A Formulation of Ambisonics in Unconventional Geometries

非球形のアレイ形状を持つAmbisonicsの定式化

by
Jorge Alberto Treviño López

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy (Information Sciences)
in Tohoku University
March 2014

Evaluation Committee:
Prof. SUZUKI Yôiti
Prof. SUGANUMA Takuo
Prof. ITO Akinori (Graduate School of Engineering)
Assoc. Prof. SAKAMOTO Shuichi

PREFACE

The human auditory system shows remarkable capabilities. We can perceive sound over a wide range of frequencies and levels. It is easy for us to concentrate on one particular sound, say, someone's voice, while ignoring other sounds that may be present. We can quickly and accurately identify the position from where sound is reaching us. These tasks are so natural to us that they appear to be trivial; however, building a machine with the capabilities of the human auditory system is still out of reach for modern technology.

Technology, on the other hand, can enhance our lives by providing new ways of listening. Two classical inventions that achieved this are the phonograph and the telephone. After their introduction, we were no longer limited to listening to the sounds of things as they take place around us. Nowadays, we can easily communicate with people all over the world or listen to some audio content, such as music, whenever we like.

Telecommunication and sound reproduction systems have been greatly improved since the early telephones and phonographs. It is still difficult to accurately measure and later reproduce sounds over the whole listening range in terms of frequency and level. However, high-quality systems are good enough for most listeners to be unable to notice any significant degradation. Despite this, there has been one aspect of sound which technology has been unable to incorporate into these communication and reproduction systems. The spatial nature of sound.

As pointed out earlier, humans can accurately identify the position where sound is being produced. This ability is known as sound localization. The process by which we localize sound is not yet fully understood; however, we know some of the important factors contributing to it. We listen to sound using two ears. When

sound arrives from one side, it will reach one ear faster and without encountering any obstacles. The other ear will listen to the same sound with a slight delay, since the sound wave had to travel a longer path. Furthermore, the sound level will be different when sound reaches the second ear since the head acts as an obstacle to sound propagation. This delay and attenuation of what one ear perceives compared to the other are called inter-aural differences. They are examples of the fundamental cues used by humans to localize sounds.

Early attempts to imbue sound recordings with spatial information rely on the inter-aural cues. Stereo systems, where two loudspeakers are placed at the front-left and front-right of the listener, have been used in this way. By changing the relative levels of the left and right channels, it is possible to induce the inter-aural differences within a certain range. More complex multi-channel systems used in movie theaters surround the listener with a few loudspeakers and can convey a wider range of inter-aural differences. However, this is not enough to fool the average listener into thinking sound is actually coming from any desired direction.

Research on human sound localization eventually led to the introduction of the Head-Related Transfer Function (HRTF). The HRTF codifies the way in which sound propagates in the presence of a specific human head. It includes the inter-aural differences, as well as some frequency-dependent effects caused by the reflection of sound by the listener's body, head and pinnae. This detailed representation makes the HRTF dependent on the specific anatomic details of the listener. Every person will exhibit a different HRTF that has to be measured. Nevertheless, some attempts at using the ideas behind the HRTF to convey spatial audio have been made. Binaural recordings, made by placing microphones in the ears of a dummy head, are an example. They can be very realistic, but they typically require the use of headphones. The need to use a fixed dummy head for recording also means that the perceived accuracy will differ among listeners, depending on how well can the dummy head match their respective HRTFs.

Until recent years, these were the only options available to record and reproduce spatial sound information. Technology has now advanced to the point

where another alternative can be considered: sound field reproduction. A sound field is the value of the sound pressure at every point in space. A conventional recording only measures sound at one point using one microphone placed there. Stereo or multi-channel systems use arrays of microphones to record sound at two or more points. Recording a sound field, however, would require filling the space with microphones and simultaneously recording all of that information. Our understanding of how sound propagates, however, allows us to calculate the sound field over an empty region from measurements made at a surface enclosing it. The extremely complex requirement of filling the space with microphones reduces to lining up microphones over a closed surface. The task is, nevertheless, a difficult one; however, it is now within the reach of modern technology. A common choice of surface to sample is the sphere. Its symmetric, compact nature leads to simpler mathematical expressions. The result of choosing this surface is the technique known as *Ambisonics*. Several microphone arrays for Ambisonics, that is, arrays that sample a spherical surface, have been built and demonstrated successfully.

The capability to record sound field information is not enough to produce a full communication or reproduction system. The sound field information must be somehow encoded for transmission or storage. The ability to synthesize sound fields described by these encodings is also required. These are the main topics covered in this dissertation.

Unlike previously discussed approaches, sound field reproduction systems do not consider human sound localization. The argument is that by re-creating an entire sound field around the listener, the sound pressure perceived by their ears will be indistinguishable from what they would hear if they were actually at the recording place. However, simple considerations of the human auditory system can help reduce the recording system's size and complexity. This is another important topic advanced in this thesis.

A critical difference between the research oulined in this dissertation and that found in previous treatments of sound field reproduction lies in the choice of a coordinate system. Conventional methods are formulated in spherical coordinates,

with a single listener assumed to be at the origin. However, this choice of coordinates makes it difficult to tackle situations where the user is unrestricted or where multiple users share a reproduction system. This dissertation approaches such scenarios by considering unconventional, that is, non-spherical geometries.

This dissertation starts by offering a brief introduction to the state of spatial audio systems. Emphasis is made on sound field reproduction technologies, particularly the one known as Ambisonics. It is here, in Chapter I, where the objectives of the present research are detailed.

Chapter II offers a review of the mathematical techniques used to study the propagation of sound. Readers familiar with the topic can skip this chapter, although it is intended to serve as a good overview. For those unfamiliar with the field, this chapter presents the basic requirements to understanding following sections of this dissertation.

Chapter III starts by reviewing some known results from considering sound propagation in cylindrical coordinates. Building upon this results, a set of new mathematical formulas that facilitate the recording of sound field information are derived. Their distinct feature is that they assume a cylinder, instead of a sphere, as the boundary surface. These formulas are applied to the problem of recording sound fields with cylindrical microphone arrays and a new encoding format, called *3D Cylindrical Ambisonics* is advanced.

Conventional Ambisonics is explored once again in Chapter IV, where the results obtained for cylindrical microphone arrays are applied to define a new method of recording Ambisonics with different horizontal and elevation resolutions. Normally, Ambisonics will have the same resolution in all directions given the use of a spherical boundary in its formulation. However, human sound localization does not exhibit this symmetry. We are more accurate at identifying the direction, in the horizontal plane, from which sound is reaching us than in estimating the sound's elevation. By reducing resolution for elevation while keeping the horizontal resolution, it is possible to design simpler systems that nevertheless offer a similar performance when presenting sound to human listeners.

The problem of synthesizing a sound field from the information recorded by a microphone array is discussed in Chapter V. This chapter first introduces the existing methods applied in Ambisonics. These methods require loudspeakers to be aligned in accordance to the foundations of Ambisonics. That is, sampling all directions with equal resolution over a sphere. A new method is introduced to allow for the reproduction of Ambisonics and other sound field recordings even if the geometry of the reproduction array does not match that of the recording one. Ambisonics can be reproduced, to some accuracy, with irregular, non-spherical arrays. Recordings made with cylindrical microphone arrays, similarly, do not require a cylindrical loudspeaker array.

Two of the main contributions of previous chapters are brought together in Chapter VI to define a full sound field reproduction system. The proposal uses 3D Cylindrical Ambisonics in the recording stage and an arbitrarily shaped loudspeaker array for reproduction. The definition of this system allows for the evaluation of the proposals considering important features such as the size of the listening region, where sound fields are reproduced with acceptable accuracy. The interaural cues perceived by a listener using the proposed system are calculated and compared with the ones they would perceive if present at the recording place. The effects of using a finite-length cylinder, which is not a closed surface, as a boundary are also discussed. Some considerations about microphone distributions on the cylinder and the possibility of using only half of the cylinder in some applications are presented.

This dissertation is wrapped up in Chapter VII, where the results achieved are briefly summarized and some conclusions are drawn. This dissertation hopes to become a pillar for the development of future sound field recording, synthesis, analysis and reproduction systems which are not constrained by the requirements of spherical symmetry or limited to a small listening region with the capacity for a couple of listeners at most.

Some final remarks regarding the implementation of an actual cylindrical microphone array are reviewed in Appendix A. There, some technical considerations needed to build a working cylindrical microphone array are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF VARIABLES

t	Time
$\vec{r}$	Spatial position vector
(x, y, z)	Cartesian coordinates
(r, θ, φ)	Spherical coordinates (radial coordinate, azimuth, elevation)
(r, θ, z)	Cylindrical coordinates (radial coordinate, polar angle, axial coordinate)
c	Speed of sound ($c \approx 343\,{}^{m}/_{s}$ at 20 °C)
$\vec{k}$	Angular wavevector
k	Angular wavenumber (related to the frequency f of sound by $k = {}^{2\pi f}/_{c}$)
∇^2	Laplacian operator
$\Psi(\vec{r}, t)$	Sound field in space and time
$\psi(\vec{r})$	Spatial part of the sound field
$\mathrm{T}(t)$	Temporal part of the sound field
$\mathrm{R}(r)$	Radial part of the solutions to the Helmholtz equation
$\Theta(\theta)$	Polar part of the solutions to the Helmholtz equation
$\mathrm{Z}(z)$	Axial part of the solutions to the Helmholtz equation
$\mathrm{R}_{\mathrm{sph}}$	Radius of a spherical microphone array
$\mathrm{R}_{\mathrm{cyl}}$	Radius of a cylindrical microphone array
$p(k, \vec{r})$	Sound pressure at point $\vec{r}$ and for angular wavenumber k
$\lvert\cdot\rvert^*$	Complex conjugate
$\lvert\cdot\rvert^+$	Pseudo-inverse
n	Polar expansion order
m	Spherical harmonic expansion degree
ξ	Cylindrical harmonic expansion damping ratio
Γ	Generalized factorial, gamma function
J_n	Bessel function of order n
Y_n	Neumann function of order n
H_n	Hankel function of the first kind and order n
j_n	Spherical Bessel function of order n
y_n	Spherical Neumann function of order n
$h_n^{(i)}$	Spherical Hankel function of the i-th kind and order n
Y_{nm}	Spherical harmonic function of order n and degree m
P_{nm}	Associated Legendre polynomial of order n and degree m

N_{nm}	Normalization weight for the spherical harmonic of order n and degree m
B_{nm}	Spherical harmonic expansion coefficient of order n and degree m
$\mathrm{Z}^{\pm}_{n,\xi}$	Cylindrical harmonic function of polar order n, damping ratio ξ and orientation $\pm$
$C^{\pm}_{n,\xi}$	Cylindrical harmonic expansion coefficient of order n, damping ratio ξ and orientation $\pm$
W	Plane-wave decomposition weights
F	Beamforming filters
G	High-Order Ambisonic decoding gains
$w^{\pm}_{n,\xi}$	Length-truncation spatial window for order n, damping ratio ξ and orientation $\pm$
α	Length-to-taper ratio for the length-truncation spatial window
$z_{\max}$	Axial coordinate for the last microphones in a finite-length cylindrical array
$\mathrm{N}_{\mathrm{mic}}$	Number of microphones in an array

CHAPTER I

Introduction

1.1 Motivation

As technology advances, users of multimedia systems have demanded the evermore realistic presentation of contents. The level of realism of a given presentation is difficult to quantify. Different users may place their priorities on different variables to judge realism. Even if only sound information is considered, some users may focus on the sound source itself and regard clarity as a most important requirement for realism, while others may consider the listening environment and deem diffusiveness to be more important. In the end, however, it should be agreed that, by definition, whatever is heard when present at an actual scene is its most realistic acoustic representation.

Sound transmission and reproduction systems have made great progress since the invention of the first devices in the 1870's [1, 2]. It is still difficult to build recording and reproduction systems that can match the sensitivity and accuracy of the human ear over the whole listening range. However, most high-end systems are good enough that the average user cannot really perceive any distortion caused by the recording, storage/transmission and reproduction processes.

Despite these advances, one crucial aspect of human auditory perception

has proved to be difficult to incorporate in audio systems: sound localization. Sound localization is the ability to identify the position in space at which a certain sound is being originated. In our daily lives, we routinely perform sound localization to facilitate a variety of tasks. For example, to engage in conversation when someone suddenly calls out our names or to prevent an accident when machinery is being operated near us.

There are a number of obstacles when trying to present sound in a way that can convey the spatial information required for sound localization. The specific hurdles one would face depend on the strategy being adopted. Two types of approach can be considered: a listener-centric approach and a sound-centric approach. The former attempts to convey the spatial cues used by the listener to localize sound; in essence, to deliver at his two ears the particular sounds he would perceive if present at the target scene. The latter, sound-centric approach, attempts instead to re-create the physical variables present at the target auditory scene. It is argued that reproducing the target acoustic conditions faithfully will convey all of the spatial cues used by the listener, irrespective of what they could be.

A listener-centric approach requires tuning of the system in accordance to the particular user. It must also either restrict user movement, or track it so that the acoustic signals being delivered can be consistently adjusted. Any user tracking loop will exhibit some delay, which may cause the inconsistent presentation of sound during heavy user movement.

The sound-centric approach does not have any of the problems of the listener-centric one. All listener-related phenomenology, such as scattering by the head and reflections at the pinna, will naturally occur within the re-created acoustic environment. No motion sensing is required either. On the other hand, measuring

or calculating the acoustic variables over a large region and then re-creating them is a difficult task. Some systems of this kind have actually been built and shown a promising performance. However, they are typically restricted to very small listening regions.

There is no doubt about the demand for spatial audio systems, those that can deliver the sound localization cues either directly or by re-creating the acoustic conditions over some region. Some promising technologies have emerged in an attempt to meet this demand. An overview of this techniques is presented in the following sections.

The problem of conveying realistic sound information, including spatial cues, is still far from solved. This dissertation attempts to contribute towards this final goal by laying out a series of mathematical techniques and formulas. It is hoped that the new methods advanced in this dissertation will accelerate the pace at which truly 3D sound systems reach end-user applications such as teleconferencing and entertainment.

1.2 Sound localization

The exact process by which the human auditory system localizes sound is not yet fully understood. Nevertheless, some of the cues we rely on for this task have been identified [3]. In this section, a brief review of what we know about the way in which humans localize sounds is presented. This is a very active research field and a full account would be beyond the scope of this dissertation. Attention is given mainly to the results that have been found to be important in the design of spatial sound presentation systems.

1.2.1 Interaural time and level differences

The interaural level difference (ILD) and interaural time difference (ITD) are two important cues for sound localization [3]. Figure 1.1 shows a simple illustration of the phenomenology behind these two cues. A sound source is present at a certain azimuth from the point of view of the listener. Since the speed of sound is finite ($c \approx 343\ \frac{m}{s}$ at 20 °C), sound from the source will reach the ipsilateral ear a few moments before reaching the contralateral one; a time difference. The path to the ipsilateral ear is a direct one, free from any obstacles. However, to reach the contralateral side, sound waves must diffract around the head. The consequence is that the ipsilateral ear picks up more energy than the contralateral one, creating a level difference.

The two types of interaural difference described complement each other allowing us to estimate the azimuth angle at which any sound source can be found. At low frequencies, for which the wavelength of the sound wave is about twice the size of the listener's head or larger, diffraction around the head does not significantly reduce the wave's energy making ILD useless for localization. However, the difference in phase between the sounds picked up by both ears makes inferring the azimuth of

arrival possible; ITD is, therefore, the main cue for localization in the horizontal plane at low frequencies, below around 800 Hz. At high frequencies, however, the situation reverses. If the wavelength is shorter than the size of the listener's head, starting at around 1.6 kHz, the difference in phase between both ears cannot be used to identify a unique azimuth for the sound source. However, the effects of diffraction around the head produce a region of low energy behind it. This is called an acoustic shadow and one of the two ears, the contralateral one, will be located somewhere inside of it. The sound source's azimuth can be inferred from the ILD caused by the head shadow. At intermediate frequencies, both cues are used to localize sounds.

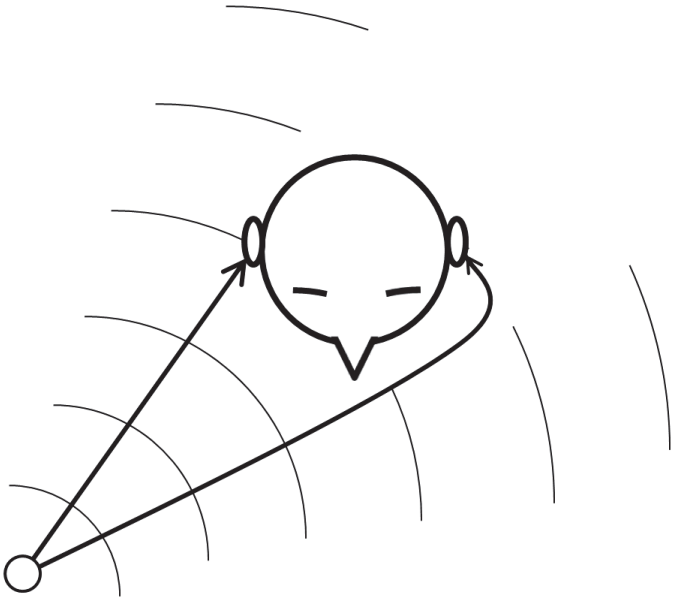


Figure 1.1: Interaural time and level differences. The sound from the source reaches the listener's right ear after travelling a short and direct path. The same sound wave must travel a larger distance and diffract around the listener's head to reach the left ear. Sound picked up by the ear ipsilateral with the source arrives earlier and appears louder than that picked by the contralateral ear.

1.2.2 Cones of confusion

Interaural differences are useful only to identify the azimuth angles of sound sources with respect to the listener. They cannot be used to fully localize sounds in 3D space. The reason behind this is illustrated in Fig. 1.2, showing what are called the *cones of confusion* [4]. Sound sources at different elevations can share the same interaural differences.

The cones of confusion can be constructed and understood by first considering a reference axis: the interaural axis. The interaural axis is defined as the line passing through both ears. This also serves as the axis for two right circular cones with their apexes located at each of the listener's ears. For any given pair of cone apertures the intersection of the two cones will define a circle. The distances between a sound source lying on this circle and the listener's ears are given by the slant heights of the cones; a value that is constant for any point on the circle. This implies that the paths that sound waves need to travel to reach the ears will have a fixed length no matter what position the sound source takes on the circle. The interaural differences, therefore, yield ambiguous results for the exact position of the sound source; however, this discussion does not consider the effects of acoustic scattering by the head.

Sound localization in humans is not constrained by the limitations of the interaural differences. We can easily identify sounds as being in front, behind, above or below us even if these four points lie on one of the circles described by the cones of confusion. For this, we rely on different kinds of spatial cues.

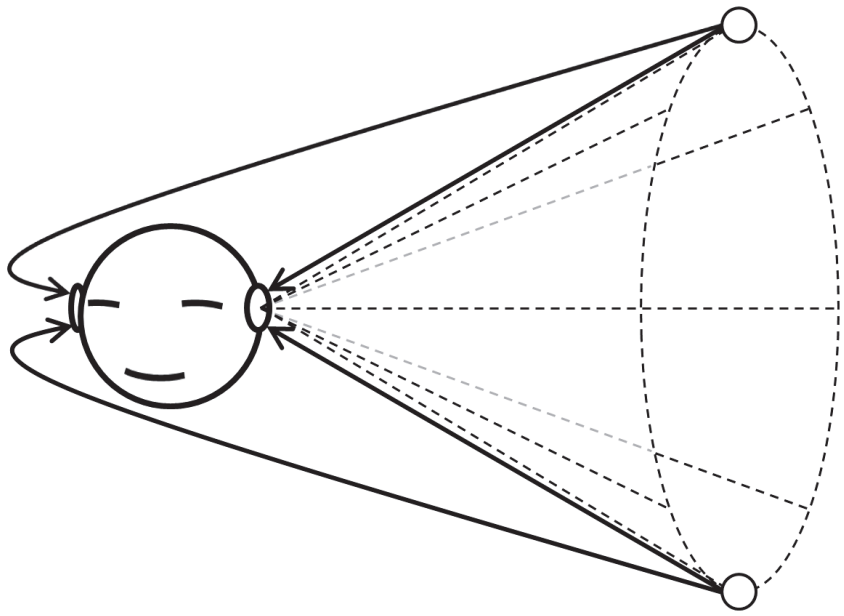


Figure 1.2: Cones of confusion. Interaural differences are constant over any circle that is centered at and perpendicular to the interaural axis. Sound waves from both sources, above and below the interaural axis, will travel similar paths to reach the listener's ears. While interaural differences exist, they are not enough to determine the position that the source occupies over the circle.

1.2.3 Head-related transfer function

Sound waves travelling from their source to our ears will invariably interact with our bodies before reaching the ear canal and finally the eardrum. The appearance of an acoustic shadow on the side of the head opposite to the sound source was discussed above. This, however, is not the only kind of acoustic diffraction used by humans to localize sound sources.

As sound reaches a listener, it is diffracted around their head and torso and reflected from their shoulders before arriving at the ears. There, it is once again scattered, this time by the intricate folds of the pinna. The process leaves a distinct mark on the sound waves that manage to enter the ear canal. Some frequencies are amplified while others are attenuated in a complicated manner. The exact details of the process depend heavily on the anatomy of the listener. However, this filtering and amplification of select frequencies encodes information about the location of the sound source.

At a coarse scale, the human head can be approximated by a sphere. This is enough to explain the head shadow, the interaural differences and gives rise to the cones of confusion. However, this coarse model is only useful for low frequencies, where the wavelengths are too large to resolve fine details. Sound interaction with the pinna, even a simplified one, is enough to break the symmetry condition that leads to the cones of confusion. The relative distances between the ears and the sources may remain the same; however, the paths travelled by the sound waves of previously ambiguous sources are no longer equal.

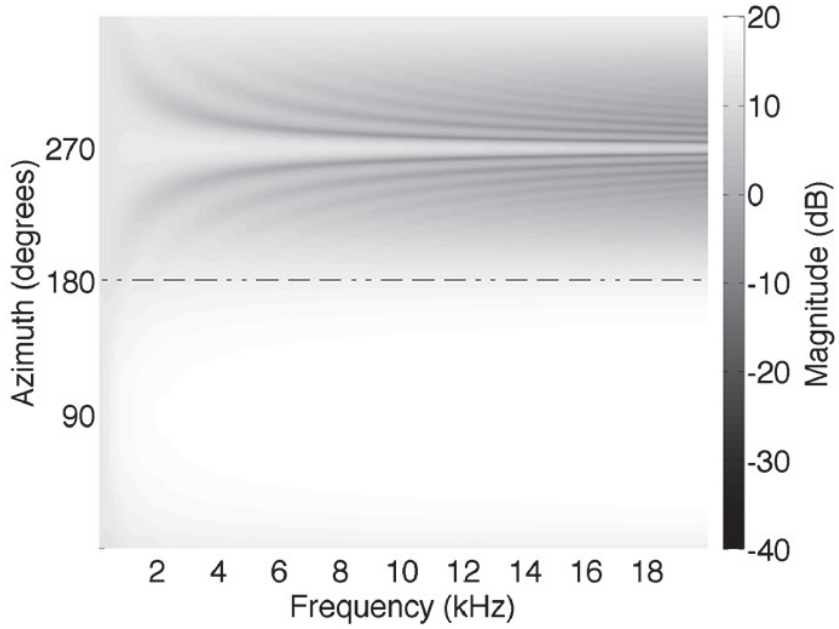
A way to quantify these effects is by defining sets of acoustic transfer functions. A transfer function expresses the relationship between one physical variable at the beginning of a linear process (the input) and an observable of interest

at its end (the output). To compare the effects of diffraction for sound sources at different positions, the input should be regarded as sound originated from a given point in space. The output would be the sound pressure at the ear canal or eardrum, after the sound waves have interacted with the listener's body.

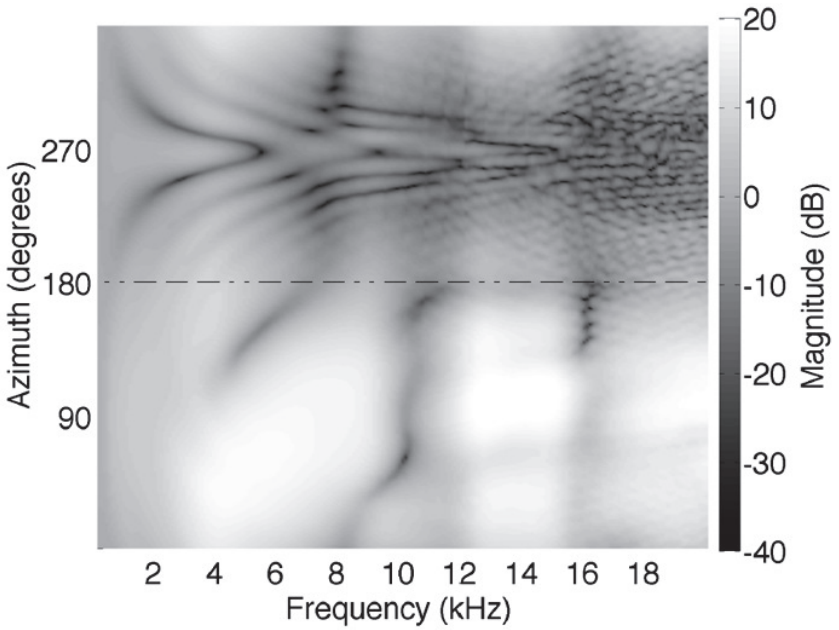
In the study of sound localization, it is common to remove the component due to propagation over the air without obstacles from the transfer function. The result of doing this is known as the Head-Related Transfer Function (HRTF), and is illustrated in Fig. 1.3 along with the coarse result of approximating the head as a sphere, or removing all fine details including the pinna.

The HRTF shown in Fig. 1.3b corresponds to a SAMRAI dummy head, a mannequin designed by KOKEN to approximate the average features of a Japanese male. Each horizontal line on the plot represents the magnitudes of one transfer function, with the input located on the horizontal plane, 1.5 m away from the center of the head. The output is measured at the mannequin's right ear, near the 90° mark. The ear canal was sealed off during the measurement.

As expected, at low frequencies the HRTF very closely matches the transfer functions for a rigid sphere. Once the wavelength becomes short enough to resolve the finer details of the dummy head, some variations in the transfer function are observed. These regions of amplified or attenuated sound transmission are used by the auditory system to localize sources.



(a) Transfer functions for sound propagation onto a rigid sphere.



(b) Transfer functions for sound propagation onto a dummy head.

Figure 1.3: Magnitude of the sound propagation transfer functions showing the effects of acoustic scattering by (a) a rigid sphere, and (b) a dummy head approximating a human head. Both models behave similarly at low frequencies. At high frequencies, however, the effects of fine structures like the pinna have a significant impact on the transfer functions.

1.2.4 Active listening

There is another way to disambiguate the position of sources that lie on the circles defined by the cones of confusion. In natural conditions, humans need not listen to sound while remaining in a fixed position and posture. Slightly rotating the head is enough to change the interaural axis, making interaural differences sufficient to differentiate between points of the previously ambiguous circle.

The process of moving while listening to sounds is called active listening [5]. It is known to be carried out by humans during sound localization tasks [6]. A moving listener introduces a significant challenge for the realistic presentation of audio imbued with spatial information. Systems must either actively track their users and adjust the presentation, or they should somehow manage to work for any possible position the user may be at and with any posture they may decide to adopt.

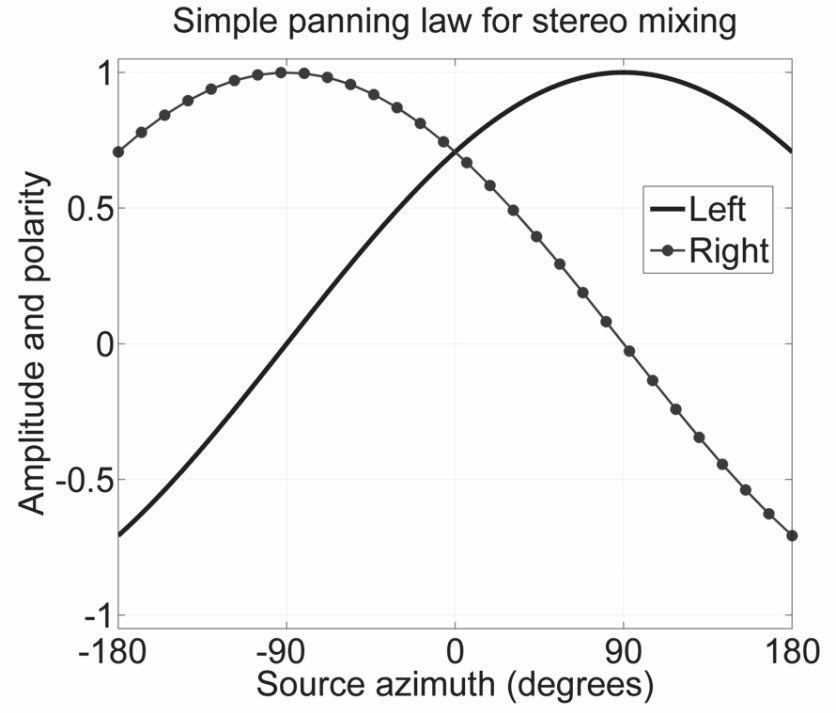


Figure 1.4: Panning law dictating the left and right channel levels for a stereo system. By driving a stereo system with similar signals at different levels for each channel, it is possible manipulate interaural differences and create the illusion of sound arriving from a position between the loudspeakers.

1.3 Presenting spatial audio

The term spatial audio is used to refer to any kind of sound that has been imbued with the spatial information needed for listeners to perform sound localization. Systems capable of presenting spatial audio will be referred in this dissertation as Three-Dimensional Virtual Auditory Displays (3D VAD). This section introduces some well-known and promising 3D VAD design approaches.

1.3.1 Stereo and surround systems

The first attempt to add the spatial component to sound recordings came in the form of stereophonic sound. A stereo system can hardly be called a VAD, however, given its limited resolution and range. Nevertheless, they represent a

turning point in the development of sound presentation technologies.

In a stereo system, two loudspeakers are placed in front of the listener, typically at 30° to the left and right of the center. Driving the loudspeakers independently will of course lead to sounds being localized at these two positions in space. However, if the loudspeakers are driven by similar signals, varying the relative levels of the two channels, it is possible to induce interaural differences that point to a sound source being somewhere in between the two loudspeakers [7]. This is known as amplitude panning; Fig. 1.4 shows a typical panning law used to set the channel levels according to the desired position for the virtual sound source.

The natural extension of stereo are the surround multi-channel audio systems. Popular configurations are the 5.1-channel and 7.1-channel layouts [8]. These systems extend the range of angles from which sounds can be presented by surrounding the user with loudspeakers. They are, however, limited to the horizontal plane and cannot truly convey sounds at any arbitrary position; they can only manipulate interaural differences, only ILD in most cases.

Multi-channel technologies are still being developed and pushed forward in consumer applications. A system incorporating elevation by distributing the loudspeakers on 3 planes of varying height is scheduled to be released with the next generation of ultra high-definition televisions [9].

1.3.2 Binaural presentation

The realistic presentation of 3D spatial audio contents needs to consider more than ILD. Our lack of knowledge regarding what are exactly the features of sound which are important for sound localization makes it difficult to design an optimal system conveying only what is needed. However, a very realistic presentation can

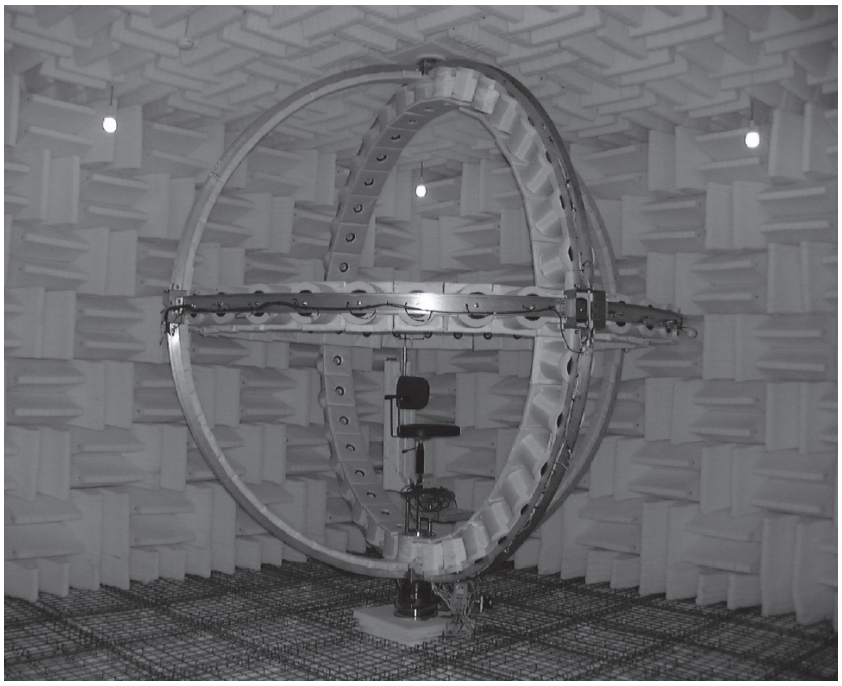


Figure 1.5: A loudspeaker array inside an anechoic chamber. The system can be used to measure the Head-Related Transfer Function. Subjects must sit still at the center of the array while the loudspeakers present sounds around them. A pair of microphones inside their ears are user for the actual measurements.

still be achieved by using the entire information available in the HRTF.

The HRTF holds the information about changes to sound as it interacts with the listener's head to reach the ear canals. Therefore, it can be used to calculate the sound pressure that must be reproduced directly at the listener's ears in order to convey the illusion of a sound source present at any desired position for which the HRTF is known. Sounds generated in this way are called binaural audio. Binaural audio can also be recorded by placing small microphones inside the listener's or a dummy head's ears and directly measuring the sound pressure as sounds reach them from different places.

Binaural sound presentation can achieve very realistic results. It, however, presents several drawbacks. The HRTFs must be measured, or binaural recordings made, for every user of the system since they are heavily dependent on the listener's

anatomy. Measuring the HRTF requires specialized equipment like the loudspeaker array shown in Fig. 1.5. The user must stay still at the center while the loudspeakers rotate around him/her and present sounds from the positions to be measured, one by one. This procedure can take several minutes up to a couple of hours and the user must remain still throughout the whole process.

Listening to binaural audio requires presenting sounds directly at the ears, which is easier to accomplish using a pair of headphones. The need to wear headphones is a significant demerit of binaural reproduction. Some attempts have been made to use loudspeaker-based systems for binaural presentation [10]. They are usually referred to as transaural systems. Transaural systems make use of inverse filtering and are very sensitive to changes in the listening conditions for which they were designed; for this reason, they cannot easily accomodate user movement.

Finally, incorporating active listening into a binaural system requires the means to continuously track the user and update the HRTF being used accordingly. This may require the user to affix a tracking sensor to their bodies and there will be an unavoidable delay between the user's movement and the adjustment of the auditory signals.

1.3.3 Sound field reproduction

Binaural sound reproduction is an example of a listener-centric strategy to convey spatial sound information. This subsection introduces a group of sound-centric approaches that are collectively known as *sound field reproduction*. A sound field refers to the sound pressure observed at all points in space. Sound field reproduction, therefore, does not attempt to re-create the sound pressure at the listener's ears directly. The goal is to reproduce the entire sound field over a region

large enough for the user to fit. If this is achieved, the diffraction effects encoded in the HRTF will take place naturally, irrespective of the posture adopted by the user. A truly realistic 3D sound presentation could be achieved without the need of user tracking or tedious measurement of the HRTF.

In this subsection, three classes of sound field reproduction systems are considered. Unfortunately, technological constraints still prevent any of these methods to be implemented in a way that the entire listening range of humans is covered faithfully. They all have shown, however, promising results and research on all of them is active.

The methods are described here in a general way without emphasizing the mathematical details behind their formulations. Chapter II reviews in greater detail the wave equation and the results that make sound field reproduction possible, particularly the Huygens-Fresnel principle and the Kirchhoff-Helmholtz integral theorem. An exception to this is a technique called Ambisonics, for which a more formal introduction will be given in a later section of this chapter.

All of the discussed methods consider an important property: sound propagation is, with very high accuracy, a linear phenomenon. Non-linear effects exist and are actively studied; however, linear acoustics are a very good approximation to real-world phenomenology for most applications, particularly those pertaining human listening. A transfer function for sound propagation can, therefore, be assumed to exist and comply with the theory of linear systems. This function, called the Green function, is formally introduced in Chapter II.

1.3.3.1 Wave-field synthesis

Wave Field Synthesis (WFS) is a technique to present 3D sound using loudspeaker arrays [11]. It works by approximating the sound waves that would

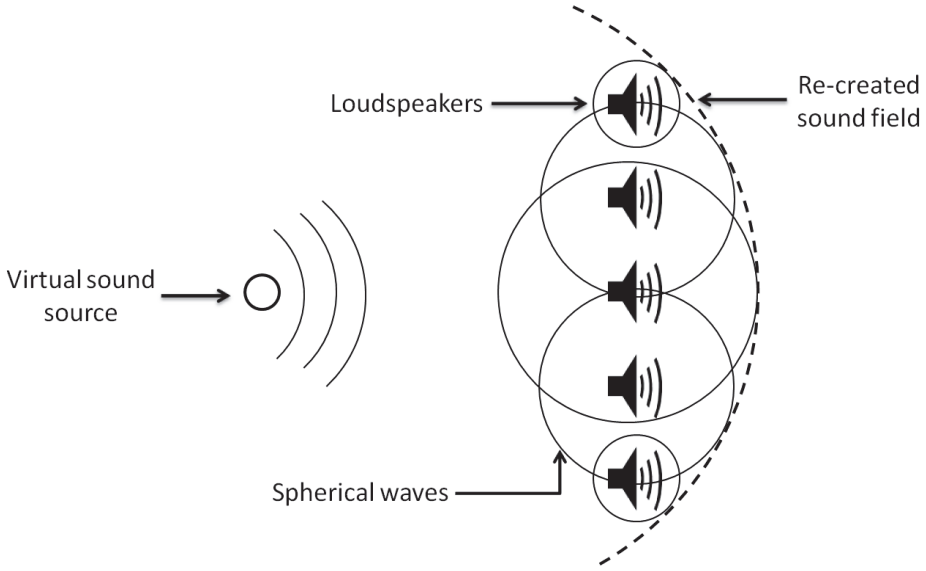


Figure 1.6: Illustration of a Wave Field Synthesis 3D sound reproduction system. A loudspeaker array is driven by properly filtered and timed signals so that the spherical waves produced by them add up into the approximation of a target wavefront. In this case, the loudspeakers approximate the sound field that would be observed if a source was present at the indicated position behind them.

be observed for a target sound source using the carefully tuned and timed sound waves from a fixed set of secondary sources (the loudspeakers). An illustration of the method is shown in Fig. 1.6.

The technique frequently relies on either a linear or a planar array of loudspeakers to densely cover all directions inside a segment or patch of space. The system then drives the loudspeakers so as to reconstruct all sound waves passing through this region and towards the listener's side. The driving functions are simply the sound pressure that would be observed at each loudspeaker position if the target acoustic scene was actually taking place behind the loudspeaker array.

In terms of computational complexity, it is simple to implement a WFS loudspeaker driving system. However, building the actual loudspeaker array makes a major obstacle of WFS evident: spatial aliasing. In WFS, the sound field is directly sampled in space at the loudspeaker positions. Sound fields, being solutions

to the wave equation, have an oscillatory nature in both time and space coordinates. The sampling theorem [12] dictates, therefore, that the average separation between loudspeakers must be half of the shortest wavelength present or less. If the full listening range for humans is considered (20 Hz to 20 kHz), a loudspeaker will be required every 8.5 mm. Modern loudspeaker manufacturing cannot yet produce loudspeakers small enough to align in such a dense grid while retaining good acoustic performance, such as a sufficiently high volume.

Despite the problems introduced by spatial alias, research surrounding WFS is very active. Reproducing sound fields that span the entire listening range may be out of reach for now; however, less ambitious projects [13] have successfully demonstrated that WFS is a viable technology for the ultra-realistic presentation of spatial sound to large audiences.

A more critical issue with WFS systems lies in the need to match specific recording and reproduction systems. There is no intermediate spatial encoding that can be used to share the same contents among users of different reproduction arrays.

1.3.3.2 Boundary surface control

Boundary Surface Control (BoSC) is a different type of technology for sound field reproduction [14]. While WFS can be used to present sounds at some desired spatial position, BoSC attempts to record the actual sound field at a given location and to later reproduce it using loudspeakers. A full BoSC system is, therefore, composed of a microphone and a loudspeaker array. This kind of system is illustrated in Fig. 1.7.

The linear nature of sound propagation is fully exploited by BoSC, which can be thought of as a multiple-input and multiple-output (MIMO) control system. The microphone array is initially placed inside the loudspeaker array to measure the

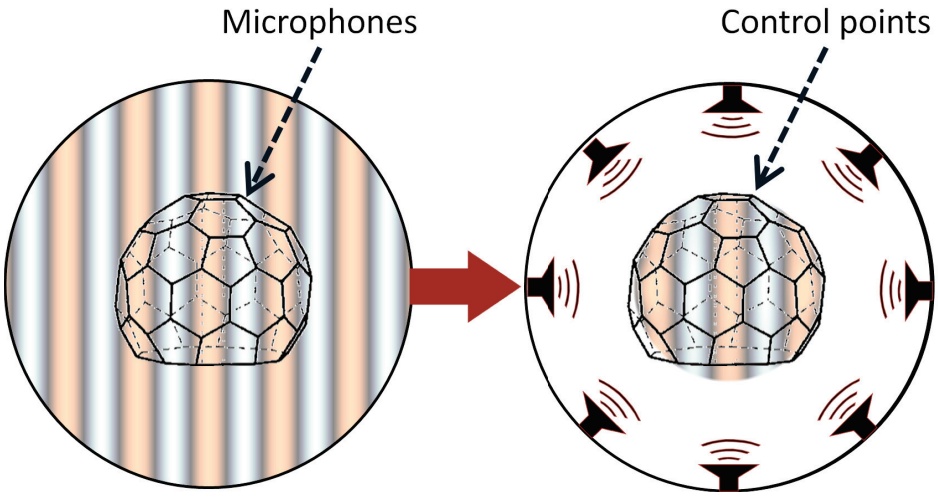


Figure 1.7: Illustration of a Boundary Surface Control system. Sound fields are sampled using a spherical microphone array. The recorded sounds are then used to drive a specific loudspeaker array so as to re-create the points sampled by the microphones. The result is an approximation to the original sound field within the region delimited by the control points.

transfer functions of each loudspeaker to each microphone. The results are condensed into a matrix which can be used to calculate the expected microphone signals when the loudspeakers are driven in some specific way.

In a BoSC system, the inverse of the transfer functions matrix is calculated and used to derive the loudspeaker signals that will most closely re-create a set of microphone measurements. Afterwards, it is possible to use the microphone array to record any desired sound field. The sound pressure measurements are then filtered by the inverse transfer functions; the results are the loudspeaker signals that will most closely re-create the acoustic conditions under which the recordings were made.

The results of a BoSC system can be as good as the equipment used allows. Very high levels of realism are possible; however, the formulation of BoSC limits it to the pairs of microphone and loudspeaker arrays which have been calibrated in advance. There is no intermediate encoding which captures the sound field information and allows for its analysis, edition or system-agnostic distribution. BoSC

is an effective technique which successfully achieves its objectives. However, a more versatile method which provides a more direct access to the spatial features of sound fields is also desirable.

1.3.3.3 Ambisonics

The third sound field reproduction technique to be reviewed in this introduction is called Ambisonics [15]. Ambisonics, as originally defined, did not attempt to reproduce sound fields directly. It was introduced as a technique to pan sound sources around the listener using an array of loudspeakers, while keeping the so-called energy and velocity vectors parallel to each other.

The velocity vector is, in essence, the acoustic intensity observed at the listening position, typically the center of the loudspeaker array. That is, the average of all wavevectors that appear in the array's presentation of a given sound. Mathematically it is defined by the following set of formulas:

$$\begin{aligned} \mathrm{P}(k) =& \sum_{\mathrm{spk}} \mathrm{G}_{\mathrm{spk}}(k), \\ \mathrm{V}_x(k) =& \frac{1}{\mathrm{P}(k)} \sum_{\mathrm{spk}} \mathrm{G}_{\mathrm{spk}}(k) \cos(\theta_{\mathrm{spk}}), \\ \mathrm{V}_y(k) =& \frac{1}{\mathrm{P}(k)} \sum_{\mathrm{spk}} \mathrm{G}_{\mathrm{spk}}(k) \sin(\theta_{\mathrm{spk}}). \end{aligned} \tag{1.1}$$

The symbols in the equation above are: k for the angular wavenumber, P denoting the total sound pressure at the listening position, $\mathrm{G}_{\mathrm{spk}}$ stands for the gain applied to the loudspeaker of label "spk", located at the azimuth angle θ_{spk}. The velocity vector, shown here for the horizontal plane only, has V_x and V_y as its x and y components, respectively. The summations are carried out over all the loudspeakers in the array under consideration. A more structured introduction to the notation used throughout this dissertation is done in Chapter II.

The energy vector is slightly more complicated; it is defined not as the average but as the 2-norm of the wavevectors, calculated component by component. Intuitively, it can be considered as indication of the direction from which most part of the acoustic energy reaches the listening position. Mathematically, however, it differs from the velocity vector simply by taking the square of the loudspeaker gains. The following expressions can be used to calculate the energy vector:

$$\begin{aligned} \mathrm{E}(k) &= \sum_{\mathrm{spk}} \mathrm{G}_{\mathrm{spk}}^2(k), \\ \mathrm{E}_x(k) &= \frac{1}{\mathrm{E}(k)} \sum_{\mathrm{spk}} \mathrm{G}_{\mathrm{spk}}^2(k) \cos(\theta_{\mathrm{spk}}), \\ \mathrm{E}_y(k) &= \frac{1}{\mathrm{E}(k)} \sum_{\mathrm{spk}} \mathrm{G}_{\mathrm{spk}}^2(k) \sin(\theta_{\mathrm{spk}}). \end{aligned} \tag{1.2}$$

The magnitude of the energy vector is denoted by E, while its x and y components are labeled as E_x and E_y, respectively. Once again, these expressions are given only for the horizontal plane for simplicity. However, they can be readily extended to three dimensions to also consider elevation.

In his original paper, Gerzon introduced Ambisonics as a way to keep both energy and velocity vectors parallel, while arguing that attaining this will reproduce the acoustic intensity in the neighborhood of the listening region [15]. It took three decades for this result to be incorporated into a sound field reproduction technology which was given the name of High-Order Ambisonics [16, 17].

1.4 High-Order Ambisonics for sound field reproduction

High-Order Ambisonics (HOA) is an extension to the ideas that had been proposed to record and reproduce Ambisonics. In Ambisonics, an omnidirectional microphone is coupled with three directional ones, all of them exhibiting a figure-of-eight polar pattern. Each of the directional microphones corresponds to one axis in the Cartesian coordinates system. This configuration allows for the direct measurement of the energy and velocity vectors observed at the measurement position.

The angular capture patterns of the four microphones used in Ambisonics correspond to a subset of the special functions referred to as the spherical harmonics. Figure 1.8 illustrates some of these functions, which have found applications in all areas of physics where rotation symmetry is somehow present. The mathematical formula to evaluate the spherical harmonic function of degree n and order m, Y_{nm}, at the azimuth angle θ and elevation angle φ is [18]:

$$\mathrm{Y}_{nm}(\theta, \varphi) = \mathrm{N}_{nm} e^{im\theta} \mathrm{P}_{nm} (\cos \varphi) . \tag{1.3}$$

The symbol P_{nm} denotes the associated Legendre polynomial of degree n and order m. Meanwhile, N_{nm} is a normalization constant that ensures that the spherical harmonics are orthonormal. A more in-depth review of the spherical harmonics and their properties is presented in Chapter II.

The original Ambisonics microphone matches the first four spherical harmonic functions as follows: the omnidirectional microphone corresponding to the zeroth-degree spherical harmonic function ($n = 0$), while the three figure-of-eight ones are associated with the three first-degree spherical harmonics ($n = 1$).

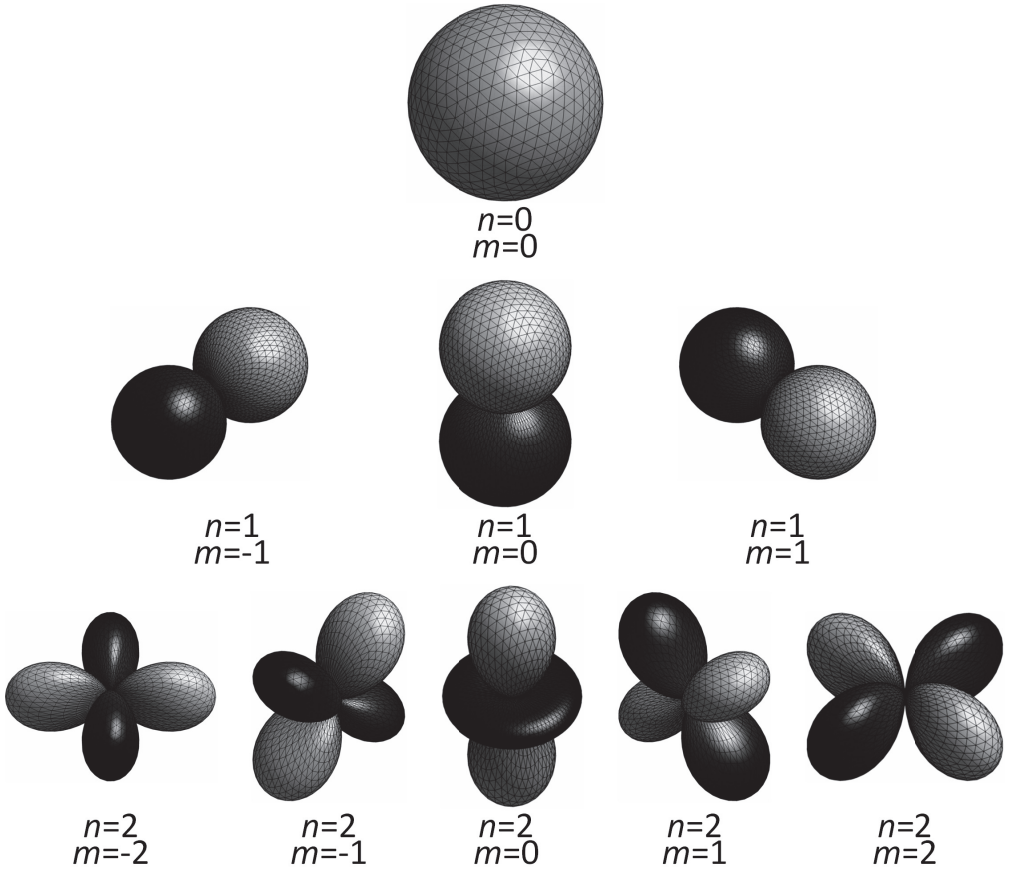


Figure 1.8: The spherical harmonic functions of degrees $n = 0$, 1 and 2. The functions are plotted for all orders m corresponding to the degrees shown. The bright lobes indicate regions where the spherical harmonics take positive values; dark lobes correspond to negative values. The amplitude of the spherical harmonics is given as the radial coordinate of the plot. The farther a point is from the center, the greater the amplitude of the spherical harmonic at that position.

High-Order Ambisonics builds upon this observation by including spherical harmonic functions of degrees higher than one. It is at this point that an important, and somehow confusing, convention must be introduced. While mathematicians and physicists use the term degree and order of the spherical harmonics in a way that it matches the degree and order of the associated Legendre polynomials, those working in the field of sound field reproduction usually reverse these terms. The degree of a given spherical harmonic is called its *Ambisonic order*. The different terms can prove confusing if not handled with care. In this dissertation the phrase *Ambisonic order*

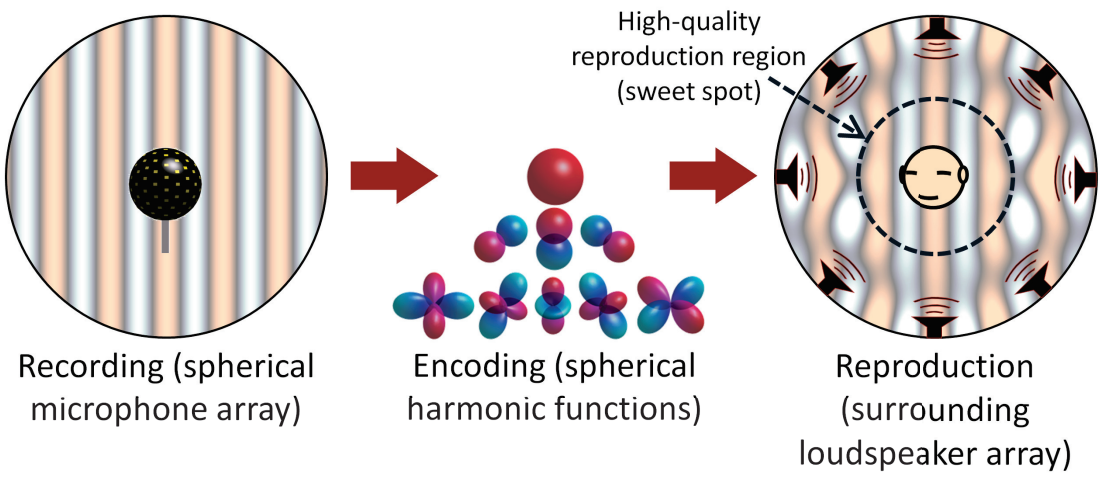


Figure 1.9: A high-order Ambisonics system. Sound fields are recorded using a spherical microphone array. The resulting signals are later encoded using the spherical harmonic functions. The result can later be decoded for reproduction using any surrounding loudspeaker array with an appropriate high-order Ambisonics decoder. This procedure results in a region at the center of the loudspeaker array in which reproduction accuracy is highest. This region is called the *sweet spot*.

is used as synonymous to the spherical harmonic degree.

1.4.1 The High-Order Ambisonics encoding

Applying the spherical harmonic functions to the problem of recording spatial sound field information has two significant consequences. First, it makes it possible to define a sound field recording scheme that has an arbitrarily high spatial resolution. There is no limit to the maximum Ambisonic order that a hypothetical system can record. The only restriction is imposed by the resolution of the recording equipment itself. Another way to state this is to say that the HOA description of sound fields is scalable. A schematic for such a recording, encoding and reproduction system appears in Fig. 1.9.

The second property is that, by using the spherical harmonic functions as a basis to describe the sound field, all references to the specific recording

or reproduction systems used vanish. The HOA description of a sound field is system-agnostic. It is defined exclusively by the spatial features of the sound field and does not make any reference to how they were measured or how are they to be used.

A description of the sound field in terms of the spherical harmonic functions is called its *HOA encoding*. Arguably, the largest advantage of HOA over other techniques like WFS or BoSC lies in the definition of this encoding. Its system-agnostic nature allows any HOA recording to be reproduced using any spatial audio system, as long as it has what is referred to as a *HOA decoder*. That is, a filter or some other method to convert the spherical harmonics description into signals for the reproduction system's loudspeakers.

Formally, the HOA encoding is defined by what is called the spherical harmonic decomposition. It is given by the following equations [19]:

$$\mathrm{B}_{nm}(k) = \int_{\theta=-\pi}^{\pi} \int_{\varphi=-\pi/2}^{\pi/2} p(k,\theta,\varphi)\mathrm{Y}^{*}_{nm}(\theta,\varphi)\sin(\varphi)d\varphi d\theta. \tag{1.4}$$

$$p(k,\theta,\varphi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \mathrm{B}_{nm}(k)\mathrm{Y}_{nm}(\theta,\varphi). \tag{1.5}$$

Equation (1.4) shows the encoding of a sound pressure distribution p as a set of Ambisonic channels B_{nm}. The inverse operation, shown in Eq. (1.5), decodes the coefficients to re-create the sound pressure over a sphere. These expressions use an infinite number of spherical harmonic functions to ensure the equalities hold. Actual implementations, however, truncate the the expansion to a given number N, the Ambisonic order of the system.

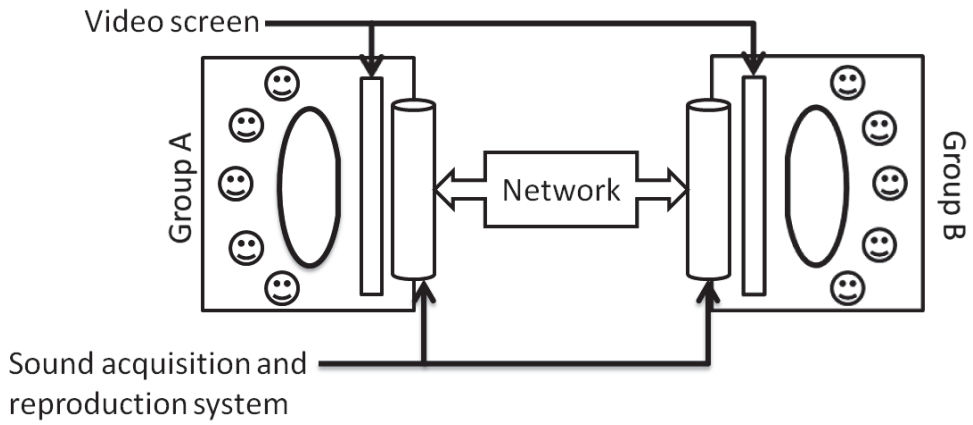


Figure 1.10: A hypothetical teleconference system. Two teams can collaborate at distance using ultra-realistic sound and video presentation. While not the goal of this dissertation, it illustrates one of the practical applications where the results achieved during this research can prove to be useful.

1.5 Research objectives

This dissertation seeks to contribute towards the development of ultra-realistic systems for the presentation of 3D spatial audio. To this end, a sound-centric approach, one that focuses on the physical variables pertaining spatial sound, is adopted. The reasoning behind this choice lies in the large number of difficulties that listener-centric approaches, such as binaural reproduction, face in adjusting to individual listeners and coping with phenomena like active listening.

In more specific terms, it is an objective to provide new tools that facilitate the design and construction of sound field recording and reproduction systems. Sound field reproduction has shown promising results despite its limited application. It elegantly sidesteps the issues of listener differences and movement. While it requires large multi-channel arrays, it is not entirely out of reach for modern technologies. An overview of the problems tackled in this dissertation and how do they stand in the frame of present technologies is shown in Fig. 1.11.

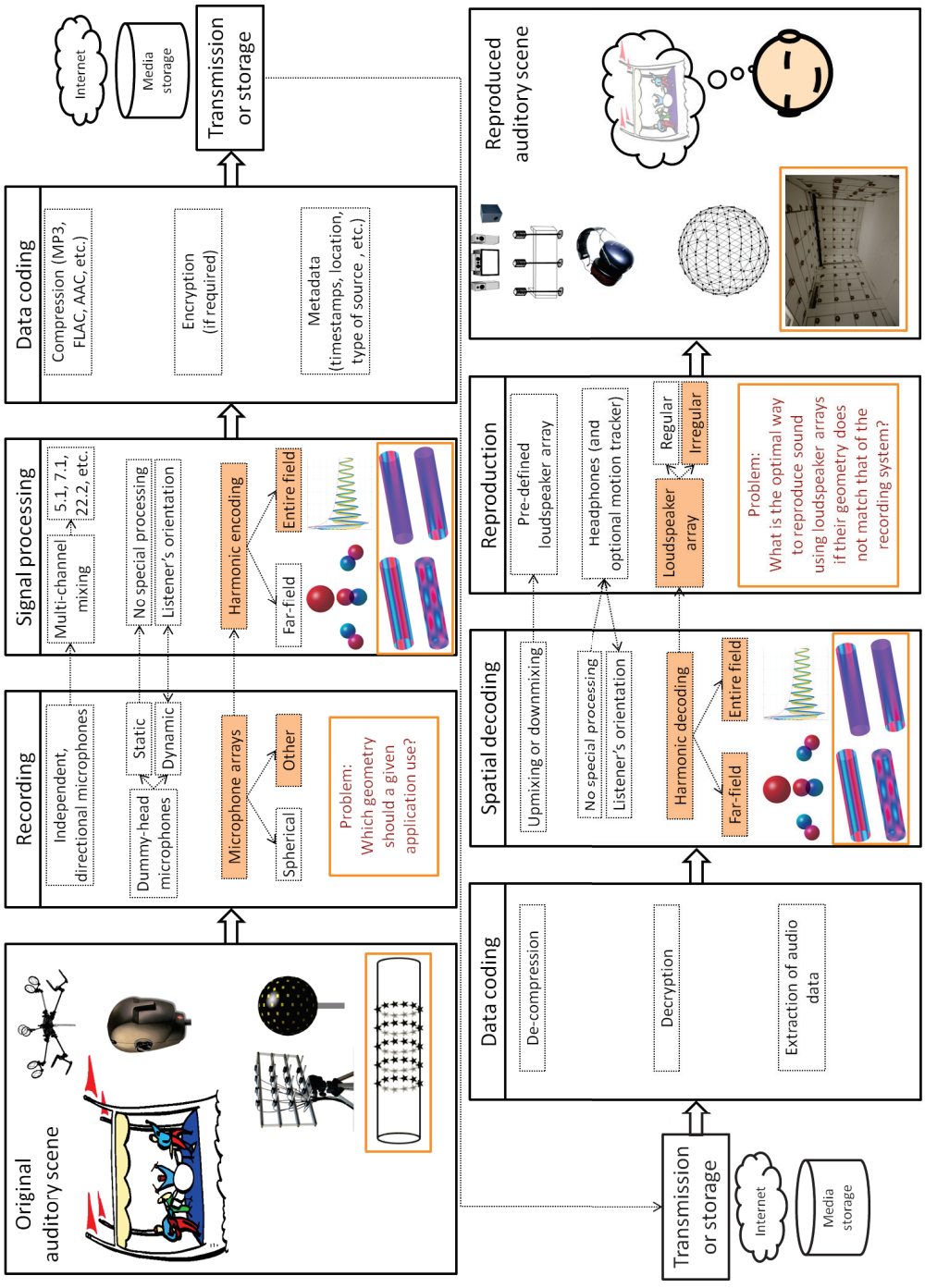


Figure 1.11: Overview of the field of spatial audio. The orange rectangles mark the areas in which this dissertation introduces innovations.

The proposal of this dissertation follow the lines of High-Order Ambisonics, a sound field reproduction technique distinguished for its scalable and system-agnostic format is considered as a starting point. The HOA format makes it possible to codify sound field information to any desired accuracy and later reproduce it using any spatial sound presentation system available. From large loudspeaker arrays to headphones, all reproduction systems can handle the contents of a HOA recording, as long as a decoder is built to match the HOA format to the end system.

The origin of HOA's flexibility and success is also the source of its largest weakness. Sound field information encoded by the spherical harmonic functions. In HOA, all directions must be treated equally and sampled with regular angular resolutions. Sampling is done by calculating or measuring sound waves as they converge on a fixed observation point. Many applications can work well with, some may even require, this constraints on the measurement. However, the spherical harmonic expansion may not be the best choice in many other conceivable scenarios.

One possible situation is exemplified in Fig. 1.10. In this figure, two teams are collaborating through an ultra-realistic teleconference system. Ideally, the users should perceive both rooms as if they were merged into a single, large space. For this, accurate sound localization cues must be conveyed to all participants. Measuring the sound field from one privileged point as done in HOA will result in a high spatial resolution for users seated close to this position, and a lower one for those who are sit farther away. A more appropriate design considers measurements of the sound field done with equal spatial resolution along an axis which closely follows the seat positions.

Up to this point, no attempts have been made to modify HOA for use in this kind of situations. The spherical harmonic decomposition is a very elegant

mathematical result which makes calculations easy and it is, therefore, hard to discard. The main objective of this dissertation is to provide an alternate set of tools which, like the spherical harmonic decomposition in HOA, lead to a scalable and system-agnostic encoding of sound field information.

The present dissertation does not attempt to build the system shown in Fig. 1.10. Many of the engineering problems in the building of such a system lie outside the scope of this research. What this study offers are the mathematical tools to record, encode and reproduce sound fields after the requirement of a spherical geometry, imposed by HOA, has been discarded. Nevertheless, some discussion regarding the practical implementation of a cylindrical microphone array is presented. Transducer misplacement and self-noise conditions are evaluated.

Recording and encoding of sound fields is not the only focus of this study. Investigations on the reproduction of HOA and other sound field encodings are also presented. The focus at this stage lies in trying to facilitate the decoding of the harmonic expansion coefficients when the target system is a loudspeaker array that does not match the geometry assumed during the encoding process. The mainstream adoption of HOA has been significantly hindered by the requirement of a surrounding array of loudspeakers with regular angular separations. In this dissertation, an alternative decoding method is introduced that can work even when this regularity condition is not met.

It is hoped that the results and guidelines laid out in this dissertation will help expand the areas of application for sound field recording and reproduction technologies. Sound field reproduction is still limited to research environments, however, with the proposals advanced here, and the efforts of the many researchers working in the field, the technology is expected to mature to the point where

individuals can enjoy ultra-realistic 3D audio contents at their homes.

CHAPTER II

Mathematical modeling of sound fields

2.1 Overview

The study of the spatial features of sound fields requires the application of some mathematical techniques that may be unfamiliar to readers of different backgrounds. The present chapter reviews the most important mathematical basis upon which the main body of this dissertation stands. Readers familiar with physical acoustics or with an otherwise solid background on the properties of the Helmholtz equation can skim through this chapter and continue to the main research body of this dissertation, starting with Chapter III. While this section attempts to be an adequate introduction to the topics and results used in following chapters, it is not intended to replace the more complete treatment of classical results that can be found in books from authors such as Williams [19] or Teutsch [20].

This chapter starts by reviewing the wave equation and some of its properties. This analysis will lead to two important results that show why sound field reproduction is possible: the Huygens-Fresnel principle and the Kirchhoff-Helmholtz integral theorem. In the process, the sound field will be divided into what are

known as the near and far fields. Discussion will focus on the far field, introducing a way to characterize sound field information in this region: the spherical harmonic decomposition.

2.2 The wave equation

Assuming an ideal propagation medium, the sound pressure field $p(\vec{r}, t)$ can be mathematically described, within a source-free region, by the scalar wave equation:

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

where $\vec{r}$ specifies a spatial point, and t stands for the time coordinate. The constant c is the speed of sound; its value depends on the atmospheric conditions. Unless otherwise stated, this dissertation will use the value of $c = 343 {}^{m}/_{s}$, which corresponds to air at approximately 20 °C. The $\Box$ operator is known as the d'Alembertian, and is a generalization of the Laplacian operator, ∇^2.

Equation (2.1), in general, describes any oscillatory phenomena in space-time. No specific spatial coordinates, such as Cartesian or spherical, have been introduced thus far. The specific choice of coordinates is not done until the Laplacian is explicitly written.

The first step towards the derivation of the solutions to the wave equation consists of separating their spatial and temporal components. The wave equation can be separated using a common multiplicative ansatz. The solutions are assumed to be of the form:

$$p(\vec{r}, t) = \psi(\vec{r})T(t). \tag{2.2}$$

Substituting the ansatz into Eq. (2.1) leads to two separate differential equations. For the temporal component, the resulting equation is that of a simple harmonic

oscillator:

$$\left[\frac{1}{c^2}\cdot\frac{d^2}{dt^2}+k^2\right]T(t)=0. \tag{2.3}$$

The constant of integration that arises from the separation of variables is expressed as k^2. This choice does not lead to any loss of generality in the solutions and will result in more manageable mathematical expressions. Furthermore, the general solutions to the simple harmonic oscillator equation are of the form:

$$T(t)=Ae^{ikc(t-t_0)}, \tag{2.4}$$

with A and t_0 being two constants of integration. It is seen that the number k is inversely proportional to the wavelength of the sound wave. This parameter is known as the angular wavenumber and is associated to the linear frequency of sound, f, by the relationship $k = {}^{2\pi f}/_{c}$.

The temporal part of the wave equation is not particularly interesting. It can be summarized as a change of phase as time progresses. Sound fields, in the absence of sources, evolve in time by simply having their phases cycle.

On the other hand, the spatial portion of Eq. (2.1) shows a more interesting behavior. After separation of variables, the spatial components of the sound field are given by the equation:

$$\left[\nabla^2+k^2\right]\psi(\vec{r})=0. \tag{2.5}$$

This is known as the Helmholtz equation, an eigenvalue problem. The spatial components of any sound field are eigenfunctions of the Helmholtz operator ∇^2+k^2.

The research topics presented in this dissertation deal exclusively with the spatial features of sound fields. The solutions to the temporal part of the wave equation are, therefore, irrelevant and will be ignored from this point. However, it

is useful to keep in mind that the time dependency can be inserted into any results through the application of the evolution operator e^{ikct}.

2.3 The Helmholtz equation

Having removed the time dependency from the analysis of sound fields, this section takes a closer look at the Helmholtz equation, the spatial portion of the wave equation. To solve the wave equation it is necessary to introduce a set of coordinates so as to explicitely express the Laplacian.

The Helmholtz equation can be solved by separation of variables through a multiplicative ansatz in a number of coordinate systems. The appropriate choice of coordinates depends on the particular problem to be tackled. In acoustics, particularly when free-field conditions are assumed, it is common to use the spherical coordinates since sound propagation does not have a priviledged direction. The main reason for this is that sound propagation is isotropic. However, acoustic problems with particular boundary conditions may be easier to solve with a different choice of coordinates.

Expressing the Laplacian explicitely in spherical coordinates leads to the following form of the Helmholtz equation:

(2.6)
$$\left[\frac{1}{r^2}\cdot\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2\sin\theta}\cdot\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\cdot\frac{\partial^2}{\partial\varphi^2}+k^2\right]\psi(r,\theta,\varphi)=0$$

To solve this equation, the spatial component of the sound field, ψ is assumed to be of the form:

(2.7)
$$\psi(r,\theta,\varphi)=j(r)\mathrm{Y}(\theta,\varphi).$$

Substituting this assumption into Eq. (2.6) leads to the following two differential

equations:

$$\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} + \left(k^2 - \frac{n}{r^2}\right)\right] j(r) = 0, \tag{2.8}$$

$$\left[\frac{1}{\sin\varphi}\frac{\partial}{\partial\varphi}\left(\sin\varphi\frac{\partial}{\partial\varphi}\right) + \frac{1}{\sin^2\varphi}\frac{\partial^2}{\partial\theta^2} + n\right] \mathrm{Y}(\theta,\varphi) = 0, \tag{2.9}$$

where n denotes the new constant of integration.

The equation governing the radial component, Eq. (2.8), is a famous expression know as the Bessel's differential equation. Its solutions are the spherical Bessel functions and can be evaluated using the following formula:

$$j_n(kr) = \sum_{z=0}^{\infty} \frac{(-1)^z}{z!\Gamma(z+n+1)}\left(\frac{kr}{2}\right)^{2z+n}. \tag{2.10}$$

In this equation, Γ represents the Gamma function, a generalization of the factorial.

The angular portion can be further separated using the ansatz

$$\mathrm{Y}(\theta,\varphi) = \Theta(\theta)\Phi(\varphi). \tag{2.11}$$

This assumption separates Eq. (2.9) into the following equations:

$$\left[\frac{d^2}{d\theta^2} + n^2\right] \Theta(\theta) = 0, \tag{2.12}$$

$$\left[n\sin^2\varphi + \sin\varphi\frac{d}{d\varphi}\left(\sin\varphi\frac{d}{d\varphi}\right) - m^2\right] \Phi(\varphi) = 0, \tag{2.13}$$

The new separation constant is labeled as m^2 since this choice will, once again, simplify some expressions.

The azimuth angle dependence results in another simple harmonic oscillator. However, an additional constraint is imposed on the solutions. Since θ is an angular coordinate, Θ must be a periodic function and, in concrete, it must repeat itself with a period of 2π. This forces the constant of integration n to take only integer values.

A second constraint on the constants of integration emerges from the consideration that the functions $\mathrm{Y}(\theta, \varphi)$ must be single-valued for elevation angles $\varphi = 0$ and $\varphi = \pi$. Equation (2.13) constraints the parameter m to lie inside the interval $[-n, n]$.

It is possible to transform Eq. (2.13) into the Legendre equation by the change of variable $x = \cos\varphi$. Consequently, the solutions to Eq. (2.13) are the associated Legendre polynomials $\mathrm{P}_{nm}(\cos\varphi)$ which can be calculated using the following formula [21]:

$$\mathrm{P}_{nm}(\cos\varphi) = \frac{(-\sin\varphi)^n}{2^m m!}\left[\frac{d}{d(\cos\varphi)}\right]^{m+n}(-\sin^2\varphi)^m. \tag{2.14}$$

The solutions to Eq. (2.9) are obtained by multiplying the solutions to Eqs. (2.12) and (2.13):

$$\mathrm{Y}_{nm}(\theta, \varphi) = \mathrm{N}_{nm} e^{in\theta} \mathrm{P}_{nm}(\cos\varphi), \tag{2.15}$$

where N^{nm} denotes a normalization constant.

The functions defined by Eq. (2.15) are known as the spherical harmonic functions of degree n and order m. One of their most important properties is that every spherical harmonic function is orthogonal to every other one. They can be made pairwise orthonormal by choosing the normalization [22]

$$\mathrm{N}_{nm} = \sqrt{\frac{(2m+1)}{4\pi}\frac{(m-n)!}{(m+n)!}}. \tag{2.16}$$

By putting together these results, the general solutions to Eq. (2.6) are:

$$\psi(r, \theta, \phi) = \sum_{n=0}^{\infty} j_n(kr) \sum_{m=-n}^{n} B_{nm}(k) \mathrm{Y}_{nm}(\theta, \varphi), \tag{2.17}$$

for all wavenumbers k. This result is valid inside any sourceless region. That means that r must be smaller than the distance to the nearest sound source for any direction.

As a reminder, the symbol j_n stands for the spherical Bessel function of order n, and Y_{mn} denotes the spherical harmonic function of degree n and order m. The constants $B_{mn}(k)$ characterize the boundary conditions for any particular sound field. They are enough to describe the field over a region free of sound sources.

2.4 Near and far fields

The solutions to the Helmholtz equation in spherical coordinates are the product of two special functions: the spherical Bessel functions and the spherical harmonic functions. The spherical Bessel functions consider only the magnitude of the position vector, that is, the effects of distance. Meanwhile, the spherical harmonic functions consider only the direction of said vector. The radial and angular components are explicitely separate, behaving independently and in different ways. In this section, the radial component, that is, the spherical Bessel functions, are considered.

The spherical Bessel functions are closely related to the Bessel functions that appear in other coordinate systems. The following formula can be used to calculate one in terms of the other:

$$j_n(kr) = \sqrt{\frac{\pi}{2kr}} J_{n+\frac{1}{2}}(kr). \tag{2.18}$$

The Bessel functions are represented here by the symbol J.

Given the relationship of Eq. (2.18), many of the properties of the Bessel functions can be directly applied to the spherical Bessel functions after scaling them and adjusting their order by 1/2.

An important result are the asymptotic forms of the Bessel functions. These are approximations to the Bessel functions when their arguments are large. The

asymptotic limit of the Bessel functions is

$$J_n(kr) \approx \frac{1}{\sqrt{2\pi kr}} e^{i\left(kr - \frac{n\pi}{2} - \frac{\pi}{4}\right)}. \tag{2.19}$$

The Bessel functions approximate oscillatory functions scaled by the square root of their argument. Additionaly, the phase of these functions is given by the order. The Bessel functions for orders 1 through 10 are shown in Fig. 2.1. In this figure, the Bessel functions have been delayed by the order-dependent phase that appears in their asymptotic forms. As the argument of the Bessel functions becomes larger, their values become closer to those of Eq. (2.19).

Similarly, the spherical Bessel functions can be approximated, for $kr \gg 1$, as follows:

$$j_n(kr) \approx \frac{1}{2kr} e^{i\left(kr - \frac{n\pi}{2} - \frac{\pi}{2}\right)}. \tag{2.20}$$

The oscillatory functions decay linearly with the argument of the spherical Bessel functions. Besides this, a phase shift of $\pi/2$ is the only difference between the asymptotic forms for both types of Bessel functions.

In the context of sound field analysis, the asymptotic forms of the Bessel functions lead to what is known as the far-field approximation. The region in space where the Bessel functions can be approximated as sinusoidals decaying with distance is called the far field. Removing the complicated Bessel functions and using a simple gain and delay leads to reduced mathematical expressions and, in practical applications, simple filtering schemes. Most sound field processing techniques make use of the far-field approximation.

On the other hand, problems that require a precise treatment of distance cannot use this approximation. For example, processing sound field data when sound sources are close to the observation point requires the full use of the Bessel functions.

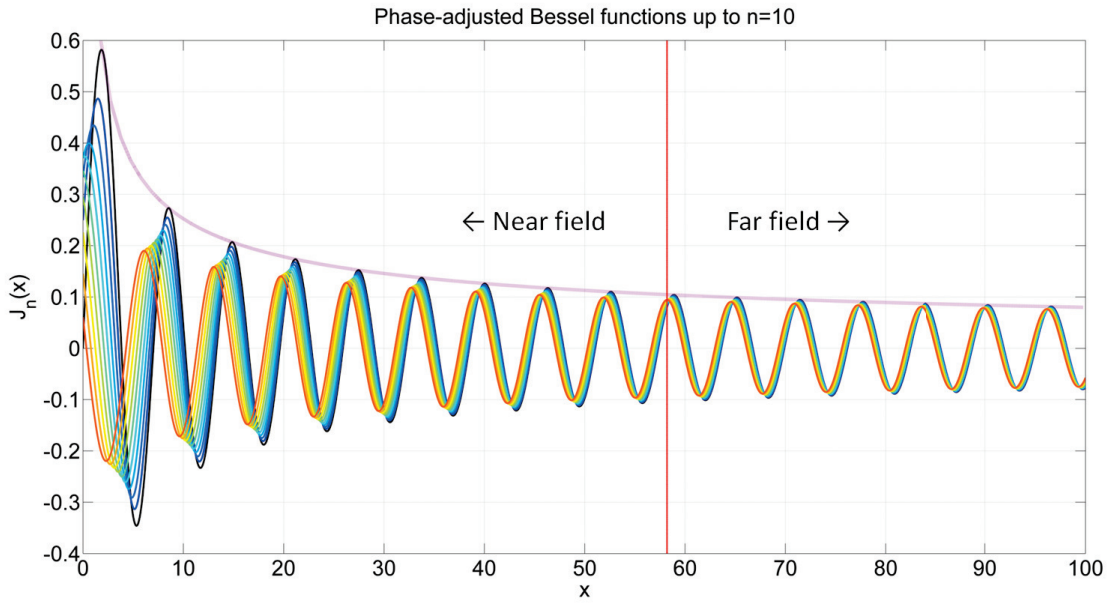


Figure 2.1: The Bessel functions of integer orders $n = 1$ to $n = 10$ after a phase shifting. The functions have been delayed according to the phase of their asymptotic limit. The violet envelope shows the amplitude of said limit. When the argument is small, there is a large variation in the value of the functions, according to their order. However, for large arguments, the functions very closely match their asymptotic limit for all orders. The region where the asymptotic limit is valid is called the far field, while the region where the Bessel functions differ according to their order is called the near field.

The regions of space where this type of complete analysis is required are called the near field.

Throughout this dissertation, as is common in the study of sound fields, most of the problems discussed will first be tackled in the far field before near field corrections are incorporated to get a general solution.

2.5 High-Order Ambisonics

The previous section discusses the radial part of the solutions to the Helmholtz equation. In this section, the angular part of the solutions is discussed. The angular solutions are the spherical harmonic functions of Eq. (2.15). Throughout

this dissertation, the normalization of Eq. (2.16) will be assumed. This choice leads to simplified expressions; however, all results can be scaled to match any different normalization if so is preferred.

The spherical harmonic functions were briefly discussed in Chapter I where they were used to define High-Order Ambisonics. A figure showing the spherical harmonics of degrees 0, 1 and 2 is listed in said chapter as Fig. 1.8. In this section, a more detailed explanation of HOA is presented by properly introducing the spherical harmonic decomposition.

2.5.1 Properties of the spherical harmonic functions

The spherical harmonics are interesting functions with several useful properties. Two of them come from the fact that they are the general solutions to a differential equation, Eq. (2.9). The general solutions to any differential equation form a functional basis for the domain of its solutions [18]. The members of a basis of any vector or functional space are orthogonal and their set is complete.

Any pair of spherical harmonic functions must be orthogonal. Furthermore, if the normalization of Eq. (2.16) is applied in their definition, the spherical harmonics are orthonormal. This means that they satisfy the following equation:

$$\int_{\theta=-\pi}^{\pi}\int_{\varphi=-\pi/2}^{\pi/2} \mathrm{Y}_{nm}(\theta,\varphi)\sin(\varphi)\mathrm{Y}^{*}_{n'm'}(\theta,\varphi)\sin(\varphi)d\varphi d\theta = \delta_{nn'}\delta_{mm'}. \tag{2.21}$$

The symbol δ_{ij} is the Kronecker delta, which takes the value of 1 if $i = j$ and 0 for any other combination of subindices.

Figure 2.2 shows the results of numerical integration for Eq. (2.21) for all pairs of spherical harmonics of degrees 0 to 7. The result should be the identity matrix, with ones on the main diagonal and zeros for all other entries. A tolerance of 10^{-6} was used to carry out the numerical integration. The result is a perfect match

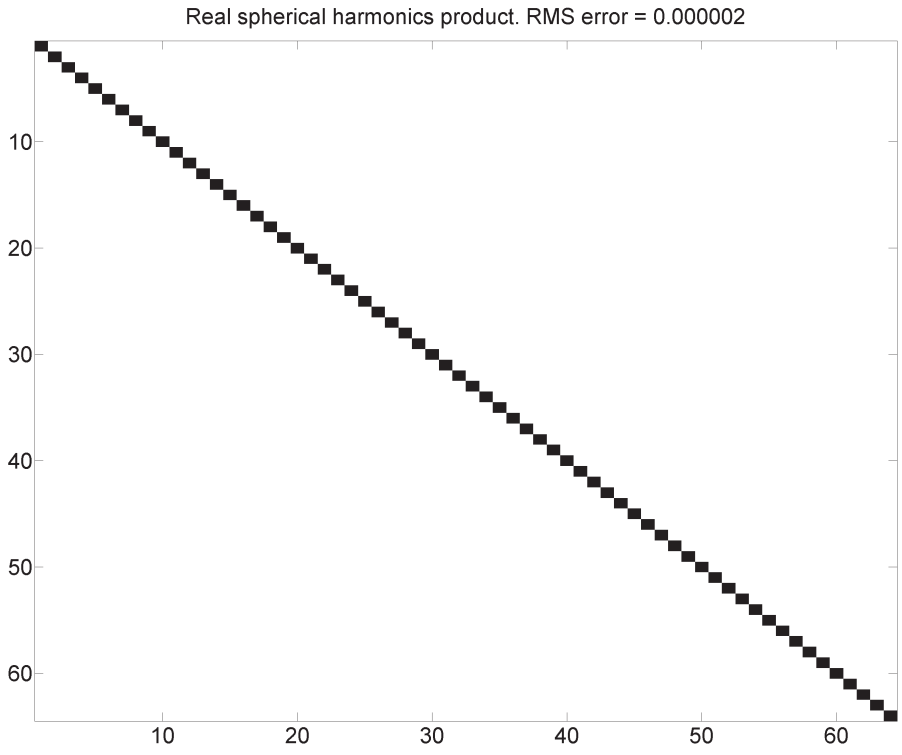


Figure 2.2: Orthogonality test of the spherical harmonic functions. All the spherical harmonics of degrees 0 to 7, 64 in total, were used in the numerical integration of the orthogonality relationship. Integration was carried out with a tolerance of 10^{-6}. The results show no deviation, outside of the tolerance margin, between the numerical computation and the Kronecker delta.

of the 64×64 identity matrix within the tolerance limit.

Orthogonality means that it is impossible to express one spherical harmonic as the weighted sum of all others. On the other hand, the spherical harmonics are complete. This means that any function on their domain, the sphere, can be written as the linear combination of spherical harmonics. The completeness property can be formalized as follows:

$$\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathrm{Y}_{nm}(\theta, \varphi) \mathrm{Y}^{*}_{nm}(\theta', \varphi') = \delta(\theta - \theta')\delta(\varphi - \varphi'). \tag{2.22}$$

The distribution $\delta(x - x')$ is the Dirac delta, which is defined as having a value of

zero for all arguments except zero, where it is undefined. The area under the Dirac delta is also defined to be equal to 1. The Dirac delta can be used to sample any function point by point by shifting it. Therefore, if the spherical harmonic functions can be combined to form a Dirac delta, they can also reconstruct any function over the sphere.

The result of evaluating the completeness formula of Eq. (2.22) for all spherical harmonic functions up to degree $n = 100$ is shown in Fig. 2.3. The large peak at the center of the distribution and small values at all other points correspond well with the Dirac delta. However, some ripples can be observed around the central peak. These are the effect of truncating the infinite summation of spherical harmonics at degree 100. Continuing the summation makes the ripples smaller compared to the peak; however, the never fully disappear since it is impossible to evaluate an infinite sum.

2.5.2 Spherical harmonic decomposition

The spherical harmonic functions are orthogonal and complete. They define a basis for all functions on the sphere. In the context of sound field recording and reproduction, High-Order Ambisonics exploits this property to define a scalable and system-agnostic encoding of sound fields.

In HOA, the far-field approximation is applied to remove the Bessel functions from the spatial description of sound fields, the solutions to the Helmholtz equation. Furthermore, the sound field is sampled only on the surface of a sphere; this means that the distance attenuation and delay of Eq. (2.20) are the same for all measurements. They can be treated as a global system delay and gain and, therefore, ignored in the encoding of the sound field.

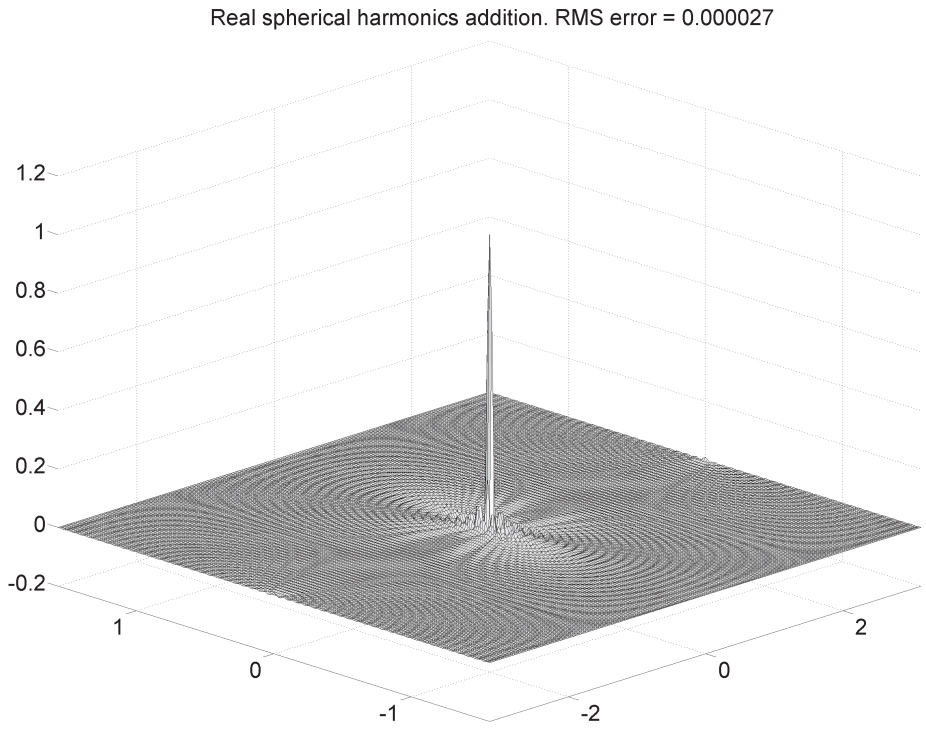


Figure 2.3: Completeness of the spherical harmonic functions. A Dirac delta distribution is approximated by the spherical harmonics of degree 0 to 100. A large peak at the origin and near-zero values elsewhere show a good correspondence with the Dirac delta; however, some ripples are visible near the peak. These are the result of truncating the summation of spherical harmonics at degree 100; only when all degrees up to infinity are considered does the distribution equals the Dirac delta.

The sound field measurements used in HOA define a sound pressure distribution on the sphere. It is, thus, susceptible to being expressed as the linear combination of spherical harmonic functions.

$$\psi^{[k,r]}(\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathrm{B}_{nm}^{[k,r]} \mathrm{Y}_{nm}(\theta,\varphi). \tag{2.23}$$

The superindex $[k,r]$ is used to explicitely note that the results are valid only at a fixed radius and for a given wavenumber. Nevertheless, the sound field ψ for these conditions is fully characterized by the expansion coefficients B.

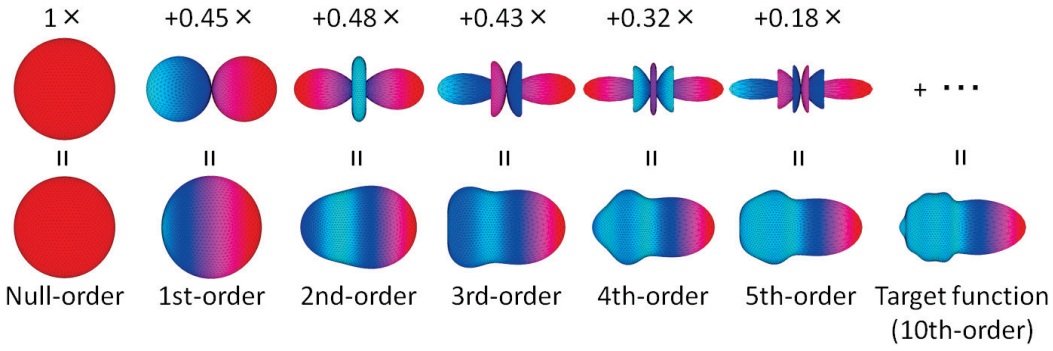


Figure 2.4: Illustration of the spherical harmonic decomposition. A function on the sphere is approximated by the sum of low-degree spherical harmonics. The approximation improves monotonically as more spherical harmonic functions are considered.

Equation (2.23) is known as the spherical harmonic expansion. It is justified by the completeness property of the spherical harmonic functions. Furthermore, it is, in essence, the same expression that was introduced in Chapter I as Eq. (1.5); however, only the spatial components of the sound field are considered here.

The spherical harmonic expansion defines a system-agnostic and scalable encoding of the sound field. The coefficients B can be calculated for all spherical harmonics of any degree and order. The summation can be truncated at any value deemed high enough to ensure the spatial resolution that a given application requires. However, Eq. (2.23) does not indicate how to measure these expansion coefficients. The inverse of this expression is needed in order to record actual sound fields.

It is straightforward to invert Eq. (2.23) by using the properties of the spherical harmonic functions. In concrete, the spherical harmonic $\mathrm{Y}^*_{n'm'}(\theta', \varphi')$ is multiplied on both sides of the equation. After integrating over the sphere, and simplifying the expressions using Eqs. (2.21) and (2.22), the result is

$$\mathrm{B}^{[k,r]}_{nm} = \int_{\theta=-\pi}^{\pi} \int_{\varphi=-\pi/2}^{\pi/2} \psi^{[k,r]}(\theta, \varphi) \mathrm{Y}^*_{nm}(\theta, \varphi) \sin(\varphi) d\varphi d\theta. \tag{2.24}$$

Recordings of the sound field made on the surface of a sphere can be encoded as their spherical harmonic expansion using Eq. (2.24). This equation defines the High-Order Ambisonics encoding of a sound field recorded using a spherical microphone array.

2.6 Huygens-Fresnel principle

Sound propagation, as described by Eq. (2.1) does not consider the presence of sound sources. To include them in the analysis it is necessary to consider the inhomogeneous equation

$$\Box p = f(\vec{r}, t), \tag{2.25}$$

where $f(\vec{r}, t)$ is any function characterizing the sound sources present. A general solution to this equation can be given by defining a Green's function [21]. The Green's function is the impulse response of the homogeneous differential equation being considered, in this case Eq. (2.25).

The Green's function for an eigenvalue problem, such as Eq. (2.1) or Eq. (2.5) can be calculated as a series in terms of the eigenfunctions that satisfy it [22]:

$$G(\vec{r}, \vec{r_0}) = \sum_{j=0}^{\infty} \frac{\psi^*_{nm}(\vec{r})\psi_{nm}(\vec{r_0})}{\lambda_{nm}}, \tag{2.26}$$

where λ_{nm} is the eigenvalue associated to the eigenfunction ψ_{nm}.

The Green's functions associated with both, the d'Alembert and Helmholtz operators, can be stated explicitely as follows [18]:

$$G(\vec{r}, \vec{r_0}) = \frac{\delta(t - \frac{|\vec{r} - \vec{r_0}|}{c})}{4\pi|\vec{r} - \vec{r_0}|}, \tag{2.27}$$

for the d'Alembert operator, and

$$G(\vec{r}, \vec{r_0}) = -\frac{e^{-ik|\vec{r} - \vec{r_0}|}}{4\pi|\vec{r} - \vec{r_0}|}, \tag{2.28}$$

for the Helmholtz operator.

The Green's function for both operators depends only on the distance between the two observation points. The direction in which sound propagates to reach $\vec{r}$ from $\vec{r_0}$ is not important. This formalizes the fact that sound propagation is isotropic, that is, sound propagates equally in all directions in space.

Any disturbance $f(\vec{r}, t)$ to the homogeneous equations Eq. (2.1) and (2.5) will propagate in all radial directions from its position of occurrence, causing new disturbances in other contiguous regions. These will again propagate radially and all contributions will add up to form the wavefronts corresponding to the original source $f(\vec{r}, t)$. This is known as the Huygens-Fresnel principle [19].

An alternative way to introduce the Huygens-Fresnel principle is by noting that the wave equation assumes that the propagation medium is linear. The value of the sound pressure at $\vec{r_s}$ caused by a sound source of angular wavnumber k and complex amplitude U located at $\vec{r_0}$ is given by the Green's function as:

$$p_k(\vec{r_s}) = U\frac{e^{ik|\vec{r_s}-\vec{r_0}|}}{|\vec{r_s}-\vec{r_0}|}. \tag{2.29}$$

The sound pressure at a different, more distant point $\vec{r}$ can be calculated not only from its distance to $\vec{r_0}$ by using Eq. (2.29), but also from the sound pressure at all points on a sphere of radius $|\vec{r_0}|$. Since the medium is linear, propagation to $\vec{r}$ is equivalent to the sum of all propagation paths going from $\vec{r_0}$ to this sphere enclosing the sound source, and then from the sphere's surface to $\vec{r}$. The result of carrying out this process is the formalization of the Huygens-Fresnel principle [19]:

$$p_k(\vec{r}) = \left[U\frac{e^{ik|\vec{r_s}-\vec{r_0}|}}{|\vec{r_s}-\vec{r_0}|}\right]\left[-\frac{ik}{4\pi}\int_{\theta=-\pi}^{\pi}\int_{\varphi=-\pi/2}^{\pi/2}\frac{e^{ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\left(1+\frac{\vec{r_0}\cdot(\vec{r}-\vec{r_s})}{|\vec{r_0}||\vec{r}-\vec{r_s}|}\right)\sin(\varphi)d\varphi d\theta\right]. \tag{2.30}$$

The angular dependency of the integrand, although not explicitely stated, occurs in

vector $\vec{r_s}$ which, in spherical coordinates, is given by (r_s, θ, φ).

Equation (2.30) explains why sound field reproduction systems are possible. The wavefronts corresponding to a given sound source can be produced not only by said source, but also by a set of *secondary sources* covering a boundary that fully separates space into two regions, one containing all sound sources and one containing all observation points.

2.7 Kirchhoff-Helmholtz integral theorem

The previous section derived an expression for the Huygens-Fresnel principle by considering a limiting surface and dividing sound propagation into two linear processes, one inside and one outside this limit. An alternative is to first consider propagation to all points inside a region of interest R and then onto the observation point. The result for an arbitrary perturbation $f(\vec{r_0})$ can be given in terms of the Green's functions as follows [19]:

$$p_k(\vec{r}) = \int_{\mathrm{R}} f(\vec{r_0}) \nabla \cdot \nabla G(\vec{r}, \vec{r_0}) + G(\vec{r}, \vec{r_0}) \nabla \cdot \nabla f(\vec{r_0}) dV. \tag{2.31}$$

This result can be simplified by making use of a result known as the second Green's identity [18]:

$$p_k(\vec{r}) = \oint_{\mathrm{S}} f(\vec{r_0}) \frac{\partial G(\vec{r}, \vec{r_0})}{\partial \hat{r_s}} + G(\vec{r}, \vec{r_0}) \frac{\partial f(\vec{r_0})}{\partial \hat{r_s}} d\mathrm{S}. \tag{2.32}$$

The new closed integral is carried out over a surface enclosing the region of interest R and the result is only valid if all sound sources are located inside this region.

Equation (2.32) is known as the Kirchhoff-Helmholtz integral theorem. It is equivalent to the Huygens-Fresnel principle introduced in the previous section and, similarly, it expresses the sound field caused by an arbitrary set of sound sources in terms of the sound pressure observed on a surface enclosing them.

The second term in the Kirchhoff-Helmholtz integral contains the derivative of the sound pressure. This term is sometimes considered problematic to sound field reproduction applications due to the difficulty to measure the derivative, also known as the particle velocity. Fortunately, in free field conditions it is possible to ignore this term by applying a 90-degrees phase shift and a $1/\lambda$ weight to all of the secondary sources. This result arises from comparing Eq. (2.32) and the equivalent Eq. (2.30). The scaling and phase shift is the result of multiplying by the coefficient ik, as it appears explicitely in the latter equation.

2.8 Summary

This chapter presented a review of classical results surrounding the wave equation. The time component of sound fields was shown to be a phase shift and removed from the analysis due to its simplicity. All results derived later in this chapter and in upcoming sections of this dissertation are valid at one instant in time, but can be easily generalized by applying the phase-shifting evolution operator e^{ikct}.

The spatial components of the sound field show a more interesting behavior governed by the Helmholtz equation. Explicit solutions were derived in spherical coordinates and used to introduce important concepts such as the near and far fields and the spherical harmonic decomposition. The latter is the core of High-Order Ambisonics, a sound field reproduction technique that defines a scalable and system-agnostic encoding of sound field information.

Two important results in the theory of the wave equation, the Huygens-Fresnel principle and the Kirchhoff-Helmholtz integral theorem were introduced. Together, they explain why sound field reproduction, that is, re-creating

the sound field of an arbitrary sound source using loudspeaker arrays, is possible. These results form the basis of all sound field reproduction systems discussed in Chapter I, and those to be presented in the following chapters of this dissertation.

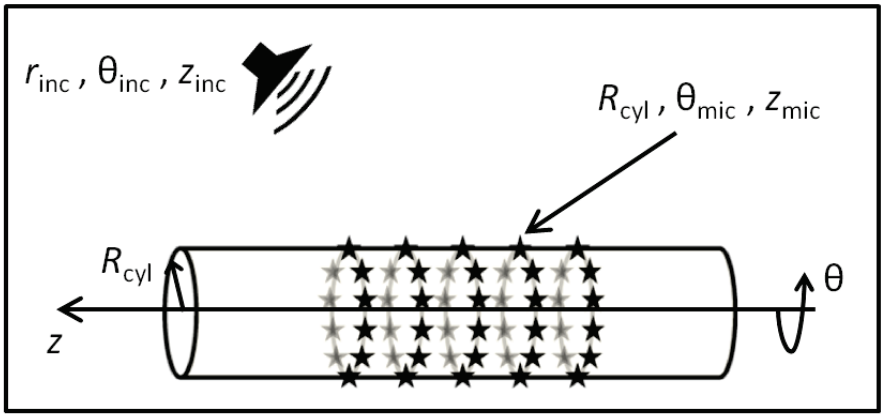
CHAPTER III

Plane wave decomposition of cylindrical sound pressure distributions

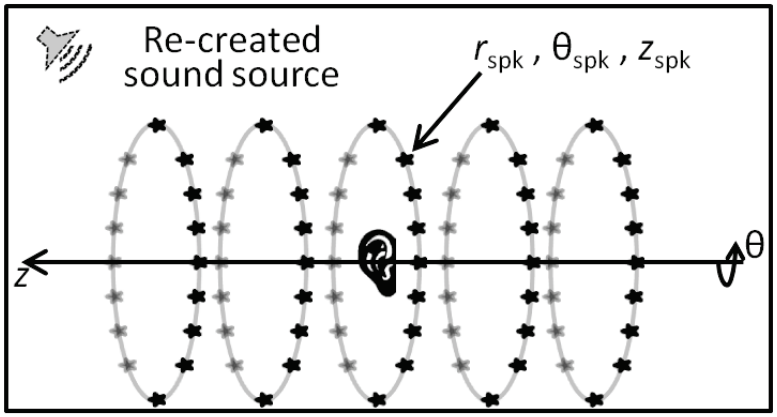
3.1 Overview

The previous chapter presented the solutions to the Helmholtz equation in spherical coordinates and used the results to formulate High-Order Ambisonics. However, the spherical geometry is not always the best choice for sampling a sound field. In particular, spherical coordinates impose a privileged point, the origin, from which all observations are made. This may be adequate to present spatial sound to a single, static listener. On the other hand, if systems to present sounds to a larger audience, or to allow the listener to move around, are to be designed, a different choice of coordinates may be more effective.

In this chapter, the cylindrical coordinates are used as a basis to observe and eventually encode sound fields. The cylindrical geometry does not present a privileged observation point; rather, it exhibits a privileged axis. This feature makes it a better choice when presenting spatial sound to a large audience who can be aligned along said axis.



(a) Recording array



(b) Reproduction array

Figure 3.1: Sound field recording and reproduction systems. (a) Microphones are distributed on the surface of a rigid cylinder as a set of parallel rings. (b) Sound fields are reproduced using a loudspeaker array surrounding the listener.

3.2 Cylindrical microphone and loudspeaker arrays

Spherical microphone arrays are commonly used to measure sound field information. Their signals can be encoded by the spherical harmonic decomposition; this forms the basis for sound field recording and reproduction technologies such as High-Order Ambisonics (HOA). Cylindrical arrays are a better choice to sample sound fields over an extended region at a fixed height.

The sound field recording and reproduction system assumed in this chapter is shown in Fig. 3.1. The microphones are aligned on the surface of a rigid cylinder of radius R_{cyl} as equidistant rings and staves. The axis of the cylinder should be parallel to the region of interest, such as the stage, and would normally lie on the horizontal plane. The microphones along each stave can be used to infer azimuth angles of incidence, while the rings can be used to approximate the elevation of the sound sources. The rigid cylinder acts as a baffle, making it easier to sense spatial information and, in general, leading to a more robust system. The merits of cylindrical baffles have been explored in the design of arrays for circular hamonics beamforming [31], and recording systems that would otherwise require directional microphones [32].

3.3 The Helmholtz equation in cylindrical coordinates

The Helmholtz equation in cylindrical coordinates can be written as follows:

$$\left[\frac{1}{r}\cdot\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\cdot\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2}+k^2\right]\psi(r,\theta,z)=0. \tag{3.1}$$

It can be solved by separation of variables following the same approach used in Chapter II to find the solutions in spherical coordinates. The ansatz used to separate the equation into the three spatial coordinates is:

$$\psi(r, \theta, z) = J(r) \cdot \Theta(\theta) \cdot Z(z). \tag{3.2}$$

Applying this ansatz to Eq. (3.1) leads to three separate ordinary differential equations that can be solved independently.

3.3.1 The radial component

The ordinary differential equation governing the radial component of the Helmholtz equation in cylindrical coordinates is:

$$\left[\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + \left(k_r^2 - \frac{n^2}{r^2}\right)\right] J(r) = 0. \tag{3.3}$$

The equation above is similar to Eq. (2.8) explored in Chapter II. This is the Bessel equation and its solutions, the Bessel functions, satisfy important properties such as the far field approximation, discussed in the previous chapter.

The Bessel functions $J_n(k_r r)$ should not be confused with the spherical Bessel functions $j_n(kr)$ that result from considering the Helmholtz equation in spherical coordinates. The two functions are closely related be the formula presented in Eq. (2.18). However, care must be taken to evaluate the correct functions depending on the coordinate system used.

The two separation constants n and k_r will be referred to as the *polar order* and the *radial wavenumber*, respectively. The polar order is related to the polar angle and is similar to the Ambisonic order used in High-Order Ambisonics. Meanwhile, the radial wavenumber is the result of projecting the wavevector onto the radial coordiante. It is given in terms of the wavenumber k and the azimuthal angle of

incidence ϕ_{inc} by the following expression:

$$k_r = k\sin(\phi_{\text{inc}}). \tag{3.4}$$

Similarly, one can project the wavevector onto the axial coordinate to obtain the complementary *axial wavenumber* given as:

$$k_z = k\cos(\phi_{\text{inc}}). \tag{3.5}$$

Both projections of the wavevector are, of course, related by the identity:

$$k_r^2 + k_z^2 = k^2. \tag{3.6}$$

In the following sections, these relationships are expressed in terms of a parameter $\xi = \sin(\phi_{\text{inc}})$, called a *damping ratio* due to its role in the differential equation for the axial coordinate.

The encoding of distance along the radial coordinate will be explored in detail in Chapter VI. For now, the analysis of sound fields in cylindrical coordinates will assume the far-field approximation introduced in Chapter II and will focus on the polar and axial coordinates instead.

3.3.2 The polar and axial components

The ansatz of Eq. (3.2) leads to two separate differential equations for the polar and axial components of the solutions to the Helmholtz equation in cylindrical coordinates. These two equations are not coupled, as was the case in spherical coordinates where they define the spherical harmonic fuctions. The solutions along each coordinate can be expressed separately.

The differential equation governing the solutions along the polar angle is

$$\left[\frac{d^2}{d\theta^2} + n^2\right]\Theta(\theta) = 0. \tag{3.7}$$

This is the equation for a simple harmonic oscillator. Its solutions are sinusoidal waves expressed by the following formula

$$\Theta(\theta) = \Theta_0 e^{\pm in\theta}. \tag{3.8}$$

Here, Θ_0 is a constant of integration and, in general, a complex number. It codifies the amplitude and phase for the sinusoid of wavenumber n. Unlike the common solutions to a harmonic oscillator, however, it must be noted that the coordinate θ, the polar angle, is cyclic. The solutions of Eq. (3.8) must satisfy the constrain $\Theta(\theta) = \Theta(\theta \pm 2a\pi)$ for all polar angles θ and natural numbers a. The constrain forces the parameter n to be a natural number. Due to its similarity to the Ambisonic order in the spherical case, this parameter will be referred to as the *polar order*.

Equation (3.8) for all naturals n solves the polar component of the Helmholtz equation. The remaining component, the axial coordinate, is governed by the following differential equation:

$$\left[\frac{d^2}{dz^2} + \left(k^2 - k_r^2\right)\right] Z(z) = 0. \tag{3.9}$$

The equation is similar in structure to the one found for the polar angle. However, the constant gain $(k^2 - k_r^2)$ can take both positive and negative values, unlike the always-positive n^2. This particular differential equation is also well-known since it describes the dynamics of a damped harmonic oscillator. Its general solutions are given as follows:

$$Z(z) = Z_0 e^{\pm iz\sqrt{k^2 - k_r^2}} = Z_0 e^{\pm ikz\sqrt{1-\xi^2}}. \tag{3.10}$$

The constant of integration Z_0 corresponds to the complex amplitude for each of the waves forming a general solution. Meanwhile, parameter ξ, known as the *damping ratio* determines the general behavior of the solutions. This parameter is also present

in the radial portion of the solutions to the Helmholtz equation, as presented in the previous subsection.

In the discussion around the radial component of the solutions to the Helmholtz equation, it was noted that the value of the damping ratio ξ indicates the relationship between the axial and radial wavenumbers k_z and k_r. From Eq. (3.10) it is further apparent that its value can also change the axial solutions from sinusoidals to exponential functions.

The general solutions to the Helmholtz equation, ignoring the radial component, are given by the product of Eqs. (3.8) and (3.10). In analogy to the results obtained in the spherical case, this result will be referred to as the *cylindrical harmonic functions*. They are given by the following expression:

$$\mathrm{Z}^{\pm}_{n,\xi}(k;\theta,z) = \mathrm{N}^{\pm}_{n,\xi} e^{\pm in\theta} \cdot e^{\pm ikz\sqrt{1-\xi^2}}. \tag{3.11}$$

The constant $\mathrm{N}_{n,\xi}$ is an arbitrary normalization constant. Unlike the spherical harmonics, the cylindrical harmonics depend on the frequency of the sound source since the wavenumber k appears in the axial-related exponential. Furthermore, the value of the damping ration ξ determines the general behavior of these functions.

There are three types of spherical harmonics, classified according to their damping ratios. The *polar harmonics* occur when $\xi = 1$. This condition results in functions that are independent of the axial coordinate. The polar harmonics are constant along the $z-$axis. A damping ratio smaller than 1 results in oscillatory behavior along the axial coordinate. The result in this case is referred to as the *axial harmonics*. Finally, when the damping ratio is larger than 1, the resulting *damped harmonics* have an exponential behavior along the $z-$coordinate. Examples for all three cases are shown in Fig. 3.2. Two variants of the damped harmonics, differing in their orientations, are shown.

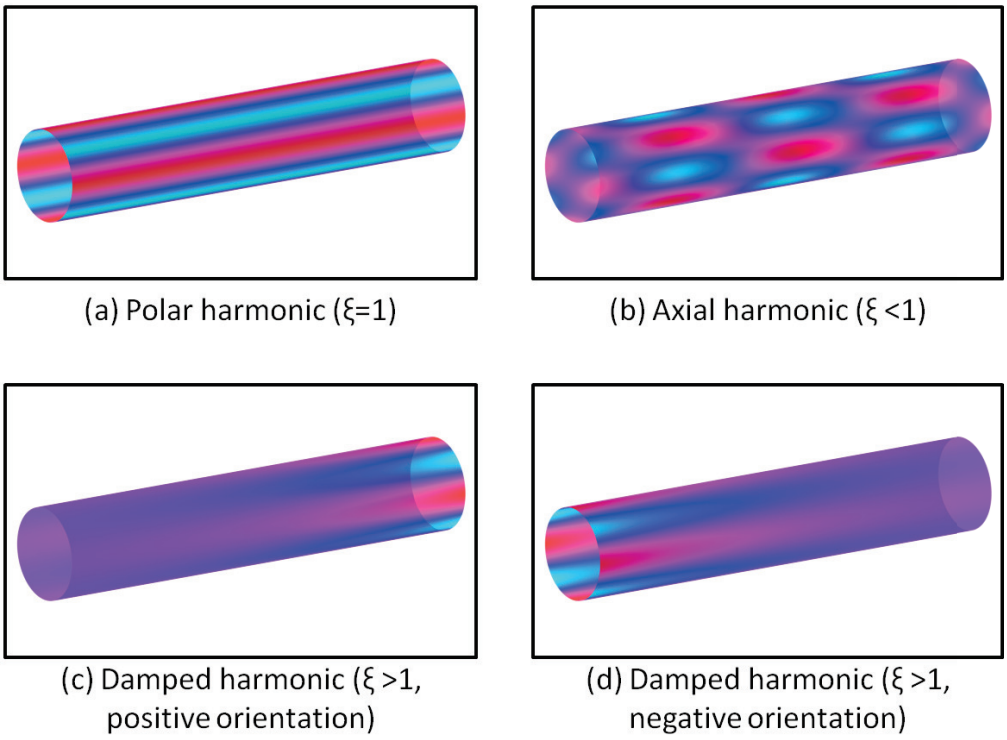


Figure 3.2: The cylindrical harmonic functions.

3.4 Plane-wave decomposition for cylindrical microphone arrays

A simple way to capture sound field information using microphone arrays is to consider the set of plane waves, incident from all directions, that would best explain the observed sound pressure. This is known as the plane-wave decomposition. This approach has been discussed to some extent in previous research for several microphone array geometries. In the cylindrical case, a least-squares solution considering only the polar coordinate has been proposed [23]. A generalization of this solution to include the axial coordinate is introduced in this section.

The process to compute the plane-wave decomposition consists of finding weights to represent the microphone measurements as a linear combination of plane waves $\mathrm{F}(k, \theta_{\mathrm{mic}}, z_{\mathrm{mic}}, \theta_{\mathrm{inc}}, \phi_{\mathrm{inc}})$ arriving from all directions. The starting point is to express the effects that a single, arbitrary plane wave has on the observations made at the microphone positions. The result of this calculation is given by the following expression:

$$\mathrm{F}(k, \theta_{\mathrm{mic}}, z_{\mathrm{mic}}, \theta_{\mathrm{inc}}, \phi_{\mathrm{inc}}) = \frac{i}{\pi^2 k \sin(\phi_{\mathrm{inc}}) R_{\mathrm{cyl}}} \sum_{n=-\infty}^{\infty} \frac{i^n}{\mathrm{H}'_n(k \sin(\phi_{\mathrm{inc}}) R_{\mathrm{cyl}})} e^{-in(\theta_{\mathrm{inc}} - \theta_{\mathrm{mic}})} e^{ik \cos(\phi_{\mathrm{inc}}) z_{\mathrm{mic}}}. \tag{3.12}$$

Here, θ_{inc} and ϕ_{inc} determine the direction of incidence. Meanwhile, θ_{mic} and z_{mic} are the polar and axial coordinates for a microphone in the array. The wavenumber is denoted by k; H'_n stands for the first derivative of the Hankel functions of the first kind and order n.

Equation (3.12) is in the form of a linear combination of the cylindrical harmonic functions introduced in Eq. (3.11). The harmonics used by this expression in particular are evaluated at angle $\theta_{\mathrm{inc}} - \theta_{\mathrm{mic}}$ and axial coordinate z_{mic}. The

cylindrical harmonics are evaluated for all polar orders n and both, positive and negative, orientations. However, the damping ratio is fixed in this expansion. The axial-related exponential leads to the following expression:

$$\sqrt{1-\xi^2} = \cos(\phi_{\text{inc}}). \tag{3.13}$$

The cosine function takes values in the interval $[-1, 1]$ which translates into a damping ratio between 0 and 1. This covers the polar and axial cylindrical harmonics, but excludes the damped harmonics. From this result it can be concluded that an arbitrary field that can be expressed as a superposition of plane waves is encoded exclusively by the polar and axial harmonics. The damped harmonics, therefore, carry the information that cannot be expressed in terms of superposed plane waves. This is called the *evanescent field*.

The plane wave decomposition provides a simple representation of the sound field. It is scalable and system-agnostic. However, it cannot characterize localized sound sources such as monopoles. These type of sources require weights that include an imaginary part; that is, an evanescent field contribution which cannot be encoded without the damped cylindrical harmonics. In general, information regarding the distance to the sound sources is partially lost in the plane-wave decomposition since it ignores the radial coordinate.

3.5 Summary

This chapter introduces the cylindrical geometry for the design of microphone and loudspeaker arrays. To this end, the solutions to the Helmholtz equation are derived in cylindrical coordinates. A significant result found in these solutions is the set of functions known as the cylindrical harmonics.

The cylindrical harmonic functions will eventually become a crucial building block for the encoding of sound fields recorded with cylindrical microphone arrays. In this sense, they are similar in importance to the spherical harmonics which lie at the core of HOA.

Finally, a least-squares approach to encode the sound fields recorded by cylindrical microphone arrays is presented. This method works only in the far field; it does not consider distance information. However, the mathematical techniques used to define this encoding, known as the plane wave expansion, can be applied to more complex encodings which include distance information. This will be one of the main topics of this dissertation and is elaborated in the following chapters.

CHAPTER IV

Mixed-Order Ambisonics for cylindrical microphone arrays

4.1 Overview

In this chapter, a new proposal is advanced to tackle the problem of efficiently encoding sound fields recorded with a cylindrical microphone array. The most salient advantage of the proposal laid out in the following sections is that the encoded sound field information can be broadcast to multiple users using different reproduction systems. That is, a system-agnostic encoding format is used. On the listener's end, a decoder is applied to generate suitable signals for his specific loudspeaker system.

The proposal introduced here does, however, make use of the spherical harmonic expansion. The reason for this is its widespread use in sound field reproduction research. Despite this, the advantages of using a cylindrical microphone array, in particular the ability to allocate different resolutions for azimuth and elevation, are throughly exploited. For this purpose, a variation of High-Order Ambisonics, known as Mixed-Order Ambisonics (MOA) will be applied.

The problem of using cylindrical microphone arrays to record and encode sound field information without restricting the sound field description to a privileged

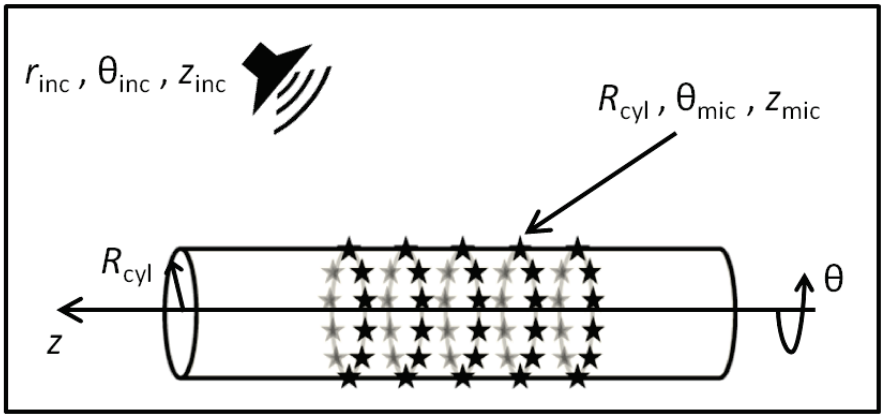


Figure 4.1: Schematic of a cylindrical microphone array. The microphones are distributed on the surface of a cylindrical baffle as a set of equidistant parallel rings. If the array is installed with the cylinder's axis in the horizontal plane, the number of microphones per stave will determine azimuthal resolution, while the number of microphones per ring will set the resolution for elevation.

observation point will be consider in Chapter VI.

4.2 Sound field recording with cylindrical arrays

Sound field recording and reproduction systems promise to deliver unprecedented levels of realism and immersion when presenting auditory information. Their main disadvantage, system complexity, can be ameliorated by an efficient choice of geometry. Specifically, a cylindrical geometry allows for an independent control of the horizontal and vertical accuracy when characterizing and rendering sound fields. The total amount of information to be stored or transmitted can be reduced, without significantly affecting perceptual quality, by limiting the vertical resolution.

Systems that employ a cylindrical geometry to capture sound field information have been proposed in the past [33]. However, current approaches require a target loudspeaker configuration to be fixed in their formulation. Spatial accuracy is fixed and all listeners must deploy the same reproduction system to enjoy a given

broadcast.

The present section provides a brief overview of existing systems and concludes by listing some of their most salient limitations which the proposal advanced in this dissertation, outlined in the next section, seeks to overcome.

4.2.1 Existing recording systems

The recording of sound fields using cylindrical microphone arrays has been explored in the past [34, 23, 32]. All current approaches involve lining up parallel rings of equidistant microphones over the surface of a rigid cylinder, as shown in Fig. 4.1. If the cylinder is placed horizontally, the number of microphones per ring will determine the spatial accuracy for elevation, while the number of rings governs the recording precision in the horizontal plane.

The microphones are placed over a rigid cylinder to enhance their spatial accuracy. The cylinder will partially block sound from a given sound source to contralateral microphones, while the sound field measured by ipsilateral ones remains mostly unaffected. The scattering effects of a rigid cylinder, including the shadowing of a sound source, are illustrated in Fig. 4.2.

Use of a scattering body can improve accuracy by magnifying the effects of the direction of arrival on the sound pressure sensed by the microphones. The end result is a better use of the microphones' dynamic range, as well as the suppression of forbidden frequencies that would arise without the scatterer. However, a side effect is the distortion of the original sound field. Existing recording systems take this into account and attempt to remove the scattering effects after the recording has been made. This is done by modeling the total sound field as a sum of an incident and an scattered field:

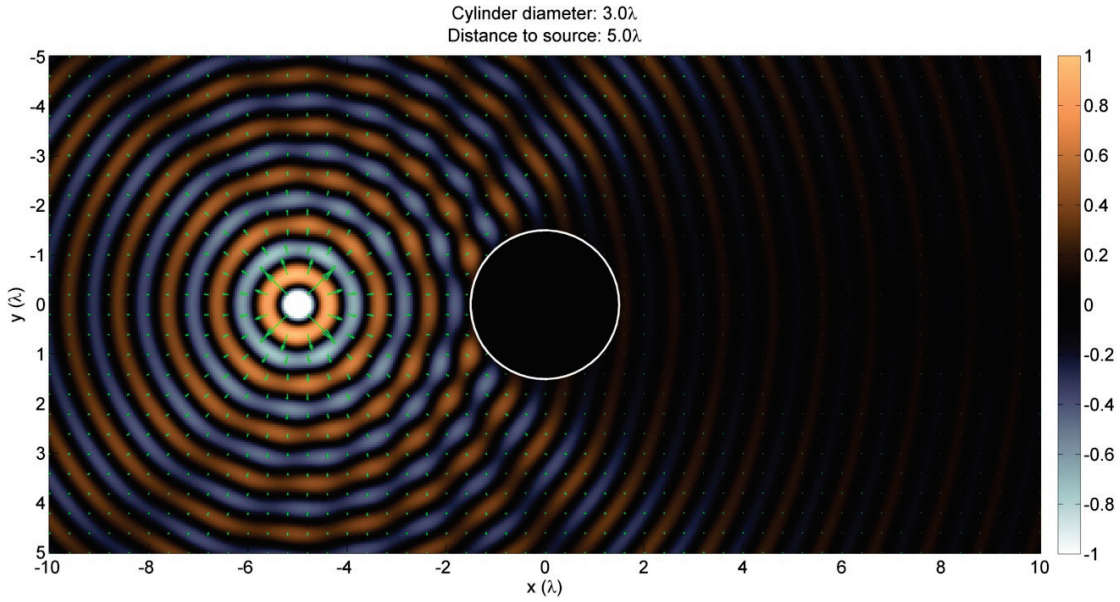


Figure 4.2: Scattering of a spherical wave by a rigid cylinder. The sound waves are perfectly reflected as they reach the surface of the cylindrical baffle causing ripples in the otherwise spherical wavefronts. Sound must diffract around the baffle to reach the side opposite to the sound source, causing a shadow effect.

$$\psi_{total}(k, \vec{r}) = \psi_{incident}(k, \vec{r}) + \psi_{scattered}(k, \vec{r}). \tag{4.1}$$

The wavenumber is denoted by the symbol k, while vector $\vec{r}$ determines the spatial point where the sound field is evaluated.

An estimate of the scattered field $\psi_{scattered}$ is calculated to recover the original sound field $\psi_{incident}$ from the recorded one psi_{total}. This generally assumes the original field can be characterized as a superposition of plane waves [23, 32]. A simplified model of a rigid cylinder of infinite length is also used in this stage. The result is the following expression for the total sound field, measured by the microphones, when the incident field consists of a plane wave incident from an elevation angle θ_{inc} and azimuth angle ϕ_{inc}:

$$\psi_{total}(k,\theta,z) = \frac{i}{\pi^2 k \sin\phi_{inc} R_{cyl}} \sum_{n=-\infty}^{\infty} \frac{i^n}{H_n^{(1)'}(k\sin\phi_{inc} R_{cyl})} e^{in(\theta-\theta_{inc})} e^{ik\sin\phi_{inc} z}. \tag{4.2}$$

R_{cyl} stands for the radius of the cylinder, and $H_n^{(1)'}$ is the derivative of the Hankel function of order n.

Other systems consider the sound field produced by a monopole source located outside of the cylinder, at a distance r from its axis [34]. In this case, the expression for the total field is:

$$\psi_{total}(k,\theta,z) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{in(\theta-\theta_{inc})} \int_{-\infty}^{\infty} \frac{-H_n^{(1)}(k_r r_{inc})}{2\pi k_r R_{cyl} H_n^{(1)'}(k_r R_{cyl})} e^{ik_z(z-z_{inc})} dk_z. \tag{4.3}$$

Here, k_z and n correspond to the coordinates for the helical wave spectrum space, while k_r denotes the projection of the wavevector onto the radial direction. In practice, the sum and integral cannot be computed over the full $(-\infty, \infty)$ intervals. The cutoffs used for these coordinates can be chosen independently, with higher cutoffs leading to greater accuracy at the expense of increased system complexity. The present document will refer to these cutoffs as the *angular and axial orders*, for n and k_z respectively. The fixed radius of the cylinder constrains the radial component of the wavevector as $k_r = \sqrt{k^2 - k_z^2}$.

Irrespective of the scattering model used, current systems attempt to recover the helical wave spectrum of the incident field from the expressions above. In the more general case considering monopole sources, this can be done using the following equation [34]:

$$\tilde{\Psi}_{incident,n}(k,k_z) = -\frac{i}{2}\pi k_r R_{cyl} J_n(k_r R_{cyl}) H_n^{(1)'}(k_r R_{cyl}) \Psi_{total,n}(k,k_z). \tag{4.4}$$

4.2.2 Existing reproduction systems

Sound fields recorded using the techniques described above can be reproduced using loudspeaker arrays. In general, present systems assume a similar geometry for both the recording and reproduction stages. The task of generating loudspeaker signals is, therefore, that of projecting the sound pressure field over the microphone cylinder onto the loudspeaker one.

Once the helical wave spectrum of the original field is known at the microphone positions, a simple and stable set of filters can be applied to project it to the loudspeaker ones. The filters were derived in [34] as:

$$\tilde{F}_n(k, k_z) = -\frac{k_r R_{cyl} H_n^{(1)'}(k_r R_{cyl})}{R_{spk} H_n^{(1)}(k_r R_{spk})} \tag{4.5}$$

They depend on the temporal and axial wavenumbers, k and k_z, only through the radial one, k_r. The filters are determined exclusively from the radius of the recording and reproduction cylinders, R_{cyl} and R_{spk}, respectively.

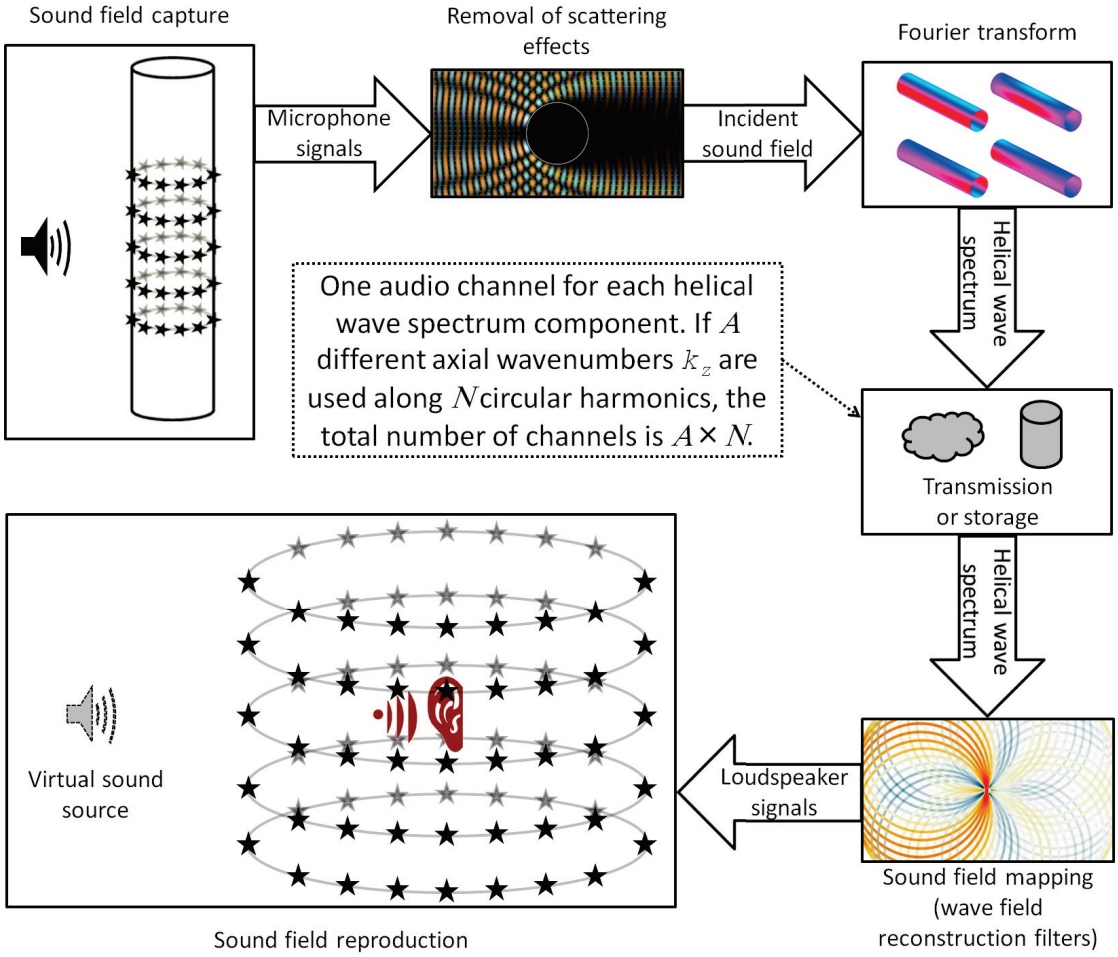


Figure 4.3: A sound field recording, transmission and reproduction system based on the helical wave spectrum. The sound field is sampled using a cylindrical microphone array. The spatial Fourier transform of the recordings is calculated and used as an encoding of the sound field. Reproduction of this encoding requires a set of filters that match the specific recording and reproduction arrays.

4.2.3 Limitations of current techniques

Current sound field recording and reproduction systems suffer from several drawbacks which hinder their practical use. Systems like those outlined in previous sections can achieve very high accuracies. However, they require a vast number of microphones, loudspeakers and a corresponding high-data-rate communication channel between them. One of the reasons for the high system complexity lies in the way in which the axial coordinate appears in Eq. (4.3). Computing independent Fourier transforms along the angular and axial coordinates leads to the requirement of densely packed transducer rings.

A broadcasting system may consider the transmission of the helical wave spectrum itself. If recorded with sufficient precision, the spectrum fully characterizes the sound field. The receiver should be able to use such an audio stream with virtually any reproduction system, as long as they have some way of calculating the sound field at the positions of their loudspeakers. Such a system is illustrated in Fig. 4.3.

The helical wave spectrum contains a very large amount of information. The use of cylindrical coordinates allows this kind of system to reduce the elevation accuracy; however, high horizontal accuracy demands a microphone/loudspeaker array with a large number of transducer rings, as well as a dense sampling of k_z.

A way to sidestep the costly transmission of the full helical wave spectrum is to generate loudspeaker signals at the recording station itself. This allows the broadcaster to transmit only the information needed by the listener instead of a full sound field characterization. Such a system is depicted in Fig. 4.4. The main disadvantage is the loss of flexibility in the reproduction stage. All users receiving the broadcast must use loudspeaker arrays of exactly the same size with equal loudspeaker distributions. Formats based on this paradigm are not future-proof since

they only preserve the sound field information required by a specific reproduction system.

A close inspection of the systems outlined above shows that the helical wave spectrum treats the sound field as a layered object and uses the same amount of information for each elevation angle. Horizontal accuracy can be modified by adding or removing transducer rings and consistently adjusting the sampling of k_z, i.e. the axial order. Elevation can be characterized with lower or higher accuracy by changing the number of transducers per ring and similarly modifying the angular order. It is, however, impossible to use different axial orders for different elevation angles.

The motivation behind the use of cylindrical coordinates is to reduce the amount of information spent in characterizing the sound field outside the horizontal plane. Ideally, the horizontal plane should be described with the highest accuracy, while other elevations (particularly near the poles) are only roughly approximated. The helical wave spectrum does not lend itself to these types of encoding.

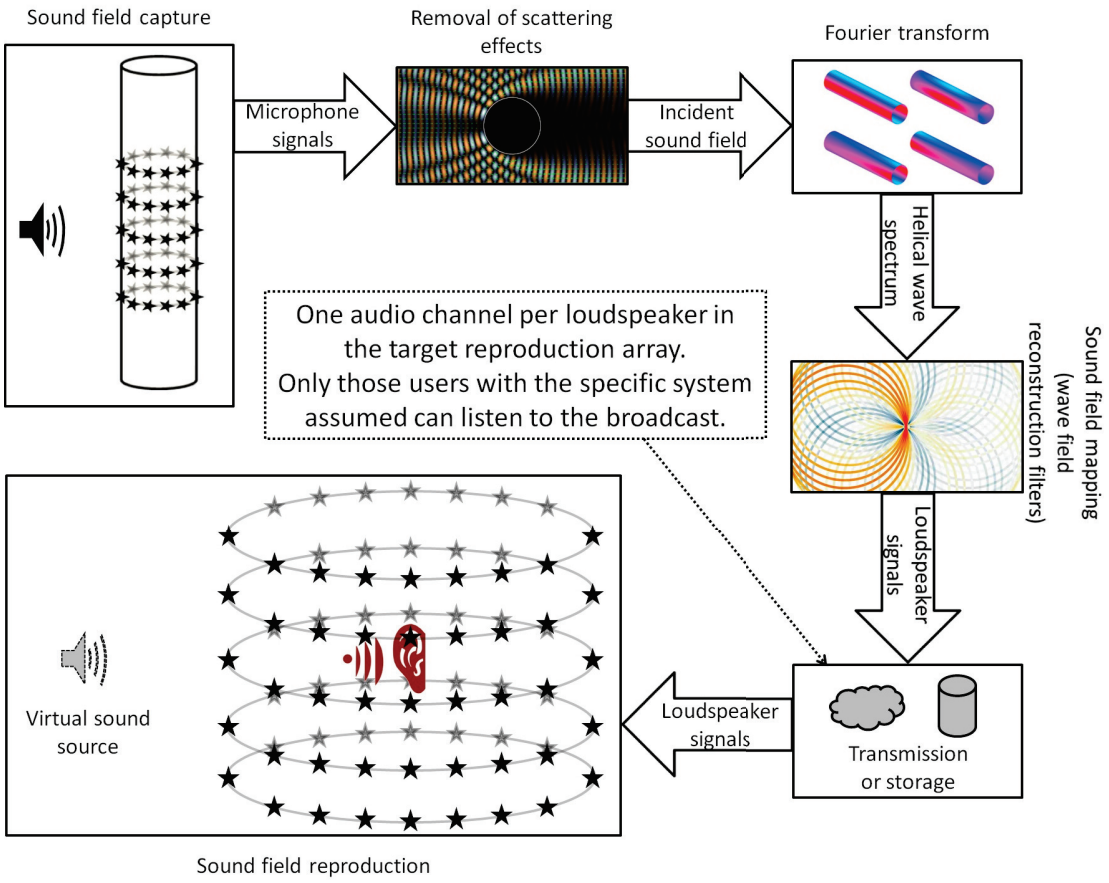


Figure 4.4: A sound field recording, transmission and reproduction system broadcasting only the information needed by a specific loudspeaker array. Recording is done using a cylindrical microphone array and its signals are matched to a target reproduction system using a set of wave field reconstruction filters. The result are the loudspeaker signals for the target array, so no further decoding is needed on the reproduction side.

Summarizing, existing technologies for sound field recording and reproduction using cylindrical arrays present the following disadvantages:

- High system complexity: The number of transducer rings required is strongly related to the temporal frequency of sound sources. High frequencies require impracticably dense transducer arrays.
- High bandwidth descriptions of sound fields: Storing or transmitting the full helical wave spectrum is, generally, wasteful. Most reproduction systems cannot fully use a complete description of the sound field.
- No standard encoding: The helical wave spectrum as recorded by the microphones or the loudspeaker signals for a target reproduction array are not system-independent encodings; thus far, no reference systems with a cylindrical geometry have been proposed.
- System-dependent format that is not future proof (in the case of systems like that shown in Fig. 4.4): Pre-computing loudspeaker signals to discard unneeded information leads to a recording that can only be used efficiently by the target loudspeaker array. Users with different reproduction systems cannot share the same broadcast, and future devices cannot take advantage of contents stored in this way.
- Equal resolution for all elevation angles: The helical wave spectrum can be computed with different axial and angular accuracies. However, the sound field is characterized with the same accuracy for all elevation angles. A cylindrical geometry was chosen to emphasize the horizontal plane, where human sound localization is most accurate, while reducing the amount of data used to encode

the sound field at different elevations. The helical wave spectrum is not an efficient choice to achieve this goal.

4.3 Mixed-Order Ambisonic encoding for cylindrical arrays

This section describes the proposal, an encoder-decoder system for sound field recording and reproduction systems of cylindrical geometry. The objective of the proposal is to overcome the drawbacks of existing systems, as outlined in the previous section. Specifically, the following properties are sought:

- System-independent encoding stage that can work with the signals recorded by any cylindrical microphone array.
- Scalable operation so that dense transducer arrays can be used optimally, while coarse ones still deliver acceptable results.
- Low-channel-count encoding allowing for the efficient use of available bandwidth and storage capacity.
- Future-proof, standard encoding compatible with Ambisonics' B-format [16]. It can be readily used with existing loudspeaker arrays and benefit from future innovations in their construction.
- An efficient characterization of the sound field that is optimized for human listeners. More resources are used in the horizontal plane, while other elevations are handled with reduced accuracy.

4.3.1 System overview

As previously mentioned, the proposed system consists of two components: an encoder and a decoder. A general view of the proposed system is shown in Fig. 4.5. Since one of the objectives is to achieve compatibility with present and future systems, the main innovation resides in the encoder. Its output requires only minor adjustments to be used with existing Ambisonic decoders [17, 28].

The encoder takes as its input the signals recorded by a cylindrical microphone array like the one depicted in Fig 4.1. A rigid scattering center, like that used in other systems [34, 23, 32], is assumed. The output is a partial description of the sound field as observed by the microphone array. The results of the encoding process are organized as to provide a coarse approximation in the lowest channels and finer details in higher ones. This makes the output of the proposed encoder compatible with B-format Ambisonic systems [16].

The decoder receives the encoder's output and generates appropriate loudspeaker signals to re-create the recorded sound field. The nature of the encoded data simplifies the decoder's task to simply choosing the appropriate channels and mapping them into a B-format stream. Any suitable Ambisonic decoder can then be applied to generate the loudspeaker signals.

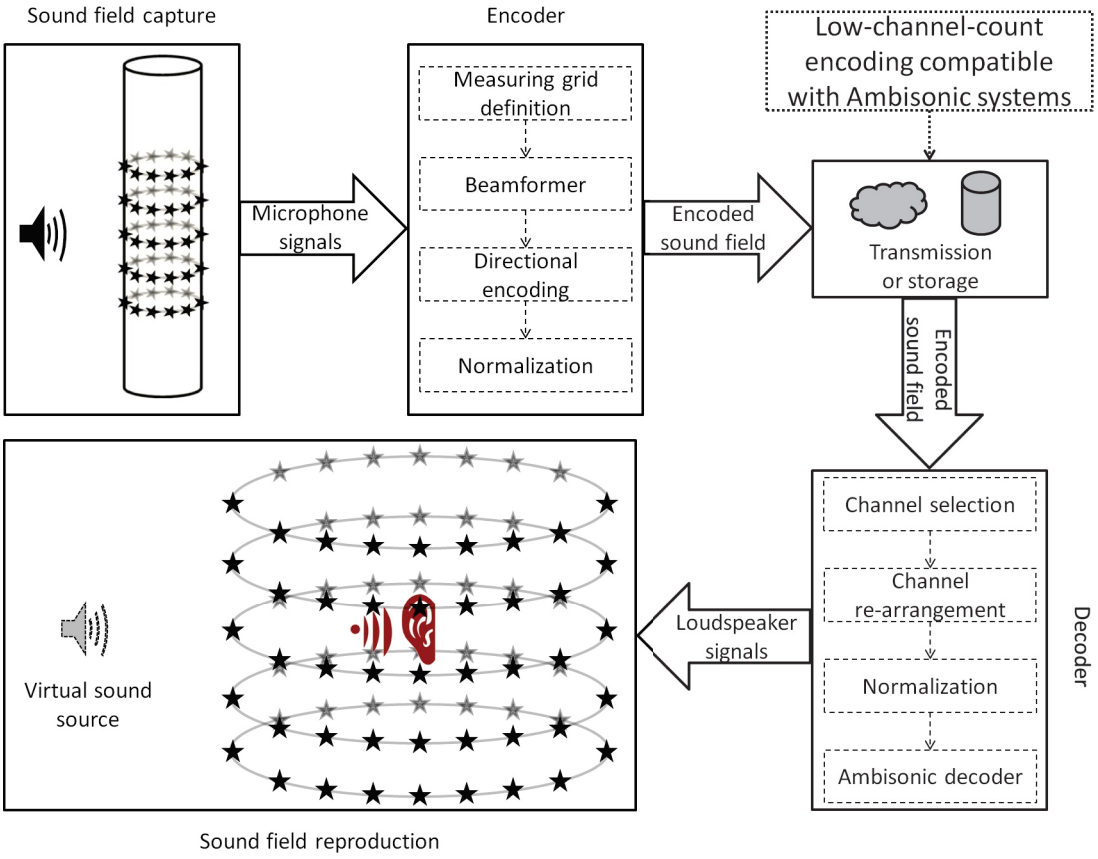


Figure 4.5: Overview of the proposed sound field recording and reproduction system. A cylindrical microphone array is used to sample the sound field. An encoding stage is used to generate a Mixed-Order Ambisonic encoding from the cylindrical microphone array recordings. The recorded sound field can be reproduced using any system capable of decoding Mixed-Order Ambisonic encodings.

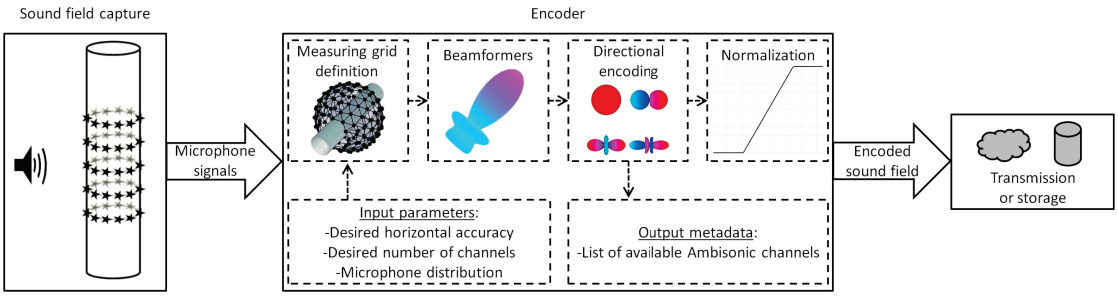


Figure 4.6: Block diagram for the proposed system's encoding stage. A cylindrical microphone array and a beamforming method is used to sample the sound field over a spherical measuring grid. The resulting sound pressure distribution is encoded using the spherical harmonic functions. Since the microphone array has different resolutions for azimuth and elevation, more horizontal spherical harmonics are used in the encoding. The result is a Mixed-Order Ambisonics encoding of the sound field.

4.3.2 Encoder

The main innovation introduced in the proposed system resides in the encoder component. Other systems rely on the helical wave spectrum to characterize the sound pressure distribution over the microphones grid. The proposal introduces an additional measuring grid which does not necessarily match the microphones one. A set of beamformers [19] are used to isolate the sound arriving from each direction in the measuring grid. These measurements are encoded according to their angles of arrival using a subset of the spherical harmonic functions. Finally, the encoding is normalized using the Furse-Malham weighting coefficients [16] to ensure optimal use of the system's dynamic range before the signals are broadcasted or recorded. A block diagram of the encoder is shown in Fig. 4.6.

4.3.2.1 The measuring grid and beamforming

The measuring grid is defined according to the application requirements and capabilities of the microphone array. An example of such a grid is shown in Fig. 4.7.

The user can indicate the desired horizontal accuracy using a positive

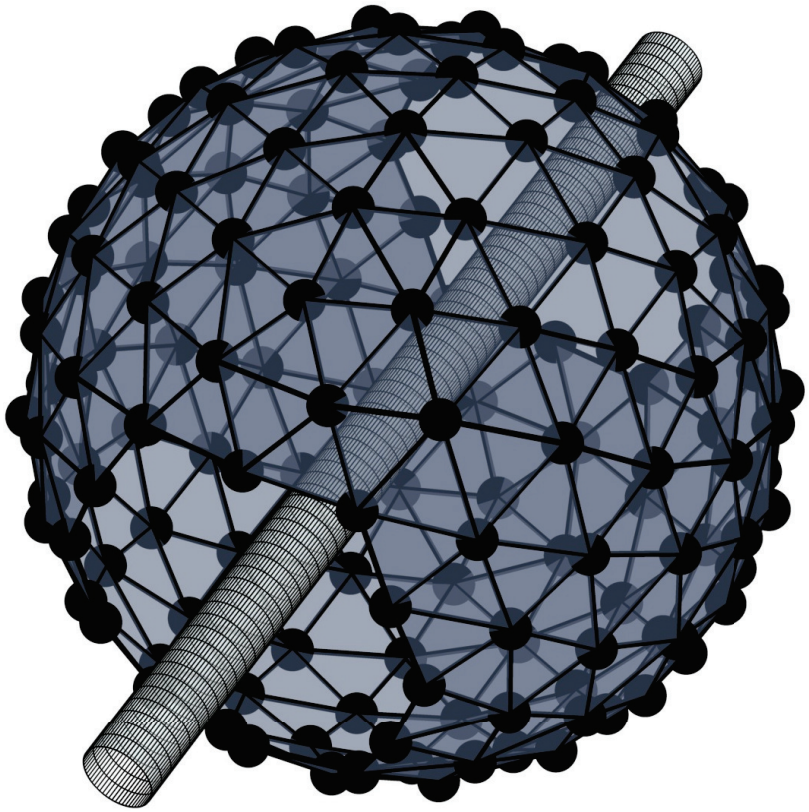


Figure 4.7: An example of a measuring grid surrounding a cylindrical microphone array. Each of the black circles represent one direction of incidence to be used in the encoding. The measuring grid should define a uniform sampling of all directions. It does not need to match the actual microphone distribution on the cylinder in any way.

integer number, the *horizontal Ambisonic order* N. This value is also used in the directional encoding stage and will result in $2N + 1$ channels of the encoding being used exclusively to characterize the horizontal component of the sound field. The microphone array must be composed of at least N rings, otherwise the user's input is ignored and the maximum value $N = \#\text{rings}$ is used instead.

It is also possible to set a desired total number of channels in the complete encoding. The value is not strictly obeyed and serves only as a guide to determine the amount of information used to encode elevation. The *mixed Ambisonic order* M is constrained by the number of channels in the encoding according to the following

expression:

$$\#\text{channels} = M^2 + 2N + 1. \tag{4.6}$$

The special case of $M = N$ corresponds to the full-sphere Ambisonic expansion. In this case both azimuth and elevation components are given equal importance. Most practical applications of this proposal will use small values for M.

The measuring grid is, for simplicity, defined as an almost uniform sampling of all directions. The density of the sampling is given by the constrain:

$$\#\text{directions} > (N+1)^2. \tag{4.7}$$

In particular, the proposed system has been evaluated using regular subdivisions of the icosahedron. Performance may improve slightly by using other measurement grids, such as Fliege distributions [40] or Lebedev quadratures [41].

The proposed system applies beamforming techniques [19] to isolate the sound arriving from each direction in the measuring grid. In particular, two kinds of beamformer are used. The first one relies on the pseudo-inverse of Eq. (4.2). The beamforming matrices are, therefore:

$$W_{\text{LS}}(k, \theta_{mic}, z_{mic}, \theta_{dir}, \phi_{dir}) = \left[\frac{i}{\pi^2 k \sin\phi_{dir} R_{cyl}} \sum_{n=-N}^{N} \frac{i^n}{H_n^{(1)'}(k \sin\phi_{dir} R_{cyl})} e^{in(\theta_{mic}-\theta_{dir})} e^{ik \sin\phi_{dir} z_{mic}} \right]^+ . \tag{4.8}$$

The above window typically provides good results; however, computation may be demanding or numerically unstable if high orders are considered. In such cases, the proposed system falls back to the following beamformer:

$$W_{\mathrm{dec}}(k, \theta_{mic}, z_{mic}, \theta_{dir}, \phi_{dir}) = -\, i\pi^2 k \sin\phi_{dir} R_{cyl}$$

$$\sum_{n=-N}^{N} i^{-n} H_n^{(1)'}(k \sin\phi_{dir} R_{cyl}) e^{-in(\theta_{mic}-\theta_{dir})} e^{-ik\sin\phi_{dir} z_{mic}}. \tag{4.9}$$

The filters of Eqs. (4.8) and (4.9) are directly applied to the signals recorded by the microphones. Their formulation already considers, and actually exploits, the scattering effects of the rigid cylinder. The result is a collection of signals, one for each direction represented in the measuring grid:

$$P_{dir}(k) = \sum_{mic} W(k, \theta_{mic}, \theta_{dir}) \psi_{total}(k, \theta_{mic}). \tag{4.10}$$

4.3.2.2 Directional encoding and normalization

Previous stages result in a set of more than $(N+1)^2$ audio signals, each corresponding to a specific direction in a quasi-regular sampling of the sphere. They can be encoded using the full set of spherical harmonic functions up to the N-th degree. However, the present proposal seeks to use higher precision in the horizontal plane and reduce the amount of information in the complete encoding.

The directional encoding stage of the proposed system considers the user-provided parameter for mixed-order Ambisonics, M. A directional encoding matrix C is defined by its elements as follows:

$$\begin{aligned} c_{(n+1)^2-n+m,dir} &= \mathrm{Y}_{nm}(\theta_{dir}, \varphi_{dir}) && \text{for } n < M, \\ \left.\begin{aligned} c_{M^2+2n,dir} &= \mathrm{Y}_{n,-n}(\theta_{dir}, \varphi_{dir}) \\ c_{M^2+2n+1,dir} &= \mathrm{Y}_{nn}(\theta_{dir}, \varphi_{dir}) \end{aligned}\right\} && \text{for } n > M. \end{aligned} \tag{4.11}$$

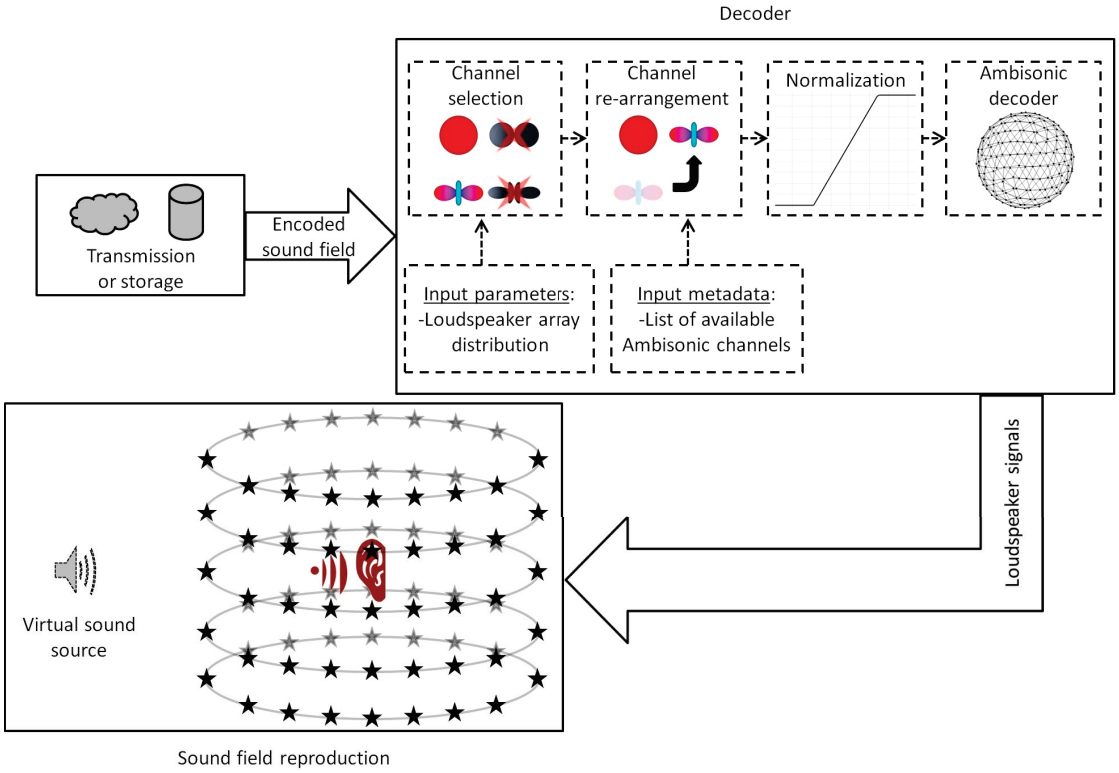


Figure 4.8: Block diagram of a Mixed-Order Ambisonics decoder which can be used with the output of the proposed encoder for cylindrical microphone arrays. The coefficients that are missing from a full spherical harmonics expansion are assumed to be zero and a full 3D High-Order Ambisonics encoding is prepared. This is later processed with any HOA decoder to produce loudspeaker signals for reproduction.

The functions Y are the spherical harmonic functions, in this case evaluated at the directions that appear in the measuring grid. The encoding matrix C has a total of $M^2 + 2N + 1$ rows and as many columns as there are directions in the measuring grid. A first encoding of the sound field, according to the user-provided parameters, can be easily calculated from C:

$$\tilde{B}(k) = C\left[P_{dir}(k)\right]. \tag{4.12}$$

The final stage performed by the encoder consists on normalizing the signals $\tilde{B}$. This step is not always necessary, but can prevent artifacts such as clipping or perceptible quantization noise under some circumstances. Normalization is done by simply multiplying the corresponding Furse-Malham coefficient [16] by each of the signals in $\tilde{B}$.

4.3.3 Decoder

The normalized encoding of the sound field, B, produced by the encoder can be easily reproduced using existing Ambisonic systems. It is important, however, to consider the fact that the encoder does not produce full-sphere descriptions of sound fields. An overview of the steps a decoder must follow to use the results of the proposed encoding system is shown in Fig. 4.8.

The encoded sound field may carry more information than a particular system is capable of reproducing. This is particularly true given the emphasis that the proposed encoder may put on the horizontal component of the sound field. A first step should determine if some channels must be discarded from the encoding. This can be achieved by evaluating the condition number for the system's decoder at the highest orders available. If the system may become numerically unstable,

high-order channels are discarded.

Secondly, the available channels must be re-arranged in the order expected by the system's decoder. If a full-sphere encoding is expected, zero-filling the unavailable channels can be done without affecting the quality of the reproduction.

The final step is to revert the Furse-Malham normalization by dividing each remaining channel by the same weights used in the encoding stage. The resulting set of signals can be run through a conventional Ambisonic decoder to reproduce the sound field observed by the cylindrical microphone array.

4.4 Summary

A new sound field encoder system for cylindrical microphone arrays was introduced. Unlike existing systems, the new proposal can flexibly devote most of the available resources to characterizing the sound field in the horizontal plane. Furthermore, the output of the encoder can be reproduced using any existing Ambisonic system with only a few modifications in the decoder system.

The new proposal satisfies the objectives set during its development. In particular, the encoder and its output are system-independent. This was achieved by defining a measuring grid which does not necessarily uses the same distribution as the microphone array. The measuring grid is defined using two parameters provided by the user, leading to a scalable encoding using considerably less channels than previous methods, such as those based on the helical wave spectrum. More specifically, the proposed method results in $M^2 + 2N + 1$ channels, where the linear dependency N sets the system accuracy in the horizontal plane. The presented system is considered to be future-proof since it is compatible with the Ambisonic B-format with little modifications.

CHAPTER V

Decoding generalized Ambisonics for arbitrary loudspeaker configurations

5.1 Overview

In this chapter, the focus changes from the recording of sound fields to their reproduction using loudspeaker arrays. Sound field reproduction systems are limited by the vast amounts of information present in sound fields. Re-creating an arbitrary sound field throughout the full listening range (20 Hz to 20 kHz) and inside a volume large enough for even a single listener would require thousands of independent audio channels. Practical systems can only approximate the sound field to the extent allowed by their bandwidth constraints and number of available loudspeakers. While it is easy to reconstruct sound fields very accurately over small regions, this chapter emphasize the need to re-create the sound field within a volume large enough for the user or users of a reproduction system to fit comfortably.

The proposals outlined in this chapter will use High-Order Ambisonics as a basis since it is a well-known technique for the approximate characterization and reproduction of sound fields using surrounding loudspeaker arrays [16, 17]. However,

the techniques described are applicable to other cases. For example, in Chapter VI these techniquies are applied to a different kind of sound field encoding.

The main objective of the reseach presented in this chapter is, therefore, to develop a method for the decoding of HOA encodings. The proposal will focus on improving the user's experience by extending the listening region. The performance of a standard HOA decoder is compared with that of the proposed method by evaluating not only the reconstruction error, but also a perceptually meaningful parameter: the interaural differences.

5.2 Conventional Ambisonic decoders

High-Order Ambisonics defines a scalable format to store and transmit sound field information. A significant advantage of HOA is that it completely separates the recording/synthesis stage from the reproduction one. An encoding process generates HOA descriptions from either microphone signals or sound field simulation results. Later, a decoder uses these descriptions to calculate proper loudspeaker signals, considering the peculiarities of the array.

A variety of methods to decode HOA exist. A full review of all these methods would go beyond the scope of this dissertation; however, two simple ones are outlined here as a simple introduction to the problem of reproducing HOA encodings.

5.2.1 Projection of the spherical harmonic functions

The basic decoding of HOA data for reproduction using a surrounding loudspeaker array is achieved by computing a weighed sum of all HOA channels for each loudspeaker. The problem of decoding HOA reduces to finding the required weights, that is, a *decoding matrix* which mixes the HOA channels to produce

loudspeaker signals.

The simplest way to find the required weights is by applying the spherical harmonic expansion, Eq. (2.23), directly. The loudspeakers are used to sample the spherical harmonic functions. This process results in the following loudspeaker signals $p_{\text{spk}}^{[k,r]}$:

$$p_{\text{spk}}^{[k,r]} = \frac{1}{\#\text{Loudspeakers}} \sum_{n=0}^{N} \sum_{m=-n}^{n} \mathrm{B}^{[}k, r]_{nm} \mathrm{Y}_{nm}(\theta_{\text{spk}}, \varphi_{\text{spk}}). \tag{5.1}$$

The reciprocal of the number of loudspeakers works as a normalization factor, ensuring that the encoding and decoding processes do not change the sound level. However, the simple gain used by the projection decoder is only valid if the loudspeakers define a truly regular sampling of all directions. This constraint is impossible to satisfy except for sets of 4, 6, 8, 12 or 20 loudspeakers. The reason behind this is that the only regular polyhedra in 3D space are the tetrahedron, hexahedron, octahedron, dodecahedron and icosahedron. Any loudspeaker distribution, besides those defined by the faces of the platonic solids, will deviate from a regular sampling of the sphere and will, therefore, introduce some error in the reproduction when using a projection decoder.

Another limitation of the projection decoder is what gives it its name. The equality of Eq. (5.1) requires the recording and reproduction arrays to have the exact same radius. This is made explicit by the superindex $[k, r]$. When the radii of the arrays differ, Eq. (5.1) will project the sounds recorded at the microphone array's radius onto that of the loudspeaker array. This process will distort sound sources making them appear smaller or larger, blurring spatial details.

Despite its limitations, the simplicity of the projection decoder makes it an attractive choice when little processing power is available and reproduction accuracy is not a major concern.

5.2.2 Least-squares approximate solution

A more robust approach to HOA decoding makes use of the least-squares approximation. The contribution of each loudspeaker towards re-creating a given spherical harmonic, as seen from the listening position, is calculated and the results are stacked into what is called a *re-encoding matrix*.

$$\mathbf{B}(k) = \mathbf{C}\mathbf{p}(k). \tag{5.2}$$

Vector $\mathbf{B}$ is the HOA encoding of a given sound field and is assumed to be truncated at Ambisonic order N. The components of vector $\mathbf{p}(k)$ are the loudspeaker signals. The re-encoding matrix $\mathbf{C}$, therefore, has $(N+1)^2$ rows and one column for every loudspeaker in the array. The elements of $\mathbf{C}$ are given by the spherical harmonic functions evaluated in the directions of the loudspeakers, $(\theta_{\text{spk}}, \varphi_{\text{spk}})$, as follows:

$$c_{n^2+n+m,\text{spk}} = \mathrm{Y}_{nm}(\theta_{\text{spk}}, \varphi_{\text{spk}}). \tag{5.3}$$

The loudspeaker signals needed to reconstruct a particular HOA-encoded sound field can be computed by inverting the linear system of Eq. (5.2). For the re-encoding matrix to be invertible, however, the number of loudspeakers in the array must match the count of ambisonic channels. In practice, it is desirable to use larger arrays to improve the reproduction accuracy; this leads to an underdetermined linear system. It is common to rely on the Moore-Penrose pseudo-inverse to invert the re-encoding matrix. The decoding equation can be written in terms of the pseudo-inverse of $\mathbf{C}$, denoted by $\mathbf{C}^+$, as [38]

$$\mathbf{p}(k) = \mathbf{C}^+\mathbf{B}(k). \tag{5.4}$$

If the number of loudspeakers in the array is larger than the number of ambisonic channels, Eq. (5.4) gives the solution that minimizes the Euclidean norm

of $\mathbf{p}(k)$. If the array has fewer loudspeakers than the number of ambisonic channels, exact reconstruction becomes impossible in general; however, Eq. (5.4) will result in the loudspeaker signals minimizing the Euclidean norm of the error vector [37]

$$\epsilon(k) = \mathbf{C}\mathbf{p}(k) - \mathbf{B}(k). \tag{5.5}$$

The Moore-Penrose pseudo-inverse is not a continuous operation and, under some circumstances, can lead to a large reproduction error. Whether or not this is the case for a given loudspeaker configuration can be determined by calculating the condition number of its re-encoding matrix.

$$\mathrm{cond}(\mathbf{C}) = \|\mathbf{C}\|\|\mathbf{C}^{+}\|, \tag{5.6}$$

where $\|\cdot\|$ denotes a matrix norm. A large condition number implies that $\mathbf{C}$ is ill-conditioned and least-squares solutions are numerically unstable.

Decoding matrices derived using the pseudo-inverse can result in a reduced listening region when a large number of loudspeakers is used. The least-squares solutions provided by the pseudo-inverse for underdetermined systems are those with minimal Euclidean norm. This means that the reproduced sound field will closely match the recorded one at a single spatial point but it will quickly vanish away from this privileged position. Minimizing the Euclidean norm of the loudspeaker signals does not ensure the best results from the perspective of a human listener who would benefit instead from a large listening region.

5.3 Optimized decoding for irregular arrays

Decoding of ambisonic data through the pseudo-inverse of a re-encoding matrix can lead to the drastic amplification of errors when targeting an irregular

loudspeaker array and a small listening region when a large number of loudspeakers is used.

In this section, a new HOA decoding method that seeks to overcome the shortcomings of conventional approaches is introduced. In the proposed method, a decoding matrix is calculated iteratively so as to stabilize the reproduction around the privileged point at the center of the listening region. To achieve this, a constraint on the spatial distribution of the reconstruction error is imposed. While approaches based on the pseudo-inverse minimize the norm of the loudspeaker signals, the proposed method attempts to minimizes the radial derivative of the reconstruction error. This leads to a constrained least-squares problem that can be solved numerically through iterative methods.

5.3.1 Radial stabilization of the reconstruction error

The proposed decoding method iteratively improves a decoding matrix by stabilizing the reconstruction around the listening position. To accomplish this, the decoding gains are perturbated in such a way that the radial derivative of the reconstruction error is minimized.

The decoding matrix obtained by applying the pseudo-inverse is a good starting point since it ensures that the reconstruction error at the center of the listening region is minimal. However, if the re-encoding matrix is ill-conditioned for a given array, other decoding matrices can be used, such as that from the projection decoder which is always stable.

Assuming ideal monopole radiators, the reconstruction error at the position $\vec{r}$ can be written as

$$\epsilon(k,\vec{r}) = \tilde{\psi}(k,\vec{r}) - \phi(k,\vec{r}) - \sum_{s}\sum_{m=0}^{N}\sum_{n=-m}^{m} G^{s}_{mn}(k)\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\mathrm{Y}_{mn}(\theta_s,\varphi_s), \tag{5.7}$$

where $\tilde{\psi}(k,\vec{r})$ represents the sound field encoded using HOA, $\phi(k,\vec{r})$ stands for sound field reproduced by the loudspeaker array when the initial decoding matrix is applied. The first sum runs over the loudspeakers in the array; the position of the $s-$th loudspeaker is given, in spherical coordinates, as $\vec{r_s} = (r_s, \theta_s, \varphi_s)$. The gains $G^s_{mn}(k)$ are initially set to zero, yielding only the initial approximation.

The proposal is to perturbate the solution of Eq. (5.7) through the gains $G^s_{mn}(k)$ in such a way that the listening region is enlarged. The behavior of the reconstruction error as the listening point moves away from the center of the array can be described by the radial derivative of Eq. (5.7)

$$\begin{aligned}\frac{\partial}{\partial r}\epsilon(k,\vec{r}) &= \nabla\epsilon(k,\vec{r})\cdot\hat{r} \\ &= \nabla[\tilde{\psi}(k,\vec{r}) - \phi(k,\vec{r})]\cdot\hat{r} - \sum_s \frac{\partial}{\partial r}\left[\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\right]\sum_{m=0}^{N}\sum_{n=-m}^{m} G^s_{mn}(k)\mathrm{Y}_{mn}(\theta_s,\varphi_s).\end{aligned} \tag{5.8}$$

The first term is the radial derivative of the reconstruction error when using the initial decoding matrix. This term can be regarded as a constant $\mathbf{d} \equiv \nabla[\tilde{\psi}(k,\vec{r}) - \phi(k,\vec{r})]\cdot\hat{r}$ since it is independent of the choice of gains $G^s_{mn}(k)$. The radial derivative of the monopole field can be expressed as:

$$\frac{\partial}{\partial r}\left[\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\right] = \mathrm{D}_s(k)\left[\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\right], \tag{5.9}$$

with the operator

$$\mathrm{D}_s(k) \equiv -\frac{|\vec{r}| - |\vec{r_s}|\cos(\vec{r},\vec{r_s})}{|\vec{r}-\vec{r_s}|}\left(\frac{1}{|\vec{r}-\vec{r_s}|} + ik\right). \tag{5.10}$$

Using these definitions, Eq. (5.8) can be rewritten as

$$\frac{\partial}{\partial r}\epsilon(k,\vec{r}) = \mathbf{d} - \sum_s\sum_{m=0}^{N}\sum_{n=-m}^{m}\left[\mathrm{D}_s(k)\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\mathrm{Y}_{mn}(\theta_s,\varphi_s)\right] G^s_{mn}(k). \tag{5.11}$$

By taking the norm of Eq. (5.11), it is possible to impose the following constraint on the radial derivative of the reconstruction error:

$$\left|\frac{\partial}{\partial r}\epsilon(k,\vec{r})\right| = \|\mathbf{L}\cdot\mathbf{G} - \mathbf{d}\| \leq \rho. \tag{5.12}$$

Here, ρ is some threshold limiting the permissible variation of the reconstruction error. The entries of $\mathbf{G}$ are the loudspeaker gains $G^s_{mn}(k)$, while the entries of the operator $\mathbf{L}$ are defined as

$$L^s_{mn}(k) = \mathrm{D}_s(k)\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\mathrm{Y}_{mn}(\theta_s,\varphi_s). \tag{5.13}$$

Equation (5.12) can be used to calculate a set of decoding gains by defining a target radius for the listening region and a maximum allowed variation for the reconstruction error. Alternatively, it is possible to find the gains that minimize the variation in the reconstruction error by successive approximations. The resulting gains

$$\begin{aligned}\mathbf{G} = \arg\min_{\mathbf{G}} \Bigg| \frac{\partial}{\partial r} \Bigg[& \tilde{\psi}_k(r,\theta,\varphi) - \phi_k(r,\theta,\varphi) \\ & - \sum_s \sum_{m=0}^{N} \sum_{n=-m}^{m} G^s_{mn}(k) \frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|} \mathrm{Y}_{mn}(\theta_s,\varphi_s) \Bigg] \Bigg|\end{aligned} \tag{5.14}$$

can be used to generate loudspeaker signals through the following decoding equation:

$$\mathbf{p}(k) = \left[\mathbf{G}^{\mathrm{T}}(k) + \mathbf{C}^{+}_{\mathrm{initial}}\right] \mathbf{B}(k). \tag{5.15}$$

In contrast with the unconstrained least-squares method, the proposed decoder does not guarantee that the reconstruction error at the center of the array will be minimal. Instead, it seeks to maintain an acceptable reconstruction accuracy over a wider region and reduce reproduction artifacts as the listener turns his head or moves slightly within the array.

5.4 Near-field corrections for non-spherical arrays

High-Order Ambisonics describe the sound field at a fixed distance from the observation point; that is, over a spherical boundary. The HOA decoders discussed thus far follow on this assumption and do not introduce any distance compensation.

If the loudspeaker array is, like the microphone one, spherical, then the two radii can be matched using a quotient of spherical Hankel functions. It is, however, very difficult to arrange vast numbers of loudspeakers at a constant distance from the position of the listener.

5.4.1 Near-field compensated high-order Ambisonics

The solutions of the Helmholtz equation derived in Chapter II are valid inside a region with no sound sources. Sound radiates from the outside and into the region under consideration. This leads to the solutions containing only the spherical Bessel functions for the radial component. However, a more general case, where a sound source is located in the region of interest and radiates sound as outwards spherical waves can also be described by the Helmholtz equation. The solutions of the Helmholtz equation under this condition include the spherical Hankel functions $h_n(kr)$ [21].

$$\psi(\vec{r},k) = \sum_{n=0}^{\infty} k \frac{h_n(kr)}{i^{n+1}} \sum_{m=-n}^{n} O_{nm}(k) \mathrm{Y}_{nm}(\theta,\varphi). \tag{5.16}$$

The coefficients O_{nm} are known as the *exterior multipole expansion* of the sound field $\psi(\vec{r},k)$ and characterize the radiation patterns of the sources present within a spherical boundary.

The exterior expansion coefficients are not a conventional HOA encoding and they are unfit for presentation using a surrounding loudspeaker array. It is possible, however, to derive an standard HOA encoding of the sound fields they encode by applying a spatial translation. The extended sources encoded by the coefficients O_nm are, then, present as if they were outside of a spherical boundary surrounding the listener.

Converting the O_{nm} coefficients into the standard HOA coefficients B_{mn} can

be done with the following equation [19]:

(5.17)

$$B_{nm}(k) = \sum_{n'} \sum_{m'} \left[\frac{k}{4\pi i^{n+n'+1} j_n(kr)} \int_{boundary} d\Omega \mathrm{Y}_{n'm'}(\theta', \varphi') \mathrm{Y}_{nm}(\theta, \varphi) h_n(kr') \right] O_{n'm'}(k).$$

The primed coordinates now represent the position from which the extended sound source will be presented. Therefore, the choice of r' must be made considering the radius of the target listening region.

Due to its explicit dependency on a fixed distance, the encoding of Eq. (5.17) is said to include *near-field corrections*. This kind of HOA recordings, combined with the information on which distance was chosen for the translation, can be referred as Near-Field Compensated High-Order Ambisonics (NFC-HOA) encodings.

5.4.2 Compensating for disparate loudspeaker distances

The decoding method introduced in the previous section cannot handle NFC-HOA encodings. A further consideration is that irregular arrays are often chosen due to the difficulty of building a regular or almost-regular one. When angular uniformity is hard to achieve, it seems unreasonable to expect that the loudspeakers can be positioned at a constant separation from the listening position. The HOA decoder previously introduced is reformulated in this section so as to handle NFC-HOA encodings.

Conventional HOA encodings use the far-field approximation and present all sound sources from a spherical boundary of an assumed large radius. In actual scenarios, however, the sound sources do not necessarily lie on this boundary. Some research attempts to correct for this by introducing distance filters; however, they exhibit an infinite bass boost [24].

On the other hand, the NFC-HOA encodings of Eq. (5.17), include an infinite

bass attenuation due to the reciprocal of the Bessel functions. It is possible to combine both effects during the decoding stage to produce a stable set of distance compensation filters. This process, which requires filtering the NFC-HOA data before decoding it, does not depend on the angular coordinates. Therefore, it has no impact on the performance of the decoding method used.

Distance filtering of the NFC-HOA encoding of a sound source located at a distance r_{src} from the listener, to be decoded using a loudspeaker array of radius r_a can be done using the following transfer function [24]:

$$H_n^{NFC}(\omega) = \frac{\sum_{m=0}^{n} \frac{(n+m)!}{(n-m)!m!} \left(\frac{-ic}{\omega r_{\text{src}}}\right)}{\sum_{m=0}^{n} \frac{(n+m)!}{(n-m)!m!} \left(\frac{-ic}{\omega r_a}\right)}. \tag{5.18}$$

When the distance to the loudspeakers is not constant, a filterbank for varying r_a can be designed. This leads to a set of HOA encodings which can then be decoded using the procedure introduced in the previous section.

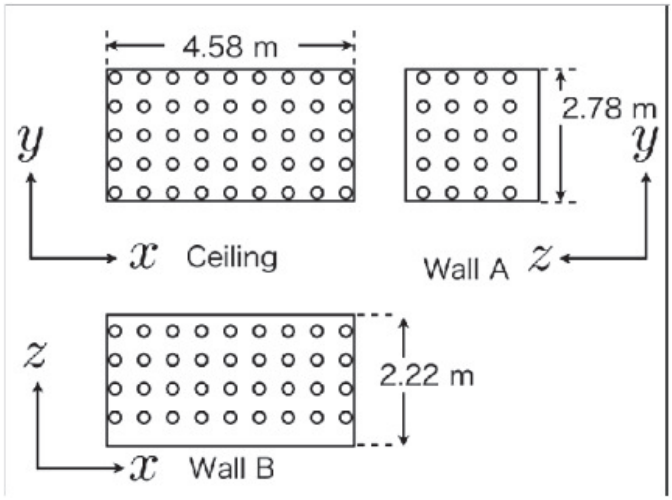
After simplification, it is possible to model the proposed decoder including corrections for near-field effects as the following constrained least-squares problem:

$$\left|\frac{\partial}{\partial r}\epsilon(k,\vec{r})\right| = ||\mathbf{L}\cdot\mathbf{G}-\mathbf{d}|| \leq \rho, \tag{5.19}$$

where the regularization operator $\mathbf{L}$ has the following components:

$$L_{nm}^{s} = \frac{(n+m)!}{m!(n-m)!}\cdot\frac{|\vec{r}|-|\vec{r_s}|\cos(\vec{r},\vec{r_s})}{(-ikr_{\text{src}})^n|\vec{r}-\vec{r_s}|}\left(\frac{1}{|\vec{r}-\vec{r_s}|}+ik\right)\frac{e^{-ik|\vec{r}-\vec{r_s}|}}{|\vec{r}-\vec{r_s}|}\mathrm{Y}_{nm}(\theta_s,\varphi_s). \tag{5.20}$$

The regularization operator depends explicitly on the distance to the sound sources. However, it is possible to establish a convention in the encoding stage that would include distance compensation filtering in order to translate all sources to a fixed distance. After settling on such convention, the decoding of NFC-HOA data can be made independent of the distance to the sources present in the re-created sound field.



(a) Layout of the 157-channel irregular loudspeaker array

(b) The 157-channel irregular loudspeaker array built ainside an anechoic chamber

Figure 5.1: Layout of a 157-channel irregular loudspeaker array used to evaluate the proposed HOA decoder. Panel (a) shows the layout for the walls and ceiling; the distance between adjacent loudspeakers is constant and equals 50 cm. Panel (b) shows a photograph of this particular loudspeaker array built inside a soundproof room covered with a sound absorbing material to reduce reflections.

5.5 Evaluation

In this section, the performance of the proposed decoding method and that of a conventional HOA decoder are compared using a computer simulation. The analysis carried out comprises both, the physical accuracy of the sound field reconstruction, and two perceptually meaningful parameters for sound localization: the interaural level and phase differences.

The loudspeaker array used for evaluation is an irregular one with loudspeakers distributed on the walls and ceiling of a rectangular room. The distance between loudspeakers is constant, resulting in an irregular angular sampling due to the differences in distance between the loudspeakers and the listener. A total of 157 loudspeakers are considered in a configuration that leads to well-conditioned re-encoding matrices up to the fifth Ambisonic order. The layout of this particular array is shown in Fig. 5.1.

High-Order Ambisonics descriptions of the sound field are synthesized by simulating plane waves of various frequencies incident from several directions. The plane waves are sampled by a spherical microphone array that uses a regular, Fliege geometry [40].

5.5.1 Sound field reconstruction error

The first evaluation consists of observing the sound field reproduction accuracy directly when a conventional least-squares decoder and the proposed constrained least-squares decoder are used. The reconstruction accuracy of both decoders was calculated over a region large enough to accommodate an average human listener. In concrete, the reconstruction error is given as the root-mean-square value of the difference between the reproduced and ideal sound fields inside a sphere

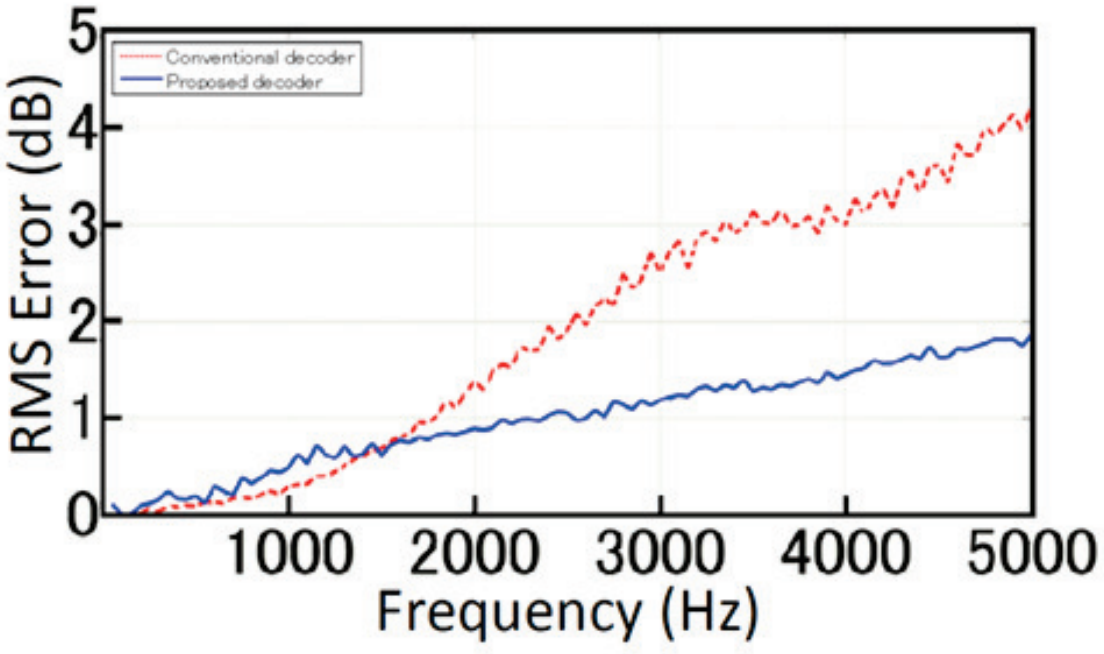


Figure 5.2: Reconstruction error for plane waves of various frequencies incident from the front. The red curve corresponds to a conventional decoder, while the blue curve shows the results achieved by the proposed decoding method.

at the center of the array. The radius of the sphere was chosen to be 8.5 cm to match that of the average human head. All results were obtained through a computer simulation that treats the loudspeakers as ideal omnidirectional radiators.

Several sound fields, consisting of single plane waves, were reconstructed. The test fields were divided into two sets. The first one consisted of simulated plane waves of frequencies between 50 Hz and 5 kHz arriving from a fixed direction $(\theta = 0, \varphi = 0)$. The second set of test fields consisted of a 2 kHz plane wave incident at azimuth angles between 0° and 90°.

The results from the first set of tests are summarized in Fig. 5.2. At low frequencies, the conventional least-squares decoder shows slightly better results, although both methods present acceptable results with small reproduction errors.

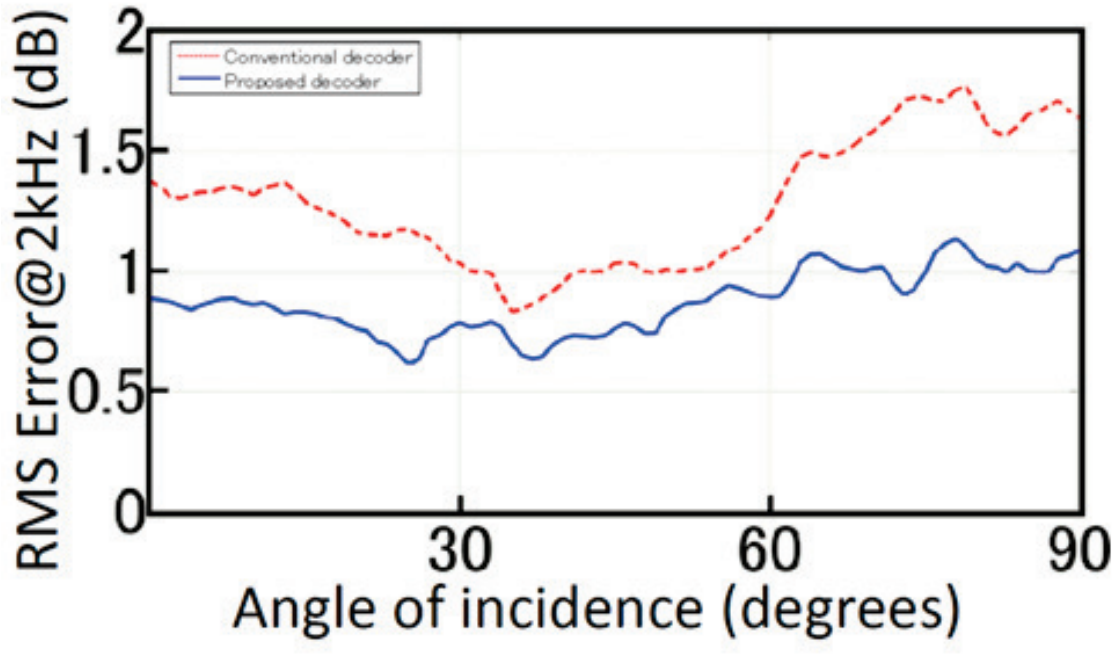


Figure 5.3: Reconstruction error for a 2 kHz plane wave as a function of the angle of incidence. The results for the conventional decoder are shown in red, and those corresponding to the proposed decoder in blue.

As frequency increases, though, the unconstrained least-squares approximation sees its reconstruction error increase rapidly. On the other hand, the proposed method manages to mantain the RMS error below 2 dB throughout the entire frequency range tested.

As for the second set of evaluation tests, the results are presented in Fig. 5.3. These tests consider a plane wave of a fixed frequency, 2 kHz. Consistently with the results of the first test, the proposed HOA decoder outperforms the conventional one, although the difference is less than 0.5 dB. However, variations in decoder performance for plane waves incident from different directions is almost twice as large when no constrains are imposed on the radial derivative of the reconstruction error.

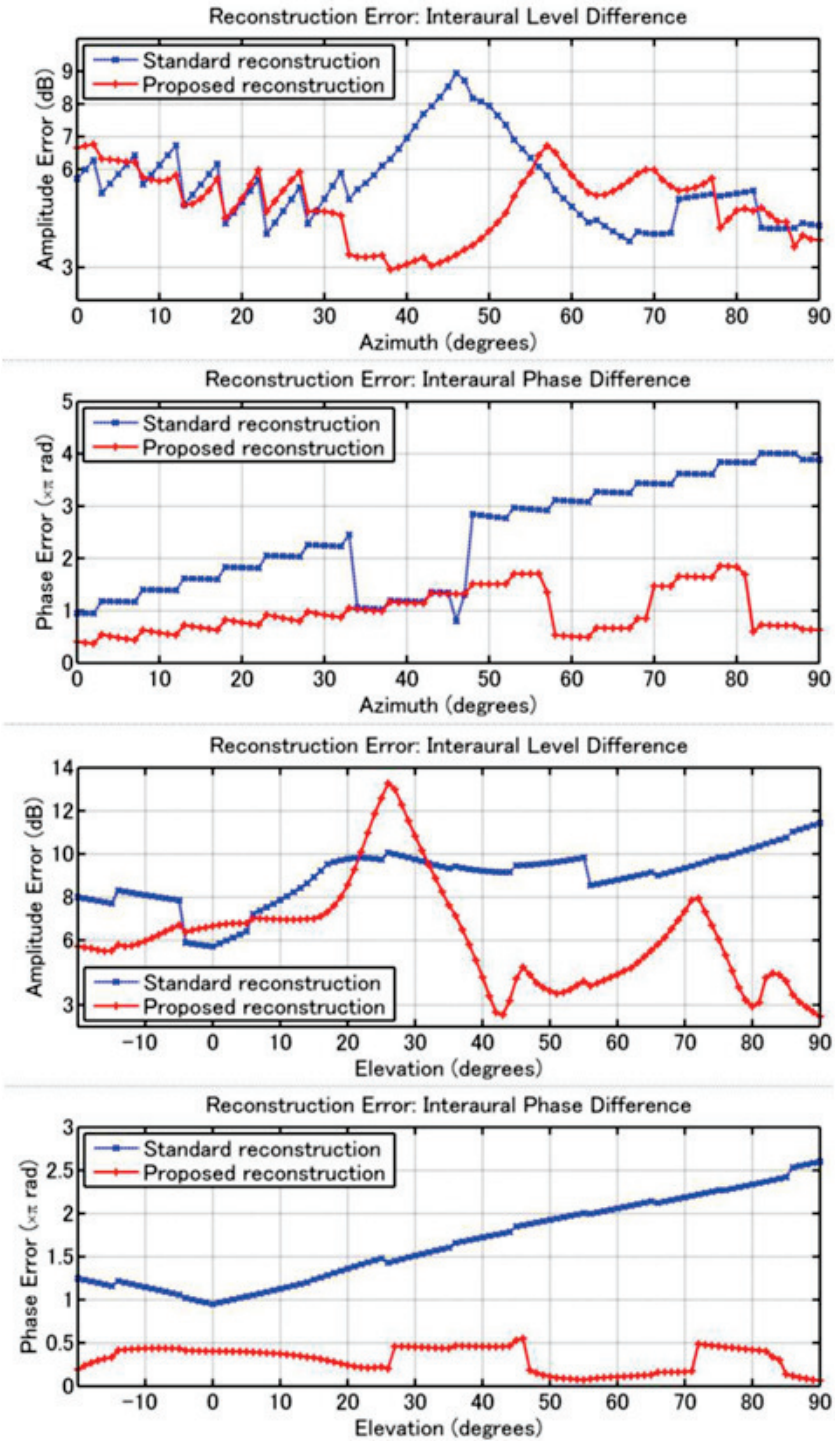


Figure 5.4: Average error in the interaural level and phase differences when presenting 5th order HOA recordings over an irregular, 157-channel loudspeaker array. The presented values represent the average for frequencies up to 5 kHz with the results for a conventional HOA decoder in blue and those for the proposed method in red. The first two panels show, from top to bottom, the error in the presented ILD and IPD for plane waves arriving at different azimuth angles and elevation 0. The last two panels show the same results but for different elevation angles at an azimuth of 0 degrees.

5.5.2 Interaural cues

A second evaluation further compares the performance of the proposed HOA decoder to that of a conventional one. Focus will now be on two perceptually meaningful variables: the interaural level difference (ILD) and the interaural phase difference (IPD). The two variables were chosen since they are important for sound localization in humans. Comparisons are drawn for a virtual loudspeaker array modeled from the Head-Related Transfer Functions of the SAMRAI dummy head. The virtual array approximates the 157 channel array used in the previous evaluation. Since the SAMRAI HRTF was not available at the precise positions defined by the array, the sample points that minimize the error in central angle were used. Distance compensation filters were also applied to approximate the parallelepiped shape of the array.

The results in conveying interaural differences for the two HOA decoders evaluated are presented in Fig. 5.4. The results show a pronounced difference in the ILD reconstruction error for azimuth angles between 30° and 55°. Incidentally, this region is also the one at which the loudspeaker array exhibits greater irregularities in its layout, as it transitions from a densely sampled region at the front to a sparsely sampled one on the left side of the listener. The performance of the proposed decoder is seen to be considerably better at elevation angles above 32°. This phenomenon is related to the lack of loudspeakers below the listener.

5.6 Summary

This chapter focused on sound field reproduction, specifically that of HOA-encoded sound fields. Conventional HOA encodings and their decoders do not consider distance explicitely since they assume that the far-field approximation

holds. Furthermore, transducers in both the recording and reproduction arrays are considered to be equidistant from the observation point.

A new HOA decoding method was introduced by taking into consideration the difficulty of building a spherical loudspeaker array, as well as the desire for a larger listening region to accommodate the listener(s). The performance of the proposal was compared with that of a conventional HOA decoder. Not only was the accuracy in the reconstruction of the field considered for evaluation, but also the interaural cues that each decoding method conveys to the listener.

CHAPTER VI

3D Cylindrical Ambisonics

6.1 Overview

A significant amount of research has focused on the design and use of spherical microphone arrays [25]. The symmetric layout, having no preferred orientation, simplifies the formulas used to characterize the sound field. In Chapter III a different kind of microphone array, a cylindrical one, was considered.

Spherical microphone arrays are inadequate when recording over large regions; for example, across a stage or a conference room. In this case, cylindrical arrays can be a good choice since the axis of the cylinder defines a privileged direction; it is thus possible, for example, to record the horizontal plane with high accuracy while using only a coarse sampling for elevation. Sound sources, such as speakers in a room or instruments on the stage, tend to be positioned at similar elevations.

Cylindrical microphone arrays are also easier to design. Any number of transducers can be spaced regularly along the cylinder's axis and circumference. Furthermore, cylindrical baffles are known to have good characteristics making the microphone array robust to errors in transducer placement as well as microphone self-noise [26, 27].

One problem of working with cylindrical arrays is the lack of an encoding

scheme, like that defined by the spherical harmonic expansion and used in HOA. This problem was considered in Chapter IV, where a method to derive HOA encodings from the recordings of cylindrical microphone arrays was presented.

In this chapter, the spherical geometry underlying HOA is discarded to make full use of the separate axial and polar coordinates introduced by the choice of a cylindrical arrangement. The basis for a new sound field recording and encoding method using cylindrical microphone arrays is outlined in the following sections. The proposed encoding, called 3D Cylindrical Ambisonics, is scalable and system-independent, like HOA in the spherical case. It treats the axial and polar coordinates independently; therefore, different resolutions can be used for azimuth and elevation. A general approach to reproduce the proposed encoding using loudspeaker arrays is briefly discussed. The computer simulation of a complete system, including a rigid cylinder baffle, recording, encoding and reproduction using loudspeakers is considered to evaluate the proposal.

6.2 Spatial encoding for cylindrical microphone arrays

There have been some efforts to reproduce sound fields recorded by cylindrical microphone arrays using loudspeakers. Chapter III introduced the plane-wave decomposition to encode and reproduce the far field. Another attempt which also works in the near field involves matching two cylinders of different radii with a propagator filter. The filters, known as the wave field reconstruction filters, are given in the helical wave spectrum domain by the following equation [33]:

$$\mathrm{G}_n(k_r) = -\frac{k_r R_{\mathrm{cyl}} \mathrm{H}'_n(k_r R_{\mathrm{cyl}})}{R_{\mathrm{spk}} \mathrm{H}_n(k_r R_{\mathrm{spk}})}. \tag{6.1}$$

The filters described in this equation depend only on the radius of the microphone array R_{cyl} and the loudspeaker array R_{spk}.

Previous attempts at using a cylindrical microphone array to characterize sound fields lack the desirable properties of methods based on spherical geometries such as HOA. The plane-wave decomposition defines a scalable and system-independent encoding scheme; however, it cannot deal with localized sound sources. In general, information regarding the distance to the sound sources is lost. On the other hand, the wave field reconstruction filters of Eq. (6.1) can accurately re-create the sound field measured by a cylindrical microphone array using a cylindrical loudspeaker array. However, the filters must be designed for every pair of recording and reproduction systems; this technique does not generate an intermediate encoding of the measured field.

In this section a new encoding format which results in a scalable representation of the sound field including distance is introduced. The proposal is independent of the recording and reproduction systems. In this sense, it is similar to HOA. By using a cylindrical microphone array, it benefits from a privileged direction and the possibility of sampling the field across a wider region. The encoding scheme follows the same paradigm as HOA. It starts from the general solutions to the Helmholtz equation in cylindrical coordinates. These were derived in Chapter III and are reproduced here.

$$\psi^{\pm}_{n,\xi}(r,\theta,z) = \mathrm{C}^{\pm}_{n,\xi}(k) J_n(\xi k r) e^{i(n\theta \pm kz\sqrt{1-\xi^2})}. \tag{6.2}$$

A set of coefficients $\mathrm{C}^{\pm}_{n,\xi}(k)$ can fully characterize any arbitrary sound field, while the general solutions formed by the product of the Bessel functions J_n and a complex exponential, the cylindrical harmonics, are a basis of the Helmholtz equation's solution space. Therefore, they can be used as to encode any possible sound pressure

distribution on the cylinder. The coefficients are classified according to their degree n, *damping ratio* ξ, and orientation (positive or negative).

The basis functions that appear in Eq. (6.2) show three distinct patterns depending on the value of ξ. These are illustrated in Fig. 3.2. When $\xi = 1$ the functions no longer depend on z and the decomposition reduces to the circular harmonic expansion. This case occurs when encoding a plane wave whose wavefronts are parallel to the cylinder axis. If $\xi < 1$ the argument for the exponential in the basis functions is purely imaginary. The basis functions will therefore exhibit an oscillatory behavior on both the axial and polar coordinates. These additional functions serve to encode plane waves incident at oblique angles. In particular, the damping ratio associated with an arbitrary plane wave is $\xi = \sin(\phi_{\text{inc}})$. Finally, the case when $\xi > 1$ leads to a real part in the argument of the exponential functions. The resulting functions diverge towards either the positive or negative direction of the z-axis. These functions help to encode the evanescent component of the sound field. They are needed to encode localized sources, such as monopoles.

The sound pressure measured by the microphones should not be encoded directly since it includes the scattering effects of the cylindrical baffle. In the ideal case of a rigid scatterer, it is possible to model the baffle effects and design a set of filters to recover the original sound pressure distribution. The general solution for the scattering of an arbitrary sound field is given by the following equation:

$$\begin{aligned} p(r,\theta,z,k) =& \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{in\theta}\int_0^{\infty}\left[\mathrm{C}^{+}_{n,\xi}(k)e^{ikz\sqrt{1-\xi^2}} + \mathrm{C}^{-}_{n,\xi}(k)e^{-ikz\sqrt{1-\xi^2}}\right] \\ & \left[J_n(\xi kr) - \frac{J'_n(\xi kR_{\text{cyl}})}{\mathrm{H}'_n(\xi kR_{\text{cyl}})}\mathrm{H}_n(\xi kR_{\text{cyl}})\right] d\xi. \end{aligned} \tag{6.3}$$

This equation is valid for all $r \geq R_{\text{cyl}}$. The divergent solutions, while unphysical, occur when calculating the field of a monopole in the vicinity of the source and are,

therefore, consistent with the wave equation.

Considering a microphone array for which all transducers are positioned precisely on the surface of the rigid cylinder, Eq. (6.3) can be simplified as follows:

$$p(R_{\text{cyl}},\theta,z,k) = \frac{1}{\pi^2 k R_{\text{cyl}}} \sum_{n=-\infty}^{\infty} e^{in\theta} \int_0^{\infty} \Big[\mathrm{C}^{+}_{n,\xi}(k) e^{ikz\sqrt{1-\xi^2}} + \mathrm{C}^{-}_{n,\xi}(k) e^{-ikz\sqrt{1-\xi^2}} \Big] \frac{i}{\xi \mathrm{H}'_n(\xi k R_{\text{cyl}})} d\xi. \tag{6.4}$$

The equation above can be easily inverted since the general solutions to the wave equation are orthogonal. The result is the encoding of the sound pressure distribution on a cylindrical boundary from the sound pressure observed over a rigid cylinder:

$$\mathrm{C}^{+}_{n,\xi}(k) = -i\pi^2 \xi k R_{\text{cyl}} \mathrm{H}'_n(\xi k R_{\text{cyl}}) \int_{z=-\infty}^{0} \int_{\theta=-\pi}^{\pi} p(\theta,z,k) e^{-in\theta} e^{-ikz\sqrt{1-\xi^2}} d\theta dz, \tag{6.5}$$

$$\mathrm{C}^{-}_{n,\xi}(k) = -i\pi^2 \xi k R_{\text{cyl}} \mathrm{H}'_n(\xi k R_{\text{cyl}}) \int_{z=0}^{\infty} \int_{\theta=-\pi}^{\pi} p(\theta,z,k) e^{-in\theta} e^{ikz\sqrt{1-\xi^2}} d\theta dz. \tag{6.6}$$

Equations (6.5) and(6.6) are the cylindrical version of the HOA encoding presented in Chapter II as Eq. (2.24) after considering the effects of the baffle. They are the defining equations of *3D Cylindrical Ambisonics*, the sound field encoding proposed in this chapter.

In practice, it is impossible to sample a continuous surface. Approximating the integrals in the equations above requires proper quadrature weights which depend on the actual array layout. Integration along the z-coordinate must also be limited to a finite interval. These contributions, along with the effects of the baffle, can be summarized in a set of filters $w^{\pm}_{n,\xi}(z,k)$. The resulting filters for a total of N_{mic} microphones uniformly distributed along both coordinates, θ and z, within the interval $[-z_{\text{M}}, z_{\text{M}}]$, and applying a Tukey window with a taper-to-length ratio of

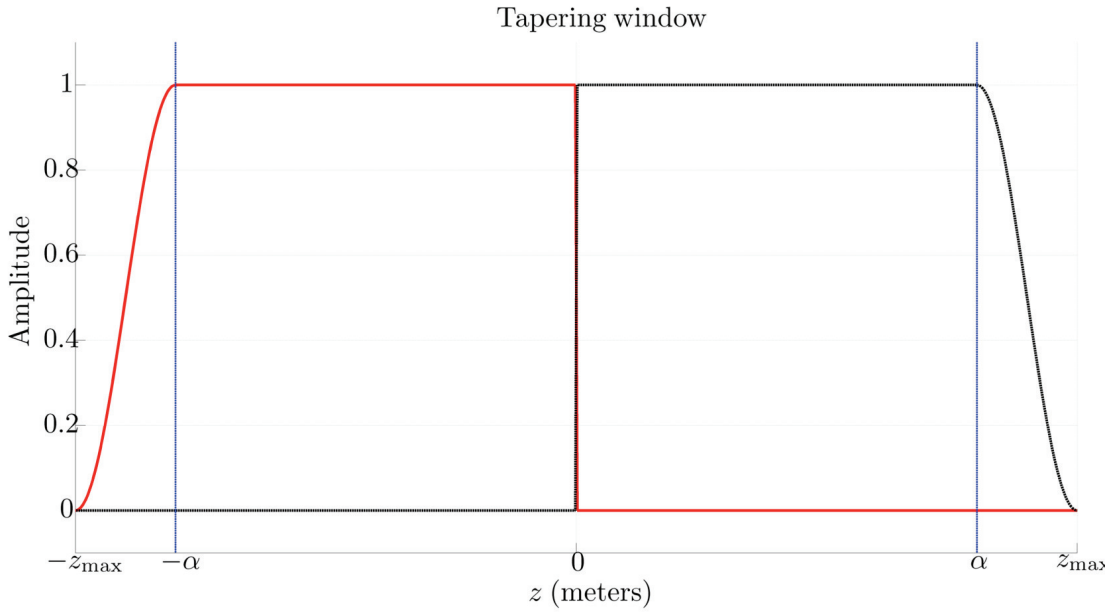


Figure 6.1: The fragmented Tukey window applied to finite-length microphone arrays. The purpose of this spatial window is to prevent large peaks due to the discontinuity in the Fourier series along the axial coordinate; the Gibbs phenomenon.

α are given by the following expression:

$$w^{\pm}_{n,\xi}(z,k)=\begin{cases} \frac{-i\pi^2\xi k R_{\text{cyl}}\mathrm{H}'_n(\xi k R_{\text{cyl}})}{2N_{\text{mic}}}\left(\cos\left[\frac{\pi}{\alpha}\left(\frac{|z|}{z_{\text{M}}}+\alpha-1\right)\right]+1\right) & \text{if } (\alpha-1)z_{\text{M}}>\pm z\geq -z_{\text{M}}, \\ \frac{-i\pi^2\xi k R_{\text{cyl}}\mathrm{H}'_n(\xi k R_{\text{cyl}})}{N_{\text{mic}}} & \text{if } 0\geq \pm z\geq (\alpha-1)z_{\text{M}}, \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

The Tukey windows used in this equation are shown, without the contribution from the baffle scattering, in Fig. 6.1.

The filters of Eq. (6.7) include all of the parameters related to the microphone array. They can be applied to the signals measured by each microphone in order to produce an encoding of the sound field. The encoding equation used to derive the coefficients $\mathrm{C}^{\pm}_{n,\xi}(k)$ from cylindrical microphone array recordings is:

$$\mathrm{C}^{\pm}_{n,\xi}(k) \approx \sum_{\text{mic}=1}^{N_{\text{mic}}} w^{\pm}_{n,\xi}(z_{\text{mic}},k)p_{\text{mic}}(\theta_{\text{mic}},z_{\text{mic}},k)e^{-in\theta_{\text{mic}}}e^{\mp ikz_{\text{mic}}\sqrt{1-\xi^2}}. \tag{6.8}$$

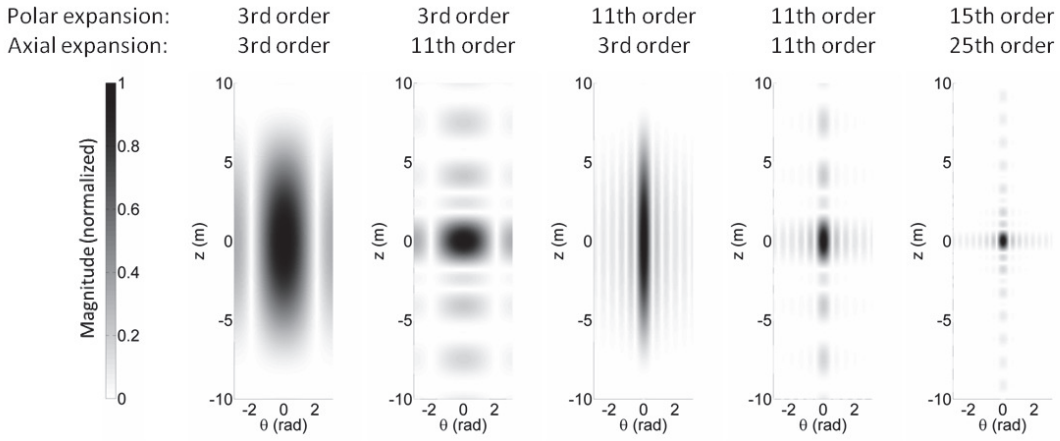


Figure 6.2: Effects of discarding high-order coefficients in the expansion of a delta function centered at the origin. The panels show the approximated delta function for different sets of harmonic expansion coefficients. The polar and axial resolutions can be chosen independently depending on the application requirements.

6.2.1 Truncation of the cylindrical harmonic expansion

The complete description of an arbitrary sound field using Eq. (6.2) requires an infinite number of coefficients $\mathrm{C}^{\pm}_{n,\xi}(k)$. Practical systems can only record and use a finite number of expansion coefficients, reducing the resolution of the encoding. In this section we consider the separate effects of discarding high-order coefficients for the polar and axial encodings.

To quantify the loss of accuracy, we consider the completeness property of the cylindrical harmonic functions:

$$\begin{aligned}
&\delta(\theta-\theta')\delta(z-z') \\
&= \frac{1}{4\pi^2}\int_{\xi=0}^{\xi=\infty}\sum_{n=-\infty}^{\infty} e^{i(n\theta\pm kz\sqrt{1-\xi^2})}e^{-i(n\theta'\pm kz'\sqrt{1-\xi^2})}d\xi \\
&= \frac{1}{4\pi^2}\sum_{n=-\infty}^{\infty} e^{in(\theta-\theta')}\int_{k_z=-\infty}^{k_z=\infty} e^{ik_z(z-z')}dk_z.
\end{aligned} \tag{6.9}$$

Here, the change of variables $k_z = k\sqrt{1-\xi^2}$ was applied. The coordniate θ is, of course, periodic every 2π. The completeness relation treats the polar and axial

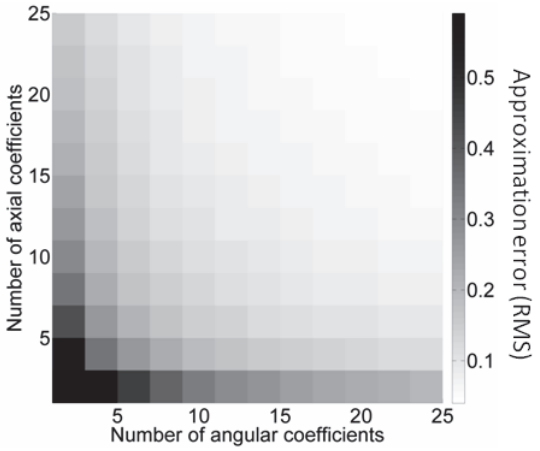


Figure 6.3: RMS of the error in the approximation of the delta function for different number of polar and axial expansion coefficients. The total number of coefficients used in the approximation is the product of both of these values.

coordinates separately. The results are two well-known [18] expressions for the delta function and the delta comb. The effects of truncating the expansion of any sound field will, at most, correspond to the deviation from the delta function exhibited by a similarly truncated Eq. (6.9).

Figure 6.2 gives a few examples illustrating the effects of an incomplete expansion. The RMS value of the error in the approximation of the delta function can be used as a measure of spatial accuracy. Figure 6.3 shows the value of this error for different numbers of expansion coefficients. As expected, the spatial resolution of the microphone arrary is closely related to the number of expansion coefficients. More coefficients always provide a higher resolution description of the field.

A final consideration is that of spatial aliasing in the description of the sound field. The truncation of Eq. (6.2) limits the maximum frequency that the microphone array is able to resolve accurately. A limiting frequency can be calculated by noting that the polar and axial coordinates are completely separated in Eq. (6.9). Direct application of the Nyquist theorem gives the following expression for the alias

frequency:

$$f_{\text{polar}}^{\text{alias}} = \frac{c(N_{\text{polar}} - 1)}{4R_{\text{cyl}}}. \tag{6.10}$$

N_{polar} stands for the number of angular coefficients used in the encoding.

The maximum frequency to be encoded along the axial coordinate is given by the maximum value of k_z. However, the maximum axial frequency that can be resolved depends on the microphone separation along the z-coordinate, Δz_{mic}.

$$f_{\text{axial}}^{\text{alias}} = \frac{c}{2\Delta z_{\text{mic}}}. \tag{6.11}$$

6.3 Reproducing 3D Cylindrical Ambisonics

A sound field encoded using Eq. (6.8) can be reproduced using most surrounding loudspeaker arrays; the reproduction system is not required to have a cylindrical geometry. The coefficients $\mathrm{C}_{n,\xi}^{\pm}(k)$ can be decoded to produce loudspeaker signals using any of the techniques used to reproduce Ambisonics. The simplest approach would use Eq. (6.2) directly to calculate the sound pressure at the loudspeaker positions and a tapering window to turn off the loudspeakers that are away from the source's direction of incidence. This, however, requires knowledge of the sound source positions. Another possibility is to define a re-encoding matrix by calculating, from Eq. (6.2), the coefficients $\hat{\mathrm{C}}_{n,\xi}^{\pm}(k)$ corresponding to the contribution of each loudspeaker towards the re-created sound field. For an ideal monopole radiator, these coefficients are given by the Hankel functions:

$$\hat{\mathrm{C}}_{n,\xi}^{\pm}(k;\Delta r) = \frac{i}{4}\mathrm{H}_n\left(\xi k|\Delta r|\right). \tag{6.12}$$

The re-encoded field over a cylindrical secondary surface for a loudspeaker located at $(r_{\text{spk}}, \theta_{\text{spk}}, z_{\text{spk}})$ is defined as the linear combination of cylindrical harmonic functions:

$$p(r,\theta,z,k) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{in(\theta_{\text{spk}}-\theta)} \int_0^{\infty} \Big[\hat{\text{C}}^{+}_{n,\xi}\left[k;(r_{\text{spk}}-r)\right] e^{ik(z_{\text{spk}}-z)\sqrt{1-\xi^2}} + \hat{\text{C}}^{-}_{n,\xi}\left[k;(r_{\text{spk}}-r)\right] e^{-ik(z_{\text{spk}}-z)\sqrt{1-\xi^2}} \Big] d\xi. \tag{6.13}$$

Equation (6.13) is the 3D cylindrical Ambisonics equivalent of Eq. (2.23). The loudspeaker signals can be chosen to minimize the square error in the re-created sound pressure distribution $p(r, \theta, z, k)$ over a secondary surface. In particular, when the loudspeakers are arranged on a cylinder surrounding the listener, the use of Eq. (6.13) to decode the sound field description will yield the same results as the wave field reconstruction filters of Eq. (6.1). It is also possible to adapt the more advanced techniques used to decode Ambisonics for reproduction using arbitrary loudspeaker arrays [36].

6.4 Numerical simulation results

The viability of the proposed encoding scheme was evaluated using the numerical simulation of a cylindrical microphone array. The recording device consisted of 100 microphones distributed in 20 rings over a 95-centimeters-long region of a rigid cylinder. The spacing between rings and angular spacing between microphones was constant at ($\Delta\theta = 72°, \Delta z = 5\ cm$). The radius of the cylinder was chosen as 5 cm. A Tukey window with a taper-to-length ratio of 0.1 was applied over the interval $z = [-0.5\ m, 0.5\ m]$ and the encoding of a 1-kHz plane wave was computed using Eqs. (6.7) and (6.8). The plane wave incided over the cylinder at a polar angle of $\theta_{\text{inc}} = 15°$ and the angle between its wavevector and the cylinder's axis was $\phi_{\text{inc}} = 45°$. The maximum polar degree calculated was $n = \pm 2$, and the damping factor was sampled at increments of 0.25 between 0 and 1. Therefore, the encoding of the sound field consisted of 45 coefficients $\mathrm{C}^{\pm}_{n,\xi}(k)$.

The sound field description was decoded using the least-squares-error criteria to generate signals for a 162-channel spherical loudspeaker array. The loudspeakers were treated as ideal monopole sources and situated on the vertices of a subdivided icosahedron, one meter away from the center. The secondary surface used to calculate the least-squares error was a 1-meter-long cylinder with a radius of 15 cm located at the center of the loudspeaker array.

Figure 6.4 shows the numerical simulation results for the recording and reproduction of a plane wave using the cylindrical microphone array and spherical loudspeaker array described above. As expected, the sound field is re-created more accurately inside the secondary control region; however, the reconstruction is acceptable even outside of it.

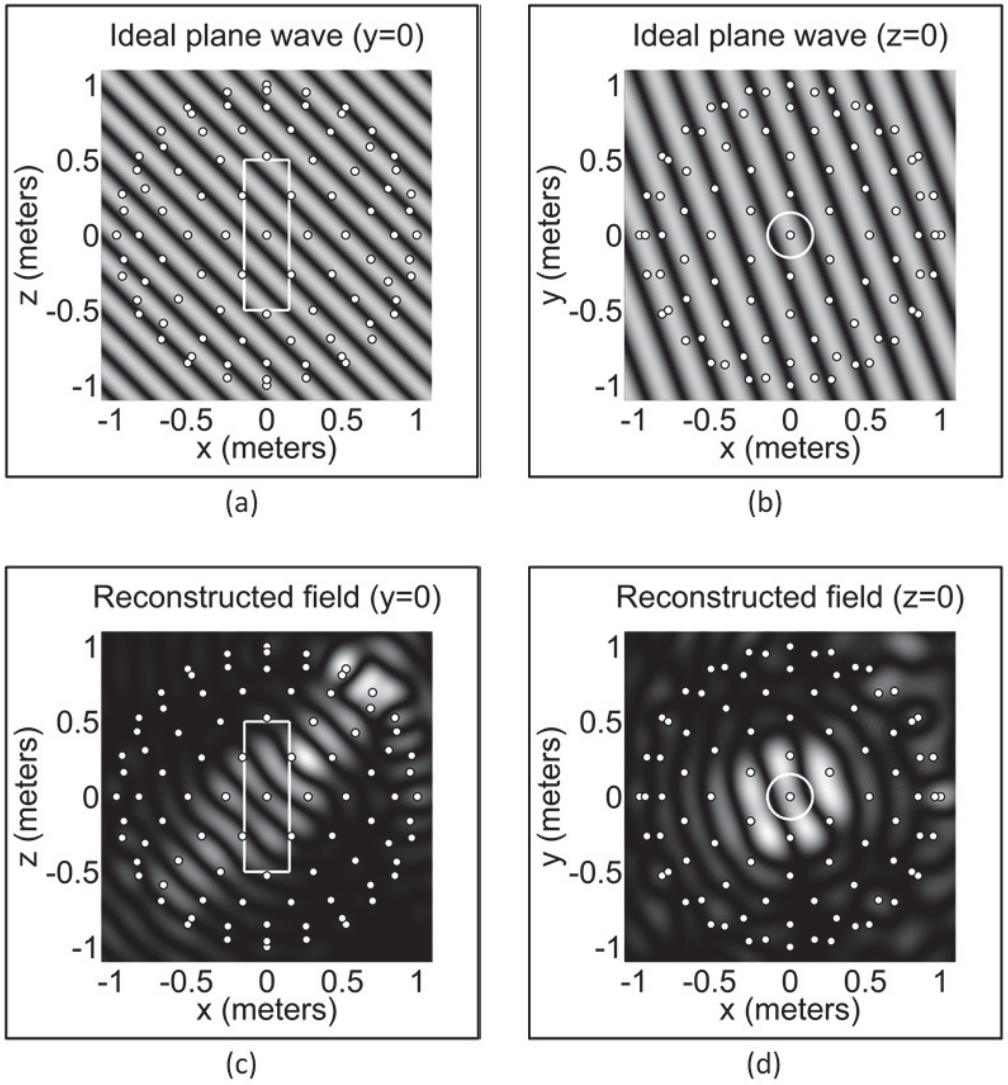


Figure 6.4: Numerical simulation results of applying the proposed encoding to sound field reproduction. (a) and (b) Original sound field consisting of an 1-kHz plane wave. (c) and (d) Re-created sound field. White circles show the loudspeaker positions; the white rectangle and circumference delineate the control surface.

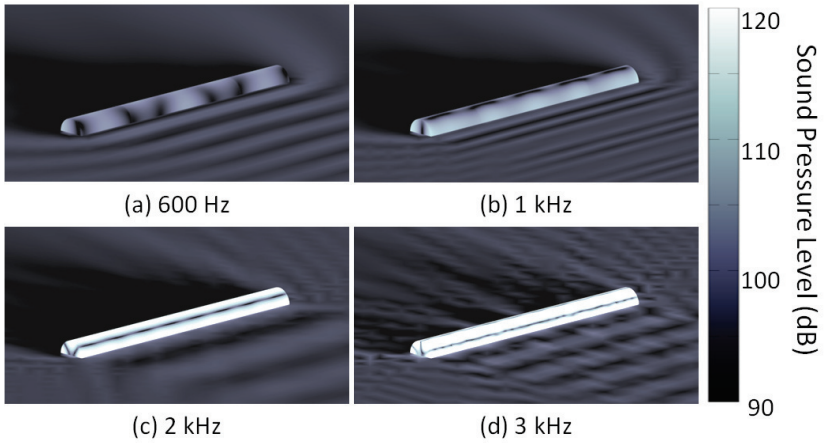


Figure 6.5: Sound pressure level on the surface of a rigid cylinder of finite length. The panels show the Boundary Element Method computation results for plane waves of different frequencies, all of them parallel to the axis of the cylinder.

6.5 Effects of a finite-length baffle

Any practical implementation of the cylindrical microphone arrays under consideration requires the baffle to be truncated. The spatial window of Eq. (6.7) was derived from the scattering model of an infinite rigid cylinder. Unfortunately, there are no similarly simple solutions for the scattering of a finite cylinder.

To estimate the effects of baffle truncation in the encoding of the sound field, a computer simulation using the Fast Multipole Boundary Element Method [29] was used. Figure 6.5 shows the results of the simulation for different frequencies. The cylinder used was 1.5 meters long with a diameter of 0.15 meters. The average size of the cells used in the 3D model was 8.8×10^{-3} m; the simulation is accurate up to around 5 kHz. A single plane wave was considered as the incident sound field. The wavevector is perpendicular to the axis of the cylinder.

The simulation shows significant effects on the scattered field over the cylinder's surface at low frequencies. In general, these edge contributions are periodic along the z-coordinate due to the symmetries involved. However, the effects are

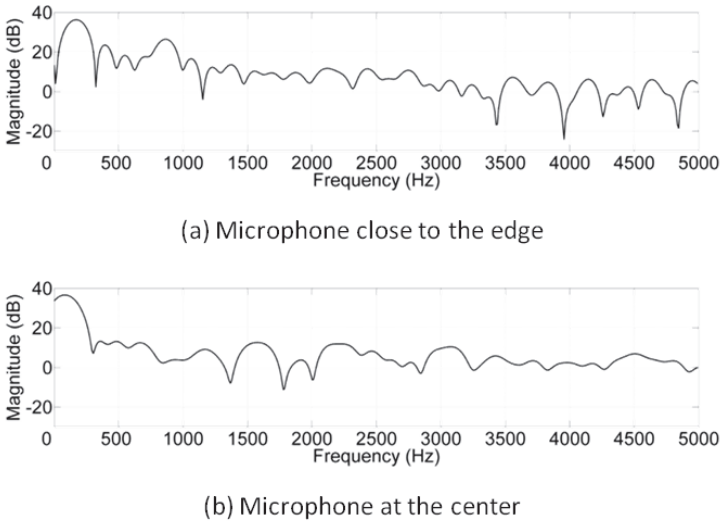


Figure 6.6: Differences between the scattered fields of a finite-length and an infinite-length cylinder. The panels show the magnitude difference in the sound field recorded by microphones on the surface of a truncated cylindrical baffle when compared to the theoretical result for an infinite rigid cylinder. Panel (a) corresponds to a microphone located 0.25 meters away from the cylinder's edge. Panel (b) is for a microphone located at the center of the 1.5-meters long cylinder.

considerably reduced when the wavelength of the incident plane wave drops below the cylinder's diameter.

The results of the BEM simulation are summarized in Fig. 6.6. Truncation effects are particularly significant at low frequencies and for microphones located close to the edges. However, once the wavelength of the sound source falls under two times the cylinder's diameter, around 1.15 kHz in our simulation, the scattering of the baffle can be considered to be close to that of an infinite cylinder.

6.6 Summary

This chapter introduced a sound field encoding method for cylindrical microphone array recording. The proposal, which has been named *3D Cylindrical Ambisonics*, can be used to store or broadcast sound field information for reproduction using most loudspeaker arrays. In this sense, it offers advantages similar

to those of HOA in spherical geometries. However, using the proposed method, it is easy to change the azimuth and elevation resolutions independently. The number of transducer rings or staves in the recording system can be chosen to fit the application requirements.

The encoding method presented considers a finite-length microphone array by introducing a set of spatial windows. This is an important requirement for the eventual application of the methods introduced to an actual microphone array.

Expressions showing the limits of cylindrical microphone arrays, due to spatial aliasing, were also presented. On the other hand, the effects of truncating the cylindrical harmonic expansion were also considered and found to be similar to those observen in HOA where the spherical harmonic expansion is truncated. In particular, the resolution increases and the error decreases monotonically as more expansion coefficients are considered.

A finite-length baffle was considered using Boundary Element Method simulations and, while truncating the baffle leads to artifacts in the recording at low frequencies, once the wavelength of the sound source is comparable to the radius of the baffle, it can be considered to be infinite with good accuracy.

CHAPTER VII

Conclusions

Through this dissertation, a series of techniques to record, analyze, encode and reproduce sound fields have been explored. The presentation starts with Chapter I offering an overview of the field of spatial audio. Stating the need for systems that can record and reproduce sounds while preserving the spatial information that humans use to determine the position of sound sources in an acoustic scene.

Chapter II offers a review of existing techniques. In particular, High-Order Ambisonics is singled out since it is the only method which introduces a scalable and system-agnositc encoding capable of characterizing the entire sound field information to any desired accuracy. These properties are extremely useful in the design of future-proof, ultra-realistic spatial audio systems. Therefore, this dissertation adopts these two properties as constraints required of all the techniques proposed by it.

Taking HOA as a basis, Chapter III of this dissertation seeks to eliminate the need for a spherical geometry in the recording and reproduction systems. In HOA, the spherical harmonic functions are used to characterize the sound field. Therefore, sound measurements made from a single privileged position and sampling all directions uniformly are required. This imposes unnecessary constraints on the recording and reproduction systems, and makes it difficult to present sound to more

than one listener. An alternative set of basis functions, the cylindrical harmonic functions, are derived in Chapter III and proposed as a replacement of the spherical harmonics. In particular, the case of encoding plane waves is considered and a new expression for the plane wave decomposition in cylindrical coordinates was derived. Previous research had only considered plane waves that are parallel to the axial coordinate; however, in this dissertation, a mathematical expression to encode all possible plane waves is derived for the first time.

Chapter IV of this dissertation takes a step back to consider the application of cylindrical microphone arrays to construct spherical harmonic encodings. In particular, it is observed that the spatial resolution of a cylindrical microphone array can be made to vary independently along the polar and axial coordinates. The present dissertation considers the properties of spatial hearing in humans and opts to develop a method that will provide higher resolution in the horizontal plane at the expense of reduced resolution in the encoding of elevation. The techniques developed in Chapter IV can produce a conventional Mixed-Order Ambisonics (MOA) encoding of sound fields having arbitrary and independent horizontal and vertical resolutions. To achieve this, a new method to create a virtual spherical microphone array from a cylindrical one is advanced. The encodings produced by this proposal are fully compatible with existing Ambisonics reproduction systems.

The problem of reproducing sound fields using loudspeakers is considered in Chapter V. Harmonic encodings of sound fields can be easily decoded and reproduced if the recording and reproduction arrays share similar geometries. However, this is hard to achieve in practice, particularly in the case of spherical geometries due to the need to distribute large numbers of loudspeakers regularly at all angles, such as below the listener. In this chapter, a new error metric is advanced as an useful

way to optimize HOA decoders when the reproduction array is irregular. The proposed metric focuses not on the reconstruction error at a single control point, like conventional decoders do, but on the change of said error as the observation point is moved away from the priviledged listening position. The result of applying this new metric in the least-squares formulation of HOA decoders is a more stable reproduction that remains free of audible artifacts over larger regions than those generated by previous methods. This result is particularly important in the reproduction of sound fields using irregular loudspeaker arrays since traditional decoders targeting them tend to be unstable. The instability is ameliorated by minimizing the proposed error metric.

Finally, Chapter VI of this dissertation introduces what may be the most salient contribution of this research. The benefits of cylindrical microphone arrays that were observed in Chapters III and IV can be fully exploited by using a new encoding method for sound field information. This method, called 3D Cylindrical Ambisonics (3DCA), can achieve optimal performance for a cylindrical microphone array and produce complete encodings that can be reproduced with loudspeaker systems of an arbitrary geometry. The proposed encoding scheme allows for an independent choice of horizontal resolution while reducing system complexity by lowering the number of channels devoted to encode elevation. The encoding equations for ideal and realistic microphone arrays are derived. Evaluation is carried out to consider the effects of microphone truncation, self-noise, uneven calibration and misplacement An outline of the decoding algorithm for 3DCA is presented in both, an analytic formula considering ideal reproduction systems and a least-squares formulation for realistic loudspeaker arrays. The decoding of 3DCA can be further improved with the addition of stabilization metrics like the one introduced in Chapter

V.

In summary, this dissertation seeks to solve three problems in the field of sound field recording and reproduction. First, it considers the solutions to the Helmholtz equation not in spherical, but in cylindrical coordinates to define a solid ground for the analysis of sound fields in geometries different to the spherical one. This led to an encoding of the far field using the plane wave decomposition of a cylindrical pressure distribution. An extension to this allows for the encoding of Mixed-Order Ambisonics descriptions of sound fields from cylindrical microphone array recordings. Second, a new HOA decoder that can be used with irregular and non-spherical loudspeaker arrays was introduced. The main difference with other approaches is the introduction of a constraint on the behavior of the reconstruction away from the listening position. This resulted in a larger listening region and better stability when the reproduction array does not match the HOA geometry. Finally, a new encoding method, 3D Cylindrical Ambisonics, is developed to encode sound field information measured with a cylindrical microphone array directly. The encoding is similar to HOA in the sense of being scalable and system-agnostic. However, unlike HOA, it allows for the independent definition of axial and polar resolutions. This feature makes it more attractive to record and present sound fields to large audiences which can be aligned along the preferred axis.

This dissertation outlines the basis for sound field recording, analysis and reproduction using non-spherical arrays and in particular a cylindrical microphone array. It is hoped that the proposal presented here, in particular 3D Cylindrical Ambisonics can become an important step towards the realization of ultra-realistic spatial audio systems.

APPENDIX

A Practical considerations for cylindrical microphone arrays

A.1 Overview

Chapter VI introduced 3D Cylindrical Ambisonics, a sound field encoding method for cylindrical microphone array recordings. The format introduces some errors in the form of a limited spatial resolution due to a finite expansion in terms of cylindrical harmonics. This error, however, can be made as small as desired by simply calculating more terms in the expansion.

In practical situations, where a microphone array is being designed, there are other sources of error which cannot be eliminated by increasing the amount of signal processing. This appendix evaluates these considerations which are an important factor for the design and use of an actual cylindrical microphone array.

A.2 Errors in microphone placement and self-noise

It is impossible to be perfectly accurate when building a physical microphone array. The encoding of Eq. (6.8), however, assumes knowledge of the microphone positions on the cylinder. Furthermore, the signals obtained from the microphones will not correspond perfectly to the sound pressure at their position since all physical microphones suffer from some degree of self-noise. The effects of these two error

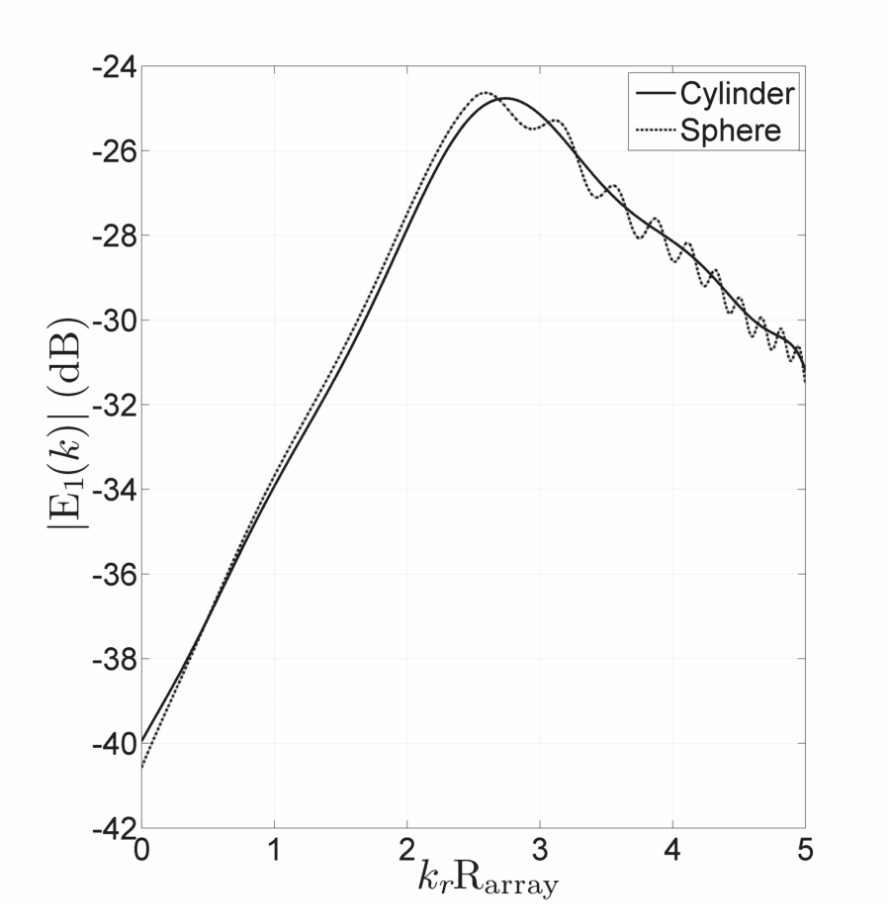


Figure A-1: Magnitude of the encoding coefficients for cylindrical and spherical arrarys when the microphone signals correspond to uncorrelated white noise. This value provides a measure for the impact of microphone misplacement and self-noise in the spatial encoding of sound fields. The graphic shows the results for arrays of 100 microphones and first-order angular expansions.

sources have been studied for arrays using spherical and cylindrical baffles. Existing results show that both factors impact the microphone array resolution in a similar way [27]. Based on this result, it is possible to analyze the effects of sensor misplacement and self-noise by introducing additive error into the sound pressure measurements $p_{\mathrm{mic}}(\theta_{\mathrm{mic}}, z_{\mathrm{mic}}, k)$ of Eq. (6.8).

In order to evaluate the encoding error introduced by the transducers, an

additive noise signal

$$\begin{aligned}\hat{\mathrm{C}}^{\pm}_{n,\xi}(k) \approx \sum_{\mathrm{mic}=1}^{N_{\mathrm{mic}}} w^{\pm}_{n,\xi}(z_{\mathrm{mic}}, k)\, [p_{\mathrm{mic}}(\theta_{\mathrm{mic}}, z_{\mathrm{mic}}, k) \\ +\mathrm{E}_{\mathrm{mic}}(\theta_{\mathrm{mic}}, z_{\mathrm{mic}}, k)]\, e^{-in\theta_{\mathrm{mic}}} e^{\mp ikz_{\mathrm{mic}}\sqrt{1-\xi^2}} \\ = \mathrm{C}^{\pm}_{n,\xi}(k) + \mathrm{E}^{\pm}_{n,\xi}(k). \end{aligned} \tag{A.1}$$

The encoding error can be easily separated due to the linearity of Eq. (6.8). The impact of microphone misplacement and self-noise in the proposed encoding can be measured as the expansion coefficients of noise, labeled here as $\mathrm{E}^{\pm}_{n,\xi}(k)$.

Our evaluation considers white noise with a signal-to-noise ratio (SNR) of 50 dB for each microphone. The same conditions were also applied to a spherical microphone array for comparison. Both arrays were chosen to have the same diameter (0.15 m) and number of transducers (100). The microphones were distributed in a rectangular grid for the cylinder and a minimum-energy grid [35] for the sphere. The spherical microphone array signals were encoded using the spherical harmonic expansion [19], while the cylindrical array used our proposal. The magnitude of the encoding error for first-ordedr angular encodings is shown in Fig. A-1.

The impact of transducer error when considering a finite-length cylindrical microphone array with an infinite cylindrical baffle is close to that observed in spherical microphone arrays. Our results are in agreement with those of existing research focusing on spheroidal baffles [27]. This analysis, however, does not consider the effects of a finite-length baffle. The truncation of the baffle is discussed in later sections.

A.3 Differences in microphone calibration

The encoding equation, Eq. (6.8), assumes ideal sound pressure measurements. The impact of self-noise in any physical microphone was considered in the previous section. However, another important source of error lies in the fact that different microphones in an array will always present different characteristics. The differences can never be completely removed, even after a careful microphone calibration procedure. In this section we evaluate the impact of these differences on the proposed encoding method and compare our results with those observed in the case of spherical microphone arrays.

The frequency responses of all transducers in an array should be fairly similar if high-quality microphones of the same model are used. However, a small gain difference will remain even after calibration. In most cases, however, calibration should reduce this gain difference between sensors to levels under 1 dB. The impact of unknown, random gain differences $\Delta \mathrm{G}_{\mathrm{mic}}$ on the encoding equation can be stated as follows:

$$\begin{aligned} \hat{\mathrm{C}}^{\pm}_{n,\xi}(k) \approx & \sum_{\mathrm{mic}=1}^{N_{\mathrm{mic}}} w^{\pm}_{n,\xi}(z_{\mathrm{mic}}, k)(1+\Delta \mathrm{G}_{\mathrm{mic}}) p_{\mathrm{mic}}(\theta_{\mathrm{mic}}, z_{\mathrm{mic}}, k) \\ & e^{-in\theta_{\mathrm{mic}}} e^{\mp ikz_{\mathrm{mic}}\sqrt{1-\xi^2}} \\ = & \, \mathrm{C}^{\pm}_{n,\xi}(k) + \mathrm{E}^{\pm}_{n,\xi}(k). \end{aligned} \tag{A.2}$$

When we considered transducer misplacement and self-noise, the encoding error $\mathrm{E}^{\pm}_{n,\xi}(k)$ was independent of the sound pressure at the ideal microphone positions $p_{\mathrm{mic}}(\theta_{\mathrm{mic}}, z_{\mathrm{mic}}, k)$.. This time, however, the encoding error includes these

measurements and is given by

$$\mathrm{E}^{\pm}_{n,\xi}(k) = \sum_{\mathrm{mic}=1}^{N_{\mathrm{mic}}} w^{\pm}_{n,\xi}(z_{\mathrm{mic}}, k)\Delta \mathrm{G}_{\mathrm{mic}} p_{\mathrm{mic}}(\theta_{\mathrm{mic}}, z_{\mathrm{mic}}, k)$$

$$e^{-in\theta_{\mathrm{mic}}} e^{\mp ikz_{\mathrm{mic}}\sqrt{1-\xi^2}}. \tag{A.3}$$

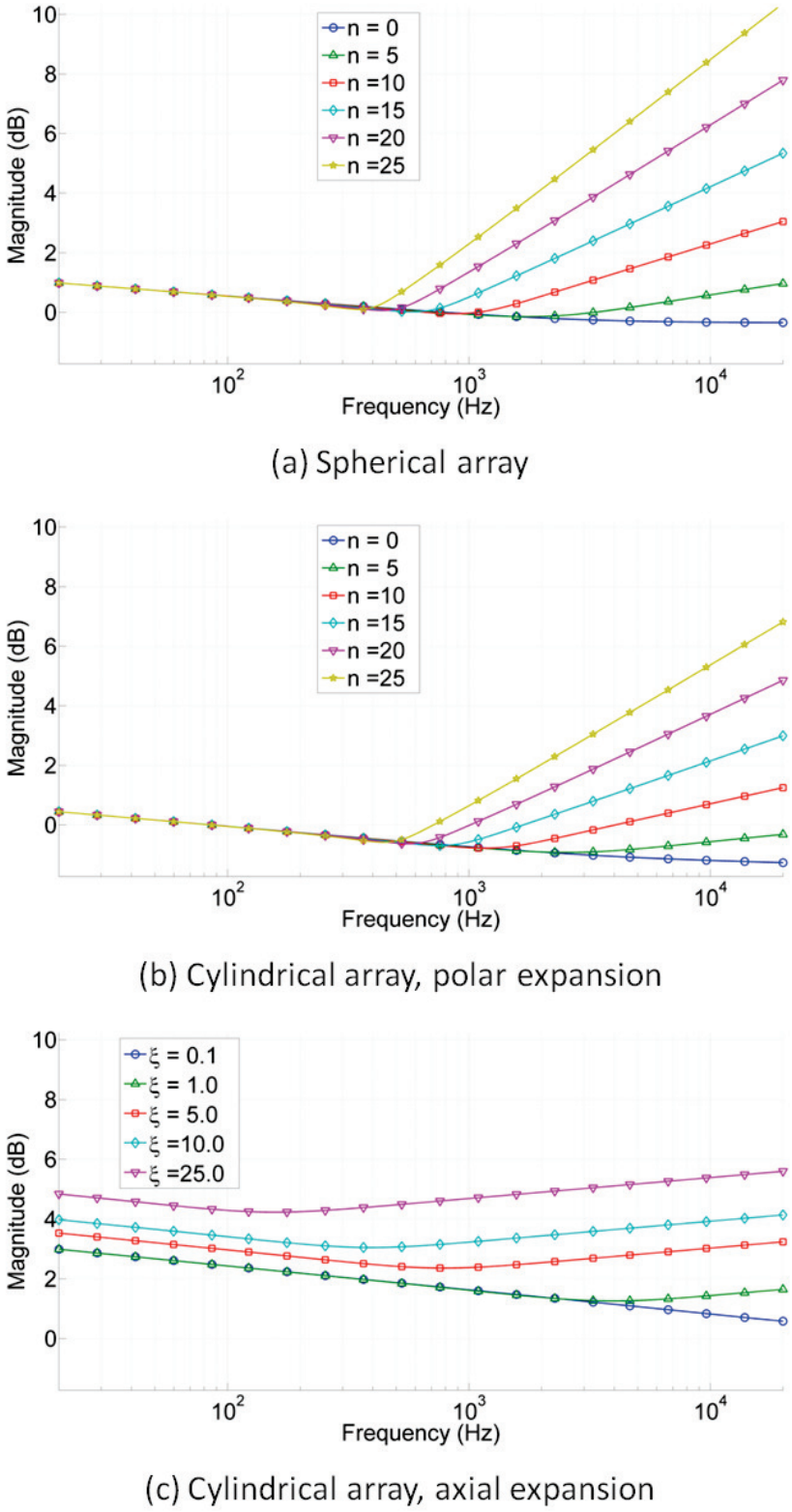


Figure A-2: Encoding error due to transducer calibration differences for cylindrical and spherical arrarys. Panel (a) shows the impact of slightly different microphone gains (± 1 dB uniformly distributed) on the accuracy of the spherical harmonic expansion. Panels (b) and (c) show the results for a cylindrical array in relation to the number of polar and axial expansion coefficients, respectively. All results assume arrays of 750 microphones.

To evaluate the impact of these gain differences, we considered two 750-channel microphone arrays: a cylindrical and a spherical one. The cylindrical array uses a uniform rectangular grid, while the spherical one uses an approximately regular distribution [35]. Uniformly-distributed random gain differences of up to 1 dB in magnitude were assumed. After computing the expansion coefficients, the inverse operation was applied to recover a sound pressure distribution. In the cylindrical case, this is done by applying Eq. (6.2). For the spherical case we used Eq. (1.5). The root-mean-square value of the resulting sound pressure distributions was calculated and the results for both, cylindrical and spherical arrays, at different expansion orders are shown in Fig. A-2.

Our results show that the impact of different microphone characteristics is negligible for low-order expansions. On the other hand, both types of microphone array, cylindrical and spherical, show a significant increase in the encoding error when high-order expansion coefficients are considered. The results are not surprising since the truncation of the harmonic expansions can be seen as the spatial low-pass filtering of the sound pressure distributions. The contribution of gain differences will appear at all spatial frequencies if they are truly random. As the order increases, so does the cutoff frequency of the spatial low-pass filter, thus more and more of the gain difference contributions are being allowed to appear in the results. Figure A-2, however, does show that the cylindrical arrangement is not significantly more susceptible to these errors than spherical ones.

A.4 Summary

This appendix reviews some important considerations for the implementation of the techniques outlined in Chapters III. IV and VI. In concrete, the effects of some inevitable sources of error when building a loudspeaker array are considered.

The effects of microphone self-noise, as well as errors in microphone placement are discussed. The impact of these on the 3D Cylindrical Ambisonics encoding are considered and found to be comparable to those faced by HOA when using a spherical microphone array. However, while the general behavior of the error is similar, cylindrical microphone arrays are more robust to these sources of error and, therefore, impose less stringent requirements on signal-to-noise ratio or manufacturing precision.

ACKNOWLEDGEMENTS

The work and results that form the body of this dissertation would not have been possible without the help and support received from those around me. I would like to express my deepest gratitude to all those who, through their assistance, made this research possible.

I would like to make a special mention of my academic advisor, Professor Suzuki Yôiti, who was been very supportive ever since I sought enrollment in Tohoku University. Besides giving me the invaluable opportunity to join the laboratory under his supervision, he has always been encouraging of my research and, through his advise, he helped me develop the ideas that form the core of this dissertation. I extend my gratitude to Professors Ito Akinori and Suganuma Takuo, members of my evaluation committee, for all the feedback they provided me during the preliminary presentations of this dissertation. Likewise, I wish to acknowledge the constant support and advise from Associate Professor Sakamoto Shuichi, also part of this dissertation's evaluation committee, and an invaluable colleague and discussion partner throughout my studies in Tohoku University.

Parts of this dissertation would have been impossible without the help and advise received from Professor Iwaya Yukio and Dr. Okamoto Takuma. They have provided me with crucial technical advise and have been irreplaceable discussion partners since my Master's studies. The results of our collaboration led to a significant part of this dissertation.

I would further like to acknowledge the help I received from Mr. Koyama Shoichi. Our technical discussions have been invaluable to my understanding of the mathematical and physical background at the center of my research. His disponibility and guidance were crucial in the accomplishment of the present research.

Finally, I extend my most sincere gratitude to the many researchers and students I have met during my graduate studies in Tohoku University and who have helped me out whenever I have encountered some difficulty. It is through all of their contributions, large and small, that this research came to a fruitful completion.

BIBLIOGRAPHY

[1] Bell, A.G. (**1876**). "Improvement in Telegraphy", US Patent **174465**.

[2] Edison, T.A. (**1878**). "Phonograph or Speaking Machine", US Patent **200521**.

[3] Blauert, J. (**1996**). "Spatial Hearing", MIT Press, rev. ed.

[4] Musicant, A.D. and Butler, R.A. (**1984**). "The influence of pinnaebased spectral cues on sound localization", J. Acoust. Soc. Am. **75**(4), 1195–1200.

[5] Suzuki, Y. (**2010**). "Auditory displays and microphone arrays for active listening", keynote lecture, 40th Int. AES Conf.

[6] Thurlow, W.R., Mangels, J.W., and Runge, P.S. (**1967**). "Head movements during sound localization", J. Acoust. Soc. Am. **42**(2), 489–493.

[7] Gerzon, M.A. (**1992**). "Panpot Laws for Multipspeaker Stereo", Proc. of the 92nd Conv. of the Audio Eng. Soc. **3309**.

[8] ITU-R Recommendation **BS.775-2** (**2006**). "Multichannel Stereophonic Sound System With and Without Accompanying Picture".

[9] Hamasaki, K., Nishiguchi, T., Okumura, R., Nakayama, Y., Ando, A. (**2007**). "22.2 Multichannel Sound System for Ultra High-Definition TV", Proc. of the SMPTE Tech. Conf.

[10] Gardner W.G. (**1997**). "3-D Audio Using Loudspeakers", Ph.D. Thesis, Massachusetts Institute of Technology.

[11] Berkhout, A.J. (**1988**). "A Holographic Approach to Acoustic Control", J.Audio Eng.Soc. **36**, 977-995.

[12] Nyquist, H. (**1928**). "Certain topics in telegraph transmission theory", Trans. AIEE, **47**, 617–644.

[13] Noisternig, M., Carpentier, T., Warusfel, O. (**2013**). "A Multichannel

Loudspeaker Array for WFS/HOA Sound Spatialization at Ircams Concert Hall", AIA-DAGA Conf. on Acoust.

[14] Ise, S. (**1999**). "A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems", Acustica **85**, 78–87

[15] Gerzon, M.A. (**1973**). "Periphony: With-Height Sound Reproduction", J. Audio Eng. Soc. **21**(1), 2–10.

[16] Malham, D. (**2003**). "Higher order Ambisonic systems. Space in Music - Music in Space", M.Phil. Thesis, University of York.

[17] Poletti, M.A. (**2005**). "Three-dimensional surround sound systems based on spherical harmonics", J. Audio Eng. Soc. **53**(11), 1004–1025.

[18] Arfken, G.B., Weber, H.J. and Harris, F.E. (**2012**). "Mathematical Methods for Physicists: A Comprehensive Guide", Academic Press, 7th ed.

[19] Williams, E.G. (**1999**). "Fourier acoustics: Sound radiation and nearfield acoustical holography", Academic Press, 1st ed.

[20] Teutsch, H. (**2007**). "Modal array signal processing: principles and applications of acoustics wavefield decomposition", Springer, 1st ed.

[21] Jackson, J.D. (**1998**). "Classical Electrodynamics", Wiley, 3rd ed.

[22] Merzbacher, E. (**1997**). "Quantum Mechanics" Wiley, 3rd ed.

[23] Zotkin, D.N., Duraiswami, R., and Gumerov, N.A. (**2010**). "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays", IEEE Trans. on Audio, Speech, and Language Processing **18**(1), 2–16.

[24] Daniel, J. (**2003**). "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format", Proc. 23rd Int. Conf. Audio Eng. Soc. **2003**(16).

[25] Rafaely, B. (**2005**). "Analysis and design of spherical microphone arrays", IEEE Trans. on Speech and Audio Processing **13**(1), 135–143.

[26] Parthy, A., Epain, N., van Schaik, A., and Jin, C.T. (**2011**). "Comparison of the measured and theoretical performance of a broadband circular microphone array", J. Acoust. Soc. Am. **130**(6), 3827–3837.

[27] Holmes, S. (**2013**). "Circular harmonics beamforming with spheroidal baffles", Proc. Int. Congr. on Acoustics 2013 **POMA 19**(055077), 1–9.

[28] Travis, C. (**2009**). "A New Mixed-Order Scheme for Ambisonic Signals", Proc. Ambisonics Symp.

[29] Liu, Y. (**2009**). "Fast Multipole Boundary Element Method: Theory and Applications in Engineering", Cambridge University Press, 1st ed.

[30] Bertilone, D.C., Killeen, D.S., and Bao C. (**2007**). "Array gain for a cylindrical array with baffle scatter effects", J. Acoust. Soc. Am. **122**(5), 2679–2685.

[31] Torres, A.M., Cobos, M., Pueo, B., and Lopez, J. (**2012**). "Robust acoustic source localization based on modal beamforming and time–frequency processing using circular microphone arrays", J. Acoust. Soc. Am. **132**(3), 1511–1520.

[32] Ahonen, J., del Galdo, G., Kuech, F., and Pulkki, V. (**2012**). "Directional analysis with microphone array mounted on rigid cylinder for directional audio coding", J. Audio Eng. Soc. **60**(5), 311–324.

[33] Koyama, S., Furuya, K., Hiwasaki, Y., Haneda, Y., and Suzuki, Y. (**2013**). "Sound field reproduction using multiple linear arrays based on wave field reconstruction filtering in helical wave spectrum domain", Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing **2013**(2842), 271–275.

[34] Koyama, S., Furuya, K., Hiwasaki, Y., Haneda, Y., and Suzuki, Y. (**2012**). "Wave Field Reconstruction Filter for Cylindrical Microphone and Loudspeaker Arrays", Proc. Autumn 2012 Meeting Acoust. Soc. of Japan 605–608.

[35] Yudin, V.A. (**1993**). "The minimum of potential energy of a system of point charges", Discrete Math. Appl. **3**(1), 75–81.

[36] Ahrens, J. and Spors, S. (**2012**). "Wave field synthesis of a sound field described by spherical harmonics expansion coefficients", J. Acoust. Soc. Am. **131**(3), 2190–2199.

[37] Golub, G.H. and Van Loan, C.F. (**1996**). "Matrix Computations", Baltimore, 3rd ed.

[38] Daniel, J. (**2000**). "Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context", Ph.D. Thesis, Université Paris.

[39] Zotter, F. (**2008**). "Sampling Strategies for Acoustic Holography/Holophony on the Sphere", Tech. Rep. Institute of Electronic Music and Acoustics University of Music and Performing Arts Graz.

[40] Fliege, J. and Maier, U. (**1999**). "The distribution of points on the sphere and corresponding curvature formulae", IMA J. Numer. Anal. **19** 317–334.

[41] Lebedev, V.I. (**1976**). "Quadratures on a Sphere", USSR Comp. Math. and Math. Phys. **16**(2) 10–24.

LIST OF PUBLICATIONS

Journal papers

1. J. Trevino, S. Koyama, S. Sakamoto and Y. Suzuki, "Mixed-order Ambisonics encoding of cylindrical microphone array signals," Acoust. Sci. and Tech.
(Chapters III and IV; accepted for publication)

2. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Sound field reproduction using Ambisonics and irregular loudspeaker arrays," IEICE Trans.
(Chapter V; in review)

3. J. Trevino, S. Koyama, S. Sakamoto and Y. Suzuki, "Encoding sound field recordings made with cylindrical microphone arrays for reproduction using loudspeaker arrays," J. Acoust. Soc. Am. (Chapter VI; in review)

International conference presentations

1. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Development of a higher order ambisonics encoder and decoder for a 157-channel surround speaker array," Proc. 5th Int. Symp. on Medical, Bio- and Nano-Electr., pp.137–138, Feb. 2010.

2. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "High order Ambisonic decoding method for irregular loudspeaker arrays," Proc. ICA 2010, Aug. 2010.

3. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Evaluation of a new ambisonic decoder for irregular loudspeaker arrays using interaural cues," Proc. 3nd Int. Symp. on Ambisonics and Spherical Acoust., June 2011.

4. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Three dimensional auditory display using Ambisonics with irregular loudspeaker arrays," Proc. The Joint Int. Conf. of the 5th Int. Symp. and the 4th Student-Organizing Int. Mini-Conf. on Inf. Electr. Syst., pp. 280–281, Feb. 2012.

5. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Ambisonic synthesis of directional sources using non-spherical loudspeaker arrays," Proc. AES 25th UK Conf. and 4th Int. Symp. on Ambisonics and Spherical Acoust., pp. 10.1–10.5, Mar. 2012.

Domestic conference presentations

1. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Decomposition of high order Ambisonic recordings for reproduction using irregular loudspeaker arrays," IEICE vol. 110, no. 71, EA2010-25, pp. 19–24, June. 2010.

2. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Comparison of Ambisonic decoding methods for an irregular, 157-channel loudspeaker array," ASJ 2010 Autumn meeting, pp. 753–756, Sept. 2010.

3. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Reproduction of sound fields using ambisonics and irregular loudspeaker arrays," 357th Acoustic Engineering meeting and 67th Ultrasonic Electronics meeting, Dec. 2010.

4. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Near-field corrections to reproduce Ambisonics using an irregular loudspeaker array," 359th Acoustic Engineering meeting and 69th Ultrasonic Electronics meeting, Dec. 2011.

5. J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Ambisonic decoder including near-field corrections for irregular loudspeaker arrays," ASJ 2012 Spring meeting, pp. 907–910, Mar. 2012.

6. J. Trevino, S. Koyama, S. Sakamoto and Y. Suzuki "Scalable encoding of sound field recordings made with cylindrical microphone arrays," ASJ 2013 Spring meeting, pp. 957–960, Mar. 2013.

7. J. Trevino, S. Koyama, S. Sakamoto and Y. Suzuki "3D cylindrical Ambisonics: Encoding sound field information using the cylindrical harmonic functions," ASJ 2013 Autumn meeting, Sept. 2013.