The Hong Kong Polytechnic University

Department of Applied Mathematics

Near-field beamformer design problems

Zhibao Li

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

September 2013

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)

_____LI Zhibao_____(Name of student)

Dedicated to my parents.

Abstract

This thesis is concerned with the near-field beamformer design problems in reverberant environment and the microphone array placement design problems.

Firstly, we study the influence of room acoustics on the design of broadband beamformer. We introduce the image source method to estimate the room impulse responses, and establish several optimization models for the broadband beamformer design. We also study the barrier beamformer design problem, and introduce the space-time conservation element and solution element method to estimate the corresponding barrier impulse responses.

Secondly, we study the time-domain beamformer design problem in the reverberant environment. We formulate the beamformer design problem as a linear system, and convert it into the least squares problem, whereas it has large scale and bad condition in the reverberant environment. Thus we introduce the Tikhonov regularization technique to improve the condition of the original problem. Moreover, we analyze the effectiveness of beamformer design as the filter length increases.

With considering the nondirectional background noise of speech enhancement, the indoor LCMV beamformer problem is studied in the next. We use the estimated room impulse responses to formulate the model of indoor LCMV beamformer design, then construct a relaxed optimization problem to solve the filter coefficients approximately. Moreover, post-filtering technique combining with indoor LCMV beamformer is studied to further improve the speech quality, and it is found that

the MSIG post-filter combining with indoor LCMV beamformer has the best performance.

Then we study the microphone array placement design problem, and formulate a composite optimization model for it. With the help of infinite length technique, we convert the subproblem on the solving of filter coefficients into the performance limit estimation problem. Moreover, we develop a hybrid descent method with genetic algorithm to solve the problem of microphone array placement design, the descent method can find the best solution around the current placement, and the genetic technique can jump out from the local solution.

Finally, we study the microphone array placement design problem in reverberant environment. We introduce the LCMV framework combining with the infinite length technique to evaluate the effectiveness of placement design. And we also introduce the hybrid descent method to find the optimal array placement design eventually.

Acknowledgements

The endeavor of carrying out research is a fascinatingly non-isolated activity. I am grateful to the several individuals who have supported me in various ways during the PhD program and would like to hereby acknowledge their assistance.

First and foremost, I would like to offer my heartfelt thanks to my supervisor, Dr. Yiu Ka-fai, Cedric, for his enthusiasm and for his extensive support throughout the years. I especially wish to thank him for helping during the whole procedure of the realization of this thesis, where he always welcomed my questions with the enlightening guidance, invaluable discussions and insightful ideas. Thank you for your time, and specially for your patience and understanding.

Furthermore, at the forefront of my PhD experience has been the guidance and kindness of Prof. Sven Nordholm who has been a constant source of inspiration and mentorship. What I have benefited most from him is the rigorous and diligent attitude to scientific research.

Moreover, I wish to thank Dr. Leung Chi-Kin, Randolph and Mr Chan Sui-Ho, Horus for their kindly co-operation works on barrier impulse responses estimation in Chapter 4.1, and the corresponding numerical data acquisition.

I wish to thank all friends and fellow students accompanying me during these hard years of study, Dr. Feng Zhiguo, Liu Jingzhen, Gao Mingjie, Deng Shirong, Hu Shenglong, Song Yisheng, Xu Yi, Xie Cong, Ma Cheng, Zhang Xingfa, Bian Chuanxin, Mr. Tian Boshi, Jin Haiyang, Zhou Yang and so on, who during these years have become my step-family, because of the time we have spent together suffering the

hard times in the library in Hong Kong.

To all my friends I have known during these year in Hong Kong. This experience helped me to discover that world has no limit for anything, great people exist everywhere and it is always a pleasure to meet people like I have met.

I gratefully acknowledge The Hong Kong Polytechnic University for the financial support which enabled me to finish my PhD program more easily. I also express my gratitude to the supporting staffs in Department of Applied Mathematics for their kindly help.

Finally, I would like to express my special thanks to my parents and my friends for their love, encouragement and support.

Achievements

In my PhD study period at Department of Applied Mathematics, The Hong Kong Polytechnic University, the following works are achieved:

1. **Zhibao Li**, Ka Fai Cedric Yiu and Zhiguo Feng, A hybrid descent method with genetic algorithm for microphone array placement design, *Applied Soft Computing*, 13(3): 1486-1490, 2013.

2. **Zhibao Li** and Ka Fai Cedric Yiu, A least-squares indoor beamformer design, *Pacific Journal of Optimization*, 9(4): 697-707, 2013.

3. **Zhibao Li**, Ka Fai Cedric Yiu and Sven Nordholm, On the indoor beamformer design with reverberation, submitted.

4. **Zhibao Li**, Ka Fai Cedric Yiu, Yui Ho Horus Chan and Chi Kin Randolph Leung, Optimal beamformer design with barriers, manuscript.

5. **Zhibao Li**, Ka Fai Cedric Yiu and Sven Nordholm, An analysis of beamformer design in reverberant environment, manuscript.

6. **Zhibao Li**, Ka Fai Cedric Yiu and Sven Nordholm, Post-filtering on LCMV beamformer in reverberant environments, manuscript.

7. **Zhibao Li**, Ka Fai Cedric Yiu and Sven Nordholm, Microphone array placement design based on LCMV beamforming technique in reverberant environment, manuscript.

Contents

Certificate of Originality **iii**

Abstract **vii**

Acknowledgements **ix**

Achievements **xi**

Contents **xiii**

List of Figures **xvii**

List of Tables **xxi**

1 Introduction **1**

1.1 Background 1

1.2 Literature Review 6

1.3 Organization of the Thesis 13

2 Common Beamformer Design **15**

2.1 Sound Wave Propagation 16

2.2 Broadband Beamformer Design 17

2.3 Beamforming Performance 21

3 Indoor Beamformer Design **25**

3.1 Room Impulse Response Estimation 27

3.2 Indoor Beamformer Design 31

3.2.1 Implementation Models 32

3.2.2 Multi-criteria Models 33
3.3 Illustration Examples 36
3.3.1 Broadband Beamformer Design in Acoustic Room 37
3.3.2 Testing Signal Suppression in Stopband 41
4 Barrier Beamformer Design **47**
4.1 Barrier Impulse Response Estimation 49
4.2 Barrier Beamformer Design 54
4.3 Illustration Example 55
5 Beamformer Design in Reverberant Environment **61**
5.1 Time Domain Beamformer Design 62
5.1.1 Common Beamformer Design Problem 62
5.1.2 Least Squares Method 64
5.2 Effectiveness of Beamformer with Increasing Filter Length 68
5.3 Illustration Examples 71
5.3.1 Measurement Indicators 71
5.3.2 Acoustic Room Configuration 74
5.3.3 Group Delay Learning 76
5.3.4 Filter Length Learning 76
5.3.5 Overall Performance Illustration 79
6 Indoor LCMV Beamformer Design **85**
6.1 Indoor LCMV Beamformer 86
6.1.1 Problem Formulation 86
6.1.2 Relaxed Beamformer Deisgn 89
6.2 Post-filtering Techniques 91
6.3 Illustration Examples 94

6.3.1 Acoustic Room Configuration 95
6.3.2 Group Delay Learning 97
6.3.3 Filter Length Learning 100
6.3.4 Overall Performance Illustration 100
6.3.5 Post-filtering Processing with LCMV Beamformer 104

7 Common Microphone Array Placement Design 109

7.1 Placement Design Problem Formulation 110
7.2 Hybrid Descent Method 115
7.3 Illustration Examples 118
7.3.1 2-D Microphone Array Placement Problem 118
7.3.2 3-D Microphone Array Placement Problem 122

8 Indoor Microphone Array Placement Design 127

8.1 Problem Formulation 128
8.2 Performance Limit of LCMV Beamformer 130
8.3 Hybrid Descent Method 133
8.4 Illustration Examples 135
8.4.1 Acoustic Room Configuration 135
8.4.2 Optimal Placement Solving 137
8.4.3 Optimal beamforming performance 138

9 Conclusions and Future Works 143

9.1 Conclusions 143
9.2 Future Works 147

Bibliography 149

List of Figures

2.1 The overall beamforming performance at different filter length L. 23

3.1 Poor performances of the beamformer designed by using the direct path transfer function (2.5) applying to an acoustic room with different reverberation time. 26

3.2 Example of the RIR. 31

3.3 Room setup for indoor beamformer design. 38

3.4 Conflicting nature of early reflections suppression and late reverberation suppression (the measurements of them are defined as ERS and LRS in the following) in the acoustic room with reverberation time $T_{60} = 0.2s$. 39

3.5 Performance of the indoor beamformering given by Model III. 42

3.6 Examples of beamforming performance on the noise suppression in the stopband. 43

3.7 Evaluation of the noise suppression in stopband. 45

4.1 Overall performance of the common designed beamformer under different conditions. 48

4.2 CE construction. 51

4.3 Definition of CE and SE. 52

4.4 A simple example of BIRs estimated by using the proposed CE/SE method: $h_{11}(t)$ and $h_{21}(t)$ are the corresponding two BIRs between microphone location M_1 and source locations S_1, S_2, respectively. 53

4.5 Microphone array setup under barrier condition. 56

4.6 Beamforming performance at specific spatial-frequency domain. 58

4.7 Overall performance of barrier beamformer. 59

5.1 The $z = 1m$ plane of the defined acoustic room. 75
5.2 The unfiltered signals at different room acoustics. 75
5.3 Learning for group delay at filter length $L = 50$. 77
5.4 Performance with respect to filter length L. 78
5.5 RIRs for SOI and INT before and after beamformer filtering at $T_{60} = 0.1s$. 81
5.6 Interferences and reverberation before and after filtering at $T_{60} = 0.1s$. . . 82
5.7 RIRs for SOI and INT before and after beamformer filtering at $T_{60} = 0.2s$. 83
5.8 Interferences and reverberation before and after filtering at $T_{60} = 0.2s$. . . 84
5.9 The beamformed signals signals under different acoustic room. 84
6.1 The $z = 1$ plane of the defined acoustic room. 96
6.2 The unfiltered signals under different acoustic room. 97
6.3 Learning for group delay τ_L at $T_{60} = 0.1s$. 98
6.4 Learning for group delay τ_L at $T_{60} = 0.2s$. 99
6.5 Performance with filter length L. 101
6.6 RIRs for SOI and INT before and after beamformer filtering. 103
6.7 Interference, reverberation and noise before and after beamformer filtering. 104
6.8 RIRs for SOI and INT before and after beamformer filtering. 105
6.9 Interference, reverberation and noise before and after beamformer filtering. 106
6.10 Filtered signals at $T_{60} = 0.1s$. 108
6.11 Filtered signals at $T_{60} = 0.2s$. 108
7.1 Configuration of Example 7.3.1. 119
7.2 Optimal placement for Example 7.3.1. 120
7.3 Placement of Example 7.3.1. 121
7.5 Optimal placement for Example 7.3.2. 122
7.4 Performance of finite length filter under optimal placement. 123

7.6 Optimal placement for Example 7.3.2. 125
7.7 Performance of Example 7.3.2. 126
8.1 Setup of the acoustic room system. 136
8.2 Common used placement configurations. 138
8.3 Limit performance of beamformer in optimal placement at $T_{60} = 0.05$s. . . 141
8.4 Limit performance of beamformer in optimal placement at $T_{60} = 0.1$s. . . 141
8.5 Limit performance of beamformer in optimal placement at $T_{60} = 0.2$s. . . 142

List of Tables

2.1 Beamforming performance with respect to filter length L. 22

3.1 Summary of the beamforming performance on desired response formulation, early reflections and late reverberation suppression. 40

3.2 Summary of off-design performance for indoor beamformer with $T_{60} = 0.1s$. 44

4.1 Summary of the performance of barrier beamformers 57

5.1 Condition number of matrices $\boldsymbol{H}^T\boldsymbol{H}$. 66

5.2 The initial values measurement. 75

5.3 Overall performance of designed beamformers at $T_{60} = 0.1s$. 79

5.4 Overall performance of designed beamformers at $T_{60} = 0.2s$. 79

6.1 The initial values measurement. 96

6.2 Overall performance of indoor LCMV beamformer at $T_{60} = 0.1s$. 102

6.3 Overall performance of indoor LCMV beamformer at $T_{60} = 0.2s$. 102

6.4 Overall performance of post-filters at $T_{60} = 0.1s$. 107

6.5 Overall performance of post-filters at $T_{60} = 0.2s$. 107

7.1 Performance of common structures. 122

8.1 Summary of the beamforming performance at different microphone array placement designs. 140

Chapter 1

Introduction

1.1 Background

We are surrounded by lots of different kinds of sounds. Among those, some carry the information that we want or need, are normally labelled as 'desired', others do not contain any useful information but interfere with the desired sound, are usually referred to as 'noise'. The term 'noise' was coined in 1905 by Einstein [20], and its definition evolves due to the penetration of the research and engineering field.

Many researchers and engineers divide noise into several categories based on the mechanism of the generation of it and conquer each category using different approaches. From the view of the physical property [7], it is divided into four basic categories: additive noise, echo, reverberation and interference. Combating these four categories of noise has led to the developments of diverse acoustic signal processing techniques. They include noise reduction (or speech enhancement), echo cancellation and suppression, speech dereverberation, and source separation, each of which is a rich subject of research, a broad coverage of these research areas can be found in [6, 54, 88]. Echo cancellation involves first recognizing the originally transmitted signal that re-appears, with some delay (usually very small), in the transmitted or received signal. Once the echo is recognized, it can be removed by 'subtracting' it from the transmitted or received signal.

If we narrow the definition of noise down to additive noise, there are lots of approaches to deal with it from the different aspect: spectrum subtraction methods [116, 9], digital short-time Fourier analysis [2, 97], statistical spectral estimation [84, 27], harmonic (or sinusoidal) model using comb filtering [74], linear predictive coding (LPC) model and the Kalman filter [96], the hidden Markov models (HMMs) [29, 26, 25], subspace methods [23, 30], and so on.

That above approaches are developed for noise reduction, and designed depending on the environment. Moreover, their effects and performances are very different. As for interference and reverberation, acoustic beamforming is the effective technique to suppress them. Beamformer combines the spatially distributed sensor collected array data linearly with the beamforming weight to achieve spatial filtering. It enhances the signal from the desired spatial direction and reduces the signal(s) from other direction(s) in addition to possible time/frequency filtering. The simplest beamforming technique is the delay-and-sum (D&S) technique developed by Johnson *et al.* in 1993 [57] and Flanagan *et al.* in 1994 [34]. In this technique all channels are equally weighted at the output, the other kinds of more complex beamforming techniques are to update their parameters to better suit the input signal, and adapt to changing noise conditions. These adaptive beamforming techniques [44, 59, 58, 60] present a higher capacity at reducing noise interference, but are much more sensitive to steering errors due to the approximation of the channel delays.

In the application of speech acquisition, such as teleconference and automobile voice pick up, microphone arrays are commonly deployed to reduce the level of localized and ambient noise from a desired location via spatial frequency filtering. Many approaches on the design of near-field beamformer have been developed in the literatures [61, 62, 103, 90, 91, 66, 118]. In [61, 62], the near-field-far-field reciprocity relationship is derived and applied to design near-field beamformers via far-field design techniques. An interesting approach is presented in [103], which makes use of

a signal propagation vector representing an ideal point source of acoustic radiation. When the desired response is known, multidimensional filter design techniques can be applied. In [90], the minimax problem is formulated as a quadratic programming problem and the sequential quadratic programming (SQP) method is applied. A penalty function method is developed in [91] to formulate the problem as an unconstrained nonlinear optimization problem. In [66], this problem is transformed into an equivalent problem by minimizing an auxiliary function, and the solution is given by the first root of this auxiliary function, then they proposed a root-catching method to find this root and solved this problem. In [118], the L_1-norm measure and the real rotation theorem are applied to formulate the problem as a semi-infinite linear programming problem.

In general, the impulse responses for describing sound propagations from the source point to microphone array have important role in the design of beamformer, and they are often described by simple but elegant mathematic formulae in the open air. However, in the reverberant environment, such as the system in a small room or with barriers around, the sound wave propagation is a very complex phenomenon in the enclosure. Where the sound conducting medium is bounded on all sides by walls, ceiling and floor, and the effect of the beamformer will likely fail if the simple transfer function is employed in the design process. Thus, some researchers have worked on speech dereverberation, noise reduction and source localization in reverberant environment [88, 71] recently. However, it is still a challenging problem on the beamformer design in reverberant environment.

Ward and Elko developed a mixed nearfield/farfield technique to solve that beamformer design problem [115], they designed a near-field beamformer to pass desired signal from a chosen near-field source location, and a far-field beamformer to suppress room reverberation. Whereas, their studies were based on the assumptions that a typical office environment has only one desired near-field source and most

interference/reverberation are generated from far-field. Moreover, the room impulse responses (RIRs) calculated in [115] are generated by Legendre function and spherical Hankel function. To improve the array performance in reverberant enclosure, Flanagan *et al.* [35] proposed the multiple beamforming and mathched-filtering techniques, and Jan *et al.* [56] also investigated the performance of the matched-filter array (MFA) processing in real room. In the MFA approach, the output of each microphone element is processed by a time inverse of the impulse response with fixed delaying truncation from the focal point to the microphone, and the array output is the summation of outputs from each matched-filter, where the RIRs are measured by using the maximal length (ML) pseudo-random sequences [82]. These pioneering works provide ideas for dereverberation, but have limitations such as inaccurate RIRs estimation. Another common idea to deal with this problem is to design beamformers based on approximating the convolutive transfer function (CTF) [36, 37], and more recently, the relative transfer function (RTF) estimator applying [109]. In the last three decades, various methods have been developed to identify the RTFs, more details can be found in [55] and the references therein.

For the specific beamforming technique using microphone array investigated in teleconferencing, hands-free communication systems, speech recognition, and even hearing aids [12, 6, 77], the minimum variance distortionless response (MVDR) beamformer is among the most popular. The MVDR beamformer, also known as Capon beamformer [14], minimizes the output power of the beamformer under a single linear constraint on the response of the array towards the desired signal, the idea of combining multiple inputs in a statistically optimum manner under the constraint of a distortionless mainlobe response toward the desired signal can be attributed to Darlington [21]. Unfortunately, the MVDR beamformer may have unacceptably low nulling level, which leads to significant performance degradation in the case of unexpected interfering signals. Specially, the performance of MVDR degrades in

rapidly moving jammer environments, because the jammer motion may bring the jammers out of the sharp notches of the adapted pattern. Several researchers developed beamformers in which additional linear constraints were imposed (e.g., Er and Cantoni [31]), these beamformers are known as linearly constrained minimum variance (LCMV) beamformers, of which the MVDR beamformer is a special case. It is worthwhile noting that the main objectives of most methods are noise reduction and interference suppression, the imposed constraint is usually constructed from the transfer function for describing sound wave propagating in free-field. However, in the reverberant environment, for instance, a simple acoustic room, the speech signals will be not only corrupted by interfering signals and background noise, but also the reverberation.

Apart from the filter coefficient factors on the design of beamformer, the placement of the microphone array also plays an important role in the overall performance. Different microphone array configuration may have a significant difference in performance. Finding good placement for the microphone array configuration has significant influence on the further design of the near-field beamformer. Therefore, the microphone array placement problem has been addressed partially in the array thinning technique [83, 105, 85, 63, 81, 92, 102], which attempts to reduce the number of elements and adjust the positions of the remaining elements to retain performance. Usually, the formulated problems for microphone array placement design are nonlinear, and there are various global optimization methods have been developed to solve them, including evolutionary programming [64], genetic algorithm [50, 19], simulated annealing algorithm [111, 110, 24] and pattern search algorithm [100]. However, the array thinning technique is essentially one-dimensional and is more suited for antenna design. In formulating the multi-dimensional design problem, a nonlinear optimization problem under the L_2-norm was proposed in [33], which allows the microphones to move around in a multi-dimensional solution space in search of better

performances. However, the objective function is highly nonlinear and is nonconvex with respect to the placement variables, it is therefore hard to tackle with traditional gradient-based methods. Moreover, the microphone array placement design problem is complicated further by the influence of different filter lengths, in which optimal designs might vary for different filter lengths. This problem can be avoided by considering the performance limit when the filter length is sufficiently long. By taking the limit of the filter length to infinity, it was shown in [33] that the filter coefficients are defined by a set of reduced one-dimensional convex optimization problems and the coefficients can be sought efficiently for a particular array configuration. By treating this as a subproblem, we can reduce the original mixed optimization problem to the placement problem with microphone placements being the only decision variables.

1.2 Literature Review

1. **Indoor beamformer design problem**

   Indeed, the sound field can be described by wave equations and appropriate boundary conditions for the walls, and it can be simulated by some useful wave-based acoustic simulators. However, in some regular configurations, many simple but efficient geometrical methods are also effect to simulate it. One way to model this process is to make use of some geometrical acoustics modelling techniques, such as ray or beam tracing method (RTM or BTM) and image-source method (ISM) [70, 108, 1, 10]. They have been widely used to model the acoustic field of enclosed spaces, especially the image-source model originally proposed for rectangular enclosures by J. Allen and D. Berkley in 1979 [1]. However, the required simulation times are growing exponentially with the reflection order, and the associated computational costs constitute a well known drawback of the original image-source implementation. E.A. Lehmann

and A.M. Johansson [68] proposed a diffuse reverberation model (fast-ISM) to achieve a reduction of the computational requirements by modelling the diffuse reverberation tail as decaying random noise, and using the ISM simulator for the computation of the early reflections.

We formulate the indoor beamformer design problem with the help of estimated RIRs, and transform it into the minimax optimization problem, where the maximum of the cost is chosen over the constrained class of position and frequency, while the minimum is taken over the set of filter coefficients. The minimax filter design problem is equivalent to a semi-infinite linear programming problem, so that we can convert it into a constrained linear programming problem by using the discretization technique. Since it usually includes the first 50ms after the direct sound as being useful early reflections [78, 11] that they usually have most of the power of the reverberation, the design criteria separating the direct path, early reflections suppression and late reverberation suppression will be meaningful. In view of this, we propose several optimization models based on the L_1-norm to design indoor beamformers by considering different combinations of the desired response and the formation of reverberation.

2. Barrier beamformer design problem

In designing beamformers for acoustic applications, obstacles in the surroundings are often neglected due to a lack of model in handling sound wave propagation around objects. In fact, existing design methods employ simple geometrical models that can not be extended to simulate the acoustics around barriers, due to the assumption of the system inside the simple enclosure. To address this problem, we employ the conservation laws directly and solve them in a domain of interest using the finite volume scheme with the conservation element and solution element (CE/SE) method developed recently for computational

aeroacoustics to simulate such complex acoustic wave field [17, 18, 79, 75, 52]. This method is also termed the space-time conservation element and solution element (CE/SE) method. It is distinguished by the simplicity of its conceptual basis: flux conservation in both space and time, and robust enough to cover the complete spectrum of inviscid flow from linear acoustic waves all of the way to high-speed flows with shocks. The great strength of the CE/SE method is that it is a very robust, intrinsically shock-capturing scheme and has no difficulty in dealing with vortex-shock interactions. Therefore, it can be expected to deal effectively with a variety of important and real-world computational aero-acoustics (CAA) problems [79, 75, 52], and can be used to simulate the sound propagations under barrier conditions.

We first demonstrate the deterioration in performance when an object is placed nearby the microphone array. Then we formulate the barrier beamformer design problem as a minimax optimization problem with the help of estimated barrier impulse responses. We also transform the beamformer design problem into the minimax optimization problem. And we convert it into a constrained linear programming problem by using the discretization technique. Then we propose optimization model based on the L_1-norm, and introduce effective numerical solver to solve the filter coefficients.

3. Time domain beamformer design in reverberant environment

In the time domain beamformer design, it is often processed on receiving a desired signal of interest arriving from a known location, whereas suppressing all the interference signals arriving from other locations. Thus, a linear system or least squares problem for the design of beamformer can be formulated accordingly [113, 118, 76, 121]. In the open air communication, the formulated linear system is with small scale and can be solved by common least squares

techniques. However, in the reverberant environment, the linear system usually has large scale and with bad condition due to the room acoustics, especially in heavier reverberant cases. Then more effective techniques on the beamformer design in the reverberant environment are necessary to be developed.

In numerical computation, many effective techniques have been developed for large scale linear systems, especially for well structured problems. For instance, the projection methods, Krylov subspace methods, preconditioning techniques and domain decomposition methods for the sparse linear systems[104]. Moreover, there are also many advanced iterative methods in optimization softwares, such as the conjugate gradient method and LSQR method [51, 95]. And in practical, the systems of linear equations are usually within highly ill-conditioned constraint matrices, it is necessary to introduce regularization methods to improve the conditions and find approximate stable solutions. It is also noted that the linear system for beamformer design has not very good sparse structure for iterative methods applying, and the idea of LSQR method is essentially a behavior of regularization approach.

The Tikhonov regularization is a powerful tool for the solution of ill-conditioned linear systems and linear least squares problems in the literature [45, 43, 40, 13]. Thus, we introduce this approach to improve the condition of the least squares problem on the design of beamformer in reverberant environment, and analyze the regularizing property on the convergence of this method. Moreover, to study the influence of the filter length on the beamformer design, we also investigate the effectiveness of the proposed beamformer design method as filter length increases.

4. Indoor LCMV beamformer design problem

The LCMV beamforming technique is effect on the speech enhancement for

nondirectional noise reduction, we also study the design of indoor LCMV beamformer problem. Notice that the output of the LCMV beamformer is a single channel signal, and there are a vast amount of short-time spectral domain speech enhancement techniques have been developed. In terms of single channel approach, the well known methods are the spectral subtraction (SS) [9, 8, 47], the minimum mean square error (MMSE) based estimator, such as the Wiener filter (WF), log-spectral amplitude (LSA) [27, 28, 106, 80]. Among these methods, SS employs only the second order statistics without utilizing the statistics and the distributions of the stochastic signal process. It was stated in the literature that the performance of the gain functions is mainly determined by the *a priori* SNR, while the *a posteriori* SNR acts only as a correction parameter for low *a priori* SNR [15], then its performance is limited, which results in audible sound artifacts known as the musical noise. However, in the MMSE based methods, *a priori* SNR is required, and the most widely used approach for estimating the *a priori* SNR is the decision-directed (DD) method [27], which preforms a linear combination of an estimate of previous *a priori* SNR and another being the maximum-likelihood (ML) SNR estimate. And the MMSE LSA estimator with the DD *a priori* SNR estimate is preferred for less musical noise and speech distortion.

Most recently, Yong *et al.* developed a low complexity gain function by using a sigmoid (SIG) function, which is a logistic function and can be viewed as a general cumulative distribution function (CDF) function [120], it provides several parameters that can be adjusted to flexibly model exponential distributions. To provide a vehicle to enhance the speech quality in high noise conditions, Yong *et al.* proposed a modified sigmoid (MSIG) gain function in [119], it combines a logistic function with a hyperbolic tangent function, which provid-

ing a more flexible gain function with three parameters, that can be optimized to match various criteria to achieve a compromised trade-off between speech distortion, noise reduction and musical noise. Moreover, a modified decision-directed (MDD) approach for estimating the *a priori* SNR has been developed in [119], which reduces the one-frame delay problem and results in an improvement of speech quality. Thus, at the post-filtering step, we can introduce these useful single channel speech enhancement techniques by using the modified decision-directed (MDD) approach to estimate the *a priori* SNR.

5. Common microphone array placement problem

The microphone array placement design problem can be formulated as a composite optimization function. Due to the objective function of the problem is highly nonlinear and nonconvex with respect to the placement variables, the local search techniques might not yield the best result. The genetic algorithm is a stochastic global search method that mimics the metaphor of natural biological evolution firstly laid down rigourously by Holland [53] in 1975, and are well described in many literatures [22, 86]. Genetic algorithms operate on a population of potential solutions applying the principle of survival of the fittest to produce (hopefully) better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals that they were created from, just as in natural adaptation.

We propose a hybrid descent method consists of a genetic algorithm combining with a gradient-based method to solve the microphone array placement

design problem. The literature in genetic algorithm is vast and it has been employed for beamformer design recently [16] to enhance speech recognition accuracy. Since the formulated nonlinear optimization problem will have many local minima, we employ the genetic algorithm here to generate the new feasible solutions in order to jump out from the local minima while retaining previous useful information. Once a descent point is identified, the gradient-based method is executed to descend to the nearest local minimum rapidly. This method has a nice descent property and the objective function is improved monotonically. Numerical experiments will be given to demonstrate that much better placements can be found.

6. Microphone array placement design in reverberant environment

In the time domain, the LCMV beamformer can be used to process some specific targets for signal of interest capturing and interference suppressing, and it can achieve good performance on noise reduction. Moreover, in reverberant environment, the cost for microphone array placement design in continuous region is high because of room acoustics.

We formulate the microphone array placement design problem based on the LCMV beamforming framework. And introduce the infinite length technique to convert the beamformer coefficients solving subproblem into performance limit estimation. Then, we establish the microphone array placement design problem based on the error between beamformer output and the given desired signal. Moreover, the composite optimization problem with respect to placement variables can be solved by the hybrid descent method with genetic algorithm effectively.

1.3 Organization of the Thesis

The thesis is structured as follows:

- Chapter 2 reviews the common broadband beamformer design problem in open air. With the help of the transfer function for describing the sound propagation, we can formulate the beamformer design problem as a minimax optimization problem. Then the discretization method can be used to approximate the corresponding semi-infinite problem.

- Chapter 3 focuses on the indoor broadband beamformer design problem, we introduce the image source method to simulate the room acoustics for the RIRs estimation. And we develop several optimization models to solve the formulated indoor beamformer design problem effectively.

- Chapter 4 is to study the beamformer design problem under barrier condition. We introduce the space-time conservation element and solution element (CE/SE) method to estimate the barrier impulse responses. We also develop effective optimization model for barrier beamformer coefficients solving.

- Chapter 5 makes a study on the time domain beamformer design problem in reverberant environment. The problem can be formulated as a linear system or least squares problem, whereas it usually has large scale and within bad condition. Thus we introduce the regularization technique to improve the condition of it. And the effectiveness of the beamformer design with respect to filter length is also studied.

- Chapter 6 develops the indoor LCMV beamformer design problem. We introduce the LCMV-like optimization model for the design of beamformer in acoustic room, and formulate the relaxed model to solve the filter coefficients

approximately. Moreover, we do numerical experiments to study the relationship between beamformer design and filter length. We also introduce kinds of effective post-filtering approaches to further reduce the remaining noise in the output of beamformer.

- Chapter 7 studies the microphone array placement design problem in space-frequency domain. We develop an optimization problem for it, and introduce infinite length technique to solve the corresponding subproblem. In consideration of nonlinear and nonconvex properties of the composite optimization problem respect to the placement variables. We develop a hybrid descent method employing a gradient-based technique for local neighbourhood improvement and the genetic algorithm technique to jump out form local minima.

- Chapter 8 deals with the design of microphone array placement problem in reverberant environment. We develop the subproblem for evaluate the placement variables under LCMV framework. We also introduce the infinite length technique to convert the subproblem into performance limit estimation. And we use the hybrid descent method to find the optimal placement.

- Chapter 9 concludes the whole thesis and plans for the future work.

Chapter 2

Common Beamformer Design

Beamforming is such a signal processing technique to provide a versatile form of spatial filtering by a processor in conjunction with an array of sensors, and in many applications, such as teleconferencing, hands-free communication systems, speech recognition, and even hearing aids, the speaker is close to the array. That is corresponding to the near field beamformer design problems.

In general, the beamforming techniques were developed for a spectrum of applications including wireless communications [99, 73, 46]. However, in audio speech signal processing, the process of sound wave propagation is very important to develop the methods for beamformer design. In this chapter, we firstly introduce a brief review on the design of near field broadband beamformer.

2.1 Sound Wave Propagation

In physics, the wave equation governs the propagation of waves through fluids (gas or liquid). The form of the equation is a second order partial differential equation. The equation describes the evolution of velocity potential or sound pressure $p(\boldsymbol{r}, t)$ as a function of position $\boldsymbol{r} = [x, y, z]$ and time t.

For a homogeneous medium undergoing inviscid fluid flow, one can linearize the equations governing the dynamic behaviour of the fluid, namely the Euler's equation (i.e., Newton's 2^{nd} law applied to fluids), the continuity equation, and the state equation, to obtain the wave equation

$$\triangle p(\boldsymbol{r}, t) - \frac{1}{c^2}\frac{\partial^2 p(\boldsymbol{r}, t)}{\partial t^2} = -s(\boldsymbol{r}, t), \tag{2.1}$$

where $\triangle$ is the *Laplacian* operator expressed in the Cartesian coordinates (x, y, z), c is the speed of sound, and $s(\boldsymbol{r}, t)$ denote the source function. By defining the Fourier transform as

$$P(\boldsymbol{r}, f) \triangleq \mathcal{F}\{p(\boldsymbol{r}, t)\}(f) = \int_{-\infty}^{\infty} p(\boldsymbol{r}, t) e^{j2\pi f t} dt, \tag{2.2}$$

applying the Fourier transform to (2.1), the time-independent Helmholtz equation can be obtained

$$\triangle P(\boldsymbol{r}, f) + \kappa^2 P(\boldsymbol{r}, f) = -S(\boldsymbol{r}, f), \tag{2.3}$$

where κ denotes the wave number that is related to the angular frequency $2\pi f$ and the wave length λ through

$$\kappa = \frac{2\pi f}{c} = \frac{2\pi}{\lambda}.$$

For a unit-amplitude harmonic point source at position $\boldsymbol{r}_s = [x_s, y_s, z_s]$, we have $S(\boldsymbol{r}, f) = \delta(\boldsymbol{r} - \boldsymbol{r}_s) = \delta(x - x_s)\delta(y - y_s)\delta(z - z_s)$, where $\delta(\cdot)$ denotes the Kronecker

delta function. The partial differential equation in (2.3) can be solved by solving the following inhomogeneous equation:

$$\triangle H(\boldsymbol{r}, \boldsymbol{r}_s, f) + \kappa^2 H(\boldsymbol{r}, \boldsymbol{r}_s, f) = -\delta(\boldsymbol{r} - \boldsymbol{r}_s), \tag{2.4}$$

where $H(\boldsymbol{r}, \boldsymbol{r}_s, f)$ is the Green's function, or called Transfer function. And for an omnidirectional point source in an unbounded space, i.e., free space, the Green's function is

$$H(\boldsymbol{r}, \boldsymbol{r}_s, f) = \frac{1}{||\boldsymbol{r} - \boldsymbol{r}_s||} e^{\frac{-j2\pi f||\boldsymbol{r}-\boldsymbol{r}_s||}{c}}. \tag{2.5}$$

2.2 Broadband Beamformer Design

Assume there is an M-elements microphone array, where behind each elements is an $L-$tap FIR filter, and the microphone elements have been set at the fixed points $\boldsymbol{r}_i,\ i = 1, 2, .., M$. If the signals received by this microphone array are sampled synchronously at the rate of f_s per second, the frequency responses of these FIR filters are defined as

$$W_i(\boldsymbol{w}, f, L) = \boldsymbol{w}_i^T \boldsymbol{d}_0(f, L), \quad i = 1, 2, ..., M, \tag{2.6}$$

where $\boldsymbol{w}_i$ is the coefficients of the i-th FIR filter

$$\boldsymbol{w}_i = [w_i(0), w_i(1), ..., w_i(L-1)]^T,$$

and $\boldsymbol{d}_0(f, L)$ is the vector defined as

$$\boldsymbol{d}_0(f, L) = \left[1, e^{\frac{-j2\pi f}{f_s}}, ..., e^{\frac{-j2\pi f}{f_s}(L-1)}\right]^T.$$

The beamformer output can be obtained by using the impulse response (IR) and FIR-filter responses $W_i(\boldsymbol{w}, f, L)$. Denote the frequency IR of the i-th microphone given by transfer function (2.5) as $H_i(\boldsymbol{r}, f)$ (here we omit the i-th microphone position),

where $\boldsymbol{r}$ is a source point, suppose the desired response is $G_d(\boldsymbol{r}, f, L)$, then our target is to find a group of coefficients $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_M]^T$ for the FIR filters, such that the beamformer output

$$G(\boldsymbol{r}, f, L) = \sum_{i=1}^{M} W_i(\boldsymbol{w}, f, L) H_i(\boldsymbol{r}, f) = \boldsymbol{H}^H(\boldsymbol{r}, f) \boldsymbol{W}(\boldsymbol{w}, f, L), \tag{2.7}$$

is close to the desired response $G_d(\boldsymbol{r}, f, L)$, where $\boldsymbol{H}(\boldsymbol{r}, f) = [H_1(\boldsymbol{r}, f), ..., H_M(\boldsymbol{r}, f)]^T$ is the frequency domain RIR vector, and $\boldsymbol{W}(\boldsymbol{w}, f, L) = [W_1(\boldsymbol{w}, f, L), ..., W_M(\boldsymbol{w}, f, L)]^T$ is the frequency filter response vector defined in (2.6), $(\cdot)^H$ denotes the Hermitian transpose.

Good criteria to measure the effectiveness of the designed beamformer is important. A simple model for the broadband beamformer design problem can be described as the following optimization problem

$$\min_{\boldsymbol{w} \in \mathbb{R}^{M \times L}} \max_{(\boldsymbol{r}, f) \in \Omega} \left\| \boldsymbol{H}^H(\boldsymbol{r}, f) \boldsymbol{W}(\boldsymbol{w}, f, L) - G_d(\boldsymbol{r}, f, L) \right\|_p^p, \tag{2.8}$$

where $p > 0$, Ω is the spatial-frequency domain for definition of $G_d(\boldsymbol{r}, f, L)$. The problem formulation (2.8) is a general form for describing the beamformer design problem, when the desired response $G_d(\boldsymbol{r}, f, L)$ is given.

It is noted that the minimax problem (2.8) is formulated in the continuous space-frequency region Ω, thus it is a semi-infinite problem. And the general numerical schemes in dealing with the semi-infinite problem are the discretization methods and the reduction based methods. For the discretization methods, in order to determine suitable multi-dimensional grids to solve the semi-infinite problems, sequences of adaptive meshes can be applied so that the meshes are refined gradually. Consider problem (2.8) with a multi-dimensional grid region Ω_K for approximating Ω using uniform grid containing K mesh points in each dimension of the spatial-frequency domain. By combining the impulse response vector $\boldsymbol{H}(\boldsymbol{r}, f)$ with the steering vector

$\boldsymbol{d_0}(f, L)$ together, we can extract the filter coefficients and rearrange the beamformer output as

$$G(\boldsymbol{r}, f, L) = \boldsymbol{H}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L) = \boldsymbol{w}^T\boldsymbol{d}(\boldsymbol{r}, f, L), \tag{2.9}$$

where $\boldsymbol{d}(\boldsymbol{r}, f, L) = \boldsymbol{H}(\boldsymbol{r}, f) \otimes \boldsymbol{d_0}(f, L)$, and $\otimes$ denotes the Kronecker product.

Expanding the complex functions as

$$\boldsymbol{d}(\boldsymbol{r}, f, L) = \boldsymbol{d_1}(\boldsymbol{r}, f, L) + j\boldsymbol{d_2}(\boldsymbol{r}, f, L), \quad G_d(\boldsymbol{r}, f, L) = G_{d_1}(\boldsymbol{r}, f, L) + jG_{d_2}(\boldsymbol{r}, f, L).$$

and denoting

$$\boldsymbol{u}(\boldsymbol{r}, f) = (\boldsymbol{w}^T\boldsymbol{d_1}(\boldsymbol{r}, f, L) - G_{d_1}(\boldsymbol{r}, f, L)), \quad \boldsymbol{v}(\boldsymbol{r}, f) = (\boldsymbol{w}^T\boldsymbol{d_2}(\boldsymbol{r}, f, L) - G_{d_2}(\boldsymbol{r}, f, L)).$$

then the minimax filter design problem (2.8) can be rewritten as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{M\times L}} \max_{(\boldsymbol{r},f)\in\Omega_K} \left|\boldsymbol{u}(\boldsymbol{r}, f, L) + j\boldsymbol{v}(\boldsymbol{r}, f, L)\right|.$$

We noticed that if the standard modulus of the complex function is used to measure the complex number $\boldsymbol{u}(\boldsymbol{r}, f) + j\boldsymbol{v}(\boldsymbol{r}, f)$, i.e.,

$$\min_{\boldsymbol{h}\in\mathbb{R}^{N\times L}} \max_{(\boldsymbol{r},f)\in\Omega_M} \boldsymbol{u}^2(\boldsymbol{r}, f) + \boldsymbol{v}^2(\boldsymbol{r}, f), \tag{2.10}$$

a quadratic term will arise which will increase the nonlinearity of the problem, this corresponds to the use of a L_2-norm to measure the vector $[\boldsymbol{u}(\boldsymbol{r}, f)\ \boldsymbol{v}(\boldsymbol{r}, f)]^T$. On the other hand, the L_1-norm has a lot of favorable properties, it provides the upper and lower bounds for the L_2-norm,

$$\frac{\sqrt{2}}{2}||\boldsymbol{x}||_1 \leqslant ||\boldsymbol{x}||_2 \leqslant ||\boldsymbol{x}||_1, \quad \forall \boldsymbol{x} \in \mathbb{R}^2.$$

Therefore, we can use the L_1-norm to measure the vector $[\boldsymbol{u}(\boldsymbol{r}, f)\ \boldsymbol{v}(\boldsymbol{r}, f)]^T$ in problem (2.10), which makes the numerical implementation easier. By using the L_1-norm, the beamformer design problem can be expressed as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{M\times L}} \max_{(\boldsymbol{r},f)\in\Omega_K} \left|\boldsymbol{u}(\boldsymbol{r}, f, L)\right| + \left|\boldsymbol{v}(\boldsymbol{r}, f, L)\right|. \tag{2.11}$$

It is noted that the above minimax problem can be converted it into a linear programming, one approach is to introduce two new variables for controlling the real part and imaginary part separately as

$$z_1 = \max_{(\boldsymbol{r},f)\in\Omega_K} |\boldsymbol{u}(\boldsymbol{r},f,L)|\,,\ z_2 = \max_{(\boldsymbol{r},f)\in\Omega_K} |\boldsymbol{v}(\boldsymbol{r},f,L)|\,,$$

then an implementation model for problem (2.8) can be established as

$$\begin{array}{ll} \min\limits_{\substack{\boldsymbol{w}\in\mathbb{R}^{M\times L}\\ \boldsymbol{z}\in\mathbb{R}^2}} & z_1+z_2 \\ s.t. & |\boldsymbol{u}(\boldsymbol{r},f,L)| \leqslant z_1, \quad \forall(\boldsymbol{r},f)\in\Omega_K, \\ & |\boldsymbol{v}(\boldsymbol{r},f,L)| \leqslant z_2, \quad \forall(\boldsymbol{r},f)\in\Omega_K. \end{array} \tag{2.12}$$

The developed optimization model (2.12) is a linear programming problem, it can be formulated as a standard linear programming model (2.13) in matrix notation.

$$\begin{array}{ll} \min\limits_{\mathcal{Z}\in R^{M\times L+2}} & c^T\mathcal{Z} \\ s.t. & \mathcal{A}(\boldsymbol{r},f,L)\mathcal{Z}-b\leqslant 0,\ \forall(\boldsymbol{r},f)\in\Omega_K, \end{array} \tag{2.13}$$

where

$$\mathcal{A}(\boldsymbol{r},f,L)=\begin{pmatrix} \boldsymbol{d_1}(\boldsymbol{r},f,L) & -1 & 0 \\ -\boldsymbol{d_1}(\boldsymbol{r},f,L) & -1 & 0 \\ \boldsymbol{d_2}(\boldsymbol{r},f,L) & 0 & -1 \\ -\boldsymbol{d_2}(\boldsymbol{r},f,L) & 0 & -1 \end{pmatrix},$$

$$\mathcal{Z}=\begin{pmatrix} \boldsymbol{w} \\ z_1 \\ z_2 \end{pmatrix}, c=\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, b=\begin{pmatrix} G_{d_1}(\boldsymbol{r},f,L) \\ -G_{d_1}(\boldsymbol{r},f,L) \\ G_{d_2}(\boldsymbol{r},f,L) \\ -G_{d_2}(\boldsymbol{r},f,L) \end{pmatrix}.$$

The above standard linear programming problem (2.13) can be solved by some optimization toolbox efficiently, such as `linprog` function provided by the optimization toolbox from MATLAB.

2.3 Beamforming Performance

In this section, we take a simple example to evaluate the proposed broadband beamformer design method. To define the microphone array, an equispaced linear array with 5-microphone elements are setup at

$$\mathsf{Mic} = \{\text{(2,3.9,1.5), (2,3.95,1.5), (2,4,1.5), (2,4.05,1.5), (2,4.1,1.5)}\}$$

in meter, respectively; the array spacing is $5cm$ to avoid spatial aliasing for the frequency of interest, and an L-tap FIR filter is fixed behind each microphone element.

To define the specific objective region for pick up, the passband is defined as

$$\Omega_p = \{(\boldsymbol{r}, f) \mid x = 1m,\ |y-4| \leqslant 0.4m,\ z = 1.5m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\},$$

and the stopband is defined as

$$\begin{aligned}\Omega_s = &\{(\boldsymbol{r}, f) \mid x = 1m,\ |y-4| \leqslant 0.4m,\ z = 1.5m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\} \ldots \\ &\cup\{(\boldsymbol{r}, f) \mid x = 1m,\ 1.5m \leqslant |y-4| \leqslant 3.0m,\ z = 1.5m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\} \ldots \\ &\cup\{(\boldsymbol{r}, f) \mid x = 1m,\ 1.5m \leqslant |y-4| \leqslant 3.0m,\ z = 1.5m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\}.\end{aligned}$$

The desired response function in the passband is given by

$$G_d(\boldsymbol{r}, f, L) = e^{-j2\pi f \frac{||\boldsymbol{r}-\boldsymbol{r}_c||}{c} + \frac{L-1}{2}}, \tag{2.14}$$

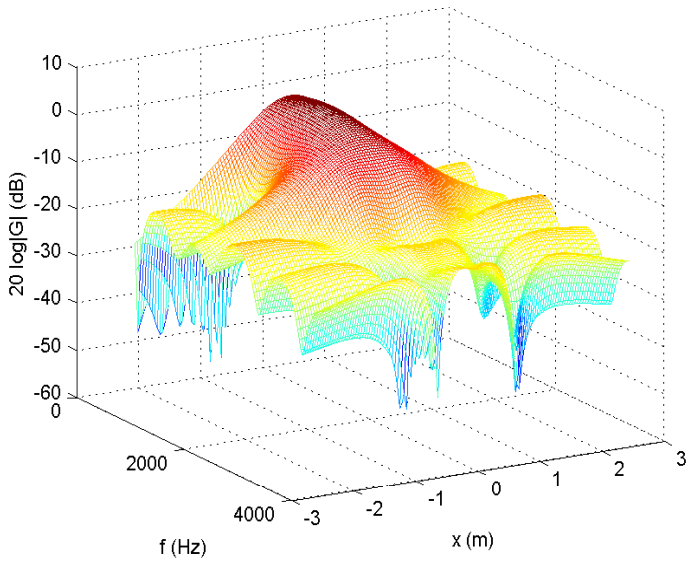
and in the stopband, we simply set $G_d(\boldsymbol{r}, f, L) = 0$ in order to filter out the interference and noise. For numerical calculation, we discrete the spatial-frequency domain of the passband and stopband into a grid of 30×30 frames. And use a more density grid scheme 120×120 to evaluate the beamforming performances.

By using the proposed beamformer design method, we can get an effective beamformer for the desired objective in passband and stopband at some feasible filter length selections. The evaluation performances on passband gain, passband ripple and stopband ripple with respect to filter length L are given in following Table 2.1.

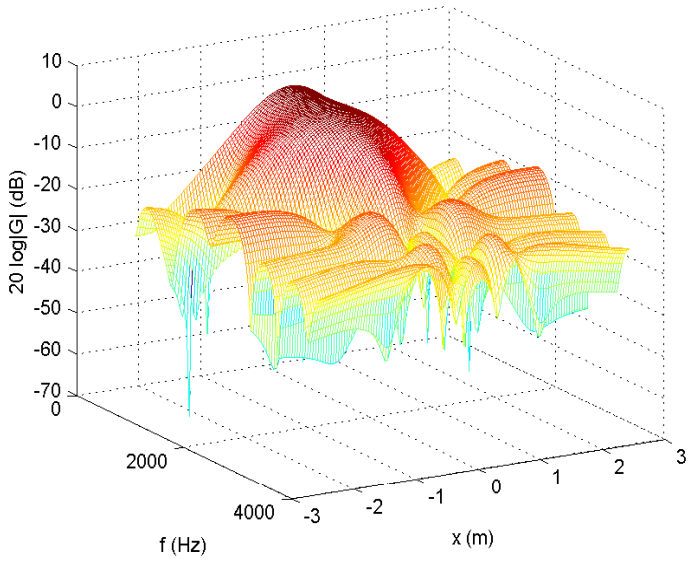
Table 2.1: Beamforming performance with respect to filter length L.

L	Passband gain	Passband ripple	Stopband ripple (dB)
5	0.8933	0.2956	-7.8355
6	0.9506	0.2424	-12.1399
7	1.0157	0.1905	-14.9641
8	0.9948	0.1893	-16.6218
9	1.0031	0.1625	-15.9187
10	0.9942	0.1468	-15.6222
11	0.9893	0.1822	-15.4857

From the above table, we can see that there are some proper filter length for our specific beamformer design, the broadband beamformer should be designed at filter length $L \geqslant 7$ to achieve some better performance. To give an illustration, two overall performances are plotted in the following Figure 2.1.



(a) $L = 7$



(b) $L = 11$

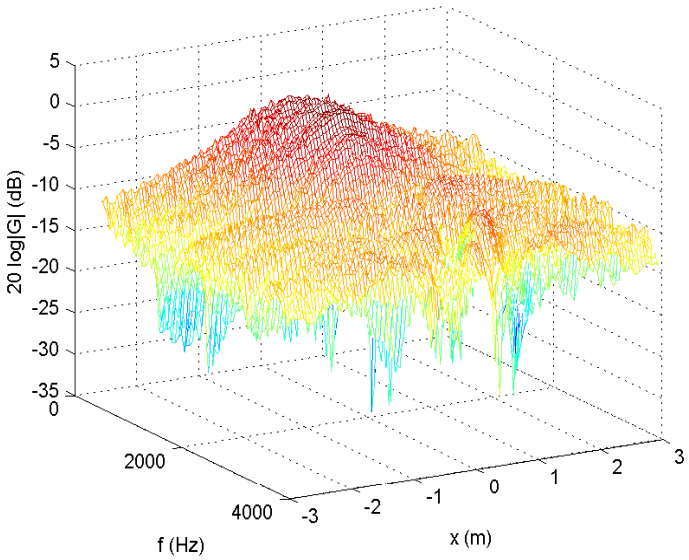
Figure 2.1: The overall beamforming performance at different filter length L.

Chapter 3

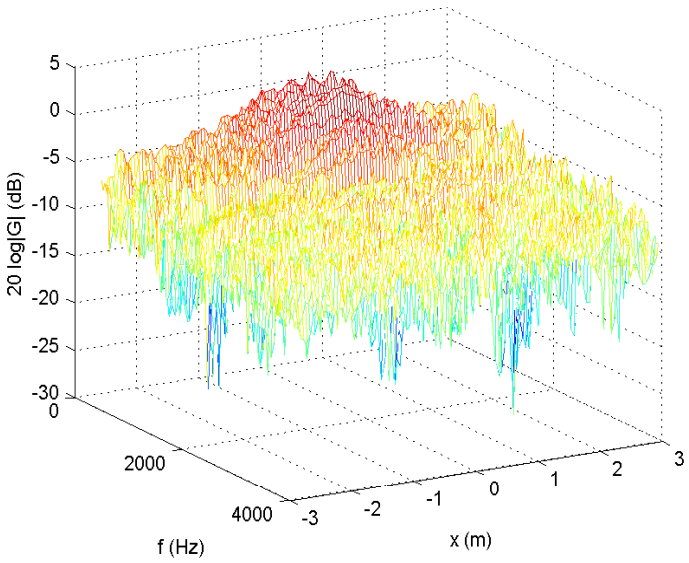
Indoor Beamformer Design

In the application of indoor beamformer design, there is no general formula of transfer function like (2.5) that can be used due to the reverberation caused by room acoustics. The common approaches used for beamformer design suffered in accuracy as a result, a simple example for illustrating the beamformer designed by the method developed from (2.13) for indoor acoustics is given in the following Figure 3.1 (the more detail on room configuration is defined in the later Section 3.3).

The cause of losing effectiveness on the design of beamformer (2.13) in acoustic room is that the transfer function (2.5) has been used directly. Therefore, more effective method is necessary to be developed on the design of indoor beamformer. In this chapter, we propose a method employing the room impulse responses (RIRs) to deal with such a problem, where the RIRs are estimated by the effective room simulator based on image-source method (ISM).



(a) $T_{60} = 0.1s$



(b) $T_{60} = 0.2s$

Figure 3.1: Poor performances of the beamformer designed by using the direct path transfer function (2.5) applying to an acoustic room with different reverberation time.

3.1 Room Impulse Response Estimation

For a model of shoebox enclosure with dimensions $\boldsymbol{L} = [L_x, L_y, L_z]^T$ containing a sound source and a microphone receiver, located at $\boldsymbol{r}_s = (x_s, y_s, z_s)^T$ and $\boldsymbol{r}_m = (x_m, y_m, z_m)^T$, respectively. The ISM technique for measuring the RIR from $\boldsymbol{r}_s$ to $\boldsymbol{r}_m$ is use the image sources on an infinite grid of mirror rooms expanding in all three dimensions. And the contribution of each image source to the receiver signal is a replica of the source signal delayed by a lag τ and attenuated by an amplitude factor A, therefore, the RIR from $\boldsymbol{r}_s$ to $\boldsymbol{r}_m$ can be denoted as

$$h(t) = \sum_{\boldsymbol{\mu}=0}^{1} \sum_{\boldsymbol{\nu}=-\infty}^{+\infty} A(\boldsymbol{\mu}, \boldsymbol{\nu}) \cdot \delta(t - \tau(\boldsymbol{\mu}, \boldsymbol{\nu})), \tag{3.1}$$

where $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)^T$ and $\boldsymbol{\nu} = (\nu_x, \nu_y, \nu_z)^T$ are triplet parameters controlling the indexing of the image sources in all dimensions, $\boldsymbol{\beta} = \{\beta_{x,i}, \beta_{y,i}, \beta_{z,i}, i = 1, 2\}$ are the reflection coefficients for each of the six enclosure surfaces, and $A(\cdot)$ is the amplitude factor is defined as

$$A(\boldsymbol{\mu}, \boldsymbol{\nu}) = \frac{\beta_{x,1}^{|x_m - x_s|} \beta_{x,2}^{|x_m|} \beta_{y,1}^{|y_m - y_s|} \beta_{y,2}^{|y_m|} \beta_{z,1}^{|z_m - z_s|} \beta_{z,2}^{|z_m|}}{4\pi \cdot d(\boldsymbol{\mu}, \boldsymbol{\nu})}, \tag{3.2}$$

$\delta(\cdot)$ denotes the Dirac impulse function and $\tau(\boldsymbol{\mu}, \boldsymbol{\nu}) = d(\boldsymbol{\mu}, \boldsymbol{\nu})/c$ is the time delay of the considered image source, c is the sound propagation velocity and $d(\cdot)$ represents the distance as

$$d(\boldsymbol{\mu}, \boldsymbol{\nu}) = \| \ \mathrm{diag}\,(2\mu_x - 1, 2\mu_y - 1, 2\mu_z - 1) \cdot \boldsymbol{r}_s + \boldsymbol{r}_m - \ \mathrm{diag}\,(2\nu_x, 2\nu_y, 2\nu_z) \cdot \boldsymbol{L} \ \| \ .$$

In (3.1), the sum over $\boldsymbol{\mu}$ (respectively, $\boldsymbol{\nu}$) is used to represent a triple sum over each of the triplet's internal indices, see [1, 68] for more details.

Notice that the above ISM approach is dealt with by using nearest-integer rounding of each image source's propagation time, resulting in a shift of the corresponding

impulse in the RIR. It thus leads to a coarse histogram-like representation of the desired RIR, which subsequently requires high-pass filtering in order to remove the nonphysical defect of it resulting at zero frequency. A more accurate solution to this problem is to carry out the ISM computations in frequency domain, which allows the representation of delays that are not necessarily integer multiples of the sampling period. In the frequency domain, a time shift τ is represented as $\exp(-j\omega\tau)$, with $j = \sqrt{-1}$ and ω denoting the frequency variable. The frequency domain RIR $H(\cdot)$ hence results from (3.1) as

$$H(\omega) = \sum_{\boldsymbol{\mu}=0}^{1} \sum_{\boldsymbol{\nu}=-\infty}^{+\infty} A(\boldsymbol{\mu}, \boldsymbol{\nu}) \cdot e^{-j\omega\tau(\boldsymbol{\mu},\boldsymbol{\nu})},$$

and the time domain RIR follows as the inverse Fourier transform of $H(\omega)$, i.e.

$$h(t) = \mathcal{F}^{-1}\left\{H(\omega)\right\}.$$

Moreover, the acoustical property of each surface in the enclosure is characterized by means of a sound absorption coefficient $\boldsymbol{\alpha}$, related to the reflection coefficient $\boldsymbol{\beta}$ according to

$$\boldsymbol{\alpha} = 1 - \boldsymbol{\beta}^2.$$

That is

$$\boldsymbol{\beta} = \pm\sqrt{1 - \boldsymbol{\alpha}}, \tag{3.3}$$

and the positive definition of the $\boldsymbol{\beta}$ parameter was adopted by the original ISM implementation. However, when used in conjunction with a frequency-domain implementation, this approach generates anomalous RIR showing a distinctively nonphysical tail decay.

An alternative approach is to use the negative definition of the parameter $\boldsymbol{\beta}$ previously studied and used by J. António *et al.* [3], this can be explained by considering the angle-dependent formula for the reflection coefficient of a boundary with impedance ζ,

$$\boldsymbol{\beta} = \frac{\zeta \cos(\theta) - 1}{\zeta \cos(\theta) + 1},$$

which can become negative for certain range of incidence angle θ. This alternative approach results in RIR represent a much better approximation of typical real-room RIR.

Since the amplitude factor (3.2) decreases as reflection order increased, the lasting RIR (3.1) should be cutoff when appropriate. The reverberation time (RT) is an important acoustical property in room acoustic simulation, and it is primarily determined by the volume of the room and how much sound absorbing material is present.

Previous literature works have made extensive use of well-known RT formulae to solve the problem under consideration, many RT expressions can be found in the acoustics literature [89], and the present work investigates some of the most commonly used definitions, namely Sabine, Eyring, Millington and Fitzroy's formulae. However, these classical RT prediction methods have significant risk that the performance results are ultimately presented for a reverberation level that does not correspond to the actual testing conditions. And some alternative approaches still do not provide much insight into the practical reverberation characteristics of the considered environment.

Use the energy decay envelope of RIR is a straightforward manner for the purpose of RT prediction, the energy decay envelope $E(t)$, known as energy decay curve (EDC), can be computed using the normalized Schroeder integration method [72,

107]:

$$E(t) = 10 \cdot \log_{10} \left(\frac{\int_t^{+\infty} h^2(\xi) d\xi}{\int_0^{+\infty} h^2(\xi) d\xi} \right),$$

where $E(\cdot)$ is expressed in dB. The work presented in [67] also provides an approximation method of the EDC in ISM-simulated RIR.

The RT T_{60} defined as the time required by the RIR energy $E(\cdot)$, to decay from -5dB to -65dB:

$$T_{60} = E^{-1}(-65) - E^{-1}(-5),$$

where $E^{-1}(\kappa)$ corresponds to the time lag t_κ for which $E(t_\kappa) = \kappa$. It is demonstrated that the accuracy of the EDC approximation method allows to generate a RT in the simulated impulse responses within a small percentage of the targeted value [69].

In addition, as the required simulation times grow exponentially with the considered order of reflection, a significant and well known drawback of the ISM implementation is still its intrinsic computational cost. Due to the tail of the simulated RIR $h(t)$ is noise like, as shown in Figure 3.2, by decomposing the RIR into early reflections and late (diffuse) reverberation with a proper cut-off selection, where the early part is simulated according to the standard ISM technique, and the diffuse RIR tail is synthesised as a decaying random noise, a diffuse reverberation ISM (DRM-ISM) proposed in [68] allows for a significant reduction of the computational costs in RIR implementation.

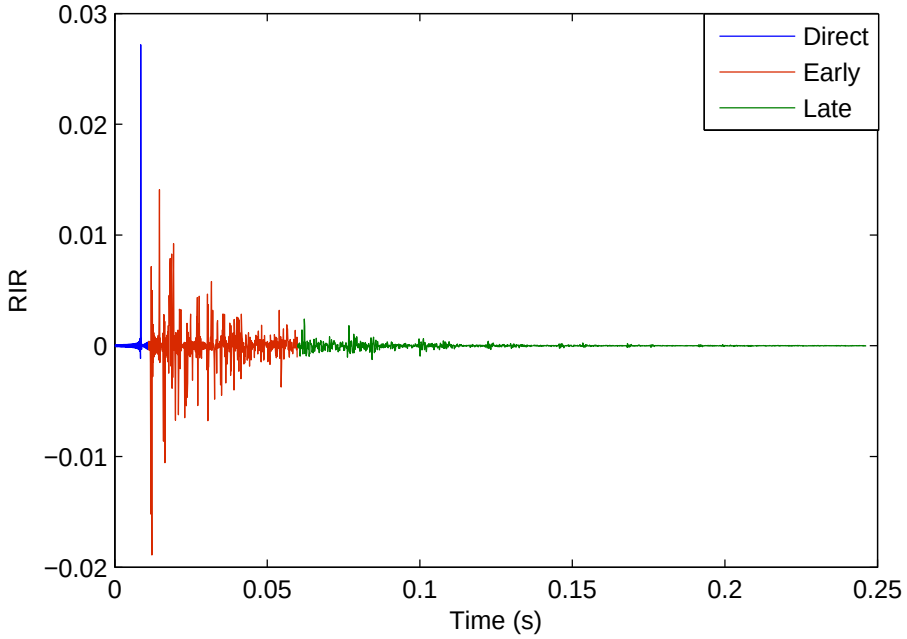


Figure 3.2: Example of the RIR.

3.2 Indoor Beamformer Design

Given a small rectangular room with M-elements microphone array, where behind each elements is an $L-$tap FIR filter, assume that the microphone elements have been set at the fixed points $\boldsymbol{r}_i,\ i = 1, 2, .., M$. With the same definition as frequency responses $W_i(\boldsymbol{w}, f, L)$, $i = 1, 2, ..., M$ and steering vector $\boldsymbol{d}_0(f, L)$ defined in Section 2.2, and the frequency domain RIR vectors $\boldsymbol{R}(\boldsymbol{r}, f) = [R_1(\boldsymbol{r}, f), ..., R_N(\boldsymbol{r}, f)]^T$ have been estimated by the ISM room simulation model. We can denote the beamformer output as

$$G(\boldsymbol{r}, f, L) = \sum_{i=1}^{M} R_i(\boldsymbol{w}, f, L) W_i(\boldsymbol{r}, f) = \boldsymbol{R}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L),$$

and establish the optimization problem for beamformer coefficients solving based on the desired response $G_d(\boldsymbol{r}, f, L)$ as:

$$\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega} \left\| \boldsymbol{R}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L) - G_d(\boldsymbol{r}, f, L) \right\|_p^p. \tag{3.4}$$

In the following, we use the discretization method to solve the minimax problem (3.4) numerically, and use a multi-dimensional grid region Ω_K for approximating Ω with uniform grid containing K mesh points in each dimension of the spatial-frequency domain.

3.2.1 Implementation Models

For the beamformer design problem (3.4), by combining the RIR vectors $\boldsymbol{R}(\boldsymbol{r}, f)$ with the steering vector $\boldsymbol{d}_0(f, L)$ together, we rearrange the beamformer as

$$\boldsymbol{R}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L) = \boldsymbol{w}^T\boldsymbol{d}(\boldsymbol{r}, f, L). \tag{3.5}$$

Expanding the complex functions as

$$\boldsymbol{d}(\boldsymbol{r}, f, L) = \boldsymbol{d}_1(\boldsymbol{r}, f) + j\boldsymbol{d}_2(\boldsymbol{r}, f), \quad G_d(\boldsymbol{r}, f, L) = G_{d_1}(\boldsymbol{r}, f, L) + jG_{d_2}(\boldsymbol{r}, f, L), \tag{3.6}$$

and denoting

$$\boldsymbol{u}(\boldsymbol{r}, f, L) = (\boldsymbol{w}^T\boldsymbol{d}_1(\boldsymbol{r}, f) - G_{d_1}(\boldsymbol{r}, f, L)), \quad \boldsymbol{v}(\boldsymbol{r}, f, L) = (\boldsymbol{w}^T\boldsymbol{d}_2(\boldsymbol{r}, f) - G_{d_2}(\boldsymbol{r}, f, L)).$$

The minimax filter design problem (3.4) can be rewritten as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega_K} \left\|\boldsymbol{u}(\boldsymbol{r}, f, L) + j\boldsymbol{v}(\boldsymbol{r}, f, L)\right\|.$$

Using the L_1-norm, the design problem (3.4) can be expressed as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega_K} \left|\boldsymbol{u}(\boldsymbol{r}, f, L)\right| + \left|\boldsymbol{v}(\boldsymbol{r}, f, L)\right|. \tag{3.7}$$

To convert it into a linear programming, one approach is to introduce two new variables for controlling the real part and imaginary part separately as

$$z_1 = \max_{(\boldsymbol{r},f)\in\Omega_K} \left|\boldsymbol{u}(\boldsymbol{r}, f, L)\right|, \; z_2 = \max_{(\boldsymbol{r},f)\in\Omega_K} \left|\boldsymbol{v}(\boldsymbol{r}, f, L)\right|,$$

then an implementation model for problem (3.4) can be established.

Model I:

$$\begin{array}{ll} \min\limits_{\substack{\boldsymbol{w}\in\mathbb{R}^{ML}\\ \boldsymbol{z}\in\mathbb{R}^2}} & z_1+z_2 \\ s.t. & |\boldsymbol{u}(\boldsymbol{r},f,L)|\leqslant z_1, \quad \forall(\boldsymbol{r},f)\in\Omega_K, \\ & |\boldsymbol{v}(\boldsymbol{r},f,L)|\leqslant z_2, \quad \forall(\boldsymbol{r},f)\in\Omega_K. \end{array} \tag{3.8}$$

The developed optimization model (3.8) is a linear programming problem, it can be formulated as a standard linear programming model (3.9) in matrix notation.

$$\begin{array}{ll} \min\limits_{\mathcal{Z}\in R^{ML+2}} & c^T\mathcal{Z} \\ s.t. & \mathcal{A}(\boldsymbol{r},f,L)\mathcal{Z}-b\leqslant 0, \ \forall(\boldsymbol{r},f)\in\Omega_K, \end{array} \tag{3.9}$$

where $\mathcal{A}(\boldsymbol{r},f,L)$, $\mathcal{Z}$, c and b are defined the same as those in problem (2.13).

The above is also a standard linear programming problem, and can be solved by `linprog` function provided by the optimization toolbox from MATLAB. This indoor beamformer design model (3.9) employs the whole RIRs to form the desired response $G_d(\boldsymbol{r},f,L)$ in both passband and stopband, whereas for dereverberation, the main task is to remove the sound reflections include early reflection and late reverberation. However, the early reflection and late reverberation are exactly generated by the early and late parts of RIRs, thus, developing criteria for reflections suppression are more reasonable, we study them in the following.

3.2.2 Multi-criteria Models

Take a look at the plot of the RIR in Figure 3.2, it is noted that it can be decoupled into three parts, namely the direct path response, early reflections and late reverberation, with different delays and different energy decay. We denote these three part of the RIRs as $\boldsymbol{R}_D(\boldsymbol{r},f)$, $\boldsymbol{R}_E(\boldsymbol{r},f)$ and $\boldsymbol{R}_L(\boldsymbol{r},f)$, that is

$$\boldsymbol{R}(\boldsymbol{r},f)=\boldsymbol{R}_D(\boldsymbol{r},f)+\boldsymbol{R}_E(\boldsymbol{r},f)+\boldsymbol{R}_L(\boldsymbol{r},f). \tag{3.10}$$

If we model the desired response with the direct path response, then the early reflections and late reverberation are expected to be suppressed.

In view of this, we can extract the reverberation part from the RIRs to modify the indoor beamformer design model (3.4) for desired response formulation and dereverberation respectively. One simple multi-criteria model is

$$\begin{aligned}\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega_K} \quad & \rho_1 \left\| \boldsymbol{R}_D^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L) - G_d(\boldsymbol{r},f,L) \right\|_p^p \\ & +\rho_2 \left\| (\boldsymbol{R}_E(\boldsymbol{r},f) + \boldsymbol{R}_L(\boldsymbol{r},f))^H \boldsymbol{W}(\boldsymbol{w},f,L) \right\|_p^p,\end{aligned} \tag{3.11}$$

where $\rho_1 > 0$ and $\rho_2 > 0$ are weighting parameters. In this model, the total reverberation is suppressed as a whole.

On the other hand, due to the early reflections have most of the energy of the reverberation, a better multi-criteria model for indoor beamformer design can be established as

$$\begin{aligned}\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega_K} \quad & \rho_1 \left\| \boldsymbol{R}_D^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L) - G_d(\boldsymbol{r},f,L) \right\|_p^p \\ & +\rho_2 \left\| \boldsymbol{R}_E^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L) \right\|_p^p + \rho_3 \left\| \boldsymbol{R}_L^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L) \right\|_p^p,\end{aligned} \tag{3.12}$$

where $\rho_1 > 0$, $\rho_2 > 0$ and $\rho_3 > 0$ are weighting parameters.

For the multi-criteria problems (3.11) and (3.12), we can also convert them into the linear programming problems like Model I based on the L_1-norm measure.

If we denote the reverberation responses for problem (3.11) as

$$\boldsymbol{R}_R(\boldsymbol{r},f) = \boldsymbol{R}_E(\boldsymbol{r},f) + \boldsymbol{R}_L(\boldsymbol{r},f), \quad \forall(\boldsymbol{r},f) \in \Omega_K,$$

and rearrange

$$\boldsymbol{R}_D^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L) = \boldsymbol{h}^T\boldsymbol{d}_D(\boldsymbol{r},f,L), \quad \boldsymbol{R}_R^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L) = \boldsymbol{h}^T\boldsymbol{d}_R(\boldsymbol{r},f,L),$$

and expand the complex functions as

$$\boldsymbol{d}_D(\boldsymbol{r},f,L) = \boldsymbol{d}_{D1}(\boldsymbol{r},f,L) + j\boldsymbol{d}_{D2}(\boldsymbol{r},f,L), \quad \boldsymbol{d}_P(\boldsymbol{r},f,L) = \boldsymbol{d}_{P1}(\boldsymbol{r},f,L) + j\boldsymbol{d}_{P2}(\boldsymbol{r},f,L).$$

Denote

$$\boldsymbol{u}_D(\boldsymbol{r},f,L)=(\boldsymbol{w}^T\boldsymbol{d}_{D1}(\boldsymbol{r},f,L)-G_{d_1}(\boldsymbol{r},f,L)),$$

$$\boldsymbol{v}_D(\boldsymbol{r},f,L)=(\boldsymbol{w}^T\boldsymbol{d}_{D2}(\boldsymbol{r},f,L)-G_{d_2}(\boldsymbol{r},f,L)),$$

$$\boldsymbol{u}_P(\boldsymbol{r},f,L)=\boldsymbol{w}^T\boldsymbol{d}_{P1}(\boldsymbol{r},f,L),\quad \boldsymbol{v}_P(\boldsymbol{r},f,L)=\boldsymbol{w}^T\boldsymbol{d}_{P2}(\boldsymbol{r},f,L),$$

the problem (3.11) can be reformulated as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{ML}}\max_{(\boldsymbol{r},f)\in\Omega_K}\rho_1\left\{|\boldsymbol{u}_D(\boldsymbol{r},f,L)+j\boldsymbol{v}_D(\boldsymbol{r},f,L)|\right\}+\rho_2\left\{|\boldsymbol{u}_P(\boldsymbol{r},f,L)+j\boldsymbol{v}_P(\boldsymbol{r},f,L)|\right\}.$$

We also use the L_1-norm and introduce the variables as

$$z_1=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{u}_D(\boldsymbol{r},f,L)\right|,\ z_2=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{v}_D(\boldsymbol{r},f,L)\right|,$$

$$z_3=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{u}_P(\boldsymbol{r},f,L)\right|,\ z_4=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{v}_P(\boldsymbol{r},f,L)\right|,$$

then the corresponding numerical model for problem (3.11) can be established as

Model II:

$$\begin{array}{ll}\min\limits_{\substack{\boldsymbol{w}\in\mathbb{R}^{ML}\\ \boldsymbol{z}\in\mathbb{R}^4}} & \boldsymbol{\rho}^T\boldsymbol{z}\\ s.t. & |\boldsymbol{u}_D(\boldsymbol{r},f,L)|\leqslant z_1,\quad\forall(\boldsymbol{r},f)\in\Omega_K,\\ & |\boldsymbol{v}_D(\boldsymbol{r},f,L)|\leqslant z_2,\quad\forall(\boldsymbol{r},f)\in\Omega_K,\\ & |\boldsymbol{u}_P(\boldsymbol{r},f,L)|\leqslant z_3,\quad\forall(\boldsymbol{r},f)\in\Omega_K,\\ & |\boldsymbol{v}_P(\boldsymbol{r},f,L)|\leqslant z_4,\quad\forall(\boldsymbol{r},f)\in\Omega_K,\end{array}\tag{3.13}$$

where $\boldsymbol{z}=[z_1\ z_2\ z_3\ z_4]^T$, and $\boldsymbol{\rho}=[\rho_1\ \rho_1\ \rho_2\ \rho_2]^T$ are the weighting parameters developed in (3.11).

For the multi-criteria problem (3.12), we can divide the reverberation responses $\boldsymbol{R}_R(\boldsymbol{r},f)$ into early reflections and late reverberation, and rearrange them as

$$\boldsymbol{R}_E^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L)=\boldsymbol{h}^T\boldsymbol{d}_E(\boldsymbol{r},f,L),\quad \boldsymbol{R}_L^H(\boldsymbol{r},f)\boldsymbol{W}(\boldsymbol{w},f,L)=\boldsymbol{h}^T\boldsymbol{d}_L(\boldsymbol{r},f,L),$$

and expand to

$$\boldsymbol{d}_E(\boldsymbol{r},f,L)=\boldsymbol{d}_{E1}(\boldsymbol{r},f,L)+j\boldsymbol{d}_{E2}(\boldsymbol{r},f,L),\quad \boldsymbol{d}_L(\boldsymbol{r},f,L)=\boldsymbol{d}_{L1}(\boldsymbol{r},f,L)+j\boldsymbol{d}_{L2}(\boldsymbol{r},f,L).$$

Denote

$$\begin{aligned}\boldsymbol{u}_E(\boldsymbol{r},f,L)&=\boldsymbol{w}^T\boldsymbol{d}_{E1}(\boldsymbol{r},f,L),\quad \boldsymbol{v}_E(\boldsymbol{r},f,L)=\boldsymbol{w}^T\boldsymbol{d}_{E2}(\boldsymbol{r},f,L).\\ \boldsymbol{u}_L(\boldsymbol{r},f,L)&=\boldsymbol{w}^T\boldsymbol{d}_{L1}(\boldsymbol{r},f,L),\quad \boldsymbol{v}_L(\boldsymbol{r},f,L)=\boldsymbol{w}^T\boldsymbol{d}_{L2}(\boldsymbol{r},f,L).\end{aligned}$$

and introduce the additional variables as

$$\begin{aligned}z_3&=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{u}_E(\boldsymbol{r},f,L)\right|,\ z_4=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{v}_E(\boldsymbol{r},f,L)\right|,\\ z_5&=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{u}_L(\boldsymbol{r},f,L)\right|,\ z_6=\max_{(\boldsymbol{r},f)\in\Omega_K}\left|\boldsymbol{v}_L(\boldsymbol{r},f,L)\right|,\end{aligned}$$

then the corresponding numerical model for problem (3.11) can be established as

Model III:

$$\begin{aligned}\min_{\substack{\boldsymbol{w}\in\mathbb{R}^{ML}\\ \boldsymbol{z}\in\mathbb{R}^6}}\quad & \boldsymbol{\rho}^T\boldsymbol{z}\\ s.t.\quad & \boldsymbol{U}(\boldsymbol{r},f,L)\leqslant\boldsymbol{z},\quad \forall(\boldsymbol{r},f)\in\Omega_K,\end{aligned}\tag{3.14}$$

where

$$\boldsymbol{U}(\boldsymbol{r},f,L)=[|\boldsymbol{u}_D(\boldsymbol{r},f,L)|\ \ |\boldsymbol{v}_D(\boldsymbol{r},f,L)|\ \ |\boldsymbol{u}_E(\boldsymbol{r},f,L)|\ \ \ldots|\boldsymbol{v}_L(\boldsymbol{r},f,L)|]^T,$$

and $\boldsymbol{z}=[z_1\ldots z_6]^T$, $\boldsymbol{\rho}=[\rho_1\ \ldots\ \rho_6]^T$.

The above multi-criteria models Model II and Model III can also be transformed into similar linear programming problems as in (3.9) of Model I.

3.3 Illustration Examples

In this section, we evaluate the performances of the beamformers designed based on the proposed models. One simple rectangular room with a microphone array is

defined to generate a reverberant environment, it is suitable for the fast-ISM room simulator to estimate the RIRs. In addition, we apply an audio signal to evaluate the performance of the designed beamformers on noise suppression in the stopband.

3.3.1 Broadband Beamformer Design in Acoustic Room

Specifically, we define a simple $8m \times 4m \times 3m$ rectangular office with adjustable absorption coefficients characterizing the room surface, which is also measured by reverberation time T_{60} for convenience. To define an microphone array for beamforming, an equispaced linear array with five microphone elements are setup at We define a simple $8m \times 4m \times 3$m rectangular office with adjustable absorption coefficients $\boldsymbol{\alpha}$ characterizing the room surface. To define the microphone array, an equispaced linear array with 5-elements microphone array is setup at

$$\mathsf{Mic} = \{\mathsf{(2,3.9,1.5)},\ \mathsf{(2,3.95,1.5)},\ \mathsf{(2,4,1.5)},\ \mathsf{(2,4.05,1.5)},\ \mathsf{(2,4.1,1.5)}\}$$

in meter, respectively. The array spacing is $0.05m$ to avoid spatial aliasing for the frequency of interest, and a 10-tap FIR filter is fixed behind each microphone element.

In the room acoustics, the RIRs to each microphone element may have big difference from different source points due to the reverberation, although they have the same distance. Therefore, its wiser to define the specific position in space to evaluate the beamforming performance. And in following example, we define the specific passband region as

$$\Omega_p = \{(\boldsymbol{r}, f) \mid x = 1m,\ |y-4| \leqslant 0.4m,\ z = 1.5m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\},$$

and the stopband are defined as

$$\begin{aligned}\Omega_s = &\{(\boldsymbol{r}, f) \mid x = 1m,\ |y-4| \leqslant 0.4m,\ z = 1.5m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\}\ldots\\ &\cup\{(\boldsymbol{r}, f) \mid x = 1m,\ 1.5m \leqslant |y-4| \leqslant 3.0m,\ z = 1.5m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\}\ldots\\ &\cup\{(\boldsymbol{r}, f) \mid x = 1m,\ 1.5m \leqslant |y-4| \leqslant 3.0m,\ z = 1.5m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\}.\end{aligned}$$

The diagram of the above specific configuration is shown in Figure 3.3.

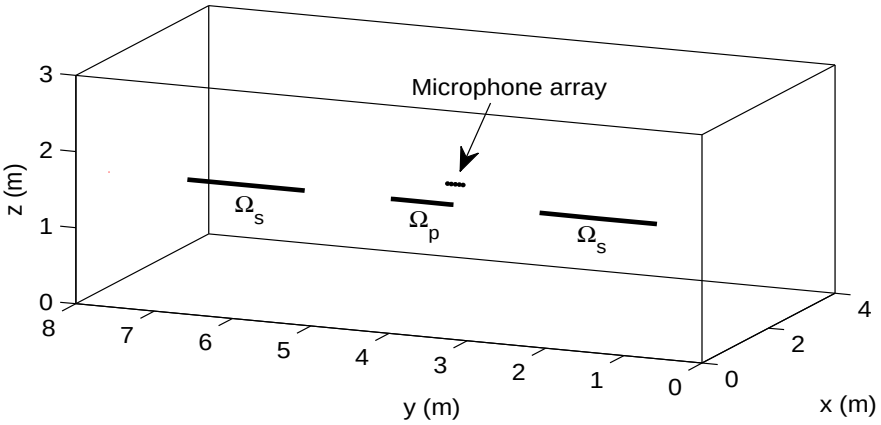


Figure 3.3: Room setup for indoor beamformer design.

The desired response function $G_d(\boldsymbol{r}, f, L)$ in the passband and stopband is defined the same as (2.14). For numerical calculation, we discrete the spatial-frequency domain of the passband and stopband into a grid of 30×30 frames. And use a more density grid scheme 120×120 to verify the beamformer performances.

In a typical room environment, a sound wave strikes a surface, a certain fraction of it is absorbed and a certain amount is transmitted into the surface; these are the energy loss from the room and the fractional loss is characterized by the absorption coefficients $\boldsymbol{\alpha}$. The overall effect can be represented by the reverberation time T_{60}. We will choose different reverberation time T_{60} to verify the proposed indoor beamformer design models with uniform absorption coefficients for all room boundaries.

The implementation of Model I is a standard linear programming problem and the beamformer coefficients can be solved from it directly, whereas the beamformers designed from the multi-criteria decision Model II and Model III are related to the weighting parameters $\boldsymbol{\phi}$. For Model III, we can vary the weighting parameters $\boldsymbol{\phi}$ to obtain the Pareto optima set corresponding to different indoor beamformer design. From Figure 3.4, clearly we cannot find a design to minimise both early reflections and late reverberation and we need to trade off the performance of both suppressions.

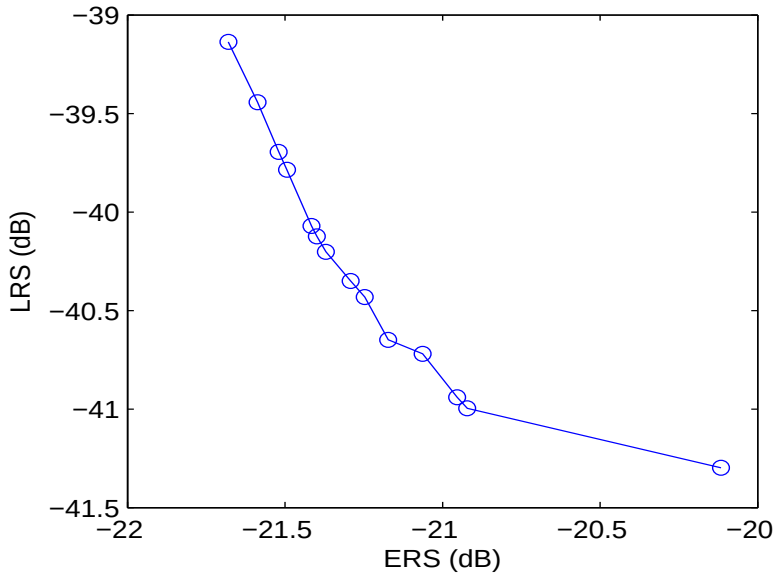


Figure 3.4: Conflicting nature of early reflections suppression and late reverberation suppression (the measurements of them are defined as ERS and LRS in the following) in the acoustic room with reverberation time $T_{60} = 0.2s$.

To explain the multi-criteria Model III, we give an example to illustrate the conflicting of the early reflections suppression and late reverberation suppression in Figure 3.4. From that figure, clearly we cannot find a design to minimize both early reflections and late reverberation and we need to trade off the performance of both suppressions. It demonstrates that the separation scheme for suppressing early reflections and late reverberation in Model III is meaningful.

Assume all the indoor beamformer design models can achieve the desired response in the passband, then the other criterion of beamforming performance is the noise reduction and reverberation suppression in the stopband. We shall use the following performance indicators to measure and compare between different models: the direct path response suppression (DRS (in dB)):

$$\frac{1}{||\Omega_s||} \sum_{(\boldsymbol{r},f)\in\Omega_s} \left\| \boldsymbol{R}_D^T(\boldsymbol{r},f)\boldsymbol{H}(\boldsymbol{h},f) \right\|_1 ,$$

the early reflections suppression (ERS (in dB)):

$$\frac{1}{||\Omega_s||}\sum_{(\boldsymbol{r},f)\in\Omega_s}\left\|\boldsymbol{R}_E^T(\boldsymbol{r},f)\boldsymbol{H}(\boldsymbol{h},f)\right\|_1,$$

and the late reverberation suppression (LRS (in dB)):

$$\frac{1}{||\Omega_s||}\sum_{(\boldsymbol{r},f)\in\Omega_s}\left\|\boldsymbol{R}_L^T(\boldsymbol{r},f)\boldsymbol{H}(\boldsymbol{h},f)\right\|_1.$$

And the overall performance of the designed beamformer will be the total reverberation suppression (TRS = DRS +ERS+LRS). We summarize the results in Table 3.1 under different acoustic room conditions with four kinds of reverberation time T_{60} = 0.05s, 0.1s, 0.15s and 0.2s, respectively.

Table 3.1: Summary of the beamforming performance on desired response formulation, early reflections and late reverberation suppression.

T_{60}	Model I			
	DRS	ERS	LRS	TRS
0.05	-14.1400	-39.2890	-61.2107	-14.1267
0.1	-14.1377	-24.4809	-55.4227	-13.7536
0.15	-14.1377	-19.4207	-47.1581	-13.0090
0.2	-14.1364	-16.6324	-38.2722	-12.1865
T_{60}	Model II			
	DRS	ERS	LRS	TRS
0.05	-21.4810	-45.1816	-65.9210	-21.4624
0.1	-20.5304	-29.4417	-59.1716	-20.0049
0.15	-20.4440	-21.9449	-41.0765	-18.0977
0.2	-20.2677	-21.1781	-38.8466	-17.6556
T_{60}	Model III			
	DRS	ERS	LRS	TRS
0.05	-22.4063	-44.4467	-67.2945	-22.3791
0.1	-21.6910	-29.8507	-59.7177	-21.0730
0.15	-21.6191	-24.7314	-50.3162	-19.8880
0.2	-20.8478	-21.2104	-40.7866	-17.9921

From Table 3.1, we can see that the proposed indoor beamformer design models are all effective for noise reduction and reverberation suppression for small reverber-

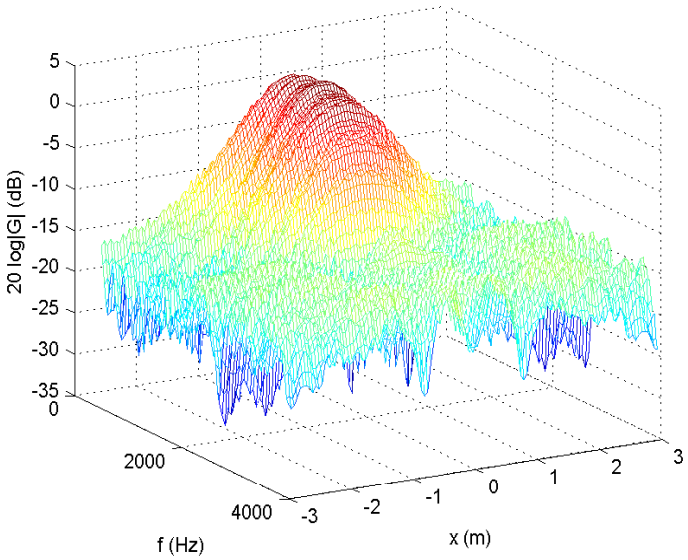
ation time T_{60} cases. But the multi-criteria decision based on Model II and Model III have better performance than Model I in general for all performance indicators, and it can be seen that Model II and III has bigger improvements on on DRS than Model I. In particular, Model III performs slightly better than Model II in the TRS, which gives credit to the flexibility introduced into Model III in controlling different weights. To take a closer look at the the overall performance of the designed beamformer, we give two performance graphs from Model III in the acoustic room with reverberation time $T_{60} = 0.1$s and $T_{60} = 0.2$s in the Figure 3.5 (in these figures, x(m)-label and f(Hz)-label denote space and frequency domains, respectively).

Moreover, to examine the robustness of the designed beamformer for estimation error in some of the parameters, we evaluate the off-design performance for the designed beamformer with $T_{60} = 0.1$s and perturb it to $T_{60} = \{0.08s,\ 0.09s,\ 0.11s,\ 0.12\}s$. We find that the designed beamformers from the proposed models are all robust to little perturbation of measurement on room acoustics, the results are summarized in Table 3.2.

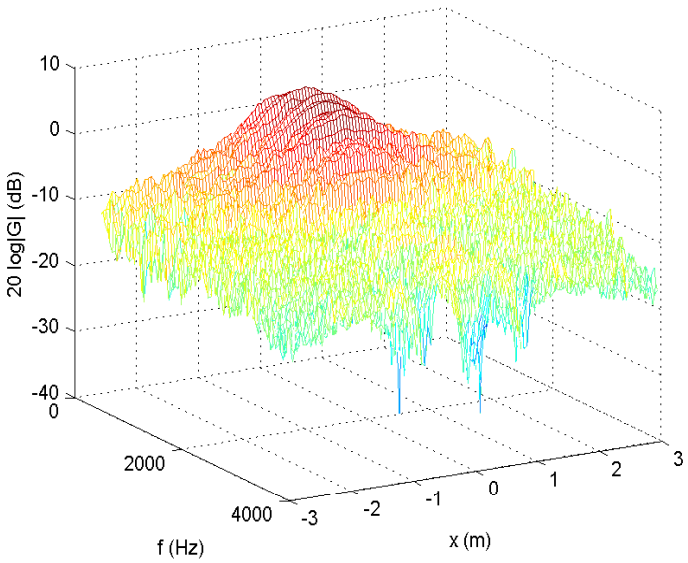
3.3.2 Testing Signal Suppression in Stopband

In this example, we want to input a testing signal at the stopband, and evaluate the performance on the suppression of the proposed indoor beamforming approaches. We firstly cut a testing signal from a broadband male speech "*Dots of light betrayed the black cat*" as the inputting signal, then we setup inputting placement at the stopband Ω_s of the acoustic room defined in the above Example 3.3.1 with reverberation time $T_{60} = 0.1s$.

Then, we can use the beamformers solved by the proposed Model I, Model II and Model III to process the signals captured by the microphone array. In comparison, we use the signal captured by the center element of the microphone array as the unprocessed reference. Simple tests are to put the inputting signal at the placement



(a) $T_{60} = 0.1s$

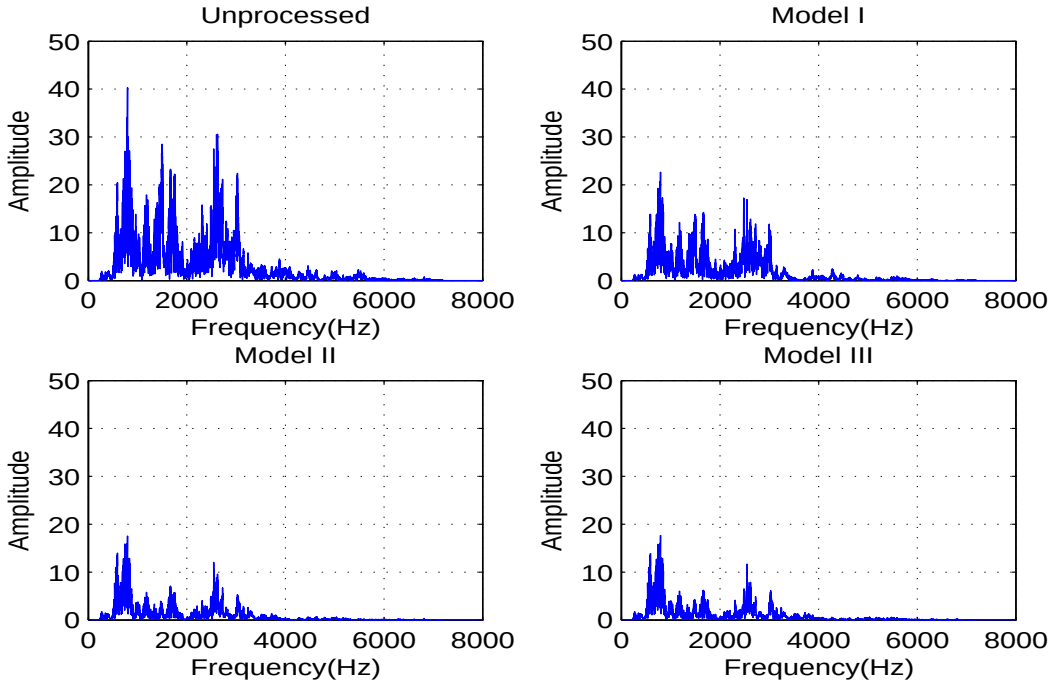


(b) $T_{60} = 0.2s$

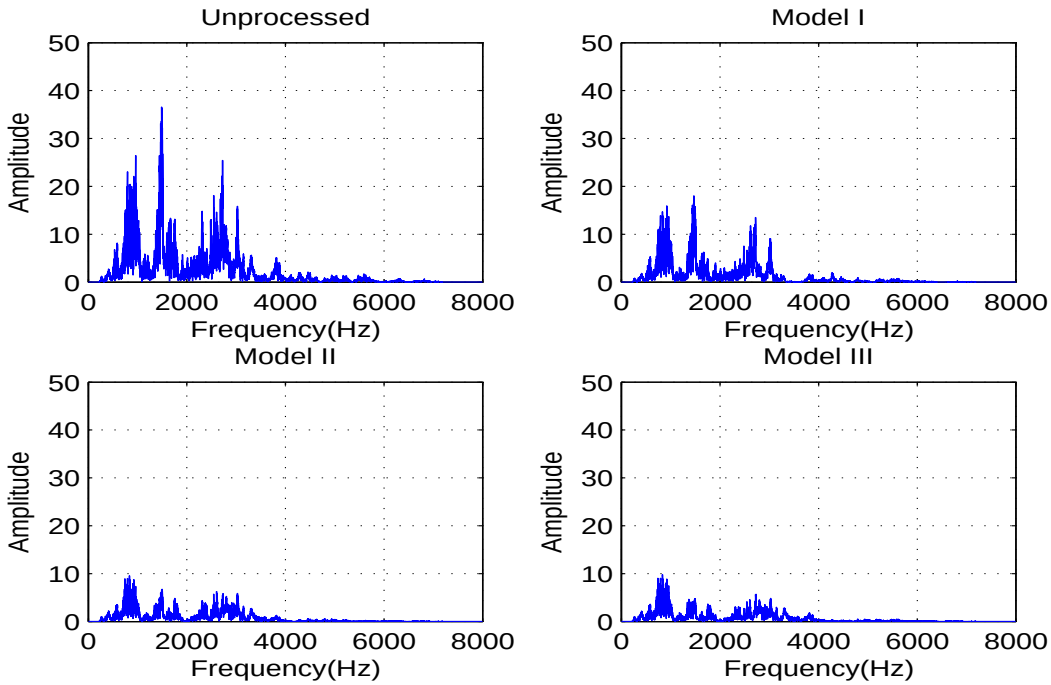
Figure 3.5: Performance of the indoor beamformering given by Model III.

$\boldsymbol{r} = (1, 2, 1.5) \in \Omega_s$ and $\boldsymbol{r} = (1, 1, 1.5) \in \Omega_s$ to verify the amplitude suppressing in the Figure 7.7, respectively.

It can be seen that all the beamformers designed by the proposed approaches are



(a) Noise at $\boldsymbol{r} = (1, 2, 1.5)$



(b) Noise at $\boldsymbol{r} = (1, 1, 1.5)$

Figure 3.6: Examples of beamforming performance on the noise suppression in the stop-band.

Table 3.2: Summary of off-design performance for indoor beamformer with $T_{60} = 0.1s$.

T_{60}	Model I			
	DRS	ERS	LRS	TRS
0.08	-14.1395	-28.1768	-57.3646	-13.9712
0.09	-14.1392	-26.1214	-56.0164	-13.8722
0.11	-14.1387	-23.1337	-54.5959	-13.6228
0.12	-14.1385	-22.0046	-53.4133	-13.4806
T_{60}	Model I			
	DRS	ERS	LRS	TRS
0.08	-21.1037	-33.8798	-62.1626	-20.8800
0.09	-22.0220	-31.7308	-60.2814	-21.5802
0.11	-20.7166	-28.2511	-57.8366	-20.0103
0.12	-20.5373	-26.9751	-56.1224	-19.6475
T_{60}	Model I			
	DRS	ERS	LRS	TRS
0.08	-20.9454	-33.4870	-62.0590	-20.7097
0.09	-21.8461	-31.6153	-60.6142	-21.4102
0.11	-22.0703	-28.5929	-58.3538	-21.1962
0.12	-21.2157	-26.9252	-56.7525	-20.1816

effect to suppress signals in the stopband, and it seems the beamformer designed from Model III is more effective than Model I and Model II. To measure the performance on stopband suppression, we define the suppression function as following

$$\text{Supp} = 10 \log_{10} \frac{||\boldsymbol{S}(f)||_2^2}{||\boldsymbol{Y}(f)||_2^2} \quad \text{(dB)}, \tag{3.15}$$

where $\boldsymbol{S}(f)$ and $\boldsymbol{Y}(f)$ represent the frequency spectrums of the inputting signal and the beamforming output. And we select the stopband placement from $\boldsymbol{r} = (1, 1, 1.5)$ to $\boldsymbol{r} = (1, 2.5, 1.5)$ with $0.1m$ spacing to evaluate the noise suppression, and for convenience, we just use the $y-$axis as the variable in denoting, a summary of results are plotted in the following Figure 3.7.

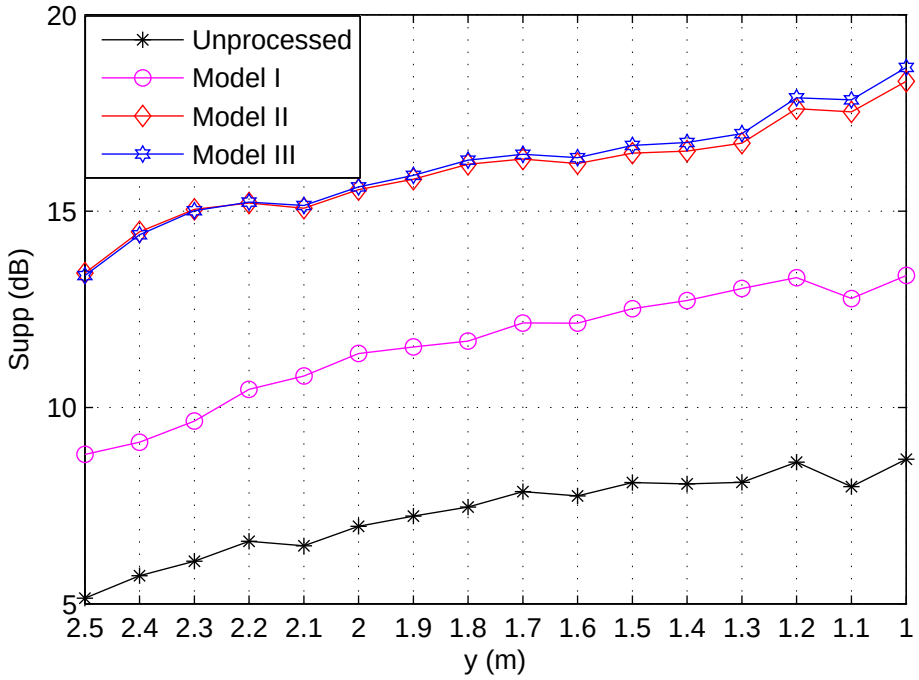


Figure 3.7: Evaluation of the noise suppression in stopband.

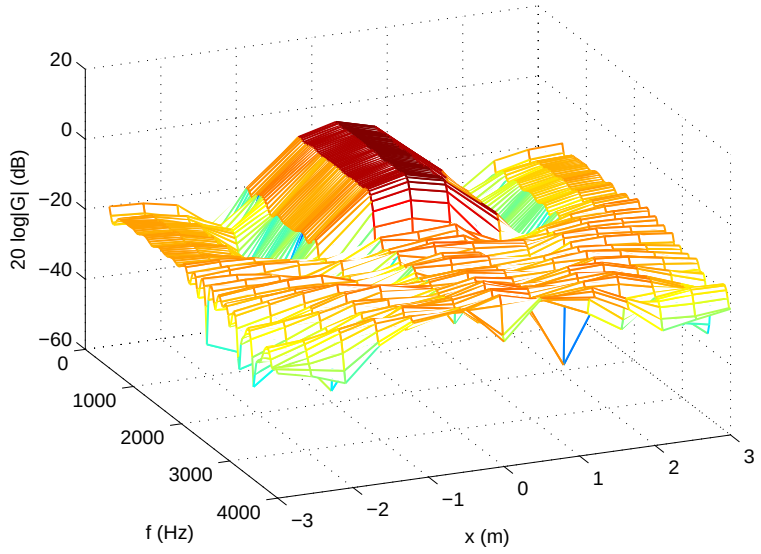
From these results, it can be seen that the proposed indoor beamformer design techniques are effect and stable, and the multi-criteria based Model II and Model III are more effective than Model I, and the proposed Model III is the best one.

Chapter 4

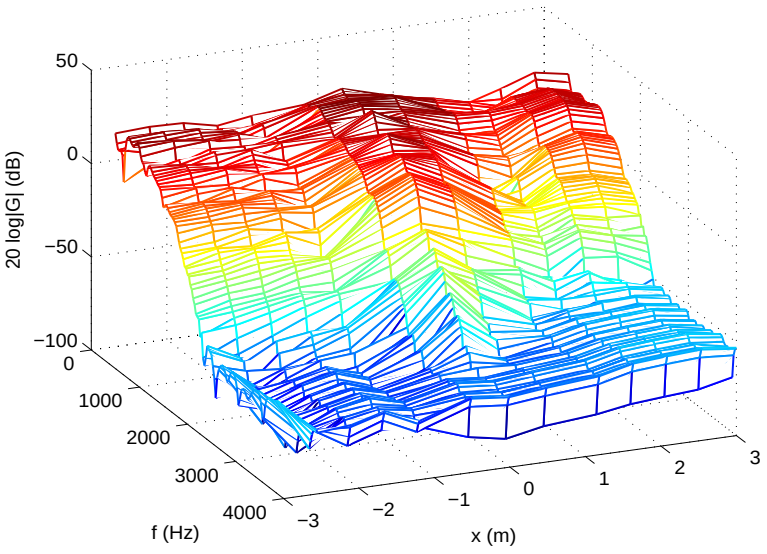
Barrier Beamformer Design

It is usually barriers exist between the sources and sensors in practice, and the common approaches used for beamformer design suffered in accuracy under barrier condition, a simple example for illustrating the same beamformer designed by the method developed from (2.13) in open air and barrier conditions are given in the following Figure 4.1 (the more detail on barrier configuration is defined in the later Section 4.3).

The main cause of failure of the beamformer under barrier condition is that the influence of barrier acoustics. Thus, find effective method to simulate the barrier acoustics and develop the beamformer design model by using the barrier impulse responses (BIRs) are necessary. In this chapter, we introduce the conservation element and solution element (CE/SE) method to estimate the BIRs and study the beamformer design problem under barrier condition for the first time.



(a) Good performance in the open air



(b) Poor performance with barrier around

Figure 4.1: Overall performance of the common designed beamformer under different conditions.

4.1 Barrier Impulse Response Estimation

Acoustic signal transmitted through fluid which is modeled as a continuum and its dynamic behavior is governed by the principles of conservation of mass, momentum and energy. The acoustic fluctuations are regarded as weak perturbation of density and pressure in the continuum. The governing equations for the fluid flow are the unsteady compressible Navier-Stokes equations with the equation of state of perfect gas. The time-domain impulse response estimation can be calculated by solving 2-D unsteady compressible N-S equations under the numerical platform CE/SE [17, 114]. The dimensionless N-S equations without source term can be written as

$$\frac{\partial U}{\partial t} + \frac{\partial \left(F - F_v\right)}{\partial x} + \frac{\partial \left(G - G_v\right)}{\partial y} = 0, \tag{4.1}$$

where

$$U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix} = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ \rho E \end{bmatrix}, \; F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ \left(\rho E + p\right) u \end{bmatrix},$$

$$G = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix} = \begin{bmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ \left(\rho E + p\right) v \end{bmatrix}, \; F_v = \begin{bmatrix} f_{v1} \\ f_{v2} \\ f_{v3} \\ f_{v4} \end{bmatrix} = C_1 \begin{bmatrix} 0 \\ \tau_{xx} \\ \tau_{xy} \\ \tau_{xx} u + \tau_{xy} v - q_x \end{bmatrix},$$

$$G_v = \begin{bmatrix} g_{v1} \\ g_{v2} \\ g_{v3} \\ g_{v4} \end{bmatrix} = C_1 \begin{bmatrix} 0 \\ \tau_{xxy} \\ \tau_{yy} \\ \tau_{xy} u + \tau_{yy} v - q_y \end{bmatrix},$$

$$\tau_{xx} = \frac{2}{3}\mu \left(2\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right), \tau_{xy} = \mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \tau_{yy} = \frac{2}{3}\mu \left(2\frac{\partial v}{\partial y} - \frac{\partial u}{\partial x}\right),$$

$$E = \frac{p}{\rho\,(\gamma-1)} + \frac{u^2+v^2}{2},\ p = \frac{\rho T}{\gamma C_2},\ q_x = -\frac{\mu}{(\gamma-1)\,PrC_2}\frac{\partial T}{\partial x},$$

$$q_y = -\frac{\mu}{(\gamma-1)\,PrC_2}\frac{\partial T}{\partial y},\ C_1 = \frac{1}{Re},\ C_2 = M^2,$$

where ρ, u, v , E and T are density, x-velocity, y-velocity, specific energy and temperature; τ, μ and γ are component of stress tensor, viscosity and specific heat ratio; Pr, Re and M are Prandtl number, Reynolds number and Mach number respectively. During the calculation, the Jacobian matrices $\frac{\partial F}{\partial U}$, $\frac{\partial G}{\partial U}$ and $\frac{\partial T}{\partial U}$ will be used to find out the spatial derivatives for the conservation.

The N-S equations in the strong conservation forms can be regarded as a set of physical conservation laws, which also governed the flux conservation in both space and time numerically. A framework must be developed to ensure the flux conservation in both space and time is strictly followed to the conservation laws. The space-time CE/SE scheme unifies the treatment of space and time dimensions, which is unique to other numerical method, and utilizes this concept to construct its numerical framework. This space-time CE/SE was employed to solve for the above N-S equations. CE is defined as a control volume for flux conservation while SE is defined to calculate the flux across element boundary in terms of both space and time. CE is constructed by both the mesh node and centroid of all mesh element as showed in Figure 4.2, ABCDEF confined the spatial boundary of the CE where points A, C, E and G are the centroid of the physical mesh while point g is the centroid of the whole CE and it is the location of the solution point. CE/SE is very similar to traditional finite volume method but except the uniform conservation treatment in both space and time, therefore the computational dimensions will be extended by the time axis. At time level n, every CE are referred as CE(g, n). Every CE is combined by 3 basic conservation elements (BCEs), ABGFF'A'B'G',

CDGBB'C'D'G' and EFGDD'E'F'G', they are defined as $\mathrm{CE}_i(g, n)$ where $i = 1, 2, 3$, having spatial area S_1, S_2 and S_3. Furthermore, SE in Figure 4.3 is located at the centroid of the entire CE, and it stored the data of the cell which their spatial and temporal derivatives are assumed constant within SE. The solution element at spatial location g and n-th time level is denoted by $\mathrm{SE}(g, n)$. All flow variables at any location X within the SE, $\phi(X) = U(X)$, $F(X)$ or $G(X)$, can be approximiated by first order Taylor expansions from $\mathrm{SE}(g, n)$, i.e. $\phi(X)_g = \phi_g + \delta x(\phi_x)_g + \delta y(\phi_y)_g + \delta t(\phi_t)_g$, where $\delta x = x - x_g$, $\delta y = y - y_g$, $\delta t = t - t^n$ and ϕ_x, ϕ_y, ϕ_t are the spatial gradients and time derivatives. Furthermore, the gradients are calculated by chain rule, i.e.

$$(\phi_x)_g = \frac{\partial \phi}{\partial U}(U_x)_g, (\phi_y)_g = \frac{\partial \phi}{\partial U}(U_y)_g, (\phi_t)_g = \frac{\partial \phi}{\partial U}(U_t)_g. \tag{4.2}$$

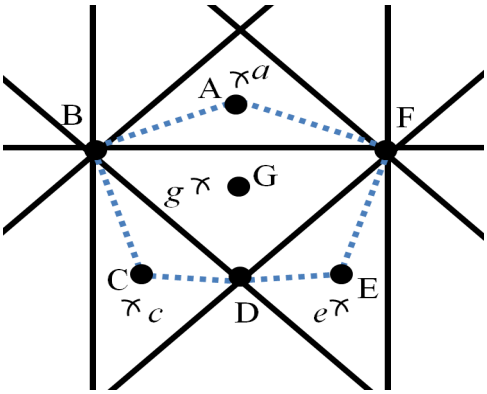
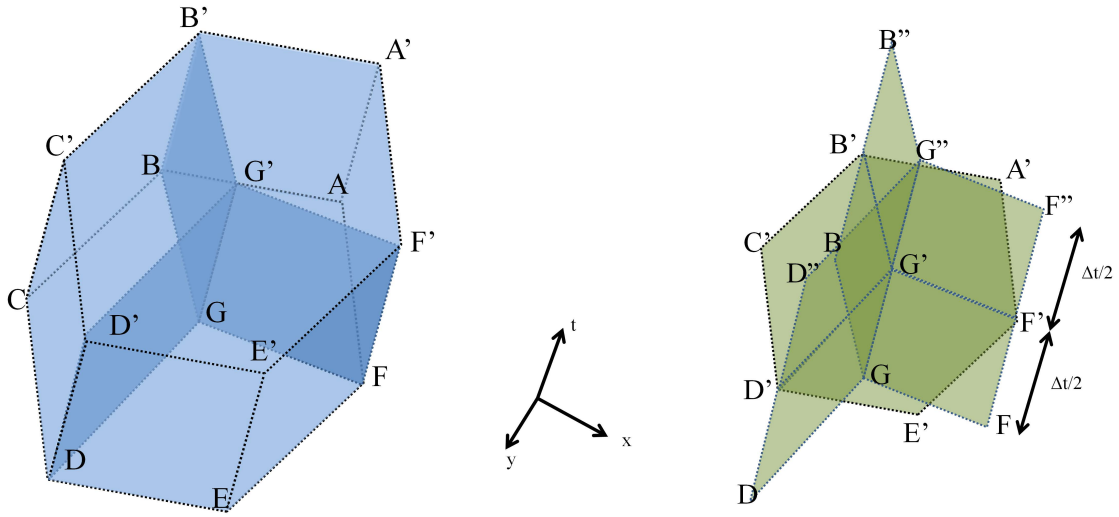


Figure 4.2: CE construction.

In order to determine the solution vector at n-th time level, U_g^n flux conservation within $\mathrm{CE}(g, n)$ must be enforced [65]. The updating procedure of U^n requires the calculation of all the flux on all the surfaces of all the basic CEs, $\mathrm{CE}_i(g, n)$, $i = 1$, 2, 3, exclude the interfaces in between due to cancellation of flux. Such that, only the flux leaving $\mathrm{CE}(g, n)$ are required. Consider the $\mathrm{CE}_1(g,n)$ in Figure 4.3a, the flux



(a) CE (b) SE

Figure 4.3: Definition of CE and SE.

entering it are through surfaces AA'F'F, BB'A'A, and BAFG. Define the known flux through these surfaces leaving $\mathrm{CE}_1(g,n)$ at $(n-1/2)$-th time level as

$$\gamma_1^{n-\frac{1}{2}} = \gamma_{AA'F'F} + \gamma_{BB'A'A} + \gamma_{BAFG}, \tag{4.3}$$

where $\gamma_{AA'F'F}$, $\gamma_{BB'A'A}$ and γ_{BAFG} are the flux leaving the corresponding surface denoted at the subscript. Besides, the evaluation of flux leaving $\mathrm{CE}_1(g,n)$ is calculated from $\mathrm{SE}(a,\, n)$. Similarly the flux leaving $\mathrm{CE}_2(g,n)$ and $\mathrm{CE}_3(g,n)$, $\gamma_2^{n-\frac{1}{2}}$ and $\gamma_3^{n-\frac{1}{2}}$, are from $\mathrm{SE}(c,\, n)$ and $\mathrm{SE}(e,\, n)$ respectively. Consider the top face A'B'C'D'E'F' of the $\mathrm{CE}(g,\, n)$, the flux γ_{top}^n is written as

$$\gamma_{top}^n = (S_1 + S_2 + S_3)\, U_g^n. \tag{4.4}$$

Applying the conservation of flux over $\mathrm{CE}(g,\, n)$,

$$\gamma_{top}^n + \gamma_1^{n-\frac{1}{2}} + \gamma_2^{n-\frac{1}{2}} + \gamma_3^{n-\frac{1}{2}} = 0, \tag{4.5}$$

which leads to the following expression,

$$U_g^n = -\frac{\gamma_1^{n-\frac{1}{2}} + \gamma_2^{n-\frac{1}{2}} + \gamma_3^{n-\frac{1}{2}}}{S_1 + S_2 + S_3}. \tag{4.6}$$

The new spatial gradients of the solution vector $\left(U_g^n\right)_x$ and $\left(U_g^n\right)_y$ will be updated having evaluated the n-th time level solution U_g^n.

Thus, we introduce this CE/SE scheme to simulate the barrier acoustics, and take a simple example for barrier impulse response estimation in configuration of Figure 4.4, we can see that the introduced CE/SE method is effective.

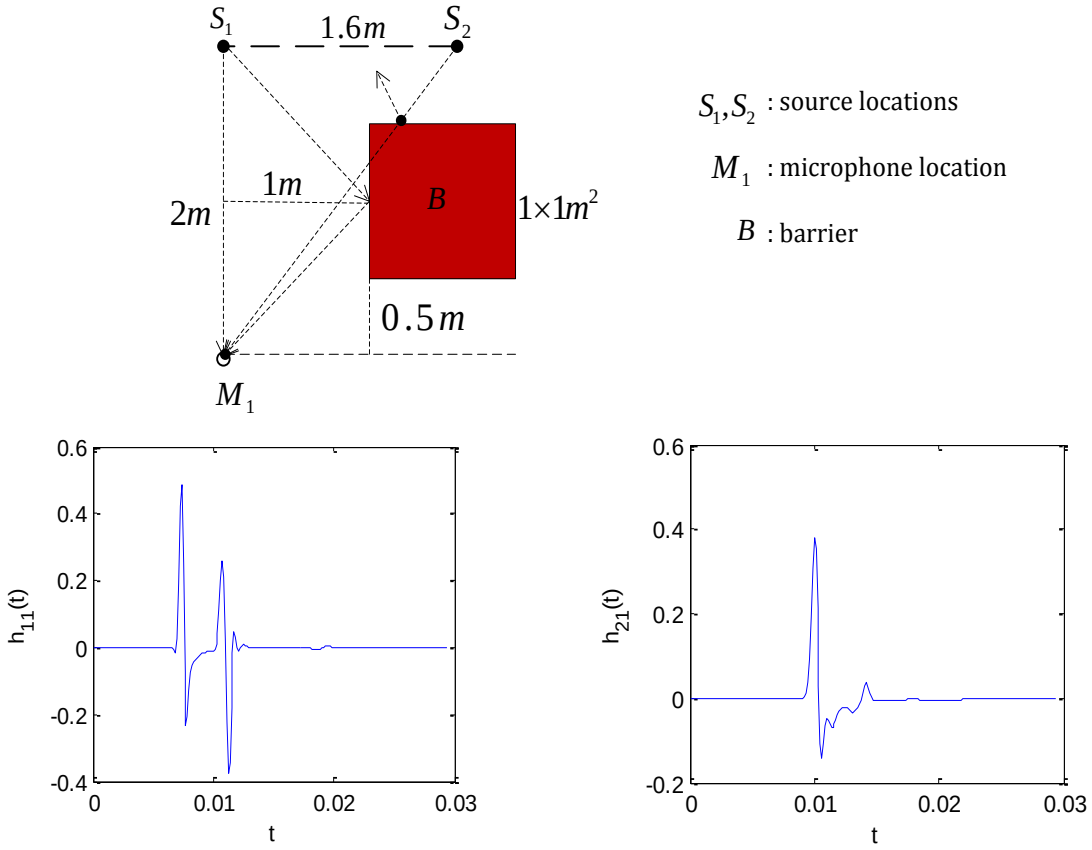


Figure 4.4: A simple example of BIRs estimated by using the proposed CE/SE method: $h_{11}(t)$ and $h_{21}(t)$ are the corresponding two BIRs between microphone location M_1 and source locations S_1, S_2, respectively.

4.2 Barrier Beamformer Design

Consider an acoustic system with M-elements microphone array, that are set up at the fixed points $\boldsymbol{r}_i,\ i = 1, 2, .., M$, and barrier B between the source region $\Omega(\boldsymbol{r})$ and sensor array. Assume there is an L-tap FIR filter behind each microphone element, if the signals received by this microphone array are sampled synchronously at the rate of f_s per second, and using the same notations for frequency responses $W_i(\boldsymbol{w}, f, L)$, $i = 1, 2, ..., M$ and vector $\boldsymbol{d}_0(f, L)$ as those defined in Section 2.2.

Denote the frequency domain BIR from the spatial point $\boldsymbol{r}$ to the i-th microphone element as $B_i(\boldsymbol{r}, f)$, and the desired response as $G_d(\boldsymbol{r}, f, L)$. Then we expect to find a group of coefficients $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_M]^T$ for the FIR filters, such that the beamformer output

$$G(\boldsymbol{r}, f, L) = \sum_{i=1}^{M} H_i(\boldsymbol{w}, f, L) B_i(\boldsymbol{r}, f) = \boldsymbol{B}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L), \tag{4.7}$$

is close to $G_d(\boldsymbol{r}, f, L)$, where $\boldsymbol{B}(\boldsymbol{r}, f) = [B_1(\boldsymbol{r}, f), ..., B_M(\boldsymbol{r}, f)]^T$ is the frequency domain BIR vector.

Thus, we can establish a similar optimization model for the barrier beamformer design problem as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega} \left|\boldsymbol{B}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L) - G_d(\boldsymbol{r}, f, L)\right|, \tag{4.8}$$

where Ω is the spatial-frequency domain for definition of $G_d(\boldsymbol{r}, f, L)$.

By applying the discrete method with a multi-dimensional grid region Ω_K for approximating Ω using uniform grid containing K mesh points in each dimension of the spatial-frequency domain, and combining the barrier impulse response vector $\boldsymbol{B}(\boldsymbol{r}, f)$ with the steering vector $\boldsymbol{d_0}(f, L)$ together. After rearrangement and new functions definition, we can get the similar optimization model based on L_1-norm for

the filter design problem (4.8) as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{ML}} \max_{(\boldsymbol{r},f)\in\Omega_K} |\boldsymbol{u}(\boldsymbol{r},f,L)| + |\boldsymbol{v}(\boldsymbol{r},f,L)|\,, \tag{4.9}$$

where $\boldsymbol{u}(\boldsymbol{r},f,L)$ and $\boldsymbol{v}(\boldsymbol{r},f,L)$ have the same definition as those in problem (3.7).

And the above barrier beamformer design problem (4.9) can also be converted into linear programming problem with two new variables introducing for controlling the real part and imaginary part as

$$z_1 = \max_{(\boldsymbol{r},f)\in\Omega_K} |\boldsymbol{u}(\boldsymbol{r},f,L)|\,,\ z_2 = \max_{(\boldsymbol{r},f)\in\Omega_K} |\boldsymbol{v}(\boldsymbol{r},f,L)|\,,$$

then the linear programming model for problem (4.9) can be established as

$$\begin{array}{ll} \min\limits_{\substack{\boldsymbol{h}\in\mathbb{R}^{ML}\\ \boldsymbol{z}\in\mathbb{R}^2}} & z_1 + z_2 \\ s.t. & |\boldsymbol{u}(\boldsymbol{r},f,L)| \leqslant z_1, \quad \forall(\boldsymbol{r},f)\in\Omega_K, \\ & |\boldsymbol{v}(\boldsymbol{r},f,L)| \leqslant z_2, \quad \forall(\boldsymbol{r},f)\in\Omega_K. \end{array} \tag{4.10}$$

Thus, it can be transformed into a standard linear programming model in matrix notation as

$$\begin{array}{ll} \min\limits_{\mathcal{Z}\in R^{ML+2}} & c^T\mathcal{Z} \\ s.t. & \mathcal{A}(\boldsymbol{r},f,L)\mathcal{Z} - b \leqslant 0,\ \forall(\boldsymbol{r},f)\in\Omega_K, \end{array} \tag{4.11}$$

where $\mathcal{A}(\boldsymbol{r},f,L)$, $\mathcal{Z}$, c and b are defined the same as those in problem (2.13).

4.3 Illustration Example

In this section, we define a simple barrier beamformer design problem to evaluate the performance of the proposed method based on the model (4.11).

A simple 1-D spatial domain is firstly defined in Figure 4.5, and the specific spatial-frequency passband Ω_p and stopband Ω_s regions are defined as

$$\Omega_p = \{(\boldsymbol{r},f)\ |\ \ |x| \leqslant 0.5m,\ 0.5kHz \leqslant f \leqslant 2.0kHz\},$$

and

$$\begin{aligned}\Omega_s = \{(\boldsymbol{r}, f) \mid\ & |x| \leqslant 0.5m,\ 2.1kHz \leqslant f \leqslant 4.0kHz\} \ldots \\ \cup\{(\boldsymbol{r}, f) \mid\ & 1.2m \leqslant |x| \leqslant 3.0m,\ 0.5kHz \leqslant f \leqslant 2.0kHz\} \ldots \\ \cup\{(\boldsymbol{r}, f) \mid\ & 1.2m \leqslant |x| \leqslant 3.0m,\ 2.1kHz \leqslant f \leqslant 4.0kHz\}.\end{aligned}$$

And a simple linear microphone array with M-elements and $2m$ in front of the X-axis is set up for capturing desired signals in passband and reducing interference and noise in stopband. The microphone elements are symmetrically arranged around the vertical line from the center of passband and spaced with $0.05m$ to avoid spatial aliasing for the frequency of interest. Then we define a square barrier between the spatial region and microphone array with $1m$ away from vertical line from the center of passband to affect the sound wave propagation, the diagram for illustration is depicted in Figure 4.5.

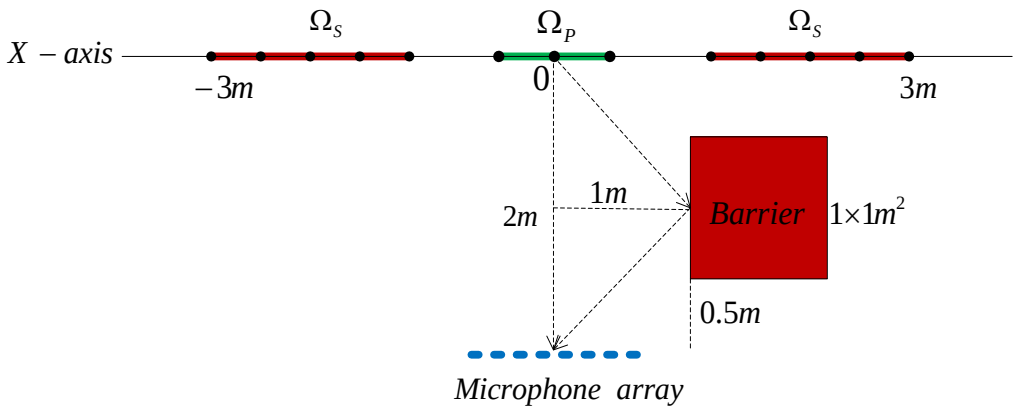


Figure 4.5: Microphone array setup under barrier condition.

For evaluation, the desired response function $G_d(\boldsymbol{r}, f, L)$ in the passband and stopband is defined the same as (2.14). Due to the introduced space-time CE/SE method for simulating barrier acoustics is time consuming, we just discrete the spatial domain into $5 \times 3 \times 5$ for stopband and passband regions, it is illustrated in above

Figure 4.5. In frequency domain, we use 60 frequency bins for discretization, and use a more density discretization scheme 120 to verify the beamforming performances.

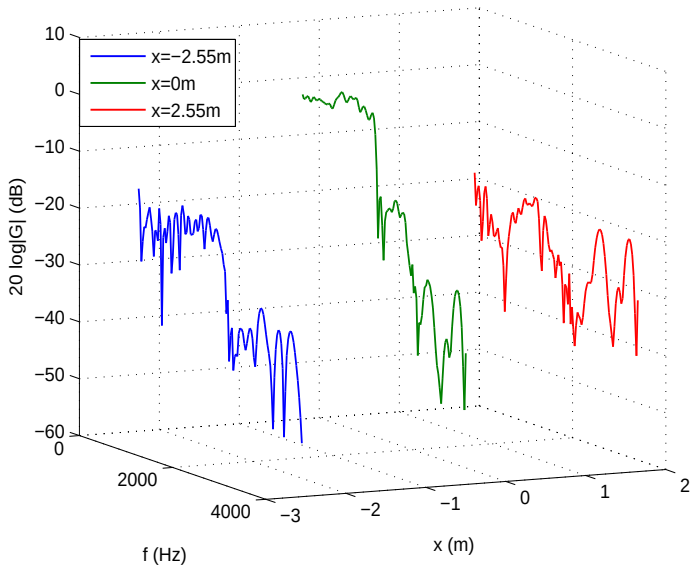
In the following simulation, we choose filter length $L = 50$ for our barrier beamformer design, and use two kinds of microphone arrays $M = 6$ and $M = 8$ to do comparison. We firstly evaluate the beamforming performance on the specific points in the passband and stopband regions, see Figure 4.6.

From the results on Figure (4.6), we can see that both the design barrier beamformers can get desired gain at $x = 0m$ in passband, and have a good suppression in the stopband at $x = \{-2.55m,\ 0m,\ 2.55m\}$. And to give an overall view, we depict the overall performance of the designed beamformers in the following Figure 4.7.

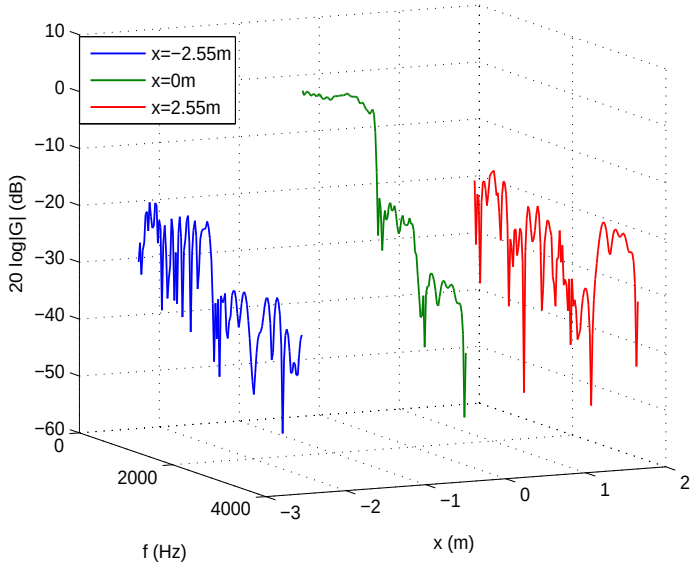
We also calculate the passband gain, passband ripple and stopband ripple values for the overall performance of the above barrier beamformers, and list the results in the following Table 4.1. The results show that both of them are effective, and the 8-elements microphone array has better performance than 6-elements microphone array.

Table 4.1: Summary of the performance of barrier beamformers

M	Passband gain	Passband ripple	Stopband ripple (dB)
6	0.9849	0.2099	-12.7563
8	1.0025	0.1874	-14.7035

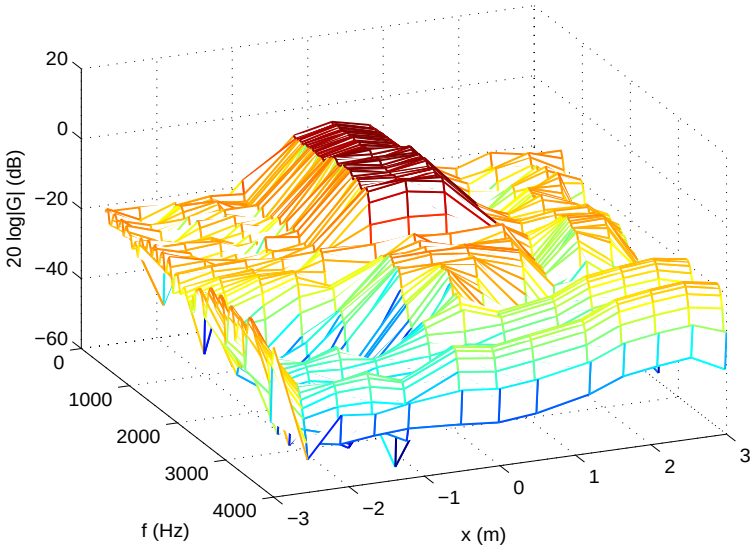


(a) $M = 6$

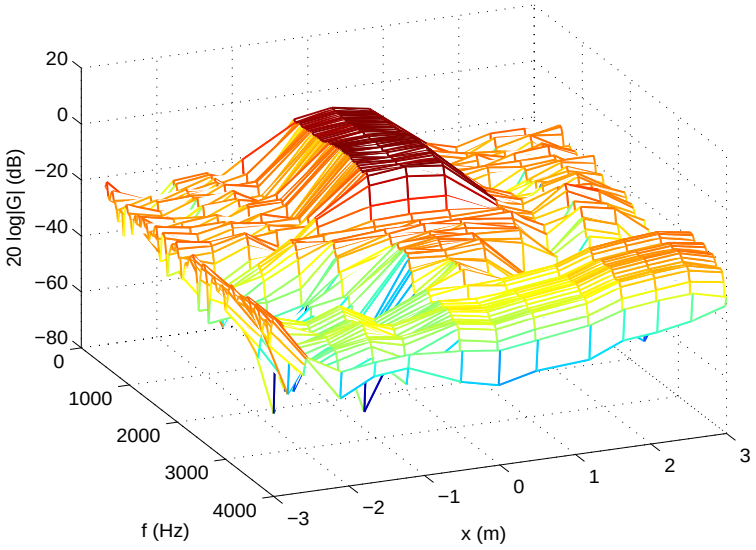


(b) $M = 8$

Figure 4.6: Beamforming performance at specific spatial-frequency domain.



(a) $M = 6$



(b) $M = 8$

Figure 4.7: Overall performance of barrier beamformer.

Chapter 5

Beamformer Design in Reverberant Environment

In general, the filter coefficients in a data independent beamformer are designed so that the beamforming response approximates a desired response independent of the array data or data statistics. In time domain beamformer design, it is often processed on receiving a signal of interest (SOI) arriving from a known location, whereas suppressing all the interference and noise signals (INTs) arriving from other locations. Thus, a linear system or least square problem for the design of beamformer can be formulated accordingly [113, 118, 76, 121].

In this chapter, we will study the time domain beamformer design problem in reverberant environment, and formulated it into a large scale linear system. We firstly introduce the Tikhonov regularization approach to improve the condition of it, then analyze the regularizing property on the convergence of this method. Moreover, to study the influence of the filter length on the beamformer design, we also study the effectiveness of the proposed beamformer design method as the filter length increases.

5.1 Time Domain Beamformer Design

5.1.1 Common Beamformer Design Problem

Consider an acoustic room with the signal of interest (SOI) generated at the source point, such as $\boldsymbol{r}_0$, and the interference signals are generated at $\boldsymbol{r}_n,\ n = 1, 2, .., N-1$. Assuming there is an M-elements microphone array fixed at $\boldsymbol{r}_m,\ m = 1, 2, .., M$, then the sensor array will capturing signals in such an indoor system. Due to the affect of room acoustics and interference, the sensor array received signals in the time domain may be expressed as

$$x_m[k] = \sum_{n=0}^{N-1} (h_{nm} * s_n)[k], \quad m = 1, 2, \ldots, M, \tag{5.1}$$

where $*$ denotes linear convolution, h_{nm} is RIR from the source location $\boldsymbol{r}_n$ to the m-th microphone element. We assume all the RIRs are estimated by the ISM based room simulator introduced in Chapter 3.

Assume there is an $L-$tap FIR filter behind each microphone element with coefficients $w_{m,l}, m = 1, 2, \ldots, M, l = 0, 1, \ldots, L-1$, then the beamformer output can be written as

$$y[k] = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_{m,l} x_m[k-l]. \tag{5.2}$$

It is noted that the convolution beamformer output can be reformulated by vector notation, and it is usually to stack the input samples and all filter coefficients into vectors as

$$y[k] = \sum_{m=1}^{M} \boldsymbol{w}_m^T \boldsymbol{x}_m[k] = \boldsymbol{w}^T \boldsymbol{x}[k], \tag{5.3}$$

where

$$\boldsymbol{w}_m = [w_{m,0}, w_{m,1}, \ldots, w_{m,L-1}]^T,\ \boldsymbol{w} = [\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_M^T]^T,$$

$\boldsymbol{x}_m[k] = [x_m[k], x_m[k-1], \ldots, x_m[k-L+1]]^T$, $\boldsymbol{x}[k] = [\boldsymbol{x}_1^T[k], \boldsymbol{x}_2^T[k], \ldots, \boldsymbol{x}_M^T[k]]^T$.

If we substitute the input signals $x_m[k]$ in (5.1), the beamformer output can be rewritten as

$$y[k] = \sum_{m=1}^{M} (w_m * h_{0m} * s_0)[k] + \sum_{n=1}^{N-1} \sum_{m=1}^{M} (w_m * h_{nm} * s_n)[k]. \tag{5.4}$$

Then our objective is to find the separation filters w_m so that the first part of the above sum constructs our desired signal, while the rest of the sum vanishes. The perfect performance on interference suppression and dereverberation need the filter coefficients satisfy the following conditions

$$\sum_{m=1}^{M} w_m * h_{0m} = \boldsymbol{g}_0, \tag{5.5}$$

$$\sum_{n=1}^{N-1} \sum_{m=1}^{M} w_m * h_{nm} = \boldsymbol{0}, \tag{5.6}$$

where $\boldsymbol{g}_0 = h_D[k - \tau_L]$ is the direct path impulse response h_D from the SOI to beamformer output point with some time delay τ_L, such as $\tau_L = L/2$, it is also known as the group delay of the sensor array.

By using the convolution matrix notation, the constant beamformer filtering response constraint towards the SOI can be expressed as

$$\boldsymbol{H}_0 \boldsymbol{w} = \boldsymbol{g}_0, \tag{5.7}$$

where $\boldsymbol{H}_0$ is the convolution matrix generated by the RIRs from the SOI to each array element corresponding to the condition (5.5), it is given by

$$\boldsymbol{H}_0 = [\boldsymbol{H}_{01} \quad \boldsymbol{H}_{02} \quad \cdots \quad \boldsymbol{H}_{0M}], \tag{5.8}$$

with

$$\boldsymbol{H}_{0i} = \begin{pmatrix} h_{0i}(0) & 0 & 0 & \cdots & 0 \\ h_{0i}(1) & h_{0i}(0) & 0 & \cdots & 0 \\ h_{0i}(2) & h_{0i}(1) & h_{0i}(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & h_{0i}(L_i) \end{pmatrix},$$

where L_i is the length of the RIR from the SOI to the i-th microphone element. And similar constraints correspond to condition (5.6) can be expressed as

$$\boldsymbol{H}_I \boldsymbol{w} = \boldsymbol{0}, \tag{5.9}$$

with $\boldsymbol{H}_I$ corresponding to the convolution matrix generated from all the interference RIRs. Writing all the constraints together and denoting $\boldsymbol{H} = [\boldsymbol{H}_0^T \ \boldsymbol{H}_I^T]^T$ and $\boldsymbol{g} = [\boldsymbol{g}_0^T \ \boldsymbol{0}^T]^T$, then the beamformer design problem can be formulated into a linear equation problem as

$$\boldsymbol{H}\boldsymbol{w} = \boldsymbol{g}. \tag{5.10}$$

Assuming all the RIRs h_{nm} have the same length $L_i = L_0$, in according to the convolution matrix construction, the constraint matrix $\boldsymbol{H}$ of above linear system (5.10) is an $2(L_0 + L - 1) \times ML$ matrix. Usually, the estimated RIRs have longer length L_0 as the room acoustics increases, whereas the filter length L of the designed beamformer is relatively short, which resulting $2(L_0 + L - 1) > ML$. Thus, the beamformer design problem (5.10) belongs to over-determined linear equation system.

5.1.2 Least Squares Method

In general, an over-determined linear equation system of $2(L_0 + L - 1)$ unknowns but ML equations has no solution if $2(L_0 + L - 1) > ML$. But it is still possible to find the optimal approximation in the least squares sense, so that the squared error is

minimized. Therefore, there is always some residual for each of the $2(L_0 + L - 1) > ML$ equations

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{w} - \boldsymbol{g}. \tag{5.11}$$

The total error can be defined as

$$e(\boldsymbol{w}) = ||\boldsymbol{r}||_2^2 = (\boldsymbol{H}\boldsymbol{w} - \boldsymbol{g})^T(\boldsymbol{H}\boldsymbol{w} - \boldsymbol{g}) = \boldsymbol{w}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{w} - 2\boldsymbol{g}^T\boldsymbol{H}\boldsymbol{w} + \boldsymbol{g}^T\boldsymbol{g}, \tag{5.12}$$

and the corresponding least squares problem can be established as

$$\min_{\boldsymbol{w}\in\mathbb{D}} \quad e(\boldsymbol{w}), \tag{5.13}$$

where $\mathbb{D} \subset \mathbb{R}^{ML}$ is some feasible convex region.

It is noted that the least squares problem (5.13) is convex, and has unique minimization value $e(\boldsymbol{w})$ on the solution set $\mathbb{S} \subset \mathbb{D}$, whereas the optimal solution $\boldsymbol{w}^*$ maybe not unique. To find an optimal $\boldsymbol{w}^*$ that minimizes $e(\boldsymbol{w})$, if we let

$$\frac{\partial e(\boldsymbol{w})}{\partial w_k} = \frac{\partial ||\boldsymbol{r}||_2^2}{\partial \boldsymbol{w}} = 2(-\boldsymbol{H}^T\boldsymbol{g} + \boldsymbol{H}^T\boldsymbol{H}\boldsymbol{w}) = 0,$$

and solve this for $\boldsymbol{w}$, this matrix equation can be solved by multiplying both sides by the inverse of $\boldsymbol{H}^T\boldsymbol{H}$, if it exists

$$\boldsymbol{w}^* = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{g} = \boldsymbol{H}^\dagger\boldsymbol{g}, \tag{5.14}$$

where $\boldsymbol{H}^\dagger = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T$ is the pseudo-inverse of the non-square matrix $\boldsymbol{H}$.

However, the matrix $\boldsymbol{H}^T\boldsymbol{H}$ is usually is ill-conditioned, and has large condition number. For instance, the condition number of matrices $\boldsymbol{H}^T\boldsymbol{H}$ formulated by the impulse responses in open air, acoustic room with reverberation time $T_{60} = 0.1s$ and $T_{60} = 0.2s$ at filter length $L = 50$ is given in Table 5.1.

Thus, the direct formulation (5.14) may lead to a solution $\boldsymbol{w}^*$ that is severely contaminated with noise in reverberant environment. Therefore, regularization is

Table 5.1: Condition number of matrices $\boldsymbol{H}^T\boldsymbol{H}$.

Open air	Acoustic room $T_{60} = 0.1s$	Acoustic room $T_{60} = 0.2s$
344	8.9×10^9	1.1×10^{11}

employed to get a more meaningful answer, such as the Tikhonov regularization [40]. The Tikhonov regularization approach improves the condition of the problem by solving the linear least squares problem

$$\min_{\boldsymbol{w}\in\mathbb{D}} \quad e_\alpha(\boldsymbol{w}) = e(\boldsymbol{w}) + \frac{\alpha}{2}||\boldsymbol{w}||_2^2 \tag{5.15}$$

instead of problem (5.13) with some $\alpha > 0$.

By adding a positive term $\frac{\alpha}{2}||\boldsymbol{w}||_2^2$, the regularized problem (5.15) becomes strictly convex, and the optimal solution $\boldsymbol{w}_\alpha$ is uniquely exist. Moreover, we can choose appropriate α such that the optimal solution $\boldsymbol{w}_\alpha$ of problem (5.15) converges to the least squares problem (5.13).

Lemma 5.1. *The optimal solution $\boldsymbol{w}_\alpha$ of problem* (5.15) *converges to the solution of least squares problem* (5.13) *as $\alpha \to 0$.*

Proof. Denote the solution set for least squares problem (5.13) as $\mathbb{S} \subset \mathbb{D}$, then for any $\boldsymbol{w}^* \in \mathbb{S}$ and $\alpha > 0$, we have

$$e(\boldsymbol{w}^*) \leqslant e(\boldsymbol{w}_\alpha) \leqslant e(\boldsymbol{w}_\alpha) + \frac{\alpha}{2}||\boldsymbol{w}_\alpha||_2^2 \leqslant e(\boldsymbol{w}^*) + \frac{\alpha}{2}||\boldsymbol{w}^*||_2^2. \tag{5.16}$$

Since $e(\boldsymbol{w})$ is a continuous function on $\mathbb{R}$, thus any cluster point of $\{\boldsymbol{w}_\alpha\}$ is a solution of least squares problem (5.13) as $\alpha \to 0$, which means

$$\lim_{\alpha\to 0} \operatorname{dist}(\boldsymbol{w}_\alpha, \mathbb{S}) = 0,$$

where $\operatorname{dist}(\boldsymbol{w}, \mathbb{D}) = \inf_{\boldsymbol{z}\in\mathbb{D}} ||\boldsymbol{w} - \boldsymbol{z}||_2$ denotes the distance function.

Let $\boldsymbol{w}_\alpha^*$ be the limit of a subsequence $\{\boldsymbol{w}_{\alpha_k}\}$ as $\alpha_k \to 0$, assume there exists a $\bar{\boldsymbol{w}}^* \in \mathbb{S}$ such that $||\bar{\boldsymbol{w}}^*||_2 \leqslant ||\boldsymbol{w}_\alpha^*||_2$ and $e(\bar{\boldsymbol{w}}^*) = e(\boldsymbol{w}_\alpha^*)$, we have

$$e(\boldsymbol{w}_{\alpha_k}) + \frac{\alpha_k}{2}||\boldsymbol{w}_{\alpha_k}||_2^2 \leqslant e(\bar{\boldsymbol{w}}^*) + \frac{\alpha_k}{2}||\bar{\boldsymbol{w}}^*||_2^2 \leqslant e(\boldsymbol{w}_\alpha^*) + \frac{\alpha_k}{2}||\boldsymbol{w}_\alpha^*||_2^2. \tag{5.17}$$

Combining the above equations (5.16) and (5.17), we have

$$0 \leqslant \frac{2}{\alpha_k}(e(\boldsymbol{w}_{\alpha_k}) - e(\bar{\boldsymbol{w}}^*)) \leqslant ||\bar{\boldsymbol{w}}^*||_2^2 - ||\boldsymbol{w}_{\alpha_k}||_2^2 \leqslant ||\boldsymbol{w}_\alpha^*||_2^2 - ||\boldsymbol{w}_{\alpha_k}||_2^2 \ \to 0 \quad \text{as} \quad \alpha_k \to 0,$$

which implies that $||\boldsymbol{w}_{\alpha_k}||_2$ converges to $||\bar{\boldsymbol{w}}^*||_2$ as $\alpha_k \to 0$. Thus, we have $||\bar{\boldsymbol{w}}^*||_2 = ||\boldsymbol{w}_\alpha^*||_2$, moreover, $\boldsymbol{w}_\alpha^*$ is the solution of least squares problem (5.13). □

Notice that the solution $\boldsymbol{w}_\alpha$ satisfies the equation

$$(\boldsymbol{H}^T\boldsymbol{H} + \alpha\boldsymbol{I})\boldsymbol{w}_\alpha = \boldsymbol{H}^T\boldsymbol{g}, \tag{5.18}$$

where $\boldsymbol{I}$ is the identity matrix, and a direct solution of $\boldsymbol{w}_\alpha$ can be obtained by

$$\boldsymbol{w}_\alpha = (\boldsymbol{H}^T\boldsymbol{H} + \alpha\boldsymbol{I})^{-1}\boldsymbol{H}^T\boldsymbol{g}. \tag{5.19}$$

Equivalently, the solution of $\boldsymbol{w}_\alpha$ can also be computed as the solution to a damped linear least squares problem

$$\min_{\boldsymbol{w}\in\mathbb{D}} \quad \left\| \begin{pmatrix} \boldsymbol{H} \\ \sqrt{\frac{\alpha}{2}}\boldsymbol{I} \end{pmatrix} \boldsymbol{w} - \begin{pmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{pmatrix} \right\|_2^2. \tag{5.20}$$

The choice of an appropriate regularization parameter α is usually crucial, there are also many methods have been proposed to select it, such as Morozov's discrepancy principle [87], the Gfrerer/Raus-method [38], the quasi-optimality criterion [4], the generalized cross-validation [41] and the L-curve criterion [49]. Moreover, the above damped linear least squares problem (5.20) also forms the basis of algorithm LSQR [96], which is a popular iterative method for the solution of large linear systems of equations and least squares problems.

On the other hand, the matrix $\boldsymbol{H}^T\boldsymbol{H} + \alpha\boldsymbol{I}$ in equation (5.18) is symmetric, then the Cholesky factorization can be applied to solve it efficiently [42]. Assume $\boldsymbol{H}^T\boldsymbol{H} + \alpha\boldsymbol{I}$ satisfies the requirement for Cholesky decomposition

$$\boldsymbol{H}^T\boldsymbol{H} + \alpha\boldsymbol{I} = \boldsymbol{R}^T\boldsymbol{R}, \tag{5.21}$$

where $\boldsymbol{R}$ is an upper triangular matrix, we can rewrite the linear system as

$$\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{w}_\alpha = \boldsymbol{H}^T\boldsymbol{g}.$$

By letting $\boldsymbol{z} = \boldsymbol{R}\boldsymbol{w}_\alpha$, we have

$$\boldsymbol{R}^T\boldsymbol{z} = \boldsymbol{H}^T\boldsymbol{g}, \tag{5.22}$$

and

$$\boldsymbol{R}\boldsymbol{w}_\alpha = \boldsymbol{z}. \tag{5.23}$$

The equations (5.22) and (5.23) are solved using backward-substitution and require $O((ML)^2)$ operations each for the solution.

5.2 Effectiveness of Beamformer with Increasing Filter Length

In general, the effectiveness of beamformer is increased as the filter length increases. However, when the filter length increases, the number of filter coefficients is increased, and the beamformer design problem (5.10) also changes, it is not sure the beamformed response $\boldsymbol{H}\boldsymbol{w}$ makes a better approximation of the desired response $\boldsymbol{g}$. In view of this, we study the influence of filter length on the design of beamformer in the following.

On the L-tap beamformer design, the size of $2(L_0 + L - 1) \times ML$ combinational convolution matrix (denoted as $\boldsymbol{H}_L$) is determined. Suppose the optimal group

delay τ_L is known, then the desired vector $\boldsymbol{g}_L$ is determined, thus, the optimal filter coefficients $\boldsymbol{w}_L$ can be denoted as

$$\boldsymbol{w}_L = \arg \min_{\boldsymbol{w}\in\mathbb{D}, 0\leqslant\tau_L\leqslant L-1} \quad ||\boldsymbol{H}_L\boldsymbol{w} - \boldsymbol{g}_L||_2^2. \tag{5.24}$$

And without loss of generality, we can denote the optimal error of the L-tap beamformer as

$$e(L, \tau_L) = ||\boldsymbol{H}_L\boldsymbol{w}_L - \boldsymbol{g}_L||_2^2. \tag{5.25}$$

Notice that the difference between the vector $\boldsymbol{g}_{L+1}$ and $\boldsymbol{g}_L$ is just the time-shifting of a desired response h_D, whereas other entries are all equal to 0, thus we can use the least squares residual $e(L, \tau_L)$ to measure the effectiveness of the beamformer design. According to this principle, we have a result on the relationship between effectiveness of the designed $L+1$-tap beamformer and L-tap beamformer under condition as follows.

Theorem 5.1. *Under the optimal criteria of beamformer design problem* (5.10)*, the* $L+1$*-tap beamformer will have better performance than* L*-tap beamformer, that is*

$$e(L+1, \tau_{L+1}) \leqslant e(L, \tau_L). \tag{5.26}$$

Proof. In order to prove the relationship (5.26), we consider the structure of problem (5.10) in different cases:

Case1: $M = 1$ *and* $N = 1$.

In this case, there is only one SOI but no INT, and we have only one microphone element. Then the convolution matrix $\boldsymbol{H}_{L+1}$ will be one more row and one more column than matrix $\boldsymbol{H}_L$, it can be rewritten as

$$\boldsymbol{H}_{L+1} = \begin{pmatrix} \boldsymbol{H}_L & \boldsymbol{b} \\ \boldsymbol{0} & h(L_0) \end{pmatrix}, \tag{5.27}$$

where $h(L_0)$ is the last element of RIR $\boldsymbol{h}$, and $\boldsymbol{b} \in \mathbb{R}^{2(L_0+L-1)}$.

Assume the optimal residual of the L-tap designed beamformer is $e(L, \tau_L)$ at the group delay τ_L for the constant vector $\boldsymbol{g}_L$ and filter coefficients $\boldsymbol{w}_L$, then the optimal residual of the $L+1$-tap designed beamformer

$$\begin{aligned} e(L+1, \tau_{L+1}) &= ||\boldsymbol{H}_{L+1}\boldsymbol{w}_{L+1} - \boldsymbol{g}_{L+1}||_2^2 \\ &\leqslant \left\| \begin{pmatrix} \boldsymbol{H}_L & \boldsymbol{b} \\ \boldsymbol{0} & h(L_0) \end{pmatrix} \begin{pmatrix} \boldsymbol{w}_L \\ 0 \end{pmatrix} - \begin{pmatrix} \boldsymbol{g}_L \\ 0 \end{pmatrix} \right\|_2^2 \\ &= ||\boldsymbol{H}_L\boldsymbol{w}_L - \boldsymbol{g}_L||_2^2 \\ &= e(L, \tau_L), \end{aligned} \tag{5.28}$$

which means the equation (5.26) holds.

Case2: $M > 1$ and $N > 1$.

In this case, there is one SOI and $N-1$ INTs, and we have M microphone elements. Notice that the convolution matrix $\boldsymbol{H}_L = [\boldsymbol{H}_{0,L}^T \ \boldsymbol{H}_{I,L}^T]^T$ is consists of SOI and INT two parts, and each of them are consist of M sub-convolution matrix like (5.7), then the convolution matrix $\boldsymbol{H}_{L+1}$ will be 2 more rows and M more columns than matrix $\boldsymbol{H}_L$. And the constant vector $\boldsymbol{g}_L$ also consists of $\boldsymbol{g}_{0,L}$ and $\boldsymbol{0}$ two parts, in according to achieve the desired response on SOI and suppression on INTs. Thus, the $L+1$-tap beamforming matrix $\boldsymbol{H}_{L+1}$ and constant vector $\boldsymbol{g}_{L+1}$ will also have two parts. Moreover, the optimal solution of a linear system will changed via column permutations of constraint matrix, whereas the optimal residual of linear system will not changed, and without loss of generality, the matrix $\boldsymbol{H}_{L+1}$ can be rewritten as

$$\boldsymbol{H}_{L+1} = \begin{pmatrix} \boldsymbol{H}_{0,L} & & \boldsymbol{b}_0 & & \\ \boldsymbol{0} & h_{01}(L_0) & h_{02}(L_0) & \dots & h_{0M}(L_0) \\ \boldsymbol{H}_{I,L} & & \boldsymbol{b}_I & & \\ \boldsymbol{0} & \sum_{n=1}^{N-1} h_{n1}(L_0) & \sum_{n=1}^{N-1} h_{n2}(L_0) & \dots & \sum_{n=1}^{N-1} h_{nM}(L_0) \end{pmatrix}, \tag{5.29}$$

where $\boldsymbol{b}_0$ and $\boldsymbol{b}_I$ denote the added columns corresponding to the RIRs from SOI to microphone array and RIRs from INTs to microphone array. Similarly, the constant vec-

tor under the optimal group delay τ_{L+1} can also be expressed as $\boldsymbol{g}_{L+1} = [\boldsymbol{g}_{0,L+1}^T \ \mathbf{0}^T]^T$.

It is noted that the optimal residual of linear system will also not changed via row permutations of constraint matrix, thus the matrix $\boldsymbol{H}_{L+1}$ can be expressed by $\boldsymbol{H}_L$ as

$$\boldsymbol{H}_{L+1} = \begin{pmatrix} \boldsymbol{H}_L & \boldsymbol{b} \\ \mathbf{0} & \boldsymbol{h}(L_0) \end{pmatrix}, \tag{5.30}$$

where $\boldsymbol{h}(L_0) = [\boldsymbol{h}_0^T(L_0) \ \boldsymbol{h}_I^T(L_0)]^T$, and $\boldsymbol{h}_0(L_0)$ and $\boldsymbol{h}_I(L_0)$ denote the added SOI and INT constraint rows corresponding to (5.29). Then optimal residual of the $L+1$-tap designed beamformer

$$\begin{aligned} e(L+1, \tau_{L+1}) &= ||\boldsymbol{H}_{L+1}\boldsymbol{w}_{L+1} - \boldsymbol{g}_{L+1}||_2^2 \\ &\leqslant \left\| \begin{pmatrix} \boldsymbol{H}_L & \boldsymbol{b} \\ \mathbf{0} & \boldsymbol{h}(L_0) \end{pmatrix} \begin{pmatrix} \boldsymbol{w}_L \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{g}_L \\ \mathbf{0} \end{pmatrix} \right\|_2^2 \\ &= ||\boldsymbol{H}_L\boldsymbol{w}_L - \boldsymbol{g}_L||_2^2 \\ &= e(L, \tau_L), \end{aligned} \tag{5.31}$$

which means the equation (5.26) holds. This completes the proof. □

5.3 Illustration Examples

5.3.1 Measurement Indicators

We first introduce some indicators to measure the performance of proposed speech enhancement method. Due to considered problem is defined in reverberant environment, we need to define indicator to evaluate the performance on dereverberation. The most direct objective measure based on the RIRs measure is the direct-to-reverberant ratio (DRR) [88] and is defined as

$$DRR = 10\log_{10}\left(\frac{\sum_{k=0}^{k_d} h^2[k]}{\sum_{k=k_d+1}^{+\infty} h^2[k]} \right) \text{dB}, \tag{5.32}$$

in which samples of the channel impulse response $h[k]$, indexed from zero up to k_d are assumed to represent only the direct-path propagation, while samples of the channel impulse response with indices greater than k_d represent only the reverberation due to reflected paths.

Based on testing signals, there are also many objective measurement methods have been derived in the literature for evaluating the performance of speech enhancement skills [80]. The mostly used methods include the SNR_{seg} measure [48], perceptual evaluation of speech quality (PESQ) measure [101], KurtR measure [112, 119], NRR and LLR measure [98], and it is noted that all these evaluation measures are proved to be consistent in [119]. Thus during our experiments, we just introduce the SNR_{seg} measure and PESQ measure. In our simulation, we define the generalized signal-to-interference and-reverberation ratio (SIRR) for measuring the performance of speech quality degraded by reverberation, interference and background noise. And the segmental SIRR measure defined like the SNR_{seg} by Loizou [80] as

$$SIRR_{seg} = \frac{1}{Q}\sum_{q=0}^{Q-1} 10\log_{10}\frac{||\boldsymbol{S}(q)||^2}{||\boldsymbol{S}(q)-\hat{\boldsymbol{S}}(q)||^2}\ \text{dB}, \tag{5.33}$$

where $\boldsymbol{S}(q)$ and $\hat{\boldsymbol{S}}(q)$ represent the direct captured speech signal when the source signal is active alone without reverberation and the enhanced speech in q-th time frame, and the calculation is thresholded in interval [−20dB, 35dB] to discard non-speech frames.

On the PESQ score measurement, it is recommended by ITU-T for speech quality assessment of $3.2kHz$ handset telephony and narrow-band speech codecs [93]. The final PESQ score is obtained by a linear combination of the average disturbance value D_{ind} and the average asymmetrical disturbance values as follows A_{ind} [94]:

$$PESQ = 4.5 - 0.1D_{ind} - 0.0309A_{ind}. \tag{5.34}$$

The range of the PESQ score is 0.5 to 4.5, although for most cases the output range will be a MOS-like score (mean opinion score), i.e., a score between 1.0 and 4.5. And in our following experimental simulations, the `comp_snr` and `pesq` coding functions from [80] are used to calculate the segmental SIRR and PESQ score.

The segmental SIRR and PESQ scores measurements are all pointed to the performance of speech quality measurement. Moreover, specific indicators for measuring the performance on dereverberation and interference suppression are also introduced. We want to introduce the normalized scheme to measure the interference suppression based on speech signal developed by Yiu *et al.* [117], that is defined as

$$Supp_{INT} = \frac{\int_{-\pi}^{\pi} P_{Y_I}(w)dw}{C_d \int_{-\pi}^{\pi} P_{S_I}(w)dw}, \tag{5.35}$$

where the constant C_d is defined as

$$C_d = \frac{\int_{-\pi}^{\pi} P_{Y_S}(w)dw}{\int_{-\pi}^{\pi} P_{S_S}(w)dw},$$

and $P_{S_S}(w)$ and $P_{S_I}(w)$ denote the spectral power estimate of the unfiltered signals when the source and interference active alone, respectively, $P_{Y_S}(w)$ and $P_{Y_I}(w)$ denote the spectral power estimate of the filtered signals when the source and interference active alone, respectively. And we wish to define the similar indicator for dereverberation as normalized reverberation suppression

$$Supp_{REV} = \frac{\int_{-\pi}^{\pi} P_{Y_R}(w)dw}{C_d \int_{-\pi}^{\pi} P_{S_R}(w)dw}, \tag{5.36}$$

where $P_{S_R}(w)$ and $P_{Y_R}(w)$ denote the spectral powers estimate of the unfiltered and filtered reverberant signals active alone, respectively.

5.3.2 Acoustic Room Configuration

A simple rectangular $12m \times 6m \times 3m$ room is defined for acoustic room modeling, all the walls, floor and ceiling are assumed to be with uniform absorption coefficients, which resulting from the desired reverberation time T_{60}. We define a 9-elements sensor array in the $z = 1m$ plane with L-tap FIR filter behind each microphone, specifically, they are placed at

$$\mathsf{Mic} = \{(5.5,3),\ (5.625,3),\ (5.75,3),\ (5.875,3),\ (6,3),\ (6.125,3),\ (6.25,3),\ (6.375,3),\ (6.5,3)\}$$

at plane $z = 1$ in meter, and with the center element as the reference receiving point, also the beamformer output point. There are also one SOI placed at $(-1, -1, 1)$, two INTs placed at $(2, -1, 1)$ and $(2, 1, 1)$ in meter with the center of the room as the origin, an illustration of the room setup is depicted in Figure 5.1. All the RIRs from the considering points to sensor array are estimated by the image source method, and the fast-ISM room simulator developed by E.A. Lehmann and A.M. Johansson [68] will be introduced for modeling room acoustics at $T_{60} = 0.1s$ and $T_{60} = 0.2s$.

In such a room setup, the inputting SOI signal is a broadband male speech ("*Dots of light betrayed the black cat*"), the INT signals are a female speech ("*she had your dark suit in greasy wash water all year*") and an excerpts of Chinese broadcast ("*zui da de shi hou ying gai jiu shi*"), all of them have $8000Hz$ sampling frequency, see the plotting in Figure 5.1.

Assume the output point of microphone array is set up at the center of it, thus before processing, the initial DRR value of the RIR from SOI to the center microphone element, the initial scores of SIRR and PESQ of the unfiltered signals under different room acoustics can be measured for comparison, and they are listed in Table 5.2. The unfiltered signals are plotted in Figure 5.2.

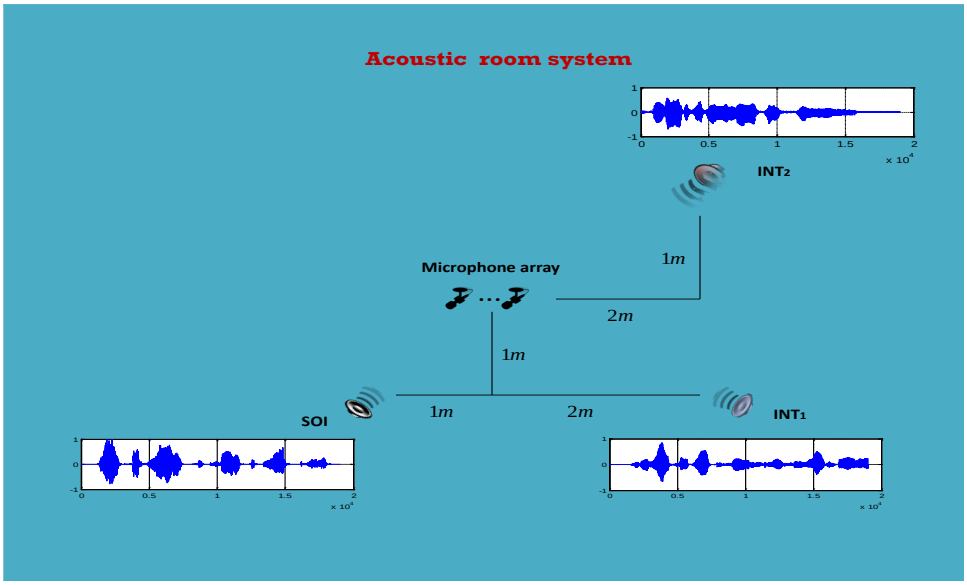
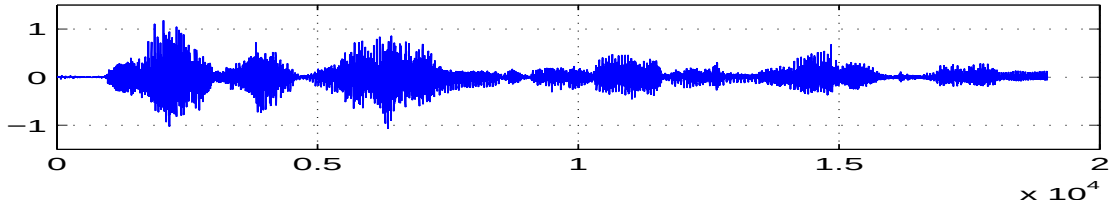


Figure 5.1: The $z = 1m$ plane of the defined acoustic room.

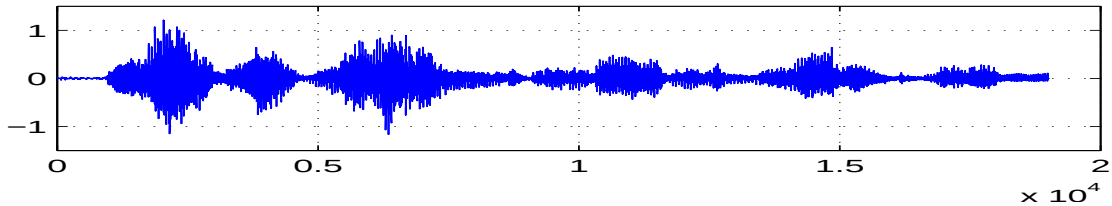
Table 5.2: The initial values measurement.

T_{60}	DRR	$SIRR_{seg}$	PESQ
0.1	9.7887	-2.9423	1.9526

T_{60}	DRR	$SIRR_{seg}$	PESQ
0.2	6.5039	-3.5692	1.9846



(a) at $T_{60} = 0.1s$



(b) at $T_{60} = 0.2s$

Figure 5.2: The unfiltered signals at different room acoustics.

5.3.3 Group Delay Learning

Notice that the group delay τ_L is important on the design of beamformer, thus we first learn the influence of it. As an example, we choose the filter length $L = 50$, and introduce the DRR, reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT_1}$ and $Supp_{INT_2}$, segmental SIRR and PESQ scores six indicators to do comparison. We learn the group delay τ_L from 1 to $L-1$ for illustration in the following, the overall performance of group delay learning is depicted in Figure 5.3.

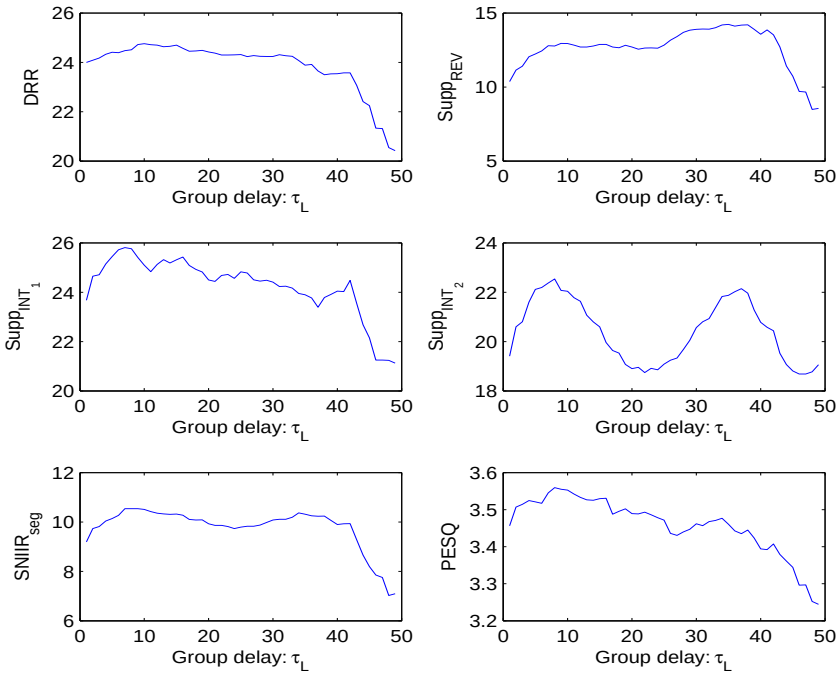
From the group delay τ_L learning results, we can see that the group delay selection do have influence on the beamformer design, too small or large group delay selection is not acceptable. And for different room acoustics, the similar results show that the better group delay should be chosen small or large than $L/2$, whereas the measurement indicators are consistent on the whole.

5.3.4 Filter Length Learning

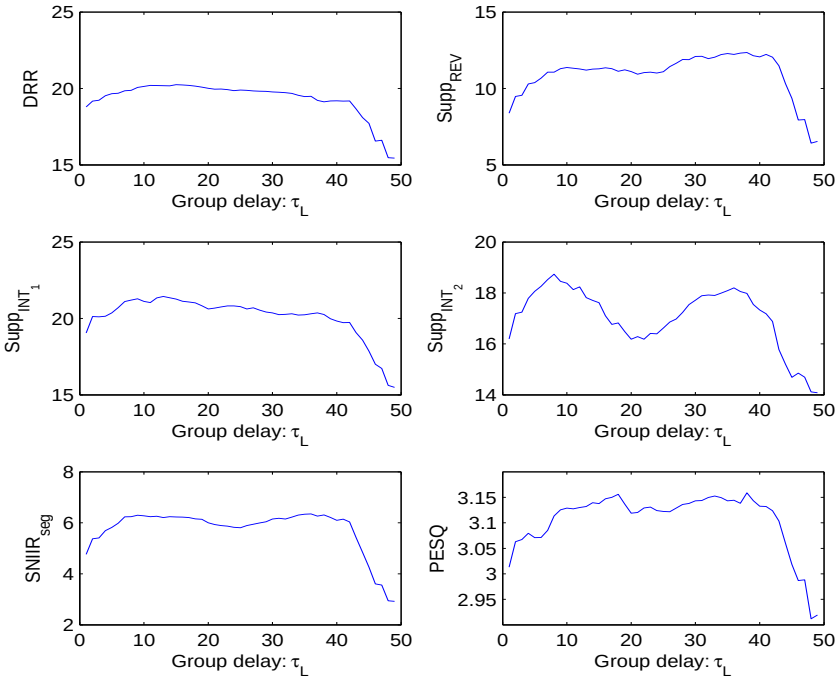
From the above group delay learning results, it is difficult to define a fixed group delay selection to achieve optimal objectives on DRR, reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT_1}$ and $Supp_{INT_2}$, segmental SIRR and PESQ scores. But for the best performance on dereverberation, we can choose the group delay τ_L for optimal $Supp_{REV}$ to learn the influence of filter length L on beamformer design in the following.

To study the filter length influence, we learn the filter length L from 2 to 100, and the overall results on filter length learning are depicted in Figure 5.4.

From the filter length learning results, we can see that the designed beamformers have effect on speech enhancement in reverberant environment when the filter length $L > 20$. And all the measurement indicators are increased as the filter length increasing, which also verifies the result on effectiveness of beamformer design with

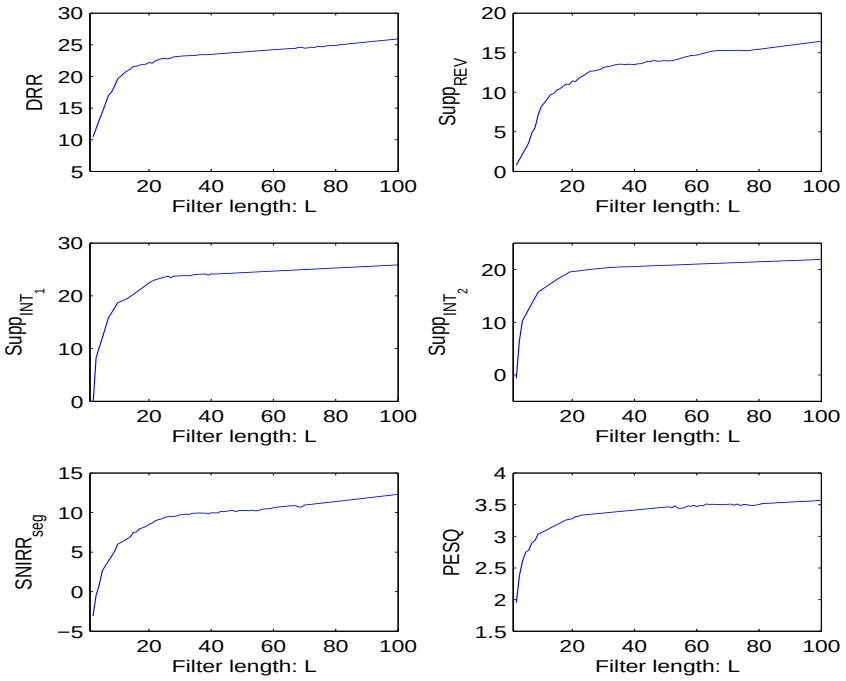


(a) at $T_{60} = 0.1s$

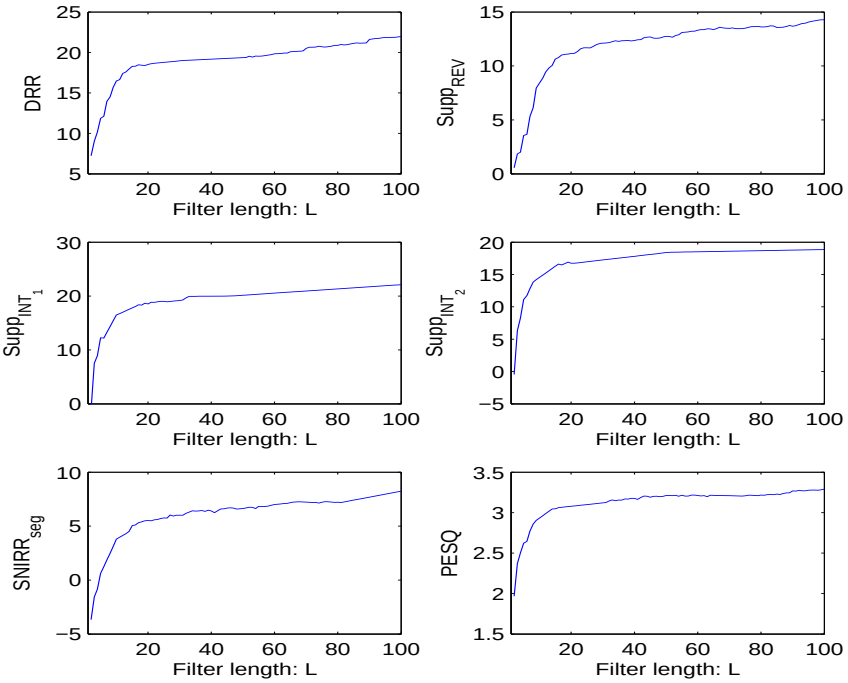


(b) at $T_{60} = 0.2s$

Figure 5.3: Learning for group delay at filter length $L = 50$.



(a) at $T_{60} = 0.1s$



(b) at $T_{60} = 0.2s$

Figure 5.4: Performance with respect to filter length L.

respect to filter length given in Theorem 5.1.

5.3.5 Overall Performance Illustration

To take an illustration on speech enhancement by the beamforming approach, in the following, we choose filter length $L = \{20,\ 30,\ 40,\ 50,\ 60\}$ to design the beamformers at the best group delay selection for dereverberation. We use the DRR, reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT_1}$ and $Supp_{INT_2}$, segmental SIRR and PESQ scores six indicators to evaluate the effect of speech enhancement, and list the results in the following Tables 5.3 and 5.4.

Table 5.3: Overall performance of designed beamformers at $T_{60} = 0.1s$.

L	DRR	$Supp_{REV}$	$Supp_{INT_1}$	$Supp_{INT_1}$	$SIRR_{seg}$	PESQ
20	22.2150	11.3649	21.3446	20.0572	8.4348	3.3081
30	23.1677	12.9396	22.9583	20.6631	9.5881	3.3079
40	23.7758	13.0727	23.4563	20.7437	9.7491	3.4350
50	24.1548	14.1747	23.6811	21.0391	10.2399	3.4672
60	24.2827	14.6807	24.5160	21.3670	10.5695	3.4793

Table 5.4: Overall performance of designed beamformers at $T_{60} = 0.2s$.

L	DRR	$Supp_{REV}$	$Supp_{INT}$	$Supp_{INT_1}$	$SIRR_{seg}$	PESQ
20	18.2787	11.0176	18.3389	16.5548	5.3293	3.0104
30	18.8396	11.8587	19.1813	17.7930	5.7156	3.1016
40	19.1368	12.3960	19.5349	17.8595	6.2091	3.1386
50	19.6442	12.8697	20.0341	18.1077	6.5640	3.1612
60	19.7076	13.3990	20.8589	18.3691	6.8774	3.2053

From the overall evaluation results on beamforming performance, we can see that all the DRR, segmental SIRR and PESQ scores have grown substantially after the beamforming processing, and all the reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT_1}$ and $Supp_{INT_2}$ indicators obtain a very good effect. Moreover, the beamformer designed at longer filter length can achieve better performance on all the DRR, reverberation suppression, interferences suppression, segmental SIRR and

PESQ scores. We can also find that all the indicators for the beamformers design at $T_{60} = 0.2s$ are small than the beamformers design at $T_{60} = 0.1s$, it demonstrates that the influence of room acoustic on speech quality enhancement.

In the following, we also depict the overall performance on the designed beamformers at filter length $L = 50$ in Figures 5.5 and 5.7 at different room acoustics for illustration. We can see that all the RIRs from the SOI to center microphone element (CMic) can be effectively beamformed to the desired responses, which make the desired SOI signals successfully formulated and reverberation suppressed. And the interference RIRs can be effectively suppressed, which make the interference signals successfully suppressed. The illustration of reverberant and interference signals suppression can also be found in Figures 5.6 and 5.8.

Moreover, the overall performance of the enhanced speech signals by using the designed beamformer at filter length $L = 50$ are depicted in Figure 5.9. On the beamforming results, we find that the beamformed signals are very close to desired SOI signals in both two room acoustics, it also demonstrates the effect of the proposed beamformer design method in reverberant environment.

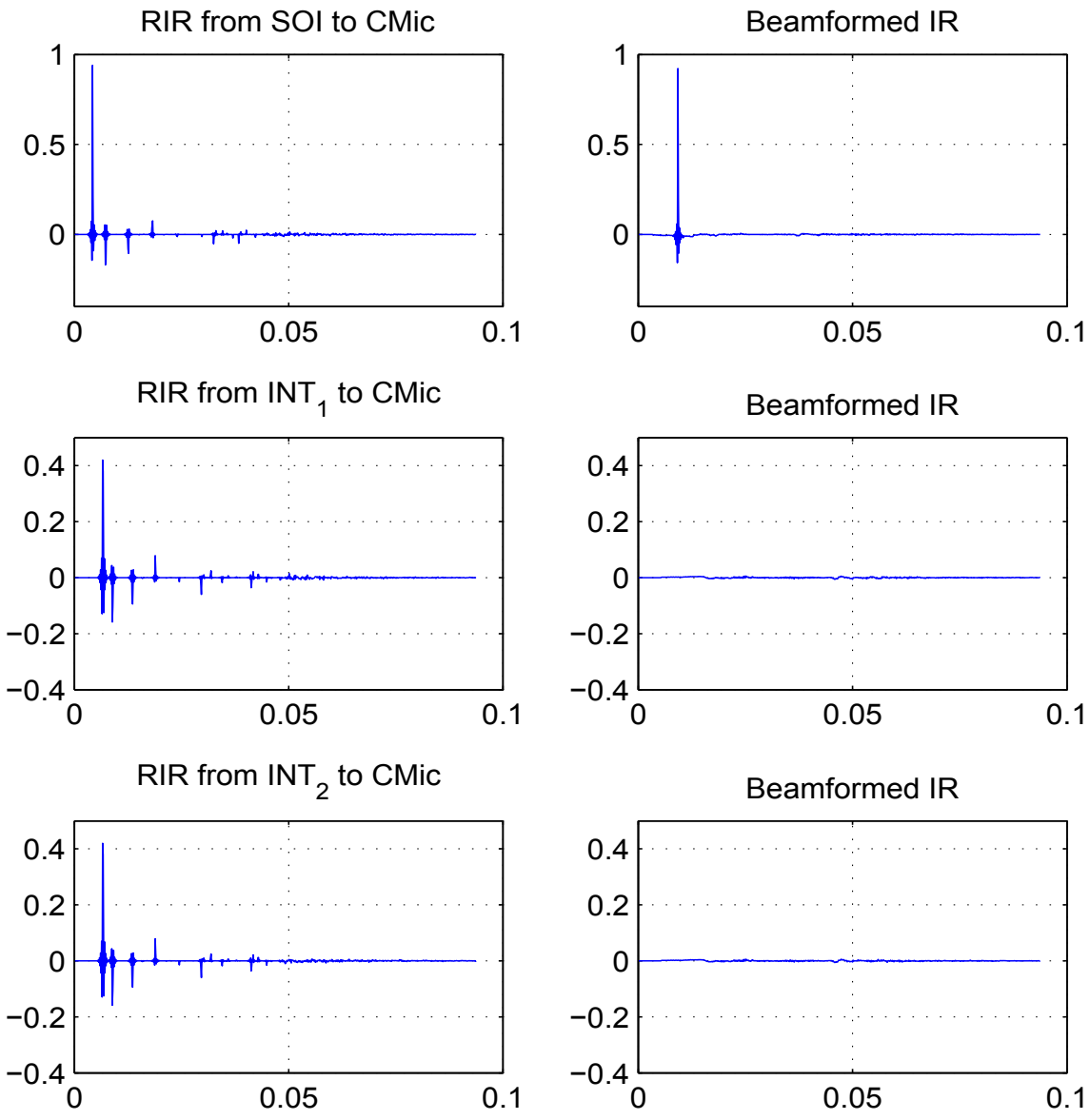


Figure 5.5: RIRs for SOI and INT before and after beamformer filtering at $T_{60} = 0.1s$.

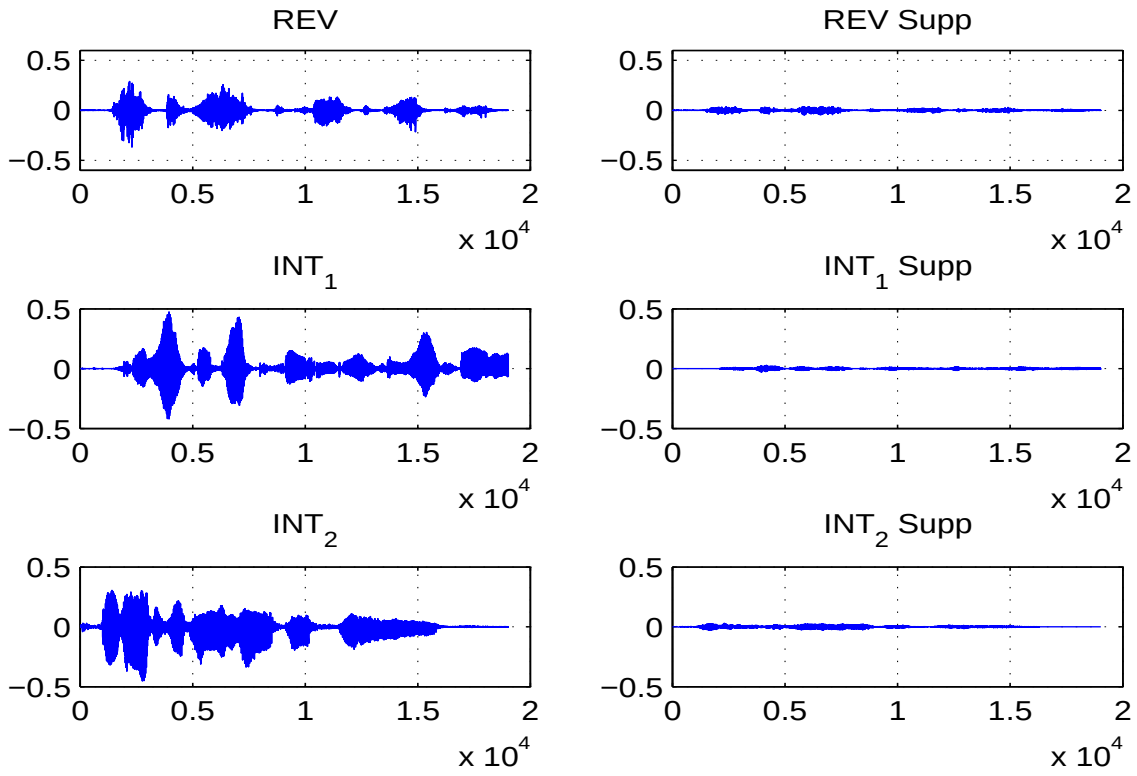


Figure 5.6: Interferences and reverberation before and after filtering at $T_{60} = 0.1s$.

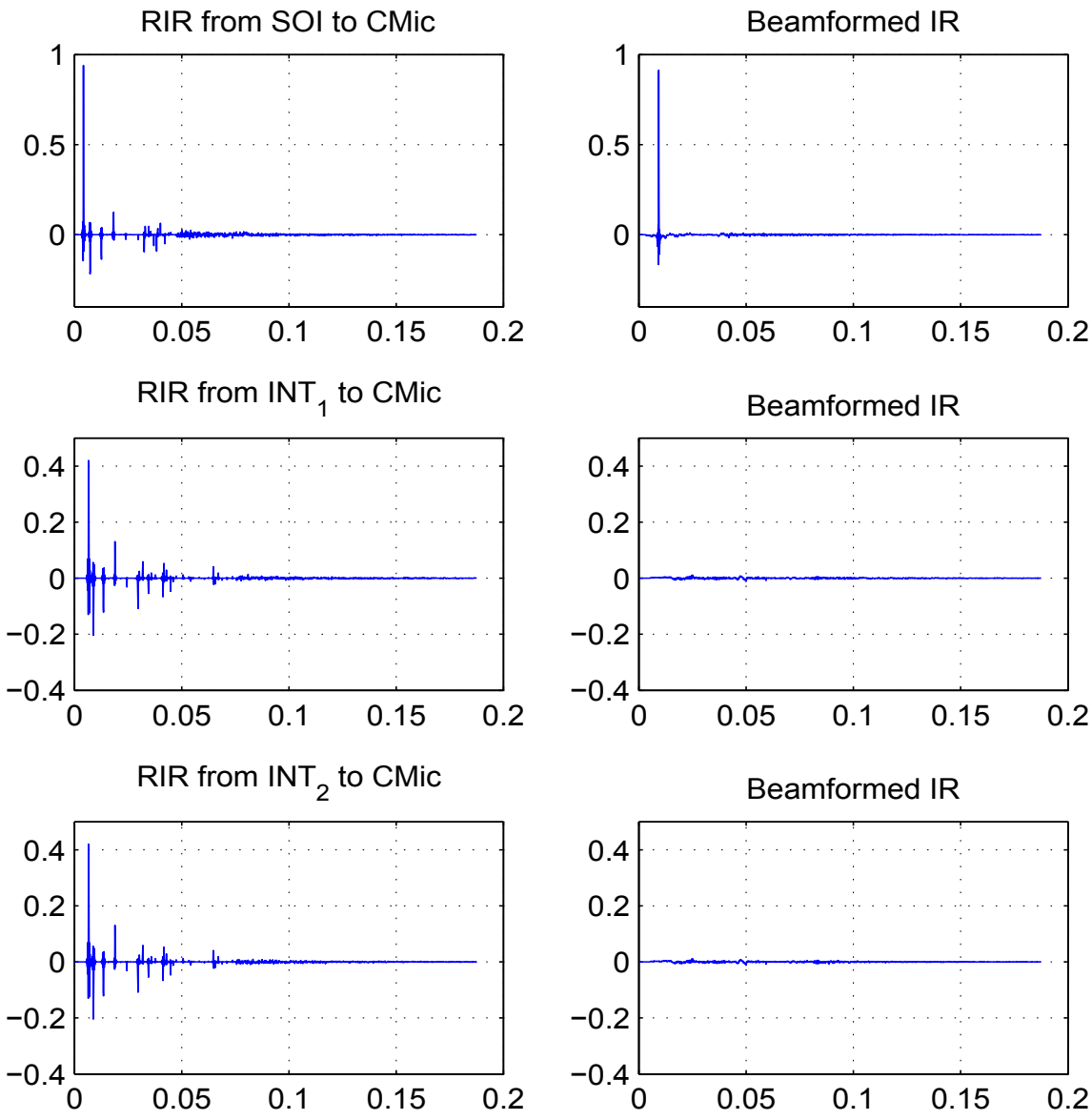


Figure 5.7: RIRs for SOI and INT before and after beamformer filtering at $T_{60} = 0.2s$.

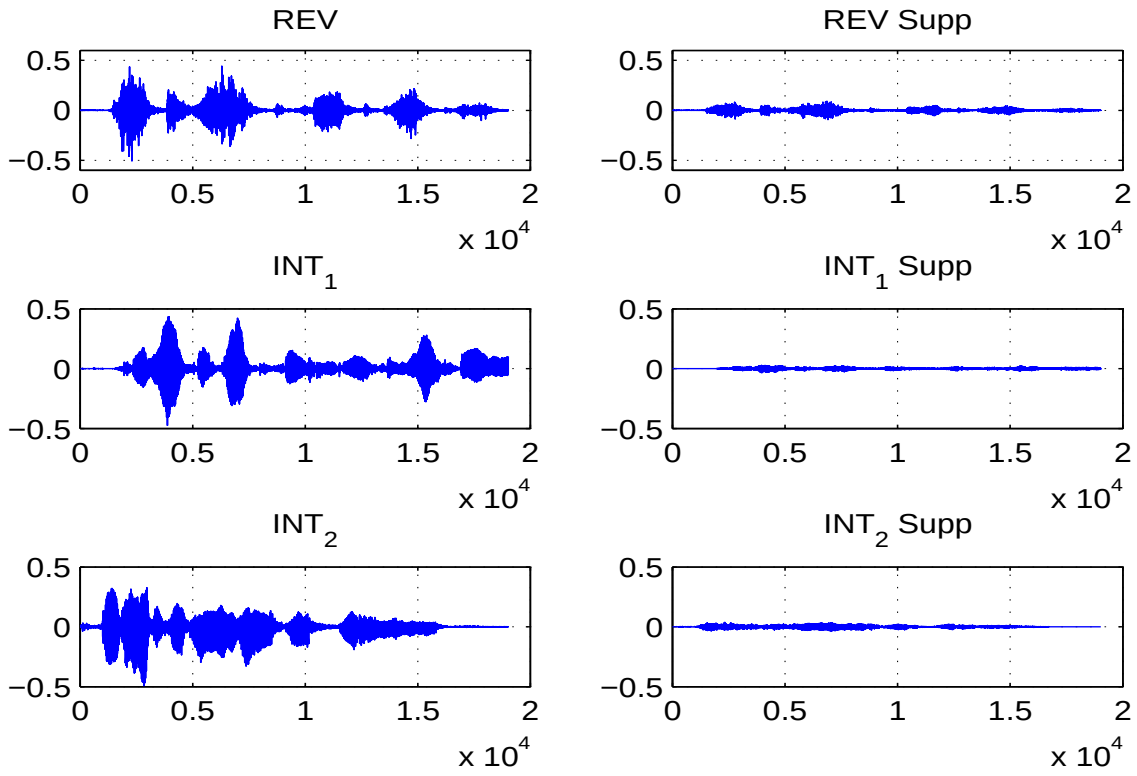
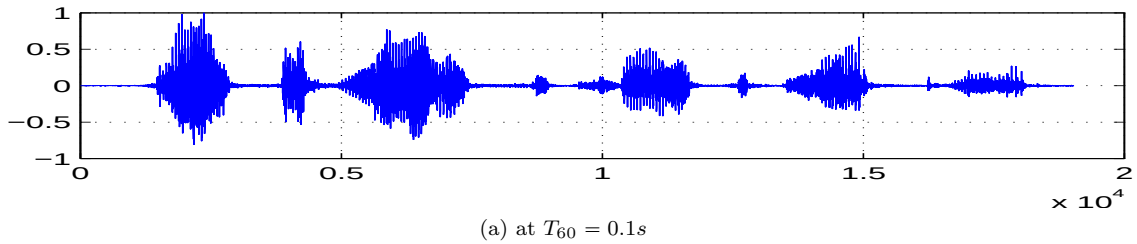
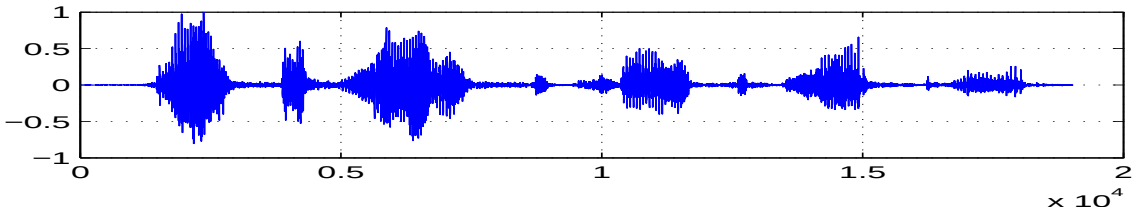


Figure 5.8: Interferences and reverberation before and after filtering at $T_{60} = 0.2s$.



(a) at $T_{60} = 0.1s$



(b) at $T_{60} = 0.2s$

Figure 5.9: The beamformed signals signals under different acoustic room.

Chapter 6

Indoor LCMV Beamformer Design

In this chapter, we study the design of linearly constrained minimum variance (LCMV) beamformer in acoustic room. It is noted that in reverberant environment, most of the conventional beamforming approaches suffer from poor performance in terms of sound reverberation caused by the room acoustics. Thus we want to enhance the speech quality by using the designed LCMV beamformer with the help of RIRs, that can be estimated by some simple but efficient room simulator.

We firstly enhance the speech quality by using the indoor LCMV beamformer designed with the help of RIRs. Then we do numerical experiments to learn the influences of filter length and group delay selection on the design of beamformer. Moreover, we study several post-filtering techniques combining with LCMV beamformer to further improve the quality of the speech. They include the Wiener filter (WF), log-spectral amplitude (LSA) estimator, and the recently proposed modified sigmoid (MSIG) filter.

6.1 Indoor LCMV Beamformer

6.1.1 Problem Formulation

Consider an acoustic room with the signal of interest (SOI) is generated at the source point, such as $\boldsymbol{r}_0$, the interferer signals (INTs) are generated at $\boldsymbol{r}_n,\ n = 1, 2, .., N-1$, and background noise (NOI) is generated at undetermined place, assuming there is an M-elements microphone array fixed at $\boldsymbol{r}_m,\ m = 1, 2, .., M$. Then the sensor array will capturing signals in such an indoor system. Due to the influence of room acoustics, interference and background noise contamination, the sensor array received signals in the time domain may be expressed as

$$x_m[k] = \sum_{n=0}^{N-1} (h_{nm} * s_n)[k] + v_m[k], \quad m = 1, 2, \ldots, M, \tag{6.1}$$

where $*$ denotes linear convolution, h_{nm} is the RIR from the source location $\boldsymbol{r}_n$ to the place $\boldsymbol{r}_m$ of m-th microphone element and v_m is the noise component received by the m-th microphone, all the RIRs are simulated by ISM room simulator.

It should be mentioned that (6.1) is a linear time-invariant model of the physical acoustic transmission channel, the interference signals and background noise are received through a linear system, and they may thus be cancelled by linearly filtering the microphone signals. In the following, we firstly propose the indoor LCMV beamforming technique to design the beamformer in reverberant environment.

Assume there is an $L-$tap FIR filter behind each microphone element with coefficients $w_{m,l}, m = 1, 2, \ldots, M, l = 0, 1, \ldots, L-1$, then the beamformer output can be written as

$$y[k] = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_{m,l} x_m[k-l]. \tag{6.2}$$

It is noted that the convolution beamformer output can be reformulated by vector notation, and it is usually to stack the input samples and all filter coefficients into

vectors as

$$y[k] = \sum_{m=1}^{M} \boldsymbol{w}_m^T \boldsymbol{x}_m[k] = \boldsymbol{w}^T \boldsymbol{x}[k], \tag{6.3}$$

where

$$\boldsymbol{w}_m = [w_{m,0}, w_{m,1}, \ldots, w_{m,L-1}]^T, \ \boldsymbol{w} = [\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_M^T]^T,$$

$$\boldsymbol{x}_m[k] = [x_m[k], x_m[k-1], \ldots, x_m[k-L+1]]^T, \ \boldsymbol{x}[k] = [\boldsymbol{x}_1^T[k], \boldsymbol{x}_2^T[k], \ldots, \boldsymbol{x}_M^T[k]]^T.$$

If we substitute the input signals $x_m[k]$ into (6.1), the beamformer output can be rewritten as

$$y[k] = \sum_{m=1}^{M} (w_m * h_{0m} * s_0)[k] + \sum_{n=1}^{N-1} \sum_{m=1}^{M} (w_m * h_{nm} * s_n)[k] + \sum_{m=1}^{M} (w_m * v_m)[k]. \tag{6.4}$$

Then our objective is to find the separation filters w_m so that the first part of the above sum constructs our desired signal, while the last two parts of the sum vanish. The perfect performance on noise reduction, interference suppression and dereverberation need the filter coefficients satisfy the following conditions

$$\sum_{m=1}^{M} w_m * h_{0m} = \boldsymbol{g}_0, \tag{6.5}$$

$$\sum_{n=1}^{N-1} \sum_{m=1}^{M} w_m * h_{nm} = \mathbf{0}, \tag{6.6}$$

$$\sum_{m=1}^{M} w_m * v_m = \mathbf{0}, \tag{6.7}$$

where $\boldsymbol{g}_0 = h_D(k - \tau_L)$ is the direct path impulse response h_D from the SOI to beamformer output with some time delay τ_L, it is also known as the group delay of the sensor array, $\mathbf{0}$ is the vector with all entries equal to zero. In LCMV beamforming, the filter coefficients are adjusted based on the statistics of the sensor array measured

signals, the expression to be minimized is the power of noise signal. Since zero-mean signals are assumed, the cost function at time k may be defined as

$$J[k] = \boldsymbol{w}^T[k]\boldsymbol{R}_{\boldsymbol{vv}}[k]\boldsymbol{w}[k], \tag{6.8}$$

where $\boldsymbol{v}[k]$ denotes the output signal with noise active alone, $\boldsymbol{E}\{\cdot\}$ is the expectation operator, $\boldsymbol{w}[k]$ is the filter coefficients at time k, and $\boldsymbol{R}_{\boldsymbol{v}v}[k] = \boldsymbol{E}\{\boldsymbol{v}[k]\boldsymbol{v}^T[k]\}$ is the inputting correlation matrix at time k. To avoid the trivial solution at $\boldsymbol{w}[k] = \boldsymbol{0}$, a constraint of constant impulse response gain towards the SOI is added, by using convolution matrix notation, the constant beamformer filtering response constraint towards the SOI can be expressed as

$$\boldsymbol{H}_0\boldsymbol{w}[k] = \boldsymbol{g}_0, \tag{6.9}$$

where $\boldsymbol{H}$ is the convolution matrix generated from the RIRs from the SOI to each array element corresponding to the condition (6.5), and the construction is similar to (5.8). And similar constraints correspond to conditions (6.6) and (6.7) can be expressed as

$$\boldsymbol{H}_I\boldsymbol{w}[k] = \boldsymbol{0}, \tag{6.10}$$

with $\boldsymbol{H}_I$ corresponding to the convolution matrix generated from all the interferer RIRs. Writing all the constraints together and denote $\boldsymbol{H} = [\boldsymbol{H}_0^T \ \boldsymbol{H}_I^T]^T$ and $\boldsymbol{g} = [\boldsymbol{g}_0^T \ \boldsymbol{0}^T]^T$, then the indoor LCMV beamformer design problem can be formulated into an optimization problem as

$$\begin{aligned} \min_{\boldsymbol{w}[k]} \quad & \boldsymbol{w}^T[k]\boldsymbol{R}_{\boldsymbol{vv}}[k]\boldsymbol{w}[k] \\ s.t. \quad & \boldsymbol{H}\boldsymbol{w}[k] = \boldsymbol{g}, \end{aligned} \tag{6.11}$$

at each time index k.

6.1.2 Relaxed Beamformer Deisgn

The above LCMV beamformer design model (6.11) is usually very hard to be solved or no solution exists due to the conditions of the equality constraints $\boldsymbol{H}\boldsymbol{w}[k] = \boldsymbol{g}$. In this part, we construct a relaxed problem to find approximation solution for problem (6.11) by using the Lagrangian relaxation technique. The Lagrangian relaxation procedure uses the idea of relaxing the explicit linear constraints by bringing them into the objective function with associated vector $\boldsymbol{\lambda}[k]$ called the *Lagrange multiplier*, and the relaxed problem is constructed as

$$\begin{array}{ll} \max\limits_{\boldsymbol{\lambda}[k]} \min\limits_{\boldsymbol{w}[k]} & \mathcal{L}(\boldsymbol{w}[k], \boldsymbol{\lambda}[k]) \\ s.t. & \boldsymbol{w}[k] \in \mathbb{D}, \end{array} \tag{6.12}$$

where $\mathbb{D}$ is an auxiliary feasible region, and $\mathcal{L}(\boldsymbol{w}[k], \boldsymbol{\lambda}[k])$ is the Lagrangian function defined as

$$\mathcal{L}(\boldsymbol{w}[k], \boldsymbol{\lambda}[k]) = \frac{1}{2}\boldsymbol{w}^T[k]\boldsymbol{R}_{\boldsymbol{vv}}[k]\boldsymbol{w}[k] - \boldsymbol{\lambda}[k]^T(\boldsymbol{H}\boldsymbol{w}[k] - \boldsymbol{g}).$$

Assuming the feasible region of problem (6.12) is $\mathbb{D} = \mathbb{R}^{ML}$, and the correlation matrix $\boldsymbol{R}_{\boldsymbol{vv}}[k]$ is positive-semidefinite and symmetric, by setting the gradient of $\mathcal{L}(\boldsymbol{w}[k], \boldsymbol{\lambda}[k])$ with respect to $\boldsymbol{w}[k]$ to be zero

$$\frac{\partial \mathcal{L}(\boldsymbol{w}[k], \boldsymbol{\lambda}[k])}{\partial \boldsymbol{w}[k]} = \boldsymbol{R}_{\boldsymbol{vv}}[k]\boldsymbol{w}[k] - \boldsymbol{H}^T\boldsymbol{\lambda}[k] = 0,$$

notice the inverse of $\boldsymbol{R}_{\boldsymbol{vv}}[k]$ is exist, we have

$$\boldsymbol{w}[k] = \boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{\lambda}[k]. \tag{6.13}$$

Then substituting it into the equality constraints $\boldsymbol{H}\boldsymbol{w}[k] = \boldsymbol{g}$, we have the subproblem for multiplier $\boldsymbol{\lambda}[k]$ as

$$\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{\lambda}[k] = \boldsymbol{g}. \tag{6.14}$$

In general, if the matrix $\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T$ is nonsingular, we can solve the multiplier $\boldsymbol{\lambda}[k]$ directly by

$$\boldsymbol{\lambda}[k] = (\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T)^{-1}\boldsymbol{g}. \tag{6.15}$$

However, the combinational convolution matrix $\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T$ has very large condition number due to the large scale and sparse of RIRs h_{nm} (see the example in Table 5.1), the solver (6.15) is difficult to be achieved in our numerical simulations. On the other hand, the multiplier design subproblem (6.14) is in fact equivalent to the least square problem

$$\min_{\boldsymbol{\lambda}[k]} \quad ||\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{\lambda}[k] - \boldsymbol{g}||_2^2. \tag{6.16}$$

The direct solution of problem (6.16) may lead to a vector $\boldsymbol{\lambda}[k]$ that is severely contaminated with noise, therefore many regularization techniques are employed to get a more meaningful solution, such as Tikhonov regularization [40]. The Tikhonov regularization approach improves the condition of the problem by solving the linear least square problem

$$\min_{\boldsymbol{\lambda}[k]} \quad ||\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{\lambda}[k] - \boldsymbol{g}||_2^2 + \alpha||\boldsymbol{\lambda}[k]||_2^2 \tag{6.17}$$

instead of (6.16), and the solution of problem (6.17) is converged to the solution of problem (6.16) in Lemma 5.1 under condition. Moreover, the solution $\boldsymbol{\lambda}_\alpha[k]$ satisfies the equation

$$(\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}[k]^{-1}\boldsymbol{H}^T + \alpha\boldsymbol{I})\boldsymbol{\lambda}_\alpha[k] = \boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{g},$$

and a positive matrix $\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T + \alpha\boldsymbol{I}$ can be obtained with an appropriate regularization parameter α, there are also many regularization criteria for choosing of it, Morozov's discrepancy principle [87], the Gfrerer/Raus-method [38],

the quasi-optimality criterion [4], the generalized cross-validation [41] and the L-curve criterion [49]. Then the approximate multiplier can be obtained

$$\boldsymbol{\lambda}_\alpha[k] = (\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T + \alpha\boldsymbol{I})^{-1}\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{g}. \tag{6.18}$$

Thus, by substituting the approximate multiplier $\boldsymbol{\lambda}_\alpha[k]$ into (6.13), we can get the optimal filter coefficients

$$\boldsymbol{w}[k] = \boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T((\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T + \alpha\boldsymbol{I})^{-1}\boldsymbol{H}\boldsymbol{R}_{\boldsymbol{vv}}^{-1}[k]\boldsymbol{H}^T\boldsymbol{g}). \tag{6.19}$$

6.2 Post-filtering Techniques

In this section, we introduce several post-filtering techniques to improve the quality of the speech enhancement after beamforming. Since the output of the beamformer is a single channel signal, we assume the speech enhancement model is $y[k] = s[k] + u[k]$, where $u[k]$ denotes all the interferer and noise parts of the beamformer output with respect to desired signal $s[k]$. By using the short-time Fourier transform (STFT) on the beamformer output $y[k]$, the spectral components of the noisy signal $Y(p,q)$ can be obtained by

$$Y(p,q) = \sum_{t=1}^{T} y[qR+t]\omega[t]e^{\frac{-j2\pi pt}{T}}, \tag{6.20}$$

where p and q are the frequency bin index and time frame index, R is the frame rate and $\omega[t]$ is a window function (such as *Hanning* window). In most of the single channel signal processing, the desired signal spectrum $S(p,q)$ is estimated from

$$\hat{S}(p,q) = G(p,q)Y(p,q), \tag{6.21}$$

where $G(p,q)$ is a nonlinear gain function. And the gain function is mostly expressed as a function of the *a priori* SNR

$$\xi(p,q) = \frac{\boldsymbol{E}\{|S(p,q)|^2\}}{\boldsymbol{E}\{|U(p,q)|^2\}} = \frac{\sigma_S^2(p,q)}{\sigma_U^2(p,q)}, \tag{6.22}$$

where $\sigma_S^2(p,q)$ and $\sigma_U^2(p,q)$ denote the desired signal power spectral density (PSD) and noise PSD, respectively.

Most of the gain functions developed for speech enhancement are based on the MMSE optimization criteria, such as the WF approach

$$G_{WF}(p,q) = \frac{\xi(p,q)}{1+\xi(p,q)}, \tag{6.23}$$

or LSA approach

$$G_{LSA}(p,q) = \min\left\{\zeta, \frac{\xi(p,q)}{1+\xi(p,q)} e^{\frac{1}{2}\int_{\mu(p,q)}^{\infty}\frac{e^{-t}}{t}dt}\right\}, \tag{6.24}$$

with

$$\mu(p,q) = \frac{\xi(p,q)}{1+\xi(p,q)}\gamma(p,q),$$

where $\gamma(p,q) = |Y(p,q)|^2/\sigma_U^2(p,q)$ denotes the *a posteriori* SNR, and ζ is a practical upper bound (such as $\zeta = 10$) used to prevent a large gain value at low *a posteriori* SNR. And the optimization of the criteria is made under certain model conditions such as stationarity and some distributions, it is desirable to get a gain function that offers optimal performance in all scenarios. However, different cost functions and gain functions are needed in different noise and interferer scenarios, and some of them involve complex mathematics equations that require heavy computational load. To design a gain function with low complexity and high flexibility, while having similar or better performance comparing to MMSE based estimators, Yong *et al.* developed a flexible sigmoid-shape gain function in [120], the rational behind using the SIG function is that it is a general cumulative distribution function (CDF) function with a shape that can be adjusted by several tunable parameters. Instead of the SIG function presented to map with the *a posteriori* SNR in [120], they proposed

a SIG gain function mapping with the *a priori* SNR estimate in [119] most recently, given as

$$G_{SIG}(p,q) = \frac{1}{1+e^{-a(\xi(p,q)-c)}}, \tag{6.25}$$

where a and c are parameters using to control the slope and mean of the gain curve, respectively, both parameters control the amount of musical noise, speech distortion and noise reduction. In order to obtain a balanced trade-off between them, the sigmoid slope has to be sensitive towards speech and less sensitive towards the variation of noise.

To provide more noise reduction at low SNR conditions, a modified sigmoid gain function (MSIG) by multiplying the SIG gain function (6.25) with a hyperbolic tangent function has been developed in [119], given as

$$G_{MSIG}(p,q) = \frac{1-e^{-a_1\xi(p,q)}}{1+e^{-a_1\xi(p,q)}} \times \frac{1}{1+e^{-a_2(\xi(p,q)-c)}}, \tag{6.26}$$

where a_1, a_2 and c are parameters defined similarly to (6.25). And during their experimental evaluation, the MISG gain functions generated the lowest speech distortion among all evaluated gain functions at large smoothing parameters. Meanwhile, Yong *et al.* [119] also proposed a modified decision-directed (MDD) approach to improve the mostly used *a priori* SNR estimation DD approach developed by Ephraim and Malah in [27]. As observed from the DD approach

$$\hat{\xi}_{DD}(p,q) = \max\left\{\beta\frac{|\hat{S}(p,q-1)|^2}{\hat{\sigma}_U(p,q)} + (1-\beta)\boldsymbol{P}[\gamma(p,q)-1], \xi_0\right\}, \tag{6.27}$$

where β and $\hat{\sigma}_U(p,q)$ denote the smoothing factor and estimated noise PSD from the preceding frame, $\boldsymbol{P}[\cdot]$ denotes the half-wave rectification and ξ_0 denotes a SNR floor. The advantage of DD approach is its capability to eliminate musical noise based on

the choice of β in the conditional smoothing procedure [15]. It was suggested to set β close to unity so that the musical noise can be eliminated, while this leads to a slow update of the *a priori* SNR estimate, resulting in speech transient distortion. Moreover, the DD approach (6.27) has an extra frame delay during speech transients since it employs the previous frame clean speech spectrum estimate. The MDD approach is developed to reduce the delay in speech transients and the *a priori* SNR estimate with the current noisy speech spectrum, it modified the first term of the DD approach such that the gain function at previous frame is mapped with the current noisy speech spectrum rather than the previous one. The MDD approach is defined as

$$\hat{\xi}_{MDD}(p,q) = \max\left\{\beta\frac{|G_{(\cdot)}(p,q-1)Y(p,q)|^2}{\hat{\sigma}_U(p,q)} + (1-\beta)\boldsymbol{P}[\gamma(p,q)-1], \xi_0\right\}, \quad (6.28)$$

where $G_{(\cdot)}$ indicates that the same gain function is used to obtain both the *a priori* SNR and the speech estimate. The advantage of MDD approach is that it has the same complexity as the DD approach while having a better enhanced speech quality, which makes it suitable for real-time implementation. More details on the properties of MDD approach are elaborated in [119].

In our post-filtering step, both the SIG and MSIG gain functions are introduced to design post-filter, and compared with the MMSE based WF and LSA methods. And we use the MDD approach to estimate the *a priori* SNR during experimental simulation.

6.3 Illustration Examples

In the following numerical experiments, we introduce the direct-to-reverberant ratio (DRR) defined in (5.32), the generalized signal-to-noise, interference and-reverberation ratio (SNIRR) defined similarly to (5.33), the PESQ score measure defined in (5.34),

the interference and noise suppression defined in (5.35) and the indicator for dereverberation defined in (5.36).

6.3.1 Acoustic Room Configuration

A simple rectangular $12m \times 6m \times 3m$ room is defined for acoustic room modeling, all the walls, floor and ceiling are assumed to be with uniform absorption coefficients, which resulting from the desired reverberation time T_{60}. We define a 9-elements sensor array in the $z = 1m$ plane with L-tap FIR filter behind each microphone, specifically, they are placed at

Mic = { (5.5,3), (5.625,3), (5.75,3), (5.875,3), (6,3), (6.125,3), (6.25,3), (6.375,3), (6.5,3)}

in meter, and with the center element as the reference receiving point, also the beamformer output point, where assume the lower right corner as the room origin. There are also one SOI placed at $(5, 2)$, one interferer placed at $(8, 2)$, and background noise placed at $(8, 4)$ at plane $z = 1m$, where the background noise placement information is not used during beamformer design, an illustration of the room setup configuration at $z = 1m$ plane is plotted in Figure 6.1. All the RIRs from the considering points to sensor array are simulated by image source method, and the fast-ISM simulator developed by E.A. Lehmann and A.M. Johansson [68] will be introduced for modeling room acoustics at $T_{60} = 0.1s$ and $T_{60} = 0.2s$.

In such a room configuration, the inputting SOI signal is a broadband male speech (*"Dots of light betrayed the black cat"*), and use another female speech (*"she had your dark suit in greasy wash water all year"*) as the interference (INT) inputting. Due to the noise signal is assumed to be independent to speech signal, we can generate a white-noise for background noise inputting as NOI, see the plotting in Figure 6.1.

Assume the output point of microphone array is set up at the center of it, thus before processing, the initial DRR value of the RIR from SOI to the center micro-

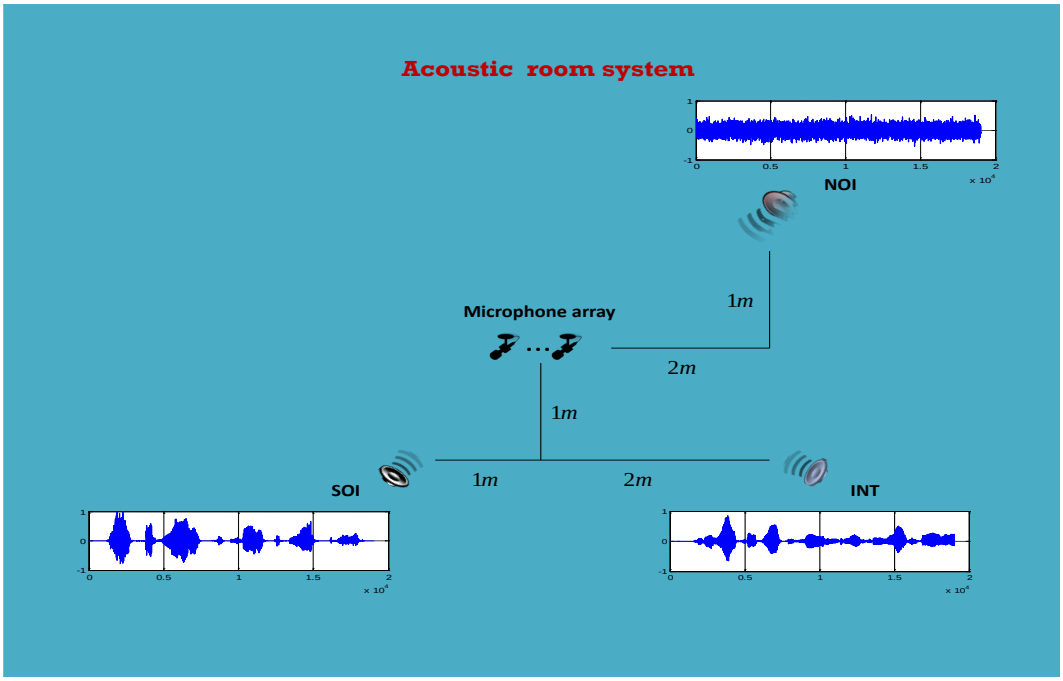


Figure 6.1: The $z = 1$ plane of the defined acoustic room.

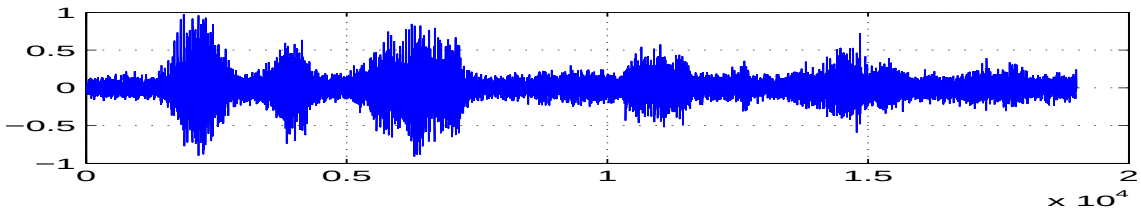
phone element, the initial values of SNIRR and PESQ of the unfiltered signals under different room acoustics can be measured for comparison, and they are listed in Table 6.1. The unfiltered signals are plotted in Figure 6.2.

Table 6.1: The initial values measurement.

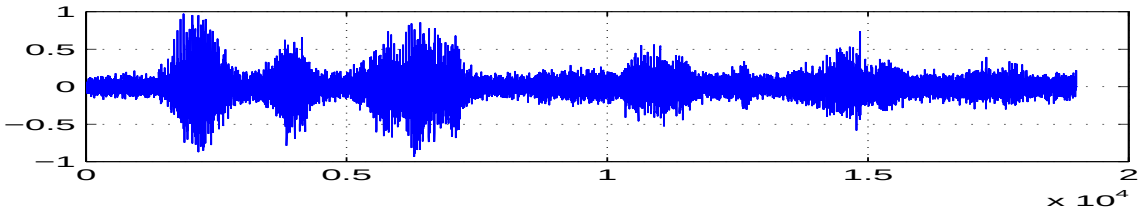
T_{60}	DRR	$SNIRR_{seg}$	PESQ
0.1	9.7839	-4.1717	1.6781

T_{60}	DRR	$SNIRR_{seg}$	PESQ
0.2	6.4927	-4.6295	1.6810

It is noted that for the different filter length L, the indoor LCMV beamformer design problem (6.11) is different, and the corresponding relaxed problem (6.12) is also different, and so is the group delay τ_L. Thus, the beamformer designed from the relaxed LCMV model (6.12) may have different performances with respect to different filter length L and group delay τ_L. In the following, we study the influence of group delay and filter length on indoor LCMV beamformer design respectively.



(a) at $T_{60} = 0.1s$



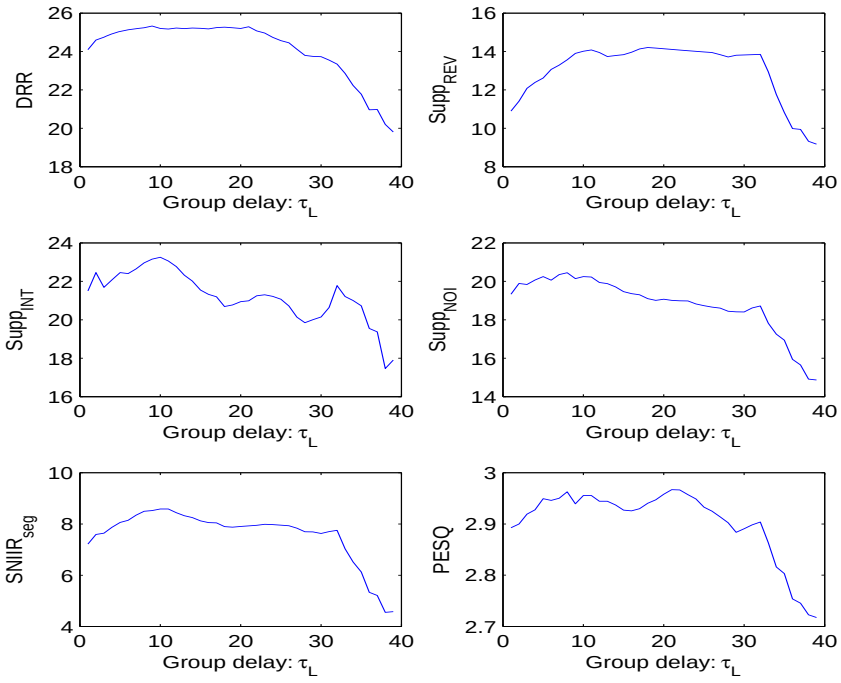
(b) at $T_{60} = 0.2s$

Figure 6.2: The unfiltered signals under different acoustic room.

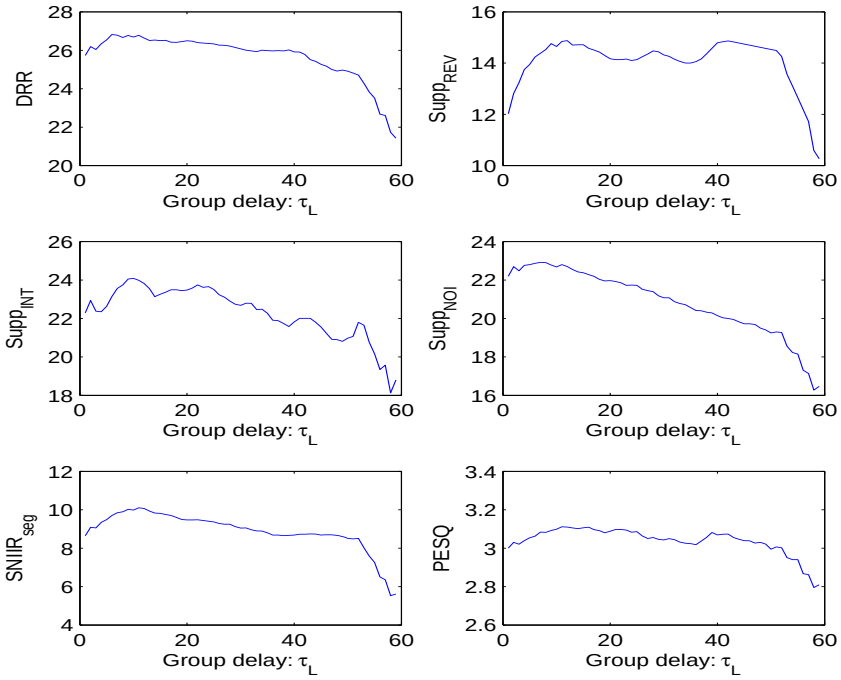
6.3.2 Group Delay Learning

Firstly, we study the influence of group delay τ_L on indoor LCMV beamformer design. For the given filter length L, we introduce the DRR, reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT}$, noise reduction $Supp_{NOI}$, segmental SNIRR and PESQ scores six indicators to do comparison. We choose $L = 40$ and $L = 60$ two cases, and learn the group delay τ_L from 1 to $L-1$ for illustration in the following. The overall performance of group delay learning is depicted in Figures 6.3 and 6.4.

From the group delay τ_L learning results, we can see that the group delay selection do have influence on the beamformer design, too small or large group delay selection is not acceptable. And for different room acoustics, the similar results show that the better group delay should be chosen around $L/2$, whereas the all the measurement indicators are consistent on the whole.

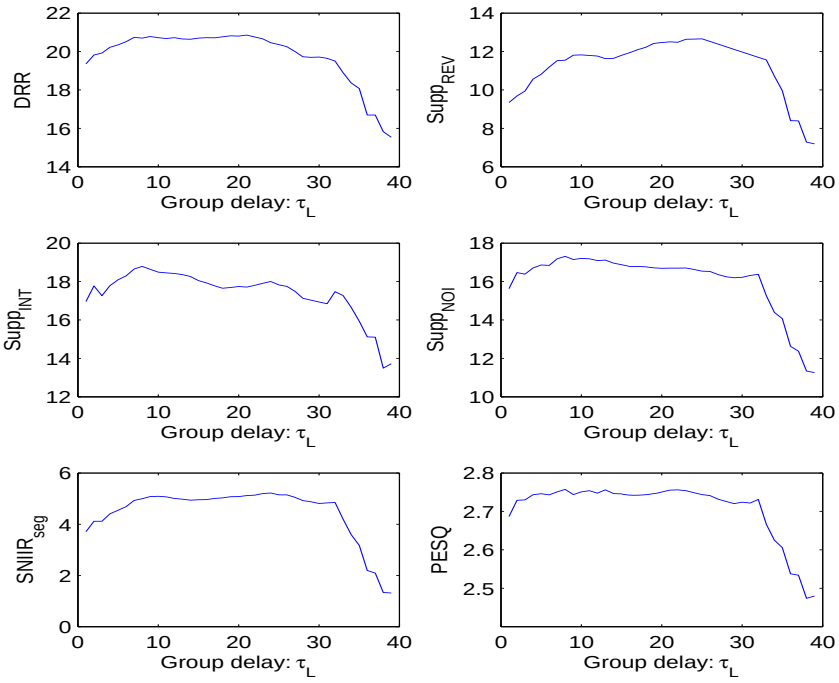


(a) for filter length $L = 40$

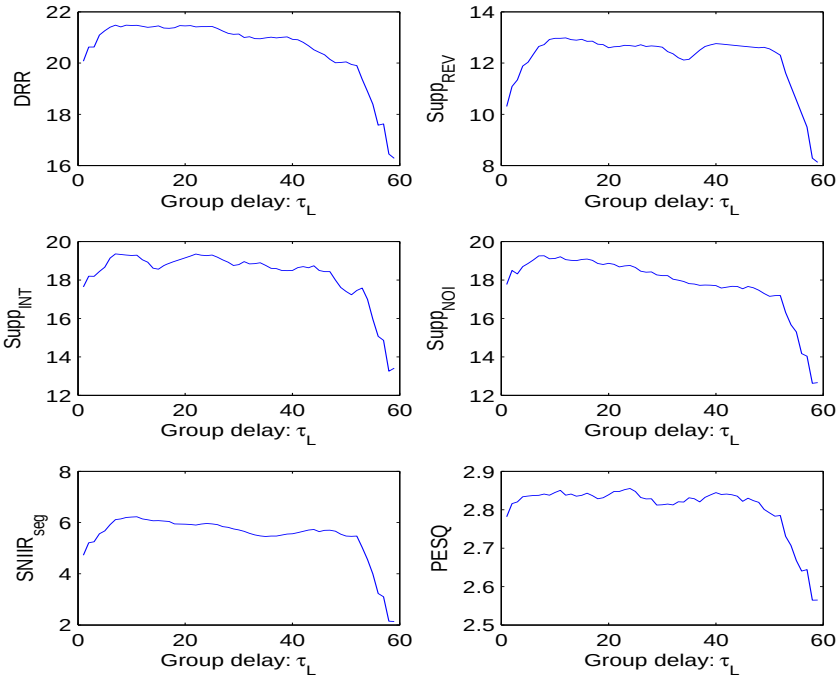


(b) for filter length $L = 60$

Figure 6.3: Learning for group delay τ_L at $T_{60} = 0.1s$.



(c) for filter length $L = 40$



(d) for filter length $L = 60$

Figure 6.4: Learning for group delay τ_L at $T_{60} = 0.2s$.

6.3.3 Filter Length Learning

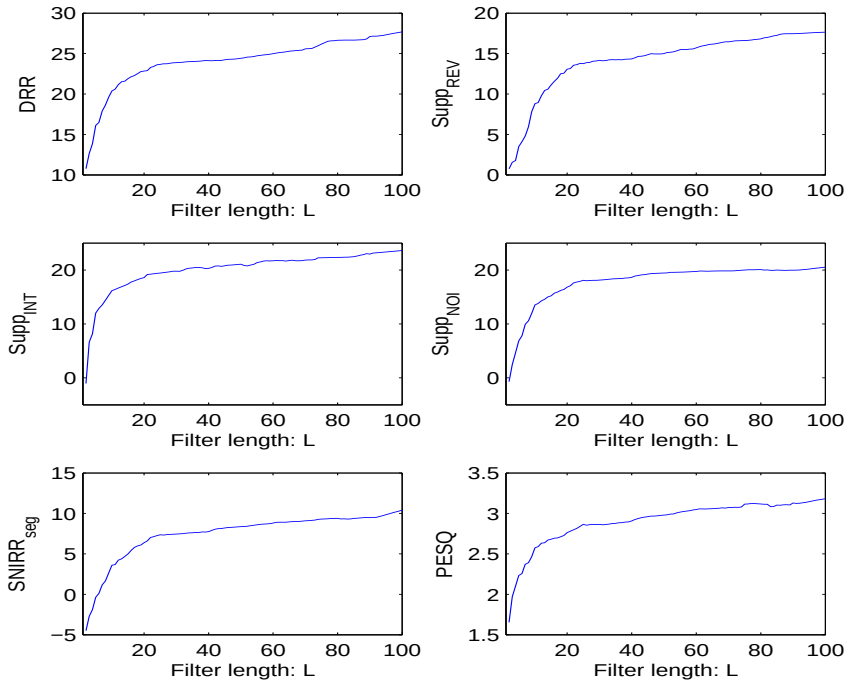
From the above group delay learning results, it is difficult to define a fixed group delay selection to achieve optimal objectives on DRR, reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT}$, noise reduction $Supp_{NOI}$, segmental SNIRR and PESQ scores. But for the best performance on dereverberation, we can choose the group delay τ_L for optimal $Supp_{REV}$ to learn the influence of filter length L on indoor LCMV beamformer design in the following.

To study the filter length learning, we learn the filter length L from 2 to 100, and the overall results on filter length learning are plotted in Figure 6.5.

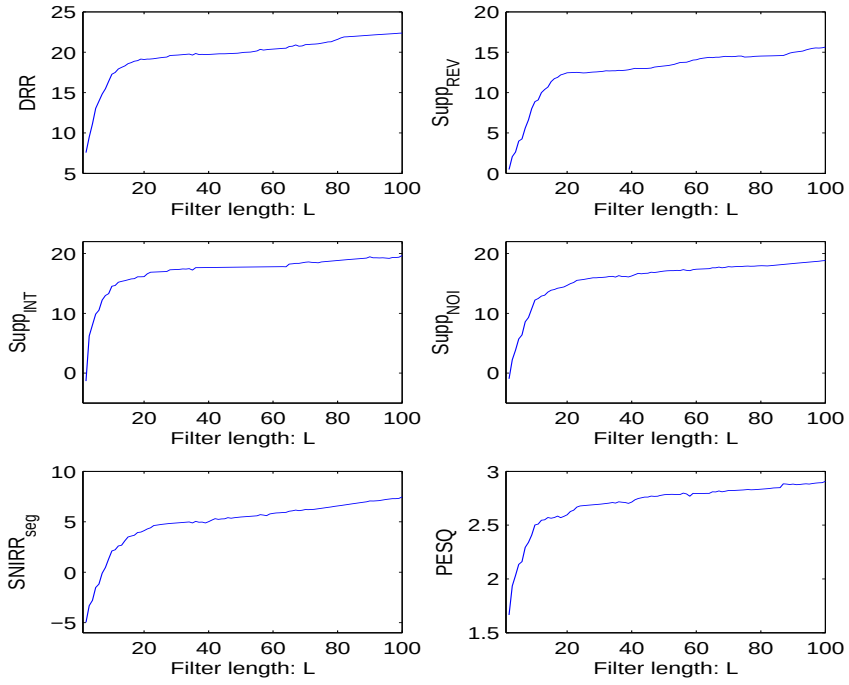
From the filter length learning results, we can see that the designed indoor LCMV beamformers have effect on speech enhancement in reverberant environment when the filter length $L > 20$. All the measurement indicators are increased as the filter length increasing, whereas they are increased slowly when the filter length $L > 40$, especially the performance on noise reduction.

6.3.4 Overall Performance Illustration

To take an illustration on speech enhancement by the beamforming approach, in the following, we choose filter length $L = \{30,\ 40,\ 50,\ 60\}$ to design the indoor LCMV beamformers at the best group delay selection for dereverberation. With the designed beamformer, we do comparison on the beamformed response, reverberant and interference responses suppression, reverberant and interference signals suppression and noise reduction, respectively. We use the DRR, reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT}$, noise reduction $Supp_{NOI}$, segmental SNIRR and PESQ scores six indicators to evaluate the effect of speech enhancement, and list the results in the following Tables 6.2 and 6.3. In the table, BF, WFB, LSAB and MSIGB denote the results from the output corresponding to LCMV beamformer,



(a) at $T_{60} = 0.1s$



(b) at $T_{60} = 0.2s$

Figure 6.5: Performance with filter length L.

WF-postfilter, LSA-postfilter and MSIG-postfilter, respectively.

Table 6.2: Overall performance of indoor LCMV beamformer at $T_{60} = 0.1s$.

L	DRR	$Supp_{REV}$	$Supp_{INT}$	$Supp_{NOI}$	$SNIRR_{seg}$	PESQ
30	23.8298	14.1449	20.0948	18.0421	7.4391	2.8507
40	24.4513	14.9424	20.7179	18.6624	7.9307	2.9249
50	24.6306	15.5551	20.8459	19.3706	8.1222	2.9575
60	24.9204	16.0928	20.8981	19.4996	8.3718	3.0298

Table 6.3: Overall performance of indoor LCMV beamformer at $T_{60} = 0.2s$.

L	DRR	$Supp_{REV}$	$Supp_{INT}$	$Supp_{NOI}$	$SNIRR_{seg}$	PESQ
30	19.4683	12.5691	17.1087	16.0111	4.7660	2.6995
40	20.0033	12.7314	17.4916	16.3633	5.0487	2.7320
50	20.0670	13.2025	17.5535	17.0516	5.4236	2.7847
60	20.1254	13.6151	17.6105	17.2993	5.5425	2.8009

From the overall evaluation results given in the Tables 6.2 and 6.3, we can see that the indoor LCMV beamformer design at longer filter length can achieve better performance on all the DRR, reverberation suppression, interference suppression, noise reduction, segmental SNIRR and PESQ scores. We can also find that all the indicators for the beamformers design at $T_{60} = 0.2s$ are small than the beamformers design at $T_{60} = 0.1s$, it is demonstrate that the influence of room acoustic on speech quality enhancement.

In the following, we also depict the overall performance on the designed beamformers at filter length $L = 60$ in Figures 6.6 and 6.8 at different room acoustics for illustration. We can see that all the RIRs from the SOI to center microphone element (CMic) can be effectively beamformed to the desired responses, which makes the desired SOI signals successfully formulated. The reverberant IRs and interference RIRs can be effectively suppressed, which makes the reverberation and interference signals successfully suppressed. The illustration of reverberant, interference and noise signals suppression can also be found in Figures 6.7 and 6.9.

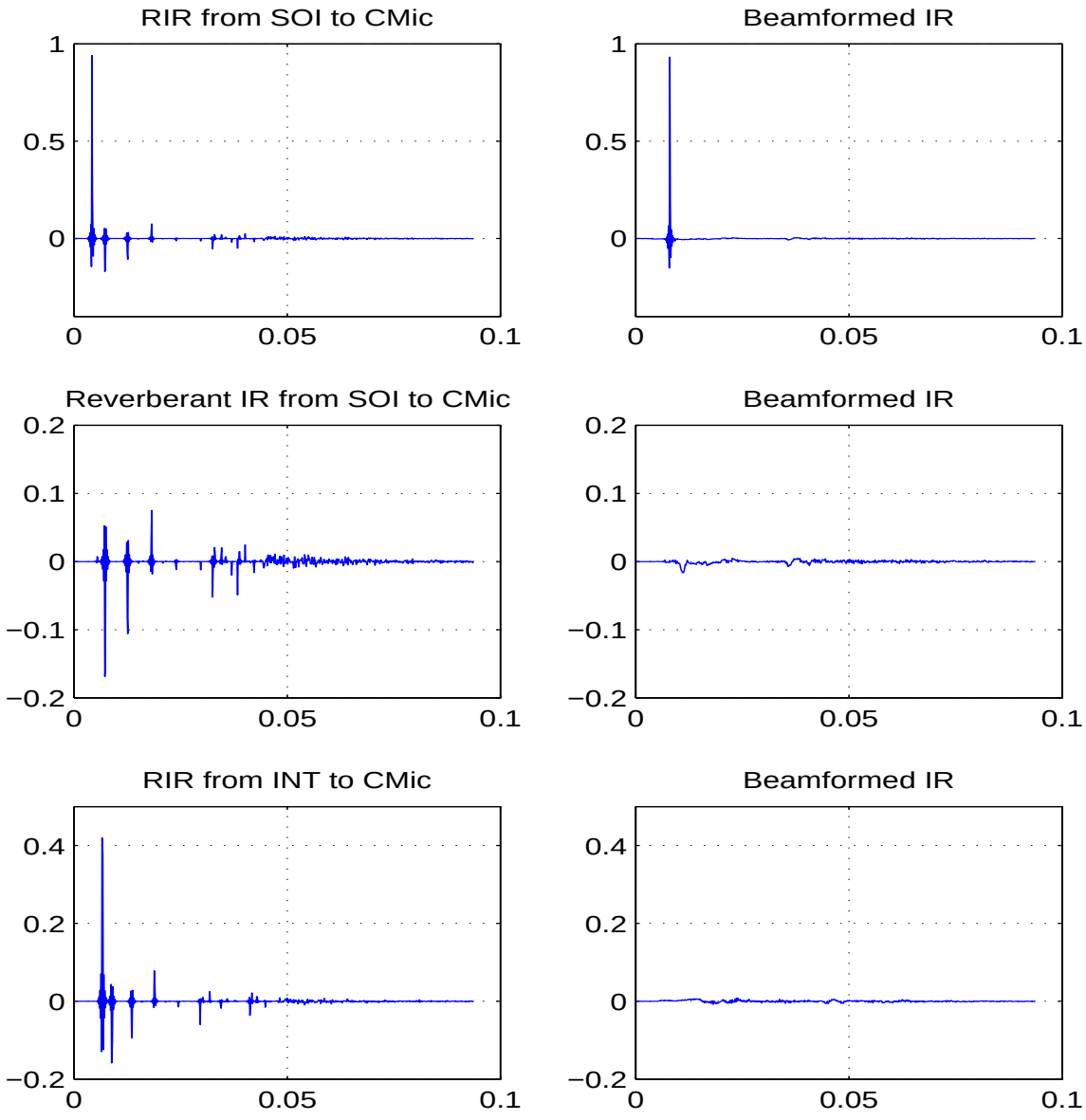


Figure 6.6: RIRs for SOI and INT before and after beamformer filtering.

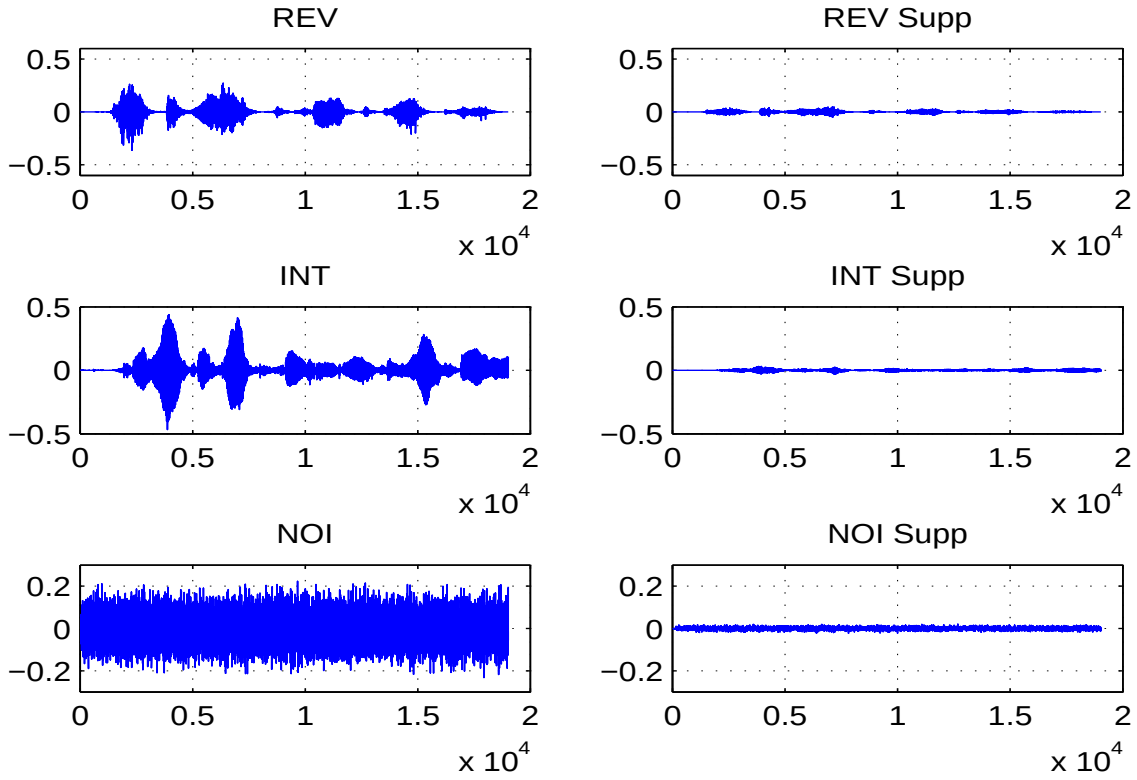


Figure 6.7: Interference, reverberation and noise before and after beamformer filtering.

From the illustration example Figures 6.7 and 6.9, we can see that under the same interference and background noise environment, the stronger room acoustics ($T_{60} = 0.2s$) has serious influence on dereverberation, and it has influence on all the reverberation, interference and noise suppression.

6.3.5 Post-filtering Processing with LCMV Beamformer

From the filter length learning results, the more improvement on dereverberation and interference suppression needs longer filter length, whereas, the background noise reduction is increased slowly as the filter length increasing when the filter length $L > 30$. Fortunately, many effective approaches based on single channel processing can achieve great improvement on noise reduction. Thus, in the following, we do experiments on the introduced post-filtering techniques for noise reduction evaluation.

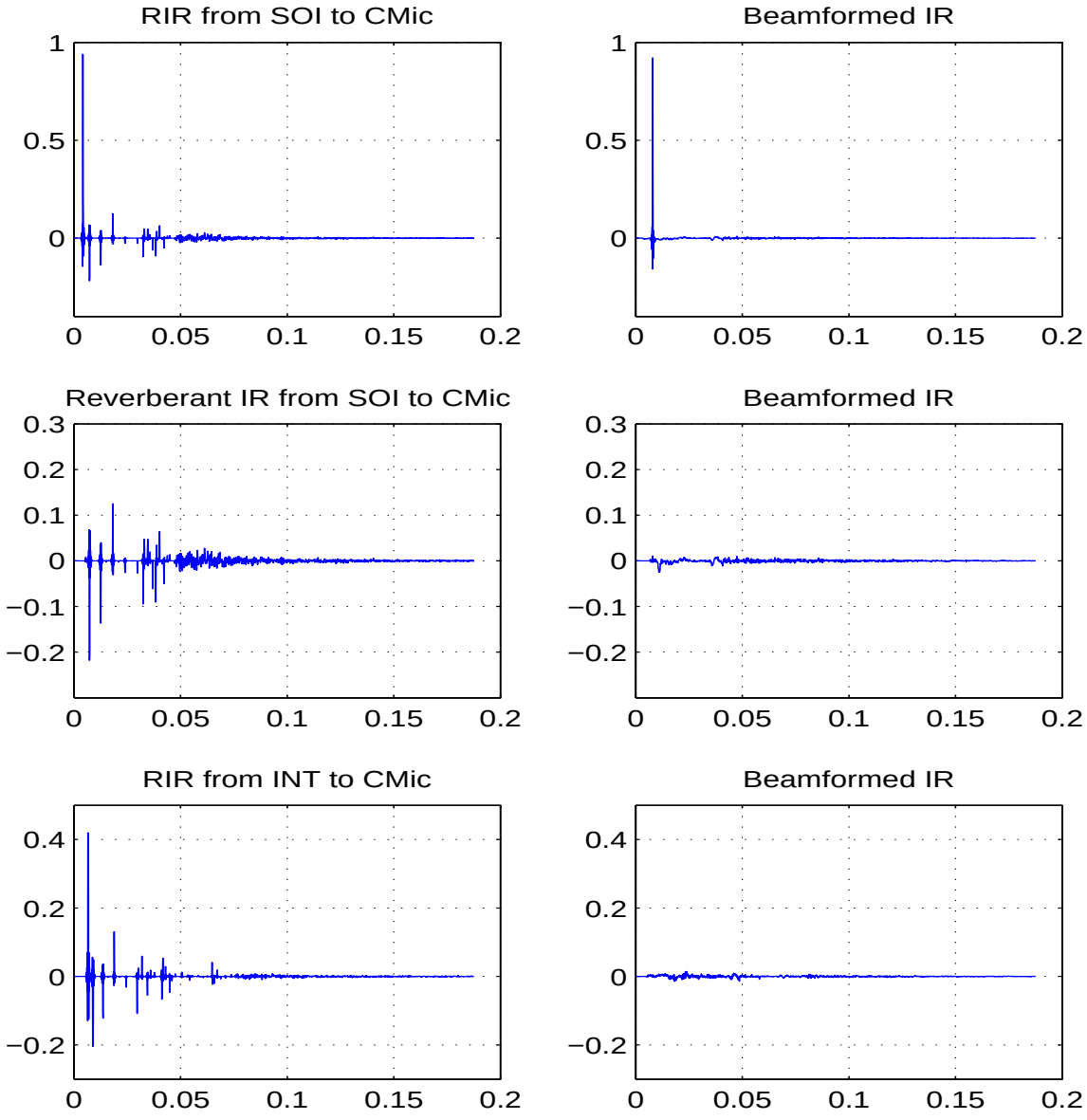


Figure 6.8: RIRs for SOI and INT before and after beamformer filtering.

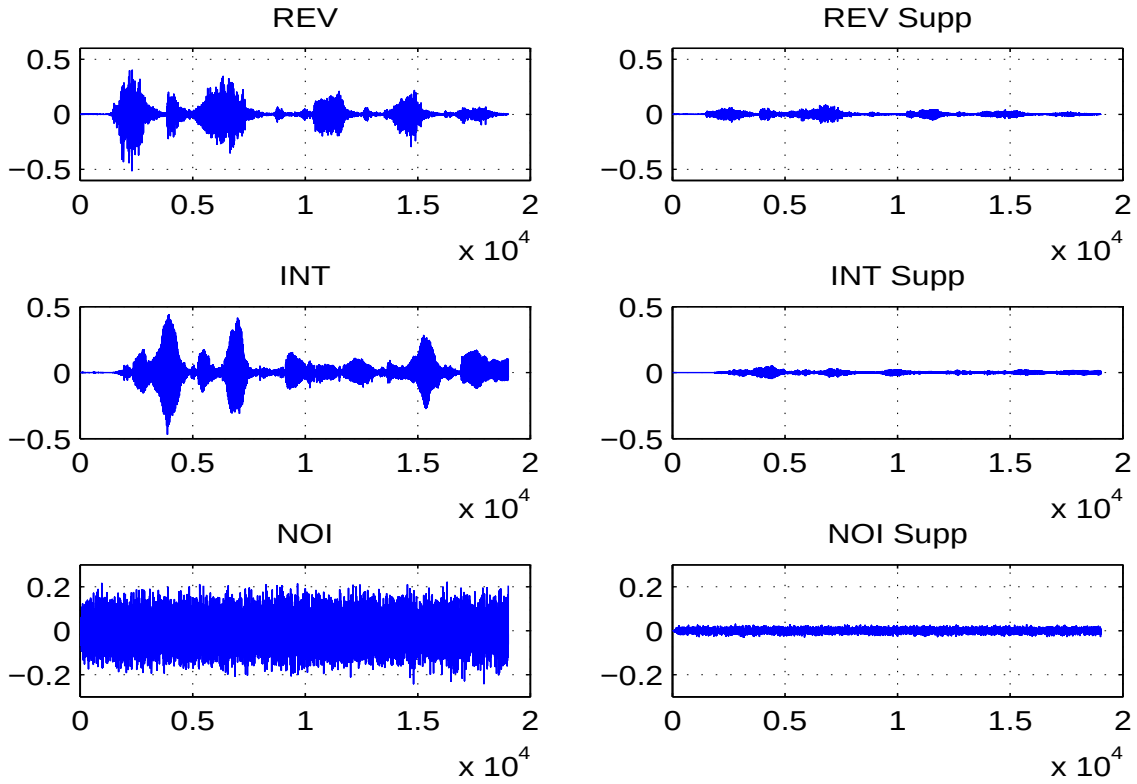


Figure 6.9: Interference, reverberation and noise before and after beamformer filtering.

To investigate the performance of post-filtering techniques on remaining background noise reduction, then further enhance the speech quality, we use four filter length $L = \{30,\ 40,\ 50,\ 60\}$ with the group delay for optimal dereverberation to design the indoor LCMV beamformers firstly, then post-filter the beamformer output using the WF-postfiltering, LSA-postfiltering and MSIG-postfiltering techniques, respectively.

We use the noise reduction $Supp_{NOI}$, segmental SNIRR and PESQ three indicators to evaluate the improvement effect of post-filtering techniques, and list the results in the following Tables 6.4 and 6.5. In the table, WFB, LSAB and MSIGB denote the results from the output corresponding to WF-postfilter, LSA-postfilter and MSIG-postfilter, respectively.

Table 6.4: Overall performance of post-filters at $T_{60} = 0.1s$.

L	$Supp_{NOI}$			$SNIRR_{seg}$			PESQ		
	WFB	LSAB	MSIGB	WFB	LSAB	MSIGB	WFB	LSAB	MSIGB
30	26.1124	26.0677	27.3918	7.4797	7.9220	8.2207	3.0187	3.0736	3.1532
40	27.0223	26.0453	27.6355	7.9927	8.1997	8.5045	3.1304	3.1806	3.2567
50	27.1258	27.2973	28.5590	8.1358	8.2642	8.5641	3.1463	3.2114	3.2709
60	27.7719	28.3020	29.1612	8.4970	8.5079	8.8693	3.2200	3.2322	3.3204

Table 6.5: Overall performance of post-filters at $T_{60} = 0.2s$.

L	$Supp_{NOI}$			$SNIRR_{seg}$			PESQ		
	WFB	LSAB	MSIGB	WFB	LSAB	MSIGB	WFB	LSAB	MSIGB
30	24.1493	23.9916	25.7384	5.0357	5.3198	5.6198	2.8869	2.9506	3.0046
40	24.1124	23.8948	25.8772	5.1188	5.4174	5.6307	2.9077	2.9592	3.0200
50	24.7411	24.1912	26.0592	5.4343	5.5131	5.8766	2.9539	3.0015	3.1068
60	24.6220	24.2172	26.1117	5.5041	5.5572	5.7703	2.9763	3.0726	3.1682

Compare with the beamforming results in Tables 6.2 and 6.3, we can see that all the introduced post-filtering techniques have good effect on noise reduction improvement. They can improve the noise reduction to some extent, then improve the segmental SNIRR and PESQ scores no matter the room acoustics or filter length changes. Moreover, among these three post-filtering techniques, the MSIG-postfilter has the best effect.

The specific filtered signals plotting for indoor LCMV beamformer at filter length $L = 60$ and after post-filtering are depicted in the following Figures 6.10 and 6.11. We find that the desired SOI signals can be effectively reformulated by the proposed beamforming methods.

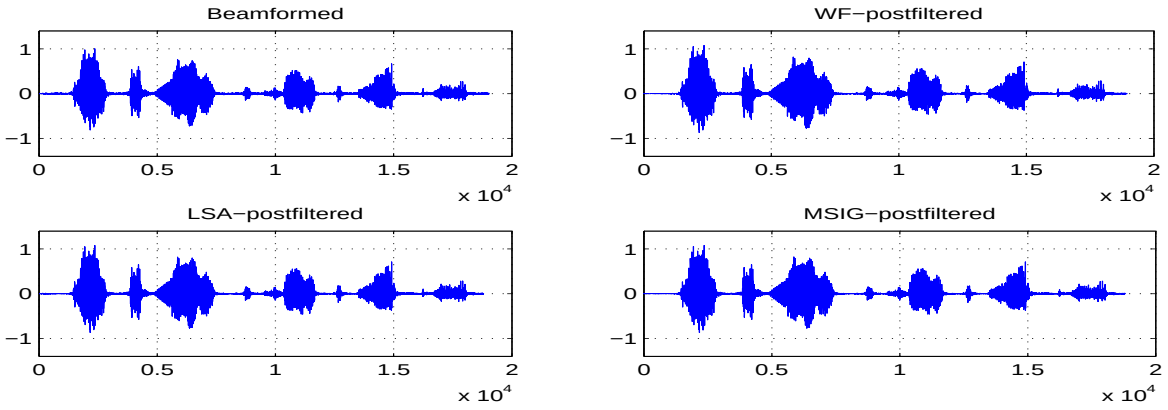


Figure 6.10: Filtered signals at $T_{60} = 0.1s$.

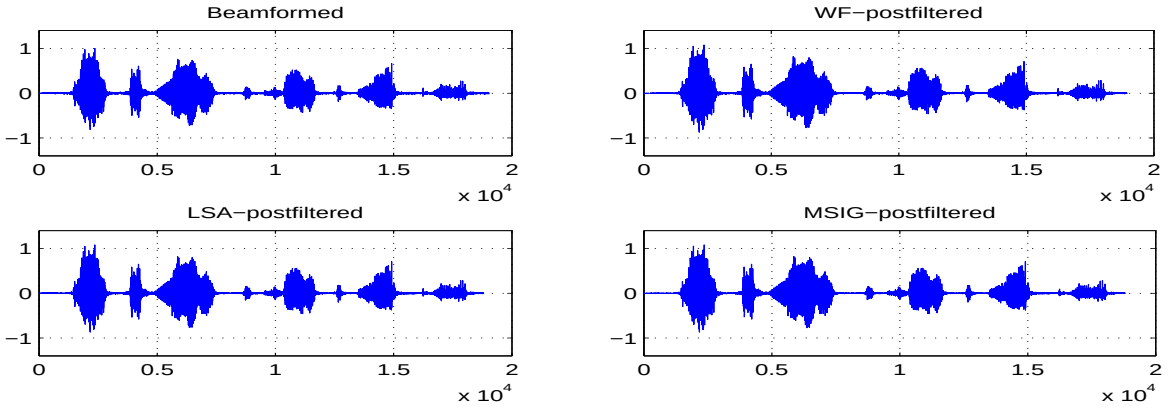


Figure 6.11: Filtered signals at $T_{60} = 0.2s$.

Chapter 7

Common Microphone Array Placement Design

In this chapter, the microphone array placement design problem on the design near-field broadband beamformer is studied. We establish a nonlinear optimization problem for the microphone array placement design. To evaluate the effect of the placement design, we also define a cost function to do measurement, the error between beamformer output and desired filter responses. Whereas this problem is with respect to the optimal beamformer coefficients solving.

In general, the optimal beamformer coefficients are according to the optimal frequency responses, and it is achieved as the filter length increasing to infinity. To avoid the optimal filter coefficients solving problem, we introduce the infinite length technique to convert the filter coefficients solving into the problem of performance limit estimation of the beamformer output. Then, we develop hybrid descent method with genetic algorithm to solve the optimal placement design.

7.1 Placement Design Problem Formulation

Given an M-elements microphone array, where behind each elements is an $L-$tap finite impulse response (FIR) filter, assume that the microphone elements have been set at the fixed points $\boldsymbol{r}_i,\ i = 1, 2, .., M$, then, the transfer function from any source point to the $i-$th microphone is given by (2.5) in the open air. If the signals received by this microphone array are sampled synchronously at the rate of f_s per second, the frequency responses $W_i(\boldsymbol{w}, f, L),\ i = 1, 2, ..., M$, of these FIR filters are given by (2.6), and the beamformer output is (2.7). Thus, optimization problem can be established based on some criteria to measure the error between beamformer output and the given desired response, such as the beamformer design model (2.8) according to previous discussion in Section 2.2 of Chapter 2.

In the previous chapters, we'd like to use the L_1-norm to measure the objective function in minimax model (2.8), whereas, there are many objective criteria exist to measure the error between $G(\boldsymbol{r}, f, L)$ and $G_d(\boldsymbol{r}, f, L)$, such as the $L_\infty-$norm designed in [5, 39], these methods have promoted the development of broadband beamformer greatly in recent decades. However, these design techniques are often very expensive, and their beamforming effects are not quite as good in general. In 2011, Feng *et al.* [32] used the L_2-norm and the infinite length technique to estimate the performance limit of the FIR filters, and quickly found the optimal FIR filter coefficients by using the active constraints during the performance limit estimate, which is called a two-stage method, and it proved to be quite effective and very fast.

In the following, we will use the L_2-norm minimization model for the design of broadband beamformer. Firstly, we define the objective function as

$$E(\boldsymbol{w}) = \frac{1}{|\Omega|}\int_{\Omega} \rho(\boldsymbol{r}, f)\left|\boldsymbol{H}^H(\boldsymbol{r}, f)\boldsymbol{W}(\boldsymbol{w}, f, L) - G_d(\boldsymbol{r}, f, L)\right|^2 d\boldsymbol{r}df, \tag{7.1}$$

where Ω is a specified spatial-frequency domain as the definition field of $G_d(\boldsymbol{r}, f, L)$,

and $\rho(\boldsymbol{r}, f)$ is a positive weighting function. Usually, the domain $\Omega = \Omega_p \bigcup \Omega_s$ is consisting of the passband region Ω_p and stopband region Ω_s. Then, the beamformer designing for a fixed microphone array is just the filter coefficients optimization problem

$$\min_{\boldsymbol{w} \in \mathbb{R}^{N \times L}} E(\boldsymbol{w}). \tag{7.2}$$

Let $\boldsymbol{\lambda} = (\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_M) \in \mathbb{R}^{3 \times M}$ be the set of the microphones' location, for the given filter length L, we can find the corresponding optimal coefficients $\boldsymbol{w}$ of the FIR filters by solving the above optimization model (7.2) at the fixed microphone array placement $\boldsymbol{\lambda}$. It is well known that the longer filter length L, the better performance, and the best performance will be achieved at the infinite length tending to infinity. However, setting the filter length $L = +\infty$ for the optimization model (7.2) is impractical, to get the performance limit of a fixed microphone array configuration, we introduce the infinite length technique next.

For the $L-$tap FIR filter, the desired response $G_d(\boldsymbol{r}, f, L)$ will have a delay term $\tau_L \in [0, L-1]$, and the total time delay of the microphone array system is $\tau_L T$, where $T = 1/f_s$. Then the desired response can be written as

$$G_d(\boldsymbol{r}, f, L) = e^{-j2\pi f \tau_L T} \hat{G}_d(\boldsymbol{r}, f), \tag{7.3}$$

where $\hat{G}_d(\boldsymbol{r}, f)$ is a response function independent of filter length L. The term of $G_d(\boldsymbol{r}, f, L)$ with respect to the filter length L is just the group delay τ_L, substituting it into the objective function (7.1), we have

$$\begin{aligned} E(\boldsymbol{w}) &= \frac{1}{|\Omega|} \int_{\Omega} \rho(\boldsymbol{r}, f) \left| \boldsymbol{H}^H(\boldsymbol{r}, f) \boldsymbol{W}(\boldsymbol{w}, f, L) - e^{-j2\pi f \tau_L T} \hat{G}_d(\boldsymbol{r}, f) \right|^2 d\boldsymbol{r} df, \\ &= \frac{1}{|\Omega|} \int_{\Omega} \rho(\boldsymbol{r}, f) \left| \boldsymbol{H}^H(\boldsymbol{r}, f) \hat{\boldsymbol{W}}(\boldsymbol{w}, f, L) - \hat{G}_d(\boldsymbol{r}, f) \right|^2 d\boldsymbol{r} df, \end{aligned} \tag{7.4}$$

where

$$\hat{\boldsymbol{W}}(\boldsymbol{w}, f, L) = e^{j2\pi f \tau_L T} \boldsymbol{W}(\boldsymbol{w}, f, L),$$

and $\boldsymbol{W}(\boldsymbol{w}, f, L)$ is the frequency response defined as (2.6).

For the objective function (7.4), we note that only $\hat{\boldsymbol{W}}(\boldsymbol{w}, f, L)$ changes with the filter length's changing, and all the frequency responses $\{\hat{W}_i(\boldsymbol{w}, f, L),\ i = 1, 2, ..., M\}$ are exactly in the space spanned by the basis

$$\Delta_L = \{e^{\frac{-j2\pi f}{f_s}(-\tau_L)}, e^{\frac{-j2\pi f}{f_s}(1-\tau_L)}, ..., e^{\frac{-j2\pi f}{f_s}(L-1-\tau_L)}\}, \tag{7.5}$$

in other word, all of them can be represented linearly by the basis Δ_L. If there are two different filter lengths L_1 and L_2, the corresponding bases Δ_{L_1} and Δ_{L_2} maybe different, however, when the filter length L increasing to $+\infty$, the space spanned by the corresponding basis $\Delta_{+\infty}$ should be the whole space containing all the space spanned by the basis Δ_L with L is finite.

Then, the infinite length idea is to extract the expression limit of the frequency response $\{\hat{W}_i(\boldsymbol{w}, f, L),\ i = 1, 2, ..., M\}$ at the filter length $L = +\infty$ equivalently, we introduce the following lemma developed in [32].

Lemma 7.1. *Suppose that*

$$\lim_{L\to+\infty} \tau_L = +\infty,\ \textit{and}\ \lim_{L\to+\infty} L - \tau_L = +\infty,$$

and $2\tau_L$ *is an integer, for any complex-valued function* $u(f) + jv(f)$ *defined in* $[0, f_s/2]$*, where* $v(0) = 0$ *and* $v(f_s/2) = 0$*, if* $u(f)$ *and* $v(f)$ *are continuous, absolute integrable, and the right-hand and left-hand derivative exist, then, there exists a real sequence* $\{c_k, k = 0, 1, ..., +\infty\}$ *such that*

$$u(f) + jv(f) = \lim_{L\to+\infty} \sum_{k=0}^{L-1} c_k e^{\frac{-j2\pi f}{f_s}(k-\tau_L)}. \tag{7.6}$$

Proof. The number $2\tau_L$ is an integer means that τ_L is an integer or an integer plus 0.5. First, in the case when τ_L is an integer, by the conditions (7.6), (7.6) is

equivalent to

$$u(f) + jv(f) = \sum_{k=-\infty}^{+\infty} c_k e^{\frac{-j2\pi kf}{f_s}}. \tag{7.7}$$

Define an even function $\hat{u}(f)$ as

$$\hat{u}(f) \quad = \quad \begin{cases} u(f), & \text{if } f \in [0, f_s/2] \\ u(-f), & \text{if } f \in [-f_s/2, 0] \end{cases}$$

and an odd function $\hat{v}(f)$ as

$$\hat{v}(f) \quad = \quad \begin{cases} v(f), & \text{if } f \in [0, f_s/2] \\ -v(-f), & \text{if } f \in [-f_s/2, 0] \end{cases}.$$

Then, $\hat{u}(f)$ and $\hat{v}(f)$ are continuous, absolute integrable, and the right-hand and left-hand derivatives exist. Since in Fourier analysis, the set of the functions

$$\{1, \cos(2\pi kf/f_s), \sin(2\pi kf/f_s),\ k = 1, \ldots, +\infty\}$$

are orthogonal and complete in $[-f_s/2, f_s/2]$, the Fourier series of $\hat{u}(f)$ and $\hat{v}(f)$ are given by

$$\hat{u}(f) = a_0 + \sum_{k=1}^{+\infty} a_k \cos(\frac{2\pi kf}{f_s}), \quad \hat{v}(f) = \sum_{k=1}^{+\infty} b_k \cos(\frac{2\pi kf}{f_s}). \tag{7.8}$$

Define c_k in (7.7) as

$$c_0 = a_0, \quad c_k \quad = \quad \begin{cases} \dfrac{a_k - b_k}{2}, & \text{if } k > 0 \\ \dfrac{a_k + b_k}{2}, & \text{if } k < 0 \end{cases}. \tag{7.9}$$

Then, by (7.7), we can verify that the c_k, which is defined in (7.9), satisfies the equation (7.8).

Similarly, we can also prove the case when τ_L is an integer plus 0.5, where the orthogonal and complete system is replaced by

$$\{\cos(2\pi(k+0.5)f/f_s), \sin(2\pi(k+0.5)f/f_s),\ k=1,\ldots,+\infty\}.$$

This completes the proof. □

For all the $u_i(f)$ and $v_i(f), i = 1, 2, ..., M$, satisfy the conditions proposed in Lemma 7.1, defined

$$\Gamma = \{\boldsymbol{u}(f) + j\boldsymbol{v}(f) = (u_i(f)) + j(v_i(f)), i = 1, 2, ..., M\}, \tag{7.10}$$

with the help of Lemma 7.1, we replace the filter frequency responses by

$$\tilde{\boldsymbol{W}}(\boldsymbol{w}, f, L) = \tilde{\boldsymbol{W}}(f) = \boldsymbol{u}(f) + j\boldsymbol{v}(f). \tag{7.11}$$

Substituting it to the objective function (7.4), we can establish the infinite length beamformer designing problem

$$\min_{\tilde{\boldsymbol{W}} \in \Gamma} E(\tilde{\boldsymbol{W}}), \tag{7.12}$$

where

$$E(\tilde{\boldsymbol{W}}) = \frac{1}{|\Omega|} \int_{\Omega} \rho(\boldsymbol{r}, f) \left| \boldsymbol{H}^H(\boldsymbol{r}, f)\tilde{\boldsymbol{W}}(f) - \hat{G}_d(\boldsymbol{r}, f) \right|^2 d\boldsymbol{r}df. \tag{7.13}$$

Notice that the performance limit is estimated for a given placement of the microphone array. Different placement can cause different performance limit. Then the placement design is to find the placement of the microphone array such that the corresponding performance limit is minimized. For this, let $\boldsymbol{\lambda} = (\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_M) \in \boldsymbol{\Lambda} \subset \mathbb{R}^{3\times M}$ denote the feasible region of the microphone array, where Λ is the set of all possible $\boldsymbol{\lambda}$. Then, the placement design problem is formulated as

$$\begin{array}{ll} \min\limits_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}, \tilde{\boldsymbol{W}}^M \in \Gamma^M} & E(\boldsymbol{\lambda}, \tilde{\boldsymbol{W}}) \\ s.t & \|\boldsymbol{r}_i - \boldsymbol{r}_j\|^2 \geqslant \varepsilon_d,\ i, j = 1, 2, .., M,\ i \neq j, \end{array} \tag{7.14}$$

where

$$E(\boldsymbol{\lambda}, \tilde{\boldsymbol{W}}) = \frac{1}{||\Omega||} \int_{\Omega} \rho(\boldsymbol{r}, f) ||\boldsymbol{H}^H(\boldsymbol{\lambda}, \boldsymbol{r}, f) \tilde{\boldsymbol{W}}(f) - \hat{G}_d(\boldsymbol{\lambda}, \boldsymbol{r}, f)||^2 d\boldsymbol{r} df \tag{7.15}$$

is the same as (7.13), but considering $\boldsymbol{\lambda}$ as the new decision vector.

The constraints $||\boldsymbol{r}_i - \boldsymbol{r}_j||^2 \geqslant \varepsilon_d,\ i, j = 1, 2, .., M,\ i \neq j$, are according with the reality, the set of microphone elements should be kept at least a certain minimum distance mutually for proper functioning, and ε_d is the square of the minimum distance between two different microphone elements.

7.2 Hybrid Descent Method

In general, the problem (7.14) is very complicated. It contains two kinds of decision vectors: placement vector and frequency response vector. Note that for a given placement vector, the optimal frequency response can be obtained by solving the quadratic problem (7.12), which can be solved very quickly by quadratic programming method, we transformed the placement design problem (7.14) into

$$\begin{array}{ll} \min\limits_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} & F(\boldsymbol{\lambda}) = E(\boldsymbol{\lambda}, \tilde{\boldsymbol{W}}^*(\boldsymbol{\lambda})) \\ s.t & ||\boldsymbol{r}_i - \boldsymbol{r}_j||^2 \geqslant \varepsilon_d,\ i, j = 1, 2, .., M,\ i \neq j, \end{array} \tag{7.16}$$

where $\tilde{\boldsymbol{W}}^*(\boldsymbol{\lambda})$ is the solution of the subproblem (7.12) with the given placement vector $\boldsymbol{\lambda}$.

Although discrete approach can be applied to transform the above semi-infinite programming problem into a large-scale constrained optimization problem similarly in dealing with problem (7.12), and the difference between them is just the additional placement variables $\boldsymbol{\lambda}$, the intrinsic characters of them are very different. The objective function (7.13) is convex with respect to the variables $\tilde{\boldsymbol{W}}$, and problem (7.12) is a convex optimization problem, most of the discrete convex optimization tools can be

used to solve it. The objective function (7.15), however, has the variables $\boldsymbol{\lambda}$ nested inside $\boldsymbol{H}(\boldsymbol{\lambda}, \boldsymbol{r}, f)$ and $\hat{G}_d(\boldsymbol{\lambda}, \boldsymbol{r}, f)$, it is nonconvex with respect to $\boldsymbol{\lambda}$, then problem (7.14) is nonconvex, and there is still no method can be used to solve the nonconvex optimization problem commonly.

The combined nonconvex optimization problem (7.14) has two kinds of variables, it is difficult to solve as a whole, note that the objective function (7.15) is nonconvex with respect to the placement variables $\boldsymbol{\lambda}$, but it is convex with respect to the filter coefficient variables $\tilde{\boldsymbol{W}}$. We can choose a placement $\boldsymbol{\lambda}$ at the first step, reduce original problem to a convex subproblem (7.12), then it can be solved easier. Next, we change the placement, resolve a subproblem (7.12) again, and repeat it until converges. Since problem (7.12) can be solved accurately and effectively, the main problem we need to do is to find an approach to update the placement effectively.

As discussed previously, the placement optimization problem (7.14) is not convex, difficult to be solved by gradient-based approach directly. But the variables have the property that the infinite length filter coefficient variables are determined by the placement, it is easy to get the idea that solving the problem (7.14) alternatively. In the first step, we adjust the placement variables independently, this will generate a subproblem (7.12) and determine the performance limit of the beamformer, then solve it for verifying the effect of this placement, and repeat it. We propose a hybrid descent method in the following.

Algorithm 7.2.1. Hybrid descent method

1. *Give two positive parameters ε_1 and ε_2, generate an initial placement point $\boldsymbol{\lambda}^0$, and solve the reduced optimization problem* (7.14) *with $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0$, get the optimal objective function value as $E(\boldsymbol{\lambda}^0)$. Set $k = 0$.*

2. *Take $\boldsymbol{\lambda}^k$ as one of the candidate point for the genetic algorithm, and execute*

N times of iterative computation until a point $\boldsymbol{\lambda}^{\bar{k}}$ *is obtained, such that the objective function has a certain degree of decline,* $E(\boldsymbol{\lambda}^{\bar{k}}) - E(\boldsymbol{\lambda}^{k}) \leqslant -\varepsilon_1$.

3. *Solve for the local minimum of* $E(\boldsymbol{\lambda}, \tilde{\boldsymbol{W}})$ *by using a gradient-based minimization method with* $\boldsymbol{\lambda}^{\bar{k}}$ *as the input point to get* $\boldsymbol{\lambda}^{k+1}$*, such that the objective function has a certain degree of decline,* $E(\boldsymbol{\lambda}^{k+1}) - E(\boldsymbol{\lambda}^{\bar{k}}) \leqslant -\varepsilon_2$.

4. *Set* $k := k + 1$*, return to Step* **2** *until convergence.*

In the Step **2** of the Algorithm (7.2.1), the fitness function value for genetic algorithm is set to be the optimal value of the subproblem (7.12), the genetic algorithm composes of five key steps, namely

1. Population representation – The real-valued placement variables $\boldsymbol{\lambda}$ are initialized using crtrp to construct the chromosomes, which storing an entire population in a single matrix with all chromosomes are of equal length.

2. Fitness assignment – The fitness values are derived from the optimal value of (7.14) through a ranking or scaling function.

3. Selection – Selection functions (reins, rws, *et al.*) select a given number of placements from the current population, according to their fitness, and return a column vector to their indices.

4. Crossover – Crossover operators (recdis, recint, *et al.*) recombine pairs of individuals with given probability to produce offspring.

5. Mutation – Mutation operators (mut, mutate, *et al.*) apply random changes to individual parents to form offspring.

In this paper, the functions are used the MATLAB default setting on the genetic algorithm, and the the major procedures of it are outlined as follows:

Genetic procedures

(a) Generate the initial placement chromosomes $\boldsymbol{\Lambda}^{k}$ with inputting the last local optimal placement $\boldsymbol{\lambda}^{k}$, evaluate (7.14) for all the individuals, get $E(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}^{k}$.

(b) Rank $E(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}^k$, if $E(\boldsymbol{\lambda}^{\bar{k}}) - E(\boldsymbol{\lambda}^k) \leqslant -\varepsilon_1$, stop; otherwise, go to next Step (c);

(c) Select the individuals $\boldsymbol{\Lambda}_1^k$ by using the selection functions, called parents;

(d) Combine two parents to form offspring $\boldsymbol{\Lambda}_2^k$ for the next generation by using the crossover operators;

(e) Apply random changes to individual parents to form offspring $\boldsymbol{\Lambda}_3^k$ by using mutation operators;

(f) Evaluate (7.14) for all the new individuals $\boldsymbol{\Lambda}_3^k$, get $E(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}_3^k$, and return to Step (b).

7.3 Illustration Examples

In our numerical experiments, the desired response function is specified over a region that would fit into a multimedia or hands free mobile phone application, including the frequency range of human voice, and a range of position that microphone elements should be directed towards. To allow for the delay of the speech to reach the microphones, the desired response function $G_d(\boldsymbol{r}, f, L)$ in the passband and stopband is defined the same as (2.14). The sampling rate is set as $8kHz$, and the maximum frequency is chosen as $f_{max} = 4kHz$, the minimum distance parameter between two different microphone elements is set as

$$\varepsilon_d = 0.015^2 m^2,$$

and the weighting function is chosen as $\rho(\boldsymbol{r}, f) = 1$.

7.3.1 2-D Microphone Array Placement Problem

We firstly consider a placement configuration problem for 2-D case, fix the microphone array with nine elements, the microphone elements and the beamforming

desired spatial region are defined on two planes respectively, and the microphone elements chosen plane is 1 meter away from the beamforming desired spatial plane, see Figure 7.1 below.

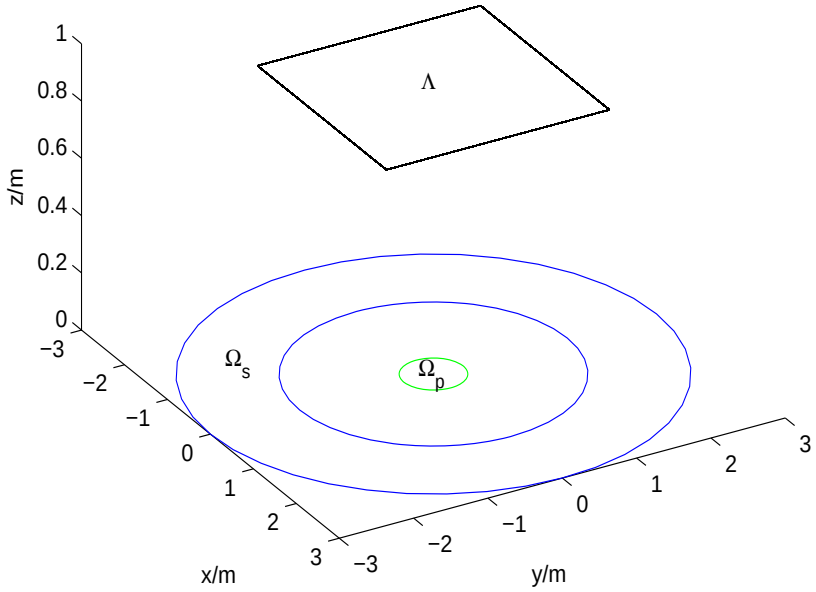


Figure 7.1: Configuration of Example 7.3.1.

From the above configuration, the beamforming desired spatial plane is defined on $z = 0m$, and the microphone elements chosen plane is fixed on $z = 1m$, the specific feasible region $\mathbf{\Lambda}$ of microphone elements, bandpass region Ω_p and band stop region Ω_s of beamforming are defined specifically as following.

Passband region:

$$\Omega_p = \{(\boldsymbol{r}, f) \mid ||(x, y)|| \leqslant 0.4m,\ z = 0m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\}.$$

Stopband region:

$$\Omega_s = \{(\boldsymbol{r}, f) \mid ||(x, y)|| \leqslant 0.4m,\ z = 0m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\} \ldots$$

$$\cup\{(\boldsymbol{r}, f) \mid 1.8m \leqslant ||(x, y)|| \leqslant 3.0m,\ z = 0m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\} \ldots$$
$$\cup\{(\boldsymbol{r}, f) \mid 1.8m \leqslant ||(x, y)|| \leqslant 3.0m,\ z = 0m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\}.$$

Placement feasible region:

$$\boldsymbol{\Lambda} = \{\boldsymbol{\lambda} \mid |x| \leqslant 1.5m,\ |y| \leqslant 1.5m,\ z = 1m\}.$$

For the discretization of $\Omega = \Omega_p \bigcap \Omega_s$, each of the frequency domain regions is generated 60 points, and the spatial domain regions are taken every $0.2m$. By applying the proposed Algorithm 7.2.1, we can find the optimal placement $\boldsymbol{\lambda}^*$ of the microphone elements at

$$\begin{aligned}\boldsymbol{\lambda}^* \ = \ & \{(0,0,1),\ (0.1693,0.0700,1),\ (0.0703,0.1692,1),\\ & (-0.0700,0.1693,1),\ (-0.1692,0.0703,1),\ (-0.1693,-0.0700,1),\ldots\\ & (-0.0704,-0.1692,1),\ (0.0700,-0.1693,1),\ (0.1692,-0.0703,1)\},\end{aligned}$$

which is plotted in Figure 7.2. Using this placement configuration, we can get the optimal objective function value under the infinite length model is -39.1635dB.

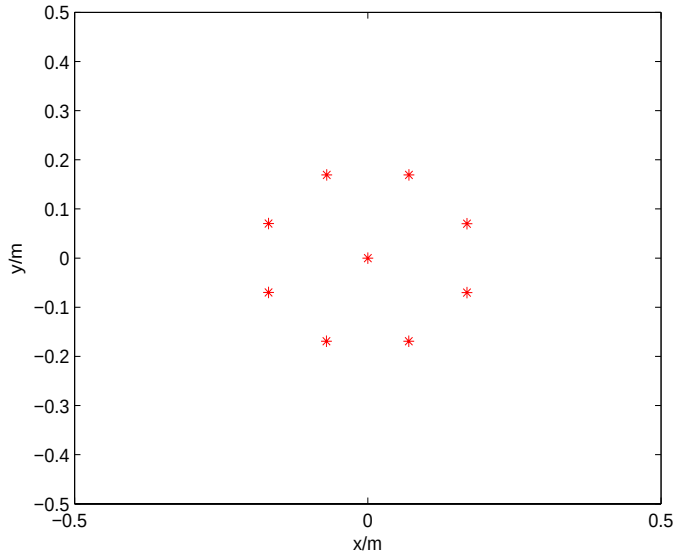


Figure 7.2: Optimal placement for Example 7.3.1.

To verify whether the placement $\boldsymbol{\lambda}^*$ is the best one or not, we compare it with four common placement structures (see Figure 7.3). For the former three regular structures, we fix the distance between every two microphones is $0.2m$, and define the radius as $0.2m$ to generate 9 microphone points for the fourth placement structure. Using these placement configurations, the optimal objective function values under the infinite length model in dB are listed in Table 7.1. It is proved that the microphone array will achieve the best performance at the placement $\boldsymbol{\lambda}^*$.

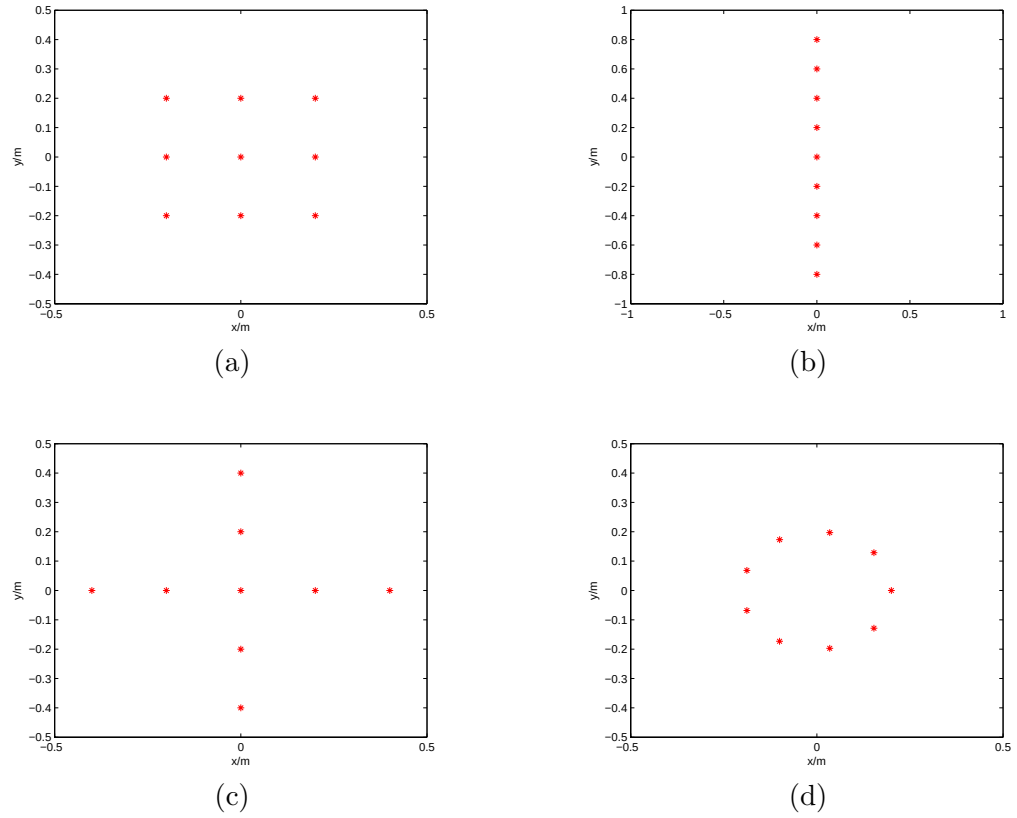


Figure 7.3: Placement of Example 7.3.1.

Taking the filter length $L = 40$, we can obtain the optimal filter coefficients and the actual response in the above microphone array placement $\boldsymbol{\lambda}^*$. To verify the effect of the beamforming, we choose an actual response performance of $X - Y$ plane at frequency $1400Hz$ and another performance of $X - F$ plane (spatial-frequency plane,

Table 7.1: Performance of common structures.

Placement	(a)	(b)	(c)	(d)
Performance limit (in dB)	-30.2153	-19.4577	-24.4524	-20.5579

$y = 0m$) in the following Figure 7.4.

7.3.2 3-D Microphone Array Placement Problem

In the second example, we consider a 3-D microphone array placement design problem, it is the spatial region and the microphone elements placement are all be setup in tridimensional space. We also consider nine microphone elements, but the microphone elements is chosen in an solid with 1 meter away from the beamforming desired cubic space, see Figure 7.5 below.

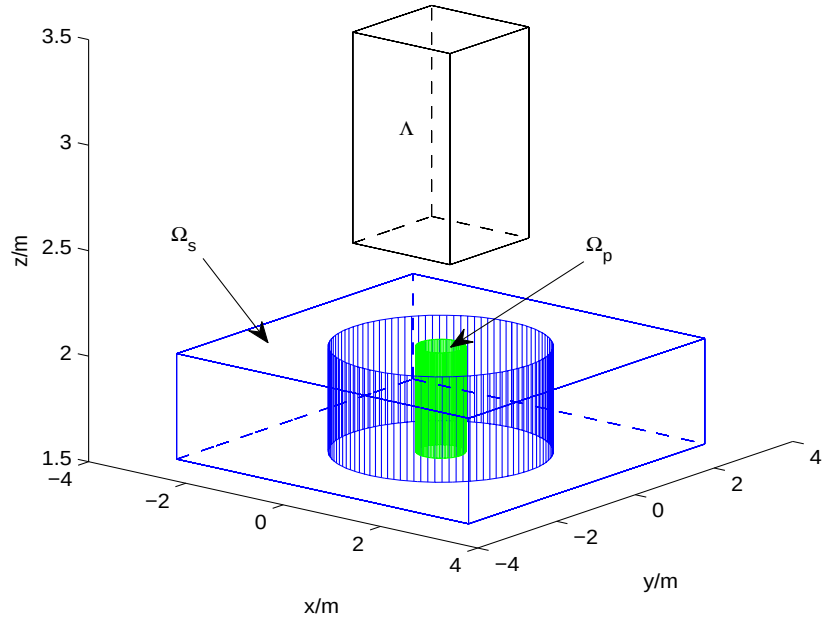
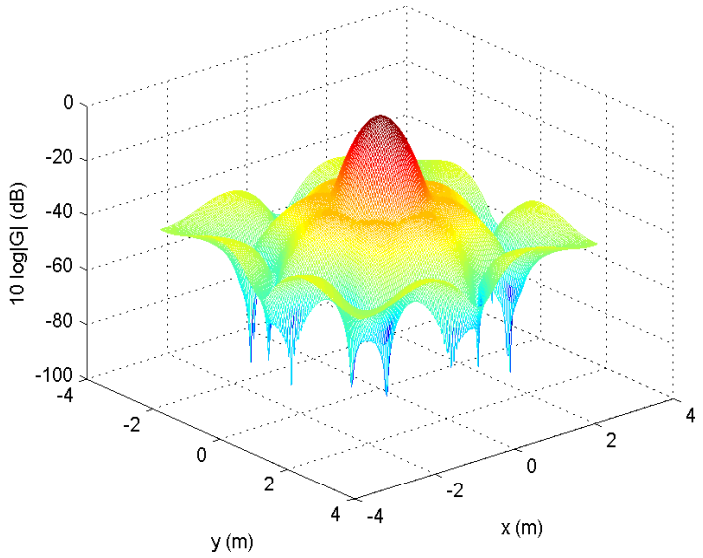
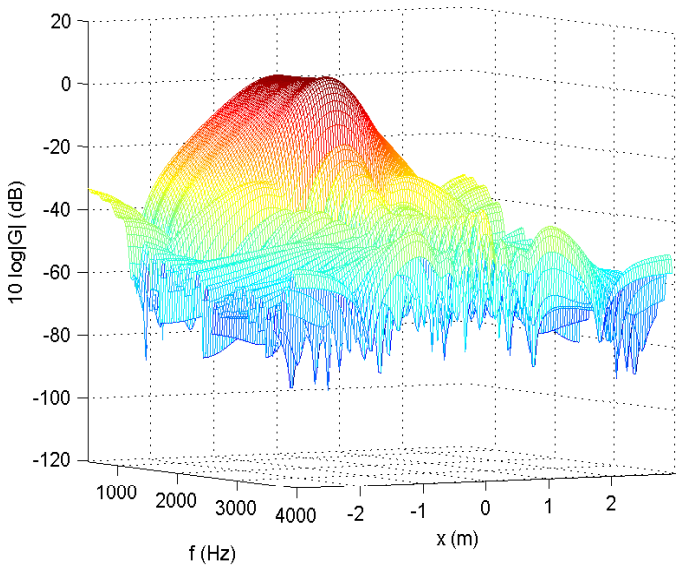


Figure 7.5: Optimal placement for Example 7.3.2.

The specific feasible region $\mathbf{\Lambda}$ of microphone elements, bandpass region Ω_p and band stop region Ω_s of beamforming are described as



(a)



(b)

Figure 7.4: Performance of finite length filter under optimal placement.

Passband region:

$$\Omega_p = \{(\boldsymbol{r}, f) \mid ||(x,y)|| \leqslant 0.4m,\ 1.5m \leqslant |z| \leqslant 2.0m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\}.$$

Stopband region:

$$\begin{aligned}\Omega_s = &\{(\boldsymbol{r}, f) \mid ||(x,y)|| \leqslant 0.4m,\ 1.5m \leqslant |z| \leqslant 2.0m,\ 2.0kHz \leqslant f \leqslant 4.0kHz\} \ldots \\ &\cup\{(\boldsymbol{r}, f) \mid 1.8m \leqslant ||(x,y)|| \leqslant 3.0m,\ 1.5m \leqslant |z| \leqslant 2.0m,\ 0.5kHz \leqslant f \leqslant 1.5kHz\} \ldots \\ &\cup\{(\boldsymbol{r}, f) \mid 1.8m \leqslant ||(x,y)|| \leqslant 3.0m,\ 1.5m \leqslant |z| \leqslant 2.0m, 2.0kHz \leqslant f \leqslant 4.0kHz\}.\end{aligned}$$

Placement feasible region:

$$\boldsymbol{\Lambda} = \{\boldsymbol{\lambda} \mid |x| \leqslant 1m,\ |y| \leqslant 1m,\ 2.5m \leqslant |z| \leqslant 3.5m\}.$$

The frequency domain and the spatial domain regions are discretized similarly to Example 7.3.1, by applying the proposed Algorithm 7.2.1, we can find the optimal placement $\boldsymbol{\lambda}^*$ of the microphone elements at

$$\begin{aligned}\boldsymbol{\lambda}^* \ = \ &\{(0,0,2.5),\ (0,0,2.883400836),\ (0,0,2.942937691),\ldots \\ &(0,0,3.125292923),\ (0,0,3.24250263),\ (0,0,3.374356638),\ldots \\ &(0,0,3.456959832),\ (0,0,3.472801381),\ (0,0,3.5)\},\end{aligned}$$

which is plotted in Figure 7.6. Using this placement configuration, we can get the optimal objective function value under the infinite length model is -34.4140dB.

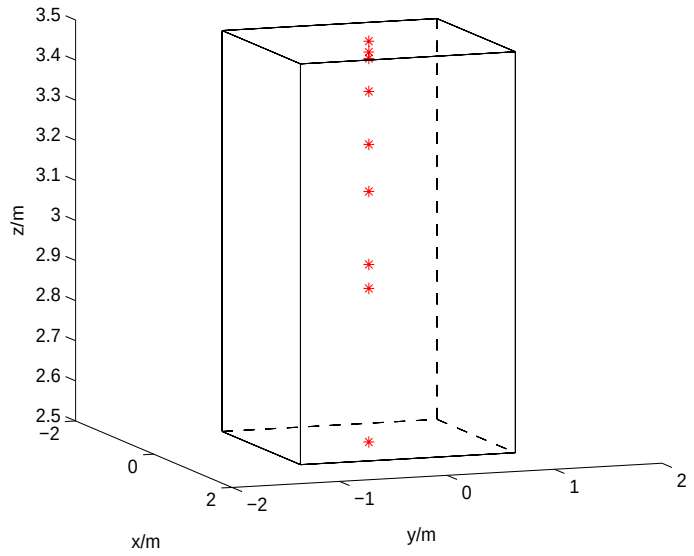
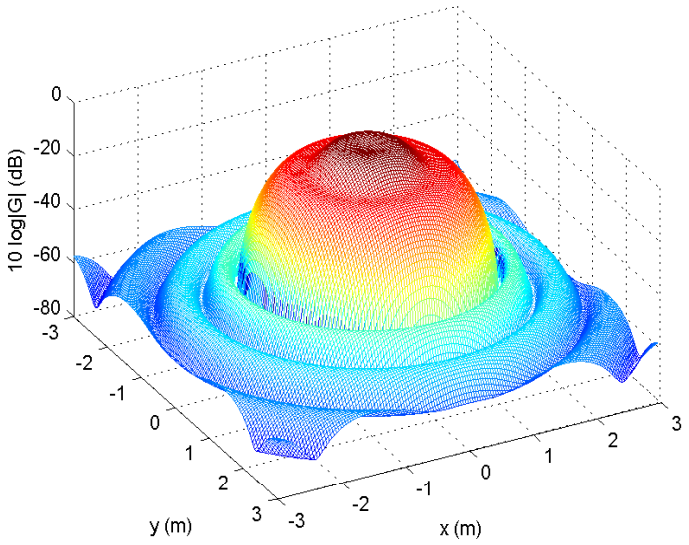
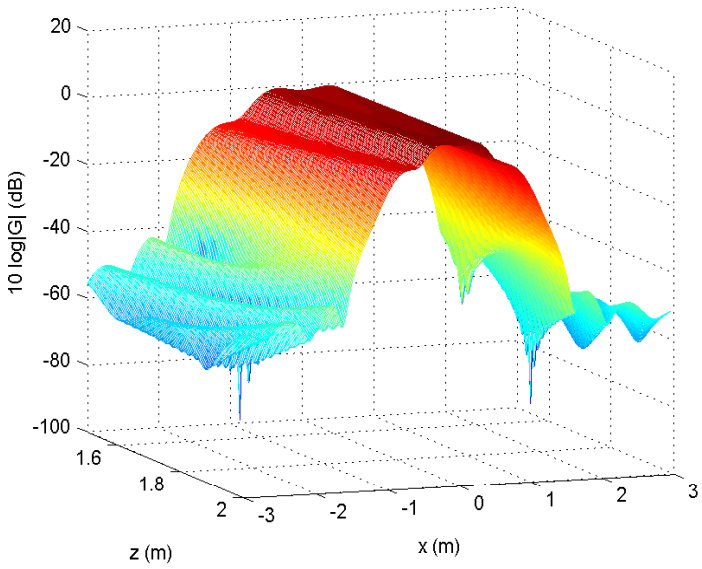


Figure 7.6: Optimal placement for Example 7.3.2.

Taking the filter length $L = 50$, we can obtain the optimal filter coefficients and the actual response in the above microphone array placement $\boldsymbol{\lambda}^*$. To verify the effect of the beamforming, we choose an actual response performance of $X - Y$ plane at frequency $1400Hz$, $z = 1.6m$, and another performance of $X - Z$ plane at frequency $1400Hz$, $y = 0m$ in the following Figure 7.7.



(a)



(b)

Figure 7.7: Performance of Example 7.3.2.

Chapter 8

Indoor Microphone Array Placement Design

The previous Chapter 7 has studied the effective method for microphone array placement design problem in the open air, whereas in reverberant environment, the room acoustics make the transfer function estimation for sound wave propagation more complicated and costly. And it is hard to design the broadband beamformer with the effective function on continuous spatial regions. In general, the LCMV beamformer consider some specific targets for signal of interest (SOI) capturing and interference (INT) suppression, and it can achieve good performance on noise (NOI) reduction.

In this chapter, we study the microphone array placement design problem in acoustic room. We employ the common LCMV beamformer design model to achieve dereverberation and interference suppression by obtaining the desired beam pattern in linear constraints, and reduce the background noise with the estimation noise power minimization, where the RIRs are estimated by the ISM based room simulator. We also introduce the infinite length technique to pursuit performance limit of the beamformer without respect to filter length. Thus, we can establish the microphone array placement design problem based on the error between beamformer output and the given desired signal. Then we apply the hybrid descent method with genetic algorithm to solve the placement vector.

8.1 Problem Formulation

Given an auditory scene (such as a rectangular acoustic room), where N number of sources are settled at $\boldsymbol{r}_n$, $n = 0,\ ....,\ N-1$, without loss of generality, the first one $\boldsymbol{r}_0$ can be denoted as the signal of interest (SOI), the others are interferences (INT) and background noise (NOI) (where the noise placement information is not used in beamformer design). Then for a given M-elements microphone array settled at $\boldsymbol{\lambda} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_M) \in \boldsymbol{\Lambda} \subset \mathbb{R}^{3\times M}$, assume the sound wave propagation between source number n and array element number m can be modeled by a transfer function $H_{n,m}(\boldsymbol{\lambda}, \omega)$ (RIR in time domain). Thus, in the frequency domain, microphone array captured signals can be denoted as

$$X_m(\omega) = \sum_{n=0}^{N-1} H_{n,m}(\omega) S_n(\omega) + V_m(\omega), \quad m = 1, \ldots, M. \tag{8.1}$$

Assume there is an $L-$tap FIR filter behind each microphone element with coefficients $w_m = [w_m(0), w_m(1), \ldots, w_m(L-1)]^T$, $m = 1, \ldots, M$. If the signals received by this microphone array are sampled synchronously at the rate of f_s per second, then the frequency responses of these FIR filters for frequency component ω can be defined as

$$W_m(\omega) = w_m^T d_0(\omega), \quad m = 1, \ldots, M, \tag{8.2}$$

where $d_0(\omega)$ is defined as

$$d_0(\omega) = [e^{\frac{-j\omega}{f_s}(-\tau_L)}, e^{\frac{-j\omega}{f_s}(1-\tau_L)}, ..., e^{\frac{-j\omega}{f_s}(L-1-\tau_L)}]^T, \tag{8.3}$$

and $0 \leqslant \tau_L \leqslant L-1$ is the delay term (such as $\tau_L = (L-1)/2$).

Then the beamformer output under such an array placement $\boldsymbol{\lambda}$ can be formed in each frequency band

$$Y(\omega) = \sum_{m=1}^{M} W_m(\omega) X_m(\omega) = \boldsymbol{W}^H(\omega) \boldsymbol{X}(\omega), \tag{8.4}$$

where $\boldsymbol{W}(\omega) = [W_1(\omega)\ \ldots\ W_M(\omega)]^T$ is the beamformer response vector and $\boldsymbol{X}(\omega) = [X_1(\omega)\ \ldots\ X_M(\omega)]^T$ is the input data vector.

Notice that this beamformer output $Y(\omega)$ is with respect to the microphone array placement $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w} = [w_1, w_2,\ \ldots,\ w_M]^T$, to measure the error between it and the desired SOI signal $S_d(\omega)$, we can use the L_2-norm criterion to define the objective function

$$\boldsymbol{F}(\boldsymbol{\lambda}, \boldsymbol{w}) = \frac{1}{||\Omega||_2} \int_{\Omega} ||\boldsymbol{W}^H(\omega)\boldsymbol{X}(\omega) - S_d(\omega)||_2^2 d\omega, \tag{8.5}$$

where Ω is the considering frequency domain. Thus the microphone array placement problem based on the above cost function (8.5) can be established as:

Definition 8.1. *Find optimal solutions of placement $\boldsymbol{\lambda}_{opt}$ in some feasible region $\boldsymbol{\Lambda}$ and the corresponding filter coefficients $\boldsymbol{w}$, make sure the beamformed signal $Y(\omega)$ satisfy $\boldsymbol{F}(\boldsymbol{\lambda}_{opt}, \boldsymbol{w}_{opt}) \leqslant \boldsymbol{F}(\boldsymbol{\lambda}, \boldsymbol{w})$ for all $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \boldsymbol{w} \in \mathbb{R}^{M \times L}$. The equivalent optimization problem is*

$$\begin{array}{ll} \min\limits_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \boldsymbol{w} \in \mathbb{R}^{M \times L}} & \boldsymbol{F}(\boldsymbol{\lambda}, \boldsymbol{w}) \\ s.t. & ||\boldsymbol{r}_i - \boldsymbol{r}_j||^2 \geqslant \varepsilon_0,\ i, j = 1, 2, \ldots, M,\ i \neq j, \end{array} \tag{8.6}$$

where $\varepsilon_0 > 0$ is a constant, which is used to make sure the set of microphone elements are kept at least a certain minimum distance mutually for proper functioning.

The above microphone array placement problem (8.6) is a composite optimization problem with respect to the placement variable $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w}$. However, the filter coefficients $\boldsymbol{w}$ is usually determined by the specific placement $\boldsymbol{\lambda}$, for each selection of $\boldsymbol{\lambda}$, determining the corresponding beamformer output $Y(\omega)$ will be a subproblem. Generally, the beamformer output $Y(\omega)$ is affected by the filter length L, and solving the optimal filter coefficients is also a challenge work at the moment, thus in this paper, we introduce the infinite length technique to estimate the optimal beamformer output $Y(\omega)$ for each of the subproblem in (8.6).

8.2 Performance Limit of LCMV Beamformer

For a given microphone array placement $\boldsymbol{\lambda}$, it is noted that the L-tap beamformer coefficient w_m is corresponding to the frequency response $W_m(\omega)$, $m = 1, \ldots, M$, and the optimal frequency response $W_{m,opt}(\omega)$ will be obtained when the filter length $L \to +\infty$. This property can also be proved by the previous Lemma 7.1.

Based on the Lemma 7.1, we can define the frequency response variables $W_m(\omega) = u_m(\omega) + jv_m(\omega)$ for our beamformer subproblems in the following. Substituting the input signals $X_m(\omega)$ with different channels (8.1), we can rearrange the beamformer output to

$$\begin{aligned} Y(\omega) = & \sum_{m=1}^{M} W_m(\omega) H_{0,m}(\omega) S_0(\omega) + \sum_{n=1}^{N-1} \sum_{m=1}^{M} W_m(\omega) H_{n,m}(\omega) S_n(\omega) \\ & + \sum_{m=1}^{M} W_m(\omega) V_m(\omega), \end{aligned} \tag{8.7}$$

Then the beamformer design subproblem is equivalent to find the optimal $W_m(\omega)$ so that the first part of the above sum construct our desired signal, while the last two parts of the sum vanish. The perfect performance on dereverberation, interference suppression and noise reduction needs the designed beamforming filter to satisfy the following conditions

$$\begin{aligned} \sum_{m=1}^{M} W_m(\omega) H_{0,m}(\omega) &= G_D(\omega), \\ \sum_{n=1}^{N-1} \sum_{m=1}^{M} W_m(\omega) H_{n,m}(\omega) &= 0, \\ \sum_{m=1}^{M} W_m(\omega) V_m(\omega) &= 0, \end{aligned} \tag{8.8}$$

where $G_D(\omega)$ denotes the direct path transfer function between the SOI point and beamformer output point. Thus, in LCMV beamforming, the frequency responses

$W_m(\omega)$ are adjusted based on the statistics of the sensor array measured signals, the expression to be minimized is the power of the background noise, and the cost function may be defined as

$$\boldsymbol{E}(\boldsymbol{W}(\omega)) = ||\boldsymbol{W}^H(\omega)\boldsymbol{V}(\omega)||^2, \tag{8.9}$$

where $\boldsymbol{V}(\omega) = [V_1(\omega)\ V_2(\omega)\ \ldots\ V_M(\omega)]^T$ is the noise inputting vector. To achieve performance on dereverberation and interference suppression, we can establish the frequency domain LCMV beamforming problem as follows

$$\begin{array}{cc} \min\limits_{\boldsymbol{W}(\omega)\in\mathbb{C}^M} & \boldsymbol{E}(\boldsymbol{W}(\omega)) \\ s.t. & \boldsymbol{H}(\omega)\boldsymbol{W}(\omega) = \boldsymbol{G}(\omega), \end{array} \tag{8.10}$$

where the constraint matrix $\boldsymbol{H}(\omega)$ is constructed by the corresponding to conditions (8.8), which is defined as

$$\boldsymbol{H}(\omega) = \begin{pmatrix} H_{0,1}(\omega) & H_{1,1}(\omega) & \cdots & H_{N-1,1}(\omega) \\ H_{0,2}(\omega) & H_{1,2}(\omega) & \cdots & H_{N-1,2}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{0,M}(\omega) & H_{1,M}(\omega) & \cdots & H_{N-1,M}(\omega) \end{pmatrix}, \tag{8.11}$$

and $\boldsymbol{G}(\omega) = [G_D(\omega)\ 0]^T$ is the response vector.

The problem (8.10) provides a general framework on the design of LCMV beamformer in frequency domain, it is defined on complex field with the complex variables $W_m(\omega)$, we can not use this model to solve the these $W_m(\omega),\ m = 1, \ldots, M$ directly. To formulate the specific real field optimization model for infinite length LCMV beamforming subproblem (8.10), we separate the necessary conditions (8.8) for complex transfer functions numbers $H_{n,m}(\omega),\ n = 0, 1, \ldots, N-1$, noise components $V_m(\omega)$ and frequency response variables $W_m(\omega)$ with $m = 1, 2, \ldots, M$ into real and

image parts, respectively. By denoting

$$\begin{array}{ll} H_{n,m}^{\mathrm{Re}}(\omega) = \mathrm{real}(H_{n,m}(\omega)), & H_{n,m}^{\mathrm{Im}}(\omega) = \mathrm{imag}(H_{n,m}(\omega)), \\ V_{m}^{\mathrm{Re}}(\omega) = \mathrm{real}(V_{m}(\omega)), & V_{m}^{\mathrm{Im}}(\omega) = \mathrm{imag}(V_{m}(\omega)), \\ W_{m}^{\mathrm{Re}}(\omega) = \mathrm{real}(W_{m}(\omega)), & W_{m}^{\mathrm{Im}}(\omega) = \mathrm{imag}(W_{m}(\omega)), \\ G_{D}^{\mathrm{Re}}(\omega) = \mathrm{real}(G_{D}(\omega)), & G_{D}^{\mathrm{Im}}(\omega) = \mathrm{imag}(G_{D}(\omega)), \end{array}$$

and stacking the real and image parts $W_m^{\mathrm{Re}}(\omega)$ and $W_m^{\mathrm{Im}}(\omega)$ of frequency response into new real variable vector $\chi_m(\omega) = (W_m^{\mathrm{Re}}(\omega)\ W_m^{\mathrm{Re}}(\omega))^T$, we can rewrite the beamforming conditions (8.8) into

$$\begin{array}{l} \sum\limits_{m=1}^{M} \begin{pmatrix} H_{0,m}^{\mathrm{Re}}(\omega) & -H_{0,m}^{\mathrm{Im}}(\omega) \\ H_{0,m}^{\mathrm{Im}}(\omega) & H_{0,m}^{\mathrm{Re}}(\omega) \end{pmatrix} \chi_m(\omega) = \begin{pmatrix} G_{D}^{\mathrm{Re}}(\omega) \\ G_{D}^{\mathrm{Im}}(\omega) \end{pmatrix}, \\ \sum\limits_{n=1}^{N-1} \sum\limits_{m=1}^{M} \begin{pmatrix} H_{n,m}^{\mathrm{Re}}(\omega) & -H_{n,m}^{\mathrm{Im}}(\omega) \\ H_{n,m}^{\mathrm{Im}}(\omega) & H_{n,m}^{\mathrm{Re}}(\omega) \end{pmatrix} \chi_m(\omega) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ \sum\limits_{m=1}^{M} \chi_m(\omega)^T \left(\begin{pmatrix} V_{m}^{\mathrm{Re}}(\omega) \\ -V_{m}^{\mathrm{Im}}(\omega) \end{pmatrix} \begin{pmatrix} V_{m}^{\mathrm{Re}}(\omega) \\ -V_{m}^{\mathrm{Im}}(\omega) \end{pmatrix}^T + \begin{pmatrix} V_{m}^{\mathrm{Im}}(\omega) \\ V_{m}^{\mathrm{Re}}(\omega) \end{pmatrix} \begin{pmatrix} V_{m}^{\mathrm{Im}}(\omega) \\ V_{m}^{\mathrm{Re}}(\omega) \end{pmatrix}^T \right) \chi_m(\omega) = 0. \end{array} \tag{8.12}$$

Then, by stacking all the matrices and vectors together, we can establish the optimization problem for the optimal frequency responses of LCMV beamformer in the field of real number as

$$\begin{array}{ll} \min\limits_{\boldsymbol{\chi}(\omega) \in \mathcal{R}^{2M}} & \boldsymbol{\chi}(\omega)^T \boldsymbol{R}_{VV}(\omega) \boldsymbol{\chi}(\omega) \\ s.t. & \bar{\boldsymbol{H}}(\omega) \boldsymbol{\chi}(\omega) = \bar{\boldsymbol{G}}(\omega), \end{array} \tag{8.13}$$

where $\boldsymbol{\chi}(\omega)$, $\boldsymbol{R}_{VV}(\omega)$, $\bar{\boldsymbol{H}}(\omega)$ and $\bar{\boldsymbol{G}}(\omega)$ are the stacked vectors and matrices corresponding to conditions (8.12).

Thus for each frequency bin ω, the above optimization model (8.13) is a standard constraint quadratic optimization problem, and we can solve the optimal frequency responses $W_{m,opt}(\omega) = \chi_m(\omega)(1) + j\chi_m(\omega)(2)$ for the microphone array by using the effective optimization tools, such as the solution `quadprog` in MATLAB. Then substitute it into (8.4), we will get the limit beamforming output $Y_{opt}(\omega)$ for such a given microphone array placement $\boldsymbol{\lambda}$.

8.3 Hybrid Descent Method

With the help of infinite length technique, the microphone array placement design problem (8.6) with respect to placement variable $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w}$ is converted into a new problem with respect to placement variable $\boldsymbol{\lambda}$ and frequency responses $\boldsymbol{W}(\omega)$,

$$\begin{array}{ll} \min\limits_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda},\boldsymbol{W}(\omega)\in\mathbb{C}^M} & \bar{\boldsymbol{F}}(\boldsymbol{\lambda},\boldsymbol{W}(\omega)) \\ s.t. & ||\boldsymbol{r}_i-\boldsymbol{r}_j||^2 \geqslant \varepsilon_0,\ i,j=1,2,\ldots,M,\ i\neq j. \end{array} \tag{8.14}$$

Moreover, by using the proposed method to solve the optimal frequency responses $\boldsymbol{W}_{opt}(\omega)$ from (8.13), we can reduce the microphone array placement problem (8.14) to

$$\begin{array}{ll} \max\limits_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} & \bar{\boldsymbol{F}}(\boldsymbol{\lambda},\boldsymbol{W}_{opt}(\omega)) \\ s.t. & ||\boldsymbol{r}_i-\boldsymbol{r}_j||^2 \geqslant \varepsilon_0,\ i,j=1,2,\ldots,M,\ i\neq j. \end{array} \tag{8.15}$$

Thus, the microphone array placement design problem can be converted into a subproblem (8.13) for optimal frequency responses $\boldsymbol{W}_{opt}(\omega)$ solving and an optimal placement solving problem (8.15). Moreover, discrete scheme can be applied to transform the above semi-infinite programming problems (8.13) and (8.15) into constrained optimization problems.

For the subproblem (8.13), the objective function of it is convex with respect to the variables $\boldsymbol{W}(\omega)$ and therefore results in a convex optimization problem, and the optimal solution can be solved by optimization tools effectively. Whereas the placement design problem (8.15) has the variables $\boldsymbol{\lambda}$ nested inside $\bar{\boldsymbol{H}}(\omega)$ and $\bar{\boldsymbol{G}}(\omega)$, it is nonconvex with respect to $\boldsymbol{\lambda}$, the problem (8.15) is thus nonconvex, the optimal solution can not be captured by any gradient based method effectively. Thus, a good strategy is required to search for better locations, and the hybrid descent method similar to Algorithm (7.2.1) is applied in the following.

Algorithm 8.3.1.

1. *Generate an initial placement point* $\boldsymbol{\lambda}^0$*, and solve the reduced optimization problem* (8.13) *with* $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0$ *to get the optimal frequency response* $\boldsymbol{W}_{opt}(\omega)$ *for the objective value as* $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}^0, \boldsymbol{W}_{opt}(\omega))$*. Set* $k = 0$*.*

2. *Take* $\boldsymbol{\lambda}^k$ *as one of the candidate point for the genetic algorithm, and execute* Q *times of iterative computation until a point* $\boldsymbol{\lambda}^{\bar{k}}$ *is obtained with the property* $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}^{\bar{k}}, \boldsymbol{W}_{opt}(\omega)) - \bar{\boldsymbol{F}}(\boldsymbol{\lambda}^k, \boldsymbol{W}_{opt}(\omega)) \leqslant -\varepsilon_1$*.*

3. *Solve for the local minimum of* $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}_{opt}(\omega))$ *by using a gradient-based minimization method with* $\boldsymbol{\lambda}^{\bar{k}}$ *as the input point to get* $\boldsymbol{\lambda}^{k+1}$*, such that the objective function has a certain degree of decline,* $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}^{k+1}, \boldsymbol{W}_{opt}(\omega)) - \bar{\boldsymbol{F}}(\boldsymbol{\lambda}^{\bar{k}}, \boldsymbol{W}_{opt}(\omega)) \leqslant -\varepsilon_2$*.*

4. *Set* $k := k + 1$*, return to Step* **2** *until convergence.*

In Step **2** of Algorithm (8.3.1), the fitness function value for genetic algorithm is set to be the optimal value of the subproblem (8.13), the genetic algorithm composes of five key steps:

1. Population representation – The real-valued placement variables $\boldsymbol{\lambda}$ are initialized to construct the chromosomes and store an entire population in a single matrix with all chromosomes are of equal length.

2. Fitness assignment – The fitness values are derived from the optimal value of (8.14) via ranking or scaling.

3. Selection – Selection functions select a given number of placements from the current population, according to their fitness, and return a column vector to their indices.

4. Crossover – Crossover operators recombine pairs of individuals with given probability to produce offspring.

5. Mutation – Mutation operators apply random changes to individual parents to form offspring.

The overall procedure is outlined as follows:

Genetic algorithm

(a) Generate the initial placement chromosomes $\mathbf{\Lambda}^k$ via the last local optimal placement $\boldsymbol{\lambda}^k$, evaluate (8.15) for all the individuals to obtain $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}_{opt}(\omega)), \boldsymbol{\lambda} \in \mathbf{\Lambda}^k$.

(b) Rank $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}_{opt}(\omega)), \boldsymbol{\lambda} \in \mathbf{\Lambda}^k$, if $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}^{\bar{k}}, \boldsymbol{W}_{opt}(\omega)) - \bar{\boldsymbol{F}}(\boldsymbol{\lambda}^k, \boldsymbol{W}_{opt}(\omega)) \leqslant -\varepsilon_1$, stop; otherwise, go to next Step (c).

(c) Select the individuals $\mathbf{\Lambda}_1^k$ by using certain selection functions to be parents.

(d) Combine two parents to form offspring $\mathbf{\Lambda}_2^k$ for the next generation by using the crossover operator.

(e) Apply random changes to individual parents to form offspring $\mathbf{\Lambda}_3^k$ by using mutation operator.

(e) Evaluate (8.15) for all of the new individuals $\mathbf{\Lambda}_3^k$, get $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}_{opt}(\omega)), \boldsymbol{\lambda} \in \mathbf{\Lambda}_3^k$, and return it to Step (b).

8.4 Illustration Examples

8.4.1 Acoustic Room Configuration

A simple rectangular room with $12m \times 6m \times 3m$ is defined for our acoustic room modeling. A certain fraction of sound wave is absorbed and a certain amount is transmitted into the walls, floor and ceiling. These are the energy loss from the room and the fractional loss is characterized by the absorption coefficients, whereas, the overall effect can be represented by the reverberation time T_{60}. We input one SOI placed at $(0, 0, 1)$, one interferer placed at $(0, 1, 1)$, and background noise placed at $(0, -1, 1)$ in meter with the center of the room as the origin, where the background

noise placement information is not used during beamformer design, an illustration of the room setup is depicted in Figure 8.1. All the RIRs from the source points to sensor array are estimated by ISM room simulator, and the fast-ISM simulator developed by E.A. Lehmann and A.M. Johansson [68] is introduced to estimate the RIRs for three kinds of room acoustics: $T_{60} = \{0.05s, 0.1s, 0.2s\}$.

In such a room setup, we use a broadband male speech ("*Dots of light betrayed the black cat*") as the inputting SOI signal, and use a female speech ("*she had your dark suit in greasy wash water all year*") as the interference (INT) inputting, both of them have $8000Hz$ sampling frequency. Due to the noise signal is assumed to be independent to the SOI signal, we can generate a white-noise for background noise inputting as the NOI signal, see the plotting in Figure 8.1.

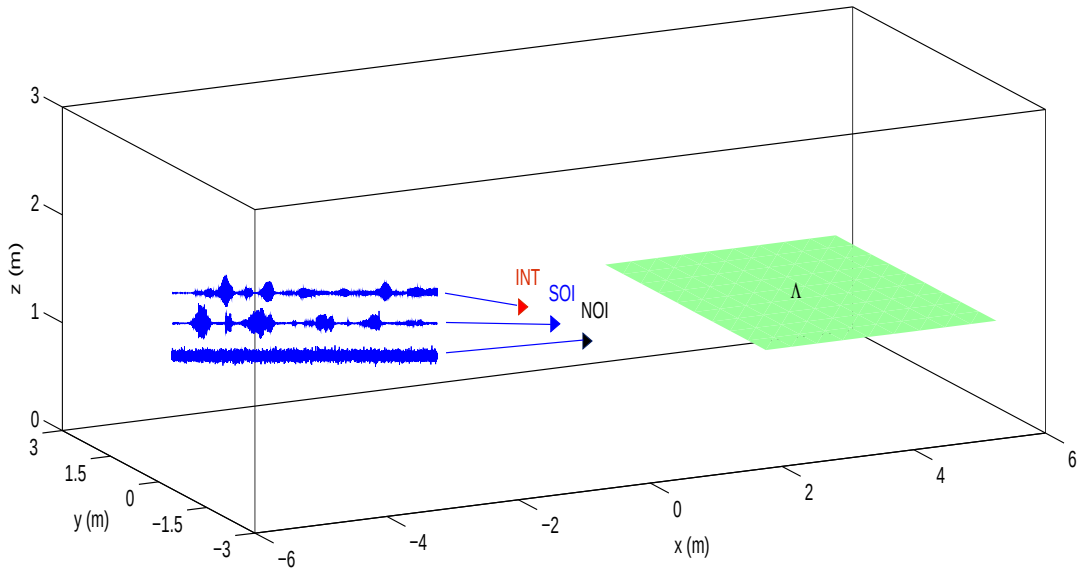


Figure 8.1: Setup of the acoustic room system.

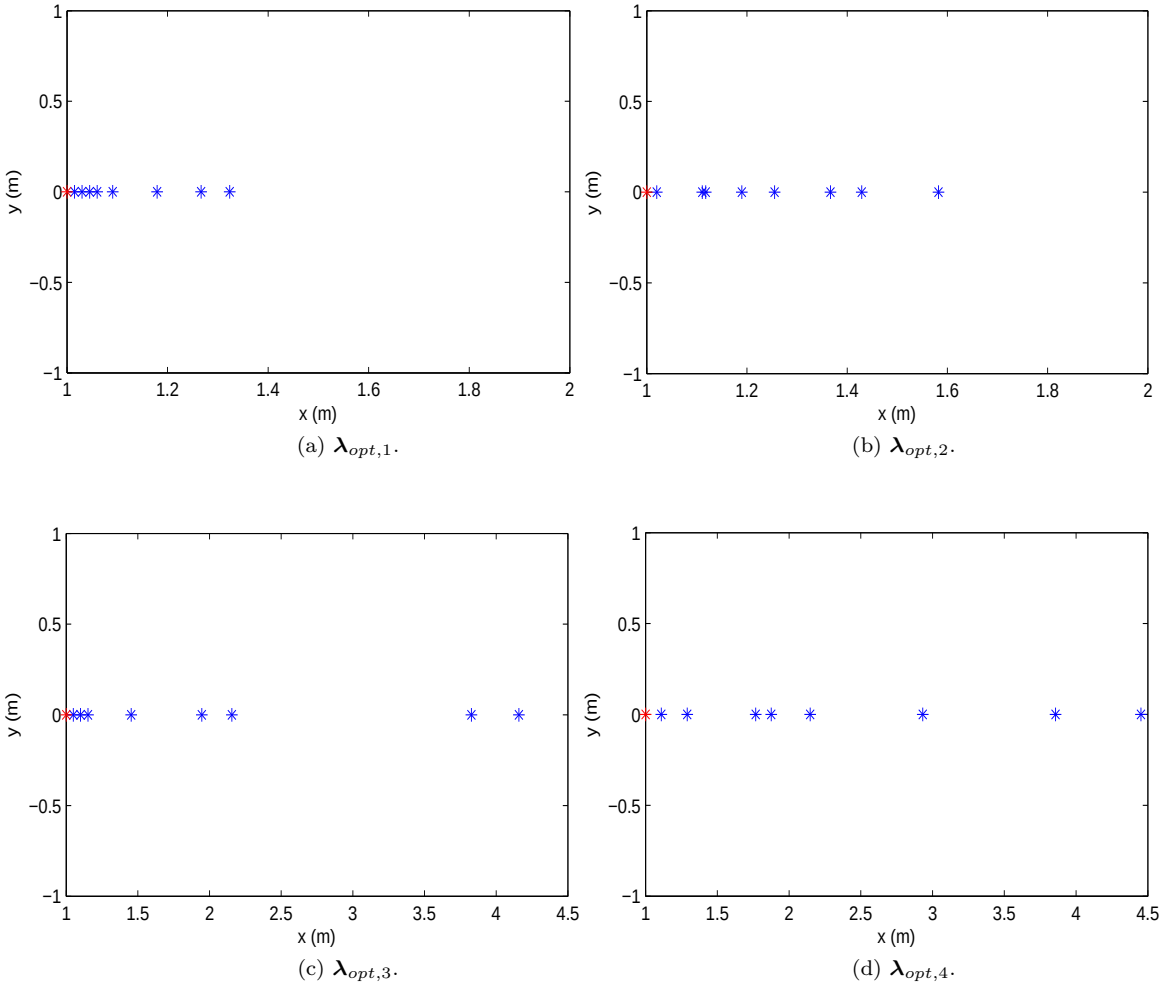
We define a 9-elements microphone array settled in a simple plane $\Lambda = \{1 \leqslant x \leqslant 5, -2.5 \leqslant y \leqslant 2.5, z = 1\}$ in meter as the feasible region to capture the SOI, INT and NOI signals, and use one of the microphone as the beamformer output point, see the Figure 8.1 for illustration.

8.4.2 Optimal Placement Solving

We firstly point out that this method is also appropriate for microphone array placement design in the free-field condition, although the proposed microphone array placement design based on LCMV beamforming technique is applied in the reverberant environment. For simplicity, we use the reverberation time $T_{60} = 0s$ to denote the sound wave propagation in the above room setup without considering the effect of reverberation, and apply the proposed method to find the optimal microphone array placement design under this condition. Therefore, we get the optimal solution $\boldsymbol{\lambda}_{opt,1}$ for this free-field microphone array placement design, see the microphone array placement configuration in Figure 8.2(a).

From the depict of the solved free-field optimal microphone array placement design in Figure 8.2(a), we can see that all the microphone elements are arranged on the vertical line in front of the source points, and the closest element is selected as the beamformer output point, see the red star in the figure, this result is also in good agreement with the experimental results in [33]. Then, we can use this solution $\boldsymbol{\lambda}_{opt,1}$ as the initial points to solve the other microphone array placement design problems in the reverberant environment. And the optimal solutions for the three kinds of indoor microphone array placement design with reverberation time $T_{60} = \{0.05s, 0.1s, 0.2s\}$ can be obtained by using the proposed method, also, the configurations of them are depicted in the Figure 8.2(b)-(d).

From the depicts of the above optimal microphone array placement designs, we can see that all the optimal placements are also arranged on the vertical line in front of the source points, and the closest elements are selected as the beamformer output points, see the red star in the figures. Moreover, the optimal microphone array placement design in the free-field is close to the source points, whereas the optimal microphone array placement designs in the reverberant environment are expended



(a) $\boldsymbol{\lambda}_{opt,1}$. (b) $\boldsymbol{\lambda}_{opt,2}$.

(c) $\boldsymbol{\lambda}_{opt,3}$. (d) $\boldsymbol{\lambda}_{opt,4}$.

Figure 8.2: Common used placement configurations.

much farther as the increase of reverberation, see the depicts in the above Figure 8.2 (a)-(d).

8.4.3 Optimal beamforming performance

Now that we compare the beamforming effects on the defined indicators for different microphone array placement designs in different environment. And we use the above defined five indicators: reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT}$, noise suppression $Supp_{NOI}$, segmental SNIRR and PESQ scores to measure the overall performance, respectively. In detail, we list the overall results on these

indicators in the following Table 8.1.

In the Table 8.1, we emphasized the best results on the enhancement in the corresponding microphone array placement design in boldface, and list the results of other microphone array placement designs for comparison. From these results, it can be seen that the best effect of speech enhancement is obtained at $\boldsymbol{\lambda}_{opt,1}$ in the free-field condition. However, as the reverberation increases, the best effect of speech enhancement can not be achieved at $\boldsymbol{\lambda}_{opt,1}$, but obtained at $\boldsymbol{\lambda}_{opt,2}$, $\boldsymbol{\lambda}_{opt,3}$ and $\boldsymbol{\lambda}_{opt,4}$ for the room acoustics with reverberation time $T_{60} = 0.05s$, $T_{60} = 0.1s$ and $T_{60} = 0.2s$, respectively.

For illustration, we also plot the beamforming results on the reverberation suppression, interference suppression, noise suppression and speech enhancement for the three kinds room acoustics in the following Figure 8.3, Figure 8.4 and Figure 8.5, respectively. From the results in these figures, it can be seen that at the optimal microphone array placement designs, the corresponding beamformers can achieve effective suppressions on reverberation, interference and noise, and finally, enhance the SOI signals.

Table 8.1: Summary of the beamforming performance at different microphone array placement designs.

T_{60}	$\boldsymbol{\lambda}_{opt}$	$Supp_{REV}$	$Supp_{INT}$	$Supp_{NOI}$	$SNIRR_{seg}$		$PESQ$	
					Unfiltered	*Filtered*	*Unfiltered*	*Filtered*
0	$\boldsymbol{\lambda}_{opt,1}$	——	**64.5225**	**55.2305**	**-4.0533**	**31.5169**	**1.3575**	**4.4891**
	$\boldsymbol{\lambda}_{opt,2}$	——	52.3383	49.1290	-4.0533	28.9000	1.3575	4.4579
	$\boldsymbol{\lambda}_{opt,3}$	——	33.2180	35.8123	-4.0526	20.6136	1.3575	3.9036
	$\boldsymbol{\lambda}_{opt,4}$	——	32.5225	34.3196	-4.0526	20.0587	1.3575	3.9224
0.05	$\boldsymbol{\lambda}_{opt,1}$	27.5460	32.4036	33.6234	-4.7940	17.2046	1.3763	3.3351
	$\boldsymbol{\lambda}_{opt,2}$	**32.4624**	**39.1435**	**39.0875**	**-4.7941**	**22.4520**	**1.3763**	**3.7075**
	$\boldsymbol{\lambda}_{opt,3}$	27.8589	33.0486	34.2669	-4.7936	18.5692	1.3764	3.4847
	$\boldsymbol{\lambda}_{opt,4}$	25.1004	30.1792	33.5946	-4.7936	17.1347	1.3764	3.5511
0.1	$\boldsymbol{\lambda}_{opt,1}$	10.4564	24.3906	27.6087	-4.6348	10.1322	1.3765	3.0990
	$\boldsymbol{\lambda}_{opt,2}$	16.4976	25.4371	26.8072	-4.6427	11.7183	1.3760	3.1170
	$\boldsymbol{\lambda}_{opt,3}$	**21.3615**	**29.0429**	**28.8280**	**-4.6310**	**14.9132**	**1.3740**	**3.1248**
	$\boldsymbol{\lambda}_{opt,4}$	21.0103	27.7977	28.0733	-4.6313	14.0199	1.3746	3.1207
0.2	$\boldsymbol{\lambda}_{opt,1}$	15.3812	20.6743	24.1998	-5.0849	8.9280	1.3739	2.8609
	$\boldsymbol{\lambda}_{opt,2}$	16.8513	20.7430	23.9908	-5.0849	9.0894	1.3739	2.8857
	$\boldsymbol{\lambda}_{opt,3}$	20.0668	24.3083	24.6788	-5.0844	10.6398	1.3740	2.8854
	$\boldsymbol{\lambda}_{opt,4}$	**20.4096**	**25.2930**	**24.8435**	**-5.0845**	**11.1968**	**1.3740**	**2.9366**

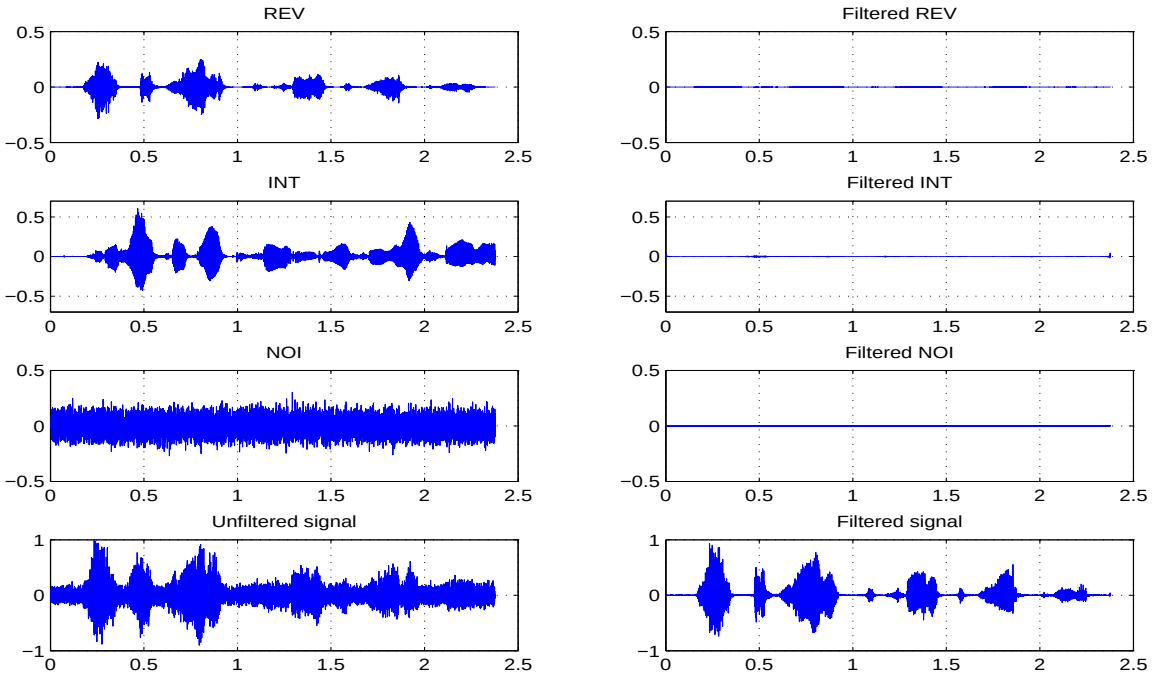


Figure 8.3: Limit performance of beamformer in optimal placement at $T_{60} = 0.05$s.

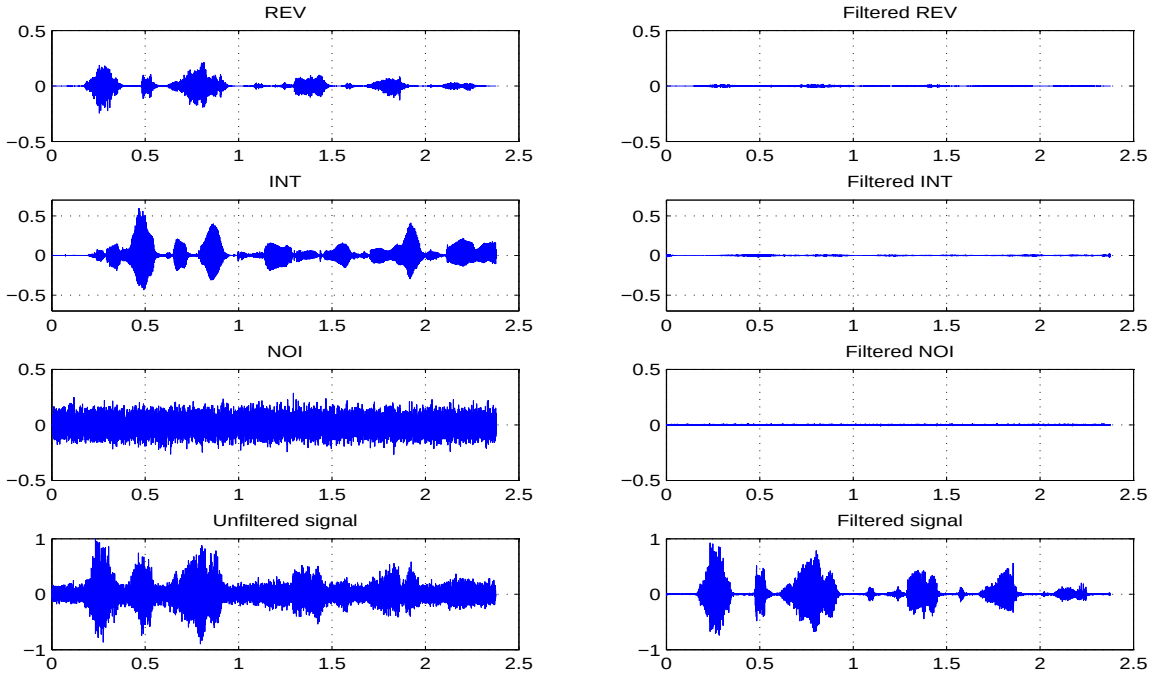


Figure 8.4: Limit performance of beamformer in optimal placement at $T_{60} = 0.1$s.

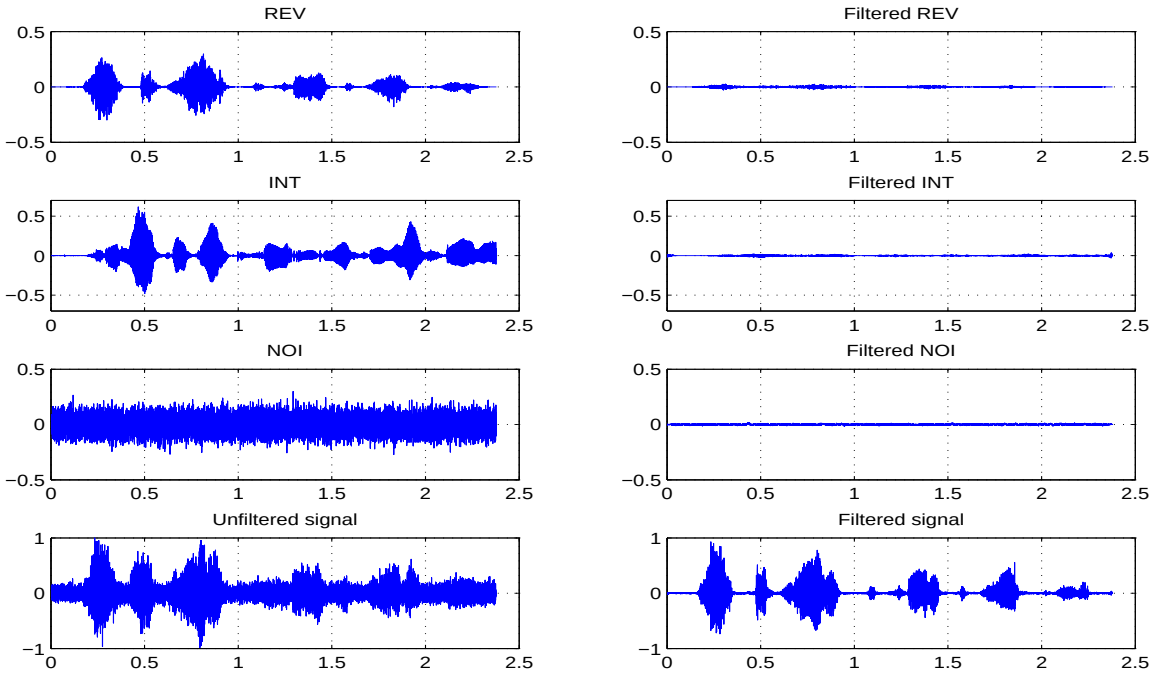


Figure 8.5: Limit performance of beamformer in optimal placement at $T_{60} = 0.2$s.

Chapter 9

Conclusions and Future Works

This chapter draws conclusions on the works studied in this thesis, and points out some possible research directions related to these works in the future.

9.1 Conclusions

The focus of this thesis is placed on various near-field beamformer design problems, including broadband beamformer design problems in reverberant environment, time domain beamformer design problems in reverberant environment and microphone array placement design problems. Specifically, six research works have been studied in the following.

1. **Indoor broadband beamformer design problem in Chapter 3.**

   In the reverberant environment, room acoustics has a big influence on the sound wave propagation. The commonly used beamformer designs based on the assumption that the considered system is in the open air will not be effective. Fortunately, there are lots of simple but efficient geometrical based method for room acoustics simulation, such as the image-source method. Thus, we establish appropriate indoor beamformer design problem with the help of RIRs estimated by the image-source method based room simulator. Then we have

proposed several numerical optimization models to solve it, and obtain effective indoor beamformer designs eventually.

2. **Barrier beamformer design problem in Chapter 4.**

   The geometrical based room simulators are usually suitable for regular office rooms, however, obstacles between the source points and sensor array often appear in practice. The wave based space-time conservation element and solution element (CE/SE) method can effectively simulate the barrier acoustics. Thus, we establish appropriate barrier beamformer design problem with the help of BIRs estimated by using the CE/SE method. Then we proposed effective numerical optimization model to obtain the better designs.

3. **A study on time domain beamformer design problem in reverberant environment in Chapter 5.**

   The time domain beamformer design problem is usually formulated as a linear system or a least squares problem. For the design of beamformer in reverberant environment, the linear system and least squares problem will be over-determined and have large scale with very poor conditions. The regularization technique is effective to improve the condition of the problem. Thus, we introduce the Tikhonov regularization scheme to deal with the least squares problem for beamformer design in reverberant environment. Moreover, we study the relationship between the beamformer design and the filter length under the optimal condition, and do numerical experiments to illustrate it.

4. **Indoor LCMV beamformer design problem in Chapter 6.**

   On the design of time domian beamformer, the LCMV beamformer technique has been widely adopted for the microphone array design. It minimizes the estimated noise power to achieve noise reduction and can be applied to non-

directional background noise generated from far-field. Therefore, we study the indoor LCMV beamformer design problem considering also the influence of the room acoustics. We carry out numerical simulations to study the influence of the group delay and the filter length on the LCMV beamformer performance. Moreover, we introduce the post-filter technique to further improve the performance of speech enhancement, and find that the MSIG-postfilter combining with LCMV beamformer has the best effect based on the results of our numerical experiments.

5. **Microphone array placement design problem in Chapter 7.**

   In general, the microphone array placement design problem is a nonconvex optimization problem. It has a composite objective function with variables of the microphone array placement and beamformer coefficients, which are usually related to filter length. Thus, we introduce the infinite length technique to transform the subproblem of solving filter coefficients to performance limit estimation, that is we reduce the filter coefficient variables related to filter length. Then, we develop a hybrid descent method with genetic algorithm to find the optimal placement design effectively. It employs a gradient-based technique for local neighbourhood improvement and a genetic algorithm technique to jump out local minima. In the numerical experiments, we apply the proposed method to the 2-D and 3-D microphone array placement design problems respectively.

6. **Indoor microphone array placement design problem in Chapter 8.**

   As a result of the complex room acoustics in the reverberant environment, the scale of the beamformer design problem is usually large, and the process of beamformer design is time consuming. Thus, it is difficult to apply the broadband microphone array placement design technique proposed in Chapter 7 for

the indoor cases. However, with the known location of signal of interest and interferences, and nondirectional background noise from far-field, we can introduce the LCMV beamforming approach to construct the optimization model for indoor microphone array placement design. The infinite length technique can also be applied to transform the subproblem of optimizing the beamformer coefficients alone to obtain performance limit estimates. Moreover, the hybrid descent method has been applied to find the optimal placement design.

9.2 Future Works

Related topics for the future research works are listed below.

1. **More effective methods for beamformer design in reverberant environment.**

   In the Chapter 3 and Chapter 4, we have studied the indoor beamformer and barrier beamformer design problems, where all the beamformer design problems have been formulated into optimization models by using L_1-norm, and solved by linear programming technique. And the numerical experiments showed the effectiveness of the proposed methods. However, the performances may not been sufficient in the heavier reverberant conditions. Thus, more effective method focus on the reverberant beamformer design is necessary in the future, so as the robust beamformers design.

2. **Time domain beamformer design in reverberant environment.**

   In the Chapter 5 and Chapter 6, the least squares technique and LCMV beamformer approach have been introduced to design beamformers in reverberant environment. And the numerical experiments show that the proposed methods are very effective. Moreover, the influences of the group delay and filter length have also been studied. However, there are also many effective beamformer design methods can be applied to reverberant environment, such as the generalized sidelobe canceller (GSC). Thus, it is meaningful to study the other beamformer design methods in the reverberant environment in the future.

3. **Microphone array placement design problems.**

   In the Chapter 7 and Chapter 8, we have studied the microphone array placement design problems in the open air and reverberant environment. The composite optimization problems have been formulated with respect to the optimal

placement variables and beamformer coefficients. The infinite length technique has been introduced to transform the subproblems of beamformer coefficients solving to performance limit estimates. The results of numerical experiments show that the proposed methods are effect. However, the developed composite optimization model is noncovex and nonlinear, the global optimal solution of the placement design is difficult to be obtained. Moreover, the proposed hybrid descent algorithm is usually very time consuming in reverberant case. It is still a challenge on the develop of effect and efficient method for microphone array placement design in the future.

Bibliography

[1] J. Allen and D. Berkley. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Amer.*, 65(4):943–950, 1979.

[2] J.B. Allen. Short-time spectral analysis, synthesis and modification by discrete Fourier transform. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-25:235–238, 1977.

[3] J. António, L. Godinho, and A. Tadeu. Reverberation times obtained using a numerical model versus those given by simplified formulas and measurements. *Acust. Acta Acoust.*, 88(2):252–261, 2002.

[4] A.B. Bakushinskii. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *USSR Comp. Math. Math. Phys.*, 24:181–182, 1984.

[5] I. Barrodale, L. Delves, and J. Mason. Linear Chebyshev approximation of complex-valued functions. *Math. Comput.*, 32(143):853–863, 1978.

[6] J. Benesty, J. Chen, and Y. Huang. *Microphone Array Signal Processing.* Springer Verlag, Berlin, 2008.

[7] J. Benesty, J. Chen, Y. Huang, and I. Cohen. *Noise Reduction in Speech Processing.* Springer Verlag, Berlin, 2009.

[8] M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. In *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP'79)*, volume 4, pages 208–211, 1979.

[9] S.F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-27(2):113–120, 1979.

[10] J. Borish. An extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Amer.*, 75:1827–1836, 1984.

[11] J.S. Bradley and H. Sato. On the importance of early reflections for speech in rooms. *J. Acoust. Soc. Amer.*, 113(6):3233–3244, 2003.

[12] M. Brandstein and D. Ward. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Verlag, Berlin, 2001.

[13] D. Calvetti and L. Reichel. Tikhonov regularization of large linear problems. *BIT Numer. Math.*, 43:263–283, 2003.

[14] J. Capon. High resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57(8):1408–1418, 1969.

[15] O. Cappe. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. Speech Audio Process.*, 2(2):345–349, 1994.

[16] K.Y. Chan, S.Y. Low, S. Nordholm, K.F.C. Yiu, and S.H. Ling. Speech recognition enhancement using beamforming and a genetic algorithm. In *Proc. IEEE Third Int. Conf. on Network System Security*, pages 510–515, Gold Coast, QLD, 2009.

[17] S.C. Chang. The method of space-time conservation element and solution element - a new approach for solving the Navier-Stokes and Euler equations. *J. Compu. Phys.*, 119:295–324, 1995.

[18] S.C. Chang, X.Y. Wang, and C.Y. Chow. The space-time conservation element and solution element method: A new high-resolution and genuinely multidimensional paradigm for solving conservation laws. *J. Compu. Phys.*, 156(1):89–136, 1999.

[19] K.S. Chen, X.H. Yun, Z.S. He, and C.L. Han. Synthesis of sparse planar arrays using modified real genetic algorithm. *IEEE Trans. Antennas Propag.*, 55(4):1067–1073, 2007.

[20] L. Cohen. The history of noise. *IEEE Signal Process. Mag.*, 22:20–45, 2005.

[21] S. Darlington. Linear least-squares smoothing and prediction with applications. *Bell Syst. Tech. J.*, 37:1121–1194, 1952.

[22] L. Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.

[23] M. Dendrinos, S. Bakamidis, and G. Carayannis. Speech enhancement from noise: a regenerative approach. *Speech Commun.*, 10:45–57, 1991.

[24] G. Doblinger. Optimized design of interpolated array and sparse array wideband beamformers. In *16th Eur. Signal Process. Conf. (EUSIPCO 2008)*, Lausanne, Switzerland, 2008.

[25] Y. Ephraim. Statistical-model-based speech enhancement systems. *Proc. IEEE*, 80:1526–1555, 1992.

[26] Y. Ephraim. A Bayesian estimation approach for speech enhancement using hidden markov models. *IEEE Trans. Signal Process.*, 40:725–735, 1992.

[27] Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Process.*, 32(6):1109–1121, 1984.

[28] Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Process.*, 33(2):443–445, 1985.

[29] Y. Ephraim, D. Malah, and B.H. Juang. On the application of hidden Markov models for enhancing noisy speech. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-37:1846–1856, 1989.

[30] Y. Ephraim and H.L. Van Trees. A signal subspace approach for speech enhancement. *IEEE Trans. Speech Audio Process.*, 3:251–266, 1995.

[31] M. Er and A. Cantoni. Derivative constraints for broad-band element space antenna array processors. *IEEE Trans. Acoust., Speech, Signal Process.*, 31(6):1378–1393, 1983.

[32] Z.G. Feng, K.F.C. Yiu, and S.E. Nordholm. A two-stage method for the design of near-field broadband beamformer. *IEEE Trans. Signal Process.*, 59(8):3647–3656, 2011.

[33] Z.G. Feng, K.F.C. Yiu, and S.E. Nordholm. Placement design of microphone arrays in near-field broadband beamformers. *IEEE Trans. Signal Process.*, 60(3):1195–1204, 2012.

[34] J. Flanagan, J. Johnson, R. Kahn, and G. Elko. Computer-steered microphone arrays for sound transduction in large rooms. *J. Acoust. Soc. Amer.*, 78:1508–1518, 1994.

[35] J. Flanagan, A.C. Surendran, and E.E. Jan. Spatially selective sound capture for speech and audio processing. *Speech Commun.*, 13(1-2):207–222, 1993.

[36] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Trans. Signal Process.*, 49(8):1614–1626, 2001.

[37] S. Gannot and I. Cohen. Speech enhancement based on the general transfer function GSC and postfiltering. *IEEE Trans. Speech Audio Process.*, 12(6):561–571, 2004.

[38] H. Gfrerer. An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Math. Comput.*, 49:507–522, 1987.

[39] K. Glashoff and K. Roleff. A new method for Chebyshev approximation of complex-valued functions. *Math. Comput.*, 36(153):233–239, 1981.

[40] G. Golub, P. Hansen, and D. O'Leary. Tikhonov regularization and total least squares. *SIAM J. Matrix Anal. Appl.*, 21(1):185–194, 1999.

[41] G. Golub, M. Heath, and G. Wahba. Generalised cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[42] G.H. Golub and C.F. Van Loan. *Matrix Computations, 3rd Edition*. Johns Hopkins University Press: Baltimore, 1996.

[43] G.H. Golub and U. von Matt. Tikhonov regularization for large scale problems. In G.H. Golub, S.H. Lui, F. Luk, and R. Plemmons, editors, *Workshop on Scientific Computing*, pages 3–26. Springer, New York, 1997.

[44] L. Griffiths and C. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antenas Propag.*, 30(1):27–34, 1982.

[45] C.W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, 1984.

[46] F.B. Gross. *Smart Antennas for Wireless Communications: With MATLAB*. McGraw-Hill, New York, 2005.

[47] H. Gustafsson, R. Martin, P. Jax, and P. Vary. A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *IEEE Trans Speech Audio Process.*, 10(5):245–256, 2002.

[48] J. Hansen and B. Pellom. An effective quality evaluation protocol for speech enhancement algorithms. In *Proc. Int. Conf. on Spoken Language Process.*, pages 2819–2822, San Francisco, CA, 1998.

[49] P.C. Hansen and D.P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14:1487–1503, 1993.

[50] R.L. Haupt. Thinned arrays using genetic algorithms. *IEEE Trans. Antennas Propag.*, 42:993–999, 1994.

[51] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, 49:409–436, 1952.

[52] C.Y. Ho, R.C.K. Leung, K. Zhou, G.C.Y. Lam, and Z. Jiang. One-step direct aeroacoustic simulation using space-time conservation element and solution element method. In *Proc. Sixth Int. Conf. on Fluid Mechanics*, volume 1376, pages 218–220, Guangzhou, China, 2011.

[53] J.H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, 1975.

[54] Y. Huang, J. Benesty, and J. Chen. *Acoustic MIMO Signal Processing*. Springer-Verlag, Berlin, 2006.

[55] Y. Huang, J. Benesty, and J. Chen. Adaptive blind multichannel identification. In J. Benesty, M.M. Sondhi, and Y. Huang, editors, *Springer Handbook of Speech Processing*, chapter 13, pages 259–282. New York: Springer-Verlag, 2007.

[56] E. Jan, P. Svaizer, and J.L. Flanagan. Matched-filter processing of microphone array for spatial volume selectivity. In *IEEE Int. Symposium on Circuits Syst.*, volume 2, pages 1460–1463, 1995.

[57] D. Johnson and D. Dudgeon. *Array Signal Processing Concepts and Techniques*. Prentice-Hall, 1993.

[58] Y. Kaneda. Directivity characteristics of adaptive microphone-array for noise reduction (amnor). *J. Acoust. Soc. Jpn.*, 12(4):179–187, 1991.

[59] Y. Kaneda and J. Ohga. Adaptive microphone-array system for noise reduction. *IEEE Trans. Acoustics, Speech, Signal Process.*, 34:1391–1400, 1986.

[60] A. Kataoka and Y. Ichirose. A microphone array configuration for anmor (adaptive microphone-array system for noise reduction). *J. Acoust. Soc. Jpn.*, 11(6):317–325, 1990.

[61] R.A. Kennedy, T.D. Abhayapala, and D.B. Ward. Broadband nearfield beamforming using a radial beampattern transformation. *IEEE Trans. Signal Process.*, 46:2147–2156, 1998.

[62] R.A. Kennedy, D.B. Ward, and T.D. Abhayapala. Nearfield beamforming using radial reciprocity. *IEEE Trans. Signal Process.*, 47(1):33–40, 1999.

[63] B.P. Kumar and G.R. Branner. Generalized analytical technique for the synthesis of unequally spaced arrays with linear, plannar, cylindrical of spherical geometry. *IEEE Trans. Antennas Propag.*, 53(2):621–634, 2005.

[64] C. Kumar, S. Rao, and A. Hoorfar. Optimization of thinned phased arrays using evolutionary programming. In *Evolutionary Programming VII*, volume 1447 of *Lecture Notes in Computer Science*, pages 157–166. Springer Berlin / Heidelberg, 1998.

[65] G.C.Y. Lam. *Aeroacoustics of Merging Flows at Duct Junctions, Ph.D. Thesis.* The Hong Kong Polytechnic University, 2011.

[66] B. Lau, Y. Leung, K.L. Teo, and V. Sreeram. Minimax filters for microphone arrays. *IEEE Trans. Circuits Syst. II*, 46(12):1522–1525, 1999.

[67] E.A. Lehmann and A.M. Johansson. Prediction of energy decay in room impulse responses simulated with an image-source model. *J. Acoust. Soc. Amer.*, 124(1):269–277, 2008.

[68] E.A. Lehmann and A.M. Johansson. Diffuse reverberation model for efficient image-source simulation of room impulse responses. *Trans. Audio, Speech, Language Process.*, 18:1429–1439, 2010.

[69] E.A. Lehmann, A.M. Johansson, and Sven Nordholm. Reverberation-time prediction method for room impulse responses simulated with the image-source model. In *2007 IEEE Workshop on Appl. Signal Process. Audio, Acoust.*, pages 159–162, New Paltz, NY, USA, 2007.

[70] H. Lehnert and J. Blauert. Principles of binaural room simulation. *Appl. Acoust.*, 36(3-4):259–291, 1992.

[71] A. Levy, S. Gannot, and E.A.P. Habets. Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments. *IEEE Trans. Audio, Speech, Language Process.*, 19(6):1540–1555, 2011.

[72] K. Li and P. Lam. Prediction of reverberation time and speech transmission index in long enclosures. *J. Acoust. Soc. Amer.*, 117:3716–3726, 2005.

[73] J.C. Liberti and T.S. Rappaport. *Smart Antennas for Wireless Communications: IS-95 and Third Generation A microphone array configuration for anmor (adaptive microphone-array system for noise reduction) Applications.* Prentice Hall, Upper Saddle River, NJ, 1999.

[74] Y.C. Lim and S.R. Parker. FIR filter design over a discrete powers-of-two coefficient space. *IEEE Trans. Acoustics, Speech, Signal Process.*, 31(3):583–591, 1983.

[75] M. Liu, J.B. Wang, and K.-Q. Wu. The direct aero-acoustics simulation of flow around a square cylinder using the CE/SE scheme. *J. Algorithms Comput. Technol.*, 1:525–537, 2007.

[76] W. Liu. Adaptive wideband beamforming with sensor delay-lines. *Signal Process.*, 89(5):876–882, 2009.

[77] W. Liu and S. Weiss. *Wideband Beamforming: Concepts and Techniques.* John Wiley and Sons, Chichester, UK, 2010.

[78] J.P.A. Lochner and J.F. Burger. The influence of reflections on auditorium acoustics. *J. Sound Vib.*, 1:426–454, 1964.

[79] C.Y. Loh, L.S. Hultgren, and S.C. Chang. Wave computation in compressible flow using space-time conservation element and solution element method. *AIAA J.*, 39(5):794–801, 2001.

[80] P. Loizou. *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton, FL, 2007.

[81] P. Lombardo, R. Cardinali, D. Pastina, M. Bucciarelli, and A. Farina. Array optimization and adaptive processing for sub-array based thinned arrays. In *Int. Conf. on Radar*, pages 197–202, Rome, Italy, 2008.

[82] F.J. MacWilliams and N.J.A. Sloane. Pseudo-random sequences and arrays. *Proc. IEEE*, 64(12):1715–1729, 1976.

[83] J.T. Mayhan. Thinned array configurations for use with satellitebased adaptive antennas. *IEEE Trans. Antennas Propag.*, 28(6):846–856, 1980.

[84] R.J. McAulay and M.L. Malpass. Speech enhancement using a soft-decision noise suppression filter. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-28:137–145, 1980.

[85] C. Metz, L.C. Stange, E. Lissel, and A.F. Jacob. Performance of thinned antenna arrays using nonlinear processing in DBF radar applications. In *IEEE MTT-S Int. Microwave Symposium Digest*, pages 275–278, Phoenix, AZ, USA, 2001.

[86] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 1992.

[87] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*. New York: Springer, 1984.

[88] P.A. Naylor and N.D. Gaubitch. *Speech Dereverberation*. Springer-Verlag, London, 2010.

[89] R. Neubauer and B. Kostek. Prediction of the reverberation time in rectangular rooms with non-uniformly distributed sound absorption. *Archives Acoust.*, 26(3):183–202, 2001.

[90] S. Nordebo, I. Claesson, and S. Nordholm. Weighted chebyshev approximation for the design of broadband beamformers using quadratic programming. *IEEE Signal Process. Lett.*, 1:103–105, 1994.

[91] S. Nordholm, V. Rehbock, K. L. Teo, and S. Nordebo. Chebyshev optimization for the design of broadband beamformers in the near field. *IEEE Trans. Circuits Syst. II*, 45(1):141–143, 1998.

[92] G. Oliveri, M. Donelli, and A. Massa. Linear array thinning exploiting almost difference sets. *IEEE Trans. Antennas Propag.*, 57(12):3800–3812, 2009.

[93] ITU-T P.835. *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm.* ITU-T Recommendation P.835, 2003.

[94] ITU-T P.862. *Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.* ITU-T Recommendation P.862, 2000.

[95] C.C. Paige and M.A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Softw.*, 8(1):43–71, 1982.

[96] K.K. Paliwal and A. Basu. A speech enhancement method based on Kalman filtering. In *IEEE Int. Conf. on ICASSP '87 Acoustics, Speech, Signal Process.*, pages 177–180, USA, 1987.

[97] M.R. Portnoff. Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *Audiology*, ASSP-28:55–69, 1980.

[98] S. Quackenbush, T. Barnwell, and M. Clements. *Objective Measures of Speech Quality.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[99] T.S. Rappaport. *Wireless Communications: Principles and Practice.* Prentice Hall, Englewood Cliffs, NJ, 1996.

[100] A. Razavi and K. Forooraghi. Progress in electromagnetics research. *IEEE Trans. Antennas Propag.*, 78:61–71, 2008.

[101] A. Rix, J. Beerends, M. Hollier, and A. Hekstra. Perceptual evaluation of speech quality (PESQ), a new method for speech quality assessment of telephone networks and codecs. In *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP'01)*, volume 2, pages 749–752, Barcelona, Spain, 2001.

[102] P. Rocca and R. Haupt. Dynamic array thinning for adaptive interference cancellation. In *The Fourth Eur. Conf. on Antennas and Propagation (EuCAP2010)*, pages 1–3, Barcelona, Spain, 2010.

[103] J. Ryan and R. Goubran. Array optimization applied in the near field of a microphone array. *IEEE Trans. Speech, Audio Process.*, 8(2):173–178, 2000.

[104] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial Mathematics, 2003.

[105] A. Sarajedini. Adaptive array thinning for space-time beamforming. In *The Thirty-Third Asilomar Conf. on Signals, Syst., Comput.*, pages 1572–1576, Pacific Grove, CA, USA, 1999.

[106] P. Scalart. Speech enhancement based on a priori signal to noise estimation. In *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP'96)*, volume 2, pages 629–632, 1996.

[107] M.R. Schroeder. New method of measuring reverberation time. *J. Acoust. Soc. Amer.*, 37:409–412, 1965.

[108] U.P. Svensson. Modelling acoustic spaces for audio virtual reality. In *Workshop on Model based Process., Coding of Audio (MPCA)*, pages 109–116, Leuven, Belgium, 2002.

[109] R. Talmon, I. Cohen, and Gannot. Relative transfer function identification using convolutive transfer function approximation. *IEEE Trans. Audio, Speech, Language Process.*, 17(4):546–555, 2009.

[110] A. Trucco. Weighting and thinning wide-band arrays by simulated annealing. *Ultrasonics*, 40:485–489, 2002.

[111] A. Trucco and V. Murino. Stochastic optimization of linear sparse arrays. *IEEE J. Ocean. Eng.*, 24(3):291–299, 1999.

[112] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo. Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics. In *Proc. Int. Workshop on Acoust., Echo, Noise Control (IWAENC'08)*, Seattle, USA, 2008.

[113] B.D. Van Veen and K.M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE Acoust., Speech, Signal Process. Mag.*, 5:4–24, 1988.

[114] X.Y. Wang and S.C. Chang. A 2-D non-splitting unstructured triangular mesh euler solver based on the space-time conservation element and solution element method. *Comput. Fluid Dyn. J.*, 8:309–325, 1999.

[115] D.B. Ward and G.W. Elko. Mixed nearfield/farfield beamforming: a new technique for speech acquisition in a reverberant environment. In *IEEE ASSP Workshop on Appl. Signal Process. to Audio, Acoust.*, 1997.

[116] M.R. Weiss, E. Aschkenasy, and T.W. Parsons. Processing speech signals to attenuate interference. In *Proc. IEEE Symposium on Speech Recognition*, pages 292–293, Pittsburgh, PA, USA, 1974.

[117] K.F.C. Yiu, N. Grbic, Teo. Kok-Lay, and S. Nordholm. A new design method for broadband microphone arrays for speech input in automobiles. *IEEE Signal Process. Lett.*, 9(7):222–224, 2002.

[118] K.F.C. Yiu, X.Q. Yang, S. Nordholm, and K.L. Teo. Near-field broadband beamformer design via multidimensional semi-infinite linear programming techniques. *IEEE Trans. Speech, Audio Process.*, 11(6):725–732, 2003.

[119] P.C. Yong, S. Nordholm, and H.H. Dam. Optimization and evaluation of sigmoid function with *a priori* SNR estimate for real-time speech enhancement. *Speech Commun.*, 55(2):358–376, 2013.

[120] P.C. Yong, S. Nordholm, H.H. Dam, and S.Y. Low. On the optimization of sigmoid function for speech enhancement. In *Proc. 19th Eur. Signal Process. Conf.(EUSIPCO'11)*, pages 211–215, Barcelona, Spain, 2011.

[121] Y. Zhao, W. Liu, and R.J. Langley. A least squares approach to the design of frequency invariant beamformers. In *Proc. Eur. Signal Process. Conf.*, pages 844–848, Glasgow, Scotland, 2009.