國 立 交 通 大 學

機械工程研究所
碩 士 論 文

微機械麥克風之最佳化設計及其陣列於 3D 聲場重建之實現

Optimal Design and Experimental Verification of a MEMS Microphone Array for 3D Hearing

研 究 生：林振邦

指導教授：白明憲

中華民國九十三年六月

微機械麥克風之最佳化設計及其陣列於 3D 聲場重建之實現

研究生 ： 林振邦　　　　　　　　指導教授 ： 白明憲

國立交通大學機械工程學系

摘要

電容式麥克風有較高的靈敏度與較平坦與寬廣的頻率響應，因此被廣泛的應用於各種場合。利用微機電技術可將體積微小化，卻不失其性能，並可降低雜散電容的影響。文中探討微機械電容式麥克風的動態行為分析，並利用田口實驗法(Taguchi method)與遺傳演算法(Genetic Algorithm)最佳化其性能與尺寸。然後利用最佳化之尺寸來設計微機械電容式麥克風，使其有最佳的靈敏度與頻寬。利用陣列訊號處理技術可提高陣列麥克風的訊噪比(Signal-to-Noise Ratio)與指向性(Directivity)。其中延遲補償法(Delay-Sum method)與超指向性法(Superdirective method)可用來做聲源方向估計(Direction of Arrival)與聲束形成(Beamforming)。設計超指向性濾波器以提高陣列麥克風的訊噪比與指向性。再搭配頭部轉移函數(Head Related Transfer Function)來重建空間聲場。實際的聽覺定位實驗中，利用耳機輸出之空間聲場有良好的立體空間感。此系統可應用於助聽器，不只可提高聲源的清晰度，也可清楚辨別聲源方向，更可讓聽者猶如置身立體空間聲場。

Optimal Design and Experimental Verification of a MEMS Microphone Array for 3D Hearing

Student: Chenpang Lin   Advisor: Mingsian Bai

Department of Mechanical Engineering,National Chiao-Tung University.

ABSTRACT

Capacitive microphones are widely used in high-quality recording because of its high sensitivity and flat frequency response. Now, it can be fabricated by Micro-Electro-Mechanical System (MEMS) technology for smaller size and better performance. Taguchi method and Genetic Algorithm (GA) are proposed to optimize the performance of sensitivity and bandwidth of the condenser microphone. Then, the parameters of diaphragm length, diaphragm thickness and air gap height can be optimized. Therefore, the MEMS condenser microphone fabrication processes are designed with the optimized parameters. Array signal processing is utilized to enhance Signal-to-Noise Ratio (SNR) and directivity of a microphone array. The comparisons between a broadside and an endfire array are illustrated. Delay-sum and superdirective methods are presented to do the Direction of Arrival (DOA) estimation and beamforming. Directivity analyses are discussed in the cases of beamwidth, directivity and SNRs. Three-dimensional (3D) spatial sound is reconstructed by inverse-filtering with Head Related Transfer Functions (HRTFs).   It is implemented using TMS320C3X Digital Signal Processor (DSP) to realize 3D spatial sound. Therefore, it can be utilized for hearing aids or other applications to enhance SNR and directivity.

誌謝

時光飛逝，短短兩年的研究生生涯轉眼就過去了。首先感謝指導教授白明憲博士的諄諄指導與教誨，使我順利完成學業與論文，在此致上最誠摯的謝意。而老師指導學生時豐富的專業知識，嚴謹的治學態度以及待人處事方面，亦是身為學生的我學習與景仰的典範。

在論文寫作上，感謝本系成維華教授、陳宗麟教授在百忙中撥冗閱讀並提出寶貴的意見，使得本文的內容更趨完善與充實，在此本人致上無限的感激。

回顧這兩年的日子，承蒙同實驗室的博士班曾平順學長、歐昆應學長、顏坤龍學長、陳榮亮學長、蘇富誠學長、林家鴻學長與李志中學長在研究與學業上的適時指點，並有幸與白淦元、董志偉、曹登傑及廖哲緯同學互相切磋討論，每在烏雲蔽空時，得以撥雲見日，獲益甚多。此外學弟曾文亮、何柏璋、周中權與林建良在生活上的朝夕相處與砥礪磨練，都是我得以完成研究的一大助因，在此由衷地感謝他們。

能有此刻，我也要感謝所有在精神上給我鼓舞支持的人，謝謝各位的幫忙與鼓勵。最後僅以此篇論文，獻給我摯愛的父親林增明先生、母親彭秀勤女士、姊姊林翠薇、哥哥林振鵬。今天我能順利取得碩士學位，要感謝的人很多，上述名單恐有疏漏，在此也一併致上我最深的謝意。

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A microphone is a transducer, which transforms acoustical energy to electrical energy and can be fabricated by Micro-Electro-Mechanical System (MEMS) technology.[1] The types of microphone consists of piezoelectric,[2] piezoresistive[3] and capacitive[4-10]. Among them, the performance of the capacitive microphone is best because of its high sensitivity, flat frequency response and low noise level. The advantages of the MEMS technology are the small size and better performance. However, the fabrication of micromachining induces a large initial stress in the diaphragm. The large initial stress of the diaphragm degrades the performance of the condenser microphone. Therefore, how to decrease the initial stress of the diaphragm is an important problem. The material of diaphragm should be chosen carefully.

The linear dynamic model of the MEMS condenser microphone is introduced.[11-13] Therefore, the dynamic performances can be acquired. In order to maximize the performances of the MEMS condenser microphone, Taguchi method[14,15] and Genetic Algorithm (GA)[16,17] are presented. The performances of the MEMS condenser microphone consist of sensitivity and bandwidth. We expect the sensitivity and bandwidth as larger as possible, but the relationship between them is trade-off. Therefore, Taguchi method and GA are utilized to optimize the values of the sensitivity and bandwidth. The parameters that need to be optimized are diaphragm length, diaphragm thickness and air gap height. From the results of optimization, sensitivity and bandwidth are improved, respectively. After optimization, the optimal parameters are acquired. Therefore, the processes of the microphone diaphragm can be designed and fabricated using the MEMS technology. The purpose to fabricate the microphone diaphragm is to know the resonant frequency in the diaphragm. Then, the initial stress of the diaphragm can be acquired and the

sensitivity of the diaphragm can be controlled.   In addition, Poly-Silicon is used for diaphragm material to degrade the initial stress.   On the other hand, the processes of the MEMS condenser microphone structures are also designed with a single silicon wafer.   The fabrication is not done, just in design steps.

Array signal processing[18-20] has been widely used in the areas such as radars, sonars, communications, and seismic exploration, and underwater imaging.   The major advantage is the enhancement of the Signal-to-Noise Ratio (SNR).   Besides, the directivity of the microphone array can be improved to be effective in eliminating background noise by beamforming techniques.   Array signal processing of Direction of Arrival (DOA) estimation and beamforming techniques are utilized.   The delay-sum and superdirective methods are utilized to do the DOA estimation and beamforming.   The locations of the sound source can be estimated.   Then, a microphone array can point to the direction of source to receive signals.   Therefore, SNR and directivity can be improved greatly.

In order to reconstruct three-dimensional (3D) spatial sound field, the microphone array with Head Related Transfer Function (HRTF)[21] are utilized to realize our goal.   Most hearing-impaired people will be able to hear speech when given sufficient amplification from their hearing aids.   However, they will hear but will not understand because of poor SNR.   In a noisy place, hearing aids will amplify the noise as well as the desired speech signal.   In a reverberant place, hearing aids will amplify late reverberation as well as the direct first arrival signal.   Furthermore, acoustical feedback also degrades the SNR and distorts the frequency response of the hearing aids.   The microphone array for hearing aids can enhance SNR.   Besides, it can reduce the effects of reverberation and acoustical feedback.   The microphone array is worn on the chest as part of a necklace to receive the sound source.   The filters between the microphone array and hearing-impaired people's ears can be

calculated by sound propagation transfer function and HRTF. The inverse filters can be implemented to reconstruct 3D spatial sound field. Then signals received by the microphone array do the DOA estimation to find the location of the sound source. Therefore, the 3D spatial sound can be reconstructed with a microphone array using HRTF. Therefore, the hearing-impaired people can hear the spatial sound in three-dimensional space. The aforementioned advantages are also included.

II. MEMS CONDENSER MICROPHONES

Most of the silicon microphones presented in the literature are based on the capacitive principles because of its high sensitivity, flat frequency response and low noise level. The capacitive microphone consists of a thin, flexible diaphragm and a rigid backplate. Mechanical sensitivity is mainly determined by the initial stress and Young's modulus of the diaphragm. Large initial stress and Young's modulus in the diaphragm can seriously degrade the sensitivity of the condenser microphones. In order to improve the sensitivity, low-stress Poly-Silicon is used as the diaphragm material.[22] First, linear dynamic model of a condenser microphone needs to be constructed. Then, dynamic performance of frequency response, sensitivity and bandwidth of a condenser microphone can be acquired. After this, Taguchi method and GA are used to maximize the sensitivity and bandwidth of a condenser microphone. The scales of diaphragm length, diaphragm thickness and air gap height of a condenser microphone can be optimized. Therefore, the performances of a MEMS condenser microphone can be improved.

A. The Linear Dynamic Model

Figure 1 shows the cross-sectional view of a single MEMS condenser microphone with a diaphragm, an air gap, a perforated backplate and a back chamber.

When a surrounding acoustic pressure is applied, a pressure difference appears in the back chamber and causes the diaphragm to deflect. The deflection of the diaphragm is measured as a change of electrical capacitance between the diaphragm and the perforated backplate. Then the changes of electrical capacitance are the received signals of the microphone. Assume that backplate is rigid and deflection is small so that the linear model applies. Electro-acoustical analogy is adopted for predicting the linear dynamic behavior of the MEMS condenser microphone. The equivalent circuits of the microphone are shown in Figure 2, wherein the acoustical, mechanical and electrical domains are coupled through ideal transformers. The acoustical system consists of resistance and mass due to radiation from the diaphragm, the air film in the air gap and the air in the acoustic holes in the backplate, and compliance of the air in the back chamber. The mechanical system consists of compliance and mass of the diaphragm and the backplate. Figure 3 shows the MATLAB Graphical User Interface(GUI). The MATLAB GUI is utilized to change the parameters of the linear dynamic model of the MEMS condenser microphone. Then the frequency response of the MEMS condenser microphone can be acquired. After linear dynamic model, the optimal design of the performances for the MEMS condenser microphone is following.

B. Optimal Microphone Design

In order to improve the performance of a condenser microphone, Taguchi method and GA is utilized to optimize the main parameters of diaphragm length, diaphragm thickness and air gap height. Then the performance indices of sensitivity and bandwidth can be improved.

1. Taguchi Method

Taguchi method is made by Dr. Genichi Taguchi after the end of the Second

World War and is an experimental design procedure for examining multi-factors in a design problem using a minimum number of observations.[14,15] In fact, it is the most powerful method available to reduce product cost, improve quality, and simultaneously reduce development interval. A general Taguchi procedure provides three kinds of functions: system design, parameter design and tolerance design. For our problem, we focus primarily on parameter design. Parameter design uses the orthogonal array to compute the optimal solution. The orthogonal array is composed of factors and levels and it can reduce the computation load. Factor is the parameter that needs to be optimized. Level is the value of each factor. The simplified procedures of parameter design by Taguchi method have 5 steps. First, choose the factors that need to be optimized. Second, choose the fitness function. Third, choose the levels of every factor. Fourth, use the orthogonal array to compute the optimal level of every factor. Fifth, repeat third and fourth step until find the optimal solutions.
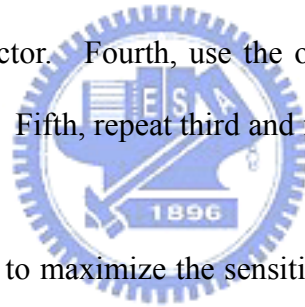
In our problem, we wish to maximize the sensitivity and bandwidth of a MEMS condenser microphone. First, parameters that need to be optimized are diaphragm length(DL), diaphragm thickness(DT) and air gap height(AH). Other parameters in the MEMS microphone model do not greatly affect the performances. Second, fitness function f is chosen

$$f = \frac{1}{0.4 \times S \times 10^3 + BW \times 10^{-4}} \tag{1}$$

where S and BW are the sensitivity and bandwidth of the MEMS condenser microphone. Third, three levels of every factor are chosen. Fourth, a $L_9 (3^3)$ orthogonal array is chosen and shown in Table I, where '1' represents the first level of a factor, '2' represents the second level of a factor, '3' represents the third level of a factor. The notation $L_9 (3^3)$ means that the experimental design requires 9

observations, 3 factors and each factor has 3 levels. The larger values of fitness function have better performances. Taguchi method is done by four times experiments to find the optimal solutions. Tables II-V are the results of four times experiments. The optimal solutions of diaphragm length, diaphragm thickness and air gap height by Taguchi method are 1.1 mm, 1.6 μm and 2.2 μm. The sensitivity and bandwidth of the MEMS condenser microphone are 7.7 mV/Pa and 18.38 kHz. Sensitivity is improved by Taguchi method but bandwidth is somewhat degraded.

2. Genetic Algorithm

Genetic Algorithm(GA) is a search algorithm based on the natural selection, genetics and evolution.[16,17] It is also a stochastic and powerful method for solving optimization problems. GA was first published by Holland in 1962. It has proven to be efficient in many areas such as function optimization and image processing. It is composed of the procedures of encoding, decoding, fitness function, reproduction, crossover and mutation. The simplified procedures of optimization by GA have 5 steps. First, choose the parameters needed to be optimized. Second, choose the fitness function. Third, choose the upper limit, lower limit and resolution of every parameter. Fourth, do encoding, decoding, reproduction, crossover and mutation for every generation. Fifth, repeat fourth step until find the optimal solutions. In the following, GA is introduced in details.

For encoding and decoding, all processes of GA are operated with binary strings encoded from original parameters. The resolution of a parameter space is dependent on the amount of bits per string and searching domain of the parameter. The resolution can be obtained as follows

$$R_{x_i} = \frac{x_i^U - x_i^L}{2^{l_i} - 1} \tag{2}$$

where x_i^U and x_i^L are the upper and lower limits of the parameter. For example,

the parameter range is $0 < x < 5.5$ cm. Then the number of bits needed for x is assumed 8. Therefore, the desired resolution is $R_x = 0.0216$. If parameter x is 2.592, then the value after encoding is [0 1 1 1 1 0 0 0]. Fitness function is the performance index for GA and chosen same as Taguchi method in Eq. (1).

Reproduction is a rule of survival of fittest. The larger value of fitness function of the factors in every generation, the higher reproduction probability can be generated. The reproduction probability S_i is shown as

$$S_i = \frac{f_{c_i}}{\sum_{k=1}^{P_l} f_c} \tag{3}$$

where f_{c_i} is the fitness function of the i factor, P_l is the population size. For instance, there are four factors, $c_1 \sim c_4$, in the zero generation with fitness functions 40, 5, 72 and 54. By substituting those values into Eq. (3), the reproduction probabilities are $S_1 = 0.2239$, $S_2 = 0.0292$, $S_3 = 0.4211$ and $S_4 = 0.3158$, respectively. The factors c_3 and c_4 have the large probability to reproduce in the next generation. Next, in the present population, four random numbers between 0 and 1 are generated. For example, if the four random numbers are 0.3675, 0.8719, 0.7697 and 0.1437, the factors of next generation will be c_3, c_4, c_4 and c_1, respectively.

Crossover is exchange the information of the factors. First, the crossover ratio C_r is defined (in general, $0.8 \leq C_r \leq 1$ and we choose $C_r = 0.85$) and two factors in the present population are selected randomly. Second, a random number is selected to check whether crossover or not. If the random number is larger than crossover ratio, then crossover needs to be done. The factors codes after the chosen point are interchanged. If the random number is less than crossover ratio, then the original factors is still the next generation. To illustrate, there are two factors c_1 and c_2

with the chosen point at third bit: $c_1 = 011_\Delta 0111$, $c_2 = 101_\Delta 1001$. With crossover, two new factors are generated: $\widetilde{c}_1 = 011_\Delta 1001$, $\widetilde{c}_2 = 101_\Delta 0111$.

However, the optimal solutions in GA may be local optimal solutions. To overcome this problem, mutation is introduced into the GA procedure. Let the mutation ratio be M_r (in general, $0 \le M_r \le 0.01$ and we choose $M_r = 0.01$). Choose a random number to determine whether mutation or not and choose mutation point. .For example, a factor c_1 with the mutation point at third bit is $c_1 = 10\,\underset{\Delta}{1}\,10100$. After mutation, the factor becomes $\widetilde{c}_1 = 10\,\underset{\Delta}{0}\,10100$. The mutation rate can not be chosen too large, or the search will be unhelpful.

In our problem, the search range of diaphragm length, diaphragm thickness and air gap height are 7-1.5 mm, 0.8-2.5 μm and 2.0-4.0 μm. The number of bits is 12. The crossover rate is 0.85. The mutation rate is 0.01. The computed generation is 150. The optimized values of diaphragm length, diaphragm thickness and air gap height are 1.0252 mm, 2.1362 μm and 2.1084 μm. In order to realize in MEMS scale, the parameters need to be chosen to minimum in MEMS process technology. Therefore, the parameters are chosen as 1.02 mm, 2.1 μm and 2.1 μm. The performances computed by GA with fitness function in Eq. (1) are 5.4 mV/Pa and 24.6 kHz. The sensitivity is too small. The reason maybe the fitness function is too simplified, because the relationship of sensitivity and bandwidth is not linear. The exact fitness function is not easy to get because there are three parameters that need to be optimized.

Then, fix the bandwidth over a range. The fitness function is sensitivity only. Figure 4 shows the optimized values of diaphragm length and diaphragm thickness using GA with constrained bandwidth at 30 kHz. Figure 5 shows the optimized values of air gap height using GA with constrained bandwidth at 30 kHz and the

frequency response.　The optimized sensitivity and bandwidth are 10.3 mV/Pa and 33.3 kHz.　The sensitivity is improved greatly.　Three optimized values are 1.24 mm, 0.8 μm and 3.4 μm.　The comparisons of frequency response of Taguchi method, GA, GA with constrained bandwidth at 16, 20, 30 and 40 kHz are shown in Figure 6. Table VI shows diaphragm length, diaphragm thickness, air gap height, sensitivity bandwidth of original, GA, Taguchi method, GA with constrained bandwidth at 16, 20, 30 and 40 kHz.　From the results, the performances of the MEMS condenser microphone can be improved greatly.

C. MEMS Diaphragm Process Design and Fabrication

　　The initial stress greatly affects the sensitivity of a MEMS condenser microphone.　In order to know the real value of initial stress in the diaphragm, we have to design and fabricate a MEMS diaphragm with Poly-Silicon.　The thickness of the diaphragm is 1 μm and the lengths of the diaphragm are designed for 1, 1.2, 1.4, 1.6, 1.8 and 2.0 mm.　Different lengths of the diaphragm can be caused different resonant frequencies.　Then, we can know the value of initial stress in the diaphragm to control the sensitivity of the diaphragm.　By the way, the resonant frequency of the diaphragm can be calculated by[13]

$$f_r = \frac{1}{2\pi\sqrt{M_D C_D}} \tag{4}$$

where M_D and C_D is mass and compliance of the diaphragm, respectively.　The imaginary part of the impedance determines the resonant frequency of the diaphragm. The real part of the impedance determines the damping effects of the diaphragm. From the frequency response of the dynamic model of the condenser microphone, the 3 dB down cut-off frequency is 42.5 kHz.　The real resonant frequency of the microphone computed by Eq. (4) is 105 kHz, not 42.5 kHz.　This is because that the

real part of acoustical impedance is large enough to oppress the resonant frequency. Therefore, the resonant frequency of the linear dynamic model of the MEMS condenser microphone can not be shown. The effects of real and imaginary parts of the impedance should be greatly realized.

The fabrication procedures of the MEMS diaphragm are shown in Figure 7. The diaphragm is square. The silicon wafer we used is N-type, <1 0 0> crystal structure, 525 μm thick, and resistivity is 1~10 Ωcm. The Si_3N_4 is deposited for 5000 Åm thick as resistant layer using LPCVD at 800℃. This layer is to resist the KOH etching and protect the bottom side of the diaphragm. The low-stress Poly-Silicon is deposited for 1 μm thick as diaphragm structure using LPCVD at 620 ℃. This layer is the main structure of the condenser microphone. The acoustic waves can be detected by the diaphragm. The Si_3N_4 is deposited for 5000 Åm as resistant layer using LPCVD at 800℃. This layer is also to resist the KOH etching and protect the top side of the diaphragm.

After all the layers are deposited, we should pattern the back chamber holes to realize the diaphragm structure. Then, RIE is used to pattern the back chamber holes of the bottom side of the wafer. The back chambers are formed by use of anisotropic etching in a KOH 30% solution at the temperature of 80℃. After etching, the Si_3N_4 layer is released by HF. Therefore, the diaphragm structure is realized. Figure 8 shows the result of the diaphragm anisotropic etching by KOH. We know that each diaphragm is etched quite square but the thickness is not very flat. The average thickness measured by surface profiler is 60 μm. Some diaphragms are etched broken and some still have 60 μm thick. The reasons are due to the etching rate of KOH is not uniform and the layers deposited are not very flat. In addition, the resistant layer of Si_3N_4 is too thin to resist the KOH. Therefore, some diaphragms are etched broken.

D. MEMS Microphone Process Design

The fabrication steps of a MEMS condenser microphone are shown in Figure 9. The silicon wafer we used is N-type, <1 0 0> crystal structure, 525 μm thick, and resistivity is 1~10 Ωcm. The SiO$_2$ is deposited for 2000 Åm thick as resistant layer to pattern the boron diffusion area as backplate using wet oxidation. Diffuse the boron in the silicon wafer and remove the SiO$_2$ using BOE. The Si$_3$N$_4$ is deposited for 2500 Åm as insulation layer to pattern the diaphragm area and backplate electrode area using LPCVD. The SiO$_2$ is deposited for 3 μm thick as sacrificial layer to define the air gap height using wet oxidation. Pattern the 3 μm thick SiO$_2$ for diaphragm area. The Poly-Silicon is deposited for 1 μm thick as diaphragm using LPCVD. Pattern the 1 μm thick Poly-Silicon for proper diaphragm area and show the backplate electrode contact pad. The SiO$_2$ is deposited for 1 μm thick as insulation layer using wet oxidation. Pattern the SiO$_2$ and Si$_3$N$_4$ to define the back chamber size. The back chamber holes are formed by use of anisotropic etching in a KOH 30% solution at the temperature of 80℃. Remove the 3 μm thick SiO$_2$ to release the air gap. Deposit Cr for 50 nm as adhesion layer and Au for 400 nm as electrode by evaporation. Then, pattern diaphragm contact pad and backplate contact pad and remove Cr in HNO$_3$+NH$_4$F and Au in KI.

III. MICROPHONE ARRAY SIGNAL PROCESSING

Array signal processing techniques are utilized for Direction of Arrival (DOA) estimation and beamforming. There is a one important thing we have to know. Avoid spatial aliasing or grating lobes, the spacing between each microphone must be less than half wavelength of the signals. In the following array signal processing, we assume that signals received at reference point are far-field and narrowband. Far-field assumes that the signals are located far enough away from the array that the

wavefronts impinging on the array can be modeled as plane waves. On the other hand, the effects of distance on field intensity can be neglected. Narrowband assumes that the incident signal that the Beamformer is trying to capture has a narrow bandwidth centered at a particular frequency. The uniform linear array (ULA) is shown in Figure 10 and the array model can be constructed.

A. Direction of Arrival (DOA) estimation and Beamforming

DOA estimation and beamforming of speech signals using a set of spatially separated microphones in an array has many practical applications such as video-conferencing system and hearing-aids. The locations of sound source can be found out by the DOA estimation. The beam of the array can be moved to the sound source to receive the signals is a beamforming technique. Two types of linear arrays with equally spaced microphones are discussed. Figure 11 shows the configuration of the broadside and endfire arrays. A "broadside" array has its microphones along a line perpendicular to the sound direction and an "endfire" array has its microphones along a line collinear with the sound direction.

1. The Delay-Sum Method

The spacing between adjacent microphones is d. Assume that the signal $r(t)$ at a reference point is a narrowband with center frequency ω_c:

$$r(t) = s(t)e^{j\omega_c t} \tag{5}$$

where $s(t)$ is the phasor of $r(t)$. The signal received at the mth array element located at $\vec{x}_m$ is denoted as $x_m(t)$, and let $\vec{r}$ be the unit vector pointing to the sound source direction. If the speed of sound is c, the signal $x_m(t)$ can be written as

$$x_m(t) = r(t + \frac{\vec{x}_m \cdot \vec{r}}{c}) + n_m(t) = s(t + \frac{\vec{x}_m \cdot \vec{r}}{c})e^{j\omega_c \frac{\vec{x}_m \cdot \vec{r}}{c}}e^{j\omega_c t} + n_m(t) \tag{6}$$

where $n_m(t)$ is the noise signal of the mth component in the array. In general, $s(t + \frac{\vec{x}_m \cdot \vec{r}}{c}) \approx s(t)$ for far field approximation. For M sensor signals $x_1(t), \cdots, x_M(t)$, the data vector can be formed as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} e^{j\omega_c \frac{\vec{x}_1 \cdot \vec{r}}{c}} \\ \vdots \\ e^{j\omega_c \frac{\vec{x}_M \cdot \vec{r}}{c}} \end{bmatrix} s(t) e^{j\omega_c t} + \begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix} = \mathbf{a}(\vec{r}) r(t) + \mathbf{n}(t) \tag{7}$$

where $\mathbf{a}(\vec{r})$ is called the array manifold vector. The unit vector $\vec{r}$ for a sound source at the θ direction is given by

$$\vec{r} = (\sin\theta, \cos\theta) \tag{8}$$

The position vector of the mth element can be expressed as

$$\vec{x}_m = ((i-1)d, 0) \tag{9}$$

Then the inner product of the position vector and the unit vector is obtained

$$\vec{x}_m \cdot \vec{r} = (i-1)d\sin\theta \tag{10}$$

The array manifold vector $\mathbf{a}(\vec{r})$ can be rewritten from Eq. (10)

$$\mathbf{a}(\omega_c, \theta) = \begin{bmatrix} 1 & e^{j\omega_c \frac{d\sin\theta}{c}} & \cdots & e^{j\omega_c \frac{(M-1)d\sin\theta}{c}} \end{bmatrix}^T \tag{11}$$

Extension from the narrowband formulation to the broadband formulation is straightforward. The center frequency ω_c is replaced by ω, where ω means a broadband frequency variable. The beamformer output $y(t)$ is the weighted sum of the delayed input signals $x_m(t), m = 1 \cdots M$,

$$y(t) = \sum_{m=1}^{M} \sum_{k=0}^{N} w_{mk} x_m(t - kT) \tag{12}$$

Equation (12) can be rewritten in the frequency domain for a particular direction θ as

$$y(\omega, \theta) = \mathbf{h}(e^{j\omega}) \mathbf{x}(\omega, \theta) = \mathbf{h}(e^{j\omega})(\mathbf{a}(\omega, \theta) r(\omega) + \mathbf{n}(\omega)) \tag{13}$$

where the manifold vector is given by

$$\mathbf{a}(\omega,\theta) = \left[1 \; e^{j\omega\frac{d\sin\theta}{c}} \; \cdots \; e^{j\omega\frac{(M-1)d\sin\theta}{c}} \right]^{T} \tag{14}$$

and $\mathbf{h}(e^{j\omega}) = [h_1(e^{j\omega}) \; h_2(e^{j\omega}) \cdots h_M(e^{j\omega})]$ is consisted of Discrete Time Fourier Transforms (DTFT) of each tapped-delay line channel. The frequency response is given by

$$h_m(e^{j\omega}) = \sum_{k=0}^{N} w_{mk} e^{-j\omega kT}, \quad m = 1 \cdots M \tag{15}$$

N and T are the filter order and the sampling period, respectively. The dimensions of $\mathbf{h}(e^{j\omega})$ and $\mathbf{a}(\omega,\theta)$ are $1 \times M$ and $M \times 1$, respectively.

In our problem, the delay-sum algorithm is tried to find $\mathbf{h}(e^{j\omega})$ that maximizes signal-to-noise ratio gain (SNRG), i.e.

$$\max_{\mathbf{h}(\omega)} \frac{\mathbf{h}(e^{j\omega})\mathbf{a}(\theta)\mathbf{a}^{H}(\theta)\mathbf{h}^{H}(e^{j\omega})}{\mathbf{h}(e^{j\omega})\mathbf{h}^{H}(e^{j\omega})} \tag{16}$$

which is equivalent to

$$\max_{\mathbf{h}(\omega)} \mathbf{h}(e^{j\omega})\mathbf{a}(\theta)\mathbf{a}^{H}(\theta)\mathbf{h}^{H}(e^{j\omega}) \quad \text{suject to } \mathbf{h}(e^{j\omega})\mathbf{h}^{H}(e^{j\omega}) = 1 \tag{17}$$

This problem can be solved by Lagrange multiplier method, the solution is obtained as below

$$\mathbf{h}(e^{j\omega}) = \mathbf{a}^{H}(\omega,\theta) \tag{18}$$

Equation (17) explains that each channel filter equals to the conjugate of each component in the manifold vector.

$$h_m(e^{j\omega}) = e^{-j\omega\frac{(m-1)d\sin\theta}{c}} = e^{-j\omega\tau_m}, \quad m = 1 \cdots M, \tag{19}$$

where $\tau_m = \frac{(m-1)d\sin\theta}{c}$ is the delay of each channel according to the difference between mth sensor and reference point. However, the delay usually is not an integer in the digital processing. There are many ways to deal with these fractional delay problems. The simplest approach is Lagrange interpolation method. Firstly we divide τ_m by sampling period T to acquire the fractional delay Ψ_m. The

14

delay is separated into two parts

$$\frac{\tau_m}{T} = \Psi_m = D_m + e_m \tag{20}$$

where D_m and e_m are the integer and fractional component of Ψ_m, respectively. For simplicity, the Finite Impulse Response (FIR) filter coefficients are obtained from Eq. (21) to realize the Lagrange interpolation.

$$w_{mk} = \prod_{\substack{l=0 \\ l \neq k}}^{N} \frac{e_m - l}{k - l}, \quad k = 0,1,2 \ldots N \tag{21}$$

The coefficients for the Lagrange filters of order $N = 1, 2$ are given in the Table VII. The case $N = 1$ corresponds to linear interpolation between two samples. Once we obtain the filters, the output signals $y(\omega, \theta)$ can be calculated from Eq. (13). The square of $y(\omega, \theta)$ is called the spatial power spectrum, which is given by

$$S(\theta) = \left| y(\omega, \theta) \right|^2 = \left| \mathbf{h}(e^{j\omega}) \mathbf{x}(\omega, \theta) \mathbf{x}^H(\omega, \theta) \mathbf{h}^H(e^{j\omega}) \right| \tag{22}$$

The maximum magnitude of the spatial power spectrum is the direction of the sound source.

2. The Superdirective Method

The superdirective method is presented.[23-28] From the delay-sum method, we can know that signals which got from microphones are

$$\mathbf{x}(e^{j\omega}) = \mathbf{s}(e^{j\omega})\mathbf{d}_s + \mathbf{v}(e^{j\omega}) \tag{23}$$

where $\mathbf{d}_s$ is the look direction vector which depends on the actual geometry of the array and the direction of the sound source signal. Then, the output is

$$y = \mathbf{w}^H \mathbf{x} \tag{24}$$

where $\mathbf{w}$ denotes the frequency-domain coefficients of the beamformer and the operator ^{H} denotes a complex conjugated transposition (Hermitian operator). The array gain is the measure which shows the improvement of the SNR between one sensor and the output of the whole array. Therefore,

$$G = \frac{SNR_{Array}}{SNR_{Sensor}} \tag{25}$$

The SNR of one sensor is given by the ratio of the power spectral densities (PSD) of the signal ϕ_{ss} and the average noise $\phi_{V_a V_a}$. The SNR at the output can be computed by deriving the PSD of the output signal.

$$\boldsymbol{\varphi}_{YY} = \mathbf{w}^H \boldsymbol{\varphi}_{XX} \mathbf{w} \tag{26}$$

where $\boldsymbol{\varphi}_{XX} = \begin{pmatrix} \phi_{X_0 X_0} & \phi_{X_0 X_1} & \cdots & \phi_{X_0 X_{N-1}} \\ \phi_{X_1 X_0} & \phi_{X_1 X_1} & \cdots & \phi_{X_1 X_{N-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{X_{N-1} X_0} & \phi_{X_{N-1} X_1} & \cdots & \phi_{X_{N-1} X_{N-1}} \end{pmatrix}$ is a power spectral density matrix of

the array input signals. When the desired signal is present only, the output is

$$\boldsymbol{\varphi}_{YY}\big|_{Signal} = \boldsymbol{\varphi}_{SS} \left| \mathbf{w}^H \mathbf{d}_s \right|^2 \tag{27}$$

and for noise only case the output is

$$\boldsymbol{\varphi}_{YY}\big|_{Noise} = \phi_{V_a V_a} \mathbf{w}^H \boldsymbol{\varphi}_{VV} \mathbf{w} \tag{28}$$

where $\boldsymbol{\varphi}_{VV}$ is a normalized cross power spectral density matrix of the noise. Therefore, array gain in Eq. (25) can be rewritten as

$$G = \frac{\left| \mathbf{w}^H \mathbf{d}_s \right|^2}{\mathbf{w}^H \boldsymbol{\varphi}_{VV} \mathbf{w}} \tag{29}$$

Assuming a homogeneous noise field can be expressed in terms of the coherence matrix

$$\boldsymbol{\Gamma}_{VV} = \begin{pmatrix} 1 & \Gamma_{V_0 V_1} & \Gamma_{V_0 V_2} & \cdots & \Gamma_{V_0 V_{N-1}} \\ \Gamma_{V_1 V_0} & 1 & \Gamma_{V_1 V_2} & \cdots & \Gamma_{V_1 V_{N-1}} \\ \Gamma_{V_2 V_0} & \Gamma_{V_2 V_1} & 1 & \cdots & \Gamma_{V_2 V_{N-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma_{V_{N-1} V_0} & \Gamma_{V_{N-1} V_1} & \Gamma_{V_{N-1} V_2} & \cdots & 1 \end{pmatrix} \tag{30}$$

where $\Gamma_{V_n V_m}(e^{jw}) = \dfrac{\phi_{V_n V_m}(e^{jw})}{\sqrt{\Gamma_{V_n V_n}(e^{jw}) \Gamma_{V_m V_m}(e^{jw})}}$ is the coherence function. Thus, array

gain can be rewritten again as

$$G = \frac{\left|\mathbf{w}^H \mathbf{d}_s\right|^2}{\mathbf{w}^H \mathbf{\Gamma}_{VV} \mathbf{w}} \tag{31}$$

This representation allows for different noise fields and they can be expressed by their coherence function.

A common quantity to evaluate beamformers is the directivity index (DI) which describes the ability of the array to suppress a diffuse noise field. Therefore, we can compute the DI by using the coherence function of a diffuse noise field

$$\left.\Gamma_{V_n V_m}\left(e^{jw}\right)\right|_{Diffuse} = \frac{\sin\left[k(n-m)d\right]}{k(n-m)d} \tag{32}$$

Then, directivity index (DI) is

$$DI = 10\log_{10}\left(\frac{\left|\mathbf{w}^H \mathbf{d}_s\right|^2}{\mathbf{w}^H \left.\mathbf{\Gamma}_{VV}\right|_{Diffuse} \mathbf{w}}\right) \tag{33}$$

The ability of the array to suppress spatially uncorrelated noise, which can be caused by self-noises of the sensors. Inserting the coherence matrix for this noise field

$$\left.\mathbf{\Gamma}_{VV}\right|_{uncorr} = \mathbf{I} \tag{34}$$

into Eg. (30) results in the white noise gain (WNG)

$$WNG = \frac{\left|\mathbf{w}^H \mathbf{d}_s\right|^2}{\mathbf{w}^H \mathbf{w}} \tag{35}$$

On a logarithmic scale positive values represent an attenuation of uncorrelated noise, whereas negative values show an amplification.

In the real implementation of superdirective beamforming, large microphone weights result in array instability and sensitivity to uncorrelated noise. The noise sensitivity (Ψ) [26]

$$\Psi = \frac{\mathbf{w}^H \mathbf{w}}{\mathbf{w}^T \mathbf{d}_s \mathbf{d}_s^H \mathbf{w}^*} \tag{36}$$

can be implied to illustrate the uncorrelated relationship between signals and noises. When $\mathbf{w}$ is determined by Eq. (38), $\mathbf{w}^T \mathbf{d}_s = 1$, and Ψ is simply $\mathbf{w}^H \mathbf{w} = \Sigma_i |w_i|^2$. Insure that the array has unity gain in the sound source, $\mathbf{w}$ must be normalized to the sound source. Figure 12 shows the relationship between noise sensitivity and DI in broadside and endfire arrays. From the results, high noise sensitivity implies the larger microphone weights. In low frequencies, the optimal weights used for implementations are larger than high ones. In order to achieve the same DI, microphone weights utilized in low frequency are larger than high ones. Therefore, the DI in low frequency is hard to improve, unless the large weights are utilized.

In order to design optimal beamformers, we have to minimize the power of the output signal of the array. Avoiding the trivial solution $\mathbf{w} = 0$, the minimization is constrained to give an undistorted signal response in the desired look direction. Therefore, the following constrained minimization problem has to be solved:

$$\min_{\mathbf{w}} \mathbf{w}^H \boldsymbol{\varphi}_{XX} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{d}_s = 1 \tag{37}$$

Since we are only interested in the optimal suppression of the noise, and we assume a perfect correspondence between the direction of the desired signal and the look direction of the array, only the noise PSD-matrix $\boldsymbol{\varphi}_{VV}$ is used. The well-know solution for Eq. (37) is called the Minimum Variance Distortionless Response (MVDR) beamformer.[23] It is given by

$$\mathbf{w} = \frac{\boldsymbol{\varphi}_{VV}^{-1} \mathbf{d}_s}{\mathbf{d}_s^H \boldsymbol{\varphi}_{VV}^{-1} \mathbf{d}_s} \tag{38}$$

and can be derived by using the Lagrange-multiplier.[29] Assuming a homogeneous noise field the solution is a function of the coherence matrix:

$$\mathbf{w} = \frac{\boldsymbol{\Gamma}_{VV}^{-1} \mathbf{d}_s}{\mathbf{d}_s^H \boldsymbol{\Gamma}_{VV}^{-1} \mathbf{d}_s} \tag{39}$$

Eq. (36) or Eq. (37) can be interpreted as a spatial decorrelation process followed by a matched filter for the desired signal. The normalization in the denominator leads to unity signal response for the look direction.

The well-known delay-sum beamformer is included for comparison purposes. It is an "optimal" beamformer for optimizing the WNG. We can derive the coefficients from Eq. (37) by inserting the coherence matrix for spatial uncorrelated noise

$$\mathbf{w} = \frac{1}{N}\mathbf{d}_s \tag{40}$$

The WNG is optimal in this case and reaches N.

The method uses a same added scalar ε to the main diagonal of the normalized PSD or coherence matrix:

$$\mathbf{w}\big|_{\text{Regularized}} = \frac{(\mathbf{\Gamma}_{VV} + \varepsilon\mathbf{I})^{-1}\mathbf{d}_s}{\mathbf{d}_s^{\,H}(\mathbf{\Gamma}_{VV} + \varepsilon\mathbf{I})^{-1}\mathbf{d}_s} \tag{41}$$

The factor ε can vary from zero to infinity, which results in the unconstrained superdirective or the delay-sum respectively. The results between delay-sum and superdirective in broadside and endfire array are illustrated in the following.

Figure 13 shows the effects of changing factor ε in 0, 0.1, 0.01 and 0.001 for DI in the broadside and endfire arrays. From the results, we know that the DI in the endfire array is larger than the broadside array. In the broadside and endfire arrays, the DI is independent of ε above the frequencies of 2 and 3 kHz. Figure 14 shows the effects of changing factor ε in 0, 0.1, 0.01 and 0.001 for WNG in the broadside array and endfire array. From the results, we know that the WNG in the broadside array is larger than the endfire array. In the broadside and endfire arrays, the WNG is independent of ε above the frequencies of 2 and 3 kHz. Figure 15 shows the effects of changing factor ε in 0, 0.1, 0.01 and 0.001 for optimum weight in the broadside and endfire arrays. From the results, we know that the optimum weights

in the broadside array are larger than the endfire array. It means that the larger values of optimum weight, the harder implementations of FIR filters. In the broadside and endfire arrays, the optimum is independent of ε above the frequencies of 2 and 3 kHz. The optimal weights can not be too large, or it is hard to be implemented. The optimal weights should be small enough to implement in FIR filters. Around the effects of ε, the best values of ε is chosen as 0.01.

After the proper ε is chosen, then the comparisons of delay-sum and superdirective in DI, WNG, and optimal weights in broadside array are illustrated. Figure 16 shows the comparisons of DI of delay-sum and superdirective with ε chosen as 0.01 in the broadside and endfire arrays. We know that the DI with the superdirective method is improved greatly both in the broadside and endfire arrays. Figure 17 shows the comparisons of WNG of delay-sum and superdirective with ε chosen as 0.01 in the broadside and endfire arrays. We know that the WNG with the delay-sum method is better than the superdirective method both in the broadside and endfire arrays. Figure 18 shows the comparisons of optimum weights of delay-sum and superdirective with ε chosen as 0.01 in the broadside and endfire arrays. From the results, we know that the optimum weights with the superdirective method are larger than the delay-sum method at low frequency both in the broadside and endfire arrays. Because the WNG is small at low frequency, the optimum weights have to compensate the weights at low frequency. The weights of superdirective method can be accepted to implement as FIR filters.

In addition, the contour plots of delay-sum and superdirective methods compared in broadside array is shown in Figure 19. We know that the directivity of superdirective method in low frequency is greatly improved. It means that the SNRs of superdirective method are larger than delay-sum method. The performances of microphone can be improved by superdirective beamforming. Figure 20 shows the

contour plots of delay-sum and superdirective methods compared in endfire array. The results are still the same with the broadside array. To compare the difference of the broadside and endfire array with the performance indices indicated above, the directivity of the endfire array is better than the broadside array, but beamwidth shown in contour plots is wider than the broadside array. The reason is that the directivity is calculated from $0^{o} \sim 360^{o}$ and the directivity of the broadside array is symmetrical at 0° and 180° and shown up from $0^{o} \sim 360^{o}$. But, the directivity of endfire array is not symmetrical and shown up just from $-90^{o} \sim 90^{o}$. That is why the differences of directivity and beamwidth in the broadside and endfire array.

Once the frequency response of the superdirective weights is obtained, the inverse Fourier transform is utilized to acquire the impulse response of the superdirective FIR filters. If the P_{w} frequencies are acquired, the discrete frequency response can be obtained as

$$H_{m}(l) = w_{m}^{*}(l), \qquad l = 1, \cdots, P_{w} \tag{42}$$

where $H_{m}(l)$ is the discrete frequency response at mth channel and $w_{m}^{*}(l)$ is the discrete frequency response of superdirective algorithm at mth channel. In order to obtain the impulse response with real coefficients, mirror $H_{m}(l)$ with conjugate operation to obtain the symmetric frequency response. Then the frequency response becomes

$$\begin{cases} H_{m}(0) = |H_{m}(1)| = |H_{m}(-1)| \\ H_{m}(l) = w_{m}^{*}(l), \quad l = 1, \cdots, P_{w} \\ H_{m}(l) = w_{m}(l), \quad l = -1, \cdots, -(P_{w}-1) \end{cases} \tag{43}$$

Subsequently, Eq. (43) is equivalent to shift $l = -1, \cdots, -(P_{w}-1)$ to $l = P_{w}+1, \cdots, 2P_{w}-1$. The discrete frequency response becomes

$$\begin{cases} H_m(0) = |H_m(1)| \\ H_m(l) = w_m^*(l), \quad l = 1, \cdots, P_w \\ H_m(l) = w_m(l), \quad l = (P_w + 1), \cdots, (2P_w - 1) \end{cases} \quad (44)$$

Finally, the impulse response can be obtained by utilizing inverse Fourier transform at each channel.

$$h_m(k) = \frac{1}{2P_w} \sum_{l=0}^{2P_w - 1} H_m(l) W_{2P_w}^{-lk}, \quad k = 1 \cdots (2P_w - 1) \quad (45)$$

where $W_{2P_w} = e^{-j(\frac{2\pi}{2P_w})} = e^{-j\frac{\pi}{P_w}}$. In general, $h_m(k)$ may be non-causal. The half of $h_m(k)$ can be moved to center as symmetrical about the center. Figure 21 shows the non-causal impulse response of superdirective FIR filters of order 64. The causal impulse response of superdirective FIR filters of order 64 is shown in Figure 22. Therefore, the digital superdirective FIR filters are implemented. From the results, the FIR filters are symmetrical. The superdirective FIR filters of Microphone 1 are the same with Microphone 4. The superdirective FIR filters of Microphone 2 are the same with Microphone 3. Besides, the relationship between Microphone 1 and Microphone 2 is like 'differential microphones'. Figure 23-26 show the simulations and measurements of the DOA estimation of delay-sum and superdirective methods with a single frequency at 500, 1000, 2000 and 4000 Hz, respectively. The sound source is indicated at 10°. From the results, the DOA estimation angle of superdirective method is more precise than delay-sum at 500, 1000 and 2000 Hz. The DOA estimation is almost the same at 4000 Hz or other high frequencies. However, the sidelobes of superdirective method are larger than delay-sum. Figure 27 shows the DOA estimation simulated by delay-sum and superdirective methods with the sampling rate 16000 Hz at a 12 cm aperture in 8 kHz broadband white noise. Obviously, the beamwidth of superdirective method is smaller than the delay-sum method and the directivity is also better than delay-sum method. However, the

sidelobes of superdirective method are larger than delay-sum.   We can know that the relationship between beamwidth and sidelobes is trade-off.   The DOA estimation of superdirective method is more precise than delay-sum method.

B. Directivity Analysis

Directivity is a quite important performance index for microphone arrays and the aperture size of microphone arrays dominates the value of directivity.   In the following simulations and measurements of directivity, the aperture sizes of microphone array are 15 mm and 12 cm and the sound source is located at 0°.   Four microphones are used to line in uniform linear array.   Figure 28 shows the configuration of a MEMS microphone array with same spacing 5 mm and the aperture size is 15 mm.   Figure 29 shows the configuration of a conventional microphone array with same spacing 4 cm and the aperture size is 12 cm.

1. Relationship between Beamwidth and Frequency

The aperture size of the line source is 15 mm in MEMS scale and 12 cm in normal scale, respectively.   Let the decay of the radiation pattern are 3, 6 and 12 dB, respectively.   Figure 30 shows the far-field radiation pattern of a continuous line source.   The beamwidth can be calculated by[13]

$$\theta = \sin^{-1}(\frac{c \cdot u}{\pi \cdot l} \cdot \frac{1}{f}) \tag{46}$$

where c = 343 m/s, u = 1.4, 1.9, 2.5 and l = 15 mm and 12 cm.   The relationship between beamwidth and frequency of a continuous line source with aperture size 15 mm and 12 cm is shown in Figure 31.   The beamwidth at line source aperture of 15 mm is greater than that at line source aperture of 12 cm.   It means that if aperture size increases, then beamwidth decreases.   We know that beamwidth decreases when frequency increases.   On the other hand, the directivity increases as frequency

increases.　The decrease of beamwidth means the increase of directivity.

2. One Microphone

The simulations and measurements of directivity of one microphone are shown in Figure 32.　The simulations of directivity are omni-directional.　The measurements of directivity at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz are almost the same with the simulation of directivity.　In 6000 Hz, the measurement of directivity is better than the simulation of directivity.　The beamwidth in 3 dB down is almost $-60^{o} \sim 60^{o}$.　This is probably due to microphone geometric structure to increase the directivity in 6000 Hz.

3. Four Microphones

The simulations and measurements of directivity of four MEMS microphones in a 15 mm aperture are shown in Figure 33.　The simulations of directivity are omni-directional because the aperture size is too small to generate directivity.　The measurement of directivity at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz are almost the same with the simulation of directivity.　In 6000 Hz, the measurement of directivity is better than the simulation of directivity.　The beamwidth in 3 dB down is almost $-70^{o} \sim 50^{o}$.　This is probably due to microphone geometric structure to increase the directivity in 6000 Hz.

Figure 34 shows the simulations and measurements of directivity of four conventional microphones in a 12 cm aperture.　The simulations of directivity at 500 Hz and 1000 Hz are omni-directional but the beamwidth in 3 dB down at 2000 Hz and 4000 Hz are $-30^{o} \sim 30^{o}$ and $-10^{o} \sim 10^{o}$, respectively. Therefore, directivity is apparent at 2000 Hz and 4000 Hz.　The measurement of directivity at 500 Hz and 1000 Hz are somewhat better than the simulation of directivity but the improvement is not much enough.　Then we can realize that a 12 cm aperture still can not improve directivity at low frequency.　In 2000 Hz and 4000 Hz, the measurements of

directivity quite match the simulations of directivity. The beamwidth in 3 dB down at 2000 Hz and 4000 Hz are $-20^{\circ} \sim 20^{\circ}$ and $-20^{\circ} \sim 10^{\circ}$, respectively.

4. 4×4 Sub-array Microphones

Figure 35 shows a 4×4 sub-array microphones scheme. The first product theorem is used for sub-array microphones.[30] The spacing of each MEMS microphone is 5 mm. MEMS microphone array arranges 4 sub-array microphones and the aperture size is 15 mm. The spacing of the array microphone is 4 cm and the total 4×4 sub-array aperture is 12 cm. The directivity of 4×4 sub-array in 12 cm aperture is shown in Figure 36(a). The beamwidth in 3 dB down at 2000 Hz and 4000 Hz are $-30^{\circ} \sim 30^{\circ}$ and $-10^{\circ} \sim 10^{\circ}$, respectively. The simulation is much like the directivity of four microphones with a 12 cm aperture, because the aperture size of sub-array is too small to generate directivity. Another case, the spacing of the array microphone is 5 cm and the total 4×4 sub-array aperture is 15 cm. The directivity of 4×4 sub-array with a 15 cm aperture is shown in Figure 36(b). The beamwidth in 3 dB down at 2000 Hz and 4000 Hz are $-25^{\circ} \sim 25^{\circ}$ and $-7^{\circ} \sim 7^{\circ}$, respectively. From the comparisons, the directivity of 4×4 sub-array with a 15 cm aperture is better than a 12 cm aperture. Therefore, the directivity is determined by the aperture size of the microphone array.

5. Signal-to-Noise Ratio (SNR) Gain

The SNR of one, four and four with array filters measured by MEMS microphones in a 15 mm aperture and conventional microphones in a 12 cm aperture are listed in Table VIII. We know that SNRs can be improved by the microphone arrays. The SNRs of MEMS microphones are higher than conventional ones, because the sensitivity of MEMS microphone is higher than conventional ones. The SNR gains improved by a 15 mm and a 12 cm aperture microphone arrays are 8.6 and 11.9 dB, respectively.

IV. SPATIAL SOUND RECONSTRUCTION WITH HRTF

The microphone arrays for hearing aids can enhance SNR, increase directivity, and reduce the effects of reverberation and acoustic feedback.   Those hearing-impaired people can hear speech and music more clear using array techniques. However, they still can not recognize the sound location.   Therefore, HRTF is added to microphone array signal processing for hearing aids.   Then, the sound received by microphone array with HRTF can be heard as 3D spatial sound field.   The array signal processing with HRTF can reconstruct 3D spatial sound, and let those people who hearing-impaired can recognize the sound location.   When watching TV or movies, the performances of 3D spatial sound can be heard.

A. Head Related Transfer Function

The 3D spatial sound can be reconstructed by the Head-Related Transfer Function (HRTF) which measured by MIT Media Lab.[21]   The measurements consist of the left and right ear impulse responses from a Realistic Optimus Pro 7 loudspeaker mounted 1.4 meters from the Knowles Electronics Manikin for Acoustic Research (KEMAR) dummy head.   Specifically, 710 different positions were measured at a sampling rate of 44.1 kHz including azimuth and elevation angles.   The inverse Fourier Transform of HRTF is termed Head-Related Impulse Response (HRIR).   For example, Figure 37 shows the HRTF frequency and HRIR at elevation $=0^{\circ}$ and azimuth $=40^{\circ}$.   From the results, the signals received by the right ear are larger than left ear.   Then, it means that the sound location is in right side.   Therefore, the HRTFs around 710 different positions can be used to reconstruct the 3D spatial sound.

B. Focusing by Inverse-Filtering

The filters with three-dimensional locations can be calculated by sound

propagation transfer function and HRTF. Therefore, spatial sound can be reconstructed by the filter. **G** is the transfer function between a point sound source and MEMS microphone array. **F** is the transfer function between microphone array and two ears. **H** is the HRTF. The effect of **FG** is equal to **H**. **G** and **H** are known and **F** can be calculated by inverse-filtering.[31] In order to reconstruct spatial sound using HRTF, then
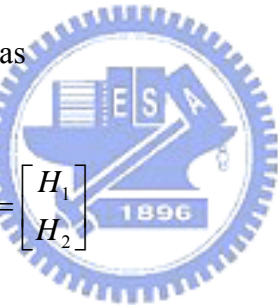
$$\mathbf{FG} = \mathbf{H} \tag{47}$$

To solve for *F*, then

$$\min\{\mathbf{FG} - \mathbf{H}\} \tag{48}$$

Therefore, Eq. (47) can be shown with M microphone sensor.

G is a M×1 matrix. **H** is a 2×1 matrix. So, **F** is a 2×M matrix. In this case, M=4. Then Eq. (47) is rewritten as

$$\begin{bmatrix} F_{11} & F_{12} & F_{13} & F_{14} \\ F_{21} & F_{22} & F_{23} & F_{24} \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \tag{49}$$

where $G_1 = \dfrac{1}{r_1}\exp(-jkr_1)$, $G_2 = \dfrac{1}{r_2}\exp(-jkr_2)$, $G_3 = \dfrac{1}{r_3}\exp(-jkr_3)$,

$G_4 = \dfrac{1}{r_4}\exp(-jkr_4)$ and r_1, r_2, r_3, r_4 are the distance from microphone to a point

sound source, respectively . $k = \dfrac{2\pi f}{c}$ is a wave number.

$$\begin{bmatrix} F_{11} & F_{12} & F_{13} & F_{14} \\ F_{21} & F_{22} & F_{23} & F_{24} \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \begin{bmatrix} G_1^+ & G_2^+ & G_3^+ & G_4^+ \end{bmatrix}$$
$$= \begin{bmatrix} H_1 G_1^+ & H_1 G_2^+ & H_1 G_3^+ & H_1 G_4^+ \\ H_2 G_1^+ & H_2 G_2^+ & H_2 G_3^+ & H_2 G_4^+ \end{bmatrix} \tag{50}$$

G can be solved by the pseudo inverse. Therefore, **F** can be calculated. Then, the filters between microphone array and two ears can be calculated. Row 1 of **F** is the left ear filters. Row 2 of **F** is the right ear filters. There are four filters in left

and right ear, respectively. Therefore, digital inverse filters are implemented in beamforming of sound pressure propagation with distance. Figure 38-41 shows the contour plots of beamforming techniques by inverse filters. Assume that the sound source is located at 60°. The microphone arrays consist of aperture size in 15 mm and 12 cm and frequencies at 500, 1000, 2000 and 4000 Hz. From the results, we know that the beamforming of sound source in large aperture size and at high frequency of the microphone array is better than small aperture and low frequency.

C. Experimental Verifications using TMS320C3X DSP

The system block diagram of 3D spatial sound reconstruction with HRTF implemented by TMS320C3X DSP is shown in Figure 42. It contains a sound source generator, a conventional microphone array, a preamplifier and two second-order low pass filter at 8 kHz, a computer with TMS320C3X DSP and AD/DA card, and a headphone for output signals. The four traditional condenser microphones are mounted on a rectangular acrylic plastic and the spacing of each microphone is 4 cm. Then, the aperture size is 12 cm. Figure 43 shows the implementation of a MEMS microphone array with a 15 mm aperture. Figure 44 shows the implementation of a conventional microphone array with a 12 cm aperture. The sound waves are acquired by the microphone array passing through a preamplifier circuit in order to enhance the sound signals and two second-order low pass filters at 8 kHz in order to decrease the noise in high frequency. Figure 45 shows the implementation of the circuits we utilized.

Figure 46 shows the flow chart of 3D spatial sound reconstruction with HRTF. It contains the DOA estimation, beamforming and HRTF. The digital signals are acquired by an A/D card, the sampling rate is 1/16000 sec and 1024 samples are computed in each block. The delay-sum method is calculated to do the DOA

estimation in the DSP platform by TMS320C3X. The superdirective method is also implemented as digital FIR filters to decrease the beamwidth and increase the directivity. Figure 47 shows the DOA estimation of delay-sum and superdirective methods results, where a sound source is located at 10°. The source is a broadband white noise. From the results, the superdirective method is more superior to the delay-sum.

After the DOA estimation, the location of sound source is known. Then, beamforming techniques can be utilized to enhance the received signals. However, the delay sample in the DSP platform is not integer. The fractional delays can not be ignored, or the error of the beamforming angle becomes larger. Figure 48 shows the effects of the delay sample to the beamforming angle with sampling rate at 8, 16, 25 and 50 kHz. From the results, 1 delay sample at sampling rate 8 kHz causes 90° of beamforming angle error. In our experimental case, 1 delay sample at sampling rate 16 kHz causes about 23° of beamforming angle error. The beamforming angle error is lower, when the sampling rate is higher. Thus, second-order Lagrange interpolation can be utilized to avoid the problem. The delay samples have to be precisely calculated to compensate the delay in each microphone. Therefore, the angle of beamforming techniques is more precise.

Finally, HRTFs are used to implement 3D spatial sound field. When the location of the sound source is changed, HRTFs which used are also changed. Therefore, 3D spatial sound can be reconstructed. Figure 49 shows the results of listening test of the 3D spatial sound reconstruction with HRTF. From the results, the source presented in the front from $-50^{\circ} \sim 50^{\circ}$ in X axis can be recognize by the listeners. The perceived angles are recognized by the listeners in Y axis. The points located in the range of the diagonal line mean the more precise in the listening test. Figure 50 shows the perceived angle errors at the presented angle of listening

test.　From the results, the angle errors of the sound source at -50°, -10°, 10° and 50° are somewhat large.　These angles are hard to recognize.　The total average angle error is 11.7°.　The small value means the listeners can recognize the sound source beside the DOA estimation angle.　The listeners recognize the sound location by headphone, when the sound source is moving around.　The test signal is random noise.　From the results, the output signals using microphone arrays with superdirective filters and HRTFs are very good in 3D spatial sound field.　The listeners can almost recognize the location of the sound source.

V. CONCLUSIONS

The linear dynamic model of a MEMS condenser microphone is constructed. The dynamic performances of sensitivity and bandwidth can be acquired.　Taguchi method and GA are used to optimize the performances of a MEMS condenser microphone.　The optimized sensitivity is 10.3 mV/Pa and bandwidth is 33.3 kHz. The performances are greatly improved by the optimization.　Three optimized parameters are diaphragm length, diaphragm thickness and air gap height.　The optimized values of them are 1.37 mm, 0.8 μm and 3.8 μm.　Therefore, the optimized parameters can be utilized to design the optimal MEMS condenser microphone.

In order to realize the effects of initial stress in the diaphragm, the MEMS diaphragm is designed and fabricated.　Resonant frequencies of the diaphragm should be considered.　We know that imaginary part of impedance determines the resonant frequency.　The real part determines the damping effects.　This is because the real part of acoustical impedance is large enough to make the resonant frequency be oppressed.　Therefore, the resonant frequency of the linear dynamic model of the MEMS condenser microphone can not be shown.　Subsequently, the fabricated

results are not very good. The fabricated diaphragm thickness is somewhat large, and some of the diaphragms are etched broken. The reasons should be due to the etching rate of KOH is not uniform and the resistant layers deposited are not very flat. In order to improve these drawbacks, the double-polished wafers should be used. The concentration and temperature of KOH should be controlled precisely. Besides, the step coverage of the deposited layers is also a great problem. Next, a MEMS condenser microphone process is fully designed in this thesis.

Array signal processing techniques are utilized to enhance the SNR and directivity of the microphone array. Delay-sum and superdirective methods are used to do the DOA estimation and beamforming. To compare the delay-sum and superdirective methods, beamwidth can be decreased and directivity can be increase by superdirective method. In others words, superdirective method is superior to delay-sum. Next, the comparisons of broadside and endfire arrays are illustrated. The directivity of endfire array is better than broadside array. The beamwidth of broadside array is less than endfire array. This is because the directivity of endfire array is only shown up in front of the array. Thus, it is hard to determine which one is better. In the case of our applications, the broadside array is utilized. Then, the directivity analysis is discussed. The directivity is mainly determined by aperture size of the array and frequency. The larger aperture size of the array, the value of directivity is larger. The higher frequencies, the value of directivity is also larger. From simulations and measurements, the directivity of four microphones is better than one microphone. Besides, the SNRs shown in Table VIII are greatly improved by array with four microphones. Thus, the SNR gains are also increased 8.6 and 11.9 dB. Therefore, the measured results show that the directivity and SNRs are really improved by microphone array.

For the purposes of 3D spatial sound reconstruction, HRTFs are utilized to

implement our goal. First, the inverse-filtering is used to find out the filters between microphone arrays and human ears. The simulated results show that the inverse filters consist of the information of the location of sound source. Therefore, the inverse filters can be implemented for beamforming to enhance the SNRs and directivity. Next, 3D spatial sound reconstruction with HRTFs is implemented in TMS320C3X DSP as following. First, delay-sum with superdirective filters is utilized to do the DOA estimation to reduce the beamwidth of the microphone array. When the DOA estimation is done, the DOA estimation angle of HRTF is used to reconstruct 3D spatial sound. The listening test results of 10 people shows that the locations of sound source can be apparently recognize with headphones. Therefore, it can be utilized for hearing aids or other applications to enhance SNR and directivity.

REFERENCES

[1] P. R. Scheeper, A. G. H. van der Donk, W. Olthus, and P. Bergveld, "A review of silicon microphones," Sensors and Actuators A, **44**, 1-11 (1994).

[2] R. P. Ried, E. S. Kim, D. M. Hong, and R. S. Muller, "Piezoelectric microphone with on-chip CMOS circuits," J. Microelectromech. Syst., **2**, 111–120 (1993).

[3] D. P. Arnold, T. Nishida, L. N. Cattafesta III, and M. Sheplak, "A directional acoustic array using silicon micromachined piezoresistive microphones," J. Acoust. Soc. Am., **113**(1), 289-298 (2003).

[4] J. Bergqvist and F. Rudolf, "A silicon condenser microphone using bond and etch-back technology," Sensors and Actuators A, **45**, 115-124 (1994).

[5] W. Kuhnel and G. Hess, "A silicon condenser microphone with structured back plate and silicon nitride membrane," Sensors and Actuators A, **30**, 251-258 (1992).

[6] X. Jin, I. Ladabaum, and B. T. Khuri-Yakub, "The microfabrication of capacitive ultrasonic transducers," IEEE J. Microelectromech. Syst., **7**(3), 295-302 (1998).

[7] I. Ladabaum, X. Jin, H. T. Soh, A.Atalar, and B.T. Khuri-Yakub, "Surface micromachined capacitive ultrasonic transducers," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., **45**(3) 678-690 (1998).

[8] Q. Zou, Z. Li, and L. Liu, "Design and fabrication of silicon condenser microphone using corrugated diaphragm technique," IEEE J. Microelectromech. Syst., **5**, 197–204 (1996).

[9] M. Pedersen, W. Olthuis, and P. Bergveld, "A silicon condenser microphone with polyimide diaphragm and backplate," Sensor and Actuators A, **63**(2), 97–104 (1997).

[10] P. R. Scheeper, A. G. H. van der Donk, W. Olthuis, and P. Bergveld, "Fabrication of silicon condenser microphones using single wafer technology," J. Microelectromech. Syst., **1**(3), 147-154 (1992).

[11]M. Pedersen, A polymer condenser microphone realized on silicon containing preprocessed integrated circuits (Ph.D. dissertation, University of Twente, The Netherlands, 1997).

[12]A. Caronti, R. Carotenuto, and M. Pappalardo, "Electromechanical coupling factor of capacitive micromachined ultrasonic transducers," J. Acoust. Soc. Am. **113**(1), 279-288 (2003).

[13]白明憲, 聲學理論與應用-主動式噪音控制, 全華書局, 2001

[14]Robert, H. C., Jr., *Fundamental Concepts in the Design of Experiments*, Hicks, C. R. and Turner, K. V., Jr., (New York, 1999)

[15]田口玄一(陳耀茂 譯), 田口實驗計畫法, 滄海書局, 1997

[16]J. H. Holland, "Outline for a logical theory of adaptive system," Journal of Association for Computing Machine, **3**(1), 297-314 (1962).

[17]C. T. Lin, and G. C. S. Lee, *Neural Fuzzy System*, (Prentice Hall, New Jersey, 1999).

[18]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques,* (Prentice-Hall PTR: Englewood Cliffs, NJ, 1993)

[19]Harry L. Van Trees, *Optimum Array Processing*, (Wiley & Sons, Inc., New York, 2002)

[20]B. A. Cray, V. M. Evora, and A. H. Nuttall, "Highly directional acoustic receivers," J. Acoust. Soc. Am. **113**(3), 1526-1532 (2003).

[21]B. Gardner and K. Martin, "HRTF measurements of a KEMER dummy-head microphone," MIT Media Lab (1994).

[22]P. C. Hsu, C. H. Mastrangelo and K. D. Wise, 1998, "A High Sensitivity polysilicon Diaphragm Condenser Microphone", Conference Record of IEEE 11th International Workshop on Micro Electro Mechanical Systems (MEMS), 580-585 (1998).

[23]H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. On Acoustics, Speech and Signal Processing, **35**, 1365-1375 (1987).

[24]M. Brandstein and D. Ward, *Microphone Arrays: signal processing techniques and applications*, (Springer-Verlag, Berlin, 2001).

[25]H. Cox et. al., "Practical Supergain", IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-**34**, 393-398 (1986).

[26]R. W. Stadler, W. Rabinowitz. "On the potential of fixed arrays for hearings aids," J. Acoust. Soc. Am., **94**(3), 1332-1342 (1993).

[27]J. Bitzer (Meyer), K.U. Simmer and K.D. Kammeyer, "An Alternative Implementation of the Superdirective Beamformer," Proc. IEEE Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA-99) , pp. 7-10, New Paltz, USA, 17-20 (1999).

[28]P. L. Chu, "Superdirective microphone array for a set-top videoconference system," in Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, (New Paltz, New York, 1997).

[29]O. L. Frost, "An algorithm for linearly constrained adaptive array processing", Proc. IEEE, **60**, 926-935 (1972).

[30]L. E. Kinsler, A. R. Frey, A. B. Coppens and J. V. Sanders, *Fundamentals of Acoustics,* (John-Wiley, New York, USA, 1982).

[31]M. Tanter, J. L. Thomas, and M. Fink, "Time reversal and inverse filter," J. Acoust. Soc. Am. **108**(1), 223-234 (2000).

TABLES

Table I. Orthogonal array L_9 (3^3) includes 3 factors, 3 levels in each factor and 9
observations in Taguchi method.

Observations	Factor 1	Factor 2	Factor 3
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

Table II. First experimental results included of sensitivity, bandwidth, fitness function and optimized parameters with orthogonal array L_9 (3^3) are calculated by Taguchi method.

3 factors								
Diaphragm Length (DL)			Diaphragm Thickness (DT)			Air gap Height (AH)		
3 levels								
A1	A2	A3	B1	B2	B3	C1	C2	C3
0.7	1.1	1.5	0.8	1.6	2.5	2	3	4

	L_9 (3^3)			Parameters			S (mV/Pa)	BW (kHz)	Fitness
Factors	A	B	C	DL	DT	AH	--	--	--
Levels	1	2	3	1	2	3	--	--	--
1	1	1	1	0.7	0.8	2	7	17.63	0.2197
2	1	2	2	0.7	1.6	3	2.3	150.38	0.0626
3	1	3	3	0.7	2.5	4	1.1	268.8	0.0366
4	2	1	2	1.1	0.8	3	10.2	27.067	0.147
5	2	2	3	1.1	1.6	4	4	136.3	0.0664
6	2	3	1	1.1	2.5	2	5.5	20.526	0.235
7	3	1	3	1.5	0.8	4	11.5	29.004	0.1332
8	3	2	1	1.5	1.6	2	19.2	5.2933	0.122
9	3	3	2	1.5	2.5	3	6.2	27.826	0.1898

Levels of parameters	Average of fitness function	Chosen level
A1	(0.2197+0.0626+0.0366)/3 = 0.1063	
A2	(0.147+0.0664+0.235)/3 = 0.1495	A2=1.1 mm
A3	(0.1332+0.122+0.1898)/3 = 0.1483	
B1	(0.2197+0.147+0.1332)/3 = 0.1666	
B2	(0.0626+0.0664+0.122)/3 = 0.0836	B1=0.8 μm
B3	(0.0366+0.235+0.1898)/3 = 0.1538	
C1	(0.2197+0.235+0.122)/3 = 0.1922	
C2	(0.0626+0.147+0.1898)/3 = 0.1331	C1=2 μm
C3	(0.0366+0.0664+0.1332)/3 = 0.0787	

Table Ⅲ. Second experimental results included of sensitivity, bandwidth, fitness function and optimized parameters with orthogonal array L_9 (3^3) are calculated by Taguchi method.

3 factors								
Diaphragm Length (DL)			Diaphragm Thickness (DT)			Air gap Height (AH)		
3 levels								
A1	A2	A3	B1	B2	B3	C1	C2	C3
0.9	1.1	1.3	1.2	1.6	2.1	2.5	3.0	3.5

	L_9 (3^3)			Parameters			S (mV/Pa)	BW (kHz)	Fitness
Factors	A	B	C	DL	DT	AH	--	--	--
Levels	1	2	3	1	2	3	--	--	--
1	1	1	1	0.9	1.2	2.5	6	32.849	0.176
2	1	2	2	0.9	1.6	3	3.8	88.91	0.0962
3	1	3	3	0.9	2.1	3.5	2.5	172.68	0.0548
4	2	1	2	1.1	1.2	3	7	39.318	0.1481
5	2	2	3	1.1	1.6	3.5	4.6	95.275	0.088
6	2	3	1	1.1	2.1	2.5	5.1	36.187	0.1766
7	3	1	3	1.3	1.2	3.5	7.8	41.555	0.1372
8	3	2	1	1.3	1.6	2.5	9.1	17.149	0.1866
9	3	3	2	1.3	2.1	3	5.7	41.555	0.1556

Levels of parameters	Average of fitness function	Chosen level
A1	(0.176+0.0962+0.0548)/3 = 0.109	
A2	(0.1481+0.088+0.1766)/3 = 0.1376	A3=1.3 mm
A3	(0.1372+0.1866+0.1556)/3 = 0.1598	
B1	(0.176+0.1481+0.1372)/3 = 0.1538	
B2	(0.0962+0.088+0.1866)/3 = 0.1236	B1=1.2 μm
B3	(0.0548+0.1766+0.1556)/3 = 0.129	
C1	(0.176+0.1766+0.1866)/3 = 0.1797	
C2	(0.0962+0.1481+0.1556)/3 = 0.1333	C1=2.5 μm
C3	(0.0548+0.088+0.1372)/3 = 0.0933	

Table Ⅳ. Third experimental results included of sensitivity, bandwidth, fitness function and optimized parameters with orthogonal array L_9 (3^3) are calculated by Taguchi method.

3 factors								
Diaphragm Length (DL)			Diaphragm Thickness (DT)			Air gap Height (AH)		
3 levels								
A1	A2	A3	B1	B2	B3	C1	C2	C3
1.1	1.3	1.5	0.8	1.2	1.6	2	2.2	2.5

	$L_9(3^3)$			Parameters			S (mV/Pa)	BW (kHz)	Fitness
Factors	A	B	C	DL	DT	AH	--	--	--
Levels	1	2	3	1	2	3	--	--	--
1	1	1	1	1.1	0.8	2	13.8	7.689	0.1591
2	1	2	2	1.1	1.2	2.2	10.5	14.13	0.1779
3	1	3	3	1.1	1.6	2.5	6.6	27.826	0.1838
4	2	1	2	1.3	0.8	2.2	35.6	6.789	0.0671
5	2	2	3	1.3	1.2	2.5	12.4	13.556	0.1584
6	2	3	1	1.3	1.6	2	12.7	8.589	0.1689
7	3	1	3	1.5	0.8	2.5	40.8	6.424	0.059
8	3	2	1	1.5	1.2	2	36.6	4.302	0.0663
9	3	3	2	1.5	1.6	2.2	15.1	7.077	0.1479

Levels of parameters	Average of fitness function	Chosen level
A1	(0.1591+0.1779+0.1838)/3 = 0.1736	
A2	(0.0671+0.1584+0.1689)/3 = 0.1315	A1=1.1 mm
A3	(0.059+0.0663+0.1479)/3 = 0.0911	
B1	(0.1591+0.0671+0.059)/3 = 0.0951	
B2	(0.1779+0.1584+0.0663)/3 = 0.1342	B3=1.6 μm
B3	(0.1838+0.1689+0.1479)/3 = 0.1669	
C1	(0.1591+0.1689+0.0663)/3 = 0.1304	
C2	(0.1779+0.0671+0.1479)/3 = 0.1309	C3=2.5 μm
C3	(0.1838+0.1584+0.059)/3 = 0.1337	

Table V. Fourth experimental results included of sensitivity, bandwidth, fitness function and optimized parameters with orthogonal array L$_9$ (3^3) are calculated by Taguchi method.

	3 factors							
Diaphragm Length (DL)			Diaphragm Thickness (DT)			Air gap Height (AH)		
	3 levels							
A1	A2	A3	B1	B2	B3	C1	C2	C3
1.1	1.2	1.3	1.2	1.4	1.6	2	2.2	2.5

	L$_9$ (3^3)			Parameters			S (mV/Pa)	BW (kHz)	Fitness
Factors	A	B	C	DL	DT	AH	--	--	--
Levels	1	2	3	1	2	3	--	--	--
1	1	1	1	1.1	1.2	2	12.3	10.569	0.1679
2	1	2	2	1.1	1.4	2.2	8.9	16.226	0.193
3	1	3	3	1.1	1.6	2.5	6.6	27.826	0.1838
4	2	1	2	1.2	1.2	2.2	12.8	11.483	0.1591
5	2	2	3	1.2	1.4	2.5	8.9	19.422	0.1812
6	2	3	1	1.2	1.6	2	10.5	10.865	0.1887
7	3	1	3	1.3	1.2	2.5	12.4	13.556	0.1584
8	3	2	1	1.3	1.4	2	15.4	7.689	0.1443
9	3	3	2	1.3	1.6	2.2	10.9	11.483	0.1818

Levels of parameters	Average of fitness function	Chosen level
A1	(0.1679+0.193+0.1838)/3 = 0.1816	
A2	(0.1591+0.1812+0.1887)/3 = 0.1763	A1=1.1 mm
A3	(0.1584+0.1443+0.1818)/3 = 0.1615	
B1	(0.1679+0.1591+0.1584)/3 = 0.1618	
B2	(0.193+0.1812+0.1443)/3 = 0.1728	B3=1.6 μm
B3	(0.1838+0.1887+0.1818)/3 = 0.1848	
C1	(0.1679+0.1887+0.1443)/3 = 0.1669	
C2	(0.193+0.1591+0.1818)/3 = 0.1779	C2=2.2 μm
C3	(0.1838+0.1812+0.1584)/3 = 0.1745	

Table VI. The optimized diaphragm length, diaphragm thickness, air gap height, sensitivity and bandwidth in original, Taguchi method, GA and GA with constrained bandwidth at 16, 20, 30 and 40 kHz.

	DL (mm)	DT (μm)	AH (μm)	S (mV/Pa)	BW (kHz)
Original	1	1	3	7.0	41.6
Taguchi	1.1	1.6	2.2	7.7	18.4
GA	1.02	2.1	2.1	5.4	24.6
GA--16kHz	1.14	0.8	2.7	12.9	17.9
GA--20kHz	1.25	0.8	3.1	12.4	21.9
GA--30kHz	1.37	0.8	3.8	10.3	33.3
GA--40kHz	1.17	0.8	3.6	8.7	43.9

Table Ⅶ. Lagrange interpolation FIR filter coefficients for $N = 1$ and $N = 2$.

	w_{m0}	w_{m1}	w_{m2}
$N = 1$	$1 - e_m$	e_m	
$N = 2$	$(e_m - 1)(e_m - 2)/2$	$-e_m(e_m - 2)$	$e_m(e_m - 1)/2$

Table Ⅷ. The SNRs and SNR gains of one and four microphones measured by a MEMS microphone array with a 15 mm aperture and a conventional microphone array with a 12 cm aperture.

Aperture size	Microphone number	SNR (dB)	SNR gain (dB)
15 mm	One	56.7	--
	Four	68.6	11.9
12 cm	One	46.1	--
	Four	54.7	8.6

FIGURES

Insulation layer   Diaphragm electrode   Back plate electrode

Diaphragm

Backplate
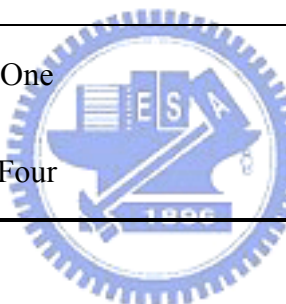
Air gap
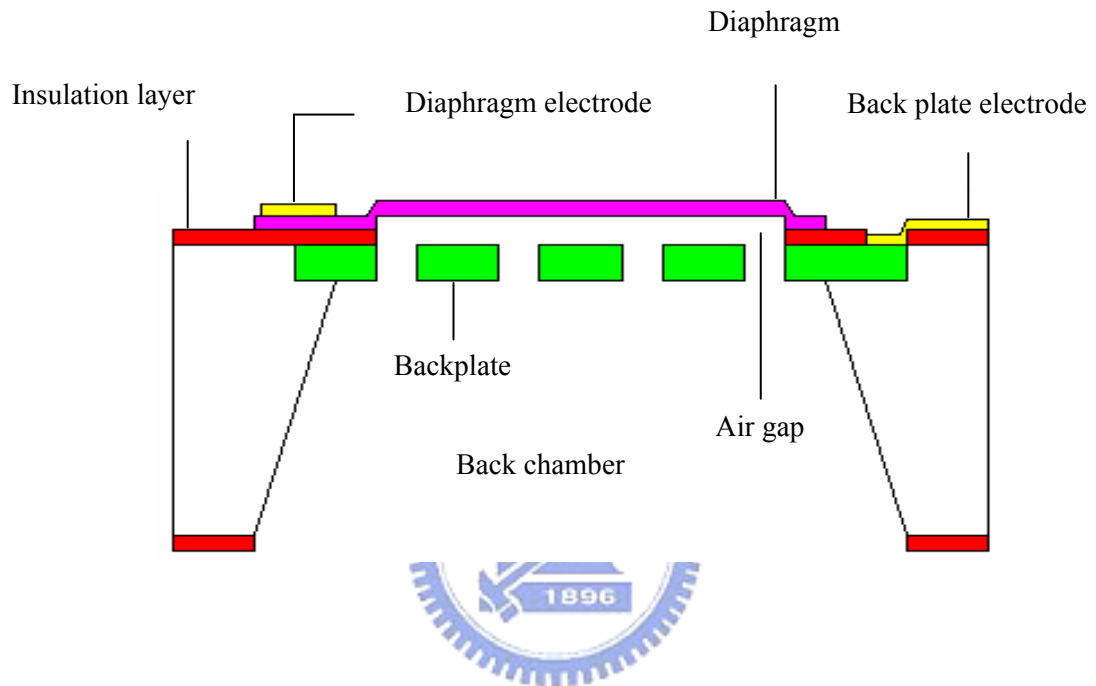
Back chamber

Figure 1.   The cross-sectional view of a single MEMS condenser microphone with a

diaphragm, an air gap, a perforated backplate and a back chamber.
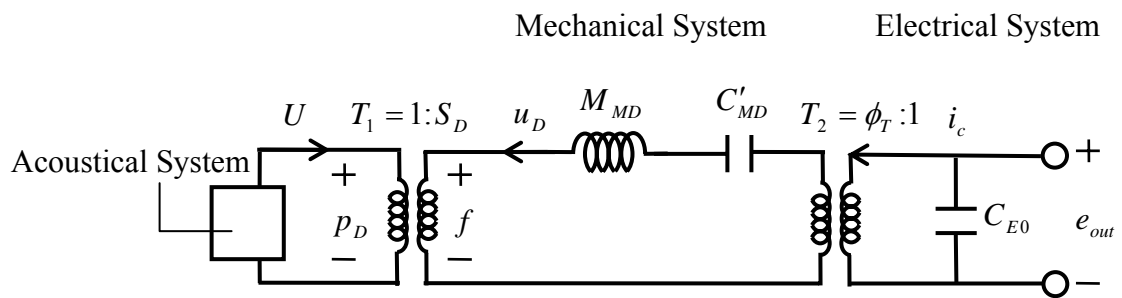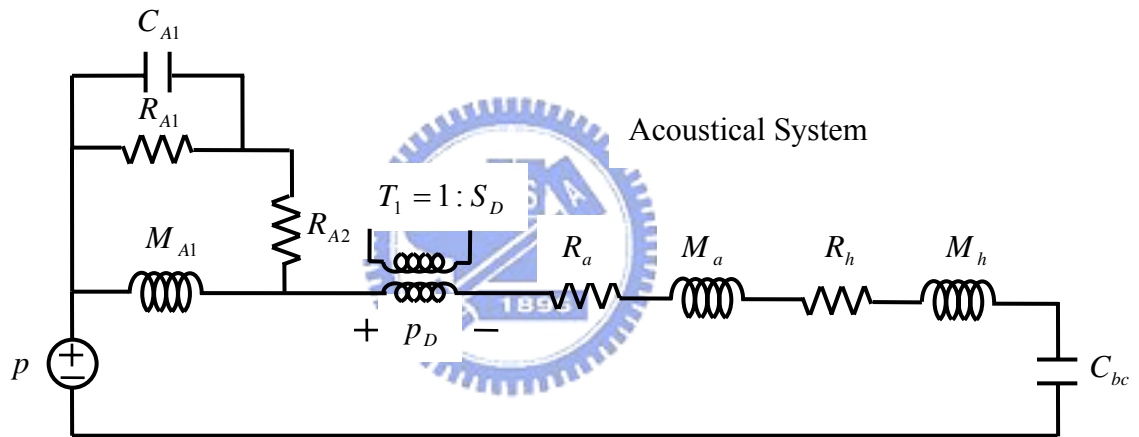
Figure 2. The equivalent circuit based on electro-acoustical analogy of the condenser microphone. (a) Complete system composed of three coupled sub-systems: the acoustical system, mechanical system and electrical system. (b) Detailed circuit representation of the acoustical subsystem.
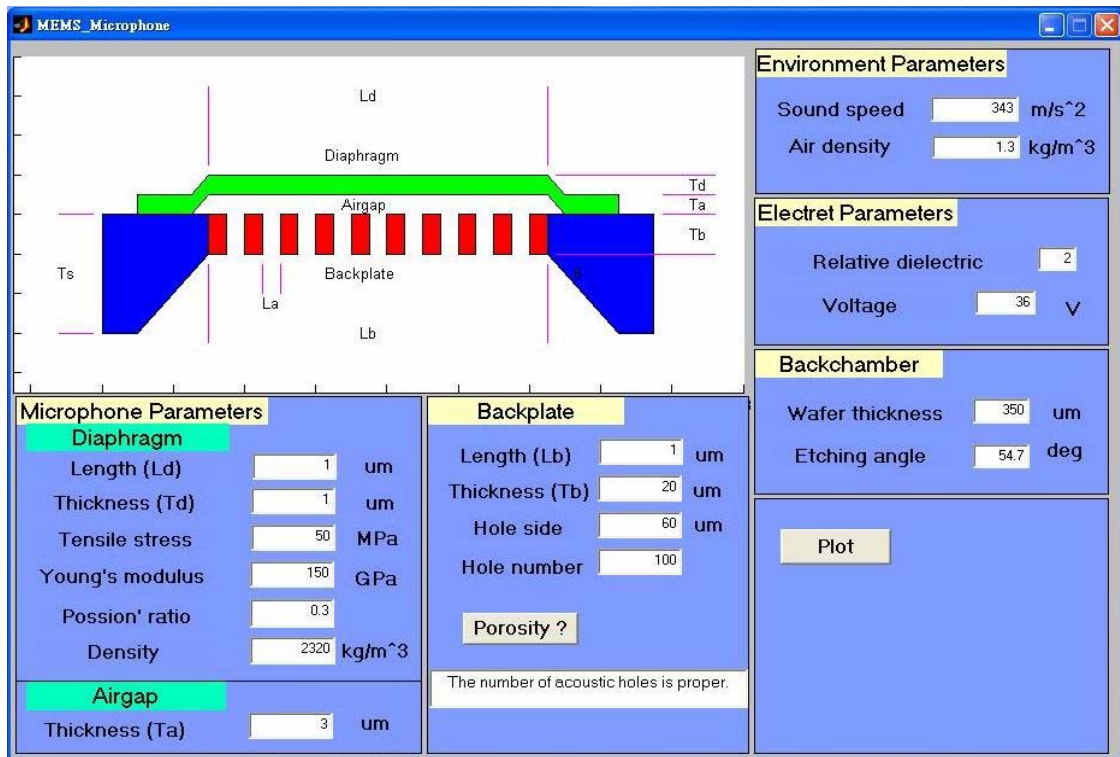
Figure 3.   The MATLAB Graphical User Interface (GUI) is utilized to change the parameters of the linear dynamic model of the MEMS condenser microphone.

The optimal solution is 1.3736 mm

(a)

The optimal solution is 0.7571 um
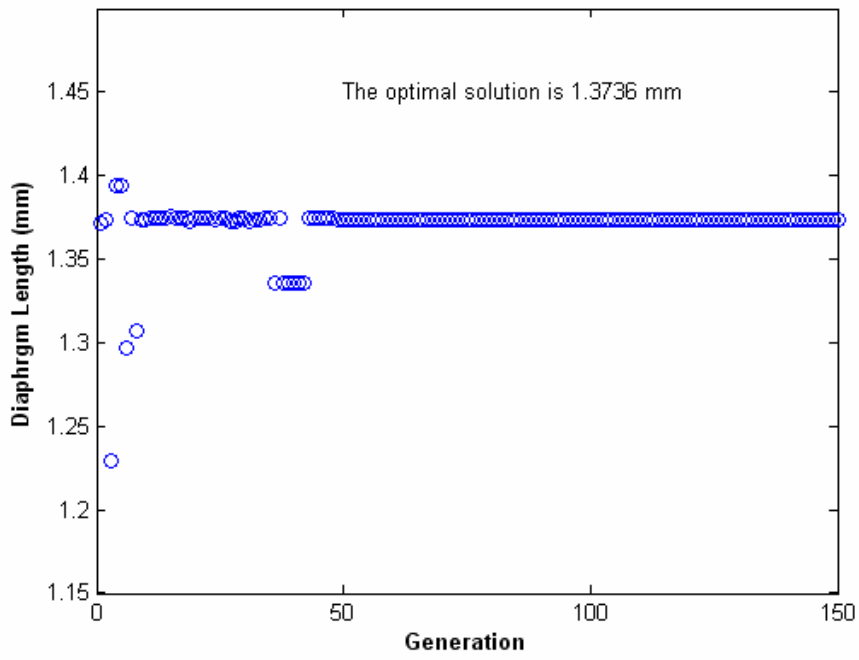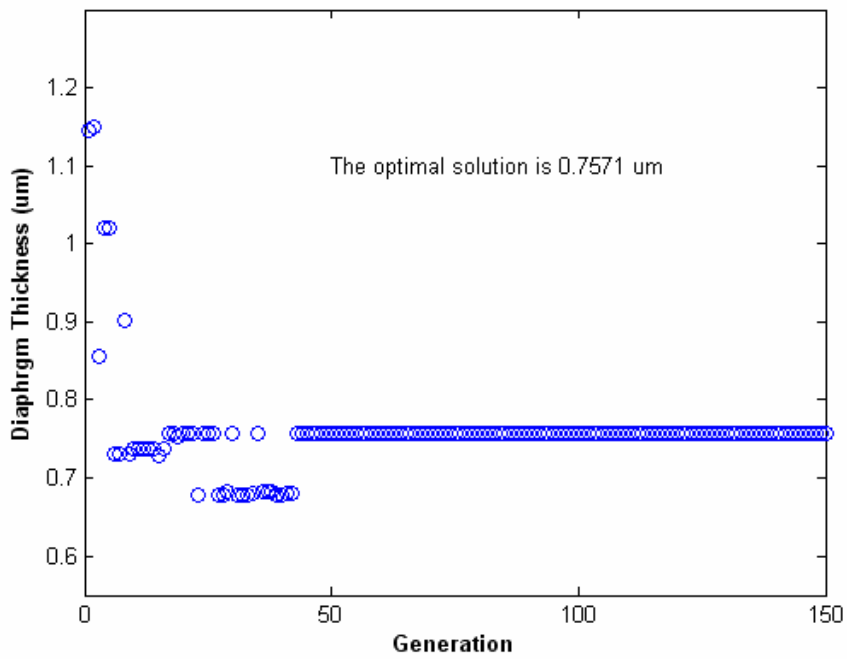
(b)

Figure 4.   The parameters optimized by GA with constrained bandwidth at 30 kHz.

(a) Optimal diaphragm length (b) Optimal diaphragm thickness.

47
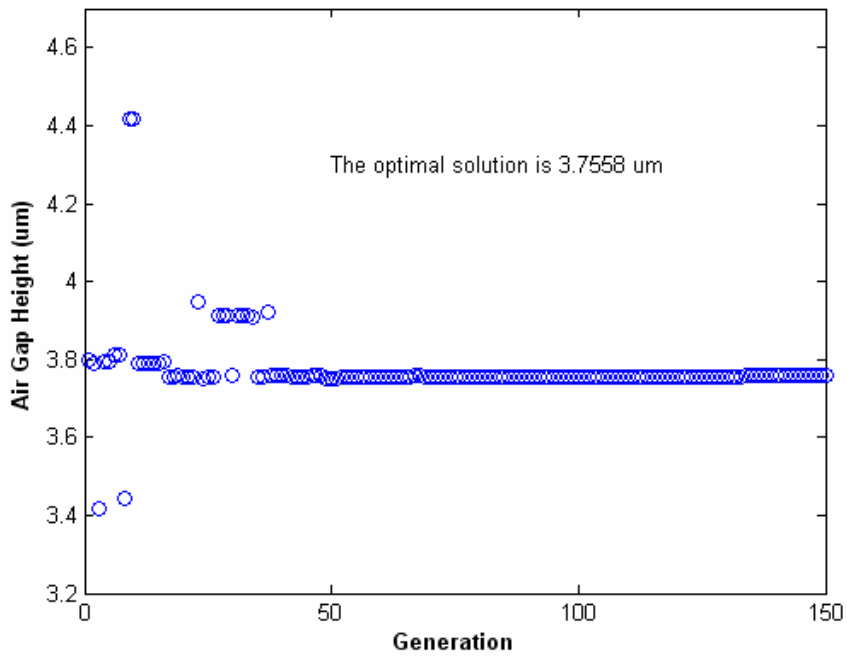
(a)
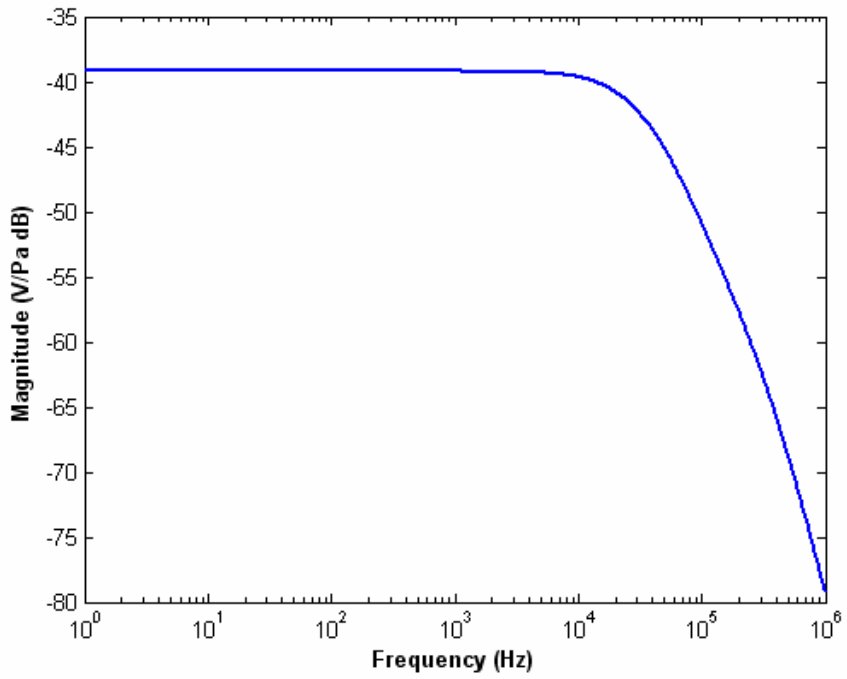
(b)

Figure 5.   The parameters optimized by GA with constrained bandwidth at 30 kHz.

(a) Optimal air gap height (b) Optimal frequency response.

48
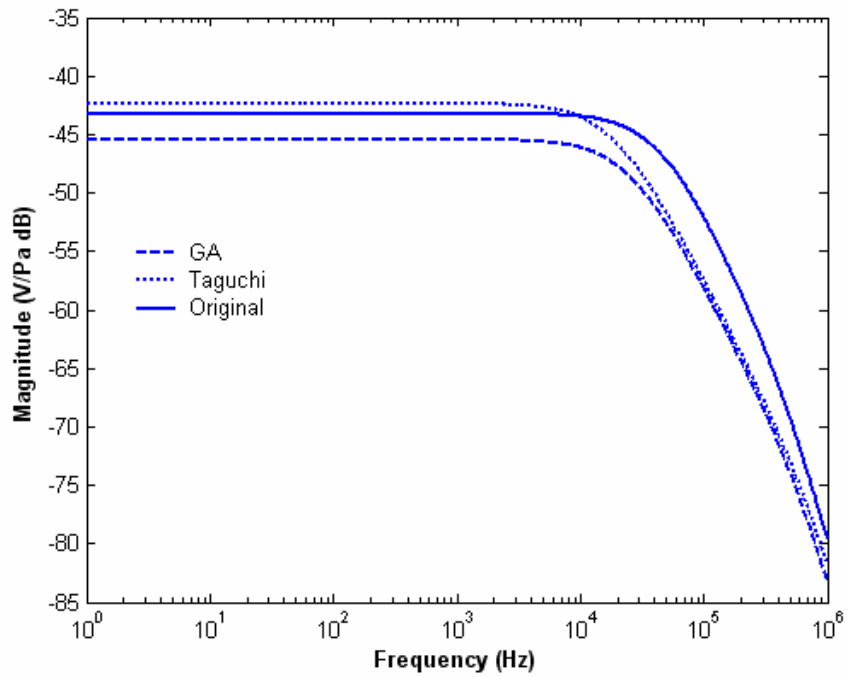
(a)
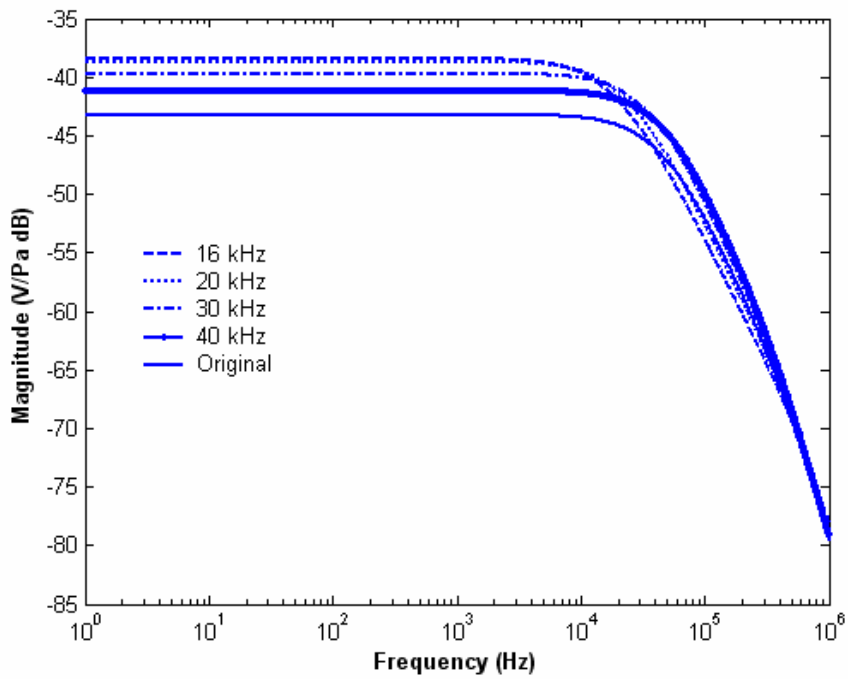
(b)

Figure 6.   The comparisons of frequency responses. (a) GA, Taguchi method and

original (b) GA with constrained bandwidth at 16, 20, 30 and 40 kHz and

original.

(a) Silicon wafer

(f) Pattern Poly-Silicon

(b) Deposit Si$_3$N$_4$

(g) Pattern Si$_3$N$_4$

(c) Deposit Poly-Silicon

(h) Etch back chamber

(d) Deposit Si$_3$N$_4$

(i) Remove Si$_3$N$_4$

(e) Pattern Si$_3$N$_4$
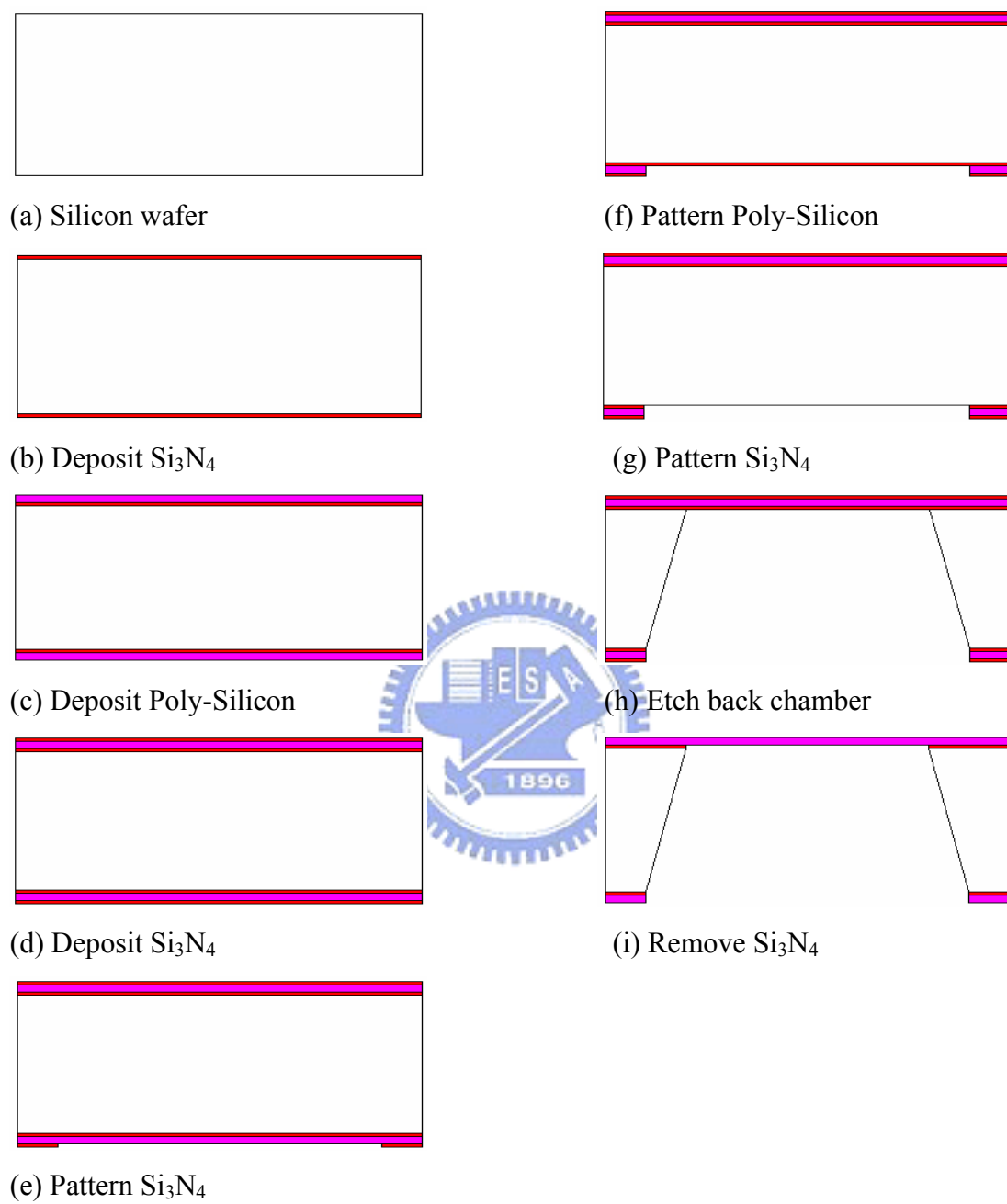
Figure 7.   The fabrication process flow of a MEMS diaphragm (a) - (i).
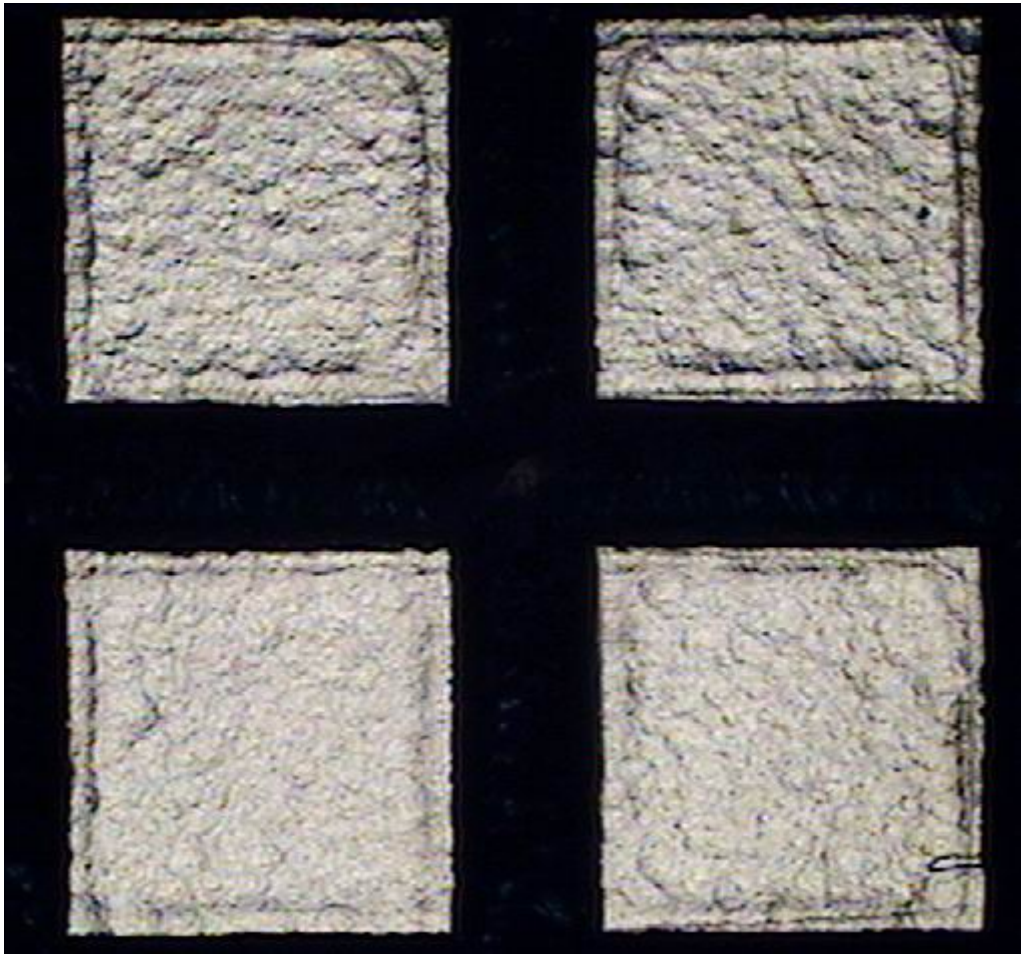
Figure 8.   The MEMS diaphragm after anisotropic etching by KOH 30％ at 80℃.
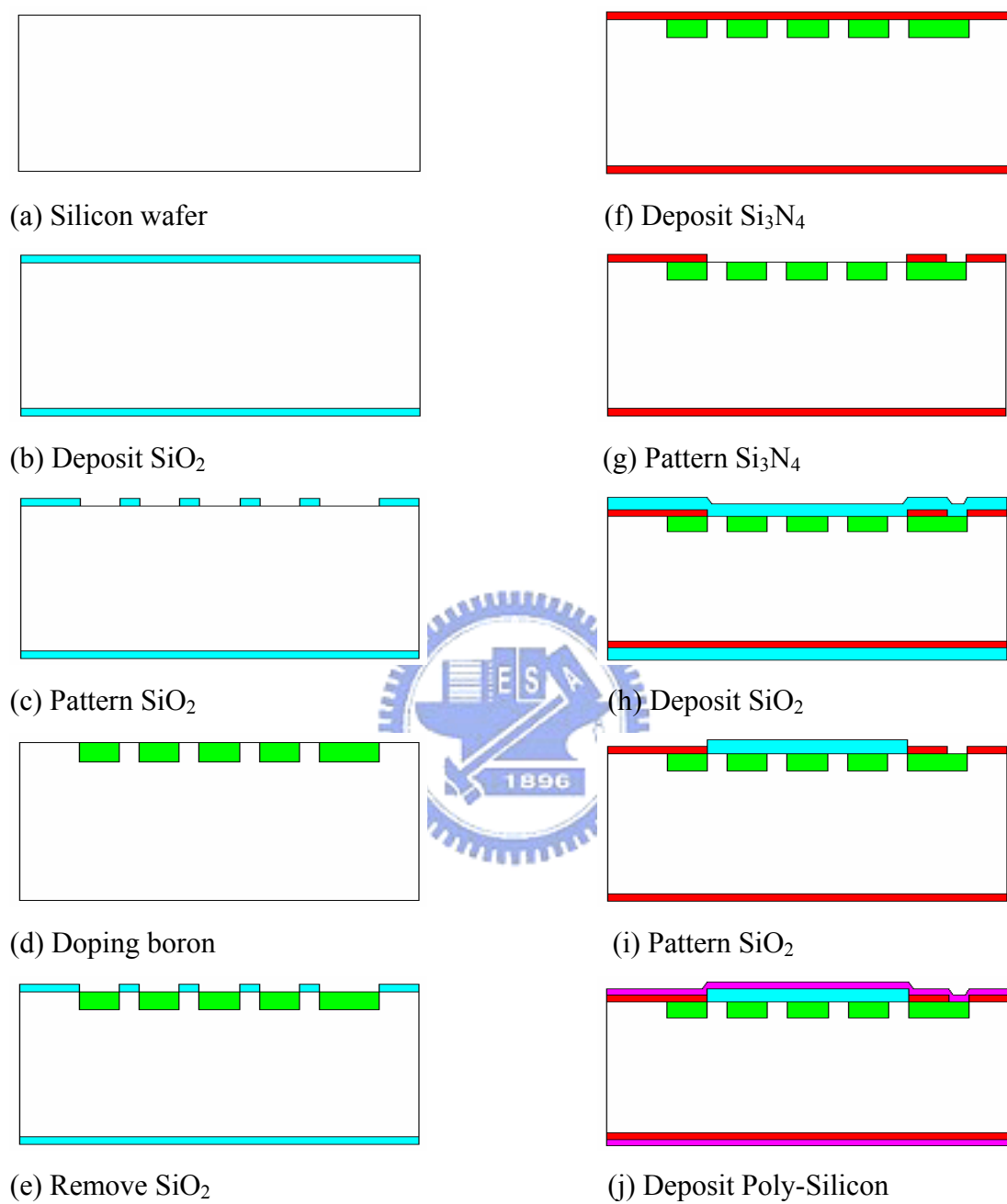
(a) Silicon wafer

(b) Deposit SiO$_2$

(c) Pattern SiO$_2$

(d) Doping boron

(e) Remove SiO$_2$

(f) Deposit Si$_3$N$_4$

(g) Pattern Si$_3$N$_4$

(h) Deposit SiO$_2$

(i) Pattern SiO$_2$

(j) Deposit Poly-Silicon

Figure 9.   The fabrication process flow of a MEMS condenser microphone (a) – (j).

(k) Pattern Poly-Silicon

(o) Remove SiO$_2$

(l) Deposit SiO$_2$

(p) Evaporate Cr and Au

(m) Pattern SiO$_2$ and Si$_3$N$_4$

(q) Pattern Cr and Au

(n) Etch Si
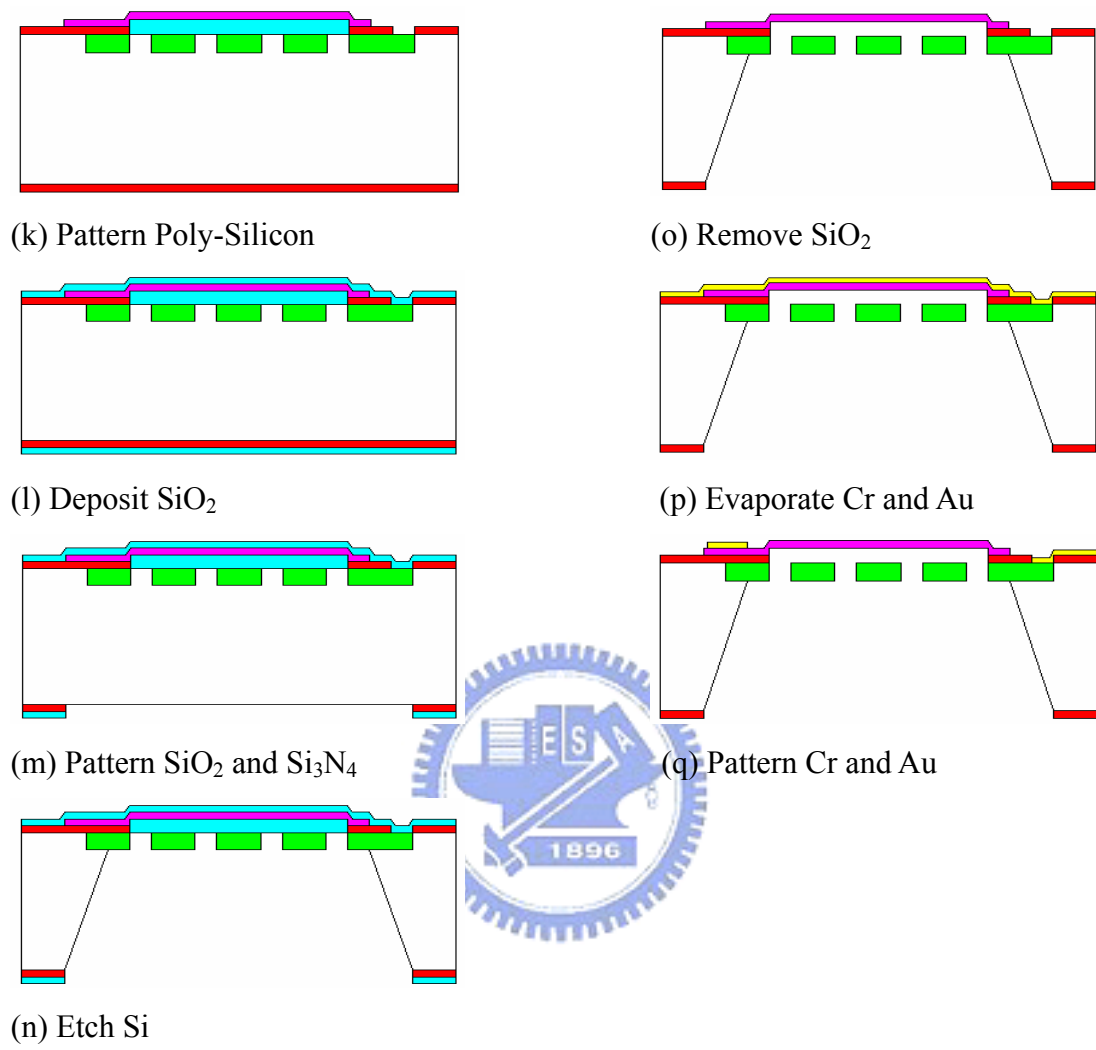
Figure 9.   The fabrication process flow of a MEMS condenser microphone (k) – (q).
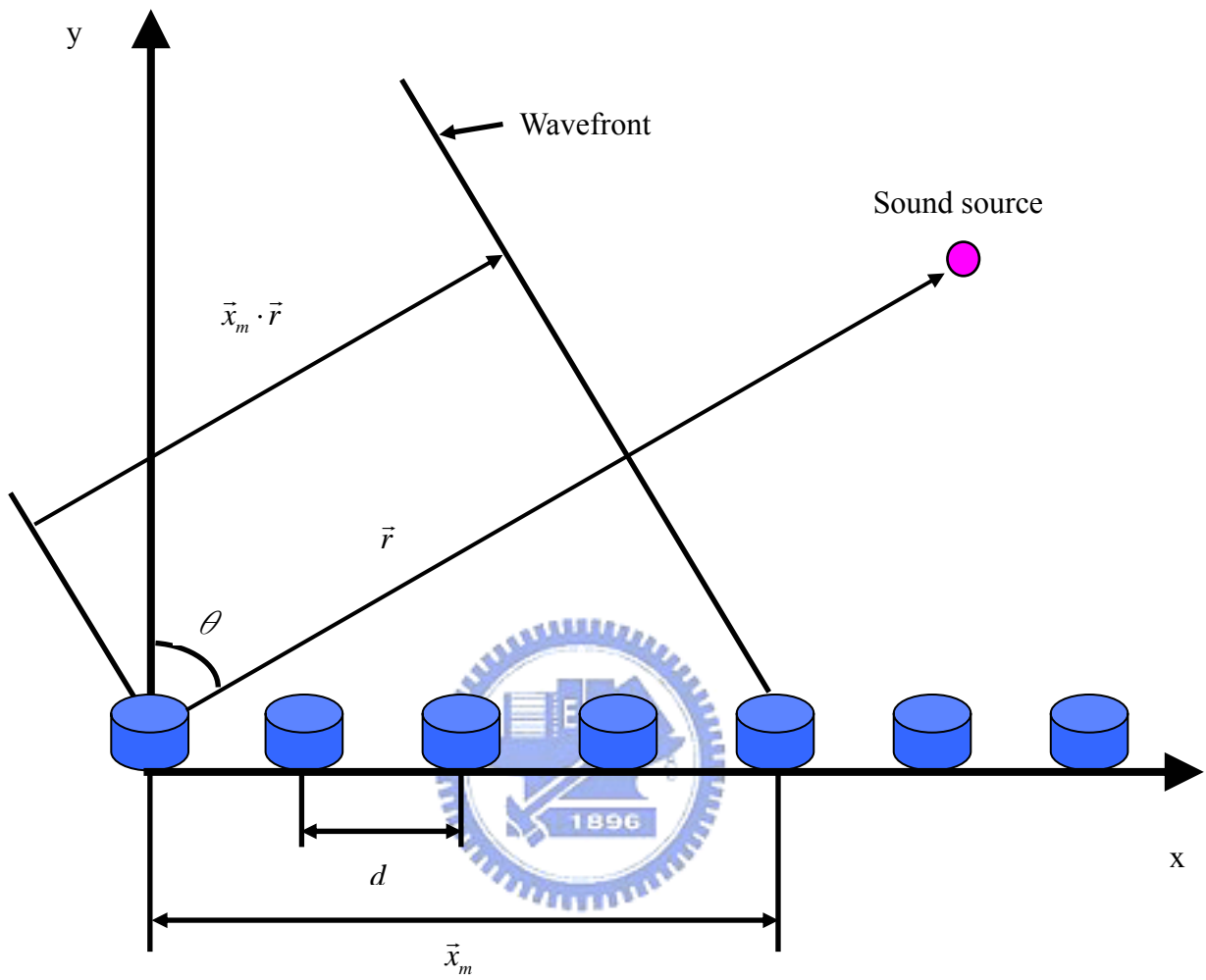
Figure 10.   The configuration of an uniform linear array (ULA).   A point source is

located at the far-field.

Sound source

(a)

Sound source

(b)
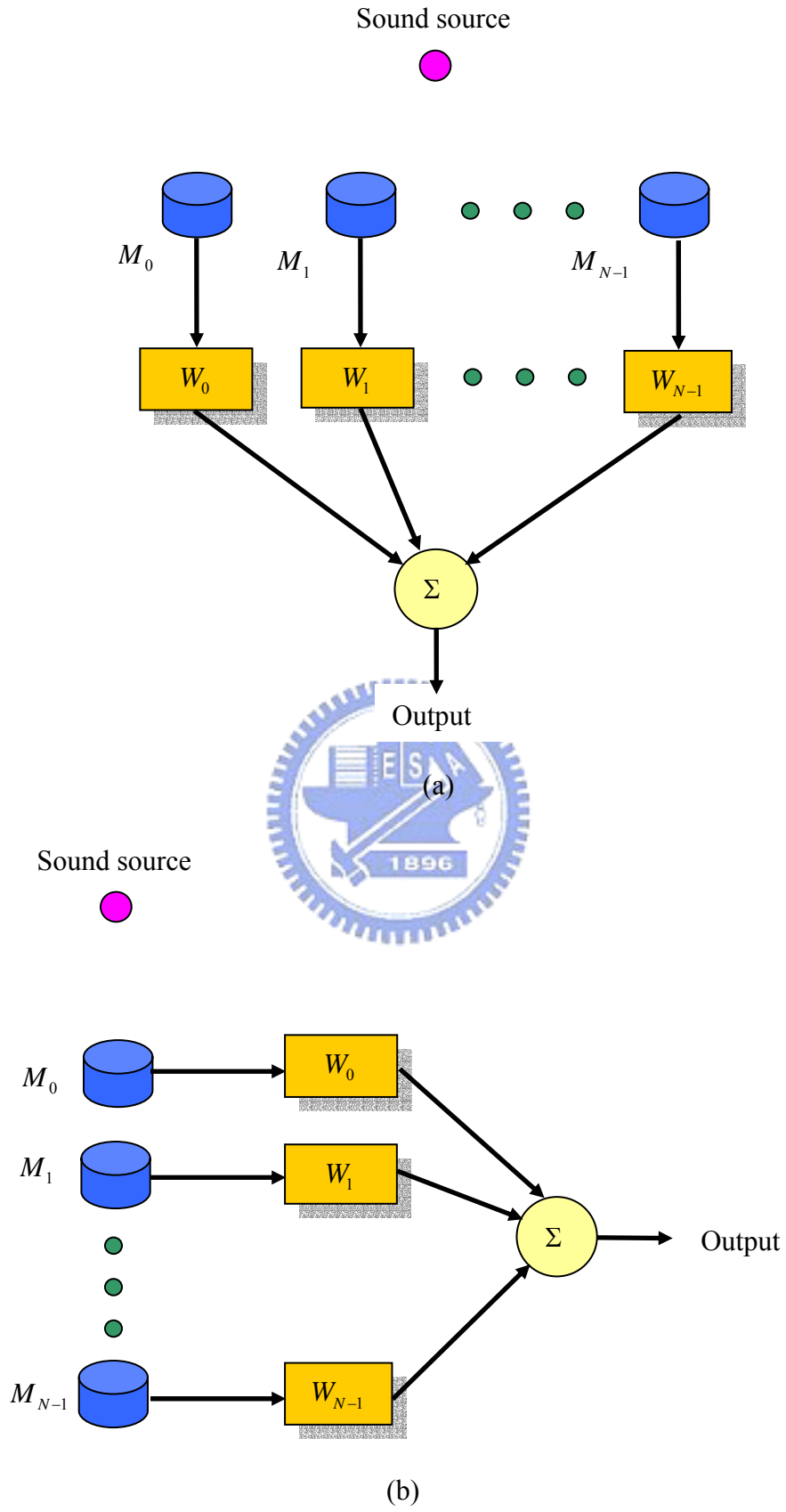
Figure 11.   The schemes of ULAs. (a) A broadside array (b) An endfire array.
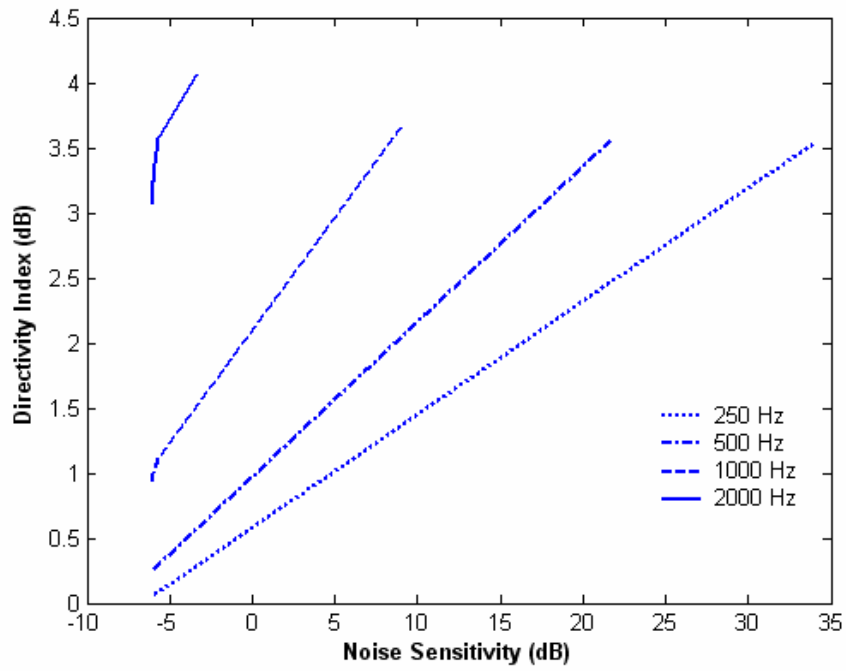
(a)
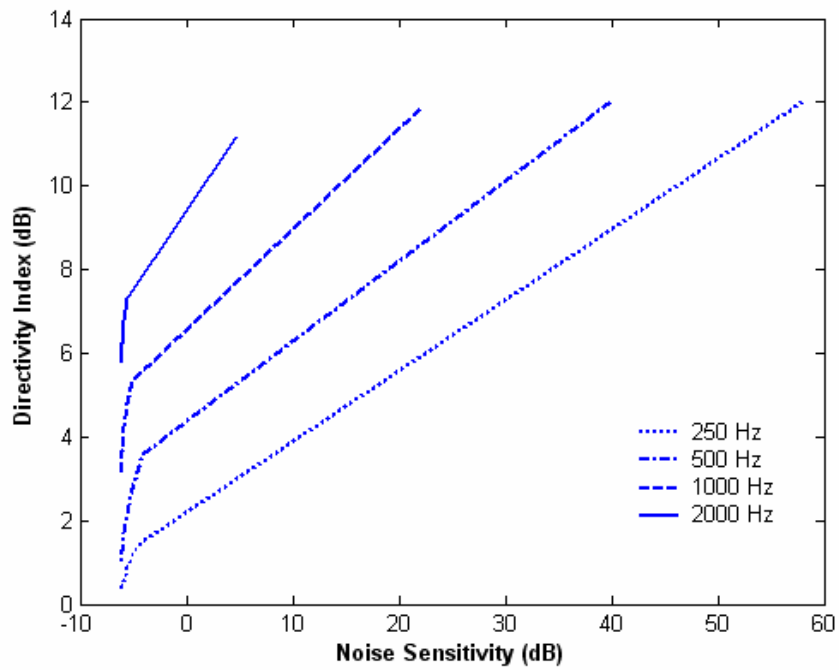
(b)

Figure 12.   The relationship between noise sensitivity and directivity index.

(a) A broadside array (b) An endfire array.
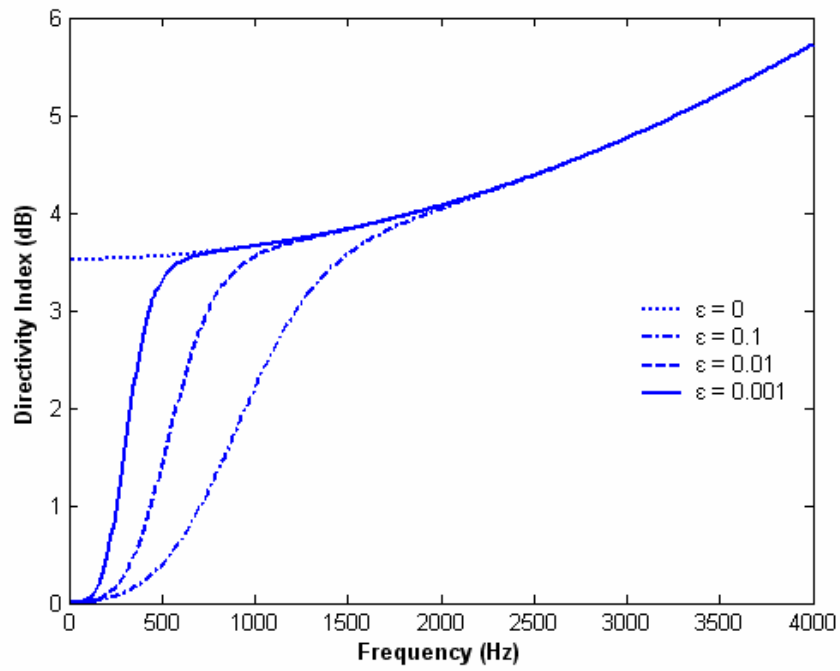
(a)
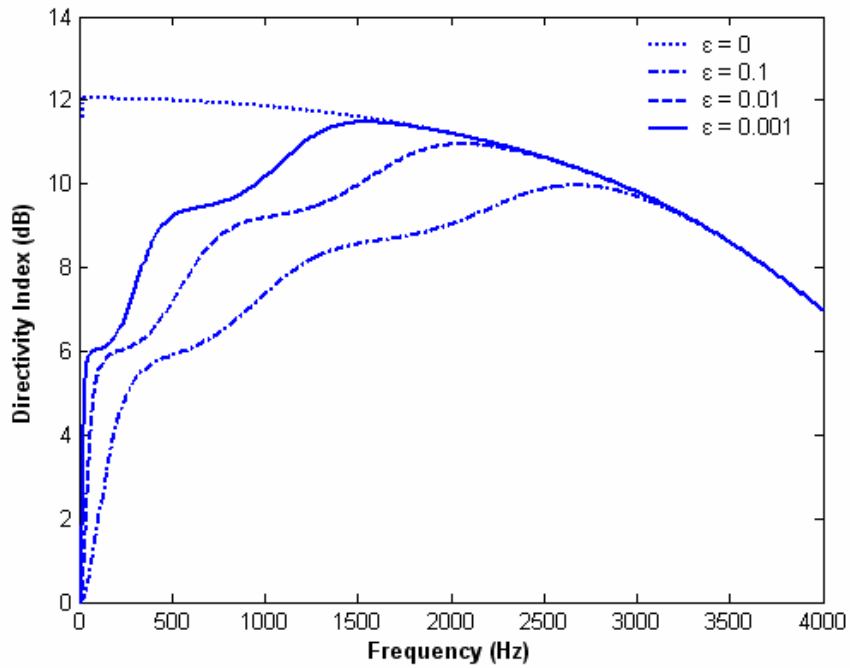
(b)

Figure 13.   The comparisons of directivity index with changing factor ε in 0, 0.1, 0.01 and 0.001 using superdirective method. (a) A broadside array (b) An endfire array.
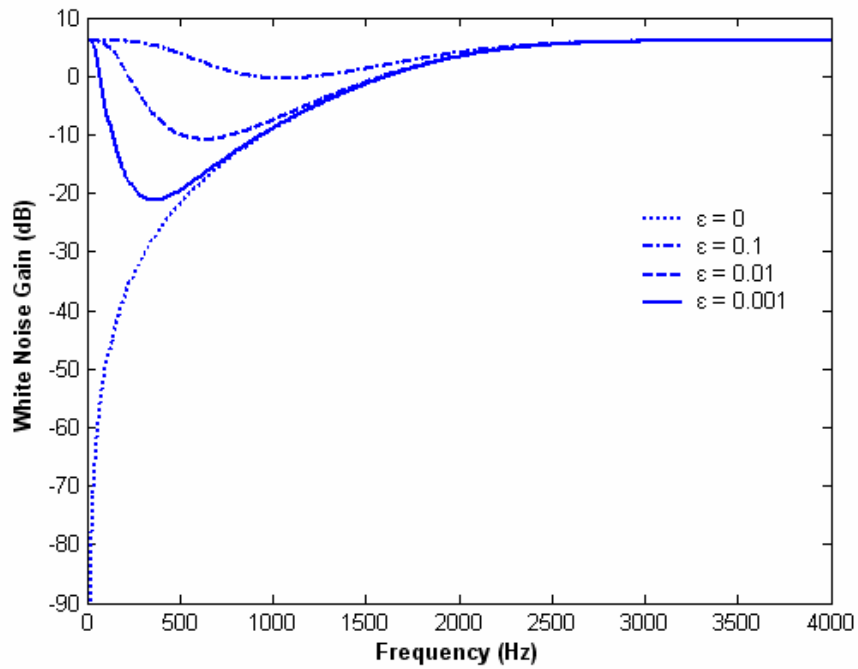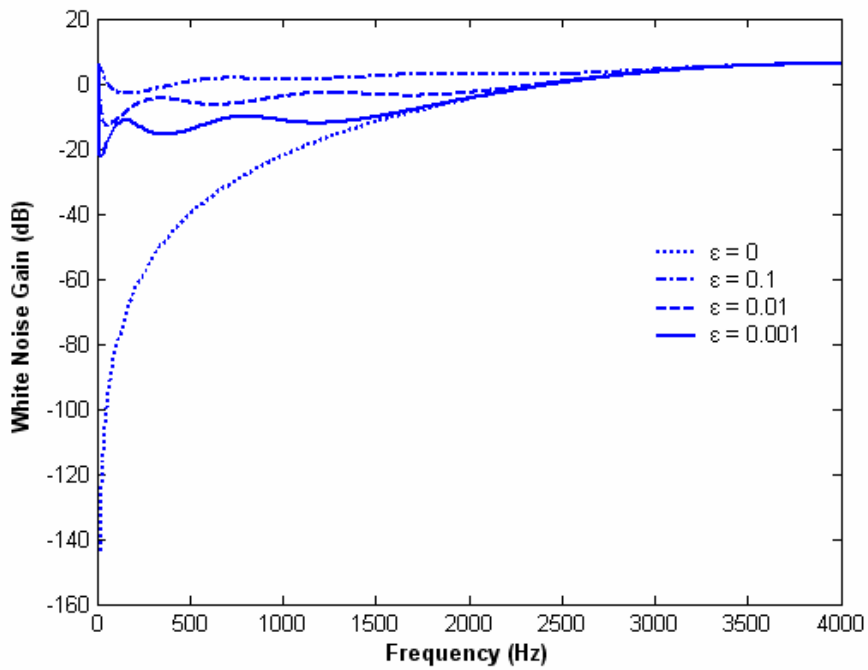
Figure 14. The comparisons of white noise gain with changing factor ε in 0, 0.1, 0.01 and 0.001 using superdirective method. (a) A broadside array (b) An endfire array.
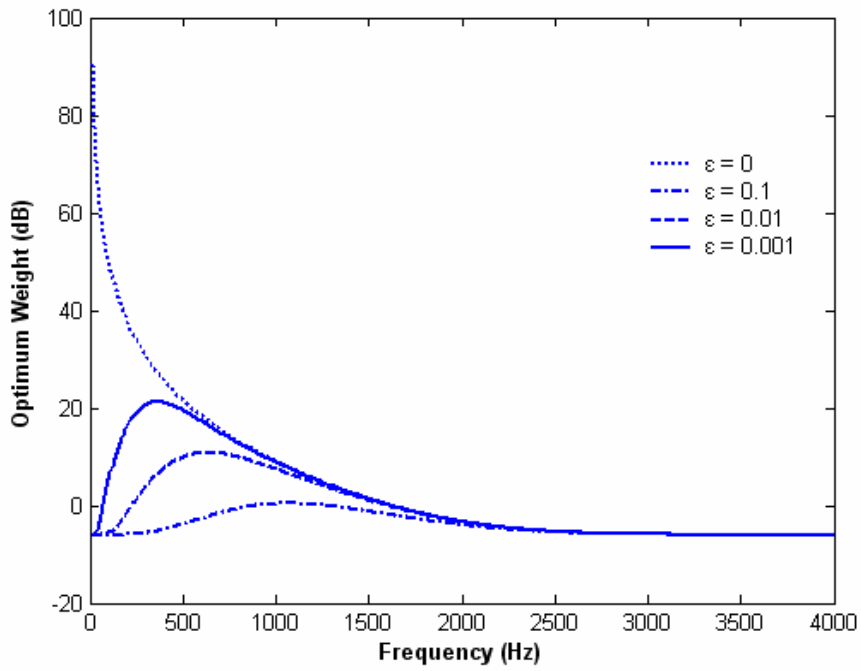
(a)
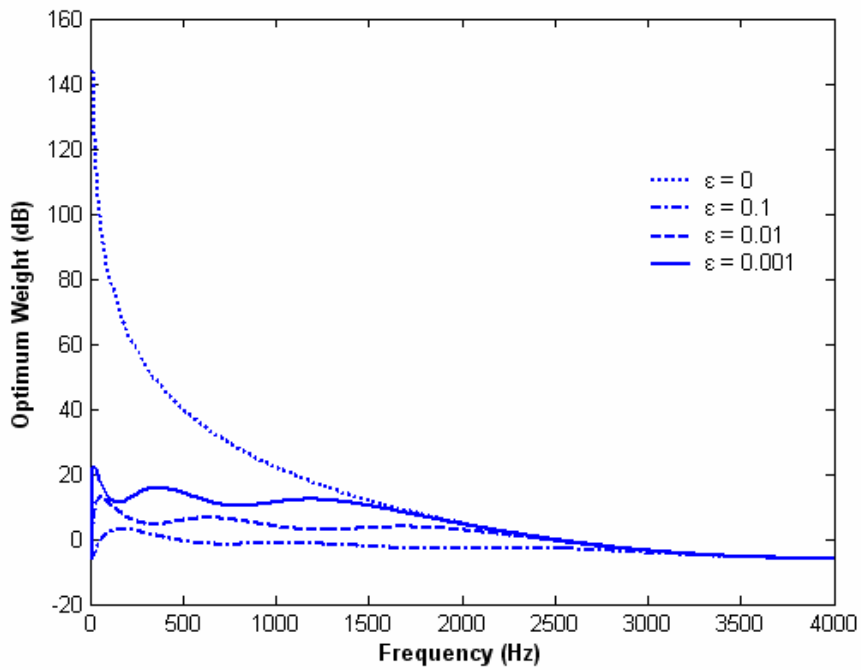
(b)

Figure 15.　The comparisons of optimum weight with changing factor ε in 0, 0.1, 0.01 and 0.001 using superdirective method. (a) A broadside array (b) An endfire array.
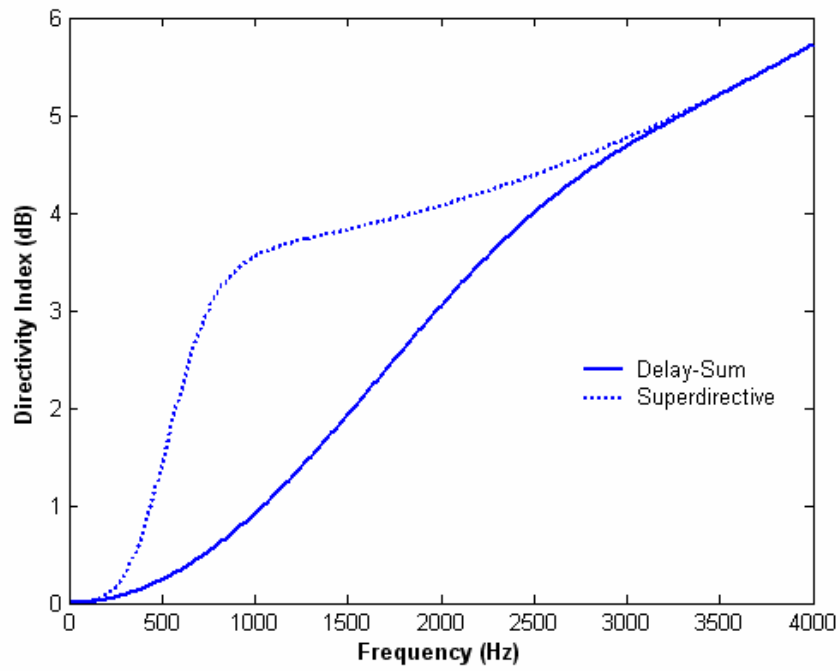
(a)
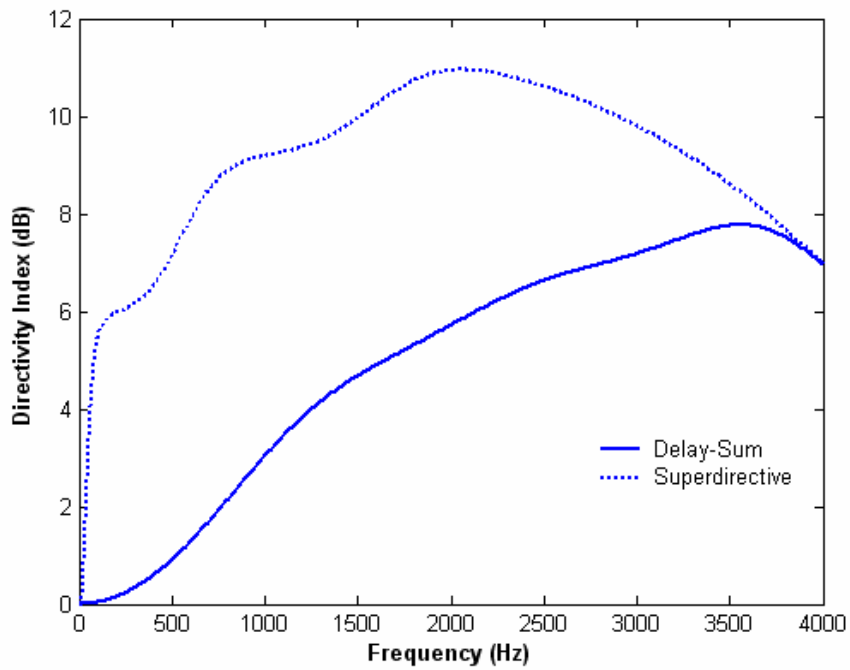
(b)

Figure 16.   The comparisons of directivity index of delay-sum and superdirective

with ε = 0.01. (a) A broadside array (b) An endfire array.

(a)

(b)

Figure 17.   The comparisons of white noise gain of delay-sum and superdirective

with $\varepsilon = 0.01$. (a) A broadside array (b) An endfire array.
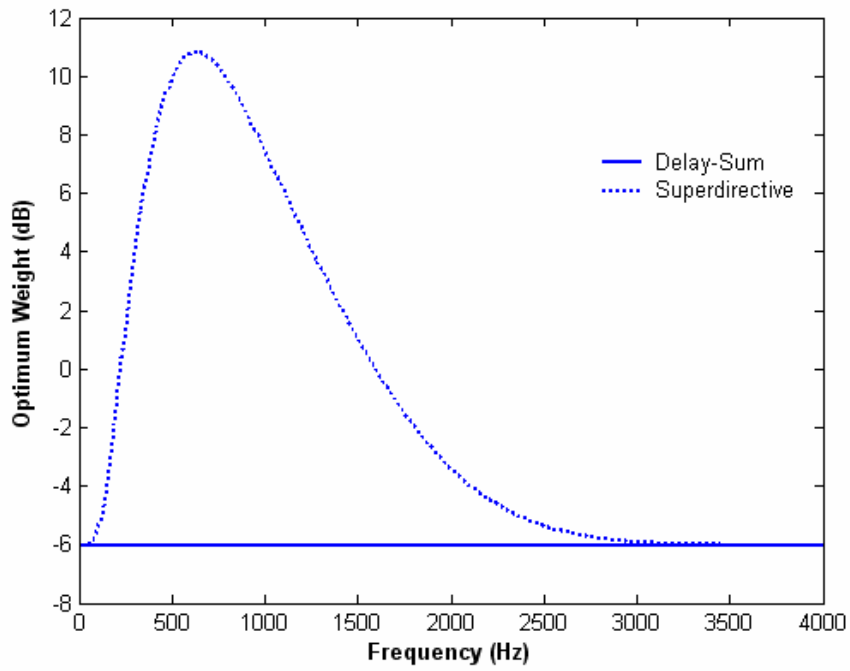
(a)
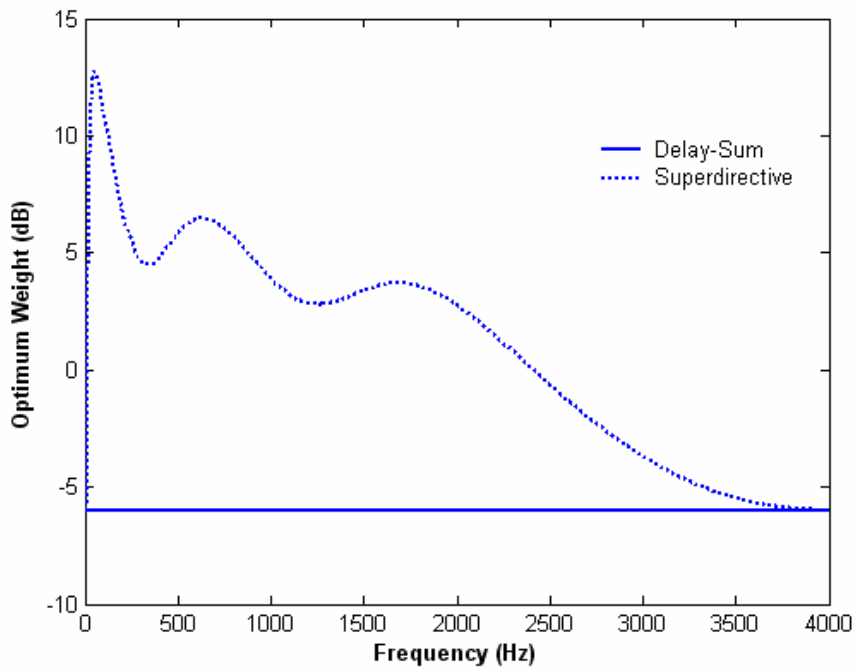
(b)

Figure 18.   The comparisons of optimum weight of delay-sum and superdirective

with ε = 0.01. (a) A broadside array (b) An endfire array.
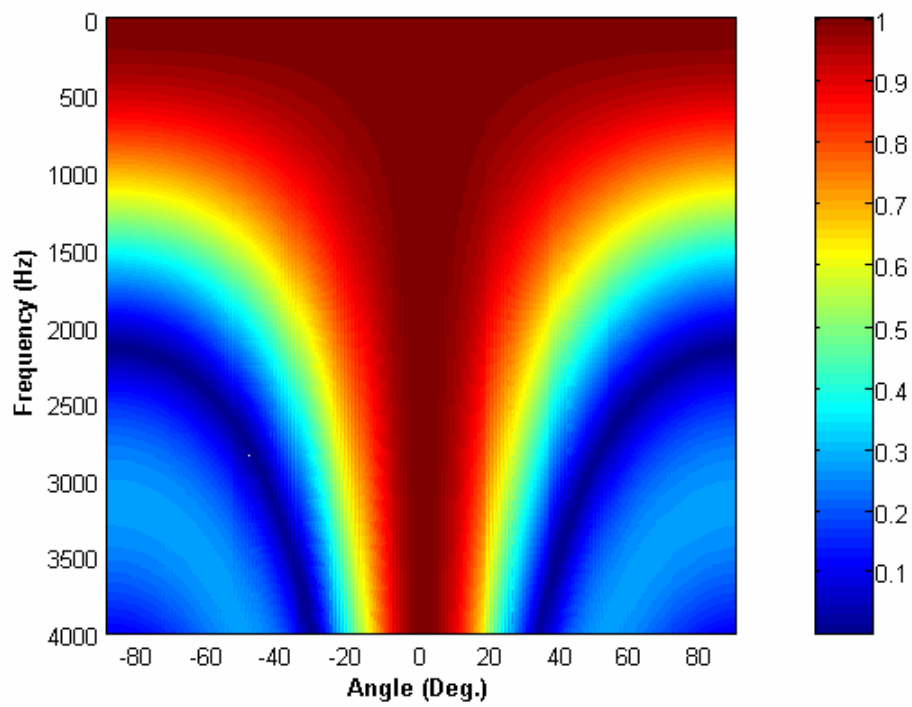
(a)
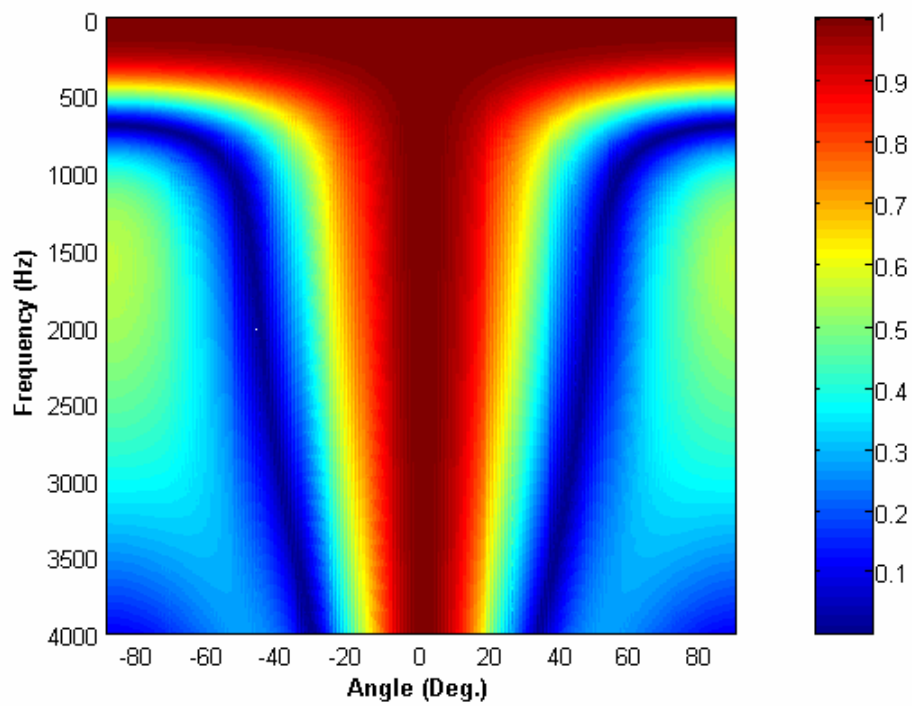
(b)

Figure 19.　The contour plots in a broadside array. (a) Delay-sum (b) Superdirective.
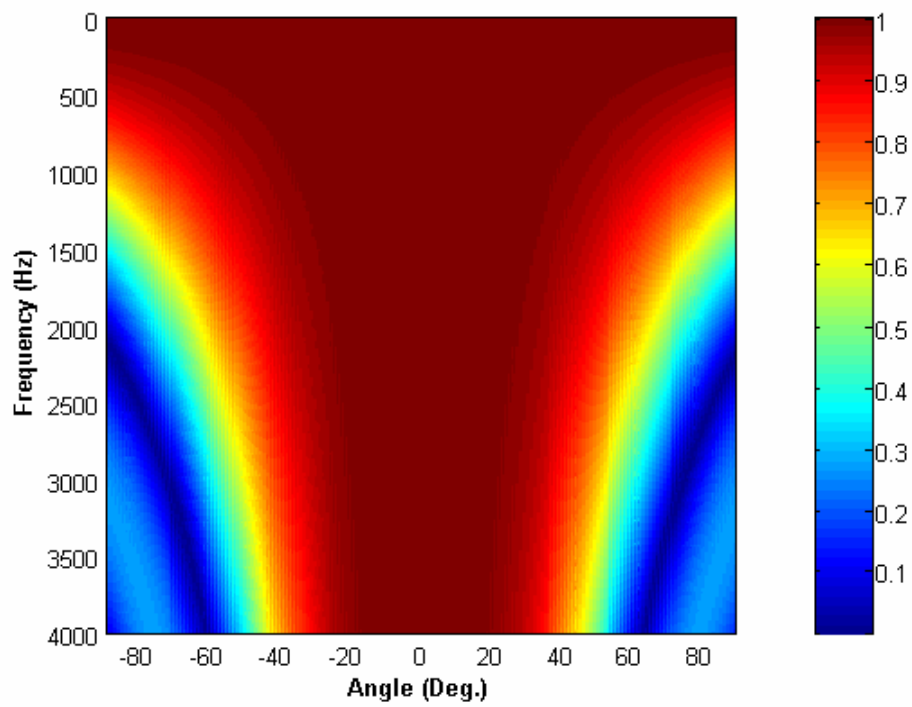
(a)
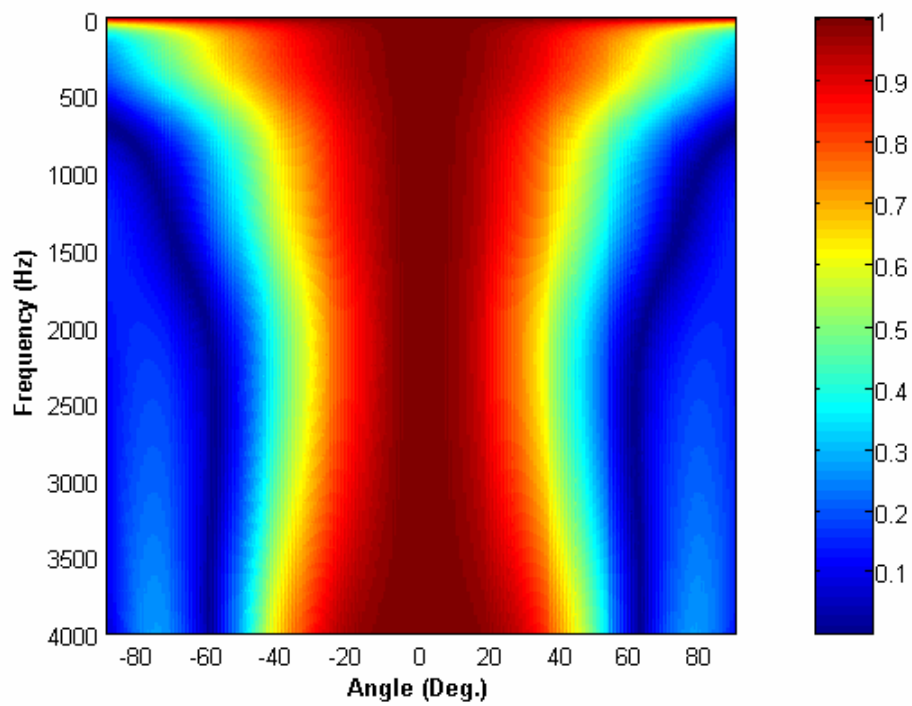
(b)

Figure 20.   The contour plots in an endfire array. (a) Delay-sum (b) Superdirective.
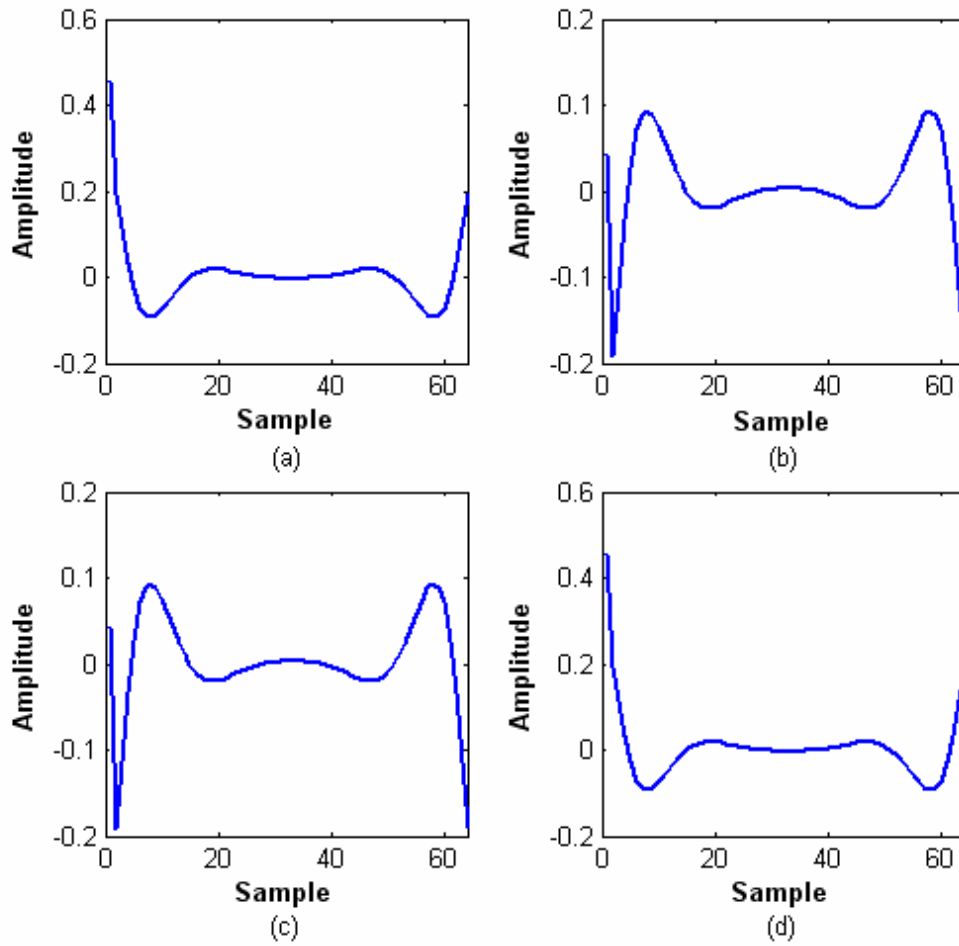
Figure 21. The non-causal impulse response of superdirective filters of order 64.

(a) Microphone 1 (b) Microphone 2 (c) Microphone 3 (d) Microphone 4.
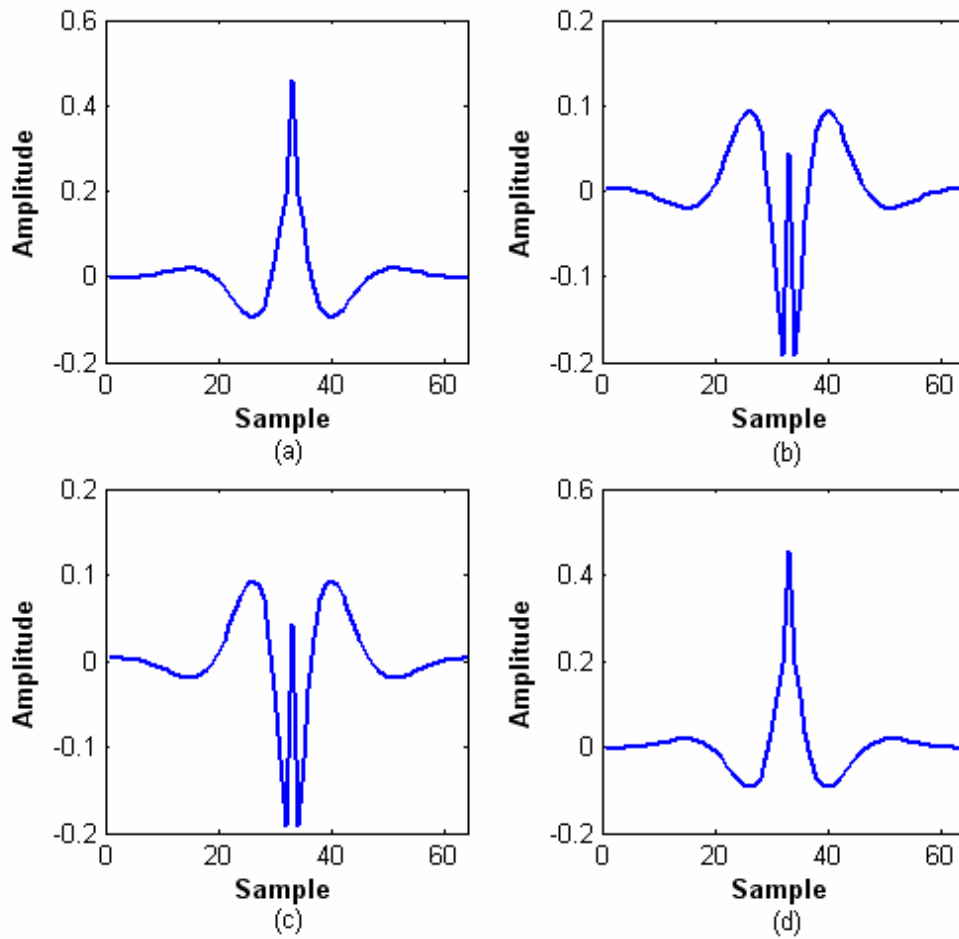
Figure 22. The causal impulse response of superdirective filters of order 64.

(a) Microphone 1 (b) Microphone 2 (c) Microphone 3 (d) Microphone 4.
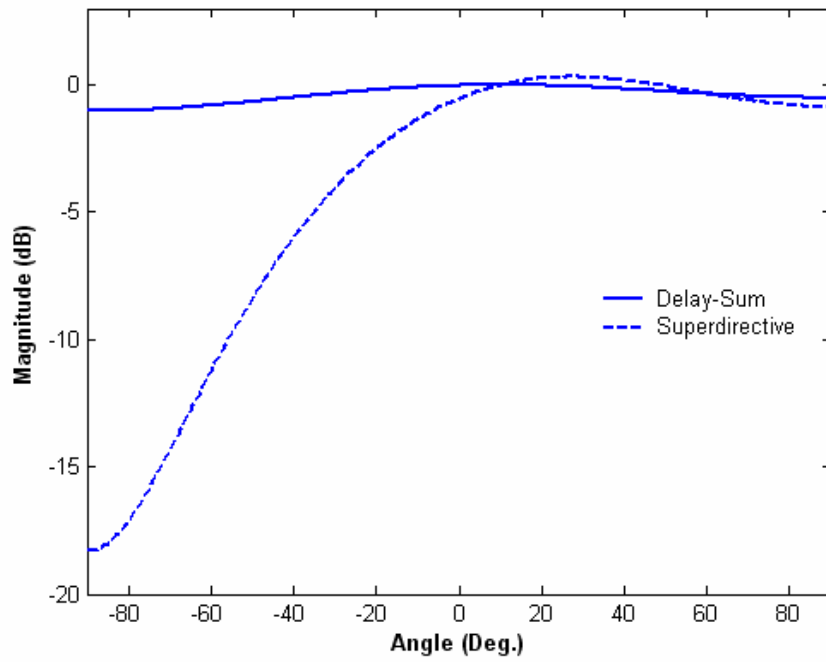
(a)
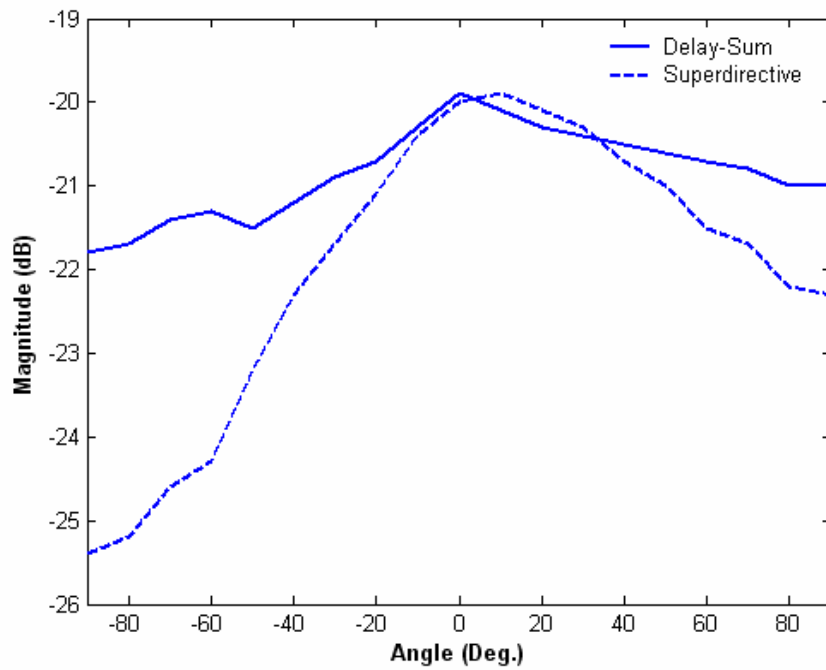
(b)

Figure 23.   The comparisons of the DOA estimation by delay-sum and superdirective

methods with a single frequency at 500 Hz.   The sound source is located

at 10° . (a) Simulations (b) Measurements.

67
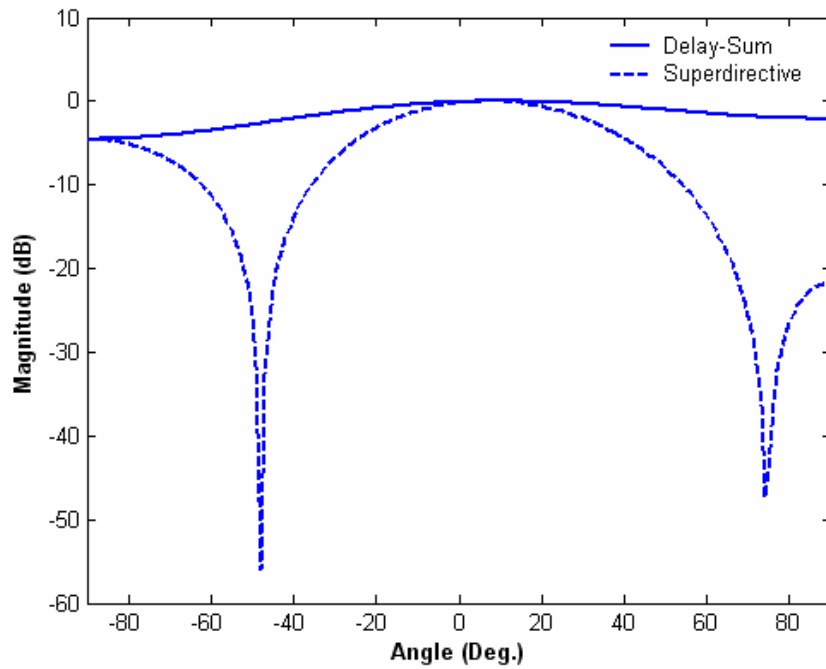
(a)
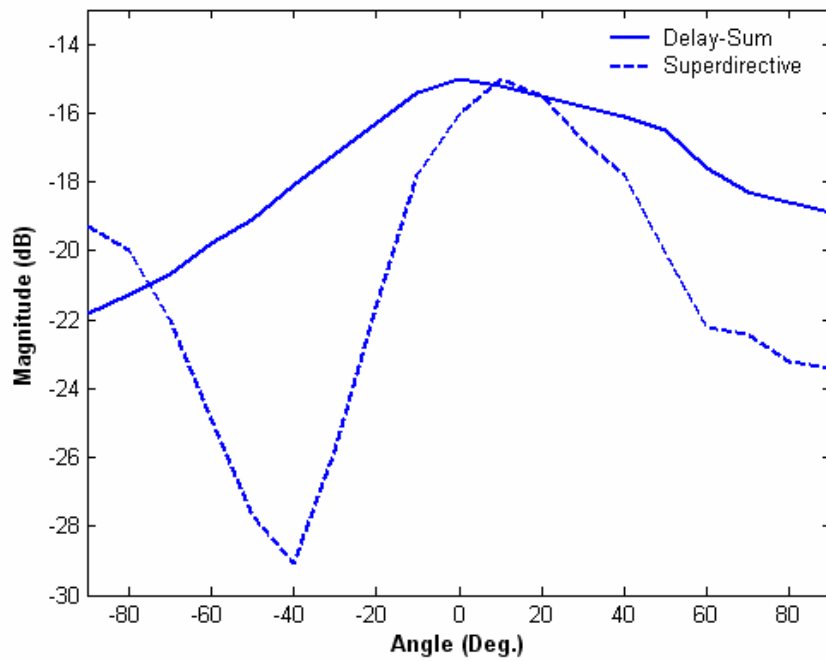
(b)

Figure 24.   The comparisons of the DOA estimation by delay-sum and superdirective

methods with a single frequency at 1000 Hz.   The sound source is

located at 10°. (a) Simulations (b) Measurements.
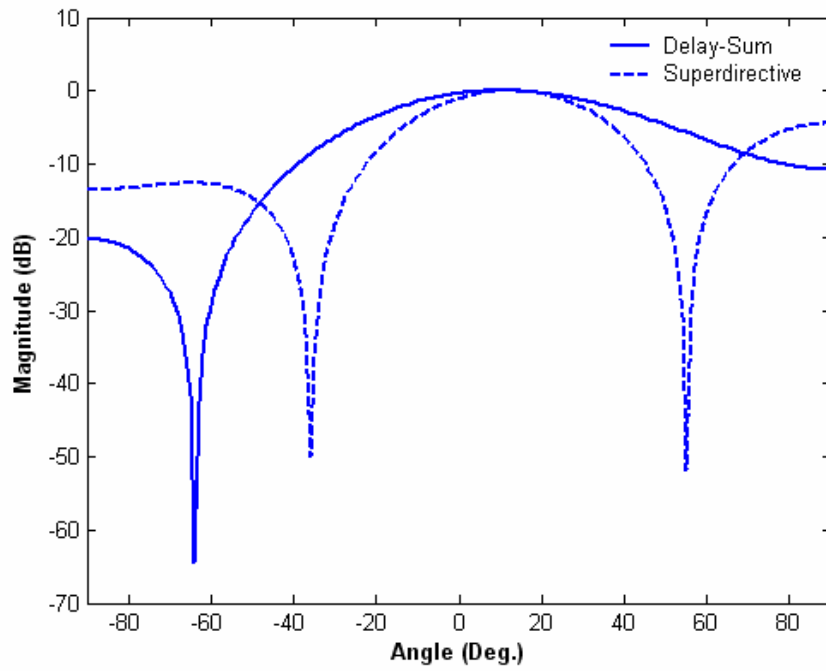
(a)
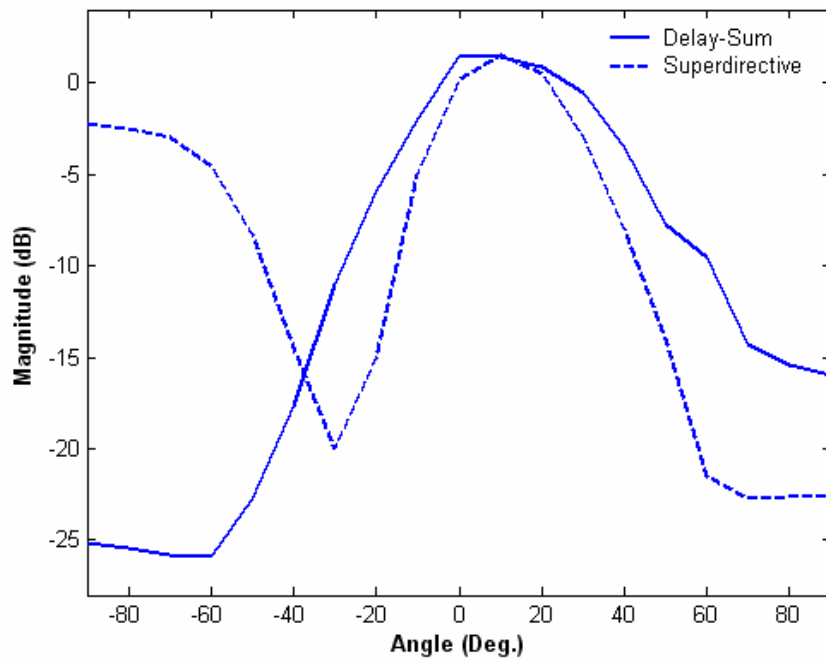
(b)

Figure 25.   The comparisons of the DOA estimation by delay-sum and superdirective

methods with a single frequency at 2000 Hz.   The sound source is

located at 10°. (a) Simulations (b) Measurements.
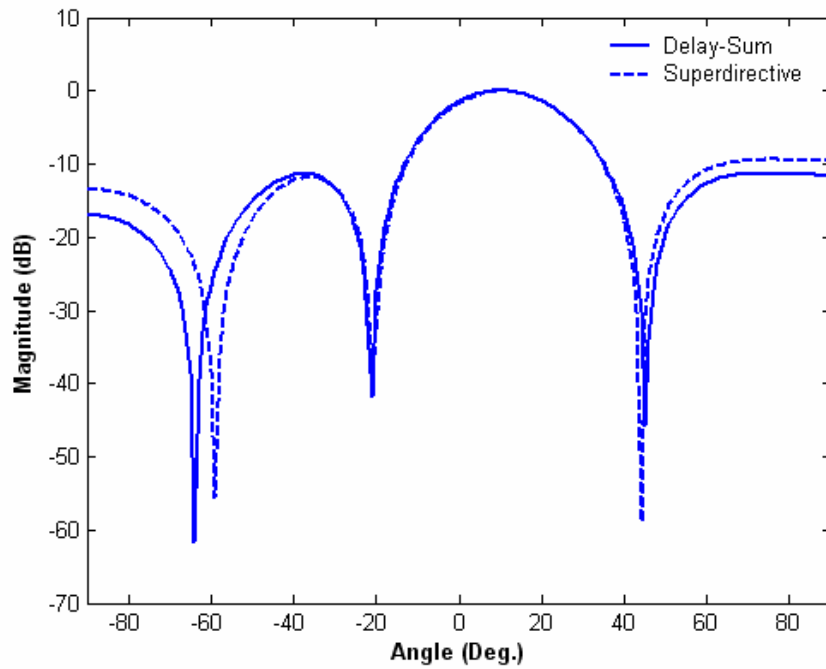
(a)
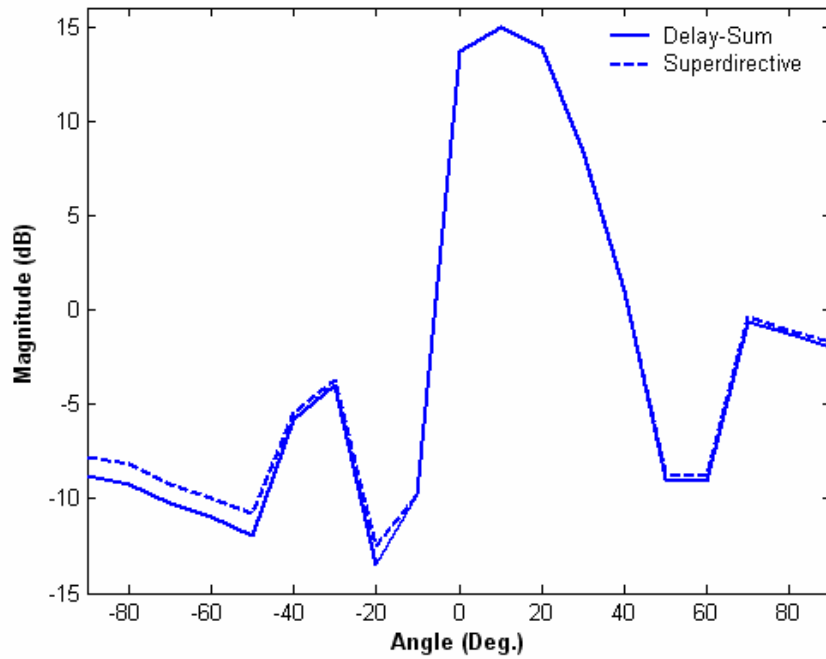
(b)

Figure 26.   The comparisons of the DOA estimation by delay-sum and superdirective

methods with a single frequency at 4000 Hz.   The sound source is

located at 10°. (a) Simulations (b) Measurements.
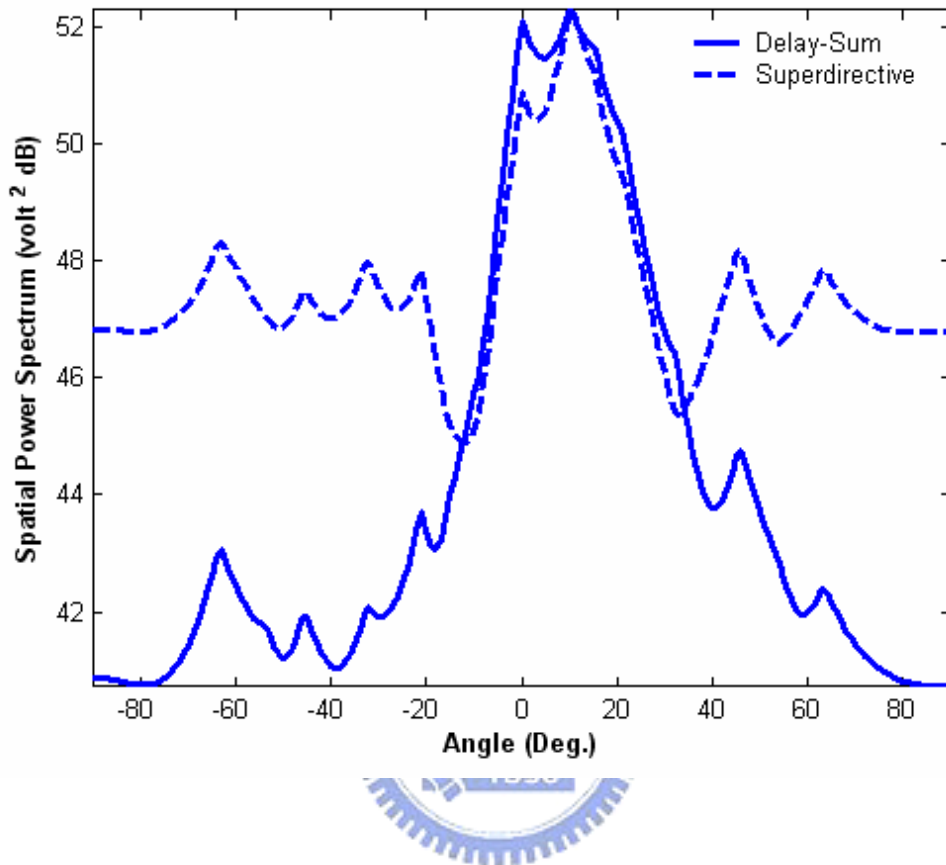
Figure 27.   The DOA estimations are simulated by MATLAB using delay-sum and

superdirective methods.   The sampling frequency is 16000 Hz.   The

number of microphones is 4 with a 12 cm aperture.
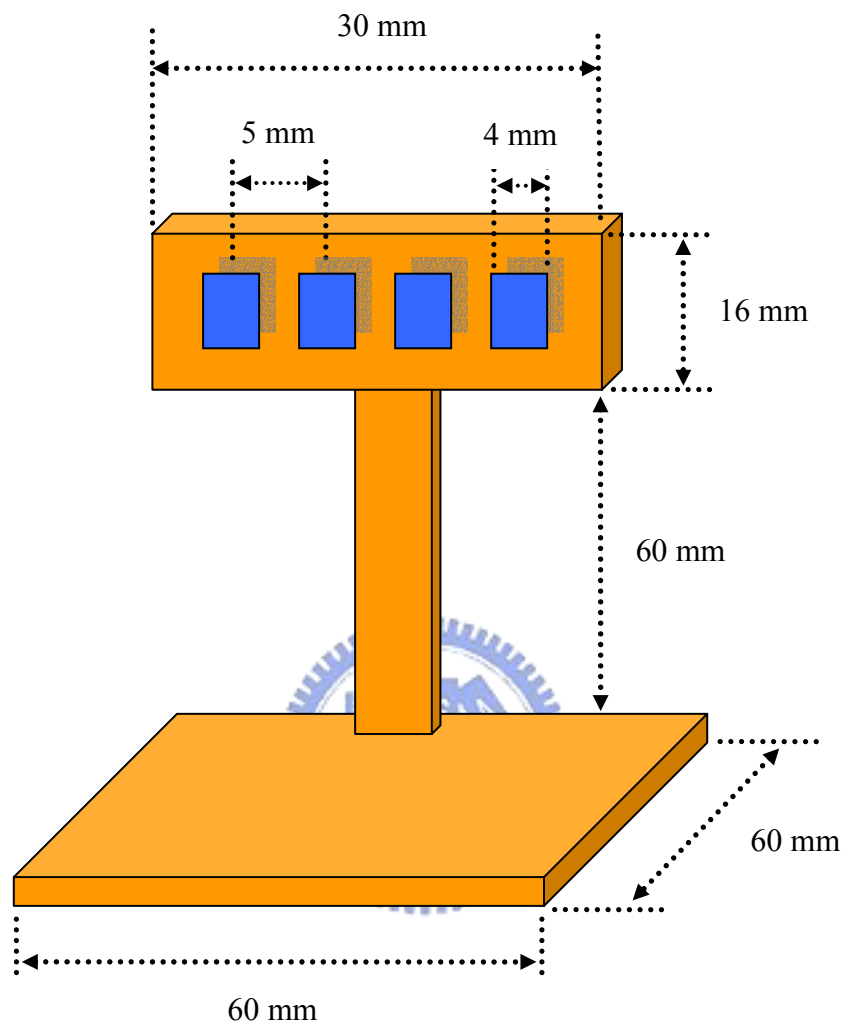
Figure 28.   The configuration of a MEMS microphone array with same spacing 5
mm and a 15 mm aperture.
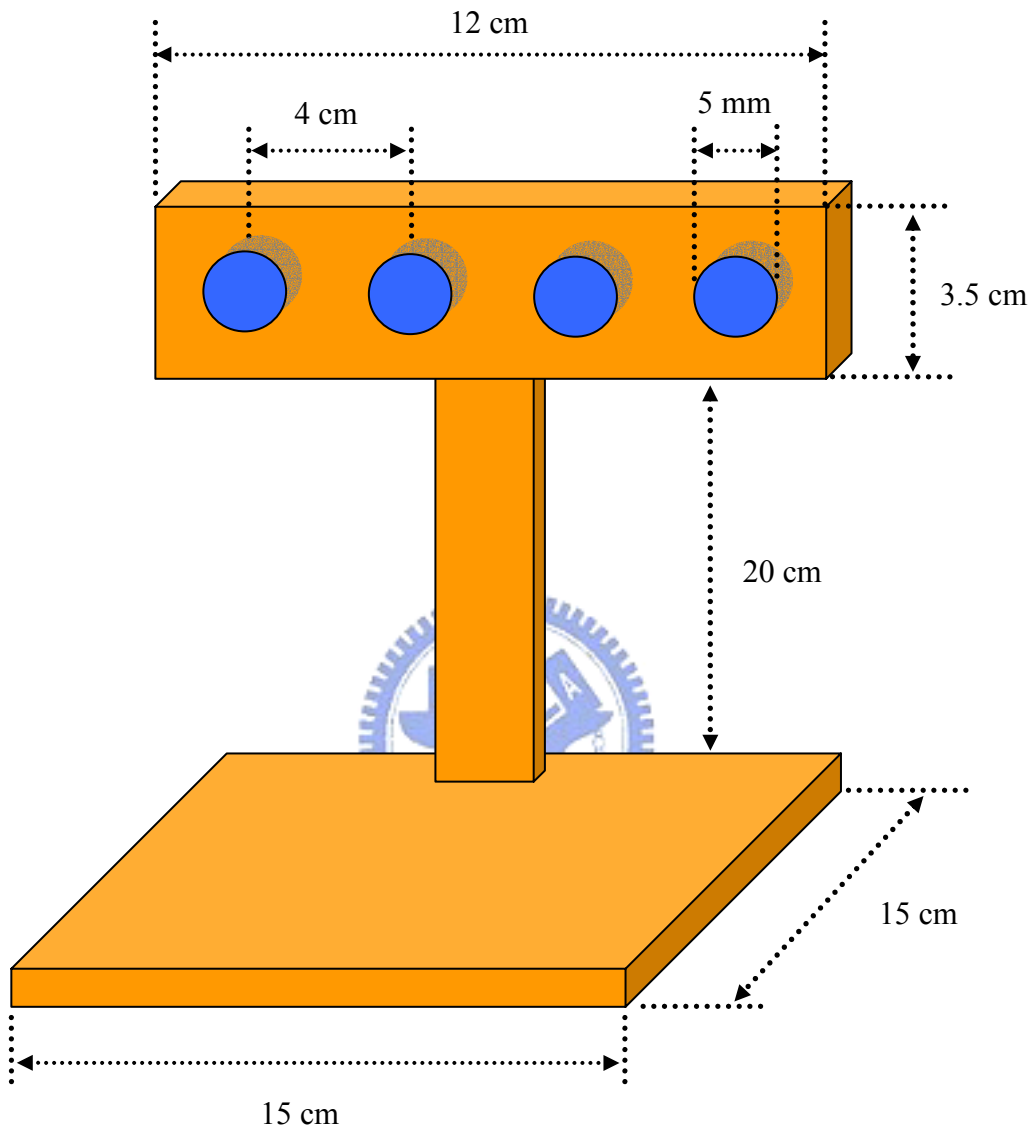
Figure 29.   The configuration of a conventional microphone array with same spacing

4 cm and a 12 cm aperture.
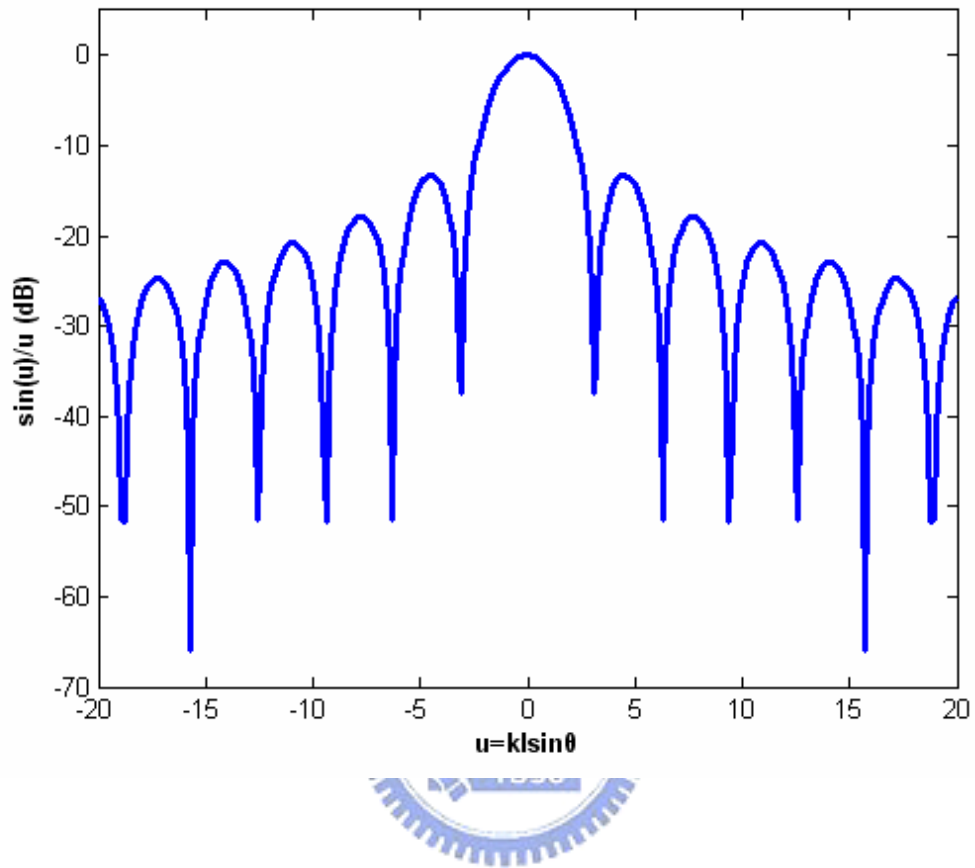
Figure 30.   The far-field radiation beam pattern of a continuous line source.
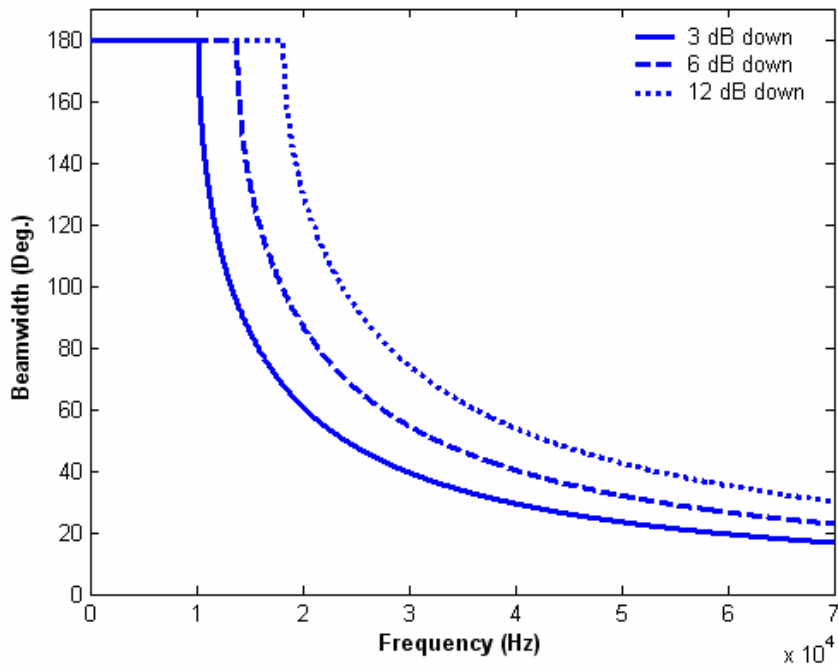
(a)
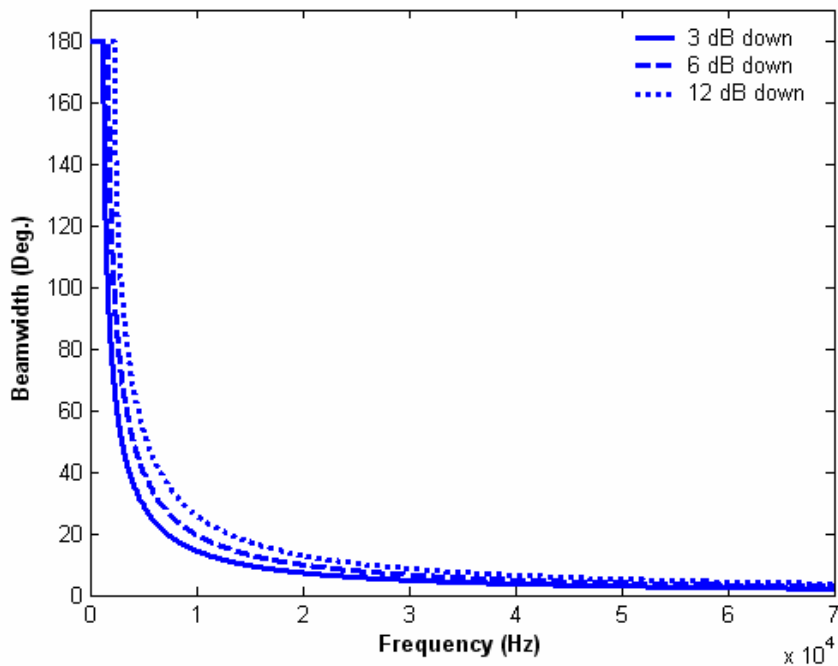
(b)

Figure 31.   The relationship between beamwidth and frequency of a continuous line

source. (a) A 15 mm aperture (b) A 12 cm aperture.
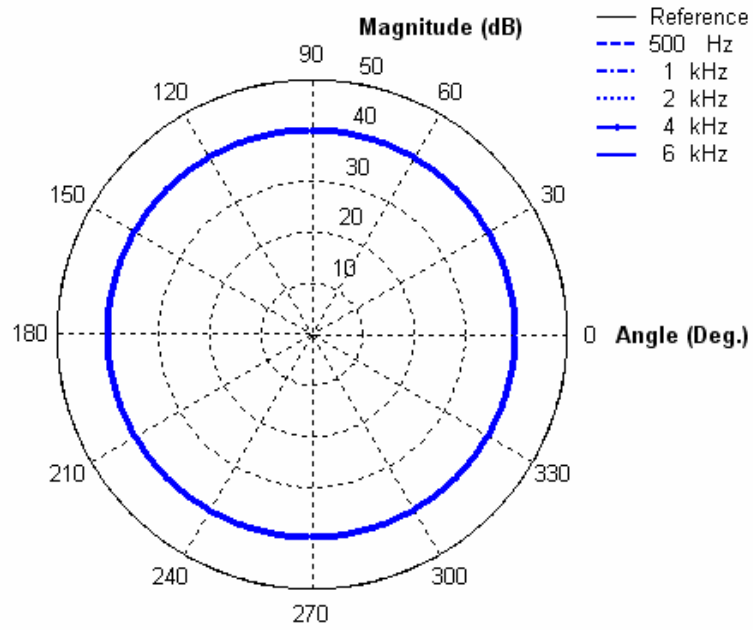
(a)
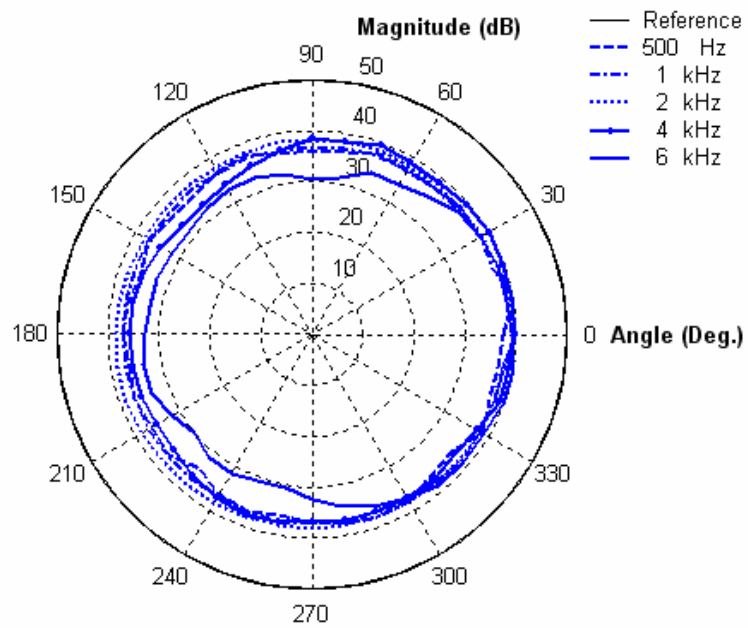
(b)

Figure 32.   The directvity of one microphone. (a) Simulations (b) Measurements.
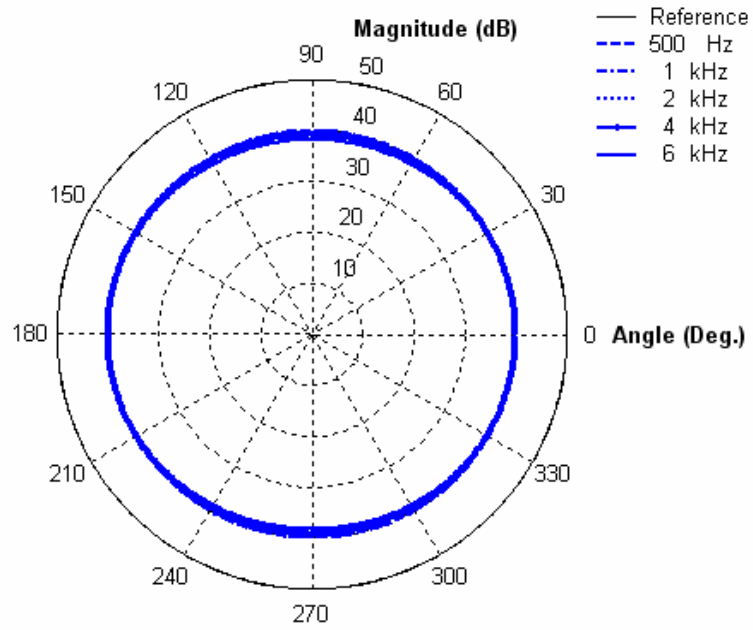
(a)
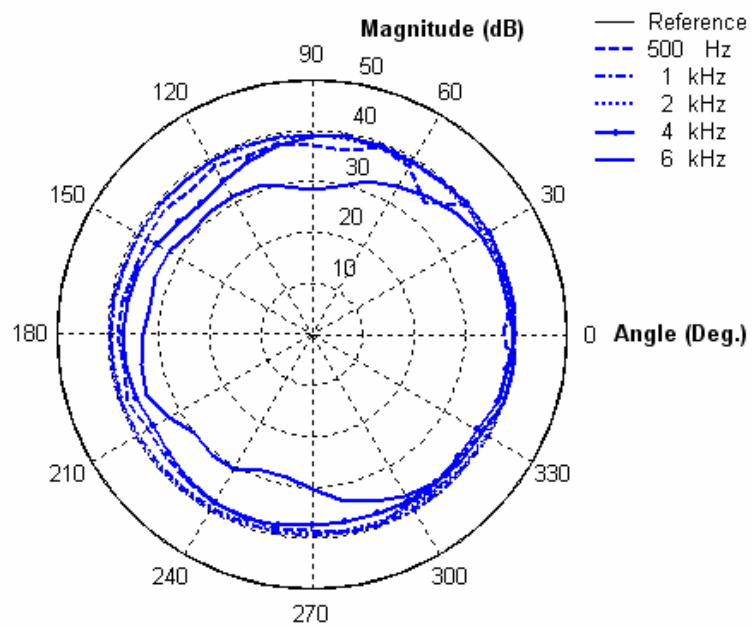
(b)

Figure 33.   The directivity of a broadside array with four MEMS microphones in a

15 mm aperture. (a) Simulations (b) Measurements.
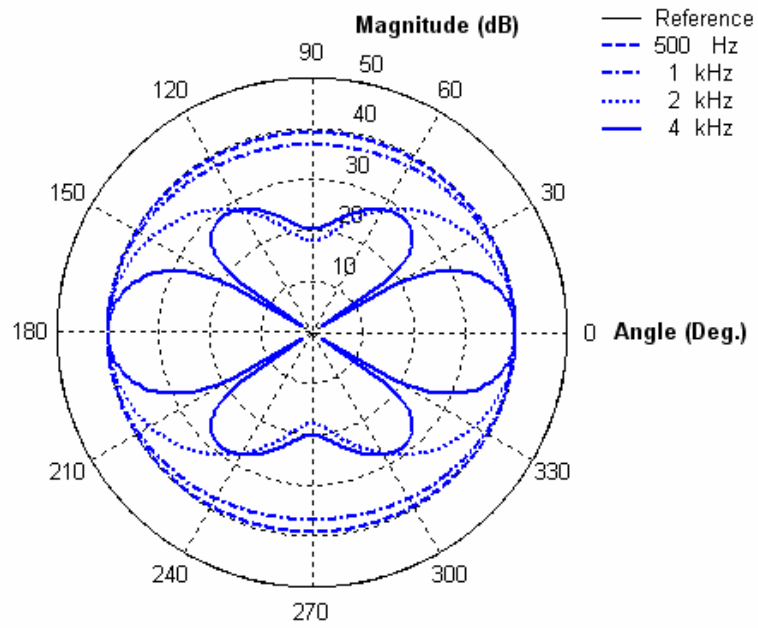
(a)
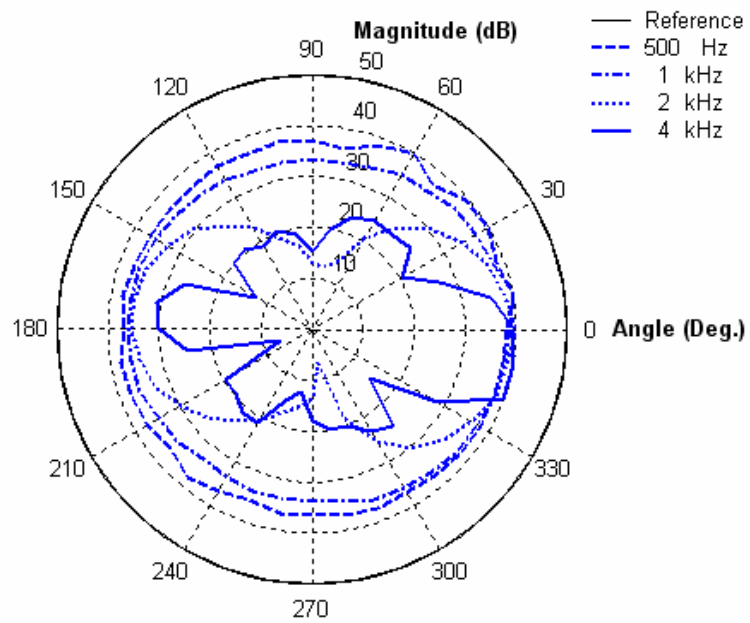
(b)

Figure 34.   The directivity of a broadside array with four conventional microphones

with in a 12 cm aperture. (a) Simulations (b) Measurements.
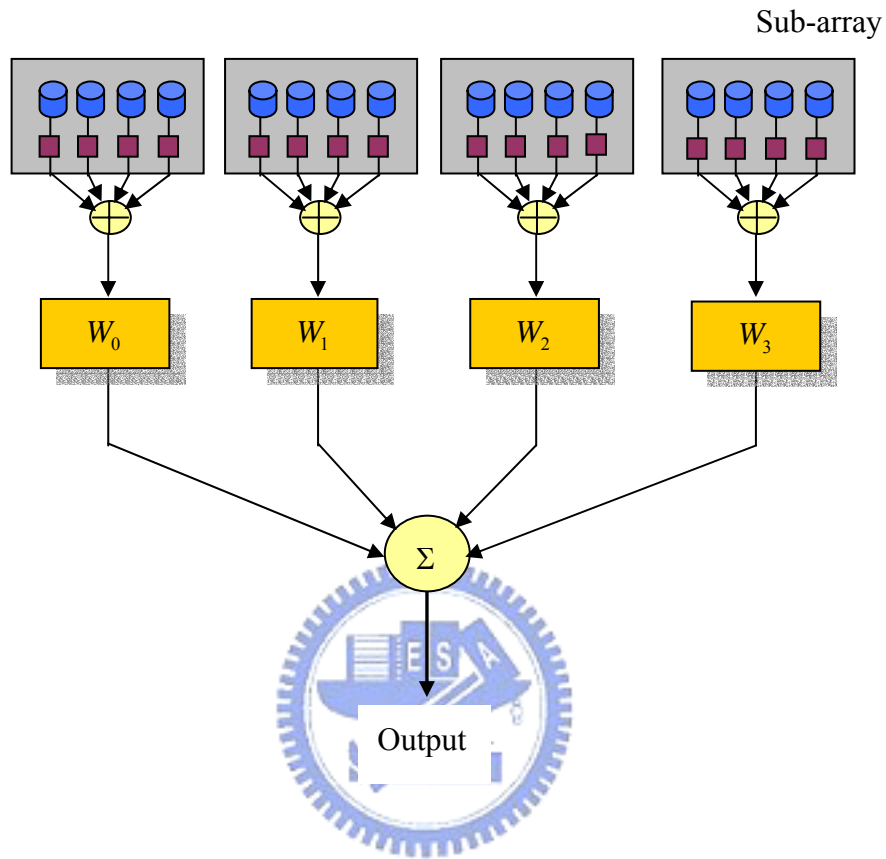
Figure 35.   A 4×4 sub-array microphones scheme.
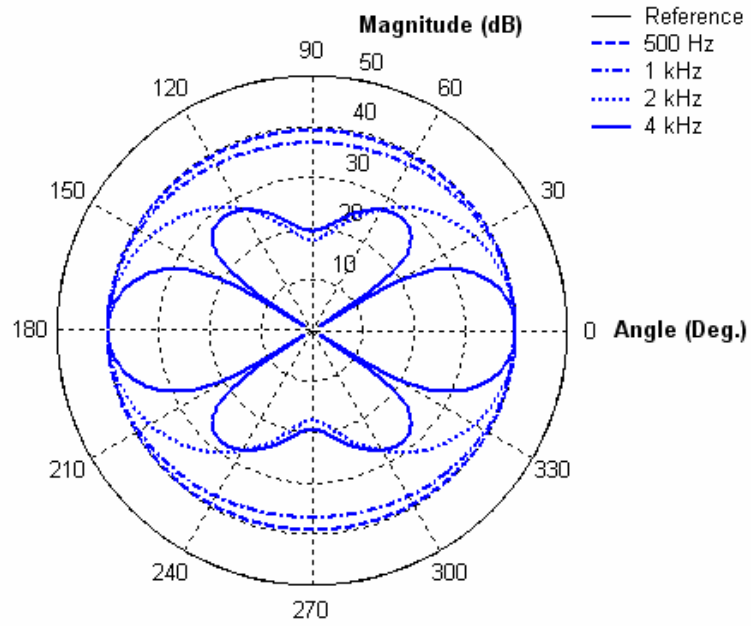
(a)
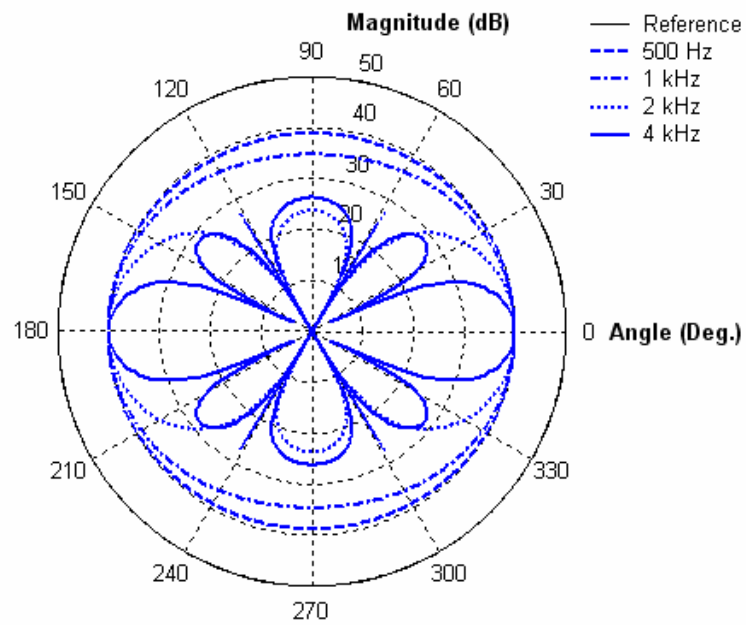
(b)

Figure 36. The directivity of 4×4 sub-array.

(a) Aperture size is 12 cm (b) Aperture size is 15 cm.
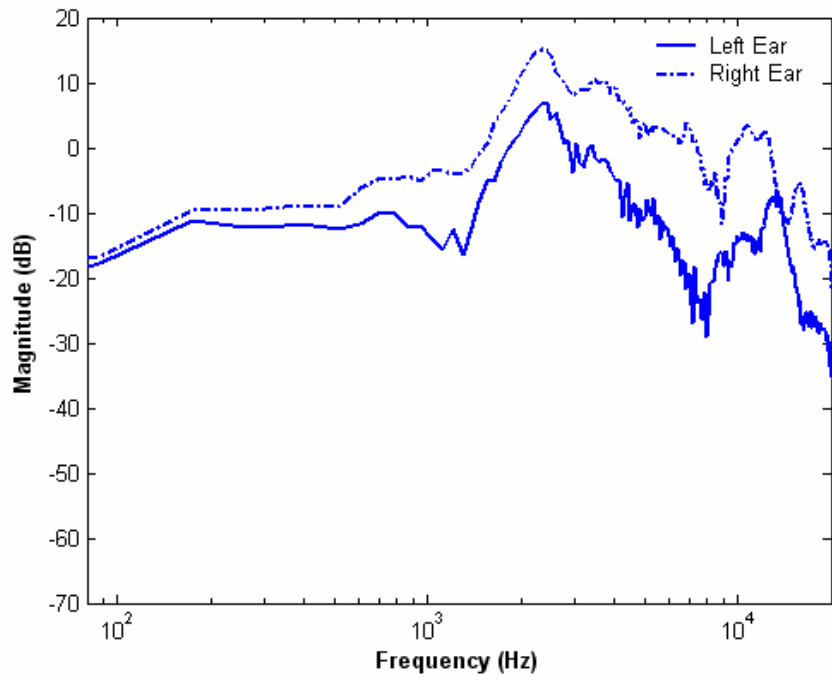
(a)
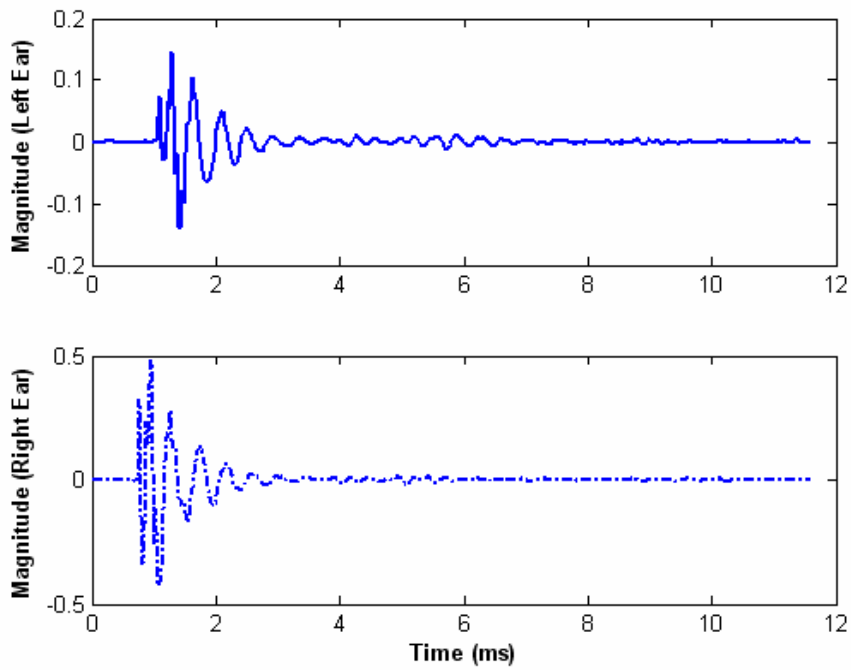
(b)

Figure 37. An HRTF example of KEMAR at elevation $=0^{\circ}$ and azimuth $=40^{\circ}$.

(a) The HRTF frequency response (b) The impulse response (HRIR).
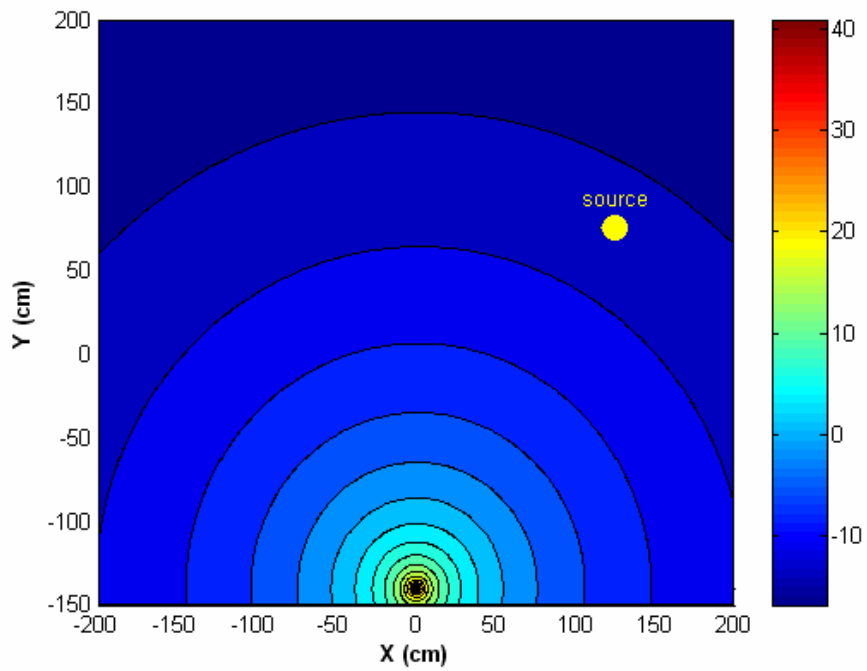
(a)
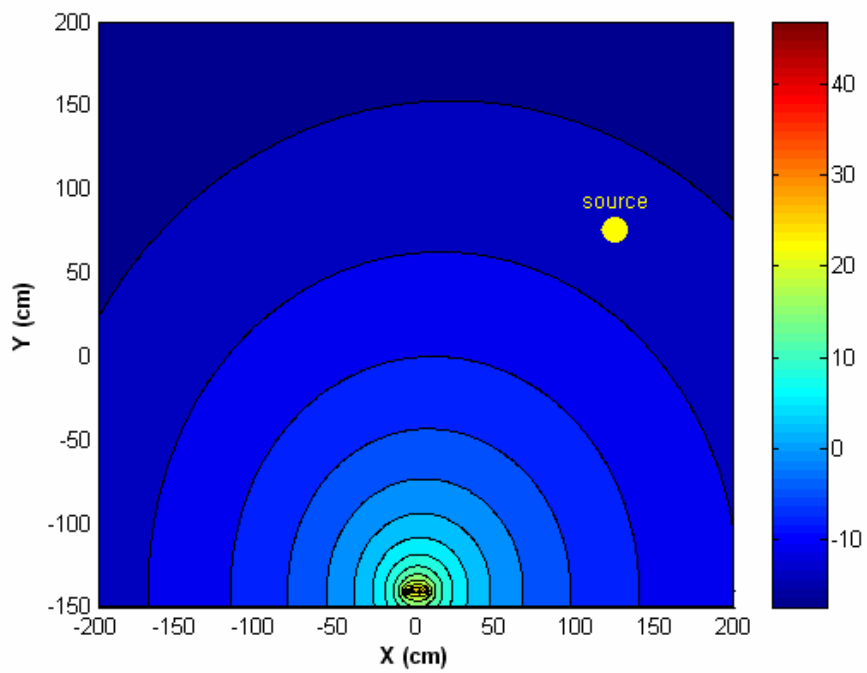
(b)

Figure 38.   The contour plots of beamforming techniques by inverse filters at 500 Hz.

The source is indicated on the plot. (a) Aperture size is 15 mm

(b) Aperture size is 12 cm.
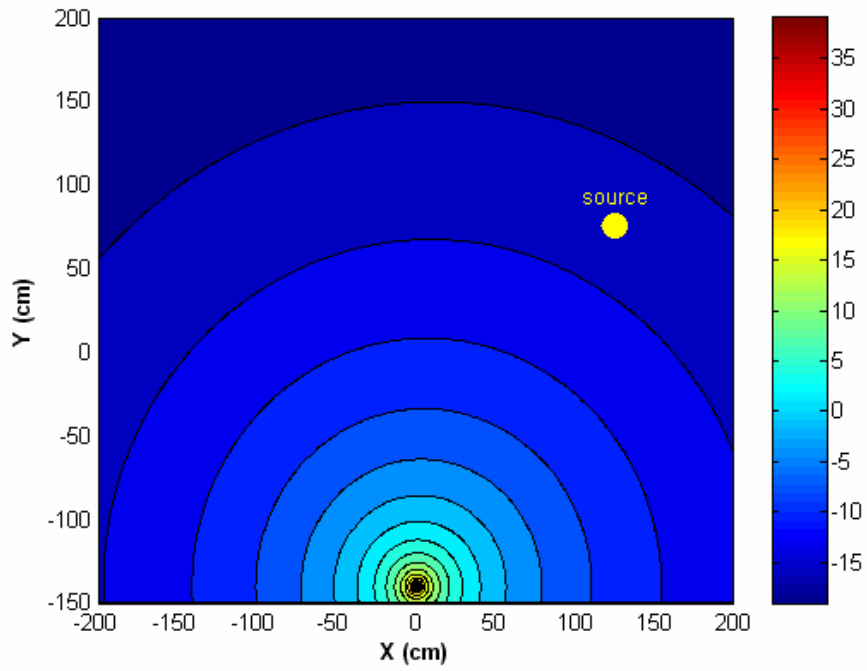
(a)
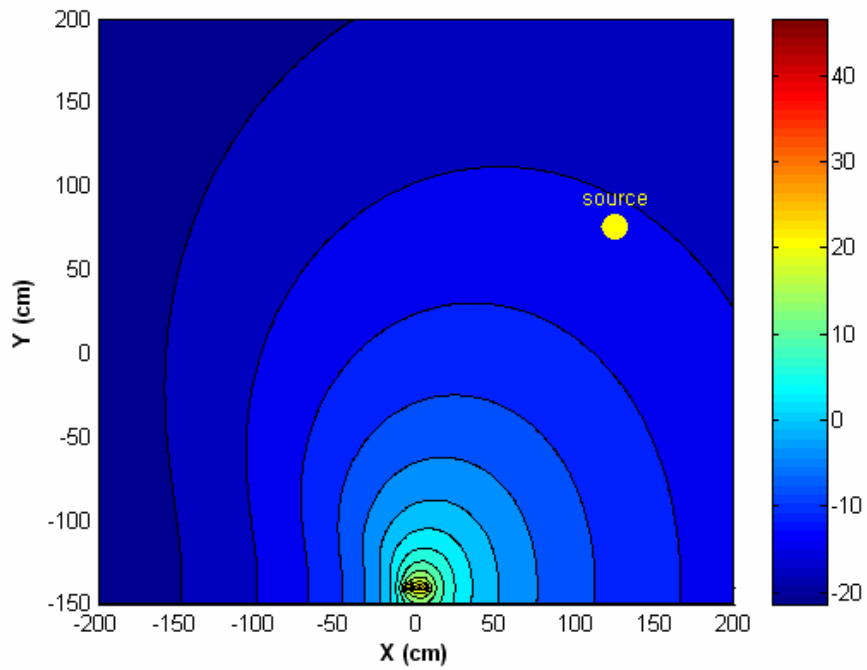
(b)

Figure 39.   The contour plots of beamforming techniques by inverse filters at 1000

Hz.   The source is indicated on the plot. (a) Aperture size is 15 mm
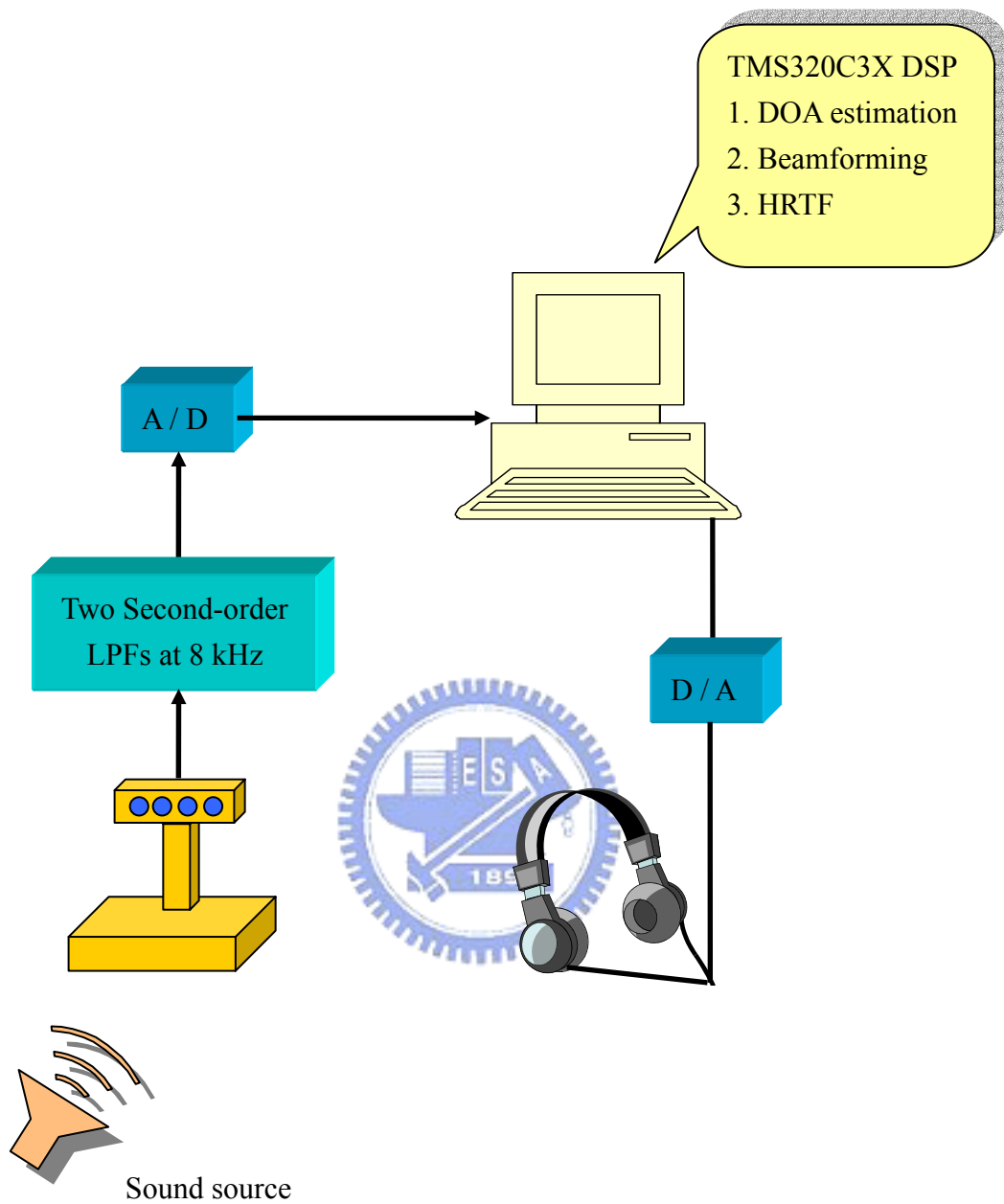
(b) Aperture size is 12 cm.
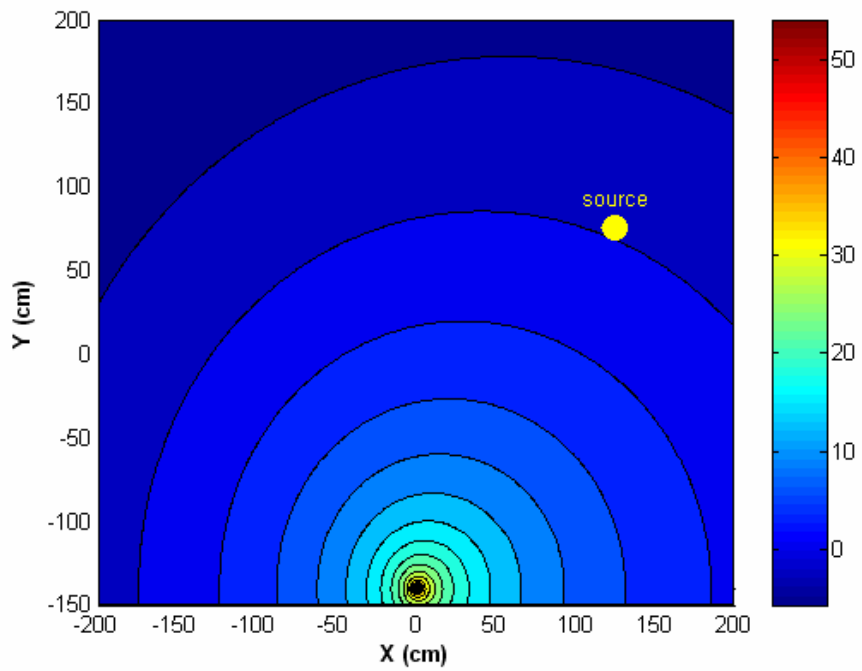
(a)
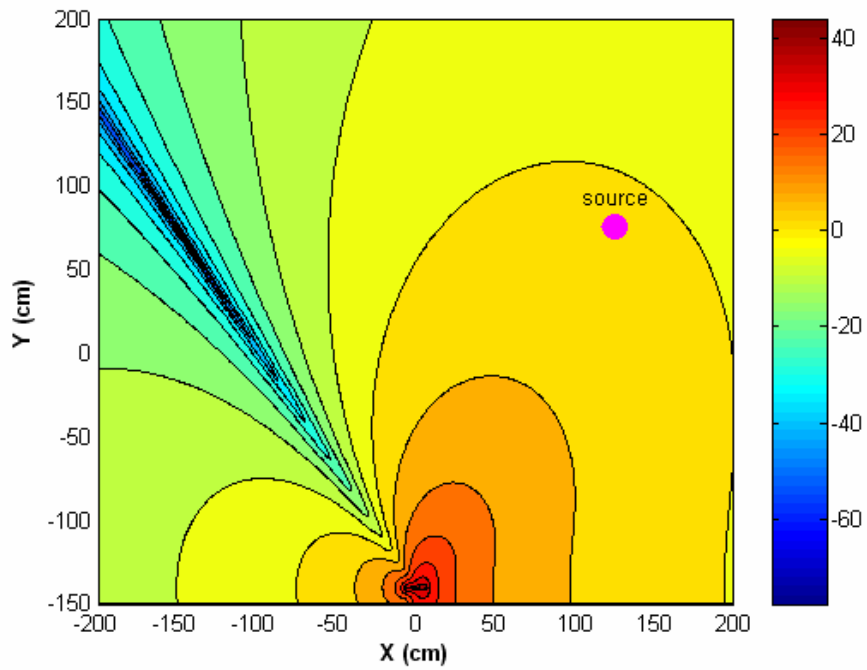
(b)

Figure 40.   The contour plots of beamforming techniques by inverse filters at 2000

Hz.   The source is indicated on the plot. (a) Aperture size is 15 mm

(b) Aperture size is 12 cm.
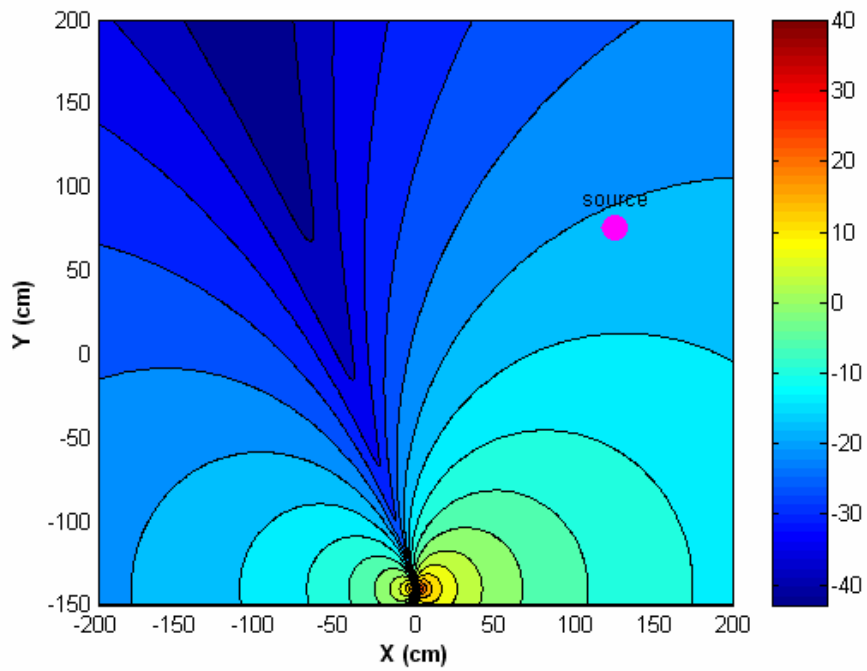
(a)
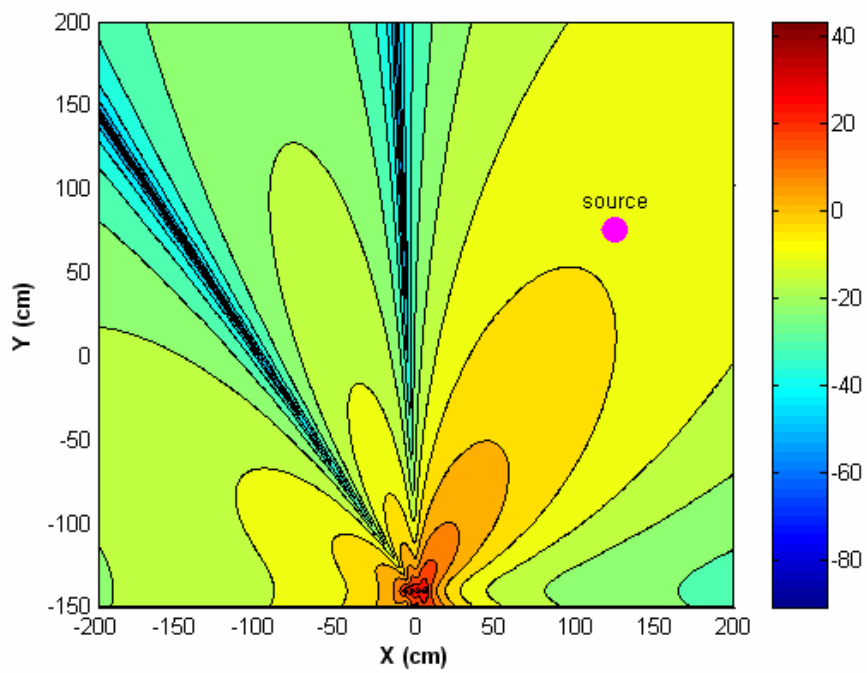
(b)

Figure 41.   The contour plots of beamforming techniques by inverse filters at 4000

Hz.   The source is indicated on the plot. (a) Aperture size is 15 mm

(b) Aperture size is 12 cm.

Figure 42.   The system block diagram of 3D spatial sound reconstruction with HRTF

implemented by TMS320C3X DSP.
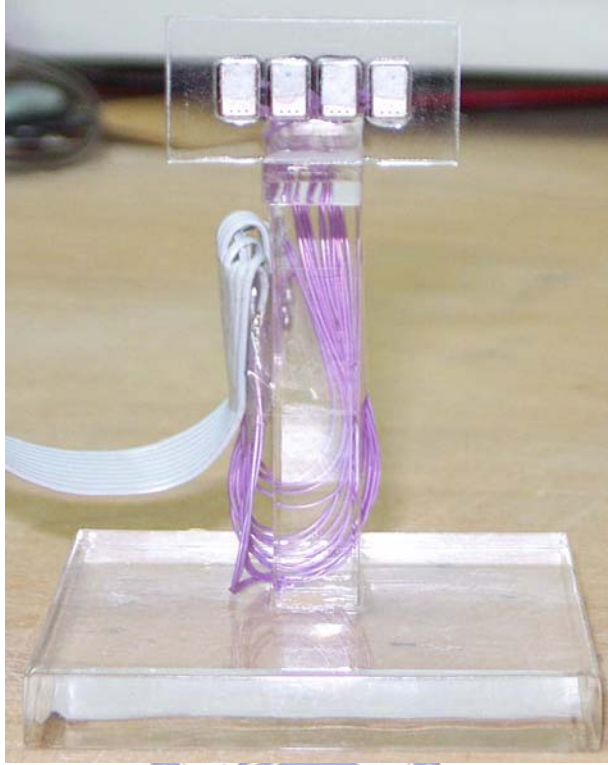
Figure 43.   The implementation of a MEMS microphone array with a 15 mm

aperture.

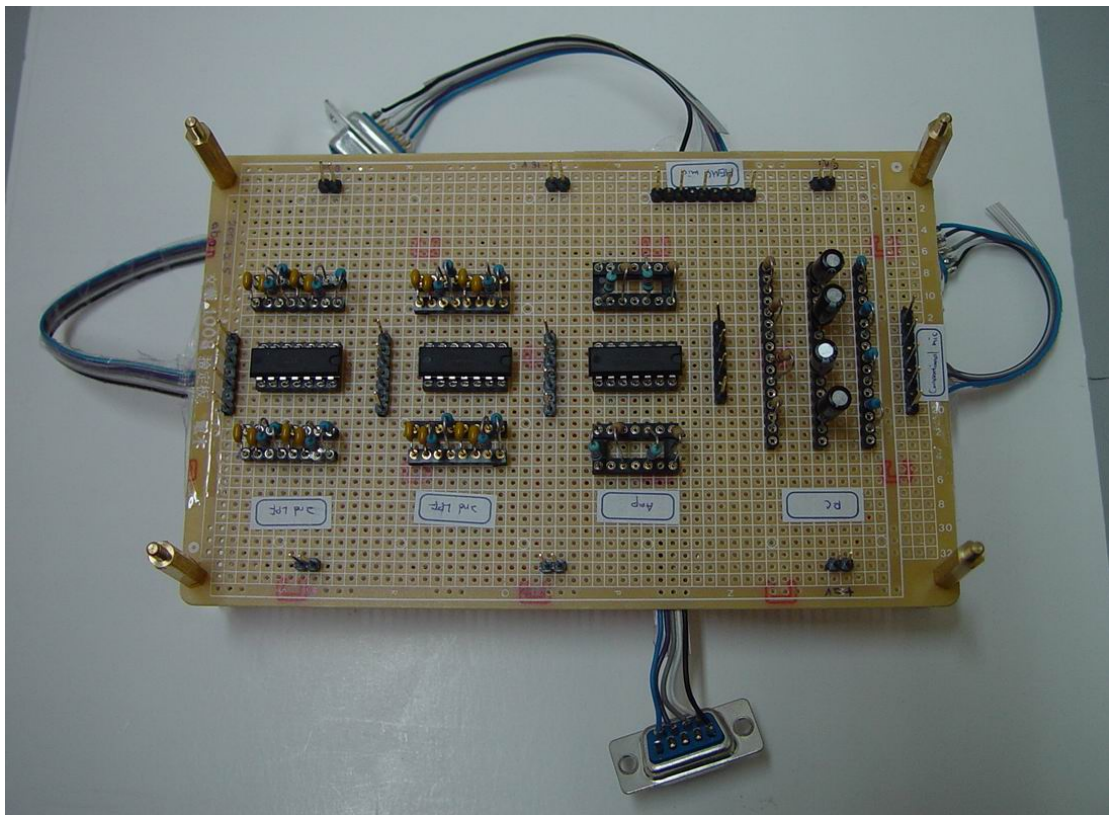Figure 44.   The implementation of a conventional microphone array with a 12 cm aperture.

Figure 45.   The implementation of the circuits with amplifiers and two second-
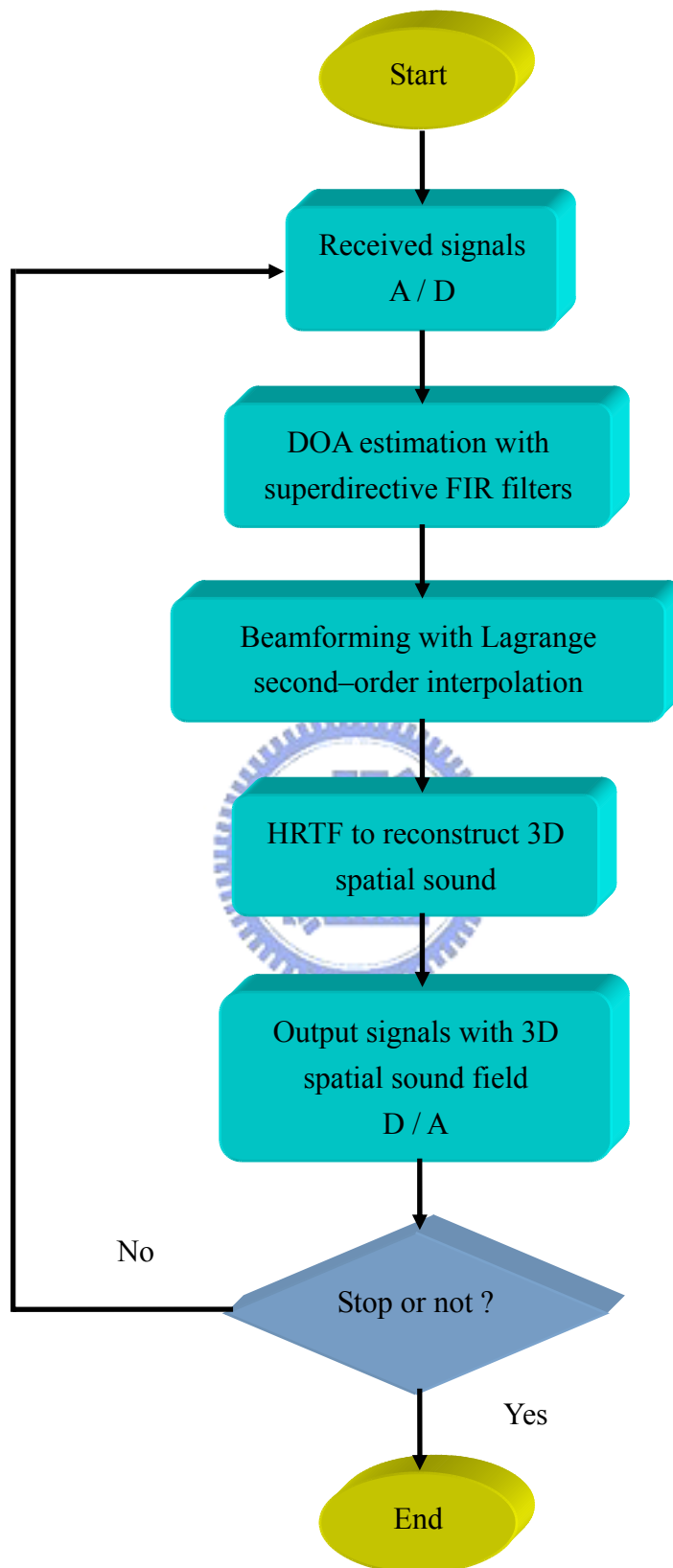order low pass filters.

Figure 46.   The flow chart of 3D spatial sound reconstruction with HRTF.
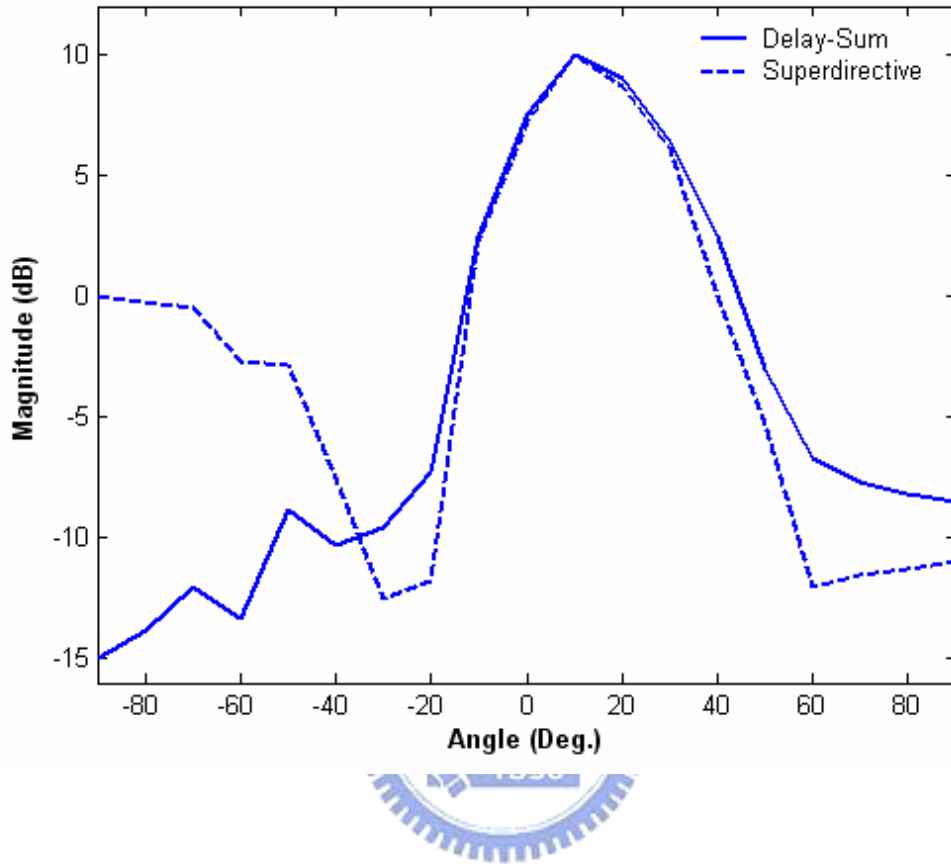
Figure 47.   The DOA estimation of the delay-sum and superdirective methods, where a sound source is located at 10° measured by the microphone array and calculated by TMS320C3X DSP.
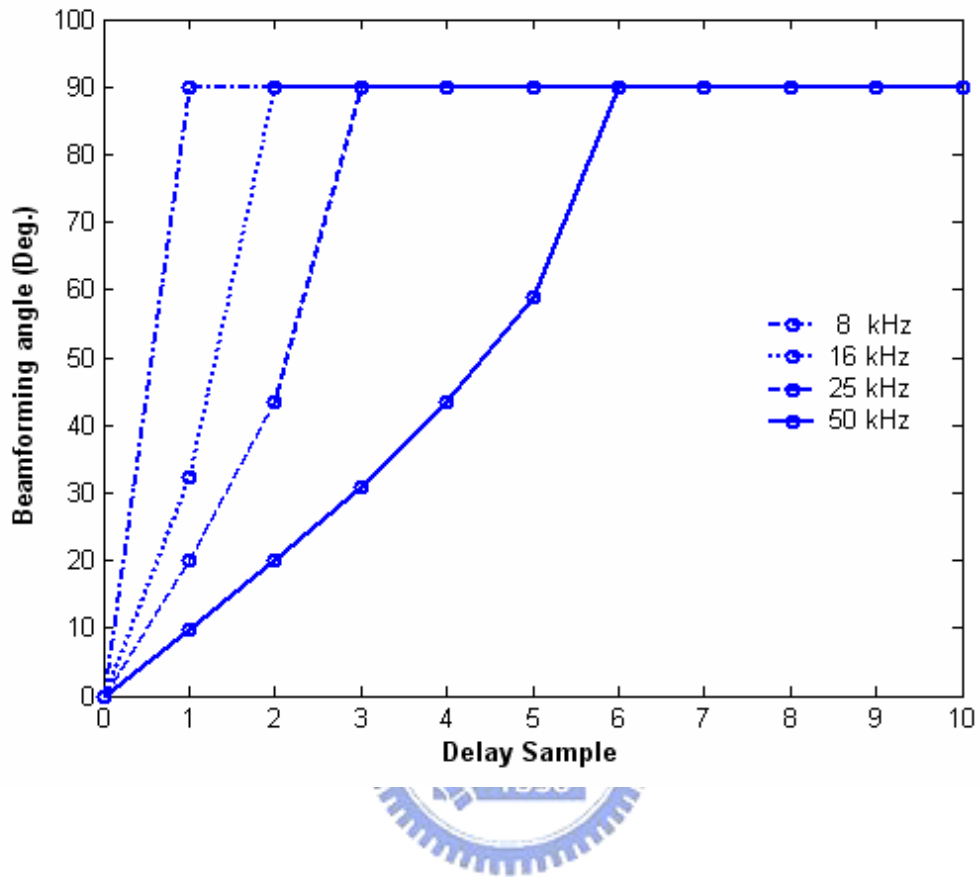
Figure 48.   The relationships of delay sample and beamforming angle with sampling
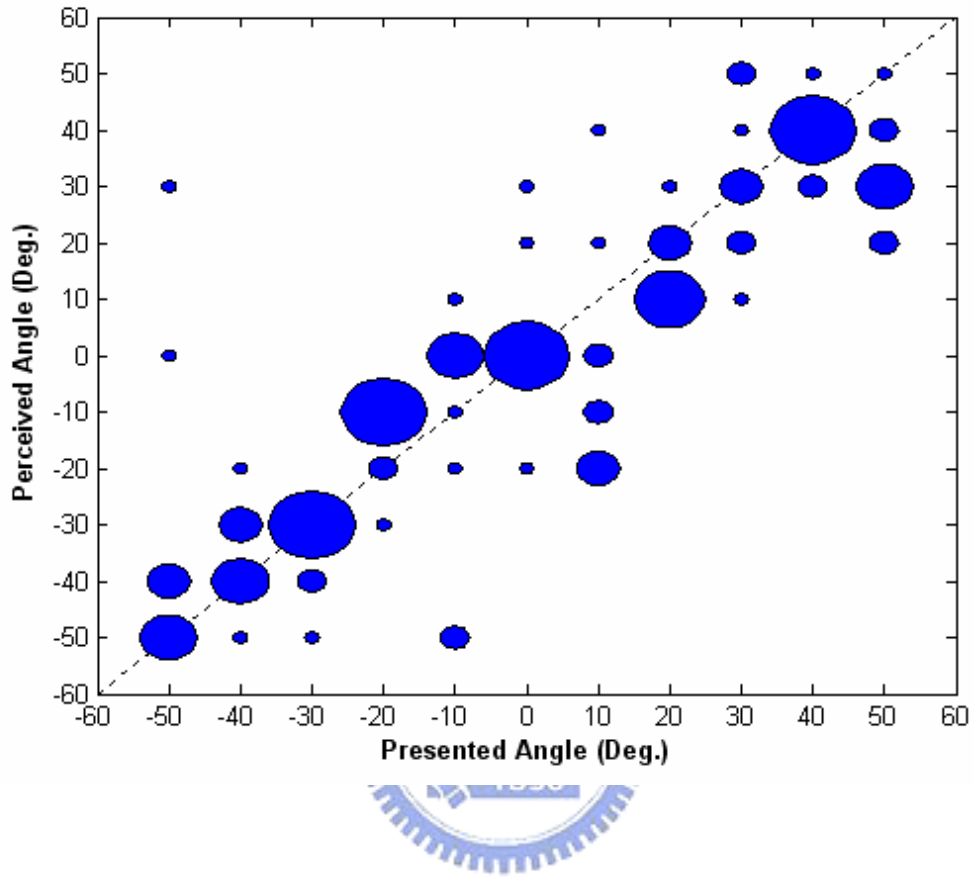
rate at 8, 16, 25 and 50 kHz.

Figure 49.   The results of listening test of the 3D spatial sound reconstruction with
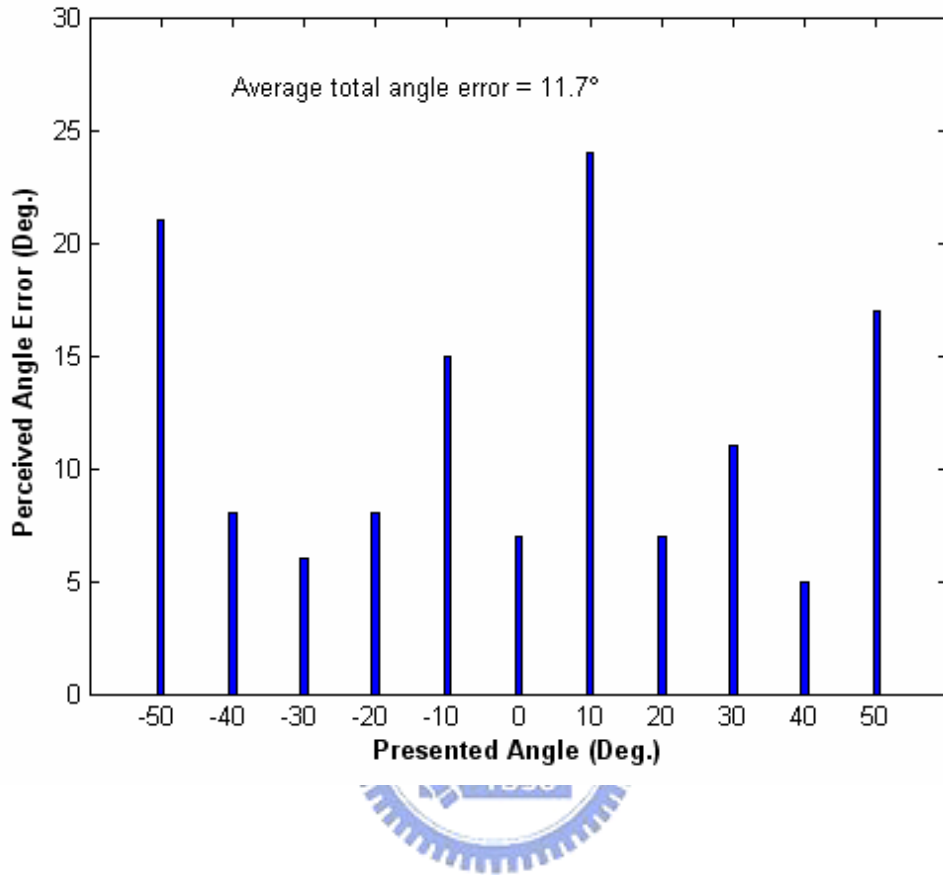
HRTF implemented by TMS320C3X DSP.

Figure 50. The perceived angle errors at the presented angle of listening test of the 3D spatial sound reconstruction with HRTF implemented by TMS320C3X DSP.