

# Theory and Design of Spatial Active Noise Control Systems

Hanchi Chen

Bachelor of Engineering (Hons 1)  
Australian National University

September 2017

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY  
OF THE AUSTRALIAN NATIONAL UNIVERSITY



Research School of Engineering  
College of Engineering and Computer Science  
The Australian National University

©Hanchi Chen 2017  
All Rights Reserved



# Declaration

The contents of this thesis are the results of original research and have not been submitted for a higher degree to any other university or institution. Much of this work has either been published or submitted for publications as journal papers and conference proceedings. Following is a list of these papers.

## Journal Publications

- H. Chen, T. D. Abhayapala, and W. Zhang, “Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis,” *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015.
- H. Chen, T. D. Abhayapala, P. N. Samarasinghe, and W. Zhang, “Direct-to-reverberant energy ratio estimation using a first order microphone,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, PP. 226–237, Feb 2017.
- P. N. Samarasinghe, T. D. Abhayapala, and H. Chen, “Estimating the Direct-to-Reverberant Energy Ratio Using a Spherical Harmonics Based Spatial Correlation Model”, in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 25, no. 2, PP. 310–319, Feb 2017.

## Conference Proceedings

- H. Chen, T. D. Abhayapala, and W. Zhang, “3D sound field analysis using circular higher-order microphone array,” in *Proc. 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 1153–1157.
- H. Chen, P. N. Samarasinghe, T. D. Abhayapala, and W. Zhang, “Estimation of the direct-to-reverberant energy ratio using a spherical microphone array.,”

in *Proc. ACE Challenge Workshop, a satellite event of WASPAA, New Paltz, NY, USA*, Oct 2015.

- H. Chen, P. N. Samarasinghe, T. D. Abhayapala, and W. Zhang, “Spatial noise cancellation inside cars: Performance analysis and experimental results,” in *Proc. 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.
- H. Chen, P. N. Samarasinghe, and T. D. Abhayapala, “In-car noise field analysis and multi-zone noise cancellation quality estimation,” in *Proc. 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2015, pp. 773–778.
- H. Chen, J. Zhang, P. N. Samarasinghe, and T. D. Abhayapala, “Evaluation of spatial active noise cancellation performance using spherical harmonic analysis,” in *Proc. 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2016, pp. 1–5.
- H. Chen, T. D. Abhayapala, and W. Zhang, “Enhanced sound field reproduction within prioritized control region,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 2014*, vol. 249, no. 3, pp. 4055–4064, Nov 2014.

The following papers are also results from my Ph.D. study, but not included in this thesis:

## Conference Proceedings

- G. Dickins, H. Chen and W. Zhang, “Soundfield control for consumer device testing”, in *Proc. 9th International Conference on Signal Processing and Communication Systems (ICSPCS’2015)*, Cairns, Australia, 2015.

The research work presented in this thesis has been performed jointly with Prof. Thushara D. Abhayapala, Dr. Wen Zhang and Dr. Prasanga Samarasinghe. Approximately 80% of this work is my own.

Hanchi Chen

Research School of Engineering  
The Australian National University  
Canberra ACT 2601  
January 2017



# Acknowledgments

Without the support of the many colleagues and friends, this work would have never been complete. I would like to acknowledge and thank each of the following.

- First and foremost, my supervisors, Prof. Thushara Abhayapala and Dr. Wen Zhang, for their professional guidance and consistent encouragement. Special thanks goes to Thushara, who had provided me with knowledge and experience not only in research, but also in many other aspects of life.
- Dr. Prasanga Samarasinghe, who had provided suggestions on many research problems, and helped in the writing and editing of many papers.
- Dr. Glenn Dickins for inviting me to visit the Dolby Labs, and sharing with me his extensive knowledge on every aspect of audio.
- The Australian National University, for the PhD scholarship and the funding and assistance for my patent application.
- My fellow students in the Applied Signal Processing Group, specially Jing, Yurui, Xiang and Aimee for their true friendship.
- Mr. Xianjun Zhen and Mr. Erasmo Scipione for providing technical support and electronics parts for my experiments.
- Mr. Yuki Mitsufuji for giving me the internship opportunity at Sony Japan.
- My parents for sending me to Australia in the first place, and supporting my study and life all these years.
- Finally, my girlfriend Mendy, for accompanying me throughout my PhD study and helping me out during the busiest days.



# Abstract

The concept of spatial active noise control is to use a number of loudspeakers to generate anti-noise sound waves, which would cancel the undesired acoustic noise over a spatial region. The acoustic noise hazards that exist in a variety of situations provide many potential applications for spatial ANC. However, using existing ANC techniques, it is difficult to achieve satisfying noise reduction for a spatial area, especially using a practical hardware setup. Therefore, this thesis explores various aspects of spatial ANC, and seeks to develop algorithms and techniques to promote the performance and feasibility of spatial ANC in real-life applications.

We use the spherical harmonic analysis technique as the basis for our research in this work. This technique provides an accurate representation of the spatial noise field, and enables in-depth analysis of the characteristics of the noise field. Incorporating this technique into the design of spatial ANC systems, we developed a series of algorithms and methods that optimizes the spatial ANC systems, towards both improving noise reduction performance and reducing system complexity.

Several contributions of this work are: (i) design of compact planar microphone array structures capable of recording 3D spatial sound fields, so that the noise field can be monitored with minimum physical intrusion to the quiet zone, (ii) derivation of a Direct-to-Reverberant Energy Ratio (DRR) estimation algorithm which can be used for evaluating reverberant characteristics of a noisy environment, (iii) propose a few methods to estimate and optimize spatial noise reduction of an ANC system, including a new metric for measuring spatial noise energy level, and (iv) design of an adaptive spatial ANC algorithm incorporating the spherical harmonic analysis technique. The combination of these contributions enables the design of compact, high performing spatial ANC systems for various applications.



# Contents

<b>Declaration</b>	<b>i</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Notations and Symbols</b>	<b>ix</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Motivation and scope . . . . .	3
1.2 Problem description . . . . .	7
1.3 Recent advancements in spatial ANC . . . . .	10
1.4 Thesis outline . . . . .	11
<b>2 Background: Spherical harmonic analysis and synthesis of sound fields</b>	<b>15</b>
2.1 Spherical harmonic expansion of a sound field . . . . .	15
2.2 Properties of the spherical harmonic expansion . . . . .	17
2.2.1 Recurrent property of associated Legendre functions . . . . .	17
2.2.2 Addition theorem . . . . .	18
2.2.3 Rotation of spherical harmonics . . . . .	20
2.2.4 Relationship between first order spherical harmonics and particle velocity . . . . .	20
2.2.5 Real-valued spherical harmonics . . . . .	23
2.3 Spatial sound recording and synthesis using spherical harmonic expansion . . . . .	24
2.3.1 Spatial sound recording using spherical microphone array . . .	24
2.3.2 Spatial sound recording using non-spherical microphone array	25
2.3.3 Spatial sound synthesis based on mode matching . . . . .	26

<b>3 Planar microphone array apertures for 3D spatial sound field analysis</b>	<b>29</b>
3.1 Introduction . . . . .	29
3.2 First order microphones for sound field acquisition . . . . .	32
3.2.1 General expression for first order microphones . . . . .	32
3.2.2 Sampling on a plane . . . . .	33
3.3 Array configuration . . . . .	34
3.3.1 Calculation of harmonic coefficients . . . . .	34
3.3.2 Discrete sensor placement: sampling of continuous aperture . .	38
3.3.3 Array design procedure . . . . .	41
3.3.4 Comments . . . . .	42
3.4 Error analysis . . . . .	43
3.4.1 Differential microphone approximation . . . . .	43
3.4.2 Spatial sampling and spatial aliasing . . . . .	44
3.5 Design examples . . . . .	44
3.5.1 Hypothetical design example . . . . .	45
3.5.2 Array implementation . . . . .	49
3.6 Summary . . . . .	51
3.7 Related patents and publications . . . . .	52
<b>4 3D sound field analysis using circular higher order microphone array</b>	<b>53</b>
4.1 Introduction . . . . .	53
4.2 Sound field model . . . . .	54
4.3 Higher-order microphone array . . . . .	55
4.3.1 Higher-order microphone . . . . .	55
4.3.2 Continuous circular higher-order microphone array . . . . .	55
4.3.3 Solving for global coefficients . . . . .	57
4.3.4 Dimensionality analysis . . . . .	58
4.4 Simulation results . . . . .	58
4.5 Experimental results . . . . .	61
4.6 Summary . . . . .	62
4.7 Related Publications . . . . .	62

<b>5 Direct-to-reverberant energy ratio estimation using a first order microphone</b>	<b>63</b>
5.1 Introduction . . . . .	64
5.2 DRR estimation based on coherence measurements . . . . .	66
5.2.1 Representation of reverberant sound field . . . . .	66
5.2.2 Representation of DRR using coherence function . . . . .	68
5.2.3 Assumptions for the reverberant sound field . . . . .	69
5.2.4 Reverberant field estimation . . . . .	72
5.2.5 DRR estimation procedure . . . . .	73
5.3 Impact of parameters on DRR estimation . . . . .	74
5.3.1 Reverberation parameter . . . . .	74
5.3.2 Nearfield sound source . . . . .	76
5.4 Validation using ACE Challenge Database . . . . .	77
5.4.1 The ACE Challenge Database . . . . .	77
5.4.2 Algorithm setup . . . . .	78
5.4.3 Full band results . . . . .	80
5.4.4 Subband results . . . . .	81
5.4.5 Impact of noise on DRR estimation . . . . .	83
5.4.6 Estimated parameters from the ACE Evaluation Dataset . .	86
5.5 Summary . . . . .	87
5.6 Related Publications . . . . .	87
5.7 Proof of Equation (5.22) . . . . .	87
<b>6 Methods for spatial ANC performance evaluation and optimization</b>	<b>89</b>
6.1 Introduction . . . . .	89
6.2 Enhanced sound field reproduction within prioritized control region .	91
6.2.1 Background . . . . .	91
6.2.2 Problem formulation . . . . .	92
6.2.3 Combined Least Mean Square Solution for sound field reproduction . . . . .	93
6.2.4 Simulation Results . . . . .	96
6.2.5 Observations and insights . . . . .	101
6.3 Evaluation of spatial active noise cancellation performance using acoustic potential energy . . . . .	102
6.3.1 Background . . . . .	102
6.3.2 Calculation of the acoustic potential energy . . . . .	103

6.3.3	Performance evaluation . . . . .	105
6.3.4	Result analysis . . . . .	106
6.3.5	Observations and insights . . . . .	110
6.4	In car spatial ANC performance analysis . . . . .	111
6.4.1	Background . . . . .	111
6.4.2	Problem Formulation . . . . .	112
6.4.3	Noise field characterization . . . . .	112
6.4.4	Residual noise level estimation . . . . .	114
6.4.5	Experiment on a single passenger seat . . . . .	115
6.4.6	Experiment with multiple passenger seats and limited loud-speaker output power . . . . .	120
6.4.7	Observations and insights . . . . .	126
6.5	Summary . . . . .	127
6.6	Related Publications . . . . .	128
<b>7</b>	<b>Spatial active noise cancellation system architectures</b>	<b>129</b>
7.1	Introduction . . . . .	129
7.2	Background theory . . . . .	131
7.2.1	Time domain multi-channel feed-forward ANC architecture . .	131
7.2.2	Frequency domain feed-forward ANC architecture . . . . .	132
7.3	Frequency domain feed-forward architecture for spatial ANC systems	136
7.3.1	Existing spatial ANC system based on circular harmonic transform . . . . .	136
7.3.2	Proposed spatial ANC system based on spherical harmonic transform . . . . .	137
7.4	Time domain feed-forward architecture for spatial ANC systems . . .	141
7.4.1	Time domain spherical harmonics representation of sound field	141
7.4.2	Spatial ANC architecture using time domain spherical harmonics analysis . . . . .	142
7.5	Experiment validation . . . . .	146
7.5.1	System setup . . . . .	146
7.5.2	Experiment results . . . . .	148
7.6	Summary . . . . .	155
<b>8</b>	<b>Conclusion and future works</b>	<b>157</b>
8.1	Conclusion . . . . .	157

8.2 Future works . . . . .	159
<b>Bibliography</b>	<b>163</b>



# Notations and Symbols

$\lceil \cdot \rceil$	ceiling operator
$\lfloor \cdot \rfloor$	floor operator
$[\cdot]^*$	complex conjugate of a matrix
$[\cdot]^T$	transpose of a matrix
$[\cdot]^H$	complex conjugate transpose of a matrix
$ \cdot $	Euclidean norm of a vector
$\ \cdot\ $	$\ell^2$ -norm of a vector
$\mathbf{A}^{-1}$	matrix psuedoinverse
$E\{\cdot\}$	Expectation operator
$Re\{\cdot\}$	real part
$Im\{\cdot\}$	imaginary part
$\delta\{\cdot\}$	Dirac delta function
$\delta_{nm}\{\cdot\}$	Kronecker delta function
$\{\ast\}$	linear convolution
$\mathcal{F}_N\{\cdot\}$	N-point Fast Fourier Transform
$\mathcal{F}_N^{-1}\{\cdot\}$	N-point Fast Fourier Transform
$\mathcal{F}^{-1}\{\cdot\}$	Inverse Fourier Transform
$i$	$\sqrt{-1}$
ANC	active noise cancellation
DRR	direct-to-reverberant energy ratio
DOA	direction of arrival
SNR	signal to noise ratio



# Chapter 1

## Introduction

### 1.1 Motivation and scope

A wide range of human activities generate unwanted noise. Acoustic noise is one of the most common hazards in the world. Exposure to acoustic noise causes discomfort and pain; long term exposure to excessive noise can also result in chronic effects to human health, especially hearing loss, which can limit one's ability to hear high frequency sounds and understand speech [1].

Methods to reduce excessive noise can be sorted into two categories: passive noise control, and active noise control. Passive noise control methods utilize sound absorbing materials, such as glasswool, acoustic foam, or other insulation materials to absorb the impinging noise (Fig. 1.1). Sometimes the material is cut into special geometries to enhance their sound absorption capabilities. The overall noise isolation capability depends on a number of factors, including sound frequency, material type, its thickness, and its geometry. In general, a common property of all passive noise isolation materials is that the sound absorption coefficient rises with sound frequency [2]. When the wavelength of the sound becomes larger than the thickness of the material, it becomes difficult for the material to absorb the sound. As a result, passive noise control systems perform well at higher frequencies, but their effectiveness reduces significantly at low frequencies. In many real-world scenarios, low frequency noise is dominant in the whole noise spectrum, in such cases, the passive noise control method becomes less effective [2].

The alternative method is active noise control. Active noise control systems rely on one or more loudspeakers, called “secondary sources”, which produce a sound wave whose magnitude is the same as the noise but is 180° out of phase, so that the

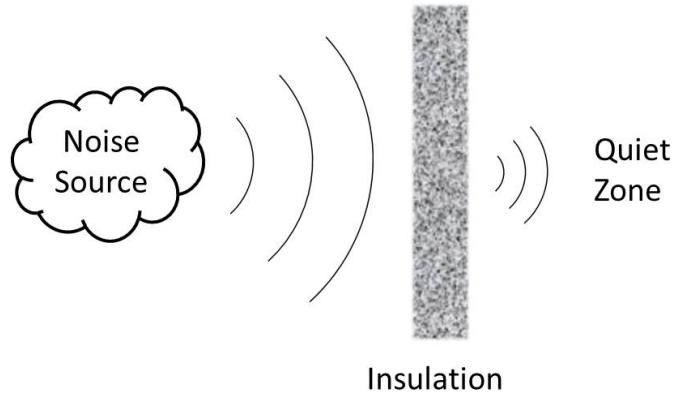


Figure 1.1: Passive noise control system.

two sound waves would cancel each other, thus reducing the noise level (Fig. 1.2). Contrary to the passive noise control strategy, the active noise control method works better at lower frequencies [2]. At lower frequencies (up to a few hundred Hz), the wavelength of the sound is longer, thus making it easier for the anti-noise signal to match with the unwanted noise.

The most commonly seen application of the active noise control technique is the active noise cancelling (ANC) headphones (Fig. 1.3). The ANC headphones typically employ a reference microphone, mounted on the outer surface of the headphone's housing. The reference microphone picks up the ambient noise, and sends the noise signal to a processing unit, which generates the anti-noise signals and plays it through the headphone driver along with the music signal [3]. In some designs, an additional error microphone is placed inside the ear cup to monitor the residual noise. It is also possible to use a feedback ANC structure, where the reference microphone is omitted, one such design is detailed in [3]. Noise cancelling headphones can yield reasonably good noise attenuation, partially due to the fact that the secondary loudspeaker and the error microphone are placed very close to the ear. According to [4], significant attenuation of sinusoidal noise signal can be achieved for frequencies up to 2 kHz. Another study on consumer ANC headphone performance [5] suggests that the noise reduction achievable by ANC headphones is typically between 10 – 25 dB, and the performance is highly dependant on the tightness of the wearing situation.

Although ANC headphones yield very good performance in terms of noise level attenuation, one of its disadvantages is that the user is required to constantly wear the headphone, which is inconvenient, or even impractical in many scenarios. In such

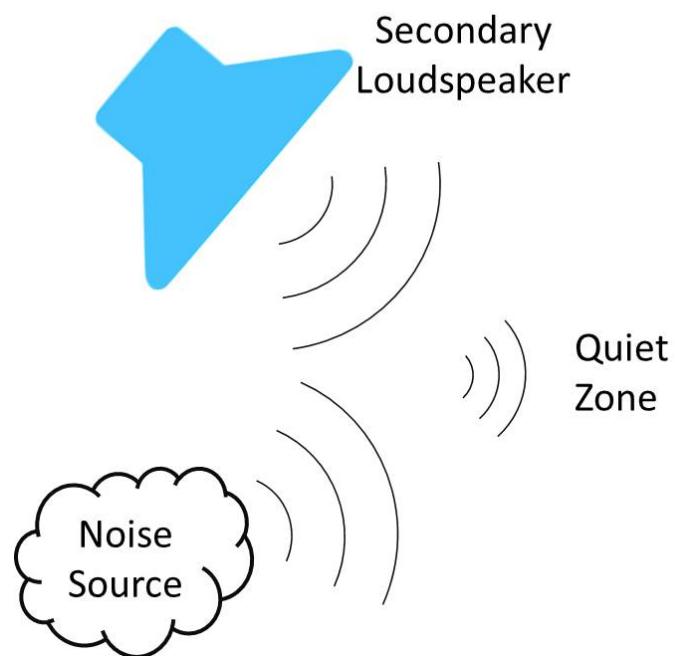


Figure 1.2: Active noise control system.

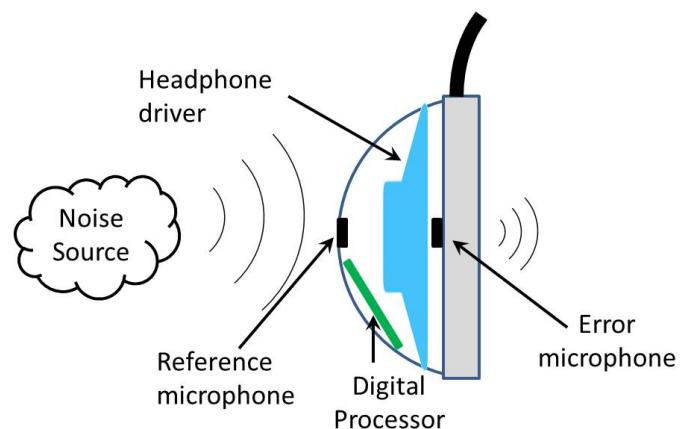


Figure 1.3: Structure of feed-forward active noise cancelling headphones.

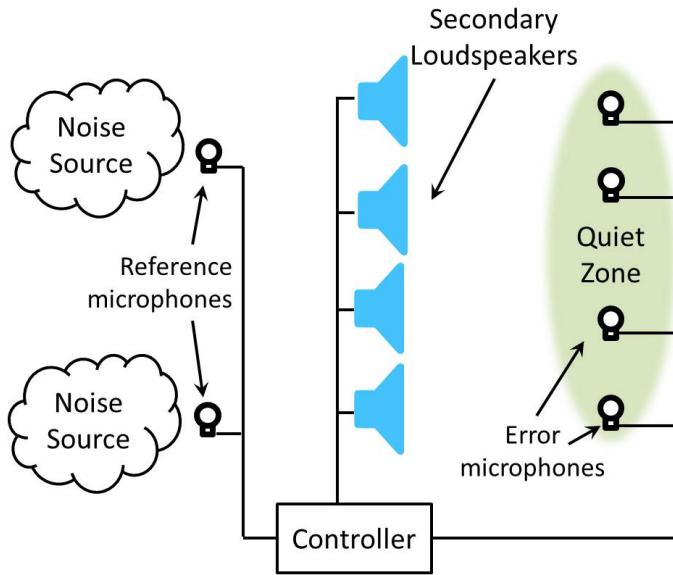


Figure 1.4: Feed-forward MIMO active noise cancelling system.

cases, it is desirable if the noise can be attenuated for a spatial area, such that people within the area can enjoy a noise-free acoustical environment. A well developed approach to achieve this goal is the Multiple-input-multiple-output (MIMO) ANC systems, or multi-channel ANC systems [6]. In these systems, multiple secondary loudspeakers are utilized to generate the anti-noise signals, while multiple error microphones are distributed in the quiet zone to monitor the residual noise level. For feed-forward systems, one or more reference microphones are placed close to the noise source to pick up the noise signal; for feedback systems, reference microphones are not needed [6]. Fig. 1.4 illustrates a feed-forward MIMO ANC system.

The MIMO ANC system has been successfully implemented to reduce the noise in environments such as vehicle cabins [7, 8] and rooms [9]. However, conventional MIMO ANC controllers minimize the sound pressure measured by the error microphones. Since the noise level is only known at the microphone locations, when a number of error microphones are randomly distributed inside the desired quiet zone, only the space in the proximity of each microphone can be expected to have significant noise reduction; in the area not covered by microphones, noise reduction cannot be guaranteed. One straightforward solution to this problem is to place a large number of microphones inside the quiet zone, however this approach greatly reduces the feasibility of MIMO ANC systems in real-life applications.

A potential way of overcoming this issue is to employ spatial sound analysis techniques, where the noise sound captured by a microphone array is transformed

into another domain, which results in a more accurate representation of the spatial noise field. One of such techniques is the spherical harmonic analysis [10], where the noise field inside a spherical region is decomposed into a series of spherical harmonic functions. This technique allows accurate representation and reconstruction of the noise field, which makes it possible to perform ANC over a continuous space, rather than at a number of sampling points. Furthermore, the transformation into spherical harmonic domain allows in-depth analysis of the noise field, such as Direction-of-Arrival Estimation (DOA) [11] and Direct-to-Reverberant Ratio (DRR) estimation [12]. However, in order to perform the spherical harmonic analysis, the error microphones need to be arranged in specific geometries, typically in a spherical arrangement [13, 14].

In general, spherical microphone arrays designed for spherical harmonic analysis of sound field can be divided into two categories: rigid sphere topology, and open sphere topology. In the former case, the microphones are mounted on a rigid sphere baffle whose radius is the same as that of the region of interest; while in the open sphere case, microphones are placed on the surface of the region of interest, without the use of a rigid baffle. However, for the open baffle topology, the microphone array may suffer from ill-conditioning, due to the inherent properties of spherical Bessel functions [15]. One way to overcome this is to use two concentric spherical arrays with similar radius [15–18]. Although the open sphere topology is easier to implement than the rigid sphere topology when the region size is large, the region of interest is still fully surrounded by microphones, which limits its feasibility in practical ANC applications.

From the discussion above, an unsolved problem regarding the active noise control technique can be summarized as follows:

*How to attenuate a complicated noise field over a space using active noise cancellation strategies, especially with a hardware system that's feasible for practical applications?*

## 1.2 Problem description

We breakdown this problem into a number of sub-tasks. As shown in Fig. 1.5, the spatial ANC problem can be divided into two major components: modelling of the noise field, and generation of anti-noise signal.

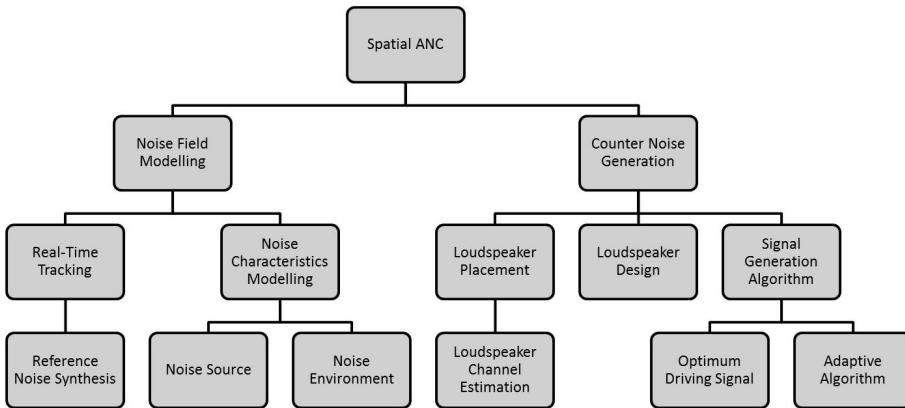


Figure 1.5: Breakdown of the spatial active noise cancellation problem.

Noise field modelling is about acquiring information about the noise field, so that an ANC algorithm can use this information to generate suitable anti-noise signals to cancel the noise. This can be further divided into two elements, namely the real-time tracking of the noise, and characterization of the noise field. The real-time tracking of the noise happens while the ANC system is online, it provides the ANC system with the instantaneous noise field information, and measures the noise attenuation achieved by the ANC system, so that the system can quickly respond to changes of the noise field and minimize the residual noise. A number of sensors, typically microphones, are usually employed to keep track of the noise in real time. The number and position of these sensors play a key role in determining the performance of an ANC system. Although distributing a large number of sensors over the entire quiet zone would provide very complete information of the noise field, practical applications demand for more compact and economic sensing solutions. In addition, in some ANC systems, the reference noise is synthesized based on measurement of noise source movement (such as engine rotation) and some prior knowledge of the noise composition (such as harmonic components), this can also be categorized into real-time noise tracking.

On the other hand, modelling of the noise characteristics can be done without a functional ANC system, it is about analyzing the nature of the spatial noise, such as its spectrum, direction of arrival, and spatial dimensionality. These information help to determine whether or not a given noise environment is suitable for spatial ANC, and whether the characteristics of the noise field can be exploited to simplify the design complexity of the spatial ANC system. Both the noise source itself and a reverberant environment contribute to the characteristics of the spatial noise field. Modelling them separately would provide more insights into the noise field.

Since the goal of every ANC system is to use a suitable sound wave to cancel the noise, generation of the optimal anti-noise signal is critical to the performance of a spatial ANC system. The position of the secondary loudspeakers, and consequently the sound field they can produce inside the quiet zone, play an important role in determining the anti-noise signal to be played. Badly positioned loudspeakers may result in very high driving signals, causing excessive noise level outside the quiet zone without achieving any significant noise attenuation inside the quiet zone, and may damage the loudspeakers themselves; on the other hand, a few well-placed loudspeakers may be able to minimize the noise level over a large region with very small output power. Once the loudspeakers are placed, it is then critical to accurately measure, and keep track of, the acoustic signal channels between each loudspeaker and the quiet zone, as inaccurate channel information can cause instability of the ANC system. Therefore studying the placement of loudspeakers and estimating the loudspeaker channels is very necessary for designing compact and efficient spatial ANC systems.

Designing loudspeakers suitable for spatial ANC is also important, and some of the design goals are different from that of consumer loudspeaker products. While consumer products aim for wide and flat frequency response, strong and deep bass, and an attractive design, loudspeakers designed for ANC purposes should have characteristics such as low harmonic distortion (especially at low frequencies), high sensitivity, and good power handling capabilities, combined with a small form factor. The frequency range can be just as wide as the target noise frequency band, and a flat response curve is not necessary, since the adaptive filter will act as an equalizer automatically. Although in some cases, the loudspeakers designed for music listening have to be employed for ANC purposes, such as in-car noise cancellation, it is still desirable to keep in mind of the properties that make a good ANC loudspeaker while selecting speakers for the ANC system.

The active noise control algorithm governs how the anti-noise signal is generated, and depending on the optimization criteria, each algorithm would result in different noise attenuation level at each position within the quiet zone. The Least-Mean-Square error algorithm, commonly used in existing multi-channel ANC systems, may not result in the best performance in a spatial noise control application. Utilizing the latest spatial sound field analysis techniques, more advanced ANC algorithms may be developed.

Many active noise control systems utilize an adaptive algorithm to estimate the noise channel, as well as generate the driving signals for loudspeakers. The use of

adaptive algorithm enables an ANC system to quickly respond to changes in the noise signal, and continuously generate the anti-noise signals most suited for the current noise signals. To yield the optimum spatial noise attenuation, it is necessary to incorporate the spatial sound processing algorithms, especially the optimum spatial ANC algorithms discussed above into the real-time adaptive algorithm, so that the adaptive ANC system can generate the optimum anti-noise signals in real time.

Motivated by the above problem, we develop a series of techniques to improve the performance and feasibility of spatial ANC systems, as well as methods to analyze the spatial noise field, with the aim of aiding development and evaluation of spatial ANC systems.

### 1.3 Recent advancements in spatial ANC

In recent years, researchers have made significant progress in the field of spatial ANC. In [19], the authors investigated the problem of cancelling the noise propagating through an open window. In this work, a number of pre-set filters are used to cancel the spatial noise through the window, an additional algorithm recognizes the nature of the impinging noise, and selects the most suitable filter for the ANC system. This open-loop ANC system is further investigated in [20]. Further more, a mixed-error approach to reduce adaptive filter complexity for the open window ANC application has been presented in [21].

ANC systems aimed to reduce in-car noise are investigated in [22–25] and [26], with [22–24] specifically target to control the road noise during driving. The results in [22] show that up to 8 dB of noise reduction is achievable for lower frequencies. In [27], a MIMO ANC system is deployed in the master cabin of a yacht to cancel the noise of the diesel engine, yielding a 23% reduction in the noise loudness.

Improvements to the adaptive algorithms for ANC systems have also been proposed. The performance of multiple subband MIMO adaptive algorithms were studied in [28]. A new feedback adaptive ANC algorithm with faster convergence rate was proposed in [29], where an adaptive notch filter is used to track the frequency components of the noise. Jihui et. al. proposed a feedback adaptive spatial ANC algorithm [30], which is capable of cancelling the impinging noise over a spatial region. For 2D spatial ANC, Spors et. al. [31] proposed a feed-forward adaptive algorithm based on circular harmonic transform, this algorithm is able to significantly reduce the computational complexity of massive ANC systems.

## 1.4 Thesis outline

This thesis is organized into 8 chapters. The key contributions of each chapter are:

### **Chapter 2 - Background: Spherical harmonic analysis and synthesis of sound fields**

This chapter briefly reviews the theory of spherical harmonic analysis for spatial sound, and presents a number of properties of the spherical harmonics. The techniques for spatial sound recording and synthesis using spherical microphone / loudspeaker arrays are also briefly reviewed. These properties and techniques are used later in the thesis for the development of various theories and techniques. In this chapter, we also derive the mathematical relationship between the first order spherical harmonics and acoustic particle velocity. Although this relationship has been assumed to exist and used in the literature, its mathematical proof has not been proposed, to the best of our knowledge.

### **Chapter 3 - Planar microphone array apertures for 3D spatial sound field analysis**

In this chapter, we present a novel method to capture 3D spatial sound fields using a 2D planar microphone array. In general, it is assumed that capturing a 3D sound field requires the use of a microphone array with 3D geometry. Here, we explain the reason of this requirement by investigating the properties of the spherical harmonics. We also show that by exploiting a property of the associated Legendre functions, it is possible to capture full 3D sound field using first order differential microphones, placed on a 2D plane. A planar microphone array structure consisting of multiple concentric circular arrays is proposed, as well as an algorithm to calculate the spherical harmonic coefficients of the sound field using this array structure.

### **Chapter 4 - 3D sound field analysis using circular higher order microphone array**

This chapter develops a method to use circular higher order microphone arrays placed on a 2D plane to capture 3D spatial sound. We use the spherical harmonic addition theorem to derive a method for calculation spherical harmonic coefficients

of a large sound field, using the local spherical harmonic coefficients captured by each higher order microphone. Compared to the method developed in Chapter 3, this method requires significantly smaller number of microphone units, due to the use of higher order microphones. This method can be seen as a generalization of the method proposed in Chapter 3.

## **Chapter 5 - Direct-to-reverberant energy ratio estimation using a first order microphone**

This chapter presents an algorithm for DRR estimation using a first order microphone system, which helps to characterize the noise environment, and the relevant room acoustics. Using the relationship between first order spherical harmonics and the acoustic particle velocity developed in Chapter 2, we derive an expression for modelling certain characteristics of the reverberation that are related to DRR estimation. Based on the estimated reverberation characteristics, we use the coherence function between sound pressure and particle velocity to estimate DRR. All the required data can be obtained using a single first order microphone. The proposed method addresses the overestimation problem observed in a previous DRR estimation algorithm.

## **Chapter 6 - Methods for spatial ANC performance evaluation and optimization**

In Chapter 6, we develop a series of methods to estimate and optimize spatial noise control performance. First, in Section 6.2, we present a method to maximize the noise reduction at certain high priority sub-regions within the global quiet zone, this technique is particularly useful when the number of secondary loudspeakers is insufficient. Then, in Section 6.3, we propose a new metric for measuring spatial noise level. This metric provides a more robust and accurate representation of the average noise energy over space compared to existing metrics. Finally, in Section 6.4, we use the proposed metric to develop a method for estimating the potential ANC performance for a given noise environment and loudspeaker setup. This method is then used to estimate in-car ANC performance for both single and multiple seats scenarios.

## **Chapter 7 - Spatial Active Noise Cancellation System Architectures**

In this chapter, we present a novel adaptive algorithm designed for spatial ANC. The proposed algorithm is based on the conventional multi-channel feed-forward adaptive algorithm, but incorporates the spherical harmonic transform, thereby achieving superior spatial ANC performance compared to existing multi-channel ANC algorithms. We present both frequency domain and time domain implementation of the algorithm, which are mathematically equivalent but each feature their own advantages. An experimental ANC system utilizing the proposed time domain algorithm is implemented in our laboratory, and we use this system to investigate the impact of secondary loudspeaker placement on the performance of the ANC system.

## **Chapter 8 - Conclusion and future works**

Chapter 8 concludes this thesis, as well as discusses a number of future works which would further improve the performance and practical feasibility of spatial active noise control systems.



# Chapter 2

## Background: Spherical harmonic analysis and synthesis of sound fields

**Overview:** This chapter provides a brief overview of the theory and techniques related to spherical harmonic analysis. We first introduce the mathematical expressions of the spherical harmonic expansion, and show how these expressions can be used to express a spatial sound field. Then, we present a number of special properties of the spherical harmonics. This is followed by a review of the techniques for recording spatial sound using microphone arrays, as well as synthesizing spatial sound using loudspeaker arrays, both of which are based on spherical harmonic analysis. The techniques described in this chapter form a foundation for the rest of the thesis.

### 2.1 Spherical harmonic expansion of a sound field

Three dimensional (3D) sound field decomposition based on spherical harmonic analysis has become a popular tool in the field of array signal processing. Applications of this technique can be found in both acoustic and radio frequency (RF) areas, such as spatial filtering and beamforming [32–38], room acoustic modeling [39–42], sound field analysis [10, 43], spatial sound field reproduction [44–46], source localization [11, 39, 47], active noise control [31, 48, 49], and phase mode processing for antenna arrays [50]. While other spatial sound field representation and reproduction techniques such as wave front synthesis [51] and plane wave decomposition [52] each have their own advantages, spherical harmonic analysis based 3D sound field

decomposition reveals the underlying characteristics of the sound field, thus allowing high accuracy manipulation and analysis of the sound field, therefore for this thesis, we choose to use spherical harmonic analysis as the fundamental tool for the development of the theories.

The essential idea of spherical harmonic analysis of a sound field is to use the weighted sum of a set of orthogonal basis functions to describe the pressure field of propagating sound. These functions, known as spherical harmonics, are solutions to the Helmholtz wave equation in the 3D space for representing the propagation modes of a sound wave.

The spherical harmonics expansion of a sound field is divided into two cases: the interior field expression, and the exterior field expression. The former is used to describe the wave field within a spatial region with no sound source inside, and all impinging sound waves are due to sources outside the region; the latter is used for the situations where the sound sources are positioned within a limited area, and the region of interest is defined as the space enclosing the source area.

In this work, only the interior field problem is considered, therefore we only describe the spherical harmonics expansion for the interior field case in this section.

Consider a sound field within a source free region, the sound pressure at a point  $(r, \theta, \phi)$  with respect to the origin  $\mathcal{O}$  can be expressed as [46]

$$P(r, \theta, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{lm}(k) j_l(kr) Y_{lm}(\theta, \phi) \quad (2.1)$$

where  $C_{lm}(k)$  are spherical harmonic coefficients,  $k = 2\pi f/c$  is the wave number,  $f$  is the frequency,  $c$  is the speed of sound propagation,  $j_l(kr)$  is the  $l$ th order spherical Bessel function of the first kind, and  $Y_{lm}(\theta, \phi)$  are the spherical harmonics, defined by

$$Y_{lm}(\theta, \phi) = \mathcal{P}_{l|m|}(\cos \theta) E_m(\phi) \quad (2.2)$$

where

$$\mathcal{P}_{l|m|}(\cos \theta) \triangleq \sqrt{\frac{(2l+1)}{2} \frac{(l-|m|)!}{(l+|m|)!}} P_{l|m|}(\cos \theta), \text{ and} \quad (2.3)$$

$$E_m(\phi) \triangleq (1/\sqrt{2\pi}) e^{im\phi} \quad (2.4)$$

are the normalized associated Legendre functions and normalized exponential functions, respectively;  $P_{l|m|}(\cos \theta)$  are the associated Legendre functions.

$Y_{lm}(\vartheta, \varphi)$  has the orthogonal property

$$\int_0^\pi \int_0^{2\pi} Y_{lm}(\theta, \phi) Y_{l'm'}^*(\theta, \phi) \sin(\theta) d\theta d\phi = \delta_{l-l', m-m'}, \quad (2.5)$$

where  $\delta_{l,m}$  is the two dimensional Dirac Delta function. The orthogonal property of the spherical harmonics is very useful in simplifying the mathematical expressions related to spatial sound, this property will be utilized later in this thesis in the derivation of many results.

It can be seen from the decomposition (2.1) and the expression (2.2) that the spherical bessel function  $j_l(kr)$  governs the radial and frequency dependant component of the basis functions, while  $\mathcal{P}_{l|m|}(\cos \theta)$  and  $E_m(\phi)$  govern the elevation and azimuth components, respectively. Due to the low pass nature of spherical Bessel functions, spherical harmonics of higher order  $l$  has very little energy when the value of  $kr$  is lower than a certain threshold. Therefore, a common practice is to truncate the infinite summation in (2.1) at a maximum order  $l = L$ , such that the finite summation provides an accurate approximation of the sound field, thus (2.1) can be approximated as

$$P(r, \theta, \phi, k) \approx \sum_{l=0}^L \sum_{m=-l}^l C_{lm}(k) j_l(kr) Y_{lm}(\theta, \phi). \quad (2.6)$$

A rule of thumb for determining the upper bound  $L$  is given by [53–55]

$$L = \lceil \frac{ekr}{2} \rceil, \quad (2.7)$$

where  $e$  is the natural exponential. Using this truncation, the number of spherical harmonics required to approximate any sound field of a certain radius and frequency is limited to  $(L + 1)^2$ .

## 2.2 Properties of the spherical harmonic expansion

### 2.2.1 Recurrent property of associated Legendre functions

A recurrent relationship between the associated Legendre function and its first order derivative is given by [56]

$$(x^2 - 1) \frac{dP_{l|m|}(x)}{dx} = nx P_{l|m|}(x) - (|m| + l) P_{(l-1),|m|}(x). \quad (2.8)$$

In the special case where  $x = 0$ , (2.8) can be simplified to

$$P'_{l|m|}(0) = (|m| + l) P_{(l-1),|m|}(0), \quad (2.9)$$

which indicates that the first order derivative of the associated Legendre functions at  $x = 0$  can be directly calculated from the same functions of a lower order.

By taking the derivative of (2.3) and setting  $\cos \theta = 0$ , expressing  $P'_{l|m|}(0)$  using (2.9) and expressing  $P_{(l-1),|m|}(0)$  with  $\mathcal{P}_{l-1|m|}(0)$  using (2.3), we derive the following relationship for the normalised associate Legendre functions

$$\mathcal{P}'_{l|m|}(0) = \sqrt{\frac{(2l+1)(l^2-m^2)}{(2l-1)}} \mathcal{P}_{(l-1)|m|}(0), \quad (2.10)$$

which illustrates a relationship between the normalized associate Legendre functions and their first order derivatives. We will show in Chapter 3 that this property can be exploited to develop compact microphone arrays for spatial sound recording.

### 2.2.2 Addition theorem

The addition theorem describes the relationship of spherical harmonic coefficients with respect to two different coordinate systems. It shows that each spherical harmonic coefficient with respect to one coordinate system can be expressed as a weighted sum of the coefficients with respect to another coordinate system. The addition theorem also has an interior field variant as well as an exterior field variant. In this section, we briefly outline the addition theorem for the interior field case.

In addition to the coordinate system with origin  $\mathcal{O}$ , we define a new coordinate system with its origin  $\mathcal{O}'$  located at  $\mathbf{R} = (R, \vartheta, \varphi)$  with respect to  $\mathcal{O}$ . The two coordinate system are defined such that they have the same orientation, i.e., the  $x$ ,  $y$  and  $z$  axes of the two coordinate systems point in the same directions.

The sound field with respect to  $\mathcal{O}'$  can also be decomposed using spherical harmonics, but using another set of weighting coefficients  $B_{\nu\mu}$ , such that the sound pressure at point  $(r', \theta', \phi')$  with respect to  $\mathcal{O}'$  can be expressed by

$$P(r', \theta', \phi', k) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} B_{\nu\mu}(k) j_{\nu}(kr') Y_{\nu\mu}(\theta', \phi'). \quad (2.11)$$

The relationship between  $C_{lm}$  and  $B_{\nu\mu}$  can be described by the spherical harmonic addition theorem [57]. The relationship can be written as [58]

$$B_{\nu\mu} = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm} \widehat{S}_{l\nu}^{m\mu}(\mathbf{R}), \quad (2.12)$$

where

$$\widehat{S}_{l\nu}^{m\mu}(\mathbf{R}) = 4\pi i^{\nu-l} \sum_{\ell=|\mu-m|}^{l+\nu+1} i^\ell (-1)^{2m-\mu} j_\ell(kR) Y_{\ell(\mu-m)}^*(\vartheta, \varphi) W, \quad (2.13)$$

$$W = \sqrt{\frac{(2l+1)(2\nu+1)(2\ell+1)}{4\pi}} W_1 W_2. \quad (2.14)$$

Here,  $W_1$  and  $W_2$  denote Wigner 3-j symbols, with

$$W_1 = \begin{pmatrix} l & \nu & \ell \\ 0 & 0 & 0 \end{pmatrix}, \quad W_2 = \begin{pmatrix} l & \nu & \ell \\ m & -\mu & \mu - m \end{pmatrix}. \quad (2.15)$$

It can be seen that by substituting (2.12) into (2.11), one can derive the sound pressure decomposition of a given point with respect to  $\mathcal{O}'$  using the spherical harmonic coefficients with respect to  $\mathcal{O}$ .

Equation (2.12) can be conveniently represented in matrix form, as

$$\mathbf{B} = \widehat{\mathbf{S}} \mathbf{C}, \quad (2.16)$$

where  $\mathbf{C} = [C_{00} \ C_{11} \ C_{10} \ \dots \ C_{LL}]^T$  and  $\mathbf{B} = [B_{00} \ B_{11} \ B_{10} \ \dots \ B_{VV}]^T$ .  $\widehat{\mathbf{S}}$  is the translation matrix that maps the coefficients  $\mathbf{C}$  to the coordinate system  $\mathcal{O}'$ .  $\widehat{\mathbf{S}}$  consists of all the  $\widehat{S}_{l\nu}^{m\mu}(\mathbf{R})$  needed to translate  $\mathbf{C}$  into  $\mathbf{B}$ , the orders of  $\widehat{S}_{l\nu}^{m\mu}(\mathbf{R})$  are arranged in correspondence with  $\mathbf{B}$  and  $\mathbf{C}$ , thus  $\widehat{\mathbf{S}}$  can be written as [58]

$$\widehat{\mathbf{S}} = \begin{bmatrix} \widehat{S}_{00}^{00} & \widehat{S}_{11}^{00} & \widehat{S}_{10}^{00} & \dots & \widehat{S}_{LL}^{00} \\ \widehat{S}_{00}^{11} & \widehat{S}_{11}^{11} & \widehat{S}_{10}^{11} & \dots & \widehat{S}_{LL}^{11} \\ \widehat{S}_{00}^{10} & \widehat{S}_{11}^{10} & \widehat{S}_{10}^{10} & \dots & \widehat{S}_{LL}^{10} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \widehat{S}_{00}^{VV} & \widehat{S}_{11}^{VV} & \widehat{S}_{10}^{VV} & \dots & \widehat{S}_{LL}^{VV} \end{bmatrix}. \quad (2.17)$$

For a given maximum order  $L$ , there are a total number of  $(L+1)^2$  spherical harmonics available, thus the size of  $\widehat{\mathbf{S}}$  becomes  $(V+1)^2$  by  $(L+1)^2$ .

### 2.2.3 Rotation of spherical harmonics

In some applications, it is convenient to perform a rotation to the coordinate system. In this case, the spherical harmonics of the sound field would also need to be rotated such that they still represent the same sound field. Here, we outline a method to perform such rotation to the spherical harmonic coefficients.

The rotation can be performed through a transformation matrix  $\mathbf{M}$ , so that the original and transformed coefficients can be expressed as

$$\begin{bmatrix} \beta_{00} \\ \beta_{11} \\ \beta_{10} \\ \beta_{1,-1} \\ \vdots \end{bmatrix} = \begin{bmatrix} M_{00}^{00} & M_{11}^{00} & M_{10}^{00} & M_{1,-1}^{00} & \dots \\ M_{00}^{11} & M_{11}^{11} & M_{10}^{11} & M_{1,-1}^{11} & \dots \\ M_{00}^{10} & M_{11}^{10} & M_{10}^{10} & M_{1,-1}^{10} & \dots \\ M_{00}^{1,-1} & M_{11}^{1,-1} & M_{10}^{1,-1} & M_{1,-1}^{1,-1} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} C_{00} \\ C_{11} \\ C_{10} \\ C_{1,-1} \\ \vdots \end{bmatrix}, \quad (2.18)$$

where  $\beta_{lm}$  and  $C_{lm}$  represent the spherical harmonic coefficients after and before rotation, respectively. The values of  $M_{lm}^{l'm'}$  can be calculated using numerical integration [59],

$$M_{lm}^{l'm'} = \int_s Y_{l'm'}(\mathcal{R}\mathbf{s}) Y_{lm}^*(\mathbf{s}) d\mathbf{s} \quad (2.19)$$

where  $\mathcal{R}$  denotes the rotation matrix for the spherical coordinates.

### 2.2.4 Relationship between first order spherical harmonics and particle velocity

The term “particle velocity” is commonly used to describe the velocity component of impinging sound, it refers to the velocity of particle movement in the medium during wave propagation. In the literature, microphones with first order beam patterns (such as cardioid and dipolar) are considered to have the capability of picking up this velocity component of sound. This suggests that the first order spherical harmonics, which form all of the first order beam patterns, should have some mathematical relationship with the particle velocity of the sound.

Here, we derive the expressions that relate the 1st order spherical harmonic coefficients to the acoustic particle velocity in the  $x$ ,  $y$  and  $z$  directions.

Defining the spherical coordinate system  $(r, \theta, \phi)$  in relation to the Cartesian coordinate system, the particle velocity at the origin is related to the spherical harmonic coefficients by the following theorem:

**Theorem 1.** *The acoustic particle velocity at the point  $\mathbf{0} \equiv (0, 0, 0)$  along the  $x, y$  and  $z$  axes at a particular frequency  $k$  can be expressed using the first order spherical harmonic coefficients,*

$$V_x(\mathbf{0}, k) = \frac{i\rho_0 c}{\sqrt{24\pi}}(C_{11}(k) + C_{1,-1}(k)) \quad (2.20)$$

$$V_y(\mathbf{0}, k) = \frac{-\rho_0 c}{\sqrt{24\pi}}(C_{11}(k) - C_{1,-1}(k)) \quad (2.21)$$

$$V_z(\mathbf{0}, k) = \frac{i\rho_0 c}{\sqrt{12\pi}}C_{10}(k), \quad (2.22)$$

where  $\rho_0$  is the density of the medium,  $c$  is the speed of sound, and  $C_{lm}(k)$  denotes the spherical harmonic coefficient of order  $l$  and mode  $m$ .

*Proof.* The particle velocity  $V_{\mathbf{x}}(\mathbf{x}_0, k)$  at position  $\mathbf{x}_0$ , in the direction  $\mathbf{x}$ , is related to the sound pressure by [60]

$$V_{\mathbf{x}}(\mathbf{x}_0, k) = \frac{i}{k\rho_0 c} \frac{\partial P(\mathbf{x}_0, k)}{\partial \mathbf{x}}. \quad (2.23)$$

For the proof of (2.20), we consider the sound pressure at a point on the  $x$ -axis, whose coordinate in the spherical coordinate system is  $(r, \pi/2, 0)$ , the sound pressure can be decomposed using (2.1),

$$P(r, \frac{\pi}{2}, 0, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kr) Y_{lm}(\frac{\pi}{2}, 0). \quad (2.24)$$

Taking the partial derivative of  $P(r, \pi/2, 0, k)$  in the direction of  $r$ , which is equivalent to  $\frac{\partial P(x,y,z)}{\partial x}$ , we have

$$\frac{\partial P(r, \frac{\pi}{2}, 0, k)}{\partial r} = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) \frac{\partial j_l(kr)}{\partial r} Y_{lm}(\frac{\pi}{2}, 0) \quad (2.25)$$

Since we consider the partial derivative at the origin, we let  $r \rightarrow 0$ . Using the recurrent relationship [61]

$$lj_{l-1}(x) - (l+1)j_{l+1}(x) = (2l+1) \frac{d j_l(x)}{dx}, \quad (2.26)$$

and the fact that

$$j_l(0) = \begin{cases} 1, & \text{if } l = 0 \\ 0, & \text{if } l = 1, 2, 3\dots \end{cases} \quad (2.27)$$

It can be shown that

$$\lim_{r \rightarrow 0} \frac{\partial j_l(kr)}{\partial r} = \begin{cases} k/3, & \text{if } n = 1 \\ 0, & \text{otherwise.} \end{cases} \quad (2.28)$$

In addition,  $Y_{10}(\pi/2, 0) = 0$ . Therefore from (2.25) we have

$$\lim_{r \rightarrow 0} \frac{\partial P(r, \frac{\pi}{2}, 0, k)}{\partial r} = \frac{k}{3} (C_{11} Y_{11}(\frac{\pi}{2}, 0) + C_{1,-1} Y_{1,-1}(\frac{\pi}{2}, 0)) \quad (2.29)$$

Substituting (2.29) into (2.23) with the values  $Y_{11}(\pi/2, 0) = Y_{1,-1}(\pi/2, 0) = \sqrt{3/8\pi}$  completes the proof.

For the proof of (2.21), we consider the partial derivative of sound pressure at  $(r, \pi/2, \pi/2)$ . The derivation is identical to that of  $\frac{\partial P}{\partial x}$ , except that  $Y_{lm}(\pi/2, 0)$  are replaced by  $Y_{lm}(\pi/2, \pi/2)$ .

In the case of (2.22), we consider the partial derivative of sound pressure at  $(r, 0, \phi)$  along  $r$ . Similar to (2.25), we can write

$$\frac{\partial P(r, 0, \phi, k)}{\partial r} = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) \frac{\partial j_l(kr)}{\partial r} Y_{lm}(0, \phi). \quad (2.30)$$

Due to the fact that  $Y_{11}(0, \phi) = 0$  and  $Y_{1,-1}(0, \phi) = 0$ , and utilizing (2.28), we can simplify (2.30), such that

$$\lim_{r \rightarrow 0} \frac{\partial P(r, 0, \phi, k)}{\partial r} = \frac{k}{3} C_{10} Y_{10}(0, \phi) \quad (2.31)$$

Substituting (2.31) into (2.23) with  $Y_{10}(0, \phi) = \sqrt{3/4\pi}$  into completes the proof.  $\square$

Theorem 1 provides a direct link between the signal received by a first order microphone and the 1st order spherical harmonic coefficients representing the sound field. For example, when placing a bi-directional microphone at the origin, with its two beams coincide with the  $z$  axis, then the signal received by the microphone is equivalent to the coefficient  $C_{10}$ , up to a constant scaling factor.

### 2.2.5 Real-valued spherical harmonics

The technique of spherical harmonic analysis is widely used in areas other than spatial audio, such as geophysics [62, 63] and computer graphics [59]. In many of these applications, the spherical functions to be analyzed are real-valued. For these applications, it is sufficient to use real-value spherical harmonics to decompose the spatial functions.

The real-value spherical harmonics can be defined as [62]

$$Y_{lm}^R(\theta, \phi) = \begin{cases} \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_{l|m|}(\cos \theta) \cos(m\phi), & \text{if } m > 0; \\ \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_{l|m|}(\cos \theta), & \text{if } m = 0; \\ \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_{l|m|}(\cos \theta) \sin(m\phi), & \text{if } m < 0. \end{cases} \quad (2.32)$$

The real-value spherical harmonics have the orthogonal property

$$\int_0^\pi \int_0^{2\pi} Y_{lm}^R(\theta, \phi) Y_{l'm'}^R(\theta, \phi) \sin(\theta) d\theta d\phi = \delta_{l-l', m-m'}. \quad (2.33)$$

Compared to the complex-value spherical harmonics, it can be seen that the only difference is that instead of using the complex exponential  $e^{im\phi}$  to express the function in the azimuth direction, the real-value spherical harmonics use the sinusoid functions. Therefore, many properties of the complex-value spherical harmonics are also valid for the real-value spherical harmonics.

It can be seen that the complex-value and real-value spherical harmonics are related through the following equation

$$Y_{lm}^R(\theta, \phi) = \frac{Y_{lm}(\theta, \phi) + Y_{l,-m}(\theta, \phi)}{2}. \quad (2.34)$$

The complex-value spherical harmonics are used for analyzing the spatial sound in the frequency domain. However, the time domain sound pressure signal is real-valued, therefore, if the spherical harmonic analysis is performed in the time domain, it is preferable to use real-value spherical harmonics instead. This is discussed in detail in Chapter 7.

## 2.3 Spatial sound recording and synthesis using spherical harmonic expansion

### 2.3.1 Spatial sound recording using spherical microphone array

The spherical microphone arrays are very suitable for capturing the spherical harmonic coefficients of a spatial sound field, since their geometry coincide with that of the spherical harmonics. The methods to capture spherical harmonics using open and rigid spherical microphone arrays have been described in [10] and [64]. The orthogonal property of the spherical harmonics is exploited in both of these methods.

For open sphere microphone arrays with radius  $R$ , the sound pressure on the surface of the spherical array can be expressed using (2.1). Multiplying both sides of (2.1) with  $Y_{lm}^*(\theta, \phi)$  and integrating over the sphere yields

$$C_{lm}(k)j_l(kR) = \int_0^\pi \int_0^{2\pi} P(R, \theta, \phi, k) Y_{lm}^*(\theta, \phi) d\theta d\phi, \quad (2.35)$$

where the orthogonal property (2.5) is used in the derivation.

The integration in (2.35) can be approximated using a finite number of microphones, placed uniformly over the sphere. The spherical harmonic coefficients can thus be calculated using [10]

$$C_{lm}(k) = \frac{1}{j_l(kR)} \sum_i P(R, \theta_i, \phi_i, k) Y_{lm}^*(\theta_i, \phi_i) \gamma_i, \quad (2.36)$$

where  $\theta_i$  and  $\phi_i$  are the elevation and azimuth angle of the  $i$ th microphone, and  $\gamma_i$  are some weighting coefficients specific to the sampling scheme of the microphone array. The number of microphones on the sphere should be no fewer than  $(L + 1)^2$ , where  $L$  is the maximum order of the spatial sound in the area, determined using (2.7).

In the case of rigid sphere microphone array, the microphones are mounted on a rigid spherical baffle. The sound field around the microphone array is affected by the baffle, and the sound pressure on the surface of the baffle can be expressed by

$$P(R, \theta, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) b_l(kR) Y_{lm}(\theta, \phi), \quad (2.37)$$

where

$$b_l(kR) = j_l(kR) - \frac{h_l^{(2)\prime}(kR)}{j_l'(kR)} h_l^{(2)}(kR), \quad (2.38)$$

and  $h_l^{(2)}(kR)$  is the spherical Hankel function of the second kind. Using the same spherical integration method, the spherical harmonics can be calculated as [64]

$$C_{lm} = \frac{1}{b_l(kR)} \sum_i P(R, \theta_i, \phi_i, k) Y_{lm}^*(\theta_i, \phi_i) \gamma_i, \quad (2.39)$$

Compared to the open sphere microphone array, the rigid sphere array avoids the ill-conditioning problem caused by  $j_l(kR)$  approaching zero at certain combinations of  $k$  and  $R$ . However, the rigid baffle completely encloses the region of interest, rendering this array format hard to implement in larger sizes, and hinders its application in fields such as spatial ANC.

### 2.3.2 Spatial sound recording using non-spherical microphone array

Non-spherical microphone array layouts have also been proposed for the purpose of spatial sound recording based on spherical harmonic analysis. In [13], it is proposed to use multiple circular microphone arrays to capture the spatial sound. This method offers superior flexibility in terms of array geometry compared to spherical microphone arrays, since the radius and position of each circular array can vary within a certain limit. We briefly outline this work in this section.

Consider a circular microphone array placed parallel to the  $x - y$  plane, with its center located on the  $z$  axis. The sound pressure at a point on the array can be expressed using (2.1) and (2.2) as

$$P(R, \vartheta, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kR) \mathcal{P}_{l|m|}(\cos \vartheta) E_m(\phi), \quad (2.40)$$

where  $R$  is the distance from the origin to the circular array, and  $\vartheta$  is the elevation angle of the array. Multiplying both sides of (2.40) by  $E_{-m}(\phi)$  and integrate with respect to  $\phi$  over  $[0, 2\pi]$ , we have

$$\alpha_m(R, \vartheta, k) = \sum_{l=|m|}^L C_{lm}(k) j_l(kR) \mathcal{P}_{l|m|}(\cos \vartheta), \quad (2.41)$$

where we define

$$\alpha_m(R, \vartheta, k) \triangleq \int_0^{2\pi} P(R, \vartheta, \phi, k) E_{-m}(\phi) d\phi. \quad (2.42)$$

For a given circular array, the maximum order of observable spherical harmonic is limited by (2.7). Equation (2.42) can be evaluated for  $m = -L, -L+1 \dots L$ .

When multiple circular arrays are deployed, each with radius and elevation angle  $(R_q, \vartheta_q)$ , the spherical harmonic coefficients of mode  $m$  can be solved through solving the LMS problem

$$\mathbf{J}_m \mathbf{C}_m = \boldsymbol{\alpha}_m, \quad (2.43)$$

where  $\mathbf{C}_m = [C_{(|m|, m)}, C_{|m|+1, m} \dots C_{Lm}]$  is a vector containing all the spherical harmonics of mode  $m$ ,  $\boldsymbol{\alpha}_m = [\alpha_m^1, \alpha_m^2 \dots \alpha_m^Q]$  is a vector containing  $\alpha_m$  from the  $q$ th circular array, and

$$\mathbf{J}_m = \begin{bmatrix} j_{|m|}(kR_1)\mathcal{P}_{|m|, |m|}(\vartheta_1) & j_{|m|+1}(kR_1)\mathcal{P}_{|m|, |m|}(\vartheta_1) & \dots & j_L(kR_1)\mathcal{P}_{|m|, |m|}(\vartheta_1) \\ j_{|m|}(kR_2)\mathcal{P}_{|m|, |m|}(\vartheta_2) & j_{|m|+1}(kR_2)\mathcal{P}_{|m|, |m|}(\vartheta_2) & \dots & j_L(kR_2)\mathcal{P}_{|m|, |m|}(\vartheta_2) \\ \vdots & \vdots & \ddots & \vdots \\ j_{|m|}(kR_Q)\mathcal{P}_{|m|, |m|}(\vartheta_Q) & j_{|m|+1}(kR_Q)\mathcal{P}_{|m|, |m|}(\vartheta_Q) & \dots & j_L(kR_Q)\mathcal{P}_{|m|, |m|}(\vartheta_Q) \end{bmatrix}. \quad (2.44)$$

The complete set of spherical harmonics can be found by solving (2.43) for every value of  $m$  which satisfies  $|m| \leq L$ . At certain combinations of array radius, position and sound frequency, the value of  $j_l(kR)\mathcal{P}_{l|m|}(\vartheta)$  may equal to zero for some  $l$  and  $m$  [13]. Ill-conditioning of  $\mathbf{J}_m$  due to this phenomenon can be avoided by employing extra circular microphone arrays [13].

Compared to spherical microphone array apertures, this method allows more flexible placement of the microphones, and the use of circular arrays can simplify the supporting structure for the microphones. Therefore this method presents a more practical solution for spatial sound recording over a larger region.

### 2.3.3 Spatial sound synthesis based on mode matching

Similar to the recording of spatial sound, a desired spatial sound field can be synthesized using a loudspeaker array, placed around the designated reproduction region. A commonly used spatial sound synthesis method is the spherical harmonic mode matching method, where the loudspeaker array often has a spherical shape, such as in [44–46].

The mode matching method aims to find driving signals for each loudspeaker

on the array so that the combined spherical harmonic coefficients due to all the loudspeakers equal to some desired value, i.e.,

$$\sum_q D_q H_{lm}^q = C_{lm}^{\text{desire}}, \quad (2.45)$$

where  $H_{lm}^q$  denotes the spherical harmonic coefficients due to the  $q$ th loudspeaker playing a unit signal,  $C_{lm}^{\text{desire}}$  denotes the spherical harmonic coefficient of the desired sound field, and  $D_q$  is the driving signal for the  $q$ th loudspeaker. This problem can be solved in a LMS manner, as

$$\mathbf{D} = \mathbf{H}^{-1} \mathbf{C}^{\text{desire}}, \quad (2.46)$$

where  $\mathbf{D} = [D_1, D_2 \dots D_Q]^T$  is the vector containing all the driving signals,  $\mathbf{C}^{\text{desire}} = [C_{00}^{\text{desire}}, C_{11}^{\text{desire}} \dots C_{LL}^{\text{desire}}]^T$  is the vector of desired spherical harmonic coefficients, and

$$\mathbf{H} = \begin{bmatrix} H_{00}^1 & H_{00}^1 & \dots & H_{00}^Q \\ H_{11}^1 & H_{11}^1 & \dots & H_{11}^Q \\ \vdots & \vdots & \ddots & \vdots \\ H_{LL}^1 & H_{LL}^1 & \dots & H_{LL}^Q \end{bmatrix} \quad (2.47)$$

is the channel matrix containing the spherical harmonic coefficients due to each loudspeaker.

Assuming a loudspeaker can be modeled as a point source, the sound field due to a loudspeaker placed at  $(R, \vartheta, \varphi)$  can be expanded as [44]

$$P(r, \theta, \phi, k) = ik \sum_{l=0}^{\infty} j_l(kr) h_l(kR) \sum_{m=-l}^l Y_{lm}(\theta, \phi) Y_{lm}(\vartheta, \varphi)^*, \quad (2.48)$$

if the loudspeaker is placed at a long distance from the reproduction region, its sound wave can be seen as plane wave, which can be expanded as [44]

$$P(r, \theta, \phi, k) = 4\pi \sum_{l=0}^{\infty} j_l(kr) i^l \sum_{m=-l}^l Y_{lm}(\theta, \phi) Y_{lm}(\vartheta, \varphi)^*, \quad (2.49)$$

Comparing (2.48) and (2.49) with (2.1), it can be seen that the spherical harmonic coefficients corresponding to a point source and a plane wave source are

$$H_{lm}^{\text{point}}(R, \vartheta, \varphi) = ik h_l(kR) Y_{lm}(\vartheta, \varphi) \quad (2.50)$$

and

$$H_{lm}^{\text{plane}}(\vartheta, \varphi) = 4\pi i^l Y_{lm}(\vartheta, \varphi) \quad (2.51)$$

respectively. If the loudspeakers are arranged in a spherical geometry around the reproduction region, with a uniform spherical sampling scheme, due to the orthogonal property of the spherical harmonics, we have

$$H^{-1} = H^H, \quad (2.52)$$

thus the driving signals for each loudspeaker can be solved using

$$\mathbf{D} = \mathbf{H}^H \mathbf{C}^{\text{desire}}. \quad (2.53)$$

Perfect reproduction of the desired sound field cannot be guaranteed if the loudspeakers are not distributed evenly around the sphere, or an insufficient number of loudspeakers are available. However, if no less than  $(L + 1)$  number of loudspeakers are used, and uniformly distributed in a spherical arrangement, high quality sound field reproduction can be achieved [44].

The technique of spherical harmonic expansion, the properties of the spherical harmonics as well as the spatial sound recording and synthesis techniques discussed in this chapter form a foundation for the algorithms and techniques to be developed in the later chapters of this thesis.

# Chapter 3

## Planar microphone array apertures for 3D spatial sound field analysis

**Overview:** Spherical harmonic analysis is a very useful tool for representing the noise field. However, a drawback of this technique is the three-dimensional microphone arrays required for recording the noise sound field. In this chapter, a method to design 2D planar microphone arrays that are capable of capturing 3D spatial sound fields is proposed. Through the utilization of both omni-directional and first order microphones, the proposed microphone array is capable of measuring sound field components that are undetectable to conventional planar omni-directional microphone arrays, thus providing the same functionality as 3D arrays designed for the same purpose. Simulations show that the accuracy of the planar microphone array is comparable to traditional spherical microphone arrays. Due to its compact shape, the proposed microphone array greatly increases the feasibility of 3D sound field analysis techniques in spatial ANC applications.

### 3.1 Introduction

We use spherical harmonic analysis as a tool to represent the 3D noise field, due to its various benefits such as accurate sound field representation and the ability to perform in-depth analysis to the noise field. In order to capture the 3D noise field in real time for the ANC system, it is necessary to use a microphone array which has the capability to capture 3D sound field, in terms of spherical harmonic coefficients

of the sound field. To the best of our knowledge, all of the previously developed microphone array structures designed for this purpose have a 3D geometry, which limits their feasibility for compact ANC systems suitable for real-life applications.

As was discussed in Chapter 2, spherical microphone array geometries are well-suited for the spherical harmonic transform, and both open and rigid sphere models have been studied [10, 43]. Both models are widely used in research applications, such as room geometry inference [65] and near field acoustic holography (NAH) [66]. An inherent drawback of the open sphere model is the numerical ill-conditioning problem, which is due to the nulls in spherical Bessel functions, thus the diameter of the microphone array has to be chosen carefully. It has been shown that such ill-conditioning problem can be overcome via methods such as using concentric spheres [67, 68], co-centered rigid/open spheres [69], or by measuring the radial velocity [43].

The placement of microphones on a spherical array has to follow a strict rule of orthogonality of the spherical harmonics [15, 70], which limits the flexibility of the array configuration. The spherical shape of the array also pose difficulties on implementation as well as practical usage.

Non-spherical microphone arrays, such as the conical microphone array aperture proposed by Gupta et.al. [71] and the multiple circular microphone array proposed by Abhayapala et. al. [13, 72] can also be used for spherical harmonic analysis. These microphone arrays offer greater geometrical flexibility compared to spherical microphone arrays, thus allowing easier implementation of larger microphone arrays. However, these apertures still occupy a 3D space, which hinders the development of compact microphone arrays for practical applications.

On the other hand, microphone arrays featuring 2D geometry are easy to implement, yet existing 2D microphone arrays are incapable of capturing complete 3D sound field information. Meyer et.al. have shown that a 2D microphone array can be used to measure certain vertical component of a 3D sound field [73]. However, due to inherent properties of the spherical harmonics, some spherical harmonic modes are invisible to omni-directional pressure microphones on the  $x - y$  plane, which explains why previously proposed 2D microphone arrays fail to extract full 3D sound field information. Measurement of these sound field components on the  $x - y$  plane calls for additional types of sensors, no such technique has been proposed to our best knowledge.

First order microphones, such as differential microphones and cardioid microphones, are known to have the capability of detecting acoustic velocity in a certain direction [74]. Kuntz et. al. have shown that through using cardioid microphones

pointed in the radial direction to replace omni-directional microphones in a circular array, the numerical ill-conditioning problem can be solved for a 2D sound field analysis system [60].

In this chapter, we first investigate using first order microphones to aid the detection of 3D sound fields, and propose a new method for 3D sound field recording using a 2D planar microphone array. In our approach, we use first order microphones in conjunction with omni-directional microphones to measure the “invisible” component of a 3D sound field on the  $x - y$  plane. Also, we propose a method of using multiple co-centered circular arrays of omnidirectional/first order microphones to compute the sound field coefficients associated with the spherical space enclosing the planar array aperture. We show that the proposed planar microphone array offers the same functionality as spherical/multiple circular arrays designed for sound field analysis.

In addition, we propose a method to capture 3D sound field using circular arrays of higher order microphones, also placed on a 2D plane. This method can be seen as a generalization of the method discussed in the first few sections of this chapter.

This chapter is arranged as follows: Section 3.2 derives the wave domain expression of sound field measured by general first order microphone. We show that the full 3D sound field can be observed on a plane with the aid of first order microphones by exploiting a property of the associated Legendre functions. Section 3.3 introduces the co-centered hybrid circular microphone array for sound field recording, and shows how the sound field coefficients can be calculated using the data measured by different components of the hybrid array. We also provide a step-by-step design procedure for determining parameters of an array based on system requirements. Section 3.4 provides an analysis on the recording accuracy of the proposed array. Two primary causes of errors are identified, and their impact on each sound field coefficient is discussed. Section 3.5 gives an hypothetical design example of the proposed microphone array, as well as an experimental microphone array built for validation of the theory. Detailed simulation results are provided for the hypothetical design example and the test results of the experimental array is compared with corresponding simulation results for performance evaluation.

## 3.2 First order microphones for sound field acquisition

In this section, we derive the general velocity of the pressure field at a point along a direction and the wave domain expression of the received signal of a general first order microphone. We also show that the 3D sound field coefficients can be divided into even and odd components, while the even modes can be measured by omnidirectional microphones, the odd components of the sound field can be observed on a plane by using a recurrent relationship of associated Legendre functions.

### Wave domain expression of pressure gradient

For reasons that will become clear later in the chapter, we consider pressure gradient of a sound field along the direction of  $\theta$ . That is, we consider either differential or velocity microphones placed in such a way that they measure pressure gradient in the direction of  $\theta$  at a given point  $(r, \theta, \phi)$ .

We define the pressure gradient of sound along the direction of  $\theta$  at a point  $(r, \theta, \phi)$  as

$$P_\theta(r, \theta, \phi, k) \triangleq \frac{\partial P(r, \theta, \phi, k)}{\partial \theta}. \quad (3.1)$$

By substituting (2.1) into (3.1) and taking the partial derivative with respect to  $\theta$ , the pressure gradient can be expressed as

$$P_\theta(r, \theta, \phi, k) = -\sin \theta \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kr) \mathcal{P}'_{l|m|}(\cos \theta) E_m(\phi), \quad (3.2)$$

where

$$\mathcal{P}'_{l|m|}(u) = \frac{d\mathcal{P}_{l|m|}(u)}{du}$$

is the first order derivative of the normalized associated Legendre function.

### 3.2.1 General expression for first order microphones

The pick-up pattern of any first order microphone can be considered as a weighted sum of an omni-directional pattern and a differential pattern. Using  $P(r, \theta, \phi, k)$  to represent the omnidirectional component of the measured sound pressure and  $P_\theta(r, \theta, \phi, k)$  for the differential component in the  $\theta$  direction at point  $(r, \theta, \phi)$ , the total sound pressure measured by an arbitrary first order microphone can be written

as

$$P_c(r, \theta, \phi, k) \triangleq \beta P(r, \theta, \phi, k) + (1 - \beta) P_\theta(r, \theta, \phi, k), \quad (3.3)$$

where  $\beta$  is a weighing factor and has a range of  $[0, 1]$ . When  $\beta = 0$ ,  $P_c(r, \theta, \phi, k)$  contains only the differential pattern, which is considered as a special case of first order pick-up patterns. Here, differential microphones are regarded as one type of first order microphones; when  $\beta = 0.5$ ,  $P_c(r, \theta, \phi, k)$  becomes the pick-up pattern of a “standard” cardioid microphone. Substituting (2.1) and (3.2) into (3.3) yields the wave domain representation of the signal received by a general first order microphone as

$$P_c(r, \theta, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kr) (\beta \mathcal{P}_{l|m|}(\cos \theta) - (1 - \beta) \sin \theta \mathcal{P}'_{l|m|}(\cos \theta)) E_m(\phi). \quad (3.4)$$

### 3.2.2 Sampling on a plane

Without loss of generality, let us place the co-ordinate system such that the plane of interest for sensor placement is the x-y plane. In the spherical co-ordinate system,  $\theta = \pi/2$  (i.e.,  $\cos \theta = 0$ ), for all points on the x-y plane. Thus, the output of an omni-directional sensor placed on the x-y plane is

$$P(r, \pi/2, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kr) \mathcal{P}_{l|m|}(0) E_m(\phi). \quad (3.5)$$

Observe that when  $l + |m|$  is an odd integer the value of  $\mathcal{P}_{l|m|}(0)$  is equal to zero [13]. Consequently, the spherical harmonics associated with these associated Legendre Functions are equal to zero. This property makes the odd mode spherical harmonics “invisible” on the  $\theta = \pi/2$  plane, which is why extraction of the complete 3D sound field information cannot be done through sampling on a single plane using omni directional microphones.

On the other hand,

$$\mathcal{P}'_{l|m|}(\cos \frac{\pi}{2}) = \begin{cases} \text{a non-zero value, when } l + |m| \text{ is an odd integer,} \\ 0, \text{ when } l + |m| \text{ is an even integer.} \end{cases}$$

Observe that the expression for the pressure gradient in (3.2) has the terms  $\mathcal{P}'_{l|m|}(\cdot)$ . Hence the ‘odd’ components of the pressure gradient along the direction of  $\theta$  is non-

zero on the x-y plane. Thus, the pressure gradient measurements contain ‘odd’  $C_{lm}(k)$  (i.e.,  $l + |m|$  odd) coefficients. We use this property in this work to propose a method to extract 3D sound field components by sampling the field on the x-y plane using differential (or first order) and omni directional microphones together.

Using the recurrent relationship of the normalized associated Legendre functions by substituting (2.10) into (3.2) and (3.4), we can write the output of the differential and general first order microphones placed at a point  $(r, \pi/2, \phi)$  on the x-y plane along the direction of  $\theta$  (i.e., perpendicular to the x-y plane) as

$$P_\theta(r, \frac{\pi}{2}, \phi, k) = - \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kr) \sqrt{\frac{(2l+1)(l^2-m^2)}{(2l-1)}} \mathcal{P}_{(l-1)|m|}(0) E_m(\phi) \quad (3.6)$$

and

$$P_c(r, \frac{\pi}{2}, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_l(kr) (\beta \mathcal{P}_{l|m|}(0) - (1-\beta) \sqrt{\frac{(2l+1)(l^2-m^2)}{(2l-1)}} \mathcal{P}_{(l-1)|m|}(0)) E_m(\phi), \quad (3.7)$$

respectively.

### 3.3 Array configuration

In this section we outline possible geometric configurations of first order and omni-directional sensors on the x-y plane to extract both the even and odd spherical harmony components of the sound field.

#### 3.3.1 Calculation of harmonic coefficients

##### Even coefficients: Omni-Array

Consider a circle placed on the x-y plane such that an arbitrary point on the circle is given by  $(R_q, \pi/2, \phi)$ . Then the output of a omni-directional microphone on the

circle at  $(R_q, \pi/2, \phi)$  is given by

$$P(R_q, \frac{\pi}{2}, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(k) j_n(kR_q) \mathcal{P}_{l|m|}(0) E_m(\phi). \quad (3.8)$$

Since sound fields over a spherical region of finite radius are mode limited (2.6), the infinite summation on right hand side of (3.8) can be approximated by a finite sum,

$$P(R_q, \frac{\pi}{2}, \phi, k) \approx \sum_{l=0}^L \sum_{m=-l}^l C_{lm}(k) j_l(kR_q) \mathcal{P}_{l|m|}(0) E_m(\phi). \quad (3.9)$$

where  $L$  denotes the maximum harmonic order at the array's radius  $R_q$  and the highest operating frequency [53]. Multiplying both sides of (3.9) by  $E_m(-\phi)$  and integrating with respect to  $\phi$  over  $[0, 2\pi]$  yields the total sound pressure received by the ring, as

$$\alpha_m(R_q, k) \triangleq \int_0^{2\pi} P(R_q, \pi/2, \phi, k) E_m(-\phi) d\phi \quad (3.10)$$

$$= \sum_{l=|m|}^L C_{lm}(k) j_l(kR_q) \mathcal{P}_{l|m|}(0). \quad (3.11)$$

Note that only the even mode harmonics are present in (3.11), since  $\mathcal{P}_{l|m|}(0) = 0$  for  $l + |m|$  odd. Let there be a total of  $Q$  circles placed at different radii but all on the  $\theta = \pi/2$  plane (x-y plane). Thus, for  $q = 1, \dots, Q$ , the relationship between the even mode sound field coefficients of mode  $m$  and the *azimuth sound pressure harmonics*  $\alpha_m(R_q, k)$  on each circle can be expressed as

$$\boldsymbol{\alpha}_m(k) = \mathbf{U}_m(k) \mathbf{C}_m^{\text{even}}(k) \quad (3.12)$$

where  $\boldsymbol{\alpha}_m(k) = [\alpha_m(R_1, k), \alpha_m(R_2, k), \dots, \alpha_m(R_Q, k)]^T$ ,

$$\mathbf{C}_m^{\text{even}}(k) = \begin{cases} [C_{mm}(k), C_{(m+2)m}(k), \dots, C_{Lm}(k)]^T, & \text{if } m \text{ and } L \text{ are both even/odd} \\ [C_{mm}(k), C_{(m+2)m}(k), \dots, C_{(L-1)m}(k)]^T, & \text{otherwise} \end{cases} \quad (3.13)$$

is the vector of the even mode coefficients of mode  $m$  and

$$\mathbf{U}_m(k) = \begin{bmatrix} j_m(kR_1)\mathcal{P}_{m|m|}(0) & j_{m+2}(kR_1)\mathcal{P}_{(m+2)|m|}(0) & \cdots & j_L(kR_1)\mathcal{P}_{L|m|}(0) \\ j_m(kR_2)\mathcal{P}_{m|m|}(0) & j_{m+2}(kR_1)\mathcal{P}_{(m+2)|m|}(0) & \cdots & j_L(kR_2)\mathcal{P}_{L|m|}(0) \\ \vdots & \vdots & \vdots & \vdots \\ j_m(kR_Q)\mathcal{P}_{m|m|}(0) & j_{m+2}(kR_Q)\mathcal{P}_{(m+2)|m|}(0) & \cdots & j_L(kR_Q)\mathcal{P}_{L|m|}(0) \end{bmatrix}, \quad (3.14)$$

for the case with both  $N$  and  $m$  are either odd or even (otherwise replace  $L$  in (3.14) by  $L - 1$ ).

We can estimate the even mode coefficients from (3.12), provided  $\mathbf{U}_m(k)$  is not singular, as

$$\mathbf{C}_m^{\text{even}}(k) = \mathbf{U}_m^\dagger(k)\boldsymbol{\alpha}_m(k) \quad (3.15)$$

where  $\mathbf{U}_m^\dagger = (\mathbf{U}_m^T \mathbf{U}_m)^{-1} \mathbf{U}_m^T$  is the pseudo inverse of  $\mathbf{U}_m$ .

Note that the calculation of even harmonic coefficients are similar to the work presented in [13]. However, we show in the following subsection how to extract odd harmonic coefficients by placing the differential microphones on the x-y plane which is a method not reported elsewhere to the best of our knowledge.

### Odd coefficients: differential microphone array

Consider a circular array of differential microphones with radius  $R_q$  placed on the x-y plane with all differential microphones pointed perpendicular to the x-y plane (i.e.,  $\theta = \pi/2$  plane). Then the output of a differential microphone on the circle at  $(R_q, \pi/2, \phi)$  is given by (3.6). Using the properties of the spherical Bessel functions, we can show that the infinite summation of (3.6) can be truncated to a finite number (similar to the case of (3.9)). The resulting equation is given below:

$$P_\theta(R_q, \frac{\pi}{2}, \phi, k) = - \sum_{l=0}^N \sum_{m=-l}^l C_{lm}(k) j_l(kR_q) \sqrt{\frac{(2l+1)(l^2-m^2)}{(2l-1)}} \mathcal{P}_{(l-1)|m|}(0) E_m(\phi). \quad (3.16)$$

By multiplying both sides of (3.16) by  $E_m(-\phi)$  and integrating with respect to  $\phi$  over  $[0, 2\pi]$ , we obtain the response of the differential microphone array, named as *azimuth pressure gradient harmonics*

$$\alpha_m^{(\text{d})}(R_q, k) \triangleq \int_0^{2\pi} P_\theta(R_q, \frac{\pi}{2}, \phi, k) E_m(-\phi) d\phi \quad (3.17)$$

$$= - \sum_{l=|m|}^L C_{lm}(k) j_l(kR_q) \sqrt{\frac{(2l+1)(l^2-m^2)}{(2l-1)}} \mathcal{P}_{(l-1)|m|}(0) \quad (3.18)$$

Note that only the odd mode harmonics are present in (3.18), since  $\mathcal{P}_{(l-1)|m|}(0) = 0$  for  $l + |m|$  even.

By evaluating (3.18) for  $q = 1, \dots, Q$ , the relationship between the odd sound field coefficients of mode  $m$  and  $\alpha_m^{(\text{d})}(R_q, k)$  on each circle can be expressed as a matrix equation:

$$\boldsymbol{\alpha}_m^{(\text{d})}(k) = \mathbf{V}_m(k) \mathbf{C}_m^{\text{odd}}(k) \quad (3.19)$$

where  $\boldsymbol{\alpha}_m^{(\text{d})}(k) = [\alpha_m^{(\text{d})}(R_1, k), \alpha_m^{(\text{d})}(R_2, k), \dots, \alpha_m^{(\text{d})}(R_Q, k)]^T$ ,

$$\mathbf{C}_m^{\text{odd}}(k) = \begin{cases} [C_{(m+1)m}(k), C_{(m+3)m}(k), \dots, C_{(L-1)m}(k)]^T, & \text{if } m \text{ and } L \text{ are both even/odd} \\ [C_{(m+1)m}(k), C_{(m+3)m}(k), \dots, C_{Lm}(k)]^T, & \text{otherwise} \end{cases} \quad (3.20)$$

and

$$\mathbf{V}_m(k) = \begin{bmatrix} V_{(m+1)|m|}^{(1)} & V_{(m+3)|m|}^{(1)} & \cdots & V_{(L-1)|m|}^{(1)} \\ V_{(m+1)|m|}^{(2)} & V_{(m+3)|m|}^{(2)} & \cdots & V_{(L-1)|m|}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{(m+1)|m|}^{(Q)} & V_{(m+3)|m|}^{(Q)} & \cdots & V_{(L-1)|m|}^{(Q)} \end{bmatrix}, \quad (3.21)$$

with

$$V_{l|m|}^{(q)} = -\sqrt{\frac{(2l+1)(l^2-m^2)}{(2l-1)}} j_l(kR_q) \mathcal{P}_{(l-1)|m|}(0) \quad (3.22)$$

for the case with both  $L$  and  $m$  are either odd or even (otherwise replace  $L-1$  in (3.21) by  $N$ ).

We can estimate the odd harmonic coefficients from (3.19), provided  $\mathbf{V}_m(k)$  is non-singular, as

$$\mathbf{C}_m^{\text{odd}}(k) = \mathbf{V}_m^\dagger(k) \boldsymbol{\alpha}_m^{(\text{d})}(k) \quad (3.23)$$

where  $\mathbf{V}_m^\dagger = (\mathbf{V}_m^T \mathbf{V}_m)^{-1} \mathbf{V}_m^T$  is the pseudo inverse of  $\mathbf{V}_m$ .

Thus the complete set of sound field coefficients can be derived through solving for the even and odd harmonics coefficients separately using the signal received from omni-directional microphones (3.12) and differential microphones (3.19), respectively.

### Cardioid or general first order microphone arrays

Alternatively, the even and odd harmonic coefficients may be calculated together in one matrix operation. This method is especially suitable for planar arrays that utilize cardioid microphones (or general first order) instead of differential microphones. According to (3.7), a first order (e.g., cardioid) microphone placed on the x-y plane picks up both the even and odd components of the sound field. For a set of finite radii circular arrays of first order microphones placed on the x-y plane, we can write a matrix equation using (3.7) and following similar steps as in the previous two subsections:

$$\boldsymbol{\alpha}_m^{(f)}(k) = \beta \mathbf{U}_m(k) \mathbf{C}_m^{\text{even}}(k) + (1 - \beta) \mathbf{V}_m(k) \mathbf{C}_m^{\text{odd}}(k) \quad (3.24)$$

where  $\boldsymbol{\alpha}_m^{(f)}(k) = [\alpha_m^{(f)}(R_1, k), \alpha_m^{(f)}(R_2, k), \dots, \alpha_m^{(f)}(R_Q, k)]^T$  with

$$\alpha_m^{(f)}(R_q, k) \triangleq \int_0^{2\pi} P_c(R_q, \frac{\pi}{2}, \phi, k) E_m(-\phi) d\phi, \quad (3.25)$$

and  $\mathbf{C}_m^{\text{even}}(k)$ ,  $\mathbf{U}_m(k)$ ,  $\mathbf{C}_m^{\text{odd}}(k)$  and  $\mathbf{V}_m(k)$  are given by (3.13), (3.14), (3.20) and (3.21), respectively.

If we have both omni-directional and first order circular arrays of microphones, then we can combine (3.12) and (3.24) to obtain

$$\begin{bmatrix} \boldsymbol{\alpha}_m(k) \\ \boldsymbol{\alpha}_m^{(f)}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{U}_m(k) & 0 \\ \beta \mathbf{U}_m(k) & (1 - \beta) \mathbf{V}_m(k) \end{bmatrix} \begin{bmatrix} \mathbf{C}_m^{\text{even}}(k) \\ \mathbf{C}_m^{\text{odd}}(k) \end{bmatrix}. \quad (3.26)$$

Equation (3.26) can be solved to calculate both the even and odd harmonics coefficients given by  $\mathbf{C}_m^{\text{even}}(k)$  and  $\mathbf{C}_m^{\text{odd}}(k)$ .

### 3.3.2 Discrete sensor placement: sampling of continuous aperture

In the previous subsection, we assumed that the pressure  $P(R_q, \pi/2, \phi, k)$ , pressure gradient  $P_\theta(R_q, \pi/2, \phi, k)$  and the first order microphone output  $P_c(R_q, \pi/2, \phi, k)$  are readily available over a continuous circular aperture in (3.10), (3.17) and (3.25), respectively. However, in practice we only have a finite set of microphones, and hence a discrete set of samples on the circular aperture. Thus, for an equally spaced

microphone arrays, we approximate the integration in (3.10), (3.17) and (3.25) by summations:

$$\alpha_m(R_q, k) \approx \frac{2\pi}{N_q} \sum_{s=1}^{N_q} P(R_q, \frac{\pi}{2}, \phi_s, k) E_m(-\phi_s) \quad (3.27)$$

$$\alpha_m^{(d)}(R_q, k) \approx \frac{2\pi}{N_q} \sum_{s=1}^{N_q} P_\theta(R_q, \frac{\pi}{2}, \phi_s, k) E_m(-\phi_s) \quad (3.28)$$

$$\alpha_m^{(f)}(R_q, k) \approx \frac{2\pi}{N_q} \sum_{s=1}^{N_q} P_c(R_q, \frac{\pi}{2}, \phi_s, k) E_m(-\phi_s) \quad (3.29)$$

where  $N_q$  are the number of microphones placed in a circle and  $\phi_s$  denotes the azimuth angle of the location of the  $s$ th microphone.

### Number of sensors per circle

Due to the spatial sampling of the sound field, one can only extract a limited number of harmonic orders by each array. In order to sample a set of circular harmonics of maximum order  $L$ , the number of microphones required is given by  $n_{mic} \geq 2L + 1$ , and  $L$  is determined using  $L \leq \lceil ekR/2 \rceil$ , where  $k$  is the wave number and  $R$  is the radius of the region of interest [53]. The exact amount of microphones to be used for each circular array thus depends on the radius of the array as well as the target frequency band.

The truncation of spherical harmonics leads to errors, which will be discussed in Section 3.4. The “rule of thumb”  $L \leq \lceil ekR/2 \rceil$  gives a sufficiently high precision for most applications [53]. For applications that require less accuracy, an alternative truncation number is given by  $L \leq \lceil kR \rceil$  [46], which truncates the order to a lower value, hence reducing system complexity at the cost of accuracy. The former rule is used in this work for higher accuracy.

Since the number of microphones on each circular array is directly linked to the wave number  $k$ , which can then be translated into the wavelength  $\lambda$ , the number of microphones needed can be easily derived from the target frequency of the application as

$$n_{mic} = 2L + 1 = 2\lceil \frac{ekR}{2} \rceil + 1 = 2\lceil \frac{e\pi R}{\lambda} \rceil + 1 = 2\lceil \frac{e\pi fR}{c} \rceil + 1, \quad (3.30)$$

where  $c$  is the speed of wave propagation, in the case of sound,  $c = 340$  m/s. Thus one can directly calculate the number of sampling points (microphones) for a given

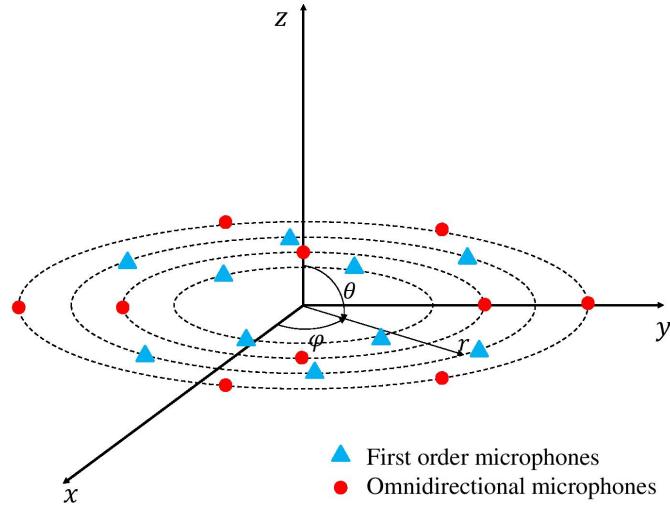


Figure 3.1: Example of omnidirectional (dot) and first order (triangle) microphone arrangement on a 2D plane for 3D sound field analysis.

array radius and a target frequency band. For example, a circular array of 0.4 m radius, designed for audio signals up to 1500 Hz would need 33 microphones.

### Configuration(s)

The array system can be configured to have multiple circular microphone arrays placed on a plane, with half of the arrays using omni-directional microphones, the other half using first order microphones placed perpendicular to the plane. The number of microphones on each array is decided by the target wave number and the radius of the array, therefore smaller arrays may have a lower amount of microphones. Figure 3.1 illustrates such a configuration.

An alternative configuration is to use closely placed omni-directional microphone pairs to realize differential microphones. In this way, each microphone pair is used in two different ways: the two microphone output signals are differentiated to create the bi-directional pick up pattern, which is used for calculation of odd numbered coefficients; in the mean time, one of the two microphone outputs is used to calculate the even-numbered coefficients. Figure 3.2 shows an example of such array arrangement.

The two microphone array configurations require the same number of microphones for the same design target, although the second option uses half the number

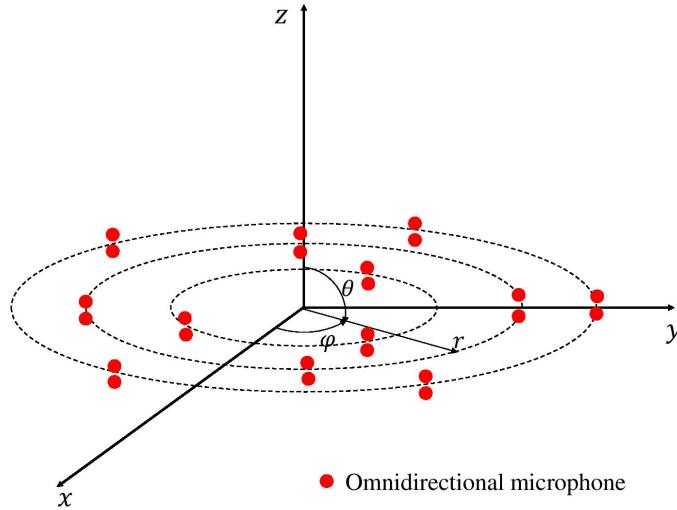


Figure 3.2: Example of omni-directional microphone pair arrangement on a 2D plane for 3D sound field analysis.

of circular arrays. However, it should be noted that the distance between the two microphones in each microphone pair should be small compared to the array radius, so as to best approximate  $P_\theta(r, \theta, \phi, k)$  in (3.1).

### 3.3.3 Array design procedure

A general guidance for designing the planar array is provided in this section. This procedure illustrates the basic steps in setting the parameters of the microphone array.

**Step 1:** Determine the desired frequency band and the radius  $R$  of the region of interest.

**Step 2:** Calculate the maximum order of the sound field using  $L = \lceil ekR/2 \rceil$ .

**Step 3:** Based on the maximum order  $L$ , decide the number of circular arrays to be implemented. For first order microphone configuration, at least  $L_{\text{omni}} = \lceil L/2 \rceil$  omnidirectional sensor arrays, and  $L_{\text{first}} = L - L_{\text{omni}}$  first order arrays are needed. For differential microphone configuration, no less than  $L_{\text{diff}} = \lceil L/2 \rceil$  arrays of microphone pairs are required.

**Step 4:** Determine the radius of each circular array. Choose the radius such that

the spherical Bessel zeros for the target frequency band are avoided. Ensure that the radii of the circular arrays have a good diversity.

**Step 5:** For each circular array, decide the maximum spherical harmonic order  $L_i$ , and estimate the number of microphones to be placed on the array, based on  $n_{\text{mic}} = 2L_i + 1$ .

After settling on a design, the parameters for sound field calculation can then be set based on the dimensions of the array.

### 3.3.4 Comments

We make the following comments and observations with the proposed array:

1. The even spherical harmonics are symmetric about the  $z = 0$  (x-y) plane, while the odd modes are not. A planar microphone array comprising only omnidirectional microphones cannot distinguish the waves that are impinging from either sides of the plane. This fact explains that why this type of array is not capable of detecting the full 3D sound field.
2. First order cardioid microphones that are placed perpendicular to the array plane can pick up a combination of even and odd mode harmonics, but are unable to separate the two components. However, if the even mode harmonic coefficients are known (which can be provided by an omnidirectional microphone array), then it becomes easy to solve for the remaining odd mode coefficients. Thus a hybrid array of both omnidirectional and first order microphones is crucial for detecting full 3D sound field using a planar array aperture.
3. The zeros in the spherical Bessel functions cause certain spherical harmonics to be “invisible” at some radius and frequency, which limits an array’s wide band capabilities. The proposed array aperture samples the sound field at multiple radii, thus improving the array’s redundancy against zero points in the spherical Bessel functions. However, the user should carefully design the array such that at each frequency, a sufficient number of circular arrays are unaffected by the Bessel zeros and are available for calculating the coefficients. In general, a properly designed planar array can avoid the Bessel zero problem for all frequencies, and thus having wideband capabilities, this is shown in Section 3.5 using a hypothetical design example.

4. Although the proposed array has a planar geometry, the free space assumption still applies to our array system, which requires that no sound source or scatterer should exist within the spherical region enveloping the planar array. For this reason, the array cannot be directly placed on walls or tables to capture the surrounding sound. However, a work-around to this problem is to place an appropriate sound absorbing material between the rigid surface (wall, table) and the planar array, which eliminates all reflections from the surface, thus the setup no longer violates the free-space assumption. Furthermore, if the reflection characteristics of the surface is known, it is possible to compensate for the reflection in the calculation. However, this is beyond the scope of this chapter, and we will investigate this in a future work.

## 3.4 Error analysis

In this section we discuss two primary sources of error, and the impact they have on the acquisition accuracy of different sound field coefficients.

### 3.4.1 Differential microphone approximation

As was mentioned in Section 3.3, a differential microphone can be realized using a pair of closely placed omni-directional microphones. However, this implementation only approximates the ideal velocity sensor, using the approximation

$$\frac{P(\mathbf{x} + d\mathbf{x}) - P(\mathbf{x})}{d\mathbf{x}} \approx \frac{\partial P(\mathbf{x})}{\partial \mathbf{x}} = V(\mathbf{x}). \quad (3.31)$$

By choosing sufficiently small value of  $d\mathbf{x}$ , the error of the approximation can be minimized. However, due to implementation constraints such as physical dimension of the microphone units, a very good approximation of (3.31) may not be achievable. We recommend choosing  $d\mathbf{x} \approx 0.1/k_{\max}$ , where  $k_{\max}$  is the wave number corresponding to the maximum operating frequency of the microphone array, so as to minimize the error due to the approximation.

Since this approximation only exists for the sampling of the odd coefficients, the accuracy of the calculated odd coefficients is expected to be slightly worse than that of the even coefficients when the differential microphone approximation is used to implement the array. This phenomenon is observed in the hypothetical design example.

### 3.4.2 Spatial sampling and spatial aliasing

One major source of error in the proposed array system is spatial sampling. By comparing (3.10) and its discrete approximation, (3.28), the error on each harmonic mode due to spatial sampling can be defined as

$$\Delta E_{\text{mode}} \triangleq \int_0^{2\pi} P(r, \theta, \phi, k) E_m(-\phi) d\phi - \frac{2\pi}{n_{\text{mic}}} \sum_{u=1}^{n_{\text{mic}}} P(r, \theta(u), \phi(u), k) E_m(-\phi). \quad (3.32)$$

The same approximation error can be defined for (3.17) and (3.25). Generally speaking, this error is small as long as the Nyquist sampling criteria is met, however, using extra microphones on each circular array can help to improve the accuracy of the system.

The truncation of spherical harmonic modes mentioned in Section 3.3 also leads to errors, as the energy of the truncated higher order harmonics are aliased into the observed harmonics during calculation. The truncation error can be expressed as

$$\begin{aligned} \Delta E_{\text{trunc}} &\triangleq \sum_{l=|m|}^{\infty} C_{lm} j_l(kr) \mathcal{P}_{l|m|}(0) - \sum_{l=|m|}^L C_{lm} j_l(kr) \mathcal{P}_{l|m|}(0) \\ &= \sum_{l=L+1}^{\infty} C_{lm} j_l(kr) \mathcal{P}_{l|m|}(0) \end{aligned} \quad (3.33)$$

Using the “rule of thumb” given in [53], the error is in the order of 1 percent. It should be noted that the truncation error will only be aliased into coefficients of the highest order, due to inherent properties of the spherical Bessel functions.

Due to the structure of the proposed design example and the nature of the spherical Bessel functions, the lower order spherical harmonic modes are sampled by multiple circular arrays, whereas the highest order ones are only visible to one or two circular arrays. As a result, when solving for the sound field coefficients using (3.12) and (3.19), the lower order coefficients are less affected by the approximation and aliasing errors than the higher order coefficients. This trend is shown in Fig. 3.6.

## 3.5 Design examples

In this section we describe (i) a hypothetical design example and (ii) an actual implementation of the proposed array. The purpose of the hypothetical example is to illustrate the procedures to design an array and to theoretically evaluate the

array's capabilities. Then the implemented array is used to validate the technique through lab experiments.

### 3.5.1 Hypothetical design example

We consider the case of recording the sound field in a spherical region with a diameter of approximately 1 m, the target frequency band is 50-850 Hz. This design example illustrates the use of pairs of omni-directional microphones to realize differential microphones in the array. We chose this array configuration because its accuracy is worse compared to the design using both omni-directional and first order microphones, due to the presence of differential pattern approximation error mentioned in Section 3.4. The radius of the array is chosen to be 0.46 m, which is close to the size of the region of interest. Thus, for the maximum frequency of 850 Hz and a radius of 0.46 m, the array can pick up sound field harmonics up to the order

$$L = \lceil \frac{ekr}{2} \rceil = 10, \quad (3.34)$$

which means that the outer ring of the array should have at least  $2L+1 = 21$  microphone pairs. Following this manner, we place a series of circular arrays of different radii inside the outer circle. Following the design procedure given in Section 3.3, the radii of the rings are set to be 0.46 m, 0.4 m, 0.34 m, 0.28 m, 0.22 m, 0.16 m and 0.1 m. Thus, the number of microphone pairs on each ring are 21, 19, 17, 13, 11, 9 and 7, respectively.

To evaluate the performance of the proposed array system, we place a single point source of frequency 150 – 1150 Hz at  $(R, \theta, \phi) = (1.6 \text{ m}, 60^\circ, 90^\circ)$ . We use the array to estimate the spherical harmonic coefficients and then reconstructed the sound field. We compare the reconstructed sound field to the original sound field and calculate the overall reproduction error of the system. Figure 3.3 depicts the error for different frequencies. Note that the error is small when the frequency is below 850 Hz, which is the desired maximum frequency for the array. Beyond the upper frequency, the error percentage increases dramatically. The reason is that as the frequency increases, the order of active spherical harmonics also grows. At frequencies above 850 Hz, the number of microphones needed to estimate the higher frequency components are greater than the number of the microphones on the array, thus causing aliasing. Also, the total number of coefficients for each mode  $m$  exceeds the number of circular arrays available, as a result the matrix inversion problems

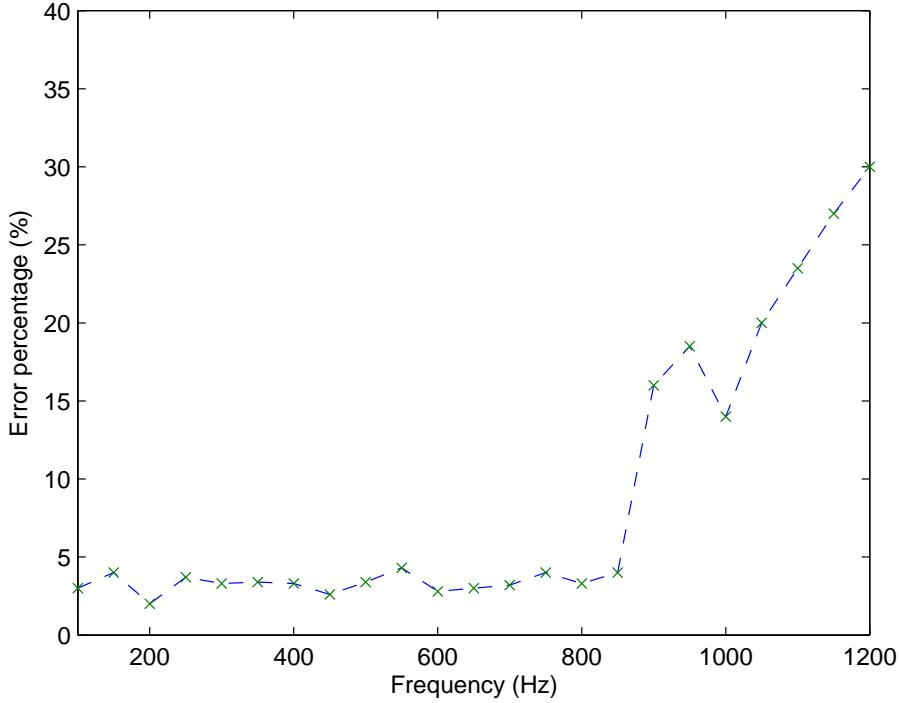


Figure 3.3: Reproduction error percentage for a point source of frequencies 150 – 1150 Hz, located at (1.6 m, 60°, 90°).

shown in (3.12) and (3.19) become under-determined, resulting in significant errors.

We plot the original and reconstructed (using captured spherical harmonic coefficients) sound fields in Fig. 3.4, where plots (a) and (c) are the actual sound field at planes  $z = 0$  and  $z = 0.2$  m, (b) and (d) are the recorded and reconstructed sound field at these two planes, respectively. We observe that the captured sound field over the region of interest in both planes are similar to the actual sound field in the same area.

To evaluate the array performance for different impinging angles, we move a plane wave source at frequency 850 Hz over different elevation angles over  $[0, 180^\circ]$  and the corresponding reproduction error is given in Fig. 3.5. As seen from Fig. 3.5, the error is less than 1.8% over all elevation angles. Due to the symmetry of the array over the azimuth angles, the performance are almost constant over different azimuth angles.

To examine the array accuracy in terms of sound field coefficients, we move a plane wave source at frequency 500 Hz over different elevation angles in the range of  $[0, 180^\circ]$  and calculate the average error for each coefficient, where the theoretical

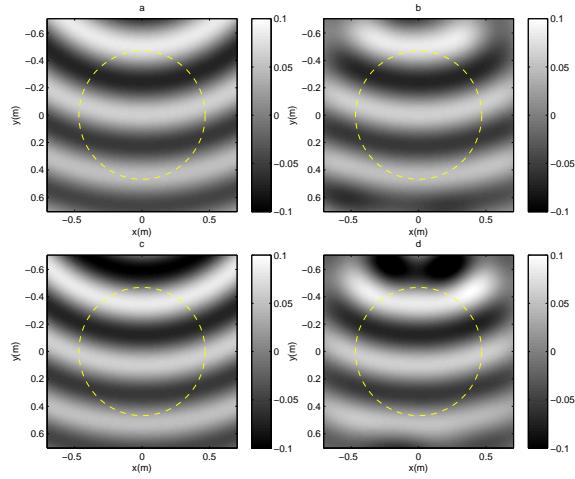


Figure 3.4: Actual (a,c) and recorded (b,d) sound field due to a 850 Hz point source located at  $\theta = 45^\circ$ ,  $R = 1.6$  m, reconstructed at  $z = 0$  (a,b) and  $z = 0.2$  m (c,d) plane.

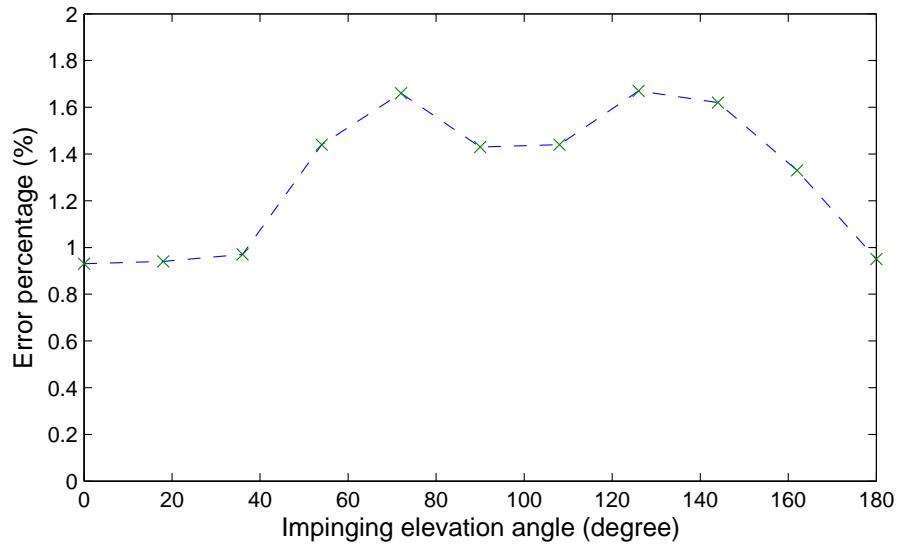


Figure 3.5: Reproduction error percentage for a plane wave source at 850 Hz, moving from  $\theta = 0$  to  $\theta = 180^\circ$ .

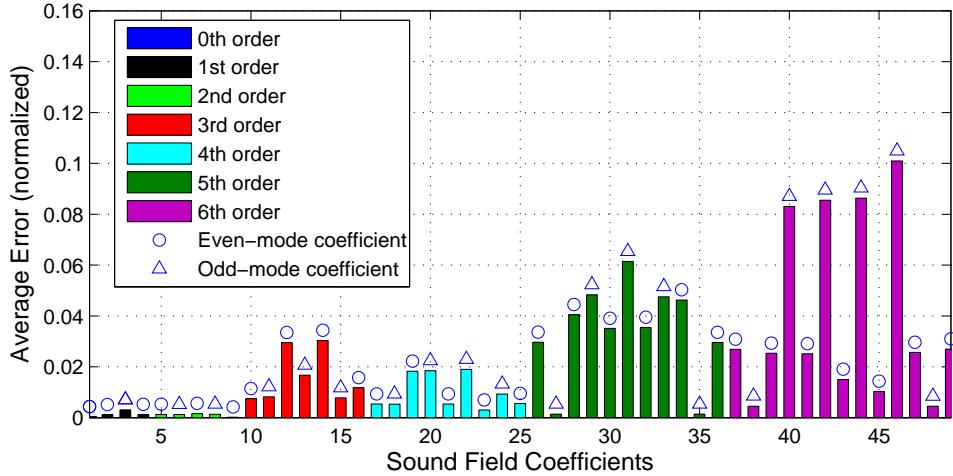


Figure 3.6: Average coefficient error due to a 500 Hz plane wave impinging from different elevation angles.

Table 3.1: Condition number of matrix  $\mathbf{U}_m$  of the hypothetical design example for frequencies 100 Hz, 200 Hz, 400 Hz and 800 Hz.

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$	$m = 10$
100Hz	5.76	1.00	1.00	/	/	/	/	/	/	/	/
200Hz	13.25	4.57	1.00	1.00	/	/	/	/	/	/	/
400Hz	46.30	19.38	15.97	6.33	1.00	1.00	/	/	/	/	/
800Hz	181.35	21.80	110.9	13.24	54.40	41.80	10.88	4.20	7.88	1.00	1.00

coefficient response to a plane wave impinging from  $(\vartheta, \varphi)$  is given by [74]

$$C_{lm} = \sqrt{4\pi} i^l Y_{lm}(\vartheta, \varphi)^*. \quad (3.35)$$

Fig. 3.6 plots the normalized average error for each coefficient. It can be observed that the lower order coefficients are more accurately measured compared to the higher order ones; also, the even mode coefficients are more accurate compared to the odd mode coefficients.

Table 3.1 shows the condition number of the matrix  $\mathbf{U}_m$  of the designed array for various frequencies. Due to the separation of the even and odd mode harmonic coefficients, the coefficients  $C_{L,\pm L}$ ,  $C_{L,\pm(L-1)}$  and  $C_{L-1,\pm(L-1)}$  are solved uniquely, therefore the matrices  $\mathbf{U}_L$  and  $\mathbf{U}_{L-1}$  are in fact vectors whose eigenvalues equal to 1. The size of  $\mathbf{U}_m$  grows as the frequency increases, and the condition number for lower modes increase correspondingly. The design example consists of the minimum number of circular arrays. We expect the condition numbers to be lower should additional circular arrays be used in the system. Also, for high order systems ( $L \geq 5$ ), regularization should be applied when inverting the matrix  $\mathbf{U}_m$ .

In general, we can see from the simulations that the design example offers good accuracy, with its error in the order of 1 percent. This is comparable to the performance of spherical microphone arrays [43] and other previously proposed array configurations such as the multiple circular microphone array [13] and the double sided cone array [71] of the same order, assuming that a similar number of microphones have been used in each array configuration.

### 3.5.2 Array implementation

In order to experimentally test the proposed array design and the associated algorithms, we built a physical array of omni-directional microphones (see Fig. 3.7). The microphones used are Panasonic WM-61B electret microphones, which have a flat frequency response for the whole audible frequency band, and a sensitivity tolerance of  $\pm 4$  dB. Due to hardware limitations, we only use 16 microphones to build the array. Therefore, the array is designed to detect up to the 2nd order sound field for up to 1000 Hz frequency. Based on the proposed design procedure, the system is built to have two co-centered circular arrays, the outer ring has the radius of 10 cm, consisting of 5 omnidirectional microphone pairs, while the inner ring is 4 cm in radius, and consists of 3 microphone pairs.

Testing of the microphone array was conducted in our acoustic lab. A series of factors contribute to the errors in the test results. First of all, although most rigid surfaces in the lab are covered by acoustic foams to reduce reverberation, the acoustic foams are relatively thin and thus reverberations still exist. Secondly, the microphone capsules used have a sensitivity variation of approximately 6 dB, and the calibration process could not guarantee high uniformity among all the microphone units. This factor has a significant impact on the performance of the differential microphone pairs. Furthermore, the position of each microphone unit has a deviation of 1-2 mm, which also leads to errors in the acquired data.

In our experiment, the impinging sound fields are due to two loudspeakers that play 850 Hz sine waves. The loudspeakers were placed at  $(R, \theta, \phi) = (1.64 \text{ m}, 45^\circ, 100^\circ)$  and  $(1.5 \text{ m}, 90^\circ, 225^\circ)$ , respectively. In order to evaluate the results of the experiments, the same loudspeaker-microphone array setup is simulated using MatLab. Figure 3.8 plots the recorded sound field (a) and the simulated sound field (b) due to a point source located at  $(R, \theta, \phi) = (1.64 \text{ m}, 45^\circ, 100^\circ)$ . It can be seen from the figure that the recorded sound field is very similar to the simulated result.

Table 3.2 lists the spherical harmonic coefficients calculated from the recorded



Figure 3.7: Implemented planar microphone array, using omni-directional microphone pairs.

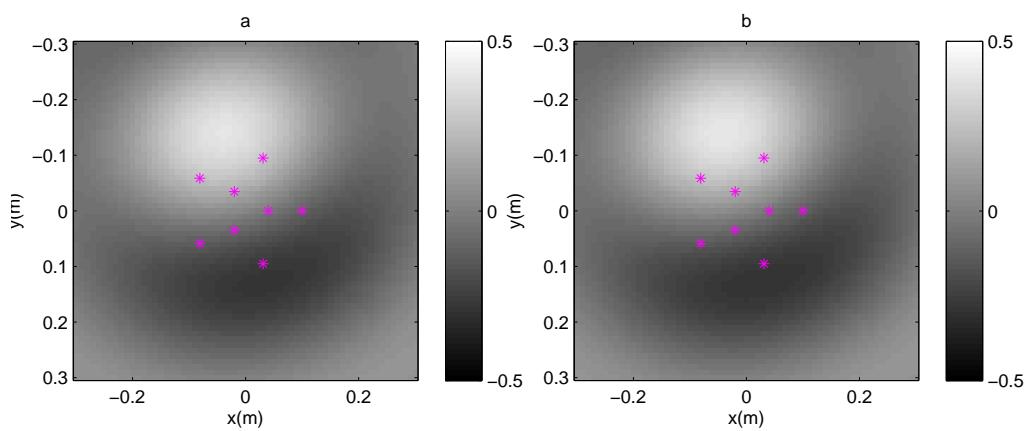


Figure 3.8: Comparison of (a) recorded and (b) simulated sound field for a 850 Hz source at  $(R, \theta, \phi) = (1.64 \text{ m}, 45^\circ, 100^\circ)$ , reconstructed at the  $z = 0.05 \text{ m}$  plane. Microphone locations are marked with “\*”.

Table 3.2: sound field coefficient comparison between simulation and experimental results, the sound fields are due to a point source located at  $(R, \theta, \phi) = (1.64 \text{ m}, 45^\circ, 100^\circ)$  and  $(1.5 \text{ m}, 90^\circ, 225^\circ)$ , respectively.

Recording 1	$C_{0\ 0}$	$C_{1\ (-1)}$	$C_{1\ 0}$	$C_{1\ 1}$	$C_{2(-2)}$	$C_{2(-1)}$	$C_{2\ 0}$	$C_{2\ 1}$	$C_{2\ 2}$
Recorded	1.5413	0.9608	0.9936	0.8351	0.9566	0.7312	1.0255	1.3735	0.5825
Simulated	1.1079	0.9569	1.1892	0.9626	0.8198	1.4140	0.5095	1.4513	0.6788
Mag. Error	0.5142	0.0040	0.1645	0.1325	0.1668	0.4829	1.1026	0.0536	0.1418
Phase Error	0.0009	0.0226	0.1510	0.0888	0.4383	0.3350	0.0145	0.1566	0.3904
Recording 2	$C_{0\ 0}$	$C_{1\ (-1)}$	$C_{1\ 0}$	$C_{1\ 1}$	$C_{2(-2)}$	$C_{2(-1)}$	$C_{2\ 0}$	$C_{2\ 1}$	$C_{2\ 2}$
Recorded	1.7838	1.5167	0.0968	1.5957	1.7211	0.1066	0.6772	0.1121	1.3902
Simulated	1.2380	1.4368	0	1.4137	1.6001	0	1.3127	0	1.8995
Mag. Error	0.5457	0.0798	/	0.1820	0.1209	/	-0.6355	/	-0.5094
Phase Error	-0.0515	0.1143	/	-0.1372	-0.1365	/	-3.9421	/	-0.4036

data as well as those acquired from the simulation results. It can be seen that although rather significant errors occur with some coefficients, the general patterns match very well. The microphone data used are raw recordings processed by microphone calibration data which was acquired before assembling the array, therefore all the errors mentioned previously are present and have an impact on the recorded coefficients. Further calibration to the system, including microphone gain calibration, array geometry adjustments and modification of algorithm parameters can be expected to greatly improve the accuracy of the system.

We would like to point out that our array system utilizes 16 microphones to capture 2nd order sound field, whereas in theory, the minimum number of microphones required to capture second order sound field is 9. Therefore, the proposed array system does not reduce the number of microphones required to sample the sound field. The highlight of our proposed array structure is that it reduces the physical dimension of a higher order microphone array system without compromising its functionality.

## 3.6 Summary

This chapter first introduces a method of measuring complete 3D sound field information on a 2D plane, through the combined use of omnidirectional microphones and first order microphones. Two options are provided for planar microphone array implementation based on the proposed sound field measuring method. Both array configurations consist of multiple co-centered circular arrays, with one option using both omni-directional microphones and first order microphones, while the other option using omni-directional microphones only. The associated algorithms to calculate sound field coefficients are also given in the chapter. We show in the

simulation example that the proposed 2D microphone array system has good accuracy within its designed operating frequency band, and both even and odd sound field coefficients can be accurately calculated. We also built an experimental planar microphone array to further validate the proposed theory.

### 3.7 Related patents and publications

The following patent is related to the work in this chapter.

- H. Chen, T.D. Abhayapala, and W. Zhang, “Planar sensor array”, International (PCT) Patent Application No. PCT/AU2015/000413.

This chapter’s work has been published in the following journal paper. [75]

- H. Chen, T. D. Abhayapala, and W. Zhang, “Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis,” *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015.

# Chapter 4

## 3D sound field analysis using circular higher order microphone array

**Overview:** This chapter proposes the theory and design of circular higher-order microphone arrays for 3D sound field analysis using spherical harmonics. Through employing the spherical harmonic translation theorem, the local spatial sound fields recorded by each higher-order microphone placed in the circular arrays are combined to form the sound field information of a large global spherical region. The proposed design reduces the number of the required sampling points and the geometrical complexity of microphone arrays. We develop a two-step method to calculate sound field coefficients using the proposed array structure, i) analytically combine local sound field coefficients on each circular array and ii) solve for global sound field coefficients using data from the first step. Simulation and experimental results show that the proposed array is capable of acquiring the full 3D sound field information over a relatively large spherical region with decent accuracy and computational simplicity, hence suitable for spatial ANC applications especially over large regions.

### 4.1 Introduction

A higher-order microphone is capable of measuring the local sound field within its proximity, and extracting the sound field coefficients up to a certain spherical harmonics order. It has been shown that the sound field over a large region can be recorded using a number of higher order microphones in a spherical geometry

[76]. Compared to using omnidirectional microphones for the same purpose, the higher order microphone array proposed in [76] requires significantly less number of individual microphone units, thereby reducing the complexity of system deployment especially for spatial sound recording over a large region.

In Chapter 3 we introduced a planar microphone array geometry consisting of differential microphone pairs, which is capable of recording 3D spatial sound. A differential microphone pair can also be seen as a special kind of higher-order microphone, since the sound pressure and pressure gradient it captures are related to the 0th order and 1st order spherical harmonic coefficients as shown in Theorem 1. Intuitively, if differential microphone arrays arranged on a plane can capture 3D sound field, then general higher-order microphones should also have this capability. In this Chapter, we present an algorithm to capture 3D sound field using circular arrays of higher order microphones, placed on a 2D plane. Compared to [76], this method requires simpler microphone geometry, thus reduces the implementation difficulty of higher order microphone arrays for the purpose of large area sound field recording. This method can be seen as a generalization of the algorithm discussed in Chapter 3.

## 4.2 Sound field model

For clarity, in this section, we refer to the sound field with origin  $\mathcal{O}$  as the global sound field, which can be expressed using spherical harmonics using 2.1; the corresponding coefficients  $C_{lm}$  are considered as the global sound field coefficients.

In addition, we define a local origin  $\mathcal{O}_q$  whose position with respect to  $\mathcal{O}$  is  $\mathbf{R}_q = (R_q, \theta_q, \phi_q)$ , then the sound pressure at a point  $\mathbf{r} = (r, \vartheta, \varphi)$  with respect to  $\mathcal{O}_q$  can be expressed by

$$P(r, \vartheta, \varphi) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} B_{\nu\mu}(k) j_{\nu}(kr) Y_{\nu\mu}(\vartheta, \varphi), \quad (4.1)$$

where  $B_{\nu\mu}(k)$  represent the sound field coefficients with respect to the local origin  $\mathcal{O}_q$ . The sound field with respect to  $\mathcal{O}_q$  is called the local sound field.

Using the spherical harmonic addition theorem (2.12), the relationship between

$B_{\nu\mu}$  and  $C_{lm}$  can be written as

$$B_{\nu\mu} = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm} \widehat{S}_{lm}^{\nu\mu}(\mathbf{R}_q). \quad (4.2)$$

In (4.2),  $B_{\nu\mu}$  are the local sound field coefficients in (4.1) and  $C_{lm}$  are the global sound field coefficients in (2.1).

## 4.3 Higher-order microphone array

### 4.3.1 Higher-order microphone

A higher-order microphone is capable of measuring the local sound field within its proximity, and extracting the sound field coefficients up to a certain spherical harmonics order. Thus if a higher-order microphone of order  $V$  is placed at a local origin  $\mathcal{O}_q$ , the sound pressure at a point close to  $\mathcal{O}_q$  can be expressed by a limited summation of spherical harmonics

$$P_q(r, \vartheta, \varphi) = \sum_{\nu=0}^V \sum_{\mu=-\nu}^{\nu} B_{\nu\mu} j_{\nu}(kr) Y_{\nu\mu}(\vartheta, \varphi). \quad (4.3)$$

A total of  $(V + 1)^2$  spherical harmonics and their respective weighing coefficients are present in the summation.

### 4.3.2 Continuous circular higher-order microphone array

A concept of continuous circular microphone array has been proposed in [13]. In this work, this concept is extended for the higher-order microphone case.

Consider a continuous distribution of  $V$ th order microphones are placed along a circle  $(R_s, \vartheta_s)$ , then each higher-order microphone, at a particular azimuth angle  $\varphi$ , is able to detect its local sound field coefficients, denoted as  $B_{\nu\mu}(\varphi)$ . The relationship between  $B_{\nu\mu}(\varphi)$  and  $C_{nm}$  is given by the following theorem:

**Theorem 2.** *Given a set of local sound field coefficients  $B_{\nu\mu}(\varphi)$  which are measured along a circle, and an integer  $m'$ , their relationship with the global sound field*

coefficients can be given by

$$\int_0^{2\pi} B_{\nu\mu}(\varphi) E_{m'}(\varphi) d\varphi = \sum_{l=|\mu-m'|}^{\infty} C_{l(\mu-m')} H_{l(\mu-m')}^{\nu\mu}(R_s, \vartheta_s). \quad (4.4)$$

where

$$H_{lm}^{\nu\mu}(R_s, \vartheta_s) = 4\pi i^{\nu-l} \sum_{\ell=|\mu-m|}^{l+\nu+1} i^\ell (-1)^{2m-\mu} j_\ell(kR_s) \mathcal{P}_{\ell|\mu-m|}(\vartheta_s) W, \quad (4.5)$$

with the definition of  $W$  given by (2.14).

*Proof.* Using (2.2), (2.13) can be rewritten with  $\mathbf{R}_s = (R_s, \vartheta_s, \varphi)$ ,

$$\widehat{S}_{lm}^{\nu\mu}(R_s, \vartheta_s, \varphi) = H_{lm}^{\nu\mu}(R_s, \vartheta_s) E_{(m-\mu)}(\varphi), \quad (4.6)$$

where  $H_{lm}^{\nu\mu}(R_s, \vartheta_s)$  is given by (4.5). Substituting (4.6) into (2.12) yields

$$B_{\nu\mu}(\varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm} H_{lm}^{\nu\mu}(R_s, \vartheta_s) E_{(m-\mu)}(\varphi). \quad (4.7)$$

Multiplying both sides of (4.7) with  $E_{m'}(\varphi)$  and integrating with respect to  $\varphi$  over  $[0, 2\pi]$ , due to the orthogonality property of complex exponential functions

$$\int_0^{2\pi} E_{(m-\mu)}(\varphi) E_{m'}(\varphi)^* d\varphi = \delta_{m-\mu, m'}, \quad (4.8)$$

the integration  $\int_0^{2\pi} C_{lm} H_{lm}^{\nu\mu}(R_s, \vartheta_s) E_{(m-\mu)}(\varphi) E_{m'}(\varphi) d\varphi$  is non-zero only when  $m = \mu - m'$ , thus (4.7) reduces to (4.4), which completes the proof.  $\square$

By replacing  $B_{\nu\mu}(\varphi)$  with  $B_{\nu\mu}(\varphi_q)$ , the discrete form of (4.4) can be written as

$$\frac{1}{Q} \sum_{q=1}^Q B_{\nu\mu}(\varphi_q) E_{\mu-m}(\varphi_q) \approx \sum_{l=|m|}^{\infty} C_{lm} H_{lm}^{\nu\mu}(R_s, \vartheta_s), \quad (4.9)$$

where  $Q$  is the number of sampling points evenly distributed on the circle. In (4.9), the variable  $m'$  has been replaced by  $(\mu - m)$  to illustrate the direct relationship between  $B_{\nu\mu}$  and  $C_{lm}$ . Due to the spatial sampling, an upper bound for the range of  $(\mu - m)$  that can be evaluated is given by

$$|\mu - m| \leq \lfloor \frac{(Q-1)}{2} \rfloor. \quad (4.10)$$

### 4.3.3 Solving for global coefficients

A method for calculating the global sound field coefficients  $C_{lm}$  up to order  $L$  using the local coefficients  $B_{\nu\mu}(\varphi_q)$  can be formulated based on (4.9).

Step 1 of the method is to evaluate the summation on the left hand side of (4.9). For each existing global sound field mode  $m$ , evaluate the summation for all combinations of  $B_{\nu\mu}(\varphi_q)$  and  $m$  that satisfy (4.10). Denote the summation as  $\alpha_{\nu\mu}^m$ , then

$$\alpha_{\nu\mu}^m = \frac{1}{Q} \sum_{q=1}^Q B_{\nu\mu}(\varphi_q) E_{(m-\mu)}(\varphi_q). \quad (4.11)$$

The second step is to solve a matrix inversion problem to find  $C_{lm}$ . Using (4.9) and (4.11), the relationship between  $C_{lm}$  and  $\alpha_{\nu\mu}^m$  can be represented in matrix form as

$$\boldsymbol{\alpha}_m = \mathbf{H}_m \mathbf{C}_m, \quad (4.12)$$

where  $\boldsymbol{\alpha}_m = [\alpha_{00}^m \ \alpha_{1(-1)}^m \ \alpha_{10}^m \ \dots \ \alpha_{\nu\mu}^m]^T$ , and  $\mathbf{C}_m = [C_{|m|m} \ C_{(|m|+1)m} \ \dots \ C_{Lm}]^T$  is the set of global coefficients of mode  $m$ .

$$\mathbf{H}_m = \begin{bmatrix} H_{|m|m}^{00} & H_{(|m|+1)m}^{00} & \dots & H_{Lm}^{00} \\ H_{|m|m}^{1(-1)} & H_{(|m|+1)m}^{1(-1)} & \dots & H_{Lm}^{1(-1)} \\ \vdots & \vdots & \ddots & \vdots \\ H_{|m|m}^{\nu\mu} & H_{(|m|+1)m}^{\nu\mu} & \dots & H_{Lm}^{\nu\mu} \end{bmatrix}$$

is the matrix that contains the weights for spherical harmonics translation. A solution for  $\mathbf{C}_m$  can be found by calculating the Moore-Penrose Pseudo Inverse of  $\mathbf{H}_m$ . The size of  $\mathbf{H}_m$  is  $(V + 1)^2$  by  $(L - |m| + 1)$ , which is significantly smaller than the  $(L + 1)^2$ -by- $(L + 1)^2$  matrix inversion proposed in [58], thus both the computational simplicity and the condition of the matrix inversion are significantly better compared to the method in [58].

The complete set of global sound field coefficients is found by solving (4.12) for  $m = [-L : L]$ , where  $L$  is the maximum order of the global sound field.

Implementing multiple circular higher-order microphone arrays in the global region can improve the robustness and precision of the microphone system. Assuming a total number of  $K$  circular arrays are implemented, then in order to calculate the global coefficients, one needs to formulate (4.12) for each circular array, denoted as

$\boldsymbol{\alpha}_{m;K} = \mathbf{H}_{m;K} \mathbf{C}_m$ , then the solution for  $\mathbf{C}_m$  can be expressed as

$$\mathbf{C}_m = (\widehat{\mathbf{H}}_m^H \widehat{\mathbf{H}}_m + \lambda \mathbf{I})^{-1} \widehat{\mathbf{H}}_m^H \widehat{\boldsymbol{\alpha}}_m \quad (4.13)$$

where  $\widehat{\mathbf{H}}_m = [\mathbf{H}_{m;1}^T \mathbf{H}_{m;2}^T \dots \mathbf{H}_{m;K}^T]^T$ ,  $\lambda$  is the regularization parameter, and  $\widehat{\boldsymbol{\alpha}}^m = [\boldsymbol{\alpha}_{m;1}^T \boldsymbol{\alpha}_{m;2}^T \dots \boldsymbol{\alpha}_{m;K}^T]^T$ . Evaluating (4.13) for  $m = [-L : L]$  yields the complete set of global sound field coefficients.

#### 4.3.4 Dimensionality analysis

Due to the nature of Spherical Bessel functions, only a number of  $j_l(kR_s)$  are active within a certain radius. From (2.7) and the range of  $\ell$  in (4.5), we can derive the maximum global spherical harmonic order detectable by a circular  $V_{th}$  order microphone array

$$L = V + \lceil \frac{ekR_s}{2} \rceil, \quad (4.14)$$

where  $L$  is the maximum global sound field order detectable and  $R_s$  is the radius of the circular array.

The minimum number of sampling points  $Q$  on a circle can be derived from eqs (4.10) and (2.7), using  $\ell \geq |\mu - m|$ ,

$$Q \geq 2 \lceil \frac{ekR_s}{2} \rceil + 1. \quad (4.15)$$

### 4.4 Simulation results

A series of simulations have been conducted to validate the performance of the proposed array structure. Two instances of the proposed array structure are used in the following simulations. Both array configurations are designed to capture sound fields up to 700 Hz within a sphere with 0.5m radius, with their dimensions determined based on eqs (4.14) and (4.15). Multiple circular arrays are employed in both cases to guarantee the quality of the matrix inversion in (4.13). One design consists of first order microphones arranged into four circular arrays, positioned at  $(R_s, \vartheta_s) = (0.4, 90^\circ), (0.34, 72^\circ), (0.28, 108^\circ)$  and  $(0.22, 72^\circ)$ , the number of first order microphones on each array is 17, 15, 13 and 11, respectively. The second design utilizes only second order microphone arrays, with two circular arrays located at  $(R_s, \vartheta_s) = (0.4, 90^\circ)$  and  $(0.2, 72^\circ)$ , with 17 and 9 second order microphones placed on each array. AWGN is added to the microphone input of all simulations with a

SNR of 40 dB. A point source is placed at  $(R, \theta, \phi) = (1.6, 60^\circ, -60^\circ)$  for all the simulation setups.

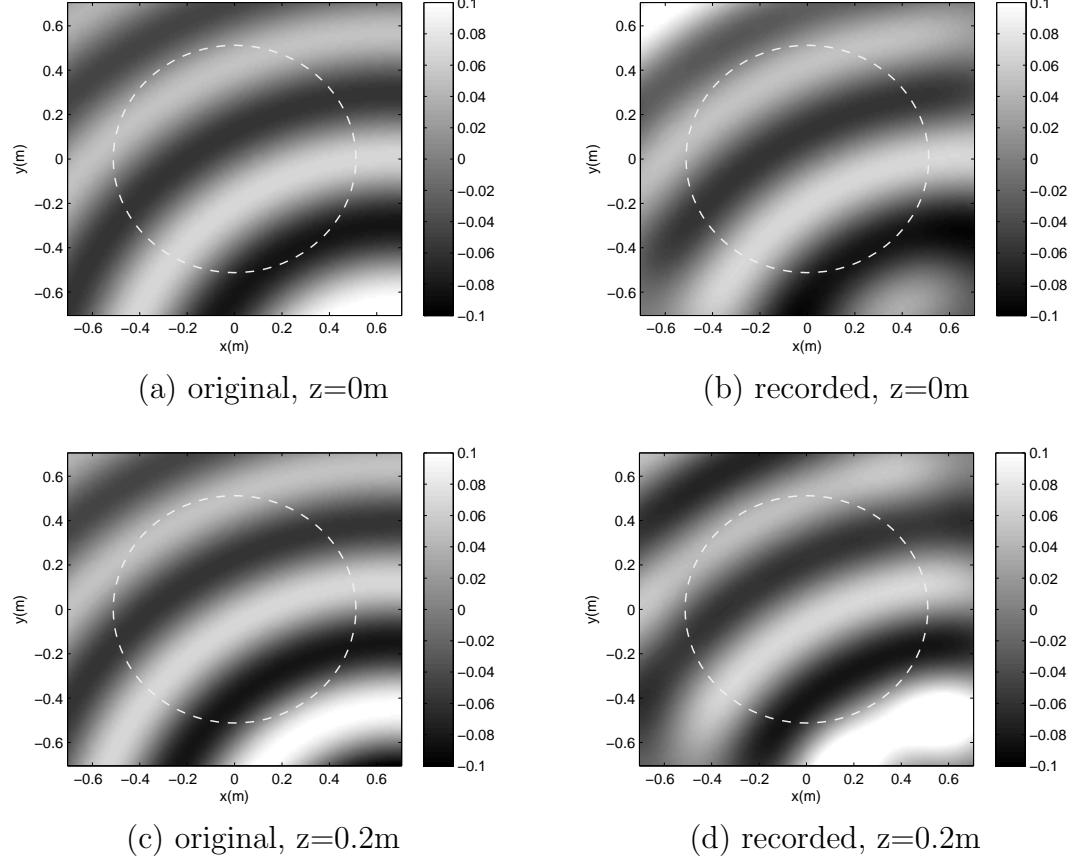


Figure 4.1: Comparison of original and recorded sound field due to 700 Hz point source, reconstructed at  $z=0$  and  $z=0.2$  m plane

Figure 4.1 shows the simulation result for the first order array configuration. In this simulation, the sound field generated by the point source is recorded by the array, and the resulting sound field coefficients are used to reconstruct the sound field. The sound field is plotted for two layers: the  $z = 0$  plane and  $z = 0.2$  m plane. Plots (a) and (c) show the original sound field at these two planes, and plots (b) and (d) show the reconstruction of the sound field coefficients obtained from the microphone array. The result shows that the microphone array is capable of accurately capture the sound field within its coverage (yellow circle).

Figure 4.2 depicts the error performance for two different array configurations at a frequency range of 100 – 1000 Hz. For this figure, the error is calculated by averaging the amplitude error over the entire region of interest, and normalizing by the average sound pressure in the same region. Since both array configurations

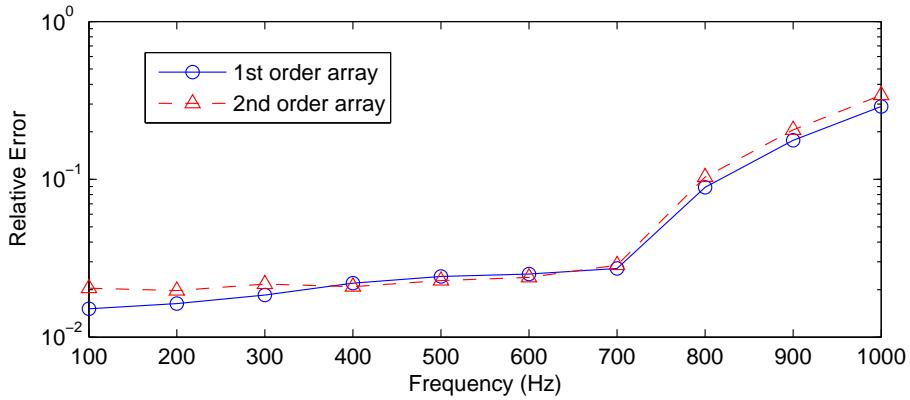


Figure 4.2: Reproduction error at different frequencies for first and second order array configurations

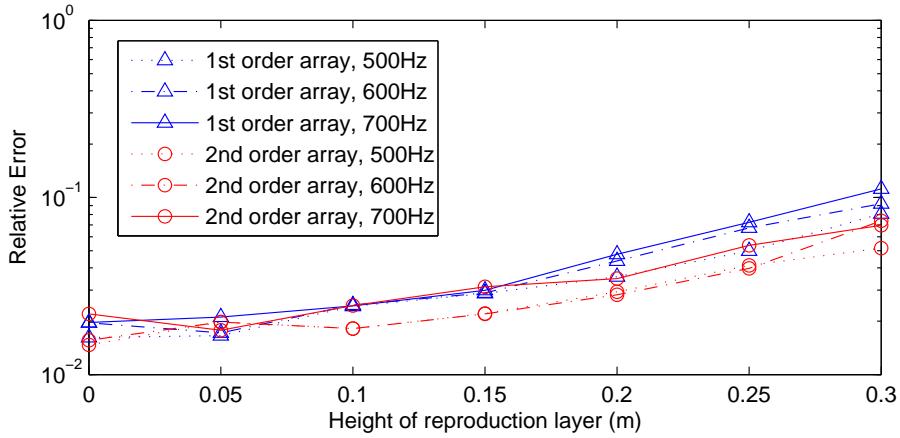


Figure 4.3: Reproduction error at different elevations and frequencies for first and second order array configurations

are designed to operate at up to 700 Hz, it can be seen from Fig. 4.2 that the reproduction error for both configurations are low for frequencies below 700 Hz, and the error increases rapidly once the frequency becomes higher than the design frequency.

The reproduction error is also evaluated at different planes using the same method, but with the region limited to horizontal planes within the spherical area. The results are shown in Fig. 4.3. The recorded sound field is reconstructed on planes of different heights, ranging from  $z = 0$  to  $z = 0.3$  m. The simulation shows that the reproduction error is smaller around the equator compared to that near the poles of the sphere, which is due to the fact that the microphones are clustered around the equator plane.

## 4.5 Experimental results

In order to further validate the proposed method, we conducted an experiment of recording a three-dimensional sound field using higher order microphone. We use a single Eigenmike as a 4th order microphone, which consists of 32 condenser microphone capsules placed on a rigid sphere of 4.2cm radius. The goal of the experiment is to test the robustness of the algorithm with the presence of noise and interference in a real-life system.

The region of interest is set to be a sphere of 25cm diameter, a loudspeaker is placed at  $(R, \theta, \phi) = (1.5\text{m}, 90^\circ, 30^\circ)$  with respect to the center of the region of interest. The Eigenmike is placed on the equatorial plane of the spherical region and is moved around a circle of 10cm radius. A total of 25 sampling points are evenly distributed along the circle. At each sampling point, the Eigenmike records a sweeping signal played by the loudspeaker, which is then converted to a set of 4th order spherical harmonic coefficients. The 25 sets of local coefficients are then combined using the proposed method to compute the 13th order global coefficients. A visualization of the reconstructed sound field at 3500 Hz within the region of interest is shown in Fig. 4.4.

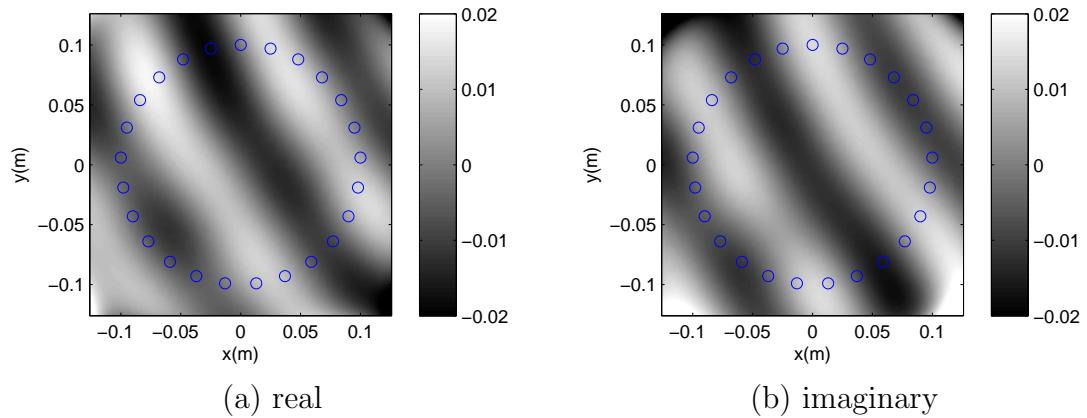


Figure 4.4: Reconstructed sound field at 3500Hz due to a loudspeaker placed at  $(1.5\text{m}, 90^\circ, 30^\circ)$ , sampling points are indicated by blue circles.

We have identified the primary causes of error to be the sensitivity variation of the microphone capsules, and the reverberation inside the laboratory. Despite said interferences, the Eigenmike was able to record the sound field with acceptable accuracy.

We believe that the proposed spatial sampling method allows for easier implementation of sound field recording systems compared to spherical sampling methods, especially when combined with the recording technique used in this experiment, and for applications such as room response modelling over a large space.

## 4.6 Summary

In this chapter, we propose a circular higher-order microphone array structure and an associated analytical algorithm for sound field analysis based on spherical harmonics decomposition. This method can be seen as a generalization of the planar microphone array proposed in Chapter 3. In this method, through employing the spherical harmonic translation theorem, the local spatial sound fields recorded by each higher-order microphone placed in the circular arrays are combined to form the sound field information of a large global spherical region. The proposed design reduces the number of the required sampling points and the geometrical complexity of microphone arrays. Simulations and experiments show that the proposed array architecture offers decent accuracy and robustness, and has the potential of simplifying sound field recording systems in certain applications.

## 4.7 Related Publications

This chapter's work has been published in the following conference proceeding [77]:

- H. Chen, T. D. Abhayapala, and W. Zhang, “3D sound field analysis using circular higher-order microphone array,” in *Proc. 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 1153–1157.

# Chapter 5

## Direct-to-reverberant energy ratio estimation using a first order microphone

**Overview:** The Direct-to-Reverberant Ratio (DRR) is an important characterization of a reverberant environment. In the context of spatial ANC, DRR helps to determine the strength of reverberation within the noise field's composition. This chapter presents a novel blind DRR estimation method based on coherence function of sound pressure and particle velocity. First, a general expression of coherence function and DRR is derived in the spherical harmonic domain, without imposing assumptions on the reverberation. In this work, DRR is expressed in terms of the coherence function as well as two parameters which are related to statistical characteristics of the reverberant environment. Then, a method to estimate the values of these two parameters using a microphone system capable of capturing first order spherical harmonics is proposed, under three assumptions which are more realistic than the diffuse field model. Furthermore, a theoretical analysis on the use of plane wave model for direct path signal and its effect on DRR estimation is presented, and a rule of thumb is provided for determining whether the point source model should be used for the direct path signal. Finally, the ACE Challenge Dataset is used to validate the proposed DRR estimation method. The results show that the average full band estimation error is within 2 dB, with no clear trend of biasing.

## 5.1 Introduction

The direct-to-reverberation energy ratio (DRR), defined as the energy ratio between direct signal and its reverberations, is an important parameter to characterize a reverberant environment, along with other parameters such as reverberation time. Since reverberation energy affects the speech signal's clarity [78], the DRR has an influence on the algorithms for various applications such as speech dereverberation [79], teleconferencing [80] and hearing aids [81], both in terms of algorithm performance and strategy. The minimum audible difference in DRR has been investigated in [82]. In [83], DRR is utilized for parametric spatial audio coding. DRR also finds its application in the field of psychoacoustics, where it is believed that DRR helps human to determine the distance of the sound source [78, 84, 85].

DRR estimation methods based on estimating room impulse responses have been presented by Larsen *et al.* [86] and Falk *et al.* [87]. However, pre-processing is required for both methods. Mosayyebpour *et al.* [88] presented a method for blind DRR estimation based on higher order statistics, where the inverse filter of the room impulse response is estimated using the skewness of the speech signal. Parada et. al. presented a single channel DRR estimation method base on a neuron network learning algorithm [89].

Methods for blind DRR estimation using multiple sensors have also been proposed in the literature. With the goal of estimating source distance, Lu [90] presented a DRR estimation algorithm using the equalization-cancellation method, where a binaural microphone system is used to capture sound signal. The coherence function framework was first introduced by Vesa [91] for estimating source distance using binaural signals, where the coherence function of the two input signals was used as a characterization of source distance. Later, the coherence function framework was also used by Jeub [92] to develop a DRR estimation algorithm. In this work, the DRR is estimated by comparing coherence value computed from two microphone inputs with theoretical coherence functions in a diffuse sound field. Thiergart [93] also developed a DRR estimation algorithm based on the complex coherence function of two omnidirectional microphones. In [94] a DRR estimation method based on spectra standard deviation of two microphones was proposed.

Directional or beam forming microphone arrays have also been used to estimate DRR, such as the methods presented in [95] and [96]. In both of these works, the power spectral density (PSD) of the reverberant field were used to estimate DRR. Another method [97] uses a circular microphone array to estimate DRR, the method

relies on the spatial correlation matrix of the microphones' received signals. The reverberation is modelled as a diffuse field in this work, while the direct path is assumed to be a plane wave. The DRR is solved using a least mean square method. Kuster [98] presented a method based on coherence function of sound pressure and particle velocity at the receiver position, measured by a differential microphone array.

In recent years, the use of higher order microphones and the technique of spherical harmonic decomposition [10] have become popular in the field of room acoustic analysis. Jarrett *et al.* [12] proposed a method to estimate Signal-to-Diffuse Ratio (SDR, equivalent to DRR when assuming diffuse reverberation field) utilizing spherical harmonic coefficients captured by a higher order microphone. It is shown that this method minimizes the SDR estimation bias. In our previous work [99], we implemented Kuster's method [98] in spherical harmonic domain, utilizing the first order spherical harmonic coefficients to estimate DRR.

In many of the previous works, such as [98], [12], [93] and our previous work [99], the direct path signal is assumed to be plane wave, and the reverberant sound field is assumed to be diffuse field. In real-life reverberant environments where these assumptions may not hold, the DRR estimation accuracy of these algorithms may degrade. For example, the DRR estimated using Kuster's method tend to be higher than ground truth in reverberant rooms [98].

In this work, we first develop a general expression for DRR estimation using the coherence function of sound pressure and particle velocity, using a point source model for the direct path signal, and without applying any assumptions for the reverberation field. Using the relationship between spherical harmonic coefficients and acoustic particle velocity, we develop the framework in the spherical harmonics domain. Then, for the direct path model, we provide a detailed analysis on the error in DRR estimation results when using the plane wave model. We propose a rule of thumb for determining whether the plane wave model can be used without introducing significant error, based on the source-to-microphone distance and target frequency. For the reverberation sound field, we show that the reverberation characteristics related to DRR estimation can be expressed using two parameters, and that under the diffused field assumptions, the values of these parameters can be determined theoretically, which results in the simplified DRR solutions in [98] and [12]. We also provide a theoretical analysis on the two parameters, their physical meanings, and their impact on the DRR estimation, which explains the positive bias phenomenon of Kuster's method [98]. Furthermore, we propose a method to

estimate these two parameters for a given reverberant environment, using a first order microphone, under a number of assumptions on the reverberant field which are less strict than the diffuse field model. The DRR can then be calculated using the estimated parameters.

The performance of the proposed DRR estimation algorithm is verified using the ACE Challenge Dataset [100]. It is shown that the results agree with the theoretical analysis, and that the proposed method addresses the positive bias problem of Kuster's method [98], and the mean DRR estimation error is less than 2 dB for all recording scenes in the ACE Challenge Dataset.

## 5.2 DRR estimation based on coherence measurements

### 5.2.1 Representation of reverberant sound field

For convenience, the spherical coordinate system is defined such that its origin is at the position of the microphone, and its positive  $z$  axis points towards the impinging direction of the direct path signal. In many scenarios, the natural coordinate system may have a different orientation than our definition. In such cases, the spherical harmonics defined under a different coordinate system (with the same origin) can be transformed into our desired coordinate system using the spherical harmonic rotation, which is described in Chapter. 2.2.3 .

The sound pressure at a point  $(r, \theta, \phi)$  close to the origin can be decomposed using (2.1). For the direct path, we have

$$P_D(r, \theta, \phi, k) = \sum_{l=0}^1 \sum_{m=-l}^l B_{lm}(k) j_l(kr) Y_{lm}(\theta, \phi), \quad (5.1)$$

where only the first order sound field is considered.

For the sound field due to reverberation, we have

$$P_R(r, \theta, \phi, k) = \sum_{l=0}^1 \sum_{m=-l}^l \alpha_{lm}(k) j_l(kr) Y_{lm}(\theta, \phi), \quad (5.2)$$

where  $B_{lm}(k)$  and  $\alpha_{lm}(k)$  represent the coefficients of the direct path and the reverberant sound field, respectively.

The following assumptions are made regarding the direct path sound:

- 1: The direct path is due to a point source located at  $(r_0, \vartheta, \varphi)$ .
- 2: The direct path signal  $P_D(r, \theta, \phi, k)$  is uncorrelated with the reverberant sound field  $P_R(r, \theta, \phi, k)$ . Using (5.1) and (5.2), this assumption can be expressed as

$$E\{B_{lm}\alpha_{l'm'}^*\} = 0, \text{ for all } l \text{ and } m. \quad (5.3)$$

where  $E\{\cdot\}$  denotes the expectation operator.

Since the direct path signal is modelled as sound waves emitted by a point source,  $B_{lm}(k)$  can be written using the following expression [101]

$$B_{lm}(k) = A_D ik h_l^{(1)}(kr_0) Y_{lm}^*(\vartheta, \varphi), \quad (5.4)$$

where  $A_D$  indicates the magnitude of the impinging sound,  $h_l^{(1)}(kr_0)$  is the  $n$ th order spherical Hankel function of the first kind,  $r_0$  is the distance between the point source and the microphone with  $r_0 > r$ ,  $(r_0, \vartheta, \varphi)$  denotes the position of the point source, and  $(\cdot)^*$  represents complex conjugate. Since the coordinate system is defined such that  $\vartheta = 0$ , and due to the fact that  $Y_{11}(0, \varphi) = Y_{1,-1}(0, \varphi) = 0$ , we have  $B_{11}(k) = B_{1,-1}(k) = 0$ . Thus the combined sound field coefficients  $C_{lm}(k)$  can be expressed as follows

$$C_{00}(k) = A_D ik h_0^{(1)}(kr_0) Y_{00}^*(0, 0) + \alpha_{00}(k), \quad (5.5)$$

$$C_{10}(k) = A_D ik h_1^{(1)}(kr_0) Y_{10}^*(0, 0) + \alpha_{10}(k), \quad (5.6)$$

$$C_{11}(k) = \alpha_{11}(k), \quad (5.7)$$

$$C_{1,-1}(k) = \alpha_{1,-1}(k). \quad (5.8)$$

Equations (5.5)-(5.8) shows that in the coordinate system defined in this section, the direct path signal is only present in  $C_{00}(k)$  and  $C_{10}(k)$ , but not in  $C_{11}(k)$  and  $C_{1,-1}(k)$ .

We note that the four coefficients  $C_{00}(k)$ ,  $C_{10}(k)$ ,  $C_{11}(k)$  and  $C_{1,-1}(k)$  can be captured by a first order microphone. Although in the general sense, microphones with certain directional beam patterns, such as cardioid microphones and differential microphones are commonly referred to as first order microphones, in the context of this section, a first order microphone is a microphone system which is capable of acquiring the 0th and 1st order spherical harmonic coefficients of its surrounding

sound field. Specific directionalities can be realized through applying beam-forming algorithms on the 0th and 1st order coefficients.

### 5.2.2 Representation of DRR using coherence function

The coherence function between the sound pressure  $P(\mathbf{0}, k)$  and particle velocity  $V_z(\mathbf{0}, k)$  along the  $z$  direction can be defined as [98],

$$\gamma^2 \triangleq \frac{|E\{P(\mathbf{0}, k)V_z(\mathbf{0}, k)^*\}|^2}{E\{|P(\mathbf{0}, k)|^2\}E\{|V_z(\mathbf{0}, k)|^2\}}. \quad (5.9)$$

Note that  $P(\mathbf{0}, k) = C_{00}Y_{00}(0, 0) = 1/\sqrt{2\pi}C_{00}$  and  $V_z(\mathbf{0}, k)$  is proportional to  $C_{10} \cdot i^1$  in (2.22). Substituting  $P(\mathbf{0}, k)$  and (2.22) into (5.9), and applying (5.5) (5.6), we have

$$\gamma^2 = \frac{|E\{C_{00}(C_{10} \cdot i)^*\}|^2}{E\{|C_{00}|^2\}E\{|C_{10}|^2\}} \quad (5.10)$$

$$= \frac{|E\{H_0(H_1i)^*\} + E\{\alpha_{00}(\alpha_{10}i)^*\}|^2}{(E\{|H_0|^2\} + E\{|\alpha_{00}|^2\})(E\{|H_1|^2\} + E\{|\alpha_{10}|^2\})}, \quad (5.11)$$

where the assumption that direct path is uncorrelated with the reverberations (5.3) is used, and we denote

$$H_0 \triangleq A_D ik h_0^{(1)}(kr_0)Y_{00}^*, \quad (5.12)$$

$$H_1 \triangleq A_D ik h_1^{(1)}(kr_0)Y_{10}^*. \quad (5.13)$$

Note that the angle arguments ( $\vartheta = 0, \varphi = 0$ ) of  $Y_{nm}(\vartheta, \varphi)$  and the frequency arguments ( $k$ ) of  $C_{nm}$  and  $\alpha_{nm}$  have been omitted for simplicity.

The linear scale direct-to-reverberant energy ratio is defined here to be the ratio of measured acoustic energy at the position of measurement due to the direct path and reverberation, since  $P(\mathbf{0}) = C_{00}Y_{00}$ , we have

$$\text{DRR} = \frac{E\{|P_D(\mathbf{0})|^2\}}{E\{|P_R(\mathbf{0})|^2\}} = \frac{E\{|B_{00}|^2\}}{E\{|\alpha_{00}|^2\}}. \quad (5.14)$$

---

<sup>1</sup>Although removing the imaginary argument  $i$  here does not affect  $\gamma^2$ , we keep  $i$  for the derivation of further expressions.

Using (5.4) to express  $B_{00}$  in (5.14), we have

$$\text{DRR} = \frac{E\{|A_D i k h_0^{(1)}(kr_0) Y_{00}^*|^2\}}{E\{|\alpha_{00}|^2\}} = \frac{E\{|H_0|^2\}}{E\{|\alpha_{00}|^2\}}. \quad (5.15)$$

Substituting (5.15) into (5.10) yields

$$\gamma^2 = \frac{\left| -\text{DRR} \cdot i \left( \frac{h_1^{(1)}(kr_0) Y_{10}^*}{h_0^{(1)}(kr_0) Y_{00}^*} \right)^* + \frac{E\{\alpha_{00}(\alpha_{10})^*\}}{E\{|\alpha_{00}|^2\}} \right|^2}{(\text{DRR} + 1)(\text{DRR} \left| \frac{h_1^{(1)}(kr_0) Y_{10}}{h_0^{(1)}(kr_0)^2 Y_{00}} \right|^2 + \frac{E\{|\alpha_{10}|^2\}}{E\{|\alpha_{00}|^2\}})} \quad (5.16)$$

which relates the coherence value  $\gamma^2$  to the DRR of the room.

For convenience, we define

$$R_1 \triangleq \frac{E\{\alpha_{00}(\alpha_{10})^*\}}{E\{|\alpha_{00}|^2\}} \frac{Y_{00}}{Y_{10}} \quad (5.17)$$

$$R_2 \triangleq \frac{E\{|\alpha_{10}|^2\}}{E\{|\alpha_{00}|^2\}} \frac{Y_{00}^2}{Y_{10}^2} \quad (5.18)$$

as the reverberation parameters, and

$$\mathcal{H} \triangleq \left( \frac{h_1^{(1)}(kr_0)}{h_0^{(1)}(kr_0)} \right)^*. \quad (5.19)$$

Then (5.16) can be simplified as

$$\gamma^2 = \frac{|-\text{DRR} \cdot i \cdot \mathcal{H} + R_1|^2}{(\text{DRR} + 1)(\text{DRR} |\mathcal{H}|^2 + R_2)} \quad (5.20)$$

$$= \frac{|\text{DRR}|^2 |\mathcal{H}|^2 + 2\text{DRR} \cdot \text{Im}\{\mathcal{H} R_1^*\} + |R_1|^2}{(\text{DRR} + 1)(\text{DRR} |\mathcal{H}|^2 + R_2)}, \quad (5.21)$$

where  $\text{Im}\{\cdot\}$  denotes imaginary part of the argument. From (5.20) it can be seen the characteristics of reverberation which affects DRR estimation using coherence method can be expressed using two parameters  $R_1$  and  $R_2$ .

### 5.2.3 Assumptions for the reverberant sound field

#### Plane wave assumption for the direct path

In previous works, the direct path signal is often assumed to be a plane wave [12,98]. Under this assumption, the following approximation can be applied (see Appendix

5.7 for the proof)

$$\lim_{r_0 \rightarrow \infty} \frac{h_1^{(1)}(kr_0)}{h_0^{(1)}(kr_0)} \approx -i, \quad (5.22)$$

and (5.20) can be simplified into

$$\gamma^2 = \frac{|DRR + R_1|^2}{(DRR + 1)(DRR + R_2)} \quad (5.23)$$

$$= \frac{DRR^2 + 2DRR \cdot Re\{R_1\} + |R_1|^2}{(DRR + 1)(DRR + R_2)}, \quad (5.24)$$

where  $Re\{\cdot\}$  denotes real part of the argument. The plane wave assumption leads to bias in the DRR estimation, primarily for lower frequencies and smaller values of  $r_0$ , which is shown in Section 5.3.2.

### Diffuse reverberation assumptions in previous works

In many previous works, the sound field due to reverberation is often modelled as diffused field [12, 98], although the exact definition of diffused field may vary. In [12], the diffuse field is defined as an infinite number of uncorrelated plane waves impinging uniformly from the sphere. Under this assumption, it is shown that  $E\{\alpha_{00}\alpha_{10}^*\} = 0$ , and  $E\{|\alpha_{lm}|^2\} = E\{|\alpha_{l'm'}|^2\}$  for all values of  $l$  and  $m$  [12]. In this case,  $R_1 = 0$ ,  $R_2 = |Y_{00}|^2/|Y_{10}|^2 = 1/3$ , and (5.24) becomes equivalent to the magnitude-squared version of Eq.(18) in [12].<sup>2</sup>

In the case of Kuster's work [98], the reverberant field is assumed to be plane waves whose impinging directions distribute uniformly over  $\theta_{in} \in [0, 2\pi)$ , where  $\theta_{in}$  is the angle between direct path and the plane wave impinging direction. This assumption differs from the reverberant field model used in [12], where plane waves are distributed uniformly over the sphere; this assumption can be fulfilled if the plane waves impinge uniformly over a circle. Under this assumption, Kuster has derived an expression for  $\gamma^2$  which takes the same form as (5.24), but with  $R_1 = 0$ , and  $R_2 = 0.5$  [98].

### Assumptions on reverberation used in this work

In many real acoustic environments, the diffused field assumptions for reverberant field made in [12] and [98] often cannot be met, which may lead to inaccuracies in the DRR estimation result. In this work, in order to improve the accuracy of DRR

---

<sup>2</sup>For  $c_{00}$  and  $c_{10}$ , with  $\Omega_{dir} = (0, 0)$ .

estimation, we relax some assumptions made on the reverberant sound field. In particular, we assume that the reverberant field satisfies the following conditions:

- 1:** The average sound intensity (product of sound pressure and particle velocity) [102] of the reverberant field has the same magnitude in  $x$ ,  $y$  and  $z$  directions.

$$|E\{P^r V_z^{r*}\}| = |E\{P^r V_x^{r*}\}| = |E\{P^r V_y^{r*}\}|. \quad (5.25)$$

where  $P^r$  and  $V^r$  denote sound pressure and particle velocity due to reverberation, respectively.

- 2:** The expected energy of the reverberant field particle velocity is constant in  $x$ ,  $y$  and  $z$  directions.

$$E\{|V_x^r|^2\} = E\{|V_y^r|^2\} = E\{|V_z^r|^2\}, \quad (5.26)$$

- 3:** The reverberant field sound intensity is zero mean when averaged over a frequency band.

$$\int_{k_1}^{k_2} E\{P^r(k)V_z^r(k)^*\} dk = 0. \quad (5.27)$$

where  $k_1$  and  $k_2$  represent the boundary of a frequency band.

In (5.27), the real part of  $P^r V_z^{r*}$  is often referred to as the active sound intensity, which represents the coherent flow of sound energy in the  $z$  direction [102]. The imaginary part of sound intensity, on the other hand, is referred to as the reactive sound intensity, which represents the coherent, but non-propagating, “standing wave” sound energy. A detailed justification of the assumption (5.27) is given in 5.3.1.

In a diffuse sound field, both active and reactive components of the sound intensity are equal to zero since the phase of particle velocity varies randomly. The energy of particle velocity can be analytically computed [12, 98]. Applying these results to (5.23) leads to simplified expressions of  $\gamma^2$  as shown in [12] and [98].

However, this work do not assume diffuse field. Hence the expected energy of particle velocity and the sound intensity cannot be directly computed without the knowledge of the reverberant field. Therefore, a method to estimate these characteristics is needed to compute the DRR. The following subsection describes one such method, using measurements from a first (or higher) order microphone system.

### 5.2.4 Reverberant field estimation

From (5.7) and (5.8), it can be observed that the spherical harmonic coefficients  $\beta_{11}$  and  $\beta_{1,-1}$  do not contain the direct path signal. In fact,  $\beta_{11}$  and  $\beta_{1,-1}$  collectively represent the particle velocity of the reverberations in the directions orthogonal to the direct path. The assumptions on the reverberation (5.25) (5.26) and (5.27) can be expressed using spherical harmonic coefficients as

$$\begin{aligned} \left| \frac{E\{\alpha_{00}(\alpha_{10}i)^*\}}{-i\sqrt{12\pi}} \right| &= \left| \frac{E\{\alpha_{00}(\alpha_{11}i + \alpha_{1,-1}i)^*\}}{-i\sqrt{24\pi}} \right| \\ &= \left| \frac{E\{\alpha_{00}(\alpha_{11}i - \alpha_{1,-1}i)^*\}}{-\sqrt{24\pi}} \right|, \end{aligned} \quad (5.28)$$

$$2 \cdot E\{|\alpha_{10}|^2\} = E\{|\alpha_{11} + \alpha_{1,-1}|^2\} = E\{|\alpha_{11} - \alpha_{1,-1}|^2\}, \quad (5.29)$$

and

$$\int_{k_1}^{k_2} E\{\alpha_{00}(k)(\alpha_{10}(k) \cdot i)^*\} dk = 0. \quad (5.30)$$

Since it is assumed that the direct path signal is uncorrelated with the reverberation signal, substituting (5.3), (5.5), (5.7) and (5.8) into (5.28), we can write

$$\begin{aligned} \sqrt{2}|E\{\alpha_{00}(\alpha_{10}i)^*\}| &= |E\{C_{00}(C_{11}i + C_{1,-1}i)^*\}| \\ &= |E\{C_{00}(C_{11}i - C_{1,-1}i)^*\}|, \end{aligned} \quad (5.31)$$

which illustrates a way to indirectly estimate the value of  $|R_1|$  in (5.24). Using (5.5) (5.6), the energy of the reverberation can be approximated by

$$E\{|\alpha_{00}|^2\} = E\{|C_{00}|^2\} - \frac{Y_{00}^2}{Y_{10}^2 |\mathcal{H}|^2} (E\{|C_{10}|^2\} - E\{|C_{10}|^2\}), \quad (5.32)$$

If the plane wave model is used for the direct path, (5.32) can be simplified using (5.22), as

$$E\{|\alpha_{00}|^2\} \approx \left( \frac{E\{|C_{00}|^2\}}{Y_{00}^2} - \frac{E\{|C_{10}|^2\}}{Y_{10}^2} + \frac{E\{|\alpha_{10}|^2\}}{Y_{10}^2} \right) Y_{00}^2, \quad (5.33)$$

where  $E\{|\alpha_{10}|^2\}$  can be estimated using (5.29). Substituting (5.29), (5.31) and

(5.32) into (5.17), the estimation expression for  $|R_1|$  can be written as

$$|R_1| \approx \frac{1}{2\sqrt{2}} \cdot \frac{|E\{C_{00}(C_{11}i + C_{1,-1})^*\}| + |E\{C_{00}(C_{11}i - C_{1,-1})^*\}|}{E\{|C_{00}|^2\} - E\{|C_{10}|^2\} \frac{Y_{00}^2}{Y_{10}^2 \mathcal{H}^2} + M_{\text{pwr}} \frac{Y_{00}^2}{Y_{10}^2 \mathcal{H}^2}}, \quad (5.34)$$

where we define

$$M_{\text{pwr}} \triangleq \frac{1}{2}(E\{|C_{11} + C_{1,-1}|^2\} + E\{|C_{11} - C_{1,-1}|^2\}) \quad (5.35)$$

similarly, by substituting (5.7) (5.8) and (5.32) into (5.18),  $R_2$  can be written as

$$R_2 \approx \frac{M_{\text{pwr}}}{E\{|C_{00}|^2\} - E\{|C_{10}|^2\} \frac{Y_{00}^2}{Y_{10}^2 \mathcal{H}^2} + M_{\text{pwr}} \frac{Y_{00}^2}{Y_{10}^2 \mathcal{H}^2}}, \quad (5.36)$$

It can be seen that all the coefficients required for the calculation can be acquired by a first order microphone array directly. The estimated values of  $|R_1|$  and  $R_2$  can be directly substituted into (5.21) or (5.24) for estimation of DRR using  $\gamma^2$ .

### 5.2.5 DRR estimation procedure

Assuming that the value of DRR is positive, the solution for DRR can be found by solving (5.20) or (5.24). For the plane wave model, the solution can be derived as

$$\text{DRR} = \frac{\gamma^2 + R_2\gamma^2 + \sqrt{4|R_1|^2(\gamma^2 - 1) + \gamma^4(R_2 - 1)^2 + 4R_2\gamma^2}}{2 - 2\gamma^2}, \quad (5.37)$$

where the assumption (5.27) is used, which leads to  $\text{Re}\{R_1\} = 0$ . The calculated DRR is in linear scale, and the more commonly used log-scale  $\text{DRR}_{\log}$  is defined as

$$\text{DRR}_{\log} = 10 \log_{10} \text{DRR}. \quad (5.38)$$

From our experience in testing the algorithm using the ACE Challenge Development Dataset [100], the estimation of  $|R_1|$  and  $R_2$  at a single frequency is often unstable. However, for typical room environments, one can assume that the characteristics of reverberation do not vary rapidly over frequencies since sound waves of similar wavelength are likely to have similar propagation modes. Therefore  $|R_1|$  and  $R_2$  can be seen as constant if the frequency band of interest is sufficiently narrow, then one can use the average values of  $|R_1|$  and  $R_2$  over a particular frequency band for the calculation of DRR for this frequency band.

For subband and full band DRR estimation, the results are obtained by taking the average of the single frequency DRR estimations within the band, then the values are converted to log scale for convenience.

We recommend the following procedures to estimate the DRR of a particular frequency band from a recording:

- 
- Step 1 Determine the direct path impinging direction using a suitable Direction-of-Arrival (DOA) algorithm, which can be done using the signal received by the first order (or higher order) microphone.
  - Step 2 Use an appropriate algorithm to detect the frames of the recording that contain speech signal and calculate the 0th and 1st order spherical harmonic coefficients for each frequency bin within the frequency band.
  - Step 3 Rotate the spherical harmonics using the method in Chapter 2.2.3, such that the z-axis is aligned with the direct path.
  - Step 4 Calculate  $|R_1|$  and  $R_2$  for each frequency bin, using (5.34) and (5.36), then average over all the frequency bins to obtain an estimation for the whole frequency band.
  - Step 5 Calculate  $\gamma$  for each frequency and using (5.20) or (5.37) with the averaged  $|R_1|$  and  $R_2$  to estimate the DRR for each frequency.
  - Step 6 Average the DRR estimations calculated from each frequency bin to obtain the subband or full band DRR estimation. Convert the result to log scale.
- 

A disadvantage of the original coherence method for DRR estimation is that the angle between the direct path and the particle velocity measurement direction is generally unknown, and in a real measurement, the microphone have to be pointed towards the direct path [98]. In our improved method, since we use a first order microphone for measurement, which records the complete sound field, it is possible to derive the velocity measurement in any direction, through rotation of the spherical harmonic coefficients. In addition, the data acquired by the microphone can be used to perform Direction-of-Arrival (DOA) estimation for the direct path, therefore there is no special requirement for positioning the microphone during measurements.

## 5.3 Impact of parameters on DRR estimation

### 5.3.1 Reverberation parameter

In order to illustrate the impact of  $R_1$  and  $R_2$  on the estimated DRR, we plot the theoretical DRR against  $\gamma^2$  using (5.24). with the diffuse field parameter setting

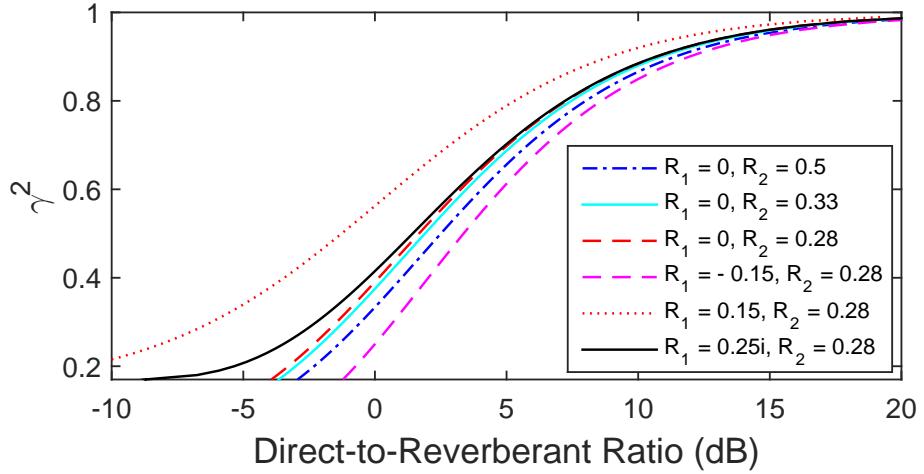


Figure 5.1: Theoretical  $\gamma^2$  versus estimated direct-to-reverberant ratio (DRR) calculated using (5.37), under various reverberation parameter settings.

proposed by Kuster [98] ( $R_1, R_2 = 0, 0.5$ ) and Jarrett [12] ( $R_1, R_2 = 0, 1/3$ ) as well as a number of other values that were commonly found in our experiment ( $R_1, R_2 = 0, 0.28; 0.15, 0.28; -0.15, 0.28; 0.25i, 0.28$ , respectively), as shown in Fig. 5.1. We note that the assumption (5.27) is not applied here, in order to illustrate the impact of  $R_1$  on the DRR estimation. It can be seen from Fig. 5.1 that depending on the values of  $R_1$  and  $R_2$ , a deviation of  $\pm 3\text{dB}$  in estimated DRR can be observed for low values of  $\gamma^2$ .

From (5.17), it can be seen that  $R_1$  is equivalent to the sound intensity in the  $z$  direction with certain normalization. Since all normalization factors are real, the real and imaginary part of  $R_1$  correspond to the active and reactive sound intensity, respectively. When  $\text{Re}\{R_1\} > 0$ , it indicates that the net energy flow of reverberation coincides with the direct path signal, and as a result the reverberation will be “added” to the direct path, and as a result contributes to coherence function  $\gamma^2$  positively. On the other hand, if  $\text{Re}\{R_1\} < 0$ , the net reverberation energy flow in the  $z$  direction opposes the direct path, essentially cancelling part of the direct path sound intensity, therefore it contributes to  $\gamma^2$  negatively. As a result of this, as can be seen in Fig. 5.1, for the same value of  $\gamma^2$ , a positive  $\text{Re}\{R_1\}$  corresponds to low value of DRR, and vice versa.

The absolute value of  $R_1$  represents the overall coherence of the reverberant field in the  $z$  direction. This includes the reactive part of  $R_1$ , which corresponds to the resonating reverberation energy. It can be seen from (5.24) that  $|R_1|$  always contributes to  $\gamma^2$  positively. Therefore, as seen in Fig. 5.1, a non-zero value of  $|R_1|$

results in lower value of DRR, for the same  $\gamma^2$ , this is especially significant at lower values of  $\gamma^2$ .

Using a first order microphone, it is possible to estimate  $|R_1|$  for each frequency bin, if it is assumed that the reverberant sound intensity is uniform in each direction. Unfortunately, the sign of  $Re\{R_1\}$ , which indicates the direction of energy flow, cannot be determined through observation of the sound field in its orthogonal directions. However, by observing the reverberation sound field from the ACE Challenge Development Dataset [100], it was found that both active and reactive sound intensity of the reverberation in the  $x$  and  $y$  directions have zero mean when averaged over each 1/3 octave subband, indicating that the energy flow of reverberation changes randomly and rapidly with frequency. Therefore it is reasonable to assume that  $PV_z^*$  is also zero mean when observed at multiple frequencies. As a result, when averaging the estimated DRR over each subband, the impact of  $Re\{R_1\}$  (and  $Im\{HR_1^*\}$  in (5.21)) on each frequency bin will be cancelled out, and the term can be removed in the derivation of (5.37), provided that appropriate frequency averaging is performed after calculating DRR for each frequency bin.

As can be seen from Fig. 5.1,  $R_2$  does not affect the estimated DRR as strongly as  $R_1$ , and a lower value of  $R_2$  results in a slightly lower estimation of DRR. From (5.18) it can be seen that  $R_2$  reflects the expected energy ratio between sound pressure and particle velocity. In Jarrett's diffuse field model [12], the value of  $R_2$  is lower ( $R_2 = 1/3$ ), therefore, we expect Jarrett's method to yield a slightly lower estimation of DRR compared to Kuster's. From our analysis to the ACE Challenge Development Dataset, the value of  $R_2$  typically varies between 0.25 – 0.33, which is close to Jarrett's model (see Table 5.2).

### 5.3.2 Nearfield sound source

In order to analyze the DRR estimation error due to using a plane wave to approximate the direct path sound field, we compute the difference in the estimated DRR using (5.20) and (5.24) ( $\Delta\text{DRR} \triangleq 10\log_{10}(\text{DRR}_{\text{plane}}/\text{DRR}_{\text{point}})$ ). It can be seen by observing (5.20) that the calculated  $\text{DRR}_{\text{point}}$  depends on the product  $kr_0$ . Fig. 5.2 plots  $\Delta\text{DRR}$  as a function of  $kr_0$ , for various values of  $\gamma^2$ . In this figure, for simplicity, we assume that  $R_1 = 0$ ,  $R_2 = 0.5$ . The selected values of  $\gamma^2$  (0.86, 0.65, 0.33 and 0.19) correspond to  $\text{DRR}_{\text{plane}} = 10\text{dB}, 5\text{dB}, 0\text{dB}$  and  $-2.5\text{dB}$ , respectively, using the parameter settings described above.

From Fig. 5.2 we can see that the plane wave model results in higher DRR

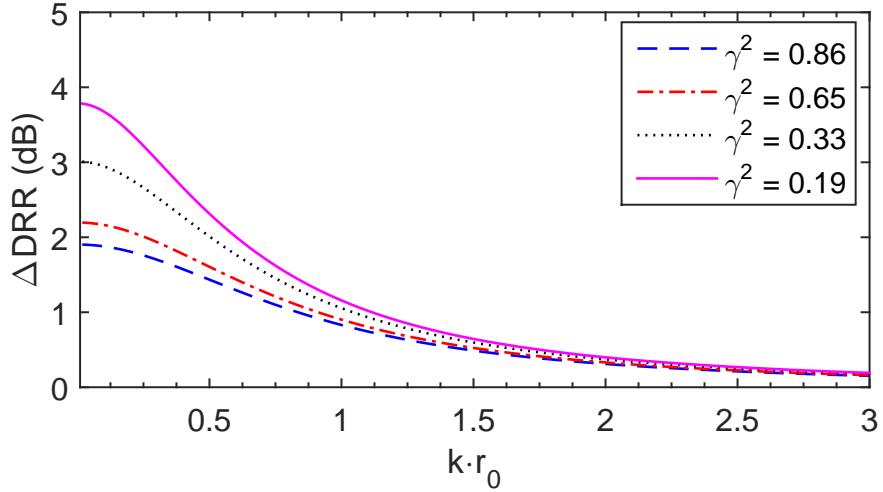


Figure 5.2: Plot of theoretical DRR versus  $kr_0$  using plane wave model (5.24) and point source model (5.20) with  $\gamma^2 = 0.86, 0.65, 0.33$  and  $0.19$ .

estimations than that of the point source model for smaller values of  $kr_0$ , where  $\Delta \text{DRR} \approx 2 - 4$  dB for  $kr_0 = 0$ , depending on the value of  $\gamma^2$ . At higher frequencies and larger source-microphone distance (higher  $kr_0$ ), the difference between the two methods reduce rapidly, at  $kr_0 > 3$ , the difference in the calculated DRR using the two models becomes negligible.

Comparing the curves corresponding to each value of  $\gamma^2$ , it can be seen that the estimation error of the plane wave model is smaller when  $\gamma^2$  is larger, corresponding to higher values of DRR. The user may select the appropriate model for their applications, based on the target frequency band and expected source distance. Here, we propose a rule of thumb for determining whether to use the point source model or the plane wave model. When  $kr_0 \geq 2$ , the error caused by plane wave model is less than 0.5 dB for all values of  $\gamma^2$ , as can be observed in Fig. 5.2. For  $kr_0 < 2$ , the use of point source model is recommended for improving DRR estimation accuracy.

## 5.4 Validation using ACE Challenge Database

### 5.4.1 The ACE Challenge Database

The ACE Challenge Database is used to validate our algorithm [100]. The database consists of two datasets: the Evaluation dataset, and the Development dataset. The Development dataset is provided to the ACE Challenge participants as a training

Table 5.1: Room dimensions (approx.) and minimum/maximum DRR for each room recording configuration

Room Name	Lecture Room 1	Lecture Room 2	Meeting Room 1	Meeting Room 2	Office 2
Length (m)	6.9	13.4	6.6	10.3	5.1
Width (m)	9.7	9.2	4.7	9.2	3.2
Height(m)	3.0	2.9	3.0	2.6	2.9
Volume (m <sup>3</sup> )	200	360	92	250	48
Setup A min DRR	-0.82	-0.37	-2.0	-2.6	-0.44
Setup A max DRR	15	13	11	11	13
Setup B min DRR	0.87	-3.7	-3.1	1.1	-2.3
Setup B max DRR	7.9	6.4	7.6	12	9.5

database, using which the participants can train and fine tune their algorithms. The Evaluation dataset is used to evaluate the performance of fine-tuned algorithms.

The Evaluation dataset consists of 4500 synthesized recordings of various configurations. A total of 5 rooms are used to record the room impulse responses, with two recording setups (positions) for each room. The room details are summarized in Table 5.1. We note that although the impulse responses of 7 rooms were recorded according to [100], only 5 of them are used to create the Evaluation dataset; the other two rooms were used to create the Development dataset. The speech and noise setup for the Development dataset differ from that of the Evaluation database, therefore in this work, the Develop dataset is only used for developing the DRR algorithm; the results presented in this section are all generated using the Evaluation dataset.

The impulse responses are recorded using an Eigenmike, and the reverberant speech recordings are synthesized by convolving the impulse responses with anechoic speech recordings [100]. The speech recordings consist of voices of 10 talkers, 5 female and 5 male, with 5 separate utterance recordings for each talker. Three different types of noise (“Ambient”, “Fan” and “Babble”) are recorded separately under the same room setup and mixed into the reverberant speech recordings, each with three SNR settings (−1 dB, 12 dB and 18 dB).

The ground truths for both full band and subband DRR have been provided. For subband DRR, the central frequencies for all bands have been chosen according to the ISO standard [100].

### 5.4.2 Algorithm setup

Since the ground truth for direct path DOA is not given, we have to estimate the DOA for each of the ten scene setups. This is done by segmenting each speech recording into multiple short frames, and selecting the frames that correspond to the beginning of each utterance (where the impinging signal is almost purely due to

the direct path). To find the frames containing speech, a simple speech detection algorithm calculates the average signal energy of each frame, and select the frames with higher energy, which are considered to contain the speech signal. If the energy of a frame is significantly higher than the previous one, then this window is considered to contain the beginning of an utterance. We then calculate the spherical harmonic coefficients for each selected frame and for frequencies between 200-2000 Hz, and perform a frequency averaged MUSIC DOA estimation in the spherical harmonic domain [11, 103]. The estimated DOA is used for further calculations.

In order to maintain the highest possible frequency resolution while at the same time to avoid violating the assumption that the direct path signal and reverberations are uncorrelated, we choose the analysis window length to be 10 ms. When fine-tuning our algorithm using the ACE Development Dataset, it was found that a window length shorter than 10 ms does not reduce the average value of  $\gamma^2$ , therefore we assume that the chosen window length is appropriate.

For each speech recording, only the windows that contain the speech signal are used for analysis. For each frequency subband, we calculate the 0th and 1st order spherical harmonic coefficients for each selected window and for all the frequency bins within each subband. We then follow steps 3 through 6 in Section 5.2.5 to estimate DRR for each subband.

Although the ground truth for subband DRR is given for all frequency bands between 20 Hz and 20 kHz [100], the recorded speech signal does not cover the complete spectrum. Therefore, we focus on the subbands with central frequency between 199.52 Hz and 2511.89 Hz, where there is sufficient energy in the speech recordings for DRR estimation. For this reason, we cannot estimate the full band DRR in the complete sense, instead, we calculate the average DRR over the selected subbands, which is used as the full band estimation. The full band ground truth DRR used for comparison is also calculated by averaging the corresponding subband ground truths, instead of using the full band DRR provided by the database.

The exact source-to-microphone distance is not provided by the database. However, according to the organizer of the ACE Challenge, the microphones are placed at no less than 1 m away from the source for all recording scenarios. Since we only focus on frequencies above 199.52, using the rule of thumb proposed in Section 5.3.2,  $kr_0 \approx 3.66 > 2$ , therefore the plane wave model is sufficiently accurate, hence used for DRR estimation.

The error of estimated DRR is defined as

$$\text{DRR}_{\text{err}} = 10 \log_{10} \left( \frac{\text{DRR}_{\text{est}}}{\text{DRR}_{\text{truth}}} \right). \quad (5.39)$$

The mean and standard deviation of DRR estimation error is then calculated using  $\text{DRR}_{\text{err}}$  from each recording.

### 5.4.3 Full band results

The full band DRR estimation results for the ACE Database are shown in Fig. 5.3. In this figure, we plot the mean and standard deviation of the DRR estimations for each of the 10 room configurations. Only the recordings with 18 dB SNR are used for this analysis. In order to better evaluate the performance of the proposed method, both Kuster's [98] and Jarrett's [12] methods were implemented for comparison. In the case of Kuster's method, since the algorithm requires a pair of omnidirectional microphones placed very close to each other for recording, which is not available in the ACE challenge (other microphone array setups used in the ACE Challenge have a minimum spacing of 60 mm [100], which is too large for accurate measurement of particle velocity), we use the 0th and 1st order spherical harmonics in place of the sound pressure and particle velocity in calculation. Since it is shown that the spherical harmonics are equivalent to sound pressure and particle velocity, this implementation is expected to be representative of Kuster's method.

It can be observed from Fig. 5.3 that all three methods yield less than 3 dB mean error for all of the 10 room setups. The method proposed by Jarrett *et al.* shows a similar trend as that proposed by Kuster, but with a slightly lower estimation of DRR in most setups, as can be expected from Fig. 5.1. The proposed method, on the other hand, results in 1 – 3 dB lower estimated DRR for most configurations.

A clear trend of DRR overestimation (estimated DRR higher than ground truth) can be observed for both methods that assume diffuse reverberant field. This is consistent with Kuster's observations from his experiments, where his method tend to overestimate DRR in real-life recording setups. The proposed method does not show any clear tendency of overestimation or underestimation, with 5 of the setups having positive mean error and the other 5 setups having negative mean error.

In terms of standard deviation, one would expect that the proposed method would yield higher standard deviation compared to Kuster's method, since in the proposed algorithm, both  $|R_1|$  and  $R_2$  need to be estimated for each frequency

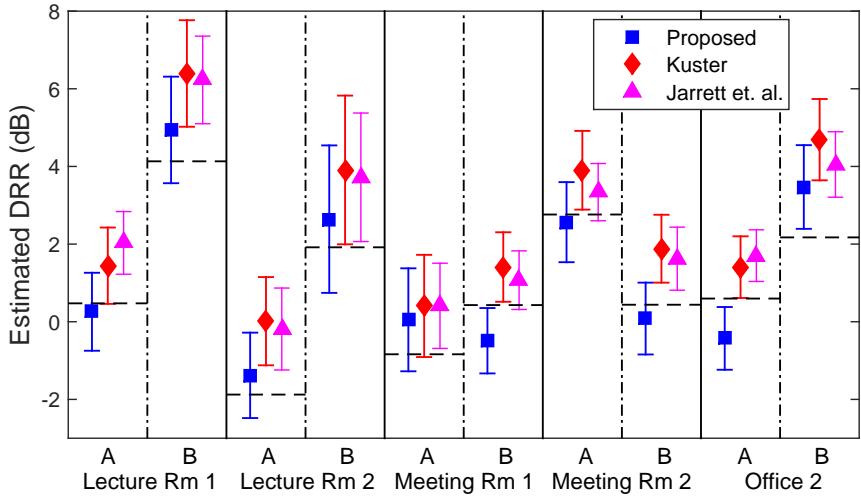


Figure 5.3: Mean and standard deviation of estimated DRR using the proposed method (blue), Kuster's method (red) and Jarrett's method (pink) for all 5 rooms and 2 locations (A and B) in each room, with 18 dB SNR, averaged over 3 noise types. Dashed lines indicate ground truth DRR.

band, which would add uncertainty to the distribution of estimated DRR. However, from Fig. 5.3 it can be seen that the proposed algorithm yields almost identical standard deviation as Kuster's method, which indicates that the primary contributor of standard deviation is the coherence function  $\gamma^2$ , which is common for both the proposed method and Kuster's method.

On the other hand, Jarrett's method results in the lowest error standard deviation for all scenarios. The reason for this is that in the other two methods, only the first order spherical harmonics are used to calculate the coherence  $\gamma^2$ , while Jarrett's method utilizes all of the available spherical harmonic coefficients to reach a more consistent estimation of  $\gamma^2$ , which reduces its deviation due to random interference and other sources of error.

#### 5.4.4 Subband results

The subband estimation results are shown in Fig. 5.4. In this figure, we plot the mean and standard deviation of the subband DRR estimation error using the proposed method as well as the two baseline methods. The error mean and standard deviation are averaged over the results from all 10 rooms, and once again only the 18 dB SNR recordings are used for the analysis. Only the DRR for the subbands with central frequency between 199 Hz and 2511 Hz are calculated.

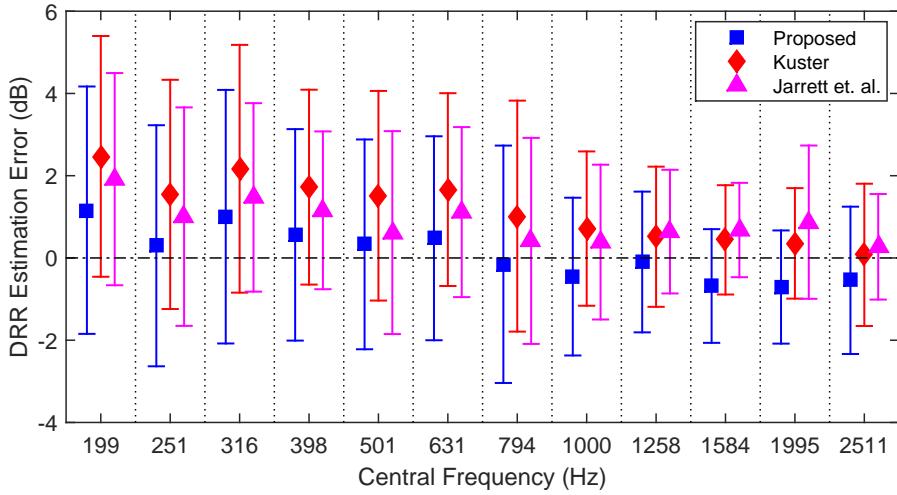


Figure 5.4: Mean and standard deviation of subband DRR estimation error for all rooms and configurations with 18 dB SNR, using the proposed method (blue), Kuster’s method (red) and Jarrett’s method (pink).

From Fig. 5.4 it can be seen that in general, the mean error of the proposed method falls within 1 dB of the ground truth for all frequency bands. furthermore, the subband results below 1000 Hz show a different pattern than the subbands above 1000 Hz. Below 1000 Hz, the mean error are all positive, indicating a slight overestimation of DRR; the error standard deviation is approximately 3 dB for these subbands. On the other hand, for frequency bands above 1000 Hz, the mean error becomes negative; the standard deviation of estimation error reduces to 2 dB at 1000 Hz, and decreases further at higher frequencies. On the other hand, both Kuster’s and Jarrett’s methods show a clear trend of overestimation, this is especially significant for Kuster’s method at lower frequencies. Jarrett’s method yields lower DRR estimations compared to Kuster’s, and in most frequency bands, have the lowest standard deviation.

Due to the geometry of the Eigenmike, only the 1st order spherical harmonics can be reliably captured for frequencies below 1000 Hz [53]. Below 1000 Hz, the 2nd order spherical harmonics are aliased onto the 1st order coefficients, and the aliasing error increases with frequency; at 1000 Hz and above, our algorithm begins to calculate the 2nd order coefficients, which removes the aliasing and improves the accuracy of the 1st order coefficients. Furthermore, at higher frequencies, the wavelength of the sound becomes closer to the dimension of the Eigenmike (8.4 cm diameter), which further increases the accuracy of 1st order spherical harmonic acquisition. This explains why the error standard deviation decreases gradually at

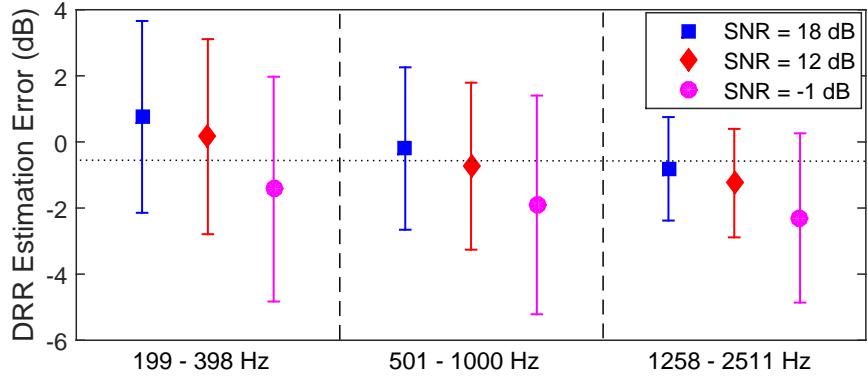


Figure 5.5: Mean and standard deviation of DRR estimation error with 18dB, 12dB and  $-1$ dB SNR.

higher frequencies.

Overall it can be seen that compared to the two baseline algorithms, the proposed method produces an unbiased DRR estimation. The standard deviation of the proposed algorithm is on par with Kuster's method, but slightly higher than Jarrett's method.

#### 5.4.5 Impact of noise on DRR estimation

In order to examine the impact of noise (interference) on the result of DRR estimation, the algorithm is run for the Evaluation dataset recordings of each SNR setting (18dB, 12dB and  $-1$ dB), and we calculate the mean and standard deviation for each SNR setting, the results are shown in Fig. 5.5. In this figure, the subband results are separated into three frequency ranges: low (199-398 Hz), medium (501-1000 Hz) and high (1258-2511 Hz). Each frequency range covers four subbands, and the subband results are averaged within each frequency range, in order to simplify the data representation.

It can be seen from Fig. 5.5 that the difference between the estimation results with 18 dB and 12dB SNR is less than 1 dB. At  $-1$  dB SNR, however, the DRR estimation becomes strongly biased towards underestimation. The cause of this phenomenon is that the interference/acoustic noise, which does not have the same impinging direction as the direct path signal, will reduce the coherence between sound pressure and gradient (particle velocity), resulting in a lower value of  $\gamma^2$ , thereby lowering the estimated DRR.

The other impact of high interference level is the increased error standard de-

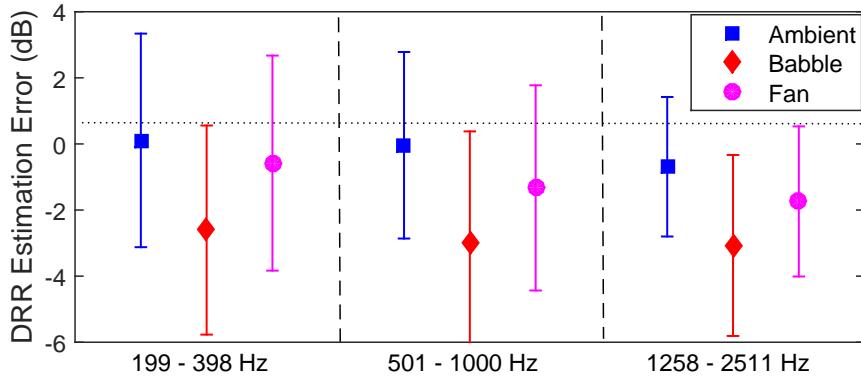


Figure 5.6: Mean and standard deviation of DRR estimation error in multiple noisy environments with  $-1\text{ dB}$  SNR.

viation. When developing and testing our algorithm using the ACE Development Dataset, we noticed that our frequency averaged MUSIC DOA algorithm became much less reliable at  $-1\text{ dB}$  SNR, compared to  $18\text{ dB}$  and  $12\text{ dB}$  SNR. A direct result of inaccurate DOA estimation is the decreased consistency of DRR estimations at different utterance/interference configurations in the same room setup, which is reflected by a higher error standard deviation. It is expected that if a more interference-robust DOA algorithm is applied, or if the DOA information can be measured directly, the proposed algorithm would produce more consistent estimations at low SNR.

How different types of interference affect the performance of the DRR estimation is also investigated. The three noise types mixed into the recordings each have different spectral characteristics, and therefore their effects on the subband DRR estimation vary. This is illustrated in Fig. 5.6, which plots the estimation results for the low, medium and high frequency ranges and for each of the three noise types. The SNR of all recordings used in this analysis are  $-1\text{ dB}$ .

From Fig. 5.6 it can be seen that the “Ambient” noise type has the least effect on DRR estimation accuracy causing only a small bias towards under estimation, while the “Babble” noise results in more than  $3\text{ dB}$  of under estimation for all frequency ranges. The “Fan” noise has slightly more impact than the “Ambient” noise type, but less than that of the “Babble” noise. The cause of this result is due to both the spectral and spatial characteristics of the different noise types.

Fig. 5.7 plots the normalized power spectrum of the three noise types, the spectra are acquired by manually selecting the sections of recordings that contain purely noise signal. It can be seen that the “Ambient” noise consists of primarily low

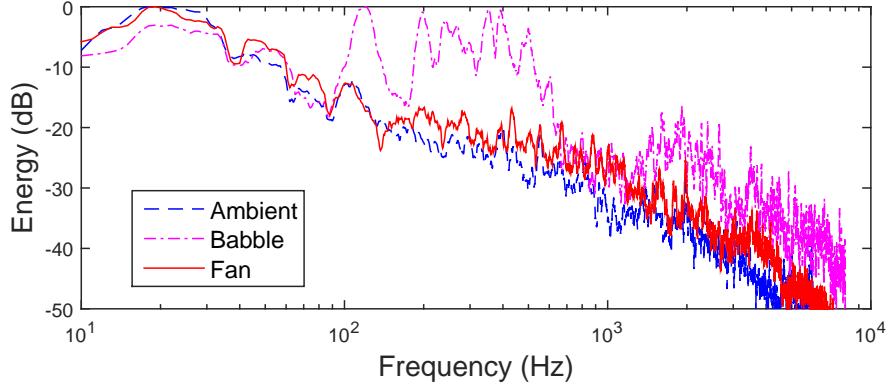


Figure 5.7: Normalized power spectrum of the “Ambient”, “Babble” and “Fan” noises in the ACE Evaluation Dataset.

frequency signals that do not overlap with the speech signal spectrum. Therefore, the subbands of interest are most likely to have higher SNR than the full band SNR of  $-1$  dB. As a result, the ambient noise has the least effect on the accuracy of DRR estimation. On the other hand, the “Babble” noise is essentially a speech recording by itself, therefore it almost completely overlaps with the spectrum of the speech of the talker, resulting in the lowest SNR in the speech spectrum of the three noise types. The “Fan” noise has very similar spectral characteristics as the “Ambient” noise type, although its higher frequency components have more energy than that of the “Ambient” noise, which leads to slightly more impact on DRR estimation.

According to the ACE Challenge description [100], the “Fan” noise is generated using one or two fans inside the recording environment, while the “Babble” noise records the voices of up to 7 people talking around the recording location. The “Ambient” noise is a recording of the ambient noise within the room. Due to the larger number of uncorrelated sources, each with a different DOA, the “Babble” noise is likely to have a lower coherence level than that of the “Fan” noise. Therefore when mixed into the speech recording, the “Babble” noise would lower  $\gamma^2$  further than the “Fan” noise. Although the nature of the “Ambient” noise is unclear, in typical room environments its source is likely to be AC vents or windows, both of which can be considered as localized sources, thus creating a more coherent sound field than the “Babble” noise. In addition, due to its spectral characteristics, its impact on DRR estimation is the smallest of all three noise types.

Table 5.2: Mean of estimated parameters in each room configuration and frequency range

Room	Setup	$ R_1 $			$R_2$		
		Low	Med	High	Low	Med	High
Lecture Room 1	A	0.280	0.194	0.219	0.288	0.251	0.265
	B	0.277	0.293	0.331	0.290	0.293	0.332
Lecture Room 2	A	0.232	0.201	0.146	0.316	0.268	0.290
	B	0.239	0.189	0.232	0.337	0.277	0.314
Meeting Room 1	A	0.191	0.120	0.157	0.294	0.248	0.253
	B	0.239	0.279	0.118	0.297	0.273	0.291
Meeting Room 2	A	0.226	0.265	0.278	0.201	0.329	0.321
	B	0.211	0.215	0.225	0.241	0.255	0.281
Office 2	A	0.268	0.167	0.174	0.252	0.269	0.286
	B	0.199	0.213	0.193	0.263	0.282	0.278

#### 5.4.6 Estimated parameters from the ACE Evaluation Dataset

The parameters  $|R_1|$  and  $R_2$  estimated for each subband of every speech recording in the ACE Evaluation Dataset has been recorded and is presented in Table. 5.2, where we have taken the average values of  $|R_1|$  and  $R_2$  for the low, medium and high frequency ranges and for all the recordings from each room configuration, only the data from recordings with 18 dB SNR are used for this calculation.

As can be seen from Table 5.2, although the values of  $R_1$  and  $R_2$  vary for each room configuration and frequency range, in general,  $|R_1|$  falls within the range of 0.15-0.25, while  $R_2$  lies in between 0.25-0.33 in the majority of cases. From Fig. 5.1, it can be seen that the values of  $|R_1|$  and  $R_2$  shown in Table 5.2 would lead to our proposed algorithm yielding lower DRR estimations than assuming  $R_1 = 0, R_2 = 0.5$ , which is indeed the case in our estimation results.

From the above results, we believe that setting  $|R_1| = 0.2$  and  $R_2 = 0.28$  provides a more reasonable and accurate model for a general reverberant sound field within room environments, compared to the diffuse model where it is assumed that  $R_1 = 0$ , and  $R_2 = 1/2$  or  $1/3$ . It is sometimes easier to acquire or implement differential microphone pairs than complete first order microphone systems (such as the Eigenmike), and when a differential array is to be used to estimate room DRR, we suggest using (5.20) or (5.37) to calculate DRR, and assume  $|R_1| = 0.2, R_2 = 0.28$ , which is likely to yield more accurate estimation results.

## 5.5 Summary

In this work, we present a novel algorithm for estimating DRR using a first order microphone system. We show that the proposed algorithm is a generalization of previous DRR estimation methods based on sound pressure-particle velocity coherence function. Using the proposed algorithm, it is possible to estimate the characteristics of a reverberant sound field which are relevant to DRR estimation, thereby improving the estimation accuracy of the method. We also show that at low frequency and small source-to-microphone distance, using the plane wave model for the direct path signal can result in a positive bias on the estimated DRR. Through validating the proposed algorithm using the ACE Challenge Dataset, it was found that the proposed algorithm provides  $\pm 2$  dB mean estimation error for the frequency range of human speech (200-2500 Hz), and shows no obvious bias.

## 5.6 Related Publications

This chapter's work has been published in the following journal paper and conference proceeding [104] [99]:

- H. Chen, T. D. Abhayapala, P. N. Samarasinghe, and W. Zhang, “Direct-to-reverberant energy ratio estimation using a first order microphone,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, PP. 226–237, Feb 2017.
- H. Chen, P. N. Samarasinghe, T. D. Abhayapala, and W. Zhang, “Estimation of the direct-to-reverberant energy ratio using a spherical microphone array.,” in *Proc. ACE Challenge Workshop, a satellite event of WASPAA, New Paltz, NY, USA*, Oct 2015.

## 5.7 Proof of Equation (5.22)

The closed form expression of spherical Hankel functions of the first kind is [61]

$$h_l^{(1)}(z) = i^{-l-1} z^{-1} e^{iz} \sum_0^l (l + \frac{1}{2}, k) (-2iz)^{-k}. \quad (5.40)$$

The expression of  $h_0^{(1)}(z)$  and  $h_1^{(1)}(z)$  can then be written as

$$h_0^{(1)}(z) = -ie^{iz}\frac{1}{z} \quad (5.41)$$

$$h_1^{(1)}(z) = -e^{iz}\frac{z+i}{z^2}, \quad (5.42)$$

substituting (5.41) and (5.42) into (5.22) yields

$$\lim_{r_0 \rightarrow \infty} \frac{h_1^{(1)}(kr_0)}{h_0^{(1)}(kr_0)} = \lim_{r_0 \rightarrow \infty} \frac{kr_0 + i}{ikr_0} = \lim_{r_0 \rightarrow \infty} \left( -i + \frac{1}{kr_0} \right) = -i, \quad (5.43)$$

which completes the proof.

# Chapter 6

## Methods for spatial ANC performance evaluation and optimization

**Overview:** *The use of spherical harmonic expansion to model noise fields enables in-depth analysis and manipulation of the sound field. In this chapter, we introduce a number of techniques for improving the spatial ANC performance. In Section 6.2, we propose an improved sound field synthesis method based on spherical harmonic mode matching. Through the use of spherical harmonic addition theorem, this method allows the user to define a number of high priority regions within the quiet zone, where greater noise attenuation can be achieved compared to the rest of the quiet zone. In Section 6.3, we propose a new metric for measuring noise energy over a spherical region, and use the new metric to evaluate the ANC performance of an experimental ANC system. Finally, in Section 6.4, we use this metric to develop a method for estimating optimum noise cancellation performance for a given noise environment, and use this method to estimate the ANC performance in a passenger car.*

### 6.1 Introduction

The goal of spatial ANC is to minimize the noise level inside a certain quiet zone. However, the exact “optimal” loudspeaker driving signals that would yield the best noise reduction depends on the loudspeaker setup as well as the characteristics of the quiet zone. For example, some regions within the quiet zone may be more

important than the others, because the users are more likely to stay within these regions. In such case, it would be beneficial to focus the ANC resources toward these “more important” regions, which would result in a better overall ANC quality than attenuating the noise evenly within the quiet zone.

Furthermore, the definition of “optimal” noise attenuation often depends on the method employed to measure the noise level. For the noise level at a single point in the space, one microphone is enough to pick up the sound pressure of the noise; however, for a spatial region, measurement of the average sound pressure level becomes much more complicated, as sampling the noise field at a few points within the quiet zone cannot accurately represent the overall noise level inside the whole quiet zone.

In addition, the ability to estimate, or predict the potentially achievable “optimal” spatial noise attenuation would be greatly beneficial to the design process of spatial ANC systems. The designer would be able to find out the amount of hardware that is necessary to achieve the desired noise attenuation, or to determine whether the available loudspeaker setup is sufficient for the ANC task, before physically implementing a complete ANC system.

In this chapter, we utilize the spherical harmonic analysis technique to develop a set of algorithms and tools to address the above problems. We show that characterization and control of the noise field can be conveniently done by appropriate manipulation of the spherical harmonic coefficients of the noise field. This chapter is organized as follows:

In **Section 6.2**, we introduce a spatial single zone sound field reproduction technique which allows for higher reproduction accuracy within certain sub-zones while maintaining a reasonable reproduction accuracy in the global region. By applying the spherical harmonics addition theorem, we connect the spherical harmonic coefficients of the global region with that of the local sub-zones, and use a weighting method to enhance the reproduction quality at the sub-zones. This technique is particularly useful when the available loudspeakers cannot provide very good sound reproduction for the whole region, but a high accuracy is desired for at least some sub-zones, such as some spatial ANC scenarios.

In **Section 6.3**, we propose a new metric for the measurement of average noise level over a region. It is formulated in terms of the spherical harmonic decomposition of sound fields. Through a series of experiments, we show that the proposed metric provides superior characterization of the noise level within the control region, compared to existing methods where a number of microphones are placed around

the control region to sample the noise level. This metric is particularly suitable for environments with irregular geometry and a fixed control region with moderate size, such as vehicle and aircraft cabins.

In **Section 6.4**, utilizing the noise level metric developed in Section 6.3, we evaluate a passenger car’s integrated loudspeakers’ noise cancelling capabilities by analyzing the in-car noise field and the loudspeaker responses. Our proposed analysis method decomposes the noise field into a number of basis sound patterns, and evaluate the loudspeakers’ capability at reproducing these basis patterns, then calculate the expected overall noise reduction based on these results. Our results show that the noise field inside a vehicle cabin has a sparse nature, and that the car’s loudspeakers are capable of cancelling the noise around the passengers’ head positions.

## 6.2 Enhanced sound field reproduction within prioritized control region

This section proposes an enhanced method for synthesizing the sound field using a relatively small number of secondary sources which allows improved synthesizing accuracy for certain subregions of the interested zone. We introduce the spherical harmonic translation into the mode matching algorithm to acquire a uniform modal-domain representation of the sound fields within different sub-regions. Then by changing the weighing of each region, the least mean squares solution can be easily controlled to cater for certain prioritized reproduction requirements. This method is shown to be especially effective in the situations where the number of secondary sources is limited.

### 6.2.1 Background

In 3D sound field synthesis, a fundamental problem arises which makes implementation very difficult: the synthesis quality is strongly related to the number and position of the loudspeakers [105–107]. The ideal placement of the loudspeakers for the mode-matching technique is to have the loudspeakers evenly distributed on a sphere surrounding the interested region [44], such structure is impractical in reality. To solve this problem, an array configuration for 3D sound field synthesis using multiple circular loudspeaker arrays was proposed by Zhang and Abhayapala [108], this method uses a functional analysis based algorithm to derive the driving signals.

Still, the trade off between the number of the loudspeakers and the size and frequency of the reproduction zone exists. The reproduction quality degrades rapidly as the number of loudspeakers becomes less than the minimal required number. In the case that the interested region can be separated and reduced into a few smaller regions, it is possible to control the sound field in these small regions through spatial multizone reproduction techniques [109]. However, the calculation involves matrix inversion, and if done without proper regularization, the results may be highly unstable.

The goal of this section is to introduce a spatial single zone sound field reproduction technique which allows for higher reproduction accuracy within certain sub-zones while maintaining a reasonable reproduction accuracy in the global region. This can be achieved by balancing between the single zone reproduction and the spatial multizone reproduction techniques. Through the use of spherical harmonic translation, the mode-matching method can be simultaneously applied to both the global interested zone and certain sub-region within it (referred to as high priority regions), and by adjusting the weighing factors in the LMS solution, one can easily control the reproduction quality of different regions. This technique is particularly useful when the reproduction region is large, but an insufficient number of loudspeakers are available, and/or in applications where a high reproduction accuracy is required for certain sub-zones, such as active noise cancellation.

### 6.2.2 Problem formulation

A common way of deriving loudspeaker driving signals to produce a certain desired sound field is by pressure matching in the modal domain, which has been briefly reviewed in 2.3.3. Given  $V$  loudspeakers and spherical harmonic coefficients up to order  $L$  representing the sound field, the channel information between the  $v^{th}$  loudspeaker and the  $n^{th}$  mode can be denoted  $H_{vn}$ , the channel matrix is then expressed as  $\mathbf{H}$ , where

$$\mathbf{H} = \begin{bmatrix} H_{00}^1 & H_{00}^2 & \dots & H_{00}^V \\ H_{11}^1 & H_{11}^2 & \dots & H_{11}^V \\ \vdots & \vdots & \ddots & \vdots \\ H_{LL}^1 & H_{LL}^2 & \dots & H_{LL}^V \end{bmatrix}, \quad (6.1)$$

where  $H_{lm}^v$  is the spherical harmonic coefficient of order  $l$  and mode  $m$ , due to the  $v^{th}$  loudspeaker playing an unit signal. The total number of coefficients is given by

$N = (L+1)^2$ . A desired sound field of the same order can be expressed as a column vector of spherical harmonic coefficients

$$\mathbf{Q} = [Q_{00}, \ Q_{11}, \ Q_{10}, \ \dots \ Q_{LL}]^T. \quad (6.2)$$

The least mean square solution for the driving signals  $\mathbf{D}$  can be written as

$$\mathbf{D} = \mathbf{H}^{-1}\mathbf{Q}, \quad (6.3)$$

where  $[\cdot]^{-1}$  denotes pseudoinverse of the matrix. This minimizes the cost function

$$\mathcal{L} = (\mathbf{Q} - \mathbf{H}\mathbf{D})^H(\mathbf{Q} - \mathbf{H}\mathbf{D}). \quad (6.4)$$

We now consider a sub-region  $\mathcal{O}_q$  within the global reproduction region. We denote the spherical harmonic coefficients of the desired sound field with respect to  $\mathcal{O}_q$  as  $\mathbf{Q}_q$ , and denote the channel matrix for this sub-region as  $\mathbf{H}_q$ . The driving signals that minimizes the reproduction error within the sub-region can be written as

$$\mathbf{D} = \mathbf{H}_q^{-1}\mathbf{Q}_q, \quad (6.5)$$

which minimizes the cost function

$$\mathcal{L}_q = (\mathbf{Q}_q - \mathbf{H}_q\mathbf{D})^H(\mathbf{Q}_q - \mathbf{H}_q\mathbf{D}). \quad (6.6)$$

Our goal is to find a driving function, which minimizes a new cost function that contains both  $\mathcal{L}$  and  $\mathcal{L}_q$ , and also has a weighting factor which could further enhance the reproduction accuracy within the sub-region. This can be expressed as

$$\min\{\mathcal{L} + \alpha\mathcal{L}_q\}, \quad (6.7)$$

where  $\alpha$  is a weighting factor.

### 6.2.3 Combined Least Mean Square Solution for sound field reproduction

Using the spherical harmonics addition theorem (2.16), the sound field coefficients  $\mathbf{C}_q$  at  $\mathcal{O}_q$  can be expressed using the global coefficients  $\mathbf{C}$  and a translation matrix  $\widehat{\mathbf{S}}_q$ , as  $\mathbf{C}_q = \widehat{\mathbf{S}}_q\mathbf{C}$ . The channel matrices also have a similar relationship  $\mathbf{H}_q = \widehat{\mathbf{S}}_q\mathbf{H}$ .

Then, a LMS solution for synthesizing the desired sound field only within the sub-region  $\mathcal{O}_q$  can be found by solving

$$\widehat{\mathbf{S}}_q \mathbf{Q} = \widehat{\mathbf{S}}_q \mathbf{H} \mathbf{D}. \quad (6.8)$$

Thus  $\mathbf{D}$  can be expressed as

$$\mathbf{D} = (\widehat{\mathbf{S}}_q \mathbf{H})^{-1} \widehat{\mathbf{S}}_q \mathbf{Q}. \quad (6.9)$$

It should be noted that the solution provided by (6.9) normally requires regularization, since the matrix  $\widehat{\mathbf{S}}_q \mathbf{H}$  may be ill conditioned, which may result in very large driving signals for the loudspeakers. This is especially true when the sub-region  $\mathcal{O}_q$  is small.

The local cost function  $\mathcal{L}_q$  can be expressed as

$$\mathcal{L}_q = (\widehat{\mathbf{S}}_q \mathbf{Q} - \widehat{\mathbf{S}}_q \mathbf{H} \mathbf{D})^H (\widehat{\mathbf{S}}_q \mathbf{Q} - \widehat{\mathbf{S}}_q \mathbf{H} \mathbf{D}), \quad (6.10)$$

which corresponds to the sum of the squared errors in all local spherical harmonic coefficients in the subregion  $\mathcal{O}_q$ .

The combined cost function (6.7) can then be written as

$$\mathcal{L}_{\text{all}} = \mathbf{E}_G^H \mathbf{E}_G + \alpha \mathbf{E}_Q^H \mathbf{E}_Q, \quad (6.11)$$

where  $\mathbf{E}_G = (\mathbf{Q} - \mathbf{H} \mathbf{D})^H (\mathbf{Q} - \mathbf{H} \mathbf{D})$  is the global error vector, and  $\mathbf{E}_Q = (\widehat{\mathbf{S}}_q \mathbf{Q} - \widehat{\mathbf{S}}_q \mathbf{H} \mathbf{D})^H (\widehat{\mathbf{S}}_q \mathbf{Q} - \widehat{\mathbf{S}}_q \mathbf{H} \mathbf{D})$  is the local error vector.  $\alpha$  controls the relative importance of the reproduction accuracy at region  $\mathcal{O}_q$ .

A LMS solution that minimizes (6.11) can be derived as

$$\begin{bmatrix} \alpha \widehat{\mathbf{S}}_q \mathbf{Q} \\ \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \alpha \widehat{\mathbf{S}}_q \mathbf{H} \\ \mathbf{H} \end{bmatrix} \mathbf{D}, \quad (6.12)$$

whose solution for  $\mathbf{D}$  is

$$\mathbf{D} = \begin{bmatrix} \alpha \widehat{\mathbf{S}}_q \mathbf{H} \\ \mathbf{H} \end{bmatrix}^{-1} \begin{bmatrix} \alpha \widehat{\mathbf{S}}_q \mathbf{Q} \\ \mathbf{Q} \end{bmatrix}. \quad (6.13)$$

The solution (6.13) considers not only the set of global coefficients, but also a linear mapping of these coefficients which correspond to the sound field in a sub-region  $O_q$

within the interested zone. By adding a weighing factor  $\alpha$ , the priority of the sub-region can be controlled. When  $\alpha = 0$ , the sub-region is ignored and the solution becomes identical to (6.3); if  $\alpha = 10$ , the local reproduction accuracy becomes 10 times more significant than the global accuracy, and as a result the driving signals  $\mathbf{D}$  would construct a sound field where the reproduction error within  $Q$  is approximately 10 times smaller than the global average.

Equation (6.13) can be extended to the multiple sub-region case, with each region controlled by a separate weighing factor

$$\mathbf{D} = \begin{bmatrix} \alpha_1 \hat{\mathbf{S}}_{q1} \mathbf{H} \\ \alpha_2 \hat{\mathbf{S}}_{q2} \mathbf{H} \\ \vdots \\ \alpha_n \hat{\mathbf{S}}_{qn} \mathbf{H} \\ \beta \mathbf{H} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \hat{\mathbf{S}}_{q1} \mathbf{Q} \\ \alpha_2 \hat{\mathbf{S}}_{q2} \mathbf{Q} \\ \vdots \\ \alpha_n \hat{\mathbf{S}}_{qn} \mathbf{Q} \\ \beta \mathbf{Q} \end{bmatrix}. \quad (6.14)$$

It can be seen that by setting  $\alpha_n = 1, \forall n$  and  $\beta = 0$ , (6.14) becomes a solution for spatial multizone sound field reproduction.

Compared to the spatial single zone sound field reproduction technique, this method allows more accurate reconstruction within some more critical areas; compared to the multizone reproduction method [109], which provides optimal reconstruction result within a few smaller regions but has no guarantee on the sound field in between these regions, this approach offers a balanced solution, where the whole zone of interest is reproduced while a few sub-regions are given higher priorities. This method is also more stable and more predictable than the multi-zone method, since the global channel matrix  $\mathbf{H}$  improves the condition of the matrix inversion in (6.14), as a result, the derived driving signal  $\mathbf{D}$  normally has limited power, and further regularization is normally unnecessary.

Another merit of this technique is when only a relatively small number of loudspeakers are accessible. Specifically, when the total number of loudspeakers is insufficient for producing the desired sound field for the whole interested zone, or the placement of the loudspeaker array disallows synthesizing sound waves impinging from certain directions, the normal LMS algorithm finds the best solution that gives the minimal average error across the whole zone. However, if the application has very high requirements on reproduction accuracy, for example active noise cancellation, this technique may fail to meet the demand, especially when insufficient loudspeakers are available. Using this new technique, it is possible to utilize the

limited amount of loudspeakers to reproduce the sound field accurately in certain critical sub-zones, such as the area where the listener is most likely positioned, while only slightly degrading the performance in the rest of the zone.

### 6.2.4 Simulation Results

The simulations are set up to synthesis a certain sound field based on microphone array recordings of the sound field. First, one or more point sources are set up to be primary sources, the sound fields in the interested region due to these primary sources are captured by a microphone array, this captured sound field becomes the desired field  $\mathbf{Q}$ . Then, the sound fields due to each secondary loudspeaker is recorded separately, which forms the channel matrix  $\mathbf{H}$ . Next, different algorithms, including the LMS mode matching, multizone and the proposed prioritized LMS mode matching are used to derive loudspeaker driving signals  $\mathbf{D}$  to best synthesize the sound field. For the latter two algorithms, one or more sub regions within the interested zone are selected as high priority area. Finally, the performance of the three methods are compared by plotting the synthesized sound fields on multiple elevations.

Figure 6.1 plots the case where 36 loudspeakers are arranged into three semi-circles, placed 1 meter away from the origin, each semi-circular array consists of 12 loudspeakers, spanning from  $\phi = 0$  to  $\phi = \pi$ . The three arrays have elevation angles  $\theta = 3\pi/8, \pi/2$ , and  $5\pi/8$ , respectively. A single point source is placed at  $(R, \theta, \phi) = (6 \text{ m}, 2\pi/5, \pi/2)$ , which generates a 540 Hz sine wave. The region of interest is a 0.7 m radius sphere centered at the origin, in addition, a priority sub-region is chosen to be a sphere of 0.4 m radius, centered at  $(0.2 \text{ m}, \pi, 0)$ .

Figure 6.1 (a) plots the real sound field produced by the primary source; (b) shows the wave field synthesis result using the LMS method; (c) plots the result using prioritized LMS method; and finally (d) plots the LMS result for the high priority region only. It can be seen that both the conventional LMS method and the proposed method yield acceptable reproduction result. Closer observation would show that the proposed method gives a more accurate reproduction of the sound field within the central circle, which represents the high priority zone. On the other hand, in the case of (d), although the number of loudspeakers provided is sufficient for an approximate reproduction in the high priority zone, due to the relatively long distance between the secondary sources and the reproduction area, the algorithm resulted in very large signal power, even with regularization parameters inserted.

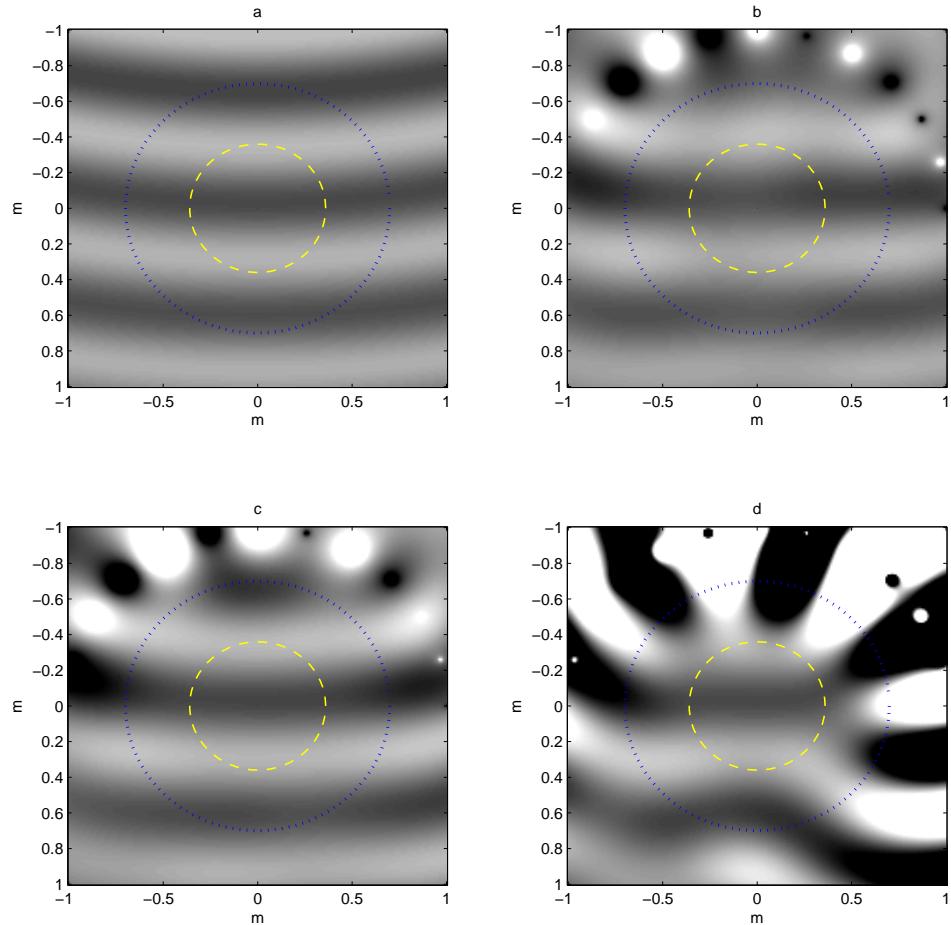


Figure 6.1: Comparison of three methods for sound field reproduction on  $\theta = \pi/2$  plane. (a) plots the desired sound field; (b) plots the synthesized sound field using LMS mode matching for global region; (c) plots the synthesized sound field using proposed prioritized region LMS method, and (d) plots the synthesized sound field using LMS mode matching for the high priority region only.

The proposed prioritized LMS algorithm, however, managed to synthesis the sound field with decent accuracy within the high priority zone, while also giving a reasonable reconstruction result for the entire reproduction region. Furthermore, the driving signals for the secondary sources are also limited, due to the global coefficients also acting as regularization parameters for the matrix solution. Clearly, in this case, the proposed method yields the best result among the three approaches.

When multiple priority regions are defined, the proposed algorithm will optimize the result for all of these regions according to the weighing factor given to these regions. A simulation result using the same setup as Fig. 6.1 is shown in Fig. 6.2. The only difference is that in Fig. 6.2, two priority sub-regions were defined, their locations were set to be  $(0.3 \text{ m}, \pi/2, 0)$ , and  $(0.4 \text{ m}, \pi/2, 0.8\pi)$ , and the radius are 0.4 m and 0.3 m, respectively. The weighing factor for both regions were set to be  $\alpha = 10$ , and the weighing for the global coefficients was  $\beta = 1$ . We note that only the proposed method was able to reconstruct the sound field. Although there are inevitable errors, the synthesized sound field was a good approximation to the desired one, especially within the two prioritized sub-regions.

It has been mentioned in Section 3 that the weighting of each priority zone can be adjusted independently to change the solution of the LMS algorithm, so as to further improve the reproduction accuracy of certain priority area. This is demonstrated in Fig. 6.3, where two priority zones of 0.3 m radius were chosen. Plot (a) shows the desired sound field, which is reconstructed using the secondary sources and the reconstruction errors are shown in plots (b) and (c). In (b), both regions were set to have the same weighing  $\alpha = 10$  while the global sound field has a weighing  $\beta = 1$ . In (c), however, the upper right priority zone has its weighting increased to  $\alpha_1 = 30$ , while the other priority zone's weighing is reduced to  $\alpha_2 = 5$ . It can be observed that compared to (b), (c) gives a more accurate reproduction in the upper right region, while the region on the left has a slightly worse accuracy. The global reproduction accuracy also degraded slightly, due to the increased weighting for priority zone  $Q_1$ .

Table 6.1 lists a series of simulated data. The normalized mean square error and average error percentage are calculated for different scenarios, covering LMS based sound field synthesis with 0, 1, 2 and 3 high priority sub-regions. In order to acquire more accurate synthesis results, a total of 60 loudspeakers are place in a semi-circle one meter away from the origin. Despite the increased number of secondary sources, there is still insufficient secondary sources to completely synthesis the sound field. This is to show the advantage of the proposed algorithm in the cases where insufficient secondary loudspeakers can be used.

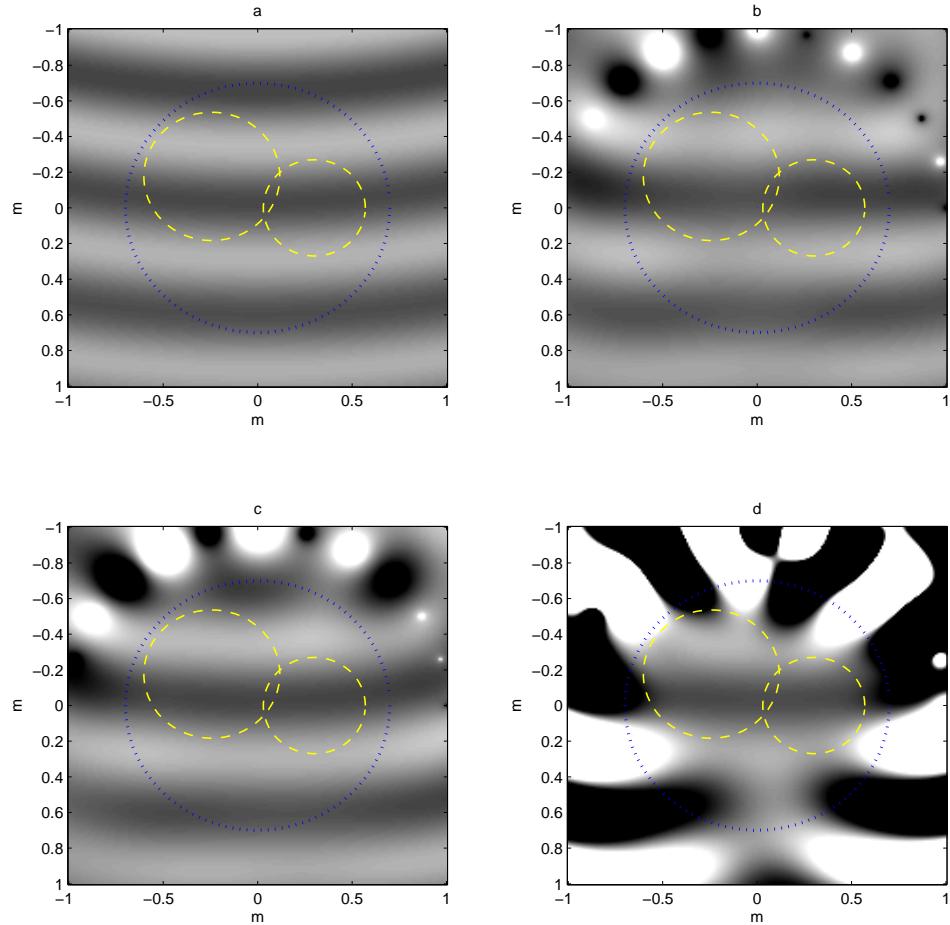


Figure 6.2: Comparison of three methods for sound field reproduction with 2 high priority zones. (a) plots the desired sound field; (b) plots the synthesized sound field using LMS mode matching for global region; (c) plots the synthesized sound field using proposed method with two high priority regions, and (d) plots the synthesized sound field using LMS mode matching for the high priority regions only.

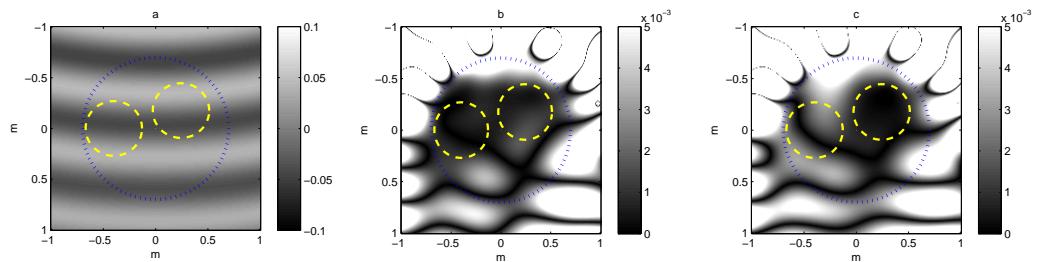


Figure 6.3: Reproduction error plots for different high priority zone weight settings;(a) desired sound field, (b) reproduction error with  $\alpha_1 = \alpha_2 = 10$ , and (c) reproduction error with  $\alpha_1 = 30, \alpha_2 = 10$ .

	Max Order	Coefficient No.	Weight	Mean Square Error	Error %
Global	10	121	1	100%	7.24%
Global	10	121	1	138.53%	7.75%
Zone 1	3	16	10	0.26%	0.48%
Global	10	121	1	131.49%	7.41%
Zone 1	4	20	5	5.43%	1.93%
Global	10	121	1	185.24%	8.53%
Zone 1	4	20	4	7.13%	2.22%
Zone 2	3	16	12	0.59%	0.70%
Global	10	121	1	135.38%	8.91%
Zone 1	3	16	3	12.75%	3.85%
Zone 2	2	9	10	0.43%	0.83%
Zone 3	4	20	2	37.21%	7.67%

Table 6.1: Reproduction accuracy for different priority zone settings

The first row in Table 6.1 shows the performance of the normal LMS algorithm as a reference. The mean square error of the LMS method without prioritized control is used as a reference for comparison of mean square error of different setups. It can be seen that an average error of 7.24% is observed from a total of 121 sound field coefficients. The next four rows show the simulation results for 1 priority zone, with its maximum order set to 3 and 4, respectively. In both situations, an increase in the global mean square error is observed; the global error percentage also saw a slight increase. Most importantly, the error percentages in the priority zones are much smaller than the global error percentage (0.48% and 1.93%), the error in row 5, Table 6.1 is greater because of the larger size of the priority region, as well as the lower weight applied to the priority zone.

The rest of Table 6.1 shows the simulation results for 2 and 3 high priority regions, in both cases each sub-region is given a different weighting factor. The effect of the weighting factors can be easily seen, as the sub region with the largest weighting assigned always result in the lowest error percentage, while the regions with low weights and large radii only see a slight improvement over the global zone. Another observations is that the global synthesis precision degrades more greatly when a large weight is given to the high priority zone. Therefore, in practice, we recommend to choose the weightings and the radii of the sub-regions according to the needs, rather than simply using overly large values.

In order to investigate the impact of the weighting factor on the prioritized sub-region and the global sound field, a series of simulations are carried out. The

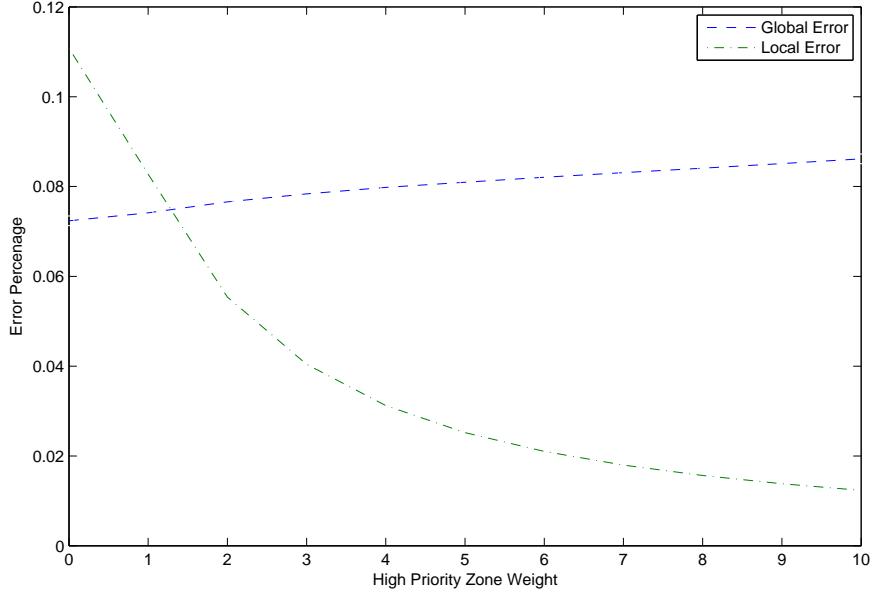


Figure 6.4: Effect of the weighting factor on the local and global error percentage

simulations evaluate the reproduction error of a sub-region, located at  $(R, \theta, \phi) = (0.1 \text{ m}, \pi/2, \pi)$  with a maximum order of 3, within the global region of interest whose maximum order is 10. The global region is given the weight  $\beta = 1$  while the weighting factor for the sub-region varied from 0 to 10. The resulting error percentages are plotted in Figure 6.4.

It is easy to identify the exponentially decaying line in Figure 6.4 as the sub-region error percentage. As the weighting factor changed from 0 to 10, the average reproduction error within this sub-region went from 11% down to a very low 1.2%. Meanwhile, the global error grew slowly from 7.2% to 8.6%. Further investigation shows that the rate at which the global error increases tend to reduce as the local weighting becomes larger.

### 6.2.5 Observations and insights

We can observe that when using the proposed method to synthesis a sound field, the global reproduction accuracy is slightly reduced, in exchange for the greatly improved reproduction quality in the prioritized sub-region. However, the tradeoff between the local and global region depends on the ratio of size between them, i.e., if the local region is very small compared to the global region, then the reproduction accuracy within it can be improved significantly, with minimum loss on the global

accuracy; if the local region is large, the accuracy gain may become smaller. In practice, Figure 6.4 can be used as a trade-off guidance when choosing the appropriate weighting factors for each region.

## 6.3 Evaluation of spatial active noise cancellation performance using acoustic potential energy

This section presents a novel metric to evaluate the performance of spatial active noise cancellation (ANC) systems. We show that the acoustic potential energy within a spherical region can be expressed by a weighted squared sum of spherical harmonic coefficients. This metric allows convenient evaluation of spatial ANC performance using a spherical microphone array.

In order to demonstrate the usefulness of the proposed metric, we carried out experiments using a MIMO system with 5 error microphones and 2 secondary loudspeakers attempting to cancel a noise field caused by a single primary source. The microphone recordings required to evaluate the spatial performance metric were obtained using a 32 channel Eigenmike [110].

### 6.3.1 Background

For the successful development of an ANC system, it is important to accurately measure their noise reduction capability over space, especially at the design stage. At present, the performance of ANC systems is analyzed in terms of (i) sound pressure at the error microphones or (ii) recordings from a secondary microphone(s) in the cancellation region. The first approach is widely used in theory, where noise reduction performance is characterized by the average noise reduction in decibels at the error microphones [111]. The second approach is mostly used with human head shaped mannequins with 2 microphones placed at the ear locations to interpret the noise reduction levels experienced by humans [112, 113]. While both of the above methods are adequate to obtain an acceptable measure of the ANC system performance, their accuracy in terms of spatial coverage is limited due to the limited number of measurement points and the sparse nature of the spatial sampling.

Here, we propose an improved metric to evaluate the noise reduction in spatial regions. It is formulated in terms of the spherical harmonic decomposition of sound-fields and requires the measurements from a secondary microphone array distributed over a spherical surface, preferably enclosing the center of the region of interest. The

spatial metric is defined as the acoustic potential energy inside a spherical region, and we formulate it in terms of the aforementioned microphone array recordings. A similar spatial metric was introduced in [114] for rectangular enclosures, where the acoustic potential energy was described in terms of room modes. However, the results were limited to simulations and the extraction of room modes is difficult in practical applications where the natural modes of a room depends greatly on the geometry of the room.

The proposed potential energy method calculates the average noise level in the entire spatial region, therefore there is no need to take multiple samplings of the control region, which simplifies the process of ANC performance evaluation. Compared to [114], our method represents the potential noise using spherical harmonic coefficients rather than room modes, which can be conveniently captured by spherical microphone arrays. The main advantage of this approach is its applicability to any arbitrary enclosure and its independence from the ANC system of use. Therefore our method is particularly suitable for environments with irregular geometry and a fixed control region with moderate size, such as vehicle and aircraft cabins.

### 6.3.2 Calculation of the acoustic potential energy

Our goal is to find a metric that best represents the average noise energy level within a spatial region. A sensible measure of the average noise level inside a region is the acoustic potential energy [114], which is defined as the integral of squared sound pressure over the entire region. In a spherical region of interest, it can be represented by

$$E_p(k) = \frac{1}{4\rho_0 c^2} \int_S |P(r, \theta, \phi, k)|^2 dS \quad (6.15)$$

where  $\int_S dS = \int_0^R \int_0^\pi \int_0^{2\pi} r^2 dr \sin(\theta) d\theta d\phi$  denote the integral over a sphere. Using (2.1), we can decompose the integral of sound energy as [115]

$$\int_S |P(r, \theta, \phi, k)|^2 dS \quad (6.16)$$

$$= \int_0^R \int_0^\pi \int_0^{2\pi} P(r, \theta, \phi, k) P^*(r, \theta, \phi, k) r^2 dr \sin(\theta) d\theta d\phi \quad (6.17)$$

$$= \sum_{l,m} C_{lm}(k) C_{lm}^*(k) \int_0^R j_l^2(kr) r^2 dr, \quad (6.18)$$

where the orthogonal property of the spherical harmonics (2.5) was used. Therefore (6.15) can be expressed using the spherical harmonic coefficients as

$$E_p(k) = \frac{1}{4\rho_0 c^2} \sum_{l,m} |C_{lm}(k)W_l(k)|^2, \quad (6.19)$$

where  $\rho_0$  denotes the density of the media and  $c$  is the speed of sound, and we define

$$W_l(k) \triangleq \left( \int_0^R j_l(kr)^2 r^2 dr \right)^{1/2}. \quad (6.20)$$

The above result shows that the acoustic potential energy within a spherical region is given by a sum of squared spherical harmonic coefficients with the weighting  $W_l(k)$ .

The commonly used criteria for ANC performance evaluation measures the attenuation of the noise energy at microphone positions, the microphones are either the error microphones themselves, or some additional microphones placed within the region of interest. In the former case, it is difficult to gain any insight into the spatial ANC performance of the system, due to lack of sampling of the noise level inside the control region. When additional microphones are utilized to measure the noise level inside the control region, it is necessary to sample the control region at multiple locations in order to have a complete evaluation of the noise attenuation.

On the other hand, the proposed potential energy method calculates the average noise level in the entire spatial region, therefore there is no need to take multiple samplings of the control region, which simplifies the process of ANC performance evaluation. In [114], the potential energy criteria is applied to evaluate the noise level inside rectangular cabins. However, in practical scenarios, the natural modes of a room depends greatly on the geometry of the room, and measuring these modes in a practical environment is very difficult. Our method, on the other hand, represents the potential noise using spherical harmonic coefficients, which can be conveniently captured by spherical microphone arrays. Therefore our method is particularly suitable for environments with irregular geometry and a fixed control region with moderate size, such as vehicle and aircraft cabins.

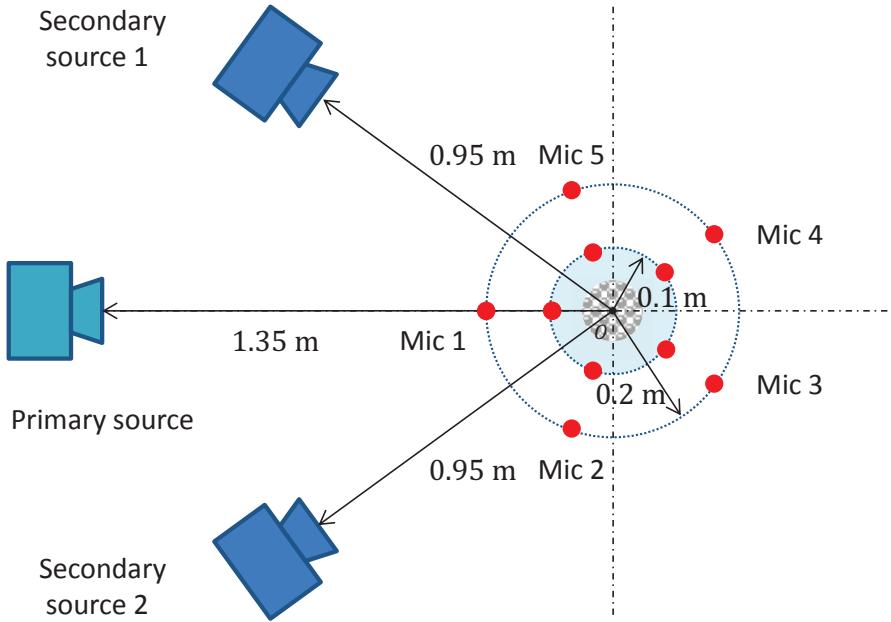


Figure 6.5: Loudspeaker and microphone placement for the experiment.

### 6.3.3 Performance evaluation

#### Experiment setup

Figure 6.5 shows the hardware configuration of the system, where the control region is defined as a spherical area with 0.1 m radius, which approximately covers the size of a human head. Five AKG CK92 omnidirectional microphones are placed evenly on the horizontal plane boundary of the region, which act as the error sensors.

In order to investigate the differences in ANC performance due to different error microphone setups, we vary the radius of the error microphone array, as well as the number of active microphones in the array. The array radius is varied between 10 cm and 20 cm, and the microphones used in each experiment are either (i) all five microphones, or (ii) “Mic 2” and “Mic 5” shown in Fig. 6.5 only. This results in a total of four combinations of array radius and microphone number.

Three TANNOY 600 loudspeakers are used as the primary and secondary sources. The two secondary speakers are placed on either sides of the primary source, forming an angle of 72 degrees.

The error microphone signals are transmitted to a PC, which performs the adaptive ANC algorithm and generates the secondary loudspeaker driving signals in real time. Since the focus of this experiment is not on the performance of the MIMO ANC algorithm itself, the reference signal is obtained directly from the electronic

signal path of the primary source, rather than using a separate reference microphone. This eliminates the feed back signal path from the secondary sources to the reference sensor which may affect the ANC performance.

An Eigenmike is placed at the center to monitor the noise field within the control region. Although the Eigenmike is capable of capturing spatial sounds up to 4 th order at 4 kHz, we only focus on the lower frequency sounds (up to 800 Hz and 1st order). This is because at higher frequencies, the second order spherical harmonics begins to have a higher contribution towards the sound energy close to the boundary of control region, but the Eigenmike is unable to capture the second order sound field at that frequency, due to its smaller radius (4.2 cm) compared to the control region.

A separate computer is used to process the audio signal recorded by the Eigenmike and calculate the potential energy while the ANC system is running. The Eigenmike is not involved in the signal path of the ANC system in any way.

Both narrow-band and wide-band signals are used as the primary noise for the experiments. The narrow-band signals are sine waves with frequencies 100 – 800 Hz; the wide-band signal is generated by filtering a in-car noise recording through a 100 – 800 Hz bandpass filter.

For narrow-band experiments, a sine wave is played through the primary speaker, then we calculate the average sound energy recorded by each error microphone with and without ANC, and calculate the attenuation of the noise energy due to ANC. The attenuation of the average sound energy within the control region is measured in the same way.

For wide-band experiments, we play the wide-band noise and record a section of signal from the error microphones as well as the Eigenmike while the ANC system is not active, then repeat the recording with ANC active and fully converged. We then calculate the average frequency spectrum of each recorded section. The playback and recordings are synchronized such that the same section of noise signal is recorded each time.

### 6.3.4 Result analysis

#### Effect of microphone array radius

In order to investigate the effect of microphone array radius on the performance of the ANC system at different frequencies, as well as the differences between the potential energy criteria and microphone pressure criteria, we conduct the narrow-

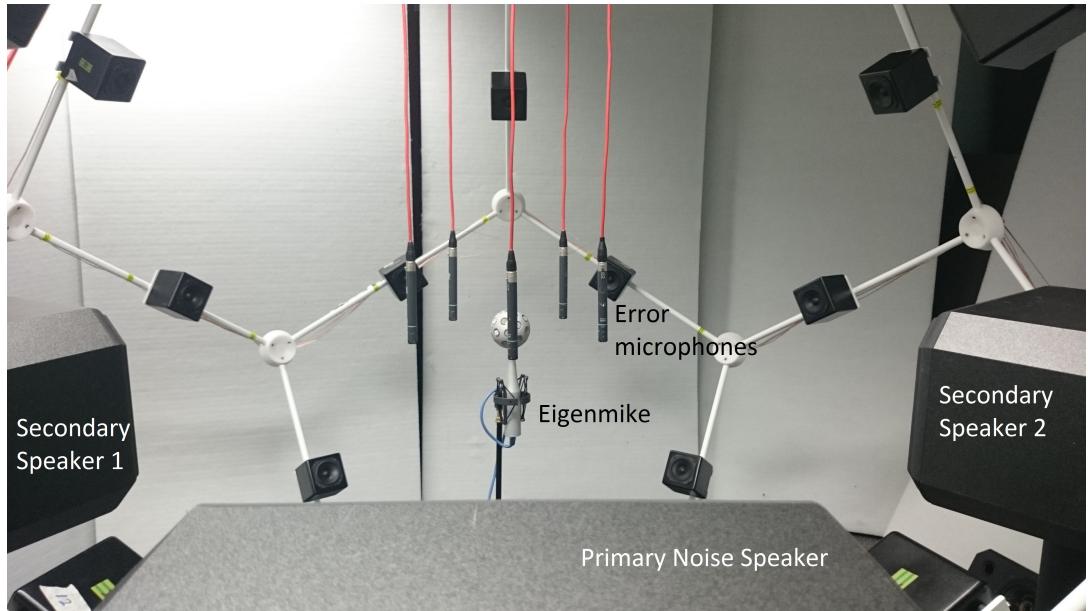


Figure 6.6: Picture of the experiment setup, the small loudspeakers in the background are not used in the experiment.

band ANC experiments for microphone array radius  $r = 10$  cm and  $r = 20$  cm, utilizing all 5 microphones. The energy attenuation of both potential energy and microphone received signal are shown in Fig. 6.7 (a).

It can be seen from Fig. 6.7 (a) that for all configurations, overall the noise attenuation becomes smaller as the frequency increases. A potential energy attenuation of over 10 dB can be achieved for frequencies up to 450 Hz if the microphones are placed on the boundary of the control region; when the microphones are placed further away, the attenuations are worse by 5 – 10 dB for frequencies above 300 Hz. This result is intuitive because when the microphones are placed further away, the correlation between the sound pressure inside the control region and the sound pressure at microphone positions becomes smaller, and this is especially true for higher frequencies, where the wavelength is shorter. However, it can be seen that at very low frequencies (below 200 Hz), both radii result in very similar ANC performance in terms of potential energy attenuation.

Comparing the curves corresponding to microphone signal attenuation with those corresponding to potential energy attenuation, it can be found that when the microphones are placed at the boundary of the region, the microphone signals are a good indication of the potential energy inside the region; when the microphones are placed further away, however, the microphone signal attenuations become much smaller than the noise attenuation observed inside the region. In this case, the mi-

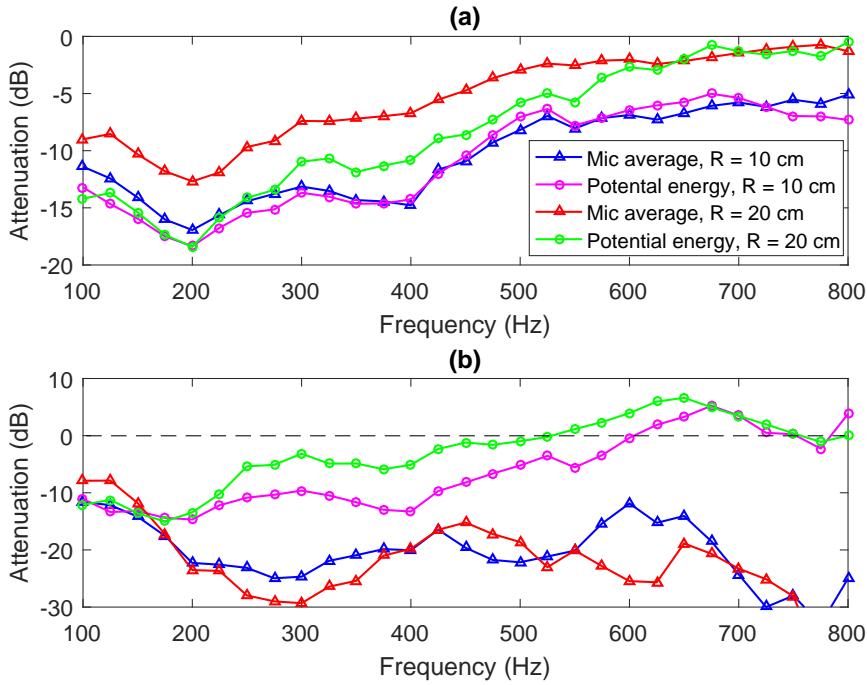


Figure 6.7: Average narrow-band noise energy attenuation at control region and microphone locations using 5 error microphones (a) and 2 error microphones (b). Legend of (b) is the same as (a).

crophone signals are no longer a good indication of the ANC systems's performance.

### Effect of microphone number

The effect of the number of error microphones is also investigated. For this purpose, we repeated the narrow-band experiments with only two error microphones active, and the noise attenuation results are shown in Fig. 6.7 (b). From this figure, it can be observed that both the microphone signal attenuations and the potential energy attenuations become very different from the case where 5 microphones are used. In particular, the microphone signals can achieve more than 10 dB attenuation for all frequencies, and the attenuation does not decay with increased frequency; on the other hand, the potential energy attenuation is significantly worse compared to the 5 channel case, and the value even became positive (higher noise level with ANC active) at some higher frequencies.

The cause of this phenomenon is that the number of error microphones is equal to the number of secondary sources, therefore a solution always exists to significantly

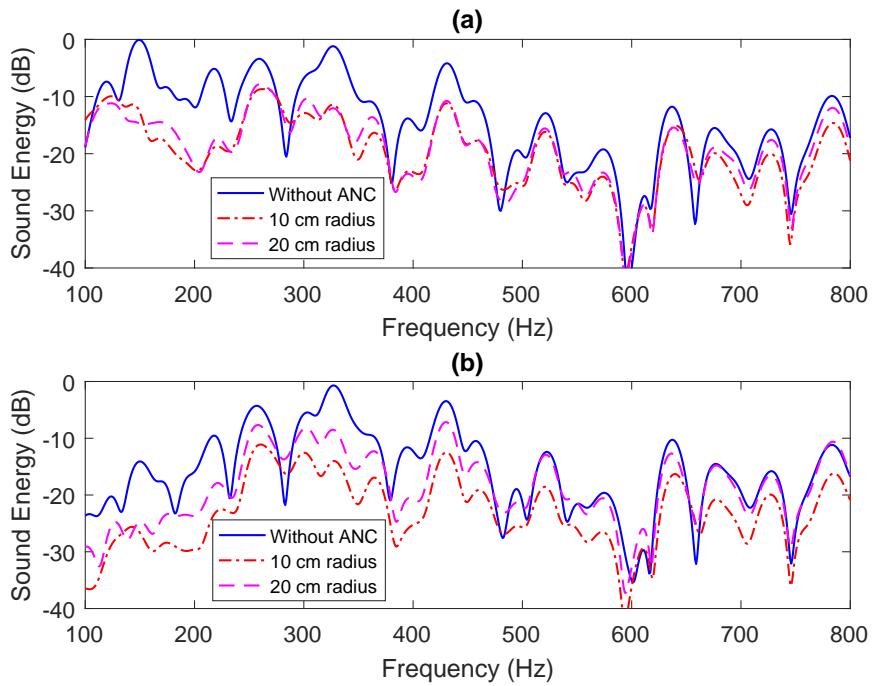


Figure 6.8: Spectrum of potential energy at control region (a) and microphone locations (b) using wide-band noise signal.

reduce the sound pressure at the microphone positions, but often at the cost of very high secondary source driving signals. Since the two microphones provide a complete coverage of the control region, the potential energy inside the region becomes less controllable compared to the 5-channel case, and in some extreme cases, this results in positive attenuation inside the region. In this case, the microphone signals are not a good indication of the ANC performance inside the control region, except at very low frequencies (below 150 Hz), as can be seen in Fig. 6.7 (b).

### Wide-band performance

In order to investigate the system's wide-band performance, we compute the average spectrum of the microphone signals with and without ANC, as well as the spectrum of observed potential energy within the control region. All five error microphones are used in this experiment. The spectrums are shown in Fig. 6.8, where Fig. 6.8 (a) plots the potential energy, and Fig. 6.8 (b) plots the spectrum of microphone signals. The difference of spectrums at low frequency between Fig. 6.8 (a) and (b) is due to the presence of high pass filters in the error microphones' amplification circuitry.

From Fig. 6.8 (a) it can be seen that overall the wide-band ANC performance agrees with the corresponding narrow-band results shown in Fig. 6.7 (a), where an attenuation of 10 dB and more can be achieved for most frequencies below 400 Hz, while at higher frequencies, the attenuation gradually reduces to 3–5 dB. Comparing the curves corresponding to 10 cm and 20 cm microphone array radius, it can be seen that the low frequency ANC performance of the two configurations are very similar, with the 10 cm configuration being superior at certain frequency ranges. At higher frequencies, the 10 cm configuration yields consistently better attenuation than the 20 cm configuration, which also agrees with Fig. 6.7 (a), although the attenuation is slight worse for the 10 cm radius case.

In the case of Fig. 6.8 (b), the attenuation observed at microphone positions differ greatly for the two radius configurations. For 10 cm radius, the attenuation is greater than 5 dB for nearly all frequencies between 100–800 Hz, while for the 20 cm setup, the attenuation becomes negligible above 475 Hz. Up on closer observation, it can be seen that the microphone attenuation at 10 cm radius is 2–3 dB greater than that of the potential energy inside the control region. Therefore, neither of the two microphone position configurations truthfully reflect the ANC performance within the control region, although the result is more accurate when the microphones are placed closer to the control region.

### 6.3.5 Observations and insights

Overall, we can see that the error microphone signals cannot be used to reliably measure the noise attenuation inside the control region, especially when the number of error microphones is small, or when the microphones are not place close to the control region. On the other hand, using the proposed method, we can use a spherical microphone array placed in the center of the region to conveniently and reliably monitor the potential energy within the region.

In addition, in terms of the performance of the ANC system, we notice that its wide-band performance largely agrees with the narrow-band performance, which confirms the capability of the implemented ANC algorithm. Also, we notice that at higher frequencies, although sometimes an attenuation of the noise cannot be observed by the microphones, the noise level within the control region still saw a small (around 3 dB) reduction. We note that this is because the wavelength of the noise is still on par with the radius of the array, if the array had an even larger radius, this phenomenon may not be observed.

## 6.4 In car spatial ANC performance analysis

In this section, we evaluate a passenger car’s integrated loudspeakers’ noise cancelling capabilities by analyzing the in-car noise field and the loudspeaker responses. We show that the noise field can be decomposed into several basis noise patterns. Noise field measurements are carried out for both single seat and multiple seats scenarios, and a series of analysis are performed to estimate the ANC capabilities of the in-car loudspeakers.

### 6.4.1 Background

The application of noise cancellation methods to minimize interior cabin noise has been a key topic of research in the automobile industry for the last 15 – 20 years [116]. Initially, this problem was approached via passive noise cancellation methods, which use acoustic treatments such as structural damping and acoustic absorption. However, with the growing need to improve fuel efficiency, there has been more preference on lighter bodies and smaller engines, which has significantly increased the structural vibration and consequent interior noise, predominantly at low frequencies (e.g. 0 – 500 Hz) [117]. As passive methods were least effective with low frequency noise, active methods were developed where secondary loudspeakers were proposed to attenuate measured noise inside the cabin [6, 117–120]. With modern in-car entertainment systems providing 4 – 6 built-in loudspeakers, the addition of an active noise cancellation systems is considered to involve no greater cost [7].

To the best of our knowledge, the existing in-car MIMO controllers are constrained to a set of arbitrary observation points. As a result, spatial control over continuous regions is limited and made worse with increased frequency. Addressing this issue, we focus this work on modeling vehicle-interior noise over a continuous spatial region such that noise control can be achieved over the region with size similar to a human head for frequencies up to  $f = 500$  Hz. We also derive the maximum attenuation levels for a given speaker arrangement so that industrial designers can investigate the potential noise cancellation capability of a given loudspeaker system for various noise sources and driving conditions. All of the analysis we perform are based on acoustic measurements taken in a real in-car environment.

### 6.4.2 Problem Formulation

Denote the unwanted noise pressure at a point  $\mathbf{x}$  as  $P_n(\mathbf{x})$ , and the sound pressure due to the loudspeakers as  $P_c(\mathbf{x})$ , the average residual noise energy within the interested region  $S$  can be expressed as

$$\int_S |P_r(\mathbf{x})|^2 dS = \int_S |P_n(\mathbf{x}) + P_c(\mathbf{x})|^2 dS. \quad (6.21)$$

A complete in-car active noise cancellation system consists of many components, each component may have an impact on the system's overall performance. In this paper, we aim to evaluate the potential performance of in-car loudspeakers on ANC applications by estimating the minimum values of  $\int_S |P_r(\mathbf{x})|^2 dS$  for various frequencies, based on the information on in-car noise field and the acoustic characteristics of the car's integrated loudspeakers, so as to see if the integrated loudspeakers are the potential bottleneck for in-car ANC systems.

Using the acoustic potential energy as the metric for spatial sound level, in the active noise cancellation scenario, the residual noise field  $P_r(\mathbf{x})$  in (6.21) thus have the average energy

$$\int_S |P_r(\mathbf{x})|^2 dS = \int_S |P_n(\mathbf{x}) + P_c(\mathbf{x})|^2 dS \quad (6.22)$$

$$= \sum_{l,m} |(C_{lm}^{(n)} + C_{lm}^{(c)})W_n|^2, \quad (6.23)$$

where the expression (6.19) is used, and  $C_{lm}^{(n)}$  and  $C_{lm}^{(c)}$  are the spherical harmonic coefficients representing the noise field and the loudspeaker anti-noise field, respectively.

We then move on to derive an estimation of  $\int_S |P_r(\mathbf{x})|^2 dS$ , by analyzing the noise field and loudspeaker channel characteristics.

### 6.4.3 Noise field characterization

For a certain driving condition, we assume that the random noise field within  $S$  can be seen as a weighted combination of a number of fixed, basis noise patterns, or noise modes [121], each driving condition may have a different set of basis. Then the noise field pressure within  $S$  at any time under a fixed driving condition can be

decomposed as

$$P_n(\mathbf{x}) = \sum_i g_i P_i(\mathbf{x}), \quad (6.24)$$

where  $P_i(\mathbf{x})$  denotes the  $i$ th basic noise pattern at  $\mathbf{x}$ , and  $g_i$  are some random weighing factors for each noise pattern. Theoretically an infinite number of modes are needed to fully describe an arbitrary noise field, however for a relatively small region and low frequencies, only a small number of noise modes are required for a good approximation of the noise field [121]. Using the spherical harmonics decomposition (2.1) to decompose the noise field  $P_n(\mathbf{x})$  and the basis patterns  $P_i(\mathbf{x})$ , we can express each noise field coefficient  $\alpha_{nm}^{(n)}$  using the corresponding coefficient  $C_{lm}^i$  of every basis pattern,

$$C_{lm}^{(n)} = \sum_i g_i C_{lm}^i. \quad (6.25)$$

We can write all the coefficients in a vector form such that  $\mathbf{C} = [C_{00}^{(n)}, C_{11}^{(n)}, \dots]^T$ , and  $\mathbf{C}_i = [C_{00}^i, C_{11}^i, \dots]^T$ , then from (6.25) we have

$$\mathbf{C} = \sum_i g_i \mathbf{C}_i. \quad (6.26)$$

In order to reflect the relative impact of each spherical harmonic coefficient on the overall noise level within  $S$ , we define the weighted coefficient vector  $\mathbf{c}_i = [C_{00}^i W_0, C_{11}^i W_1, \dots]^T$ , and by multiplying both sides of (6.26) with a diagonal matrix  $\mathbf{W}$  with  $\text{diag}\{\mathbf{W}\} = [W_0, W_1, W_1, W_1, W_2, \dots]^T$ , we have

$$\mathbf{c} = \sum_i g_i \mathbf{c}_i, \quad (6.27)$$

where  $\mathbf{c}$  represents the random noise field in  $S$  and  $\|\mathbf{c}\|^2 = \int_S |P_n(\mathbf{x})|^2 dS$ . Similar to the modal domain MUSIC DOA algorithm [122], we can find a set of  $\mathbf{c}_i$  by calculating the autocorrelation matrix  $E\{\mathbf{c}\mathbf{c}^H\}$ , and then decompose  $E\{\mathbf{c}\mathbf{c}^H\}$  to acquire a set of orthonormal eigenvectors and their corresponding eigenvalues. Unlike the MUSIC DOA method which utilizes the noise subspace eigenvectors, we select the signal subspace eigenvectors to be  $\mathbf{c}_i$ , which correspond to the eigenvalues  $\lambda_i$  whose values are significant. The eigenvalues indicate the energy distribution of the overall noise field among the basis noise patterns, and  $E\{|g_i|\} = \lambda_i$ .

Therefore the expected average noise power within  $S$  can be calculated as

$$E\left\{\int_S |P_n(\mathbf{x})|^2 dS\right\} = E\{\|\mathbf{c}\|^2\} = \sum_i \|\lambda_i \mathbf{c}_i\|^2. \quad (6.28)$$

Through decomposing the noise field into basis noise patterns, we gain more insight in the dimensionality/sparsity of the noise field. A noise field of high order may have a compact representation using (6.24). Furthermore, additional signal analysis methods such as direction-of-arrival (DOA) estimation may be applied on the basis noise patterns to identify principal noise sources, which helps in determining optimal loudspeaker placement for ANC purposes when designing the vehicle.

#### 6.4.4 Residual noise level estimation

We model the characteristics of loudspeakers at each frequency bin through the wave domain channel matrix  $\mathbf{H}$ , which describes the loudspeakers' response over the entire region of interest,

$$\mathbf{H} = \begin{bmatrix} H_{00}^1 & H_{00}^2 & \dots & H_{00}^V \\ H_{11}^1 & H_{11}^2 & \dots & H_{11}^V \\ \vdots & \vdots & \ddots & \vdots \\ H_{LL}^1 & H_{LL}^2 & \dots & H_{LL}^V \end{bmatrix} \quad (6.29)$$

with  $H_{lm}^v$  being the spherical harmonic coefficient of order  $l$  and mode  $m$ , defined in the region of interest  $S$ , due to the  $v$ th loudspeaker playing a unit signal at one frequency. For an  $L$ th order region  $S$  and an array of  $V$  independent loudspeakers, the size of  $\mathbf{H}$  is  $(L + 1)^2$ -by- $V$ .

Since the noise field can be completely described by its eigenvectors  $\mathbf{c}_1, \mathbf{c}_2\dots$ , we can estimate the noise cancellation performance by comparing the eigenvectors with the loudspeaker channels. In particular, we define the weighted channel matrix  $\mathcal{T} = \mathbf{W}\mathbf{H}$ , where  $\mathbf{W}$  is the diagonal matrix defined in Section 6.4.3.

Then we can solve for the loudspeaker driving signal solution that minimizes (6.22) for each basis noise field pattern defined by  $\mathbf{c}_i$ , which can be derived as

$$\mathbf{D} = -(\mathcal{T}^H \mathcal{T})^{-1} \mathcal{T}^H \mathbf{c}_i. \quad (6.30)$$

The residual error vector is then

$$\mathbf{e}_i = \mathbf{c}_i + \mathcal{T}\mathbf{D} = (\mathbf{I} - \mathcal{T}(\mathcal{T}^H\mathcal{T})^{-1}\mathcal{T}^H)\mathbf{c}_i. \quad (6.31)$$

The driving signal solution (6.30) is essentially the Least Mean-Square Error (LMS) solution over the continuous space  $S$ , instead of the LMS solution based on a number of discrete spatial sampling points which is commonly used in existing car ANC systems.

We use the eigenvalues  $\lambda_i$  as well as the original and residual noise field vectors,  $\mathbf{c}_i$  and  $\mathbf{e}_i$ , respectively, to express the noise cancelling performance, and the overall expected noise power reduction ratio can be given using (6.28)

$$e = \frac{E\{\int_S |P_r(\mathbf{x})|^2 dS\}}{E\{\int_S |P_n(\mathbf{x})|^2 dS\}} = \frac{\sum_i \|\lambda_i \mathbf{e}_i\|^2}{\sum_i \|\lambda_i \mathbf{c}_i\|^2}, \quad (6.32)$$

where the term  $\mathbf{c}_i$  in (6.32) can be omitted since  $\mathbf{c}_i$  are orthonormal.

#### 6.4.5 Experiment on a single passenger seat

##### Experiment setup

In this experiment, we use the method developed in the previous sections to analyze the potential noise cancellation performance of the loudspeakers installed in a car (2005 Ford Falcon XR6 sedan).

We use an Eigenmike to measure the in-car noise field; the region of interest is chosen to be a spherical area with 10 cm radius, located at the head position of the frontal passenger seat. The radius of the region is larger than that of the EigenMike (4.2 cm), therefore we only analyze the sound field for frequencies below 500 Hz, within this frequency range, only the 0th and 1st order sound field harmonics are active inside the region of interest [53], which can be reliably measured by the Eigenmike placed in the center of the region. Also, spectral analysis of the in-car noise indicate that the majority of the noise power lie below 500 Hz (an example of the noise spectrum is shown in Fig 6.11), thus the noise cancelling performance within this frequency band is indicative of the overall cancelling quality.

The vehicle has four full-band loudspeakers installed, two of which are integrated in either of the front doors, while the other two are placed behind each rear seat. Unfortunately, the car's audio system can only play stereo signals, which means the two loudspeakers on either side cannot be driven separately, and always play the



Figure 6.9: Picture of the EigenMike installed in a Ford Falcon XR6.

same signal.

In order to characterize the noise field, we record the in-car noise under various driving conditions. We also recorded the noise fields due to engine and air-conditioner while the car is stationary. For each driving condition, a 10-second-long recording is separated into 100 snapshots, we then calculate the sound field coefficients for each snapshot and at every frequency bin, and finally calculate the coefficient covariance matrix of all the 100 snapshots. The covariance matrix used as the estimation of  $\mathbf{cc}^H$ , and is used for further data analysis.

The loudspeaker channel matrix is obtained by measuring the spatial response at the region of interest due to the left channel and right channel separately using the Eigenmike, the sound field coefficients for each frequency bin are calculated in the same way as the noise field samples. The 1st order sound field and the stereo speaker system result in a 4-by-2 channel matrix for each frequency bin.

When calculating the residual noise field vector  $\mathbf{e}_i$ , we include a small regular-

100 km/h	100 Hz	200 Hz	300 Hz	400 Hz	500 Hz
$\lambda_1$	137.7	29.58	18.43	15.02	9.360
$\lambda_2$	6.578	1.049	1.286	1.270	1.134
$\lambda_3$	2.610	0.651	0.814	0.800	0.679
$\lambda_4$	1.475	0.418	0.645	0.596	0.414
Engine Only	100 Hz	200 Hz	300 Hz	400 Hz	500 Hz
$\lambda_1$	74.50	50.97	17.65	8.697	5.029
$\lambda_2$	2.217	1.807	0.930	0.997	0.578
$\lambda_3$	0.862	0.706	0.496	0.574	0.442
$\lambda_4$	0.557	0.400	0.341	0.369	0.172

Table 6.2: Table of noise field eigenvalues for freeway driving condition and pure engine noise

ization parameter  $\beta$ , such that (6.31) becomes

$$\mathbf{e}_i = (\mathbf{I} - \boldsymbol{\mathcal{T}}(\boldsymbol{\mathcal{T}}^H \boldsymbol{\mathcal{T}} + \beta \mathbf{I})^{-1} \boldsymbol{\mathcal{T}}^H) \mathbf{c}_i, \quad (6.33)$$

with  $\beta = 0.01$ . The regularization prevents severe ill-conditioning of the matrix inversion, thereby preventing the occurrence of very high secondary loudspeaker volumes.

### Experimental data analysis

By diagonalizing the estimated coefficient covariance matrices acquired from the recordings from various driving conditions, we obtained the eigenvalues for every case and each frequency bin. The eigenvalues are given in Table 6.2 for the freeway driving condition and a pure engine noise recording. For the freeway recording, the car was driven on a freeway at 100 km/h, with air conditioning turned to low; for the engine noise recording, the car was parked in a quiet place with air conditioning switched off, and the engine ran at various rpm during the recording.

From Table 6.2 we see that the eigenvalues decay quickly as the frequency increases, which indicates the shape of the noise spectrum, where the lower frequencies are more dominant.

Furthermore, one may notice that the first eigenvalues for both cases and each frequency bin are much larger than the other 3 eigenvalues, this is particularly noticeable at lower frequencies, and the same phenomenon is observed in all the other driving scenarios. In Section 6.4.3 we showed that each eigenvector of the covariance matrix corresponds to a specific spatial sound field pattern, with the relative importance of each pattern indicated by its eigenvalue. This result shows that there is one dominant noise pattern in the region of interest for each frequency

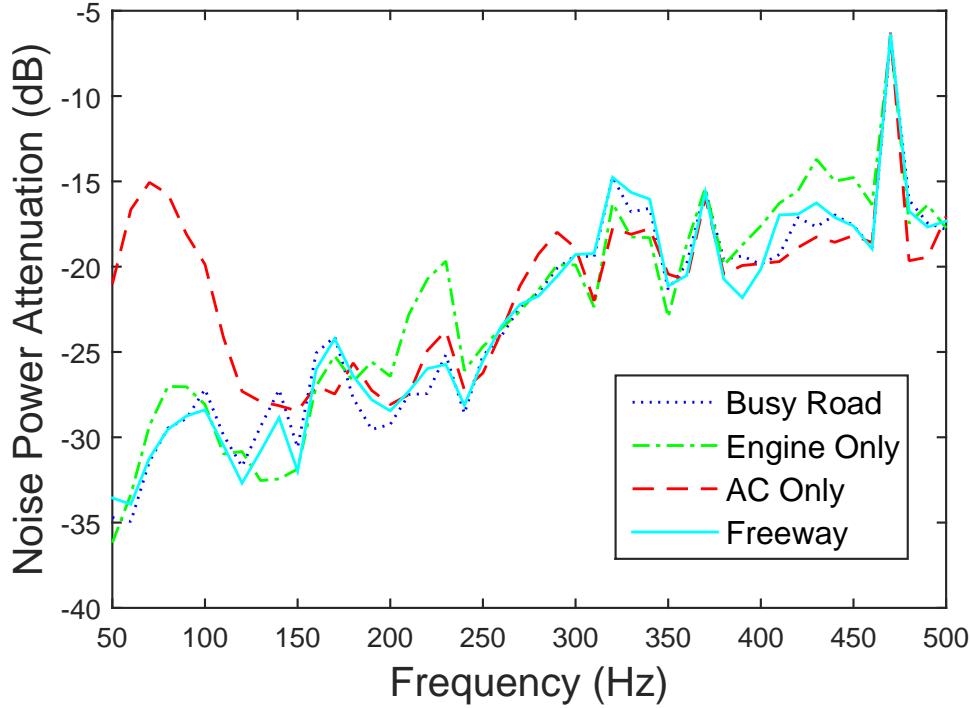


Figure 6.10: Noise power spectrum attenuation for 4 different driving conditions.

bin. Therefore we expect that the lower frequency noise fields can be seen as sparse, thus controlling such sound fields may require only a small number of well-placed loudspeakers which can nicely reproduce the dominant noise pattern.

Figure 6.10 plots the noise power attenuation for four different driving conditions, with the values calculated using (6.32). In addition to the freeway recording and the engine noise recording, the “Busy Road” recording was taken while driving on a 3-lane road at moderate speed with multiple vehicles passing by; while the “AC only” recording was taken with the car parked in a quiet place and engine idle, the air conditioning turned to maximum.

Figure 6.10 indicates that for most cases, the noise cancelling performance is relatively consistent, with the attenuation reducing gradually from 30 – 35 dB at 50 Hz to 15 – 20 dB at 500 Hz. This frequency-dependent performance is expected since the noise field is expected to be more complicated and harder to reproduce/cancel when the wavelength is shorter. We also notice that the noise field due to air conditioning is particularly difficult to cancel at 50 – 100 Hz, compared to other scenarios. We expect this is because the noise field due to AC is less similar to that of the loudspeakers, compared to other noise sources. One may also notice the

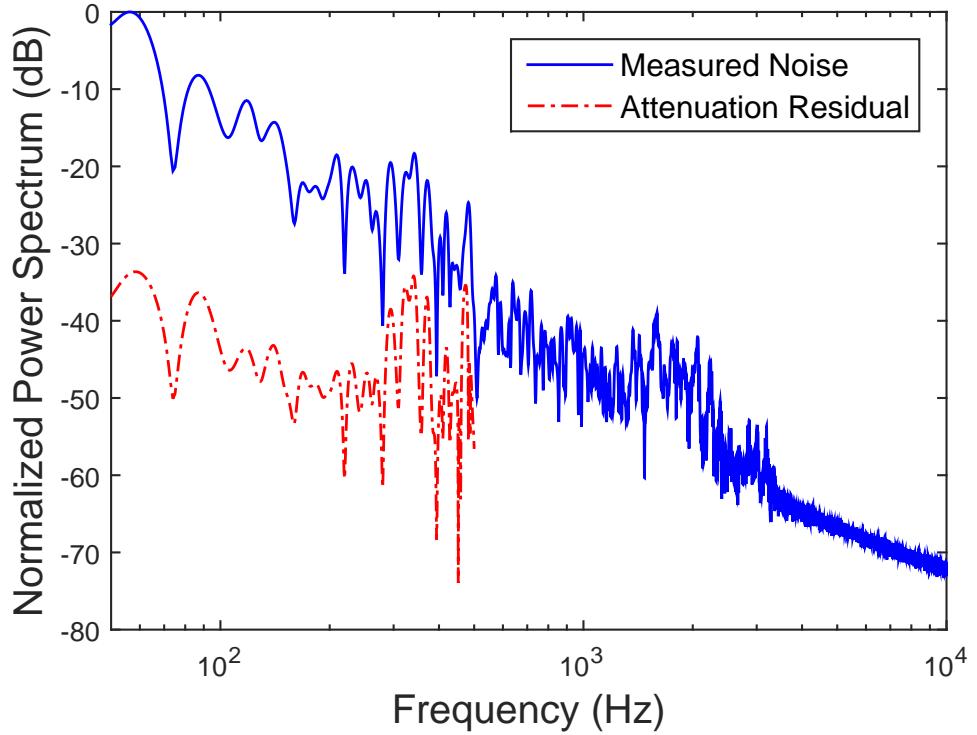


Figure 6.11: Comparison of average noise field power spectrum before and after cancellation.

common peak in all cases at 470 Hz, clearly at this frequency, the loudspeakers are unable to reproduce the noise fields very well.

We also include Fig. 6.11 which depicts the overall noise spectrum without attenuation, and the expected residual noise spectrum if the in-car loudspeakers are employed to cancel the noise field. The original noise spectrum is recorded while driving at 70 km/h with air conditioning at minimum. The attenuation is cut off at 500 Hz. We can see from the figure that the most dominant noise frequencies can be effectively cancelled by the integrated loudspeakers, resulting in a much quieter sound field within the region of interest.

In general, we can conclude that the integrated loudspeakers are capable of cancelling the noise field within our defined region of interest at the front passenger seat. However, we would expect the performance to degrade should the noise cancellation be carried out for multiple seats. Nevertheless, a proper in-car ANC system would be able to drive the four loudspeakers separately, which provides extra degrees of freedom for the loudspeaker channels, thereby promoting the overall performance of the system.

### 6.4.6 Experiment with multiple passenger seats and limited loudspeaker output power

The theory developed in Section 6.4.3 can be easily extended to multi-zone case. Assuming that a number of adjacent regions are defined inside the car cabin, Then, considering one of the control regions  $S_j$ , we can use the spherical harmonics decomposition (2.1) to decompose the noise field  $P_n(\mathbf{x}), x \in S_j$  as well as the basis patterns  $P_i(\mathbf{x}), x \in S_j$ , we can then express the noise field coefficients belonging to the  $j$ th control region  $C_{lm}^j$  using the corresponding coefficient  $C_{lm}^{j,i}$  of every basis pattern,

$$C_{lm}^j = \sum_i g_i C_{lm}^{j,i}. \quad (6.34)$$

We have shown that the average energy of a noise field is related to the spherical harmonic coefficients that represent the noise field by  $W_l$ , substituting  $c_{lm} = C_{lm} W_l$  into (6.34), we have

$$c_{lm}^j = \sum_i g_i c_{lm}^{j,i}. \quad (6.35)$$

Since we are considering the overall noise field over all of the control regions, it is convenient to write the coefficients of all regions in vector form, such that  $\mathbf{c} = [c_{00}^1, c_{11}^1 \dots c_{00}^2 \dots c_{LL}^j]^T$ , and  $\mathbf{c}_i = [c_{00}^{1,i}, c_{11}^{1,i} \dots c_{11}^{2,i} \dots c_{LL}^{j,i}]^T$ . Then from (6.35) and combining the coefficient of all control regions we have the vector representation

$$\mathbf{c} = \sum_i g_i \mathbf{c}_i. \quad (6.36)$$

A limitation of the mode matching method for deriving loudspeaker driving signals is that the amplitude of the loudspeaker driving signal is unbounded. Although a regularization can be added to the matrix inversion in (6.30) to avoid extremely high driving signals, there is no strict upperbound to the loudspeaker output power. From a practical point of view, driving a loudspeaker beyond its linear operating range would result in harmonic distortions, which introduces additional noise in the control regions. In order to avoid this problem, we define the optimization problem

$$\min f(\mathbf{D}) = \|\mathbf{c} + \mathcal{T}\mathbf{D}\|, \text{ subject to } |D_i| \leq K, i = 1, 2 \dots \quad (6.37)$$

where  $D_i$  are the elements of  $\mathbf{D}$  and represent the driving signal for the  $i$ th loudspeaker,  $\|\cdot\|$  denotes  $\ell^2$ -norm,  $K$  is a constant which sets the volume upper bound

for each loudspeaker. The noise energy attenuation can be represented as

$$A = \frac{\int_{S_1, S_2..} |P_r(\mathbf{x})|^2 d\mathbf{x}}{\int_{S_1, S_2..} |P_r(\mathbf{x})|^2 d\mathbf{x}} = \frac{\|\mathbf{c} + \mathcal{T}\mathbf{D}\|^2}{\|\mathbf{c}\|^2} \quad (6.38)$$

where  $\mathbf{D}$  is the solution to (6.37).

The loudspeaker channel matrix can be expressed in a similar manner, as

$$\mathbf{H} = \begin{bmatrix} H_{00}^{1,1} & H_{00}^{1,2} & H_{00}^{1,3} & \dots \\ H_{11}^{1,1} & H_{11}^{1,2} & H_{11}^{1,3} & \dots \\ \vdots & \vdots & \vdots & \dots \\ H_{LL}^{1,1} & H_{LL}^{1,2} & H_{LL}^{1,3} & \dots \\ H_{00}^{2,1} & H_{00}^{2,2} & H_{00}^{2,3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (6.39)$$

where  $H_{lm}^{j,v}$  being the spherical harmonic coefficient of order  $l$  and mode  $m$ , associated with the  $j$ th control region, due to the  $v$ th loudspeaker playing a unit signal.

## Experiment Setup

In this experiment, we aim to investigate the noise field complexity within a 2005 Ford Falcon XR6 sedan, under various driving conditions; as well as examine the noise cancelling potential of the multimedia loudspeakers installed in the car. The regions of interest are chosen to be spherical regions located at the head position of each of the four seats, the radius of each region is set to 10 cm, which covers the size of a human head.

For this experiment, we focus on the noise below 200 Hz. Using (6.19), we can calculate the relative contribution of each spherical harmonic mode towards the total noise energy within the control regions, at  $f = 200$  Hz we have

$$\frac{\int_S |P_{00}(\mathbf{x})|^2 dS}{\int_S |P(\mathbf{x})|^2 dS} = \frac{|\alpha_{00} W_0|^2}{\sum_{l,m} |\alpha_{lm} W_l|^2} \approx 0.972 \quad (6.40)$$

thus the 0th order spherical harmonic accounts for the vast majority of the noise energy within the control regions, for frequencies below 200Hz, the contribution of the 0th mode is even higher (99.3% at 100 Hz). Therefore, in our experiments, we only monitor the 0th order spherical harmonic for each control region, which can be done by placing a single omni-direction microphone at the center of each

region. We note that we measure only the 0th mode spherical harmonic because at low frequencies, the 0th mode contributes to the majority of the noise energy, not because we believe the noise field is isotropic. For noise field analysis of large region and higher frequencies, higher-order microphones are required, such as the Eigenmike.

The recording system we use consists of four AKG CK92 omnidirectional condenser microphones, connected to a TubeFire 8 audio interface via four AKG SE300B microphone pre-amps. The synchronous audio streams are recorded using a Macbook, which is connected to the TubeFire 8 via firewire.

We record the noise field at the four control regions simultaneously for various driving conditions, including the pure engine noise recording, where the car is parked in a relatively quiet place and the engine ran at 2000 rpm. For each driving condition, we record the noise for 10 seconds. The recording is then split into 100 frames and transformed into spherical harmonic coefficients  $\alpha_{00}^j(k)$  at different frequency bins for further analysis.

The Ford sedan has four full-band loudspeakers installed, two of which are integrated at the bottom of either of the front doors, while the other two are placed behind each rear seat. However, the car's audio playback system only supports stereo signals, which means the two loudspeakers on the left side simultaneously play the left channel of the stereo signal, and the same goes for the right channel.

We obtain the loudspeaker channel matrix by measuring the impulse response at the region of interest due to the left channel and right channel separately, and then calculating the corresponding sound field coefficients for each frequency bin, in the same way as we obtain the noise field measurements. The channel matrix takes the form of (6.39). The 0th order sound fields at 4 regions and the stereo speaker system result in a 4-by-2 channel matrix for each frequency bin.

In order to estimate the noise cancellation capability of the in-car loudspeakers in each driving condition, we solve (6.37) for each of the 100 snapshots in every recording, and calculate the expected residual noise energy for each snapshot. The value of  $K$  is chosen such that the sound energy at the regions of interest due to each loudspeaker is no more 3 times more than that due to the noise. We then calculate the average noise energy attenuation using

$$A = \frac{\sum_{n=1}^{100} \|\mathbf{c}_n + \mathcal{T}\mathbf{D}_n\|^2}{\sum_{n=1}^{100} \|\mathbf{c}_n\|^2}, \quad (6.41)$$

where  $\mathbf{c}_n$  and  $\mathbf{D}_n$  are the weighted coefficient vectors and the optimal driving signals

Table 6.3: Noise field eigenvalues for freeway driving condition and pure engine noise

100 km/h	40 Hz	80 Hz	120 Hz	160 Hz	200 Hz
$\lambda_1$	1.000	1.000	1.000	1.000	1.000
$\lambda_2$	0.292	0.282	0.498	0.476	0.292
$\lambda_3$	0.062	0.207	0.181	0.372	0.102
$\lambda_4$	0.007	0.139	0.049	0.092	0.053
Engine Only	40 Hz	80 Hz	120 Hz	160 Hz	200 Hz
$\lambda_1$	1.000	1.000	1.000	1.000	1.000
$\lambda_2$	0.033	0.315	0.108	0.293	0.042
$\lambda_3$	0.005	0.095	0.018	0.106	0.031
$\lambda_4$	0.000	0.018	0.003	0.045	0.015

for the snapshots in each recording, respectively.

## Data Analysis

We first investigate the dimensionality of the combined noise field over the four control regions by observing the eigenvalues of the estimated covariance matrix of the spherical harmonic coefficients. We normalize the eigenvalues and sort them from the largest to the smallest, the results for pure engine noise and the noise when driving at 100 km/h are shown in Table 6.3. We can see from Table 6.3 that the eigenvalues of the engine noise are almost always smaller than the corresponding eigenvalues of the freeway driving condition (100 km/h). In the case of engine noise, the fourth eigenvalue is in the order of 0.01 for most frequencies, therefore the noise field may be modelled using 3 noise modes in (6.25), without significant loss of accuracy. As a result, in order to effectively cancel the engine noise over the four control regions simultaneously, a minimum of 3 loudspeakers would be sufficient, assuming that the loudspeaker channels have sufficient diversity.

On the other hand, the noise field of the freeway driving condition is more complicated, the fourth eigenvalues are above 0.01 for all frequencies above 40 Hz. Therefore at least four independent loudspeakers are required to effectively cancel the noise within the control regions simultaneously.

Since the car's loudspeakers can only play stereo signals, and that the combined noise fields require no less than 4 independent loudspeaker channels to effectively control, we do not expect a high noise energy attenuation over 3 or 4 seats. However, we expect the loudspeakers to simultaneously cancel the noise over two control regions with good results. In order to validate our expectations, we use (6.41) to calculate the expected noise attenuation for simultaneous noise cancellation for 2, 3 and 4 seats, the results are shown in Figs. 6.12-6.15. The noise cancellation

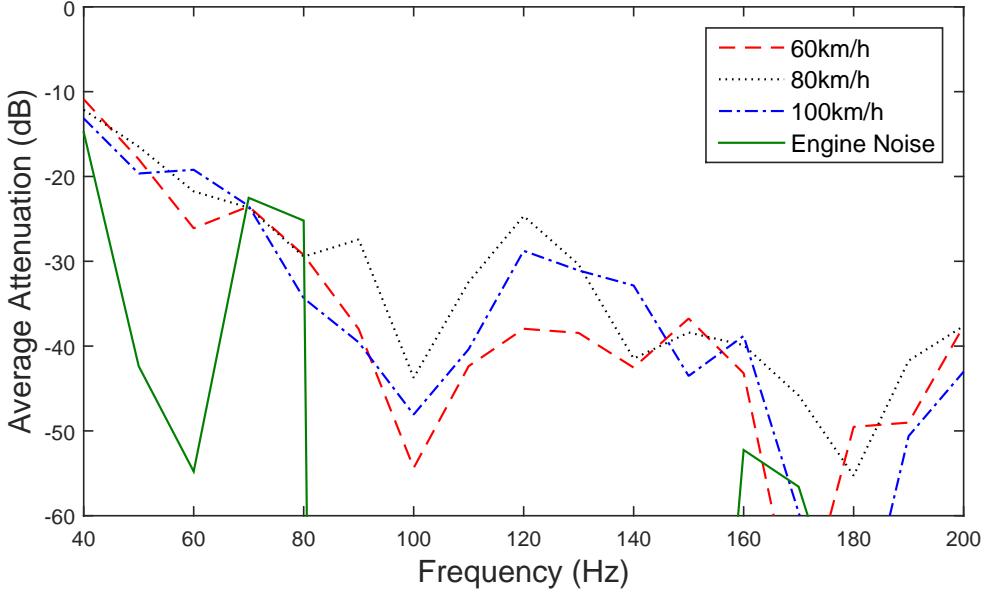


Figure 6.12: Expected noise power attenuation after noise cancellation in the two front seats only.

performance for the two front seats only is shown in Fig. 6.12. The attenuations are calculated for frequencies from 40 Hz to 200 Hz, and for driving speeds at 60 km/h, 80 km/h and 100 km/h. The attenuation for the engine noise is also included in the figure. We can see from Fig. 6.12 that the attenuations for all three driving speeds are very similar. The residual noise level is highest at 40 Hz, and gradually reduces to around -40 dB for all three driving speeds. The engine noise, on the other hand, can be effectively cancelled at most frequency bins. We believe this is because of the low dimensionality of the engine noise field, as is shown in Table 6.3.

Since we are only considering the 0th order coefficients in our calculations, while ignoring the other coefficients which contribute to approximately 1 percent of total noise energy, the upper bound of actual achievable attenuation would be around 20 dB, depending on the loudspeakers' ability to attenuate the higher order coefficients.

Figure 6.13 shows the results for simultaneous noise control for the two right side seats. A trend similar to that in Fig. 6.12 can be observed. We believe that the reason for the increasing attenuation over frequency is due to the impact of wavelength on loudspeaker channels, where at low frequency, the sound pressure at two different seats due to one particular loudspeaker is very similar. Therefore the loudspeaker channel matrix is highly coupled at low frequencies, resulting in less

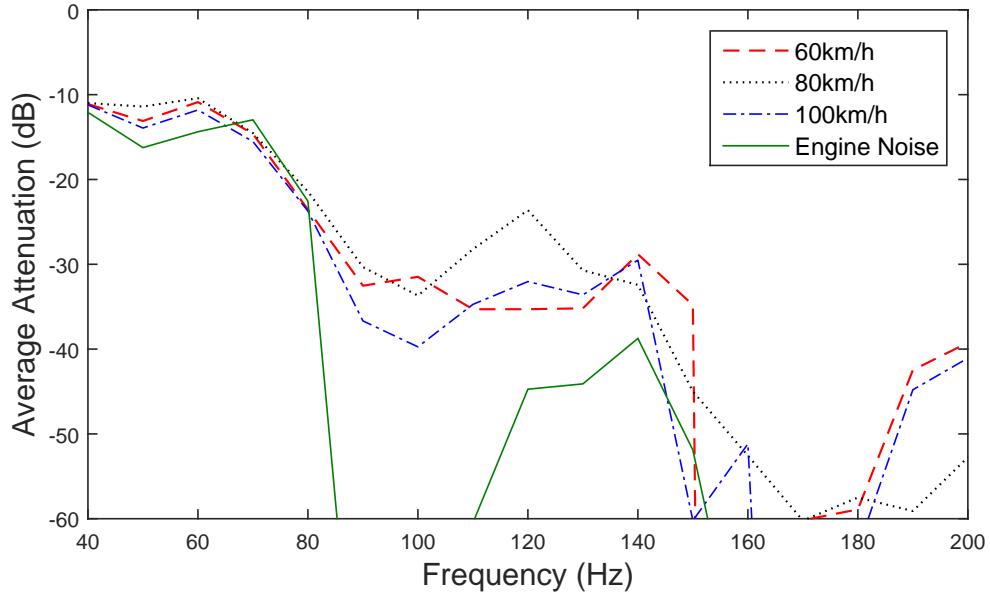


Figure 6.13: Expected noise power attenuation after noise cancellation in the two right side seats only.

noise attenuation under the same output power constraint. Figure 6.14 illustrates the expected ANC performance for simultaneous 3-seat noise control (two front seats and left passenger seat). As expected, the noise energy reduction is significantly worse than the two-seat cases, with around 10 dB reduction across all frequency bins of interest. We also notice that the engine noise is no longer easier to cancel than the other noise fields apart from a few frequency bands (40-60 Hz). This is consistent with Table 6.3, where the third and fourth eigenvalues of engine noise at 40 Hz are very small, indicating a sparse noise field with 2 degrees of freedom, therefore the noise field can be controlled by a stereo system. We also include Fig. 6.15 which depicts the four-seat ANC performance. Compared to Fig. 6.14, the attenuation is even smaller at around 6-7 dB. However, the ANC performance is once again consistent over different driving speeds. From this observation, we estimate that the noise field at different driving speeds are similar, and that a loudspeaker array's capability of controlling in-car noise does not vary greatly at different driving speeds.

The attenuation of the engine noise is often lower than that of the noise fields under various driving conditions. However, from our subjective tests, the majority of the noise in the car cabin came from the tires and suspension, the engine noise only plays a small part in the overall perceived noise. Therefore, it is understandable

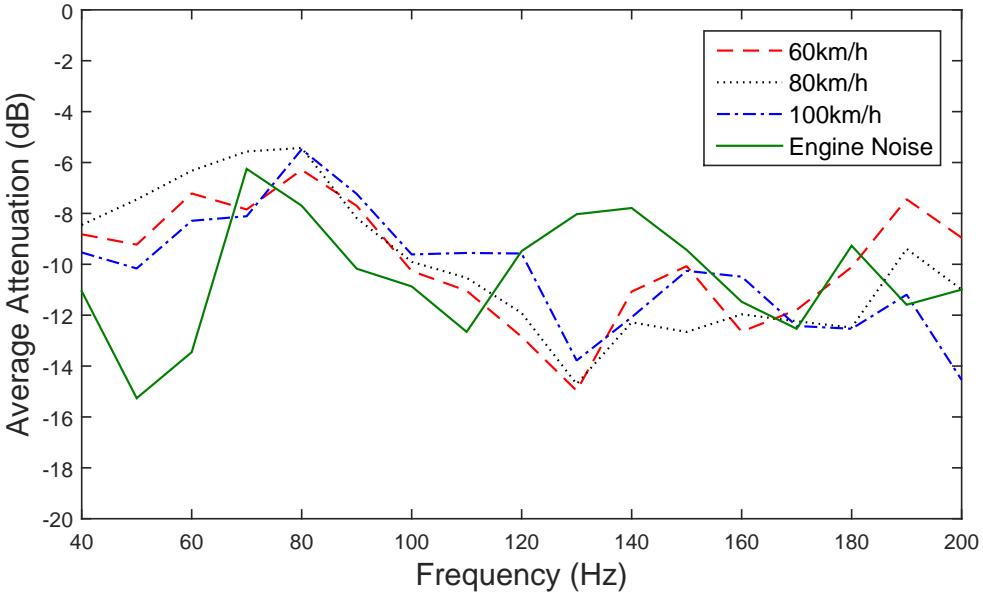


Figure 6.14: Expected noise power attenuation after noise cancellation in the two front seats and the left passenger seat.

that the overall noise reduction is different from the engine noise suppression under the same conditions.

We would like to point out that the analysis of multiple-seat ANC performance is limited to 200 Hz and 0th order only because of the limitations in the hardware setup, more specifically, the lack of synchronized higher order microphones. In order to obtain the analysis results for higher frequency and/or larger control regions, it is necessary to replace the omni-directional microphones that are used in this experiment with suitable higher order microphones, so that sound field components of higher orders can be captured.

#### 6.4.7 Observations and insights

In general, we can conclude that the integrated loudspeakers, when used as a stereo system, are capable cancelling the in-car noise field at the head position of a single seat for frequencies up to 500 Hz, or simultaneously cancelling the noise fields at two seats for frequencies up to 200 Hz. It can be seen that when the number of control regions exceed the number of loudspeaker channels, the ANC performance reduces significantly. In order to control the noise over more regions or at higher frequencies, additional independent loudspeakers are required. We expect the multi-zone ANC performance of the four integrated loudspeakers to improve significantly,

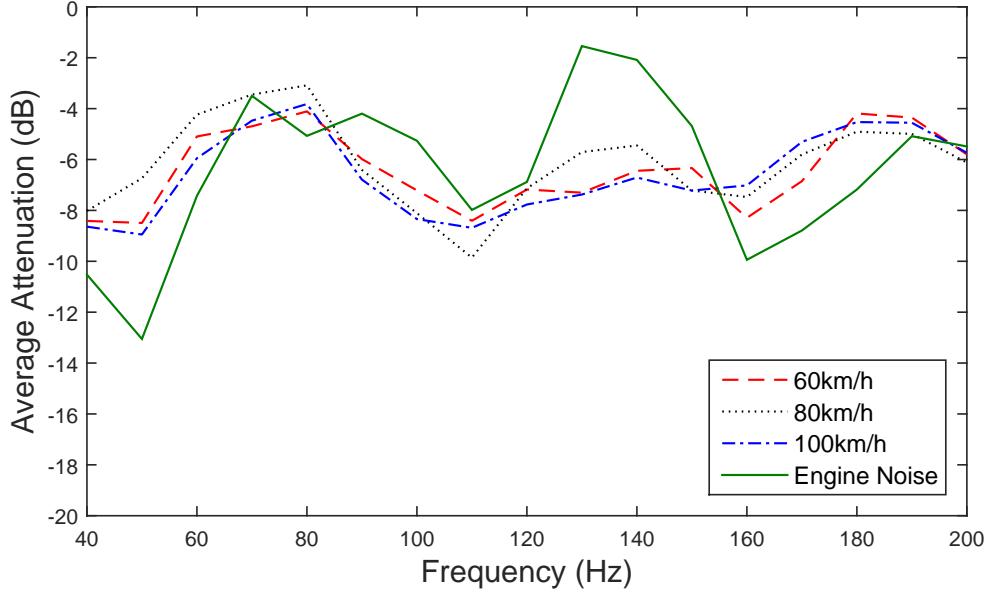


Figure 6.15: Expected noise power attenuation after noise cancellation in all four seats.

if they could be driven separately.

## 6.5 Summary

In this chapter, we proposed one method to enhance the sound field reproduction quality over a large region by prioritizing the reproduction at some smaller, sub-zones. This method improves the sound field reproduction accuracy at the smaller sub-zones at the cost of slight worse overall reproduction accuracy. This method is especially useful when there is insufficient number of loudspeakers available for sound field reproduction.

We also proposed a new metric for measuring average noise level within a region. It is shown that this metric is more robust and accurate than the commonly used method where the noise level is determined by averaging the noise pressure measured by microphones.

This metric is then utilized to develop a method to estimate the potential performance of a spatial ANC system. We use this method to evaluate the in-car loudspeakers' capability of cancelling the in-car noise at the passengers' head positions, it was shown that the loudspeakers have the capability to attenuate the noise level at lower frequencies for the given region of interest.

## 6.6 Related Publications

This chapter's work has been published in the following conference proceedings [115] [123] [124] [125]:

- H. Chen, P. Samarasinghe, T. D. Abhayapala, and W. Zhang, “Spatial noise cancellation inside cars: Performance analysis and experimental results,” in *Proc. 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.
- H. Chen, P. Samarasinghe, and T. D. Abhayapala, “In-car noise field analysis and multi-zone noise cancellation quality estimation,” in *Proc. 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2015, pp. 773–778.
- H. Chen, J. Zhang, P. N. Samarasinghe, and T. D. Abhayapala, “Evaluation of spatial active noise cancellation performance using spherical harmonic analysis,” in *Proc. 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2016, pp. 1–5.
- H. Chen, T. D. Abhayapala, and W. Zhang, “Enhanced sound field reproduction within prioritized control region,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 2014*, vol. 249, no. 3, pp. 4055–4064, Nov 2014.

# Chapter 7

## Spatial active noise cancellation system architectures

**Overview:** The adaptive filtering algorithm is a critical component of an ANC system. For spatial ANC, we believe that incorporating spatial signal processing techniques into the adaptive filtering algorithm would yield improved performance. In this chapter, we propose an adaptive filtering algorithm developed for spatial ANC applications. This algorithm is based on the existing multi-channel feed-forward adaptive algorithm, but with the input signals transformed into spherical harmonic domain before entering the adaptive filtering process. Both frequency domain and time domain architectures are proposed for the adaptive algorithm. We also implement an experimental spatial ANC system using the proposed time domain adaptive algorithm, and investigate the impact of loudspeaker number and placement on the performance of the spatial ANC system.

### 7.1 Introduction

Active noise cancellation (ANC) over space has been a hot topic of research in the last two decades. Typically, ANC systems targeting spatial noise reduction are realized by Multi-Input Multi-Output (MIMO) systems [126] employing a feed-forward or feedback control algorithm [6]. Most popular applications of such systems include aircraft cabin noise reduction [127] and automobile noise reduction [7, 128].

The most widely used MIMO ANC systems employ a number of microphones, placed within the spatial area where noise attenuation is desired. The algorithms are designed to minimize the average noise level captured by the error microphones

in the least-mean-square sense [6], through playing back counter-noise signals from a number of loudspeakers. Using these algorithms, the noise attenuation is only maximized at the positions of the error microphones, while the overall noise attenuation quality within the region of interest cannot be guaranteed.

A spatial ANC algorithm utilizing circular harmonics analysis has proposed in [31]. In this work, both the noise field and the secondary loudspeaker driving signals are converted into the circular harmonics domain, and it was shown that this method reduces computational complexity of massive multichannel ANC systems [31]. However, one drawback of this method is that the method requires circular arrangement of the secondary loudspeaker array, and that the algorithm only performs well in 2D sound fields. In [30] a feedback spatial ANC algorithm was proposed, and it was shown that this algorithm has a faster convergence speed than existing MIMO algorithms. Both of these algorithms were only validated in simulations.

In this chapter, we propose an improved spatial feed-forward ANC algorithm. It is formulated in terms of the spherical harmonic decomposition of the noise field and requires the measurements from a secondary microphone array distributed over the boundary of a spherical region of interest. After converting the noise signal into the spherical harmonic domain, an additional weighting is applied to each spherical harmonic coefficient, which are then used to update an adaptive filter through a LMS adaptive algorithm. We show that through the spherical harmonic transform and the weighting process, the ANC algorithm is able to maximize the average noise attenuation over the entire region.

We develop the algorithm in the frequency domain, and since most ANC algorithms are implemented in the time domain, we also present the time domain equivalent of the algorithm, which can be realized through time domain filtering of microphone signals. In order to validate the the proposed algorithm, a prototype spatial ANC system is built inside our laboratory. The system is used to investigate the spatial ANC performance of the proposed algorithm under various hardware configurations.

## 7.2 Background theory

### 7.2.1 Time domain multi-channel feed-forward ANC architecture

The time domain FxLMS algorithm, which is the most popular adaptation algorithm in the ANC application is briefly described here.

The system diagram of the multi-channel Filtered-X LMS algorithm is shown in Fig. 7.1. The wide arrows represent a vector of signals (acoustic or electrical).  $\mathbf{H}(z)$  represents the primary channel.  $\mathbf{S}(z)$  and  $\widehat{\mathbf{S}}(z)$  represent the secondary channel from the loudspeakers to the error microphones and the their estimations, respectively.

Assume  $n$  is the time index, and the system consists of  $U$  reference signals,  $V$  secondary loudspeakers and  $Z$  error microphones, the reference input signals are  $x_u(n)$ ,  $u = 1, \dots, U$  and the instantaneous error microphone measurements are  $e_z(n)$ ,  $z = 1, \dots, Z$ . The primary noise field and the sound field due to the secondary loudspeakers at the error microphone positions are represented by  $d_z(n)$  and  $y'_z(n)$ , respectively. The error function in our system can be written as

$$e_z(n) = d_z(n) + y'_z(n), \quad (7.1)$$

where

$$y'_z(n) = \sum_v y_v(n) * \mathbf{s}_{vz}, \quad \text{for } z = 1, 2, \dots, Z \quad (7.2)$$

$y_v(n)$  is the driving signal for the  $v$ th loudspeaker,  $\mathbf{s}_{vz}$  is the secondary channel between the  $v$ th loudspeaker and  $z$ th error microphone, and “ $*$ ” denotes linear convolution.

The secondary source driving signals in each iteration can be represented by

$$y_v(n) = \sum_u \mathbf{w}_{uv}^T(n) \mathbf{x}_u(n), \quad \text{for } v = 1, 2, \dots, V \quad (7.3)$$

where  $\mathbf{w}_{uv}(n) = [w_{uv,0}(n), w_{uv,1}(n), \dots, w_{uv,L-1}(n)]^T$  are the adaptive filter coefficients in the  $n$ th iteration,  $L$  is the length of the FIR adaptive filters.

The update equation of the multi-channel FxLMS algorithm is derived by

$$\mathbf{w}_{uv}(n+1) = \mathbf{w}_{uv}(n) - \mu \sum_u \mathbf{x}'_{uvz}(n) e_z(n), \quad \text{for } v = 1, 2, \dots, V, \text{ and } u = 1, \dots, U \quad (7.4)$$

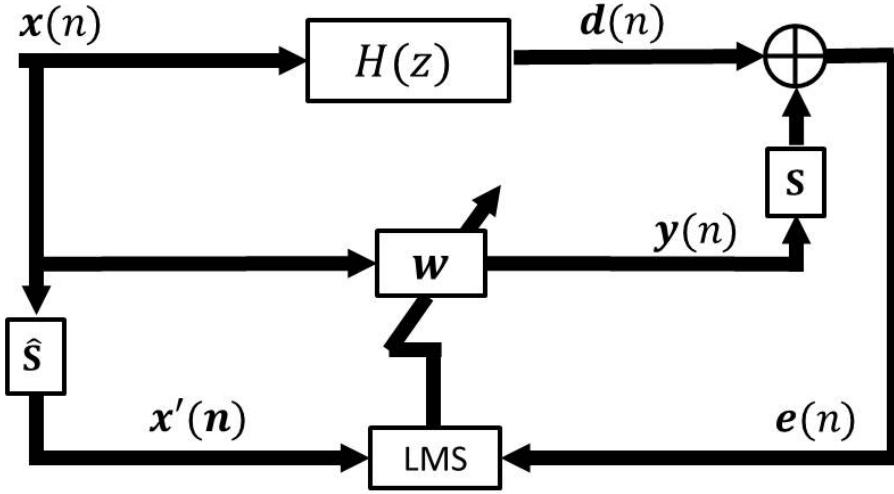


Figure 7.1: Block diagram of the time domain feedforward ANC system.

where  $\mu$  is the step size,  $\mathbf{x}'_{uvz}(n) = [x'_{uvz}(n), \dots, x'_{uvz}(n - \mathcal{L} + 1)]^T$  is the vector of the latest  $\mathcal{L}$  filtered reference signals, and the filtered reference signals can be obtained by

$$x'_{uvz}(n) = x_u(n) * \hat{\mathbf{s}}_{vz}. \quad (7.5)$$

Filtering the reference signal  $x_u(n)$  by the secondary channel estimation helps to improve the convergence speed of the adaptive algorithm, especially when the secondary path has a long delay [6].

## 7.2.2 Frequency domain feed-forward ANC architecture

The ANC algorithm can be efficiently implemented by computing the time-domain linear convolutions using FFT techniques, which results in a system with an equivalent system but with much lower computational cost. A single-channel frequency domain implementation has been introduced in [129]. The data flow is shown in Fig. 7.2.

In the frequency domain implementation proposed in [129], the data is processed frame by frame, rather than sample by sample as in the time domain architecture. Assuming the frame size to be  $N$ , the frequency domain reference signal is calculated by taking the  $2N$ -point FFT of two consecutive input data frames, i.e.

$$\mathbf{X}(j+1) = \mathcal{F}_{2N} \begin{bmatrix} \mathbf{x}(j) \\ \mathbf{x}(j+1) \end{bmatrix}, \quad (7.6)$$

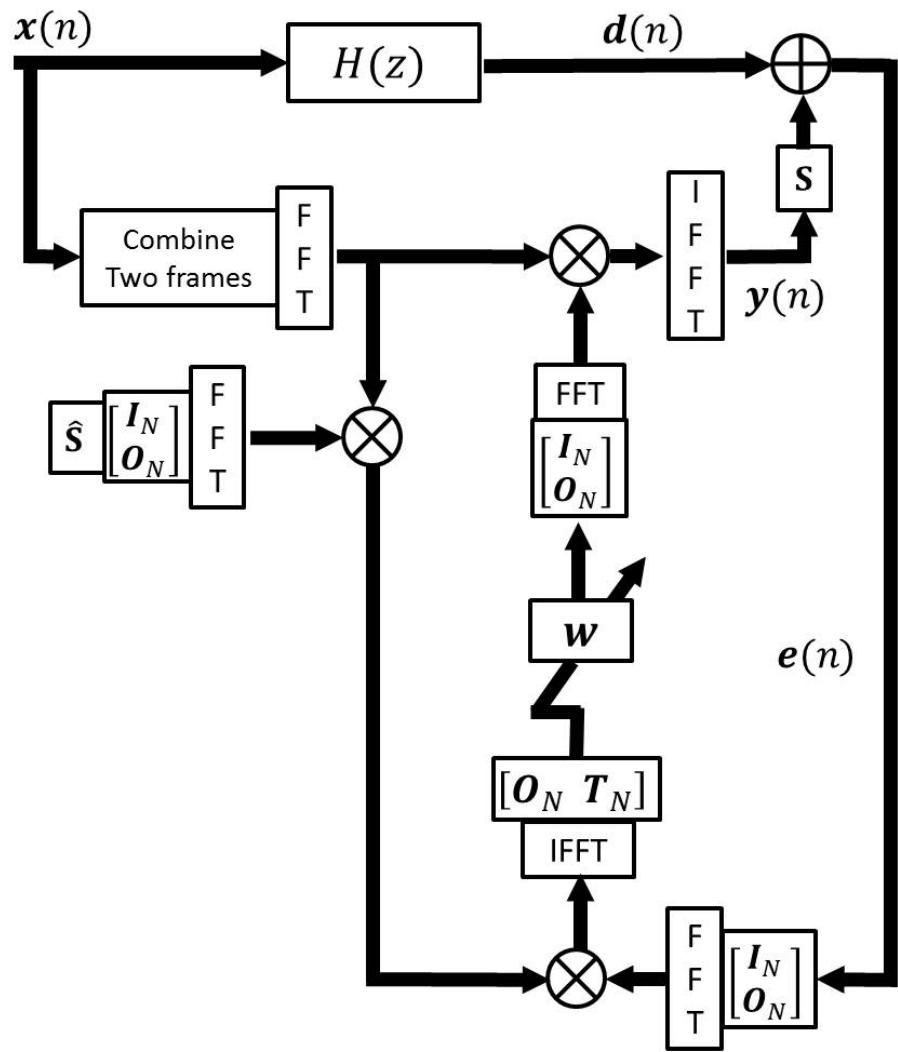


Figure 7.2: Block diagram of the Frequency Domain feedforward ANC system.

where  $\mathcal{F}_{2N}$  denotes  $2N$ -point FFT,  $j$  is frame index,  $\mathbf{x}(j)$  and  $\mathbf{x}(j + 1)$  denote the previous and current frame of input data, respectively.

The cancelling signal is generated by convolving the reference signal and the filter  $\mathbf{w}(\mathbf{k})$  in the frequency domain, and discarding the first  $N$  samples of the IFFT output, which can be expressed as

$$\mathbf{y}(j + 1) = [\mathbf{O}_N \ \mathbf{I}_N] \mathcal{F}_{2N}^{-1} [\mathbf{X}(j + 1) \otimes \mathbf{W}(j)], \quad (7.7)$$

where “ $\otimes$ ” denotes pair-wise multiplication of two vectors or matrices,  $\mathbf{y}(k + N)$  is the latest frame of secondary source driving signal,  $\mathbf{O}_N$  and  $\mathbf{I}_N$  denote  $N$ -by- $N$  zero matrix and identity matrix, respectively.  $\mathbf{W}(j)$  is the  $2N$  point FFT of the time domain adaptive filter  $\mathbf{w}(j)$  with zero padding, expressed as

$$\mathbf{W}(j) = \mathcal{F}_{2N} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{O}_N \end{bmatrix} \mathbf{w}(j). \quad (7.8)$$

The filtering of the reference signal with the secondary path estimation is implemented in a similar manner. Denote the estimated secondary path impulse response of length  $N$  as  $\hat{\mathbf{s}}$ , the frequency domain filtered reference signal  $\mathbf{X}'(j + 1)$  can be calculated as

$$\mathbf{X}'(j + 1) = \mathcal{F}_{2N} \begin{bmatrix} \mathbf{x}(j) \\ \mathbf{x}(j + 1) \end{bmatrix} \otimes \hat{\mathbf{S}}, \quad (7.9)$$

where  $\hat{\mathbf{S}}$  is the  $2N$  point FFT of  $\hat{\mathbf{s}}$ ,

$$\hat{\mathbf{S}} = \mathcal{F}_{2N} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{O}_N \end{bmatrix} \hat{\mathbf{s}}. \quad (7.10)$$

The adaptive filter  $\mathbf{w}(k)$  is implemented in the time domain, in the form of a vector of length  $N$ . The filter is updated when a new frame of reference signal and error signal is available, and the update equation can be written as

$$\mathbf{w}(j + 1) = \mathbf{w}(j) + 2\mu \mathbf{c}(j + 1) \quad (7.11)$$

where  $\mu$  represents step size, and

$$\mathbf{c}(j + 1) = [\mathbf{O}_N \ \mathbf{T}_N] \mathcal{F}_{2N}^{-1} [\mathbf{X}'(j + 1) \otimes \mathbf{E}(j + 1)], \quad (7.12)$$

with  $\mathbf{T}_N$  being a time reversal matrix, with its secondary diagonal equal to 1, and

other entries equal to 0. The frequency domain error signal  $\mathbf{E}(j + 1)$  is obtained by taking the  $2N$  point FFT of the latest frame of the time domain error signal, expressed as

$$\mathbf{E}(j + 1) = \mathcal{F}_{2N} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{O}_N \end{bmatrix} \mathbf{e}(j + 1). \quad (7.13)$$

The algorithm is guaranteed to converge if [129]

$$0 \leq \mu \leq \frac{1}{N\lambda_{max}}, \quad (7.14)$$

where  $\lambda_{max}$  is the maximum eigenvalue of the autocorrelation matrix of the input signal, and

$$E[\mathbf{X}'^T(k)\mathbf{X}'(j)] = 0, k \neq j. \quad (7.15)$$

The single channel frequency domain implementation can be easily extended to a multi-channel system, with  $U$  reference sources,  $V$  secondary sources and  $Z$  error microphones. In this case, the size of the data matrices would be  $2N \times U$  for  $\mathbf{X}(k)$ ,  $2N \times U \times V \times Z$  for  $\mathbf{X}'(k)$ ,  $2N \times Z$  for  $\mathbf{E}(k)$ , and  $N \times U \times V$  for  $\mathbf{w}(k)$ . The adaptive filter  $\mathbf{w}(k)$  needs to be updated for each  $u$  and  $v$ , i.e.,

$$\mathbf{w}_{uv}(j + 1) = \mathbf{w}_{uv}(j) + \sum_{z=1}^Z 2\mu \mathbf{c}_{uvz}(j + 1), \quad (7.16)$$

where

$$\mathbf{c}_{uvz}(j + 1) = [\mathbf{O}_N \ \mathbf{T}_N] \mathcal{F}_{2N}^{-1} [\mathbf{X}'_{uvz}(j + 1) \otimes \mathbf{E}_z(j + 1)], \quad (7.17)$$

and the driving signals for each loudspeaker is generated as

$$\mathbf{y}_v(j + 1) = \sum_{u=1}^U [\mathbf{O}_N \ \mathbf{I}_N] \mathcal{F}_{2N}^{-1} [\mathbf{X}_u(j + 1) \otimes W_{uv}(j)]. \quad (7.18)$$

It can be seen that the computational complexity grows quickly as the number of reference signals, loudspeakers and error microphones increase.

## 7.3 Frequency domain feed-forward architecture for spatial ANC systems

### 7.3.1 Existing spatial ANC system based on circular harmonic transform

Spatial ANC systems using circular harmonic transform have been proposed in [31]. The overall data flow in this method is similar to the multi-channel ANC algorithm. However, in this method, the error microphones form a circular array, which is surrounded by the secondary loudspeaker array, also taking a circular geometry. The reference microphones surround the loudspeaker array, and form a third circular array. The extensive use of circular array geometries allows for transforming both the reference and error signals into circular harmonic coefficients; in addition, the secondary loudspeaker channels  $\mathbf{H}_v(k)$  are also transformed into circular harmonic domain, under the assumption that the loudspeakers are point sources.

In this method, the filtered reference signals are generated by filtering the reference circular harmonic coefficients through the secondary channel circular harmonic coefficients, therefore  $\mathbf{X}'(k)$  is also in the circular harmonic domain. Furthermore, since the error signals are also transformed into circular harmonics, the adaptive filter, which takes the same form as in the MIMO adaptive algorithm, operates in the circular harmonic domain, and updates the adaptive filter  $\mathbf{W}(k)$  which contains circular harmonic coefficients that mimics the primary channel.

The loudspeaker driving signals are first generated by filtering the reference coefficients through the adaptive filter, also in the form of circular harmonic coefficients. Then, an inverse circular harmonic transform maps the coefficients to each individual loudspeaker, and produces the final output signal for each speaker.

Due to the use of circular harmonic transform, this method is able attenuate the noise within the 2D space covered by the error microphone array, while significantly reducing the computational complexity, compared to a multi-channel algorithm using the same loudspeaker-microphone setup [31]. However, one disadvantage of this method is that in order to express the secondary sources using circular harmonics, the loudspeakers have to be arranged as a circular array.

### 7.3.2 Proposed spatial ANC system based on spherical harmonic transform

The adaptive algorithm proposed in [31] can be extended to 3D space by replacing the circular harmonic transforms with spherical harmonic transforms. However, this would require a hardware setup consisting of three concentric spherical arrays surrounding the quiet zone, which makes deployment difficult, especially for real-life applications.

In this section, we propose an alternative adaptive algorithm based on spherical harmonic transform. In order to maintain the flexibility of the MIMO adaptive algorithm, we only require that the error microphone array has the capability of capturing 3D spatial sound field in the form of spherical harmonics, but not require any specific geometry for the reference microphones and secondary loudspeakers.

Due to not using spherical reference microphone array, we do not perform spherical transform for the reference signal. However, both the error signal and the secondary loudspeaker channels can be transformed into spherical harmonic domain. Consequently, the filtered reference signal  $\mathbf{X}'(k)$  can also expressed as spherical harmonic coefficients. This allows the adaptive algorithm to operate in the spherical harmonic domain, since all of its inputs are spherical harmonic coefficients. The inverse transform to derive loudspeaker driving signals can be omitted, due to not using spherical loudspeaker array geometry.

Figure 7.3 illustrates the structure of the proposed frequency domain spatial ANC architecture. Compared to Fig. 7.2, it can be seen that the structure of the spatial ANC system is similar to that of the multi-channel system, except that the filtered reference signals and the error signals are transformed into spherical harmonic domain before being passed to the adaptive filter.

The spherical harmonics transform operates on each frequency bin of all the error microphone signals, the operation may be expressed in matrix form as

$$\boldsymbol{\beta}(k, j) = \mathcal{T}(k) \mathbf{E}(k, j), \quad (7.19)$$

where  $\boldsymbol{\beta}(k, j)$  is a vector of length  $(L + 1)^2$ , containing all the spherical harmonic coefficients for the error signals at frequency  $k$  and frame index  $j$ ,  $\mathbf{E}(k, j)$  is a vector of length  $Z$ , containing the signals of all error microphones at frequency  $k$  and frame index  $j$ .  $\mathcal{T}(k)$  is the transformation matrix specific for the frequency bin and microphone array geometry. For a uniform spherical error microphone array of

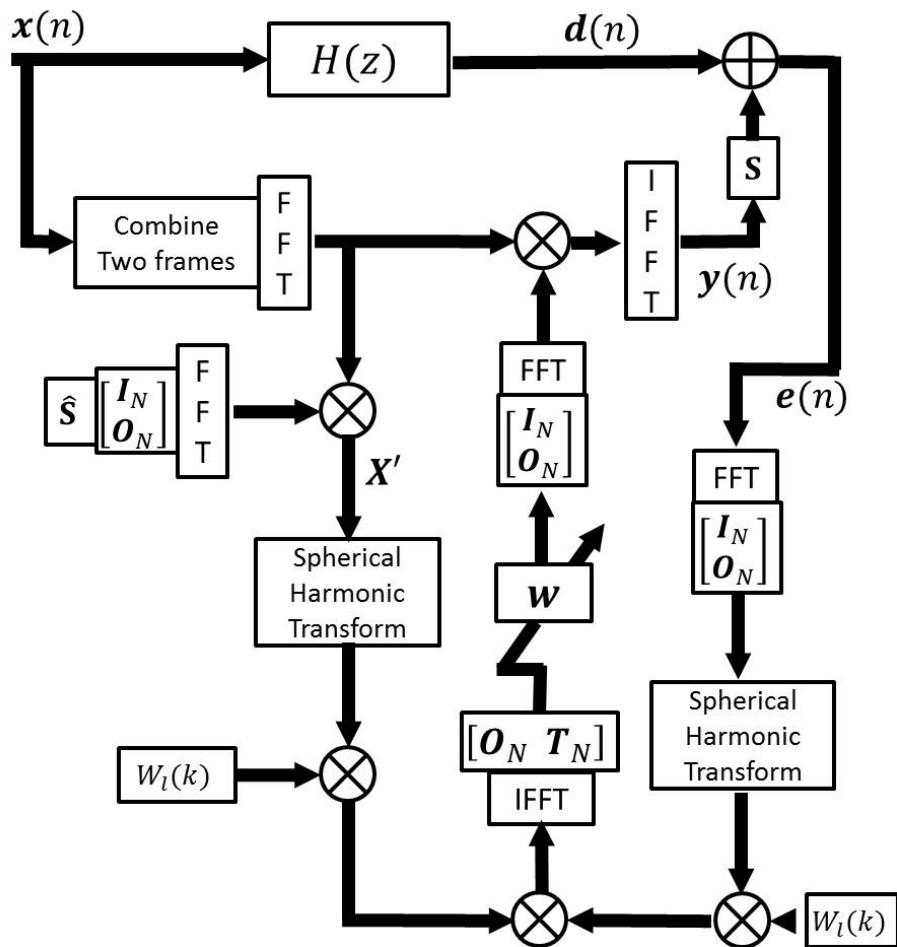


Figure 7.3: Block diagram of the frequency domain feedforward spatial ANC system.

radius  $R$ ,  $\mathcal{T}(k)$  takes the form

$$\mathcal{T}(k) = \begin{bmatrix} \frac{Y_{00}^*(\theta_1, \phi_1)}{j_0(kR)} & \frac{Y_{00}^*(\theta_2, \phi_2)}{j_0(kR)} & \cdots & \frac{Y_{00}^*(\theta_z, \phi_z)}{j_0(kR)} \\ \frac{Y_{11}^*(\theta_1, \phi_1)}{j_1(kR)} & \frac{Y_{11}^*(\theta_2, \phi_2)}{j_1(kR)} & \cdots & \frac{Y_{11}^*(\theta_z, \phi_z)}{j_1(kR)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{Y_{LL}^*(\theta_1, \phi_1)}{j_L(kR)} & \frac{Y_{LL}^*(\theta_2, \phi_2)}{j_L(kR)} & \cdots & \frac{Y_{LL}^*(\theta_z, \phi_z)}{j_L(kR)} \end{bmatrix}. \quad (7.20)$$

The spherical harmonics transform for the filtered reference signals can be defined similarly as

$$\boldsymbol{\alpha}_{uv}(k, j) = \mathcal{T}(k) \mathbf{X}'_{uv}(k, j), \quad (7.21)$$

where  $\boldsymbol{\alpha}_{uv}(k, j)$  are the spherical harmonic coefficients of reference signal  $u$  filtered through the channel responses of secondary source  $v$ .

For a microphone array suitable for spherical harmonics analysis, the transformation of the signal to the spherical harmonic domain changes the number of channels from  $Z$  to  $(L+1)^2$ , and since  $Z \geq (L+1)^2$  when no spatial aliasing occurs, the transform reduces the complexity of the adaptive algorithm by a factor of approximately  $(L+1)^2/Z$ .

The Least-Mean-Square algorithm employed in the system aims to minimize the mean square of all the error inputs. For a multi-channel ANC system without spherical harmonics transformation, the optimization goal is

$$\min \left\{ \sum_z |E_z(k, j)|^2 \right\} \quad (7.22)$$

which minimizes the average signal energy at each of the error microphone. After applying the spherical harmonic transform to both  $\mathbf{X}'(k)$  and  $\mathbf{E}(k)$ , the LMS minimization criteria becomes

$$\min \left\{ \sum_l \sum_m |\beta_{lm}(k, j)|^2 \right\} \quad (7.23)$$

which approximately reduces the noise level within a sphere, rather than at a finite number of points. However, in order to achieve minimum residual acoustic potential energy, an additional weighing needs to be applied to each of the obtained spherical harmonic coefficients, such that

$$\min \left\{ \sum_l \sum_m |W_l \beta_{lm}(k, j)|^2 \right\} = \min \left\{ \int_S P(r, \theta, \phi, k) dS \right\}, \quad (7.24)$$

where the expression of the weights  $W_l$  is given in (6.20). The same weighing needs to be applied to both the filtered reference signal coefficients and the error signal coefficients, as shown in Fig. 7.3. The weighing procedure can be combined with the spherical harmonics transform, and the weighted transform matrix can be expressed as

$$\mathcal{T}_W(k) = \mathbf{W}(k)\mathcal{T}(k) \quad (7.25)$$

where  $\mathbf{W}(k)$  is a  $(L + 1)^2$ -by- $(L + 1)^2$  diagonal matrix, with its diagonal elements arranged as  $\text{diag}\{\mathbf{W}(k)\} = [W_0(k), W_1(k), W_1(k) \dots W_L(k)]^T$ .

The adaptive filter  $\mathbf{w}_{u,v}$  for the  $u$ th reference signal and  $v$ th secondary source is updated according to

$$\mathbf{w}_{uv}(j+1) = \mathbf{w}_{uv}(j) + 2\mu \sum_{l,m} \mathbf{c}_{lm}^{uv}(j+1) \quad (7.26)$$

where  $\mathbf{c}_{lm}^{uv}(j+1)$  is generated by taking the IFFT of the product of  $\boldsymbol{\alpha}_{uv}(k, j+1)$  and  $\boldsymbol{\beta}(k, j+1)$  for all  $k$ ,

$$\mathbf{c}_{lm}^{uv}(j+1) = [\mathbf{O}_N \ \mathbf{T}_N] \mathcal{F}_{2N}^{-1} [\boldsymbol{\alpha}_{uv}(j+1) \otimes \boldsymbol{\beta}(j+1)]. \quad (7.27)$$

The driving signal for the  $v$ th secondary source is generated as

$$\mathbf{y}_v(j+1) = \sum_u [\mathbf{O}_N \ \mathbf{I}_N] \mathcal{F}_{2N}^{-1} [\mathbf{X}_u(j+1) \otimes \mathbf{W}_{uv}(j)], \quad (7.28)$$

where  $\mathbf{W}_{uv}(j)$  is the FFT of  $\mathbf{w}_{uv}(j)$  using (7.8).

It can be seen that due to the frame-based data processing scheme, generation of the secondary driving signals as well as updating of the adaptive filter are delayed by at least  $N$  samples, which is undesirable for ANC applications. One way to reduce the delay of the secondary path signal is to implement (7.28) in the time domain, i.e., use the time domain convolution method to generate the secondary signals [129], this way, the latency of the secondary path signal can be reduced to the same level of time domain ANC algorithms, at the cost of computational efficiency. However, updating of the adaptive filter still needs to be processed frame-by-frame, therefore the updating latency of the adaptive filter is always higher than time domain implementations.

## 7.4 Time domain feed-forward architecture for spatial ANC systems

Although the frequency domain implementation is straight forward and computationally efficient due to its use of FFT, it is sometimes desirable to implement ANC systems in the time domain for lower latency and easier realization in existing embedded systems. In this section, we propose a time domain spatial ANC architecture based on spherical harmonics analysis.

### 7.4.1 Time domain spherical harmonics representation of sound field

The sound pressure on the surface of a spherical region of radius  $r$  at a certain frequency may be represented using spherical harmonics as

$$P(r, \theta, \phi, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \tau_{lm}(k, r) Y_{lm}(\theta, \phi). \quad (7.29)$$

By comparing (7.29) with (2.1), it can be seen that  $\tau_{lm}(k, r)$  is related to the commonly used spherical harmonic coefficients by

$$\tau_{lm}(k, r) = C_{lm}(k) j_l(kr) \quad (7.30)$$

Taking the inverse Fourier transform of (7.29), we have

$$p(r, \theta, \phi, n) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{\tau}_{lm}(n, r) Y_{lm}(\theta, \phi), \quad (7.31)$$

where  $p(r, \theta, \phi, n)$  is the sound pressure at a discrete time index  $n$ , and  $\hat{\tau}_{lm}(n, r)$  is a set of coefficients defined in the time domain. The physical meaning of (7.31) is that the sound pressure on the surface of a spherical region at a certain time instant can be represented by an infinite summation of spherical harmonics. We note that  $p(r, \theta, \phi, n)$  is the sum of spatial sound of all frequencies, and since the higher frequency sound has shorter wave length and hence more complicated pressure field, it is necessary to use the higher order spherical harmonics to represent the higher frequency spatial sound components.

However, since for a given frequency, only a finite number of spherical harmonics

are needed to represent the sound field [53], if the sound signal is band limited, the infinite summation in (7.31) may be truncated to a finite summation up to order  $L$  given by (2.7). When sampling the sound field using a finite number of microphones, in order to avoid spatial aliasing, it is necessary to low pass filter the input signal before applying the time domain spherical harmonic transform, so that the higher order spherical harmonics associated with higher frequency sound components would be removed from the input signal.

Rearranging (7.30) and taking its inverse Fourier transform, we have

$$\widehat{C}_{lm}(n) = \widehat{\tau}_{lm}(n, r) * \mathcal{F}^{-1}\left\{\frac{1}{j_l(kr)}\right\}, \quad (7.32)$$

where  $\widehat{C}_{lm}(n) = \mathcal{F}^{-1}\{C_{lm}\}$ . From (7.32), it can be seen that the time domain spherical harmonic coefficients  $\widehat{C}_{lm}(n)$  can be obtained by filtering the corresponding  $\widehat{\tau}_{lm}(n, r)$  with a filter whose frequency response is equal to  $1/j_l(kr)$ .

Since time domain sound pressure signals are real-valued, it is sufficient to use real-valued spherical harmonics (2.32) as the basis functions. The sound pressure at a certain location  $(r', \theta', \phi')$  and time index  $n$  can be obtained by

$$p(r', \theta', \phi', n) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \widehat{C}_{lm}(n) * \mathcal{F}^{-1}\{j_l(kr')\} Y_{lm}^R(\theta, \phi). \quad (7.33)$$

In order to realize the minimization criteria (7.24), an additional filter is required to obtain the weighted coefficients, this can be expressed as

$$\widetilde{C}_{lm}(n) = \widehat{C}_{lm}(n) * \mathcal{F}^{-1}\{W_l(k)\}. \quad (7.34)$$

We note that the filters corresponding to  $1/j_l(kr)$  and  $W_l(k)$  can be designed such that their frequency responses are accurate only for the interested frequency band, so as to reduce the difficulty and complexity of the filter design problem.

#### 7.4.2 Spatial ANC architecture using time domain spherical harmonics analysis

We propose a time domain ANC system based on the time domain multichannel ANC architecture and the time domain spherical harmonic analysis techniques. The system structure is illustrated in Fig. 7.4.

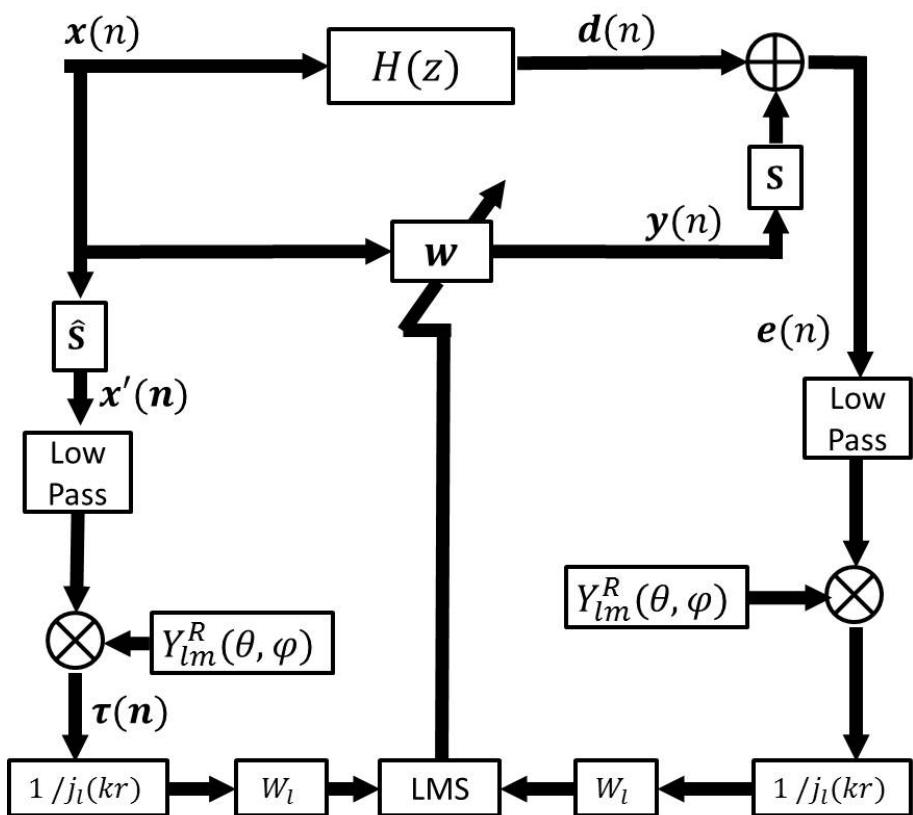


Figure 7.4: Block diagram of the time domain feedforward spatial ANC system.

In the proposed system, the filtered reference signals  $x'_{uvz}(n)$  are obtained by

$$x'_{uvz}(n) = x_u(n) * \hat{s}_{vz}(n). \quad (7.35)$$

The filtered reference signal is then passed through a low pass filter, whose cut off frequency equals to the maximum operating frequency of the ANC system. The maximum frequency needs to agree with the capability of error microphone array, i.e., the array should be able to capture spatial sound field at this frequency without spatial aliasing.

For a uniform spherical error microphone array, the time domain coefficients  $\hat{\tau}_{lm}^{uv}(n)$  of order  $l$  and mode  $m$ , due to reference signal  $u$  and secondary source  $v$ , are obtained by

$$\hat{\tau}_{lm}^{uv}(n) = \sum_z x'_{uvz}(n) Y_{lm}^R(\theta_z, \phi_z), \quad (7.36)$$

where  $(\theta_z, \phi_z)$  denote the angular position of the  $z$ th error microphone. For spherical microphone arrays that do not have a uniform spatial sampling scheme, the alternative option is to solve for the coefficients in a Least-Mean-Square manner, which can be expressed by

$$\hat{\boldsymbol{\tau}}_{uv(n)} = (\mathbf{Y}^R)^{-1} \mathbf{x}'_{uv}(n), \quad (7.37)$$

where  $\hat{\boldsymbol{\tau}}_{uv(n)} = [\hat{\tau}_{00}^{uv}(n), \hat{\tau}_{11}^{uv}(n), \hat{\tau}_{10}^{uv}(n), \dots]^T$  is the vector of all the coefficients  $\hat{\tau}_{lm}^{uv}$  at time instant  $n$ ,  $\mathbf{x}'_{uv}(n) = [x'_{uv1}(n), x'_{uv2}(n), \dots]^T$  is the vector containing the filtered reference signals for all the error microphones at time instant  $n$ , and  $(\mathbf{Y}^R)^{-1}$  is the Moore-Penrose pseudo inverse of the matrix  $\mathbf{Y}^R$ , which is given by

$$\mathbf{Y}^R = \begin{bmatrix} Y_{00}^R(\theta_1, \phi_1) & Y_{11}^R(\theta_1, \phi_1) & Y_{10}^R(\theta_1, \phi_1) & \cdots \\ Y_{00}^R(\theta_2, \phi_2) & Y_{11}^R(\theta_2, \phi_2) & Y_{10}^R(\theta_2, \phi_2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ Y_{00}^R(\theta_Z, \phi_Z) & Y_{11}^R(\theta_Z, \phi_Z) & Y_{10}^R(\theta_Z, \phi_Z) & \cdots \end{bmatrix}. \quad (7.38)$$

The pseudo inverse of the matrix  $\mathbf{Y}^R$  can be obtained offline, therefore it does not increase the computation complexity of the algorithm.

For clarity, instead of using  $\hat{C}_{lm}(n)$ , we use  $\hat{\alpha}_{lm}^{uv}(n)$  and  $\hat{\beta}_{lm}(n)$  to represent the time domain spherical harmonic coefficients for the filtered reference signals and error signals, respectively. We note that both  $\hat{\alpha}_{lm}^{uv}(n)$  and  $\hat{\beta}_{lm}(n)$  are real value coefficients, due to the use of real-valued spherical harmonics. The time domain spherical harmonic coefficients can be obtained according to (7.32). Next, the weighted

spherical harmonic coefficients  $\tilde{\alpha}_{lm}^{uv}(n)$  are obtained by passing the coefficients  $\hat{\alpha}_{lm}^{uv}(n)$  through the weighing filter according to (7.34).

As can be seen in Fig. 7.4, the error signals from the microphone array are processed in the same way as the filtered reference signal, and it is necessary that the same low pass filter is used for both reference signal and error signal. We denote the weighted error coefficients as  $\tilde{\beta}_{lm}(n)$ .

Since the inputs to the Least-Mean-Square algorithm are spherical harmonic coefficients, the LMS algorithm operates on the spherical harmonics domain. The adaptive filter bank  $\mathbf{w}(n)$  has the size  $U \times V \times \mathcal{L}$ . At each time instant,  $\mathbf{w}(n)$  is updated using the equation

$$\mathbf{w}_{uv}(n+1) = \mathbf{w}_{uv}(n) + \mu \sum_l \sum_m \tilde{\alpha}_{lm}^{uv}(n) \otimes \tilde{\beta}_{lm}(n), \quad (7.39)$$

where  $\tilde{\alpha}_{lm}^{uv}(n) = [\tilde{\alpha}_{lm}^{uv}(n), \tilde{\alpha}_{lm}^{uv}(n-1), \tilde{\alpha}_{lm}^{uv}(n-2) \dots \tilde{\alpha}_{lm}^{uv}(n-\mathcal{L}+1)]$  is the vector of the latest  $\mathcal{L}$  samples of the reference signal coefficients, and  $\tilde{\beta}_{lm}(n) = [\tilde{\beta}_{lm}(n), \tilde{\beta}_{lm}(n-1), \tilde{\beta}_{lm}(n-2) \dots \tilde{\beta}_{lm}(n-\mathcal{L}+1)]$  is the vector of the latest  $\mathcal{L}$  samples of the error coefficients.

The driving signal for the  $v$ th loudspeaker is generated the same way as existing multi-channel algorithm, given by

$$y_v(n) = \sum_u \mathbf{w}_{uv}(n) \mathbf{x}_u(n)^T. \quad (7.40)$$

The steps of filtering the reference signal and converting the filtered signal into spherical harmonic domain may be simplified. Consider the equation

$$\hat{\alpha}_{lm}^{uv}(n) = \left( \sum_z (x_u * \hat{s}_{vz}) Y_{lm}^R(\theta_z, \phi_z) \right) * \mathcal{F}^{-1}\left\{ \frac{1}{j_l(kr)} \right\} \quad (7.41)$$

$$= x_u * \left( \sum_z \hat{s}_{vz} Y_{lm}^R(\theta_z, \phi_z) \right) * \mathcal{F}^{-1}\left\{ \frac{1}{j_l(kr)} \right\}, \quad (7.42)$$

we may define the spherical harmonic domain secondary channel impulse response as

$$\hat{\mathcal{S}}_{lm}^v \triangleq \sum_z \hat{s}_{vz} Y_{lm}^R(\theta_z, \phi_z) * \mathcal{F}^{-1}\left\{ \frac{1}{j_l(kr)} \right\}, \quad (7.43)$$

which essentially transforms the secondary channel impulse responses into the spherical harmonic domain, and  $\hat{\mathcal{S}}_{lm}^v$  represents the spherical harmonic impulse response of order  $l$  and mode  $m$ , due to secondary source  $v$ . Since  $\hat{\mathcal{S}}_{lm}^v$  can be calculated of-

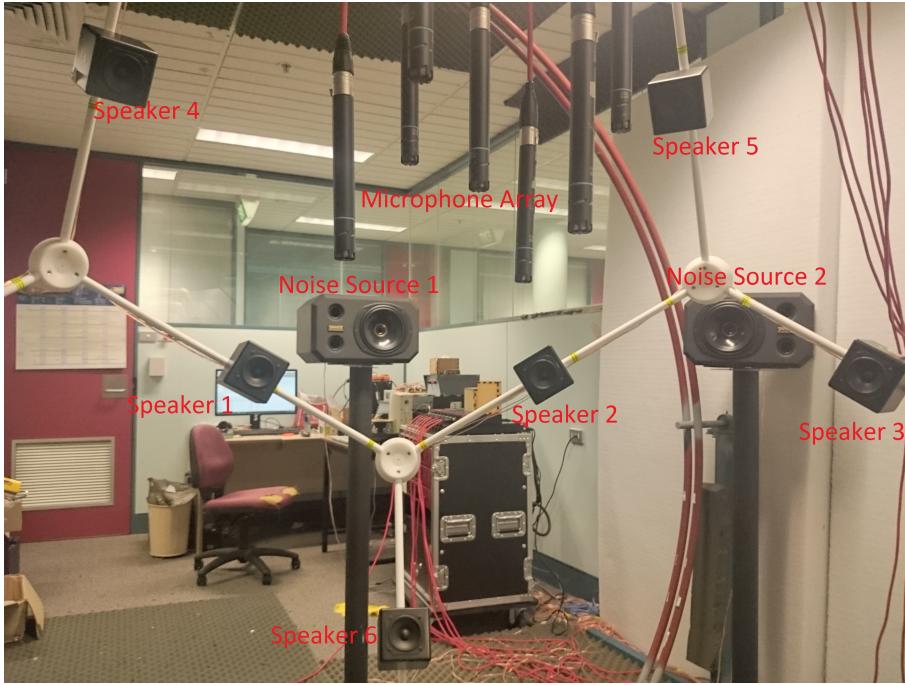


Figure 7.5: Setup of the experimental ANC system.

ffine without knowledge of the reference signal, the computational cost of obtaining  $\hat{\alpha}_{lm}^{uv}(n)$  can be greatly reduced through directly filtering  $x_u(n)$  by  $\hat{\mathcal{S}}_{lm}^v$ , however it should be noted that  $x_u(n)$  needs to be low pass filtered first in order not to result in spatial aliasing.

## 7.5 Experiment validation

### 7.5.1 System setup

An experimental ANC system is set up in the laboratory to validate the proposed spatial ANC algorithm. Fig. 7.5 shows the hardware setup of the ANC system.

In this system, the ANC control region is defined as a spherical area of 29 cm diameter, located at the center of a dodecahedron loudspeaker array system. Two Tannoy 600 loudspeakers are used as the primary source, which are placed 1.5 meters away from the center of the control region, outside of the dodecahedron array. Although the dodecahedron array consists of 30 loudspeakers, only 6 of them are utilized in this experiment as secondary loudspeakers, with their respective numbering shown in Fig. 7.5.

In order to monitor the noise field within the control region, 6 omni-directional

microphones are placed on the boundary of the region. The angular positions of the microphones are  $(60^\circ, 0)$ ,  $(60^\circ, 120^\circ)$ ,  $(60^\circ, 240^\circ)$ ,  $(120^\circ, 60^\circ)$ ,  $(120^\circ, 180^\circ)$  and  $(120^\circ, 300^\circ)$ , respectively. An additional microphone is placed at the center of the array, however this microphone is only used for monitoring purposes and is not part of the ANC system.

The audio playback / record as well as real time signal processing are handled by a desktop PC, with the ANC algorithm implemented using MatLab R2016. The proposed time domain feed-forward spatial ANC system is implemented; we also implemented the time domain MIMO ANC algorithm for comparison. The spatial ANC algorithm is implemented in a frame based manner, with a frame size of 384 samples at a sampling rate of 44100 Hz. At each frame time, the program receives audio input from the microphones, perform the ANC algorithm and generates the loudspeaker driving signals to be played during the next frame. The noise signal is also generated by the program, in order to create a controlled experimental environment.

Normally, in a feed-forward ANC system, the reference noise is picked up by a reference microphone / sensor, sometimes attached to the primary noise source. The digital signal processing system then processes the reference noise and generates the anti-noise signals while the noise sound propagates towards the control region. Ideally, the anti-noise sound is played by the secondary speakers before the noise reaches the control region. In order to achieve this, the signal processing latency must be smaller than the propagation time of the noise, from the error microphone to the control region. Unfortunately, the signal path round trip latency of our system is more than 2000 samples or 45 ms due to the buffering in the computer's data path, which means that the primary sources need to be placed more than 15 meters away from the control region, should a reference microphone be used, which is not possible in our lab condition<sup>1</sup>.

Due to this reason, the reference noise is directly picked up in the electronic path instead of being captured by a microphone, which eliminates the delay in the capture of reference signal completely. In addition, the feedback from secondary sources to the reference microphone is also avoided.

The aim of the experiment is to evaluate the performance of the proposed ANC system under various system configurations. The target frequency band is 200 – 500

---

<sup>1</sup>Using embedded signal processing systems such as microcontrollers, DSPs or FPGAs to implement the adaptive algorithm and AD/DA conversion would significantly reduce the round trip latency, down to only a few milliseconds or less, in which case primary noise source distance would not be a problem.

Hz, this frequency band is chosen because it was found that the typical noise energy inside cars is below 500 Hz, and that the loudspeakers being used as secondary sources have limited low frequency capabilities. Within this frequency band, only the 0th and 1st order spherical harmonic modes exist within the noise field, and the microphone array can reliably pick up the noise field.

In order to evaluate the system's performance at each frequency, for each experiment, a sine wave of a certain frequency is played through one or two of the primary noise sources; after a small period of time, the ANC algorithm begins to function and gradually cancels the noise. The sound pressure received by the microphones is recorded throughout each experiment, and the level of noise attenuation at the microphone positions is calculated by

$$A_{\text{mic}} = 10 \log_{10} \frac{\sum_z E\{|e_z^{\text{bef}}(n)|^2\}}{\sum_z E\{|e_z^{\text{aft}}(n)|^2\}}, \quad (7.44)$$

where  $e_z^{\text{bef}}(n)$  and  $e_z^{\text{aft}}(n)$  denote the sound pressure received by the  $z$ th error microphone before ANC begins and after ANC algorithm converges, respectively.

Compared to the noise attenuation at the microphone positions, the attenuation of the average noise level within the control region provides a better characterization of the ANC performance. In order to calculate this, the noise signals received by the error microphones are converted to spherical harmonic coefficients using (7.37), and then filtered through the weighing filter  $W_l$  to obtain  $\tilde{\beta}_{lm}(n)$ . The attenuation of the average noise energy within the control region is then defined as

$$A_{\text{avg}} = 10 \log_{10} \frac{\sum_{l,m} E\{|\tilde{\beta}_{lm}^{\text{bef}}(n)|^2\}}{\sum_{l,m} E\{|\tilde{\beta}_{lm}^{\text{aft}}(n)|^2\}}, \quad (7.45)$$

where  $\tilde{\beta}_{lm}^{\text{bef}}(n)$  and  $\tilde{\beta}_{lm}^{\text{aft}}(n)$  represent the weighted error signal coefficients before ANC begins and after ANC algorithm fully converges, respectively.

### 7.5.2 Experiment results

#### Comparison between MIMO ANC algorithm and spatial ANC algorithm

In this experiment, we investigate the differences in the performance between the proposed spatial ANC algorithm and the existing MIMO ANC algorithm. For the MIMO algorithm, all of the 6 microphones on the boundary of the control region are utilized as the error microphones. For both algorithms, only the noise source 1

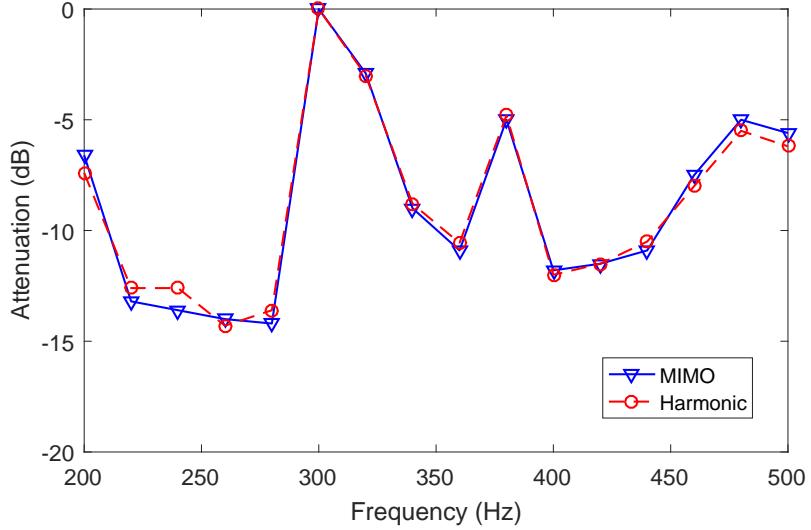


Figure 7.6: Comparison of spatial noise attenuation using MIMO algorithm (blue) and the proposed spatial ANC algorithm (red).

is activated, and we only use speaker 1 (in Fig.7.5) as the secondary source.

Figure 7.6 plots the average noise attenuation within the control region, using the existing MIMO ANC method as well as the proposed spatial ANC method. It can be seen that overall, the two methods result in similar noise attenuations, and the attenuation at lower frequencies are better than that at higher frequencies. Since the 6 microphones are placed evenly over the spherical boundary of the control region (in order to capture the spherical harmonic coefficients), and that their distances are much smaller than the wave length of the noise, they provide a very good representation of the noise level within the field. As a result, the proposed method does not show a clear advantage over the MIMO method.

However, it can be seen from Fig. 7.6 that at higher frequencies (460 Hz and above), the spatial ANC method begins to yield consistently better attenuation than the MIMO method. The reason for this is that as the wavelength shortens, sampling of noise pressure on the boundary of the control region begins to have less correlation with the noise field inside the region. As a result, minimizing the noise pressure at the microphone positions no longer guarantees minimization of the noise level inside the region. The proposed method, on the other hand, is able to control the entire region through converting the microphone signals into the spherical harmonic domain. Should the region size be larger, this phenomenon would be more pronounced.

Of special notice is the peak at 300 Hz in Fig. 7.6, where the attenuation is 0

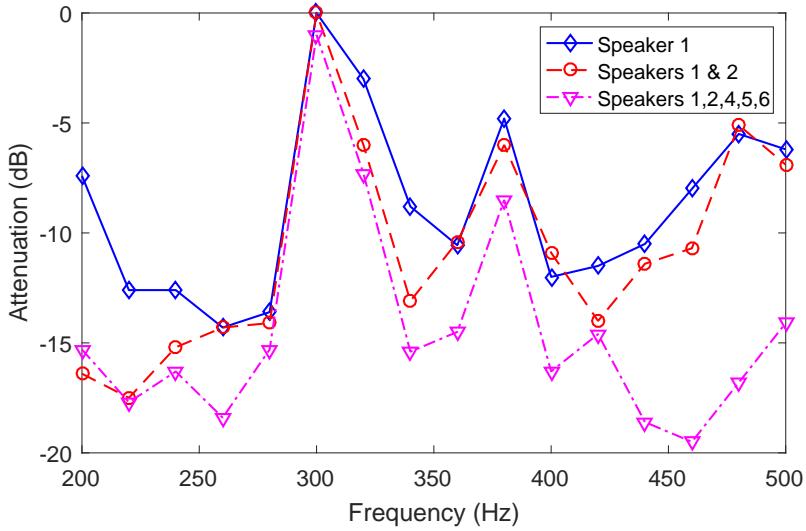


Figure 7.7: Spatial noise attenuation using primary noise source 1 and various number of secondary loudspeakers.

dB for both MIMO algorithm and the spatial algorithm. A careful investigation reveals that at this frequency, a standing wave is formed between the noise source and the walls of the lab, with the control region located at a “minimum” point of the standing wave, i.e., the amplitude of the standing wave is very small. Therefore, the noise field essentially cancels itself without need of the secondary loudspeakers, which results in minimum attenuation gain for the ANC system.

### Impact of secondary source number on spatial ANC performance

In a multi channel ANC system, more than one secondary sources may be employed in order to improve the attenuation level of the system. When the control region is large, or the target frequency band is high, it is expected that a larger number of secondary sources are required to achieve sufficient noise attenuation, due to the increased complexity of the noise field. In this experiment, we validate this assumption using the experimental ANC system.

First, we use only one primary noise source (noise source 1 in Fig. 7.5), and perform the ANC experiment using the proposed algorithm with (i) speaker 1, (ii) speakers 1 & 2, and (iii) speakers 1,2,4,5,6. The average noise attenuation for the three cases at frequencies from 200-500 Hz is shown in Fig. 7.7.

From Fig. 7.7, we can see that indeed the noise attenuation is greater when more secondary sources are active. Using only speaker 1, the system is able to

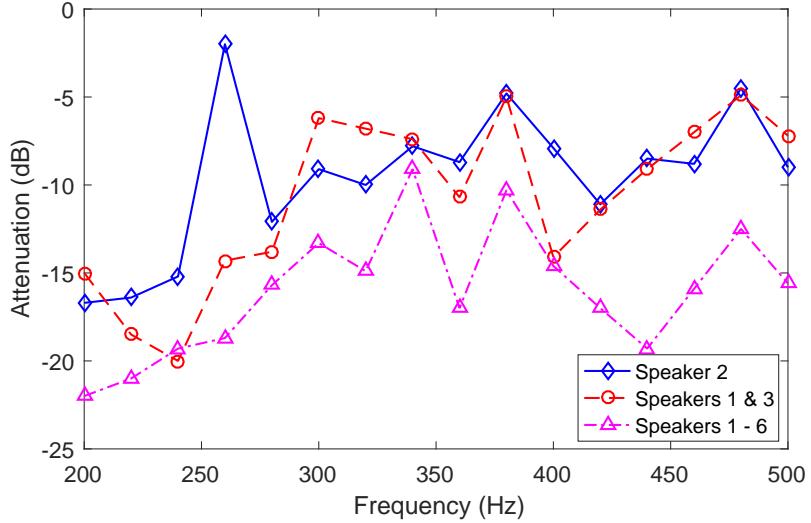


Figure 7.8: Spatial noise attenuation using primary noise source 1 and various number of secondary loudspeakers.

achieve around 10 dB attenuation for most frequencies, and overall, the attenuation is better at lower frequencies (below 300 Hz) than at higher frequencies (above 400 Hz). When speaker 2 is added to the system, a much better attenuation is observed at the lowest frequencies, while a smaller performance gain is achieved at the higher frequencies. Neither of the two configurations were able to attenuate the noise at 300 Hz.

When five secondary sources are used in the ANC system, compared to the case with only speakers 1 and 2, the noise attenuation from 200 Hz to 240 Hz are almost identical. As frequency increases, the difference between the two configurations becomes more significant. At 440 Hz and above, the 5-speaker configuration results in approximately 10 dB higher attenuation than the 2-speaker case. Furthermore, unlike the other two cases, the 5-speaker setup is able to maintain more than 15 dB noise reduction throughout the whole frequency band, with the only exception at 300 Hz, where only 1 db of noise reduction is observed.

The experiment is repeated with both primary noise sources active, with each playing a sine wave at the same frequency, but different phase. This time, the secondary sources being activated are (i) speaker 2, (ii) speakers 1 & 3, and finally (iii) speakers 1-6. We plot the noise attenuation in Fig. 7.8.

The overall trend shown in Fig. 7.8 is similar to that of Fig. 7.7, where the single secondary source case results in the least noise attenuation, and the case where all 6 speakers are used has the most attenuation. The overall trend of decreasing

ANC performance with increased frequency is also observed for the single and duo secondary source cases. At 420 Hz and above, using two speakers (speakers 1 and 3) does not provide significant improvement over using speaker 2 alone. However, when all 6 speakers are used, the noise attenuation can be further improved by 5 – 10 dB.

We also note that when both primary sources are active, the system no longer experience difficulty at 300 Hz, this is because the speakers are capable of cancelling the noise field due to the second noise source, whose noise field does not exhibit the self-cancelling behavior like that of noise source 1.

It can be seen from both Fig. 7.7 and Fig. 7.8 that at lower frequencies, a small number of secondary sources are sufficient to provide more than 15 dB of noise attenuation, and the benefit of adding more secondary sources is only marginal. This is because given the radius of our control region (0.145 m), at around 200 Hz, the 0th order spherical harmonic mode is dominant within the control region. Since the 0th mode is uniform and isotropic, any secondary source located in any direction is capable of producing this mode within the control region, therefore only one or two secondary speakers is sufficient to reproduce the noise field generated by the noise source, hence resulting in high attenuation to the noise energy.

On the other hand, at higher frequencies, the 1st order spherical harmonic modes begin to have more impact on the sound field. Since 1st order modes are directional, a single secondary source can produce good results only if the sound field it produces within the control region is very similar to that of the primary noise source. Because this is very unlikely to be the case, using a single or a small number of secondary sources generally cannot provide high attenuation at higher frequencies. The combined use of multiple secondary sources, however, can substantially improve the spatial ANC system's sound field reproduction capability, hence leads to better results.

### **Impact of secondary source position on spatial ANC performance**

In Section 7.5.2, it was stated that the noise attenuation is related to the secondary source's ability to reproduce the noise field, which in turn is related to the placement of secondary sources in relation to the primary noise. This is investigated in more detail in this section.

In this experiment, we first use primary source 1 to generate the noise field, and compare the ANC performance using secondary speaker 1 and secondary speaker 3. We plot the experiment results in Fig. 7.9.

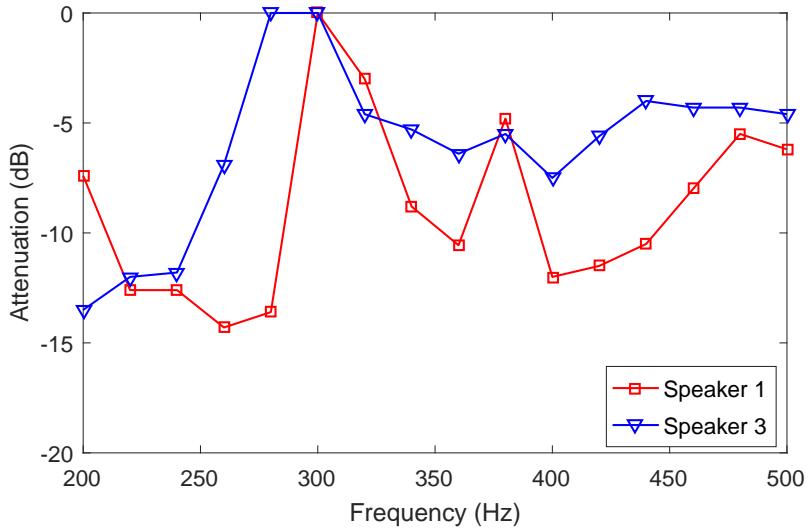


Figure 7.9: Spatial noise attenuation using primary noise source 1 with secondary speaker 1 (red) and secondary speaker 3 (blue).

It can be observed from Fig. 7.9 that the overall noise attenuation using speaker 1 is superior to that of speaker 3. Below 250 Hz, the difference of the two is not very significant; however, at higher frequencies, speaker 1 begins to consistently outperform speaker 3.

From Fig. 7.5, it can be seen that if viewed from the center of the control region, secondary speaker 1 lies approximately in the same direction as primary source 1; on the other hand, secondary speaker 3 forms a 45 degree angle with primary source 1. Since the loudspeakers employed in this experiment (both primary sources and secondary sources) can be approximately seen as point sources, it can be expected that the sound field produced by secondary speaker 1 would be very close to that of primary source 1. Secondary speaker 3, on the other hand, will produce a very different sound field due to its different impinging direction.

We note that in the case of speaker 3, the ANC system fails to converge at 280 Hz, in addition to 300 Hz. The failure at 280 Hz is not due to a standing wave of the noise field, but likely due to the secondary channel being very different from the primary channel, thus causing difficulty with the convergence of the algorithm.

The experiment is also repeated with multiple secondary sources. Instead of changing the position of secondary sources, we examine the impact of primary source position on ANC performance, given a fixed set of secondary loudspeakers. The secondary sources used in this experiment are speakers 1,2,4,5 and 6. We plot the noise attenuation results achieved using primary source 1 and primary source 3 in

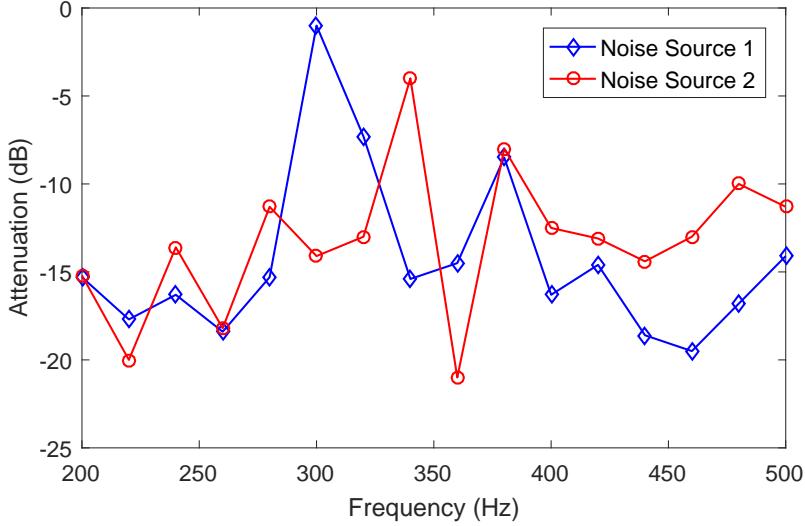


Figure 7.10: Spatial noise attenuation using secondary speakers 1,2,4,5,6 with primary source 1 (blue) and primary source 2 (red).

Fig. 7.10.

In Fig. 7.10 we can see that the noise attenuation for both noise sources are very similar at lower frequencies; at 400 Hz and above, however, the noise attenuation for primary source 1 becomes significantly better than that of primary source 2, with more than 5 dB extra attenuation at some frequencies. Also, we note that primary source 2 does not produce the standing wave like primary source 1 does at 300 Hz, and therefore our system is able to yield a noise attenuation consistent with other frequencies.

It can be seen from Fig. 7.5 that the five selected secondary sources essentially “surrounds” primary source 1, if viewed from the control region, with secondary speakers 1 and 2 being the closest to the noise source. The other three secondary speakers have a very different elevation angle than the noise source, however their elevation angles approximately coincides with the reflected waves from the ceiling and the floor, emitted by primary source 1. Therefore, the secondary source setup provides a very good coverage of the noise field generated by primary source 1, hence the high, consistent ANC performance across the whole frequency band. In the case of primary source 2, due to its different azimuth angle, the secondary array’s capability to reproduce its noise field is limited and degrades gradually as the frequency increases and the sound field becomes more complex, which results in a slowly decaying attenuation level.

Overall, it can be concluded that the performance of the spatial ANC system

is affected by both the number of secondary sources and the relative position between secondary sources and primary sources. For a small control region and low frequencies, even a small number of secondary sources is sufficient to provide adequate noise attenuation, and the system performance is not very sensitive to the placement of secondary sources. As the frequency increases, the system's performance becomes more sensitive to the number and position of secondary sources. If the secondary sources are placed such that they cover the impinging directions of the primary noises, it is possible to achieve consistent noise attenuation of over 15 dB for frequencies up to 500 Hz within a spherical region of 0.145 m radius, using our proposed spatial ANC system.

## 7.6 Summary

In this chapter we propose a spatial active noise cancellation algorithm based on spherical harmonics decomposition of the noise field. Both the frequency domain implementation and the equivalent time domain implementation of the algorithm are discussed. The proposed algorithm allows flexible placement of secondary sources, and is able to optimize the noise attenuation for a given spherical region. Through a series of experiments with the proposed ANC system, we show that the spatial noise cancelling quality of the system depends on both the secondary source numbers and placement, and an average noise attenuation of over 15 dB is achievable for a spherical region of 0.145m radius, at a frequency range of 200-500 Hz, using the proposed spatial ANC system.



# Chapter 8

## Conclusion and future works

### 8.1 Conclusion

For the purpose of effectively attenuating the noise level inside a spatial region in a practical environment, using active noise control methods, a number of problems and difficulties have to be addressed. These include accurate acquisition of noise field information, as well as generation of optimal noise cancelling signals. The goal of this thesis was to address these problems by proposing a number of signal processing algorithms, which includes algorithms for spatial sound recording, noise environment modelling, as well as spatial adaptive ANC system architecture.

Spherical harmonic analysis has been shown to be an efficient and accurate tool for representing spatial sound field. However, existing microphone array layouts suitable for spherical harmonic analysis all exhibit 3D geometries. In Chapter 3, we proposed a 2D planar microphone array layout which has the capability of capturing 3D spatial sound field. Through the use of vertically placed first order microphone units, the proposed planar array is able to detect sound field components that are “invisible” on a plane. The proposed array geometry is shown to have the same capability as a spherical array of the same radius. However, it is desirable to use high precision microphones to implement the proposed array, since its robustness is inferior to spherical arrays. In Chapter 4, we also propose a generalization of this method, which allows the use of planar higher order microphone arrays to sample 3D sound field. This method reduces the total number of microphones required for sound field recording, hence promotes the feasibility of spherical harmonic analysis in real-life ANC applications.

In a reverberant environment, the noise field due to a single source can be re-

flected multiple times, thereby creating a more complex noise field. Such kind of noise field is harder to control, due to its wide range of impinging directions. It is therefore critical to have a method to estimate the level of reverberation in a given environment, so as to aid the design of the ANC system. In Chapter 5, we developed an algorithm for Direct-to-Reverberant Ratio estimation. Compared to existing methods for DRR estimation, the proposed method provides a more accurate modelling of the reverberant field, therefore its estimation of DRR is more accurate. The sound field information required by the algorithm can be captured by a first order microphone system.

In order to develop spatial noise cancellation techniques, one first needs a metric to measure the average noise level inside a spatial area. We proposed one such metric in Chapter 6, which can be calculated by taking the weighted squared sum of the spherical harmonic coefficients of the sound field. Using this metric, we developed a method to predict the optimum spatial ANC performance in a given noise environment, before physically implementing an ANC system. This method is then applied to estimate the performance of in-car loudspeakers for the purpose of cancelling the car cabin noise under various driving conditions. In this chapter, we also show that by appropriately weighing the spherical harmonic coefficients, it is possible to optimize the ANC performance at a number of sub-regions within the desired quiet zone, only at a small sacrifice to the global noise reduction performance. This technique is shown to be especially useful when the number of secondary sources is insufficient.

In Chapter 7, we developed an adaptive ANC algorithm based on the spherical harmonic analysis technique. Through transforming the microphone signals into the spherical harmonics domain, the algorithm is able to optimize the noise attenuation within a spherical region. Both frequency domain and time domain implementations are discussed. An experimental spatial ANC system based on the proposed algorithm has been implemented, and we used this system to investigate the performance of ANC system under various system configurations.

Overall, it can be concluded that despite the challenges that still remain unsolved, the spatial active noise cancellation technique has been developed to the point where it does not only exist in theory and simulations, but has become feasible and practical to be deployed in many applications to solve real-world problems. Further development of the algorithms related to spatial ANC would surely lead to improvements in the performance of spatial ANC systems, as well as identifying more potential applications for the technique.

## 8.2 Future works

A number of problems that arose from the work presented in this thesis are listed below.

### Planar microphone array for mounting on rigid surfaces

The planar microphone array aperture described in this work reduces the space requirement for capturing 3D spatial sound using a microphone array. However, the proposed array is developed under the free field assumption, i.e., there should not be sound sources or reflectors within the spherical region which the array covers. This means that the planar array cannot be mounted on walls or ceilings for convenient deployment. A possible approach to address this problem is to model the sound field in the proximity of a planar reflector, and incorporate this sound field model into the calculation of the spherical harmonic coefficients. Since in this case, the sound impinging direction is limited to within a hemisphere, which reduces the complexity or dimensionality of the sound field, the total number of microphones may be reduced.

### Compressive sampling of noise field

In Chapter 6, we have shown that the noise field inside a vehicle cabin shows a tendency of sparseness, where a small number of basis functions is sufficient to represent the noise field at any time instant. This is likely to be the case especially when the number of independent noise sources is small. In such cases, a relatively small number of well placed microphones may be sufficient to monitor the noise field, given prior knowledge of the basis functions that could describe the noise field. This would allow much more flexible placement of error microphones for a spatial ANC system.

### Secondary channel transformation

The adaptive ANC algorithm described in this work only transforms the error microphone signals to spherical harmonics domain; the secondary loudspeaker model used here is the same as the MIMO ANC algorithms. Since the convergence speed of the algorithm depends on the secondary channel information, when the channels of multiple loudspeakers show a strong correlation, convergence speed will become

slower. In a previous work, the secondary driving signals are transformed into circular harmonics, which overcomes this problem. The drawback of this method is that it requires a circular loudspeaker array, which is often impractical. For non-circular speaker arrays, it may be possible to define a different transformation which does not require specific loudspeaker placement, thus improving the convergence speed of the adaptive algorithm.

## Online secondary channel estimation

The adaptive algorithms presented in this work all require the secondary channel information. Although this information can be obtained offline by playing a sweep signal from each secondary loudspeaker, it is still desirable to be able to update the secondary path while the ANC system is functioning. This is because the secondary path changes with the slightest movement of objects inside and around the control region, when the pre-recorded secondary path deviates too far from the current secondary path, the adaptive system would become unstable. With online secondary channel estimation, the ANC system can keep track of the changes of the secondary path, thus promoting system stability and noise attenuation level.

## Feedback adaptive spatial ANC algorithm

The feedback ANC algorithm has a number of advantages over its feed-forward counterpart, such as the omission of reference microphones. Incorporating the spatial noise control techniques developed in this work into feedback ANC systems would certainly bring benefits such as a simplified system structure. However, the limitations of feedback algorithms would also apply, which includes its inability to cancel non-periodic noise, and its inferior stability compared to feed-forward ANC algorithm. Furthermore, the feedback algorithm has a higher computational cost than the feed-forward algorithm, due to its need to synthesize reference signals. These challenges should be investigated to determine the feasibility of a spatial feedback ANC system in real-life applications.

## Active Noise Control in reverberant environments

In a reverberant environment, the sound field due to a simple sound source can become very complex due to the reverberations. The long impulse response also poses a challenge to the active noise control systems deployed in such environments.

It is therefore worthwhile to investigate how badly the reverberation can impact ANC performance, and what modifications can be done to the ANC algorithm to improve its performance in such environments. The DRR estimation method proposed in this thesis would become very useful for this purpose, because it allows the ANC system to identify the strength of the reverberation, so that the algorithm can adjust accordingly to optimize noise attenuation.



# Bibliography

- [1] Sharon G. Kujawa and M. Charles Liberman, “Adding insult to injury: Cochlear nerve degeneration after “temporary” noise-induced hearing loss,” vol. 29, no. 45, pp. 14077–14085, 2009.
- [2] S. J. Elliott and P. A. Nelson, “Active noise control,” *IEEE Signal Processing Magazine*, vol. 10, no. 4, pp. 12–35, Oct 1993.
- [3] S. M. Kuo, S. Mitra, and Woon-Seng Gan, “Active noise control system for headphone applications,” *IEEE Transactions on Control Systems Technology*, vol. 14, no. 2, pp. 331–335, March 2006.
- [4] S. M. Kuo and S. Mitra, “Design of noise reduction headphone,” in *Proc. 2006 Digest of Technical Papers International Conference on Consumer Electronics*, Jan 2006, pp. 457–458.
- [5] M. Guldenschuh, A. Sontacchi, M. Perkmann, and M. Opitz, “Assessment of active noise cancelling headphones,” in *Proc. 2012 IEEE Second International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Sept 2012, pp. 299–303.
- [6] S. M. Kuo and D. R. Morgan, “Active noise control: a tutorial review,” *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–973, 1999.
- [7] H. Sano, T. Inoue, A. Takahashi, K. Terai, and Y. Nakamura, “Active control system for low-frequency road noise combined with an audio system,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, pp. 755–763, 2001.
- [8] R. J. Bernhard, “Active control of road noise inside automobiles,” in *Inter-noise and Noise-con Congress and Conference Proceedings*. Institute of Noise Control Engineering, 1995, vol. 1995, pp. 21–32.

- [9] A. Barkefors, S. Berthilsson, and M. Sternad, “Extending the area silenced by active noise control using multiple loudspeakers,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 325–328.
- [10] T. D. Abhayapala and D. B. Ward, “Theory and design of high order sound field microphones using spherical microphone array,” in *Proc. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, vol. 2, pp. II–1949–II–1952.
- [11] D. Khaykin and B. Rafaely, “Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009, pp. 221–224.
- [12] D. P. Jarrett, O. Thiergart, E. A. P. Habets, and P. A. Naylor, “Coherence-based diffuseness estimation in the spherical harmonic domain,” in *Proc. 2012 IEEE 27th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, Nov 2012, pp. 1–5.
- [13] T. D. Abhayapala and A. Gupta, “Spherical harmonic analysis of wavefields using multiple circular sensor arrays,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1655–1666, 2010.
- [14] G. W. Elko and J. M. Meyer, “Using a higher-order spherical microphone array to assess spatial and temporal distribution of sound in rooms,” *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1912–1912, 2012.
- [15] B. Rafaely, “The spherical-shell microphone array,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 740–747, 2008.
- [16] C. T. Jin, N. Epain, and A. Parthy, “Design, optimization and evaluation of a dual-radius spherical microphone array,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 193–204, Jan 2014.
- [17] T. D. Abhayapala and C. T. Chan, “Limitation and error analysis of spherical microphone arrays,” in *Proc. 14th International Congress on Sound and Vibration (ICSV14)*, 2007.

- [18] M. Chan, “Theory and design of higher order sound field recording,” *Department of Engineering, FEIT, ANU, Honours Thesis*, 2003.
- [19] Rishabh RANJAN, Jianjun HE, Tatsuya MURAO, Lam BHAN, and Woon Seng GAN, “Selective active noise control system for open windows using sound classification,” in *Proc. Inter.noise 2016*, Nov 2016.
- [20] Chuang Shi, Tatsuya Murao, Dongyuan Shi, Bhan Lam, and Woon-Seng Gan, “Open loop active control of noise through open windows,” *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3313–3313, 2016.
- [21] Tatsuya Murao, Chuang Shi, Woon-Seng Gan, and Masaharu Nishimura, “Mixed-error approach for multi-channel active noise control of open windows,” *Applied Acoustics*, vol. 127, pp. 305 – 315, 2017.
- [22] Jordan Cheer and Stephen J. Elliott, “Multichannel control systems for the attenuation of interior road noise in vehicles,” *Mechanical Systems and Signal Processing*, vol. 6061, pp. 753 – 769, 2015.
- [23] S. J. Elliott, W. Jung, and J. Cheer, “The spatial properties and local active control of road noise,” in *Proc. of Euro-noise*, 2015, pp. 2189–2194.
- [24] J. Cheer and S. J. Elliott, “Mutlichannel feedback control of interior road noise,” in *Proceedings of Meetings on Acoustics ICA2013*. ASA, 2013, vol. 19, p. 030118.
- [25] Kosuke Sakamoto and Toshio Inoue, “Development of feedback-based active road noise control technology for noise in multiple narrow-frequency bands and integration with booming noise active noise control system,” *SAE International Journal of Passenger Cars-Mechanical Systems*, vol. 8, no. 2015-01-0660, pp. 1–7, 2015.
- [26] Akira Takahashi, Toshio Inoue, Kosuke Sakamoto, and Yasunori Kobayashi, “Integrated active noise control system for low-frequency noise in automobiles,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering, 2011, vol. 2011, pp. 2105–2113.
- [27] J. Cheer and S. J. Elliott, “Active noise control of a diesel generator in a luxury yacht,” *Applied Acoustics*, vol. 105, pp. 209–214, 2016.

- [28] J. Cheer and S. Daley, "An investigation of delayless subband adaptive filtering for multi-input multi-output active noise control applications," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 359–373, Feb 2017.
- [29] Tongwei Wang, Woon-Seng Gan, and Sen M. Kuo, "New feedback active noise control system with improved performance," in *Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, pp. 6662–6666.
- [30] J. Zhang, W. Zhang, and T. D. Abhayapala, "Noise cancellation over spatial regions using adaptive wave domain processing," in *Proc. 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.
- [31] S. Spors and H. Buchner, "Efficient massive multichannel active noise control using wave-domain adaptive filtering," in *Proc. 3rd International Symposium on Communications, Control and Signal Processing*, March 2008, pp. 1480–1485.
- [32] T. D. Abhayapala, *Modal Analysis and Synthesis of Broadband Nearfield Beamforming Arrays*, Ph.D. thesis, The Australian National University and Telecommunications Engineering Group, 12 1999.
- [33] E. Tiana-Roig, F. Jacobsen, and E. Fernandez-Grande, "Beamforming with a circular array of microphones mounted on a rigid sphere," *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1095–1098, 2011.
- [34] Shefeng Yan, Haohai Sun, U. P. Svensson, Xiaochuan Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 361–371, 2011.
- [35] C. Lai, S. Nordholm, and Y. Leung, "Design of steerable spherical broadband beamformers with flexible sensor configurations," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 427–438, 2013.
- [36] T. D. Abhayapala, R. A. Kennedy, and R. C. Williamson, "Nearfield broadband array design using a radially invariant modal expansion," *The Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 392–403, 2000.

- [37] R. A. Kennedy, T. D. Abhayapala, D. B. Ward, and R. C. Williamson, “Nearfield broadband frequency invariant beamforming,” in *Proc. 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*. IEEE, 1996, vol. 2, pp. 905–908.
- [38] T. D. Abhayapala, R. A. Kennedy, R. C. Williamson, and D. B. Ward, “Nearfield broadband adaptive beamforming,” in *Proc. of the Fifth International Symposium on Signal Processing and Its Applications (ISSPA'99)*. IEEE, 1999, vol. 2, pp. 839–842.
- [39] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, “Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing,” *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2828–2840, 2012.
- [40] D. Khaykin and B. Rafaely, “Acoustic analysis by spherical microphone array processing of room impulse responses,” *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 261–270, 2012.
- [41] P. N. Samarasinghe, T. D. Abhayapala, M. A. Poletti, and T. Betlehem, “An efficient parameterization of the room transfer function,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2217–2227, 2015.
- [42] P. N. Samarasinghe, T. D. Abhayapala, M. A. Poletti, and T. Betlehem, “On room impulse response between arbitrary points: An efficient parameterization,” in *Proc. 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*. IEEE, 2014, pp. 153–156.
- [43] B. Rafaely, “Analysis and design of spherical microphone arrays,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.
- [44] M. A. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.
- [45] D. B. Ward and T. D. Abhayapala, “Performance bounds on sound field reproduction using a loudspeaker array,” in *Proc. of Workshop on App. of Sig. Proc. to Audio and Acoustics, Mohonk*, 2001.

- [46] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707, September 2001.
- [47] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2008*, 2008, pp. 5276–5279.
- [48] Y. Peled and B. Rafaely, "Method for dereverberation and noise reduction using spherical microphone arrays," in *Proc. 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 113–116.
- [49] D. P. Jarrett and E. A P Habets, "On the noise reduction performance of a spherical harmonic domain tradeoff beamformer," *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 773–776, 2012.
- [50] E. de Witte, H. D. Griffiths, and P. V. Brennan, "Phase mode processing for spherical antenna arrays," *Electronics Letters*, vol. 39, no. 20, pp. 1430–1431, 2003.
- [51] A. J. Berkhouit, D. de Vires, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, no. 5, pp. 2764–2778, 1993.
- [52] Boaz Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, 2004.
- [53] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2542–2556, 2007.
- [54] H. M. Jones, R. A. Kennedy, and T. D. Abhayapala, "On dimensionality of multipath fields: Spatial extent and richness," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2002, vol. 3, pp. III–2837.
- [55] T. D. Abhayapala, T. S. Pollock, and R. A. Kennedy, "Spatial decomposition of mimo wireless channels," in *Proc. Seventh International Symposium on Signal Processing and Its Applications*. IEEE, 2003, vol. 1, pp. 309–312.

- [56] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, p. 955, Academic Press, 2000.
- [57] P. A. Martin, *Multiple scattering: interaction of time harmonic waves with N obstacles*, Cambridge Univ., 2006.
- [58] P. N. Samarasinghe, T. D. Abhayapala, and M. A Poletti, “3D spatial sound-field recording over large regions,” in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2012, pp. 1–4.
- [59] J. Kautz, J. Snyder, and P. J. Sloan, “Fast arbitrary BRDF shading for low-frequency lighting using spherical harmonics,” *Rendering Techniques*, vol. 2, pp. 291–296, 2002.
- [60] R. Rabenstein A. Kuntz, “Cardioid pattern optimization for a virtual circular microphone array,” in *Proc. of the EAA Symposium on Auralizatio, Espoo, Finland*, 2009.
- [61] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, p. 439, Number 55. Courier Corporation, 1964.
- [62] Frederik J. Simons, “Slepian functions and their use in signal estimation and spectral analysis,” in *Handbook of Geomathematics*, Willi Freeden, M. Zuhair Nashed, and Thomas Sonar, Eds., pp. 891–923. Springer Berlin Heidelberg, 2010.
- [63] R. H. Rapp and N. K. Pavlis, “The development and analysis of geopotential coefficient models to spherical harmonic degree 360,” *Journal of Geophysical Research: Solid Earth*, vol. 95, no. B13, pp. 21885–21911, 1990.
- [64] J. Meyer and G. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2002, vol. 2, pp. II–1781–II–1784.
- [65] E. Mabande, K. Kowalczyk, H. Sun, and W. Kellermann, “Room geometry inference based on spherical microphone array eigenbeam processing,” *The Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 2773–2789, 2013.

- [66] F. Jacobsen, G. Moreno-Pescador, E. Fernandez-Grande, and Jørgen Hald, “Near field acoustic holography with microphones on a rigid sphere,” *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 3461–3464, 2011.
- [67] I. Balmages and B. Rafaely, “Open-sphere designs for spherical microphone arrays,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 727–732, 2007.
- [68] T. D. Abhayapala and M.C.T. Chan, “Limitation and error analysis of spherical microphone arrays,” in *Proc. 14th International Congress on Sound and Vibration (ICSV14)*, Cairns, Australia, July 2007.
- [69] C. Jin, A. Parthy, and A. Van Schaik, “Optimisation of co-centred rigid and open spherical microphone arrays,” in *Proc. 120th Audio Engineering Society Convention*, Paris, France, May 2006, p. 6 pages, Audio Engineering Society.
- [70] Z. Li and R. Duraiswami, “Flexible and optimal design of spherical microphone arrays for beamforming,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 702–714, 2007.
- [71] A. Gupta and T. D. Abhayapala, “Double sided cone array for spherical harmonic analysis of wavefields,” in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 77–80.
- [72] T. D. Abhayapala, A. Gupta, et al., “Non-spherical microphone array structures for 3d beamforming and spherical harmonic analysis,” in *Proc. of the 11th International Workshop on Acoustic Echo and Noise Control*, 2008.
- [73] J. M. Meyer and G. W. Elko, “Spherical harmonic modal beamforming for an augmented circular microphone array,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 5280–5283.
- [74] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, chapter 3, pp. 53–54, Springer, Mar. 2007.
- [75] H. Chen, T. D. Abhayapala, and W. Zhang, “Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis,” *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015.

- [76] P. N. Samarasinghe, T. D. Abhayapala, and M. A Poletti, “Wavefield analysis over large areas using distributed higher order microphones,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 3, pp. 647–658, Mar. 2014.
- [77] H. Chen, T. D. Abhayapala, and W. Zhang, “3D sound field analysis using circular higher-order microphone array,” in *Proc. 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 1153–1157.
- [78] D. Griesinger, “The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment,” in *Proc. Audio Engineering Society Convention 126*. Audio Engineering Society, 2009.
- [79] K. Lebart, J. M. Boucher, and P. N. Denbigh, “A new method based on spectral subtraction for speech dereverberation,” *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 359–366, 2001.
- [80] C. Marro, Y. Mahieux, and K. U. Simmer, “Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering,” *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, 1998.
- [81] D. B. Hawkins and W. S. Yacullo, “Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation,” *Journal of Speech and Hearing Disorders*, vol. 49, no. 3, pp. 278–286, 1984.
- [82] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, “On the minimum audible difference in direct-to-reverberant energy ratio,” *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, 2008.
- [83] M. Laitinen and V. Pulkki, “Utilizing instantaneous direct-to-reverberant ratio in parametric spatial audio coding,” in *Proc. Audio Engineering Society Convention 133*. Audio Engineering Society, Oct 2012.
- [84] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, “Auditory distance perception in humans: A summary of past and present research,” *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.
- [85] A. J. Kolarik, S. Cirstea, and S. Pardhan, “Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-

- to-reverberant cues,” *Experimental brain research*, vol. 224, no. 4, pp. 623–633, 2013.
- [86] E. Larsen, C. D. Schmitz, C. R. Lansing, W. D. O’Brien Jr, B. C. Wheeler, and A. S. Feng, “Acoustic scene analysis using estimated impulse responses,” in *Proc. IEEE Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, 2003, vol. 1, pp. 725–729.
- [87] T. H. Falk and W. Chan, “Temporal dynamics for blind measurement of room acoustical parameters,” *IEEE Trans. on Instrumentation and Measurement*, vol. 59, no. 4, pp. 978–989, 2010.
- [88] S. Mosayyebpour, H. Sheikhzadeh, T. A. Gulliver, and M. Esmaeili, “Single-microphone LP residual skewness-based inverse filtering of the room impulse response,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1617–1632, 2012.
- [89] P. P. Parada, D. Sharma, T. Waterschoot, and P. A. Naylor, “Evaluating the non-intrusive room acoustics algorithm with the ACE challenge,” in *Proc. ACE Challenge Workshop, a satellite event of WASPAA, New Paltz, NY, USA*, Oct 2015.
- [90] Y. Lu and M. Cooke, “Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1793–1805, 2010.
- [91] S. Vesa, “Sound source distance learning based on binaural signals,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct 2007, pp. 271–274.
- [92] M. Jeub, C. Nelke, C. Beaugeant, and P. Vary, “Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals,” in *Proc. 19th European Signal Processing Conference*, Aug 2011, pp. 1347–1351.
- [93] O. Thiergart, G. Del Galdo, and E. A. P. Habets, “Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones.,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 309–312.

- [94] E. Georganti, J. Mourjopoulos, and S. van de Par, “Room statistics and direct-to-reverberant ratio estimation from dual-channel signals,” in *Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4713–4717.
- [95] O. Thiergart, T. Ascherl, and E. A. P. Habets, “Power-based signal-to-diffuse ratio estimation using noisy directional microphones,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 7440–7444.
- [96] Y. Hioka and K. Niwa, “PSD estimation in beamspace for estimating direct-to-reverberant ratio from a reverberant speech signal,” in *Proc. ACE Challenge Workshop, a satellite event of WASPAA, New Paltz, NY, USA*, Oct 2015.
- [97] Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, and Y. Haneda, “Estimating direct-to-reverberant energy ratio using D/R spatial correlation matrix model,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2374–2384, 2011.
- [98] M. Kuster, “Estimating the direct-to-reverberant energy ratio from the coherence between coincident pressure and particle velocity,” *The Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 3781–3787, 2011.
- [99] H. Chen, P. N. Samarasinghe, T. D. Abhayapala, and W. Zhang, “Estimation of the direct-to-reverberant energy ratio using a spherical microphone array.,” in *Proc. ACE Challenge Workshop, a satellite event of WASPAA, New Paltz, NY, USA*, Oct 2015.
- [100] J. Eaton, A. H. Moore, N. D. Gaubitch, and P. A. Naylor, “The ACE challenge - corpus description and performance evaluation,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA*, Oct 2015.
- [101] E. G. Williams, *Fourier Acoustics: Sound Radiation and Near field Acoustical Holography*, USA: Academic, 1999.
- [102] F.J. Fahy, *Sound Intensity*, Elsevier Applied Science, London, 1989.
- [103] T. D. Abhayapala and H. Bhatta, “Coherent broadband source localization by modal space processing,” in *Proc. 10th International Conference on Telecommunications (ICT 2003)*, 2003, vol. 2, pp. 1617–1623.

- [104] H. Chen, T. D. Abhayapala, P. Samarasinghe, and W. Zhang, “Direct-to-reverberant energy ratio estimation using a first order microphone,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. PP, no. 99, pp. 1–1, 2016.
- [105] N. Epain and E. Friot, “Active control of sound inside a sphere via control of the acoustic pressure at the boundary surface,” *J. Sound Vibr.*, vol. 299, no. 3, pp. 587–604, 2007.
- [106] M. Naoe, T. Kimura, and M. Katumoto, “Performance evaluation of 3D sound field reproduction system using a few loudspeakers and wave field synthesis,” in *Proc. 2nd Int. Symp. Universal Commun., Osaka, Japan*, 2008.
- [107] F. M. Fazi, P. A. Nelson, Christensen J. E. N., and J. Seo, “Surround system based on three-dimensional sound field reconstruction,” in *Proc. 125th Conv. Audio Eng. Soc., San Francisco, CA, USA*, 2008.
- [108] W. Zhang and T. D. Abhayapala, “Three dimensional sound field reproduction using multiple circular loudspeaker arrays: Functional analysis guided approach,” *IEEE Trans. Audio, Speech, Lang. Process*, vol. 22, no. 7, pp. 1184–1194, 2014.
- [109] Y. J. Wu and T. D. Abhayapala, “Spatial multizone sound field reproduction: Theory and design,” *IEEE Trans. Audio, Speech, Lang. Process*, vol. 19, no. 6, pp. 1711–1715, 2011.
- [110] “em32 eigenmike microphone array release notes,” [www.mhacoustics.com/sites/default/files/ReleaseNotes.pdf](http://www.mhacoustics.com/sites/default/files/ReleaseNotes.pdf), 2013.
- [111] S. C. Douglas, “Fast implementations of the filtered-X LMS and LMS algorithms for multichannel active noise control,” *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, pp. 454–465, Jul 1999.
- [112] M. de Diego, A. Gonzalez, M. Ferrer, and G. Pinero, “An adaptive algorithms comparison for real multichannel active noise control,” in *Proc. 12th European Signal Processing Conference*, Sep 2004, pp. 925–928.
- [113] M. de Diego, A. Gonzalez, M. Ferrer, and G. Pinero, “Multichannel active noise control system for local spectral reshaping of multifrequency noise,” *Journal of Sound and Vibration*, vol. 274, no. 1C2, pp. 249 – 271, 2004.

- [114] A. Montazeri, J. Poshtan, and M. H. Kahaei, “Analysis of the global reduction of broadband noise in a telephone kiosk using a MIMO modal ANC system,” *International Journal of Engineering Science*, vol. 45, no. 2C8, pp. 679 – 697, 2007.
- [115] H. Chen, P. Samarasinghe, T. D. Abhayapala, and W. Zhang, “Spatial noise cancellation inside cars: Performance analysis and experimental results,” in *Proc. 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.
- [116] P. N. Samarasinghe, W. Zhang, and T. D. Abhayapala, “Recent advances in active noise control inside automobile cabins: Toward quieter cars,” *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 61–73, 2016.
- [117] P. A. Nelson and S. J. Elliott, *Active control of sound*, Academic press, 1991.
- [118] S. Hasegawa, T. Tabata, A. Kinoshita, and H. Hyodo, “The development of an active noise control system for automobiles,” Tech. Rep., SAE Technical Paper, 1992.
- [119] C. Bohn, A. Cortabarria, V. Härtel, and K. Kowalczyk, “Active control of engine-induced vibrations in automotive vehicles using disturbance observer gain scheduling,” *Control Engineering Practice*, vol. 12, no. 8, pp. 1029–1039, 2004.
- [120] S. J. Elliott and P. A. Nelson, “The active control of sound,” *Electronics & communication engineering journal*, vol. 2, no. 4, pp. 127–136, 1990.
- [121] S. J. Elliott, “A review of active noise and vibration control in road vehicles,” Technical Report 981, ISVR Technical Memorandum, 2008.
- [122] Xuan Li, Shefeng Yan, Xiaochuan Ma, and Chaohuan Hou, “Spherical harmonics MUSIC versus conventional MUSIC,” *Applied Acoustics*, vol. 72, no. 9, pp. 646 – 652, 2011.
- [123] H. Chen, P. Samarasinghe, and T. D. Abhayapala, “In-car noise field analysis and multi-zone noise cancellation quality estimation,” in *Proc. 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2015, pp. 773–778.

- [124] H. Chen, J. Zhang, P. N. Samarasinghe, and T. D. Abhayapala, “Evaluation of spatial active noise cancellation performance using spherical harmonic analysis,” in *Proc. 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2016, pp. 1–5.
- [125] H. Chen, T. D. Abhayapala, and W. Zhang, “Enhanced sound field reproduction within prioritized control region,” in *Proc. Inter.noise 2014*, Nov 2014, p. 596.
- [126] Y. Kajikawa, W. S. Gan, and S. M. Kuo, “Recent advances on active noise control: Open issues and innovative applications,” *APSIPA Transactions on Signal and Information Processing*, vol. 1, pp. 21, Apr 2012.
- [127] S.J. Elliott, P.A. Nelson, I.M. Stothers, and C.C. Boucher, “In-flight experiments on the active control of propeller-induced cabin noise,” *Journal of Sound and Vibration*, vol. 140, no. 2, pp. 219 – 238, 1990.
- [128] J. Cheer, *Active control of the acoustic environment in an automobile cabin*, Ph.D. thesis, University of Southampton, 2012.
- [129] D. P. Das, G. Panda, and S. M. Kuo, “New block filtered-x lms algorithms for active noise control systems,” *IET Signal Processing*, vol. 1, no. 2, pp. 73–81, June 2007.