ABSTRACT

Title of dissertation: THE CAPTURE AND RECREATION OF 3D AUDITORY SCENES

Zhiyun Li, Doctor of Philosophy, 2005

Dissertation directed by: Professor Ramani Duraiswami
Department of Computer Science
Institute of Advanced Computer Studies

The main goal of this research is to develop the theory and implement practical tools (in both software and hardware) for the capture and recreation of 3D auditory scenes. Our research is expected to have applications in virtual reality, telepresence, film, music, video games, auditory user interfaces, and sound-based surveillance.

The first part of our research is concerned with sound capture via a spherical microphone array. The advantage of this array is that it can be steered into any 3D directions digitally with the same beampattern. We develop design methodologies to achieve flexible microphone layouts, optimal beampattern approximation and robustness constraint. We also design novel hemispherical and circular microphone array layouts for more spatially constrained auditory scenes.

Using the captured audio, we then propose a unified and simple approach for recreating them by exploring the reciprocity principle that is satisfied between the two processes. Our approach makes the system easy to build, and practical. Using this approach, we can capture the 3D sound field by a spherical microphone array and recreate it using a spherical loudspeaker array, and ensure that the recreated sound field matches the recorded field up to a high order of spherical harmonics. For some regular or semi-regular microphone layouts, we design an efficient parallel implementation of the multi-directional spherical beamformer by using the rotational symmetries of the beampattern and of the spherical microphone array. This can be implemented in either software or hardware and easily adapted for other regular or semi-regular layouts of microphones. In addition, we extend this approach for headphone-based system.

Design examples and simulation results are presented to verify our algorithms. Prototypes are built and tested in real-world auditory scenes.

THE CAPTURE AND RECREATION OF 3D AUDITORY SCENES

by

Zhiyun Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Commmittee:

Dr. Ramani Duraiswami, Chair/Advisor
Dr. Dennis M. Healy, Dean's Representative
Dr. Larry S. Davis
Dr. Amitabh Varshney
Dr. David Mount
Dr. Nail A. Gumerov

PREFACE

路漫漫其修远兮，吾将上下而求索。

—— 屈原《离骚》(B.C.313-312)

This famous ancient Chinese motto has been driving countless people to search after truths for more than two thousand years. While there is no exact English translation, different people have different understandings. For me, it is a perfect summary of my PhD study: *the road ahead is endless and unpredictable, yet up and down I'll unyieldingly explore with my ever increasing curiosity.*

When I was still in primary school, I was taught that scientific research is like mountain climbing. Well, I never climbed a mountain, so I just simply assumed that is because the mountain is high. Many years later, I gradually understand more. The height is actually not that important because in most research, we just don't know how high we can achieve or it is just an endless journey. Instead, what really challenges me is that I often find myself in the middle of a dark and dangerous path. I can't see where the light is to allocate my limited resources. Yet I can't look back either since I can't afford to give up. Even worse, I am not sure this is the right path. At the beginning, I repeatedly asked myself: should I change a path or move forward? Later I learned that it is more than a choice. In either case, I will install some hooks and ropes there in the hope that some time in the future, I may come across this point again and make ends meet, and hopefully become wiser for the next move.

In this sense, my whole research experience is a process of exploration and installing hooks and ropes, and this dissertation is such a converging point in my past long journey. In this dissertation, I am finally able to combine my knowledges, ideas, skills and tools from physics, mathematics, signal processing, mechanical engineering and computer science into one piece. Thank God!

The only pity in my PhD journey here is that I have just too many things to try, but my time is relatively too limited.

DEDICATION

谨以此文感谢我的父母及孪生兄弟

To my parents for their support and encouragement, also to my twin brother for our unique learning experience since the first minute in our lives.

ACKNOWLEDGMENTS

I owe my gratitude to all the people who have made this dissertation possible and because of whom my PhD journey has converged to this starting point for the rest of my life.

First and foremost I'd like to thank my advisor, Professor Ramani Duraiswami for giving me an invaluable opportunity to work on the spherical microphone array project. His clear foresight and deep insight have led this project on its right track. It has been a great pleasure to work with and learn from such an extraordinary individual.

I would also like to thank my previous advisor, Professor Amitabh Varshney. He enlightened me on how to do research, otherwise, I may still walk aimlessly in darkness. And I thank Dr. Nail A. Gumerov for his remarkable physics and mathematics skills. Thanks are due to Professor Larry S. Davis, Professor Dave Mount and Professor Dennis Healy for agreeing to serve on my dissertation committee and for sparing their invaluable time reviewing the manuscript.

My colleagues and friends deserve a special mention. They are Dmitry N. Zotkin, Elena Grassi, Kexue Liu, Zhihui Tang, Vikas C. Raykar, Changjiang Yang and Ryan Farell from Perceptual Interfaces and Reality Lab, and Xuejun Hao, Chang Ha Lee, Aravind Kalaiah, Thomas Baby from Graphics Lab, and other friends including Xiaoming Zhou, Xinhui Zhou, Zhenyu Zhang, Waiyian Chong and Xinhua

He from Computer Science Department or Electrical & Computer Engineering Department. Without their help in my research and life, I may have stopped anywhere in this long journey.

I owe my deepest thanks to my family - my parents and twin brother. They gave me the unique life experience.

I would like to acknowledge financial support from NSF Award 0205271.

TABLE OF CONTENTS

List of Figures xiii

1 Research Motivation and Overview 1

1.1 Auditory Interfaces 1

1.2 Capture of 3D Auditory Scenes 4

1.3 Recreation of 3D Auditory Scenes 6

2 Introduction to 3D Audio 8

2.1 3D Audio Capture 8

2.1.1 Directivity of Microphone 8

2.1.2 Spatial Recording with Multiple Channels 11

2.1.3 Spatial Filtering with Microphone Array 12

2.2 3D Audio Recreation 12

2.2.1 Recreation of 3D Audio over Headphones 12

2.2.2 Recreation of 3D Audio over Loudspeakers 15

2.3 3D Virtual Audio Modeling 16

2.3.1 Room Reverberation and Distance Cues 16

2.3.2 Air Absorption 18

2.3.3 Object Occlusion and Diffraction 18

3 Brief Tutorial on Theoretical Acoustics 19

3.1 Acoustic Wave Equation 19

3.2 Helmholtz Equation 20

3.3 Solution Using Green's Function 21

3.4 Solution Using Separation of Variables 23

3.4.1 Separation of Variables in Cylindrical Coordinate System . . . 23

3.4.2 Separation of Variables in Spherical Coordinate System 24

3.5 Special Functions 25

3.5.1 Legendre Polynomials 25

3.5.2 Spherical Harmonics 25

3.5.3 Bessel Functions 26

3.6 Wave Scattering from Rigid Surface 26

3.6.1 Scattering from a Rigid Cylinder 27

3.6.2 Scattering from a Rigid Sphere 28

4 Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming[1] 30

4.1 Introduction 31

4.2 Principle of Spherical Beamformer 32

4.3 Discrete Spherical Array Analysis 34

4.3.1 Previous Work 36

4.3.2 Orthonormality Error Noise Analysis 37

4.3.3 Design Examples 40

4.4 Simplified Optimization of Desired Beampattern for Discrete Array . 43

4.4.1 Discrete Spherical Beamformer as Finite Linear System 45

[1] This chapter is based on our original work in [63][61][60][64].

4.4.2 Maximum Beampattern Order for Robust Beamformer 47

4.4.3 Design Examples 49

4.5 Optimal Approximation Subject to the WNG Constraint 51

4.5.1 Constrained Optimization 52

4.5.2 Adaptive Implementation 53

4.5.3 Convergence and Optimal Step Size 57

4.5.4 Simulation Results 58

4.6 Conclusions 62

5 Hemispherical Microphone Arrays for Sound Capture and Beamforming[2] 63

5.1 Introduction 63

5.2 Design of A Hemispherical Microphone Array 65

5.2.1 Acoustic Image Principle 65

5.2.2 A Symmetric and Uniform Layout 66

5.2.3 Discrete Hemispherical Beamforming 67

5.3 Effective Calibration 68

5.4 Simulation and Experimental Results 71

5.4.1 Verification of Hemispherical Beamformer 71

5.4.2 Verification of Flexible and Optimal Spherical Array Design . 72

5.4.3 Test in Real-World Auditory Scenes 74

5.5 Precomputed Fast Spherical Beamforming 75

5.5.1 Time Domain Mixing via Orthonormal Decomposition 78

[2] This chapter is based on our original work in [59].

5.5.2 Verification 79

5.6 Conclusions 80

6 Cylindrical Microphone Arrays for Sound Capture and Beamforming[3] 81

6.1 Scattering from a rigid cylinder 82

6.2 Cylindrical Beamforming 83

6.3 Practical Design and Analysis 87

6.3.1 Discreteness 87

6.3.2 Practical Beampattern 88

6.3.3 Robustness 89

6.3.4 Beampattern in 3D Sound Field 91

7 Recreate 3D Auditory Scenes Via Reciprocity[4] 95

7.1 Loudspeaker Array-Based System 95

7.1.1 Recording as 3D Sound Field Projection 96

7.1.2 Recreation as 3D Sound Field Unprojection 98

7.1.3 Extension to Point-Source Form 101

7.1.4 Design Example and Simulations 103

7.1.5 Efficient Multi-directional Beamformer 106

7.2 Headphone-Based System 108

7.2.1 Ideal HRTF Selection 108

7.2.2 HRTF Approximation In Orthogonal Beam-Space 110

[3] This chapter is based on our original work in [62].

[4] This chapter is based on our original work in [65].

7.2.3 Recreation Algorithm 112

7.2.4 Design Examples 112

7.2.5 Summary 113

7.3 A Discrete Huygens Principle Interpretation 114

7.3.1 Solutions by Discrete Huygens Principle 115

7.3.2 Simulation Results 117

8 Recreate 2D Auditory Scenes[5] 120

8.1 Loudspeaker Array-Based System 120

8.1.1 Plane Wave Case 120

8.1.2 Spherical Wave Case 122

8.1.3 Simulation 126

8.2 Headphone-Based System 129

8.2.1 Ideal HRTF Selection 129

8.2.2 HRTF Approximation Via Orthonormal Decomposition 130

8.2.3 Reproduction from Recordings 132

8.2.4 Surround Audio Design Example 133

9 Experiment in Traffic Auditory Scenes[6] 135

10 Conclusions 142

11 On-going and Future Work 144

11.1 Multiple Array Data Fusion 144

[5] This chapter is based on our original work in [62].

[6] This chapter is based on our original work in [64].

11.2 3D Sound Texture Resynthesis 145

11.3 Complex Environment Analysis 145

11.4 3D Auditory Scene Editing 145

11.5 Real-Time Implementation 146

Bibliography 148

LIST OF FIGURES

2.1 The 3D supercardioid directivity pattern. 10

2.2 Microphone directivity patterns. The angle means that inside it the microphone picks up at least half (or within 3dB drop) of the peak value. 10

2.3 Measurement of HRTFs. 14

2.4 Recreation of virtual spatial sound using HRTFs. 14

2.5 Direct and crosstalk transmission paths from loudspeakers to the ears of a listener. 16

2.6 The real world 3D acoustic environment. 17

3.1 Plane wave incident on a rigid sphere. 29

4.1 $b_n(ka)$ for orders from 0 to 30. Given ka, $b_n(ka)$ decays very quickly with respect to n. 39

4.2 (a) Fliege's 64 nodes. (b) The orthonormality errors. 42

4.3 (a) Same as in Fig. 4.2(a) except the bottom four nodes are removed. (b) The orthonormality errors. 42

4.4 $(a) - (c)$ show the beampatterns from order 3 to 5 using the 60-node array in Fig. 4.3 with radius of 10cm. The top row uses Fliege's cubature coefficients. The bottom row uses the coefficients solved by (4.23). Simulated at 1KHz. 43

4.5 Using the 60-nodes array in Fig. 4.3, the horizontal lines show the mode bounds for robust beamforming at given ka. 50

4.6 The beampatterns of order 3 to 5 for the 60-nodes array in Fig. 4.3 using simplified optimization. 50

4.7 The random layout of 64 microphones on a sphere of radius 10cm. . . 51

4.8 Unconstrained beampatterns from order 3 to 5 for the array in Fig. 4.7. 51

4.9 Mode bounds for the array in Fig. 4.7. 52

4.10 The geometric interpretation. 56

4.11 Iteration process for beampattern of order 3. The red (thick) curves use the left scale, blue (thin) curves right scale. 59

4.12 Constrained optimal beampattern of order 3. 59

4.13 Precision comparisions for order 4 beamforming: (a) Comparison of unconstrained and constrained beampattern coefficients with regular order 4 beampattern coefficients $c_4\mathbf{B}_4$. (b) Residue comparison between unconstrained and constrained beampattern coefficients. Both plots show the absolute values. 60

4.14 Constrained optimal beampattern of order 4. 60

4.15 Constrained optimal beampattern of order 5. It is actually a superdirective beampattern. 61

4.16 The iteration process optimally approximates the DI of the regular beampattern of order 5. The red (thick) curves use the left scale, blue (thin) curves right scale. 61

4.17 The beampattern of the regular implementation of superdirective beamformer. 62

5.1 A hemispherical microphone array built on the surface of a half bowling ball. Its radius is 10.925cm. 65

5.2 The hemispherical array with a rigid plane is equivalent to a spherical array in free space with real and image sources. 66

5.3 The symmetric and uniform layout of 128 nodes on a spherical surface. The blue (dark) nodes are for real microphones. The yellow (light) nodes are images. 68

5.4 Discrete orthonormality errors. Plot shows absolute values. 69

5.5 Simulation and Experimental results: (a) simulation of 3D scanning result with two sound sources, the beamformer is of order 8; (b) experimental result using the calibrated beamfomer of order 8; (c) simulation result after sound sources are moved; (d) experimental result using the same calibrated beamformer. 72

5.6 (a) The 88-node layout is generated by removing 40 symmetric nodes in the 128-node layout in Fig. 5.3. (b) The order 7 beampattern using equal quadrature weights at 2.5KHz. 73

5.7 (a) The simulated source localization result using the array in Fig. 5.6. The beamformer at each direction is made with equal quadrature weights. (b) The experimental result using equal quadrature weights. Real sound source is from $(\pi/4, -\pi/2)$, at 2.5kHz. 74

5.8 The simulated optimal localization result without WNG constraint. . 75

5.9 Localization results using optimal beamformers under different WNG constraints. Top left: $0dB < WNG < 5dB$. Top right: $5dB < WNG < 10dB$. Bottom left: $10dB < WNG < 15dB$. Bottom right: $WNG > 20dB$. 76

5.10 Moving sound source tracking. Plots show the azimuth and elevation angles of the energy peak at each frame. The tracking is performed in the frequency band from 3kHz to 6kHz, with beamformers of order seven (3kHz˜4kHz) and eight (4kHz˜6kHz). 77

5.11 Sound separation using hemispherical beamformer. The signals in the right column correspond to segments in the original signals in the left column. 77

5.12 Gain comparision of direct beamforming and fast mixing in time domain. Plot shows the absolute value for each beamforming direction. Simulated in order eight with $\boldsymbol{\theta}_k = (\pi/4, \pi/4)$. 79

6.1 A plane wave incident on a rigid cylinder. 83

6.2 $A_n(ka)$ for n from 0 to 20. 86

6.3 $B_n(ka)$ for n from 0 to 20. 87

6.4 Beampatterns for $M = 1, 5, 10, 16$. 89

6.5 Spatial aliasing. 90

6.6 White noise gain. 91

6.7 Polar Beampatterns for different elevation angles at 5 kHz. The beampattern for $\theta = \pi/2$ is of order 13 with $WNG \approx -5.96$ dB. 92

6.8 3D beampattern of order 13 at 5 kHz, looking at $\varphi = 0$. (lighted to show 3D effect.) 92

7.1 Left: the microphone array captures a 3D sound field remotely. Right: the loudspeaker array recreates the captured sound field to achieve the same sound field. 96

7.2 System workflow: for each chosen direction $\boldsymbol{\theta}_i = (\theta_i, \varphi_i)$, we first beamform the 3D sound field into that direction using our N-order beamformer, then simply playback the resulted signal from the loudspeaker in that direction. 97

7.3 The layout of 32 microphones on a rigid sphere with radius of 5 cm. (The sphere is lighted to show the 3D effect, NOT the sound scattering.) 102

7.4 The layout of 32 loudspeakers in free space arranged as a sphere of large radius. 103

7.5 The plane wave of 4 kHz incident from positive direction of X-axis scattered by the microphone array. Plot shows the pressure on the equator. 104

7.6 The recreated plane wave to order 4. Plot shows the $2 \times 2\,\mathrm{m}^2$ area on the $z = 0$ plane. 105

7.7 The rotational symmetry of icosahedron. 107

7.8 The efficient structure of the multi-directional spherical beamformer. 108

7.9 HRTF approximations to orders 5 and 10. Plot shows the magnitude in dB scale. 113

7.10 Phases of the approximations to orders 5 and 10. 114

7.11 1KHz sound source at 2a, recorded using a=10cm spherical array with 64 Fliege nodes. Plot shows the recreated sound field using 64 virtual plane waves. 117

7.12 Same setup, but using only 60 microphones and 60 virtual plane waves.118

7.13 recorded using spherical array with 324 Fliege's nodes with a=8cm. 8KHz source from 2a. Plot shows the recreated sound field using 324 virtual plane waves. 118

7.14 Recreated plane wave of 1KHz, using 64 nodes mic array in free space with $a = 10cm$. Recreated by 64 virtual plane waves. 119

7.15 Free space, 8KHz plane wave, recorded using free space 32-nodes mic array with radius $a = 3cm$. Recreated by 32 virtual plane waves. . . . 119

8.1 H_5, H_{10} and H_{16} at frequency 2 kHz. 124

8.2 H_5, H_{10} and H_{16} at frequency 5 kHz. 126

8.3 Reproduce the 5 kHz plane wave to order 5 using plane waves. 126

8.4 Reproduce the 5 kHz plane wave to order 10 using plane waves. . . . 127

8.5 Reproduce the 5 kHz plane wave to order 16 using plane waves. . . . 127

8.6 Reproduce the 5 kHz plane wave to order 5 using spherical waves. . . 127

8.7 Reproduce the 5 kHz plane wave to order 10 using spherical waves. . . 128

8.8 Reproduce the 5 kHz plane wave to order 16 using spherical waves. . . 128

8.9 Reproduce the 2 kHz plane wave to order 5 using spherical waves. . . 128

8.10 HRTFs in spatial frequency domain. 134

9.1 The snapshot of our spherical microphone array system. 136

9.2 The scenario of the experiment. A car is moving from left to right. The spherical microphone array is placed on the roadside. 136

9.3 Tracking one vehicle using the spherical microphone array. 138

9.4 3D localization of the first chosen frame. The car is now on the left side of the array. 139

9.5 The car is now roughly in front of the array. 140

9.6 The car now moves to the right side of the array. 140

9.7 Tracking two vehicles using the spherical microphone array. 141

Chapter 1
Research Motivation and Overview

Our research is originally motivated by the goal to create *customizable audio user interfaces* (CAUI) for the visually impaired and the sighted although our contributions will enrich the 3D audio technologies in general. There are approximately 4 million blind or visually impaired people in the US. CAUI aims to help these people interact with surrounding environment [37][68][96][76].

Our research motivation is best illustrated by a hypothetical application: *creating a CAUI system to train a blind or visually impaired to cross a street or even drive a vehicle safely.* To implement this system, therefore, our research is focused on the capture and recreation of 3D auditory scenes.

In this chapter, we first briefly review perceptual interfaces, especially auditory interfaces, then we overview our capture and recreation technologies, respectively.

1.1 Auditory Interfaces

Since the human sight accounts for a large portion of all perceptual information, most *human-computer interfaces* (HCI) and *virtual reality* (VR) systems are based on graphics technologies. To extend that, perceptual interfaces aim to achieve effective human-computer interactions using full human perceptions: visual, tactile and auditory, etc. In [13], the concept of voice and gesture commanding was introduced to accommodate more natural human-machine interaction. Some examples include two-handed input [20], direct manipulation and natural language [22], image manipulation using speech and gestures [49], facial expression [87], using ultrasonic pointer and displaywall [67][66].

We are interested in auditory interfaces in our research. Auditory information plays an important role in human's interaction with the world, especially for the visually impaired people. With the advance in 3D sound recording and playback technologies, it is possible to build a sound-based VR system [97]. In this section, we briefly review the advantages, techniques and challenges of auditory interfaces.

The advantages of auditory interfaces haven't been fully exploited. In summary, there are three major occasions where auditory interfaces may be desirable:

1. when the information can be conveyed both auditorily and visually, or only auditorily;
2. when the users want to perform more than one task at a time;
3. when visual displays are not easily available, for safety reasons or for visually impaired individuals.

The auditory interfaces can be implemented by two general approaches: speech based and non-speech based. The speech based techniques are relatively well studied [15]. However, they also have serious limits. Because speech interfaces are basically to translate information into text and then pronounce it, they are usually slow and even impossible sometimes. The non-speech based techniques partially solved those difficulties. They include three main techniques: *auditory icons*, *earcons* and *sonification*. Each technique has been successfully applied to certain types of applications.

The first technique is auditory icons, which are analogous to graphical icons [75][44]. The auditory icons use everyday sounds, such as water pouring and door opening, to represent the operations that users would do with graphic icons. Auditory icons are tightly limited in both the sound dimension and the mapping dimension. The sounds have to exist in everyday world and the mapping has to convey the meaning of everyday events. The expected advantage of using auditory icons is a intuitive mapping between the sounds and their meaning. This advantage, however, is also its weakness because of the difficulty in more complicated interfaces.

In contrast with auditory icons, earcons take advantage of several dimensions of sound, such as pitch, timbre, rhythm and spatial location [10]. Typically, earcons are short musical sounds. More information such as computer object, operation, or interaction can be efficiently communicated to the user. In addition, earcons can be combined to represent composite meanings. The effectiveness of an earcon-based auditory interface largely depends on how well the sounds are designed. However, the mapping is implicit and much less directly connected with human perceptual capabilities than that of auditory icons.

Since sound is multidimensional, it has the potential to convey multidimensional data efficiently. Therefore, the task of sonification is the translation from data dimensions to sound dimensions. Sonification has a wide range of applications such as the visualization of DNA sequences [12], aid in economic forecasting [72] and visualization of geographical data [96][95].

The auditory interface research is still in its early stages. One major challenge is to fully understand and exploit the multidimensional properties of sound such as pitch, timbre, loudness, duration and spatial cues. In our research, we are mostly interested in spatial cues. Spatial cues have been attracting a great deal of research effort in psychoacoustics to understand [92][79][5][19][18][17][57][80] and simulate them [73][8][70][25]. To capture and recreate spatial sound, we develop the theory and implement practical tools.

1.2 Capture of 3D Auditory Scenes

This is apparently the first step of the implementation if we desire the training to be as realistic as possible. We wish the captured auditory scene has high resolutions to provide reliable training materials. That means we must use a microphone array since an individual microphone cannot capture the sound's spatial information.

There are various designs for microphone arrays to capture and analyze sound fields. Geometrically, most such arrays fall into one of three designs: linear, circular and spherical. Perhaps the most straightforward design is the linear microphone array. In [85], general linear sensor arrays for beamforming are elaborated. A more symmetric and compact configuration is the circular microphone array such as the one for speech acquisition described in [58]. To capture the 3D sound field, we prefer a 3D symmetric configuration: the spherical microphone array. Spherical microphone arrays are recently becoming the subject of some study as they allow omnidirectional sampling of the sound-field, and may have applications in soundfield capture. In [3], the sound field was captured using a spherical microphone array in free space. Microphones can also be positioned on the surface of a rigid sphere to make use of the scattering effect. The paper [71] presented a preliminary analysis of such arrays, and showed how sound can be analyzed using them. This paper performed an elegant separation of the analysis and beamforming parts by using a modal beamformer structure. The beamformer using such configuration has the same shape of beampattern in all directions.

In practice, however, we have to consider the cable outlet and the mounting base. In Chapter 4, we propose an extension of this approach that allows flexible microphone placements with minimal performance compromise.

If the sound sources are bounded within 3D halfspace, we can design a hemispherical microphone array using the acoustic image principle. This will be detailed in Chapter 5. In this chapter, we also develop a precomputed fast beamforming algorithm.

In our hypothetical application, the traffic auditory scene is largely two dimensional or "surround", with all sound sources roughly moving on a plane. Other such examples include cocktail party, roundtable conference, surround music, etc. So to build a full spherical microphone array seems redundant. Our third problem is to design and build a microphone array for capturing and spatial filtering of higher order surround sound field.

The circular microphone arrays should be the obvious choice in this case. In previous work, the circular microphone array is assumed to be placed in free space [58]. Although that seems simpler to model, it has some drawbacks in practice, especially the loss of spatial information with the presence of noise. To overcome this drawback, we parallel the design of our spherical microphone array and propose a circular microphone array mounted on the surface of a rigid cylinder. We use this array to record the scattered surround sound field. The spatial filtering, or beamforming, is based on the orthonormal decomposition of the recorded sound field by this array. Detailed description is in Chapter 6.

1.3 Recreation of 3D Auditory Scenes

With the captured auditory scene, the next step is to recreate it as exactly to the original scene as possible so that the trainee believes she is physically there. There are two general ways: recreation over headphone or loudspeakers. Therefore, our next problem to develop the theory of recreation from captured scenes.

People have proposed several schemes to build the microphone-loudspeaker array system. In [9], based on the Kirchhoff-Helmholtz integral on a plane, the sound field is captured by a directive microphone array, and recreated by a loudspeaker array. That is called the *Wave Field Synthesis* (WFS) method. While this system works well in an auditorium environment where the listening area can be separated from the primary source area by a plane, it is hard to render an immersive perception in a 3D sound field. Further, a quantification of the approximations made is not presented. In [82], a general framework was proposed, which uses a microphone array beamformer with a localization and tracking system to identify the sound sources, then uses the loudspeaker array to recreate them with the correct spatial cues. To work properly, however, it requires a robust and accurate localization and tracking system and a highly directive beamformer which are usually expensive, if available.

Loudspeaker arrays have similar configurations as microphone arrays to recreate the sound field. In [27][9], a linear loudspeaker array is designed recreate the sound field on a 2D plane. In [88], a theoretical analysis for using a spherical loudspeaker array to recreate 3D sound field was provided.

In Chapter 7, we apply this approach to analyze and design our microphone-loudspeaker integrated system. We explore the reciprocity between capturing and recreating sound and used it to propose a simple and unified way to capture and recreate a 3D sound field to higher orders of spherical harmonics. Besides using loudspeaker arrays which are usually expensive, we also can build a personalized 3D audio system over headphones with HRTFs. We present this approach in Chapter 7. In addition, we present an alternative interpretation using the discrete Huygens principle.

In addition, we develop the theory for recreating the recorded 2D auditory scenes, by both headphone and loudspeaker array. This will be in Chapter 8.

The effectiveness of our algorithms is verified by various design examples and simulations throughout this dissertation. Our algorithms are further demonstrated by experimental results using our prototypes as shown in Fig. 5.1 and Fig. 7.39.

Chapter 2
Introduction to 3D Audio

The goal of 3D audio is to capture and recreate spatial sound field accurately so that the listener feels she is actually there. It also includes creating virtual 3D auditory scenes. This chapter gives a tutorial on related 3D audio technologies for microphones, headphones and loudspeakers, and also discusses the modelling of 3D virtual audio.

2.1 3D Audio Capture

The device to record sound signal is microphone. In this section, we first present a brief introduction of microphones with emphasis on the 3D spatial recording properties. Then we review some popular spatial recording technologies using microphones as individual components. To achieve more flexible spatial recording and filtering, we need the microphone array. Several geometric designs for different applications will be introduced here.

2.1.1 Directivity of Microphone

The basic principle of microphone is to convert variations in air pressure into equivalent electrical variations, in current or voltage. While there are many ways to convert sound into electrical energy, the two most popular methods are *dynamic* and *condenser*. For more details, please refer to [32]. In our research, we are more concerned about the *directivity* (or *polar*) *patterns* (in dB scale) of a microphone, that is, how well it picks up sound from various directions. Most microphones can be placed in one of two main groups: *omnidirectional* and *directional.*

Omnidirectional Microphones

Omnidirectional microphones are supposed to pick up sound from every 3D directions equally. The output should only depend on the relative distance between the sound source and the microphone. However, even the best omnidirectional microphones tend to have distorted directivity patterns, especially at higher frequencies. This is caused by many factors. One of them is the physical geometry of the omnidirectional microphone. When the wave length of sound is comparable or smaller than the microphone size, the scattering and shadowing effects will distort the directivity pattern.

Directional Microphones

Directional microphones are specially designed to selectively pick up the sound from one direction (usually from the front) and suppress the sound incident from all other directions. Again, the real-world directivity patterns depend on frequencies. This directional ability is usually achieved with external openings and internal passages in the microphone to allow sound to reach both sides of the diaphragm corresponding to the desired directivity pattern. The basic directivity patterns include cardioid, supercardioid, hypercardioid and bidirectional. A 3D plot of supercardioid pattern is shown in Fig. 2.1.

A comparison is shown in 2D in Fig. 2.2.

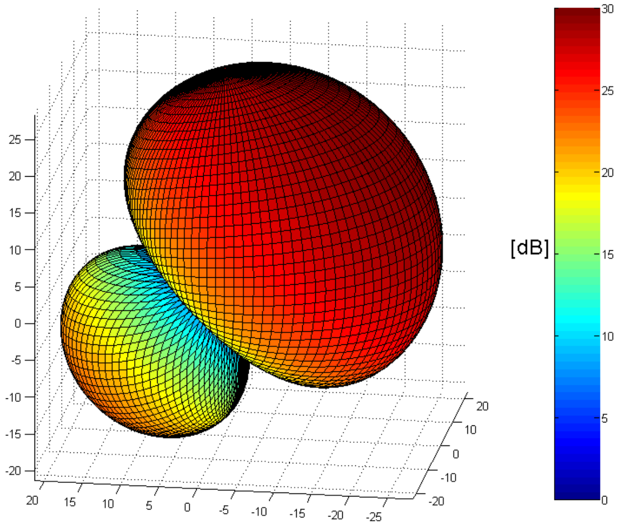


Figure 2.1: The 3D supercardioid directivity pattern.

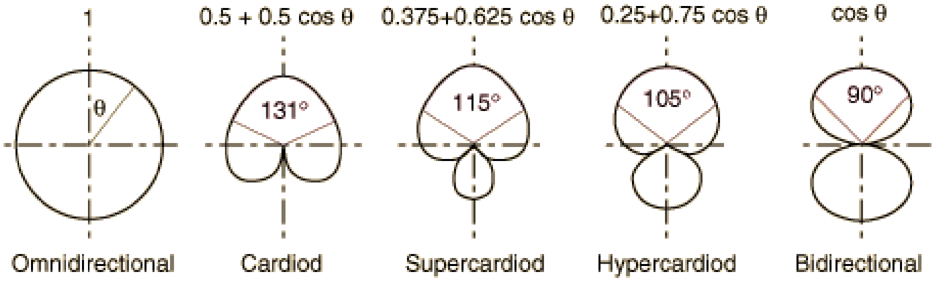


Figure 2.2: Microphone directivity patterns. The angle means that inside it the microphone picks up at least half (or within 3dB drop) of the peak value.

2.1.2 Spatial Recording with Multiple Channels

Probably the simplest recording is monaural recording. It uses a single microphone so the spatial cues are missed. To record at least some spatial cues, however, multiple recording channels are used.

In two-channel recording, or *stereo*, there are two widely used methods: *coincident* and *spaced* microphones. In coincident microphone recording, the spatial sound is recorded by two directional microphones pointing in different directions but as close to each other as possible. Therefore, the time difference is expected to be eliminated while capturing the intensity differences. In spaced microphone recording, two identical microphones are spaced some distance apart to capture the time difference. These two stereo recording techniques can be combined together to capture both the intensity and time differences.

To capture the same complete spatial cues as a pair of human ears do, a natural method is to place two small microphones in each ear canal of a dummy head, or even a person, to record an auditory scene. This is called the *binaural* recording. To playback, it simply feeds each channel to each ear through headphones.

The surround sound systems extended this goal further using more than two channels. In *quadraphonic* system, four microphones are placed at the four corners of a room, with each channel feeding a loudspeaker. However, this technology only lasted for a short time, technically because it was inaccurate and non-realistic in presenting a 3D sound source. To overcome those difficulties, a more advanced system, the *Ambisonics*, was developed to provide a relatively high resolution surround sound. we show this is only a special case of our approach in later chapters.

2.1.3 Spatial Filtering with Microphone Array

In multiple channel recording, although various coding schemes are designed to accommodate efficient data transmission, they are transparent to the input and the corresponding output channels. In other words, every channel actually works independently. To make multiple channels collaborate more efficiently, microphone arrays arranged in different geometries are designed to perform specific spatial sampling and filtering. For example, microphones arranged along a line samples one dimensional space, circular microphone array samples two dimensional space, and a spherical microphone array is for three dimensional space sampling which is appropriate for our research. More details on spherical microphone arrays will be presented in later chapters.

2.2 3D Audio Recreation

There are two general ways to recreate 3D audio: by headphone and by loudspeakers.

2.2.1 Recreation of 3D Audio over Headphones

To recreate 3D audio over headphones, we need to understand how humans can localize sounds using only two ears. Suppose there is a sound source in 3D space generating a sound wave that propagates to the ears of the listener. When the sound source is to the left of the listener, the sound reaches the left ear before the right ear which causes an *interaural time difference* (ITD). In addition, because of the acoustic scattering by the listener's torso, head and pinna, the intensity received at the left ear is different from that at the right ear, which is called the *interaural intensity difference* (IID). Therefore, signal received by each ear is the result of a different and complicated filtering process depending on the direction of the incident sound. The human auditory system then compares these two channel signals, extracts the spatial cues and derives the spatial location of the sound source. The exact mechanism is still not completely known. More detailed information about sound localization by human listeners can be obtained in [46][11].

A source-filter model is used to describe this process. Specifically, the transformation from the sound signal generated by the source to the signal received by an ear is a filter called the *head-related transfer function* (HRTF). HRTFs are usually measured by inserting miniature microphones into the ear canals of a human subject or a manikin. A measurement signal is played by a loudspeaker from many different directions and recorded by the microphones. This process is shown in Fig. 2.3. By reciprocity, the locations of microphone and speech can be exchanged without effecting the results. Using a microphone array to record multiple HRTFs at a time, this approach makes the tedious measurement process significantly faster [98].

Since the spatial cues have been captured by the measurement for the listener, the recreation of 3D audio is to reproduce the same spatial cues at the ears of the listener using a pair of measured HRTFs. When a sound is filtered by the HRTFs and fed into the headphones, the listener will perceive the sound is from the location specified by the HRTFs. This process is called *binaural synthesis* (binaural signals are defined as the signals at the ears of a listener) as shown in Fig. 2.4.

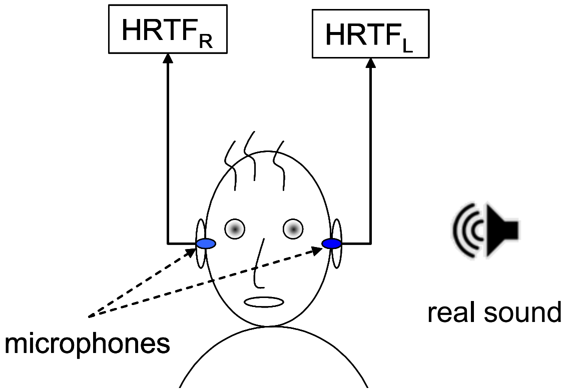


Figure 2.3: Measurement of HRTFs.

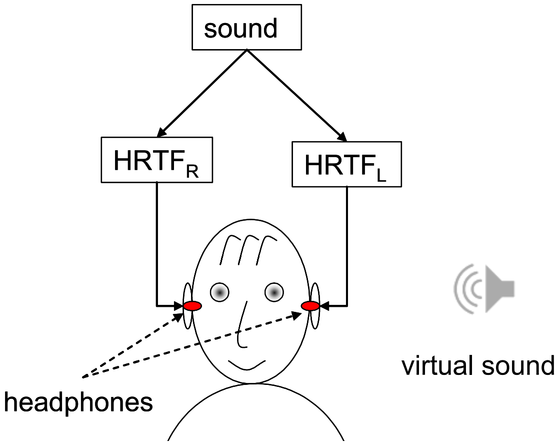


Figure 2.4: Recreation of virtual spatial sound using HRTFs.

Binaural synthesis works extremely well when the listener's own HRTFs are used to synthesize the spatial cues [92][93]. In practice, since HRTFs are not easily available for every listeners, 3D audio systems typically use a single set of HRTFs previously measured from a particular human or manikin subject. Some HRTF databases are KEMAR [2][43] and CIPIC [6]. However, HRTFs depend on the geometries of the subject's torso, head and especially pinna, this makes HRTFs highly individual. Therefore, spatial cues are not always correctly recreated for a listener if using HRTFs measured from a different individual. Common errors are front/back confusions and elevation errors [91]. This is the major limit of headphone-based 3D audio systems.

2.2.2 Recreation of 3D Audio over Loudspeakers

The simplest loudspeaker-based spatial audio system is to use two loudspeakers. In contrast with headphone-based systems, sound from the two loudspeakers are not separately perceived by the listener's ears. Instead, both ears receive the signals from both loudspeakers in direct and crosstalk paths as shown in Fig. 2.5.

To correctly recreate spatial cues, the crosstalk must be eliminated by a carefully designed digital filter, often called a crosstalk canceller [41][40]. This filter cancels the crosstalk at a pre-specified location for the listener, or called *sweet spot*, thus separates the two channels. However, when listening, the listener must be facing forward in the sweet spot to get the best spatial cues. An extension is the surround sound system which uses more than two channels. As with binaural synthesis, accurate crosstalk canceller is also very individual.

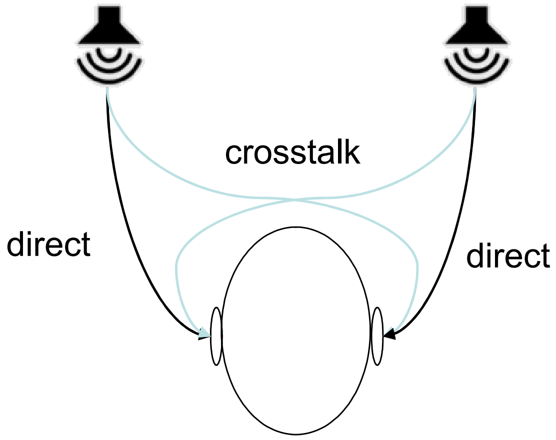


Figure 2.5: Direct and crosstalk transmission paths from loudspeakers to the ears of a listener.

To solve this problem, one approach is to recreate the 3D sound field around the listener using loudspeaker arrays as described in section 1.3.

2.3 3D Virtual Audio Modeling

In addition to recreate 3D audio from real recordings, it can also be simulated virtually. 3D virtual audio aims to create complete 3D auditory scenes, by modelling 3D acoustic environments in real world such as room reverberation and distance cues, air absorption, object occlusion and diffraction, etc. as shown in Fig. 2.6. One such system is described in [97].

2.3.1 Room Reverberation and Distance Cues

Room reverberation is caused by the reflection of sound waves. The listener will hear the sound wave from the source via a direct path if it exists, followed by reflections off nearby surfaces, or called early reflections, then late reflections in all directions, or called diffuse reverberation. The reverberation time is defined as the

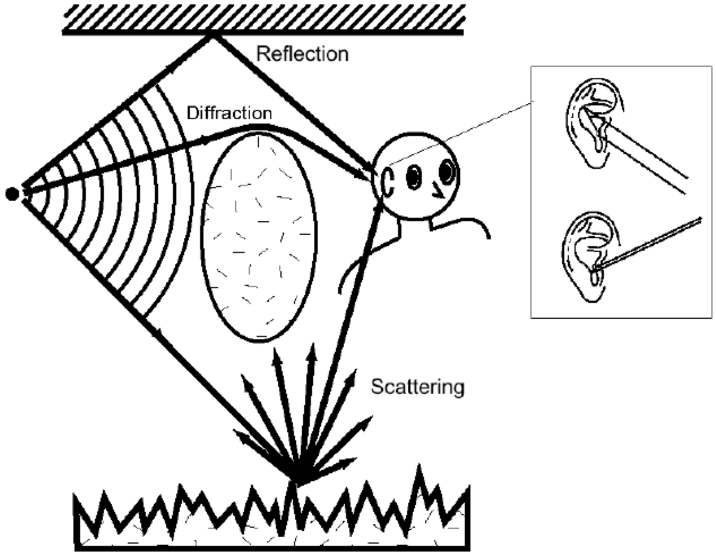


Figure 2.6: The real world 3D acoustic environment.

duration between the initial level and 60 dB decay.

Apparently, the loudness of the sound provides the main distance cue if the listener knows the normal loudness of that sound. Another important distance cue is the relative loudness of reverberation since it provides the relative spatial information about the source location and the environment. A simple example is that a distant source sounds more reverberant than close ones.

In practice, the geometry of acoustic environment is modelled as a simple room with reflective surfaces, then the early reflection can be easily computed, such as by beam tracing [39]. The computation of late reverberation is expensive because of the large number of reflections in all directions. Some alternative methods include recursive filters [42], feedback delay networks [54], and *fast multipole method* (FMM) [30], etc.

2.3.2 Air Absorption

Air absorption also provides spatial cues of the acoustic environment. When sound wave propagates through air, part of its energy is absorbed along the path. Air absorption depends on sound frequency and atmospheric conditions such as temperature and humidity etc. In general the more distant the sound source is, the more energy will be absorbed. In addition, low frequency sound propagates further than high frequency sound. The resulting effect is actually a lowpass filter.

2.3.3 Object Occlusion and Diffraction

As shown in Fig. 2.6, the sound path may be bent around occluding objects. This can be modelled by acoustic scattering theory. If the size of the occluding object is comparable to the wavelength, scattering is strong; if much smaller, the object is approximately transparent to the wave; if much bigger, then the wave will be reflected. Again, the occluding effect can be modelled as a lowpass filter. However, the computation is expensive and the cost increases dramatically with more occluding objects. For simple geometric objects like spheres, the scattering can be computed efficiently using FMM [47].

Chapter 3
Brief Tutorial on Theoretical Acoustics

In this chapter, we give a brief tutorial on theoretical acoustics. We emphasize on the basic concepts and results instead of strict derivations. This will provide a self-contained background for the works in later chapters. More details can be easily found in any acoustics textbooks, e.g. [74].

3.1 Acoustic Wave Equation

We introduce the wave motion in air, the most important type of wave motion in acoustics. We first study the motion of *plane waves* of sound. Plane waves have the same direction of propagation everywhere in space whose "crests" are in planes perpendicular to the direction of propagation. The detailed properties of the acoustic wave motion in a fluid depend on the ratios between the amplitude and frequency of the acoustic motion and the molecular mean-free-path and the collision frequency, on whether the fluid is in thermodynamic equilibrium or not, and on the shape and thermal properties of the boundaries enclosing the fluid. In our work, we use a simplified, yet practical model of acoustic wave motion in an ideal fluid which is uniform and continuous, at rest in thermodynamic equilibrium and without nonlinear effects.

The two basic observations of acoustic waves are:

1. a pressure gradient produces an acceleration of the fluid;

2. a velocity gradient produces a compression of the fluid.

In one-dimensional case, they are formulated as:

$$\rho\frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x}, \tag{3.1}$$

$$\kappa\frac{\partial p}{\partial t} = -\frac{\partial u}{\partial x}, \tag{3.2}$$

where u is the velocity flow of fluid in the x-direction, p is the pressure in the fluid, ρ is the mass per unit volume of the fluid, and κ is the *compressibility* of the fluid. Eliminating u, we have:

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \qquad c^2 = \frac{1}{\kappa\rho}, \tag{3.3}$$

which is the one-dimensional *wave equation.* It is easy to extend this to three-dimensional case:

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}, \tag{3.4}$$

where ∇^2 is the *Laplacian operator* defined as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{3.5}$$

It is usually to express p as the *velocity potential* ϕ, so the wave equation becomes:

$$\frac{\partial^2 \phi}{\partial t^2} = c^2\nabla^2\phi, \tag{3.6}$$

where c is the acoustic wave speed in air, it is approximately $343\,\mathrm{m/\,s}$.

3.2 Helmholtz Equation

If we assume the solutions of wave equation are time harmonic standing waves of frequency ω

$$\phi(\mathbf{r}, t) = e^{-i\omega t}\psi(\mathbf{r}), \tag{3.7}$$

substitute (3.7) into (3.6), we find $\psi(\mathbf{r})$ satisfies the *homogeneous Helmholtz equation*:

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0, \tag{3.8}$$

where $k = \omega/c$ is the wave number. The solutions of the Helmholtz equation are the solutions of the wave equation in frequency domain.

If there is a harmonic wave source $f(\mathbf{r})e^{-i\omega t}$ then we have the *inhomogeneous Helmholtz equation*:

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = f(\mathbf{r}). \tag{3.9}$$

3.3 Solution Using Green's Function

Instead of solving $\psi(\mathbf{r})$ in (3.9) directly, we first introduce the Green's function $G(\mathbf{r}_1, \mathbf{r}_2)$ satisfying:

$$(\nabla^2 + k^2)G(\mathbf{r}_1, \mathbf{r}_2) = \delta^3(\mathbf{r}_1 - \mathbf{r}_2), \tag{3.10}$$

then we can construct the solution for (3.9):

$$\psi(\mathbf{r}) = \int G(\mathbf{r}_1, \mathbf{r}_2) f(\mathbf{r}_2) d\mathbf{r}_2. \tag{3.11}$$

So our problem becomes to solve $G(\mathbf{r}_1, \mathbf{r}_2)$ in (3.10).

Because of the spherical symmetry of this problem, the solution is only dependent on $r = |\mathbf{r}_1 - \mathbf{r}_2|$. Using:

$$\nabla^2 G = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial G}{\partial r}\right) = \frac{1}{r}\frac{\partial^2}{\partial r^2}(rG), \tag{3.12}$$

the problem is then:

$$\frac{1}{r}\left(\frac{\partial^2}{\partial r^2}(rG) + k^2(rG)\right) = \frac{\delta(r)}{4\pi r^2}. \tag{3.13}$$

So for $r > 0$, we have

$$\frac{\partial^2}{\partial r^2}(rG) + k^2(rG) = 0, \tag{3.14}$$

which has a well-known general form solution:

$$G = \frac{A}{4\pi r}e^{ikr} + \frac{B}{4\pi r}e^{-ikr}. \tag{3.15}$$

Under the radiation condition, the waves move away from the source. So we must take $B = 0$.

$$G = \frac{A}{4\pi r}e^{ikr}, \qquad r > 0. \tag{3.16}$$

We extend this to all values of r by defining G to be the generalized function

$$G = \lim_{\epsilon \to 0}\left(\frac{AH(r-\epsilon)}{4\pi r}e^{ikr}\right), \tag{3.17}$$

where

$$H(x) = \begin{cases} 1 & x > 0 \\ \frac{1}{2} & x = 0 \\ 0 & x < 0 \end{cases} \tag{3.18}$$

is the Heaviside step function. Then we have:

$$\nabla^2 G = \frac{-Ak^2 e^{ikr}}{4\pi r} - A\delta(\mathbf{r}). \tag{3.19}$$

So

$$(\nabla^2 + k^2)G = -A\delta(\mathbf{r}). \tag{3.20}$$

Hence, we take $A = -1$. The solution for G is:

$$G(r) = -\frac{1}{4\pi r}e^{ikr} = -\frac{1}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|}e^{ik|\mathbf{r}_1 - \mathbf{r}_2|}. \tag{3.21}$$

Therefore, the solution of the inhomogeneous Helmholtz equation (3.9) satisfying the outward radiation condition is:

$$\psi(\mathbf{r}_1) = -\frac{1}{4\pi}\int \frac{f(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} e^{ik|\mathbf{r}_1-\mathbf{r}_2|} d\mathbf{r}_2. \tag{3.22}$$

3.4 Solution Using Separation of Variables

While the solution using Green's function has a compact form of integral, sometimes, we desire a series solution so that each component can be analyzed individually.

3.4.1 Separation of Variables in Cylindrical Coordinate System

In cylindrical coordinate system, the Helmholtz equation (3.8) becomes:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \psi}{\partial \varphi^2} + \frac{\partial^2 \psi}{\partial z^2} + k^2\psi = 0. \tag{3.23}$$

Let

$$\psi(\mathbf{r}) = R(r)\Phi(\varphi)Z(z), \tag{3.24}$$

(3.23) then becomes three ordinary differential equations:

$$Z'' + \mu Z = 0, \tag{3.25}$$

$$\Phi'' + n^2\Phi = 0, \tag{3.26}$$

$$r^2R'' + rR' + (k^2r^2 - n^2)R = 0, \tag{3.27}$$

where μ, n^2, k^2 are constants decided by boundary conditions. (3.25) and (3.26) can be solved easily. To solve (3.27), let $x = kr$ and $y(x) = R(r)$, (3.27) then becomes:

$$x^2y'' + xy' + (x^2 - n^2)y = 0, \tag{3.28}$$

which is the nth-order Bessel's equation. The solution is a series of Bessel functions.

3.4.2 Separation of Variables in Spherical Coordinate System

Similarly, in spherical coordinate system, the Helmholtz equation (3.8) becomes:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\varphi^2}+\lambda u=0. \tag{3.29}$$

Let

$$u(r,\theta,\varphi)=R(r)\Theta(\theta)\Phi(\varphi), \tag{3.30}$$

we have:

$$\Phi''+m^2\Phi=0, \quad m=0,1,2,... \tag{3.31}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\left[l(l+1)-\frac{m^2}{\sin^2\theta}\right]\Theta=0, \tag{3.32}$$

$$r^2\frac{d^2R}{dr^2}+2r\frac{dR}{dr}+[k^2r^2-l(l+1)]R=0. \tag{3.33}$$

Let $x=\cos\theta$ and $y(x)=\Theta(\theta)$, then (3.32) becomes:

$$(1-x^2)y''-2xy'+\left[l(l+1)-\frac{m^2}{1-x^2}\right]y=0, \tag{3.34}$$

which is the associated Legendre differential equation. The solutions are associated Legendre polynomials. The combined solution of Θ and Φ is the spherical harmonics.

For (3.33), let $x=kr$ and $\frac{1}{\sqrt{x}}y(x)=R(r)$, it becomes:

$$x^2y''+xy'+\left[x^2-\left(l+\frac{1}{2}\right)^2\right]y=0, \tag{3.35}$$

which is the spherical Bessel equation. The solutions are the spherical Bessel functions.

3.5 Special Functions

The following are the special functions we mentioned in the last section. These functions will be used to generate the solutions to the problems of wave scattering, from a rigid sphere and a rigid cylinder, which are the cases for our research.

3.5.1 Legendre Polynomials

The Legendre polynomial of order l is defined as:

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l, \tag{3.36}$$

or as an expansion:

$$P_l = \frac{1}{2^l}\sum_{k=0}^{\lfloor l/2 \rfloor}\frac{(-1)^k(2l-2k)!}{k!(l-k)!(l-2k)!}x^{l-2k}, \tag{3.37}$$

where $\lfloor l/2 \rfloor$ is the floor of $l/2$.

The associated Legendre polynomials are the solutions to the associated Legendre differential equation. They can be derived from Legendre polynomials:

$$\begin{aligned} P_l^m(x) &= (-1)^m(1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x) \\ &= \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l. \end{aligned} \tag{3.38}$$

3.5.2 Spherical Harmonics

The spherical harmonics are defined as:

$$Y_l^m(\theta,\varphi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\varphi}, \tag{3.39}$$

and they are orthonormal to each other on the spherical surface Ω:

$$\int_\Omega Y_n^{m*}(\theta,\varphi)Y_{n'}^{m'}(\theta,\varphi)d\Omega = \delta_{nn'}\delta_{mm'}. \tag{3.40}$$

3.5.3 Bessel Functions

The Bessel function of the first kind $J_v(x)$ can be expressed as a series of gamma functions:

$$J_v(x) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{v+2k}}{k!\Gamma(v+k+1)}. \tag{3.41}$$

The Bessel function of the second kind $N_v(x)$ (or sometimes called $Y_v(x)$) is defined as:

$$N_v(x) = \frac{J_v(x)\cos v\pi - J_{-v}(x)}{\sin v\pi}, \quad v \notin \mathbb{Z} \tag{3.42}$$

$$N_n(x) = \lim_{v \to n} N_v(x), \quad n = 0, 1, 2, ... \tag{3.43}$$

The Bessel functions of the third kind, or Hankel functions of the first and second kinds, are defined as:

$$H_n^{(1)}(x) = J_n(x) + iN_n(x), \tag{3.44}$$

$$H_n^{(2)}(x) = J_n(x) - iN_n(x), \tag{3.45}$$

The spherical Bessel functions of all three kinds are:

$$j_l(x) = \sqrt{\frac{\pi}{2x}} J_{l+\frac{1}{2}}(x), \tag{3.46}$$

$$n_l(x) = \sqrt{\frac{\pi}{2x}} N_{l+\frac{1}{2}}(x), \tag{3.47}$$

$$h_l^{(1)}(x) = \sqrt{\frac{\pi}{2x}} H_{l+\frac{1}{2}}^{(1)}(x), \tag{3.48}$$

$$h_l^{(2)}(x) = \sqrt{\frac{\pi}{2x}} H_{l+\frac{1}{2}}^{(2)}(x). \tag{3.49}$$

3.6 Wave Scattering from Rigid Surface

The problem of plane wave scattering from a rigid surface is to find that solution to the Helmholtz equation which satisfies

1. the Neumann boundary condition on the rigid surface,
2. the boundary condition at infinity as when $r \to \infty$, the wave is a plane wave, and
3. the outward radiation condition.

By linearity, the solution has two parts: the incident plane wave and the scattered wave:

$$\psi = \psi_{in} + \psi_{scat} \tag{3.50}$$

3.6.1 Scattering from a Rigid Cylinder

Suppose the incident plane wave perpendicular to an infinite rigid cylinder of radius a is:

$$\psi_{in} = e^{i\mathbf{k}\cdot\mathbf{r}} = e^{ikr\cos\varphi}. \tag{3.51}$$

This plane wave can be expressed in terms of cylindrical waves:

$$\psi_{in} = \sum_{n=0}^{\infty} \epsilon_n i^n J_n(kr)\cos(n\varphi) \tag{3.52}$$

where $J_n(x)$ is the nth-order Bessel function of x, and ϵ_n are Neumann symbols defined as:

$$\epsilon_n = \begin{cases} 1, & n = 0; \\ 2, & n \geq 1; \end{cases} . \tag{3.53}$$

The boundary condition that the solution represent a plane wave plus a scattered wave, outgoing only, demands that the scattered wave have the form:

$$\psi_{scat} = \sum A_n i^n H_n(kr)\cos(n\varphi) \tag{3.54}$$

where $H_n(x)$ is the nth-order Hankel function of x which ensures that all the scattered wave is outward. To satisfy the Neumann boundary condition, we must have:

$$\left.\frac{\partial\psi}{\partial r}\right|_{r=a} = 0 \tag{3.55}$$

which is:

$$\sum_{n=0}^{\infty} \epsilon_n i^n J_n'(ka)\cos(n\varphi) + \sum_{n=0}^{\infty} A_n i^n H_n'(ka)\cos(n\varphi) = 0 \tag{3.56}$$

So the solution for A_n is:

$$A_n = -\epsilon_n \frac{J_n'(ka)}{H_n'(ka)} \tag{3.57}$$

and the scattered wave is:

$$\psi_{scat} = -\sum_{n=0}^{\infty} \epsilon_n i^n \frac{J_n'(ka)}{H_n'(ka)} H_n(kr)\cos(n\varphi) \tag{3.58}$$

The complete solution for the Helmholtz equation under boundary conditions is:

$$\begin{aligned}\psi &= \psi_{in} + \psi_{scat} \\ &= \sum_{n=0}^{\infty} \epsilon_n i^n \left[J_n(kr) - \frac{J_n'(ka)H_n(kr)}{H_n'(ka)}\right]\cos(n\varphi)\end{aligned} \tag{3.59}$$

3.6.2 Scattering from a Rigid Sphere

Similarly, we can derive the solution in the rigid sphere case.

As shown in Fig. 3.1, for a unit magnitude plane wave with wavenumber k incident from direction $\boldsymbol{\theta}_k = (\theta_k, \varphi_k)$, the incident field at an observation point $\mathbf{r}_s = (\boldsymbol{\theta}_s, r_s) = (\theta_s, \varphi_s, r_s)$ can be expanded as

$$\psi_{in}(\mathbf{r}_s, \mathbf{k}) = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr_s) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s), \tag{3.60}$$

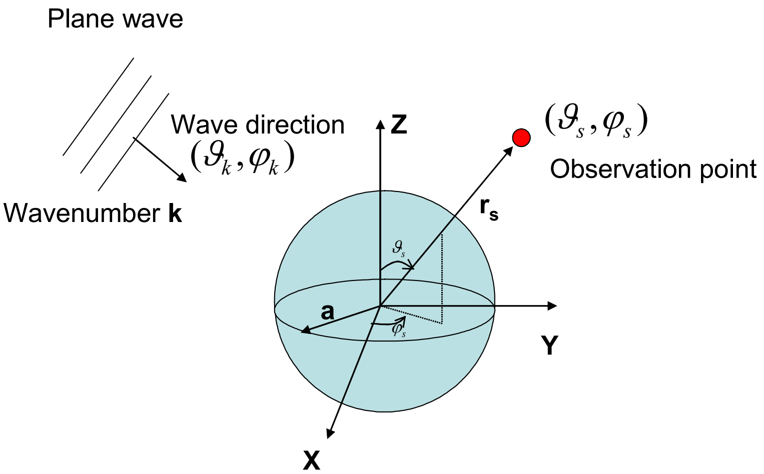


Figure 3.1: Plane wave incident on a rigid sphere.

where j_n is the spherical Bessel function of order n, Y_n^m is the spherical harmonics of order n and degree m. $*$ denotes the complex conjugation. At the same point, the field scattered by the rigid sphere of radius a is [47]:

$$\psi_{scat}\left(\mathbf{r}_s, \mathbf{k}\right) = -4\pi \sum_{n=0}^{\infty} i^n \frac{j_n^{'}(ka)}{h_n^{'}(ka)} h_n(kr_s) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s). \tag{3.61}$$

The total field on the surface ($r_s = a$) of the rigid sphere is:

$$\begin{aligned} \psi\left(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka\right) &= \left[\psi_{in}\left(\mathbf{r}_s, \mathbf{k}\right) + \psi_{scat}\left(\mathbf{r}_s, \mathbf{k}\right)\right]\big|_{r_s=a} \\ &= 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s), \end{aligned} \tag{3.62}$$

$$b_n(ka) = j_n(ka) - \frac{j_n^{'}(ka)}{h_n^{'}(ka)} h_n(ka), \tag{3.63}$$

where h_n are the spherical Hankel functions of the first kind.

Chapter 4
Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming[1]

This chapter describes a methodology for designing a flexible and optimal spherical microphone array for beamforming. Using the approach, a spherical microphone array can have very flexible layouts of microphones on the spherical surface, yet optimally approximate a desired beampattern of higher orders within a specified robustness constraint. Depending on the specified beampattern order, our approach automatically achieves optimal performances in two cases: when the specified beampattern order is reachable within the robustness constraint, we achieve a beamformer with optimal approximation of the desired beampattern; otherwise we achieve a superdirective beamformer, both robustly. For efficient implementation, we also developed an adaptive algorithm. It converges to the optimal performance quickly while exactly satisfying the specified frequency response and robustness constraint in each adaptive step without accumulated roundoff errors. One direct advantage is it makes much easier to build a real world system, such as those with cable outlets and a mounting base, with minimal effects on the performance. Simulation results are presented in this chapter. In the next chapter, we use a hemispherical microphone array to verify our algorithms. Experimental results will be presented in section 5.4.2.

[1] This chapter is based on our original work in [63][61][60][64].

4.1 Introduction

Spherical arrays of microphones are recently becoming a subject of interest as they allow omnidirectional sampling of the soundfield, and may have applications in soundfield capture [65]. The paper [71] presented a first analysis of such arrays, and showed how sound can be analyzed using them. This paper performed an elegant separation of the analysis and beamforming parts by using a modal beamformer structure. One implicit aspect of the analysis is that the distribution of microphones on the surface of the sphere seems to be redundant considering the results achieved. This is because the beamforming relies on numerical quadrature of spherical harmonics. In [71] this is done using a specified regular distribution of points, which has two issues:

1. For practical arrays, it may not be possible to place microphones precisely at all the quadrature locations. Moving even one microphone destroys the quadrature.

2. If higher order beamformers are necessary, quadrature points may be unavailable.

We discuss these issues further in the chapter. Here, we propose an extension of this approach that allows flexible microphone placements. Then we show how the array can achieve optimal performance.

This chapter is organized into four sections. In section 4.2, we present the basic principle of beamforming using a spherical microphone array. In section 4.3, we give a theoretical analysis of the discrete system. This part includes a summary of previous work and an analysis of orthonormality error: how it is introduced into the system, how it gets amplified and how it affects performance. To cancel the error noise optimally, we propose an improved solution and compare several design examples including a practical one. In section 4.4, we formulate our optimization problem into a finite linear system. We simplify the optimization by using reduced *degrees of freedom* (DOFs) for specified beamforming direction. The resulting beamformer then is checked against the robustness constraint. The upper bound of the beampattern order is derived theoretically. We again use the example from section 4.3 to demonstrate our simplified optimization. Nevertheless, we also point out its unrobustness especially for ill-conditioned layouts. This limitation will be solved in section 4.5 by a controlled trade-off between the accuracy of approximation and robustness. We formulate it as a constrained optimization problem and develop an adaptive implementation. We will show our algorithm automatically optimizes in two different situations: the superdirective beampattern or the desired beampattern of pre-specified order. Our adaptive implementation inherits the advantages of the classical ones in [38] and [24].

4.2 Principle of Spherical Beamformer

The basic principle of spherical beamformer is to make use of the orthonormality of spherical harmonics to decompose the soundfield arriving at a spherical array. The orthogonal components of the soundfield are then linearly combined to approximate a desired beampattern [71].

Using the notations in section 3.6.2 and Fig. 3.1, suppose we capture the sound field using a spherical microphone array, each microphone at the spherical surface point $\boldsymbol{\theta}_s$ samples the complex pressure of the total field as:

$$\psi\left(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka\right) = \left[\psi_{in}\left(\mathbf{r}_s, \mathbf{k}\right) + \psi_{scat}\left(\mathbf{r}_s, \mathbf{k}\right)\right]\Big|_{r_s=a} = 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s), \tag{4.1}$$

$$b_n(ka) = j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka). \tag{4.2}$$

Following [71], if we assume that the pressure recorded at each point $\boldsymbol{\theta}_s$ on the surface of the sphere Ω_s, is weighted by

$$W_{n'}^{m'}(\boldsymbol{\theta}_s, ka) = \frac{Y_{n'}^{m'}(\boldsymbol{\theta}_s)}{4\pi i^{n'} b_{n'}(ka)}, \tag{4.3}$$

applying the orthonormality of spherical harmonics

$$\int_{\Omega_s} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) d\Omega_s = \delta_{nn'}\delta_{mm'}, \tag{4.4}$$

the total output from a pressure-sensitive spherical surface is

$$\int_{\Omega_s} \psi\left(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka\right) W_{n'}^{m'}(\boldsymbol{\theta}_s, ka) d\Omega_s = Y_{n'}^{m'}(\boldsymbol{\theta}_k). \tag{4.5}$$

This shows the response of the plane wave incident from $\boldsymbol{\theta}_k$, for a continuous pressure-sensitive spherical microphone, is $Y_{n'}^{m'}(\boldsymbol{\theta}_k)$. Since any square-integrable function $F(\boldsymbol{\theta})$ can be expanded in terms of complex spherical harmonics, we can implement arbitrary beampatterns. For example, an ideal beampattern looking at the direction $\boldsymbol{\theta}_0$ can be modeled as a delta function

$$F(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{4.6}$$

which can be expanded into an infinite series of spherical harmonics [4]:

$$F(\boldsymbol{\theta},\boldsymbol{\theta}_0) = 2\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}). \tag{4.7}$$

So the weight at each point $\boldsymbol{\theta}_s$ to achieve this beampattern is

$$w(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) = \sum_{n=0}^{\infty} \frac{1}{2i^n b_n(ka)} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}_s). \tag{4.8}$$

The advantage of this system is that it can be steered into any 3D direction *digitally* with the same beampattern. This is for an ideal continuous microphone array on a spherical surface. Further, to achieve this we need infinite summations.

4.3 Discrete Spherical Array Analysis

This section also follows [71], but we make the band limit restrictions explicit. For a real-world system, however, we have a discretely sampled array with S microphones mounted at $(\boldsymbol{\theta}_s), s = 1, 2, ..., S$. To adapt the spherical beamforming principle to the discrete case, the continuous integrals are approximated by weighted summations, or quadratures[2] [84, p. 71]:

$$\frac{4\pi}{S} \sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) C_{n'}^{m'}(\boldsymbol{\theta}_s) = \delta_{nn'}\delta_{mm'}, \tag{4.9}$$

$$(n = 0, ..., N_{eff}; m = -n, ..., n;$$

$$n' = 0, ..., N; m' = -n', ..., n'),$$

where $C_{n'}^{m'}(\boldsymbol{\theta}_s)$ is the quadrature coefficient for $Y_{n'}^{m'}$ at $\boldsymbol{\theta}_s$. N_{eff} is the maximum order with significant strength, i.e. the band limit of spatial frequency in terms of

[2] While cubature is sometimes used for representing nodes and weights in 2D, we prefer to use the word quadrature, which is in any case used for still higher dimensions.

spherical harmonics orders. N is the order of beamformer. (4.9) can be solved in the *least squares* sense to minimize the 2-norm of the residues for (4.9).

Therefore, to approximate a regular beampattern of order N, which is bandlimited in the sense that it has no components of order greater than N,

$$F_N(\boldsymbol{\theta},\boldsymbol{\theta}_0) = 2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}), \tag{4.10}$$

the weights for the beamformer are:

$$w_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) = \sum_{n=0}^{N} \frac{1}{2i^n b_n(ka)} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}_s) C_n^m(\boldsymbol{\theta}_s). \tag{4.11}$$

To evaluate the robustness of a beamformer, we use the *white noise gain* (WNG) [16], usually in dB scale:

$$WNG(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) = 10 \log_{10} \left(\frac{|\mathbf{dW}|^2}{\mathbf{W}^H \mathbf{W}} \right), \tag{4.12}$$

where $\mathbf{d}$ is the row vector of complex pressure at each microphone position produced by the plane wave of unit magnitude from the desired beamforming direction $\boldsymbol{\theta}_0$ and $\mathbf{W}$ is the column vector of complex weights for each microphone. WNG defines the sensitivity on the white noise including the environmental noise, the device noise and implicitly, the microphone position mismatches among other perturbations. Positive WNG means an attenuation of white noise, whereas negative means an amplification.

To evaluate the directivity of a beampattern, we use the *directivity index* (DI) [16], also in dB:

$$DI(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) = 10 \log_{10} \left(\frac{4\pi \left| H(\boldsymbol{\theta}_0) \right|^2}{\int_{\Omega_s} \left| H(\boldsymbol{\theta}) \right|^2 d\Omega_s} \right), \tag{4.13}$$

where $H(\boldsymbol{\theta})$ is the actual beampattern and $H(\boldsymbol{\theta}_0)$ is the component in the desired look direction. For regular beampattern of order N, the DI is $20\log_{10}(N+1)$. DI represents the ability of the array to suppress a diffuse noise field. It is the ratio of the gain for the look direction $\boldsymbol{\theta}_0$ to the average gain over all directions.

4.3.1 Previous Work

The previous work can be summarized according to the choices made for $C_n^m(\boldsymbol{\theta}_s)$ and optimization.

In [71], $C_n^m(\boldsymbol{\theta}_s)$ are intuitively chosen to be unit to provide relative accuracy for some "uniform" layouts such as the 32 nodes defined by a truncated icosahedron. This straightforward choice simplifies the computation, however, it is not for "non-uniform" layouts and it doesn't leave any other DOF for optimization subject to the WNG constraint except the beampattern order N. In addition, we will see that even small errors can destroy the beampattern. In [3], several options are mentioned including equiangular grid layout [50] and an intuitive equidistance layout [36]. The common limitation of those schemes is that they are inflexible. If a patch of the spherical surface is inappropriate for mounting microphones, the orthonormality error may be large, thereby destroying the beampattern as the quadrature relation will not hold.

The approach in [52, Chapter 3] is equivalent to choosing $C_n^m(\boldsymbol{\theta}_s)$ to be independent of $\boldsymbol{\theta}_s$. The remaining $(N+1)^2$ DOFs are not used to satisfy (4.9) but maximize the directivity within WNG constraint. The optimization is performed by using an undetermined Lagrangian multiplier. Since there is no simple relation between the multiplier and the resulting WNG, the implementation uses a straightforward trial-and-error strategy.

4.3.2 Orthonormality Error Noise Analysis

Unfortunately, (4.9) can't be satisfied exactly for over-determined or rank-deficient system in general, which is usually the case. In addition, the number of equations in (4.9) for each pair of n' and m' depends on N_{eff}. Then for any choice of $C_{n'}^{m'}(\boldsymbol{\theta}_s)$, we always have:

$$\frac{4\pi}{S}\sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s)Y_{n'}^{m'}(\boldsymbol{\theta}_s)C_{n'}^{m'}(\boldsymbol{\theta}_s) = \delta_{nn'}\delta_{mm'} + \epsilon_{nn'}^{mm'}, \tag{4.14}$$

where $\epsilon_{nn'}^{mm'}$ is usually the non-zero error caused by discreteness.

Now, we see how this error could degrade the performance of soundfield decomposition. To extract the component of order n' and degree m' from the soundfield (3.62), we consider the quadratures of (4.5) for $\psi_n^m(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka)$, denoting one component of ψ at order n and degree m:

$$P_n^m(\boldsymbol{\theta}_k, ka) = \int_{\Omega_s} \psi_n^m(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka)\frac{Y_{n'}^{m'}(\boldsymbol{\theta}_s)C_{n'}^{m'}(\boldsymbol{\theta}_s)}{4\pi i^{n'} b_{n'}(ka)} d\Omega_s \tag{4.15}$$

where:

$$\psi_n^m(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka) = 4\pi i^n b_n(ka) Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s). \tag{4.16}$$

Using S discrete points, we have:

$$P_n^m(\boldsymbol{\theta}_k, ka) = Y_n^m(\boldsymbol{\theta}_k)\left[\frac{i^n b_n(ka)}{i^{n'} b_{n'}(ka)}\right](\delta_{nn'}\delta_{mm'} + \epsilon_{nn'}^{mm'}). \tag{4.17}$$

We notice that:

$$\left[\frac{i^n b_n(ka)}{i^{n'} b_{n'}(ka)}\right]\delta_{nn'}\delta_{mm'} = \delta_{nn'}\delta_{mm'} \tag{4.18}$$

So, (4.17) can be rewritten as:

$$P_n^m(\boldsymbol{\theta}_k, ka) = Y_n^m(\boldsymbol{\theta}_k)\{\delta_{nn'}\delta_{mm'} + \left[\frac{i^n b_n(ka)}{i^{n'} b_{n'}(ka)}\right]\epsilon_{nn'}^{mm'}\} \tag{4.19}$$

The second term is the noise caused by scaled orthonormality error $\epsilon_{nn'}^{mm'}$. We call it the *orthonormality error noise* (OEN) which is possible with any discrete microphone array layout. To prevent it from damaging the orthonormality, we must have:

$$\left|\frac{i^n b_n(ka)}{i^{n'} b_{n'}(ka)}\epsilon_{nn'}^{mm'}\right| \ll 1 \tag{4.20}$$

So, we get:

$$|\epsilon_{nn'}^{mm'}| \ll \left|\frac{b_{n'}(ka)}{b_n(ka)}\right|, \quad \forall n, n', m, m' \tag{4.21}$$

Since b_n decays very quickly with respect to n as shown in Fig.4.1, for a given microphone number and layout, we cannot decompose the high order component of soundfield if (4.21) fails. In addition, we can see (4.21) is independent of magnitude of the incoming sound wave. That means: even if the microphones have recorded the high order components, the system may be unable to decompose them.

To prevent errors from being amplified, we include the frequency-dependent scale factor

$$B_{n'}^n(ka) = \frac{b_n(ka)}{b_{n'}(ka)} \tag{4.22}$$

into (4.9). The linear system for quadrature coefficients then becomes a *weighted*

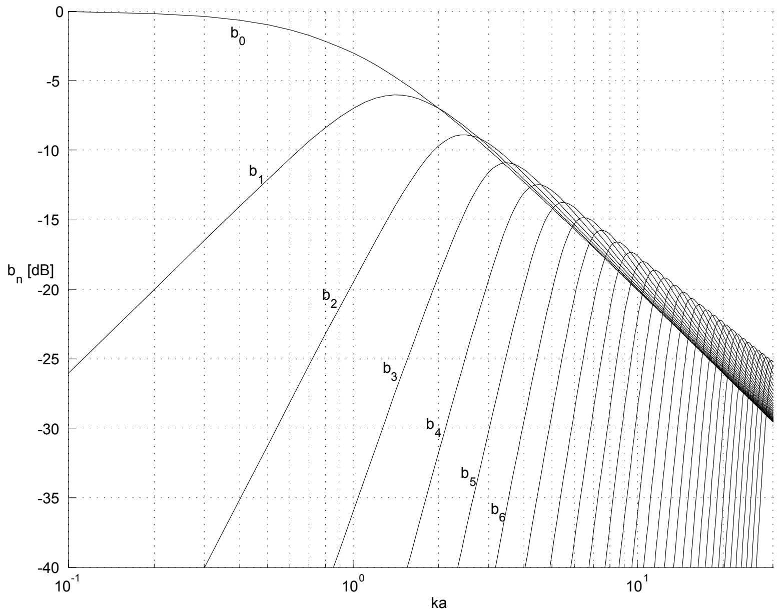


Figure 4.1: $b_n(ka)$ for orders from 0 to 30. Given ka, $b_n(ka)$ decays very quickly with respect to n.

least squares problem:

$$\min_{C_{n'}^{m'}(\boldsymbol{\theta}_s,ka)} \left\| \frac{4\pi}{S} \sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) C_{n'}^{m'}(\boldsymbol{\theta}_s, ka) B_{n'}^{n}(ka) - \delta_{nn'}\delta_{mm'} \right\|_2^2, \qquad (4.23)$$

$$(n = 0, ..., N_{eff}; m = -n, ..., n;$$

$$n' = 0, ..., N; m' = -n', ..., n').$$

The discrete orthonormalities in (4.23) weighted with smaller $|B_{n'}^n(ka)|$ are less important than those with larger $|B_{n'}^n(ka)|$. Since $B_{n'}^n(ka)$ converges to zero when n increases for given n', compared with (4.9), (4.23) suppresses the orthonormalities in higher orders and adds more accuracy to the lower orders. In addition, since $B_{n'}^n(ka)$ converges to zero rapidly with respect to n, this weighting also significantly lessens the sensitivity of the solutions on the choice of N_{eff} only if $\left|b_{N_{eff}}(ka)\right|$ is small enough compared with $|b_N(ka)|$, e.g. for $ka = 1.83$, $N_{eff} = N + 3$ will make

$\left|b_{N_{eff}}(ka)\right|$ at least 40dB below $|b_N(ka)|$ for $N = 3, 4, 5$.

4.3.3 Design Examples

A quadrature formula provides locations at which we evaluate the function and weights to multiply and sum up to obtain the integral. It was proven that any quadrature formula of order N over the sphere should have more than $S = (N+1)^2$ quadrature nodes [81][48]. If the quadrature functions have the bandwidth upto order N, to achieve the exact quadrature using equiangular layout, we need $S = 4(N+1)^2$ nodes [28]. This is too large and redundant for our application. For special layouts, S can be made much smaller. For example, for a spherical grid which is a Cartesian product of equispaced grid in φ and the grid in θ in which the nodes distributed as zeros of the Legendre polynomial of degree $N+1$ with respect to $\cos\theta$ we need $S = 2(N+1)^2$ [77]. Another special design with equal quadrature coefficients is called the *spherical t-design* [48]. While these designs achieve exact quadratures, however, they are for strictly band-limited functions and S is still large considering our quadrature function is the multiplication of *two* band-limited functions. This means that to achieve a beampattern of order five we need at least 144 microphones.

If we use approximate quadrature formula, then it may be possible to reduce S. Intuitively, we want the microphones distributed "uniformly" on the spherical surface. Unfortunately, it has been proven that only five regular polyhedrons (or called "Platonic solids") exist: tetrahedron, cube, dodecahedron, icosahedron and octahedron [34]. Semi-regular polyhedrons can be used also such as the truncated icosahedron used in [71] to layout 32 microphones. The general problem to distribute arbitrary number of points approximately "uniformly" on a spherical surface is numerically solved by Fliege in [36] by minimizing the potential energy of a distribution of movable electrons on a perfect conducting sphere. Then, one set of optimal quadrature coefficients are solved.

Fig. 4.2(a) shows Fliege's 64-node layout in [36]. Fig. 4.2(b) shows the errors of orthogonal conditions of spherical harmonics using those optimal quadrature coefficient for each nodes. However, those optimal quadrature coefficients are not generally available for other flexible layouts. More importantly, for a given layout, especially for a "non-uniform" one, we will explain in the next section that there is no such a single optimal set of coefficients to satisfy all orthogonal conditions.

For example, Fig. 4.3(a) is the layout of our array using the angular positions of those 64 nodes with four nodes at the bottom removed because of the cable outlet. Fig. 4.3(b) shows the orthonormality errors using the 60 nodes and quadrature coefficients. It is less accurate compared with Fig. 4.2(b). The top row in Fig. 4.4 shows the beampatterns from order three to five using this configuration. At order three, the beampattern is distorted. At order four, the orthonormality errors significantly damage the beampattern. At order five, the beampattern is almost completely destroyed. The bottom row in Fig. 4.4 shows the beampatterns using the quadrature coefficients solved by (4.23), which optimally approximate the regular beampatterns.

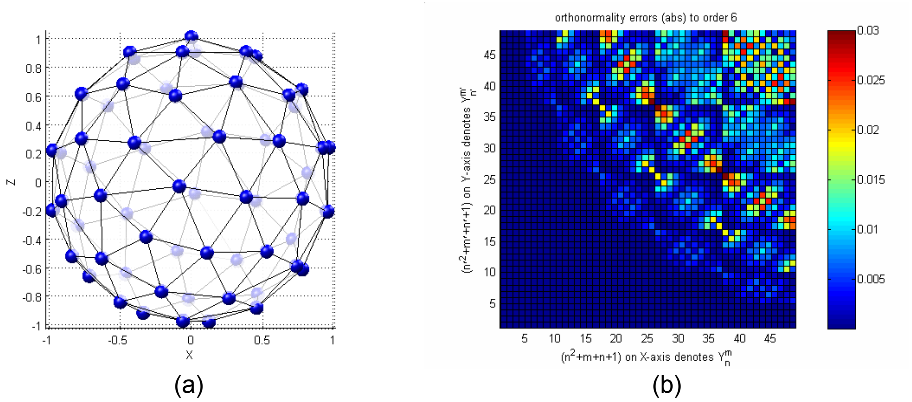


Figure 4.2: (a) Fliege's 64 nodes. (b) The orthonormality errors.

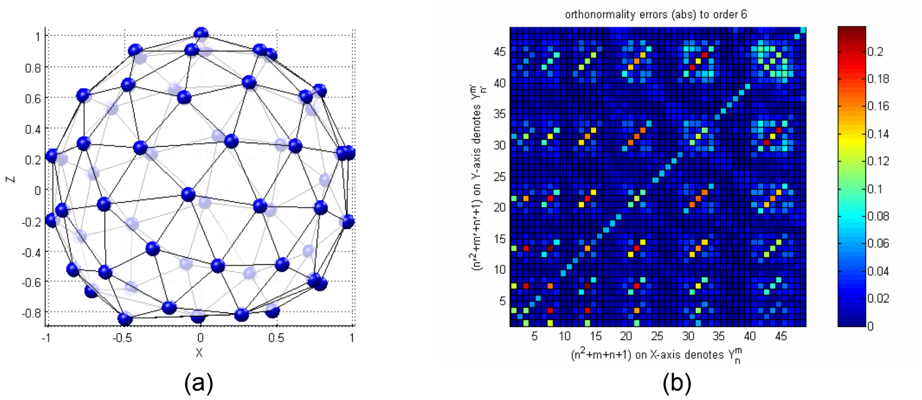


Figure 4.3: (a) Same as in Fig. 4.2(a) except the bottom four nodes are removed. (b) The orthonormality errors.

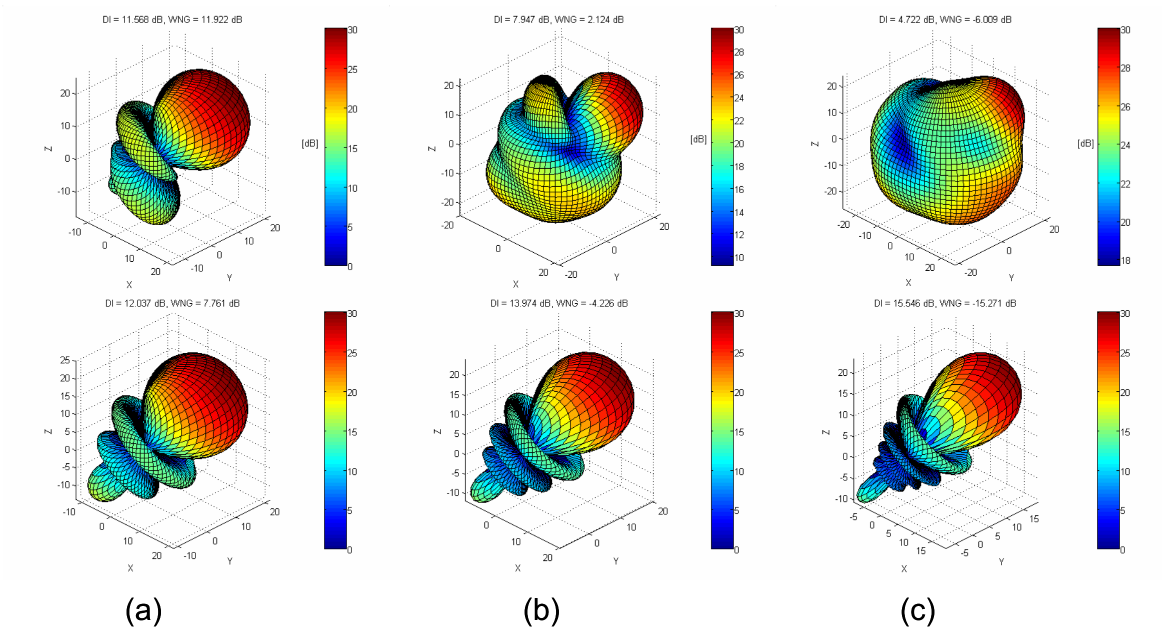


Figure 4.4: $(a)-(c)$ show the beampatterns from order 3 to 5 using the 60-node array in Fig. 4.3 with radius of 10cm. The top row uses Fliege's cubature coefficients. The bottom row uses the coefficients solved by (4.23). Simulated at 1KHz.

4.4 Simplified Optimization of Desired Beampattern for Discrete Array

In total, we have $S \times (N+1)^2$ quadrature coefficients for each frequency. However, those coefficients are not directly related to the WNG constraint, which can't lead to an explicit constrained optimization easily. In addition, the $S \times (N+1)^2$ DOFs are intuitively redundant, specifically, given only S microphones. That means $C_{n'}^{m'}(\boldsymbol{\theta}_s, ka)$ should somehow be independent of n' and m'. Plausibly, we have the following least squares problem with only S variables:

$$\frac{4\pi}{S}\sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) C(\boldsymbol{\theta}_s, ka)\frac{b_n(ka)}{b_{n'}(ka)} = \delta_{nn'}\delta_{mm'}. \tag{4.24}$$

This method, however, is infeasible especially for ill-conditioned layouts since in practice we aim to use as few microphones as possible, so the linear equations will largely outnumber the microphones which makes the optimization less meaningful. Instead of satisfying (4.24), we can use $C(\boldsymbol{\theta}_s, ka)$ to maximize the directivity subject to the WNG constraint, which simply returns to a classical solved problem of designing a superdirective and robust beamformer [24]. This method doesn't aim to optimally approximate a desired beampattern of a specified order N. It invalidates the main advantage of using a spherical microphone array, i.e. the same beampattern in all steering directions. An example using this feature is the system described in [65].

An alternative explanation is: because the optimization on (4.24) is independent on the beamforming direction, it is too ambitious to find such a single set of optimal weights for every beamforming directions. This, in turn, also implies that the optimization using S DOFs should be performed for each beamforming direction. In this section, we follow this approach and develop algorithms to find the optimal weights for regular beampatterns. We will first formulate the discrete spherical beamformer into a linear system. The optimal solution of this linear system subject to the WNG constraint will be the optimal approximation of the desired beampattern. Then we present straightforward solutions. Design examples are provided to demonstrate our approach.

4.4.1 Discrete Spherical Beamformer as Finite Linear System

To achieve a regular beampattern of order N (4.10), a discrete spherical beamformer with S microphones can be formulated as a finite linear system:

$$\begin{aligned} \mathbf{AW} &= c_N \mathbf{B}_N, && (4.25) \\ \mathbf{dW} &= 1, && (4.26) \end{aligned}$$

where (4.25) defines the beampattern, and (4.26) the frequency response to the sound from the beamforming direction. Without loss of generality, here we consider an all-pass filter. In (4.25), $\mathbf{A}$ are the coefficients of the spherical harmonics expansion of the soundfield in (3.62):

$$\mathbf{A} = [\ \mathbf{A}_1 \quad \mathbf{A}_2 \quad \cdots \quad \mathbf{A}_S\ ], \tag{4.27}$$

$$\mathbf{A}_s = 4\pi \begin{bmatrix} i^0 b_0(ka) Y_0^{0*}(\boldsymbol{\theta}_s) \\ i^1 b_1(ka) Y_1^{-1*}(\boldsymbol{\theta}_s) \\ \vdots \\ i^N b_N(ka) Y_N^{N*}(\boldsymbol{\theta}_s) \\ i^{(N+1)} b_{(N+1)}(ka) Y_{(N+1)}^{-(N+1)*}(\boldsymbol{\theta}_s) \\ \vdots \\ i^{N_{eff}} b_{N_{eff}}(ka) Y_{N_{eff}}^{N_{eff}*}(\boldsymbol{\theta}_s) \end{bmatrix}, \tag{4.28}$$

$$(s = 1, ..., S.)$$

$\mathbf{W}$ is the vector of complex weights to be assigned to each microphone at $(\boldsymbol{\theta}_s, a)$:

$$\mathbf{W} = \begin{bmatrix} W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, ka) \\ W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_2, ka) \\ \vdots \\ W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_S, ka) \end{bmatrix}. \tag{4.29}$$

$\mathbf{B}_N$ is the vector of coefficients of the beampattern of order N steered to $(\boldsymbol{\theta}_0)$ in (4.10):

$$\mathbf{B}_N = 2\pi \begin{bmatrix} Y_0^{0*}(\boldsymbol{\theta}_0) \\ Y_1^{-1*}(\boldsymbol{\theta}_0) \\ \vdots \\ Y_N^{N*}(\boldsymbol{\theta}_0) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{4.30}$$

In (4.26), $\mathbf{d}$ is the row vector of the complex pressure at each microphone position produced by a plane wave of unit magnitude from the desired beamforming direction $\boldsymbol{\theta}_0$:

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_S \end{bmatrix}^T = \begin{bmatrix} \psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0, ka) \\ \psi(\boldsymbol{\theta}_2, \boldsymbol{\theta}_0, ka) \\ \vdots \\ \psi(\boldsymbol{\theta}_S, \boldsymbol{\theta}_0, ka) \end{bmatrix}^T. \tag{4.31}$$

In (4.25), c_N is a normalizing coefficient to satisfy the all-pass frequency response (4.26). The least squares solution of (4.25) is

$$\mathbf{W} = \left[(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\right] c_N \mathbf{B}_N. \tag{4.32}$$

Then c_N can be determined using (4.26). If we assume (4.25) has small residues, from (4.10), the *a priori* estimate of c_N is

$$c_N \approx \frac{1}{2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_0) Y_n^{m*}(\boldsymbol{\theta}_0)}. \tag{4.33}$$

According to the spherical harmonic addition theorem, c_N is independent of $(\boldsymbol{\theta}_0)$ and (4.33) can be simplified easily

$$c_N \approx \frac{1}{\sum_{n=0}^{N} \frac{2n+1}{2} P_n(\cos 0)} = \frac{2}{(N+1)^2}. \tag{4.34}$$

We will use it to predict the maximum order in the next subsection.

Note that because we have absorbed all frequency dependence (ka) into the linear system, this must be solved for each frequency.

4.4.2 Maximum Beampattern Order for Robust Beamformer

A robust beamformer requires a minimum WNG of δ^2 (such as -6 dB in [16]):

$$\frac{|\mathbf{dW}|^2}{\mathbf{W}^H\mathbf{W}} \geq \delta^2. \tag{4.35}$$

Substituting (4.26) into (4.35), we have a spherical constraint on $\mathbf{W}$:

$$\mathbf{W}^H\mathbf{W} \leq \delta^{-2}. \tag{4.36}$$

Assume the maximum order we can possibly decompose robustly is $N_{\max}$, then the linear system (4.25) becomes:

$$\mathbf{AW} = c_{N_{\max}} \mathbf{B}_{N_{\max}}, \tag{4.37}$$

where

$$c_{N_{\max}} \approx \frac{2}{(N_{\max}+1)^2}. \tag{4.38}$$

Suppose we have a least squares solution of $\mathbf{W}$ to (4.37), considering the following equations of order $N_{\max}$:

$$4\pi \begin{bmatrix} i^{N_{\max}} b_{N_{\max}}(ka) Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_1) \\ i^{N_{\max}} b_{N_{\max}}(ka) Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_2) \\ \ldots \\ i^{N_{\max}} b_{N_{\max}}(ka) Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_S) \end{bmatrix}^T \begin{bmatrix} W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, ka) \\ W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_2, ka) \\ \vdots \\ W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_S, ka) \end{bmatrix}$$

$$\approx 2\pi c_{N_{\max}} Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_0), \tag{4.39}$$

$$(M = -N_{\max}, ..., N_{\max})$$

Using Cauchy's Inequality, we have:

$$\begin{aligned} & \left| \sum_{s=1}^{S} i^{N_{\max}} b_{N_{\max}}(ka) Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_s) W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) \right|^2 \\ \leq & \left(\sum_{s=1}^{S} \left| i^{N_{\max}} b_{N_{\max}}(ka) Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_s) \right| \left| W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) \right| \right)^2 \\ \leq & \sum_{s=1}^{S} \left| i^{N_{\max}} b_{N_{\max}}(ka) Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_s) \right|^2 \sum_{s=1}^{S} \left| W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) \right|^2 \\ = & \left| b_{N_{\max}}(ka) \right|^2 \sum_{s=1}^{S} \left| Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_s) \right|^2 \sum_{s=1}^{S} \left| W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) \right|^2 \end{aligned}$$

So:

$$\begin{aligned} \mathbf{W}^H \mathbf{W} &= \sum_{s=1}^{S} \left| W(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) \right|^2 \\ &\gtrsim \frac{\left| c_{N_{\max}} Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_0) \right|^2}{2 \left| b_{N_{\max}}(ka) \right|^2 \sum_{s=1}^{S} \left| Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_s) \right|^2}. \end{aligned} \tag{4.40}$$

From (4.36), we have:

$$\delta^{-2} \gtrsim \frac{\left|c_{N_{\max}} Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_0)\right|^2}{2\left|b_{N_{\max}}(ka)\right|^2 \sum_{s=1}^{S} \left|Y_{N_{\max}}^{M*}(\boldsymbol{\theta}_s)\right|^2}, \tag{4.41}$$

which is:

$$\left|b_{N_{\max}}(ka)\right| \gtrsim \frac{\sqrt{2}\left|Y_{N_{\max}}^{M}(\boldsymbol{\theta}_0)\right|\delta}{(N_{\max}+1)^2\sqrt{\sum_{s=1}^{S}\left|Y_{N_{\max}}^{M}(\boldsymbol{\theta}_s)\right|^2}}, \quad \text{for all } M = -N_{\max}, ..., N_{\max}. \tag{4.42}$$

Therefore, given a spherical microphone array and the beamforming direction, for each frequency, we know the upper bound order $N_{\max}$ of a robust beamformer.

4.4.3 Design Examples

We still use the 60-nodes layout at 1 kHz as an example. We set $\delta = 10^{-0.3}$, i.e. the minimum WNG is -6dB. Fig. 4.5 shows the bounds for $\boldsymbol{\theta}_0 = (\pi/4, \pi/4)$, where the red (dashed) lines show the bounds. For example, the black (dash-dot) line shows 1 kHz ($ka = 1.8318$) in our case. Since the intersection of the black line and the $n = 4$ mode is above the $n = 4$ bound, while its intersection with $n = 5$ mode is below the $n = 5$ bound, we predict the maximum order of a robust beamformer is four. Fig. 4.6 shows resulted beampatterns of order from three to five using the simplified optimization. The WNG of order 5 beamformer falls below the minimum of -6dB as predicted. These results are almost identical to the bottom row in Fig. 4.4.

As a general example, Fig. 4.7 shows a random layout of 64 nodes. Fig. 4.8 shows the beampatterns from order three to five at 1 kHz. Their WNGs are all below -6dB. It seems to be contradictory to the bounds shown in Fig. 4.9, however,

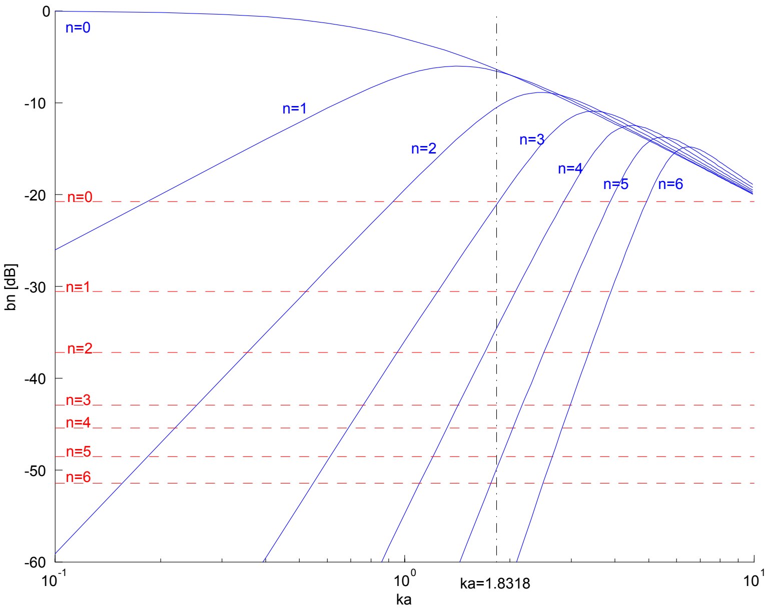


Figure 4.5: Using the 60-nodes array in Fig. 4.3, the horizontal lines show the mode bounds for robust beamforming at given ka.

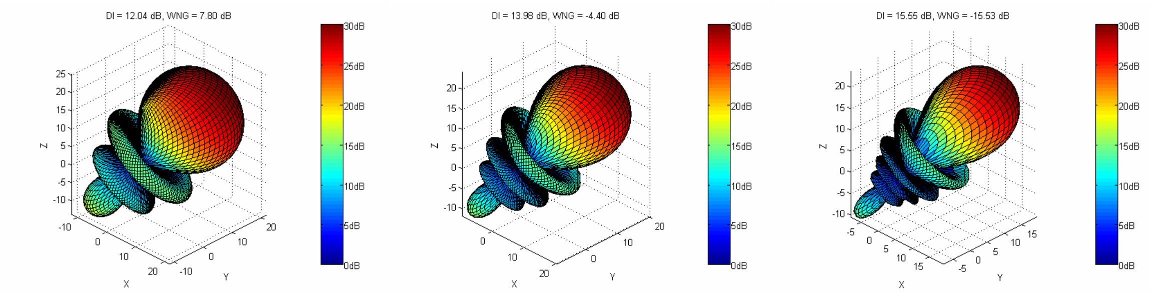


Figure 4.6: The beampatterns of order 3 to 5 for the 60-nodes array in Fig. 4.3 using simplified optimization.

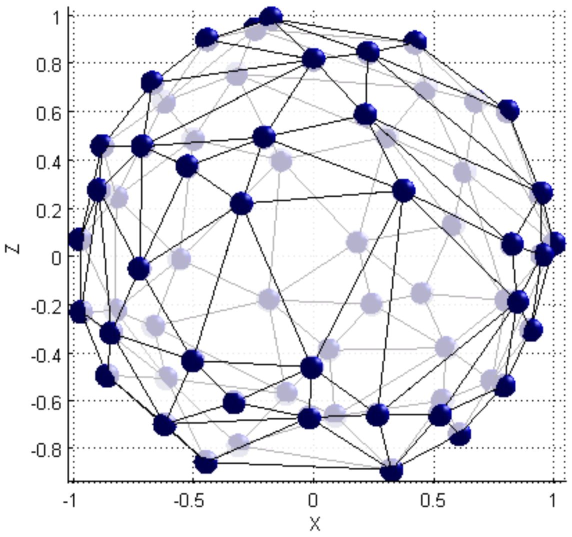


Figure 4.7: The random layout of 64 microphones on a sphere of radius 10cm.

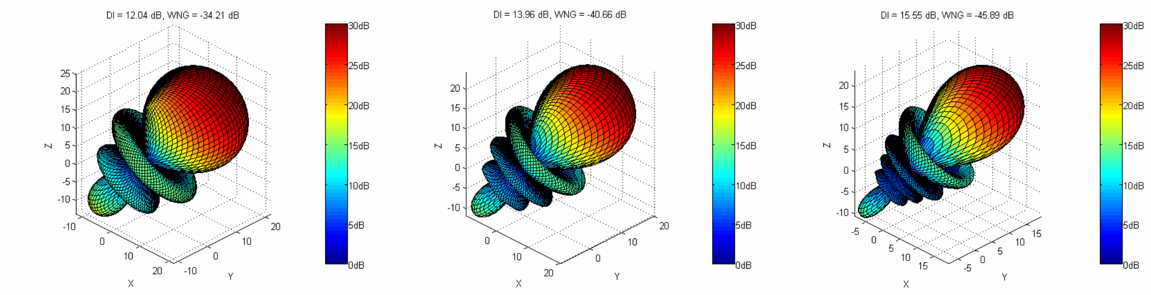


Figure 4.8: Unconstrained beampatterns from order 3 to 5 for the array in Fig. 4.7.

we will show in the next section that the robust beamformers up to order four are still achievable, with constrained relaxation on the least squares solution (4.32).

4.5 Optimal Approximation Subject to the WNG Constraint

In the previous section, we minimize the residue of a finite linear system and check the resulted beamformer against the WNG constraint. In this section, we extend the algorithm further to address the following two aspects:

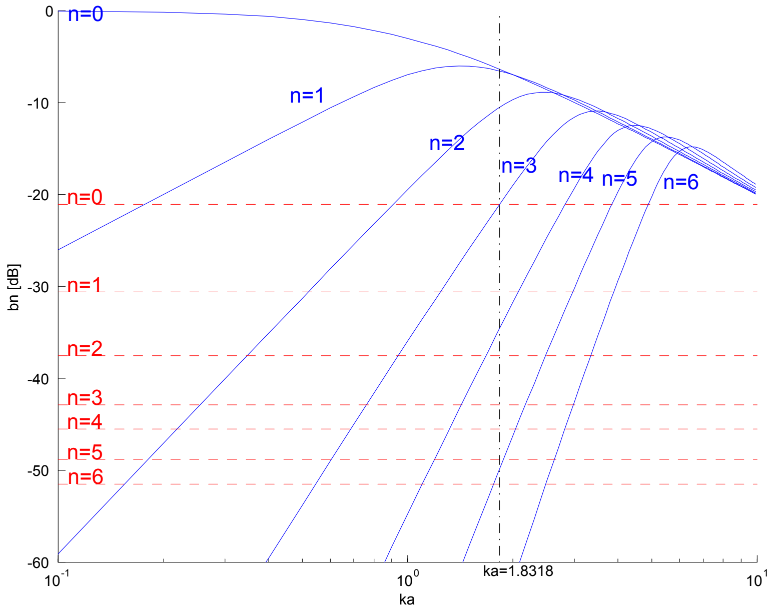


Figure 4.9: Mode bounds for the array in Fig. 4.7.

1. we need to relax the approximations to stay within the constraint such as in Fig. 4.7.

2. we need a robust and superdirective beamformer with maximum directivity index.

The two problems are closely related to each other and can be formulated as a unified constrained optimization.

4.5.1 Constrained Optimization

To design a robust spherical beamformer with finite microphones, yet optimally approximate the desired beampattern to certain order (e.g. the ideal beampattern in our case), we need to optimize the following 2-norm function:

$$\min_{\mathbf{W}} \|\mathbf{AW} - c_N \mathbf{B}_N\|_2^2 \tag{4.43}$$

subject to:

$$\begin{aligned} \mathbf{dW} &= 1, & (4.44) \\ \mathbf{W}^H\mathbf{W} &\leq \delta^{-2}. & (4.45) \end{aligned}$$

This optimization can be numerically solved by some blackbox software packages, such as MATLAB function `fmincon`, etc. Another way is to use Tikhonov regularization. Specifically, we place a 2-norm constraint on $\mathbf{W}$ by appending a damping matrix with the regularization parameter λ:

$$\begin{bmatrix} \mathbf{A} \\ \lambda\mathbf{I} \end{bmatrix} \mathbf{W} = \begin{bmatrix} c_N\mathbf{B}_N \\ \mathbf{0} \end{bmatrix}, \tag{4.46}$$

The solution is:

$$\mathbf{W} = \left[(\mathbf{A}^H\mathbf{A} + \lambda^2\mathbf{I})^{-1}\mathbf{A}^H\right] c_N\mathbf{B}_N.$$

This regularization parameter λ, however, is not directly related to the WNG constraint. A trial-and-error strategy can be used in implementation.

4.5.2 Adaptive Implementation

The most straightforward way to implement this system is to precompute all the weights for each pre-defined 3D direction and store them in a lookup table. This method, however, is not very efficient because of the obvious trade-off between the spatial resolution and the cost of storage. In this subsection, we reformulate our problem so that we can parallel the method in [24] to design an adaptive implementation which automatically and robustly converges to the desired beamformer of a specified order in any steering directions.

We rewrite the object function into an ellipsoidal form:

$$\min_{\mathbf{W}} \|\mathbf{A}\mathbf{W} - c_N \mathbf{B}_N\|_2^2 = \min_{\widetilde{\mathbf{W}}} \widetilde{\mathbf{W}}^H \mathbf{R} \widetilde{\mathbf{W}}, \tag{4.47}$$

subject to

$$\begin{aligned} \mathbf{C}^H \widetilde{\mathbf{W}} &= \mathbf{g}, & (4.48) \\ \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} &\leq \delta^{-2} + 1, & (4.49) \end{aligned}$$

where

$$\begin{aligned} \widetilde{\mathbf{W}} &= \begin{bmatrix} \mathbf{W} \\ W_0 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \mathbf{A}^H \\ c_N \mathbf{B}_N^H \end{bmatrix} \begin{bmatrix} \mathbf{A} & c_N \mathbf{B}_N \end{bmatrix}, \\ \mathbf{C} &= \begin{bmatrix} d_1^* & 0 \\ d_2^* & 0 \\ \vdots & \vdots \\ d_S^* & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \end{aligned}$$

We know $W_0 = -1$ from (4.47), however, we include it as an extra variable into $\widetilde{\mathbf{W}}$ and its actual value is automatically determined by the constraint (4.48) in the process of optimization.

To solve this optimization, we first decompose $\widetilde{\mathbf{W}}$ into its orthogonal components:

$$\widetilde{\mathbf{W}} = \mathbf{W}_c + \mathbf{V}, \tag{4.50}$$

$$\mathbf{W}_c = \mathbf{C}[\mathbf{C}^H \mathbf{C}]^{-1} \mathbf{g}. \tag{4.51}$$

$\mathbf{W}_c$ is the least squares solution to satisfy the linear constraint (4.48). The residue is expected to be zero since usually (4.48) is a highly under-determined system. Substituting (4.51) into (4.49), we have:

$$\mathbf{V}^H\mathbf{V} \leq \delta^{-2} + 1 - \mathbf{g}^H[\mathbf{C}^H\mathbf{C}]^{-1}\mathbf{g} = b^2. \tag{4.52}$$

Thus, the WNG constraint becomes a spherical constraint on $\mathbf{V}$. Since $\mathbf{R}\widetilde{\mathbf{W}}(t)$ is the gradient of the object function (4.47) at step t, the tentative update vector is:

$$\widetilde{\mathbf{V}}(t+1) = \widetilde{\mathbf{P}}_c[\mathbf{V}(t) - \mu\mathbf{R}\widetilde{\mathbf{W}}(t)], \tag{4.53}$$

$\mathbf{V}(t)$ is the scaled projection of $\widetilde{\mathbf{V}}(t)$ into the sphere surface of radius b:

$$\mathbf{V}(t) = \begin{cases} \widetilde{\mathbf{V}}(t) & \text{for } \left|\widetilde{\mathbf{V}}\right|^2 \leq b^2 \\ b\frac{\widetilde{\mathbf{V}}(t)}{\left|\widetilde{\mathbf{V}}(t)\right|} & \text{for } \left|\widetilde{\mathbf{V}}\right|^2 > b^2 \end{cases}, \tag{4.54}$$

μ is the step size, and $\widetilde{\mathbf{P}}_c$ is the null space of $\mathbf{C}^H$:

$$\widetilde{\mathbf{P}}_c = \mathbf{I} - \mathbf{C}[\mathbf{C}^H\mathbf{C}]^{-1}\mathbf{C}^H. \tag{4.55}$$

The weights are updated as

$$\begin{aligned} \widetilde{\mathbf{W}}(t+1) &= \mathbf{W}_c + \begin{cases} \widetilde{\mathbf{V}}(t+1) & \text{for } \left|\widetilde{\mathbf{V}}\right|^2 \leq b^2 \\ b\frac{\widetilde{\mathbf{V}}(t+1)}{\left|\widetilde{\mathbf{V}}(t+1)\right|} & \text{for } \left|\widetilde{\mathbf{V}}\right|^2 > b^2 \end{cases} && (4.56) \\ &= \mathbf{W}_c + \mathbf{V}(t+1). && (4.57) \end{aligned}$$

We set the initial guess as:

$$\begin{aligned} \widetilde{\mathbf{W}}(0) &= \mathbf{R}^{-1}\mathbf{C}[\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C}]^{-1}\mathbf{g}, && (4.58) \\ \widetilde{\mathbf{V}}(0) &= \widetilde{\mathbf{W}}(0) - \mathbf{W}_c, && (4.59) \end{aligned}$$

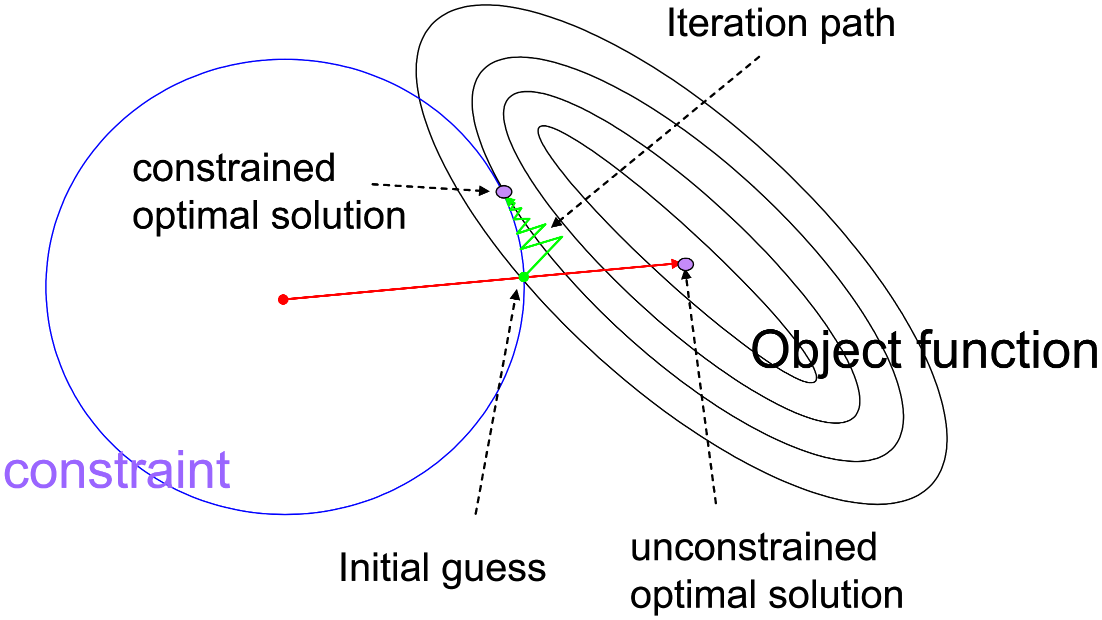


Figure 4.10: The geometric interpretation.

which is equivalent to the solution we get in section 4.4. If the resulting WNG is within constraint, the iteration will stay with this solution, otherwise, it will start the constrained optimization process, both automatically. At each step, the constraints (4.48) and (4.49) are satisfied exactly. In addition, similar to the methods in [24] and [38], round-off errors don't accumulate. This iteration is independent of the actual signal processing rate, so it may be implemented more efficiently as a parallel unit with other processors.

The geometric interpretation of this constrained optimization is shown in Fig. 4.10.

4.5.3 Convergence and Optimal Step Size

From (4.53) and (4.56), we have

$$\begin{aligned}\widetilde{\mathbf{V}}(t+1) &= \widetilde{\mathbf{P}}_c\left[\mathbf{V}(t) - \mu\mathbf{R}\left[\mathbf{W}_c + \mathbf{V}(t)\right]\right] \\ &= \widetilde{\mathbf{P}}_c\left[\mathbf{I}-\mu\mathbf{R}\right]\mathbf{V}(t) - \mu\widetilde{\mathbf{P}}_c\mathbf{R}\mathbf{W}_c. \qquad (4.60)\end{aligned}$$

Let

$$\widetilde{\mathbf{V}}(t+1) = \gamma(t+1)\mathbf{V}(t+1), \qquad (4.61)$$

$$\gamma(t+1) = \begin{cases} 1 & \text{for } \left|\widetilde{\mathbf{V}}\right|^2 \leq b^2 \\ \frac{\left|\widetilde{\mathbf{V}}(t+1)\right|}{b} & \text{for } \left|\widetilde{\mathbf{V}}\right|^2 > b^2 \end{cases} \geq 1, \qquad (4.62)$$

we then have:

$$\begin{aligned}\mathbf{V}(t+1) &= \frac{1}{\gamma(t+1)}\widetilde{\mathbf{P}}_c\left[\mathbf{I}-\mu\mathbf{R}\right]\mathbf{V}(t) - \frac{1}{\gamma(t+1)}\mu\widetilde{\mathbf{P}}_c\mathbf{R}\mathbf{W}_c \\ &= \frac{\widetilde{\mathbf{P}}_c\left[\mathbf{I}-\mu\mathbf{R}\right]^{t+1}}{\prod_{i=0}^{t}\gamma(i+1)}\mathbf{V}(0) - \left[\sum_{k=0}^{t}\frac{\widetilde{\mathbf{P}}_c^k\left[\mathbf{I}-\mu\mathbf{R}\right]^k}{\prod_{i=0}^{k}\gamma(i+1)}\right]\mu\widetilde{\mathbf{P}}_c\mathbf{R}\mathbf{W}_c. \qquad (4.63)\end{aligned}$$

To guarantee convergence, we need

$$0 < \mu < \frac{2}{\sigma_{\max}}, \qquad (4.64)$$

where $\sigma_{\max}$ is the maximum eigenvalue of $\mathbf{R}$.

Although it is difficult to precisely model and control γ if possible, it won't cause divergence from (4.62) and will be very close to one in practice. Thus the optimal step size μ_{opt} can be roughly estimated as the least squares solution of

$$\mu_{opt}\boldsymbol{\sigma} = \mathbf{1}, \qquad (4.65)$$

which is

$$\mu_{opt} = \left[\left(\boldsymbol{\sigma}^T\boldsymbol{\sigma}\right)^{-1}\boldsymbol{\sigma}^T\right]\mathbf{1}, \qquad (4.66)$$

where $\boldsymbol{\sigma}$ is the column vector of eigenvalues of $\mathbf{R}$, and $\mathbf{1}^T \triangleq [1, ..., 1]$ is the vector of ones with the length of $\boldsymbol{\sigma}$.

4.5.4 Simulation Results

We use our algorithm to solve the two cases we mentioned at the beginning of this section.

We go back to the example in Fig. 4.7 and Fig. 4.8. We first simulate the beamformer of order three. Fig. 4.11 shows the iteration process using the optimal step size. As can be seen from this figure, the optimization goal is not to maximize the DI, instead it converges to the regular beampattern of order three. The resulted beampattern is shown in Fig. 4.12. Fig. 4.14 shows the optimal approximations of the regular beampattern of order four subject to the WNG constraint. There is minimal difference between the beampatterns in Fig. 4.14 and Fig. 4.7(b). The comparisons of residues are shown in Fig. 4.13.

If we desire optimal directivity, we can approximate the ideal beampattern as (4.7). In practice, we just need to approximate an order above the theoretical upper bound derived in (4.42), such as order 5 in this case. It is best to demonstrate this via simulations. Fig. 4.16 clearly shows the actual DI is approaching the regular DI of order five. Fig. 4.15 shows the resulted beampattern. Fig. 4.17 shows the regular implementation of superdirective beamformer, which results in the nearly identical beampattern as Fig. 4.15. These simulations also demonstrates our algorithm can robustly reconfigure itself after microphone reorganization.

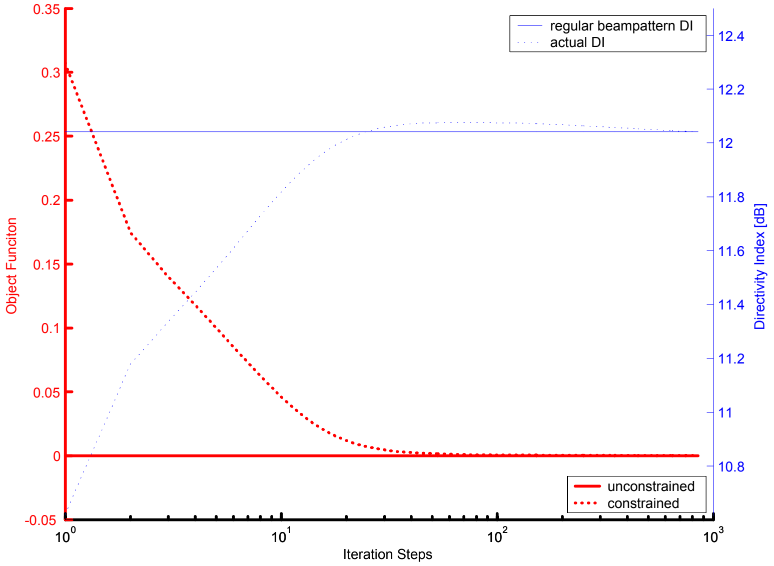


Figure 4.11: Iteration process for beampattern of order 3. The red (thick) curves use the left scale, blue (thin) curves right scale.

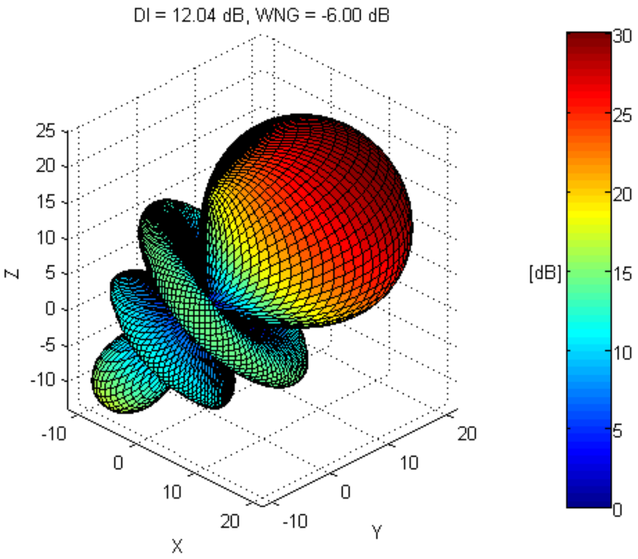


Figure 4.12: Constrained optimal beampattern of order 3.

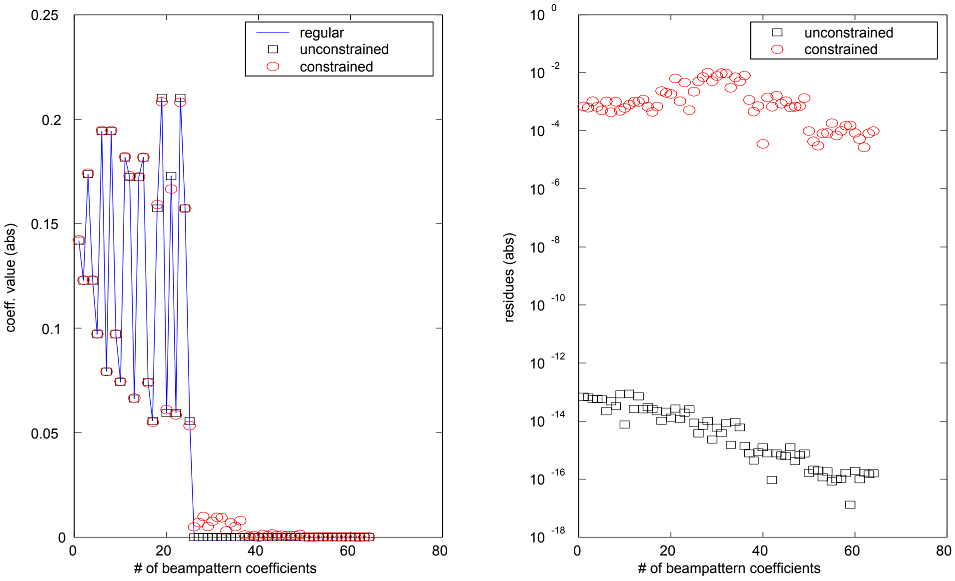


Figure 4.13: Precision comparisions for order 4 beamforming: (a) Comparison of unconstrained and constrained beampattern coefficients with regular order 4 beampattern coefficients $c_4\mathbf{B}_4$. (b) Residue comparison between unconstrained and constrained beampattern coefficients. Both plots show the absolute values.

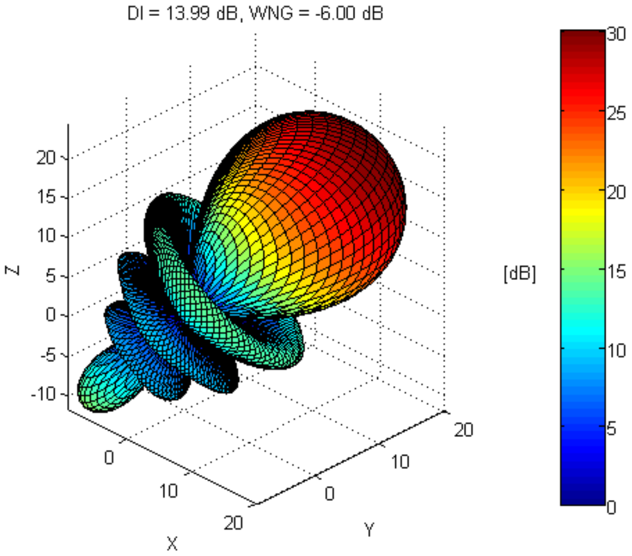


Figure 4.14: Constrained optimal beampattern of order 4.

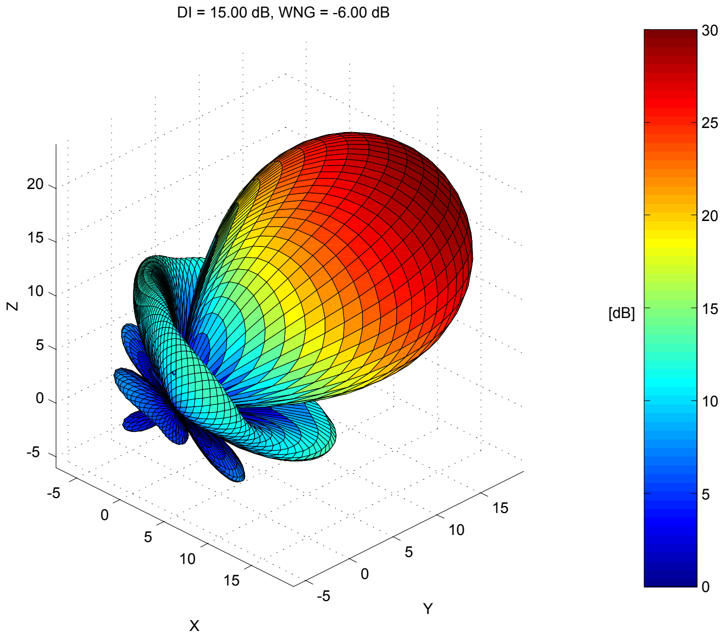


Figure 4.15: Constrained optimal beampattern of order 5. It is actually a superdirective beampattern.

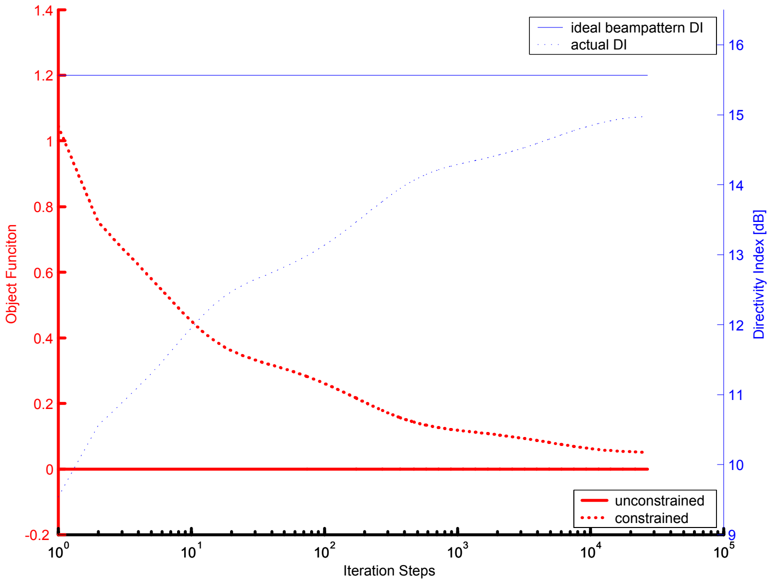


Figure 4.16: The iteration process optimally approximates the DI of the regular beampattern of order 5. The red (thick) curves use the left scale, blue (thin) curves right scale.

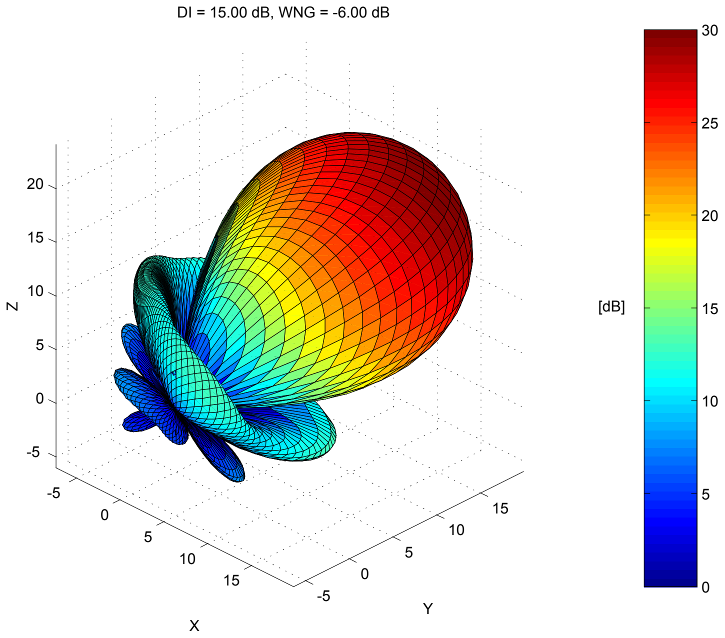


Figure 4.17: The beampattern of the regular implementation of superdirective beamformer.

4.6 Conclusions

This chapter describes a flexible and optimal design of spherical microphone arrays for beamforming. We analyzed the effects of discrete orthonormality error and proposed a constrained optimization approach to achieve optimal beamformer, either with regular or superdirective beampattern. An adaptive implementation is developed. Various design examples are presented. Experimental results will be presented in the next chapter.

Chapter 5
Hemispherical Microphone Arrays for Sound Capture and Beamforming[1]

In this chapter, we design and demonstrate a hemispherical microphone array for spatial sound acquisition and beamforming. Our design makes use of the acoustic image principle. It is especially appropriate for a half 3D acoustic environment where all sound sources are constrained on one side of a rigid plane. It avoids the difficulties of building a full spherical microphone array yet keeps the advantage of achieving a direction-invariant beampattern. A special microphone layout is designed for simple implementation. We also propose an approach to effectively calibrate data-independent coefficients of the system. Simulation and experimental results are presented. We also use the hemispherical array to verify the optimal design algorithms in Chapter 4. In addition, we present a pre-computed fast beamforming algorithm.

5.1 Introduction

Spherical microphone arrays are attracting increasing interest since they can capture a 3D soundfield and provide direction-invariant beampatterns in all directions [71][3]. In practice, using only a finite number of microphones, various layouts have been designed to optimally cancel the error caused by discreteness. The microphone positions can be either carefully selected to achieve optimal performances [77] or quite flexible with minimal performance compromise [61].

[1] This chapter is based on our original work in [59].

However, to physically build a full spherical microphone array on a rigid sphere is a challenging or impossible task. More importantly, in numerous real-world scenarios where sound sources are located in a constrained acoustic environment instead of a full 3D space, a full spherical array is either uneconomic or redundant. For example, in a conference room environment, all sound sources are usually above the table surface which forms a half 3D space. In this case, a hemispherical array may be a better choice because:

1. The table surface is usually rigid and inevitably creates acoustic images of the real sound sources, which validates the design of a hemispherical array.
2. Given a specified number of microphones and a sphere of given radius, a hemispherical array will have a denser microphone arrangement, thereby allowing for analysis of a wider frequency range. Even in an acoustic environment without image sources, using a hemispherical array mounted on a rigid plane to create images may be appropriate since it provides higher order beampatterns.
3. A hemispherical array is easier to build and maintain, it can be mounted on a rigid surface such as table surface or wall, and wires can be conveniently placed to the microphones.

Fig. 5.1 shows our prototype of a hemispherical array with 64 microphones installed on the surface of a half bowling ball. This chapter is organized into three parts. We first briefly review the theories of spherical beamformer and propose a methodology to design a hemispherical microphone array using the principle of

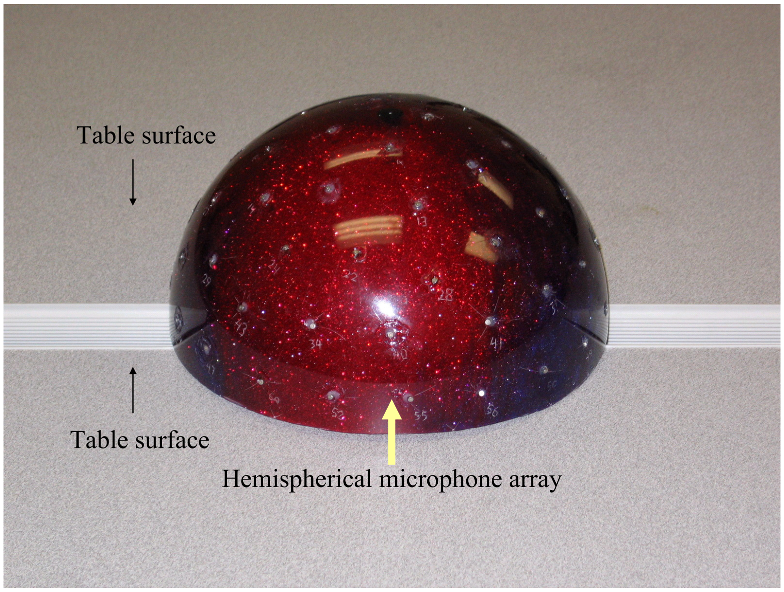


Figure 5.1: A hemispherical microphone array built on the surface of a half bowling ball. Its radius is 10.925cm.

images. Next, an effective calibration algorithm is proposed. Finally, our algorithms are demonstrated by simulation and experimental results using our prototype.

5.2 Design of A Hemispherical Microphone Array

In this section, we make use of the acoustic image principle to design a hemispherical microphone array.

5.2.1 Acoustic Image Principle

Suppose a sound source is placed on one side of a perfectly rigid plane, then in any point on the same side of the plane, the sound pressure is the combined result of unbounded irradiations of this sound source and the image sound source which lies on the other side of the plane symmetrically with respect to the plane. This is the *acoustic image principle* [74]. If we attach a rigid plane to the bottom of the

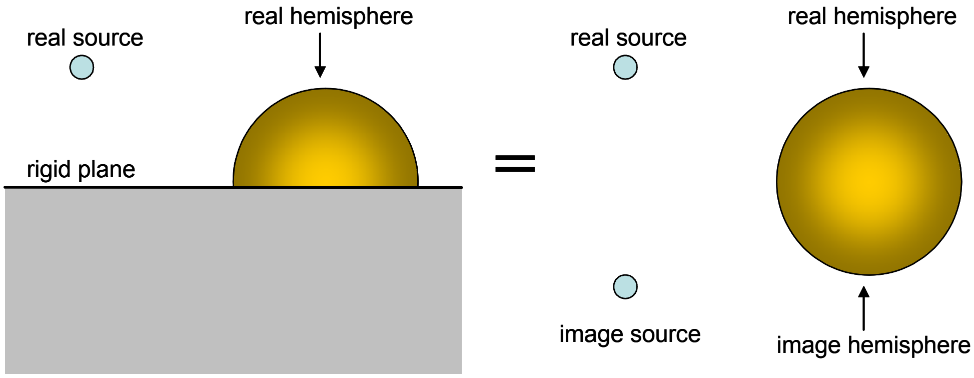


Figure 5.2: The hemispherical array with a rigid plane is equivalent to a spherical array in free space with real and image sources.

hemispherical array (see Fig. 5.2), the pressure at each real microphone position $\boldsymbol{\theta}_s = (\theta_s, \varphi_s)$ can be easily solved by removing the rigid plane and adding the image source at $\tilde{\boldsymbol{\theta}}_k = (\tilde{\theta}_k = \pi - \theta_k, \varphi_k)$ and the image hemispherical array. In addition, the image microphone at $\tilde{\boldsymbol{\theta}}_s = (\tilde{\theta}_s = \pi - \theta_s, \varphi_s)$ receives the same pressure as its counterpart. In other words, the rigid plane acts as an acoustic mirror. The solution on $\boldsymbol{\theta}_s$ is:

$$\psi_h(\boldsymbol{\theta}_s) = 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} A_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s), \tag{5.1}$$

$$A_n^m(\boldsymbol{\theta}_k) = Y_n^m(\boldsymbol{\theta}_k) + Y_n^m(\tilde{\boldsymbol{\theta}}_k). \tag{5.2}$$

5.2.2 A Symmetric and Uniform Layout

For a discrete hemispherical microphone array, the rigid plane creates a symmetric spherical layout of microphones with respect to the plane. Intuitively, we desire this symmetric layout also be “uniform" over the whole spherical surface, especially in the neighborhood of the rigid plane. Although the spherical layout can be made more flexible at the cost of extra computation [61], we want a layout with minimal implementation overhead.

To find a "uniform" layout, we use the simulation described in [36]. It minimizes the potential energy of a distribution of movable electrons on a perfectly conducting sphere. It is obvious that if the simulation starts with a symmetric initial layout, the resulting layout in each iteration step is guaranteed symmetric, so is the final layout. We use this approach to obtain a 128-node layout that is both symmetric and uniform. Specifically, to make it repeatable, we start with Fliege's 64-node layout [36]. We flip it upside down and add it on the original nodes to create 128 nodes, then perform the simulation until all nodes are optimally separated. The resulting layout is shown in Fig. 5.3. No nodes are too close to the $z = 0$ plane, or the rigid plane, which also helps avoid technical construction difficulties.

5.2.3 Discrete Hemispherical Beamforming

For a hemispherical array with S microphones uniformly mounted at $\boldsymbol{\theta}_s, s = 1, 2, ..., S$, the image microphones are at $\tilde{\boldsymbol{\theta}}_s$. To adapt the spherical beamforming principle as described in section 4.2 to the discrete case, the continuous integrals (4.4) are approximated by:

$$\frac{2\pi}{S}\sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s)Y_{n'}^{m'}(\boldsymbol{\theta}_s) + Y_n^{m*}(\tilde{\boldsymbol{\theta}}_s)Y_{n'}^{m'}(\tilde{\boldsymbol{\theta}}_s) \approx \delta_{nn'}\delta_{mm'}, \tag{5.3}$$

where $(n = 0, ..., N_{\max}; m = -n, ..., n;\ n' = 0, ..., N; m' = -n', ..., n')$, $N_{\max}$ is the band limit of spatial frequency in terms of spherical harmonics orders, N is the order of beamforming. In general, more precise approximation can be achieved by using appropriate cubature weights [36]. For simplicity, we just use equal cubature

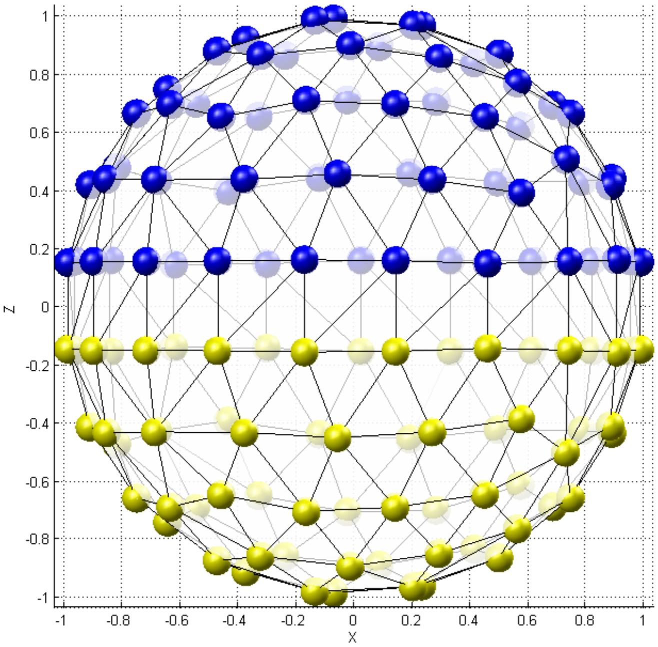


Figure 5.3: The symmetric and uniform layout of 128 nodes on a spherical surface. The blue (dark) nodes are for real microphones. The yellow (light) nodes are images.

weights without significantly affecting the results. To verify this, Fig. 5.4 shows the absolute errors of (5.3) using the layout in Fig. 5.3.

Therefore, to approximate the regular beampattern of order N as in (4.10), the weight for the s-th microphone is:

$$w_s = \sum_{n=0}^{N} \frac{\sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0)\left[Y_n^m(\boldsymbol{\theta}_s) + Y_n^m(\tilde{\boldsymbol{\theta}}_s)\right]}{2i^n b_n(ka)}. \tag{5.4}$$

5.3 Effective Calibration

In section 5.2, we derived the ideal beamformer from the ideal solution in (5.1). In practice, many factors will affect the complex pressure captured by the real-world hemispherical microphone array, such as table surface geometry and impedance,

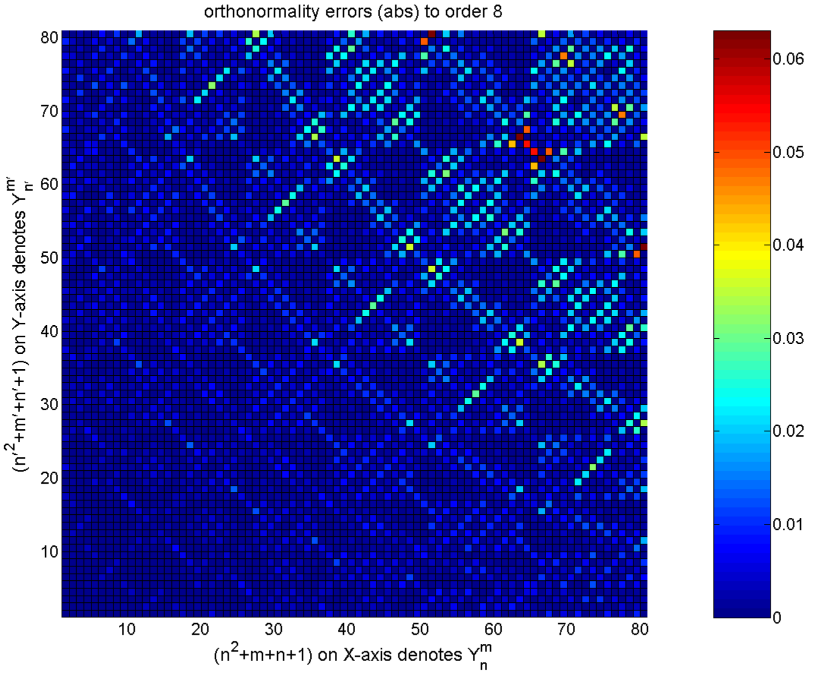


Figure 5.4: Discrete orthonormality errors. Plot shows absolute values.

array placement, microphone positions and characteristics, etc. To achieve better results, the array has to be calibrated to match the theoretical solution.

According to our system settings and the beamforming algorithm, a complete calibration is unnecessary for our prototype. Instead, by examining (5.4), we propose a simple yet effective calibration. We notice that $b_n(ka)$ describes the theoretical strength of the order n expansion in (5.1). The actual $b_n(ka)$ captured by the real-world microphone array has to match the theoretical value in (5.4) so that they can cancel each other to synthesize the desired 3D beampatterns, such as (4.10). This is especially important for high order beamforming since high order components of captured soundfield are increasingly weak because of the convergence of (5.1). This can be clearly observed in Fig. 4.1. Another advantage to calibrate $b_n(ka)$ lies in its independence of the incident waves in calibration, if they are known in advance. In this case, the calibration can be made in one measurement.

We describe the actual soundfield captured by our hemispherical array as:

$$\bar{\psi}_h = 4\pi \sum_{n=0}^{N_{\max}} i^n \bar{b}_n(ka) \sum_{m=-n}^{n} C_n^m(ka) Y_n^{m*}(\boldsymbol{\theta}_s) + \epsilon, \tag{5.5}$$

where $\bar{b}_n(ka)$ denotes the captured $b_n(ka)$. $C_n^m(ka)$ is the soundfield coefficient, which is known during calibration. For a single plane wave of unit amplitude incident from $\boldsymbol{\theta}_k$, in our hemispherical array setting, $C_n^m(ka) = A_n^m(\boldsymbol{\theta}_k)$ as in (5.2). ϵ is the residual error not included in $\bar{b}_n(ka)$. To calibrate $\bar{b}_n(ka)$ at order n', we assign one component from the hemispherical beamforming weight in (5.4) for the s-th microphone:

$$w_s^{n'} = \sum_{m'=-n'}^{n'} Y_{n'}^{m'*}(\boldsymbol{\theta}_0) \left[Y_{n'}^{m'}(\boldsymbol{\theta}_s) + Y_{n'}^{m'}(\tilde{\boldsymbol{\theta}}_s) \right], \tag{5.6}$$

and the total output is:

$$\begin{aligned} \bar{\Psi}^{n'} &= \frac{2\pi}{S} \sum_{s=1}^{S} w_s^{n'} \bar{\psi}_h. \\ &\approx 4\pi i^{n'} \bar{b}_{n'}(ka) \sum_{m'=-n'}^{n'} C_{n'}^{m'}(ka) Y_{n'}^{m'*}(\boldsymbol{\theta}_0), \end{aligned} \tag{5.7}$$

We then have:

$$\bar{b}_{n'}(ka) = \frac{\bar{\Psi}^{n'}}{4\pi i^{n'} \sum_{m'=-n'}^{n'} C_{n'}^{m'}(ka) Y_{n'}^{m'*}(\boldsymbol{\theta}_0)} + \varepsilon, \tag{5.8}$$

where ε contains ϵ and the orthonormality errors from (5.3). If ε is zero-mean Gaussian with respect to the beamforming direction $\boldsymbol{\theta}_0$, then we can estimate $\bar{b}_{n'}(ka)$ by averaging over every $\boldsymbol{\theta}_0$ with only one measurement if the calibration environment defined by $C_{n'}^{m'}(ka)$ can be precisely modeled.

5.4 Simulation and Experimental Results

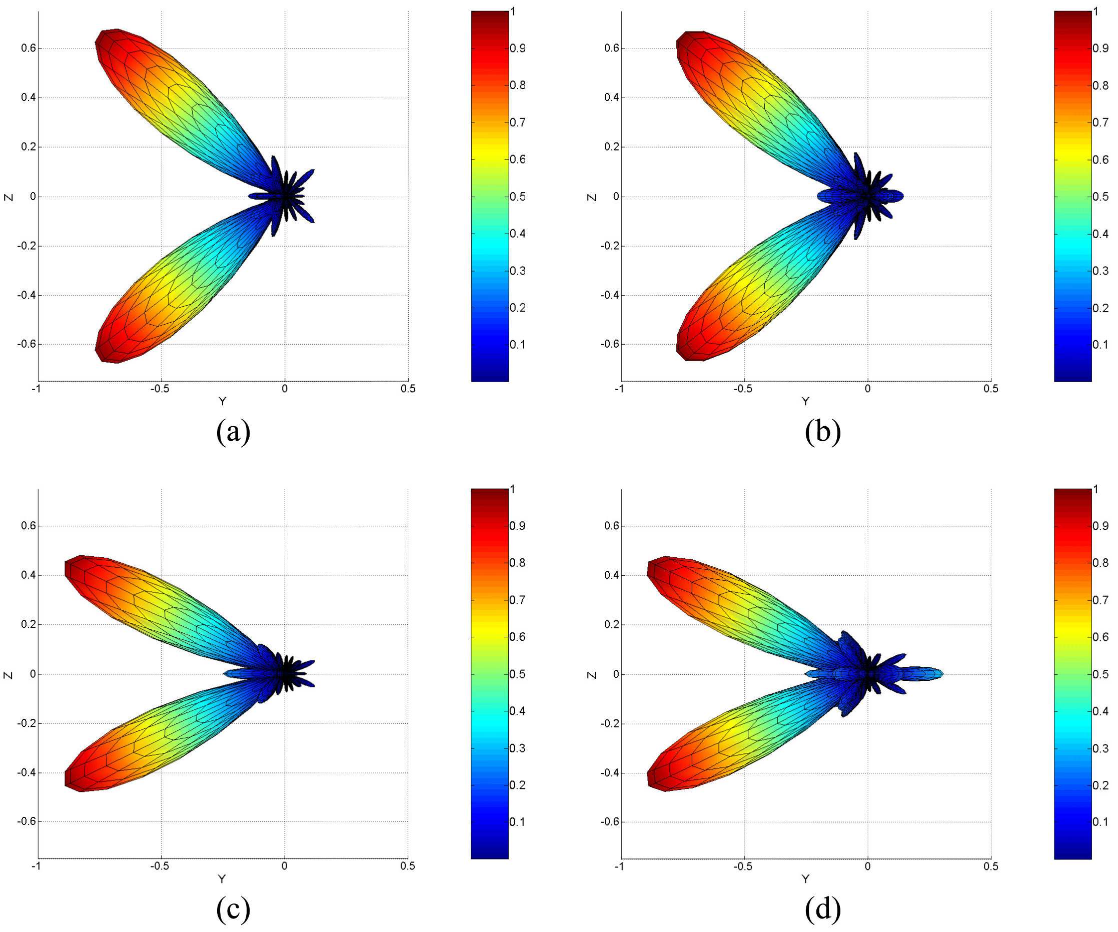
This section has two parts. First, we verify our hemispherical beamformer design. Second, we verify our flexible and optimal design described in Chapter 4.

5.4.1 Verification of Hemispherical Beamformer

We first show the simulation result. Suppose in free space, there are two plane waves of 4 kHz incident from $(\pi/4, -\pi/2)$ and $(3\pi/4, -\pi/2)$, respectively. Using the 128-node spherical microphone array, we scan every 3D direction using the beamformer of order eight. For each direction, we plot the amplitude of the output from this beamformer as a 3D point with the amplitude value as the distance to the origin. The 3D scanning result is shown in Fig. 5.5(a). Please do not confuse this with the beampattern of the 128-node array. In fact, this scanning result is the sum of the *two* 3D beampatterns of order eight, steered respectively to $(\pi/4, -\pi/2)$ and its image $(3\pi/4, -\pi/2)$. If the beamformer is steered to $(\pi/4, -\pi/2)$, the sound from $(3\pi/4, -\pi/2)$ will be significantly suppressed. The plot is shown in linear scale for clearer separation of two sources.

In calibration, we play the same sound from a real sound source from $(\pi/4, -\pi/2)$. The hemispherical array is set up on the table surface as shown in Fig. 5.1. We then use the approach in section 5.3 to calibrate. The 3D scanning result using the calibrated beamformer is shown in Fig. 5.5(b). Ideally, it should be the same as Fig. 5.5(a), but it also detected some reflections from other surfaces of the room, especially from the back direction $(\pi/2, \pi/2)$.

To demonstrate this calibration is independent of the sound source locations,



(a)

(b)

(c)

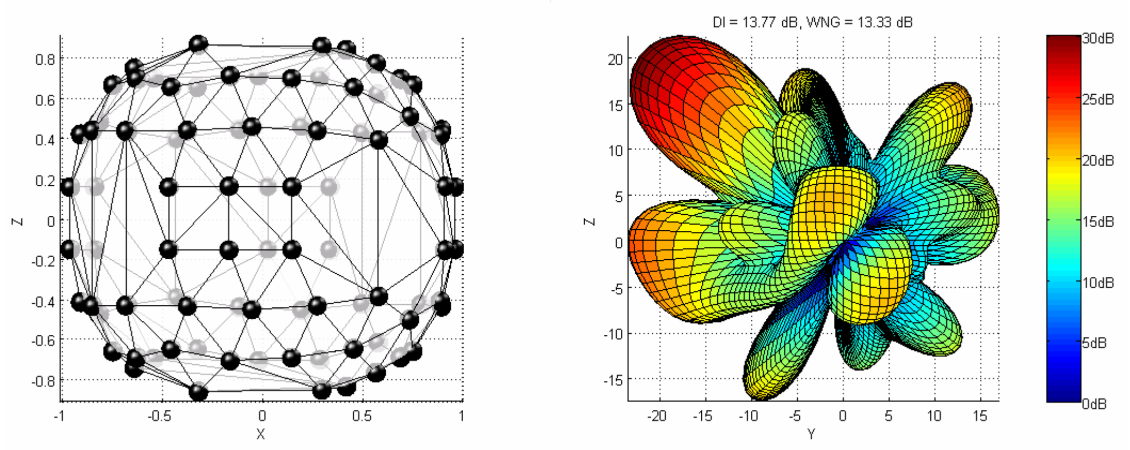
(d)

Figure 5.5: Simulation and Experimental results: (a) simulation of 3D scanning result with two sound sources, the beamformer is of order 8; (b) experimental result using the calibrated beamfomer of order 8; (c) simulation result after sound sources are moved; (d) experimental result using the same calibrated beamformer.

we move the real sound source to about $(\pi/3, -\pi/2)$. Fig. 5.5(c) shows the simulation result. In our experiment, we use the same calibrated beamformer to scan. The scanning result is shown in Fig. 5.5(d). As expected, the back reflection is stronger in this case.

5.4.2 Verification of Flexible and Optimal Spherical Array Design

We use the hemispherical array to verify our optimal design algorithm in Chapter 4. Instead of building another spherical microphone array with a non-uniform



(a) (b)

Figure 5.6: (a) The 88-node layout is generated by removing 40 symmetric nodes in the 128-node layout in Fig. 5.3. (b) The order 7 beampattern using equal quadrature weights at 2.5KHz.

layout, we simply block 20 channels in our 64-node hemispherical array. The resulting 44-node (88-node if including images) layout is shown in Fig. 5.6(a). As expected, the straightforward implementation of beamformer using equal quadrature weights has distorted beampatterns such as the one shown in Fig. 5.6(b). Please note that since the array layout is no longer uniform on the spherical surface, the beampattern will have different shapes in different steering directions. Therefore, the 3D localization result won't have the same shape as the beampattern in Fig. 5.6(b). Using the straightforward implementation, the simulated 3D localization result is shown in Fig. 5.7(a), and the experimental result is shown in Fig. 5.7(b). Using the simplified optimization algorithm in Chapter 4, without WNG constraint, the simulation result is shown in Fig. 5.8, which is the ideal result for our 44-node hemispherical array. In practice, the implementation must be under WNG constraint. Fig. 5.9 shows the experimental results of localization under dif-

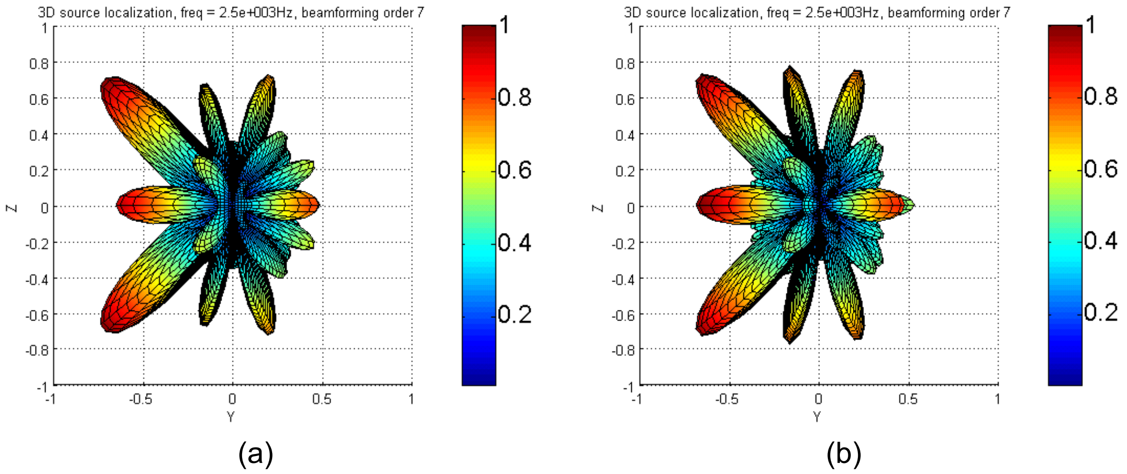


Figure 5.7: (a) The simulated source localization result using the array in Fig. 5.6. The beamformer at each direction is made with equal quadrature weights. (b) The experimental result using equal quadrature weights. Real sound source is from $(\pi/4, -\pi/2)$, at 2.5kHz.

ferent WNG constraints. It clearly shows the beamformer is less sensitive to noise with higher WNG.

5.4.3 Test in Real-World Auditory Scenes

We test the hemispherical microphone array in real-world auditory scenes. In the first experiment, a person stands in front of the hemispherical array about two meters away. He holds a loudspeaker and moves it periodically along the right-up-left-up-right path. The loudspeaker is playing music. Fig. 5.10 shows the tracking results. In the left of Fig. 5.10, it clearly shows when the real sound source is close to the table surface ($\theta = \pi/2$), because of the image sound source, the energy peaks tend to lie on the $\theta = \pi/2$ plane. In the right figure, the azimuth angles are not affected by the image sound source.

In the second experiment, we evaluate its noise reduction ability. We use two

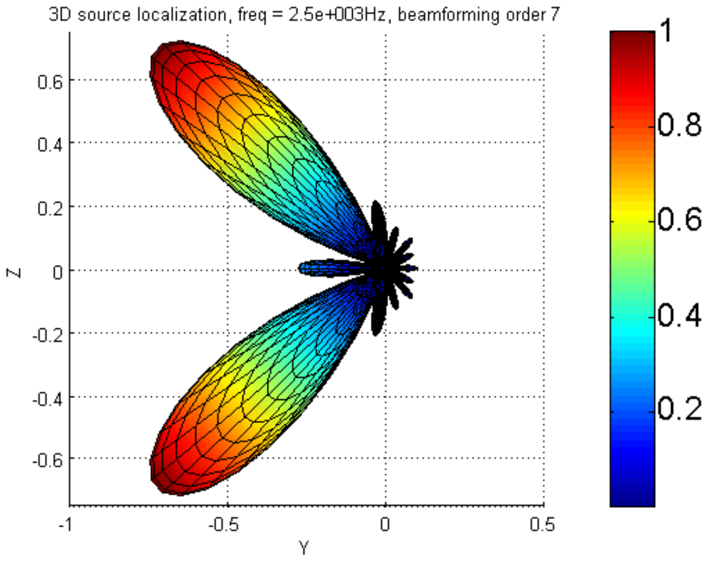


Figure 5.8: The simulated optimal localization result without WNG constraint.

loudspeakers facing the hemispherical array from $(1.075, 2.7594)$ and $(1.1456, -0.2408)$ respectively. The two loudspeakers are playing different sounds, one is music, the other speech. Using the hemispherical beamformer, the two sounds are separated. Fig. 5.11 shows the results.

5.5 Precomputed Fast Spherical Beamforming

Suppose we perform N-order spherical beamforming to L uniform 3D directions $(\boldsymbol{\theta}_i)$, $i = 1, ..., L$. Then, according to (4.10), for the ith beamformed result, the incident wave from $\boldsymbol{\theta}_k$ has the gain of

$$\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_k) Y_n^m(\boldsymbol{\theta}_i). \tag{5.9}$$

Here we omit the constant coefficient for clarity.

Now we want a new beamformed result at the direction $\boldsymbol{\theta}_j$, which is different from any $(\boldsymbol{\theta}_i)$, $i = 1, ..., L$. Apparently, the straightforward approach is to make a

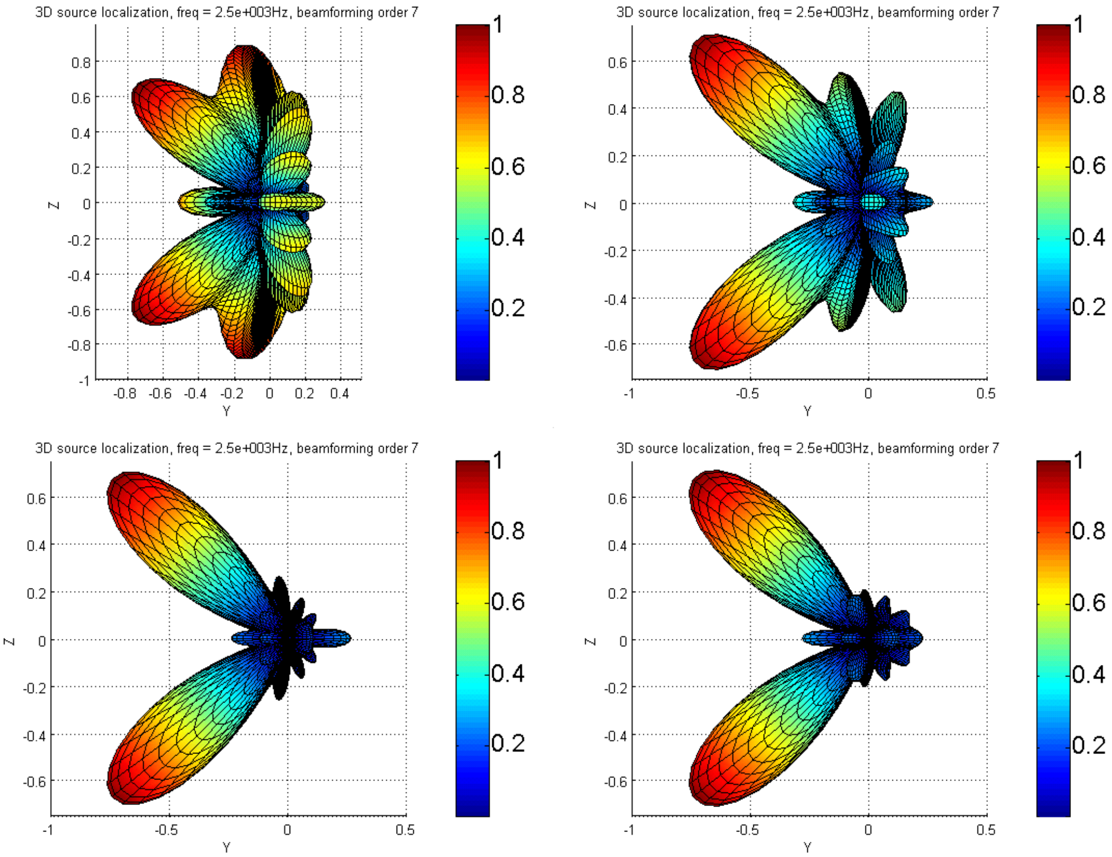


Figure 5.9: Localization results using optimal beamformers under different WNG constraints. Top left: $0dB < WNG < 5dB$. Top right: $5dB < WNG < 10dB$. Bottom left: $10dB < WNG < 15dB$. Bottom right: $WNG > 20dB$.

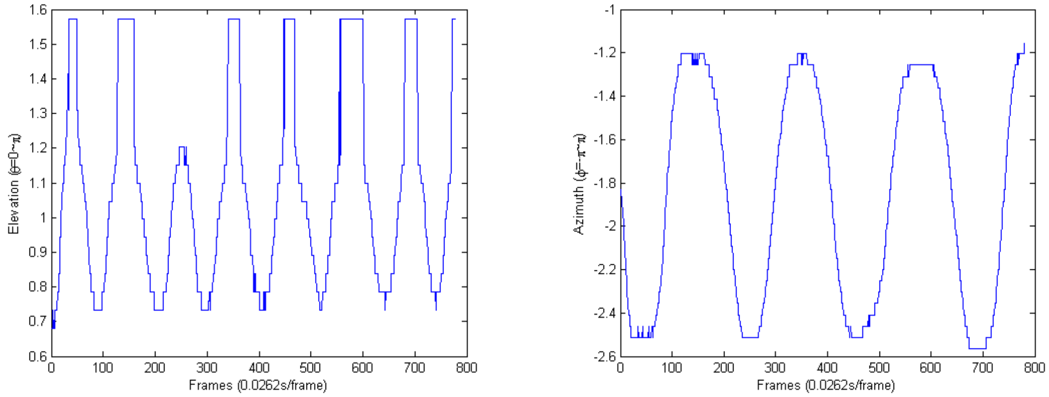


Figure 5.10: Moving sound source tracking. Plots show the azimuth and elevation angles of the energy peak at each frame. The tracking is performed in the frequency band from 3kHz to 6kHz, with beamformers of order seven (3kHz~4kHz) and eight (4kHz~6kHz).

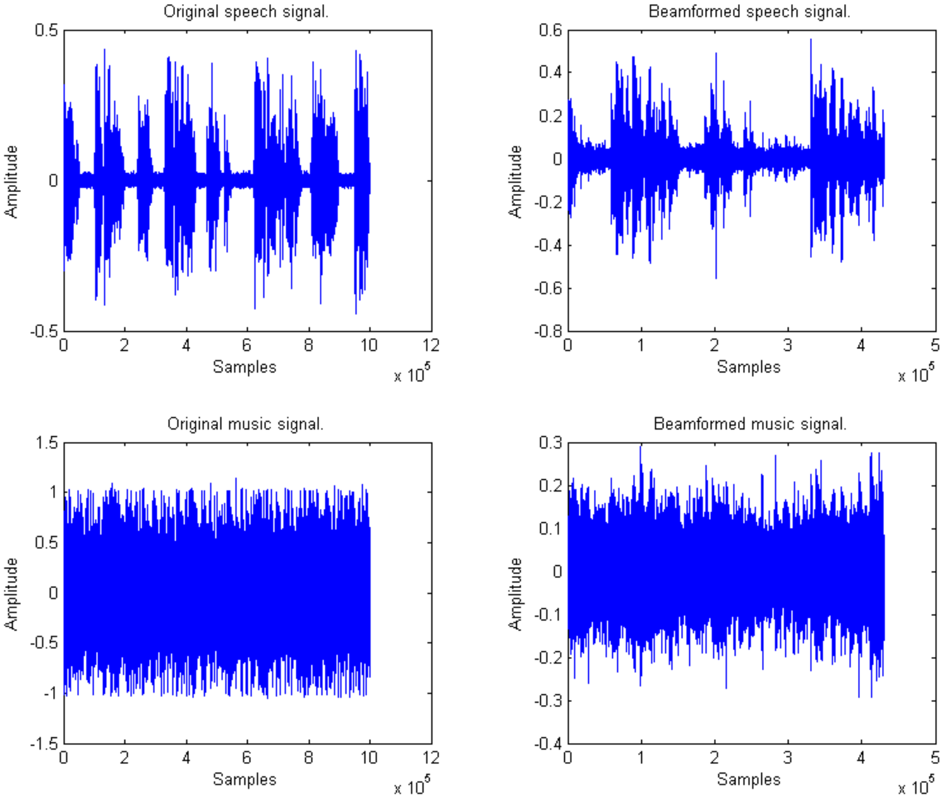


Figure 5.11: Sound separation using hemispherical beamformer. The signals in the right column correspond to segments in the original signals in the left column.

regular spherical beamforming at the direction $\boldsymbol{\theta}_j$. However, since we already have L beamformed results, we can design an efficient way to achieve this in time domain.

5.5.1 Time Domain Mixing via Orthonormal Decomposition

We notice that the following identity

$$\begin{aligned}
&\int_\Omega \left\{ \left[\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_k) Y_n^m(\boldsymbol{\theta}) \right] \times \left[\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m}(\boldsymbol{\theta}_j) Y_n^{m*}(\boldsymbol{\theta}) \right] \right\} d\Omega \\
= &\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_k) Y_n^m(\boldsymbol{\theta}_j) \qquad (5.10)
\end{aligned}$$

can be rewritten into a discrete approximation:

$$\begin{aligned}
&\sum_{i=1}^{L} \left\{ \left[\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_k) Y_n^m(\boldsymbol{\theta}_i) \right] \times \left[\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m}(\boldsymbol{\theta}_j) Y_n^{m*}(\boldsymbol{\theta}_i) \right] \right\} \\
\approx &\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_k) Y_n^m(\boldsymbol{\theta}_j). \qquad (5.11)
\end{aligned}$$

This clearly means if we apply the weight

$$W_i = \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m}(\boldsymbol{\theta}_j) Y_n^{m*}(\boldsymbol{\theta}_i) \qquad (5.12)$$

to the ith beamformed result, and sum all L results up, we get the new beamformed result as if we steer the beamformer to $\boldsymbol{\theta}_j$.

Now, the last confusion remains: if W_i is a complex number, that means for each frequency, the phase shift is the same. In another word, W_i is an all-pass filter but not in linear phase. That can't lead an efficient implementation in time domain.

Fortunately, it is easy to prove that W_i is always a real number using addition theorom of spherical harmonics:

$$\sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m}(\boldsymbol{\theta}_j) Y_n^{m*}(\boldsymbol{\theta}_i) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\cos\gamma), \qquad (5.13)$$

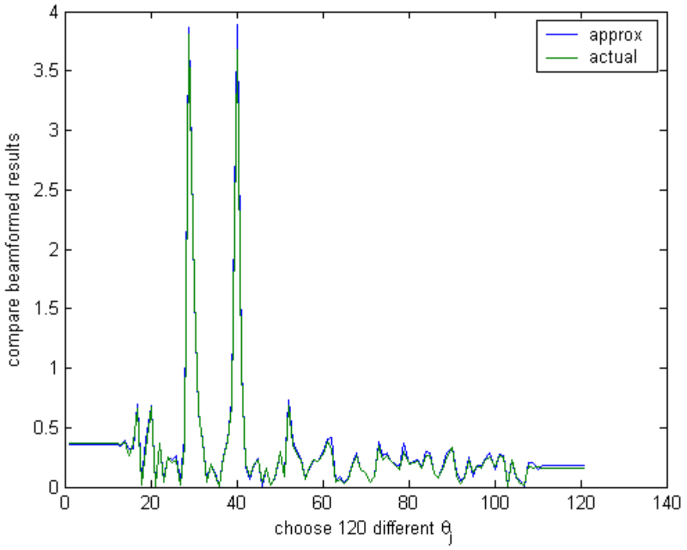


Figure 5.12: Gain comparision of direct beamforming and fast mixing in time domain. Plot shows the absolute value for each beamforming direction. Simulated in order eight with $\boldsymbol{\theta}_k = (\pi/4, \pi/4)$.

where γ is the angle between $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_i$.

Therefore, the fast spherical beamforming algorithm is:

1. Precompute the beamformed results in L uniform directions.

2. To fast beamform at a new direction $\boldsymbol{\theta}_j$, we multiply W_i to each precomputed channel in time domain and then sum them up.

5.5.2 Verification

To verify this algorithm, we use the 128 nodes in Fig. 5.3 as both microphone positions and beamforming directions. Using 120 equiangular $\boldsymbol{\theta}_j$ in 3D space, we compare the results of (5.10) and (5.11). Fig. 5.12 verifies its effectiveness.

5.6 Conclusions

We designed a hemispherical microphone array using the acoustic image principle. It can be seen as a symmetric spherical microphone array across the edge of real and image space, which is easy to build and install. It is especially appropriate in numerous scenarios where sound sources are constrained in a half 3D space. We built an array using 64 microphones positioned on a hemispherical surface. The layout of real and image microphones is both symmetric and uniform, which leads to a simple implementation of hemispherical beamformer. An effective calibration method is proposed to extract the data-independent coefficients of the system from the captured soundfield. The simulation and experimental results demonstrate the effectiveness of our design. In addition, we use the hemispherical array to verify the optimal algorithms described in Chapter 4 with real data. A precomputed fast spherical beamforming algorithm is also presented and verified.

Chapter 6
Cylindrical Microphone Arrays for Sound Capture and Beamforming[1]

The previous two chapters have established the theories for 3D audio system using spherical microphone array. However, sometimes we only desire the spatial filtering in one angular dimension, such as in surround sound field.

In this chapter, we propose a novel circular microphone array system for recording and spatial filtering high order surround sound field. In our array, omni-directional microphones are uniformly mounted along the cross-section circle on the side surface of a sound-rigid cylinder. We will use this array to record the scattered surround sound field, which can capture more spatial information than other designs without scatterer [58]. The spatial filtering, or beamforming, is based on the orthogonal decomposition of the recorded sound field by this array. Design example and performance analysis are presented to demonstrate its effectiveness.

The main advantages of our system are:

1. It is easier to build and more practical to use for surround auditory scenes, especially compared with the spherical array;

2. It is highly efficient and effective in recording and spatial filtering surround auditory scenes;

3. It is highly scalable.

In the rest of this chapter, we first introduce the theory of acoustic scattering

[1] This chapter is based on our original work in [62].

from a rigid cylinder. We then develop the theory of cylindrical beamforming in section 6.2. In addition, we explain why we need a rigid cylinder instead of using the microphone array in free space. To demonstrate our design, we analyze the performance of a practical example in section 6.3. We address the practical design issues such as discreteness, practical beampattern, robustness, and its performance in 3D sound field. Both theoretical analyses and simulation results are presented.

6.1 Scattering from a rigid cylinder

Considering a general scenario as shown on the left side in Fig. 6.1, where a unit magnitude plane sound wave of wave number $\mathbf{k}$ is incident from (θ_k, φ_k), an infinite length rigid cylinder of radius a is positioned vertically at the origin of the coordinate system, and the continuous circular microphone array is on the cross-section circle at $z = 0$ plane. The complex pressure at the observation point $\mathbf{r} = (r, \pi/2, \varphi)$ on the $z = 0$ plane is:

$$\psi_{in} = e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{6.1}$$

If $\mathbf{k}$ is perpendicular to the axis of the cylinder (z-axis), i.e. $\theta_k = \pi/2$, we then have:

$$\psi_{in} = e^{ikr\cos(\varphi-\varphi_k)}, \tag{6.2}$$

where $k = \|\mathbf{k}\|$. Now it is reduced to a 2D problem as shown on the right side in Fig. 6.1. We will come back to the 3D issues in section 6.3.4.

From section 3.6.1, this plane wave can be expressed in terms of cylindrical waves [74]:

$$\psi_{in} = \sum_{n=0}^{\infty} \epsilon_n i^n J_n(kr) \cos\left[n(\varphi - \varphi_k)\right], \tag{6.3}$$

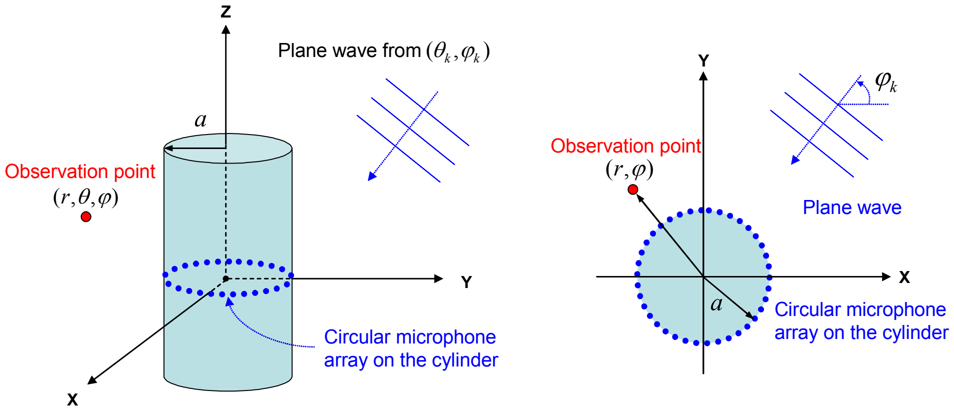


Figure 6.1: A plane wave incident on a rigid cylinder.

The total complex pressure at the circle on the cylinder surface is:

$$\begin{aligned}\psi(\varphi, \varphi_k) &= (\psi_{in} + \psi_{scat})|_{r=a} \\ &= \sum_{n=0}^{\infty} B_n(ka) \cos\left[n(\varphi - \varphi_k)\right], \end{aligned} \tag{6.4}$$

where:

$$B_n(ka) = \epsilon_n i^n \left[J_n(ka) - \frac{J_n'(ka) H_n(ka)}{H_n'(ka)} \right]. \tag{6.5}$$

6.2 Cylindrical Beamforming

The basic principle of cylindrical beamforming is to make orthogonal decompositions of the recorded surround sound field, and the decomposed components are combined to approximate a desired beampattern.

If we assign the weight $\frac{\cos(n'\varphi)}{\pi B_{n'}(ka)}$ to each continuous point on the circle and make an integral over φ, we have:

$$\begin{aligned}&\int_0^{2\pi} \psi \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi \\ &= \int_0^{2\pi} \sum_{n=0}^{\infty} B_n(ka) \cos\left[n(\varphi - \varphi_k)\right] \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi. \end{aligned} \tag{6.6}$$

Since

$$\cos\left[n(\varphi-\varphi_k)\right] = \cos(n\varphi)\cos(n\varphi_k) + \sin(n\varphi)\sin(n\varphi_k), \tag{6.7}$$

and using the following integral identities:

$$\begin{aligned}
\int_0^{2\pi} \cos(n\varphi)\cos(n'\varphi)d\varphi &= \pi\delta_{nn'}, && (6.8)\\
\int_0^{2\pi} \sin(n\varphi)\cos(n'\varphi)d\varphi &= 0, && (6.9)\\
\int_0^{2\pi} \cos(n'\varphi)d\varphi &= 0, && (6.10)
\end{aligned}$$

for $n, n' \neq 0$, where $\delta_{nn'}$ is the Kronecker delta function defined as:

$$\delta_{nn'} = \begin{cases} 1, & n = n' \\ 0, & n \neq n' \end{cases}. \tag{6.11}$$

We then have:

$$\int_0^{2\pi} \psi \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi = \cos(n'\varphi_k), \qquad (n' > 0), \tag{6.12}$$

That means the wave from φ_k direction has the gain of $\cos(n'\varphi_k)$.

Similarly, if the weight is $\frac{\sin(n'\varphi)}{\pi B_{n'}(ka)}$, we have:

$$\int_0^{2\pi} \psi \frac{\sin(n'\varphi)}{\pi B_{n'}(ka)} d\varphi = \sin(n'\varphi_k), \qquad (n' > 0), \tag{6.13}$$

In addition, when the weight is $\frac{1}{2\pi B_0(ka)}$, we have:

$$\begin{aligned}
&\int_0^{2\pi} \psi \frac{1}{2\pi B_0(ka)} d\varphi \\
=\ & \frac{1}{2\pi B_0(ka)} \left\{ \int_0^{2\pi} B_0(ka) d\varphi + \int_0^{2\pi} \sum_{n=1}^{\infty} B_n \cos\left[n(\varphi-\varphi_k)\right] d\varphi \right\} \\
=\ & 1. && (6.14)
\end{aligned}$$

If the ideal beampattern is a peak at φ_0 and zero everywhere else, it can be modeled as the delta function $\delta(\varphi - \varphi_0)$[2]. Its Fourier series expansion is:

$$\delta(\varphi - \varphi_0) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}[\cos(n\varphi_0)\cos(n\varphi) + \sin(n\varphi_0)\sin(n\varphi)]. \tag{6.15}$$

Therefore, to achieve the ideal beampattern $f(\varphi, \varphi_0) = \delta(\varphi - \varphi_0)$, the weight for the microphone at point φ is:

$$\begin{aligned} w(\varphi, \varphi_0) &= \frac{1}{4\pi^2 B_0(ka)} + \frac{1}{\pi}\sum_{n=1}^{\infty}\left[\cos(n\varphi_0)\frac{\cos(n\varphi)}{\pi B_n(ka)} + \sin(n\varphi_0)\frac{\sin(n\varphi)}{\pi B_n(ka)}\right] \\ &= \frac{1}{4\pi^2 B_0(ka)} + \frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{\cos[n(\varphi - \varphi_0)]}{B_n(ka)}. \end{aligned} \tag{6.16}$$

Some careful readers may have noticed the similarity between the fomulas (6.3) and (6.4) in the sense that both are the infinite series of cosine functions except the coefficients are different. We rewrite (6.3) as:

$$\psi_{in} = \sum_{n=0}^{\infty} A_n \cos\left[n(\varphi - \varphi_k)\right], \tag{6.17}$$

where

$$A_n = \epsilon_n i^n J_n(kr). \tag{6.18}$$

It is plausible that the cosine components recorded by a circular microphone array in free space can also be extracted similarly as we did in section 6.2. So why bother to scatter the plane wave by a rigid cylinder? The intuition is: the scattering makes the pressure distribution along the circular microphone array more complicated, especially for higher frequency wave, so more spatial information is captured. This is illustrated more clearly in Fig. 6.2 and Fig. 6.3. When ka is relatively small, the cylinder is approximately transparent to the sound wave, so $A_n(ka)$ and $B_n(ka)$

[2] It is also called "Dirac's delta function" or the "impulse symbol". Refer [74, pp. 31] for the formal definition. Don't be confused with the Kronecker delta function as (6.11).

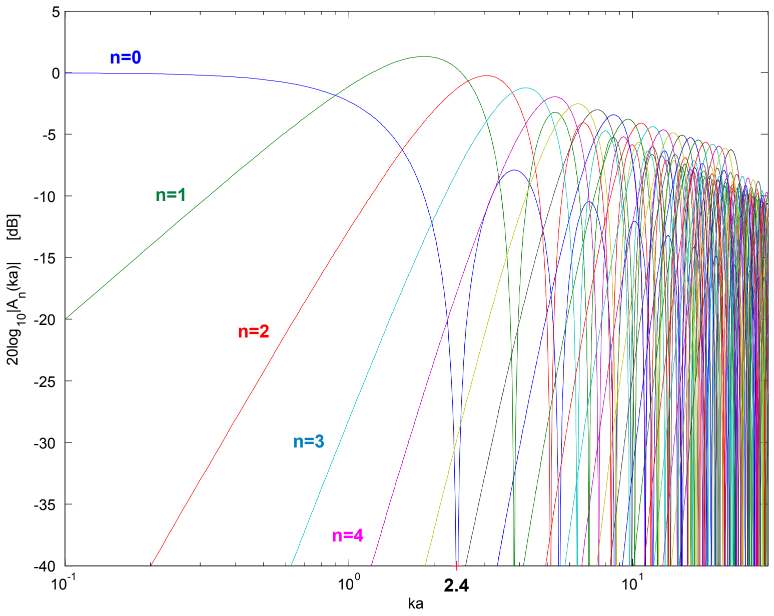


Figure 6.2: $A_n(ka)$ for n from 0 to 20.

have similar behaviors. When ka increases, the scattering effect is stronger, so the irregular pressure distribution along the circular array contains richer spatial frequency content. For example, in Fig. 6.2, when ka lies in the first notch of $A_0(ka)$ which is about 2.4, apparently the spatial frequency component of $n = 0$ cannot be extracted robustly since it is too weak compared to the $n = 1$ component and hidden by noise. While at the same ka in Fig. 6.3, the component of $n = 0$ is boosted to about -6 dB which is much easier to be extracted. There are more notches for larger ka in Fig. 6.2 although $A_n(ka)$ finally converges with respect to n for given ka. On the contrary, in Fig. 6.3, for fixed ka, $B_n(ka)$ exhibits smoother convergence with the increase of n.

In summary, with the help of the scattering from the cylinder, it is much easier to extract the desired spatial components.

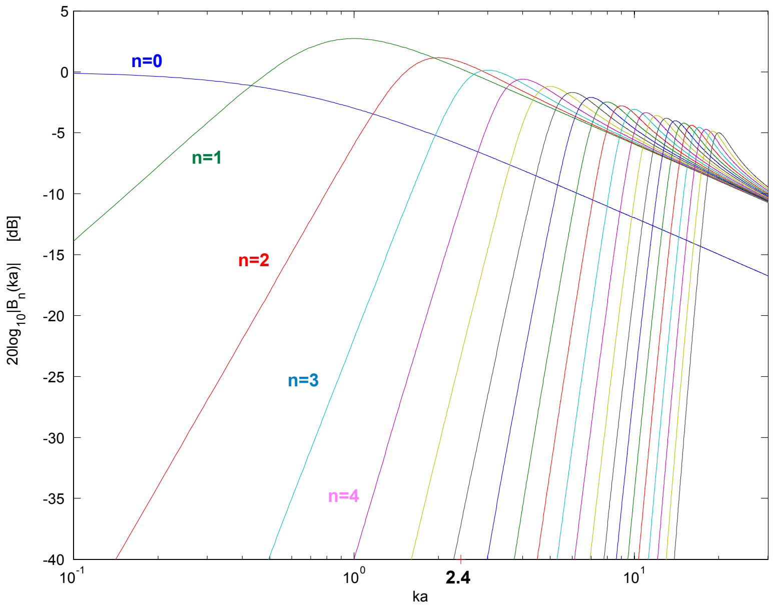


Figure 6.3: $B_n(ka)$ for n from 0 to 20.

6.3 Practical Design and Analysis

As an example, suppose we have a cylinder with radius of 10 cm, our array has 32 microphones positioned uniformly on the circle. We will analyze the performance of this system. Since we only care about the pressure on the microphones which is on the side surface of the cylinder, in practice, the cylinder can be of finite length and the microphone array can be placed away from both endpoints. The scattered sound field distortion is negligible at the microphone positions.

6.3.1 Discreteness

So far we have assumed a continuous circular microphone array. In practice, however, we have to use finite number of microphones positioned discretely on a circle. According to the well-known sampling theorem, to fully reconstruct a band-limited signal, it must be sampled at a rate at least twice as fast as the highest frequency. This sampling rate is called the *Nyquist frequency.* This theorem is for both spatial and temporal sampling.

In our system, we assume the temporal sampling rate is already fast enough. Our concern is the spatial sampling rate with respect to the number of microphones on a circle. Using N microphones positioned uniformly on the circle, the sampling rate is N. So we can fully capture the sound field with highest spatial frequency of $n = N/2$. For example, in Fig. 6.3, when $ka = 1$, the amplitude of the spatial frequency component of $n = 4$ is about -40 dB which is negligible in practice, so we may just say the highest spatial frequency is $n = 3$ and we need $N = 6$ microphones.

For the array with N microphones, if the highest spatial frequency with significant amplitude is $M(M \leq N/2)$, the weight is:

$$w_M(\varphi, \varphi_0) = \frac{1}{4\pi^2 B_0(ka)} + \frac{1}{\pi^2} \sum_{n=1}^{M} \frac{\cos[n(\varphi - \varphi_0)]}{B_n(ka)}. \tag{6.19}$$

With fixed a, the spatial frequency content depends on the temporal frequency. This is also illustrated in Fig. 6.3 when ka increases, the sound field along the circular array contains higher spatial frequencies. Spatial aliasing occurs when the sound field contains spatial frequency greater than $N/2$.

6.3.2 Practical Beampattern

The ideal beampattern is defined as (6.15). In the real array with finite number of microphones, using the practical weight as (6.19), the actual beampattern is an M-truncated version of (6.15):

$$f_M(\varphi, \varphi_0) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{M} \cos[n(\varphi - \varphi_0)]. \tag{6.20}$$

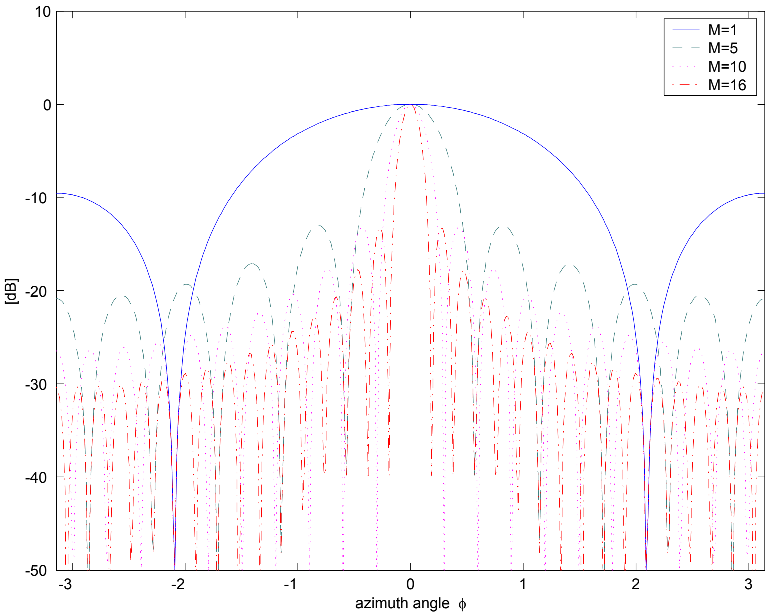


Figure 6.4: Beampatterns for $M = 1, 5, 10, 16$.

We say this beampattern is of order M. In Fig. 6.4, we plot a few beampatterns of different orders. To show the spatial aliasing effects, we compare the beampatterns at high frequencies against the theoretical beampattern of order 16, which is the maximum order for our example. The plot is for $\theta = \pi/2$ and $\varphi = [-\pi, \pi]$ as shown in Fig. 6.5.

6.3.3 Robustness

A robust beamformer requires the minimum *white noise gain* (WNG) is -6 dB, that is [16]:

$$WNG = 10\log_{10}\left(\frac{\left|\mathbf{W}^H\mathbf{d}\right|^2}{\mathbf{W}^H\mathbf{W}}\right) \geq -6 \text{ dB}, \tag{6.21}$$

where $\mathbf{W}$ is the weight vector applied to the circular microphone array to beamform at the direction $(\pi/2, 0)$:

$$\mathbf{W} = [w_M(\varphi_1, 0), w_M(\varphi_2, 0), ..., w_M(\varphi_N, 0)], \tag{6.22}$$

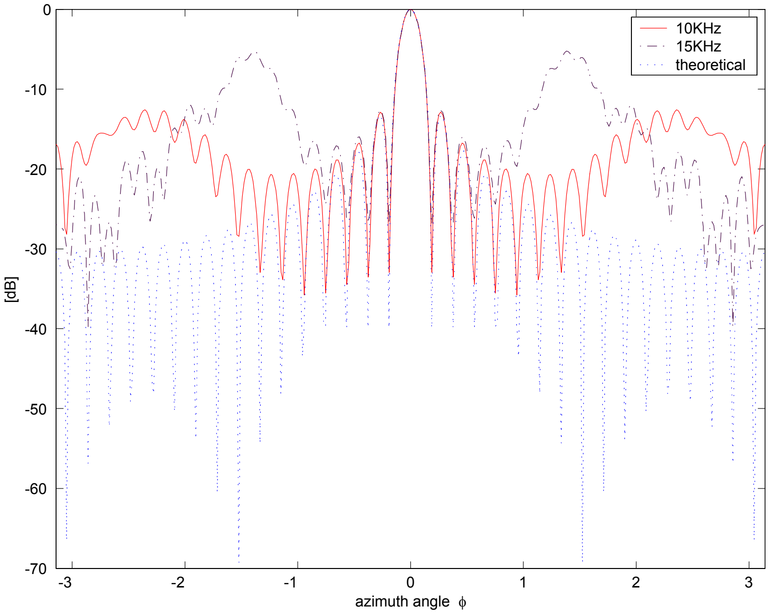


Figure 6.5: Spatial aliasing.

and $\mathbf{d}$ is the vector of the complex pressure produced at the array by a plane wave of unit magnitude incident from the same direction:

$$\mathbf{d} = [\psi(\varphi_1, 0), \psi(\varphi_2, 0), ..., \psi(\varphi_N, 0)]. \tag{6.23}$$

Here φ_i specifies the individual microphone position:

$$\varphi_i = 2\pi(i-1)/N, \quad i = 1, ..., N. \tag{6.24}$$

Fig. 6.6 shows the WNGs at different orders using our configuration. For example, when ka is in the range of $[0.13, 0.55]$, the beamformer can only work robustly at order 1; when ka is in the range of $[0.55, 1.14]$, the beamformer can work robustly at order 2, etc.

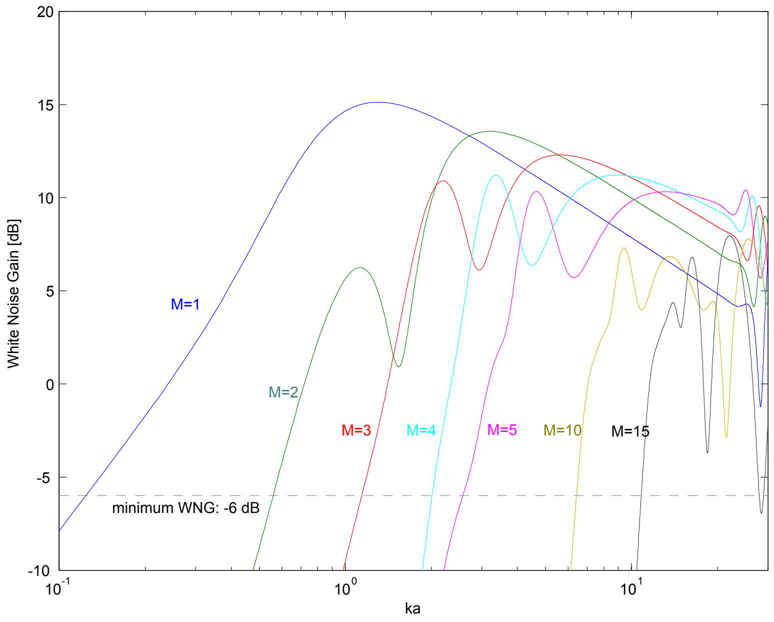


Figure 6.6: White noise gain.

6.3.4 Beampattern in 3D Sound Field

Our system is designed to make spatial filtering of two dimensional auditory scenes where sound sources are roughly on the same plane specified by the circular microphone array. In the real world, however, undesired sound can come from all 3D directions other than on this plane, such as the room reverberations or other noise.

In Fig. 6.7, we plot the polar beampatterns for different elevation angles. It clearly shows when the wave direction is moving away from the plane specified by the microphone array, the beampattern tends to be less directive. This can be explained theoretically. Suppose the sound wave is from (θ_k, φ_k) and the component of the wave number $\mathbf{k}$ perpendicular to the cylinder axis (z-axis) is $\mathbf{k}_{\perp}$, the incident

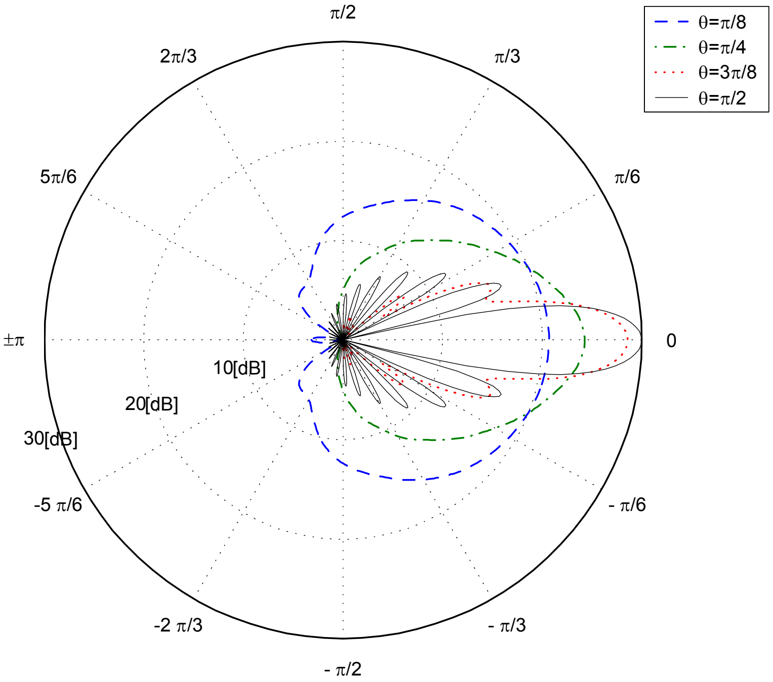


Figure 6.7: Polar Beampatterns for different elevation angles at 5 kHz. The beampattern for $\theta = \pi/2$ is of order 13 with $WNG \approx -5.96$ dB.

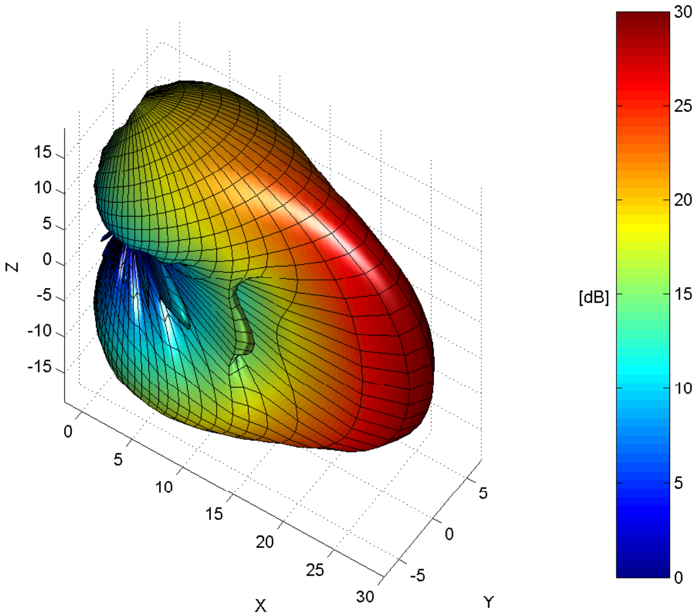


Figure 6.8: 3D beampattern of order 13 at 5 kHz, looking at $\varphi = 0$. (lighted to show 3D effect.)

wave is:

$$\psi_{in} = e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}_\perp\cdot\mathbf{r}} = e^{ik_\perp r\cos(\varphi-\varphi_k)}, \tag{6.25}$$

where:

$$k_\perp = k\sin\theta_k. \tag{6.26}$$

Accordingly, (6.12), (6.13) and (6.14) become:

$$\begin{aligned}
\int_0^{2\pi}\psi\frac{\cos(n'\varphi)}{\pi B_{n'}(ka)}d\varphi &= \frac{B_{n'}(k_\perp a)}{B_{n'}(ka)}\cos(n'\varphi_k), && (6.27)\\
\int_0^{2\pi}\psi\frac{\sin(n'\varphi)}{\pi B_{n'}(ka)}d\varphi &= \frac{B_{n'}(k_\perp a)}{B_{n'}(ka)}\sin(n'\varphi_k), && (6.28)\\
\int_0^{2\pi}\psi\frac{1}{2\pi B_0(ka)}d\varphi &= \frac{B_0(k_\perp a)}{B_0(ka)}, && (6.29)\\
(n' &> 0).
\end{aligned}$$

Therefore, the actual M-truncated beampattern for θ_k is:

$$f_M(\varphi,\varphi_0,\theta_k) = \frac{1}{2\pi}\frac{B_0(k_\perp a)}{B_0(ka)} + \frac{1}{\pi}\sum_{n=1}^{M}\frac{B_n(k_\perp a)}{B_n(ka)}\cos[n(\varphi-\varphi_0)]. \tag{6.30}$$

From Fig. 6.3, we notice that except $B_0(ka)$ is monotonically decreasing, each $B_n(ka)$ $(n>0)$ has only one peak. On the left side of its peak, it is monotonically increasing and on the right side monotonically decreasing[3]. Since $k_\perp a < ka$ for $\theta_k \neq \pi/2$, there exists a threshold integer m $(0<m<M)$ such that:

$$\begin{cases} \dfrac{B_n(k_\perp a)}{B_n(ka)} \geq 1, & 0 \leq n \leq m \\ \dfrac{B_n(k_\perp a)}{B_n(ka)} < 1, & m < n \leq M \end{cases}. \tag{6.31}$$

That means the higher spatial frequency components are suppressed while the lower spatial frequency components are amplified. Apparently, this low-pass filter makes the resulted beampattern less directive.

[3] Strict theoretical proof omitted here.

The full 3D beampattern is shown in Fig. 6.8. It has a strawberry-like shape. Therefore, our array can also be used in 3D auditory scenes where spatial filtering in the azimuthal dimension is desired.

Chapter 7
Recreate 3D Auditory Scenes Via Reciprocity[1]

In this chapter, we develop the theories of building a high order 3D audio capture and recreation system. As we have discussed in Section 1.3, the previous designs are either unsuitable for 3D immersive applications [9], or unstable in complex auditory scenes [82]. Our contribution is to use the reciprocity between capturing and recreating to design a simple, robust and unified 3D audio system in high orders of spherical harmonics. We will apply this approach in loudspeaker array-based and headphone-based systems.

7.1 Loudspeaker Array-Based System

We use a microphone array mounted on a rigid sphere for the capture and a spherical loudspeaker array in free space for playback. A scenario is shown in Fig. 7.1. A design example and simulation results are presented to demonstrate the effectiveness of our system.

The main advantages of our system are:

1. the reciprocity between capturing and recreating processes makes the system easy to build;

2. the performance is robust and optimal with the given number of microphones and loudspeakers;

[1] This chapter is based on our original work in [65].

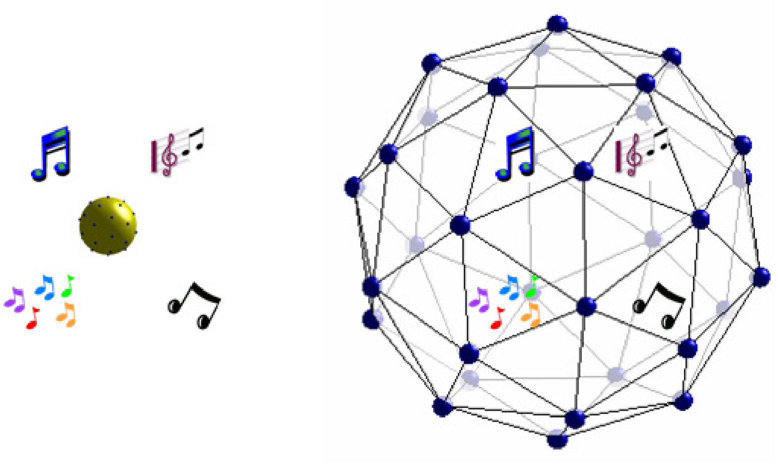


Figure 7.1: Left: the microphone array captures a 3D sound field remotely. Right: the loudspeaker array recreates the captured sound field to achieve the same sound field.

3. the capturing part is compact and portable which is convenient to record the immersive 3D sound field;

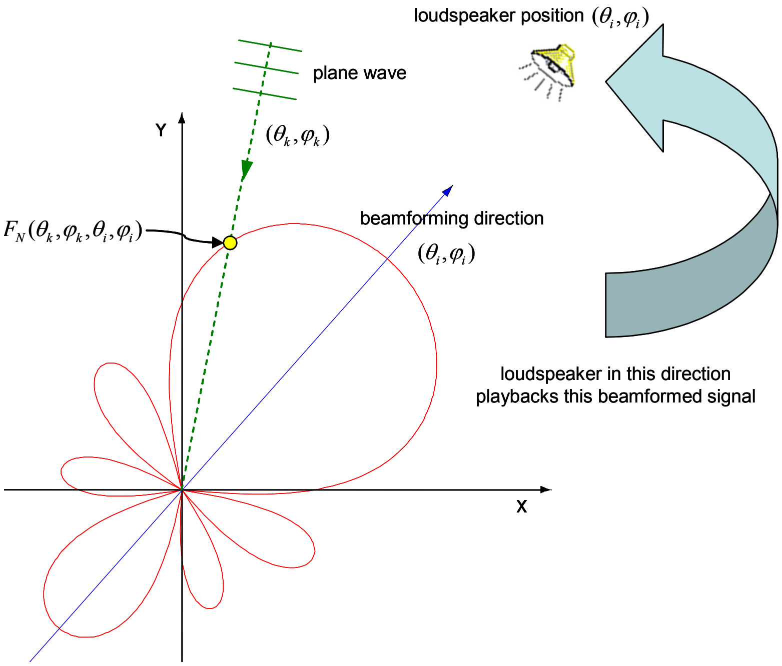
4. the system is highly scalable.

In addition, for some regular or semi-regular microphone layouts, there exists efficient parallel implementations of the multi-directional spherical beamformer. We will illustrate this in Section 7.1.5.

Our system can be seen as an extension to higher orders of the Ambisonics system, which only captures and recreates the 3D sound field to the first order of spherical harmonics [45][1].

7.1.1 Recording as 3D Sound Field Projection

We present a concept of viewing the spherical beamforming as a form of projection which leads to the design of the proposed system.



loudspeaker position (θ_i,φ_i)
plane wave
Y
(θ_k,φ_k)
$F_N(\theta_k,\varphi_k,\theta_i,\varphi_i)$
beamforming direction
(θ_i,φ_i)
loudspeaker in this direction
playbacks this beamformed signal
X

Figure 7.2: System workflow: for each chosen direction $\boldsymbol{\theta}_i = (\theta_i, \varphi_i)$, we first beamform the 3D sound field into that direction using our N-order beamformer, then simply playback the resulted signal from the loudspeaker in that direction.

In a real-world system, we only have a finite number of microphones distributed on the spherical surface discretely. With S microphones the following discrete version of (4.4) can be approximately satisfied:

$$\frac{4\pi}{S}\sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) = \delta_{nn'}\delta_{mm'}, \tag{7.1}$$

where the equation holds to order N. The achieved beampattern of order N is a truncated version of (4.7):

$$F_N(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \quad = \quad 2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} \; Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^{m}(\boldsymbol{\theta}). \tag{7.2}$$

Each beamformed signal not only picks up the sound from the desired direction $(\boldsymbol{\theta}_0)$, but also from other directions. We assume the 3D sound field is composed of plane waves. The beamforming can be viewed as a *projection* process. This is illustrated in Fig. 7.2, where it is shown in 2D for clarity. In this example, the beamformer ($N = 3$) is pointing to ($\boldsymbol{\theta}_i$). We can think of the plane wave from ($\boldsymbol{\theta}_k$) as weighted by

$$F_N(\boldsymbol{\theta}_k, \boldsymbol{\theta}_i) = 2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_i) Y_n^m(\boldsymbol{\theta}_k) \tag{7.3}$$

before it is projected into the ($\boldsymbol{\theta}_i$) direction. We view this projected plane wave as a *subsource*.

Suppose we beamform the recorded signals in D directions. This creates D subsources for the 3D sound field.

7.1.2 Recreation as 3D Sound Field Unprojection

As shown in Fig. 7.2, if the D beamforming directions are well chosen, we can simply playback the D subsources from their respective directions. We wish to recreate the original 3D sound field to order N.

In this section, we first derive a theoretical condition for achieving the recreation. We then find an approximate and optimal solution that provides us a simple and unified way to recreate the 3D sound field from the subsources.

Theoretical Condition for Recreation

We assume all loudspeakers are positioned in free space and arranged as a spherical array with a radius large enough to produce plane waves at the observation points (this will be relaxed in the next section). Each loudspeaker plays back the identical signal weighted by a complex coefficient $a_l(k)$. The resulted sound field inside this sphere is:

$$\psi_c = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr_s) \sum_{m=-n}^{n} \sum_{l=1}^{L} a_l(k) Y_n^m(\boldsymbol{\theta}_l) Y_n^{m*}(\boldsymbol{\theta}_s), \tag{7.4}$$

where $(\boldsymbol{\theta}_l), l = 1, ..., L$ are the angular positions of L loudspeakers and $(\boldsymbol{\theta}_s, r_s)$ is the observation point inside the sphere.

To recreate the 3D sound field exactly, we let (3.60) equal (7.4):

$$\psi_{in} = \psi_c \tag{7.5}$$

Let Ω_s be the spherical surface of radius r_s. Here we only require r_s to be less than the radius of the loudspeaker array and satisfy the plane wave assumption, not necessarily equal to the radius of the microphone array. We have:

$$\int_{\Omega_s} \psi_{in} Y_{n'}^{m'}(\boldsymbol{\theta}_s) d\Omega_s = \int_{\Omega_s} \psi_c Y_{n'}^{m'}(\boldsymbol{\theta}_s) d\Omega_s. \tag{7.6}$$

Using (4.4), we get:

$$Y_{n'}^{m'}(\boldsymbol{\theta}_k) = \sum_{l=1}^{L} a_l(k) Y_{n'}^{m'}(\boldsymbol{\theta}_l), \tag{7.7}$$

$$(n' = 0, ..., \infty, \quad m' = -n', ..., n').$$

Please note (7.7) is independent of the observation point. To recreate the 3D sound field to order N, we need (adapted from [88]):

$$\mathbf{Pa} = c\mathbf{b} \tag{7.8}$$

where c is a constant, $\mathbf{a}$ is the vector of unknown weights to be assigned to each loudspeaker;

$$\mathbf{P} = \begin{bmatrix} Y_0^0(\boldsymbol{\theta}_1) & Y_0^0(\boldsymbol{\theta}_2) & \cdots & Y_0^0(\boldsymbol{\theta}_L) \\ Y_1^{-1}(\boldsymbol{\theta}_1) & Y_1^{-1}(\boldsymbol{\theta}_2) & \cdots & Y_1^{-1}(\boldsymbol{\theta}_L) \\ Y_1^0(\boldsymbol{\theta}_1) & Y_1^0(\boldsymbol{\theta}_2) & \cdots & Y_1^0(\boldsymbol{\theta}_L) \\ Y_1^1(\boldsymbol{\theta}_1) & Y_1^1(\boldsymbol{\theta}_2) & \cdots & Y_1^1(\boldsymbol{\theta}_L) \\ \vdots & \vdots & \ddots & \vdots \\ Y_N^{-N}(\boldsymbol{\theta}_1) & Y_N^{-N}(\boldsymbol{\theta}_2) & \cdots & Y_N^{-N}(\boldsymbol{\theta}_L) \\ \vdots & \vdots & \ddots & \vdots \\ Y_N^N(\boldsymbol{\theta}_1) & Y_N^N(\boldsymbol{\theta}_2) & \cdots & Y_N^N(\boldsymbol{\theta}_L) \end{bmatrix}; \tag{7.9}$$

and

$$\mathbf{b} = \begin{bmatrix} Y_0^0(\boldsymbol{\theta}_k) \\ Y_1^{-1}(\boldsymbol{\theta}_k) \\ Y_1^0(\boldsymbol{\theta}_k) \\ Y_1^1(\boldsymbol{\theta}_k) \\ \vdots \\ Y_N^{-N}(\boldsymbol{\theta}_k) \\ \vdots \\ Y_N^N(\boldsymbol{\theta}_k) \end{bmatrix}. \tag{7.10}$$

Here $(\boldsymbol{\theta}_k)$ is the direction from which the original plane wave is incident.

Reproduction as the Reciprocal of Beamforming

If we choose the beamforming directions $(\boldsymbol{\theta}_i)$ and loudspeaker angular positions $(\boldsymbol{\theta}_l)$ to all be the same L angular positions, in such a way that the discrete orthonormality of spherical harmonics can be satisfied to order N, we have:

$$\frac{4\pi}{L}\sum_{i=1}^{L} F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_i)Y_{n'}^{m'}(\boldsymbol{\theta}_i) = Y_{n'}^{m'}(\boldsymbol{\theta}_k). \tag{7.11}$$

So we have the solution of $\mathbf{a}$ as:

$$\mathbf{a} = \begin{bmatrix} F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_1) \\ F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_2) \\ \vdots \\ F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_L) \end{bmatrix}. \tag{7.12}$$

This clearly shows that each loudspeaker plays back exactly the beamformed signal with the beampattern pointing to itself as shown in Fig. 7.2.

7.1.3 Extension to Point-Source Form

Thus far, we have assumed the loudspeakers are far enough to generate plane waves. In this section, we extend the algorithms to the general point-source model.

If the loudspeakers can be modeled as point sources, the soundfield recreated by the spherical loudspeaker array with radius r becomes [23]:

$$\psi_c = 4\pi\sum_{n=0}^{\infty} i^n j_n(kr_s)R_n(kr)\sum_{m=-n}^{n}\sum_{l=1}^{L} a_l(k)Y_n^m(\boldsymbol{\theta}_l)Y_n^{m*}(\boldsymbol{\theta}_s),$$

where

$$R_n(kr) = -ikre^{ikr}i^{-n}h_n(kr). \tag{7.13}$$

The condition for recreation, (7.7), then becomes:

$$Y_{n'}^{m'}(\boldsymbol{\theta}_k) = R_{n'}(kr)\sum_{l=1}^{L} a_l(k)Y_{n'}^{m'}(\boldsymbol{\theta}_l), \tag{7.14}$$

$$(n' = 0, ..., \infty, \quad m' = -n', ..., n').$$

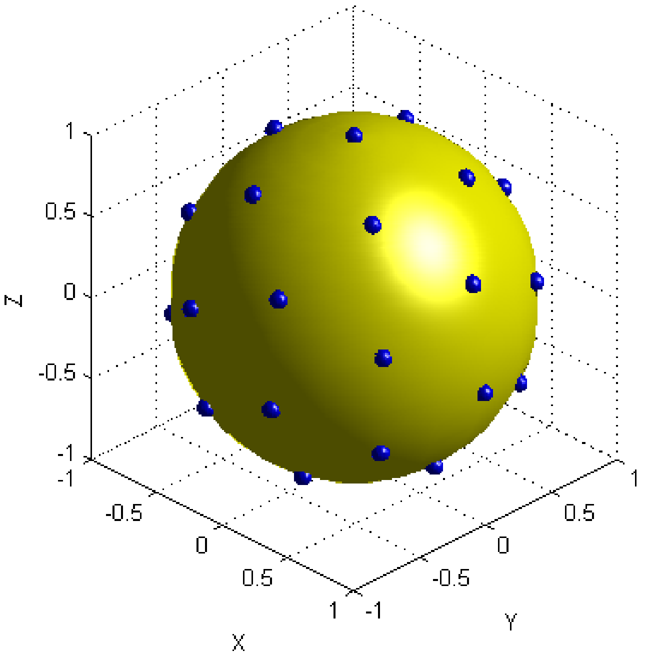


Figure 7.3: The layout of 32 microphones on a rigid sphere with radius of 5 cm. (The sphere is lighted to show the 3D effect, NOT the sound scattering.)

To have an optimal solution to order N, we have to modify our spherical beamformer. Specifically, we normalize the expansion of the beampattern (7.3) to:

$$\tilde{F}_N(\boldsymbol{\theta}_k, \boldsymbol{\theta}_i) = 2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_i) \frac{Y_n^m(\boldsymbol{\theta}_k)}{R_n(kr)}. \tag{7.15}$$

The solution of (7.14) is then:

$$\mathbf{a} = \begin{bmatrix} \tilde{F}_N(\boldsymbol{\theta}_k, \boldsymbol{\theta}_1) \\ \tilde{F}_N(\boldsymbol{\theta}_k, \boldsymbol{\theta}_2) \\ \vdots \\ \tilde{F}_N(\boldsymbol{\theta}_k, \boldsymbol{\theta}_L) \end{bmatrix}. \tag{7.16}$$

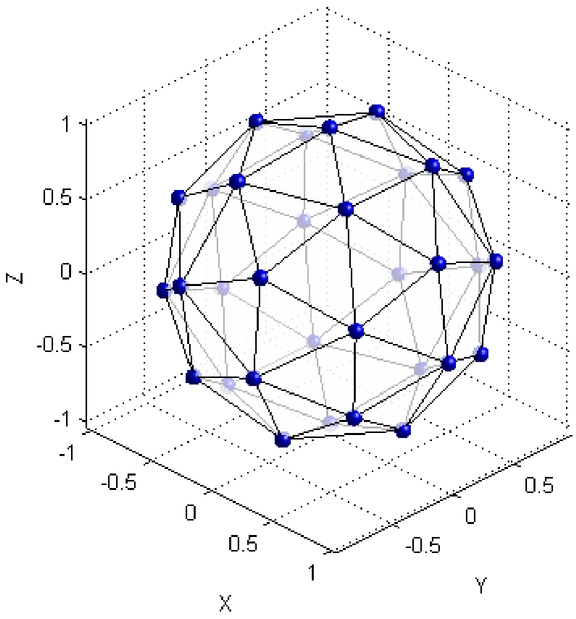


Figure 7.4: The layout of 32 loudspeakers in free space arranged as a sphere of large radius.

7.1.4 Design Example and Simulations

To start our design, we need a layout of discrete points on a spherical surface which satisfy (7.1) to some order. The angular positions specified by the centers of the faces of a truncated icosahedron can satisfy (7.1) to order 4 according to [71]. Other layouts can be used also such as in [35][50][63]. We just use these 32 angular positions to design our system. Our microphone array is as shown in Fig. 7.3 where the 32 omnidirectional microphones are positioned on a rigid sphere of radius 5 cm. Our loudspeaker array is as shown in Fig. 7.4 with the same 32 angular positions. We capture the 3D sound field using the microphone array, and then use the (normalized) spherical beamformer to steer the signals to these 32 directions. Finally, we recreate the 3D sound field using the loudspeaker array.

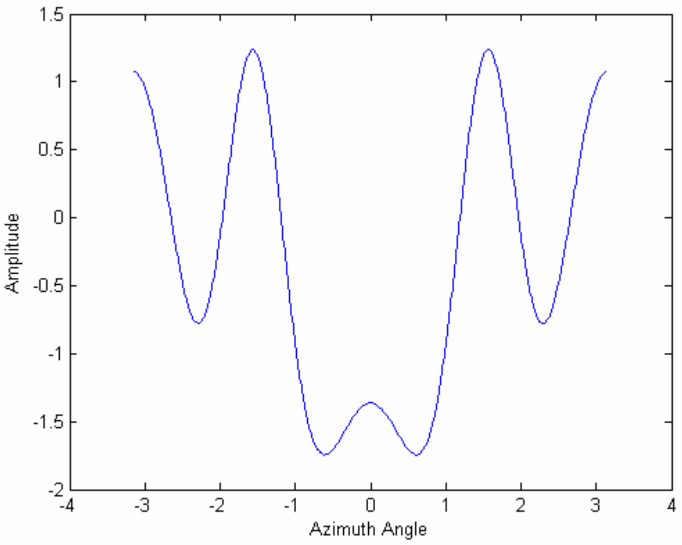


Figure 7.5: The plane wave of 4 kHz incident from positive direction of X-axis scattered by the microphone array. Plot shows the pressure on the equator.

We simulate our system by a plane wave of 4 kHz incident from the positive direction of X-axis in Fig. 7.3. Fig. 7.5 shows the scattered sound field on the equator of the spherical microphone array. Fig. 7.6 is the recreated plane wave to order 4 on the $z = 0$ plane.

Please note that a spherical microphone array can also be rotated arbitrarily without affecting the results. Because of the reciprocity between the beamformer and the loudspeaker arrays, the beamforming directions and the loudspeaker arrays must have the same angular positions. However, there may exist efficient parallel implementation of the multi-directional beamformer if the microphone array locations can be made consistent with the beamforming directions in some way. We will illustrate this in Section 7.1.5.

When we only decompose the 3D sound field to the first order of spherical

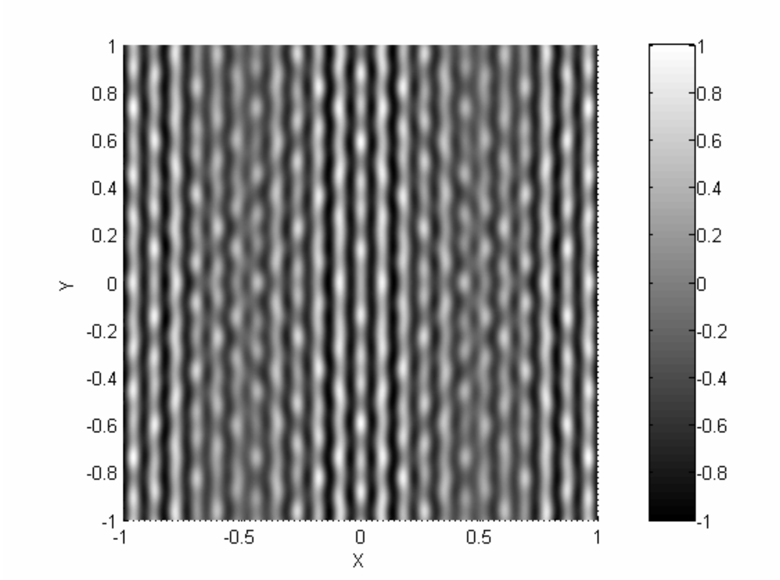


Figure 7.6: The recreated plane wave to order 4. Plot shows the $2 \times 2\,\mathrm{m}^2$ area on the $z = 0$ plane.

harmonics, our system mimics the Ambisonics system by pointing the bidirectional beampattern (the first order spherical harmonics) to the left-right, front-back, and up-down along with the omnidirectional beampattern (the zeroth order spherical harmonics) [45][1].

Like most microphone arrays that sample the 3D space discretely, the spherical microphone array has a spatial aliasing problem [71]. Low frequency requires larger sphere to decompose higher orders of spherical harmonics while high frequency requires denser microphones to avoid spatial aliasing. To broaden the frequency band, we may need a larger sphere with more microphones. Another solutions may be to use nested spherical microphone arrays, or multiple arrays.

7.1.5 Efficient Multi-directional Beamformer

If the layout of microphones is regular or semi-regular, we can make use of rotational symmetries of the microphone array and of the beampattern to design an efficient parallel multi-directional spherical beamformer. For specificity, we consider a regular icosahedron to illustrate the idea. It can be easily adapted to other regular or semi-regular microphone layouts.

A regular icosahedron, as shown in Fig. 7.7, is rotationally symmetric around any line connecting the origin and an arbitrary node. If the symmetric axis is as shown in the figure, there are two groups of nodes $\{2, 3, 4, 5, 6\}$ and $\{7, 8, 9, 10, 11\}$ on two different cones along with the two endpoints $\{1\}$ and $\{12\}$ of the line.

The rotational symmetry of the beampattern is formulated as the spherical harmonic addition theorem. Let γ be the angle between the two spherical coordinates $(\boldsymbol{\theta}_0)$ and $(\boldsymbol{\theta})$. We have:

$$\sin\theta \sin\theta_0 \cos(\varphi - \varphi_0) + \cos\theta \cos\theta_0 = \cos\gamma, \tag{7.17}$$

which defines a cone with respect to $(\boldsymbol{\theta})$. According to the spherical harmonic addition theorem, we have:

$$\sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}) = \frac{2n+1}{4\pi} P_n(\cos\gamma), \tag{7.18}$$

where $P_n(x)$ is the Legendre polynomial of degree n. So the value of the beampattern on the cone is defined by (7.17) as a function of γ:

$$2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}) = \sum_{n=0}^{N} \frac{2n+1}{2} P_n(\cos\gamma). \tag{7.19}$$

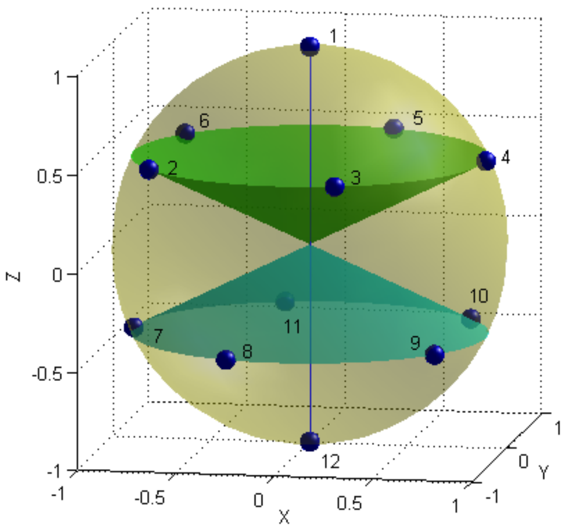


Figure 7.7: The rotational symmetry of icosahedron.

From (4.8), the weight for each microphone at $(\boldsymbol{\theta}_s)$ is:

$$f(\boldsymbol{\theta}_s) = f(\gamma_j) = \sum_{n=0}^{N} \frac{2n+1}{8\pi i^n b_n(ka)} P_n(\cos \gamma_j), \qquad (7.20)$$

$$(s = 1, ...12; \quad j = 1, 2, 3, 4).$$

For each beamforming direction specified by the spherical coordinates of the icosahedron's nodes, there are only four γ values for the icosahedron layout: $\{0, 1.1071, \pi - 1.1071, \pi\}$. For example, if the beampattern is pointing to node 1, node 1 has $\gamma_1 = 0$ and nodes 2, 3, 4, 5 and 6 have $\gamma_2 = 1.1071$, etc.; if the beampattern is pointing to node 2, node 2 has $\gamma_1 = 0$ and nodes 1, 3, 6, 7 and 8 have $\gamma_2 = 1.1071$, etc.. Thus, to build a twelve-directional beamformer, for each microphone, we only need to make four multiplications corresponding to four γ values instead of twelve. The structure of the multi-directional spherical beamformer is shown in Fig. 7.8.

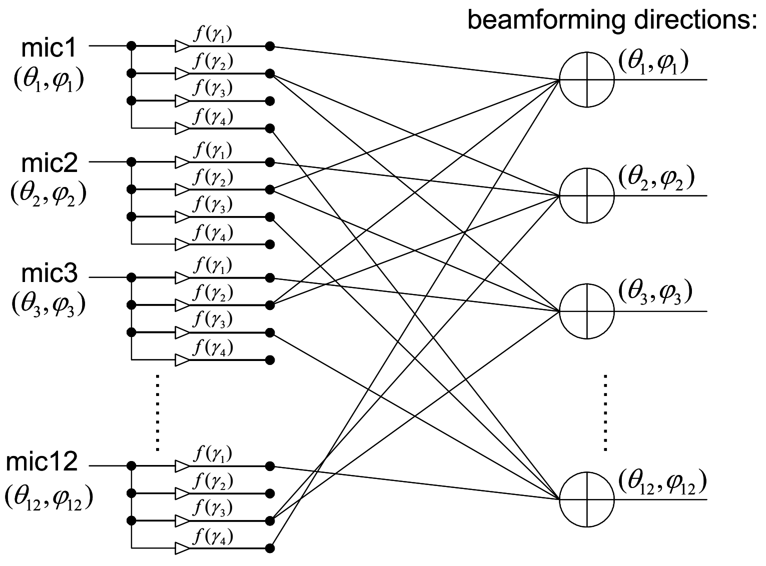


Figure 7.8: The efficient structure of the multi-directional spherical beamformer.

7.2 Headphone-Based System

In this section, we develop a coupled theory of reproducing *real* 3D auditory scenes by headphone from 3D recordings by spherical microphone array through HRTF approximation. The currently available HRTF databases are described in [2][6]. For simple auditory scene with only a few sound sources, we can first use the spherical beamformer to locate the sound sources, then filter the beamformed signals with the HRTFs in those corresponding directions. For complex auditory scenes, however, we need algorithm which is independent with the source locations.

7.2.1 Ideal HRTF Selection

In an ideal case, we assume the HRTF is already measured continuously on the spherical surface of radius r. Our goal is to select the correct HRTF for a specified direction. Although it seems trivial for an ideal case, we use this as a starting point and extend it to more practical cases in the following sections.

We drop the arguments k and r for simplicity, the HRTF for the sound of wave number k from the point $(r, \boldsymbol{\theta})$ is [31]:

$$\psi(\boldsymbol{\theta}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\boldsymbol{\theta}), \tag{7.21}$$

where h_n and Y_n^m have the same definitions as in the last section, and α_{nm} are the fitting coefficients which can be determined using real-world discrete HRTF measurements [31].

Suppose we want to select the HRTF for the direction $(\boldsymbol{\theta}_k)$, we apply the following delta function (ideal beampattern) to each measured HRTF:

$$F(\boldsymbol{\theta}, \varphi) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k), \tag{7.22}$$

we have:

$$\int_{\Omega_s} \psi(\boldsymbol{\theta}) F(\boldsymbol{\theta}) d\Omega_s = \psi(\boldsymbol{\theta}_k), \tag{7.23}$$

where Ω_s is the spherical surface. Obviously, the delta function simply selects the value we need and discards everything else.

To present another viewpoint of HRTF selection, we rewrite (7.23) into a more "complicated" form by using (7.21) and (4.7):

$$\begin{aligned}
&\int\limits_{\Omega_s} \left[\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\boldsymbol{\theta}) \right] \times \left[2\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta}_k) \right] d\Omega_s \\
&= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\boldsymbol{\theta}_k).
\end{aligned} \tag{7.24}$$

Alternatively, this can be easily proven by using the orthonormalities of spherical harmonics (4.4).

7.2.2 HRTF Approximation In Orthogonal Beam-Space

In practice, however, HRTFs are measured on discrete points. In this case, (7.24) and (4.4) can only hold approximately and in finite orders. In addition, using a practical spherical array with finite number of microphones, the beampattern is (4.10).

The HRTF for the sound of wave number k from the measurement point $(r, \boldsymbol{\theta}_i)$ is:

$$\psi(\boldsymbol{\theta}_i) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\boldsymbol{\theta}_i), \tag{7.25}$$

$$(\ i = 1, ..., B \ )$$

where B is the number of HRTF measurements.

The weighted combination of HRTFs then becomes:

$$\sum_{i=1}^{B} \psi(\boldsymbol{\theta}_i) F_N(\boldsymbol{\theta}_k, \boldsymbol{\theta}_i). \tag{7.26}$$

If the HRTF measurement points $(\boldsymbol{\theta}_i), i = 1, ...B$, are approximately uniformly distributed on a spherical surface so that the orthonormality of spherical harmonics holds up to order N', then the HRTF can be expanded into two groups:

$$\psi(\boldsymbol{\theta}_i) = \psi_0^{N'}(\boldsymbol{\theta}_i) + \psi_{N'+1}^{\infty}(\boldsymbol{\theta}_i), \tag{7.27}$$

where

$$\psi_0^{N'}(\boldsymbol{\theta}_i) = \sum_{n=0}^{N'} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\boldsymbol{\theta}_i), \tag{7.28}$$

$$\psi_{N'+1}^{\infty}(\boldsymbol{\theta}_i) = \sum_{n=N'+1}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\boldsymbol{\theta}_i). \tag{7.29}$$

So (7.26) can be rewritten as:

$$\sum_{i=1}^{B}\left[\psi_0^{N'}(\boldsymbol{\theta}_i)+\psi_{N'+1}^{\infty}(\boldsymbol{\theta}_i)\right]F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_i)$$

$$=\sum_{i=1}^{B}\psi_0^{N'}(\boldsymbol{\theta}_i)F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_i) \tag{7.30}$$

$$+\sum_{i=1}^{B}\psi_{N'+1}^{\infty}(\boldsymbol{\theta}_i)F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_i) \tag{7.31}$$

$$=\psi_0^{\min(N',N)}(\boldsymbol{\theta}_k)+\epsilon \tag{7.32}$$

which is the approximation of HRTF up to the order $\min(N', N)$. Here the error ϵ consists of two parts: one is the orthonormality error from (7.30) which is supposed to be small according to the discrete orthonormalities; the other is from (7.31) which is also small with well-chosen N' because of the convergence of the series expansion in (7.25).

If HRTF are not measured on uniformly distributed angular points, which is the case for all currently available measurements, we can first obtain a uniform version via interpolation [31]. In addition, we can interpolate as many points as we want to satisfy the orthonormality to arbitrarily high orders. In this case, the HRTF approximation at $(\boldsymbol{\theta}_k)$ depends only on the order of beampattern N, which is:

$$\sum_{i=1}^{B}\psi(\boldsymbol{\theta}_i)F_N(\boldsymbol{\theta}_k,\boldsymbol{\theta}_i)=\psi_0^{N}(\boldsymbol{\theta}_k)+\epsilon. \tag{7.33}$$

Therefore, apparently if there is a plane wave incident from $(\boldsymbol{\theta}_k)$ in the original auditory scene, it will be automatically filtered with the corresponding HRTF, in the approximation of order N.

7.2.3 Recreation Algorithm

Suppose we have built a spherical microphone array to record a 3D auditory scene. The spherical beamformer for this array has the beampattern as in (4.10). To reproduce the 3D auditory scene from the recordings, there are three steps:

1. beamform the recordings to $(\boldsymbol{\theta}_i)$ for $i = 1, ..., B$ (the "uniformly" interpolated point);

2. filter the beamformed signal at $(\boldsymbol{\theta}_i)$ with the personalized HRTF $\psi(\boldsymbol{\theta}_i)$ for $i = 1, ..., B$;

3. superimpose the resulted signals for $i = 1, ..., B$.

With the currently available HRTF measurements, the only factor that determines reproduction quality is the beampattern order N of the spherical microphone array.

7.2.4 Design Examples

We use the KEMAR HRTF measurements [2] to demonstrate our design. In Fig. 7.9, the red (solid) line shows the HRTF measurement at the position just in front of the manikin. The green (dot) line shows the approximation to order five supposing we have a spherical microphone array of order five. It is a good approximation for frequencies until about 2KHz. It is also a relatively close approximation until 4KHz which may be used in spatial speech acquisition and reproduction. The blue (dash) line shows the approximation to order 10, which closely matches the measurement until about 6KHz. The phases are compared in Fig. 7.10.

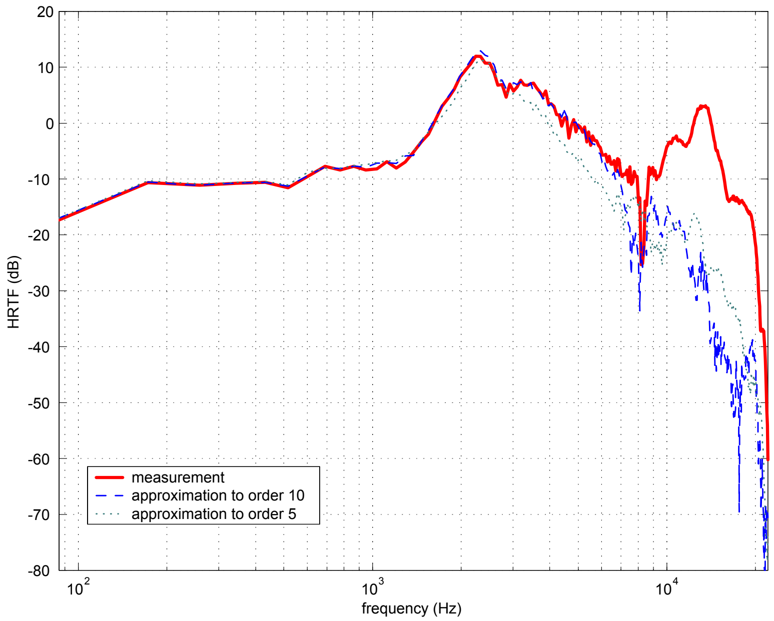


Figure 7.9: HRTF approximations to orders 5 and 10. Plot shows the magnitude in dB scale.

For efficient implementation in practice, the beamformer should be in different orders for different frequency bands.

7.2.5 Summary

In summary, we have developed the theory of reproducing 3D auditory scene using headphone from recordings of a spherical microphone array. We made use of the spherical microphone array since it provides a natural way to decompose the 3D soundfield in orthogonal beam space which will be used to approximate the HRTF measurements by using the orthonormalities of spherical harmonics. The advantage of our method lies in its independence of the sound source locations and the surrounding environment, only if under the far-field assumption. Preliminary design examples are presented to justify our approach. Future work may include

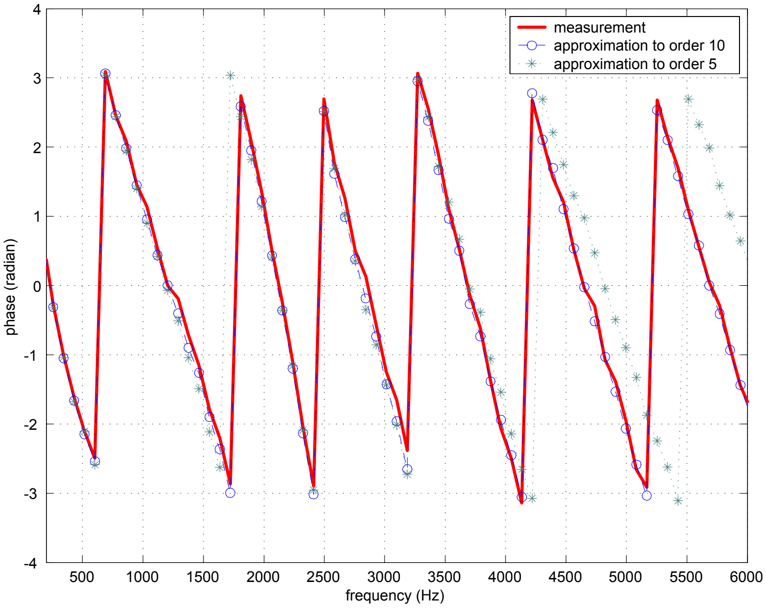


Figure 7.10: Phases of the approximations to orders 5 and 10.

reduced-dimensional description of HRTF measurements, calibration of spherical microphone array, efficient data structure, extension to near-field case, etc.

7.3 A Discrete Huygens Principle Interpretation

This section describes a simple recreation algorithm by using discrete Huygens' principle. It provides a clear and simple insight of the recreation algorithm. It also completely removes the trouble caused by quadratures of spherical harmonics, especially for unbalanced microphone and virtual plane wave layouts. In addition, it provides extremely flexible microphone array and virtual plane waves designs. Since the weights are directly contributed to the final recreated field, it is easy to include the robustness constraints, such as the White Noise Gain.

It is obviously that this algorithm can be derived from wave equations directly without using Huygens principle. In this sense, this section will try to provide an integrated viewpoint of different recreation algorithms and more insights among them.

7.3.1 Solutions by Discrete Huygens Principle

For band-limited 3D sound field, *Huygens Principle* can be approximated by discrete sampling points, limited by Nyquist spatial frequency.

We first consider a unit magnitude plane wave $\mathbf{k}$ in free space, incident from direction $\boldsymbol{\theta}_k$, at the observation point $(a, \boldsymbol{\theta}_j), i = 1, ..., M$, the complex pressure is:

$$\psi_{in}(\boldsymbol{\theta}_k,a,\boldsymbol{\theta}_j) = 4\pi \sum_{n=0}^{p} i^n j_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_j). \tag{7.34}$$

M is the number of observation points or microphones. p is the band limit. a is the radius of the observation point. It doesn't have to be the same value for all observation points. We will extend this later in this section.

This can be easily extended into arbitrarily complicated band-limited sound field by linearity. We will see later our algorithm only depends on the recordings on microphones, not on the complexity of sound field.

To recreate the original real sound field using L virtual plane waves in free space from directions $\boldsymbol{\theta}_i, i = 1, ..., L$, we only need to recreate the sound field at the finite sampling points for band-limited sound field:

$$\psi_{in}(\boldsymbol{\theta}_k,a,\boldsymbol{\theta}_j) = \sum_{i=1}^{L} w_i \psi_{in}(\boldsymbol{\theta}_i, a, \boldsymbol{\theta}_j). \tag{7.35}$$

$$(j = 1, ..., M.)$$

where w_i is the complex weight assigned to each virtual plane wave incident from

$\boldsymbol{\theta}_i$. We then have:

$$\begin{aligned}
&\sum_{n=0}^{p} i^n j_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_j) \\
&= \sum_{i=1}^{L} w_i \left[\sum_{n=0}^{p} i^n j_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_i) Y_n^{m*}(\boldsymbol{\theta}_j) \right] \\
&= \sum_{i=1}^{L} w_i A_p(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j), \qquad (7.36)
\end{aligned}$$

To insert a rigid spherical array of radius a in both real and virtual cases, we then have:

$$\psi_{total}(\boldsymbol{\theta}_k,a,\boldsymbol{\theta}_j) = \sum_{i=1}^{L} w_i \psi_{total}(\boldsymbol{\theta}_i, a, \boldsymbol{\theta}_j). \qquad (7.37)$$

That is:

$$\begin{aligned}
&\sum_{n=0}^{p} i^n \left[j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka) \right] \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_j) \\
&= \sum_{i=1}^{L} w_i \left[\sum_{n=0}^{p} i^n \left[j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka) \right]) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_i) Y_n^{m*}(\boldsymbol{\theta}_j) \right] \\
&= \sum_{i=1}^{L} w_i B_p(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j). \qquad (7.38)
\end{aligned}$$

(7.38) can be formed into a linear system:

$$\begin{bmatrix}
B_p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1) & B_p(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1) & \cdots & B_p(\boldsymbol{\theta}_L, \boldsymbol{\theta}_1) \\
B_p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) & B_p(\boldsymbol{\theta}_2, \boldsymbol{\theta}_2) & \cdots & B_p(\boldsymbol{\theta}_L, \boldsymbol{\theta}_2) \\
\cdots & \cdots & \ddots & \cdots \\
B_p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_M) & B_p(\boldsymbol{\theta}_2, \boldsymbol{\theta}_M) & \cdots & B_p(\boldsymbol{\theta}_L, \boldsymbol{\theta}_M)
\end{bmatrix}
\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix}
=
\begin{bmatrix} \psi_{total}(\boldsymbol{\theta}_k, a, \boldsymbol{\theta}_1) \\ \psi_{total}(\boldsymbol{\theta}_k, a, \boldsymbol{\theta}_2) \\ \vdots \\ \psi_{total}(\boldsymbol{\theta}_k, a, \boldsymbol{\theta}_M) \end{bmatrix}. \qquad (7.39)$$

It is obvious to me that the solution of $w_i, i = 1, ..., L$ also satisfies (7.36).

Please note the right side of (7.39) is just a sampling of the sound field, *i.e.* a recording. It can be multiple sources in complicated acoustic environment.

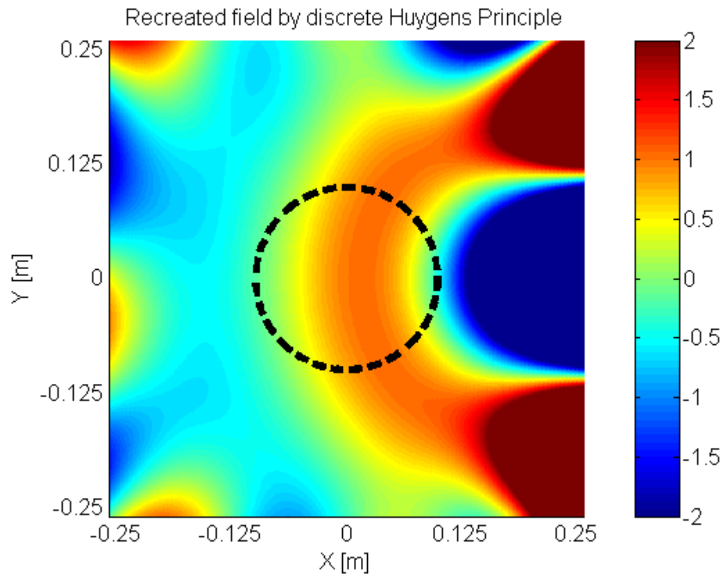


Figure 7.11: 1KHz sound source at 2a, recorded using a=10cm spherical array with 64 Fliege nodes. Plot shows the recreated sound field using 64 virtual plane waves.

7.3.2 Simulation Results

We first use microphone array on rigid sphere to record the sound field. Figure 7.11-7.13 show the recreation results. Please note Fig. 7.12 uses 60 mics and 60 virtual plane waves.

We can also use microphone array in free space as in Fig. 7.14. In this case, we can use multiple arrays easily. For example, we can use two small microphone arrays apart to sample the regions around each ear, each with radius $3cm$ and 32 nodes (face centers of truncated icosahedron). Then we can use two sets of 32 virtual plane waves to recreate sound field at those two regions sampled by two mic arrays, respectively. Therefore, we recreate the sound field around left and right ears, then filtered by HRTF. A simulation result for the region around one ear is shown in Fig. 7.15.

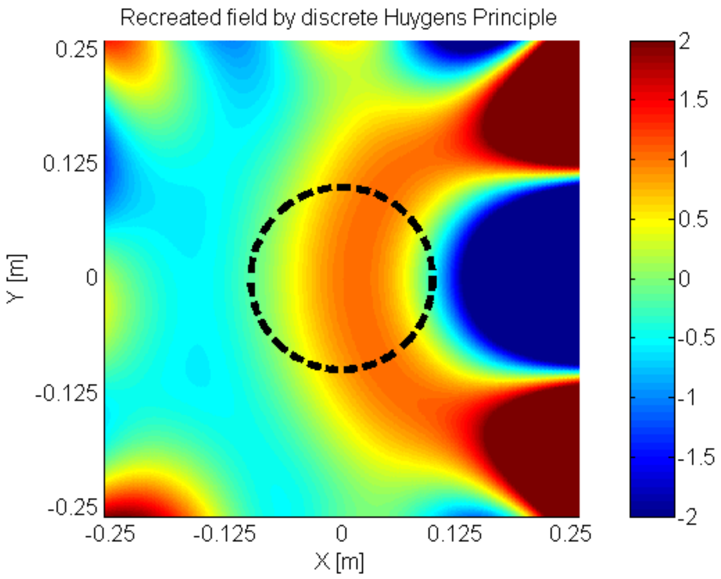


Figure 7.12: Same setup, but using only 60 microphones and 60 virtual plane waves.

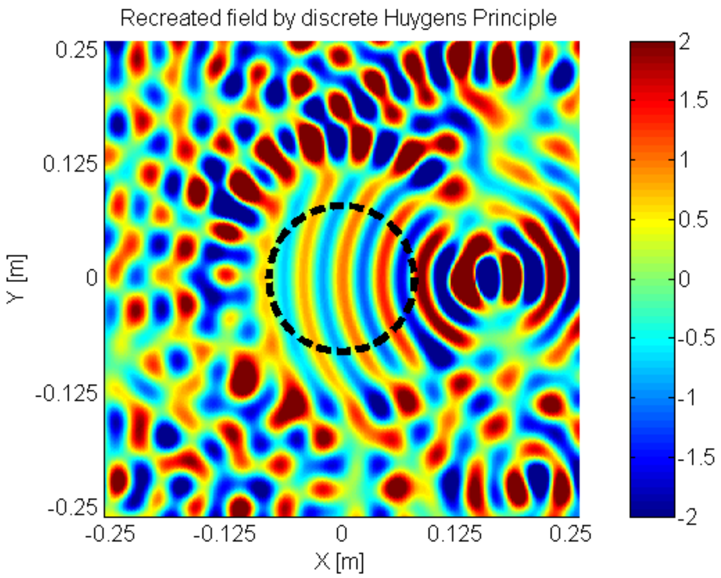


Figure 7.13: recorded using spherical array with 324 Fliege's nodes with a=8cm. 8KHz source from 2a. Plot shows the recreated sound field using 324 virtual plane waves.

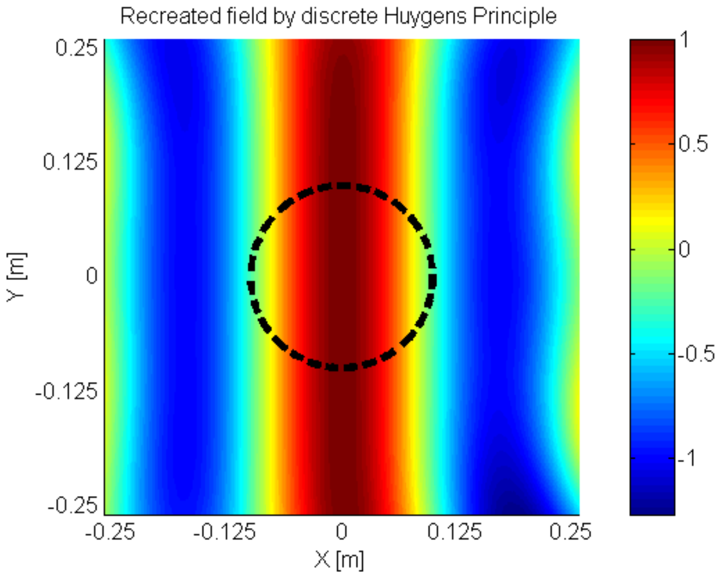


Figure 7.14: Recreated plane wave of 1KHz, using 64 nodes mic array in free space with $a = 10cm$. Recreated by 64 virtual plane waves.

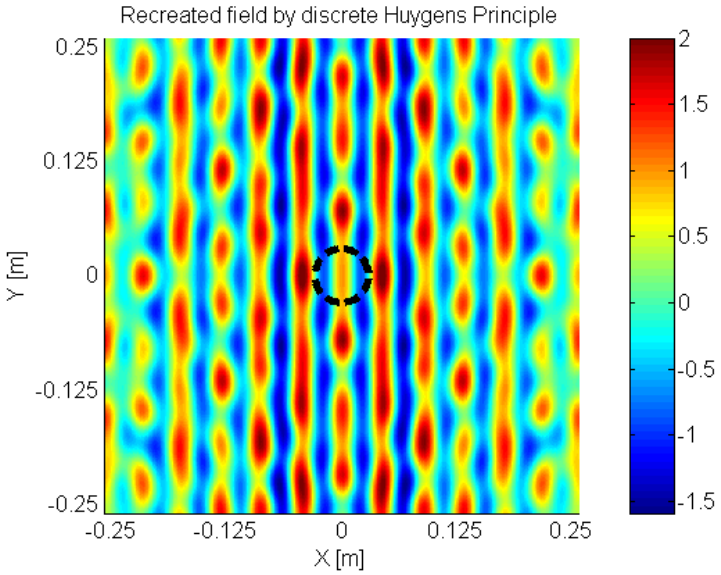


Figure 7.15: Free space, 8KHz plane wave, recorded using free space 32-nodes mic array with radius $a = 3cm$. Recreated by 32 virtual plane waves.

Chapter 8
Recreate 2D Auditory Scenes[1]

This chapter parallels Chapter 7 to develop the theories of reproducing the recorded surround auditory scenes through loudspeaker array and headphone.

8.1 Loudspeaker Array-Based System

Suppose we have recorded the surround auditory scene using our circular microphone array, the next step is to reproduce it with a circular loudspeaker array. In this array, L loudspeakers are positioned uniformly around a circle in free space. To achieve the optimal performance given the number of microphones (N) used in recording, we let L equal to N. For clarity, we still use the notation L for the number of loudspeakers. In this section, we first consider the plane wave case where the radius of the circular loudspeaker array is large enough to produce plane waves at the observation points. We will derive the theoretical condition of reproduction, and give a simple solution to ensure the optimal performance. Then we extend this to the spherical wave case where the circle of the loudspeaker array is small so that every loudspeaker can be modeled as a point sound source.

8.1.1 Plane Wave Case

We denote the L loudspeaker positions as $\varphi_l, l = 1, ..., L$, which are also the plane wave incident directions. Suppose we feed every loudspeaker with the identical signal of wave number k, and the output of the l-th loudspeaker is weighted by a

[1] This chapter is based on our original work in [62].

complex number $a_l(k)$, the reproduced sound field at the observation point φ is a weighted combination of plane waves:

$$p_c = \sum_{n=0}^{\infty} \epsilon_n i^n J_n(kr) \sum_{l=1}^{L} a_l(k) \cos\left[n(\varphi - \varphi_l)\right]. \tag{8.1}$$

To perfectly reproduce the original sound field, it requires (6.3) equal (8.1). Thus the condition for reproduction is:

$$\cos\left[n(\varphi - \varphi_k)\right] = \sum_{l=1}^{L} a_l(k) \cos\left[n(\varphi - \varphi_l)\right], \tag{8.2}$$

$$(n = 0, ..., \infty).$$

That is:

$$\cos\left(n\varphi_k\right) = \sum_{l=1}^{L} a_l(k) \cos\left(n\varphi_l\right), \tag{8.3}$$

$$\sin\left(n\varphi_k\right) = \sum_{l=1}^{L} a_l(k) \sin\left(n\varphi_l\right), \tag{8.4}$$

$$(n = 0, ..., \infty).$$

However, since we only can record the sound field up to order M, the reproduction can be achieved to order M at most. If we let:

$$a_l(k) = f_M(\varphi_k, \varphi_l) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^{M} \cos[n'(\varphi_k - \varphi_l)], \tag{8.5}$$

then (8.3) and (8.4) can be satisfied to order M as following:

$$\begin{aligned} & \sum_{l=1}^{L} a_l(k) \cos\left(n\varphi_l\right) \\ = \; & \sum_{l=1}^{L} \left\{ \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^{M} \cos[n'(\varphi_k - \varphi_l)] \right\} \cos\left(n\varphi_l\right) \\ = \; & \cos(n\varphi_k) \qquad (n = 0, ..., M), \end{aligned} \tag{8.6}$$

and

$$\begin{aligned}
& \sum_{l=1}^{L} a_l(k)\sin(n\varphi_l) \\
= & \sum_{l=1}^{L}\left\{\frac{1}{2\pi}+\frac{1}{\pi}\sum_{n'=1}^{M}\cos[n'(\varphi_k-\varphi_l)]\right\}\sin(n\varphi_l) \\
= & \sin(n\varphi_k) \qquad\qquad (n=0,...,M). \qquad (8.7)
\end{aligned}$$

The physical meaning is clear: if each loudspeaker plays back the beamformed signal in this direction, the loudspeaker array will recreate the original surround sound field to order M.

8.1.2 Spherical Wave Case

Thus far, we have assumed the loudspeakers are far enough from the observation points to generate plane waves. In this section, we extend the algorithm to the general point-source model where the reproduced sound field is a weighted combination of spherical waves.

Suppose the radius of the circular loudspeaker array is r_0. If the loudspeakers can be modeled as point sources, the sound field at the observation point (r, θ, φ) reproduced by the circular loudspeaker array becomes [23]:

$$p_c = 4\pi\sum_{n=0}^{\infty} i^n j_n(kr)R_n(kr_0)\sum_{m=-n}^{n}\sum_{l=1}^{L} a_l(k)Y_n^m(\theta_l,\varphi_l)Y_n^{m*}(\theta,\varphi), \qquad (8.8)$$

where

$$R_n(kr_0) = -ikr_0e^{ikr_0}i^{-n}h_n(kr_0). \qquad (8.9)$$

Normally, the listener is at the center of the circular ring of the loudspeakers, so

$\theta_l = \pi/2$ and $\theta = \pi/2$. Since

$$\begin{aligned} Y_n^m(\frac{\pi}{2}, \varphi) &\equiv \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\frac{\pi}{2}) e^{im\varphi} \\ &= \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{im\varphi}, \end{aligned} \tag{8.10}$$

the recreated sound field is:

$$p_c = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) R_n(kr_0) \sum_{l=1}^{L} a_l(k) \sum_{m=-n}^{n} \left[\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}\right] \left[P_n^m(0)\right]^2 e^{im(\varphi_l - \varphi)}. \tag{8.11}$$

Using the identity relating associated Legendre polynomials with negative m to the corresponding functions with positive m:

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x), \tag{8.12}$$

we have:

$$\begin{aligned} &\sum_{m=-n}^{n} \left[\tfrac{2n+1}{4\pi}\tfrac{(n-m)!}{(n+m)!}\right] \left[P_n^m(0)\right]^2 e^{im(\varphi_l - \varphi)} \\ &= \sum_{m=0}^{n} \epsilon_m \left[\tfrac{2n+1}{4\pi}\tfrac{(n-m)!}{(n+m)!}\right] \left[P_n^m(0)\right]^2 \cos\left[m(\varphi_l - \varphi)\right]. \end{aligned} \tag{8.13}$$

let

$$E_n = 4\pi i^n j_n(kr) R_n(kr_0), \tag{8.14}$$

$$F_n^m = \epsilon_m \left[\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}\right] \left[P_n^m(0)\right]^2. \tag{8.15}$$

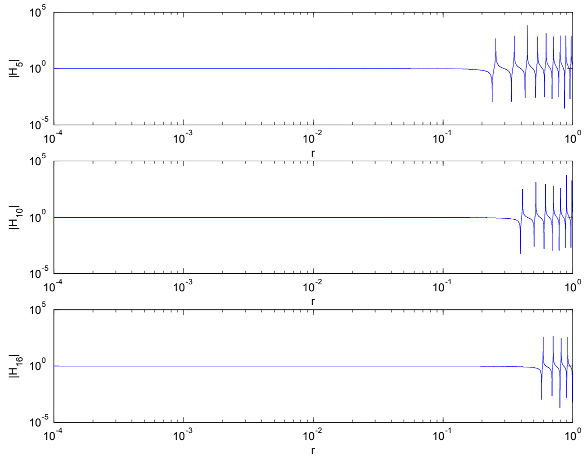


Figure 8.1: H_5, H_{10} and H_{16} at frequency 2 kHz.

So we have:

$$\begin{aligned}
p_c &= \sum_{l=1}^{L} a_l(k) \sum_{n=0}^{\infty} E_n \sum_{m=0}^{n} F_n^m \cos\left[m(\varphi_l - \varphi)\right] \\
&= \sum_{l=1}^{L} a_l(k) \sum_{m=0}^{\infty} \left(\sum_{n=m}^{\infty} E_n F_n^m \right) \cos\left[m(\varphi_l - \varphi)\right] \\
&= \sum_{l=1}^{L} a_l(k) \sum_{n=0}^{\infty} \left(\sum_{m=n}^{\infty} E_m F_m^n \right) \cos\left[n(\varphi_l - \varphi)\right] \\
&= \sum_{n=0}^{\infty} \left(\sum_{m=n}^{\infty} E_m F_m^n \right) \sum_{l=1}^{L} a_l(k) \cos\left[n(\varphi_l - \varphi)\right]. \qquad (8.16)
\end{aligned}$$

Comparing (8.16) against (6.3) or (6.17), the condition for reproduction is:

$$\left(\sum_{m=n}^{\infty} E_m F_m^n \right) \sum_{l=1}^{L} a_l(k) \cos\left[n(\varphi_l - \varphi)\right] = A_n \cos\left[n(\varphi - \varphi_k)\right], \qquad (n = 0, ..., \infty). \qquad (8.17)$$

That is:

$$\left(\sum_{m=n}^{\infty} E_m F_m^n\right) \sum_{l=1}^{L} a_l(k) \cos\left(n\varphi_l\right) = A_n \cos\left(n\varphi_k\right), \tag{8.18}$$

$$\left(\sum_{m=n}^{\infty} E_m F_m^n\right) \sum_{l=1}^{L} a_l(k) \sin\left(n\varphi_l\right) = A_n \sin\left(n\varphi_k\right), \tag{8.19}$$

$$(n = 0, ..., \infty).$$

Similarly, (8.18) and (8.19) can be satisfied for $n = 0$ to M with the solution of $a_l(k)$ as:

$$\begin{aligned} a_l(k) &= \frac{A_n}{\sum_{m=n}^{\infty} E_m F_m^n} f_M(\varphi_k, \varphi_l) \\ &= H_n(k, r, r_0) \left\{ \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^{M} \cos[n'(\varphi_k - \varphi_l)] \right\}, \end{aligned} \tag{8.20}$$

which is still the beampattern but multiplied by a variable coefficient:

$$H_n(k, r, r_0) = \frac{A_n}{\sum_{m=n}^{\infty} E_m F_m^n}. \tag{8.21}$$

To analyze its effects on the reproduced sound field, we plot $H_n(k, r, r_0)$ with $r_0 = 1\,\mathrm{m}$. Figure 8.1 plots H_5, H_{10} and H_{16} at 2 kHz with respect to r. Figure 8.2 plots the same functions at 5 kHz. Several interesting observations are:

1. If r is small compared with r_0, i.e. the observation points are in the far field area, $H_n \approx 1$. This is roughly the same as the plane wave case;

2. At fixed frequency, with larger n, the far field area stretches. This is because the first zero of H_n occurs at the same r as the first zero of $J_n(kr)$, the n-th order Bessel function of the first kind;

3. The far field area shrinks as the frequency increases, which is obvious since the first zero of $J_n(kr)$ for fixed n is a constant number, when k increases, r has to decrease to stay in the far field area.

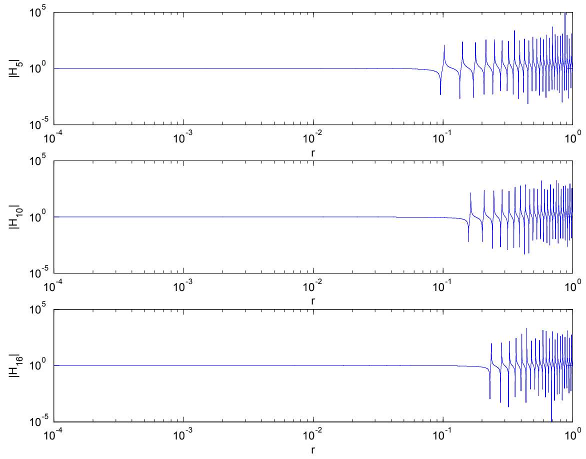


Figure 8.2: H_5, H_{10} and H_{16} at frequency 5 kHz.

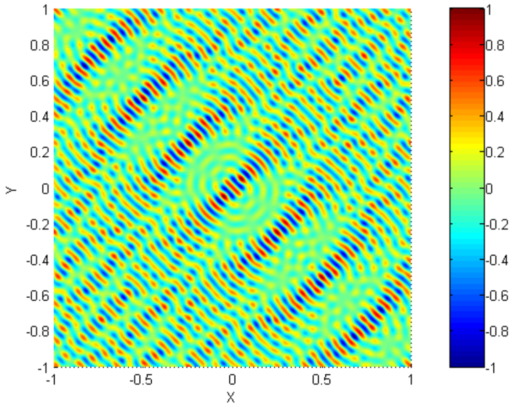


Figure 8.3: Reproduce the 5 kHz plane wave to order 5 using plane waves.

8.1.3 Simulation

We simulate the reproduction of a plane wave of 5 kHz incident from $\varphi = \pi/4$. Figure 8.3-8.5 show the reproduction by plane waves to the order 5, 10, 16, respectively. Figure 8.6-8.8 show the reproduction by spherical waves at the same orders with $r_0 = 1\,\mathrm{m}$. Figure 8.9 shows the reproduction to order 5 at 2 kHz. As we have predicted, the reproduction is effective in a larger area than in Figure 8.6.

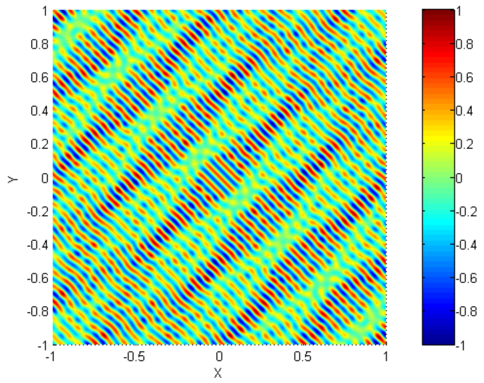


Figure 8.4: Reproduce the 5 kHz plane wave to order 10 using plane waves.

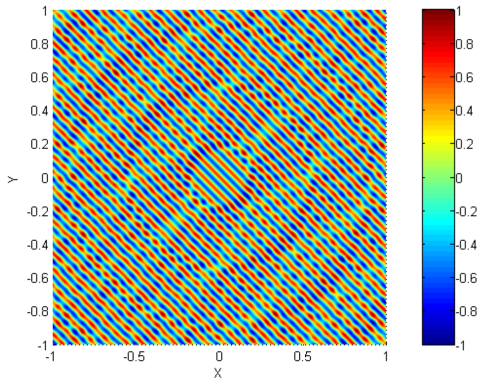


Figure 8.5: Reproduce the 5 kHz plane wave to order 16 using plane waves.

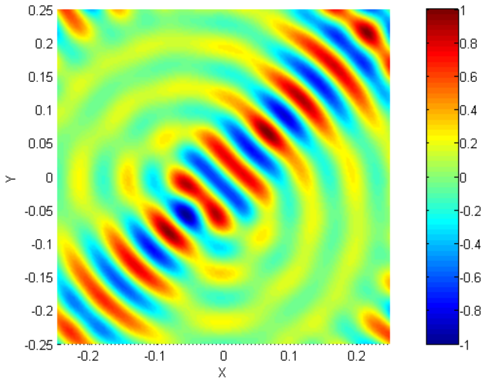


Figure 8.6: Reproduce the 5 kHz plane wave to order 5 using spherical waves.

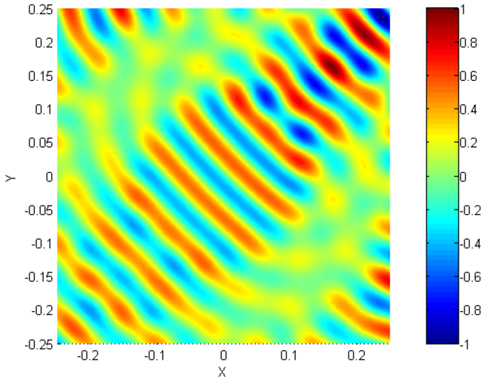


Figure 8.7: Reproduce the 5 kHz plane wave to order 10 using spherical waves.

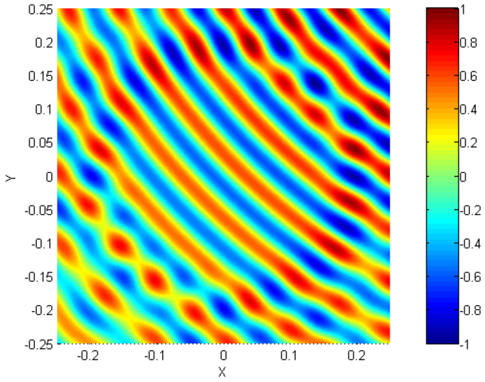


Figure 8.8: Reproduce the 5 kHz plane wave to order 16 using spherical waves.

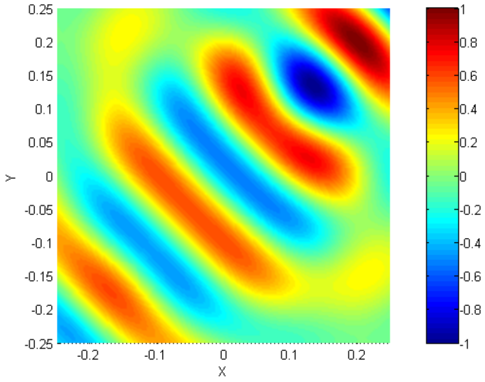


Figure 8.9: Reproduce the 2 kHz plane wave to order 5 using spherical waves.

8.2 Headphone-Based System

In this section, we achieve optimal personalized reproduction from the recording of our circular microphone array, with the cylindrical beamformer as the bridge.

8.2.1 Ideal HRTF Selection

In an ideal case, we assume the HRTF is already measured continuously on the spherical surface of radius r. Our goal is to select the correct HRTF for a specified direction. Although it seems trivial for an ideal case, we use this as a starting point and extend it into more practical cases in the following sections.

We drop the arguments k and r for simplicity, the HRTF for the sound of wave number k from the point (r, θ, φ) can be modeled as [31]:

$$\psi(\theta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta, \varphi), \tag{8.22}$$

where h_n is the first kind spherical Hankel function of order n, Y_n^m is the spherical harmonics of order n and degree m, and α_{nm} is the fitting coefficients.

In surround audio case, $\theta = \pi/2$, we have:

$$\begin{aligned} Y_n^m(\frac{\pi}{2}, \varphi) &\equiv \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\frac{\pi}{2}) e^{im\varphi} \\ &= \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{im\varphi}, \end{aligned} \tag{8.23}$$

So:

$$\psi(\frac{\pi}{2}, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{nm} h_n(kr) e^{im\varphi}, \tag{8.24}$$

where

$$A_{nm} = \alpha_{nm}\sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(0). \tag{8.25}$$

Suppose we want to select the HRTF for the spatial point $(r, \pi/2, \varphi_k)$, we apply the following weighting function (ideal beampattern) to each measured HRTF:

$$\delta(\varphi - \varphi_k) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{in(\varphi-\varphi_k)}. \tag{8.26}$$

According to the property of delta function, we have:

$$\int_0^{2\pi} \psi(\frac{\pi}{2}, \varphi)\delta(\varphi - \varphi_k)d\varphi = \psi(\frac{\pi}{2}, \varphi_k), \tag{8.27}$$

Obviously, the weighting function simply selects the value we need and discards everything else.

To present another viewpoint of HRTF selection, we rewrite (8.27) into a more "complicated" form:

$$\begin{aligned}&\int_0^{2\pi}\left[\sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}h_n(kr)e^{im\varphi}\right]\left[\frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{in(\varphi-\varphi_k)}\right]d\varphi\\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}h_n(kr)e^{im\varphi_k}.\end{aligned} \tag{8.28}$$

Alternatively, this can be easily proved using the orthonormality of exponential functions:

$$\frac{1}{2\pi}\int_0^{2\pi} e^{im\varphi}e^{in\varphi}d\varphi = \delta_{mn}. \tag{8.29}$$

8.2.2 HRTF Approximation Via Orthonormal Decomposition

In practice, however, HRTFs are measured on discrete points. In this case, (8.28) and (8.29) can only hold approximately and in finite orders. In addition, using a practical cylindrical array with finite number of microphones, the beampattern is an M-truncated version of (6.15).

Suppose the HRTF for the sound of wave number k from the measurement point $(r, \pi/2, \varphi_i)$ is:

$$\psi(\frac{\pi}{2}, \varphi_i) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{nm} h_n(kr) e^{im\varphi_i}, \tag{8.30}$$

$$(\ i = 1, ..., B\ )$$

where B is the number of HRTF measurements.

When the beamformer is looking at φ_i, the gain for sound wave from φ_k is the truncated beampattern to order M:

$$\begin{aligned} f_M(\varphi_k, \varphi_i) &= f_M(\varphi_i, \varphi_k) \\ &= \frac{1}{2\pi} \sum_{n=-M}^{M} e^{in(\varphi_i - \varphi_k)}. \end{aligned} \tag{8.31}$$

The weighted combination of HRTFs then becomes:

$$\sum_{i=1}^{B} f_M(\varphi_i, \varphi_k) \psi(\frac{\pi}{2}, \varphi_i). \tag{8.32}$$

If the HRTF measurement points $(\theta_i, \varphi_i), i = 1, ...B$, are uniformly distributed on a circle so that the orthonormality of Fourier exponential functions holds up to order N', then the HRTF can be expanded into two groups:

$$\psi(\frac{\pi}{2}, \varphi_i) = \psi_0^{N'}(\frac{\pi}{2}, \varphi_i) + \psi_{N'+1}^{\infty}(\frac{\pi}{2}, \varphi_i), \tag{8.33}$$

where

$$\psi_0^{N'}(\frac{\pi}{2}, \varphi_i) = \sum_{n=0}^{N'} \sum_{m=-n}^{n} A_{nm} h_n(kr) e^{im\varphi_i}, \tag{8.34}$$

$$\psi_{N'+1}^{\infty}(\frac{\pi}{2}, \varphi_i) = \sum_{n=N'+1}^{\infty} \sum_{m=-n}^{n} A_{nm} h_n(kr) e^{im\varphi_i}. \tag{8.35}$$

So (8.32) can be rewritten as:

$$
\begin{aligned}
&\sum_{i=1}^{B} f_M(\varphi_i, \varphi_k) \left[\psi_0^{N'}(\frac{\pi}{2}, \varphi_i) + \psi_{N'+1}^{\infty}(\frac{\pi}{2}, \varphi_i) \right] \\
&= \sum_{i=1}^{B} f_M(\varphi_i, \varphi_k) \psi_0^{N'}(\frac{\pi}{2}, \varphi_i) \qquad (8.36) \\
&\quad + \sum_{i=1}^{B} f_M(\varphi_i, \varphi_k) \psi_{N'+1}^{\infty}(\frac{\pi}{2}, \varphi_i) \qquad (8.37) \\
&= \psi_0^{\min(N', M)}(\frac{\pi}{2}, \varphi_k) + \epsilon, \qquad (8.38)
\end{aligned}
$$

which is the approximation of HRTFs up to the order $\min(N', M)$. Here the error ϵ consists of two parts: one is the orthonormality error from (8.36) which is supposed to be small according to the discrete orthonormalities; the other is from (8.37) which is also small with well-chosen N' because of the convergence of the series expansion in (8.22) [31]. We will provide a design example in the next section.

If HRTFs are not measured on uniformly distributed angular points, we can first obtain a uniform version via interpolation [31]. In addition, given the number of microphones used in recording, we need to make at least the same number of HRTF measurements so that the recorded spatial information can be recreated fully. In this case, the HRTF approximation at $(\pi/2, \varphi_k)$ depends only on the order of beampattern M, which is:

$$\psi_0^{M}(\frac{\pi}{2}, \varphi_k) + \epsilon. \qquad (8.39)$$

8.2.3 Reproduction from Recordings

From the analysis of HRTF approximation, the algorithm to reproduce the surround sound through headphone becomes quite obvious and intuitive: we first beamform the recordings to $(\pi/2, \varphi_i)$, and filter the beamformed signal with the HRTF measured at the same direction, then sum them up for $i = 1, ..., B$. After that, the resulted two-channel signals are fed into the headphone.

With the currently available HRTF measurements, the only factor that determines reproduction quality is the beampattern order M of the cylindrical microphone array.

8.2.4 Surround Audio Design Example

To illustrate our design, we use the KEMAR HRTF data [2]. In our case, we only consider the measurements on the equator where $\theta = \pi/2$. Fig. 8.10 shows the magnitudes of spatial frequency components at different temporal frequencies. For example, the HRTF at 861.3 Hz has significant spatial frequency components for $|n| \leq 4$ (negative n actually is $72 + n$ in Fig. 8.10.). From Fig. 6.3, ka should have the value of 2.4 roughly to have the beampattern of order 4, which means the radius of the cylinder is approximately 15 cm. Similarly, at 2584.0 Hz, the significant spatial frequency components are for $|n| \leq 9$, so we desire $ka \approx 7$ to achieve a beampattern of order 9. Again we estimate the radius of the cylinder as $a \approx 15$ cm. We can do this at other temporal frequencies plotted in Fig. 8.10. As the temporal frequency increases, the required radius a decreases. At 9474.6 Hz, the required a is well less than 10 cm. In practice, we can just set the radius a as 15 cm.

To determine the minimum number of microphones required in this array, we first have to know the temporal frequency band according to specific applications. For example, in surround speech acquisition, the frequency is from 500 Hz to 4 kHz.

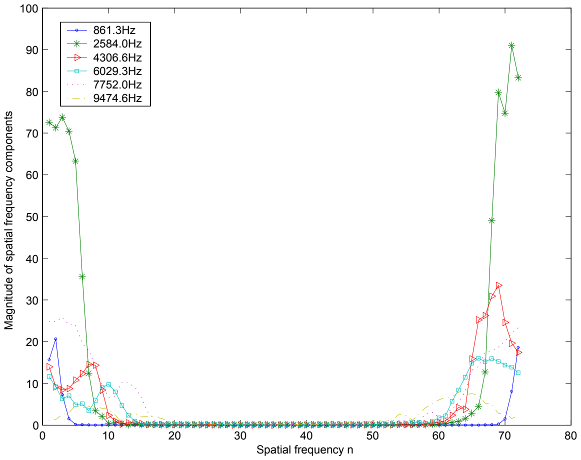


Figure 8.10: HRTFs in spatial frequency domain.

From Fig. 8.10, the significant spatial frequency components are $|n| \leq 16$. Therefore, 32 microphones are enough for this application.

Chapter 9
Experiment in Traffic Auditory Scenes[1]

Traffic surveillance is usually based on computer vision and has been attracting many researchers [21][78][86][55]. On the other hand, the sound from a running vehicle also provides rich information about the vehicle. Researchers have used analogous methods from computer vision for vehicle sound signature recognition [94]. To extract spatial information from the auditory scene, however, we use our spherical microphone array. This chapter presents our preliminary 3D auditory scene analysis in simple traffic environment. The experiment also tests the performance of our spherical microphone array.

Our system is as shown in Fig. 9.1. The main part is the spherical microphone array, which consists of 60 omnidirectional microphones (Knowles FG-3629) mounted on the rigid spherical surface of radius 10 cm. The positions of the 60 microphones are decided as the 64 nodes in [35] with four nodes removed because of the cable outlet and the mounting base. The sound signals received by the array will flow through a 64 channel pre-amplifier before acquired by two 32 channel NI-DAQ cards on the computer.

We took the spherical microphone array system to roadside to record real-world traffic auditory scenes. A simple scenario is as shown in Fig. 9.2 where a car is moving from the left side of the array to the right side along the street.

To track the car in this scenario from the recordings, for each frame, we simply

[1] This chapter is based on our original work in [64].

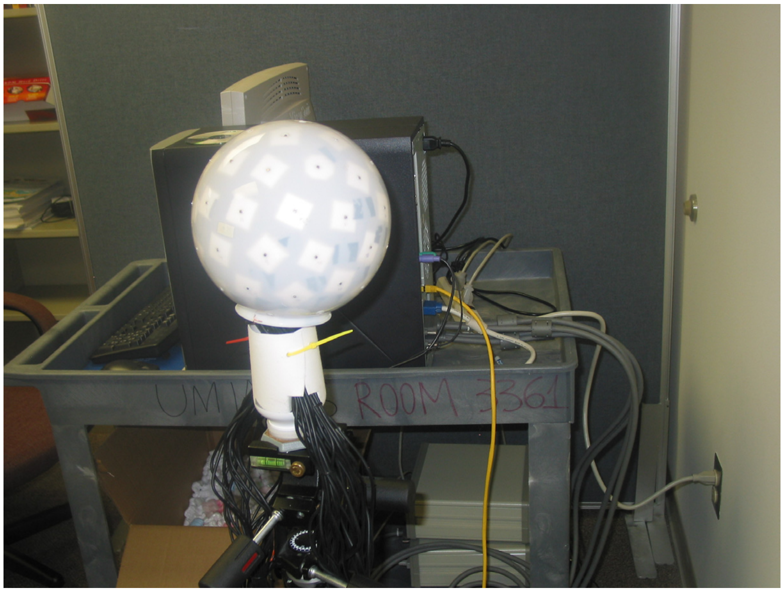


Figure 9.1: The snapshot of our spherical microphone array system.

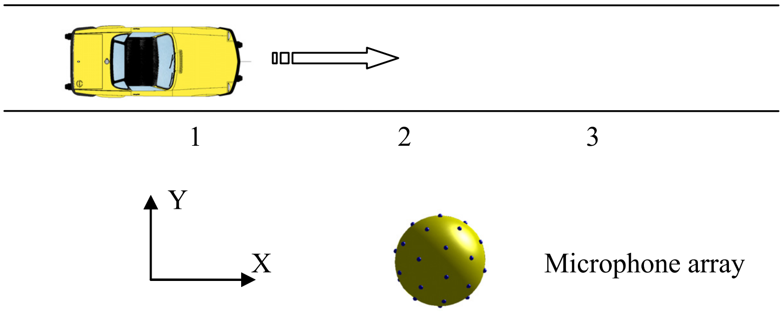


Figure 9.2: The scenario of the experiment. A car is moving from left to right. The spherical microphone array is placed on the roadside.

steer the spherical beamformer to every 3D direction to search for the peaks. We assume the sound from the car is stronger than the sound from all other directions. So the peaks indicate the car locations. Since the spectrum of the sound from the car is not constant, the localization is performed at multiple frequencies. We will extend this to multiple sound source localization later in this section. The tracking path is shown in Fig. 9.3 where the left and right sides of the array correspond to azimuth angle π and 0, respectively. The jitters on the tracking path may be caused by several factors:

1. the sound spectrum from the vehicle is not constant across the recording time;
2. the sound source is not simply a point source since many parts of a running vehicle can make sound;
3. the environment itself is very noisy with wind blowing, people walking and talking, a building being constructed on the other side of the street, the reflections from walls and the ground surface in addition to self noise of the system.

When the car is far away from the array, its sound fades into the environmental noise. So the jitters tend to be more salient at the two endpoints of the path. We will see this more clearly in the two-vehicle case later.

To show the actual 3D localization performance of the spherical beamformer, we pick three typical frames of the recorded sound as numbered 1-3 in Fig. 9.2. The localization result is a 3D surface, where each surface point represents a localization

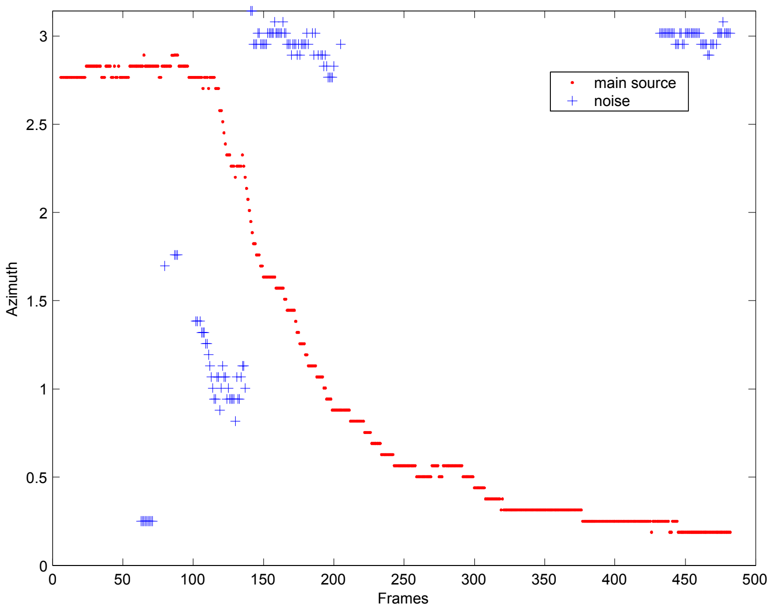


Figure 9.3: Tracking one vehicle using the spherical microphone array.

angle and its distance from the origin represents the amplitude of the beamformed signal. In an ideal free space environment under plane wave assumption, using a standard beampattern, the localization surface will have the same shape of this beampattern pointing to the plane wave direction. In a real-world environment, however, those factors listed above will affect the final result which may in turn provide some spatial cues about the sound source and the environment. For clarity, we only show the localizations using a single frequency at about 2 kHz, and plot in normalized linear scale instead of dB scale. We provide two different viewpoints for each figure, the left plot is from the same viewpoint as in Fig. 9.2 while the right one is from an appropriate 3D viewpoint. The first 3D localization is shown in Fig. 9.4. The localization for the second chosen frame is shown in Fig. 9.5, which indicates the effect of a spreading sound source at this frequency. In addition, the irregular sidelobes are likely caused by all kinds of noise. Fig. 9.6 shows the 3D localization

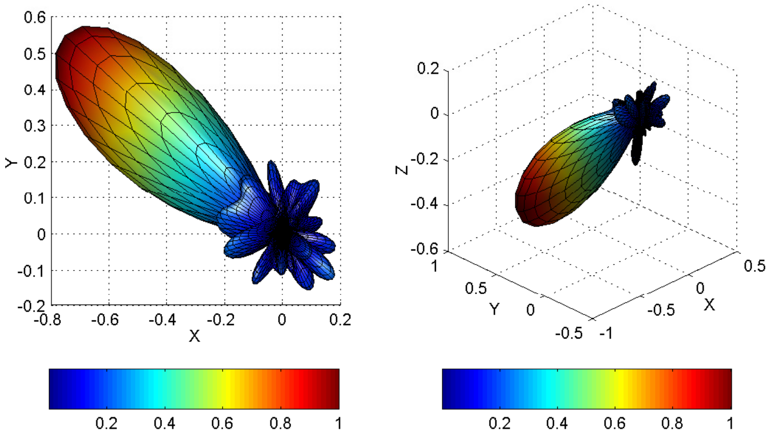


Figure 9.4: 3D localization of the first chosen frame. The car is now on the left side of the array.

for the third chosen frame. It clearly shows a second sound source at this frequency, real or image, but weak. We can also see this sound is from the lower part of the car, possibly from the running wheels or brakes.

We also tested our system in a more challenging scenario. In another experiment, we recorded two cars moving from right to left successively in different speed. Fig. 9.7 shows the tracks of two cars.

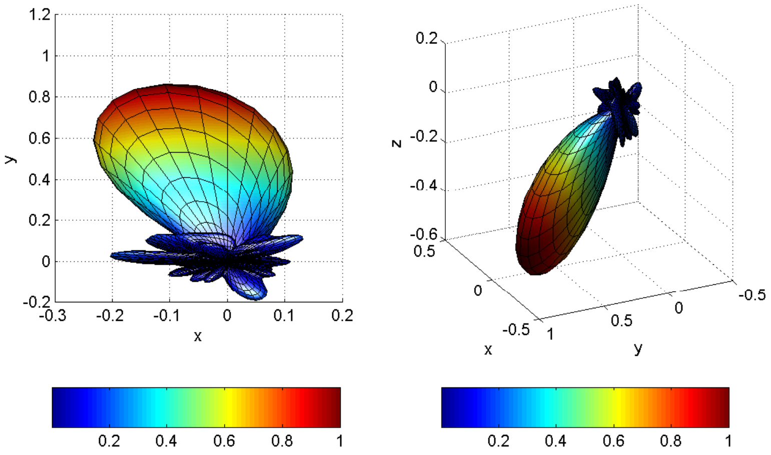


Figure 9.5: The car is now roughly in front of the array.

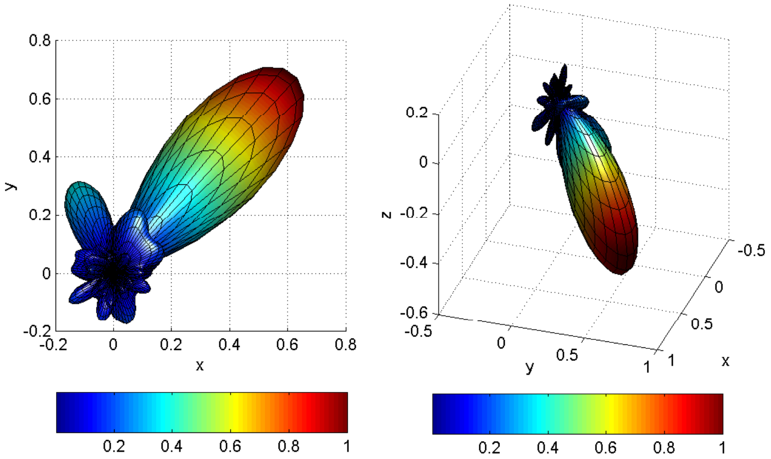


Figure 9.6: The car now moves to the right side of the array.

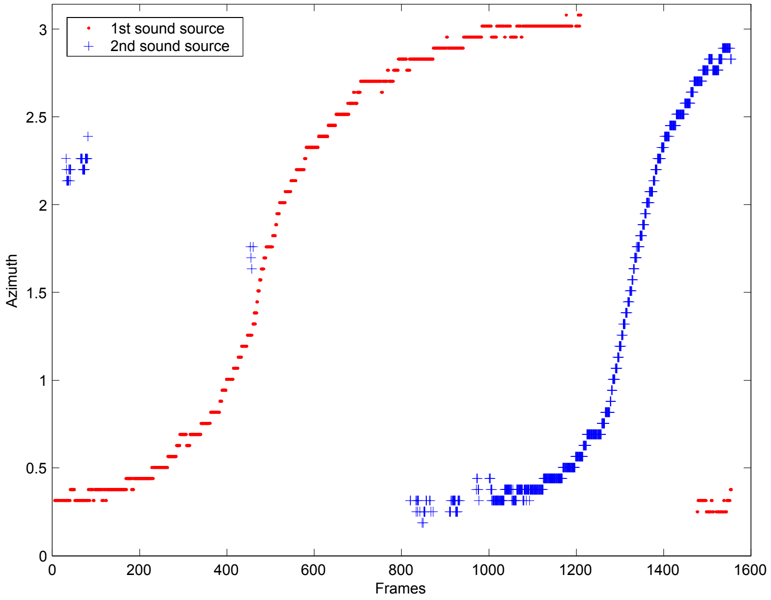


Figure 9.7: Tracking two vehicles using the spherical microphone array.

Chapter 10
Conclusions

We have developed the theories and tools for capture and recreation of 3D auditory scenes. Our contributions are:

1. We propose a flexible and optimal design of spherical microphone arrays for beamforming. We analyze the discrete orthonormality error of spherical harmonics and cancel them by several methods. A simplified algorithm is designed to easily control the robustness of the resulting beamformer. An adaptive implementation is presented. It optimally converges to desired beampatterns under pre-specified robustness constraint. Design examples and simulation results are provided.

2. We design and build a hemispherical microphone array using the acoustical image principle. Simulation and experimental results demonstrate that a hemispherical microphone array with a hard plane is equivalent to a full spherical array with image microphones and image sound sources. We also use the hemispherical array to verify our methodology in 1.

3. We propose a pre-computed fast spherical beamforming algorithm. It achieves fast mixing in time domain.

4. We design the cylindrical beamforming algorithm using a circular microphone array mounted on a rigid cylinder. It is practical in 2D auditory scenes.

5. We design loudspeaker and headphone based algorithms for optimal recreation of spatial auditory scenes. It exploits the reciprocity between the capture and recreation processes. In addition, we present an alternative interpretation using the discrete Huygens principle.

6. We test our spherical microphone array in real-world traffic scenes.

Chapter 11
On-going and Future Work

The on-going and future work will focus on extending our established work. Here we just give a short list of some possible researches.

11.1 Multiple Array Data Fusion

Up to now, we only use a single spherical microphone array to capture auditory scenes. Just like people using multiple cameras in computer vision [26], we can use more than one array to capture the auditory scene from different locations. The recordings are expected to deliver more information about the sound source and the surrounding acoustical environment.

For example, the performance of the spherical microphone array we have discussed is related to the microphone numbers and the sphere size. To work with broader band sound signals, we may need a nested spherical microphone array with each layer has different number of microphones and different radius. Another example is that we can use multiple arrays to localize the sound source in 3D space, not just an angular position.

Since our ultimate goal is to recreate a dynamic 3D auditory scene from recordings, we should allow the listener to move in the scene. In [31], HRTFs can be interpolated and range extrapolated. We believe that we can implement a system integrating multiple microphone arrays, recreation algorithms, head tracking and HRTF interpolation and range extrapolation to achieve the dynamic workthrough in 3D auditory scenes.

11.2 3D Sound Texture Resynthesis

Probably the best way to describe sound texture is by examples: crackling fire, running water, raining, applause, moving vehicle, large group of people chatting, etc. It is usually monotonic and contains a lot of self-similiar sound frames. The concept of sound texture is from its counterpart in graphics: image texture.

The analysis and resynthesis of image texture have been extensively explored in [33][14][51][89][90][83], etc. Recently, sound texture has started to attract some attentions, such as in [7][69][29]. We plan to follow this direction further into 3D sound texture using our 3D auditory scene capture and recreation technologies and tools.

11.3 Complex Environment Analysis

In our experiments with our spherical microphone array, we found the results are more or less affected by the surrounding environment such as undesired noise, reflection, etc. While the difference between ideal free space and real-world environment adds some irregularities in the experimental results, this also provides some clue about the environment.

11.4 3D Auditory Scene Editing

In a video game, there are only limited set of graphics components created by laser scanner, camera, or/and modeling software. The editing and resynthesis of those components give the object astonishing details and dynamics. In addition, to achieve some special visual effects in a movie which is impossible to happen in real scenes, the editing and resynthesis are the keys.

We want to implement the same functionalities in auditory scenes. In that way, we only need to make limited set of recordings, then we can generate as many desired auditory scenes as we need. In addition to beamforming, there are many algorithms of *blind signal separation* (BSS) to separate one sound source from the other [16]. Those techniques may let us be able to edit the individual auditory objects in the recorded scenes. Another application is to migrate the recorded auditory scenes into a new acoustic environment.

11.5 Real-Time Implementation

For some applications, it is important to process the data in real time. Therefore, we have to design a highly efficient signal processing unit. This will be addressed in two parts:

1. Efficient data structure and codec. The raw data recorded by the microphone array has a large size even for a short time recording. This may be a bottleneck in the band-limited remote applications such as teleconference. Fortunately, the multichannel data are apparently redundant in the sense they are highly correlated to each other. We need to implement an efficient data structure and codec to store and transmit the recordings.

2. Efficient rendering pipeline on graphics hardware. In previous chapters, we have made some efforts to design an efficient parallel implementation of multi-

directional spherical beamformer and pre-computed fast spherical beamforming. However, the multichannel signal processing is still expensive. To solve this problem, we will make use of the parallel architecture of the modern graphics hardware, i.e. the *graphics processing unit* (GPU). Some examples of using GPU for general purposes are [56][53].

BIBLIOGRAPHY

[1] Ambisonics website. http://www.ambisonic.net.

[2] KEMAR website. http://sound.media.mit.edu/KEMAR.html.

[3] Abhayapala, T. D., and Ward, D. B. Theory and design of high order sound field microphones using spherical microphone array. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02)* (May 2002), vol. 2, pp. 1949–1952.

[4] Abramowitz, M., and Stegun, I. A., Eds. *Handbook of Mathematical Functions*. U.S. Government Printing Office, 1964.

[5] Algazi, V. R., Avendano, C., and Duda, R. O. Low-frequency ILD elevation cues. *J. Acoust. Soc. Am. 106* (1999), 2237.

[6] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. The CIPIC HRTF database. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, Oct. 2001), pp. 99–102.

[7] Athineos, M., and Ellis, D. Sound texture modelling with linear prediction in both time and frequency domains. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003)* (2003), vol. 5, pp. 648–651.

[8] Begault, D. R., Wenzel, E. M., Lee, A. S., and Anderson, M. R. Direct comparison of the impact of head-tracking, reverberation and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc. 49*, 10 (2001), 904–916.

[9] Berkhout, A., Vries, D., and Vogel, P. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America 93*, 5 (May 1993), 2764–2778.

[10] Blattner, M., Greenberg, R., and Sumikawa, D. Earcons and icons: Their structure and common design principles. *Human Computer Interaction 4*, 1 (1989), 11–44.

[11] Blauert, J. *Spatial Hearing.* MIT Press, Cambridge, MA., 1983.

[12] Bly, S., Frysinger, S. P., Lunney, D., Mansur, D. L., Mezrich, J. J., and Morrison, R. C. *Readings in Human-Computer Interaction: A Multi-Disciplinary Approach.* Los Altos: Morgan-Kauffman, 1987, ch. Communicating with Sound, pp. 420–424.

[13] Bolt, R. A. "put-that-there": Voice and gesture at the graphics interface. In *ACM 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'80)* (Seattle, WA, 1980), pp. 262–270.

[14] Bonet, J. D. Multiresolution sampling procedure for analysis and synthesis of texture images. In *Proceedings of SIGGRAPH 1997* (1997).

[15] Bradford, J. H. The human factors of speech-based interfaces: A research agenda. In *ACM SIGCHI Bulletin* (Apr. 1995), vol. 27, pp. 61–67.

[16] Brandstein, M., and Ward, D., Eds. *Microphone Arrays.* Springer-Verlag, New York, 2001.

[17] Brungart, D. S. Auditory localization of nearby sources III: Stimulus effects. *J. Acoust. Soc. Am. 106* (1999), 3589–3602.

[18] Brungart, D. S., and Durlach, N. I. Auditory localization of nearby sources II: Localization of a broadband source in the near field. *J. Acoust. Soc. Am. 106* (1999), 1956–1968.

[19] Brungart, D. S., and Rabinowitz, W. M. Auditory localization of nearby sources i: Head-related transfer functions. *J. Acoust. Soc. Am. 106* (1999), 1465–1479.

[20] Buxton, W., and Myers, B. A study in two-handed input. In *CHI '86 Conference on Human Factors in Computing Systems* (Apr. 1986), pp. 321–326.

[21] Chen, Y. Highway overhead structure detection using video image sequences. *IEEE Transactions on Intelligent Transportation Systems 4*, 2 (June 2003), 67–77.

[22] Cohen, P. R., and Sullivan, J. W. Synergistic use of direct manipulation and natural language. In *CHI'89* (1989), pp. 227–234.

[23] Colton, D., and Kress, R. *Inverse Acoustic and Electromagnetic Scattering Theory.* Springer-Verlag, New York, 1997.

[24] Cox, H., Zeskind, R. M., and Owen, M. M. Robust adaptive beamforming. *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-35*, 10 (Oct. 1987), 1365–1376.

[25] Culling, J. F., Hodder, K. I., and Toh, C. Y. Effects of reverberation on perceptual segregation of competing voices. *J. Acoust. Soc. Am. 114* (2003), 2871–2876.

[26] Cutler, R., Duraiswami, R., Qian, J., and Davis, L. Design and implementation of the university of maryland keck laboratory for the analysis of visual movement. Tech. Rep. CS-TR-4329, Computer Science Dept., Univ. of Maryland, College Park, MD 20742.

[27] de Vries, D., and Boone, M. Wave field synthesis and analysis using array technology. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (Oct. 1999), pp. 15–18.

[28] Driscoll, J. R., and D. M. Healy, J. Computing Fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math. 15* (1994), 202–250.

[29] Dubnov, S., Bar-Joseph, Z., El-Yaniv, R., Lischinski, D., and Werman, M. Synthesizing sound textures through wavelet tree learning. *Computer Graphics and Application* (July/August 2002), 38–48.

[30] Duraiswami, R., Gumerov, N., Zotkin, D., and Davis, L. Efficient evaluation of reverberant sound fields. In *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY., 2001), pp. 203–206.

[31] Duraiswami, R., Zotkin, D., and Gumerov, N. Interpolation and range extrapolation of HRTFs. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004)* (Montreal, Canada, May 17-21 2004), pp. IV45–IV48.

[32] Eargle, J. *Handbook of Recording Engineering.* Van Nostrand Reinhold, New York, NY, USA, 1996.

[33] Efros, A. A., and Leung, T. K. Texture synthesis by non-parametric sampling. In *Proceedings of IEEE International Conference on Computer Vision(ICCV99)* (Corfu, Greece, Sept. 1999), pp. 1033–1038.

[34] Euclid. *The Thirteen Books of the Elements: Books X-XIII.*, vol. 3. Dover, New York, 1956.

[35] Fliege, J., and Maier, U. A two-stage approach for computing cubature formulae for the sphere. Tech. Rep. Ergebnisberichte Angewandte Mathematik, No. 139T, Fachbereich Mathematik, Universität Dortmund, 44221 Dortmund, Germany, Sept. 1996.

[36] Fliege, J., and Maier, U. The distribution of points on the sphere and corresponding cubature formulae. *IMA Journal on Numerical Analysis 19* (1999), 317–334.

[37] Frauenberger, C., and Noisternig, M. 3d audio interfaces for the blind. In *Proceedings of the 9th International Conference on Auditory Display* (July 2003), pp. 280–283.

[38] Frost, O. L. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE 60*, 8 (Aug. 1972), 926–935.

[39] Funkhouser, T., Tsingos, N., Carlbom, I., Elko, G., Sondhi, M., West, J. E., Pingali, G., Min, P., and Ngan, A. A beam tracing method for interactive architectural acoustics. *J. Acoust. Soc. Am. 115*, 2 (Feb. 2004), 739–756.

[40] Gardner, W. G. *3-D Audio Using Loudspeakers.* PhD thesis, Dept. of Media Arts and Sciences, MIT, 1997.

[41] Gardner, W. G. Head tracked 3-d audio using loudspeakers. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY., 1997).

[42] Gardner, W. G. *Reverberation Algorithms.* Kluwer Academic, Norwell, MA., 1998.

[43] Gardner, W. G., and Martin, K. D. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am. 97*, 6 (1995), 3907–3908.

[44] Gaver, W. W. Synthesizing auditory icons. In *SIGCHI Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands, 1993), pp. 228–235.

[45] Gerzon, M. Periphony: With-height sound reproduction. *J. Audio Eng. Soc. 21* (Jan. 1973), 2–10.

[46] Gilkey, R., and Anderson, T. R., Eds. *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum Associates, Mahwah, NJ., 1997.

[47] Gumerov, N. A., and Duraiswami, R. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions.* Elsevier Science, 2005. ISBN:0080443710.

[48] Hardin, R. H., and Sloane, N. J. A. McLaren's improved snub cube and other new spherical designs in three dimensions. *Discrete and Computational Geometry 15* (1996), 429–441.

[49] Hauptmann, A. G. Speech and gestures for graphic image manipulation. In *ACM CHI 89 Human Factors in Computing Systems Conference* (1989), pp. 241–245.

[50] Healy, D., Rockmore, D., and Moor, S. An FFT for the 2-sphere and applications. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96)* (May 1996), vol. 3, pp. 1323–1326.

[51] Heeger, D., and Bergen, J. Pyramid based texture Analysis/Synthesis. In *Proceedings of SIGGRAPH 95* (1995).

[52] Huang, Y., and Benesty, J., Eds. *Audio Signal Processing For Next-Generation Multimedia Communication Systems*. Kluwer Academic Publishers, 2004.

[53] J. Bolz, I. Farmer, E. G., and Schroeder, P. Sparse matrix solvers on the GPU: Conjugate gradients and multigrid. In *SIGGRAPH2003* (2003), pp. 917–924.

[54] Jot, J., and Chaigne, A. Digital delay networks for designing artificial reverberators. In *Proc. Audio Eng. Soc. Conv.* (1991).

[55] Jung, Y., Lee, K., and Ho, Y. Content-based event retrieval using sematic scene interpretation for automated traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems 2*, 3 (Sept. 2001), 151–163.

[56] Krueger, J., and Westermann, R. Linear algebra operators for GPU implementation of numerical algorithms. In *SIGGRAPH2003* (2003), pp. 908 – 916.

[57] Kulkarni, A., and Colburn, H. S. Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. Am. 115* (2004), 1714–1728.

[58] Li, Y., Ho, K., and Kwan, C. Design of broad-band circular ring microphone array for speech acquisition in 3-d. In *Proceedings of IEEE International Con-*

ference on Acoustics, Speech, and Signal Processing (ICASSP'03) (Apr. 2003), vol. 5, pp. V221–V224.

[59] Li, Z., and Duraiswami, R. Hemispherical microphone arrays for sound capture and beamforming. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'05)* (New Paltz, New York, Oct. 2005).

[60] Li, Z., and Duraiswami, R. Robust and flexible design of spherical microphone arary. *IEEE Trans. on Speech and Audio Processing* (2005). (submitted).

[61] Li, Z., and Duraiswami, R. A robust and self-reconfigurable design of spherical microphone array for multi-resolution beamforming. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)* (Mar. 2005), vol. IV, pp. 1137–1140.

[62] Li, Z., Duraiswami, R., and Davis, L. S. Recording and reproducing high order surround auditory scenes for mixed and augmented reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality* (Arlington, VA, Nov. 2004), pp. 240–249.

[63] Li, Z., Duraiswami, R., Grassi, E., and Davis, L. S. Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (Montreal, Quebec, Canada, May 17-21 2004), pp. IV41–44.

[64] Li, Z., Duraiswami, R., Grassi, E., and Davis, L. S. A spherical microphone array system for traffic scene analysis. In *Proceedings of the 7th IEEE International Conference on Intelligent Transportation Systems (ITSC2004)* (Washington, DC, October 3-6 2004).

[65] Li, Z., Duraiswami, R., and Gumerov, N. A. Capture and recreation of higher order 3d sound fields via reciprocity. In *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)* (Sydney, Australia, July 6-9 2004).

[66] Li, Z., Lee, C. H., and Varshney, A. A real-time seamless tiled display system for 3d graphics with user interaction. Tech. Rep. CS-TR-4696, CS Dept., University of Maryland, 2005.

[67] Li, Z., and Varshney, A. A real-time seamless tiled display for 3d graphics. In *Proceedings of 7th Annual Symposium on Immersive Projection Technology (IPT 2002)* (Orlando, FL, March 24 - 25 2002).

[68] Loomis, J. M. Basic and applied research relating to auditory displays for visually impaired people. In *Proceedings of the 9th International Conference on Auditory Display* (July 2003), pp. 300–302. invited talk.

[69] Lu, L., Li, S., Liu, W., Zhang, H., and Mao, Y. Audio textures. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2002)* (May 2002), vol. 2, pp. 1761–1764.

[70] Macpherson, E. A., and Middlebrooks, J. C. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am. 111* (2002), 2219–2236.

[71] Meyer, J., and Elko, G. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02)* (May 2002), vol. 2, pp. 1781–1784.

[72] Mezrich, J., Frysinger, S., and Slivjanovski, R. Dynamic representation of multivariate time series data. *Journal of the Americal Statistical Association 79* (1984), 34–40.

[73] Middlebrooks, J. C. Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am. 106* (1999), 1493–1510.

[74] Morse, P., and Ingaard, K. *Theoretical Acoustics.* McGraw Hill, New York, 1968.

[75] Mynatt, E. D., and Edwards, W. K. Mapping GUIs to auditory interfaces. In *ACM Symposium on User Interface Software and Technology (UIST '92).* (1992), pp. 61–88.

[76] Ohuchi, M., Iwaya, Y., Suzuki, Y., and Munekata, T. A game for visually impaired children with a 3-d virtual auditory display. In *Proceedings*

of the 9th International Conference on Auditory Display (July 2003), p. 309. abstract.

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

[78] Schoepflin, T., and Dailey, D. Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. *IEEE Transactions on Intelligent Transportation Systems 4*, 2 (June 2003), 90–98.

[79] Shaw, E. A. G. Acoustical features of the human external ear. In *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, Eds. Erlbaum, New York, 1997, pp. 25–48.

[80] Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. Localizing nearby sound sources in a classroom: Binaural room impulse responses. *J. Acoust. Soc. Am. 117*, 5 (May 2005), 3100–3115.

[81] Taylor, M. Cubature for the sphere and the discrete spherical harmonic transform. *SIAM J. Numer. Anal. 32*, 2 (1995), 667–670.

[82] Teutsch, H., Spors, S., Herbordt, W., Kellermann, W., and Rabenstein, R. An integrated real-time system for immersive audio applications. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA, Oct. 2003).

[83] Turk, G. Texture synthesis on surfaces. In *Proceedings of SIGGRAPH* (2001).

[84] Ueberhuber, C. W. *Numerical Computation 2: Methods, Software, and Analysis*. Springer-Verlag, Berlin, 1997.

[85] Veen, B., and Buckley, K. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine* (Apr. 1988), 4–24.

[86] Veeraraghavan, H., Masoud, O., and Papanikolopoulos, N. Computer vision algorithms for intersection monitoring. *IEEE Transactions on Intelligent Transportation Systems 4*, 2 (June 2003), 78–89.

[87] Walker, J., Sproull, L., and Subramani, R. Using a human face in an interface. In *CHI'94 Human Factors in Computing Systems* (Boston, Apr. 1994), pp. 85–91.

[88] Ward, D. B., and Abhayapala, T. D. Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Transactions on Speech and Audio Processing 9*, 6 (Sept. 2001), 697–707.

[89] Wei, L.-Y., and Levoy, M. Fast texture synthesis using tree-structured vector quantization. In *Proceedings of SIGGRAPH* (2000), pp. 479–488.

[90] Wei, L.-Y., and Levoy, M. Texture synthesis over arbitrary manifold surfaces. In *Proceedings of SIGGRAPH* (2001), pp. 355–360.

[91] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am. 94*, 1 (1993), 111– 123.

[92] Wightman, F. L., and Kistler, D. J. Headphone simulation of free-field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am. 85*, 2 (1989), 858–867.

[93] Wightman, F. L., and Kistler, D. J. Headphone simulation of free-field listening. II: Psychophysical validation. *J. Acoust. Soc. Am. 85*, 2 (1989), 868–878.

[94] Wu, H., Siegel, M., and Khosla, P. Vehicle sound signature recognition by frequency vector principal component analysis. *IEEE Transactions on Instrumentation and Measurement 48*, 5 (Oct. 1999), 1005–1009.

[95] Zhao, H., Plaisant, C., Shneiderman, B., and Duraiswami, R. Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. In *10th International Conference on Auditory Display* (Sydney, Australia, July 2004).

[96] Zhao, H., Plaisant, C., Shneiderman, B., Zotkin, D., and Duraiswami, R. Sonification of dynamic choropleth maps: Geo-referenced data exploration for the vision-impaired. In *Proceedings of the 9th International Conference on Auditory Display* (July 2003), p. 307. abstract.

[97] Zotkin, D. N., Duraiswami, R., and Davis, L. S. Rendering localized spatial audio in a virtual auditory space. *IEEE Transactions on Multimedia 6*, 4 (Aug. 2004), 553–564.

[98] Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A. Fast head related transfer function measurement via reciprocity. *J. Acoust. Soc. Am.* (2005). submitted.