracy was evaluated. To emphasize the preservation of spectral information used in human auditory localization, we gave special attention to the synthesis of monaural and interaural spectral cues.

Only one example virtual loudspeaker radius $b = 1.5$ m was used. Most of the available HRTF datasets are measured at this typical radius, beyond which the HRTFs hardly depend on distance [34]. The sound source to be recorded was also assumed placed at a 1.5 m distance. An exhaustive evaluation at different distances close to the head, while important, is outside the intended scope of this study.

5.5.1 Conditions of the evaluation

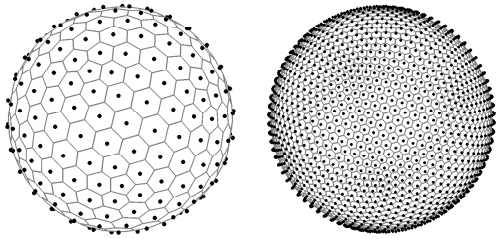
By $\hat{B} = \{\hat{B}_{\text{left}}(\Omega_i, f_j), \hat{B}_{\text{right}}(\Omega_i, f_j)\}$, we denote a set of *binaural transfer functions*, where $i = 1, 2, ..., I$, and $j = 1, 2, ..., J$. The set $\hat{B}$ was calculated using (5.12) and (5.19) with $\lambda = 1 \times 10^{-3}$, for the particular case of a number I of point sources placed at a 1.5 m distance, in the directions $\Omega_i = (\theta_i, \phi_i)$. The point sources were radiating non-simultaneously. Each point source was radiating a single sinusoidal signal at a time, and this case was repeated for a number J of single frequencies f_j in the full audible range. Similarly, by $H = \{H_{\text{left}}(\Omega_i, f_j), H_{\text{right}}(\Omega_i, f_j)\}$, we denote the reference HRTF datasets (target) calculated using the boundary element method (BEM) [43] for the head model described in Fig. 5.10. The mesh grid of this head model consists of $14,096$ points with an average cell length of 5.1 mm, which limited our evaluations to an average frequency of 16.6 kHz. Evaluations were based on comparisons of $\hat{B}$ and H.

To calculate $\hat{B}$, microphone signals for the non-simultaneous point sources (input) were first calculated with the algorithm in [122], assuming a rigid sphere of radius $a = 8.5$ cm. Then, HRTF datasets for a virtual loudspeaker array of radius $b = 1.5$ m (user parameter) were calculated using BEM [43] and the head model shown in Fig. 5.10. The positions of microphones and virtual loudspeakers were decided using spherical grids constructed by subdividing the edges of the icosahedron into equal segments. Examples of such grids are shown in Fig. 5.11, where dots indicate the positions and lines enclose their Voronoi cells.



(a) Lateral view. (b) Front view.

Figure 5.10: Head model used for numerical experiments.



(a) 252 points. (b) 1962 points.

Figure 5.11: Examples of icosahedral grids used for the arrangement of microphones and virtual loudspeakers.

Although not explicitly mentioned in (5.17) and (5.18), encoding and decoding fundamentally lie on numerical integrations on the sphere. The required quadrature weights were thus determined to be proportional to the cell areas.

To select the maximum order N_j required to approximate a frequency f_j, we used the bound proposed in [95]. Maximum orders N_j are determined by setting a constant truncation error within a region of interest enclosed by a given radius. In our simulations, we set a truncation error equal to 10^{-5} within the radius $a = 8.5$ cm of the microphone array. Source distance information is also considered by this rule; we set this variable equal to 1.5 m. Under these conditions, approximations up to an average limit frequency of 16.6 kHz would require an order $N = 43$ and, hence, at least $Q = (43 + 1)^2 = 1936$ microphones. Conversely, a given number of microphones would limit the directional resolution up to a maximum frequency.

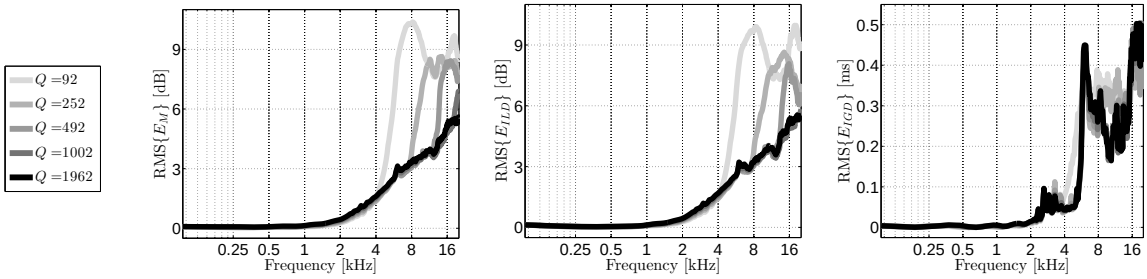


Figure 5.12: Overall accuracy for sources on the sphere calculated using (5.26), for synthesis with $L = 1962$ virtual loudspeakers and different numbers of microphones (Q).

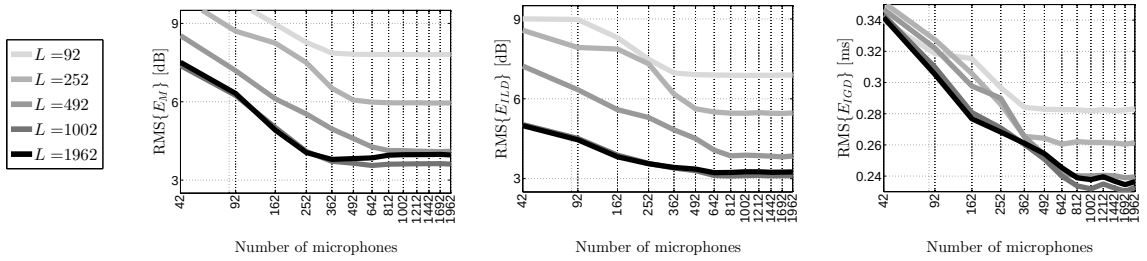


Figure 5.13: Overall accuracy for sources on the sphere calculated using (5.27) up to 16.6 kHz, for synthesis with different numbers of virtual loudspeakers (L).

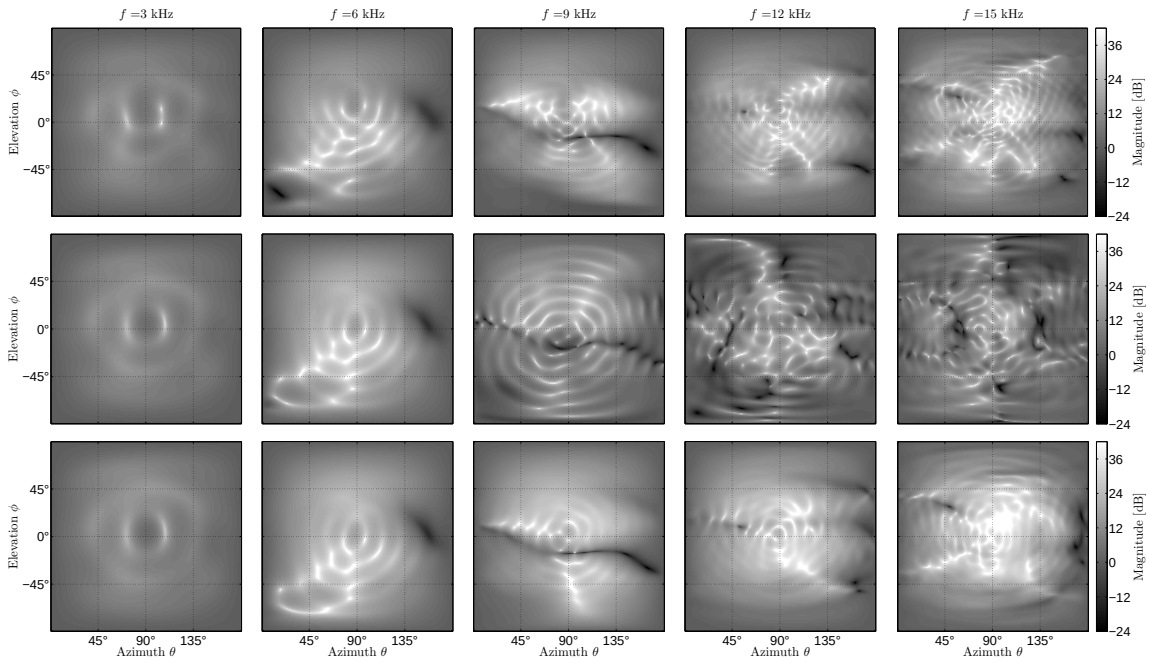
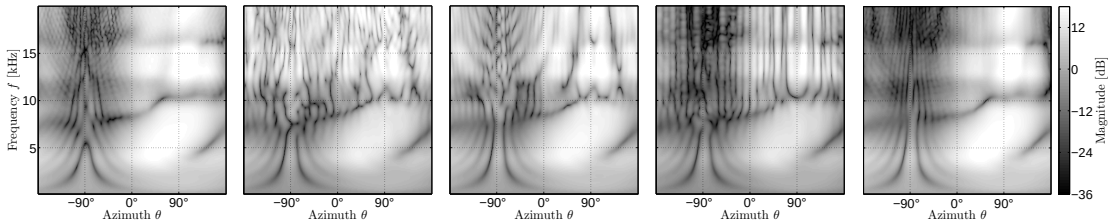
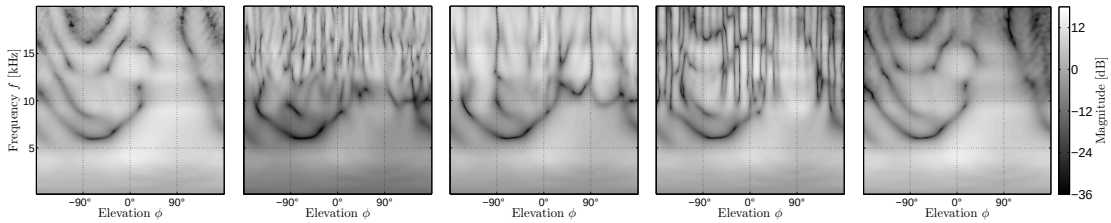


Figure 5.14: Interaural level differences (ILDs) on the sphere. Top row shows the reference ILDs. Middle row shows the ILDs synthesized using $Q = 252$ microphones and $L = 1962$ virtual loudspeakers. Bottom row shows the ILDs synthesized using $Q = 1962$ microphones and $L = 1962$ virtual loudspeakers.



(a) Reference (b) $Q = 252$, $L = 252$ (c) $Q = 252$, $L = 1962$ (d) $Q = 1962$, $L = 252$ (e) $Q = 1962$, $L = 1962$



(f) Reference (g) $Q = 252$, $L = 252$. (h) $Q = 252$, $L = 1962$. (i) $Q = 1962$, $L = 252$. (j) $Q = 1962$, $L = 1962$.

Figure 5.15: Left ear HRTFs on the horizontal (top row) and median (bottom row) planes, synthesized using Q microphones and L virtual loudspeakers.

5.5.2 Objective measures of accuracy

Sets $\hat{B}$ and H were compared giving special attention to the spectral monaural and interaural localization cues. The monaural local error in decibels is defined by [111]:

$$E_{\mathrm{M}}(\Omega_i, f_j) = 20 \log_{10} \left| \frac{\hat{B}_{\mathrm{left}}(\Omega_i, f_j)}{H_{\mathrm{left}}(\Omega_i, f_j)} \right|. \tag{5.21}$$

To highlight the capability of synthesizing the main peaks and notches of the HRTFs, which provide important features for auditory localization, the overall gain mismatch was further removed from (5.21) by subtracting its overall mean value $\bar{E}_{\mathrm{M}}$.

Because interaural information is important in sound localization, as opposed to monoaural phase [123], interaural measures of accuracy were also considered.

The reference interaural HRTFs are defined as [29]

$$H_{\mathrm{interaural}}(\Omega_i, f_j) = \frac{H_{\mathrm{left}}(\Omega_i, f_j)}{H_{\mathrm{right}}(\Omega_i, f_j)}, \tag{5.22}$$

and the synthesized interaural transfer functions as

$$\hat{B}_{\text{interaural}}(\Omega_i, f_j) = \frac{\hat{B}_{\text{left}}(\Omega_i, f_j)}{\hat{B}_{\text{right}}(\Omega_i, f_j)}. \tag{5.23}$$

The interaural level differences (ILDs) corresponding to the reference and synthesized transfer functions are defined by the magnitude in decibels of (5.22) and (5.23), respectively. We calculated the ILD local error correspondingly in decibels by

$$E_{\text{ILD}}(\Omega_i, f_j) = 20 \log_{10} \left| \frac{\hat{B}_{\text{interaural}}(\Omega_i, f_j)}{H_{\text{interaural}}(\Omega_i, f_j)} \right|. \tag{5.24}$$

On the other hand, the interaural phase differences (IPDs) associated with the reference and synthesized transfer functions correspond to the phase in radians of (5.22) and (5.23), respectively. In particular, phase information can be displayed by means of its group delay to highlight spectral information related to peaks and notches with a better resolution [124]. We calculated the interaural group delay (IGD) local error correspondingly in seconds according to

$$E_{\text{IGD}}(\Omega_i, f_j) = \frac{\Delta \arg \left(\frac{\hat{B}_{\text{interaural}}(\Omega_i, f_j)}{H_{\text{interaural}}(\Omega_i, f_j)} \right)}{2\pi \Delta f_j}, \tag{5.25}$$

where arg denotes the unwrapped phase and Δ is the finite difference operator along the discrete variable f_j.

We examined the overall accuracy of our method based on the root mean square (RMS) values of the errors defined in (5.21), (5.24) and (5.25). Norms of accuracy equivalent to the RMS value have been evaluated through listening tests in [111], where the suitability of these norms for predicting audible differences between measured and synthesized HRTFs was verified. We calculated the RMS value along directions based on the following expression:

$$\underset{\Omega_i}{\text{RMS}} \{E\} (f_j) = \left(\sum_{i=1}^{I} E(\Omega_i, f_j)^2 w_i \right)^{\frac{1}{2}}. \tag{5.26}$$

Here, E can be one of the errors E_{M} in (5.21), E_{ILD} in (5.24), or E_{IGD} in (5.25), and w_i are

normalized quadrature weights (area of Voronoi cells) for numerical integration over all of the sound source directions on a sphere with a radius of 1.5 m. Normalization is done in such a way that $\sum_i w_i = 1$, where $w_i > 0$. Similarly, we extended the calculation of the RMS value to cover all directions and frequencies according to:

$$\underset{\Omega_i, f_j}{\mathrm{RMS}}\{E\} = \left(\frac{1}{J} \sum_{i,j=1}^{I,J} E(\Omega_i, f_j)^2 w_i \right)^{\frac{1}{2}}. \tag{5.27}$$

5.5.3 Synthesis on the sphere

We considered $I = 5762$ sound sources almost uniformly distributed on a sphere with a radius of 1.5 m and frequency bins in the full audible range for a sampling frequency of 48 kHz. The results based on (5.26) and (5.27) are displayed in Figs. 5.12 and 5.13, respectively.

In Fig. 5.12, it can be observed that when a number of virtual loudspeakers sufficient to cover the audible frequency range was used, excellent monaural and interaural overall accuracies were obtained up to 2 kHz, even with a limited number of microphones. Over 2 kHz, accuracies of monaural levels and interaural level differences (left and middle panels) gradually decreased with increasing frequency and a decreasing number of microphones. Nevertheless, increasing the number of microphones beyond 1002 did not yield a significant improvement in these overall accuracies. Regarding the interaural group delay (right panel), good performance was also obtained at low frequencies, precisely where these cues are known to be important.

In Fig. 5.13, it can be observed that both monaural and interaural accuracies were also degraded when the number of virtual loudspeakers decreased below 1002. However, increasing this number over 1002 did not improve the overall accuracies. On the other hand, overall accuracies were significantly improved by increasing the number of microphones up to a certain limit, after which adding more microphones did not lead to a decrease in the RMS errors. For a number of virtual loudspeakers greater than 1002, the left panel shows

that increasing the number of microphones beyond 362 did not improve the overall accuracy of the monaural cues. For the same condition, the middle panel shows that the overall accuracy in ILD did not benefit from additional microphones beyond 642. Furthermore, for more than 1002 virtual loudspeakers, the right panel shows that increasing the number of microphones beyond around 1002 did not improve the overall IGD accuracy. The limits were different depending on the type of spectral cue under consideration.

The overall accuracies obtained for 5762 sources can be used to predict, to some extent, the local accuracy when synthesis is performed for denser distributions of sources on the sphere. To exemplify this, the spectral cues for dense datasets and some single frequencies have been synthesized. A qualitative and visual comparison between the reference and synthesized datasets is briefly described below.

Fig. 5.14 shows some examples of reference (top row) and synthesized ILDs, for 360×180 sound sources equiangularly distributed on the sphere. Given the symmetry of the head model, only the left hemisphere is shown. The middle row shows the synthesized ILDs when the array of 252 microphones available in our institute was assumed [74,81,82]. Good accuracies were obtained for the cases of 3 and 6 kHz, as expected from the curve for $Q = 252$ in the center panel of Fig. 5.12, where RMS values below 3 dB are observed up to around 8 kHz. The bottom row shows the synthesized ILDs when an array of 1962 microphones was assumed. In theory, this provides a directional resolution sufficiently high to cover frequencies up to the limit of 16.6 kHz imposed by the head model. In this case, good performance was obtained at 3 and 6 kHz. The performance at 9 kHz was also acceptable. However, local distortions were observed in the ILDs for the cases of 12 and 15 kHz, specially in the lateral region around azimuth $\theta = 90^\circ$ and elevation $\phi = 0^\circ$.

5.5.4 Synthesis on the horizontal and median planes

Figure 5.15 shows the reference and synthesized HRTFs for the left ear and $I = 360$ sound sources equiangularly distributed on the horizontal plane (top row) and the median plane (bottom row). A visual comparison with the reference HRTFs in the first column

shows that transfer functions in the second column, synthesized with $Q = 252$ microphones and $L = 252$ virtual loudspeakers, show fair agreement up to around 10 kHz. However, a detailed inspection of the horizontal plane shows that spectral distortions start to appear at 7.5 kHz on the contralateral side. The third column shows that an increase in the number of virtual loudspeakers up to $L = 1962$, which would be sufficient to cover the audible frequency range, slightly improves the accuracy by smoothing the artifacts at high frequencies, specially on the ipsilateral side and the median plane. However, given that accuracy improves for one side only, there is no guarantee that the interaural accuracy also improves. The fourth column shows, on the other hand, that increasing the number of microphones up to $Q = 1962$ has a better smoothing effect in both the ipsilateral and contralateral sides, as well as on the median plane. Nevertheless, as shown in the last column, 1962 microphones and virtual loudspeakers would be needed to faithfully maintain the monaural spectral cues. In all cases, a smooth and bounded behavior of the synthesized transfer functions along frequency and direction is observed.

Based on the simulation of a head model valid up to 16.6 kHz, the effects of using different numbers of microphones and virtual loudspeakers on binaural synthesis were evaluated. The placement of microphones and virtual loudspeakers was determined by following a sampling of the sphere based on the geometry of an icosahedron. Accuracy was evaluated for dense sets of sound source directions. Overall error metrics for monaural and interaural spectral cues were used.

In general terms, bounded and smooth synthesis of monaural and interaural spectral cues was possible by frequency-dependent order limitation and regularization. However, the performance of the spherical harmonic expansion significantly affected the accuracy in regions where the binaural signals showed rapid variations as a function of frequency and direction. The performance of binaural synthesis based on spherical microphone arrays was found to depend mainly on the number of microphones; this determines the maximum frequency that can be resolved by the system. Nevertheless, our results showed a limit after which increasing the number of microphones does not lead to an improvement in

accuracy. Furthermore, different limits were found depending on the type of spectral cue under consideration.

The number of microphones required to synthesize the interaural cues was higher than that required to synthesize the monaural cues. The reason is that interaural synthesis requires the control of sound pressures at two ears, while monoaural synthesis requires only one ear. However, it is not simple to provide a quantitative relation between the limit to the number of microphones depending on the type of spectral cue under consideration. Such relation would depend on several parameters such as the frequency and position of the sound sources, and the distributions of the microphones. Under the particular conditions for the evaluation considered in this study, the following ordering relations have been found. When the number of virtual loudspeakers was sufficiently large to cover frequencies up to 16.6 kHz, we found that the number of microphones required to improve the overall synthesis accuracy of the interaural level difference cues was higher than the number required to improve the overall synthesis accuracy of the monaural cues. Furthermore, the number of microphones required to accurately synthesize the interaural group delay cues was greater than the number required by the interaural level difference cues.

Further considerations regarding the synthesis of individual features in the HRTFs, as well as perceptual evaluations by means of detectability of differences, and localization tests along azimuth and elevation angles, could provide more insight into the validity of the present results.

5.6 Propagation of noise through the model

Accurate binaural synthesis requires to capture an auditory scene with high spatial resolution. This implies the use of a large number of microphones, together with the complications that arise when controlling a large number of signals.

Predicting the performance of a microphone array in real conditions is a crucial stage for its design. There is a particular necessity of models for predicting the robustness of an

array to the transducer noise, the microphone positioning error, and the effects of space discretization, among other sources of perturbation. These issues have been widely addressed in the theory of beamforming [125], where sensor arrays are used to synthesized spatial patterns. Because binaural synthesis can also be formulated as a beamforming problem, where the spatial patterns are given by the HRTF datasets, beamforming constitutes an adequate framework for investigating the effects of noise in binaural synthesis.

In beamforming, the white noise gain is defined as the output power due to spatially uncorrelated white noise at the sensors [125]. This variable is used as a general measure for robustness to such kind of noise. The kind of arrays examined in conventional beamforming, however, are typically limited to low spatial resolutions [49]. To predict a more precise improvement in the signal-to-noise ratio for higher resolution arrays, additional analyses are required.

In this section, the propagation of noise through a binaural system is investigated based on the gain in signal-to-noise ratio from the input of the microphone array to the output of the binaural system. For this purpose, a linear model of a system in arbitrary geometries is formulated. The model takes into consideration the contributions of additive white noise, which is assumed spatially uncorrelated and with a uniform distribution of energy around the array.

Spherical arrays and spherical HRTF datasets are of interest in binaural synthesis. The model is therefore tested in spherical geometries. Numerical experiments consider models of high resolution arrays. The results can be used as an objective design recommendation to identify the number of microphones that are necessary to synthesize the spectral information of virtual sound sources with a specified accuracy in binaural systems.

5.6.1 White noise gain

The white noise gain of a beamformer is defined as the output power due to spatially uncorrelated, unit variance white noise of at the sensors. The norm squared of the weight

vectors represents the white noise gain [125]:

$$\text{WNG}=\|\mathbf{w}\|^2. \tag{5.28}$$

For simplicity, $\mathbf{w}$ will represent either $\mathbf{w}^{\text{left}}$ or $\mathbf{w}^{\text{right}}$ in all of what follows. The norm is defined by

$$\|\mathbf{u}\| = \left(\mathbf{u}^{\dagger}\mathbf{u}\right)^{\frac{1}{2}} = \left(\sum_{q=1}^{Q} |u_q|^2\right)^{\frac{1}{2}} < \infty. \tag{5.29}$$

where $\mathbf{u}$ is an arbitrary vector of size $Q \times 1$. This norm is used throughout this section.

If the white noise gain is large, it is expected a poor signal-to-noise ratio at the output of the beamformer due to white noise contributions. The inverse of the white noise gain, WNG^{-1}, should therefore be as high as possible. This variable is used as a general measure for robustness. However, to predict a more precise improvement in the signal-to-noise ratio, an additional analysis is presented below.

5.6.2 Gain in signal-to-noise ratio

The propagation of noise through the synthesis model in (5.12) can be calculated based on the gain in signal-to-noise ratio from the input to the output:

$$G^{\text{SNR}} = \frac{\text{SNR}^{\text{output}}}{\text{SNR}^{\text{input}}}. \tag{5.30}$$

This quantity is expected to be greater than unit and as high as possible. To calculate G^{SNR}, a signal model with noise is considered below.

The vector of microphone array recordings affected by additive, spatially uncorrelated white noise is modeled as follows:

$$\mathbf{p} = \mathbf{s} + \boldsymbol{\nu}, \tag{5.31}$$

where the vector $\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \cdots & s_Q \end{bmatrix}^{\mathsf{T}}$ contains the microphone signals without noise, and the vector $\boldsymbol{\nu} = \begin{bmatrix} \nu_1 & \nu_2 & \cdots & \nu_Q \end{bmatrix}^{\mathsf{T}}$ contains the noise only. For this model, binaural

synthesis results in

$$\begin{aligned}\hat{\mathbf{b}} &= \mathbf{w}^{\dagger}\left[\mathbf{s}+\boldsymbol{\nu}\right], \\ &= \underbrace{\mathbf{w}^{\dagger}\mathbf{s}}_{\hat{\mathbf{b}}^{(s)}} + \underbrace{\mathbf{w}^{\dagger}\boldsymbol{\nu}}_{\hat{\mathbf{b}}^{(\nu)}},\end{aligned} \tag{5.32}$$

where $\hat{\mathbf{b}}^{(s)}$ denote the binaural signals from recordings without noise, and $\hat{\mathbf{b}}^{(\nu)}$ represents the binaural signals due to noise only.

The multichannel signal-to-noise ratio at the input can be defined by

$$\mathrm{SNR}^{\mathrm{input}} := \frac{\|\mathbf{s}\|^2}{\|\boldsymbol{\nu}\|^2} \tag{5.33}$$

with the norm in (5.29). The signal-to-noise ratio at the output, for the left or right ear, can also be defined in consistency with (5.29) but for the case of a single-channel signal:

$$\mathrm{SNR}^{\mathrm{output}} := \frac{|\hat{b}^{(s)}|^2}{|\hat{b}^{(\nu)}|^2} = \frac{|\mathbf{w}^{\dagger}\mathbf{s}|^2}{|\mathbf{w}^{\dagger}\boldsymbol{\nu}|^2}. \tag{5.34}$$

When $\boldsymbol{\nu}$ has a uniform distribution of energy,

$$\boldsymbol{\nu} = \nu_0 \cdot \boldsymbol{\varphi}, \tag{5.35}$$

where $\boldsymbol{\varphi} = \begin{bmatrix} e^{j\varphi_1} & e^{j\varphi_2} & \dots & e^{j\varphi_Q} \end{bmatrix}^{\mathsf{T}}$ is a vector with random phases φ_q, (5.30) results in

$$G^{\mathrm{SNR}} = \frac{\frac{|\mathbf{w}^{\dagger}\mathbf{s}|^2}{|\nu_0|^2|\mathbf{w}^{\dagger}\boldsymbol{\varphi}|^2}}{\frac{\|\mathbf{s}\|^2}{|\nu_0|^2\|\boldsymbol{\varphi}\|^2}} = \frac{|\mathbf{w}^{\dagger}\mathbf{s}|^2}{|\mathbf{w}^{\dagger}\boldsymbol{\varphi}|^2\frac{\|\mathbf{s}\|^2}{\|\boldsymbol{\varphi}\|^2}}. \tag{5.36}$$

By virtue of the Cauchy-Schwarz inequality for the norm in (5.29), the left factor in the denominator of (5.36) has the following upper bound:

$$|\mathbf{w}^{\dagger}\boldsymbol{\varphi}|^2 \leq \|\mathbf{w}\|^2\|\boldsymbol{\varphi}\|^2. \tag{5.37}$$

Equating (5.37) and (5.36), it is shown that

$$
\begin{aligned}
G^{\mathrm{SNR}} &\geq \frac{|\mathbf{w}^{\dagger}\mathbf{s}|^2}{\|\mathbf{w}\|^2\|\mathbf{s}\|^2}, \\
&\geq \frac{|\sum_{\ell,q} h_{\ell}^{*}\kappa_{\ell q}s_q|^2}{\left(\sum_q|\sum_{\ell} h_{\ell}^{*}\kappa_{\ell q}|^2\right)\left(\sum_q |s_q|^2\right)},
\end{aligned}
\tag{5.38}
$$

where * denotes complex conjugate. When the recording signals are affected by additive noise that is spatially uncorrelated and has a uniform distribution of energy, the right side of (5.38) provides a theoretical lower bound for the propagation of this kind of noise through the synthesis model in (5.12).

To analyze the performance of (5.28) and (5.38) on predicting improvement in signal-to-noise ratio, a physical model for $\mathbf{K}$ would be useful. We describe a model for binaural synthesis in spherical geometries.

5.6.3 Evaluation of robustness to noise

Figure 5.16 shows the predictions of robustness obtained with WNG^{-1} from (5.28), and with the proposed lower bound for G^{SNR} in (5.38). The binaural beamformer was modeled by using (5.19) with $\lambda = 1\times10^{-3}$. The minimum of G^{SNR} was obtained from the simulation of 5000 sources randomly distributed around the array at 1.5 m distance. The microphones and virtual loudspeakers were distributed using spherical grids based on the geometry of an icosahedron.

The top panel shows the results for WNG^{-1}, where it can be observed that increasing the number of microphones improved the expected robustness at lower and middle frequencies. However, when the number of microphones increased, its contribution to robustness only increased slightly. This was specially true at higher frequencies.

The bottom panel shows the results for G^{SNR}. At lower frequencies up to around 2 kHz, the results confirmed the tendency predicted by WNG^{-1}. Nevertheless, the results showed that adding more microphones to the system does not necessarily imply an improvement in robustness at higher frequencies.

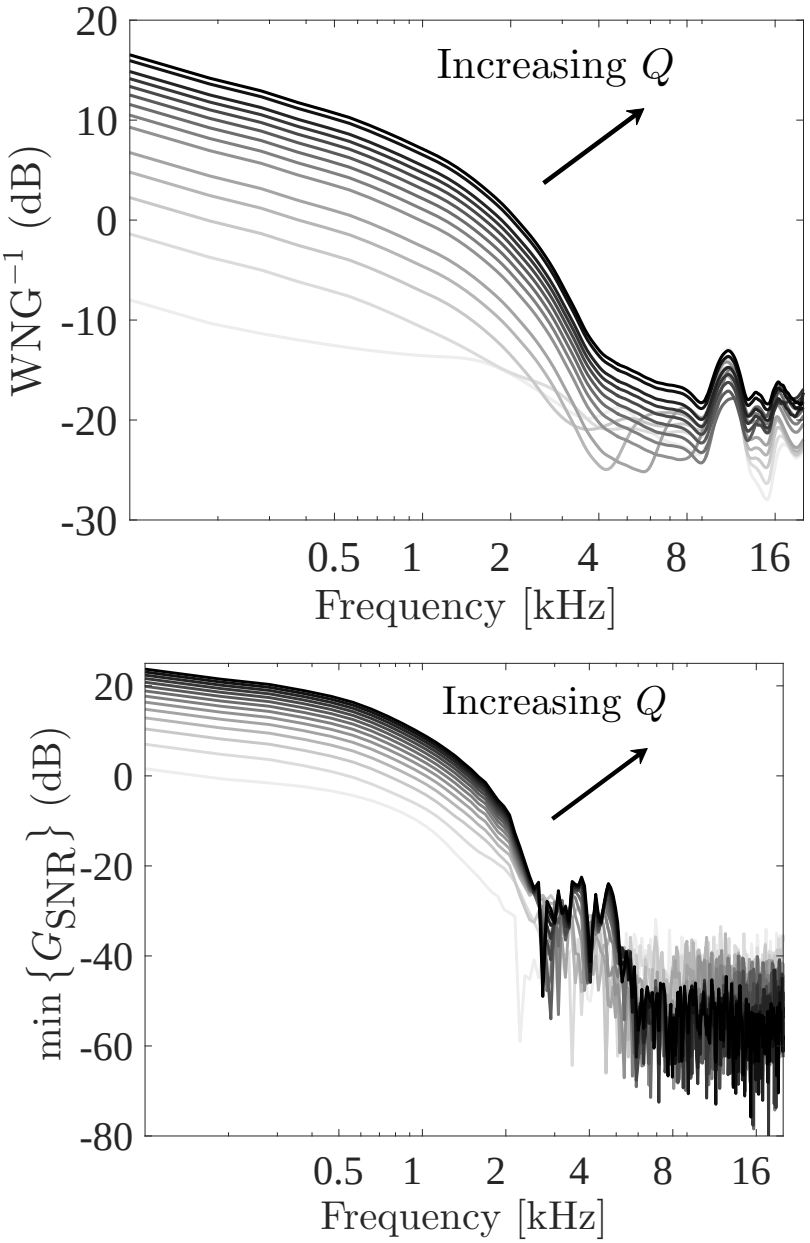


Figure 5.16: Evaluation of inverse of white noise gain (WNG^{-1}) and gain in signal-to-noise ratio (G^{SNR}) for a binaural beamformer in spherical geometries. Results were obtained for a = 8.5 cm, b = 1.5 m, L = 1962 virtual loudspeakers, and different numbers of microphones (Q =12, 42, 92, 162, 252, 362, 492, 642, 812, 1002, 1212, 1442, 1692, 1962.).

5.7 Summary

A general binaural synthesis model for far and near distances has been introduced. The model can combine and edit distinct types of spatial sound information on the sphere, such as microphone array recordings, acoustic transfer functions, and angular windows.

A linear model for binaural systems in arbitrary geometries was also formulated. The model takes into consideration the contributions of additive white noise, which is assumed spatially uncorrelated and with a uniform distribution of energy around the array. The propagation of such kind of noise through the model was investigated based on two predictors of robustness: 1) the inverse of white noise gain used in beamforming, and 2) a proposed lower bound for the gain in signal-to-noise ratio.

Numerical experiments considering a binaural system in spherical geometries showed that similar predictions at lower frequencies can be obtained with the white noise gain and the gain in signal-to-noise ratio. However, results at higher frequencies showed that the white noise gain might not be sufficient to predict robustness in this region. In this regard, the estimates obtained with the lower bound to the gain in signal to noise ratio predicted much lower levels of robustness at higher frequencies.

Additional experiments based on a more precise model of noise, accompanied with objective validations in real world conditions, could give more insight into the findings reported in this work.

Chapter 6

Conclusions

Stable spherical acoustic models at far and near distances were presented in this thesis. Specifically, the models are used for:

- Chapter 3: Synthesis of HRTFs
- Chapter 4: Sound field recording with spherical arrays
- Chapter 5: Generalized binaural synthesis

In chapter 3, distance-varying filters to synthesize HRTFs at arbitrary distances from a dataset of HRTFs at a single distance around the listener were presented and evaluated.

When synthesis was limited to the horizontal plane, a new set of distance-varying filters was derived from the solution of the acoustic wave equation for sound fields that are assumed to be invariant with respect to elevation angles in spherical coordinates. Results were free of the lateral side discontinuities that appear when sound fields are assumed to be invariant along polar angles in interaural coordinates. Moreover, to properly account for HRTF datasets obtained for discrete distributions that limit the angular bandwidth, we introduced a new magnitude-dependent band-limiting threshold (MBT) to restrict the action of distance-varying filters to limited angular bandwidths.

The performance was also evaluated numerically using three models of a human head valid up to 16.6 kHz. Results showed that overall accuracies obtained with the proposed distance-varying filters in spherical coordinates outperform the existing distance-varying filters in interaural coordinates. Furthermore, overall accuracies obtained with the proposed MBT outperformed overall accuracies achieved with the traditional frequency-dependent threshold, especially at low frequencies and distances close to the head.

A perceptual evaluation of the distance-varying filters by means of detectability of differences, and localization tests along azimuth and distances, could provide more insight into the validity of the results obtained in chapter 3.

In chapter 4, boundary matching filters (BMFs) for obtaining free-field representations of sound fields in the vicinity or outside a rigid spherical microphone array were presented. By including a hypothetical sphere for reproduction, a new set of BMFs was proposed to cope with the excessive amplification of existing encoding filters. The principle of boundary matching filtering might also be applied to other simple geometries.

In chapter 5, a general binaural synthesis model for far and near distances has been introduced. The model can combine and edit distinct types of spatial sound information on the sphere, such as microphone array recordings, acoustic transfer functions, and angular windows.

A linear model for binaural systems in arbitrary geometries was also formulated. The model takes into consideration the contributions of additive white noise, which is assumed spatially uncorrelated and with a uniform distribution of energy around the array. The propagation of such kind of noise through the model was investigated based on two predictors of robustness: 1) the inverse of white noise gain used in beamforming, and 2) a proposed lower bound for the gain in signal-to-noise ratio.

Numerical experiments considering a binaural system in spherical geometries showed that similar predictions at lower frequencies can be obtained with the white noise gain and the gain in signal-to-noise ratio. However, results at higher frequencies showed that the white noise gain might not be sufficient to predict robustness in this region. In this

regard, the estimates obtained with the lower bound to the gain in signal to noise ratio predicted much lower levels of robustness at higher frequencies. Additional experiments based on a more precise model of noise, accompanied with objective validations in real world conditions, could give more insight into the findings reported in chapter 5.

Bibliography

[1] A. Steed and R. Schroeder, "Collaboration in immersive and non-immersive virtual environments," in *Immersed in Media: Telepresence Theory, Measurement & Technology* (M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, eds.), pp. 263–282, Cham: Springer International Publishing, 2015.

[2] K. Lee, "Augmented Reality in Education and Training," *TechTrends*, vol. 56, no. 2, pp. 13–21, 2012.

[3] G. Riva, C. Botella, R. Baños, F. Mantovani, A. García-Palacios, S. Quero, S. Serino, S. Triberti, C. Repetto, A. Dakanalis, D. Villani, and A. Gaggioli, "Presence-inducing media for mental health applications," in *Immersed in Media: Telepresence Theory, Measurement & Technology* (M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, eds.), pp. 283–332, Cham: Springer International Publishing, 2015.

[4] M. Lombard and M. T. Jones, "Defining presence," in *Immersed in Media: Telepresence Theory, Measurement & Technology* (M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, eds.), pp. 13–34, Cham: Springer International Publishing, 2015.

[5] J. Freeman, "Implications for the measurement of presence from convergent evidence on the structure of presence," in *Information Systems Division at the annual meeting of the International Communication Association, New Orleans, LA*, 2004.

[6] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence: Teleop. Virt. Env.*, vol. 5, pp. 290–301, Jan. 1996.

[7] M. Kobayashi, K. Ueno, and S. Ise, "The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study," *Presence: Teleop. Virt. Env.*, vol. 24, pp. 163–174, May 2015.

[8] D. Västfjäll, "The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments," *Cyberpsychol. Behav.*, vol. 6, no. 2, pp. 181–188, 2003.

[9] Y. Suzuki, T. Okamoto, J. Treviño, Z.-L. Cui, Y. Iwaya, S. Sakamoto, and M. Otani, "3d spatial sound systems compatible with human's active listening to realize rich high-level *kansei* information," *Interdiscipl. Inform. Sci.*, vol. 18, no. 2, pp. 71–82, 2012.

[10] B. E. Riecke, D. Feuereissen, J. J. Rieser, and T. P. McNamara, "Spatialized sound enhances biomechanically-induced self-motion Illusion (vection)," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (New York, NY, USA), pp. 2799–2802, ACM, 2011.

[11] W. Teramoto, S. Sakamoto, F. Furune, J. Gyoba, and Y. Suzuki, "Compression of auditory space during forward self-motion," *PLoS ONE*, vol. 7, p. e39402, June 2012.

[12] B. Keshavarz, L. J. Hettinger, D. Vena, and J. L. Campos, "Combined effects of auditory and visual cues on the perception of vection," *Exp. Brain. Res.*, vol. 232, no. 3, pp. 827–836, 2014.

[13] B. E. Riecke and J. Schulte-Pelkum, "An integrative approach to presence and self-motion perception research," in *Immersed in Media: Telepresence Theory, Measurement & Technology* (M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, eds.), pp. 187–235, Cham: Springer International Publishing, 2015.

[14] T. Nasir and J. C. Roberts, "Sonification of spatial data," in *Proc. ICAD*, (Montreal, Quebec, Canada), pp. 112–119, June 2007.

[15] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "Navigation System for the Blind: Auditory Display Modes and Guidance," *Presence: Teleop. Virt. Env.*, vol. 7, pp. 193–203, Apr. 1998.

[16] S. Holland, R. D. Morse, and H. Gedenryd, "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface," *Pers. Ubiquit. Comput.*, vol. 6, no. 4, pp. 253–259, 2002.

[17] M. Geronazzo, A. Bedin, L. Brayda, C. Campus, and F. Avanzini, "Interactive spatial sonification for non-visual exploration of virtual maps," *Int. J. Hum.-Comput. St.*, vol. 85, pp. 4 – 15, 2016. Data Sonification and Sound Design in Interactive Systems.

[18] A. Berkhout, "A holographic approach to acoustic control," *J. Audio Eng. Soc.*, vol. 36, pp. 977–995, Dec. 1988.

[19] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *23rd Int. Conf. Audio Eng. Soc.*, May 2003.

[20] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust. Speech & Sig. Proc.*, vol. 36, pp. 145–152, Feb. 1988.

[21] M. Kolundzija, C. Faller, and M. Vetterli, "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion," *J. Audio Eng. Soc.*, vol. 59, no. 10, pp. 721–734, 2011.

[22] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *Acta Acust. united Ac.*, vol. 85, pp. 78–87, 1999.

[23] E. Corteel, "Equalization in an Extended Area Using Multichannel Inversion and Wave Field Synthesis," *J. Audio Eng. Soc.*, vol. 54, no. 12, pp. 1140–1161, 2006.

[24] D. G. Malham, "Toward reality equivalence in spatial sound diffusion," *Comp. Mus. J.*, vol. 25, no. 4, pp. 31–38, 2001.

[25] F. A. Firestone, "The phase difference and amplitude ratio at the ears due to a source of pure tone," *J. Acoust. Soc. Am.*, vol. 2, no. 2, pp. 260–270, 1930.

[26] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, vol. 1, no. 3, pp. 167–174, 1980.

[27] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone, "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," *J. Acoust. Soc. Jpn. (J)*, vol. 45, pp. 756–766, 1989. (in Japanese), English translation: J. Acoust. Soc. Jpn. (E), vol. 12, pp. 203-216, 1991.

[28] H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, no. 3-4, pp. 171–218, 1992.

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

[30] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3–19, 1989.

[31] C. Han, T. Okamoto, Y. Iwaya, and Y. Suzuki, "Loudspeaker distributions suitable for crosstalk cancellers robust to head rotation," *Acoust. Sci. Technol.*, vol. 33, no. 4, pp. 266–269, 2012.

[32] V. Algazi and R. Duda, "Headphone-Based Spatial Sound," *IEEE Signal Process. Mag.*, vol. 28, pp. 33–42, Jan. 2011.

[33] M. Morimoto, Y. Ando, and Z. Maekawa, "On head-related transfer function in distance perception," in *Proc. Congress Acoust. Soc. Jpn.*, (Japan), pp. 137–138, 1975. (in Japanese).

[34] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, pp. 1465–1479, Sept. 1999.

[35] H.-Y. Kim, Y. Suzuki, S. Takane, and T. Sone, "Control of auditory distance perception based on the auditory parallax model," *Appl. Acoust.*, vol. 62, pp. 245–270, Mar. 2001.

[36] D. Brungart, "Auditory parallax effects in the HRTF for nearby sources," in *Proc. IEEE WASPAA*, pp. 171–174, 1999.

[37] D. S. Brungart, "Near-Field Virtual Audio Displays," *Presence: Teleop. Virt. Env.*, vol. 11, pp. 93–106, Feb. 2002.

[38] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoust. Sci. Technol.*, vol. 35, no. 3, pp. 159–165, 2014.

[39] M. Morimoto, N. Johren, Y. Ando, and Z. Maekawa, "On head-related transfer functions," *Transaction on Technical Committee of Psychological and Physiological Acoustics*, vol. H-31-1-2, 1976. (in Japanese).

[40] H. Wallach, "On sound localization," *J. Acoust. Soc. Am.*, vol. 10, no. 4, pp. 270–274, 1939.

[41] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *J. Acoust. Soc. Am.*, vol. 42, no. 2, pp. 480–488, 1967.

[42] Y. Iwaya, Y. Suzuki, and D. Kimura, "Effects of head movement on front-back error in sound localization," *Acoust. Sci. Technol.*, vol. 24, no. 5, pp. 322–324, 2003.

[43] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, vol. 119, pp. 2589–2598, May 2006.

[44] I. Toshima, H. Uematsu, and T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement: TeleHead," *Acoust. Sci. Technol.*, vol. 24, no. 5, pp. 327–329, 2003.

[45] I. Toshima, S. Aoki, and T. Hirahara, "Sound localization using an acoustical telepresence robot: TeleHead II," *Presence: Teleoperators and Virtual Environments*, vol. 17, pp. 392–404, July 2008.

[46] V. R. Algazi, R. O. Duda, and D. M. Thompson, "Motion-tracked binaural sound," *J. Audio Eng. Soc.*, vol. 52, no. 11, pp. 1142–1156, 2004.

[47] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE ICASSP*, vol. II, (Orlando, FL, USA), pp. 1781–1784, May 2002.

[48] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, pp. 2149–2157, Oct. 2004.

[49] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech, Audio Process.*, vol. 13, pp. 135–143, Jan. 2005.

[50] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Trans. Signal Process.*, vol. 55, pp. 1003–1010, Mar. 2007.

[51] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, pp. 702–714, Feb. 2007.

[52] E. Fisher and B. Rafaely, "The nearfield spherical microphone array," in *Proc. IEEE ICASSP*, pp. 5272–5275, Mar. 2008.

[53] E. Fisher and B. Rafaely, "Near-field spherical microphone array processing with radial filtering," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, pp. 256–265, Feb. 2011.

[54] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Trans. Aud. Speech & Lang. Proc.*, vol. 18, pp. 2–6, Jan. 2010.

[55] S. Yan, H. Sun, U. Svensson, X. Ma, and J. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, pp. 361–371, Feb. 2011.

[56] T. Rettberg and S. Spors, "Time-domain behaviour of spherical microphone arrays at high orders," in *Proc. DAGA*, 2014.

[57] S. Lösler and F. Zotter, "Comprehensive radial filter design for practical higher-order Ambisonic recording," in *Proc. DAGA*, 2015.

[58] D. Alon and B. Rafaely, "Beamforming with Optimal Aliasing Cancellation in Spherical Microphone Arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, pp. 196–210, Jan. 2016.

[59] 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 *AES 119*, (New York, USA), Oct. 2005.

[60] W. Song, W. Ellermeier, and J. Hald, "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *J. Acoust. Soc. Am.*, vol. 123, pp. 910–924, Feb. 2008.

[61] W. Song, W. Ellermeier, and J. Hald, "Psychoacoustic evaluation of multichannel reproduced sounds using binaural synthesis and spherical beamforming," *J. Acoust. Soc. Am.*, vol. 130, no. 4, pp. 2063–2075, 2011.

[62] H. Sun, S. Yan, and U. P. Svensson, "Optimal higher order ambisonics encoding with predefined constraints," *IEEE Trans. Aud. Speech & Lang. Proc.*, vol. 20, pp. 742–754, Mar. 2012.

[63] C. D. Salvador, S. Sakamoto, J. Treviño, J. Li, Y. Yan, and Y. Suzuki, "Accuracy of head-related transfer functions synthesized with spherical microphone arrays," *Proc. Mtgs. Acoust.*, vol. 19, Apr. 2013.

[64] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *J. Acoust. Soc. Am.*, vol. 133, pp. 2711–2721, May 2013.

[65] N. R. Shabtai and B. Rafaely, "Generalized spherical array beamforming for binaural speech reproduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, pp. 238–247, Jan. 2014.

[66] N. R. Shabtai, "Optimization of the directivity in binaural sound reproduction beamforming," *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 3118–3128, 2015.

[67] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.*, vol. 118, pp. 3094–3103, Nov. 2005.

[68] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 261–270, 2012.

[69] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 2828–2840, 2012.

[70] S. Tervo and A. Politis, "Direction of Arrival Estimation of Reflections from Room Impulse Responses Using a Spherical Microphone Array," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, pp. 1539–1551, Oct. 2015.

[71] H. Morgenstern, B. Rafaely, and F. Zotter, "Theory and investigation of acoustic multiple-input multiple-output systems based on spherical arrays in a room," *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 2998–3009, 2015.

[72] F. M. Fazi and P. A. Nelson, "Nonuniqueness of the solution of the sound field reproduction problem with boundary pressure control," *Acta Acust. United Ac.*, vol. 98, pp. 1–14, Jan. 2012.

[73] E. Rasumow, M. Blau, S. Doclo, M. Hansen, S. Van de Par, D. Püschel, and V. Mellert, "Least squares versus non-linear cost functions for a vitual artificial head," *Proc. Mtgs. Acoust.*, vol. 19, no. 1, 2013.

[74] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya, and Y. Suzuki, "Sound-space recording and binaural presentation system based on a 252-channel microphone array," *Acoust. Sci. Technol.*, vol. 36, no. 6, pp. 516–526, 2015.

[75] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, "Binaural reproduction of plane waves with reduced modal order," *Acta Acust. United Ac.*, vol. 100, pp. 972–983, Sept. 2014.

[76] J. Sheaffer, M. van Walstijn, B. Rafaely, and K. Kowalczyk, "Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, pp. 2125–2135, Dec. 2015.

[77] D. L. Alon, J. Sheaffer, and B. Rafaely, "Plane-wave decomposition with aliasing cancellation for binaural sound reproduction," in *Audio Engineering Society Convention 139*, Oct. 2015.

[78] M. V. Laitinen and V. Pulkki, "Binaural reproduction for Directional Audio Coding," in *Proc. IEEE WASPAA*, (Espoo, Finland), pp. 337–340, Oct. 2009.

[79] S. Delikaris-Manias, J. Vilkamo, and V. Pulkki, "Parametric binaural rendering utilizing compact microphone arrays," in *Proc. IEEE ICASSP*, pp. 629–633, Apr. 2015.

[80] E. Rasumow, M. Hansen, S. v. d. Par, D. Püschel, V. Mellert, S. Doclo, and M. Blau, "Regularization approaches for synthesizing HRTF directivity patterns," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, pp. 215–225, Feb. 2016.

[81] 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 *Proc. 2nd Int. Symp. Universal Comm.*, pp. 429–434, 2008.

[82] S. Sakamoto, S. Hongo, and Y. Suzuki, "3d sound-space sensing method based on numerous symmetrically arranged microphones," *IEICE Trans. Fundamentals*, vol. E97-A, pp. 1893–1901, Sept. 2014.

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

[84] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *AES 124*, (Amsterdam, The Netherlands), May 2008.

[85] S. Spors and J. Ahrens, "Towards a theory for arbitrarily shaped sound field reproduction systems," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3930–3930, 2008.

[86] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, vol. 93. Springer Science & Business Media, 2012.

[87] F. M. Fazi and P. A. Nelson, "Sound field reproduction as an equivalent acoustical scattering problem," *J. Acoust. Soc. Am.*, vol. 134, no. 5, pp. 3721–3729, 2013.

[88] J. A. Hargreaves and Y. W. Lam, "An energy interpretation of the Kirchhoff-Helmholtz boundary integral equation and its application to sound field synthesis," *Acta Acust. United Ac.*, vol. 100, pp. 912–920, Sept. 2014.

[89] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Sound field analysis along a circle and its applications to HRTF interpolation," *J. Audio Eng. Soc.*, vol. 56, no. 3, pp. 156–175, 2008.

[90] M. Thomas, J. Ahrens, and I. Tashev, "A method for converting between cylindrical and spherical harmonic representations of sound fields," in *Proc. IEEE ICASSP*, pp. 4723–4727, May 2014.

[91] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *Proc. IEEE ICASSP*, vol. 4, pp. 45–48, May 2004.

[92] W. Zhang, T. Abhayapala, and R. A. Kennedy, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *J. Acoust. Soc. Am.*, vol. 127, pp. 2347–2357, Apr. 2010.

[93] 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 Acust. United Ac.*, vol. 98, pp. 72–82, Jan. 2012.

[94] N. A. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions*. Elsevier Series in Electromagnetism, Maryland, USA: Elsevier, 2004.

[95] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, and D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *J. Acoust. Soc. Am.*, vol. 127, pp. 370–386, Jan. 2010.

[96] J. Ahrens and S. Spors, "A modal analysis of spatial discretization of spherical loudspeaker distributions used for sound field synthesis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, pp. 2564–2574, Nov. 2012.

[97] G. Enzner, M. Krawczyk, F.-M. Hoffmann, and M. Weinert, "3d reconstruction of HRTF-fields from 1d continuous measurements," in *Proc. IEEE WASPAA*, pp. 157–160, Oct. 2011.

[98] S. Koyama, K. Furuya, K. Wakayama, S. Shimauchi, and H. Saruwatari, "Analytical approach to transforming filter design for sound field recording and reproduction using circular arrays with a spherical baffle," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1024–1036, 2016.

[99] J. R. Driscoll and D. M. Heally, "Computing Fourier transforms and convolutions on the 2-sphere," *Adv. Appl. Math.*, vol. 15, pp. 202–250, 1994.

[100] D. Healy, Jr., D. Rockmore, P. Kostelec, and S. Moore, "FFTs for the 2-Sphere-Improvements and Variations," *J. Fourier Anal. Appl.*, vol. 9, pp. 341–385, July 2003.

[101] P. N. Aronszajn, "La théorie des noyaux reproduisants et ses applications Première Partie," *Math. Proc. Cambridge Philos. Soc.*, vol. 39, no. 03, pp. 133–153, 1943.

[102] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[103] F. J. Simons, F. A. Dahlen, and M. A. Wieczorek, "Spatiospectral concentration on a sphere," *SIAM Rev.*, vol. 48, no. 3, pp. 504–536, 2006.

[104] R. A. Kennedy, P. Sadeghi, Z. Khalid, and J. D. McEwen, "Classification and construction of closed-form kernels for signal representation on the 2-sphere," in *SPIE Proceedings*, vol. 8858, (San Diego, California, USA), pp. 88580M–88580M, International Society for Optics and Photonics, Sept. 2013.

[105] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *J. Acoust. Soc. Am.*, vol. 120, pp. 2202–2215, Oct. 2006.

[106] S. Smale, "Mathematical problems for the next century," *Math. Intell.*, vol. 20, no. 2, pp. 7–15, 1998.

[107] A. Bates, Z. Khalid, and R. Kennedy, "Novel sampling scheme on the sphere for head-related transfer function measurements," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, pp. 1068–1081, June 2015.

[108] J.-M. Lee, J.-W. Choi, and Y.-H. Kim, "Reproduction of a higher-order circular harmonic field using a linear array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 137, no. 3, pp. EL227–EL233, 2015.

[109] S. Spors and J. Ahrens, "Efficient range extrapolation of head-related impulse responses by wave field synthesis techniques," in *Proc. IEEE ICASSP*, pp. 49–52, May 2011.

[110] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *J. Acoust. Soc. Am.*, vol. 125, pp. 2233–2242, Apr. 2009.

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

[112] M. Shin, P. A. Nelson, F. M. Fazi, and J. Seo, "Velocity controlled sound field reproduction by non-uniformly spaced loudspeakers," *J. Sound Vib.*, vol. 370, pp. 444 – 464, 2016.

[113] P. A. Nelson, "Active control of acoustic fields and the reproduction of sound," *J. Sound Vib.*, vol. 177, no. 4, pp. 447 – 477, 1994.

[114] S. H. Schot, "Eighty years of Sommerfeld's radiation condition," *Hist. Math.*, vol. 19, no. 4, pp. 385 – 401, 1992.

[115] E. Fernandez-Grande, "Sound field reconstruction using a spherical microphone array," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1168–1178, 2016.

[116] J. Sheaffer, S. Villeval, and B. Rafaely, "Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction," in *Proc. of the EAA Joint Symp. Auralization and Ambisonics*, (Berlin, Germany), Mar. 2014.

[117] D. A. Russell, J. P. Titlow, and Y.-J. Bemmen, "Acoustic monopoles, dipoles, and quadrupoles: An experiment revisited," *Am. J. Phys.*, vol. 67, no. 8, pp. 660–664, 1999.

[118] G. H. Koopmann, L. Song, and J. B. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," *J. Acoust. Soc. Am.*, vol. 86, no. 6, pp. 2433–2438, 1989.

[119] L. G. Copley, "Fundamental Results Concerning Integral Representations in Acoustic Radiation," *J. Acoust. Soc. Am.*, vol. 44, no. 1, pp. 28–32, 1968.

[120] V. A. Morozov, *Regularization Methods for Ill-posed problems*. CRC Press, 1993.

[121] G. Romigh, D. Brungart, R. Stern, and B. Simpson, "Efficient Real Spherical Harmonic Representation of Head-Related Transfer Functions," *IEEE J. Sel. Topics Signal Process.*, vol. 9, pp. 921–930, Aug. 2015.

[122] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, pp. 3048–3058, Nov. 1998.

[123] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2821–2840, 1999.

[124] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 364–374, 2005.

[125] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP*, vol. 5, pp. 4–24, Apr. 1988. This Magazine ceased production in 1990. The current retitled publication is IEEE Signal Processing Magazine.

List of publications

Journal papers

1. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values,” accepted, Acoust. Sci. & Tech.

2. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Spatial accuracy of binaural synthesis from rigid spherical microphone array recordings,” accepted, Acoust. Sci. & Tech.

International conferences

3. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Numerical evaluation of binaural synthesis from rigid spherical microphone array recordings,” in AES International Conference on Headphone Technology, Aalborg, Denmark, August 2016.

4. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Analytic filters for sound field recording with a rigid spherical microphone array,” in RIEC International Symposium on Ultra Realistic Interactive Acoustic Communications (ISURAC 2016), Sendai, Japan, May 2016.

5. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Embedding distance information in binaural renderings of far field recordings,” in Proc. EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 2014, pp. 133–139.

6. C. Salvador, S. Sakamoto, J. Trevino, J. Li, Y. Yan, and Y. Suzuki, "Accuracy of head-related transfer functions synthesized with spherical microphone arrays," in Proc. Mtgs. Acoust., vol. 19, no. 1, April 2013.

National conferences

7. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Evaluation of white noise gain in a binaural system for microphone arrays,” in Autumn Meeting of the Acoustical Society of Japan, Toyama, Japan, September 2016.

8. C. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “A new signal processing procedure for stable distance manipulation of circular HRTFs on the horizontal plane,” in Spring Meeting of the Acoustical Society of Japan, Yokohama, Japan, April 2016.

9. C. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, “Editing distance information in compact microphone array recordings for its binaural rendering,” IEICE technical report, vol. 114, no. 3, pp. 13–18, April 2014.

10. C. 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 Spring Meeting of the Acoustical Society of Japan, Tokyo, Japan, April 2013.