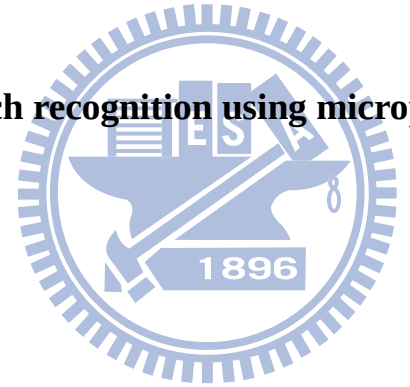
國 立 交 通 大 學

機械工程學系

碩士論文

應用麥克風陣列技術來提昇語音辨識



Robust speech recognition using microphone arrays

研 究 生: 劉嬰婷

指導教授: 白明憲

中華民國九十九年六月

應用麥克風陣列技術來提昇語音辨識

Robust speech recognition using microphone arrays

研 究 生：劉嬰婷　　Student：Ying-Ting Liou

指導教授：白明憲　　Advisor：Mingsian R. Bai

國 立 交 通 大 學

機械工程學系

碩 士 論 文

A thesis
Submitted to Department of Mechanical Engineering
Collage of Engineering
National Chiao Tung University
In Partial Fulfillment of Requirements
For the Degree of Master of Science
in
Mechanical Engineering
June 2010
HsinChu, Taiwan, Republic of China

中華民國九十九年六月

應用 ESIF 陣列技術來改善語音的品質

研究生：劉嬰婷　　指導教授：白明憲 教授

國立交通大學 機械工程學系 碩士班

摘　要

本論文提供一能提昇語音辨識率的麥克風陣列。利用指向性比一般陣列高的超指向性麥克風陣列（其為端射陣列,endfire array），能夠達到減噪的效果，特別是當噪音在陣列背後的時候。評估表現的客觀參數有三種：指向性因子（directivity index）、前後比（front-to-back ratio）以及不變的波束寬（constant beam-width），利用將上述三種客觀參數最佳化，便能得到設計出超指向性麥克風陣列的濾波器。另一方面，如果噪音並不是在陣列背後的方向，相反地，是在靠近語音訊號的位置，則需使用另一種稱之為相位差評估（phase difference estimation）的演算法來解決這種問題，此方法能在不使語音訊號失真的情況下消除噪音。本研究發現在相位差評估內的 ITD threshold 對於語音辨識提昇的效果扮演著重要的角色，因此必須要將其作一最佳化的設計，在此本研究是使用 GSS（Golden Section Search）來將其最佳化。此外，音量亦會影響辨識率，因此也必須要加以控制。如果目標訊號並不是在設計的主軸上，則必須要使用 beam-steering 的技術將主波束轉置目標訊號的位置上。最後將會分析實驗結果，並驗證本研究所提出之演算法能夠使語音辨識率大幅提昇。

Robust speech recognition using microphone arrays

Student: Ying-Ying Liou **Advisor: Mingsian R. Bai**

Department of Mechanical Engineering

National Chiao-Tung University

ABSTRACT

This paper proposes microphone array techniques aimed at enhancing speech recognition. By using super-directive microphone arrays steering to endfire, where the directivity is higher than convention arrays, the noise reduction can be reached, especially when noise is at the rear. There are three objective functions, directivity index, front-to-back ratio, and constant beam-width, to be optimized and therefore the filter can be designed. If the noise doesn't come from the rear, on the contrary, it comes from the direction closing to the target source, then the phase difference estimation is used to solve this problem, which can reduce the noise without distortion even when the angle between noise and target source is small. It is found that the ITD threshold in the phase difference estimation plays an important role in enhancing the speech recognition, and hence it has to be optimized. In this paper, GSS is used to search the optimal threshold. Moreover, the volume also affects the performance, and needs to be controlled. If the target source is not from the direction of main lobe, beam steering technique has to be applied to the system. Finally, experiment results are discussed to demonstrate that the performance of the proposed algorithm is better than conventional methods.

誌謝

短短兩年的研究生生涯轉眼即逝。在此感謝白明憲教授的諄諄教誨與照顧，在白明憲教授的指導期間，深刻的感受到教授對於追求學問的熱忱，更是佩服教授淵博的學問與解決問題的方法。在教授豐富的專業知識以及嚴謹的治學態度下，使我能夠順利完成學業與論文，在此致上最誠摯的謝意。

在論文寫作方面，感謝冀泰石教授與桑梓賢教授在百忙中撥冗閱讀，並提出寶貴的意見與指導，使得本文的內容更趨完善與充實，在此學生致上無限的感激。

在這兩年的研究生生涯中，承蒙博士班林家鴻學長、陳勁誠學長，以及已畢業的何克男學長、王俊仁學長、郭育志學長、艾學安學長、劉冠良學長在研究與學業上的適時指點，並有幸與曾智文同學、廖國志同學、桂振益同學、廖士涵同學、張濬閣同學、陳俊宏同學互相切磋討論，讓我獲益甚多。此外學弟徐偉智、王俊凱、吳俊慶、衛帝安、許書豪、學長劉志傑以及學姐李雨容在生活上的朝夕相處與砥礪磨練，亦值得細細回憶。因為有了你們，讓實驗室裡總是充滿歡笑。能順利取得碩士學位，要感謝的人很多，上述名單恐有疏漏，在此一併致上我最深的謝意。

最後僅以此篇論文，獻給我摯愛的家人及男友，爺爺劉錦雲先生、奶奶劉元嬌女士，您們慈祥的笑容及呵護，總是讓我有勇氣繼續前進。感謝母親鍾貴娥女士、父親劉得鏡先生、姊姊劉芷芸、哥哥劉康鼎、弟弟劉康佑，你們對我無微不至的包容與諄諄教誨，讓我不至於迷失了方向。感謝男友韓忠諭，總是陪伴在我身邊，聽我大吐苦水並給我最真摯的加油鼓勵。這一路上，因為有你們的付出與支持，給了我最大的精神支柱，也讓我有勇氣面對更艱難的挑戰。

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Automatic speech recognizers (ASRs) have significantly improved in recent years but the performance degrades rapidly in noisy or reverberant environments [1]. Therefore, noisy speech needs to be processed by speech improvement algorithms. For instance, the delay-and-sum (DAS) beamformer is a well known algorithm which is computational efficiency. However, it only performed well for uncorrelated noise [2]. The one-channel noise reduction (NR) technology has been widely applied in the communication community, and was expected to enhance speech recognition. Nevertheless, the improvement of one-channel NR in speech enhancement does not always translate into substantial gains in speech recognition performance, because too aggressive NR destroys the speech features. The one-channel NR encounters the dilemma of noise reduction or distortion. Therefore, microphone array is used in the proposed algorithm, which can ease the tradeoff of the above situation.

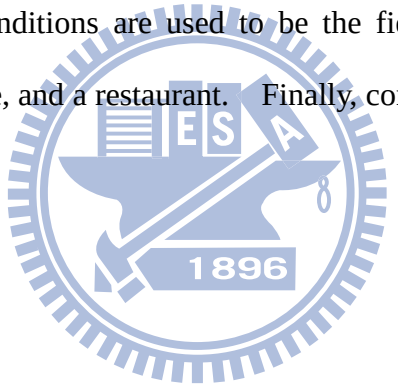
Lately, a missing-data approach was suggested to enhance speech recognition in noisy environments, based on designing whether data are reliable[3]. The performance of the missing-data approach is significantly improved comparing to that of the DAS beamformer. Nevertheless, the success of this technique depends on the sufficiency of reliable data and errors in imputation procedures affect the performance [4]. The speech recognition in the environments with non-stationary noise still remains a tough problem [3]. An alternative is the binaural processing which is well known for separating speech signals [5]. Several algorithms were discussed the phenomena of binaural system, such as interaural time difference (ITD) and interaural intensity difference (IID) [6], [7]. Recently, computational auditory scene analysis (CASA) systems were developed to construct an ideal binary mask by comparing the signals at the two microphones in binaural systems [7]-[9]. Both voice and unvoiced

speech signals could be segregated by CASA systems from a noisy environment. However, the computation of the CASA systems is quite complex.

In this study, microphone arrays are used for enhancing speech recognition in noisy and reverberant environments. Typically, there are two types of microphone array— the broadside and endfire arrays. When the maximum of the array beam pattern (the mainlobe) is alone a line perpendicular to the axial direction of the microphone array, the array is called a broadside array. On the contrary, an endfire array means that the mainlobe is in the direction diaphragm to the microphone axis, "off the end" rather than off the side and consequently the name is endfire array [10]. In Section I, super-directive microphone arrays with the type of endfire array are discussed [10], [11]. Since the directivity of super-directive microphone arrays is higher than that of a uniformly summed array in the same condition, it can not only suppress noise and reverberation coming from all directions well but also keep the feature of the target signal from the principal direction. Furthermore, although in many applications the direction of the target signal can't be predetermined, it is usually in front of the array and disturbances are at the rear. In these cases, the endfire array is suitable than a broadside array.

Section II introduces phase-difference estimation (PD), which is based on differences of arrival time of the signal that microphones received and is a kind of broadside array [12]. With the aid of phase-difference estimation, speech signal can be separated well without distortion and the recognition rate is enhanced. Because this algorithm is very sensitive to the choice of ITD threshold in binary masking criterion, how to choose ITD threshold becomes an important problem. An automatic selection of ITD threshold proposed by Kim et al is based on minimizing the cross correlation between the target and the interference signals. However, the performance of the automatic selection algorithm degrades significantly when

signal-to-noise ratio (SNR) and the subtending angle between speech and noise signal are small. Hence, this paper proposes an optimal threshold varying with the subtending angle, which is based on finding the minimum of the WER by GSS. Using the optimal ITD threshold proposed in this paper, PDE algorithm can perform well with small SNR and subtending angle. Furthermore, the selection of volume affects the performance and needs to be adjusted, which is also discussed in this paper. The speech recognition will decrease when the sound source isn't on the main lobe of microphone arrays, and the system needs the beam-steering technique to change the main lobe of array pattern by electronic compensation. The experiment results are showed in Section III in order to evaluate the proposed optimized algorithm. Three different environmental conditions are used to be the field tests which include an anechoic chamber, an office, and a restaurant. Finally, conclusion is given in Section IV.



II. SUPER-DIRECTIVE MICROPHONE ARRAYS

Super-directive microphone arrays are introduced in this section. It begins with first-order differential microphone array (DMA), a simple kind of super-directive microphone array. Second, a method of optimization of array beampattern is introduced. There are three objective functions to be maximized, including directive index (DI), front-to-back ratio (FBR) and constant beamwidth (CBW). Due to their directional and close-talking properties, they have proven essential for the reduction al feed-back in public address systems. In telephone applications, such as speakerphone teleconferencing, directional microphones are very useful but at present are seldom utilized. Since small differential arrays can offer significant improvement in typical teleconferencing configurations, it is expected that they will become more prevalent in years to come.

A. First-order difference microphone array

First-order DMAs have been discussed for more than 50 years [11], [13], [14]. Owing to the small size of 1st order DMAs, they can be used in hands-free telecommunications where the distance between microphones and speakers are quite short. Another benefit is that the directivity of 1st order DMAs is independent of frequency. The block diagram of 1st order DMA is shown in Fig. 1. For a plane wave with amplitude A and wave number k incident on a two-element array, the magnitude of output can be written as

$$|P_d| = A\omega\left(\tau + \frac{d\cos\theta}{c}\right) = A\omega\left(\tau + \frac{d}{c}\right)\left[\frac{\tau}{\tau + \frac{d}{c}} + \frac{\frac{d}{c}}{\tau + \frac{d}{c}}\cos\theta\right] \tag{1}$$

where τ is the incorporated delay, c is the speed of sound, and θis the polar angle. It

is found from Eq. (1) that the response of 1st order DMAs is in direct proportion with frequency, which means 1st order DMAs need an equalizer to balance the response.

Let $\alpha_1 = a_0 = \dfrac{\tau}{\tau + \dfrac{d}{c}}$ and $1\text{-}\alpha_1 = a_1 = \dfrac{\dfrac{d}{c}}{\tau + \dfrac{d}{c}}$ (2)

Then $a_0 + a_1 = 1$ (3)

Thus, the normalized directional response is

$$p_{Nd}(\theta) = a_0 + a_1 \cos\theta \tag{4}$$

Accordingly, the directional response can be designed by adjusting the parameter a_0.

In order to get a better directivity, there should be objective measures for analyzing the array performance. One possible measure is FBR, the microphone gain for signals propagating to the front of the microphone relative to the rear, and it is defined as

$$FBR(\omega) = \frac{\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\pi/2}\left|H(\omega,\theta,\phi)\right|^2 \sin\theta d\theta d\phi}{\frac{1}{2\pi}\int_0^{2\pi}\int_{\pi/2}^{\pi}\left|H(\omega,\theta,\phi)\right|^2 \sin\theta d\theta d\phi} \tag{5}$$

where the angles θ and ϕ are the spherical coordinate angles and $H(\omega,\theta,\phi)$ is the frequency response of the array. Fig. 2 shows the relativity between the parameter α_1 and FBR. The maximum FBR occurs when α_1 is equal to 0.366, and in this situation, the array can reject the noise from rear well.

The other measure is DI, the ratio of intensity of the acoustic beam in the measured axis to that of the entire distributed omnidirectional sound energy. It is defined as

$$DI(\omega,\theta_0,\phi_0)=10\log_{10}\left(\frac{\left|H(\omega,\theta_0,\phi_0)\right|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\left|H(\omega,\theta,\phi)\right|^2\sin\theta d\theta d\phi}\right) \quad (6)$$

where θ_0 *and* ϕ_0 are the angles at which DI is being measured. The DI of 1st order DMAs varies with the parameter α_1, and the maximum DI reaches at $\alpha_1=0.25$, which is shown in Fig. 3.

Fig. 4 is the polar plot of the absolute value of the responses. The 1st order DMAs that correspond to the maximum DI is given the name hypercardioid, and the maximum FBR value corresponds to the supercardioid design. When $\alpha_1=0$, the 1st order differential system is a dipole. At $\alpha_1=1$, the microphone is an omnidirectional microphone with 0 dB DI. A special case of $\alpha_1=0.5$ is the cardioids pattern. Although the cardioids microphone is not optimal in directional gain or front-to-back ratio, it is the most commonly manufactured differential microphone. Table I is the summary of the results for first-order microphone.

FIG. 5 is the directivity pattern of 1st order DMA in different frequency. When α_1 is fixed, the shape of the directivity pattern is almost the same no matter what frequency it is, but the gain increase as frequency increasing.

1. First order adaptive DMA

Whenever undesired noise sources are spatially non-stationary, conventional DMA has its limits in terms of interference suppression. Adaptive differential microphone arrays (ADMAs) with their null steering capabilities promise better performance. The utilization of conventional directional microphones with fixed directivity is a limited solution to this problem because the undesired noise is often not fixed to a certain angle. A better approach to solving this problem is to take advantage of the adaptive noise cancellation capabilities of DMAs in combination

with digital signal processing.

A way to circumvent the necessity to generate the delay T directly in order to obtain the desired directivity response is to utilize an adaptive back-to-back cardioids system as shown in Fig. 6. This system can be used to adaptively adjust the response of the backward facing cardioids (see Fig. 7) in order to track a possibly moving noise source in the back half plane. By choosing $T = d/c$, the back-to-back cardioids can be formed directly by appropriately subtracting the delayed microphone signals.

The output response of the first-order ADMA can be obtained as follows:

$$c_F(f) = A\left(1 - e^{-j\omega\left(\tau + \frac{d\cos\theta}{c}\right)}\right) = 2jAe^{-j\omega\left(\tau + \frac{d\cos\theta}{c}\right)/2}\sin\left(\frac{\omega\left(\tau + \frac{d\cos\theta}{c}\right)}{2}\right) \tag{7}$$

$$c_B(f) = A\left(e^{-j\omega\tau} - e^{-j\omega\frac{d\cos\theta}{c}}\right) = 2jAe^{j\omega\left(\tau + \frac{d\cos\theta}{c}\right)/2}\sin\left(\frac{\omega\left(\frac{d\cos\theta}{c} - \tau\right)}{2}\right) \tag{8}$$

$$\begin{aligned} y(f) &= \beta c_B(f) - c_F(f) \\ &= 2jAe^{-jkd/2(1-\cos\theta)}\left(\beta\sin\left(\frac{kd(\cos\theta - 1)}{2}\right) - \sin\left(\frac{kd(\cos\theta + 1)}{2}\right)\right) \end{aligned} \tag{9}$$

The single independent null angle θ_1 of the first-order ADMA, which in this work is assumed to be placed into the back half plane of the array $(90° \le \theta_1 \le 180°)$, can be found by setting Eq. (9) to zero and solving for $\theta = \theta_1$. Therefore,

$$\theta_1 \approx \arccos\frac{\beta - 1}{\beta + 1} \tag{10}$$

for small spacing and delay are assumed. A selection of directivity patterns that can be obtained by a first-order ADMA is depicted in Fig. 8.

In an time-varying environment, it is advisory to use an adaptive algorithm in order to obtain the update of the parameter β. For this matter, the normalized least-mean-square (NLMS) adaptive algorithm is utilized, which is computationally

inexpensive, easy to implement and which offers reasonably fast tracking capabilities. Here, the real valued time-domain one-tap NLMS algorithm can be written as

$$y(i) = c_F(i) - \beta(i) c_B(i) \tag{11}$$

$$\beta(i+1) = \beta(i) + \frac{\mu}{a + \|c_B(i)\|^2} c_B(i) y(i) \tag{12}$$

where $c_F(i)$ and $c_B(i)$ are the values for the forward and the backward facing cardioids signals at time instance *i*, $0 < \mu < 2$ is the adaptation constant and $a > 0$ is a small constant.

B. Optimization of array beampattern

Another kind of super-directive microphone array is optimization of array beampattern. The signal model is shown in Fig. 9. In the frequency domain, the output signal is given by

$$y(\omega) = \mathbf{w}^H \mathbf{x} = \mathbf{w}^H \mathbf{a}(\omega, \theta, \phi) s(\omega) \tag{13}$$

where **w** denotes the frequency-domain coefficients of the beamformer, the operator *H* denotes a conjugated transposition, **x** denotes the microphone signal and $\mathbf{a}(\omega, \theta, \phi)$ is the manifold (steering) vector.

Let $H(\omega, \theta, \phi) = \mathbf{w}^H(\omega) \mathbf{a}(\omega, \theta, \phi)$ (14)

which is the frequency response of the microphone array.

then $y(\omega) = H(\omega, \theta, \phi) s(\omega)$ (15)

The aim here is to estimate the coefficient of the filter **w** by using the objective functions, which include DI, FBR and CBW.

1. Maximum for directive index (MDI)

The first objective function is DI, the equation introduced in previous section can be written as a Rayleigh quotient for two Hermitian quadratic forms as

$$DI\left(\omega,\theta_0,\phi_0\right)=10\log_{10}\left(\frac{\mathbf{w}^H\mathbf{A}\mathbf{w}}{\mathbf{w}^H\mathbf{B}\mathbf{w}}\right) \tag{16}$$

where

$$\mathbf{A}=\mathbf{a}\left(\omega,\theta_0,\phi_0\right)\mathbf{a}^H\left(\omega,\theta_0,\phi_0\right) \tag{17}$$

$$\mathbf{B}=\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\mathbf{a}\left(\omega,\theta,\phi\right)\mathbf{a}^H\left(\omega,\theta,\phi\right)\sin\theta d\theta d\phi \tag{18}$$

w is complex weighting applied to the microphones and *H* is the complex conjugate transpose. For spherically isotropic noise (diffuse field),

$$\mathbf{a}(\omega,\theta,\phi)=[1\quad e^{jkd\cos\theta}\cdots e^{j(M-1)kd\cos\theta}]^T \tag{19}$$

and

$$B_{mn}=\frac{\sin[(m-n)kd]}{(m-n)kd}=\mathrm{sinc}[(m-n)kd] \tag{20}$$

The maximum of the Rayleigh quotient is reached at a value equal to the largest generalized eigenvalue of the equivalent generalized eigenvalue problem, i.e.,

$$\max_{\mathbf{w}} DI\left(\omega,\theta_0,\phi_0\right)=\max_{\mathbf{w}}10\log_{10}\left(\frac{\mathbf{w}^H\mathbf{A}\mathbf{w}}{\mathbf{w}^H\mathbf{B}\mathbf{w}}\right) \tag{21}$$

$$\Leftrightarrow \quad \min_{\mathbf{w}}\mathbf{w}^H\mathbf{B}\mathbf{w} \qquad \text{st. } \mathbf{w}^H\mathbf{A}\mathbf{w}=1 \tag{22}$$

$$\Leftrightarrow \quad \lambda_{\max} \text{ of } \mathbf{A}\mathbf{w}=\lambda\mathbf{B}\mathbf{w} \tag{23}$$

where λ is the general eigenvalue and **w** is the corresponding general eigenvector. The eigenvector corresponding to the largest engenvalue will contain the coefficients attaining MDI.

2. Maximum for front-to-back ratio (MFBR)

In many applications the target sources are in front of the microphone array and

almost all interferences are in the rear, hence FBR is appropriate to describe the influence of the array. The definition of FBR given in previous section can be written as

$$FBR(\omega) = \frac{\mathbf{w}^H \mathbf{A} \mathbf{w}}{\mathbf{w}^H \mathbf{B} \mathbf{w}} \tag{24}$$

where

$$\mathbf{A} = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi/2} \mathbf{a}(\omega,\theta,\phi) \mathbf{a}^H(\omega,\theta,\phi) \sin\theta d\theta d\phi \tag{25}$$

$$\mathbf{B} = \frac{1}{2\pi} \int_0^{2\pi} \int_{\pi/2}^{\pi} \mathbf{a}(\omega,\theta,\phi) \mathbf{a}^H(\omega,\theta,\phi) \sin\theta d\theta d\phi \tag{26}$$

The value of MFBR is the same as the largest eigenvalue of the equivalent generalized eigenvalue problem $\mathbf{A}\mathbf{w} = \lambda \mathbf{B}\mathbf{w}$ as in the case of MDI.

3. Maximum for constant beamwidth (MCBW)

The beamwidth of most conventional arrays decreases when frequency increases and therefore the received signal varies with position in the beam [15]. The inverse proportionality of beamwidth relating to frequency is usually insignificant in narrowband beamformers but causes the outer portion of the main lobe subjected to lowpass filtering in broadband beamformer. Unfortunately, many applications like telecommunication systems need broadband operation. As a result, it is necessary to design a broadband directional array whose beamwidth is independent of frequency. The constant beamwidth (CBW) can be expressed as

$$CBW(\omega) = \frac{\int_0^{2\pi} \int_0^{\theta_1} |H(\omega,\theta,\phi)|^2 \sin\theta d\theta d\phi}{\frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} |H(\omega,\theta,\phi)|^2 \sin\theta d\theta d\phi} \tag{27}$$

which can be written as the ratio of two Hermitian quadratic forms as

$$CBW(\omega) = \frac{\mathbf{w}^H \mathbf{A} \mathbf{w}}{\mathbf{w}^H \mathbf{B} \mathbf{w}} \tag{28}$$

where

$$\mathbf{A} = \int_0^{2\pi} \int_0^{\theta_1} \mathbf{a}(\omega,\theta,\phi)\mathbf{a}^H(\omega,\theta,\phi)\sin\theta d\theta d\phi \tag{29}$$

$$\mathbf{B} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \mathbf{a}(\omega,\theta,\phi)\mathbf{a}^H(\omega,\theta,\phi)\sin\theta d\theta d\phi \tag{30}$$

MCBW can be obtained by calculating the maximum eigenvalue of the equivalent generalized eigenvalue problem as shown in previous sessions.

C. Super-directive microphone array with equalizer

FIG. 10 shows the contour plot of super-directive microphone arrays, where the x-axis represent the direction of source, y-axis is the frequency, and the maximum of the contour plot equals to 1. Namely, if the direction is fixed, the gain of the contour plot in different frequency means the frequency response. On the other hand, if the frequency is fixed, the gain in different direction represents the directivity pattern. It is noticed that the gain in low frequency is lower than that in high frequency whichever kind of super-directive microphone array it is. Moreover, the gain of all methods in the rear is smaller than that in the front, which corresponds with the characteristics of endfire array.

To observe the behavior of super-directive microphone arrays, white noise is used as the input signal. The definition of white noise is that it has constant power per unit frequency and exhibits a flat spectral density. Therefore, the power spectral density (PSD) of white noise has a zero slope [16]. Fig. 11 is the PSD of the output response in the look direction of super-directive microphone arrays. It can be found that the magnitudes of all methods are not very “flat”, that is, it dependents on frequency. As a result, the output signal may get some distortion. To solve this problem, the equalizer is applied into the system,

Let $G = \frac{1}{|H|}$ (31)

which is the equalizer suggested here, and G is converted to the time-domain finite-impulse-response (FIR) filters with the aid of IFFT and circular shift in real-time implementation, which is introduced in FIG. 12.

D. Simulated and experimental results

The super-directive microphone arrays are applied to the ASR system to reduce the noise and reverberation. A 2-channel endfire array is used for these simulations. The spacing between microphones is 0.5cm and the sampling rate is 8 kHz. The input sources are 50 comments (547 wave files) which are served as plane wave. The target signal is placed in 0 degree (look direction), and the noise signal (white noise) is placed in 0, 45, 90, and 180 degrees. The conventional recognizer is used to refer to a continuous density Hidden Markov Model (HMM) based ASR using MFCCs as features [19]-[21]. Fig. 13 compares recognition rate for several of the algorithms discussed in this chapter. "Original" refers to the input noisy signal. "max. DI", "max. FBR", "const. BW" refer to the design of optimal beampattern maximum for DI, FBR, and CBW that mensioned in previous section. "Delay&Subtract" refers to the conventional 1st order DMA. As can be seen, the recognition rate increases about 0 to 40% when the noise is located at 180 degrees and it increases about 20-60% when the noise is placed at 90 degrees. The results reveal that the noise reduction can be achieved by using the super-directive microphone arrays.

FIG. 14 is the result with different noise signal— car noise. It indicates that all of the super-directive microphone arrays perform better than the original noisy signal when noise is located at 180° and 90°. A special outcome is that the 1st order DMA gets the best performance when the noise comes from 90°.

III. PHASE-DIFFERENCE ESTIMATION (PDE)

The block diagram of PDE algorithm is illustrated in FIG. 15[12]. The noisy signal received in two microphones is first segmented to frames by applying a moving Hamming window and then transferred to the time-frequency domain by Short-Time Fourier transform (STFT) as follows:

$$P_1(k,l) = X(k,l) + \sum_{i=1}^{V} N_i(k,l) \tag{32}$$

$$P_2(k,l) = X(k,l) + \sum_{i=0}^{V} e^{-j\omega_k d_i(k,l)} N_i(k,l) \tag{33}$$

where k is the frequency index and l is the frame index, $X(k,l)$ and $N_i(k,l)$ represent the speech and the ith noise signals, respectively, $P_1(k,l)$ and $P_2(k,l)$ are the signals at the first and second microphone, and $\omega_k = 2\pi k / N$ for $0 \le k \le N/2-1$, where N is the STFT size. The frame length here is 75ms and the hub size is half of frame length. It is assumed that the target signal is at the location alone the perpendicular bisector of the line between two microphones, and therefore its ITD is equal to zero. On the other hand, $d_i(k,l)$ is the ITD of the ith noise signal dependent on time and frequency. If a time-frequency bin (k_m, l_m) is controlled by a strongest interference source n, the above equations can be approximated as

$$P_1(k_m, l_m) \approx N_n(k_m, l_m) \tag{34}$$

$$P_2(k_m, l_m) \approx e^{-j\omega_{k_m} d_n(k_m, l_m)} N_n(k_m, l_m) \tag{35}$$

and the ITD of this bin can be estimated by calculating the unwrapped phase difference between two microphones:

$$\left|d_n\left(k_m,l_m\right)\right| \approx \frac{1}{\left|\omega_{k_m}\right|}\min_r\left|\angle P_1\left(k_m,l_m\right)-\angle P_2\left(k_m,l_m\right)-2\pi r\right| \tag{36}$$

Then, a binary mask can be formulated as

$$B\left(k_m,l_m\right)=\begin{cases}1, & \text{if } \left|d_n\left(k_m,l_m\right)\right|\le\tau\\ 0.01, & \text{otherwise}\end{cases} \tag{37}$$

where τ is the ITD threshold. It means that only bins with its ITD smaller than τ are supposed to belong to the target signal. Correspondingly, the speech signal $S\left(k,l\right)$ is re-established from multiplying the average signals of the two microphones $\overline{P}\left(k,l\right)$ by the mask $B\left(k_j,l_j\right)$ got in above formula.

$$\overline{P}\left(k,l\right)=\frac{1}{2}\left\{P_1\left(k,l\right)+P_2\left(k,l\right)\right\} \tag{38}$$

$$S\left(k,l\right)=B\left(k,l\right)\overline{P}\left(k,l\right) \tag{39}$$

Finally, the enhanced speech signal is converted to the time-domain with the aid of inverse fast Fourier transform (IFFT) and overlap addition (OLA) method. In this paper, three approaches of technical refinement are exploited to enhance the aforementioned PDE algorithm. As shown in Fig. 16, after the received signal is transformed to the time-frequency domain, the system estimates the speech and noise location. The subtending angle between speech and noise is used to select the corresponding optimal ITD threshold searched by GSS. If the target source is not from the designed direction, beam steering technique is applied to orient the main lobe to the target source location. After IFFT and OLA, the time domain signal is scaled to the optimal volume to further increase the WRR.

A. Optimization of the ITD threshold using GSS

As mentioned previously, the parameter τ is used in the binary mask principle as

the ITD threshold having profound impact on mask estimation and hence on the performance of the speech recognition. As expected, it is found that this parameter is related to the included angle between speech and noise sources. Therefore, it is worth exploring how to adjust this parameter such that the recognition rate can be maximized. In the following, a procedure based on the GSS is presented for automated tuning of the ITD threshold.

1. Golden section search

The goals of GSS are to get an optimal reduction factor for a search interval and to minimize the number of the iterations [20]. By GSS, the minimum can be searched efficiently within a finite number of steps, and do not need to evaluate numerical gradients. Assume a function $f(x)$ is continuous and having only one minimum over the interval [*a*, *b*]. An interior point *c* is between *a* and *b*, and

$$\frac{c-a}{b-a}=w, \qquad \frac{b-c}{b-a}=1-w \tag{40}$$

where $0<w<\frac{1}{2}$. Suppose another interior point *d* is over [*c*, *b*], and

$$\frac{d-c}{b-a}=z \tag{41}$$

Notice that the choosing of *d* is applied the same strategy as that of *c*, which means

$$\frac{z}{1-w}=w \tag{42}$$

For minimizing the number of the iterations, the fraction 1-*w* musts equal to *w*+*z*, i.e. the new point *d* is the symmetric point of *c* in the interval[*a*, *b*], namely

$$z=1-2w \tag{43}$$

Comparison of Eqs. (42) and (43) yields the following quadratic equation

$$w^2 - 3w + 1 = 0 \tag{44}$$

and the root

$$w = \frac{3-\sqrt{5}}{2} \approx 0.382 \tag{45}$$

is used. Note that the number is related to the golden ratio *g*, where

$$g = \frac{\sqrt{5}+1}{2} = \frac{1}{1-w} \tag{46}$$

Therefore it's called "golden section search". Now comparing $f(c)$ and $f(d)$, if $f(c) < f(d)$, then the new interval is [*a*, *d*]; otherwise, it becomes [*c*, *b*]. The rule at each stage is to keep a center point lower than the two outside points. The process above iterates until the interval is tolerably small, and the question here is how to decide the time to stop the iteration. According to Taylor's theorem, the value of the function $f(x)$ near x_m is approximately

$$f(x) \approx f(x_m) + \frac{1}{2} f''(x_m)(x - x_m)^2 \tag{47}$$

If $f(x)$ is enough close to $f(x_m)$, then the second term can be quite small and negligible, which can be represented as

$$\frac{1}{2} f''(x_m)(x - x_m)^2 < \varepsilon \left| f(x_m) \right| \tag{48}$$

where εis usually set to 10^{-2} for single precision.

2. The Optimal ITD threshold varying with the included angle

The searching process of the optimal ITD threshold by GSS is shown in FIG. 17, where the noise type is babble at SNR 6dB and the included angle is 15 degrees. The SNR here were conducted according to ITU P.56 standard, which defined as

$$SNR = 10 * \log_{10}\left(\frac{x^2}{n^2}\right) \tag{49}$$

where x and n represent the speech signal and noise respectively. Fig. 18(a) shows the performance of PD algorithm where the ITD threshold τ varies from 0.1 to 1.5. It can be found that the recognition rate gets better by increasing τ but decreases sharply whenτ exceeds a value which differs with the included angle. It turns out that there is a relation between τ and the included angle. To find the optimal ITD threshold, GSS is used in this paper, which can quickly search the local minimal of a function in an interval. The result of the optimal τ found by GSS is shown in Fig. 18(b). The included angle is from 15 degrees to 90 degrees at SNR 0dB and 6dB, and "babble" noise is used as the noise source. It indicates that the optimal thresholds τ at SNR 0dB and 6dB are similar to each other, which means the influence of SNR is small and can be disregard. Because the curve of the optimal τ has an obvious trend, it can be fitted by a polynomial of low degree easily. A polynomial fitting of degree 2 is shown in Fig. 18(b), which is found to be

$$\tau(i) = \left(-7.76 * 10^{-5}\right) i^2 + \left(1.69 * 10^{-2}\right) i - \left(5.45 * 10^{-2}\right) \tag{50}$$

where i is the included angle. It revealed that, by using a polynomial fitting, it can use only 3 parameters to represent the optimal τ varying with the included angle very well. The relations between the effective beamwidth corresponding to the optimal ITD threshold τ of PDE algorithm and the real subtending angles are summarized in TABLE II. By comparing the effective beamwidth and the real spanning angles, the effective beamwidth is smaller but the differences become smaller as the subtending angle decreases. The reason is that, the effective beamwidth has to be smaller than the real subtending angle, or the noise will be received in the binary mask, while if the effective beamwidth is too small, some speech signals will not be picked up in the

binary mask and some feature will lose. For ASR, preservation of speech features is crucial. Loss of speech features causes the WRR to markedly decrease. Even if the noise is close to the target source, the effective beamwidth can not be too small.

B. Volume scaling

During the simulation, it is found that the speech recognition relates to the volume, especially at low SNR, and it needs to be adjusted. The adjustments start with normalization of the signal by the maximum in time-domain, and then multiply a gain to the signal. As showed in Fig. 19, when SNR is at 0dB, the recognition rate can vary from 82.8% to 91.6% for different signal volumes. It indicates that the volume indeed affects the recognition rate, and no matter what SNR it is, all results have the same tendency— the maximum recognition rate happens when the largest gain of the time domain signal is 0.125. For searching the optimal volume precisely, again, GSS is used here. The results searching by GSS show that the largest gain of the time domain signal should be 0.079 at SNR 0dB, and then the arrays get best performance. On the other hand, the largest gain equals to 0.11 at SNR 6dB is the best choice. That means, when SNR becomes smaller, the largest gain should be smaller, too.

C. Beam steering

The beam steering technique is discussed in this section to overcome the problem about the movement of the target source. With the aid of electronic compensation, the direction of the main lobe of the microphone array pattern can be changed. Assume the angle to be moved is θ_M, then the beam steering filters are given as

$$w_n = e^{-j\,n\,k\,d\sin\theta_M} = e^{\frac{-j\omega\ f_s\ nd\sin\theta_M}{c}} \tag{51}$$

where n is array index, ω is the frequency index, and f_s is the sampling rate, d is the spacing between microphones. In time domain, the beam steering filter can be written as a delay:

$$delay = \frac{f_s \; nd \sin\theta_M}{c} \tag{52}$$

That is, by applying different delays to the signal received in every microphone, the direction of main lobe can be controlled and steered to any desired directions. One thing has to be noticed is that these delays are not integer delays, hence Lagrange interpolation [18] is used here to interpolate fractional delay values, which is easier to achieve and more flexible. Simplicity, it can approximate a fractional delay by a FIR filter,

$$h(n) = \prod_{\substack{k=0 \\ k \neq n}}^{N} \frac{D-k}{n-k} \quad \text{for } n = 0,1,2...N \tag{53}$$

where N is the order of the filter. The case N=1 corresponds to linear interpolation between two samples, which suffices when the sampling frequency is high enough. The result is in Fig. 19 with the target source angle from 15° to 75° aside the main lobe. When the target source is far from the main lobe, the recognition rate degrades correspondingly. By using beam steering technique, the performance is enhanced obviously, as shown in Figure 20.

D. Simulated and experimental results

The simulated and experimental results are presented in this section. The input stimuli are 50 commends (547 wave files) rendered from a point source placed at 90 degrees (the look direction). The speech recognizer is based on continuous density Hidden Markov Model (HMM) with Mel-Frequency Cepstral Coefficients (MFCCs) as features. As shown in Fig. 21, the interelement spacing is 5 cm and the sampling

rate is 8 KHz, the distance between microphone array and the speakers is 30 cm. Assume a room of dimensions $12\times12\times9$ m, with the microphone located at the center of the room. The SNR is from 0 to 15dB and the subtending angle is from 15 to 90 degrees. Babble noise is used as the noise source. FIG. 22 compares the performance of the original noisy signal, PDE algorithm with fixed ITD threshold, automatic ITD threshold selection algorithm, and the proposed PDE-based enhancement algorithm. The subtending angle between target source and interference signal is 15°, 45°, and 75°, and there is no reberberation. The volume gain here is set to be 0.0945. The original noisy signal is the signal received in one microphone, and PDE algorithm with fixed ITD threshold is the result of the basic PDE system, where the ITD threshold is chose to be 0.4. Automatic ITD threshold selection algorithm is organized as follows: First, two complementary masks are constructed using the binary threshold, one for the target signal, the other for interference signal. After that, the short-time power for the target and the interference is calculated. Finally, the ITD threshold is obtained by minimizing the cross-correlation of the target and interfering signals after a compressive nonlinearity, as shown below:

$$\hat{\tau}_0 = \arg\min_{\tau_0} \left| \frac{\frac{1}{N}\sum_{l=1}^{L} R_T[l \mid \tau_0) R_I[l \mid \tau_0) - \mu_{R_T}\mu_{R_I}}{\sigma_{R_T}\sigma_{R_I}} \right| \tag{54}$$

where $R_T[l \mid \tau_0)$ and $R_I[l \mid \tau_0)$ are the power of the target and the interference signals after nonlinearity, σ_{R_T} and σ_{R_I} are the standard deviations of $R_T[l \mid \tau_0)$ and $R_I[l \mid \tau_0)$, respectively, and μ_{R_T} and μ_{R_I} are the means of $R_T[l \mid \tau_0)$ and $R_I[l \mid \tau_0)$, respectively. From FIG. 18 can find that, the proposed PDE-based enhancement algorithm gets excellent performance no matter what subtending angle it is, which enhances WRR

about 50-60% at SNR 0dB and all the accuracies in different subtending angles are above 90% even if the noise is very close to the target source like 15 degrees, whereas the fixed-threshold PDE and the automatic-threshold selection algorithm degrade at low SNR. Furthermore, the automatic-threshold selection algorithm performs as well as the proposed algorithm when the subtending angle is large, like 75 degrees, but significantly degrades if the subtending angle is small and SNR is low.

The effect of reverberation presents in FIG. 23. The Room Impulse Response (RIR) software is used here to simulate reverberation effects. T60 represents the reverberation time, which is the time it takes for the reverberation level to drop by 60 dB. When the reverberation time T60 is small, the effect of reverberation is not obvious, and the performance after the proposed algorithm is almost above 85% at SNR 0dB. One thing to be noticed is that, PDE technique doesn't work if noise and speech come from the same direction, as shown in FIG. 23. It even gets worse WRR than the original signal when the reverberation time is long because of the distortion of speech signal. The performance decreases quickly when T60 is larger than 2 seconds. Even with the aid of the proposed PDE-based enhancement algorithm, WRR only increases to about 60% at SNR 0dB, and the result is worse than the original signal at high SNR because of the distortion of speech signal. FIG. 24 is recognition rate of record wave files. The recording is at an anchor chamber, and therefore the effect of reverberation can be neglected. SNR is 0dB in this case, and the noise source is babble noise. It indicates that, all WRR of original signals are low, between 10% and 30%, and after the proposed PDE-based enhancement algorithm, the performance is excellent even when SNR is low.

IV. **CONCLUSIONS**

The enhancement of speech recognition using microphone arrays is presented in this paper. The super-directive microphone arrays steering to endfire performs well when the noise source is from the rear. An equalizer is applied in the super-directive microphone arrays to prevent the distortion in speech signal. When the noise signal is close to the speech, PD is proposed to solve this problem. Using GSS to find the optimal ITD threshold differing with the included angle and the optimal volume can further improve the speech recognition. Finally, simulated and experimental results are discussed to prove effective in enhancement of speech recognition.

REFERENCES

1. Y. Gong, “Speech recognition in noisy environments: a survey”, Speech Commun. 16(1995), 261-291.
2. J. Bitzer, K. U. Simmer and K. D. Kammeyer, “Multi-microphone noise reduction techniques for hands-free speech recognition –a comparative study-,” in *Robust Methods for Speech Recognition in Adverse Conditions (ROBUST99)*, 171–174, Tampere, Finland, May 1999.
3. M. Cooke, P. Green, L. Josifovski, A. Vizinho, “Robust automatic speech recognition with missing and unreliable acoustic data,” Speech Commun. 34(2001), 267-285.
4. S. Srinivasan, N. Roman, D.L. Wang, “Binary and ratio time-frequency masks for robust speech recognition,” Speech Commun. 48(2006), 1486-1501.
5. R. M. Stern, E. Gouvea, C. Kim, K. Kumar, and H. Park, “Binaural and multiple-microphone signal processing motivated by auditory perception”, in Hands-Free Speech Communication and Microphone Arrys, pages 98–103, May. 2008.
6. R. M. Stern and C. Trahiotis, “Models of binaural interaction,” in Hearing, B. C. J. Moore, Ed. Academic Press,2002, pp. 347–386.
7. H. Park, and R. M. Stern, “Spatial separation of speech signals using amplitude estimation based on interaural comparisons of zero crossings,” Speech Communication, vol. 51, no. 1, pp. 15–25, Jan. 2009.
8. K.J. Palomaki, G.J. Brown, D.L. Wang, “A binaural processor for missing data speech recognition in the presence of noise and small-room

reverberation," ,Speech Commun. 43(2004), 361-378.

9. N. Roman, D.L. Wang, G.J. Brown, "Speech segregation based on sound localization," J. Acoust. Soc. Am. 114, 2236-2252, 2003.
10. M. Brandstein and D. Ward, *Microphone arrays* (Springer, New York, 2001).
11. S.L. Gay, J. Benesty, *Acoustic signal processing for telecommunication*, (Kluwer Academic Publishers, 2000).
12. C. Kim, K. Kumar, B. Raj, and R. M. Stern, "Signal Separation for Robust Speech Recognition Based on Phase Difference Information Obtained in the Frequency Domain," in INTERSPEECH-2009, pages 2495–2498, Sept. 2009.
13. H. Teutsch, G.W. Elko, "First- and Second-order adaptive differential microphone arrays," 2001.
14. H. Song, J. Liu, "First-Order Differential Microphone Array for Robust Speech Enhancement," Language and Image Processing, 2008.
15. P. H. Rogers, A. L. V. Buren, "New approach to a constant beamwidth transducer," J. Acoust See. Am. 64(1), July 1978.
16. W. Marshall Leach, Jr., *Introduction to electroacoustics and audio amplifier design* (Kendall/Hunt publishing company,2003).
17. J.G.Wilpon, L.R.Rabiner, C.H.Lee, E.R.Goldmn, "Automatic recognition of keyword in unconstrained speech using hidden Markov models," IEEE Trans. ASSP. Nov 1990.
18. H. Ney, "The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition," IEEE Trans. Acoustics, Speech, Signal Proc., vol.32, no2, pp.263-271, April 1984.
19. Chin-Hui Lee, Frank K. Soong and Kuldip K. Paliwal. "Automatic Speech and Speaker Recognition," Kluwer Academic Publishers. 1995.
20. numerical recipes in C: the art of scientific computing, 2nd Edition, 1993.

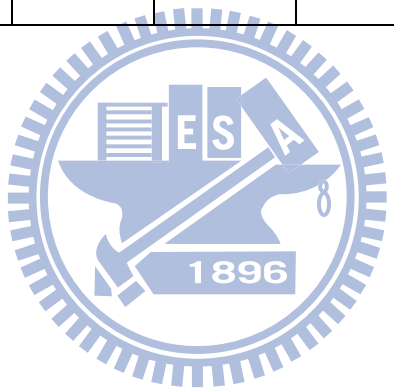
21. J. Bergqvist and F. Rudolf, “A silicon condenser microphone using bond and etch-back technology,” Sensors and Actuators A, 45, 115-124 (1994).
22. C. Kim, R.M. Stern, K. Eom, J. Lee, “Automatic selection of thresholds for signal separation algorithm based on interaural delay,” 2010.

TABLE I Table of first-order differential designs.

Microphone type	**DI (dB)**	**FBR (dB)**	**3dB Beamwidh**	**Nulls (degrees)**
Dipole	4.77	0.00	90.00°	90.00
Cardioid	4.77	8.45	131.06°	180.00
Hypercardioid	6.02	8.45	104.90°	109.47
Supercardioid	5.72	11.44	114.90°	125.26

TABLE II Comparing the effective beamwidth corresponding to the optimal ITD threshold and the subtending angle.

Average τ	0.9909	0.9597	0.9025	0.7055	0.4676	0.2653
Corresponding effective beamwidth	58.72	55.9	51.1	37.5	23.8	13.2
The subtending angle	90	75	60	45	30	15



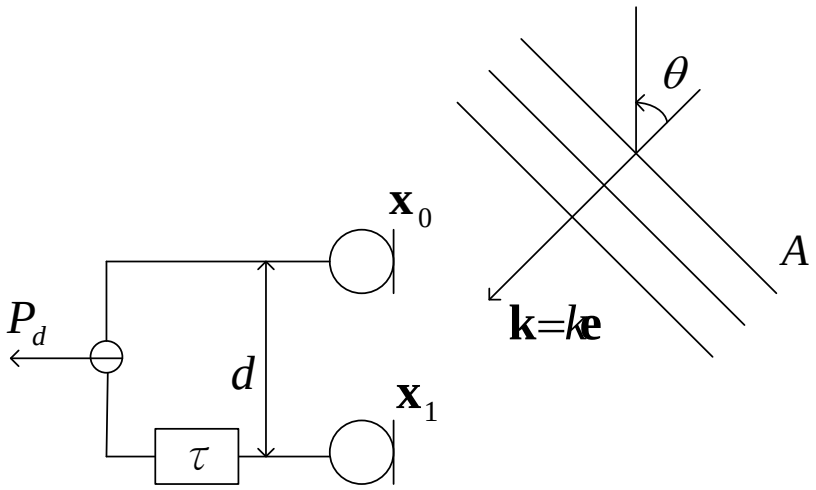


FIG. 1 Diagram of first-order microphone composed of two microphones.

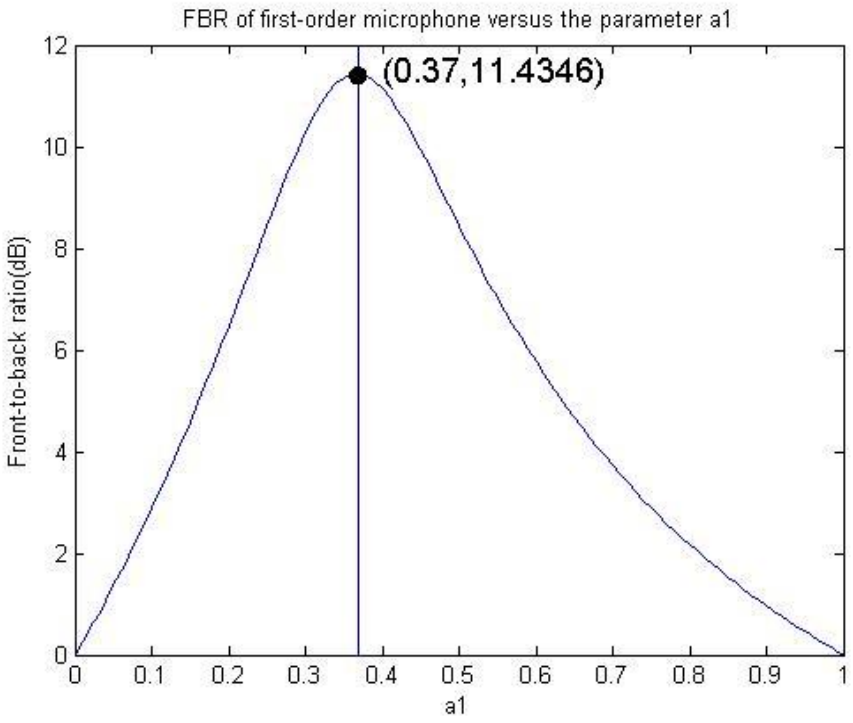


FIG. 2 Front-to-back ratio of first-order microphone versus the first-order differential parameter $\alpha_{1..}$

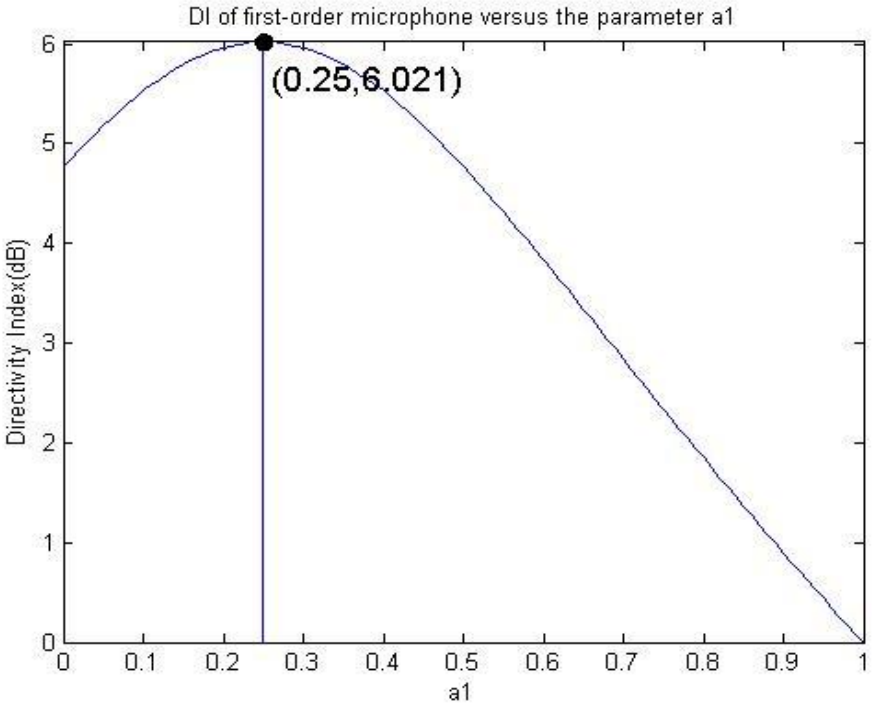


FIG. 3 Directivity index of first-order microphone versus the first-order differential parameter α_1.

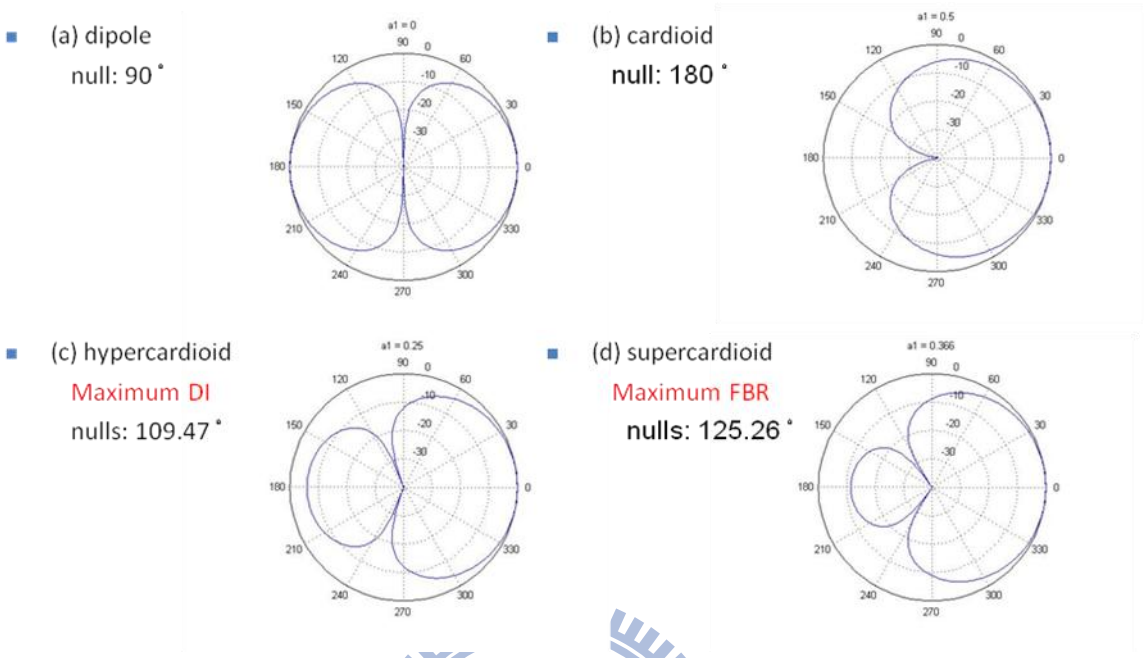


FIG. 4 Various first-order directional responses (a) dipole, (b) cardioids, (c) hypercardioid, (d) supercardioid.

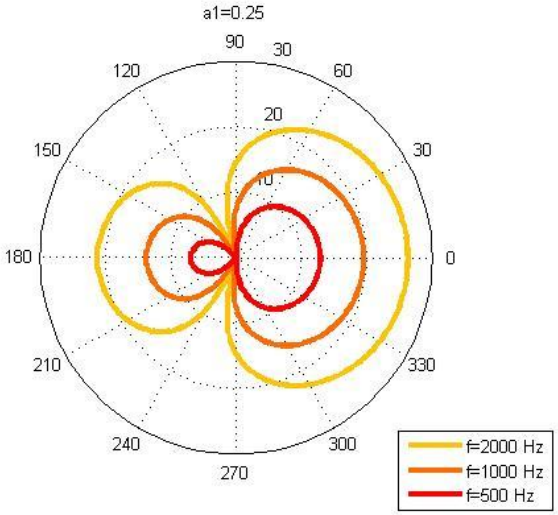


FIG. 5(a) α_1=0.25

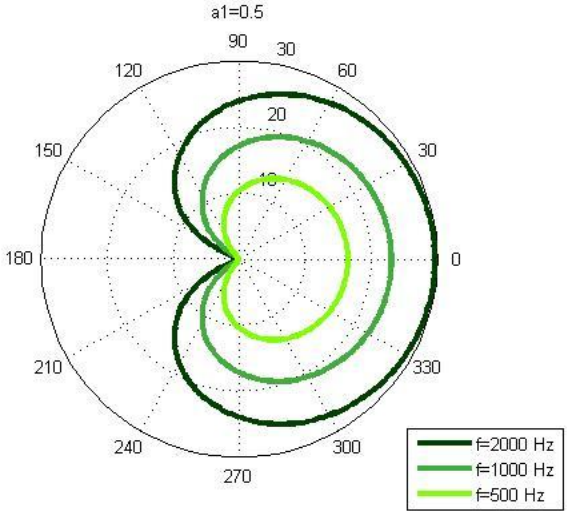


FIG. 5(b) α_1=0.5

FIG. 5 The directivity pattern of 1st order DMAs.

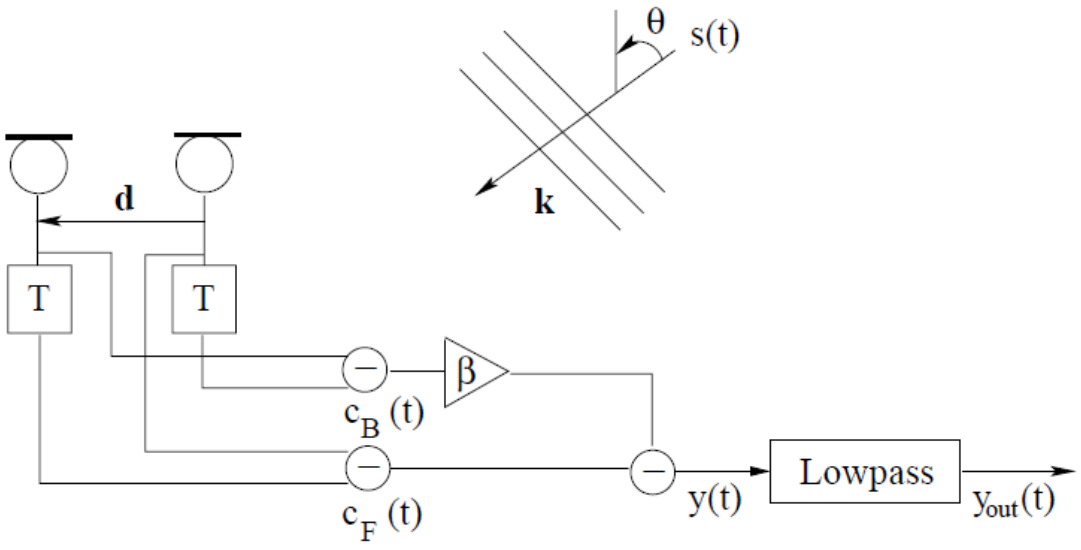


FIG. 6 The block diagram of First-order ADMA

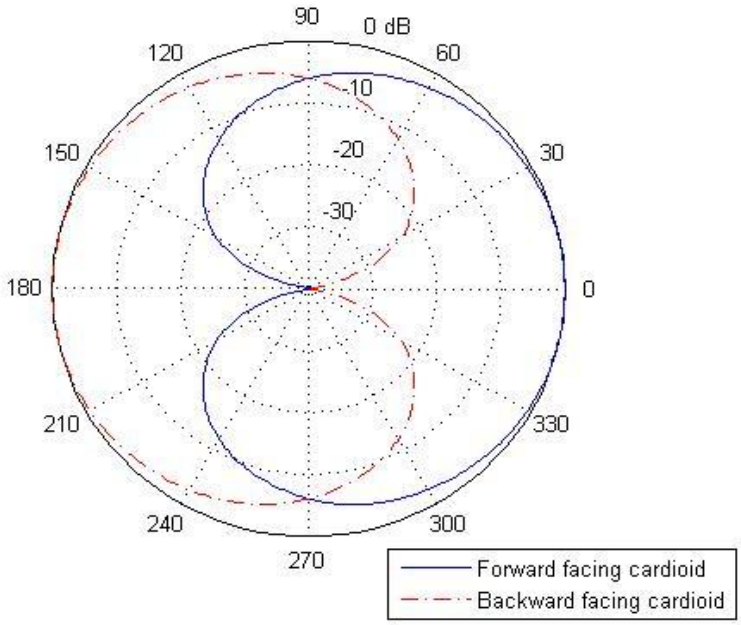


FIG. 7 Directivity pattern of the first-order back-to-back cardioids system.

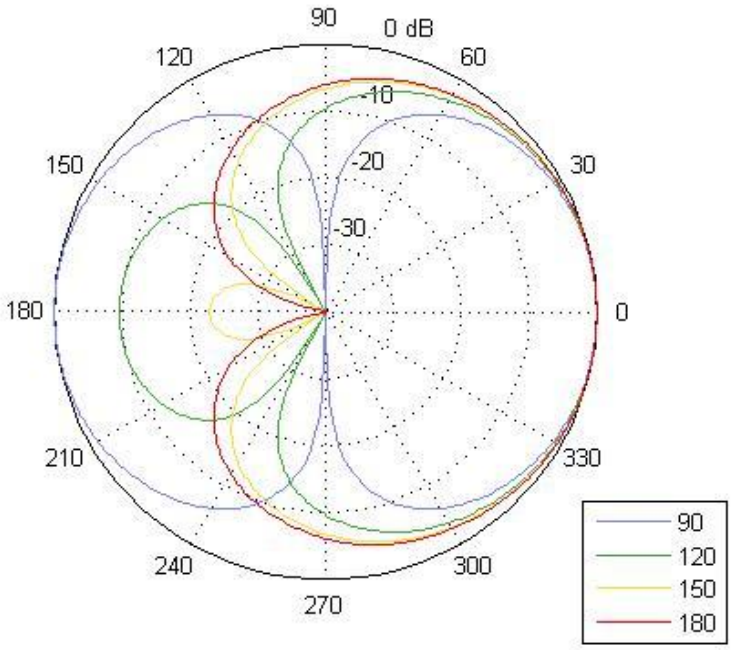


FIG. 8 Various directivity patterns for a first-order ADMA

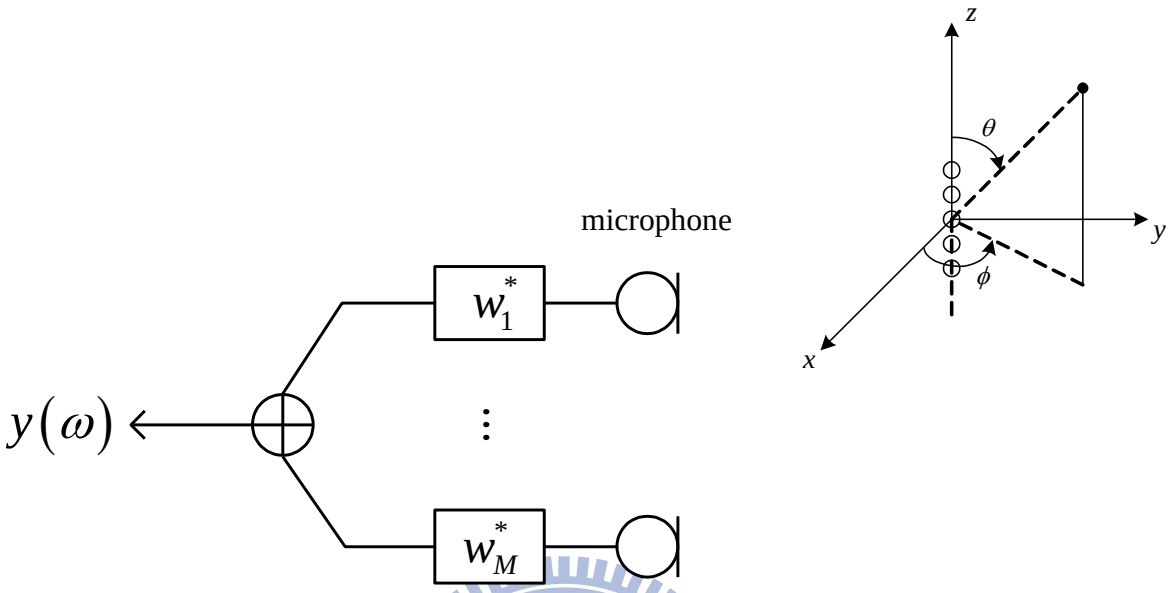


FIG. 9 The model of the optimal beamformer, which is a filter and sum system.

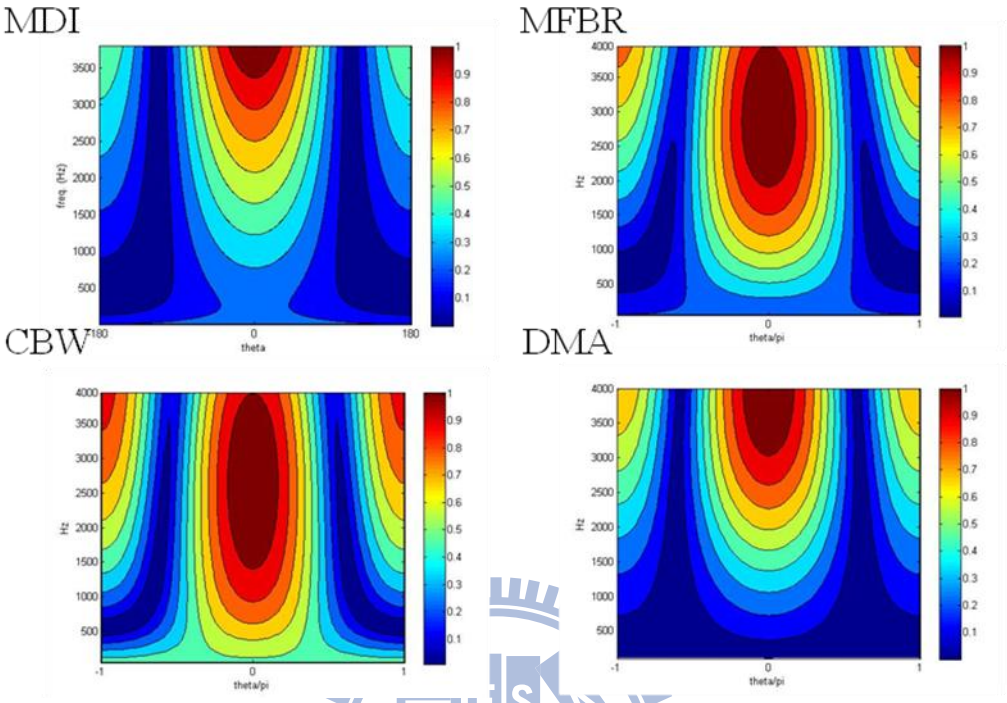


FIG. 10 The contour plot of super-directive microphone arrays, the four plots represent maximum for DI, maximum for FBR, maximum for constant beamwidth, and 1st DMA respectively.

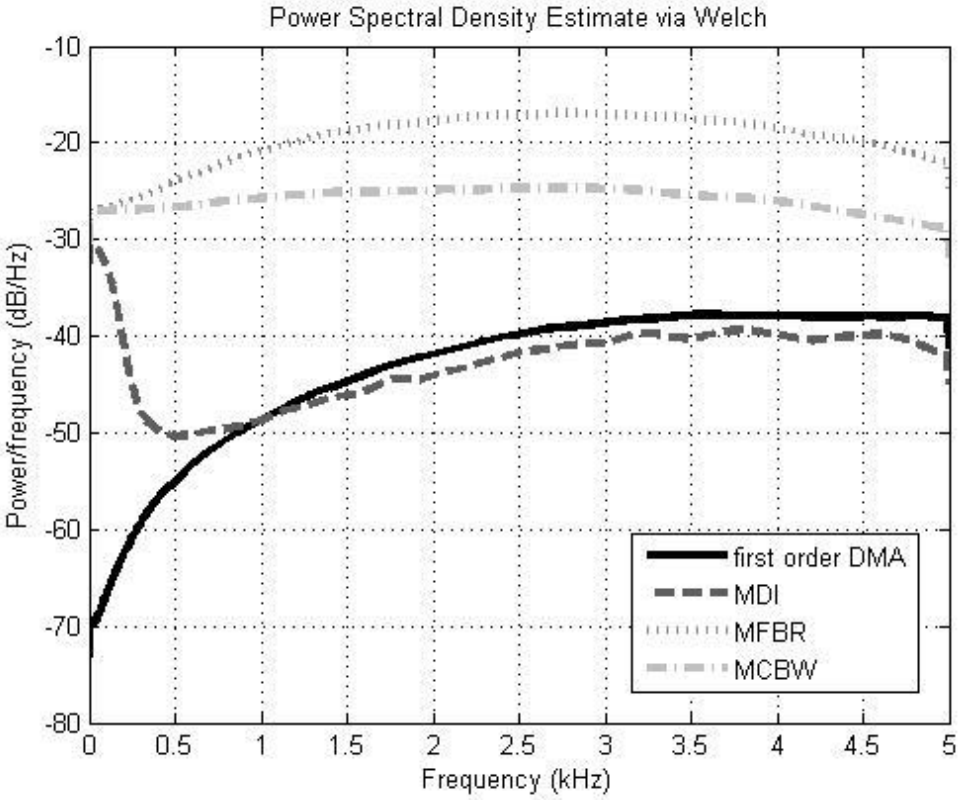
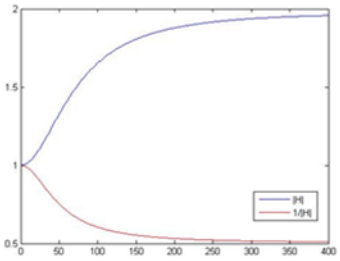
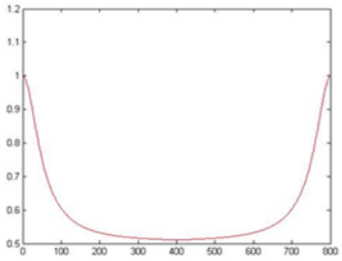


FIG. 11 The power spectral density of super-directive microphone arrays.

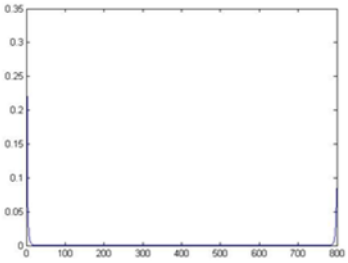
❖ 1. 1/|H|



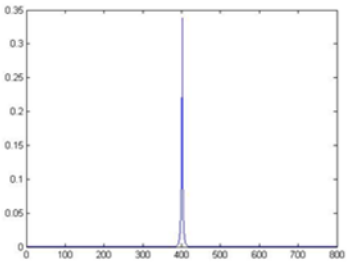
2. 1/|H| + symmetric 1/|H|



❖ 3. ifft



4. shift



❖ 5. convolution

FIG. 12 The processing of applying an equalizer in super-directive microphone arrays.

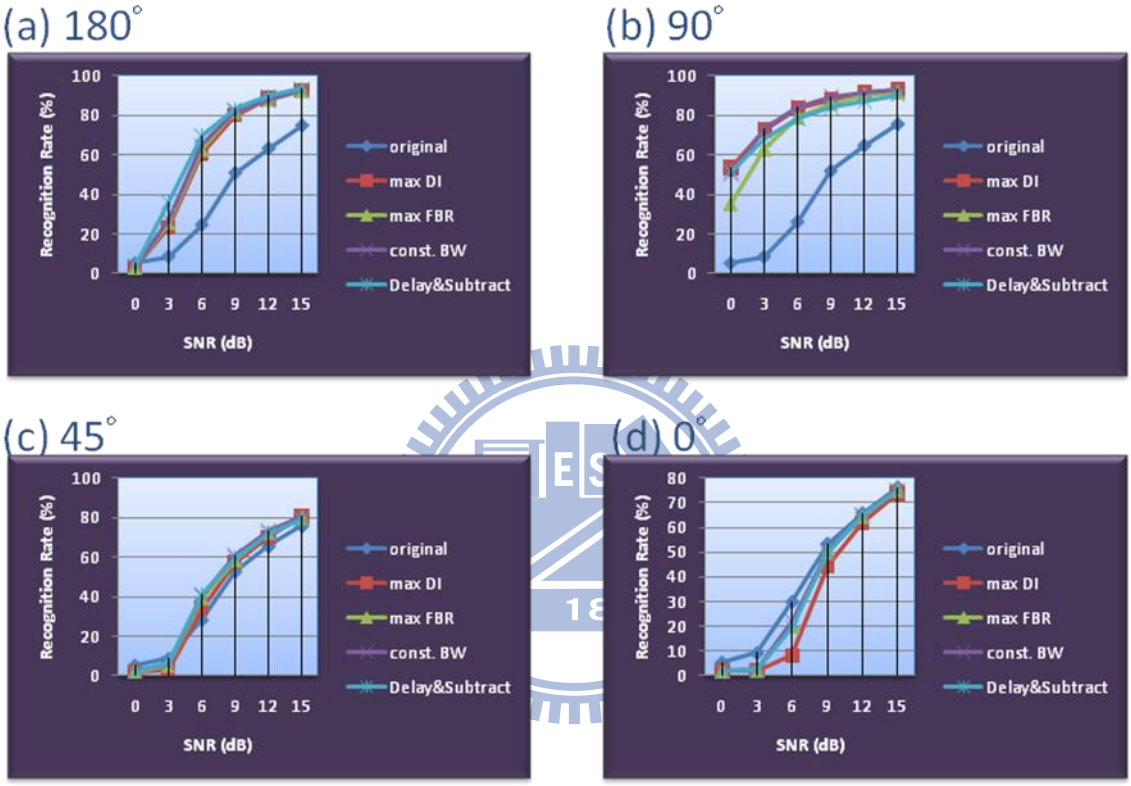


FIG. 13 The recognition rate (%) of the noisy speech (white noise) using different algorithms, where the noise signal is located at (a) 180 degrees, (b) 90 degrees, (c) 45 degrees, (d) 0 degree.

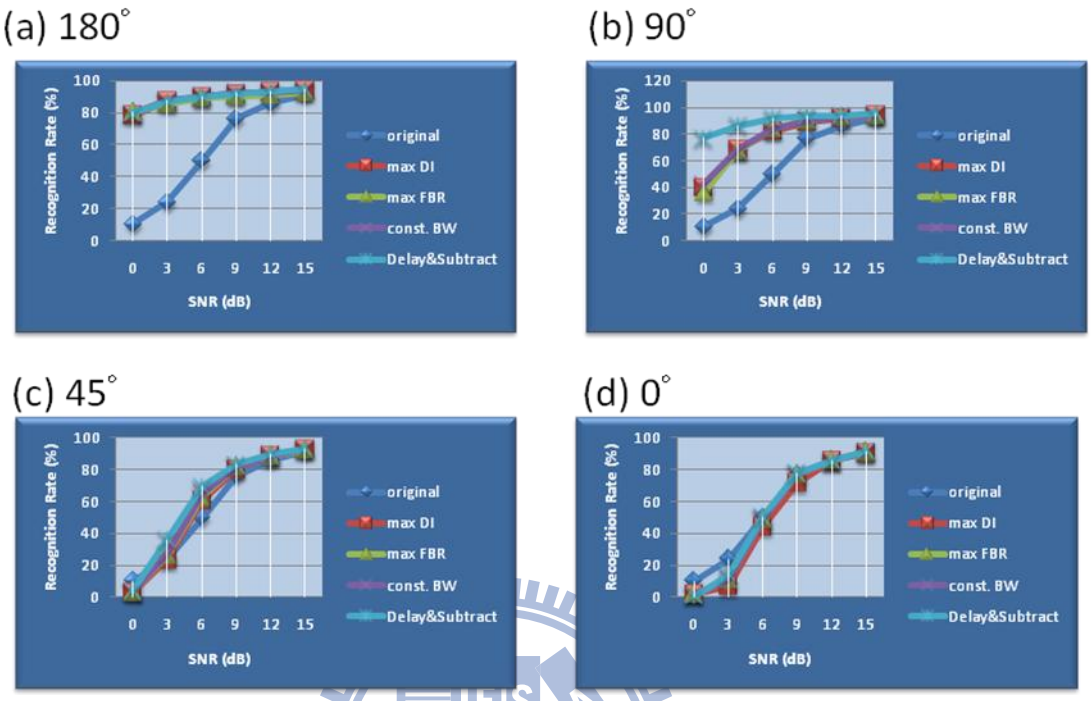


FIG. 14 The recognition rate (%) of the noisy speech (car noise) using different algorithms, where the noise signal is located at (a) 180 degrees, (b) 90 degrees, (c) 45 degrees, (d) 0 degree.

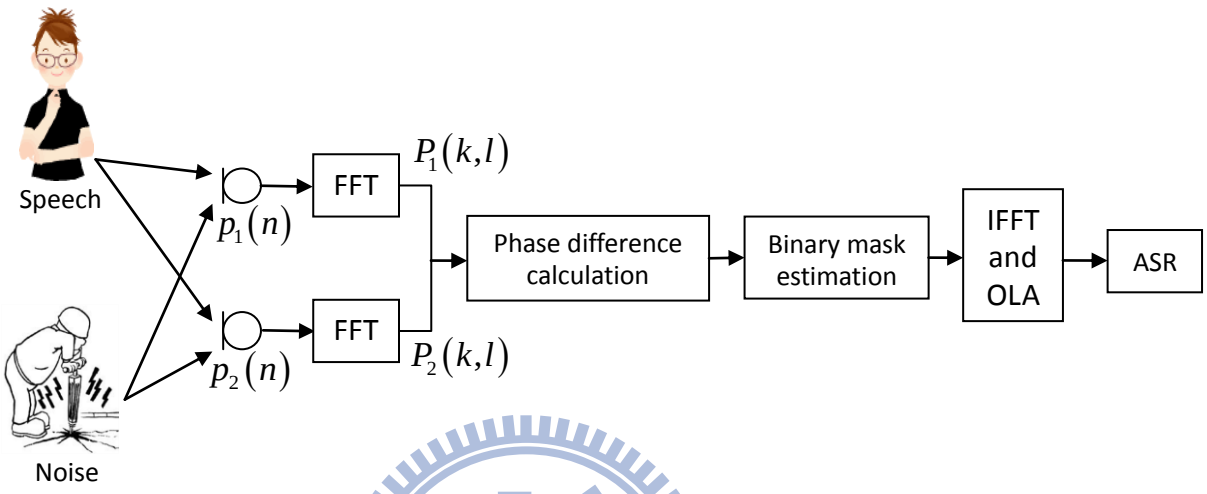


FIG. 15 The block diagram of phase-difference estimation.

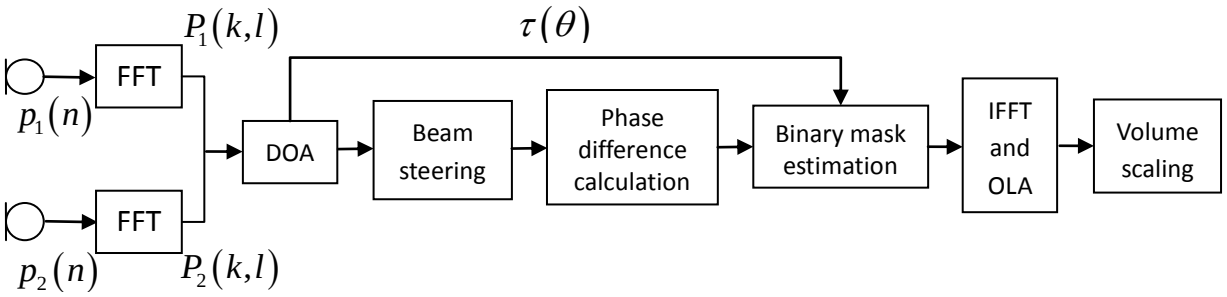
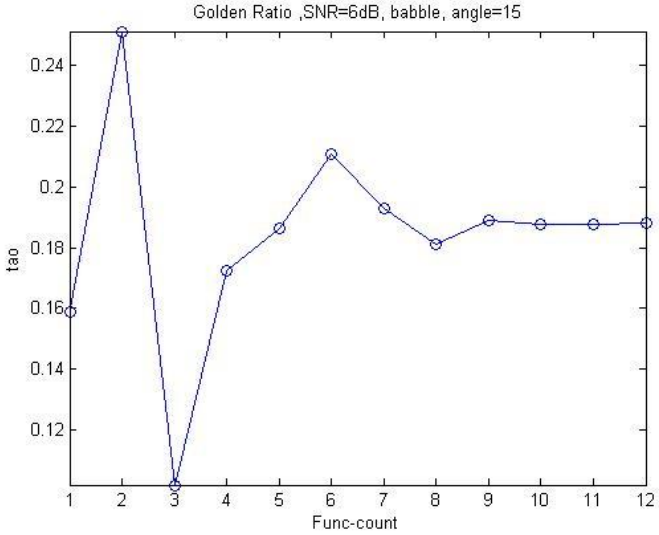
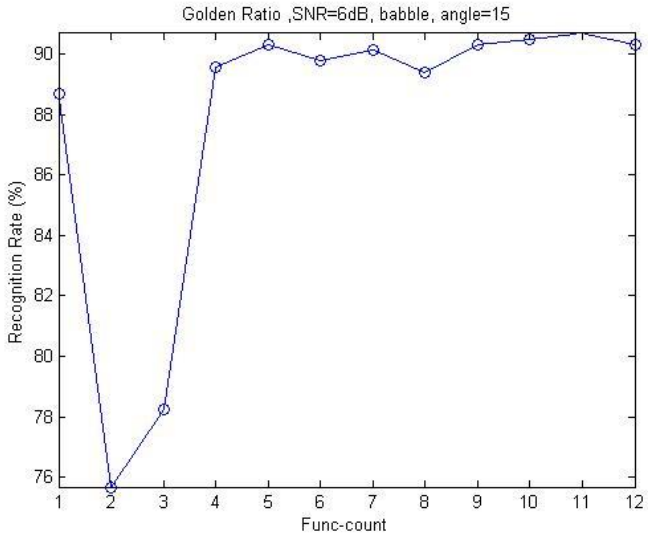


FIG. 16 The block diagram of the proposed PDE-based enhancement algorithm, where θ is the subtending angle estimated by DOA.



(a) The searching process of τ



(b) Relative recognition rate

FIG. 17 The searching process of the ITD threshold by GSS.

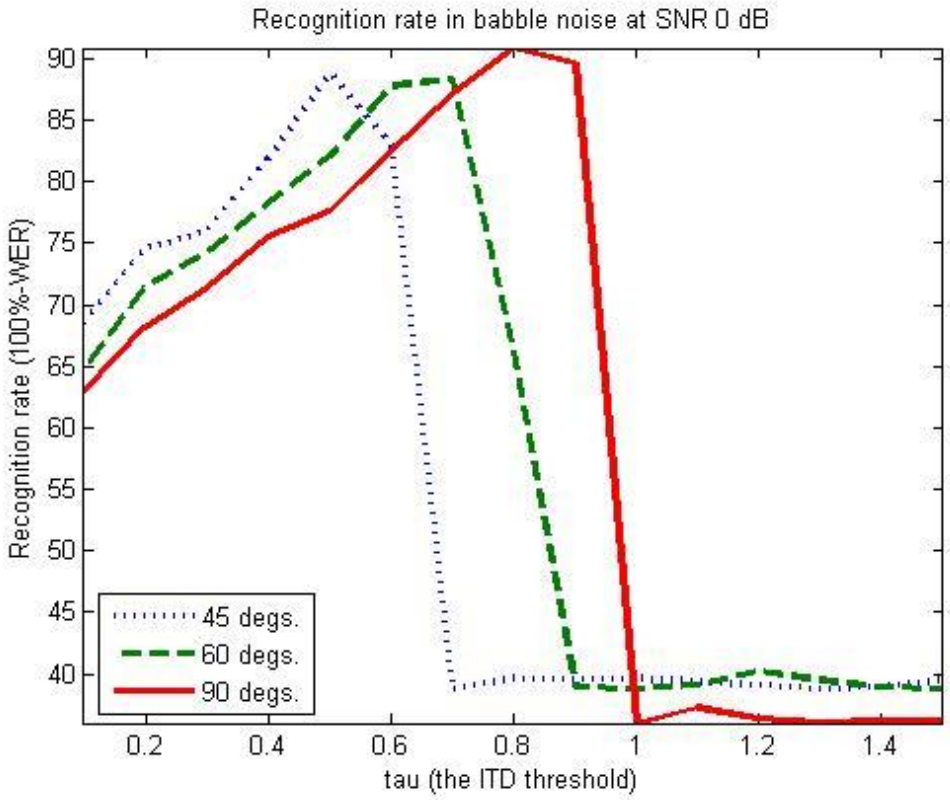


FIG.18 (a)

FIG. 18 (a) Recognition rate in babble noise at SNR 0dB. (b) The optimal ITD threshold tau and the polynomial fitting.

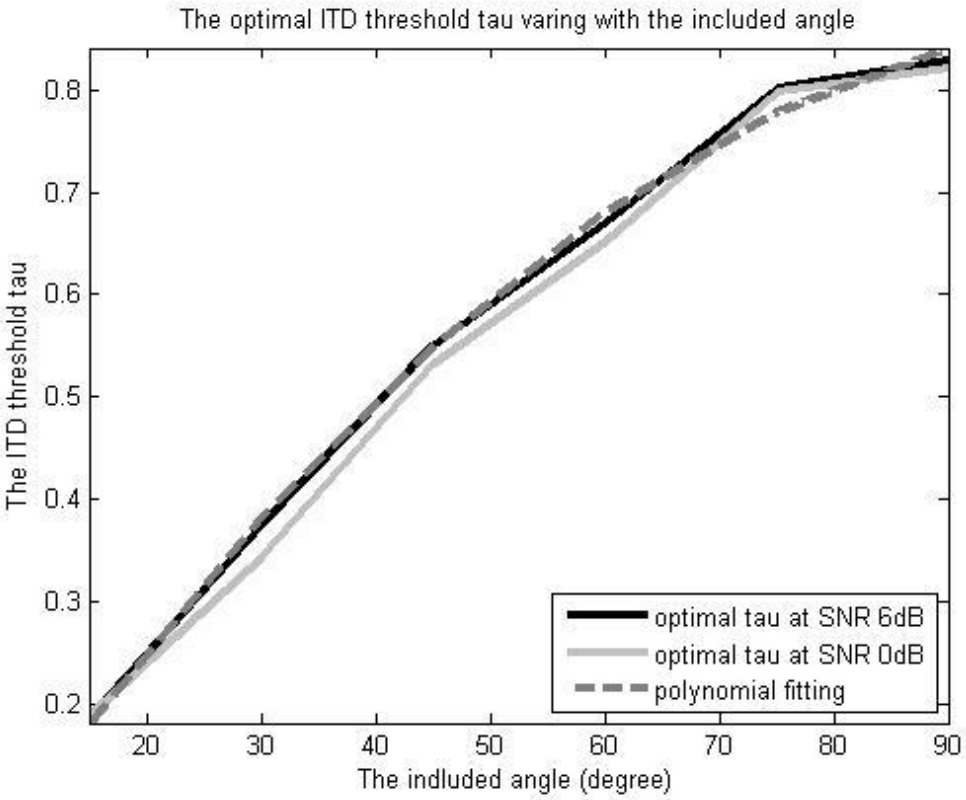

The optimal ITD threshold tau varing with the included angle
The ITD threshold tau
The indluded angle (degree)
0.2
0.3
0.4
0.5
0.6
0.7
0.8
20
30
40
50
60
70
80
90
optimal tau at SNR 6dB
optimal tau at SNR 0dB
polynomial fitting

FIG.18 (b)

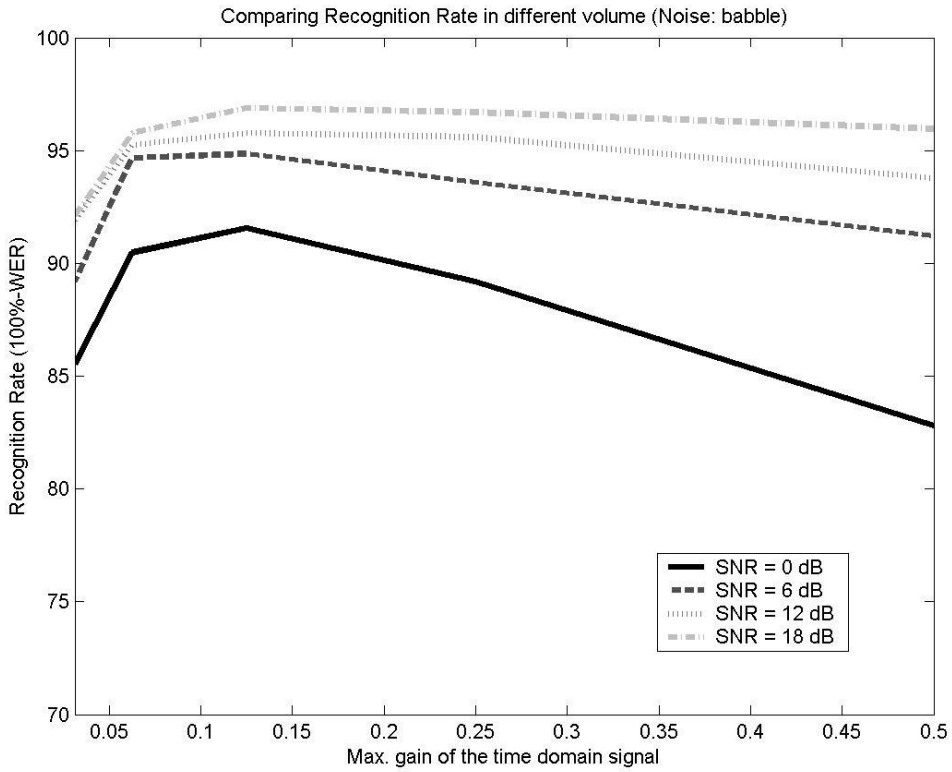


FIG. 19 Comparing recognition rate in different volume.

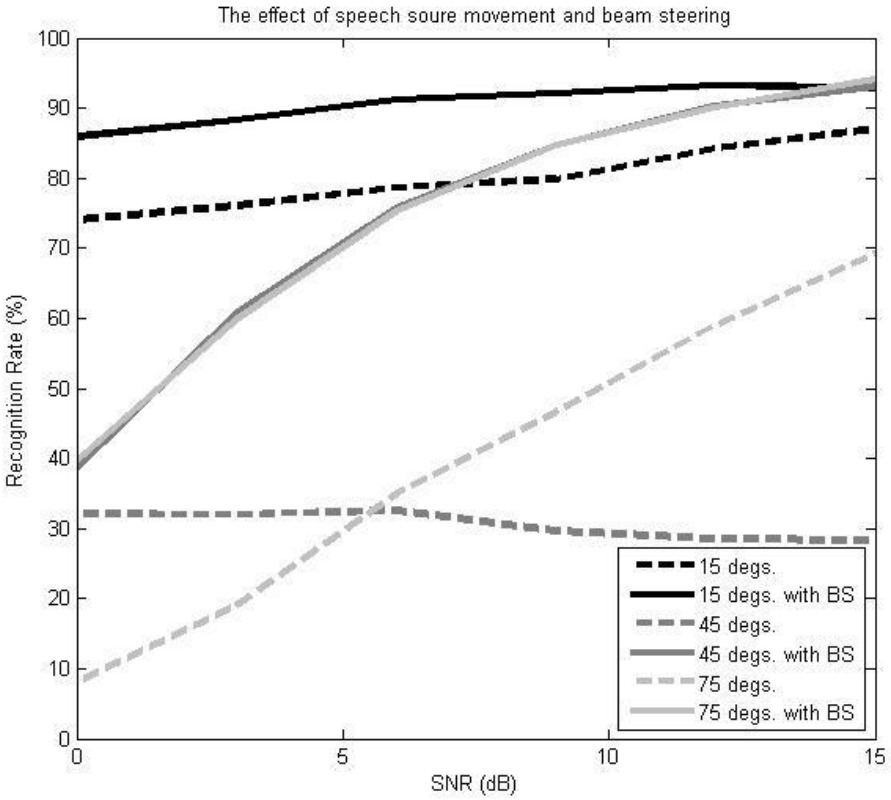


FIG. 20 Comparing the recognition rate when the source is not at the direction of the designed mainlobe and the effect of beam steering, where "15degs." means the source is aside the desired main axis 15 degrees.

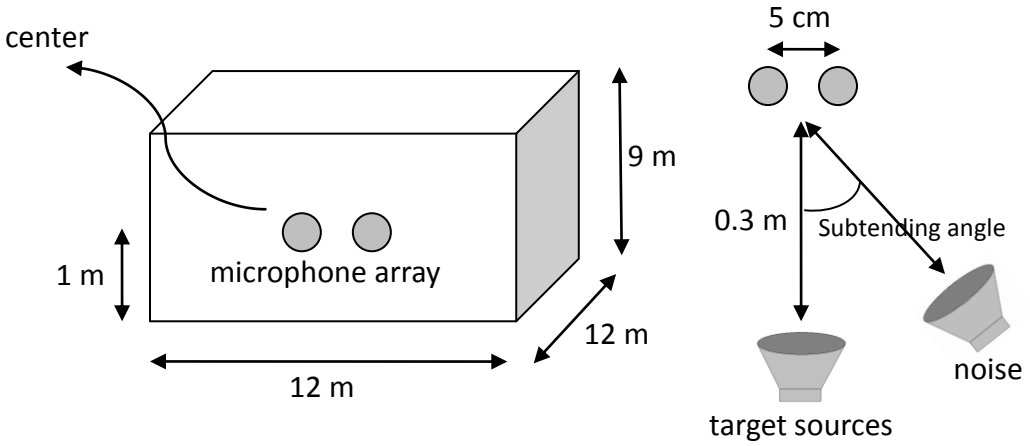


FIG. 21 The simulated and experimental environments.

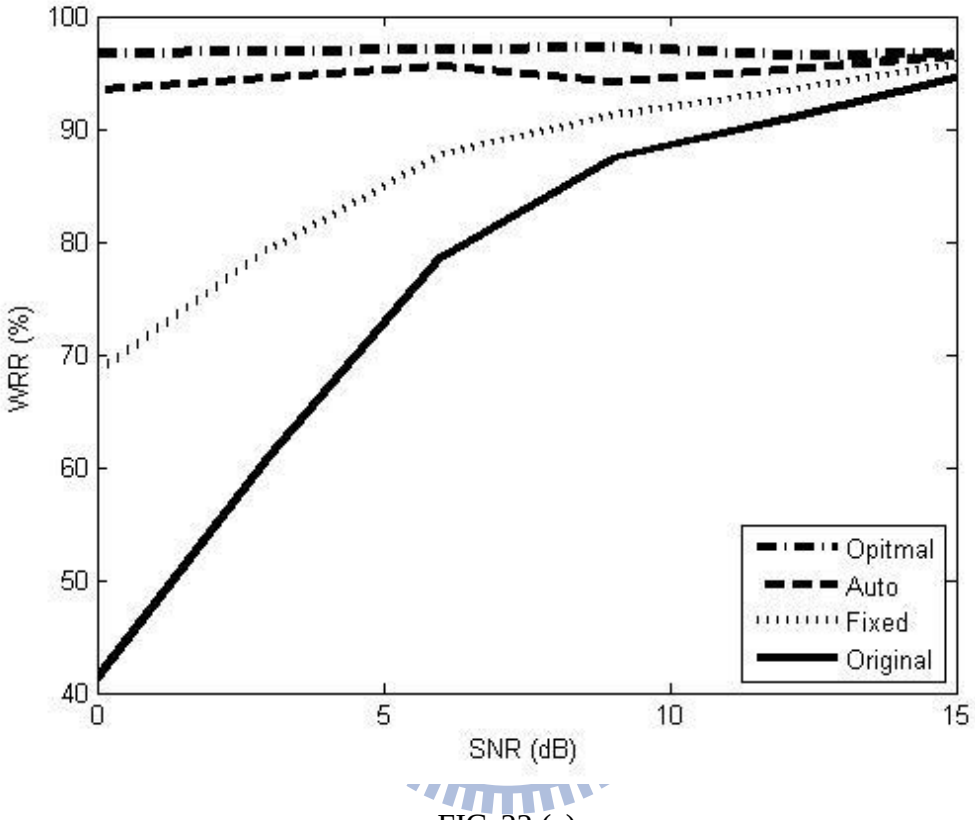


FIG. 22 (a)

FIG. 22 Comparing the performance of the original noisy signal, PDE algorithm with fixed ITD threshold, automatic ITD threshold selection algorithm, and the proposed PDE-based enhancement algorithm (a) Subtending angle = 75°. (b) Subtending angle = 45°. (c) Subtending angle = 15°.

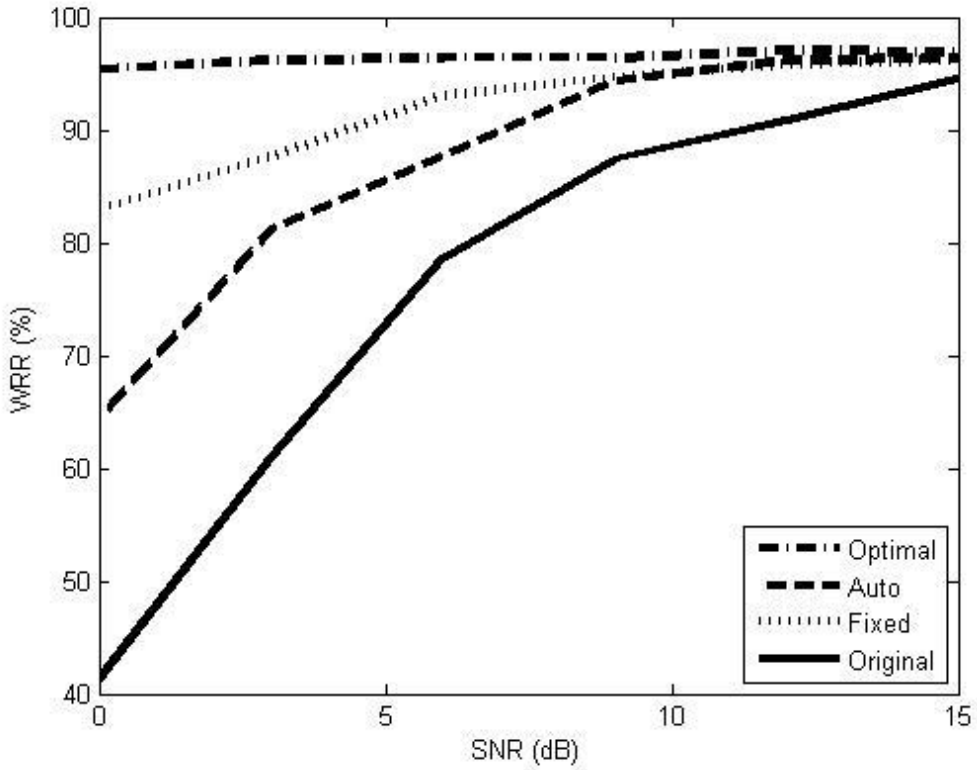

100
90
80
70
60
50
40
WRR (%)
0
5
10
15
SNR (dB)
Optimal
Auto
Fixed
Original

FIG. 22 (b)

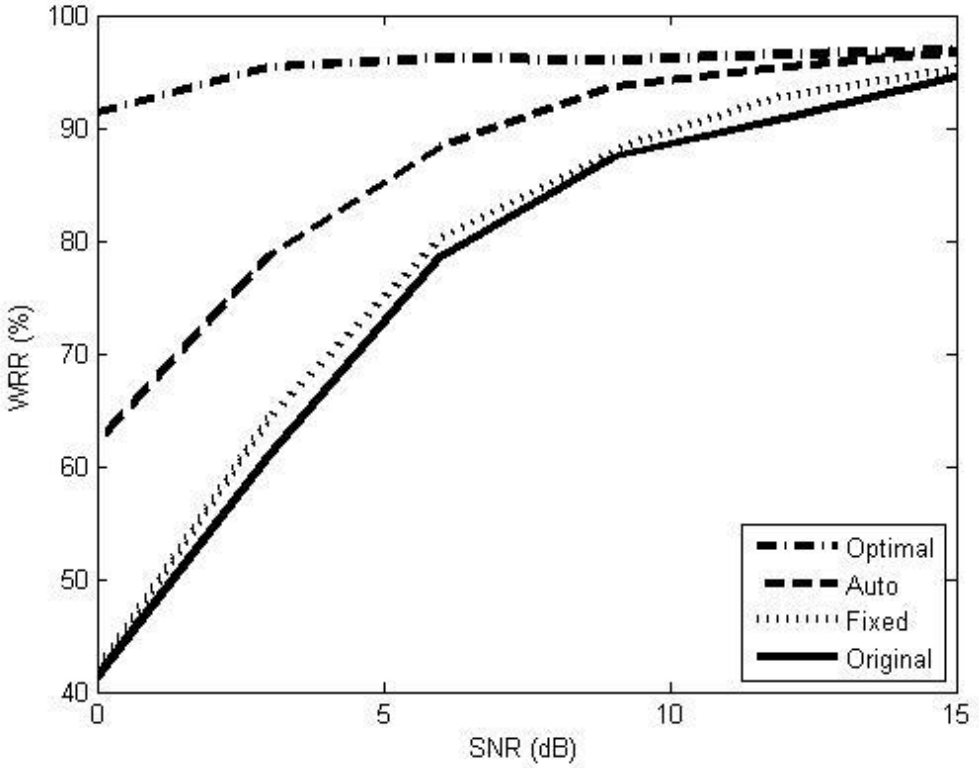

100
90
80
70
60
50
40
WRR (%)
0
5
10
15
SNR (dB)
Optimal
Auto
Fixed
Original

FIG. 22 (c)

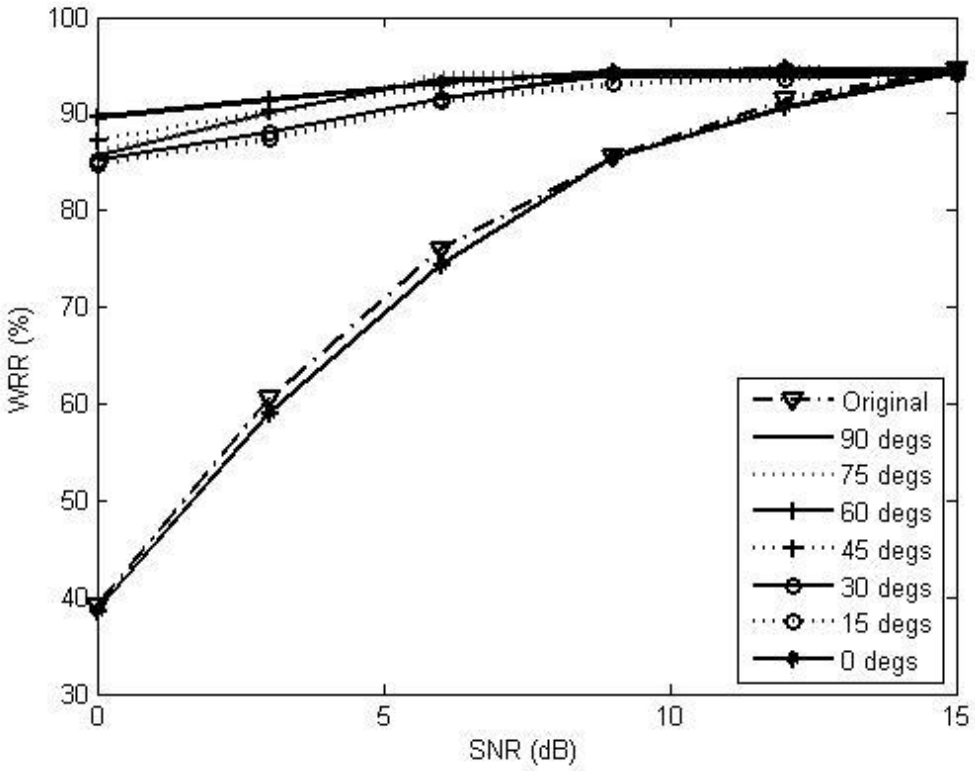


FIG. 23 (a)

FIG. 23 The effect of reverberation, where the subtending angle is from 0 to 90 degrees. (a)T_{60}=0.138 secs. (b)T_{60}=0.966 secs. (c)T_{60}=2.898 secs.

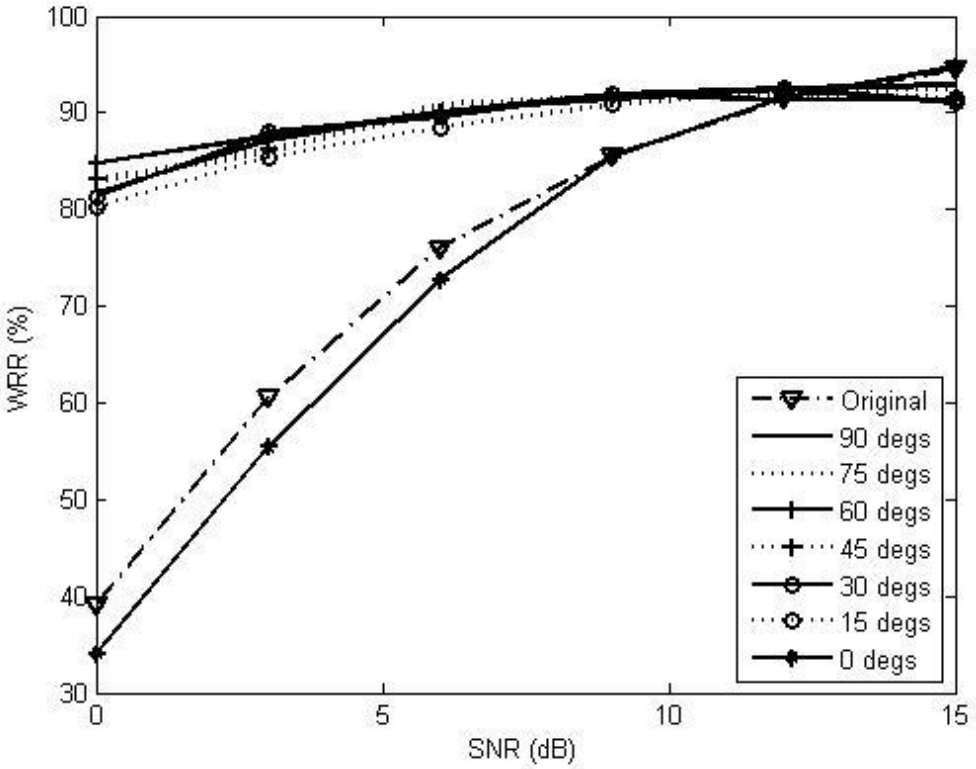

100
90
80
70
60
50
40
30
WRR (%)
0
5
10
15
SNR (dB)
Original
90 degs
75 degs
60 degs
45 degs
30 degs
15 degs
0 degs

FIG. 23 (b)

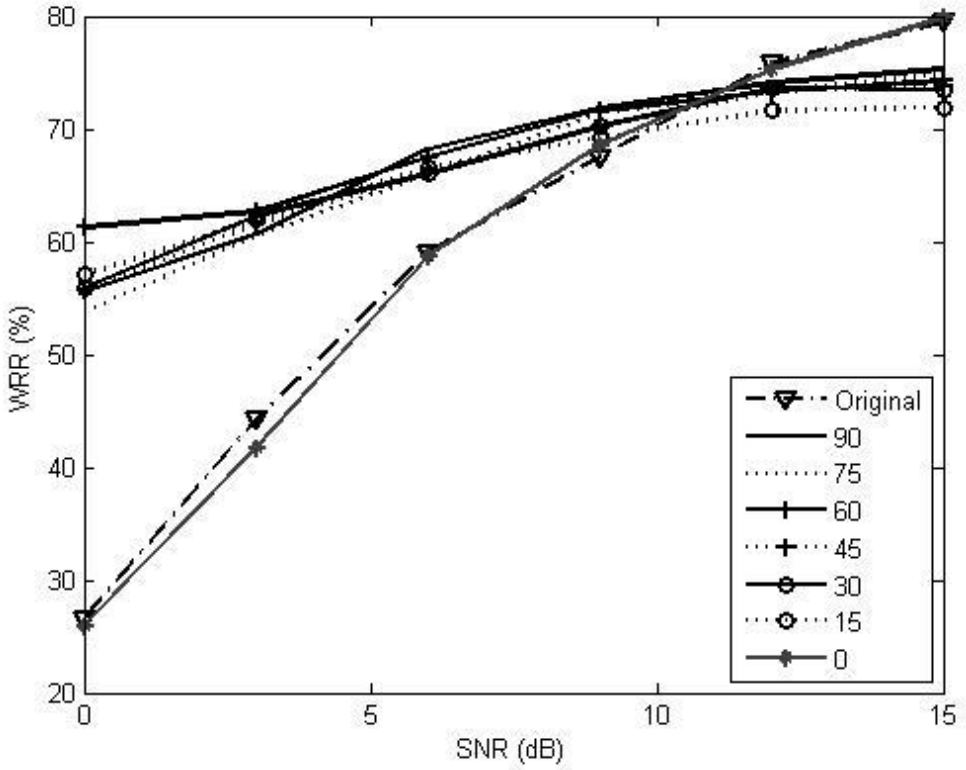


FIG. 23 (c)

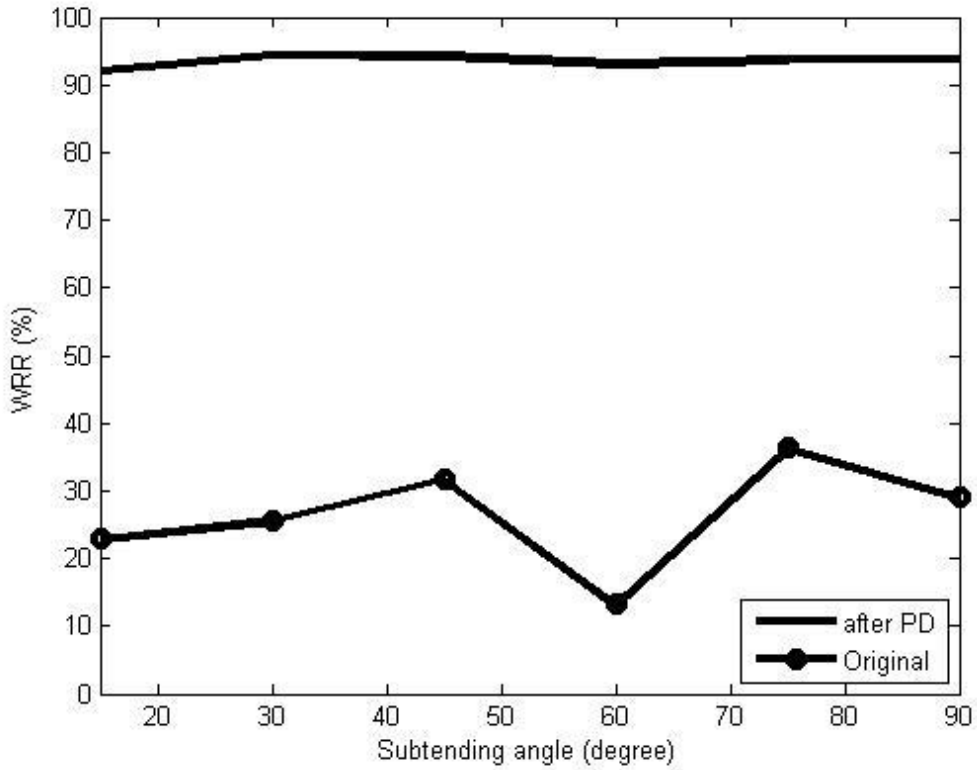


FIG. 24 The recognition rate with the optimal threshold of record wave file.