國 立 交 通 大 學

電控工程研究所

碩 士 論 文

多通道語音強化
使用相對轉移函數建構之零波束形成

Multichannel Speech Enhancement
Using Relative Transfer Function Based Nullforming

研 究 生： 蔡 沛 錡

指導教授： 胡 竹 生 博士

中 華 民 國 九 十 九 年 七 月

多通道語音強化

使用相對轉移函數建構之零波束形成

研究生：蔡　沛　錡　　　　　　　指導教授：胡　竹　生　博士

國立交通大學

電控工程研究所碩士班

摘要

本論文提出一個針對穩態或是非穩態干擾聲源消除的語音強化方法。消除非穩態雜訊是目前語音純化研究中相當重要的問題，本論文提出一個以適應性濾波器為基礎並結合零波束形成演算法的空間濾波器。零波束形成演算法是利用奇異值分解法找出干擾聲源的零空間當作零波束形成器。論文中零波束形成器分為固定式和可變式。可變式零波束形成器以階數回歸最小平方誤差估計的方法找出當前麥克風收到訊號的子空間，並使用子空間相似度的演算法，剔除目標聲源子空間並利用正交上三角分解產生一組獨立基底。這些基底組成了干擾聲源的子空間。零波束形成演算法可應用在不同適應性濾波器上，本論文將固定式零波束形成器應用在廣義旁辦對消器和參考訊號架構為基礎之濾波器；可變式零波束形成器則應用在廣義旁辦對消器。所提出的零波束形成演算法同時可對目標聲源做語音活動偵測以加強適應性濾波器的效能。本論文最後以線型麥克風陣列在實際環境下的實驗結果說明本演算法的效能。

Multichannel Speech Enhancement Using Relative Transfer Function Based Nullforming

Student： Pei-Chi Tsai Advisor： Prof. Jwu-Sheng Hu

Institute of Electrical and Control Engineering
National Chiao-Tung University

ABSTRACT

The thesis proposed a speech enhancement method for stationary and nonstationary interfering sources. To effectively eliminate nonstationary intereferences is an important research topic for speech enhancement. This thesis proposed an adaptive nullforming spatial filter. The nullforming algorithm uses singular value decomposition (SVD) to find the null space of interfering sources. Both fixed and adaptive nullforming algorithms are studied. The adaptive nullforming uses order recursive least square estimation (ORLS) to find the subspace of presently received signal. The algorithm assumes that the relative transfer functions (RTFs) of sources from different direction can be obtained. The estimated subspaces from these RTF's contain the subspace of the desired signal. They are sorted according to the distance to the subspace of source from every direction. Then the bases of desired signal subspace from estimated subspace could be removed and a set of independent basis are derived using the orthogonal triangular decomposition (QRD). The basis then comprises of the subspaces of the interfering sources. The fixed nullforming algorithm could be appiled to generalized sidelobe canceler (GSC) and reference signal based adaptive beamformer (RSAB) while the adaptive one can be applied to GSC. Further, it can also be used as directional voice activity detection (VAD) to enhance the performance. Finally, experiments using a linear microphone array under real environment are conducted to demonstrate the performance of proposed algorithm.

Contents

誌謝

記得兩年前的一個下午，當老師決定收我為 XLAB 的一份子時，便開啟了我多采多姿的研究生活。兩年的研究生涯裡，首先感謝胡老師辛勤的指導。在論文方向不定時，老師總能指引一條正確的路。感謝我的良師益友同時也是我的最佳球友-熱心的明唐學長，總是不吝於解答我研究上的問題，幫助我了解如何擁有正確的研究態度。我真的很高興能和你一起研究、一起打球。興哥總是能用最簡單的方式讓我了解艱深的理論。這篇論文的完成非常感謝兩位學長的幫助。感謝建安學弟，我的電腦在快要口試之前意外地掛了，感謝你在這個時候即時伸出援手。感謝阿法學長，常常在研究煩悶的時候能有人一起吃飯聊天。還有常常一起打球的肉鬆。感謝 XLAB 的同梯們-常常被我串門子的阿 him、讓我英文能力增進不少的的 Rodolfo、衝浪男孩 Simon、看起來有點兇但是人很好的小蔡；還有 XLAB 的學長學姐們學弟妹-在實驗室做好吃點心的鏗元學姐、常常提供我美食資訊的阿吉、修課時幫助我許多的助教大師兄、常常提醒我要認真的永融、辦出遊一流的學妹、為實驗室帶來歡樂的 Macaca、樂理超強的昀軒、常常一起打球的新文、深藏不露的學文和喜歡桌遊的育成。和你們一起工作讓我學習了許多。

特別感謝我從小到大的好朋友孟璋，我們一起進大學、一起進研究所，終於我們要一起畢業了。雖然我們從國小之後就不在同校了，但在漫長的求學過程中我們常常互相勉勵對方，期許彼此能更進一步。希望未來我們都能完成自己的夢想。還有我的大學朋友柏志，先恭喜你成為了研究新鮮人。在工作之外能一起吃個飯逛個街聊一聊彼此的夢想真的很高興。這樣能讓我暫時忘記工作的壓力。還有智哲，也恭喜你輔成為研究生。希望你們接下來的研究生涯一切順利。還有我的高中朋友碩詣，快要考研究所了，希望你能多加油。期許你能考上自己理想的科系。另外要感謝我的神師老弘，一次又一次的談話總讓我有許多的啟發，期許自己成為一個懂得生活的人。

最後我要感謝生命中最重要的人-我的父母和我的妹妹。不管我做什麼決定，他們常常在背後默默的支持我。爸媽總是為了我的教育而辛苦的賺錢，還不時要處理我的意外狀況。記得大學時曾經一度想放棄學業，是爸媽及時從高雄趕來新竹勸我不要放棄繼續撐下去。幸好當初有他們的堅持，現在我才有機會在這邊寫下這篇誌謝。

漫長的求學生涯在此告一個段落，這段旅程的結束揭示了另一段旅程的開始。前方的道路上還有許多新奇的事等著，期許自己未來能好好利用自己在實驗室所學到的技能貢獻社會。

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Objective

Speech enhancement in a noisy environment is an important research issue for speech signal processing. There are various kinds of interferences in the environment and they are usually classified into stationary noises and nonstationary ones. One of the approaches to solve this problem is to use microphone array where the spatial characteristics of sound waves are exploited. For stationary noises, the multichannel adaptive Wiener filter and its variations [1-4] were proposed and proved to be quite effective. However, they do not perform well in real practice when nonstationary noises such as competing speech are present.

In spatial signal processing of a microphone array, blocking one of the sound sources is equivalent to finding the corresponding null space within the multi-dimensional signal space formed by the microphone measurements. To effectively obtain the subspaces and process their signals accordingly for interference reduction are two major focuses of the research in recent years. The difficulty is the subspaces are usually unknown in advance and become time-varying when environment changes. This provides the motivation of this thesis to study and propose innovative methods to compute the subspaces for nonstationary interference reduction. The primary target of interference considered in this thesis is competing speech. It is a common issue for speech communication as well as recognition under multi-person scenarios.

1.2 Literature Review

Speech enhancement using microphone array has been widely used in noisy environment. Generally speaking, microphone array uses the spatial response of the signals received by different microphones to separate the signal from different directions. These kinds of signal enhancement methods are generally called beamforming. Beamforming technique has been studied for many years. In sonar system [5], beamforming has been used since 1960s. The earliest beamforming is delay-and-sum (DS) beamforming, which is also called conventional beamforming. The DS beamforming adds the signals with delay compensation but it is not effective under reverberant environment and requires a large amount of array elements for higher performance.

The adaptive beamforming was originally proposed by Griffiths [6]. This beamforming algorithm is an unconstrained minimum mean square error (MMSE) method. After that, the concept of constrained beamforming was proposed in several research works. The most famous one is the constrained least mean square (LMS) algorithm derived by Frost [7]. The performance of speech enhancement is greatly influenced by the mismatch of microphones. Cox, H et al. [1] proposed a robust adaptive filter to avoid the problem of mismatch. Griffiths and Jim reconsidered Frost's algorithm and proposed the generalized sidelobe canceler (GSC) [8]. GSC comprises of three parts. The first part is a fixed beamformer, the second one is a blocking matrix and the third one is an adaptive noise canceller. The architecture of GSC satisfies the criterion of LCMV. To cope with wide-band signals, Nordholm *et al.* [9] proposed the wide-band Wiener solutions under the Griffiths-Jim beamformer architecture. Speech enhancement methods in a reverberant room using GSC are suggested by some authors. Hoshuyama *et al.* [3] proposed an adaptive beamformer similar to the architecture of GSC with a modified blocking matrix to work adaptively.

To deal with the nonstationary signal, Gannot *et al*. proposed generalized sidelobe canceler (GSC) with nonstationary desired source using relative transfer function (RTF) [11]. The RTF could be used to describe the relative transfer function of room impulse response (RIR) between microphones. The purpose of using RTF on GSC is to let the blocking matrix blocks the nonstationary signal. Reuven *et al*. [10] proposed dual source transfer function GSC (DTF-GSC), which would eliminate a single nonstationary interfering source. In [10], the fixed beamformer (FBF) and blocking matrix (BM) are modified to block the nonstationary signal. Therefore, the GSC would eliminate the residual stationary noise only. For the case with two or more interfering sources, the method cannot effectively eliminate all the interfering signals.

The RTF based BM can be used to eliminate the nonstationary signal, that is, the BM is a nullformer of nonstationary sources. To enhance the desired signal, applying the nullformer to adaptive filter seems to be a feasible method. However, in practical environment, it's difficult to know the number of emitting sources. The method to estimate BM by [10] for dual sources is inflexible. Therefore, it is necessary to generate nullformer on-line in order to eliminate the unknown number of interfering sources.

Dahl *et al.* [2] proposed an adaptive filter using normalized least mean square (NLMS) criterion to perform indirect microphone calibration and minimize the speech distortion due to the channel effect (using pre-recorded speech signals). Chen *et al.* [11] proposed reference signal based frequency domain adaptive beamformer (RSAB) using NLMS. The required computational effort would be simplified in frequency domain.

1.3 Thesis Scope and Contribution

The thesis focuses on eliminating multiple directional nonstationary signals using nullformer with adaptive filter. In comparison with beamformer, nullformer makes a null space to the interfering signal which could be used to eliminate the interfering sources.

The scope of the thesis can be divided into two parts: 1. applying nullforming to the adaptive filter, 2. adaptive nullforming technique. The fixed nullformer constructs the null space to the interfering sources before executing the adaptive filter for target speech enhancement. In this case, the interfering sources are assumed unchanged during the adaptation. The adaptive nullformer updates the nullspace in a period to trace the change of interfering sources and corresponding nullspace.

Nullformer is applied to two different adaptive filters in the thesis; they are reference signal based adaptive beamformer (RSAB) and generalized sidelobe canceler (GSC). RSAB uses normalized least mean square (NLMS) to find the weighting of filter. The FBF and BM should be modified to satisfy the architecture when the nullformer is applied to GSC.

For the fixed nullformer, The RTFs of interfering sources are used to find the null space by using singular value decomposition (SVD). For the adaptive nullformer, the RTFs from different directions are estimated before executing the enhancement procedure. These RTFs are used to find the subspace distance between previously known RTFs and estimated signal subspace in real-time. Therefore the existence of desired source in each frequency can be found and processed accordingly.

The proposed adaptive nullformer on GSC is implemented and the experiment compares the performance between the proposed speech enhancement and conventional adaptive beamformer.

1.4 Outlines of Thesis

The thesis can be divided into two parts: The adaptive filter with nullformer and adaptive nullforming algorithm. The topics of each chapter are described as follows.

Chapter 2: The problems are formulated in this chapter. Then the reference signal based adaptive filter (RSAB) would be reviewed, including the architecture and mathematical descriptions

Chapter 3: The linear constrained minimum variance (LCMV) problem would be described. The Frost algorithm would solve the problem. Finally the generalized sidelobe canceler (GSC) using relative transfer function would be derived based on Frost algorithm

Chapter 4: Introducing differential microphone and finding null space of interfering signal using singular value decomposition (SVD). Then Appling fixed nullformer to RSAB and GSC

Chapter 5: The variable nullforming algorithm using order recursive least square estimation (ORLS) and subspace distance. Then a voice activity detection method using the algorithm was proposed. Finally, the variable nullforming is applied to GSC to .

Chapter 6: Experiment results shows the performance of RSAB, GSC, RSAB with nullforming and GSC with nullforming

Chapter 7: Conclusion and future study

Chapter 2

Reference Signal Based Adaptive Beamforming

2.1 Introduction

The time domain reference signal based adaptive beamforming (RSAB) was introduced by Dahl *et al.* [2]. The work in [11] proposed frequency domain RSAB, which optimize the performance at each frequency bin. From RSAB, filter weighting adjustment has two purposes: one is to minimize the interfering sources and noises another is to equalize the channel effect. The architecture of RSAB is discussed in the following section.

2.2 Problem Formulation

Consider an array with M sensors in a noisy reverberant environment receiving one nonstationary desired source and some stationary interfering signals. The received signal in time domain would be

$$x_m(n) = a_m^D(n) \otimes s^D(n) + n_m(n); \quad m = 1, \dots, M \tag{2-1}$$

where each symbol represents:

$\otimes$	convolution operation
$x_m(n)$	signal received by *m*th sensor
$a_m^D(n)$	the transfer function (TF) between desired source and *m*th microphone
$s^D(n)$	desired source
$n_m(n)$	the noise received by *m*th sensor.

The received signal is analyzed frame by frame in frequency domain so the short time Fourier transform (STFT) can be approximately written as

$$x_m(k,\omega) \approx a_m^D(k,\omega)s^D(k,\omega) + n_m(k,\omega); \quad m = 1,\ldots,M \tag{2-2}$$

where ω denotes frequency under kth frame. The approximation is justified for the FFT size be sufficiently large. Assuming that the environment does not change severely thus $a_m^D(\omega) \approx a_m^D(k,\omega)$. The vector formulation of the equation set (2-2) can be written as

$$\mathbf{x}(k,\omega) = \mathbf{a}^D(\omega)s^D(k,\omega) + \mathbf{n}(k,\omega). \tag{2-3}$$

where

$$\mathbf{x}(k,\omega) = \begin{bmatrix} x_1(k,\omega) & x_2(k,\omega) & \cdots & x_M(k,\omega) \end{bmatrix}^T$$

$$\mathbf{a}^D(\omega) = \begin{bmatrix} a_1^D(\omega) & a_2^D(\omega) & \cdots & a_M^D(\omega) \end{bmatrix}^T$$

$$\mathbf{n}(k,\omega) = \begin{bmatrix} n_1(k,\omega) & n_2(k,\omega) & \cdots & n_M(k,\omega) \end{bmatrix}^T.$$

For the case with two or more interfering sources, the TFs of desired source and interfering sources are independent. Therefore the received signal in frequency domain with one desired source and N interfering sources from different directions can be formulated as

$$\begin{aligned}\mathbf{x}(k,\omega) &= \mathbf{a}^D(\omega)s^D(k,\omega) + \sum_{i=1}^{N}\mathbf{a}_i^I(\omega)s_i^I(k,\omega) + \mathbf{n}(k,\omega) \\ &= \mathbf{A}(\omega)\mathbf{s}(k,\omega) + \mathbf{n}(\omega)\end{aligned} \tag{2-4}$$

Where

$$\mathbf{A}(\omega) = \begin{bmatrix} \mathbf{a}^D(\omega) & \mathbf{a}_1^I(\omega) & \cdots & \mathbf{a}_N^I(\omega) \end{bmatrix}$$

$$\mathbf{s}(k,\omega) = \begin{bmatrix} s^D(k,\omega) & s_1^I(k,\omega) & \cdots & s_N^I(k,\omega) \end{bmatrix}^T$$

$$\mathbf{a}_i^I(\omega) = \begin{bmatrix} 1 & a_{i2}^I(\omega) & \cdots & a_{iM}^I(\omega) \end{bmatrix}^T \quad i = 1,\ldots,N$$

is the vector form of TFs between interfering sources and microphone array and $s_i^I(k,\omega)$ is the ith interfering source.

2.3 Reference Signal Based Adaptive Filter

RSAB requires prior information before executing the beamformer. The prior information is pre-recorded signals received by microphone array- $s_1(k,\omega),\ldots,s_M(k,\omega)$ and the reference signal- $r(k,\omega)$. A set of pre-recorded speech signals are collected by placing a source on the desired position and letting the source emit for a short while under quiet environment. The pre-recorded signals provide a priori information between desired source and the microphone array. The reference signal could be the original source or original source received by another microphone in good quality.

After collecting the pre-recorded signal and reference signal, the procedure of the RSAB is divided into two phases- training phase and filtering phase. Figure 2-1 shows the overall system architecture.

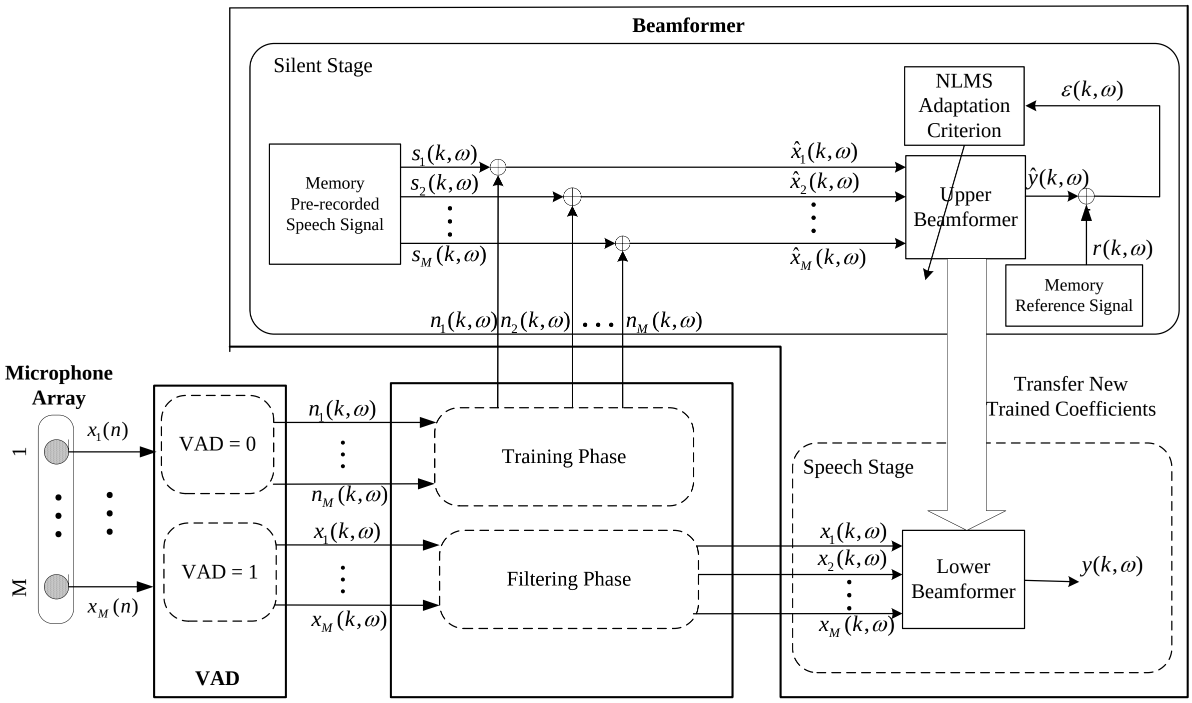


Figure 2-1 Reference signal based adaptive beamformer

In the algorithm, the voice activity detection (VAD) is used to detect the activity of desired signal. When VAD shows that desired signal is inactive, the system started training phase using normalized least mean square (NLMS). For the training phase, the error signal at frequency ω is written as

$$\varepsilon(k,\omega) = r(k,\omega) - \mathbf{w}^{\dagger}(k,\omega)\left[\hat{\mathbf{x}}(k,\omega) + \mathbf{s}(k,\omega)\right] \tag{2-5}$$

where

$$\mathbf{w}(k,\omega) = \left[w_1(k,\omega) \quad w_2(k,\omega) \quad \cdots \quad w_M(k,\omega)\right]^T$$

$$\hat{\mathbf{x}}(k,\omega) = \left[\hat{x}_1(k,\omega) \quad \hat{x}_2(k,\omega) \quad \cdots \quad \hat{x}_M(k,\omega)\right]^T$$

$$\mathbf{s}(k,\omega) = \left[s_1(k,\omega) \quad s_2(k,\omega) \quad \cdots \quad s_M(k,\omega)\right]^T$$

and † denotes complex conjugate transpose. $\varepsilon(k,\omega)$ is error signal. $r(k,\omega)$ is the pre-recorded reference signal. $\mathbf{w}(k,\omega)$ is the filter weighting for adaption. $\hat{\mathbf{x}}(k,\omega)$ is the received signal from microphone array in training phase. And $\mathbf{s}(k,\omega)$ is pre-recorded desire source.

The purpose of RSAB is to minimize the mean square error between received signal and the desired signal. The mean square error is

$$J_{LMS} = \varepsilon^*(k,\omega)\varepsilon(k,\omega).$$

Then minimize the mean square error

$$\begin{aligned}\min_{\mathbf{w}} J_{LMS} &= \min_{\mathbf{w}} \varepsilon^*(k,\omega)\varepsilon(k,\omega) \\ &= \min_{\mathbf{w}} \left[r(k,l) - \mathbf{w}^{\dagger}(k,\omega)\hat{\mathbf{x}}(k,\omega)\right]^* \left[r(k,l) - \mathbf{w}^{\dagger}(k,\omega)\hat{\mathbf{x}}(k,\omega)\right]\end{aligned} \tag{2-6}$$

The optimal solution would be obtained by taking the derivative to previous equation to find a local minimum. But the optimal solution is not practical for implementation. Therefore, the adaptive solution is introduced. For adaptive solution, the weighting

$\mathbf{w}(k,\omega)$ is updated in the steepest direction thus

$$\mathbf{w}(k+1,\omega) = \mathbf{w}(k,\omega) + \mu\left[\frac{\partial J_{LMS}}{\partial \mathbf{w}}\right]. \tag{2-7}$$

From (2-6) and (2-7) , using NLMS algorithm to achieve a stable solution in each frequency. Therefore, the filter weighting update procedure is

$$\mathbf{w}(k+1,\omega) = \mathbf{w}(k,\omega) + \mu\frac{\varepsilon(k,\omega)\hat{\mathbf{x}}^*(k,\omega)}{\gamma + \hat{\mathbf{x}}^\dagger(k,\omega)\hat{\mathbf{x}}(k,\omega)}. \tag{2-8}$$

When VAD detected that the received sound signal contains desired speech signal, the system switched to the filtering phase. The system starts to filter the received signal with $\mathbf{w}$ trained in training phase so

$y(k,\omega) = \mathbf{w}^\dagger(\omega)\mathbf{x}(k,\omega)$.

Where $y(k,\omega)$ denotes output signal, and $\mathbf{x}(k,\omega)$ denotes received signals in filtering phase. The flow of the procedure is described in Figure 2-2.

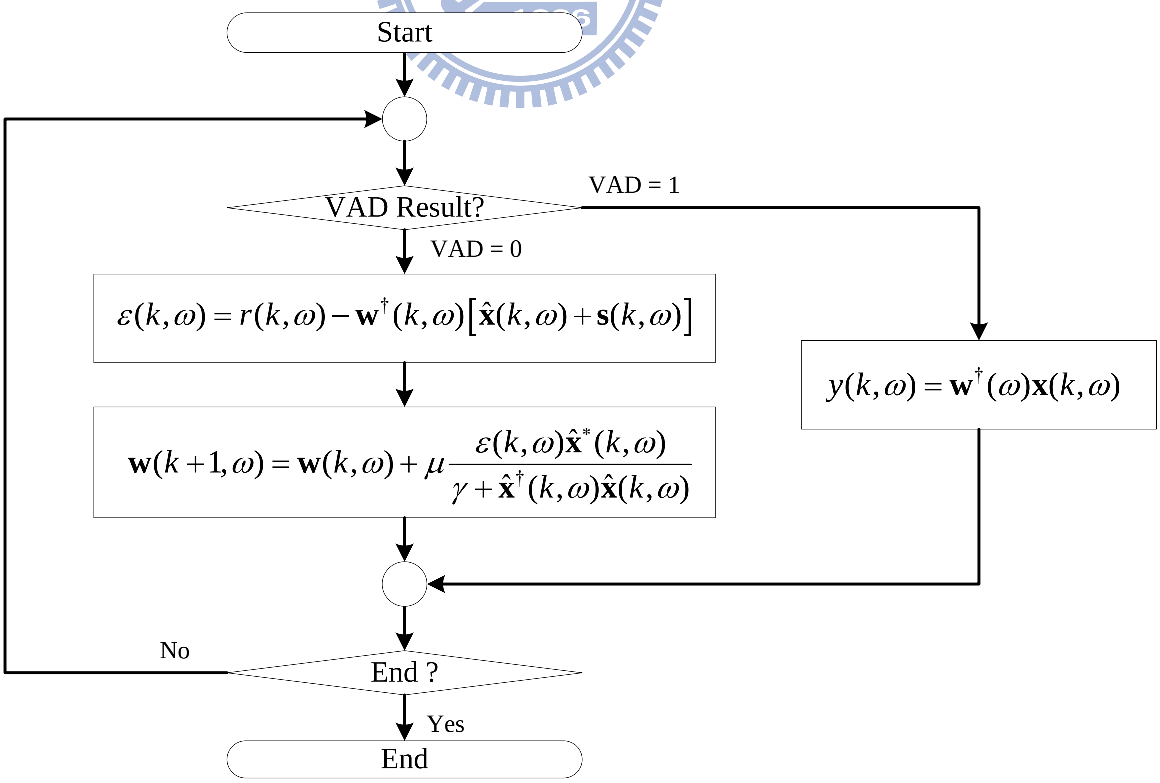


Figure 2-2 Flow of the reference signal based domain adaptive beamformer

Chapter 3

Linear Constrained Minimum Variance Beamforming

3.1 Introduction

Frost [7] proposed a method to minimize the target signal power under constraint. Griffiths and Jim [8] reconsidered the Frost's algorithm and obtained generalized sidelobe canceler (GSC). GSC is widely used to cope with interference signal. Gannot *et al.* [4] applied relative transfer function (RTF) to GSC to enhance the performance when there's a nonstationary desired source in a reverberant room. In this chapter the Frost algorithm is introduced and then RTF GSC.

3.2 Frequency Domain Frost Algorithm

3.2.1 Optimal Solution

Starting from the same problem formulated in section 2.2. The purpose is to find a set of weighting that filter the received signal and obtain the original desired source. The filter weighting in vector form is

$$\mathbf{w}(k,\omega)=\begin{bmatrix} w_1(k,\omega) & w_2(k,\omega) & \cdots & w_M(k,\omega) \end{bmatrix}^T.$$

The set of filter weighting can be used to filter the received signal so the output would be

$$\begin{aligned} y(k,\omega) &= \mathbf{w}^\dagger(k,\omega)\mathbf{x}(k,\omega) \\ &= \mathbf{w}^\dagger(k,\omega)\mathbf{a}(k,\omega)s(k,\omega)+\mathbf{w}^\dagger(k,\omega)\mathbf{n}(k,\omega) \\ &\triangleq y_s(k,\omega)+y_n(k,\omega). \end{aligned} \tag{2-1}$$

Where $y(k,\omega)$ denotes output signal, $y_s(k,\omega)$ denotes signal part of filtered signal and $y_n(k,\omega)$ denotes interfering part of filtered signal. The output power would be

$$\begin{aligned} E\left\{y(k,\omega)y^*(k,\omega)\right\} &= E\left\{\mathbf{w}^\dagger(k,\omega)\mathbf{x}(k,\omega)\mathbf{x}^\dagger(k,\omega)\mathbf{w}(k,\omega)\right\} \\ &= \mathbf{w}^\dagger(k,\omega)\Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega) \end{aligned} \tag{2-2}$$

where $\Phi_{\mathbf{xx}}(k,\omega)$ denotes power spectral density of input signal. The goal is to minimize output power. If there's no constraint for the problem, the trivial solution would be zero. Therefore, a constraint is set as

$$\begin{aligned} y_s(k,\omega) &= \mathbf{w}^\dagger(k,\omega)\mathbf{a}^D(\omega)s(k,\omega) \\ &= f^*(k,\omega)s(k,\omega) \end{aligned} \tag{2-3}$$

where $f^*(k,\omega)$ is a prescribed filter, usually let it a delay. Therefore, the linear constrained minimum variance (LCMV) problem can be formulated as:

$$\min_{\mathbf{w}}\left\{\mathbf{w}^\dagger(k,\omega)\Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega)\right\} \text{ subject to } \mathbf{w}^\dagger(k,\omega)\mathbf{a}(\omega) = f^*(k,\omega) \tag{2-4}$$

Using complex Lagrange multipliers to solve the problem

$$\begin{aligned} \mathcal{L}(\mathbf{w}) = &\ \mathbf{w}^\dagger(k,\omega)\Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega) \\ &+\lambda\left[\mathbf{w}^\dagger(k,\omega)\mathbf{a}(\omega) - c^*(k,\omega)\right] \\ &+\lambda^*\left[\mathbf{a}^\dagger(\omega)\mathbf{w}(k,\omega) - c(k,\omega)\right] \end{aligned}$$

where λ is the Lagrange multiplier. Set the derivative of $\mathcal{L}(\mathbf{w})$ with respect to $\mathbf{w}$ to be zero yields

$$\frac{\partial\mathcal{L}(\mathbf{w})}{\partial\mathbf{w}} = \Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega) + \lambda\mathbf{a}(\omega) = 0 \tag{2-5}$$

By (2-3) and (2-5) , the optimal solution of LCMV problem would be

$$\mathbf{w}(k,\omega) = \left[\mathbf{a}^\dagger(\omega)\Phi_{\mathbf{xx}}{}^{-1}(k,\omega)\mathbf{a}(\omega)\right]^{-1}\Phi_{\mathbf{xx}}{}^{-1}(k,\omega)\mathbf{a}(\omega)\,f(k,\omega) \tag{2-6}$$

3.2.2 Adaptive Solution

The constrained form of the optimal solution is impractical in the real world. It's difficult to find the room impulse response by using system identification method. This constrained form can't tract changes in the environment [4]. So by Frost [7], the adaptive form was introduced, which would be more useful in practical environment. Consider the steepest descent adaptive algorithm:

$$\begin{aligned}\mathbf{w}(k+1,\omega) &= \mathbf{w}(k,\omega) - \mu\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}} \\ &= \mathbf{w}(k,\omega) - \mu\left[\Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega) + \lambda\mathbf{a}(\omega)\right].\end{aligned} \tag{2-7}$$

Imposing the constraint on $\mathbf{w}(k+1,\omega)$. Then

$$\begin{aligned}f(\omega) &= \mathbf{a}^{D\dagger}(\omega)\mathbf{w}(k+1,\omega) \\ &= \mathbf{a}^{D\dagger}(\omega)\mathbf{w}(k,\omega) - \mu\mathbf{a}^{D\dagger}(\omega)\Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega) - \mu\mathbf{a}^{D\dagger}(\omega)\mathbf{a}^{D}(\omega)\lambda\end{aligned}.$$

Solving the Lagrange multiplier yields

$$\mathbf{w}(k+1,\omega) = \mathbf{P}(\omega)\mathbf{w}(k,\omega) - \mu\mathbf{P}(\omega)\Phi_{\mathbf{xx}}(k,\omega)\mathbf{w}(k,\omega) + \mathbf{f}(\omega) \tag{2-8}$$

Where

$$\mathbf{P}(\omega) = \mathbf{I} - \frac{\mathbf{a}^{D}(\omega)\mathbf{a}^{D\dagger}(\omega)}{\left\|\mathbf{a}^{D}(\omega)\right\|^2} \in \mathcal{N}(\mathbf{a}^{D\dagger}(\omega)) \tag{2-9}$$

$$\mathbf{f}(\omega) = \frac{\mathbf{a}^{D}(\omega)}{\left\|\mathbf{a}^{D}(\omega)\right\|^2} f(\omega) \in \mathcal{R}(\mathbf{a}^{D\dagger}(\omega)) \tag{2-10}$$

$\mathbf{P}(\omega)$ is the projection matrix that project vector to the null space of $\mathbf{a}^{D\dagger}(\omega)$. And $\mathcal{N}(\mathbf{a}^{D\dagger}(\omega))$ represents the null space of $\mathbf{a}^{D\dagger}(\omega)$. $\mathbf{f}(\omega)$ is the range space of $\mathbf{a}^{D\dagger}(\omega)$ and $\mathcal{R}\left(\mathbf{a}^{D\dagger}(\omega)\right)$ represents the range space of $\mathbf{a}^{D\dagger}(\omega)$. From (2-2), replacing $\Phi_{\mathbf{xx}}(k,\omega)$ by $E\left\{\mathbf{x}(k,\omega)\mathbf{x}^{\dagger}(k,\omega)\right\}$ and rearrange (2-7), the adaptive Frost algorithm would be

$$\mathbf{w}(k+1,\omega) = \mathbf{P}(\omega)\left[\mathbf{w}(k,\omega) - \mu\mathbf{x}(k,\omega)y^{*}(k,\omega)\right] + \mathbf{f}(\omega) \tag{2-11}$$

3.3 Generalized Sidelobe Canceler

From the frost algorithm, the filter weighting could be separated into two parts; the first part is the range space of $\mathbf{a}^D(\omega)$ and the second part is the null space of $\mathbf{a}^D(\omega)$. Hence

$$\mathbf{w}(k,\omega)=\mathbf{w}_{\mathrm{FBF}}(k,\omega)-\mathbf{w}_{\mathrm{ANC}}(k,\omega) \tag{2-12}$$

where

$$\mathbf{w}_{\mathrm{FBF}}(k,\omega)\in\mathcal{R}(\mathbf{a}^{D\dagger}(\omega)) \quad \text{and} \quad -\mathbf{w}_{\mathrm{ANC}}(k,\omega)\in\mathcal{N}(\mathbf{a}^{D\dagger}(\omega))$$

comparing the filter weighting with adaptive Frost algorithm, let

$$\mathbf{w}_{\mathrm{FBF}}(k,\omega)=\mathbf{f}(\omega)=\frac{\mathbf{a}^D(\omega)}{\left\|\mathbf{a}^D(\omega)\right\|^2}f(\omega) \tag{2-13}$$

and

$$\mathbf{w}_{\mathrm{ANC}}(k,\omega)=\mathbf{P}(\omega)\mathbf{g}(k,\omega)\,. \tag{2-14}$$

From (2-12), (2-13) and (2-14),the output signal would be

$$\begin{aligned} y(k,\omega)&=\mathbf{w}^\dagger(k,\omega)\mathbf{x}(k,\omega)\\ &=\mathbf{w}^\dagger_{\mathrm{FBF}}(k,\omega)\mathbf{x}(k,\omega)-\mathbf{w}^\dagger_{\mathrm{NC}}(k,\omega)\mathbf{x}(k,\omega)\\ &=\mathbf{f}^\dagger(\omega)\mathbf{x}(k,\omega)-\mathbf{g}^\dagger(k,\omega)\mathbf{P}^\dagger(\omega)\mathbf{x}(k,\omega)\\ &\triangleq y_{\mathrm{FBF}}(k,\omega)-y_{\mathrm{ANC}}(k,\omega)\end{aligned} \tag{2-15}$$

The architecture of GSC could be separated into three parts, fixed beamformer (FBF), blocking matrix (BM) and adaptive noise canceler (ANC). The purpose of FBF is to obtain the signal that contains the desired source and a stationary noise. The BM blocks the desired source to extract the stationary noise. Then the ANC uses multichannel wiener filter to estimate the noise of FBF and cancel the noise. Figure 3-1 shows the entire architecture of GSC. Detail discussion on each element in the architecture is given as follows.

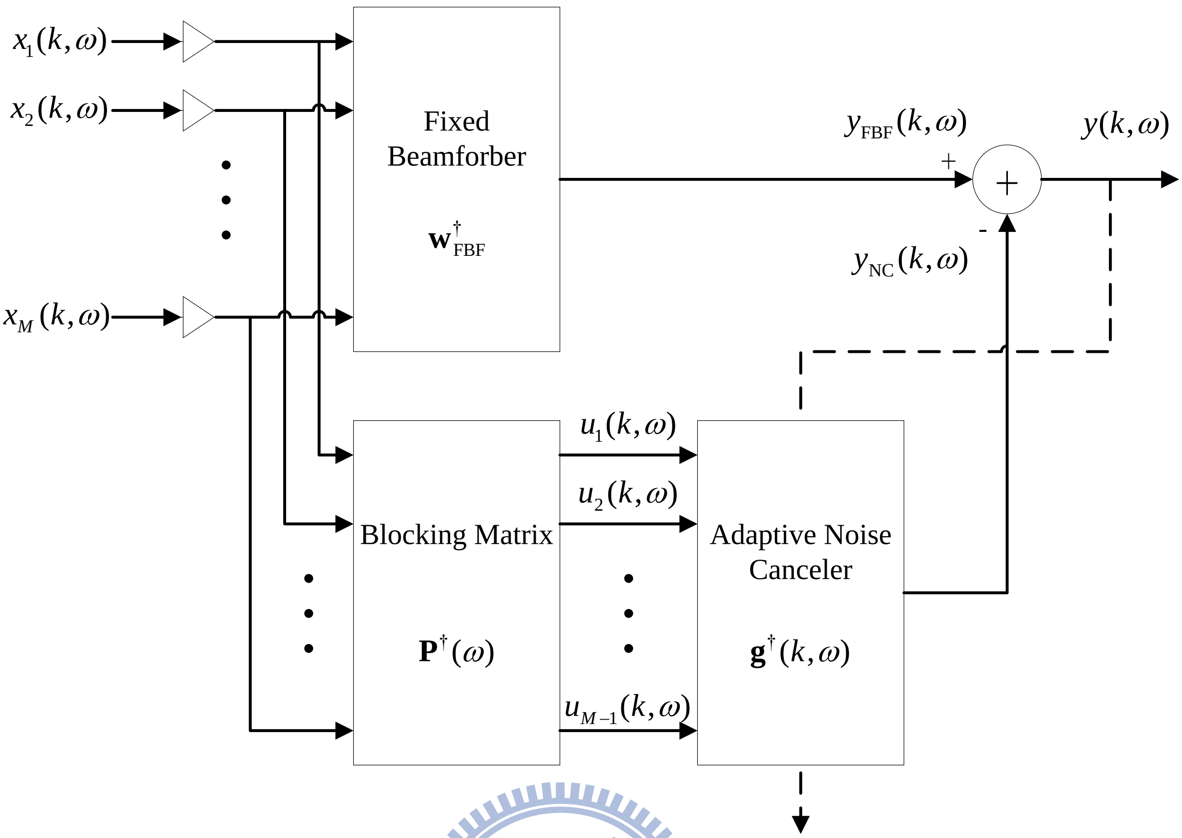


Figure 3-1 Generalized sidelobe canceler

1. *Fixed Beamformer (FBF)*:

From (2-15), the output of FBF is

$$
\begin{aligned}
y_{\mathrm{FBF}}(k,\omega) &= \mathbf{f}^{\dagger}(\omega)\mathbf{x}(k,\omega) \\
&= \frac{\mathbf{a}^{D\dagger}(\omega)}{\left\|\mathbf{a}^{D}(\omega)\right\|^{2}} f^{*}(\omega)\left[\mathbf{a}^{D}(\omega)s(k,\omega)+\mathbf{n}(k,\omega)\right] \\
&= f^{*}(\omega)s(k,\omega)+\frac{\mathbf{a}^{D\dagger}(\omega)\mathbf{n}(k,\omega)}{\left\|\mathbf{a}^{D}(\omega)\right\|^{2}} f^{*}(\omega)
\end{aligned}
\qquad (2\text{-}16)
$$

Because $f^{*}(\omega)$ is just a simple delay, the output of FBF contains undistorted desired source and noise. This is an optimal solution where the desired source is just a simple delay from output of FBF. The issue of the optimal solution is that the actual TFs are difficult to find so Gannot *et al*. applied relative transfer function (RTF) to GSC to approach the suboptimal solution [4]. The RTF is easy to obtained by system identification method proposed by [12]. RTF is the ratio of RIR between two microphones. Let the first microphone be the reference microphone then the RTF is

$$h_m^D(k,\omega)=\frac{a_m^D(\omega)}{a_1^D(\omega)} \quad m=1,\ldots,M\,. \tag{2-17}$$

Take the vector form

$$\mathbf{h}^D=\begin{bmatrix}1 & h_1^D(\omega) & \cdots & h_M^D(\omega)\end{bmatrix}^T=\frac{\mathbf{a}^{DT}(\omega)}{a_1^D(\omega)}$$

If the actual TFs in (2-13) are replaced by RTFs, the FBF would be

$$\mathbf{w}_{\mathrm{FBF}}(k,\omega)=\frac{\mathbf{h}^D(\omega)}{\left\|\mathbf{h}^D(\omega)\right\|^2}f(\omega)\,. \tag{2-18}$$

By (2-15) and (2-18) ,the output of FBF would be

$$\begin{aligned}y_{\mathrm{FBF}}(k,\omega)&=\mathbf{w}_{\mathrm{FBF}}^{\dagger}(\omega)\mathbf{x}(k,\omega)\\&=\frac{\mathbf{h}^{D\dagger}(\omega)}{\left\|\mathbf{h}^D(\omega)\right\|^2}f^*(\omega)\left[\mathbf{a}^D(\omega)s(k,\omega)+\mathbf{n}(k,\omega)\right]\\&=a_1(\omega)f^*(\omega)s(k,\omega)+\frac{\mathbf{a}^{D\dagger}(\omega)\mathbf{n}(k,\omega)}{\left\|\mathbf{a}^D(\omega)\right\|^2}f^*(\omega)\end{aligned} \tag{2-19}$$

Therefore, a suboptimal solution is obtained with signal distorted by the transfer function of desired source to the first microphone.

2. Blocking Matrix(BM):

The BM using RTFs could properly block the desired signal. Therefore, the columns of BM are the bases of desired signal null space. Therefore, considering the following matrix

$$\mathbf{P}^*(\omega)=\begin{bmatrix} -\frac{a_2^*(\omega)}{a_1^*(\omega)} & -\frac{a_3^*(\omega)}{a_1^*(\omega)} & \cdots & -\frac{a_M^*(\omega)}{a_1^*(\omega)} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Then the output of BM is

$$\begin{aligned} u_m(k,\omega) &= x_m(k,\omega)-\frac{a_m^D(\omega)}{a_1^D(\omega)}x_1(k,\omega) \\ &= a_m^D(\omega)s^D(k,\omega)+n_m(\omega)-\frac{a_m^D(\omega)}{a_1^D(\omega)}\left[a_1^D(\omega)s^D(k,\omega)+n_1(\omega)\right] \\ &= n_m(\omega)-\frac{a_m^D(\omega)}{a_1^D(\omega)}n_1(\omega) \qquad\qquad m=1,\ldots,M-1. \end{aligned} \tag{2-20}$$

Therefore, output of BM would be the noise only signal.

From the criterion of GSC, the output of blocking matrix should be independent of k because the noise is assumed stationary. But in practical, the BM cannot block the entire desired signal. Thus the output of BM would be changed under nonstationary source. Thus the vector form of output from BM is

$$\begin{aligned} \mathbf{u}(k,\omega) &= \mathbf{P}^\dagger(\omega)\mathbf{x}(k,\omega) \\ &= \mathbf{P}^\dagger(\omega)\left[\mathbf{a}(\omega)s(k,\omega)+\mathbf{n}(k,\omega)\right] \\ &= \mathbf{P}^\dagger(\omega)\mathbf{n}(k,\omega). \end{aligned} \tag{2-21}$$

3. *Adaptive noise canceler (ANC)*:

The output of ANC would be noise only signal because $\mathbf{P}^\dagger(\omega)$ is the null space of desired signal. Thus by (2-15) and (2-21), the output of ANC is

$$\begin{aligned} y_{\mathrm{ANC}}(k,\omega) &= \mathbf{g}^\dagger(k,\omega)\mathbf{u}(k,\omega) \\ &= \mathbf{g}^\dagger(k,\omega)\mathbf{P}^\dagger(\omega)\mathbf{x}(k,\omega) \\ &= \mathbf{g}^\dagger(k,\omega)\mathbf{P}^\dagger(\omega)\left[\mathbf{a}(\omega)s(k,\omega)+\mathbf{n}(k,\omega)\right] \\ &= \mathbf{g}^\dagger(k,\omega)\mathbf{P}^\dagger(\omega)\mathbf{n}(k,\omega) \end{aligned} \tag{2-22}$$

Recalling (2-4), the purpose is to minimize the output power so minimize

$$E\left\{\left\|y_{\mathrm{FBF}}(k,\omega)-\mathbf{g}^{\dagger}(k,\omega)\mathbf{u}(k,\omega)\right\|^{2}\right\}$$

Then the multichannel Wiener filter would be

$$\mathbf{g}(k,\omega)=\Phi_{\mathbf{uu}}^{-1}(k,\omega)\Phi_{\mathbf{u}y}(k,\omega) \tag{2-23}$$

where

$$\Phi_{\mathbf{uu}}(k,\omega)=E\left\{\mathbf{u}(k,\omega)\mathbf{u}^{\dagger}(k,\omega)\right\}$$

$$\Phi_{\mathbf{u}y}(k,\omega)=E\left\{\mathbf{u}(k,\omega)y_{\mathrm{FBF}}^{*}(k,\omega)\right\}.$$

To track the change of environment and achieve a stable solution, Gannot *et al.* use NLMS algorithm to recursively update the weighting of ANC [4], thus the weighting of ANC would be

$$g_m(k+1,\omega)=g_m(k,\omega)+\mu\frac{u_m(k,\omega)y^{*}(k,\omega)}{P_{est}(k,\omega)}\quad m=1,\ldots,M-1. \tag{2-24}$$

By [4], let

$$P_{est}(k,\omega)=\rho P_{est}(k-1,\omega)+(1-\rho)\sum_{m}\left|x(k,\omega)\right|^{2} \tag{2-25}$$

where ρ is a forgetting factor. $P_{est}(k,\omega)$ can be $\mathbf{u}^{\dagger}(k,\omega)\mathbf{u}(k,\omega)$, which also normalize the power and make the recursion more stable.

Chapter 4

Nullforming

4.1 Introduction

In this chapter, several methods are introduced to achieve nullforming and the associated algorithms are explained for the adaptive filters to enhance the desired source.

4.2 Differential Microphone

Delay and sum beamformer is commonly used under both the far field and free field assumptions. The method enhances the signal from desired direction

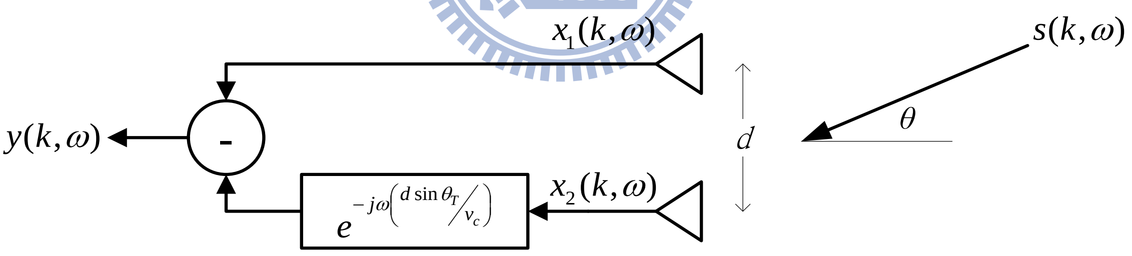


Figure 4-1 Differential microphone

Elko *et al*. [13] proposed differential microphone to reduce the signal from target direction. Figure 4-1 shows the architecture of differential microphone. The method makes a nullforming using two microphones subtraction with delay compensation. The signal is assumed a far field plain wave and the microphones are perfectly matched thus the output of differential microphone pairs can be written as,

$$y_{\mathrm{DM}}(k,\omega)=\frac{1}{2}\left[x_1(k,\omega)-e^{-j\omega\left(d\sin\theta_T/v_c\right)}x_2(k,\omega)\right]$$
$$=\frac{1}{2}\left[e^{-j\omega\left(r/v_c\right)}-e^{-j\omega\left(d\sin\theta_T/v_c\right)}e^{-j\omega\left(r+d\sin\theta/v_c\right)}\right]s(k,\omega) \tag{3-1}$$

where $y_{\mathrm{DM}}(k,\omega)$ is the output of differential microphone pairs, $s(k,\omega)$ is the source, d is the distance between two microphones, r is the distance between source and microphones, v_c is the speed of sound, θ is the direction of source, θ is the target direction of differential microphone, $x_1(k,\omega)$ and $x_2(k,\omega)$ are signals received by microphones. Rearrange the formula, the magnitude of output would be

$$\left|y_{DM}(k,\omega)\right|=\frac{1}{2}\left|x_1(k,\omega)-e^{-j\omega\left(d\sin\theta_T/v_c\right)}x_2(k,\omega)\right|$$
$$=\frac{1}{2}\left|e^{-j\omega\left(r/v_c\right)}\right|\left|1-e^{-j\omega\left(d(\sin\theta+\sin\theta_T)/v_c\right)}\right|\left|s(k,\omega)\right| \tag{3-2}$$
$$=\left|\sin\left[\frac{\omega d}{2v_c}\left(\sin\theta+\sin\theta_T\right)\right]\right|\left|s(k,\omega)\right|$$

Figure 4-2 shows the beam patterns of differential microphone plotted by using equation (3-2). The beam patterns are plotted under different distance of microphones by letting the magnitude of source be 1, the speed of sound be 343 m/s and the target direction be at 0˚.

The differential microphone works like high pass filter so differential microphone would enhance the noises in high frequencies. Different distance of microphones would affect the ability to deal with different frequency band. For the short distance differential microphone, lower band of frequencies would be eliminated from almost every direction.

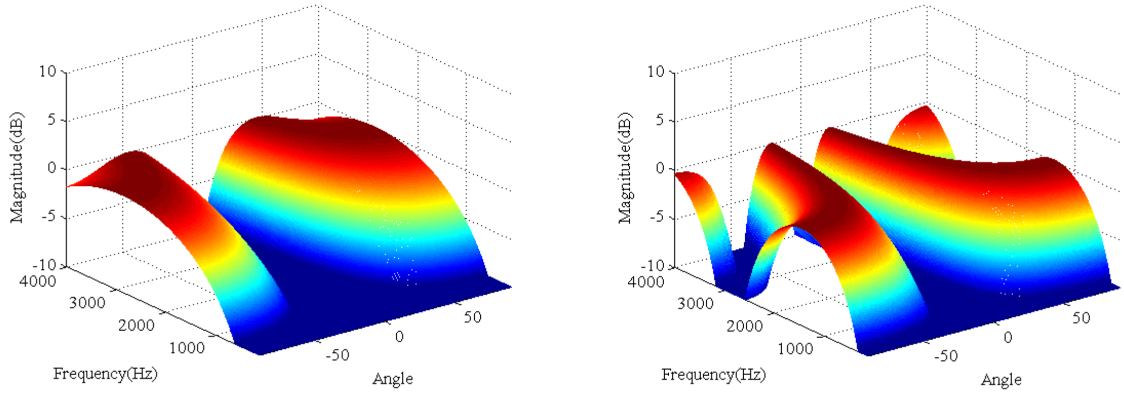


Figure 4-2 Beam pattern of differential microphone with *d*=0.12 m (left) and *d*=0.24 m (right)

4.3 Nullforming Using Null Space of Interfering Signal

Previous section shows a nullformer for one interfering source. There may be two or more interfering sources in practical environment. The thesis uses singular value decomposition (SVD) to find the null space of the entire interfering signal. Assume there are N interfering sources in the environment and we have the RTFs of them as described by (2-17) in chapter 3. The RTFs of interfering sources are

$$\mathbf{h}_i^I(\omega) = \begin{bmatrix} 1 & h_{i2}^I(\omega) & \cdots & h_{iM}^I(\omega) \end{bmatrix}^T = \frac{\mathbf{a}_i^{IT}(\omega)}{a_{i1}^I(\omega)} \quad i = 1, \ldots, N$$

and take the complex conjugate of these RTFs in matrix form

$$\mathbf{H}^I(\omega) = \begin{bmatrix} \mathbf{h}_1^{I*}(\omega) & \mathbf{h}_2^{I*}(\omega) & \cdots & \mathbf{h}_N^{I*}(\omega) \end{bmatrix}$$

which are the bases of interfering signal subspace. Then applying singular value decomposition (SVD) to $\mathbf{H}^I(\omega)$

$$\mathbf{H}^I(\omega) = \mathbf{U}(\omega)\mathbf{S}(\omega)\mathbf{V}^\dagger(\omega) \tag{3-3}$$

Where

$$\mathbf{U}(\omega) = \begin{bmatrix} \mathbf{u}_1(\omega) & \mathbf{u}_2(\omega) & \cdots & \mathbf{u}_M(\omega) \end{bmatrix}$$

are the eigenvectors of $\mathbf{H}(\omega)\mathbf{H}^\dagger(\omega)$

$$\mathbf{V}(\omega)=\begin{bmatrix}\mathbf{v}_1(\omega) & \mathbf{v}_2(\omega) \cdots & \mathbf{v}_N(\omega)\end{bmatrix}$$

are the eigenvectors of $\mathbf{H}^{I\dagger}(\omega)\mathbf{H}^{I}(\omega)$

$$\mathbf{S}(\omega)=\begin{bmatrix}\sigma_1(\omega) & 0 & \cdots & 0 \\ 0 & \sigma_2(\omega) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_M(\omega)\end{bmatrix}$$

are the singular values. These singular values are eigenvalues of $\mathbf{H}^{I\dagger}(\omega)\mathbf{H}^{I}(\omega)$ and $\mathbf{H}^{I}(\omega)\mathbf{H}^{I\dagger}(\omega)$. From [14], the zero eigenvectors of $\mathbf{H}^{I\dagger}(\omega)\mathbf{H}^{I}(\omega)$ corresponds to the zero singular values

$$\sigma_i(\omega)=0, \quad \mathbf{v}_i(\omega)=0 \quad i=N+1,\ldots,M.$$

$\mathbf{u}_i(\omega)$ corresponds to zero singular value for $i=N+1,\ldots,M$ thus

$$\mathbf{H}^{I\dagger}(\omega)\mathbf{u}_i(\omega)=\sigma_i(\omega)\mathbf{v}_i(\omega)=0 \quad i=N+1,\ldots,M.$$

Therefore, $\mathbf{u}_i(\omega)$ are the null space bases of $\mathbf{H}(\omega)^{\dagger}$ for $i=N+1,\ldots,M$, That is

$$\begin{aligned}\mathbf{H}^{I\dagger}(\omega)\mathbf{u}_i(\omega)&=\begin{bmatrix}\mathbf{h}_1^I(\omega) & \mathbf{h}_2^I(\omega) & \cdots & \mathbf{h}_N^I(\omega)\end{bmatrix}^T\mathbf{u}_i(\omega)\\ &=\begin{bmatrix}\dfrac{\mathbf{a}_1^I(\omega)}{a_{11}^I(\omega)} & \dfrac{\mathbf{a}_2^I(\omega)}{a_{21}^2(\omega)} & \cdots & \dfrac{\mathbf{a}_N^I(\omega)}{a_{N1}^N(\omega)}\end{bmatrix}^T\mathbf{u}_i(\omega)\\ &=\mathbf{0} \qquad\qquad\qquad\qquad i=N+1,\ldots,M\end{aligned}$$

Where $\mathbf{0}=\begin{bmatrix}0 & \cdots & 0\end{bmatrix}^T$ is a zero vector. Therefore,

$$\mathbf{u}_i(\omega)\in\mathcal{N}\left(\mathbf{a}_1^{I*}(\omega) \quad \mathbf{a}_2^{I*}(\omega) \quad \cdots \quad \mathbf{a}_N^{I*}(\omega)\right) \quad i=N+1,\ldots,M$$

This null space is a fixed nullformer where

$$\mathbf{U}_{FN}(\omega)=\begin{bmatrix}\mathbf{u}_{N+1}(\omega) & \mathbf{u}_{N+2}(\omega) \cdots & \mathbf{u}_M(\omega)\end{bmatrix} \tag{3-4}$$

is an M input and N output filter.

In the following section, the fixed nullformer would be applied to RSAB and GSC.

4.4 Reference Signal Based Adaptive Filter with Fixed Nullforming

The nullformer could be used to block the interfering signals thus applying the nullformer to RSAB would eliminate the residual noise and reconstruct the desired source. Figure 4-3 shows the architecture of adaptive filter with nullformer. The effect of adaptive filter with nullformer could be considered as the convolution of room impulse response and impulse response of nullformer system.

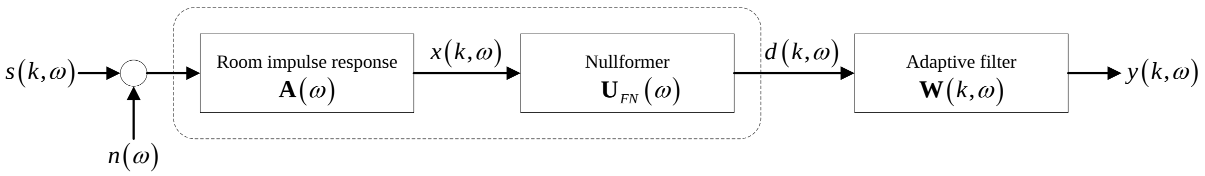


Figure 4-3 System of RSAB with fixed nullformer

For the case with multiple interfering sources described in (2-3).Let the multiplication of RIR and nullformer be a new room impulse

$$\mathbf{R}(\omega)=\mathbf{U}_{FN}^{\dagger}(\omega)\mathbf{A}(\omega) \tag{3-5}$$

and the new input of adaptive filter would be

$$\begin{aligned}\mathbf{d}(k,\omega)&=\mathbf{U}_{FN}^{\dagger}(\omega)\mathbf{x}(k,\omega)\\&=\mathbf{U}_{FN}^{\dagger}(\omega)\left[\mathbf{A}(\omega)\mathbf{s}(k,\omega)+\mathbf{n}(\omega)\right]\\&=\left[\mathbf{U}_{FN}^{\dagger}\mathbf{a}^{D}s^{D}(k,\omega)+\sum_{i=1}^{N}\mathbf{U}_{FN}^{\dagger}\mathbf{a}_{i}^{I}s_{i}^{I}(k,\omega)\right]+\mathbf{U}_{FN}^{\dagger}(\omega)\mathbf{n}(\omega)\\&=\mathbf{U}_{FN}^{\dagger}\mathbf{a}^{D}s^{D}(k,\omega)+\mathbf{U}_{FN}^{\dagger}(\omega)\mathbf{n}(\omega)\end{aligned} \tag{3-6}$$

where

$$\mathbf{d}(k,\omega)=\left[d_{1}(k,\omega)\quad d_{2}(k,\omega)\quad\cdots\quad d_{M-N}(k,\omega)\right]^{T} \tag{3-7}$$

is the output of nullformer. Therefore, the input of adaptive filter would be *M*-*N* channels.

The nullformr would cause a great distortion for it's a high pass filter. Therefore, the reference signal of RSAB would be used to reconstruct the desired signal. In the pre-recording procedure showed in Figure 4-4, the pre-recorded signals- $s_1(k,\omega),\ldots,s_{M-N}(k,\omega)$ are received by the output of nullformer and the reference signal- $r(k,\omega)$ would be the desired signal with good quality.

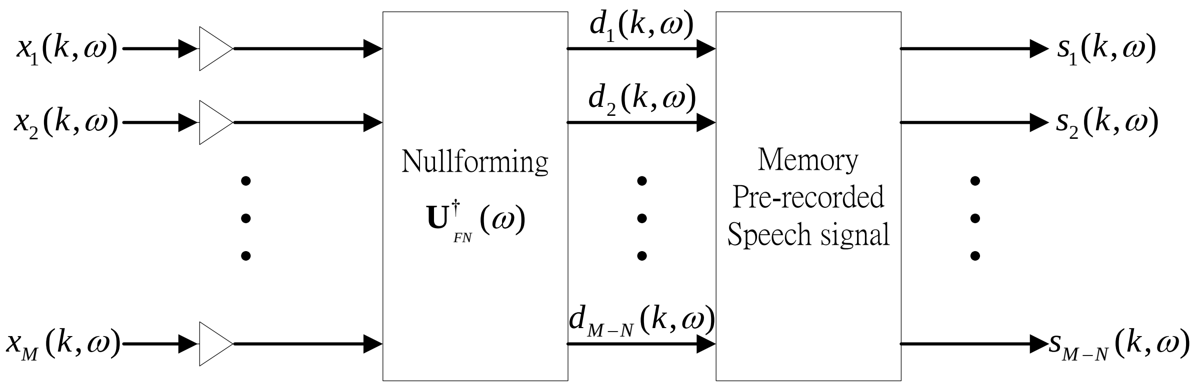


Figure 4-4 Pre-recording procedure of RSAB

The procedures of training phase and filtering phase are the same as described in section 2.3. The only difference is that there's a nullformer before the input of adaptive filter. Figure 4-5 shows the architecture of RSAB.

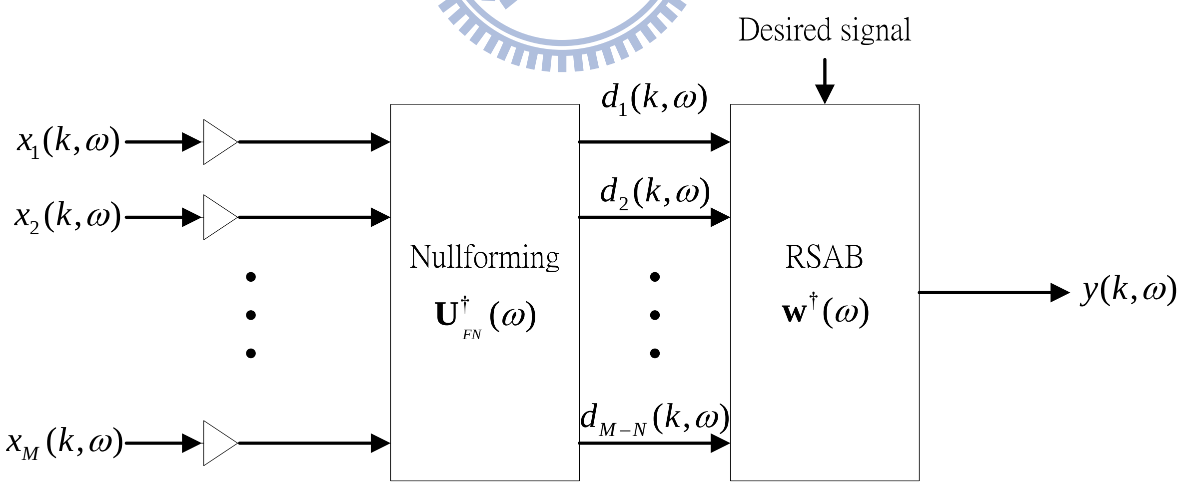


Figure 4-5 System architecture of RSAB

4.5 Generalized Sidelobe Canceler with Fixed Nullforming

Ordinary GSC does not work for the condition with nonstationary interfering signal in the environment. The existence of nonstationary signal does not satisfy the criterion of GSC. The work in [10] proposed a dual-source transfer function GSC (DTF-GSC) method to eliminate a directional nonstationary source by modify the FBF and BM. DTF-GSC could block one nonstationary source. But when there are two or more interfering sources, DTF-GSC is not effective in blocking all these sources.

There are some features when applying the nullformer to GSC. Figure 4-2 shows that the nullformer is a high pass filter. The high pass feature would cause the received signal a great distortion. Therefore, the fixed beamformer and blocking matrix must be modified to satisfy the architecture of GSC with nullformer.

From (3-5), the effect of nullforming is the multiplication of RIR and impulse response of nullformer. Multiply the nullformer weighting from (3-8) with desired signal RTF. Then the new RTF is

$$\mathbf{h}^{Null}(\omega) = \mathbf{U}_{FN}^{\dagger}(\omega)\mathbf{h}^{D}(\omega) \tag{3-9}$$

Where

$$\mathbf{h}^{D}(\omega) = \begin{bmatrix} 1 & \frac{a_2^D(\omega)}{a_1^D(\omega)} & \cdots & \frac{a_2^D(\omega)}{a_1^D(\omega)} \end{bmatrix}^T = \frac{\mathbf{a}^D(\omega)}{a_1^D(\omega)}$$

is the desired signal RTF and

$$\mathbf{h}^{Null}(\omega) = \begin{bmatrix} h_1^{Null}(\omega) & h_2^{Null}(\omega) & \cdots & h_{M-N}^{Null}(\omega) \end{bmatrix}^T$$

is the new RTF, which is the null space of interfering signals, from (3-4)

$$\mathbf{h}^{Null\dagger}(\omega)\mathbf{a}_1^{I\dagger}(\omega) = \mathbf{h}^{D\dagger}(\omega)\mathbf{U}_{FN}(\omega)\mathbf{a}_1^{I\dagger}(\omega) = 0 .$$

Apply SVD to the new obtained RTF

$$\mathbf{h}^{Null}(\omega) = \boldsymbol{\mathcal{U}}(\omega)\boldsymbol{\mathcal{S}}(\omega)\boldsymbol{\mathcal{V}}^{H}(\omega) \tag{3-10}$$

Where

$$\boldsymbol{\mathcal{U}}(\omega) = \begin{bmatrix} \boldsymbol{\mu}_1(\omega) & \boldsymbol{\mu}_2(\omega) & \cdots & \boldsymbol{\mu}_{M-N}(\omega) \end{bmatrix}^T$$

are the eigenvectors of $\mathbf{h}^{Null}(\omega)\mathbf{h}^{Null\dagger}(\omega)$.

$$\mathcal{S}(\omega)=\begin{bmatrix}\delta_1(\omega) & 0 & \cdots & 0\\ 0 & \delta_2(\omega) & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & \delta_{M-N}(\omega)\end{bmatrix}$$

are the singular values. These singular values are eigenvalues of $\mathbf{h}^{Null}(\omega)\mathbf{h}^{Null\dagger}(\omega)$ and $\mathbf{h}^{Null\dagger}(\omega)\mathbf{h}^{Null}(\omega)$.

$$\boldsymbol{\mathcal{V}}(\omega)=\left[\boldsymbol{\nu}_1(\omega)\quad \boldsymbol{\nu}_2(\omega)\quad \cdots\quad \boldsymbol{\nu}_{M-N}(\omega)\right]^T$$

are the eigenvectors of $\mathbf{h}^{Null\dagger}(\omega)\mathbf{h}^{Null}(\omega)$. The zero singular values correspond to zero eigenvectors so

$$\mathbf{h}^{Null\dagger}(\omega)\boldsymbol{\mu}_i(\omega)=\begin{cases}\delta_1(\omega)\boldsymbol{\nu}_1(\omega)\\ \delta_i(\omega)\boldsymbol{\nu}_i(\omega)=\mathbf{0} \quad i=2,\ldots,M-N\,.\end{cases} \tag{3-11}$$

For $\boldsymbol{\nu}_1(\omega)$ is an 1×1 vector, let

$$\mathbf{w}_n(\omega)=\frac{\boldsymbol{\mu}_1(\omega)}{\delta_1(\omega)\boldsymbol{\nu}_1(\omega)} \tag{3-12}$$

And multiply the RTF $\mathbf{h}^{Null\dagger}(\omega)$ with $\mathbf{w}_u(\omega)$ (3-13)

$$\begin{aligned}\mathbf{h}^{Null\dagger}(\omega)\mathbf{w}_n(\omega)&=\mathbf{h}^{D\dagger}(\omega)\mathbf{U}_{FN}(\omega)\mathbf{w}_n(\omega)\\ &=\frac{\mathbf{a}^{D\dagger}(\omega)}{a_1^{D*}(\omega)}\mathbf{U}_{FN}(\omega)\frac{\boldsymbol{\mu}_1(\omega)}{\delta_1(\omega)\boldsymbol{\nu}_1(\omega)}=1\end{aligned}$$

Thus

$$\mathbf{a}^{D\dagger}(\omega)\mathbf{U}_{FN}(\omega)\frac{\boldsymbol{\mu}_1(\omega)}{\delta_1(\omega)\boldsymbol{\nu}_1(\omega)}=a_1^{D*}(\omega) \tag{3-14}$$

Therefore, the new FBF would be obtained

$$\mathbf{w}_{NFBF}(\omega)=\mathbf{U}_{FN}(\omega)\mathbf{w}_n(\omega) \tag{3-15}$$

By (3-14) (3-15), The output of FBF would be

$$
\begin{aligned}
y_{\mathrm{FBF}}(k,\omega) &= \mathbf{w}^{\dagger}_{NFBF}(\omega)\mathbf{x}(k,\omega) \\
&= \mathbf{w}^{\dagger}_{u}(\omega)\mathbf{U}^{\dagger}_{FN}(\omega)\mathbf{x}(k,\omega) \\
&= \frac{\boldsymbol{\mu}^{\dagger}_{1}(\omega)}{\delta^{*}_{1}(\omega)\boldsymbol{\nu}^{\dagger}_{1}(\omega)}\mathbf{U}^{\dagger}_{FN}(\omega)\left[\mathbf{a}^{D}(\omega)s^{D}(k,\omega)+\sum_{i=1}^{N}\mathbf{a}^{I}_{i}(\omega)s^{I}_{i}(k,\omega)+\mathbf{n}(\omega)\right] \\
&= \left[\mathbf{a}^{D\dagger}(\omega)\mathbf{U}_{FN}(\omega)\frac{\boldsymbol{\mu}_{1}(\omega)}{\delta_{1}(\omega)\boldsymbol{\nu}_{1}(\omega)}\right]^{\dagger}s^{D}(k,\omega)+\frac{\boldsymbol{\mu}^{\dagger}_{1}(\omega)}{\delta^{*}_{1}(\omega)\boldsymbol{\nu}^{\dagger}_{1}(\omega)}\mathbf{U}^{\dagger}_{FN}(\omega)\mathbf{n}(\omega) \\
&= a^{D}_{1}(\omega)s^{D}(k,\omega)+\frac{\boldsymbol{\mu}^{\dagger}_{1}(\omega)}{\delta^{*}_{1}(\omega)\boldsymbol{\nu}^{\dagger}_{1}(\omega)}\mathbf{U}^{\dagger}_{FN}(\omega)\mathbf{n}(\omega)
\end{aligned}
\tag{3-16}
$$

The FBF would block the interfering signal and make a beam to the desired signal. Let

$\mathbf{P}_{n}(\omega)=\left[\boldsymbol{\mu}_{2}(\omega)\quad \boldsymbol{\mu}_{2}(\omega)\quad \cdots \quad \boldsymbol{\mu}_{M-N-1}(\omega)\right]$.

The blocking matrix would be

$$
\mathbf{P}_{NBM}(\omega)=\mathbf{U}_{FN}(\omega)\mathbf{P}_{n}(\omega) \tag{3-17}
$$

Then the output of BM is

$$
\begin{aligned}
\mathbf{u}(k,\omega) &= \mathbf{P}^{\dagger}_{NBM}(\omega)\mathbf{x}(k,\omega) \\
&= \mathbf{P}^{\dagger}_{n}(\omega)\mathbf{U}^{\dagger}_{FN}(\omega)\left[\mathbf{a}^{D}(\omega)s^{D}(k,\omega)+\sum_{i=1}^{N}\mathbf{a}^{I}_{i}(\omega)s^{I}_{i}(k,\omega)+\mathbf{n}(\omega)\right] \\
&= \mathbf{P}^{\dagger}_{n}(\omega)^{\dagger}\mathbf{U}^{\dagger}_{FN}(\omega)\left[\mathbf{a}^{D}(\omega)s^{D}(k,\omega)+\mathbf{n}(\omega)\right] \\
&= \left[\mathbf{a}^{D\dagger}(\omega)\mathbf{U}_{FN}(\omega)\boldsymbol{\mu}_{2}(\omega)\quad \cdots \quad \mathbf{a}^{D\dagger}(\omega)\mathbf{U}_{FN}(\omega)\boldsymbol{\mu}_{M-N-1}(\omega)\right]^{\dagger}s^{D}(k,\omega) \\
&\quad +\mathbf{P}^{\dagger}_{n}(\omega)^{\dagger}\mathbf{U}^{\dagger}_{FN}(\omega)\mathbf{n}(\omega) \\
&= \mathbf{P}^{\dagger}_{n}(\omega)^{\dagger}\mathbf{U}^{\dagger}_{FN}(\omega)\mathbf{n}(\omega)
\end{aligned}
\tag{3-18}
$$

where

$$
\mathbf{u}(k,\omega)=\left[u_{1}(k,\omega)\quad u_{2}(k,\omega)\quad \cdots \quad u_{M-N-1}(k,\omega)\right]
$$

Therefore, the BM can block the desired signal and interfering signal and obtain the stationary noise. The ANC can be used to eliminate the residual stationary noise. Recalling section 3.3, the ANC is the same one as described in (2-24). The architecture of GSC with nullforming is showed in Figure 4-6.

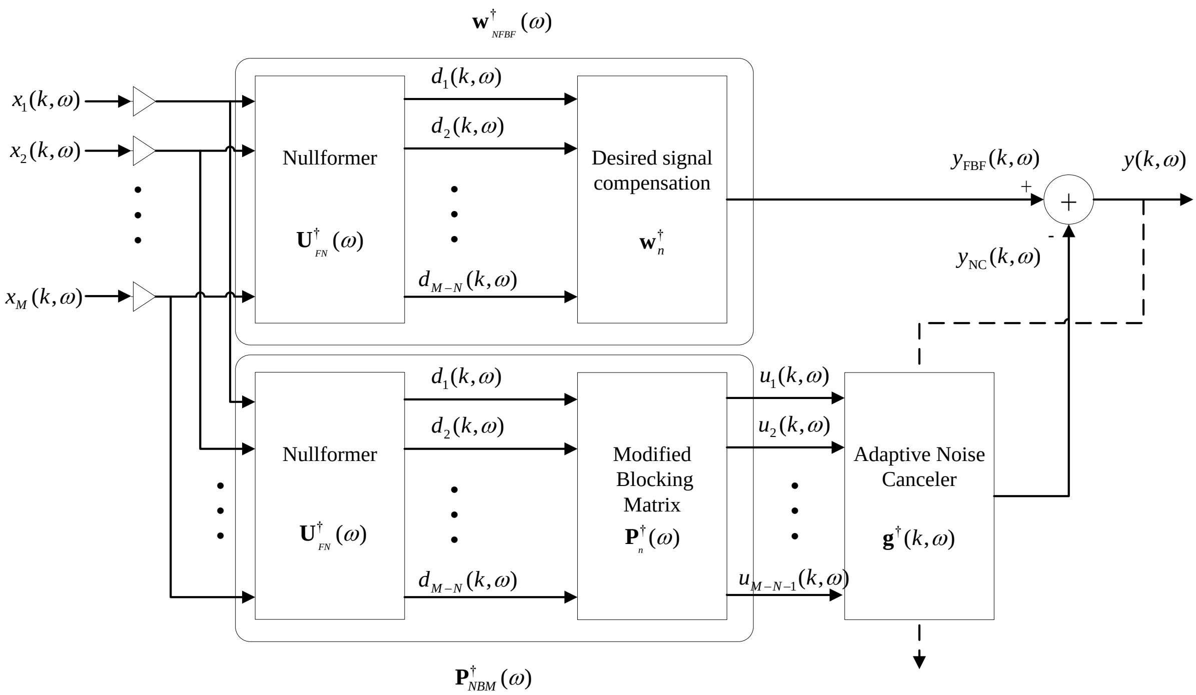


Figure 4-6 Generalized sidelobe canceler with nullforming

Chapter 5

Variable Nullforming Adaptive Filter

5.1 Introduction

In previous chapters, several methods to approach nullforming are introduced. These nullforming methods are fixed so they are not able to track the interfering sources. For example, the weighting of nullformer was previously set to one desired direction so the nullformer works well when interfering sources emit in the exact direction. When there are new interfering sources from other direction or the original interfering source change the direction, these kinds of fixed nullformer are unable to block the interfering sources.

In this chapter, a novel method to construct a variable nullformer is proposed. The nullforming algorithm could trace the change of sources. Then the algorithm applies the variable nullforming to generalized sidelobe canceler to obtain the reconstructed desired signal.

5.2 Variable Nullforming

5.2.1 Estimate Signal Subspace Using Order Recursive Least Square

Starting from the estimation of RTF vector, the estimation of RTF vector is from the output of blocking matrix [4]. Blocking matrix contains the bases spanning the null space of RTF vector. When there's only one sound source in action, rearranging the term (2-20) in chapter 3, the signal received from microphones are

$$x_m(k,\omega) = h_m^1(k,\omega)\, x_1(k,\omega) + u_m(k,\omega)\,. \tag{4-1}$$

Assume that the statistics of desired signal is stationary in each frame and the RIR changes slowly in a short period. Considering the cross PSD of *k*th frame

$$\Phi_{x_m x_1}^{(t)}(k,\omega) = h_m^1(k,\omega)\Phi_{x_1 x_1}^{(t)}(k,\omega) + \Phi_{u_m x_1}(k,\omega) \quad t = 1,\ldots,K \tag{4-2}$$

where K is the number of frames for estimating RTFs. $n_m(k,\omega)$, $m = 1,\ldots,M$ are assumed stationary, $s(k,\omega)$ and $u_m(k,\omega)$ are independent thus $\Phi_{u_m x_1}(k,\omega)$ is independent of the frame index *k*. [12] proposed a system identification method with nonstationary signal by applying least square (LS) estimation to the following equation

$$\begin{bmatrix} \Phi_{x_m x_1}^{(1)}(k,\omega) \\ \Phi_{x_m x_1}^{(2)}(k,\omega) \\ \vdots \\ \Phi_{x_m x_1}^{(K)}(k,\omega) \end{bmatrix} = \begin{bmatrix} 1 & \Phi_{x_1 x_1}^{(1)}(k,\omega) \\ 1 & \Phi_{x_1 x_1}^{(2)}(k,\omega) \\ \vdots & \vdots \\ 1 & \Phi_{x_1 x_1}^{(K)}(k,\omega) \end{bmatrix} \begin{bmatrix} \Phi_{u_m x_1}(k,\omega) \\ h_m^1(k,\omega) \end{bmatrix} + \begin{bmatrix} \epsilon_m^{(1)}(k,\omega) \\ \epsilon_m^{(2)}(k,\omega) \\ \vdots \\ \epsilon_m^{(K)}(k,\omega) \end{bmatrix}. \tag{4-3}$$

The RTFs are estimated when there's only one source. When there is more than one source in the environment, the method could not describe the RTF properly. The work in [10] proposed a method to estimate the blocking matrix when there are two sources emitting simultaneously in the environment. When number of sources increases, the number of reference microphones increases. Increasing number of reference microphones would increase the number of nullforming directions. Considering *N* sources emitting simultaneously, the linear equation would become

$$\begin{bmatrix} \Phi_{x_m x_1}^{(1)}(k,\omega) \\ \Phi_{x_m x_1}^{(2)}(k,\omega) \\ \vdots \\ \Phi_{x_m x_1}^{(K)}(k,\omega) \end{bmatrix} = \begin{bmatrix} 1 & \Phi_{x_1 x_1}^{(1)}(k,\omega) & \Phi_{x_2 x_1}^{(1)}(k,\omega) & \cdots & \Phi_{x_N x_1}^{(1)}(k,\omega) \\ 1 & \Phi_{x_1 x_1}^{(2)}(k,\omega) & \Phi_{x_2 x_1}^{(2)}(k,\omega) & \cdots & \Phi_{x_N x_1}^{(2)}(k,\omega) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \Phi_{x_1 x_1}^{(K)}(k,\omega) & \Phi_{x_2 x_1}^{(K)}(k,\omega) & \cdots & \Phi_{x_N x_1}^{(K)}(k,\omega) \end{bmatrix} \begin{bmatrix} \Phi_{u_m x_1}(k,\omega) \\ h_m^1(k,\omega) \\ h_m^2(k,\omega) \\ \vdots \\ h_m^N(k,\omega) \end{bmatrix} + \begin{bmatrix} \epsilon_m^{(1)}(k,\omega) \\ \epsilon_m^{(2)}(k,\omega) \\ \vdots \\ \epsilon_m^{(K)}(k,\omega) \end{bmatrix} \tag{4-4}$$

An excessive number of reference microphones are not needed when there are not many sources in action. It's difficult to know the number of active sources in each time so the signal subspace is estimated by order recursive least square estimation (ORLS). The method works by increasing the number of reference microphones when number of emitting sources grows. ORLS was originally used in line-fitting by increasing the order [15]. Rewrite (4-4) to the linear equation

$$\mathbf{m}_m(k,\omega)=\mathbf{M}_i(k,\omega)\boldsymbol{\theta}_i^m(k,\omega)+\mathbf{e}_m(k,\omega) \tag{4-5}$$

Where

$$\mathbf{m}_m(k,\omega)=\begin{bmatrix}\Phi_{x_m x_1}^{(1)}(k,\omega) & \Phi_{x_m x_1}^{(2)}(k,\omega) & \cdots & \Phi_{x_m x_1}^{(K)}(k,\omega)\end{bmatrix}^T$$

$$\mathbf{M}_i(k,\omega)=\begin{bmatrix}1 & \mathbf{m}_1(k,\omega) & \mathbf{m}_2(k,\omega) & \cdots & \mathbf{m}_{k-1}(k,\omega)\end{bmatrix}^T$$

$$\boldsymbol{\theta}_i^m(k,\omega)=\begin{bmatrix}\Phi_{u_m x_1}(k,\omega) & h_m^1(k,\omega) & h_m^2(k,\omega) & \cdots & h_m^{i-1}(k,\omega)\end{bmatrix}^T$$

$$\mathbf{e}_m(k,\omega)=\begin{bmatrix}\epsilon_m^{(1)}(k,\omega) & \epsilon_m^{(2)}(k,\omega) & \cdots & \epsilon_m^{(K)}(k,\omega)\end{bmatrix}^T$$

$\mathbf{m}_m(k,\omega)$ is PSD from the *m*th microphone, $\mathbf{M}_i(k,\omega)$ are the PSDs from the first to the (*k*-1)th microphone and $\boldsymbol{\theta}_k^m$ is the *m*th entry of signal subspace under *k*th order. $\mathbf{e}_m(k,\omega)$ is the estimation error. For *k*=1, The first entry of the estimator is $\boldsymbol{\theta}_1^m=\Phi_{u_m x_1}(k,\omega)$.

In each order, estimate the parameters by using LS estimation then the estimator would be

$$\begin{aligned}\hat{\boldsymbol{\theta}}_i^m(k,\omega)&=\left[\mathbf{M}_i^\dagger(k,\omega)\mathbf{M}_i(k,\omega)\right]^{-1}\mathbf{M}_i^\dagger(k,\omega)\mathbf{m}_m(k,\omega)\\ &\quad i=1,\ldots,N+1 \quad m=i,\ldots,M.\end{aligned} \tag{4-6}$$

To simplify the representation, the following derivation excludes the representation of parameters-(k,ω). The previous equation can be reformulated to recursive form by

[15]. The estimator $\hat{\boldsymbol{\theta}}_i^m$ updated in the $(i+1)$th order is

$$\hat{\boldsymbol{\theta}}_{i+1}^m = \begin{bmatrix} \hat{\boldsymbol{\theta}}_i^m - \dfrac{\left(\mathbf{M}_i^\dagger \mathbf{M}_i\right)^{-1} \mathbf{M}_i^\dagger \mathbf{m}_{i+1} \mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_m}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} \\ \dfrac{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_m}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} \end{bmatrix} \tag{4-7}$$

where $\mathbf{P}_i^\perp = \mathbf{I} - \mathbf{M}_i \left(\mathbf{M}_i^\dagger \mathbf{M}_i\right)^{-1} \mathbf{M}_i^\dagger$ is the projection matrix onto the subspace being orthogonal to that spanned by the columns of $\mathbf{M}_i$. The data under $(i+1)$th order is $\mathbf{M}_{i+1} = \left[\mathbf{M}_i \quad \mathbf{m}_{i+1}\right]$. To enhance the efficiency, the inverse operation is formulated to

$$\mathbf{T}_{i+1} = \begin{bmatrix} \mathbf{T}_i + \dfrac{\mathbf{T}_i \mathbf{M}_i^\dagger \mathbf{m}_{i+1} \mathbf{m}_{i+1}^\dagger \mathbf{M}_i \mathbf{T}_i}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} & -\dfrac{\mathbf{T}_i \mathbf{M}_i^\dagger \mathbf{m}_{i+1}}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} \\ -\dfrac{\mathbf{m}_{i+1}^\dagger \mathbf{M}_i \mathbf{T}_i}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} & \dfrac{1}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} \end{bmatrix} \tag{4-8}$$

where $\mathbf{T}_i = \left(\mathbf{M}_i^\dagger \mathbf{M}_i\right)^{-1}$ and $\mathbf{P}_i^\perp = \mathbf{I} - \mathbf{M}_i \mathbf{T}_i \mathbf{M}_i^\dagger$ is the projection matrix. The least square error of mth microphone under ith order is

$$J_{LS}^{m,i} = \left(\mathbf{m}_m - \mathbf{M}_i \hat{\boldsymbol{\theta}}_i^m\right)^\dagger \left(\mathbf{m}_m - \mathbf{M}_i \hat{\boldsymbol{\theta}}_i^m\right) \quad m = i, \ldots, M$$

the least square error could be updated by using the recursive form

$$J_{LS}^{m,i+1} = J_{LS}^{m,i} - \frac{\left(\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}\right)^2}{\mathbf{m}_{i+1}^\dagger \mathbf{P}_i^\perp \mathbf{m}_{i+1}} . \tag{4-9}$$

Figure 5-1 shows that the least square error shrinks gradually as the order of recursion increases. The recursion should be stopped when the error is less than a threshold.

Stop ORLS when $J_{LS}^{m,i} < J_{LS}^{threshold}$

Because the scale of least square error would depend on the present data, a robust stopping condition is set the threshold be the scale of the error from the first recursion.

Stop ORLS when $J_{LS}^{m,i} < \eta J_{LS}^{M,1}$

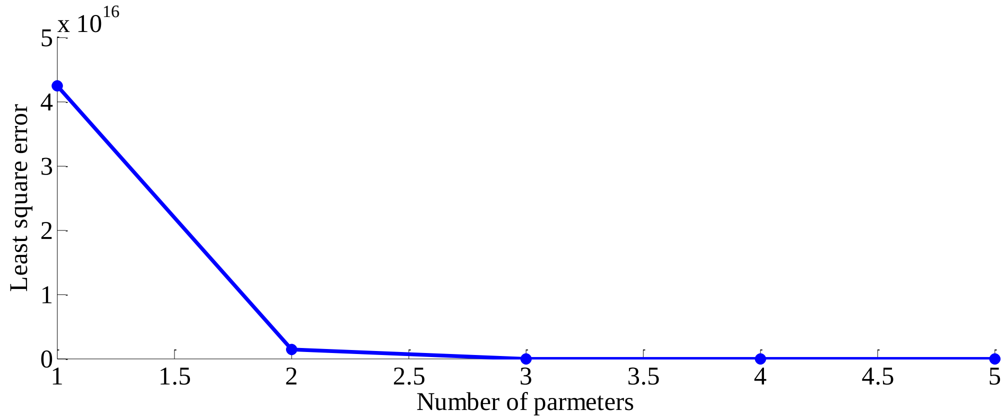


Figure 5-1　Effect of choosing time of iteration on least square error

The incoming data from small-size microphone array are highly correlated thus $\mathbf{M}_i^{\dagger}\mathbf{M}_i$ would become singular. The singularity would cause the estimation result being unstable. Tikhonov regularization [16], which is also called ridge regression is used to avoid the ill-condition. Tikhonov regularization reduces the singularity by imposing a penalty term, generally the penalty term is $\lambda\mathbf{I}$. This method may result in a biased estimation but stabilize the estimator. The nonnegative complexity parameter λ controls the amount of bias. Therefore, by the definition of Tikhonov regularization, the new form of linear equation in (4-6) becomes

$$\hat{\boldsymbol{\theta}}_i^m = \left(\mathbf{M}_i^{\dagger}\mathbf{M}_i + \lambda\mathbf{I}\right)^{-1}\mathbf{M}_i^{\dagger}\mathbf{m}_m . \tag{4-10}$$

Applying Tikhonov regularization to ORLS, the result is

$$\mathbf{P}_k^{\perp} = (\lambda+1)\mathbf{I} - \mathbf{M}_k\mathbf{T}_k\mathbf{M}_k^{\dagger}$$

$$\mathbf{T}_{i+1} = \begin{bmatrix} \mathbf{T}_i + \left(\dfrac{\mathbf{T}_i\mathbf{M}_i^{\dagger}\mathbf{m}_{i+1}\mathbf{m}_{i+1}^{\dagger}\mathbf{M}_i\mathbf{T}_i}{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_{i+1}}\right)\left(\dfrac{\mathbf{m}_{i+1}^{\dagger}\mathbf{m}_{i+1}}{\lambda+\mathbf{m}_{i+1}^{\dagger}\mathbf{m}_{i+1}}\right) & -\left(\dfrac{\mathbf{T}_i\mathbf{M}_i^{\dagger}\mathbf{m}_{i+1}}{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_{i+1}}\right)\left(\dfrac{\mathbf{m}_{i+1}^{\dagger}\mathbf{m}_{i+1}}{\lambda+\mathbf{m}_{i+1}^{\dagger}\mathbf{m}_{i+1}}\right) \\ -\left(\dfrac{\mathbf{m}_{i+1}^{\dagger}\mathbf{M}_i\mathbf{T}_i}{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_{i+1}}\right)\left(\dfrac{\mathbf{m}_{i+1}^{\dagger}\mathbf{m}_{i+1}}{\lambda+\mathbf{m}_{i+1}^{\dagger}\mathbf{m}_{i+1}}\right) & \dfrac{1}{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_{i+1}} \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}}_{i+1}^m = \begin{bmatrix} \hat{\theta}_i^m - \dfrac{\mathbf{T}_i\mathbf{M}_i^{\dagger}\mathbf{m}_{i+1}\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_m}{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_{i+1}} \\ \dfrac{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_m}{\mathbf{m}_{i+1}^{\dagger}\mathbf{P}_i^{\perp}\mathbf{m}_{i+1}} \end{bmatrix} \tag{4-11}$$

Assume the order of the estimation is N+1, the estimated RTFs in mth microphone would be

$$\hat{\boldsymbol{\theta}}_{N+1}^{m}=\begin{bmatrix}\hat{\Phi}_{u_m x_1}(k,\omega) & \hat{h}_m^1(k,\omega) & \hat{h}_m^2(k,\omega) & \cdots & \hat{h}_m^N(k,\omega)\end{bmatrix}^T \quad m=N+1,\ldots,M\,.$$

Rearranging the form to exclude $\hat{\Phi}_{u_m x_1}(k,\omega)$ and then the RTFs only vector would be obtained as

$$\hat{\mathbf{h}}_{N+1}^{m}=\begin{bmatrix}\hat{h}_m^1(k,\omega) & \hat{h}_m^2(k,\omega) & \cdots & \hat{h}_m^N(k,\omega)\end{bmatrix}^T \quad m=N+1,\ldots,M$$

Take the matrix form

$$\begin{aligned}\hat{\mathbf{H}}&=\begin{bmatrix}\hat{\mathbf{h}}_{N+1}^{N+1} & \hat{\mathbf{h}}_{N+1}^{N+2} & \cdots & \hat{\mathbf{h}}_{N+1}^{M}\end{bmatrix}\\&=\begin{bmatrix}\hat{h}_{N+1}^1(k,\omega) & \hat{h}_{N+2}^1(k,\omega) & \cdots & \hat{h}_M^1(k,\omega)\\ \hat{h}_{N+1}^2(k,\omega) & \hat{h}_{N+2}^2(k,\omega) & \cdots & \hat{h}_M^2(k,\omega)\\ \vdots & \vdots & & \vdots\\ \hat{h}_{N+1}^N(k,\omega) & \hat{h}_{N+2}^N(k,\omega) & \cdots & \hat{h}_M^N(k,\omega)\end{bmatrix}\end{aligned}$$

Because the number of microphones is limited, the number of estimated TFs in each basis reduces as the order increases. The matrix $\hat{\mathbf{H}}$ must be extended to make the length of bases being the same as the number of microphones. Therefore,

$$\begin{aligned}\mathbf{H}&=\begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_N\end{bmatrix}\\&=\begin{bmatrix}\mathbf{I} & \hat{\mathbf{H}}\end{bmatrix}^T\\&=\begin{bmatrix}1 & 0 & \cdots & 0 & \hat{h}_{N+1}^1(k,\omega) & \hat{h}_{N+2}^1(k,\omega) & \cdots & \hat{h}_M^1(k,\omega)\\ 0 & 1 & & \vdots & \hat{h}_{N+1}^2(k,\omega) & \hat{h}_{N+2}^2(k,\omega) & \cdots & \hat{h}_M^2(k,\omega)\\ \vdots & & \ddots & 0 & \vdots & \vdots & & \vdots\\ 0 & \cdots & 0 & 1 & \hat{h}_{N+1}^N(k,\omega) & \hat{h}_{N+2}^N(k,\omega) & \cdots & \hat{h}_M^N(k,\omega)\end{bmatrix}^T\end{aligned} \tag{4-12}$$

Where

$$\mathbf{h}_i=\begin{bmatrix}\underbrace{0 \ \cdots \ 0}_{i-1} & 1 & \underbrace{0 \ \cdots \ 0}_{N-i} & \hat{h}_{N+1}^i(k,\omega) & \hat{h}_{N+2}^i(k,\omega) & \cdots & \hat{h}_M^i(k,\omega)\end{bmatrix}^T$$

$$i=1,\ldots,N$$

are the estimated subspaces of present acting sources.

After the estimation, the number of estimated parameter shows the number of sources to describe the subspace under each frequency. The estimated bases are linear combination of subspaces from each source.

5.2.2 Estimate Subspace of Interfering Sources from Signal Subspace

The goal is to find out the null space of interfering sources. Therefore, after estimating the signal subspace from ORLS, the desired signal subspace must be excluded from the estimated subspace. Assume the subspaces of sources from each direction are obtained, the similarity between estimated subspace and each previously known subspace can be estimated by evaluating subspace distance by using the definition in [17]. The definition of subspace distance is

$$\begin{aligned}\mathcal{D}\left(\mathbf{H},\mathbf{b}_j\right) &= \sqrt{\max\left(\mathbf{H},\mathbf{b}_j\right)-\sum_{i=1}^{N}\mathbf{h}_i^{\dagger}\mathbf{b}_j} \\ &= \sqrt{N-\sum_{i=1}^{N}\mathbf{h}_i^{\dagger}\mathbf{b}_j} \qquad j=1,\ldots,P \end{aligned} \tag{4-13}$$

$\mathcal{D}(\ )$ is the definition of subspace distance and $\max(\ )$ is the maximum dimension between two subspaces. The subspace distance measurement works like doing inner product. The work in [17] proves that the measurement has the following properties.

1. *Nonnegativity*: $\mathcal{D}(U,V)\geq 0$, and $\mathcal{D}(U,V)=0$ if and only if $U=V$
2. *Symmetry*: $\mathcal{D}(U,V)=\mathcal{D}(V,U)$
3. *Upper boundness*: $\mathcal{D}(U,V)\leq\sqrt{\max(U,V)}$, and $\mathcal{D}(U,V)=\sqrt{\max(U,V)}$ if and only if $U\perp V$. That is, all bases in U are orthogonal to bases in V.

And $\mathbf{b}_j=\left[1\quad b_1^j(k,\omega)\quad\cdots\quad b_M^j(k,\omega)\right]^T$ is the previously estimated RTFs in P directions. Figure 5-2 shows the subspace distance of desired source and undesired source.

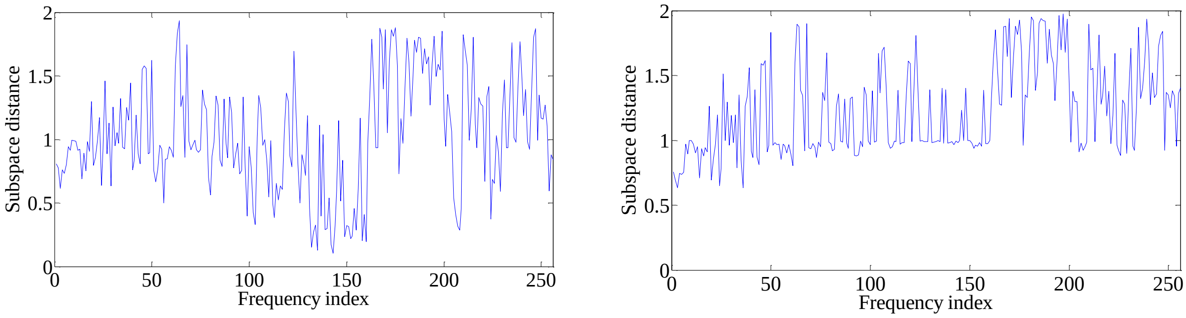


Figure 5-2 Subspace distance of desired signal (left) and interfering signal (right)

There are P previously estimated subspaces so P values of subspace distance would be obtained. Then by sorting these values and choose the least N RTF vectors

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_{j_1} & \mathbf{b}_{j_2} & \cdots & \mathbf{b}_{j_N} \end{bmatrix} \quad \{j_1, \ldots, j_N\} \in \{1, \ldots, P\}$$

Rearrange the index of these N vectors to $1, \ldots, N$

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_N \end{bmatrix} \tag{4-14}$$

where the columns of $\mathbf{B}$ are the most probably bases of present sources. These bases are linear combination of N RTF vectors. Rearranging the index $j_1, \ldots, j_N$ to $1, \ldots, N$, each estimated basis is

$$\mathbf{h}_i = \sum_{j=1}^{N} c_j^i \mathbf{b}_j; \quad i = 1, \ldots, N, \tag{4-15}$$

where c_j^i is the coefficient of each RTF vector. Assume the RTF of desired source is the dth vector in $\mathbf{B}$. Then eliminating the component of desired signal yields

$$\begin{aligned} \hat{\mathbf{o}}_i &= \mathbf{h}_i - c_d^i \mathbf{b}_d \\ &= \mathbf{h}_i - \underbrace{\left[\mathbf{B}\mathbf{e}\right]}_{\mathbf{b}_d} \underbrace{\left[\mathbf{e}^\dagger \left(\mathbf{B}^\dagger \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \mathbf{h}_i\right]}_{c_d^i} \\ &= \left[\mathbf{I} - \mathbf{B}\mathbf{e}\mathbf{e}^\dagger \left(\mathbf{B}^\dagger \mathbf{B}\right)^{-1} \mathbf{B}^\dagger\right] \mathbf{h}_i; \qquad i = 1, \ldots, N \end{aligned} \tag{4-16}$$

where $\mathbf{I}$ is a P by P identity matrix and $\mathbf{e} = \begin{bmatrix} \underbrace{0 \ \cdots \ 0}_{d-1} & 1 & \underbrace{0 \ \cdots \ 0}_{N-d} \end{bmatrix}^T$

Take the matrix form of $\hat{\mathbf{o}}_i$

$$\hat{\mathbf{O}} = \left[\mathbf{I} - \mathbf{B}\mathbf{e}\mathbf{e}^{\dagger}\left(\mathbf{B}^{\dagger}\mathbf{B}\right)^{-1}\mathbf{B}^{\dagger}\right]\mathbf{H}$$

Where

$$\hat{\mathbf{O}} = \left[\hat{\mathbf{o}}_1 \quad \hat{\mathbf{o}}_2 \quad \cdots \quad \hat{\mathbf{o}}_N\right].$$

The bases may contain dependent basis thus the orthogonal triangular decomposition (QRD) is used to exclude the dependent basis. So

$$\hat{\mathbf{O}}\mathbf{E} = \mathbf{Q}\mathbf{R} \tag{4-17}$$

where $\mathbf{Q}$ is a unitary matrix, $\mathbf{R}$ is a upper triangular matrix and $\mathbf{E}$ is a permutation matrix. The permutation matrix is used to make the diagonal terms of upper triangular matrix decreasing. The dependent bases correspond to the zero diagonal terms. Therefore, let

$$\mathbf{O} = \hat{\mathbf{O}}\mathbf{E} \tag{4-18}$$

$$\mathbf{O} = \left[\mathbf{o}_1 \quad \mathbf{o}_2 \quad \cdots \quad \mathbf{o}_D\right]$$

Where D is the number of independent bases and $\mathbf{O}$ contains the independent bases of interfering signal subspace.

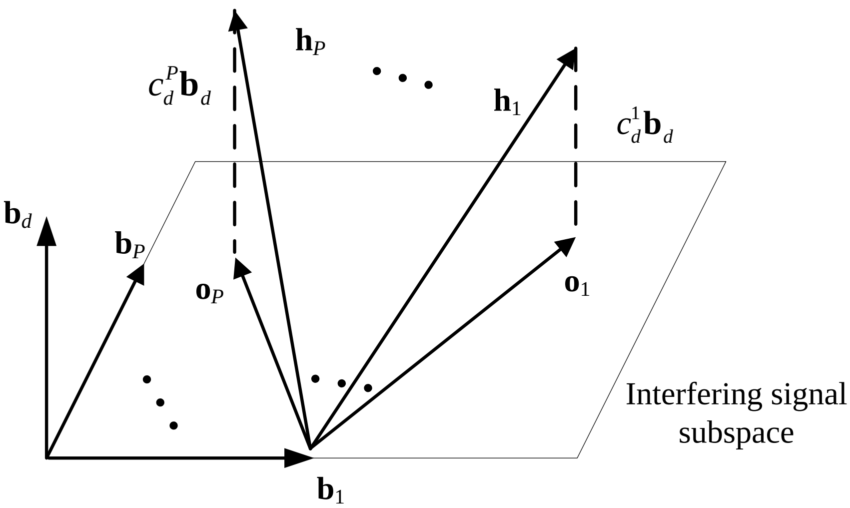


Figure 5-3 Relation of desired signal subspace and interfering signal subspace

For the accurate estimation of signal subspace, order recursive least square estimation needs a number of data. The alogorithm repeats the procedures to update the nullformer. When the nonstationary signal changes rapidly, the update frequency would be raised. Figure 5-4 shows update procedure of proposed algorithm. Assume the algorithm uses K frames to estimate the procedure of collecting data. The uptade procedure would be stopped when the number of collected data is K.

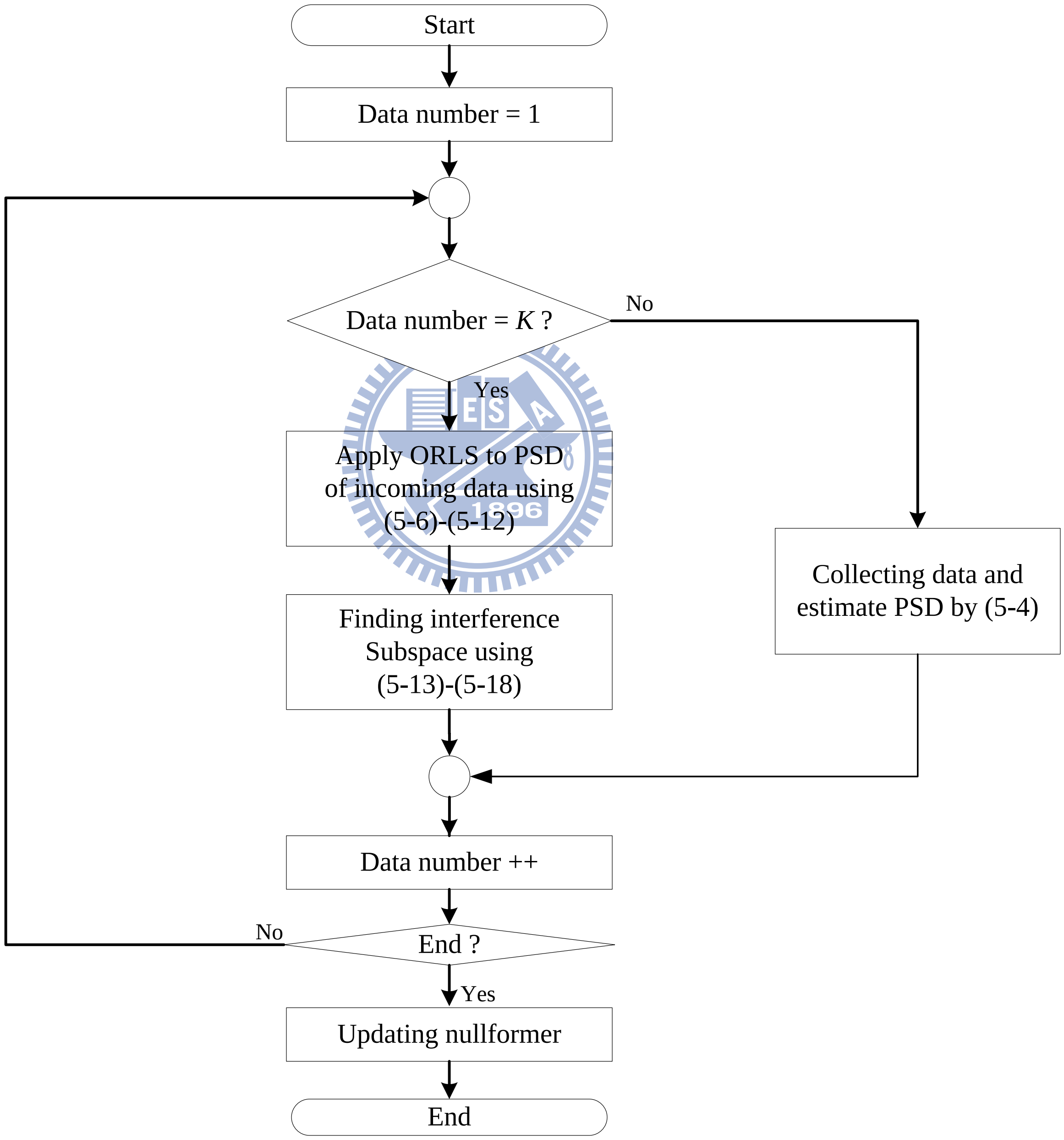


Figure 5-4 The updating flow of variable nullforming

The subspace of interfering signals could be used to find the null space of interfering sources. The null space of interfering signals can be found by using SVD as described in section 4.3,

$$\mathbf{U}_{VAR}(\omega)=\left[\mathbf{u}_{D+1}(\omega) \quad \mathbf{u}_{D+2}(\omega) \ \cdots \ \mathbf{u}_{M}(\omega)\right] \tag{4-19}$$

where $\mathbf{U}_{VAR}(\omega)$ is the null space of interfering signal subspace. That is

$$\mathbf{U}_{VAR}(\omega)\in\mathcal{N}\left(\mathbf{a}_1^{I\dagger}(\omega) \quad \mathbf{a}_2^{I\dagger}(\omega) \quad \cdots \quad \mathbf{a}_N^{I\dagger}(\omega)\right).$$

5.2.3 Directional Voice Activity Detection

In this section, a voice activity detection (VAD) method using the algorithm described above is proposed. Most of the VAD algorithms estimate input power to decide whether the desired source is in action. When there are new interfering sources from other directions with a similar magnitude of power comparing with the desired source, conventional VAD cannot distinguish this new source. The equation (4-15) shows that each frequency bin may contain the desired signal component. Therefore, the statistics of desired signal existence in each segment are compiled to decide the existence of desired signal. From Figure 5-2, the subspace distances in lower band are of no difference between desired source and interfering source thus choosing part band of the frequencies to compile the statistics would be proper.

Figure 5-5 shows the statistics of the target source. In Figure 5-5, the desired source and interfering signal overlap during 5 sec to 9 sec. The values of statistics between 5 sec to 9 sec are less than the desired signal-only period. When the number of sources increase, the threshold should be set lower to assure the accuracy of voice activity detection.

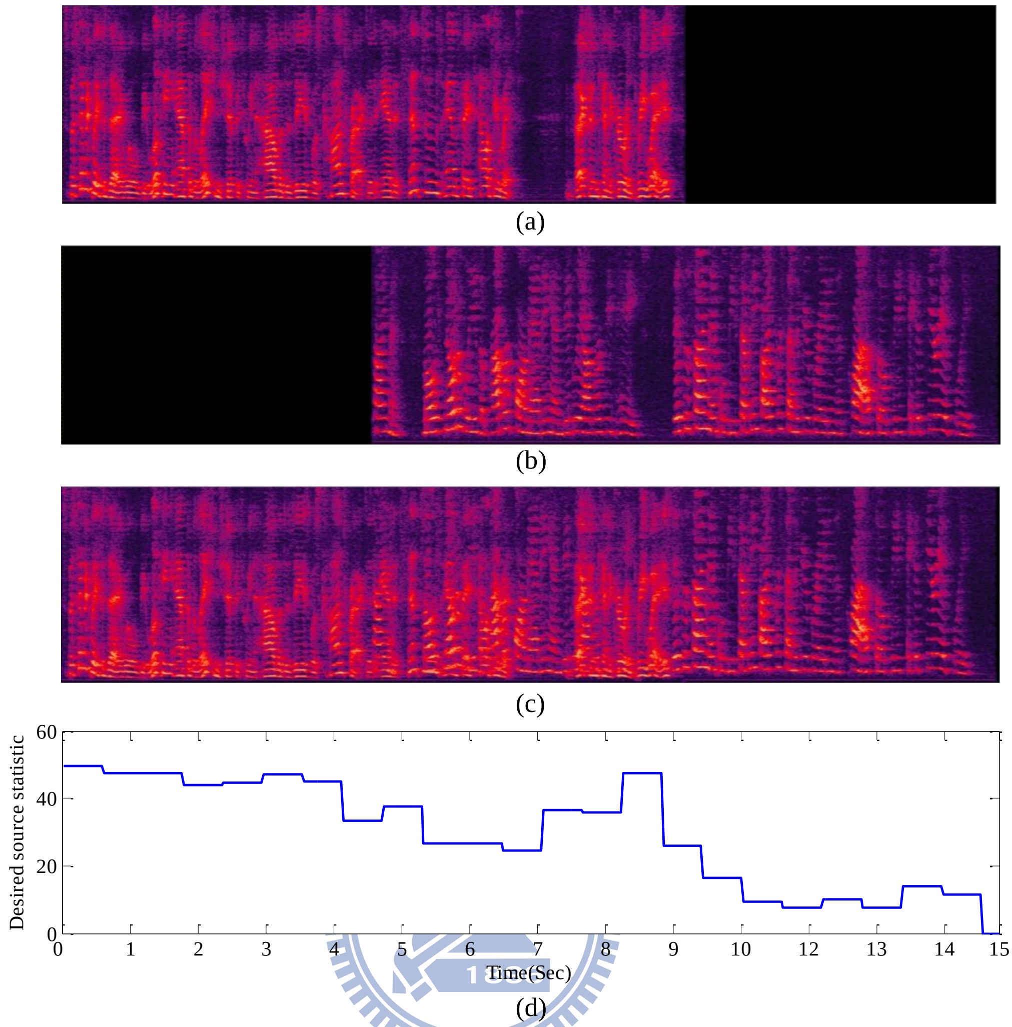


Figure 5-5 (a) The target source (b) Interfering source (c) Received signal (d) desired source statistics

The VAD method could be used to attain more accurate interfering signal subspace. In each frequency, the subspace distance may shows that the the desired signal is inactive in such frequency. But actually, the desired source is active. Therefore, the VAD could be used to show that whether the desired signal is active. The statistics would show if the estimated subspace contains the desired signal in such frequency bin. The VAD would decide whether the estimated subspace should be formulated to (4-16). If VAD shows the desired source is active, the component of desired signal should be excluded from the estimated subspace by (4-16).

5.3 Generalized Sidelobe Canceler with Variable Nullforming

The variable nullforming could be applied on GSC to eliminate the nonstationary interfering signal. The interfering signal subspace is obtained from the algorithm described in previous section. Therefore, the fixed beamformer and blocking matrix could be found by the same method described in section 4.5. Because the nullspace of present emitting signal is obtained, the method to find blocking matrix is simplified.

From (4-12), the estimated subspace $\mathbf{H}$ contains the subspace of present emitting signal so the nullspace of estimated subspace would block all present sources. Therefore, the blocking matrix would be the nullspace of $\mathbf{H}^{\dagger}$, that is,

$$\mathbf{P}_{\mathrm{BM}_{VAR}}(\omega) \in \mathcal{N}\left(\mathbf{a}^{D\dagger}(\omega) \quad \mathbf{a}_1^{I\dagger}(\omega) \quad \mathbf{a}_2^{I\dagger}(\omega) \quad \cdots \quad \mathbf{a}_N^{I\dagger}(\omega)\right),$$

where $\mathbf{P}_{\mathrm{BM}_{VAR}}(\omega)$ is the blocking matrix.

The fixed beamformer would block the interfering sources therefore applying the nullspace of interfering signal to GSC by the method described by (3-9)-(3-15) in section 4.5 then the fixed beamformer would be obtained. The fixed beamformer is

$$\mathbf{w}_{FBF_{VAR}}(\omega) = \mathbf{U}_{VAR}(\omega)\mathbf{w}_{\mathrm{v}}(\omega)$$

Where $\mathbf{w}_{\mathrm{v}}(\omega)$ is the weighting estimated by (3-12)

In Figure 5-4, every time the variable nullformer algorithm updates the FBF and BM, the ANC would be reset to initial value and then restart the adaption. Recalling (2-24), the ANC is same as the one in GSC.

Chapter 6

Experimental Result

6.1 Experimental Environment and Test Scenario

The proposed algorithm was tested by using the data recorded in a real environment with a uniform linear microphone array of eight un-calibrated microphones. The distance between each microphone is 6 cm. The size of the room is 5m × 4 m × 3 m and the microphone array was placed on a table at a distance of 0.5 m from the wall. The arrangement of microphone array and sound sources is showed in Figure 6-1.

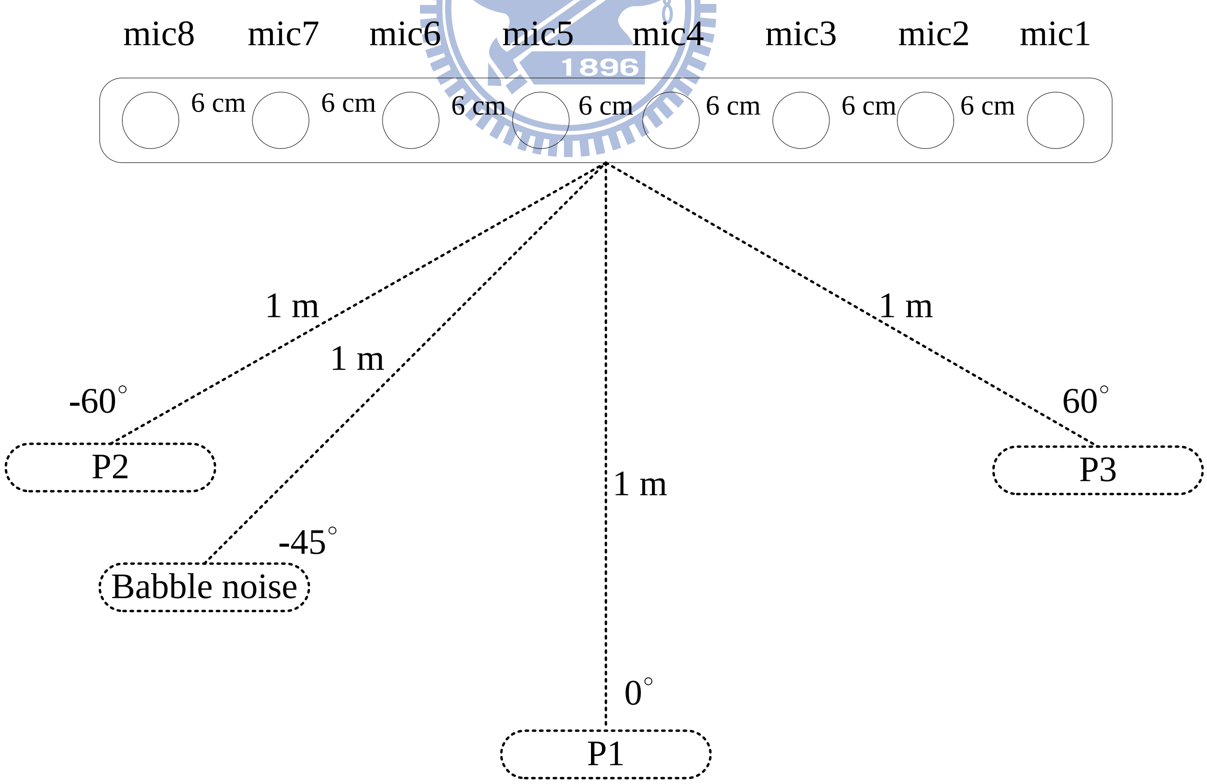


Figure 6-1 The location of microphone array and sources

The desired speech signal is located at 0° consists of sentences spoken by male. Two nonstationary interfering sources are located at 60° and -60°, one male voice from 60° and one frmale voice from -60°. One babble noise is located at -45°. The distance between these sources and the middle of microphone array is 1m. The training data from 7 directions are babble noise or sentences spoken by male and female, and the locations of these training sources are described in Table 1.

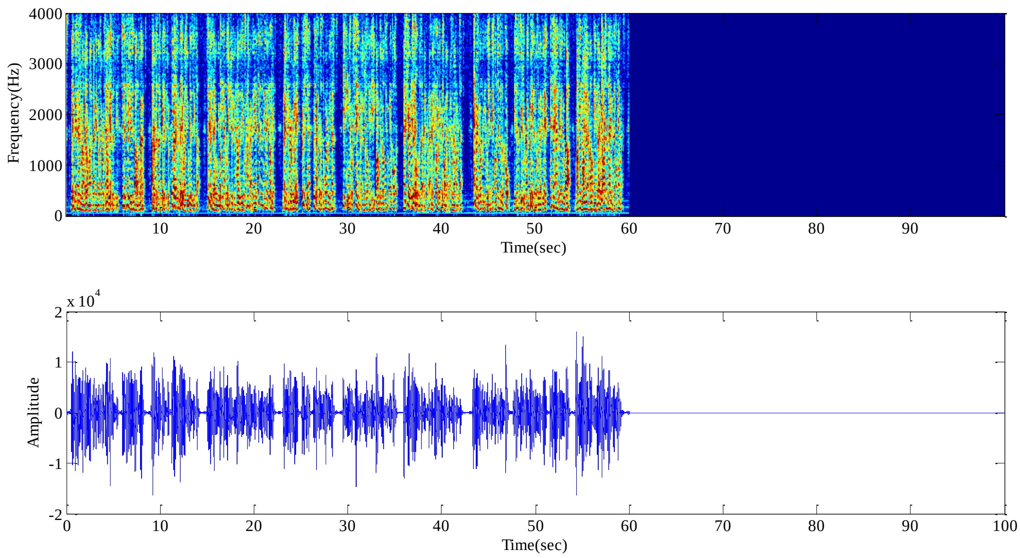
Degree	-60°	-45°	-30°	0°	30°	45°	60°
Type	Female1	Babble	Female1	Male1	Male2	Female2	Male2

Table 1 Sources for training data

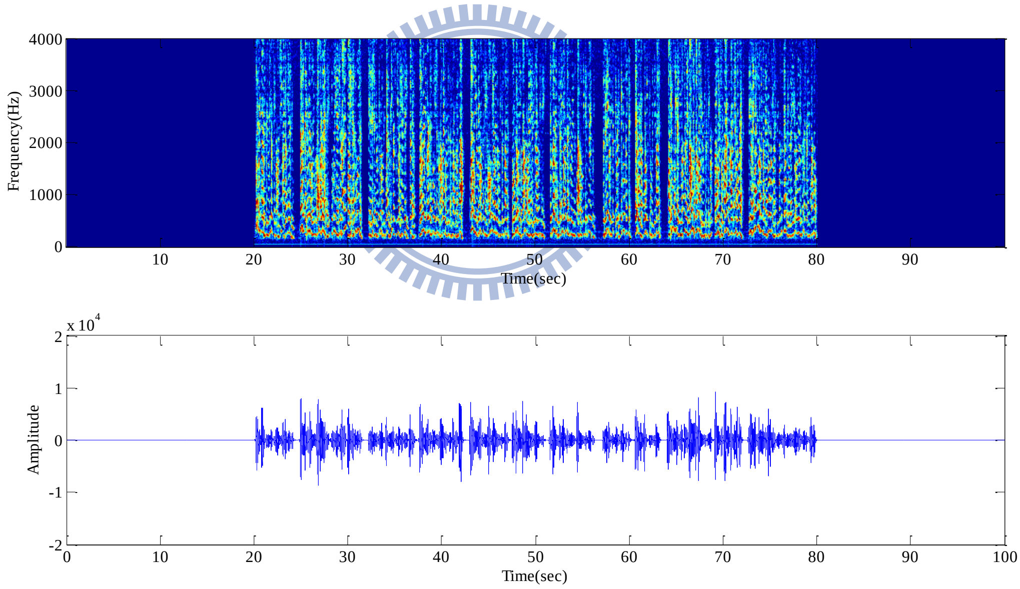
The sound sources are recorded independently for the purpose of estimating the segmental noise level (segNL) and log spectrum distortion (LSD). The data for testing purification is obtained by combining these sources with different time shifting. The testing data for estimating the performance of beamforming methods can be separated into five different scenarios for different time intervals. These five test scenarios are depicted in Table 2. The waveform and frequency spectrum of original desired source, interfering sources and noise are showed in Figure 6-2.

Scenario	Time(Sec)	Source	Interference	Stationary
C1	0~20	P1	None	-45°
C2	20~40	P1	P2	-45°
C3	40~60	P1	P2P3	-45°
C4	60~80	None	P2P3	-45°
C5	80~100	None	P3	-45°

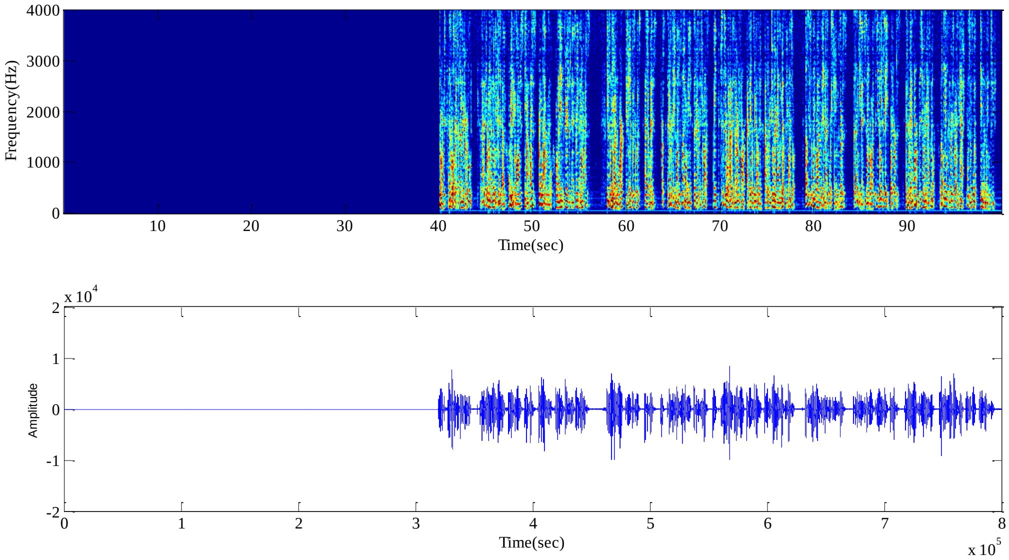
Table 2 Test scenario



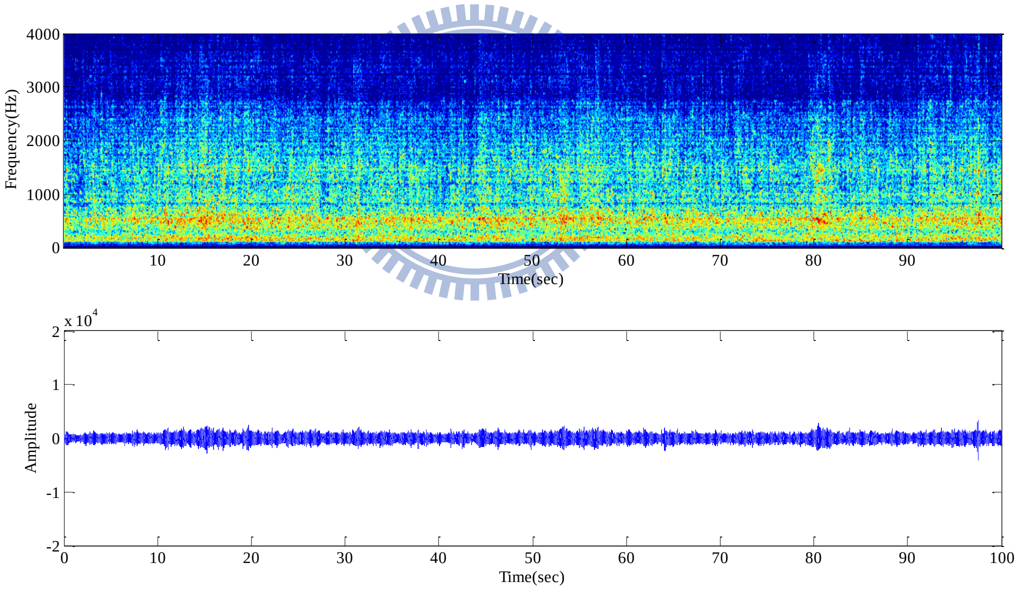
(a)Desired signal P1



(b)Interfering signal P1



(c)Interfering signal P2



(d)Babble noise

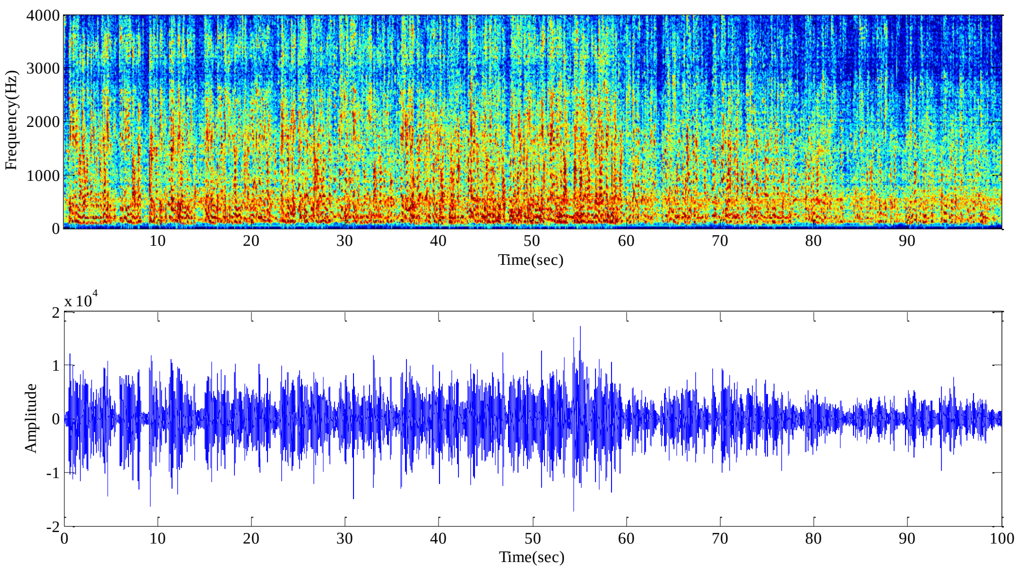
Figure 6-2 Frequency spectrum and waveform of sound sources

6.2 Experimental Results of GSC with Variable Nullforming

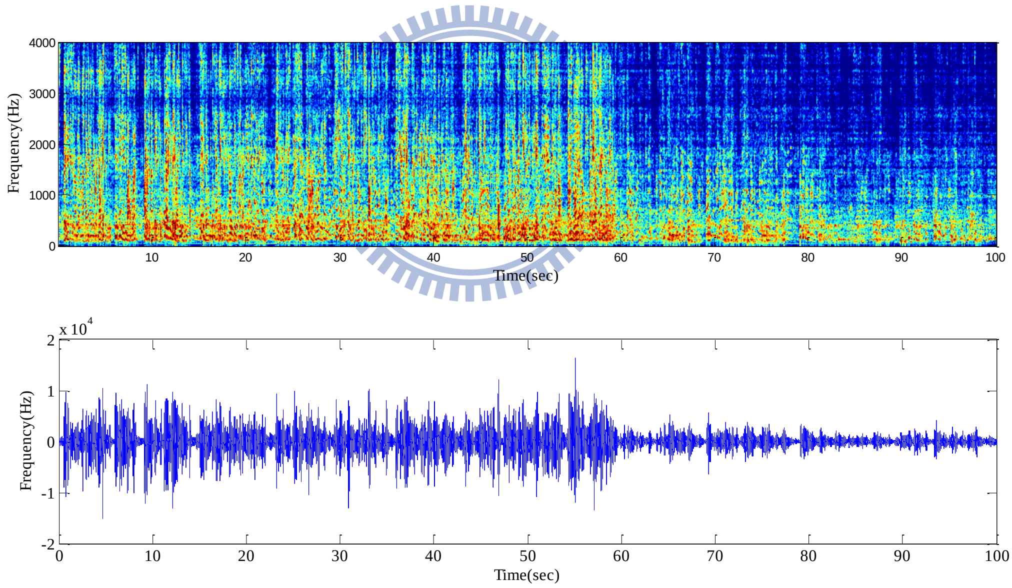
Before executing the variable nullforming algorithm, there are some parameters to be set. These parameters are described in Table 3. Figure 6-3 shows the comparison of received noisy signal and the signal enhanced by GSC with variable nullforming.

Parameter	Descrition	Value
f_s	Sampling rate	8000 Hz
N_{FFT}	FFT size	512 taps
FFTOVLP	The overlap size of FFT	256 taps
L_{RTF}	The training length of RTFs	20 frames
L_{SEG}	The length for estimating signal subspace	10 frames
T_{SEG}	Period of weighting update	2 sec
η	The ratio of least square error threshold for stopping ORLS	0.7
λ	The factor of Tikhonov regularization	10^{10}
T_{VAD}	The frequency band for VAD Statistics	1000 Hz~3000 Hz
TH_{VAD}	VAD threshold	50
μ	The NLMS step size	0.05

Table 3 The parameters used for GSC with variable nullforming



(a)Received signal from first microphone



(b)Purified signal by GSC with variable nullformer

Figure 6-3　Frequency spectrum and waveform of received signal and purified signal

6.3 Performance Estimation

The performance of adaptive filter would be affected by the existence of desired source. Therefore, the performance estimation would be tested under noncausal fixed beamformer. The adaptive filter would approach it's best performance after a while so the experiment copy the weighting of adaptive beamformer to fixed beamformer when the adaptive filter attach it's best performance in each scenario. Then the weightings of beamformer are used to filter the signal under such scenario. Figure 6-4 shows the test procedure of the experiment.

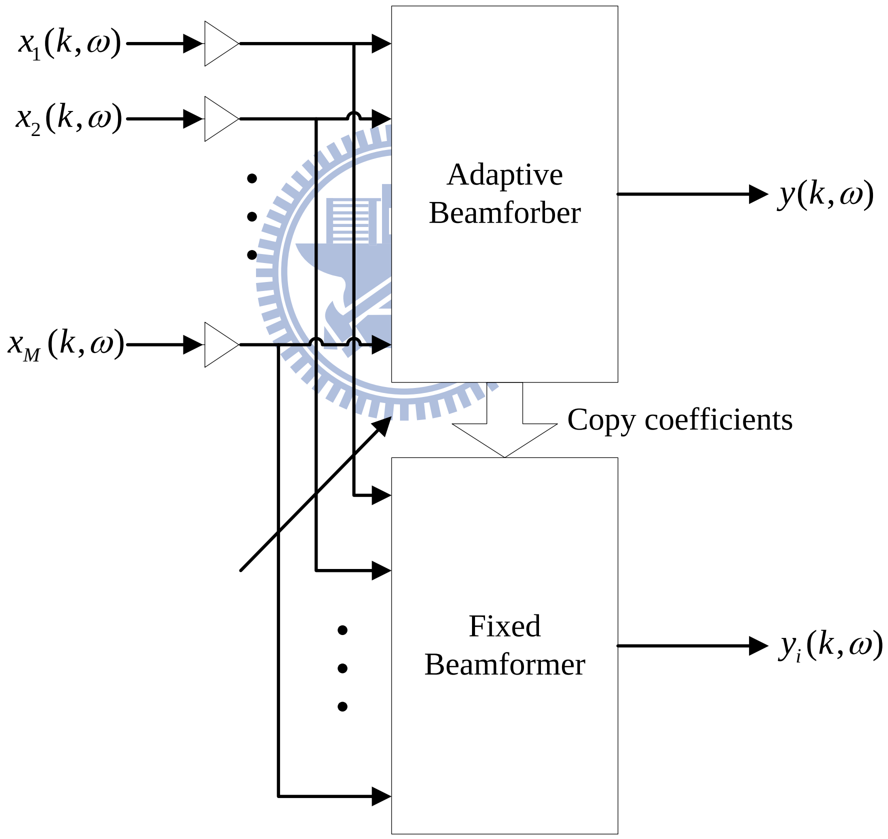


Figure 6-4 Test procedure for evaluating the performance

For the experiment, six speech enhancement methods are used for comparison. These methods are RSAB, GSC, RSAB with nullforming (NRSAB), GSC with one fixed nullforming to P2 (NGSC1), GSC with two fixed nullforming (NGSC2) and GSC with variable nullforming (VNGSC).

For RSAB, the weightings of beamformer are trained when P2 and P3 are emitting simultaneously in training phase. The training length is 40 frames. The experiment does not apply VAD to the RSAB method. Therefore, the beamformer is fixed in filtering phase. The NRSAB method pre-trained the interfering sources independently to obtain the RTFs of these interfering sources. Then use these RTFs to generate the null space of interfering sources. The test method of NRSAB is the same as used in RSAB.

The RTF of desired source is obtained for the GSC method. For NGSC1, the nullformer is the null space of P2 only; while for NGSC2, the nullformer is the null space of P2 and P3. Two different methods for the performance index are performed as follows.

1. *Segmental noise level (segNL):*

One quality measure for evaluating the performance of noise reduction is segmental noise level, which is defined as follows

$$Seg\ NL\ (\mathrm{dB}) = \frac{1}{L}\sum_{l=1}^{L}\left(10\log_{10}\left(\sum_{n=1}^{N}\hat{n}^2\left(n+lN\right)\right)\right). \tag{5-1}$$

SegNL evaluates the noise-only signal so the desired source must be segmented from the data and leaving the noise-only signal for testing. After filtering the noise-only signal, the output of the beamformer would be the reduced noise signal. The reduced noise signal is used to evaluate the segmental noise level.

The comparison of segmental noise level between different methods is summarized in Table 4. Channel 1 is the original interfering sources received from channel 1. Condition-C2 contains one desired source-P1 and one interfering source-P2. NGSC1 blocks the first interfering source P1 only so the nullity of NGSC1 would be larger than NGSC2. Comparing the noise level under condition C2, NGSC1 performs better than NGSC2. There's only one interfering source-P1 in the environment.

Therefore, the larger nullity would make the spatial filter sharper.

For condition C3, there are two interfering sources-P2 P3, one desired source- P1 and one ststionary noise in the environment. NGSC2 would perform better under such condition because the nullformer of NGSC2 has initially blocked two interfering sources from two directions. The nullformer of NGSC1 does not block another interfering source P3. Because the nullspace interfering sources is estimated and updated real-timely, VNGSC performs better than NGSC1 but worse than NGSC2 under C3.

There are two interfering sources and one ststionary noise under condition C4; one interfering source and one ststionary noise under condition C5. NGSC2 performs better than NGSC1 cause the interfering signal- P3 is blocked by the nullformer of NGSC2. Because the nullformer would enhance the signal from directions other than the desired direction, NGSC1 performs even worse than GSC under C5. The VNGSC performs better than NGSC1 and NGSC2 when there's no desired source.

NRSAB performs best under every condition because the existence of interfering sources is already known and the weightings of beamformer are well trained.

	C1	C2	C3	C4	C5
Channel 1	78.1432	80.8256	83.0169	83.9901	83.9080
RSAB	70.0936	75.9180	79.0773	79.0442	76.4045
GSC	72.6745	80.1337	81.9863	80.5743	77.1528
NRSAB	69.9616	75.2822	77.6048	77.5365	74.2547
NGSC1	72.8694	79.4318	81.6690	80.3390	77.7297
NGSC2	73.8355	79.4956	80.0839	79.1322	74.9946
VNGSC	75.4000	79.4519	80.5347	78.4874	74.8816

Table 4　　Segmental noise level of different speech enhancement methods

2. *Log spectrum distortion (LSD)*

The performance of noise reduction and distortion is a trade-off of a beamformer. The better noise reduction performance may cause more distortion. Therefore, another quality measure for evaluating the performance of distortion is log-spectral distortion, which is defined as

$$LSD = \frac{1}{L}\sum_{l=0}^{L-1}\left\{\frac{1}{\frac{K}{2}+1}\sum_{k=0}^{\frac{K}{2}}\left[\log_{10}\Psi S(k,l)-\log_{10}\Psi\hat{S}(k,l)\right]^{2}\right\}^{\frac{1}{2}} \quad (5\text{-}2)$$

where

$$\Psi S(k,l) \equiv \max\left\{\left|S(k,l)\right|^{2},\delta\right\}$$

is used to confine the log-spectrum dynamic rang about 50 dB i.e.

$$\delta = 10^{-50/10}\max_{k,l}\left\{\left|S(k,l)\right|^{2}\right\}.$$

The LSD compares the original desired signal with the enhanced signal. LSD of different methods are evaluated by using the original desired source recorded by microphone one and the enhanced signal by each method. The condition C4 and C5 are noise-only cases so the first three conditions are compared.

The comparison of LSD between different methods is summarized in Table 5. For C1, the environment exists desired source and a ststionary noise. The recorded desired source contains background noise and the signal enhancement methods would eliminate the noise. Therefore, the LSD of the enhancement method may be worse than the original received signal under highly SNR condition. Because GSC with nullformer eliminates more interfering signal, the LSD of NGSC1, NGSC2 and VNGSC are worse than GSC under C1.NGSC1 performs better than NGSC2 under C2, while NGSC2 performs better in C3. The VNGSC performs better than NGSC2 under C2 but worse in C3, which shows that the nullforming algorithm would be change according to the present acting interfering sources.

	C1	C2	C3
Channel1	0.3347	0.6301	0.7425
RSAB	0.4328	0.5522	0.5323
GSC	0.3782	0.5938	0.5781
NRSAB	0.3994	0.5173	0.5193
NGSC1	0.4019	0.5761	0.5780
NGSC2	0.4697	0.5965	0.5524
VNGSC1	0.4141	0.5731	0.5934

Table 5 Log spectrum distortion of different methods

Chapter 7

Conclusion and Future Study

The proposed variable nullforming algorithm would be used to eliminate the nonstationary interfering signal. The algorithm uses order recursive least square to approach the signal subspace. Because the existence of sources from different directions is unknown, the subspace distance would be used to find the similarity of present estimated subspace and pre-estimated subspace from different directions.

The subspace distance would be used to find the existence of emitting sources. Therefore, the algorithm could be used on multiple sources localization. Assume the RTFs from each direction are obtained, the statistics of subspace distance in each frequency bin would show the existence of sources.

The nullforming algorithm could be used as directional VAD. In the thesis, the threshold of VAD is not discussed. Figure 5-5 shows that the statistics of VAD would be affected by the number of sources. Therefore, a method for finding the entropy of received signal could be used on the factor of threshold.

There are several areas for improvement. The variable nullforming is updated from the past received data so if the existence of desired signal changed severely, the nullformer would make a distortion to the desired signal. If a perfect VAD applied on the algorithm, the algorithm doesn't need the previously estimated RTFs. The VAD could tell whether the desired source is in action alone or not in action. The interfering signal subspace would be updated when desired source is inactive and update the RTF of desired source when desired source is in action only.

References

1. Cox, H., R. Zeskind, and M. Owen, *Robust adaptive beamforming.* IEEE Transactions on Acoustics, Speech and Signal Processing, 1987. **35**(10): p. 1365-1376.
2. Dahl, M. and I. Claesson, *Acoustic noise and echo cancelling with microphone array.* Vehicular Technology, IEEE Transactions on, 1999. **48**(5): p. 1518-1526.
3. Hoshuyama, O., A. Sugiyama, and A. Hirano, *A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters.* Signal Processing, IEEE Transactions on, 1999. **47**(10): p. 2677-2684.
4. Gannot, S., D. Burshtein, and E. Weinstein, *Signal enhancement using beamforming and nonstationarity with applications to speech.* Signal Processing, IEEE Transactions on, 2001. **49**(8): p. 1614-1626.
5. Lo, Y., S. Lee, and Q. Lee, *Optimization of directivity and signal-to-noise ratio of an arbitrary antenna array.* Proceedings of the IEEE, 1966. **54**(8): p. 1033-1045.
6. Griffiths, L.J., *A simple adaptive algorithm for real-time processing in antenna arrays.* Proceedings of the IEEE, 1969. **57**(10): p. 1696-1704.
7. Frost, O.L., III, *An algorithm for linearly constrained adaptive array processing.* Proceedings of the IEEE, 1972. **60**(8): p. 926-935.
8. Griffiths, L. and C. Jim, *An alternative approach to linearly constrained adaptive beamforming.* Antennas and Propagation, IEEE Transactions on, 1982. **30**(1): p. 27-34.
9. Nordholm, S., I. Claesson, and P. Eriksson, *The broad-band Wiener solution for Griffiths-Jim beamformers.* Signal Processing, IEEE Transactions on, 1992. **40**(2): p. 474-478.
10. Reuven, G., S. Gannot, and I. Cohen, *Dual-Source Transfer-Function Generalized Sidelobe Canceller.* Audio, Speech, and Language Processing, IEEE Transactions on, 2008. **16**(4): p. 711-727.
11. Hu, J. and C. Cheng, *Frequency domain microphone array calibration and beamforming for automatic speech recognition.* IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2005. **88**(9): p. 2401-2411.
12. Shalvi, O. and E. Weinstein, *System identification using nonstationary signals.* Signal Processing, IEEE Transactions on, 1996. **44**(8): p. 2055-2063.
13. Elko, G.W. and P. Anh-Tho Nguyen. *A simple adaptive first-order differential*

microphone. in *Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on*. 1995.

14. Strang, G., *Introduction to linear algebra*. 2003: Wellesley Cambridge Pr.
15. Kay, S., *Fundamentals of statistical signal processing: estimation theory*. 1993.
16. Tikhonov, A. *On the stability of inverse problems*. 1943.
17. Wang, L., X. Wang, and J. Feng, *Subspace distance analysis with application to adaptive Bayesian algorithm for face recognition.* Pattern Recognition, 2006. **39**(3): p. 456-464.