Theory and Design of Feasible Active Noise Control Systems for 3D Regions

Huiyuan Sun

B.Sc. Engineering (Hons 1)
The Australian National University

January 2022

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

Declaration

The content of this thesis is 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. Sun, T. D. Abhayapala, and P. N. Samarasinghe, "A Realistic Multiple Circular Array System for Active Noise Control Over 3D Space," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 3041 - 3052, 2020.

- H. Sun, T. D. Abhayapala, and P. N. Samarasinghe, "Time Domain Spherical Harmonic Processing with Open Spherical Microphones Recording," *Applied Sciences*, vol. 11, no. 3, pp. 1074, Jan. 2021.

- H. Sun, N. Murata, J. Zhang, T.Magariyachi, T. D. Abhayapala, S. Hayashi and P. N. Samarasinghe, T. Itabashi, "Secondary channel estimation in spatial active noise control systems using a single moving high order microphone" *The Journal of the Acoustical Society of America*, vol. 151, no.3, pp. 1922-1931, March, 2022.

Conference Proceedings

- H. Sun, T. D. Abhayapala and P. N. Samarasinghe, "Time Domain Spherical Harmonic Analysis for Adaptive Noise Cancellation over a Spatial Region,"

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 516-520, Brighton, UK, May, 2019.

- H. Sun, T. D. Abhayapala and P. N. Samarasinghe, "Active Noise Control Over 3D Space with Multiple Circular Arrays," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 135-139, New York, USA, October, 2019.

- H. Sun, J. Zhang, T. D. Abhayapala and P. N. Samarasinghe, "Active noise control over 3D space with remote microphone technique in the wave domain," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October, 2021.

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

Conference Proceedings

- H. Sun, J. Zhang, P. N. Samarasinghe and T. D. Abhayapala, "Spatial Active Noise Control with the Remote Microphone Technique: An approach with a moving higher order microphone," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8707-8711, Singapore, Singapore, May, 2022.

- J. Zhang, H. Sun, P. N. Samarasinghe and T. D. Abhayapala, "Active Noise Control Over Multiple Regions: Performance Analysis," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8409-8413, Barcelona, Spain, May, 2020.

- A. Bastine, T. D. Abhayapala, J. Zhang and H. Sun, "Modelling Room Reverberation Directivity using von Mises-Fisher Mixture Distribution," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 694-700, Auckland, New Zealand, December, 2020.

The research work presented in this thesis has been performed jointly with Prof. Thushara D. Abhayapala, Dr. Prasanga N. Samarasinghe and Dr. Jihui Zhang.

Approximately 75% of this work is my own.

Huiyuan Sun
Research School of Engineering
The Australian National University
Canberra ACT 2601
January 2022

Acknowledgments

Without the support of many colleagues, friends, and other people, 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. Prasanga Samarasinghe, for their professional, patient, unconditional guidance, and consistent encouragement. Their father-like supervision significantly promoted the progress of my growth, both academically and personally. Special thanks go to Prasanga, who had paid great effort in getting me back to campus under the travel restriction of COVID-19 such that I could complete my Ph.D. study on time.
- Dr. Hanchi Chen for convincing me to pursue a Ph.D. degree and leading me onto my academic path.
- Dr. Jihui Zhang for providing suggestions and cooperating on many research problems.
- Naoki Murata for providing an opportunity for me to involve with the collaboration research project with Sony Japan.
- The Australian National University and the Australian Research Council, for the Ph.D. scholarship, the research grant, and the Vice Chancellor's travel grant.
- My fellow students in the Acoustic Audio Signal Processing Group, especially Fei Ma, Huanyu Zuo, Yonggang Hu, and Fahim for sharing their researches and lives all the time.

- My long-time friends, Mengya He, Zhixin Jin, Yuzhen Wang, Haolei Ye, for their continuous friendships and cheering me through the tough times.
- My mother, grandmother, and aunt Fang, for their constant encouragement and inspiration.
- My boyfriend Lachlan Birnie for his kindly accompany, helping in the writing and editing of many papers and helping me out during the busiest days.
- Finally, a Japanese idol group, Arashi, who provided strong emotional and inspirational support, consistently encouraging me to overcome all the failures and rejections throughout my Ph.D..

Abstract

This thesis advances Active Noise Control (ANC) over three-dimensional (3D) space using feasible loudspeaker and microphone array systems. By definition, ANC reduces unwanted acoustic noise by generating an anti-noise signal(s) from secondary loudspeakers. The concept of spatial ANC aims to reduce unwanted acoustic noise over a continuous 3D region, by utilizing multiple microphones and multiple secondary loudspeakers to create a large-sized quiet zone for listeners in three-dimensional space. However, existing spatial ANC techniques are usually impractical and difficult to implement due to their strict hardware requirements and high computation complexity. Therefore, this thesis explores various aspects of spatial ANC, seeking algorithms and techniques to promote the reliability and feasibility of ANC over space in real-life applications.

The spherical harmonic analysis technique is introduced as the basis of conventional spatial ANC systems. This technique provides an accurate representation of a given spatial sound field using higher-order microphone (spherical microphone array) recordings. Hence, the residual noise field in a spatial ANC system can be effectively captured spatially by applying the spherical harmonic technique. Incorporating conventional spatial ANC methods, we developed a series of algorithms and methods that optimize conventional methods regarding array geometries and ANC algorithms, towards improving the feasibility of a conventional spatial ANC system involving the spherical harmonic analysis.

Overall, motivated by feasible and realistic designs for spatial ANC systems, work included in this thesis mainly solves the three problems of: (i) the impracticality of realizing spherical microphone and loudspeaker arrays, (ii) achieving secondary channel estimation with microphones remote from their desired locations, and (iii) unreasonable delays inherent to frequency domain spatial ANC methods. Based on our work, we have stepped towards achieving a spatial ANC system in a real-world environment for people to enjoy silence in the control region with the

reliable usage of resources and algorithms.

Several contributions of this work are: (i) designing a 3D spatial ANC system using multiple circular microphone and loudspeaker arrays instead of spherical arrays, (ii) proposing a 3D spatial ANC method with remote microphone technique such that noise reduction over a region is achieved with microphones remote from the region, (iii) proposing a secondary channel estimation method using a moving higher-order microphone such that usage of an error microphone array is not necessary, (iv) deriving a time domain spherical harmonic analysis method for open spherical microphone array recording with less delay than in the frequency domain, and (v) designing a feed-forward adaptive spatial ANC algorithm incorporating the time domain spherical harmonic analysis technique to better minimize the noise in the region of interest.

Contents

Chapter 1

Introduction

1.1 Motivation and Scope

A wide range of human activities generate perceptible noise. These acoustic noises are usually undesired and uncomfortable for the listener. Excessive amounts of acoustic noise pose a health risk to a majority of the population and it is the most significant single cause of hearing loss (37%) in Australia [1]. The need of eliminating acoustic noises keeps increasing over the years with the development of industries.

There are two categories of methods to reduce excessive noise: passive noise control, and Active Noise Control (ANC) [2]. Passive noise control methods isolate the noise with sound absorbing materials, such as acoustic foam, sponge, or glass-wool [3]. Using a pair of earplugs to enjoy a quiet sleep is a common application of passive noise control. The performance of passive noise control depends on various factors, including the noise frequency, the material type, shape, and thickness. In general, the wavelength of the noise needs to be less than the thickness of the material to achieve an effective reduction of noise. As a result, passive noise control performs better at higher frequencies than lower frequencies.

In contrast, for low frequency noises, an alternative method, named active noise control becomes more effective than passive noise control [2]. In its simplest form, an ANC system aims to reduce the unwanted acoustic noise by generating an anti-noise using loudspeakers which maintains the same magnitude but is 180° out

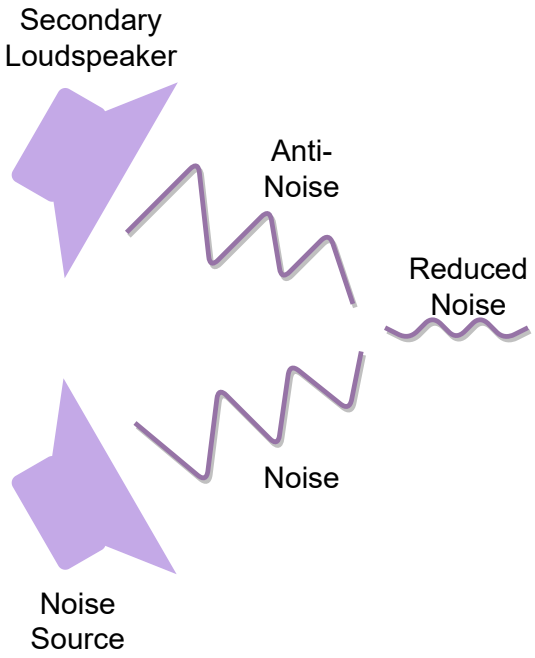


Figure 1.1: An ANC system.

of phase with the noise (Fig. 1.1 [4]). These anti-noise producing loudspeakers are named secondary loudspeakers or secondary sources. The two sound waves corresponding to the noise and the anti-noise cancel each other, resulting in a reduced noise level [5]. ANC is more efficient at lower frequencies, because the wave-length of the noise is longer, which makes it easier for the anti-noise to match with the undesired noise [6].

In recent years, ANC has been applied commercially through ANC headphones [7]. While ANC headphones have gained much popularity among users by providing them with satisfactory noise reduction levels, they are not immune to fundamental limitations. For example, the users are required to constantly wear the headphones, which is inconvenient, potentially uncomfortable, and impractical in some situations. Hence, it is desirable for the noise to be attenuated over a spatial region, such that people inside the region can enjoy a headphone-free quiet environment. The common solution for achieving such a quiet region is via a well developed approach: Multiple-input-multiple-output (MIMO) ANC system. In a MIMO ANC system, multiple secondary loudspeakers are applied to generate an anti-noise field, while

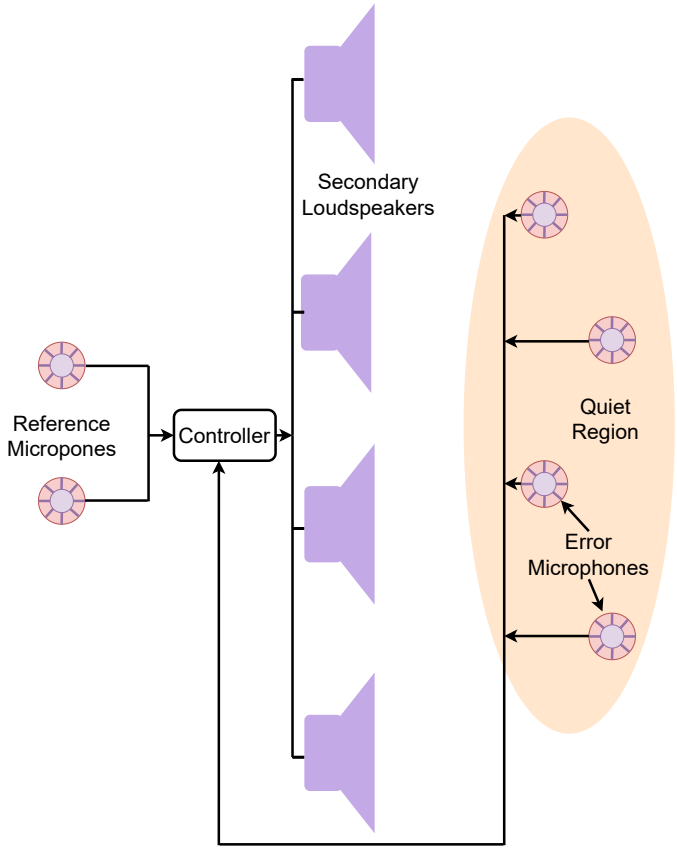


Figure 1.2: A feed-forward MIMO ANC system.

multiple microphones are applied in the region of interest to monitor the residual noise field [2]. For a feed-forward system, reference microphones are applied close to the noise sources to record the noise signals. In a feed-back system, the reference microphones are not necessary [8]. Figure 1.2 [9] shows a feed-forward MIMO ANC system.

MIMO ANC systems successfully reduce the noise in a range of applications including noise control in cars [10,11], in-flight [12,13], through an open window [14], and inside other enclosures [15, 16]. However, this conventional system is only able to minimize the sound pressure at the error microphones because the noise level is only known at these microphones' locations. As a result, when the error microphones are randomly located within a desired quiet region, the significant noise reduction can be only obtained in the proximity of each microphone, leav-

ing the area not covered by microphones to remain noisy. To achieve a uniform noise reduction performance over the region of interest, a large number of error microphones needs to be applied evenly throughout the region. Commercial MIMO ANC systems in cars released recently overcome this problem either by focusing on extremely low frequencies (under 125 Hz) [17], or by using virtual sensing techniques [18]. However, these applications constrain the noise reduction to be at a limited number of discrete points (such as at 2 human ears or a very small region), which limits the movement of listeners, thus limiting their applications in many other real world applications of spatial ANC.

Wave domain signal processing [19] has thus been introduced into ANC systems. The noise can be captured and analysed over space by a microphone array with the spherical harmonic decomposition technique. This technique achieves an accurate recording of the noise field throughout a continuous region. Hence, it is possible for an ANC system to achieve noise reduction spatially over a continuous region, rather than at a number of distributed points [20]. Note here that sound field processing involving wave field synthesis is also called wave domain processing in some papers, in this work, by 'wave domain processing' we strictly mean spherical harmonic based sound field processing. In order to apply spherical harmonic analysis to a spatial ANC system, the error microphones need to be in specific geometries. Conventionally, an evenly distributed spherical microphone array is necessarily placed on the boundary of the region of interest to record the residual sound field inside the region in the wave domain [21, 22]. These error microphone arrays need to be open arrays surrounding the region of interests which is not often practical due to the physical obstruction of entering and exiting the region. Compared to rigid arrays where the microphones are mounted on a rigid spherical baffle, open arrays which are built as cage like structures are also more difficult to design and implement. On top of array complexities, the frequency domain definition of the spherical harmonic analysis also introduces undesirable system delays and adds computational costs to spatial ANC systems.

From the discussion above, we pose the following research problem to address in this thesis:

How to achieve a feasible spatial ANC system for practical applications rather

than theoretical models?

1.2 Problems and Solutions

We breakdown the above research problem into three sub-tasks (the array geometry, the secondary channel estimation, and spatial ANC algorithms) where each subtask optimizes the conventional spatial ANC system from a different aspect and provides intended solutions to each of the following tasks:

(i) Feasible array geometries

Given that the open spherical microphone array is impractical and blocks people entering and exiting the quiet region, an alternative error microphone array geometry for wave domain ANC is necessary to be developed, as well as the secondary loudspeaker array. Hence, we propose a spatial ANC system with multiple circular arrays, where the error microphones and loudspeakers are all located on or above the x-y plane of the region of interest. The circular arrays are more friendly to build and use, and the lower hemi-sphere in the proposed method is left empty for users to enjoy the noise reduction.

Alternatively, virtual sensing techniques can be a useful tool in spatial ANC. These techniques allow the error signals to be estimated at locations that are remote from the physical error microphones [23]. In other words, virtual sensing algorithms are able to reduce the residual noise over the region of interest with some knowledge of systems while physical microphones record error signals away from the region. We develop a wave domain ANC system based on one of the virtual sensing techniques, named remote microphone technique [24], for noise reduction over a spherical region. In this case, we avoided the usage of a spherical error microphone array, or even any error microphones inside the region of interest when the noise is being controlled such that users can move freely in the region.

(ii) Feasible secondary channel estimations

Since typical noise is often time-varying, adaptive filtering is considered to be a beneficial tool in ANC. Adaptive filters typically adjust their coefficients to generate the secondary loudspeaker driving signals such that the signals recorded by the error microphones are minimized as they vary. Hence, the secondary channels, or the impulse responses (IRs) from the secondary loudspeakers to the error microphones,

are required as pre-knowledge before the ANC system begins to work [25]. In task (i), we have proposed the wave domain remote microphone technique to avoid the usage of a spherical error microphone array. However, an impractical error microphone array is still necessary for the secondary channel estimation. To solve this problem, we introduce spatial sound field recording methods with higher order microphones into secondary channel estimations in spatial ANC systems. We also discuss the potential of further replacing these higher order microphones with a single moving higher order microphone.

(iii) Feasbile spatial ANC algorithms

Other than improving the feasibility of the hardware implementation in a spatial ANC system, we also investigate a more practical spatial ANC algorithm. For a feed-forward ANC system, the system delay which includes the filter group delay (signal processing algorithm), the A/D and D/A converters, and the data processing delay, should all be less than the acoustic delay from the reference microphones to the secondary loudspeakers in order to satisfy causality. Conventionally, the spherical harmonic analysis based wave domain spatial ANC system is developed in the frequency domain, where a Fourier transform and an inverse Fourier transform are necessary. These time-frequency domain transforms introduce undesired system delay. We develop a time-wave domain spatial ANC system, where Finite Impulse Response (FIR) filters are designed based on the inverse Fourier transform of the frequency domain spatial functions. Furthermore, we replace these inverse Fourier transform based filters by proposing a pure time domain spherical harmonic decomposition method for sound field recording in a region of interest. The proposed method achieves less latency than the conventional frequency domain methods.

Beyond the three aspects that we address in this thesis, the wave domain spatial ANC system has some other limitations and disadvantages. The major concern is the hardware complexity. The usage of a large amount of transducers (microphones and loudspeakers) leads to a heavy workload in constructing and operating the system. Additionally, the necessary multi-channel audio interface with a higher number of channels is costly. Although there is a few commercially available audio interface, they are not yet enough popular. Nevertheless, the distances between secondary loudspeakers and error microphones are longer in spatial ANC systems

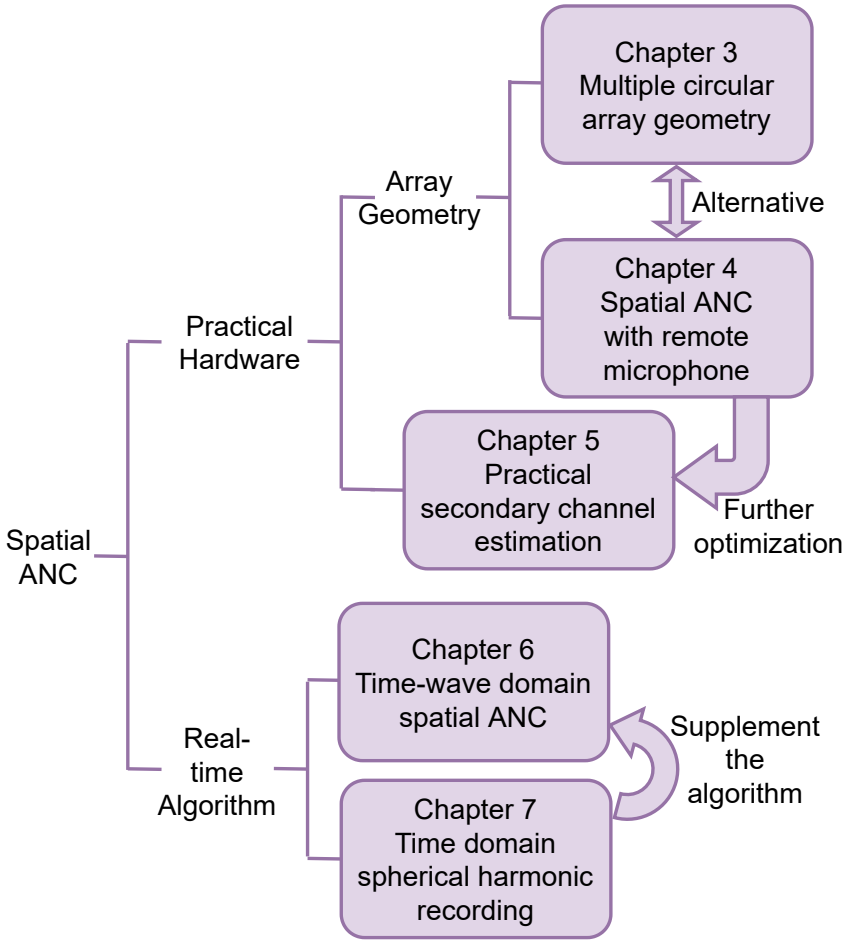


Figure 1.3: Thesis outline.

compared to well-developed ANC headphone models. As a result, the stability and the accuracy of the secondary channel estimation become a harder problem and are still to be solved.

1.3 Thesis Outline

Based on the discussions above, this thesis aims to improve the feasibility of a conventional spatial ANC system. Figure 1.3 presents an overview the thesis outline and the logical flow. The detailed structure of the thesis is presented as follow:

Chapter 2: Literature Review and Background Theory

We present an extensive literature review of past and present spatial ANC methods. We first review the past work on the development and advancement of ANC methods including recent methods. Thereafter, we introduce background knowledge of spherical harmonic based wave domain signal processing along with some helpful mathematical properties. Finally, we introduce basic feed-forward and feed-back adaptive filtering structures for both MIMO ANC systems and wave domain spatial ANC systems. These theories reviewed in this chapter are the fundamental elements for building up this thesis.

Chapter 3: A Realistic Multiple Circular Array System for Active Noise Control over 3D Space

This chapter presents an array geometry optimization method for spatial ANC systems. We first introduce a spatial sound field recording method and a reproduction method using multiple circular arrays instead of the impractical spherical arrays, respectively. Then, we apply these two methods in a feed-forward spatial ANC system. Both the microphone arrays and loudspeaker arrays are located at the upper hemi-sphere of the region of interest, which allows the users to easily enter the quiet region from the lower hemi-sphere. The human head scattering effect is also considered in the proposed method. We construct a second order system in both simulation and experiment to demonstrate the effectiveness of the proposed method on noise reduction over space, and illustrate by experiment that the scattering effect by the human head (user) is negligible. By conducting the experiment in a real-world environment, the proposed method is also proved to be more feasible than conventional methods.

Chapter 4: Active Noise Control over 3D Space with Remote Microphone Technique in the Wave Domain

Conventionally, the error microphone array in a MIMO system or a wave-domain spatial ANC system is not practical. In addition to the method we proposed in chapter 3, in this chapter, we overcome the impractical error microphone problem by introducing virtual sensing techniques to a spatial ANC system. The error signals are estimated at locations that are remote from the physical error microphones. We develop a wave domain ANC system based on this remote microphone technique to achieve noise reduction over a spherical region where all the microphones are located away from the region of interest during the noise reduc-

tion processing. By simulation, we demonstrate that the proposed method can obtain comparable ANC performance to conventional methods, while introducing new freedoms to user movement within the wider region of interest.

Chapter 5: Secondary Channel Estimation in Spatial Active Noise Control Systems Using a Single Moving Higher Order Microphone

In this chapter, we further optimize the array geometry of spatial ANC systems by proposing an alternative method for the secondary channel estimation. In chapter 4, we proposed a method where the impractical error microphone array is avoided when the noise reduction is achieved in the region of interest. However, the system still requires an error microphone array inside the region of interest for the secondary channel estimation. In this chapter, we introduce a spatial sound field recording method using a moving higher order microphone to estimate the secondary channels in a spatial ANC system. With this approach, Impulse Responses (IRs) over the entire region within the desired frequency range can be measured with one higher order microphone, including the required secondary channels. Hence, the usage of the impractical error microphone array can be fully avoided. By simulation, we demonstrate that the proposed method is able to record the desired IRs with tolerable error.

Chapter 6: Active Noise Control over a Spatial Region in the Time Domain

In this chapter, we optimize the adaptive algorithm in a spatial ANC system. Conventionally, the wave domain spatial ANC system is investigated based on the spherical harmonic decomposition method. This method is developed in the frequency domain, where a Fourier transform is necessary before the recorded signal is further processed. This Fourier transform increases the system latency, which is undesirable in an ANC system. We develop a time domain spherical harmonic based spatial ANC system in this chapter. Two time-wave domain feed-forward adaptive filters are designed with different cost functions. These time domain methods do not require a time-frequency domain transform of the recorded signals. Hence, the Fourier transform related buffers are not necessary. The proposed method is also more practical to be applied on an embedded system or other resource limited hardware. Through simulations we show that the proposed method can achieve noise reduction over the control region with less system delay.

Chapter 7: Time Domain Spherical Harmonic Processing with Open Spherical Microphones Recording

We present a time domain spherical harmonic processing approach for open spherical microphone array recording in this chapter. When we were developing the time-wave domain feed-forward adaptive filters for sptial ANC systems in chapter 6, we met the problem that the residual sound field recorded by the error microphones are hard to be analysed in the time-wave domain. The spherical harmonic related functions are developed in the frequency domain. In chapter 6, we avoid this problem by using filters given by the inverse Fourier transform of spherical Bessel functions. In this chapter, we develop the spherical harmonic decomposition algorithm in the time domain, where microphone signals are processed by a series of pre-designed finite impulse response (FIR) filters to obtain a set of time domain spherical harmonic coefficients. These filters do not rely on the inverse Fourier transform. We corroborate the time domain algorithm with a numerical simulation of a fourth order system, and show the proposed recording method to have lower delay than existing approaches, which can be beneficial to a spatial ANC system.

Chapter 8: Conclusions and Future Work

This chapter concludes this thesis, as well as discusses potential future work to further improve the feasibility or the performance of a spatial ANC system.

Chapter 2

Literature review and background

Overview: *This chapter provides a literature review on the development of spatial ANC, where the advances towards problems out of the scope of this thesis are also included. Afterwards, detailed background knowledge on the spherical harmonic decomposition method and its properties are provided. By the end, we introduce existing multi-channel ANC methods and wave domain spatial ANC methods based on the spherical harmonic decomposition.*

2.1 Literature Review: Active Noise Control

Back in the 1930s, ANC was first achieved by a microphone and an electronically driven loudspeaker in [26]. From then on, researchers have shown considerable interests in ANC and have made significant progress. In 1992, Nelson and Elliott published a book [27] detailing the theory of ANC with emphasis on the acoustical aspects. They mentioned that the frequency content, amplitude, and phase of a given noise is often non-stationary as the noise signal and the environment are both time varying. Therefore, adaptive algorithms are necessary in ANC systems to cope with such variations [28]. During the same period, the development of more powerful digital signal processors enabled the low-cost implementation of complex algorithms and encouraged the application of ANC systems based on digital chips [29]. The usage of adaptive filtering was hence possible, which was one of the important foundations to develop the concept of ANC.

Up until now, other than ANC headphones which have become popular in this decade, there are few commercial ANC applications. By the time this thesis is complete, spatial ANC will still be acting as an academic exercise more than something that can be applied in practice. The challenges of the high requirements on resources (loudspeakers and microphones) and computation (Digital Signal Processors) barricade the real-world implementation. However, researchers have been working for years to gradually advance ANC systems from various aspects. We are hoping that there will be a day that other ANC applications, including spatial ANC, can be as practical as ANC headphones such that more people can enjoy the world of quiet.

We now review the relevant literature to have a brief introduction of the past and current development regarding ANC systems. There are mainly two types of adaptive filtering structures for ANC: feed-forward and feed-back. A typical single channel feed-forward adaptive filter requires both a reference microphone to capture the original noise and an error microphone to record the residual noise for generating the driving signal of the secondary loudspeaker [30], which is suitable for both narrow-band and wide-band noise. On the contrary, a feed-back filter does not require a reference microphone, but is only effective with narrow-band noise [31].

While ANC with single channel adaptive filtering using one error microphone and one secondary loudspeaker can only achieve noise reduction in narrow ducts or at one point, researchers are not satisfied with it. To achieve ANC for more complex enclosures or larger areas, algorithms using multiple microphones and loudspeakers (MIMO ANC) are developed [2]. To improve the robustness and the performance of MIMO ANC systems, other than the conventional FxLMS method [32], Normalized LMS [33], adjoint LMS [34], and recursive LMS [35] are developed for more stable and effective filtering. An Eigenvalue equalization FxLMS algorithm is developed in [36] to increase the convergence speed, while in [37] and [38], a fast affine projection algorithm is proposed, which provides a better convergence speed and steady-state control performance in adaptive systems. With these optimizations of MIMO systems, successful applications of ANC have been developed including noise control in cars [10], [11], [39], in flights [12], [13], in rooms [40], through an open window [14], and inside other enclosures [15], [16].

Beyond the investigation of adaptive algorithms, researchers are also working on advancing the ANC system from other aspects. Analysis on the acoustic energy density provides a more robust algorithm against number and location of error microphones [41], [42]. The sparsity of the noise signal is considered in [43] to further increase the robustness of convergence in reverberant environments. A source localization technique is added in reference signals recording in [44] to separate the target noise and the disturbance noise such that a better noise reduction performance is achieved. A neural-network based training process with particle swarm optimization is investigated to accelerate the convergence speed [45] and improve the scalability [46]. A selective fixed-filter ANC method based on convolutional neural network is addressed in [47]. The estimation of secondary channels is another important topic where the accuracy of the estimation highly influences the stability of the system [48]. On-line secondary channel estimation approach is proposed in [49] and analysed in [50]. In [51], sparsity based nonlinear modeling is given to reduce the computational complexity in a complex environment. Moreover, in a feed-forward adaptive filtering system, the feedback from the secondary loudspeakers to the reference microphones is settled in [52]. Later, this effect has been push to online processing in [53], [54], and decoupled from the feed-forward system in [55].

Generally, noise reduction can only be achieved around an error microphone with the spatial limit being approximately $\lambda/10$, where λ is the wavelength of the undesired noise [56]. Hence, to achieve noise reduction over a large region, we need a large amount of microphones to cover the entire region of interest, which is difficult and costly [57].

One of the solutions for these impractical microphone arrays is virtual sensing techniques [58]. These virtual sensing algorithms estimate the error signals at the remote virtual locations using the signals obtained at physical microphones, and the control signals and knowledge of the system. Instead of minimising the physical error signals, the estimated error signals are minimised at the virtual locations [23]. These techniques are widely applied to MIMO ANC in enclosures [59], [60], in mobile phones [61], in vehicles [11], [62] and proved to be especially beneficial in head-sets [63], [64], [65], [66]. In [67] and [68], the causality and robustness of two different virtual sensing techniques are optimized, respectively, while the guidance

of choosing a suitable method is provided in [60]. Other than these methods, non-acoustic sensors are applied in virtual sensing ANC systems to provide more feasible design [69], [70].

Although virtual sensing techniques improve the feasibility of ANC over a region, the analysis of the sound field is still point based. In other words, the noise reduction is achieved at multiple points of interest instead of a continuous area. Wave field synthesis [71] has been introduced into ANC systems in [72], [73]. This technique not only provides the spatial and temporal information of the noise field, but also is proved to be more efficient for non-stationary noises [74] and impulsive noises [75]. Additionally, Kernel interpolation has been applied to improve the performance of noise reduction [76], [77].

Spherical harmonic analysis [19] is considered to be another useful tool for spatial ANC. It can provide a more insightful and efficient spatial sound field recording [21], [78], and reproduction [79], [80] using discrete arrays. A wide range of applications related to the spatial sound field processing have been developed based on this wave domain technique including echo cancellation [81], room equalization for sound field reproductions [82], [83], higher order loudspeaker array design [84], [85], source separation [86] and personal audio systems [87], [88]. Spherical harmonics also show its efficiency in spatial ANC systems for noise reduction of both outgoing sound fields [89] and incoming noise fields [90], [91], [92], and is optimized regarding computational complexity recently [93]. The spherical harmonic based spatial ANC algorithm in [90] is considered to be the fundamental work of this thesis. In this thesis, we provide multi-faceted optimization of the spherical harmonic based spatial ANC system in [90] for a more feasible design.

Conventionally, wave domain sound field recording and reproduction approaches are developed using spherical arrays, which is not practical in most of ANC real-world applications. In [94], deep learning is applied for spherical harmonic based sound field recording for less microphone usage. In [95], [96], [97], [98], [99], several approaches are provided for wave domain sound field reproduction using circular or irregular loudspeaker arrays. On the other hand, [100], [101], [102] provide the sound field recording methods with non-spherical microphone arrays. From these papers, we can see a potential to optimize the array geometry in a wave domain spatial ANC system.

ANC systems have another important requirement regarding system delay [2]. This is because the system delay including the filter group delay (signal processing algorithm), the A/D and D/A converter, and the data processing delay, are required to be less than the acoustic delay from the reference microphones to the secondary loudspeakers in order to satisfy causality [27]. Moreover, ANC systems, especially ANC systems employing adaptive filtering, are usually real-time processing, where lower latency is always desired for time-varying noises. In [103], [104], sound field reproduction over space is achieved by the wave field synthesis in the time domain. The time domain spherical harmonic descriptions are provided in [105], [106], which have been applied in beam-forming [107], and scattering problems [108], [109]. These researches show that the time domain processing of the spatial sound field is often computationally efficient and thus benefits real-time applications, including spatial ANC.

2.2 Background: Spherical harmonic Properties

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 are given by the fundamental solutions of the Helmholtz wave equations [19]. In general, there are two classes: the interior sound field expression, and the exterior field expression. The interior sound field expression describes the sound field within a spatial region with absent of sound sources inside; while the exterior expression is for the sound sources are within a limited area, and the region of interest is the source-free space enclosing the source area. We only focus on the interior expression in this thesis.

We start with defining the coordinate system used in this thesis to represent the spatial attributes of a given sound field. In 3-D space, it is convenient to use a spherical coordinate system as shown in Figure 2.1. The spherical coordinates of a point in the space can be represented in terms of the position vector $\boldsymbol{x} = (r, \theta, \phi)$ where $r = \|\boldsymbol{x}\|$ denotes the distance from the origin with $\|\cdot\|$ implying the Euclidean distance, θ denotes the polar angle from the vertical axis with $0 \leq \theta \leq \pi$, and ϕ denotes the azimuths angle in the horizontal plane containing the origin with $0 \leq \phi \leq 2\pi$.

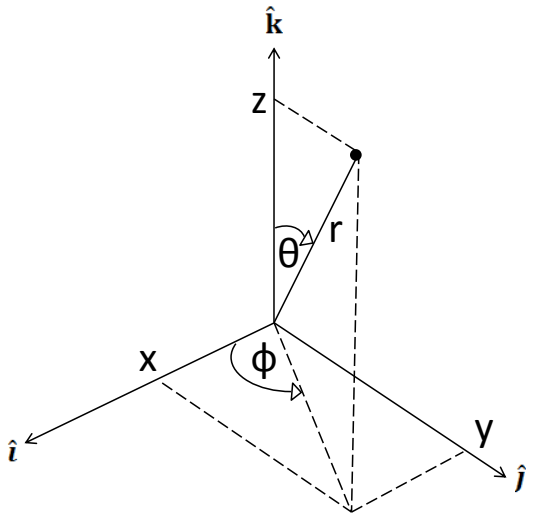


Figure 2.1: Representation of a position vector $\boldsymbol{x}$ in the 3-D spherical coordinate system.

2.2.1 Spherical harmonic expansion

Assuming a sound field within a source-free region Ω of radius r_Ω, the sound pressure at a point $\boldsymbol{x} = (r, \theta, \phi)$ inside the region with respect to the origin Ω_O can be expressed as [19]

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{2.1}$$

where order n and mode m are integers, $\alpha_{nm}(k)$ is a set of spherical harmonic coefficients representing the sound field inside Ω, $k = 2\pi f/c$ is the wave number, f is the frequency of the sound signal, c is the speed of sound, $j_n(kr)$ is the n-th order spherical Bessel function of the first kind [110], $Y_{nm}(\theta, \phi)$ is the spherical harmonic function, defined by

$$Y_{nm}(\theta, \phi) = \mathcal{P}_{nm}(\cos\theta) E_m(\phi), \tag{2.2}$$

where

$$\mathcal{P}_{nm}(\cos\theta) = \sqrt{\frac{2n+1}{2}} \sqrt{\frac{(n-|m|)!}{(n+|m|)!}} P_{n|m|}(\cos\theta) \tag{2.3a}$$

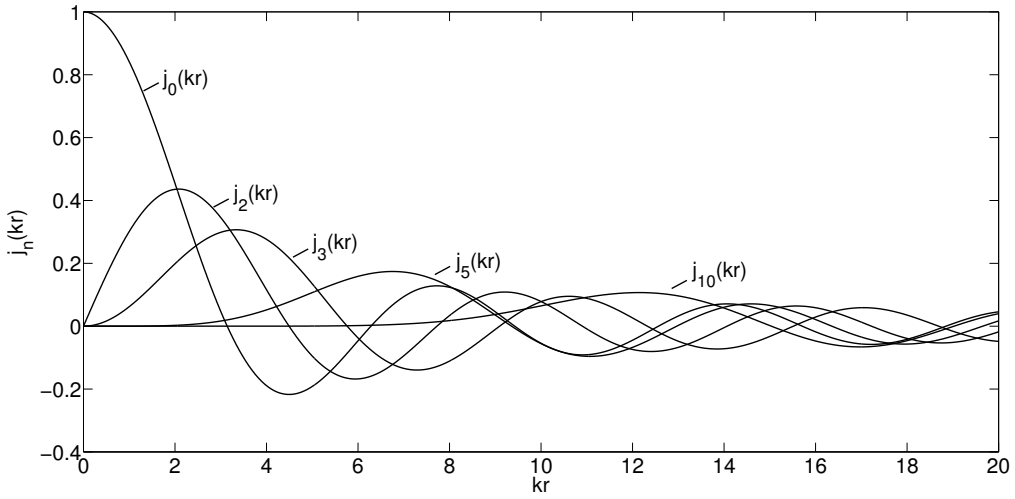


Figure 2.2: The spherical Bessel function $j_n(kr)$ for $n = 0, 1, 2, 5$ and 10.

$$E_m(\phi) = \sqrt{(1/2\pi)}e^{im\phi}, \tag{2.3b}$$

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

The spherical harmonic functions have the important property of orthogonality that

$$\int_0^{\pi}\int_0^{2\pi} Y_{nm}(\theta,\phi)Y^*_{n'm'}(\theta,\phi)\sin(\theta)d\theta d\phi = \delta_{n-n',m-m'}, \tag{2.4}$$

where $(\cdot)^*$ refers to the conjugation operator, and $\delta_{n-n',m-m'}$ is the two dimensional Dirac Delta function.

In (2.1), the summation of order up to an infinite order is not practical. Hence, a truncation of this summation at a maximum order N is common, due to the low pass nature of the spherical Bessel functions, as shown in Fig. 2.2. Thus, (2.1) can be approximated as

$$z(\boldsymbol{x},k) \approx \sum_{n=0}^{N}\sum_{m=-n}^{n} \alpha_{nm}(k)j_n(kr)Y_{nm}(\theta,\phi), \tag{2.5}$$

where $N = \lceil kr \rceil$ [80].

In the free space, the Green's function of a point source at $\boldsymbol{x}_s$ can be written

by

$$g(\boldsymbol{x}|\boldsymbol{x}_s, k) = \frac{e^{(ik|\boldsymbol{x}-\boldsymbol{x}_s|)}}{4\pi|\boldsymbol{x}-\boldsymbol{x}_s|}, \tag{2.6}$$

which can be expanded into the wave domain with (2.5) as

$$g(\boldsymbol{x}|\boldsymbol{x}_s, k) = ik\sum_{n=0}^{N}\sum_{m=-n}^{n} h_n(kr_s)Y_{nm}^*(\theta_s, \phi_s)j_n(kr)Y_{nm}(\theta, \phi), r_s > r, \tag{2.7}$$

where $h_n(\cdot)$ refers to the spherical Hankel function of the n-th order.

Therefore, there is an analytical solution of the spherical harmonic coefficients of Green's function in the free space that

$$\alpha_{nm}^{(g)}(k) = ikh_n(kr_s)Y_{nm}^*(\theta_s, \phi_s). \tag{2.8}$$

2.2.2 Properties of the associated Legendre function

We list some useful properties of the associated Legendre function in this section. When $\theta = \pi/2$, we have

$$\mathcal{P}_{n|m|}(\cos\theta) = \mathcal{P}_{n|m|}(0)\begin{cases} = 0 & n+|m| \text{ is odd,} \\ \neq 0 & n+|m| \text{ is even,} \end{cases} \tag{2.9}$$

while when $\theta = 0$, we have

$$\mathcal{P}_{nm}(\cos\theta) = \mathcal{P}_{nm}(1)\begin{cases} = 0 & m \neq 0 \\ \neq 0 & m = 0 \end{cases} \tag{2.10}$$

A recurrent relationship between the associated Legendre function and its first order derivative is given by [111]

$$(x^2-1)\frac{dP_{n|m|}(x)}{dx} = nxP_{n|m|}(x) - (|m|+n)P_{(n-1)|m|}(x). \tag{2.11}$$

When $x = 0$, (2.11) can be simplified to

$$P'_{n|m|}(0) = (|m|+n)P_{(n-1)|m|}(0). \tag{2.12}$$

A similar equation can be derived for the normalized associated Legendre function that

$$\mathcal{P}'_{n|m|}(0) = \sqrt{\frac{(2n+1)(n^2-m^2)}{(2n-1)}}\mathcal{P}_{(n-1)|m|}(0). \tag{2.13}$$

Moreover, we have the Fourier relationship between the spherical Bessel function $j_n(kr)$ and the Legendre function $P_n(x)$ given by [112]

$$\int_{-\infty}^{\infty} e^{ikrx} j_n(kr)dk = \pi i^n P_n(x). \tag{2.14}$$

These properties helps us to develop algorithms and design the array geometry in the following chapters.

2.2.3 Addition theorem

The addition theorem describes the relationship of spherical harmonic coefficients with respects to two different coordinate systems. This enables the transformation in the wave domain decomposed sound fields, such that the obtained spherical harmonic coefficients can be represented with respects to desired locations. The derivation of 3-D addition theorems uses the Hobson's theorem, the Erdelyi operator and the linearisation formula for spherical harmonics, as given in [113], [114]. While there exist a number of different addition theorems derived in the above manner, we only discuss the ones that are useful in the current context, namely the addition theorems for Bessel functions and the addition theorems for Hankel functions for an interior field.

Consider two origins O and O_1. Let $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ be the position vectors of a general point $\boldsymbol{S}$ with respect to O and O_1 respectively. Let $\boldsymbol{x}_2$ represent the position vector of O_1 with respect to O, so that $\boldsymbol{x}_0 = \boldsymbol{x}_1 + \boldsymbol{x}_2$. See Figure 2.3 for a sketch of the geometry. Let $\boldsymbol{x}_0 = (r_0, \theta_0, \phi_0)$, $\boldsymbol{x}_1 = (r_1, \theta_1, \phi_1)$ and $\boldsymbol{x}_2 = (r_2, \theta_2, \phi_2)$ in spherical coordinates. The addition theorem for spherical Bessel functions is then given by [113]

$$j_n(kr_0)Y_{nm}(\theta_0, \phi_0) = \sum_{\nu=0}^{\infty}\sum_{\mu=-\nu}^{\nu} T_{n\nu}^{m\mu}(\boldsymbol{x}_2) j_\nu(kr_1) Y_{\nu\mu}(\theta_1, \phi_1) \tag{2.15}$$

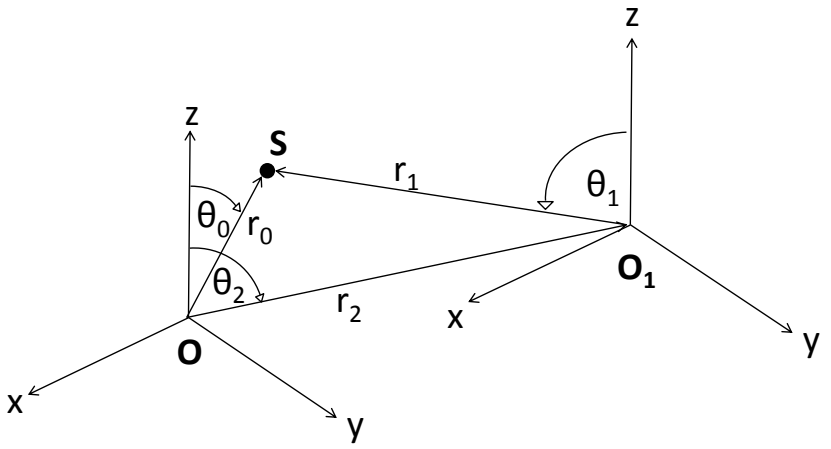


Figure 2.3: Geometry for three dimensional addition theorems.

where ν and μ denote higher order indices and

$$\begin{aligned} T_{n\nu}^{m\mu}(\boldsymbol{x}_2) =& 4\pi i^{\nu-n} \sum_{u=0}^{n+\nu} i^u (-1)^{2m-\mu} \sqrt{\frac{(2n+1)(2\nu+1)(2u+1)}{4\pi}} \\ & \times j_u(kr_2) Y_{u(\mu-m)}^*(\theta_2, \phi_2) W_1 W_2, \end{aligned} \tag{2.16}$$

with

$$W1 = \begin{pmatrix} n & \nu & u \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \qquad W2 = \begin{pmatrix} n & \nu & u \\ m & -\mu & (\mu - m) \end{pmatrix} \tag{2.17}$$

denoting Wigner $3-j$ symbols [113].

2.2.4 Real-valued spherical harmonics

Real valued functions are necessary in many applications of spherical harmonic analysis. The real-valued spherical harmonic is defined as [115]

$$\begin{aligned} Y_{nm}(\theta, \phi) =& (-1)^{|m|} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} \\ & \times \begin{cases} P_{nm}(\cos\theta)\cos(m\phi) & m \geq 0 \\ P_{nm}(\cos\theta)\sin(m\phi) & m < 0 \end{cases}, \end{aligned} \tag{2.18}$$

which has the orthogonality property of

$$\int_0^{2\pi}\int_0^{\pi} Y_{nm}(\theta,\phi)Y_{n'm'}(\theta,\phi)\sin\theta d\theta d\phi = \delta_{nn'}\delta_{mm'}. \tag{2.19}$$

The complex value spherical harmonic functions are widely applied in the frequency domain signal processing. In the time domain, sound pressures are usually real-valued. Hence, it is preferable to use real-valued spherical harmonic functions for the time domain analysis. In this thesis, we use the complex valued spherical harmonic function in most chapters unless explicit mentioned.

2.2.5 Spatial sound field recording with spherical microphone arrays

The conventional way of capturing the spherical harmonic coefficients $\alpha_{nm}(k)$ is by applying a spherical microphone array surrounding the region. The methods with open and rigid spherical microphone arrays for interior sound field recording are given in [21] and [116], respectively.

For an open spherical microphone array of radius $R_Q > R_\Omega$, $\alpha_{nm}(k)$ can be extracted by integrating the recording over the spherical surface while exploiting the orthogonality property of $Y_{nm}(\cdot)$, which gives

$$\alpha_{nm}(k) = \frac{1}{j_n(kR_Q)}\int_0^{2\pi}\int_0^{\pi} z(R_Q,\theta,\phi,k)Y_{nm}^*(\theta,\phi)\sin\theta d\theta d\phi. \tag{2.20}$$

In practice, this integration is achieved using an equivalent discrete summation of spatial samples over the sphere, where at least $Q = (N+1)^2$ samples are necessary. We rewrite (2.20) as

$$\alpha_{nm}(k) \approx \frac{1}{j_n(kR_Q)}\sum_{q=1}^{Q} z(R_Q,\theta_q,\phi_q,k)Y_{nm}^*(\theta_q,\phi_q)\chi(q), \tag{2.21}$$

where θ_q, ϕ_q are the elevation and azimuth angle of the q-th microphone, respectively, and $\chi(q)$ is the weighting coefficients specific to the sampling scheme of the microphone array. For regularly distributed sampling, $\chi(q) = 1$ for all the

microphone samples.

In the case of the rigid spherical array, the sound pressure on the surface of the rigid baffle is derived with

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) b_n(kr) Y_{nm}(\theta, \phi), \tag{2.22}$$

where

$$b_n(kr) = j_n(kr) - \frac{j_n'(kr_0)}{h_n'(kr_0)} h_n(kr) \tag{2.23}$$

where r_0 ($r_0 \leq R_\Omega$) denoting the radius of the rigid spherical baffle, $h_n(\cdot)$ is the spatial Hankel function, $(\cdot)'$ refers to the first order derivative. Using the same integration method, the spherical harmonic coefficients of a region recorded by a rigid array can be calculated by

$$\alpha_{nm}(k) \approx \frac{1}{b_n(kR_Q)} \sum_{q=1}^{Q} z(R_Q, \theta_q, \phi_q, k) Y_{nm}^*(\theta_q, \phi_q) \chi(q). \tag{2.24}$$

Comparing with the open spherical array, the rigid array provides more robustness on implementation. Additionally, rigid array avoids potential ill-conditioning problem caused by $j_n(kr) = 0$ at certain combinations of k and r by using $1/b_n(kR_Q)$ in (2.24) instead of $1/j_n(kR_Q)$ in (2.21). However, the rigid baffle completely encloses the region of interest, rendering this array hard to implement in a large area, and barricades its application in spatial ANC systems. In the spatial ANC system involving wave-domain processing, open spherical microphone arrays are commonly used, which will be introduced in the next section.

2.3 Background: Active Noise Control

After reviewing the spherical harmonic analysis for sound field recording and analysis in the wave domain, we now move to the introduction of basic ANC structures which provide us conventional solutions of noise reduction. As we mentioned in Sec. 2.1, we use feed-forward adaptive filtering in this thesis for its benefits on wide-band noise reduction performance.

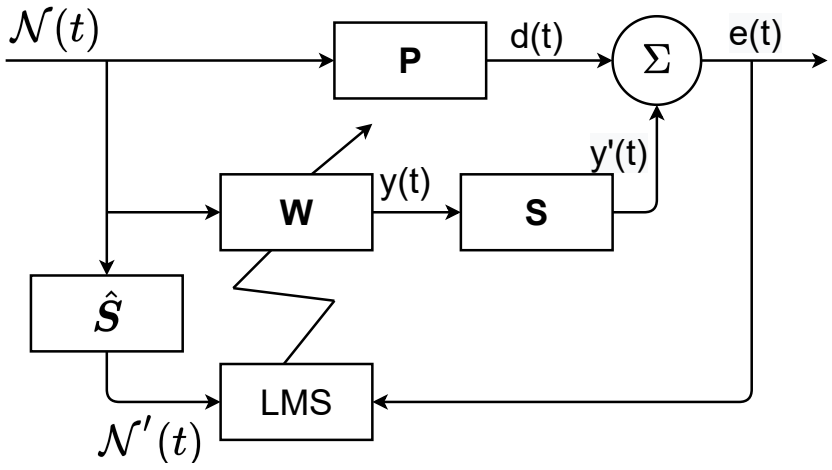


Figure 2.4: Block diagram of a Fx-LMS feed-forward ANC system.

2.3.1 Single channel Feed-forward system

A feed-forward adaptive filtering system for ANC contains an error microphone, a reference microphone, a secondary source, and an adaptive algorithm. We use the Fx-LMS algorithm to implement the system. The block diagram is shown in Fig. 2.4. In the figure, $\boldsymbol{P}$ represents the primary path, which is the acoustic impulse responses from the reference sensor to error microphone. Secondary channel and its estimation are given by $\boldsymbol{S}$ and $\widehat{\boldsymbol{S}}$, respectively. The reference input signal is $\mathcal{N}(t)$, and the instantaneous error microphone measurement is $e(t)$.

In this system, the controller is the adaptive filter using the LMS algorithm. The adaptive filter $\boldsymbol{W}$ continuously tracks variations in the primary noise field. The error signal recorded by the error microphone is given by

$$e(t) = d(t) + y'(t), \tag{2.25}$$

where t refers to the time interval. In general, this t is given by discrete time sampling. $y'(t)$ is the anti-noise at the error microphone generated by the secondary loudspeaker, which is given by

$$y'(t) = y(t) * \boldsymbol{S}, \tag{2.26}$$

where $y(t) = \mathcal{N}(t) * \boldsymbol{W}(t)$ is the driving signal of the loudspeaker, $*$ denotes linear convolution, $\boldsymbol{S}$ refers to the secondary channel from the secondary loudspeaker to the error microphone. $\boldsymbol{W}(t) = [w_0(t), w_1(t), \cdots, w_{L_w}(t)]$ is the adaptive filter weight coefficients with filter length L_w at the time interval t, and $\mathcal{N}(t)$ is a vector containing the reference noise recorded by the reference microphone with a length of L_w.

Minimizing the mean square value of the error signal $e(t)$, the cost function is

$$\xi(t) = e(t)^2. \tag{2.27}$$

The update equation is hence derived based on the steepest descent method, which is

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) - \frac{\lambda}{2}\nabla\xi(t), \tag{2.28}$$

where λ is the step size, and ∇ is the gradient operator.

Taking the derivative of ξ with respect to $\boldsymbol{W}(t)$, we can derive

$$\begin{aligned} \nabla\xi(t) &= \nabla(e(t)^2) \\ &= 2e(t)\nabla e(t) \\ &= 2e(t)\mathcal{N}(t) * \widehat{\boldsymbol{S}}, \end{aligned} \tag{2.29}$$

where $\widehat{\boldsymbol{S}}$ is the estimated value of the secondary channel $\boldsymbol{S}$, which is usually obtained by taking an impulse response measurement from the secondary loudspeaker to the error microphone.

Substituting (2.29) to (2.28), we derive the final update equation of the Fx-LMS feed-forward adaptive filtering algorithm as:

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) - \lambda e(t)\mathcal{N}(t) * \widehat{\boldsymbol{S}}. \tag{2.30}$$

We observe from (2.30) that the filter coefficients $\boldsymbol{W}$ in the next iteration are related to the error signal and the filtered reference signal vector at current iteration.

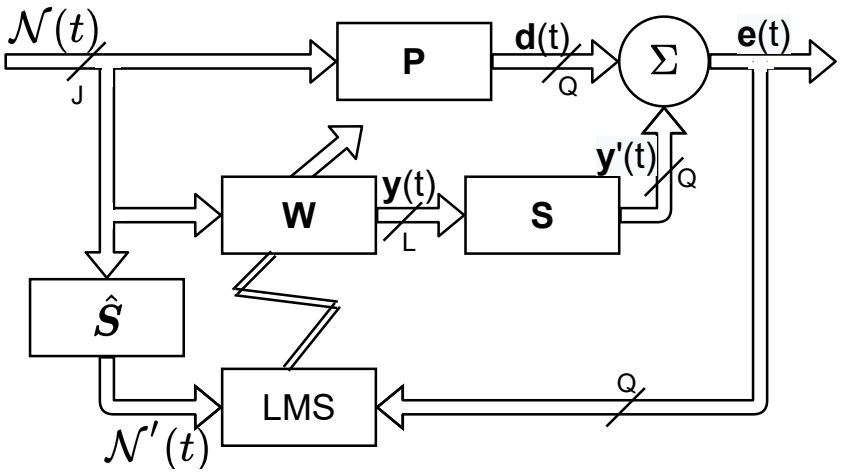


Figure 2.5: Block diagram of a multi-channel Fx-LMS feed-forward ANC system.

2.3.2 Multi-channel Feed-forward system

After the discussion for signal channel feed-forward adaptive filtering algorithm, which is only able to reduce the noise at a single point, we now move on to the multi-channel case, where multiple error microphones and secondary loudspeakers are employed in the system [35].

A block diagram of the commonly used Fx-LMS multi-channel algorithm is shown in Fig. 2.5. In the figure, $\boldsymbol{P}$ are the primary paths, which are the acoustic impulse responses from the reference sensors to error microphones. Secondary channels and their estimation are given by $\boldsymbol{S}$ and $\widehat{\boldsymbol{S}}$, respectively. The reference input signals are $\mathcal{N}_j(t)$, $j \in J$, and the instantaneous error microphone measurements are $e_q(t)$ $q \in Q$, where J and Q is the number of reference microphones and error microphones, respectively. The number of secondary loudspeakers is given by L.

The error signal at the q-th error microphone is given by

$$e_q(t) = d_q(t) + y'_q(t), \tag{2.31}$$

where $d_q(t)$ is the primary noise at the q-th microphone, $y'_q(t)$ is the secondary

field, which can be written as

$$y'_q(t) = \sum_{\ell=1}^{L} y_\ell(t) * \boldsymbol{S}_{\ell q}(t), \tag{2.32}$$

where $\boldsymbol{S}_{\ell q}(t)$ is the impulse response from the ℓ-th secondary loudspeaker to the q-th error microphone, $y_\ell(t)$ is the driving signal of the ℓ-th secondary loudspeaker. This driving signal is generated as

$$y_\ell(t) = \sum_{j=1}^{J} \boldsymbol{w}_{j\ell}^T(t)\mathcal{N}_j(t), \tag{2.33}$$

for $\ell \in L$, where $\boldsymbol{w}_{j\ell}(t)$ are the adaptive filter coefficients of the ℓ-th loudspeaker at the t iteration, $\mathcal{N}_j(t)$ is the reference signal at the j-th reference microphone with the same length of $\boldsymbol{w}_{j\ell}(t)$.

For a multi-channel feed-forward ANC system, the cost function is given by the sum of mean squared error signals:

$$\xi(t) = \sum_{q=1}^{Q} |e_q(t)|^2. \tag{2.34}$$

Therefore, we write the update equation of the multi-channel feed-forward algorithm for $\boldsymbol{w}_{j\ell}(t)$ as

$$\boldsymbol{w}_{j\ell}(t+1) = \boldsymbol{w}_{j\ell}(t) - \lambda \sum_{q=1}^{Q} e_q(t)\mathcal{N}_j(t) * \widehat{S}_{\ell q}(t). \tag{2.35}$$

Comparing with the single channel ANC system, multi-channel systems provide better noise control over multiple points, which enable the application of spatial ANC. The computation cost is also significantly increased. However, the noise reduction of multi-channel system is limited by the position of error microphones, which is inconvenient to be implemented in real life applications.

2.3.3 Wave domain ANC

We introduce the wave domain approach for spatial ANC in this section. Being different from the conventional multi-channel ANC system, the wave domain ANC system is developed based on the spherical harmonic analysis, hence allows the noise reduction over a continuous region instead of discrete points. Several kinds of spatial ANC algorithms are developed in the wave domain. We only detail one of them in this thesis, which is also used as the conventional method to compare with in the following chapters.

In [90], the conventional wave domain spatial ANC system is proposed in 2-D space. In this thesis, we directly introduce the corresponding spatial ANC system in 3D space. We first transform necessary components in multi-channel ANC system into the wave domain using spherical harmonic based wave equation that we have addressed in Sec. 2.2, and then introduce the wave domain feed-forward ANC system structure. To introduce the wave domain processing of the sound field, spherical arrays of the error microphone and the secondary loudspeaker are used. The array geometry of a wave domain spatial ANC system in 3-D space is shown in Fig.2.6. We use single reference microphone as an example such that it is clearer to focus on the wave domain processing. The wave domain algorithm can be also applied for multiple reference microphones.

We rewrite (2.31) in the frequency domain that the residual noise at the q-th microphone located at $\boldsymbol{x}_q$ as

$$e_q(\boldsymbol{x}_q, k) = d_q(\boldsymbol{x}_q, k) + y'_q(\boldsymbol{x}_q, k), \tag{2.36}$$

where $k = 2\pi f/c$ is the wave number.

Consider a source free region Ω with a radius of r_Ω. As described in (2.21), if we have Q microphones evenly distributed over a sphere with radius $r_q \geq r_\Omega$, the spherical harmonic coefficients in the region of interest Ω can be express as

$$\alpha_{nm}^{(e)}(k) \approx \frac{1}{j_n(kr_q)} \sum_{q=1}^{Q} e(\boldsymbol{x}_q, k) Y_{nm}^*(\theta_q, \phi_q) \chi(q), \tag{2.37}$$

where at least $Q \geq (N+1)^2$ microphones are required. $N = \lceil k_{max} r_\Omega \rceil$ is the

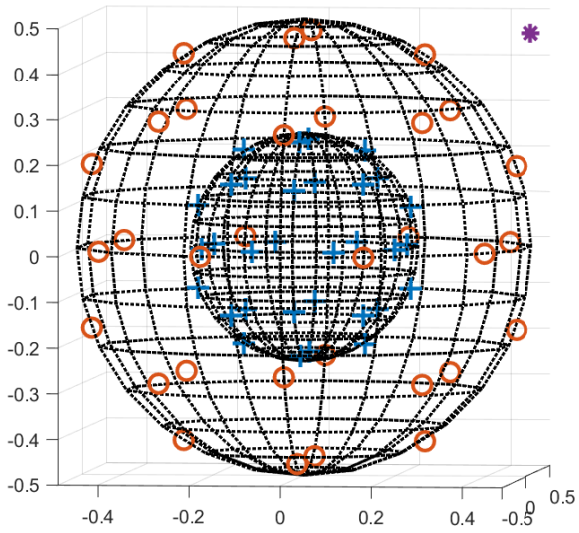


Figure 2.6: A spatial ANC system in 3D space with a spherical error microphone array in '+' (blue); a spherical secondary loudspeaker array in 'o' (orange); and a noise source/reference microphone in '*' (purple).

maximum order that the system is recording, k_{max} is the maximum noise frequency that we are targeting.

These wave domain coefficients of the residual noise can be written as the summation of the corresponding coefficients for the noise field and the secondary sound field, given by

$$\alpha_{nm}^{(e)}(k) = \alpha_{nm}^{(d)}(k) + \alpha_{nm}^{(s)}(k). \tag{2.38}$$

In the wave domain ANC system, the adaptive algorithm minimises the instantaneous squared error of the residual wave domain coefficients, such that the cost function at the t-th iteration becomes

$$\xi(k,t) = \sum_{n=0}^{N_{max}} \sum_{m=-n}^{n} |\alpha_{nm}^{(e)}(k,t)|^2 = \boldsymbol{\alpha}^{(e)}(k)^H \boldsymbol{\alpha}^{(e)}(k). \tag{2.39}$$

where $\boldsymbol{\alpha}^{(e)}(k) = [\alpha_{00}^{(e)}(k), \alpha_{1-1}^{(e)}(k), \cdots, \alpha_{NN}^{(e)}(k)]$ is a vector contains all the spherical harmonic coefficients of the residual noise field, $(\cdot)^H$ denotes the conjugate

transpose.

Given the update equation in (2.28) that

$$\boldsymbol{W}_\ell(k,t+1) = \boldsymbol{W}_\ell(k,t) - \frac{\lambda}{2}\nabla\xi(k,t), \tag{2.40}$$

we take the derivative of $\xi(k,t)$ with respect to $\boldsymbol{W}$, which gives

$$\begin{aligned}\nabla\xi(k,t) &= \nabla(\boldsymbol{\alpha}^{(e)}(k)^H\boldsymbol{\alpha}^{(e)}(k)) \\ &= 2\mathcal{N}'_\ell(k)^H\boldsymbol{\alpha}^{(e)}(k),\end{aligned} \tag{2.41}$$

where $\mathcal{N}'_\ell(k)$ is the vector of the spherical harmonic coefficients of the filtered reference signal, with each of the elements given by

$$\mathcal{N}'_{\ell nm}(k) = \mathcal{N}(k)\widehat{\boldsymbol{S}}_{\ell nm}(k), \tag{2.42}$$

where $\widehat{\boldsymbol{S}}_{\ell nm}(k)$ is the spherical harmonic coefficients of the ℓ-th loudspeaker's secondary channel, which can be pre-measured and calculated by measuring the impulse responses from the secondary loudspeaker to the error microphone array.

Hence, the update equation is given by

$$\boldsymbol{W}_\ell(k,t+1) = \boldsymbol{W}_\ell(k,t) - \lambda\mathcal{N}'_\ell(k)^H\boldsymbol{\alpha}^{(e)}(k). \tag{2.43}$$

By replacing the LMS filter by the normalized LMS filter, the final update equation of the normalized wave domain feed-forward algorithm can be written as

$$\boldsymbol{W}_\ell(k,t+1) = \boldsymbol{W}_\ell(k,t) - \frac{\lambda_0}{||\mathcal{N}'_\ell(k)^H||^2}\mathcal{N}'_\ell(k)^H\boldsymbol{\alpha}^{(e)}(k), \tag{2.44}$$

where $\lambda_0 \in [0,1]$ is the normalized step size.

2.4 Summary

In this chapter, we reviewed the literature regarding developments and advances of spatial ANC systems over the past years, and gave a brief introduction to sound field properties in the wave domain, as well as conventional ANC feed-forward

adaptive algorithms. We started with the conventional ANC method of the single channel and multi-channels, and then introduced the wave domain approach in spatial ANC systems. Standing on top of other researchers work, and based on the conventional wave domain ANC algorithm, we propose multiple novel methods and approaches from different aspects to improve the feasibility of the spatial ANC system in Chapters 3, 4, 5, and 6.

Chapter 3

A realistic multiple circular array system for ANC over 3D space

Overview: *Conventionally, spatial ANC systems are proposed using point based multi-channel systems and recently wave-domain methods have been developed using spherical harmonic analysis of spatial sound fields, as introduced in Chapter 2. A major limitation for implementing the latter approach is the requirement of regularly distributed microphones and loudspeakers over spherical surfaces. In this chapter,the above constraint is relaxed by constructing a system utilizing multiple circular microphone and loudspeaker arrays and designing a corresponding ANC algorithm. By simulation, we show that the proposed method can achieve comparable ANC performance to conventional spherical array methods. By experiment, we demonstrate the feasibility of implementing the multiple circular array structure and verify its effectiveness given practical constraints.*

3.1 Introduction

As reviewed in Sec. 2.3, an ANC system aims to reduce unwanted acoustic noise by generating anti-noise with secondary loudspeakers [6]. A spatial ANC system aims to reduce the unwanted noise inside a continuous region of interest, and it is often achieved by a multi-input-multi-output (MIMO) system utilizing error microphones and secondary loudspeakers [117], [118].

Adaptive filtering is considered to be a beneficial tool in ANC systems because the noise we are facing is often time-varying [119], [120]. The adaptive filters adjust their coefficients to generate the secondary loudspeaker driving signals such that the signals recorded by the error microphones are minimized [121], [122]. Conventionally, a MIMO spatial ANC system requires the error microphones to be uniformly distributed inside the control region [19], which is one of the main drawbacks prohibiting practical implementation. Space domain signal processing using harmonic (cylindrical/spherical) based sound field processing has been recently applied in ANC to decrease the use of resources by placing error microphones on the boundary of the region of interest [118], [73]. While space domain methods are proven to be effective for 3D space noise reduction [90], there is a prerequisite of regularly spacing the error microphones [123] and the secondary loudspeakers [80] about a certain sphere. Conventional spherical arrays surrounding the region of interest still complicate the implementation of an ANC system due to the restrictions of regular spherical sampling as well as mobility constraints of users entering and exiting the region. Hence, a more practical array geometry for spherical harmonic based spatial ANC is desirable.

Okamoto [124] has addressed an external noise control system with a planar loudspeaker array based on spherical harmonic analysis. Zhang et al [125] have developed a 2.5D reproduction method to reproduce sound field inside a 2D region using 3D-distributed point sources. Although a cumbersome 3D spherical array is avoided in [125], the reproduced sound field is restricted to be height invariant. Later on, Maeno et al [126] have achieved spatial ANC with a circular array design. However, this method only considers the even spherical harmonics, hence it can only achieve noise reduction within the x-y plane. Zhang et al [97] have developed a spatial sound field reproduction method based on Hilbert space spherical harmonic analysis. This method has achieved a 3D sound field reproduction with multiple circular loudspeaker arrays. However, it requires the distance between two adjacent circular arrays to be within a given range. As a result, the arrays are placed over the entire region which highly restricts the users movement. In [127] and [128], Gupta and Abhayapala introduce 3D spatial sound field recording and reproduction methods that use multiple circular arrays. Based on azimuth harmonic analysis, these two papers do not restrict the positioning of arrays in order to cover the

entire region, which provides us the building blocks for constructing a realizable 3D spatial ANC system with multiple circular arrays.

The previous work on spatial ANC always focus on an empty region of interest where free-space is assumed inside. However, we design ANC systems for people to enjoy the silence. Therefore, the presence of users should be taken into consideration. The scattering effect of a human head has been studied in [129], [130] for personal audio system purposes. Betlehem and Poletti have developed a spatial sound field reproduction system with a rigid spherical scatter inside the region of interest in space domain based on mode matching [131]. While most of the previous work focus on modelling and reducing the scattering sound field, there are few studies on the scattering effect in a spatial ANC system. In this chapter, we focus on analysing the scattered sound field in the region of interest in space domain. As we are often not able to perfectly cancel the noise field, it is valuable to study the influence of the scattering effect of the users in a spatial ANC system to increase the noise reduction performance.

In this chapter, we propose a spatial ANC structure with multiple circular microphone and loudspeaker arrays (shown in Fig. 3.1), along with a feed-forward adaptive filtering algorithm in space domain to achieve ANC over the whole 3D region of interest which includes a human head modeled as a rigid scatter. The ANC method utilizes a novel flexible array geometry isolated in the upper hemisphere such that does not obstruct users entering and exiting the region of interest.

3.2 Problem Formulation

In this section, we review the basic ideas of recording a noise field and performing noise reduction inside a 3D source free region Ω, and relate these algorithms to the current problem.

We use the same spherical coordinates $\boldsymbol{x} = (r, \theta, \phi)$ to describe a location of a point $\boldsymbol{x}$ where r is the distance to the point from the origin, θ is the elevation angle downwards from the z-axis from 0 to π, ϕ denotes the azimuth angles, and the origin refers to the centre of Ω. Specifically, the radius of a circular array is defined as the distance from any point on the circle to the origin of the region, instead of to the centre of the circle.

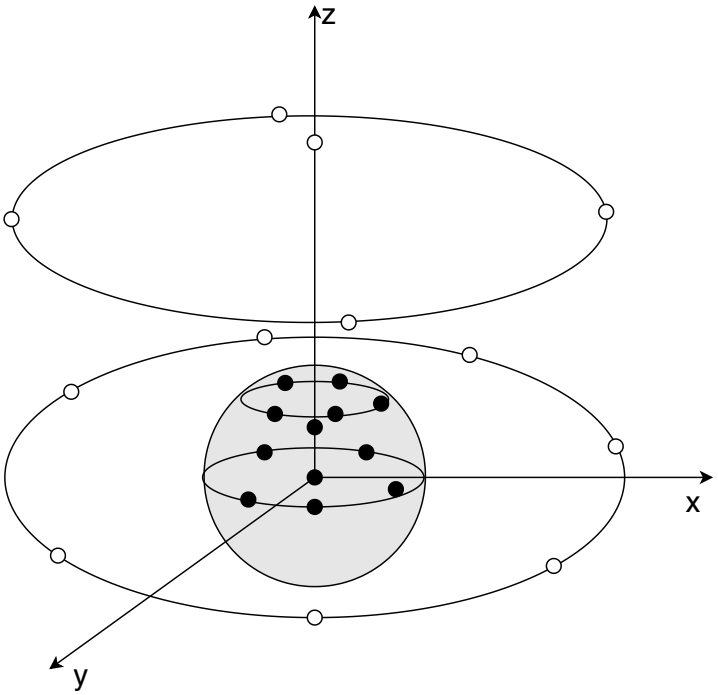


Figure 3.1: Geometry of microphones (•) and loudspeakers (○) with the region of interest being the lower hemi-sphere of the shaded area for the MCAD method.

The residual noise field at an arbitrary point $\boldsymbol{x} = (r, \theta, \phi)$ inside a spherical region of interest Ω of radius r_Ω can be expressed as [19]

$$e(\boldsymbol{x}, k) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}^{(e)}(k) j_n(kr) \mathcal{P}_{nm}(\cos\theta) E_m(\phi), \tag{3.1}$$

where order n and mode m are integers, $N = \lceil kr_\Omega \rceil$ [80], $\alpha_{nm}^{(e)}(k)$ is a set of spherical harmonic coefficients representing the residual sound field inside Ω, $k = 2\pi f/c$ is the wave number, f is the frequency of the noise signal, c is the speed of sound, $j_n(kr)$ is the spherical Bessel function [110], $E_m(\phi) = \sqrt{(1/2\pi)} e^{im\phi}$ is the normalized exponential function and $\mathcal{P}_{nm}(\cos\theta)$ is the normalized associated Legendre function given by

$$\mathcal{P}_{nm}(\cos\theta) = \sqrt{\frac{2n+1}{2}} \sqrt{\frac{(n-|m|)!}{(n+|m|)!}} P_{n|m|}(\cos\theta), \tag{3.2}$$

where $P_{n|m|}(\cos\theta)$ refers to the associated Legendre function.

If we have a continuous spherical microphone array, then we can integrate the sound pressure $e(\boldsymbol{x}, k)$ with respect to θ and ϕ to obtain [21]

$$\alpha_{nm}^{(e)}(k) = \frac{\int_0^{2\pi}\int_0^{\pi} e(\boldsymbol{x}, k)\mathcal{P}_{nm}^*(\cos\theta)E_{-m}(\phi)\sin\theta d\theta d\phi}{j_n(kr)}, \tag{3.3}$$

where $(\cdot)^*$ refers to the complex conjugation.

However, when there are only circular microphone arrays, we can only integrate the sound pressure over ϕ but not over θ in (3.3). Therefore, we are not able to obtain the spherical harmonic coefficients $\alpha_{nm}^{(e)}(\cdot)$ of the sound field inside Ω. To propose an efficient circular array geometry, we first obtain azimuth harmonic coefficients on the $\hbar$-th circular microphone array with $r = r_\hbar$ and $\theta = \theta_\hbar$ as [127]

$$a_m(r_\hbar, \theta_\hbar, k) \triangleq \int_0^{2\pi} e(r_\hbar, \theta_\hbar, \phi, k)E_{-m}(\phi)d\phi. \tag{3.4}$$

By substituting (3.1) to (3.4), we obtain the relationship between the spherical harmonic coefficients $\alpha_{nm}^{(e)}(\cdot)$ and the azimuth harmonic coefficients $a_m(\cdot)$ as

$$a_m(r_\hbar, \theta_\hbar, k) \approx \sum_{n=|m|}^{N} \alpha_{nm}^{(e)}(k)j_n(kr_\hbar)\mathcal{P}_{nm}(\cos\theta_\hbar). \tag{3.5}$$

To reduce the noise inside Ω, we need to reproduce an anti-noise field with its spherical harmonic coefficients given by $\alpha_{nm}^{(s)}(\cdot) = -\alpha_{nm}^{(e)}(\cdot)$. From previous work [80], we know the aperture function of a continuous spherical loudspeaker array at the point $(R, \theta_\ell, \phi_\ell)$ can be expressed as

$$\zeta_s(\theta_\ell, \phi_\ell, k) \approx \sum_{n=0}^{N}\sum_{m=-n}^{n} \sigma_{nm}(k)\mathcal{P}_{nm}(\cos\theta_\ell)E_m(\phi_\ell), \tag{3.6}$$

and the spherical harmonic coefficients due to this loudspeaker array are

$$\alpha_{nm}^{(s)}(k) = 4\pi ikh_n^{(1)}(kR)\sigma_{nm}(k), \tag{3.7}$$

where $h_n^{(1)}(\cdot)$ is the spherical Hankel function of the first kind. This method is called

mode-matching since it matches the desired sound field $\alpha_{nm}^{(s)}(\cdot)$ and the aperture function of loudspeakers $\sigma_{nm}(\cdot)$ mode by mode.

However, with these $\sigma_{nm}(\cdot)$ in hand, it is not possible to obtain the aperture function of a horizontal continuous circular loudspeaker in (3.6). We have these circular aperture functions at the τ-th loudspeaker array with $\theta = \theta_\tau$ defined as [128]

$$\zeta^{(\tau)}(\phi_\ell^{(\tau)}, k) \triangleq \sum_{m=-\infty}^{\infty} \beta_m^{(\tau)}(k) E_m(\phi_\ell^{(\tau)}), \tag{3.8}$$

where

$$\beta_m^{(\tau)}(k) = \int_0^{2\pi} \zeta^{(\tau)}(\phi_\ell^{(\tau)}, k) E_{-m}(\phi_\ell^{(\tau)}) d\phi_\ell^{(\tau)}, \tag{3.9}$$

is defined as circular harmonics of the aperture function. The resulting secondary sound pressure at a point $\boldsymbol{x}$ due to this circular loudspeaker aperture is

$$y'(\boldsymbol{x}, k) = \int_0^{2\pi} \zeta^{(\tau)}(\phi_\ell^{(\tau)}, k) \frac{e^{ik\|\boldsymbol{Y}-\boldsymbol{x}\|}}{\|\boldsymbol{Y}-\boldsymbol{x}\|} d\phi_\ell^{(\tau)}, \tag{3.10}$$

where $\boldsymbol{Y}$ is a point on the circular loudspeaker aperture. With (3.8), (3.10) and the addition theorem of the Green's function, the sound field due to $\beta_m^{(\tau)}(k)$ is derived as

$$\begin{aligned} y'(\boldsymbol{x}, k) \approx & \sum_{n=0}^{N} \sum_{m=-n}^{n} 4\pi i k h_n^{(1)}(kR_\tau) \mathcal{P}_{nm}(\cos\theta_\tau) \\ & \times \beta_m^{(\tau)}(k) j_n(kr) \mathcal{P}_{nm}(\cos\theta) E_m(\phi). \end{aligned} \tag{3.11}$$

Hence, the spherical harmonic coefficients of the sound field in Ω due to the circular harmonics of multiple circular loudspeaker arrays in the free space can be expressed as

$$\alpha_{nm}^{(s)}(k) = \sum_{\tau=1}^{F} 4\pi i k h_n^{(1)}(kR_\tau) \mathcal{P}_{nm}(\cos\theta_\tau) \beta_m^{(\tau)}(k) \text{ for } n \geq |m|. \tag{3.12}$$

According to the discussion above, we are not able to directly record the spherical harmonic coefficients $\alpha_{nm}^{(e)}(\cdot)$ of the residual sound field in Ω. Instead, we need to record a series of $a_m(\cdot)$ for solving $\alpha_{nm}^{(e)}(\cdot)$. Similarly, we are not able to obtain the aperture function of loudspeakers from spherical aperture coefficients $\sigma_{nm}(\cdot)$ by

mode matching. We need to obtain the circular harmonics of the aperture functions $\beta_m(\cdot)$ from the desired sound field given by $\alpha_{nm}^{(s)}(\cdot) = -\alpha_{nm}^{(e)}(\cdot)$. At first glance, we can achieve these two requirements by solving the least mean square problems in (3.5) and (3.12). However, we risk our system having singular values associated with matrix inversion and spatial aliasing if we randomly place the arrays. Instead, a carefully designed array geometry is necessary to achieve unrestricted user movement and a guaranteed existence of the solutions for $\alpha_{nm}^{(e)}(\cdot)$ and $\beta_m(\cdot)$.

We utilize the natural properties of Bessel functions and Legendre functions to obtain $\alpha_{nm}^{(e)}(\cdot)$ from a given set of $a_m(\cdot)$ in (3.5) with a specific microphone array geometry in Sec. 3.3.1. In Sec. 3.3.2, a loudspeaker array is designed to derive the solution of the circular aperture function $\beta_m(\cdot)$ based on a desired set of $\alpha_{nm}^{(s)}(\cdot)$. In addition to the array design, we address other problems including; 1) In reality, the user of our system causes scattering in the region of interest; 2) For real time ANC, we need an adaptive filtering algorithm in the spherical harmonic domain. We detail the solutions of these two problems in Sec. 3.4 and Sec. 3.5.1, respectively.

3.3 Array Configuration

In this section we continue to design a N-th order spatial ANC system employing circular arrays. We first discuss the placement of continuous circular microphone and loudspeaker arrays for recording and reproduction, respectively, and then discretize those circular arrays for realistic implementation. For brevity, we assume N is even for the subsequent derivations. If N is odd, some of the derivation needs to be adjusted correspondingly.

3.3.1 Microphone array design

We detail the design procedure of obtaining the spherical harmonic coefficients $\alpha_{nm}(k)$ of a given region of interest with multiple circular array in this subsection.

To configure the placement of microphones, we separate spherical harmonic coefficients $\alpha_{nm}(k)$ into even-coefficients ($n + |m|$ is even) and odd coefficients ($n+|m|$ is odd). Furthermore, we separate the even coefficients into $\alpha_{00}(k)$ and the

remaining even coefficients $\boldsymbol{\alpha}_m^{\mathrm{e}} = [\alpha_{|m|m}, \alpha_{(|m|+2)m}, \cdots, \alpha_{Nm}]^T$, and separate the odd coefficients into zero mode odd coefficients $\boldsymbol{\alpha}_0^{\mathrm{o}} = [\alpha_{10}, \alpha_{30}, \cdots, \alpha_{(N-1)0}]^T$ and the remaining odd coefficients $\boldsymbol{\alpha}_m^{\mathrm{o}} = [\alpha_{(|m|+1)m}, \alpha_{(|m|+3)m}, \cdots, \alpha_{(N-1)m}]^T$. For the remainder of this subsection we discuss how to record each sub-set of coefficients.

$\alpha_{00}(k)$ *estimation:* By inspecting (3.5), we observe that at $r = 0$, the only active coefficient is $\alpha_{00}(k)$ since $j_n(0) = 0$ for $n > 0$. Hence, we can obtain $\alpha_{00}(k)$ by placing a single microphone at the origin as

$$\alpha_{00}(k) = \frac{a_0(0,0,k)}{\mathcal{P}_{00}(\cos 0)} = \sqrt{2}a_0(0,0,k), \tag{3.13}$$

where $a_0(0,0,k)$ is calculated by (3.4).

$\boldsymbol{\alpha}_m^e$ *estimation:* We notice that the normalized associated Legendre function has a property of

$$\mathcal{P}_{n|m|}\left(\cos\frac{\pi}{2}\right) = \mathcal{P}_{n|m|}(0)\begin{cases} = 0 & n+|m| \text{ is odd}, \\ \neq 0 & n+|m| \text{ is even}, \end{cases} \tag{3.14}$$

such that only even coefficients are active with $\theta = \pi/2$ (on the x-y plane) in (3.5). For a N-th order system, we have maximum $N/2-1$ elements in $\boldsymbol{\alpha}_m^e$ with a specific mode m, excluding $\alpha_{00}(k)$. Hence, we need $D_{\mathrm{e}} = N/2-1$ arrays on the x-y plane with $\theta_\hbar = \pi/2$ and radius $r_\hbar$, $\hbar = 1, 2, \cdots, D_{\mathrm{e}}$, respectively, to obtain all the elements in $\boldsymbol{\alpha}_m^{\mathrm{e}}$, given by

$$\boldsymbol{\alpha}_m^{\mathrm{e}} = \boldsymbol{J}_m^{\mathrm{e}(-1)}\boldsymbol{a}_m, \tag{3.15}$$

where $\boldsymbol{a}_m = [a_m(r_1, \frac{\pi}{2}, k), a_m(r_2, \frac{\pi}{2}, k), \cdots, a_m(r_{D_{\mathrm{e}}}, \frac{\pi}{2}, k)]^T$ is obtained from (3.4), and $\boldsymbol{J}_m^{\mathrm{e}(-1)}$ is the inverse matrix of

$$\boldsymbol{J}_m^{\mathrm{e}} = \begin{bmatrix} J_{|m|m}^{(1)} & J_{(|m|+2)m}^{(1)} & \cdots & J_{Nm}^{(1)} \\ \vdots & \ddots & \vdots & \\ J_{|m|m}^{(D_{\mathrm{e}})} & J_{(|m|+2)m}^{(D_{\mathrm{e}})} & \cdots & J_{Nm}^{(D_{\mathrm{e}})} \end{bmatrix}, \tag{3.16}$$

where $J_{nm}^{(\hbar)} = j_n(kr_\hbar)\mathcal{P}_{nm}(\cos\theta_\hbar)$.

We select the radii of each array to avoid spherical Bessel zeros and to make sure that $j_n(kr_\hbar) \neq 0$ for $k \in [k_l, k_h]$, where k_l and k_h are the lowest and the highest wave number of the noise signal we focus on, respectively. Meanwhile, we uniformly space the arrays in the range of $(0, r_\Omega)$ as best as possible while avoiding Bessel zeros. Therefore, we avoid the singularity of matrix $\boldsymbol{J}_m^{\mathrm{e}}$ so that the system does not rely on the ability of the least squares for matrix inversion. One of the arrangements can be:

$$r_\hbar = \frac{2}{k_0}, \frac{4}{k_0}, \cdots, \frac{N}{k_0}, \tag{3.17}$$

where $k_0 = N/r_\Omega$. We also note that more circular arrays can be applied for improving the accuracy of the measurements. In this case, we use $\boldsymbol{J}_m^{\mathrm{e}\dagger}$, the Moore-Penrose inverse of $\boldsymbol{J}_m^{\mathrm{e}}$ instead of $\boldsymbol{J}_m^{\mathrm{e}(-1)}$ in (3.15).

$\boldsymbol{\alpha}_0^o$ *estimation:* When $\theta = 0$, which refers points along the z-axis, we have

$$\mathcal{P}_{nm}(\cos 0) = \mathcal{P}_{nm}(1) \begin{cases} = 0 & m \neq 0 \\ \neq 0 & m = 0 \end{cases} \tag{3.18}$$

such that the sound field on the z-axis only contains the components with mode $m = 0$. A similar property is found at $\theta = \pi$. Since we only want the microphones above the head (in the upper hemi-sphere), we do not use $\theta = \pi$. For a N-th order system, there are $N/2$ elements in $\boldsymbol{\alpha}_0^{\mathrm{o}}$. Hence, we put $D_{\mathrm{z}} = N/2$ microphones on the z-axis ($\theta_{\mathrm{z}\hbar} = 0$) at $(r_{\mathrm{z}\hbar}, 0, 0)$, $q = 1, 2, \cdots, Q_{\mathrm{z}}$. In addition, we have already acquired even $\boldsymbol{\alpha}_0^{\mathrm{e}} = [\alpha_{00}, \alpha_{20}, \cdots, \alpha_{N0}]^T$ by (3.15), which allows us to derive

$$\boldsymbol{\alpha}_0^{\mathrm{o}} = \boldsymbol{J}_0^{z,\mathrm{o}\dagger} \boldsymbol{a}_0^{\mathrm{z}} - \boldsymbol{J}_0^{\mathrm{z},\mathrm{o}\dagger} \boldsymbol{J}_0^{\mathrm{z,e}} \boldsymbol{\alpha}_0^{\mathrm{e}}, \tag{3.19}$$

where $\boldsymbol{a}_0^{\mathrm{z}} = [a(r_{\mathrm{z}1}, 0, k), a(r_{\mathrm{z}2}, 0, k), \cdots, a(r_{\mathrm{z}D_{\mathrm{z}}}, 0, k)]^T$, $\boldsymbol{J}_0^{\mathrm{z},\mathrm{o}\dagger}$ is the Moore-Penrose inverse of $\boldsymbol{J}_0^{\mathrm{z,o}}$ given by

$$\boldsymbol{J}_0^{\mathrm{z,o}} = \begin{bmatrix} J_{10}^{(\mathrm{z}1)} & J_{30}^{(\mathrm{z}1)} & \cdots & J_{(N-1)0}^{(\mathrm{z}1)} \\ \vdots & \ddots & \vdots & \\ J_{10}^{(\mathrm{z}D_{\mathrm{z}})} & J_{30}^{(\mathrm{z}D_{\mathrm{z}})} & \cdots & J_{(N-1)0}^{(\mathrm{z}D_{\mathrm{z}})} \end{bmatrix}, \tag{3.20}$$

and $\boldsymbol{J}_0^{\mathrm{z,e}}$ is of the same expression of $\boldsymbol{J}_0^{\mathrm{z,o}}$ but the order n is even, varying from 0

to N. The selection of $r_{z\hbar}$ follows the same rules in choosing $r_\hbar$ on the x-y plane.

$\boldsymbol{\alpha}_m^o$ *estimation:* For a 1-st order system, we have already finished estimating all the spherical harmonic coefficients after the previous three steps. For a N-th order system where $N > 1$, there are some odd coefficients that remain unknown. We define $D_\mathrm{o} \triangleq \max\{n - |m|\}$ where $\alpha_{nm}(k)$ remains unknown. We place D_o circles parallel to the x-y plane at $(r_{\mathrm{o}\hbar}, \theta_{\mathrm{o}\hbar}, \phi)$, where $q = 1, 2, \cdots, D_\mathrm{o}$. For the $\hbar$-th circle, we select the radius $r_{\mathrm{o}\hbar}$ to be $N_{\mathrm{o}\hbar}/k_0$, where $N_{\mathrm{o}\hbar}$ is the highest order of the unknown $\alpha_{nm}(k)$ with $n - |m| = \hbar$. $\theta_{\mathrm{o}\hbar}$ is chosen such that $\mathcal{P}_{n|m|}(\cos\theta_{\mathrm{o}\hbar}) \neq 0$ for all remaining modes where $n - |m| = \hbar$, which can be $\pi/3$ for $\hbar = 1$ and $\pi/6$ for $\hbar = 2$. As the normalized associate Legendre function is symmetric to $\theta = \pi/2$, we can always find a $\theta_{\mathrm{o}\hbar} < \pi/2$ to meet the requirement, which refers to this array being in the upper hemi-sphere of Ω. With the given $\boldsymbol{\alpha}_m^\mathrm{e}$ in (3.15) and $\boldsymbol{\alpha}_0^\mathrm{o}$ in (3.19), along with (3.5) we derive

$$\boldsymbol{\alpha}_m^\mathrm{o} = \boldsymbol{J}_m^{\mathrm{o}\dagger}\boldsymbol{a}_m - \boldsymbol{J}_m^{\mathrm{o}\dagger}\widehat{\boldsymbol{J}}_{\boldsymbol{m}}^\mathrm{e}\boldsymbol{\alpha}_m^\mathrm{e}, \tag{3.21}$$

where $\boldsymbol{J}_m^\mathrm{o}$ is given by

$$\boldsymbol{J}_m^\mathrm{o} = \begin{bmatrix} J^{(\mathrm{o}1)}_{(|m|+1)m} & J^{(\mathrm{o}1)}_{(|m|+3)m} & \cdots & J^{(\mathrm{o}1)}_{(N-1)m} \\ \vdots & \ddots & \vdots & \\ J^{(\mathrm{o}D_\mathrm{o})}_{(|m|+1)m} & J^{(\mathrm{o}D_\mathrm{o})}_{(|m|+3)m} & \cdots & J^{(\mathrm{o}D_\mathrm{o})}_{(N-1)m} \end{bmatrix}, \tag{3.22}$$

$\boldsymbol{J}_m^\mathrm{e}$ is similar to $\boldsymbol{J}_m^\mathrm{o}$ when the order is even and is varying from $|m|$ to N, and $\boldsymbol{a}_m = [a(r_{\mathrm{o}1}, \theta_{\mathrm{o}1}, k), a(r_{\mathrm{o}2}, \theta_{\mathrm{o}2}, k), \cdots, a(r_{\mathrm{o}D_\mathrm{o}}, \theta_{\mathrm{o}D_\mathrm{o}}, k)]^T$.

With (3.15), (3.19) and (3.21), we obtain the full set of spherical harmonic coefficients $\alpha_{nm}(k)$ with circular arrays of microphones isolated to the upper hemi-sphere, on and above the x-y plane, and along the z-axis.

3.3.2 Loudspeaker array design

Since the spherical harmonic decomposition is valid only in a source free region, we are unable to place loudspeakers inside Ω. Hence, for loudspeaker array design, it is not practical to consider the same method that we proposed in Sec.3.3.1 with the spherical harmonics separated into even and odd components.

We observe in (3.12) that $\alpha_{nm}^{(s)}(k)$ is only influenced by $\boldsymbol{\beta}_m(k) = [\beta_m^{(1)}(k), \beta_m^{(2)}(k), \cdots, \beta_m^{(F)}(k)]$ when $n \geq |m|$. For example, $\alpha_{NN}^{(s)}(k)$ and $\alpha_{N-N}^{(s)}(k)$ are only influenced by $\boldsymbol{\beta}_N(k)$. Additionally, $\boldsymbol{\beta}_m(k)$ controls the sound field along the degree m but not order n. Based on this observation, we address the location of loudspeaker arrays to calculate the azimuth aperture function $\boldsymbol{\beta}_m(k)$ with a given set of $\alpha_{nm}^{(s)}(k)$.

For $|m| = N$ coefficients: Consider (3.12) with $|m| = N$, we have

$$\alpha_{NN}^{(s)}(k) = \sum_{\tau=1}^{F} 4\pi ikh_N^{(1)}(kR_\tau)\mathcal{P}_{NN}(\cos\theta_\tau)\beta_N^{(\tau)}(k), \tag{3.23}$$

where there is only one coefficient $\alpha_{NN}^{(s)}(k)$ to be controlled by $\beta_N^{(\tau)}(k)$. In other words, we only need one circular array of loudspeakers to have one aperture function $\beta_N^{(1)}(k)$. Hence, with a desired $\alpha_{NN}^{(s)}(k)$, we calculate $\beta_N^{(1)}(k)$ by

$$\beta_N^{(1)}(k) = \frac{\alpha_{NN}^{(s)}(k)}{4\pi ikh_N^{(1)}(kR_1)\mathcal{P}_{NN}(\cos\theta_1)}. \tag{3.24}$$

By examining (3.24), we observe that the placement of this loudspeaker array is required to avoid the zeros of the spherical Hankel function $h_N^{(1)}(kR_1)$ and the associated Legendre function $\mathcal{P}_{NN}(\cos\theta_1)$. One solution is placing the first circle on the $\theta_1 = \pi/2$ x-y plane because we know that $\mathcal{P}_{NN}(\cos\pi/2) \neq 0$ from (3.18).

For $|m| = N-1$ coefficients: We observe that $\beta_{N-1}^{(\tau)}(k)$ influences two coefficients $\alpha_{N-1N-1}^{(s)}(k)$ and $\alpha_{NN-1}^{(s)}(k)$. Hence, we need at least two circular aperture functions $\beta_{N-1}^{(1,2)}(k)$ to realize the desired sound field coefficients. We can reuse the first circle ($\tau = 1$), and place another circle at an elevation angle θ_2 such that $\mathcal{P}_{NN-1}(\cos\theta_2) \neq 0$. From (3.12) we know

$$\begin{aligned}\alpha_{N-1N-1}^{(s)}(k) = &\sum_{\tau=1}^{2} 4\pi ikh_{N-1}^{(1)}(kR_\tau) \\ &\times \mathcal{P}_{N-1N-1}(\cos\theta_\tau)\beta_{N-1}^{(\tau)}(k),\end{aligned} \tag{3.25}$$

and

$$\alpha_{NN-1}^{(s)}(k) = \sum_{\tau=1}^{2} 4\pi ikh_N^{(1)}(kR_\tau)\mathcal{P}_{NN-1}(\cos\theta_\tau)\beta_{N-1}^{(\tau)}(k). \tag{3.26}$$

Therefore, we obtain $\boldsymbol{\beta}_{N-1}(k)$ as

$$\begin{bmatrix} \beta_{N-1}^{(1)} \\ \beta_{N-1}^{(2)} \end{bmatrix} = \boldsymbol{H}_{N-1}^{-1} \begin{bmatrix} \alpha_{(N-1)(N-1)}^{(s)}(k) \\ \alpha_{N(N-1)}^{(s)}(k) \end{bmatrix}, \tag{3.27}$$

where $\boldsymbol{H}_m$ is given by

$$\boldsymbol{H}_m = \begin{bmatrix} H_{|m|m}^{(1)} & \cdots & H_{Nm}^{(N-m+1)} \\ \vdots & \ddots & \vdots \\ H_{|m|m}^{(1)} & \cdots & H_{Nm}^{(N-m+1)} \end{bmatrix}, \tag{3.28}$$

and $H_{nm}^{(\tau)} = 4\pi i k h_n^{(1)}(kR_\tau)\mathcal{P}_{nm}(\cos\theta_\tau)$. To ensure that the second array does not influence the $|m| = N$ series coefficients, we set $\beta_N^{(2)} = 0$.

For $|m| = N-2$ coefficients: Similar to the discussion we had for $|m| = N-1$ series coefficients, now we need to calculate three unknown aperture functions for three desired sound field coefficients $\alpha_{N-2N-2}^{(s)}(k)$,$\alpha_{N-1N-2}^{(s)}(k)$ and $\alpha_{NN-2}^{(s)}(k)$. Reusing the first and the second array, we set one more circular array at θ_3 such that $\mathcal{P}_{NN-2}(\cos\theta_3) \neq 0$. Hence, we have

$$\begin{bmatrix} \beta_{N-2}^{(1)} \\ \beta_{N-2}^{(2)} \\ \beta_{N-2}^{(3)} \end{bmatrix} = \boldsymbol{H}_{N-2}^{-1} \begin{bmatrix} \alpha_{(N-2)(N-2)}^{(s)}(k) \\ \alpha_{(N-1)(N-2)}^{(s)}(k) \\ \alpha_{N(N-2)}^{(s)}(k) \end{bmatrix}. \tag{3.29}$$

Additionally, we set $\beta_N^{(3)} = 0$ and $\beta_{N-1}^{(3)} = 0$ to avoid the influence on previous series of coefficients. Repeat the process for $|m| = N-3, N-4, \cdots, 1$ coefficients.

For $|m| = 0$ coefficients: There are $N+1$ coefficients in this series. We reuse the previous N circles and place an additional circle at θ_{N+1} such that $\mathcal{P}_{N0}(\cos\theta_{N+1}) \neq 0$ to calculate $\beta_0^{(N+1)}$ with (3.12). The simplest solution is $\theta_{N+1} = 0$, which refers to a single loudspeaker on the z-axis. We set $\beta_m^{(N+1)} = 0$ for $m = 1, 2, \cdots, N$.

In total, we utilize $F = N+1$ circular arrays to reproduce the full set of desired sound field coefficients $\alpha_{nm}^{(s)}(k)$. These arrays are also used to estimate $\beta_m^{(\tau)}(k)$ when m is negative, starting from $m = -N$, with the same equations as (3.24), (3.27) and (3.29).

Since the associated Legendre function is symmetric about $\theta = \pi/2$, we are

able to choose $\theta_\tau \leq \pi/2$ for all F circular arrays, which allows for no array to be below the x-y plane. Furthermore, since spherical Hankel functions have no zeros, the radii of these loudspeaker arrays are not restricted by our proposed method. These loudspeaker radii share the requirements of traditional MIMO ANC systems. Secondary loudspeakers should be put outside the region of interest while also remaining closer to the origin than the reference microphones or the noise sources. For a reverberate room, we suggest placing the secondary loudspeakers closer to the region of interest for a higher direct to reverberate radio as this increases the robustness of the system [9]. One of the benefits of our proposed loudspeaker array design is that the placement of circular loudspeaker arrays is flexible. Both radii and elevation angles are flexible to within a range for the convenience of construction, making the system more practically viable.

3.3.3 Discretization and sensor placement

We have assumed the arrays of microphone and loudspeaker to be continuous in the previous discussion. In this sub-section, we discretize them into a finite number of sensors for the purpose of practical realization following the Nyquist-Shannon's sampling theorem.

For the $\hbar$-th microphone array with radius of $r_\hbar$, it can record up to $N_\hbar = \lceil kr_\hbar \rceil$ order of coefficients. We need at least $Q_\hbar = 2N_\hbar + 1$ microphones uniformly placed on the circle. Thus, we approximate (3.4) as

$$a_m(r_\hbar, \theta_\hbar, k) \approx \frac{2\pi}{Q_\hbar} \sum_{q=1}^{Q_\hbar} S(r_\hbar, \theta_\hbar, \phi_q, k) E_{-m}(\phi_q), \tag{3.30}$$

For loudspeaker arrays, we define the highest active mode of one circular array as $N_\tau = N - \tau + 1$, where $\beta_m^{(\tau)}(k) = 0$ for $m > N_\tau$. At least $L_\tau = 2N_\tau + 1$ loudspeakers are required uniformly on the circle. We approximate (3.9) as

$$\beta_m^{(\tau)}(k) \approx \frac{2\pi}{L_\tau} \sum_{\ell=1}^{L_\tau} \rho(\phi_\ell, k) E_{-m}(\phi_\ell). \tag{3.31}$$

Hence, the aperture function of the ℓ-th loudspeaker at the τ-th circle is given by

$$\zeta(\phi_\ell^\tau, k) \approx \frac{2\pi}{L_\tau} \sum_{m=-N_\tau}^{N_\tau} \beta_m^{(\tau)}(k) E_m(\phi_\ell^\tau). \tag{3.32}$$

3.3.4 Array design procedure

We review the design process and give a general guidance of setting up the proposed spatial ANC system in this section. We note here that we focus on the geometric design of secondary loudspeakers and error microphones. Although reference microphones are often necessary in ANC systems, the placement of reference microphones is out of the scope of this work. The design procedure is as follows:

1. Determine the highest desired frequency band k and the radius of the region of interest r_Ω.

2. Calculate the maximum order of the system, given by $N = \lceil kr_\Omega \rceil$.

3. Determine the number of microphone arrays. At least $N/2$ arrays on the x-y plane (including one at the origin), $N/2$ microphones on the z-axis and $N/2$ arrays parallel to the x-y plane are necessary.

4. Determine the radii of the microphone arrays. Make sure these radii avoid the spherical Bessel zeros for the target frequency band and have a good diversity over $(0, r_\Omega)$.

5. Determine the elevation angle $\theta_\hbar$ for each array parallel to the x-y plane. Select $\theta_\hbar$ such that $\mathcal{P}_{n(n-\hbar+1)}(\cos\theta_\hbar) \neq 0$ for all $1 < n < N$. If the lower hemi-sphere of the region of interest is desired to be empty, choose $\theta_\hbar < \pi/2$.

6. For each circular array, calculate the maximum spherical harmonic order $N_\hbar = \lceil kr_\hbar \rceil$ and uniformly place at least $2N_\hbar + 1$ microphones on the circle.

7. Determine the number of loudspeaker arrays, given by $N + 1$.

8. Determine the elevation angle θ_τ of those loudspeaker arrays. For the τ-th array, choose θ_τ such that $\mathcal{P}_{N(N-\tau+1)}(\cos\theta_\tau) \neq 0$. We suggest $\theta_1 = \pi/2$ and

$\theta_{N+1} = 0$. If the lower hemi-sphere of the region of interest is desired to be empty, choose $\theta_\tau < \pi/2$.

9. Determine the radii of the loudspeaker arrays. Those radii should be larger than r_Ω, but smaller than the radii of the noise sources and radii of the reference microphones.

10. For each circular array, calculate the maximum active spherical harmonic order $N_\tau = N - \tau + 1$ and uniformly place at least $2N_\tau + 1$ loudspeakers on the circle.

3.4 Influence of human head scattering on sound field recording

In the previous sections, we assume the region of interest Ω is a free-space region without sources or scatters. However, this assumption is not always true since users (human heads) need to stay inside the region to enjoy the noise reduction. In other words, in the application of a spatial ANC system, we often have to deal with the scattering sound field instead of the free-space sound field. In this section, we modify our proposed method to consider the human head as a spherical rigid scatterer inside Ω.

For a spherical rigid scatterer of radius r_a at the center of Ω, the expression (3.1) changes to [19]

$$e(\boldsymbol{x}, k) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}^{(e)}(k) b_n(kr) \mathcal{P}_{nm}(\cos\theta) E_m(\phi), \tag{3.33}$$

where

$$b_n(kr) = j_n(kr) - \frac{j_n(kr_a)'}{h_n^{(1)}(kr_a)'} h_n(kr), \tag{3.34}$$

and $(\cdot)'$ refers to the differential operator, r is defined with respect to the centre of the scatter. However, it is not convenient to fix the user's head at the origin of Ω. Moreover, as detailed in Sec.3.3.1, we place one error microphone at the origin of Ω for residual sound field recording, making it impossible for the user to place their

head at the same position. Therefore, we investigate the case that the scatterer is inside the region of interest but not co-centered.

Define the origin of the region of interest Ω as O and the origin of the scatterer as O_1. As the scatterer is located inside the region, we have $r_\Omega > |\widehat{\boldsymbol{x}}|$, where $\widehat{\boldsymbol{x}} = O - O_1 = (|\widehat{\boldsymbol{x}}|, \theta_x, \phi_x)$. Thus, the location at an arbitrary point inside Ω can be written as $\boldsymbol{x} = (r, \theta, \phi)$ with respect to O and $\boldsymbol{x}_1 = (r_1, \theta_1, \phi_1)$ with respect to O_1. With the addition theorem [113], we have

$$j_n(kr_1)Y_{nm}(\theta_1, \phi_1) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} T_{n\nu}^{\mu m}(k\widehat{\boldsymbol{x}}) j_\nu(kr) Y_{\nu\mu}(\theta, \phi), \tag{3.35}$$

and

$$h_n^{(1)}(kr_1)Y_{nm}(\theta_1, \phi_1) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} T_{n\nu}^{\mu m}(k\widehat{\boldsymbol{x}}) h_\nu^{(1)}(kr) Y_{\nu\mu}(\theta, \phi), \tag{3.36}$$

where

$$\begin{aligned} T_{n\nu}^{\mu m}(k\widehat{\boldsymbol{x}}) = 4\pi i^{\nu-n} \sum_{u=0}^{n+\nu} i^u (-1)^\mu \sqrt{\frac{(2n+1)(2\nu+1)(2u+1)}{4\pi}} \\ \times j_u(k|\widehat{\boldsymbol{x}}|) Y_{u(\mu-m)}^*(\theta_x, \phi_x) W1 W2, \end{aligned} \tag{3.37}$$

with

$$W1 = \begin{pmatrix} n & \nu & u \\ 0 & 0 & 0 \end{pmatrix}, \; W2 = \begin{pmatrix} n & \nu & u \\ m & -\mu & \mu - m \end{pmatrix}, \tag{3.38}$$

denoting Wigner $3-j$ symbols [113]. With (3.35) and (3.36), we derive

$$b_n(kr_1)Y_{nm}(\theta_1, \phi_1) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} T_{n\nu}^{\mu m}(k\widehat{\boldsymbol{x}}) b_\nu(kr) Y_{\nu\mu}(\theta, \phi). \tag{3.39}$$

As a result, the sound pressure at a point with respect to O_1 and with respect to

O is expressed as

$$\begin{aligned} z(\boldsymbol{x}_1, k) &\approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) b_n(kr_1) \\ &\quad \times \mathcal{P}_{nm}(\cos\theta_1) E_m(\phi_1) \\ &= \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \alpha_{nm}(k) T_{n\nu}^{\mu m}(k\widehat{\boldsymbol{x}}) \\ &\quad \times b_\nu(kr) \mathcal{P}_{\nu\mu}(\cos\theta) E_\mu(\phi) = z(\boldsymbol{x}, k). \end{aligned} \tag{3.40}$$

By examining (3.40), we obtain the sound pressure at the point with respect to O as

$$z(\boldsymbol{x}, k) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \alpha_{\nu\mu}^{'}(k) b_\nu(kr) \mathcal{P}_{\nu\mu}(\cos\theta) E_\mu(\phi), \tag{3.41}$$

where

$$\alpha_{\nu\mu}^{'}(k) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) T_{n\nu}^{\mu m}(k\widehat{\boldsymbol{x}}). \tag{3.42}$$

Comparing (3.33) and (3.41), we observe that the spherical harmonic decomposition utilizes $b_n(\cdot)$ instead of $j_n(\cdot)$ when there is a scatterer inside the region of interest, regardless of whether the scatterer is co-centered with the region of interest or not. Therefore, we are able to model the human head scattering by modifying all of the spherical Bessel functions $j_n(\cdot)$ in our proposed method to $b_n(\cdot)$ with respect to the origin of Ω regardless of the location of the human head. Furthermore, the movement of the scatter only influences the secondary channels (impulse responses from the secondary loudspeakers to the error microphones) but does not influence the algorithm for noise reduction. This supports that if online secondary channel estimation [132] is applied, users are able to move freely inside the region of interest.

3.5 Design example

3.5.1 Feed-forward adaptive filtering for spatial ANC

In this section, we design a feed-forward adaptive filter in the spherical harmonic domain based on filtered-X LMS algorithm, for the proposed geometry of secondary loudspeakers and error microphones. A feed-forward adaptive filter is selected because, in practice, we are more likely to face wide-band noise, where a feedback adaptive filter is not effective [9].

The residual noise inside the region of interest Ω is the summation of the noise field by outside noise sources, and the anti-noise field by loudspeaker arrays, where we desire

$$\boldsymbol{\alpha}^{(s)} = -\boldsymbol{\alpha}^{(d)}, \tag{3.43}$$

where $\boldsymbol{\alpha}^{(d)}$ is the noise field coefficients vector and $\boldsymbol{\alpha}^{(s)}$ is the desired anti-noise sound field coefficients vector generated by secondary loudspeaker arrays. We omit the frequency component k in the following derivation for simplicity. The residual sound field coefficients in Ω are given by

$$\boldsymbol{\alpha}^{(e)} = \boldsymbol{\alpha}^{(d)} + \boldsymbol{\alpha}^{(s)} = \boldsymbol{\alpha}^{(d)} + \mathcal{N}_{ref}\widetilde{\boldsymbol{S}}\boldsymbol{W}, \tag{3.44}$$

where $\widetilde{\boldsymbol{S}}$ is the spherical harmonic coefficients matrix of the secondary channel transfer function from each secondary loudspeaker to the error microphones, $\mathcal{N}_{ref}$ is the known reference signal, and $\boldsymbol{W}$ is the vector with elements $w(\ell, t)$ as the filter weight for the ℓ-th loudspeaker at the t iteration. In the free-space, the channel coefficient $\widetilde{S}_{nm}(\ell, k)$ of the ℓ-th loudspeaker at $(R_\ell, \theta_\ell, \phi_\ell)$ for order n, mode m, and wave number k, is given by [19]

$$\widetilde{S}_{nm}(\ell, k) = 4\pi i k h_n^{(1)}(kR_\ell)\mathcal{P}_{nm}(\cos\theta_\ell)E_{-m}(\phi_\ell). \tag{3.45}$$

In a simulated room, $\widetilde{S}_{nm}(\ell, k)$ can be obtained with the image source method in the spherical harmonic domain [133]. For practical implementation, the transfer functions can be obtained by measuring the impulse response between the ℓ-th loudspeaker and the error microphones in advance. The coefficients $\widetilde{S}_{nm}(\ell, k)$ can

then be calculated using the method we address in Sec. 3.3.1 and Sec. 3.4.

We define the cost function as

$$\xi = \boldsymbol{\alpha}^{(e)*}\boldsymbol{\alpha}^{(e)}. \tag{3.46}$$

Taking the derivative of ξ with respect to $\boldsymbol{W}$, we obtain

$$\nabla\xi = \frac{\partial\xi}{\partial\boldsymbol{W}} = 2\boldsymbol{\alpha}^{(e)}\frac{\partial\boldsymbol{\alpha}^{(e)}}{\partial\boldsymbol{W}} = 2\boldsymbol{\alpha}^{(e)}(\mathcal{N}_{ref}\widetilde{\boldsymbol{S}})^*. \tag{3.47}$$

Hence, we obtain the filter weights update equation as [9]

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) - \frac{\lambda}{2}\nabla\xi = \boldsymbol{W}(t) - \mu\boldsymbol{\alpha}^{(e)}(\mathcal{N}_{ref}\widetilde{\boldsymbol{S}})^*, \tag{3.48}$$

where t is the iteration index and λ is the step size. Thus, the updated driving signal of the ℓ-th loudspeaker at time iteration t is given by

$$\boldsymbol{Y}(\ell, t) = \mathcal{N}_{ref}w(\ell, t), \tag{3.49}$$

where $w(\ell, t)$ is the ℓ-th element of $\boldsymbol{W}$ at iteration t.

This adaptive algorithm also works with multiple noise sources and multiple reference signals, where $\mathcal{N}_{ref}$ in (3.44) becomes a vector and $\boldsymbol{W}$ becomes a matrix with each row representing the loudspeaker weights with regards to one reference signal [9].

We note that the adaptive filtering algorithm remains the same with different loudspeaker geometries since the update equations for obtaining the loudspeaker weights (3.48) do not rely on the spherical harmonic or azimuth harmonic decomposition of the loudspeaker aperture functions. Thus, the proposed algorithm can also work with conventional spherical loudspeaker arrays or any geometry of loudspeakers. Furthermore, the proposed geometry in Sec. 3.3.2 is guaranteed to be able to control any noise field with its recorded spherical harmonic coefficients $\boldsymbol{\alpha}^{(e)}$. However, using fewer loudspeakers or a different geometry may not reach the same ideal ANC performance over the region of interest. The highest performance of a certain loudspeaker geometry relies on its sound field reproduction ability in spherical harmonic domain, which can be estimated by [134].

3.5.2 Simulation results

In this section, we first evaluate the noise reduction performance of the proposed multiple circular array geometry (MCAD) over the x-y, x-z, and y-z planes. Then, we compare its overall noise reduction performance and the convergence speed inside the full 3D region against the conventional spherical array geometry (SAD) [90], with the same adaptive algorithm detailed in Sec. 3.5.1. Finally, we test the robustness of the proposed MCAD method by evaluating the overall noise reduction performance inside the region with random positioning errors and calibration errors.

Simulation Setting: The radius of the region of interest is set to be $r_\Omega = 0.25$ m, where a human head can fit within the lower hemi-sphere. In general, ANC systems are targeting noise signals up to 400 Hz, with passive control being more effective with high-frequency noises. Hence, we construct an $N = \lceil kr \rceil = 2$-nd order system. A single noise source is located at (3 m, $\pi/4$, $\pi/4$), with a reference microphone nearby. The noise is set to be a combination of 250 Hz and 350 Hz mono-tone signals. A 60 dB signal to noise ratio (SNR) white Gaussian noise is added to each microphone, where the 'signal' in SNR refers to the residual noise signal at the microphone, and the 'noise' in SNR refers to the thermal noise of the microphone itself. We test the system in both free-space and a room environment. The room environment is simulated by the image source method [133], with 5 m $\times$ 4 m $\times$ 3 m dimensions and wall reflection coefficients of 0.7, which is considered to be a reasonable simulation of an office room.

Array Description: We apply multiple circular arrays to construct the proposed MCAD method according to our discussion in Sec. 3.3. A circular array with 5 microphones is set on the x-y plane with a radius of 0.25 m. A second circular array with 5 microphones is set at $\theta = \pi/3$ with a radius of 0.25 m. A single microphone is placed on the z-axis with a height of 0.12 m, and the final single microphone is set at the origin. A circular loudspeaker array with 7 loudspeakers is set on the x-y plane with a radius of 2 m. A circular array with 4 loudspeakers is set at $\theta = \pi/3$ with a radius of 2 m. A single loudspeaker is placed on the z-axis with a height of 1.5 m. In total, we use 12 microphones and 12 loudspeakers, all of them are within the upper hemi-sphere, leaving the lower hemi-sphere free of microphones and loudspeakers, as shown in Fig. 3.1.

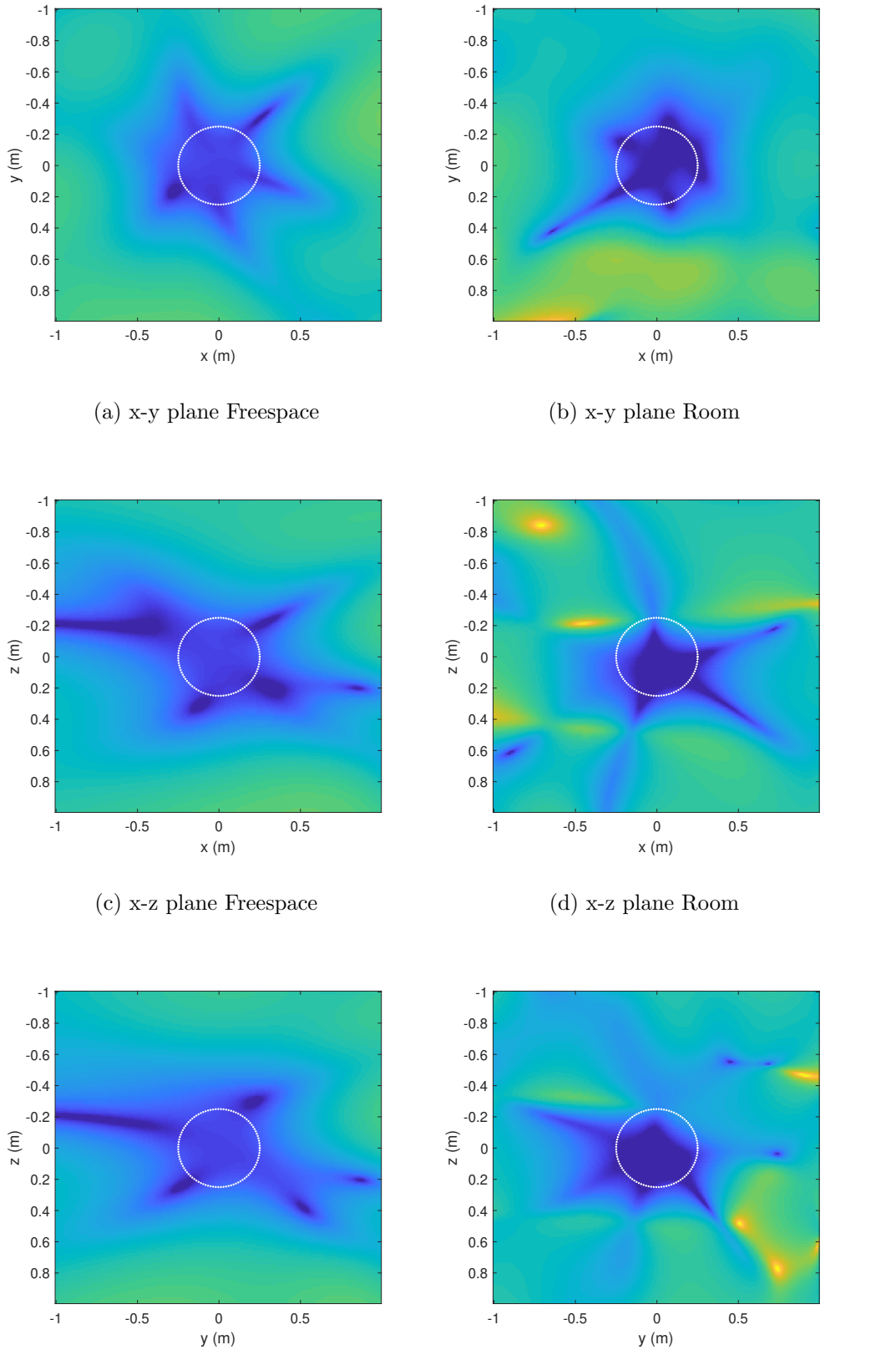
For the SAD method, we use a spherical microphone array with 9 microphones regularly distributed on a sphere of 0.25 m radius. We place 9 loudspeakers with the same geometry of microphones on a concentric sphere with a radius of 1 m.

Noise Reduction Performance: We first evaluate the noise reduction performance of the MCAD method at different planes. Figure 3.2 shows the noise reduction performance at the x-y plane (a,b), x-z plane (c,d), and y-z plane (e,f) in a free-space and room environment, respectively. The noise reduction performance is defined as

$$\eta(\boldsymbol{x}) = 10\log_{10}\frac{\sum_k E\{|e(\boldsymbol{x},k)|^2\}}{\sum_k E\{|d(\boldsymbol{x},k)|^2\}}, \tag{3.50}$$

where $e(\boldsymbol{x},k)$ is the residual noise pressure at the point $\boldsymbol{x}$, $d(\boldsymbol{x},k)$ is the original noise pressure without ANC at the point $\boldsymbol{x}$, and $E\{\cdot\}$ is referred to as the expectation. We observe that the MCAD method can reduce the noise level in the x-y, y-z, and x-z planes within the focus radius for both the free-space and the simulated room. Hence, the MCAD method is considered to achieve ANC over the whole region of interest. Performance inside the room is worse than in free-space, as the sound field is more complicated due to reverberation.

Robustness to Microphone Noise: Figure 3.3 shows the variation of average noise reduction performance over the region of interest during the noise reduction process over time (iterations), with the microphone SNR to be 40 dB and 60 dB. This result further supports that both the SAD method and the MCAD method can achieve ANC inside the whole 3D region in both free-space and in a room. While the performance of the MCAD method is not as stable as the SAD method in a room, it still achieves a more than 20 dB noise reduction after convergence, which is considered to be satisfactory performance. The convergence speed of the MCAD method becomes slower after a few iterations than the SAD method. Meanwhile, the MCAD method is more sensitive to the SNR, as we observe that the noise reduction level of the MCAD method keeps fluctuating during and after convergence. The fluctuation of the performance is relatively large with a high SNR. This is because the MCAD method solves more minimization problems than the SAD method. In the MCAD method, (3.15), (3.19) and (3.21) are required to be solved in order to obtain a set of spherical harmonic coefficients. Each equation involves matrix inversion, which is more sensitive to measurement errors.



(a) x-y plane Freespace

(b) x-y plane Room

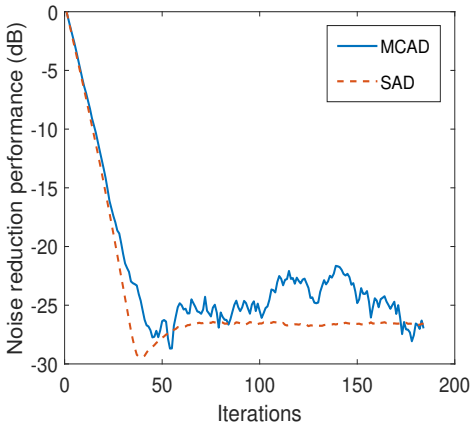
(c) x-z plane Freespace

(d) x-z plane Room

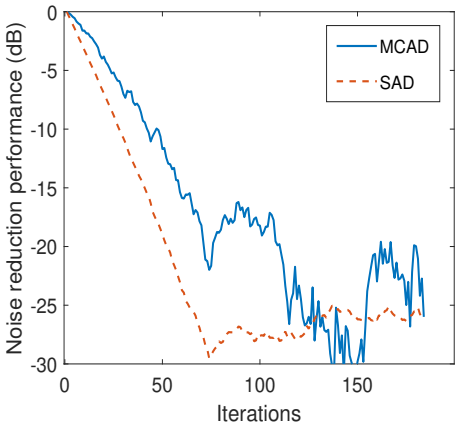
(e) y-z plane Freespace

(f) y-z plane Room

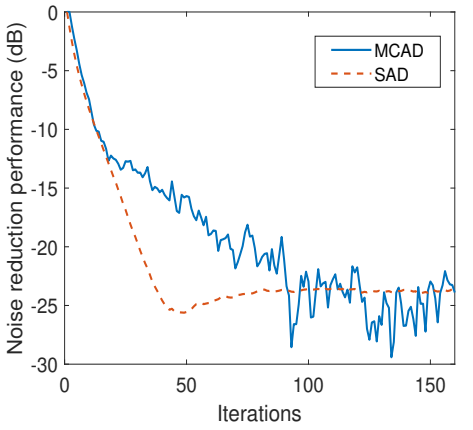
Figure 3.2: Noise reduction performance of the MACD method in the free-space



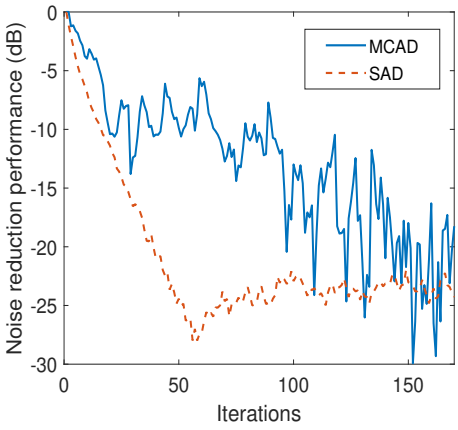
(a) Free space with SNR = 60 dB



(b) Free space with SNR = 40 dB

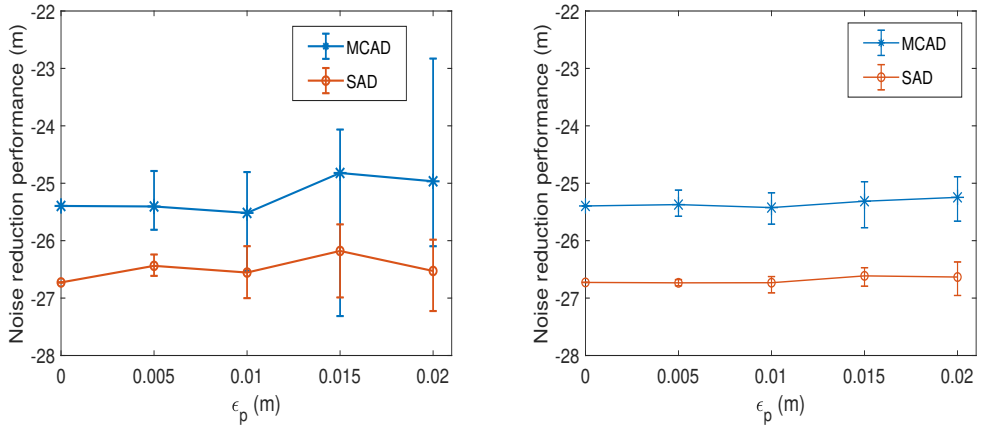


(c) Room with SNR = 60 dB



(d) Room with SNR = 40 dB

Figure 3.3: Average noise reduction performance of MCAD method and SAD method inside the whole region in free-space (a,b) and in room (c,d) over iterations with different SNR.



(a) Position error on microphones

(b) Position error on loudspeakers

Figure 3.4: Average noise reduction performance of the MCAD method and the SAD method against the positioning error with a length of ϵ_p and a random direction added to microphones (a) and loudspeakers (b).

Moreover, the calculation of odd coefficients in the MCAD method is reliant on the correct calculation of even coefficients in (3.21). A higher SNR leads to a higher error when we estimate even coefficients by (3.15), and hence a higher error in odd coefficient estimation by (3.21). Whereas, the SAD method calculation of different order/mode coefficients are independent to others.

Robustness to Positioning errors: Figure 3.4 shows the change of average noise reduction performance against positioning errors. We add a positioning error of length ϵ_p with a random direction on the location of each microphone or loudspeaker. We then estimate the average noise reduction performance over the region of interest after convergence. We repeat the simulation several times to evaluate the average performance. We observe that the average noise reduction performance is about the same when ϵ_p increases in a limited range. However, the distribution spreads more as ϵ_p increases, resulting in more randomness on the noise reduction performance. Compared to the SAD method, the noise reduction performance of the MCAD method is influenced more by the positioning error. We argue that although the proposed MCAD method has less tolerance on positioning error, the noise reduction performance still varies within an acceptable range

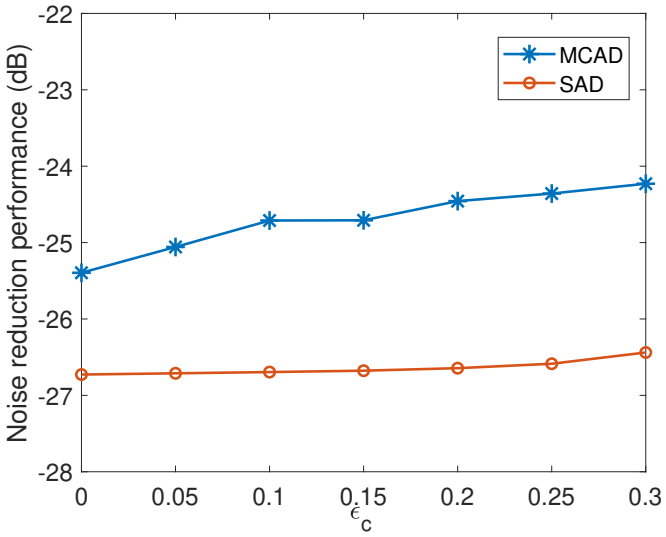


Figure 3.5: Average noise reduction performance of the MCAD method and the SAD method against the Gaussian distributed random microphone calibration errors between $[1-\epsilon_c, 1+\epsilon_c]$.

(more than 20 dB with 2 cm positioning error on random direction). Meanwhile, the positioning error of loudspeakers is not as significant as the positioning error of the microphones, for both the MCAD method and the SAD method. This is within our expectation as the proposed adaptive filtering algorithm does not rely on the circular/spherical aperture function of the loudspeaker arrays, but instead highly relies on the recording accuracy of the spherical harmonic coefficients of the residual sound field. Thus, accurate positioning of error microphones contributes more towards noise reduction performance.

Robustness to Microphone Calibration errors: Figure 3.5 shows the average noise reduction performance against microphone calibration error. We estimate the microphone calibration error by adding a Gaussian distributed random weight between $[1-\epsilon_c, 1+\epsilon_c]$ to the recorded signal. We observe that the calibration error results in a decrease of the noise reduction performance for both methods. The performance of the proposed MCAD method decreases more when ϵ_c increases, compared to the SAD method. However, this decrease on performance is tolerable when the calibration error is in a reasonable range, less than 30% in this simulation.

Overall, we observe that the proposed MCAD method is more sensitive with

measuring error (higher SNR, higher positioning error, and higher calibration error) than the SAD method. The noise reduction performance of the MCAD method decreases and fluctuates more than the SAD method with higher system errors. However, this decrease and fluctuation of the performance is within a acceptable range. The noise reduction performance remains to be satisfactory at more than 20 dB. As the proposed MCAD method has a realistic geometry of microphones and loudspeakers, which is an important advantage over the SAD method, we argue that the proposed method is preferred to a conventional method. In the next section, we will present experimental results with uncalibrated microphones, to further support the proposed system's feasibly and capability in achieving spatial noise reduction in a realistic environment.

3.5.3 Real world experiment in a room

In this section, we evaluate our proposed method with a practical experiment. We build a second order multiple circular array spatial ANC system in a lab-room at the Australian National University with the dimension of $[3.6, 6.7, 2.8]$ m. Following the discussion in Sec. 3.3, we implement a circular array with 5 microphones on the x-y plane with a radius of 0.25 m. Another circular array with 5 microphones is set at $\theta = \pi/6$ with a radius of 0.25 m. A single microphone is set on the z-axis with a height of 0.22 m, and another single microphone is set at the origin. A circular loudspeaker array with 10 loudspeakers is set on the x-y plane with a radius of 1 m. Another two circular arrays each with 5 loudspeakers are set at $\theta = \pi/6$ and $\theta = \pi/3$ with a radius of 1 m, respectively. In total, we use 12 microphones and 20 loudspeakers, where all of them are located within the upper hemi-sphere of the region of interest, leaving the lower hemi-sphere free for users. Figure 3.6 shows the experiment setup, where the active microphones and loudspeakers are circled in red and blue, respectively. We use Grover Notting CR1 speakers (Classic Audio Designs Pty Ltd.) as secondary sources, and Dayton Audio EMM-6 Electret Measurement Microphones (Dayton Audio) for recording.

Since we are not able to place a loudspeaker on the ceiling, we instead use a circular array to replace the single loudspeaker on the z-axis. Therefore, we use more loudspeakers than required. Furthermore, with the highest frequency of the

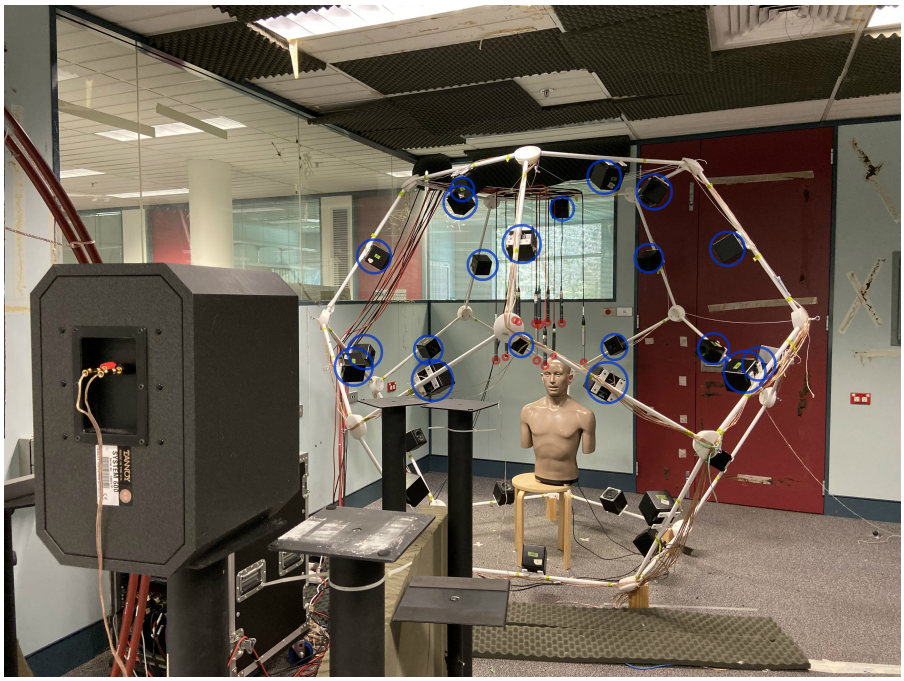


Figure 3.6: The proposed set up for spatial ANC with 12 error microphones (red circle) and 20 secondary loudspeakers (blue circle) in a lab room.

noise signal to be limited below 300 Hz, the radius of the region of interest can be up to $r_\Omega = 0.36$ m. However, it is not convenient for us to place microphone arrays on the boundary of the region of interest due to hardware constraints. We find a circle-shaped piece of wood with a radius of $r = 0.25$ m to fix the microphone positions in our lab. Therefore, we set the radius of the microphone array on the x-y plane to be $r = 0.25$ m. We reiterate here that one of the advantages of our proposed method is that the positioning of arrays are flexible within a given range, and that other configurations may be viable.

Before the experiment, we first measure the system delay by generating a clip signal with a small loudspeaker near one of the error microphones. As the acoustic delay from this small loudspeaker to the error microphone can be omitted, the delay we obtained in the measurement is considered to be purely system delay. With our set up, this delay is usually around 1219 samples. Then, we estimate the secondary channels for adaptive filtering by measuring the impulse responses from each secondary loudspeaker to the error microphone arrays with sinusoidal sweep

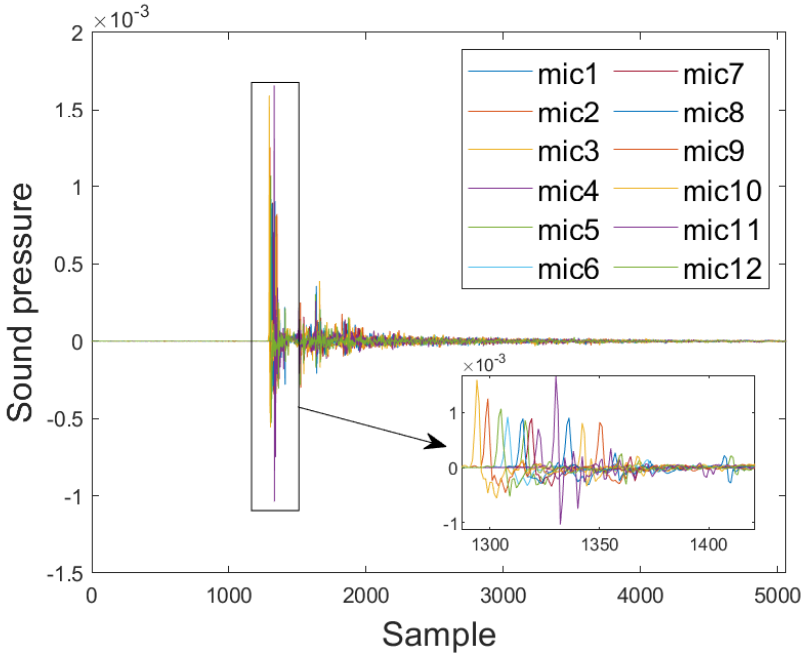


Figure 3.7: Secondary channel measurements from one secondary loudspeaker to 12 error microphones with a scatter in the region of interest.

signals. In total we get 12×20 channels. The first 1219 samples are discarded from each measurement result to remove the system delay. Figure 3.7 shows the impulse responses from one of the secondary loudspeakers to all of the error microphones. We note that as we are unable to calibrate our microphones, the amplitude of these impulse responses have more disparity than expected. The noise is generated by a single loudspeaker located 3 m away. We assume the noise signal is known instead of placing a reference microphone near the noise loudspeaker because the suppression of the feedback signal from the secondary loudspeakers to the reference microphone is out of the scope of this work. The noise is generated with a combination of 150 Hz and 200 Hz mono-tone signals. Sampling frequency is set to be 48000 Hz.

We evaluate our proposed method with the above set-up in two scenarios: 1) Free-space inside the region of interest; 2) A human head model (dummy head) [135] is placed inside the region of interest.

Figure 3.8 shows the energy of the signal (square of the average amplitude) recorded by the error microphones in scenario one. The proposed method is enabled

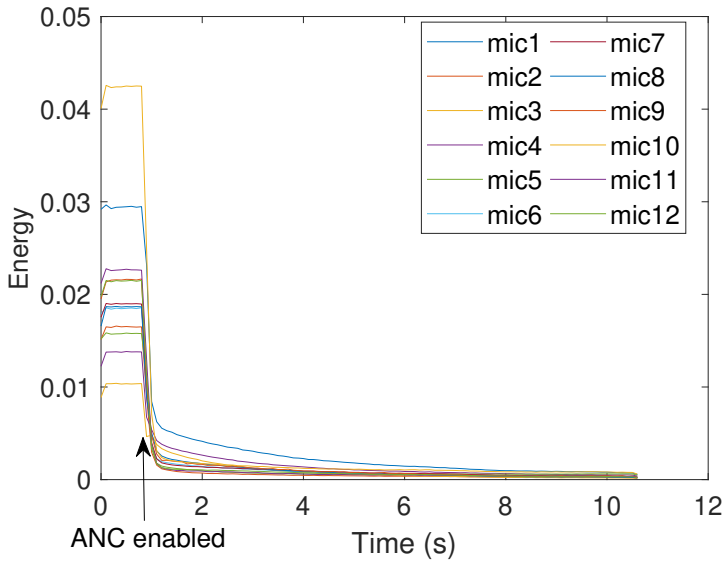
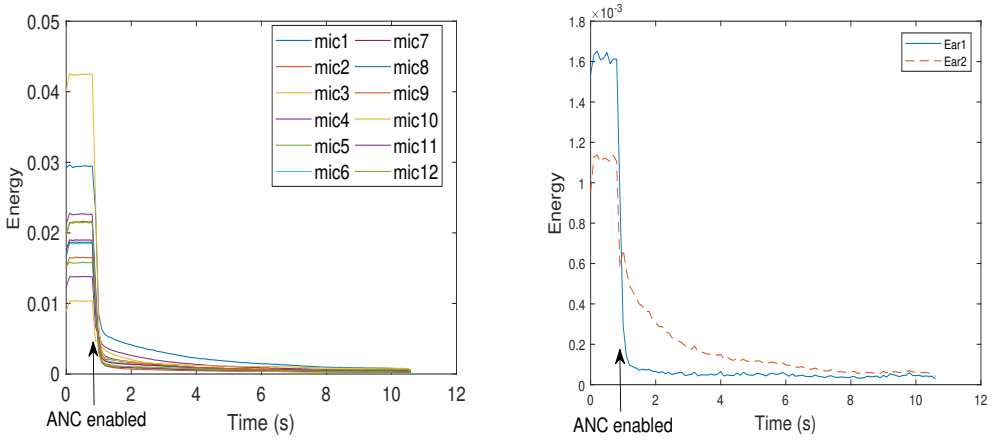


Figure 3.8: Sound energy measurements at 12 error microphones in a free-space region of interest with ANC enabled at $t = 1$ s.

at $t = 1$ s. We observe that the energy of the recorded signals decreases at all of the error microphones, which illustrates the effectiveness of our proposed method.

We then place a dummy head in the region of interest and modify our proposed method with scattering considerations according to Sec. 3.4. Figure 3.9 shows the energy of the signal recorded (a) by the error microphones and (b) by the microphones in the ears of the dummy head. The proposed method achieves a reduction in the noise signals at both error microphones and dummy head ears. Figure 3.10 shows the frequency spectrum amplitude of the original noise signals and the reduced signals recorded at the ears of the dummy head. We observe an average 9.5 dB noise reduction at the focused frequency bins, proving that the proposed method is reliable for spatial ANC in the real world environment. The ANC performance in the experiment is lower than the simulation. We consider this reduction of performance to be reasonable due to the error introduced by the secondary channel measurements and the unstable environment (temperature, people movement, etc..).

The influence of the scattering effects of the dummy head is then investigated by deriving the secondary loudspeaker driving signals with the scattering method



(a) (b)

Figure 3.9: Sound energy measurements (a) at 12 error microphones and (b) at ears on a dummy head with ANC enabled at $t = 1$ s.

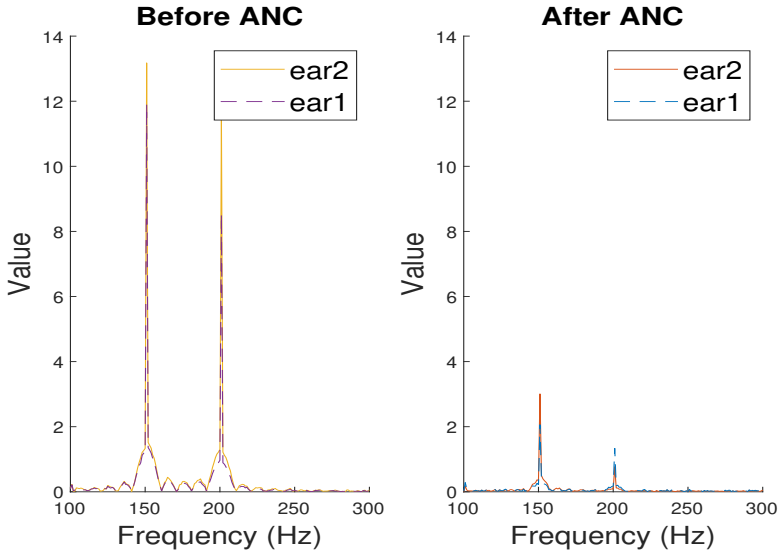


Figure 3.10: Frequency spectrum of the noise signals before ANC of $150, 200$ Hz and the residual signals after ANC at both ears of the dummy head.

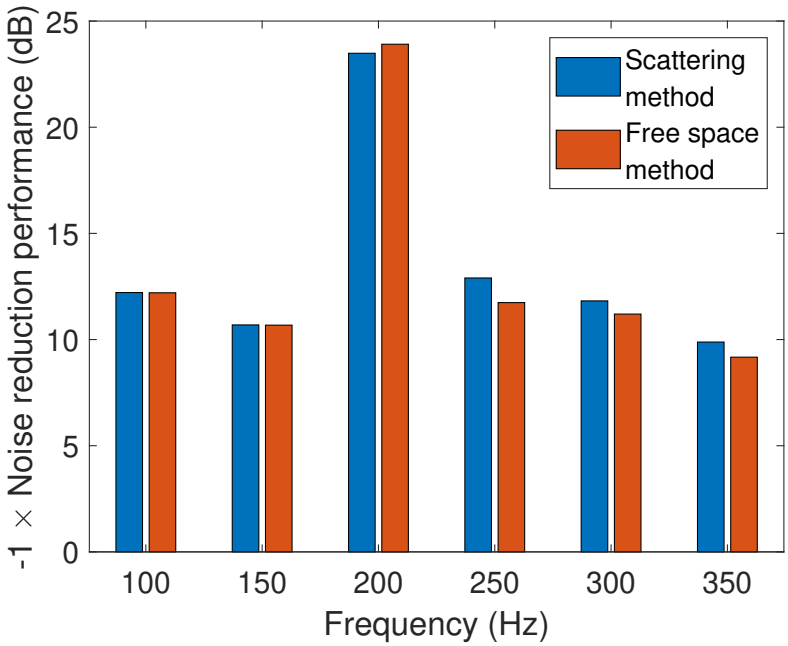


Figure 3.11: Noise reduction performance with the scattering method and the free space method averaged over two ears of the dummy head at different frequencies.

detailed in Sec. IV and the free space method, respectively. We estimate the residual noise level at the dummy head ears in the region of interest with both methods over different frequencies by using noise signals of $100, 150, 200, 250, 300,$ and 350 Hz, respectively. Figure 3.11 shows the noise reduction performance with the scattering method and the free space method averaged over the two ears of the dummy head. Note that we show the noise reduction performance with positive numbers in Fig. 3.11 such that higher bars mean better performances. We observe from Fig. 3.11 that the noise reduction performance at the dummy head ears with the scattering method is higher than the free space method at most of the frequencies. However, this difference between the two methods is not significant. We observe a discrepancy at 200 Hz, which we consider that it is due to the experimental errors. Hence, we believe that the scattering effect of the human head is negligible such that the region of interest can be considered as a free space. The scattering effect by the human head is not significant because 1) the size of the human head is smaller than the wavelength of the noise signal's frequency contents;

2) the residual noise level in the region of interest is low after ANC, resulting in little scattering effect. However, we think the proposed work regarding scattering effect is still valuable because it provides us the confidence to omit the scattering due to the human head.

3.6 Conclusions

In this chapter, we address the spatial ANC problem for a 3D region with multiple circular microphone arrays and loudspeaker arrays. We construct a second order system in both simulation and experiment to demonstrate the effectiveness of the proposed method on noise reduction over space, and illustrate by experiment that the scattering effect of the human head is negligible. Although we use a few more microphones and loudspeakers than conventional spherical array systems, the proposed method is advantageous in its circular array design that is isolated to the upper hemi-sphere. This provides a practical design that is easier to position and implement, while also freeing the lower hemi-sphere for user movement. As a result, the proposed method in this chapter is a step forward in realizing spatial ANC for real world applications.

3.7 Related Publications

Much of this chapter's work has been published in the following journal papers and conference proceedings.

- H. Sun, T. D. Abhayapala, and P. N. Samarasinghe, "A Realistic Multiple Circular Array System for Active Noise Control Over 3D Space," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 3041 - 3052, 2020.

- H. Sun, T. D. Abhayapala and P. N. Samarasinghe, "Active Noise Control Over 3D Space with Multiple Circular Arrays," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 135-139, New York, USA, October, 2019.

Chapter 4

ANC over 3D space with remote microphone technique in the wave domain

Overview: *Conventionally, a spatial ANC system is achieved by using error microphones evenly distributed over the region of interest, or evenly on the surface of a spherical region with the spherical harmonic analysis of spatial sound fields in the wave domain. A major disadvantage for these systems comes from the error microphone arrays, which have bulky cumbersome designs that block the user movement into the region. In Chapter 3, we proposed a method using multiple circular arrays to over come this problem. In this chapter, we propose an alternative method to overcome this problem by introducing the remote microphone technique where the error signals are estimated at locations that are remote from the physical error microphones. We develop a wave domain ANC system based on the remote microphone technique for noise reduction over a spherical region. The error microphones are located away from the region of interest when the noise reduction is achieved. By simulation, we demonstrate that the proposed method can obtain comparable ANC performance to the conventional methods without remote microphone technique, while being more feasible in practice.*

4.1 Introduction

Conventionally, a spatial ANC system is achieved by a multi-input-multi-output (MIMO) adaptive filtering system utilizing multiple error microphones and multiple secondary loudspeakers, where the error microphones are required to be uniformly distributed inside the control region [19]. Wave domain signal processing using harmonic (cylindrical/spherical) based sound field processing has been recently applied in ANC to decrease the use of resources by placing error microphones on the boundary of the region of interest [73]. While the wave domain methods are effective for 3D space noise reduction [90], there is a prerequisite of regularly spacing the error microphones on a certain sphere [123]. Whereas, this spherical microphone array surrounding the region of interest blocks the users entering and exiting the region. Hence, an advanced method with more realistic error microphone placements to achieve spatial ANC is worth considering.

Virtual Sensing is one of the solutions [23]. Given the physical error signals recorded by the physical microphones, the noise signals, and knowledge of the system, virtual sensing algorithms minimize the noise signals at the virtual microphone locations instead of the physical microphone locations. Noise reduction is achieved based on MIMO feed-forward adaptive filtering at human ears [66] and over 2D space [136] with microphones away from the desired noise reduction point (region) during the noise control stage.

There are two popular virtual sensing algorithms applied to ANC systems: Auxiliary Filter based Virtual Sensing (AF-VS) method [67] and Remote Microphone (RM) method [137]. AF-VS method shows a better noise reduction performance in some cases with wide band noise [138]. However, this method utilizes an auxiliary filter which relies on the secondary channels between the secondary loudspeakers and the physical microphones. In a wave domain Fx-LMS adaptive filtering ANC system, these physical secondary channels are required to be measured in the wave domain. In this case, if we want to develop a wave domain (spherical harmonic based) AF-VS method, the physical microphones, as well as the virtual microphones, have to be uniformly distributed spherical arrays. These spherical arrays are against our motivation to have a reliable geometry for users' movements. On the other hand, RM method does not rely on the spatial modelling of physical sec-

ondary channels [139]. A filter is pre-measured in the tuning stage to compensate the differences between the primary channels of the physical microphones and the virtual microphones, which also shows a capacity of wide-band noise reduction [24]. Very recent work [140] has mentioned a RM method for ANC in a 2D environment, which supports our idea that RM method has higher potential in 3D spatial ANC than AF-VS method.

In this chapter, we develop a wave-domain remote microphone method to achieve ANC over a spherical region of interest. The virtual microphones are located as a spherical array surrounding the region of interest such that the recorded signals can be obtained in the wave-domain by spherical harmonic analysis. By applying the remote microphone technique, the noise signals recorded by physical microphones and the virtual microphones are related with a pre-measured wave-domain filter in the turning stage. As a result, noise reduction can be achieved over the region of interest in the control stage without the existence of the microphones inside the region. Hence, these microphones do not obstruct the movement of the users in and out of the quiet region, which is the main novelty of this work. We compare the noise reduction performance of the proposed method to the conventional method without remote microphone technique by simulation, and demonstrate that the proposed method can achieve comparable noise reduction performance over the 3D region of interest.

4.2 Problem formulation

Consider a source-free region of interest Ω in 3D space, where noise sources are outside of Ω. The noise field at an arbitrary point $\boldsymbol{x} = (r, \theta, \phi)$ inside Ω can be expressed as [19]

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{4.1}$$

where order n and mode m are integers, $N = \lceil kR \rceil$ [21], R is the radius of Ω, $\alpha_{nm}(k)$ is a set of location independent spherical harmonic coefficients representing the sound field inside Ω, $k = 2\pi f/c$ is the wave number, f is frequency, c is

the speed of sound, $j_n(\cdot)$ is the spherical Bessel function, $Y_{nm}(\cdot)$ is the spherical harmonic function [110].

The conventional way of obtaining $\alpha_{nm}(k)$ is by applying a spherical microphone array that surrounds the region, where at least $Q = (N+1)^2$ microphones are uniformly distributed over the sphere. The corresponding $\alpha_{nm}(k)$ can be extracted by integrating (summation for discrete sampling) the recording over the spherical surface while exploiting the orthogonality property of $Y_{nm}(\cdot)$ [21]

$$\alpha_{nm}(k) \approx \frac{1}{j_n(kR)} \sum_{q=1}^{Q} z(R, \theta_q, \phi_q, k) Y_{nm}^*(\theta_q, \phi_q) \chi(q), \tag{4.2}$$

where $z(R, \theta_q, \phi_q, k)$ refers to the recorded sound pressure of the q-th microphone located at (R, θ_q, ϕ_q), and $\chi(q)$ is the sampling weight of the q-th microphone, where $\chi(q)$ for uniformly sampling.

Hence, to achieve spatial ANC, we need to measure the spherical harmonic coefficients of the residual sound field with error microphones to obtain $\alpha_{nm}^{(e)}(k)$, and to produce a cancelling secondary sound field given by $\alpha_{nm}^{(s)}(k)$ in the region of interest such that the residual sound field coefficients given by

$$\alpha_{nm}^{(e)}(k) = \alpha_{nm}^{(d)}(k) + \alpha_{nm}^{(s)}(k), \tag{4.3}$$

are to be minimized through loudspeakers driving signals, where $\alpha_{nm}^{(d)}(k)$ represents the primary noise field coefficients.

Consider a spherical array of L loudspeakers located at a distance R_L from the origin of Ω with a driving signal $y(\ell, k)$ at the ℓ-th loudspeaker located at $(R_L, \theta_\ell, \phi_\ell)$. The resulting sound pressure at the point $\boldsymbol{x}$ due to this loudspeaker array is

$$z(\boldsymbol{x}, k) = \sum_{\ell=1}^{L} y(\ell, k) S(\ell, \boldsymbol{x}, k), \tag{4.4}$$

where $S(\ell, \boldsymbol{x}, k)$ is the frequency response at frequency bin k from the location of the ℓ-th loudspeaker to the point $\boldsymbol{x}$.

In a typical feed-forward adaptive ANC system, the driving signal $y(\ell, k)$ is obtained by filtering the reference noise signal $\mathcal{N}(t)$ using a FIR adaptive filter

$w_\ell(k)$ such that

$$y(\ell, k) = \mathcal{N}(k) w_\ell(k). \tag{4.5}$$

Hence, we have the secondary sound field coefficients $\alpha_{nm}^{(s)}(k)$ given by

$$\alpha_{nm}^{(s)}(k) = \sum_{\ell=1}^{L} \mathcal{N}(k) w_\ell(k) \widetilde{S}_{nm}(\ell, k), \tag{4.6}$$

where $\widetilde{S}_{nm}(\ell, k)$ refers to the nm-th mode of the spherical harmonic decomposition of the pre-recorded $S(\ell, q, k)$ at the error microphones.

Define the cost function of the adaptive filter to be

$$\xi(t) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \|\alpha_{nm}^{(e)}(t, k)\|^2, \tag{4.7}$$

where $\alpha_{nm}^{(e)}(t, k)$ is the recorded spherical harmonic coefficients by the error microphone array at the t-th iteration. With (4.3, 4.6, 4.7), we have the update equation of the weight $w_\ell(k)$ to be [90]

$$w_\ell(t+1, k) = w_\ell(t, k) - \lambda \mathcal{N}(k) \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}^{(e)}(t, k) \widetilde{S}_{nm}(\ell, k), \tag{4.8}$$

where λ is the step size.

In (4.5) and (4.8), it is shown that the driving signal of a wave domain spatial ANC system relates to the coefficients of the residual signal recorded by the error microphone array and the secondary channels. However, the spherical microphone array is not reliable as it will block the users to enter or exit the region. In the following sections, we solve this problem by introducing the RM technique into the system.

4.3 Wave domain remote microphone method

To achieve noise reduction with RM, the system is separated into two stages: the Tuning Stage and the Control Stage, as shown in the Fig. 4.1. A group of filters

which contains the information of the primary channels of the physical microphone locations and the virtual microphone locations (points of interest) are estimated in the Tuning Stage. Then, this filter is applied in the Control Stage to reduce the noises at the virtual microphone locations with the recordings of the residual noise at physical microphones.

The block diagram of an adaptive feed-forward Fx-LMS filtering ANC system with the wave domain RM method is shown in Fig.4.1. The 'SH' block, colored yellow, refers to the spherical harmonic decomposition of the input signals, given in Eq.(4.2). By this block, the input signals are decomposed into a series of spherical harmonic coefficients with respect to the origin (centre of the region of interest). According to the discussion in Sec.4.2, these wave domain coefficients are location independent, which contain the information of the input sound field in the region of interest. We use $\widetilde{(\cdot)}$ to represent that a variable is given in the wave domain by spherical harmonic coefficients, and $\widehat{(\cdot)}$ to represent the estimation value of a variable.

In the tuning stage, a group of filters $\widetilde{\boldsymbol{C}}$ is estimated in the wave domain to compensate the difference between the primary channels of the virtual microphones and the primary channels of the physical microphones. The optimal solution is given by

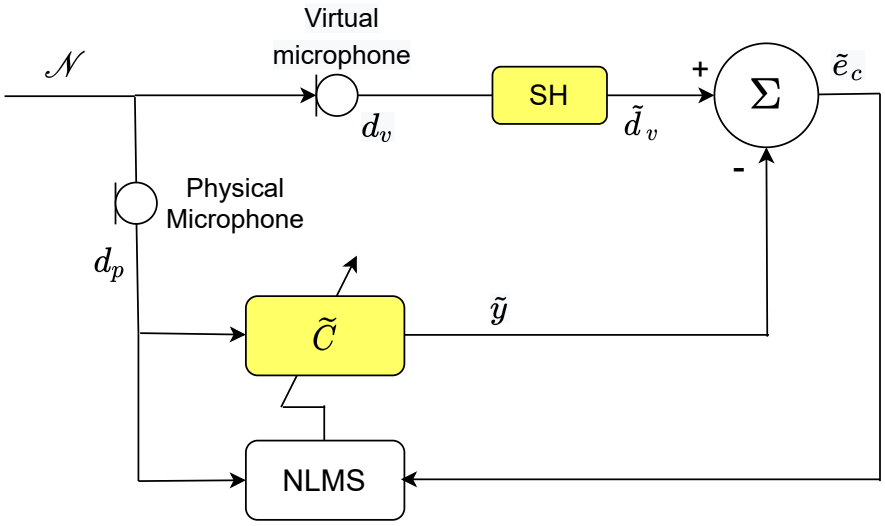
$$\widetilde{\boldsymbol{C}} = \boldsymbol{d}_p^{\dagger}\widetilde{\boldsymbol{d}}_v, \tag{4.9}$$

where $\boldsymbol{d}_p$ is the vector of the recorded signals from the physical microphones, $\widetilde{\boldsymbol{d}}_v$ is the vector of the wave domain recording of the virtual microphones, and $(\cdot)^{\dagger}$ refers to the Moore-Penrose inverse. We omit the frequency component k from now on for simplicity.

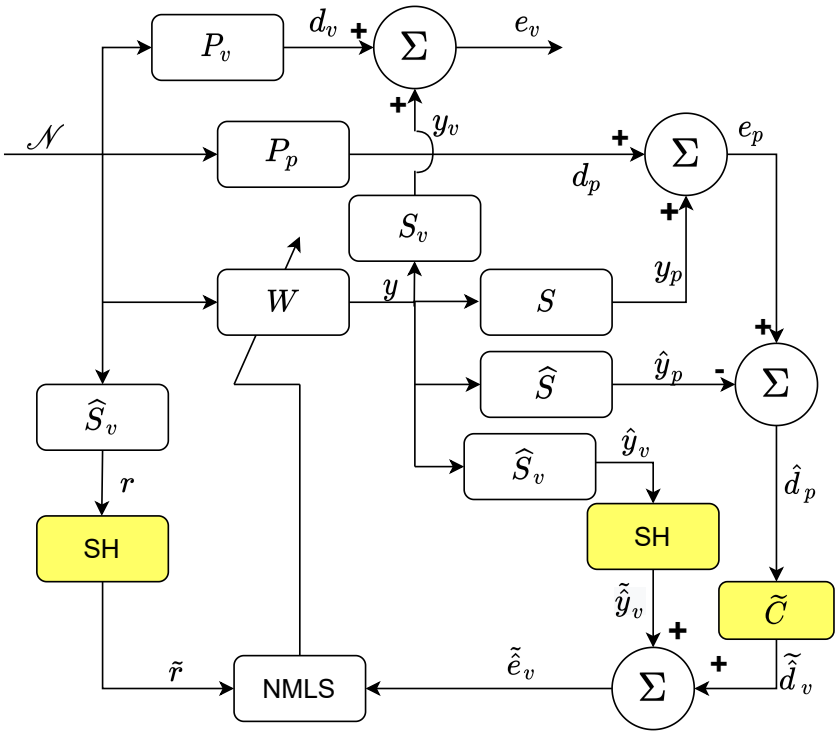
In the control stage, we can not obtain the residual sound field in the region of interest by recording $\widetilde{\boldsymbol{e}}_v$, due to the lack of the microphones at virtual positions. Hence, we record the sound pressure at the physical microphone locations as $\boldsymbol{e}_p$ to estimate the primary noise at the physical locations as

$$\widehat{\boldsymbol{d}}_p = \boldsymbol{e}_p - \widehat{\boldsymbol{y}}_p = \boldsymbol{e}_p - \mathcal{N}W\widehat{\boldsymbol{S}}, \tag{4.10}$$

where $\widehat{\boldsymbol{y}}_p$ is the estimation of the secondary sound pressure at the physical locations, $\mathcal{N}$ refers to the noise signal given by the reference microphone, $W =$



(a) Tuning stage



(b) Control stage

Figure 4.1: Block diagram of an adaptive feed-forward spatial ANC system with wave domain remote microphone technique, to minimize the error signals at virtual microphones e_v from the signals recorded at physical microphones e_p.

$[w_1, w_2, \cdots, w_L]^T$ is a group of adaptive filter weights for the driving signals of the secondary loudspeakers, $\widehat{\boldsymbol{S}}$ is the estimation of the secondary channels of the physical microphones. With (4.9), the estimated primary noise coefficients of the region of interest is given by:

$$\widetilde{\widehat{\boldsymbol{d}}}_v = \widetilde{\boldsymbol{C}}\widehat{\boldsymbol{d}}_p = \widetilde{\boldsymbol{C}}\boldsymbol{e}_p - \widetilde{\boldsymbol{C}}\mathcal{N}W\widehat{\boldsymbol{S}}. \tag{4.11}$$

Therefore, we have the estimated residual sound field coefficients of the region of interest $\widetilde{\widehat{\boldsymbol{e}}}_v$ with the estimated virtual secondary channel coefficients $\widetilde{\widehat{\boldsymbol{S}}}_v$ as

$$\widetilde{\widehat{\boldsymbol{e}}}_v = \widetilde{\widehat{\boldsymbol{d}}}_v + \mathcal{N}W\widetilde{\widehat{\boldsymbol{S}}}_v = \widetilde{\boldsymbol{C}}\boldsymbol{e}_p - \widetilde{\boldsymbol{C}}\mathcal{N}W\widehat{\boldsymbol{S}} + \mathcal{N}W\widetilde{\widehat{\boldsymbol{S}}}_v. \tag{4.12}$$

Define the cost function of the adaptive filter as

$$\xi(t) = \widetilde{\widehat{\boldsymbol{e}}}_v\widetilde{\widehat{\boldsymbol{e}}}_v^T, \tag{4.13}$$

where $(\cdot)^T$ refers to the conjugate transpose of a matrix. The update equation is given by

$$w_\ell(t+1) = w_\ell(t) - \frac{\lambda}{2}\nabla\xi(t), \tag{4.14}$$

where $\nabla\xi(t)$ is the derivative of $\xi(t)$ with respect to $w_\ell(t)$, which is given by

$$\begin{aligned}\nabla\xi(t) &= \frac{\partial\xi(t)}{\partial w_\ell(t)} = 2\widetilde{\widehat{\boldsymbol{e}}}_v\frac{\partial(\widetilde{\widehat{\boldsymbol{e}}}_v)^T}{\partial w_\ell(t)} = 2\widetilde{\widehat{\boldsymbol{e}}}_v\frac{\partial(\widetilde{\widehat{\boldsymbol{d}}}_v + \mathcal{N}W\widetilde{\widehat{\boldsymbol{S}}}_v)^T}{\partial w_\ell(t)} \\ &= 2\mathcal{N}\widetilde{\widehat{\boldsymbol{e}}}_v\widetilde{\widehat{\boldsymbol{S}}}_v^T.\end{aligned} \tag{4.15}$$

Hence, we have

$$w_\ell(t+1) = w_\ell(t) - \lambda\mathcal{N}\widetilde{\widehat{\boldsymbol{e}}}_v\widetilde{\widehat{\boldsymbol{S}}}_v^T. \tag{4.16}$$

We observe that the update equation of the proposed remote microphone method (4.16) is similar to the update equation of the conventional method (4.8). However, in the proposed method, the spherical harmonic coefficients of the residual noise field is given by (4.12), which is related to the recorded sound pressures $\boldsymbol{e}_p$ instead of the direct recording at virtual positions $\boldsymbol{e}_v$.

Additionally, by introducing the RM technique into spatial ANC systems, a

causality problem appears. Thus, the distance between the physical microphones and the noise source should be smaller than the distance between the virtual microphones and the noise source such that the filters $\widetilde{\boldsymbol{C}}$ are causal. In the proposed method, a spherical array is applied as virtual microphones which surround the region of interest. The physical microphones to sample over all directions is also desired such that the noise can be properly recorded wherever the noise source is. In this case, the causality of filters $\widetilde{\boldsymbol{C}}$ may not be satisfied. The solution of this problem is given in [68], where a delay block is added in the filters $\widetilde{\boldsymbol{C}}$ and at the virtual secondary channels. It is also shown in [68] that the use of feed-forward adaptive filtering has its advantage over feed-backs on the causality issue of the RM technique.

4.4 Simulation and results

In this section, we evaluate the noise reduction performance of the proposed method by simulation, and compare the result with the conventional method where error microphones are physically located at the virtual microphone locations as a spherical array.

To control the noise below 600 Hz, a 4-th order system is constructed with a region of interest of the radius $r = 0.3$ m. The noise source is located at $(1.5, 1.7, 0.5)$ m, with a reference microphone nearby. We assume the reference microphone records the reference noise signal without any disturbance[1]. A 40dB signal to noise ratio (SNR) white Gaussian noise is added to each microphone (virtual in the Tuning stage and physical in both stages). We test the system in a room environment with 5 m $\times$ 6 m $\times$ 4 m in size and with 0.8 for wall reflection coefficients. The room environment is simulated by the image source method [133]. A spherical loudspeaker array concentric with the region of interest of radius $R = 1.5$ m with 32 loudspeakers evenly distributed on the surface is applied as secondary loudspeakers.

For the proposed method, we need both physical microphone array and virtual microphone array. We use 32 microphones uniformly distributed on the surface of the region of interest as virtual microphone array. Note that this microphone

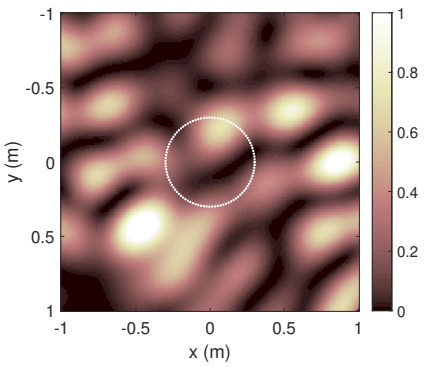
[1]The controlling of the feedback signal from the secondary loudspeakers is out of the scope of this work.

array is not applied in the control stage during the noise reduction. At least the same amount of physical microphones should be applied to make the filter $\widetilde{\boldsymbol{C}}$ a under-determined system. However, a system with less physical microphones is able to work when the noise environment is simple, as long as the adaptive filter at tuning stage is able to reach convergence with little error. We apply 8 physical microphones at the 8 corners of a 0.8 m side length cubic concentric with the region of interest. For the conventional method, we apply 32 error microphones uniformly distributed on the surface of the region of interest. This error microphone array shares the same location as the virtual microphone array in the proposed method.

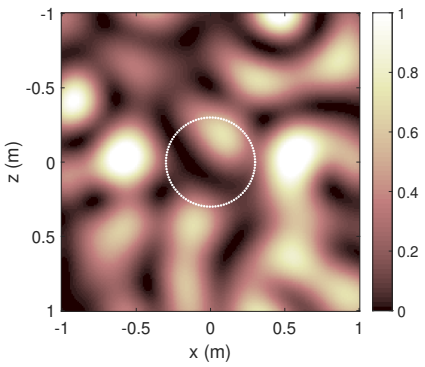
Figure 4.2 shows the sound field over the x-y plane and the x-z plane of the original noise field, the residual noise field after noise reduction by the proposed method, and the residual noise field after noise reduction by the conventional method, respectively. A single tone sine wave of 500 Hz is applied as the noise signal. The region of interest is marked by a white circle. We observe that the residual noise energy decreases inside the region of interest with both the proposed method and the conventional method, while the noise energy increases outside the region of interest due to the secondary sound field.

Figure 4.3 shows the average power spectral density of the noise inside the region of interest before and after control by the proposed method and the conventional method against the frequency of the noise signal. The frequency of the noise signal varies from 40 Hz to 1000 Hz. In total 3000 evaluation points are examined in the region of interest. We note from Fig. 4.3 that both the proposed method and the conventional method can achieve noise reduction over the region with the noise signal frequency in the range of focus. The noise reduction performance begins to decrease when the frequency is higher than 600 Hz. This is due to the 4-th order system is not able to record enough active spherical harmonic coefficients of the residual sound field when the frequency of the noise signal is higher than 600 Hz. The average performance of the proposed method is slightly worse than the conventional method due to the error introduced in the tuning stage. However, this performance difference is not significant comparing with the more than 20 dB noise reduction.

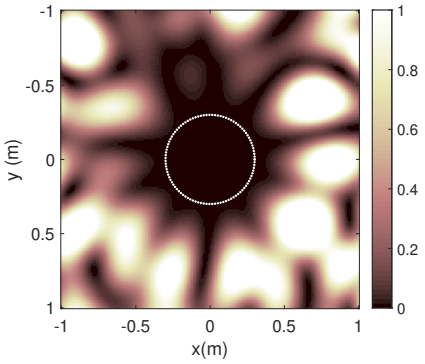
In Fig. 4.4, we further examine the average noise reduction performance in the region of interest against time (iterations) with different adaptive filtering step size



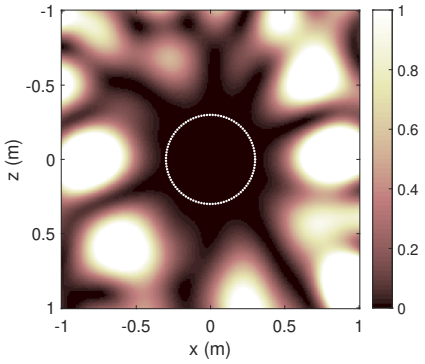
(a) Noise x-y plane



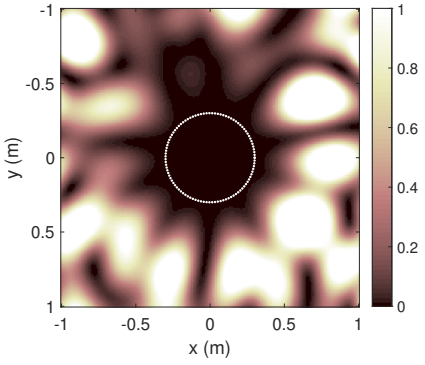
(b) Noise x-z plane



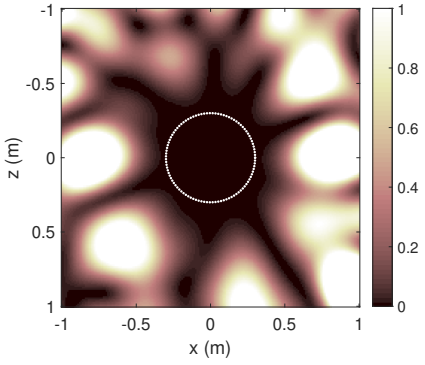
(c) Proposed x-y plane



(d) Proposed x-z plane



(e) Conventional x-y plane



(f) Conventional x-z plane

Figure 4.2: Noise field over (a) the x-y plane and (b) the x-z plane of the original noise field; the residual noise field after noise reduction by the proposed method over (c) the x-y plane and (d) the x-z plane; and the residual noise field after noise reduction by the conventional method over (e) the x-y plane and (f) the x-z plane.

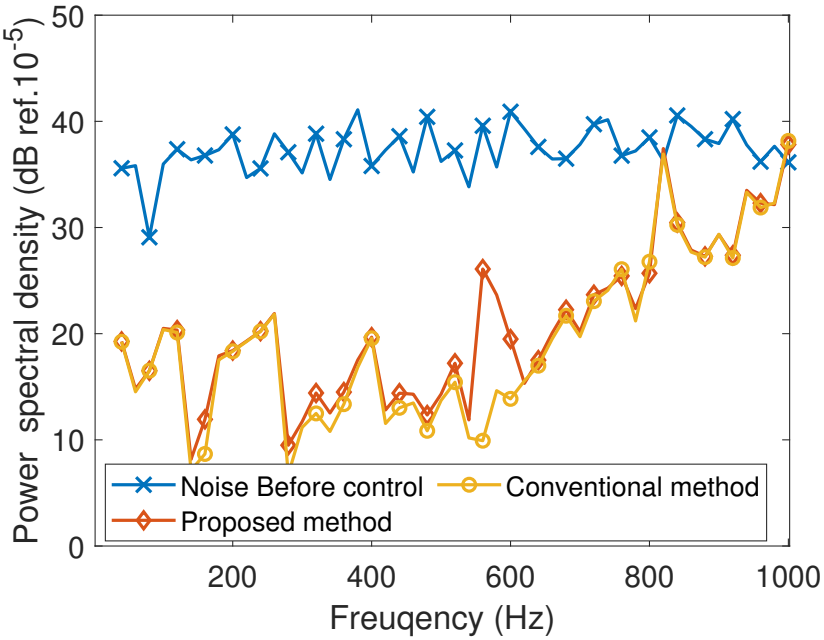


Figure 4.3: Average power spectral density of the noise inside the region of interest before control and after control by the proposed method and the conventional method with the frequency of the noise signal varies from 40 Hz to 1000 Hz.

λ. We define the average noise reduction performance as

$$\eta(k) = 10 \log_{10} \frac{\sum_{\boldsymbol{x}} E\{|e(\boldsymbol{x}, k)|^2\}}{\sum_{\boldsymbol{x}} E\{|d(\boldsymbol{x}, k)|^2\}}, \tag{4.17}$$

where $e(\boldsymbol{x}, k)$ is the residual noise pressure at point $\boldsymbol{x} \in \Omega$ and $d(\boldsymbol{x}, k)$ is the original noise without ANC. A noise signal of 500 Hz is applied as an example. The performance after convergence of the proposed method is slightly worse than the conventional method, which matches the result we obtain from Fig. 4.3. Meanwhile, the proposed method has the same speed of convergence with the conventional method if the same step size is applied to the adaptive filtering.

While the proposed method provides a similar ANC performance as the conventional method and has advantages in sensor placements, there are some drawbacks. The introduction of the remote sensing technique requires an accurate estimation of both the virtual and the physical secondary channels. Any change in the acoustic environment may result in higher performance degradation or an unstable system. To this end, investigation of extra constraints in the update equations may contribute to a more stable system, which can be one direction of future work.

4.5 Conclusions

In this chapter, we address a spatial ANC problem for a spherical region with the remote microphone technique. By simulation, we show that the proposed method provides a comparable noise reduction performance and speed of convergence with the conventional wave domain ANC method. Meanwhile, the proposed method shows its advantage that the noise reduction over space is achieved with the error microphone absent of the inside or on the surface of the region. Comparing to the conventional method where a spherical error microphone array is necessary surrounding the region of interest, the proposed method does not block the users to enter/exit the quiet region. Hence, we consider the proposed method is more realistic in implementing spatial ANC for real-world applications. Future work can include further investigations on the noise reduction performance with multiple noise sources as well as real-world experiments.

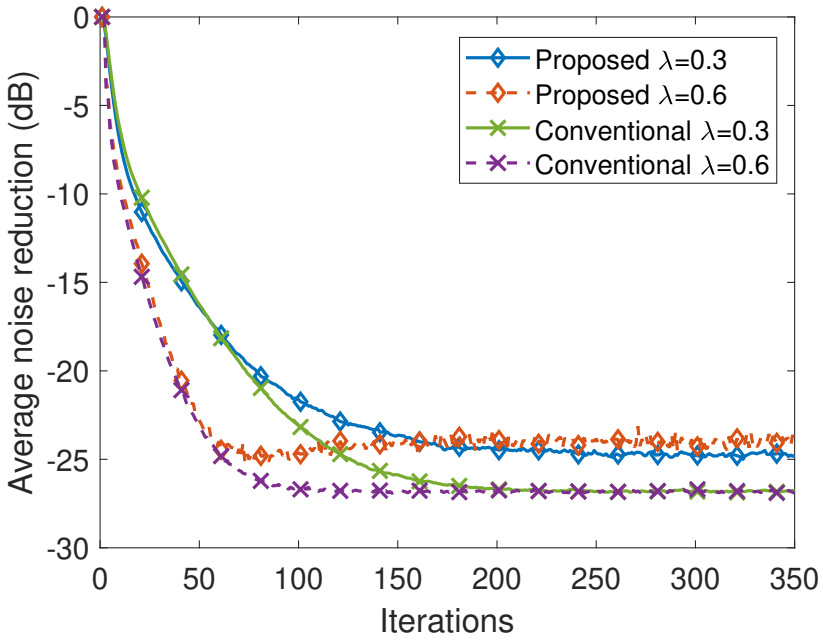


Figure 4.4: Average noise reduction performance in the region of interest with the proposed method and the conventional method over iterations with the step size of the adaptive filter $\lambda = 0.3$ and $\lambda = 0.6$.

4.6 Related Publications

Much of this chapter's work has been published in the following conference proceedings.

- H. Sun, J. Zhang, T. D. Abhayapala and P. N. Samarasinghe, "Active noise control over 3D space with remote microphone technique in the wave domain," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October, 2021.

Chapter 5

Secondary channel estimation in spatial active noise control systems using a single moving higher order microphone

Overview: *Spatial ANC systems focus on minimizing unwanted acoustic noise over continuous spatial regions by generating anti-noise fields with secondary loudspeakers. Conventionally, an error microphone array is necessary inside the region to measure the channels from the secondary loudspeakers to the error microphones, and record the residual sound fields during the noise control. These error microphone arrays highly limit the implementation of spatial ANC systems because of their impractical geometry and obstruction to the users from accessing the quiet regions. Recent advances, such as virtual sensing, focus on obtaining the residual sound field with microphones placed away from the region. While these techniques relax the usage of error microphones during the noise control, an error microphone array remains necessary during the secondary channel estimation. In this chapter, we propose a method to estimate the secondary channels without using an error microphone array. Instead, a moving higher order microphone, such as an Eigenmike, is applied to obtain the impulse responses within the desired frequency range from the secondary loudspeakers to the entire region of interest, which includes all desired*

error microphone locations. By simulation, we show that the proposed method is robust against various measuring errors introduced by the movement of the higher order microphone, and is suitable for the secondary channel estimation in spatial ANC systems.

5.1 Introduction

Conventionally, a MIMO (multiple-input multiple-output) spatial ANC system requires the error microphones to be uniformly distributed inside the control region [19] or regularly distributed on a surface enclosing the desired region. [118]. This error microphone array records the residual sound field at the multiple receiver points of interest, and also measures the secondary channels from the secondary loudspeakers to the target error microphone locations.

However, these error microphone arrays are difficult to assemble and often obstruct users from accessing the quiet zones, which is one of the main drawbacks prohibiting practical implementation of a spatial ANC system. To address this problem, virtual sensor techniques have recently been applied in spatial ANC systems [24]. By applying the remote microphone methods or virtual sensing methods, noise reduction can be achieved at human ears [138] and over space [67] while the microphones are located remotely away from these desired noise reduction points (or region). However, prior to the noise reduction, error microphone arrays are still necessary in the region of interest for secondary channel estimations [119]. The secondary channels are typically measured by directly measuring the impulse responses (IRs) from the secondary loudspeakers to the error microphone at the desired noise reduction points [121]. To completely omit the unpractical error microphone array from a spatial ANC system, it is worth developing a new secondary channel estimation method which does not require an error microphone array at the desired points. Since spatial ANC systems usually focus on low frequency noise [6], the secondary channel measurements only need to focus on the limited frequencies, resulting in only a limited number of microphones being used in a spatial recording method.

Several previous studies have addressed spatial sound field recording over a region of interest. Abhayapala et al [21] have achieved sound field recording with a

spherical array on the surface of the region of interest. Although microphones are not necessary at all desired recording points, the usage of a spherical microphone array in this work is not practical either. Rafaely et al [141] achieve the spatial sound field recording using a higher order microphone (HOM). The single HOM applied for spatial sound field recording can only achieve the recording within a limited region, while the region of interest in a spatial ANC system is usually large. Samarasinghe et al [142] achieve sound field recording over a large region with distributed HOMs by local-to-global spherical harmonic coefficient translation. Ueno et al [100] achieve sound field recording over a large region with a similar sound field translation theorem but without a global origin. These multiple HOM methods can be beneficial for secondary channel estimations.

Therefore, in this chapter, we carry out a detailed exploration of the potential of applying spatial sound field recording methods to secondary channel estimations. Furthermore, we investigate replacing the multiple fixed-position HOMs with a single moving HOM to further improve the feasibility of a spatial ANC system. The potential relative scattering diffraction between multiple microphones can be also avoided by using a single compact HOM. An Eigenmike (a commercially available 4-th order microphone) [143] is used as an example. We examine the robustness of the sound field recording methods against various measurement errors to approximate the usage of a moving Eigenmike. We propose a solution where the desired secondary channels in a spatial ANC system can be estimated with a single moving HOM. By simulation analysis of the IR measurements performance over a region, we show that the proposed method is able to estimate the secondary channels in a desired frequency range over the entire spatial region.

We organise the main body of this chapter as follows: In Section 5.2 we detail the spatial sound field recording method with a spherical microphone array. In Section 5.3, we introduce the spatial sound field recording methods with arrays of HOMs (Eigenmike) based on the addition theorem and analyze the potential of these methods on secondary channel estimation in spatial ANC systems. The influence of substituting the fixed-position microphone array with a moving microphone is given in Sec. 5.4. Section 5.5 presents simulation results for the proposed method on sound field recording and secondary channel estimation, verifying the effectiveness of the proposed method. We conclude the findings and insights gained

from this work in Section 5.6.

5.2 Problem formulation

Consider a source free region of interest Ω with radius R. The impulse response from a loudspeaker outside the region to an arbitrary point $\boldsymbol{x} = (r, \theta, \phi)$ inside the region is given by $h(\boldsymbol{x}, t)$, where r refers to the distance between the point $\boldsymbol{x}$ and the origin, θ and ϕ denote elevation and azimuth angles, respectively. Using the solution to the wave equation, the frequency response at this point can be express as [19]

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{5.1}$$

where order n and mode m are integers, $\alpha_{nm}(k)$ is a set of spherical harmonic coefficients representing the sound field inside Ω, $k = 2\pi f/c$ is the wave number, f is the frequency of the noise signal, c is the speed of sound, $j_n(\cdot)$ is the spherical Bessel function, and $Y_{nm}(\cdot)$ is the spherical harmonic function [110].

A spatial ANC system usually focuses on limited frequency range, where wave number k has a highest limit $k_{\max}$. In this case, (5.1) can be truncated as

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}^*(\theta, \phi), \tag{5.2}$$

where $N = \lceil k_{max} R \rceil$ [21]. Hence, we can obtain the frequency response at any point $\boldsymbol{x}$ inside the region by obtaining the finite number of $\alpha_{nm}(k)$.

The conventional way of obtaining $\alpha_{nm}(k)$ is by applying a spherical microphone array surrounding the region, where the radius of the array $R_Q \geq R$. The corresponding $\alpha_{nm}(k)$ can be extracted by integrating the recording over the spherical surface while exploiting the orthogonality property of $Y_{nm}(\cdot)$, which gives [21]

$$\begin{aligned} \alpha_{nm}(k) = & \frac{1}{j_n(kR_Q)} \int_0^{2\pi} \int_0^{\pi} z(R_Q, \theta, \phi, k) \\ & \times Y_{nm}^*(\theta, \phi) \sin\theta d\theta d\phi. \end{aligned} \tag{5.3}$$

In practice, this integration is realized using an equivalent discrete summation of

spatial samples over the sphere, where at least $(N+1)^2$ samples are necessary. In other words, we need at least $(N+1)^2$ microphones on a sphere of radius R_Q to obtain the $\alpha_{nm}(k)$ up to desired order N, in order to calculate the frequency response $S(\boldsymbol{x}, k)$ up to a desired frequency. For example, if the radius of the region of interest is $R = 0.5$ m, we need at least 36 and 100 microphones on the sphere to record $S(\boldsymbol{x}, k)$ up to 500 Hz and 1000 Hz, respectively. However, such a spherical microphone array over a large region is usually not realistic. Hence, we need to develop an alternative way to measure $\alpha_{nm}(k)$ of the region of interest.

5.3 Addition theorem based sound field recording over a region using higher order microphones

HOMs are a useful tool to record the spherical harmonic coefficients over large regions. We outline two spatial sound field recording methods [100,142] using HOM arrays in this section. In the following sections, we further investigate the potential of replacing those multiple HOMs with a single moving HOM and evaluate their performance, considering that the usage of multiple HOMs is difficult and expensive to achieve. An Eigenmike is used as an example of a HOM, where 32 channels of recording can be taken on a sphere of radius 0.042 m and up to 4-th order spherical harmonic coefficients can be obtained from the given recording below 3000 Hz.

5.3.1 Spatial sound field recording with a spherical array with HOMs

Conventionally, the spatial sound field recording over a large region with HOM array is developed based on the addition theorem [142].

Consider an Eigenmike located at $\boldsymbol{Y} = (R_E, \theta_E, \phi_E)$ inside the region of interest Ω, where $R_\Omega > R_E$. The spherical harmonic coefficients $\iota_{nm}(k)$ up to 4-th order of a given sound field with respect to the centre of the rigid sphere can be recorded

as

$$\iota_{nm}(k) = \frac{\sum_{q=1}^{32} z(r_q, \theta_q, \phi_q, k) Y_{nm}(\theta_q, \phi_q) \chi(q)}{b_n(r_q, k)}, \tag{5.4}$$

where

$$b_n(r_q, k) = j_n(kr_q) - \frac{j_n'(kr_q)}{h_n'^{(1)}(kr_q)} h_n^{(1)}(kr_q), \tag{5.5}$$

$h_n^{(1)}(\cdot)$ refers to the spherical Hankel function of the first kind, $\chi(q)$ refers to the weight of the q-th microphone such that the summation in (5.5) can sufficiently approximate the integration in (5.3). The frequency limit is given by $k_{\max} = N/R_E$. These spherical harmonic coefficients recorded by the Eigenmike $\iota_{nm}(k)$ can be related to the spherical harmonic coefficients $\alpha_{\nu\mu}(k)$ up to the desired order N_0 with respect to the centre of the region of interest using the addition theorem of spherical Bessel functions as:

$$\iota_{nm}(k) = \sum_{\nu=0}^{N_0} \sum_{\mu=-\nu}^{\nu} \alpha_{\nu\mu}(k) T_{n\nu}^{m\mu}(\boldsymbol{Y}, k), \tag{5.6}$$

where

$$\begin{aligned} T_{n\nu}^{m\mu}(\boldsymbol{Y}, k) =& 4\pi i^{\nu-n} \sum_{u=0}^{\infty} i^u (-1)^{2m-\mu} \\ & \times \sqrt{\frac{(2n+1)(2\nu+1)(2u+1)}{4\pi}} \\ & \times j_u(kR_E) Y_{u(\mu-m)}^*(\theta_E, \phi_E) W_1 W_2, \end{aligned} \tag{5.7}$$

order ν, mode μ, and u are integers, W_1 and W_2 refers to the Wigner $3-j$ symbol.

Q Eigenmikes in the region of interest, where the q-th Eigenmike is at $O_q = (R_q, \theta_q, \phi_q)$, are able to capture local sound field coefficients $\iota_{nm}^{(q)}(k)$ with respect to O_q up to the 4-th order. With these local coefficients, we can write (5.6) in matrix form

$$\boldsymbol{\iota} = \boldsymbol{\mathcal{T}} \boldsymbol{\alpha}, \tag{5.8}$$

where $\boldsymbol{\iota} = [\iota_{00}^{(1)} \cdots \iota_{44}^{(1)} \cdots \iota_{00}^{(Q)} \cdots \iota_{44}^{(Q)}]^T$, $\boldsymbol{\alpha} = [\alpha_{00} \cdots \alpha_{N_0 N_0}]^T$, $\boldsymbol{\mathcal{T}} = [\boldsymbol{T}(O_1), \cdots, \boldsymbol{T}(O_Q)]$ is the translation matrix with the size of $(4+1)^2 Q \times (N_0+1)^2$, $\boldsymbol{T}(O_q)$ is the translation matrix of the q-th Eigenmike, with each element given by $T_{n\nu}^{m\mu}(O_q, k)$.

Hereafter, frequency k is omitted for notational simplicity. Since $\boldsymbol{\mathcal{T}}$ is known with the location of the Eigenmikes, we can obtain the global coefficients $\boldsymbol{\alpha}$ by solving Eq.(5.8) as

$$\boldsymbol{\alpha} = \boldsymbol{\mathcal{T}}^{\dagger}\boldsymbol{\iota}, \tag{5.9}$$

where $(\cdot)^{\dagger}$ is the psuedo-inverse of a matrix.

Exact solutions to the (5.8) exist when $\boldsymbol{\mathcal{T}}$ is square or wide, i.e, $(4+1)^2Q \geq (N_0+1)^2$, which results in an over-determined system. But for the purpose of recording a given sound field, we need at least the same number of spatial samples. The resulting over-determined system can be approximately solved using the method of least squares. Hence, we need $Q > (N_0+1)^2/(4+1)^2$ Eigenmikes to obtain the global coefficients $\alpha_{nm}(k)$ up to order N_0. By applying Q Eigenmikes in the region of interest, we can first record the impulse responses from the loudspeaker to the microphones on the Eigenmikes, calculate $\iota_{nm}^{(q)}(k)$ for the q-th Eigenmike, then obtain the global coefficients $\alpha_{nm}(k)$ of the region of interest. In this way, the desired secondary channels in a spatial ANC system can be obtained by (5.1) up to $k_{\max} = N_0/R_E$.

Regularly sampling with Eigenmikes is necessary in this method to make the matrix inversion in (5.8) to be effective and to avoid spatial alising. Hence, the accuracy of the Eigenmike locations highly influences the performance of the spatial recording. However, we prefer the recording to be achieved by a moving Eigenmike, which introduces more error on the positioning of Eigenmikes. Therefore, the aforementioned method may not achieve a good performance. Hence, in the next sub-section, we introduce a recent advanced method on spatial sound field recording, which provides better robustness with asymmetric/irregular Eigenmike arrays.

5.3.2 Spatial sound field recording with distributed HOMs

We have introduced a method for spatial sound field recording over a region based on the addition theorem in the last section. To achieve that, a series of global spherical harmonic coefficients with respect to the global origin $\alpha_{nm}(\boldsymbol{O})$ is obtained from the recorded local coefficients. In this section, we introduce an alternative method using the addition theorem for spherical harmonic coefficients recording

over a region [100], which enables us to estimate the harmonic coefficients at an arbitrary desired position without setting a global origin, and hence is more flexible with Eigenmike geometry.

It can be shown that the translation matrix $\boldsymbol{T}$ satisfies [113]

$$\begin{aligned} \boldsymbol{T}(-\boldsymbol{Y}) &= \boldsymbol{T}(\boldsymbol{Y})^H, \\ \boldsymbol{T}(\boldsymbol{Y}+\boldsymbol{Y}') &= \boldsymbol{T}(\boldsymbol{Y})\boldsymbol{T}(\boldsymbol{Y}'), \end{aligned} \tag{5.10}$$

where $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ are the location of two different points. Let the directivity pattern of a microphone be represented as $G(\theta,\phi)$, and the spherical harmonic coefficients at this point $\boldsymbol{Y}$ is given by $\boldsymbol{\alpha}(\boldsymbol{Y}) = [\alpha_{00}(\boldsymbol{Y}), \alpha_{1-1}(\boldsymbol{Y}), \cdots, \alpha_{NN}(\boldsymbol{Y})]$, then the recording of the microphone at the point $\boldsymbol{Y}$ can be express by

$$z(\boldsymbol{Y}) = \widetilde{\mathbf{G}}^H \boldsymbol{\alpha}(\boldsymbol{Y}) \tag{5.11}$$

where $\widetilde{\mathbf{G}} = [\widetilde{G}_{00}, \widetilde{G}_{1-1}, \cdots, \widetilde{G}_{NN}]$, with $\widetilde{G}_{nm}$ refering to the spherical harmonic expansion of $G(\theta,\phi)$ at order n and mode m.

Consider Q microphones with arbitrary directivity given by $G_1, \cdots, G_Q$, and with an arbitrary geometry within a certain region at $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_Q$, respectively. The signal observed by the q-th microphone is given by [100]

$$z(\boldsymbol{Y}_q) = \widetilde{\boldsymbol{G}}_q^H \boldsymbol{\alpha}(\boldsymbol{Y}_q) = \widetilde{\boldsymbol{G}}_q^H \boldsymbol{T}(\boldsymbol{Y}_q - \boldsymbol{Y}_0)\boldsymbol{\alpha}(\boldsymbol{Y}_0), \tag{5.12}$$

where $\boldsymbol{Y}_0$ is defined as the global origin.

By defining $\boldsymbol{\Xi}(\boldsymbol{Y}_0) = [\boldsymbol{T}(\boldsymbol{Y}_0 - \boldsymbol{Y}_1)\widetilde{\boldsymbol{G}}_1, \cdots, \boldsymbol{T}(\boldsymbol{Y}_0 - \boldsymbol{Y}_Q)\widetilde{\boldsymbol{G}}_Q]$, Eq. (5.12) can be written in matrix form as

$$\boldsymbol{z} = \boldsymbol{\Xi}(\boldsymbol{Y}_0)^H \boldsymbol{\alpha}(\boldsymbol{Y}_0). \tag{5.13}$$

Using Eq. (5.10), (5.11), and (5.13), the spherical harmonic coefficients $\boldsymbol{\alpha}(\boldsymbol{Y})$ with respect to an arbitrary point $\boldsymbol{Y}$ in the region of interest is given by [100]

$$\boldsymbol{\alpha}(\boldsymbol{Y}) = \boldsymbol{\Xi}(\boldsymbol{Y})(\boldsymbol{\Psi} + \sigma^{-2}\boldsymbol{\Sigma})^{-1}\boldsymbol{z}, \tag{5.14}$$

where $\boldsymbol{\Xi}(\boldsymbol{Y}) = \boldsymbol{T}(\boldsymbol{Y} - \boldsymbol{Y}_0)\boldsymbol{\Xi}(\boldsymbol{Y}_0)$, $\sigma^{-2}\boldsymbol{\Sigma}$ refers to the regularization parameter of

the microphone's noise, σ^{-2} is the average signal power of coefficients and $\mathbf{\Sigma}$ is the covariance of noise, $\sigma^{-2}\mathbf{\Sigma}$ could be a noise variance times identity matrix for eigenmikes, and $\mathbf{\Psi}$ is a matrix with the element at the q-th column and the q'-th row as

$$\Psi_{(q,q')} = \widetilde{\boldsymbol{G}}_q^H \boldsymbol{T}(\boldsymbol{Y}_q - \boldsymbol{Y}_{q'})\widetilde{\boldsymbol{G}}_{q'}, \tag{5.15}$$

which does not rely on the position of the origin $\boldsymbol{Y}_0$. Furthermore, we note that $\widetilde{\boldsymbol{G}}_q$ is usually given by a finite order of non-zero coefficients, where the calculation of matrix $\mathbf{\Psi}$ and hence $\boldsymbol{\alpha}(\boldsymbol{Y})$ are with a finite number of elements.

This method achieves sound field recording over the region by calculating the spherical harmonic coefficients with respect to each desired point $\boldsymbol{Y}$ instead of the global origin. Since only a limited order of coefficients can be recorded by an HOM, there is a limitation of the frequency range and the distance between each microphone to achieve spatial sound field recording. Comparing with the method introduced in Sec. 5.3.1, this method is more flexible on the HOM geometry. As a result, we consider this method to be more suitable to apply for secondary channel estimation in a spatial ANC system with a moving HOM. The comparison of the performance will be given in Sec. 5.5.

5.4 Influence of replacing fix-positioned HOMs with a moving HOM

In the previous sections, we have introduced secondary channel estimation methods with distributed HOM arrays. In the next step, we aim to replace these distributed fix-positioned HOMs by a moving HOM to obtain the necessary IR measurements. By replacing the fixed HOMs with a moving HOM, we face three new problems: (1) The recording at the microphones are not taken at the exactly same points as desired; (2) The Doppler effect introduced by the movement of a HOM; (3) The microphone trajectories may not accurately cover the desired locations, while the tracking of the measurement locations may have error. We analysis the influences of the first problem in this section and demonstrate the influence of the second problem in Sec. 5.5.

Consider a moving microphone with a constant speed of $\boldsymbol{v}$ m/s along a certain

direction. The location of the microphone depends on the time of travel, given by $\boldsymbol{x}(t) = \boldsymbol{x}(0) + \boldsymbol{v}t$ $(t \geq 0)$. At each time interval t, the sound pressure recorded by the microphone is given by the convolution of the source signal denoted as $z(t)$ with the corresponding impulse response $S(\boldsymbol{x}, t)$. We assume the sound source to be at a fixed location. Therefore, the sound pressure recorded by the microphone is given by

$$\begin{aligned} z_m(t) &= z(t) * S(\boldsymbol{x}(t), t) \\ &= \int_{-\infty}^{\infty} z(t_0) S(\boldsymbol{x}(t), t - t_0) dt_0, \end{aligned} \tag{5.16}$$

where $*$ denotes a convolution in the time domain. Comparing to the fixed microphone located at $\boldsymbol{x}(0)$, whose recording is given by

$$\begin{aligned} z_f(t) &= z(t) * S(\boldsymbol{x}(0), t) \\ &= \int_{-\infty}^{\infty} z(t_0) S(\boldsymbol{x}(0), t - t_0) dt_0, \end{aligned} \tag{5.17}$$

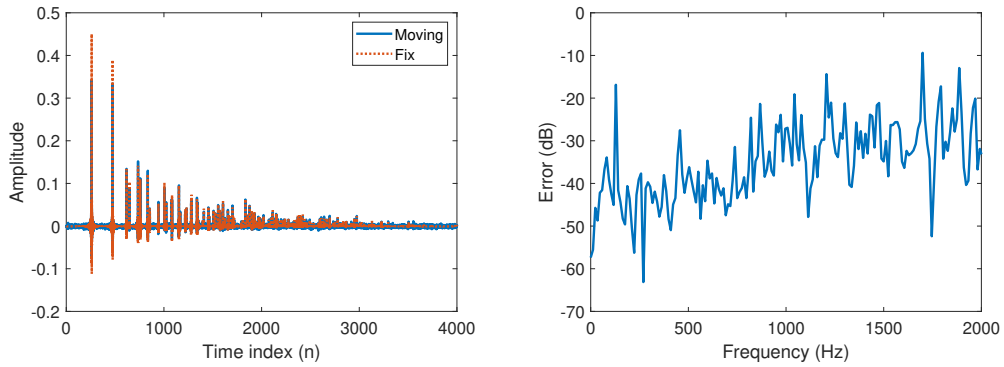
the recording error is given by

$$\begin{aligned} e(t) &= z_m(t) - z_f(t) \\ &= z(t) * [S(\boldsymbol{x}(0) + \boldsymbol{v}t, t) - S(\boldsymbol{x}(0), t)]. \end{aligned} \tag{5.18}$$

We observe from Eq. (5.18) that the error of recording depends on the moving speed of the microphone and the time of travel t. The impulse response error between the moving microphone and the fixed microphone depends on $S(\boldsymbol{x}(0) + \boldsymbol{v}t, t) - S(\boldsymbol{x}(0), t)$. In practice, the impulse response is given by a vector with finite length of discrete samples, and $\boldsymbol{v} \ll c$, where c is the speed of sound. Hence, the error introduced by the moving microphone into the system is limited. Additionally, the Doppler Effect caused by the moving microphone is given by

$$\omega = \omega_0 (\frac{c + \boldsymbol{v}}{c}), \tag{5.19}$$

where the ω_0 is the original frequency, ω is the resulting frequency. The Doppler Effect in our case is very weak and negligible because ω_0 is small and $\boldsymbol{v} \ll c$ [144]. The same result can be extended to the Eigenmike recording, as well as other



(a) Impulse response (b) Frequency response error

Figure 5.1: (a) Impulse response measurements and (b) the frequency response error of $[0, 2000)$ Hz by a moving microphone along a line with the speed of 5 cm/s and a fixed microphone at the centre of the trajectory.

HOMs, as an Eigenmike contains 32 microphones where these microphones can record the impulse responses independently.

We examine the moving error by simulating a single microphone moving along a line while recording. Fig. 5.1(a) compares the impulse response recorded by a moving microphone with a speed of 5 cm/s and a fixed microphone at the centre of the microphone trajectory. The measurement error in the frequency domain up to 2000 Hz of the moving microphone is also shown in Fig. 5.1(b), which is given by:

$$\eta(k) = 10 \log_{10} \frac{E\{|S_m(k) - S_f(k)|^2\}}{E\{|S_f(k)|^2\}}, \tag{5.20}$$

where $S_m(k)$ is the frequency response recorded by the moving microphone at the frequency k, and $S_f(k)$ is the frequency response recorded by the fixed microphone. A maximum length sequence signal of length 4095 samples is used. The signal is repeated 3 times for a single impulse response recording.

It is shown in Fig. 5.1 that the estimation error within the frequency of interest is lower than –20 dB. We consider that this error is tolerable. Furthermore, a microphone recording usually consists of thermal noise, which is measured by the Signal to Noise Ratio (SNR). In Fig. 5.2, the error between the frequency response

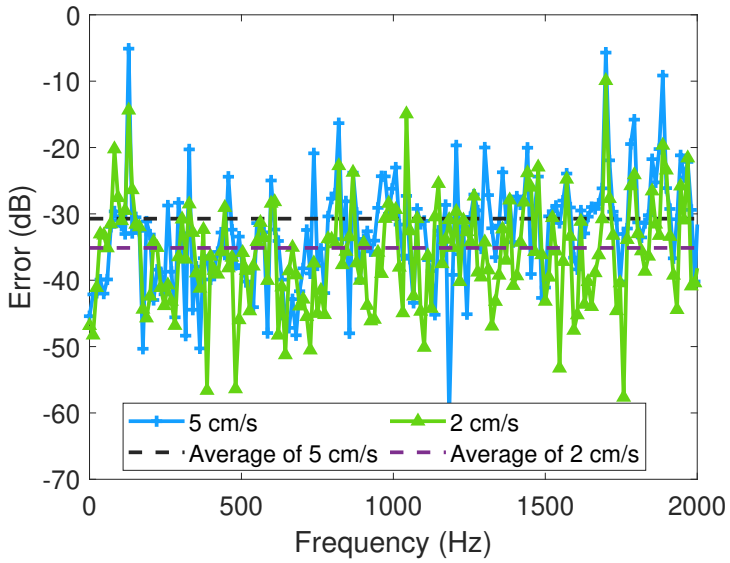
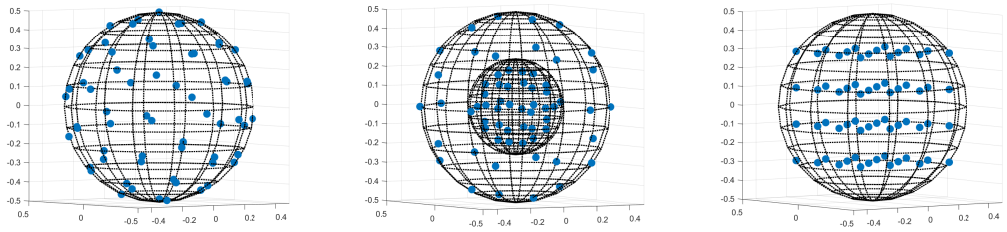


Figure 5.2: Frequency response error by a moving microphone along a line with a speed of 5cm/s and 2cm/s with 40 dB SNR.

by the moving microphone and by the fixed microphone with a SNR of 40 dB is shown. The speed of the microphone is 2cm/s and 5cm/s, respectively. We conclude that in both cases the moving microphone can measure the impulse response under 2000 Hz with a tolerable error. The average error is around –30 dB for the speed of 5 cm/s and –35 dB for the speed of 2 cm/s. Hence, in the following sections, we assume that a moving microphone of speed 5 cm/s with 40 dB SNR is equivalent to a fixed microphone with 30 dB SNR.

5.5 Evaluation examples and analysis

In this section, we evaluate the proposed methods by simulation using Matlab 2019a. We use the Eigenmike's microphone geometry as an example in this section for our model. In Sec. 5.4, we have shown that the fixed Eigenmike can be replaced with a moving Eigenmike through the desired points with slightly higher noise. Hence, we apply fixed Eigenmikes with 30 dB SNR unless we explicitly mention simulating a moving Eigenmike with 40 dB SNR. According to the discussion in Sec.5.3, we consider three setups in a region of interest of the radius of 0.5 m with



(a) Set up I (b) Set up II (c) Set up III

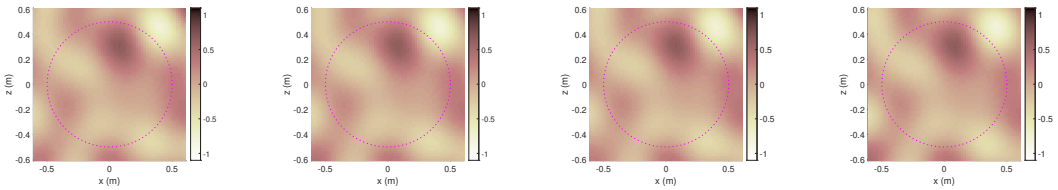
Figure 5.3: The proposed Eigenmike (blue dots) (a) set up (i) 64 Eigenmikes regularly distributed on the spherical surface of the region with $r = 0.5$ m, (b) set up (ii) concentric two spherical arrays with $Q_1 = 32$ at $r = 0.25$ m and $Q_2 = 32$ with $r = 0.5$ m, and (c) set up (iii) 64 Eigenmikes uniformly distributed on a grid as a cubic shape having 4 Eigenmikes on each side.

(i) $Q = 64$ Eigenmikes regularly distributed on the spherical surface of the region; (ii) concentric two spherical arrays with $Q_1 = 32$ at $r = 0.25$ m, and $Q_2 = 32$ with $r = 0.5$ m; (iii) $Q = 64$ Eigenmikes uniformly distributed on a grid as a cubic shape having 4 Eigenmikes on each side. The Eigenmike setups are shown in Fig. 5.3. The centre of the region located at $(2, 3.2, 3)$ m in a room with 4 m height, 6 m length and 5 m width. A single loudspeaker (sound source) is located at $(3, 4.4, 4)$ m in the same room. The reflection coefficients of the 4 walls, roof and ground are set to $[1, 0.7, 0.7, 0.5, 0.2, 1]$. This reflection coefficients are chosen as an example to generate the impulse responses, while in real-world scenarios the impulse responses are given by measurements. The sampling frequency is $F_s = 48000$ Hz and the speed of sound is $c = 343$ m/s.

First, we evaluate the performance of the method introduced in Sec.5.3.2 on sound field recording over the x-z plane with the three set-ups. The driving signal of the loudspeaker is a single tone sine-wave at $f = 750$ Hz. The error of the recording at $\boldsymbol{x}$ is defined by

$$\eta(\boldsymbol{x}) = 10 \log_{10} \frac{E\{|z_r(\boldsymbol{x}) - z(\boldsymbol{x})|^2\}}{E\{|z(\boldsymbol{x})|^2\}}, \tag{5.21}$$

where $z_r(\boldsymbol{x})$ refers to the recorded sound at $\boldsymbol{x}$, and $z(\boldsymbol{x})$ refers to the original sound

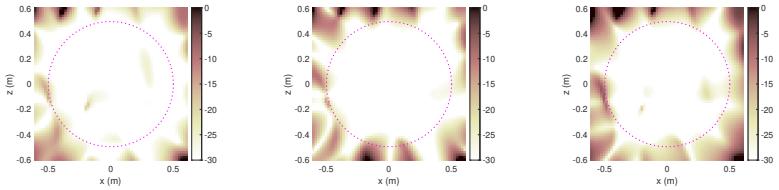


(a) Original sound field

(b) Recorded sound field Setup 1

(c) Recorded sound field Setup 2

(d) Recorded sound field Setup 3



(e) Estimation error Setup 1

(f) Estimation error Setup 2

(g) Estimation error Setup 3

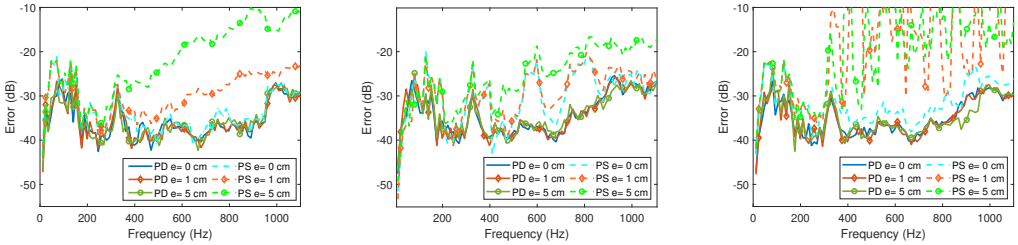
Figure 5.4: (a) The original sound field, (b) (c) (d) the recorded sound field by the proposed three set-ups on the x-z plane and (e) (f) (g) the estimation error with the region of interest circled by pink.

field at $\boldsymbol{x}$.

Figure 5.4 shows the original sound field (5.4(a)), the recorded sound field (5.4(b),5.4(c),5.4(d)) by the method in Sec.5.3.2 and the error (5.4(e),5.4(f),5.4(g)) with the three proposed set-ups. The region of interest is circled by pink lines in the figures. We can obtain that the proposed method can achieve sound field recording over region with little recording error for all three set-ups. Note that the Eigenmike set-up can be flexible depending on the applications and environments.

We further examine the capacity of the proposed methods on secondary channel estimation over the region. We know from Sec.5.4 that the trajectory of a moving microphone may not accurately cover the desired location. There is a potential location error between the true trajectory of the moving Eigenmike and the desired array.

Hence, we estimate the average error of the frequency response recorded by the three Eigenmike set-ups with the proposed method using distributed Eigenmike



(a) Set up I (b) Set up II (c) Set up III

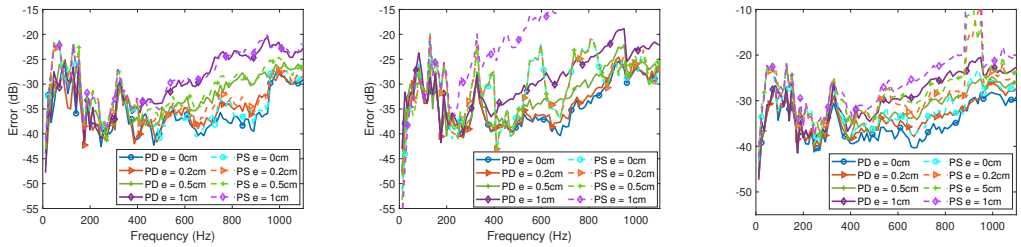
Figure 5.5: The average frequency response error over the region recorded with two proposed methods with location error $e = 0, 1, 5$ cm by the Eigenmike setups (i-iii) for (a-c), respectively.

array (PD) in Sec. 5.3.2 and the proposed method using spherical Eigenmike array (PS) in Sec. 5.3.1 over the region with some location error in Fig. 5.5. Due to physical limitations and obstructions, we sometimes face difficulties in placing error microphones at desired locations. Hence, we place microphones at the nearest available locations with some location error. In the simulation, this location error is added on each Eigenmike with a given distance e towards a random direction. Thus, the true locations of Eigenmikes are e away from the desired locations. The average error of the frequency response measurement over the region of interest at the frequency is given by

$$\eta(k) = 10 \log_{10} \frac{\sum_{\boldsymbol{x} \in \Omega} E\{|S_r(\boldsymbol{x}, k) - S(\boldsymbol{x}, k)|^2\}}{\sum_{\boldsymbol{x} \in \Omega} E\{|S(\boldsymbol{x}, k)|^2\}}, \tag{5.22}$$

We observe from Fig. 5.5 that the PD method can achieve the secondary channel estimation over the region of interest within the target frequency range with all the three set-ups. The estimation error is not significantly influenced by the position error of the Eigenmike. In contrast, the results of the PS method drastically degrade as position error increases, especially at frequencies above 300 Hz. Comparing with the PS method, the PD method achieves a better performance with a higher position error. We also observe that the PD method outperforms the PS method for the other setups.

In spite of location error, the tracking error of a moving microphone is also



(a) Set up I (b) Set up II (c) Set up III

Figure 5.6: The average frequency response error over the region recorded with two proposed methods with Eigenmike tracking error $e = 0, 0.2, 0.5, 1$ cm by the Eigenmike setups (i-iii) for (a-c), respectively.

expected to be significantly larger than a fix-positioned microphone. The location information of a fixed microphone can be accurately pre-measured before the acoustic measurement, while the location of a moving microphone has to be simultaneously observed and updated, which makes it very difficult to be as precise. Being different from the location error, tracking error is the misplacement of microphones that we are not able to notice.

Fig. 5.6 shows the average error of the frequency response recorded by the three Eigenmike set-ups with the PD method and the PS method over the region with Eigenmike tracking error. This tracking error is added on each Eigenmike with a given distance e towards a random direction. Thus, the true locations of Eigenmikes are e away from the measured locations. We obtain from Fig.5.6 that the PD method can achieve the secondary channel estimation over the region of interest within the target frequency range with all three set-ups with Eigenmike tracking error. Comparing with the PS method, the PD method shows the more robustness against the tracking error in all three setups. The performance of the PS method can only remain roughly the same for set up II with $e \leq 0.5$ cm, and degrades in all other scenarios when the tracking error is added, while the performance of the PD method remains roughly the same. Hence, the PD method is more suitable for secondary channel estimation in a spatial ANC system against the tracking error of HOMs.

Overall, we consider the PD method is more robust when the accurate Eigen-

mike location is not available with a high SNR such that it is applicable to estimate the secondary channel with a moving HOM in a spatial ANC system.

5.6 Conclusions

We have shown in this work that a moving higher order microphone can record impulse responses within the target frequency range over the region with a limited movement speed. Hence, the proposed method can be a realistic solution to the secondary channel estimation problem in a spatial ANC system with the absence of an error microphone array. Hence, an error microphone array may not be necessary in a spatial ANC system that applies virtual sensing techniques. As a result, we relax the constraints of building up the unpractical error microphone array in the region of interest, while avoiding these microphones to obstruct users into the quiet zone. We leave the implementation of the proposed method, potentially cooperating with virtual sensing techniques, in a real-world spatial ANC system as future work.

5.7 Related Publications

Much of this chapter's work has been submitted in the following journal proceedings.

- H. Sun, N. Murata, J. Zhang, T.Magariyachi, T. D. Abhayapala, S. Hayashi and P. N. Samarasinghe, T. Itabashi, "Secondary channel estimation in spatial active noise control systems using a single moving high order microphone" *The Journal of the Acoustical Society of America*, vol. 151, no.3, pp. 1922-1931, March, 2022.

Chapter 6

ANC over a spatial region in the time domain

Overview: *The wave domain based spatial ANC is a recently introduced concept that focuses on continuous spatial regions. The design of adaptive filters for spatial ANC is often based on the frequency domain spherical harmonic decomposition method. Given that the performance of an ANC system is highly correlated to the system delay, a major limitation of the frequency-wave domain spatial ANC system is the increased system latency. In this chapter, we develop a time-domain spherical harmonic based signal decomposition method and use it to develop two time-space domain feed-forward adaptive filters for spatial ANC. Through simulations we show that the proposed methods can achieve higher noise reduction performance over the control region with microphones located on the surface of the region compared to the conventional time-domain adaptive filter.*

6.1 Introduction

Conventional multi-channel ANC requires the error microphones uniformly distributed inside the control region [90], which is one of the main drawbacks. Space domain signal processing, using harmonics (cylindrical/spherical) based sound field processing, is recently applied in ANC [73], [145] to increase the performance with microphones on the surface of control region. As space domain solution of wave

equation is mainly developed in the frequency domain [19], most of space domain adaptive filter designs are also done in the frequency-domain [90], [73], [146]. Given ANC systems are very sensitive to time latency [9], significant delays from time-frequency transform thus becomes a problem.

Barkefors and Berthilsson [40] have proposed a time-domain multi-channel spatial ANC system, but it requires the error microphones to be distributed uniformly inside the control region, which limits its usage scenario. Work by Zhang et al [90] achieves noise control over a region with error microphones placed on the boundary of the control region. However, this algorithm is designed for a 2D region in the frequency domain. Chen et al [147] achieves noise control for a 3D region in a car. Their approach is also based on frequency-domain signals, where latency problem introduced by time-frequency transform is still an issue.

In this chapter, we first derive a new spherical harmonic based signal decomposition method in the time-wave domain. Then based on this decomposition method, we propose two new feed-forward adaptive filter designing methods using filtered-x LMS algorithm [9]. These new methods not only have the advantage of the wave domain signal processing, but also avoid the latency caused by time-frequency domain transform. We compare noise reduction performance of time-domain conventional multi-channel method [9] and our proposed two methods by simulation, finding the proposed methods achieve a higher noise reduction over the control region.

6.2 System model

Consider the region of interest Ω as a spherical region of radius r_Ω without any sound sources inside. To measure the residual sound field and to generate a secondary sound field, an array of Q omni-directional microphones and an array of L loudspeakers are uniformly placed on the surface of this control region and on a sphere with radius R_L ($R_L > r_\Omega$), respectively, as shown in Fig. 6.1.

Inside the control region, the residual sound pressure at an arbitrary observation point $\boldsymbol{x} = (r_x, \phi_x, \theta_x)$ is influenced by both the noise sources and the loudspeakers. Let $\mathcal{N}(t)$ be the noise measured by a reference microphone near the noise source, $y_\ell(t)$ be the driving signal of the ℓ^{th} loudspeaker, $g(\boldsymbol{x}|\mathcal{N}, t)$ be the propagation function between the point $\boldsymbol{x}$ and the noise source. Then, the noise sound field

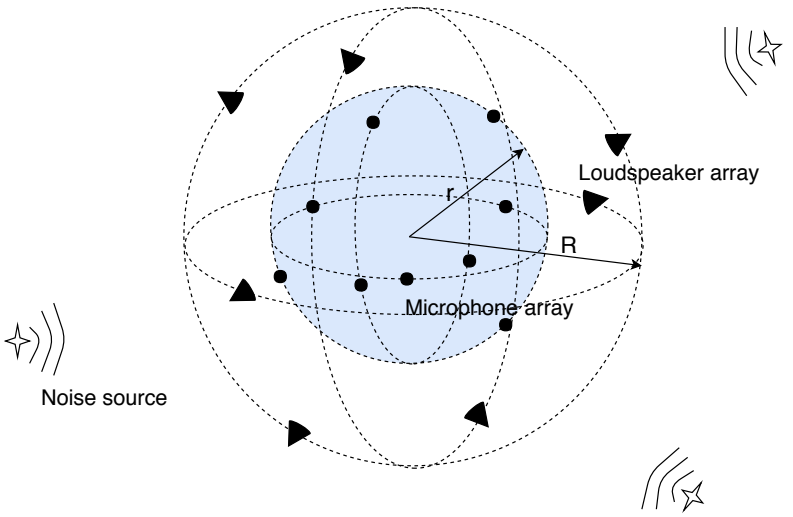


Figure 6.1: ANC system setup with a spatial control region (shaded area) consisting a microphone array of radius r_Ω and a loudspeaker array of radius R_L.

present at $\boldsymbol{x}$ can be given by

$$d(\boldsymbol{x}, t) = \mathcal{N}(t) * g(\boldsymbol{x}|\mathcal{N}, t), \tag{6.1}$$

where $*$ denotes convolution operation. Similarly, the secondary sound field generated by the loudspeakers at point $\boldsymbol{x}$ is

$$y'(\boldsymbol{x}, t) = \sum_{\ell=1}^{L} y_\ell(t) * S(\boldsymbol{x}|\boldsymbol{\ell}, t), \tag{6.2}$$

where $S(\boldsymbol{x}|\boldsymbol{\ell}, t)$ is the propagation function between the point $\boldsymbol{x}$ and the ℓ-th loudspeaker. Therefore, the total or residual sound field $e(\boldsymbol{x}, t)$ as observed at point $\boldsymbol{x}$ is

$$e(\boldsymbol{x}, t) = d(\boldsymbol{x}, t) + y'(\boldsymbol{x}, t). \tag{6.3}$$

In a typical feed-forward adaptive system, the driving signal is obtained by filtering the reference signal $\mathcal{N}(t)$ using a FIR adaptive filter with an impulse response of $w_\ell(t)$, i.e.,

$$y_\ell(t) = \mathcal{N}(t) * w_\ell(t). \tag{6.4}$$

More details on the process of developing this filter is discussed later in Sec.6.4.

The novelty of this method lies in the introduction of a time-wave domain spherical harmonic decomposition of the wave-field, which facilitates the design of the aforementioned filter characteristics $w_\ell(t)$.

6.3 Time-domain analysis of spherical harmonic decomposition

Typically, spherical harmonic based signal decomposition is formulated in the frequency domain. In this section, we develop the corresponding time domain decomposition method. Note that the time domain signal decomposition can be used not only for ANC systems, but also other applications involving spatial sound.

Let $z(\boldsymbol{x}, t)$ be the sound pressure measured at a point $\boldsymbol{x}$ with respect to an origin at time t, and let $Z(\boldsymbol{x}, f)$ be the Fourier transform of $z(\boldsymbol{x}, t)$, where f is the frequency in Hz. Note that $Z(\boldsymbol{x}, f)$ is a solution to the Helmholtz wave equation [19] and can be expressed as

$$Z(\boldsymbol{x}, f) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \alpha_\nu^\mu(f) j_\nu(\frac{2\pi f r_x}{c}) Y_\nu^\mu(\theta_x, \phi_x), \tag{6.5}$$

where $\alpha_\nu^\mu(k)$ is frequency-dependent spherical harmonic coefficients, $j_\nu(\cdot)$ is the n^{th} order spherical Bessel function of the first kind, and $Y_\nu^\mu(\cdot)$ are the real valued spherical harmonic function of the order ν and the degree μ [148]. For any $r_x < r_\Omega$, we can truncate the infinite summation in (6.5) at $V = \lceil k r_x \rceil$ [80].

Due to the Fourier transform relationship between $z(\boldsymbol{x}, t)$ and $Z(\boldsymbol{x}, f)$, we can use (6.5) to write

$$z(\boldsymbol{x}, t) = \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \gamma_\nu^\mu(t) * p_\nu(t) Y_\nu^\mu(\theta_x, \phi_x), \tag{6.6}$$

where $\gamma_\nu^\mu(t)$ is the inverse Fourier transform of $\alpha_\nu^\mu(f)$ and $p_\nu(t)$ is the inverse Fourier transform of $j_\nu(2\pi f r_x/c)$, which is given by [149]

$$p_\nu(t) = \frac{i^\nu c}{2 r_x} P_\nu(\frac{tc}{r_x}), \tag{6.7}$$

where $P_\nu(\cdot)$ is the Legendre function of order ν. A similar truncation to (6.5) of order V can be obtained in (6.6) since $\gamma_\nu^\mu(t) * p_\nu(t)$ and $\alpha_\nu^\mu(k) j_\nu(2\pi f r_x/c)$ are Fourier transform pairs.

By integrating (6.6) over the sphere of radius r_x and using the orthogonal property of $Y_\nu^\mu(\cdot)$, we derive:

$$\gamma_\nu^\mu(t) * p_\nu(t) = \int_0^{2\pi} \int_0^{\pi} z(\boldsymbol{x}, t) Y_\nu^\mu(\theta_x, \phi_x) \sin\theta_x d\theta_x d\phi_x. \tag{6.8}$$

Then, by convolving (6.8) with $\rho_\nu(t)$[1], where $\rho_\nu(t) * p_\nu(t) = \delta(t)$, we have

$$\gamma_\nu^\mu(t) = \rho_\nu(t) * \int_0^{2\pi} \int_0^{\pi} z(\boldsymbol{x}, t) Y_\nu^\mu(\theta_x, \phi_x) \sin\theta_x d\theta_x d\phi_x. \tag{6.9}$$

We can approximate the integration in (6.9) with a finite summation to estimate $\gamma_\nu^\mu(t)$ [21]. When sound field $z(\boldsymbol{x_q}, t)$ is measured for $q = 1, \cdots, Q$ with the error microphones, we can calculate $\gamma_\nu^\mu(t)$ by

$$\gamma_\nu^\mu(t) \approx \rho_\nu(t) * \sum_{q=1}^{Q} z(\boldsymbol{x_q}, t) Y_\nu^\mu(\theta_q, \phi_q) \chi_q, \tag{6.10}$$

where χ_q is a correction factor depending on the geometry of the microphones.

6.4 Time-space domain adaptive algorithms

In this section, we develop two time-space domain feed-forward adaptive methods based on the time-domain Fx-LMS algorithm [9]. We use discrete-time signals and discrete time-space domain coefficients in the following sections, thus the time variable t representing the discrete time samples hereafter.

[1]Note that $\rho_\nu(t)$ can be constructed by taking the inverse Fourier transform of $1/j_\nu(2\pi f r/c)$ and using a suitable band stop filter to avoid Bessel zeros.

6.4.1 Formulation of time-space domain signal coefficients

In Sec.6.3, we introduced the spherical harmonic coefficients of the time-space signal $z(\boldsymbol{x}, t)$ as $\gamma_\nu^\mu(t)$. By applying this relationship to (6.3), we have

$$\gamma_\nu^{\mu(e)}(t) = \gamma_\nu^{\mu(d)}(t) + \gamma_\nu^{\mu(s)}(t), \tag{6.11}$$

where $\gamma_\nu^{\mu(e)}(t), \gamma_\nu^{\mu(d)}(t)$ and $\gamma_\nu^{\mu(s)}(t)$ are the spherical coefficients of $e(\boldsymbol{x}, t)$, $d(\boldsymbol{x}, t)$ and $y'(\boldsymbol{x}, t)$, respectively. By substituting for (6.2) from (6.4), we have

$$\gamma_\nu^{\mu(s)}(t) = \sum_{\ell=1}^{L} \mathcal{N}(t) * \widetilde{S}_{\nu,\ell}^{\mu}(t) * w_\ell(t), \tag{6.12}$$

where $\widetilde{S}_{\nu,\ell}^{\mu}(t)$ is the spherical harmonic coefficient of $S(\boldsymbol{x}|\ell, t)$.

Based on (6.11) and (6.12), we design two time-space domain feed-forward adaptive filters in the following sections with the block diagram shown in Fig. 6.2.

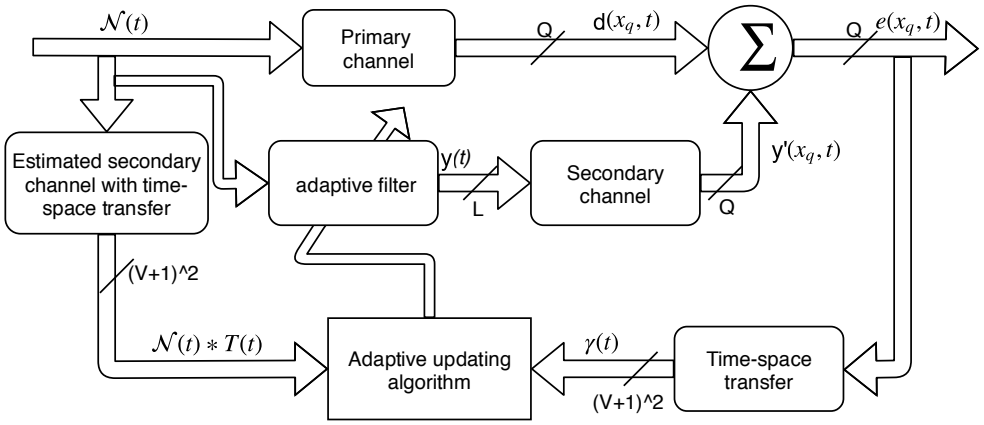


Figure 6.2: Block diagram of time-space domain adaptive algorithms.

By deriving the adaptive update algorithm of $w_\ell(t)$, we aim to minimize $\gamma_\nu^{\mu(e)}(t)$, which represents the residual sound field $e(\boldsymbol{x}, t)$ over the control region.

6.4.2 Minimizing Squared Residual Sound Field Coefficient Error over Region (MSE-R)

Let $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \cdots, \boldsymbol{w}_L]$, where $\boldsymbol{w}_\ell$ is the vector of filter taps for the ℓ^{th} loudspeaker, with order of L_w. To minimize the residual sound field, we define the adaptive algorithm cost function as

$$\xi(t) = \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \|\gamma_\nu^{\mu(e)}(t)\|^2. \tag{6.13}$$

Taking the derivative of ξ with respect to $\boldsymbol{W}(t)$, and by using (6.11) and (6.12), we derive

$$\begin{aligned} \nabla\xi(t) = \frac{\partial \xi(t)}{\partial \boldsymbol{W}(t)} &= \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} 2\gamma_\nu^{\mu(e)}(t)[\frac{\partial \gamma_\nu^{\mu(e)}(t)}{\partial \boldsymbol{W}(t)}] \\ &= 2\sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \gamma_\nu^{\mu(e)}(t)[\mathcal{N}(n) * \widetilde{S}_{\nu,\ell}^{\mu}(t)]. \end{aligned} \tag{6.14}$$

For conventional multi-channel adaptive filters, the update equation is typically given by [9]

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(t) - \frac{\lambda}{2}\nabla\xi(t), \tag{6.15}$$

where λ is the step size. Hence, for each secondary loudspeaker, the τ^{th} element of time-space domain update equation is

$$\begin{aligned} \boldsymbol{w}_{\ell,\tau}(n+1) =& \boldsymbol{w}_{\ell,\tau}(t) - \lambda \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \gamma_\nu^{\mu(e)}(t) \\ & \times [\mathcal{N}(n) * \widetilde{S}_{\nu,\ell}^{\mu}(t)]. \end{aligned} \tag{6.16}$$

We implement our adaptive algorithm (6.16) by the following steps: a) pre-estimate the impulse response of secondary channel $S(\boldsymbol{x}_q|\ell, t)$ from each loudspeaker to each microphone, b) measure the error signal $e(\boldsymbol{x}_q, t)$ by error microphones, and reference signal $\mathcal{N}(t)$ by the reference microphone, c) estimate $\gamma_\nu^{\mu(e)}(t)$ and $\widetilde{S}_{\nu,\ell}^{\mu}(t)$.

To obtain $\gamma_\nu^\mu(t)^{(e)}$ and $\widetilde{S}_{\nu,\ell}^\mu(t)$ using (6.10), an inverse Fourier transform based function $\rho_\nu(t)$ is involved, where its group delay can lower performance and slow down convergence and its Bessel zeros can influence stability, which are not desirable in the system. Therefore, in the next section, we study an alternate cost function to avoid these drawbacks.

6.4.3 Minimizing Squared Residual Sound Field Error on the Region Boundary (MSE-B)

Instead of minimizing the residual sound field over the whole control region as in Sec. 6.4.2, here we only minimize the residual sound field on the boundary. As there are no noise sources inside the region, this method should still achieve an acceptable noise reduction within the region.

We define the adaptive algorithm cost function as

$$\xi = \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \|\gamma_\nu^{\mu(b)}(t)\|^2, \tag{6.17}$$

where

$$\begin{aligned} \gamma_\nu^{\mu(b)}(t) &= \gamma_\nu^{\mu(e)}(t) * p_\nu(t) \\ &= \int_0^{2\pi} \int_0^{\pi} e(\boldsymbol{x}, t) Y_\nu^\mu(\theta_x, \phi_x) \sin\theta_x d\theta_x d\phi_x \\ &\approx \sum_{q=1}^{Q} e(\boldsymbol{x_q}, t) Y_\nu^\mu(\theta_q, \phi_q) \chi_q. \end{aligned} \tag{6.18}$$

By taking the derivative of ξ in (6.17) with respect to $\boldsymbol{W}(t)$ and using (6.11) and (6.12), the gradient of this cost function can be derived as

$$\begin{aligned} \nabla\xi = \frac{\partial \xi}{\partial \boldsymbol{W}(t)} &= \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} 2\gamma_\nu^{\mu(b)}(t) [\frac{\partial \gamma_\nu^{\mu(b)}(t)}{\partial \boldsymbol{W}(t)}] \\ &= 2\gamma_\nu^{\mu(b)}(t) [\mathcal{N}(n) * \widetilde{S}_{\nu,\ell}^\mu(t) * p_\nu(t)]. \end{aligned} \tag{6.19}$$

With (6.15) and (6.19), for each loudspeaker, the τ^{th} element of the update equation

is

$$\begin{aligned}\boldsymbol{w}_{\ell,\tau}(n+1) =& \boldsymbol{w}_{\ell,\tau}(t) - \lambda \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} [\gamma_{\nu}^{\mu(e)}(t) * p_{\nu}(t)] \\ & \times \{[\mathcal{N}(n) * \widetilde{S}_{\nu,\ell}^{\mu}(t)] * p_{\nu}(t)\} \\ =& \boldsymbol{w}_{\ell,\tau}(t) - \lambda \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \sum_{q=1}^{Q} e(\boldsymbol{x_q}, t) Y_{\nu}^{\mu}(\theta_q, \phi_q) \\ & \times [\mathcal{N}(n) * \sum_{q=1}^{Q} S(\boldsymbol{x}_q|\ell, t) Y_{\nu}^{\mu}(\theta_q, \phi_q)].\end{aligned} \tag{6.20}$$

In this case, we avoid calculating $\rho_{\nu}(t)$ to obtain the update equation, hence avoid latency of inverse Fourier transform and Bessel zeros problem.

6.5 Simulation Results and analysis

In this section, we compare the performances of the proposed methods (MSE-R, MSE-B) against the conventional multi-channel adaptive filtering method (MP) [2] in both the free space and the room environment.

We simulate a feed-forward ANC system consisting of 9 error microphones and 9 loudspeakers uniformly spaced on two concentric spheres of radius 0.16 m and 0.48 m [150], respectively. A single noise source is located at $(2, 90^{\circ}, 90^{\circ})$, where a reference microphone is placed nearby to obtain reference signals. We consider four different noise signals in this simulation, each lasting 1 s.

Scenario 1: Multiple superimposing sine waves of frequency 100 Hz, 170 Hz, and 250 Hz;

Scenario 2: Single sine wave of frequency 210 Hz;

Scenario 3: filtered Gaussian distributed random signal with a 600 Hz cut-off low-pass filter;

Scenario 4: filtered real noise recorded in a computer room with a 600 Hz cut-off low-pass filter.

A signal-to-noise ratio (SNR) of 60 dB is added to the microphone recordings for the first three scenarios. Sampling rate is 48 kHz, yet we down sampled at a rate of 10 to reduce computational cost. In order to simulate the reverberant room

environment, the image-source method [151] is employed, where the room size is set to be 4 m × 5 m × 3 m with reflection coefficients of 0.9, 0.7, 0.8, 0.6, 0.5 and 0.8 of 4 walls and floor and roof respectively. The noise of scenario 4 is recorded in a room full of computers. The majority of the noise is contributed by the computer fan noise, and the air-conditioner noise exists, too. There are no people walking or talking.

We define a metric for noise reduction at point $\boldsymbol{x}$ inside the control region as

$$\eta(\boldsymbol{x}) = 10\log_{10}\frac{\sum_n e(\boldsymbol{x},n)^2}{\sum_n d(\boldsymbol{x},n)^2}, \tag{6.21}$$

where the summation is over the last 480 samples of the signals, to ensure that the adaptive filter has reached convergence.

We first plot the performance of scenario 1 on the x-y plane with red and blue circles indicating the sphere where the error microphones and the loudspeakers are located, respectively.

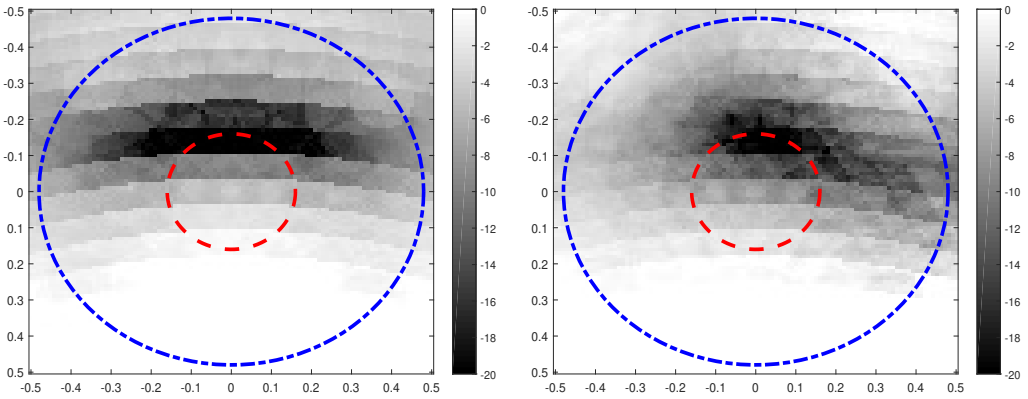
As shown in Fig. 6.3, all methods can achieve some level of noise reduction in both free-space and room environment. However, its clearly observed that the noise reduction within the entire control region is better with the proposed methods while MSE-B method achieves the highest performance. With MSE-B method, We can see that almost every test point inside the control region is dark, which refers to around 20 dB noise reduction.

Performance of these methods over the whole region are theoretically evaluated by averaging $\eta(\boldsymbol{x})$ over the whole control region with 2103 uniformly placed points.

Table 6.1: Average performance (dB) over the whole region in free-space.

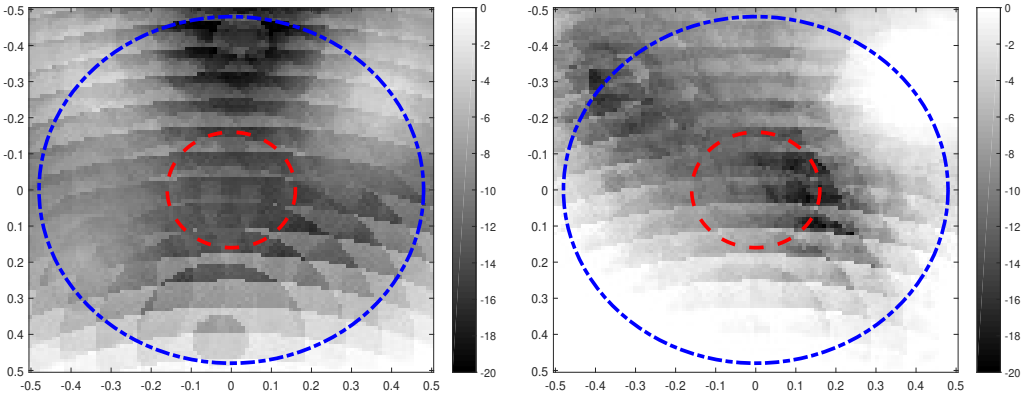
	MP	MSE-R	MSE-B
Single sine wave	5.55	18.76	21.27
Multi sine wave	6.52	14.06	20.92
Random noise	16.11	16.85	21.85
Real noise	9.00	10.03	13.50

Table 6.1 shows the results of average performance over the whole region for four scenarios and three different methods in free space. From Table 6.1, we note that



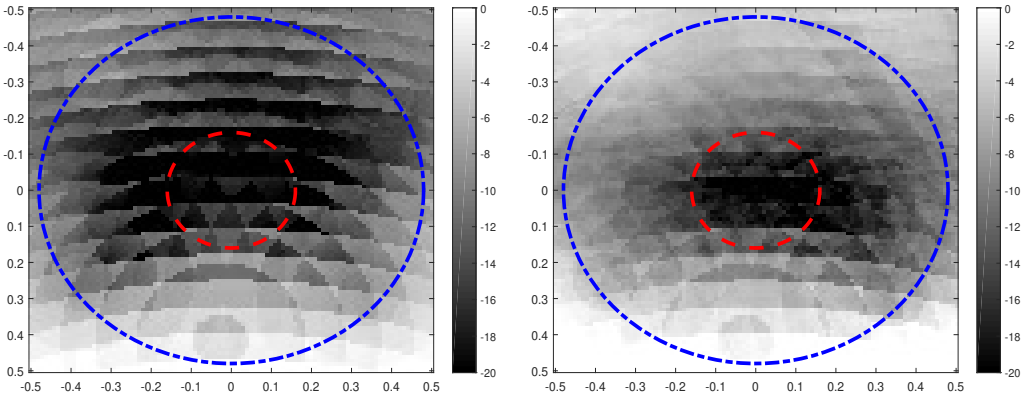
(a) MP-freespace

(b) MP-room



(c) MSE-R-freespace

(d) MSE-R-room



(e) MSE-B-freespace

(f) MSE-B-room

Figure 6.3: Noise reduction performance in free field with a) MP, c) MSE-R, e) MSE-B, and in a reverberant room with b) MP, d) MSE-R, f) MSE-B.

the proposed two methods achieve higher performance than MP method. MSE-B method achieves the highest noise reduction with all scenarios. For stationary signals, MSE-R method performs significantly better than MP method.

Table 6.2 shows the results of average performance over the whole region for the four scenarios mentioned earlier and three different methods in a room environment. We find the same trends as in free-space with sinusoidal noises, that MSE-B method

Table 6.2: Average performance (dB) over the whole region in room.

	MP	MSE-R	MSE-B
Single sine wave	8.60	20.09	23.67
Multi sine wave	8.85	13.43	17.51
Random noise	5.56	10.83	9.56
Real noise	4.40	5.72	4.41

achieves the highest performance. When the noise signals are non-stationary, MSE-R method achieves the highest performance, which is different from what we found in free space.

Theoretically, MSE-R method should achieve the highest performance since it minimizes the coefficients over the whole control region. However, we note the different results in free space and for stationary signals in room from the simulation. This phenomenon is due to the group delay as described in Sec.6.4.2. In free space, the system delay is mainly caused by filtering methods, hence we obtain an obvious difference on performance between MSE-R method and MSE-B method. Using an advanced windowing method can help to decrease the delay [152].

However in the room environment, as the reverberation from walls makes the impulse responses between loudspeakers and microphones longer, the latency of the whole system becomes much more longer than in free-space. Non-stationary noises have time-varying frequency responses, hence are more sensitive with the delays. We can observe in the room environment, performance of non-stationary signals with all three methods are smaller than stationary signals. With those non-stationary noises, as the delays caused by filtering methods are much more shorter than channel delays in the room, MSE-R achieves the best performance with its strong control over the whole control region.

6.6 Conclusions

In this chapter, we derive a spherical harmonic based time-wave domain signal decomposition method, providing a tool for spatial sound field analysis without transforming signals to the frequency domain. Based on that, we proposed two time-wave domain methods for feed-forward adaptive filtering to achieve noise reduction over a spherical region. We compared the noise reduction performance of these proposed methods against the conventional time-domain multi-channel ANC system in both the free-space and the reverberant room environments, finding that the proposed two methods are effective with both narrow-band and wide-band noise signals. In simple environments, such as the free-space, MSE-B method achieves the best performance because of its short filtering latency. In a reverberant room with long and complex channels between microphones and loudspeakers, MSE-R method achieves a better performance with non-stationary noises since it minimizes the coefficients of residual signals over the whole control region.

6.7 Related Publications

Much of this chapter's work has been published in the following conference proceedings.

- H. Sun, T. D. Abhayapala and P. N. Samarasinghe, "Time Domain Spherical Harmonic Analysis for Adaptive Noise Cancellation over a Spatial Region," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 516-520, Brighton, UK, May, 2019.

Chapter 7

Time domain spherical harmonic processing with open spherical microphones recording

Overview: *Spherical harmonic analysis has been a widely used approach for spatial audio processing in recent years. Among all applications that benefit from spatial processing, spatial ANC remains unique with its requirement for open spherical microphone arrays to record the residual sound field throughout the continuous region. Ideally, a low delay spherical harmonic recording algorithm for open spherical microphone arrays is desired for real-time spatial ANC systems. Currently, the frequency domain algorithms for spherical harmonic decomposition of microphone array recordings are applied in a spatial ANC system, where a Short Time Fourier Transform is required, which introduces undesirable system delay for ANC systems. We have developed the time-wave domain ANC system in Chapter 6 with Inverse Fourier Transform based filters. In this chapter, we develop a time domain spherical harmonic decomposition algorithm for the application of spatial audio recording mainly with benefit to ANC with an open spherical microphone array. Microphone signals are processed by a series of pre-designed finite impulse response (FIR) filters to obtain a set of time domain spherical harmonic coefficients. The time domain coefficients contain the continuous spatial information of the residual sound field. We corroborate the time domain algorithm with a numerical simulation of a fourth*

order system, and show the proposed method to have lower delay than existing approaches.

7.1 Introduction

Spherical harmonic analysis has been widely used for spatial acoustic signal processing for years [80]. Sound field recordings can be decomposed into a set of orthogonal spatial basis functions and respective coefficients when an appropriately designed spherical microphone array is used [21, 78]. The spherical harmonic decomposition has the advantage that a given sound field can be analyzed over a continuous spatial region rather than a set of distributed points [153]. This has embraced a wide range of algorithms in three-dimensional (3D) audio signal processing such as: sound intensity analysis [154], sound field diffusive analysis [155], beamforming [156, 157], source localization [158, 159], and spatial ANC [90, 146].

A spatial ANC system aims to reduce the unwanted acoustic noise [160] over a space in order to create a silent zone for people. Multiple microphones are used to record the residual noise, and multiple loudspeakers are used to generate the anti-noise field. The recording's accuracy of the residual sound field can highly influence the performance of an ANC system. Furthermore, recording efficiency is also important, as ANC usually focuses on low frequency and time-variant noise. As a result, an accurate and low latency algorithm for residual sound field recording is desired [9].

The sound field recording step in a spatial ANC system focuses on obtaining the location independent spherical harmonic coefficients that represent the residual sound field inside a region of interest. This is different to real time spherical harmonic beamforming or directivity analysis which focuses on extracting source location information from the spatial recording. Moreover, spatial ANC mainly focuses on reducing the sound field inside the spherical microphone array (the region of interest). While other spatial recording applications may focus on analysing the sound field exterior to the array. Additionally, although most of the spatial audio applications utilize a rigid spherical array [161–163] for its convenience to build and use, an open spherical array is considered to be more suitable for a spatial ANC system. This is because users should be able to enter and move

within the ANC region of interest that is surrounded by the spherical microphone array [90, 164]. Furthermore, there exists previous work focusing on optimising the open array for spherical harmonic recording [102, 127], and for spatial ANC systems [165]. However, we consider the optimisation of the open microphone array design to be outside of the scope of this work, and instead focus on a time-domain recording algorithm.

Real-time spatial beamforming systems illustrate that applications with strict delay requirements can highly benefit from the small latency and efficient computation of time domain processing [166,167]. By posing the signal processing algorithm in the time domain, system performance can be optimized with real-valued lower order filters [168], and lower modeling delays [169]. Specifically, for a spatial ANC system, the system delay which includes the filter group delay (signal processing algorithm), the A/D and D/A converter, and the data processing delay, should be less than the acoustic delay from the reference microphones to the secondary loudspeakers in order to satisfy causality [27]. Furthermore, a longer signal processing delay slows down the convergence speed of the adaptive filtering and may lead to an unstable system [170,171]. Therefore, it is worthwhile to consider a time domain spherical harmonic decomposition method to achieve sound field recording with an open spherical array for the application of spatial ANC.

Frequency domain spherical harmonic recording has been well developed with various optimised filters [172–174]. One benefit of developing the method in the frequency domain is that the influence of the spherical Bessel zeros can be easily removed by avoiding the estimation of the coefficients at these erroneous frequency bins [100,102,175,176]. However, when we consider a time domain method, we can not simply avoid the Bessel zeros because we do not apply a Fourier Transformation to separate the Bessel zero frequency components from the others.

Meanwhile, there are also several work relate to time domain spatial audio signal processing. In [105], Poletti and Abhayapala give a time domain description of the free-space Green's function in the spherical harmonic domain. This provides a solution to decompose the free-space channel between a loudspeaker and microphone into the time-space domain. This work only targets the free-space Green's function, and as a result, the method is highly limited to the application of free space sound field reproduction. In [104], a time domain wave field synthesis method

is presented. Although an IFFT is applied to derive the time domain solution, the work still demonstrates that time-domain wave field synthesis can be beneficial to time-varying spatial acoustic applications. In [177], Hahn and Spors offer a time domain representation of the spherical harmonic equation. They relate the time domain spherical harmonic coefficients to the sound pressure, but do not include the method of obtaining the time domain coefficients from a given recording. Time domain beamformers are designed in [107, 178] with the IFFT of spherical harmonics. These papers show certain advantages for finite impulse response (FIR) filtering based signal processing systems. Overall, these time domain approaches illustrate the advantages of time domain signal processing, however, they remain unable to obtain location-independent spherical harmonic coefficients. This makes them ill-suited for spatial ANC systems, as these location-independent coefficients provide necessary information about the continuous residual sound field inside the region of interest.

In this chapter, we propose a FIR filter based time domain spherical harmonic analysis method to accurately record spatial sound fields with an open spherical microphone array for the purpose of spatial ANC. We note that this work focuses solely on the problem of sound field recording, and that the spatial ANC application acts purely as motivation to our problem. Therefore, with spatial ANC in mind, the recording method prioritizes a minimum processing delay, a bandwidth of interest (low frequencies for typical noise scenarios), and a practical array geometry (open sphere surrounding a quiet zone). Employing the recording method in an actual ANC system, and its evaluation, is out of the scope of this work. The novelty of the presented work is the investigation of time domain spherical harmonic coefficients. These time domain coefficients match the properties of conventional frequency domain spherical harmonic coefficients. That is, the coefficients are location independent within the region of interest, and they represent the continuous sound field over the space. Additionally, these coefficients are obtained in the time domain, which relieves the block processing constraint (and can do sample-by-sample processing) and results in lower system delay. Hence, the proposed method is considered to be highly beneficial to spatial ANC systems.

We organise the main body of this chapter as follows: In Section 7.2 we detail the background of the frequency domain spherical harmonic algorithm for spatial sound

field recording. Additionally, we introduce the time domain equation of spherical harmonic decomposition, while addressing the challenges of recording time domain spherical harmonic coefficients. The filter's design and implementation to obtain the time domain spherical harmonic coefficients is presented in Section 7.3, along with error analysis. Effects of truncation and filter length are shown in Section 7.4 via initial simulations of filter performance. Section 7.5 presents simulation results for the proposed method's estimation of spherical harmonic coefficients, as well as sound field reconstruction performance at a point and over space, verifying the effectiveness of the proposed theory and design. We conclude the findings and insights gained from this work in Section 7.6.

7.2 Problem Formulation

We begin this section by reviewing the well-known frequency domain spherical harmonic decomposition method. We then introduce the corresponding time domain formulation, and detail the Fourier Transform relationship between the components in the frequency domain equation and the time domain equivalent. Finally, we show the difficulties in obtaining spherical harmonic coefficients in the time domain.

7.2.1 Spherical Harmonic Decomposition of Sound Field in Frequency-Space Domain

An incident sound field at any arbitrary point $\boldsymbol{x} = (r, \theta, \phi)$ inside a source free 3D spherical region Ω, where r refers to the distance between the point $\boldsymbol{x}$ and the origin, θ and ϕ denote elevation and azimuth angles, respectively [**?**], can be expressed in the frequency domain as [19, 80]

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{7.1}$$

where order n ($n \geq 0$) and mode m are integers, $N = \lceil kR \rceil$ [80], $k = 2\pi f/c$ is the wave number, f is frequency, c is the speed of sound, R is the radius of Ω, $\alpha_{nm}(k)$ is a set of spherical harmonic coefficients representing the sound field inside Ω,

$j_n(kr)$ is the n^{th} order spherical Bessel function of the first kind, $Y_{nm}(\theta, \phi)$ are the spherical harmonic functions. For convenience, we use real spherical harmonics in this thesis, given by [148]

$$\begin{aligned} Y_{nm}(\theta, \phi) =& (-1)^{|m|} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} \\ & \times \begin{cases} P_{nm}(\cos\theta)\cos(m\phi) & m \geq 0 \\ P_{nm}(\cos\theta)\sin(m\phi) & m < 0 \end{cases}, \end{aligned} \tag{7.2}$$

where $P_{nm}(\cdot)$ is the associated Legendre function. Real spherical harmonics have the orthogonality property of

$$\int_0^{2\pi} \int_0^{\pi} Y_{nm}(\theta, \phi) Y_{n'm'}(\theta, \phi) \sin\theta d\theta d\phi = \delta_{nn'}\delta_{mm'}. \tag{7.3}$$

If the spherical harmonic coefficients $\alpha_{nm}(k)$ are available for a sound field, then these coefficients can fully describe the sound field over the continuous spatial region of interest. Traditionally, when spatial harmonic processing is used to record a spatial sound field $z(\boldsymbol{x}, k)$, it is recorded over a spherical surface of radius R_Q ($R_Q \geq r$). The corresponding $\alpha_{nm}(k)$ are extracted by integrating (7.1) over the spherical surface while exploiting the orthogonality property of $Y_{nm}(\cdot)$ in (7.3), which gives [21]

$$\alpha_{nm}(k) = \frac{1}{j_n(kr)} \int_0^{2\pi} \int_0^{\pi} z(r, \theta, \phi, k) Y_{nm}^*(\theta, \phi) \sin\theta d\theta d\phi. \tag{7.4}$$

In practice, this integration is realized using an equivalent discrete summation of spatial samples over the sphere.

7.2.2 Equivalent Spherical Harmonic Decomposition of a Sound Field in Time-Space Domain

While the frequency domain spatial sound field capture is well established as explained in Section 7.2.1, in this work, our objective is to investigate the possibility of an analogous spherical harmonic analysis in the time domain. In a similar fash-

ion to (7.1) and (7.4), we now consider the relationship between sound pressure $z(\boldsymbol{x}, t)$ recorded by a spherical microphone array and the time domain spherical harmonic coefficients, denoted as $\gamma_{nm}(t)$. It is desirable to have these time domain coefficients $\gamma_{nm}(t)$ independent of the measurement radius. Thus, we only need to record $\gamma_{nm}(t)$ to obtain the sound field over the entire region of interest Ω. A time domain method can directly extract $\gamma_{nm}(t)$, thus avoiding the Fourier transformation of signals.

As a time domain analysis is usually with real-valued components, we rewrite (7.1) in the form of

$$z(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} i^n \alpha_{nm}(k) \frac{j_n(kr)}{i^n} Y_{nm}(\theta, \phi), \tag{7.5}$$

where $i = \sqrt{-1}$, in order to make the inverse Fourier transform of all terms to be real. Taking the inverse Fourier transformation of (7.5), we obtain

$$z(\boldsymbol{x}, t) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \gamma_{nm}(t) * p_n(t, r) Y_{nm}(\theta, \phi), \tag{7.6}$$

where $*$ denotes the convolution operation,

$$\gamma_{nm}(t) \xrightarrow{\mathcal{F}} i^n \alpha_{nm}(k), \tag{7.7}$$

where $\xrightarrow{\mathcal{F}}$ denotes the Fourier transform operator,

$$p_n(t, r) \xrightarrow{\mathcal{F}} \frac{j_n(kr)}{i^n}, \tag{7.8}$$

which is given by

$$p_n(t, r) = \begin{cases} \frac{c}{2r} P_n(\frac{tc}{r}) & -\frac{r}{c} \le t \le \frac{r}{c} \\ 0 & |t| > \frac{r}{c} \end{cases}, \tag{7.9}$$

where $P_n(\cdot)$ is the Legendre function. We note that every component in (7.6) is real valued. The proof of (7.9) is given below:

We have the Fourier relationship between the spherical Bessel function $j_n(kr)$

and the Legendre function $P_n(t)$ given by (Eq.4) in [112] that

$$\int_{-\infty}^{\infty} e^{ikrt} j_n(kr) dkr = \pi i^n P_n(t). \tag{7.10}$$

With (7.10), $p_n(t,r)$ in (7.8) can be express as

$$\begin{aligned} p_n(t,r) &= \frac{c}{i^n 2\pi r} \int_{-\infty}^{\infty} j_n(kr) e^{\frac{itckr}{r}} dkr \\ &= \begin{cases} \frac{c}{2r} P_n(\frac{tc}{r}) & -\frac{r}{c} \leq t \leq \frac{r}{c} \\ 0 & \pm t > \frac{r}{c} \end{cases} . \end{aligned} \tag{7.11}$$

This completes the proof of (7.9).

Equation (7.6) shows how to reconstruct the sound pressure at $\boldsymbol{x} = (r, \theta, \phi)$ with the recorded time domain spherical harmonic coefficients $\gamma_{nm}(t)$. We consider an alternative time domain filter to obtain $\gamma_{nm}(t)$ from the recorded signals rather than taking the inverse Fourier transform of (7.4) since $1/j_n(kr)$ is unbounded when $j_n(kr) = 0$. Note that $j_n(kr)$ as a filter has order dependent zeros when $j_n(kr) = 0$. As a result, $1/j_n(kr)$ approaches infinity at these frequencies, making it unstable to have an inverse Fourier transform. In other words, the z-transform of $p_n(t,r)$ given in (7.9), has zeros on the unit circle because of Bessel zeros, refers to a non-minimum phase system. In this case, the inverse system of $p_n(t,r)$, with the frequency response of $1/j_n(kr)$ is not stable. As a result, we first define

$$\Gamma_{nm}(t,r) \triangleq \gamma_{nm}(t) * p_n(t,r), \tag{7.12}$$

which has a frequency response of

$$\Gamma_{nm}(t,r) \xrightarrow{\mathcal{F}} i^n \alpha_{nm}(k) \frac{j_n(kr)}{i^n} = \alpha_{nm}(k) j_n(kr). \tag{7.13}$$

Since $Y_{nm}(\theta, \phi)$ is independent to both frequency and time, $\Gamma_{nm}(t,r)$ can be obtained by integrating (7.6) over a sphere of radius r such that

$$\Gamma_{nm}(t,r) = \int_0^{2\pi} \int_0^{\pi} z(\boldsymbol{x}, t) Y_{nm}(\theta, \phi) \sin(\theta) d\theta d\phi. \tag{7.14}$$

If we regularly place $Q \geq (N+1)^2$ omni-directional microphones on a sphere of radius R_Q, we can estimate the integration in (7.14) with a finite summation such that

$$\Gamma_{nm}(t, R_Q) \approx \sum_{q=1}^{Q} z(\boldsymbol{x}_q, t) Y_{nm}(\theta_q, \phi_q). \tag{7.15}$$

To simplify the implementation, we sample the signals with a sampling frequency of F_s. Hence, the variable t refers to discrete time intervals hereafter.

We rewrite (7.12) as

$$\begin{aligned} \Gamma_{nm}(t, R_Q) &= \gamma_{nm}(t) * p_n(t, R_Q) \\ &= \sum_{u=-L_p}^{L_p} p_n(u, R_Q) \gamma_{nm}(t-u), \end{aligned} \tag{7.16}$$

where

$$p_n(t, R_Q) = \begin{cases} \frac{c}{2R_Q} P_n(\frac{tc}{R_Q F_s}) & -\frac{R_Q F_s}{c} \leq t \leq \frac{R_Q F_s}{c} \\ 0 & |t| > \frac{R_Q F_s}{c} \end{cases}, \tag{7.17}$$

is a time limited function with $p_n(t, r) \neq 0$ when $-L_p \leq t \leq L_p$, $L_p = \lceil R_Q F_s / c \rceil$ such that the length of $p_n(t, R_Q)$ is $2L_p + 1$.

With (7.16) in hand, our problem reduces to obtaining $\gamma_{nm}(t)$ from the measured $\Gamma_{nm}(t, R_Q)$. This is not achievable since it is an under-determined problem. We always have $2L_p+1$ more unknowns ($\gamma_{nm}(t)$) than knowns ($\Gamma_{nm}(t, R_Q)$). Moreover, this is not practically feasible because the z-transform of $p_n(t, R_Q)$ has zeros on the unit circle, resulting in poles on the unit circle in its direct inverse, making the system unstable. Alternatively, $\gamma_{nm}(t)$ can be extracted from $\Gamma_{nm}(t, R_Q)$ using an appropriately designed filter.

In this chapter, we attempt to design a filtering solution while overcoming the above challenges. It is important to note that the Fourier transform relationship discussed in this section were solely used to formulate the definition of the time-domain spherical harmonic decomposition of a sound field. From this point onward, we will focus on signal processing of the captured sound field only in the time domain.

7.3 Filter Design for Obtaining Time Domain Spherical Coefficients

In Section 7.2, we have presented a method to obtain $\Gamma_{nm}(t, R_Q)$ from recorded sound pressure $z(\boldsymbol{x}_q, t)$ with a spherical microphone array. In this section, we design a series of FIR filters to obtain $\gamma_{nm}(t)$ from given $\Gamma_{nm}(t, R_Q)$.

7.3.1 Stability of Ideal Inverse Filter

Due to the challenges mentioned in Section 7.2, rather than directly using (7.16), we pre-design a series of filters $\rho_n(t, r)$ such that

$$\begin{aligned} &\Gamma_{nm}(t, R_Q) * \rho_n(t, R_Q) \\ &= \gamma_{nm}(t) * p_n(t, R_Q) * \rho_n(t, R_Q) = \gamma_{nm}(t), \end{aligned} \tag{7.18}$$

where

$$p_n(t, R_Q) * \rho_n(t, R_Q) \approx \delta(t). \tag{7.19}$$

We note here that $\rho_n(t, r)$ should be order n dependent but mode m independent, as is the same property with $p_n(t, r)$.

However, we can never achieve a precise $\delta(t)$ in (7.19), as the energy of measured sound pressure at the frequency bins of Bessel zeros has been filtered to zero by $p_n(t, R_Q)$. Therefore, we refrain from designing the inverse filter at these zero positions. Instead, we modify $\delta(t)$ to $\mathcal{Z}_n(t)$ such that its frequency response $\widehat{\mathcal{Z}}_n(f)$ is given by

$$\widehat{\mathcal{Z}}_n(f) = \begin{cases} 1 & |j_n(kr)| \geq \varepsilon \\ 0 & |j_n(kr)| < \varepsilon \end{cases}, \tag{7.20}$$

where ε is a small positive constant threshold which satisfies $j_n(kr) \approx 0$ when $|j_n(kr)| < \varepsilon$. For a fixed R_Q, both $j_n(2\pi f R_Q/c)$ and $\widehat{\mathcal{Z}}_n(f)$ can be seen as a function of f. Figure 7.1 shows $j_n(2\pi f R_Q/c)$ and $\widehat{\mathcal{Z}}_n(f)$ with $\varepsilon = 1/40$ (shown as red and orange dash lines) of the first four orders of n, the black dash line represents the maximum frequency of interest.

From Figure 7.1, we can see that $\widehat{\mathcal{Z}}_n(f)$ is a superposition of a series of rectangu-

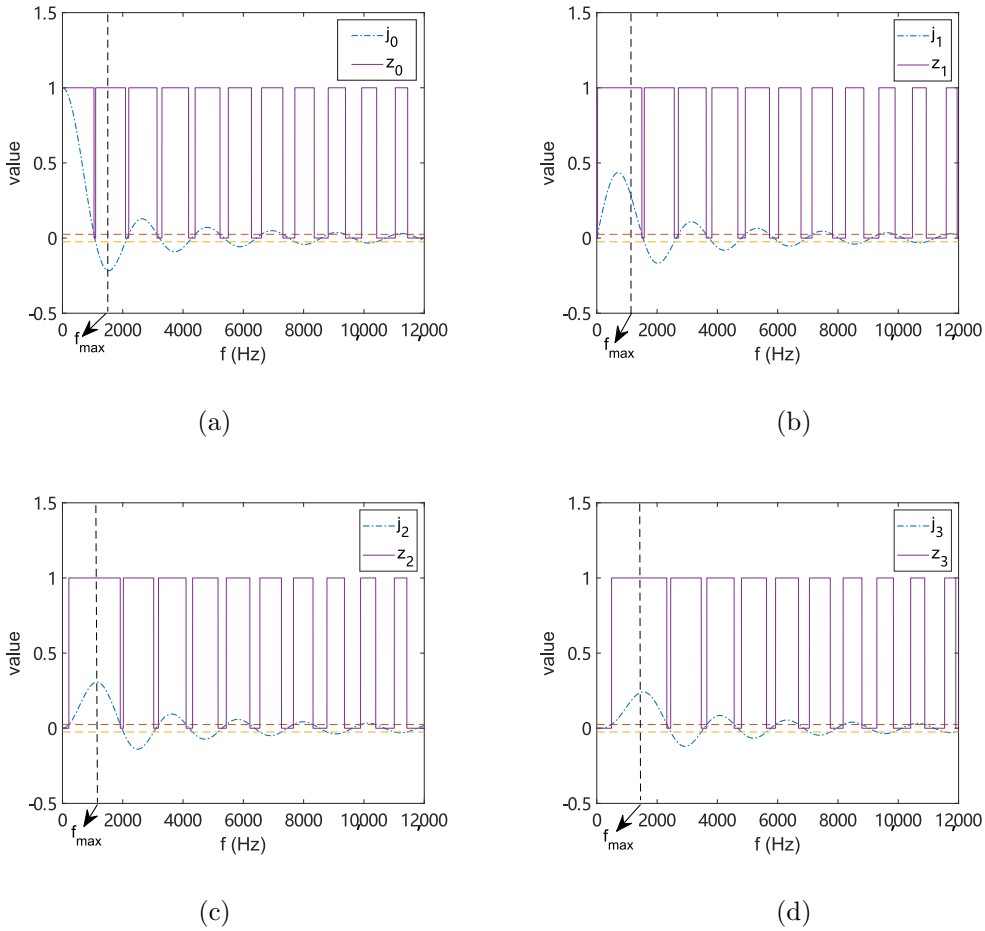


Figure 7.1: The spherical Bessel function $j_n(2\pi f R_Q/c)$ and $\widehat{\mathcal{Z}}_n(f)$ of order (**a**) $n = 0$, (**b**) $n = 1$, (**c**) $n = 2$, (**d**) $n = 3$ with $f_{\max} \approx 1360$ Hz, $\varepsilon = 1/40$, $R_Q = 0.16$ m and $c = 343$ m/s.

lar windows, meaning its inverse Fourier transformation, $\mathcal{Z}_n(t)$, should be a superposition of sinc functions. In practice, due to inherent properties of $j_n(2\pi f R_Q/c)$, for a given maximum frequency $f_{\max}$, the number of active spherical harmonic orders is up to $N \approx \lceil kR_Q \rceil$ [80]. We use the same truncation limit when designing $\widehat{\mathcal{Z}}_n(f)$, resulting in $\mathcal{Z}_n(t)$ to be a superposition with a finite number of sinc functions. The necessity and the influence of this truncation on frequency $f_{\max}$ will be further discussed in Section 7.4.1.

Let us define $\omega^{(n)}$ in radian (rad), such that $j_n(\omega^{(n)} F_s R_Q/c) = \varepsilon$, where ε is the positive threshold we explained in the last paragraph. Therefore, $\omega^{(n)}$ can be considered as the edges of window in $\widehat{\mathcal{Z}}_n(f)$ (see Figure 7.1). Given the vector of $[\omega_1^{(n)}, \omega_2^{(n)}, \omega_3^{(n)}, \cdots]$, we can write $\mathcal{Z}_n$ as

$$\begin{aligned}\mathcal{Z}_n(t) = &\sum_{\imath=1}^{I} (\omega_{2s}^{(n)} - \omega_{2s-1}^{(n)}) \mathrm{sinc}\Big(\frac{\omega_{2s}^{(n)} - \omega_{2s-1}^{(n)}}{2} t\Big) \\ &\times \cos\Big(\frac{\omega_{2s}^{(n)} + \omega_{2s-1}^{(n)}}{2} t\Big),\end{aligned} \tag{7.21}$$

where I is the number of rectangular windows in $\widehat{\mathcal{Z}}_n(f)$ for $-f_{\max} \leq f \leq f_{\max}$. Furthermore, $\omega^{(n)}$ are dependent on the radius of the microphone array R_Q, sampling frequency F_s and the speed of sound c, but the value of $\omega^{(n)} F_s R_Q/c$ remains constant for each order n such that $|j_n(\omega^{(n)} F_s R_Q/c)| = \varepsilon$. The first four order of $\omega^{(n)}$ is given in Table 7.1 with the highest frequency limit of $f_{\max} = 2047$ Hz and sampling frequency $F_s = 48,000$ Hz. Note that for the zero-th order, we set $\omega_1 = 8.9 \times 10^{-4}$ to block DC component in practice.

If we have a series of concentric spherical microphone arrays with the radii of $r_1, r_2, \cdots$, the value of $\omega^{(n)} F_s r_q/c$ would be different from a single sphere model, which can be calculated by $|j_n(\omega^{(n)} r_1/F_s c) + j_n(\omega^{(n)} r_2/F_s c) + \cdots| = \varepsilon$.

7.3.2 Modified Inverse Filter

Now that the design for $\mathcal{Z}_n(t)$ is established, our next step is to design filters $\rho_n(t, R_Q)$ which satisfies

$$p_n(t, R_Q) * \rho_n(t, R_Q) = \mathcal{Z}_n(t). \tag{7.22}$$

Table 7.1: The first four order of $\omega^{(n)}$ to derive $\mathcal{Z}_n(t)$ with $f_{\max} \approx 2047$ Hz, $\varepsilon = 1/40$, $R_Q = 0.16$ m and $c = 343$ m/s.

n (order)	$\omega_1^{(n)}$	$\omega_2^{(n)}$	$\omega_3^{(n)}$	$\omega_4^{(n)}$
0	0.0009	0.1369	0.1439	0.2680
1	0.0033	0.1957	0.2059	0.2680
2	0.0276	0.2509	0.3643	0.2680
3	0.0640	0.2680	-	-

We notice in (7.22) that $p_n(t, R_Q)$ is a finite length vector and we would like $\rho_n(t, R_Q)$ also to be a finite length vector. However, $\mathcal{Z}_n(t)$ is infinitely long with a series of sinc functions. If we perform linear convolution of $p_n(t, R_Q)$ with $\rho_n(t, R_Q)$, we would obtain a vector with the length of $2(L_\rho + L_p) + 1$ samples, where $2L_\rho + 1$ is the filter length of ρ_n, such that $\rho_n(t, R_Q)$ has none-zero values for $-L_\rho \leq t \leq L_\rho$. Thus, we need to truncate the infinite length $\mathcal{Z}_n(t)$ to $2(L_\rho + L_p) + 1$ samples for every order of n where

$$\mathcal{Z}'_n(t) \triangleq \begin{cases} \mathcal{Z}_n(t) & -(L_\rho + L_p) \leq t \leq L_\rho + L_p \\ 0 & \text{otherwise} \end{cases}. \tag{7.23}$$

We can then write (7.22) in a finite summation form as

$$\begin{aligned} \mathcal{Z}'_n(t) &= p_n(t, R_Q) * \rho_n(t, R_Q) \\ &= \sum_{u=-L_\rho}^{L_\rho} p_n(t-u, R_Q)\rho_n(u, R_Q). \end{aligned} \tag{7.24}$$

We rewrite (7.24) into matrix form

$$\boldsymbol{\psi}_n = \boldsymbol{P}_n \boldsymbol{\rho}_n, \tag{7.25}$$

where $\boldsymbol{\psi}_n = [\mathcal{Z}_n(-(L_\rho + L_p)), \mathcal{Z}_n(-(L_\rho + L_p) + 1), \cdots, \mathcal{Z}_n((L_\rho + L_p))]^T$, $\boldsymbol{\rho}_n = [\rho_n(-L_\rho, R_Q), \rho_n(-L_\rho + 1, R_Q), \cdots, \rho_n(L_\rho, R_Q)]^T$, and $\boldsymbol{P}_n$ is the convolution matrix based on the Toeplitz structure of $p_n(t, R_Q)$, given in (7.26).

The size of matrix $\boldsymbol{P}_n$ is $[2(L_\rho + L_p) + 1, 2L + 1]$, where we choose the filter length $2L_\rho + 1$ of $\rho_n(t, R_Q)$ to be significantly larger than both $2L_p + 1$ and the

$$\boldsymbol{P}_n = \begin{bmatrix} p_n(-L_p, R_Q) & 0 & \dots & \dots & \dots & \dots & 0 \\ p_n(-L_p+1, R_Q) & p_n(-L_p, R_Q) & 0 & \dots & \dots & \dots & 0 \\ \vdots & & & \ddots & & & \vdots \\ p_n(L_p, R_Q) & \dots & p_n(-L_p, R_Q) & 0 & \dots & \dots & 0 \\ 0 & p_n(L_p, R_Q) & \dots & p_n(-L_p, R_Q) & 0 & \dots & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & \dots & 0 & p_n(L_p, R_Q) & \dots & p_n(-L_p+1, R_Q) & p_n(-L_p, R_Q) \\ 0 & \dots & \dots & 0 & p_n(L_p, R_Q) & \dots & p_n(-L_p+1, R_Q) \\ \vdots & & & \ddots & & & \vdots \\ 0 & \dots & \dots & \dots & 0 & p_n(L_p, R_Q) & p_n(L_p-1, R_Q) \\ 0 & \dots & \dots & \dots & \dots & 0 & p_n(L_p, R_Q) \end{bmatrix}. \tag{7.26}$$

main lobe width of function $\mathcal{Z}'_n(t)$, to avoid $\boldsymbol{P}_n$ being ill-conditioned and minimize the error of truncating $\mathcal{Z}'_n(t)$ into a finite length signal. The influence of choosing L_ρ will be detailed in Section 7.4.2.

Since (7.25) is an over-determined system of equations, we apply LMS method to (7.25) to obtain

$$\boldsymbol{\rho}_n = \boldsymbol{P}_n^+ \boldsymbol{\psi}_n, \tag{7.27}$$

where $\boldsymbol{P}_n^+$ refers to the Moore-Penrose inverse of $\boldsymbol{P}_n$. As a result, with (7.18) and (7.27), $\gamma_{nm}(t)$ can be estimated by

$$\begin{aligned} \gamma_{nm}(t) &\approx \Gamma_{nm}(t, R_Q) * \rho_n(t, R_Q) \\ &= \big(\sum_{q=1}^{Q} \mathcal{Z}(\boldsymbol{x}_q, t) Y_{nm}(\theta_q, \phi_q) \big) * \rho_n(t, R_Q). \end{aligned} \tag{7.28}$$

In this way we obtain $\gamma_{nm}(t)$ while overcoming the challenges listed in Section 7.2.

7.3.3 Practical Considerations of Filter Implementation

In (7.28), $\gamma_{nm}(t)$ is obtained by filtering $\Gamma_{nm}(t, R_Q)$ with $\rho_n(t, R_Q)$, where we get

$$\Gamma_{nm}(t, R_Q) * \rho_n(t, R_Q) = \gamma_{nm}(t) * \mathcal{Z}'_n(t) \approx \gamma_{nm}(t). \tag{7.29}$$

Naturally $\gamma_{nm}(t)$ at time index t is only influenced by $[\Gamma_{nm}(t-L_p, R_Q), \Gamma_{nm}(t-L_p+1, R_Q), \cdots, \Gamma_{nm}(t+L_p, R_Q)]$ because of the Legendre function in $p_n(t, R_Q)$. However, with the influence of sinc functions in $\mathcal{Z}'_n(t)$ in our proposed filters $\rho_n(t, R_Q)$, we now need the past L_ρ samples and the future L_ρ samples of $\Gamma_{nm}(t, R_Q)$ to obtain $\gamma_{nm}(t)$ at time index t. For offline signal processing, L_ρ samples of zeros should be added both in the beginning and the end of the vector of Γ_{nm} before filtering it with pre-designed $\rho_n(t, R_Q)$. Moreover, an overlap of $2L_\rho + 1$ samples is needed for frame based signal processing. For on-line real time signal processing, we cannot obtain future samples of $\Gamma_{nm}(t, R_Q)$. As a result, we add L_ρ samples of zeros in front of the filter $\rho_n(t, R_Q)$, and create a buffer of the past $2L_\rho + 1$ samples of $\Gamma_{nm}(t, R_Q)$. At time index t, we obtain $\gamma_{nm}(t - L_\rho)$ with the buffer of $[\Gamma_{nm}(t - 2L_\rho, R_Q), \cdots, \Gamma_{nm}(t, R_Q)]$. Thus, there is a L_ρ samples of group delay of the system. We further discuss and compare the group delay with frequency domain method in Section 7.5.5.

7.3.4 Error Analysis

We define the error $\epsilon_{nm}(t)$ as the difference between the desired time domain spherical harmonic and the coefficients we obtained by the proposed method, which can be decomposed to:

$$\epsilon_{nm}(t) = \epsilon_{\text{filter}}(n, m, t) + \epsilon_{\text{position}}(n, m) + \epsilon_{\text{truncation}}(n), \tag{7.30}$$

where $\epsilon_{\text{filter}}(n, m, t)$ is filtering error introduced by $\rho_n(t, R_Q)$, $\epsilon_{\text{truncation}}(n)$ is the truncation error of order N, and $\epsilon_{\text{position}}(n, m)$ is due to the microphones position error. The qualitative analysis of $\epsilon_{\text{truncation}}(n)$ and $\epsilon_{\text{position}}(n, m)$ based on the frequency domain method are addressed in [78], where we draw a similar conclusion in the time domain that with increasing number of microphones and fixed N, $\epsilon_{\text{truncation}}(n)$ decreases. Meanwhile, $\epsilon_{\text{position}}(n, m)$ depends on the nature of inaccurate microphone positioning, referring to the distance between the desired point and microphone location. We mainly focus on $\epsilon_{\text{filter}}(n, m, t)$ here as it is the main error contribution due to the proposed filtering approach.

According to (7.29), $\epsilon_{\text{filter}}(n, m, t)$ at a specific order n and mode m can be expressed

as

$$\begin{aligned}\epsilon_{\text{filter}}(n,m,t) &= |\gamma_{nm}(t) * \mathcal{Z}'_n(t) - \gamma_{nm}(t)| \\ &= |\gamma_{nm}(t) * \epsilon_n(t)|,\end{aligned} \tag{7.31}$$

where

$$\epsilon_n(t) \triangleq \delta(t) - \mathcal{Z}'_n(t). \tag{7.32}$$

Using (7.20) and (7.23), the Fourier transform of $\epsilon_n(t)$ is

$$\epsilon_n(t) \xrightarrow{\mathcal{F}} \Delta_n(f) = \begin{cases} 1 & |j_n(kr)| \leq \epsilon, kr < N \\ 0 & \text{otherwise} \end{cases}, \tag{7.33}$$

with the same truncation in frequency as $\widehat{\mathcal{Z}}_n(f)$. Thus, $\epsilon_n(t)$ can be expressed as

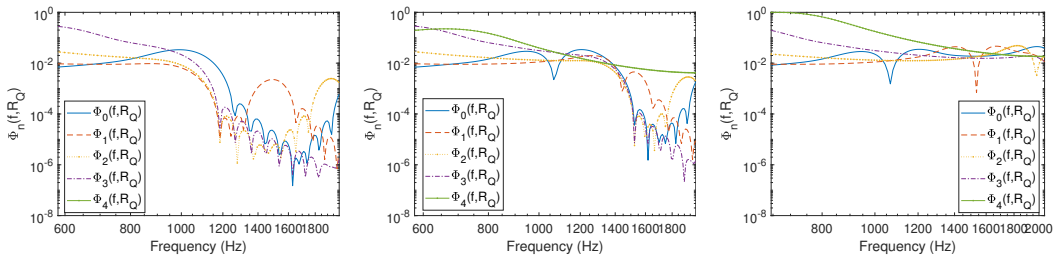
$$\begin{aligned}\epsilon_n(t) = &\sum_{i=0}^{I-1} \big(w_{2s+1}^{(n)} - w_{2s}^{(n)}\big)\text{sinc}\big(\frac{w_{2s+1}^{(n)} - w_{2s}^{(n)}}{2}t\big) \\ &\times \cos\big(\frac{w_{2s+1}^{(n)} + w_{2s}^{(n)}}{2}t\big),\end{aligned} \tag{7.34}$$

where I and $w^{(n)}$ have the same definition as in (7.21) and $w_0^{(n)} = 0$.

With (7.31) and (7.34) we can quantitatively calculate $\epsilon_{\text{filter}}(n,m,t)$ introduced by the filter $\rho_n(t, R_Q)$. The total error caused by filtering can be calculated by a summation of $\epsilon_{\text{filter}}(n,m,t)$ over every order of n and mode of m. As this filtering error is mainly due to Bessel zeros, it can be reduced by limiting the highest order N of the system, where a smaller N results in lower Bessel zeros hence a smaller $\epsilon_{\text{filter}}(n,m,t)$. Also, N depends on the highest wave number k and the radius of the microphone array R_Q. By choosing N with a pre-knowledge of the frequency limit of the input signals and R_Q also helps to minimize the filtering error $\epsilon_{\text{filter}}(n,m,t)$.

7.4 A Filter Design Example

To provide a further understanding of the filter design process, we present a design example of a fourth ($N = 4$) order spherical microphone array of $R_Q = 0.16$ m, designed to record the time domain spherical harmonic coefficients within the spatial region enclosed by the array with a desired frequency band of $[20, 1360]$ Hz.



(a) (b) (c)

Figure 7.2: Frequency response of up to 4-th order of the pre-designed order dependent FIR filter $\rho_n(t, R_Q)$ with $\mathcal{Z}'_n(t)$ frequency truncated at (**a**) $f_1 = 1023.6$ Hz, (**b**) $f_2 = 1364.8$ Hz, and (**c**) $f_3 = 2047.1$ Hz.

Let sampling frequency to be $F_s = 48,000$ Hz and $c = 343$ m/s. Before we apply the proposed method to recording signals, we first analyze the influence of several steps in designing the proposed filter $\rho_n(t, R_Q)$.

7.4.1 Effect of Frequency Truncation of $Z_n(t)$

As audio signals are often band limited in ANC applications [9], we can have a finite truncation on spherical harmonic decomposition with order $N = \lceil kR_Q \rceil$. In other words, if we have a fixed N-th order system, the highest frequency that the system can successfully capture is given by $f_{\max} = Nc/(2\pi R_Q) \approx 1360$ Hz. Figure 7.2 shows the frequency response of $\rho_n(t, R_Q)$, refers to $\Phi_n(f, R_Q)$, which is designed using (7.27) with $\mathcal{Z}'_n(t)$ truncated at $f_1 = 1023.6$ Hz (Figure 7.2a), $f_2 = 1364.8$ Hz (Figure 7.2b), $f_3 = 2047.1$ Hz (Figure 7.2c), respectively. The filter length is set to be 500. To obtain the frequency response of $\rho_n(t, R_Q)$, a FFT of 4096 points is applied with zero padding to $\rho_n(t, R_Q)$. We recall here that $\mathcal{Z}'_n(t)$ is given by (7.21) in the time domain, which does not rely on any frequency domain processing.

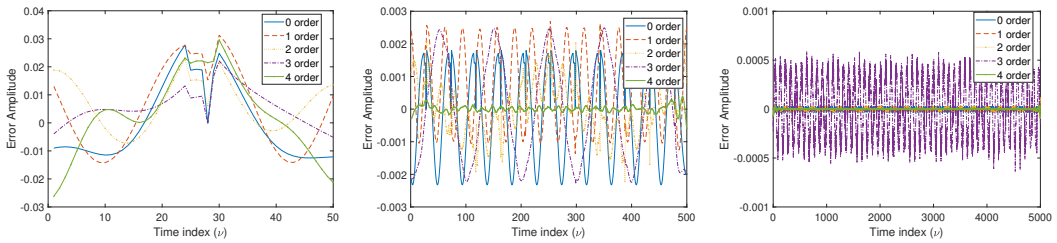
We observe that for a $N = 4$th order system, the truncation at f_1 is not enough to get an accurate frequency response of $\rho_n(t, R_Q)$, as the frequency response $\Phi_n(f, R_Q)$ begins to decline at f_1. In this case, $\rho_n(t, R_Q)$ can not provide an acceptable filtering result with signals containing higher frequency components. Truncation at both f_2 and f_3 can give a satisfied frequency response when $f < f_{\max}$,

that the frequency response remains at the desired amplitude for $f < f_{\max}$. As the frequency range of the system is also limited by $N = \lceil kR_Q \rceil$, it is not necessary to look at the frequency response when $f > f_{\max}$. So in both cases $\rho_n(t, R_Q)$ can give an acceptable filtering output. As a result, we choose to truncate $\mathcal{Z}'_n(t)$ at $f_{\max}$, where $2\pi f_{\max} R_Q/c = N$. If the recorded signal is known as a band limited signal where its highest frequency component is less than $f_{\max}$, an alternative choice of the frequency truncation of $\mathcal{Z}'_n(t)$ is at this highest frequency to reduce the computation complexity. Meanwhile, if $\mathcal{Z}'_n(t)$ has been designed with a higher frequency truncation, it can also be used in a lower order system with a lower requirement of frequency truncation.

7.4.2 Choice of Filter Length of $\rho_n(t, R_Q)$

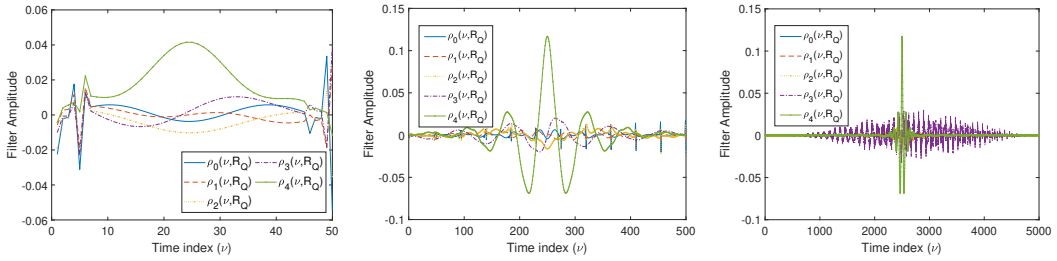
Intuitively, a longer filter often brings us less error and better performance. Figure 7.3 supports this idea by showing the result of $\rho_n(t, R_Q) * p_n(t, R_Q) - \mathcal{Z}'_n(t)$ with different choices of L_ρ, which refers to the error introduced into the system by the filtering processing. We observe that the filtering error decreases across all of the orders with a higher L. This is due to the time truncation of $\mathcal{Z}'_n(t)$ (length of vector $\boldsymbol{z}_n$ in (7.27)), being related to L_ρ. Thus, a higher L_ρ leads to less information loss in the time truncation of $\mathcal{Z}'_n(t)$, hence smaller error in $\rho_n(t, R_Q)$. However, Figure 7.4 shows the time domain filter $\rho_n(t, R_Q)$ with different lengths. We observe that a longer filter results in a higher group delay of filtering. This is not desirable because it leads to a higher system delay of our proposed method, while lowering the system delay is one of the most important motivations that we develop the proposed time domain method.

As a result, we need to balance the noise tolerance, group delay, and the filtering error when we choose L. We suggest that filter length $2L_\rho+1$ should be significantly larger than the main lobe width of $\mathcal{Z}'_n(t)$ and $2L_P+1$, the length of $p_n(t, R_Q)$, but no more than 50 times of $2L_P + 1$. Additionally, L should be less than the maximum tolerance of the delay of the system. Based on these guidelines, for the current example, we choose $2L_\rho + 1 = 501$.



(a) (b) (c)

Figure 7.3: Error of $\rho_n(t, R_Q) * p_n(t, R_Q)$ with length $L_\rho =$ (**a**) 25, (**b**) 250 (**c**) 2500.



(a) (b) (c)

Figure 7.4: Time representation of the pre-designed order dependent FIR filter $\rho_n(t, R_Q)$ with length $L_\rho =$ (**a**) 25, (**b**) 250 (**c**) 2500.

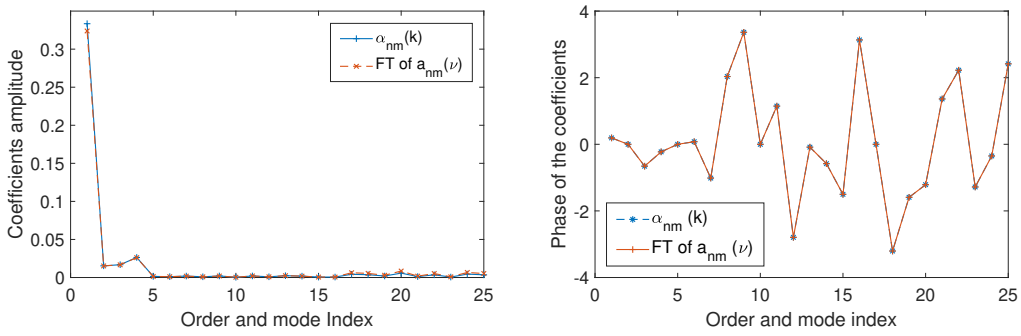
7.5 Simulation Results and Analysis

In this section, we evaluate the result of the proposed algorithm for time domain spherical harmonic analysis using a fourth order ($N = 4$) system. We consider 32 microphones regularly placed on an open spherical array of $R_Q = 0.16$ m, where the analysis region of interest is inside the array. A point source is placed at $[1, 2, 1]$ m with respect to the origin which coincides with the origin of the microphone array. The sampling frequency is 48,000 Hz, and the filter length $2L_\rho + 1$ is 501. A noise signal at 40 dB SNR is added to each microphone to reflect thermal noise. Considering the application of the proposed method to be spatial ANC, we construct the desired frequency band to cover the target noise band, and construct the radius of the region to be wide enough to fit one human head.

It is difficult to validate our method in the time domain directly because the coefficients are time dependent and no ground truth has been given. Therefore, we first validate our proposed time domain spherical harmonic coefficients in the frequency domain. Thus, we compare the Fourier transformation of the time domain coefficients to the theoretical frequency domain coefficients given in (7.4). Next, to clarify that our proposed method has the ability to record a sound field in the region of interest in the time domain, we reconstruct sound pressure at an arbitrary point as well as over a plane inside the region of interest with the captured time domain spherical harmonic coefficients by (7.6). Finally, the time delay of the proposed method is given.

7.5.1 Comparison between the Time Domain and the Frequency Domain Spherical Harmonic Coefficients

We use a narrow band signal at 1200 Hz to test if our proposed method can obtain the time domain spherical harmonic coefficients $\gamma_{nm}(t)$ correctly with (7.28). In (7.13), we give the relationship between $\gamma_{nm}(t)$ and $\alpha_{nm}(k)$. We compare the Fourier transformation result of our obtained time domain spherical harmonic coefficients $\mathcal{FT}\{\gamma_{nm}(t)\}$ with the desired frequency domain spherical harmonic coefficients $\alpha_{nm}(k)$, obtained by Equation (7.4) in frequency domain. Fourier transformations use 1024 points. We do not compare the phase of these coefficients since



(a) (b)

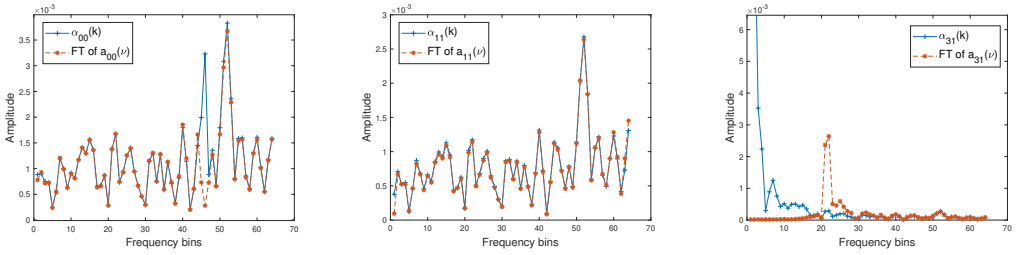
Figure 7.5: (**a**) Amplitude and (**b**) phase difference comparison between the Fourier Transform of the time domain spherical harmonic coefficients $\gamma_{nm}(t)$ and frequency domain spherical harmonic coefficients $\alpha_{nm}(k)$ at a single frequency $f = 1200$ Hz.

the group delay of the time domain method and the frequency domain method is different. Instead, we compare the phase difference, given by $\alpha_{nm}(k) - \alpha_{n(m-1)}(k)$. The results of both amplitude and phase difference are shown in Figure 7.5.

In Figure 7.5 we see that there is little to no difference on both amplitude and phase difference between the Fourier Transformed time domain coefficients and the frequency domain coefficients over all the order and modes. Thus, our proposed time domain method successfully obtained the time domain spherical harmonic coefficients, which can be related to the frequency domain coefficients by Fourier transformation.

Next, we compare the coefficients over different frequencies with a wide band test signal within the frequency limited of $[20, 1300]$ Hz. In Figure 7.6, we show the comparison of amplitude at $FT\{a_{00}(t)\}$ and $\alpha_{00}(k)$, $FT\{a_{11}(t)\}$ and $\alpha_{11}(k)$, and $FT\{a_{31}(t)\}$ and $\alpha_{31}(k)$ over frequencies respectively while Figure 7.7 shows the phase difference.

A huge error is observed in Figure 7.6a at the 46th frequency bin. This error is due to the fact that there is a Bessel zero of the zeroth order at this frequency bin (around 1072 Hz). We see the frequency domain spherical harmonic coefficients $\alpha_{00}(k)$ has a much higher amplitude, while our proposed method suppressed the amplitude at this certain frequency bin.



(a) (b) (c)

Figure 7.6: Amplitude comparison between the Fourier Transform of the time domain spherical harmonic coefficients $\gamma_{nm}(t)$ and frequency domain spherical harmonic coefficients $\alpha_{nm}(k)$ at mode (**a**) 00, (**b**) 11 and (**c**) 31 with a white Gaussian noise.

Meanwhile, we can see in Figures 7.6 and 7.7 that the error at $a_{31}(t)$ is higher compared to the other two modes. As order increases, the error increases. This error can be decreased by applying more microphones on the array. We also obtain a non-negligible error before the 30th frequency bin of the coefficients amplitude for $(n, m) = (3, 1)$ in Figures 7.6c and 7.7c. This error is because our time domain proposed method and conventional frequency domain method have different processing approaches for suppressing Bessel zeros. During the reconstruction process, the high pass property of spherical Bessel function removes the information at this frequency bin. Thus, this error will not influence the reconstruction of the sound field.

7.5.2 Sound Pressure Comparison at a Point Of Interest

In this section, we reconstruct the sound field with the captured time domain spherical harmonic coefficients at a point in the region of interest, and compare it with the desired sound field at the same point of interest. We use a signal containing three frequency components of 600 Hz, 850 Hz, and 1300 Hz. Figure 7.8 shows the desired sound pressure and the reconstructed sound pressure calculated by $\gamma_{nm}(t)$ at the point $[-0.13, 0.07, 0.02]$ m and $[-0.03, 0.01, 0.1]$ m inside the region of interest in the time domain. The desired sound field has been manually delayed for 272 samples to match the group delay of the reconstructed sound field, where

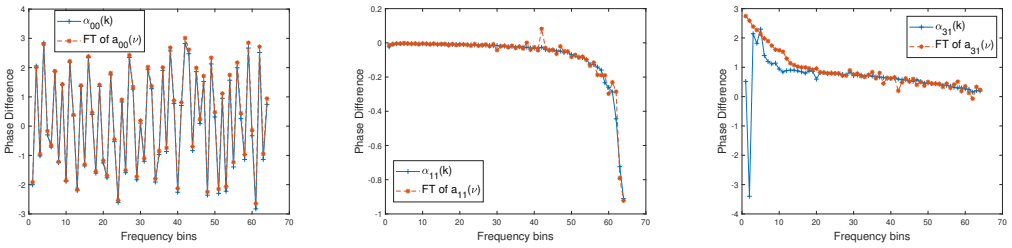
the details of this delay will be shown in Section 7.5.5.

We note here that when reconstructing the sound-field with (7.6), we face the problem that at a point $\boldsymbol{x} = (r, \theta, \phi)$ where the radius r is very small, the filter $p_n(t, r)$, whose filter length dependent on rF_s/c, is too short to perform efficient filtering. To overcome this problem, we up-sample the obtained $\gamma_{nm}(t)$ with a rate of R_Q/r and construct corresponding $p_n(t, r)$ with the same length of $L_p = 2 * R_Q F_s/c + 1$. We then down-sample the resulting $\Gamma_{nm}(t, r)$ with a rate of r/R_Q to keep the sampling frequency consistent with F_s. The effective construction of the sound field with small r will be shown in the next section. We can see from Fig. 7.8 that the obtained $\gamma_{nm}(t)$ by our proposed method can successfully reconstruct the sound pressure at a point inside the region of interest with a tolerable error. This supports that our time domain coefficients contain certain spatial information of the sound field that the sound pressure at an arbitrary point inside the region of interest can be properly calculated with the measurements only being taken on the boundary of the region.

7.5.3 Sound Field Comparison over a Plane

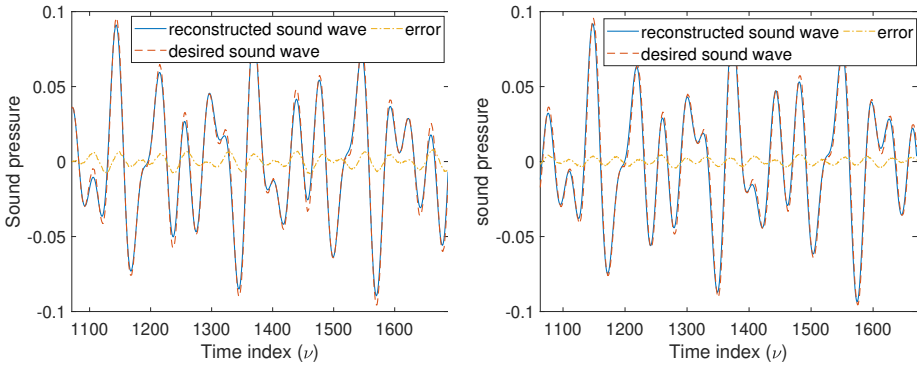
To further evaluate our method on reconstructing sound field over space, we now reconstruct the sound field by $\gamma_{nm}(t)$ over a plane. We use a narrow band signal of 1200Hz here such that the sound field in the region of interest is simple and clearly understood. Although the sound field is reconstructed over time, a 2D plot can only show the result of one time index. Figure 7.9 shows the reconstructed sound field and the desired sound field over the plane parallel to the x-y plane, with $z = 0.02$ m at $t = 0.3$ s. The 272 samples group delay is manually fixed and will be discussed later in the next subsection.

The white line in Figure 7.9 bounds the region of interest. We can see that the reconstructed sound field inside this region in Figure 7.9a is roughly the same as the desired sound field in Figure 7.9b. This confirms that the coefficients recorded by our proposed method are able to capture the sound field inside the region of interest.



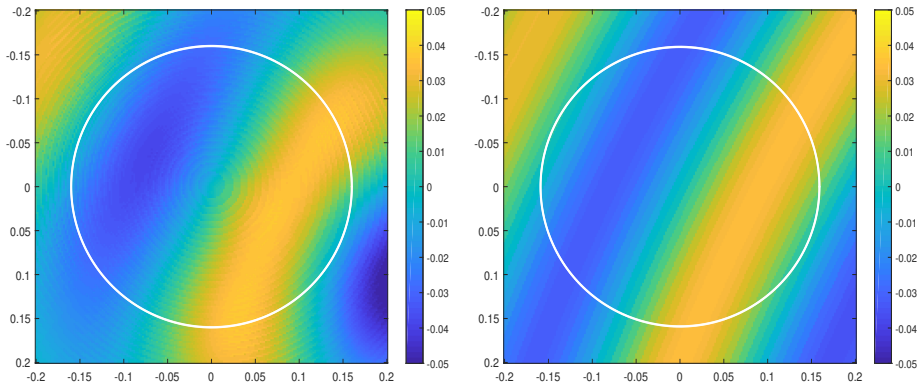
(a) (b) (c)

Figure 7.7: Phase difference comparison between the Fourier Transform of the time domain spherical harmonic coefficients $\gamma_{nm}(t)$ and frequency domain spherical harmonic coefficients $\alpha_{nm}(k)$ at mode (**a**) 00, (**b**) 11 and (**b**) 31 with a white Gaussian noise.



(a) (b)

Figure 7.8: Comparison between reconstructed sound pressure and desired sound pressure at the point (**a**) $(-0.13, 0.07, 0.02)$ m and (**b**) $(-0.03, 0.01, 0.1)$ m.



(a) (b)

Figure 7.9: Comparison between (**a**) reconstructed sound field and (**b**) desired sound field at the horizontal plane with a height of z = 0.02 m.

7.5.4 Sound Field Error Estimation over The Region

To evaluate the reconstructed sound field over time, we calculate the instantaneous average squared spatial error over time, which is defined by

$$\eta(t) \triangleq \frac{\sum_{\Omega} \|z_r(\boldsymbol{x}, t) - z_d(\boldsymbol{x}, t)\|^2}{\sum_{\Omega}}, \tag{7.35}$$

where $z_r(\boldsymbol{x}, t)$ and $z_d(\boldsymbol{x}, t)$ refers to the reconstructed and the original sound pressure at the point $\boldsymbol{x}$, respectively.

Figure 7.10 shows how the error fluctuates with time in a tolerable range (no more than 5×10^{-4}) with a 900 Hz tone and a 1072 Hz tone. We have already observed in Figure 7.8 that the error of the sound pressure at a point of interest is proportional to the desired sound pressure. We observe the same trend when we evaluate the error over the region that the error increases when the sound field inside the region of interest is at peak amplitude. We also observe in Figure 7.10 that the error with 1072 Hz signal is higher than 900 Hz signal. This is due to the fact that there is a Bessel zero of the zeroth order ($j_0(kr)$) at 1072 Hz in the proposed spatial ANC system. Hence, the amplitude of $a_{00}(t)$ is suppressed by the proposed method, leading to a higher error in reconstructing the sound field.

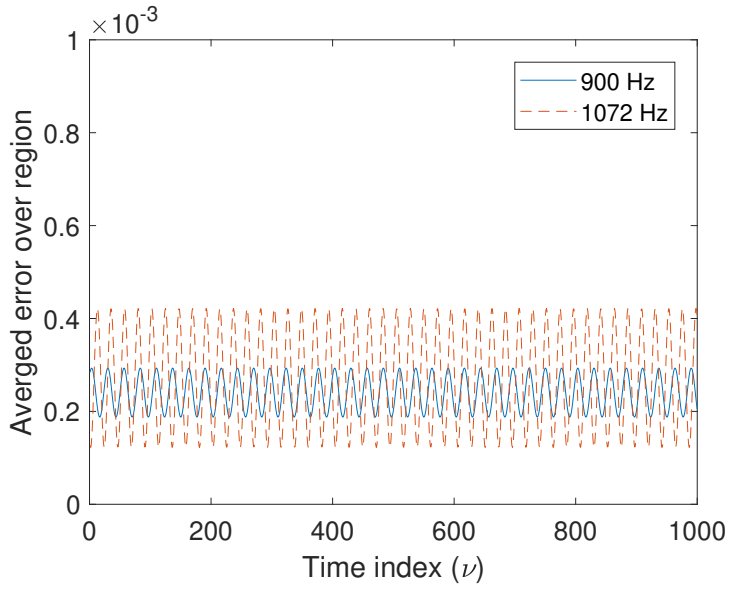


Figure 7.10: Instantaneous region averaged squared spatial error of the proposed method for sound field reconstruction over space at 900 Hz and 1072 Hz (Bessel zero).

7.5.5 Processing Delay Analysis

In this section, we indicate the group delay of our method. Figure 7.11 shows the desired sound pressure and the reconstructed sound pressure of a signal containing three frequency components of $[600, 850, 1300]$ Hz at the point $[-0.13, 0.07, 0.02]$ m. We can obtain from Figure 7.11 that the processing delay of the system is $1046 - 774 = 272$ samples, which equals to $L_\rho + \lceil R_Q F_s/c \rceil$. The L_ρ samples of the delay is from the group delay of the proposed filter $\rho_n(t, R_Q)$, while $R_Q F_s/c$ is the delay introducing by the Legendre function within filter $p_n(t, r)$ to reconstruct the sound pressure at a point with the time domain spherical harmonic coefficients. Comparing to a conventional frequency domain scenario with 512 frame-size and 75% of overlap Short Time Fourier transformation with overlap save method, which refers to a 2048 samples (4 frames) of delay [179], our proposed method can significantly reduce the processing delay. Additionally, the frame size (and the number of FFT points) can not be shorter in most of the scenarios to maintain the frequency resolution in the frequency domain.

Comparing to one of the start-of-art frequency domain spherical harmonic filter

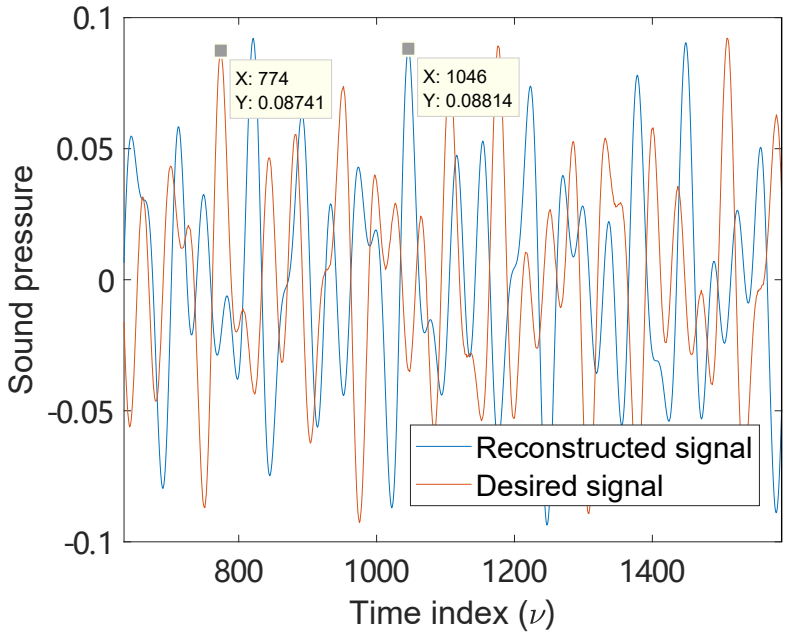


Figure 7.11: Delay analysis between the desired signal and the reconstructed signal at a point inside the region of interest.

designs [180], which states a 75 ms delay with a 900 sample long filter, our method can achieve a 972 samples (20.25 ms with 48k Hz sampling frequency) delay with the same length of filter. Meanwhile, as our method is processed in the time domain, there is nothing to stop us from doing a sample by sample signal processing instead of frame based signal processing. This sample based processing considerably extends the application of spherical harmonic analysis.

7.6 Conclusions

In this chapter, a time domain spherical harmonic analysis method for spatial sound field recording over 3D space has been developed with the goal to minimize processing delay. This favours the specific application of spatial ANC. With the proposed FIR filter designing, the time domain spherical harmonic coefficients can be obtained from sound pressure measurements of an open spherical microphone array. The filters are designed based on the inverse of the Legendre function. Additionally, the filters are modified with considerations of stability and practical

implementation. We have provided simulation results demonstrating the validity of the proposed method.

We note that by obtaining the proposed time domain spherical harmonic coefficients, the desired sound field can be efficiently captured and reconstructed over space. The proposed time domain spherical harmonic coefficients can be related to the conventional frequency domain coefficients, and both of the coefficients have the same location independent property. The proposed method has the prominent advantage of lower delay since it is developed in the time domain without the introduction of a Fourier transformation or inverse Fourier transformation. Furthermore, the proposed time domain filtering method can support the sample based signal processing instead of the frame based, which indicates that the frame size can be one sample if necessary. As a result, we consider the proposed time domain spherical harmonic analysis method to be highly suitable for a spatial ANC system where accurate spatial recording with low delay is desired.

7.7 Related Publications

Much of this chapter's work has been published in the following journal proceedings.

- H. Sun, T. D. Abhayapala, and P. N. Samarasinghe, "Time Domain Spherical Harmonic Processing with Open Spherical Microphones Recording," *Applied Sciences*, vol. 11, no. 3, pp. 1074, Jan. 2021.

Chapter 8

Conclusions and Future Work

In this chapter, we state the general conclusions drawn from this thesis. We also outline some future research directions arising from this work.

8.1 Conclusions

To answer the question we raised in Chapter 1, several feasible designs have been addressed for spatial ANC in this thesis, aiming for people to enjoy silence.

The array geometry of spatial ANC, especially the error microphone arrays, has been optimized in Chapter 3 and Chapter 4. These methods not only avoid the usage of impractical spherical arrays but also allow human heads to freely move in to and out of the region of interest. Hence, compared to the conventional methods where microphone arrays block the region of interest, the proposed ANC systems are able to better serve people a quieter region around their heads.

In chapter 5, a secondary channel estimation method with a moving higher-order microphone is detailed. We achieve secondary channel estimation without the usage of the impractical error microphone array. This method, combined with the remote microphone technique addressed in Chapter 4, can be applied to spatial ANC systems to fully avoid the use of impractical spherical microphone arrays. Moreover, the combined system has flexible physical array geometry, which makes it possible for the system to suit different environments where the potential microphone positions may be constrained.

Beyond the optimization of practical array geometry, we also work on minimizing the system delay and hardware resource usage of a spatial ANC system. Chapter 6 and 7 provide the time-wave domain spatial ANC system and the method for the time domain spherical harmonic coefficients recording, respectively. While the wave domain processing enlarges the application of ANC systems, the development of the time-wave domain processing with less delay and fewer hardware resources can further improve the chance of a spatial ANC system to be realized in a real-world environment.

Overall, this thesis mainly works on spatial ANC systems which can be realized in a real-world environment. Nowadays, the concept of ANC is becoming more and more popular in the public, thanks to the commercialization of ANC headphones and earphones. More and more people are paying attention to products with ANC functions built in. The author believes that there is a significant potential for spatial ANC to be commercially applied in wider scenarios (e.g., in virtual and augmented physical spaces or inside autonomous cars/ flights) and for spatial ANC to benefit a larger population. This thesis is also a part of the effort toward commercializing spatial ANC, with the expectation for a brighter future.

8.2 Future Research

There are a number of future research problems arising from the work presented in this thesis. We list some possible problems here.

Implement the time-wave domain spatial ANC with more feasible array geometry.

While the time-wave domain spatial ANC is developed in Chapter 7, the usage of an impractical open spherical microphone array remains necessary. To be more feasible, a time-wave domain method can be developed for the multiple circular array system we proposed in Chapter 3.

Experimental validation of the spatial ANC system with the remote microphone technique.

In Chapter 4, we introduced a spatial ANC system using the remote microphone technique. In this method, the spherical error microphone array is not necessary during the noise reduction. Meanwhile, we have the secondary channel estimation method in Chapter 5, which estimates the secondary channels with a moving HOM instead of an error microphone array. With these two methods in hand, a spatial ANC system fully avoiding the usage of spherical microphone arrays can be proposed and implemented. Without the requirement of impractical arrays, experiments of spatial ANC in a real world environment become possible. Experimental validation can strongly support the reliability of a spatial ANC system in the real life.

Perceptual validation of a spatial ANC system

Optimizing for the feasibility of spatial ANC always introduces more errors into the system. In other words, we are sacrificing the noise reduction performance to make the implementation of the system more feasible. It would be nice to have an experiment validation about the sensitivity of human ears regarding the spatial ANC performance. Can people tell the difference between a 20 dB noise reduction and a 25 dB noise reduction? If so, can people tell the difference between a 20 dB noise reduction and a 21 dB noise reduction? Where is the boundary? Is this boundary related to the type (narrow-band or wide-band) and the frequency spectrum of the noise? Additionally, is the result the same with the wave domain ANC systems and the conventional ANC systems? A perceptual test can provide us with the answer. Similar research can be found for binaural room impulse response in [181]. In Chapter 3, a spatial ANC system providing satisfied noise reduction performance over a 3D region with a human head in the region has been addressed. With this experiment, we are able to control the noise reduction performance by controlling the amplitude of the driving signals of the secondary loudspeakers or the amplitude of the noise signal.

Secondary channel estimation in non-stationary environments

As the spatial ANC system is becoming reliable, there are more problems that should be taken into consideration. One of them is, what's the influence of the user's movement? There are several kinds of movement: 1) moving inside the region; 2) moving from outside the region to inside or vice versa; 3) moving outside the region. These movements can lead to changes in secondary channels. The secondary channels are pre-measured, and are usually assumed to be stable during noise control. However, this assumption is not always true. In extreme cases, for example, what would happen if a table is moved out of the room while a spatial ANC system is active? Online secondary channel estimations [132] can partly solve this problem. However, the remote microphone technique, as well as other virtual sensing techniques, are not able to fit online secondary channel estimation. We need alternative solutions for non-stationary secondary channels.

Multi-zone spatial ANC

In this thesis, the spatial ANC system focuses on creating a single quiet region in 3D space. In real life, multiple quiet regions in the same room can be necessary. For example, in cars or aircraft, one region is necessary for each passenger. These regions may share the same noise sources and secondary loudspeaker arrays. In this case, a special algorithm is needed to take care of multiple regions. Moreover, a quiet region with another 'no influence' region may also be necessary in some cases. As a spatial ANC system influences the sound field not only inside the region of interest, but also the sound field in the entire environment, there can be another region in the same environment requiring no change of the original noise field. Researches (such as [182]) in blocking the noise in one region while keeping the original sound in another region are necessary. Hence, designing and implementing the adaptive filtering algorithm for the multi-zone ANC system is worth being achieved.

Bibliography

[1] Access Economics Pty Ltd, *Listen hear! the economic impact and cost of hearing loss in Australia*, Access Economics, 2006.

[2] 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.

[3] C. M. Harris, *Handbook of acoustical measurements and noise control*, McGraw-Hill New York, 1991.

[4] "Figure: what-is-anc-technology-how-does-it-work," https://www.cardinalpeak.com/blog/what-is-anc-technology-how-does-it-work, 2022.

[5] S. J. Elliott and P. A. Nelson, "Active noise control," *IEEE signal processing magazine*, vol. 10, no. 4, pp. 12–35, 1993.

[6] C. C. Fuller, S. Elliott, and P. A. Nelson, *Active control of vibration*, Academic Press, 1996.

[7] Apple Pty Ltd, "Airpods max," https://www.apple.com/airpods-max/, 2021.

[8] S. M. Kuo and D. Morgan, *Active noise control systems: algorithms and DSP implementations*, John Wiley Sons, Inc., 1995.

[9] S. M. Kuo and D. Morgan, *Active noise control systems: algorithms and DSP implementations*, John Wiley & Sons, Inc., 1995.

[10] 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.

[11] W. Jung, S. J. Elliott, and J. Cheer, "Local active control of road noise inside a vehicle," *Mechanical Systems and Signal Processing*, vol. 121, pp. 144–157, 2019.

[12] 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.

[13] T. Haase, O. Unruh, S. Algermissen, and M. Pohl, "Active control of counter-rotating open rotor interior noise in a dornier 728 experimental aircraft," *Journal of Sound and Vibration*, vol. 376, pp. 18–32, 2016.

[14] T. Murao, C. Shi, W. S. Gan, and M. Nishimura, "Mixed-error approach for multi-channel active noise control of open windows," *Applied Acoustics*, vol. 127, pp. 305–315, 2017.

[15] P. Belanger, A. Berry, Y. Pasco, O. Robin, Y. St-Amant, and S. Rajan, "Multi-harmonic active structural acoustic control of a helicopter main transmission noise using the principal component analysis," *Applied Acoustics*, vol. 70, no. 1, pp. 153–164, 2009.

[16] A. Montazeri and C. J. Taylor, "Modeling and analysis of secondary sources coupling for active sound field reduction in confined spaces," *Mechanical Systems and Signal Processing*, vol. 95, pp. 286–309, Oct. 2017.

[17] Hyundai Motor Group, "Hyundai's world's first road-noise active noise control, ranc - hyundai motor group tech," https://tech.hyundaimotorgroup.com/article/hyundais-worlds-first-road-noise-active-noise-control-ranc/, 2022.

[18] Silentium, "Enhance your driving experience with active noise control," https://www.silentium.com/automotive-2/, 2022.

[19] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, academic press, 1999.

[20] J. Zhang, T. D. Abhayapala, W. Zhang, P. N. Samarasinghe, and S. Jiang, "Active noise control over space: A wave domain approach," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 4, pp. 774–786, 2018.

[21] T. D. Abhayapala, D. B. Ward, et al., "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, vol. 2, pp. 1949–1952.

[22] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Audio Speech Lang. Processing*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[23] D. Moreau, B. Cazzolato, A. Zander, and C. Petersen, "A review of virtual sensing algorithms for active noise control," *Algorithms*, vol. 1, no. 2, pp. 69–99, 2008.

[24] S. Elliott, C.K. Lai, T. Vergez, and J. Cheer, "Robust stability and performance of local active control systems using virtual sensing," in *23rd International Congress on Acoustics, integrating 4th EAA Euroregio 2019*, 2019.

[25] Y. Xiao, L. Ma, and K. Hasegawa, "Properties of fxlms-based narrowband active noise control with online secondary-path modeling," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2931–2949, 2009.

[26] P. Lueg, "Process of silencing sound oscillations," *US pat ent 2043416*, 1936.

[27] P. A. Nelson and S. J. Elliott, *Active control of sound*, Academic press, 1991.

[28] B. Widrow and E. Walach, "Adaptive signal processing for adaptive control," *IFAC Proceedings Volumes*, vol. 16, no. 9, pp. 7–12, 1983.

[29] S. Kuo and C. Chen, "Implementation of adaptive filters with the tms320c25 or the tms320c30, indigital," 1991.

[30] J. C. Burgess, "Active adaptive sound control in a duct: A computer simulation," *The Journal of the Acoustical Society of America*, vol. 70, no. 3, pp. 715–726, 1981.

[31] T. Wang and W. S. Gan, "Stochastic analysis of fxlms-based internal model control feedback active noise control systems," *Signal Processing*, vol. 101, pp. 121–133, 2014.

[32] S. J. Elliott and P. A. Nelson, "The application of adaptive filtering to the active control of sound and vibration," *NASA STI/Recon Technical Report N*, vol. 86, pp. 32628, 1985.

[33] C. H. Hansen, S. D. Snyder, X. Qiu, L. A. Brooks, and D. J. Moreau, *Active control of noise and vibration*, Spon London, 1997.

[34] E. A. Wan, "Adjoint lms: An efficient alternative to the filtered-x lms and multiple error lms algorithms," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. IEEE, 1996, vol. 3, pp. 1842–1845.

[35] S. Elliott, I. Stothers, and P. Nelson, "A multiple error lms algorithm and its application to the active control of sound and vibration," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1423–1434, 1987.

[36] J. K. Thomas, S. P. Lovstedt, J. D. Blotter, and S. D. Sommerfeldt, "Eigenvalue equalization filtered-x algorithm for the multichannel active noise control of stationary and nonstationary signals," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4238–4249, 2008.

[37] M. Bouchard, "Multichannel affine and fast affine projection algorithms for active noise control and acoustic equalization systems," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 1, pp. 54–60, 2003.

[38] S. Pradhan, X. Qiu, and J. Ji, "Affine combination of the filtered-x lms/f algorithms for active control," in *Vibration Engineering for a Sustainable Future*, pp. 313–319. Springer, 2021.

[39] J. Cheer and S. J. Elliott, "Multichannel control systems for the attenuation of interior road noise in vehicles," *Mechanical Systems and Signal Processing*, vol. 60, pp. 753–769, 2015.

[40] A. Barkefors, S. Berthilsson, and M. Sternad, "Extending the area silenced by active noise control using multiple loudspeakers," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 325–328.

[41] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, "Error analysis of a practical energy density sensor," *The Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 211–222, 2000.

[42] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, "Narrowband and broadband active control in an enclosure using the acoustic energy density," *The Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 192–203, 2000.

[43] J. Xie, D. Jin, W. Zhang, X. Zhang, J. Chen, and D. Wang, "Robust sparse multichannel active noise control," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 521–525.

[44] K. Iwai, S. Kinoshita, and Y. Kajikawa, "Multichannel feedforward active noise control system combined with noise source separation by microphone arrays," *Journal of Sound and Vibration*, vol. 453, pp. 151–173, 2019.

[45] R. P. Monteiro, G. A. Lima, J. Oliveira, D. Cunha, and C. Bastos-Filho, "Improving adaptive filters for active noise control using particle swarm optimization," *International Journal of Swarm Intelligence Research (IJSIR)*, vol. 9, no. 4, pp. 47–64, 2018.

[46] N. V. George and G. Panda, "A particle-swarm-optimization-based decentralized nonlinear active noise control system," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 12, pp. 3378–3386, 2012.

[47] D. Shi, B. Lam, K. Ooi, X. Shen, and W. S. Gan, "Selective fixed-filter active noise control based on convolutional neural network," *Signal Processing*, vol. 190, pp. 108317, 2022.

[48] I. T. Ardekani and W. H. Abdulla, "Effects of imperfect secondary path modeling on adaptive active noise control systems," *IEEE transactions on control systems technology*, vol. 20, no. 5, pp. 1252–1262, 2011.

[49] Y. Xiao, L. Ma, and K. Hasegawa, "Properties of fxlms-based narrowband active noise control with online secondary-path modeling," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2931–2949, 2009.

[50] C. Chang, S. M. Kuo, and C. Huang, "Secondary path modeling for narrowband active noise control systems," *Applied Acoustics*, vol. 131, pp. 154–164, 2018.

[51] X. Guo, Y. Li, J. Jiang, C. Dong, S. Du, and L. Tan, "Sparse modeling of nonlinear secondary path for nonlinear active noise control," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 482–496, 2018.

[52] J. Lu, X. Qiu, and B. Xu, "The application of adaptive iir filter in active noise control with consideration of strong acoustic feedback," in *Proc. Int. Conf. Acoust*, 2004, vol. 3.

[53] M. T. Akhtar, M. Abe, and M. Kawamata, "On active noise control systems with online acoustic feedback path modeling," *IEEE transactions on audio, speech, and language processing*, vol. 15, no. 2, pp. 593–600, 2007.

[54] T. Bai, Z. Wang, Y. Xiao, Y. Ma, L. Ma, and K. Khorasani, "A multi-channel narrowband active noise control system with simultaneous online secondary- and feedback-path modeling," in *2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. IEEE, 2019, pp. 289–292.

[55] L. Wu, X. Qiu, I. S. Burnett, and Y. Guo, "Decoupling feedforward and feedback structures in hybrid active noise control systems for uncorrelated

narrowband disturbances," *Journal of Sound and Vibration*, vol. 350, pp. 1–10, 2015.

[56] J. Lorente, M. Ferrer, D. M. De, and A. González, "Gpu implementation of multichannel adaptive algorithms for local active noise control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 11, pp. 1624–1635, 2014.

[57] D. Li and M. Hodgson, "Optimal active noise control in large rooms using a "locally global" control strategy," *The Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3653–3661, 2005.

[58] J. Garcia-Bonito, S.J. Elliott, and C.C. Boucher, "Generation of zones of quiet using a virtual microphone arrangement," *The journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3498–3516, 1997.

[59] D. Halim, L. Cheng, and Z. Su, "Virtual sensors for active noise control in acoustic–structural coupled enclosures using structural sensing: robust virtual sensor design," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1390–1399, 2011.

[60] J. Zhang, S. J. Elliott, and J. Cheer, "Robust performance of virtual sensing methods for active noise control," *Mechanical Systems and Signal Processing*, vol. 152, pp. 107453, 2021.

[61] J. Cheer, S. J. Elliott, E. Oh, and J. Jeong, "Application of the remote microphone method to active noise control in a mobile phone," *The Journal of the Acoustical Society of America*, vol. 143, no. 4, pp. 2142–2151, 2018.

[62] J. Diaz, J.M. Egaña, and J. Vinolas, "A local active noise control system based on a virtual-microphone technique for railway sleeping vehicle applications," *Mechanical systems and signal processing*, vol. 20, no. 8, pp. 2259–2276, 2006.

[63] J. Garcia-Bonito, S.J. Elliott, and C. C. Boucher, "A virtual microphone arrangement in a practical active headrest," in *Inter-Noise 96 (Noise control: the next 25 years, Liverpool, 30 July-2 August 1996)*, 1996, pp. 1115–1120.

[64] D. Das, D. Moreau, and B. S. Cazzolato, "Nonlinear active noise control for headrest using virtual microphone control," *Control Engineering Practice*, vol. 21, no. 4, pp. 544–555, 2013.

[65] N. Miyazaki and Y. Kajikawa, "Head-mounted active noise control system with virtual sensing technique," *Journal of Sound and Vibration*, vol. 339, pp. 65–83, 2015.

[66] W. Jung, S. J. Elliott, and J. Cheer, "Combining the remote microphone technique with head-tracking for local active sound control," *The Journal of the Acoustical Society of America*, vol. 142, no. 1, pp. 298–307, 2017.

[67] D. Shi, B. Lam, and W. Gan, "Analysis of multichannel virtual sensing active noise control to overcome spatial correlation and causality constraints," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8499–8503.

[68] S. J. Elliott, W. Jung, and J. Cheer, "Causality and robustness in the remote sensing of acoustic pressure, with application to local active sound control," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8484–8488.

[69] J. Chambers, D. Bullock, Y. Kahana, A. Kots, and A. Palmer, "Developments in active noise control sound systems for magnetic resonance imaging," *Applied Acoustics*, vol. 68, no. 3, pp. 281–295, 2007.

[70] T. Xiao, X. Qiu, and B. Halkon, "Ultra-broadband local active noise control with remote acoustic sensing," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[71] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *In 124th Convention of the AES*. Citeseer, 2008.

[72] A. Kuntz and R. Rabenstein, "An approach to global noise control by wave field synthesis," in *2004 12th European Signal Processing Conference*. IEEE, 2004, pp. 1999–2002.

[73] P. Peretti, S. Cecchi, L. Palestini, and F. Piazza, "A novel approach to active noise control based on wave domain adaptive filtering," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 307–310.

[74] A. Lapini, F. Borchi, M. Carfagni, and F. Argenti, "Application of wave field synthesis to active control of highly non-stationary noise," *Applied Acoustics*, vol. 131, pp. 220–229, 2018.

[75] A. Lapini, M. Biagini, F. Borchi, M. Carfagni, and F. Argenti, "Active noise control for pulse signals by wave field synthesis," in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 883–887.

[76] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Spatial active noise control based on kernel interpolation with directional weighting," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8404–8408.

[77] J. Brunnström and S. Koyama, "Kernel-interpolation-based filtered-x least mean square for spatial active noise control in time domain," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 161–165.

[78] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on speech and audio processing*, vol. 13, no. 1, pp. 135–143, 2004.

[79] 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.

[80] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on speech and audio processing*, vol. 9, no. 6, pp. 697–707, 2001.

[81] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wave-field syn-

thesis," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2004, vol. 4, pp. iv–iv.

[82] D. S. Talagala, W. Zhang, and T. D. Abhayapala, "Efficient multi-channel adaptive room compensation for spatial soundfield reproduction using a modal decomposition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1522–1532, 2014.

[83] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Room reflections assisted spatial sound field reproduction," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 1352–1356.

[84] P. N. Samarasinghe, M. A. Poletti, S. Salehin, T. D. Abhayapala, and F. Fazi, "3d soundfield reproduction using higher order loudspeakers," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 306–310.

[85] M.A. Poletti, T.D. Abhayapala, and P. Samarasinghe, "Interior and exterior sound field control using two dimensional higher-order variable-directivity sources," *The Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 3814–3823, 2012.

[86] Y. Mitsufuji, N. Takamune, S. Koyama, and H. Saruwatari, "Multichannel blind source separation based on evanescent-region-aware non-negative tensor factorization in spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 607–617, 2020.

[87] S. J. Elliott, J. Cheer, J. W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2123–2133, 2012.

[88] Q. Zhu, X. Qiu, P. Coleman, and I. Burnett, "A comparison between two modal domain methods for personal audio reproduction," *The Journal of the Acoustical Society of America*, vol. 147, no. 1, pp. 161–173, 2020.

[89] B. Bu, C. Bao, and M. Jia, "Design of a planar first-order loudspeaker array for global active noise control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 2240–2250, 2018.

[90] J. Zhang, T. D. Abhayapala, W. Zhang, P. N. Samarasinghe, and S. Jiang, "Active noise control over space: A wave domain approach," *IEEE/ACM Transactions on audio, speech, and language processing*, vol. 26, no. 4, pp. 774–786, 2018.

[91] Y. Dong, J. Chen, and W. Zhang, "Distributed wave-domain active noise control based on the diffusion adaptation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2374–2385, 2020.

[92] Y. Maeno, Y. Mitsufuji, P. N. Samarasinghe, N. Murata, and T. D Abhayapala, "Spherical-harmonic-domain feedforward active noise control using sparse decomposition of reference signals from distributed sensor arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 656–670, 2019.

[93] Y. Dong, J. Chen, and W. Zhang, "Wave-domain active noise control over distributed networks of multi-channel nodes," *Signal Processing*, vol. 184, pp. 108050, 2021.

[94] L. Zhang, X. Wang, R. Hu, D. Li, and W. Tu, "Estimation of spherical harmonic coefficients in sound field recording using feed-forward neural networks," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 6187–6202, 2021.

[95] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 8, pp. 2038–2050, 2010.

[96] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1149–1159, 2010.

[97] W. Zhang and T. D. Abhayapala, "Three dimensional sound field reproduction using multiple circular loudspeaker arrays: functional analysis guided approach," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 7, pp. 1184–1194, 2014.

[98] H. Zuo, P. N. Samarasinghe, and T. D. Abhayapala, "Intensity based spatial soundfield reproduction using an irregular loudspeaker array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1356–1369, 2020.

[99] Y. Maeno, Y. Takida, N. Murata, and Y. Mitsufuji, "Array-geometry-aware spatial active noise control based on direction-of-arrival weighting," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8414–8418.

[100] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 135–139, 2017.

[101] 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, 2009.

[102] 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.

[103] N. Hahn, F. Winter, and S. Spors, "Synthesis of a spatially band-limited plane wave in the time-domain using wave field synthesis," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 673–677.

[104] F. Winter, N. Hahn, and S. Spors, "Time-domain realisation of model-based rendering for 2.5 d local wave field synthesis using spatial bandwidth-limitation," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 688–692.

[105] M. Poletti, T. D. Abhayapala, and P. D. Teal, "Time domain description of spatial modes of 2d and 3d free-space greens functions," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*. Audio Engineering Society, 2016.

[106] N. Hahn, F. Schultz, and S. Spors, "Time domain sampling of the radial functions in spherical harmonics expansions," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4502–4512, 2021.

[107] S. Yan, H. Sun, X. Ma, U. P. Svensson, and C. Hou, "Time-domain implementation of broadband beamformer in spherical harmonics domain," *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 5, pp. 1221–1230, 2010.

[108] J. Li, D. Dault, and B. Shanker, "A quasianalytical time domain solution for scattering from a homogeneous sphere," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1676–1685, 2014.

[109] P.A. Martin, "Acoustic scattering by a sphere in the time domain," *Wave Motion*, vol. 67, pp. 68–80, 2016.

[110] F. Bowman, *Introduction to Bessel functions*, Courier Corporation, 2012.

[111] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Academic press, 2014.

[112] H. Chang, T. K. Sarkar, and O. M. Pereira-Filho, "Antenna pattern synthesis utilizing spherical bessel functions," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 6, pp. 853–859, 2000.

[113] P.A. Martin, *Multiple scattering: interaction of time-harmonic waves with N obstacles*, Number 107. Cambridge University Press, 2006.

[114] G.N. Watson, *A treatise on the theory of Bessel functions*, p. 361, Cambridge University Press, London, UK, 2 edition, 1995.

[115] F. J. Simons, "Slepian functions and their use in signal estimation and spectral analysis," *arXiv preprint arXiv:0909.5368*, 2009.

[116] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on*, 2002, vol. 2, pp. 1781–1784.

[117] 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, Aug. 2012.

[118] J. Zhang, T. D. Abhayapala, P. N. Samarasinghe, W. Zhang, and S. Jiang, "Multichannel active noise control for spatially sparse noise fields," *The Journal of the Acoustical Society of America*, vol. 140, no. 6, pp. 510–516, 2016.

[119] 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 (TASLP)*, vol. 25, no. 2, pp. 359–373, 2017.

[120] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Feedforward spatial active noise control based on kernel interpolation of sound field," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 511–515.

[121] B. Rafaely and S. J. Elliott, "An adaptive and robust feedback controller for active control of sound and vibration," *Proc. UKACC International Conference on Control*, pp. 1149–1153, 1996.

[122] J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, vol. 2, pp. 789–792.

[123] B. Rafaely, "Analysis and design of spherical microphone arrays," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 1, pp. 135–143, 2005.

[124] T. Okamoto, "3d localized sound zone generation with a planar omnidirectional loudspeaker array," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019, pp. 110–114.

[125] W. Zhang and T. D. Abhayapala, "2.5 d sound field reproduction in higher order ambisonics," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2014, pp. 342–346.

[126] Y. Maeno, Y. Mitsufuji, P. N. Samarasinghe, and T. D. Abhayapala, "Mode-domain spatial active noise control using multiple circular arrays," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 441–445.

[127] 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.

[128] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1149–1159, 2011.

[129] M. Olsen and M. Møller, "Sound zones: Scattering study with head and torso simulator," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control-Engineering and Perception*. Audio Engineering Society, 2013.

[130] J. Chang, J. Park, and Y. Kim, "Scattering effect on the sound focused personal audio system," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3060–3066, 2009.

[131] T. Betlehem and M. A. Poletti, "Sound field reproduction around a scatterer in reverberation," in *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 89–92.

[132] W. Zhang, C. Hofmann, M. Buerger, T. D. Abhayapala, and W. Kellermann, "Spatial noise-field control with online secondary path modeling: A

wave-domain approach," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 12, pp. 2355–2370, 2018.

[133] P. N. Samarasinghe, T. D. Abhayapala, Y. Lu, H. Chen, and G. Dickins, "Spherical harmonics based generalized image source method for simulating room acoustics," *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1381–1391, 2018.

[134] J. Zhang, T. D. Abhayapala, W. Zhang, and P. N. Samarasinghe, "Active noise control over space: A subspace method for performance analysis," *Applied Sciences*, vol. 9, no. 6, pp. 1250, 2019.

[135] "Neumann—current microphones, dummy head ku-100 description," https://en-de.neumann.com/ku-100, 2013, Accessed: 2019-12-19.

[136] R. Xie, C. Shi, and H. Li, "Virtual sensing technique for a multi-reference and multi-error active noise control system," in *Proceedings of the 23rd International Congress on Acoustics (2019)*, pp. 6891–6897.

[137] A. Roure and A. Albarrazin, "The remote microphone technique for active noise control," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering, 1999, vol. 1999, pp. 1233–1244.

[138] R. Maeda and Y. Kajikawa, "Comparisons of two virtual sensing methods for broadband noise," in *Proceedings of the 23rd International Congress on Acoustics (2019)*, pp. 6883–6890.

[139] E. Friot, A. Roure, and M. Winninger, "A simplified remote microphone technique for active noise control at virtual error sensors," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering, 2001, vol. 2001, pp. 414–417.

[140] Q. Zhu, X. Qiu, and I. Burnett, "An acoustic modelling based remote error sensing approach for quiet zone generation in a noisy environment," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8424–8428.

[141] B. Rafaely, "The Spherical-Shell microphone array," *IEEE Trans. Audio Speech Lang. Processing*, vol. 16, no. 4, pp. 740–747, May 2008.

[142] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 647–658, 2014.

[143] MH Acoustics, "Em32 eigenmike microphone array release notes (v17.0)," *25 Summit Ave, Summit, NJ 07901, USA*, 2013.

[144] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1636–1645, 2007.

[145] J. Zhang, T. D. Abhayapala, P. N. Samarasinghe, W. Zhang, and S. Jiang, "Sparse complex fxlms for active noise cancellation over spatial regions," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 524–528.

[146] S. Spors and H. Buchner, "An approach to massive multichannel broadband feedforward active noise control using wave-domain adaptive filtering," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2007, pp. 171–174.

[147] H. Chen, P. Samarasinghe, T. D. Abhayapala, and W. Zhang, "Spatial noise cancellation inside cars: Performance analysis and experimental results," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2015, pp. 1–5.

[148] H. H. H. Homeier and E. O. Steinborn, "Some properties of the coupling coefficients of real spherical harmonics and their relation to gaunt coefficients," *Journal of Molecular Structure: THEOCHEM*, vol. 368, pp. 31–37, Sep. 1996.

[149] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55, Courier Corporation, 1965.

[150] N. J. A. Sloane, R. H. Hardin, W. D. Smith, et al., "Tables of spherical codes," *See http://www. research. att. com/njas/packings*, 2000.

[151] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1979.

[152] F. Zhang, Z. Geng, and W. Yuan, "The algorithm of interpolating windowed fft for harmonic analysis of electric power system," *IEEE transactions on power delivery*, vol. 16, no. 2, pp. 160–164, 2001.

[153] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3094–3103, 2005.

[154] H. Zuo, P. N. Samarasinghe, T. D. Abhayapala, and G. Dickins, "Spatial sound intensity vectors in spherical harmonic domain," *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. EL149–EL155, 2019.

[155] N. Epain and C. T. Jin, "Spherical harmonic signal covariance and sound field diffuseness," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 24, no. 10, pp. 1796–1807, 2016.

[156] A. Koretz and B. Rafaely, "Dolph–chebyshev beampattern design for spherical arrays," *IEEE transactions on Signal processing*, vol. 57, no. 6, pp. 2417–2420, 2009.

[157] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 1980–1991, 2002.

[158] S. Argentieri, P. Danes, and P. Soueres, "Modal analysis based beamforming for nearfield or farfield speaker localization in robotics," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 866–871.

[159] L. Birnie, T. D. Abhayapala, H. Chen, and P. N. Samarasinghe, "Sound source localization in a reverberant room using harmonic based music," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 651–655.

[160] L. Cassina, L. Fredianelli, I. Menichini, C. Chiari, and G. Licitra, "Audio-visual preferences and tranquillity ratings in urban areas," *Environments*, vol. 5, no. 1, pp. 1, 2018.

[161] R. Baumgartner, H. Pomberger, and M. Frank, "Practical implementation of radial filters for ambisonic recordings," *Proc. of ICSA, Detmold*, 2011.

[162] M. Sébastien, D. Jérôme, and B. Stéphanie, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, 2006, pp. 20–23.

[163] F. Zotter, "A linear-phase filter-bank approach to process rigid spherical microphone array recordings," *Proc. IcETRAN*, 2018.

[164] H. Sun, T. D. Abhayapala, and P. N Samarasinghe, "Time domain spherical harmonic analysis for adaptive noise cancellation over a spatial region," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 516–520.

[165] H. Sun, T. D. Abhayapala, and P. N Samarasinghe, "Active noise control over 3d space with multiple circular arrays," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019, pp. 135–139.

[166] S. Bilbao, J. Ahrens, and B. Hamilton, "Incorporating source directivity in wave-based virtual acoustics: Time-domain models and fitting to measured data," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2692–2703, 2019.

[167] A. Farina, A. Capra, L. Chiesi, and L. Scopece, "A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and

in post production," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. Audio Engineering Society, 2010.

[168] E. Mabande, A. Schad, and W. Kellermann, "A time-domain implementation of data-independent robust broadband beamformers with lowfilter order," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*. IEEE, 2011, pp. 81–85.

[169] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 11, pp. 1869–1878, 2015.

[170] G. Long, F. Ling, and J. G. Proakis, "The lms algorithm with delayed coefficient adaptation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 9, pp. 1397–1405, 1989.

[171] G. Long, F. Ling, and J. G. Proakis, "Corrections to'the lms algorithm with delayed coefficient adaptation'," *IEEE Transactions on Signal Processing*, vol. 40, no. 1, pp. 230–232, 1992.

[172] S. Lösler and F. Zotter, "Comprehensive radial filter design for practical higher-order ambisonic recording," *Fortschritte der Akustik, DAGA*, pp. 452–455, 2015.

[173] 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, 2013.

[174] A. Politis and H. Gamper, "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2017, pp. 224–228.

[175] 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.

[176] G. Chardon, W. Kreuzer, and M. Noisternig, "Design of spatial microphone arrays for sound field interpolation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 780–790, 2015.

[177] N. Hahn and S. Spors, "Time-domain representations of a plane wave with spatial band-limitation in the spherical harmonics domain," in *Proc. 45th German Annu. Conf. Acoust.(DAGA)*, 2019, pp. 1434–1439.

[178] W. Ren, H. Chen, and W. Gao, "On the design of time-domain implementation structure for steerable spherical modal beamformers with arbitrary beampatterns," *Applied Acoustics*, vol. 122, pp. 146–151, 2017.

[179] A. V. Oppenheim, *Discrete-time signal processing*, Pearson Education India, 1999.

[180] C. Langrenne, E. Bavu, and A. Garcia, "A linear phase iir filterbank for the radial filters of ambisonic recordings," in *EAA Spatial Audio Signal Processing Symposium*, 2019, pp. 127–132.

[181] W. Hahne, V. Erbes, and S. Spors, "On the perceptually acceptable noise level in binaural room impulse responses," in *Proc. 45th German Annu. Conf. Acoust.*, 2019, pp. 1–4.

[182] J. Tan, J. Cheer, and Daley S., "Wave-based control for nonreciprocal acoustics using a planar array of secondary sources," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering, 2021, vol. 263, pp. 5446–5456.