2013

MICROPHONE ARRAY OPTIMIZATION IN IMMERSIVE ENVIRONMENTS

Jingjing Yu
University of Kentucky, jyu5@uky.edu

Recommended Citation

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

<div align="right">

Jingjing Yu, Student

Dr. Kevin D. Donohue, Major Professor

Dr. Zhi Chen, Director of Graduate Studies

</div>

MICROPHONE ARRAY OPTIMIZATION IN IMMERSIVE ENVIRONMENTS

_____

DISSERTATION
_____

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering at the
University of Kentucky

By
Jingjing Yu

Lexington, Kentucky

Director: Dr. Kevin D. Donohue, Professor of Electrical Engineering Department

Lexington, Kentucky

2013

ABSTRACT OF DISSERTATION

MICROPHONE ARRAY OPTIMIZATION IN IMMERSIVE ENVIRONMENTS

The complex relationship between array gain patterns and microphone distributions limits the application of traditional optimization algorithms on irregular arrays, which show enhanced beamforming performance for human speech capture in immersive environments. This work analyzes the relationship between irregular microphone geometries and spatial filtering performance with statistical methods. Novel geometry descriptors are developed to capture the properties of irregular microphone distributions showing their impact on array performance. General guidelines and optimization methods for regular and irregular array design are proposed in immersive (near-field) environments to obtain superior beamforming ability for speech applications. Optimization times are greatly reduced through the objective function rules using performance-based geometric descriptions of microphone distributions that circumvent direct array gain computations over the space of interest. In addition, probabilistic descriptions of acoustic scenes are introduced to incorporate various levels of prior knowledge for the source distribution. To verify the effectiveness of the proposed optimization methods, simulated gain patterns and real SNR results of the optimized arrays are compared to corresponding traditional regular arrays and arrays obtained from direct exhaustive searching methods. Results show large SNR enhancements for the optimized arrays over arbitrary randomly generated arrays and regular arrays, especially at low microphone densities. The rapid convergence and acceptable processing times observed during the experiments establish the feasibility of proposed optimization methods for array geometry design in immersive environments where rapid deployment is required with limited knowledge of the acoustic scene, such as in mobile platforms and audio surveillance applications.

Jingjing Yu

_____
Student's Signature

April 18, 2013
_____
Date

MICROPHONE ARRAY OPTIMIZATION IN IMMERSIVE ENVIRONMENTS

By

Jingjing Yu

Kevin D. Donohue

Director of Dissertation

Zhi Chen

Director of Graduate Studies

April 18, 2013

Date

To my loving family and dearest friends

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Microphone Array Signal Processing

Microphone arrays use spatial diversity of element positions to capture acoustic signals in higher quality and reduce degradation brought on by reverberation and noise. Inspired by traditional radar and sonar technology, as early as 1970's microphone arrays have been involved in the research of audio signal processing to capture speech signals, and then achieved significant performance improvement by applying digital signal processing technology in 1990's. Nowadays, microphone arrays have been widely applied in many applications, such as speech enhancement, talker tracking, teleconference, multi-party telecommunication, hands-free human-machine interfaces, acoustic surveillance systems, and computer games [1, 2].

The main objective of microphone array processing is to use the temporal-spatial-frequency information captured by dispersed microphones to extract interested signal components from interferences, and estimate relevant acoustic parameters [2]. Depending on applications, performance of array is usually assessed by its ability to locate, track, and separate sound sources in the Field of View (FOV) [1]. Critical factors affecting performance include acoustic environment, source spectral content, processing algorithm, and microphone geometry. The main focus of previous research is usually on the improvement of algorithm to process the information available at the output of array, which contains two aspects, estimation of source parameters and beamforming technology. Source parameter estimations use the received multichannel signal to estimate the direction, position, and frequency contents of single or multiple sound

sources. Generally, they can be divided into three classes: maximum output power beamforming [3], high resolution spectral estimation [4], and localization based on TDOA [5, 6]. Because the localization based on TDOA has less computational complexity and is more applicable for real-time system, it has been widely studied and applied. Other approaches such as maximum likelihood, linear prediction, Music and ESPRIT are also considered as the classical spectral estimation algorithms. Beamforming technology considers spatial filter to sum weighted signals from microphones, and provides enhanced Signal-to-noise Ratio (SNR) in desired fields by generating distinct responses of array output pattern in different spatial directions or points. The common used criteria of beamforming includes Minimum Mean Square Error (MMSE), Least Square (LS), maximum SNR, Linear Constrained Minimum Variance (LCMV), and maximum likelihood [7, 9]. However, the actual performance of array might be greatly different from the theoretical evaluation, due to the various errors existing in real applications, such as source location errors, quantization errors of weight, and finite sampling errors. One possible solution for this problem is accurately estimating and correcting these errors, while the other is applying robust adaptive algorithms. For example, paper [10] applies diagonal loading to correct the errors of covariance matrix, and paper [11, 12] applies eigenspace and orthogonal projection as the adaptive algorithm. The impact of array placement errors is also evaluated in paper [13, 14].

Most of these approaches for microphone array signal processing are directly derived from radar, sonar, antenna array theories with the assumption of narrow-band, far-field, and stationary data mode, which have been studied for a long period of time [15-20]. However, microphone array applications work in a different way. For example, it is known that the far field analysis is used to approximate the wave field where the distances from source to sensors are longer than $r = 2 d^2 / \lambda$. (d is the array aperture, and λ is the signal wavelength.) For some cases requiring lower sidelobes and zero nulls [21], this distance limitation needs to be extended as $r = 10 d^2 / \lambda$. If applying 8 microphones with the inter distance as $\lambda / 2$ for a 2000Hz acoustic signal, which is the most important frequency band for human understanding of speech according to the Speech Intelligibility Index (SII), the far field assumption is only applicable for the source located farther than 4.2 meters, which is not common for most applications of

microphone array in indoor environments. In addition, other main differences between microphone array problems and the traditional antenna array analyses are listed as below [2]:

- Speech is a broad band signal.

- The spectrum of noise and desired speech signal are usually overlapped.

- The number of microphone and the size of array are usually restricted by the room size.

- High reverberation or severe multipath effect could occur in the acoustic environment.

- Signals and propagation environments are highly non-stationary.

- The extremely wide dynamic range of human hearing and the hearing sensitivity for week tails require high performance of processing filter.

These differences limit the performance of traditional algorithms, which cannot work well in real acoustic problems. Therefore, more studies focus on microphone array processing in immersive environments are necessary.

From the studies in paper [24-26], it has been demonstrated that array geometry plays an important role in the formulation of processing algorithms. For a fixed number of microphones it is the dominant factor for performance. As discussed before, in order to simplify the problem of estimation and directly borrow the traditional narrow-band array processing technologies, previous works about array geometry have largely focused on regular arrays in far-field. These results are not as useful for immersive geometries that typically occur for surveillance and smart room applications. Our work, therefore, focuses on the relationship between microphone distribution properties and spatial filtering performance that is more suited for cases when the focal point is close to the arrays and the arrays have irregular placements. Classes of irregular and regular geometries for immersive environments are statistically analyzed to identify key geometric characteristics related to array performance. Then, effective optimization algorithms are proposed to optimize/design the array geometries based on the knowledge of acoustic environment providing superior SNR performance.

As mentioned before, regular arrays (elements arranged under a regular spacing constraint) have been studied for a long time, such as uniformly spaced linear, planar, and

circular arrays [2]. Due to the regularity of element arrangements, their geometries can be specified by a simple parameter set, such as aperture and number of elements or their spacings, which are directly related to aspects of performance [2, 27]. In general, most of these analyses have been done for narrow-band far-field cases where spatial aliasing is directly related to microphone spacing and resolution to aperture. Irregular arrays, which diversify microphone positions, can potentially achieve better performance as demonstrated in [24, 25, 28]. Instead of limited optimal range of signal frequencies for regular array applications, irregular arrays can result in a more consistent performance over a broader range of frequencies, such as those associated with speech [27].

Although special arrays that deviate from simple Cartesian arrangements have been studied for better performance, such as spherical arrays to better capture and render sound fields [29, 30], and minimum redundancy arrays to achieve maximum spatial resolution with fixed number of microphones [31-33], they still retain certain regularity of element placements and are restricted by previous limitations of regular arrays. This kind of array configurations mutated from traditional regular structures will be generalized and developed in Chapter 4 to meet the requirements of microphone array applications in indoor environments. Furthermore, in order to completely break the symmetry/regularity of microphone placements as in traditional regular structures, which results in severe sidelobe leakage on interferences, our study focuses on the arrays with totally random microphone distribution. We constrain the microphone geometries to a plane, but allow for any arrangement of elements and compare geometries with similar relationships to the focal point. For example, Figure 1.1 shows 3 planar arrays with the same centroid and dispersion focused on the center source 0.2 meter below the array centroid. (Dispersion is analogous to aperture.) Array gains over the FOV were computed via simulation by moving a colored noise source of unit power with speech-like frequency distribution over the grid points of the FOV, and then computing the received power from the beamformed focal point as described in [28, 34]. The microphone positions are superimposed over their array gains showing the irregular array in Figure 1.1(b) having larger gains at the non-focal points than regular array in Figure 1.1(a), while the irregular array in Figure 1.1(c) shows lower gains at non-focal points. These performance differences cannot be explained by previous analyses of regular geometries. Although paper [35, 36] have

introduced optimization approaches for irregular geometries by minimizing the residues between desired gain pattern and actual pattern computed from each microphone position, it is still not clear which geometric properties are crucial for the performance of the irregular arrays (such as aperture and element spacing for regular arrays). Direct and effective design/optimization methods for irregular geometries according to given environmental limitations are still lacking. Therefore, our study will identify the novel geometric descriptors with a relationship to performance that are useful for explaining the performance differences of irregular arrays, and further develop feasible array design methods without the restriction of regular placements.
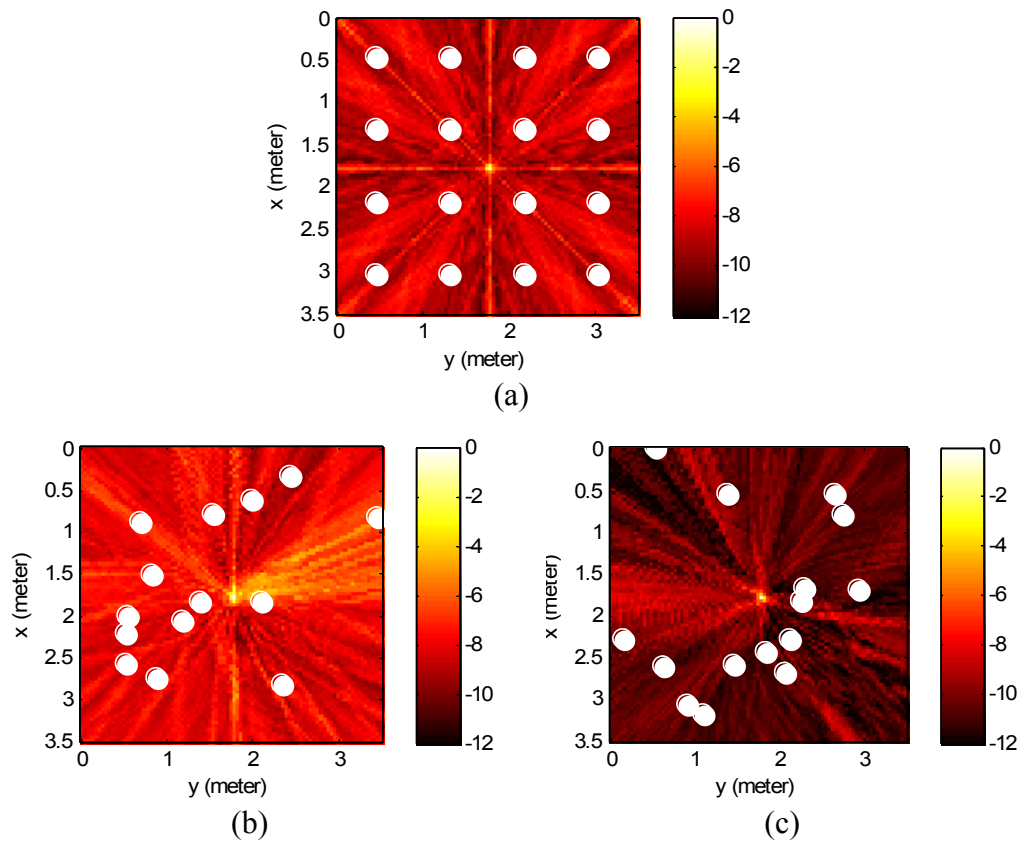
Figure 1.1: Top view gain patterns of arrays with 16 microphones. Each array has the same centroid and dispersion, focused on the same point at the center. The white dots represent microphone positions. (a) Regular planar array. (b) Irregular array with inferior performance. (c) Irregular array with superior performance.

1.2 Beamforming

Beamforming is a spatial filtering technology originally derived from antenna arrays. It can modify the enhanced signal source region without physical movement of microphone positions by combining the received signals in a coherent manner according to the target positions. Generally, it contains two processes. The first is synchronization, which adds proper compensating time delays to the signals received by elements to synchronize captured target components and decorrelate the noise components. And the second step is to weight and combine the aligned coherent signals to form the output providing enhanced SNR [1, 2]. Normally, the first process decides the steering direction of one array, and the second process controls the mainlobe width and sidelobe shape of the array gain pattern, which is the hotspot of most previous research to adaptively adjust the weighting coefficients based on the features of target and noise [2].

The effect of beamforming relates on the accuracy of sound source localization and the incoherent level of received signals from the target and interferences. The array geometry plays an important role on deciding the relevant array ability. An optimal distribution of microphone positions for certain acoustic scenario should provide stronger incoherence of received signals over all possible occurrences of source distribution, and bring enhanced SNR performance no matter which beamforming algorithm is applied. Therefore, in our work, to focus on the optimal microphone distribution, the simplest Delay and Sum Beamforming (DSB) technology using an inverse distance weighting is applied to generate the array gain pattern.

In order to reveal the impact of microphone distribution on beamforming performance, the following two sections present the formulations to compute the three dimensional array gains of microphone arrays relative to a focal point. Parametric performance metrics are directly computed from these gain patterns.

1.2.1 General Formulation of Sound Propagation and Array Capture

Consider microphones and sound sources distributed in a three dimensional space. Let $u(t;\boldsymbol{r}_s)$ be the target source located at position $\boldsymbol{r}_s$, and $n(t;\boldsymbol{r}_k)$ be the kth noise

source located at position r_k. r_s and r_k are the position vectors denoting the x, y, and z coordinates. The waveform received by the pth microphone can be expressed as:

$$v(t;r_s,r_p)=u(t;r_s)*h(t;r_s,r_p)+\sum_{k=1}^{K} n(t;r_k)*h(t;r_k,r_p) \quad, \quad (1)$$

where $h(.)$ represents the impulse response of propagation path from source to microphone. For a reverberant room, the impulse response can be given by:

$$h(t;r_s,r_p)=a_{sp0}(t-\tau_{sp0})+\sum_{n=1}^{\infty} a_{spn}(t-\tau_{spn}) \quad, \quad (2)$$

where $a_{spn}(t)$ is the response related to the nth propagation path of target signal, τ_{spn} is the corresponding time delay, and $n=0$ represents the direct path from source to microphone. In frequency domain the received signal of Eq. (1) over a finite time frame can now be expressed as:

$$\hat{V}(\omega;r_s,r_p)=\hat{U}(\omega;r_s)\hat{A}_{sp0}(\omega)\exp(-j\omega\tau_{sp0})+\hat{U}(\omega;r_s)\sum_{n=1}^{\infty}\hat{A}_{spn}(\omega)\exp(-j\omega\tau_{spn})$$
$$+\sum_{k=1}^{K}\hat{N}(\omega;r_k)\sum_{n=0}^{\infty}\hat{A}_{kpn}(\omega)\exp(-j\omega\tau_{kpn})$$
$$, \quad (3)$$

where the hat notation expresses the Fourier transform of corresponding time domain quantity.

1.2.2 The Delay and Sum Beamformer

With the purpose to suppress undesired noise terms by incoherence (the second and third additive terms on the right side of Eq. (3)), the beamformer time-shifts the received signal of each microphone according to the direct path propagation delay from the target source, then weight-sums these aligned signals to obtain the recovered target signal.

Denote the desired focal point as r_i, which is ideally identical with r_s but usually slightly dislocated due to the source localization errors. The impulse response of DSB is given by:

$$w(t;\boldsymbol{r}_i,\boldsymbol{r}_p)=B_{ip}\boldsymbol{\delta}(t+\tau_{ip0}) \quad, \tag{4}$$

Then the recovered target signal from DSB's output is computed as:

$$y(t;\boldsymbol{r}_s,\boldsymbol{r}_i)=\sum_{p=1}^{P}w(t;\boldsymbol{r}_i,\boldsymbol{r}_p)*v(t;\boldsymbol{r}_s,\boldsymbol{r}_p) \quad, \tag{5}$$

If the exact position of target source is known, $\boldsymbol{r}_i=\boldsymbol{r}_s$. Eq. (5) is rewritten as:

$$y(t;\boldsymbol{r}_s,\boldsymbol{r}_i)=u(t;\boldsymbol{r}_s)*\sum_{p=1}^{P}B_{ip}a_{sp0}(t)+u(t;\boldsymbol{r}_s)*\sum_{p=1}^{P}(B_{ip}\sum_{n=1}^{\infty}a_{spn}(t+\tau_{ip0}-\tau_{spn}))$$
$$+\sum_{k=1}^{K}n(t;\boldsymbol{r}_k)*\sum_{p=1}^{P}(B_{ip}\sum_{n=0}^{\infty}a_{kpn}(t+\tau_{ip0}-\tau_{kpn})) \quad, \tag{6}$$

where P is the total number of microphones. B_{ip} is a scalar representing filter coefficient related to focal point $\boldsymbol{r}_i$ and microphone position $\boldsymbol{r}_p$. For results in this dissertation the coefficient was set to the inverse distance from microphone to the focal point as $B_{ip}=1/d_{ip}$, $d_{ip}=\|\boldsymbol{r}_i-\boldsymbol{r}_p\|$, which is considered as a compensation to cancel out the direct-path-propagation attenuation $a_{sp0}(t)$.

It can be seen that in the recovered signal, the first term of Eq. (6) represents the desired signal component almost identical with the original target signal $u(t;\boldsymbol{r}_s)$. The second and third terms represent the noise components which are expected to be canceled out by the incoherent time delays derived from distributed microphones. Therefore, the SNR of the beamforming output can be given as:

$$SNR(\boldsymbol{r}_s,\boldsymbol{r}_i)=$$

$$\frac{\left|u(t;\boldsymbol{r}_s)*\sum_{p=1}^{P}B_{ip}a_{sp0}(t)\right|^2}{\left|u(t;\boldsymbol{r}_s)*\sum_{p=1}^{P}(B_{ip}\sum_{n=1}^{\infty}a_{spn}(t+\tau_{ip0}-\tau_{spn}))+\sum_{k=1}^{K}n(t;\boldsymbol{r}_k)*\sum_{p=1}^{P}(B_{ip}\sum_{n=0}^{\infty}a_{kpn}(t+\tau_{ip0}-\tau_{kpn}))\right|^2} \quad, \tag{7}$$

where $\overline{|.|^2}$ denotes the expected time average of signal power. If the beamforming algorithm is determined, the numerator of SNR expression is only related to the target signal power, while the denominator depending on the sums of time delays derived from microphone positions and the power ratio of target and noise sources (the input SNR of beamformer). Therefore, the objective in selecting a optimal microphone distribution can be stated as increasing the incoherent level of the time delays of microphone signals to minimize the sum of the noise terms and provide enhanced SNR performance. This optimization criterion of array geometry can be expressed as:

$$G_{opt} = \underset{G \in \text{mic space}}{\text{argmax}} \left\langle \underset{\substack{r_s \in target \\ space}}{E} \left[SNR(r_s, r_i, G) \right] \right\rangle, \tag{8}$$

where G is the set of P microphone positions representing a particular geometry. $E[.]$ denotes the probabilistic average of SNRs over all target sources with localization errors, given as [1]:

$$\underset{\substack{r_s \in target \\ space}}{E} \left[SNR(r_s, r_i, G) \right] = \int_{r_s \in \text{target space}} \left\{ \int SNR(r_s, r_i, G,) p(r_i | r_s) d\, r_i \right\} p(r_s) d\, r_s, \tag{9}$$

where $p(r_s)$ are the probability density functions representing interested target region. And $p(r_i|r_s)$ are the probability density functions for dislocating focal point at r_i when the actual interested source located at r_s. This criterion will be applied later in this dissertation for our works of array optimization, which usually set the goal to find microphone geometries maximizing beamformer SNR for given distribution of target and specified acoustic scenes.

1.2.3 Generating and measuring 3D Beampatterns

The traditional far-field analyses of array pattern only apply one dimensional parameter of the steering angle of beamformer, which cannot resolve the sound waves from collinear or rotationally symmetrical sources with respect to the array center. In order to analyze the beamforming behavior of microphone array in common three dimensional space, the three dimensional beampattern as the function of the Cartesian coordinates is generated and assessed.

9

Because our major concern is the array ability to capture and extract interested human speech from other interferences, which is the common scenario for the immersive applications such as audio surveillance and cocktail party, the source signal is simulated by a colored noise with a spectrum equivalent to the band importance function from SII, which emphasizes the frequency bands most important to human understanding of speech [26]. The applied SII band importance function is shown in Figure 1.2.
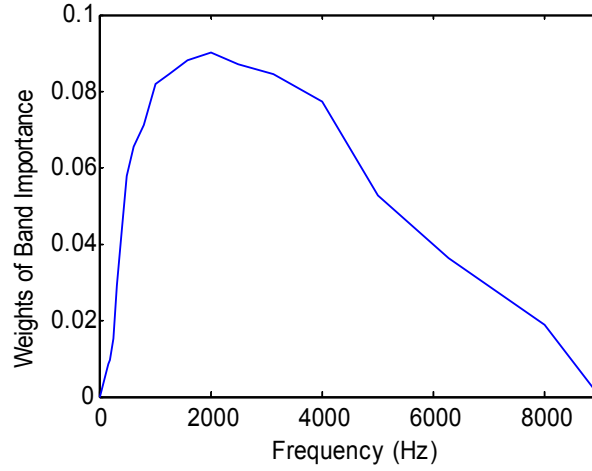
Figure 1.2: SII band importance spectrum

The 3D beampattern of an array geometry is generated by moving a sound source with constant power over all spatial points in FOV, while the focal point is fixed at the target. The DSB output power is computed and normalized as the beamforming gain for the source at each spatial point to form the 3D beampattern. In order to ensure that no more than a 3dB change occurred at adjacent spatial points [26, 37], the resolution of beampattern should be chosen smaller than

$$\Delta_{grid} = \frac{0.4422\,c}{\sqrt{d}\,f_{max}} = \frac{0.4422}{\sqrt{d}}\lambda_{min} \quad , \tag{10}$$

where d is the dimension of FOV (3 for our case). f_{max} is the highest target signal frequency (or one that bounds most of the relevant energy), and λ_{min} is the corresponding wavelength. In our later study of array beampattern, the size of spatial grid in FOV is chosen as 0.04m covering the highest peak of SII band importance spectrum as 2000Hz.

In addition, paper [26] has indicated that for generating beampattern via DSB in a small room, where sound attenuation through air can be neglected according to the 3dB threshold of beamforming gain variation for the spatial grids, the operation of holding the focal point and moving a sound source in FOV is equivalent with the operation of holding the sound source and moving the focal point in FOV. Because the second operation avoids repeating the simulation of signal propagation from each source, which is very time consuming, it is the actual operation applied in our experiments to generate the beampattern for a specified array geometry.

In order to visually inspect the beamfield of one microphone array, paper [26] provides methods to visualize the three dimensional beampattern in a Cartesian coordinate system by applying intensity-dependent transparency on the volumetric plot. A Matlab GUI is developed based on these methods giving the audiences a direct view and corresponding measures about whether this array geometry have superior beamforming ability for the specified target space, as shown in Figure 1.3.
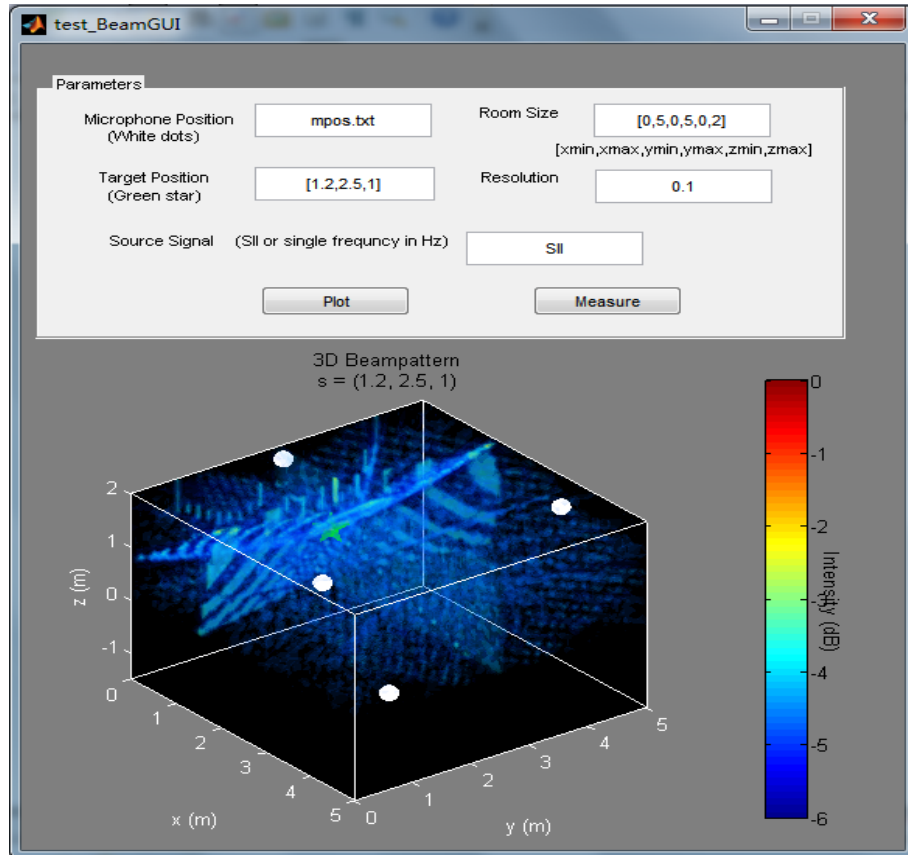
Figure 1.3: Matlab GUI of volumetric beampattern plot through DSB

Two metrics are usually applied to quantitatively assess the array beamforming performance based on the spatial gain pattern of DSB: Mainlobe Width (MLW) associated with resolution, and Mainlobe-to-peak-sidelobe Ratio (MPSR) associated with noise suppression ability [34]. In our works, the size of mainlobe is characterized by the dimensions of the surface consisting of spatial points with gains 3 dB below that at the focal point (maximum gain). Let x_δ, y_δ, and z_δ denote the projections of the 3dB mainlobe contour onto the x, y, and z axes, respectively. The MLW can then be expressed as:

$$B_{3dB} = \sqrt{(x_\delta^2 + y_\delta^2 + z_\delta^2)} \quad , \tag{11}$$

Let $S(\boldsymbol{r}_i, \boldsymbol{r}_s)$ denote the power gain of the beamformer focused on $\boldsymbol{r}_i$ with a unit power source at $\boldsymbol{r}_s$. The MPSR for a beamformer focused on $\boldsymbol{r}_i$ can be expressed as:

$$\Gamma_i = \frac{S(\boldsymbol{r}_i, \boldsymbol{r}_i)}{\max_{\boldsymbol{r}_s \notin ML} \left[S(\boldsymbol{r}_i, \boldsymbol{r}_s) \right]} \quad , \tag{12}$$

where the maximum is taken over all possible source positions $\boldsymbol{r}_s$ outside the 3dB mainlobe region (ML) in FOV. This metric represents the worst case leakage. Because there is normally a tradeoff between B_{3dB} and Γ_i, the common criterion to decide the optimal array beampattern is to limit the MLW to a tolerable spatial resolution and maximize the MPSR in FOV. In the later chapter of this thesis, statistical analyses of Monte Carlo simulation results will be presented to assess the impacts of important geometric properties on these performance metrics and demonstrate their relationship in immersive or near-field applications to guide the optimization of microphone array geometry.

1.3   Hardware Resources of Audio Lab

All our works about the distributed microphone arrays in immersive environment are developed at the Audio Lab in the Center for Visualization and Virtual Environments of the University of Kentucky. The main focus of this lab is to develop technologies for enhancing and extending distributed audio system applications, such as the smart room and audio surveillance system. The hardware resources applied in the study of geometry

optimization involve the audio recording system conducting the experimental evaluation, and the supercomputer cluster which performing algorithm development, signal processing, and simulations.

The audio recording system includes a audio cage and a audio server. As shown in Figure 1.4, all our experiments are performed in this 3.58x3.58x2.29m aluminum cage, where the absorbing foam walls can be mounted on the cage's faces to change the reverberation characteristics. Microphone capsules, composed of 6mm omnidirectional Panasonic electret condenser microphones (WM-61A), can be attached to the Velcro panels of the ceiling or mounted on the beams of the cage to form arbitrary array geometry. The audio server can operate up to 48 channels of data from/to the distributed microphone/speakers. The recording sample rate is up to 96 kHz and the resolution up to 24 bit. Other detailed information of applied equipments is shown as below.

- Low-noise computer
- Ubuntu Studio Linux Distribution
- 2 RME HDSP9652 Sound Cards
- 6 RME Octamic-D Microphone Preamps
- 6 RME ADI 8-DS A/D-D/A Converters
- Apogee Big Ben Reference Clock Generator
- JACK Audio Server

The computer cluster consists of 11 Dell Poweredge 2950 computers in a ROCKS Linux cluster configuration, providing 88 CPUs, 188G RAM and 1.5TB disk space. The installed Sun Grid Engine enables us to run parallel jobs in the cluster, which greatly reduces the time consuming of Monte Carlo experiments and makes the real-time processing become possible for high quality audio signals.
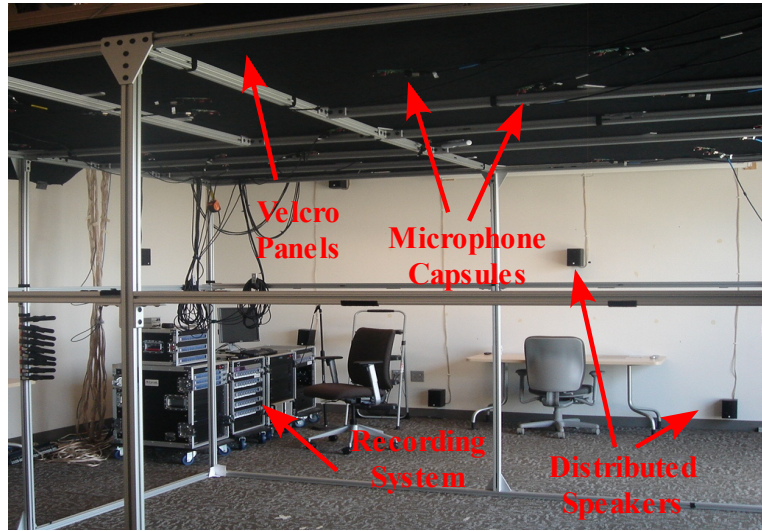
Figure 1.4: Photos of audio cage and recording system

1.4   Organization of Dissertation

The dissertation is organized as follows: Chapter 2 analyzes important geometric features of microphone distribution related to irregular array beamforming performance. The relationships between proposed geometry descriptors and key performance metrics are concluded as general guidelines for the irregular array design in various scenarios. Chapter 3 presents the basic principles and criteria for the microphone array optimization procedure. Chapters 4 generalizes and develops the array design methods based on the mutations of regular geometry. Chapter 5 proposes novel optimization methods for the irregular geometry design. Finally, the conclusions and direction of further research are presented in Chapter 6.

Chapter 2   Geometry Factors related to Beamforming Performance

It has been introduced in Chapter 1 that some irregular arrays show better noise suppression ability than regular arrays. Performance analysis for arrays with irregular geometries typically requires direct computation of beamforming gains over the spatial and frequency ranges of interest. The computation can be very time consuming and limit synthesis methods for applications that require rapid answers, as in the case of surveillance and mobile platforms. A better understanding of microphone arrangements and their impact on performance can result in more efficient objective functions for optimizing array performance. This chapter, therefore, analyzes the relationship between irregular microphone geometries and spatial filtering performance with Monte Carlo simulations. Novel geometry descriptors are developed to capture the properties of both regular and irregular microphone distributions showing their impact on array performance. Performance metrics are computed from three-dimensional beampatterns through DSB with a fixed number of microphones for irregular arrays and comparable regular arrays. Statistical analysis and Multi-way Analysis of Variance (ANOVA) establish relationships between key performance metrics and proposed geometry descriptors, which can be applied as the objective functions for the optimization methods of microphone array and provide the general guideline and insight for the irregular microphone cluster design.

2.1 Problem Formulation

As discussed in section 1.2.1 and 1.2.2, the optimal array with enhanced beamforming SNR should increase the incoherent level of the time delays provided by distributed microphones to decorrelate the noise components. In order to further explore the noise suppression ability of microphone arrays on specified source distributions, the received signal of *pth* microphone (as Eq.(1)) can be decomposed as the superposition of the signal from each source (including the target and noises) in FOV. In frequency domain, the received signal derived from source $u(t;r_s)$ at r_s is expressed as:

$$\hat{V}(\omega;r_s,r_p)=\hat{U}(\omega;r_s)\hat{A}_{sp0}(\omega)\exp(-j\omega\tau_{sp0})+\hat{U}(\omega;r_s)\sum_{n=1}^{\infty}\hat{A}_{spn}(\omega)\exp(-j\omega\tau_{spn})$$

$$, \quad (13)$$

where the hat notation expresses the Fourier transform of corresponding time domain quantity. And $u(t;r_s)$ can represent arbitrary sound source in FOV. Denote the desired focal point as r_i and express the DSB output as:

$$\hat{G}(\omega;r_i,r_s)=\sum_{p=1}^{P}B_{ip}\hat{V}(\omega;r_s,r_p)\exp(j\omega\tau_{ip}); \quad , \quad (14)$$

where P is the total number of microphones. B_{ip} is a scalar representing the filter coefficient, which in this dissertation was set to the inverse distance from r_i to r_p. The total output power of this filtered sum is computed by:

$$S(r_i,r_s)=\int\sum_{p=1}^{P}\sum_{q=1}^{P}B_{ip}B_{iq}\hat{V}(\omega;r_s,r_p)\hat{V}^*(\omega;r_s,r_q)\exp(j\omega(\tau_{ip}-\tau_{iq}))d\omega \quad , \quad (15)$$

In order to obtain simplified formulation that is useful for analysis and understanding the geometric relationship, consider only the direct paths in Eq. (13). With the assumption that the beamformer coefficients and propagation attenuation product factors are uncorrelated with the path differentials, $S(r_i,r_s)$ can be rewritten as:

$$S(\boldsymbol{r}_i,\boldsymbol{r}_s)=P^2\int|\hat{U}(\omega;\boldsymbol{r}_s)|^2\mathrm{E}[B_{ip}B_{iq}\hat{A}_{sp}(\omega)\hat{A}_{sq}^{*}(\omega)]$$
$$\mathrm{E}[\exp(j\omega((\tau_{sq}-\tau_{sp})+(\tau_{ip}-\tau_{iq})))]d\omega \quad , \qquad (16)$$

where E[·] denotes the expected value operator over all microphone pairs generated by the double summation of Eq. (15), and $S(\boldsymbol{r}_i,\boldsymbol{r}_s)$ is the output power of beamformer targeting $\boldsymbol{r}_i$ with actual sound source at $\boldsymbol{r}_s$. Therefore, considering all the sources in FOV, the total output power of beamformer can be obtained from the superposition of $S(\boldsymbol{r}_i,\boldsymbol{r}_s)$ of each source. When $\boldsymbol{r}_s=\boldsymbol{r}_i$, $S(\boldsymbol{r}_i,\boldsymbol{r}_i)$ represents the power component at beamformer output derived from the target source. And $S(\boldsymbol{r}_i,\boldsymbol{r}_s)$, $\boldsymbol{r}_s\neq\boldsymbol{r}_i$, represents the power derived from the noise source.

To investigate the beamforming performance in relation to array geometry, the time delays are expressed in terms of spatial distances and signal wavelengths:

$$S(\boldsymbol{r}_i,\boldsymbol{r}_s)=P^2\int|\hat{U}(\omega;\boldsymbol{r}_s)|^2\mathrm{E}[B_{ip}B_{iq}\hat{A}_{sp}(\omega)\hat{A}_{sq}^{*}(\omega)]$$
$$\mathrm{E}[\exp(j2\pi(\frac{d_{sq}-d_{sp}}{\lambda}+\frac{d_{ip}-d_{iq}}{\lambda}))]d\omega \quad , \qquad (17)$$

where d_{sp} denotes the distance from sound source $\boldsymbol{r}_s$ to microphone position $\boldsymbol{r}_p$, and d_{ip} denotes the distance from focal point $\boldsymbol{r}_i$ to microphone position $\boldsymbol{r}_p$. Therefore, the formulation of beamforming gain for sources in FOV is separated into three parts; the signal power in the time window, the propagation environment and beamforming algorithm, the microphone distributions which is useful for array geometry design. For arrays with fixed number of microphones and constant coefficients of beamformer, $S(\boldsymbol{r}_i,\boldsymbol{r}_s)$ only depends on the average of exponential terms over all microphone pairs derived from the microphone positions and source signal frequencies. For the case where a signal source is located at the beamformer focal point, $\boldsymbol{r}_s=\boldsymbol{r}_i$, the arguments of the exponents are 0 and the signal is enhanced by the coherent addition of complex exponential terms. Sources not located at the focal point, $\boldsymbol{r}_s\neq\boldsymbol{r}_i$, will have reduced power due to the incoherent phases of exponential terms. The objective in selecting a microphone distribution is to minimize the average value of the exponential terms in Eq. (17) when $\boldsymbol{r}_s\neq\boldsymbol{r}_i$ while maximizing the average when $\boldsymbol{r}_s=\boldsymbol{r}_i$ for all possible target and noise positions in the FOV. If the weights of the DSB are fixed, the

summations will always be maximized when $r_s = r_i$ (exponential arguments are all zero). Therefore, a more practical optimization strategy would be stated as minimizing the maximum value of $S(r_i, r_s)$ when $r_s \neq r_i$ for all r_s and r_i in the FOV. The metric based on this notion, MPSR, will be used in later simulations to assess performance.

Eq. (17) identifies the phase terms responsible for minimizing the power gain when $r_s \neq r_i$, related to the source wavelength and the Differential-path-distance (DPD) distribution of all (p, q) microphone pairs given by:

$$\Delta_{pq}(r_i, r_s) = (d_{sq} - d_{sp}) + (d_{ip} - d_{iq}) \quad, \tag{18}$$

where r_i is the focal point of beamformer (target position), and r_s is the interfering source position. Note that $\Delta_{pq}(r_i, r_s)$ is the exponential argument in Eq. (17) without the wavelength scaling.

Ideally, if the DPDs of a given microphone geometry result in the complex exponential arguments distributing uniformly from $-\pi$ to π over all pairwise microphones, the expected power is zero when targeting r_i. That is to say, in order to minimize gains for the interference/noise sources $(r_s \neq r_i)$, the corresponding DPDs should be distributed as widely as possible relative to the source wavelength (incoherence). For the case beamforming at the desired source, all the phase terms in Eq. (17) will be close to zero (coherent), and result in a maximum power gain in the target position. Even if the sound source localization errors might bring some dislocation between the target source position and the focal point of beamformer, $r_s = r_i + \Delta r_{error} \approx r_i$, because the variance of DPDs derived from the localization errors is usually much smaller than the wavelengths of significant speech signal frequencies, the phases of exponential arguments are still limited to a small range and result in significant coherent sums. Therefore, Eq. (17) demonstrates the impact of the DPD distribution over all microphone pairs on the array's ability to enhance target and suppress noise signals. The beamforming gain for each source when steering at target is related to the spread and uniformity of corresponding DPD distribution. The optimal microphone geometry should provide a widely spread and even distributed DPDs for the noise source positions to decorrelate the noise from target

18

signals. Statistics assessing the uniformity of DPD distributions are proposed in the next section as the novel geometric descriptors to explain the variations of array beamforming performances, especially for irregular arrays.

2.2 Proposed Geometry Descriptors related to Performance

Analysis in previous section suggests a correlation between array beamforming gains and microphone distributions. This section proposes several geometric characterizations applicable to irregular arrays and related to array performance. In addition, descriptors for regular arrays, such as the aperture size and microphone spacings, will be generalized for irregular array geometries.

2.2.1 Centroid Offset and Dispersion

Derived from regular array research, the important array properties impacting performance are its distance from the focal point (determining near/far field behaviors) and the spread of its elements (related to aperture). The array centroid offset is defined as the distance between array focal point $r_i = (x_i, y_i, z_i)$ and the centroid of array elements given by [34]:

$$L = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2 + (z_0 - z_i)^2} \quad , \tag{19}$$

where $r_0 = (x_0, y_0, z_0)$ denotes array centroid as:

$$r_0 = (x_0, y_0, z_0) = \left(\frac{1}{P} \sum_{p=1}^{P} x_p, \frac{1}{P} \sum_{p=1}^{P} y_p, \frac{1}{P} \sum_{p=1}^{P} z_p \right) \quad , \tag{20}$$

where P is the number of microphone and $r_p = (x_p, y_p, z_p)$ denotes the position of the *pth* microphone.

Array dispersion, analogous to the aperture size, is a measure of average microphone spread about the centroid, computed by [34]:

$$a = \sqrt{\frac{1}{P} \sum_{p=1}^{P} [(x_p - x_0)^2 + (y_p - y_0)^2 + (z_p - z_0)^2]} \quad , \tag{21}$$

19

Note that L and a can be applied to characterize both regular and irregular geometries, as shown in Figure 1.1. For regular arrays a directly impacts resolution (MLW), and determines the microphone spacing in conjunction with P, which affects sidelobe behavior. The distance L indicates whether sound sources are effectively located in the near field (*small L* for immersive application), or far field (*large L*), where the terms *small* and *large* are used relative to the source wavelengths. The typical near and far fields derived from antenna array are shown in Figure 2.1 [38], where D represents array aperture analogous to dispersion defined in Eq. (21). For acoustic fields with microphone arrays, there will be no non-radiative reactive zone considered. Array beamforming behaviors according to microphone distributions will change greatly when moving from near field (Fresnel zone) to far field. In the research of this dissertation focusing on the broadband human speech applications, which is simulated by the colored noise generated by the SII band importance spectrum, L_0 approximately equals to 3 times of array dispersion. When the distances from sources to array centroid are larger than L_0, the array takes on more characteristics of a far-field application, and the impact of proposed geometry descriptors on array performance changes dramatically. Relative analysis will be given in Section 2.3.
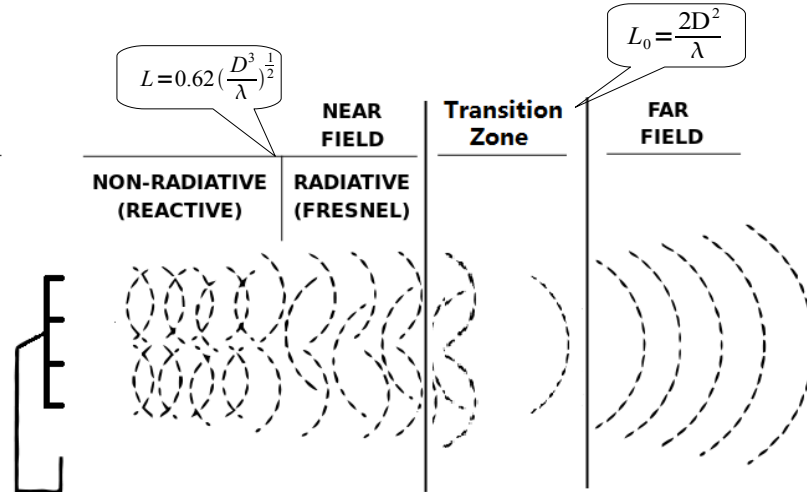
Figure 2.1: Typical antenna field regions (adapted from [38])

2.2.2 Statistics of Differential-path-distance

Traditional descriptors L and a can be applied to characterize both regular and irregular geometries. Arrays with the same P, L and a are usually considered as the same class to compare the performance of regular and irregular geometries. However, as illustrated by the examples in Figure 1.1, these descriptors are limited in their ability to explain the beamforming behavior when additional degrees of freedom are allowed as in the case of irregular arrays. Therefore, additional descriptors involving DPD distribution for all microphone pairs to points in FOV are proposed as metrics.

From the analysis of Section 2.1, a limited DPD distribution increases the likelihood of unexpected coherence at non-target locations, especially when DPDs are less than a quarter wavelengths at significant signal frequencies. DPD distributions can be examined via histograms or characterized with various statistics. One potentially useful statistic is the standard deviation of the DPDs over all microphone pairs. In [39, 40] closed form expressions were presented for the expected value of the exponential terms in Eq. (17). With a normal distributed DPDs over all microphone pairs, the expected value of the exponential terms is given by:

$$E\left[\exp\left(j2\pi\left(\frac{\Delta_{pq}(\boldsymbol{r}_i,\boldsymbol{r}_s)}{\lambda}\right)\right)\right]=\exp\left(-2\left(\pi\frac{\sigma_\Delta(\boldsymbol{r}_i,\boldsymbol{r}_s)}{\lambda}\right)^2\right) \quad , \qquad (22)$$

where σ_Δ presents the DPD standard deviation. If the DPDs are uniformly distributed, the expected value becomes

$$E\left[\exp\left(j2\pi\left(\frac{\Delta_{pq}(\boldsymbol{r}_i,\boldsymbol{r}_s)}{\lambda}\right)\right)\right]=\mathrm{sinc}\left(\pi\frac{\sqrt{12}\,\sigma_{\Delta(r_i,r_s)}}{\lambda}\right) \quad , \qquad (23)$$

In both cases the expected value of the exponential terms approaches zero for increasing σ_Δ . When $\boldsymbol{r}_i=\boldsymbol{r}_s$, the DPDs are zero for all microphone pairs resulting a DPD variance of 0. Thus, the scaling provided by the DPD exponential factor of Eq. (17) is at a maximum of 1, which is desired when the source and focal point are identical. Consistent with previous conclusion, the more widely spread of DPDs (large σ_Δ), the better ability of the array to extract target signal at $\boldsymbol{r}_i$ and decorrelate signals from noise

source at $\boldsymbol{r}_s$. Therefore, standard deviation can be applied as an effective measure that describes performance of irregular arrays. For particular focal and noise source locations, the DPD standard deviation is defined as:

$$\sigma_\Delta(\boldsymbol{r}_i, \boldsymbol{r}_s) = \sqrt{\frac{1}{P^2} \sum_{p=1}^{P} \sum_{q=1}^{P} \left(\Delta_{pq}(\boldsymbol{r}_i, \boldsymbol{r}_s) \right)^2} \quad , \tag{24}$$

In addition, with the same standard deviation, the expected value of the exponential terms approaches zero for decreasing λ , representing better noise suppressing ability for the signals in high frequency bands. Wider spread of DPDs are needed to decorrelate the signal source with low frequencies, such as in the case of male voices. In this dissertation, in order to focus on the impact of DPDs derived from array geometry, colored noise generated by SII mode is applied as the excitation of the simulations to compute performance metrics.

From Eq. (22) and Eq. (23), different DPD distributions can also impact the incoherence level of beamforming. Figure 2.2 provides a real case example of linear arrays. Figure 2.2(a) shows 2 linear arrays in a planar FOV with microphone positions denoted by O markers. Two sound sources represented by X markers are located in the FOV, while one source is considered as the target (focal point of beamformer) and the other is the noise source. Colored noise from each source is recorded separately, and the received signals of microphones are normalized by the average rms value over all channels and superimposed. The SNR is computed as the power ratio of beamformed signal from target source over that from noise source. The DPD histograms of both arrays are shown in Figure 2.2(b) and Figure 2.2(c), respectively. The beamforming SNR results are provided in Table 2.1. An analogous simulation of the array recording was also performed and presented in this table. For both the real and simulated recordings it can be seen that although these two arrays have the same σ_Δ , array 2 shows a 2 to 3 dB SNR improvement over array 1 for both targets due to the reason that array 2 provides a more uniform DPD distribution over the source spectrum, thus demonstrating a need for another statistic related to DPD diversity. Therefore, Pielou's evenness index [41], which is a normalized Shannon entropy, is introduced to numerically assess the diversity of DPD distribution as:

$$J(\boldsymbol{r}_i, \boldsymbol{r}_s) = \frac{H(\boldsymbol{r}_i, \boldsymbol{r}_s)}{H_{max}(\boldsymbol{r}_i, \boldsymbol{r}_s)} = \frac{-\sum_{k=1}^{K}(p_k \ln p_k)}{-\sum_{k=1}^{K}(\frac{1}{K}\ln(\frac{1}{K}))} = \frac{-\sum_{k=1}^{K}(p_k \ln p_k)}{\ln K}, \qquad (25)$$

where K is the total number of DPD bins for the histogram estimate, and p_k is the percentage of DPDs within the kth bin, $H(\boldsymbol{r}_i, \boldsymbol{r}_s)$ is the Shannon entropy, and $H_{max}(\boldsymbol{r}_i, \boldsymbol{r}_s)$ is the maximum possible entropy for the given number of bins, which represents an ideal uniform distribution of DPDs. This normalization avoids the variations from different ranges of DPD distributions and different numbers of microphones. Note that, the DPD range is binned by constant intervals whose size should be associated with the quarter wavelengths of significant signal frequencies to result in reasonably smooth histograms of DPDs related to the incoherent level of phase terms of beamforming gain. For the results in this dissertation bin size is set to 0.1 meter, which is less than a quarter wavelength of the important frequency band around 800 Hz for male voice intelligibility [42].
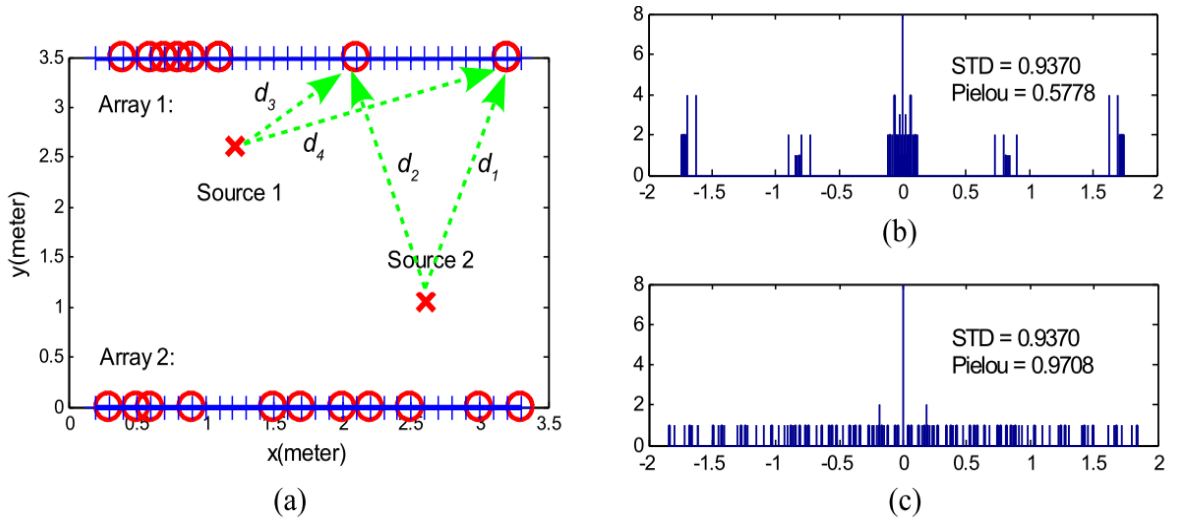
Figure 2.2: Linear arrays with the same standard deviations of DPD distributions. O's represent microphones. X's represent sound sources. (a) FOV diagram. (b) Histogram of DPDs for array 1 with lower entropy distribution. (c) Histogram of DPDs for array 2 with higher entropy distribution.

Table 2.1: SNR results of linear arrays

		Target at Source1	Target at Source 2
Simulations	Array 1	7.70 dB	7.28 dB
	Array 2	10.20 dB	9.11 dB
Real Recordings	Array 1	4.10 dB	3.52 dB
	Array 2	6.33 dB	6.48 dB

2.2.3 Interrelations of Descriptors

Therefore, four geometry descriptors $\{L, a, \sigma, J\}$ are proposed to characterize both regular and irregular microphone distributions and show their impact on array beamforming performance. As summarized in Table 2.2, these descriptors depend on various geometric aspects of the application environment. Descriptors $\{L, a\}$ are related to microphone coordinates or beamforming focal point. They are usually applied together as a basis for comparing similar arrays. The descriptors $\{\sigma, J\}$ vary with each array geometry instance and also depend on the characteristics of possible target and noise source distributions. This dependency brings the expectation of stronger correlation with array performance based on different acoustic scenes.

Table 2.2: Dependencies of geometry descriptors

	Mic Coordinates	Target Space	Noise Space
L	$\checkmark$	$\checkmark$	
a	$\checkmark$		
σ	$\checkmark$	$\checkmark$	$\checkmark$
J	$\checkmark$	$\checkmark$	$\checkmark$

In addition, because $\{L, a\}$ and $\{\sigma, J\}$ are both related to the microphone positions, they are not independent with each other. The interactions among them should also be studied and considered when analyzing array performance. In order to show a clear

relationship between DPD statistics and traditional descriptors $\{L, a\}$, consider array centroid as the origin of coordinate system. Assuming microphones are uniformly distributed about centroid, the coordinates of pth microphone can be expressed as $\boldsymbol{r}_p = (x_p, y_p, z_p)$, where $x_p, y_p, z_p \sim U(0, a^2)$. The DPD for microphone pair (p, q) associated with target $\boldsymbol{r}_i$ and noise source $\boldsymbol{r}_s$ can be given as:

$$\Delta_{pq}(\boldsymbol{r}_i, \boldsymbol{r}_s) = (d_{sq} - d_{sp}) + (d_{ip} - d_{iq}) = (|\boldsymbol{r}_s - \boldsymbol{r}_q| - |\boldsymbol{r}_s - \boldsymbol{r}_p|) + (|\boldsymbol{r}_i - \boldsymbol{r}_p| - |\boldsymbol{r}_i - \boldsymbol{r}_q|) \quad , (26)$$

Then, by applying multi-variable Taylor series expansion about origin in Cartesian coordinate system to approximate the distance from microphone to source point, Eq. (26) can be simplified as:

$$\Delta_{pq}(\boldsymbol{r}_i, \boldsymbol{r}_s) \approx (\frac{x_s}{|\boldsymbol{r}_s|} - \frac{x_i}{|\boldsymbol{r}_i|})(x_p - x_q) + (\frac{y_s}{|\boldsymbol{r}_s|} - \frac{y_i}{|\boldsymbol{r}_i|})(y_p - y_q) + (\frac{z_s}{|\boldsymbol{r}_s|} - \frac{z_i}{|\boldsymbol{r}_i|})(z_p - z_q) \quad , (27)$$

Therefore, for a specified source pair of target and noise where $\boldsymbol{r}_i$ and $\boldsymbol{r}_s$ are fixed, DPD can be approximated by a linear function of microphone coordinates. The mean and standard deviation of the DPD distribution containing all microphone pairs can be computed as

$$E(\Delta(\boldsymbol{r}_i, \boldsymbol{r}_s)) = (\frac{x_s}{|\boldsymbol{r}_s|} - \frac{x_i}{|\boldsymbol{r}_i|}) E(x_p - x_q) + (\frac{y_s}{|\boldsymbol{r}_s|} - \frac{y_i}{|\boldsymbol{r}_i|}) E(y_p - y_q) + (\frac{z_s}{|\boldsymbol{r}_s|} - \frac{z_i}{|\boldsymbol{r}_i|}) E(z_p - z_q) = 0$$

$$, (28)$$

$$\sigma^2(\Delta(\boldsymbol{r}_i, \boldsymbol{r}_s)) = (\frac{x_s}{|\boldsymbol{r}_s|} - \frac{x_i}{|\boldsymbol{r}_i|})^2 \sigma^2(x_p - x_q) + (\frac{y_s}{|\boldsymbol{r}_s|} - \frac{y_i}{|\boldsymbol{r}_i|})^2 \sigma^2(y_p - y_q) + (\frac{z_s}{|\boldsymbol{r}_s|} - \frac{z_i}{|\boldsymbol{r}_i|})^2 \sigma^2(z_p - z_q)$$

$$, (29)$$

$$\sigma^2(\Delta(\boldsymbol{r}_i, \boldsymbol{r}_s)) = 4 a^2 (1 - \frac{\boldsymbol{r}_s \cdot \boldsymbol{r}_i}{|\boldsymbol{r}_s||\boldsymbol{r}_i|}) = 4 a^2 (1 - \cos\theta_{r_s r_i}) \quad , \tag{30}$$

Eq. (30) shows a general relation between array dispersion and the standard deviation of DPDs, where the proportion related to the angle between target and noise source

position vectors. Arrays with widely spread microphones have more chance to generate a high incoherent level between received target and noise signals (large σ). With the same array dispersion, large spread of DPDs (large σ) results from large directivity difference between target and noise positions according to the array centroid, while independent with the distances from each source to the array. This interrelationship limits the capability of DPD standard deviation to explain the performance variance between similar regular and irregular geometries classified by $\{L, a\}$.

Another DPD statistic J, as the entropy of DPD diversity, shows more independence with a, which means increasing the spread of microphones does not necessarily enrich the diversity of DPD distribution or further result in better ability to suppress noise. Take the linear arrays in Figure 2.2(a) for example, which have different noise suppression abilities. Figure 2.3 shows the variation of DPD statistics when scaling microphone coordinates to increase dispersion. It can be seen that even with different array geometries, the σ values of bad and good arrays are similar and changed together with the values of dispersion, which cannot reflect the difference of performance. However, J is not sensitive to the change of array dispersion and still showing relative stable difference of DPD diversity between these two arrays, and can better explain different noise suppression abilities in this case. Therefore, the entropy based statistic J is not restricted by the value of array dispersion, which are expected to better explain the performance differences for the class of irregular and regular arrays with similar traditional descriptors.
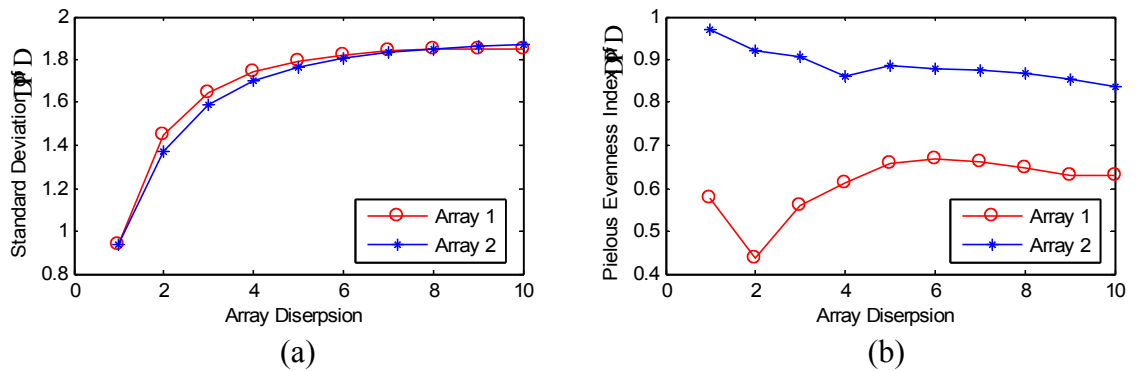
Figure 2.3: DPD statistics with different dispersions of scaled linear arrays. Array 2 has superior noise suppression ability. (a) DPD standard deviation. (b) Pielou's evenness index of DPD.

26

The next section presents further discussion about the impact of proposed descriptors on array performance related to the acoustic environment, such as field regions and the number of microphone. The proposed geometry descriptors are applied to characterize different stochastic array geometries, and their relationships with key performance metrics of 3D beampattern are analyzed with Monte Carlo simulations.

2.3 Impacts of Geometry Descriptors on Performance

To reveal the relationship between proposed geometry descriptions and array performance, experiments are performed using Monte Carlo simulations to analyze 3D beampatterns by uniformly distributed microphones over a planar design space. The simulation flow chart is shown in Figure 2.4.
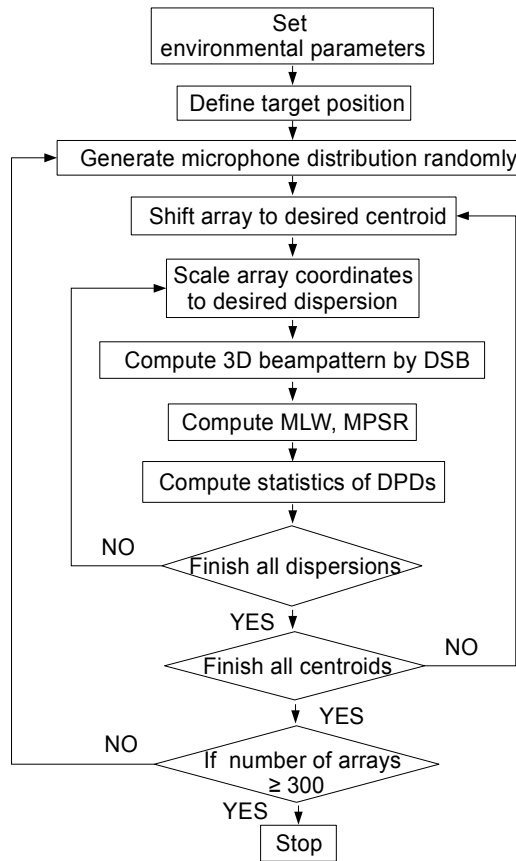
Figure 2.4: Flow chart of Monte Carlo experiments

27

The FOV is a 10x10x2 m^3 room and microphone positions are randomly generated with a uniformly distribution on the ceiling plane. Then the microphone coordinates are shifted and scaled to obtain desired array centroid and dispersion. For immersive environment or near-field study, array centroid values range from the center of ceiling to the edge at 1 meter intervals along x axis, while 5 levels of dispersion are applied with each centroid. For each combination of centroid offset and dispersion level, 300 independent arrays are generated by Monte Carlo experiments. The 3D beampattern of each array is obtained by moving a sound source with constant power over all spatial points in FOV, while the focal point is fixed in the center of room. The DSB output power is computed for the source at each spatial point in FOV to form the 3D beampattern. As presented in Section 1.2.3, two metrics are applied to assess array performance, MLW associated with resolution and MPSR associated with noise suppression ability. Their relationship with proposed geometry descriptors is established through statistical analyses, which will be applied directly to guide *ad hoc* (not computer aided) optimal microphone placements in immersive environment in the Section 2.4.

Since the main applications considered for our study involve speech (as in the case of surveillance in a cocktail party environment), the excitation of the arrays need to compute the beampattern and performance metrics is colored noise with the same spectral distribution as the band importance function used SII. This provides a compact summary statistic that is relevant for application where speech intelligibility is important. Because the impact of each geometry descriptor also depends on microphone number, irregular arrays with 16, 25, 36, 49 and 64 microphones are examined with comparable regular arrays and logarithmic arrays. The logarithmic array consists of 3 superimposed regular subarrays used for octaves from 800Hz~3200Hz to generate a relative uniform frequency response over the important frequency bands. Statistical analyses of simulation results are presented in the next two sections to assess the impacts of proposed geometry descriptors and demonstrate their relationship with performance metrics in immersive or near-field applications. For comparison, these relationships in far field are also studied in the same manner by assess the arrays with centroid offset over 10 times of the longest important wavelength of 300Hz signal. (The high-pass pre-filter of beamformer to suppress background noise are usually set to 300Hz).

2.3.1 Relationship Plots

Plots from Monte Carlo simulations are presented to reveal relationships between each geometry descriptor and performance metrics. Figures 2.5-2.8 present the geometry descriptors versus MLW and MPSR, where the error bars span ± one standard deviation about the mean. For comparison sake a regular planar array and logarithmically spaced array with the same geometry descriptors are also marked in the figures.

Figure 2.5 indicates the impact of centroid offset on array performance. From Figures 2.5(a)(b), it can be seen that for fixed array dispersion, increasing the centroid offset increases the MLW and reduces MPSR, representing degradation of array performance. The standard deviation of MLW increases with the growing of centroid offset, while ±1dB variance of MPSR is observed for each centroid offset value with fixed dispersion. Logarithmic arrays show much larger increases in MLW than regular and irregular arrays because the microphone density is high near array centroid causing a longer mainlobe in the direction of the offset. Although better MPSR can be observed for logarithmic arrays with large centroid offset, it does not necessarily represent superior ability to suppress non-target sources. The lower sidelobe levels are primarily the result of FOV being included in a huge mainlobe. Therefore, logarithmic array has a major limitation on target space, and cannot adjust well to focal points away from array centroid. Figures 2.5(c)(d) show variations of performance metrics along centroid offset when dispersion is fixed at a small value. For the centroid offset values below 2.5m, the trends of MLW and MPSR over centroid offset levels are as expected with more sensitivity for arrays with smaller dispersion (microphones closer together on average) when compared to Figures 2.5(a)(b). For the centroid offset values beyond 2.5m (exceeding 5 times that of the dispersion), the MLW becomes very large relative to the size of FOV. The apparent improvement in the MPSR after this is artefactual because the mainlobe dominates the FOV pushing the significant sidelobes outside the FOV. The observed high MPSR values, therefore, cannot be associated with superior beamforming performance when the centroid offset is large relative to the dispersion. In every case there is a significant portion of randomly generated arrays that perform better than the logarithmic and regular arrays as seen by their marker positions relative to the standard deviation range of the irregular arrays.
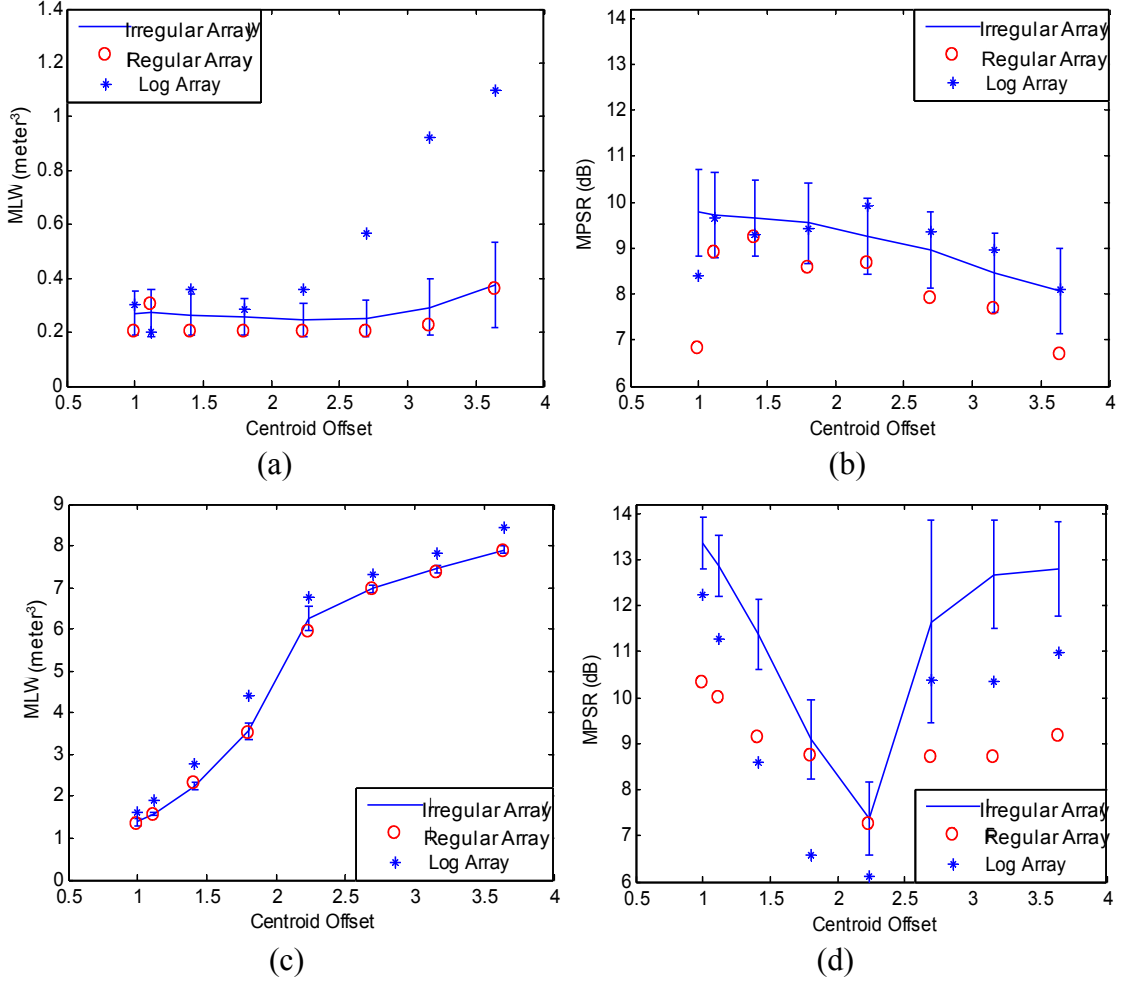
Figure 2.5: Centroid offset (in meters) and performance metrics with fixed dispersions, showing error bars at ±1 standard deviation. (a) MLW for a dispersion of 3.5. (b) MPSR for a dispersion of 3.5. (c) MLW for a dispersion of 0.5. (d) MPSR for a dispersion of 0.5.

Figure 2.6 presents the impact of array dispersion for a fixed centroid at the center of ceiling. It can be noted that small dispersions result in better MPSR for all geometries (closer average spacings between microphones); however, most of the irregular arrays perform better than either the regular or logarithmic arrays. With the centroid offset fixed, when array dispersion increases in the horizontal microphone plane, the MLW decreases along the horizontal direction; however, the MLW along vertical direction grows. This phenomenon is illustrated in Figure 2.7. When moving microphones away from the array centroid/target, the differential distances from each microphone to target point and the nearby locations reduce, resulting in higher coherent power for these points in Z

direction, thus extending the mainlobe. The sensitivity of these variations to dispersion is inversely related to the centroid offset. As the centroid offset becomes large relative to the dispersion, beamforming on a focal point is not practical (no longer an immersive environment). The array takes on more characteristics of a far-field array where the vertical direction MLW is so large that one only considers the angle or look direction instead of a focal point. In summary, for a fixed number of microphones there is tradeoff between MLW and MPSR that is dependent on the dispersion, as would be expected given the similarities between dispersion and aperture. In addition, by inspecting the standard deviation of error bars along each level of dispersion when array centroid is fixed, it can be seen that the variance of MLW increases with growing dispersion. A MPSR variance of $\pm 1 \sim 1.5$ dB is observed for each dispersion level with fixed centroid. Therefore, additional geometry descriptors based on the DPD distribution are expected to explain part of these variations of array performance.
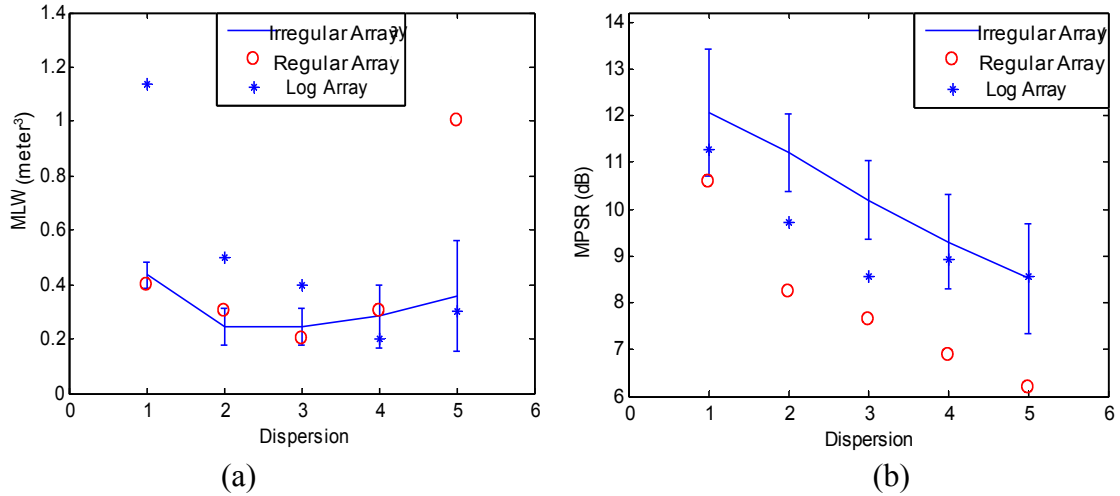
Figure 2.6: Array dispersion (in meters) and performance metrics with array centroid at the center of ceiling and centroid offset equal to 1, showing error bars at ± 1 standard deviation. (a) Dispersion vs. MLW. (b) Dispersion vs. MPSR.

$$\Delta d_1 = |d_1 - d_1'|$$
$$\Delta d_2 = |d_2 - d_2'|$$
$$\Delta d_1 > \Delta d_2$$
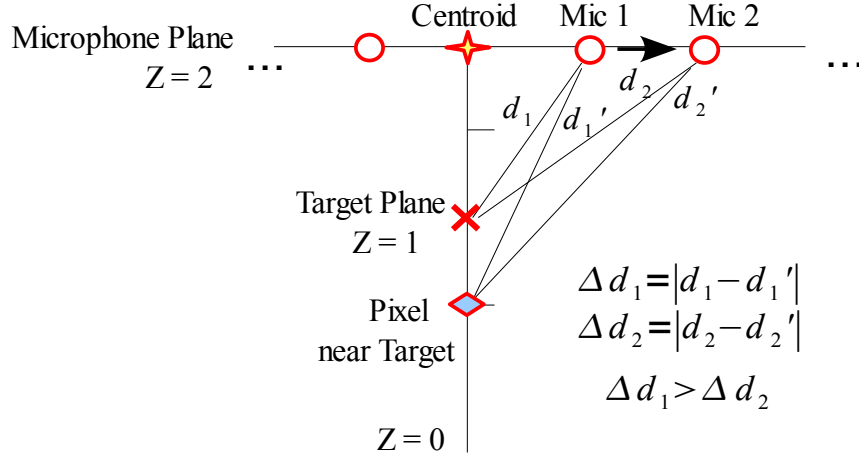
Figure 2.7: Differential distances of microphones in vertical direction of FOV

Results in Figures 2.5 and 2.6 demonstrate the impact of geometry descriptors related to aperture and array distance from focal points, which are largely consistent with the expectations. In all cases a portion of the randomly generated irregular arrays were superior to the regular arrays. In order to resolve between classes of irregular arrays, the following paragraphs analyze geometry descriptors based on DPD statistics with fixed centroids and dispersions, and demonstrate their ability to identify classes of irregular geometries with similar performance properties.

For a fixed centroid offset and dispersion, Figure 2.8 shows a relationship between array geometry DPD statistics and performance. Figures 2.8(a)(b) present the results for the arrays with similar centroid offset and dispersion values, while Figures 2.8(c)(d) present arrays with small centroid offset and large dispersion. And Figures 2.8(e)(f) present arrays applied in far field with large centroid offset over 10 times of the longest significant wavelength of 300 Hz. The results for the regular and logarithmic arrays are also plotted for reference. Figures 2.8(a)(b) demonstrate that larger DPD standard deviations and Pielou's evenness indices result in improved MPSR. These results are consistent with theoretical analysis indicating that wider and more evenly distributed DPDs create more incoherence in the phase terms of Eq. (17) and suppress noise better. Pielou's evenness index shows more sensitivity to the MPSR than the standard deviation, primarily because with a fixed dispersion, the standard deviation has limited range. Note

that the relative performance of logarithmic array in Figure 2.8(a) shows it with a very high standard deviation but not consistent with the trends of the irregular array, while for Pielou's index the MPSR of both the regular and logarithmic array are more consistent with irregular array performances.

When the array dispersion becomes much larger than centroid offset in Figure 2.8(c)(d), improvements of MPSR with increasing standard deviation or Pielou's index are not as dramatic. That is because arrays with large dispersion and small centroid offset typically generate a large DPD distribution spread (demonstrated by the increasing range of DPD standard deviation in Figure 2.8(c)) extending over many wavelengths in the useful frequency range. In these cases, Pielou's evenness index does not correlate as well with the beamforming gain as in Figure 2.8(b) because of the 2π modularity of the exponential argument. For a frequency of interest, the DPDs scaled by the wavelength are mapped to the $[-\pi, \pi]$ range by the modulo operation. The evenness index can be computed after this operation for frequency specific measures related to beamforming gains. In addition, results of Figures 2.8(c)(d) show that almost any irregular distribution will perform better than the regular geometry, and approximately 50% will perform better than the logarithmic array. Also, the relative performance of regular and logarithmic arrays is more consistent with the trends of the irregular array according to Pielou's evenness index than to standard deviation.

When the centroid offset becomes larger than 3 times of dispersion, the array takes on more characteristics of a far-field application. These cases do not fit with the primary focus of this analysis for immersive environments. As shown in Figure 2.8(e)(f), the DPD variations are limited and inappreciable over the FOV relative to the signal wavelength and large centroid offset (indicated by the observed dropping range of DPD standard deviations and Pielou's evenness indices). Variations in the microphone distributions will have little impact on performance, unlike for near-field applications. Centroid offset becomes the dominating factor affecting array beamforming performance, and the behavior of microphone array approaches the behavior of a single element in these far-field cases.

The results analyzed above demonstrate the impact of DPD distribution on array beamforming performance. Geometry descriptors based on the statistics of DPD

distribution show a correlation with array performance when the focal points and microphone distributions are typical for immersive or near-field applications. These DPD statistics explained the variations in performance when array centroid offset and dispersion were fixed. For a fixed number of microphones, increases in dispersion improved resolution, but also degraded noise suppression, while increases in centroid offset degraded both of these performance metrics. However, as shown in Figure 2.8, with fixed centroid and dispersion, ± 0.5~1dB variances of performance metrics are observed for each bin of DPD statistics. Although these variations of performance metrics partly result from the quantization errors of DPD statistics, other geometry parameters may exist that can further reduce these variations.

In addition, it is noted that during our simulations, the main processing errors result from three sources, quantization errors of DPD statistics, missing data in some levels of geometry descriptors, and pseudo-randomness of source signal generator based on SII. In order to reduce these errors, quantization steps should be set smaller than 10% of the mean values of corresponding geometry descriptors. And all data levels with missing cells are removed before applying any statistical analysis.
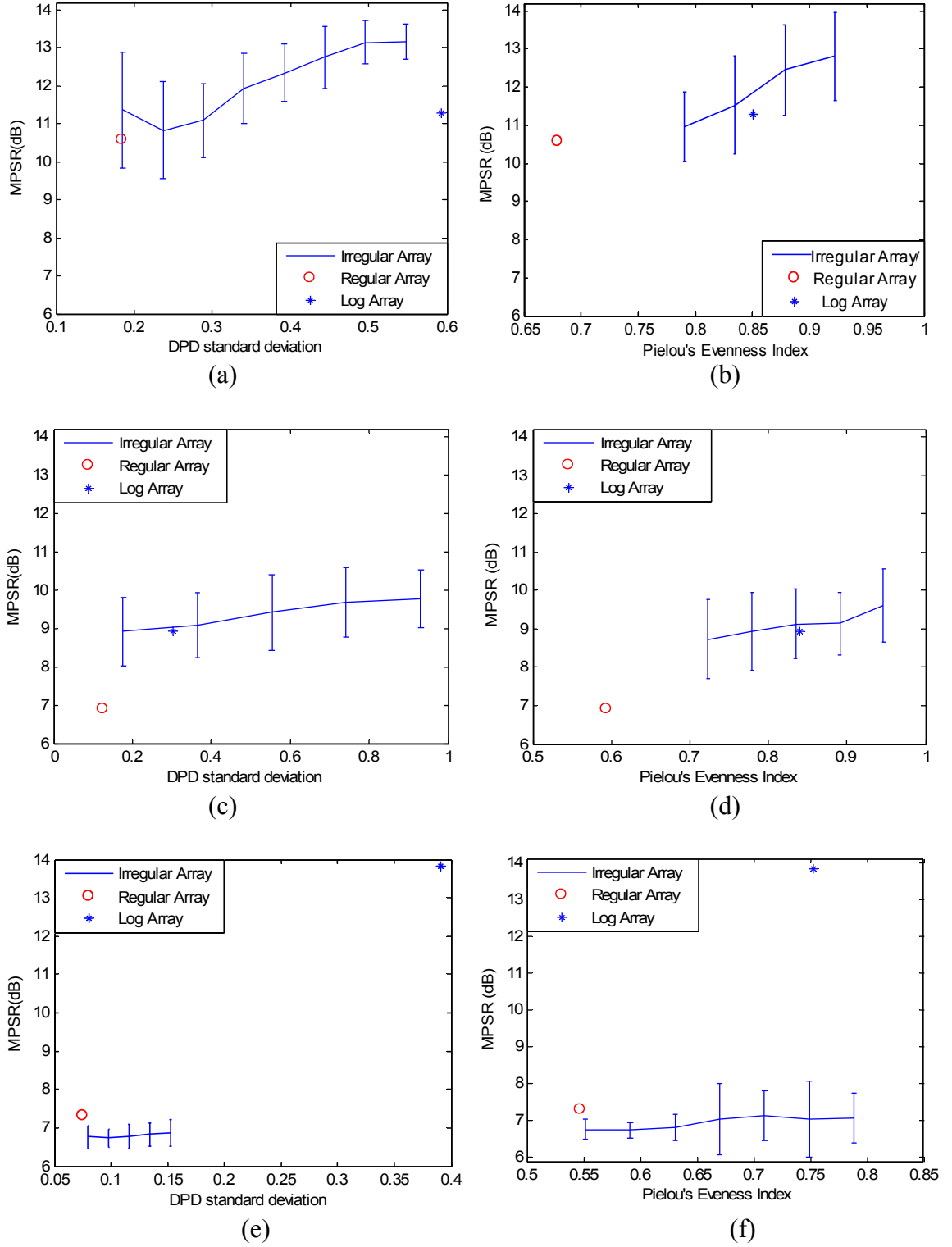
Figure 2.8: DPD statistics and performance metrics for fixed dispersions and centroids, showing error bars at ±1 standard deviation. Each bin of DPD statistics shown in the

35

figure contains over 10% samples (over 30 arrays) of the total experiments. (a) DPD standard deviation vs. MPSR for a centroid of 1 and dispersion of 1. (b) Pielou's evenness index vs. MPSR for a centroid of 1 and dispersion of 1. (c) DPD standard deviation vs. MPSR for a centroid of 1 and dispersion of 4. (d) Pielou's evenness index vs. MPSR for a centroid of 1 and dispersion of 4. (e) DPD standard deviation vs. MPSR for for a centroid of 13 and a dispersion of 3.5. (f) Pielou's evenness index vs. MPSR for a centroid of 13 and a dispersion of 3.5.

2.3.2   Analysis of Variance

To further investigate the significance of the proposed geometry descriptors' impact on performance, ANOVA is applied, which is useful for investigating the effect of independent factors on observations [43]. The performance metric variation is partitioned into portions attributed to the effect of independent factor (between-group variation) and portions attributed to random error (within-group variation). An F statistic is computed using the ratio between these variances and tested for significance. Tables 2.3 and 2.4 show the 3-way ANOVA results for MLW and MPSR values, respectively. Centroid offset, dispersion, DPD statistics, and their interactions are considered as the independent factors impacting the performance metrics. By examining the results, it can be seen that the p values for these geometry descriptors and their interactions are all highly significant (all less than 0.01) for their impact on MLW and MPSR. In addition, high R-Square values indicate that 99.7% of the variation in MLW data can be accounted for by these independent factors, so does 82% data of MPSR. Therefore, it is demonstrated that proposed geometry descriptors, including centroid offset, dispersion and DPD statistics, have strong correlations with array performance.

Finally, through statistical analysis and ANOVA the relationships between proposed

geometry descriptors and array performance are established and demonstrated. However, because the number of microphone determines the number of DPDs, the impact of each geometry descriptor varies with the number of microphone. In order to analyze these differences, data collected from Monte Carlo experiments of irregular arrays with 16, 25, 36, 49 and 64 microphones are compared. All the experiments were performed in immersive environments with comparable values of centroid offset and dispersion. Table 2.5 provides the R-Square results of least squares method by fitting General Linear Model (GLM) of selected geometry descriptors on MPSR. It is noted that even with this simplest regression model, over 50% variation of MPSR can be accounted for by GLM$\{L, a, \sigma, J\}$. This percentage increases to 70% ~ 90% when applying higher order fitting functions of geometry descriptors (nonlinear regression models). With increasing microphone number, better R-Square values are obtained.

By comparing the results of GLM$\{L, a\}$ derived from array apertures and positions with GLM$\{L, a, \sigma, J\}$ taking account of DPD distributions' impact, at least 10% improvements of R-Square values are observed. Especially for the arrays with microphone density larger than 0.5 mic/m^2, the impact of $\{L, a\}$ are reduced greatly due to the increasing possibilities of microphone arrangements with fixed centroid and dispersion, while the DPD statistics show stronger correlation with array performance. Furthermore, by comparing the trends of R-Square values of GLM$\{L, a, \sigma\}$ and GLM$\{L, a, J\}$ with increasing microphone number, DPD standard deviation assessing the spread of DPD distribution shows a little stronger correlation with MPSR for arrays with microphone density less than 0.2 mic/m^2, while Pielou's evenness index assessing the diversity of DPD distribution has greater impact on MPSR for array with density larger than 0.2 mic/m^2. The reason for this phenomenon is that low microphone density cannot

provide enough DPD samples to measure the entropy (Pielou's evenness index), and DPD

standard deviation representing the average spread of DPDs about zero is more reflective

for characteristics of the DPD distribution related to the beamforming gain.

Table 2.3: 3-way ANOVA results of MLW

	Factors	F Value	p value (Pr>F)
Main	Centroid Offset	1955.98	<0.0001
Effects	Dispersion	3327.62	<0.0001
	Pielou's Evenness Index	2.20	0.0031
Interactions	Centroid*Dispersion	2725.46	<0.0001
	Centroid*Pielou	1.63	0.0004
	Dispersion*Pielou	2.20	<0.0001
R-Square		0.9967	

Table 2.4: 3-way ANOVA results of MPSR

	Factors	F Value	p value (Pr>F)
Main	Centroid Offset	25.60	<0.001
Effects	Dispersion	1.99	<0.0366
	Pielou's Evenness Index	26.21	<0.0001
Interactions	Centroid*Dispersion	52.55	<0.0001
	Centroid*Pielou	3.83	<0.0001
	Dispersion*Pielou	5.50	<0.0001
R-Square		0.8169	

Table 2.5: R-Square results for general linear models of geometry descriptors on MPSR

Number of Mics	Mic Density (mic/m^2)	R-Square of GLM$\{L, a\}$	R-Square of GLM$\{L, a, \sigma\}$	R-Square of GLM$\{L, a, J\}$	R-Square of GLM$\{L, a, \sigma, J\}$
16	0.16	41.93%	46.72 %	46.56 %	50.33%
25	0.25	53.84%	58.25 %	58.81 %	62.72%
36	0.36	60.81%	64.80 %	66.49 %	69.32%
49	0.49	68.77%	73.20 %	74.04 %	77.31%
64	0.64	63.61%	80.09 %	81.69 %	86.97%

2.4   General Guidelines for the Design of Optimal Array

Previous analyses identified the important characteristics for irregular microphone arrays that directly related to beamforming performance for human speech applications. Combined with descriptors analogous to traditional geometry parameters for regular arrays (i.e. array centroid and dispersion), novel geometry descriptors involving DPD statistics described both regular and irregular arrays. Simulations demonstrated that irregular microphone geometries typically exceed the performance of regular geometries, and arrays with high DPD entropy and wide DPD spread correspond to arrays with better noise suppression ability. These results are primarily applicable for microphone arrays in near-field applications, such as in immersive environments.

The relationships between geometry descriptors and beamforming performance developed in previous sections can be applied to predict the array SNR performance in given acoustic environments, and further act as the objective functions (as in Eq. (8)) in the optimization procedure to search for the optimal microphone distributions. Note that the results of this dissertation were based on Monte Carlo experiments with planar microphone arrays, which are more applicable for indoor applications, such as audio surveillance systems. So far, the DPD statistics do not have simple geometric interpretations and must be computed based on all the microphone positions and desired focal points. In order to directly generate a microphone array with good values of proposed geometry descriptors or guide *ad hoc* microphone placements, this section

discusses the methods and guidelines for the design of array pattern with desired geometry descriptors $\{L, a, \sigma, J\}$, providing good performance in specified scenarios.

2.4.1   Methods to Build Arrays with Desired Geometry Features

From previous analyses, *good* values of geometry descriptors $\{L, a, \sigma, J\}$ result in enhanced SNR performance in corresponding environment, where the term *good* is assessed based on previous conclusions about relationship between geometry descriptors and performance metrics. With the knowledge of acoustic scene, such as possible source distribution or desired FOV, methods to directly design a microphone distribution providing desired $\{L, a, \sigma, J\}$ levels are needed to guide *ad hoc* (not computer aided) optimal microphone array design in immersive environment.

As applied in the Monte Carlo simulations, different values of array centroid offset and dispersion can be generated by shifting and scaling microphone coordinates according to the origin. Since array dispersion is related to centroid, the transfer function from array with centroid r_0 and dispersion a_0 to array with centroid r_0' and dispersion a' can be given as:

$$[r_1', .., r_p', .., r_P'] = \frac{[r_1, .., r_p, .., r_P] - r_0}{\dfrac{a_0}{a'}} + r_0' \quad , \tag{31}$$

where r_p and r_p' are the original and transferred positions of pth microphone.

DPD statistics have been demonstrated as the novel geometry features showing strong correlation with the noise suppression ability of microphone array. Unlike traditional descriptors, $\{\sigma, J\}$ are the statistics for DPD distribution according to specified target and noise positions, which do not have simple geometric interpretations and were computed based on pre-generated microphone positions in the Monte Carlo experiments. Usually, for specified source distributions arrays with large $\{\sigma, J\}$ are discovered by random or heuristic searching methods, which can be applied in the computer aided optimization cases as discussed in Section 5.2. For *ad hoc* (not computer aided) optimal microphone array design, although the interrelationship between array dispersion and DPD standard deviation (discussed in Section 2.2.3) can control σ in some degree by

changing array dispersion, direct geometric design methods related to good values of entropy J are still lacking. Therefore, a novel cluster design method based on hyperbola curve is proposed to directly generate geometries with large values of DPD statistics.

The definition of DPD in Eq. (18) can be rewritten as:

$$\Delta_{pq}(\boldsymbol{r}_i, \boldsymbol{r}_s) = (d_{sq} - d_{iq}) - (d_{sp} - d_{ip}) \quad , \tag{32}$$

where the DPD is explained as the difference of the differential distances from each mic to two spatial positions $\{\boldsymbol{r}_i, \boldsymbol{r}_s\}$ in FOV. Because a hyperbola curve can be defined equivalently as the locus of points where the absolute value of the difference of the distances to the two foci is a constant (equal to the distance between its two vertices), it can be applied in here to distinguish microphones with different values of $(d_{sq} - d_{iq})$. As shown in Figure 2.9(a), $\{\boldsymbol{r}_i, \boldsymbol{r}_s\}$ are considered as the focuses. Microphones located on the same hyperbola curve (marked in the same color) have identical values of $(d_{sq} - d_{iq})$, while microphones located inside the hyperbola curve show larger absolute values of differential distance. Take the light red curve for example, microphones on the curve have $d_{sq} - d_{iq} = 1.5$, while microphones located in the dark red area have $d_{sq} - d_{iq} > 1.5$. Therefore, with specified $\{\boldsymbol{r}_i, \boldsymbol{r}_s\}$, in order to obtain good statistics of DPD distribution over all microphone pairs, some microphones should be distributed inside the dark red hyperbola area and some inside dark blue hyperbola area to generate a set of DPDs with possible largest spread. The DPD values between this range are obtained by the nearby microphone pairs located in the same area, providing a smooth entropy. Note that there is no need to put microphones in the middle area of $\{\boldsymbol{r}_i, \boldsymbol{r}_s\}$. Figure 2.9(b) gives a irregular array clustered according to the hyperbola theory. One target and three interferences are considered for this scene. The hyperbola areas for each target-noise pair are marked in dashed lines with different color. Microphones are divided into four clusters uniformly distributed in these areas. Simulation results demonstrated that this hyperbola clustered array has comparable or even better SNR performance than arrays picked by heuristic searching methods. And large SNR improvement is observed when comparing with similar regular arrays. More information of the hyperbola cluster array design will be discussed in Section 5.3.1.
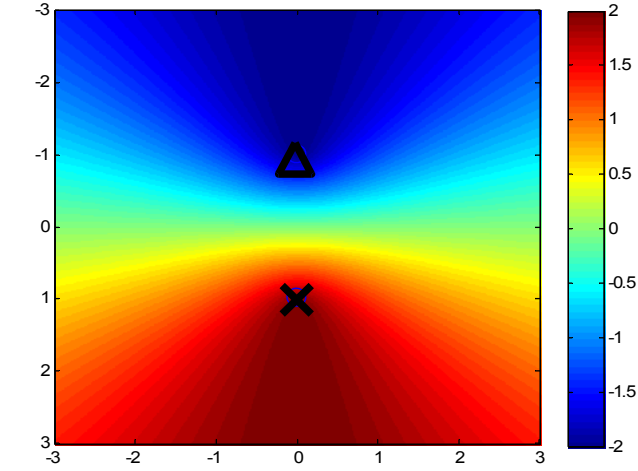
(a)
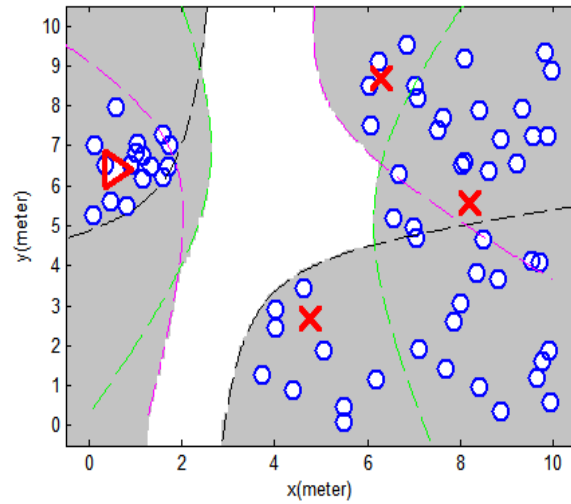
(b)

Figure 2.9: Top view diagrams of hyperbola theory. Triangles represent target. X's represent noise sources. Microphones are marked as O's. (a) Hyperbola areas. (b) Hyperbola clustered array.

2.4.2 Guidelines for Good Geometries in Scenarios

To summarize previous analyses about important geometry features related to superior performance, general guidelines to create good microphone distributions for common scenarios are presented in this section, which can be directly implemented or visualized by humans in the interested fields. Note that because all our conclusions are derived from

Monte Carlo experiments with ceiling arrays, there are several limitations for the application of these guidelines:

(1) Indoor environments, such as audio surveillance systems.

(2) Near-field applications, where array distances from the Sources of Interest (SOI) have comparable values (less than three times) with array apertures. (Centroid offset should be smaller than three times of array dispersion.)

(3) Human speech applications, such as cocktail party.

Three acoustic scenarios are created to represent common cases in real application, where an approximate rectangular/cubic room is assumed as the FOV. Guidelines for the optimal microphone distributions in immersive environment are provided as below.

(1) When SOI positions/regions and major interferences are known

- High microphone density should be clustered near the target and interference positions. If the interferences are located in certain regions, some microphones should be placed to surround these regions, instead of distributing over the regions. Relative examples of the optimal irregular geometries are given in Figure 2.10.

- If higher resolution is required, the spread of microphones from targets should be increased, and the centroid of microphone array should be close to the target region. However, the noise suppression ability will be degraded with the increase of dispersion.

- To be specified, microphones should be placed in the hyperbola areas of each pair of target and interference positions to generate a DPD distribution with rich entropy, and further improve array noise suppression ability. Relative optimal geometry is given in Figure 2.9(b), where microphones (marked by O's) are clustered inside the hyperbola areas as many as possible, according to each pair of target and interference in FOV (marked by triangles and X's).
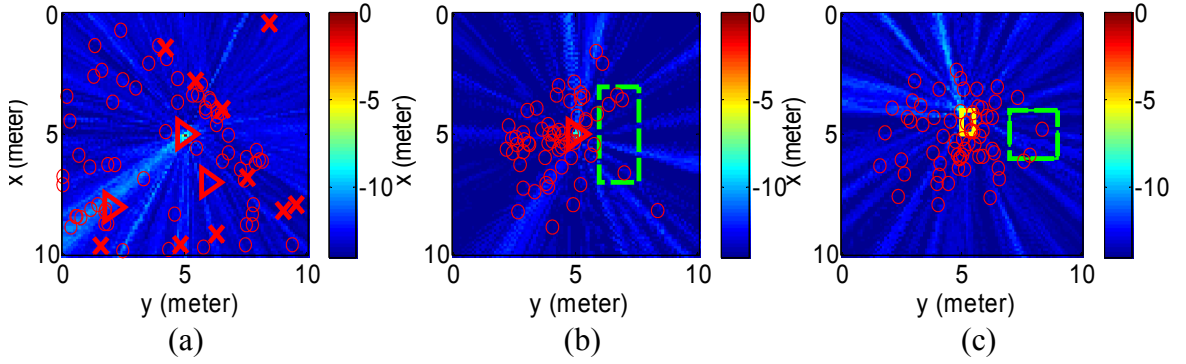
Figure 2.10: Top view gain patterns on target plane of optimal irregular arrays. Red circles represent microphone positions. Red cross or green square is the possible noise space. Red triangle or yellow square is the desired target space. (a) and (b) focus on the center target in the room. (c) focuses on the top left corner of target space.

(2) When SOI positions/regions are known
- High microphone density should be clustered near the target positions/regions.
- If high resolution for each position in the target region is required, the spread of microphones should be increased, and the array centroid should be close to the target region. However, the noise suppression ability will be degraded with the increase of dispersion. If the target region and microphone design space are parallel square planes, such as ceiling array in cocktail party, rough functions derived from statistical analyses can be given to compute the minimum dispersion to obtain desired 3dB resolution as:

$$a_0 = \frac{\Gamma_{3dB} - 0.7d - 0.047 L_0 - 0.784}{-0.077} \quad , \tag{33}$$

or

$$a_0 = \frac{\Gamma_{3dB} + 0.17 d^2 - 1.12 d - 0.045 L_0^3 + 0.02 L_0 - 0.565}{-0.04} \quad , \tag{34}$$

where a_0 represents required minimum array dispersion. Γ_{3dB} represents desired 3dB resolution (minimum distance with more than 3dB change of array

gains). *d* represents dimension of target plane. L_0 represents distance between target and microphone planes.

(3) When neither are known

- Microphones should be spread all over the FOV with an uniform distribution to ensure a full coverage.

- If the FOV has an irregular shape with protruding walls, the half-blind areas should have higher microphone density than uniform distribution, while the blind areas should have lower microphone density than uniform distribution. As shown in Figure 2.11, the red protruding walls block the direct propagation paths of sound wave from/to the dark blue area. The optimal irregular geometries generally have uniform distribution of microphones over the entire FOV, while with higher mic density in the triangle half-blind areas on the right and lower mic density in the triangle blind areas on the left.
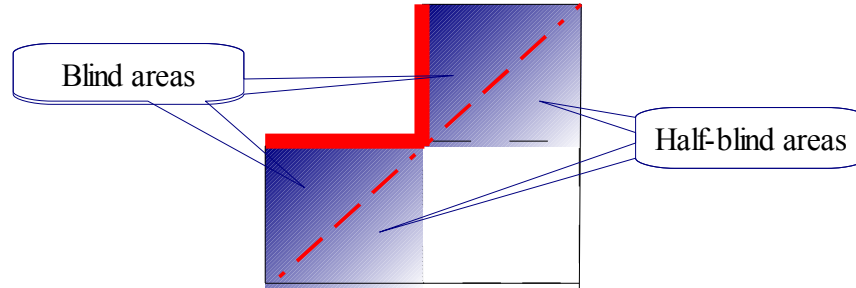
Figure 2.11: Top view diagram of irregular FOV

Therefore, by following these guidelines, good microphone placements which have high possibility to show enhanced beamforming performance in specified acoustic environments can be directly visualized or set by humans, such as indoor audio surveillance or mobile platform applications where rapid deployment is required with limited knowledge of the acoustic scene. However, these guidelines only provide general instructions for the microphone placements applied in simple scenes. If more information of environment is known, the computer aided heuristic searching methods can be more

effective to sort out the superior irregular geometries, which require closed-form objective functions derived from the relationship between important geometry features and performance. Relative discussion will be provided in the next section.

2.5   Relationship Function of Geometry Descriptors with Performance

This section presents the closed-from relationship functions between important geometry descriptors and key performance metrics from statistical analyses. They can be effectively applied to predict the SNR performance based on the geometry features, and provide a simple and feasible solution for the optimization of microphone arrays.

2.5.1   Data Collection and Fitting Steps

As mentioned in Section 2.3, the data sets applied in the fitting procedures are derived from the Monte Carlo experiments by uniformly distributed microphones over a ceiling planar design space in a 10x10x2 m^3 room. Since the main applications considered for our study involve speech (as in the case of surveillance in a cocktail party environment), the excitation of the arrays is colored noise with the same spectral distribution as the band importance function used SII. Arrays with 16, 25, 36, 49 and 64 microphones are examined. As immersive environment or near-field study, array centroid values range from the center of ceiling to the edge at 1 meter intervals along x axis, while 5 levels of dispersion are applied with each centroid. For each combination of centroid offset and dispersion level, 300 independent arrays are generated by Monte Carlo experiments, while the DPD statistics are computed directly from the microphone positions. The 3D spatial gains computed over the FOV are used to directly estimate the MLW and MPSR. And their relationships with proposed geometry descriptors are explored through nonlinear regression procedures, respectively. The details are shown as below.

 Step 1: Apply multi-way ANOVA to investigate the significance of the impact of proposed geometry descriptors and their interactions on each performance metric, as well as the number of microphone. Pick the significant ones as the independent factors applied in corresponding fitting procedure.

Step 2: Examine the mean plot of each independent factor vs. the performance metric to obtain the qualitative conclusion about the relationship function, such as the general shape of curve between each independent factor and performance metric, or interactions among independent factors affecting the shape of curve.

Step 3: Apply nonlinear regression to search for the optimal transformation from each independent variable to dependent variable (performance matrix), based on the highest R-Square value of the regression model and the results of residual analysis (such as moments, skewness index, and normality plots of residuals to validate the assumptions of regression analysis that residuals are normal with constant variance over the values of dependent variable).

Step 4: Ordered by the significance of independent variables, combine the nonlinear functions of each independent variable by additive and multiplicative operators, while the weight of each nonlinear term is expressed as the function of related interactions discovered in step 2. Use the resulting coefficients of each nonlinear transformation in step 3 as the initial conditions of Gauss nonlinear regression procedure to obtain the final model by iteration.

Step 5: Determine the best model and combination stepwise, according to R-Square, p-value, moments, and normality of integrated model residuals.

2.5.2 Closed-form Relationship Functions

The fitting results of closed-form relationship function for each performance metric based on geometry features are presented in this section. For MPSR, the mean plots in step 2 indicate the following characteristics of $\Gamma_i(L, a, \sigma, J, P)$:

- $\Gamma_i()$ is a monotone increasing function of the number of microphone P.

- $\Gamma_i()$ is a decreasing function of centroid offset L. The slope is related to array dispersion a, and the parallel displacement is related to the number of microphone P. (Smaller a results in sharp slope, while more microphones bring higher Γ_i .)

- $\Gamma_i()$ is a decreasing function of array dispersion a. The parallel displacement of curve is related to the number of microphone P, where more microphones bring higher Γ_i . The slope of curve is related to DPD statistics, where Pielou's evenness

index J shows greater impact than DPD standard deviation σ . When DPD statistics decrease, the value of slop decreases from near zero to negative levels.

- $\Gamma_i()$ is a increasing function of DPD standard deviation σ . The parallel displacement of curve is related to the number of microphone P, where more microphones bring higher Γ_i .

- $\Gamma_i()$ is a increasing function of Pielou's evenness index J. When the number of microphone increases, slop increases from near zero to positive levels. When array dispersion increases, slop decreases from positive to negative levels.

For 3D MLW, the characteristics of $B_{3dB}(L, a, P)$ are developed in the same manner, as provided below:

- $B_{3dB}()$ is a decreasing function of the number of microphone P.

- $B_{3dB}()$ is a increasing function of centroid offset L, where the slope is related to array dispersion a and the parallel displacement is related to the number of microphone P. (Smaller a results in sharp slope, while more microphones bring smaller B_{3dB} .)

- $B_{3dB}()$ is a decreasing function of array dispersion a, where the slope of curve is related to P and L. (Smaller P and large L result in sharp slope.)

The closed-form functions derived from corresponding subset data are provided in Table 2.6 and 2.7, which will be applied to form the objective functions of microphone array optimization in different application cases.

Table 2.6: Results of nonlinear regression procedure for MPSR

Formulation	R-Square	Mean of Residuals	Validated Applicable Areas
Model 1 : $\Gamma_i = \text{Exp1}[L] + \text{Cubic}[a] + \text{Cubic}[J]$ $+ \text{Cubic}[\sigma]$	78.35%	2.24E-8	Planar arrays with mic density of $0.04 \sim 4$ mic/m^2. Linear arrays with mic density of $0.4 \sim 3.6$ mic/m.
Model 2: $\Gamma_i = \text{Cubic}[P] + \text{Cubic}[P]\text{Exp1}[L]$ $+ \text{Cubic}[P]\text{Cubic}[a]$ $+ \text{Cubic}[P]\text{Cubic}[\sigma]$ $+ \text{Cubic}[P]\text{Cubic}[J]$	84.22%	-1.73E-10	The best model for planar array with mic density of $0.16 \sim 7.84$ mic/m^2. Linear arrays with mic density of $1.6 \sim 3$ mic/m. Spherical ceiling arrays with mic density of $0.16 \sim 7.84$ mic/m^2.
Model 3: $\Gamma_i = \text{Cubic}[P] + \text{Cubic}[a]\text{Cubic}[L]$ $+ \text{Cubic}[J]\text{Exp2}[a] + \text{Exp2}[\sigma]$ $+ \text{Cubic}[P]\text{Exp1}[a]\text{Cubic}[J]$	78.54%	-1.22E-3	Planar arrays with mic density from 0.25 to 0.4 mic/m^2 .

Note: The optimal nonlinear transformations in step 3 are listed as below: Cubic[x]: a+bx+cx^2+dx^3; Exp1[x]: (a+bx+cx^2);   Exp2[x]: exp(a+b/x).   The coefficients {a,b,c,d} in each nonlinear base are obtained during iteration.

Table 2.7: Results of nonlinear regression procedure for MLW

Formulation	R-Square	Mean of Residuals	Validated Applicable Areas
Model 1: $B_{3dB} = \text{Cubic}[L] + \text{Exp1}[a]$	99.13%	-4.87E-7	Derived from arrays with 64 microphones.
Model 2: $B_{3dB} = \text{Cubic}[a]\text{Cubic}[P]$ $+ \text{Cubic}[a]\text{Exp1}[L]$ $+ \text{Cubic}[P]\text{Cubic}[L]\text{Cubic}[a]$	69.86%	2.67E-8	Applied for arrays with mic density smaller than 0.5 mic/m²

Note: The optimal nonlinear transformations in step 3 are listed as below: Cubic[x]: a+bx+cx^2+dx^3; Exp1[x]: (a+bx+cx^2). The coefficients {a,b,c,d} in each nonlinear base are obtained during iteration.

Table 2.6 and 2.7 show the fitting results of several nonlinear models. From the R-Square values it can be seen that about 70% ~ 90% of the variation of performance metrics can be explained by these models. Note that moments, skewness index, and normality plots of residues were checked for every regression model to validate the assumptions of regression analysis (residuals are normal with constant variance over the values of predicted Y). These conditions were met for all the resulting models provided in Tables. Therefore, the proposed geometry features provide a reasonable correlation with array performance, and the key metrics are successfully expressed by the functions of proposed descriptors. Combined with the prior knowledge of acoustics scene (possible distributions of interested target and major interferences), these geometry-based performance functions can be used as the objective functions in the computer aided optimization procedure to search for the optimal irregular arrays with enhanced SNR performance. The details of relative optimization strategies involving heuristics searching will be proposed in Chapter 5.

Chapter 3   Principles of Microphone Array Optimization

3.1   Problem Formulation

The goal of microphone array optimization in our study is to find microphone geometries maximizing beamformer SNR for given distribution of target and specified acoustic scenes. As analyzed in Section 1.2, the important factors affecting array SNR performance include microphone positions, target and noise source regions, significant signal frequency bands, propagation transfer functions of acoustic environment and the beamforming algorithm. Let Ω represents the 3D acoustics space. $\boldsymbol{R}_T \subset \Omega$ and $\boldsymbol{R}_N \subset \Omega$ denotes the target and noise space containing all possible target positions $\boldsymbol{r}_i$ and noise source positions $\boldsymbol{r}_k$. Note that $\boldsymbol{R}_T$ and $\boldsymbol{R}_N$ can be continuous or discrete space derived from the preknowledge of acoustic scene. Arrays with P microphones are distributed in the design space $\boldsymbol{R}_M$, where a specified geometry is represented by $\boldsymbol{G} = \{\boldsymbol{r}_{1,}.., \boldsymbol{r}_p, .., \boldsymbol{r}_P\}$.
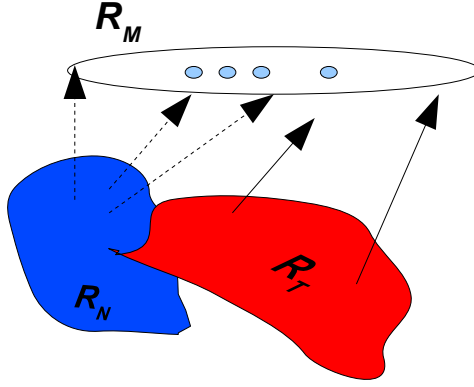
Figure 3.1: Propagation environment (adapted from [44]). $\boldsymbol{R}_M$ is the microphone design space, including all the possible microphone positions. $\boldsymbol{R}_T$ represents the interested target space. $\boldsymbol{R}_N$ represents the noise space. The solid arrows represent target signal propagation paths. The dashed arrows represent the noise signal propagation paths.

Therefore, the signals received by the array with geometry $\boldsymbol{G}$ can be expressed in frequency domain as [44] :

$$\hat{V}(\omega;\boldsymbol{G}) = \int_{\boldsymbol{r}_i \in \boldsymbol{R}_T} \hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})\, d\,\hat{U}(\omega;\boldsymbol{r}_i) + \int_{\boldsymbol{r}_k \in \boldsymbol{R}_N} \hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_k,\boldsymbol{G})\, d\,\hat{N}(\omega;\boldsymbol{r}_k) \quad , \quad (35)$$

where $\hat{V}(\omega;\boldsymbol{G}) = [\hat{V}(\omega;\boldsymbol{r}_1),..,\hat{V}(\omega;\boldsymbol{r}_p),..,\hat{V}(\omega;\boldsymbol{r}_P)]^T$ is the vector denoting signal received by each element of array.

$\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_s,\boldsymbol{G}) = [\hat{H}(\omega;\boldsymbol{r}_s,\boldsymbol{r}_1),..,\hat{H}(\omega;\boldsymbol{r}_s,\boldsymbol{r}_p),..,\hat{H}(\omega;\boldsymbol{r}_s,\boldsymbol{r}_P)]^T, \boldsymbol{r}_s \in \Omega$ is the transfer function corresponding two spatial points in Ω , which is supposed to be known by the propagation laws applied in specified environment. $d\,\hat{U}(\omega;\boldsymbol{r}_i), d\,\hat{N}(\omega;\boldsymbol{r}_k)$ are the target and noise signal components in frequency domain, respectively. Note that in Eq. (35) the hat notation represents corresponding frequency domain component and the bold notation represents vector. If single target and uncorrelated noise sources are considered. Eq.(35) can be rewritten to the common used model as:

$$\hat{V}(\omega;\boldsymbol{r}_i,\boldsymbol{G}) = \hat{U}(\omega;\boldsymbol{r}_i)\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G}) + \hat{N}(\omega;\boldsymbol{G}) \quad , \qquad (36)$$

Then the output of beamforming filter $\hat{W}(\omega)$ can be expressed as:

$$\hat{Y}(\omega;\boldsymbol{r}_i,\boldsymbol{G})=\hat{\boldsymbol{W}}^H(\omega)\hat{V}(\omega;\boldsymbol{r}_i,\boldsymbol{G})$$
$$\hat{Y}(\omega;\boldsymbol{r}_i,\boldsymbol{G})=\hat{\boldsymbol{W}}^H(\omega)\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})\hat{U}(\omega;\boldsymbol{r}_i)+\hat{\boldsymbol{W}}^H(\omega)\hat{N}(\omega;\boldsymbol{G}) \quad, \qquad (37)$$

Usually, the beampattern is defined as the gain between the output and the original target signal as [44]:

$$\hat{\Psi}(\omega;\boldsymbol{r}_i,\boldsymbol{G})=\hat{\boldsymbol{W}}^H(\omega)\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G}); \quad \boldsymbol{r}_i\in R_T \quad, \qquad (38)$$

And the output of beamformer $\hat{Y}(\omega;\boldsymbol{r}_i,\boldsymbol{G})$ can be considered as the linear estimator of the original target signal $\hat{U}(\omega;\boldsymbol{r}_i)$, where the the estimated target signal are expressed as [44]:

$$\hat{U}_E(\omega;\boldsymbol{r}_i,\boldsymbol{G})=\frac{\hat{\boldsymbol{W}}^H(\omega)\hat{V}(\omega;\boldsymbol{r}_i,\boldsymbol{G})}{\hat{\boldsymbol{W}}^H(\omega)\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})} \quad, \qquad (39)$$

In our study, in order to focus on the impact of microphone geometry, the beamforming algorithm are fixed to DSB. And all the source signals applied in the experiments are derived from normalized colored noise of SII model to simulate the important frequency bands of human speech. Combined with the analysis of Eq (7), it can be seen that the array gain pattern is the dominate factor determining the array ability to recover target signal, and has strong correlation with SNR performance in specified acoustic scene. Therefore, the relationship function derived from the statistical analyses of the array gain pattern measures according to array geometry features can be applied as the objective function in the microphone optimization problems, to represent superior target extraction ability of arrays. In the next section, based on these system models, several criteria applied in microphone array optimization are proposed to assess array performance in the search of optimal microphone geometries.

3.2   Optimization Criteria and Constrains

Because different application environment has different requirements for array performance, selecting suitable optimization criterion is the first step to search for the optimal array geometry. Several common used criteria, which can reflect different

capabilities of microphone array, are introduced in this section.

C1: Array pattern distances [25, 44]

C2: Function of performance metrics [28]

C3: Minimum noise power of beamforming estimator [2, 44]

C4: Information capacity between source positions and array [44]

3.2.1 *C1:* Array Pattern Distances

As discussed in Section 2.1, the optimal microphone arrangements should provide strong coherent addition in the target regions and near zero power gain over the noise regions to enhance target signals and filter out the noises. The objective in selecting a microphone distribution, therefore, can be stated to minimize the difference between desired ideal pattern and actual array pattern at the interested points in the FOV.

Let $\hat{\Psi}_0(\omega;r_s), r_s \in \Omega$ denote the desired beampattern, which normally is defined as a set of impulses in the target positions $\hat{\Psi}_0(\omega;r_s) = \delta(r_s - r_i); r_i \in R_T$. The residual distance between this ideal pattern and actual array pattern given in Eq.(38) is expressed as [25, 44]:

$$d_q(\omega;\boldsymbol{G}) = \left[\int_{r_s \in (R_T \cup R_N)} \|\hat{\Psi}(\omega;r_s,\boldsymbol{G}) - \hat{\Psi}_0(\omega;r_s)\|^q d\eta(\omega;r_s) \right]^{1/q} , \qquad (40)$$

where $\eta(\omega;r_s)$ is the weight to combine the pattern difference at each spatial point and frequency band. It can be derived from the Power Spectrum Density (PSD) measures over source areas to assign different importance to specified spatial region or signal frequency band [44]. Or it can denote the probability density function representing the likelihood of positions for the desired target and possible noise sources locations, which is related to the behavior pattern of interested sources. In the case when no prior knowledge is available, $\eta(\omega,r_s)$ is normally set to uniform values, and contains larger values for some region expecting low errors. Therefore, the criterion to select optimal array geometry can be written as:

$$\boldsymbol{G}_{opt} = \underset{\boldsymbol{G} \subset \mathbf{R_M}}{argmin} \langle d_q(\omega;\boldsymbol{G}) \rangle , \qquad (41)$$

54

Or it can be stated as a minmax problem over spatial positions in FOV, as show below.

$$G_{opt} = \underset{G \subset \mathbf{R_M}}{argmin} \langle \underset{r_s \in (R_T \cup R_N)}{max} [d_q'(\omega; \boldsymbol{r_s}, \boldsymbol{G})] \rangle \quad , \tag{42}$$

where $d_q'(\omega; \boldsymbol{r_s}, \boldsymbol{G}) = \eta(\omega; \boldsymbol{r_s}) \cdot \|\hat{\Psi}(\omega; \boldsymbol{r_s}, \boldsymbol{G}) - \hat{\Psi}_0(\omega; \boldsymbol{r_s})\|_q$ measuring the residues between desired and actual patterns to reflect how well the actual pattern approximates to the desired pattern at interested spatial sample points.

3.2.2   *C2:* **Function of Performance Metrics**

The optimization in this dissertation seeks to find a microphone distribution within predefined design space that maximizes the beamformer SNR over a given distribution of target and noise sources. As defined in Section 1.2.3, two performance metrics that related directly to key aspects of the array gain pattern are applied in our study, MLW associated with resolution and MPSR associate with noise suppression ability. In order to avoid the expensive computations associated with a direct array gain computation over the space of interest, important geometric descriptors showing strong correlation with array performance are proposed in Chapter 2. The nonlinear relationships of proposed descriptors on each performance metric for human speech applications in immersive environments are developed and demonstrated in Section 2.5. These relationship functions can be combined together as the objective function for array optimization, which directly considers the features of gain pattern related to the SNR performance and avoids the redundant definition of desired pattern in traditional optimization method.

From the conclusions of previous research, there is usually a trade-off between these two performance metrics. For example, increase the dispersion of microphones can result in better resolution (small MLW), but bring higher sidelobes (lower MPSR) degrading noise suppression ability. In order to consider these two criteria simultaneously, a limit on MLW is set by enforcing a penalty on the function of MPSR through a multiplier λ to form the objective function of optimization. For a given focal point $\boldsymbol{r_i}$ and noise source at $\boldsymbol{r_k}$, the objective function can be written as:

$$F(L, a, \sigma, J, \boldsymbol{r_i}, \boldsymbol{r_k}) = -\hat{\Gamma}(L, a, \sigma, J, \boldsymbol{r_i}, \boldsymbol{r_k}) + \lambda \cdot max \left[\hat{B}_{3dB}(L, a, \sigma, J, \boldsymbol{r_i}) - \eta, 0 \right] \quad , \tag{43}$$

where η represents the limit on the maximum MLW, and $\{L, a, \sigma, J\}$ is the set of geometric descriptors measured from a particular array geometry G. The maximum operation ensures a penalty added when MLW exceeds limit.

For a specified acoustic scene, probability density functions derived from the behavior pattern of speakers can be used to model the possible source distributions. Then the objective function of Eq. (43) can be expanded as

$$F'(\boldsymbol{G}) = \int_{r_i \in R_T} \left\{ \int_{r_k \in R_N} F(\boldsymbol{G}, \boldsymbol{r}_i, \boldsymbol{r}_k) p(\boldsymbol{r}_k | \boldsymbol{r}_i) d\boldsymbol{r}_k \right\} p(\boldsymbol{r}_i) d\boldsymbol{r}_i , \qquad (44)$$

where $p(\boldsymbol{r}_i)$ and $p(\boldsymbol{r}_k | \boldsymbol{r}_i)$ are the probability density functions representing the likelihood of positions for the desired target and noise sources. In the case when no prior knowledge is available, these can simply be set to uniform distributions. Therefore, the criterion to search for optimal array geometry is represented by

$$\boldsymbol{G}_{opt} = \underset{\boldsymbol{G} \subset R_M}{argmin} \langle F'(\boldsymbol{G}) \rangle , \qquad (45)$$

This criterion combining performance-based geometric descriptors and probabilistic descriptions of acoustic scenes will be applied as the objective function for the irregular array optimization strategies in Section 5.2, which greatly reduces the optimization processing time. Also the large SNR improvements observed in the experiments and real recordings for the optimized array geometries indicate the strong correlation between proposed objective function rules and array beamforming SNR performance for human speech in immersive environments.

3.2.3 *C3:* **Minimum Noise Power of Beamforming Estimator**

As discussed in Section 3.1, for a unbiased linear estimator, the recovered target signal can be given by:

$$\hat{U}_E(\omega; \boldsymbol{r}_i, \boldsymbol{G}) = \frac{\hat{\boldsymbol{W}}^H(\omega) \hat{\boldsymbol{V}}(\omega; \boldsymbol{r}_i, \boldsymbol{G})}{\hat{\boldsymbol{W}}^H(\omega) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r}_i, \boldsymbol{G})}$$

$$= \hat{U}(\omega; \boldsymbol{r}_i) + \int_{r_k \in R_N} \frac{\hat{\boldsymbol{W}}^H(\omega) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r}_k, \boldsymbol{G})}{\hat{\boldsymbol{W}}^H(\omega) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r}_i, \boldsymbol{G})} d\hat{N}(\omega; \boldsymbol{r}_k)$$

$$\hat{U}_E(\omega;\boldsymbol{r}_i,\boldsymbol{G})=\hat{U}(\omega;\boldsymbol{r}_i)+\int_{r_k\in R_N}\frac{\hat{\boldsymbol{H}}^H(\omega;\boldsymbol{r}_i,\boldsymbol{G})\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_k,\boldsymbol{G})}{\hat{\boldsymbol{H}}^H(\omega;\boldsymbol{r}_i,\boldsymbol{G})\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})}d\hat{N}(\omega;\boldsymbol{r}_k) \quad , \quad (46)$$

With the assumption that the target and noise signal propagation functions are orthogonal over $\boldsymbol{R}_N$, $\hat{U}_E(\omega;\boldsymbol{r}_i,\boldsymbol{G})$ are perfectly led to $\hat{U}(\omega;\boldsymbol{r}_i)$ [44], while the integral noise part of Eq.(46) approaching zero. For the case without this orthogonal assumption, the optimal array should minimize the noise component at the output of estimator to better recover the target signal. Therefore, for the case with single target $\boldsymbol{r}_i$, the noise power in the output of estimator can be considered as the criterion to assess array beamforming performance, as defined by [44]:

$$\hat{S}_n(\omega;\boldsymbol{r}_i,\boldsymbol{G})=\int_{r_k\in R_N}\left|\frac{\hat{\boldsymbol{H}}^H(\omega;\boldsymbol{r}_i,\boldsymbol{G})\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_k,\boldsymbol{G})}{|\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})|^2}\right|^2 dn(\omega;\boldsymbol{r}_k) \quad , \quad (47)$$

where $dn(\omega;\boldsymbol{r}_k)$ is the PSD measure over the noise space $\boldsymbol{R}_N$. For multi-source applications, Eq.(47) is rewritten as:

$$\hat{S}_n(\boldsymbol{G})=$$

$$\int_{\omega\in Speech}\int_{r_i\in R_T}\int_{r_k\in R_N}\left|\frac{\hat{\boldsymbol{H}}^H(\omega;\boldsymbol{r}_i,\boldsymbol{G})\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_k,\boldsymbol{G})}{|\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})|^2}\right|^2 dn(\omega;\boldsymbol{r}_k)du(\omega;\boldsymbol{r}_i)d\omega \quad ,$$

$$(48)$$

where $du(\omega;\boldsymbol{r}_i)$ is the PSD measure over the interested target space $\boldsymbol{R}_T$. The criterion for the optimal array selection is expressed as:

$$\boldsymbol{G}_{opt}=\underset{\boldsymbol{G}\subset \boldsymbol{R}_M}{argmin}\langle\hat{S}_n(\boldsymbol{G})\rangle \quad , \quad (49)$$

For the Minimum Variance Distortionless Respond (MVDR) filter application, which is the most widely used adaptive beamformer, the beamforming algorithm is defined based on the maximum likelihood rule as [2, 44]:

$$\hat{W}(\omega) = \frac{C_{NN}^{-1}(\omega; \boldsymbol{G}) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r_i}, \boldsymbol{G})}{\hat{\boldsymbol{H}}^H(\omega; \boldsymbol{r_i}, \boldsymbol{G}) C_{NN}^{-1}(\omega; \boldsymbol{G}) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r_i}, \boldsymbol{G})} \quad , \tag{50}$$

where $C_{NN}(\omega; \boldsymbol{G})$ is the auto covariance matrix of received noise signals, computed as:

$$C_{NN}(\omega; \boldsymbol{G}) = E[\hat{\boldsymbol{N}}^H(\omega; \boldsymbol{G}) \hat{\boldsymbol{N}}(\omega; \boldsymbol{G})] = \int_{\boldsymbol{r_k} \in R_N} \hat{\boldsymbol{H}}_H(\omega; \boldsymbol{r_k}, \boldsymbol{G}) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r_k}, \boldsymbol{G}) \, d n(\omega; \boldsymbol{r_k})$$

$$, \tag{51}$$

The estimated target signal is computed from:

$$\hat{U}_E(\omega; \boldsymbol{r_i}, \boldsymbol{G}) = \hat{W}(\omega) \hat{V}(\omega; \boldsymbol{r_i}, \boldsymbol{G}) \quad , \tag{52}$$

which has

$$E[\hat{U}_E(\omega; \boldsymbol{r_i}, \boldsymbol{G})] = \hat{U}(\omega; \boldsymbol{r_i})$$

$$\sigma^2[\hat{U}_E(\omega; \boldsymbol{r_i}, \boldsymbol{G})] = \frac{1}{\hat{\boldsymbol{H}}^H(\omega; \boldsymbol{r_i}, \boldsymbol{G}) C_{NN}^{-1}(\omega, \boldsymbol{G}) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r_i}, \boldsymbol{G})} \quad , \tag{53}$$

Note that Eq. (53) is independent with the beamforming algorithm. Therefore, the minimum variance of the output of estimator indicates the smallest error between recovered and original target signals and represents better target signal extraction ability of array. The criterion to search for the optimal array geometry can be expressed as [44]:

$$\hat{S}_{MVDR}(\boldsymbol{G}) = \int_{\omega \in speech} \int_{\boldsymbol{r_i} \in R_T} \frac{1}{\hat{\boldsymbol{H}}^H(\omega; \boldsymbol{r_i}, \boldsymbol{G}) C_{NN}^{-1}(\omega, \boldsymbol{G}) \hat{\boldsymbol{H}}(\omega; \boldsymbol{r_i}, \boldsymbol{G})} \, d u(\omega; \boldsymbol{r_i}) d\omega$$

$$\boldsymbol{G}_{opt} = \underset{\boldsymbol{G} \subset R_M}{argmin} \langle \hat{S}_{MVDR}(\boldsymbol{G}) \rangle \quad , \tag{54}$$

3.2.4 *C4:* Information Capacity of Array

In order to represent superior signal extraction ability against noise, previous criteria focus on minimizing the power of the noise components at the output of beamformer by varying array geometry according to specified beamforming algorithm and acoustic environment. Another thought to represent enhanced SNR performance of array is to

maximize the proportion of target component in the received signals of array by varying microphone positions.

As shown in Eq.(35) and Eq.(36), the signals received by the array can be divided into two parts. One is the transmitted target signal component, and the other is the transmitted additive noise component. Better target extraction ability in the presence of noise can be specified by high similarity or dependency between the transmitted target component and the total received signal of array. Take the single target and uncorrelated noise case for example (modeled by Eq. (36)), high dependency observed between $\hat{U}(\omega;\boldsymbol{r}_i)\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})$ and $\hat{V}(\omega;\boldsymbol{r}_i,\boldsymbol{G})$ received by specified array $\boldsymbol{G}$ represents the superior noise suppression ability of this array. In order to assess this dependency, mutual information derived from the Shannon information theory in communication system are applied, as [44]:

$$I\{\hat{U}\,\hat{\boldsymbol{H}};\hat{V}\}=\mathcal{H}\{\hat{U}\,\hat{\boldsymbol{H}}\}-\mathcal{H}\{\hat{V}\mid\hat{U}\,\hat{\boldsymbol{H}}\}=\mathcal{H}\{\hat{U}\,\hat{\boldsymbol{H}}\}-\mathcal{H}\{N\} \quad, \tag{55}$$

where $\mathcal{H}\{.\}$ denotes the entropy per sample of the corresponding random vector. For the single target at $\boldsymbol{r}_i$ it can be rewritten as [44]:

$$I(\boldsymbol{r}_i,\boldsymbol{G})=I\{\hat{U}\,\hat{\boldsymbol{H}};\hat{V}\}$$
$$=\int_{\omega\in Speech}\log[1+\hat{\boldsymbol{H}}^H(\omega;\boldsymbol{r}_i,\boldsymbol{G})C_{NN}^{-1}(\omega;\boldsymbol{G})\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})C_{UU}(\omega;\boldsymbol{r}_i)]\,d\omega$$
$$\tag{56}$$

where $C_{UU}(\omega;\boldsymbol{r}_i)=E[|\hat{U}(\omega;\boldsymbol{r}_i)|^2]$ denotes the auto covariance of the original target signal. Therefore, considering all the interested targets, the criterion to search for the optimal array geometry can be expressed as [44]:

$$I(\boldsymbol{G})=\int_{\omega\in Speech}\int_{\boldsymbol{r}_i\in R_T}\log[1+\hat{\boldsymbol{H}}^H(\omega;\boldsymbol{r}_i,\boldsymbol{G})C_{NN}^{-1}(\omega;\boldsymbol{G})\hat{\boldsymbol{H}}(\omega;\boldsymbol{r}_i,\boldsymbol{G})d\,u(\omega;\boldsymbol{r}_i)]\,d\omega$$

$$\boldsymbol{G}_{opt}=\underset{\boldsymbol{G}\subset R_M}{argmax}\langle I(\boldsymbol{G})\rangle \quad, \tag{57}$$

3.2.5   Constrains

Four criteria used to optimize and assess array beamforming performance are introduced in the previous sections. In applications, according to specific needs and prior knowledge of environment, the most appropriate criterion need to be selected as the objective function applied to evaluate the array performance in the search of optimal geometries. In addition, the microphone array optimization problems normally come with several constrains derived from the physical environment and subjective conditions of application, such as the largest array size, minimum spacing between elements to avoid unwanted coupling, and specific element placing area and topography. Also in order to maintain a constant output power, the coefficients of beamformer should be normalized according to the propagation functions. During the optimization procedure, the array pattern need to be controlled in certain ways to minimize or maximize the selected criterion. For example, when MLW is chosen as the object function, the system might have constrain that the sidelobe levels should be always smaller than a specific threshold, or some pattern nulls should be located in specific locations [45].

The ideal to address these constrained optimization problems is to build a new objective function combining the original criterion and the constrain conditions. Then the problems are turned into the normal optimizing problems without constrains. Lagrange multipliers can be used in developing this kind of optimization criteria for both equality and inequality constrained problems [46, 47]. As shown in Eq. (43), Lagrange multiplier is applied to combine the criterion of maximizing MPSR and adding penalty for violating MLW constrain. By adjusting the value of λ, more emphasis can be assigned to the desired part. In addition to the Lagrange multiplier approach, subspace theory is another common solution for the constrained optimization problems. Paper [48] solved the constrains problem by projecting the desired pattern function into the intersection lines of the subspace containing solution set of minimizing pattern residuals in the ML region and the subspace satisfying the sidelobe constrains.

3.3   Optimization Approaches

From the study of engineering optimization theory [46, 47], it is known that if the

expression of objective function and constrains are the simple, continuous, differentiable functions of design variables (microphone placements), the classical analytical methods using differential calculus in locating the optimum points can be applied, such as the regular array optimization problems in far field. In contrast, if the objective function and constrains cannot be stated as a clear function of design variables or they are too complicated to manipulate, such as the irregular microphone array problems in immersive environment, numerical optimum seeking methods are more useful than the classical ones. In general, numerical methods of nonlinear programming cases can be expressed by the following scheme [47]:

Step 1: Set a initial point of design vector X_1, which represents one specified array geometry in our cases.

Step 2: Search for a suitable direction S_i^*, pointing to the general optimum direction, i is the number of iteration. In the microphone array optimization problems, this optimum searching direction represents the way moving microphones to reach superior SNR performance.

Step 3: Find a suitable step length λ_i^* for moving X_1 along the direction S_i^*, which is related to the minimum distance between microphones and the number of microphone being moved in one iteration.

Step 4: Get the new approximation X_{i+1}, as $X_{i+1} = X_i + \lambda_i^* S_i^*$.

Step 5: Evaluate whether X_{i+1} is the optimal solution by the objective function. If yes, stop the iteration. Otherwise, let $i = i + 1$ and return to step 2.

Note that this optimization process contains two key aspects, the optimal search direction and the optimal step size. If the search direction is specified, the optimal step size can be considered as a one dimensional optimization problem. Increasing the step size can reduce the computational work, while reducing step size can be applied after bracketing the optimum range to increase accuracy. According to the searching method of the optimal direction, the optimization approaches can be divided into two classes: the direct searching methods and descent methods. Descent methods require the first/second order derivatives of the objective function to seek the optimum point along gradient direction, such as the Newton's method applied to search for the optimal irregular array by adding spatial perturbations to the regular placements (will be discussed in Section

5.1) [25, 49-58]. They have fast convergence speed, but might converge on local extreme point instead of global. The direct searching methods, such as the Genetic Algorithm (GA) methods applied to search for the optimal random microphone arrangements (will be discussed in Section 5.2), require only the objective function values, and are more suitable for simple problems involving small number of design variables. They have lower converge speed and higher computational complexity. They are less efficient than the descent methods, due to the reason that the moving direction of design variables are chosen all around, regardless of the gradient of the objective function. On the other hand, this kind of random search also bring benefit of less possibility to miss the global optimum point. The lower converge rate and higher computational complexity can be improved by adding certain guide to choose the moving direction. Therefore, it can be seen that there is a tradeoff between efficiency and reliability in search of optimum. The randomness of searching direction ensures to obtain global optimum, but brings more computing complexity and lower convergence speed. Searching along the descent direction can get fastest convergence speed but might miss the global optimal solutions. In order to take advantage of both methods, they are combined together in our study by the following principles:

(1)   Parallel combination, which means adding general direction guide for direct search method, or adding random perturbation of searching direction to each iteration of descent methods.

(2)   Series Combination. First applying direct method and large step size to search for the global optimum range, then applying descent method in this range to obtain a fast convergence rate. For example, for the irregular microphone array problems, GA can be used to search globally with a large step size and take the resulted array geometry as the initial condition of Gaussian iteration, which has a fast converging speed for the local optimum, to add small perturbations on the element positions of initial array to obtain the best SNR performance.

   By following these principles, feasible optimization methods for the design of regular and irregular arrays in immersive environments for speech applications will be proposed and evaluated in the Chapter 4 and Chapter 5.

3.4 Performance Evaluation

After applying the optimization approaches, a class of arrays providing good values of selected objective function are sorted out, which are expected to show optimal beamforming performance based on the probabilistic rules in corresponding acoustic environment. In order to demonstrate this prospect and assess the real performance of these arrays for speech signals, Monte Carlo experiments are performed to simulate the SNR performance of each array from the last iteration of the optimization searching procedure with various target and noise occurrences according to the knowledge of acoustic scene. Human voice samples or colored noise generated by the SII model which emphasizes the frequency bands most important to speech intelligibility are applied as the excitations for both target and noise sources. The typical choices [14] of the physical conditions of acoustic environment are given in Table 3.1, and the evaluation procedure is describe as below:

Step 1: Choose the environmental parameters according to the preknowledge of acoustic scene, including room size, reverberation level, probability density functions $p(\boldsymbol{r}_i)$ or PSDs of the target and noise positions.

Step 2: Pick one array from the last iteration of the optimization procedure. Measure the geometry descriptors of this array for later discussion about the optimal pattern of microphone clusters.

Step 3: Generate a set of target sources with SII spectrum, while the source positions are chosen according to predefined statistic measures of target space, such as $p(\boldsymbol{r}_i)$ and PSD.

Step 4: Distribute the noise sources: (1) the interfering noise is simulated by a set of speech signal sources distributed in the noise space according to $p(\boldsymbol{r}_k|\boldsymbol{r}_i)$, the probability density function representing the likelihood of positions for the noise sources. The noise power can be obtained by the predefined power ratio of target to noise source. In our experiments, the power of target and noise sources are all normalized to 1; (2) the reverberation noise is simulated by using the image method to add additional

noise sources as shown in Figure 3.2. The amplitude of reverberated signal is attenuated by *1/d* and the reflection coefficient of corresponding wall, where *d* is the propagation distance.

Step 6: With the beamformer focused at each target position (ignore the localization errors), estimate the received target signals by microphones and the output of beamformer derived from the target source.

Step 7: With the beamformer focused at each target position (ignore the localization errors), estimate the received noise signals by microphones and the output of beamformer derived from all the noise sources.

Step 7: Compute the SNR when beamformer focusing on each target. Then compute the average and the variance of SNRs over all target sources by probabilistic rules according to $p(\boldsymbol{r}_i)$ and PSD.

Step 8: Discuss the SNR performance of each array according to the application requirements (details will be discussed in Section 5.2.2), choose three top performed arrays as the optimal ones for real recording.

Table 3.1: Physical parameters of acoustic environment

Room Size（meter^3）		Reflection Coefficient		Power Ratio of Target and Noise Source	Multi-path Simulation Level
Small room	3*3*2	Low	0.2	Choose form the range -50dB to 50dB	8th images as shown in Figure 3.2
Medium room	5*5*2	Medium	0.5		
Large room	10*10*2	High	0.8		
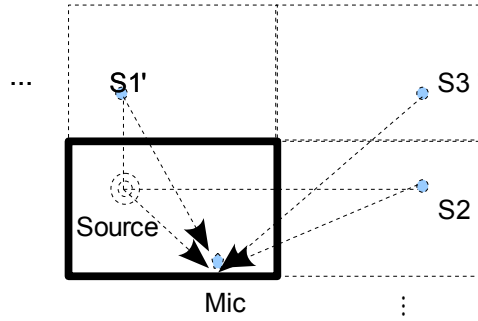
Figure 3.2: Image method to simulate multi-path effect. S1', S2', and S3' represent the virtual reverberation noise sources.

This chapter introduced the basic principles about the microphone array optimization problems, including the general problem formulation, common used criteria and constrains, general optimization approaches, and the experimental evaluation. Based on these principles, the next two chapters will discuss the optimization strategies for revised regular arrays and irregular arrays, providing enhanced SNR performances in immersive environment applications.

Chapter 4   Array Design Based on Regular Geometry

After a general review of microphone array processing technique, building the problem formulation, and applying appropriate criteria and constrains to describe concerns and restrictions in real applications, common used array geometries and optimization approaches derived from the regular arrangements are discussed in this chapter. Most of previous optimization approaches are directly derived from antenna array theories based on far-field propagation mode, which usually assume a uniform and regular spaced microphone arrangement to simplify the problem formulation and apply analytical solutions to compute the optimal microphone positions. Since the regular equispaced arrays are limited in their ability in speech capture, these results are not as useful for immersive environment that typically occur for surveillance and smart room applications. This chapter, therefore, introduces geometries mutated from the traditional equispaced arrays to overcome their limits and provide superior SNR performance for speech capture in immersive environments, while certain regularity of element arrangement is reserved for easy installation and operation.

4.1   Limitations of Equispaced Arrays

The traditional equispaced array suffers from several significant problems. First of all, due to the spatial aliasing problem caused by the regularity of element placements, it only works for narrow-band signals. According to Nyquist sampling theory, in order to avoid

spatial aliasing, the inter-element spacing cannot exceed $\lambda_{min}/2$ [59, 60]. Considering human speech signals with the frequency range from 50Hz to 8000Hz, elements should be placed as closed as 2 centimeters to each other. Obviously, it is not achievable in real cases. Even if this condition can be met, the spatial resolution for the low frequency bands will be sacrificed greatly. In addition, since the array aperture should be no smaller than $\lambda_{max}/\theta_{MLW}$, where the denominator is the mainbeam width in radians, the number of microphone should be greater than $2\lambda_{max}/\theta_{MLW}\lambda_{min}$ [44, 59], which might be too large for most indoor applications.

Secondly, because the traditional equispaced linear arrays for far-field applications only have one parameter θ for steering and the elements spread all in one dimension, they cannot resolve sources located at the rotationally symmetrical positions of array or the sources from the same direction [26, 61]. In other words, these arrays have only one degree of spatial selectivity, which usually not sufficient for real applications. Therefore, extension to 2D and 3D arrays with unequal spacing is necessary for human speech applications in immersive environments.

4.2   Mutation of Regular Configurations

This section discusses several array geometries mutated from the traditional equispaced arrays, that overcome the limitations of equispaced arrangements and keep certain regularity of microphone positions for easy installation.

4.2.1   Harmonic Nested Array

In order to eliminate the problem of spatial aliasing and extend applicable bandwidth for broad-band signal capture, harmonic nested array is applied [14, 61]. The main ideal of this array geometry is to divide the broad signal band into several subbands and use equispaced subarrays to capture each narrow-band signal. Through a proper division of subbands and changing the inter-mic distance in subarrays, the beam width of recomposed pattern can remain constant over frequencies. And the total number of microphones can be reduced by sharing elements in the overlapping positions of each subarray. Figure 4.1 shows 3 subarrays used for each octave, and inter-mic spacing is halved across next band.

The simulation results in paper [14] shows that for the harmonic array with 41 elements, the beam width remains constant over 6 octave interval, which is sufficient to cover 50Hz to 7000Hz signals without spatial aliasing. In addition, this array can be expanded to 2D and 3D arrays to achieve a better spatial selectivity.
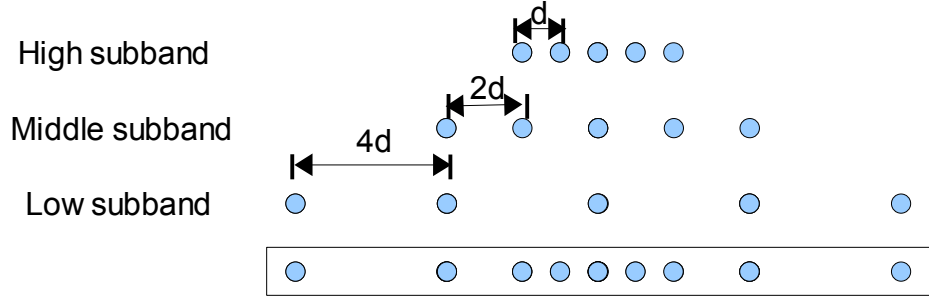
Figure 4.1: A harmonic array with 9 microphones (adapted from [14]). *d* is the unit distance.

The procedure to build this type of logarithmically spaced array can be described as below. First, set up the uniform spaced array for highest subband. Then progressively add more element for lower band on both end. The general expression of this procedure is given as [27]:

$$r_p = p\frac{\lambda_{min}}{2}, \ when \ 0 \le p \le \frac{Q}{2}$$

$$r_{p+1} = \frac{Q}{Q-1}r_p \ , \ when \ p > \frac{Q}{2}, r_p < (Q-1)\frac{\lambda_{max}}{2}$$

$$r_{-p} = -r_p \qquad\qquad , \qquad (58)$$

where the origin is located in the center of array. r_p is the distance of the pth microphone from origin. r_{-p} represents the symmetric microphone of r_p. Q is a constant related to the division of subband. When Q = 2, the harmonic nested array is produced.

The simulated gain patterns of harmonic nested arrays with 9 microphones and

68

corresponding equispaced arrays are shown in Figure 4.2. Figure 4.2(a)(c)(e) present the gain patterns for 1500Hz signal, while Figure 4.2(b)(d)(f) present gain patterns for human speech signals simulated by SII model. It can be seen that the harmonic nested array containing three subarrays (Figure 4.2(c)) eliminated the severe aliasing problem for 1500Hz signal, as shown for the equispaced arrays in Figure 4.2(a). Figure 4.2(e)(f) give the results of harmonic nested array with the same dispersion of equispaced array. When comparing with Figure 4.2(a)(b), a comparable spatial resolution (MLW) is provided by the harmonic array, while the spatial aliasing problem are successfully eliminated. Therefore, with fixed array centroid and aperture, harmonic nested array can provide better beamforming performance for speech applications. In the later section, it will be applied in 2D and 3D space for extra degree of spatial selectivity.
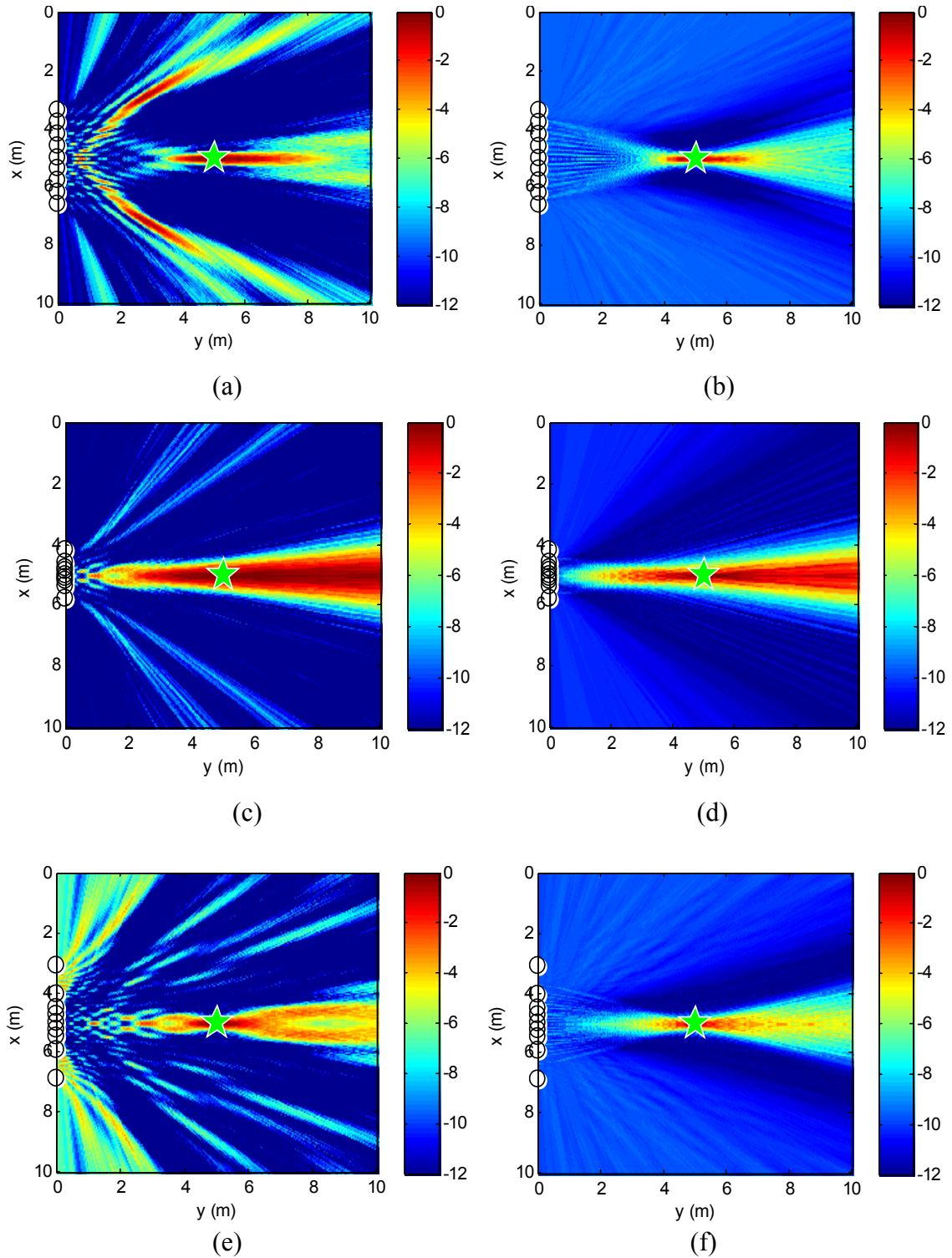
Figure 4.2: Top view gain patterns of equispaced arrays and corresponding harmonic nested arrays with 9 microphones. The beamformer is focused on the same point at the center marked by star, where white dots represent microphone positions. (a) Equispaced array with 10cm inter-mic distance and 1500Hz signal. (b) Equispaced array with 10cm

inter-mic distance and SII signal. (c) Harmonic nested array with the largest inter-mic distance of 10cm and 1500Hz signal. (d) Harmonic nested array with the largest inter-mic distance of 10cm and SII signal. (e) Harmonic nested array with the same dispersion of the equispaced array and 1500Hz signal. (f) Harmonic nested array with the same dispersion of the equispaced array and SII signal.

4.2.2   Minimum Redundancy Array

From previous study, the size of main beam and the maximum intensity of gain patterns relatively won't be affected by the precise element positions, while the array aperture and number of element is fixed [2, 45, 62]. The goal of Minimum Redundancy Array (MRA) is to achieve maximum spatial resolution for a given number of elements by reducing redundant inter-mic spacings as suggested in [31-33, 59, 63-68]. This kind of arrays focus on the pairwise distance of elements based on the assumption that the inter-mic spacings always be a multiple of unit distance, while having the covariance matrix with the least repeated entries. As shown in Figure 4.3(b), for a array with N elements, the zero redundancy array samples the spatial frequency spectrum at uniform intervals with uniform distribution, while providing identical lags with a equispaced linear array of

$$\frac{1}{2}N(N-1)+1 \quad \text{elements.}$$

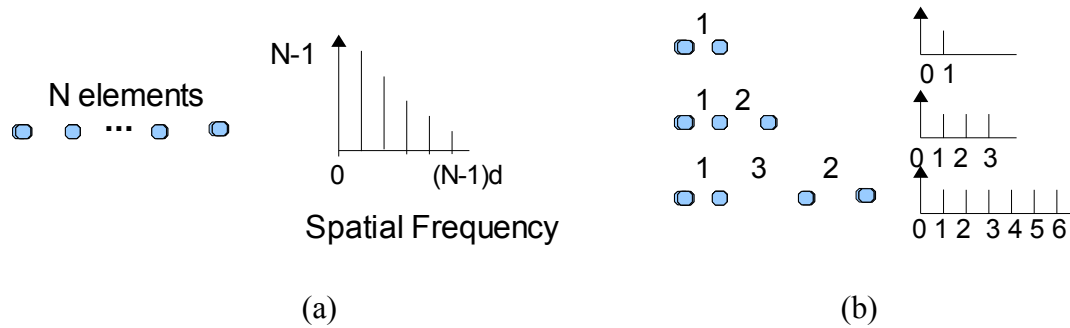(a)   (b)

Figure 4.3: Spatial frequencies of MRA and corresponding equispaced array (adapted from [31]). d is the unit distance. (a) Equispaced linear array. (b) Zero redundancy linear array.
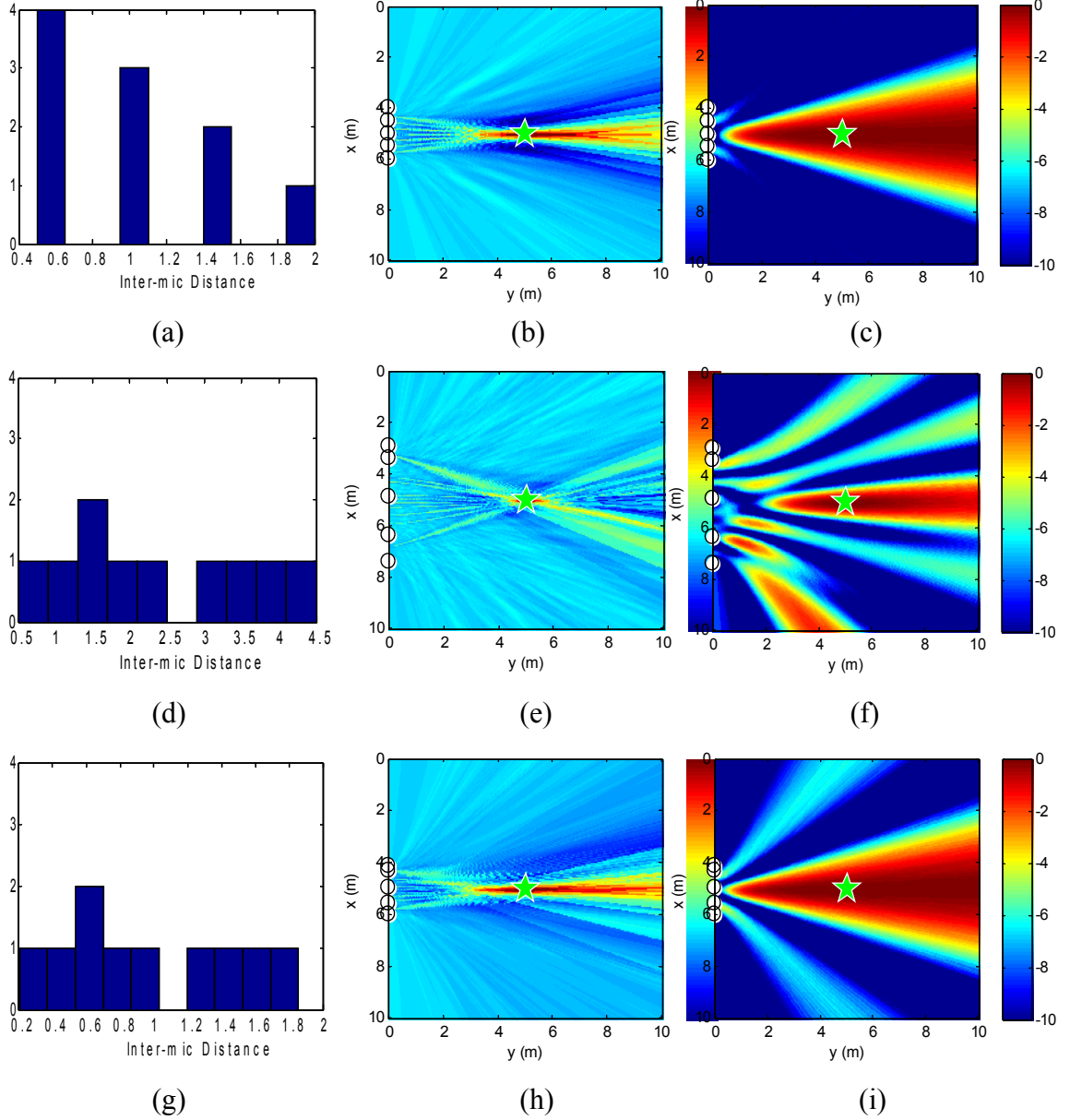
71

The methods to search for microphone placements with low redundancy have been well studied and published in many papers. In [31], Moffet shows several possible sampling schemes of MRA. A fast computing method to search for linear MRAs is proposed by Linebarger [64]. In paper [65], a table of MRAs without missing lags has been compiled by Leech. Christopher [67] uses a numerical implementation of annealing process to guide the random search for the low redundancy linear array and shorten the running time for the arrays with large number of elements. Therefore, in our experiment, the microphone placements of MRA according to the predefined unit distance are selected directly from a table including possible arrangements of MRA with different number of microphones. A part of this table is shown in Table 4.1 [31, 68].

Paper [32, 66] compared the performance of MRA with uniformly spaced linear array for different number of elements in terms of interference cancellation. Take the array with 5 microphones for example. Figure 4.4 gives the gain patterns and histograms of inter-mic distance of MRA and corresponding equispaced linear array. Figure 4.4(d)(e)(f) show the MRA with same unit distance of equispaced array. Figure 4.4(g)(h)(i) show scaled MRA with the same dispersion of equispaced array. It is proved that by minimizing the redundancy of inter-mic distance, MRA can greatly reduce the number of elements while following the ML region of uniformly spaced array, but the cost of this improvement is a higher sidelobe level of beampattern. In conclusion, for a given number of elements, MRA can provide maximum spread of microphone to achieve best spatial resolution by reducing redundant inter-mic spacings. Other techniques, such as adaptive beamformer aiming at suppressing sidelobe level, need to be considered as the compensation when applying MRA.

Table 4.1: MRA configurations

Number of Mics.	Aperture	Configuration (inter distance between nearby mics)
5	9	1 3 3 2
	11	1 3 5 2
	….	
6	13	1 5 3 2 2
	17	1 3 6 2 5
	….	
7	17	1 3 6 2 3 2
	25	1 3 6 8 5 2
	….	
8	23	1 3 6 6 2 3 2
	34	1 3 5 6 7 10 2
9	29	1 3 6 6 6 2 3 2
	44	1 4 7 13 2 8 6 3
	….	
10	36	1 2 3 7 7 7 4 4 1
	55	1 5 4 13 2 8 6 3
	….	
11	43	1 2 3 7 7 7 7 4 4 1
	….	

Note: All distances are expressed as the multiples of unit distance *d*.

Figure 4.4: Spatial frequencies and gain patterns of MRA and corresponding equispaced array with 5 microphones. The unit distance is set to 0.5 meter. The beamformer is focused on the same point at the center marked by star, where white dots represent microphone positions. (a) Histogram of inter-mic distance of equispaced linear array. (b) Gain pattern of equispaced array for SII signal. (c) Gain pattern of equispaced array for 300 Hz signal. (d) Histogram of inter-mic distance of MRA. (e) Gain pattern of MRA for SII signal. (f) Gain pattern of MRA for 300 Hz signal. (g) Histogram of inter-mic distance of MRA with the same dispersion of equispaced array. (h) Gain pattern of MRA with the same dispersion of equispaced array for SII signal. (i) Gain pattern of MRA with the same

74

dispersion of equispaced array for 300 Hz signal.

4.2.3   Spatial Extension of One Dimensional Configurations

Previous sections introduced the limitation of equispaced linear array and two mutated regular linear arrays with performance improvements. In order to obtain extra degree of spatial selectivity, these linear arrays are extended to 2D and 3D space for the speech applications in immersive environments.

The basic regular one dimensional configurations applied for array spatial extension in our study are listed as below:

- Equispaced linear arrays (ELA)
- Harmonic nested arrays (HNA)
- Minimum redundancy arrays (MRA)

In order to show similar performance improvements in multidimensional space, these one dimensional configurations are extended to 2D and 3D design space to result in extra degree of spatial selectivity. Several common used configurations are shown as below:

(1)   Two dimensional extending arrays.

- Perimeter arrays, as in Figure 4.5(a).
- Planar arrays, as in Figure 4.5(b).
- Cross arrays, as in Figure 4.5(c).
- Wheel arrays, as in Figure 4.5(d), where the linear configurations are arranged as the spokes in a wheel and tilted the same angle from the radial direction of the center [69].

Note that the linear configurations (as shown in the black box) applied in these two dimensional arrays can be selected from any of the basic one dimensional arrays to obtain different performance improvement. The typical gain patterns of these two dimensional extending arrays for speech signals are shown in Figure 4.6 and 4.7. From Figure 4.6, it can be seen that the perimeter array derived from ELA exhibits high sidelobes along the directions of symmetric lines of microphone arrangements. And the perimeter array of MRA eliminates these high sidelobe ridges by breaking the uniformity of microphone

placements of ELA while following the area of main lobe. Figure 4.7 compared the gain patterns of wheel array and corresponding cross array. Since each spoke of wheel array is tilted the same angle away from the radial direction, their positions can be defined by the vertical distance d from spoke to the center of wheel. And the cross array shown in Figure 4.7(b) can be considered as a special case of four-spokes wheel array when $d=0$. As show in Figure 4.7, by adding the lateral offset d of each linear spoke from the array center, the maximum sidelobe ridges of cross array protruding along the vertical direction of microphone lines are successfully suppressed by the wheel array, while the MLW of wheel array is only slightly broaden.

(2) Three dimensional extending arrays.
- 3D arrays by applying regular 1D configuration in each direction of Cartesian coordinate axes, as shown in Figure 4.8(a).
- Platonic solid arrays, such as equilateral triangle, tetrahedron in Figure 4.8(b)

Figure 4.9 gives the top view and 3D gain patterns of tetrahedron arrays for example. It can be seen that when the target is located outside the tetrahedron, due to the symmetry of microphone placements of tetrahedron array, high sidelobes are observed in the symmetrical positions according to the microphone arrangements. Meanwhile, this kind of arrays show superior beamforming ability for the targets inside the tetrahedron space near the centroid, as shown in Figure 4.9(c).
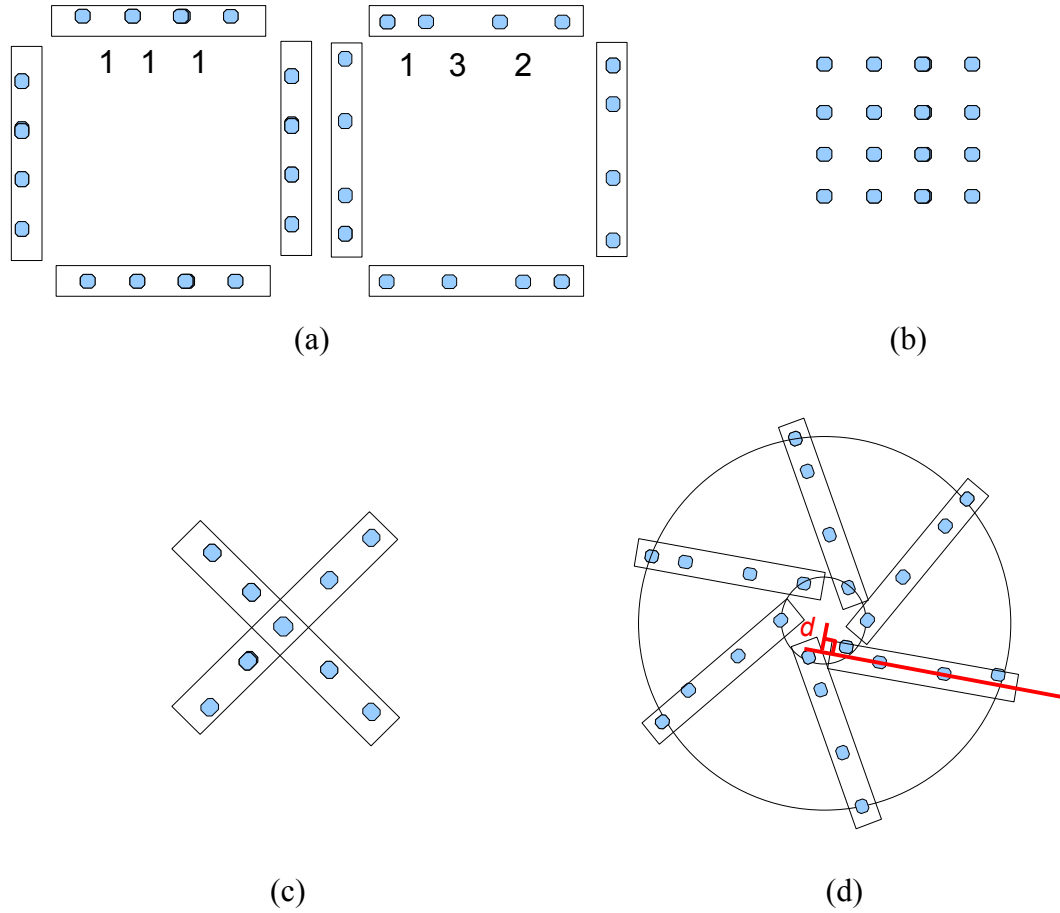
Figure 4.5: Two dimensional extending arrays. (a) Perimeter arrays derived from equispaced linear array and MRA. (b) Planar array derived from equispaced linear array. (c) Cross array derived from MRA. (d) Wheel array derived from MRA.
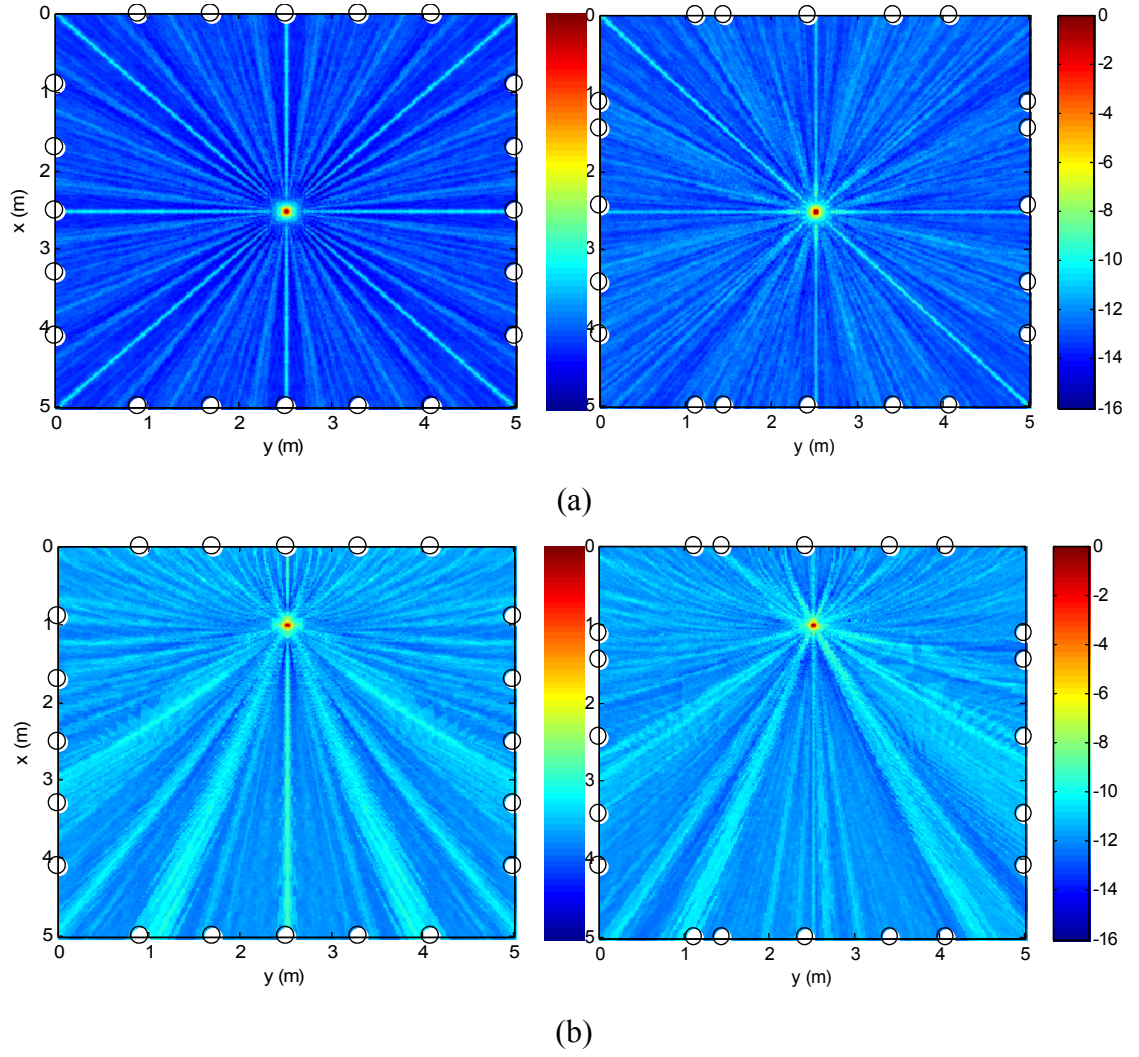
Figure 4.6: Top view gain patterns of perimeter arrays derived from ELA (left) and MRA (right) with the same dispersion for speech applications. Black circles represent microphones. (a) Steering at the center of FOV. (b) Steering at the spatial point [1, 2.5] in FOV.
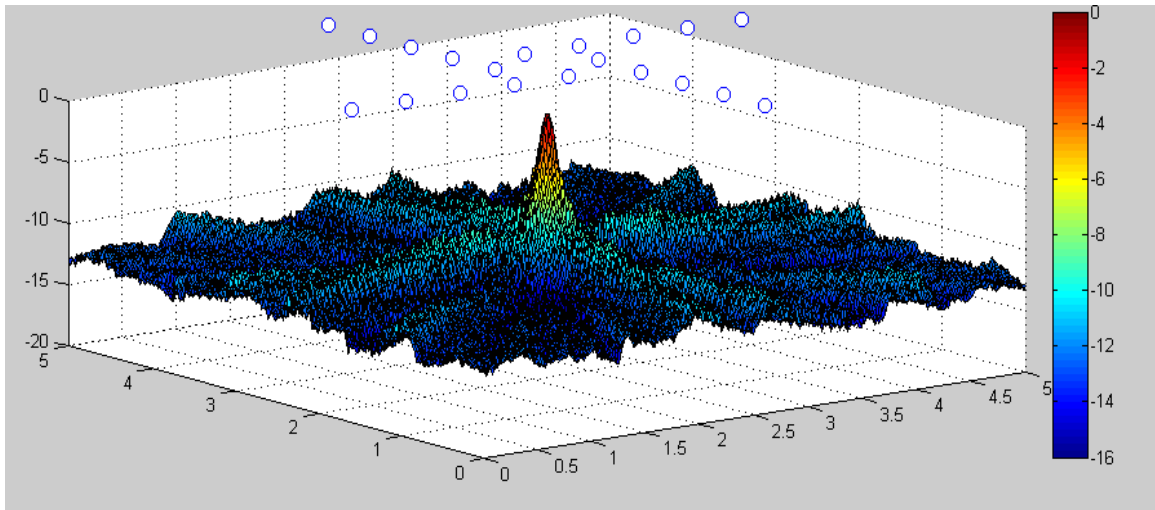
(a)
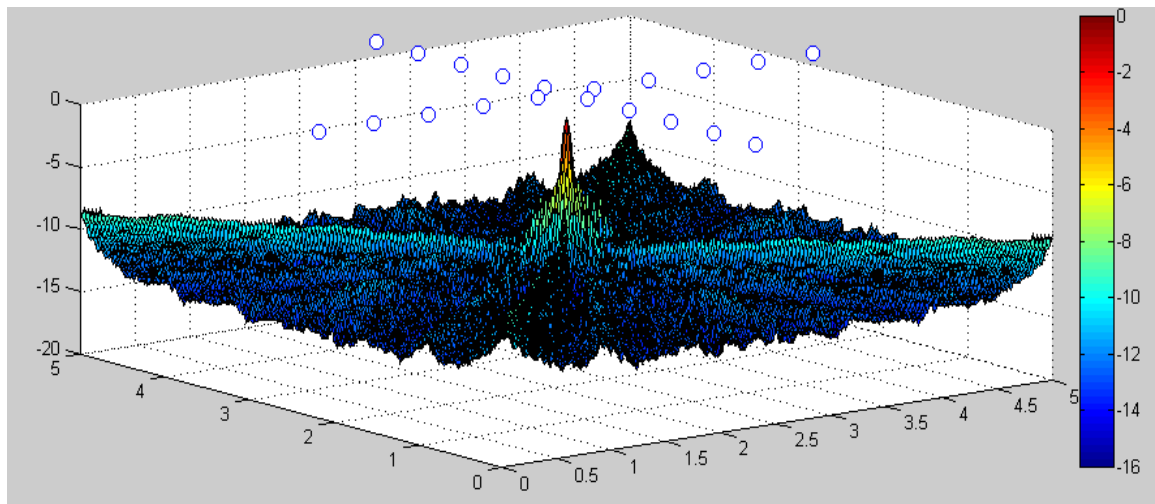
(b)

Figure 4.7: Gain pattern comparison. (a) wheel array with four spokes. (b) Corresponding cross array. Blue circles represent microphones.
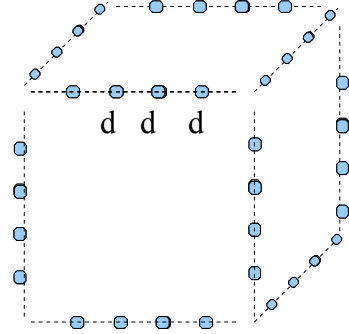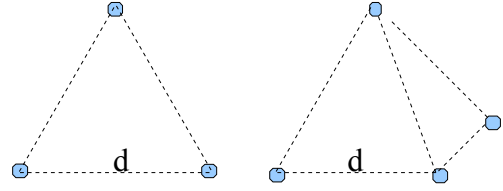
Figure 4.8: Three dimensional extending arrays. *d* is the unit distance of microphone spacing. (a) Regular 3D extending array. (b) Platonic solid array.

(a)
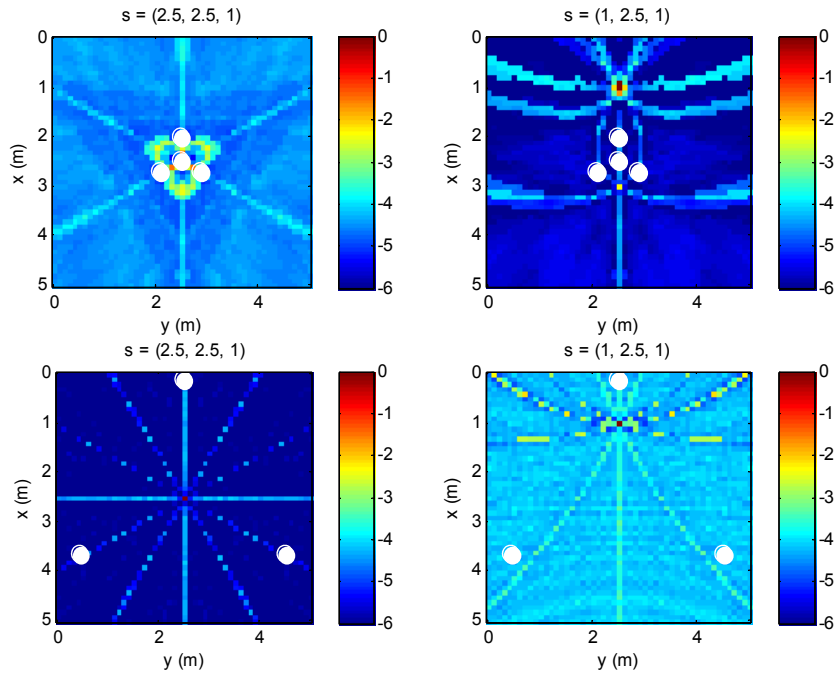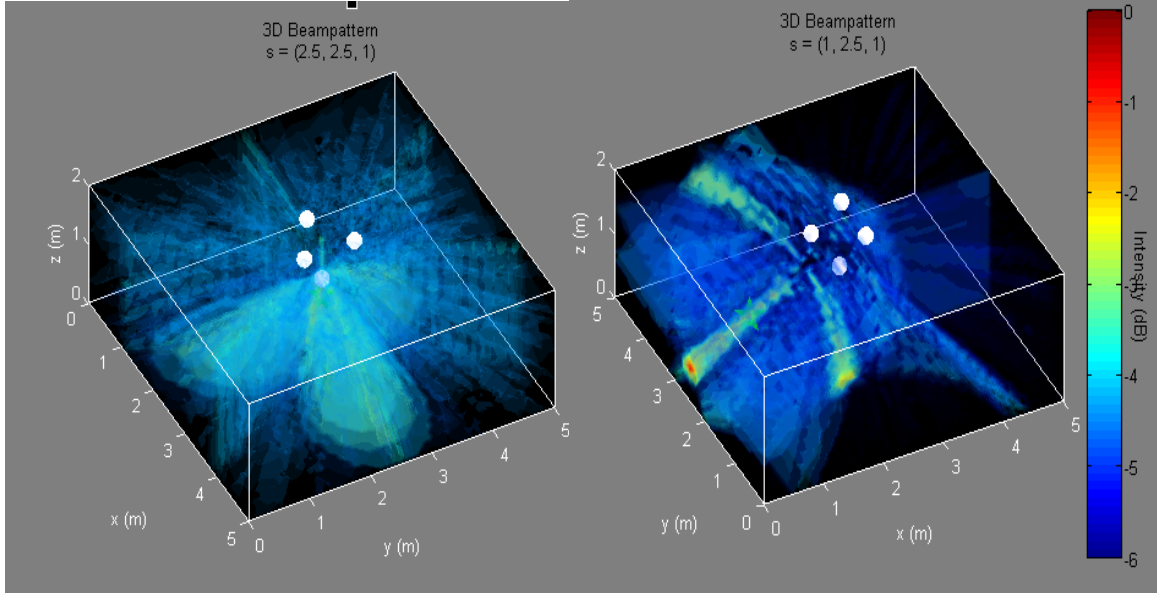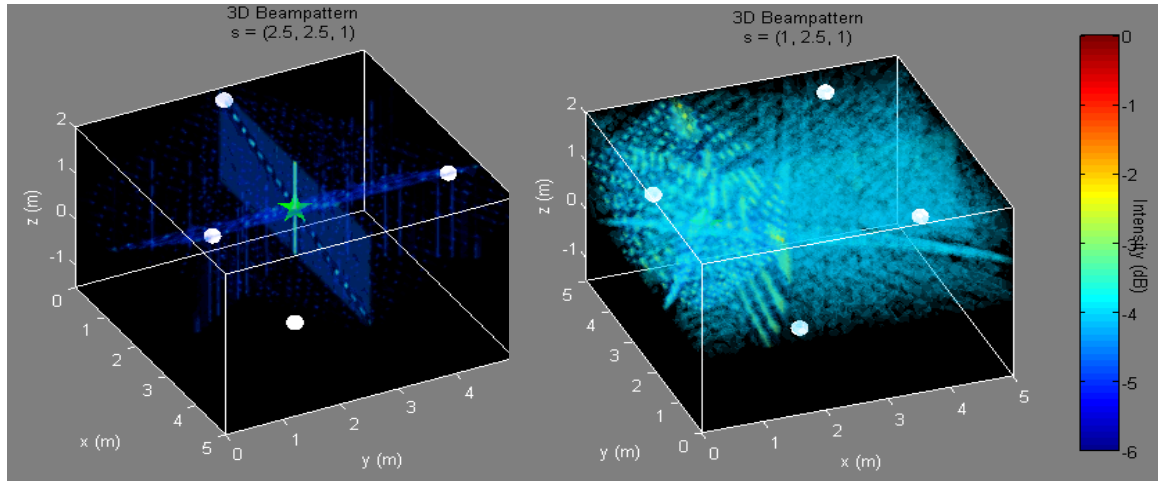
Figure 4.9: Gain patterns of tetrahedron array. White dots represent microphones. Stars represent target points. (a) Four top view gain patterns: small ceiling array steering at the center of FOV, small ceiling array steering at the left corner of FOV, large array with the centroid in the center of room when steering at the center of FOV, large array with the centroid in the center of room when steering at the left corner of FOV. (b) 3D View of small ceiling array, when steering at the center of FOV (left) and the left corner of FOV (right). (c) 3D View of large array with the centroid in the center of room, when steering at the center of FOV (left) and the left corner of FOV (right).

4.2.4 Edge Cutting of Regular Configurations

Inspired by the theory of exciting current tapering in antenna array, it is proved that by cutting some fringe or corner microphones of planar array can reduce the maximum sidelobe level greatly [54]. GA are applied in paper [54] to "turn on/off" the elements in the fringe areas and search for the theoretical optimal geometries with the mini-max sidelobe levels. Two similar approaches are applied in our study to adjust the edge elements of regular arrays to obtain the best MPSR performance.

- Based on the regular configurations proposed in the previous two sections, directly cut the corner elements as shown in Figure 4.10. Then assign a set of binary wights for each elements located in the fringe areas, while 1 represents "turn on" the microphone and 0 represents "turn off" the microphone. Apply GA to search for the optimal set of these binary wights until the sidelobe levels of array gain pattern meet required threshold. Finally, according to the optimal choice of the binary wights, a new array configuration are generated.

- Based on the traditional theory of adaptive beamformer, according to the assigned weights of elements in the regular configurations as proposed in previous two sections, remove the elements whose weights are lower than the predefined threshold, and keep the others with the higher weights.
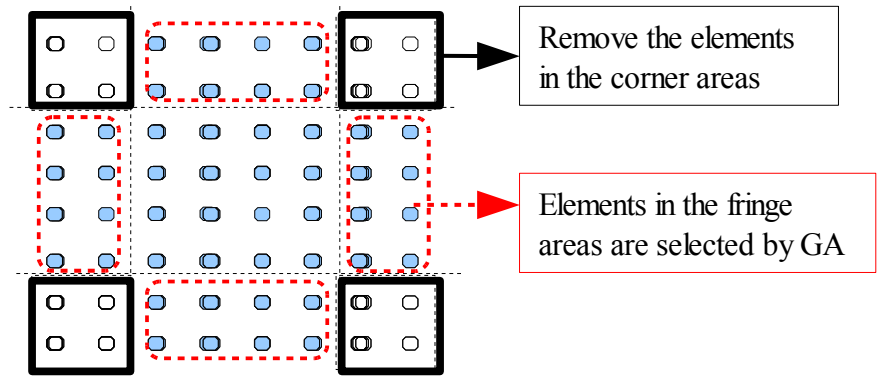

Figure 4.10: Edge cutting of regular planar arrays (adapted from [54]). The blue dots represent microphones. Black boxes represent corner areas, while red dashed boxes represent fringe areas.

In order to evaluate the performance of edge cutting arrays, Table 4.2 gives the simulated results of edge cutting arrays based on the regular planar configuration with 17*17 elements. The corresponding microphone arrangements are given in Figure 4.11, and the gain patterns are shown in Figure 4.12. Four corner areas of 3*3 elements are cut for the corner cutting array, while four triangle corner areas of 6 elements are removed to form the half corner cutting array. The four fringe areas of 3*11 elements are optimized by GA to minimize maximum sidelobe level while steering at zenith. It can be seen that, even with less elements than equal spaced planar arrays, the corner cutting array and GA arrays can efficiently reduce the sidelobe level at desired steering angle, while only slightly broaden the mainbeam width due to the decrease of array dispersion. In other steering direction (large array centroid offset), because the equal spaced planar array has the most microphones and largest spread, it shows a more stable performance than the others for the targets away from the centroid, but these edge cutting arrays can still obtain several good results at some angles. Furthermore, their performance can be improved by changing the optimization objective function of GA, or increasing the number of elements.

Table 4.2: Performance comparisons of edge cutting planar arrays

Steering Angle	Performance Metrics	Equispaced Planar Array	Half-corner-cutting Array	Corner-cutting Array	Corner-cutting & GA Array
0°	MLW	6.1276^o	6.2840^o	6.2623^o	6.7306^o
	MPSR	18.82dB	19.21dB	19.3dB	19.63dB
15°	MLW	6.3345^o	6.5256^o	6.6962^o	7.1917^o
	MPSR	17.92dB	17.95dB	18.24dB	17.39dB
30°	MLW	6.3223^o	8.0005^o	7.8390^o	8.3250^o
	MPSR	17.55dB	17.1dB	15.57dB	16.09dB

3 mics

17 mics

3 mics

17 mics

Fringe area with 3*11 mics, selected by GA

17 mics

(a)   (b)   (c)
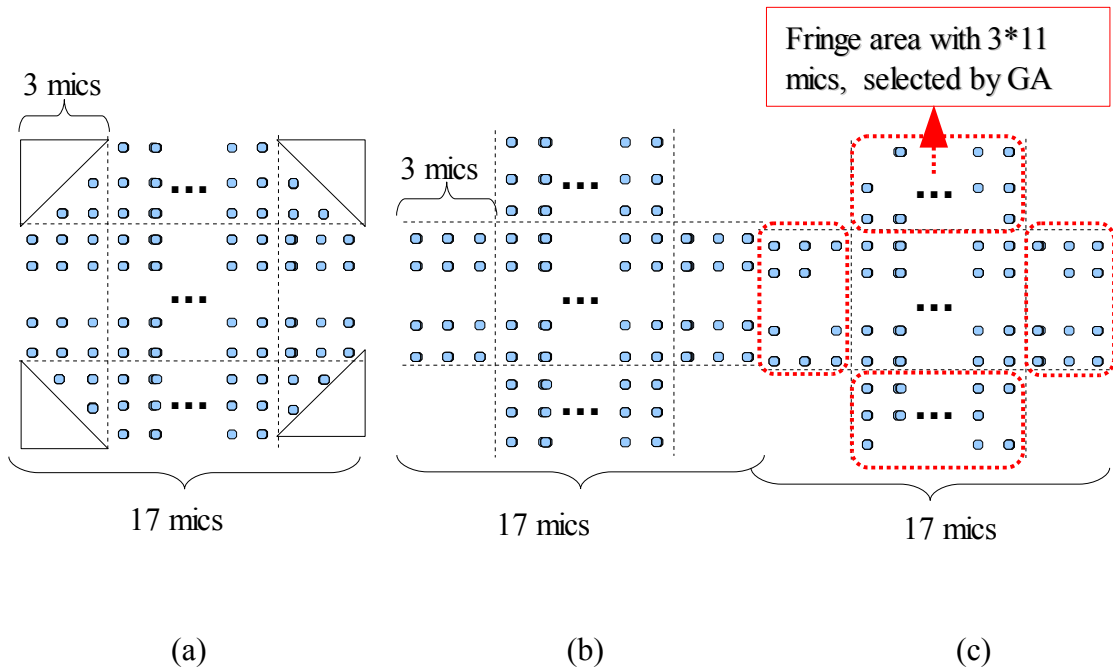
Figure 4.11: Examples of edge cutting arrays for the performance comparison based on 17*17 regular planar array. (a) Half corner cutting array. (b) Corner cutting array. (c) Corner cutting & GA array.

(a)

(b)

(c)

(d)
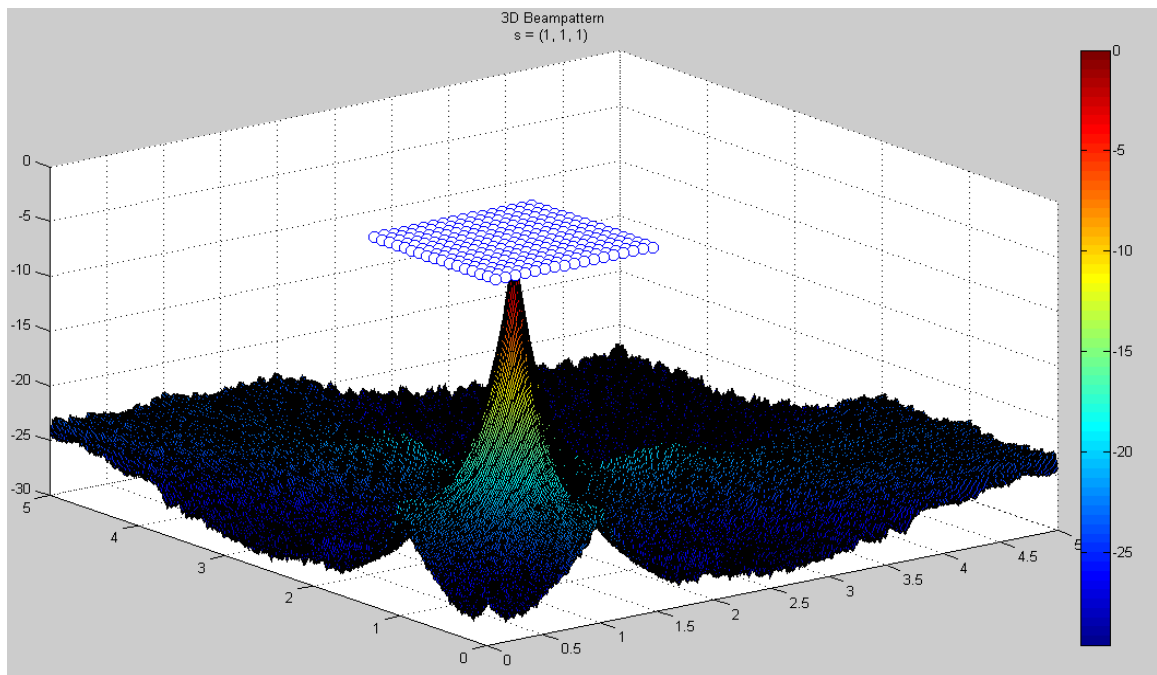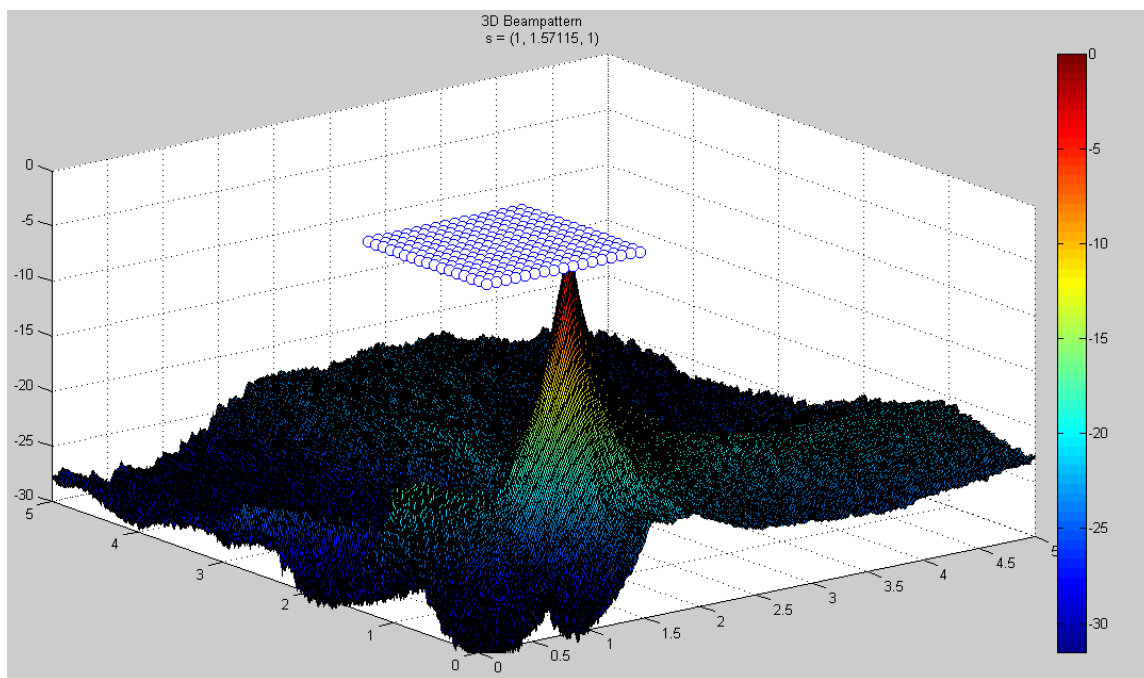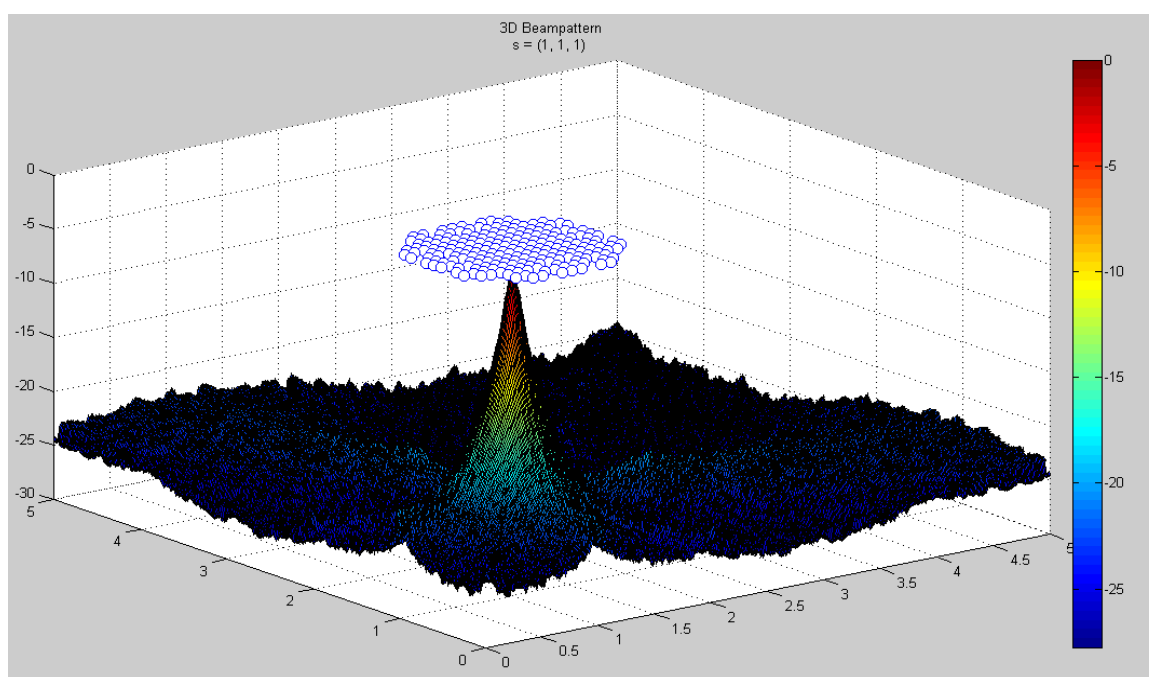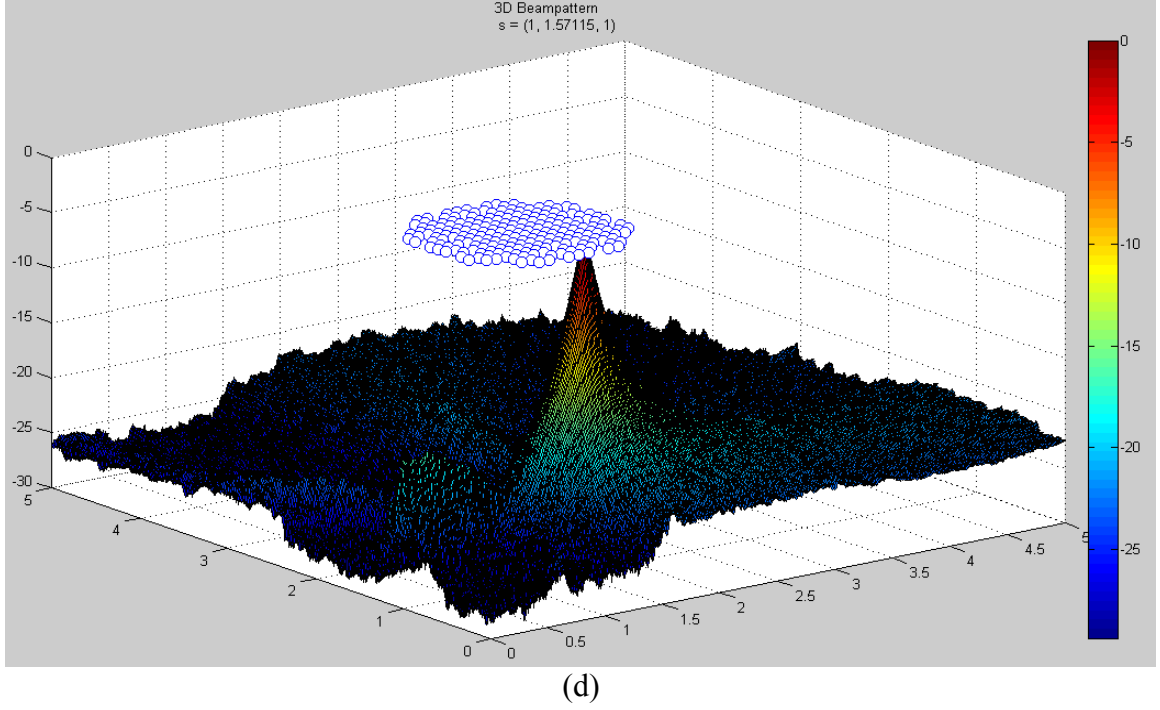
Figure 4.12: Gain pattern comparisons of equispaced planar array vs. corner cutting & GA array on the 5*5 m^2 target plane. Blue dots represent microphones. (a) Equispaced planar array with 0° steering angle. (b) Equispaced planar array with 30° steering angle. (c) Corner cutting & GA array with 0° steering angle. (d) Corner cutting & GA array with 30° steering angle.

4.2.5 Spiral Array

It has been discussed before that due to the periodicity of element placements, the application of regular equispaced arrays for broadband speech signals is limited. Generally, the highest frequency band forces the inter-mic spacing to be smaller to avoid spatial aliasing problem, while the lowest frequency band determining the tolerant minimum array aperture for required spatial resolution. Also, it has been demonstrated from previous gain pattern analyses that the high sidelobe ridges occur in the direction of symmetric/periodic line of microphone arrangements. Therefore, the irregular arrays which break the periodicity of element placements are expected to show better

86

beamforming performance for speech applications. Figure 1.1 gave the example of completely randomly distributed array with superior performance when comparing with corresponding regular planar array. Chapter 2 identified the important statistic geometry descriptors of these irregular arrays to predict or control the beamforming performance for speech signals. However, since the microphone positions are totally random, this kind of irregular arrays are difficult to manufacture and operate in some application cases [69]. Spiral arrays, which maintaining certain regularity for easy manufacturing and eliminating the periodicity/redundancy of microphone spacing of traditional regular planar array to suppress sidelobe levels, are proposed in this section.

Inspired by the natural growing pattern of sunflower head, several spiral arcs are developed by following two basic principles, the most uniform angular distribution of branches/flowers (by applying the golden/Fibonacci angle), and most uniform area distribution of seeds (by applying different spiral arcs) [70], where the seeds and flowers are considered as the microphones in our study. These principles of the uniform coverage eliminate the possibility of redundant or symmetric microphone placements according to the target, and suppress the worse-case sidelobes greatly. The general formulation of the spiral arcs in polar coordinates (r, φ) can be expressed as:

$$\varphi_p = p\,\delta; \quad r_p = f(\varphi_p); \quad p = 1, 2, \dots, P \quad , \tag{59}$$

where p is the order of microphone counting outward from the center. P is the number of microphone. δ is the radiation angle difference from center between nearby microphones, which is usually set to $137.5°$ as the golden angle to uniformize the element distribution over the circle [70]. $f(.)$ is the function defining the shape of spiral arc. Normally, there are two classes of spiral arc definitions. The first class can be expressed as:

$$r_p = a + b\,\varphi_p^{1/k} \quad , \tag{60}$$

where the value of k determines the type of spiral arc, affecting microphone density distribution.

when k = 1, it represents Archimedean spiral with equal space between successive turnings as shown in Figure 4.13(a), where changing a turns the spiral and b controls the

distance between successive turnings [71]. The area for single seed/microphone can be computed as:

$$\frac{S}{p} = \frac{\pi (a + b \varphi_p)^2}{p} = \pi (\frac{a^2}{p} + b^2 \delta^2 p) + 2ab \pi \delta \quad , \tag{61}$$

Because the microphone area is a monotonic increasing function of the microphone number p, arrays generated by this spiral have the highest microphone density at the center and decreasing density when moving outward from the center, as shown in Figure 4.14(a).

when k > 1, Eq. (60) represents spirals with decreasing space between successive turnings, which add more microphones in periphery areas of the arrays. One important example is Fermat's spiral with $a = 0$ and $k = 2$, as:

$$r_p = b \sqrt{\varphi_p} \quad , \tag{62}$$

Then, the area for single seed/microphone is computed as:

$$\frac{S}{p} = \frac{\pi (b \sqrt{\varphi_p})^2}{p} = \pi b^2 \delta \quad , \tag{63}$$

Therefore, the area provided for each microphone is the same regardless of the microphone order p, which represents a uniform microphone density over the design space. A corresponding array is shown in Figure 4.14(b).

In addition, if 0 < k < 1 in Eq. (60), it reprents increasing space spirals which show ever higher microphone density at the center. When k < 0, it also represents increasing space spirals, but the end of curve will not converge as shown in Figure 4.13(b). This kind of arrays can be applied for the design cases with irregular room or design space.

(a)   (b)
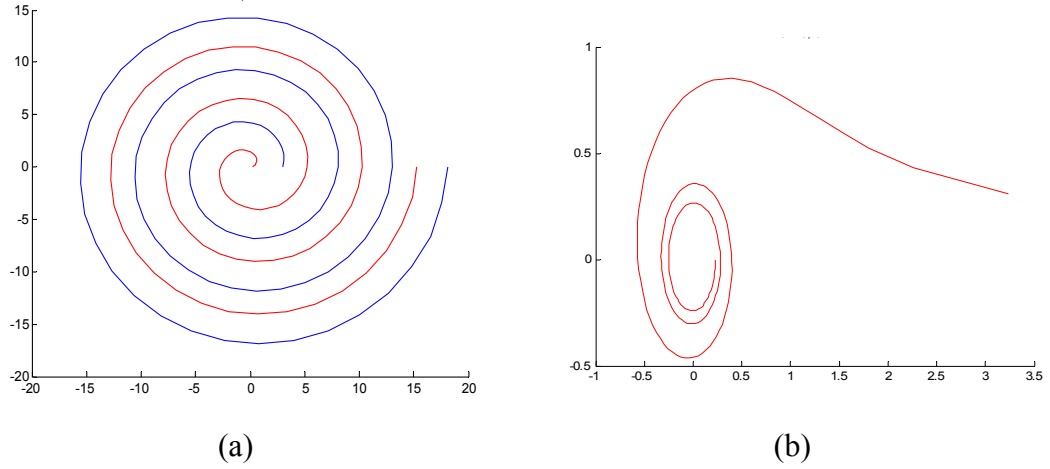
Figure 4.13: Spirals. (a) Archimedean spiral with b = 0.8. The red arc is with a = 0.2, and the blue arc is with a = 3. (b) Increasing space spiral with k = -2, while the end of curve cannot converge (a = 0, b = 1).

(a)   (b)
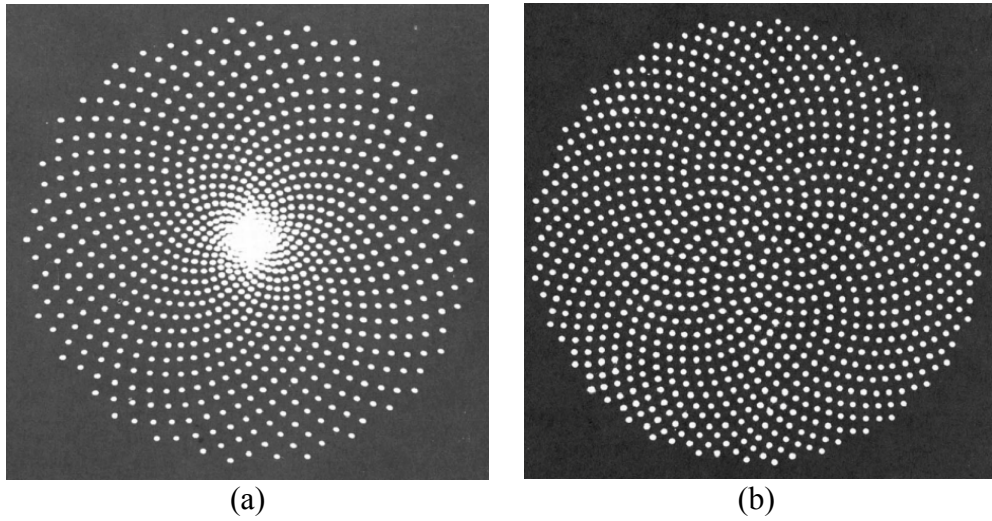
Figure 4.14: Arrays derived from spiral arcs (adapted from [70]). White dots reprent microphones/seeds. (a) Archimedean spiral. (b) Fermat's spiral.

89

The second class of spiral arcs is the logarithmic spiral (equiangular spiral or growth spiral) which is a special kind of spiral curve often appearing in nature. It can be expressed as:

$$r_p = a e^{b \varphi_p} \, ,$$ (64)

where b controls the shape of spiral and a controls the size of spiral. Normally, it is a increasing space spiral and approaches a Archimedean spiral $r_p \approx a(1 + b \varphi_p)$ when b is small. The area for each microphone can be computed as

$$\frac{S}{p} = \pi a e^{2b\delta} \cdot \frac{e^p}{p} \, ,$$ (65)

Therefore, the microphone density of this kind of arrays is decreased from the center to the periphery of array. It shows an even higher decreasing rate of microphone density when comparing with corresponding Archimedean spiral array, as given in Figure 4.15.
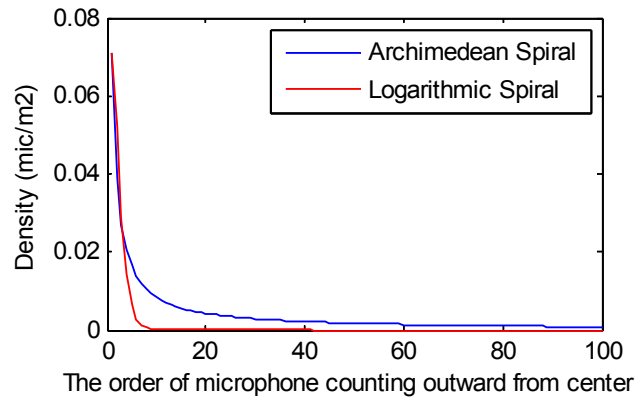
Figure 4.15: Microphone density comparison.

Previous discussions introduced the formulations of common used spirals applied for the microphone array design. The important parameters controlling the geometric shapes of spiral related to the microphone density are analyzed, and further affect the performance of beamformer. Generally, there is a trade off between MLW and MPSR. Arrays with microphones concentrating near the periphery areas show smaller ML area

and higher sidelobe levels, while arrays with high microphone density near center will have broader ML and lower sidelobe levels. Take the logarithmic spiral arrays for example. As shown in Figure 4.16, the array with more microphones in the center provides a narrower mainlobe, while broader area with relative high sidelobes are observed inside the FOV. Therefore, the proper choice of microphone density distribution controlled by the formulation of spiral arcs and the values of corresponding parameters should be made according to the knowledge of acoustic scenes and source behaviors in the real case applications.

The simulated beamforming result of a spiral array is given in Figure 4.17. Its performance is compared to corresponding regular planar array with identical microphone density and design space. It can be seen that, by eliminating the translational periodicity of element placements according to the focal point, the spiral array suppresses the highest sidelobe ridges in the symmetrical directions of regular planar array. It performs better than the traditional regular array in noise suppression, while maintaining the good directivity of regular array by providing a uniform coverage over the FOV. Although the optimized irregular arrays based on the descriptors proposed in Chapter 2 can perform better than the spiral arrays, they are difficult for installation [28, 69]. Therefore, as a mutated regular array, the spiral array combines the enhanced beamforming performance with easy operation which is feasible for the speech applications in immersive environments.

In conclusion, this chapter presented array geometries derived from the regular structures. By adding mutation to the traditional regular microphone arrangements, proposed geometries overcame the limitations of regular structures and provided superior SNR performance for speech captures in immersive environments. In addition, because these mutated arrays still retain certain regularity of element arrangements, they can be easily installed and operated. Simple geometry parameters were proposed for each mutated array to identify exact microphone positions. By controlling these parameters, good microphone density distribution providing superior noise suppression ability in specified scene can be easily generated according to the prior knowledge of possible source distribution.
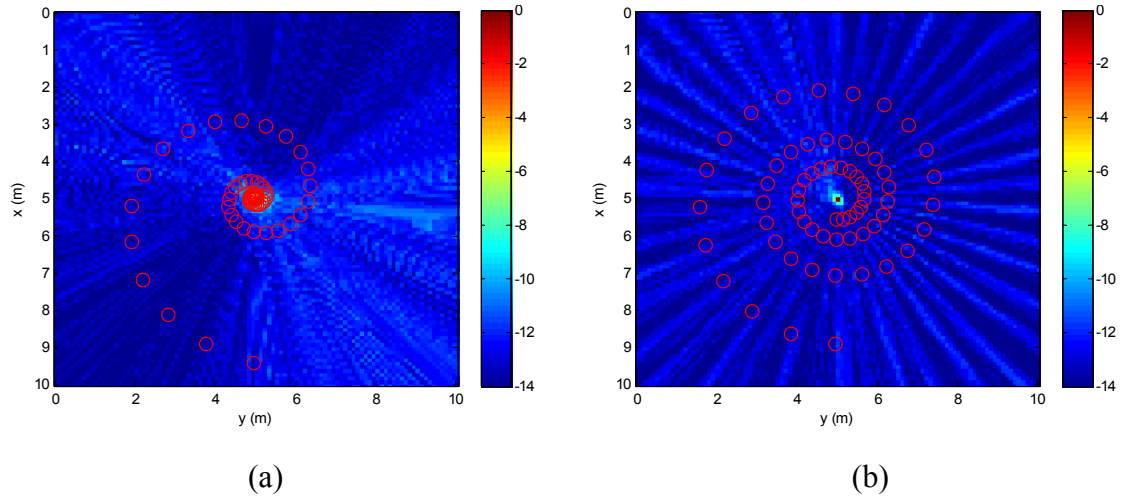
Figure 4.16: Top view gain pattern comparison for arrays derived from logarithmic spirals. (a) The Array with higher microphone density in the center. (b) The Array with lower microphone density in the center.
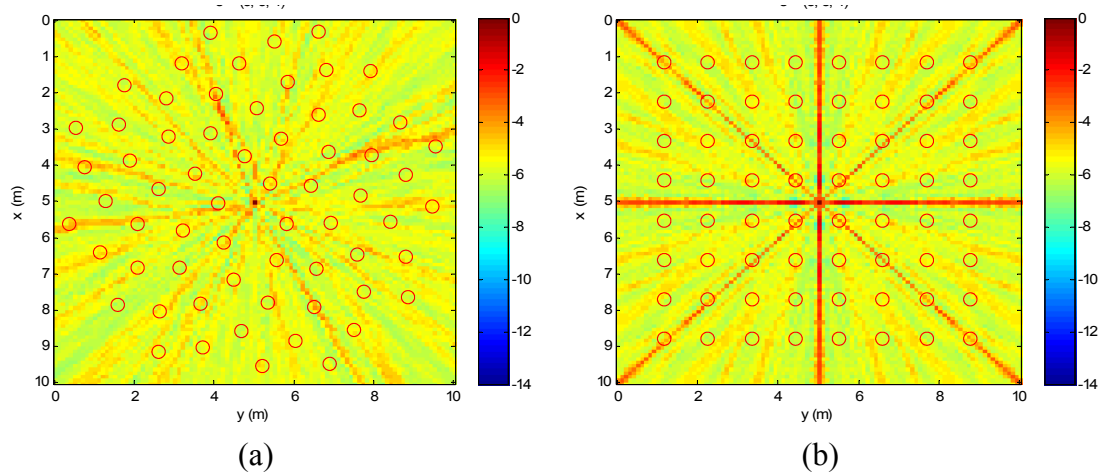
Figure 4.17: Top view gain pattern comparison when steering at the center of FOV. Red circles represent microphones. (a) Spiral array. (b) Regular planar array.

Chapter 5   Irregular Microphone Array Design

The limitations of traditional regular equispaced arrays for broadband speech signals are discussed in previous chapter. Several array geometries which mutated from the regular configurations to break the periodicity of microphone arrangements according to the FOV are proposed, showing enhanced beamforming performance. However, these arrays still retain certain regularity of microphone placements, which is demonstrated to restrict the uniformity of DPD distribution and result in severe sidelobe leakage on the interferences. Some irregular arrays with near random or completely random distribution of microphone are expected to achieve better performance for the broadband speech signals.

Due to the randomness of microphone placements of irregular array, it is difficult to control or predict the array performance. The traditional optimization methods of irregular array design are numerically very demanding, resulting from the large degree of freedom of element spacing [69]. In our study, as discussed in Chapter 2, statistic measures are applied to identify the array geometry properties related to performance, instead of unit distance and aperture applied for regular array analysis. By controlling these important geometry features of array configurations related to performance, irregular arrays with superior performance evaluated by the preselected objective function can be sorted out. By following the principles introduced in Chapter 3, several optimization methods for the irregular microphone array design are proposed in this

chapter.

5.1 Spatial Perturbation based on Regular Configurations

From previous discussion, it can be seen that the regular structured arrays have predictable gain patterns but suffering from the spatial aliasing problem for broadband signals. This problem results from the periodic Direction of Arrival (DOA) of sound wave from the source to microphones according to the inter-mic spacing. Therefore, nonuniform irregular arrays breaking the periodic nature of DOA are considered to be more effective to suppress the coherent level of received signals and further reduce sidelobe levels.

The basic ideal of spatial perturbation method for irregular microphone array design is adding perturbation to each element position of uniformly spaced regular array to suppress sidelobe levels. By applying the first two terms of Taylor series expansion to approximate the complex and nonlinear pattern function, the residuals between desired array gain pattern and an initial regular pattern are expressed as the linear combinations of spatial perturbations from regular positions. Then the optimal perturbations to form the irregular geometry can be calculated by a linear iteration procedure [53-58, 72, 73].

Consider the ideal situation for traditional far-field array design, where this optimization method is originally developed, the radiation field pattern of an even equispaced linear array with uniform excitation can be described as:

$$E_r(\theta) = \sum_{p=1}^{P} A_p \exp(j x_p u) = \frac{2}{P} \sum_{p=odd}^{P-1} \cos(\frac{p}{2} u) \quad, \tag{66}$$

where $u = \beta d (\sin\theta - \sin\theta_0)$. $\beta = 2\pi/\lambda$. θ_0 is the array steering angle. d is the unit inter spacing of elements. By adding spatial perturbation vector e_p to the pth microphone position of regular array, the new element position can be expressed as:

$$x_p = (p + e_p)d \quad, \tag{67}$$

Then the gain pattern is rewritten as:

$$E(\theta) = \frac{2}{P} \sum_{p=odd}^{P-1} \cos\left(\left(\frac{p}{2} + e_p\right)u\right)$$

$$E(\theta) = E_r(\theta) + \frac{2}{P} \sum_{p=odd}^{P-1} \left[\cos\left(\frac{p}{2}u\right) \cdot \cos(e_p u) - \sin\left(\frac{p}{2}u\right) \cdot \sin(e_p u) - \cos\left(\frac{p}{2}u\right)\right]$$

$$, (68)$$

If all the perturbations are very small, after picking up the first item of Taylor series, the approximation can be made as $\cos(e_p u) = 1 ; \sin(e_p u) = e_p u$. Then the residuals between the pattern of perturbed microphones and the reference pattern can be written as:

$$E - E_r = \frac{-2u}{P} \sum_{p=odd}^{P-1} e_p \sin\left(\frac{p}{2}u\right) , \qquad (69)$$

which is a linear combination of e_p.

Analytically, e_p can be computed by the formula of Fourier coefficients

$e_p = \frac{P}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{u}(E_r - E)\sin\left(\frac{pu}{2}\right)du$, as Harrington proposed in [73], where the approximation is restricted to small perturbation and minimizes the mean square error of pattern. Hodjat [55] and Kogan [56] directly chose the residual between the desired pattern and the original pattern of equispaced regular array as the objective function, then defined a multiplier $f(u)$ to designate a arbitrary percentage of residual reduction for each iteration, as $E - E_r = -E_r f(u) = R$. After defining the important spatial points $u_1, u_2, .. u_N$ for gain pattern control (eg. possible noise and target positions) and corresponding sidelobe reduction levels $R_1, R_2, .. R_N$, N/2 linear equations of e_p are listed. (Usually let $N=P$.) Due to the symmetry of array, these equations can be solved to get the values of perturbation (e_p , p = 1, 2, ..., P) as the results of the first iteration. Then the new gain pattern can be calculated by Eq. (68). In the next iteration, this new pattern are considered as the reference pattern E_r , and the computation can be continued in the same manner until the desired pattern is achieved.

In addition, other optimization criteria, such as minmax pattern distance [72] (as

discussed in 3.2.1) and minimum square difference $C = \dfrac{1}{2}\sum_{n=1}^{N}[E(u_n) - E_r(u_n)]^2$ [53, 57, 58], can also be applied as the objective functions for the linear programming routine (eg. Gauss-Newton algorithm) to compute the optimal perturbations to generate desired gain pattern. The constrains of these objective functions can be transferred by the subspace theory and Lagrange multipliers as mentioned in Section 3.2.5.

In conclusion, the spatial perturbation methods apply linear approximation to transfer the nonlinear nonuniform spacing problem of irregular array pattern to a linear combination of microphone perturbations based on the regular geometry. Then apply traditional linear programming algorithms to compute the proper values of perturbation providing desired gain pattern. Simulation results show that these approaches can reduce the first sidelobe level significantly, while only slightly broaden the MLW.

However, it can be seen that the traditional spatial perturbation methods are derived with two limitations: the far-field propagation mode and the initial regular array geometry. Since the typical scene for the speech capture of microphone array is near-field and immersive environment, far-field approximation might bring serious pattern errors. In addition, the approximation of the nonlinear pattern has the limitation that the element spatial perturbations should be very small, which means these optimization methods can only slightly adjust the configuration of the initial regular array. The choice of initial array is very important for resulting in global optimal positions, instead of local optimum. In previous optimization procedures, only regular equispaced arrays are applied as the initial arrays for the pattern approximation to simplify the problem. Further diversification of the microphone arrangements is restricted. Therefore, a new problem function for completely arbitrary microphone placements in immersive environment is needed.

By applying DSB, the general gain pattern expression based on the spherical wave propagation is derived as:

$$
\begin{aligned}
E(r,\theta,\phi,\omega) &= \sum_{p=1}^{P} \frac{B_p}{d_p} \exp\left[-j\omega(\tau_p - \tau'_p)\right] \\
&= \sum_{p=1}^{P} \frac{B_p}{d_p} \exp\left\{-j\beta[d_p(r,\theta,\phi) - d'_p(r_0,\theta_0,\phi_0)]\right\} ,
\end{aligned}
\tag{70}
$$

where $\beta = 2\pi/\lambda$, θ and ϕ are elevation and azimuth angles. $\boldsymbol{r}_p = (r_p, \theta_p, \phi_p)$ is the position vector of the *pth* microphone. $\boldsymbol{r} = (r, \theta, \phi)$ is the spatial position vector representing assumed source position. $\boldsymbol{r}_i = (r_i, \theta_i, \phi_i)$ is the focal point of beamformer (maximum radiation position). d_p is the distance between *pth* microphone and the source, which equal to

$$\|\boldsymbol{r}_p - \boldsymbol{r}\| = \sqrt{r_p^2 + r^2 - 2 r_p r [\cos\theta_p \cos\theta + \sin\theta_p \sin\theta \cos(\phi_p - \phi)]} \quad , \qquad (71)$$

d'_p is the distance between *pth* microphone and focal point $\boldsymbol{r}_i$, which equal to

$$\|\boldsymbol{r}_p - \boldsymbol{r}_i\| = \sqrt{r_p^2 + r_i^2 - 2 r_p r [\cos\theta_p \cos\theta_i + \sin\theta_p \sin\theta_i \cos(\phi_p - \phi_i)]} \quad , \qquad (72)$$

With fixed focal point $\boldsymbol{r}_i = (r_i, \theta_i, \phi_i)$ and signal frequency ω, any spatial point of beampattern can be expressed as the function of microphone position vectors $\boldsymbol{r}_p = (r_p, \theta_p, \phi_p), p = 1,2,\ldots, P$. Since the objective functions of array optimization are usually expressed as the functions of beampattern, they can be transferred to the functions of elements positions, such as pattern distances proposed in section 3.2.1. Therefore, the optimization procedure can be described as varying the design vector $(r_p, \theta_p, \phi_p), p = 1 \ldots P$ in possible design space to search for the extreme point of the objective functions. For regular linear array along x axes with equal inter-mic space d,

$$\boldsymbol{r}_p = (r_p, \theta_p, \phi_p) = (pd, \frac{\pi}{2}, 0) \quad , \quad \text{and} \quad d_p = pd \sin\theta \cos\phi \quad \text{in far-field mode. The}$$

pattern function turns into $E = \sum_{p=1}^{P} B_p \exp\{-j\beta p d (\sin\theta \cos\phi - \sin\theta_i \cos\phi_i)\}$, which is consistent with the formulation used in traditional spatial perturbation approaches. For irregular array, the design vector can be assigned in any possible position sets without the limitation of classical regular array geometry.

However, by abandoning the simplification mode of far field and regular array geometry expression which is not very suitable for microphone array applications, the objective function become very complicated when applying analytical optimization approaches. Therefore, instead of far-field mode, multi-variable Taylor series expansion is applied to simplify Eq. (71) and Eq. (72) to the polynomials of design vectors. For

example, for a specified spatial point r (interested spatial sample of beampattern), the second order Taylor series approximation of Eq. (71) about origin $r* = (0,0,0)$ is given by:

$$f(r_p) = \|r_p - r\| \approx f(r*) + d\, f(r*) + \frac{1}{2!} d^2 f(r*)$$
$$\approx r - r_p \cos(\theta) - r_p \theta_p \sin(\theta) \cos(\phi) + r_p^2 \frac{\sin^2 \theta}{2r}$$
$$\approx a * r_p^2 + b * r_p \theta_p + c * r_p + d \qquad , \qquad (73)$$

where a, b, c, d are the coefficients of polynomial, computed from $r = (r, \theta, \phi)$. In Cartesian coordinate system, this Taylor approximation is given by:

$$f(r_p) = \|r_p - r\| \approx f(r*) + d\, f(r*) + \frac{1}{2!} d^2 f(r*)$$
$$\approx \sqrt{x^2 + y^2 + z^2} - \frac{x}{\sqrt{x^2 + y^2 + z^2}} x_p - \frac{y}{\sqrt{x^2 + y^2 + z^2}} y_p - \frac{z}{\sqrt{x^2 + y^2 + z^2}} z_p$$
$$\approx a * x_p + b * y_p + c * z_p + d \qquad , \quad (74)$$

where a, b, c, d are also expressed as the functions of spatial sample position $r = (x, y, z)$. Therefore, for a fixed spatial point, the coefficients a, b, c, d is consistent, and the differential distance can be approximated by a polynomial of r_p . Then the complex objective functions of optimization can be transferred to a simple mode of design vectors, which can be solved by traditional optimization algorithms as discussed earlier in this section. It is noted that the expansion point $r*$ should be near the variable r_p to reduce the approximation error. In real cases, $r*$ is usually set as the centroid of microphone positions, or other suitable spatial point associated with specified type of Taylor series expression and the shape of possible microphone arrangement region.

In order to compare the approximation errors between traditional far field mode and proposed approach, Figure 5.1 gives the plots of approximation errorbars for 8 spatial samples in the circle of $(r=4, 0 < \theta < \pi/2, \phi = \pm \pi/2)$ with the microphone design space in $(0 < x_p < 1.5, -1.5 < y_p < 1.5, 0 < z_p < 1.5)$. It can be seen that, because the microphone design space can be expressed in a relative small range of (x_p, y_p, z_p) but large range of (θ_p, ϕ_p) , the multi-variable Taylor expansion in Cartesian

coordinate system has the smallest approximation error, while the Taylor expansion in spherical coordinate system doesn't work well. In order to avoid this problem, in Figure 5.1(c) Taylor approximation for single variable r_p in spherical coordinate system is applied to minimize approximation error, while the second order expansion is actually the same with the traditional far field expression. Therefore, by choosing suitable expansion point and Taylor series, the pairwise distances in beampattern computation can be expressed by a polynomial of design variables r_p within a acceptable approximation error range, which makes it possible to apply traditional analytical optimization approaches to solve the objective function of microphone array optimization in an immersive environment.
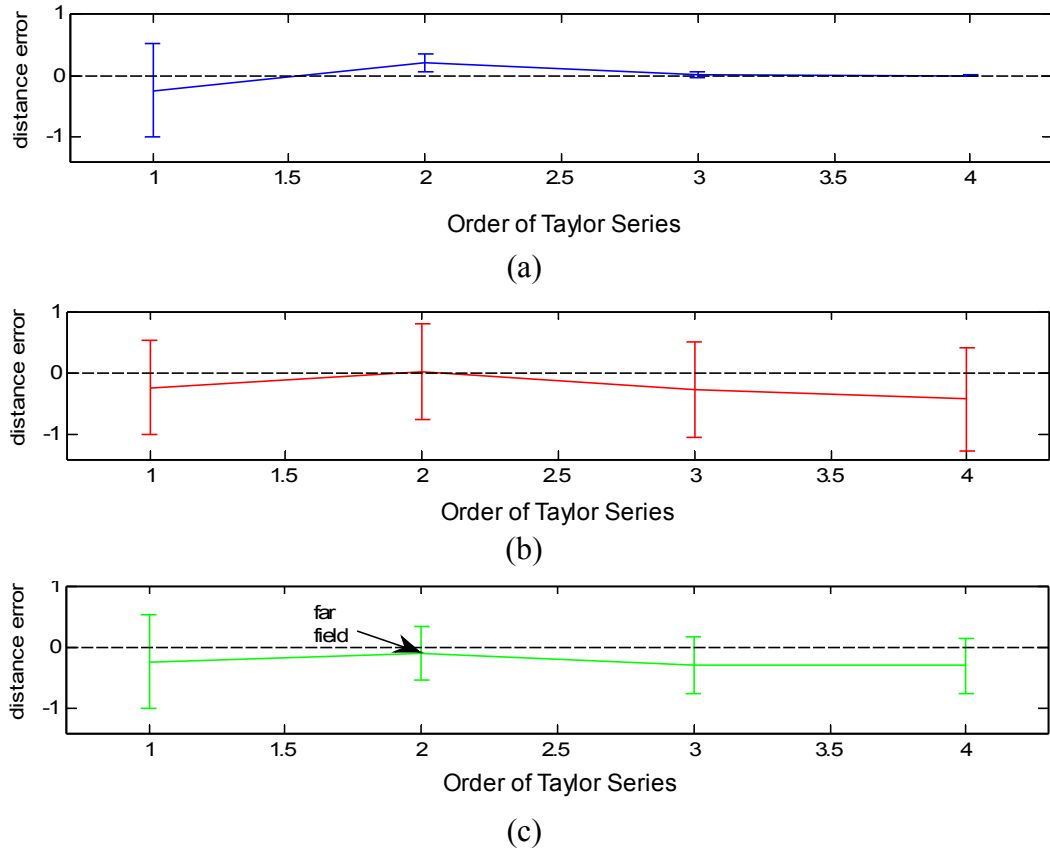

Figure 5.1: Errorbar plots of Taylor series approximation. (a) Errorbars of multi-variable Taylor expansion in Cartesian coordinate system. (b) Errorbars of multi-variable Taylor expansion in spherical coordinate system. (c) Errorbars of Taylor expansion for signal variable r_p in spherical coordinate system, where the second order expansion is identical with tradition far-field mode.

5.2 Heuristic Optimization Strategy

5.2.1 Introduction

The beamforming performance of microphone array is often assessed in terms of array spatial gain pattern, which is a complex and nonlinear function of the microphone geometry. This complex relationship limits the application of traditional optimization algorithms on irregular arrays. Previous optimization approaches apply linear approximation and spatial perturbations based on regular geometries to simplify the problem. The residuals between desired array gain pattern and an initial regular pattern are expressed as the linear combinations of spatial perturbations from regular positions to calculate the optimal geometry by linear iteration procedure [53-57, 72, 73]. Although this method can effectively sculpture gain pattern shape, it is limited to small perturbations to ensure prerequisite of linear approximations. Several new works [74, 75] introduce numerical approaches to minimize the residues between desired gain pattern and the actual pattern, which has to be computed point to point based on microphone positions. Random or exhaustive search methods can also be used in search of optimal geometry, which evaluate each candidate via Monte Carlo simulations. These methods are very flexible, but time consuming for large spaces and complex acoustic scenes [1]. This limits their feasibility for applications where rapid deployment is required, such as in the case of mobile platforms with changing acoustic scenes and surveillance applications. To address this limitation, this section proposes a genetic algorithm for the array optimization problem in conjunction with efficient objective functions and flexible acoustic scene descriptions, as discussed in Section 2.5 and Section 3.2.2.

Derived from the theory of natural selection, GA exploits the historical information of evolution procedure to guide the searching path. It predicts the new generation with expected better performance based on the probabilistic rules, and has been demonstrated as an effective tool in the area of nonlinear optimization problems [76-79]. This section introduces GA to the microphone array optimization problem with the purpose of obtaining superior interference reduction for speech applications. Instead of computing the beamforming performance of every candidate via Monte Carlo simulations, functions

based on microphone geometric descriptors showing strong correlations with array performance are applied as the objective functions (C_2 proposed in Section 3.2.2). Experiments in three acoustic scenes with different possible noise and target distributions are performed to validate the effectiveness and feasibility of proposed method. Three types of design surface for microphone distribution considered here are the 1D (linear design space), 2D (square surface), and 3D arrays (spherical surface). Results in terms of SNR are compared to comparable regular arrays, randomly generated irregular arrays, and optimal arrays obtained by a direct exhaustive search method to assess the performance of proposed method. The comparisons for the impacts of GA operators are also presented to improve the robustness of the optimization for different applications. Evaluation of reliability and results from real recordings are also provided to demonstrate feasibility.

5.2.2   Settings of GA

Chapter 2 analyzed the relationship between irregular microphone array geometries and spatial filtering performance with Monte Carlo simulations. Novel geometry descriptors were developed to capture the key properties of microphone distributions showing their impact on array performance. It has been demonstrated that in conjunction with array centroid offset and dispersion, statistics of DPD distribution can explain variations of performance metrics when steering at targets for immersive or near-field microphone applications. The optimization of our study seeks to find a microphone distribution within predefined design space that maximizes the beamformer SNR over a given distribution of target and noise sources. In order to avoid the expensive computations associated with a direct array gain computation over the space of interest, these geometric descriptors are applied to characterize irregular microphone distributions with similar performance. By applying the objective functions (C_2 proposed in Section 3.2.2) using performance-based geometric descriptions of microphone distribution that circumvent direct array gain computations over the space of interest, the optimization time of GA can be greatly reduced. In addition, probabilistic descriptions of acoustic scenes are introduced to incorporate various levels of prior knowledge for the source distribution.

After determining the objective function of optimization experiments, the settings of

101

GA procedure are introduced in this section. It is noted that the GA approach exploits historical information of the evolution procedure to predict new generation with better fitness (smaller value of the objective function). Following the rules of "survival of the fittest" [76-79], the genes of individuals (the microphone coordinates of arrays) with higher fitness will have more chance to be inherited by offspring, while perturbations are introduced randomly to the population to enhance diversity of the evolution. In our case, all coordinates of specified array are considered as an individual, and the fitness value is assessed by the objective function C_2. Parents are selected based on fitness values undergoing crossover and mutation to give birth to the new generation. The evolution procedure continues until reaching acceptable fitness value or the limitation of iteration number. The general flow chart is shown in Figure 5.2.
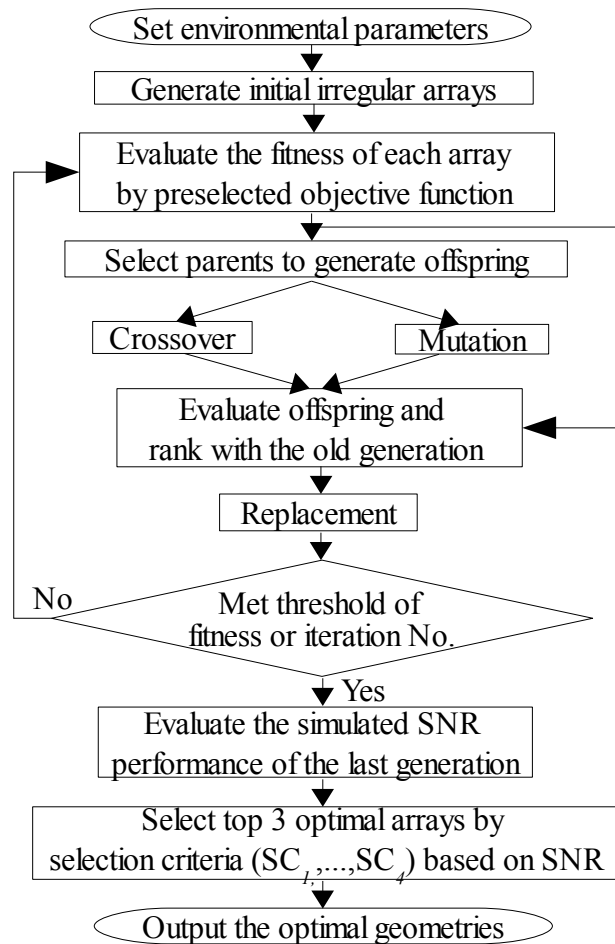

Figure 5.2: The flow chart of GA

Different with traditional optimization algorithms, which search from one single solution to another, GA is based on the group evolution of possible solutions, effectively reducing the possibility of being trapped in local optima [76]. Also the collective searching of GA, where the change of each individual affects multiple offspring, enhances the effectiveness of evolution, and all required information of solution is included in the coding of individuals to ensure the simplicity of the algorithm [76-79]. In order to control the optimization procedure properly for various microphone array problems, different choices of GA scheme are introduced and compared in three typical experiments to search for relative optimal microphone distributions. The important steps of GA applied in our study are discussed as below.

Coding of Individual: Different coding selection of GA can greatly affect the evolution path and convergence characteristics. Usually two types of coding scheme are applied, the multi-variable integer coding (IC) and the binary coding (BC) [76, 77]. In IC scheme, the coordinates of each microphone can be used directly as the individual. Or they can be discretized into the sequence number of locations by the minimum tolerable distance of adjacent microphones. In BC scheme the microphone locations are represented by a n bit binary code, while $2^n \geqslant M$ and M is the number of all possible microphone arrangements. The performance of each coding scheme for the microphone optimization problem will be assessed later in the experimental section.

Parent Selection: With the idea of "survival of the fittest", parent selection is trying to assign a selection probability function with the bias toward these high fitness individuals, who are more likely to generate good offspring. The probability to select the *ith* individual can be computed as [76]:

$$P(i) = \frac{f_i}{\sum_{i=1}^{N} f_i} \quad , \tag{75}$$

where f_i is the fitness value of *ith* individual and N is the population of each generation. In order to avoid high-fitness domination problem during evolution, this selection probability can also be assigned as a function of the fitness ranks, while different functions reflect different tendencies to select top ranked individuals independent with the fitness values [76, 77].

Crossover: In our experiments, to maintain a reasonable balance between inheritance and exploration of the offspring, 60% of parents are devoted into crossover process, which is implemented by randomly copying information from the corresponding parents' coordinates in each dimension. However, the impact of this process greatly depends on the choice of coding schemes. As show in Figure 5.3, if applying IC whose element exclusively represents one dimension of microphone coordinates, the offspring will be a random combination of parent coordinates. For BC, by exchange binary bits the coordinates of parents will not be retained in the offspring. Crossover under IC ensures the inheritance of good genes from parents, while crossover under BC adds diversity to the inheritance procedure but might lose focus on fittest. Therefore, experimental comparison is performed to explore the benefits of these coding schemes on microphone array problems.

Mutation: Other than crossover, the remaining 40% of parents undergo mutation, which adds random perturbations to each bit of the parents' codes as shown in Figure 5.3. In IC case, the perturbations directly represent spatial displacements to each dimension of parents' coordinates, and are generated from a normal distribution $N(0, \sigma^2(n))$ in our study, as [78]:

$$\sigma(n) = \sigma_0 \rho^{(n-1)} \quad , \tag{76}$$

where n is the iteration number, and σ_0 is the initial standard deviation related to the wavelengths of signal. $0 \leqslant \rho \leqslant 1$ is a constant to control the rate of shrinkage of perturbation along generations. The idea here is to add large perturbation at the beginning of iteration to increase searching diversity avoiding GA trapped in local optimum. Because large perturbations interrupt a search path from an optimal direction and reduce convergence speed, $\sigma(n)$ is reduced to speed up the convergence when the optimal searching direction becomes more specified along with the iteration. Mutation of BC is performed in a similar manner by flipping bits according to the probability of $P(n)$, where $P(n)$ plays the same role with $\sigma(n)$ to adaptively control the level of added diversity during iteration.
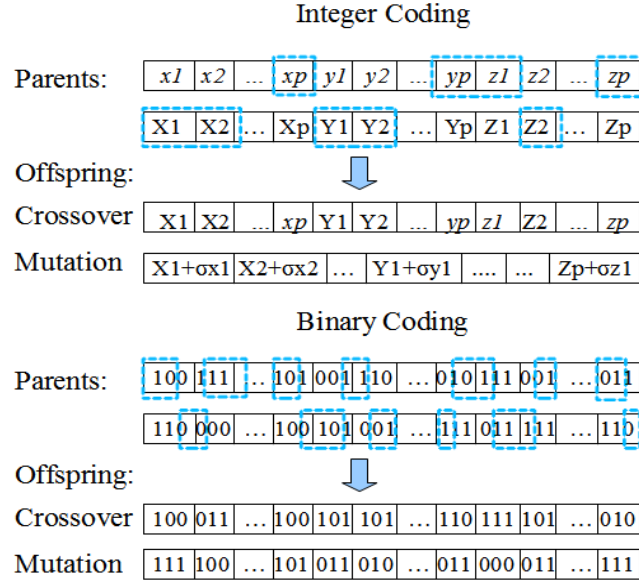
Figure 5.3: Crossover and mutation for arrays with P microphones. The black box represents each dimension of microphone coordinate. The blue dashed box represents the bit of parent's coding used for crossover and mutation. The mutations are based on the second parents.

Balance of Inheritance and Diversity: The art of a successful GA optimization is to maintain the balance between inheritance and exploration to suit different optimization problems. A tradeoff exists between inheritance pressure of good genes from parents and searching diversity of distributing offspring into the problem space. Inheritance pressure can be assessed by the bias to select high-fitness individuals as parents, and the ratios of crossover and mutation. Searching diversity is measured by the level of distractions from current searching path, and is related to the population size and standard deviation of perturbations of GA. High inheritance pressure increases concentration on high-fitness individuals while losing diversity of offspring and risking trapping in local optimal solution. However, too much searching diversity could blur the guided direction from the information of older generation, and slow down the convergence speed of evolution. Therefore, in order to balance inheritance and exploration, appropriate choice of GA parameters are necessary. Several methods to control the inheritance pressure and searching diversity are applied in this paper as below.

(1) Immortal elite

Although information of parents are exploited to guide the searching path, there is no guarantee that the high-fitness individuals will survive during the generation transition [76]. The potential best solution may be lost forever. Therefore, immortal elite method ranks the fitness of the offspring together with the old generation and sorts out the new generation with the same population to ensure the survival of elite.

(2) Adaptive selection probability function

During the parent selection, bias to select high-fitness individuals is controlled by the selection probability function based on fitness value or fitness rank of individuals. For specified optimization problem, this probability function needs to be adjusted to meet the requirement of inheritance and searching diversity. Adaptive selection probability function assigns a moderate bias at the beginning of GA to enhance diversity, and adaptively increases this tendency along iterations to speed up convergence when the optimal searching direction becomes more specified through evolution. Our experiments applies Rank Adaptive Probability Function (RAP) for the parent selection, and compares its performance with the traditional Constant Probability Functions based on Fitness Values (FCP) and Fitness Rank (RCP). In RAP, the probability to select ranked *ith* individual is defined as:

$$P(i) = \frac{1}{1 + (\frac{i}{i_E})^{\textbf{ceil}(\frac{n}{10})+1}} \quad , \tag{77}$$

where n is the number of iterations and i_E is the cutoff number of elite, which can be computed from the population of each generation and the predefined percentage of elite. As shown in Figure 5.4, the shape of probability function is adjusted every 10 generations, and converge to RCP when the iteration number becomes very large.

(3) Forced mutation

Because of the limitation derived from the physical conditions of acoustic environment, illegal individuals may be generated during iteration, which are outside defined microphone design space, or repeat with the other individuals. Forced mutation [76, 77] directly replaces these illegal individuals by random generated individuals to increase mutation rate and add more searching diversity to ensure focusing on global optimum.
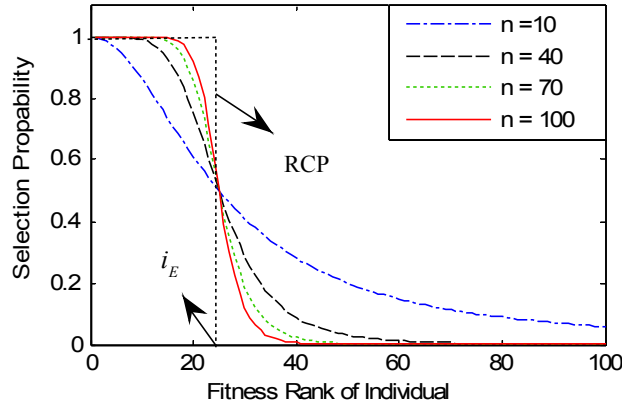
Figure 5.4: Rank adaptive probability (RAP) function for parent selection, where *n* is the iteration number.

Selection Criteria based on SNRs: After the iteration, the last generation of GA contains individuals with the achievable highest fitness assessed by proposed objective function, which is derived from the experiential relations between important geometric characteristics of microphone array and performance metrics. These individuals in the last generation representing specified microphone distributions are expected to show optimal beamforming performances based on the probabilistic rules. In order to demonstrate this prospect and select the optimal arrays for specified acoustic scene, Monte Carlo experiments (as discussed in section 3.4) are performed to simulate the SNR performance of each array in the last generation with various target and noise occurrences. In this dissertation, simulated sources consist of colored noise generated by the band important function of the SII model [34], which emphasizes the frequency bands most important to speech intelligibility. For each array in one specified acoustic scene, 30 second simulation is performed to compute the SNR while the target and noise sources are reselected every 0.5 second under predefined corresponding probability density function of source distribution.

The criterion to select top 3 optimal geometries from the last generation in terms of simulated 60 SNR results of each array depends on specified requirement of application, such as strong robustness, optimal average performance and high stability. Four selection

criteria are proposed to fit different applications, as shown below:

SC_1: Maximum average SNR over the 30s Monte Carlo experiment, to pick up the array geometry with best overall beamforming performance.

SC_2: Smallest standard deviation of SNRs along different target and noise occurrences in the 30s simulation, related to the stability of beamforming performance in different situations.

SC_3: Maximum minimum SNR or 10% SNR percentiles over the 30s simulation, to pick up the array with the optimal worst case performance.

SC_4: Limitations of the SNR percentiles (10%, 25%, 50% and 75%) over 30s simulation, combining the previous three criteria together to meet multiple requirements of array beamforming performance.

In the next section, SC_1 is applied in all the experiments to pick up the optimal array from the last generation of GA, where the beamforming performance of each array is assessed by the average SNR over the 30s simulation.

5.2.3   Experiments and Discussions

This section evaluates the effectiveness of optimizing irregular microphone distributions with the GA using the geometric-based objective functions. Optimization experiments over 3 design spaces are performed, which include the 1D linear array, 2D planar irregular array, and 3D spherical array. Results in terms of SNR are compared to regular arrays with the same number of microphone and dispersion, as well as the optimal array obtained through a direct exhaustive search (representing the upper limit for performance). In addition, comparisons for different GA parameters are also presented to assess their impact on optimization performance, and provide general guidelines for GA settings based on the application.

5.2.3.1   Experimental Settings

The space of interest for all the experiments was within a $10\times10\times2$ m³ room. The goal of the optimization procedure was to find the distribution with superior noise suppression in given acoustic scene (source and interferer distribution and activity). Each source and interferer consisted of colored noise generated from the band importance function of the

SII to simulate the frequency range of human speech. The details of each optimization problem are given in Table 5.1. The basic settings related to GA inheritance and diversity are given in Table 5.2. In order to test the performance of the resulting arrays, Monte Carlo simulations are applied over each acoustic scene to estimate the SNR for the proposed geometries, as discussed in previous section. A 30 second recording is simulated for each scene, where a target/noise occurrence is randomly changed every 0.5 seconds under predefined probability density function of the source distributions introduced in Eq. (44). To assess the effectiveness of the proposed objective functions C_2, the average SNR over all the geometries in the last generation are compared to randomly generated irregular geometries of the first generation, in addition to the regular and optimal geometries picked by exhaustive search method. The convergence speed, reliability, and computational complexity of this optimization method are also examined.

Table 5.1: Optimization scenarios.

	Design Spaces	Acoustic Scenes
Case 1:	Linear over middle beam of ceiling	3 uniformly generated source positions in the vertical plane of middle beam. One is arbitrarily selected as target, while others act as interference. $$p(r_i)=\frac{1}{3}, i=1,\cdots,3;$$ $$p(r_s|r_i)=1,\ s=1,\cdots,3,\ s\neq i$$ (Target and noise space are completely overlapped.)
Case 2:	Planar over ceiling	10 uniformly generated source positions. The target is arbitrarily selected from 4 of them. Except for the target position, the other 9 are considered as interferences. Each has 4/5 chance to make sound. $$p(r_i)=\frac{1}{4}, i=1,\cdots,4;$$ $$p(r_s|r_i)=\frac{4}{5}, s=1,\cdots,10, s\neq i$$ (Target and noise space are partly overlapped.)
Case 3:	Spherical cap ceiling (radius of the sphere= 36m; center = [5,5,37.6]; height of the cap = 0.7013m)	5 uniformly generated source positions. The target is arbitrarily selected from 2 of them (predefined interested positions), while the other 3 acting as interferences. Each one has a 3/4 chance to be active. $$p(r_i)=\frac{1}{2}, i=1,2;$$ $$p(r_s|r_i)=\frac{3}{4}, s=3,\cdots,5$$ (Separated target and noise space.)

Note: The minimum inter-mic distance for all geometries was limited to 0.1m.

Table 5.2: Fixed GA settings

	Pop.	Elite Percentage	Ratio of Crossover	Ratio of Mutation	Iteration Threshold
Case 1:	80	25%	60%	40%	Iter. No. = 100
Case 2:	100	25%	60%	40%	Iter. No. = 120
Case 3:	100	25%	60%	40%	Iter. No. = 120

5.2.3.2 Results in terms of SNR

Table 5.3 shows the SNR results of each optimization scenario. The mean SNRs with ± one standard deviation over the arrays in the first and last generation of GA are presented. It can be seen that the randomly selected irregular arrays in the first generation contain both superior and inferior arrays relative to the regular array. By comparing the first and last generation of the GA large SNR improvements are observed as a result of optimization under the geometry-based objective function rules, no matter which coding scheme is applied. Also for all the three problems, the SNR variance along arrays in the last generation are reduced, indicating that the optimization procedure sorts out the good irregular arrays with increased performance consistency. The optimal array obtained by an exhaustive search method provides a benchmark to evaluate the SNR improvement resulting from GA. By comparing the SNR of the top performing arrays in the last generation (estimated by the 3rd quartile SNR representing performance of upper half generation arrays) with the optimal array from exhaustive searching, about 60% of the maximum possible SNR improvement is achieved by GA for each case. Take case 1 GA-BC for example, the 3rd quartile SNR of the last generation arrays = 22.88 dB,

$$\frac{22.88-17.51}{25.87-17.51}=64.23\%$$.

Table 5.3: SNR comparison (dB)

	Last Generation of GA-IC	Last Generation of GA-BC	First Generation of GA	Regular Planar	Exhaustive Search
Case 1:	20.62 (±2.10)	21.27 (±2.06)	17.51 (±2.90)	18.32	25.87
Case 2:	6.93 (±0.75)	6.57 (±0.75)	3.02 (±1.59)	2.20	10.24
Case 3:	12.91 (±1.82)	11.44 (±1.79)	8.78 (±2.52)	9.32	17.09

Note: the results are for arrays with 16 microphones.

Figure 5.5 provides the SNR improvements vs. microphone densities for the three optimization scenarios with different coding scheme. Figures 5.5(a)(c)(e) show the SNR results when applying GA-IC. Generally, the SNR improvements brought about by the GA optimization reduce with increasing microphone density. Microphone placements are more critical for low densities. If the microphone density is very high, it is not necessary to perform the optimization procedure, since an arbitrary array can provide a good beamforming SNR. While the critical density values are dependent on the design spaces and acoustic scenes, the numbers from Figure 5.5 provide a general idea of when optimizing the microphone placement is important. Each graph plots the SNR gain between the last generation and the first as well as the last generation and the regular array. The critical microphone density for the linear array is 2.5 mic/m, for the planar array is 6 mic/m^2, and for the spherical array in problem 3 is 4 mic/m^2.

Figures 5.5(b)(d)(f) show the SNR results when applying BC for generational changes. It can be seen that, because BC doesn't copy the coordinates of old generation directly, it adds more perturbations to the evolution, while losing some inheritable information. Its SNR improvements degrade severely with the increase of microphone density. However, due to the same reason, it also provides more chance to catch optimal points outside the original searching path, obtaining unexpected SNR improvements in certain cases, such as the 1.5 mic/m case in Figure 5.5(b) and 2.5 mic/m^2 case in Figure 5.5(f). These results demonstrate that proposed optimization method provides important SNR improvement

when the FOV needs to be covered by small number of microphones in immersive scenarios. If microphone density is in the comfortable range (smaller than the critical density value), IC is the proper and effective choice. If the diversity of microphone coverage is restricted (high microphone density) or the population size is limited, BC is a good choice for finding singularities with superior beamforming performance. Note for each case, according to the type of array and microphone number, proper geometric relationship function derived from the simulation data of corresponding array type is selected to form the objective function [28, 80], which makes the SNR gain trend varied for each plot.
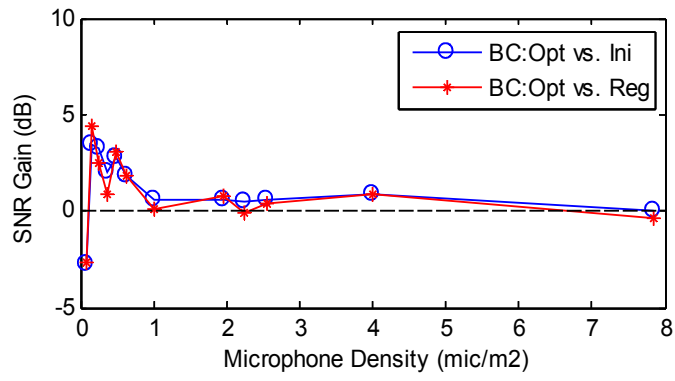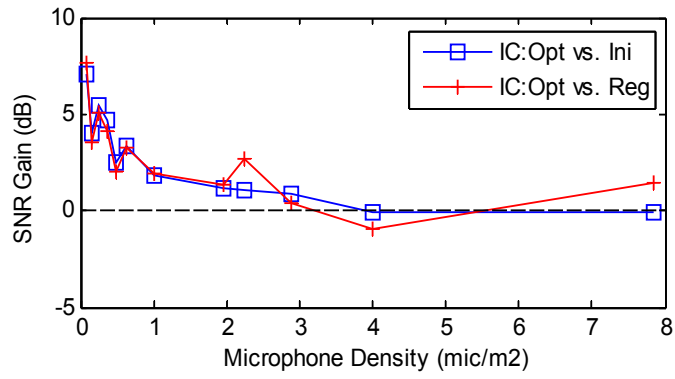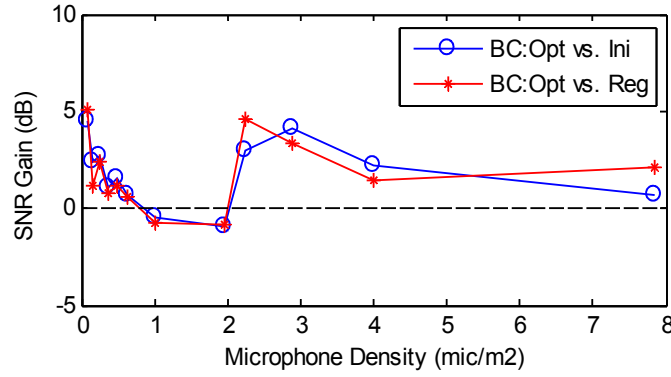

(a)

(b)

(c)

(d)

(e)

114
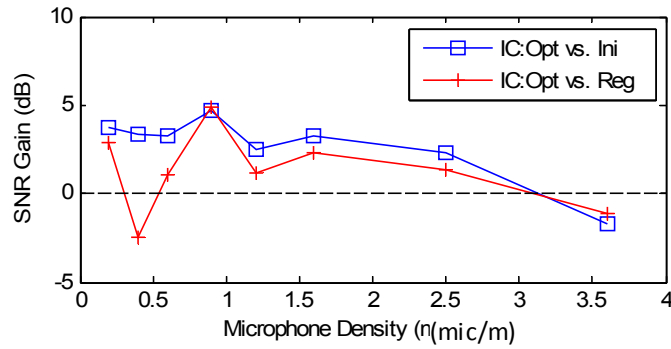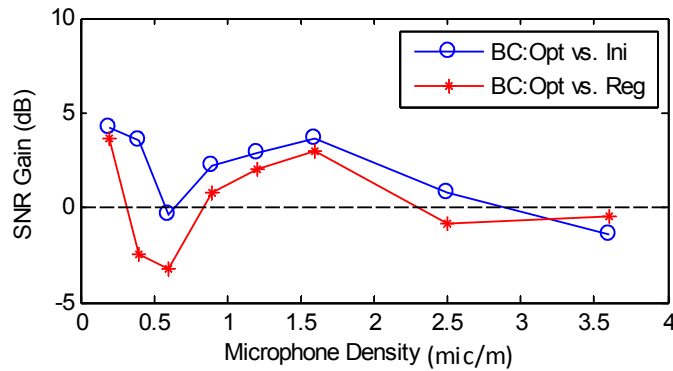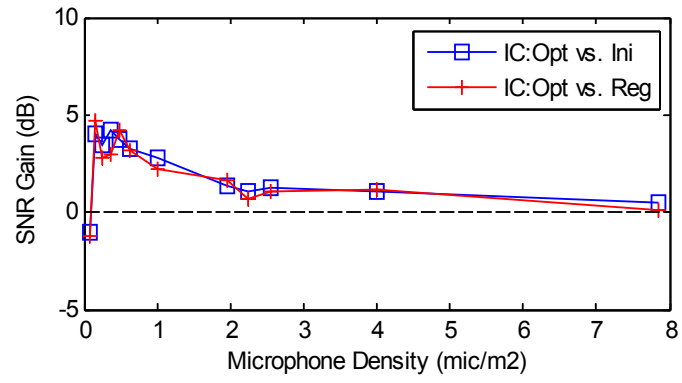
(f)

Figure 5.5: SNR gains through GA optimization vs. microphone densities. (a) Case 1 with IC. (b) Case 1 with BC. (c) Case 2 with IC. (d) Case 2 with BC. (e) Case 3 with IC. (f) Case 3 with BC.

5.2.3.3 Observation of GA Evolution

As mentioned before, the balance between inheritance and exploration is crucial for the success of GA. The size of population, the coding scheme, and the parent selection probability function are the important factors impacting the searching diversity and convergence.

Figure 5.6 shows the GA convergence with 16 microphones for the optimization problem 3. Figure 5.6(a)(b) give the results using IC-RCP and IC-FCP schemes. It can be noted that, insufficient size of population (eg. 30 in Figure 5.6(a)) cannot be compensated for by increasing iterations. It traps the evolution in a local optimum. On the other hand, excessive population sizes only result in limited improvements, while greatly increasing the computational complexity (eg. 160 in Figure 5.6(a)). Therefore, the traditional GA using RCP is very sensitive to the population size. Choosing a moderate size of population is critical for the success of the optimization. For the FCP parent selection function, it can be seen that although FCP based on the fitness value of each individual shows slower convergence speed than RCP, it has more robustness with different choices of population size.

Figure 5.6(c) shows the convergence of IC-RAP, as defined in Eq. (77). It can be seen

that the new adaptive selection function results in a fast convergence speed as RCP, and meanwhile shows enhanced robustness along different population sizes. Therefore, in conclusion, RAP is the better choice for the parent selection in microphone array optimization problem. With sufficient population size, RCP has the fastest convergence, while FCP can be applied to bring more iteration diversity to compensate for limited size of array population.

Figure 5.6(d) shows the convergence of the BC scheme. When compared with IC (Figure 5.6(a)), the size of population has less impact on the BC convergence, while more iterations are needed to reach the optimum. In addition, from Figure 5.6(e), the convergence of BC is very sensitive to the microphone density. In the optimization of problem 3, with the same size of population and iteration number, the converged value of BC degrades severely with the increase of microphone density, which explains the relative small SNR improvement in Figure 5.5(f).

Table 5.4 summarizes the general comparisons of GA schemes for microphone array problem. In order to further demonstrate our conclusions, Figure 5.7 gives the average repeat rate of individuals of the first 20 iterations and 80 iterations with different GA schemes. By comparison, IC-RAP is the most efficient scheme providing the lowest repeating rate to rapidly update new geometries with better fitness in the early generations. BC-RCP and IC-FCP also perform effective evaluation by providing the lowest repeat rate of individuals along 80 generations. These results are consistent with the previous conclusions of GA schemes. The optimal GA settings combine the coding scheme and parent selection function with complementary features of robustness and convergence. Take IC-FCP for example, the slow convergence speed of FCP is offset by the fast speed of IC, while FCP brings more robustness with insufficient population size.

116
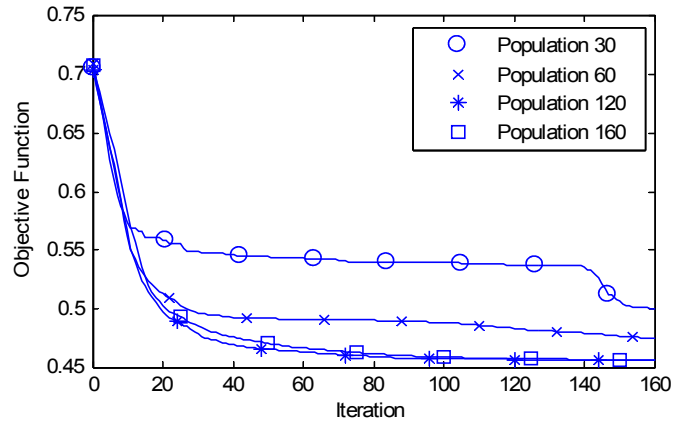
(a)
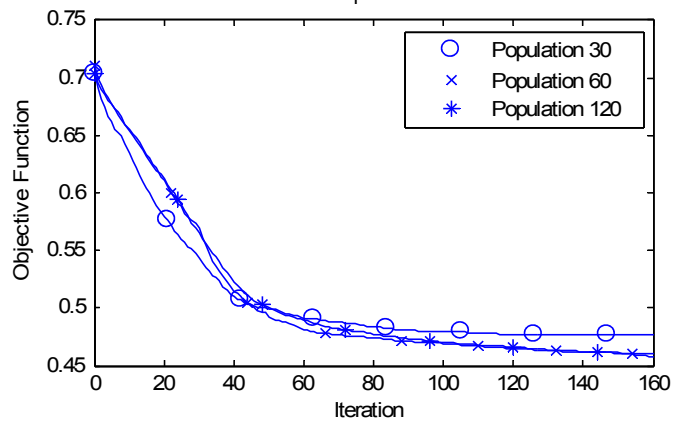
(b)
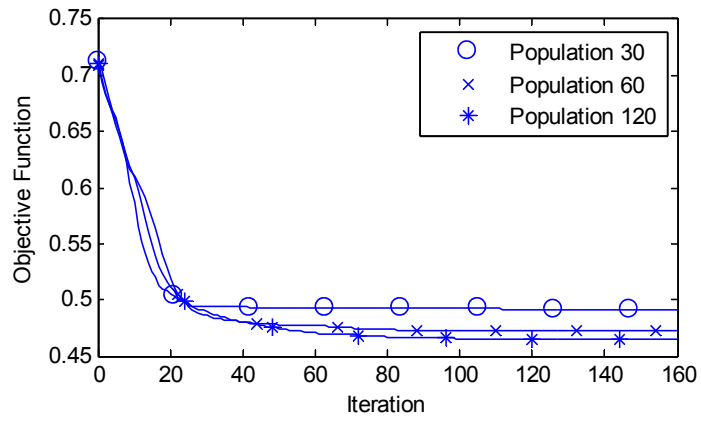
(c)

(d)
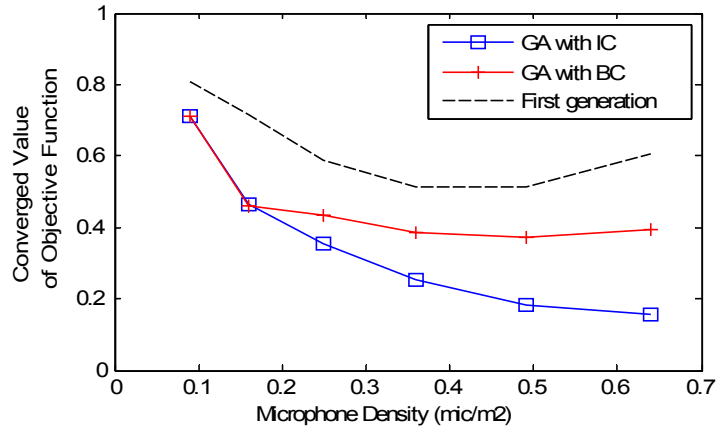
(e)

Figure 5.6: The convergence of GA for arrays with 16 microphones in problem 3. (a) IC-RCP. (b) IC-FCP. (c) IC-RAP. (d) BC-RCP. (e) Converged values with RCP vs. microphone density. The smaller the converged value of the objective function, the higher the fitness is.

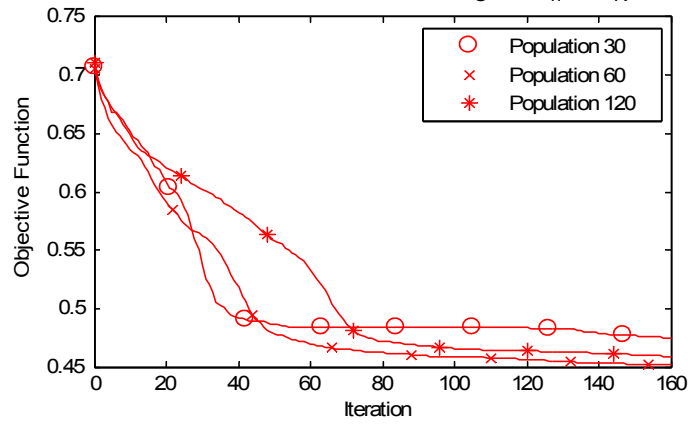Table 5.4: Comparisons of GA schemes

		Convergence Speed	Robustness along Pop. Size	Robustness along Mic Density	Ability to Discover Singularity
Coding	IC	fast	bad	good	bad
	BC	slow	good	bad	good
Parent Select	RCP	faster	bad	good	good
	FCP	slow	better	good	bad
	RAP	fast	good	good	good

Note: Two types of coding are compared to each other, providing relative good/bad results. Results of 3 parent selection functions are provided from the similar manner.
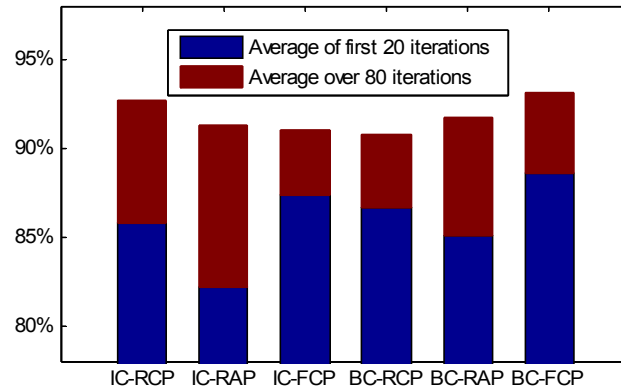
Figure 5.7: Repeat rate of individuals along iterations

5.2.3.4 Reliability

For all optimization problems, all the possible source positions were randomly chosen in FOV. In order to evaluate the reliability of this optimization algorithm for arbitrary source distribution in FOV, 30 experiments with 16 microphones are performed in the similar optimization scenes, where the source positions in each problem are randomly shifted in FOV. The results are provided in Table 5.5.

For the optimization results of planar arrays (case 2) and spherical arrays (case 3), the

large majority of GA optimized irregular arrays have better beamforming performance than regular arrays, as well as the randomly generated irregular arrays, representing high successful rate of the optimization algorithm over the similar acoustic scenes. In addition, IC shows higher reliability than BC in these cases. For the linear array problem (case 1), where the one dimensional design space lacks resolution for rotationally symmetrical positions, the optimization won't be successful if sources are placed in these blind positions. Also, with fixed number of microphone, the limited choices of possible microphone positions reduce the diversity of possible solutions, and restrict the potential improvement thorough the objective functions. In these cases, consistent with previous SNR analysis, BC shows stronger ability to find the optimal array with superior beamforming performance.

The average number of iterations needed to reach convergence are also given in Table 5.5. Similar with previous conclusion, IC has faster convergence speed than BC, while BC has larger chance to discover super good arrays for problem with limited diversity of solution (as in case 1). The cpu processing times of GA are compared to exhaustive random searching methods with the same number of assessed array geometries. Results show a reduction in time by about 3 orders of magnitude.

Table 5.5: Optimization reliability and processing speed

Similar Scene with:	Case 1		Case 2		Case 3	
Coding of GA:	IC	BC	IC	BC	IC	BC
Outperform Initial Generation	33%	57%	100%	93%	90%	80%
Outperform Regular Array	33%	20%	100%	100%	100%	97%
Convergence Speed (in iterations)	138	82	59	200	126	200
Computation Time	0.50% of the random search method					

Note: take case 3-BC for example, 97% means that in 29 experiments out of 30, the average SNR of the last generation arrays of GA outperforms corresponding regular array.

5.2.3.5 Experimental Recording Results

To further demonstrate the beamforming performance of GA optimized arrays, experimental recordings of planar ceiling arrays using 25 omnidirectional microphones are performed in a 3.58x3.58x2.29m space of interest bounded by aluminum struts as shown in Figure 1.4. The microphone density for both arrays is 1.95 mic/m^2. The microphone signals were amplified with RME Octamic-D preamplifiers, sampled with RME HDSP9652 sound card at 44.10 kHz, and downsampled to 8kHz for processing. Four loud speakers were placed in the FOV to play prerecorded human speeches (including two male voices and two female voices) as the acoustic sources. These voices were normalized for power and the duty ratios for each source are 0.62, 0.56, 0.61, 0.54, so each speaker had similar volume levels. The optimization was performed assuming the sources as SII colored noise signals. The acoustic distribution was limited to 4 source positions, where each one is considered as a target and the others as interferers over 4 possible combinations. The optimized array was expected to show the best target signal enhancement relative to the interferers summed over the 4 possible cases.

Three separate recordings of 28 seconds were made by each microphone geometry, where the optimal irregular array was selected by 50 iterations of GA-IC-RCP. The Steering Response Coherent Power (SRCP) images [40] averaged over all frames are shown in Figure 5.8. It can be seen that the optimized irregular array shows stronger responses in the source positions, which is primarily a result of optimizing this array to provide the best beamforming performance over all sources. In addition, the recovered waveforms from the output of beamformer when targeting at the right source are shown in Figure 5.9. By inspecting the waveforms and listening to the outputs of beamformer, it can be seen that in the time slots when the target source is active (marked by red points), the recovered signal of optimized array shows enhanced beamforming gain than the regular array.

In order to assess the SNR enhancement relative to each target focal point over the interferers, the inactivity periods for the target speaker were segmented out from the output of beamformer, and the average power values were computed to denote the noise leaking into beamformed signal from the other speakers. The average power values for the active segments were computed as well. Results of SNR evaluations are presented in

Table 5.6. Note that when targeting at the top and right source the optimized irregular array shows better SNR performance, while regular array provides better SNR when targeting at the bottom and left source. The optimized irregular array provides a 0.94 dB overall SNR enhancement over the regular array. This result is consistent with the optimization criteria since it aimed at the best average performance and did not require the enhancement be evenly distributed over all the sources. In this instance it configured most of the microphones between the top source and all others. This pattern typically happens when one target position is selected against the others. In addition to these numeric results, the listening comparisons between beamformed signals can also confirm that the GA optimized geometry has better overall noise suppression ability, most notable in the top and right target sources. This example, therefore, demonstrates the ability and behavior of the optimization approach on a real recording, as well as the ability of the geometry-based objective functions derived from the simulation data to improve performance through irregular microphone placements.
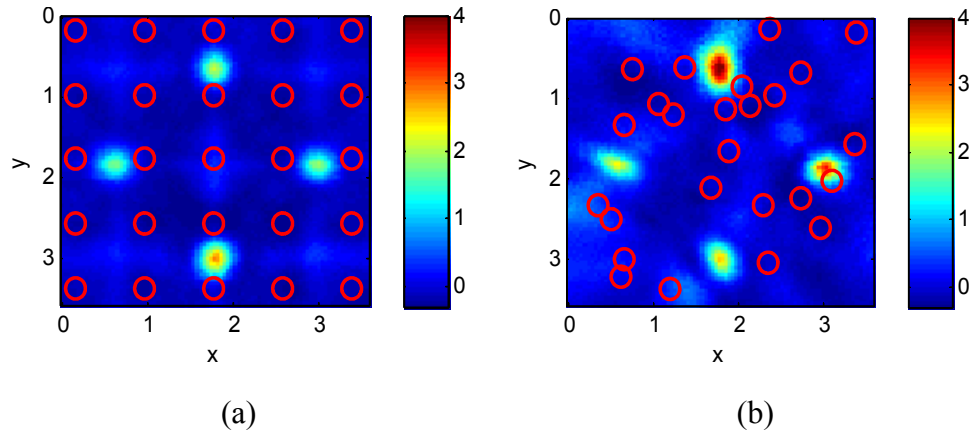

(a)   (b)

Figure 5.8: Top view power images. Red circles represent microphones. Four sources with the elevated power levels are located at the top, bottom, right and left in the FOV. (a) Regular array. (b) Optimized irregular array.
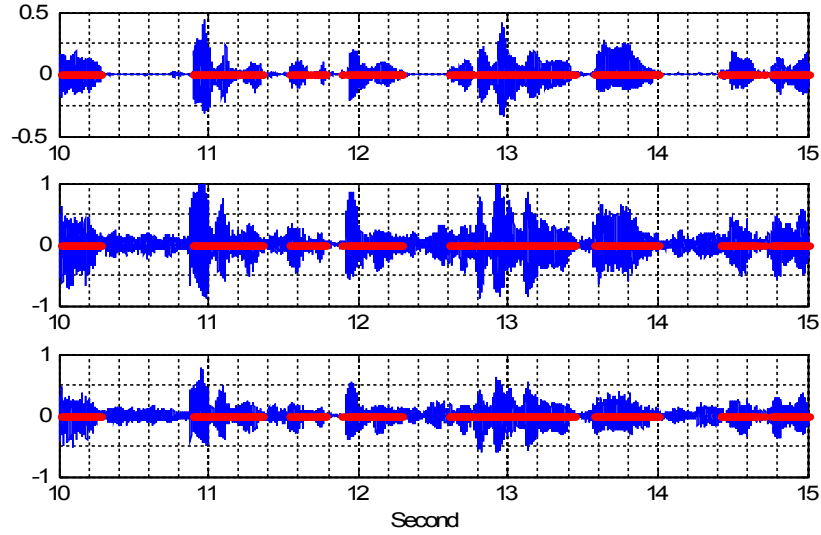
Figure 5.9: Waveforms of the original signal from the right source (row 1), beamforming output of optimized irregular array (row 2) and regular array (row 3) when targeting at the right source. The red dots represent the time slots when the target source is active (making sound).

Table 5.6: SNR comparison (dB)

Arrays	Right Target	Bottom Target	Left Target	Top Target
Regular	6.06	6.07	1.65	4.54
Optimized Irregular	6.64	4.06	1.11	7.45

5.2.4 Conclusion

This section applied geometry-based objective functions for optimizing the irregular microphone distributions with superior SNR performance for specified acoustic scene. Several GA schemes controlling the balance between inheritance and exploration of searching procedure were evaluated and compared in the real optimization problems of microphone array.

Simulation results demonstrate that proposed optimization method effectively sorts out

123

these superior irregular geometries with large SNR improvements when comparing with randomly generated irregular arrays and regular arrays. The proper setting of GA should choose the population size, coding scheme, parent selection function, array type, and other parameters with complementary features of convergence and robustness of iterations for different optimization problems. In addition, the acceptable processing time observed during the experiments and the validation of real recordings indicate strong correlation between proposed objective function rules and array beamforming SNR performance for human speech. Also, the feasibility of proposed method is established for array geometry design in immersive environments where rapid deployment is required with limited knowledge of the acoustic scene, such as in mobile platforms and audio surveillance applications.

5.3   Cluster Design

5.3.1   Hyperbola Cluster

By visually inspecting optimal random array geometries obtained from the heuristic optimization methods in the previous section, similar patterns of microphone cluster distribution are observed. Figure 2.10 [28] has provided several GA optimized irregular arrays with top SNR performance in specified acoustic scenes. It can be noted that for the case with discrete noise sources, high microphone densities exist in the area near target and noise positions. For the cases with continuous noise space, the optimal irregular arrays have sparse microphones over the noise space, but place microphones surrounding the noise area from the edge to the other side of room to generate a set of DPDs with large dispersion. This kind of spatial density distribution of optimal microphone arrangements can be successfully explained by the Hyperbola Area proposed in Section 2.4.1. It concluded that for one pair of target and noise positions, the largest spread DPDs will be generated if microphones are clustered inside the hyperbola areas with target and noise positions as focuses. For example, Figure 5.10(a) gives a irregular array with top SNR   performance   resulted   from   GA   optimization,   where   one   target   and   three

interferences are considered for this scene. The hyperbola areas for each target-noise pair are marked in dashed line with different color. It can be seen that most of microphones in the GA-optimized array are clusted in these hyperbold areas (grey areas). In order to demonstrate this conlcusion, Figure 5.10(b) provides a example for the irregular array clustered based on the hyperbola theory in the same scene. Microphones are divided into four clusters uniformly distributed in these four hyperbola areas. Simulations are performed with human speech signals in corresponding acoustic scenes. The SNR results as in Table 5.7 demonstrate previous conclusion by showing that the hyperbola clustered arrays have comparable or even better SNR results than GA optimized arrays, while great SNR improvements are observed in both of these geometries compared with corresponding regular arrays.
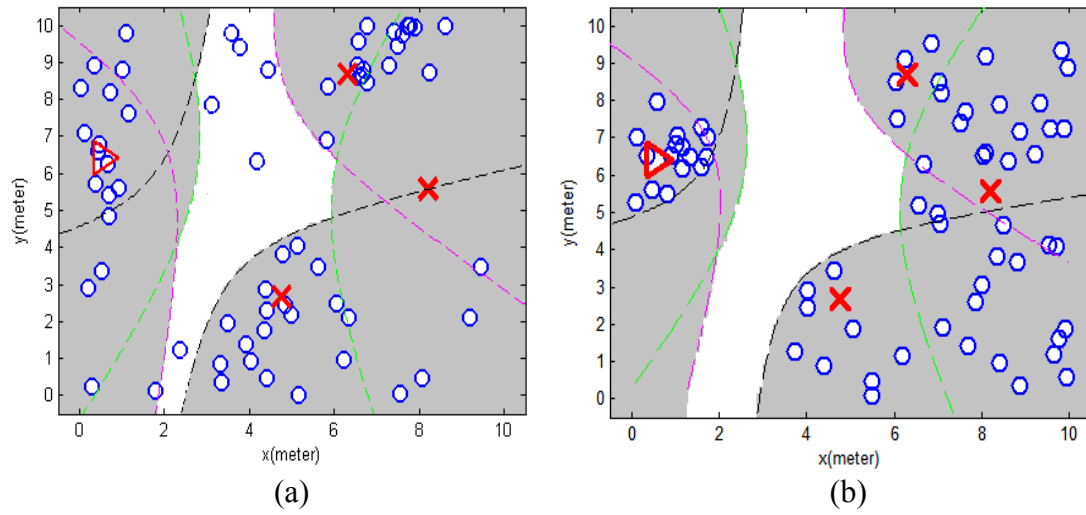

Figure 5.10: Top view of GA-optimized irregular array and hyperbola clustered array. Blue circles represent microphones. Red crosses represent the possible noise space. Red triangle is the desired target space. (a) GA-optimized irregular array. (b) Hyperbola clustered array.

Table 5.7: SNR comparison of hyperbola clustered arrays and corresponding GA optimized arrays

Acoustic Scenes	Hyperbola cluster array set (50 arrays)		GA optimized irregular array set (50 arrays in the last generation of GA)		Randomly distributed irregular array set (50 arrays in the 1st generation of GA)		Regular array
	Top 3 SNR (dB)	Average SNR (dB)	Top 3 SNR (dB)	Average SNR (dB)	Top 3 SNR (dB)	Average SNR (dB)	SNR (dB)
64 mics arrays with continuous noise space	8.92 8.87 8.71	6.81	9.03 8.99 8.98	8.45	6.47 5.67 5.49	3.83	3.40
64 mics array with discrete noise sources	28.04 27.38 27.03	24.76	26.33 25.78 25.63	23.83	22.61 22.18 22.09	17.96	17.28
9 mics array with discrete noise sources	21.83 21.07 20.69	17.93	18.24 17.71 17.65	16.56	17.78 17.62 17.20	9.01	8.89

Five separate recordings with different signal power levels were performed for the GA-optimized array with 9 microphones, the hyperbola clustered array and regular planar arrays over the ceiling of the aluminum cage, where colored noise generated by the band importance function from the SII model are played through the speakers and varied for each recording. As shown in Figure 5.11, the SRCP images are created to validate the superiority of GA-optimized irregular array and hyperbola clustered array when comparing with regular array.

(a)

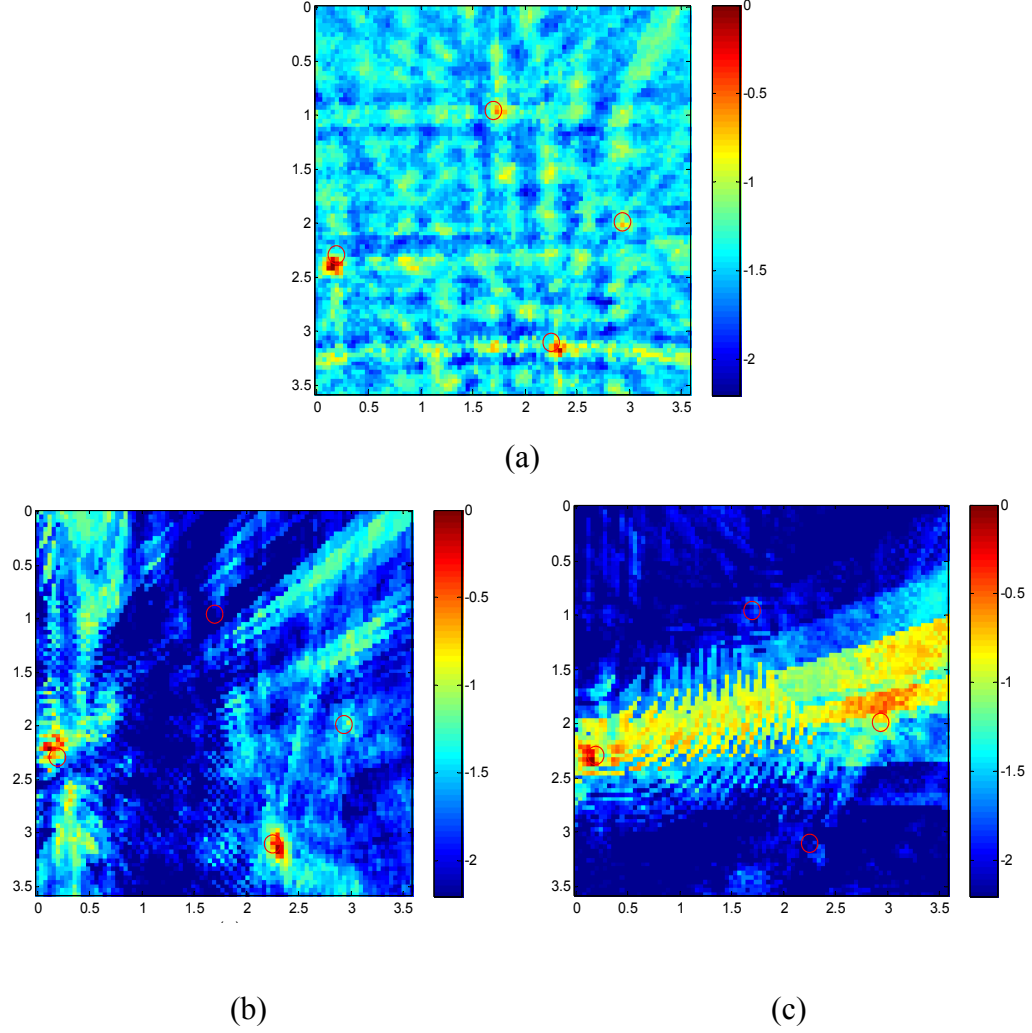(b)   (c)

Figure 5.11: Top view SRCP images averaged overall time slots. The red circles represent source positions. (a) Regular planar array. (b) GA-optimized irregular array. (c) Hyperbola clustered arrays.

In conclusion, the hyperbola theory successfully explains the optimal microphone density distribution from the results of optimization approaches in some degree. To be specified, microphones should be clustered in the hyperbola areas of each target and interference position pair to generate a DPD distribution with rich entropy, and further improve array noise suppression ability. By following this conclusion, hyperbola clustered arrays can be directly generated according to the prior knowledge of acoustic scene, and provide comparable SNR performance with the GA-optimized arrays. The hyperbola cluster design method has been demonstrated to be a easy and feasible method

for the *ad hoc* (not computer aided) microphone array design.

5.3.2   Multilayer Cluster

Previous section proposed a array design method with the element clustered in the hyperbola areas, while microphones are uniformly distributed in each cluster (subarray). Inspired by the theory of cell division and differentiation, this cluster design method can be extended to different types of subarray and cluster distribution, which can be generalized as the multilayer cluster design method. The details are provided as below.

   Step 1: According to the prior knowledge of the acoustic scene, choose several initial configurations as the bases of array geometry, which can be either irregular or regular arrangements as discussed in section 4.2. The basic configurations set can be expresses as $\{ \boldsymbol{G} \}$.

   Step 2: Based on the predefined objective function, search for the optimal combination of $\{g, L, a, \sigma, J\}, g \in \boldsymbol{G}$ as the initial layer of array, where g represents a basic array configuration in $\boldsymbol{G}$. L, a, σ, J are the key geometry descriptors to identify the microphone positions in g .

   Step 3: If the total number of microphone P is larger than the number of element P' in the initial layer with $\{g, L, a, \sigma, J\}$, consider each element as the centroid of one subarray. This problem turns into P' subarrays design problems with the same objective function. With fixed centroids, pick the optimal combination of $\{g, a, \sigma, J\}$ for each subarrays to form the final array configuration. This division process will continue until all microphones are used.

   Table 5.8 generalizes the possible initial configurations of $\boldsymbol{G}$ and related key geometrical parameters. The characteristics of beampattern of these configurations have be discussed in section 4.2. The configurations provided in this table can be considered as the bases of array geometry in each layer. Moreover, when applying Clustering Analysis (CA) to evaluate the beampattern of a complex array with large number of microphone, Table 5.8 can also be considered as the definitions of the type of small clusters to decompose the original array, which is actually the inverse process of multilayer design approach.

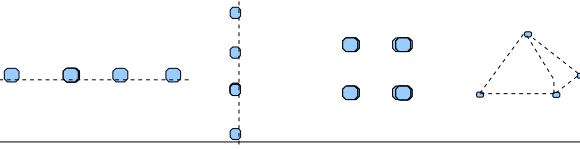Table 5.8: Several basic array configurations and related geometry descriptors

Number of Microphones	Array Configuration	Geometry Descriptor
2		$\{L, a\}$ or directly use microphone coordinates
3		$\{L, a\}$
4		$\{L, a\}$
Large number	Uniformly distributed in the design space	$\{L, a, \sigma, J\}$

Figure 5.12 and Figure 5.13 provide a simple example for a 2 layer 2D array design case of 6 microphones. In this example, a pairwise array and related geometry parameters are chosen as the initial layer to meet the requirement of MLW in interested steering angle. Then each element in the initial layer is transferred to a triangle subarray to reduce the sidelobe levels. It can be seen that the final result of beampattern generally keeps the initial width of main lobe, while greatly reducing the sidelobe levels and adding one dimension spatial resolution along y axis. Moreover, because the optimizing procedures only search for two geometrical parameters of two elements in the initial layer to meet MLW requirement, and one geometrical parameter of two sub-triangle arrays in the final layer to meet MPSR, the computational complexity has been greatly reduced when comparing with the heuristic searching method.
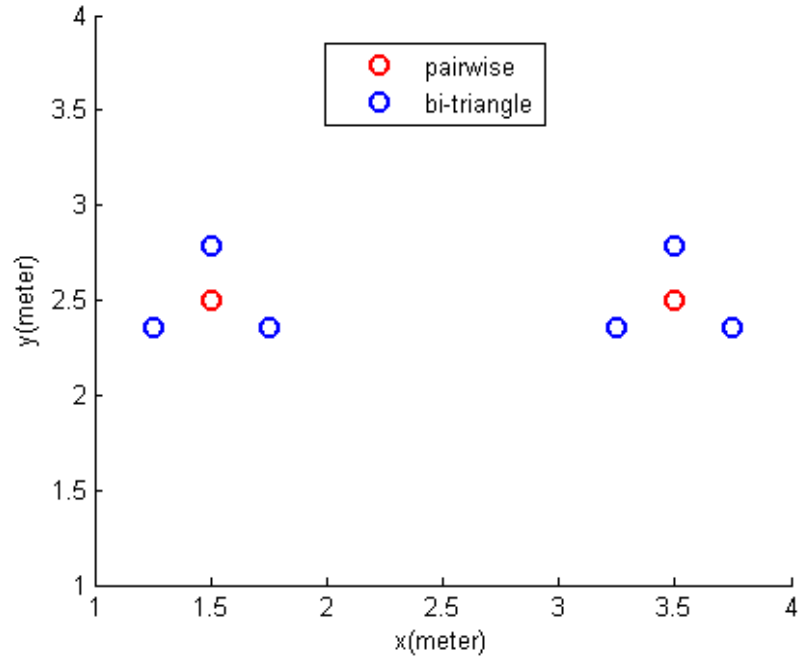
Figure 5.12: Two layer array design. Circles represent microphones. Step 1: obtain the optimal red positions as initial layer. Step 2: extend each red dot to a subarray marked as the blue positions.
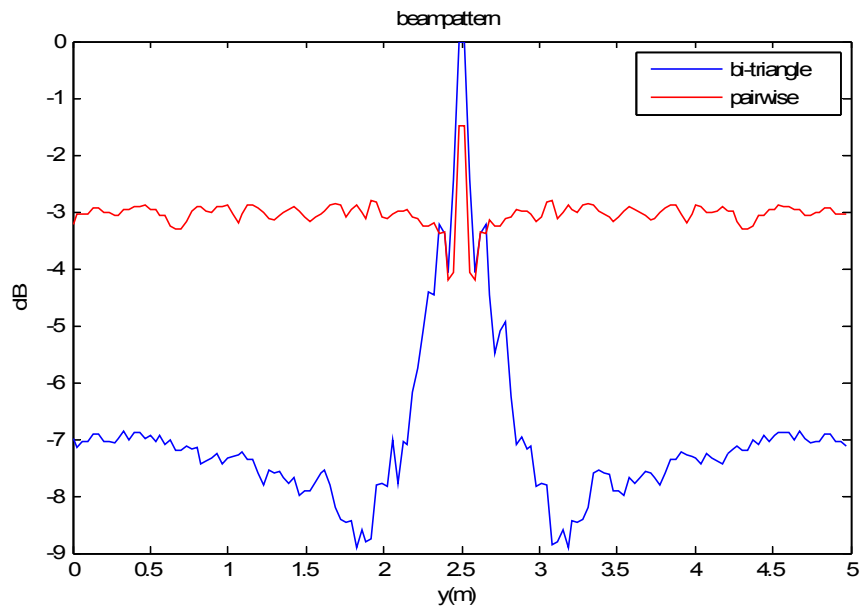

Figure 5.13: Resulted beampatterns of the arrays in Figure 5.12

Figure 5.14 gives another example [81] for the multilayer array design, where spiral arrays are considered as the initial layer and each element of spiral is transferred to a circular subarray as the final layer. By controlling the $\{ \varphi, a, b \}$ of spiral array (geometry parameters as discussed in section 4.2.5) in the first step and the number of elements in each circular subarray, the microphone density distribution can be adjusted according to the source distribution. As discussed in [81], this kind of arrays with circularly symmetry and zero redundancy of inter-mic spacing can substantially eliminate the grating lobes over a broad range of frequencies in near field, while the required number of microphone is greatly reduced than a regular structure.

(a)   (b)
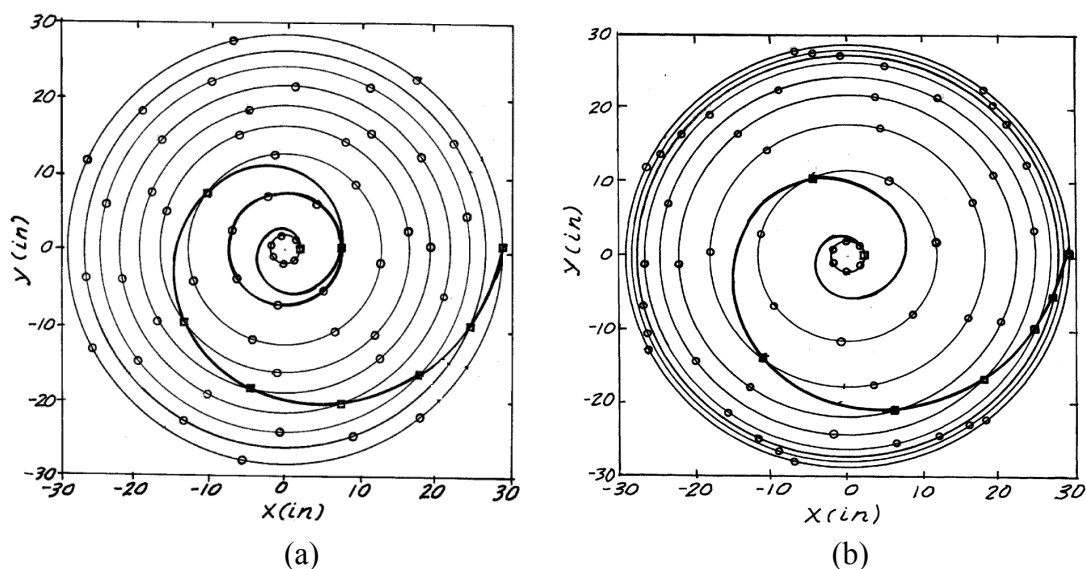
Figure 5.14: Circularly symmetric and non-redundant planar arrays (adapted from [81])

In conclusion, the basic ideal of multilayer cluster design method is to use combinations of regular basic configurations to represent or decompose random arrangements, which brings the benefit of irregular element placements. By breaking the design problem into multilayer subarrays, the variables of optimization are greatly reduced. In addition, in terms of searching direction of optimization, these array design approaches search the optimal microphone placements by the guide of basic configurations while the centroids of subarray are obtained from previous step. It will greatly reduce the computational complexity.

131

The complex relationship between array gain patterns and microphone distributions limits the application of traditional optimization algorithms on irregular arrays. Due to the randomness of microphone placements, it is difficult to control or predict the irregular array performance. This chapter introduced three kinds of optimization methods for the irregular array design in immersive environments with broadband signals. The traditional analytical methods based on spatial perturbation were extended to the arbitrary irregular array in near-field applications. Based on the relationships of important geometry descriptors with performance measures, heuristic searching methods were proposed with the probabilistic functions of acoustic scene to incorporate various levels of prior knowledge of the source distribution. By successfully controlling the statistic geometry features of array configurations according to the source distribution, cluster design methods were also introduced to reduce the degree of freedom in the optimization procedure of irregular array, which make it possible for the *adhoc* array design (not computer aided) with easy operation and installation. Simulation and experimental results have demonstrated that these optimization methods are feasible for the irregular microphone array applications in immersive environments where rapid deployment is required with limited knowledge of the acoustic scene, such as in mobile platforms and audio surveillance applications.

Moreover, as discussed before, no matter which optimization approach is applied, due to the restrict of small perturbations and possible ill-condition of iteration, the proper choice of initial array configuration is crucial for resulting in global optimum solutions, as well as reducing computational complexity and improving the converging speed of the optimization procedure. Instead of equispaced linear and planar arrays in traditional methods, various basic array configurations were applied in our research to meet the requirements of different applications. Based on the conclusions of Chapter 3-5 for the array design, in the real cases the initial array configurations and related optimization approaches can be chosen flexibly or combined together. For example, (1) MRA can be applied first to obtain desired spatial resolution, then use spatial perturbation method to slightly adjust the element placements to obtain minmax sidelobe level. (2) Apply GA to search globally, and take the resulted array as the initial condition of Gaussian iterations, which has a fast converging speed for local optimum. In summary, a smart initial guess of

array configurations can provide a faster convergence to the optimal solutions, improve the robustness, and efficiently reduce the computational complexity of optimization procedure.

Chapter 6   Conclusions and Future Work

Irregular arrays, which diversify microphone positions, can achieve better performance than traditional regular arrays, especially for human speech applications in immersive environments. Due to the randomness of element positions, it is difficult to control or predict the behavior of irregular arrays, while the geometry parameters applied for regular array analyses are limited for explaining the performance differences of irregular arrays.

Our work, therefore, analyzed the relationship between irregular microphone geometries and spatial filtering performance with statistical methods. Combined with descriptors analogous to traditional parameters for regular arrays (i.e. array centroid and dispersion), novel geometry descriptors involving DPD statistics were developed to capture the properties of both irregular and regular microphone distributions showing their impact on array performance. It has been demonstrated that arrays with enhanced noise suppression ability should provide a DPD distribution with wide spread and high entropy to decorrelate the noise from target signals. Feasible optimization algorithms were proposed with the objective function rules using established relationship functions and probabilistic descriptions of acoustic scenes to incorporate various levels of prior knowledge of the source distribution. General guidelines in real scenarios and cluster design methods were also introduced to effectively control the key geometry descriptors related to microphone density distributions, and directly build arrays with superior SNR performance. In addition, arrays mutated from the regular configurations were also

introduced to overcome the limitations of regular arrays, and provide good SNR results for speech signals with easily installation. To verify the effectiveness of these proposed methods, simulated gain patterns and real SNR results of the optimized arrays were compared to corresponding traditional regular arrays and arrays obtained from direct exhaustive search method. Large SNR enhancements and acceptable processing time were observed in most of these cases, which make the proposed design methods applicable for the environments required rapid deployment, such as in mobile platforms and audio surveillance applications.

The results of this dissertation were based on the data derived from Monte Carlo experiments with ceiling arrays, most of the conclusions are more applicable for the human speech signals in indoor environments and near-field applications, where array distances from SOI have comparable values (less than three times) with array aperture. (Centroid offset should be smaller than three times of array dispersion.) In addition, although proposed DPD measures explained most of the performance differences of irregular arrays with fixed centroid and dispersion, there are still $\pm$ 0.5~1dB variances of performance metrics observed for each level of DPD statistics. Other more direct geometry parameters with simple geometric interpretations may exist that can further reduce these variations and be operated easily.

References

[1]   D. Rabinkin, R. Renomeron, and J. French, "Optimum sensor placement for array sound capture", *Proceedings of SPIE*, 3162, 227-239, 1997.

[2]   J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing,* vol. 1, ed. by J Benesty, W Kellermann, Springer, Berlin Heidelberg, 2008.

[3]   P. Howells, "Intermediate frequency side-lobe canceler", *US. Patent 3,202,990*, 1965.

[4]   B. Widrow, "A microphone array for hearing aids", *IEEE Circuits and Systems Magazine*,1(2): 26-32, 2001.

[5]   C. Knapp, G. Carter, "The generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4): 320-327, 1976.

[6]   D. Youn, N. Ahmed, and G. Carter, "On using the LMS algorithm for time delay estimation", *IEEE Transactions on Acoustics,Speech and Signal Processing*, 30(5): 798-801, 1982.

[7]   S. Applebaum, "Adaptive array", *IEEE Transactions on Antennas and Propagation*, 24(2): 585-598, 1976.

[8]   O. Frost, "An algorithm for linearly constrained adaptive array processing", *Proceedings of the IEEE*, 60(8): 926-935, 1972.

[9]   H. Pan, *Research on near field beam forming*, Thesis, University of Electronic Science and Technology of China, 2007.

[10]   B. Carlson, "Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays", *IEEE Transactions on Aerospace and electronic Systems*, 24(4): 397-401, 1988.

[11]   J. Kim, ets, "A robust adaptive array based on signal subspace approach", *IEEE*

Transactions on Signal Processing, 41(11), 3166-3171, 1992.

[12]   S. Yu, J. Lee, "The statistical performance of eigenspace-based adaptive array beamformers", *IEEE Transactions on Antennas and Propagation*, 44(5): 665-671, 1996.

[13]   A. Muthukumarasamy, K. Donohue, "Impact of microphone placement errors on speech Intelligibility", in *Proceedings of the IEEE Southeastcon* 2009, Atlanta, GA, USA, March 2009.

[14]   D. Rabinkin, *Optimum sensor placement for microphone arrays,* PhD thesis, Rutgers the State University of New Jersey, 1998.

[15]   D. Cheng, "Optimization techniques for antenna arrays", *Proceedings of IEEE,* 59(12): 1664-1674, 1971.

[16]   W. Stutzman, G. Thiele, *Antenna theory and design*, Wiley, 2th edition, 1997.

[17]   F. HodJat, S. Hovahessian, "Nonuniformly spaced linear and planar array antennas for sidelobe reduction" *IEEE Transactions on Antennas and Propagation,* AP-26 (2), March 1978.

[18]   I. Haya, B. Petersen, B. Colpitts, "Optimum 2-D LOS MIMO performance using omni-directional antennas attained through genetic algorithms", in *Communication Networks and Services Research Conference*, Halifax, NS, Canada, 2008.

[19]   M. Ma, *Theory and application of antenna arrays,* Wiley, New York, 1974.

[20]   B. Steinberg, *Principles of aperture and array system design: including random and adaptive arrays*, Wiley, New York, 1976.

[21]   R. Kennedy, T. Abhayapala, and D. Ward, "Broadband nearfield beamforming using a radial beampatten transformation", *IEEE Transactions on Signal Processing*, 46(8): 2147-2156, 1998.

[22]   C. Knapp, G. Carter, "The generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustics, Speech and Signal Processing,* 24(4): 320-327, 1976.

[23]   D. H. Youn, N. Ahmed, and G. C. Carter, "On using the LMS algorithm for time delay estimation", *IEEE Transactions on Acoustics,Speech and Signal Processing*, 30(5): 798-801, 1982.

[24] S. Shanan, C. Pomalaza-Raez, "The use of nonuniform element spacing in array processing algorithms", *J. Acoust. Soc. Am*, 86, p1416-1418, 1989.

[25] H. Schjaer-Jacobsen, K. Madsen, "Synthesis of nonuniformly spaced arrays using a general nonlinear minmax optimization method", I*EEE Trans. Antennas Propagat.*, AP-24, 501-506, 1976.

[26] P. Townsend, *Enhancements to the generalized sidelobe canceler for audio beamforming in an immersive environment*, MS thesis, Department of Electrical Engineering, University of Kentucky, 2009.

[27] M. Brandstein, D. Ward, *Microphone Arrays Signal Processing Techniques and Applications,* ed. by A Lacroix, A Venetsanopoulos, M Brandstein, D Ward, p. 3, Springer, Berlin Heidelberg New York, 2001.

[28] J. Yu, K. Donohue, "Performance for randomly described arrays", in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* 2011.

[29] J. Meyer, G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the sound field". Paper presented at the *IEEE ICASSP-02*, Orlando, Florida, USA, 13-17, May 2002.

[30] J. Daniel, R. Nicol, S. Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging", in *114th Convention of Audio Engineering Society,* 2003.

[31] A. Moffet, "Minimum-redundancy linear arrays", *IEEE Trans. Antennas Propagat.*, AP-16, pp. 172-175, 1968.

[32] W. Chen, Y. Bar-Ness, "Minimum redundancy array structure for interference cancellation", in *Antennas and Propagation Society International Symposium*, 1, pp. 121-124, 1991.

[33] S. Pillai, Y. Bar-Ness, F. Haber, "A new approach to array geometry for improved spatial spectrum estimation", *Proc. IEEE*, 73, pp. 1522-1524, Oct. 1985.

[34] P. Townsend, K. Donohue, "Beamfield analysis for statistically described planar microphone arrays"*, in Proceedings of the IEEE Southeastcon* 2009, Atlanta, GA, USA, March 2009.

[35] Z. Li, K. Yiu, and Z. Feng, "A hybrid descent method with genetic algorithm for microphone array placement design", *Applied Soft Computing* , 2012.

[36]   Z. Feng, K. Yiu, and S. Nordholm, "Placement design of microphone arrays in near-field broadband beamformers", *IEEE Trans. on Signal processing*, 60(3): 1195-1204, 2012.

[37]   K. Donohue, J. Hannemann, and H. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments", *Elsevier Science Direct,* 2007.

[38]   Near and Far field (2013). *http://en.wikipedia.org/wiki/Near_and_far_field*. Accessed 2012.

[39]   K. Donohue, K. McReynolds, A. Ramamurthy, "Sound source detection threshold estimation using negative coherent power", in *IEEE Southeastcon* , 2008.

[40]   K. Donohue, S. SaghanianNejadEsfahani, J. Yu, "Constant false alarm rate sound source detection with distributed microphones", *EURASIP J Adv Signal Process,* 2011. doi:10.1155/2011/656494

[41]   E. Pielou, "The measurement of diversity in different types of biological collections", *J. Theoret. Biol.*, 13, 131–144, 1966.

[42]   NTI Audio. Introduce to speech intelligibility (2008). *http://www.nti-audio.com Portals/0/Products/Minstruments/AL1/AppNotes/NTI_App_Note_Introducing_ST I-PA.pdf*. Accessed 2011.

[43]   SAS Institute Inc. SAS/STAT(R) 9.22 user's guide (2010). *http://support.sas.com/ documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_anova_sect 022.htm*. Accessed 2011.

[44]   S. Gazor. Y. Grenier, "Criteria for positioning of sensors for a microphone array", *IEEE. Transaction on Speech and audio processing,* vol 3(4), 1995.

[45]   D. K. Cheng, "Optimization techniques for antenna arrays", *Proceedings of IEEE,* vol 59(12), 1971.

[46]   G. V. Reklaitis, A. Ravindran, K. M. Ragsdell, *Engineering optimization methods and applications*. John Wiley & Sons, 1983.

[47]   S. Rao, *Engineering optimization theory and practic*e. John Wiley & Son. 1996.

[48]   H. Pan, *Research on near field beam forming.* Thesis, University of Electronic Science and Technology of China, 2007.

[49]   C. L. Dolph, "A current distribution for broadside arrays which optimizes the

relationship between beam width and side-lobe level", *Proceedings of I.R.E. And Waves and Electrons,* June 1946.

[50]   S. A. Schelkunoff, "A mathematical theory of linear arrays", *Bell Syst. Tech. J. ,* 22. pp. 80-107, 1943.

[51]   R. P. Dooley, "The optimum design of small nonumiformly spaced arrays", *IEEE transactions on Antennas and Propagation,* September 1972.

[52]   C. H. Tang. "On the optimum performance of nonuniformly spaced arrays", *Communications*. Sep. 1966.

[53]   J. Bae, K. Kim, J. Lee, H. Kim, J. Choi, "Design of steerable nonuniform linear array geometry for side-lobe reduction", *Microwave and optical Technology Letters*, 36 (5), pp. 362-367, 2003.

[54]   J. Bae, K. Kim, C. Pyo, and J. Chae, "Design of scannable nonuniform planar array structure for maximum side-lobe reduction", *ETRI Journal*, 26 (1)**,** pp. 53-56, Feb 2004.

[55]   F. HodJat, S. A. Hovahessian, "Nonuniformly spaced linear and planar array antennas for sidelobe reduction". *IEEE Transactions on Antennas and Propagation*. AP 26 (2), pp. 198-204 , March 1978.

[56]   L.Kogan. "Optimizing a large array configuration to minimize the sidelobes", *IEEE Transactions on Antennas and Propagation*, 48 (7), pp. 1075-1078, July 2000.

[57]   C. Yu, "Sidelobe reduction of asymmetric linear array by spacing perturbation". *IEEE Electronics Letters*, 33 (9), pp. 730-732, 1997.

[58]   C. Yu, "Sidelobe reduction of a symmetric broadside array by spacing perturbation", *Microwave and Optical Technology Letters*. 13 (3), pp. 147-149, 1996.

[59]   D. E. Dudgeon, "Fundamental of digital array processing", *Proceedings of IEEE,* vol 65, 1977.

[60]   W. L. Stutzman, G. A. Thiele, *Antenna theory and design*. Wiley. 2th edition. 1997.

[61]   D. Giuliani, M. Matassoni, M. Omologo, P. Svaizer. "Use of different microphone array configurations for hands-free speech recognition in noisy and reverberant

environment", *Proc. Eurospeech*, 1997.

[62]   M. I. Skolnik, G. Nemhauser, J. W. Sherman, "Dynamic programming applied to unequally spaced arrays", *IEEE Transactions on antennas and propagation,* 1964.

[63]   D. Pearson, S. U. Pillai, Y. Lee, "An algorithm for near-optimal placement of sensor elements", *IEEE Transactions on Information Theory,* vol 36(6), 1990.

[64]   D. A. Linebarger, "A fast method for computing the coarray of sparse linear arrays", *IEEE Trans. Antenn. Propagat.,* vol. AP40 (9), 1992.

[65]   J. Leech, "On the representation of 1,2,,...,n by differences", *J. London Math. Society,* vol 31 pp. 160-169, 1956.

[66]   L. Presti, G. Montalbano, G. Povero, "Beamforming design with partially adaptive and minimum redundancy arrays", in *IEEE International Conference on* Acoustics, Speech, and Signal Processing, 1996.

[67]   S. R. Christopher, "Numberical annealing of low redundancy linear arrays", *IEEE Transaction on Antennas and propagation,* vol 41(1), 1993.

[68]   E. Vertatschitsch, S. Haykin, "Nonredundant arrays", *Proceedings of the IEEE,* vol 74(1), Jan. 1986.

[69]   J. Christensen, J. Hald, *Technical review beamforming.* Bruel&Kjaer Sound & Vibration Measurement A/S, DK-2850 Naerum, Denmark, 2004.

[70]   H. Vogel, "A better way to construct the sunflower head", *Mathematical biosciences,* 44:179-189, 1978.

[71]   Archimedean spiral (2013). *http://en.wikipedia.org/wiki/Archimedean_spiral*. Acccessed 2012.

[72]   H. Schjaer-Jacobsen, K. Madsen, "Synthesis of nonuniformly spaced arrays using a general nonlinear minmax optimization method". *IEEE Transactions on Antennas and Propagation*, 24 (4), pp. 501-506, July 1976.

[73]   R. F. Harrington, "Sidelobe reduction by nonuniform element spacing". *IRE Transactions on Antennas and Propagation*, 9 (2), pp. 187-192, March 1961.

[74]   Z. Li, K. Yiu and Z. Feng. "A hybrid descent method with genetic algorithm for microphone array placement design", *Applied Soft Computing,* 2012. doi:10.1016/j.bbr.2011.03.031

[75] Z. G. Feng, K. Yiu, and S. Nordholm, "Placement design of microphone arrays in near-field broadband beamformers". *IEEE Trans. on Signal processing*, 60 (3), pp. 1195-1204, 2012.

[76] M. Simpson, C. Hansen, "Use of genetic algorithms to optimize vibration actuator placement for active control of harmonic interior noise in a cylinder with floor structure"*, Noise Control Eng. J,* 44 (4), pp. 169-184, 1996.

[77] A. Montazeri, J. Poshtan, and M.H. Kahaei, "Optimal placement of loudspeakers and microphones in an enclosure using genetic algorithm", in *Proceedings of 2003 IEEE Conference on Control Applications*, Istanbul, Turkey, 2003.

[78] I. Haya, B. Petersen, and B. Colpitts, "Optimum 2-D LOS MIMO performance using omni-directional antennas attained through genetic algorithms", in *Com. Networks and Services Research Conference*, Halifax, NS, Canada, 2008.

[79] R. Akbari, K. Ziarati, "A multilevel evolutionary algorithm for optimizing numerical functions", *International J. of Industrial Eng. Computations*, 2(2011), pp. 419-430, 2011.

[80] J. Yu, K. Donohue, "Geometry descriptors of irregular microphone arrays related to beamforming performance", *EURASIP Journal on Advances in Signal Processing*, 2012**:**249, 2012. doi:10.1186/1687-6180-2012-249

[81] J. R. Underbrink. Circularly symmetric, zero redundancy, planar array having broad frequency range application. (2001), *http://www.google.com/patents? id=hTgGAAAAEBA*J&pg=PA5&hl=zhCN&source=gbs_selected_pages&cad=4# v*=onepage&q&f=false.* Accessed 2011.

Vita

Jingjing Yu was born in Henan, China. She received the B.S. and M.S. degrees in communication engineering from Beijing Jiaotong University, Beijing, China, in 2005 and 2007, respectively. She is currently a Ph.D. candidate in Electrical Engineering Department at the University of Kentucky, Lexington, KY. And she is also a research assistant in the Center for Visualization & Virtual Environments at the University of Kentucky. Her current research interests include communication and information system, audio signal processing, signal processing, and engineering optimization. Ms. Yu is a student member of the IEEE.