ON THE BEAMFORMING DESIGN

HE QI

PhD

The Hong Kong Polytechnic University

2023

The Hong Kong Polytechnic University
Department of Applied Mathematics

On the Beamforming Design

HE QI

A thesis submitted in partial fulfilment o f t he r equirements for
the degree of Doctor of Philosophy
August 2022

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

HE Qi (Name of student)

Dedication

Dedicate to my parents.

Abstract

Beamforming is an essential technique in speech enhancement, which has been applied to many applications, including wireless communication, hands-free communication, and speech recognition. This thesis mainly focuses on the beamformer design and sensor array localization problem via optimization techniques in the distributed network.

Firstly, we consider the microphone array localization problem with time-difference-of-arrival (TDOA) measurements. We formulated the problem which applied the known source locations to identify the wireless array configuration and estimate the location for each array. The proposed method formulates a mixed semidefinite programming (SDP) and second-order cone programming (SOCP) relaxation model. Then the acoustic geometry calibration is obtained by solving a convex optimal programming. The characteristics of the optimal solution are studied, and exact relaxation conditions are given. In addition, two methods are proposed to give an offset of the random internal error caused by the recording procedure. Experimental results demonstrate the proposed mixed model in 2-dimensional and 3-dimensional space, which outperforms other relaxation methods.

Then, with the given locations, a near-field broadband beamformer based on IIR filters performing spatial and frequency filtering is designed. The coefficients of the beamformer can be found from an optimal minimax problem which minimizes the error between the desired response and the actual response. We proposed a decom-

position method to solve the stability problem, and two optimization algorithms are considered to obtain a global solution. Since the design problem is very complex and highly nonlinear due to the need for stability constraints, a specific structure is proposed to simplify the problem, Furthermore, the performance limit of the general and specific structure is analyzed. Results show that significantly fewer coefficients are needed than for FIR filter designs, and the corresponding computational load in the implementation decreases.

Finally, two adaptive beamformers are designed in the modulation domain, bypassing the need of spatial information. The beamformers are designed based on the least square (LS) error between the desired and estimated signals and the maximum signal-to-noise ratio (SNR). The proposed methods have been evaluated by three indicators, including STOI, noise suppression, and signal distortion. The results show that the beamformers designed in the modulation domain outperform the counterparts developed in the frequency domain.

List of Publications

Here lists the publication during my PhD study period:

- He, Q., Low, S.Y. and Yiu, K.F.C. An Optimized Fixed Equalizer for Speech Enhancement. Circuits, Systems, and Signal Processing 41, 5743–5764 (2022). https://doi.org/10.1007/s00034-022-02051-1

Submitted

- He, Qi, Zhi Guo Feng, Zhibao Li, Ka Fai Cedric Yiu, and Sven Nordholm. Design and performance limit of near-field broadband IIR beamformers. IEEE Transactions on Signal Processing (Under review)

- He Qi, Mingjie Gao, Ka Fai Cedric You, and Sven Nordholm "On the SDP-SOCP method for solving microphone array localization problem. " IEEE/ACM Transactions on Audio Speech and Language Processing

To be submitted

- He, Qi, Siow Yong Low, and Ka Fai Cedric Yiu. "Modulation-domain multi-channel filtering." Speech Communication

Acknowledgements

Achieving the Ph.D. degree is a non-isolated process, and I am grateful to receive much support from my supervisor, partners, friends, and my family. I want to acknowledge them hereby.

First of all, I would like to express my most profound appreciation to my supervisor, Prof. Yiu Ka-Fai, Cedric, for his guidance and inspiration during the whole procedure of my M. Phil and Ph.D. study. Thank you for providing this study opportunity at PolyU and generously providing knowledge and expertise. Throughout the past few years, I have learned a lot, which will influence me greatly in my future work and life.

I am grateful to Prof. Sven Nordholm, Prof. Zhiguo Feng, and Prof. Siow Yong Low, whose insights and suggestions on the subject steered me through my study period. I have benefited a lot from their valuable comments. Thanks for your time and patience. I would also like to acknowledge Prof. Zhibao Li and Dr. Mingjie. Gao for their support of my work.

Additionally, I would like to thank The Hong Kong Polytechnic University for the financial support, without which I couldn't finish this work. Thanks should also go to the staff in the AMA for your help.

I am thankful for the accompanying of my friends in PolyU, Dr. Changyu Liu, and Ms. Jing Li. I want to thank my friend Ms. Zhongyi Su for her reassurance and help when I suffered in life.

I wish to thank my darling, Mrs. Xin Tong. Although you are not by my side, you give me the most emotional support.

I would also like to thank my boyfriend, Dr. Renqiang Zhu, for your accompanying, love and support in recording and plotting figures. It is also my most extraordinary luck to meet you during this journey.

Finally, I would like to express my deepest gratitude to my parents, Mr. Ruifeng HE and Mrs. Hongyan ZHANG, for your all-enduring and selfless love! I will love you all now and forever.

Contents

Dedication iv

Abstract v

List of Publications vii

Acknowledgements viii

List of Figures xiii

List of Tables xv

1 Introduction 1

1.1 Background 1

1.2 Literature Review 9

1.2.1 Microphone Array Localization Problem 9

1.2.2 Fixed Beamformer Design Problem 16

1.2.3 Adaptive Beamformer Design Problem 19

2 Distributed Microphone Array Localization Problem via SDP-SOCP Method 31

2.1 Introduction 31

2.2 Fundamentals of Localization Problem 32

2.2.1 Distance Measurements 33

2.2.2 Estimation of TDOA 35

2.3 Problem Statement 39

2.3.1 Direct Problem . . . 39
2.3.2 Inverse Problem . . . 40
2.4 Relaxation Models . . . 42
2.4.1 SDP Relaxation Model . . . 43
2.4.2 SOCP Relaxation Model . . . 45
2.4.3 SDP-SOCP Relaxation Model . . . 46
2.5 Analysis of the Mixed SDP-SOCP Relaxation . . . 49
2.6 Offset of TDOA with Real Data . . . 58
2.6.1 Method 1 . . . 59
2.6.2 Method 2 . . . 60
2.7 Experimental Results . . . 61
2.7.1 Numerical Results . . . 61
2.7.2 Experiments in simulated rooms . . . 64
2.7.3 Experiments with Real Data . . . 64
3 Design of Near-Field Broadband Beamformer Based on IIR Filters 71
3.1 Introduction . . . 71
3.2 Problem Statement . . . 72
3.3 Stability Condition . . . 75
3.3.1 A Root of Multiplicity 1 or 2 . . . 76
3.3.2 General Roots . . . 77
3.4 Optimization Algorithms . . . 82
3.5 Special Structure . . . 85
3.6 Performance Limit Analysis . . . 90
3.6.1 Performance Limit of a Special Structure . . . 91
3.6.2 Performance Limit of a General Structure . . . 98

3.7 Simulation Results . . . 100
3.7.1 Experiments with a free-field model . . . 101
3.7.2 Experiments with room simulation . . . 105

4 Design of Modulation-Domain Based Beamformers 110

4.1 Introduction . . . 110
4.2 Problem Formulation . . . 112
4.2.1 Signal Model in Frequency Domain . . . 112
4.2.2 Signal Model in Modulation Domain . . . 114
4.3 Beamformer Design . . . 115
4.3.1 Beamformer with Least Squares Criterion . . . 116
4.3.2 Beamformer with SNR Criterion . . . 118
4.3.3 Hybrid Method . . . 121
4.4 Performance Measurement Indicators . . . 122
4.4.1 Signal Distortion and Noise Suppression . . . 123
4.4.2 A Short-Time Objective Intelligibility Measure(STOI) . . . 124
4.5 Experimental Results . . . 125
4.5.1 Effect of Different Number of Frequency Bin and Modulation Spectrum Length . . . 126
4.5.2 Effect of Different Noise Levels . . . 128
4.5.3 Results of Hybrid Design Method . . . 131
4.5.4 Results of STOI values . . . 132

5 Conclusions and Suggestions for Future Research 135

5.1 Conclusions . . . 135
5.2 Suggestions for Future Research . . . 137

Bibliography 139

List of Figures

1.1 Noisy environment 2
1.2 The structure of the beamformer 3
1.3 Three fundamental sensor arrangements 10
1.4 The four different measurements 12
1.5 Sound waves in the near field and far field 17
1.6 The framework of centralized algorithms 19
1.7 The framework of distributed algorithms 21
1.8 Data model of the network 22
2.1 Two different measurements used in localization problem 36
2.2 Geometric setup of direct problem 41
2.3 Geometric setup of inverse problem 43
2.4 The result of Example a 56
2.5 The result of Example b 58
2.6 Comparison with different methods 63
2.7 The results of room simulation 65
2.8 The recording setup 66
2.9 The configurations of real data 67
2.10 The results with real data C1 68
2.11 The results with real data C2 69
2.12 The results with real data C3 70

3.1 Beamforming structure applied IIR filters 74
3.2 Beamforming structure using common feedback 87
3.3 The structure of the 3 feedback sections 102
3.4 Cost function value for the first example 103
3.5 The magnitude of the actual response of the first example in free-field 104
3.6 Magnitude of the actual response and desired response of example 1 in free field 104
3.7 Cost function value for the second example 106
3.8 Magnitude of the actual response and desired response of example 2 in free field 107
3.9 Magnitude of the actual response and desired response of example 1 in simulated room 108
3.10 Magnitude of the actual response and desired response of example 2 in simulated room 109
4.1 The framework of the adaptive beamformer 113
4.2 The framework of the modulation transform 116
4.3 The framework of the hybrid system 122
4.4 The configuration of the experiments 126
4.5 Clean speech, noisy speech and denoised speech signal in time domain(Fb=512,MB=4) 129
4.6 Spectrogram analysis of clean speech, noisy speech and denoised speech signal (Fb=512,MB=4) 130
4.7 Noise suppression and signal distortion of different noise levels 131
4.8 Trade off between the noise suppression and speech distortion 133
4.9 STOI values of different noise levels 134

List of Tables

2.1 Errors of different methods 63

2.2 Errors of different configuration 66

3.1 Cost function value (dB) with different number of IIR 102

3.2 Cost function value (dB) in simulated room 106

4.1 Effect of the number of FFT in frequency domain 128

4.2 Effect of the number of FFT in modulation domain 129

4.3 Effect of the SNR (Fb=512,Mb=4) 131

4.4 The results of hybrid method 132

4.5 STOI values with different SNR (Fb=512,Mb=4) 133

Chapter 1

Introduction

1.1 Background

Beamforming is an essential technique in speech signal processing, which has been applied in many areas such as wireless communication, hands-free communication, speech recognition, voice control devices, and hearing aids [19]. For example, consumers would like to speak commands to their devices via systems of automatic speech recognition and natural language processing like Siri, Google Now, Alexa, and Cortana [171]. Popular consumer products like Amazon Echo and Google Home are drawing much attention from significant manufacturers and product developers. However, in the real-world situation, the received speech signals can be polluted by a lot of noises, such as babble noise, white noise, traffic noise, or reverberation, as depicted in Fig. 1.1. As a result, systems for voice control suffer from severe performance degradation. A critical technique to enhance the received signal and reduce noise is to apply beamforming over a set of sensors, which analyzes the spatial properties of received signals to improve desired source locations over others. A general structure of the beamformer is given in Fig. 1.2. Compared with a single-point microphone observation of a speech signal, multichannel observations can obtain more information about the noise and desired signals, including both the spatial domain and time domain information. This technique applies a linear filter-and-sum opera-

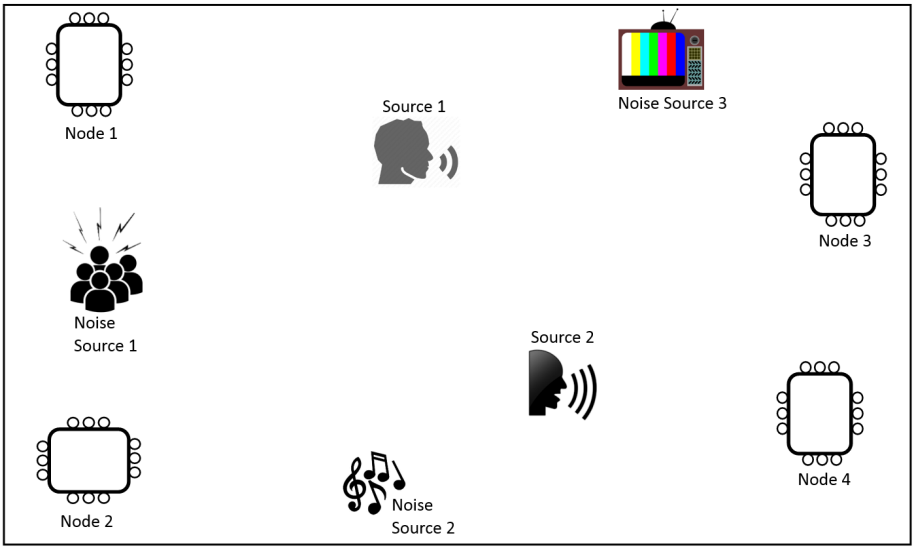


Figure 1.1: Noisy environment

tion on the received signals, where the filters are designed based on specific criteria.

The conventional beamfomering techniques are always applied in a compact microphone array. However, with the development of micro-electro-mechanical systems (MEMS) over the last few decades, small smart acoustic sensors have appeared. They are equipped with powerful embedded processors and can achieve multi-task, like data processing, communication, and sensing, with low cost and power. Following the miniaturization of sensor technologies, the use of portable devices increases significantly, and many are networked via various protocols. This allows information to flow between sensors and be controlled remotely by users. At the same time, deploying many microphones is gaining popularity for devices with acoustic capabilities, such as smartphones and portable computers. Being distributed in the environment, if properly employed, can form a robust acoustic sensor network to give another dimension in carrying out various speech processing tasks [7]. This kind of network is referred to as wireless sensor networks (WSNs) [139, 129], have been widely used in many areas, with applications ranging from locating and tracking sound sources and enhancing speech signals [26, 152, 163, 6, 111]. It is an emerging scheme comprising

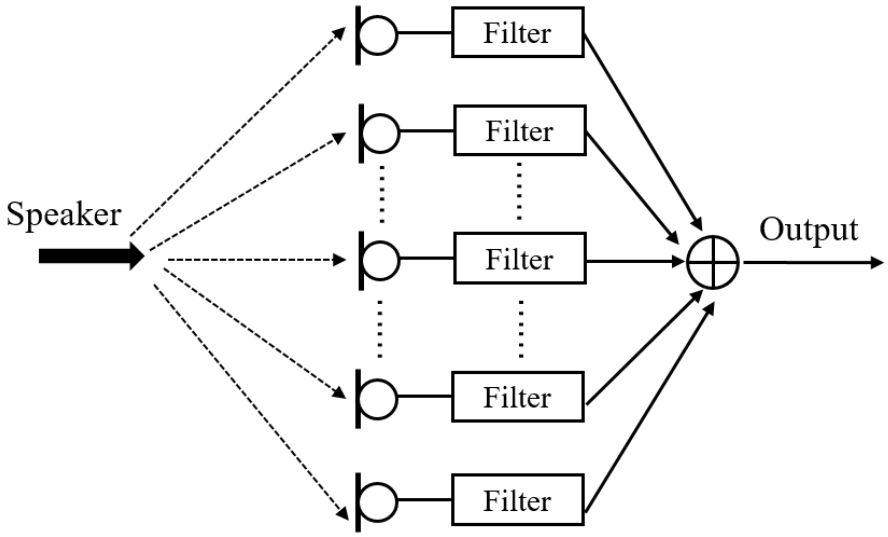


Figure 1.2: The structure of the beamformer

many wireless sensor nodes; each node consists of a single microphone or a single microphone array. Compared with the conventional compact microphone arrays, the array configuration of WSNs is no longer limited to standard structure, which can provide comprehensive spatial coverage. The microphones can even locate meters away from one another, enabling the various acoustic scenes to be obtained, resulting in a more accurate estimation of the users' environment. Some problems should be addressed to better apply the distributed signal processing techniques; the one main challenge is the unknown array structure.

To allow the formation of microphone networks by the sensors finding themselves in the same enclosure, the location of the microphones and the array configuration are required for many applications. Research shows that beamformers are sensitive to systematic errors because of the differences in the design conditions. One uncertainty is in the transfer function from the source to the sensors, which is related to the source locations and the sensor array configuration [82]. Errors in the sensor coordinates can affect the transfer function, which will further have a severe negative effect on the performance of the beamformer [179, 86]. The regular microphone array structure is

often deployed for the ease of manufacture, which results in the linear array being the predominant way to arrange microphones. This imposes many restrictions on the beamforming algorithms in localizing sound sources and enhancing speeches. Besides, with advanced wireless technology, microphone sensors can be distributed arbitrarily, and regular structure is unnecessary in the wireless world. As microphone locations are critical to speech enhancement and source localization accuracy, a thorough study of microphone locations and arrangements is necessary.

As reviewed in [116, 112], much recent research on acoustic geometry calibration is to determine the locations of the microphones from the received speech signals. This is to explore the spatial differentiation hidden in the signals. Typical indicators extracted are the time difference of an arrived signal at two sensor nodes or the direction under which the signal is detected. The obtained information is applied in a cost function measuring the discrepancy between the predicted indicators by the assumed geometry and the actual measured quantities. Since the extracted location information often relies on measuring time or time difference, accurate time synchronization is essential. To compound the difficulty, for a distributed system, the sampling process in each node is closely related to its local clock oscillator and hardware structure. Inevitably there will be a delay in the sampling start, and there is a sampling rate offset between the signals received by different nodes, severely affecting localization and speech enhancement performance. The system should compensate for the offset in the calculations to avoid this performance reduction.

Assuming that the array configuration is estimated, beamforming algorithms can be performed over the network. There are various beamformers, which can be categorized as fixed and adaptive beamformers. As its name indicated, the parameters of fixed beamformers are fixed during the process; and the filter is independent of the microphones' signal data. In essence, the filters can be considered as multidimensional filters, also known as broadband beamformers, which can extract the signal of

a particular beam-width and bandwidth while restraining the signals that are not in the desired space or frequency. Fixed beamforming techniques enhance noisy signals by calculating and summing the delay. We can divide the fixed beamformers into near-field and far-field beamformers in terms of the distance between the source and microphone. In some traditional applications, such as sonar, antennas, and radar, the source is assumed to be located at an infinite distance from the array, receiving plane waves [76, 167, 132]. However, the applications we focus on in this thesis, such as hands-free mobile in cars, and voice-commanded systems, typically have short distances between sources and arrays. Traditional far-field algorithms always perform poorly in this near-field situation, so designing algorithms suitable for near-field applications is necessary. There exists many near-field broadband beamforming algorithms, such as [101, 77, 69, 25, 42, 100, 172, 102, 83, 170, 43]. The main goal of a beamformer is to enhance the audio signal from the parts of interest and reduce the undesired elements. One general method is constructing the near-field beamformer design problem as a minimax program in a quadratic form of the weighted Chebyshev approximation problem [101].

Another type of beamformer is the adaptive beamformer. The parameters of adaptive beamformers are changeable during the operation, and the beamformers depend on static characteristics of desired, noisy speech signals received by the microphone. Typical adaptive beamformers include minimum variance distortionless response (MVDR), linear constraint minimum variance (LCMV) and generalized side lobe canceller (GSC). They always formulate the design problem as an optimization problem by different criteria, and the parameters can be obtained by solving the problem. The classic adaptive beamformers always require spatial cues. As mentioned above, the spatial cues are hard to estimate due to the dynamic geometries. Thus, some blind algorithms, such as generalized eigenvalue decomposition and max signal to noise ratio method, are needed, which can bypass the need for spacial

information. Another problem that should be considered is that the classic centralized beamformer needs a central processing unit to calculate the data accepted by all the sensors, which is undesirable in actual applications. Therefore, distributed algorithms are necessary to distribute the processing load over different nodes, as each node only contains partial data with limited energy supplies. In addition, the distributed algorithms should have the same outputs as the centralized methods.

Above all, the formation of a microphone array network is still an active research area with many possibilities. In the current state, there are still many problems to be overcome. This thesis mainly focuses on designing acoustic beamformers via optimization techniques and auditory sensor network localization. A convex optimization techniques-based microphone array localization algorithm is proposed first. Then, with the given locations, a fixed near-field broadband beamformer based on IIR filters is designed, where significantly fewer coefficients are needed than FIR-based algorithms. It is satisfying in WSNs, due to the battery and computation load limitations. Finally, adaptive beamformers are designed in the modulation domain, which has no requirements for spatial information and performs better than the frequency-based counterpart algorithms in terms of four widely used indicators. The structure of the thesis is sketched below.

- Chapter 2 considers the microphone array localization problem in a distributed acoustic network based on a relaxation model with time-difference-of-arrival (TDOA) measurements. In multimedia applications, it is common to employ acoustic sensors collectively to enhance signals and to locate sound sources. A direct problem can be formulated to locate sound sources from a set of known sensors. In order to form the acoustic sensor network, it is important to locate the sensor array locations first. However, unlike other networks in which direct TOA measurements might be possible, acoustic distributed network can

only obtain time-difference-of-arrival (TDOA) measures indirectly from various sound source anchors. While it is common to employ convex optimization techniques to localize sensor locations in a network with TOA information, it has not been studied properly when it comes to TDOAs. This chapter considers the microphone array localization problem in a distributed acoustic network with time-difference-of-arrival (TDOA) measurements. We formulate the inverse problem which applied the known source locations to identify the wireless array configuration and estimate the location for each array. The proposed method formulates a mixed semidefinite programming (SDP) and second-order cone programming (SOCP) relaxation model. Then the acoustic geometry calibration is obtained by solving convex optimal programming. The characteristics of the optimal solution are studied, and exact relaxation conditions are given. Furthermore, offset algorithms are proposed to decrease the random error raised from the recording procedure. Experimental results demonstrate the proposed mixed model in 2-dimensional and 3-dimensional space, which outperforms other relaxation methods. The efficiency of the offset algorithms has been proved in real applications.

- Chapter 3 considers the design of a near-field broadband beamformer based on an IIR filter performing spatial and frequency filtering. Many near-field broadband beamforming models are achieved by an FIR filter attached to each channel. This chapter investigates the characteristics of the IIR-based beamformer. By using IIR filters in the design, significantly fewer coefficients are needed than for FIR filter designs, and the corresponding computational load in the implementation decreases. To solve the stability problem better, we further decompose the feedback part into a sum of low-order sections and add stability constraints in the optimal problem resulting in a stable filter struc-

ture. Two global optimization algorithms are given to solve the non-convex optimization problem. However, due to the need for stability constraints in the optimization, the full IIR filter problem is very complex and highly nonlinear. A specific structure is proposed in which all the elements in the beamformer share the same feedback section to simplify the problem. Furthermore, we study the performance limit of the proposed method, where the filter length can be chosen arbitrarily, and the performance limit can be obtained efficiently by solving a series of subproblems. We prove that the specific structure have the same limit performance as the general structure. Numerical experiments showed that the optimal value of the IIR design method could approach the limit much faster than FIR-based beamformers. The proposed method was evaluated utilizing a room simulation model for varying reverberation times. The design degrades consistently with increasing reverberation time.

- Chapter 4 considers the design of beamformers in the modulation domain without the requirements of spatial cues. The concept of modulation domain is developed from Short Time Fourier transform (STFT), which analyzes the modification synthesis framework. In essence, modulation domain processing is the evolution of the signal's temporal and spectral information. These modulation domain characteristics promote the application of the modulation domain processing in speech enhancement. This chapter extends two popular beamformers, including least square (LS) and signal-to-noise ratio (SNR), into the modulation domain. The proposed methods have been evaluated by four indicators, including noise suppression, signal distortion, and STOI. The results show that the beamformers designed in the modulation domain outperform the counterpart methods developed in the frequency domain.

- Chapter 5 summarises the whole thesis and gives insights into future work.

1.2 Literature Review

1.2.1 Microphone Array Localization Problem

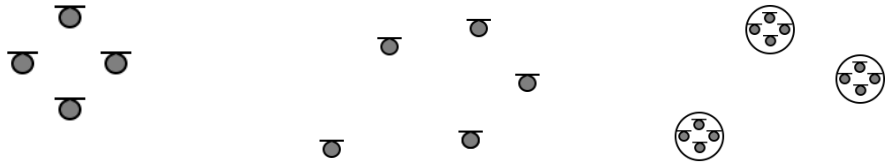
This part reviews the most recent algorithms for microphone array localization problems, also called position self-calibration or microphone self-localization. Microphone localization algorithms have been proposed for various scenarios. In general, there are three kinds of basic microphone arrangements for different application scenarios: single compact array, distributed individual sensors, and distributed arrays. The last two are for wireless acoustic sensor networks. Follows characterizes the three arrangements in detail, and Figure 1.3 depicts the three scenarios.

1. The first scenario considers the arrangement containing only one single microphone array and all the microphones located in the array of small geometric dimensions. In application, we can always assume that all the microphones have the same time base. In this situation, the acoustic location calibration problem is also known as array shape calibration. The configuration characteristic is that all the microphones are close to each other. Therefore, there is some acoustic coherence among the signals received by the sensors. Many algorithms are developed based on the coherence information.

2. In the second scenario, the microphones are not in a compact array while they are distributed in the network, and each node only contains one microphone. The calibration task is to determine the locations of all the microphones distributed in the whole room, called microphone configuration calibration. Due to the distribution of sensors, we cannot generally know the time synchronization among nodes.

3. The difference between the second and third scenarios is the number of the microphone in each node. In the second scenario, each node only consists of

one microphone, while in the third scenario, each node contains a compact microphone array (more than one microphone). This acoustic geometry calibration task is known as array configuration calibration. In this task, the inner configuration of the microphone array is known, and the microphones in the same array always share the same time base. Thus, we can calculate the microphone locations with the knowledge of the center microphone's position and the sensor nodes' orientation. Then, the sensor localization problem is simplified, as we only need to estimate the center positions of arrays and the orientations of each node.



(a) single compact array (b) distributed sensors (c) distributed sensor arrays

Figure 1.3: Three fundamental sensor arrangements

According to different application situations, acoustic geometry calibration algorithms are developed. The general idea of acoustic geometry calibration algorithm is to extract geometric arrangement-related information from the received signals [84], such as the time of arrival between sources and sensors, the time difference of arrival between two different microphones [138], and the direction of arrival [106]. Then, one objective function can be formulated with the obtained geometric arrangement information and the measurements as predicted. In general, four kinds of acoustic measurements are categorized by the geometric arrangement of the sensors in a connection to other microphones or an active speaker. The four kinds of measurements include pairwise distance (PD), the direction of arrival (DOA), time of arrival

(TOA), and time difference of arrival (TDOA), which are depicted in Fig. 1.4.

(a) Pairwise distance (PD)

PD measurements measure the distance between a particular pair of sensors in an acoustic network. It can be obtained by calculating the noise coherence between the signals received by the two microphones.

(b) Time of arrival (TOA)

This measurement describes the distance between the source and the sensor, which is generally obtained by receiving the sources with known positions. It means the TOA measurement requires the emission time of the source signal. This measurement is also known as the time of flight, measuring source to sensor distance in a direct propagation.

(c) Time difference of arrival (TDOA)

TDOA measures the difference between the time delay of two different sensors. Compared with TOA, TDOA requires no emission time of source signals.

(d) Direction of arrival (DOA)

Microphone localization algorithms based on DOA only for the cases where each node contains a sensor array rather than a single sensor. A unit norm vector measures the direction in which the source reached the node.

The following part reviews the recent literature on microphone localization, which is ordered based on the four geometric measurements above.

Pairwise Distance (Noise Coherence)

The pair-wise distance can be obtained by calculating the noise coherence, the normalized cross-power spectrum of two received signals. The assumption of diffuse

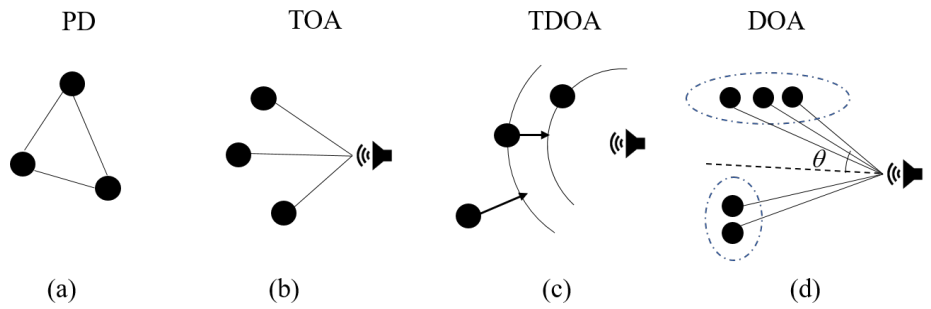


Figure 1.4: The four different measurements

noise field limits microphones with small distances, so this kind of method only works for a compact array. The fundamental insight of this kind of method is to match the theoretical diffuse noise coherence function to the measurements [99]. Then, we can formulate an optimization problem by minimizing the difference between the theoretical distance and the measurements.

We can solve the optimization problem in a closed form by applying multidimensional scaling [15], which can obtain the spatial arrangement of the sensor. The general idea of this method is that the sensor configuration can be found by eigenvalue decomposition from the productive scalar matrix, and for more details in [36]. Instead of directly fitting the diffuse noise coherence, a model for the generalized cross-correlation with phase transform (GCC-PHAT) was proposed in [155], which achieves minor estimation error compared with the previous methods. Another problem with the PD measurements is that the measurements matrix has a higher dimension than its rank. A low rank matrix completion strategy is applied in [142] and a nonnegative matrix factorization method is used in [3] to solve the problem.

TOA

Methods based on TOAs focus on the arrival time of an active source. If a point source propagates in a direct path, the TOA is related to the distance between the sensor and the speaker. Multiply a given speed of sound in the air on the TOAs.

Then, we can get the distances from the sensors to the source. In essence, the estimation of TOA can enable localization since it indicates the measurement of distances between the sensors and targets, which induces a high accuracy. With the known positions of the sound sources, the configuration of the sensor network can be estimated by triangulation. If all the nodes are in a coplanar scenario, at least three base nodes and three TOA measurements are needed to find the location of one target node. If a known base node position and a TOA r are given, the potential target location will be on a circle with a radius r. If there are two known base nodes, there would be two TOA circles, and the potential target would be the two points of intersection of the TOA circles. One more node is needed to find out the exact position. The final target location is the point that goes through the third TOA circle. In a non-coplanar scenario, an additional one node is required, that is, four base nodes, to localize one target due to the increasing spatial dimension.

The TOA is given by the pairwise distance between the source and sensor positions, the onset time of the source, and the internal recording delay. The calculation of onset time requires the same time base between the sensor and sound, while the internal delay requires the signal at the source. The first requirement means the source and sensor should be synchronization, and the second requirement assumes the source is a loudspeaker. If the offset and delay are known, we can derive the sensor positions by minimizing the error between the pairwise distance and the estimated measurements. [16], where the estimation of TOA can be calculated by cross-correlating the sensor signal. Some relevant methods are proposed in [29, 30, 125].

However, if the knowledge of internal recording delay and an onset delay is unknown, some more complex estimation algorithms should be considered. A two-stage method is proposed in [50], which estimates the timing information first and then solves the positions. These two steps exploit the low-rank structure of the location matrices $\boldsymbol{S}$ and $\boldsymbol{A}$. Then, the problem becomes to estimate matrices with much lower

dimensions, which can be considered as a nonlinear least square problem. Once timing parameters are solved, the second step can be solved by methods proposed in [30].

TDOA

TDOA measurements refer to the time difference between two sensors to the source when there is a direct path from the source to the sensor. The arrival time difference can be presented in a hyperbola, a locus of a point in a plane. Then, the distance difference between two fixed anchors is a constant. Three base nodes and two TDOA measures are needed to determine one unknown target node in a coplanar scenario. Suppose that the first base node that received the signal emitted from the target node is the reference node. Then the potential target location is on a hyperbola formulated by the TDOA with one base node and the reference node. To find out the exact position, we need another TDOA information, and the final place is the point of the intersection of two hyperbolas. While in a non-coplanar situation, one additional base node is required, that is, four base nodes and three TDOA measures. In this situation, there isn't an onset time of the source, which means this measurement does not need the time synchronization between the microphone and source.

Assuming that the system is time-synchronized or the equivalent knowledge of delays is known, we can derive the TDOAs directly according to the estimation coordinates. The steered response power with phase transform is always used for this method. It is the same as a delay and sum beamformer when finding the position with a maximal output power [19]. A sensor localization algorithm in reverberant environments is proposed in [65] to perform a local coordinate mapping based on coherence analysis. In addition, in a situation where each node contains a sensor array, acoustic images can be obtained by the delay and sum beamformer in each

array [151]. This makes positions can be extracted from a camera model in Cartesian coordinates. If the sources are located far away form the sensors, which means the sensors are in a far field of the speaker. Then, the microphone localization problem can be simplified [148].

If the system is unsynchronized, the recording delays are unavailable. More complex joint estimation algorithms should be considered. The method in [107] formulate the localization problem as the nonlinear least square problem, which can be solved by the algorithms based on auxiliary function. Besides, research showed that the alternating optimization has better convergence properties than gradient descent. A two-stage method is given in [157] by using a rank approximation method. Thrun [148] first proposed the approximation rank-based algorithms with TDOA measurements and known onset time and delays. Wang [157] extended this method to unsynchronized systems, where the time delays should be estimated. Another insight of TDOA is that the TDOA can be applied to calculate pairwise distance [114]. The advantage of this is that the time synchronization between sensors is unnecessary.

DOA

DOA-based methods can only be applied where each node contains a microphone array rather than a single sensor. For DOA measurement, a cosine distance measure is better than a Euclidean distance measure, as we focus on directions. The optimization problem is formulated by comparing the direction of the source, as measured by the sensor.

A Newton algorithm is applied to solve the optimization problem in which only a synchronization among the same node is required [73]. The advantage of this method is that we can eliminate the synchronization problem between different nodes and the synchronization between sensors and sources. However, there is a drawback: methods based on DOA can not estimate the scale of the acoustic network. This

drawback can be overcome by applying TDOA measurements to the localization problem. The method in [115] showed that a known inter-array arrangement of a circular array could also solve the scale uncertainty problem.

1.2.2 Fixed Beamformer Design Problem

Fixed beamforming techniques enhance noisy signals by calculating and summing the delay. The fixed beamformer design problem is related to a multidimensional digital filter design problem with an arbitrarily specified amplitude and phase. We can set the desired frequency and place as the pass region, while the undesired elements are the stop region. After the beamformer, the desired parts' audio signal remains, and the undesired components are reduced. The pass and stop regions can be discretized into a finite number of grid points. Then, we can formulate a minimax problem with a quadratic formulation of the weighted Chebyshev approximation problem to derive the parameters of the beamformer, which can be solved by the linear programming technique [103].

The fixed beamformers contain far-field and near-field beamformers according to the distance between the source and microphone array. If the microphones are far enough from the source, we can assume the wavefronts are planar and will decay at a rate of 6dB per distance doubled from the start. The signals received by the sensors can be equally considered attenuated, and the source can be viewed as a point source. Most transducers are characterized and operate in the far-field as behavior is consistent across a range of frequencies required by specific imaging applications in sonar, non-destructive testing, or biomedical industries. Suppose the microphone is near the source. In that case, the model becomes a complex constructive and destructive interference pattern as the waves are generated from an aperture of a set geometric size. The sound waves in near field and far field are depicted in Fig. 1.5.

Many works have been done to design the beamformers. The traditional research

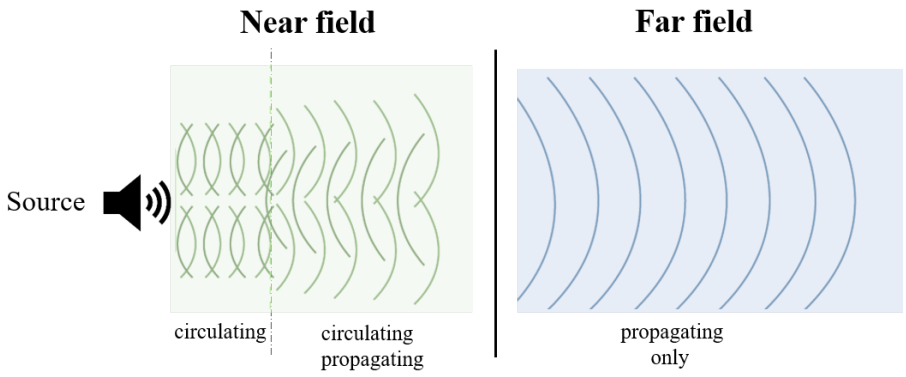


Figure 1.5: Sound waves in the near field and far field

mainly focus on the one-dimensional or two-dimensional FIR filters, which are designed by the minimax metod and applied on a single grid with the requirement of the linear phase [71, 137, 60]. Then, many general algorithms have been proposed for multidimensional filters design with arbitrary phases and amplitudes. A far-field beamformer design algorithm has been proposed in [44], and they analyzed the influence of the performances concerning the filter length and the number of microphones. Ward et al. proposed a frequency variant fixed far-field beamformer in [160], which is suitable for distributed arrays with any bandwidths and has no requirement of the beam shape. Mars et al. [97] further improved the frequency variant beamformer with lower signal distortion and less computation load compared with the adaptive ones.

There also exist many nearfield broadband beamforming algorithms. Rodney et al. [78] proposed the reciprocity relationship between the nearfield and far-field beam patterns and designed a nearfield beamforming array using the relationship. The method in [77] transformed the wanted nearfield model to an equivalent radial pattern by applying the spherical harmonic solution to the wave equation. The desired nearfield beamformer can be obtained by using the far-field design methods. Ryan et al. [124] proposed a method using a signal propagation vector to describe a

point source. This method can be applied when the desired source is in the nearfield region, but the noises are far away from the sensors.

Another design method is constructing the near-field beamformer design problem as a minimax program with a quadratic formulation of the weighted Chebyshev approximation problem [101]. This approach is related to a multidimensional digital filter design problem with an arbitrarily specified amplitude and phase. It performs on a discrete domain where frequency and spatial domains are discretized into a finite number of grid points. Then, it can be solved by the linear programming technique. Many conventional near-field broadband beamforming models are achieved by an FIR filter attached to each channel [69, 43]. The FIR filter is used to generate a frequency-dependent magnitude and phase shift over the array operating bandwidth. The wider the operating bandwidth is, the larger the number of taps required to obtain a given level of broadband interference injection [98, 123]. The method in [83] introduced an auxiliary function to solve the continuous space design problem. The auxiliary function's first root gives the solution, which can be found by a root-catching method. To improve the optimization processing, the researchers propose a penalty function method to reduce or get rid of the constraints [102]. In [170], they use l_1-norm and actual rotation theorem to reduce the nonlinearity in optimization, and the design problem becomes a semi-infinite linear programming problem. To decrease memory usage and computational complexity, a two-stage algorithm is given in [43]. It is vitally important in high dimension problems, as the number of discrete points needed is significantly large, resulting in a large-scale optimization problem. The FIR filter is used for all the methods above to formulate a frequency-dependent magnitude and phase shift over the network operating bandwidth. The wider the operating bandwidth is, the larger the number of taps required to obtain a given level of broadband interference injection [98, 123].

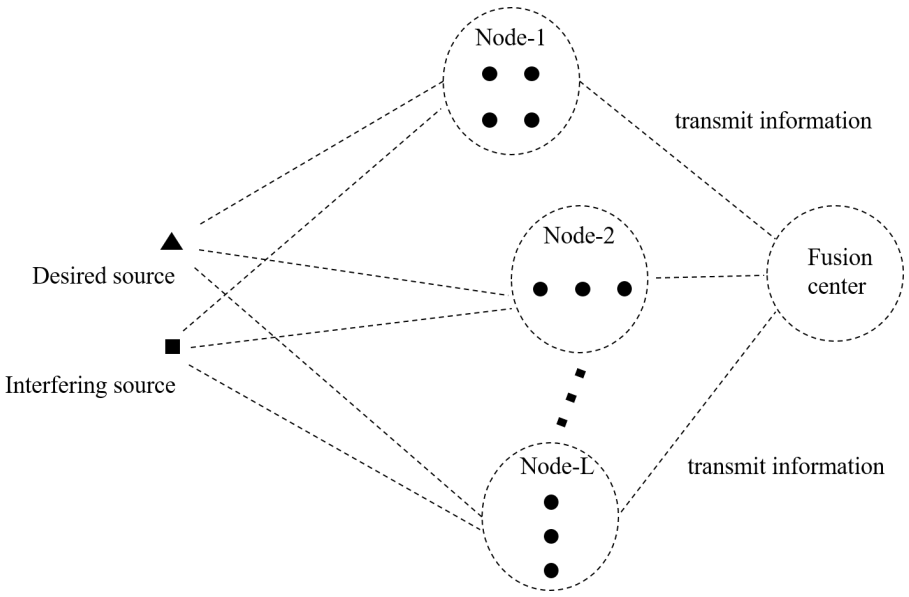


Figure 1.6: The framework of centralized algorithms

1.2.3 Adaptive Beamformer Design Problem

In this section, the most recent theoretical and practical literature on adaptive beamforming techniques has been reviewed. Compared with the fixed beamformer, such adaptive algorithms can adapt to the changeable signals. It is more robust and suitable for a changeable situation. The common idea of all methods is estimating the target signal, and the design problem is finding the beamformer's proper parameters. Many beamforming techniques have been studied, and the main difference among them is their different criteria. For example, multi-channel Wiener filter (MWF) beamforming [35, 9, 80] optimizes the mean square error between the desired and output signals. Minimum variance distortionless response (MVDR) beamformer [88, 51] aims to minimize the system's output power with only one linear constraint on the sensor's response to the target signal. In contrast, linear constrained minimum variance (LCMV) [1] is an extension of MVDR, which have constraints on both desired and noisy signal. Another method is the maximization of signal-to-noise-ratio (MAX-SNR) [147], which aims to optimize the SNR criterion.

However, there is a problem that many beamformers require the spatial information or equivalent knowledge, which could be hard to obtain in a wireless sensor network. This problem also exists in the fixed beamformer design problem. Some methods assume that the transfer functions between microphones and sources are known [68]. Some methods assume that spatial clues are not available and have to be estimated based on statistical properties of the microphone signals, such as subspace methods and the generalized eigenvalue decomposition method (GEVD).

Another problem is that, in the conventional methods, the communication of wireless microphones is committed by a central processor known as a fusion center used to collect all data for further processing. A centralized algorithm framework is given in Fig. 1.6. This model is applied in conventional centralized multi-channel noise reduction algorithms, but there is a drawback that the need for energy or transmission bandwidth is overmuch. It is unsuitable in WANs due to each array's limited battery life, communication, and computational load. An alternative solution known as distributed arrays is always applied in the WASNs algorithms, which can decrease the cost of communication and spread the processing burden to different nodes. A framework of the distributed algorithms is given in Fig. 1.7.

The acoustic model is formulated first to give a better understanding of the algorithms. Suppose that there are L nodes in a acoustic sensor network, and each node $l \in \{1, \cdots, L\}$ contains m_l sensors. Then, the number of sensors in the whole network is $m = \sum_{l=1}^{L} m_l$. We consider that in the network there are n speakers, some interfering and some desired; and the desired and interfering parts are assumed to be uncorrelated. The γth source captured by the ith senor in the lth node at the time index t is modelled as

$$y_{li}^{\gamma}(t) = p_{li}^{\gamma} * s_{\gamma}(t) + v_{li}(t), \tag{1.1}$$

where p_{li}^{γ} is the propagation steering vector from the γth source position to the ith

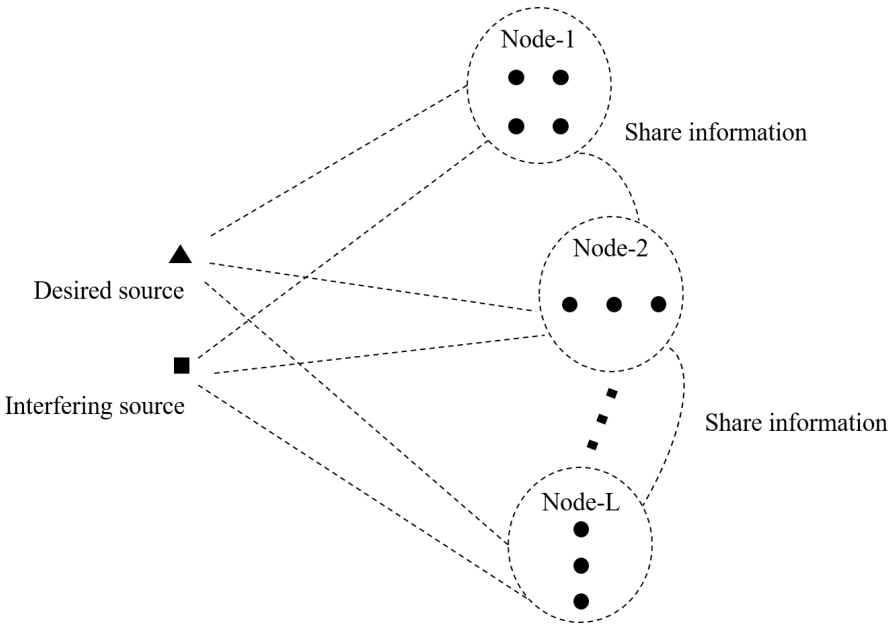


Figure 1.7: The framework of distributed algorithms

sensor in the lth array, s_γ is the clean signal at the γth source location, and v_{li} is the noise signal. Then, the received speech at the ith senor in the lth node is expressed as

$$y_{li}(t) = \sum_{\gamma=1}^{n} (p_{li}^{\gamma} * s_\gamma)(t) + v_{li}(t). \tag{1.2}$$

Denote $\mathbf{y}_l = [y_{l1}, \cdots, y_{lm_l}]$ as a stack of the received signals of the lth node; and $\mathbf{y} = [\mathbf{y}_1, \cdots, \mathbf{y}_L]$ as a stack of the all signals of the network. By using the short time discrete Fourier transform (STFT), we can window and transform the signals into frequency domain. As we assume that STFT coefficients are independent in time and frequency, we can omit indices for the brevity of notation. Let $\mathbf{Y}_l = [Y_{l1}, \cdots, Y_{lm_l}]^T$ be a stack of signals received by the lth node, where $(\cdot)^T$ means the matrix transposition, and let $\mathbf{Y} = [\mathbf{Y}_1, \cdots, \mathbf{Y}_L]^T$ be a stack of all signals received by all the nodes. Then, the $m \times 1$ vector $\mathbf{Y}$ can be described as

$$\mathbf{Y} = \mathbf{p}\boldsymbol{S} + \boldsymbol{V}, \tag{1.3}$$

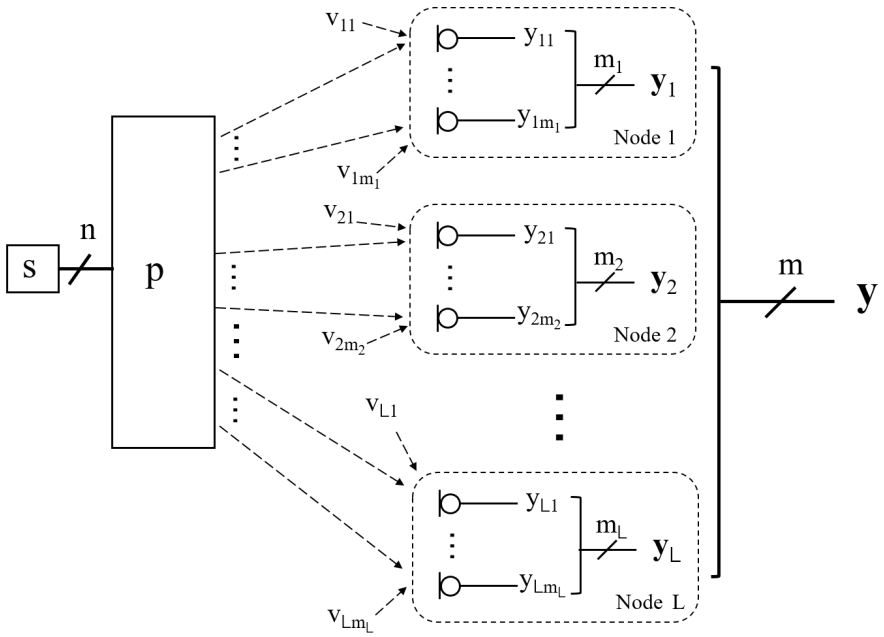


Figure 1.8: Data model of the network

where $\mathbf{p}$ is an $m \times n$ mixing matrix, containing the propagation steering vectors from the n source locations to the m microphones, $\boldsymbol{S}$ is an $n \times 1$ vector consisting source signals DFT coefficients, and $\boldsymbol{V}$ represents the noise. The auto-correlation matrix of the vector $\mathbf{Y}$ can be denoted as $\boldsymbol{R}_{YY} = E(\mathbf{Y}\mathbf{Y}^H)$, where $E(\cdot)$ is the statistical expectation operator. Similarly, the spectral covariance matrix of desire and noise signal are $\boldsymbol{R}_{XX}$ and $\boldsymbol{R}_{VV}$, respectively.

In the following, a review of algorithms for speech enhancement applying beamforming techniques is presented, which is ordered in terms of the criteria used. For each beamformer, the corresponding optimization problem will be formulated first, and then the solution set is given. Furthermore, for each kind of method, some distributed algorithms are given.

Minimum Mean Square Error (MMSE)

MMSE method is to minimize the mean square error between the desired signals and the estimated signals. For a node l, the estimation of the desired signal is given

by

$$\hat{\mathbf{x}}_l = \mathbf{w}_l^H \mathbf{y},$$

where $\mathbf{w}_l$ is a node specific estimator. According to the criterion of the MMSE method, we can formulate the design problem of the estimator as an square optimization problem as

$$\mathbf{w}_l = \arg\min_{\mathbf{w}_l} E\{|\mathbf{x}_l - \mathbf{w}_l^H \mathbf{y}|^2\}, \tag{1.4}$$

where $E\{\cdot\}$ is the expected value operator, and $\mathbf{x}_l$ represents a node specific desired signal. Suppose that the correlation matrix of the output signal $\boldsymbol{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\}$ has a full rank. When the signals received by different sensors are independent, the assumption could be satisfied. Then, the unique solution of (1.4) is given by [53]

$$\mathbf{w}_l = \boldsymbol{R}_{\mathbf{yy}}^{-1} \mathbf{r}_{\mathbf{x}}, \tag{1.5}$$

where $\mathbf{r}_{\mathbf{x}} = E\{\mathbf{y}\mathbf{x}_l^*\}$ and $(\cdot)^*$ represents the conjugate of the matrix, and $\mathbf{r}_x$ can be estimated by applying training sequences or voice activity detection mechanism.

In [35], an iterative multi-channel Wiener filter was introduced in a binaural hearing aid for the MMSE estimation by using a pruned version of the distributed adaptive node-specific signal estimation (DANSE) algorithm. It proves the distributed beamformer converges to the centralized solution in the situation of a rank-1 correlation matrix, i.e., with a single desired source. Then, the algorithms with different desired sources are proposed. Bertrand et al. [8] introduced a batch-mode DANSE algorithm to calculate an MMSE estimator for multiple desired signals. A robust DANSE algorithm was proposed in [9], and the efficiency and convergence of the method have been proven in a simulated room with multiple speakers. In [10], more details of the DANSE algorithm are given, including the convergence proof, a truly adaptive version, and the simulation results in a dynamic scenario.

The DANSE algorithm was further extended to greedy algorithm [140], and cooperative adaptive algorithm [61] for the MMSE estimation. While an optimal node

selection strategy is NP-hard, the greedy selection procedure reduces the computation cost by adding and iteratively removing nodes. The method in [61] considers node-specific DOA estimation. However, all the methods based on DANSE require a fully connected broadcasting sensor network or a tree topology. In [158], the authors proposed an MMSE beamformer based on a gossip algorithm, which has no requirements on the topology of the networks. In addition, blockchain technology is applied to protect data integrity during transmission and give more reliable connections between different nodes.

Minimum Variance Distortionless Response Beamforming (MVDR)

MVDR is an important technique in speech enhancement, which optimizes the power of the beamformer output with only one constraint on the response of the sensor to the target speech. Firstly, we consider the centralized situation. Assume that there exists only one desired signal. The filter coefficient vector $\mathbf{w}$ can be obtained by solving the following problem [19]

$$\begin{aligned} &\min_{\mathbf{w}} \ \mathbf{w}^H \boldsymbol{R}_{YY} \mathbf{w} \\ &s.t. \ \mathbf{w}^H \mathbf{p} = 1, \end{aligned} \tag{1.6}$$

where $\boldsymbol{R}_{YY}$ is the auto-correlation matrix. If we suppose that the speech signal is independent with the noise signal, then $E(\boldsymbol{X}\boldsymbol{V}^H) = E(\boldsymbol{V}\boldsymbol{X}^H) = 0$, and $\boldsymbol{R}_{\boldsymbol{YY}} = \boldsymbol{R}_{\boldsymbol{XX}} + \boldsymbol{R}_{\boldsymbol{VV}}$. Mathematically, the optimal problem in (1.6) is equivalent to minimize the cost function $\mathbf{w}^H \boldsymbol{R}_{VV} \mathbf{w}$. According to the matrix inversion lemma and Langrange multiplier method [20, 153], we can get the solution of (1.6) as

$$\mathbf{w}^*_{MVDR} = \frac{\boldsymbol{R}_{VV}^{-1} \mathbf{p}}{\mathbf{p}^H \boldsymbol{R}_{VV}^{-1} \mathbf{p}}. \tag{1.7}$$

The estimation of the desired speech is given by

$$\hat{S} = \mathbf{w}^H \mathbf{Y}. \tag{1.8}$$

The main challenge of the method is how to compute $\boldsymbol{R}_{VV}^{-1}$.

If we assume that the noise in network is spatially uncorrelated across microphones, which can be achieved in a diffuse noise field or/and that the distance between the arrays is quite large. With the assumption, the matrix $\boldsymbol{R}_{VV}$ is simplified by setting all the off-diagonal number to be 0, and the calculation of the matrix inversion is bypassed. In addition, the PSDs of each microphone can be different. Accordingly, the output of the beamformer is rewritten as

$$\hat{S} = \frac{\sum_{l=1}^{L} \sum_{i=1}^{m_l} p_{li} \sigma_{v_{li}}^{-2} Y_{li}}{\sum_{l=1}^{L} \sum_{i=1}^{m_l} p_{li} \sigma_{v_{li}}^{-2} p_{li}}, \tag{1.9}$$

where the power spectral density $\sigma_{v_{li}}^{-2}$ can be estimated by an ideal voice activity detector [63], and p_{li} can be estimated by the method in [47]. This kind of beamformer is also known as delay and sum beamformer (DSB).

The general idea of mixing the multi-channel signals in an optimization problem with the restriction on signal distortion was first proposed by Darlington [32]. Then, many works have been done to improve it. Frost introduced an adaptive scheme in [45] by using a constrained least mean-squares algorithm. However, this original method is quite sensitive to the errors. Thus, many algorithms have been proposed to achieve the robust beamforming, such as [22, 28], which applied diagonal loading techniques. Another type of robust method applied the steering vector, in which the output power is optimized according to the estimation of difference between the actual and presumed steering vectors [62]. To further balance the trade-off between noise reduction and signal distortion, a system in terms of AMNOR was proposed, which used a soft constraint [74]. A generalized sidelobe canceller (GSC) structure was proposed to eliminate the constrained adaptation [56], which is an extension of [45]. Then, the MVDR beamformer was further used to dereverberation, but there is a trade-off between noise reduction and dereverberation, which has been rigorously

analyzed in [58].

Another alternative model of the MVDR algorithm is the pseudo-coherence based signal model, witch is quite flexible and suitable for changeable geometry. Tavakoli et al. [145] proposed a blind speech enhancement method, which is modeled by the pseudo coherence vector, without the requirements of the sensor position information exploiting an orthogonal decomposition of the target signal. Furthermore, a primal and dual method of multipliers is used to tackle the distributed optimization problem formulated in the pseudo-coherence based model [146]. A frame work of speech enhancement methods based on pseudo-coherence vectors and matrices is given in [144].

Yuan et al. [173, 174] proposed a distributed DSB via asynchronous randomized gossip algorithm. The goal of the distributed algorithm is to calculate the $\tilde{Y}_{ave} = 1/m \sum_{l=1}^{L} \sum_{i=1}^{m_l} \tilde{Y}_{li}$ and $\tilde{p}_{ave} = \sum_{l=1}^{L} \sum_{i=1}^{m_l} \tilde{p}_{li}$ in a distributed manner. In each iteration, a random pair of nodes is active; and the pair of nodes exchange and update the values. Comparing with the full MVDR beamformer, although the calculation load of the DSBs decreases, the performance is also degraded. To balance the calculation load and the performance, a trade-off parameter is given in [67]; and a message-passing algorithm is applied to exchange information among different nodes. The inversion of the matrix is calculated by formulating a quadratic optimization problem, which is solved by the generalized linear-coordinate descent message passing algorithm [177, 176]. However, this algorithm suppose that the noise covariance matrix is diagonally dominant, that means the non-neighboring nodes are needed to be uncorrelated.

Matt et al. [104] proposed a diffusion-based distributed MVDR beamformer, that can approach to a full MVDR centralized beamformer with no requirements on the structure of the noise covariance matrix nor the prior knowledge of the covariance matrix. As the diffusion adaptation can adapt to changing data in real

applications[90, 127], this method is particularly suitable for dynamic environments. However, there is a drawback that each node eventually ends up with a vector consisting all weight value of the beamformer, limiting the true distributed nature of the algorithm, especially in the extremely large scale networks. In addition, all the methods above need a global averaging in each time slot to give a beamformer output. To preserve privacy, Yuan et al. [175] provided a distributed beamformer in which each array can estimate its own target source without sharing the steering vector with the other elements in the network. In [175], the Sherman-Morrison formula [133] is applied to enable the estimation of the inverse correlation matrix to be a consensus problem, which can be solved by a gossip algorithm [18]. However, this method is either not suitable for large scale networks, as it requires each node contains a large vector of covariances with all other nodes. In [105], Matt et al. further proposed a beamformer operating in a fully distributed and asynchronous manner and applying sparsity implementation by an l_1 penalty of the weight vector. Instead of focusing on the entire network, this method aims to formulate optimization problems with a subset of nodes in beamformer calculation.

Linearly Constrained Minimum Variance (LCMV)

An LCMV beamformer can be considered an extension of MVDR, which imposes constraints on both desired and noisy signals. The beamformer parameters $\mathbf{w}$ design minimizes the noise power while protecting the target responses. If all the nodes can transmit information to a fusion center for further processing, i.e., the centralized LCMV beamformer, the design problem can be formulated as

$$\begin{aligned} \min_{\mathbf{w}} \ & \mathbf{w}^H \boldsymbol{R}_{VV} \mathbf{w} \\ s.t. \ & \mathbf{p}^H \mathbf{w} = \mathbf{s}, \end{aligned} \tag{1.10}$$

where $\mathbf{s}$ represents the desired responses of the speech signals. We can solve this problem by Lagrange multiplier and Karush-Kuhn-Tucker (KKT) conditions, and

the closed form solution of problem (1.10) is given by [154]

$$\mathbf{w}_{LCMV}^{*} = \boldsymbol{R}_{VV}^{-1}\mathbf{p}^{T}(\mathbf{p}\boldsymbol{R}_{VV}^{-1}\mathbf{p}^{T})^{-1}\mathbf{s}. \tag{1.11}$$

Note that this closed form solution needs the knowledge of acoustic transfer functions, which can be estimated by the methods in [93, 52, 143].

An adaptive scheme of LCMV was first proposed in [40] by adding additional constraints of MVDR beamformers. One drawback of the method in [40] is that the beamformer is sensitive to the error of the covariance matrix and the steering vector. Then, many works have been done to improve the traditional LCMV beamformer. One method to tackle the problem is adding a phase constraint to the target output response, which makes the magnitude response and phase response less distorted than the conventional LCMV [164]. Thus, the proposed method is robust against the covariance matrix error and steering vector mismatch. To solve the problem, some other techniques have also been considered, such as convex quadratic constraints [121], Bayesian approach [5], and diagonal loading technique [23]. In addition, research proves that the generalized sidelobe canceler (GSC) form can also be applied in this multiple constraints case [21].

Distributed LCMV algorithms have been proposed to decrease the data exchanged among arrays, which avoids the computation of the network-wide covariance matrix. As summarized in [94], the generic formulation of distributed algorithms always contains two parts. One is a compression matrix, which fuses the multiple local signals into a signal with fewer channels and then broadcasts to the other nodes. The other is a local beamformer, which is designed by using the local signal and the compressed information from other nodes to give the desired output signal. In this way, each node can continuously adapt its compression matrix and the parameters of the local beamformer to the changes in the compression matrix at the other nodes. The main differences among the algorithms are the strategies of information chances.

The beamformer algorithm proposed in [13] fused the multiple signals of the same node into a single-channel signal and considered the case in which all nodes share the same constraints set. A DANSE algorithm is applied on the LCMV beamformer, which allows specific constraints at each node of the network [12]. It considers both the steering vector known case and the blind beamforming case similar to [93, 11], and these two versions' convergence and optimality have been proved. However, these two methods need a fully connected or tree-shaped network. Sherson et al. [134] further proposed a robust LCMV by adding a regularisation term for both cyclic and acyclic implementations and then cast the Beamformer design problem into a distributed convex optimization problem. Then, we can solve the problem by PDMM and ADMM algorithms. Compared with [13, 12], this method avoids multiple updates to calculate the optimal BF response, decreasing data transmitted during computation. It has no limitation on the topology of the network. The distributed beamformer in [81] reduces the cross power spectral density matrix to a block-diagonal form with some linear equality constraints and arbitrary topology. Li. et al. [87] first introduced the augmented Lagrangian method to design a centralized Beamformer, and then a distributed beamformer was proposed by using ADMM. The convergence is proved without any other additional conditions. Distributed multiple constraints GSC algorithms for a fully connected sensor network are proposed in [95, 96], designed for a reverberant environment.

Max-SNR Beamformer

As its name indicated, a max-SNR beamformer aims to optimize the SNR criterion. Compared with the MVDR and LCMV, the max-SNR technique bypasses the need for any spatial information, such as the steering vector, which may be hard to estimate accurately in real applications. Suppose that there is only one desired signal. According to the objective of the beamformer, the optimization problem for

a centralized beamformer can be formulated as

$$\max_{\mathbf{w}} \frac{\mathbf{w}^H \boldsymbol{R}_{XX} \mathbf{w}}{\mathbf{w}^H \boldsymbol{R}_{VV} \mathbf{w}}. \tag{1.12}$$

However, problem (1.12) is not a standard concave maximization problem, which can be solved by a Lagrange multiplier, and the optimal solution is achieved by making the gradient of its Lagrangian zero. The optimal weights matrix is the generalized eigenvector, which can be solved by the generalized Schur algorithm.

A distributed max-SNR speech enhancement method using the PDMM algorithm is proposed in [147]. This method decouples the optimization problem into local elements, and then the distributed convex optimization problem is applied. Similar to [178], the optimization problem is systematically solved by PDMM.

Chapter 2

Distributed Microphone Array Localization Problem via SDP-SOCP Method

2.1 Introduction

Acoustic geometry calibration estimates the locations of the microphones from the received signals, which explores the spatial differentiation hidden in the signals. Typical indicators extracted are the time of arrived signal [16, 30], the time difference of an arrived signal at two nodes [117, 151], or the direction under which the signal is detected [73, 115]. As reviewed in the first chapter, the methods based on TOA measurements are always biased by the unknown source onset times and unknown device capture times [148]. Compared with the TOA, TDOA gains fewer stringent requirements on time synchronization between sensors and the source locations [157]. Then, the extracted information is used in a cost function measuring the discrepancy between the predicted indicators by the assumed geometry, and the actual measured quantities [70]. This kind of method usually leads to optimizing a non-convex cost function. Researchers proposed relaxation models transforming the non-convex function into convex to solve this problem. In [135], semidefinite programming (SDP) based relaxation model for the position estimation problem in sensor network local-

ization is analyzed, and it proved that networks could be localized. Furthermore, a second-order cone programming (SOCP) model is studied in [150] due to its simpler structure and faster calculation speed. However, these models used the DOA measurements.

In the proposed method, the TDOA between each pair of microphones is obtained and transformed into distance measurements between microphones. Then, we can get a series of nonlinear hyperbolic equations, and assess the location by solving the set of equations. However, there is a difficulty that the formulated problem is highly nonlinear and nonconvex. To solve this problem, we extensively employ a convex relaxation method to solve the group microphone localization problem and propose a novel mixed relaxation model. This approach has been studied extensively in [49] by exploring the properties of SDP and SOCP, which outperforms the other existing methods. To formulate a novel localization method based on convex relaxation, we have also studied the optimal solution's theoretical structure and characteristics. However, in real applications, the sensors in wireless sensor networks are inherently asynchronous, resulting in a temporal offset in each channel. Two offset compensation algorithms are proposed to increase accuracy in real data estimates. Then, in the experimental part, we test the three relaxation models, i.e., SDP, SOCP, and SDP-SOCP relaxation model, in numerical and room simulation methods. Results show that the SDP-SOCP outperforms the other two methods. In addition, examples with real data are given, and the two offset algorithms are applied. Results show the efficiency of the offset algorithms.

2.2 Fundamentals of Localization Problem

This section discusses the fundamentals of the localization problem, and prepare some knowledge for the proposed sensor array localization algorithm.

2.2.1 Distance Measurements

One of the most fundamental techniques in the positioning system is to extract measurements from the received signals. There are four distance measurements in general. The pairwise distance can only work for a compact array, and the direction of arrival can only be applied where each node contains a microphone array rather than a single sensor. In contrast, TOA and TDOA are suitable for all the application scenarios. This part further discusses the differences between these two measurements. The signal models based on these two measurements and their fundamental localization principles are presented below.

Time of Arrival (TOA)

TOA is the signal propagation time from one source to a receiver. If a point source propagates in a direct path, the TOA is related to the distance between the sensor and the speaker. Suppose that one sound is emitted at time t_1, and one set of sound waves was captured by on sensor by t_2. Then, the TOA between the sensor and the sound is $t_2 - t_1$. Let $\boldsymbol{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_m]$ be a stack of all the sensors' locations, where $\mathbf{a}_i,\ i = 1, \cdots, m$ is the location coordinates of the ith microphone. Suppose that the sources' locations are gathered in $\boldsymbol{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_n]$, where $\mathbf{s}_\gamma, \gamma = 1, \cdots, n$ is the position of the γth source. Then, the TOA of the γth source at the jth microphone is given by

$$t_{\gamma,i} = \frac{\|\mathbf{a}_i - \mathbf{s}_\gamma\|}{c} + o_\gamma - o_i, \tag{2.1}$$

where o_γ is an onset time of the source and o_i is an internal recording delay, c represents the speed of source, and $\|\cdot\|$ represents the l_2 norm.

The calculation of onset time requires the same time base between the sensor and sound, while the internal delay requires the signal at the source. If this knowledge

is known, the distance between the γth source at the jth sensor is

$$\hat{d}_{\gamma,i} = (\hat{t}_{\gamma,i} - o_\gamma + o_i) \cdot c, \tag{2.2}$$

where $\hat{t}_{\gamma,i}$ is the measured TOA, and $\hat{t}_{\gamma,i}$ can be calculted by the cross-correlating of the microphone signal. The sensors' postions can be estimated by

$$\min_{\boldsymbol{A}} \sum_{i=1}^{m} \sum_{\gamma=1}^{n} (\|\mathbf{a}_i - \mathbf{s}_\gamma\| - \hat{d}_{\gamma,i})^2. \tag{2.3}$$

A base point multidimensional scaling can be used to give the direct solution of problem (2.3).

Although TOA has high accuracy, this technique has some drawbacks. TOA measurement requires precise synchronization; even a tiny timing error can result in a significant error in calculating distance. Besides, TOA needs the signal emission time, and this additional time measure can result in another error.

Time Difference of Arrival (TDOA)

Whereas similar in name, TDOA measures the time difference between the signals arriving at a pair of sensors. The TDOA can be formulated in mathematics as follows. Assume that there is a signal emitted from the source k at the unknown time t_0. The ith sensor receives the signal at time t_i, while the jth sensor receives the signal at time t_j. A distinct TDOA is given as $\tau_{ij} = (t_i - t_0) - (t_j - t_0) = t_i - t_j$. There are total $C_m^2 = m(m-1)/2$ possible pairs, where m is the total number of sensors in the distributed network. However, if the error of TDOAs is absent, there are some redundant items among all the $C_m^2 = m(m-1)/2$ possible pairs. For example if there are 3 sensors, then $\tau_{23} = t_2 - t_3 = (t_1 - t_3) - (t_1 - t_2) = \tau_{13} - \tau_{12}$, which is redundant. Thus, we can reduce the $m(m-1)/2$ distinct TDOAs to $m-1$ non-redundant ones in a noise-free situation, decreasing calculation load without falling estimation accuracy. Without loss of generality, we can choose the first sensor in

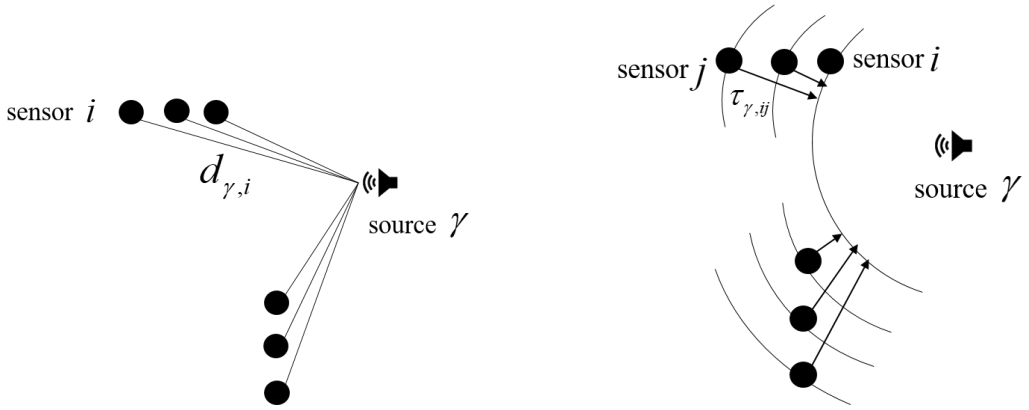
the network as the reference sensor, and the $m-1$ non-redundant TDOAs can be given as τ_{1i}, where $i = 2, \cdots, m$. In contrast, if there are some random errors in the measures, we can not reduce the TDOA measures.

The distance difference between the source and two sensors can be obtained by multiplying the speed of sound in the air. Measurements related to the coordinates can be given as

$$\begin{aligned} \tau_{\gamma,ij} &= t_{\gamma,i} - t_{\gamma,j} \\ &= \frac{\|\mathbf{a}_i - \mathbf{s}_\gamma\| - \|\mathbf{a}_j - \mathbf{s}_\gamma\|}{c} - o_i + o_j, \end{aligned} \tag{2.4}$$

where $-o_i + o_j$ is the time offsets of between two sensors, and c is the speed of sound.

Compared with the TOA measurement, this method overcomes one drawback of TOA. As in the TDOA measures, all microphones accept the same signals emitted by the sources; there is no need to synchronize the unknown anchors' clock with the base anchors, which means a lower hardware cost. Consequently, this method only requires the synchronization of base anchors. Due to this advantage, this chapter considers a localization algorithm based on TDOA measures.



(a) the TOA information (b) the TDOA information

Figure 2.1: Two different measurements used in localization problem

2.2.2 Estimation of TDOA

The TDOA measures can not be obtained directly in the real world, and they need to be estimated from the real data received by the sensors. This part gives a widely used estimation method, the Maximum likelihood generalized cross-correlation (GCC) method, which is considered the classical method for TDOA estimation from microphone pairs.

In this method, a maximum likelihood estimator can determine the time delay between the audio signals received by a pair of microphones. The maximum likelihood estimator can be considered two receiver prefilters followed by a cross-correlator. The delay estimate is obtained when the correlator reaches a maximum, where the time delay is usually considered a prominent peak. The goal of the prefilters is to guarantee that the signals before passing the correlator are at the highest SNR to decrease the noise power. This procedure can be formulated in mathematics as follows.

Firstly, the signals received by a pair of sensors are formulated as

$$\begin{aligned} y_1(t) &= x(t) + n_1(t) \\ y_2(t) &= x(t+d) + n_2(t), \end{aligned} \tag{2.5}$$

where $n_1(t)$ and $n_2(t)$ represents the noise, $x(t)$ is the pure signal emitted from a remote source and d is the delay between the pair of sensors, also known as the time difference of arrival (TDOA). In this model, we assume that the $x(t)$ is uncorrelated with $n_1(t)$ and $n_2(t)$. One general method estimating d is to calculate the cross correlation function between $y_1(t)$ and $y_2(t)$,

$$R_{y_1y_2}(\tau) = E[y_1(t)y_2(t-\tau)] \tag{2.6}$$

where E represents expectation. As the observation time is finite, we can only get an estimation of $R_{y_1y_2}(\tau)$ and the estimation in an observation interval T is presented

by

$$\hat{R}_{y_1y_2}(\tau) = \frac{1}{T-\tau}\int_{\tau}^{T} y_1(t)y_2(t-\tau)dt. \tag{2.7}$$

When $\hat{R}_{y_1y_2}(\tau)$ reaches a maximum, the value of τ is the delay estimation. However, $y_1(t)$ and $y_2(t)$ contain noise, decreasing estimation accuracy. A prefilter part is considered to solve this problem.

Because of the Fourier transform, we can have a function about $R_{y_1y_2}(\tau)$ and the cross power spectral density, formulated as

$$R_{y_1y_2}(\tau) = \int_{-\infty}^{+\infty} P_{y_1y_2}(f)e^{j2\pi f\tau}df. \tag{2.8}$$

The the cross power spectral density is

$$P_{y_1y_2}(f) = Y_1(f)Y_2^*(f) \tag{2.9}$$

where $*$ means the complex conjugate. The $Y_1(f)$ and $Y_2(f)$ are received signals after Fourier transform

$$\begin{aligned} Y_1(f) &= X(f) + N_1(f) \\ Y_2(f) &= X(f)e^{j2\pi fd} + N_2(f), \end{aligned} \tag{2.10}$$

where $X(f)$, $N_1(f)$ and $N_2(f)$ are $x(t)$, $n_1(t)$ and $n_2(t)$ in frequency domain. If there are prefilters before correlator, the filter outputs are given as

$$\begin{aligned} W_1(f) &= G_1(f)Y_1(f) \\ W_2(f) &= G_2(f)Y_2(f), \end{aligned}$$

where $H_1(f)$ and $H_2(f)$ are the transfer functions of filters. Then, the cross power spectrum between the outputs is

$$\begin{aligned} P_{w_1w_2}(f) &= W_1(f)W_2^*(f) \\ &= (G_1(f)Y_1(f))(G_2(f)Y_2(f))^* \\ &= G_1(f)G_2^*(f)P_{y_1y_2}(f). \end{aligned} \tag{2.11}$$

The generalized correlation between $y_1(t)$ and $y_2(t)$ is

$$R^{(g)}_{w_1w_2}(\tau) = \int_{-\infty}^{+\infty} \Omega_g(f) P_{y_1y_2}(f) e^{j2\pi f\tau} df. \tag{2.12}$$

where $\Omega_g(f) = G_1(f)G_2^*(f)$. Then, the TDOA estimation is solving the following optimization problem.

$$\tau^* = \arg\max_{\tau} = R^{(g)}_{w_1w_2}(\tau). \tag{2.13}$$

However, there is still a problem with how to choose the weighting $\Omega_g(f)$? Various weighting schemes of $\Omega_g(f)$ are proposed, such as Roth Processor, the smoothed coherence transform (SCOT), the phase transform(PHAT), and the Echart filter [162, 72, 161]. Research shows that GCC-PHAT is a suitable approach for TDOA estimation in real practice [57, 113]. In GCC-PHAT method, the weighting is given as

$$\Omega_p(f) = \frac{1}{|P_{y_1y_2}(f)|}, \tag{2.14}$$

and the generalized correlation becomes

$$R^{(p)}_{w_1w_2}(\tau) = \int_{-\infty}^{+\infty} \frac{P_{y_1y_2}(f)}{|P_{y_1y_2}(f)|} e^{j2\pi f\tau} df. \tag{2.15}$$

In an ideal noiseless situation,

$$|P_{y_1y_2}(f)| = \alpha P_{x_1x_2}(f),$$

resulting in

$$\frac{P_{y_1y_2}(f)}{|P_{y_1y_2}(f)|} = e^{j2\pi fd}. \tag{2.16}$$

Then, we have

$$R^{(p)}_{w_1w_2}(\tau) = \delta(t-d), \tag{2.17}$$

and the time delay estimation problem becomes

$$\tau^* = \arg\max_{\tau} = R^{(p)}_{w_1w_2}(\tau). \tag{2.18}$$

2.3 Problem Statement

We consider an acoustic network with a total of n sources and m sensors. For convenience, we denote the spatial coordinates of the γth source as $\mathbf{s}_\gamma = (s_{\gamma 1}, \cdots, s_{\gamma d})^T$, $\gamma = 1, \cdots, n$, where d is the dimension of the network. Accordingly, define the source location matrix as $\mathbf{S} = (\mathbf{s}_1, \cdots, \mathbf{s}_n)$. Consider that there are L sensor nodes distributed in the environment, each of which may contain a single microphone or a microphone array. Suppose that each node $l \in \{1, \cdots, L\}$ has m_l sensors, and position coordinates of the sensors in node l are denoted by a $d \times m_l$ matrix $\boldsymbol{A}_l = (\mathbf{a}_{l1}, \cdots, \mathbf{a}_{lm_l})$. In particular, if $L = 1$ and $m > 1$, it is a compact single array. If $L > 1$, and $\forall l$ $m_l = 1$, then the type of microphone arrangement is distributed individual sensors. Let $m = \sum_{l=1}^{L} m_l$, and define the sensor location matrix as $\mathbf{A} = (\boldsymbol{A}_1, \cdots, \boldsymbol{A}_L)$, where $\mathbf{A}$ is a $d \times m$ matrix. Acoustic localization algorithms can be characterized as a direct problem and inverse problem according to the estimation of the geometric arrangement of the source or sensor, as illustrated in Fig.2.2 and Fig.2.3.

2.3.1 Direct Problem

As given in Section 4.2.1, the TDOA can be formulated mathematically as follows. Assume that C is the set of all microphones. Suppose that there is a signal emitted from the source γ at the unknown time t_γ and the ith ($i \in C$) sensor receives the signal at time t_i, and another sensor j ($j \in C, j \neq i$) received the signal at time t_j. We can measure a distinct TDOA, denoted as $\tau_{\gamma,ij} = (t_i - t_\gamma) - (t_j - t_\gamma)$. There are a total of $m(m-1)/2$ possible pairs, where m is the total number of sensors in the distributed network.

In the direct problem, the measurements from unknown sound sources are first obtained and transformed into distance measures between sensors, formulating a

series of nonlinear hyperbolic equations. The estimation of TDOAs, denoted as $\tau_{\gamma,ij}$, can be calculated from the signals received by a pair of sensors, via the method given in Section 2.2.2. In addition, we can get a set of TDOAs correlating with the spacial coordinates as

$$\hat{\tau}_{\gamma,ij} = (\|\hat{\mathbf{s}}_\gamma - \mathbf{a}_i\| - \|\hat{\mathbf{s}}_\gamma - \mathbf{a}_j\|), \tag{2.19}$$

where $\mathbf{a}_i$ is the known microphone locations, and $\hat{\mathbf{s}}_\gamma$ is estimated positions of sources. Then, we can formulate an optimization problem

$$\min_{\hat{\mathbf{s}}_\gamma} = \sum |\tau_{\gamma,ij} - \hat{\tau}_{\gamma,ij}|^2. \tag{2.20}$$

By solving problem (2.20), we can obtained the estimated sensor locations. This is the formulation of the direct problem and the geometric figure is given in Fig. 2.2.

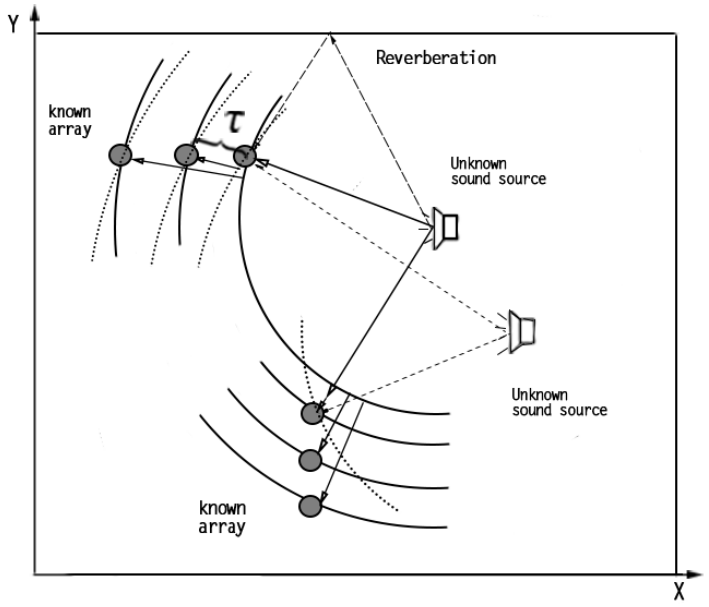


Figure 2.2: Geometric setup of direct problem

2.3.2 Inverse Problem

Inversely, if some of the sound source locations are known, we can use them to identify the wireless array configuration and estimate the location for each array. The

purpose of the inverse problem is to calibrate the realization of microphone locations such that the distance measurement equations are satisfied. In this problem, there are a total of $m(m-1)/2$ possible node pairs for TDOA information. Similarly, a set of $\hat{\tau}_{\gamma,ij}$ can been calculated by

$$\hat{\tau}_{\gamma,ij} = (\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| - \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_j\|), \quad \forall i,j \in C, i > j, \gamma = 1, \cdots, n \tag{2.21}$$

where $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{a}}_j$ are the microphone locations that we need to estimate, $\mathbf{s}_\gamma$ is the given source locations, and C is the set of all microphones. An estimation of TDOA can also be obtained by GCC method from the received signals, denoted by $\tau_{\gamma,ij}$. The purpose of the inverse problem is to find out the sensor locations by minimizing the error between $\tau_{\gamma,ij}$ and $\hat{\tau}_{\gamma,ij}$.

In application, the inner configuration of one microphone array is fixed. It is reasonable to assume the relative locations of microphones within an array are known. Thus, for each array, we have the distance between each pair of microphones in the same node. When solving the localization problem, the follows constraints should be satisfied

$$d_{ij}^2 = \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2, \ \forall i,j \in C_l, i > j, \ l = 1, \cdots, L,$$

where d_{ij} is the known distance and C_l is the set of microphones within the lth array. Then, the inverse problem is

$$\begin{aligned} \min_{\hat{\mathbf{a}}_i} \quad & \sum |\tau_{\gamma,ij} - \hat{\tau}_{\gamma,ij}|^2 \\ s.t. \quad & d_{ij}^2 = \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2, \ \forall i,j \in C_l, i > j, \ l = 1, \cdots, L. \end{aligned} \tag{2.22}$$

One can identify suitable locations for placing sound sources and calibrate the array configuration. Once the signals are recorded, the inverse problem can be solved by nonlinear optimization techniques. This inverse problem can be illustrated in Fig. 2.3.

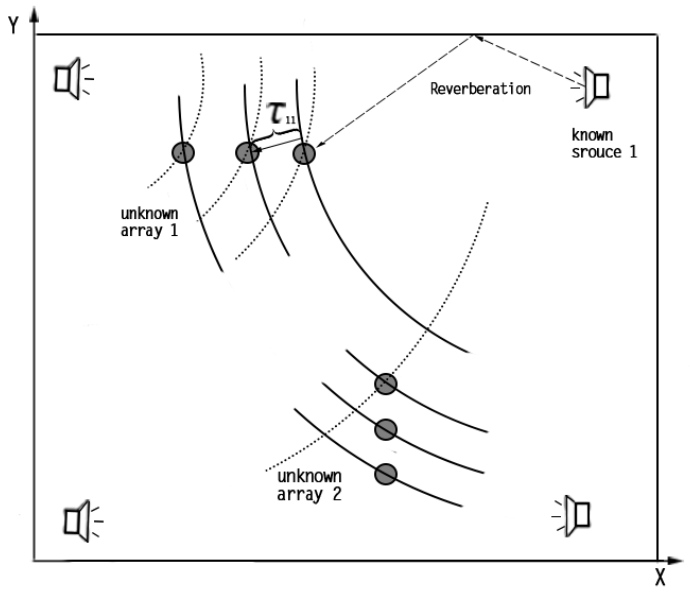


Figure 2.3: Geometric setup of inverse problem

However, the problem (2.22) is highly nonlinear and nonconvex. One common idea to tackle is to employ convex relaxation, which can convert the problem into a convex optimization problem. This relaxation technique has been considered to solve the inverse problem in [135, 150] based on TOA measurements. As far as our knowledge is concerned, there is no relaxation method used to solve the inverse problem based on the TDOA measurements, and in this chapter, we aim to fill this blank.

2.4 Relaxation Models

The sensor localization problem problem (2.22) is highly nonlinear and nonconvex. In tackling the problem, efficient relaxation models have been proposed based on TOA measurements, such as SDP [135] and SOCP relaxation [150]. A relaxation model for the direct problem with TDOA measures has been proposed in [49]. After relaxation, the optimization problem become convex. In this section, both SDP and SOCP relaxation models will be extended for TDOA measurements. Then, a mixed

SDP-SOCP relaxation model will be proposed and described.

2.4.1 SDP Relaxation Model

The purpose the optimal problem in (2.22) is to make $\hat{\tau}_{\gamma,ij} \approx \tau_{\gamma,ij}$. It is equivalent to find a set of suitable $\hat{\mathbf{a}}_i$ satisfying

$$\begin{aligned} &\tau_{\gamma,ij} = (\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| - \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_j\|)/c && \forall i,j \in C, i > j, \gamma = 1, \cdots, n, \\ &d_{ij}^2 = \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L. \end{aligned} \tag{2.23}$$

Set $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i}$ and $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_j\| = \beta_{\gamma,j}$. Then, the equations in (2.23) can be expressed as

$$\begin{aligned} &\beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1, \cdots, n, \\ &\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\ &d_{ij}^2 = \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2, && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L. \end{aligned} \tag{2.24}$$

Let $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i}, \forall i \in C, \gamma = 1, \cdots, n$. We can rewrite the squared distance between sensor and source with respect to the target matrix $\boldsymbol{A}$. For all $i \in C, \gamma = 1, \cdots, n$, we have

$$\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_i^T \end{pmatrix} \begin{pmatrix} I_d \\ \boldsymbol{A}^T \end{pmatrix} \begin{pmatrix} I_d & \boldsymbol{A} \end{pmatrix} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_i \end{pmatrix}, \tag{2.25}$$

where e_i is a vector of all zeros except an -1 at the ith position and I_d is identity matrix. To simplify the equation (2.25), we introduce a symmetric matrix $\boldsymbol{Y} \in R^{n \times n}$ and $\boldsymbol{Y} = \boldsymbol{A}^T\boldsymbol{A}$. Equation (2.25) is equivalent to

$$\begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_i^T \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{Y} \end{pmatrix} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_i \end{pmatrix} = \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2. \tag{2.26}$$

Then, the SDP relaxation is to replace the equivalent constraint $\boldsymbol{Y} = \boldsymbol{A}^T\boldsymbol{A}$ by an inequivalent constraint $\boldsymbol{Y} \succeq \boldsymbol{A}^T\boldsymbol{A}$, which is equal to

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{Y} \end{pmatrix} \succeq 0. \tag{2.27}$$

Similarly, we can rewrite the squared distance between the sensors in the same array as

$$\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = \mathbf{e}_{ij}^T \boldsymbol{A}^T \boldsymbol{A} \mathbf{e}_{ij}, \tag{2.28}$$

where e_{ij} is the vector with 1 at the ith position and an -1 at the j the position and zeros the other positions. It can be further written as

$$\begin{pmatrix} \mathbf{0} & \mathbf{e}_{ij}^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{ij} \end{pmatrix} = d_{ij}^2 \quad \forall i, j \in C_l, i > j,\ l = 1, \cdots, L. \tag{2.29}$$

As $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i}$ and $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i}$, then $\beta_{\gamma,i}^2 = \alpha_{\gamma,i}$. We can also relax this equation to

$$\begin{pmatrix} 1 & \beta_{\gamma,i} \\ \beta_{\gamma,i} & \alpha_{\gamma,i} \end{pmatrix} \succeq 0.$$

If there isn't a relaxation, $\alpha_{\gamma,i}$ should be equal to $\beta_{\gamma,i}$. After relaxation, we should make sure the value of $\alpha_{\gamma,i}$ is as small as possible. Therefore, we minimize $\sum \alpha_{\gamma,i}$ in the objective function, where $i = 1, \cdots, n,\ \forall i \in C$. Thus, the relaxed model becomes a standard SDP problem

$$\begin{aligned}
\min_{\boldsymbol{W}, \alpha_{\gamma,i}, \beta_{\gamma,i}} \quad & \sum \alpha_{\gamma,i} \\
\mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,i} = c\tau_{\gamma,ij}, && \forall i, j \in C, i > j, \gamma = 1, \cdots, n, \\
& \begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_i^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_i \end{pmatrix} = \alpha_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\
& \begin{pmatrix} \mathbf{0} & \mathbf{e}_{ij}^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{ij} \end{pmatrix} = d_{ij}^2, && \forall i, j \in C_l, i > j,\ l = 1, \cdots, L, \\
& \begin{pmatrix} 1 & \beta_{\gamma,i} \\ \beta_{\gamma,i} & \alpha_{\gamma,i} \end{pmatrix} \succeq 0, && \forall i \in C, \gamma = 1, \cdots, n, \\
& \boldsymbol{W}_{1:d,1:d} = I_d, \boldsymbol{W} \succeq 0.
\end{aligned} \tag{2.30}$$

Compared with TOA measurements, when applying TDOA, there are two folds SDP relaxation, and the relaxation becomes quite weaker, and numerical results also prove it.

2.4.2 SOCP Relaxation Model

For the SOCP relaxation model, we can also rewrite the equations in (2.23) to

$$\begin{aligned}
&\beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1, \cdots, n, \\
&\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\
&\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L.
\end{aligned}$$

Inspired by the idea in [150], we can relax the second equivalent constraint into " $\leq$ " inequivalent constraints and the relaxation model is as follows

$$\begin{aligned}
\min_{\hat{\mathbf{a}}_i, \beta_{\gamma,i}} \quad & \sum \beta_{\gamma,i} \\
\mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1, \cdots, n, \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| \leq \beta_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\
& \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L.
\end{aligned} \tag{2.31}$$

By the similar method, we further relax the third equation as

$$\begin{aligned}
\min_{\hat{\mathbf{a}}_i, \beta_{\gamma,i}, z_{ij}} \quad & \sum \beta_{\gamma,i} + \sum |z_{ij} - d_{ij}^2| \\
\mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1, \cdots, n, \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| \leq \beta_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\
& \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 \leq d_{ij}, && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L.
\end{aligned} \tag{2.32}$$

Then, problem becomes an SOCP which is a convex problem. Obviously, this model has a simpler structure. Many optimization algorithms can be applied on

SOCP problem such as an interior-point algorithm. We can get positions of microphones by solving the SOCP problem over the entire sensor network. Based on the knowledge of [150], if the microphone location leaves the convex hull, the SOCP model can not work well. To generate a more robust model, we further combine these two models.

2.4.3 SDP-SOCP Relaxation Model

Similarly, rewrite the equations in (2.23) as

$$\begin{aligned} &\beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1, \cdots, n \\ &\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n \\ &\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L. \end{aligned} \tag{2.33}$$

Set $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i},\ \forall i \in C, \gamma = 1, \cdots, n$. Then, we can get the relationship between $\alpha_{\gamma,i}$ and $\beta_{\gamma,i}$ and the constraints becomes

$$\begin{aligned} &\beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1, \cdots, n, \\ &\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\ &\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i}, && \forall i \in C, \gamma = 1, \cdots, n, \\ &\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1, \cdots, L, \\ &\alpha_{\gamma,i} = (\beta_{\gamma,i})^2, && \forall i \in C, \gamma = 1, \cdots, n. \end{aligned} \tag{2.34}$$

Rewrite the equality constraint $\alpha_{\gamma,i} = (\beta_{\gamma,i})^2$ into following inequalities

$$\begin{aligned} &\alpha_{\gamma,i} \geq (\beta_{\gamma,i})^2, && \forall i \in C, \gamma = 1, \cdots, n, \\ &\alpha_{\gamma,i} \leq (\beta_{\gamma,i})^2, && \forall i \in C, \gamma = 1, \cdots, n. \end{aligned} \tag{2.35}$$

The inequality constraints $\alpha_{\gamma,i} \geq (\beta_{\gamma,i})^2$ is equivalent to $\begin{pmatrix} 1 & \beta_{\gamma,i} \\ \beta_{\gamma,i} & \alpha_{\gamma,i} \end{pmatrix} \succeq 0$, and $\alpha_{\gamma,i} \leq (\beta_{\gamma,i})^2$ is equivalent to $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| \leq \beta_{\gamma,i}$. In addition, equality constraints $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| =$

$\beta_{\gamma,i}$ can be relaxed into " $\leq$ " inequality constraints, which yields a second order cone problem. If without the relaxation, $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|$ should be equal to the $\beta_{\gamma,i}$, therefore, $\beta_{\gamma,i}$ should be as small as possible after relaxation. Thus, we add $\beta_{\gamma,i}$ to the objective function, and the objective function becomes

$$\min \quad \sum \beta_{\gamma,i}.$$

If without the relaxation, $(\beta_{\gamma,i})^2$ equals to $\alpha_{\gamma,i}$. As $\beta_{\gamma,i}$ is positive, minimizing $\beta_{\gamma,i}$ is equivalent to making $\alpha_{\gamma,i}$ as small as possible. However, $\alpha_{\gamma,i} = \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2$ is a quadratic function which is satisfied in optimization problem, so that we replace $\sum \beta_{\gamma,i}$ with $\sum \alpha_{\gamma,i}$ in the objective function. Then the original microphone localization problem could be transformed into

$$\begin{aligned}
\min_{\hat{\mathbf{a}}_i,\alpha_{\gamma,i},\beta_{\gamma,i}} \quad & \sum \alpha_{\gamma,i} \\
\mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1,\cdots,n \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| \leq \beta_{\gamma,i}, && \forall i \in C, \gamma = 1,\cdots,n, \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i}, && \forall i \in C, \gamma = 1,\cdots,n, \\
& \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1,\cdots,L, \\
& \begin{pmatrix} 1 & \beta_{\gamma,i} \\ \beta_{\gamma,i} & \alpha_{\gamma,i} \end{pmatrix} \succeq 0, && \forall i \in C, \gamma = 1,\cdots,n.
\end{aligned} \tag{2.36}$$

For the equivalent constraints $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i}$ and $\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2$, based on SDP relaxation rule in section 2.3, we can relax it into following constraints

$$\begin{aligned}
& \begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_i^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_i \end{pmatrix} = \alpha_{\gamma,i}, && \forall i \in C, \gamma = 1,\cdots,n, \\
& \begin{pmatrix} \mathbf{0} & \mathbf{e}_{ij}^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{ij} \end{pmatrix} = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1,\cdots,L,
\end{aligned} \tag{2.37}$$

$$\boldsymbol{W}_{1:d,1:d} = I_d,$$

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{Y} \end{pmatrix} \succeq 0,$$

where e_i is a vector of all zeros except an -1 at the ith position; and e_{ij} is the vector with 1 at the ith position and an -1 at the j the position and zero the other positions. Finally, the relaxed version of the original problem (2.21) can be represented as the following mixed SDP-SOCP relaxation model

$$\begin{aligned}
\min_{\hat{\mathbf{a}}_i,\alpha_{\gamma,i},\beta_{\gamma,i}} \quad & \sum \alpha_{\gamma,i} \\
\mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij}, && \forall i,j \in C, i > j, \gamma = 1,\cdots,n \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| \le \beta_{\gamma,i}, && \forall i \in C, \gamma = 1,\cdots,n, \\
& \begin{pmatrix} 1 & \beta_{\gamma,i} \\ \beta_{\gamma,i} & \alpha_{\gamma,i} \end{pmatrix} \succeq 0, && \forall i \in C, \gamma = 1,\cdots,n, \\
& \begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_i^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_i \end{pmatrix} = \alpha_{\gamma,i}, && \forall i \in C, \gamma = 1,\cdots,n, \\
& \begin{pmatrix} \mathbf{0} & \mathbf{e}_{ij}^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{ij} \end{pmatrix} = d_{ij}^2, && \forall i,j \in C_l, i > j,\ l = 1,\cdots,L, \\
& \boldsymbol{W}_{1:d,1:d} = I_d, \\
& \boldsymbol{W} = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{Y} \end{pmatrix} \succeq 0.
\end{aligned} \tag{2.38}$$

2.5 Analysis of the Mixed SDP-SOCP Relaxation

In this section, we analyze when the SDP-SOCP model has an exact relaxation. This occurs when $\hat{\tau}_{\gamma,ij} = \tau_{\gamma,ij},\ \forall i,j \in C, i > j, \gamma = 1,\cdots,n$. In the following section, the properties of the mixed relaxation model will be studied, and some lemmas to verify the optimal solution of (2.38) will be given.

For convenience, we can firstly simplify the problem (2.38). Let y_i be the ith

diagonal element of matrix $\mathbf{Y}$, and $\mathbf{y} = [y_1, \cdots, y_m]$. Then, we have

$$\begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_i^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_i \end{pmatrix} = \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 + (y_i - \|\hat{\mathbf{a}}_i\|^2) = \alpha_{\gamma,i},$$

and

$$\begin{pmatrix} \mathbf{0} & \mathbf{e}_{ij}^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{ij} \end{pmatrix} = \|\mathbf{a}_i - \mathbf{a}_j\|^2 + (y_i - \|\mathbf{a}_i\|^2) + (y_j - \|\mathbf{a}_j\|^2) = d_{ij}^2.$$

As $\boldsymbol{W} \succeq 0$, we can get $y_i \geq \|\hat{\mathbf{a}}_i\|$. Rewrite the constraint $\begin{pmatrix} 1 & \beta_{\gamma,i} \\ \beta_{\gamma,i} & \alpha_{\gamma,i} \end{pmatrix} \succeq 0$ as $\beta_{\gamma,i}^2 \leq \alpha_{\gamma,i}$. The simplified model is formulated as follows

$$\begin{aligned} \min_{\hat{\mathbf{a}}_{li}, \alpha_{\gamma,i}, \beta_{\gamma,i}, y_i} \quad & \sum \alpha_{\gamma,i} \\ \mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij} \quad \forall i, j \in C, i > j, \gamma = 1, \cdots, n, \\ & \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 + (y_i - \|\hat{\mathbf{a}}_i\|^2) = \alpha_{\gamma,i} \quad \forall i \in C, \gamma = 1, \cdots, n, \\ & \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 + (y_i - \|\hat{\mathbf{a}}_i\|^2) + (y_j - \|\hat{\mathbf{a}}_j\|^2) = d_{ij}^2 \\ & \forall i, j \in C_l, i > j, \ l = 1, \cdots, L, \\ & y_i \geq \|\hat{\mathbf{a}}_i\|^2, \ \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 \leq \beta_{\gamma,i}^2 \leq \alpha_{\gamma,i} \quad \forall i \in C, \gamma = 1, \cdots, n. \end{aligned} \tag{2.39}$$

Denote one feasible solution of (2.39) as $(\boldsymbol{A}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{y})$, where $\boldsymbol{\alpha} = (\alpha_{\gamma,i})$ and $\boldsymbol{\beta} = (\beta_{\gamma,i})$. The optimal solution and the real true solution is denoted as $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\mathbf{y}})$ and $(\boldsymbol{A}^*, \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \mathbf{y}^*)$, respectively. We can use the following lemmas to describe the relationship between the real solution and optimal solution.

Lemma 2.1. *Suppose that $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\mathbf{y}})$ is an optimal solution of (2.39). Then there exists a λ_γ, $\gamma = 1, \cdots, n$ such that*

$$\hat{\beta}_{\gamma,i} = (\beta_{\gamma,i})^* - \lambda_\gamma \quad \forall i \in C.$$

Proof. If $(\boldsymbol{A}^*, \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \mathbf{y}^*)$ is the true solution of problem (2.39), it satisfies that

$$(\beta_{\gamma,i}^*)^2 = \alpha_{\gamma,i}^* = \|s_\gamma - a_i^*\|^2, \ \forall i \in C, \gamma = 1, \cdots, n.$$

Notice that both $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ are the solutions of a series of linear equations

$$\beta_{\gamma,1} - \beta_{\gamma,j} = c\tau_{\gamma,1j}, \quad \gamma = 1, \cdots, n, \forall j \in C, j \neq 1.$$

We divide these equations into n groups according to different source locations. For each source γ, there are $m-1$ equations with m variables, and the rank of the coefficient matrix is $m-1$. Then, there is λ_i such that

$$\hat{\beta}_j = \beta_j^* - \lambda_\gamma \quad \forall j \in C.$$

□

Lemma 2.2. *Suppose that* $(\hat{\boldsymbol{A}}, \hat{\beta}, \hat{\alpha}, \hat{\mathbf{y}})$ *is an optimal solution of (2.39) and* $(\boldsymbol{A}^*, \beta^*, \alpha^*, \mathbf{y}^*)$ *is a true solution. Then, if* $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$ *and the number of sources is larger than 2, then* $\hat{\boldsymbol{A}} = \boldsymbol{A}^*$.

Proof. If $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$ and $\beta_{ij}^* = \|s_i - a_j^*\|, \forall i, j$, then $\hat{\beta}_{ij} = \|s_i - \hat{a}_j\| = \|s_i - a_j^*\|, \forall i, j$. We define that $D(\mathbf{s}, r) = \{\boldsymbol{b} || \boldsymbol{b} - \mathbf{s}| = r\}$. Thus, if the number of sources $p > 2$, for all $j = 1, \cdots, m$, we have $\hat{\mathbf{a}}_j = \cap_{i=1}^p D(\mathbf{s}_i, \hat{\beta}_{ij}) = \cap_{i=1}^p D(\mathbf{s}_i, \beta_{ij}^*) = \mathbf{a}_j^*$, which means $\hat{\boldsymbol{A}} = \boldsymbol{A}^*$.

□

Lemma 2.3. *In a distributed acoustic network with more than 2 sources, if one microphone is localizable, then all the microphones are localizable.*

Proof. If the zth microphone is localizable, then for all $\gamma = 1, \cdots, n$, it has

$$\hat{\beta}_{\gamma,z} = (\beta_{\gamma,z})^*.$$

According to Lemma 2.1, we can get $\forall \gamma = 1, \cdots, n$, $\lambda_\gamma = 0$, and $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$. If $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$ and $\beta_{\gamma,i}^* = \|\mathbf{s}_\gamma - \mathbf{a}_i^*\|, \forall \gamma, i$, then $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| \leq \hat{\beta}_{\gamma,i} = \|\mathbf{s}_\gamma - \mathbf{a}_i^*\|, \forall \gamma, i$. We define that $D(\mathbf{s}, r) = \{\boldsymbol{b} |\ \|\boldsymbol{b} - \mathbf{s}\| = r\}$ and $F(\mathbf{s}, r) = \{\boldsymbol{b} |\ \|\boldsymbol{b} - \mathbf{s}\| \leq r\}$. For all $i \in C$,

we have $\hat{\mathbf{a}}_i = \cap_{\gamma=1}^{n} F(\mathbf{s}_\gamma, \hat{\beta}_{\gamma,i}) \supseteq \cap_{\gamma=1}^{n} D(\mathbf{s}_\gamma, \hat{\beta}_{\gamma,i}) = \mathbf{a}_i^*$. If the number of sources $n > 2$ and $\cap_{\gamma=1}^{n} D(\mathbf{s}_\gamma, \hat{\beta}_{\gamma,i}) \neq \emptyset$, we have $\cap_{\gamma=1}^{n} F(\mathbf{s}_\gamma, \hat{\beta}_{\gamma,i}) = \cap_{\gamma=1}^{n} D(\mathbf{s}_\gamma, \hat{\beta}_{\gamma,i})$. Thus, for all i, $\hat{\mathbf{a}}_i = \mathbf{a}_i^*$, which means $\hat{\boldsymbol{A}} = \boldsymbol{A}^*$. We can have that all the microphones are localizable. □

Lemma 2.4. *Let $(\hat{\boldsymbol{A}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ be an optimal solution of (2.39). Then, for every s_i, $\|s_i - a_j\| < \hat{\beta}_{ij}, j = 1, \cdots, m$ cannot be satisfied simultaneously.*

Proof. For a certain s_i, if $\forall j = 1, \cdots, m$, we have $\|s_i - a_j\| < \hat{\beta}_{ij}$. Set $\varepsilon = min_{j=1,\cdots,m}\{\hat{\beta}_{ij} - \|\mathbf{s}_i - \hat{\mathbf{a}}_j\|\}$, and set $\bar{\beta}_{ij} = \hat{\beta}_{ij} - \frac{\varepsilon}{2}$. Given $\xi > 0$, such that $\mathbf{y} - \|\hat{\mathbf{a}}\| \geq \xi$. Then, we can have $(\hat{\beta}_{ij} - \frac{\varepsilon}{2})^2 < \hat{\alpha}_{ij} - \xi$. Set $\bar{\boldsymbol{A}} = \hat{\boldsymbol{A}}$, $\bar{\mathbf{y}} = \hat{\mathbf{y}} - \xi$ and $\bar{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}} - \xi$. Then, $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{y}})$ is also a feasible solution, and $\sum \bar{\alpha}_{ij} < \sum \hat{\alpha}_{ij}$. Thus, $(\hat{\boldsymbol{A}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ is not the optimal solution. □

Lemma 2.5. *If the network is localizable, then there is at least one microphone in the convex hull of the sources.*

Proof. We prove this lemma by contradiction. Denote the convex hull of the sources as $conv\{s_\gamma\}_{\gamma=1,\cdots,n}$. Suppose that the convex hull fails to hold for all the sensors. According to Lemma 2.3, if the network is localizable, then all the sensors are localizable, which means $(\boldsymbol{A}^*, \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \mathbf{y}^*)$ is the optimal solution of (2.39). Let $\mathbf{p}_i$ be the nearest point projection of $\mathbf{a}_i^*$ onto $conv\{s_\gamma\}_{\gamma=1,\cdots,n}$. Then, we have $\mathbf{p}_i \neq \mathbf{a}_i^*$, and $\forall \gamma = 1, \cdots, n$, we have $(\mathbf{a}_i^* - \mathbf{p}_i)^T(\mathbf{p}_i - \mathbf{s}_\gamma) \geq 0$, which means

$$\begin{aligned} \|\mathbf{s}_\gamma - \mathbf{a}_i^*\|^2 &= \|\mathbf{s}_\gamma - \mathbf{p}_i + \mathbf{p}_i - \mathbf{a}_i^*\|^2 \\ &= \|\mathbf{s}_\gamma - \mathbf{p}_i\|^2 + \|\mathbf{p}_i - \mathbf{a}_i^*\|^2 + 2(\mathbf{s}_\gamma - \mathbf{p}_i)^T(\mathbf{p}_i - \mathbf{a}_i^*) \\ &> \|\mathbf{s}_\gamma - \mathbf{p}_i\|^2. \end{aligned}$$

Define $\mathbf{a}_i^\varepsilon = (1-\varepsilon)\mathbf{a}_{li}^* + \varepsilon \mathbf{p}_i$, where $\varepsilon \in (0, 1)$. As $(\boldsymbol{A}^*, \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \mathbf{y}^*)$ is the true solution, we have $\|\mathbf{s}_\gamma - \mathbf{a}_i^*\|^2 = \beta_{\gamma,i}^*$, $\forall i \in C, \gamma = 1, \cdots, n$. Due to the convexity and continuity

of $\|\cdot\|^2$, for all ε is sufficiently small, we have

$$\|\mathbf{s}_\gamma - \mathbf{a}_i^\varepsilon\|^2 < (\beta_{\gamma,i}^*)^2,\ \forall i \in C, \gamma = 1, \cdots, n.$$

Set $\varepsilon = min\{\beta_{\gamma,i} - \|\mathbf{s}_\gamma - \mathbf{a}_i^\varepsilon\|$, and $\bar{\beta}_{\gamma,i} = \beta_{\gamma,i}^* - \frac{\varepsilon}{2}$. Given a $\bar{\mathbf{y}}$, such that $\|\mathbf{s}_\gamma - \mathbf{a}_i^\varepsilon\|^2 + (\bar{y}_i - \|\mathbf{a}_i^\varepsilon\|^2) = \alpha_{\gamma,i}^*,\ \forall i \in C, \gamma = 1, \cdots, n$. As $\|\mathbf{s}_\gamma - \mathbf{a}_i^\varepsilon\|^2 < (\beta_{\gamma,i}^*)^2 = \alpha_{\gamma,i}^*,\ \forall i \in C, \gamma = 1, \cdots, n$, we can always find a $\xi > 0$, such that $\bar{y}_i - \|\mathbf{a}_i^\varepsilon\|^2 \geq \xi$, and $(\beta_{\gamma,i}^* - \frac{\varepsilon}{2})^2 \leq \alpha_{\gamma,i}^* - \xi$. Set $\bar{\boldsymbol{A}} = \{\mathbf{a}_i^\varepsilon\}$, and $\bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^* - \xi$. Then, $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{y}})$ is also a feasible solution, and $\sum \bar{\alpha}_{\gamma,i} < \sum \alpha_{\gamma,i}^*$. Thus, it contradicts the assumption that $(\boldsymbol{A}^*, \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \mathbf{y}^*)$ is the optimal solution. □

Lemma 2.6. *If the optimal solution* $\hat{\boldsymbol{W}} = \begin{pmatrix} \boldsymbol{I}_d & \hat{\boldsymbol{A}} \\ \hat{\boldsymbol{A}}^T & \hat{\mathbf{Y}} \end{pmatrix}$ *satisfies* $\hat{\mathbf{Y}} = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$*, which is equivalent to the rank of* $\hat{\boldsymbol{W}}$ *is* d *(the dimension of the network), then* $\hat{\boldsymbol{A}}$ *is also a solution of original problem (2.21).*

Proof. For the matrix $\hat{\boldsymbol{W}}$, we have $\hat{\boldsymbol{W}} = \begin{pmatrix} \boldsymbol{I}_d & \hat{\boldsymbol{A}} \\ \hat{\boldsymbol{A}}^T & \hat{\mathbf{Y}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_d & \mathbf{0} \\ \hat{\boldsymbol{A}}^T & \boldsymbol{I}_d \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_d & \hat{\boldsymbol{A}} \\ \mathbf{0} & \hat{\mathbf{Y}} - \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}. \end{pmatrix}$, where I_d is a d-dimension matrix. We can see that the rank of $\hat{\boldsymbol{W}}$ is at least d. If the rank of $\hat{\boldsymbol{W}}$ is d, then $Trace\{\hat{\mathbf{Y}} - \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}\}$ must equals to 0, which means $\hat{\mathbf{Y}} = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$. Inversely, if $\hat{\mathbf{Y}} = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$, we have $rank\{\hat{\boldsymbol{W}}\} = rank\{\begin{pmatrix} \boldsymbol{I}_d & \hat{\boldsymbol{A}} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\} = d$. The rank of $\hat{\boldsymbol{W}}$ is d.

If $\hat{\mathbf{Y}} = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$, then $\forall i$, we have $\hat{y}_i = \|\hat{\mathbf{a}}_i\|^2$, which means $\forall i \in C, \gamma = 1, \cdots, n$, $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \hat{\alpha}_{\gamma,i}$, and $\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j = d_{ij}$. As $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 \leq \beta_{\gamma,i}^2 \leq \alpha_{\gamma,i}$ and $\|s_\gamma - a_i\|^2 = \hat{\alpha}_{\gamma,i}$, $\forall i \in C, \gamma = 1, \cdots, n$, we can get $\hat{\beta}_{\gamma,i}^2 = \hat{\alpha}_{\gamma,i}$ and $\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \hat{\beta}_{\gamma,i}, \forall i \in C, \gamma = 1, \cdots, n$.

Then, the problem (2.39) becomes

$$\begin{aligned}
\min_{\hat{\mathbf{a}}_i,\alpha_{\gamma,i},\beta_{\gamma,i},y_i} \quad & \sum \alpha_{\gamma,i} \\
\mathbf{s}.t. \quad & \beta_{\gamma,i} - \beta_{\gamma,j} = c\tau_{\gamma,ij} \quad \forall i,j \in C, i > j, \gamma = 1,\cdots,n, \\
& \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2 \quad \forall i,j \in C_l, i > j,\ l = 1,\cdots,L, \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\|^2 = \alpha_{\gamma,i} \quad \forall i \in C, \gamma = 1,\cdots,n, \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| = \beta_{\gamma,i} \quad \forall i \in C, \gamma = 1,\cdots,n.
\end{aligned} \tag{2.40}$$

Then, the constraints of (2.40) can be written as

$$\begin{aligned}
&\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| - \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_j\| = c\tau_{\gamma,ij} && \forall i,j \in C, i > j, \gamma = 1,\cdots,n, \\
&\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 = d_{ij}^2 && \forall i,j \in C_l, i > j,\ l = 1,\cdots,L.
\end{aligned}$$

It means $\hat{\tau}_{\gamma,ij} = \tau_{\gamma,ij},\ \forall \gamma,i,j$. We can prove that the optimal solution of problem (2.40) is also a solution of original problem without realxation.

□

Given a set of $\hat{\tau}_{\gamma,ij}$, we can get an calibration of the geometry $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{W}})$ by solving (2.38). The lemmas above can check whether the estimation is correct. Lemma 2.3 establishes that, as long as there is an $\hat{\mathbf{a}}_i \in \hat{\boldsymbol{A}}$ satisfying $\hat{\mathbf{a}}_i = \mathbf{a}_i^*$, then the whole network is localizable. It means that, in a distributed acoustic network, if one microphone's location is known, we can use this location information to check the correctness of the other positions. Lemma 2.5 gives a localizable condition that at least one sensor should be in the convex hull of the sources. Moreover, Lemma 2.6 implies that if problem (2.23) has a solution for $\hat{\boldsymbol{A}}$ with rank d, then the network is localizable.

Given a set of $\hat{\tau}_{\gamma,li,kj}$, we can get an calibration of the geometry $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{W}})$ by solving (2.38). The lemmas above can be used to check whether the estimation is correct. Lemma 2.3 establishes that, as long as there is a $\hat{\mathbf{a}}_{li} \in \hat{\boldsymbol{A}}$ making $\hat{\mathbf{a}}_{li} = \mathbf{a}_{li}^*$, then

the whole network is localizable. It means that, in a distributed acoustic network, if one microphone's location is known, we can use this location information to check the correctness of the other positions. Moreover, Lemma 2.5 implies that if problem (2.23) has a solution for $\hat{\boldsymbol{A}}$ being rank d, then the network is localizable.

To understand the lemmas better, we give two examples. Suppose there are four sources $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4]$, two microphone arrays each with only one sensor $\boldsymbol{A} = [\mathbf{a}_1, \mathbf{a}_2]$, and the dimension of the network is 2. Then, we have the following equations:

$$\begin{aligned}
\|\mathbf{s}_1 - \hat{\mathbf{a}}_1\| - \|\mathbf{s}_1 - \hat{\mathbf{a}}_2\| = \hat{\beta}_{1,1} - \hat{\beta}_{1,2} &= \tau_{1,12}c \\
\|\mathbf{s}_2 - \hat{\mathbf{a}}_1\| - \|\mathbf{s}_2 - \hat{\mathbf{a}}_2\| = \hat{\beta}_{2,1} - \hat{\beta}_{2,2} &= \tau_{2,12}c \\
\|\mathbf{s}_3 - \hat{\mathbf{a}}_1\| - \|\mathbf{s}_3 - \hat{\mathbf{a}}_2\| = \hat{\beta}_{3,1} - \hat{\beta}_{3,2} &= \tau_{3,12}c \\
\|\mathbf{s}_4 - \hat{\mathbf{a}}_1\| - \|\mathbf{s}_4 - \hat{\mathbf{a}}_2\| = \hat{\beta}_{4,1} - \hat{\beta}_{4,2} &= \tau_{4,12}c
\end{aligned}$$

Set $\boldsymbol{\tau} = [\tau_{1,12}, \tau_{2,12}, \tau_{3,12}, \tau_{4,12}]$. Based on Lemma 2.1, for the optimal $\hat{\beta}$ and the real β^*, we have

$$\begin{cases} \hat{\beta}_{1,1} = \beta^*_{1,1} - \lambda_1 \\ \hat{\beta}_{1,2} = \beta^*_{1,2} - \lambda_1 \end{cases}$$

$$\begin{cases} \hat{\beta}_{2,1} = \beta^*_{2,1} - \lambda_2 \\ \hat{\beta}_{2,2} = \beta^*_{2,2} - \lambda_2 \end{cases}$$

$$\begin{cases} \hat{\beta}_{3,1} = \beta^*_{3,1} - \lambda_3 \\ \hat{\beta}_{3,2} = \beta^*_{3,2} - \lambda_3 \end{cases}$$

$$\begin{cases} \hat{\beta}_{4,1} = \beta^*_{4,1} - \lambda_4 \\ \hat{\beta}_{4,2} = \beta^*_{4,2} - \lambda_4 \end{cases}$$

If $\mathbf{a}_1$ is localizable, it means $\hat{\beta}_{\gamma,1} = \beta^*_{\gamma,1},\ \gamma = 1,2,3,4$. Then, we have $\lambda_\gamma = 0,\ \gamma = 1,2,3,4$. Accordingly, $\hat{\beta}_{\gamma,2} = \beta^*_{\gamma,2},\ \gamma = 1,2,3,4$. Then, $\hat{\mathbf{a}}_2 = \cap_{\gamma=1}^4 D(\mathbf{s}_\gamma, \hat{\beta}_{\gamma,2}) = \cap_{\gamma=1}^4 D(\mathbf{s}_\gamma, \beta^*_{\gamma,2}) = \mathbf{a}^*_2$, where $D(\mathbf{s}, r) = \{\boldsymbol{b} |\ \|\boldsymbol{b} - \mathbf{s}\| = r\}$. It is in accordance with Lemma 2.3.

- **Example 1**

  In this example, we set the location of the sources as {(0, 0), (0, 1), (1, 0), (1, 1)}, and two microphones needed to be estimated are located at {(0.2, 0.4), (0.7, 0.6)}. By solving the optimal problem (2.38), we can get an optimal solution as

$$\hat{\boldsymbol{W}} = \begin{pmatrix} \boldsymbol{I}_d & \hat{\boldsymbol{A}} \\ \hat{\boldsymbol{A}}^T & \hat{\mathbf{Y}} \end{pmatrix} = \begin{bmatrix} 1.00 & 0.00 & 0.20 & 0.70 \\ 0.00 & 1.00 & 0.40 & 0.60 \\ 0.20 & 0.40 & 0.20 & 0.38 \\ 0.70 & 0.60 & 0.38 & 0.85 \end{bmatrix}.$$

  The microphone location matrix is given as

$$\hat{\boldsymbol{A}} = \begin{bmatrix} 0.20 & 0.70 \\ 0.40 & 0.60 \end{bmatrix}.$$

  The optimal matrix $\hat{\boldsymbol{W}}$ satisfies that $\hat{\mathbf{Y}} = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$, and rank of $\hat{\boldsymbol{W}}$ is 2. According to Lemma 2.5, the estimation location should be the real location. The result is depicted in Fig. 2.4. It is clear that the estimation is correct.

- **Example 2**

  In this example, we set the location of the sources as {(0, 0), (0, 1), (1, 0), (1, 1)}, which is the same as Example a. We move the two microphones outside the convex hull formulated by the sources, and the locations of the microphones are {(−0.2, 0.2), (−0.2, 0.6)}. By solving the optimal problem (2.38), we can

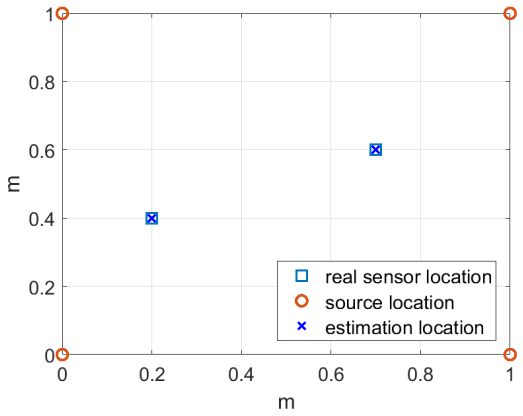


Figure 2.4: The result of Example a

get an optimal solution as

$$\hat{\boldsymbol{W}} = \begin{bmatrix} 1.00 & 0.00 & 0.31 & 0.32 \\ 0.00 & 1.00 & 0.31 & 0.67 \\ 0.31 & 0.31 & 0.32 & 0.31 \\ 0.32 & 0.67 & 0.31 & 0.62 \end{bmatrix}.$$

The decision matrix $\hat{\boldsymbol{W}}$ does not satisfy $\hat{\mathbf{Y}} = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$, and rank of $\hat{\boldsymbol{W}}$ is 4.

The microphone location matrix is estimated as

$$\hat{\boldsymbol{A}} = \begin{bmatrix} 0.31 & 0.32 \\ 0.31 & 0.67 \end{bmatrix},$$

while the true solution is

$$\boldsymbol{A}^* = \begin{pmatrix} -0.20 & -0.2 \\ 0.20 & 0.6 \end{pmatrix}.$$

It is clear that there is an error between the estimation location and the true location, which is also in accord with Lemma 2.5. The optimal value of $\hat{\boldsymbol{\beta}}$ is given as

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 0.44 & 0.79 \\ 0.83 & 0.45 \\ 0.83 & 0.95 \\ 0.97 & 0.79 \end{pmatrix}.$$

The real value of $\boldsymbol{\beta}^*$ is

$$\boldsymbol{\beta}^* = \begin{pmatrix} 0.28 & 0.63 \\ 0.82 & 0.44 \\ 1.21 & 1.34 \\ 1.44 & 1.26 \end{pmatrix}.$$

We can see that there is a gap between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ and this gap induces an error between the estimation location and real location.

The source locations, real sensor locations and estimation locations are depicted in Fig. 2.5. It is clear that both of the sensors are not localizable.

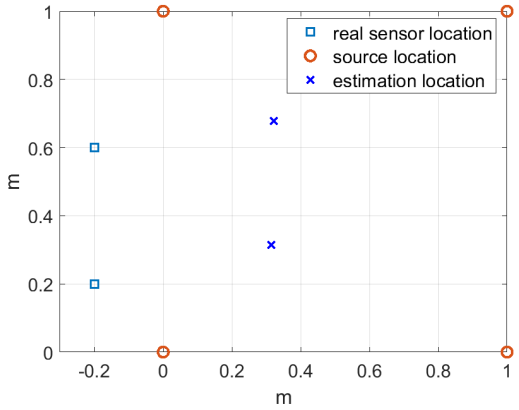


Figure 2.5: The result of Example b

2.6 Offset of TDOA with Real Data

Although the inverse problem is well developed above, there is still another problem that we need to consider when using real data. In application, the TDOAs are estimated by the real signals received by the sensors. However, the distributed acoustic network sensors are inherently asynchronous, which reduces a temporal offset in each channel. Therefore, the real data can not directly be applied to the inverse problem. In this section, the task is to investigate the optimization of the temporal offset estimation coupled with the inverse problem.

Denote the first microphone in the first array as the reference microphone. According to the definition, the real TDOA concerning the source and sensor locations is given as

$$\tau_{\gamma,1j} = \frac{\|\mathbf{s}_\gamma - \mathbf{a}_1\| - \|\mathbf{s}_\gamma - \mathbf{a}_j\|}{c},$$

where c is the sound speed, $\mathbf{s}_\gamma$ is the location of the γth source, $\mathbf{a}_1$ location of the reference microphone, and $\mathbf{a}_j$ is the location of the jth microphone. For each pair of microphones, we estimate TDOA calculated by the received signals in the real world according to the GCC-PHAT method proposed in section 2. The estimation of TDOA is denoted as $\bar{\tau}_{\gamma,1j}$, which can be obtained by (2.18) with the received signals. Because of the inherently asynchronous, there is an error between the real TDOA $\tau_{\gamma,1j}$ and the estimation TDOA $\bar{\tau}_{\gamma,1j}$. We add some time offsets on the estimation values to decrease the error.

Suppose that there is an unknown time offset between the reference sensor and each other sensor. Denote o_j as the offset between the jth microphone and the reference sensor. Then, the TDOA with the offset is given as

$$\tau^*_{\gamma,1j} = \bar{\tau}_{\gamma,1j} + o_j.$$

The task of this section is to estimate the offsets and adjust the assessed value to approach the actual value. This section proposes two different methods to solve the offset problem. The first one tries to add the offsets as the additional variables in the optimization problem (2.38), while the second method considers a pre-training procedure.

2.6.1 Method 1

For the first method, we consider the offsets as part of the decision variables and formulate a final optimization problem for the calibration of array configuration.

Therefore, the final calibrated system should be less sensitive to offset errors. If we apply the TDOAs with the offset in the SDP-SOCP relaxation model, the set of linear equation constraints corresponding to the TDOAs become

$$\beta_{\gamma,1} - \beta_{\gamma,j} = c(\bar{\tau}_{\gamma,1j} + o_j), \quad \gamma = 1, \cdots, n, \forall j \in C, j \neq 1.$$

The final optimization problem becomes

$$\begin{aligned}
\min_{\hat{\mathbf{a}}_j, \mathbf{o}, \alpha_{\gamma,j}, \beta_{\gamma,j}} \quad & \sum \alpha_{\gamma,j} \\
\mathbf{s}.t. \quad & \beta_{\gamma,1} - \beta_{\gamma,j} = c(\bar{\tau}_{\gamma,1j} + o_j), \quad \gamma = 1, \cdots, n, \forall j \in C, j \neq 1, \\
& \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_j\| \leq \beta_{\gamma,j}, \qquad \gamma = 1, \cdots, n, \forall j \in C, \\
& \begin{pmatrix} 1 & \beta_{\gamma,j} \\ \beta_{\gamma,j} & \alpha_{\gamma,j} \end{pmatrix} \succeq 0, \qquad \gamma = 1, \cdots, n, \forall j \in C, \\
& \begin{pmatrix} \mathbf{s}_\gamma^T & \mathbf{e}_j^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{s}_\gamma \\ \mathbf{e}_j \end{pmatrix} = \alpha_{\gamma,j}, \qquad \gamma = 1, \cdots, n, \forall j \in C, \\
& \begin{pmatrix} \mathbf{0} & \mathbf{e}_{ij}^T \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{ij} \end{pmatrix} = d_{ij}^2, \quad \forall i, j \in C_l, i > j, l = 1, \cdots, L, \\
& \boldsymbol{W}_{1:d,1:d} = I_d, \\
& \boldsymbol{W} = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{Y} \end{pmatrix} \succeq 0.
\end{aligned} \tag{2.41}$$

By solving the optimization problem, we can obtain the final estimated sensor locations with time offsets.

2.6.2 Method 2

In the second method, we use some data to train the offsets for a system. As the internal random delay is only related to the devices, we can train the delay in advance. For the training set, the signals are recorded by the sensors with known positions. As the purpose of the algorithm is to minimize the errors between the estimated

TDOAs and real TDOAs, for each case in the training set, we can formulate the follows optimization problem

$$\begin{aligned} \min_{\mathbf{o}} \ & \sum (\bar{\tau}_{\gamma,1j} + o_j - \tau_{\gamma,1j})^2 \\ & s.t. -\varepsilon \leq \ o_j \leq \varepsilon \ \forall j \end{aligned} \tag{2.42}$$

where $\mathbf{o} = [o_1, \cdots, o_m]$ and ε is a fixed small value. With the known sensor locations and source locations, the real TDOAs can be calculated by

$$\tau_{\gamma,1j} = \frac{\|\mathbf{s}_\gamma - \mathbf{a}_1\| - \|\mathbf{s}_\gamma - \mathbf{a}_j\|}{c}.$$

The $\bar{\tau}_{\gamma,1j}$ is obtained by the received signals based on GCC method. Then, the problem (2.42) can be solved by any gradient based optimization algorithm. The average offsets of the training set is considered as offsets of the system. Then, the TDOAs with offset can be applied in SDP-SOCP relaxation model.

2.7 Experimental Results

This section gives some results to test the SDP-SOCP relaxation model. Firstly, the mixed model is compared with the SDP model and SOCP model, and the results show that the mixed model outperforms the other two models. Then, a rectangular room is defined for the fast ISM room simulator to calculate the RIRs. Results show the proposed method is a robust in-room simulation model for both 2-D and 3-D space. Finally, an example with real data is given, and the offset algorithm is applied. Results show the efficiency of the proposed algorithms. All the conic programming is solved by SDPT3 [149] software package in Matlab.

2.7.1 Numerical Results

In the first example, assume that we have 2 microphone arrays, and we have 3 microphones in the linear equispaced beamforming array with inter-element distance

5cm. In addition, there are 4 sources with known locations. In this example, we suppose that there is no TDOA estimation error, and the TDOA is directly derived by

$$\|\mathbf{s}_\gamma - \hat{\mathbf{a}}_i\| - \|\mathbf{s}_\gamma - \hat{\mathbf{a}}_j\| = c\tau_{\gamma,ij},$$

where $c = 340m/s$ is the speed of sound in the air.

Firstly, we set source location as $\mathbf{S} = \{(0,\ 0),\ (0,\ 1),\ (1,\ 0),\ (1,\ 1)\}$, and give the true microphones' locations as $\boldsymbol{A}^* = \{(0.75,\ 0.7)(0.8,\ 0.7)(0.85,\ 0.7)(0.35,\ 0.5)(0.4,\ \ 0.5)(0.45,\ \ 0.5)\}$. With $\boldsymbol{A}^*$ and $\mathbf{S}$, we can calculate $\tau_{\gamma,li,kj}$. Then, the SDP-SOCP, SDP, and SOCP models are applied to estimate the sensor locations, respectively. The results are presented in Fig. 2.6(a). It shows that SOCP and the mixed SDP-SOCP can give the correct estimates, while the SDP method can not. It can be seen that the two folds SDP relaxation is weak. In addition, the rank of $\hat{\boldsymbol{W}}$ from the SDP-SOCP solution is 2 ($d = 2$), and the numerical results show the optimal solution after relaxation is the same as the original problem. It is accordant with Lemma 2.5.

Then, we move one of the array outside the convex hull slightly, and the new microphone location matrix is $\boldsymbol{A}^* = \{(0.75,\ \ 0.7)(0.8,\ \ 0.7)(0.85,\ \ 0.7)(-0.25,\ 0.5)(-0.2,\ \ 0.5)(-0.15,\ \ 0.5)\}$, and the results are given in Fig.2.6(b). As the localization region of the SOCP relaxation model is the convex hull formulated by the known sources, in this example, the SOCP relaxation can either not give an exact estimation. Then, we move the outside array further away from the convex hull to $\{(-2.25, 0.5)(-2.2, 0.5)(-2.15, 0.5)\}$, and results show that the SDP-SOCP model can also give the exact location. According to the above all, we can see that the mixed SDP-SOCP relaxation model has a larger localizable region. Table 3.2 illustrates the estimation errors of these three models.

To investigate the model better, we further move both of the arrays outside the

convex hull and set the microphone location as $\boldsymbol{A}^* = \{(0.75,\ 1.2)(0.8,\ 1.2)(0.85,\ 1.2)(-0.25,\ 0.4)(-0.2,\ 0.4)(-0.15,\ 0.4)\}$. In this situation, the rank of the optimal solution (SDP-SOCP) is $8 \neq 2$. From Fig 2.6(d), we can observe that there is an error between the true location and the estimated location, which is also in accord with Lemma 2.5. Besides, we can see that all the microphones are not localizable, and it follows the Lemma 2.3.

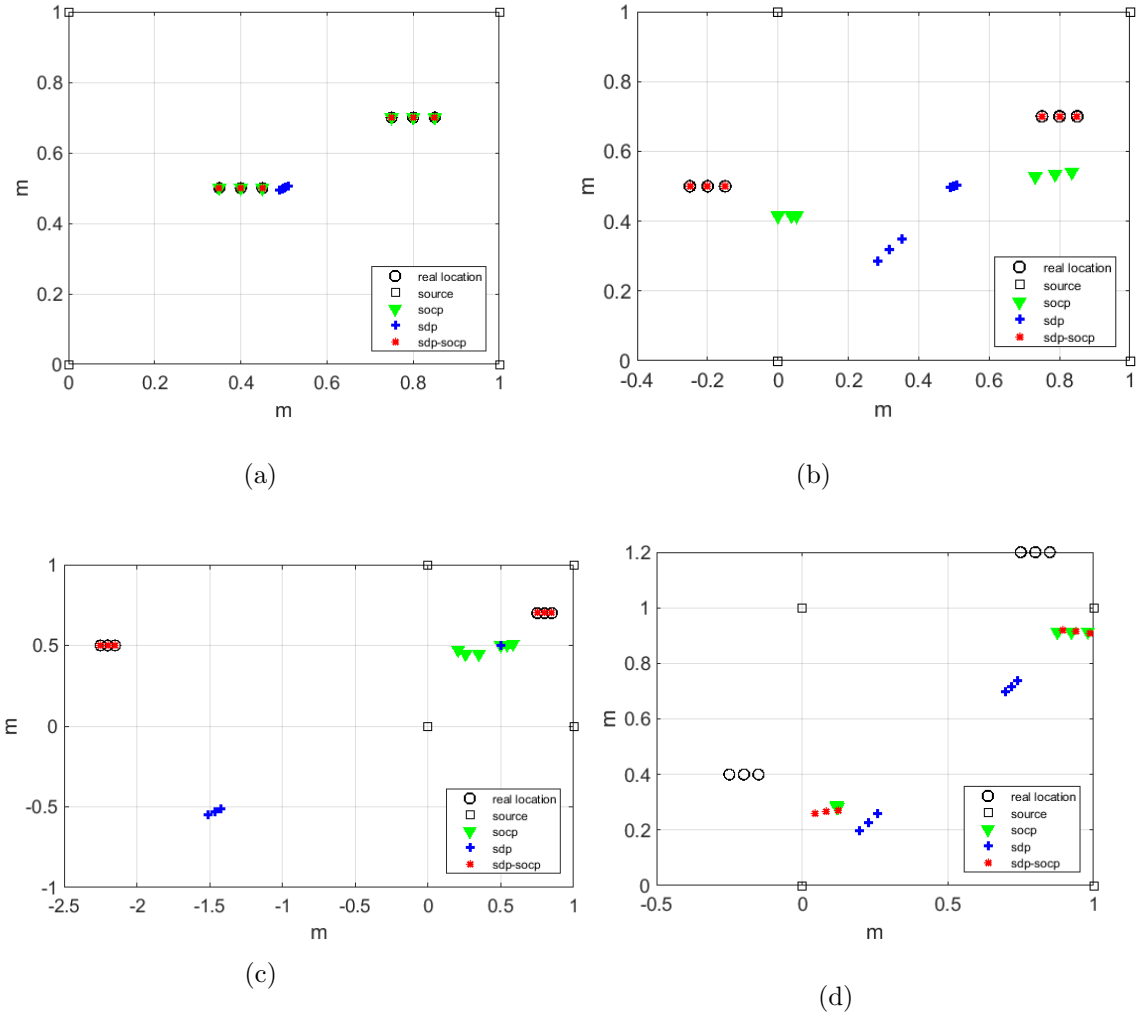


Figure 2.6: Comparison with different methods

Table 2.1: Errors of different methods

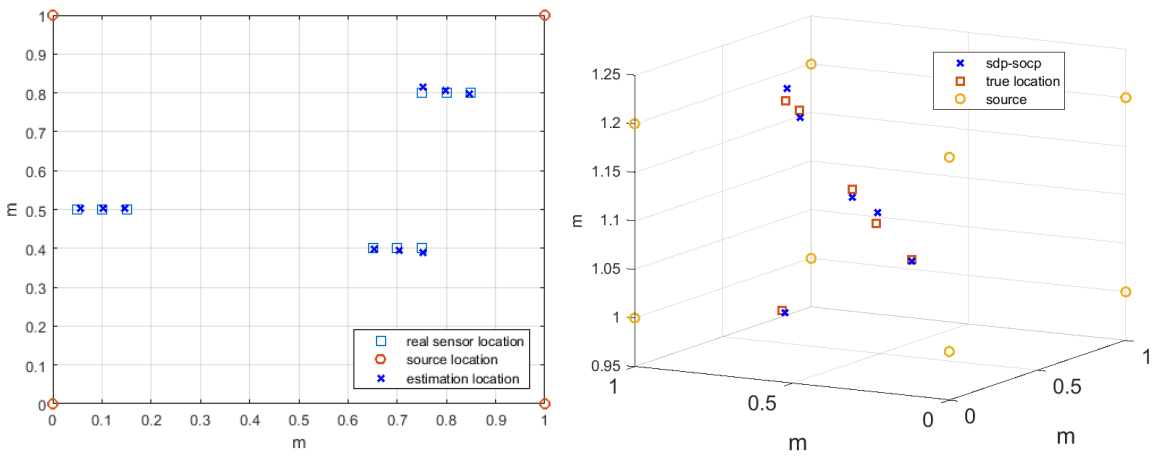
Methods	SDP	SOCP	SDP-SOCP
Error (a)	1.80	4.49e-14	3.61e-15
Error (b)	3.60	1.60	2.47e-14
Error (c)	6.79	8.91	1.14e-6
Error (d)	3.49	2.56	2.40

2.7.2 Experiments in simulated rooms

In this subsection, we consider the room acoustics. Suppose that there are four sources located at $\{(0,\ 0),\ (0,\ 1),\ (1,\ 0),\ (1,\ 1)\}$ and three linear arrays. Each of them contains 3 elements with inter-element distance 5cm. True locations of the three sensor arrays are $\{(0.75,\ 0.80),\ (0.80,\ 0.80),\ (0.85,\ 0.80),\ (0.05,\ 0.50),\ (0.10,\ 0.50),\ (0.15,\ 0.50),\ (0.65,\ 0.40),\ (0.70,\ 0.40),\ (0.75,\ 0.40)\}$. The audio data is generated from the reflection impulse responses (RIRs) generator using the image method in [85]. The simulation reverberant room is $4m \times 4m \times 3m$, and the reverberation time $T_{60} = 0.3$. Firstly, we generate a set of synthetic RIRs by the image source method technique. According to the GCC method, we can obtain τ by (**??**). Then, by solving the problem (2.38), we can calibrate the locations of the microphones. The result is illustrated in Fig. 2.7(a), and the absolute error sum between the estimated locations and real locations is 0.07.

Furthermore, we consider a 3-dimension example in a simulated room. Suppose that there are 8 known sources and 6 unknown microphones. The locations of sources are $\{(0,\ 0,\ 1),\ (0,\ 1,\ 1),\ (1,\ 0,\ 1),\ (1,\ 1,\ 1),\ (0,\ 0,\ 1.2),\ (0,\ 1,\ 1.2),\ (1,\ 0,\ 1.2),\ (1,\ 1,\ 1.2)\}$, and the microphones are located at $\{(0.7,\ 0.7,\ 1.1),\ (0.5,\ 0.4,\ 1.05),\ (0.4,\ 0.7,\ 1.2),\ (0.3,\ 0.4,\ 1.1),\ (0.5,\ 0.8,\ 1.2),\ (0.3,\ 0.7,\ 1)\}$. We consider the room reflection with the same room settings as in the 2-D example. By solving the problem (2.38), we can get the estimated location matrix of rank 3, which equals the network's dimension. The simulation results with room reflections are shown in Fig.

2.7(b), and the corresponding error is 0.13.



(a) 2-D example (b) 3-D example

Figure 2.7: The results of room simulation

2.7.3 Experiments with Real Data

In this example, we further investigate the application of SDP-SOCP model in real data. For real data collecting, Microsemi's audio processing eval board is used with a ZLE38000 model, containing four microphones and a USB interface converter, and a Raspberry Pi applied to receive signals. The audio signals are recorded with the program Audacity on a MacBook Pro with 3GHz Intel Core i7 processor. The experiment is conducted in a room with $3.75 \times 7.5 \times 2.75$m, and the total number of sensors used for recording is 4. The four microphones are divided into two groups and each group contains two microphones. The experiments contain three different configurations, depicted in Fig. 2.9. The recording setup is given in Fig. 2.8.

The estimated TDOA $\bar{\tau}$ is calculated by the real signals received by the sensors, and the algorithm used for estiamting the TDOA is GCC method. To improve the estimation accuracy, the proposed TDOA offset algorithms are applied. The errors of the two different offset algorithms are compared with the model without the offsets.

The results are given in Table 2.2, where E_{TDOA} is the sum of the square errors between the real TDOAs and the estimated TDOAs, and E_{mics} represents the sum of absolute error between the real sensor locations and estimated sensor locations. For offset 1, we consider the offsets as part of decision variable, and the estimated sensor locations are achieved by solving problem (2.42). For offset 2, when estimating one configuration, we consider the other two configurations as the training set. For example, when calibrating configuration 1, the configuration 2 and 3 are considered as the training set. The offsets for configuration 1 are the average value of the training set. Clearly, after the offset algorithms, the sum of errors decreases. The locations are given in Fig. 2.10.

Table 2.2: Errors of different configuration

Configuration	Without offset		Method 1		Method 2	
	E_{TDOA}	E_{mics}	E_{TDOA}	E_{mics}	E_{TDOA}	E_{mics}
1	0.1308	0.5709	0.1151	0.4394	0.1178	0.4438
2	0.0597	0.2143	0.0176	0.1798	0.0432	0.1861
3	0.0610	0.3698	0.0405	0.3162	0.0519	0.3224

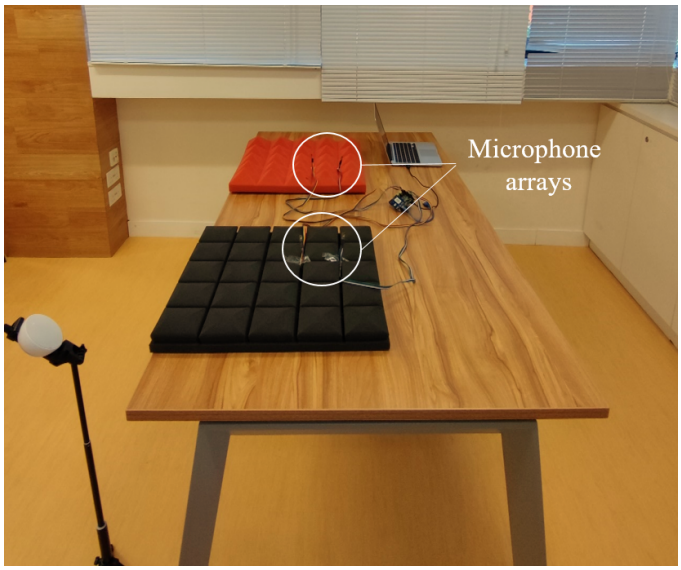
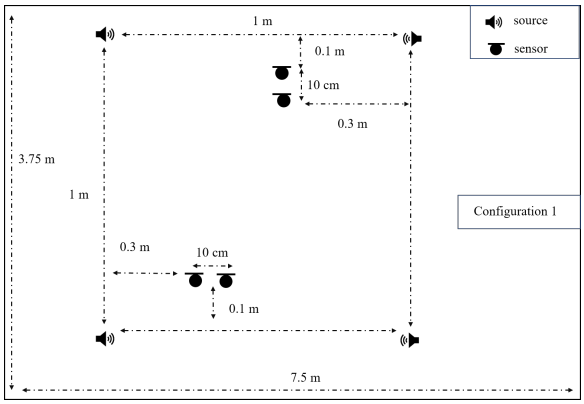
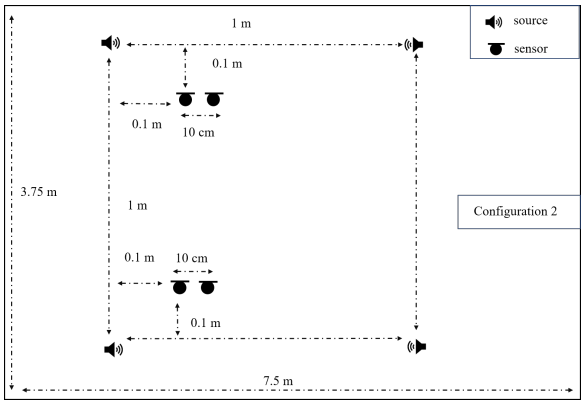


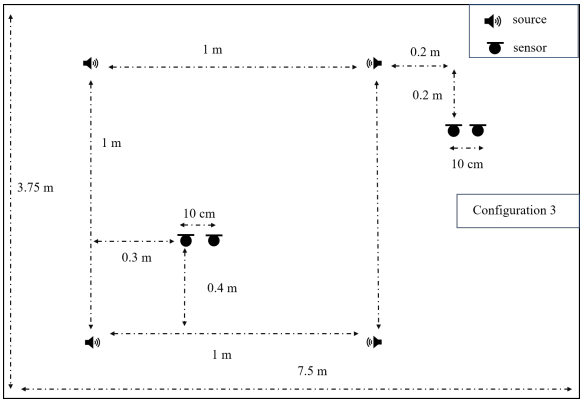
Figure 2.8: The recording setup



(a) Configuration 1

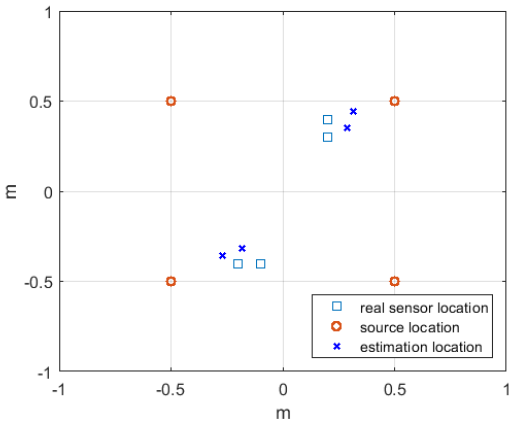


(b) Configuration 2

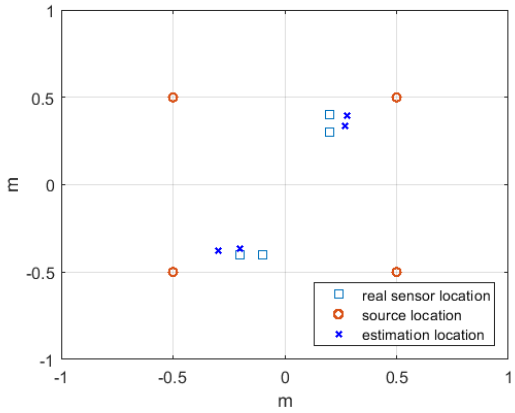


(c) Configuration 3

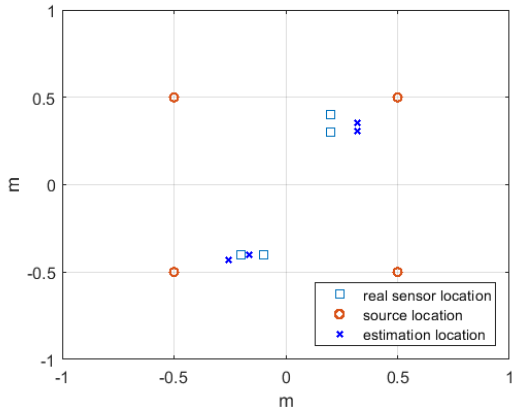
Figure 2.9: The configurations of real data



(a) Without offset

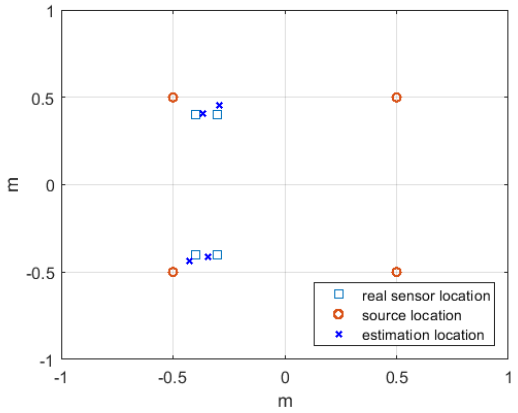


(b) Method 1

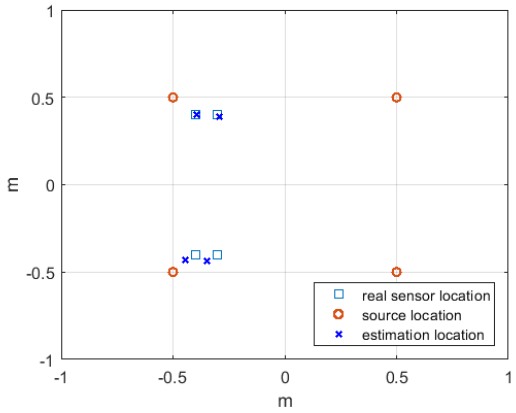


(c) Method 2

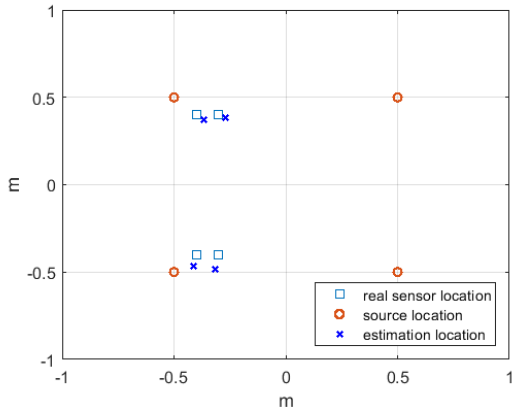
Figure 2.10: The results with real data C1



(a) Without offset

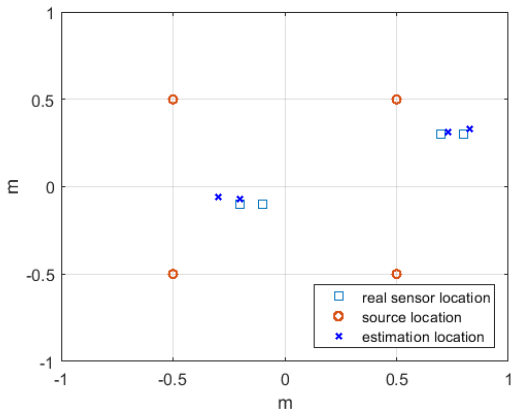


(b) Method 1

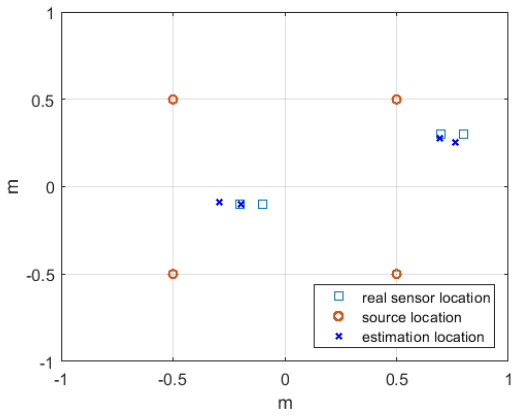


(c) Method 2

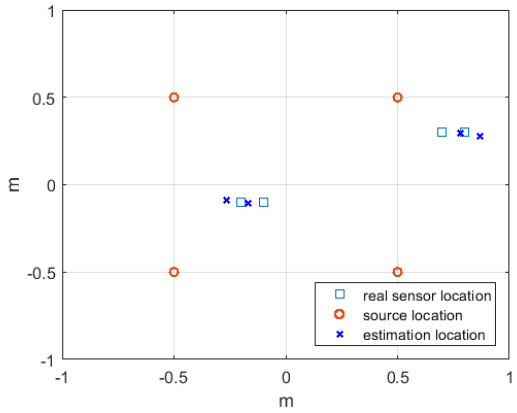
Figure 2.11: The results with real data C2



(a) Without offset



(b) Method 1



(c) Method 2

Figure 2.12: The results with real data C3

Chapter 3

Design of Near-Field Broadband Beamformer Based on IIR Filters

3.1 Introduction

Assuming that the array configuration is given, beamforming algorithms can be performed over the network. This chapter motivates the use of infinite impulse response (IIR) filters in near-field broadband beamformer design in wireless sensor networks. The main idea of beamforming is to enhance the audio signal from the desired parts and reduce the undesired elements. Many conventional near-field broadband beamforming models are achieved by an FIR filter attached to each channel [69, 43]. The general approach is related to a multidimensional digital filter design problem with an arbitrarily specified amplitude and phase. It performs on a discrete domain where frequency and spatial domains are discretized into a finite number of grid points. Then, it can be solved by a linear programming technique. Motivated by the desire to decrease the number of coefficients, a new structure of tapped delay line processor with both feedback and feedforward digital filtering, known as IIR filter, is proposed in this chapter. Compared to the FIR filter, it is shown that the optimal frequency dependent array weighting could be more efficiently approximated by IIR filters in broadband beamformers [54, 37, 131]. In the proposed method, this

improvement in efficiency is also shown to be true for the case of nearfield fixed array processing. Furthermore, we analyse the performance limit of the proposed beamformer. The performance limit will be reached with fewer coefficients when IIR filters are used compared to FIR. This results a decrease in the computational load in the implementation.

This chapter motivates the use of IIR filters in beamformer design. Section 3 formulates the problem. Since the IIR filters contain feedback sections, they may have stability problems. Section 4 addresses the stability problem by decomposing the direct form into a sum of partial fractions with low orders. Section 5 gives the algorithms to obtain the optimal solution to the formulated optimization problem. Section 6 proposes a specific structure to simplify the stability problem, in which all the filters share the same poles. Section 7 analyzes the limited performance of the proposed beamforming. Finally, we give two examples to verify the performance of the beamforming designs. From those examples, it can be seen that the results follow the theory. Example 1 shows that the proposed method reduces the number of parameters to half compared to FIR filters when approaching the same limit performance.

3.2 Problem Statement

Assume that there are m sensors distributed in an acoustic network, and after each sensor there is an Infinite-Impulse-Response (IIR) filter with an M order denominator and $N-1$ order numerator. The frequency response of these IIR filters can be defined as

$$R_k(f) = \frac{H_k(f)}{W_k(f)} = \frac{\mathbf{h}_k^{\mathsf{T}} \boldsymbol{d}_0(f)}{1 + \mathbf{w}_k^{\mathsf{T}} \boldsymbol{d}_1(f)}, \quad k = 1, \cdots, m, \tag{3.1}$$

where

$$\begin{aligned}
\mathbf{h}_k &= [h_k(0), h_k(1), \cdots, h_k(N-1)]^{\mathsf{T}} \\
\mathbf{w}_k &= [w_k(1), \cdots, w_k(M)]^{\mathsf{T}} \\
\boldsymbol{d}_0(f) &= [1, e^{\frac{-j2\pi f}{fs}}, \cdots, e^{\frac{-j2\pi f(N-1)}{fs}}]^{\mathsf{T}} \\
\boldsymbol{d}_1(f) &= [e^{\frac{-j2\pi f}{fs}}, \cdots, e^{\frac{-j2\pi f(M)}{fs}}]^{\mathsf{T}},
\end{aligned}$$

and the coefficients matrices are defined as

$$\begin{aligned}
\mathbf{h} &= [\mathbf{h}_1^{\mathsf{T}}, \mathbf{h}_2^{\mathsf{T}}, \cdots, \mathbf{h}_m^{\mathsf{T}}]^{\mathsf{T}}, \\
\mathbf{w} &= [\mathbf{w}_1^{\mathsf{T}}, \mathbf{w}_2^{\mathsf{T}}, \cdots, \mathbf{w}_m^{\mathsf{T}}]^{\mathsf{T}}.
\end{aligned}$$

Denote the corresponding transfer function in frequency domain from the source position $\mathbf{r}$ to the kth microphone as $A_k(\mathbf{r}, f)$, suppose the array response vector $\boldsymbol{A}(\mathbf{r}, f) = [A_1(\mathbf{r}, f), \cdots, A_m(\mathbf{r}, f)]^{\mathsf{T}}$, where T means the transposition. Then, we can obtain the actual model as

$$G(\mathbf{r}, f) = \sum_{k=1}^{m} R_k(f) A_k(\mathbf{r}, f), \tag{3.2}$$

and the structure of the beamformer is depicted in Fig.3.1. For the transfer function depicting sound wave propagation $A_k(\mathbf{r}, f)$, if a simple spherical free field model is applied, the transfer function in the free field is written as

$$A_k(\mathbf{r}, f) = \frac{1}{\|\mathbf{r} - \mathbf{r}_k\|} e^{\frac{-j2\pi f \|\mathbf{r} - \mathbf{r}_k\|}{c}}, \tag{3.3}$$

where c is the sound speed in the air, $\mathbf{r}_k$ is the position of the kth sensor. However, in the case of a reverberant environment, this free field model can not estimate the complicated propagation accurately. Thus, for indoor beamformer design, the image-source method (ISM) is applied, where the room impulse response (RIRs)

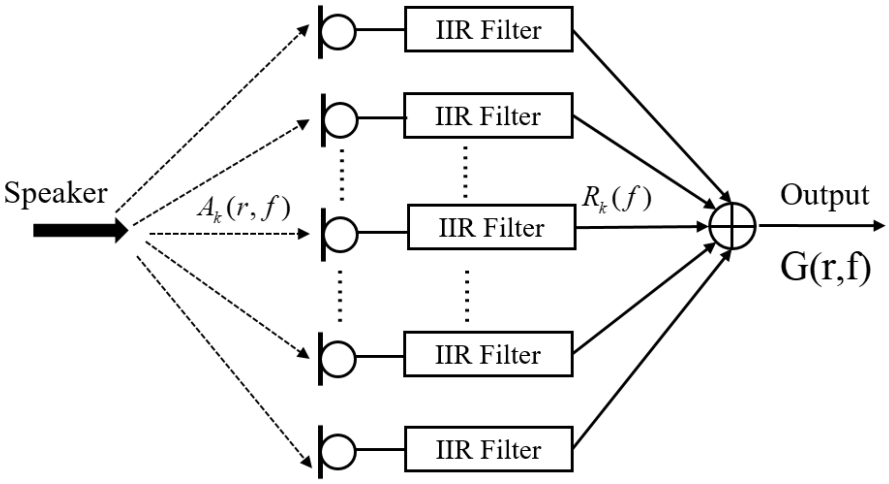


Figure 3.1: Beamforming structure applied IIR filters

are calculated by an efficient room simulator developed in [85]. The RIRs measure responses in the design, and give $\boldsymbol{A}(\mathbf{r}, f)$.

Denote the specific desired response of the microphone system as $G_d(\mathbf{r}, f)$ in frequency and space. Give a space-frequency region Ω as the definition field of $G_d(\mathbf{r}, f)$, which contains two parts, stop region Ω_s and pass region Ω_p. We can set the pass and stop region according to real applications. The main task of this problem is to find the suitable coefficients such that the actual response $G(\mathbf{r}, f)$ fits a desired response $G_d(\mathbf{r}, f)$ optimally. The error between the actual and desired response is defined as

$$E(\mathbf{h}, \mathbf{w}, \mathbf{r}, f) = \alpha(\mathbf{r}, f)|G(\mathbf{r}, f) - G_d(\mathbf{r}, f)|^2, \tag{3.4}$$

where α is a positive weighting function to measure the importance of pass region and stop region. As there are feedback sections, we have to consider the stability problem. Denote $S^{'} = \{\mathbf{w} : \mathbf{w}_k(f)\ is\ stable, k = 1, \cdots, m\}$. Then, we can formulate a minimax beamforming problem as

$$\min_{\mathbf{w} \in S^{'}, \mathbf{h} \in R^{NP}} \max_{(\mathbf{r}, f) \in \Omega} E(\mathbf{h}, \mathbf{w}, \mathbf{r}, f). \tag{3.5}$$

The minimax problem can be transformed into an equivalent semi-infinite problem

constituting infinite inequality constraints as

$$\begin{aligned} &\min_{\mathbf{w}\in S',\mathbf{h}\in R^{NP},\delta\in R^{+}} \quad \delta \\ &s.t.\ E(\mathbf{h},\mathbf{w},\mathbf{r},f)-\delta\leq 0,\ \forall(\mathbf{r},f)\in\Omega. \end{aligned} \tag{3.6}$$

By solving the optimization problem, we can obtain the parameters of the filters. However, two problems still need to be considered. The first problem is how to formulate a stable region. As the stability region is hard to be obtained directly, we need to decompose it into different partial fractions. The other problem is how to solve the optimization problem. The problem (3.6) is nonconvex, and some global optimization algorithms should be applied. The following two sections will tackle these two problems.

3.3 Stability Condition

The stability problem of the beamformer was raised from the feedback sections of the IIR filters. Thus, the beamformer is stable only if the IIR filters satisfy the stability condition, which means we only need to guarantee that the frequency response of the IIRs are stable. Then, the problem becomes how to find the stability condition. The sufficient and necessary stable condition for the IIR filter is that all the filter poles should be inside the unit circle [122]. If the poles of IIR filters are placed outside the unit circle, the system is unstable. Therefore, the stability domain must be determined by solving all the poles. When the denominator is a polynomial of order 1 or 2, we can get the poles directly. However, when the order is large, the poles cannot be obtained easily with a transfer function given in (3.34). It is difficult to formulate and solve the optimization problem. One efficient method to solve the problem is decomposing the direct form into first and second-order sections. In the cascaded form, the poles of filters can be easily obtained, and coefficient quantization

errors also reduced that occur because of the using a finite number of bits to represent the filter coefficients. In the following part, we discuss the stability condition for these two cases.

3.3.1 A Root of Multiplicity 1 or 2

For each channel k, the frequency response of the IIR is given as

$$R_k(f) = \frac{H_k(f)}{W_k(f)}.$$

When the denominator is a polynomial of order 1 or 2, it is much easier to obtain the poles. In terms of the first order case, the feedback coefficient is the unique pole directly. For the stability requirement, we must restrict the feedback coefficients between a range from -1 to 1, that is

$$S^{'} = \{\mathbf{w} : -1 < w_k(1) < 1;\ \forall k = 1, \cdots, m\}. \tag{3.7}$$

If the order of feedback section is 2, we can obtain from [120] that the stable domain is

$$S^{'} = \{\mathbf{w} : 1+w_k(1)+w_k(2) > 0; 1-w_k(1)+w_k(2) > 0; 1-w_k(2) > 0; \forall k = 1, \cdots, m\}. \tag{3.8}$$

Denote

$$\mathbf{D} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}; \qquad \mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then, (3.8) is equivalent to

$$S^{'} = \{\mathbf{w} :\ \mathbf{D}\mathbf{w}_k < \mathbf{e};\ \forall k = 1, \cdots, m\}. \tag{3.9}$$

3.3.2 General Roots

However, when the order of denominator is greater than 2, the stable domain of an IIR filter in a direct form is exceptionally complicated. Thus, we consider to decompose the direct form into a sum of partial fractions with lower orders. A decomposition of rational polynomial depends on the poles of the denominator. Then, for the polynomial where the coefficients are real numbers, each pole or each pair of pole can have four different situations.

(i) Single Real Pole

If there is only one single real pole z_0, then the decomposition must have a term $\frac{A}{z-z_0}$. If there are at least two single real poles z_1 and z_2, the decomposition have two terms $\frac{A_1}{z-z_1}$ and $\frac{A_2}{z-z_2}$. We can rewrite it as

$$\frac{cz+d}{z^2+az+b} = \frac{A_1}{z-z_1} + \frac{A_2}{z-z_2} \tag{3.10}$$

where $a = -z_1 - z_2$, $b = z_1 z_2$, $c = A_1 + A_2$, $d = -A_1 z_2 - A_2 z_1$.

(ii) Double Real Pole

If there is a double real pole z_0, then the decomposition must have the terms $\frac{A_1}{z-z_0} + \frac{B_0}{(z-z_0)^2}$, this can be rewritten as

$$\frac{cz+d}{z^2+az+b} = \frac{A_1}{z-z_0} + \frac{B_0}{(z-z_0)^2} \tag{3.11}$$

where $a = -2z_0$, $b = z_0^2$, $c = A_1$, $d = B_0 + A_1 z_0$.

(iii) Single Complex Pole

Note that the coefficients of the polynominal are real, the complex pole appear conjugately, which means the decomposition must contain the terms $\frac{A}{z-(\alpha+\beta j)}$

and $\frac{A}{z-(\alpha-\beta j)}$. By combining the conjugate poles, we can get the combined part:

$$\frac{cz+d}{z^2+az+b} = \frac{A}{z-(\alpha+\beta j)} + \frac{A}{z-(\alpha-\beta j)} \tag{3.12}$$

where $a=-2\alpha$, $b=\alpha^2+\beta^2$, $c=2A$, $d=-2A\alpha$.

(iv) Repeated Pole

If there exist repeated real pole z_0 of order $M_0(M_0 \geq 3)$, then except the terms in (i) and (ii), there are serveral terms in the decomposition of filters

$$\frac{A_1}{(z-z_0)^{M_0}} + \cdots + \frac{A_{N-2}}{(z-z_0)^3}$$

Similarly, if there exists repeated complex poles of order $M_1(M_1 \geq 2)$, then there are several terms in the decomposition as follows

$$\frac{c_1 z+d_1}{(z^2+az+b)^2} + \cdots + \frac{c_{n-1}z+d_{N-1}}{(z^2+az+b)^N} \tag{3.13}$$

Hence, we can decompose the frequency response of the IIR, denoted by $R_k(f)$, for each channel k in the following possible cases.

(1) The multiplicities of all the poles are in the cases of (i), (ii), (iii).

In this case, the function $R_k(f)$ can be decomposed into a uniform equation

$$R_k(f) = \sum_{l=1}^{M/2} \frac{b_{l1}^k z + b_{l2}^k}{z^2+\alpha_l^k z+\beta_l^k} + \sum_{l=0}^{N-M} c_l^k z^l \tag{3.14}$$

when M is even, and

$$R_k(f) = \frac{a_1^k}{z-d_1^k} + \sum_{l=1}^{(M-1)/2} \frac{b_{l1}^k z + b_{l2}^k}{z^2+\alpha_l^k z+\beta_l^k} + \sum_{l=0}^{N-M} c_l^k z^l \tag{3.15}$$

when M is odd.

(2) If there exists an M_0-repeated real pole a_0 of $R_k(f)$ ($M_0 \geq 3$). Besides the pole a_0, there are $M - M_0$ times of the other poles. Then, if the $M - M_0$ times of the poles are the case of (a) above,the decomposition is also the same. Hence, the decomposition is

$$R_k(f) = \sum_{l=0}^{N-M-1} c_l^k z^l + H_1(z) + r(z), \tag{3.16}$$

where

$$H_1(z) = \sum_{l=1}^{(M-M_0)/2} \frac{b_{l1}^k z + b_{l2}^k}{z^2 + \alpha_l^k z + \beta_l^k} + \sum_{l=0}^{N-M-1} c_l^k z^l \tag{3.17}$$

when $M - M_0$ is even, and

$$H_1(z) = \frac{a_1^k}{z - d_1^k} + \sum_{l=1}^{(M-M_0-1)/2} \frac{b_{l1}^k z + b_{l2}^k}{z^2 + \alpha_l^k z + \beta_l^k} + \sum_{l=0}^{N-M-1} c_l^k z^l \tag{3.18}$$

when $M - M_0$ is odd. In addition, $r(z)$ is given by

$$r(z) = \frac{b_1 z + b_2}{z^2 - 2dz + d^2} + \sum_{l=3}^{M_0} \frac{b_k}{(z-d)^l}. \tag{3.19}$$

(3) If there are M_1-repeated ($M_1 \geq 2$) complex conjugate poles z_0 and $\bar{z}_0$ of $R_k(f)$. Besides, the poles z_0 and $\bar{z}_0$, there are $M - 2M_1$ times of the poles, which is similar to the situation (2). The decomposition is the same as (3.16), where $H_1(z)$ is shown as (3.17) or (3.18) and $r(z)$ is given by

$$r(z) = \sum_{l=2}^{M_1} \frac{b_{1l-1} z + b_{1l}}{(z^2 + 2\alpha z + \beta)^l}. \tag{3.20}$$

As the order of denominator m increases, the number of the decompositions increases. Hence, it is required to simplify the problem. One way is to approximate higher order terms suitably. We have the following lemma.

Lemma 3.1. *For any transfer function* $Q(z) = \frac{A}{(z-a_0)^{M_0}}$ $(M_0 \geq 3)$*, where* $(|a_0| < 1)$ *is a real number, then* $\forall \varepsilon > 0$*, there is a stable* $P(z)$ *satisfies* $|Q(z) - P(z)| < \varepsilon$*, where* $P(z) = \frac{1}{(z-a_0)^{M_0-\delta}}$ *and the multiplicity of* $P(z)$ *is 1.*

The proof of the lemma are given in [120]. It can be seen that the multiplicities of poles of $P(z)$ are 1, which can be decomposed into the following form

$$P(z) = \sum_{r=1}^{M_0/2} \frac{b_{r1}z + b_{r2}}{z^2 + \alpha_r z + \beta_r} \tag{3.21}$$

when M_0 is even and

$$P(z) = \sum_{r=1}^{(M_0-1)/2} \frac{b_{r1}z + b_{r2}}{z^2 + \alpha_r z + \beta_r} + \frac{a_1}{z - d_1} \tag{3.22}$$

when M_0 is odd.

Based on the results from Lemma 1, we can easily obtain the following results.

Corollary 3.1. *For any transfer function* $Q(z) = \frac{bz+c}{(z^2+\alpha z+\beta)^{M_1}} (M_1 \geq 2)$*, where* α, β *satisfy the stability triangle in (3.8), there is a* $P(z)$ *satisfies* $|Q(z) - P(z)| < \varepsilon$*.*

By Lemma 3.1 and Corollary 3.1, we have the following conclusion that for any $\delta > 0$, and any decomposition of $R_k(f)$ which contains $M_0(M_0 \geq 3)$ repeated real roots or $M_1(M_1 \geq 2)$ complex roots, there is a decomposition of first kind which approached to $R_k(f)$. It means that we can always solve the first kind of decomposition to find the optimal solution and the other kind of decompositions are unnecessary. Hence, we simplify the problem by only considering one subproblem. We have the following two theorems to summarize.

Theorem 3.1. *For any* $\varepsilon > 0$*, and any* $R_k(f)$ *which contains* $M_0(M_0 \geq 3)$ *repeated real roots or* $M_1(M_1 \geq 2)$ *complex roots, there exists* $\bar{R}_k(z)$ *which is of the form (3.14) or (3.15), such that*

$$|R_k(f) - \bar{R}_k(f)| < \varepsilon, \quad \forall z.$$

Then, for each channel k, we have the parallel form of $R_k(f)$ as

$$\bar{R}_k(f) = \begin{cases} \sum_{l=1}^{M/2} \dfrac{b_{l1}^k z^{-1} + b_{l0}^k}{1 + \alpha_l^k z^{-1} + \beta_l^k z^{-2}}, & \text{if } N-1 < M \\ \sum_{l=1}^{M/2} \dfrac{b_{l1}^k z^{-1} + b_{l0}^k}{1 + \alpha_l^k z^{-1} + \beta_l^k z^{-2}} + \sum_{l=0}^{N-M} c_l^k z^{-l}, & \text{if } N-1 \geq M \end{cases} \tag{3.23}$$

when M is even, and

$$\bar{R}_k(f) = \begin{cases} \dfrac{a_1^k}{1 - d_1^k z^{-1}} + \sum_{l=1}^{(M-1)/2} \dfrac{b_{l1}^k z^{-1} + b_{l0}^k}{1 + \alpha_l^k z^{-1} + \beta_l^k z^{-2}}, & \text{if } N-1 < M \\ \dfrac{a_1^k}{1 - d_1^k z^{-1}} + \sum_{l=1}^{(M-1)/2} \dfrac{b_{l1}^k z^{-1} + b_{l0}^k}{1 + \alpha_l^k z^{-1} + \beta_l^k z^{-2}} + \sum_{l=0}^{N-M-1} c_l^k z^{-l}, & \text{if } N-1 \geq M \end{cases} \tag{3.24}$$

when M is odd.

The original feedback part becomes a cascade of several first or second-order parts. The whole feedback section is stable if and only if every low order part is stable. Then, based on the stable condition of low order cases in (3.7) and (3.9), we can obtain the stability domain as

$$S' = \{(\alpha_l^k, \beta_l^k) : \mathbf{D} \begin{pmatrix} \alpha_l^k \\ \beta_l^k \end{pmatrix} < \mathbf{e},\ l = 1, \cdots, \frac{M}{2},\ k = 1, \cdots, m\} \tag{3.25}$$

when M is even and

$$S' = \{(d_1^k, \alpha_l^k, \beta_l^k) : -1 < d_1^k < 1; \mathbf{D} \begin{pmatrix} \alpha_l^k \\ \beta_l^k \end{pmatrix} < \mathbf{e},\ l = 1, \cdots, \frac{M-1}{2},\ k = 1, \cdots, m\} \tag{3.26}$$

when M is odd.

Then, for a given $\bar{R}_k(f)$, where the coefficients are $\mathbf{h}_k$, $\mathbf{w}_k$, there are the corresponding coefficients a_1^k, b_{l0}^k, b_{l1}^k, c_l^k, d_1^k, α_l^k, β_l^k such that $\bar{R}_k(f)$ approaches to $R_k(f)$.

Hence, we denote a vector $\mathbf{y} = \{a_1^k,\ b_{l0}^k,\ b_{l1}^k,\ c_l^k,\ d_1^k\},\ k = 1, \cdots, m$ as a stack of all these variables. The transfer function becomes

$$G_s(\mathbf{r}, f) = \sum_{k=1}^{m} \bar{R}_k(f) A_k(\mathbf{r}, f), \tag{3.27}$$

where $\bar{R}_k(f)$ is given by (3.23) and (3.24). The error between the real actual and desired response becomes

$$E_s(\mathbf{y}, \mathbf{r}, f) = \alpha(\mathbf{r}, f)|G_s(\mathbf{r}, f) - G_d(\mathbf{r}, f)|^2. \tag{3.28}$$

We can transfer the problem (3.5) into the problem below.

$$\begin{aligned} &\min_{\mathbf{y} \in S', \delta \in R^+} \quad \delta \\ &s.t.\ E_s(\mathbf{y}, \mathbf{r}, f) - \delta \leq 0,\ \forall (r, f) \in \Omega. \end{aligned} \tag{3.29}$$

Problem (3.29) can be optimized directly by any gradient-based algorithm. After $\mathbf{y}$ is obtained by solving (3.29), the corresponding $\mathbf{h}$ and $\mathbf{w}$ can also be obtained by unifying all the partial fractions into one fraction. Then, the stability problem is well solved. The remaining issue is how to find the global optimal solution to the problem (3.29).

3.4 Optimization Algorithms

This section considers the algorithms use to solve the optimization problem. As problem (3.29) is a smooth nonlinear constrained optimization problem, it can be solved by any gradient-based algorithm. Sequential quadratic programming (SQP) is an efficient one to solve this kind of problem with a fast convergence rate. Based on the works of Biggs [14], Han [59] and Powell [118], an SQP method mimics Newton's method for constraint optimization. It is an iterative method of starting from some initial point and converging to a constrained local minimum. At each iteration, one

obtains search directions from a quadratic program (QP) that is a quadratic model of a specific Lagrangian function subject to the constraints. We consider the unknown of Problem (3.29) in one vector

$$\mathbf{x} = \begin{bmatrix} \delta \\ \mathbf{y} \end{bmatrix}.$$

We can rewrite the objective function and the constraints as

$$\begin{aligned} f(\mathbf{x}) &= \delta \\ c(\mathbf{x}) &= \delta - E_s(\mathbf{y}, \mathbf{r}, f). \end{aligned}$$

The Lagrangian function associated with problem (3.29) is defined as

$$L(\mathbf{x}, \lambda) \triangleq f(\mathbf{x}) - \lambda^T c(x). \tag{3.30}$$

Then, the SQP search direction $d_{(i)}$ for iteration i can be calculated by solving a sub-problem described as

$$\begin{aligned} \min_{d} \quad & d_{(i)} + \frac{1}{2} d_{(i)}^T B_{(i)} d_{(i)}, \\ s.t. \quad & c(\mathbf{x}_{(i)}) + \nabla c(\mathbf{x}_{(i)})^T d_{(i)} \geq 0, \end{aligned} \tag{3.31}$$

where ∇ means gradient operation, and B is an approximation of the Hessian of the Lagrangian function. For the calculation of B, we can consider the Quasi-Newton method. The related works have been proposed in [17] [27], and research shows that a simple Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is efficient in practice. In this problem, we use a BFGS approximation, and a BFGS update for each iteration t is given as

$$B_{(i+1)} = B_{(i)} + \frac{\mathbf{p}_{(i)} \mathbf{p}_{(i)}^T}{\mathbf{p}_{(i)}^T \mathbf{s}_{(i)}} - \frac{B_{(i)} \mathbf{s}_{(i)} \mathbf{s}_{(i)}^T B_{(i)}^T}{\mathbf{s}_{(i)}^T B_t \mathbf{s}_{(i)}}, \tag{3.32}$$

where $\mathbf{p}_{(i)} = \nabla L(\mathbf{x}_{(i+1)}, \lambda_{(i+1)}) - \nabla L(\mathbf{x}_{(i)}, \lambda_{(i)})$ and $\mathbf{s}_{(i)} = \mathbf{x}_{(i+1)} - \mathbf{x}_{(i)}$. Then, for iteration i, we have

$$\mathbf{x}_{(i+1)} = \mathbf{x}_{(i)} + \alpha d_{(i)},$$

where α is the step length.

The pseudo-code of the SQP algorithm is given in Algorithm 1. The step length parameter α is required to enforce global convergence of the SQP method, and it is usually chosen to satisfy a certain Armijo [2] condition. It guarantees a reduced direction at each step of a procedure. Besides, κ is used to estimate whether the current solution is convergent. The algorithm stops if sufficient descent is not observed after a certain number of iterations. If the tested stepsize falls below machine precision or the accuracy by which model function values are computed, the merit function cannot decrease further.

Algorithm 1 SQP Algorithm

Require: Choose an initial guess $\mathbf{x}_{(0)}$, and initial matrix B_0. Set $i = 0$.
Ensure: $\mathbf{y} \in S^{'}$ and $\delta \geq 0$

1: Compute the optimum update $d_{(t)}$ by solving QP problem (3.31).
2: Set $\mathbf{x}_{(i+1)} = \mathbf{x}_{(i)} + \alpha d_{(i)}$, with $\alpha \in (0, 1)$, which is a suitable steplength parameter. Update $B_{(i+1)}$ according to (3.32).
3: If $\| \mathbf{x}_{(i)} \|_2 \ / \ \| \mathbf{x}_{(i+1)} \|_2 < \kappa$, stop, where $\|\mathbf{x}_{(i)}\|_2$ is the Euclidean norm. Otherwise, set $i = i + 1$ and then go to (1)

However, there is another problem. Since problem (3.29) is not convex, we cannot find a global minimum by a general constrained optimization algorithm. Generally, there are two methods to find out the global optimal solution. One method is to try different initial guesses, and another is to consider a global optimization algorithm. The idea of the first method is quite simple. The initial guesses can be chose arbitrarily in the feasible region. We can choose different $\mathbf{y}_{(0)}$ and repeat Algorithm 1 to get a set of local optimal solutions. Finally, we can find the best solution among them. To ensure the solution is global, we also consider a hybrid global optimization method, which combines a simulated annealing algorithm and SQP algorithm. We tried these two methods in the experimental section and obtained the same optimal solution.

The details of the hybrid algorithm are given as follows. If we set a function

$f(\mathbf{y}) = \max\{E_s(\mathbf{y}, \mathbf{r}, f)\}$, then problem (3.29) is equivalent to

$$\min_{\mathbf{y} \in S'} \ f(\mathbf{y}), \tag{3.33}$$

where $f(\mathbf{y})$ is a non-convex function on the set $\Psi = \{\mathbf{y} \in S'\}$. Theoretically, a global minimum solution of $f(\mathbf{y})$ can be found by a simulated annealing algorithm alone. This algorithm contains three main steps:

1. generate the next trial point in space S' by random perturbations,
2. choose a probability distribution to manage the accepting of uphill steps,
3. an annealing schedule.

For the choice of a probability distribution, the classical Boltzmann distribution is applied here [79, 24]. In the annealing schedule, there are some determinants. Denote $t_e,\ \alpha,\ n_p,\ n_c$ as the initial temperature, cooling speed, the number of random perturbations for each temperature, and the number of cooling steps; and the value of the parameters is given in [119]. For the simulated annealing algorithm, there is also a drawback that the convergence rate is always relatively slow, while the SQP method is much faster in converging to a stationary point. Inspired by the method proposed in [169], we consider solving this non-convex optimal problem by a hybrid descent method to find a global optimal solution efficiently. The algorithm is described in Algorithm 2.

3.5 Special Structure

Compared with the conventional FIR implementations, the structure with IIR filters can be more efficient for the same filtering accuracy. However, the design problem formulated in (3.5) is complicated, especially when m is large, as it contains m feedback parts resulting in many stability problems. In addition, it is hard to

Algorithm 2 A hybrid global optimization algorithm

Require: Choose an initial guess $\mathbf{y}_{(0)}$ and $\delta_{(0)}$ randomly. Set $l = 0$.

Ensure: $\mathbf{y} \in S^{'}$ and $\delta \geq 0$

1: Set $\mathbf{x}_{(l)} = [\delta_{(l)}; \mathbf{y}_{(l)}]$.
2: Solve problem (3.29) by an SQP algorithm with the initial guess $\mathbf{x}_{(l)}$ to obtain a local minimal solution $\mathbf{x}_{(l*)}$, such that $f(\mathbf{y}_{(l*)}) - f(\mathbf{y}_{(l)}) \leq -\varepsilon_l$.
3: Set $\mathbf{y}_{(l*)}$ as the initial guess of simulating annealing algorithm, and choose $t_e,\ \alpha,\ n_p,\ n_c$. Evaluate $f(\mathbf{y}_{(l*)})$
4: **for** $j = 1;\ j < n_c;\ j = j + 1$ **do**
5: **for** $i = 1;\ i < n_p;\ i = i + 1$ **do**
6: $q \leftarrow random\{1, 2, 3\}$
7: **if** q=1 **then**
8: choose one element of $\mathbf{y}$ as $\hat{\mathbf{y}}$
9: **else if** q=2 **then**
10: $p \leftarrow random\{1, \cdots, n_p\}$
11: choose p elements of $\mathbf{y}$ as $\hat{\mathbf{y}}$
12: **else**
13: choose the whole vector of $\mathbf{y}$ as $\hat{\mathbf{y}}$
14: **end if**
15: $b \leftarrow f(\hat{\mathbf{y}}) - f(\mathbf{y})$
16: **if** $b < -\sigma_l$ or $random[0, 1] < t_e \exp(-b/t_e)$ **then**
17: $\mathbf{y} = \hat{\mathbf{y}}$
18: **end if**
19: $j \leftarrow j + 1$
20: **end for**
21: $t_e \leftarrow \alpha t_e$
22: **end for**
23: Set $\mathbf{y}_{(l+1)} = \mathbf{y}$, $\delta_{(l+1)} = f(\mathbf{y})$, and $l = l + 1$.
24: Return to step 1 until convergence.

find globally optimal solutions. To simplify the problem, we consider a specific structure in which we suppose that all the IIR filters share the same poles. Then, the beamforming structure applied IIR filters can be depicted in Fig. 3.2 with m all-zero filters and a standard all-pole filter. The multi-channel FIR part mainly concentrates on spatial filtering, while the common IIR filter part guarantees the frequency filtering efficiency. Thus, this specific structure improves efficiency and reduces design complexity as well.

For the proposed structure, we can rewrite the actual response of the beamformer

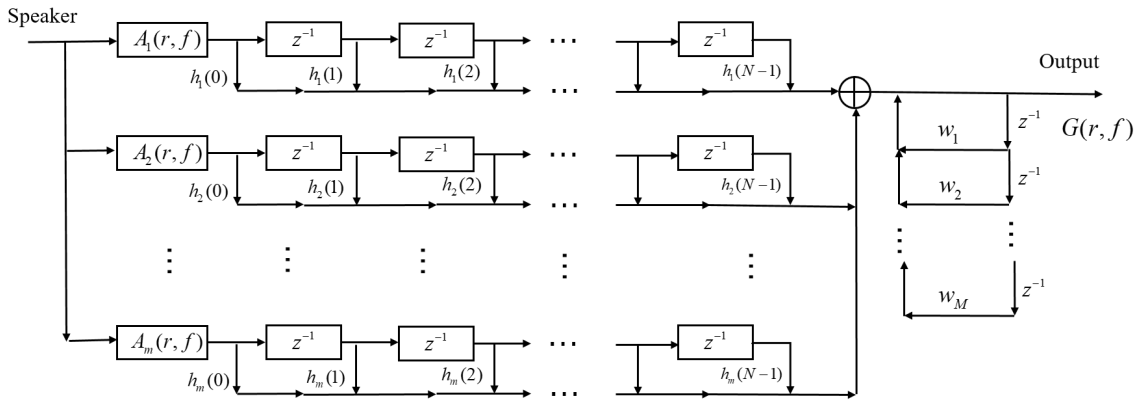


Figure 3.2: Beamforming structure using common feedback

as

$$\begin{aligned} G(\mathbf{r},f) =& \frac{\sum_{k=1}^{m} H_k(f)A_k(\mathbf{r},f)}{W(f)} \\ =& \frac{\mathbf{h}^{\mathsf{T}}[\boldsymbol{A}(\mathbf{r},f)\boldsymbol{d}_0^{\mathsf{T}}(f)]}{1+\mathbf{w}^{\mathsf{T}}\boldsymbol{d}_1(f)}, \end{aligned} \tag{3.34}$$

where $\mathbf{h}$ and $\mathbf{w}$ are vectors of the filter coefficients, denoted as

$$\begin{aligned} \mathbf{h} &= [\mathbf{h}_1^{\mathsf{T}}, \mathbf{h}_2^{\mathsf{T}}, \cdots, \mathbf{h}_P^{\mathsf{T}}]^{\mathsf{T}}, \\ \mathbf{w} &= [w(1), \cdots, w(M)]^{\mathsf{T}}. \end{aligned}$$

It would be much easier to consider the stability problem of transfer function in (3.34). The stable region becomes $S = \{\mathbf{w} : \mathbf{w}(f)\ is\ stable\}$.

We can reformulate the error between the real actual and desired response as

$$E(\mathbf{h},\mathbf{w},\mathbf{r},f) = \alpha(\mathbf{r},f)|G(\mathbf{r},f) - G_d(\mathbf{r},f)|^2, \tag{3.35}$$

where α is a positive weighting function. For an N-tap filter design problem, the desired response always contains a delay term τ_N, which always takes value in $[0, N-$

1]. Then, the desired response can be denoted as

$$G_d(\mathbf{r}, f) = e^{\frac{-j2\pi f\tau_N}{fs}} \bar{G}_d(\mathbf{r}, f). \tag{3.36}$$

Accordingly, we can rewrite $G(\mathbf{r}, f)$ as

$$G(\mathbf{r}, f) = e^{\frac{-j2\pi f\tau_N}{fs}} \bar{G}(\mathbf{r}, f), \tag{3.37}$$

where

$$\begin{aligned} \bar{G}(r, f) =& \frac{\mathbf{h}^T[\boldsymbol{A}(\mathbf{r}, f)\bar{\boldsymbol{d}}_0^T(f)]}{\mathbf{w}^T\boldsymbol{d}_1(f)} \\ \bar{\boldsymbol{d}}_0(f) =& e^{\frac{-j2\pi f(-\tau_N)}{fs}} \boldsymbol{d}_0(f) \\ =& [e^{\frac{-j2\pi f(-\tau_N)}{fs}}, e^{\frac{-j2\pi f(1-\tau_N)}{fs}}, \cdots, e^{\frac{-j2\pi f(N-1-\tau_N)}{fs}}]^{\mathsf{T}} \end{aligned} \tag{3.38}$$

Then, the cost function (3.35) is equivalent to

$$\bar{E}(\mathbf{h}, \mathbf{w}, \mathbf{r}, f) = \alpha(\mathbf{r}, f)|\bar{G}(\mathbf{r}, f) - \bar{G}_d(\mathbf{r}, f)|^2.$$

The original problem is transformed into an equivalent semi-infinite problem constituting infinite inequality constraints as

$$\begin{aligned} &\min_{\mathbf{w}\in S, \mathbf{h}\in R^{NP}, \delta\in R^+} \delta \\ &s.t.\ \bar{E}(\mathbf{h}, \mathbf{w}, \mathbf{r}, f) - \delta \leq 0,\ \forall(\mathbf{r}, f) \in \Omega \end{aligned} \tag{3.39}$$

The stable region of the specific structure is similar as the case discussed in section 3. When the order of $W(f)$ is 1 or 2, we have the stable region

$$S = \{\mathbf{w} : -1 < w(1) < 1\}. \tag{3.40}$$

If the order of feedback section is 2, the stable domain is

$$S = \{\mathbf{w} :\ \mathbf{D}\mathbf{w} < \mathbf{e}\}, \tag{3.41}$$

where

$$\mathbf{D} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}; \qquad \mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

For a general case where the order of $W(f)$ is greater than 2, we can also equivalently decompose the polynomial into a sum of partial fractions with lower orders. After the decomposition, for each k, we have

$$\frac{H_k(z)}{\bar{W}(z)} = \begin{cases} \sum_{l=1}^{M/2} \dfrac{b_{l0}^k + b_{l1}^k z^{-1}}{1 + \alpha_l z^{-1} + \beta_l z^{-2}}, & \text{if } N-1 < M \\ \sum_{l=1}^{M/2} \dfrac{b_{l0}^k + b_{l1}^k z^{-1}}{1 + \alpha_l z^{-1} + \beta_l z^{-2}} + \sum_{l=0}^{N-M-1} c_l^k z^{-l}, & \text{if } N-1 \geq M \end{cases} \tag{3.42}$$

when M is even, and

$$\frac{H_k(z)}{\bar{W}(z)} = \begin{cases} \dfrac{a_1}{1 - d_1 z^{-1}} + \sum_{l=1}^{(M-1)/2} \dfrac{b_{l0}^k + b_{l1}^k z^{-1}}{1 + \alpha_l z^{-1} + \beta_l z^{-2}}, & \text{if } N-1 < M \\ \dfrac{a_1}{1 - d_1 z^{-1}} + \sum_{l=1}^{(M-1)/2} \dfrac{b_{l0}^k + b_{l1}^k z^{-1}}{1 + \alpha_l z^{-1} + \beta_l z^{-2}} + \sum_{l=0}^{N-M-1} c_l^k z^{-l}, & \text{if } N-1 \geq M \end{cases} \tag{3.43}$$

when M is odd. Then, we can obtain the stability domain as

$$S = \{(\alpha_l, \beta_l) : \mathbf{D} \begin{pmatrix} \alpha_l \\ \beta_l \end{pmatrix} < \mathbf{e}\}, \ l = 1, \cdots, \frac{M}{2}, \tag{3.44}$$

when M is even and

$$S = \{(d_1, \alpha_l, \beta_l) : -1 < d_1 < 1; \mathbf{D} \begin{pmatrix} \alpha_l \\ \beta_l \end{pmatrix} < \mathbf{e}\}, \ l = 1, \cdots, \frac{M-1}{2}, \tag{3.45}$$

when M is odd.

Then, for a given $\frac{H_k(z)}{W(z)}$, where the coefficients are $\mathbf{h}_k$, $\mathbf{w}$, there are the corresponding coefficients a_1, b_{l0}^k, b_{l1}^k, c_l^k, d_1, α_l, β_l such that $\frac{H_k(z)}{\bar{W}(z)}$ approaches to $\frac{H_k(z)}{W(z)}$. Hence, we denote all these variables by $\mathbf{y}$ and optimize $\mathbf{y}$ directly. The transfer function becomes

$$G_s(\mathbf{r}, f) = \frac{\sum_{k=1}^{m} H_k(f) A_k(\mathbf{r}, f)}{\bar{W}(f)}, \tag{3.46}$$

where $\frac{H_k(z)}{\bar{W}(z)}$ is given by (3.42) and (3.43). The error between the real actual and desired response becomes

$$E_s(\mathbf{y}, \mathbf{r}, f) = \alpha(\mathbf{r}, f)|G_s(\mathbf{r}, f) - G_d(\mathbf{r}, f)|^2. \tag{3.47}$$

We can transfer the problem (3.39) into the problem below.

$$\begin{aligned} &\min_{\mathbf{y}\in S, \delta \in R^+} \quad \delta \\ &s.t.\ E_s(\mathbf{y}, \mathbf{r}, f) - \delta \leq 0,\ \forall (r, f) \in \Omega. \end{aligned} \tag{3.48}$$

Problem (3.48) can be optimized by the algorithms given in Section 4.

3.6 Performance Limit Analysis

For the proposed beamforming, there are two parameters, N and M, which are the numbers of parameters in the numerator and denominator, respectively. If N and M change, the optimal value also changes. As in many applications N and M can be chosen arbitrarily, it is necessary to analyze the change rule between the optimal value and the orders N and M to help the designers make a proper choice.

For this problem, we denote the optimal value in the case of N and M by $V(N, M)$. To avoid a waste of calculations and memory usage, it is necessary to find the specific value of $\inf_{N,M} V(N, M)$. Once this limit value is reached, there is

no need to increase the filter order. We can see that there are two variables in $\inf_{N,M} V(N,M)$, and it would be difficult to define the infimum value. For the order M, it is related to the complexity of the design processing. If M is large, the nonlinearity and the instability of the design problem are also severe, and the design problem formulated in (3.48) becomes very complicated. In addition, a large M is not favorable in actual applications. Therefore, for the calculation of infimum value, we choose M as a fixed small number. Then, the infimum value is transformed from $\inf_{N,M} V(N,M)$ to $\inf_{N} V(N,M)$, and the problem is simplified. The performance limit is the same with different fixed M, which will be proved later. The difference between different M is that the larger M can achieve the limit performance with a smaller N. This is in accordance with the property of the IIR filter: it could be more efficiently approximating the optimal frequency part comparing with the FIR filter. The numerical examples also prove it. Designers can choose the suitable N and M according to their demands.

The followings are the calculation of infimum values. As the specific structure is relatively more uncomplicated, we first analyze the performance limit of the specific structure. Then we prove that the general structure has the same infimum value as the specific structure.

3.6.1 Performance Limit of a Special Structure

For a special structure, as depicted in Fig. 3.2, denote $R_k(f) = H_k(f)/W(f)$, $k = 1, \cdots, m$. Then, the actual response function is rewritten as

$$G(\mathbf{r}, f) = A^{\top}(\mathbf{r}, f)\boldsymbol{R}(f), \text{ where } \boldsymbol{R}(f) = [R_1(f), \cdots, R_m(f)]^{\mathsf{T}}. \tag{3.49}$$

Denote a set functions as $\Delta_N = \{e^{-j2\pi f(i-\tau_N)} : i = 0, \cdots, N-1\}$. Then, for each $k = 1, \cdots, m$, $H_k(f)$ can be expressed as the linear combination of the elements in Δ_N. If Δ_N is monotonically increasing as N increases, then $\inf_{N} V(N,M)$ can be

simplified as a limit $\lim_{N\to\infty} V(N,M)$. However, this is always not the case, because of the existence of τ_N. Basically, if N increases, the delay τ_N always increases. Then, for the subsequences are monotonically increasing and then the limit of each subsequences can be computed.

For example, if $\tau_N = (N+1)/3$, we can find three subsequences $\{\Delta_1, \Delta_4, \Delta_7, \cdots\}$, $\{\Delta_2, \Delta_5, \Delta_8, \cdots\}$, $\{\Delta_3, \Delta_6, \Delta_9, \cdots\}$, such that they are monotonically increasing. That is,

1. $N = 3\bar{N} + 1$,

$$\begin{aligned}\Delta_N =& \{e^{-j2\pi f(i-\tau_{\bar{N}})} : i = 0, \cdots, N-1\} \\ =& \{e^{-j2\pi fi} e^{j\frac{4}{3}\pi f} : i = -\bar{N}, \cdots, 2\bar{N}\} \\ \subset& \{e^{-j2\pi fi} e^{j\frac{4}{3}\pi f} : i = -\bar{N} - 1 \cdots, 2\bar{N} + 2\} \\ =& \Delta_{N+1}\end{aligned}$$

2. $N = 3\bar{N} + 2$,

$$\begin{aligned}\Delta_N =& \{e^{-j2\pi f(i-\tau_{\bar{N}})} : i = 0, \cdots, N-1\} \\ =& \{e^{-j2\pi fi} : i = -\bar{N} - 1, \cdots, 2\bar{N}\} \\ \subset& \{e^{-j2\pi fi} : i = -\bar{N} - 2 \cdots, 2\bar{N} + 2\} \\ =& \Delta_{N+1}\end{aligned}$$

3. $N = 3\bar{N}$,

$$\begin{aligned}\Delta_N =& \{e^{-j2\pi f(i-\tau_{\bar{N}})} : i = 0, \cdots, N-1\} \\ =& \{e^{-j2\pi fi} e^{j\frac{2}{3}\pi f} : i = -\bar{N}, \cdots, 2\bar{N}\} \\ \subset& \{e^{-j2\pi fi} e^{j\frac{2}{3}\pi f} : i = -\bar{N} - 1 \cdots, 2\bar{N} + 2\} \\ =& \Delta_{N+1}\end{aligned}$$

It can be seen that for each subsequences Δ_{N_l}, there exists $\eta \in [0,1)$, such that $L-\tau_{N_l} = \eta + i_l, \forall l = 1, \cdots$, where i_l is some integer. Then, $\eta = -\tau_{N_l} - [-\tau_{N_l}]$, where $[-\tau_{N_l}]$ is the largest integer less than or equals to $-\tau_{N_l}$. Denote

$\Gamma_0 = \{\bar{H}(f) : \bar{H}(f) = u(f) + jv(f), f \in [0, f_s/2], u(f)$ and $v(f)$ are continuous and their left derivatives and right derivatives exist, $v(0) = 0, v(f_s/2) = 0\}$.

Then,

$$\begin{aligned}\Gamma_\eta =& e^{-j2\pi f\eta/f_s/2} \cdot \Gamma_0 \\ =& \{\bar{H}(f) : \bar{H}(f) = u(f) + jv(f), t \in [0, f_s/2]), u(f) \text{ and } v(f) \text{ are continuous and} \\ & \text{their left derivatives and right derivatives exist}, v(0) = 0, \bar{H}(f_s/2) = a \cdot e^{j\pi\eta}, \\ & a \text{ is a real number}\}\end{aligned}$$

Then, for Γ_0, we have the following theorem.

Theorem 3.2. *Suppose that τ_{N_l} is an integer for any l, $\lim_{l\to\infty} \tau_{N_l} = +\infty$ and $\lim_{l\to\infty} N_l - \tau_{N_l} = +\infty$. Then, for any complex function $R(f)$ defined in $[0, f_s/2]$, if $R(f) \in \Gamma_0$, then there exists a series of coefficients $\{c_i : i = 0, \cdots\}$, such that*

$$R(f) = \lim_{l\to+\infty} \sum_{i=0}^{N_l-1} c_i e^{-j2\pi f(i-\tau_{N_l})/f_s} / W(f), \tag{3.50}$$

where $W(f)$ is an arbitrary stable filter.

Proof. Since $W(f) = \sum_{i=0}^{M-1} w(k) e^{-j2\pi if/f_s}$, then $W(0)$ and $W(f_s/2)$ are real numbers. Let $H(f) = R(f) \cdot W(f)$, then $H(0)$ and $H(f_s/2)$ are real numbers. Hence, $H(f) \in \Gamma_0$. Denote $H(f) = u(f) + jv(f)$, then we generate $H(f)$ from $[0, f_s/2]$ to $(-\infty, +\infty)$ by

$$\begin{cases} u(f) = u(-f), \quad v(f) = -v(-f), \text{ if } f \in [-f_s/2, 0] \\ v(f) = u(f - if_s), \quad v(f) = v(f - if_s), \text{ if } f \in [(i-0.5)f_s, (i+0.5)f_s]. \end{cases}$$

By Fourier series approximation, $H(f)$ can be expressed by the basis functions

$$\{1,\ \cos(2\pi if/f_s),\ \sin(2\pi if/f_s),\ i=1,2,\cdots,\infty\}. \tag{3.51}$$

That is, there exist $\{a_i,\ i=0,1,\cdots\}$, $\{b_i,\ i=1,2,\cdots\}$, such that

$$\begin{cases} u(f) = a_0 + \sum_{i=1}^{\infty} a_i \cos(2\pi if/f_s) \\ v(f) = \sum_{i=1}^{\infty} b_i \sin(2\pi if/f_s). \end{cases}$$

Denote coefficients d_i as

$$d_0 = a_0, \quad d_i = \begin{cases} (a_i - b_i)/2, & \text{if } i > 0 \\ (a_i + b_i)/2, & \text{if } i < 0, \end{cases}$$

and $d_i = c_{(i+\tau_{N_l})}$. Then, we have

$$\begin{aligned} H(f) &= a_0 + \sum_{i=1}^{+\infty} a_i \cos(2\pi if/f_s) + j \sum_{i=1}^{+\infty} b_i \sin(2\pi if/f_s) \\ &= a_0 + \sum_{i=1}^{+\infty} \frac{a_i + b_i}{2} \cos(2\pi if/f_s) + \sum_{i=1}^{+\infty} \frac{a_i - b_i}{2} \cos(2\pi if/f_s) \\ &\quad + j \sum_{i=1}^{+\infty} \frac{a_i + b_i}{2} \sin(2\pi if/f_s) - j \sum_{i=1}^{+\infty} \frac{a_i - b_i}{2} \sin(2\pi if/f_s) \\ &= a_0 + \sum_{i=1}^{+\infty} \frac{a_i + b_i}{2} (\cos(2\pi if/f_s) + j \sin(2\pi if/f_s) \\ &\quad + \sum_{i=1}^{+\infty} \frac{a_i - b_i}{2} (\cos(2\pi if/f_s) - j \sin(2\pi if/f_s) \\ &= c_0 + \sum_{i=1}^{+\infty} d_{(-i)} e^{j2\pi if/f_s} + \sum_{i=1}^{+\infty} d_i e^{-j2\pi if/f_s} \\ &= \sum_{i=-\infty}^{\infty} d_i e^{-j2\pi if/f_s} \end{aligned}$$

Note that $\lim_{l\to\infty} \tau_{N_l} = \infty$, $\lim_{l\to\infty} N_l - \tau_{N_l} = \infty$, we have

$$\sum_{i=-\infty}^{\infty} d_i e^{-j2\pi i f/f_s} = \lim_{l\to\infty} \sum_{i=0}^{N_l-1} c_i e^{-j2\pi f(i-\tau_{N_l})/f_s}.$$

Hence,

$$H(f) = \lim_{l\to\infty} \sum_{i=0}^{N_l-1} c_i e^{-j2\pi f(i-\tau_{N_l})/f_s}.$$

Since $W(f)$ is stable, then $W(f) \neq 0$, $\forall f \in [0, f_s/2]$, then we have

$$\begin{aligned} R(f) &= H(f)/W(f) \\ &= (\lim_{l\to\infty} \sum_{i=0}^{N_l-1} c_i e^{-j2\pi f(i-\tau_{N_l})/f_s})/W(f). \end{aligned}$$

The proof completes. □

Similarly, we have the result for Γ_η as follows.

Corollary 3.2. *Suppose that $\tau_{N_l} = \eta + i_l$, $\forall l$, where i_l is an integer, and $\lim_{l\to+\infty} \tau_{N_l} = +\infty$, $\lim_{l\to+\infty} N_l - \tau_{N_l} = +\infty$. Then, for any complex function $R(f)$ defined in $[0, f_s/2]$, if $R(f) \in \Gamma_\eta$, then there exists a series $\{c_i : i = 0, \cdots\}$ such that*

$$R(f) = \lim_{l\to+\infty} \sum_{i=0}^{N_l-1} c_i e^{-j2\pi f(i-\tau_{N_l})/f_s}/W(f) \tag{3.52}$$

where $W(f)$ is an arbitrary stable filter.

Proof. Let $\bar{R}(f) = e^{j2\pi f\eta/f_s} \cdot R(f)$, $\bar{\tau}_{N_l} = \tau_{N_l} - \eta$. Then, $\bar{R}(f) \in \Gamma_0$. Since $\bar{\tau}_{N_l}$ is an integer for any l, and $\lim_{l\to+\infty} \bar{\tau}_{N_l} = \tau_{N_l} - \eta = +\infty$, $\lim_{l\to+\infty} N_l - \bar{\tau}_{N_l} = N_l - \tau_{N_l} + \eta = +\infty$.

It follows by Theorem 3.2 that there exists a series $\{c_i : i = 1, \cdots\}$ such that

$$\bar{R}(f) = \lim_{l \to +\infty} \sum_{i=0}^{N_l - 1} c_i e^{-j2\pi f(i - \bar{\tau}_{N_l})/f_s} / W(f).$$

Multiplied by $e^{-j2\pi f\eta/f_s}$ in two sides, we obtain (3.52). The proof completes. □

By Theorem 3.2 and Corollary 3.2, for any subsequences N_l of the same η, the set function ΔN_l can approach to $\{e^{-j2\pi f\eta} \cdot e^{-j2\pi if} : i = -\infty, \cdots, +\infty\}$. Hence, any $R(f)$ defined in $[0, f_s/2]$ and belongs to Γ_η can be approximated by (3.50). Then, the infimum of the optimal value can be decomposed as many limits, where each limit is related to one fixed η and is computed by

$$\inf_{N_l} V(N_l, M) = \lim_{l \to +\infty} V(N_l, M).$$

Find all possible η, and denote the subsequences for η by $N_l(\eta)$, the infimum of the optimal value is computed as

$$\inf_{N} V(N, M) = \min_{\eta} (\lim_{N_l \to \infty} V(N_l(\eta), M)). \tag{3.53}$$

To simplify the computation of (3.53), we formulate a functional optimization problem for each subsequences $N_l(\eta)$ as

$$\text{Problem}(\eta): \quad \begin{aligned} &\min_{R(f) \in \Gamma_\eta, z} \quad z \\ &s.t. \quad \alpha(\mathbf{r}, f)|A^\top(\mathbf{r}, f)R(f) - G_d(\mathbf{r}, f)|^2 \le z, \quad \forall (\mathbf{r}, f) \in \Omega \end{aligned} \tag{3.54}$$

For each $f \in [0, f_s/2]$, denote the corresponding space set for $f \in \Omega$ by Ω_f. Problem (3.54) can be decomposed by many subproblems as

$$\text{Problem}(\eta, f): \quad \begin{aligned} &\min_{R(f)} \quad z \\ &s.t. \quad \alpha(\mathbf{r}, f)|A^\top(\mathbf{r}, f)R(f) - G_d(\mathbf{r}, f)|^2 \le z, \quad \forall \mathbf{r} \in \Omega_f, \end{aligned} \tag{3.55}$$

where $R(f)$ take value in a suitable space.

Remark 1. *Note that $f \in [0, f_s/2]$ and there may exist f such that Ω_f is an empty set. It means that Problem (η, f) is an empty problem. For this case, the solution $R(f)$ is not determined and can be any value such that $R(f) \in \Gamma_\eta$ is satisfied.*

Remark 2. *If $f \in [0, \bar{f}_s] \subset [0, f_s/2]$, there is no difference between any two different Γ_η. The difference between two different Γ_η is the function value at $f_s/2$. Hence, if $f \in [0, \bar{f}_s] \subset [0, f_s/2]$, Problem (η_1, f) is the same as Problem (η_2, f), for any $\eta_1 \neq \eta_2$. In this case, we only need to solve Problem $(0, f)$.*

If $f = 0$, the function value is a real number, then Problem$(\eta, 0)$ becomes

$$\begin{aligned} &\min_{R_0 \in R^{N_l}, \delta} \quad \delta \\ &s.t. \quad \alpha(\mathbf{r}, 0)|A^\top(\mathbf{r}, 0)R_0 - G_d(\mathbf{r}, 0)|^2 \leq \delta, \quad \forall \mathbf{r} \in \Omega_0. \end{aligned} \tag{3.56}$$

If $f \in (0, f_s/2)$, the function value is a complex number, then Problem(η, f) is

$$\begin{aligned} &\min_{R_0 \in C^{N_l}, \delta} \quad \delta \\ &s.t. \quad \alpha(\mathbf{r}, f)|A^\top(\mathbf{r}, f)R_f - G_d(\mathbf{r}, f)|^2 \leq \delta, \quad \forall \mathbf{r} \in \Omega_f. \end{aligned} \tag{3.57}$$

If $f = f_s/2$, we denote $G_d^\eta(\mathbf{r}, f) = G_d(\mathbf{r}, f)e^{j\pi\eta}$, then Problem$(\eta, f_s/2)$ becomes

$$\begin{aligned} &\min_{R_{f_s/2} \in R^{N_l}, \delta} \quad \delta \\ &s.t. \quad \alpha(\mathbf{r}, f_s/2)|A^\top(\mathbf{r}, f_s/2)R_{f_s/2} - G_d^\eta(\mathbf{r}, f_s/2)|^2 \leq \delta, \quad \mathbf{r} \in \Omega_{f_s/2}. \end{aligned} \tag{3.58}$$

After Problem (η, f) is solved for each f in $[0, f_s/2]$, where the optimal value is δ_f and then the optimal value of Problem (η) is computed as

$$\delta_\eta = \max_{f \in [0, f_s/2]} \delta_f. \tag{3.59}$$

For each η, we can obtain z_η; then the infimum value can be computed as

$$\inf_N V(N, M) = \min_\eta \delta_\eta. \tag{3.60}$$

Finally, for the infimum value $\inf_N V(N, M)$, we have follows theorem.

Theorem 3.3. *For different orders M_1 and M_2, $\inf_N V(N, M_1) = \inf_N V(N, M_2)$.*

Proof. Suppose that $W_1(f)$ is an arbitrary stable filter with order M_1, and $W_2(f)$ is an arbitrary stable filter with order M_2. Then, by Corollary 3.2, any $R(f)$ in Γ_η can be expressed as

$$\begin{aligned} R(f) &= \lim_{l\to+\infty} \sum_{i=0}^{N_l-1} c_i^{(1)} e^{-j2\pi f(i-\tau_{N_l})/f_s} / W_1(f) \\ &= \lim_{l\to+\infty} \sum_{i=0}^{N_l-1} c_i^{(2)} e^{-j2\pi f(i-\tau_{N_l})/f_s} / W_2(f), \end{aligned} \tag{3.61}$$

where $c_i^{(1)}$ and $c_i^{(2)}$ are the corresponding coefficients. Hence, for different orders M_1 and M_2, the space Γ_η is the same for any $\eta \in [0, 1)$. Then, $\inf_N V(N, M_1)$ and $\inf_N V(N, M_2)$, which both are the optimal value of (3.54), are the same. The proof completes. □

It can be seen from Theorem 3.3 that $\inf_N V(N, M)$ is independent of M. The difference between different M is the convergence rate. If M increases, the convergence rate also increases, which can be seen from the numerical experiment below.

3.6.2 Performance Limit of a General Structure

This part considers the case of the general structure, as depicted in Fig. 3.1. Suppose that $V(N, M)$ is the optimal value of Problem (3.6) with the coefficients N and M. That is

$$V(N, M) = \max_{(\mathbf{r},f)\in\Omega} E_1(\mathbf{h}^{N,*}, \mathbf{w}^{M,*}, \mathbf{r}, f),$$

where $\mathbf{h}^{N,*}$ and $\mathbf{w}^{M,*}$ are the optimal solutions. We have Problem (η) as (3.54)

$$\text{Problem}(\eta): \quad \min_{R(f)\in\Gamma_\eta} \max_{(r,f)\in\Omega} \alpha(\mathbf{r}, f)|A^\top(\mathbf{r}, f)R(f) - G_d(\mathbf{r}, f)|^2.$$

The convergence of the problem can be found in the following theorem.

Theorem 3.4. *Suppose that N_l is a subsequences with the same η. Then, $\lim_{N_l\to\infty} V(N,M) = V^*$, where V^* is the optimal value of Problem (η).*

Proof. Suppose that the optimal solution and the optimal value of problem (η) are $R^*(f)$ and v^*, that is

$$V^* = \max_{(r,f)\in\Omega} \alpha(r,f)|A^{\mathsf{T}}(r,f)R^*(f) - G_d(r,f)|^2.$$

Let $H^{N_l,*}(f)$ and $W^{M,*}(f)$ are the corresponding response functions with $\mathbf{h}^{N_l,*}$ and $\mathbf{w}^{M,*}$. Then, we have

$$R_k^{N_l,M,*}(f) = H_k^{N_l,*}(f)/W_k^{M,*}(f) \in \Gamma_n, \quad \forall k.$$

By the optimality of Problem (η), we have

$$V(N,M) \geq V^*, \quad \forall N_l.$$

Then, we obtain

$$\inf V(N,M) \geq V^*.$$

By the monotonicity of $V(N,M)$, as N_l increases, we have

$$\lim_{N_l\to\infty} V(N,M) = \inf_{N_l} V(N,M) \geq V^*. \tag{3.62}$$

On the other hand, it follows by Theorem 3.2 that there exists the coefficients $\mathbf{h}^{N_l}$, $\mathbf{w}^M$ and the corresponding response functions $H^{N_l}(f)$ and $W^M(f)$, such that

$$R_k^*(f) = \lim_{l\to\infty} \sum_{i=0}^{N_l-1} H_k^{N_l}(f)/W_k(f),$$

where $W(f) = W^M(f)$ can be chosen the same as N_l increases. By the continuity of the function

$$\max_{(r,f)\in\Omega} \alpha(r,f)|A^{\mathsf{T}}(r,f)R(f) - G_d(r,f)|^2,$$

we have

$$\lim_{l\to+\infty}\max_{(r,f)\in\Omega}\alpha(r,f)|A^{\mathsf{T}}(r,f)\frac{H^{N_l}(f)}{W(f)}-G_d(r,f)|^2=\max_{(r,f)\in\Omega}\alpha(r,f)|A^{\mathsf{T}}(r,f)R^*(f)-G_d(r,f)|^2.$$

That is

$$V^*=\lim_{l\to+\infty}\max_{(r,f)\in\Omega}\alpha(r,f)|A^{\mathsf{T}}(r,f)\frac{H^{N_l}(f)}{W(f)}-G_d(r,f)|^2.$$

Since

$$\begin{aligned}V(N_l,M)&=\lim_{l\to+\infty}\max_{(r,f)\in\Omega}\alpha(r,f)|A^{\mathsf{T}}(r,f)\frac{H^{N_l}(f)}{W^M(f)}-G_d(r,f)|^2\\&\leq\max_{(r,f)\in\Omega}\alpha(r,f)|A^{\mathsf{T}}(r,f)\frac{H^{N_l}(f)}{W(f)}-G_d(r,f)|^2.\end{aligned}$$

Then, we have

$$\lim_{l\to+\infty}V(N_l,M)\leq V^*. \tag{3.63}$$

Combined with (3.62) and (3.63), we have

$$\lim_{l\to+\infty}V(N_l,M)=V^*.$$

□

From the theorem above, we can see that the specific structure and the general structure have the same limit performance with the same fixed M. The difference between the two structures is that the general structure can converge to the limited performance with a faster speed. The results are also following the theorem.

3.7 Simulation Results

This part shows the simulation results of two examples designed by the proposed method in the free field and room simulation model. First, the direct transfer function depicting the sound wave propagation in an acoustic-free field is applied.

The performance of beamformers with different filter lengths is compared with FIR beamformers. Then a rectangular room is defined for the fast ISM room simulator to calculate the RIRs [85]. The limit performances of different models are compared.

3.7.1 Experiments with a free-field model

Example 1

The first example is the same as the first example in [43], so it is convenient to compare filter performances. We design an equispaced linear array with 5 elements and a 5cm element spacing, and the central microphone is located as $(0, 0)$. Set the sampling rate as 8kHz. The beamformer is specified on an x-axis parallel with 1m in front of the array. The beamformer is suitable for the frequency range of human voice, which can be applied in hand-free equipment. The pass region is defined as

$$\{(x, f) : 0.3kHz \leq f \leq 2kHz, -0.4m \leq x \leq 0.4m\},$$

and stop region is defined as

$$\{(x, f) : 2.5kHz \leq f \leq 4kHz, -0.4m \leq x \leq 0.4m\}$$
$$\{(x, f) : 0.3kHz \leq f \leq 2kHz, 1.5m \leq |x| \leq 2.5m\}$$
$$\{(x, f) : 2.5kHz \leq f \leq 4kHz, 1.5m \leq |x| \leq 2.5m\}.$$

Then, we can consider the pass region and stop regions as four rectangular regions. The discrete points of the frequency domain are taken every 0.01 kHz, and for the spatial domain, we take discrete points every 0.01m.

Firstly, we consider the free-field model and compare the performance of the general and specific structures. Table 3.1 gives the optimal value of the cost function with different N and a fixed $M = 2$, where N is the order of numerator, and M is the order of denominator. The first row represents the special structure, as depicted in Fig. 3.2, where all the IIR filters share the same feedback section. The results

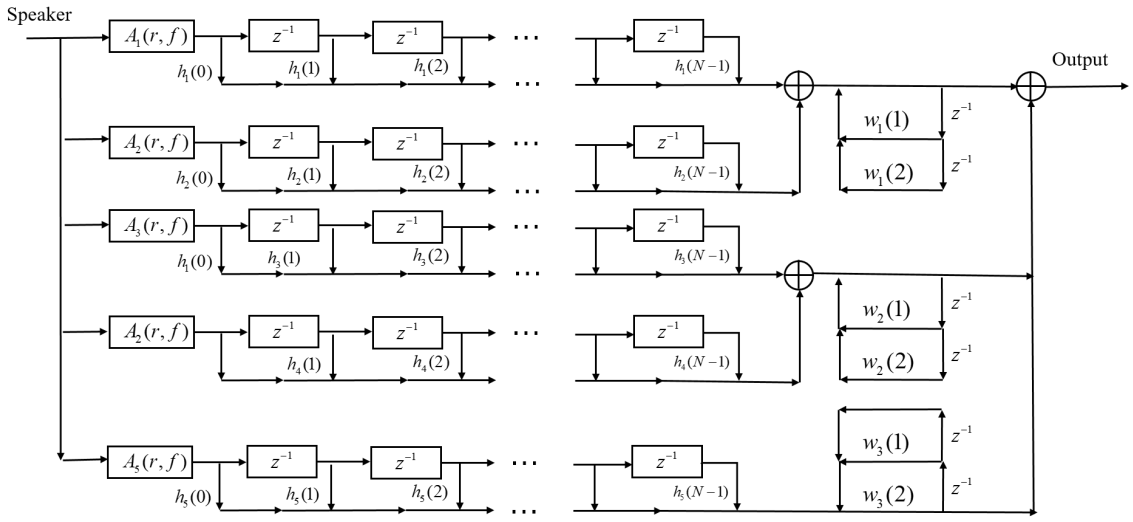


Figure 3.3: The structure of the 3 feedback sections

are obtained by solving the problem (3.48). The third row represents the results of the general structure, as depicted in Fig. 3.1, which contains five different feedback sections. The results are obtained by solving the problem (3.29). The second row considers a structure that contains three different feedback sections, as given in Fig. 3.3, in which the five filters are divided into three groups, and the same group shares the same feedback section. This resembles the situation that we have 3 distributed arrays, with each array having a common feedback structure. We can see that when N is small, the general structure has a better performance. When N becomes more significant, the three structures have similar performances.

Table 3.1: Cost function value (dB) with different number of IIR

$No.$	$N = 6$	$N = 10$	$N = 14$	$N = 18$	$N = \infty$
1	-11.44	-14.40	-18.44	-20.80	-21.40
3	-11.97	-17.59	-18.60	-20.80	-21.40
5	-16.32	-17.95	-18.66	-20.82	-21.40

Then, we compare the results of the FIR and IIR-based patterns. We apply the specific structure in the free-field model to solve the problem in (3.48) with a series of

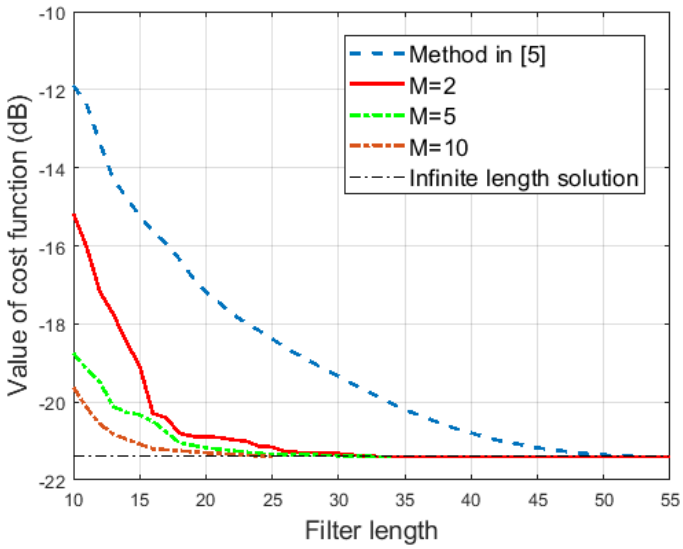


Figure 3.4: Cost function value for the first example

different N from 10 to 55. For all N, we set three different fixed order of denominator as $M = 2$, $M = 5$ and $M = 10$. For the proposed method, when the length of the filter is 30 with a fixed $M = 2$, the cost function value approaches infinite length solution, while the beamformers designed by the method in [43] converge to the same value at a larger length of 55. It can be seen that under the same length of filters, the proposed method gains a much better performance. The comparing results are shown in Fig.3.4, which describes the change of cost function value while the filter length increases. Both methods converge to the same infinite length solution, and the infinite cost function value is -21.40dB. The beamformers with different lengths of denominator also converge to the same performance limit, which is following Theorem 3.3. In addition, with the same fixed N, the cost function value decreases when M increases. The magnitude of the actual response when the $N = 22$ and $M = 2$ is shown in Fig.3.5. The resulting magnitude of actual response for frequency $f = 1$kHz and $x = 0.2$m are given in Fig.3.6(a) and Fig.3.6(b).

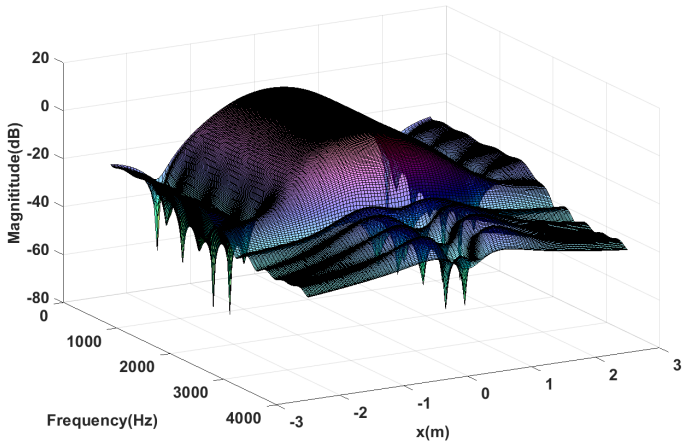
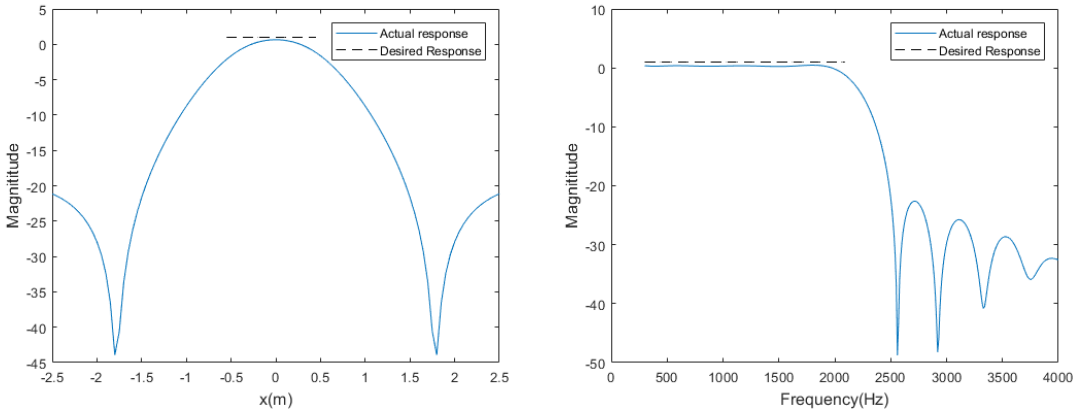


Figure 3.5: The magnitude of the actual response of the first example in free-field



(a) $f = 1$kHz

(b) $x = 0.2$m

Figure 3.6: Magnitude of the actual response and desired response of example 1 in free field

Example 2

In the second example, we consider a 2-dimensional region situation. Suppose that there is a microphone array with five microphones, locating at $\{[(-0.05, 0, 2); (0, -0.05, 2); (0, 0, 2); (0, 0.05, 2); (0.05, 0, 2)]\}$. Given the passband and stopband region on an $(x, y)-$plane with $z = 1$ under the microphone array. The pass region is defined

as

$$\Omega_p = \{(r, f) : -0.4m \leq x \leq 0.4m; -0.4m \leq y \leq 0.4m; 0.5kHz \leq f \leq 1.5kHz\},$$

and the stopband is defined the following three parts

$$\begin{aligned}
\{(r, f) : \quad & x \in [-4m, -2m] \cup [-0.4m, 0.4m] \cup [2m, 4m]; \\
& y \in [-4m, -2m] \cup [-0.4m, 0.4m] \cup [2m, 4m]; f \in [2kHz, 4kHz]\} \\
\{(r, f) : \quad & x \in [-0.4m, 0.4m]; \\
& y \in [-4m, -2m] \cup [-0.4m, 0.4m] \cup [2m, 4m]; f \in [0.5kHz, 1.5kHz]\} \\
\{(r, f) : \quad & x \in [-4m, -2m] \cup [-0.4m, 0.4m] \cup [2m, 4m]; \\
& y \in [-0.4m, 0.4m]; f \in [0.5kHz, 1.5kHz]\}
\end{aligned}$$

The same as the Example 1, the weighting function $\alpha(\mathbf{r}, f)$ is 1, and the sampling rate is 8kHz. Then, we solve the optimal solution with a length from 7 to 17. Fig. 3.7 gives the cost function of different filter lengths, and it can be seen that when the length is 16, the cost function converges to the limit performance -14.43dB. When comparing FIR-based beamformers, the beamformers designed by the proposed method still gain better cost function values with the same filter length. The actual response of the beamforming, whose filter length is 16, is depicted in Fig. 3.8.

3.7.2 Experiments with room simulation

Example 1

To investigate the method better, we further evaluate the performance of the beamformers designed in a simulated room. The settings of the microphones and desired response are the same the example 1 in Section 3.7.1. One rectangular room with a size of $8m \times 6m \times 3m$ is defined for the fast-ISM room simulator, and we set the relevant reverberation time T_{60} as 0.1, 0.2, 0.3 and 0.4 which characterizes the room surface. Compared with the free field transfer function, the limit cost function value

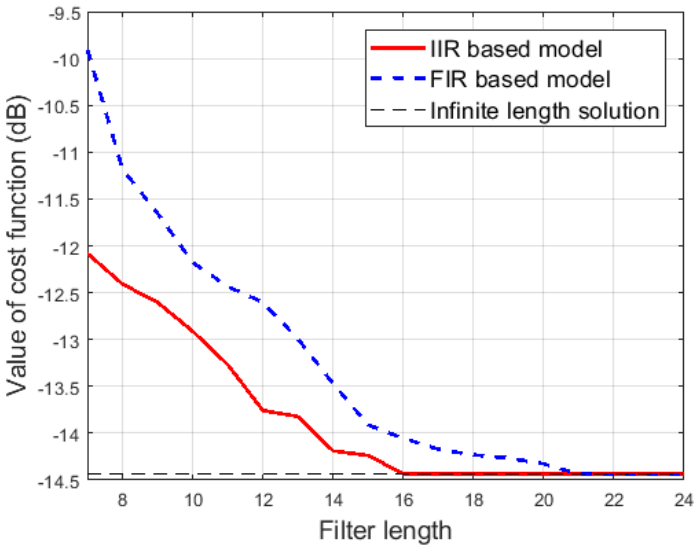


Figure 3.7: Cost function value for the second example

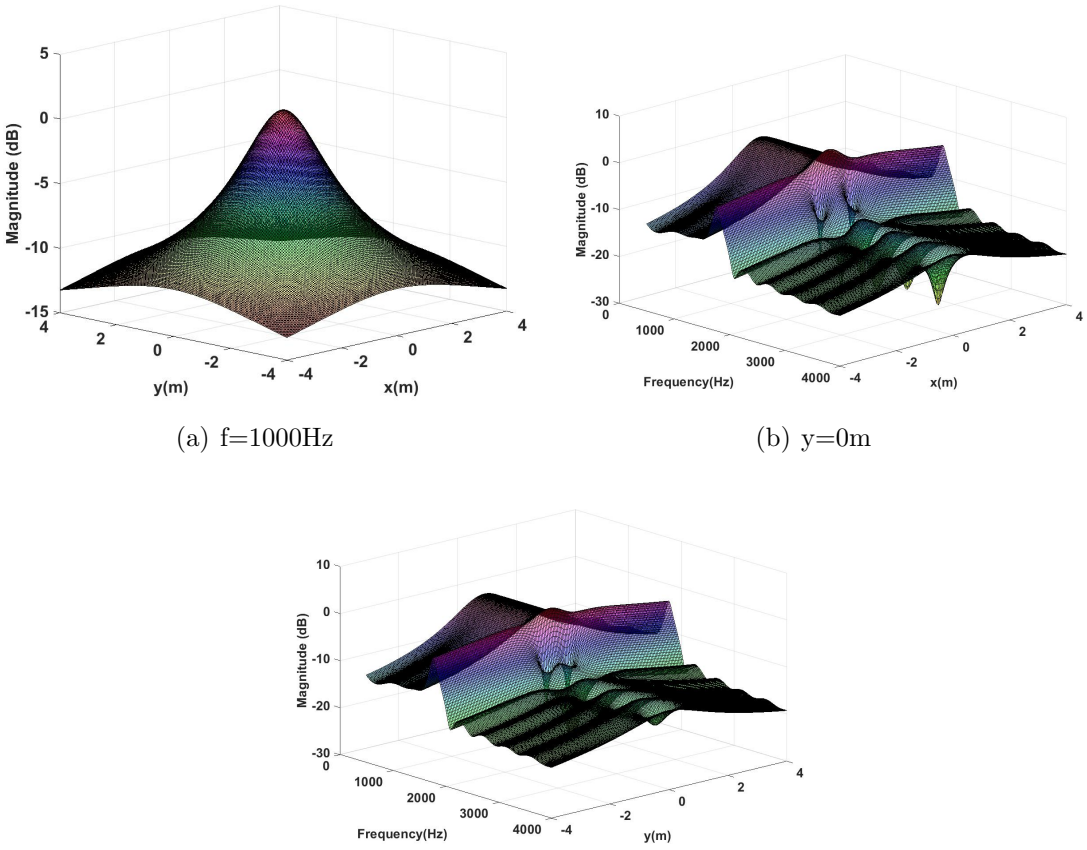
designed by the image model has a slight decrease, and as the T_{60} grows, the decline in limit performance increases. When $T_{60} = 0.1$, the performance limit is -17.03dB, while $T_{60} = 0.4$, the performance limit reduced to -11.02dB. In this example, we set the fixed $M = 2$. The cost function value of different reverberation time (T_{60}) with different filter length is given in Table 3.2. The magnitude of the actual response with different T_{60} when $N = 22$ and $M = 2$ is given in Fig.3.9.

Table 3.2: Cost function value (dB) in simulated room

T_{60}	Limit value	$N = 4$		$N = 10$		$N = 16$		$N = 22$	
		IIR	FIR	IIR	FIR	IIR	FIR	IIR	FIR
0.1	-17.03	-8.12	-6.86	-14.15	-11.94	-16.98	-14.94	-17.03	-16.04
0.2	-14.35	-7.01	-5.98	-11.86	-10.69	-13.99	-13.74	-14.34	-14.34
0.3	-12.63	-6.93	-4.90	-10.43	-9.38	-11.51	-11.17	-12.29	-11.91
0.4	-11.02	-6.10	-4.43	-9.82	-8.95	-10.61	-10.25	-11.02	-10.73

Example 2

In this example, we also test the beamformers designed by the image model. The settings of the microphones and desired response are the same as example 2 in



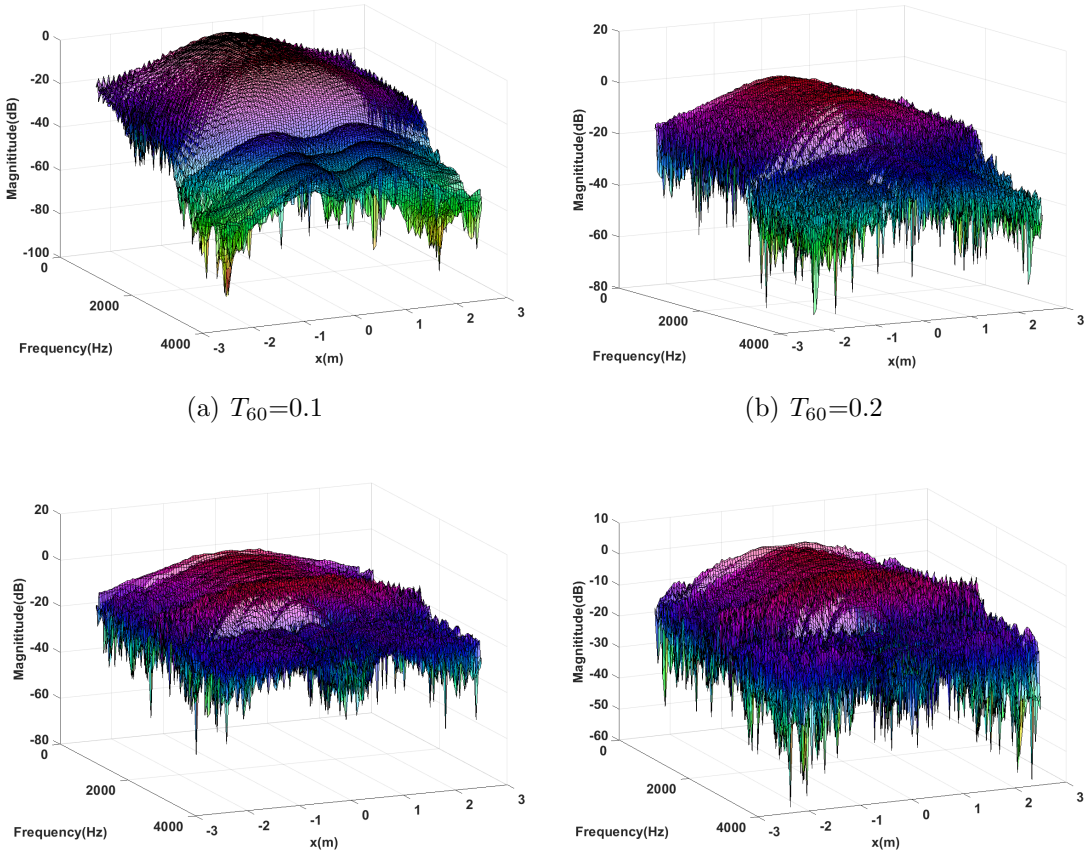
(a) f=1000Hz

(b) y=0m

(c) x=0.4m

Figure 3.8: Magnitude of the actual response and desired response of example 2 in free field

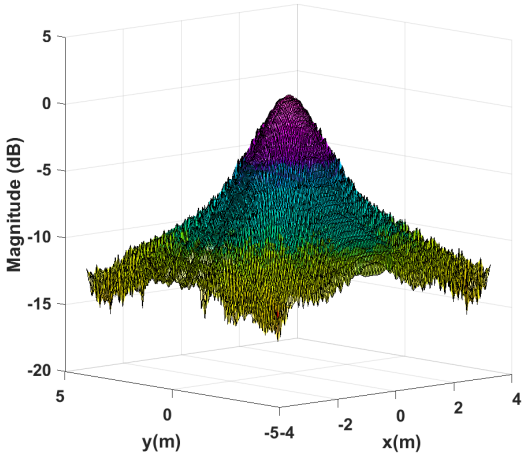
Section 3.7.1. The simulated room's settings are the same as example 1 in Section 3.7.2. When $T_{60} = 0.1$, the limit performance is -14.03dB. Fig.3.10 shows the actual response of the designed beamformer with a 16 filter length.



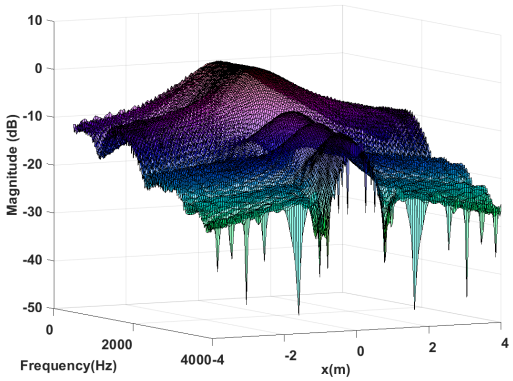
(a) T_{60}=0.1 (b) T_{60}=0.2

(c) T_{60}=0.3 (d) T_{60}=0.4

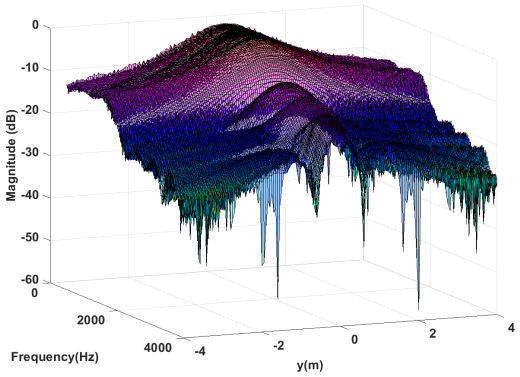
Figure 3.9: Magnitude of the actual response and desired response of example 1 in simulated room



(a) f=1000Hz



(b) y=0m



(c) x=0.4m

Figure 3.10: Magnitude of the actual response and desired response of example 2 in simulated room

Chapter 4

Design of Modulation-Domain Based Beamformers

4.1 Introduction

Speech signals can generally be considered low-frequency modulators modulating high-frequency carriers, similar to the amplitude modulation [38, 39]. It means speech information combines slow-changing modulations with a fast-changing carrier element. The speech intelligibility and quality are shown in the slow-changing modulation [46]. The concept of modulation domain is developed from STFT, which focuses on the modification synthesis framework. The modulation domain processing compactly represents the evolution of the spectral, temporal information of speech [4, 156, 91]. These findings led to the interest in applying the modulation domain as a substitute to the frequency domain for noise reduction, and the efficiency has been shown in many algorithms [108, 159, 128].

The aim of speech enhancement is to reduce noise from a polluted signal with intelligibility damage as small as possible. The carrier frequency modulator is the most crucial part of preserving linguistic information, which means we can process the speech more accurately. In the following, we briefly introduce some speech enhancement methods in the modulation domain. In [66, 41, 92], FIR bandpass filters

are proposed to enhance noisy speech, which is utilized on time trajectories of the short-term power spectrum of the polluted speech. This kind of bandpass filter has some drawbacks, such as the strict stationarity requirements for noise and speech signals, and lack of consideration of the noise properties in the design process. To overcome the disadvantages, the spectral subtraction method was proposed on the modulation spectrum, in which noises were considered to be quasi-stationary, instead of the conventional bandpass filters that assumed stationarity for all time [109]. Another well-developed processing technique in the acoustic domain, named Kalman filtering, was alsp applied [136, 159, 34]. Researchers show that the Kalman filter in the modulation domain performs better in processing non-stationary signals and estimating the phase and magnitude spectrum.

However, all these above methods are for single-channel speech enhancement. Compared with the single-channel algorithms, multichannel approaches can additionally apply spatial information to improve performance. One conventional technique for multichannel speech enhancement is beamforming, which generally exploits the correlation between multiple sensor signals to estimate the desired signal [110, 21, 130, 48, 64]. Inspired by the conventional methods, some researchers extended the single-channel speech enhancement algorithms in the modulation domain to multichannel cases. In [165], a novel multichannel Kalman filter was designed to operate in both STFT and modulation domains. With a certain assumption, the novel system is a concatenation of the MVDR beamformer and a single channel modulation domain Kalman filter. Thus, the novel system can jointly apply inter-frame temporal and inter-channel spatial correlation. Then, the system was further extended to parametric Kalman filtering, which contains a constant to control flexibly the speech enhancement behavior in each time-frequency bin [166]. Based on the minimum power distortionless response beamformer (MPDR), a preprocessing algorithm is proposed to filter the short-time spectral subtraction of the modulator.

It can improve the performance of the beamformer in the situation with the presence of reverberation [75].

Our work in this chapter considers speech enhancement algorithms in the modulation domain for the multi-channel case that extends the most widely used optimal beamformers, including LS [55] and SNR beamformers [31]. It is demonstrated in [168] that the LS method concentrates more on distortion control with a deficiency in noise suppression. In contrast, the SNR technique can consistently achieve a higher noise suppression level with a high distortion cost. We analyze the performance of the two techniques in the modulation domain by three performance measurement indicators and further exploit the combination of existing optimal designs. Results show that the proposed designs outperform the conventional beamformers in most indicators.

4.2 Problem Formulation

In this section, we first formulate the acoustic model in the frequency domain. Then, we further extend the formulated model into the modulation domain.

4.2.1 Signal Model in Frequency Domain

Suppose that there are m microphones randomly located in the acoustic network. Denote the noisy signal received by the ith sensor as

$$y_i(n) = x_i(n) + v_i(n), \quad \forall i \tag{4.1}$$

where $x_i(n)$ is the clean signal and $v_i(n)$ is the noise signal. The noise signal can be a mixture of several fixed point noises, and the noises can be coherent or incoherent. There is a filter after each sensor, and the parameter vector of the filter is given as $\boldsymbol{g}$ (Fig. 4.9). Then, after sum and delay operations, the output of the system is

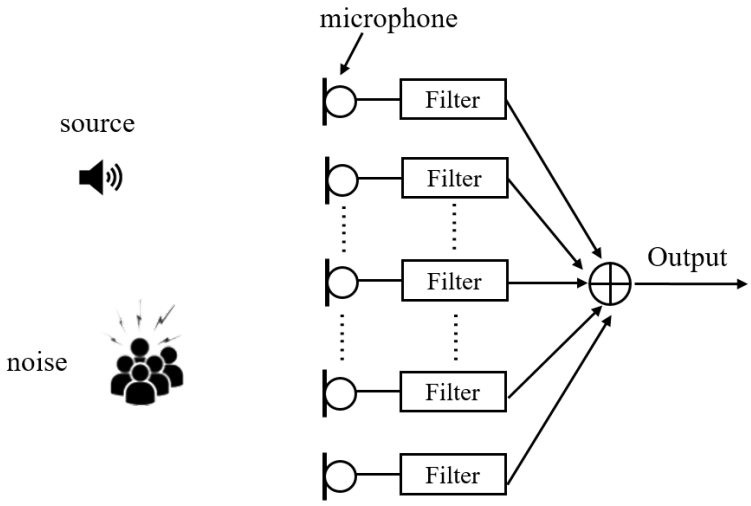


Figure 4.1: The framework of the adaptive beamformer

represented as

$$s(n) = \sum_{i=1}^{m} \sum_{k=0}^{N-1} \alpha_i(k) y_i(n-k) \tag{4.2}$$

where N is the filter length, i is the channel number, k represents tap number, and α is the coefficient of the filter.

As described in [168], we can equivalently depict the signal model in frequency domain by applying a sub-band scheme. In this case, for each frequency bin, there are specific weights. Then, we have the output of the frequency ω as

$$s^{(\omega)}(n) = \sum_{i=1}^{m} \alpha_i^{(\omega)} y_i^{(\omega)}(n), \tag{4.3}$$

where $y_i^{(\omega)}(n)$ is the narrow band signal for frequency ω, and $\alpha_i^{(\omega)}$ is the specific weight vector. The received signal is obtained by

$$y_i^{(\omega)}(n) = x_i^{(\omega)}(n) + v_i^{(\omega)}(n).$$

To calculate the weight vector, we can formulate the following problem for each

subband

$$\max_{\boldsymbol{\alpha}^{(\omega)}} \| \sum_{i=1}^{m} \alpha_i^{(\omega)} y_i^{(\omega)}(n) \| \quad \forall \omega. \tag{4.4}$$

4.2.2 Signal Model in Modulation Domain

The clean signal $x_i(n)$ for the ith sensor can be represented by the product of a modulator $m(n)$ and a carrier $c(n)$ as

$$x_i(n) = m(n)c(n). \tag{4.5}$$

The modulator is given by an envelope detector

$$m(n) = O_{env}\{x_i(n)\}, \tag{4.6}$$

where O_{env} represents the envelope detector operator. Equivalently, we can transfer the speech signal into frequency domain by STFT as

$$X_i(\omega, l) = M(\omega, l) * C(\omega, l), \tag{4.7}$$

where $*$ represents the convolution operator, ω is the frequency index and l is the time index. Accordingly, the envelope $M(\omega, l)$ in frequency domain is given by

$$\begin{aligned} M(\omega, l) &= O_{env}\{X_i(\omega, l)\} && (4.8) \\ &= O_{env}\{\sum_{n=0}^{N-1} x_i(n) w(n - lR_1) e^{-j\omega n}\}, && (4.9) \end{aligned}$$

where R_1 is the hop size and N is the window length, and l is the time index. Then, we can have the short time modulation spectrum of $s(n)$ as

$$\begin{aligned} X_i^{(mod)}(\omega, \gamma, k) &= F\{M(\omega, k)\} \\ &= F\{O_{env}\{X_i(\omega, l)\}\} \\ &= \sum_{l=0}^{L-1} |X_i(\omega, l)| w(l - kR_2) e^{-j\gamma k}, \end{aligned} \tag{4.10}$$

where F means the Fourier transform operator $|\cdot|$ means absolute value operator. In the modulation spectrum, the limited window has a hop size of R_2 and a length of L, and the modulation frequency is presented by $\gamma \in \gamma_0, \cdots, \gamma_{L-1}$.

Then, the short time modulation spectrum of the noisy signal presented in (4.1) for ω signal frequency and the γ modulation frequency at time k is

$$\begin{aligned} Y_i^{(mod)}(\omega, \gamma, k) &= F\{O_{env}\{Y_i(\omega, l)\}\} \\ &= \sum_{l=0}^{L-1} |Y_i(\omega, l)| w(l - kR_2) e^{-j\gamma k}. \end{aligned} \tag{4.11}$$

Similarly, the short time modulation spectrum of the noise signal presented in (4.1) for ω signal frequency and the γ modulation frequency at time k is

$$\begin{aligned} V_i^{(mod)}(\omega, \gamma, k) &= F\{O_{env}\{V_i(\omega, l)\}\} \\ &= \sum_{l=0}^{L-1} |V_{li}(\omega, l)| w(l - kR_2) e^{-j\gamma k}. \end{aligned} \tag{4.12}$$

For a specific acoustic frequency ω of the kth time instant, the output can be represented as

$$S_l^{(\omega,k)}(\gamma) = \sum_{i=1}^{m} \alpha_i^{(\omega,k)} Y_i^{(mod)}(\omega, \gamma, k). \tag{4.13}$$

Finally, we convert the enhanced signal back to the time domain. The framework of the modulation transform is given in Fig. 4.2. Then, the remaining problem is to design the parameter. In the next section, we propose two efficient design methods aiming to reduce noise.

4.3 Beamformer Design

This section presents two methods to design the filter weights with different criteria. The first method minimizes the error between the source and noisy signals. The

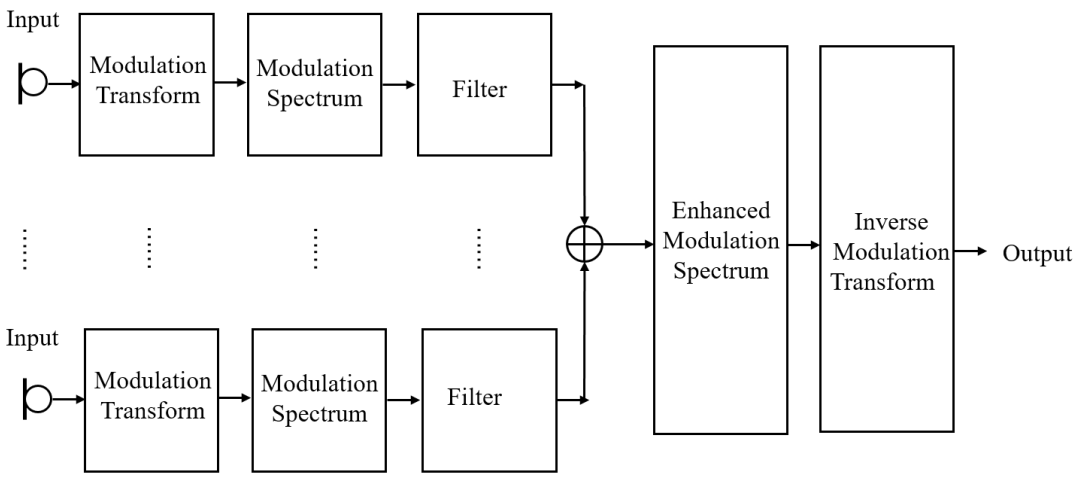


Figure 4.2: The framework of the modulation transform

filter weights can be obtained by solving the formulated least-squares optimal problem. The second criterion is the signal-to-noise power ratio (SNR), which measures the SNR of the output. The formulated optimal problem maximizes a ratio between two matrices, and a linear variable transformation can find the optimal solution.

4.3.1 Beamformer with Least Squares Criterion

The task of the filter design is to determine the weighting matrix $\boldsymbol{g}$. We can obtain an calibration sequence in a quiet environment. It can be considered as the reference source signal, which contains the temporal and spatial information of the source signal. Denoted the reference signal as $s_r(n)$, and the short time modulation spectrum of it is $S_r^{(\omega,k)}(\gamma)$. If a least squares (LS) criterion is used to measure the error for each modulation spectrum subset, we can formulate an objective as

$$\min_{\boldsymbol{\alpha}^{(\omega,k)}} \{\sum_{\gamma=0}^{L-1} [|S^{(\omega,k)}(\gamma) - X_r^{(\omega,k)}(\gamma)|^2]\}, \tag{4.14}$$

where $S_r^{(\omega,k)}(\gamma)$ is the output of the network for the ω frequency at time k in modulation domain, and $X_r^{(\omega,k)}(\gamma)$ is the modulation spectrum of the calibration signal.

The problem (4.14) can be solved as

$$\boldsymbol{\alpha}_{ls}^{(\omega,k)}(\gamma) = [\hat{\boldsymbol{R}}_{XX}^{(\omega,k)}(\gamma) + \hat{\boldsymbol{R}}_{YY}^{(\omega,k)}(\gamma)]^{-1}\hat{\mathbf{r}}^{(\omega,k)}(\gamma), \tag{4.15}$$

where $\hat{\boldsymbol{R}}_{XX}^{(\omega,k)}$ and $\hat{\mathbf{r}}^{(\omega,k)}$ are the source signal correlation estimations and $\hat{\boldsymbol{R}}_{YY}^{(\omega,k)}$ is the cross correlation matrix of the received signal. Let $\boldsymbol{R}^{(\omega,k)}(\gamma) = \hat{\boldsymbol{R}}_{XX}^{(\omega,k)}(\gamma) + \hat{\boldsymbol{R}}_{YY}^{(\omega,k)}(\gamma)$. Then, we have

$$\boldsymbol{\alpha}_{ls}^{(\omega,k)} = \boldsymbol{R}^{(\omega,k)}(\gamma)^{-1}\hat{\mathbf{r}}^{(\omega,k)}(\gamma), \tag{4.16}$$

The source correlation estimations can be pre-calculated in the calibration phase as

$$\begin{aligned}
\hat{\boldsymbol{R}}_{XX}^{(\omega,k)}(\gamma) &= \frac{1}{L}\sum_{\gamma=0}^{L-1}\boldsymbol{X}^{(\omega,k)^H}(\gamma)\boldsymbol{X}^{(\omega,k)^H}(\gamma) \\
\hat{\mathbf{r}}_{X}^{(\omega,k)}(\gamma) &= \frac{1}{L}\sum_{\gamma=0}^{L-1}\boldsymbol{X}^{(\omega,k)^H}(\gamma)X_r^{(\omega,k)^*}(\gamma)
\end{aligned}$$

where $*$ represents the conjugation operator, H represents the Hermitian transpose. The matrix $\boldsymbol{X}^{(\omega,k)}$ represents the sensor observation signals when the calibration source signal is active alone, denoted as

$$\boldsymbol{X}^{(\omega,k)} = [\boldsymbol{X}_1^{(\omega,k)}, \cdots, \boldsymbol{X}_m^{(\omega,k)}]^T.$$

The correlation matrix of the noisy signal can be directly calculated by the sensor received signal

$$\hat{\boldsymbol{R}}_{YY}^{(\omega,k)}(\gamma) = \frac{1}{L}\sum_{\gamma=0}^{L-1}\mathbf{Y}^{(\omega,k)^H}(\gamma)\mathbf{Y}^{(\omega,k)}(\gamma)$$

where

$$\mathbf{Y}^{(\omega,k)}(\gamma) = [\mathbf{Y}_1^{(\omega,k)}(\gamma), \cdots, \mathbf{Y}_m^{(\omega,k)}(\gamma)]^T.$$

For the problem (4.16), we can further decompose the correlation matrix into

$$\boldsymbol{R}^{(\omega,k)}(\gamma) = \beta\boldsymbol{R}^{(\omega,k)}(\gamma-1) + \mathbf{Y}^{(\omega,k)}(\gamma)\mathbf{Y}^{(\omega,k)^H}(\gamma) + \boldsymbol{Q}^{(\omega,k)^H}\Lambda^{(\omega,k)}\boldsymbol{Q}^{(\omega,k)} \tag{4.17}$$

where $\boldsymbol{Q}$ is the eigenvector matrix of the matrix $\hat{\boldsymbol{R}}_{XX}^{(\omega,k)}(\gamma)$ and Λ is the corresponding eigenvalue matrix

$$\begin{aligned} \boldsymbol{Q}^{(\omega,k)} &= [\mathbf{q}_1^{(\omega,k)}, \cdots, \mathbf{q}_m^{(\omega,k)}] \\ \Lambda^{(\omega,k)} &= diag([\lambda_1^{(\omega,k)}, \cdots, \lambda_m^{(\omega,k)}]). \end{aligned}$$

After a rank-1 approximation, we have

$$\boldsymbol{R}^{(\omega,k)}(\gamma) = \beta \boldsymbol{R}^{(\omega,k)}(\gamma-1) + \mathbf{Y}^{(\omega,k)}(\gamma)\mathbf{Y}^{(\omega,k)^H}(\gamma) + (1-\beta)\mathbf{q}_i^{(\omega,k)^H}\lambda_i^{(\omega,k)}\mathbf{q}_i^{(\omega,k)}, \quad (4.18)$$

where $i = 1 + (\gamma \ mod \ m)$. By the matrix inversion lemma, we can further compute the inversion of $\boldsymbol{R}^{(\omega,k)}(\gamma)$ iteratively

$$\boldsymbol{R}^{(\omega,k)}(\gamma)^{-1} = \boldsymbol{R}_1^{(\omega,k)} - \frac{\lambda_i^{(\omega,k)}(1-\beta)\boldsymbol{R}_1^{(\omega,k)}\mathbf{q}_i^{(\omega,k)}\mathbf{q}_i^{(\omega,k)^H}\boldsymbol{R}_1^{(\omega,k)}}{1+\lambda_i^{(\omega,k)}(1-\beta)\mathbf{q}_i^{(\omega,k)^H}\boldsymbol{R}_1^{(\omega,k)}\mathbf{q}_i^{(\omega,k)}} \quad (4.19)$$

where

$$\boldsymbol{R}_1^{(\omega,k)} = \beta^{-1}\boldsymbol{R}^{(\omega,k)}(\gamma-1)^{-1} - \frac{\beta^{-2}\boldsymbol{R}^{(\omega,k)}(\gamma-1)^{-1}\mathbf{Y}^{(\omega,k)}(\gamma)\mathbf{Y}^{(\omega,k)^H}(\gamma)\boldsymbol{R}^{(\omega,k)}(\gamma-1)^{-1}}{1+\beta^{-1}\mathbf{Y}^{(\omega,k)^H}(\gamma)\boldsymbol{R}^{(\omega,k)}\mathbf{Y}^{(\omega,k)^H}}.$$

Then, we have the weight vector

$$\boldsymbol{\alpha}_{ls}^{(\omega,k)}(\gamma) = \beta\boldsymbol{\alpha}_{ls}^{(\omega,k)}(\gamma-1) + (1-\beta)\boldsymbol{R}^{(\omega,k)}(\gamma)^{-1}\hat{\mathbf{r}}^{(\omega,k)}(\gamma). \quad (4.20)$$

The output of each modulation bin is

$$\boldsymbol{S}^{(\omega,k)}(\gamma) = \boldsymbol{\alpha}_{ls}^{(\omega,k)^H}(\gamma)\mathbf{Y}^{(\omega,k)}(\gamma). \quad (4.21)$$

4.3.2 Beamformer with SNR Criterion

The signal-to-noise power ratio (SNR) is an important criterion to measure the noise suppression level in speech enhancement, which is defined as

$$SNR = \frac{P_x}{P_y} \quad (4.22)$$

where P_x is the power of the clean signal and P_y is the power of the noisy signal. The output of the desired signal for frequency ω at time k is given as

$$\boldsymbol{\alpha}^{(\omega,k)^H} \boldsymbol{R}_{XX}^{(\omega,k)} \boldsymbol{\alpha}^{(\omega,k)}$$

and the output of the noisy signal is

$$\boldsymbol{\alpha}^{(\omega,k)^H} \boldsymbol{R}_{YY}^{(\omega,k)} \boldsymbol{\alpha}^{(\omega,k)}.$$

If SNR is considered as a criterion, the beamformer design problem becomes to maximize a ratio between two quadratic forms of positive definite matrices as

$$\boldsymbol{\alpha}_{snr}^{(\omega,k)} = \arg\max_{\boldsymbol{\alpha}^{(\omega,k)}} \{ \frac{\boldsymbol{\alpha}^{(\omega,k)^H} \boldsymbol{R}_{XX}^{(\omega,k)} \boldsymbol{\alpha}^{(\omega,k)}}{\boldsymbol{\alpha}^{(\omega,k)^H} \boldsymbol{R}_{YY}^{(\omega,k)} \boldsymbol{\alpha}^{(\omega,k)}} \}, \tag{4.23}$$

where $\boldsymbol{R}_{XX}^{(\omega,k)}$ is the correlation matrix of source signal in each sub modulation spectrum, and $\boldsymbol{R}_{YY}^{(\omega,k)}$ is the correlation matrix of the observed data. The correlation matric are calculated by

$$\begin{aligned} \boldsymbol{R}_{XX}^{(\omega,k)} &= E\{\boldsymbol{X}^{(\omega,k)}(\gamma)\boldsymbol{X}^{(\omega,k)}(\gamma)^H\} \\ \boldsymbol{R}_{YY}^{(\omega,k)} &= E\{\mathbf{Y}^{(\omega,k)}(\gamma)\mathbf{Y}^{(\omega,k)}(\gamma)^H\}, \end{aligned}$$

where $E\{\cdot\}$ represents the mean value operator and

$$\begin{aligned} \boldsymbol{X}^{(\omega,k)}(\gamma) &= [\boldsymbol{X}_1^{(\omega,k)}(\gamma), \cdots, \boldsymbol{X}_m^{(\omega,k)}(\gamma)]^T, \\ \mathbf{Y}^{(\omega,k)}(\gamma) &= [\mathbf{Y}_1^{(\omega,k)}(\gamma), \cdots, \mathbf{Y}_m^{(\omega,k)}(\gamma)]^T. \end{aligned}$$

The problem (4.23) is a maximization of a ratio between two positive definite matrices with quadratic forms, which is related to the generalized eigenvector problem. We can obtain an optimal solution by solving

$$\mathbf{v}_{opt}^{(\omega,k)} = \arg\max_{\mathbf{v}} \frac{\mathbf{v}^{(\omega,k)^H} \boldsymbol{R}_{YY}^{(\omega,k)^{-\frac{1}{2}^H}} \boldsymbol{R}_{XX}^{(\omega,k)} \boldsymbol{R}_{YY}^{(\omega,k)^{-\frac{1}{2}}} \mathbf{v}^{(\omega,k)}}{\mathbf{v}^{(\omega,k)^H} \mathbf{v}^{(\omega,k)}}, \tag{4.24}$$

where $\mathbf{v}_{opt}^{(\omega,k)}$ is the eigenvector corresponding to the largest eigenvalue satisfying the following equation

$$\boldsymbol{R}_{YY}^{(\omega,k)^{-\frac{1}{2}^H}} \boldsymbol{R}_{XX}^{(\omega,k)} \boldsymbol{R}_{YY}^{(\omega,k)^{-\frac{1}{2}}} \mathbf{v}_{opt}^{(\omega,k)} = \lambda \mathbf{v}_{opt}^{(\omega,k)}. \tag{4.25}$$

Accordingly, the optimal solution of problem (4.23) can be obtained by the inverse of the linear variable transformation

$$\boldsymbol{\alpha}_{snr}^{(\omega,k)} = \boldsymbol{R}_{YY}^{(\omega,k)^{-\frac{1}{2}}} \mathbf{v}_{opt}^{(\omega,k)}. \tag{4.26}$$

Then, the problem becomes to find the eigenvector $\mathbf{v}_{opt}^{(\omega,k)}$, and a widely used method to solve problem (4.25) is an iterative power method. The update rule follows

$$\mathbf{v}_{opt}^{(\omega,k)}(\gamma+1) = \frac{\boldsymbol{R}_{YY}^{(\omega,k)^H} \boldsymbol{R}_{XX}^{(\omega,k)} \mathbf{v}_{opt}^{(\omega,k)}(\gamma)}{\| \boldsymbol{R}_{YY}^{(\omega,k)^H} \boldsymbol{R}_{XX}^{(\omega,k)} \mathbf{v}_{opt}^{(\omega,k)} \|}. \tag{4.27}$$

The initial guess of $\mathbf{v}_{snr}^{(\omega,k)}(0)$ can be an arbitrary vector. Then, the remaining problem is to find out the inverse of $\boldsymbol{R}_{YY}^{(\omega,k)}$ with a fast speed. One efficient method is Cramer's rule, which contains three main steps

- Formulate a cofactor matrix by calculating co-factors of $\boldsymbol{R}_{YY}^{(\omega,k)}$.
- Compute the determinant of the cofactor matrix and form a matrix denoted as A.
- Multiply A by the determinant's reciprocal.

This method outperforms the Gaussian elimination method, especially when the given matrix has a small dimension. As the matrix in each modulation bin is relatively small, we apply Cramer's rule.

4.3.3 Hybrid Method

We can compute the filter weights separately from the last two parts by LS and SNR methods. As indicated in [168], the LS method focuses more on speech distortion control, while the SNR technique concentrates more on noise suppression. As each method has its own property when enhancing the speech, we attempt to combine the two methods to adjust the speech quality in different aspects. We first separately calculate two sets of optimal coefficients by LS and SNR methods. Because of the linearity of the filtering system , we try to give a hybrid weights of these two techniques, which can adjust the performances according to the decision makers' requirements. However, the different ranges of the coefficients may result in a disparity when combining the weights. Thus, before the combination, we normalized the coefficients to their maximum (i.e., the largest coefficient value) in each set of coefficient individually, so that the coefficients are ranged from zero to one. A framework of the hybrid system is given in Fig. 4.3.

Let the normalized optimal coefficients of the ωth frequency bin at time k be $\boldsymbol{\alpha}_{ls}^{(\omega,k)}$ and $\boldsymbol{\alpha}_{snr}^{(\omega,k)}$. The hybrid weight vector is given as

$$\boldsymbol{\alpha}_{hybrid}^{(\omega,k)} = a\boldsymbol{\alpha}_{ls}^{(\omega,k)} + (1-a)\boldsymbol{\alpha}_{snr}^{(\omega,k)}, \tag{4.28}$$

where a is a real value from 0 to 1. Then, the output of the hybrid method for a specific acoustic frequency ω of the kth time instant is

$$Y^{(\omega,k)}(\gamma) = \sum_{i=1}^{m} \alpha_i^{(\omega,k)} X_i^{(mod)}(\omega,\gamma,k), \tag{4.29}$$

where $\alpha_i^{(\omega,k)}$ is given by (4.28).

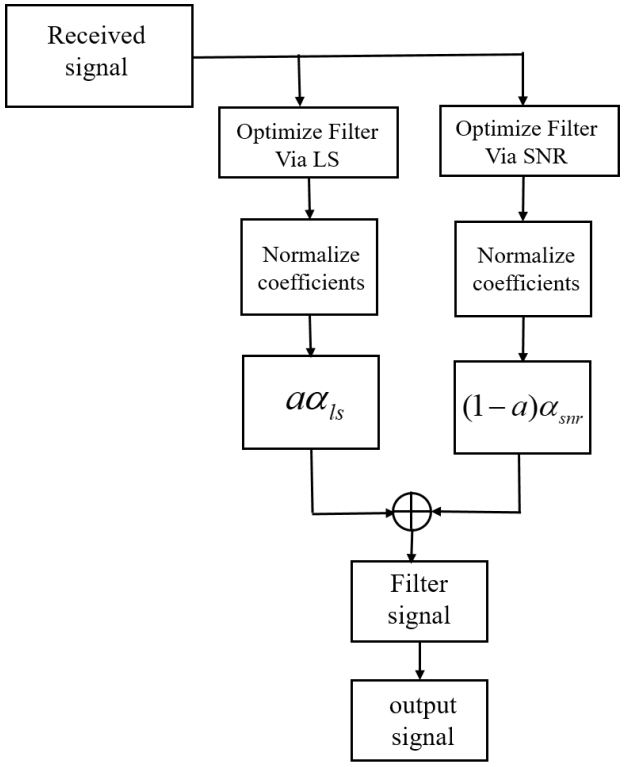


Figure 4.3: The framework of the hybrid system

4.4 Performance Measurement Indicators

To evaluate the performance of the beamformers designed by different methods, in this section, we discuss performance measurement indicators. Generally speaking, the assessment of speech quality includes subjective and objective evaluation. Subjective evaluation relies on listeners' subjective listening tests, which could be pretty accurate but costly and time-consuming. Objective evaluation measures the numerical distance between the reference signal and the processed signals, predicting the speech quality with high correlation [89]. One of the most popular objective measure for speech enhancement is the STOI, which are closely related to human auditory perception and widely used in speech enhancement as evaluation criteria. STOI is the latest popular measure, computing the correlation of short-time temporal envelopes.

Another two performance measures are noise suppression, considered an engineering design, and speech distortion, thought a quality design. The output signal requires the highest possible SNR for engineering design, which describes noise suppression capability. As for the quality design, it requires to protect the perceptual features, which means minimal speech quality degradation. In fact, it isn't easy to optimize the engineering and quality design simultaneously. The reason is that when improving SNR, the speech quality continuously degrades, leading to a natural trade-off. In particular, we will study for two-stream streams of applications. The first one is on signal quality for human perception. In this case, speech quality is essential, and higher objective evaluation scores are required. The other important application is to feed the beamformer output to speech recognition engines. For this application, minimizing noise could be more critical, and SNR becomes a more important criterion.

In this section, four indicators are given to evaluate the output signal. The first criterion tests the distortion between the original speech signal and the output signal, and the second concentrates on noise suppression. Another criterion is STOI, which measure the numerical distance between the reference signal and the processed signals, predicting the speech quality with high correlation [89].

4.4.1 Signal Distortion and Noise Suppression

Firstly, we introduce an indicator to measure the signal distortion, and a normalized quantity can be given as:

$$S_d = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mid C_d \hat{P}_{y_x}(\omega) - \hat{P}_{s_x}(\omega) \mid d\omega \tag{4.30}$$

where $\hat{P}_{y_x}(\omega)$ is a mean spectral power of the clean signal received by sensors and $\hat{P}_{s_x}(\omega)$ is the mean spectral power of the output when the desired signal is active

alone. The constant C_d can be defined as

$$C_d = \frac{\int_{-\pi}^{\pi} \hat{P}_{s_x}(\omega)d\omega}{\int_{-\pi}^{\pi} \hat{P}_{y_x}(\omega)d\omega}. \tag{4.31}$$

Another indicator is given to measure the noise suppression, and this measure is described as

$$N_S = C_s \frac{\int_{-\pi}^{\pi} \hat{P}_{s_n}(\omega)d\omega}{\int_{-\pi}^{\pi} \hat{P}_{y_n}(\omega)d\omega} \tag{4.32}$$

where

$$C_s = \frac{1}{C_d}.$$

In (4.32), the $\hat{P}_{y_n}$ represents the spectral power of the noise signal received by microphones and $\hat{P}_{s_n}$ is the spectral power of the output when only the noise signal is active.

4.4.2 A Short-Time Objective Intelligibility Measure(STOI)

STOI is the latest popular objective machine-driven intelligibility measure, which evaluates the correlation of short-time temporal envelopes by a function of a time-frequency-dependent intermediate intelligibility measure. It compares the temporal envelopes of the reference and polluted signal in short-time regions by a correlation coefficient. Experiments showed that STOI correlates better with speech intelligibility than other reference objective intelligibility models. Thus, the STOI is applied in evaluating the performance in speech intelligibility [126], [33].

Denote the clean and polluted speech in the time domain as $\mathbf{s}$ and $\mathbf{y}$. As STOI is based on short time segments, we denote the short-time temporal envelope for clean and polluted signals as $\mathbf{s}_{j,m}$ and $\mathbf{y}_{j,m}$, respectively. The intermediate intelligibility is then defined as the sample correlation coefficient between the two vectors, denoted

as

$$d_{j,m} = \frac{(s_{j,m} - \mu_{s_{j,m}})^T \tilde{\mathbf{y}}_{j,m}}{\|\mathbf{s}_{j,m} - \mu_{s_{j,m}}\| \|\tilde{\mathbf{y}}_{j,m} - \mu_{\tilde{y}_{j,m}}\|},$$

where $\mu_{(\cdot)}$ is the mean value of the corresponding vector, and $\tilde{\mathbf{y}}_{j,m}$ is the corresponding modulation vector of

$$\tilde{Y}_j(m) = \min(Y_j(m), 6.33 \cdot \frac{\|\mathbf{y}_{j,m}\|}{\|\mathbf{s}_{j,m}\|} S_j(m)),$$

where $S_j(m)$ and $Y_j(m)$ are the time-frequency cell amplitudes of the clean and polluted speech. The final average measure of all bands and frames is given by

$$d = \frac{1}{JM} \sum_{j,m} d_{j,m} \tag{4.33}$$

where M represents the total number of frames and J the number of one-third octave bands. The details of the calculation of the STOI can be found in [141].

4.5 Experimental Results

We evaluate the performances of different speech enhancement methods in this section. The real data is recorded by Sven Nordholm in Curtin University [49]. The acoustic network contains eight microphones, which have been calibrated before use. The sound card connected to the speakers is Delta 1010LT, and the sound driver is ASIO Hammerfall DSP. We choose five clean signals from the NOIZEUS database, and a babble noise signal from NOISEX-92 database. All speeches are recorded with a sampling rate of 48000Hz, and then the signals are resampled to 8000Hz for speech enhancement. We recorded the noise signal and the clean signal separately, and the noisy signal is generated by adding the noise to the clean signal, and we choose one single channel observation as the reference channel observation. We first analyze the effect of a different number of frequency bins, modulation bins, and different SNR

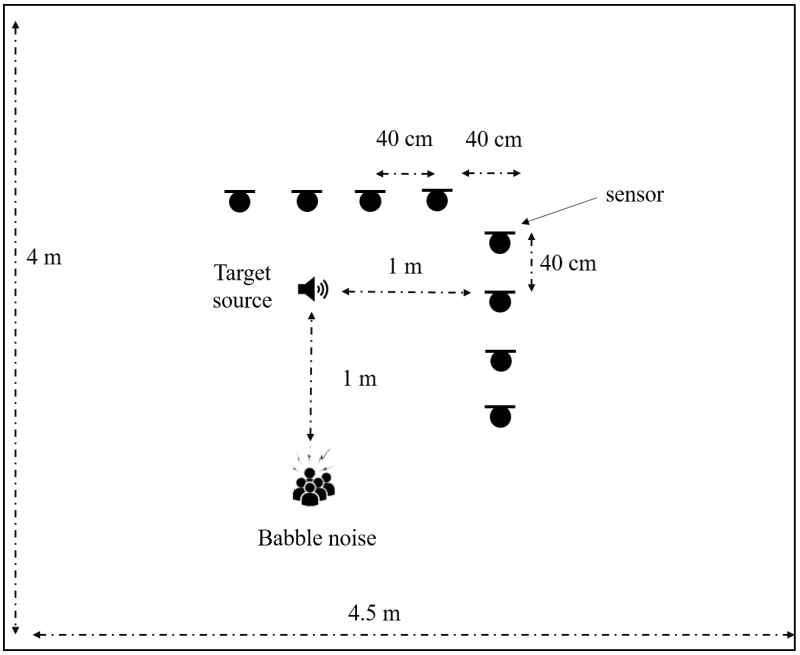


Figure 4.4: The configuration of the experiments

levels on the beamforming performances. Then, we test the performances of the hybrid method, and the results prove that it can trade off signal distortion and noise suppression.

4.5.1 Effect of Different Number of Frequency Bin and Modulation Spectrum Length

This part compares the LS and SNR methods' experimental results in the frequency and modulation domains. There are total 8 sensors ranged L shape in the network. The signals are recorded in a room with $4 \times 4.5 \times 2.4m$ $(L \times W \times H)$, and the height of the sensor are fixed at around 1.5m above the floor. The locations of the sources and sensors are given in Fig. 4.4.

The desired signal and the interfering signal are recorded separately. The noisy signal is created by adding the interfering signal to the desired signal. The noisy signal is given by adding the noise signal to the clean signal, denoted as

$$y(n) = x(n) + \alpha v(n),$$

where $x(n)$ is clean signal, $v(n)$ is the noise signal, and α is a scalar value. To evaluate the power of noise signal, we introduce SNR in decibels(dB), defined as

$$SNR_{dB} = 10\log_{10}(\frac{p_x}{p_v}),$$

where p_x is the power of the clean signal, and p_v is the power of the noise signal. According to the definition of SNR, the value of α can then be obtained by

$$\alpha = \sqrt{\frac{\|x(n)\|_2^2}{\|SNR \cdot v(n)\|_2^2}},$$

where $\|\cdot\|_2$ represents the Euclidean norm. In this part, the SNR of the noisy signal is $0dB$.

The polluted signal is filtered by the LS method in the frequency domain (F-LS) and the modulation domain (M-LS). We further apply the SNR method in the frequency domain (F-SNR) and the modulation domain (M-SNR). Then, performances of the four methods by the four indicators presented in the previous section are compared, including signal distortion S_d, and noise suppression N_s.

The number of frequency bins is directly related to the number of parameters of the beamformer. We designed various beamformers with different frame sizes to test the performance for speech enhancement based on the different frequency bins. Table 4.2 shows the results of the different frequency bins from 64 to 512, which is the size of the first STFT that we use to transform the time domain data to the frequency domain. All the results of this section show average values. Also, we compare the results in the frequency domain with the number of subbands from 64 to 512. Then, we further evaluate the performance of different lengths of modulation spectrum, which is the number of second STFT in the modulation operator. When investigating the effect of the frequency bins, we set the fixed number of modulation spectrum bins as 4. When investigating the effect of the number of modulation bins,

we set the fixed number of the frequency bins as 512.

For the LS method, the beamformers designed in the modulation domain perform better in speech distortion, which means less quality degradation. For the SNR method, the modulation domain filters considerably improve noise suppression. It can be seen that when the number of frequency bins increases, the noise suppression increases. The increasing length of the modulation spectrum can also help suppress noise. To analyze the differences between the four methods, we give the time domain signals and spectrograms of one speech as an example. The time-domain signal and spectrogram analysis of clean speech, noisy speech, and denoised signals with 512 frequency bin and 4 modulation bin are given in Fig. 4.5 and Fig. 4.6, where 'Fb' represents the number of frequency bins, and 'Mb' means the number of modulation bins.

Table 4.1: Effect of the number of FFT in frequency domain

Noisy signal (SNR:0dB)				
No. frequency bin	F-LS		F-SNR	
	N_s	S_d	N_s	S_d
64	19.38	-10.77	32.99	-3.35
128	19.56	-9.38	34.44	-3.21
256	19.79	-8.70	35.46	-3.18
512	19.95	-7.97	36.45	-3.17

4.5.2 Effect of Different Noise Levels

In this part, we further investigate the effect of SNR, and five SNR levels are chosen as -10dB, -5dB, 0dB, 5dB, 10dB. The recording setting is the same as the last part. For the methods in the modulation domain, we set the fixed first STFT number as 512 and the second STFT number as 4. For the frequency domain methods, the size of subbands is 512. We also compare the proposed methods with a modulation domain based Wiener filter.

Table 4.2: Effect of the number of FFT in modulation domain

Noisy signal (SNR:0dB)				
No. frequency bin	M-LS		M-SNR	
(4 modulation bin)	N_s	S_d	N_s	S_d
64	6.24	-15.06	28.83	-3.55
128	10.72	-13.67	33.72	-3.53
256	16.82	-11.90	38.11	-3.64
512	21.32	-10.05	41.07	-3.78
No. modulation bin	M-LS		M-SNR	
(512 frequency bin)	N_s	S_d	N_s	S_d
4	21.32	-10.05	41.07	-3.78
8	21.91	-9.09	42.27	-3.79
16	20.42	-9.22	43.86	-3.74
32	18.20	-9.40	46.22	-3.69

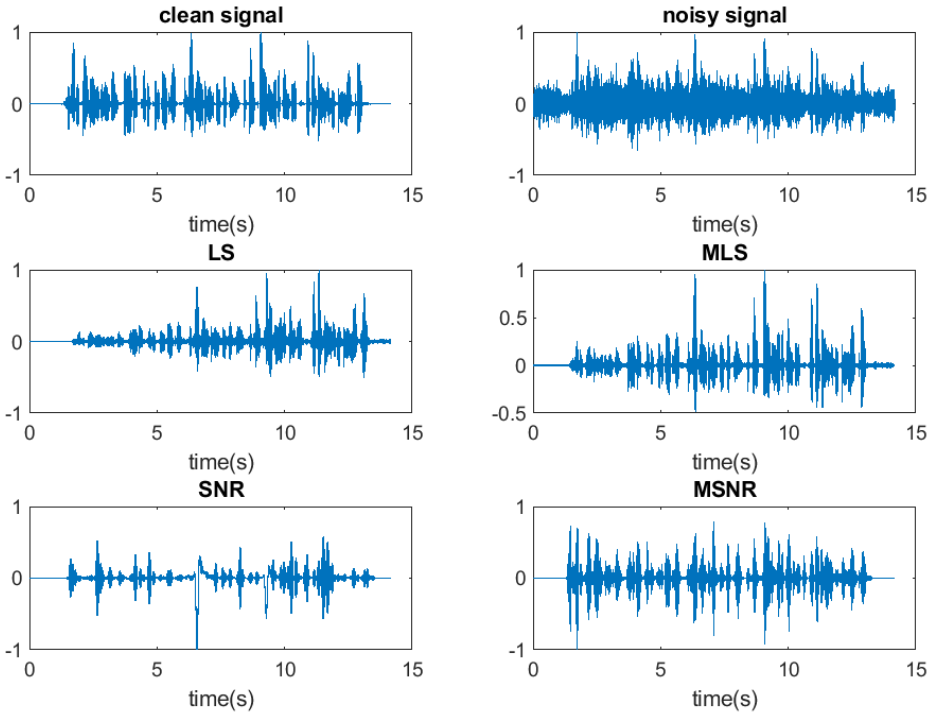


Figure 4.5: Clean speech, noisy speech and denoised speech signal in time domain(Fb=512,MB=4)

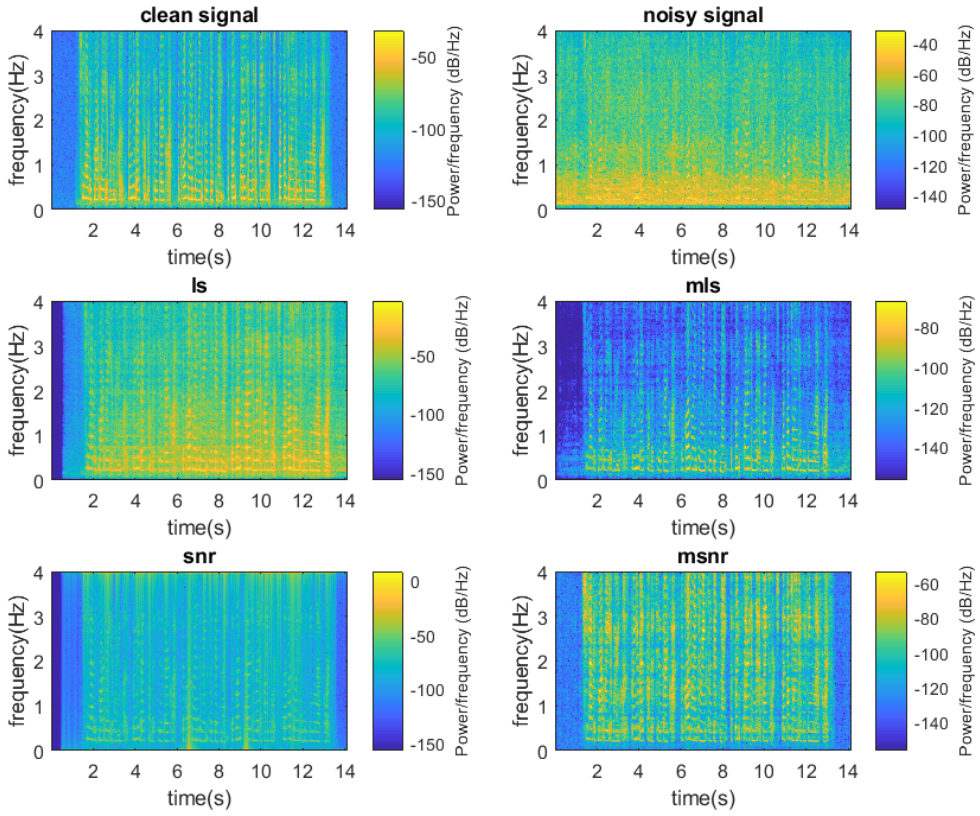


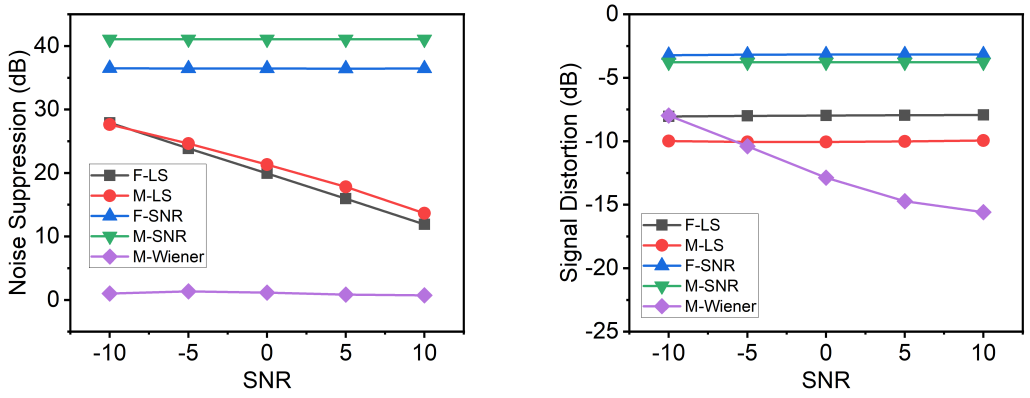
Figure 4.6: Spectrogram analysis of clean speech, noisy speech and denoised speech signal (Fb=512,MB=4)

Table 4.3 shows the results with different noise levels. We can see that the performances of the SNR-based methods are similar for all the noise levels, which means the noise levels have a more negligible effect on the performances of the F-SNR and M-SNR methods. For the LS-based methods, when the noise levels go up, the noise suppression increases with a similar distortion. Similar to the last part's results, M-SNR generally has better noise suppression than the other three methods, while the M-LS method has minor noise distortion. The speech quality is protested well in the M-LS method. However, for the modulation domain-based Wiener method, although the signal distortion is less than the other methods, the noise suppression

is little.

Table 4.3: Effect of the SNR (Fb=512,Mb=4)

SNR		F-LS	M-LS	F-SNR	M-SNR	M-Wiener
-10	N_s	27.91	27.64	36.50	41.07	1.00
	S_d	-8.04	-9.98	-3.22	-3.78	-7.98
-5	N_s	23.86	24.63	36.46	41.08	1.34
	S_d	-8.00	-10.06	-3.18	-3.78	-10.39
0	N_s	19.95	21.32	36.45	41.07	1.16
	S_d	-7.97	-10.05	-3.17	-3.78	-12.87
5	N_s	15.96	17.81	36.44	41.06	0.85
	S_d	-7.95	-10.01	-3.17	-3.78	-14.73
10	N_s	11.91	13.66	36.45	41.07	0.71
	S_d	-7.93	-9.94	-3.17	-3.78	-15.59



(a) Noise suppression (b) Signal distortion

Figure 4.7: Noise suppression and signal distortion of different noise levels

4.5.3 Results of Hybrid Design Method

From the previous results, we can see that the optimal filter weights for LS and SNR have unique properties in signal distortion and noise suppression. To find a balance between the two criteria, we try to form a linear combination of the two optimal weights. The hybrid weights are given as

$$\boldsymbol{\alpha}_{hybrid} = a\boldsymbol{\alpha}_{ls} + (1-a)\boldsymbol{\alpha}_{snr}, \tag{4.34}$$

where $\boldsymbol{\alpha}_{ls}$ is the normalized coefficients of M-LS method, $\boldsymbol{\alpha}_{snr}$ is the normalized coefficients of M-SNR method, and a is a scalar value from 0 to 1. We give the results of different a with different size of frequency bins in Table 4.4. It can be seen that with the increase of a, the signal distortion decreases, while the noise suppression goes down. The less signal distortion and the more noise suppression, the better speech quality. Comparing with the single objective based method, this hybrid method allows a trade-off between the noise suppression and signal distortion. For the LS method, it has less signal distortion. However, for the SNR method, it performs better on the noise suppression. If two criteria are comprised and a suitable parameter is chosen, we can obtain a beamformer with both favourable signal distortion and noise suppression. Designers can choose a proper α according to their specific requirements.

Table 4.4: The results of hybrid method

Fb=128,Mb=4			Fb=256,Mb=4			Fb=512,Mb=4		
a	N_s	S_d	a	N_s	S_d	a	N_s	S_d
0	34.05	-5.29	0	39.09	-5.39	0	41.85	-5.58
0.1	33.28	-5.27	0.1	38.18	-5.36	0.1	40.93	-5.55
0.2	30.54	-5.25	0.2	35.47	-5.35	0.2	38.37	-5.51
0.3	27.13	-5.25	0.3	32.12	-5.35	0.3	35.15	-5.49
0.4	23.71	-5.30	0.4	28.75	-5.40	0.4	31.85	-5.52
0.5	20.42	-5.44	0.5	25.48	-5.56	0.5	28.60	-5.64
0.6	17.25	-5.83	0.6	22.31	-5.99	0.6	25.46	-6.04
0.7	14.30	-6.78	0.7	19.40	-7.09	0.7	23.64	-7.08
0.8	12.96	-8.94	0.8	18.55	-9.49	0.8	22.70	-9.32
0.9	11.92	-12.58	0.9	18.27	-12.76	0.9	22.25	-12.03
1	11.90	-16.55	1	18.04	-14.14	1	21.81	-12.89

4.5.4 Results of STOI values

This part further evaluates a widely used objective evaluation measure, STOI. We add the noise with different SNR conditions, and process the noisy signals by the four methods with 512 frequency bins and 4 modulation bins. Five speech signals are

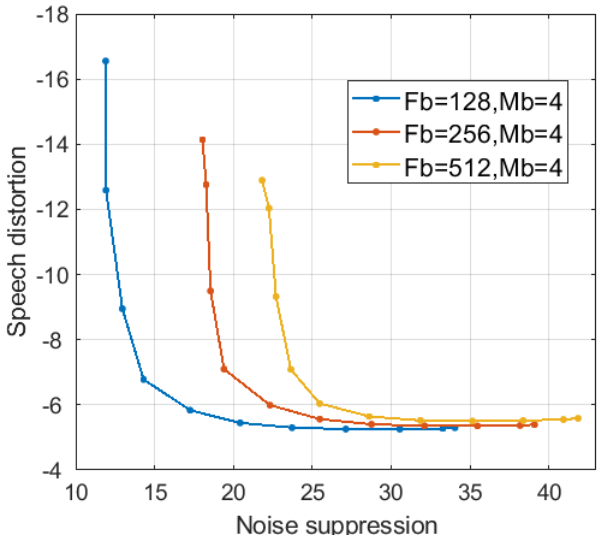


Figure 4.8: Trade off between the noise suppression and speech distortion

tested, and the same babble noise is added on the clean speeches. Table 4.5 give the average STOI values for test samples corrupted under different SNR conditions and enhanced by different methods. As observed, the modulation domain filters always have a better STOI than the frequency domain methods, and M-LS has the most significant improvements in STOI cores. Significantly when the number of frequency bins is large, the SOTI values decrease severely for the frequency-based methods.

Table 4.5: STOI values with different SNR (Fb=512,Mb=4)

SNR	noisy	F-LS	M-LS	F-SNR	M-SNR
-10	0.46	0.06	0.80	0.02	0.74
-5	0.56	0.06	0.85	0.02	0.74
0	0.68	0.05	0.88	0.02	0.76
5	0.79	0.05	0.91	0.02	0.77
10	0.88	0.04	0.94	0.03	0.79

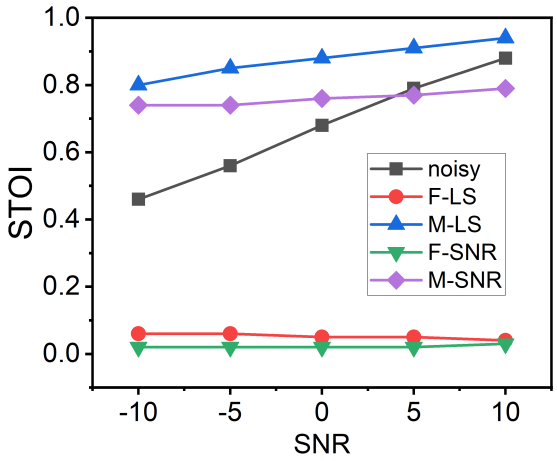


Figure 4.9: STOI values of different noise levels

Chapter 5

Conclusions and Suggestions for Future Research

In this chapter, a summary of this thesis will be given, and some possible works for future research will be shown.

5.1 Conclusions

This thesis mainly focuses on the beamforming design problem in the wireless acoustic network and some related problems during the design procedure, such as the sensor localization problem. The transfer function from the source to the sensor is sensitive to the microphone locations, and even minor errors can severely affect the beamformers' performances. Thus, an efficient sensor localization algorithm is necessary, and Chapter 2 proposed the relaxation model-based methods to estimate the microphone locations and structure. Then, with the given array configuration, beamforming algorithms can be performed. Chapter 3 proposed a design method for the fixed beamformers, which requires the transfer function. Compared with the previous FIR-based methods, the proposed IIR-based algorithm is much more efficient and can achieve the same performance with less filter length. As sensors in the WSNs are always miniaturized, this improvement in efficiency is essential. Then, some adaptive algorithms are proposed, which have no requirements on prior

knowledge of the sensor locations, also known as blind estimation algorithms. These methods are applied in the modulation domain, and four indicators are used to test the performances.

The three research works in this thesis have been concluded as follows.

1. Chapter 2 presents a method to calibrate the microphone locations in a distributed acoustic network. The proposed method formulates a linear optimization problem by an SDP-SOCP relaxation model with TDOA information, which can be solved in a polynomial time. Then, we study the characteristics of the solution to the relaxation model, and several theorems are given to check the correctness of the model. In addition, two simple offsets algorithms are proposed to eliminate the random noise in real data. Numerical results show that the mixed relaxation model is more accurate than the SDP or SOCP relaxation models. We further applied the proposed model in simulation rooms and real situations. Results show that the sensor locations can be estimated accurately, and the offset algorithms can improve the accuracy.

2. In Chapter 3, an IIR-based broadband beamforming design method is investigated. Since there is a feedback section in an IIR filter that results in an intractable stability problem, a specific structure in which all the elements share the same feedback section is proposed. This structure efficiently simplifies the stability problem. We can ensure the whole system is stable by adding constraints on the poles of the feedback part. Based on this structure, an optimal design beamformer method is given, and the performance limit of this design method has been studied. Furthermore, according to the simulation results, it has been proven that our method can converge much faster to the same limiting cost function value compared with an FIR-based beamformer.

3. Chapter 4 presents two popular beamformer design methods in the modulation

domain: the least square and maximization SNR. We compared the methods in the modulation domain and frequency domain by three indicators. Experimental results have shown that modulation domain-based methods outperform the frequency-domain-based techniques, especially in SOTI, which is an essential measurement for human perception. In addition, SNR-based methods have higher noise suppression in the modulation domain than the frequency domain, and noise suppression is a critical indicator in speech recognition engines. In contrast, the LS-based method has better speech quality with less noise suppression. Then, a hybrid method is given to trade off the speech quality and noise suppression. Users can choose suitable methods according to different applications.

5.2 Suggestions for Future Research

Speech signal processing in wireless acoustic networks is a rich research subject. Here are some possible directions for the future works.

- **Estimation of TOA by TDOAs**

  Compared with TDOAa, TOAs are much powerful measurements in sensor localization problem. However, acoustic distributed network can only obtain TDOAs indirectly from various sound source anchors. Thus, if we can estimate the TOAs from the obtained TDOAs, more powerful sensor localization algorithms can be applied.

- **Calibration of impulse response**

  Except for the sensor localization problem, there are still some other problems that can affect the performances of the beamformers in the wireless acoustic networks, such as impulse responses of microphones. The impulse responses

are required in many beamformer design methods, such as MVDR and LCMV. In addition, the room impulse responses used in Chapter 2 are generated by a room simulator. If we can obtain a more accurate impulses response in actual situations, the beamformer designed in Chapter 2 can be more robust in real applications.

- **Design of distributed beamformer algorithm**

  For example, the adaptive algorithms proposed in Chapter 4 are classic centralized adaptive beamformers, which require a central processing unit to calculate the data received by all the nodes in the network. It could be undesirable in actual applications because of privacy problems, transmission range, and battery limitations. Thus, distributed algorithms should be considered in future work, decreasing the energy for transmitting data to the center fusion. The distributed algorithms should enable the processing load distribution over different nodes, as each node only contains partial data with limited energy supplies. In addition, it is possible to use compressed sensing in transmitting data and derive algorithms with optimality in the sense that distributed methods have the same beamformer outputs as its centralized counterpart method.

Bibliography

[1] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen. Spatially correct rate-constrained noise reduction for binaural hearing aids in wireless acoustic sensor networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2731–2742, 2020.

[2] L. Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.

[3] A. Asaei, N. Mohammadiha, M. J. Taghizadeh, S. Doclo, and H. Bourlard. On application of non-negative matrix factorization for ad hoc microphone array calibration from incomplete noisy distances. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2694–2698. IEEE, 2015.

[4] L. Atlas and S. A. Shamma. Joint acoustic and modulation frequency. *EURASIP Journal on Advances in Signal Processing*, 2003(7):1–8, 2003.

[5] K. L. Bell, Y. Ephraim, and H. L. Van Trees. A bayesian approach to robust adaptive beamforming. *IEEE Transactions on Signal Processing*, 48(2):386–398, 2000.

[6] J. Benesty, S. Makino, and J. Chen. *Speech enhancement*. Springer Science & Business Media, 2006.

[7] A. Bertrand. Applications and trends in wireless acoustic sensor networks: A signal processing perspective. In *2011 18th IEEE symposium on communications and vehicular technology in the Benelux (SCVT)*, pages 1–6. IEEE, 2011.

[8] A. Bertrand and M. Moonen. Distributed adaptive estimation of correlated node-specific signals in a fully connected sensor network. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2053–2056, 2009.

[9] A. Bertrand and M. Moonen. Robust distributed noise reduction in hearing aids with external acoustic sensor nodes. *EURASIP Journal on Advances in Signal Processing*, 2009:1–14, 2009.

[10] A. Bertrand and M. Moonen. Distributed adaptive node-specific signal estimation in fully connected sensor networks—part I: Sequential node updating. *IEEE Transactions on Signal Processing*, 58(10):5277–5291, 2010.

[11] A. Bertrand and M. Moonen. Distributed LCMV beamforming in wireless sensor networks with node-specific desired signals. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2668–2671, 2011.

[12] A. Bertrand and M. Moonen. Distributed node-specific LCMV beamforming in wireless sensor networks. *IEEE Transactions on Signal Processing*, 60(1):233–246, 2012.

[13] A. Bertrand and M. Moonen. Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission. *IEEE Transactions on Signal Processing*, 61(13):3447–3459, 2013.

[14] M. Biggs. Constrained minimization using recursive equality quadratic programming. *Numerical Methods for Nonlinear Optimization*, pages 411–428, 1972.

[15] S. T. Birchfield. Geometric microphone array calibration by multidimensional scaling. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, volume 5, pages V–157. IEEE, 2003.

[16] S. T. Birchfield and A. Subramanya. Microphone array position calibration by basis-point classical multidimensional scaling. *IEEE transactions on Speech and Audio Processing*, 13(5):1025–1034, 2005.

[17] H. G. Bock and K.-J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. *IFAC Proceedings Volumes*, 17(2):1603–1608, 1984.

[18] R. Bovenkamp, F. Kuipers, and P. V. Mieghem. Gossip-based counting in dynamic networks. In *International conference on research in networking*, pages 404–417. Springer, 2012.

[19] M. Brandstein and D. Ward. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Science & Business Media, 2001.

[20] D. Brandwood. A complex gradient operator and its application in adaptive array theory. In *IEE Proceedings H-Microwaves, Optics and Antennas*, volume 130, pages 11–16. IET, 1983.

[21] B. Breed and J. Strauss. A short proof of the equivalence of LCMV and GSC beamforming. *IEEE Signal Processing Letters*, 9(6):168–169, 2002.

[22] B. Carlson. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Transactions on Aerospace and Electronic Systems*, 24(4):397–401, 1988.

[23] B. D. Carlson. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Transactions on Aerospace and Electronic systems*, 24(4):397–401, 1988.

[24] V. **v**Cernỳ. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of optimization Theory and Applications*, 45(1):41–51, 1985.

[25] S. Chan and H. Chen. Uniform concentric circular arrays with frequency invariant characteristics theory, design, adaptive beamforming and DOA estimation. *IEEE Transactions on Signal Processing*, 55(1):165–177, 2006.

[26] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee. A survey of sound source localization methods in wireless acoustic sensor networks. *Wireless Communications and Mobile Computing*, 2017, 2017.

[27] A. R. Conn, G. Gould, and P. L. Toint. *LANCELOT: a Fortran Package for Large-Scale Nonlinear Optimization*, volume 17. Springer Science & Business Media, 2013.

[28] H. Cox, R. Zeskind, and M. Owen. Robust adaptive beamforming. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(10):1365–1376, 1987.

[29] M. Crocco, A. Del Bue, M. Bustreo, and V. Murino. A closed form solution to the microphone position self-calibration problem. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2597–2600. IEEE, 2012.

[30] M. Crocco, A. Del Bue, and V. Murino. A bilinear approach to the position self-calibration of multiple sensors. *IEEE Transactions on Signal Processing*, 60(2):660–673, 2011.

[31] M. Dahl and I. Claesson. Acoustic noise and echo cancelling with microphone array. *IEEE transactions on Vehicular Technology*, 48(5):1518–1526, 1999.

[32] S. Darlington. Linear least-squares smoothing and prediction, with applications. *Bell System Technical Journal*, 37(5):1221–1294, 1958.

[33] V. H. Diaz-Ramirez and V. Kober. Robust speech processing using local adaptive non-linear filtering. *IET Signal Processing*, 7(5):345–359, 2013.

[34] N. Dionelis and M. Brookes. Phase-aware single-channel speech enhancement with modulation-domain Kalman filtering. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(5):937–950, 2018.

[35] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters. Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(1):38–51, 2009.

[36] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli. Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6):12–30, 2015.

[37] H. Duan, B. P. Ng, C. M. S. See, and J. Fang. Broadband beamforming using TDL-form IIR filters. *IEEE Transactions on Signal Processing*, 55(3):990–1002, 2007.

[38] H. Dudley. Remaking speech. *The Journal of the Acoustical Society of America*, 11(2):169–177, 1939.

[39] H. Dudley. The carrier nature of speech. *Bell System Technical Journal*, 19(4):495–515, 1940.

[40] M. Er and A. Cantoni. Derivative constraints for broad-band element space antenna array processors. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(6):1378–1393, 1983.

[41] T. H. Falk, S. Stadler, W. B. Kleijn, and W.-Y. Chan. Noise suppression based on extending a speech-dominated modulation band. In *Eighth Annual Conference of the International Speech Communication Association*, 2007.

[42] Z. G. Feng and K. F. C. Yiu. The design of multi-dimensional acoustic beamformers via window functions. *Digital Signal Processing*, 29:107–116, 2014.

[43] Z. G. Feng, K. F. C. Yiu, and S. E. Nordholm. A two-stage method for the design of near-field broadband beamformer. *IEEE Transactions on Signal Processing*, 59(8):3647–3656, 2011.

[44] Z. G. Feng, K. F. C. Yiu, and S. E. Nordholm. Performance limit of broadband beamformer designs in space and frequency. *Journal of Optimization Theory and Applications*, 164(1):316–341, 2015.

[45] O. L. Frost. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935, 1972.

[46] F. Gallun and P. Souza. Exploring the role of the modulation spectrum in phoneme recognition. *Ear and hearing*, 29(5):800, 2008.

[47] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 49(8):1614–1626, 2001.

[48] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov. A consolidated perspective on multimicrophone speech enhancement and source separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(4):692–730, 2017.

[49] M. Gao, K.-F. C. Yiu, S. Nordholm, and Y. Ye. On a new SDP-SOCP method for acoustic source localization problem. *ACM Transactions on Sensor Networks (TOSN)*, 12(4):1–26, 2016.

[50] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens. Auto-localization in ad-hoc microphone arrays. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 106–110. IEEE, 2013.

[51] S. M. Golan, S. Gannot, and I. Cohen. A reduced bandwidth binaural MVDR beamformer.

[52] S. M. Golan, S. Gannot, and I. Cohen. Subspace tracking of multiple sources and its application to speakers extraction. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 201–204. IEEE, 2010.

[53] G. H. Golub and C. F. Van Loan. *Matrix Computations*. JHU press, 2013.

[54] R. Gooch and J. Shynk. Wide-band adaptive array processing using pole-zero digital filters. *IEEE transactions on antennas and propagation*, 34(3):355–367, 1986.

[55] N. Grbić and S. Nordholm. Soft constrained subband beamforming for hands-free speech enhancement. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages I–885. IEEE, 2002.

[56] L. Griffiths and C. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, 30(1):27–34, 1982.

[57] T. Gustafsson, B. D. Rao, and M. Trivedi. Source localization in reverberant environments: Modeling and statistical analysis. *IEEE Transactions on Speech and Audio Processing*, 11(6):791–803, 2003.

[58] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski. New insights into the mvdr beamformer in room acoustics. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(1):158–170, 2009.

[59] S. P. Han. Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Mathematical Programming*, 11(1):263–282, 1976.

[60] D. Harris and R. Mersereau. A comparison of algorithms for minimax design of two-dimensional linear phase FIR digital filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(6):492–500, 1977.

[61] A. Hassani, A. Bertrand, and M. Moonen. Cooperative integrated noise reduction and node-specific direction-of-arrival estimation in a fully connected wireless acoustic sensor network. *Signal Processing*, 107:68–81, 2015.

[62] A. Hassanien, S. A. Vorobyov, and K. M. Wong. Robust adaptive beamforming using sequential quadratic programming: An iterative solution to the mismatch problem. *IEEE Signal Processing Letters*, 15:733–736, 2008.

[63] R. C. Hendriks, R. Heusdens, and J. Jensen. MMSE based noise PSD tracking with low complexity. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4266–4269. IEEE, 2010.

[64] R. C. Hendriks, R. Heusdens, U. Kjems, and J. Jensen. On optimal multi-channel mean-squared error estimators for speech enhancement. *IEEE Signal Processing Letters*, 16(10):885–888, 2009.

[65] M. Hennecke, T. Plotz, G. A. Fink, J. Schmalenstroer, and R. Hab-Umbach. A hierarchical approach to unsupervised shape calibration of microphone array

networks. In *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, pages 257–260, 2009.

[66] H. Hermansky, E. A. Wan, and C. Avendano. Speech enhancement based on temporal processing. In *1995 International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 405–408. IEEE, 1995.

[67] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn. Distributed MVDR beamforming for (wireless) microphone networks using message passing. In *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, pages 1–4. VDE, 2012.

[68] I. Himawan, I. McCowan, and S. Sridharan. Clustered blind beamforming from ad-hoc microphone arrays. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):661–676, 2011.

[69] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Transactions on Signal Processing*, 47(10):2677–2684, 1999.

[70] D. Hu, Z. Chen, and F. Yin. Analytical geometry calibration for acoustic transceiver arrays. *IEEE Signal Processing Letters*, 27:1979–1983, 2020.

[71] J. Hu and L. Rabiner. Design techniques for two-dimensional digital filters. *IEEE Transactions on Audio and Electroacoustics*, 20(4):249–257, 1972.

[72] J. Ianniello. Time delay estimation via cross-correlation in the presence of large estimation errors. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(6):998–1003, 1982.

[73] F. Jacob, J. Schmalenstroeer, and R. Haeb-Umbach. Microphone array position self-calibration from reverberant speech input. In *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, pages 1–4. VDE, 2012.

[74] Y. Kaneda and J. Ohga. Adaptive microphone-array system for noise reduction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(6):1391–1400, 1986.

[75] S. Karimian-Azari and T. H. Falk. Modulation spectrum based beamforming for speech enhancement. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 91–95. IEEE, 2017.

[76] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo. Far-field multicast beamforming for uniform linear antenna arrays. *IEEE Transactions on Signal Processing*, 55(10):4916–4927, 2007.

[77] R. Kennedy, T. Abhayapala, and D. Ward. Broadband nearfield beamforming using a radial beampattern transformation. *IEEE Transactions on Signal Processing*, 46(8):2147–2156, 1998.

[78] R. A. Kennedy, D. B. Ward, and T. D. Abhayapala. Nearfield beamforming using radial reciprocity. *IEEE Transactions on Signal Processing*, 47(1):33–40, 1999.

[79] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[80] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters. Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues. *IEEE Transactions on Signal Processing*, 55(4):1579–1585, 2007.

[81] A. I. Koutrouvelis, T. W. Sherson, R. Heusdens, and R. C. Hendriks. A low-cost robust distributed linearly constrained beamformer for wireless acoustic sensor networks with arbitrary topology. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(8):1434–1448, 2018.

[82] H. Kuttruff. *Room Acoustics*. Crc Press, 2016.

[83] B. K. Lau, Y. H. Leung, K. L. Teo, and V. Steeram. Minimax filters for microphone arrays. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 46(12):1522–1524, 1999.

[84] T.-K. Le and K. C. Ho. Uncovering source ranges from range differences observed by sensors at unknown positions: Fundamental theory. *IEEE Transactions on Signal Processing*, 67(10):2665–2678, 2019.

[85] E. A. Lehmann and A. M. Johansson. Diffuse reverberation model for efficient image-source simulation of room impulse responses. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1429–1439, 2009.

[86] J. Li and P. Stoica. *Robust Adaptive Beamforming*. John Wiley, Inc. Hoboken, New Jersey, 2006.

[87] Z. Li, K. F. C. Yiu, Y.-H. Dai, and S. Nordholm. Distributed LCMV beamformer design by randomly permuted ADMM. *Digital Signal Processing*, 106:102820, 2020.

[88] W. Lobato and M. H. Costa. Worst-case-optimization robust-MVDR beamformer for stereo noise reduction in hearing aids. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2224–2237, 2020.

[89] P. C. Loizou. Speech quality assessment. In *Multimedia Analysis, Processing and Communications*, pages 623–654. Springer, 2011.

[90] C. G. Lopes and A. H. Sayed. Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, 56(7):3122–3136, 2008.

[91] S. Y. Low. Compressive speech enhancement in the modulation domain. *Speech Communication*, 102:87–99, 2018.

[92] J. G. Lyons and K. K. Paliwal. Effect of compressing the dynamic range of the power spectrum in modulation filtering based speech enhancement. In *Ninth Annual Conference of the International Speech Communication Association*. Citeseer, 2008.

[93] S. Markovich, S. Gannot, and I. Cohen. Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6):1071–1086, 2009.

[94] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot. Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks. *Signal Processing*, 107:4–20, 2015.

[95] S. Markovich-Golan, S. Gannot, and I. Cohen. Distributed GSC beamforming using the relative transfer function. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 1274–1278, 2012.

[96] S. Markovich-Golan, S. Gannot, and I. Cohen. Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(2):343–356, 2013.

[97] R. Mars, V. G. Reju, and A. W. H. Khong. A frequency-invariant fixed beamformer for speech enhancement. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pages 1–6, 2014.

[98] J. Mayhan, A. Simmons, and W. Cummings. Wide-band adaptive antenna nulling using tapped delay lines. *IEEE Transactions on Antennas and Propagation*, 29(6):923–936, 1981.

[99] I. McCowan, M. Lincoln, and I. Himawan. Microphone array shape calibration in diffuse noise fields. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3):666–670, 2008.

[100] L. Nahma, H. H. Dam, and S. Nordholm. Robust Beamformer design against mismatch in microphone characteristics and acoustic environment. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pages 76–80, 2018.

[101] S. Nordebo, I. Claesson, and S. Nordholm. Weighted chebyshev approximation for the design of broadband beamformers using quadratic programming. *IEEE Signal Processing Letters*, 1(7):103–105, 1994.

[102] S. Nordholm, V. Rehbock, K. Tee, and S. Nordebo. Chebyshev optimization for the design of broadband beamformers in the near field. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 45(1):141–143, 1998.

[103] S. E. Nordholm, H. H. Dam, C. C. Lai, and E. A. Lehmann. Broadband beamforming and optimization. In *Academic Press Library in Signal Processing*, volume 3, pages 553–598. Elsevier, 2014.

[104] M. O'Connor and W. B. Kleijn. Diffusion-based distributed MVDR beamformer. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 810–814. IEEE, 2014.

[105] M. O'Connor, W. B. Kleijn, and T. Abhayapala. Distributed sparse MVDR beamforming using the bi-alternating direction method of multipliers. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 106–110. IEEE, 2016.

[106] O. Olgun, E. Erdem, and H. Hacıhabibo***u***glu. Rotation calibration of rigid spherical microphone arrays for multi-perspective 6DoF audio recordings. In

2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), pages 1–7. IEEE, 2021.

[107] N. Ono, H. Kohno, N. Ito, and S. Sagayama. Blind alignment of asynchronously recorded signals for distributed microphone array. In *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 161–164. IEEE, 2009.

[108] K. Paliwal, B. Schwerin, and K. Wójcicki. Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. *Speech Communication*, 54(2):282–305, 2012.

[109] K. Paliwal, K. Wójcicki, and B. Schwerin. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. *Speech Communication*, 52(5):450–475, 2010.

[110] C. Pan, J. Chen, and J. Benesty. Performance study of the mvdr beamformer as a function of the source incidence angle. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(1):67–79, 2013.

[111] S. Pasha, J. Lundgren, C. Ritz, and Y. Zou. Distributed microphone arrays, emerging speech and audio signal processing platforms: A review. *Advances in Science, Technology and Engineering Systems*, 5(4):331–343, 2020.

[112] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *IEEE Signal Processing Magazine*, 22(4):54–69, 2005.

[113] J. Perez-Lorenzo, R. Viciana-Abad, P. Reche-Lopez, F. Rivas, and J. Escolano. Evaluation of generalized cross-correlation methods for direction of arrival estimation using two microphones in real environments. *Applied Acoustics*, 73(8):698–712, 2012.

[114] P. Pertilä, M. Mieskolainen, and M. S. Hämäläinen. Passive self-localization of microphones using ambient sounds. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 1314–1318. IEEE, 2012.

[115] A. Plinge and G. A. Fink. Geometry calibration of multiple microphone arrays in highly reverberant environments. In *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pages 243–247. IEEE, 2014.

[116] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink. Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. *IEEE Signal Processing Magazine*, 33(4):14–29, 2016.

[117] M. Pollefeys and D. Nister. Direct computation of sound and microphone locations from time-difference-of-arrival data. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2445–2448. IEEE, 2008.

[118] M. J. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical Analysis*, pages 144–157. Springer, 1978.

[119] W. H. Press, S. A. Teukolsky, B. P. Flannery, and W. T. Vetterling. *The Art of Scientific Computing*. Cambridge university press, 1992.

[120] H. Qi, Z. G. Feng, K. F. C. Yiu, and S. Nordholm. Optimal design of IIR filters via the partial fraction decomposition method. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 66(8):1461–1465, 2018.

[121] F. Qian and B. D. Van Veen. Quadratically constrained adaptive beamforming for coherent signals and interference. *IEEE Transactions on Signal Processing*, 43(8):1890–1900, 1995.

[122] L. R. Rabiner and B. Gold. *Theory and Application of Digital Signal Processing*. Prentice-Hall, 1975.

[123] W. Rodgers and R. Compton. Adaptive array bandwidth with tapped delay-line processing. *IEEE Transactions on Aerospace and Electronic Systems*, (1):21–28, 1979.

[124] J. Ryan and R. Goubran. Array optimization applied in the near field of a microphone array. *IEEE Transactions on Speech and Audio Processing*, 8(2):173–176, 2000.

[125] J. M. Sachar, H. F. Silverman, and W. R. Patterson. Microphone position and gain calibration for a large-aperture microphone array. *IEEE Transactions on Speech and Audio Processing*, 13(1):42–52, 2004.

[126] S. Samui, I. Chakrabarti, and S. K. Ghosh. Improved single channel phase-aware speech enhancement technique for low signal-to-noise ratio signal. *IET Signal Processing*, 10(6):641–650, 2016.

[127] A. H. Sayed. Diffusion adaptation over networks. In *Academic Press Library in Signal Processing*, volume 3, pages 323–453. Elsevier, 2014.

[128] B. Schwerin and K. Paliwal. Using STFT real and imaginary parts of modulation signals for MMSE-based speech enhancement. *Speech Communication*, 58:49–68, 2014.

[129] J. Segura-Garcia, S. Felici-Castell, J. J. Perez-Solano, M. Cobos, and J. M. Navarro. Low-cost alternatives for urban noise nuisance monitoring using wireless sensor networks. *IEEE Sensors Journal*, 15(2):836–844, 2014.

[130] M. L. Seltzer, B. Raj, and R. M. Stern. Likelihood-maximizing beamforming for robust hands-free speech recognition. *IEEE Transactions on Speech and Audio Processing*, 12(5):489–498, 2004.

[131] S. R. Seydnejad and R. Ebrahimi. Broadband beamforming using laguerre filters. *Signal Processing*, 92(4):1093–1100, 2012.

[132] N. Sharaga, J. Tabrikian, and H. Messer. Optimal cognitive beamforming for target tracking in MIMO radar/sonar. *IEEE Journal of Selected Topics in Signal Processing*, 9(8):1440–1450, 2015.

[133] J. Sherman and W. J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[134] T. Sherson, W. B. Kleijn, and R. Heusdens. A distributed algorithm for robust LCMV beamforming. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 101–105, 2016.

[135] A. M.-C. So and Y. Ye. Theory of semidefinite programming for sensor network localization. *Mathematical Programming*, 109(2-3):367–384, 2007.

[136] S. So and K. K. Paliwal. Modulation-domain kalman filtering for single-channel speech enhancement. *Speech Communication*, 53(6):818–829, 2011.

[137] K. Steiglitz, T. W. Parks, and J. F. Kaiser. METEOR: A constraint-based FIR filter design program. *IEEE Transactions on Signal Processing*, 40(8):1901–1909, 1992.

[138] D. Su, H. Kong, S. Sukkarieh, and S. Huang. Necessary and sufficient conditions for observability of SLAM-based TDOA sensor array calibration and source localization. *IEEE Transactions on Robotics*, 37(5):1451–1468, 2021.

[139] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong. *Wireless Sensor Networks: Signal Processing and Communications Perspectives*. John Wiley & Sons, 2007.

[140] J. Szurley, A. Bertrand, P. Ruckebusch, I. Moerman, and M. Moonen. Greedy distributed node selection for node-specific signal estimation in wireless sensor networks. *Signal Processing*, 94:57–73, 2014.

[141] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2125–2136, 2011.

[142] M. J. Taghizadeh, R. Parhizkar, P. N. Garner, H. Bourlard, and A. Asaei. Ad hoc microphone array calibration: Euclidean distance matrix completion algorithm and theoretical guarantees. *Signal Processing*, 107:123–140, 2015.

[143] R. Talmon, I. Cohen, and S. Gannot. Identification of the relative transfer function between sensors in the short-time Fourier transform domain. In *Speech Processing in Modern Communication*, pages 33–47. Springer, 2010.

[144] V. M. Tavakoli, J. R. Jensen, M. G. Christensen, and J. Benesty. A framework for speech enhancement with ad hoc microphone arrays. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(6):1038–1051, 2016.

[145] V. M. Tavakoli, J. R. Jensen, M. G. Christenseny, and J. Benesty. Pseudo-coherence-based MVDR beamformer for speech enhancement with ad hoc microphone arrays. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2659–2663. IEEE, 2015.

[146] V. M. Tavakoli, J. R. Jensen, R. Heusdens, J. Benesty, and M. G. Christensen. Ad hoc microphone array beamforming using the primal-dual method of multipliers. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 1088–1092, 2016.

[147] V. M. Tavakoli, J. R. Jensen, R. Heusdens, J. Benesty, and M. G. Christensen. Distributed max-SINR speech enhancement with ad hoc microphone arrays. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 151–155, 2017.

[148] S. Thrun. Affine structure from sound. *Advances in Neural Information Processing Systems*, 18, 2005.

[149] K. C. Toh, M. J. Todd, and R. H. Tütüncü. On the implementation and usage of SDPT3–a Matlab software package for semidefinite-quadratic-linear programming, version 4.0. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 715–754. Springer, 2012.

[150] P. Tseng. Second-order cone programming relaxation of sensor network localization. *SIAM Journal on Optimization*, 18(1):156–185, 2007.

[151] S. D. Valente, M. Tagliasacchi, F. Antonacci, P. Bestagini, A. Sarti, and S. Tubaro. Geometric calibration of distributed microphone arrays from acoustic source correspondences. In *2010 IEEE International Workshop on Multimedia Signal Processing*, pages 13–18. IEEE, 2010.

[152] J.-M. Valin, F. Michaud, and J. Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55(3):216–228, 2007.

[153] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons, 2004.

[154] B. Van Veen and K. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, 1988.

[155] J. Velasco, M. J. Taghizadeh, A. Asaei, H. Bourlard, C. J. Martín-Arguedas, J. Macias-Guarasa, and D. Pizarro. Novel GCC-PHAT model in diffuse sound field for microphone array pairwise distance based calibration. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2669–2673. IEEE, 2015.

[156] M. S. Vinton and L. E. Atlas. Scalable and progressive audio codec. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, volume 5, pages 3277–3280. IEEE, 2001.

[157] L. Wang, T.-K. Hon, J. D. Reiss, and A. Cavallaro. Self-localization of ad-hoc arrays using time difference of arrivals. *IEEE Transactions on Signal Processing*, 64(4):1018–1033, 2015.

[158] Q. Wang, S. Guo, and K.-F. C. Yiu. Distributed acoustic beamforming with blockchain protection. *IEEE Transactions on Industrial Informatics*, 16(11):7126–7135, 2020.

[159] Y. Wang and M. Brookes. Model-based speech enhancement in the modulation domain. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 26:580–594, 2018.

[160] D. B. Ward, R. A. Kennedy, and R. C. Williamson. Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns. *The Journal of the Acoustical Society of America*, 97(2):1023–1034, 1995.

[161] E. Weinstein and A. Weiss. Fundamental limitations in passive time-delay estimation–Part II: Wide-band systems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(5):1064–1078, 1984.

[162] A. Weiss and E. Weinstein. Fundamental limitations in passive time delay estimation–Part I: Narrow-band systems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(2):472–486, 1983.

[163] K. J. Woods and J. H. McDermott. Attentive tracking of sound sources. *Current Biology*, 25(17):2238–2246, 2015.

[164] J. Xu, G. Liao, and S. Zhu. Robust LCMV beamforming based on phase response constraint. *Electronics Letters*, 48(20):1304–1306, 2012.

[165] W. Xue, A. H. Moore, M. Brookes, and P. A. Naylor. Modulation-domain multichannel kalman filtering for speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(10):1833–1847, 2018.

[166] W. Xue, A. H. Moore, M. Brookes, and P. A. Naylor. Speech enhancement based on modulation-domain parametric multichannel kalman filtering. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:393–405, 2020.

[167] P. S. Yedavalli, T. Riihonen, X. Wang, and J. M. Rabaey. Far-field RF wireless power transfer with blind adaptive beamforming for internet of things devices. *IEEE Access*, 5:1743–1752, 2017.

[168] K. F. C. Yiu. A parallel beamforming system with real-time implementation. *Multimedia Tools and Applications*, 78(16):23581–23595, 2019.

[169] K. F. C. Yiu, Y. Liu, and K. L. Teo. A hybrid descent method for global optimization. *Journal of Global Optimization*, 28(2):229–238, 2004.

[170] K. F. C. Yiu, X. Yang, S. Nordholm, and K. L. Teo. Near-field broadband beamformer design via multidimensional semi-infinite-linear programming techniques. *IEEE Transactions on Speech and Audio processing*, 11(6):725–732, 2003.

[171] D. Yu and L. Deng. *Automatic Speech Recognition.* Springer, 2016.

[172] Z. L. Yu, W. Ser, M. H. Er, Z. Gu, and Y. Li. Robust adaptive Beamformers based on worst-case optimization and constraints on magnitude response. *IEEE Transactions on Signal Processing*, 57(7):2615–2628, 2009.

[173] Y. Zeng and R. C. Hendriks. Distributed delay and sum beamformer for speech enhancement in wireless sensor networks via randomized gossip. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4037–4040, 2012.

[174] Y. Zeng and R. C. Hendriks. Distributed delay and sum beamformer for speech enhancement via randomized gossip. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(1):260–273, 2014.

[175] Y. Zeng and R. C. Hendriks. Distributed estimation of the inverse of the correlation matrix for privacy preserving beamforming. *Signal Processing*, 107:109–122, 2015.

[176] G. Zhang and R. Heusdens. Convergence of generalized linear coordinate-descent message-passing for quadratic optimization. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1997–2001, 2012.

[177] G. Zhang and R. Heusdens. Linear coordinate-descent message passing for quadratic optimization. *Neural Computation*, 24(12):3340–3370, 2012.

[178] G. Zhang and R. Heusdens. Distributed optimization using the primal-dual method of multipliers. *IEEE Transactions on Signal and Information Processing over Networks*, 4(1):173–187, 2017.

[179] Q. Y. Zou, Z. L. Yu, and Z. P. Lin. A robust algorithm for linearly constrained adaptive beamforming. *IEEE Signal Processing Letters*, 11(1):26–29, 2004.