Spatial Multizone Soundfield Reproduction Design

by

Wenyu Jin

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Engineering and Computer Science.

Victoria University of Wellington
2015
Abstract

It is desirable for people sharing a physical space to access different multimedia information streams simultaneously. For a good user experience, the interference of the different streams should be held to a minimum. This is straightforward for the video component but currently difficult for the audio sound component. Spatial multizone soundfield reproduction, which aims to provide an individual sound environment to each of a set of listeners without the use of physical isolation or headphones, has drawn significant attention of researchers in recent years. The realization of multizone soundfield reproduction is a conceptually challenging problem as currently most of the soundfield reproduction techniques concentrate on a single zone.

This thesis considers the theory and design of a multizone soundfield reproduction system using arrays of loudspeakers in given complex environments. We first introduce a novel method for spatial multizone soundfield reproduction based on describing the desired multizone soundfield as an orthogonal expansion of formulated basis functions over the desired reproduction region. This provides the theoretical basis of both 2-D (height invariant) and 3-D soundfield reproduction for this work. We then extend the reproduction of the multizone soundfield over the desired region to reverberant environments, which is based on the identification of the acoustic transfer function (ATF) from the loudspeaker over the desired reproduction region using sparse methods. The simulation results confirm that the method leads to a significantly reduced number of required microphones for an accurate multizone sound reproduction compared with the state of the art, while it also facilitates the reproduction over a wide frequency range.
In addition, we focus on the improvements of the proposed multizone reproduction system with regard to practical implementation. The so-called 2.5D multizone soundfield reproduction is considered to accurately reproduce the desired multizone soundfield over a selected 2-D plane at the height approximately level with the listener’s ears using a single array of loudspeakers with 3-D reverberant settings. Then, we propose an adaptive reverberation cancelation method for the multizone soundfield reproduction within the desired region and simplify the prior soundfield measurement process. Simulation results suggest that the proposed method provides a faster convergence rate than the comparative approaches under the same hardware provision. Finally, we conduct the real-world implementation based on the proposed theoretical work. The experimental results show that we can achieve a very noticeable acoustic energy contrast between the signals recorded in the bright zone and the quiet zone, especially for the system implementation with reverberation equalization.
Declaration

The content of this thesis are the results of original research and has not been submitted for a higher degree at any other university or institution. Much of this work has either been published or submitted for publications as journal papers, conference proceedings and patents. These publications are:

**Peer-reviewed Publications:**


W. Jin, W. B. Kleijn, "Design and Implementation of Practical Personal Sound Rendering System", to be submitted to *IEEE Transactions on Audio, Speech and Language Processing*.

**Patents:**


The research represented in this thesis has been performed jointly with Prof. W. Bastiaan Kleijn. The majority, approximately 80%, of this work is my own.
Acknowledgements

First and foremost, I would like to express my sincere appreciation to my supervisor, Professor Bastiaan Kleijn for assisting and guiding me toward the completion of this thesis. Without his help, this thesis will never reach the current stage and I greatly thank him for his precious time and efforts committed to this thesis. In particular, his profound erudition, prudence and humor make him a fantastic mentor to work with. I would also like to thank the Faculty of Graduate Research for providing me with the Victoria Doctoral Scholarship.

The thesis can not be completed without me mentioning the following friends I have made during my PhD studies, they are: Bing, Fatih, Feng, Ghifary, Harsh, Kelson and Ramoni. I am very grateful for their friendship and continuous support.

Last but not the least, I would like to thank my wife Helen for her accompaniment and generous encouragement on this journey to complete the thesis. No words can express my heartfelt gratitude and love for her.
## Contents

### List of Figures

x

### 1 Introduction

1.1 Background .................................................. 1
  1.1.1 Spatial Sound .......................................... 2
  1.1.2 Reverberation ........................................ 5
1.2 Motivations and Approaches ................................. 5
1.3 Thesis Outline .............................................. 8
1.4 Summary of Thesis Contributions .......................... 10

### 2 Literature Review and Background Theory

2.1 Coordinate Systems .......................................... 14
2.2 Helmholtz Wave Equation .................................... 14
2.3 General Solution to the Helmholtz Equation .............. 15
  2.3.1 3-D Case ................................................ 15
  2.3.2 2-D Case ................................................ 17
2.4 The Green’s Function and Soundfield Superposition ...... 18
2.5 Existing Soundfield Reproduction Techniques ............ 23
  2.5.1 Wave Field Synthesis .................................. 23
  2.5.2 Higher Order Ambisonics .............................. 25
  2.5.3 Comparison of Higher Order Ambisonics (HOA)
       and Wave Field Synthesis (WFS) ...................... 27
  2.5.4 The Least Squares Method .............................. 28
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6 Acoustic Channel Effects in Complex Environments</td>
<td>33</td>
</tr>
<tr>
<td>2.6.1 Introduction to Reverberation</td>
<td>33</td>
</tr>
<tr>
<td>2.6.2 The Scattering Effects Resulted From Obstacles</td>
<td>40</td>
</tr>
<tr>
<td>2.7 Works Related to Multizone Soundfield Reproduction</td>
<td>43</td>
</tr>
<tr>
<td>2.8 Summary and Links to next Chapters</td>
<td>47</td>
</tr>
</tbody>
</table>

3 Multizone Soundfield Reproduction Using Orthogonal Basis Expansion  49

| 3.1 Introduction                                                      | 50 |
| 3.2 Problem Formulation and Notation                                  | 51 |
| 3.3 Describing Arbitrary Soundfields with a Basis Expansion           | 53 |
| 3.4 Mode Limitedness                                                  | 56 |
| 3.5 The Reproduction of the Linear-combined Plane Waves                | 57 |
| 3.6 Smooth Weighting Functions                                         | 60 |
| 3.7 Results and Discussion                                            | 62 |
| 3.7.1 Basis Function Approach                                         | 62 |
| 3.7.2 Desired Multizone Reproduction using Non-circular Loudspeaker Array | 65 |
| 3.7.3 Reproduction with Smooth Weighting Function                     | 67 |
| 3.7.4 Comparison with Prior Work                                       | 70 |
| 3.8 Conclusion and Contribution                                       | 72 |

4 Multizone Soundfield Reproduction In Reverberant Rooms Using Sparse Approximation Methods  75

| 4.1 Introduction                                                      | 76 |
| 4.2 Problem Formulation and Notation                                  | 80 |
| 4.3 Evaluation Measures                                               | 80 |
| 4.4 Green’s Function Modeling                                         | 81 |
| 4.4.1 Estimation of the ATFs in Reverberant Rooms                     | 82 |
| 4.4.2 Approximate Cramér-Rao Bound for General Green’s Function Modeling Approach | 87 |
| 4.5 Orthogonal Basis Function Approach                                | 90 |
## CONTENTS

4.6  Extension to 3-D Reproduction ........................................ 92
4.7  Results and Discussion .................................................. 94
    4.7.1  Experimental Setup ............................................. 94
    4.7.2  Estimation of the Desired ATFs ............................... 95
    4.7.3  The Cramér-Rao Bound ........................................ 99
    4.7.4  Desired 2-D Multizone Soundfield Reproduction ............ 102
    4.7.5  On the Multizone Soundfield Reproduction with Practical Settings ........................................ 105
    4.7.6  Desired 3-D multizone soundfield reproduction ........... 109
4.8  Conclusion and Contribution ............................................ 110
4.9  Appendix: Algorithms for the Reconstruction of the Sparse
    signals ........................................................................ 111

5  Design of Practical Multizone Soundfield Reproduction .......... 117
    5.1  Introduction .......................................................... 118
    5.2  Overview .............................................................. 123
    5.3  2.5D Multizone Soundfield Reproduction Using Sparse Methods .................................................. 123
    5.4  Adaptive Reverberation Cancellation For Multizone Soundfield Reproduction .................................. 127
        5.4.1  Soundfield Basis Expansion ................................. 128
        5.4.2  Soundfield Characterization Using Sparse Methods ............................................................... 128
        5.4.3  Channel Estimation and Reverberation Cancellation .............................................................. 130
        5.4.4  Room Channel Modeling and Estimation .................. 130
        5.4.5  Loudspeaker Updating Signals .............................. 131
    5.5  Results and Discussion .................................................. 133
        5.5.1  Adaptive Reverberation Cancellation for Multizone Soundfield Reproduction .......................... 136
        5.5.2  2.5D Multizone Soundfield Reproduction ............... 139
    5.6  Conclusion and Contribution ............................................ 143
CONTENTS

5.7 Appendix: Investigation on the Required Number of Loudspeakers For 2.5D Reproduction ........................................... 145

6 Real-world Implementation of Multizone Soundfield Reproduction .............................................................................. 149
  6.1 Introduction ............................................................................................................................................. 150
  6.2 Overview .............................................................................................................................................. 151
  6.3 Implementation of Multizone Soundfield Reproduction with Free-field Assumption ........................................... 153
    6.3.1 Experimental Setup ...................................................................................................................... 153
    6.3.2 Loudspeaker Gain Calibration ........................................................................................................ 154
    6.3.3 Experimental Procedure – Narrowband Case ............................................................................. 156
    6.3.4 Experimental Procedure – Wideband Case ................................................................................. 159
    6.3.5 Performance Analysis ...................................................................................................................... 162
  6.4 Implementation of Multizone Soundfield Reproduction with Reverberation Equalization .................................... 165
    6.4.1 Real-world Estimation of Desired Loudspeaker ATF ..................................................................... 167
    6.4.2 Multizone Soundfield Reproduction with Reverberation Equalization – Narrowband .............. 171
    6.4.3 Comments on Multizone Soundfield Reproduction with Reverberation Equalization – Wideband ...... 175
  6.5 Conclusion and Contribution ......................................................................................................................... 177

7 Conclusions ......................................................................................................................................................... 179
  7.1 Conclusion .................................................................................................................................................. 179
  7.2 Future Research Directions ........................................................................................................................ 182
List of Figures

1.1 Demonstration of three types of spatial sound systems: i) Binaural System, ii) Transaural System and iii) Soundfield Reproduction. ......................................................... 3

1.2 Thesis structure and research areas. Main research contributions and the possible future research directions of this thesis are indicated in solid blocks and dashed blocks respectively. ................................................................. 11

2.1 The position of a source in (a) 3-D and (b) 2-D coordinate systems. ................................................................. 14

2.2 Sound transmission in a room is modeled as an linear time invariant system. ................................................................. 34

2.3 Demonstration of image source method for modeling the 2-D rectangular room reverberation. The solid rectangle in the center represents the room, the real source is displayed as (●) and the image sources are denoted as (◦). The lattice is formed by infinitely repeating the block bounded by dashed lines. ................................................................. 36

2.4 Soundfield reproduction around a human head. ............ 41

3.1 The 2-D multizone soundfield reproduction in the desired reproduction region of radius $r$ with $Q$ loudspeakers. ...... 52
3.2 Describing the desired multizone soundfields with a basis expansion. (a) and (b) are for the case when $\phi_d = 45^\circ$ and $\phi_d = 60^\circ$ respectively. ..................................................... 62

3.3 Describing the desired soundfield due to a virtual point source within the bright zone using the basis expansion. (a) and (b) represents the real and imaginary part, respectively. “o” marks the position of the virtual source. ................................. 63

3.4 Performance as the desired plane wave is panned, comparing reproductions as parameters $b$ are varied (solid curves), at different frequencies (dashed). Plotted are (a) an multizone sound reproduction design with one bright zone and one quiet zone (b) the MSE in $D_b$ and (c) acoustic contrast between the bright and quiet zone. ................................. 64

3.5 Desired multizone reproduction using the approach of part-circle with an angular window of $\phi_0 = \pi$. 39 loudspeakers are used and the red circles demonstrate the positions of loudspeakers. (a) and (b) represent the cases with $\phi_d = 45^\circ$ and $\phi_d = 60^\circ$ respectively. ..................................................... 65

3.6 Desired multizone reproduction using a square array of loudspeakers. 75 loudspeakers are used and the red circles demonstrate the positions of loudspeakers. (a) and (b) represent the cases with $\phi_d = 45^\circ$ and $\phi_d = 60^\circ$ respectively. ................................. 66

3.7 Multizone Sound reproduction using the proposed smooth weighting function at 1500 Hz. (a) maps the generated smooth weighting function, and (b) illustrates the corresponding reproduced soundfield. ..................................................... 68

3.8 Two complementary scenarios to Fig. 3.7(b): Fig. 3.8(a) shows the situation where the bright and quiet zone in Fig. 3.7(b) are exchanged for $\phi_d = 30^\circ$, while Fig. 3.8(b) features two bright zones and the two plane waves with the arriving direction of $60^\circ$ and $30^\circ$ are superimposed. ................................. 69
3.9 Multizone sound reproduction using a semi-circle loudspeaker array at the frequency of 1500 Hz. (a) demonstrates the reproduction with the proposed smooth weighting function, and (b) shows the reproduction with the previous weighting function which assigns a fixed value to each of the pre-defined zones. 70

3.10 Wide-band multizone soundfield reproduction using the proposed basis function approach and the convex optimization method in [1]. 71

4.1 The 2-D multizone soundfield reproduction over the desired reproduction region using $Q$ loudspeakers in a reverberant rooms. 79

4.2 The estimated performance of the desired loudspeaker ATF is plotted as a function of the number of microphones used at 1 kHz with different settings. We ran 100 trials for each iteration with an assigned value of $m$. 96

4.3 Wide-band desired ATF estimation with 20 noisy measurements, using the proposed method and the method in [2]. Estimation error curves have been averaged over 50 trial runs. 97

4.4 MSE over the selected estimation region with various variances of the positioning inaccuracy. Estimation error curves have been averaged over 50 trial runs. 98

4.5 The CRB on the variance of $\hat{R}_q(x, k)$ to the desired ATF over the selected zones. We examined the white-noise setting with the variance $\sigma^2 = 5 \times 10^{-4}$ with 15 (left) and 16 (right) measurements. The white crosses represents the position of the microphones. 100
LIST OF FIGURES

4.6 Reproduction of the desired multizone sound using 16 noiseless measurements for each selected zone in a reverberant room. (a) and (b) demonstrate the case for $\gamma = 0$ and $\gamma = 0.02$ respectively. The red crosses represent the positions of the microphones. 102

4.7 Reproduction of the desired multizone sound with various values of the regularization parameter. 104

4.8 Wide-band multizone soundfield reproduction with 64 noiseless pressure samples, using our proposed method, the reproduction method in [3] and the multi-point method in [4]. 105

4.9 The estimated performance of the desired loudspeaker ATF is plotted as a function of the number of microphones used at 1 kHz (left) and 5 kHz (right) with single-wall setting. . 106

4.10 The required number of measurements for 100% accurate estimation of the desired loudspeaker ATF with the single-wall setting. 106

4.11 Reproduction of the desired three-zone soundfield using 7 microphones for each selected zone in a reverberant room. (a) and (b) demonstrate the real part and imaginary part of the soundfield respectively. The red crosses represent the positions of the microphones. 107

4.12 3-D Reproduction of the desired multizone sound using 30 microphones for each selected zone in a reverberant room at 500 Hz. (a) and (b) demonstrate the real part and imaginary part of the soundfield respectively. The results show the pressure field at the cut slice of $z = 1.2$ m in Cartesian coordinates. 108

5.1 Overview of the 2.5D active reverberation cancellation system including the loudspeaker array and microphone configurations. 122
LIST OF FIGURES

5.2 The model of an 2.5D multizone soundfield reproduction system in reverberant environments. 124

5.3 Reproduction of the desired multizone sound using 20 noisy measurements for each selected zone in a reverberant room. (a) and (b) demonstrate the real and imaginary part respectively. The red crosses represent the positions of the microphones. 134

5.4 Equalization error at the microphone measurements vs. the number of adaption steps. The results are averaged over 10 trial runs. 135

5.5 Comparison of reproduction performance between our method and the approach in [5]. (a) and (b) represent the performance of acoustic contrast between $D_b$ and $D_q$ and MSE over $D_b$ respectively. 137

5.6 Wide-band multizone soundfield reproduction with 64 noisy pressure samples, using the proposed method, the adaptive reproduction method in [5] and the non-adaptive method in [6]. The results are averaged over 10 trial runs. 138

5.7 2.5D Reproduction of the desired multizone sound using 24 microphones for each selected zone in a reverberant room at 1000 Hz. The red crosses represent the location of the measurements. (a) and (b) demonstrate the non-regularized and the regularized results respectively. 139

5.8 2.5D Reproduction of the desired multizone sound with loudspeakers located at a different height in a reverberant room at 1000 Hz. The red crosses represent the location of the measurements. (a) and (b) demonstrate the non-regularized and the regularized results respectively. 140
5.9 The variation of the soundfield above and underneath (within a reasonable range) the desired horizontal region. (a) and (b) demonstrate the vertical consistency over the bright zone and the quiet zone respectively. ........................................ 141

5.10 2.5D Reproduction of the desired multizone sound using 25 noisy measurements for each selected zone in a reverberant room. (a) and (b) demonstrate the real and imaginary part respectively. The red crosses represent the positions of the microphones. .................................................. 143

6.1 Illustration of basic multizone soundfield rendering system implementation. ......................................................... 152

6.2 Arrangement of loudspeakers and microphones in the VUW electroacoustic lab. From [7], with permission. ................. 154

6.3 Arrangement of setup of loudspeaker array. From [7], with permission................................................................. 155

6.4 Soundfield reproduction of 1000 Hz signal at desired region by 24 loudspeakers. From [7], with permission. ................. 157

6.5 Recorded signals at $q_1$, $q_2$, $q_3$ and $b$. From [7], with permission. ............................................................. 158

6.6 The performance in terms of the acoustic contrast between the bright zone and quiet zone are plotted for the centered frequencies. ................................................................. 161

6.7 The simulated measurements of the sound at the centers of the bright zone and the quiet zone are plotted. .................. 162

6.8 The real measurements of the sound at the centers of the bright zone and the quiet zone are plotted. ......................... 163

6.9 Soundfield reproduction of 1000 Hz signal at desired region by 24 loudspeakers given the measured the reflection coefficients and the geometry of the VUW electroacoustic lab. From [7], with permission. ............... 164

6.10 The media lab in Huawei European Research Center (Munich). ................................................................. 166
6.11 Microphone arrangements for the loudspeaker ATF estimation experiment. The red crosses represent the locations of microphone measurements. 167
6.12 Room impulse response of the testing loudspeaker. 169
6.13 Average MSE to the the actual measured loudspeaker ATFs at 18 selected points from 100 to 1 kHz. 169
6.14 The measured and estimated frequency response at the point 17 and 18. 170
6.15 System layout and microphone arrangement. The red crosses represent the position of microphone measurements. 171
6.16 Performance of the acoustic contrast between the bright zone and quiet zone at four different pairs of locations. (a) represents the acoustic contrast between (0.6,0) and (-0.6,0) in Fig 6.15, (b) represents the acoustic contrast between (-0.7,-0.25) and (0.7,-0.25), (c) represents the acoustic contrast between (-0.65,0.15) and (0.65,0.15) and (d) represents the acoustic contrast between (-0.42,-0.2) and (0.42,-0.2). 172
6.17 The acoustic contrast between the center of the bright zone and quiet zone at the height level of 0 cm (a) and -6 cm (b). 173
Overview: In this chapter, we present an introduction to the problems the thesis aims to solve. It is followed by the research motivations, a statement of approaches adopted, the goals of the research and finally the outline of the thesis.

1.1 Background

In physics, sound is a vibration that propagates as a typically audible traveling wave with an oscillation of pressure, through a medium such as air, water or a solid. In air, sound waves can be measured using pressure sensors, which we refer to as microphones. Therefore, a soundfield can be considered as a pressure field, which is measured by microphones in practice.

Sound reproduction systems that aim to provide high-quality playback of sound and audio streams have been continuously developed over the last century. The original systems consisted of only one loudspeaker, which are referred as monophonic systems. A limited extent of spatial impres-
sion can be recreated using monophonic systems. It was followed by the stereophonic reproduction system in 1930’s, in which the spatial effect was created based on amplitude differences between the loudspeakers [8]. The mainstream of current commercial surround sound systems include Dolby Digital [9] and DTS\textsuperscript{1} [10], which have been widely employed in home-based entertainment systems, cinemas, personal computers and various portable devices. Nowadays, the technology of spatial sound reproduction has also been applied in many realms, such as medicine [11] [12], education, data visualization [13] [14], virtual reality [15] [16], and various entertainment applications [17].

1.1.1 Spatial Sound

Spatial sound is the subtle characteristic of a soundfield that provides awareness of the environment and a sense of presence. More specifically, it creates the impression that the sounds are coming from sound sources placed anywhere in the space [18]. Currently, research on spatial sound mainly focuses on two categories: binaural/transaural systems and systems based on spatial soundfield reproduction. A demonstration of the three spatial sound systems are given in Fig. 1.1 and a brief introduction to these two categories of spatial sound approaches is listed in this section.

Binaural literally means “having or relating to two ears”. Binaural hearing, along with frequency cues, allows humans to determine the perception of complex spatial sound [19]. The sound is delivered directly to the ears through headphones in typical binaural systems (as shown in Fig. 1.1(i)), in which only two channels of information are considered and have a small sound reconstruction region.

Transaural systems generally deliver the binaural sound signals through a small number of speakers (as shown in Fig. 1.1(ii)), which implies

\textsuperscript{1}DTS is the trademark for the digital audio format created by Digital Theater System Inc.
that only the soundfield at the location of user’s ear is rendered [20].
The main challenges of the transaural systems are due to the unique
HRTFs (Head Related Transfer Functions) for various individuals [21–23].
Additionally, such systems are very sensitive to the listener’s position and
performance is significantly undermined when room echoes are involved
[4]. At present, transaural systems can only be effective if the systems are
uniquely designed and tailored for individual users, which hinders the
approach from widespread adoption.

The principle of spatial soundfield reproduction systems is to recreate the
desired soundfield\(^2\) over an extended region of interest using loudspeaker
arrays (as shown in Fig. 1.1(iii)). Soundfield reproduction systems are
generally well suited for the delivery of a surround sound over a large
audience as it is independent of the listeners and their HRTFs [18]. In this
thesis, the main focus is on spatial soundfield reproduction.

Although the 5.1 multichannel audio system has been a consolidated

---

\(^2\)The desired soundfield can be any arbitrary wavefield that is physically feasible and
satisfies the wave equation.
1.1. BACKGROUND

standard in surround sound nowadays, there is an increasing interest in emerging rendering systems based on spatial soundfield reproduction. Various, often-related techniques exist for soundfield reproduction techniques, including wave field synthesis (WFS), ambisonics and the least squares methods.

WFS was initially introduced by Berkhout et al. [24] [25], who used the method of “acoustic holography” to reproduce a desired soundfield over a relatively large area with highly intense secondary source distributions. The WFS approaches [26–28] are based on the Kirchhoff-Helmholtz integral [29] [30], which states that a continuous distribution of appropriately driven secondary monopole and dipole sources (e.g. loudspeakers) arranged on the boundary of the desired listening area is capable of reproducing any virtual wave field inside that area but zero wavefield outside [29].

Ambisonics, was initially developed by Gerzon in 1973 [31], who focused on the zeroth and first order spherical harmonic decomposition of the original soundfield at a single point in space. Ambisonics was further developed as higher order ambisonics (HOA) for larger reproduction areas [32] [33]. The principle of the Ambisonics system is based on matching the desired soundfield with the soundfields resulted from the secondary sources (i.e. loudspeakers) in terms of the harmonics-based expansion for 3-D (spherical harmonics representation) or 2-D (cylindrical harmonics representation) cases [29] [34]. This is called the “mode-matching” approach [34].

The least squares method, is another popular method for the reproduction of soundfields. In the 1990s, Kirkeby and Nelson [35] [36] proposed the least squares techniques to theoretically determine the minimum number of loudspeakers required to recreate a local soundfield. Recently, the least squares method has been widely used in active sound control applications [5,37,38]. However, this approach involves a matrix inversion. The solution may not exist if the matrix is poorly conditioned [39]. A discussion of
possible solutions to this ill-conditioning problem is presented in [39].

1.1.2 Reverberation

The above mentioned techniques were usually implemented under the free-field assumption, which is an idealized description of a reflection-free environment. However, the received sound from microphone recordings generally includes a component that is indirectly propagated from the sources in a typical listening environment where the sound is scattered off various surfaces (e.g. the room walls and scatterers) in the environment. The collective effect of the sound interaction with the surroundings is known as reverberation [40].

Reverberation resulting from the environment usually leads to inferior performance compared with ideal reproduction of the desired soundfield using the free-field assumption. In general, methods that aim to minimize the reverberation effects of the soundfield reproduction can be categorized into three groups [41] [42]: passive techniques that reduce reflections, equalization approaches based on models of the reverberant soundfield or channel behavior, and adaptive room cancellation methods.

The soundfield reproduction methods and reverberation compensation approaches is investigated in details in Chapter 2.

1.2 Motivations and Approaches

It is desirable for people sharing a physical space to access various multimedia information streams simultaneously. This is straightforward for the video component but is currently very difficult for the audio sound component. Spatial soundfield reproduction with a multizone setting over an extended spatial region has drawn researchers’ attention in recent years. At present, the majority of the existing spatial soundfield reproduction techniques focus on a single zone, which feature a single sound
environment inside a physical enclosure with the employed loudspeaker array.

Multizone soundfield reproduction is an advanced application of spatial sound that aims at providing an individual sound environment to each of a set of listeners without physical isolation or the use of headphones. Potential applications of spatial multizone soundfield reproduction include but are not limited to:

- simultaneous entertainment systems in cars;
- personal audio system on portable devices;
- The creation of individual quiet zones in noisy environments (e.g. passenger planes);
- surround sound systems in performance venues.

The realization of such multizone soundfield reproduction systems is a conceptually challenging problem and only few works related to this area have been reported in the previous literature [3, 4, 43–47]. It is usually more complicated than other signal processing problems due to the multi-frequency nature of sound. In practical spatial multizone soundfield reproduction, the performance is always adversely affected due to the limitation on the hardware provision. Furthermore, the system performance can be exacerbated by unintended interference and reverberant environments in the real world.

The overall goal of this thesis is to develop a practical multizone soundfield reproduction system that enables users to control the reproduction of the desired multizone soundfield over each of a set of selected zones using a loudspeaker array under complex environments (e.g. reverberant environments). We divide this problem into three further objectives:

- to find the best trade-off between precisely reproducing the desired soundfield in the pre-defined acoustic "bright" zones and the control of sound leakage into other zones;
• to consider the multizone soundfield reproduction under complex environments such as reverberant environments;

• to facilitate the practical implementation of the proposed system, i.e. the reduction of hardware requirements, stable reproduction performance over a wide frequency range, etc.

In this thesis, we start with a novel method for spatial multizone soundfield reproduction by describing the desired multizone soundfield as an orthogonal expansion of basis functions over the desired reproduction region. The proposed soundfield basis expansion enables us to control the reproduction of the desired soundfield over each of a set of pre-defined zones individually. It reduces the complicated multizone sound reproduction problem to the recreation of a set of elementary wavefields over the desired region. Then, we extend the orthogonal basis function approach to the case with reverberant settings, which is based on the identification of the acoustic transfer function (ATF) between the loudspeaker over the desired reproduction region using a limited number of microphones. We assume that the soundfield is sparse in the domain of plane wave decomposition and identify the ATF using sparse methods. Simulations confirm that the method leads to a significantly reduced number of required microphones for accurate multizone sound reproduction compared with the state-of-the-art techniques, while it also facilitates the reproduction over a wide frequency range. Additionally, we improve the practical implementation of the multizone soundfield reproduction system by incorporating the concept of the so-called 2.5D reproduction and active echo cancellation. Finally, the performance of the proposed systems and techniques have been verified through real-world implementation.
1.3 Thesis Outline

The content of this thesis is summarized chapter by chapter, as follows

**Chapter 2**: This chapter presents the literature review and background theory on spatial soundfield reproduction. The introduction of the acoustic wave equation and the *Helmholtz* equation provides a foundation for further discussions. Existing soundfield reproduction techniques such as WFS, HOA and the least squares method are investigated. The potential ill-conditioning problems in the spatial soundfield reproduction are introduced and possible solutions are discussed. We then investigate the acoustic channel effects that alter the reproduced soundfield in complex environments. Discussions on techniques to improve the soundfield reproduction performance under those circumstances are presented. A literature review on existing multizone sound systems is also presented. The chapter ends with a summary and links to the following chapters.

**Chapter 3**: In this chapter, we introduce a method for 2-D spatial multizone soundfield reproduction based on describing the desired multizone soundfield as an orthogonal expansion of basis functions over the desired reproduction region. This approach finds the physically feasible solution that is closest to the desired soundfield in a weighted least squares sense. The basis orthogonal set is formed using QR factorization with a suitable set of plane wave functions as input. The coefficients of the plane waves can then be calculated. The method facilitates its application with a practical loudspeaker configuration. The proposed soundfield basis expansion, which provides the theoretical basis for an efficient parametrization of any feasible soundfield throughout the thesis, enables us to control the reproduction of the desired soundfield over each of a set of pre-defined zones individually.

**Chapter 4**: In this chapter, we introduce a method of reproducing a multizone soundfield within a desired region in reverberant environments. It is based on the identification of the acoustic transfer function (ATF)
between the loudspeakers and the desired reproduction region using a limited number of microphones. The Cramér-Rao bound on the variance of the desired ATF estimator using sparse microphones is also derived. The ATFs are then used to derive the optimal least-squares solution for the loudspeaker filters that minimize the reproduction error over the entire reproduction region. Simulations confirm that the method leads to a significantly reduced number of required microphones for accurate multizone sound reproduction compared with the prior art, while it also facilitates the reproduction over a wide frequency range. Extension to the 3-D multizone soundfield reproduction is also discussed.

Chapter 5: In this chapter, we mainly focus on improvements of the proposed multizone reproduction system with regard to practical implementation. We first discuss the implementation of multizone sound reproduction over a desired horizontal region at the same level as the listener’s ears using a circular loudspeaker array located in the same or a different plane, in 3-D reverberant settings. Comparing with the 2-D case, this so-called 2.5D multizone soundfield reproduction features more accurate models to represent characteristics of the employed loudspeakers, as well as the reverberation characteristics in practice. Secondly, we propose an adaptive reverberation cancellation method for the multizone soundfield reproduction within a desired region to reduce the complexity of the soundfield measuring process. We consider the reproduced soundfield as a linear transformations of the desired soundfield, where the basis function coefficient set for the desired soundfield can be computed under the free-field assumption. We then introduce the adaptive channel estimation process using sparse methods to identify these transformation and derive the required loudspeaker updating signals. Comparing with the classical cylindrical/spherical harmonics-based systems, the proposed method facilitates a more accurate multizone soundfield reproduction, as well as a faster convergence rate given the same provision of hardware.
Chapter 6: In this chapter, we apply our proposed methods from theory to real-world implementation. The experimental work was conducted in the Electroacoustic Lab of the Victoria University of Wellington (VUW), as well as in the media lab of Huawei European Research Center. We start with the implementation in a normal listening room based on the system design with ideal free-field settings. It can be clearly seen that the room reverberation undermines the reproduction performance with free-field design, which is an ideal setting for the soundfield rendering. We then extend the multizone soundfield system with the reverberation equalization method (as discussed in Chapter 4) based on feedbacks from a limited number of microphone measurements. The chapter is ended with an analysis on the possible factors that limit the system performance.

Chapter 7: This chapter presents the general conclusions that are drawn from the study of this research project. Additionally, we list some future research directions arising from this work.

Fig. 1.2 summarizes the thesis structure and research areas. The main research contributions and the possible future research directions are denoted in solid blocks and dashed blocks, respectively.

1.4 Summary of Thesis Contributions

The key contributions of the thesis are:

- Formulation of the set of basis functions that is orthogonal over the desired reproduction region using a modified Gram-Schmidt procedure with a suitable set of plane wave functions as input. The proposed soundfield basis expansion enables us to control the reproduction of the desired soundfield over each of a set of predefined zones individually.

- A generalization of the reproduction method proposed in [48] that facilitates a reduced number of loudspeakers in a practical con-
Figure 1.2: Thesis structure and research areas. Main research contributions and the possible future research directions of this thesis are indicated in solid blocks and dashed blocks respectively.

- The identification of the acoustic transfer function from the loudspeakers over the desired reproduction region using the concept of sparse approximation, which facilitates a significant reduction on the required number of microphones, flexible microphone placements within the desired region, and a stable estimation performance over a wide frequency range.
- The derivation of the Cramér-Rao lower bound on the variance of the...
desired soundfield estimator with a small number of microphones.

- A generalized framework on 2-D multizone soundfield reproduction in reverberant rooms using a limited number of microphones is proposed based on the orthogonal basis function approach. The extensions to 3-D multizone soundfield reproduction using a spherical loudspeaker array in reverberant environments are presented.

- Design of the 2.5D multizone soundfield reproduction over a desired horizontal region using a circular array of loudspeakers located in the same plane or a plane at different height, in a 3-D reverberant environment (i.e. the reverberation from the floor and ceiling is also taken into account).

- Adaptive reverberation cancellation method for the multizone soundfield reproduction within the desired region that allows a parallel implementation and does not require a prior measurement of the loudspeaker ATFs over the desired region. It further improves the reproduction performance of practical applications. The 2.5D multizone soundfield reproduction with the adaptive reverberation cancellation system is also included.

- Real-world implementation of the proposed multizone soundfield system under the free-field assumption in a non-anechoic environment. The experiments are conducted for both narrowband and wide-band cases (up to 4 kHz).

- Real-world experiments of the loudspeaker ATF estimation based on sparse approximation methods in a typical listening environment.

- Real-world implementation of the proposed multizone soundfield system in a real listening room with the reverberation equalization design based on a limited number of microphone measurements. The experiments are conducted at the frequency of 1 kHz.
Overview: This chapter provides a comprehensive literature review and background theory on spatial soundfield reproduction. We first introduce the Helmholtz equation and the general solutions to the homogeneous wave equation. These form the foundations for describing any physically feasible soundfield and the effects of the acoustic channel, which can be modeled as an expansion of the solutions to the Helmholtz equation. Existing soundfield reproduction techniques such as wave field synthesis (WFS), higher order ambisonics (HOA) and the least squares method are then presented. The techniques to improve the soundfield reproduction performance in complex environments are investigated. A review of existing multizone soundfield rendering systems is also presented.
2.1 Coordinate Systems

Both three dimensional (3-D) and two dimensional (2-D) coordinate systems are used throughout this thesis. For the 3-D case, the spherical coordinate system is used as shown in Fig. 2.1(a). We can express the translation between the spherical coordinates \((r, \theta, \phi)\) (where \(r\) is the distance from the origin, \(\theta\) is the elevation angle and \(\phi\) is the azimuth angle) of the position vector \(\mathbf{x}\) and the Cartesian coordinates \((x, y, z)\) as \((x, y, z) = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)\). Similarly, the coordinates are represented using a polar-coordinate system for 2-D case as shown in Fig. 2.1(b). The conversion between polar coordinates \((r, \phi)\) of a position vector \(\mathbf{x}\) and Cartesian coordinates \((x, y)\) can be written as: \((x, y) = (r \cos \phi, r \sin \phi)\), where \(r\) is the distance from the origin and \(\phi\) is the polar angle.

![Coordinate Systems Diagram](image)

Figure 2.1: The position of a source in (a) 3-D and (b) 2-D coordinate systems.

2.2 Helmholtz Wave Equation

During the propagation of soundwaves, the instantaneous sound pressure varies around the mean pressure. We can model the soundwave prop-
agation as \( p(x, t) \), which represents the sound pressure recorded at the position \( x \) and time \( t \) given an acoustic event in the space. This is the solution of the classical acoustic wave equation in the absence of sources [49] given by:

\[
\nabla^2 p(x, t) - \frac{1}{c^2} \frac{\partial^2 p(x, t)}{\partial t^2} = 0
\]

(2.1)

where \( c \) represents the speed of sound and \( \nabla^2 \) is the Laplacian operator. By separating the time dependent solution \( e^{i\omega t} \), it leads to the homogenous Helmholtz Equation [49]:

\[
\nabla^2 p(x) - k^2 p(x) = 0
\]

(2.2)

where \( \omega \) is the radian frequency and \( k = \frac{\omega}{c} = \frac{2\pi f}{c} \) is the wave number. For simplicity, \( k \) is used to represent frequency throughout this thesis as we assume the speed of sound \( c \) is a constant.

### 2.3 General Solution to the Helmholtz Equation

The general solutions to the Helmholtz equation can be used to characterize any feasible soundfield as the pressure and velocity of a homogeneous wavefield following Eq. (2.2). In this section, the general solutions to the Helmholtz equation are derived for both the 3-D and 2-D cases.

#### 2.3.1 3-D Case

We can express the Helmholtz equation (2.2) for 3-D case with respect to spherical coordinates using the appropriate \( \nabla^2 \) [50]:

\[
\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} p(x, t) \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} p(x, t) \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} p(x, t) - \frac{1}{c^2} \frac{\partial^2 p(x, t)}{\partial t^2} = 0.
\]

(2.3)
2.3. GENERAL SOLUTION TO THE HELMHOLTZ EQUATION

By taking the Fourier transform of the solution to (2.3), the spatial solution to the 3-D Helmholtz wave equation can be written as an expansion of spherical harmonics, which is an orthogonal basis set for characterizing any arbitrary soundfield inside the source-free spatial region

\[ S(x, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kx) Y_{nm}(\hat{x}), \quad (2.4) \]

where \( \alpha_{nm} \) are the coefficients for the spherical harmonics \( Y_{nm}(\hat{x}) \). Note that \( Y_{nm}(\hat{x}) \) are functions of the unit vector \( \hat{x} \) in the direction of position vector \( x \). The subscripts \( n \) and \( m \) denote the order and the mode of the spherical harmonics, respectively. The mode \( m \) varies over the range from \(-n\) to \( n\) for each order \( n \). \( j_n(\cdot) \) is the \( n \)th order spherical Bessel function of the first kind.

The spherical harmonics are defined as [51]:

\[ Y_{nm}(\hat{x}) = \sqrt{\frac{(2n + 1)(n - |m|)!}{4\pi(n + |m|)!}} P_{n|m|}(\cos \theta)e^{im\phi}, \quad (2.5) \]

where \( P_{n|m|}(\cdot) \) is the associated Legendre function. According to [52], the associated Legendre function features the orthogonality property:

\[ \int_{-1}^{1} P_{n|m|}(\cos \theta)d(\cos \theta) = \frac{2(n + |m|)!}{(2n + 1)(n - |m|)!}\delta_{nm}. \quad (2.6) \]

Additionally, the exponential function also has the orthogonality property:

\[ \int_{0}^{2\pi} e^{im\phi}e^{im'\phi}d\phi = 2\pi\delta_{mm'}. \quad (2.7) \]

Overall, we have the following orthogonality property for the spherical harmonic functions \( Y_{nm}(\hat{x}) \):

\[ \int Y_{nm}^{*}(\hat{x})Y_{pq}(\hat{x})d(\hat{x}) = \delta_{np}\delta_{mq}. \quad (2.8) \]
CHAPTER 2. LITERATURE REVIEW AND BACKGROUND THEORY

Note that integration is over the unit sphere. Using this orthogonality property, the harmonic coefficients for the sound-field within a source-free region of radius \( r \) can be derived

\[
\alpha_{nm}(k) = \frac{1}{j_n(kr)} \int Y_{nm}^*(\hat{x}) S(x, k) d(\hat{x}), \tag{2.9}
\]

where the integration is over a sphere of radius \( r \). Note that (2.9) is only valid if \( j_n(kR) \neq 0 \) for the wavenumber \( k \), otherwise the harmonic coefficients at those invalid frequencies (i.e. the vicinity of \( j_n(kR) = 0 \)) cannot be determined.

2.3.2 2-D Case

Similar to the 3-D case, we can express the Helmholtz Equation for 2-D case in terms of polar coordinates using the appropriate \( \nabla^2 \) [51]:

\[
\frac{\partial^2 p(x, t)}{\partial r^2} + \frac{1}{r} \frac{\partial p(x, t)}{\partial r} + \frac{1}{r^2} \frac{\partial^2 p(x, t)}{\partial \phi^2} - \frac{1}{c^2} \frac{\partial^2 p(x, t)}{\partial t^2} = 0, \tag{2.10}
\]

The spatial solution to the 2-D Helmholtz wave equation can be decomposed into cylindrical harmonics:

\[
S(x, k) = \sum_{m=-\infty}^{+\infty} \alpha_m(k) J_m(\|x\|) e^{im\phi}, \tag{2.11}
\]

where \( \alpha_m \) are the coefficients for the Fourier-Bessel series expansion and \( J_m(\cdot) \) is the \( m \)th order Bessel function of the first kind. The 2-D soundfield coefficients can be determined by using the orthogonality property of the exponential function:

\[
\alpha_m(k) = \frac{1}{2\pi J_m(kR)} \int_0^{2\pi} S(x, k) e^{-im\phi} d\phi, \tag{2.12}
\]

where \( J_m(kR) \neq 0 \).
The modal expansion (2.11) provides an efficient parametrization of the soundfield, which has been widely used in the prior art [2, 46, 48]. However, it can be seen that it leads to problems for systems with a wide frequency range as the coefficients for the defined basis function cannot be determined at a number of frequencies due to the zeros of the Bessel function terms.

2.4 The Green’s Function and Soundfield Superposition

The Green’s function represents the acoustic transfer function (ATF) of the channel between the source and any other location in the free-field space. In a 3-D space, we define \( p(x, y, z, t) \) as the sound pressure recorded at the position \((x, y, z)\) and time \(t\) given an acoustic event in the space. It is the solution of the scalar acoustic-wave equation given by

\[
\nabla^2 p(x, y, z, t) - \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2} = s(x, y, z, t). \tag{2.13}
\]

Let a point source emit a Dirac pulse at position \((x_s, y_s, z_s)\) at instant \(t_s\), the soundfield resulted from this source can be expressed as

\[
\nabla^2 p(x, y, z, t) - \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2} = \delta(x - x_s)\delta(y - y_s)\delta(z - z_s)\delta(t - t_s). \tag{2.14}
\]

For simplicity, we assume the source position is \(s = (0, 0, 0)\) and it emits the pulse at \(t_s = 0\). The expression (2.14) can be simplified as

\[
\nabla^2 p(x, t) - \frac{1}{c^2} \frac{\partial^2 p(x, t)}{\partial t^2} = \delta(x)\delta(t). \tag{2.15}
\]
where \( x = (x, y, z) \). Taking the Fourier transform in time in (2.15) leads to

\[
\nabla^2 \tilde{p}(x, k) - k^2 \tilde{p}(x, k) = \delta(x),
\]

(2.16)

where \( \tilde{p}(x, k) \) represents the Fourier transform of \( p(x, t) \). Using spherical coordinates, we can rewrite (2.16) as

\[
\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d \tilde{p}(r, k)}{dr} \right) + k^2 \tilde{p}(r, k) = \frac{\delta(r)}{4\pi r^2}.
\]

(2.17)

Note that \( \tilde{p}(x, k) \) has been replaced with \( \tilde{p}(r, k) \) as \( \tilde{p}(x, k) \) only depends on \( r = \|x\| \) due to symmetry.\(^1\)

Multiplying both sides by \( r \) and considering \( r > 0 \) leads to

\[
\frac{d^2}{dr^2} [r \tilde{p}(r, k)] + k^2 [r \tilde{p}(r, k)] = 0.
\]

(2.18)

Eq. (2.18) represents the one dimensional Helmholtz equation whose solution is known as the d’Alembert solution [53]

\[
\tilde{p}(r, k) = ae^{-ikr} + be^{ikr},
\]

(2.19)

where \( a \) and \( b \) are undetermined constants. The second term in (2.19) can be omitted in this case as the wave is assumed to travel only towards \( +r \) direction from the source that is placed at the origin. The value of \( a \) can be obtained by substituting (2.19) into (2.16) and then integrating (2.16) within a sphere \( V \) of radius \( r \) that includes the origin. The result of

\[
a = \frac{1}{4\pi}
\]

can be obtained if we take the limit \( r \to 0 \).\(^2\)

\(^1\)We assume free-field settings.

\(^2\)For ease of demonstration, we skip this derivation and readers can refer to [53] for more details.
Therefore, (2.19) becomes

\[ \tilde{p}(x, \omega) = \frac{e^{-ikr}}{4\pi r}. \]  

(2.20)

Replacing \( r \) with the norm \( \|x\| \) that represents distance between the source and the observation point in (2.20), we have

\[ \tilde{p}(x, \omega) = \frac{e^{-ik\|x\|}}{4\pi \|x\|}. \]  

(2.21)

In the case that the source is not located at the origin but at position \( s \), we have \( r = \|x - s\| \), which leads to

\[ \tilde{p}(x, \omega) = \frac{e^{-ik\|x-s\|}}{4\pi \|x - s\|}. \]  

(2.22)

Taking the inverse Fourier transform of (2.22) as follows

\[ p(x, y, z, t) = \frac{\delta(t - \sqrt{(x-x_s)^2 + (y-y_s)^2 + (z-z_s)^2})}{4\pi \sqrt{(x-x_s)^2 + (y-y_s)^2 + (z-z_s)^2}} \]  

(2.23)

The next step is to obtain the Green’s function for the 2-D case in the plane formed by axes \( x \) and \( y \) while the direction \( z \) is perpendicular to the plane of interest. With a 3-D setup, we consider the excitation as a unique spherical point source. For the 2-D case, we assume an infinite long point cylindrical wave line source along the direction of \( z \). The wavefield due to this excitation is written as

\[ \tilde{p}(x, y, k) = \int_{-\infty}^{\infty} \frac{e^{-ik\sqrt{(x-x_s)^2 + (y-y_s)^2 + z^2}}}{4\pi \sqrt{(x-x_s)^2 + (y-y_s)^2 + z^2}} dz. \]  

(2.24)
The integral definition of the Hankel function is provided in [54] as

\[ H^{(2)}_0(k \xi) = \frac{i}{\pi} \int_{-\infty}^{\infty} \frac{e^{-ik\sqrt{\xi^2 + z^2}}}{\sqrt{\xi^2 + z^2}} dz, \]  

(2.25)

where \( H^{(2)}_0 \) denotes the order zero Hankel function of the second kind. Therefore, we can write (2.24) as

\[ \tilde{p}_{2D}(x, y, k) = -\frac{i}{4} H^{(2)}_0(k \sqrt{(x - x_s)^2 + (y - y_s)^2}). \]  

(2.26)

(2.26) corresponds to the Green’s function for the 2-D wave equation.

Given a single source with a representation in frequency domain \( s(k) \), we can represent a 2-D (height-invariant) soundfield \( S(x, k) : \mathbb{R}^2 \times \mathbb{R} \mapsto \mathbb{C} \) at the position \( x \) based on the superposition principle

\[ S(x, s, k) = s(k) \tilde{p}_{2D}(x, s, k). \]  

(2.27)

We can regard the presence of multiple sources as the superposition of single sources due to the linearity of the wave equation, so we have

\[ S(x, k)_{\text{multi}} = \sum_{q=1}^{Q} s_q(k) \tilde{p}_{2D}(x, s_q, k), \]  

(2.28)

where \( Q \) is the number of sources. Substituting (2.26) into (2.28), we have

\[ S(x, k)_{\text{multi}} = \sum_{q=1}^{Q} s_q(k) \frac{-i}{4} H^{(2)}_0(k \|x - s_q\|). \]  

(2.29)

Theoretical results in [55] indicate that a linear combination of plane wave functions provides good approximations to solutions of the Helmholtz equation under any type of boundary conditions. Importantly, both the set of cylindrical wave functions and the set of plane wave functions provide good approximations to any feasible soundfield under certain
2.4. THE GREEN’S FUNCTION AND SOUNDFIELD SUPERPOSITION

conditions. Generally, the spectrum of the Helmholtz solutions consists of two parts: the propagating part and the evanescent waves [51]. The set of plane waves or cylindrical waves including evanescent components formulates a complete set. That is, an arbitrary 2-D (height-invariant) soundfield function \( S(x, k) \) satisfying the wave equation can be considered as a superposition of elementary plane wave and/or cylindrical wave functions with the consideration of the evanescent waves. For the set of cylindrical wave functions and the set of plane wave functions to be complete without considering the evanescent parts, we have to restrict the space of solutions to the solution space where the sources are sufficiently far away due to the exponential decay of the evanescent contribution (i.e. the far field assumption [51]). For the rest of the thesis, the evanescent waves are not considered as it was shown that the evanescent contributions of the desired soundfield cannot be reconstructed without accepting significantly high levels in the driving signals of the secondary sources [56] [57]. Hence, in practice it is favorable to neglect the evanescent contributions of the soundfield. Additionally, the perceptual properties of evanescent contributions are still unclear at the current state of research.

Therefore, we can write \( S(x, k) : \mathbb{R}^2 \times \mathbb{R} \mapsto \mathbb{C} \) as a weighted series of elementary wavefields \( F_n \)

\[
S(x, k) = \sum_n C_n F_n(x, k),
\]

(2.30)

where \( C_n \) are a set of coefficients for individual elementary wavefields. This basis expansion provides a way to characterize the soundfield and it builds up a theoretical basis of both 3-D and 2-D (height invariant) soundfield reproduction in this thesis. More details will be discussed in Chapter 3.
2.5 Existing Soundfield Reproduction Techniques

The principle of soundfield reproduction systems is to recreate the desired soundfield over an extended region of interest using a set of loudspeakers. In this section, we describe the three major approaches to spatial soundfield reproduction: wave field synthesis (WFS), higher order ambisonics (HOA), and the least squares method.

2.5.1 Wave Field Synthesis

The concept of WFS was first introduced by Berkhout [25], in which he used this technique to reproduce a desired soundfield over a relatively large area with intense secondary source distributions. The theoretical basis of WFS approaches [26] [27] is the Kirchhoff-Helmholtz integral [29] [30].

The Kirchhoff-Helmholtz Integral

An secondary source surrounding contour can be considered as an inhomogeneous boundary condition for the wave equation [26]. The Kirchhoff-Helmholtz integral provides the solution of the homogeneous wave equation for a bounded region \( V \) in terms of inhomogeneous boundary conditions [51]:

\[
S(x, k) = -\int_{\partial V} G(x|y, k) \frac{\partial}{\partial n} S(y, k) - S(y, k) \frac{\partial}{\partial n} G(x|y, k) dS_y, \tag{2.31}
\]

where \( S(x, k) \) represents the soundfield within an enclosed listening area \( V \) inside the boundary \( \partial V \), \( x \) is a observation point within \( V \), \( y \) is a point on the boundary \( \partial V \) and \( S_y \) is an element of surface area on \( \partial V \). We assume that no sound source is placed inside \( V \). We denote the inward pointing normal vector of the boundary \( \partial V \) as \( n \), \( \frac{\partial}{\partial n} \) represents the normal derivative in the direction of the vector \( n \) (i.e. directional gradient).
2.5. EXISTING SOUNDFIELD REPRODUCTION TECHNIQUES

$G(x|y, k)$ denotes the free-space Green’s function. According to [30], the soundfield $S(x, k)$ outside $V$ is zero.

The Kirchhoff-Helmholtz integral (2.31) states that the soundfield $S(x, k)$ within $V$ inside the boundary $\partial V$ can be fully controlled by the soundfield $S(x, k)$ and the normal derivative of the pressure (i.e. velocity) on the boundary $\partial V$. If the desired pressure $S^d(x, k)$ is realized by a continuous distribution of secondary monopole and dipole sources arranged on the boundary $\partial V$ with appropriately designed driving signals, the soundfield over the whole region inside $V$ is fully determined by these monopole and dipole sources.

In practice, since loudspeakers generally feature monopole sources at low frequencies [58] and dipole sources are less practical to implement, the dipole sources should be eliminated\(^3\). This is essentially to eliminate the second term in the Kirchoff-Helmholtz integral (2.31) that represents the dipole secondary sources. There are various approaches to simplify (2.31) and derive monopole-only versions of the Kirchhoff-Helmholtz integral, such as the simple source approach and the WFS-based approach with a modification of the free-field Green’s function [51]. These approaches are shown to produce similar results [59].

2-D Wave Field Synthesis

Due to the elimination of the dipole secondary sources, the 2-D soundfield recreated by the surrounding secondary source contour that consists of only the monopole secondary sources with any arbitrary shape is given by

$$S(x, k) = -\int_{\partial V} D_{\text{WFS}}(y, k) G(x|y, k) dS_y.$$  \hspace{1cm} (2.32)

\(^3\)Under this circumstance, the exterior soundfield is no longer zero.
The driving function $D_{WFS}(y, k)$ is given as

$$D_{WFS}(y, k) = 2a(y) \frac{\partial}{\partial n} S^d(y, k) G(x|y, k),$$

(2.33)

where $a(y)$ denotes a predefined window function. This window function allows only those relevant secondary sources where the local propagation direction of the desired soundfield at the position $y$ has a positive component with respect to the direction of the normal vector $n$ of the secondary source to be active [59].

### 2.5.2 Higher Order Ambisonics

Initially invented in the early 1970s by Gerzon [31], Ambisonics is a series of sound recording and rendering techniques using multichannel mixing technology. A two-dimensional planar or three-dimensional full-sphere soundfield can be characterized by encoding and decoding the sound information on multiple channels. The traditional ambisonics system reproduced the zeroth and first order spherical/cylindrical harmonics at a single point and then it was extended to higher orders as higher order ambisonics (HOA) [32], which enlarges the reproduction area and the soundfield inside is accurately reproduced with a superior overall quality of spatial localization.

The principle of the ambisonics system is based on the decompositions of the desired soundfield and the reproduced soundfields due to the secondary sources (i.e. loudspeakers) into spherical harmonics (3-D) or cylindrical harmonics (2-D) [29] [34]. The loudspeakers are typically placed on the boundary of a sphere or a full circle around the desired listening area for the 3-D and 2-D case, respectively. The decompositions of the relevant soundfields into spatial basis harmonic functions facilitate the method of “mode-matching” that leads to an equation system. The optimal solution for the loudspeaker filter gains $D_{HOA}(y_q, k)$ can be derived using the mode matching approach and it is solved to best match
the desired soundfield $S^d(x, k)$ [60]:

$$S^d(x, k) \approx \sum_{q=1}^{Q} D_{\text{HOA}}(y_q, k) G(x|y_q, k),$$

(2.34)

where $q = 1, \ldots, Q$ denote the loudspeaker indices, $y_q$ represents the loudspeaker locations and $G(x|y, k)$ is the acoustic transfer function from the $q$th loudspeaker. In the context of ambisonics, the derivation of loudspeaker driving signals are based on microphone recordings in the frequency domain. The recording procedure and the sound rendering are referred to encoding and decoding, respectively. Generally, the loudspeaker driving signals can be derived directly from the initial description of the desired soundfield and its extension to the encoding and decoding of the soundfield is demonstrated in [60].

For the traditional HOA, which is an amplitude panning approach, the loudspeakers are required to be placed far away from the listening area to meet the assumption that soundfields radiated from the sources are plane waves at the listening position [32]. This is often impractical in real-world applications. Daniel et al. [61] introduced a near-field compensation approach to HOA for more accessible loudspeaker arrangements, in which the spatial temporal transfer function $G(x|y, k)$ was set to the 2-D Green’s function for the near-field compensation rather than using the plane wave assumption. Similar approaches can also be found in cylindrical/spherical harmonics based systems in [2, 3, 45, 46, 48]. Due to the truncation properties of the Bessel functions [62], these optimized technique do not require large source-listener distance and thus are more applicable for implementation in small rooms. Note that these reproduction systems are often confined to circular and spherical loudspeaker contours.

In [56] [60], the authors formulated the problem of a plane wave soundfield reproduction in the spatial wavenumber domain following the similar procedure as used in HOA. Therefore, these methods can be considered as an extension of HOA to planar and linear secondary source distribu-
tions. The underlying difference lies on the fact that these approaches assume a continuous spatial spectrum due to the infinite extension of the secondary source distribution, whereas the conventional HOA features a finite domain and therefore leads to a discrete spectrum [63].

2.5.3 Comparison of Higher Order Ambisonics (HOA) and Wave Field Synthesis (WFS)

The theoretical foundations of WFS and HOA are quite different. As we discussed above, WFS is based on the Kirchhoff-Helmholtz integral while HOA is typically built on the basis of the mode-matching approach. Many studies have been conducted to investigate the differences and connections between WFS and HOA [32, 37, 64]. In 2008, Spors et al. [59] presented a comprehensive comparison between the two approaches in terms of physical properties and spatial sampling. The comparison was conducted based on 2-D reproduction using a circular secondary source arrangement, in which monopoles were used to model secondary sources. A conclusion was drawn that WFS and HOA are not similar with respect to physical properties, and the sampling artifacts of both approaches differ significantly [59]. In terms of physical properties, a list of differences is provided as below [59]:

- For WFS systems, the analytical driving functions are applied to any arbitrary convex geometries while conventional HOA systems are often restricted to circular and spherical loudspeaker contours;

- HOA systems are designed to achieve exact reproduction over the whole predefined reproduction region while WFS systems can only achieve it with a certain source geometry (e.g. linear/planar systems);

- All secondary sources contribute to reproduce the desired sound-field for HOA systems while it is only the selected active secondary
sources for WFS systems.

With respect to the sampling artifacts, the differences are listed as follows [59]:

- HOA systems feature regular spatial structure due to the regular loudspeaker contours, whereas WFS systems feature irregular spatial structure;

- The loudspeaker driving functions for HOA systems are band-limited in the angular frequency domain while WFS systems are not band-limited;

For more details, please refer to [59].

### 2.5.4 The Least Squares Method

Another prevalent soundfield reproduction approach is the least squares method. Initially introduced in 1964 by Trott [65], the main idea of the least squares approach is to calculate the loudspeaker weights by solving the least square solution in the area of sound reproduction. In 1990s, Kirkeby and Nelson [35] [36] proposed a method to theoretically determine the minimum number of loudspeakers required to recreate a soundfield for the least squares techniques. Recently, the least squares method has been widely used in adaptive control applications [5, 37, 38]. The least squares approach generally involves a matrix inversion, which can become ill-conditioned in scenarios when the zones are not carefully positioned or due to certain loudspeaker arrangements [39]. The ill-conditioning problem is extensively discussed in [39], which states that if the matrix is poorly conditioned, the solution may not exist.

A multizone soundfield reproduction system based on the least squares method can be found in [4] [39] with \( N \) non-overlapping control zones. The system requires the use of both loudspeakers and microphones.
The objective is to recreate a soundfield \( S(x, k) \) that best matches the desired soundfield \( S^d(x, k) \), in the least-squares sense. \( M_1 \) microphones generates a measurement vector of the reproduced soundfield \( \hat{p} \), as well as a measurement vector for the desired soundfield \( p \). The Q loudspeaker driving signals are also written in vector form as \( l(k) \). In the multizone case, \( M_1 = \sum_n M_n \) \((n = 1, \ldots, N)\) and \( M_n \) represents the number of sampling points in each of the selected zone. \( \hat{p} \) and \( l(k) \) have the following relationship based on the matrix product

\[
\hat{p} = H(k)l(k),
\]  

(2.35)

where \( H(k) \) is a \( M_1 \times Q \) matrix featuring the acoustic transfer functions (ATFs) from the loudspeakers to the listener zones. Its \((p, q)\)th element represents the linear ATF between the \((q, p)\)th loudspeaker-microphone pair. Note that the variation of ATFs depends on the characteristics and arrangement of the loudspeaker-microphone pairs [4]. In real listening environments, the effects of the room reverberation that occur in the reproduction should also be taken into account.\(^4\)

We aim to find the loudspeaker filter gain solution \( l(k) \) that minimizes the error between the reproduced soundfield samplings \( \hat{p} \) and the desired soundfield samplings \( p \) in a least-squares sense:

\[
\min_l \|p - Hl\|^2.
\]  

(2.36)

The least-square solution to the problem is

\[
l_0(k) = H^\dagger(k)p,
\]  

(2.37)

where \([\cdot]^\dagger\) is the Moore-Penrose Pseudo Inverse\(^5\) and when \( H(k) \) is full-

---

\(^4\)See Section “Reverberation” of this Chapter for more details.

\(^5\)For more information on the Moore-Penrose pseudo-inverse, the reader is referred to [66]
rank
\[ H^\dagger(k) = (H(k)^H H(k))^{-1} H(k)^H. \] (2.38)

The symbol \([·]^H\) denotes the Hermitian transpose of a matrix. This system formulation can be categorized as the method of pressure matching in soundfield reproduction, which equalizes the transfer functions over a discrete set of points. The above-mentioned mode matching method also uses the least squares method to derive the optimal loudspeaker solutions. The difference lies on the fact that mode matching matches the reproduced soundfield with the desired soundfield over the whole control region in the mode domain. The individual soundfield modes are coefficients of the orthogonal basis functions. The ill-conditioning problem also exists among the works using mode matching as it involves an inversion of transfer function matrix of the loudspeakers.

The Ill-Conditioning Problem

The solution to the inverse problem in (2.37) is strongly sensitive to the conditioning of \(H(k)\), which can be quantified by the condition number of the matrix [39]. Either large loudspeaker gains may be required which leads to large levels of sound outside the control zone or large reproduction errors may result from small perturbations of the ATFs from the loudspeakers to the listener zones if the propagation matrix is ill-conditioned [39]. The ill-conditioning problem may occur in these scenarios: 1) soundfield reproduction using a nonuniform arrangement of loudspeakers [67] and 2) creating multiple zones of independent spatial audio in the same acoustic space, when one zone is occluded by another [1].

In [39], the concept of singular value decomposition (SVD) is introduced to analyze the ill-conditioning problem that involves the calculation of the pseudo-inverse \(H^\dagger(k)\). If we assume that \(Q > M\), which is a common setup in practice, the SVD of the channel matrix \(H\) can be represented
as [39]:

$$H = U \Sigma V^H,$$

(2.39)

where $V \in \mathbb{C}^{M_1 \times M_1}$ and $U \in \mathbb{C}^{Q \times Q}$ are unitary matrices that represent an algebraic basis change. $\Sigma$ is a $M_1 \times Q$ matrix consisting of a $T \times T$ ($T = \min(M_1, Q)$) diagonal submatrix, whose elements are the singular values $\sigma_n$ in decreasing order. The remaining $|M_1 - Q|$ columns contain only zeros. In the case that the number of loudspeakers is larger than the number of observation points, the matrix $\Sigma$ has the following form:

$$\Sigma_{M_1 \times Q} = \begin{bmatrix} \sigma_1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_{M_1} & 0 & \ldots & 0 \end{bmatrix}.$$  

(2.40)

Furthermore, the Moore-Penrose pseudo-inverse matrix $H^\dagger$ can be expressed as [68]:

$$H^\dagger = V \Sigma^+ U^H,$$

(2.41)

where the matrix $\Sigma^+$ is derived by transposing $\Sigma$ and setting each non-zero element to its reciprocal.

The condition number of the propagation matrix, which is represented by the ratio between the largest and the smallest singular value of $H$, is a good index of the conditioning of the matrix. Considering the inverse problem in (2.37), it is concluded in [39] that the soundfield reproduction requires a large amount of energy (proportional to $1/\sigma_n^2$) if the so-called microphone array mode$^6$ corresponds to an inefficient loudspeaker array mode$^7$. This implies that an almost infinite amount of energy is required by the loudspeaker array for exact reconstruction if $\sigma_n$ is close to 0, thereby leading to poor-conditioning of $H(k)$. In the case that the inverse

---

$^6$A combination of $M_1$ orthogonal microphone array modes $u_n$ can be used to describe the sampled soundfield from microphone recordings.

$^7$A combination of $Q$ orthogonal loudspeaker array modes $v_n$ can be used to describe the soundfield generated by the loudspeaker array.
problem is ill-posed due to some inefficient loudspeaker array modes, the small errors in the estimation of the propagation channel matrix or the presence of trivial noises in the microphone recordings can lead to “devastating” effects in the reproduction system as it requires a large amount of loudspeaker effort for those inefficient modes [39]. Generally, we can diminish the effects of the inefficient loudspeaker array modes by constraining the total loudspeaker array energy under this circumstance, which will be discussed in the following section.

Possible Solutions to Ill-conditioning Problem

One technique known as truncated singular value decomposition was introduced to regularize an ill-posed problem in [49] [69]. This is achieved by deliberately avoiding the reconstruction of the microphone array modes corresponding to inefficient loudspeaker array modes. The approach implies that the smaller singular values are not inverted in the computation of $\Sigma^+$ due to the truncation (i.e. converted into zero). Alternatively, it is possible to limit the total loudspeaker array energy by using the approach of Tikhonov regularization, which is essentially achieved by adding a small quantity $\beta$ to all singular values $\sigma_n$ [70]. The Tikhonov regularization method implies a trade-off between the accuracy of the desired soundfield reproduction and the limit on the total energy of the signals driving the loudspeakers.

As we mentioned, the ill-conditioning problem results from matrix inversion which may lead to errors if the matrix is poorly conditioned. In [48], the author introduced the concept of theoretical continuous loudspeakers to the soundfield reproduction to derive the discretized loudspeaker aperture functions by avoiding matrix inversion, which provides an alternative solution to address the ill-conditioning problem. In addition, this continuous aperture method reveals the underlying structure of the solution as a function of the desired soundfield, the loudspeaker positions, and the frequency. However, the approach only applies to a
full circular loudspeaker array. In Chapter 3, we present a generalization of the reproduction method proposed in [48] that facilitates practical loudspeaker array arrangements.

2.6 Acoustic Channel Effects in Complex Environments

In this thesis, we extend the spatial multizone soundfield reproduction to complex environments. We consider two types of acoustic channel effects that alter the received soundfield: room reverberation and scattering effects. This section describes the background concepts of modeling these channel effects.

2.6.1 Introduction to Reverberation

Reverberation is the collection of reflected sounds from the surfaces in an enclosure [71]. It is created when a sound or signal leads to a large number of reflections and then gradually decay as the sound is absorbed by walls, scatterers and air. This is most noticeable when the sound source stops but the reflections continue to exist till they reach zero amplitude. The majority of the soundfield reproduction techniques are designed with free-field assumption, but this is not the case in most real implementations. Room reverberation poses a major challenge in soundfield reproduction and it generally leads to inferior performance. As the impact of reverberation on the soundfield reproduction performance will be considered in this thesis, a study of the reverberation characteristics and an introduction to dereverberation techniques are presented in this section.

Image-Source Method (ISM)

To reveal the general reverberation properties, we demonstrate the model of reverberant soundfields using the image source method in this section.
2.6. ACOUSTIC CHANNEL EFFECTS IN COMPLEX ENVIRONMENTS

Figure 2.2: Sound transmission in a room is modeled as a linear time invariant system.

The soundfield in a reverberant environment can be modeled as a linear and time-invariant system as shown in Fig. 2.2. The sound source signal \( s(t) \) is filtered by a position-dependent acoustic impulse response \( h(x, t) \), representing the collection of reflected sounds between the source and the position \( x \), to generate the actual soundfield \( s_a(x; t) \) at a point \( x \) and time instance \( t \).

Taking the Fourier transform in terms of the wavenumber \( k \), we have the following expression:

\[
S^a(x, k) = H(x, k)s^a(k),
\]

where \( H(x, k) \) is defined as the complex acoustic transfer function between source signal \( s(k) \) in the wavenumber domain and the actual soundfield \( S^a(x, k) \). With the assumptions of free-field space and the point source be-
CHAPTER 2. LITERATURE REVIEW AND BACKGROUND THEORY

ing an omnidirectional model, \( H(x, k) \) can be replaced with the following transfer function in space:

\[
G(x|y; k) = \frac{e^{-ik\|x-y\|}}{4\pi\|x - y\|},
\]

(2.43)

where \( y \) is the position of the source. This is referred to the 3-D Green’s function, which is the same as (2.22).

In this thesis, we use the image source method (ISM) to model room reverberation. The conventional image source method, initially proposed by Allen and Berkley [72], is a widely-used algorithm for simulating room impulse responses (RIRs) with the provided room geometries. Some improvement work to Allen and Berkley’s original ISM implementation were conducted in [73] [74]. The basic idea of the image source method is that the reverberation can be characterized as the superposition effect of an infinite number of image sources by assuming that the walls are specular (mirror-like) sound reflectors and the wall reflection coefficients are independent of incidence angle and frequency [72]. Note that the image sources are simulated by reflecting the real sound source in the given room that is generally set to be a 2-D rectangular chamber\(^8\) to simplify the computation of the image source positions. A demonstration of image source method for modeling the 2-D rectangular room reverberation is shown in Fig. 2.3.

Similar to the 2-D case, we define a 3-D Cartesian coordinate system with its origin corresponding to one of the room corners. Let \( y = [x_s, y_s, z_s]^T \) and \( x = [x_r, y_r, z_r]^T \) denote the position vectors of a source and a receiver respectively. Similarly, let \( r_{\text{imag}} = 2[L_x, mL_y, nL_z]^T \) represent the vector of room dimensions, in which \( L_x, L_y, L_z \) represent length, width and height respectively \((l, m, n \in \mathbb{Z})\). The sound reflection coefficient \( \beta_{\text{wall}} \), which characterize the acoustical property of each surface in the enclosure, is

\(^8\)For the 3-D case, the room features a cubical shape.
Figure 2.3: Demonstration of image source method for modeling the 2-D rectangular room reverberation. The solid rectangle in the center represents the room, the real source is displayed as (●) and the image sources are denoted as (○). The lattice is formed by infinitely repeating the block bounded by dashed lines.

related to the absorption coefficient $\alpha_{\text{wall}}$ according to [72]

$$\alpha_{\text{wall}} = 1 - \beta_{\text{wall}}^2.$$  \hspace{1cm} (2.44)

The reflection coefficients for each of the surfaces are denoted $\beta_{x,i}$, $\beta_{y,i}$, $\beta_{z,i}$ and $\beta_{x,i}$ with $i = 1,2$.

The RIR from the source to the receiver can be characterized by considering image sources on an infinite grid of mirror rooms expanding in all
The Green’s function for the ISM hence is
\[
G(x; k) = \sum_{u=0}^{1} \sum_{l=-\infty}^{\infty} A(u, l) e^{-ik\|d(u, l)\|/\|d(u, l)\|},
\]
where the scalars \(u = (u, v, w)\) and \(l = (l, m, n)\) are parameters controlling the indexing of the image sources in the space. More specifically, \(u\) indexes the basic four sources (displayed in Fig. 2.3 for 2-D case) and \(l\) denotes the room-dimension offset \(r_{imag}\). The attenuation factor \(A(u, l)\) and the distance from the image source to the receiver \(d(u, l)\) are defined as follows:
\[
A(u, l) = \frac{1}{4\pi} \beta_{x,1}^{(u-1)} \beta_{x,2}^{(u-1)} \beta_{y,1}^{(m-v)} \beta_{y,2}^{(m-v)} \beta_{z,1}^{(n-w)} \beta_{z,2}^{(n-w)} ,
\]
\[
d(u, l) = \|\text{diag}(2u-1, 2v-1, 2w-1) \cdot y + x - \text{diag}(r_{imag})\|,
\]
where \(\|\cdot\|\) represents the Euclidean norm and \(\text{diag}(\cdot)\) denoting a diagonal matrix with the arguments as diagonal entries.

**Dereverberation Techniques**

The majority of existing works in the area of spatial soundfield reproduction do not take into account the reverberant environments that practical multizone sound reproduction systems will encounter. The ambient reverberation usually leads to an inferior performance compared with reproduction of the desired soundfield using the free-field assumption. The reverberation compensation process is difficult to handle because the reverberant room channel is unknown and a large number of loudspeakers and microphones required by existing soundfield reproduction systems. Reverberation cancelation techniques are indispensable for a reproduction system with real-world settings. In general, the methods that aim to minimize the reverberation effects on the soundfield reproduction can be
categorized into three groups [41] [42]: passive techniques that reduce room reflections, equalization approaches based on models of the reverberant soundfield or channel behavior, and adaptive room cancelation methods.

The most natural way among the passive techniques is to equip the room with acoustic absorption materials, so that a modest attenuation of sound reflection is provided. However, the related costs pose a major challenge for this method and it is difficult to realize in many real-world application scenarios (e.g. soundfield reproduction in an office or home environment). A more technically advanced passive approach may use fixed or variable directivity higher order loudspeakers in order to minimize the sound radiation directing towards the walls of a room [75] [76]. However, it requires some specific sound reproduction apparatus, which is difficult to achieve in practice.

To equalize the reverberation effects, a common way is inverting the channel response prior to the derivation of the loudspeaker driving signals. The traditional approach for spatial sound reproduction in a reverberant setting is pressure matching, which equalizes the transfer functions over a discrete set of points [77] [78]. This technique leads to poor robustness in regions further away from the design points and inaccurate reproduction [2]. The work by Brannmark et al. [79] [80] proposed a polynomial based MIMO formulation of the equalization problem to improve the robustness by applying a probabilistic model of the channel variability. A technique based on mode matching to reproduce a single-zone soundfield accurately over a control region in reverberant rooms was proposed in [2]. The method determines the ATF between each loudspeaker and the circular control region by evenly sampling sound pressure along the boundary. An approach of reproducing a multizone soundfield within a desired region using sparse methods was introduced in [6]. A limited number of randomly placed measurements were employed to estimate the transfer functions between the loudspeakers over the desired region in reverberant
environments based on sparse approximation. The estimates were then used to derive the optimal least-squares solution for the loudspeaker filter gains. Although the approach was shown to significantly reduce the number of required microphones for an accurate multizone soundfield reproduction, a prior measurement of the room transfer function for all the employed loudspeakers was still needed. This is time-consuming to implement in practice and it can be further exacerbated by changes in the ambient environment conditions during the measurement process.

Wave Domain Adaptive Filtering (WDAF) is a more practical approach to the application of reverberation cancellation in soundfield reproduction. Initially proposed by Buchner et al. [81] [82], it has been introduced to active listening room compensation in Wave Field Synthesis systems [83–85]. Comparing with the traditional time and frequency domain adaptive filtering methods [86] [87], it features a better convergence behavior by decoupling the high correlation between multiple reproduction channels. The wave-domain representation of the soundfield was described using two transformations on the microphone array input and the loudspeaker output respectively [83]. The work by Schneider et al. [84] [88] has further reduced the computational complexity of the basic WDAF adaptation process. It was achieved by considering that the dominant couplings between the soundfield modes limit only in the vicinity of the diagonal line of the transformations and neglecting the weaker ones. A similar adaptive method was proposed in [5], in which the reverberant soundfield was described and estimated by exploiting the orthogonality of the Fourier-Bessel expansion to simplify the listening room compensation problem within a region of interest. The work of [5] [88] provides some insights into the underlying structure of the reverberant soundfield in the mode-domain. In Chapter 5, we use this inspiration to propose an adaptive reverberation cancelation system for the desired multizone soundfield reproduction using scarce microphone measurements.
2.6. ACOUSTIC CHANNEL EFFECTS IN COMPLEX ENVIRONMENTS

2.6.2 The Scattering Effects Resulted From Obstacles

In practice, the desired sound waves can be scattered with by presence of the head and body of the user in the pre-defined reproduction region. Similar to the effects due to reverberation, the scattering effects generally degrades the reproduction performance of the proposed system. The objective of this section is to obtain a preliminary understanding of the change in the soundfield by the scattering effects.

An incident wavefield is scattered when it meets a scatterer or an obstacle. The scattered soundfield is normally defined as the difference between the total field (or the changed field) and the incident field [89]. The total field due to a monopole secondary source and a rigid object can be derived by applying the boundary condition on the surface of the rigid circle, that is the particle velocity perpendicular to the surface is zero on the surface of the rigid object [51]. For ease of demonstration, the boundary condition is represented and applied only with respect to a circular object of radius \( r = \|x\| \) in the 2-D (height-invariant) cylindrical coordinate. Therefore, soundfields are expressed by the expansion of cylindrical harmonics in this part of the work.

Figure 2.4 illustrates the system model of soundfield reproduction around an obstacle (e.g. human head). For simplicity, we assume that no reflection from walls is included and the user’s head is regarded as a rigid circular object whose radius \( a \) is \( R_1 \). The rigid circle is placed in the center of the desired reproduction region with the radius of \( R_2 \). The loudspeaker array is placed along a concentric circle with the radius of \( R_0 \). The incident field generated by the \( l \)th loudspeaker \( H_l(x, k) \) with unit gain in the frequency
CHAPTER 2. LITERATURE REVIEW AND BACKGROUND THEORY

Figure 2.4: Soundfield reproduction around a human head.

domain \(^9\) can be expressed as

\[
H_I(x, k) = \frac{-i}{4} H_0^{(2)}(k\|y_l - x\|) = \sum_{m=-M}^{M} H_m^{(1)}(k\|y_l\|) e^{-im\phi_l} J_m(k\|x\|) e^{im\phi_x},
\]

(2.48)

where \(M = \lceil kR_2 \rceil\) is the truncation length and \(y_l\) is the position of the \(l\)th loudspeaker.

The scattered soundfield is the wavefield that propagates outward from the inside to the outside of the scatterer and it can be represented as [51]

\[
H_{sca}^{(l)}(x, k) = \sum_{m=-M}^{M} \alpha_m^{(l)}(k) H_m^{(1)}(k\|x\|) e^{lm\phi_x},
\]

(2.49)

where \(\alpha_m^{(l)}(k)\) denotes a set of unidentified coefficients. Since the total field

\(^9\)Loudspeakers are regarded as monopole secondary sources.
2.6. ACOUSTIC CHANNEL EFFECTS IN COMPLEX ENVIRONMENTS

\( H_{tot}^{(l)}(x, k) \) is the sum of the incident and scattered fields. We have

\[
H_{tot}^{(l)}(x, k) = H_{inc}^{(l)}(x, k) + H_i(x, k).
\]

(2.50)

Now we consider the boundary condition, which explicitly states that the particle velocity is zero on the surface of the rigid scatterer, can be represented by taking the derivative with respect to \( r \) and setting it to 0:

\[
\frac{\partial}{\partial r} [H_{tot}^{(l)}(x, k)]_{r=R_1, \|y\|=R_2} = 0.
\]

(2.51)

From this equation, we can derive the coefficient \( \alpha_m^{(l)}(k) \) as

\[
\alpha_m^{(l)}(k) = -\frac{i}{4} \frac{H_m^{(1)}(kR_0)e^{-im\phi_1}J_m'(kR_1)}{H_m^{(1)\prime}(kR_1)},
\]

(2.52)

where \( J_m'(\cdot) = J_m^{m-1}(\cdot) + J_m^{m+1}(\cdot) \) and \( H_m^{(1)\prime}(\cdot) = \frac{H_m^{(1)}(\cdot) + H_m^{(1)\prime}(\cdot)}{2} \). Using the coefficients \( \alpha_m^{(l)}(k) \), we can then calculate the total fields due to the loudspeaker and the rigid object.

Several research works have been conducted to alleviate the scattering effects resulting from obstacles in the soundfield. Naturally, the scattering effect can be reduced if more loudspeakers are employed and the loudspeakers are placed at longer distances from the reproduction region. Under this circumstance, the size of the scatterers becomes relatively smaller and its impact diminishes consequently. Betlehem and Poletti et. al [90] proposed a soundfield reproduction method around a scatter in a reverberant room based on the soundfield model around a solid sphere using mode-matching. Similar work can also be found in [91] by employing spherical harmonic decomposition. However, soundfield reproduction around arbitrary scattering objects has not been well-studied yet.
2.7 Works Related to Multizone Soundfield Reproduction

Spatial multizone soundfield reproduction has recently drawn attention due to its various applications. It aims to provide an individual sound environment to each of a set of listeners without the provision of physical isolation or the use of headphones. However, the majority of the existing works in spatial soundfield reproduction focuses on a single zone, which only features a single sound environment within a set of loudspeakers. The realization of multizone soundfield reproduction is a conceptually challenging problem and only a limited number of works are reported in the literature.

Intuitively, multizone sound systems can be set up in principle using separate loudspeaker arrays for each of the selected zones. The Kirchhoff-Helmholtz integral equation in (2.31) indicates that the soundfield inside the boundary of the spatial region can be fully determined by the pressure and its normal derivative (velocity) on the boundary. In the meanwhile, the pressure outside the region of space is zero. It implies that an infinitely dense distribution of secondary monopole and dipole sources surrounding one zone of interest (with arbitrary shape) can reproduce any feasible soundfield inside that zone without contributing any interference to other non-overlapping zones. Therefore, a soundfield with multizone settings could be theoretically generated using multiple monopole and dipole speaker arrays for each of the selected zones. However, we wish to create multiple soundfields using a single array of monopole loudspeakers that allows free movements for the listeners between zones. In addition, the dipole sources are difficult to implement and the ignorance of dipole sources leads to the sound interference outside the enclosure being non-zero. Therefore, the Helmholtz integral-based approach is not applicable.

The origin of work in the area of multizone sound rendering can be traced to 1997, when Druyvesteyn and Garas [92] sought to reproduce
sound in some desired region while reducing the sound elsewhere using conventional beamforming methods. Beamforming is an acoustic signal processing technique used for directional sound transmission or reception, which is achieved by combining elements in a phased array in such a way as to receive the sound radiating from a specific location and attenuate the sound from other locations [93]. Tashev et al. [94] from Microsoft realized the cancelation of sound waves in one area and the amplification in another simultaneously using a linear loudspeaker array. The system was based on a steerable beamforming algorithm with an enhanced robustness. However, this project is still in its early stages.

In 2002, Choi and Kim [95] proposed an acoustic contrast control method to maximize the ratio of the average acoustic energy in the so-called bright and quite zones. Improved robustness was obtained by the work in [96] [43]. In [96], the author states that the acoustic contrast control method can be seen as a generalization of the conventional beamforming problem. Comparing with the conventional beamforming, the systems based on the contrast maximization approach aim to steer the main beam over a range of angles instead of just a single direction. It can be shown that such systems are reduced to conventional optimal beamforming if only one microphone is selected in the bright zone [97]. However, the focus of the above-mentioned approaches is limited to the control of acoustic energy contrast between two different zones and the outcome of this approach fails to control the soundfield and provide a sense of localization for the listener in the bright zone.

In [32], higher order ambisonics (HOA) was introduced to reproduce soundfield in multi-zones based on mode matching. In 2008, Poletti [4] proposed a least squares method to generate a 2-D monochromatic multizone soundfield. In [1], a method was proposed to control the sound in each zone independently, while also controlling the leakage into other listeners’ zones. The system was based on a non-linear convex optimization approach for determining the loudspeaker weights, in which
the authors also incorporated a constraint on the summed square value of the loudspeaker weights to improve the system robustness. Ahrens et al. proposed a soundfield reproduction method for planer or linear arrays of loudspeakers, which analytically derives loudspeaker filters by using the spectral division method (SDM) [56]. The method was extended to the multizone sound case in [98] [99] by modeling sound pressures as a rectangular window corresponding to bright and dark zones and deriving the spatial filter at the wavenumber domain. However, the work of [56, 98, 99] can only render the desired soundfield precisely for receiver positions on a straight line parallel to the linear loudspeaker array (or multiple reference lines using multiple loudspeaker array). In 2013, Coleman et al. proposed the planarity control method in [47] [100], which can be seen as a combination of the acoustic contrast control and the pressure matching method. A cost function was formulated to optimize sound attenuation in the quiet zone and the reproduction of the plane wave with a limited flexibility of incoming direction specification in the bright zone. In [44], a method of describing the desired multizone soundfield as an orthogonal expansion of basis functions over the reproduction region was proposed. The basis function set was formed by implementing a modified Gram-Schmidt process on a suitable set of plane wave functions.

The fore-mentioned approaches do not take into account reverberant environments in multizone sound reproduction. The reverberant case is difficult to handle because of the rapid variation of the ATF over the room [101]. In practical scenarios, the performance of multizone soundfield reproduction techniques is commonly degraded by the effects of reverberation.

The traditional approach for spatial sound reproduction in a reverberant setting is pressure matching, which equalizes the transfer functions over a discrete set of points using least squares techniques [77] [102]. This technique leads to poor robustness in regions further away from the design points and inaccurate reproduction [2]. A technique to reproduce
2.7. WORKS RELATED TO MULTIZONE SOUNDFIELD REPRODUCTION

a single-zone soundfield accurately over a control region in reverberant rooms was proposed in [2]. The method is based on mode matching, and uses a method for determining the ATF from each loudspeaker over the desired control region by sampling the sound pressure along the boundary. This work was extended to the multizone soundfield case in [3] using the translation of spatial harmonic coefficients between coordinate systems to achieve a global control of the multizone soundfield. To reproduce a spatial soundfield with a designated quiet zone, a method that uses spatial band stop filters to suppress the unintended interzone interference was proposed in [45] [46]. This was achieved by using the higher order spatial harmonics of one zone to cancel the undesirable effects of the lower order harmonics of the same zone on the other zones. However, this method is applicable only for two zones.

The methods of [2,3,45,46] are based on describing the desired soundfield by the cylindrical/spherical harmonic expansion representation and have some practical drawbacks. Firstly, they require a relatively large number of measurements to estimate the ATFs of the loudspeakers with the needed accuracy. In addition, these approaches lead to problems for systems with a wide frequency range as the coefficients for the defined basis function cannot be determined at a number of frequencies due to the so-called large error scaling [2]. This irregularity problem has been extensively studied in [103–105] and it is associated with the so-called nonuniqueness of the solution of the soundfield description at the Dirichlet eigenfrequencies. A solution for this is to double the number of pressure measurements and sample over two concentric circles/spheres about these frequencies [2] [106], which makes the method cumbersome to implement for the wide-band multizone sound reproduction. The issue can be overcome by using alternate systems [2,107–109]. However, these methods impose additional constraints on the hardware and its arrangement. For example, either using dual spherical microphone arrays (which doubles the number of required microphones), or mounting
the microphones around a rigid sphere (which disables an open-sphere configurations), or using multiple high-order or directional microphones have been proposed in the literature. These additional constraints make the systems less practical to implement.

2.8 Summary and Links to next Chapters

In this chapter, we presented the literature reviews and related background theory on spatial soundfield reproduction. First, we investigated the Helmholtz equation and the general solutions to the homogeneous wave equation. This forms the theoretical foundation in Chapter 3 for an efficient parametrization of any feasible soundfield, which is modeled using the spatial basis functions obtained from the solutions to the wave equation. Existing soundfield reproduction techniques such as wave field synthesis (WFS), higher order ambisonics (HOA) and least squares method were presented. This was followed by a comparison between HOA and WFS. Some of the existing soundfield reproduction techniques will be considered and applied in the multi-zone system proposed in Chapter 3. Two types of acoustic channel effects that alter the received soundfield in complex environments were considered: room reverberation and scattering effects. How we can alleviate the impacts on the multizone sound system due to reverberation effects will be the main focus in this thesis. Reverberation characteristics were described and a review of the major dereverberation techniques was also presented. In Chapter 4 and Chapter 5, we will incorporate the design of reverberation equalization to extend the application of multizone soundfield reproduction to reverberant environments using sparse methods. We also investigated the existing spatial multizone sound rendering system. The approaches and drawbacks in the prior art were analyzed, which motivates our work in this thesis. Note that the majority of the existing work on multizone soundfield reproduction is
limited to theoretical results. In Chapter 6, we will conduct a real-world implementation based on the proposed theoretical work in this thesis and aim to create multiple bright and quiet zone in practical listening rooms.
Overview: In this chapter, we introduce a method for 2-D spatial multizone soundfield reproduction based on describing a desired multizone soundfield as an orthogonal expansion of basis functions over the desired reproduction region. This approach will find the solution to the Helmholtz equation that is closest, in the weighted least squares sense to the desired soundfield. The basis orthogonal set is formed using QR factorization with as input a set of solutions of the Helmholtz equation (e.g. plane waves or circular waves). The coefficients of the Helmholtz solution wavefields can then be calculated, which reduces the multizone sound reproduction problem to reconstructing basis wavefields over the entire desired region. The method facilitates applications with practical loudspeaker configurations. The approach is shown effective on the control of both accurately reproducing sound in the selected bright zone and trivial sound...
3.1 Introduction

Multizone soundfield reproduction is a technique that aims at providing an individual sound environment to each listener using a set of loudspeakers. Currently, the main research goals in this area focus on reproducing the soundfield accurately in the selected zone, the creation of acoustically bright and quiet zones simultaneously with high level of energy contrast, improving the system robustness, facilitating practical applications, low computational complexity etc.

In 2002, an acoustic contrast control method was proposed in [95] to generate an acoustically bright zone which features higher acoustic energy than in another space. A sound-focused personal audio system was implemented in [110], which obtained a channel separation of over 20 dB in the bright zone. The research focus of [95] [110] is limited to the control of acoustic contrast between different zones and the outcome of this approach fails to render the soundfield in the bright zone.

In [32], higher order ambisonics (HOA) was introduced to reproduce soundfield in multi-zones based on mode matching. In 2008, Poletti [4] proposed an approach of using a standard least-squares method to reproduce a 2-D monochromatic multizone soundfield. Further investigation was made in [111] which applied the work to narrowband speech signals. Global control of the multizone soundfield was proposed in [3] using spatial harmonic coefficients translation between coordinate systems.

The fore-mentioned approaches do not touch on a controllable limit of the sound leakage from one zone into other specified zones. A method of using spatial band stop filtering to suppress the inter-zone sound interference was proposed in [45]. It can be observed that acoustic attenuation can only be achieved outside the desired control region, which makes...
the method not useful for practical applications. In [1], multizone sound reproduction was formulated as a constrained optimization problem to control the sound in each zone individually, while also controlling the leakage into the zones of other listeners. The works of [1,3,4,45] involve a matrix inversion of the acoustic transfer function, which can easily become non-robust in the scenarios when the zones are not carefully positioned [39]. This ill-conditioned problem in the soundfield reproduction can be generally improved by introducing regularization such as Tikhonov method [39] [112].

In this chapter, we propose a method of describing a desired multizone soundfield to control the reproduction in each specified zone independently. Our approach consists of two distinct stages: 1) To match the desired multizone soundfield using an orthogonal expansion of basis functions formed by QR factorization. 2) To reproduce the feasible soundfield with an array of loudspeakers. We find the optimal solution of the Helmholtz equation at the first stage and the derived coefficients of the basis wavefields enable us to apply existing reproduction methods such as those in [48, 113, 114] to reproduce the desired multizone soundfield. As we are interested in methods that facilitate practical applications, we extended the methods of [48] to include reproduction from a semi-circle array of loudspeakers. The simulation results confirm the effectiveness of the multizone system in terms of accurately recreating the desired soundfield in the bright zone while also achieving manipulable acoustic attenuation in the selected quiet zone. The concept of a smooth weighting function is also introduced to improve system performance in practical settings.

3.2 Problem Formulation and Notation

We seek to control the reproduction of the desired 2-D (height-invariant) multizone soundfield over each of a set of pre-defined zones. In this work,
we consider disc-shaped zones. The theory we develop in this chapter is readily extended to 3-D space. Throughout this chapter, we use the following notation: matrices and vectors are represented by upper and lower bold face respectively, e.g., \( \mathbf{C} \) and \( \mathbf{y} \). \( \mathbf{x} \in \mathbb{R}^2 \) denotes an arbitrary spatial observation point. \( \| \cdot \| \) denotes the Euclidean norm and \( [\cdot]^H \) denotes the Hermitian transpose. The imaginary unit is denoted by \( i = \sqrt{-1} \). The superscripts “d” and “a” are used to represent the desired soundfield and the actual soundfield, respectively.

As illustrated in Fig. 3.1, the desired reproduction region \( \mathbb{D} \) is the entire control zone of interest with a radius of \( r \), which includes both the acoustic bright zone \( \mathbb{D}_b \) and the quiet zone \( \mathbb{D}_q \) of radius \( r_q \). The target bright and quiet zones are located at angles \( \phi_1 \) and \( \phi_2 \) respectively. We define the remaining area in \( \mathbb{D} \) as the unattended zone \( \mathbb{D}_u \). \( k = \frac{2\pi f}{c} \) is the wavenumber, where \( f \) is the frequency and \( c \) is the speed of sound.
propagation. We assume no sound sources or scattering objects are present inside the reproduction area. The number of employed loudspeakers is \( Q \).

For simplicity, the loudspeakers are assumed to behave as omnidirectional point sources.

Our main objective is to accurately reproduce the desired soundfield inside the zone \( \mathbb{D}_b \) while controlling the acoustic energy leaked into the zone \( \mathbb{D}_q \). To evaluate the performance of our system we use the following two measures:

- The acoustic brightness contrast between the bright zone \( \mathbb{D}_b \) and the quiet zone \( \mathbb{D}_q \) to quantify sound leakage between the two zones [110]:

\[
\zeta(k) = \frac{1}{\|\mathbb{D}_b\|} \int_{\mathbb{D}_b} |S^a(x, k)|^2 \, dx
\]

\[
\frac{1}{\|\mathbb{D}_q\|} \int_{\mathbb{D}_q} |S^a(x, k)|^2 \, dx.
\]

(3.1)

where \( x \) denotes an arbitrary spatial observation point. \( \|\mathbb{D}_b\| \) and \( \|\mathbb{D}_q\| \) mark the sizes of \( \mathbb{D}_b \) and \( \mathbb{D}_q \).

- The mean square error (MSE) between the desired sound \( S^d(x, k) \) and the actually rendered sound \( S^a(x, k) \) over the zone \( \mathbb{D}_b \) to gauge the reproduction accuracy:

\[
\epsilon_M(k) = \frac{\int_{\mathbb{D}_b} |S^d(x, k) - S^a(x, k)|^2 \, dx}{\int_{\mathbb{D}_b} |S^d(x, k)|^2 \, dx}.
\]

(3.2)

3.3 Describing Arbitrary Soundfields with a Basis Expansion

In this section, we introduce a method to describe a desired soundfield as an orthogonal expansion of basis functions for the reproduction region.

An arbitrary 2-D (height-invariant) soundfield function \( S^a(x, k) \) satisfying the wave equation can be written as a superposition of an orthogonal set of solutions of the the homogeneous Helmholtz equation (e.g. plane
3.3. DESCRIBING ARBITRARY SOUNDFIELDS WITH A BASIS EXPANSION

wave functions) [51].\(^1\) Therefore, we can write \(S^a(x, k) : \mathbb{D} \times \mathbb{R} \rightarrow \mathbb{C}\) as a weighted series of basis functions \(\{G_n\}_{n \in A}\) (where \(A\) is a set of indices)

\[
S^a(x, k) = \sum_n E_n G_n(x, k),
\]

where \(E_n\) is a set of coefficients for the basis functions. We assume that all feasible solutions on \(\mathbb{D}\) fall in the space spanned by the orthogonal set \(\{G_n\}_{n \in A}\).

To define the basis functions, we use the weighted inner product

\[
\langle Y_1, Y_2 \rangle_w = \int_{\mathbb{D}} Y_1(x) Y_2^*(x) w(x) dx,
\]

where \(Y_1\) and \(Y_2\) are functions of the form \(Y_1 : \mathbb{R}^2 \rightarrow \mathbb{C}\) and \(Y_2 : \mathbb{R}^2 \rightarrow \mathbb{C}\). The weighting function \(w(x)\) specifies the relative importance of the reproduction accuracy for each point in space. The weighted inner product induces a squared norm, which we denote as \(\|Y_1\|_w^2 = \langle Y_1, Y_1 \rangle_w\).

In the present case, where we split \(\mathbb{D}\) into discrete bright, quiet and unattended zones, it is reasonable to assign a fixed weight to each zone:

\[
w(x) = \begin{cases} 
w_b, & x \in \mathbb{D}_b \\
w_q, & x \in \mathbb{D}_q \\
w_u, & x \in \mathbb{D}_u \end{cases}
\]

Given a desired soundfield \(S^d(x, k)\), computing the expansion coefficients \(E_n = \langle S^d(x, k), G_n(x, k) \rangle_w\) then gives the coefficients \(E_n\) that minimize:

\[
\min_{E_n \in A} \| \sum_n E_n G_n(x, k) - S^d(x, k) \|_w^2.
\]

To construct the orthogonal set of basis functions that can be used to describe the desired multizone soundfield, it is convenient to start with

\(^1\)Note that we did not consider the evanescent waves in this work.
a set of functions \( F_n(x, k) : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{C} \) that represent \( N \) plane waves arriving from angles \( \hat{\phi}_n = (n - 1)\Delta \phi, \quad n = 1, \ldots, N \) where \( N = 2\pi/\Delta \phi \):

\[
F_n(x, k) = e^{ik\cdot \hat{\phi}_n},
\]

(3.7)

where \( \hat{\phi}_n \) is the unit vector in the direction of the \( n \)th plane wave. We can then implement an orthogonalization to find a set of orthogonal basis functions \( G_n(x, k) \) over \( \mathcal{D}, n \in \mathcal{A} \).

In practice this orthogonalization can be performed using a QR factorization by means of a modified Gram-Schmidt process [68] on discrete functions:

\[
G_n(x, k) = F_n(x, k) - \sum_{i=0}^{n-1} \frac{\langle F_n(x, k), G_i(x, k) \rangle_w}{\langle G_i(x, k), G_i(x, k) \rangle_w} G_i(x, k),
\]

(3.8)

which results in

\[
G_n(x, k) = \sum_{j \in \mathcal{A}} R_{jn} F_j(x, k),
\]

(3.9)

where \( R \) is a lower triangular matrix.

Using (3.9) and (3.3), it is now straightforward to express the approximation of the actual soundfield:

\[
S^a(x, k) = \sum_{n,j \in \mathcal{A}} E_n R_{jn} F_j(x, k),
\]

(3.10)

where \( R_{jn} \) denotes the \((j, n)\)th component of matrix \( R \). Therefore, \( P_j = \sum_n C_n(R)_{jn} \) specifies the coefficient for the \( j \)th \((j \in \{1, \ldots, N\})\) plane wave function to construct the desired multizone soundfield.
3.4 Mode Limitedness

The $F_n(x, k)$ in (3.7) are spatially independent on $\mathbb{R}^2$ and, therefore, also on $\mathbb{D}$. As the set of physically feasible independent solutions is finite, selecting $\Delta \phi$ sufficiently small and $N$ sufficiently large leads to a set of functions that can describe any feasible solution.

Here we investigate the minimum value for $N$ required to describe any soundfield satisfying the wave equation with reasonable accuracy. Any arbitrary 2-D soundfield $S^a : \mathbb{D} \times \mathbb{R} \mapsto \mathbb{C}$ generated by any number of sound sources outside a region of radius $r$ can be expressed by the following representation $[48]$

$$S^a(x, k) = \sum_{m=-\infty}^{\infty} \alpha_m(k) J_m(k \|x\|) e^{im\phi},$$  \hspace{1cm} (3.11)

where $J_m(\cdot)$ is the Bessel function of order $m$ and $\alpha_m(k)$ are a set of coefficients for the soundfield.

Previous results $[62] [113]$ have shown that a reasonable number of basis functions to be used in (3.11) is $2M + 1$ with the truncational error less than 16%, where $M = \lfloor kr \rfloor$. Only the “active” basis functions with order $-M \leq m \leq M$ contribute significantly to the soundfield inside $\mathbb{D}$. The remaining basis functions are referred to as “inactive” in $\mathbb{D}$. This implies that the desired sound is mode-limited over $\mathbb{D}$. The mode-limited set of basis functions spans a subspace, which we refer to as the mode-limited subspace.

Now consider $N$ linearly independent plane waves in (3.7). Each plane wave can be expanded into $2M + 1$ mode-limited basis expansion with a reasonably small error. Hence we can express any combination of plane waves that describes the desired soundfield using the same set of $2M + 1$ mode-limited basis. Conversely this means we can use any independent set of $N = 2M + 1$ plane waves to determine the $2M + 1$-order basis expansion with sufficient accuracy.
For the case of 3-D reproduction, the 2-D basis functions described above need to be replaced with 3-D basis functions. The orthogonal set of basis functions \( \{ G_n \}_{n \in A} \) can be constructed using a modified Gram-Schmidt process similar to (3.8) on \( N_{3D} = (\lceil ekr/2 \rceil + 1)^2 \) plane wave functions [113] except the weighted inner product is calculated over the desired 3-D reproduction volume of radius \( r \). The directions of the \( N_{3D} \) discrete plane wave functions can be based on the library of 3-D packings in [115].

### 3.5 The Reproduction of the Linear-combined Plane Waves

In the first stage, we map the desired multizone soundfield onto the closest Helmholtz solution, in the least-squares sense, and express it as an expansion of plane wave functions in Eq. (3.10). With the outcome of the first stage, which is the coefficient set \( \{ P_n \} \), the multizone sound reproduction problem is reduced to the reproduction of plane waves \( F_n(x, k) \) over the region \( \mathcal{D} \).

We now present a generalization of the reproduction method proposed in [48] that facilitates practical implementations. The essence of this problem is to find a set of Fourier coefficients for the aperture function, such that it can be used to approximate the desired soundfield, and such that the aperture vanishes over an angular window. The latter property facilitates a reduced number of loudspeakers in a practical configuration (e.g. a semi-circle or even quarter-circle of loudspeaker arrangement).

We first write the loudspeaker aperture function \( \rho(\phi, k) \) on a full circle as a Fourier series expansion as it is a periodic function of the angle \( \phi \):

\[
\rho(\phi, k) = \sum_{m=-M}^{M} \beta_m(k) e^{im\phi},
\]

where \( M = \lceil kr \rceil \) is the length of truncation modes [113]. \( \{ \beta_m(k) \} \) are
the Fourier coefficients and \( \{ \beta_m^0(k) \} = \frac{2}{\pi H_1^m(kR)} \alpha_m^d(k) \) is derived for the full circular continuous aperture function of radius \( R \) in [48]. \( \alpha_m^d(k) \) uniquely represents the desired soundfield and \( H_1^m(\cdot) \) is a \( m \)th-order Hankel function of the first kind [49]. To illustrate, the desired soundfield coefficients for a plane wave source, \( \alpha_m^d(k) \) are given by:

\[
\alpha_m^d(k) = i^m e^{-im\phi_{pw}},
\]

where \( \phi_{pw} \) is the polar angle of the plane wave. Here in our multizone case, \( \alpha_m^d(k) \) is defined as the set of linear combined coefficients for the desired multizone soundfield based on the coefficient set \( \{ P_n \} \):

\[
\alpha_m^d(k) = \sum_j P_j \alpha_j^m(k),
\]

where \( \alpha_j^m(k) \) represents the desired soundfield coefficients for the \( j \)th plane wave function we introduced.

In our scenario we have a constraint on the aperture function. It is then natural to find the coefficients \( \{ \beta_m(k) \} \) that minimize the error

\[
\Gamma(\{ \beta_m(k) \}) = \sum_{m=-M}^{M} |\beta_m(k) - \beta_m^0(k)|^2
\]

subject to an inequality constraint on the power of the aperture function \( \rho(\phi,k) \) for the angular window \( 0 \leq \phi < \phi_0 \) where the aperture function should be zero (e.g. we set \( \phi_0 = \pi \) for the a semi-circle arrangement):

\[
\int_0^{\phi_0} \left| \sum_{m=-M}^{M} (\beta_m(k)e^{im\phi}) \right|^2 d\phi \leq \eta_c.
\]

Note that the corresponding number of required loudspeakers for an angular window of \( \phi_0 \) is \( Q_0 = \left[ \frac{\phi_0(2M+1)}{2\pi} \right] \).

We solve the optimization problem using Lagrange multipliers. Assuming
we are at the inequality bound, we minimize

\[ \eta_0 = \Gamma(\{\beta_m(k)\}) + \lambda \eta_c, \quad (3.17) \]

where the multiplier \(\lambda \geq 0\). From an alternative viewpoint, it can be seen that (3.17) defines a weighting between the constraint and the function \(\Gamma\) that is determined by \(\lambda\). Let \(e_m = [e^{-iM\phi}, \ldots, e^{iM\phi}]\). Then, the Fourier coefficients that optimize (3.17) are

\[ \beta_d = (I - \lambda \int_0^{\phi_0} (e_m^H e_m) d\phi)^{-1} \beta_0. \quad (3.18) \]

The solution of \(\beta_d = [\beta_{-M}(k), \ldots, \beta_M(k)]^T\) can then be used to calculate the \(q\)th loudspeaker weights \(l_q(k)\) in the desired non-zero aperture region required for approximately reproducing the desired soundfield [48]:

\[ l_q(k) = \sum_{m=-M}^{M} \beta_m(k) e^{im\phi_s} \Delta\phi_s, \quad (3.19) \]

where \(\Delta\phi_s = (2\pi - \phi_0)/Q\) is the angular spacing of the loudspeakers and \(\phi_q = \phi_0 + (q - 1)\Delta\phi_s\) is the angle of the \(q\)th loudspeaker from x-axis. Let \(S^a_{\text{disc}}(x, k)\) be the reproduced soundfield using the part-circle method with weights provided in (3.19). Then

\[ S^a_{\text{disc}}(x, k) = \sum_q l_q(k) i e^j H^{(1)}_0(k\|R\phi_q - x\|) \quad (3.20) \]

where \(R\phi_q\) denotes \(q\)th loudspeaker position and \(R\) is the radius of the part-circle where the loudspeakers are located.

Note that the loudspeaker arrangement is limited to a circular shape for this approach. In the following simulations, we also seek to recreate the desired multizone soundfield using a non-circular loudspeaker array. It is based on the least square design method in the mode domain of the
soundfield.\textsuperscript{2} The basic assumption for the Helmholtz equation is that the desired reproduction region is source-free [51]. As long as this basic assumption is satisfied, flexible loudspeaker array configurations can be applied to the desired multizone soundfield reproduction.

### 3.6 Smooth Weighting Functions

The weighting function $w(x)$ enables us to control the reproduction accuracy over various types of zones by different settings. In the previous section, we assigned fixed weight values for the discrete bright, quiet and unattended zones in (3.5). A so-called "non-robustness" problem of multizone sound reproduction was discussed in [4] and the author observed very obvious redundant sound between the two selected regions with its amplitude even greater than the sound in the bright zone. This might bring about unpleasant experiences for the listeners if they coincidentally get into these areas. Naturally, this issue can be alleviated by increasing the corresponding values of the weighting function for these unattended region. However, that is an ad-hoc way to do so. In this section, we attempt to address this issue by introducing a smoothly and continuously changing weighting function over the desired reproduction region.

In practice, the position of the listener’s head (ears) is not guaranteed to be stationary within the selected region due to the movement of its body.\textsuperscript{3} For each of the bright and quiet zone, we assume that the distribution of the listener’s head position follows a Gaussian distribution. Within $D$, we can represent the distribution of listener’s head position as a Gaussian distribution function of its distance to the center of the bright zone $d_b$. We

---

\textsuperscript{2}For more details of this method, please refer to [48].

\textsuperscript{3}Note that the scattering effects resulted from obstacles in the soundfield (e.g. human head) is not considered.
have:
\[ f(d_b) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(d_b - \mu)^2}{2\sigma^2}}, \] (3.21)

where the mean or expectation \( \mu = 0 \). If we assign the standard deviation \( \sigma = r/2 \), it implies that the probability that the listener’s head positioned within a circle of radius \( r/2 \) m from the center of the bright zone (which is defined as central region of the bright zone) is 68.3%. The probability that the listener’s head is within the pre-defined bright zone is 95.4%, which means that there is still a chance of 4.6% that the listener’s head can be found outside the bright zone and the listener might experience a blasting of sound in this case with our previous settings of weighting function. This model also applies to the listener in the quiet zone \( \mathbb{D}_q \).

To avoid the large excursions in signal amplitude in the unattended zone, we can represent the weighting function over the desired reproduction region as a combination of weighted Gaussian distribution functions:
\[ w(x) = \frac{a}{\sigma \sqrt{2\pi}} e^{-\frac{(\|x - O_b\|)^2}{2\sigma^2}} + \frac{b}{\sigma \sqrt{2\pi}} e^{-\frac{(\|x - O_q\|)^2}{2\sigma^2}}, \] (3.22)

where \( O_b \) and \( O_q \) are the center of \( \mathbb{D}_b \) and \( \mathbb{D}_q \) respectively. Note that we can set up a minimum value for the weighting function \( w(x) \) in \( \mathbb{D} \) to avoid approaching to zero over the unattended zone. With this setting of the weighting function, the system will distribute the importance of the reproduction accuracy over different zones in a more flexible and efficient manner due to the introduction of the smoothly and continuously changing weighting function. More emphasis will be attached to the region where the listener’ ears are more likely to appear (e.g. the central region of the bright and quiet zone), while the reproduction effort might be distracted in some region (e.g. the edge of the bright and quiet zone) in order to alleviate the sound blasting in the unattended zone. The simulation results in Sec. 3.7.3 verify that the smooth weighting functions leads to improved system performance in practical setting.
3.7 Results and Discussion

We consider a multizone reproduction with a bright zone and a quiet zone, each of radius 0.3 m within $\mathbb{D}$ of radius $r = 1$ m at a frequency of 2 kHz. The centers of $\mathbb{D}_b$ and $\mathbb{D}_q$ lie on a circle of radius $d = 0.6$ m. The target bright and quiet zone are located at $\phi_1$ and $\phi_2$ respectively as shown in Fig. 3.1. We first set the weighting function $w(x)$ as: $a = 1$, $b = 2.5$ and $c = 0.05$. We attempt to recreate a plane wave arriving from angle $\phi_d$ in $\mathbb{D}_b$, while also attenuating the sound in $\mathbb{D}_q$.

3.7.1 Basis Function Approach

We start with simulations in Fig. 3.2 illustrating how the desired multizone soundfield is described by a feasible Helmholtz solution using the method introduced in Sec. 3.3. We set $\Delta \phi = \pi/40$ in (3.7) and the number of degrees of freedom (the number of orthogonal basis in the set) is 80. In

Figure 3.2: Describing the desired multizone soundfields with a basis expansion. (a) and (b) are for the case when $\phi_d = 45^\circ$ and $\phi_d = 60^\circ$ respectively.
Figure 3.3: Describing the desired soundfield due to a virtual point source within the bright zone using the basis expansion. (a) and (b) represents the real and imaginary part, respectively. “o” marks the position of the virtual source.

Fig. 3.2(a), $\phi_1 = 180^\circ$ and $\phi_2 = 0^\circ$, while a plane wave at $\phi_d = 45^\circ$ is created in $D_b$. The synthesized multizone soundfield features an acoustic contrast of 40.5 dB between $D_b$ and $D_q$, while the MSE of reproduction in $D_b$ is -29.7 dB. In Fig. 3.2(b), we have $\phi_d = 60^\circ$, $\phi_1 = 225^\circ$ and $\phi_1 = 45^\circ$. The sound between $D_b$ and $D_q$ was strongly occluded, which made this multizone reproduction scenario more challenging. We achieved an acoustic contrast between the zone of $D_b$ and $D_q$ of 21.4 dB, while still maintaining an accurate reproduction of the desired plane wave in $D_b$ with a MSE of -13.8 dB.

Additionally, we also examined the reproduction of the desired soundfield over the selected bright zone due to a virtual point source using the proposed plane wave expansion\(^4\). In Fig. 3.3, a circular wave due to a virtual point source at (-0.2,0.3) is reproduced over $D_b$, while cancelling \(^{4}\text{The evanescent contribution of the desired soundfield due to the virtual source is ignored.}\)
3.7. RESULTS AND DISCUSSION

the sound leakage over $\mathbb{D}_q$ simultaneously. The synthesized multizone soundfield features an acoustic contrast of 26.5 dB between $\mathbb{D}_b$ and $\mathbb{D}_q$, while the MSE of reproduction in $\mathbb{D}_b$ is -18.8 dB. Fig. 3.3 demonstrates the capability of reproducing desired soundfields other than plane waves using the the proposed basis expansion.

![Figure 3.4: Performance as the desired plane wave is panned, comparing reproductions as parameters $b$ are varied (solid curves), at different frequencies (dashed). Plotted are (a) an multizone sound reproduction design with one bright zone and one quiet zone (b) the MSE in $\mathbb{D}_b$ and (c) acoustic contrast between the bright and quiet zone.](image)

Fig. 3.4 studies system performance as the plane wave angle $\phi_d$ over the bright zone in Fig. 3.2(a) is panned. We can observe that the panning angle of the desired plane wave affects the system performance and that the worst performance is achieved when the plane wave is in-line with both zones, while the best performance is obtained when it is perpendicular with the line drawn through the centres of the zones. We vary the weighting parameter $b \in \{10, 2.5, 1, 0.1\}$ for $\mathbb{D}_q$ (we assign a fixed value of $a = 1$ to $\mathbb{D}_b$ and $c = 0.05$ to the unattended zone) to investigate the design trade-off between the MSE of reproduction in $\mathbb{D}_b$ and the acoustic contrast. The results show that as we increase the relative reproduction importance
over $\mathcal{D}_q$, it improves the acoustic contrast between the two zones at the expense of a decline in the reproduction accuracy over $\mathcal{D}_b$. In addition, we also investigated the performance of our approach at different frequencies: 1, 3 and 5 kHz. We can see that the system provides similar performance at these frequencies.

![Figure 3.5: Desired multizone reproduction using the approach of part-circle with an angular window of $\phi_0 = \pi$. 39 loudspeakers are used and the red circles demonstrate the positions of loudspeakers. (a) and (b) represent the cases with $\phi_d = 45^\circ$ and $\phi_d = 60^\circ$ respectively.](image)

3.7.2 Desired Multizone Reproduction using Non-circular Loudspeaker Array

Fig. 3.5 demonstrates the desired multizone reproduction of the above-mentioned two scenarios using a part-circle with an angular window of $\phi_0 = \pi$ at the frequency of 2 kHz while $R = 1.5$m. We set the Lagrangian $\lambda = 10$ in (3.17). Overall, the number of the employed loudspeakers is 39 and only the lower part of loudspeakers are used, while a circular array of 77 loudspeakers is required using the existing reproduction method.
3.7. RESULTS AND DISCUSSION

proposed in [48]. Note that we merely adopt the orthogonal set which consists of basis plane waves arriving from 0 to $\pi$. The rationale of doing this is that physically we are not able to render sound waves travel towards the semicircle of loudspeakers and the introduction of the other half set of basis plane waves would lead to large reproduction errors overall.

![Figure 3.6: Desired multizone reproduction using a square array of loudspeakers. 75 loudspeakers are used and the red circles demonstrate the positions of loudspeakers. (a) and (b) represent the cases with $\phi_d = 45^\circ$ and $\phi_d = 60^\circ$ respectively.](image)

We can observe that the reproduced multizone soundfields in Fig. 3.5 correspond well to the desired sound over $\mathbb{D}$. For the first case with a $\phi_d = 45^\circ$, the acoustic contrast between the two zones is 34.7 dB while the MSE in $\mathbb{D}_b$ is -26.8 dB. For the other case with a $\phi_d = 60^\circ$, the acoustic contrast drops to 19.5 dB while the MSE in the bright zone is now -13.7 dB. By using
the semi-circle method, we achieve the same reproduction accuracy over $D_b$ compared with the results using the reproduction method with a full continuous aperture function [48] in the second step. We note a reasonable decline in terms of the acoustic contrast between the two zones with the semi-circle method, which is due to the reduced number of employed loudspeakers, as well as the input effort [96].

Fig. 3.6 demonstrates the desired multizone reproduction of the above-mentioned two scenarios using a non-circular loudspeaker array at the frequency of 2 kHz. For demonstration purposes, we employed a square-shape loudspeaker array with $Q = 2M + 1 = 75$ loudspeakers. The size of the square-shaped loudspeaker array was $1.2 \text{ m} \times 1.2 \text{ m}$. We equally placed 18 loudspeakers on one side of the square-shaped loudspeaker array and 19 loudspeakers on each of the other three sides respectively. The resulting reproduced field is shown in Fig. 3.6.

We can observe that the reproduced multizone soundfields in Fig. 3.6 correspond well to the desired sound over $D$. The synthesized multizone soundfield for the first scenario features an acoustic contrast of $40.4 \text{ dB}$ between $D_b$ and $D_q$, while the MSE of reproduction in $D_b$ is $-27.5 \text{ dB}$. The results for the second scenario are $21.35 \text{ dB}$ and $-13.8 \text{ dB}$ respectively. It means the system performance is exactly the same with the results for the two scenarios shown in Fig. 3.2, which is an ideal description of the multizone soundfield. However, an enormous amount of redundant sound could be observed outside the desired reproduction region due to a high level of total energy of the signals driving the loudspeakers. This is due to the ill-conditioning problem in the sound reconstruction, which was discussed in Chapter 2.

### 3.7.3 Reproduction with Smooth Weighting Function

To investigate the system performance with the proposed smooth weighting function discussed in Sec. 3.6, we took the second scenario in Fig. 3.2 with a $\phi_d = 60^\circ$ for example at the frequency 1500 Hz. The parameters $a$,
3.7. RESULTS AND DISCUSSION

$b$ and $c$ enable us to adjust the design trade-off between the reproduction over different regions. In this case we have $a = 2.5$, $b = 5$ in (3.22). A minimum value of 0.01 for the weighting function $w(x)$ was introduced over $\mathbb{D}$. We map the generated weighting function $w(x)$ in Fig. 3.7(a). The corresponding reproduced soundfield is shown in Fig. 3.7(b).

As we can see from Fig. 3.7(b), the reproduced multizone soundfield corresponds well to the desired sound over $\mathbb{D}$. More specifically, the MSE in the central region of $\mathbb{D}_b$ is -14.35 dB and the acoustic contrast between the central regions of the bright and quiet zone is 22.5 dB, which are even better performance compared with the results with previous settings of weighting function (i.e. -13.47 dB and 20.64 dB, respectively). This means we can achieve a better system performance over the region in which the listener’s ears are most likely to be located. In the meantime, the ambience in the unattended zone becomes much less noisy and the excursions in loudness are greatly alleviated compared with Fig. 3.2(b). Note that the reproduction accuracy at the edge of both the bright zone and the quiet zone was sacrificed in order to facilitate the system performance in a more
practical manner.

Figure 3.8: Two complementary scenarios to Fig. 3.7(b): Fig. 3.8(a) shows the situation where the bright and quiet zone in Fig. 3.7(b) are exchanged for $\phi_d = 30^\circ$, while Fig. 3.8(b) features two bright zones and the two plane waves with the arriving direction of $60^\circ$ and $30^\circ$ are superimposed.

Fig. 3.8 illustrates two complementary scenarios to Fig. 3.7(b): Fig. 3.8(a) shows the situation where the bright and quiet zone in Fig. 3.7(b) are exchanged for $\phi_d = 30^\circ$ (representing the other listener in the top bright zone); Fig. 3.8(b) features two bright zones and the two plane waves with the arriving direction of $60^\circ$ and $30^\circ$ are superimposed, which represents the scenario in a real room with two listeners (e.g. the amplitudes of the sound over the two bright zones are determined by the music they are playing out). For Fig. 3.8(b), the MSE in the central region of top and bottom bright zones are -13.59 dB and -14.16 dB respectively, which represent accurate reproduction of the desired soundfield over the two regions.

Fig. 3.9(a) demonstrates the desired multizone reproduction with the proposed smooth weighting function using a semi-circle loudspeaker array at the frequency of 1500 Hz. The number of the employed loudspeakers was 56. Fig. 3.9(b) shows the multizone reproduction with the previous
weighting function which assigns a fixed value to each of the pre-defined zone. From the two figures, we can see that our proposed smooth weighting function also decreases the presence of sound leakage outside the desired reproduction region.

![Figure 3.9: Multizone sound reproduction using a semi-circle loudspeaker array at the frequency of 1500 Hz. (a) demonstrates the reproduction with the proposed smooth weighting function, and (b) shows the reproduction with the previous weighting function which assigns a fixed value to each of the pre-defined zones.](image)

### 3.7.4 Comparison with Prior Work

In this section, the desired multizone soundfield reproduction using the proposed basis function approach is compared with the work in [1], which is considered as the state-of-the-art. In [1], the multizone sound reproduction was formulated as a convex optimization problem to control the sound in each zone individually (including the rendering of an acoustically quiet zone).

The arrangements of the selected bright zone and quiet zone follows the settings in Fig. 3.2(b). We investigated the wide-band multizone
soundfield reproduction from 100 to 2 kHz. Overall, a full-circular array of 60 loudspeakers was employed for both approaches. For the proposed basis function approach, the weighting function $w(x)$ was set as: $a=1$, $b=5$ and $c=0.01$. For the method in [1], the attenuation factor for the quiet zone rendering was 0.001 and the loudspeaker weight energy constraint parameter was 40. Three measures were introduced for the performance comparison: the acoustic contrast between $D_b$ and $D_q$, the MSE over $D_b$ and the total loudspeaker weight energy (LWE) [1].

From Fig. 3.10, we can see that the constrained optimization in [1] is shown able to hold the acoustic contrast between the bright and quiet zone and the total LWE constant within the frequency range of interests. The constrained optimization method outperforms the proposed basis func-

Figure 3.10: Wide-band multizone soundfield reproduction using the proposed basis function approach and the convex optimization method in [1].
tion approach below 800 Hz at the expense of higher LWE consumption. For frequencies beyond 800 Hz, the MSE over $\mathbb{D}_b$ using the constrained optimization method drops significantly but the system consumes less LWE comparing with the proposed basis function approach. Overall, both methods are shown to be capable of creating the zone of quiet and rendering the desired soundfield over the bright zone effectively. The constrained optimization in [1] limits the sound energy leakage into other listener zones to fixed levels and the constrained consumption of LWE leads to an improved system robustness at high frequencies. However, the method shows inferior performance in terms of the reproduction accuracy over $\mathbb{D}_b$ at high frequencies. Additionally, the work in [1] is based on the equalization on discrete sets of multiple points, which has been shown to be vulnerable to reverberant settings [2]. Our proposed basis function approach is extended to the applications in reverberant environments in the following chapters.

3.8 Conclusion and Contribution

We proposed a novel approach of describing a desired 2-D multizone soundfield as an orthogonal expansion of basis functions over the desired reproduction region. It provides the theoretical foundation for an efficient parametrization of any feasible soundfield throughout the thesis. The proposed soundfield basis expansion facilitates sparse representations, as well as a stable reproduction over a wide frequency range (which will be shown in Chapter 4). The approach was shown to be effective both in controlling the rendering accuracy in the bright zone and the leakage into the quiet zone, particularly when occlusions occur between the two zones. The method also enables us to tailor the overall system performance by adjusting the parameters in the weighting function. In addition, we presented a generalization of the reproduction method of [48]. This approach allows us to diminish the number of employed loudspeakers
and obtain a more practical configuration.

The major contributions made in this chapter are:

- Formulation of the set of basis functions that is orthogonal over the desired reproduction region using modified Gram-Schmidt with a suitable set of plane wave functions as input.

- A generalization of the reproduction method proposed in [48] that facilitates a reduced number of loudspeakers in a practical configuration (e.g. a semi-circle or even quarter-circle of loudspeaker arrangement).

- The concept of a smoothly and continuously changing weighting function over the desired reproduction region, which leads to improved system performance in practical settings.
Overview: Multizone soundfield reproduction over an extended spatial region is a challenging problem in acoustic signal processing. We introduce a method of reproducing a multizone soundfield within a desired region in reverberant environments. It is based on the identification of the acoustic transfer function (ATF) from the loudspeaker over the desired reproduction region using a limited number of microphone measurements. We assume that the soundfield is sparse in the domain of plane wave decomposition and identify the ATF using sparse methods. The estimates of the ATFs are then used to derive the optimal least-squares solution for the loudspeaker filters that minimize the reproduction error over the entire reproduction region. Simulations confirm that the method leads to
4.1 Introduction

Spatial multizone soundfield reproduction, which aims to provide an individual sound environment to each of a set of listeners without the use of physical isolation or headphones, has drawn researchers’ attention recently. However, the majority of the existing works in spatial soundfield reproduction mainly concentrate on a single zone. Various, often-related techniques exist for single-zone reproduction, including ambisonics [31] [32], least squares techniques [35, 36, 67], wave field synthesis approach [25–27, 30] and the systems based on spherical harmonics [2, 48, 113]. The realization of multizone soundfield reproduction is a conceptually challenging problem and only a limited quantity of work is reported in the literature.

The origin of work in the multizone area can be traced to 2002, when [95] proposed an acoustic contrast control method to maximize the ratio of the mean square sound pressure in the so-called bright and dark zones. Further improvements of this work were made in [43] to enhance the computational robustness. These approaches are not able to reproduce a desired soundfield. In 2008, Poletti [4] proposed a least squares method to generate a 2-D monochromatic multizone soundfield. In [98] [99], the authors devised a multizone sound system by modeling sound pressures as a rectangular window corresponding to bright and dark zones and deriving the spatial filter based on the spectral division method (SDM) [56]. In [44], a method of describing the desired multizone soundfield as an orthogonal expansion of basis functions over the reproduction region was proposed. The basis function set was formed by implementing a modified Gram-Schmidt process on a suitable set of plane wave functions.
The fore-mentioned approaches do not take into account the reverberant environments in multizone sound reproduction. The reverberant case is difficult to handle because of the rapid variation of the ATF over the room [101]. In practical scenarios, the performance of multizone sound reproduction techniques is commonly degraded by the effects of reverberation. The traditional approach for spatial sound reproduction in a reverberant setting is pressure matching, which equalizes the transfer functions over a discrete set of points using least squares techniques [102] [77]. This technique leads to poor robustness in regions further away from the design points and inaccurate reproduction [2].

The soundfield reproduction techniques in [2, 3, 45, 46] were proposed based on describing the desired soundfield by the cylindrical/spherical harmonic expansion representation and have some practical drawbacks. Firstly, they require a relatively large number of measurements to estimate the ATFs of the loudspeakers with the needed accuracy. In addition, these approaches lead to problems for systems with a wide frequency range as the coefficients for the defined basis function cannot be determined at a number of frequencies due to the so-called large error scaling [2]. This irregularity problem has been studied extensively in [103–105] and it is associated with the nonuniqueness of the solution of the soundfield description using the boundary sampling techniques when the wavenumber is one of the Dirichlet eigenvalues, which adversely affects the wide-band multizone soundfield reproduction. The issue can be overcome by using alternate systems [2, 107–109]. However, these methods impose additional constraints on the hardware and its arrangement. For example, either using dual spherical microphone arrays (which doubles the number of required microphones), or mounting the microphones around a rigid sphere (which disables an open-sphere configurations), or using multiple high-order or directional microphones have been proposed in the literature. These additional constraints make the systems less practical to implement.
Recently, the technique of sparse approximation has been introduced to the area of spatial acoustics. In [116], a soundfield reproduction system in an anechoic room using the least-absolute shrinkage and selection operator (Lasso) was proposed by exploiting the sparsely-distributed loudspeakers in space. The results suggest that Lasso-based pressure matching yields better performance than the traditional least-squares method especially at a relatively high frequency when the number of provided loudspeakers is limited. However, the optimal selectivity of the loudspeaker placement using Lasso varies for different desired soundfields and at various frequencies, which makes the Lasso-based system less practical to implement in real world. Furthermore, the work in [117] shows that the Lasso approach is less accurate than least-squares method for wideband sound as Lasso does not employ all selected speakers to reproduce the wavefields for all frequencies. A similar sparse concept was used in [118] [119] to improve the soundfield rendering accuracy with a HOA system when an abundance of loudspeakers is employed. The work in [120] suggests that using sparsity makes it possible to reconstruct the desired soundfield over a planar region from a small number microphone measurements. Mignot et al. [121] extended a similar method to the interpolation of room impulse responses at low frequencies. In these works, the key step is to find a sparse representation for the desired soundfield, so that the ill-posedness of the original under-determined system can be eliminated. Similar assumptions will also be made in our work.

In this chapter, we present a novel method of reproducing a desired multizone soundfield over the reproduction region in reverberant environments. The proposed approach consists of two distinct stages. We first identify the ATFs from the loudspeakers over the reproduction region using sparse approximation techniques. The method is based on separating the actual loudspeaker ATF into a basic component that consists of the free-field Green’s function and a corrective soundfield
due to room reverberation. We assume that the corrective soundfield is sparse in the domain of plane wave decomposition, i.e. the corrective soundfield results from only a relatively small number of plane wave functions. This sparseness assumption facilitates an accurate corrective soundfield reconstruction based on scarce sound pressure measurements at randomly-selected locations within the desired regions.\textsuperscript{1} The estimates of the ATFs from the loudspeakers are then used to derive the optimal loudspeaker filter solution that minimizes the mean squared error (MSE) between the desired multizone sound and the actual reproduced sound over the entire reproduction region. Simulation results confirm that the approach provides accurate soundfield reproduction with a smaller number of microphones than existing methods while also facilitating the wide-band reproduction of the desired multizone soundfield without imposing additional constraints on the microphone array. We also extend the theory we develop in this chapter to 3-D multizone soundfield reproduction.

\textsuperscript{1}In this work, we assume that the microphones are location-informed.

Figure 4.1: The 2-D multizone soundfield reproduction over the desired reproduction region using $Q$ loudspeakers in a reverberant rooms.
4.2 Problem Formulation and Notation

We seek to control the reproduction of the desired multizone soundfield over each of a set of pre-defined zones within a reverberant enclosure. In this work, we consider disc-shaped zones for 2-D case. The theory we develop in this chapter is readily extended to 3-D space. Throughout this chapter, we use the following notation: matrices and vectors are represented by upper and lower bold face respectively, e.g., $C$ and $y$. $x \in \mathbb{R}^2$ denotes an arbitrary spatial observation point. $\| \cdot \|$ denotes the Euclidean norm and $[\cdot]^H$ denotes the Hermitian transpose. The imaginary unit is denoted by $i = \sqrt{-1}$. The superscripts “d” and “a” are used to represent the desired soundfield and the actual soundfield, respectively.

As illustrated in Fig. 4.1, the desired reproduction region $D$ is the entire control zone of interest with a radius of $r$, which includes both the acoustic bright zone $D_b$ and the quiet zone $D_q$ of radius $r_q$. The target bright and quiet zones are located at angles $\phi_1$ and $\phi_2$ respectively. The remaining area in $D$ is define as the unattended zone $D_u$. $k = 2\pi f/c$ is the wavenumber, where $f$ is the frequency and $c$ is the speed of sound propagation. The air absorption of sound is neglected in this work as we focus on indoor sound propagation in a reverberant room with finite size.

We assume no sound sources or scattering objects being present inside the reproduction area. The number of employed loudspeakers is $Q$. For simplicity, the loudspeakers are assumed to behave as omnidirectional point sources.

4.3 Evaluation Measures

The goal is to describe any related soundfield using the basis function set $\{G_n\}_{n \in A}$ and determine the loudspeaker filter weights required to reproduce the desired multizone sound in a reverberant room based on the characterization of the ATF for each of the loudspeaker. To evaluate
the performance of our work we use the following three measures:

- The acoustical brightness contrast between the bright zone \( \mathcal{D}_b \) and the quiet zone \( \mathcal{D}_q \) to quantify acoustic contrast between the two zones:

\[
\zeta(k) = \frac{\int_{\mathcal{D}_b} |S^a(x, k)|^2 \, dx / S_b}{\int_{\mathcal{D}_q} |S^a(x, k)|^2 \, dx / S_q},
\]

where \( S_b = \int_{\mathcal{D}_b} 1 \, dx \) and \( S_q = \int_{\mathcal{D}_q} 1 \, dx \) mark the sizes of \( \mathcal{D}_b \) and \( \mathcal{D}_q \) respectively.

- The mean square error (MSE) between the desired sound \( S^d(x, k) \) and the actually rendered sound \( S^a(x, k) \) over the zone \( \mathcal{D}_b \) to gauge the reproduction accuracy:

\[
\epsilon_M(k) = \frac{\int_{\mathcal{D}_b} |S^d(x, k) - S^a(x, k)|^2 \, dx}{\int_{\mathcal{D}_b} |S^d(x, k)|^2 \, dx}.
\]

- The number of required microphones.

### 4.4 Green’s Function Modeling

The accurate characterization of the loudspeaker ATF is essential for the multizone sound reproduction system as large reproduction errors may result from small perturbations of the ATFs from the loudspeakers to the listener zones in scenarios when one zone is obscured by another [4]. In this section, we introduce a novel method to determine the ATFs from the loudspeakers over the desired regions using the concept of compressed sensing. Compressed sensing [122] [123] is a signal processing technique for efficiently acquiring and reconstructing a signal from an under-determined linear system, assuming the observed phenomenon is known a priori to be sparse.
4.4. GREEN’S FUNCTION MODELING

4.4.1 Estimation of the ATFs in Reverberant Rooms

The objective of this section is to estimate the ATF from the \( q' \)-th loudspeaker \( T_{q'}(x, k) \). Repeating the estimation procedures discussed in this section \( Q \) times, the ATFs from all the loudspeakers can be interpolated over the desired regions using scarce measurements.

The soundfield \( T_{q'}(x, k) \) is first separated into a basic component, the 2-D free-field Green’s function, and an unknown corrective soundfield \( R_{q'}(x, k) \)

\[
T_{q'}(x, k) = -\frac{i}{4} H_0^{(2)}(k \| h_{q'} - x \|) + R_{q'}(x, k),
\]

where \( h_{q'} \) represents the position of the \( q' \)-th loudspeaker and \( H_0^{(2)} \) is the zeroth-order Hankel function of the second kind [49]. For the 3-D case, the basic component of ATF from the \( q' \)-th loudspeaker equals to \( e^{-ik \| h_{q'} - x \|} / (4\pi \| h_{q'} - x \|) \).

An arbitrary soundfield function satisfying the wave equation can be written as a superposition of a set of \( N' \) solutions of the homogeneous Helmholtz equation (the solutions can be non-orthogonal) [51]. Therefore, we can write \( R_{q'} : \mathbb{R}^2 \times \mathbb{R} \mapsto \mathbb{C} \) as a weighted series of the function set \( \{ P_n \}_{n \in A} \) (where \( A = \{ 1 \ldots N' \} \) and \( N' \) is selected to be sufficiently large)

\[
R_{q'}(x, k) = \sum_{n=1}^{N'} y_n P_n(x, k),
\]

where the coefficient set \( y \in \mathbb{C}^{N'} \). Theoretical results in [55] indicate that plane waves provide good approximations to any feasible soundfield on any disc-shaped zone under any type of boundary conditions.

In the well-known image source method [72], which provides a good model of room reverberation, walls correspond to virtual image sources that are far away from the observation points in many practical scenarios. Therefore, the propagated wavefields from the image sources resemble [82]
plane waves over the observation points. This suggests that plane wave representations provide a more efficient and sparse parametrization of the corrective soundfield than the conventional harmonic expansion representation in reverberant environments.

The basic principle of the proposed method is to assume that \( R_q'(x, k) \) can be sparsely identified in the domain of plane wave decomposition, i.e. the corrective soundfield results from only a relatively small number of plane waves. Therefore, we refer to \( y \) in (4.4) as a \( K \)-sparse signal, which means that \( y \) has only \( K \) (\( K \ll N' \)) non-zero entries at unknown locations while the other entries are zero (or very close to zero). The value of \( K \) depends on how complicated the reverberant environment is. Note that we only estimate the corrective part of \( T_q'(x, k) \) since the elimination of the free-field Green’s function generally increases the sparsity level of the signal in the plane wave function domain and leads to a better estimation using a limited number of measurements.

Based on the assumptions listed above, we put forward a linear system

\[
v = \Phi y, \tag{4.5}
\]

where the dictionary \( \Phi \) is an \( m \times N' \) sensing matrix (\( N' \gg m \)) whose columns are the normalized measurement vectors of an overcomplete set of plane wave functions \( \{P_n\}_{n \in A} \). \( v \) is an \( m \times 1 \) observation vector which contains the values of the desired corrective soundfield \( R_q'(x, k) \) at \( m \) randomly chosen locations within the desired region. From (4.3), we see that the values of \( v \) are the difference between the original measurements of the soundfield and the value of the free-field Green’s function. In practice, the ATF measurements of the loudspeakers at certain wavenumber \( k \) are based on the room impulse responses (RIR) of the loudspeakers at the sampling position, which can be captured by using the logarithm swept chirp method [124].

In [123], the number of required measurements for exact reconstruction of the original signal with high probability is generalized as a function of the
sparsity level $K$. Although the value of $K$ is always unknown in this work, theoretical study in the field of compressed sensing has shown that the so-called Restricted Isometry Property (RIP) [125] of the matrix $\Phi$ guarantees that sparse approximation will provide an accurate and stable estimate of $y$ from the insufficient observation $v$, when $y$ is sufficiently sparse. It is suggested that various families of random matrices satisfy the RIP with very high probability. In the area of spatial soundfield, however, it is very unlikely to find a $\Phi$ that fully verifies the RIP due to the limited freedom in the design of the acquisition scheme [120]. In this work, $\{P_n\}_{n \in A}$ is selected to be $N'$ independent plane waves arriving from various angles. Therefore, the measurements in $v$ are the products of rows of the sensing matrix and the sparse signal $y$. Randomness is introduced by randomly locating the microphones within the desired region. The RIP condition essentially implies that the observation value is the linear projection of the sparse signal onto an incoherent basis (see, e.g. [126] for an explanation of incoherence). Our formulation is consistent with this requirement as the random samplings of the sound pressure field in $v$ is incoherent with the original basis of $y$. The estimation results in Sec. 4.7 confirm the validity of the proposed approach with a significantly reduced number of microphones.

An $l^p$ norm (where $0 < p < 1$) non-convex optimization problem is considered to produce an accurate estimate of $y$ in (4.5) with scarce measurements [127]:

$$\min_y \|y\|^p_p, \text{ s.t. } v = \Phi y,$$

(4.6)

where the $l^p$ norm is defined as $\|y\|_p = (\sum_{i=1}^{N'} |y_i|^p)^{1/p}$. It was shown in [128] that the optimization in (4.6) requires fewer measurements for precise reconstruction than the $l^1$ norm optimization using the same sensing matrix. Therefore, the total number of degrees of freedom for (4.5) is decreased from $N'$ to $K$ given that the set of basis wavefields with
nonzero entries is well-fitted.

In any practical situation, the measured signal is corrupted by noise, so a more practical model is to reconstruct \( y \) from noisy under-sampled measurements \( v \). In such a case, the equality condition \( v = \Phi y \) should be relaxed to

\[
\min_y \| y \|_p, \quad \text{s.t.} \quad \| v - \Phi y \|_2^2 \leq \epsilon.
\] (4.7)

The additive noise is assumed to be Gaussian noise of zero mean and variance \( \sigma^2(k) \), which makes the Euclidean norm for the second term appropriate [129]. (4.7) can be converted to the following unconstrained form:

\[
\min_y \frac{1}{p} \| y \|_p^p + \frac{\lambda}{2} \| v - \Phi y \|_2^2,
\] (4.8)

where \( \lambda \) is related to the error allowance \( \epsilon \). Experimental results suggest that choosing \( \lambda \) in the range \( [\sigma^2(k) \sqrt{\log(N')} , \sigma^2(k) \sqrt{2 \log(N')} ] \) yields a fair reconstruction [129].

We can apply the regularized Iteratively Reweighted Least Squares (IRLS) algorithm [127] [129] to solve (4.36) and derive the optimal solution \( \hat{y} \) that estimates the corrective component of the soundfield generated by each loudspeaker in reverberant environments:

\[
R_{q'}(x, k) \approx \sum_{n=1}^{N'} \hat{y}_n P_n(x, k),
\] (4.9)

where \( \hat{y} \) is a sparse coefficient vector that has only \( m' (m' \leq m) \) non-zero components. Therefore, the ATF from the \( q' \)th loudspeaker over the desired regions can be approximated as follows according to (4.3) and (4.9):

\[
T_{q'}(x, k) \approx -\frac{i}{4} H_0^{(2)}(k\|h_{q'} - x\|) + R_{q'}(x, k).
\] (4.10)
4.4. GREEN’S FUNCTION MODELING

The above-mentioned estimation procedures is then repeated for the ATFs from all the \( Q \) loudspeakers and the estimates will be used for the loudspeaker filter design to reproduce the desired multizone soundfield in Sec. 4.5.

Since plane wave functions provide good approximations to any feasible soundfield that satisfy the wave equation on any disc-shaped zone under any type of boundary conditions, the Green’s function modeling approach with the over-complete set of plane wave functions can also be applied to the estimation in more complex environments that include the scattering effects resulting from objects (e.g. human head). However, the increase in the complexity of the desired ATF would lead to an increase of the required number of microphones for accurate estimation.

Comparing with the works in [120] [121], which also explored sparse approximation methods to reconstruct the desired soundfield over a planar region from a relatively smaller number microphone measurements, there are four major differences in our work: 1. The works in [120] [121] were interpolating the plenacoustic function (PAF) of the room impulse response, while the proposed method estimates the corrective part of the ATF in spatial-frequency domain; 2. The reference methods employed orthogonal matching pursuit (OMP) to reconstruct the sparse signal, while the proposed method is formulated to use the \( p \)-norm IRLS method, which generally leads to an accurate sparse estimation with less measurements; 3. The reference methods were shown to be effective for sampling the desired wavefield in a room with a reduced number of microphones only at low frequencies (below 400 Hz), while the proposed method facilitates a more sparse representation of the reverberant soundfield up to 5 kHz (as shown in Sec. 4.7); 4. The reference methods were limited to the sparse estimation of desired soundfield in reverberant rooms, while the proposed system further uses the derived ATF information to conduct multizone soundfield reproduction.
4.4.2 Approximate Cramér-Rao Bound for General Green’s Function Modeling Approach

In estimation theory and statistics, the Cramér-Rao bound (CRB) [130] [131] expresses a lower bound on the variance of estimators of a deterministic parameter. In Sec. 4.4.1, we make the approximation that $R_q(x, k)$ is $K$-sparse in the domain of plane wave decomposition, which means that $y$ is sparse and has maximal support (i.e. $\|y\|_0 = K$). It is shown in [132] that the unbiased CRB equals that which would be obtained if the sparsity pattern were known; this can also be considered as an “oracle bound”. However, the derivation of the oracle estimator requires a sufficient number of measurements so that it matches for the sparsity level $K$. In this section, we apply the oracular knowledge to derive the CRB on the variance of the corrective soundfield estimator $\hat{R}_q(x, k)$ with an insufficient number of microphones, which can be considered as an approximation to the oracle bound. As we aim at minimizing the number of employed microphones, the approximate CRB will underestimate the true CRB. Note that the proposed approach enables the microphones to be randomly placed within the desired region, so the CRB allows us to observe if the estimation performance of the desired ATF at different places varies with various arrangements of the microphone placement.

The oracle bound is based on the support of $y$ being identified, which means that the awareness of locations (but not the values) of the nonzero representation elements is provided. Therefore, we can construct a $m \times K$ sub-dictionary $\Phi_s$ whose columns contain the normalized values of the $K$ plane waves at the corresponding measurement location. Given the measurement vector $v$ and $\Phi_s$, we can derive an estimator $\hat{y}'$ (which only contains $K'$ non-zero entries where $K' \leq m$) using the IRLS algorithm [127] [128] and construct a submatrix $\Phi'$ from the columns of $\Phi_s$ corresponding to the $K'$ non-zero entries of $\hat{y}$: $\theta = [\theta_1, \ldots, \theta_{K'}]$.

Taking the noise corruption into account, the measurement process can be
modeled as [129],

$$v = \Phi'\hat{y}' + \eta,$$

(4.11)

where $\eta$ is independent and identically distributed (i.i.d.) complex Gaussian noise with zero mean and variance $\sigma^2$ for each microphone. In short, $v \sim \mathbb{C}\mathcal{N}(\mu(\theta), U)$, where $U = \sigma^2 I$ ($I'$ is an identity matrix) and

$$\mu_v = \sum_{k' = 1}^{K'} \Phi'_{vk'} \theta_{k'}, \quad v = 1 \ldots m.$$  

(4.12)

With (4.12), we can then write the probability density function (pdf) $v$ as

$$p(v; \theta) = \frac{1}{(2\pi\sigma^2)^{m/2}} e^{-\frac{1}{2\sigma^2} \sum_{v=1}^{m} (v - \sum_{k'=1}^{K'} \Phi'_{vk'} \theta_{k'})^2}. $$

(4.13)

Taking the natural logarithm of the pdf, we have

$$\ln p(v; \theta) = -\frac{m}{2} (\ln(2\pi) - \ln(\sigma^2))$$

$$- \frac{1}{2\sigma^2} \sum_{v=1}^{m} (v - \sum_{k'=1}^{K'} \Phi'_{vk'} \theta_{k'})^2.$$  

(4.14)

The pdf $p(v; \theta)$ may be transformed into a reparameterized version $p(v; \gamma)$ with

$$\gamma = (\alpha_1, \beta_1, \ldots, \alpha_{K'}, \beta_{K'})^T$$

(4.15)

by substituting $\alpha_{k'} + \beta_{k'} j$ for $\theta_{k'}$, where $\alpha_{k'}$ and $\beta_{k'}$ are the real part and the imaginary part of $\theta_{k'}$, respectively.

The CRB states that the variance of any unbiased estimator is at least as high as the inverse of the Fisher information [133]. The Fisher information
matrix (FIM) can be derived using the Slepian-Bang formula [134] as:

\[
[I(\gamma)]_{pq} = E\left[\frac{\partial \ln p(v; \gamma)}{\partial \gamma_p} \frac{\partial \ln p(v; \gamma)}{\partial \gamma_q}\right] = 2Re\left[\frac{\partial \mu^H}{\partial \gamma_p} U^{-1} \frac{\partial \mu}{\partial \gamma_q}\right] + \text{Tr}\{U^{-1} U_p U^{-1} U_q\}
\]

(4.16)

where \(U_p = \frac{\partial U}{\partial \gamma_p}\). Note that the second part of Eq. (4.16) is equal to 0 in this work.

For the estimation of the corrective part of the desired room transfer function \(R_q'(x, k) = g(\gamma) = \sum_{k'=1}^{K'} (\alpha_{k'} + \beta_{k'} j) F_{k'}(x)\) (where \(F_{k'}(x)\) represents the corresponding \(k'\)th plane wave function), we have [133]:

\[
\text{Var}{\hat{R}_q'(x)} - \frac{\partial g(\gamma)}{\partial \gamma} I^{-1}(\gamma) \frac{\partial g(\gamma)^H}{\partial \gamma} \geq 0,
\]

(4.17)

where \(\text{Var}{\hat{R}_q'(x)}\) defines the variance of \(\hat{R}_q'(x, k)\). Therefore, given the number of measurements (as well as the measurement location), the lower bound on the variance of estimators of \(\hat{R}_q'(x, k)\) is given by (4.17). In the case that \(I(\gamma)\) is singular (or very nearly so), the CRB should be modified by replacing \(I(\gamma)^{-1}\) with the Moore-Penrose pseudoinverse of \(I(\gamma)\) [135].

Eq. (4.17) reveals the optimal performance on the variance of the estimator \(\hat{R}_q'(x, k)\) when an insufficient number of microphones are used. Note that the CRB is a function of the location \(x\) and the wavenumber \(k\), which provides us with a good knowledge of the estimation accuracy that can be achieved for the desired corrective soundfield at the certain position. This is very critical for desired soundfield reproduction in reverberant rooms as the estimates of the loudspeaker ATFs are directly used to design the optimal loudspeaker filter gains.
4.5 Orthogonal Basis Function Approach

With the outcome of the first stage in Sec. 4.4.1, which is the estimate of the ATFs between loudspeakers over the desired reproduction region, we now present an orthogonal basis function approach that enables us to derive the optimal loudspeaker filter weights for the desired multizone sound reproduction.

In the basis function approach, we express the desired multizone soundfield $S^d(x, k)$, the actual reproduced soundfield $S^a(x, k)$ and the ATF for a given loudspeaker $T_{q'}(x, k)$ in Eq. (4.10) as a weighted series of basis orthonormal functions $\{G_n\}_{n \in A}$, which is constructed using the approach proposed in Sec. 3.3. Provided all sound sources lie outside $D$, we can use the following representations:

$$S^d(x, k) = \sum_{n=1}^{N} A_n G_n(x, k), \quad (4.18)$$

$$S^a(x, k) = \sum_{n=1}^{N} B_n G_n(x, k), \quad (4.19)$$

$$T_{q'}(x, k) = \sum_{n=1}^{N} C^q_n G_n(x, k), \quad (4.20)$$

where $N = 2M + 1$ ($M = \lceil kr \rceil$ is the truncation length for $D$) according to the discussion in Sec. 3.4. The coefficients for $S^d(x, k)$, $T_{q'}(x, k)$ can be derived as $A_n = \langle S^d(x, k), G_n(x, k) \rangle_w$, $C^q_n = \langle T_{q'}(x, k), G_n(x, k) \rangle_w$ respectively. The reproduced soundfield $S(x, k)$ resulting from the Q loudspeakers is equal to

$$S^a(x, k) = \sum_{q'=1}^{Q} l_{q'}(k) T_{q'}(x, k). \quad (4.21)$$
Substituting (4.19) and (4.20) into (4.21), the coefficients of the reproduced soundfield are related to $C_n^q$ through $B_n = \sum_{q'=1}^{Q} l_{q'}(k) C_n^{q'}$. 

With the reverberant setting, the main task of soundfield reproduction is to choose filter weights $l_{q'}$ to minimize the weighted squared norm over $D$: 

$$
\eta = \| S^a(x, k) - S^d(x, k) \|^2_{(w)}. 
$$

(4.22)

Substituting (4.18) and (4.19) into (4.22) and use the orthonormal property of $G_n(x, k)$ over $D$: $\langle G_i(x, k), G_j(x, k) \rangle_w = \delta_{ij}$, we see that the weighted squared error reduces to

$$
\eta = \sum_n |B_n - A_n|^2. 
$$

(4.23)

The optimal least-square solution for (4.23) is expressed in terms of the basis function coefficients. Let us define the vector of loudspeaker filter weights $l = \left[ l_1(k), \ldots, l_Q(k) \right]^T$, the vector of the coefficients of the reproduced soundfield $b = [B_1, \ldots, B_N]$ and the $N \times Q$ matrix of the coefficients of the room responses of all loudspeakers

$$
C = \begin{bmatrix}
C_1^1 & \cdots & C_1^Q \\
\vdots & \ddots & \vdots \\
C_N^1 & \cdots & C_N^Q
\end{bmatrix}. 
$$

(4.24)

We then have $b = Cl$. Additionally, we define the vectors of the coefficients of the desired multizone soundfield, $a = [A_1, \ldots, A_N]$.

In practical applications, it is advantageous to write (4.23) in matrix form and add a regularization term, both to constrain loudspeaker effort and ensure a robust solution:

$$
\eta = (b - a)^H (b - a) + \tau ||l||^2, 
$$

(4.25)

where $\tau \geq 0$. The optimal solution is then
where $I$ is an identity matrix. Once $(C^HC + \tau I)^{-1}C^Ha$ is computed for the complex acoustical environment, the reproduced soundfield can be changed by modifying $a$, which specifies the desired multizone soundfield.

### 4.6 Extension to 3-D Reproduction

The previous work is based on a 2-D setup. However, we are immersed in a 3-D environment in reality (i.e. the reverberation from the floor and ceiling should also be taken into account). The fundamental solution to the wave equation in the 3-D scenario is different from the 2-D case, which leads to a different set of optimal solutions for the loudspeaker filter design. The theory we developed above is readily extended to the 3-D space.

As shown in Chapter 2, a soundfield can be represented using a set of basis spherical harmonics functions in the 3-D spherical coordinates system. The spherical harmonics functions are functions of the elevation angle $\theta$ and the azimuth angle $\phi$, which are orthonormal over a unit sphere. An arbitrary soundfield within a source-free region can be expressed as an expansion of these spherical harmonics:

$$S(r, \theta, \phi; k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) P_{nm}(\cos \theta) E_m(\phi),$$

(4.27)

where $m$ and $n \ (n \geq 0)$ are integers, $\alpha_{nm}(k)$ are the spherical harmonic coefficients of the soundfield, $j_n(\cdot)$ are the spherical Bessel functions [136] of order $n$, $E_m(\phi) = (1/\sqrt{2\pi})e^{im\phi}$ are the normalized exponential functions and $P_{nm}(\cos \theta)$ are the normalized associated Legendre function [137].

Similar to the 2-D scenario, the spherical harmonic expansion of the soundfield (4.27) can also be truncated to a limited number of orders.
As we are interested in controlling the multizone soundfield restricted inside a region \( D \) of a defined radius \( R_r \), the infinite summation of order \( n \) in (4.27) can be truncated to the upper bound of \( \lceil keR_r/2 \rceil \) due to the property of the Bessel functions [62]. The spherical harmonic coefficients also vary over modes \( m \) ranging from \(-n\) to \( n\). Therefore, the total number of coefficients required to describe the desired soundfield is given by \( N_{3D} = (\lceil keR_r/2 \rceil + 1)^2 \) [113]. If we can control all these \( N_{3D} \) spatial coefficients, we can accurately recreate the desired multizone soundfield inside the desired region \( D \). It implies that \( Q \geq N_{3D} \) loudspeakers are needed to approximately reproduce the 3-D desired multizone soundfield within \( D \) with certain errors [113].

We can use the similar method for 2-D multizone sound reproduction to reproduce the desired multizone soundfield with 3-D settings. The following remarks list the differences between 2-D and 3-D techniques:

- To estimate the ATF between the \( q^{th} \) loudspeaker \( T_{q}^{3D}(x, k) \) in 3-D space, we first separate the actual soundfield into a basic component, the 3-D free-field Green’s function and an unknown corrective soundfield \( R_{q}^{3D}(x, k) \)

\[
T_{q}^{3D}(x, k) = \frac{e^{-ik\|Y_q' - x\|}}{4\pi\|Y_q' - x\|} + R_{q}^{3D}(x, k).
\] (4.28)

The basic principle of the method still holds and we assume that \( R_{q}^{3D}(x, k) \) is sparse in the 3-D plane wave function domain;

- For each of the selected zones, the active mode of the basis functions is assumed to be limited by \( N_{3D} = (\lceil ekr_q/2 \rceil + 1)^2 \), as we are only interested in the soundfield within a finite volume of bright and quiet zones;

- At the second stage (i.e. the basis function approach), the orthonormal set of basis functions \( \{G_n\}_{n \in A} \) over the desired reproduction region \( D \) is constructed using the modified Gram-Schmidt process.
4.7. RESULTS AND DISCUSSION

on $(\lceil ekr_q/2 \rceil + 1)^2$ plane wave functions. The directions of the basis plane wave functions are given by the library of 3-D packings in [115].

4.7 Results and Discussion

The main objective is to determine the ATFs of the loudspeakers accurately using a small number of microphones, so that it leads to an accurate multizone soundfield reproduction over $\mathbb{D}$. To define our multizone soundfield requirements, one bright zone and one quiet zone are selected.

4.7.1 Experimental Setup

As the work mainly focuses on the 2-D multizone soundfield reproduction, the reverberant room is an enclosed rectangle (size $6 \times 5$ m) with a wall absorption coefficient of 0.3 and the sound reflections from the floor and the ceiling are not considered. We assumed that the speed of sound $c$ is 340 m/s in our simulations. $\mathbb{D}$ has a radius of $r = 1$ m with its center located at $(2m, 2.5m)$ and the employed loudspeakers are evenly distributed along a concentric circle with a radius of 1.5 m. We used the image source method [72] to simulate the soundfield created by the loudspeaker in the reverberant room. In the simulations, a total of 60 sources (including both real and image sources) for each loudspeaker were included. The centers of $\mathbb{D}_b$ and $\mathbb{D}_q$ lie on a circle of radius $d = 0.6$ m within the source-free $\mathbb{D}$.

The target bright and quiet zones are located at $\phi_1 = 225^\circ$ and $\phi_2 = 45^\circ$, respectively, with $r_q = 0.3$ m, as shown in Fig. 4.1. In order to numerically observe how the number of microphones used would affect the estimation accuracy of the desired ATFs from the

\footnote{The scattering effects due to objects will not be included as the boundary condition discussed in Chapter 2 is only valid for a rigid circular object (which is rarely encountered in practice). Existing methods to simulate the effects due to irregular non-rigid objects are very limited.}

94
loudspeakers, we randomly selected $m/2$ locations within each zone and measured the value of the corrective soundfield at those positions. The size of the overcomplete dictionary $\Phi$ in (4.36) is $m \times 1000$. For the approach of IRLS in (4.36), the sparsity promoting norm is $p = 0.4$.

In this section, a comparison is made between the proposed work with prior art in two aspects. Firstly, we compared the Green’s function modeling approach in Sec. 4.4 with the estimation method proposed in [2] in terms of the interpolation of the the desired ATF. Note that the method of [2] requires the microphones to be evenly placed along the boundary of each selected zone. Secondly, the proposed approach was compared with the multizone soundfield reproduction method in [3] and the pressure matching method over multiple control points [4] with respect of the multizone soundfield reproduction (with a designated quiet zone) based on the three evaluation measures in Sec. 4.3. The method in [3] uses the estimation method of [2] to characterize the ATFs from the loudspeakers over the desired reproduction region.

### 4.7.2 Estimation of the Desired ATFs

We first illustrate the ATF estimation of a desired loudspeaker placed outside $D$ at $f = 1$ kHz. The loudspeaker requirement of this scenario is specified by the truncation length for $D$: $M = \lceil kr \rceil = 19$, which suggest the use of $2M + 1 = 39$ loudspeakers. The layout setup yields an average direct-to-reverberant energy ratio from each loudspeaker of -4.0 dB at the center of $D$. In Fig. 4.2, we demonstrate the performance with the following three different settings: noiseless setting, noisy setting with optimization constraint $\|v - \Phi y\|_2^2 \leq \epsilon$ (inequality constraint) and with optimization constraint of $v = \Phi y$ (equality constraint). For this experiment, we defined the reconstruction of the desired ATF over the two selected zones with an estimation error no greater than -20 dB as an accurate estimate.

For the ideal noiseless measurement case, we set $\lambda = 0$ in (4.36). From Fig.
4.7. RESULTS AND DISCUSSION

Figure 4.2: The estimated performance of the desired loudspeaker ATF is plotted as a function of the number of microphones used at 1 kHz with different settings. We ran 100 trials for each iteration with an assigned value of $m$.

4.2 we see that the use of 14 microphones for each selected zone facilitates an accurate estimate of the desired loudspeaker ATF with a probability of over 90%. In contrast, the estimation method proposed in [2] requires approximately 20 microphones to obtain the same level of accuracy with the same settings. The estimated performance is also plotted for several additive noise Signal-to-noise ratios (SNRs) while the values of $\lambda$ are selected accordingly. Naturally, the performance of reconstructions with the inequality constraint is superior to the estimation using the equality constraint. For the estimation result with the equality constraint, we can observe a decline after it reaches a peak as the number of noisy measurements increases.

Importantly, the estimation performance improves as the level of the added noise decreases. The results show that a 30 dB SNR would lead to an accurate reconstruction of the desired ATF at 1 kHz using a limited
number of noisy measurements. In practice, especially for the case with a relatively lower SNR on the ATF measurements, the noise corruption can be alleviated by averaging over a long sequence signal that contains multiple room responses from the microphone recording.

In Fig. 4.3(a), the proposed method is compared with the estimation method proposed in [2] in terms of the wide-band desired ATF estimation
in the frequency range from 100 Hz to 1 kHz with noisy samplings. We used 20 noisy measurements for each zone for the proposed method while a total of 40 measurements were employed for the comparative methods. The additive noise SNR was 30 dB. As we can see, the proposed method consistently outperforms the estimation method proposed in [2], typically up by about 8 dB. More importantly, the general trend of the proposed method is that the error smoothly increases along with frequency. This trend is due to the increase in the complexity of the desired ATF with frequency. In contrast, we can observe obvious peaks in the result of [2]. These peaks occur in the vicinity of the zeros of the Bessel terms due to the fact that the approximation error of the defined modal coefficients in [2] is weighted by the Bessel function term $1/J_{n'}(kR')$, where $n'$ is the number of mode order and $R'$ is the radius of sampling circle. As stated in Sec. 4.1, alternate systems can be applied to alleviate this “Bessel-zero problem” but additional hardware constraints on the microphone array are required. The error bars of the estimation performance at numerous frequencies are plotted in Fig. 4.3(b).

In this section, we also investigate the estimation robustness to microphone position mismatches in the case that the microphones are not ideally aligned in practice. A desired reproduction region of radius 0.3
m was defined and 17 microphone measurements were placed at the predefined locations within the region. To observe how the position mismatches affect the estimation performance, the microphone measurements were assumed to be noiseless. We simulated positioning errors by introducing 2-D normally distributed noise to the positions of the microphone measurements. In Fig. 4.4 (a), we demonstrate the desired ATF estimation accuracy from 100 Hz to 1 kHz with different positioning error settings. The variances of the positioning tolerance for each of the 17 measurements were chosen to be $\sqrt{2\text{mm}^2 + 2\text{mm}^2}$, $\sqrt{5\text{mm}^2 + 5\text{mm}^2}$ and $\sqrt{1\text{cm}^2 + 1\text{cm}^2}$, respectively. Comparing with the estimation performance in Fig. 4.4 (b), which applies zero positioning error, we can clearly see that the estimation performance deteriorates as the variance of the positioning error increases and good robustness to the position mismatch with the variance up to $\sqrt{5\text{mm}^2 + 5\text{mm}^2}$ can be observed.

The simulation results in this section verify the underlying assumption of the proposed Green’s function modeling approach. That is the corrective soundfield in reverberant rooms can be interpolated with sufficient accuracy using only a relatively small number of plane waves. Although soundfields can be generally characterized using the harmonic-based decomposition [51], the higher-order cylindrical/spherical harmonics cannot be created spontaneously. In contrast, plane waves appear readily in various environments in real world, which facilitate a more sparse soundfield representation. The results confirm that the number of the sparsely-selected plane waves is significantly less than both the number of degree of freedom using the cylindrical/spherical harmonics-based systems and the number of image sources included in the simulations.

### 4.7.3 The Cramér-Rao Bound

Experiments for desired ATF estimation were implemented at the frequency of 1 kHz. As we discussed in Sec. 4.4.2, awareness of the identity of
Figure 4.5: The CRB on the variance of $\hat{R}_q(x, k)$ to the desired ATF over the selected zones. We examined the white-noise setting with the variance $\sigma^2 = 5 \times 10^{-4}$ with 15 (left) and 16 (right) measurements. The white crosses represents the position of the microphones.

In Fig. 4.5, we plotted the derived CRB (Eq. (4.17)) on the variance of the estimator $\hat{R}_q(x, k)$ to the desired ATFs using an insufficient number of microphones located at selected positions. The results were averaged for all 39 loudspeakers. We examined the noisy setting with the variance $\sigma^2 = 5 \times 10^{-4}$.

We can clearly observe that the approximate CRB on the estimation error increases with the distance of the observation points to the location of the measurements. The bound approaches to zero at the position of the measurements and its neighborhood (within a small range). However,
the value of the CRB increases with the distance from the measurements, especially at the edge of each selected zones as there are no microphones outside the zones. This is due to the fact that the CRB is computed for a limited number of measurements, which determines only $K'$ entries out of the total degrees of freedom. The CRB provides a bound on the optimal performance of the MSE over the two selected zones for the case of 15 (left) and 16 (right) measurements with -22.4 dB and -27 dB, respectively. When we compared the estimator $\hat{R}_q(x,k)$ derived from the general Green’s function modeling approach (Sec. 4.4.2) with the CRB, it obtained very close estimation performance. The MSE over the two selected zones using 15 and 16 microphones (with the same spatial arrangement) are -19.2 dB and -25.5 dB on average, respectively. We can observe a trend of convergence to the derived lower bound as we increase the number of measurements from 15 to 16.

In this thesis, we do not attempt to find such an optimum geometry. Standard acquisition of signals relies on a regular and uniform sampling of space with respect to Shannon-Nyquist theory. At a given frequency, the space sampling has to be dense enough to avoid spatial aliasing in reconstruction and interpolation [138], i.e., uniform sampling steps $\sigma_v$ are defined to satisfy the sampling theorem to avoid spatial aliasing: $\sigma_v < \frac{n}{k_{max}} \forall \in x, y$ ($k_{max}$ represents the maximum wavenumber). For relatively low frequencies (e.g. 1 kHz), this sampling theorem can be easily satisfied in real-world. However, such an uniform sampling often requires an extremely high number of microphones when frequency goes higher. The reduction of the number of employed microphones is important in practice. One of the advantages of the proposed system over the prior art is that it allows flexible microphone arrangements without rigid restrictions on the spatial sampling steps and arrangements. That is, the microphone measurements can be randomly placed over the selected region and the system is shown to optimizes to achieve a superior and more stable estimation performance than the comparative methods using
the same microphone provision, as shown in the results in Sec. 4.7.2. To the author’s best knowledge, this part of the work is the first reported study on the investigation of the CRB on the variance of room ATF sparse estimation given the arrangements of the microphone array.

![Figure 4.6](image)

Figure 4.6: Reproduction of the desired multizone sound using 16 noiseless measurements for each selected zone in a reverberant room. (a) and (b) demonstrate the case for $\gamma = 0$ and $\gamma = 0.02$ respectively. The red crosses represent the positions of the microphones.

### 4.7.4 Desired 2-D Multizone Soundfield Reproduction

Fig. 4.6(a) demonstrates the reproduction of the desired multizone sound using 16 noiseless measurements for each selected zone at 1 kHz. The desired soundfield over $\mathbb{D}_b$ is a plane wave arriving from $60^\circ$. This is a challenging multizone sound reproduction scenario due to the occlusions of sound between $\mathbb{D}_b$ and $\mathbb{D}_q$. The values of weighting function $w(x)$ assigned to $\mathbb{D}_b$, $\mathbb{D}_q$ and $\mathbb{D}_u$ are $w_a = 1$, $w_b = 10$ and $w_c = 0.01$ respectively. The average estimation error of the ATFs between all loudspeakers over the two selected zones is -35.2 dB.
We can see that the reproduced soundfield matches the desired multizone sound well. The acoustic contrast between $D_b$ and $D_q$ is 24.5 dB and the MSE over $D_b$ is -14.7 dB. We can observe high-amplitude sound outside $D$ due to a high level of loudspeaker array effort. This can be alleviated by increasing the value of $\tau$ in (4.25) at the expense of decreasing the overall reproduction accuracy [39] (as shown in Fig. 4.6(b)). In Fig. 4.6(b), the acoustic contrast between $D_b$ and $D_q$ has dropped to 21.5 dB and the MSE over $D_b$ is -12.9 dB with a regularization parameter $\tau = 0.01$. In Fig. 4.7, the multizone soundfield reproduction results are plotted for various values of the regularization parameter $\tau$. From Fig. 4.7, we can clearly observe the trade-off between the accuracy of the desired multizone sound rendering (which is represented by the acoustic contrast between $D_b$ and $D_q$ in Fig. 4.7(a) and the MSE over $D_b$ in Fig. 4.7(b)) and the limit on the total loudspeaker weight energy (Fig. 4.7(c)).

In Fig. 4.8, the proposed approach is compared with the reproduction method in [3] and [4] in terms of the wide-band multizone soundfield reproduction from 100 Hz to 2 kHz using noiseless measurements. An array of 75 loudspeakers were employed to satisfy the truncation length [62] for $D$ at a maximum frequency of 2 kHz. The same system model and multizone weighting assignments were applied. Overall, we employed 32 microphone measurements for each of the selected zone for the proposed method and the multi-point method [4], while 64 measurements in total were made along the boundaries of the smallest circle that encloses the two spatial zones of interest for the method proposed in [3]. Note that the regularization term to constrain the loudspeaker effort was not included among these simulations to ensure the optimal reproduction performance for all the three approaches.

As we can see in Fig. 4.8, under the circumstance that the same number of loudspeakers and pressure samples are provided for all frequencies, the proposed method significantly outperforms the method proposed in [3] in both the acoustic contrast between the two zones (Fig. 4.8(a))
and the reproduction accuracy over the bright zone (Fig. 4.8(b)). This is not surprising as the superior estimation performance of the desired loudspeaker ATFs in the reverberant room generally leads to better reproduction performance and it is especially critical for the case of multizone soundfield reproduction. We can observe obvious peaks in the red curves for the method proposed in [3] due to the so-called large error scaling [2] at certain frequencies. In contrast, the performance for the proposed method maintains a stable level throughout the selected frequency range. We can also see that the multiple-point method [4] (black
curve) only achieves decent reproduction accuracy at low frequencies and the performance degrades drastically as the frequency increases. This is due to the fact that the equalization of the desired soundfield is limited to a discrete set of measurement points and the performance deteriorates away from the control points, whereas the proposed approach prevails by performing the least squares design over the whole reproduction region.

4.7.5 On the Multizone Soundfield Reproduction with Practical Settings

The numerical results listed in Sec. 4.7.2 suggest that at least 14 measurements are needed to accurately characterize the desired ATF with high probability within a circular region of radius 0.3 m at the frequency of 1 kHz. Note that the previous simulations were implemented in a four-wall reverberant room with relatively reflective wall settings. However, the reverberant environments we may encounter in practice may be simpler than this, especially under the circumstance that the desired region and the loudspeakers are placed close to one wall while the reverberant room

\(^3\)Note that we define the estimation with MSE the over -20 dB as an accurate one.
4.7. RESULTS AND DISCUSSION

Figure 4.9: The estimated performance of the desired loudspeaker ATF is plotted as a function of the number of microphones used at 1 kHz (left) and 5 kHz (right) with single-wall setting.

Figure 4.10: The required number of measurements for 100% accurate estimation of the desired loudspeaker ATF with the single-wall setting.

is large (which is a common scenario in reality).

In Fig. 4.9, we demonstrate the performance of the desired ATF estimation with the same settings except that we set up a space with single-wall setting (the wall absorption coefficient is 0.3) at the frequency of 1 kHz and 5 kHz. Both the noiseless case and the noisy measurement case (with 30 dB SNR) are illustrated. We ran 100 trials for each iteration with an
assigned number of microphone measurements. From Fig. 4.9, we can see that only 6 randomly-selected measurements within the selected zone would generally lead to an accurate estimate of the desired loudspeaker ATF at 1 kHz. The required number of measurements for completely accurate estimation at 5 kHz is 17. In Fig. 4.10, the numbers of required measurements for 100% accurate estimation of the desired loudspeaker ATF from 1 kHz to 5 kHz are plotted with the single-wall setting. We can see that the required number of microphones to guarantee an accurate estimation in the proposed formulation does not rise linearly with the increase of the frequency. It makes our approach more advantageous than the prior methods that the required number of microphones increases linearly with the working frequency [2] [3]. Overall, it is reasonable to expect that significantly fewer measurements are required for the case with a simple environment in practice than for the case when it features complex environments. The reduction of the required number of measurements under the simple
reverberant assumption also facilitates the multizone (more than two zones of interests) soundfield reproduction using a limited number of microphones. In Fig. 4.11, we illustrate the reproduction of the desired three-zone soundfield using only 7 noiseless measurements for each selected zone at 1 kHz. The desired soundfield over Zone 1 is a plane wave arriving at $0^\circ$ while for Zone 2 is $-150^\circ$. Zone 3 is set to be a quiet region. This is a challenging multizone sound reproduction scenario due to the strong occlusions of sound in each of the individual zones. The values of the weighting function $W(x)$ assigned to Zone 1, 2, 3 and the unattended zone are 1, 1, 5 and 0.01, respectively. The average estimation error of the ATFs between all loudspeakers over the three selected zones is -36 dB. We can see that the reproduced soundfield matches the desired multizone sound well. The acoustic contrast between the defined bright zones (Zone 1 and 2) and Zone 3 is 20.8 dB and the average MSE over Zone 1 and 2 is -11.3 dB.

![Figure 4.12: 3-D Reproduction of the desired multizone sound using 30 microphones for each selected zone in a reverberant room at 500 Hz. (a) and (b) demonstrate the real part and imaginary part of the soundfield respectively. The results show the pressure field at the cut slice of $z = 1.2$ m in Cartesian coordinates.](image)
4.7.6 Desired 3-D multizone soundfield reproduction

Now we turn to the simulations of 3-D multizone soundfield reproduction. The reverberant room was cubical (size 6 m × 5 m × 4 m). The desired reproduction region \( \mathbb{D} \) was a full sphere of radius \( R_r = 1 \) m with its center located at \((2, 2.5, 1.2)\) m. The simulations were implemented at 500 Hz. Following the rule of thumb proposed in [113], we have \( Q = (\left\lceil k e R_r / 2 \right\rceil + 1)^2 = 196 \) secondary sources (i.e. loudspeakers) on the sphere and all loudspeakers were equi-distant from the origin (we set the radius of the loudspeaker array 1.5 m). The location of the loudspeakers was given by the library of 3-D packings (pack.196) in [115]. We used the image source method [72] to simulate the soundfield created by the loudspeaker in the reverberant room. In the simulations below, a total of 360 image sources were included for each of the employed loudspeakers.

Fig. 4.12 demonstrates the reproduction of the desired multizone sound using 30 noiseless measurements for each selected spherical zone of radius 0.3 m. The desired soundfield over \( \mathbb{D}_b \) is a monochromatic plane wave arriving from \([15^\circ, 75^\circ]\). This is again a challenging multizone sound reproduction scenario due to the occlusions of sound between \( \mathbb{D}_b \) and \( \mathbb{D}_q \). The average estimation error of the ATFs between all loudspeakers over the two selected zones was -33.8 dB. We can see that the reproduced soundfield matches the desired multizone soundfield well. The acoustic contrast between \( \mathbb{D}_b \) and \( \mathbb{D}_q \) is 24.7 dB and the MSE over \( \mathbb{D}_b \) is -15.1 dB. The system performance is consistent with the 2-D multizone soundfield reproduction listed in the previous section.

Based on the discussions listed above, we note that an accurate soundfield reproduction with 3-D settings is still considered as a difficult problem unless the loudspeakers are placed on a sphere that encloses the 3-D spatial region of interest. However, a spherical loudspeaker array is not commonly available in practice, which makes the 3-D multizone soundfield reproduction very cumbersome to implement. Being motivated by these issues, we aim to obtain the best performance for multizone
4.8 Conclusion and Contribution

We proposed a general Green’s function modeling approach for determining the ATFs between loudspeakers over the desired regions in reverberant environments using a sparse approximation method. The proposed approach facilitates practical implementations in the following aspects: i) a significant reduction over existing methods on the number of required measurements for accurate characterization of the loudspeaker ATFs, ii) consistent performance of the multizone soundfield reproduction over a wide frequency range, iii) the flexibility of the spatial microphone arrangement.

The CRB on the variance of the desired ATF estimator using an insufficient number of microphones was also derived, which shows that the estimation performance deteriorates as the distance of the observation points to the location of the selected measurements increases. The optimal loudspeaker filter solution for the desired multizone sound reproduction can be derived based on the estimate of loudspeaker ATFs using an orthogonal basis function approach. The results verify that the proposed multizone soundfield reproduction method consistently outperforms the existing works with the same hardware settings (i.e. employed loudspeakers and microphones) in reverberant rooms. We also extend our work to the reproduction of the desired 3-D multizone soundfield.

The major contributions made in this chapter are:

- The interpolation of the acoustic transfer function from the loudspeakers over the desired reproduction region using the concept of sparse approximation, which facilitates a significant reduction of the required number of microphones, flexible microphone placements
within the desired region and a stable estimation performance over a wide frequency range.

- The derivation of the Cramér-Rao lower bound on the variance of the desired soundfield estimator with an insufficient number of microphones.

- The 2-D multizone soundfield reproduction in reverberant rooms using a limited number of microphones based on the orthogonal basis function approach.

- The investigation of the multizone soundfield reproduction with practical settings, which shows that a simple reverberant environment would further reduce the required number of measurements.

- The extensions to 3-D multizone soundfield reproduction using a spherical loudspeaker array in reverberant environments.

4.9 Appendix: Algorithms for the Reconstruction of the Sparse signals

Our objective is to find the optimal solution \( \hat{y} \) that identifies the corrective component of the soundfield generated by each loudspeaker in reverberant environments. We first present an algorithm by finding the sparsest representation of the signal \( y \). According to [123], one can reconstruct \( y \) exactly, with very high probability, as the solution of the following optimization problem:

\[
\min_y \|y\|_0, \text{ s.t. } v = \Phi y. \tag{4.29}
\]

Importantly, this result continues to hold if we replace the \( l^0 \) norm with the \( l^1 \), which leads to a convex problem and it is referred as the so-called
4.9. APPENDIX: ALGORITHMS FOR THE RECONSTRUCTION OF THE SPARSE SIGNALS

Basis Pursuit:

\[
\min_{y} \|y\|_1, \text{ s.t. } v = \Phi y. \tag{4.30}
\]

The problem (4.29) is NP-hard [139], whereas the convex problem (4.30) can be solved efficiently.

A family of iterative greedy algorithms – Matching Pursuit [140–142] has been shown to generally feature less computational complexity with a similar exact reconstruction property. However, these algorithms require more measurements for exact reconstruction than the basis pursuit method [143].

Alternatively, it has been shown in [127] that a nonconvex variant of basis pursuit will produce an exact reconstruction with fewer required measurements, which is more relevant to our needs for this work. Chartrand [127] also demonstrated that this nonconvex variant of basis pursuit is less computationally intractable than the optimization problem in (4.29). More specifically, the $l^1$ could be replaced with the $l^p$ norm, where $0 < p < 1$.

Recall that for $0 < p < +\infty$,

\[
\|y\|_p = \sum_{i=1}^{N} |y_i|^p. \tag{4.31}
\]

Note that $\|\cdot\|_p$ is not an actual norm when $0 < p < 1$, but $\|\cdot\|_p$ satisfies the triangle inequality and induces a metric [144]. Therefore, we have the following nonconvex problem:

\[
\min_{y} \|y\|_p^p, \text{ s.t. } v = \Phi y. \tag{4.32}
\]

The observation that fewer measurements were required for exact reconstruction than when $p = 1$ was demonstrated by numerical experiments in [127] with the same sensing matrix.

Early works in [145] [146] proposed the iteratively reweighted least
squares (IRLS) approaches. However, the problem of (4.32) using IRLS has been mostly considered for the case of $p \geq 1$. The case of $0 < p \leq 1$ was studied by Rao et al. [144]. The approach was to transform the $l^p$ objective function in (4.32) to a weighted $l^2$ norm,

$$\min_y \sum_{i=1}^{N} w_i y_i^2, \text{ s.t. } v = \Phi y,$$  \hspace{1cm} (4.33)

where the weights are derived from the former iterate $y^{n-1}$, so that the objective in (4.33) is a first-order approximation to the $l^p$ objective: $w_i = |y_i^{n-1}|^{p-2}$.

In any practical situation, the measured signal is corrupted by noise, so a more practical model for the measurement process is [129],

$$v = \Phi y + \eta,$$  \hspace{1cm} (4.34)

where $\eta$ features the additive white noise. Now the task is to reconstruct $y$ from noisy under-sampled measurements $v$. In such a case, the equality condition $v = \Phi y$ should be relaxed to [129]

$$\min_y \|y\|^p_p, \text{ s.t. } \|v - \Phi y\|^2_2 \leq \epsilon.$$  \hspace{1cm} (4.35)

The first term $\|y\|^p_p$ is the modeling term and the second term $\|v - \Phi y\|^2_2$ represents an allowance for an error.

We assume that the additive noise is Gaussian noise of zero mean and variance $\sigma^2(k)$, which makes the $l^2$-norm for the second term appropriate [129]. (5.13) can be converted to the following unconstrained form:

$$\min_y \frac{1}{p} \|y\|^p_p + \frac{\lambda}{2} \|v - \Phi y\|^2_2,$$  \hspace{1cm} (4.36)

where $\lambda$ is related to the error allowance $\epsilon$. Experimental results suggest that choosing $\lambda$ in the range $[\sigma^2(k)\sqrt{\log(N)}, \sigma^2(k)\sqrt{2\log(N)}]$ yields a fair reconstruction [129].
In the IRLS method the modeling term \( (l^p\text{-norm}) \) is approximated by a weighted \( l^2\text{-norm} \): \( \frac{1}{p} \| y \|_p^p \approx \frac{1}{2} \| W_m y \|_2^2 \). The weighting matrix is updated at each iteration and it is given by

\[
W_m(n) = \text{diag}(2^{p} |y(n-1) + \delta(n)|^{p/2-1}). \tag{4.37}
\]

The perturbation \( \delta \) interlaces with the elements of the weighting matrix. At each iteration, the perturbation \( \delta \) is reduced so that when the solution converges, the weighted \( l^2\text{-norm} \) is a good approximation of the desired \( l^p\text{-norm} \). The idea of perturbing the IRLS was proposed in [128]. It showed that the perturbed method was significantly better than the unperturbed ones such as [144].

With the approximations made above, (4.36) can be represented (approximately) in the following form [129]:

\[
\min_y \frac{1}{2} \| W_m y \|_2^2 + \frac{\lambda}{2} \| v - \Phi y \|_2^2. \tag{4.38}
\]

The closed-form solution to (4.38) is

\[
\hat{y} = (\lambda W_m^H W_m + \Phi^H \Phi)^{-1} \Phi^H v. \tag{4.39}
\]

(4.39) is applied iteratively by updating the weight matrices till the solution converges (the change in relative \( l^2\text{-norm} \) from the previous iterate approaches to 0) or \( \delta \) is less than the minimum value. In our work, we define \( \delta_{\text{min}} = 10^{-6} \).

In terms of the computational consumption, the IRLS method generally calculates the pseudo-inverse of the dictionary \( \Phi \) for multiple times until the residual error reaches the threshold. The algorithm works reasonably fast. On the laptop with Intel\textsuperscript{TM} Core\textsuperscript{TM} i5 four-core and 4GB RAM, it takes approximately 7 seconds to derives the final estimate of the sparse coefficient set in Matlab, given the size of the dictionary is \( M \times 1000 \) (where \( M \) is the number of microphone measurements). In contrast,
the orthogonal matching pursuit (i.e. a greedy-fashion approach and was used in [120] [121]), takes approximately 12 seconds with the same settings.
4.9. APPENDIX: ALGORITHMS FOR THE RECONSTRUCTION OF THE SPARSE SIGNALS
Overview: In this chapter, we aim to improve the practical implementation of the multizone soundfield reproduction system by incorporating concepts of 2.5D reproduction and active echo cancellation. We first extend the existing multizone soundfield reproduction method in Chapter 4 to 2.5D reproduction in a 3-D reverberant environment (i.e. the reverberation from the floor and ceiling is also taken into account). We also introduce an adaptive reverberation cancellation method for multizone soundfield reproduction using sparse methods. The reproduced soundfield is described as a weighted series of orthogonal basis functions over the desired reproduction region, which is then used to adaptively equalize the desired multizone soundfield in terms of the basis function coefficients. The sparse methods result in a significantly reduced number of required microphones for the measuring process of the reproduced soundfield. Simulation results verify the efficient room reverberation compensation for desired multizone soundfield reproduction. The proposed method facilitates reproduction over a wide frequency
5.1 Introduction

Reproduction of a desired multizone soundfield over a region of interest has drawn the attention of researchers in recent years due to its various potential applications. It aims to provide an individual sound environment to each of a set of listeners without the use of headphones. The majority of the existing works in spatial soundfield reproduction mainly concentrate on a single zone [18]. Common single-zone sound reproduction techniques include wave field synthesis (WFS) approach [25–27, 30], higher order ambisonics (HOA) [31, 32, 60] and the systems based on spherical harmonics [2, 48, 67, 113]. The realization of multizone soundfield reproduction is a conceptually challenging problem and the main difficulties lie in two aspects: requirement on the loudspeaker array arrangement and the reverberation effects from the ambient environment. Spatial soundfield reproduction generally requires a number of loudspeakers to be placed at discrete locations outside or on the boundary of the desired reproduction region. The required number of loudspeakers is proportional to the size of the control region and the working frequency. Thus, the number of the employed loudspeakers and the placement of the loudspeaker array have a significant impact on the performance of the desired soundfield reproduction, especially for the case of 3-D soundfield reproduction [62] [113]. Comparing with the 2-D case that focuses on a height-invariant soundfield reproduction, 3-D spatial soundfield reproduction enables an enhanced immersive acoustic experience for the users. In addition, the 2-D soundfield reproduction is designed by approximating secondary sources (i.e. loudspeakers) as line sources that exhibit 2-D acoustic transfer functions. However, a 3-D point source is naturally a more accurate model to represent characteristics.
of the loudspeaker employed in practice. Relatively little work relating to the implementation of 3-D soundfield rendering systems has been conducted. This is particularly true for WFS-based work due to the fact that WFS systems are generally applied for large areas and a highly dense distribution of secondary sources surrounding the selected area is required [147] [148].

Even though there are a few systems developed for 3-D sound reproduction based on HOA/spherical harmonics [34, 48, 148–150], a critical problem in terms of loudspeaker placement still exists. To accurately reproduce an incident soundfield using spherical harmonics, the current 3-D soundfield reproduction systems generally require the employed loudspeakers to be placed equidistantly on a sphere that encloses the desired reproduction region [34]. In reality, this spherical loudspeaker array placement is not practical to achieve. The possibility of non-spherical loudspeaker array deployment was investigated in [151] based on the least-squares approach to match the desired soundfield in terms of the spherical-harmonic decomposition. Gupta et al. proposed a method of 3-D sound field reproduction using multiple circular loudspeaker arrays at different colatitudes [152]. The authors provided a technique for spherical harmonic mode-selection to control the reproduced soundfield. However, the number of loudspeakers increases quadratically with the working frequency, which hinders such systems from wide adoption.

The so-called 2.5D soundfield reproduction, which aims to reproduce the desired soundfield over a horizontal region at the height approximately level with the listener’s ears using a single array of loudspeakers (modeled as 3-D point sources), has been introduced as an alternative soundfield reproduction approach in the 3-D coordinates system. The WFS techniques have been applied to 2.5D soundfield reproduction using a discrete linear distribution of point sources based on the spectral division method (SDM) in the wavenumber domain [56] [153]. Analytical derivations for

---

1For simplicity, the loudspeakers are assumed to be omnidirectional.
5.1. INTRODUCTION

the 2.5D HOA soundfield systems were also investigated in [60, 113, 154] using a circular loudspeaker arrangement. However, the main drawback of the existing 2.5D SDM and HOA systems is that the reproduction of the desired soundfield can only be rendered precisely for receiver positions on a straight line parallel to the linear loudspeaker array and a reference point within the selected region, respectively. The analysis in [56] [154] suggests that a noticeable reproduction accuracy decay can be observed when it moves away from the reference positions.

Importantly, the fore-mentioned approaches do not take into account the reverberant environments that practical multizone sound reproduction systems will encounter. The ambient reverberation leads to inferior performance compared with ideal reproduction of the desired soundfield using the free-field assumption. The reverberation compensation process is difficult to handle because the reverberant room channel is unknown and because a large number of loudspeakers and microphones is required by existing soundfield reproduction systems.

Reverberation cancellation techniques are indispensable for a reproduction system with real-world settings. To equalize the reverberation effects, a common way is inverting the channel response prior to the derivation of the loudspeaker driving signals. The traditional approach for spatial sound reproduction in a reverberant setting is pressure matching, which equalizes the transfer functions over a discrete set of points [77] [78]. This technique leads to poor robustness in regions further away from the design points and inaccurate reproduction [2]. A technique based on mode matching to reproduce a single-zone soundfield accurately over a control region in reverberant rooms was proposed in [2]. The method determines the ATF between each loudspeaker and the circular control region by evenly sampling sound pressure along the boundary. An approach of reproducing a multizone soundfield within a desired region using sparse methods was introduced in [6]. A limited number of randomly placed measurements were employed to estimate the transfer
functions between the loudspeakers over the desired region in reverberant environments based on a sparse approximation. The estimates were then used to derive the optimal least-squares solution for the loudspeaker filter gains. Although the approach was shown to significantly reduce the number of required microphones for an accurate multizone soundfield reproduction, a prior measurement of the room transfer function for all the employed loudspeakers was still needed. This is time-consuming to implement in practice and it can be further exacerbated by changes in the ambient environment conditions during the measurement process.

Wave domain adaptive filtering (WDAF) is a more practical approach to the application of reverberation cancellation in soundfield reproduction. Initially proposed by Buchner et al. [81] [82], it has been introduced to active listening room compensation in wave field synthesis systems [83–85]. Comparing with the traditional time and frequency domain adaptive filtering methods [86] [87], it features better convergence behavior by decoupling the high correlation between multiple reproduction channels. The wave-domain representation of the soundfield was described using two transformations on the microphone array input and the loudspeaker output, respectively [83]. The work by Schneider et al. [84] [88] has further reduced the computational complexity of the basic WDAF adaptation process. It was achieved by considering that the dominant couplings between the soundfield modes are strong only in the vicinity of the diagonal line of the transformations and neglecting the weaker ones. A similar adaptive method was proposed in [5], in which the reverberant soundfield was described and estimated by exploiting the orthogonality of the Fourier-Bessel expansion to simplify the listening room compensation problem within a region of interest. The work of [5] [88] provides some insights into the underlying structure of the reverberant soundfield in the mode-domain.

In this chapter, we aim to improve the practical implementation of the multizone soundfield reproduction system by incorporating the concept of
2.5D reproduction and active echo cancellation. We first apply the existing 2-D multizone soundfield reproduction method discussed in Chapter 4 to 2.5D reproduction system in a 3-D reverberant environment. To fully fulfill the 3-D reverberant settings, the reverberation from the floor and ceiling is also taken into account. The proposed approach expresses the soundfield as an orthonormal basis function expansion and matches the desired multizone soundfield and the actual reproduced soundfield in the basis-function domain over the desired reproduction region. Additionally, we use the inspiration from [5] [88] to propose an adaptive reverberation cancellation system for the desired multizone soundfield reproduction using sparse microphone measurements. We consider the reproduced soundfield as a linear transformation of the desired soundfield, where the coefficient set for the desired soundfield can be computed under the free-field assumption. We then introduce the adaptive channel estimation process using sparse methods to identify these transformations and derive the required loudspeaker updating signals.

Figure 5.1: Overview of the 2.5D active reverberation cancellation system including the loudspeaker array and microphone configurations.
5.2 Overview

The vertical view of the proposed 2.5D active multizone soundfield reproduction system is shown in Fig. 5.1. It consists of a circular array of $Q$ loudspeakers and $M$ microphones. Note that the circular loudspeaker array can be placed at the same or different plane with different height level. The adaptive reverberation cancellation system aims to rectify the reverberation effects based on iterative feedback from sparse microphone measurements. To define our multizone soundfield requirements, one bright zone $D_b$ and one quiet zone $D_q$ are selected in this work. We define the remaining area in the desired reproduction region $D$ as the unattended zone $D_u$.

The rest of the chapter is organized as follows. In Sec. 5.3, we first introduce the 2.5D reproduction using a circular loudspeaker array outside $D$ in 3-D reverberant environments. In Sec. 5.4, we propose an adaptive reverberation cancellation system for the desired multizone soundfield reproduction using sparse microphones. Finally, performance of the adaptive reverberation cancellation and the 2.5D multizone soundfield reproduction over the selected zones are verified by the simulation results in Sec. 5.5.

5.3 2.5D Multizone Soundfield Reproduction Using Sparse Methods

The 3-D soundfield reproduction is generally very cumbersome to implement due to the fact that a large number of loudspeakers are required to be placed on an enclosed sphere. In this section, we implement the so-called 2.5D multizone soundfield reproduction using a single circular array of loudspeakers. The reproduction design takes the microphone measurements of the loudspeaker ATFs over the selected zones the into
consideration. The model of an 2.5D multizone soundfield reproduction system in 3-D reverberant environments is shown in Fig. 5.2. Note that the system allows flexibility in terms of the height level of the loudspeaker array and the microphones are located in the same plane of the desired reproduction region.

![Loudspeaker Array](image)

Figure 5.2: The model of an 2.5D multizone soundfield reproduction system in reverberant environments.

We start with the estimation of the acoustic transfer function (ATF) from the loudspeakers over the horizontal reproduction region in 3-D reverberant environments using a sparse approximation. It has been shown effective in Chapter 4 that the sparse methods facilitate a significantly reduced number of required measurements for accurate soundfield interpolation, as well as a stable performance over a wide frequency range, compared with the classical the cylindrical/spherical harmonics-based systems.

To estimate the ATF between the $q$’th loudspeaker $T_q(x, k)$, we first separate the actual soundfield into a basic component, the 3-D free-field Green’s function and an unknown corrective soundfield $R_q(x, k)$.
where $h_{q'}$ represents the position of the $q'$th loudspeaker.

The basic principle of our method is to assume that $R_{q'}(x, k)$ is sparse in the domain of plane wave decomposition, i.e. the soundfield results from only a relatively small number of plane wave functions. Based on this assumption, we consider the following $p$ norm (where $0 < p < 1$) non-convex optimization problem

$$\min_y \|y\|^p_p, \text{ s.t. } v = \Phi y, \quad (5.2)$$

where $\|y\|^p_p = \sum_{i=1}^{N'} |y_i|^p$. The dictionary $\Phi$ is an $M \times N'$ sensing matrix ($N' \gg M$) whose columns are the normalized measurement vectors of an overcomplete set of plane wave functions $\{P_n\}_{n \in A}$ (where $A$ is a set of indice) arriving from various azimuth and elevation angles. $y$ is a sparse signal (i.e. $y$ has a limited number of non-zero entries at unknown locations) that represents the coefficients for $\{P_n\}_{n \in A}$. $v$ is an $M \times 1$ observation vector which contains the values of the corrective soundfield at $M$ randomly chosen locations within the desired region.

From (5.1), we see that the values of $v$ are the difference between the original measurements of the soundfield and the value of the free-field Green’s function.

We can apply the regularized Iteratively Reweighted Least Squares (IRLS) algorithm [127] [129] to solve (5.2) and derive the optimal solution $\hat{y}$ that estimate the reproduced soundfield in reverberant environments:

$$T_{q'}(x, k) \approx \frac{e^{-ik\|h_{q'}-x\|}}{4\pi\|h_{q'}-x\|} + \sum_{n=1}^{N'} \hat{y}_n P_n(x, k), \quad (5.3)$$

where $\hat{y}$ has only $m'$ ($m' \leq M$) non-zero components.

The estimate of the ATFs from loudspeakers over the desired reproduction region is then used to derive the optimal loudspeaker filter gains for the
desired multizone sound reproduction. We express the desired multizone soundfield \( S^d(x, k) \), the actual reproduced soundfield \( S^a(x, k) \) and the ATF for a given loudspeaker \( T_q'(x, k) \) as a weighted series of basis orthonormal functions \( \{G_i\} \). The orthonormal set of basis functions \( \{G_i\} \) over the desired reproduction region \( \mathbb{D} \) is constructed using the modified Gram-Schmidt process on \( N = \lceil \text{ker}/2 \rceil + 1 \) plane wave functions [44]. The directions of the selected plane wave functions are given by the library of 3-D packings for the specific value of \( N \) in [115]. Provided all sound sources lie outside \( \mathbb{D} \), we can use the following representations:

\[
S^d(x, k) = \sum_{n=1}^{N} A_n G_n(x, k), \tag{5.4}
\]

\[
S^a(x, k) = \sum_{n=1}^{N} B_n G_n(x, k), \tag{5.5}
\]

\[
T_q'(x, k) = \sum_{n=1}^{N} C_n^q G_n(x, k), \tag{5.6}
\]

The coefficients for \( S^d(x, k), T_q'(x, k) \) can be derived as \( A_n = \langle S^d(x, k), G_n(x, k) \rangle_w \), \( C_n^q = \langle T_q'(x, k), G_n(x, k) \rangle_w \) respectively. The weighted inner product \( \langle \cdot \rangle_w \) is defined as

\[
\langle Y_1, Y_2 \rangle_w = \int_{\mathbb{D}} Y_1(x) Y_2^*(x) w(x) dx, \tag{5.7}
\]

where the weighting function \( w(x) \) specifies the relative importance of the reproduction accuracy for each point in space [44].

The reproduced soundfield \( S(x, k) \) resulting from the \( Q \) loudspeakers is equal to
\( S^a(x, k) = \sum_{q'}^{Q} l_q'(k) T_{q'}(x, k). \)  

(5.8)

Substituting (5.5) and (5.6) into (5.8), the coefficients of the reproduced soundfield are related to \( C_{q'}^n \) through \( B_n = \sum_{q'=1}^{Q} l_q'(k) C_{q'}^n. \) It has been shown in [113] that the number of loudspeakers required for good approximation of a 2.5D soundfield reproduction is \( Q \geq 2 \lceil \text{ker}/2 \rceil + 1. \)

We can see that the required number of loudspeakers increases linearly with the working frequency and the size of \( \mathbb{D} \) for the 2.5D soundfield reproduction.

The optimal least-square solution for the loudspeaker gains that minimizes the sum squared error between the desired multizone soundfield and the reproduced soundfield coefficients is given by [6]

\[ l = (C^H C + \tau I)^{-1} C^H a, \]  

(5.9)

where \( I \) is an identity matrix and \( \tau \) is a regularization term on the loudspeaker array effort. \( C \) is the \( N \times Q \) matrix of the coefficients of the room responses of all loudspeakers

\[ C = \begin{bmatrix} C_1^1 & \cdots & C_1^Q \\ \vdots & \ddots & \vdots \\ C_N^1 & \cdots & C_N^Q \end{bmatrix} . \]  

(5.10)

5.4 Adaptive Reverberation Cancellation For Multizone Soundfield Reproduction

In this section, we introduce an adaptive reverberation cancellation method for multizone soundfield reproduction using sparse methods. The model

\[ \text{For this part of work, please see the appendix of this chapter for more details.} \]
of the soundfield rendering system is shown in Fig. 5.1. The loudspeakers are placed outside the desired reproduction region while the microphones are randomly placed within the selected zones of interest to iteratively record the generated soundfield at the assigned position.

5.4.1 Soundfield Basis Expansion

The received measurements at the microphones can be expressed in matrix form as

\[ \mathbf{v}(k) = \mathbf{H}(k)\mathbf{l}(k), \]  

(5.11)

where \( \mathbf{l}(k) = [l_1(k), \ldots, l_Q(k)]^T \) are the loudspeaker driving signals, \( \mathbf{v}(k) = [v_1(k), \ldots, v_M(k)]^T \) are the received signals at the microphone measurements, and \( \mathbf{H}(k) \) represents the channel between the \((m,q)\)th microphone-loudspeaker pair at the wavenumber \( k \). Note that we can separate the channel effects \( \mathbf{H}(k) \) into the direct and reverberant paths and we have \( \mathbf{H}(k) = \mathbf{H}^d(k) + \mathbf{H}^r(k) \), where \( \mathbf{H}^d(k) \) and \( \mathbf{H}^r(k) \) represent the direct and reverberant channels between the \((m,q)\)th microphone - loudspeaker pair, respectively.

Applying our orthogonal basis function approach in Chapter 3, the measurements in (5.11) can be expressed as

\[ v_m(k) = \sum_{n=1}^{N} B_n(k) G_n(x_m, k), \]  

(5.12)

where \( N \) is the truncation length for desired reproduction region \( \mathbb{D} \) [62] and \( B_n(k) \) are the coefficients for the reproduced soundfield. \( x_m \) represents the \( m \)th microphone location. The coefficients \( B_n(k) \) can be derived from the soundfield measurements \( \mathbf{v}(k) \), as described in the next section.

5.4.2 Soundfield Characterization Using Sparse Methods

In this section, we apply the estimation method in [6] to calculate \( B_n(k) \) from the randomly-placed measurements \( v_m(k) \) within the selected zones
of interest.

The basic principle of our method is to assume that the reproduced soundfield \( S(x, k) \) is sparse in the domain of plane wave decomposition. In this part of work, we also take the noisy measurement into account and propose the following \( l^p \) norm (where \( 0 < p < 1 \)) non-convex optimization problem

\[
\min_{\hat{y}} \| y \|_p^p, \quad \text{s.t.} \quad \| v - \Phi y \|_2^2 \leq \epsilon,
\]

where the error allowance \( \epsilon \) is related to the noise level and we assume that the additive noise is complex Gaussian noise of zero mean and variance \( \sigma^2(k) \). Importantly, the columns of dictionary \( \Phi \) are built up to contain the values of \( \{ G_n(x, k) \}_{n \in A} \) at \( M \) locations. In this case the columns themselves are each determined by a small number of basis vectors. By applying the IRLS approach, we can derive an estimator \( \hat{y} \) based on the noisy measurements \( v \), which directly represents the basis function coefficients \( B_n(k) \) for the reproduced soundfield. This is a more efficient approach of characterizing the reproduced soundfield using the feedback from the microphone measurements as we skip the integral calculation \( \langle \cdot \rangle_w \). It is particularly well suited for the adaptive system in this chapter, in which the estimation of the reproduced soundfield needs to be conducted iteratively while the formulation of the orthogonal basis functions can be implemented offline.

Overall, we formulate the calculation of the soundfield coefficients \( B_n(k) \) based on the soundfield measurements in the following matrix form

\[
B(k) = T H(k) l(k),
\]

where \( B(k) = [B_1(k), \ldots, B_N(k)] \). \( T (N \times M) \) is a transformation matrix expressing the derivation of \( B(k) \) from \( v(k) \), which can be seen as the projection from the scarce measurements onto the subspace \( V \in \mathbb{C}^N \) spanned by the orthogonal set \( \{ G_n \}_{n \in A} \).
5.4.3 Channel Estimation and Reverberation Cancellation

In this section, we describe a model of the reverberant room channel, the channel estimation processes in the basis-function domain and the computation of the loudspeaker updating signals required for active reverberation cancellation.

5.4.4 Room Channel Modeling and Estimation

We can express the desired multizone soundfield $S^d(x, k)$, the actual reproduced soundfield in a reverberant room $S^a(x, k)$ as a weighted series of the orthogonal basis functions $\{G_n\}_{n \in A}$. The coefficients for $S^d(x, k)$ can be derived as $\langle S^d(x, k), G_n(x, k) \rangle_w$.

We assume that a linear transformation can be formulated between the reproduced soundfield and the desired soundfield. As $\{G_n\}_{n \in N}$ are orthonormal with each other in the defined basis-function domain over $\mathbb{D}$ (i.e. $\langle G_n(x, k), G_m(x, k) \rangle_w = \delta_{n,m}$) [44], the couplings between the coefficients for basis functions having different index can be set to zero. This can be expressed by a linear transformation of the soundfield coefficient set for the basis function, which features a diagonal structure:

$$B(k) = U(k)B^d(k), \quad (5.15)$$

where $U(k) = \text{diag}[U_1(k), \ldots, U_N(k)]$. The room channel transformation $U(k)$ can be estimated in an iterative fashion, and it is also interlaced with the necessary active updating signals $\sigma(k) (Q \times 1)$ on the loudspeaker array.

We define $\tilde{B}(k)$ as the measured soundfield coefficient at the microphones after updating the loudspeaker signals. An accurate estimate of the room channel transformation $\hat{U}(k)$ can be achieved if the value of the residual error $\|\tilde{B}'(k)\|^2$ (where $\tilde{B}'(k) = \tilde{B}(k) - B^d(k)$ is the difference between the measured and desired soundfield coefficients) is minimized [84, 85, 88]. This is a classical adaptive filtering problem and $U(k)$ can be estimated actively by using algorithms such as Least Mean Squares (LMS) filter and
Recursive Least Squares (RLS) filter.

Let $\hat{U}(k)$ be the estimate of $U(k)$ at the $\tau$th adaption step. Due to the diagonal structure of $U(k)$, calculating the diagonal entries $\hat{U}_n(k)$ can be further simplified as a single-tap adaptive filtering problem [155]:

$$\hat{U}_n(k) = \hat{U}_n(k) - \frac{1}{\phi_n^2(k)} B^d_n(k) \tilde{B}_n'(k),$$

(5.16)

where $\phi_n^2(k)$ is the gain factor $\phi_n^2(k) = \lambda \phi_n^2(k-1) + |B^d_n(k)|^2$. $\lambda$ is the forgetting factor. We choose the RLS algorithm due to the fact that $U(k)$ can be seen as a FIR filter on the desired soundfield coefficients which has a relatively small order $N$ in our work. The method of RLS provides a fast convergence rate under this circumstance. Therefore, (5.16) can be applied to obtain an iterative estimate of $U(k)$ based on the residual error $\tilde{B}_n'(k)$ at the $\tau$th adaption step. The updating signal on the loudspeaker array in each iteration is based on the estimator $\hat{U}(k)$. It is designed to minimize the residual error and ensure the estimation convergence, as described in the following section.

### 5.4.5 Loudspeaker Updating Signals

In this section, we derive the optimal loudspeaker updating signals based on the knowledge of the estimated room channel transformation, so that the reverberation cancellation performance is maximized.

We precondition the initial loudspeaker array signals to reproduce the desired multizone soundfield under free-field assumption following the method proposed in Chapter 3. Therefore, the coefficients for the desired soundfield $B^d(k)$ can be expressed by replacing $H(k)$ with the direct channel $H^d(k)$ in (5.14):

$$B^d(k) = TH^d(k)I(k).$$

(5.17)

Incorporating the room channel model in (5.15) and the estimator $\hat{U}(k)$ in Sec. 5.4.4, we have
5.4. ADAPTIVE REVERBERATION CANCELLATION FOR MULTIZONE SOUNDFIELD REPRODUCTION

\[ B(k) = \hat{U}(k)TH^d(k)l(k). \]  

(5.18)

Following (5.18), the measured soundfield coefficients \( \tilde{B}_n(k) \) after adding updating signals \( \sigma(k) \) to the loudspeakers can be given by

\[ \tilde{B}(k) = \hat{U}(k)TH^d(k)[l(k) + \sigma(k)]. \]  

(5.19)

We can write the difference between the measured and desired soundfield coefficients \( \tilde{B}'(k) \) using (5.17) and (5.19):

\[ \tilde{B}'(k) = [\hat{U}(k) - I]TH^d(k)l(k) + \hat{U}(k)TH^d(k)\sigma(k), \]  

(5.20)

where \( I \) is an identity matrix.

Importantly, the problem of matching the actual reproduced soundfield with the desired multizone soundfield over the desired reproduction region in the weighted least-squares sense can be reduced to the minimization of \( \|\tilde{B}'(k)\|^2 \) [6]. Therefore, an efficient reverberation compensation and accurate soundfield reproduction can be achieved by finding the optimal loudspeaker updating signals \( \sigma(k) \). The least squares solution is given by

\[ \sigma(k) = [\hat{U}(k)TH^d(k)]^\dagger[I - \hat{U}(k)]TH^d(k)l(k) \]

\[ = [\hat{U}(k)TH^d(k)]^\dagger[I - \hat{U}(k)]B^d(k), \]  

(5.21)

where \([::]^\dagger\) is the Moore-Penrose pseudoinverse. As we can see, the calculation of the pseudoinverse in each iteration is needed, which leads to additional computational complexity for the derivation of the solution in (5.21).

We can neglect the reverberation effect of \( \sigma(k) \) in the right hand side of (5.20) (i.e. assume that \( TH^r(k)\sigma(k) \rightarrow 0 \)), which is small if \( \sigma(k) \) is small and can be mitigated by the adaptive process. This is especially true under simple reverberent environments that features high direct to reverberation path ratios. Therefore, (5.20) can be simplified as

\[ \tilde{B}'(k) = [\hat{U}(k) - I]TH^d(k)l(k) + TH^d(k)\sigma(k). \]  

(5.22)
In practice, we can also apply the Tikhonov regularization method [70] and add a limiting term $\gamma$ ($\gamma \geq 0$) on the energy of updating signals to ensure a small value of $\sigma(k)$ so that the algorithm converges. It leads to a more efficient loudspeaker updating solution:

$$\sigma(k) = \left[ D(k)^H D(k) + \gamma I \right]^{-1} D(k)^H [ I - \hat{U}(k) ] B^d(k),$$

(5.23)

where $D(k) = TH^d(k)$ represents the coefficient matrix of the transfer functions for all loudspeakers assuming free-field propagation and it can be calculated offline.

Overall, the algorithm for the proposed reverberation cancellation system can be summarized as follows:

- **Step 1:** Initialize the updating signals $\sigma(k)_0 = 0$, the residual error $\tilde{B}_n'(k)_0 = 0$ and $\hat{U}(k)_0$ is a zero matrix. Let iteration counter be $\tau = 1$.

- **Step 2:** With the updated loudspeaker signals $l(k) + \sigma(k)_{\tau-1}$, obtain the measured soundfield coefficients for the $\tau$th adaption step following the measuring process in Sec. 5.4.2. Update the new residual $\tilde{B}_n'(k)_\tau$.

- **Step 3:** Calculate the diagonal entries $\hat{U}_n(k)_{\tau}$ in Eq. (5.16).

- **Step 4:** With $\hat{U}(k)_{\tau}$, derive the loudspeaker updating signals $\sigma(k)_{\tau}$ in Eq. (5.22) for reverberation compensation.

- **Step 5:** Increment $\tau$, return to Step 2.

### 5.5 Results and Discussion

We start with a discussion of simulation parameters. The speed of sound $c$ is 340 m/s in our simulations. We used the image source method [72] to simulate the soundfield created by the loudspeakers in the reverberant room. The centers of $\mathbb{D}_b$ and $\mathbb{D}_q$ lie on a circle of radius $d = 0.6$ m.
within $\mathbb{D}$. The target bright and quiet zones were located at $225^\circ$ and $45^\circ$, respectively, with $r_q = 0.3$ m, as shown in Fig. 5.1. We randomly selected $M/2$ locations within each zone and measured the value of the soundfield at those positions. The desired soundfield over $\mathbb{D}_b$ is selected to be a plane wave arriving from $90^\circ$ for the following simulations. The values of weighting function $w(x)$ assigned to $\mathbb{D}_b$, $\mathbb{D}_q$ and $\mathbb{D}_u$ were 1, 7.5 and 0.01 respectively. For the approach of IRLS, the sparsity promoting norm was $p = 0.3$. We used (5.23) to calculate the active loudspeaker updating signals in each iteration and the limiting term $\gamma = 0.005$. The pressure power of the desired soundfield was normalized to 0 dB at the centre of the bright zone, and complex Gaussian noise was introduced in order to maintain specific Signal-to-Noise Ratios (SNRs) with respect to this location.

![Figure 5.3](image)

Figure 5.3: Reproduction of the desired multizone sound using 20 noisy measurements for each selected zone in a reverberant room. (a) and (b) demonstrate the real and imaginary part respectively. The red crosses represent the positions of the microphones.

We compared the performance of the proposed method with the adaptive approach in [5] that requires the estimation of the reverberant component of the room channel, as well as the basis function approach in [6] (Chapter 4) which is a non-adaptive reverberation equalization method for multizone soundfield reproduction using sparse methods. We extended
Figure 5.4: Equalization error at the microphone measurements vs. the number of adaption steps. The results are averaged over 10 trial runs.

The work in [5] to the multizone case by applying the same multizone formulation method in [44], i.e., consider the desired multizone soundfield as a linear combination of plane waves arriving from various angles. The method of [5] requires the microphones to be evenly placed along the boundaries of the smallest circle that encloses the two selected zones. RLS was used as the adaptation algorithm of both the proposed technique and comparative method, where the forgetting factor was set to $\lambda = 0.95$. The following two measures will be used to evaluate the performance: i) The reproduction accuracy of the desired multizone soundfield, which is specified by the acoustic energy contrast between $D_b$ and $D_q$, as well as the MSE between the desired and the actual reproduced soundfield over $D_b$ [44]; ii) The equalization error to the desired soundfield coefficients at the microphone measurements after a number of adaptation steps, which is defined as

$$\varepsilon'(k) = 10\log_{10}|\tilde{B}'(k)H\tilde{B}'(k)|. \quad (5.24)$$
5.5. RESULTS AND DISCUSSION

5.5.1 Adaptive Reverberation Cancellation for Multizone Soundfield Reproduction

Due to the fact that the comparative method [5] is not capable of being extended to the 2.5D case, 2-D settings are applied to the simulations in this section for the purpose of better comparison. The reverberant room is set to be rectangular (size 6 m × 5 m) with wall reflection coefficients of 0.7. The center of $\mathbb{D}$ is located at (2 m, 2.5 m) and the loudspeakers are evenly distributed along a circle with a radius of 1.5 m. In the simulations, a total of 60 sources for each loudspeaker were included.

Fig. 5.3 demonstrates the reproduction of the desired multizone soundfield after 100 adaption steps using 20 noisy measurements (at the noise level of SNR 40 dB) for each selected zone and 40 loudspeakers at 1 kHz. We can see that the reproduced soundfield matches the desired multizone sound well. The acoustic contrast between $\mathbb{D}_b$ and $\mathbb{D}_q$ is 30.6 dB and the MSE over $\mathbb{D}_b$ is -25.4 dB. The equalization error at the microphone measurements is plotted in Fig. 5.4 as a function of the adaption step. The results are averaged over 10 trial runs. We tested three different settings with various noise levels. It can be seen that the equalization error gradually converges to the noise floor created by the external noise, especially for the case with relatively higher SNR. Naturally, the equalization performance improves as the level of the added noise decreases.

Given the same number of loudspeakers and microphones, we compare our proposed method with the reproduction approach in [5] in Fig. 5.5. The reproduction performance is plotted as a function of the adaption step with two noise settings. The results are averaged over 10 trial runs. From Fig. 5.5, we can see that the proposed method outperforms the reproduction approach in [5] in the aspects of both bright/quiet zone acoustic contrast and the MSE over $\mathbb{D}_b$ after 100 adaption steps. Additionally, the proposed method features a faster convergence rate to
Figure 5.5: Comparison of reproduction performance between our method and the approach in [5]. (a) and (b) represent the performance of acoustic contrast between $D_b$ and $D_q$ and MSE over $D_b$ respectively.

A good reproduction performance than the approach in [5].

The better performance of our method is due to the following reasons: i) the sparse estimation method facilitates a more accurate characterization of the reverberant room channel than classical estimation approaches, given the same provision of microphones [6], ii) The coefficient weighting function [2] attached to various modes in the cylindrical harmonic decomposition was not considered in [5] when minimizing the error between the desired soundfield and the reproduced soundfield coefficient set. In contrast, our formulation does not suffer from this issue as the employed basis function set $\{G_n\}_{n \in A}$ is formulated to be orthonormal over the desired reproduction region and it facilitates finding the optimal updating signal solution so that a more efficient reverberation cancellation is achieved.

In Fig. 5.6, we compare our proposed method with the reproduction approaches in [5] and [6] in terms of the wide-band multizone soundfield reproduction after 100 adaption steps from 100 Hz to 2 kHz. For our proposed method and the adaptive approach in [5], noisy measurements at the noise level of SNR 40 dB were introduced while the method of [6] employed noiseless measurements. A circular array of 75 loudspeakers
5.5. RESULTS AND DISCUSSION

Figure 5.6: Wide-band multizone soundfield reproduction with 64 noisy pressure samples, using the proposed method, the adaptive reproduction method in [5] and the non-adaptive method in [6]. The results are averaged over 10 trial runs.

were employed to satisfy the truncation length [62] for $\mathbb{D}$ at a maximum frequency of 2 kHz. Overall, we use 64 noisy measurements for all the three methods. As we can see in Fig. 5.6, our method outperforms the method proposed in [5] in terms of the desired multizone soundfield reproduction performance over the selected frequency range. The difference is even more obvious when the frequency gets close to 2 kHz. This is because the channel estimation process of the method in [5] becomes under-sampled for determining the soundfield coefficients with sufficient accuracy at relatively higher frequency, while the proposed method prevails thanks to the employment of sparse methods. We can also observe obvious peaks in the red curves for the method proposed in [5] due to the so-called large error scaling [2] at certain frequencies. A solution for this is to double the number of pressure measurements and sample over two concentric circles about these frequencies [2] [106], which makes the method cumbersome to implement. In contrast, the performance for the proposed method (blue curve) smoothly varies with frequency. The performance of the proposed method in this chapter also
approaches to the performance of the method in [6] that employs noiseless measurements, which is an ideal setup and can be difficult to meet in practice. Note that the method in [6] requires the pre-measurement of the transfer function over the selected zones of all the employed loudspeakers.

### 5.5.2 2.5D Multizone Soundfield Reproduction

With a sufficient number $Q$ of loudspeakers that are evenly placed in a circular shape, we can apply the reproduction method in Sec. 5.3 to implement the 2.5D multizone soundfield reproduction. In the following simulations, the 3-D reverberant room features a cubical shape (size $6 \text{ m} \times 5 \text{ m} \times 3 \text{ m}$). The circular reproduction region $\mathcal{D}$ has a radius of $r = 1\text{ m}$ with its center located at $(2, 2.5, 1.2)$. The simulations are implemented at $1\text{ kHz}$. Following the requirement on the number of loudspeakers for accurate 2.5D reproduction at $1\text{ kHz}$, we have $Q = 2\lceil k/r/2 \rceil + 1 = 53$ loudspeakers placed outside $\mathcal{D}$ on a full circle (we set the radius of loudspeaker array $R = 1.5\text{ m}$). In the simulations below, a total of 360 image sources are included for each of the loudspeakers.

![Figure 5.7](image_url)

**Figure 5.7:** 2.5D Reproduction of the desired multizone sound using 24 microphones for each selected zone in a reverberant room at 1000 Hz. The red crosses represent the location of the measurements. (a) and (b) demonstrate the non-regularized and the regularized results respectively.
5.5. RESULTS AND DISCUSSION

Figure 5.8: 2.5D Reproduction of the desired multizone sound with loudspeakers located at a different height in a reverberant room at 1000 Hz. The red crosses represent the location of the measurements. (a) and (b) demonstrate the non-regularized and the regularized results respectively.

Fig. 5.7 demonstrates the reproduction of the desired multizone sound using 25 noiseless measurements for each selected spherical zone of radius 0.3 m. The desired soundfield over $D_b$ is a monochromatic plane wave arriving from $60^\circ$. We created a challenging multizone sound reproduction scenario due to the occlusions of sound between $D_b$ and $D_q$. The average estimation error of the ATFs between all loudspeakers over the two selected zones was -34.7 dB. In Fig. 5.7(a), we can see that the reproduced soundfield matches the desired multizone sound well. The acoustic contrast between $D_b$ and $D_q$ is 24.3 dB and the MSE over $D_b$ is -18.4 dB. Comparing with the loudspeaker requirements in 3-D sound reproduction (at least $Q_{3D} = ([\text{ker}/2] + 1)^2 = 676$ loudspeakers with a spherical arrangement instead of the 53 for the 2.5D case), the 2.5D method is significantly more practical to reproduce the desired multizone soundfield over a desired horizontal region with 3-D settings. In Fig. 5.7(b), we regularized our system by adding an extra constraint on the loudspeaker array effort ($\tau = 0.01$ in Eq. (5.9)), the system performance was decreased to 19.8 dB and -13.7 dB in terms of acoustic contrast and
the MSE over $D_b$, respectively. Meanwhile, we note that the unwanted high-amplitude sound leakage is greatly alleviated, especially in the top left and around the desired reproduction region.

In Fig. 5.8, we have 53 loudspeakers placed on a full circle (we set the radius of loudspeaker array to 1.5 m) at the height of 2.5 m while the desired reproduction plane region is at the height of 1.2 m. Therefore, $h = 1.3 m$ in Fig. 5.2. This is to simulate a real-world scenario in which the loudspeaker array is placed near the ceiling of the room and it delivers the multizone audio toward the plane where the listeners’ ears are located.

The average estimation error of the ATFs between all loudspeakers over the two selected zones was -33.8 dB. In Fig. 5.8(a), we can see that the reproduced soundfield matches the desired multizone sound well. The acoustic contrast between $D_b$ and $D_q$ is 23.1 dB and the MSE to the desired plane wave function over $D_b$ is -15.2 dB. In Fig. 5.8(b), we regularized our system by adding an extra constraint on the loudspeaker array effort, the system performance was decreased to 18.7 dB and -11.2 dB in terms of acoustic contrast and the MSE over $D_b$ respectively.

Figure 5.9: The variation of the soundfield above and underneath (within a reasonable range) the desired horizontal region. (a) and (b) demonstrate the vertical consistency over the bright zone and the quiet zone respectively.
In Fig. 5.9, we also examine the variation of the soundfield above and underneath (within a reasonable range) the desired horizontal region for the regularized case (Fig. 5.8(b)). The motivation for this is to see if any unpleasant experiences would occur when the listener (or the microphone) is slightly misplaced vertically. The work in [156] demonstrates that a least-squares based multizone sound system can achieve satisfactory soundfield reproduction over the selected zones at heights between zero and 0.5 m from the loudspeakers’ plane under the free-field assumption. However, it might not be the case in reverberant environments because of the rapid variation of the ATF over the room. To evaluate the performance consistency over a vertical range (from -0.1 m to 0.1 m with respect to the desired horizontal plane), we introduced the following two measures: i) the MSE to the desired plane wave function over the bright zone at various height levels, ii) the ratio of the average acoustic sound energy (i.e. the square of sound amplitude) over the quiet zone at various height levels to the average energy of the desired plane wave over the bright zone. The results are plotted in Fig. 5.9.

From Fig. 5.9, we can see that, within a vertical misplacement distance of 2-4 cm, the soundfield variation is fairly acceptable with decent reproduction accuracy over the bright zone, as well as little sound leakage over the quiet zone. However, when the distance goes beyond 4 cm, it gradually undermines the listening experience for both of the selected zones with a poor reproduction accuracy over the bright zone and obvious sound leakage over the quiet zone. This simulation result is also consistent with the observation from the real-world implementation of our system, which will be discussed in Chapter 6.

Lastly, we conducted the 2.5D multizone soundfield reproduction with the adaptive reverberation cancellation system. We had 53 loudspeakers placed on a full circle at the height of 2.5 m. Fig. 5.10 demonstrates the reproduction of the desired multizone soundfield after 100 adaption steps using 25 noisy measurements (at the noise level of SNR 30 dB) for each
selected zone at 1 kHz. The desired soundfield over $D_b$ is the plane wave with the arriving angle of 90°. We can see that the reproduced soundfield matches the desired multizone sound well. The acoustic contrast between $D_b$ and $D_q$ is 22.9 dB and the MSE over $D_b$ is -15.87 dB.

5.6 Conclusion and Contribution

In this chapter, our main objective is to facilitate the practical implementation of the multizone soundfield reproduction system. We first introduced the concept of 2.5D reproduction to our existing multizone soundfield rendering system. It was shown that the desired multizone soundfield over a selected 2-D plane at the height approximately the same level of the listener’s ears can be reproduced using an enclosed array of loudspeakers with 3-D reverberant settings. The required number of loudspeakers is linearly proportional to the working frequency and the size of the desired reproduction region. Additionally, we presented a multizone soundfield reproduction system with the active reverberation cancellation approach,
which avoids the requirement for a prior measurement of the room channels for all loudspeakers. The concept of sparse approximation was applied to the adaptive channel estimation process using a limited number of randomly placed noisy measurements and the diagonal structure of the modeled channel transformation facilitates to reduce the computational complexity. The optimal loudspeaker updating signals that maximize the reverberation cancellation was also derived based on the estimate of the transformation matrix. Simulation results suggest that the proposed method provides a faster convergence rate than the comparative approach given the same hardware provision, as well as a consistently accurate reproduction of the desired soundfield over a wide frequency range.

The major contributions made in this chapter are:

- Design of multizone soundfield reproduction over a desired horizontal region using an enclosed array of loudspeakers located in the same plane or a plane at different height, with 3-D reverberant settings (the so-called 2.5D reproduction). The sparse methods are introduced to facilitate a reduction of the required number of microphones.

- Adaptive reverberation cancellation method for the multizone soundfield reproduction within the desired region that allows a parallel implementation and does not require a prior measurement of the loudspeaker ATFs over the desired region. It further improves the reproduction performance in terms of practical applications.

- The 2.5D multizone soundfield reproduction with the adaptive reverberation cancellation system is also presented.
5.7 Appendix: Investigation on the Required Number of Loudspeakers For 2.5D Reproduction

The objective of this appendix is to identify the required number of loudspeakers for 2.5D desired soundfield reproduction (e.g. plane wave). For a circular loudspeaker array of radius $R$ located on a horizontal plane with free-field setting, the actual reproduced soundfield can be written as

$$ S^a(x, k) = \sum_{q=1}^{Q} w_q(k) e^{-ik\|Y_q-x\|} \frac{4\pi}{\|Y_q-x\|}, \quad (5.25) $$

where $x \in \mathbb{R}^3$ specifies the position of the observation point. $w_q(k)$ is the filter gain for the $q$th loudspeakers and $Y_q = R\hat{\phi}_q$ specifies the location of the $q$th loudspeaker and $\hat{\phi}_q$ is the unit directional vector of the $q$th loudspeaker. Provided that the desired reproduction region is source-free, we can apply the addition theorem to rewrite the 3-D free-space Green’s equation in (5.25) [49]:

$$ e^{-ik\|Y_q-x\|} = \sum_{m=-\infty}^{\infty} \sum_{n=-|m|}^{\infty} e^{im(\phi_x-\phi_q)} j_n(k\|x\|) h_n^{(2)}(kR) Y_n^m(\phi_x, \frac{\pi}{2}) Y_n^m(\phi_q, \frac{\pi}{2})^* $$

$$ = \sum_{m=-\infty}^{\infty} e^{im(\phi_x-\phi_q)} \sum_{n=|m|}^{\infty} (-ik)j_n(k\|x\|) h_n^{(2)}(kR) 2n+1 \frac{(n-m)!}{(n+m)!} P_n^m(0)^2, \quad (5.26) $$

where $j_n(\cdot)$ denotes the $n$-th order spherical Bessel function, $h_n^{(2)}(\cdot)$ represents the $n$-the order spherical Hankel function of second kind and $P_n^m(\cdot)$ is the normalized associated Legendre function. Note that we set the elevation angle in all position vectors to $\pi/2$ as we are interested on the 2.5D soundfield reproduction that focuses on the horizontal plane. Using the Jacobi-Anger expression [49], we can also translate the desired plane
wave $F(x, \theta_{pw}, k)$ to a spherical harmonic expansion:

$$F(x, \theta_{pw}, k) = \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} 4\pi (-i)^n j_n(k||x||) Y_n^m(\phi_x, \frac{\pi}{2}) Y_n^m(\theta_{pw}, \frac{\pi}{2})^*$$

$$= \sum_{m=-\infty}^{\infty} e^{i m (\phi_x - \theta_{pw})} \sum_{n=|m|}^{\infty} 4\pi (-i)^n j_n(k||x||) \frac{2n + 1 (n - m)!}{4\pi (n + m)!} P_n^m(0)^2, \quad (5.27)$$

where $\theta_{pw}$ represents the arriving angle of the incident plane wave.

Putting (5.26) into (5.25) and equating the desired plane wave field (5.27) to the actual soundfield, we have:

$$\Psi_m(||x||, k) = \sum_{q=1}^{Q} w_q(k) \Omega_m(||x||, k) e^{i m (\theta_{pw} - \phi_q)}, \quad (5.28)$$

where

$$\Psi_m(||x||, k) = \sum_{n=|m|}^{\infty} 4\pi (-i)^n j_n(k||x||) \frac{2n + 1 (n - m)!}{4\pi (n + m)!} P_n^m(0)^2 \quad (5.29)$$

$$\Omega_m(||x||, k) = \sum_{n=|m|}^{\infty} (-ik) j_n(k||x||) h_n^{(2)}(kR) \frac{2n + 1 (n - m)!}{4\pi (n + m)!} P_n^m(0)^2. \quad (5.30)$$

For a desired soundfield within $\mathbb{D}$, the infinite superposition of mode $m$ can be truncated to the upper bound of $M' = \lceil ke||x||/2 \rceil$ with sufficient accuracy [113]. Therefore, we construct a linear system of equations from (5.28):

$$\Psi(||x||, k) = \Omega(||x||, k) w(k), \quad (5.31)$$

where $\Psi(||x||, k) = [\Psi_{-M'}(||x||, k), \ldots, \Psi_{M'}(||x||, k)]^T$, $\Omega = [w_1(k), \ldots, w_Q(k)]^T$. 

146
CHAPTER 5. DESIGN OF PRACTICAL MULTIZONE SOUNDFIELD REPRODUCTION

and

\[ \Omega(||x||, k) = \begin{bmatrix} \Omega_{-M'}(||x||, k)e^{im(\theta_{pw} - \phi_1)} & \ldots & \Omega_{-M'}(||x||, k)e^{im(\theta_{pw} - \phi_Q)} \\ \vdots & \ddots & \vdots \\ \Omega_{M'}(||x||, k)e^{im(\theta_{pw} - \phi_1)} & \ldots & \Omega_{M'}(||x||, k)e^{im(\theta_{pw} - \phi_Q)} \end{bmatrix}. \tag{5.32} \]

From (5.31), we can see that the system can only be satisfied exactly if \( Q \geq 2M' + 1 \) for each of position \( x \) with the distance \( ||x|| \). The Kirchhoff-Helmholtz theorem [51] suggest that as long as we can satisfy (5.31) on the boundary of the desired reproduction region, then reproduction will be accurate for all the points within the region. Therefore, the number of loudspeakers required for accurate reproduction of the desired soundfield over the entire region is \( Q \geq 2\lceil ker/2 \rceil + 1 \) (where \( r \) is the radius of the reproduction region of interest). Similarly to the 2-D soundfield reproduction, this rule also applies to the desired soundfield reproduction with reverberant and multizone settings [2] [157].
5.7. APPENDIX: INVESTIGATION ON THE REQUIRED NUMBER OF LOUDSPEAKERS FOR 2.5D REPRODUCTION
Overview: In this chapter, we describe the implementation of our proposed multizone soundfield reproduction approaches into a real-world system. The experiments were conducted in real listening environments in the electroacoustic lab of Victoria University of Wellington and the Huawei Munich Media Lab. The experimental results show that we can achieve a significant acoustic energy contrast between the signals recorded in the bright zone and the quiet zone, especially for the system implementation with reverberation equalization. An analysis of the possible factors that limit the system performance is also provided.
6.1 Introduction

Spatial multizone soundfield reproduction is a conceptually challenging problem in acoustic signal processing. Real-world implementations of the sound rendering systems, which aim to provide various soundfield environments over different spatial areas using only loudspeaker arrays, have rarely been reported in the prior art. In 2002, Choi and Kim [95] proposed an acoustic contrast control method to maximize the ratio of the mean square sound pressure in the bright and quite zones. By using this acoustic contrast control method, a personal audio system for a mono-sound was implemented in [110] and a 20 dB pressure level difference between the bright and quiet zone was achieved for the frequency up to 5 kHz in an anechoic chamber. Coleman et al. proposed the planarity control method in [47] [100], in which a cost function was formulated to optimize the acoustic cancellation within the quiet zone and the reproduction of the incident plane wave over a limited range of incoming azimuths in the bright zone. Therefore, this approach can be seen as a combination of the acoustic contrast control and the multiple-point pressure matching method. For the implementation of multi-point matching methods, the spacing between the measurement points is required to be less than half of the minimum wavelength (corresponding to the highest frequency), which is $\Delta x < \lambda_{\text{min}}/2$ [67].

Tashev et al. [94] employed a linear loudspeaker array to demonstrate the cancellation of sound in one area and the amplification in another simultaneously, which was based on an audio beamforming algorithm with the feedback from a sound localizer array (e.g. Kinect™ sensors). In [99], the author extended the spectral division method (SDM) [56] to the multizone soundfield case by modeling sound pressures as a rectangular window corresponding to bright and quiet zones. The system implementation was conducted in an anechoic room using a linear array of 64 loudspeakers.

\[ \Delta x = \lambda/2.5 \] is suggested [67].
Note that majority of the existing spatial soundfield reproduction systems are implemented in anechoic chambers; that is the sound rendering systems are implemented under an ideal free-field assumption. Very few studies considered the implementation of multizone soundfield reproduction systems in reverberant listening rooms. The reverberant case is difficult to handle because of the rapid variation of the transfer function over the room [101]. In practical scenarios, the performance of multizone sound reproduction techniques is commonly degraded by the effects of reverberation, especially for the multi-point based methods that require a large number of control points for fully characterizing the complicated soundfield [2].

In this chapter, we describe a real-world implementation based on the theory described in earlier chapters. We start with the implementation of the multizone sound system in a non-anechoic room based on the system design with an ideal free-field setting for soundfield rendering.² It can be clearly seen that the room reverberation undermines the soundfield reproduction performance of the system design under the free-field assumption. We then extend the multizone soundfield system with the reverberation equalization methods (as discussed in Chapter 4 and Chapter 5) based on the feedbacks from a limited number of microphone measurements.³ An analysis of the possible factors that limit the system performance is also provided for both cases.

### 6.2 Overview

The design and results of real multizone soundfield reproduction systems are presented in this section, as well as an analysis on the system performance. In Fig. 6.1, a flowchart that demonstrates how the theories of the proposed multizone soundfield systems in this thesis could be

---

²This part of work was conducted in the VUW Electroacoustic Lab.
³This part of work was conducted in Huawei Munich Media Lab.
6.2. OVERVIEW

Figure 6.1: Illustration of basic multizone soundfield rendering system implementation.

implemented in practice is presented. The system takes the coefficients for the desired multizone soundfield and the layout information of each loudspeaker and microphone as inputs. Note that for the reproduction systems with reverberation equalization, the recording signals from the microphone array should also be taken into account as an input for the estimation of the reverberant room channel. With these inputs, the system derives the optimal loudspeaker filter gains with the proposed algorithms. These pre-filtered loudspeaker signals are then passed through a series of devices (e.g. converters, amplifiers and etc.) and finally played through the loudspeaker array.

As the research focus of this work is on soundfield reproduction in a multizone setting, a natural demonstration of a multizone sound system is the creation of the desired bright zone and quiet zone simultaneously. For the implementation part of the work, we use the microphone recordings of the actual sound signals to evaluate the multizone sound rendering performance. By analyzing the sound recordings from microphones, it allows us to numerically quantify the acoustic energy contrast between the bright zone and the quiet zone, which has been widely used in the literature of spatial multizone soundfield reproduction. To illustrate, the authors in [110] suggest that an acoustic energy contrast of 20 dB between the bright and quiet zone denotes a perceivable multizone sound implementation. Note that the spatial sound systems introduced in
this thesis are designed to create multizone soundfield to a relatively large audience and are independent of the listeners or Head Related Transfer Functions. Therefore, the sound recordings from microphones are assumed to be the same with the sound perceived by the listener at the specified location.

### 6.3 Implementation of Multizone Soundfield Reproduction with Free-field Assumption

In this section, we describe the implementation of the proposed multizone soundfield reproduction approach (Chapter 3) into a real-world system that consists of 24 loudspeakers in the electroacoustic lab of Victoria University of Wellington (VUW). We designed a FIR filter for the source signal of each loudspeaker based on the design and theory of multizone soundfield reproduction with the free-field assumption in Chapter 3. Then, the filtered signals were played through the loudspeaker and recorded by two microphones that were located in the predefined bright zone and quiet zone, respectively. The output recorded signals were then analyzed and compared. The performance for both the narrowband and the wideband cases was examined.

#### 6.3.1 Experimental Setup

Overall, our system has 24 loudspeakers and one microphone. The loudspeaker model was Mackie\textsuperscript{TM} MR5 Studio Monitor and the microphone model was DPA\textsuperscript{TM} 4061 Omnidirectional, Lo-Sens microphone. The employed loudspeakers and microphone were driven by three Presonus\textsuperscript{TM} FP10 recording systems. Each Presonus\textsuperscript{TM} FP10 provides eight analog microphone inputs and eight analog outputs, which means each device can support up to eight loudspeakers and eight microphones [158]. It also features preamplification to supply 48V phantom power to microphones.
6.3. IMPLEMENTATION OF MULTIZONE SOUNDFIELD REPRODUCTION WITH FREE-FIELD ASSUMPTION

Figure 6.2: Arrangement of loudspeakers and microphones in the VUW electroacoustic lab. From [7], with permission.

for recording [158]. In our work, we connected three FP10 via firewire cables so that the entire system supports up to 24 loudspeakers and 24 microphones.\(^4\) The laboratory is a 3.86 m \(\times\) 3.86 m square room with acoustic absorbers being placed on all four sides of its walls, as well as on the ceiling. The arrangement and the orders of the loudspeakers are shown in Fig. 6.2.

In the practical experiments, the 24 loudspeakers were placed along an enclosed circle with a radius of 0.75 m. Both the bright zone and the quiet zone were inside the desired reproduction region that was a concentric circle with a radius of 0.6 m. The bright zone and the quiet zone also had a disc-shape of radius 0.2 m. The configuration of the system is shown in Fig. 6.3.

6.3.2 Loudspeaker Gain Calibration

In reality, the loudspeakers differ from each other with respect to frequency response. Therefore, it is indispensable to normalize the filtered

\(^4\)Each of the recording system drives 8 loudspeakers and the three channels were synchronized in advance.
signals before they are played through loudspeakers. The normalization process can generally eliminate the variations caused by various loudness levels among different loudspeakers, as well as the changes of single loudspeaker loudness level over time [159]. Similarly, microphone gain calibration can also be conducted using the approach in this section if necessary.\(^5\)

Let \(s_{\text{ref}}[n]\) (where \(n\) is the sequential number) be a reference signal. A gain compensator \(c_t\) is designed to adjust the energy level of the test signal \(s_t[n]\),

\(^5\)Since only one microphone was used, the microphone gain calibration process is skipped for this part of work.
6.3. IMPLEMENTATION OF MULTIZONE SOUNDFIELD REPRODUCTION WITH FREE-FIELD ASSUMPTION

which can be derived as

\[ c_t = \sqrt{\frac{\sum_{n=0}^{\infty} |s_{\text{ref}}[n]|^2}{\sum_{n=0}^{\infty} |s_{t}[n]|^2}}. \]  

(6.1)

To compute the gain compensator for all 24 employed loudspeakers, we placed a microphone at the center of the circular loudspeaker array. This is to ensure that the distances between the microphone and each of the loudspeakers are identical. The reference signal was chosen to be the signal played by a selected loudspeaker (e.g. the first loudspeaker). Then, we played a sinewave signal at the specified frequency through each of the employed loudspeakers and recorded through the microphone. The gain compensator of each loudspeaker was then computed by using Eq. 6.2. It is suggested to repeat these steps and take the average value to ensure more accurate and consistent gain compensators. In our work, a four-time repetition was performed.

6.3.3 Experimental Procedure – Narrowband Case

In this section, we discuss the implementation procedures of the narrowband experiment. This part of work was based on the theory and design in Chapter 3.\(^6\)

For the narrowband case, the source signal for each loudspeaker was selected to be a 1000 Hz sine wave that only contains one frequency component. Firstly, we normalized the source signal by taking into account the corresponding loudspeaker gain compensator (as discussed in Sec. 6.3.2). Then, the compensated input signals were filtered by the loudspeaker weights designed by the 2-D multizone soundfield reproduction system in Chapter 3, in which the acoustic transfer function from loudspeakers was chosen to be 2-D Green’s function model with free-field assumption.

\(^{6}\)The experimental work for this narrowband case in Sec. 6.3.3 was conducted by Kelson Jiawen Chua with the author’s support. The rest of the experimental activities reported in this chapter were solely conducted by the author.
Subsequently, the filtered signals were passed to the recording systems FP10s and assigned to the corresponding loudspeakers. Finally, the signals were then played through the loudspeaker array and recorded by the microphone measurements in the bright zone and quiet zone. In terms of the microphone position in the bright zone, it was located at the center of the bright zone \( b (-0.3, 0) \) to record the reproduced sound pressure. For the "quiet" zone, we selected three locations, which are \( q_1 (0.3271, 0.0626) \), \( q_2 (0.3, 0) \) and \( q_3 (0.3271, -0.0626) \).

Figure 6.4: Soundfield reproduction of 1000 Hz signal at desired region by 24 loudspeakers. From [7], with permission.

Given the system model in Fig. 6.3, we first implemented the theoretical multizone soundfield reproduction at 1000 Hz under the free-field assumption. We defined the desired soundfield over the bright zone to be a plane wave function arriving from the angle of \( 15^\circ \). The simulation was based on the reproduction method in Chapter 3. We can clearly see from Fig. 6.4 that a region with attenuated acoustic energy has been created and the acoustic contrast between the bright zone and quiet zone is 23.02 dB. The sound energy difference between the center point of the bright zone
and quiet zone is 28.6 dB.

For the practical results, the pre-filtered signals were played through the 24 employed loudspeakers. The reproduced sound signals were recorded at $b$, $q_1$, $q_2$ and $q_3$ accordingly. In Fig. 6.5, the recording signals at these four points are plotted, in which the blue curve represents the signal recorded at the center of the bright zone $b$.

![Figure 6.5: Recorded signals at $q_1$, $q_2$, $q_3$ and $b$.](image)

Figure 6.5: Recorded signals at $q_1$, $q_2$, $q_3$ and $b$. From [7], with permission.

From the results in Fig. 6.5, we can clearly see that the signal amplitude that represents the acoustic energy at $b$ is significantly higher than those at $q_1$, $q_2$ and $q_3$ in the quiet zone. A delay can be observed within the period from 0 to 0.1 s, which is mainly due to the internal delay of the recording systems FP10s and the duration of sound propagation from the loudspeakers to the microphone position. Therefore, we eliminated the time interval between 0 and 0.1s and the acoustic energy of the signals at the selected points were calculated over a time window from 0.1 to 2 s. The acoustic energy contrast between the signal recorded in the bright zone $s_b[n]$ and the signals recorded in the quiet zone $s_q^i[n]$ (where $i = [1, 2, 3]$)
can be calculated as

\[ \epsilon_i^j = 10 \log_{10} \frac{\sum_{n=0.1f_s}^{2f_s} |s_b[n]|^2}{\sum_{0.1f_s}^{2f_s} |s_q^j[n]|^2}, \tag{6.2} \]

where the sampling rate \( f_s = 44100 \) Hz. The energy contrast between the signal recorded at \( b \) and the signals at \( q_1, q_2, q_3 \) are 12.67 dB, 12.21 dB and 14.23 dB, respectively. Comparing with the theoretical results, we can clearly observe a performance loss in terms of the acoustic contrast between the bright zone and the quiet zone. A discussion on this performance loss is provided in Sec. 6.3.5.

6.3.4 Experimental Procedure – Wideband Case

We also implemented the wide-band multizone soundfield reproduction with free-field assumption in the VUW electroacoustic lab. The source signals are first filtered into \( N \) sub-bands in the frequency domain. The band-pass filter for the \( n \)th sub-band \( l_n(k) \) is derived based on its corresponding center frequency using the method in Chapter 3. As the band-pass filtered signals are to be recombined at a later stage, a filter-bank allowing perfect signal reconstruction is desirable. In our work, the method of Gabor filter banks is used.\(^7\)

In the following work, the source signal is a music segment that features the frequency range from 100 to 4000 Hz with a length of 5 seconds. We divided the signal into \( N = 78 \) sub-bands, which implies that each sub-band covers a frequency range of 50 Hz. To define the parameters for the Gabor filter banks, as we have 78 sub-bands in the frequency domain, we selected the Hamming window function with a length of \( 2 \times 78 - 2 = 154 \) due to the conjugate symmetric property. An oversampling of two was used for perfect reconstruction and the elimination of window edge distortion. We first implemented a forward transform of the Gabor filters with

\(^7\)Please refer to [160] for more details of Gabor filters.
the input of the source signal, then we multiplied the corresponding filter
 gains (which were derived offline) and loudspeaker gain compensator
to each sub-band. Finally, the filtered signal for each of the employed
loudspeakers in the time domain was obtained by performing a backward
transform.

In terms of the configuration of the selected zones, a similar system
model to Fig. 6.3 was used but we defined a relatively smaller desired
reproduction region of radius 0.32 m following the linear rule between the
number of required loudspeakers and frequency [62]. One bright zone
and one quiet zone were included in the reproduction region, each with a
radius of 0.1 m and the distance between the centers of the two selected
zones was 0.4 m.

To derive the loudspeaker gain compensator for all 24 loudspeakers within
the frequency range, we once again put a microphone at the center
point of the loudspeaker circular array to ensure the distances between
microphone and each loudspeaker were identical. Then, a sequence of
six concatenated chirp signals was played through each loudspeaker and
recorded at the microphone. The gain compensator of each loudspeaker
for each frequency band was then computed by averaging the last five
chirp signals in the frequency domain. Finally, we normalized the gain
compensator for the 24 loudspeakers to avoid clipping. For more details
of the procedures of the gain compensator computation, refer to Sec.
6.3.2.

Given the system setup provided above, we first implemented the theoret-
ical simulations of the multizone soundfield reproduction with the same
wideband settings. In Fig. 6.6, the performance of the acoustic contrast
between the bright zone and quiet zone are plotted for the centered
frequencies. From Fig. 6.6, we can see that when the frequency goes
beyond 500 Hz, the acoustic contrast performance remains stable over
24 dB. The relatively inferior performance at low frequencies (lower than
500 Hz) is due to the fact that the soundfield rendering over both the

160
Figure 6.6: The performance in terms of the acoustic contrast between the bright zone and quiet zone are plotted for the centered frequencies.

bright zone and quiet zone is physically more challenging when the sound wavelength is long and the distance between the two zones is low. The sawtooth characteristic of the performance may be caused by the fact that the continuous aperture function is not efficiently sampled at those notch frequencies, given the arrangements of the 24 loudspeakers.

In Fig. 6.7, we simulate the measurements of the sound at the centers of the bright zone and the quiet zone by assuming that the acoustic transfer function from the loudspeakers follows the 2-D Green’s function in the free-field. A 27 dB acoustic energy contrast between the selected two points is obtained, which represents that theoretically our system can also be extended to the wideband settings.

For practical results, the measurements of the recorded signals at the center of the bright zone and the quiet zone are plotted in Fig. 6.8 as obtained under the free-field assumption, which is not accurate for

---

8This artifact can be observed in various studies using different methods (e.g. [18] [98]), the reason is yet to be explored.
6.3. IMPLEMENTATION OF MULTIZONE SOUNDFIELD REPRODUCTION WITH FREE-FIELD ASSUMPTION

Figure 6.7: The simulated measurements of the sound at the centers of the bright zone and the quiet zone are plotted.

this scenario. A 12.1 dB acoustic energy contrast between the selected two points was obtained. We repeated the recording procedure at three different pairs of microphone location within the selected zone, and the averaged acoustic contrast between the bright zone and quiet zone was 11.5 dB. Comparing Fig. 6.8 with Fig. 6.7, once again we can observe a performance loss of the desired multizone soundfield reproduction.

6.3.5 Performance Analysis

There are a number of potential limitations of the performance of our system. First and foremost, the system we implemented in this section is based on the design with an ideal free-field setting. However, the listening room is not fully anechoic. The performance of spatial soundfield reproduction techniques is generally degraded by the effects of ambient reverberation (i.e. reflection from walls, floor and ceiling). In [7], Chua
measured the reflection coefficients for the four walls\textsuperscript{9} in the Victoria electroacoustic lab. Given the measured reflection coefficients and the geometry of the room, we used the image source method\textsuperscript{[72]} to simulate room reverberation and implemented the multizone soundfield reproduction in Fig. 6.4 with the same set of loudspeaker filter gains. 100 image sources were included for each of the loudspeaker positions. The simulation of the acoustic energy of the reproduced soundfield over the desired reproduction region is plotted in Fig. 6.9. The solid line encircles the desired reproduction region. The dashed lines encircle the bright zone and the quiet zone with the marker “B” and “Q”, respectively. From Fig. 6.9, we can see that the performance of the desired multizone soundfield reproduction is significantly undermined comparing with the case of the ideal free-field settings of Fig. 6.5. In Fig. 6.9, the average acoustic contrast between the bright zone and the quiet zone is 12.43 dB, which

\textsuperscript{9}The reflections from the floor and ceiling were not considered for simplicity in the 2-D case.
6.3. IMPLEMENTATION OF MULTIZONE SOUNDFIELD REPRODUCTION WITH FREE-FIELD ASSUMPTION

Figure 6.9: Soundfield reproduction of 1000 Hz signal at desired region by 24 loudspeakers given the measured the reflection coefficients and the geometry of the VUW electroacoustic lab. From [7], with permission.

is very close to the results obtained from the practical implementation in Sec. 6.3.3. Therefore, we can conclude that the performance of the free-field multizone soundfield reproduction system is mostly limited by the room reverberation. The effects of reverberation should be taken into account to improve our system performance, which will be introduced in the following section.

The 3-D directivity pattern of the employed loudspeakers was also studied and measured in [7]. It was shown that the employed loudspeaker features a directivity pattern between the spherical cap model [53] and the point source omnidirectional model within the frequency range of interest from 100 Hz to 5000 Hz. Comparing with the ideal point source model that radiates sound equally towards all directions, the sound radiation at the front side of the spherical cap based source model is stronger than its sidelobe and this beam-like directional property gets more obvious as the working frequency goes up to 20 kHz. The results in [7] show
that even though the directivity pattern of the employed loudspeakers within the frequency range of interest (100 Hz-5000 Hz) is non-ideal, it does not adversely affect the performance of the multizone soundfield reproduction system. For more details of the loudspeaker directivity pattern, refer to [7].

The following measures can be considered to further improve the performance for this task:

- To make the listening room less reverberant. As this part of work is based on the assumption that the ATF from loudspeakers follows the 2-D Green’s function in free-field, the reverberation in the listening room inevitably undermines the reproduction performance. A natural way to make the room less reverberant is to hang more absorptive curtains. Now only two sides of walls are covered by absorption panels.

- To make the loudspeaker placement more accurate. In our system, the loudspeakers are assumed to be placed uniformly along a circle of radius 0.75 m. Currently, the loudspeaker arrangement is only roughly placed.

- To replace the defective loudspeakers.

### 6.4 Implementation of Multizone Soundfield Reproduction with Reverberation Equalization

In this section, we implement our proposed multizone soundfield reproduction system with reverberation equalization in the Huawei media lab. The goal was to reproduce the desired soundfield in a pre-defined acoustic bright zone and minimize the sound pressure in the specified quiet zone in a real listening environment.
The multi-channel audio reproduction system in Fig. 6.10 consisted of 48 loudspeakers and two microphones. The loudspeakers were evenly placed on a circle of radius 2 m. The model of the loudspeaker was Canton CD 1020 and the microphone model was a DPA 4061 Omnidirectional miniature microphone. All loudspeakers were driven by two RME Fireface 800 soundcards,\textsuperscript{10} four RME ADI 8DS AD converters, and six IMG STA-1508 8-channel amplifiers.

The testing environment was an office room with rectangular shape of dimensions 5.9 m length by 4.6 m width by 3 m height. The ceiling was treated with acoustic absorbers to reduce vertical reflections and the floor was covered by carpet. One wall of the testing room was completely covered by a sound-absorptive curtain and one wall featured large glass windows. No acoustic treatment was applied to the remaining two walls (as shown in Fig. 6.10). The testing room had features that created some diffusion in the lateral dimension: a doorway, a window, a PC, a wheel

\textsuperscript{10}Each of the soundcard drives 24 loudspeakers and the two channels were perfectly synchronized in advance.
chair, two shelves that host soundcards/converters, etc. Overall, the reverberation time $T_{60}$ was approximately 0.4 s at 1 kHz. Considering the circular setup of the loudspeaker array, we set the desired reproduction region $\mathbb{D}$ on the 2-D plane inside the loudspeaker array. The origin of the coordinates was the center point of the circular loudspeaker array.

In this section, we start with the real-world implementation work of the proposed Green’s function modeling approach in Chapter 4. Then, the narrow-band multizone soundfield reproduction using microphone feedbacks at the frequency of 1 kHz is presented. The setup and implementation instructions of the wide-band multizone soundfield reproduction with adaptive reverberation cancellation system are also introduced.

### 6.4.1 Real-world Estimation of Desired Loudspeaker ATF

![Microphone arrangements for the loudspeaker ATF estimation experiment. The red crosses represent the locations of microphone measurements.](image)

Figure 6.11: Microphone arrangements for the loudspeaker ATF estimation experiment. The red crosses represent the locations of microphone measurements.

In this section, we present the experiments that examine the ATF estimation of a specified loudspeaker in the array system. The goal is to examine the effectiveness of the loudspeaker ATF estimation using the proposed
sparse methods in a real listening environment and to see if the practical results are consistent with the simulation analysis presented in Sec. 4.7.

As shown in Fig. 6.11, we defined an circular estimation area of radius 0.3 m and the locations of microphone measurements were randomly selected inside. As we confined the estimation of the loudspeaker ATFs to the 2-D plane where the loudspeaker array was located, the 18 measurement points were strictly placed at the same height level with the loudspeakers (i.e. we set the height level of the loudspeakers to \( z = 0 \) m). The testing loudspeaker was located at the position of \((0, -2)\) in Fig. 6.11.

The samples of the ATFs of the loudspeakers were based on the impulse responses from the loudspeakers to the microphones. The room impulse responses were measured by using the logarithm swept chirp method [124] with periodically five repeated chirp sequences. Since natural noises from the stereo system (i.e. soundcards, AD converters, amplifiers and etc.) and other loudspeakers were present, multiple periods of the chirp were used to improve the SNR of the ATF measurements by assuming time invariance. The length of each of the chirp signal was 44100 samples at a sample rate of 44.1 kHz. Fig. 6.12 shows the RIR of the testing loudspeaker when the microphone is placed at the center of the loudspeaker array.

A cross-validation of the estimation of desired loudspeaker ATF from 100 to 1 kHz over the area of interest was conducted, in which we estimated the loudspeaker ATF at one single point based on the remaining 17 ATF measurements in Fig. 6.11 and compared the estimate of the ATF with the actual measured frequency response. Then we repeated the estimation process for all 18 points and the averaged MSE was shown in Fig. 6.13. As we can see, accurate loudspeaker ATF estimation can be achieved over the desired region based on only 17 microphone measurements at lower frequencies in this real-world experiment and the performance gradually degrades as the frequency increases up to 1 kHz, a result that is similar to the simulation results in Fig. 4.3.
In Fig. 6.14, the estimated frequency responses at point 17 and 18 based the ATF measurements of points 1-16 are plotted. For point 17, which was close to other microphone measurements (especially point 13), the estimated loudspeaker ATF matches the actual measurement accurately for the entire testing frequency range. In contrast, the estimation perfor-
Figure 6.14: The measured and estimated frequency response at the point 17 and 18.

Performance was inferior at high frequencies for point 18, which was distant from the employed 16 measurement points. This is consistent with the results derived from the CRB analysis in Chapter 4, which states that the desired loudspeaker ATF can be estimated more accurately when it is close to the locations of microphone measurements.
6.4.2 Multizone Soundfield Reproduction with Reverberation Equalization – Narrowband

The implementation of this part of work is based on the 2.5D multizone soundfield reproduction in reverberant environments discussed in Chapter 5, which implies that the reflections due to the walls, floor and ceiling are all considered. We used the microphone measurements to sample the soundfield at the frequency of 1 kHz at the same plane of the loudspeaker array, which can then be used to estimate the ATFs of the loudspeakers over the desired region for filter design. In terms of the configuration of the selected zones, we define a desired reproduction region of radius 1 m. One bright zone and one quiet zone are included in the reproduction region, each with a radius of 0.3 m and the distance between the centers of the two selected zones is 1.1 m. Overall, we sampled the ATFs of the loudspeakers at 15 selected measurement points for each of the selected zones (as shown in Fig. 6.15).

Figure 6.15: System layout and microphone arrangement. The red crosses represent the position of microphone measurements.

The samples of the ATFs of are derived by finding the values of the room impulse responses at the frequency bin of 1 kHz for all 48 loudspeakers.
We played out the filtered input signal (two-second sinewave signal at 1 kHz) through the loudspeaker array while two microphones were used to record the reproduced sound signal within the selected bright zone and quiet zone.\textsuperscript{11} The results show that we can achieve as good as 17.9 dB

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure616.png}
\caption{Performance of the acoustic contrast between the bright zone and quiet zone at four different pairs of locations. (a) represents the acoustic contrast between (0.6,0) and (-0.6,0) in Fig 6.15, (b) represents the acoustic contrast between (-0.7,-0.25) and (0.7,-0.25), (c) represents the acoustic contrast between (-0.65,0.15) and (0.65,0.15) and (d) represents the acoustic contrast between (-0.42,-0.2) and (0.42,-0.2).}
\end{figure}

\textsuperscript{11}The different microphone gains at 1 kHz are measured and considered in further
acoustic contrast between the bright zone and the quiet zone (as shown in Fig. 6.16(a)). We found that the best acoustic contrast was obtained when we placed the microphone and recorded the reproduced sound signal right at (or very close to) the position where the loudspeaker ATFs were sampled. When we moved the microphones to a position that is not close to any of the sampling points, the performance decreased (e.g., the acoustic contrast is dropped to 9.8 dB in Fig. 6.16(b)). This is due to the fact that the more accurate estimation of the loudspeaker ATFs generally leads to better performance in terms of the reproduction of the desired soundfield in the bright zone and quiet zone. As we mentioned above, the desired loudspeaker ATF can be estimated more accurately at the location of measurements and its neighborhood and the performance goes down with the distance from the measurements. We repeated the recording procedure at 10 different pairs of microphone locations within the selected zone, the averaged acoustic contrast between the bright zone and quiet zone was 13.5 dB.

![Figure 6.17: The acoustic contrast between the center of the bright zone and quiet zone at the height level of 0 cm (a) and -6 cm (b).](image)

Additionally, we also tested the variation of the acoustic contrast performance above and underneath (within a reasonable vertical range) the
desired plane. We first placed microphones at the centers of the bright zone and the quiet zone (relative height level is 0 cm), at which the acoustic contrast is 15.9 dB (Fig. 6.17(a)). Note that no loudspeaker ATF sampling is measured at the centers of the selected zones, thus it was reasonable that the performance was not as good as the optimal performance (i.e. 17.9 dB). Then we moved up or moved down the two microphones at the same time with the same distance and observed the variation of the soundfield quality. The acoustic contrast at the height of +6, +4, +2, -2, -4, -6 centimeters are -7.5 dB, 1.1 dB, 8.8 dB, 9.7 dB, 1.8 dB and -9.5 dB (Fig. 6.17(b)) respectively. We can see that within a vertical misplacement distance of 2 cm to 4 cm, the soundfield variation is fairly acceptable with decent acoustic contrast between the bright zone and quiet zone. However, when the distance goes beyond this range, the performance gradually degrades. The results are consistent with the simulation results that we derived in Chapter 4.

In order to further improve the performance for this part of work, we have the following suggestions:

- Take more measurements to sample the desired loudspeaker ATFs. For this work, 15 microphone measurements were sampled within each of the selected zone to estimate the desired loudspeaker ATFs for the 2.5D case. As shown in Chapter 4, the employment of more measurements improves the fidelity of the overall estimation over the selected zones thereby leading to more accurate soundfield reproduction.

- Minimize the background noise in the listening room. Currently, the stereo system (i.e. soundcards, AD converters, amplifiers and etc.) in the Huawei media lab is placed next to the loudspeakers and microphones, which create a fair amount of background noise. The background noise deteriorates the accuracy of the loudspeaker ATF estimation. It was shown in [6] that a high measurement SNR
(preferably over 30 dB) is required for an accurate reconstruction of the desired ATF at 1 kHz using a limited number of noisy measurements. Naturally, the estimation performance improves as the level of the added noise decreases.

• To measure the locations of the microphone placement more precisely.

6.4.3 Comments on Multizone Soundfield Reproduction with Reverberation Equalization – Wideband

The real-world implementation of the multizone soundfield reproduction with reverberation equalization can also be extended to the wideband case based on the theory of adaptive reverberation cancellation method discussed in Chapter 5. The following text provides a brief instruction on how the experiments could be approached.

**Experimental setup** In terms of the configuration of the selected zones, we can define a desired reproduction region of radius 0.4 m. One bright zone and one quiet zone are included in the reproduction region, each with a radius of 0.15 m and the distance between the centers of the two selected zones is 0.44 m. The desired soundfield over the bright zone is a plane wave arriving from the angle of $90^\circ$. Multiple microphone measurements are required for each of the selected zone to ensure accurate active estimation. 15 microphones for each of the selected zones are suggested.\(^{12}\)

**Calibration** Since multiple measurements are needed simultaneously to provide active feedback of the reproduced soundfield, a microphone gain calibration procedure must be pre-conducted. We can first select

---

\(^{12}\text{We may also reduce the number of microphones for each zone to see how would the system performance be affected.}\)
a loudspeaker with normal functionality. Then a microphone is placed at the center point of the loudspeaker array and play a chirp signal through the selected loudspeaker. Repeating this process for all the microphones and recording signals can be used to calculate the calibration coefficients following the calibration approach in Sec. 6.3.2. Note that a loudspeaker gain calibration procedure is not needed in this case, as the active microphone feedback of the reproduced soundfield covers the information about various gains of different loudspeakers in the adaption procedure.

**Main routine: the adaption process** The main routine of the wideband multizone soundfield reproduction is the adaptive reverberation cancellation process. The test signal used during the adaption process can be a two-second chirp signal sweeping from 100 Hz to 4 kHz. This is to ensure that all frequency components are covered within the frequency range of interest. Using the frequency-domain design in Chapter 5, the FIR filters can be populated by considering a bin-by-bin approach. Note that the frequency bin width is suggested to be less or equal than $4.4/T_{60}$ [161], in order to address the frequency correlation and make sure the estimated room transfer functions are valid across each bin.

Following the adaptive reverberation cancellation method, we can start with the loudspeaker gain solution that is pre-derived under the free-field assumption at the center frequencies of the bins. Gabor filter banks are used to decompose and recombine the band-pass filtered signals. The filtered signals are then played back through the loudspeaker array and the recording signals from the microphone measurements are used to estimate the reverberant room channel based on sparse approximation methods. We can then determine the updating signals on the loudspeaker array for the following adaption step, which is interlaced with the active estimate of the reverberant room channel. We can examine how the performance of the acoustic contrast between the bright zone and the
quiet zone evolves. Finally, the updated loudspeaker gains will be used
to design the filter for playing back a music segment that features the
frequency range from 100 to 4 kHz with a length of 5 seconds.

6.5 Conclusion and Contribution

The majority of the existing spatial soundfield reproduction systems are
implemented in anechoic chambers under the ideal free-field assumption.
Being motivated by this, we conducted real-world implementations based
on the previously developed theories. We started with the implementation
of our system in a non-anechoic room with the free-field design. An
average acoustic contrast between the bright zone and quiet zone of
12.9 dB was achieved with the multizone sound settings at 1 kHz. We
concluded that the room reverberation was the main impact factor that
undermines the performance of the reproduction system based on the
design with free-field assumption, which is an ideal setting for soundfield
rendering. We then extended the multizone soundfield system with
the reverberation equalization method based on the feedback a limited
number of microphone measurements. From the practical results for the
narrowband case, the acoustic energy contrast between the two selected
zones can reach up to 17.9 dB. The optimal performance was obtained at
those points that were close to the location where the loudspeaker ATFs
were sampled. The system performance drops as the distance of the
observation points to the location of measurements is increased.
The investigation on the variation of the acoustic contrast performance at
different height level also suggests that the system allows a reasonable
vertical misplacement distance (up to 4 cm) while the sound rendering
performance is still acceptable within this range.
The major contributions made in this chapter are:

• Real-world implementation of the proposed multizone soundfield
  system under the free-field assumption in a non-anechoic environ-
6.5. CONCLUSION AND CONTRIBUTION

- Real-world implementation of the proposed multizone soundfield system in a real listening room with the reverberation equalization design based on a limited number of microphone measurements. The experiments are conducted at a single frequency.

- Analysis on the possible impact factors that limit the multizone soundfield rendering performance for both the case with the free-field design and reverberation equalization.
Overview: In this chapter, we state the general conclusions drawn from this thesis. Some future research directions that arises from our work are also outlined.

7.1 Conclusion

In this thesis, we have addressed the following problem:

The development of a practical multizone soundfield reproduction system that enables us to control the reproduction of the desired multizone soundfield over each of a set of selected zones using a loudspeaker array in a given complex environment.

To address the problem, several multizone soundfield reproduction techniques in complex environments (i.e. reverberant environments) have been proposed:

- Multizone soundfield reproduction using orthogonal basis expansion
7.1. CONCLUSION

We introduced a method for spatial multizone soundfield reproduction based on describing a desired multizone soundfield as an orthogonal expansion of basis functions over the desired reproduction region, which provides the theoretical basis of both 2-D (height-invariant) and 3-D multizone soundfield reproduction for this work. This approach finds the solution to the Helmholtz equation that is closest, in the weighted least squares sense, to the desired soundfield. The basis orthogonal set was formed using a modified Gram-Schmidt process with a set of plane wave functions as input. The coefficients of the Helmholtz solution wavefields can then be calculated, which reduces the multizone sound reproduction problem to single zone reproduction problems over the entire desired region. The approach is shown effective for the control of the soundfield in the selected bright zone and minimization of sound interference into the predefined quiet zone.

- Multizone soundfield reproduction in reverberant rooms using sparse approximation methods

We introduced a method of reproducing a multizone soundfield within a desired region in reverberant environments. The key of this method is the identification of the acoustic transfer function (ATF) between the loudspeaker over the desired reproduction region using a limited number of microphones. The approach assumes that the soundfield is sparse in the domain of plane wave decomposition and identifies the ATF using sparse methods. The estimated ATFs were then used to derive the optimal least-squares solution for the loudspeaker filters that minimizes the reproduction error over the entire reproduction region in terms of the coefficient set for the orthogonal basis functions. Simulations confirm that the method significantly reduces the number of microphones needed for an accurate multizone soundfield reproduction, while it also facilitates
the sound reproduction over a wide frequency range. We also discussed the 3-D multizone soundfield reproduction based on the proposed method and extended the theory to the so-called 2.5D multizone soundfield reproduction.

- **Multizone soundfield reproduction with adaptive reverberation cancellation**

  We introduced an adaptive reverberation cancellation method for multizone soundfield reproduction using sparse methods. The proposed method does not require a prior measurement of the room transfer function for the employed loudspeakers. The proposed approach expresses the soundfield as an orthonormal basis function expansion in the space-frequency domain over the desired reproduction region. The method considers the reproduced soundfield as a linear transformation of the desired soundfield, where the coefficient set for the desired soundfield can be computed under the free-field assumption. The method uses an adaptive channel estimation process using sparse methods to identify these transformations. The required loudspeaker updating signals are interlaced with the active estimate of the reverberant room channel. Simulation results confirm that the proposed method for room reverberation compensation for multizone reproduction provides a better convergence behavior than the comparative methods.

The above mentioned techniques show that the reproduction of multiple independent soundfields in separate zones using a loudspeaker array is feasible. The main challenge for multizone soundfield reproduction is the reverberation equalization in complex environments over a wide frequency range. The improvements of the proposed multizone reproduction system with regard to practical implementation has also been investigated. It has been shown that the proposed approaches facilitate to the reduction of the required number of loudspeakers and microphones.
for an accurate multizone sound reproduction and a flexible placement of the microphone array is allowed. The implementation of the proposed multizone soundfield reproduction approaches into a real-world system is described. We introduce the experimental setup, design, results and performance analysis of real-world multizone soundfield reproduction systems in this thesis.

7.2 Future Research Directions

In this section we outline several future research directions that arise from the work proposed in this thesis.

- **Practical Implementation of 3-D Multizone Soundfield Reproduction**

  The practical implementations of multizone soundfield reproduction techniques presented in this thesis are based on 2-D (height-invariant) and 2.5D soundfield model, which focus on the control of the desired multizone soundfield over a planar region. The extension to 3-D model will be of great value, since it would enable an enhanced immersive acoustic experience for a listener, as well as the possibility of free movements within a 3-D reproduction region for the users. However, 3-D soundfield reproduction methods generally require the placement of a large number of loudspeakers on a sphere that encloses the desired reproduction region. In [162], the authors applied the functional analysis framework and proposed the design of placing multiple circular loudspeaker arrays only over the 3-D desired reproduction region. Non-spherical and non-uniform loudspeaker placement is well fitted in the devised system and thereby allowing a flexible arrangement of the loudspeaker array. More investigations are still highly desirable on the reduction of the required number of loudspeakers for an accurate 3-D soundfield reproduction.
• **Real-time Implementation on Graphical Processing Units**

The real-time implementation of the proposed multizone soundfield reproduction techniques is very challenging due to their high computational demands. The usage of Central Processing Units (CPUs) for computation may not meet the real-time requirement. Recently, Graphics Processing Units (GPUs) have drawn increasing attention for data-intensive applications because they offer high parallelism with up to hundreds of processing cores compared to CPUs [163]. Spatial multizone soundfield reproduction is highly suitable for parallel processing, as different frequencies are independent. Therefore, achieving an efficient implementation of multizone soundfield reproduction using GPUs could be an interesting research direction.

• **Effect of the Listener on the Soundfield**

In this work, we make an assumption that the sound recordings from microphones with the listener being absent are considered to be the same with the sound perceived by the listener at the specified location. However, this assumption may not be strictly valid in reality. In practice, when a user enters the pre-defined reproduction region of a sound reproduction system, the reproduced soundfield is scattered by the user’s head. It leads to a decrease in the reproduction accuracy of the proposed system to some extent. At present, soundfield reproduction around arbitrary scattering objects have not been analytically studied yet. How we can assess the impact of a random scatterer on the reproduction of the desired soundfield is an open problem. It would also be interesting to include the subjective listening test (i.e. mean opinion score test), which obtains the multiple users’ view of the quality of the perceive multizone sound, as a evaluation measure for the multizone sound system.

• **2-D Multizone Soundfield Reproduction with HRTF**
7.2. FUTURE RESEARCH DIRECTIONS

A head-related transfer function (HRTF) is a response that can be used to synthesize a binaural sound for characterizing how an ear receives a sound from a virtual point in space [164]. Recently, the authors in [165] applied the spectral elevation cues of HRTF to the existing WFS system and created the sensation of elevated virtual sources within the specified control region. Listening tests were carried out in [166] for assessing listeners’ discrimination capability between sources located at different elevation angles. It would be of great value to explore the possibility to deliver slight elevation effects in multizone soundfield reproduction with HRTF using a 2-D loudspeaker array in the future. How can we combine the spectral filtering that produces the sensation of source elevation with the existing 2-D multizone sound systems that control the azimuth angle of the reproduced wavefield? How can we create the sensation of various elevation angles for different zones of interest to match the multizone settings more properly? These problems could be interesting to explore.

- Microphone Array Self-Localization

The microphone array feedback is essential to the reverberation equalization for multizone soundfield reproduction methods proposed in this thesis. Note that the proposed sound rendering systems allow a flexible placement of the microphone measurement and the location of the randomly-placed microphones needs to be informed. However, it can be cumbersome to conduct in practice. In [167], the author proposed a mechanism for localizing a microphone array when the location of sound sources in the environment is known. The approach is built on the spatial observability function (SOF) based microphone array integration technique. In [168] [169], the authors present a method for microphone localization in adhoc microphone arrays based on time-of-arrival (TOA) measurements.
from spatially distributed acoustic events or a moving acoustic source. Such a system does not require the knowledge of sound source locations. However, a particular periodic click waveform needs to be played out through speakers and the authors suggest that a narrowband signal with a Gaussian envelope achieves the best performance in practice using smartphone speakers and microphones. Is it possible to use existing filtered signals for the desired multizone soundfield reproduction, so that the microphone calibration process can be implemented automatically and in parallel with the sound rendering process? In addition, the TOAs of the source signals are required to be manually labeled in the recorded data. However, peak picking might be a difficult task in the practical implementation when room echoes are involved. A more efficient microphone array calibration is yet to be explored.
7.2. FUTURE RESEARCH DIRECTIONS
Bibliography


190


[80] L. J. Brannmark, A. Bahne, and A. Ahlen, “Compensation of loudspeaker room responses in a robust mimo control framework,”


[137] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and


