國 立 交 通 大 學
電控工程研究所

碩 士 論 文

適應性波束形成器於寬頻語音純化
使用具二階約束之卡曼濾波器

Adaptive Beamformer for Speech Enhancement
Using Second-Order Constrained Kalman Filter

研 究 生： 李 哲 宇
指導教授： 胡 竹 生 博士

中 華 民 國 一 百 零 二 年 七 月

適應性波束形成器於寬頻語音純化
使用具二階約束之卡曼濾波器

Adaptive Beamformer for Speech Enhancement Using Second-Order Constrained Kalman Filter

研 究 生：李　哲　宇　　Student: Che-Yu Li

指導教授：胡　竹　生　博士　　Advisor: Jwu-Sheng Hu, Ph.D.

國立交通大學
電控工程研究所碩士班
碩 士 論 文

A Thesis
Submitted to Institute of Electrical and Control Engineering
College of Electrical Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of Master
in

Electrical and Control Engineering

July 2013

Hsinchu, Taiwan, Republic of China

中 華 民 國 一 百 零 二 年 七 月

適應性波束形成器於寬頻語音純化
使用具二階約束之卡曼濾波器

研究生：李　哲　宇　　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

當演算法中與環境及麥克風陣列的相關假設是不成立的時候，適應性波束形成方法的性能會有大幅的衰減。當預設目標訊號出現在訓練的樣本數中，即使在預設的指向向量和實際的指向向量有輕微的失配發生，陣列演算法的效果也會變得相當靈敏，相關的失配問題發生於變動的環境以及遠近效應、聲源散佈、與局部散射等。

本論文提出一套穩健型寬頻波束形成器，基於最佳化中在最糟情況下解決任意且受規範的目標訊號之指向向量失配的問題。利用麥克風陣列訊號的空間資訊，由最小變異無失真響應(MVDR)的波束形成器以空間濾波的方式針對聲源方向純化語音，同時壓抑來自於其它方向的雜訊。在實際應用的例子，波束形成器可以表示為狀態觀測器於二階展開的卡曼濾波器(SOE-KF)。然而窄頻波束形成器沒有考慮低頻訊號的空間指向性，目標聲源會因此受到壓抑而造成語音的失真。為了在指向向量失配的情況下輸出較高的輸出訊號與干擾加雜聲比，演算法根據不同頻帶下訊號的特性來選擇適當的指向向量的限制範圍。此外，當目標聲源不存在時，演算法對雜訊做追蹤，並透過零波束形成限制式於 SOE-KF 中以進一步提升語音純化的效果。本論文所提出的方法不僅改善了壓抑雜訊的效果並且提升了語音的品質。透過模擬和實驗驗證，本論文所提出的演算法有效地提升語音品質於吵雜以及有迴響的環境，並與其它已知的方法進行比較和分析。

Adaptive Beamformer for Speech Enhancement Using Second-Order Constrained Kalman Filter

Student: Che-Yu Li Advisor：Jwu-Sheng Hu, Ph.D.

Institute of Electrical and Control Engineering
National Chiao-Tang University

ABSTRACT

Adaptive beamforming methods are known to degrade significantly if some of underlying assumptions on the environment, sources, or sensor array are violated. The array performance may become sensitive even for a slight mismatch between the presumed and actual signal steering vectors when the desired signal is present. Such kind of mismatch occurs due to the dynamic environment, near-far mismatch, source spreading, and local scattering.

This paper presents a novel approach to design the robust broadband beamformer against arbitrary steering vector mismatch based on the optimization of worst-case performance. Using the spatial information from the microphone arrays, the desired source is enhanced while suppressing the directive noises via the robust minimum variance distortionless response (MVDR) beamformer. In practice, the beamformer is formulated into state-space observer form of the second-order extended (SOE) Kalman filter. However, the narrowband beamformer won't consider the signal directivity in the low frequency-bands and the desired source would be distorted. For maintaining higher OutputSINR under steering mismatch, the broadband selection of the steering vector bound is investigated. Furthermore, the noise tracking is utilized as null constraints into the SOE Kalman filter for speech enhancement when the source is absent. The proposed algorithm in this thesis not only improves the performance of

noise suppression but also enhances the speech equality. Simulations and experiments demonstrate the effectiveness of the proposed algorithm in a noisy and reverberant environment by comparing with existing algorithms.

Index Terms—wideband beamformer, constrained Kalman filter, robust MVDR beamformer, signal mismatch problem.

誌　謝

時光總是無形的流轉，很快地兩年的研究生涯要畫下句點了！過程中的點點滴滴都將成為往後日子中無可取代的回憶。首先要感謝我的指導教授胡竹生老師，謝謝老師給學生這個磨練自己的機會，在老師創新的思維的引導下，碰到困難的時候學生學習獨立思考與解決研究問題的方法，跳脫舊有的模式，訓練自己具備一位研究生該有的態度以及觀念，並對自己的研究內容負責。讓我以身為 XLAB 的成員感到驕傲，向老師獻上最誠摯的謝意。

另外感謝我的父母親家人們以及女朋友小璇，很多時候都能體諒我的無心抱怨以及支持我做最好的自己，讓我無太多的後顧之憂，這些日子以來有你們的陪伴和鼓勵，一直都是我努力向上的強力支柱。你們都是我最愛的人，感謝妳們相信我的想法與目標，我一定不會讓你們失望的，謝謝妳們。

再來要感謝 XLAB 的成員們：文武雙全的唐哥，感謝您始終耐心地教導我許多研究上的知識，祝您畢業後鵬圖大展；聲音組的棟樑耕維，祝您學業和事業一切順利;實驗室最強的男人Judo,大學專題中感謝有您領導一同奮戰的難忘時光，祝您愛情與學業兩兩得意;兩年來室友兼戰友的罐頭,研究生涯中一同切磋知識，謝謝您生活中大大小小不辭辛苦地幫忙，祝您愛情早日開花。還有率直的勁源學長、理論基礎很扎實的阿吉學長、總是帶來驚喜和歡笑的阿法學長、專注和對理想堅持的男哥、溫柔體貼的新好男人鳴哥、幽默風趣的廷哥、打球總是跳最高速度最快的翰哥、社交能力一流的小山東，托你的福氣可以認識許多朋友、3C 產品達人大夢，引領我進入蘋果的世界、毅力十足的阿文，認真的你最帥、研究專注與謹慎的鳴遠、及總能跟上日韓潮流的期元、以及已經畢業的學長姐們和學弟們，很榮幸加入 XLAB 這個大家庭，感恩大家對我的指導與包容。

最後要感謝交通大學，從大學到碩士生涯伴隨著我的回憶，提供許多學子進修的管道以及資源，飲水思源的傳承是我往後受用無窮的做人本質，謝謝交大，感謝在這裡的一切美好回憶。

CONTENT

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Motivation and Objective

Speech enhancement in a noisy environment is an important research issue for speech signal processing. It will cause a great impact on both respects of voice recognition and communication. Although the hearing of human beings is able to recognize desired speeches even under noisy environment, it is still regarded as a difficult task for computers or machines.

The common sensor for receiving sound waves is the microphone. Single microphone can collect spectral information but not the spatial information. The advantage of microphone arrays is applied to catch not only spectral information but also spatial information among the sound waves. Adaptive spatial filter, which is called beamformer, is one of the most effective methods and are extensively studied for hands-free speech communication or recognition among several existing microphone-array-based speech enhancement algorithms in recent years.

The background noise and reverberation from undirected diffused noises or directed interferences are the most dominant reasons for the degradation of signal quality. The noises and reverberation level will determine the distortion level of the desired signal. Although the methods of multichannel speech enhancement are used to reduce the effect of noise and reverberation, they do not perform well in real practice when the pre-assumption of adaptive spectral filter violates the environment conditions. This provides the motivation of this thesis to study and propose innovative methods to handle both interference suppression and desired source mismatch problems, which is useful in a scenario like a real life conference in a meeting room or communication in the living room, where the equality of sound is deteriorated by human beings' talking noise and reverberation in the space of the room.

1.2 Literature Review

Microphone arrays can be used to achieve the effect of spatial filtering, which is generally called Beamformer (BF) [1]. The beamformers can be categorized in two types, fixed beamformers and adaptive beamformers. Although the implementation costs of fixed beamformers are often lower than the adaptive beamformers, the beamforming effect is not robust enough due to there is no update mechanism in the algorithm.

Fixed beamformers include delay-and-sum beamformer (DAS) [2], constant directivity beamformer (CDB) [3] and fixed superdirective beamformers [4]. The fixed weights are utilized to form a spatial filter according to the pre-known spatial information. The DAS is the simplest structure in beamformer. It compensates to the relative time delay between distinct microphone signals and then sums the steered signals with a fixed weighting in every channel to form a single output. The CDB maintains the spatial response over a wide frequency-band; and the fixed superdirective beamformer keeps desired source distortionless at a pre-defined direction while attempting to suppress the noises from the other directions. These approaches assume the desired source and interferences are at pre-known location in stationary environment. Hence, these algorithms are sensitive to steering mismatches, which degrade the capacity of noise reduction and result to desired source distortion and signal self-cancellation.

Instead of using fixed beamformer, an adaptive beamformer can generate a beam response to the desired source direction and null at undesired signals to suppress the noises and interferences. Many adaptive beamformer techniques were extensively studied in the last three decade. The linearly constrained minimum variance (LCMV) beamformer was proposed in [5] to minimize the array output power under a look direction constraint. A special case similar to LCMV is the minimum variance

distortionless response (MVDR) proposed by Capon in [6]. Another popular technique is the generalized sidelobe canceller (GSC) algorithm which essentially transforms the LCMV constrained minimization problem into an unconstrained one [7].

The formulation of MVDR is implemented with Kalman filter using the state-space observer form. Owing to the undesirable mismatch between the actual desired signal steering vector and the presumed one in single steered constraint, various adaptive beamformers were proposed to improve the performance. The signal mismatches can be induced by signal point error [8], imperfect array calibration [9], or channel effect. In the presence of these effects, an adaptive beamformer suppresses the desired signal instead of maintaining distortionless response. Such phenomenon is commonly referred as signal self-nulling [10]. To strengthen the robustness against steering vector error, various methods are investigated [17], [19]. The Kalman filter can also be substituted by second-order extended Kalman filter [18], [20] and constrained Kalman filter [12], [13] to improve its robustness and reducing non-linearity against mismatch problems.

Among adaptive beamformers which are realized by Kalman filter, the usage of constraint projected method and steering vector bound regulation in wideband concept is a solution to the signal mismatch problem. The relative theory can be found in [17], [18], [21], [22].

1.3 Thesis Scope and Contribution

The contribution of this thesis is to propose and implement an innovative algorithm against signal mismatch problem for speech enhancement. The scope of thesis can be divided to two parts. The first part is to formulate a constrained adaptive beamformer considering the multiple arrays directivity and spatial coherence of spatial filtering. The second part is to handle the beamforming constraints given by the information of voice activity detection to achieve better performance of speech enhancement.

In the first part, the formulation using MVDR structure with signal mismatch problems is given. To obtain the solution, the nonlinear second-order extended Kalman filter is applied to deal with inequality nonlinear constraints as well as constraining the state prediction. In the optimal minimum mean-square error (MMSE) algorithm, the selection of parameters is to avoid suppression of the desired signal component (signal self-nulling) in broadband sense. Each selection has different result in different signal mismatch situation. The principle of selection is investigated and explained.

In the second part, the noise tracking can be utilized as null constraints for further enhancement when the desired source is not present. We incorporate the equality-constraints (Hard-Constraints) into the Kalman filter by projecting the updated state estimate onto the constrained region. The robustness of performance against signal mismatch for directive noise and dereveberation is achieved by choosing appropriate parameters in different conditions (ex: microphone arrays number, mismatch angle). In particular, the information given by the voice activity detector can also be reused to select appropriate parameters in beamforming. The performance of the algorithm is discussed and explained.

1.4 Outline of Thesis

The remainder of this thesis is organized as follows.

Chapter 2: The beamformers of adaptive spatial filtering which are based on the robust beamforming technique Minimum Variance Distortionless Response (MVDR) are introduced. The ideal steered linear inequality constraint is incorporated into the steepest gradient method and state space formulation to implement MVDR. By comparisons, the pros and cons construct the foundation of proposed algorithm.

Chapter 3: The detailed formulation of second-order extended Kalman filter with nonlinear inequality constraints is presented. It includes the solutions to signal mismatch problem and beamformer null constraint for suppression of interferences, given the information of voice activity detection (VAD). The technique of choosing the appropriate parameter in wideband beamforming and its effect are also discussed. Finally, the overall flowchart and architecture are illustrated and explained.

Chapter 4: The results of simulation and experiment are shown. It contains comparison between adaptive spatial beamformers and the capability of beamforming against signal mismatch problem in Room Impulse Response (RIR) and real room respectively. Some objective indices are calculated to compare the performance between proposed algorithm and other existing algorithms.

Chapter 5: The conclusion of this thesis and some issue that is discussed for future studies in this chapter.

Chapter 2. Adaptive Beamformer

2.1 Array Signal Processing

In the traditional digital signal processing, Most of the techniques focus on the processing skills in the time and frequency domain. Generally, sampling the continuous signals first, and then converting to the frequency domain to analyze the signals or discriminate different components by passing the filter.

Multiple microphones are in a fixed shape to receive the signals which are transported in the spatial. As a result of the microphones in the different position, microphones will retrieve different energy change and time delay from the same signal in the same source. Then the processing analysis for exacting a desired source out of spatial distinct sources from multiple microphones is called microphone array signal processing. The field of the research can be classified to two categories:

First：Focus on the number of signals or the spatial direction, generally called (Direction of Arrivals Estimation, DOA).

Second：Using the spatial relationship of signals can find out different gain of different direction of signals to have the spatial filtering effect. Generally, the way to separate the different direction signals is beamformer, and is also one kind of spatial filter.

Assign a microphone array and a reference point (generally one microphone of the array), array manifold vector defines the relationship between the source signal retrieved from microphones and the reference point of time. In the thesis of beamformer, array manifold vector is used to compensate the input signal phase delay between different microphones. And, in the thesis of Direction of Arrivals Estimation,

many methods find out the source direction by comparing the similarity of array manifold vector and the eigenspace of source signals (Eigenstructure Method DOA). The array processing converts the raw data into frequency domain to compute by Short-Time Fourier Transform (STFT). In frequency domain, the time delay will become to the phase delay. Thus, the resolution of computation for source angle improves effectively in frequency domain.

Usually using the different array arrangement in different situation, following is the common (Uniform Linear Array, ULA). The structure is as Fig. 1. Assume the source signal is the (Far field plane wave), s(t) is initial source, n(t) is noise, and the output of M microphones can represented as the following vector notation：

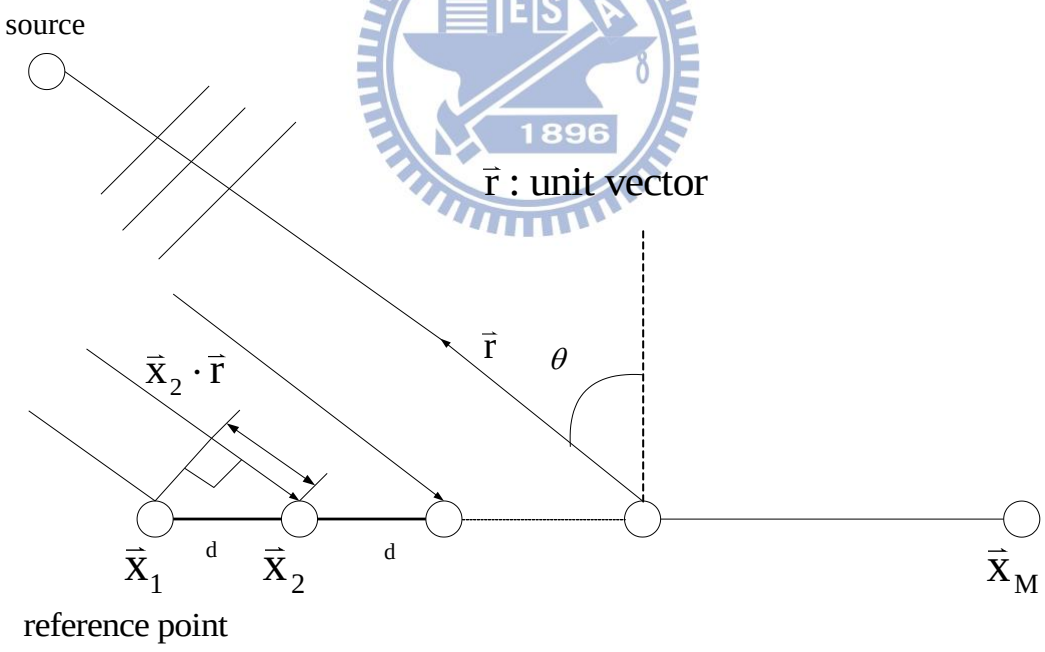


Fig. 1 Uniform Linear Array structure (ULA).

This Thesis is based on the Uniform Linear Array structure.

$$x(t)=\begin{bmatrix} x_1(t) \\ \vdots \\ x_2(t) \end{bmatrix}=\begin{bmatrix} s(t)e^{jw_c\frac{\vec{x_1}\cdot\vec{r}}{c}} \\ \vdots \\ s(t)e^{jw_c\frac{\vec{x_M}\cdot\vec{r}}{c}} \end{bmatrix}+\begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix}$$
$$=\begin{bmatrix} e^{jk_c\vec{x_1}\cdot\vec{r}} \\ \vdots \\ e^{jk_c\vec{x_M}\cdot\vec{r}} \end{bmatrix}s(t)+\begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix}=a(\vec{r})s(t)+n(t) \quad (2.1)$$

$k_c=\frac{w_c}{c}=\frac{2\pi}{\lambda_c}$, k_c is called wavenumber, and w_c is the wavelength, c is the wavespeed.

$a(\vec{r})$ is the array manifold vector, including the time relationship of signals transported to the microphones, which can simplified as following :

$$a^T(\theta)=\begin{bmatrix} 1 & e^{jk_cd\sin\theta} & \cdots & e^{jk_c(\mathrm{M}-1)\sin\theta} \end{bmatrix} \quad (2.2)$$

In addition to the assumption to the source model, the equality of the microphones has to be checked to a certain extent. Theoretically, we will assume the characteristics between microphones are identical completely (No difference in gain or phase) to ensure the signal from desired source to the microphone arrays is just a relationship of time delay. Once there is difference between arrays, the estimation of spatial equivalent relationship will be influenced considerably (ex, Direction of Achieving Angle). Therefore, we have to confirm the gain between microphones is limited to a range. When the distance from desired source to microphone arrays is smaller, we have to consider the near-field assumption. Opposite to the near-field assumption is the far-field assumption which is used in array signal processing. Moreover, the shadowing effect of arrays disposal and directivity of array itself will both influence the spatial relationship that arrays received.

Fig. 2 is the uniform circular array arrangement. Its source angle searching ability has 2-D dimensions. Setting the center of the Fig. 2 as reference point, the array manifold vector can deduced to the following vector notation：

$$a^T(\theta)=\left[1 \quad e^{jk_c*R\sin\phi\cos\theta} \quad e^{jk_c*R\sin\phi\cos(\theta-\frac{2\pi}{M})} \quad e^{jk_c*R\sin\phi\cos(\theta-\frac{2(M-1)\pi}{M})}\right] \tag{2.3}$$

R：Center of the circle　　M：Number of arrays

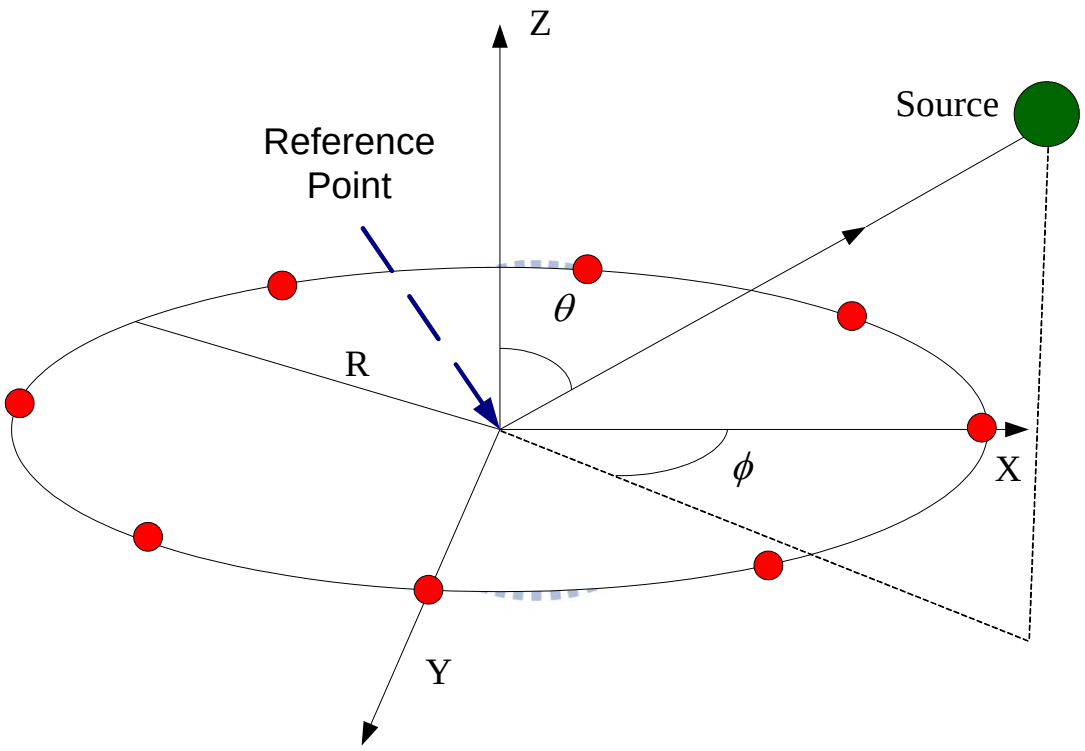


Fig. 2 Uniform Circular Array structure (UCA).

2.2 Beamformer under MVDR Structure

The minimum variance distortionless response (MVDR) beamformer, also known as Capon beamformer [6], minimizes the output power of the beamformer under a single linear constraint on the response of the array towards the desired signal.

Consider the conventional signal model in which an M-element microphone array captures a convolved desired signal (speech source) in some noise field. The received signals are expressed as [3]

$$x_m(k) = a_m * s(k) + v_m(k) \qquad \text{m} = 1, 2, \ldots \tag{2.4}$$

where a_m is the impulse response from the unknown (desired) source $s(k)$ to the m_{th} microphone, * stands for convolution, and $v_m(k)$ is the noise at the microphone m. The signals $s(k)$ and $v_m(k)$ are assumed as uncorrelated and zero mean. In the frequency domain, (2.4) can be written as

$$X_m(jw) = A_m(jw) * S(jw) + V_m(jw) \qquad \text{m} = 1, 2, \ldots \tag{2.5}$$

where $A_m(jw)$, $S(jw)$, $X_m(jw)$, $V_m(jw)$ are the discrete-time Fourier transforms (DTFTs) of $a_m(k)$, $s(k)$, $x_m(k)$, $v_m(k)$, respectively, at angular frequency w $(-\pi < w \le \pi)$ and j is the imaginary unit ($j^2 = -1$).

These M microphone signals in the frequency domain are summarized in a vector notation as

$$\mathbf{X}(jw) = \mathbf{A}(jw)S(jw) + \mathbf{V}(jw) \tag{2.6}$$

where

$$\mathbf{X}(jw) = [X_1(jw)\ X_2(jw)\ \cdots X_M(jw)]^T$$
$$\mathbf{A}(jw) = [A_1(jw)\ A_2(jw)\ \cdots A_M(jw)]^T$$
$$\mathbf{V}(jw) = [V_1(jw)\ V_2(jw)\ \cdots V_M(jw)]^T$$

and superscript ^{T} denotes transpose of a vector or a matrix.

Consider finding a weighting vector w_{MV} which satisfies the look direction constraint

$$w_{MV}^{H}(jw)a(\theta_s, jw)=1 \tag{2.7}$$

while attempting to minimize beamformer output power

$$E\{|Y(jw)|^2\}=E\{|w_{MV}^{H}(jw)X(jw)|^2\}\equiv w_{MV}^{H}(jw)R_{XX}(jw)w_{MV}(jw) \tag{2.8}$$

in order to suppress undesired interference from $\theta\neq\theta_s$ and noise. $Y(jw)$ is the beamformer output given by

$$\mathbf{Y}(jw)=w_{MV}^{H}(jw)X(jw). \tag{2.9}$$

$a(\theta_s, jw)$ is the array manifold vector that points to the source direction.

With the consideration above, the following constrained optimization problem can be formulated:

$$\min w_{MV}^{H}(jw)R_{XX}(jw)w_{MV}(jw) \text{ subject to } w_{MV}^{H}(jw)a(\theta_s, jw)=1 \tag{2.10}$$

To solve this problem, the Lagrange Multiplier is incorporated.

$$\begin{cases}\nabla_{W_{MV}(jw)}w_{MV}^{H}(jw)R_{XX}(jw)w_{MV}(jw)-\lambda\nabla_{W_{MV}(jw)}[w_{MV}^{H}(jw)a(\theta_s, jw)-1]=0\\ w_{MV}^{H}(jw)a(\theta_s, jw)=1\end{cases} \tag{2.11}$$

(2.8) can be reduced to

$$\begin{cases}R_{XX}(jw)w_{MV}(jw)=\lambda a(\theta_s, jw)\\ w_{MV}^{H}(jw)a(\theta_s, jw)=1\end{cases} \tag{2.12}$$

Assuming R_{XX} is nonsingular. Then

$$w_{MV}(jw)=\frac{R_{XX}^{-1}(jw)a(\theta_s, jw)}{a^{H}(\theta_s, jw)R_{XX}^{-1}(jw)a(\theta_s, jw)} \tag{2.13}$$

which is the optimal solution to MVDR problem proposed by Capon [6].

2.3 Beamformer Using Equality State Constrained Kalman Filter

Kalman filtering [26] is a method to make real-time prediction for systems with some known dynamics in control theory. Traditionally, problems requiring Kalman Filtering have been complex and nonlinear. Many advances have been made in the direction of dealing with nonlinearities (e.g., Extended Kalman Filter, Unscented Kalman Filter). These problems also tend to have inherent state space equality constraints and state space inequality constraints. In this thesis, the use of constrained Kalman filter in signal processing is more concerned. A constraint on the microphone array response along the look direction is added to the measurement equation of the Kalman filter. The weight vector of the constrained Kalman beamformer is derived and shown to converge to that of the minimum-variance distortionless-response beamformer (MVDR). The technique of incorporating state space concept and Kalman filter to solve the MVDR problem is presented in subsection 2.3.1 and subsection 2.3.2 by two forms of constraints. In later Sections, another formulation to maintain the distortionless constraint will be presented and investigated.

2.3.1 Soft-Constraint Kalman Filter under MVDR Structure

The traditional formulation and solution to MVDR is presented in Section 2.2. In this Section, The Kalman filter is introduced to solve the MVDR problem in a new formulation by Y.H. Chen and C.T. Chiang [12].

With the same formulation as MVDR structure above(2.10), state equations describing such formulation can be written as：

Measurement Equation：

$$\begin{bmatrix} 0 \\ f(\theta_s)=1 \end{bmatrix} = \begin{bmatrix} X^H(k, jw) \\ a^H(\theta_s, jw) \end{bmatrix} w(k, jw) + \begin{bmatrix} v_1(k, jw) \\ v_2(k, jw) \end{bmatrix} \rightarrow \mathbf{Y} = \mathbf{X}^H(k, jw)w(k, jw) + \mathbf{V}(k, jw) \quad (2.14)$$

Process Equation

$$w(k+1, w) = w(k, w) + Q(k, w) \quad (2.15)$$

where k is the frame index and the superscript "^{H}" means conjugate-transpose. A cure for the signal-distortionless problem is to constrain the gain of the system as unity on the desired signal from look direction and minimize the system output power subject to this constraint as Equation (2.7). The noise $\mathbf{V}(k,w)$ and $\mathbf{Q}(k,w)$ are assumed with Gaussian distribution and thus the covariance matrix can be written as

$$\begin{aligned}\mathbf{Q}(k,w) &\sim N(0,\sigma_Q^2 I)\\ \mathbf{V}(k,w) &\sim N(0,\begin{bmatrix}\sigma_{v1}^2 & 0\\ 0 & \sigma_{v2}^2\end{bmatrix})\end{aligned} \tag{2.16}$$

where "N" means Normal Distribution and σ_Q^2, σ_{v1}^2, σ_{v2}^2 are parameters to be chosen.

$\mathbf{X}(k,w)$ is the received signal and $a(\theta_s, jw)$ is the array manifold vector.

Let the state estimation error is

$$e(k|k-1,w) = w(k,w) - \hat{w}(k|k-1,w) \tag{2.17}$$

and the error covariance matrix is

$$R_{ee}(k|k-1,w) = E[e(k|k-1,w)e^T(k|k-1,w)] \tag{2.18}$$

In the first step, no new observation is used. To predict $w(k)$ by using the state equation, the best possible predictor would be as below which is given that no new information is available.

$$\hat{w}(k|k-1,w) = \hat{w}(k-1|k-1,w) \tag{2.19}$$

The estimation error is

$$\begin{aligned}e(k|k-1,w) &= w(k,w) - \hat{w}(k|k-1,w)\\ &= w(k-1,w) + Q(k,w) - \hat{w}(k-1|k-1,w)\\ &= e(k-1|k-1,w) + Q(k,w)\end{aligned} \tag{2.20}$$

If requiring that $E[\mathbf{e}(k-1|k-1,w)] = 0$ (this zero-mean condition states that there is no

constant bias in the optimal linear estimation [7]), $E[e(k|k-1,w)]=0.$ Since $e(k-1|k-1,w)$ is uncorrelated with $\mathbf{Q}(k,w)$,

$$R_{ee}(k|k-1,w)=R_{ee}(k-1|k-1,w)+\sigma_Q^2 I \text{ (Riccati Equation)} \tag{2.21}$$

In the second step, the new observation, $\begin{bmatrix}0\\ f(\theta_s)=1\end{bmatrix}=\mathbf{Y}(k,w)$ is incorporated to estimate $\mathbf{w}(k,w)$. A linear estimate that is based on $\hat{w}(k|k-1,w)$ and $\mathbf{Y}(k,w)$ has the form

$$\hat{w}(k|k,w)=\mathbf{K}'(k,w)\hat{w}(k|k-1,w)+\mathbf{k}(k,w)\mathbf{Y}(k,w) \tag{2.22}$$

where $\mathbf{K}'(k,w)$ and $\mathbf{k}(k,w)$ are some matrix and vector to be determined. The vector $\mathbf{k}(k,w)$ is called the Kalman gain. Now, the estimation error is

$$\begin{aligned}\mathbf{e}(k|k,w)&=w(k,w)-\hat{w}(k|k,w)\\&=w(k,w)-\mathbf{K}'(k,w)\hat{w}(k|k-1,w)-\mathbf{k}(k,w)\mathbf{Y}(k,w)\\&=w(k,w)-\mathbf{K}'(k,w)[w(k,w)-e(k|k-1,w)]-\mathbf{k}(k,w)[\mathbf{X}^H(k,w)w(k,w)+\mathbf{V}(k,w)]\\&=[I-\mathbf{K}'(k,w)-\mathbf{k}(k,w)\mathbf{X}^H]w(k,w)+\mathbf{K}'(k,w)e(k|k-1,w)-\mathbf{k}(k,w)\mathbf{V}(k,w),\end{aligned} \tag{2.23}$$

where $\mathbf{X}^H(k,w)=\begin{bmatrix}X^H(k,w)\\ a^H(\theta_s, jw)\end{bmatrix}$.

Since $E[e(k|k-1,w)]=0,$ then $E[e(k|k,w)]=0$ only if

$$\mathbf{K}'(k,w)=I-\mathbf{k}(k,w)\mathbf{X}^H \tag{2.24}$$

with this constraint, it follows that

$$\begin{aligned}\hat{w}(k|k,w)&=[I-\mathbf{k}(k,w)\mathbf{X}^H]\hat{w}(k|k-1,w)+\mathbf{k}(k,w)\mathbf{Y}(k,w)\\&=\hat{w}(k|k-1,w)+\mathbf{k}(k,w)[\mathbf{Y}(k,w)-\mathbf{X}^H\hat{w}(k|k-1,w)]\end{aligned} \tag{2.25}$$

and

$$\begin{aligned}e(k|k,w)&=\mathbf{K}'(k,w)e(k|k-1,w)-\mathbf{k}(k,w)\mathbf{V}(k,w)\\&=[I-\mathbf{k}(k,w)\mathbf{X}^H]e(k|k-1,w)-\mathbf{k}(k,w)\mathbf{V}(k,w).\end{aligned} \tag{2.26}$$

Since $\mathbf{V}(k,w)$ is uncorrelated with $\mathbf{Q}(k,w)$ and with $\mathbf{Y}(k-1,w)$, then $\mathbf{V}(k,w)$ will be uncorrelated with $w(k,w)$ and with $\hat{w}(k|k\text{-}1,w)$; as a result $E[e(k|k,w)\mathbf{V}(k,w)]$ =0. Therefore, the error covariance matrix for $e(k|k,w)$ is

$$\begin{aligned} R_{ee}(k|k,w) &= E[e(k|k,w)e^T(k|k,w)] \\ &= [I-\mathbf{k}(k,w)\mathbf{X}^H]R_{ee}(k|k-1,w)[I-\mathbf{k}(k,w)\mathbf{X}^H]^T + \mathbf{k}(k,w)R_v(k,w)\mathbf{k}^T(k,w), \end{aligned} \tag{2.27}$$

where $R_v(k,w) = \sigma_{v1}^2\begin{bmatrix} 1 & 0 \\ 0 & \rho_v \end{bmatrix}$.

The final task is to find the Kalman gain vector $\mathbf{k}(k,w)$, that minimizes the MSE

$$J(k) = tr[R_{ee}(k|k,w)] \tag{2.28}$$

Differentiating $J(k)$ with respect to $\mathbf{k}(k,w)$, we get

$$\frac{\partial J(k)}{\partial \mathbf{k}(k,w)} = -2[I-\mathbf{k}(k,w)\mathbf{X}^H]R_{ee}(k|k-1,w)\mathbf{X} + 2\mathbf{k}(k,w)R_v(k,w) \tag{2.29}$$

and equating it to zero, we deduce the Kalman gain

$$\mathbf{k}(k,w) = R_{ee}(k|k-1,w)\mathbf{X}\left[\mathbf{X}^H R_{ee}(k|k-1,w)\mathbf{X} + R_v(k,w)\right]^{-1} \tag{2.30}$$

The expression for the error covariance matrix can be simplified as

$$\begin{aligned} R_{ee}(k|k,w) &= [I-\mathbf{k}(k,w)\mathbf{X}^H]R_{ee}(k|k-1,w) - \\ &\{[I-\mathbf{k}(k,w)\mathbf{X}^H]R_{ee}(k|k-1,w)[I-\mathbf{k}(k,w)\mathbf{X}^H] + R_v(k,w)\mathbf{k}(k,w)\}\mathbf{k}^T(k,w), \end{aligned} \tag{2.31}$$

Where, by using (2.29), the second term in (2.31) is equal to zero. Hence

$$R_{ee}(k|k,w) = [I-\mathbf{k}(k,w)\mathbf{X}^H]R_{ee}(k|k-1,w) \tag{2.32}$$

In conclusion, the Soft-Constraint Kalman filter can be summarized as following. The signal-flow graph of the constrained Kalman algorithm can be plotted as Fig. 3.

Algorithm: Soft-Constraint Kalman filter

State Equation:

$$\hat{w}(k+1|k,w) = \hat{w}(k|k,w) + \mathbf{Q}(k,w)$$

Measurement Equation (Cost Equation):

$$\mathbf{Y}(k,w) = \begin{bmatrix} 0 \\ f(\theta_s) = 1 \end{bmatrix} = \begin{bmatrix} X^H(k,w) \\ a^H(\theta_s, jw) \end{bmatrix} w(k,w) + \mathbf{V}(k,w) = \mathbf{X}^H(k,w) w(k,w) + \mathbf{V}(k,w)$$

Computation for $k = 1,2,\cdots$

$$\hat{w}(k|k-1,w) = \hat{w}(k-1|k-1,w)$$

$$R_{ee}(k|k-1,w) = R_{ee}(k-1|k-1,w) + \sigma_Q^2 I$$

The Kalman gain:

$$\mathbf{k}(k,w) = R_{ee}(k|k-1,w)\mathbf{X}\left[\mathbf{X}^H R_{ee}(k|k-1,w)\mathbf{X} + R_v(k,w)\right]^{-1}$$

$$\hat{w}(k|k,w) = \hat{w}(k|k-1,w) + \mathbf{k}(k,w)[\mathbf{Y}(k,w) - \mathbf{X}^H(k,w)\hat{w}(k|k-1,w)]$$

$$R_{ee}(k|k,w) = [I - \mathbf{k}(k,w)\mathbf{X}^H(k,w)]R_{ee}(k|k-1,w)$$

Table 1. Soft-Constraint Kalman filter algorithm.

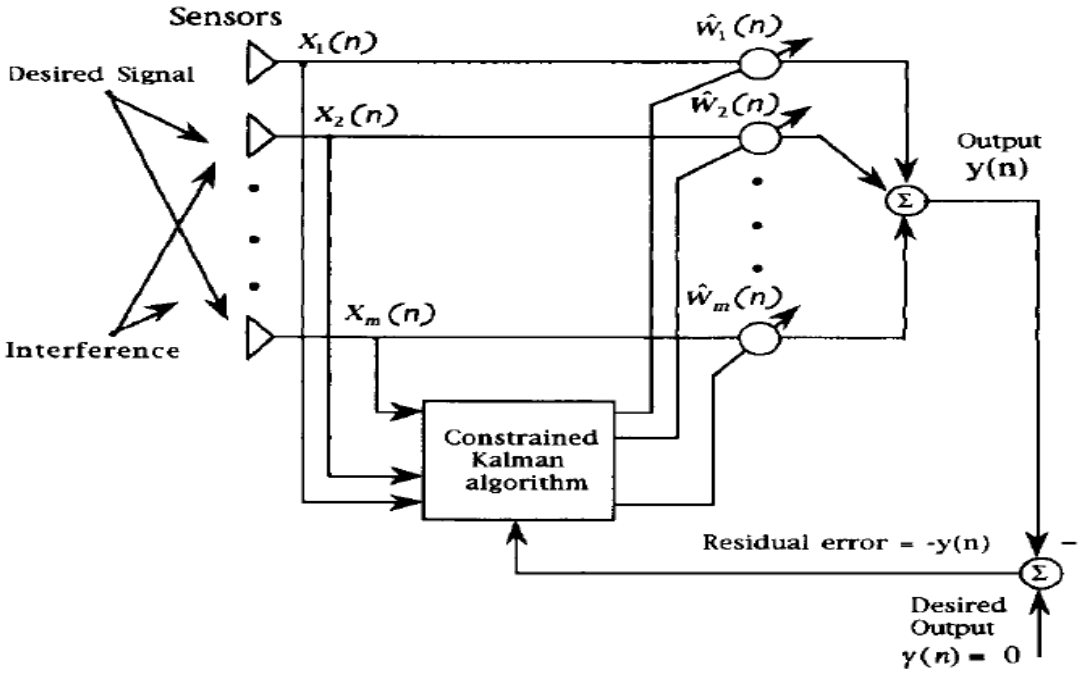


Fig. 3 Block diagram of the Soft-Constrained Kalman filter algorithm[12].

Soft equality constraints are constraints that are only required to be approximately satisfied rather than exactly satisfied. The perfect measurement approach (Kalman-Filter) can be extended to soft constraints by adding small nonzero measurement noise to the perfect measurement as ρ_v. It makes the optimal Lagrange Multiplier solution be a trade-off between the residual error and the constraint error. The methods are implemented in cases where the constraints are heuristic rather than exactly satisfied. But in thesis cases, we want to use Hard-constraints, as opposed to Soft-constraints, to solve the signal mismatch problem which is discussed in next Chapter.

2.3.2 Hard-Constraint Kalman Filter under MVDR Structure

A number of approaches have been proposed for solving the constrained Kalman Filtering problem [13], [14], [15], [26]. Analogous to the way that a Kalman filter can be extended to solve problems containing nonlinearities, linear equality constrained filtering can be extended to problems with nonlinear constraints by linearizing locally. The accuracy achieved by methods dealing with nonlinear constraints will naturally depend on the structure and curvature of the nonlinear function itself. We would want to implement the equality constraints that are exactly satisfied specifically.

In this subsection, there are two distinct approaches which are discussed to generalize an equality constrained Kalman Filter. The first approach is to run an unconstrained Kalman Filter and project the estimate down to the equality constrained space in each iteration. The second approach will start with a constrained prediction, and restrict the Kalman Gain so that the estimate will lie in the constrained space. Finally, we will show the numerical preservation of the updated error covariance with the feedback loop in the projection framework. The equality constraint in this subsection can be defined as below, where A is a $q \times w$ matrix, b is a q-vector, and w_k the state is a w-vector.

$$A(w_k) = b \tag{2.33}$$

So the updated state estimate and state prediction to be constrained at each iteration, which would allow a better forecast in the system, as below.

$$\begin{cases} A\hat{w}_{k|k} = b & \text{state} - \text{estimate} \\ A\tilde{w}_{k|k-1} = b & \text{state} - \text{prediction} \end{cases} \tag{2.34}$$

In the following two approaches, we will discuss the constraining updated state estimate.

A. Projecting the Unconstrained Estimate

This method projects the state to lie in the constrained space each iteration, feeds the new constrained estimate back to the unconstrained Kalman Filter and continues this process. Such method can be described by the following minimization problem for a given time-step k, in which $\hat{w}_{k|k}^{P}$ is the constrained estimate, $\hat{w}_{k|k}$ is the unconstrained estimate from the Kalman Filter equation, and W_k is any positive definite symmetric weighting matrix, where the superscript 'P' is used to denoted the "Projected" constrained filter and 'U' is denoted as "Unconstrained".

$$\hat{w}_{k|k}^{P} = \arg\min_{w \in \mathbb{R}^n} \left\{ (w - \hat{w}_{k|k}^{U})^T W_k (w - \hat{w}_{k|k}^{U}) : Aw = b \right\} \tag{2.35}$$

The constrained estimate is then solved by the Lagrange Multiplier as equation (2.11~2.13), which is given as below:

$$\hat{w}_{k|k}^{P} = \hat{w}_{k|k}^{U} - W_k^{-1} A^T (A W_k^{-1} A^T)^{-1} (A\hat{w}_{k|k}^{U} - b) \tag{2.36}$$

In a general case, we can find out the updated error covariance as a function of the unconstrained Kalman Filter's updated error covariance matrix as before. Define the matrix Υ as below first.

$$\Upsilon = W_k^{-1} A^{'} (A W_k^{-1} A^{'})^{-1} \tag{2.37}$$

Equation (2.36) can be re-written as follows:

$$\hat{w}_{k|k}^{P} = \hat{w}_{k|k}^{U} - \Upsilon(A\hat{w}_{k|k}^{U} - b) \tag{2.38}$$

The reduced form for $w_k - \hat{w}_{k|k}^{P}$ as below:

$$\begin{aligned} w_k - \hat{w}_{k|k}^{P} &= w_k - \hat{w}_{k|k} + \Upsilon(Aw_{k|k} - b - (Aw_k - b)) \\ &= w_k - \hat{w}_{k|k} + \Upsilon A(w_k - \hat{w}_{k|k}) \\ &= -(I - \Upsilon A)(\hat{w}_{k|k} - w_k) \end{aligned} \tag{2.39}$$

According to the definition of the error covariance matrix, we arrive at the following expression.

$$\begin{aligned} P_{k|k}^{P} &= E\left[(w_k - \hat{w}_{k|k}^{P})(w_k - \hat{w}_{k|k}^{P})^{'}\right] \\ &= E\left[(I - \Upsilon A)(\hat{w}_{k|k} - x_k)(\hat{w}_{k|k} - x_k)^{'}(I - \Upsilon A)^{'}\right] \\ &= (I - \Upsilon A)P_{k|k}(I - \Upsilon A)^{'} \\ &= P_{k|k} - \Upsilon A P_{k|k} - P_{k|k}A^{'}\Upsilon^{'} + \Upsilon A P_{k|k}A^{'}\Upsilon^{'} \\ &= P_{k|k} - \Upsilon A P_{k|k} \\ &= (\mathrm{I} - \Upsilon A)P_{k|k} \end{aligned} \tag{2.40}$$

Note that the ΥA in equation (2.39) is a projection matrix, as is $(I - \Upsilon A)$, so we can deduce the equation (2.40) to the result. It can be shown that there is smallest updated error covariance when $W_k = P_{k|k}^{-1}$. It also provides a measure of the information in the state k.

B. Restricting the optimal Kalman Gain

Alternatively, for restricting the optimal Kalman gain so the updated state estimate lies in the constrained space, we can expand the updated state estimate term in Equation (2.34) using Equation (2.25), the state update is $\hat{w}_{k|k} = \hat{w}_{k|k-1} + K_k V_k$, where $V_k = (Y_k - H_k \hat{w}_{k|k-1})$.

$$A(\hat{w}_{k|k-1} + K_k V_k) = b \tag{2.41}$$

The MMSE of the estimate $\hat{w}_{k|k}$ is $E\left[(w_k - \hat{w}_{k|k})^T (w_k - \hat{w}_{k|k})\right]$, which is equivalent to the trace of error covariance matrix of $\hat{w}_{k|k}$. Then, we can choose a Kalman Gain K_k^P that forces the updated state estimate to be in the constrained space. We choose the optimal Kalman Gain K_k which yields Equation (2.30) in the unconstrained case by solving the minimization below

$$K_k = \arg\min_{K \in \Re^{n*m}} trace\left[(I - K_k H_k) P_{k|k-1} (I - K_k H_k)' + K_k R_k K_k'\right] \tag{2.42}$$

Now we seek the optimal K_k^P that satisfies the constrained optimization problem written below for a time-step k.

$$\begin{aligned} K_k^P = \arg\min_{K \in \Re^{n*m}} & trace\left[(I - K_k H_k) P_{k|k-1} (I - K_k H_k)' + K_k R_k K'\right] \\ s.t. & A(\hat{w}_{k|k-1} + K_k V_k) = b \end{aligned} \tag{2.43}$$

Using the method of Lagrange Multiplier technique to solve the optimal minimization,

$$\begin{aligned} J_k = & Tr((I - K_k H_k) P_{k|k-1} (I - K_k H_k)' + K_k R_k K_k') \\ & + \lambda_k^T (A(\hat{w}_{k|k-1} + K_k V_k) - b) \end{aligned} \tag{2.44}$$

where the λ_k is the Lagrange multiplier. Taking the derivative of J_k with respect to K_k and setting it to zero yields

$$-2P_{k|k-1} H_k^T + 2K_k C_k + A_k^T \lambda_k V_k^T = 0 \tag{2.45}$$

where $C_k = H_k P_{k|k-1} H_k^T + R_k$. Then we find the following Kalman gain.

$$K_k^P = P_{k|k-1} H_k^T C_k^{-1} - \frac{1}{2} A_k^T \lambda_k V_k^T C_k^{-1} \tag{2.46}$$

Applying Equation (2.41), after some manipulations, the optimal Lagrange multiplier is obtained as below.

$$\lambda_k = 2(V_k^T C_k^{-1} V_k A_k A_k^T)^{-1} (A_k P_{k|k-1} H_k^T C_k^{-1} V_k + A_k^T \hat{w}_{k|k-1} - b) \tag{2.47}$$

Finally, substituting the optimal Kalman Gain and Lagrange multiplier into the Equation (2.41) that yields the following constrained updated state estimate:

$$\hat{w}_{k|k}^{P} = v_k^U - A_k^T (A_k A_k^T)^{-1} (A_k \hat{w}_k^U - b) \tag{2.48}$$

This is of course equivalent to the result of Equation (2.36) with the weighting matrix W_k which is chosen as the identity matrix in open loop scheme. The error covariance is given by Equation (2.40). That is, the Kalman Filter can be run in real-time, and as a post-processing step, the unconstrained estimate and updated error covariance matrix can be reformulated in the constrained space; or alternatively, the constrained estimate and its updated error covariance matrix can be fed back into the system in real-time. A large benefit of incorporating constraints can be realized in both techniques, though the feedback system should generally outperform the system without feedback.

Numerical Preservation of the Updated Error Covariance

These methods are shown to be mathematically equivalent under the assumption that the weighting matrix W_k is chosen appropriately. In [15], with the augmentation method, the soft equality constraints can be deduced to be equivalent to hard equality constraints into a Kalman filter by adding a proportionate amount of noise to the bottom right error covariance matrix (see Equation (2.16)). However, the numerical round-off error will be ignored possibly in implementations. We will not see the exact same result while these methods can be mathematical equivalent. The round-off error that causes the most trouble occurs when the updated error covariance matrices lose symmetry or positive definiteness. According to feedback loop of the projection frame, we should find a form of state estimate and state error covariance that preserves symmetry and positive definiteness better.

[15] uses the numerical preservation of updated error covariance to find a simplified form for the constrained updated error covariance as below:

$$P_{k|k}^{P} = (I - K_k^P H_k^P) P_{k|k-1}^{U} (I - K_k^P H_k^P)' + K_k^P R_k K_k^{P'} \tag{2.49}$$

Define the projection matrix Γ_k,

$$\Gamma_k = I - P_{k|k}^{U} A' (A P_{k|k}^{U} A')^{-1} A \tag{2.50}$$

In term of Γ_k, the following are true.

$$\begin{cases} P_{k|k}^{P} = \Gamma_k P_{k|k}^{U} \\ I - K_{k|k}^{P} H_{k|k}^{P} = \Gamma_k (I - K_{k|k}^{P} H_{k|k}^{P}) \end{cases} \tag{2.51}$$

Finally, in order to maintain numerical stability we can find out the simplified form of constrained updated error covariance by using Equation (2.49).

$$\begin{aligned} P_{k|k}^{P} &= (I - K_k^P H_k^P) P_{k|k-1}^{U} (I - K_k^P H_k^P)' + K_k^P R_k K_k^{P'} \\ &= \Gamma_k (I - K_k^U H_k^U) P_{k|k-1}^{U} (I - K_k^U H_k^U)' \Gamma_k' \\ &\quad + \Gamma_k K_k^U R_k K_k^{U'} \Gamma_k' \\ &= \Gamma_k \left[(I - K_k^U H_k^U) P_{k|k-1}^{U} (I - K_k^U H_k^U)' + K_k^U R_k K_k^{U'} \right] \Gamma_k' \\ &= \Gamma_k P_{k|k}^{U} \Gamma_k' \end{aligned} \tag{2.52}$$

In conclusion, the Hard-Constraint Kalman Filter can be summarized under MVDR structure as following :

Algorithm: Hard-Constraint Kalman filter

State Equation:

$$\hat{w}(k+1|k,w) = \hat{w}(k|k,w) + \mathbf{Q}(k,w)$$

Measurement Equation (Cost Equation):

$$\mathbf{Y}(k,w) = [0] = \mathbf{X}^H(k,w)w(k,w) + \mathbf{V}(k,w)$$

Computation for $k = 1,2,\cdots$

$$\hat{w}(k|k-1,w) = \hat{w}(k-1|k-1,w)$$

$$R_{ee}(k|k-1,w) = R_{ee}(k-1|k-1,w) + \sigma_Q^2 I$$

The Kalman gain :

$$k(k,w) = R_{ee}(k|k-1,w)\mathbf{X}\left[\mathbf{X}^H R_{ee}(k|k-1,w)\mathbf{X} + R_v(k,w)\right]^{-1}$$

Update the state estimate :

$$\hat{w}(k|k,w) = \hat{w}(k|k-1,w) + \mathbf{k}(k,w)[\mathbf{Y}(k,w) - \mathbf{X}^H(k,w)\hat{w}(k|k-1,w)]$$

$$\hat{w}(k|k,w) = \hat{w}(k|k,w) - a^H(\theta_s, jw)(a(\theta_s, jw)a^H(\theta_s, jw))^{-1}(a^H(\theta_s, jw)\hat{w}(k|k,w) - b)$$

Update the error covariance :

$$R_{ee}(k|k,w) = [I - \mathbf{k}(k,w)\mathbf{X}^H(k,w)]R_{ee}(k|k-1,w)$$

$$R_{ee}(k|k,w) = [I - a^H(\theta_s, jw)(a(\theta_s, jw)a^H(\theta_s, jw))^{-1}a(\theta_s, jw)]*$$
$$R_{ee}(k|k,w)[I - a^H(\theta_s, jw)(a(\theta_s, jw)a^H(\theta_s, jw))^{-1}a(\theta_s, jw)]'$$

b : Response value

Table 2. Hard-Constraint Kalman filter algorithm.

2.4 Adaptive Constrained Least Mean Square Beamformer

The optimal Linearly-Constrained Minimum-Variance Filter (LCMV) in the concept of the minimum mean output energy (MOE) that minimizes the same objective function as MVDR beamformer, and projects to the set of the linear constraints. In [5] Frost has proposed an algorithm to estimate w_{opt} based on the least-mean-square (LMS) algorithm for adaptive filtering. Under the MVDR structure, linearly-constrained adaptive filters are deduced as below.

Core of the LMS algorithm is to find a weighting to minimize the error covariance between desired source and filtered output. Assume the desired signal, that we want to achieve, is zero-mean and the variance is σ_d^2. Auto-correlation and cross-correlation matrices definition of the input signal are the following.

$$E\{d(k)\} = 0, \sigma_d^2 = E\left\{|d(k)|^2\right\}$$
$$R_{xx} = E\{x(k) * x(k)\} \quad \text{Auto} - \text{correlation matrix}$$
$$R_{dx} = E\left\{\mathrm{d}(k)x(k)^*\right\} \quad \text{Cross} - \text{correlation matrix}$$

Then the Cost function is as Equation (2.53).

$$J(k) = \min_w E\left\{d(k) - x(k) * w(k)\right\}^2 = E(d(k) - x(k) * w(k))(d(k) - x(k) * w(k))^* \quad (2.53)$$

Then the method to find the weighting is the Steepest-Descend Method as

$$w(k) = w(k-1) + \mu p \quad (2.54)$$

μ is the proportion called step-size (or convergence factor), and the p choosing is deduced from (2.53). Expanding the Equation (2.53):

$$J(k) = \sigma_d^2 - R_{dx}^* w(k) - w^*(k) R_{dx} + w^*(k) R_{xx} w(k) \quad (2.55)$$

We take the ∇_w of (2.55) to find the minimization and we get

$$p = R_{dx} - R_{xx} w(k-1), \quad (2.56)$$

For the purpose to let $w(k)$ on the lowest direction and strength of $J(k)$,

and rewritten the (2.54)

$$w(k) = w(k-1) + \mu\left[R_{dx} - R_{xx}w(k-1)\right] \quad k \geq 1 \tag{2.57}$$

Thus, LMS algorithm can be organized as below

Filter out : $y(k) = x^*(k)w(k)$

Error function : $e(k) = d(k) - y(k)$ (2.58)

Update weight : $w(\mathrm{k}-1) + \mu x^*(k)e(k)$

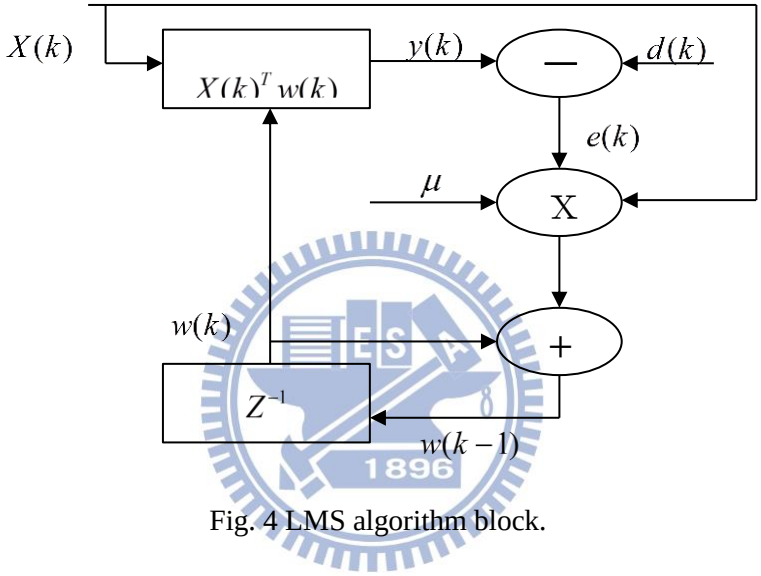


Fig. 4 LMS algorithm block.

Since the frequency response of look-direction is fixed by the desired source constraint under MVDR structure, minimization of the non-look-direction noise power is the same as minimization of the total output power as Equation (2.10). The gradient-descent constrained LMS algorithm presented as below using the same method of Lagrange multipliers, which is discussed in Equation (2.11).

$$\begin{aligned} w(k) &= w(k-1) - \mu\nabla_w H\left[w(k-1)\right] \\ &= w(k-1) - \mu\left[R_{xx}w(k-1) - a(\theta_s)\lambda(k)\right] \end{aligned} \tag{2.59}$$

The Lagrange multipliers are chosen by requiring $w(k)$ to satisfy the constraint

$$f(\theta_s) = a^T(\theta_s)w(k) = a^T(\theta_s)w(k) - \mu a^T(\theta_s)R_{xx}w(k) - \mu a^T(\theta_s)a(\theta_s)\lambda(k) = 1 \tag{2.60}$$

Solving for the Lagrange multipliers $\lambda(k)$ and substituting back to the weight

iteration Equation (2.59),

$$w(\mathrm{k}) = w(k-1) - \mu\left[I - a(\theta_s)(a^T(\theta_s)a(\theta_s))^{-1}a^T(\theta_s)a(\theta_s) \right] R_{xx} w(\mathrm{k}) \\ + a(\theta_s)(a^T(\theta_s)a(\theta_s))^{-1}a^T(\theta_s)\left[1 - a^T(\theta_s)w(\mathrm{k}) \right] \tag{2.61}$$

Defining the M dimension vector (F), and $M \times M$ matrix (P),

$$P \triangleq I - a(\theta_s)(a^T(\theta_s)a(\theta_s))^{-1}a^T(\theta_s)$$

is the projection matrix onto the subspace orthogonal to the subspace spanned by the constraint matrix.

$$F \triangleq a(\theta_s)(a^T(\theta_s)a(\theta_s))^{-1}$$

The algorithm could be rewritten as

$$w(k) = P\left[w(k-1) - \mu R_{xx} w(k) \right] + F \tag{2.62}$$

where the algorithm requires the prior knowledge of the input correlation matrix R_{xx}, which, however, is unavailable a priori in the array problem. Using the outer product of tap voltage vector to approximate R_{xx} at k_{th} iteration with itself : $x(k)x^T(k)$, gives the stochastic constrained LMS algorithm.

$$w(0) = F \\ w(k) = P\left[w(k-1) - \mu y(k)x(k) \right] + F \tag{2.63}$$

Note that [11] in (2.63) the term multiplied by the projection matrix corresponds to the unconstrained LMS solution, which is projected onto the homogeneous hyperplane $a^T(\theta_s)w(k) = 0$, and moved back to the constraint hyperplane by adding the vector F.

In the LMS algorithm, for sure the convergence of the algorithm, the range of step-size μ must to be in $0 < \mu < \frac{2}{\lambda_{\max}}$, $\lambda_{\max}$ is the maximum eigenvalue of the R_{xx}. If the order of the spatial filter is higher, it is more difficult to solve the problem. Thus, for simplifying the computation load, there is normalized version derived.

$$w(k) = w(k-1) + \frac{\alpha x(k) w(k)}{\gamma + x^*(k)x(k)} \quad \text{(NLMS)} \tag{2.64}$$

In Equation (2.64), the difference between NLMS and LMS is weight updating. The step-size is replaced by the $\frac{\alpha}{\gamma + x^*(k)x(k)}$. $0 < \alpha < 2$, γ is the number for sure that the denominator is not equal to zero. In that way, the algorithm of NLMS will converge. Another normalized version of CLMS is shown in [11], which uses the result of NLMS for convergence speed need. Since the instantaneous error is given in Equation (2.53), the instantaneous squared error (posterior error)[11] can be written as.

$$\begin{aligned} e'^2(k) &= (d(k) - x^T(k)w(k+1))^2 \\ &= (d(k) - x^T(k)(w(k) + \mu_k x(k)))^2 \end{aligned} \tag{2.65}$$

Take the partial derivative of the $e'^2(k)$ with respect to μ_k and make it to zero,

$$\mu_k = \frac{d(k) - x^T(k)w(k)}{x^T(k)x(k)} \tag{2.66}$$

and start with the NLMS algorithm as same as the CLMS algorithm.

$$\begin{aligned} w(k+1) &= Pw_{NLMS}(k+1) + F \\ &= P\left[w(k) + \mu_k x(k)\right] + F \end{aligned} \tag{2.67}$$

Remember that $w(k)$ has to satisfy the constraint in Equation (2.60) which means that $w(k) = Pw(k) + F$, and the Equation (2.67) can be rewritten as

$$w(k+1) = w(k) + \mu_k P x(k) \tag{2.68}$$

From Equation (2.58) and (2.68), (2.66), we substitute the input vector by a rotated version of $x'(k) = Px(k)$. Moreover, recalling that $P^2 = P$, it follows that

$$\begin{aligned} e(k) &= d(k) - x^T(k)w(k) \\ w(k) &= P\left[w(k-1) + \frac{e(k)x(k)}{x^T(k)Px(k)} \right] + F \quad \text{(NCLMS)} \end{aligned} \tag{2.69}$$

And in Equation (2.69), the step-size, which is normalized with respect to the energy of projected input vectors, makes sure the convergence of the algorithm.

Chapter 3. Robust Adaptive Beamformer to the Signal Mismatch Problem

3.1 Introduction

Robust speech enhancement algorithm arises in many practical applications where the desired source is usually contaminated by background noise and influenced by reverberation in the beamformer training data. (e.g., mobile communications, passive source location, microphone arrays speech processing, medical imaging, and radio astronomy).

One of the key issues in adaptive beamformers is the sensitivity due to the mismatch between the presumed signal steering vector of desired signal and the actual one (e.g., mismatches due to array perturbations, array manifold vector mismodeling, wave-front distortions, or source local scattering). Several approaches have been developed to overcome arbitrary mismatches since past three decades, such as diagonal loading of the sample covariance matrix [24] and the eigenspace beamformer [25]. However, for the former approach, it is not clear how to obtain the optimal value of the diagonal loading factor based on the known level of uncertainty of the signal steering vector; For the latter, it is limited to the high signal-to-noise ratios (SNRs) and the dimension of signal-plus-interference subspace. Then one of the theoretically rigorous and efficient approaches to robust beamforming in the presence of an arbitrary unknown steering signal mismatch is based on worst-case performance optimization [17]. It optimizes the weighting vector by minimizing the output interference-plus-noise power while maintaining a distortionless response in the worst case. The robust minimum variance distortionless response (MVDR: the dominant structure of this thesis) was formulated in [17] as a second-order cone programming (SOCP) problem, which can be solved in polynomial time using interior point method. In further works, several extensions of the robust MVDR beamformer have been

developed in [19], [20], and in alternative Newton-type iterative algorithms. The dominant shortcoming of these algorithms is that they do not have a computationally efficient online implementation.

In this thesis, the constrained Kalman filter is used to solve the linearly equality constrained MVDR beamforming problem, which is first proposed in [12]. Following the similar idea, the second-order extended (SOE) Kalman filter for recursive implementation uses the robustness constraint, which is incorporated into the measurement equation instead of the conventional distortionless constraint used in [12]. To derive the SOE Kalman filter based on the MVDR beamformer in the frequency domain [17], the second-order Taylor series expansion is used to approximate the nonlinear function involved in the beamformer under the assumption that the dynamics of the signal generating processes are known. However these assumptions limit the performance on the practical application. To solve the limitation of performance, this thesis combines the voice activity detection information and the constraint projected method with the SOE Kalman filter to strengthen the performance of speech enhancement against steering mismatch problem.

3.2 Formulation of Signal Mismatch Problem

According to the Equation (2.10), the well-known MVDR beamformer minimizes the output power of the interference-signals-plus-stationary-noise while maintaining a distortionless response to the desired signal, which is given in the frequency domain (the imaginary unit j is omitted for equation simplification),

$$\min w_{MV}^{H}(w)R_{XX}(w)w_{MV}(w) \text{ subject to } w_{MV}^{H}(w)a(\theta_s, w) = f(\theta_s) = 1$$

where

$$R_{xx}(w) = E\left\{X(k,w)X^{H}(k,w)\right\}$$

$R_{xx}(w)$ is the $M \times M$ correlation matrix and the $a(\theta_s, w) \in \mathbb{C}^{M\times 1}$ is the presumed

steering vector. The solution of the MVDR problem is given by Equation (2.13)

$$w_{MV}(w) = \frac{R_{XX}^{-1}(w)a(\theta_s, w)}{a^H(\theta_s, w)R_{XX}^{-1}(w)a(\theta_s, w)}$$

in real application, the correlation matrix is unavailable and usually approximated by

$$R_{xx}(w) = \frac{1}{N}\sum_{k=1}^{N} X(k, w)X^H(k, w) \tag{3.1}$$

where N is the number of frames available, which is called sample covariance matrix. Using the Equation (3.1) to replace the true correlation matrix and the resulting solution is commonly referred to the sample matrix inversion (SMI) algorithm. The main disadvantage of the SMI algorithm is that it is not robust to the mismatch between the presumed steering vector and the actual one, and the algorithm degrades dramatically if the desired signal is present in the training snapshots. The signal self-cancellation phenomenon observed in such case is commonly known as signal self-nulling. The phenomenon is shown in Fig. 5. Obviously, there is a severe null at the signal of interest (SOI) when the presumed steering vector is different to the actual steering vector. The goal of solutions to the signal mismatch problem is to constrain more steering vectors to exceed unity gain while suppressing the interferences as the red solid dotted line which is expressed in Fig. 5.

For the sufficient robustness against the desired signal mismatch problem, the norm of the steering vector distortion can usually be bounded by some known constant $\varepsilon > 0$ [18], and the actual steering vector belongs to the set

$$\Lambda(\varepsilon) \equiv \left\{ C(w) \mid C(w) = a(\theta_s, w) + e(w), \|e(w)\| \le \varepsilon \right\} \tag{3.2}$$

the robust formulation of the MVDR beamformer can be written as [18].

$$\min w_{MV}^H(w)R_{xx}(w)w_{MV}(w) \ \text{ subject to } \ \left| w_{MV}^H(w)C(w) \right| \ge 1 \ \ \textit{for all } C(w) \in \Lambda(\varepsilon) \tag{3.3}$$

Such distortionless response is maintained by means of inequality constraints for a continuum of all possible steering vectors by the set $\Lambda(\varepsilon)$. For each choice of

$C(w) \in \Lambda(\varepsilon)$, the condition $\left|w_{MV}^H(w)C(w)\right| \geq 1$ represents a nonlinear and nonconvex constraint on w. There are infinite vectors $C(w)$ in $\Lambda(\varepsilon)$. Hence, the Equation (3.3) is a semi-infinite nonconvex quadratic program. It is well known that the nonconvex constrained quadratic programming problem is NP-hard. Such problems are difficult to solve in engineering. For simpler formulation of Equation (3.3), we convert the semi-infinite nonconvex constraints to a single constraint that corresponds to the worst-case constraint. The Equation (3.3) can be equivalently described as below.

$$\min w_{MV}^H(w)R_{xx}(w)w_{MV}(w) \text{ subject to } \min_{C(w)\in\Lambda(\varepsilon)} \left|w_{MV}^H(w)C(w)\right| \geq 1 \tag{3.4}$$

According to Equation (3.2), we can rewrite the constraint of Equation (3.4) as

$$\min_{C(w)\in\Lambda(\varepsilon)} \left|w_{MV}^H(w)a(\theta_s, w) + w_{MV}^H(w)e(w)\right| \geq 1$$

Applying the triangle and Cauchy-Schwarz inequalities along with the inequality $\|e(w)\| \leq \varepsilon$, we know that

$$\left|w_{MV}^H(w)a(\theta_s, w) + w_{MV}^H(w)e(w)\right| \geq \left|w_{MV}^H(w)a(\theta_s, w)\right| - \left|w_{MV}^H(w)e(w)\right| \geq \left|w_{MV}^H(w)a(\theta_s, w)\right| - \varepsilon\|w\| \tag{3.5}$$

Moreover, we can deduce that

$$\left|w_{MV}^H(w)a(\theta_s, w) + w_{MV}^H(w)e(w)\right| = \left|w_{MV}^H(w)a(\theta_s, w)\right| - \varepsilon\|w\| \tag{3.6}$$

Then, the semi-infinite nonconvex quadratic constrained problem can be replaced as the following quadratic minimization problem with a single nonlinear constraint:

$$\min w_{MV}^H(w)R_{xx}(w)w_{MV}(w) \text{ subject to } \left|w_{MV}^H(w)a(\theta_s, w)\right| - \varepsilon\left\|w_{MV}(w)\right\| \geq 1 \tag{3.7}$$

The nonlinear constraint in Equation (3.7) is still nonconvex due to the absolute value operation on the left-hand side. We can observe that the cost function in Equation (3.7) is a complex number which is unchanged when the w undergoes an arbitrary phase rotation. Hence, choose w such that without any loss of generality.

$$\mathrm{Re}\left\{w_{MV}^{H}(w)a(\theta_s,w)\right\}\geq 0 \quad \mathrm{Im}\left\{w_{MV}^{H}(w)a(\theta_s,w)\right\}=0 \tag{3.8}$$

Using this observation and employing Equation (3.8) as additional constraints, the constraint in Equation (3.7) can be rewritten as

$$w_{MV}^{H}(w)a(\theta_s,w)-\varepsilon\left\|w_{MV}(w)\right\|\geq 1 \tag{3.9}$$

Note that the formulation is much simpler than the Equation (3.3) and is convex. It can be proved that the inequality constraint in (3.9) can be presented by an equality one under some assumption in [18].

$$\begin{aligned}&\min w_{MV}^{H}(w)R_{xx}(w)w_{MV}(w)\\&\text{subject to}\quad \left|w_{MV}^{H}(w)a(\theta_s,w)-1\right|^2=\varepsilon^2 w_{MV}^{H}(w)w_{MV}(w)\end{aligned} \tag{3.10}$$

The problem in (3.4) has been solved by second-order cone programming in [17]. In the next Section, we will briefly review the SOE Kalman filter solution and combine with a new approach based on it for speech enhancement.

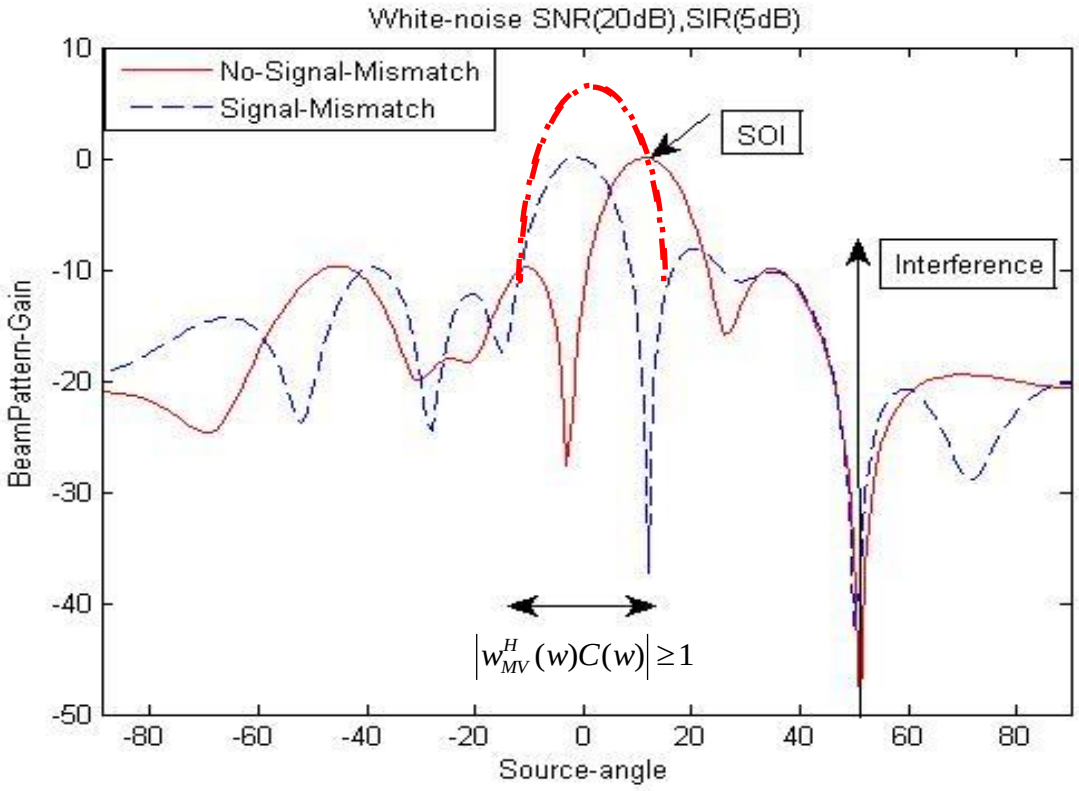


Fig. 5 Beampattern of signal mismatch and non-mismatch condition.

3.3 Solution to the Signal Mismatch Problem Using SOE Kalman Filter

The constraint in Equation (3.10) can be rewritten as

$$f_2(w_{MV}(k,w)) = f(\theta_s) = 1 \tag{3.11}$$

where

$$\begin{aligned} f_2(w_{MV}(k,w)) = {} & \varepsilon^2 w_{MV}^H(k,w) w_{MV}(k,w) \\ & - w_{MV}^H(k,w) a(\theta_s,w) a^H(\theta_s,w) w_{MV}(k,w) \\ & + w_{MV}^H(k,w) a(\theta_s,w) + a^H(\theta_s,w) w_{MV}(k,w) \end{aligned} \tag{3.12}$$

For the sake of the convenience of analysis, the mean square error (MSE) between the zero signals and the beamformer output is presented as following

$$E\left[\left|0 - X^H(k,w) w_{MV}(k,w)\right|^2\right] = w_{MV}^H(k,w) R_{xx}(w) w_{MV}(k,w) \tag{3.13}$$

where $E(\cdot)$ denotes the expectation operation. Therefore, the robust MVDR beamformer problem can be formulated as

$$\begin{aligned} & \min_{w_{MV}} E\left[\left|0 - X^H(k,w) w_{MV}(k,w)\right|^2\right] \\ & \text{subject to } f_2(w_{MV}(k,w)) = 1 \end{aligned} \tag{3.14}$$

The Kalman filter is a minimum mean square error (MMSE) estimator, using the state-space model to model an unknown dynamic system of the constraint minimization problem in Equation (3.14) is:

State Equation

$$w_{MV}(k,w) = w_{MV}(k-1,w) + \mathbf{v}_s(k,w) \tag{3.15}$$

Measurement Equation

$$\begin{aligned} \bar{y} &= \begin{bmatrix} X^H(k,w) w_{MV}(k,w) \\ f_2(w_{MV}(k,w)) \end{bmatrix} + \begin{bmatrix} v_1(k,w) \\ v_2(k,w) \end{bmatrix} \\ &= \boldsymbol{f}(w_{MV}(k,w)) + \mathbf{v}_m(\mathrm{k,w}) \end{aligned} \tag{3.16}$$

where $\mathbf{v}_s(k,w)$ and $\mathbf{v}_m(k,w)$ are the process and measurement noise respectively, and modeled as zero-mean, independent white noise sequences with the covariance matrices Q and R.

$$E\left[\boldsymbol{v}_s(k,w)\boldsymbol{v}_s^H(k,w)\right]=\sigma_s^2 I=\tilde{Q}$$
$$E\left[\boldsymbol{v}_m(k,w)\boldsymbol{v}_m^H(k,w)\right]=\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}=\tilde{R} \tag{3.17}$$

The SOE Kalman filter expands the nonlinear function around the last estimate of the state vector $w_{MV}(k,w)$ by using the second-order Taylor series and finds out the unbiased estimate $\hat{w}_{MV}(k,w)$ to minimize the variance of estimation error which is presented as below

$$\text{MSE}=E\left[\left|w_{MV}(k,w)-\hat{w}_{MV}(k,w)\right|^2\right] \tag{3.18}$$

the estimated weight vector $w_{MV}(k,w)$ can be found by evaluating the first derivative of $f(w_{MV}(k,w))$ which is denoted as Jacobian $F_w(k,w)$ and the second derivative of $f(w_{MV}(k,w))$ which is denoted as the Hessian matrices $F_{ww}^{(1)}(\text{w})$ and $F_{ww}^{(2)}(\text{w})$ below

$$\begin{aligned}F_w(\text{k},\text{w}) &= \left\{\nabla_{w_{MV}}\boldsymbol{f}^T(w_{MV}(k,w))\right\}^T\\ &=\begin{bmatrix} X^H(\text{k},\text{w})\\ \varepsilon^2 w_{MV}^H(k,w)-(a(\theta_s,w)a^H(\theta_s,w)w_{MV}(k,w))^H+a^H(\theta_s,w)\end{bmatrix}\end{aligned} \tag{3.19}$$

$$F_{ww}^{(1)}(\text{w})=\nabla_{w_{MV}}\nabla_{w_{MV}}^H\left\{X^H(\text{k},\text{w})w_{MV}(k,w)\right\}=0 \tag{3.20}$$

$$F_{ww}^{(2)}(\text{w})=\nabla_{w_{MV}}\nabla_{w_{MV}}^H\left\{f_2(w_{MV}(k,w))\right\}=\varepsilon^2 I-a(\theta_s,w)a^H(\theta_s,w) \tag{3.21}$$

where the I is the identity matrix. The SOE Kalman filter solution is given by [18], [21] based on the state space model (3.15) and (3.16)

$$w_{MV}(k,w)=w_{MV}(k-1,w)+\tilde{K}(k,w)\left[\bar{y}-\boldsymbol{f}(w_{MV}(k,w))\right]+\boldsymbol{\pi}(k,w) \tag{3.22}$$

where

$$\boldsymbol{\pi}(k,w)=\frac{1}{2}\tilde{\text{K}}(k,w)\cdot\begin{bmatrix}0\\1\end{bmatrix}\cdot tr\left\{F_{ww}^{(2)}(w)\tilde{P}'^{-}(k,w)\right\} \tag{3.23}$$

is the correction term to make the state estimate unbiased. The filter gain and the predicted weight error covariance matrices are given by

$$\tilde{K}(k,w) = \tilde{P}^{-}(k,w)F_w^H(\mathrm{k},\mathrm{w})\left[F_w(\mathrm{k},\mathrm{w})\tilde{P}^{-}(k,w)F_w^H(\mathrm{k},\mathrm{w}) + \Lambda(\mathrm{k},\mathrm{w}) + \tilde{R}\right]^{-1} \quad (3.24)$$

$$\tilde{P}^{-}(k,w) = \tilde{P}^{+}(k-1,w) + \tilde{Q} \quad (3.25)$$

$$\tilde{P}^{+}(k,w) = \left[I - \tilde{K}(k,w)F_w(\mathrm{k},\mathrm{w})\right]\tilde{P}^{-}(k,w) \quad (3.26)$$

where

$$\Lambda(\mathrm{k},\mathrm{w}) = \frac{1}{2}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} tr\left\{F_{ww}^{(2)}(\mathrm{w})\tilde{P}^{-}(k,w)F_{ww}^{(2)}(\mathrm{w})\tilde{P}^{-}(k,w)\right\} \quad (3.27)$$

is the correction term deduced from the $\pi(k,w)$; $\tilde{K}(k,w)$ is the Kalman gain; $\tilde{P}^{-}(k,w)$ is the a priori error covariance matrix and $\tilde{P}^{+}(k,w)$ is the a posteriori error covariance matrix. After some algebra operation the Kalman gain and the covariance matrices can be rewritten as below

$$\tilde{K}(k,w) = \tilde{P}^{-}(k,w)\left[I + F_w^H(\mathrm{k},\mathrm{w})(\Lambda(\mathrm{k},\mathrm{w}) + \tilde{R})^{-1} \times F_w(\mathrm{k},\mathrm{w})\tilde{P}^{-}(k,w)\right]^{-1} F_w^H(\mathrm{k},\mathrm{w})(\Lambda(\mathrm{k},\mathrm{w}) + \tilde{R})^{-1} \quad (3.28)$$

$$\tilde{P}^{-}(k,w) = \tilde{P}^{-}(k-1,w)\left[I + F_w^H(\mathrm{k}-1,\mathrm{w})(\Lambda(\mathrm{k}-1,\mathrm{w}) + \tilde{R})^{-1} \times F_w(\mathrm{k}-1,\mathrm{w})\tilde{P}^{-}(k-1,w)\right]^{-1} + \tilde{Q} \quad (3.29)$$

Algorithm: Second-Order Extended Kalman filter

State Equation:

$$\hat{w}(k+1|k,w) = \hat{w}(k|k,w) + \tilde{\mathbf{Q}}(k,w)$$

Measurement Equation (Cost Equation):

$$\mathbf{Y}(k,w) = \begin{bmatrix} 0 \\ f(\theta_s) = 1 \end{bmatrix} = \begin{bmatrix} X^H(k,w) w_{MV}(k,w) \\ f_2(w_{MV}(k,w)) \end{bmatrix} + \mathbf{V}(k,w)$$

Computation for $k = 1,2,\cdots$

$$S_k = H_w^H P_k^- H_w^H + \frac{1}{2}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} tr\left\{H_{ww}^{(2)} P_k^- H_{ww}^{(2)} P_k^-\right\} + \tilde{R}$$

Update the state estimate :

$$K_k = P_k^- H_w^H S_k^{-1}$$

$$\hat{z}_k^- = \begin{bmatrix} \gamma X_k^H \hat{w}_k^- \\ h_2(\gamma \hat{w}_k^-) + \frac{1}{2} tr\left\{H_{ww}^{(2)} P_k^-\right\} \end{bmatrix}$$

$$\hat{w}_k = \hat{w}_k^- + K_k(z_k - \hat{z}_k^-)$$

$$H_w(\mathrm{w}_k) = \left\{\nabla_w h^T(\mathrm{w}_k)\right\}^T = \begin{bmatrix} X_k^H \\ \varepsilon^2 w_k^H - (\mathrm{CC}^H \mathrm{w}_k)^H + C^H \end{bmatrix}$$

$$\text{Jacobian}: H_{ww}^{(1)} = \nabla_w \nabla_w^H \left\{X_k^H w_k\right\} = 0$$

$$\text{Hessian}: \quad H_{ww}^{(2)} = \nabla_w \nabla_w^H \left\{h_2(w_k)\right\} = \varepsilon^2 I - \mathrm{CC}^H$$

Update the error covariance :

$$P_k = (I - K_k H_w) P_k^- (I - K_k H_w)^H + K_k \tilde{R} K_k^H$$

Table 3. Second-Order Extended Kalman filter algorithm.

In Section 3.4, we will use the structure as Equation (3.17) $\tilde{Q}$ and $\tilde{R}$, where the parameters σ_s^2, σ_1^2, σ_2^2, and ε are discussed.

3.4 Parameter Selection and Tradeoff

In Section 3.3, the steering vector bound that discriminates the speech enhancement is not determined. In subsection 3.4.1 the parameters selection and the relationship between the error covariance matrices $\sigma_s^2, \sigma_1^2, \sigma_2^2$ of Kalman filter will be presented. In subsection 3.4.2 the steering vector bound ε selection in wideband concept are further discussed in different source direction mismatch conditions.

3.4.1 Covariance Matrix of Kalman Filter

Firstly, the selection of the parameters σ_s^2 and ξ are determined in the measurement equation and state equation rewritten as below:

State Equation

$$w_{MV}(k,w) = w_{MV}(k-1,w) + \mathbf{Q}(k,w) \tag{3.30}$$

Measurement Equation

$$\bar{y} = \begin{bmatrix} X^H(k,w) w_{MV}(k,w) \\ f_2(w_{MV}(k,w)) \end{bmatrix} + \boldsymbol{V}(k,w) = \boldsymbol{X}(k,w) + \boldsymbol{V}(k,w) \tag{3.31}$$

$$\boldsymbol{Q}(k,w) \sim N(0, \sigma_s^2 I)$$

$$\boldsymbol{V}(k,w) \sim N(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = N(0, \sigma_1^2 \begin{bmatrix} 1 & 0 \\ 0 & \xi \end{bmatrix}) \tag{3.32}$$

In the weighting matrix $\boldsymbol{Q}$, σ_s^2 controls the error covariance of the random walk for the weights updating in the process equation. Since the environment surrounded is assumed to be a Linearly Time-Invariant (LTI) system, σ_s^2 is set to zero. Next, σ_1^2 and σ_2^2 represent the variance of filtered output error and constraint error respectively in the measurement equation. Then the value $\frac{\sigma_s^2}{\sigma_1^2}$ which is the ratio between σ_s^2 and

σ_1^2, controls the weighting matrix adaption speed. If the $\frac{\sigma_s^2}{\sigma_1^2}$ is large, the filter adapts to the variation in environment faster. In (3.30), it can be observed that if σ_s^2 is large, the change between $w_{MV}(k,w)$ and $w_{MV}(k-1,w)$ will larger and leads to faster adaption in $w_{MV}(k,w)$. In (3.31), it can also be observed that if σ_1^2 is small, the measurement error $\boldsymbol{V}(k,w)$ has small variations between each step, which means $w_{MV}(k,w)$ has to adapt fast if $\boldsymbol{X}(k,w)$ varies fast.

In [18], the authors proposed that σ_1^2 should be chosen of the same order as the optimal output power of the array. It can be approximated as $\|w_{MV}\|^2(M\sigma^2+\sigma_n^2)$, where σ^2 and σ_n^2 are the desired signal and sensor noise power, respectively. The definition

$$\rho=\frac{\sigma_1^2}{M\sigma^2+\sigma_n^2} \tag{3.33}$$

is discussed with the steering vector bound in the subsequent examples. The latter σ_2^2 should be chosen small enough to satisfy the beam constraint robustness with a high accuracy. We choose the ratio $\xi=\frac{\sigma_2^2}{\sigma_1^2}$ as (3.32) for the following numerical and mathematical reasons, rather than setting the parameter σ_2^2.

1) ξ determines the condition number of the weighting matrix $\boldsymbol{V}(k,w)$.
2) ξ controls the tradeoff between noise reduction and dereverberation.
3) The constrained Kalman algorithm converges to optimum MVDR filter if ξ is small enough.

Hence, large ξ leads to strong noise reduction and little dereveberation while small ξ leads to strong dereveberation and little noise reduction. If ξ is small, that means the error variation in the lower line of (3.32) is relatively small compared with the upper line, which leads to closer tracing in the lower line and looser tracing in the upper line, achieving strong dereveberation and weak noise reduction. Then $\xi < 10^{-8}$ is recommended to fulfill the requirements of the signal mismatch problem. (note that ξ should be set small enough to keep the matrix $\boldsymbol{V}(k,w)$ well-conditioned).

3.4.2 Steering Vector Bound Wideband Selection

Second, the selection of the steering vector bound ε of the SOE Kalman filter in different frequency bands with different array number are studied. We use the same parameters of the Kalman filter and choose the $\sigma_1^2 = 50(M\sigma^2 + \sigma_n^2)$, which the output SINR of Kalman filter remains close to the optimal in a wide range of the values of ρ proven in [18][21]. Then [18][21] choose the fixed steering vector bound ε in narrow band approach to ensure the spatial characteristics, and the performances are discussed under the frequency whose wavelength is chosen to be twice as the microphones' spacing(i.e., f = c/2d, where c is the sound velocity). Different to that, we design different steering vector bound according to different subbands to keep outputSINR higher while maintaining the distortionless level in the wideband analysis.

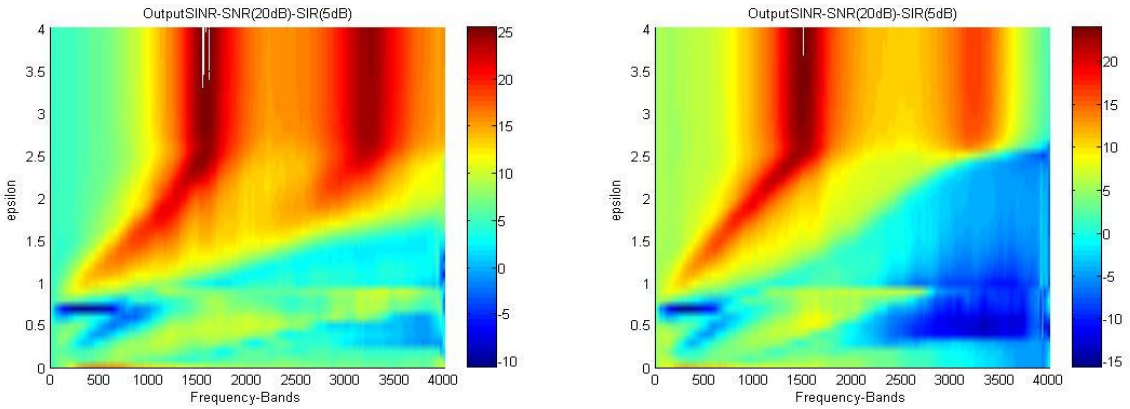


Fig. 6 (a) Output SINR of wideband epsilon selection $\Delta\theta = 8^o$(b) $\Delta\theta = 16^o$(SOE-KF).

Fig. 6 shows an example when the signal mismatch becomes severe when $\Delta\theta = 8^o$ and 16^o with six omnidirectional microphones. It can be observed that the characteristics of the OutputSINR maintain higher in the lower frequency-bands when the steering vector bound tends to be smaller. Although there is a wide range of fixed epsilon selection in higher frequency-bands, whose wavelength is chosen to be twice as the microphones' spacing, the performance of narrowband selection is not good in speech enhancement as a whole. Compared to the narrowband epsilon selection, if there are steering vector mismatches, the result of wideband selection is closer to our destination obviously. Despite the fact that the conceptual relationship can be imagined, there is still no concrete equation to describe the relationship between them. Hence we propose to choose the steering vector bound functionally by the sigmoid function which is a mathematical function used in modeling system as below.

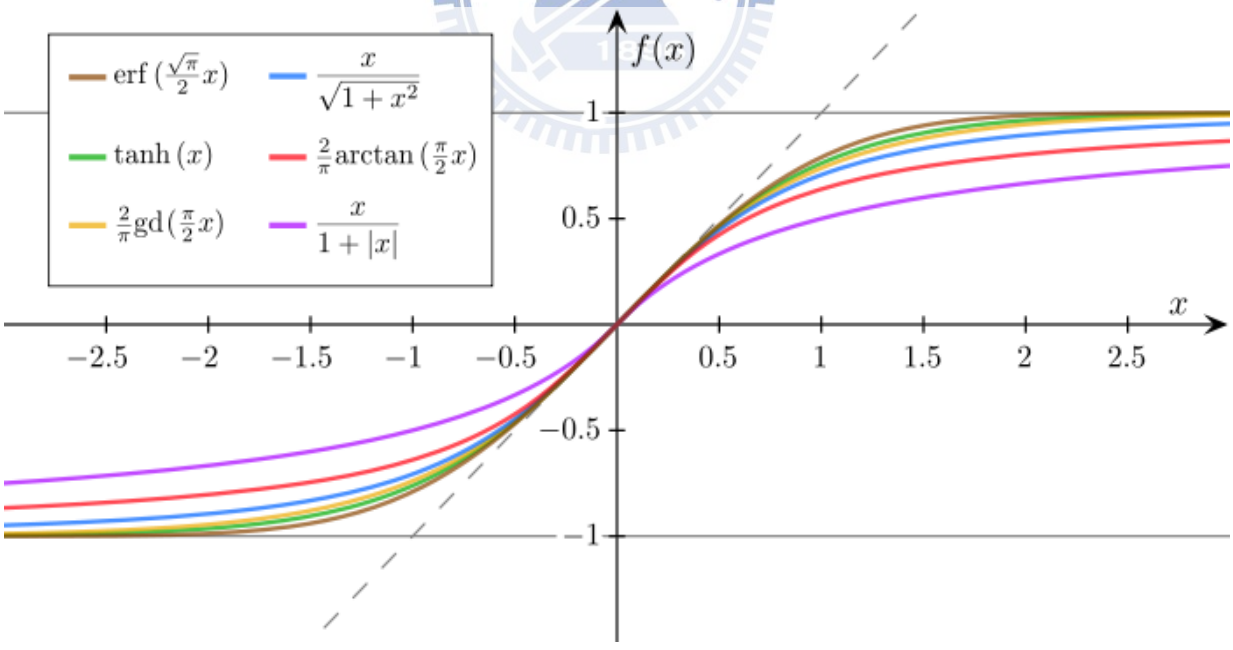
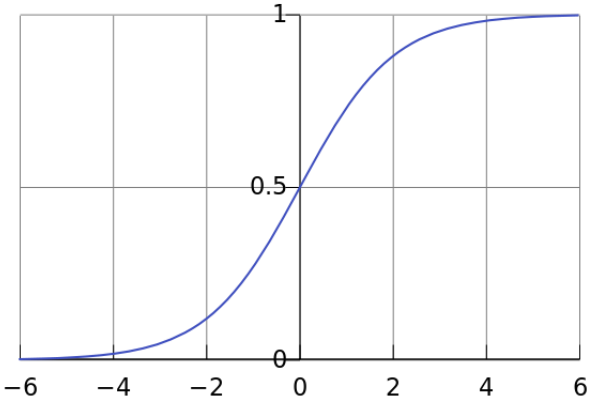


Fig. 7 The Sigmoid functions compared.

In the Fig. 7, all of the sigmoid functions are normalized in such a way that their slope at origin is 1. Sigmoid functions that refer to the special case of the logistic function shown on the Fig. 8 are mathematical function having an "S" shape (sigmoid curve). Besides the logistic function, sigmoid functions include the ordinary **arctangent**, the **hyperbolic tangent**, and the **error function**, but also the **generalized logistic function** and **algebraic functions** in Fig. 7. The initial stage of growth is approximately exponential; then, as saturation begins, the growth slows, and at maturity, growth stops. For values of x in the range of real positive numbers from 0 to $+\infty$, the tendency is similar to the Fig. 6. Hence we use the sigmoid function with equation as below:

$$f(x) = \frac{1}{1+e^{(-a*(x-d))}} \begin{cases} d : shift \\ a : slope \end{cases} \tag{3.34}$$

where e is the Euler's number. In practice, due to the nature of the exponential function e^{-x}, it is sufficient to compute $f(x)$ over a wide range of real numbers with corresponding slope. In Fig. 9, the pre-training procedure changes a from 0.01 to 0.1 with summation of 0.01 and finds the appropriate slope a .The output signal to interference-plus-noise ratio (OutputSINR) and log spectrum distortion (LSD) are used to be the criterions of selection in Fig. 10 and Fig. 11 respectively.



$$f(x) = \frac{1}{1+e^{-x}}$$

$$\frac{d}{dx} f(x) = f(x)(1-f(x))$$

It has the property that

$$1-f(x) = f(-x)$$

Thus, the function is **odd**

Fig. 8 Standard logistic sigmoid function.

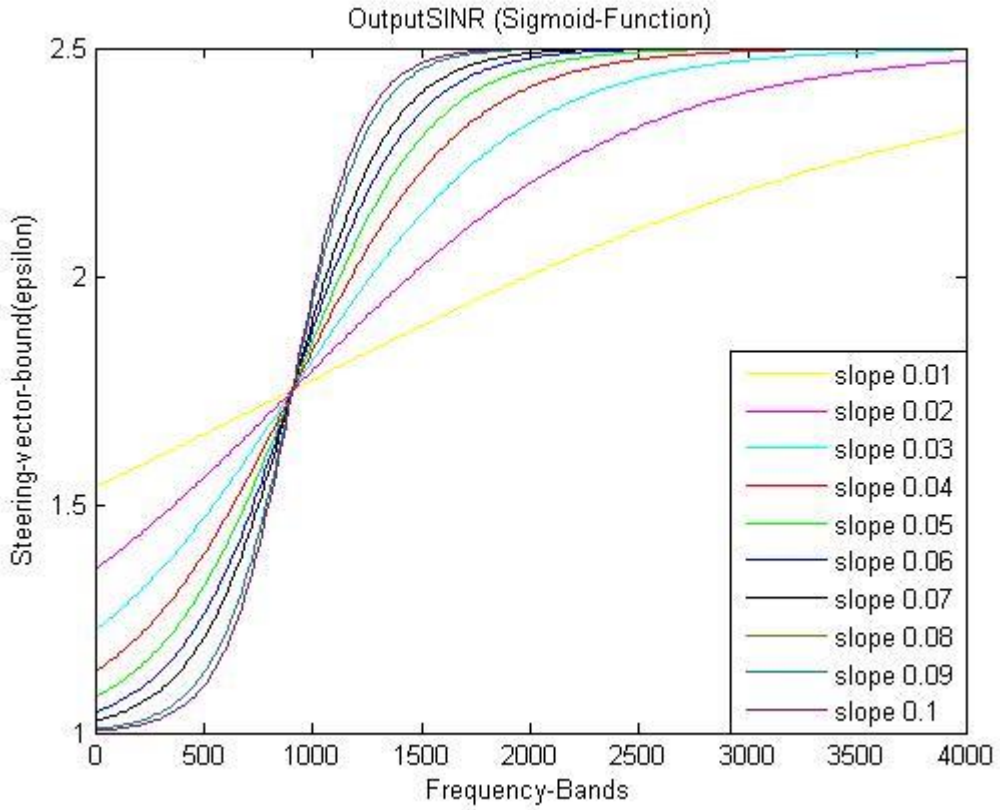


Fig. 9 The Sigmoid function of different slope.

The goal of wideband selection is to maintain the higher OutputSINR and distortionless when there are source direction mismatches to the assumption of algorithms. The relationship between them is mutually exclusive. The higher outputSINR leads to higher distortion and vice versa. In Fig. 10 and Fig. 11, the variation of OutputSINR in different signal mismatches has significant differences. The strategy of determination is based on higher OutputSINR to choose the distortionless response. The best sigmoid function slope is chosen as $a = 0.05$ and the maximum and minimum bound of sigmoid-function are 2.5 and 1.0 respectively according to the Fig. 6.

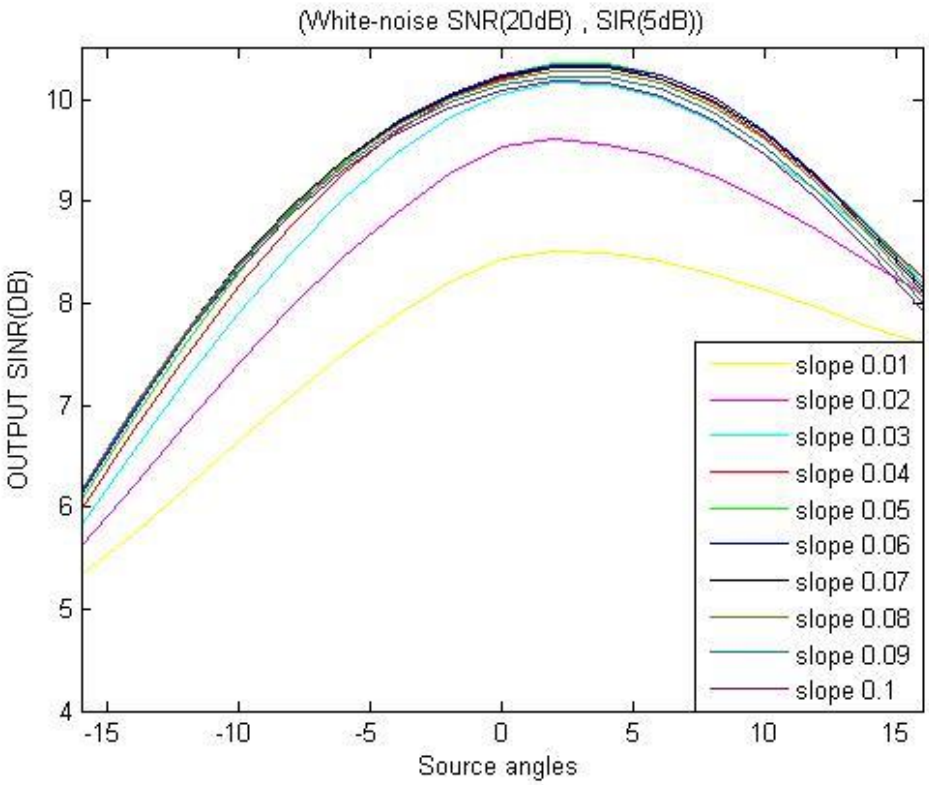


Fig. 10 The OutputSINR of different slopes.

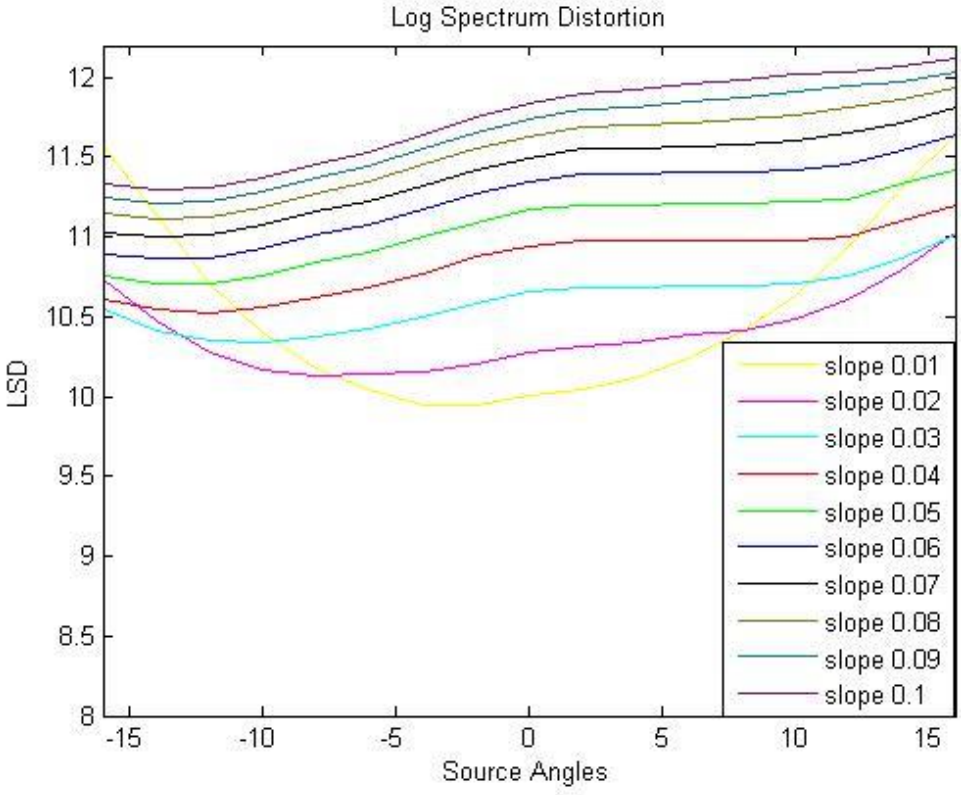
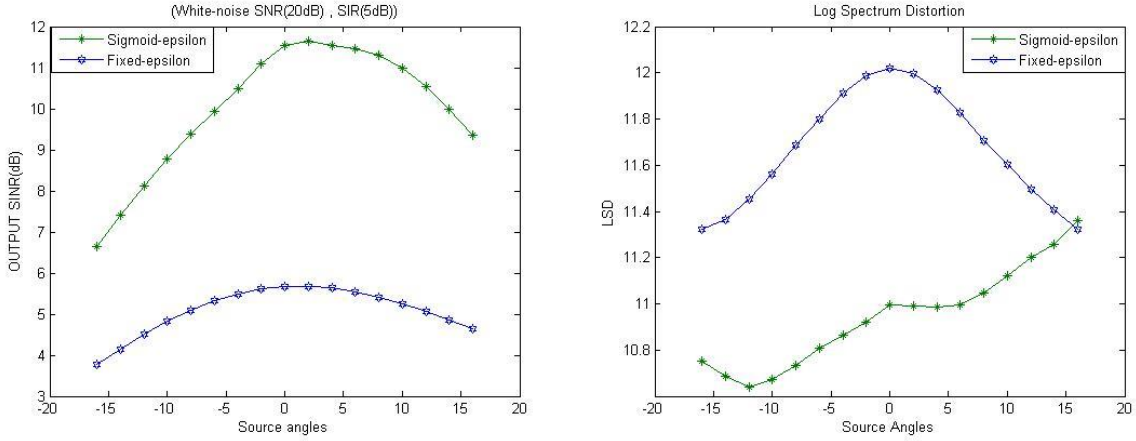


Fig. 11 The Log Spectrum Distortion of different slopes.

Using the appropriate selection of sigmoid function slope in Fig. 10-11 to compare the effect of wideband epsilon selection shows the significant difference. In Fig. 12, the proposed steering vector bound leads to not only higher outputSINR but also the lower LSD when the mismatch becomes severe with M=6. The result is the goal we want to achieve. In the next Section, we will introduce constraints projected method to combine with the proposed steering vector bound selection.



(a) OutputSINR (b) LSD

Fig. 12 Comparison between wideband and narrowband epsilon selection.

3.5 Speech Enhancement under Proposed Constrained Formulation

In beamforming, we estimate the signal of interest arriving from some specific directions in the presence of noise and interference signals with the aid of an array of sensors. These sensors are located at different spatial positions and sample the propagating waves in space. The specific spatial response of the array system is achieved with “beams” pointing to the desired signals and ‘nulls’ towards the interfering ones. Ideally, for beamforming, we aim to formulate a fixed response to the signal of interest and zero response to the interfering signals. Note for simplicity, we do not consider the effect of noise in this Section. This requirement can be expressed with array manifold vector as the following matrix equation:

$$\begin{pmatrix} 1 & e^{-jw\tau_1(\theta_0)} & \cdots & e^{-jw\tau_{M-1}(\theta_0)} \\ 1 & e^{-jw\tau_1(\theta_1)} & \cdots & e^{-jw\tau_{M-1}(\theta_1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-jw\tau_1(\theta_{M-1})} & \cdots & e^{-jw\tau_{M-1}(\theta_{M-1})} \end{pmatrix} \begin{pmatrix} w_0^* \\ w_1^* \\ \vdots \\ w_{M-1}^* \end{pmatrix} = \begin{pmatrix} constant \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{3.35}$$

Obviously, as long as the matrix on the left has full rank, we can always find a set of array weights to cancel the M-1 interfering signals and the exact value of the weights for complete cancellation of the interfering signals is dependent on the signal frequency (certainly also on their directions of arrival).

For wideband signals, since each of them consists of infinite number of different frequency components, the value of the weights should be different in different frequencies. That is the reason that narrowband beamforming structure with a single constant for each received signal from arrays will not work effectively in a wideband environment. In subsection 3.5.1, the directive interfering signal tracking when only noises are present can yield the direction of interference and the corresponding beamformer null response value. In subsection 3.5.2, using the information from the subsection 3.5.1 as constraints, which are combined with the proposed second-order extended Kalman filter to improve the performance of speech enhancement, compared to the existing algorithms is shown.

3.5.1 Beamformer Null Tracking when Source is Absent

Firstly, we make the noise tracking problem to be the beamformer null tracking of the adaptive beamformer when the source is not present. The weightings of spatial filter in different frequency-bands can be expressed as beampattern gain according to the steering array manifold vector. Comparison between the two algorithms (HC-KF and SOE-KF) in beamformer null tracking ability with the same conditions is shown in Fig. 13 and Fig. 14. In Fig. 15 there are two interferences at 30^o SIR = 5(dB) and

60^o SIR = 0(dB). In Fig. 16, two interferences at 50^o SIR = 5(dB) and - 50^o SIR = 5(dB) in Fig. 14 when desired source is at 0^o. In the low frequency, the HC-KF algorithm tracking is better than SOE-KF. Due to the desired source is not present, the algorithm focus on the noises reduction and no need to put more emphasis on the source direction distortionless. On the other hand, there is no trade-off between the covariance matrices which is discussed in the subsection 3.4.1. However, the beam directivity of the both algorithms degrades significantly in the low-frequency. For reducing the number of constraints to increase the degrees of freedom of the beamformer weightings for interferences suppression, we propose the HC-KF to do beamformer null tracking given the information of voice activity detection.

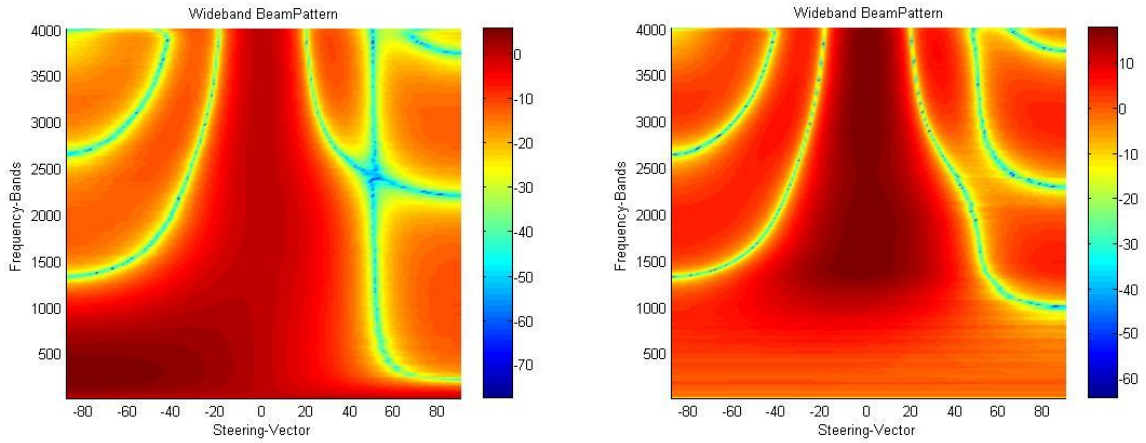


Fig. 13 (a) Wideband beampattern gain of one-interference 50^o(HC-KF)(b) (SOE-KF).

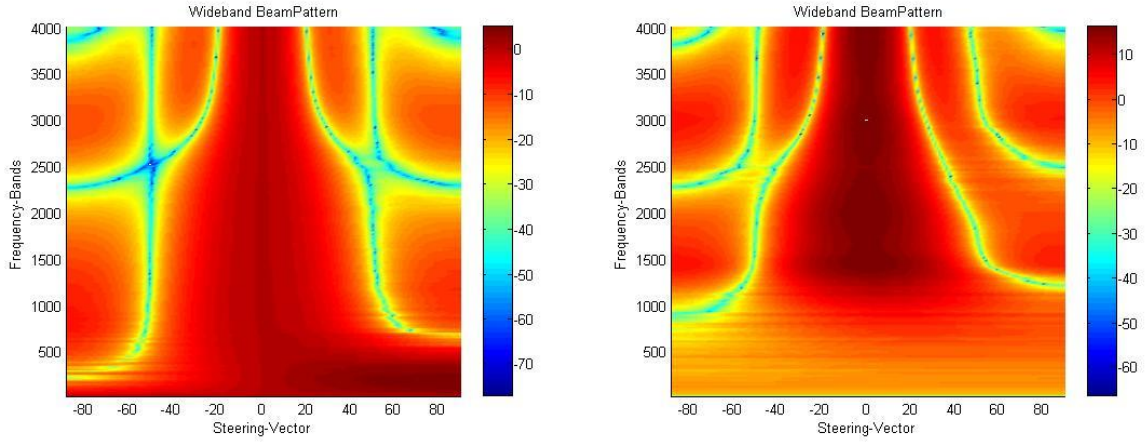


Fig. 14 (a) Wideband beampattern gain of two-interferences $50^o, -50^o$(HC-KF)(b) (SOE-KF).

Parametric methods for sources localization search physical space for local maximum(s) of a "point-based" measure of activity, which is estimated for each point of space, from the recorded multichannel signals. On the contrary, [32] proposes a Sector-based Activity Measure (SAM) method that partitions the space into sectors, and defines an activity measure for each sector. Each SAM method defines an activity function $A_k(X_1,\ldots,X_M)\in\mathbb{R}$ for each sector $\breve{s}=1\ldots N_{bins}$. The higher this value is, the more likely the sector to contain active source, which can be used to take a hard decision, for example by applying a threshold on A_k. And [33] uses the sparsity assumption, similar to the one in (Roweis, 2003), which implies that all other sectors $S_{\breve{s}}, \breve{s}\neq \breve{s}_{\min}(k)$ are attributed to a zero posterior probability of containing acoustic activity at the discrete frequency k. The posterior probability of having at least one active source in sector $S_{\breve{s}_{\min}(k)}$ and at frequency k is modeled as:

$$P(\text{sector } S_{\breve{s}} \text{ active at discrete frequency } k \mid u^{(t)}(k)) = \delta_{Kr}(\breve{s}-\breve{s}_{\min}(k)) \tag{3.36}$$

$$\breve{s}_{\min}(k) = \arg_{\breve{s}} \min \overline{D}_{\breve{s}}^{(t)}(k) \tag{3.37}$$

where $\delta_{Kr}(\xi)$ is the Kronecker function, equal to 1 iff ξ=0, and zero otherwise; $\overline{D}_{\breve{s}}^{(t)}$ is defined as the function of standard comparison. Then, in order to measure the wideband acoustic activity within each sector of space, [33] proposes the SAM-SPARSE-MEAN method. For a given sector frame and a given frequency bin, it is the average probability where the sector is dominant. Similar to the concept and according to the adaptive beamforming result in Fig. 13-14, the space around a uniform linear microphone array would be partitioned to sectors S_t:

$$S_t = \left\{(\theta,k)\in\Re^2 \mid -\frac{\pi}{2}\le\theta\le\frac{\pi}{2}, 500\le k<2000(Hz)\right\} \tag{3.38}$$

where θ, k designate azimuth and discrete frequency-bands. The high-frequencies signals degrade significantly as the distance between source and arrays is farther, so choose the middle-frequencies bins in the visible region. Within each frequency bin k, only one sector $h_{\min}(k,\theta)$ is judged as active:

$$h_{\min}(k,\theta) \equiv \arg\min_{k}(P_{k,\theta}) \tag{3.39}$$

where $P_{k,\theta}(X(k))$ is the beamformer null activity function in different frequency bins.

$$P_{k,\theta} = \sum_{i=1}^{M} W_i \cdot e^{j(\frac{2\pi k}{c})\cdot(i-1)d\sin\theta} \tag{3.40}$$

where W_i is the weighting of adaptive spatial filter for speech enhancement. Hence, we could use the mean of the beamformer null probability to search which direction of interference is dominant the same as the Equation (3.36).

$$\zeta_{\bar{s},t} = \sum_{k_{\min}}^{k_{\max}} P(\text{sector } S_{\bar{s}} \text{ active at selected frequency } k \,|\, u^{(t)}(k)) \tag{3.41}$$

If there are more than one interference, we can take a hard decision by applying threshold on $h_{\min}(k,\theta)$ to do multiple interferences tracking.

In the Fig. 15-16, there are two blocks used to determine the best estimation of the direction of interference in two cases. The first block separates the beampattern into three specific sections as shown below. Mainlobe section is the possible direction range of desired source, which is not regarded as the discussed section. And then the left and right minimum beamformer null statistical probability analyses are individual. With the minimum point statistical analyses, the null degrees will determine whether there is directive interference or not through thresholds. After the calculation and numerical analysis, the direction of interferences can be found with the response values to be null-constraints in next subsection.

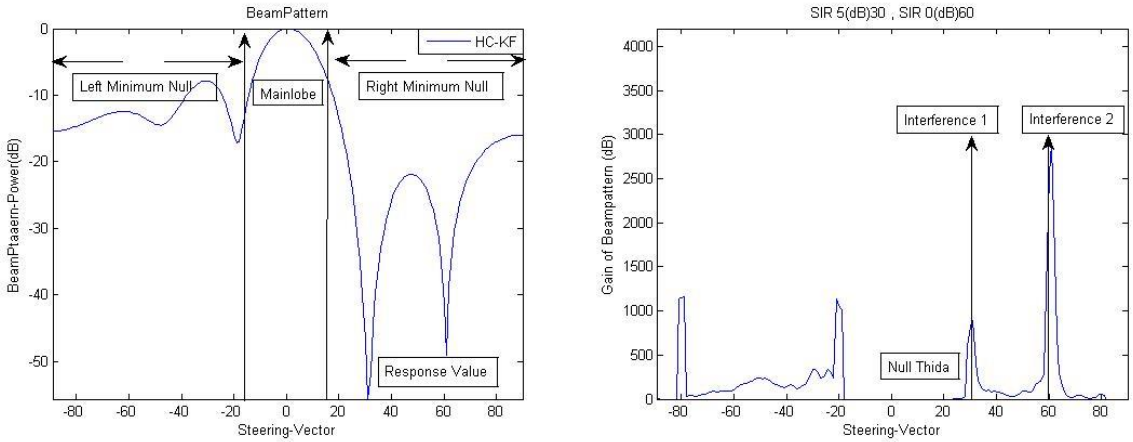


Fig. 15 Beampattern null tracking (two-interferences SIR 5(dB) 30^o, SIR 0(dB) 60^o).

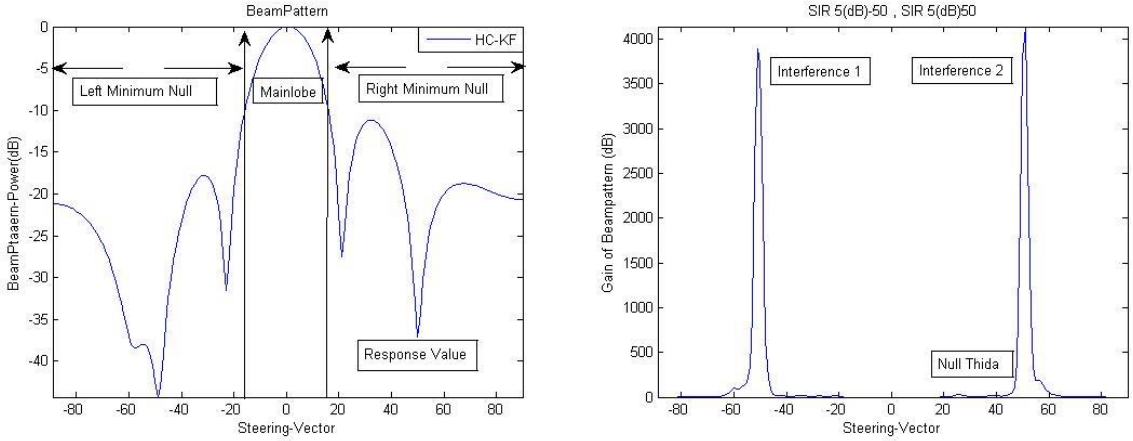


Fig. 16 Beampattern null tracking (two-interferences SIR 5(dB) -50^o, SIR 5(dB) 50^o).

Compared to the eigenspace-based beamformer, which is known as one of the most powerful robust techniques applicable to source tracking, the proposed method isn't limited to high input avgSINR cases. The disadvantages of eigenspace-based beamformer are that the estimation of the projection matrix onto the signal-plus-interference subspace breaks down due to subspace swaps greatly at low input avgSINR and the dimension of signal-plus-interference subspace are low and unknown exactly. In particularly, if the calibration of arrays and compensation of phase delay do well, the beamformer null tracking outperforms than other algorithms.

3.5.2 Beamformer Null Constraint when Source is Present

In this subsection, using the proposed SOE-KF method to handle the null constraints given with the information of voice activity detection in previous subsection has better performance in speech enhancement applications. In the subsection 3.5.1, we can yield the beamformer null thida θ_n and the corresponding response values $f(\theta_n)$ according to the weighting of constrained Kalman filter which is based on robust MVDR beamformer. The goal of this subsection is to use the direction of interferences and corresponding response values as constraints to be combined with the constraint projected method, which is discussed in subsection 2.3.2, when the source is not present.

$$\min_{w} w^H \hat{R} w \quad \text{subject to} \begin{Bmatrix} C_N^{'} \hat{w}_k = f(\theta_n) \\ C_N^{'} = a_{SV}^n(\theta_n) \end{Bmatrix} a_{SV}^n(\theta_n) = \begin{bmatrix} 1 & e^{jk_c d \sin\theta_n} & \cdots & e^{jk_c (M-1) d \sin\theta_n} \end{bmatrix} \quad (3.42)$$

The steering vector to interference direction is applied to be constraints. For SOE-KF method, if there are noisy interferences existing, there are beam and null constraints integrated into the algorithm which is the proposed method (HC-SOE-KF). The state and error covariance matrix update equation of proposed method can be expressed briefly as below:

$$\hat{w}_k = \hat{w}_k^- + K_k (z_k - \hat{z}_k^-)$$

$$\hat{w}_k = \hat{w}_k - C_N^{'} (C_N C_N^{'})^{-1} (C_N \hat{w}_k - f(\theta_n)) \quad (3.43)$$

$$P_k = (I - K_k H_w) P_k^- (I - K_k H_w)^H + K_k R K_k^H$$

$$P_k = (\mathrm{I} - \mathrm{C}_N^{'} (\mathrm{C}_N \mathrm{C}_N^{'})^{-1} \mathrm{C}_N^{'}) \mathrm{P}_k (\mathrm{I} - \mathrm{C}_N^{'} (\mathrm{C}_N \mathrm{C}_N^{'})^{-1} \mathrm{C}_N^{'})^{'} \quad (3.44)$$

$$C_N^{'} = a_{SV}^n(\theta_n) \quad (3.45)$$

In Equation (3.43) and (3.44), the close loop updating form of Kalman filter is based on the optimal langrage multipliers solution. Although the proposed method can suppress the interference directively, the number of the null constraints is limited to the degrees of freedom of the beamformer weighting and the characteristic of spatial coherence. In Fig. 17 and Fig. 18, the OutputSINR of HC-SOE-KF shows the relationship between the steering vector bound and the frequency-bands in different mismatch conditions. Due to the wideband beamforming concept, the source and noises of the training data are white noises at direction $0^o$and 50^o respectively and the signal to interference ratio is 5(dB); signal to noise ratio is 20(dB). Obviously, there is enormous degradation in the low frequencies under 1K (Hz) approximately and different phenomenon in different microphone number. It is important to mention that no matter the number of arrays, smaller steering vector bound will have better performance in lower frequency-bands with null constraint. In array processing, the spectrum coherence influences the tracking of interferences direction and the increase in signal bandwidth will result in the degradation of steered response. In Fig. 19, the peaks of the OutputSINR corresponding to the epsilons selection in frequency-bands is shown. It can be observed that for the higher OutputSINR, the much lower epsilon in low frequency-bands. Hence the regulation of steering vector bound is in need to compensate such phenomenon of signal self-cancellation in low frequency-bands.

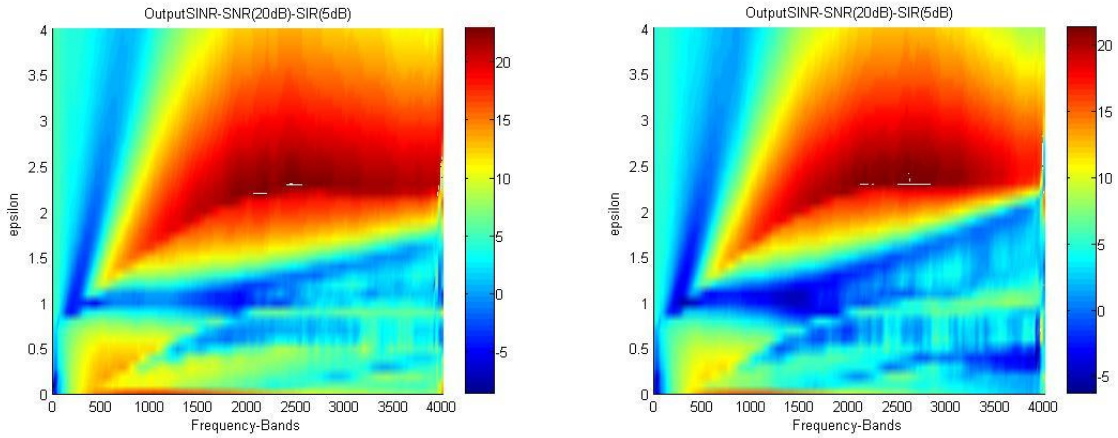


Fig. 17 (a) Output SINR of null-constraint wideband epsilon selection $\Delta\theta = 8^o$(b) $\Delta\theta = 16^o$(HC-SOE-KF M=4).

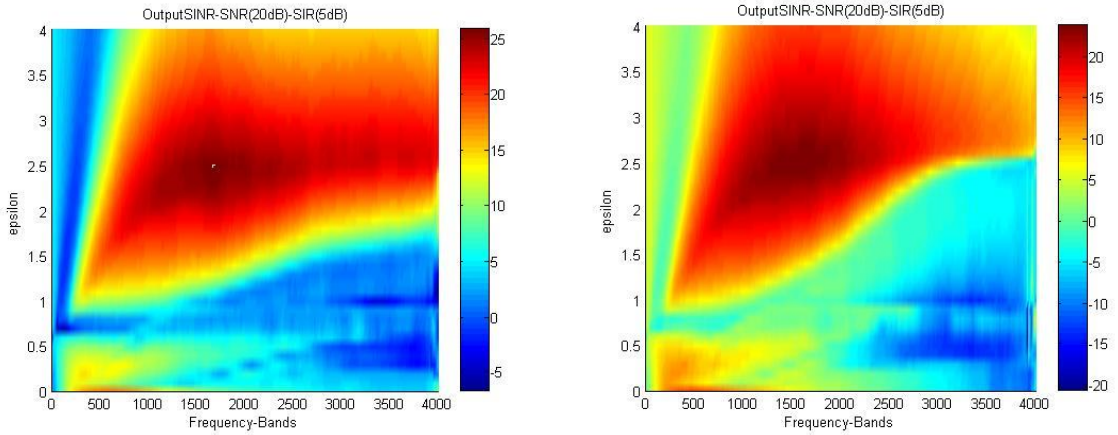


Fig. 18 (a) Output SINR of null-constraint wideband epsilon selection $\Delta\theta = 8^o$(b) $\Delta\theta = 16^o$(HC-SOE-KF M=6).

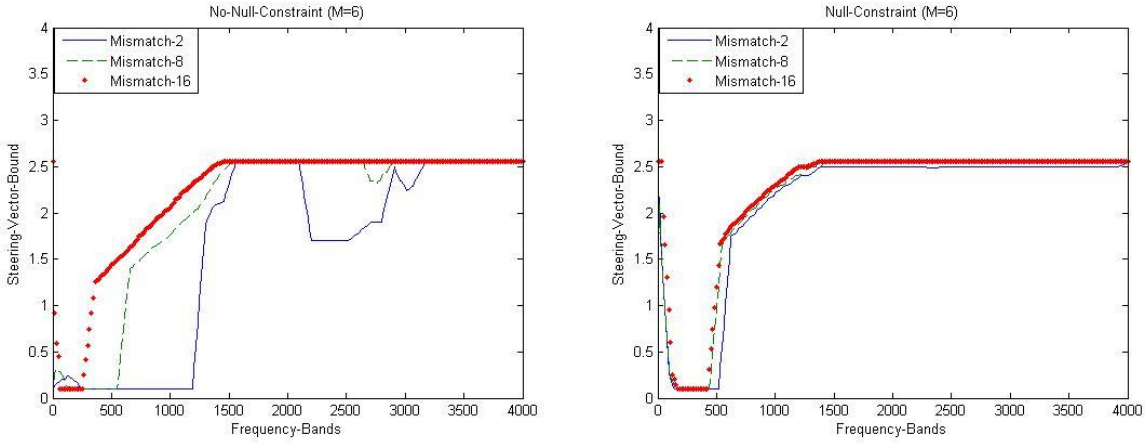


Fig. 19 (a) Peaks of OutputSINR with epsilon selection (no null-constraint) (b) null-constraint.

Algorithm: Second-Order Extended Constrained Kalman filter

State Equation:

$$\hat{w}(k+1|k,w) = \hat{w}(k|k,w) + \mathbf{Q}(k,w)$$

Measurement Equation (Cost Equation):

$$\mathbf{Y}(k,w) = \begin{bmatrix} 0 \\ f(\theta_s) = 1 \end{bmatrix} = \begin{bmatrix} X^H(k,w) w_{MV}(k,w) \\ f_2(w_{MV}(k,w)) \end{bmatrix} + \mathbf{V}(k,w)$$

Computation for $k = 1,2,\cdots$

$$S_k = H_w^H P_k^- H_w^H + \frac{1}{2}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} tr\left\{ H_{ww}^{(2)} P_k^- H_{ww}^{(2)} P_k^- \right\} + R$$

Update the state estimate :

$$K_k = P_k^- H_w^H S_k^{-1}$$

$$\hat{z}_k^- = \begin{bmatrix} \gamma X_k^H \hat{w}_k^- \\ h_2(\gamma \hat{w}_k^-) + \frac{1}{2} tr\left\{ H_{ww}^{(2)} P_k^- \right\} \end{bmatrix}$$

$$\hat{w}_k = \hat{w}_k^- + K_k(z_k - \hat{z}_k^-)$$

$$\hat{w}_k = \hat{w}_k - C_N^{'}(\mathrm{C}_N C_N^{'})^{-1}(\mathrm{C}_N \hat{w}_k - f(\theta_n))$$

$$H_w(\mathrm{w}_k) = \left\{ \nabla_w h^T(\mathrm{w}_k) \right\}^T = \begin{bmatrix} X_k^H \\ \varepsilon^2 w_k^H - (\mathrm{CC}^H \mathrm{w}_k)^H + C^H \end{bmatrix}$$

$$\text{Jacobian}: H_{ww}^{(1)} = \nabla_w \nabla_w^H \left\{ X_k^H w_k \right\} = 0$$

$$\text{Hessian}: \quad H_{ww}^{(2)} = \nabla_w \nabla_w^H \left\{ h_2(w_k) \right\} = \varepsilon^2 I - \mathrm{CC}^H$$

$$C_N^{'} = a_{SV}^n(\theta_n)$$

Update the error covariance :

$$P_k = (I - K_k H_w) P_k^- (I - K_k H_w)^H + K_k R K_k^H$$

$$P_k = (I - C_N^{'}(\mathrm{C}_N C_N^{'})^{-1} C_N^{'}) P_k (I - C_N^{'}(\mathrm{C}_N C_N^{'})^{-1} C_N^{'})$$

Table 4. Second-Order Extended Constrained Kalman filter algorithm.

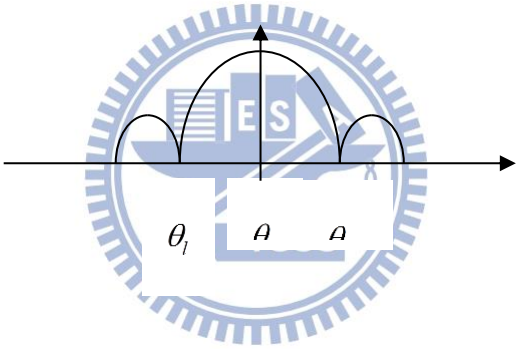
In Fig. 20, to illustrate two points, consider the beampattern near the origin whose period is $\frac{\lambda_c}{d}$ just same as the concept of "grating lobe". Grating lobe is a lobe of the same height as the mainlobe in array processing. We refer to Δu_2 as the null-to-null beamwidth and denote it by BW_{NN}. One-half of the BW_{NN} is the distance to the first null ($0.5BW_{NN}$ half-power beamwidth (HPBW)).

$$\sin\left(\frac{k_c Nd}{2}(\sin\theta - \sin\theta_0)\right) = 0$$

$$\because \theta_r - \theta_0 = \theta_0 - \theta_l = \Delta\theta$$

$$\Rightarrow BW_{NN} = \sin\theta_r - \sin\theta_l = \frac{2\lambda_c}{Nd} \approx 2\Delta\theta\cos\theta_0 \qquad (3.46)$$

$$\Rightarrow \Delta\theta \approx \frac{\lambda_c}{Nd\cos\theta_0}$$



The wavenumber k_c can be represented as following.

$$k_c = \frac{w}{c} = \frac{2\pi}{\lambda_c}, w = 2\pi f \qquad \begin{array}{l} d \text{ : array distance} \\ N \text{ : array number} \\ c \text{ : voice speed} \end{array} \qquad (3.47)$$

From Equation (3.46), beamwidth is directly proportional to $\frac{\lambda_c}{Nd}$. Thus, the beamwidth will be narrower when the number of array elements is larger or the distance between them is too far. That is to say that the spatial resolution of array is higher. We can enhance the resolution of an array system form by the increase of array number or array dimension; then another effect of beamwidth (Beam Broadening Effect). The effect is that the beamwidth will increase as the direction of beam tends to

be bigger. In Equation (3.46), beamwidth increases when $\theta_0 \uparrow \Rightarrow \cos(\theta_0) \downarrow \Rightarrow \Delta\theta \downarrow$.

In order to find out the frequency-bands corresponding to the first null-to-null bandwidth in different array number, we use the delay and sum (DAS) beamformer and Fig.19 (b) as the regulation standard of steering vector bound with beamformer-null constraint. The beampattern is shown in Fig. 21.

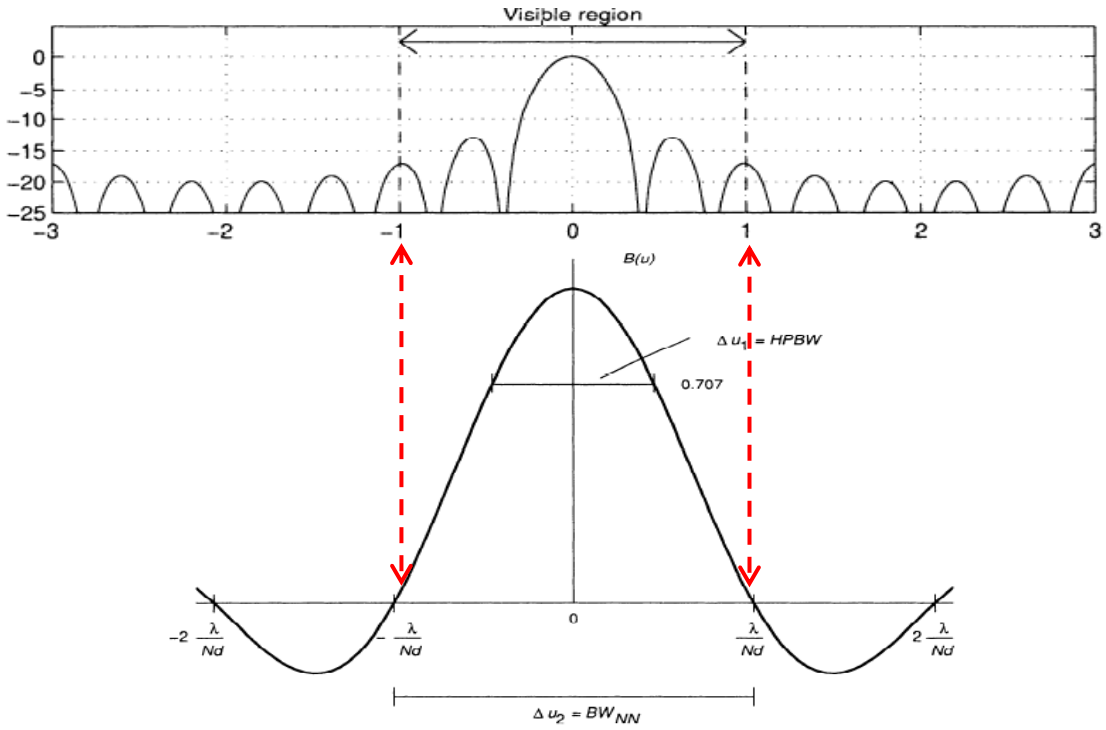


Fig. 20 Visible mainlobe of beampattern and bandwidth.

In Fig. 21, the steered response degrades with the decrease in frequency-bands. In high frequency whose wavelength is chosen to be the twice as the microphones' spacing, the bandwidth of the impinging signals should be narrow enough to make sure the signals received by the opposite ends of the array are still correlated with each other which is termed a narrowband beamformer. In other words, the steering vector bound should be wide enough to avoid the signal self-cancellation when there is mismatch existing. However when the bandwidth of the impinging signal is too wide to lose the spatial directivity, the steering vector bound should be narrow enough to

keep the beam directivity. Thereafter the use of the functional steering vector bound selection which is based on the wideband concept, if there is in need to use the null constraints for suppressing the interferences, choosing the smaller one as same as the Fig. 19(b) below the first null-to-null frequency-bands to strengthen the performance of speech enhancement from the DAS beamformer.

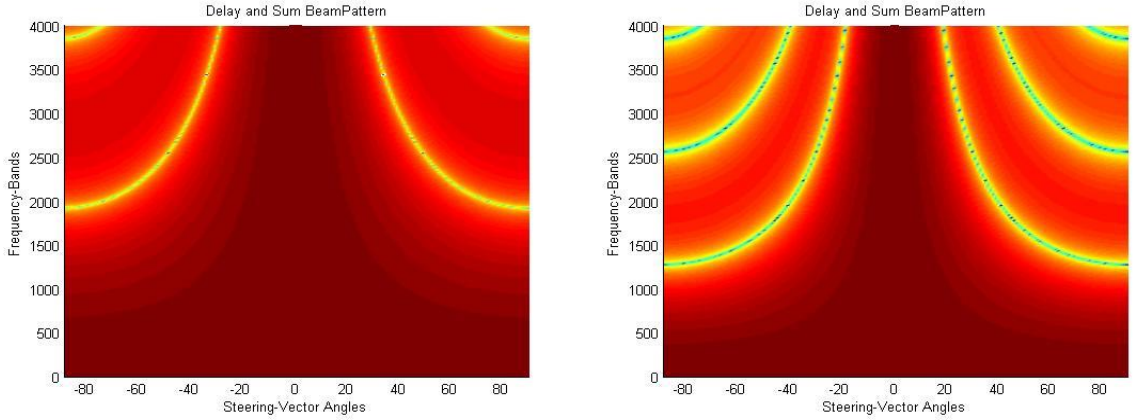


Fig. 21 (a) Delay and Sum (DAS) beamformer wideband beampattern M=4 (b) M=6.

Dramatically degradation on the low frequency-bands is shown in Fig. 22 which is compared to the Fig. 23. After the approximate epsilon regulation, there is a great improvement in the result of speech enhancement. In other words, the proposed method can achieve the noise reduction and solve the signal mismatch problem at meanwhile in the wideband concept. The results of filtered waveform are presented in the Fig. 24-25.

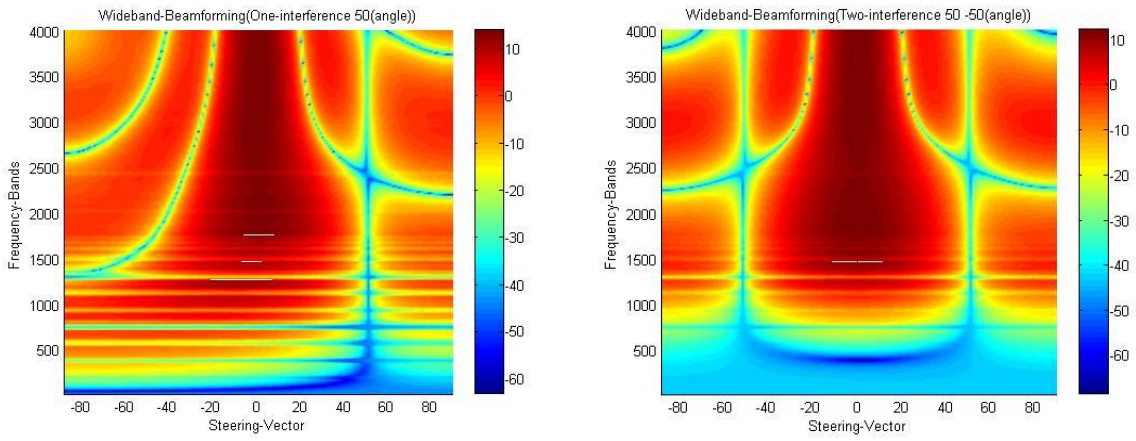


Fig. 22 (a) Narrowband epsilon selection of one-interference 50^o (b) two-interference $50^o - 50^o$ (HC-SOE-KF).

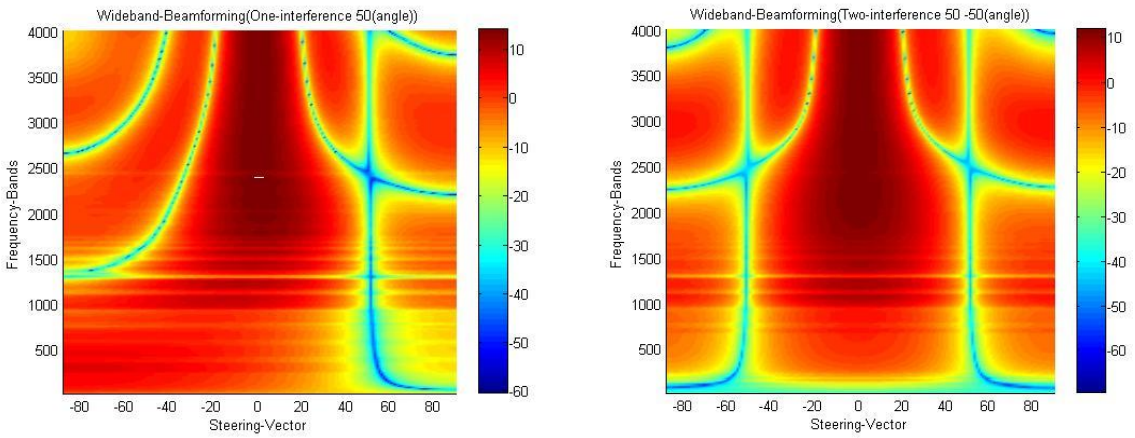


Fig. 23 (a) Wideband epsilon selection of one-interference 50^o (b) two-interference$50^o, -50^o$ (HC-SOE-KF).

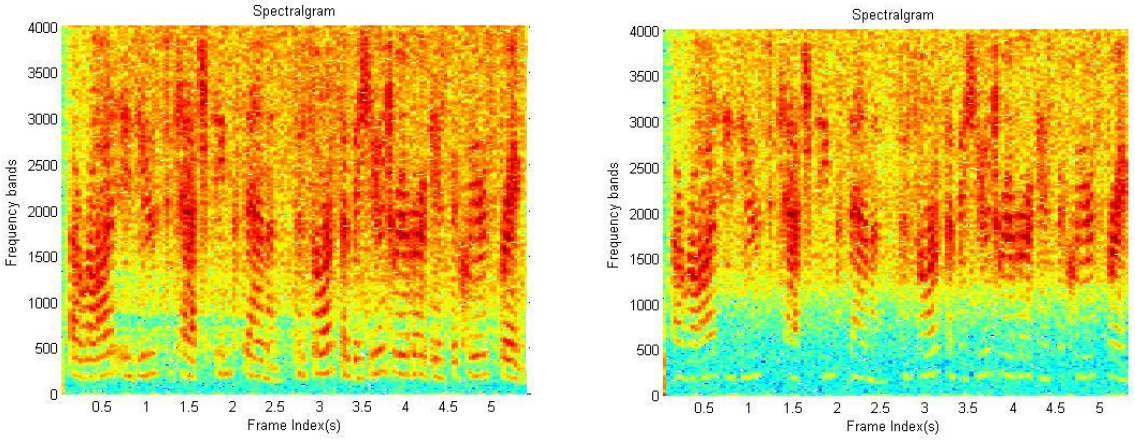


Fig. 24 (a) Narrowband epsilon selection speech enhancement of one-interference (b) two-interferences.

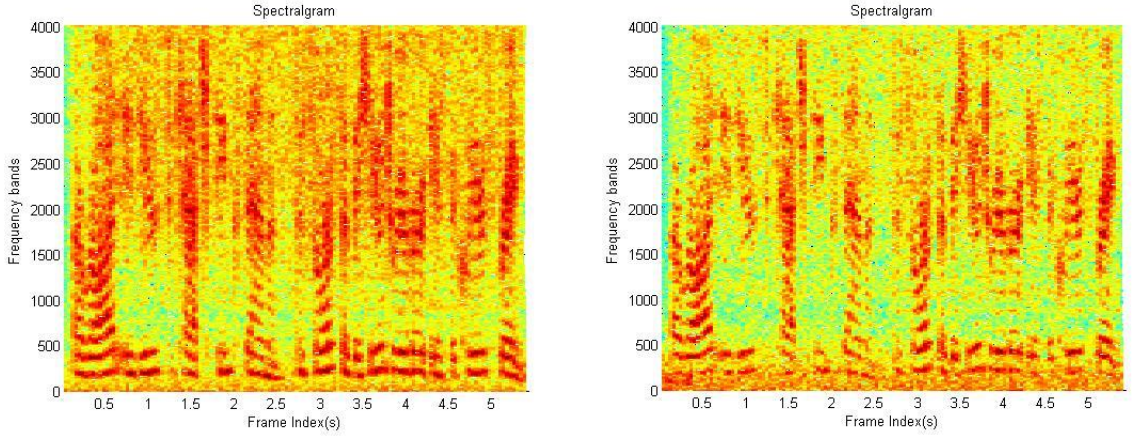


Fig. 25 (a) Wideband epsilon selection speech enhancement of one-interference (b) two-interferences.

3.6 Overall System Architecture

Combining the beamforming technique proposed in Section 3.3, the parameter determinism in Section 3.4, and the beamformer null-tracking and null-constraints in Section 3.5 for speech enhancement, the overall system architecture is summarized briefly in this Section.

The flowchart is plotted in Fig. 27 to elaborate the overall system architecture. The main processing can be separated into three Kalman filters, written as Kalman filter 1, 2, 3 in Fig. 27. The Kalman filter 1 is operated as the beamformer null tracking estimator, thus its error covariance of measurement equation should be chosen small enough to achieve better efforts on noise reduction. Note that there is no tradeoff between noise reduction and dereverberation in the silent stage. The Kalman filter 2 serves as the robust beamformer for solving the signal mismatch problem in speech stage, so the steering vector bound should be chosen appropriately to maintain the steered response in different frequency-bands. The Kalman filter 3 uses the null response value and dominant null direction as constraints to strengthen the noise reduction. The primary structure of the overall system is shown in Fig. 26.

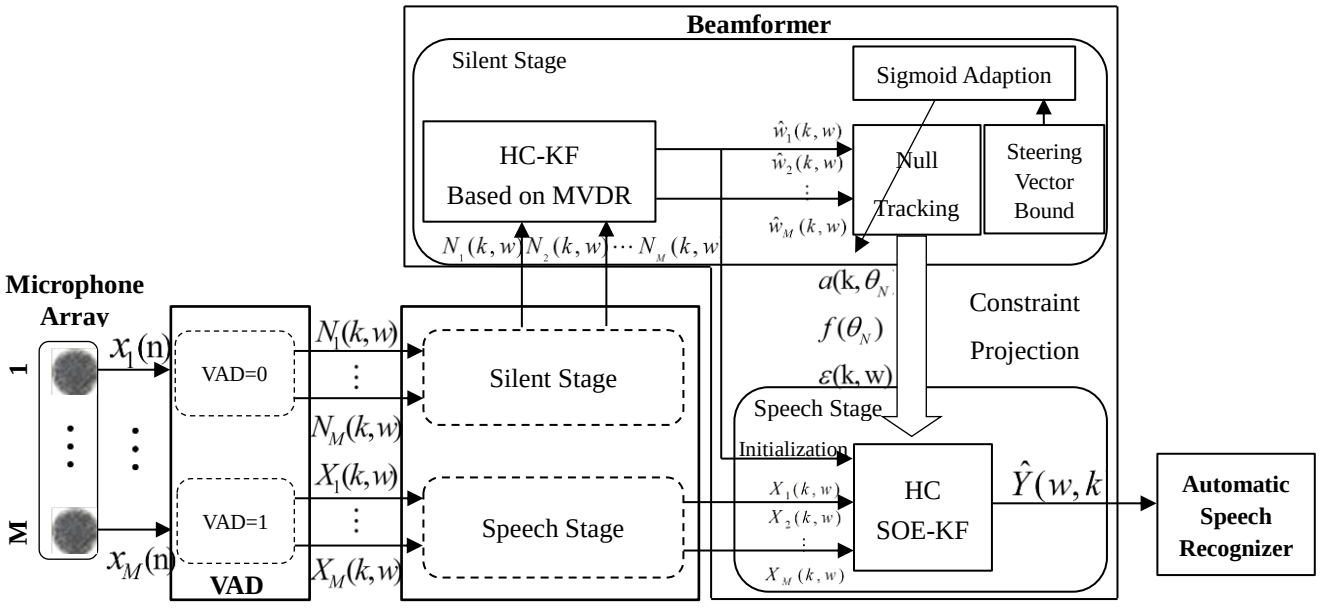


Fig. 26 The Speech Enhancement structure of the Overall System.

To start with, new speech samples in time domain are collected in frames with fixed overlap to the previous frame and transformed to frequency domain after zero padding and Hanning windowing. Assume the information of voice activity detection is given, we can determine the new frame which is desired signal active or inactive.

If the new frame is determined as desired signal inactive, feeding the spatial filtering weights which is given by Kalman filter 1, into beamformer null-tracking. As mentioned before, the Kalman filter 1 serves as noise reduction beamformer and it could find out the direction of undesired interferences and the corresponding beamformer null response values. The parameters of noise tracking should be loaded into the parameters before adapting to the new frame, since the new frame that contains desired signal won't be adapted by Kalman filter 1.

If the new frame is determined as desired signal active, it should to check the parameters which are loaded from noise-tracking. If there are no noisy interferences in the undesired direction, the Kalman filter 2 is used to filter out the undesired noise and solve the signal mismatch problem for maintaining the desired signal undistorted. Using the steering vector bound to cover more steering vectors in sigmoid mathematical model wideband selection works, but if there are noisy interferences existing, the parameters will be fed into the Kalman filter 3 as null constraints which are projected into the updated equation for suppressing the interferences dramatically. And the epsilon selection should be regulated according to the characteristic of (DAS) beamformer and Fig. 9 in low frequency-bands with different array number.

To sum up with, the overall algorithm contains three Kalman filters to handle the two issues of beamformer null tracking and robust beamforming against signal mismatch problems. The Kalman filters differ in its constraints for maintaining the desired signal and its crucial steering vector bound ε parameter thus render different functions and speech enhancement results. The matrices of state and error covariance

in the measurement equation should also be chosen approximately according to subsection 3.4.1. Finally, given the information of voice activity detection incorporated to keep performance of speech enhancement more robust, and the more accurate performance on automatic speech recognizer or relative voice applications is.

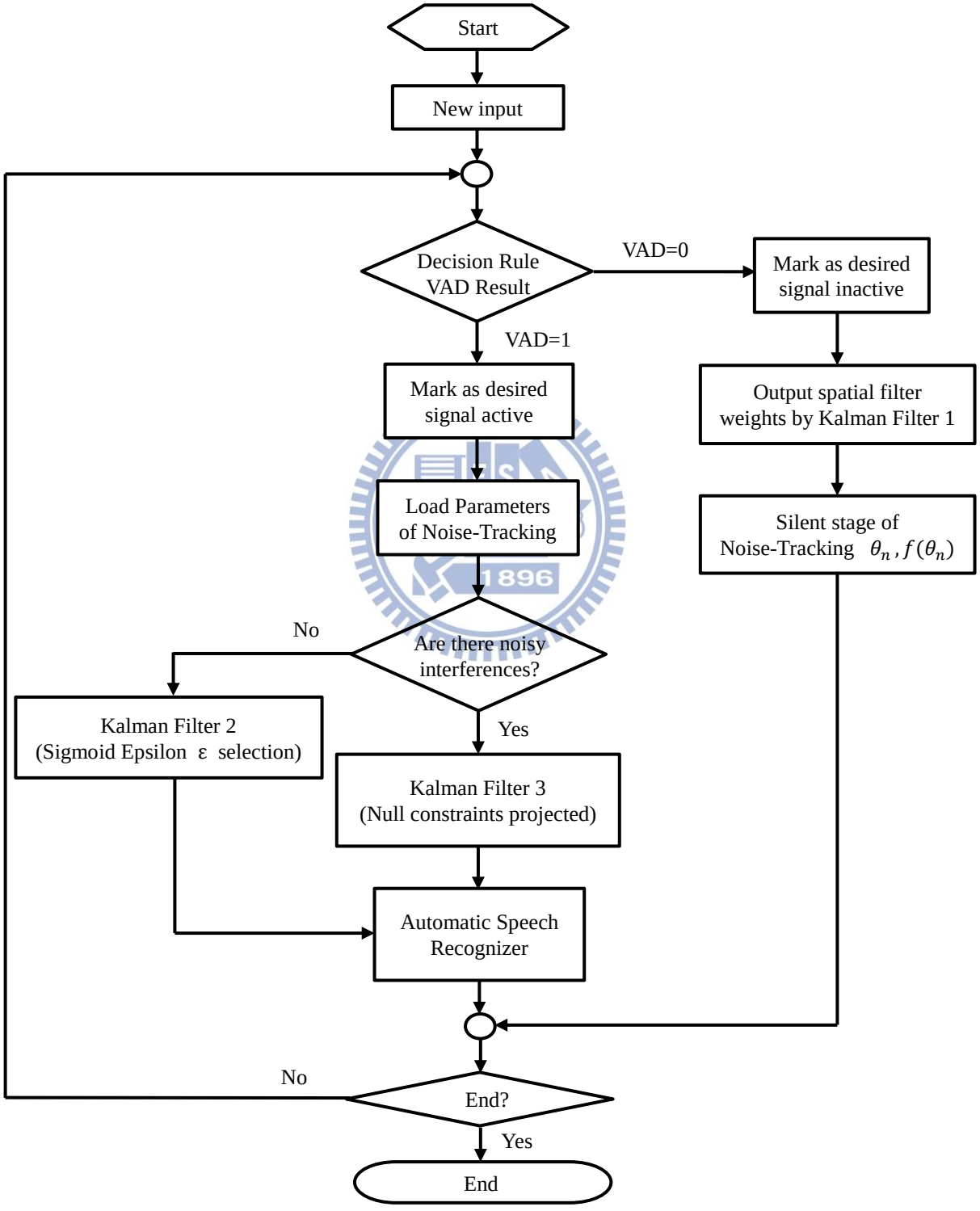


Fig. 27 The Flowchart of the Overall System.

Chapter 4. Simulation and Experiment Results

4.1 Introduction of the Simulation and Experiment Condition

In this chapter, the comparisons between algorithms, which are discussed in chapter 2, are shown in Section 4.2 firstly. The second, we will choose the best performance one but use the proposed method to solve the main shortcoming of it in Section 4.3. The third, experiments in Room Impulse Response (RIR) environment and a real meeting room are presented for comparison. Finally, the pros and cons are explained and discussed with the experiment results.

The proposed algorithm was tested with a uniform linear microphone array of four and six omnidirectional microphones. The distance between each microphone is 4.5 cm. The Size of the room is 10 m × 10 m × 3 m and the microphone array was placed on a table at a distance of 1 m from the wall. The arrangement of microphone array and sound sources is shown in Fig. 28.

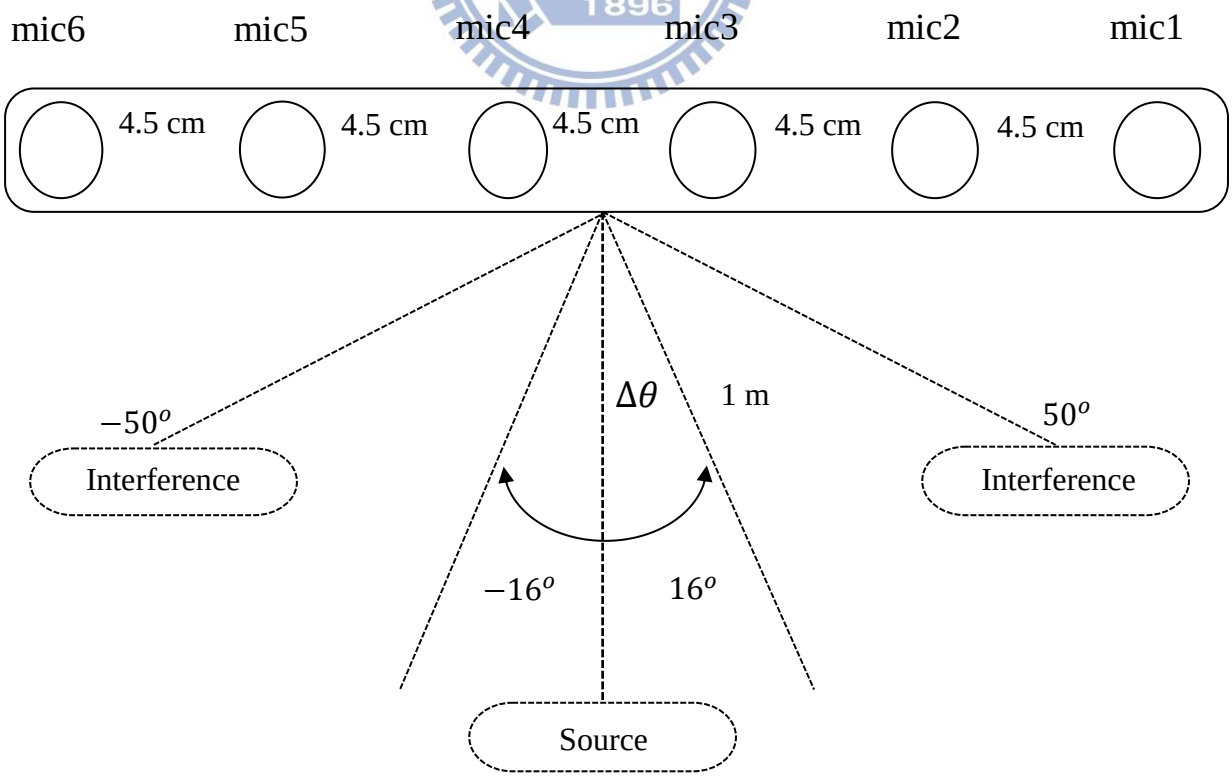


Fig. 28 The location of microphone array and sources.

The desired source is located at $-16^o \sim 16^o$ which consists of sentences spoken by man or woman. Two white interferences are located at 50^o and -50^o. The distance between these sources and middle of the microphone array is 1 m. And the sound sources are recorded independently for the purpose of estimating the output signal to interference-plus-noise ratio (OutputSINR) and log spectrum distortion (LSD).

The sound sources are played by the Head and Torso Simulator (HATS) and recorded in the anechoic chamber without reverberation; a real meeting room with reverberation in Section 4.2 and Section 4.3 respectively. The HATS is used to simulate the characteristic of human being acoustic output. And the sound sources are recorded by a digital microphone array, which are digital microphones to receive signal and collect 16-bits array data in an Altera FPGA development board. The received data is visible for embedded network hardware Net-Burner through shared memory. Finally, the array data is transferred to PC or Laptop through Local Area Network (LAN).

The speech data is extracted from a listening comprehension test by an English learning center, thus the input SNR could be very high. The detail specifications about the pre-processing of raw data by Short-time Fourier transform (STFT) are presented in Table 5 and Fig. 29. The Fig. 30-32 illustrate the recording environment and experiment equipment (including Laser Range Finder, Decibel meter, Altera FPGA development board, digital microphone arrays, ULA setting block, and loud speaker). Three objective performance indices are used to measure the performance of noise reduction and waveform property directly as following:

Microphone Number	4 , 6	Microphone Displacement	4.5 cm
Sampling rate	8000 Hz	FFT size	512 samples
Shift number	160 samples	Zero padding	32 samples

Table 5. Parameters in simulations and experiments.

1. *Output signal to interference-plus-noise ratio (OutputSINR)*

One equality in the frequency weighted measurement for evaluating the performance of noise reduction, which is defined as below:

$$SINR_{avg} = 10 \cdot \log_{10} \frac{\sum_{k}^{\varepsilon} \sum_{w=w_1-5}^{w_1+5} Y_s^{\,2}(k,w)}{\sum_{k}^{\varepsilon} \sum_{w=w_1-5}^{w_1+5} Y_n^{\,2}(k,w)} \tag{4.1}$$

where the Y_s is the filtered output of the desired signal, and Y_n is the filtered output of interferences plus noise; (w_1-5, w_1+5) is the windowing of frequency-bands.

2. *Log spectrum distortion (LSD)*

The performance of noise reduction and distortion is a trade-off of beamforming. The better performance of noise reduction may cause more distortion. Hence, another equality measurement for evaluating the performance of distortion, which is defined as

$$LSD = \frac{1}{K} \times \sum_{k=1}^{K} \sqrt{\frac{1}{W} \sum_{w=1}^{W} (20 \cdot \log_{10} |A_{11}(w) S_1(k,w)| - 20 \log_{10} |Y(k,w)|)^2} \tag{4.2}$$

where the $Y(k,w)$ is the STFT of the filtered output and the $S_1(k,w)$ is the STFT of the original sound played by the HATS. 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 from methods.

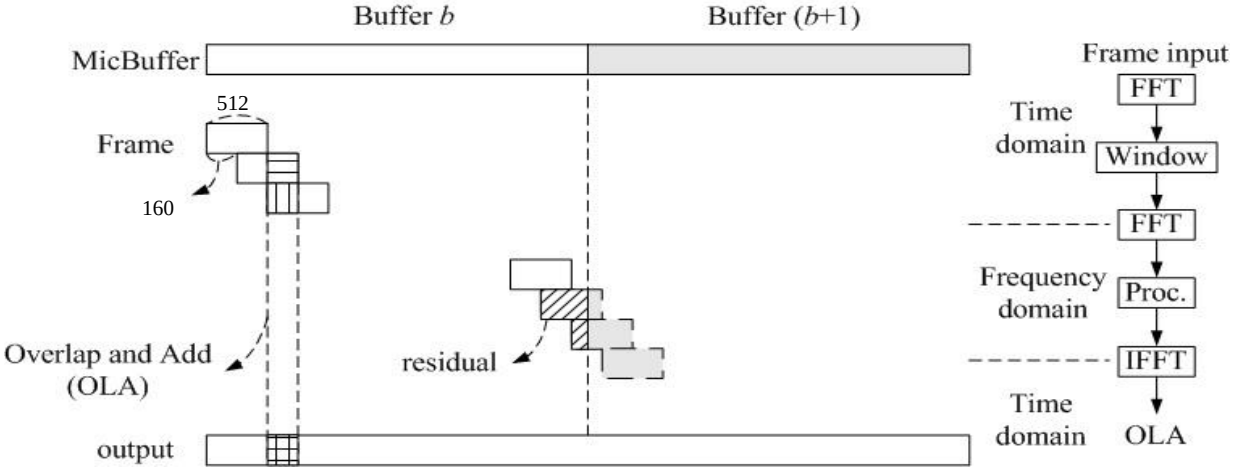


Fig. 29 Pre-processing of raw data in beamforming (STFT)

3. Perceptual Evaluation of Speech Quality (PESQ)

The last equality measurement for evaluating the performance of speech equality is PESQ [27]. It is a widely accepted industry standard for objective voice quality testing and used to estimate the whole speech enhancement performance at different input avgSINR. The automatic objective estimation of speech equality is substituted to the judgment system before. The PESQ result will show the difference between the original input signal and the degrading output signal which is played by the HATS. The recorded sources contain background noise and diffused noise. PESQ can be regarded as the subjective hearing equality of human being to a certain extent.

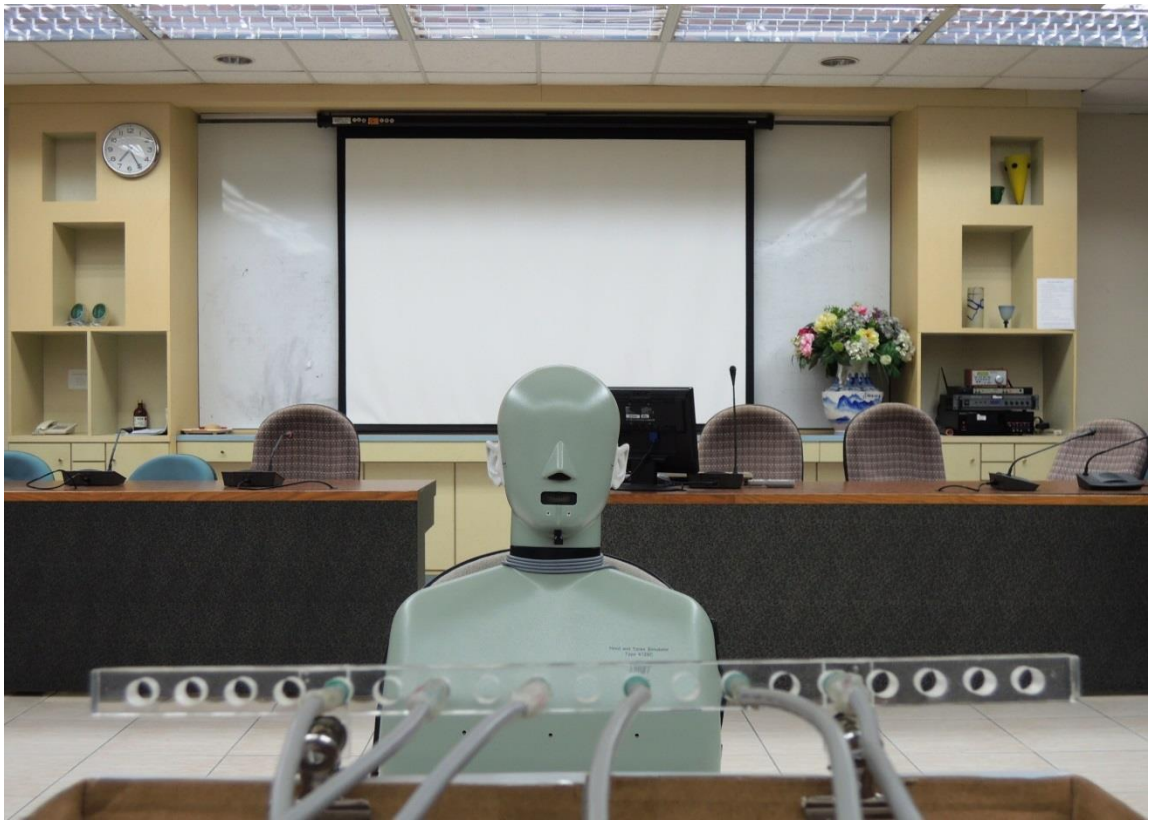


Fig. 30 The HATS and the digital microphone arrays (ULA).

Fig. 31 The real meeting room environment.

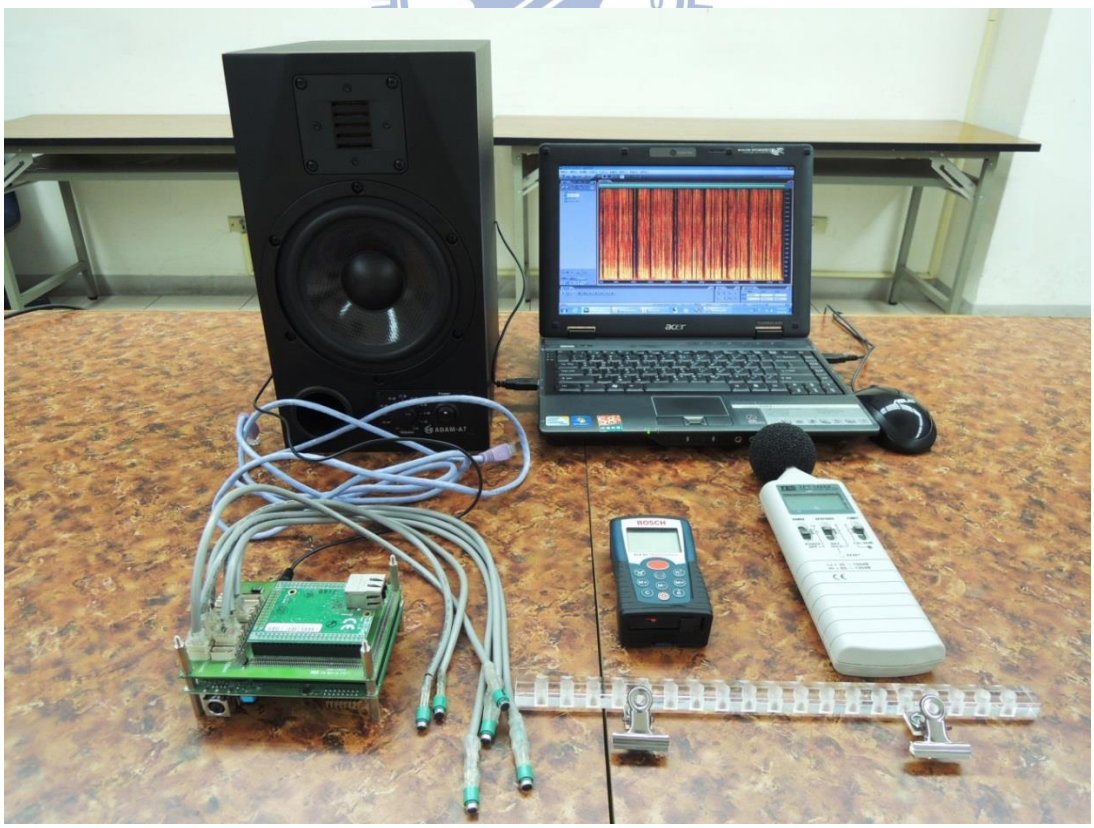
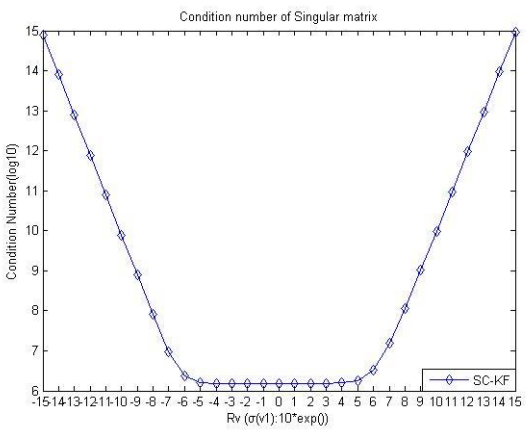


Fig. 32 The illustration of experiment equipments.

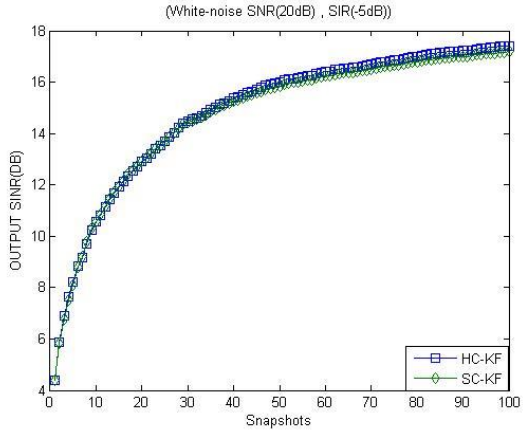
4.2 Adaptive Spatial Filter Comparison Result

In this Section, the comparisons of noise reduction between adaptive spatial filters, which are discussed in Section 2 based on the robust MVDR beamformer, is exhibited. The weights of spatial filter update only when the desired source is not present. When the desired source is present, we use the weights as the fixed beamformer to investigate the performance. The data is recorded at Anechoic Chamber without reverberation. In Fig. 33, the recorded source is at 0^o SNR 20(dB) and interfering source at 50^o SIR -5(dB). We change the variance of filtered output error covariance matrix in the measurement equation from 10^{-15}~10^{15} to see the relationship between the SC-KF and HC-KF. Obviously, when the σ_1^2 is chosen too small, the OutputSINR and LSD of SC-KF will degrade significantly in Fig. 33(c), (d). The condition number in Fig. 33(a) which is the maximum eigenvalue divided by the minimum eigenvalue of the input correlation matrix will be too big to cause the matrix singular. Although choosing the appropriate parameter of σ_1^2 and σ_2^2 , the HC-KF and SC-KF can be equivalent, there is a tradeoff relationship which is discussed in subsection 3.4.1 between the parameters in SC-KF. Hence for the robust performance consideration, the HC-KF is used as reference adaptive algorithm. In Fig. 34, under different input SIR, the HC-KC is closer to SMI solution in no signal mismatch problem. The CLMS and CNLMS are also close to the optimal solution, but the step-size, which is selected as $10^{-11}, 10^{-1}$ respectively, should be chosen approximately according to the input data and the variable environment.

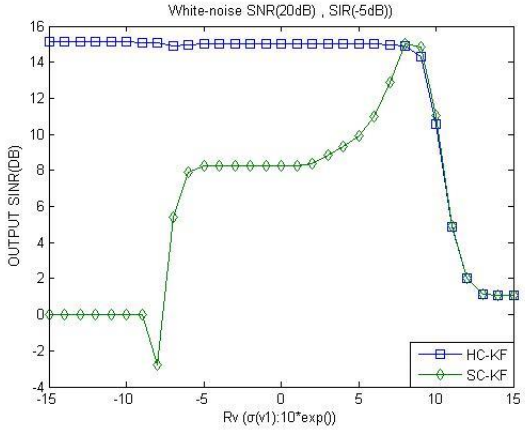
For subjective evaluation, Fig. 35, 36, 37 show the waveforms and spectrograms at different SIR -7 dB, 0 dB, 7 dB with same SNR (0 dB). It can be observed that the diffused noise is rather hard to suppress compared to directive interference. In the final experiments, the proposed method focuses on the suppression of directive interfering source and solves the signal mismatch problem at the same time.



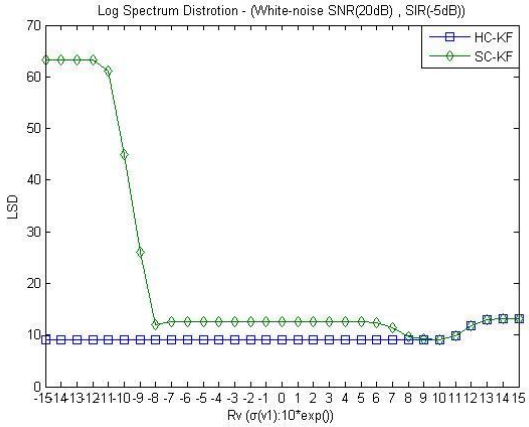
(a) Condition number



(b) Convergence speed $\sigma_1^2 = 10^{10}$

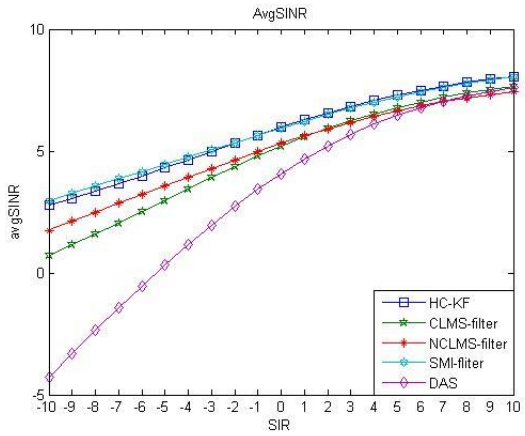


(c) OutputSINR result

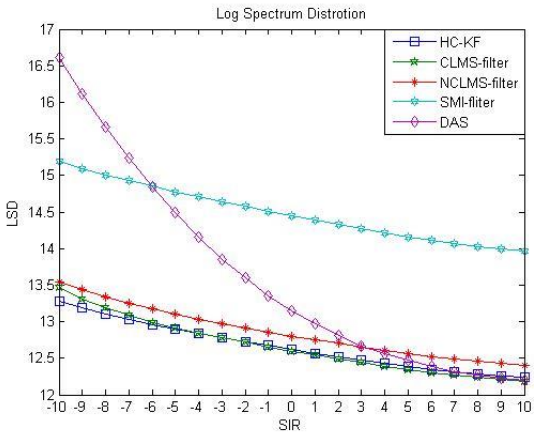


(d) LSD result

Fig. 33 The difference between SC-KF and HC-KF

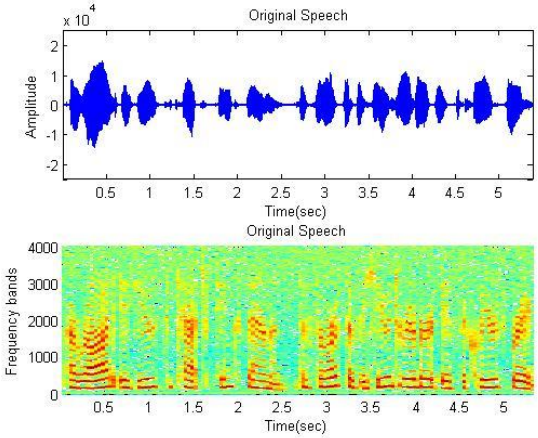


(a) AvgSINR in algorithms

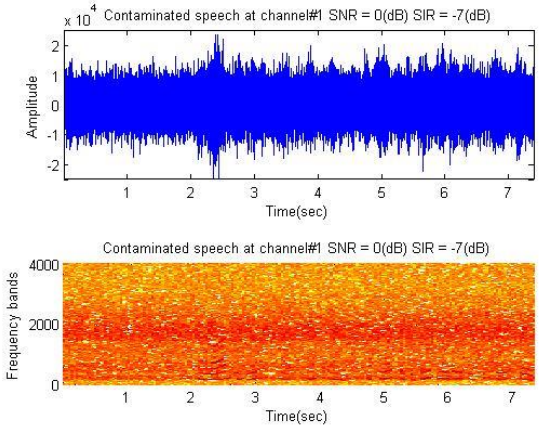


(b) LSD in algorithms

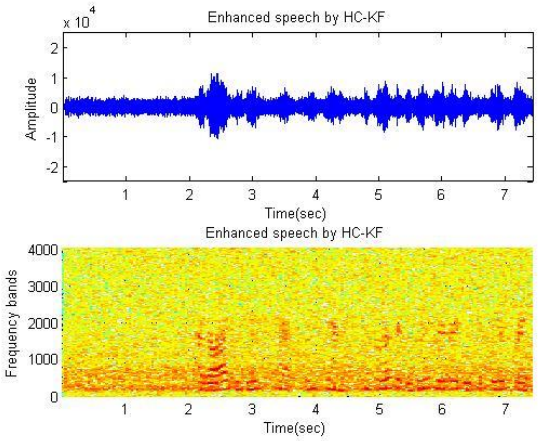
Fig. 34 Comparison between different adaptive algorithms.



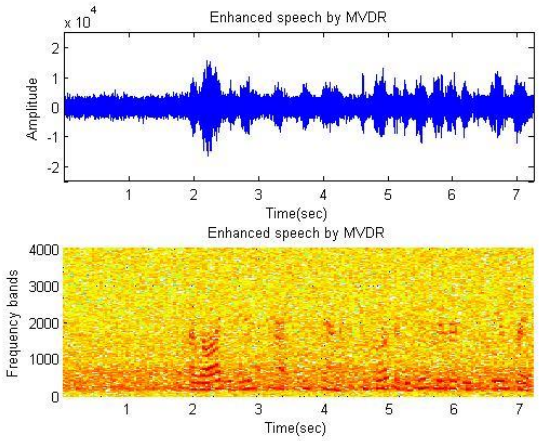
(a) Original Speech



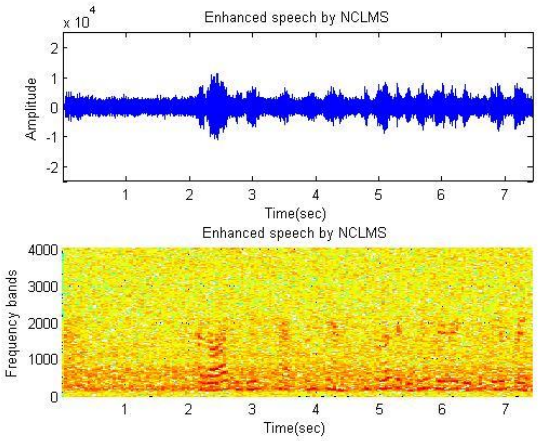
(b) Contaminated Speech



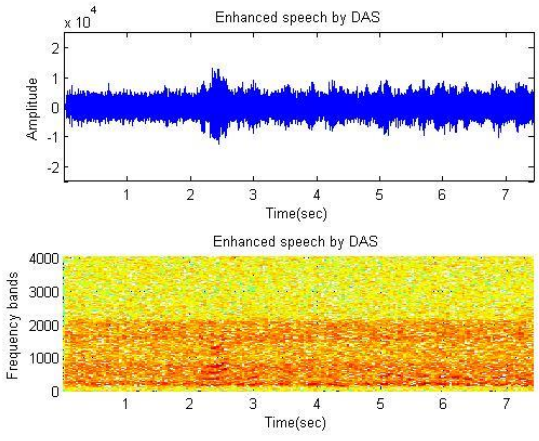
(c) HC-KF algorithm



(d) SMI algorithm

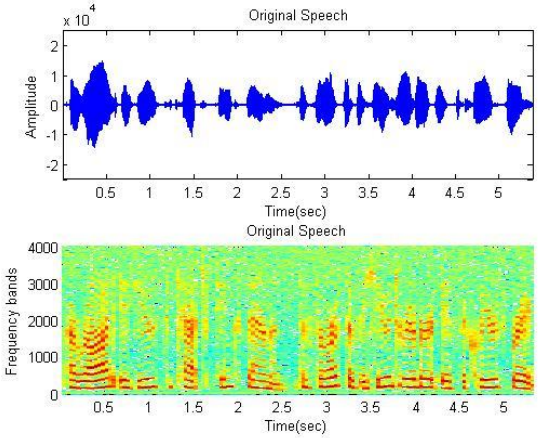


(e) CNLMS algorithm

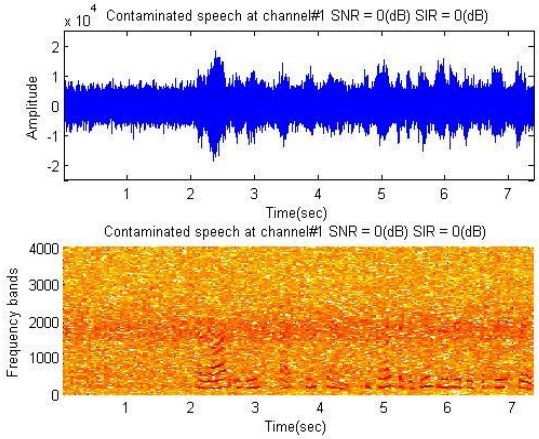


(f) DAS algorithm

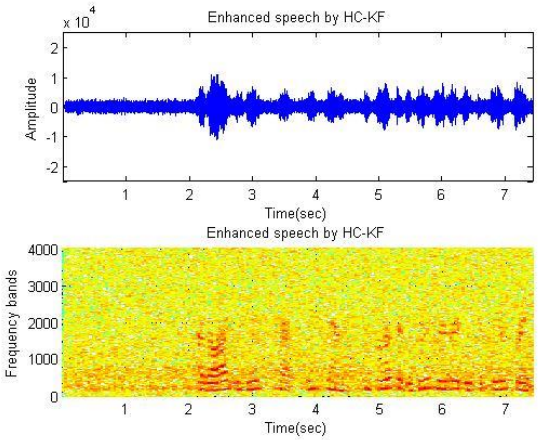
Fig. 35 Experiment results with input SNR 0(dB), SIR -7(dB).



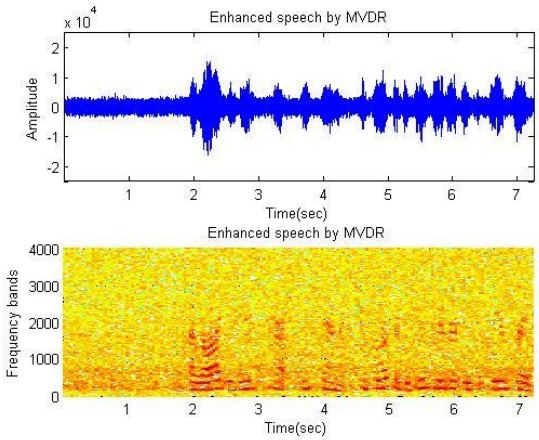
(a) Original Speech



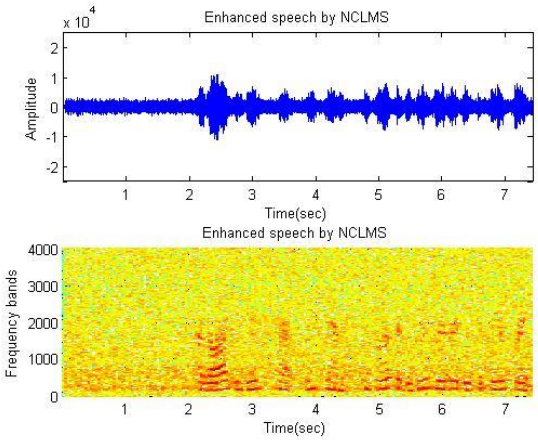
(b) Contaminated Speech



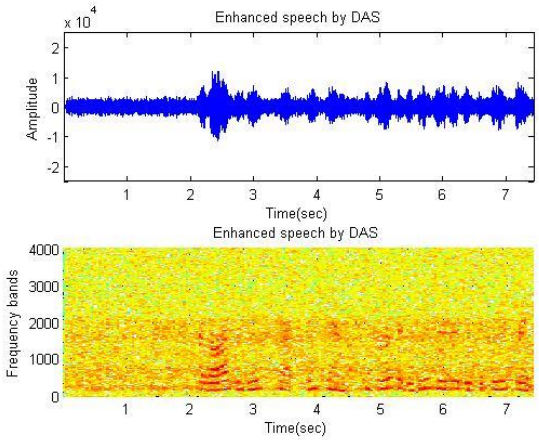
(c) HC-KF algorithm



(d) SMI algorithm

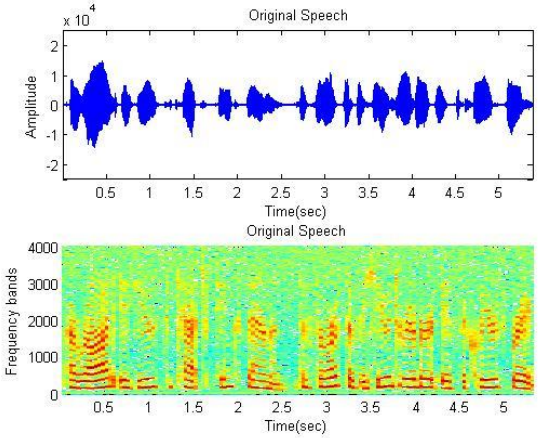


(e) CNLMS algorithm

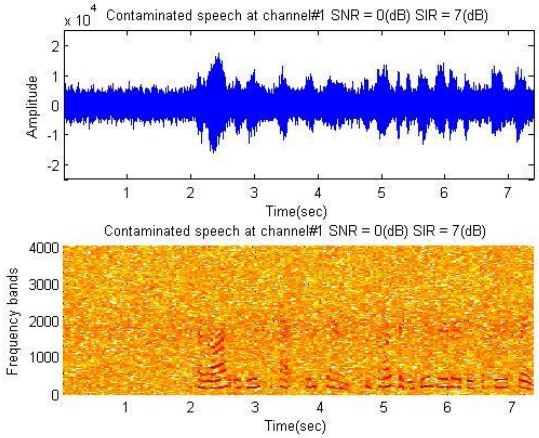


(f) DAS algorithm

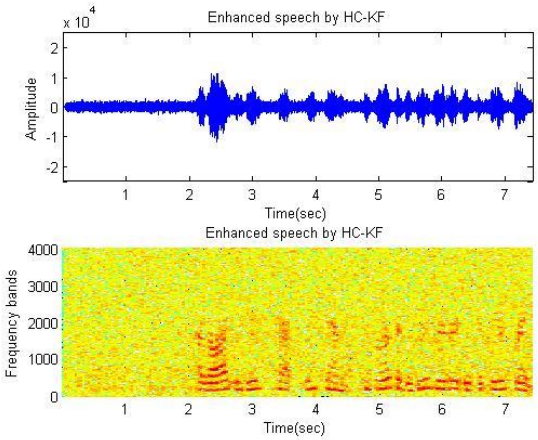
Fig. 36 Experiment results with input SNR 0(dB), SIR 0(dB).



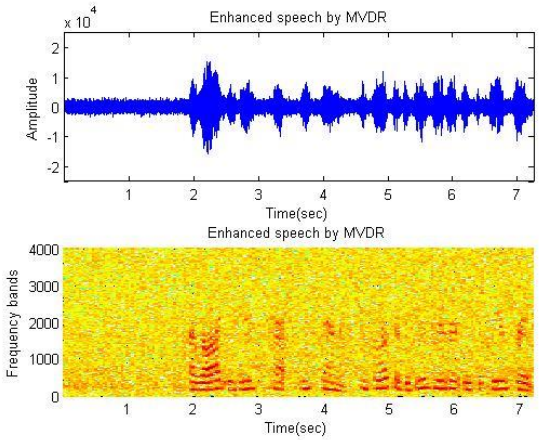
(a) Original Speech



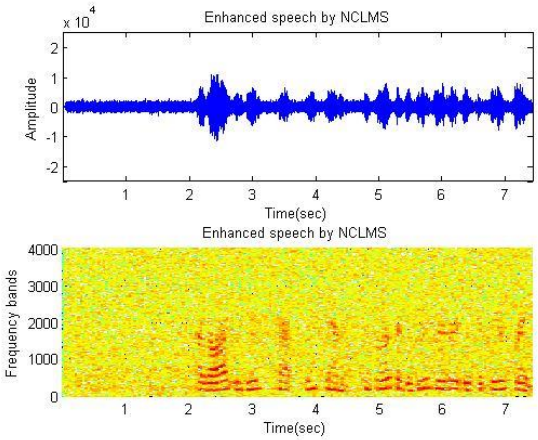
(b) Contaminated Speech



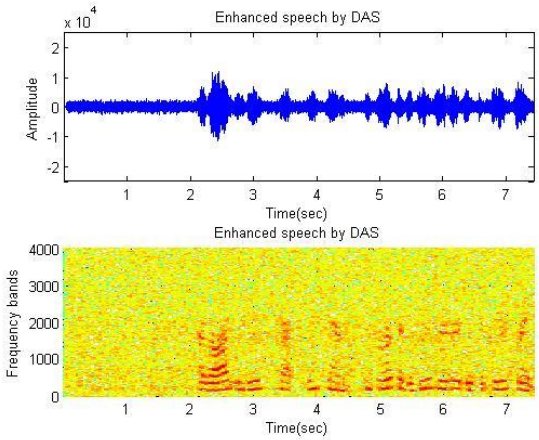
(c) HC-KF algorithm



(d) SMI algorithm



(e) CNLMS algorithm



(f) DAS algorithm

Fig. 37 Experiment results with input SNR 0(dB), SIR 7(dB).

4.3 Experiment on Performance to the Signal Mismatch Problem

In this Section, the comparisons of experiments in the Room Impulse Response (RIR) and a real meeting room are shown. Signal Mismatch problem could easily happen in real environment for speech enhancement applications. For examples, the reverberation is the most known reason that causes signal mismatches. The second, the assumption for spatial relationship of arrays is different to the real condition that results in arrays calibration procedure to compensate for channel amplitude and phase mismatch. The third, the violation of the spatial Nyquist theorem at low frequencies degrades the performance and limits the effective observations in finite time.

(A) RIR Experiments Performance

Three performance indices, which are introduced in Section 4.1, are used to compare the performance of speech enhancement. We use the HC-KF as beamformer null tracking when source is not present and the constrained SOE-KF as modification of HC-KF to solve the signal mismatch problems. The parameter of the HC-KF is $\sigma_1^2 = 10^{10}$, which is the same as the SOE-KF filtered output covariance and the $\sigma_s^2 = 0, \xi = 10^{-12}$. The steering vector bound ε is selected as 2.55 [18] in narrow-band selection, and slope of sigmoid function is 0.05 in wideband selection for ε.

In Fig. 38-41, we analyze the different desired signal mismatch conditions from $\Delta\theta = -16$ to $\Delta\theta = 16$ in the RIR experiment. The average input signal to noise-plus-interference ratio is 5 (dB) approximately. Comparisons between SOE-KF and HC-SOE-KF are also shown in LSD and OutputSINR. In Fig. 38-39, the OutputSINR and LSD of HC-SOE-KF are more non-sensitive against different mismatches than HC-KF and SMI with impressive performance. In particular, since the mismatch conditions differ from frequency-bands, the proposed mathematical functional ε selection in HC-SOE-KF has a better performance in average.

The robustness of beamforming in speech enhancement is what we want to achieve despite the HC-KF has the best result when there is no signal mismatch. The corresponding results of waveform of frequency spectrum in original desired source and interfering sources with diffused noise are shown in Fig. 40-41.

(B)(C) Experiments in Real Meeting Room

The meeting room dimension is 10 m × 10 m × 3 m and a uniform linear array (ULA) with six omnidirectional microphones was placed on a table at a distance of 1.5 m from the HATS. In Fig. 42-44;48-50, the results of performance indices, which is collected in the Table 6 and Table 7 as the same as the RIR experiment, are shown when the signal mismatch $\Delta\theta = 16^{o}$. It can be observed that SMI and HC-KF beamformer don't take the effect of steering error into account, so the OutputSINR is prone to be proportional to the distortion to a certain extent. Hence, it is possible to reduce the interferences while nulling the desired source, most especially in the low frequencies-bands (Fig. 45-47;51-53). Using the beamformer null-constraints, which are found out in the training snapshots, to strengthen the suppression of interferences is available on both the HC-SOE-KF (soft-beam and hard-null constraints) and HC-KF (hard-beam and hard-null constraints) when the desired source is present. The proposed method not only improves the shortcoming of robust MVDR beamformer which is based on Kalman filter for solving the signal mismatch problem but also enhances the OutputSINR as the same results in the RIR experiments.

It has to be mentioned that, although the distortionless of proposed method is prone to be lower at lower avgSINR like -3 (dB), the performances of OutputSINR and the filtered speech quality PESQ are more robust than others. The reason is that the beam directivity will degrade in low frequency-bands due to hard null-constraints. It enhances the OutputSINR at the cost of signal self-nulling; and the beam constraints

by steering vector bound would amplify the desired source and diffused noise simultaneously in the worst-case performance optimization. Because the direction of source is unknown in algorithm, the distortionless performance is a challenging task under signal mismatch problem. The post-filtering of diffused noise and superdirective beamforming are in need as the future work.

(A) Room Impulse Response

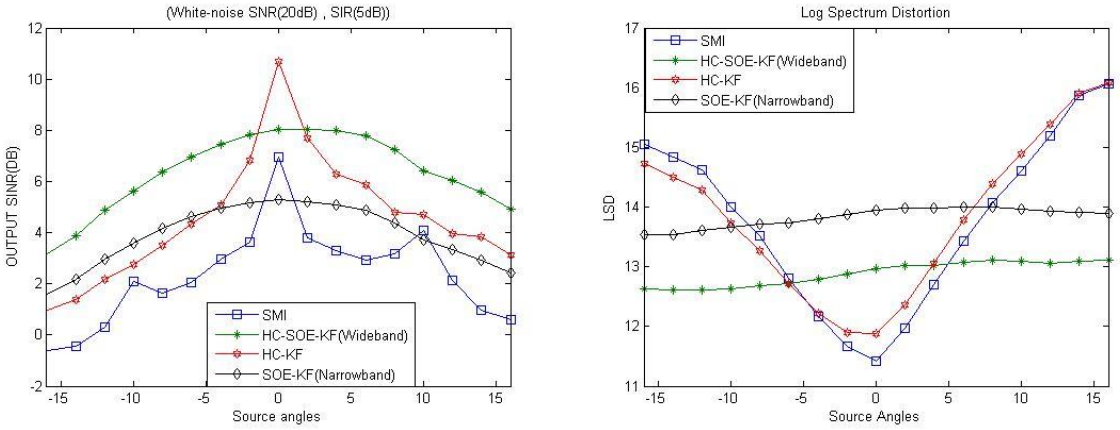


Fig. 38 Comparison of mismatch conditions (No Null-constraint) (a) OutputSINR (b) LSD.

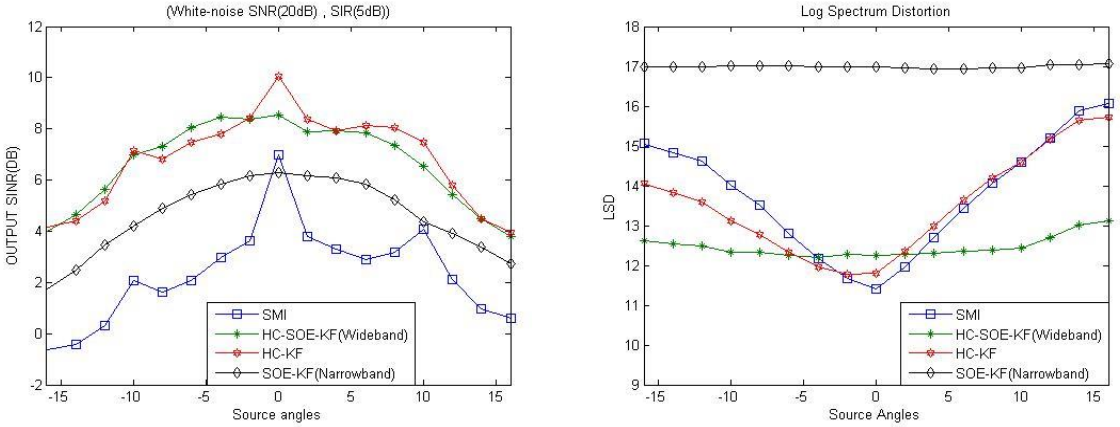
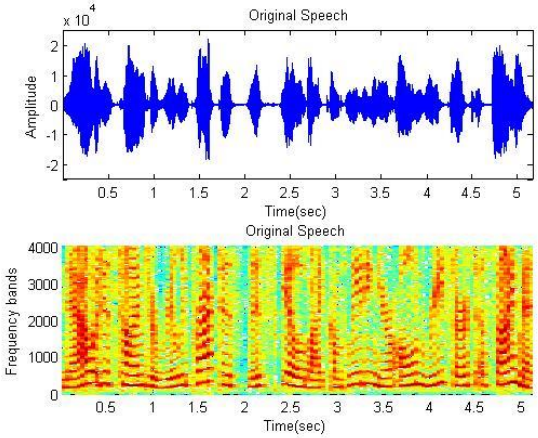
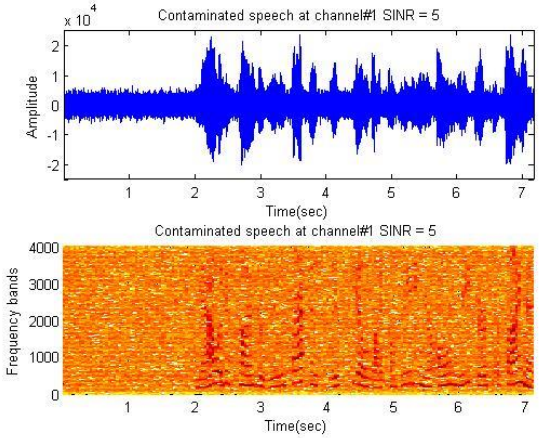


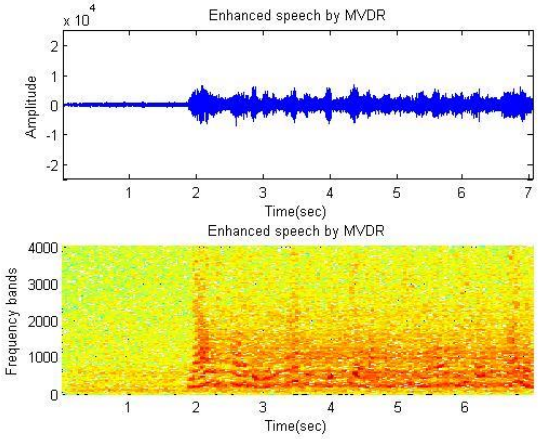
Fig. 39 Comparison of mismatch conditions (Null-constraint) (a) OutputSINR (b) LSD.



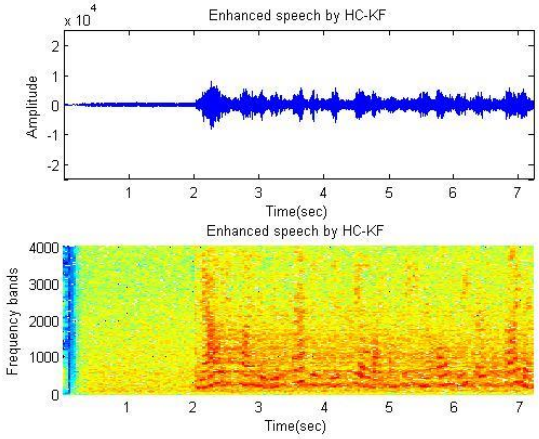
(a) Original Speech



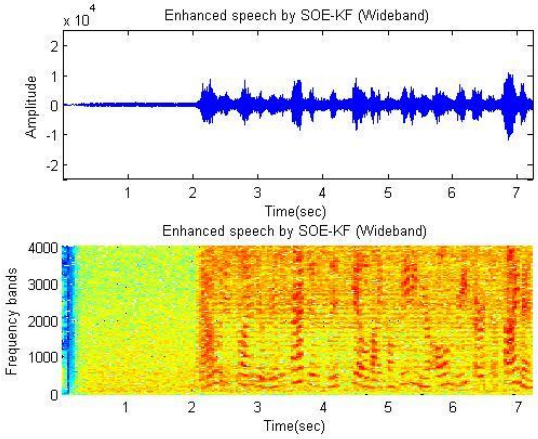
(b) Contaminated Speech



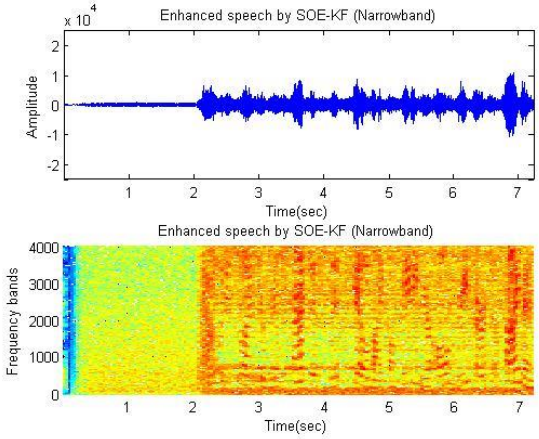
(c) SMI algorithm



(d) HC-KF algorithm

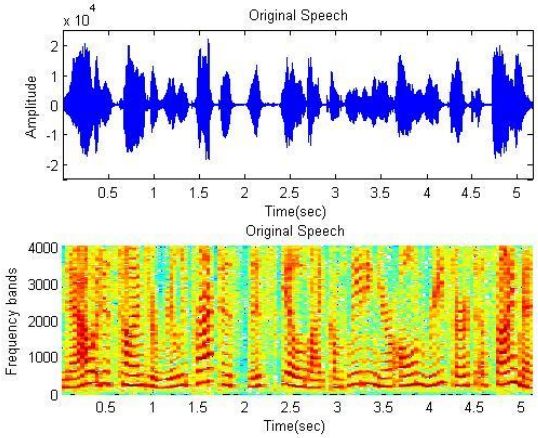


(e) HC-SOE-KF (Wideband ε)

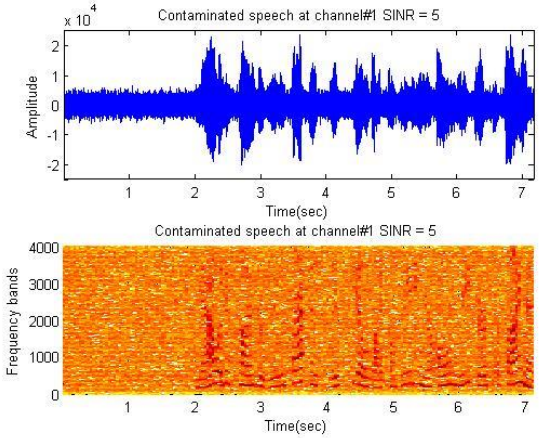


(f) SOE-KF (Narrowband ε)

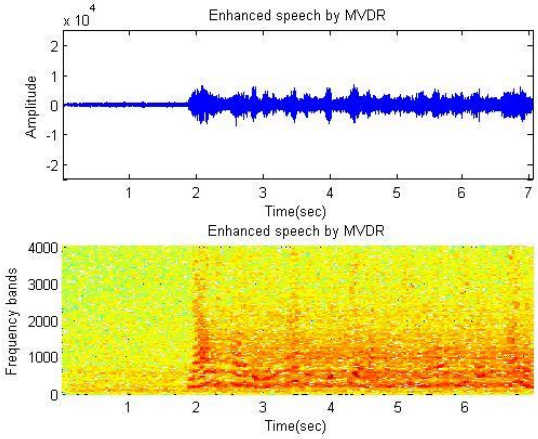
Fig. 40 Filtering results in different mismatch conditions (No Null-constraint).



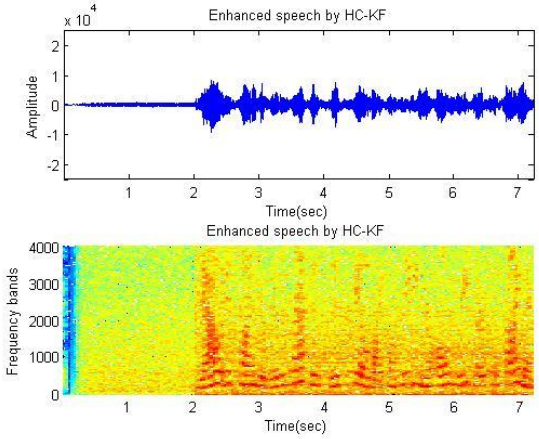
(a) Original Speech



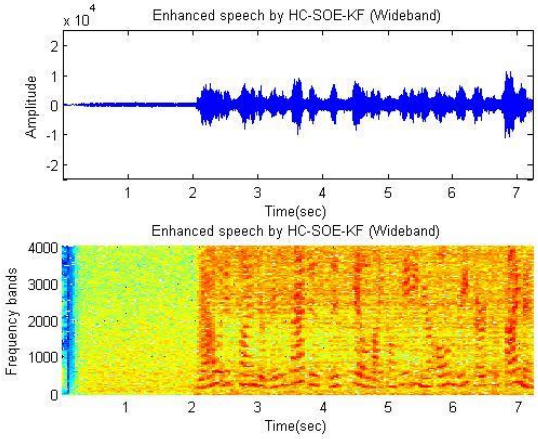
(b) Contaminated Speech



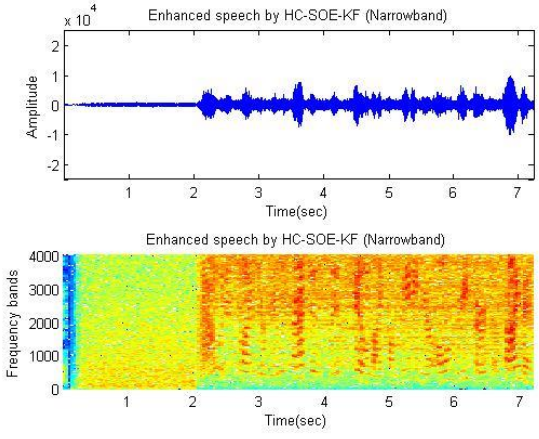
(c) SMI algorithm



(d) HC-KF algorithm



(e) HC-SOE-KF (Wideband ε)



(f) SOE-KF (Narrowband ε)

Fig. 41 Filtering results in different mismatch conditions (Null-constraint).

(B) (Speech source (woman) + White interference)

Input AvgSINR = -5 (dB)			
Algorithms	OutputSINR	LSD	PESQ
Proposed	**5.775**	10.93	**1.86**
HC-KF	5.144	10.21	1.621
SOE-KF	2.135	12.71	1.636
SMI	1.821	**10.17**	1.319
Input AvgSINR = 0 (dB)			
Proposed	**9.618**	**10.55**	**1.986**
HC-KF	8.531	10.89	1.755
SOE-KF	5.226	12.3	1.741
SMI	4.002	10.88	1.288
Input AvgSINR = 5 (dB)			
Proposed	**13.07**	**10.41**	**2.109**
HC-KF	11.78	11.47	1.851
SOE-KF	8.296	12.08	1.839
SMI	5.729	11.54	1.321

Table 6. Results of different algorithms for speech enhancement based on MVDR(V+W).

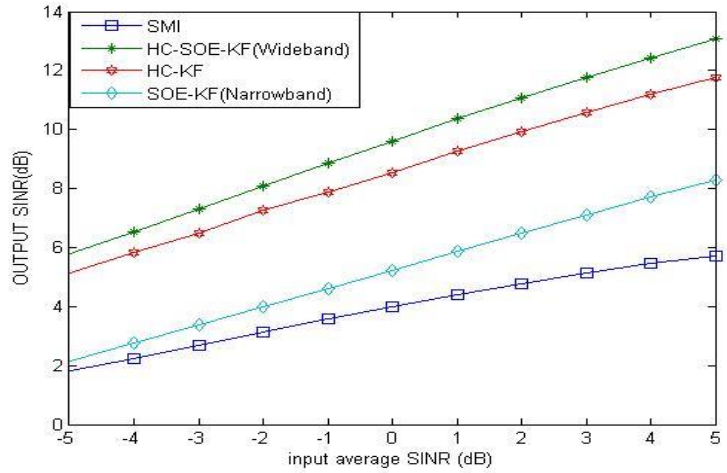


Fig. 42 OutputSINR Comparison of algorithms (V+W).

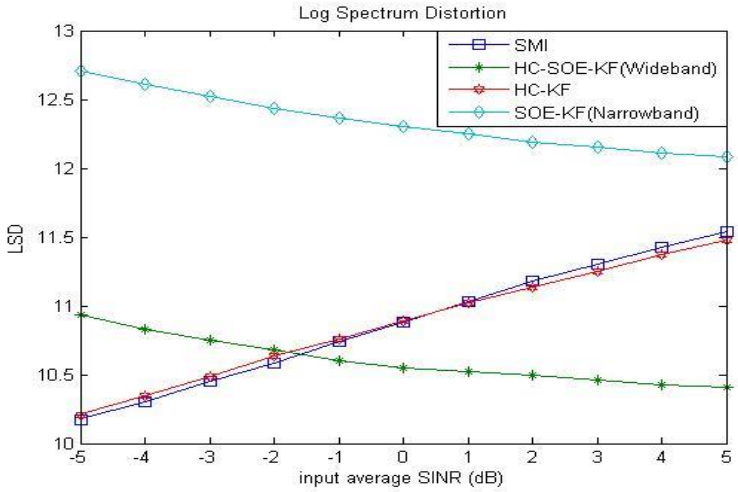


Fig. 43 LSD Comparison of algorithms (V+W).

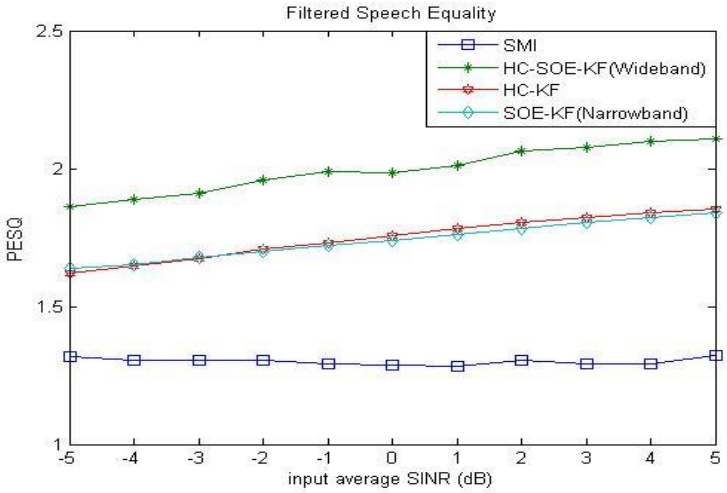
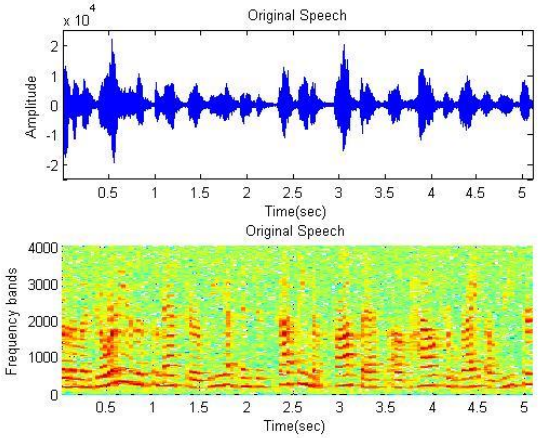
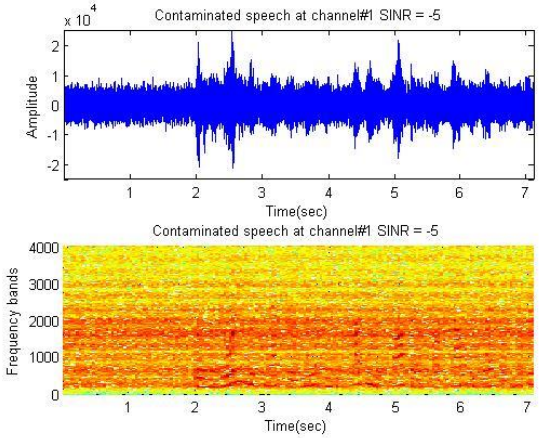


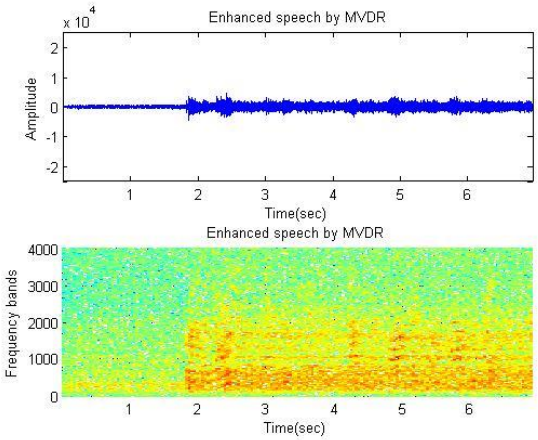
Fig. 44 PESQ Comparison of algorithms (V+W).



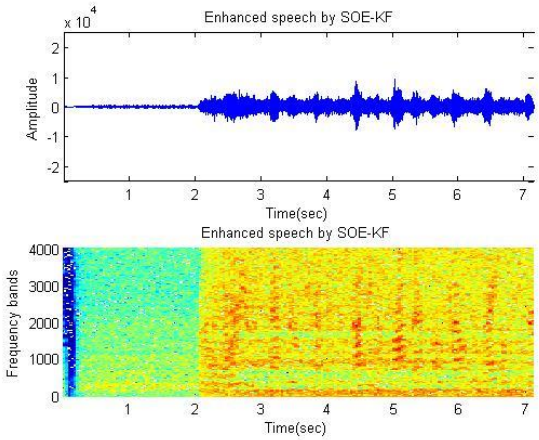
(a) Original Speech



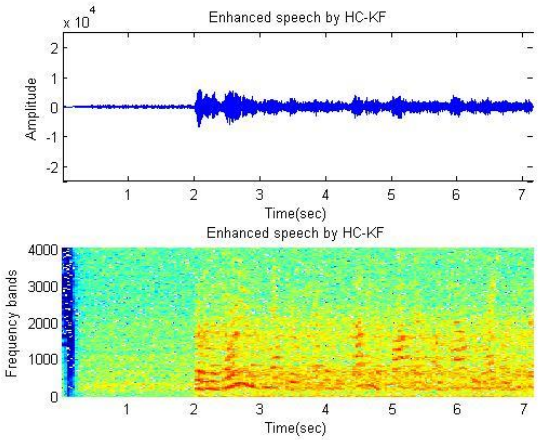
(b) Contaminated Speech



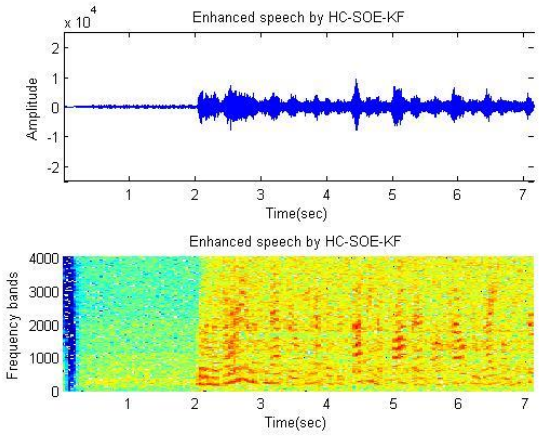
(c) SMI algorithm



(d) SOE-KF algorithm

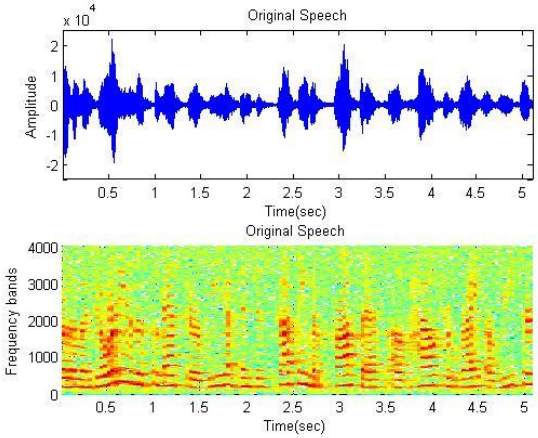


(e) HC-KF algorithm

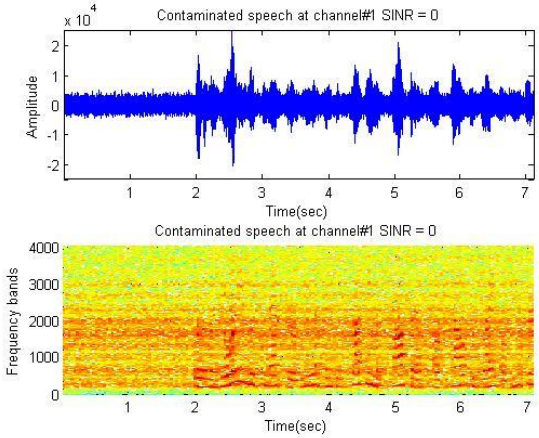


(f) HC-SOE-KF algorithm

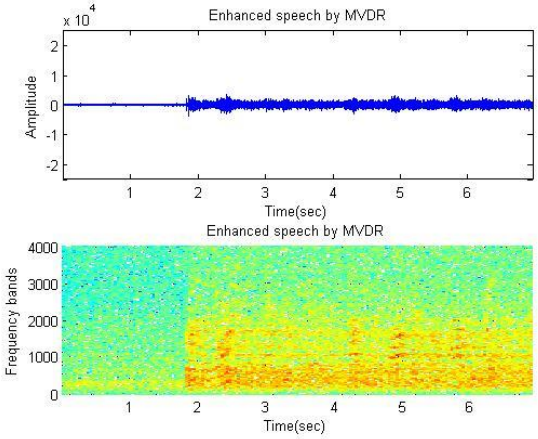
Fig. 45 Filtering results in input avgSINR -5(dB)(V+W).



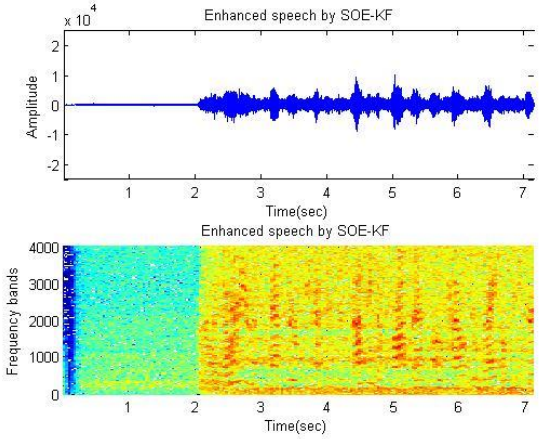
(a) Original Speech



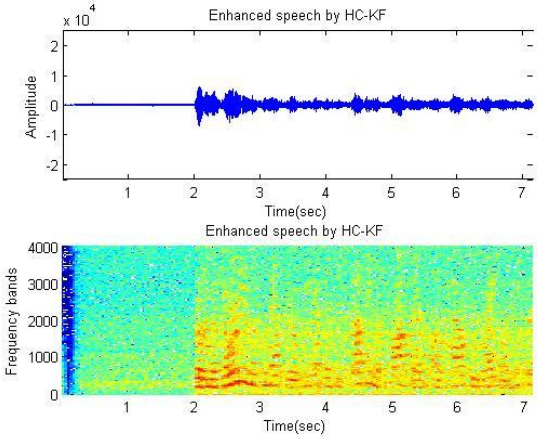
(b) Contaminated Speech



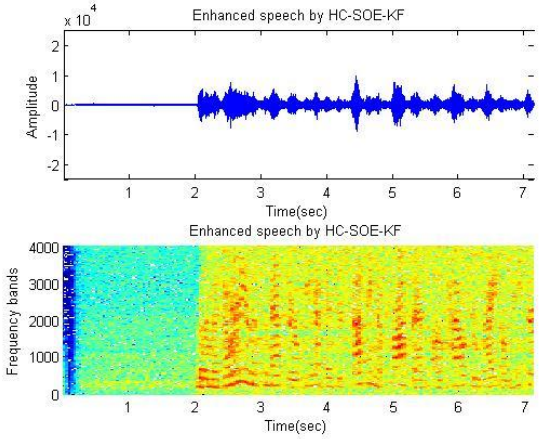
(c) SMI algorithm



(d) SOE-KF algorithm

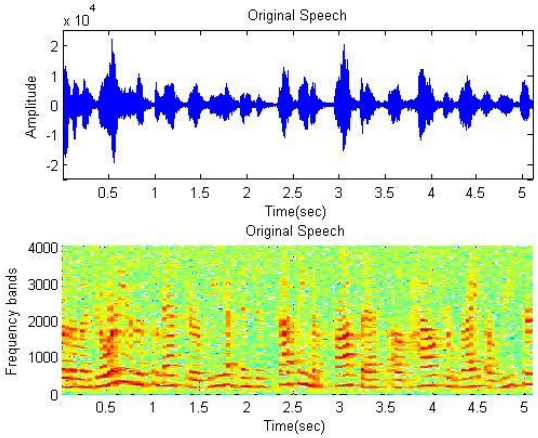


(e) HC-KF algorithm

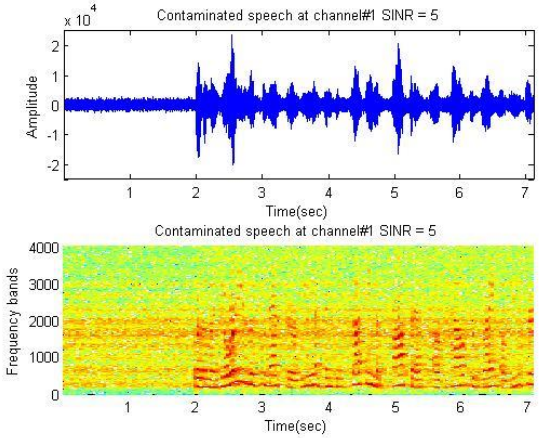


(f) HC-SOE-KF algorithm

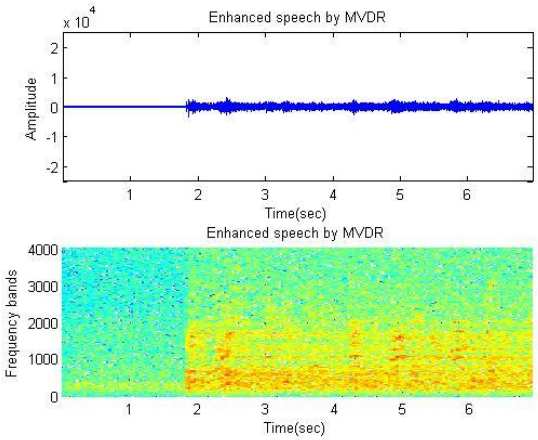
Fig. 46 Filtering results in input avgSINR 0(dB)(V+W).



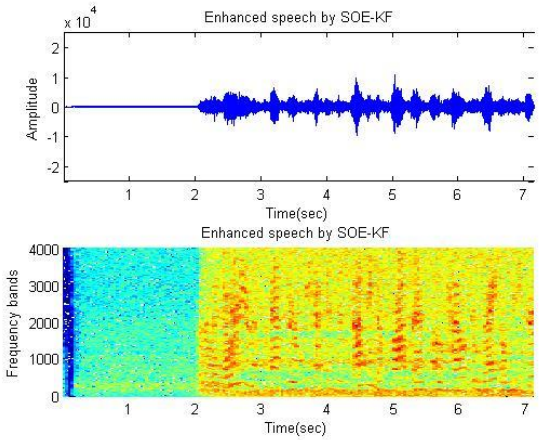
(a) Original Speech



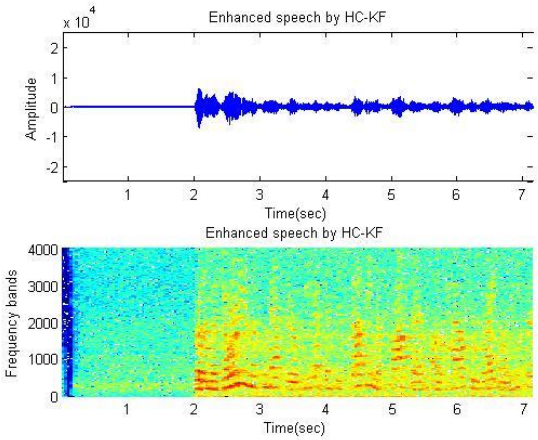
(b) Contaminated Speech



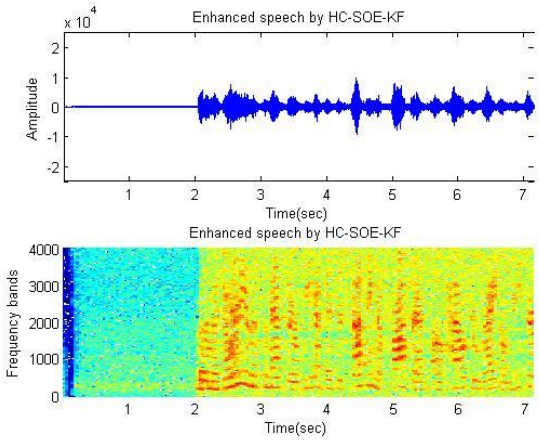
(c) SMI algorithm



(d) SOE-KF algorithm



(e) HC-KF algorithm



(f) HC-SOE-KF algorithm

Fig. 47 Filtering results in input avgSINR 5(dB)(V+W).

(C) (Speech source (woman) + Speech interference (man))

Input AvgSINR = -5 (dB)			
Algorithms	OutputSINR	LSD	PESQ
Proposed	**4.007**	10.87	**1.882**
HC-KF	0.5431	10.2	1.824
SOE-KF	2.93	13.04	1.761
SMI	0.0197	**10**	1.566
Input AvgSINR = 0 (dB)			
Proposed	**7.771**	**10.52**	**2.016**
HC-KF	3.049	10.92	1.943
SOE-KF	5.447	12.52	1.839
SMI	2.352	10.69	1.512
Input AvgSINR = 5 (dB)			
Proposed	**11.46**	**10.43**	**2.098**
HC-KF	6.48	11.56	2.048
SOE-KF	7.534	12.2	1.9
SMI	4.482	11.41	1.53

Table 7. Results of different algorithms for speech enhancement based on MVDR(V+V).

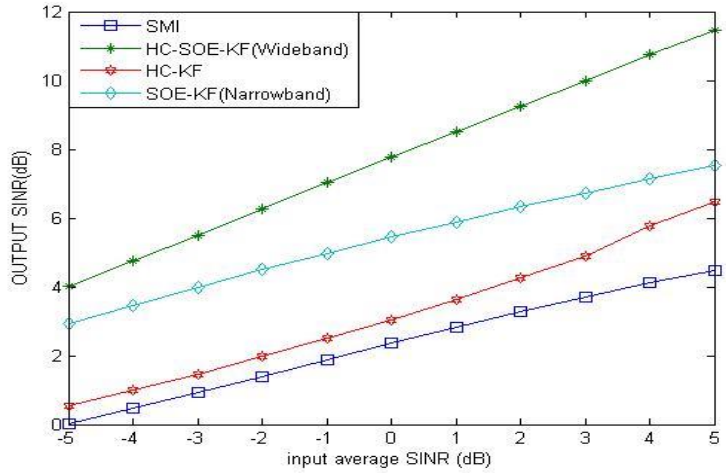


Fig. 48 OutputSINR Comparison of algorithms (V+V).

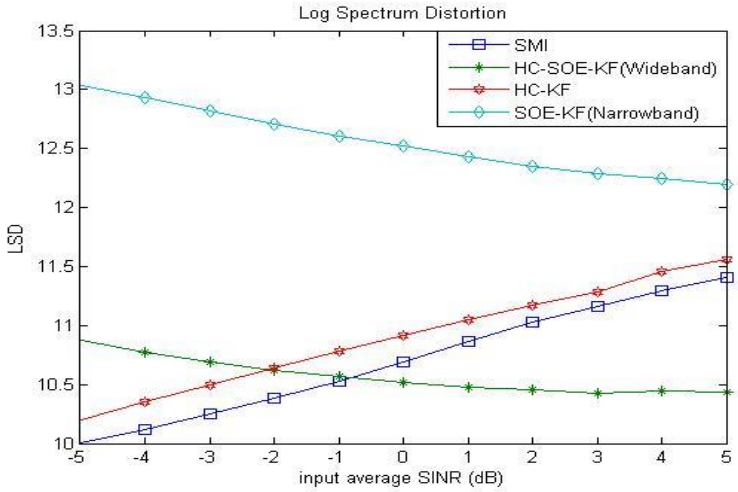


Fig. 49 LSD Comparison of algorithms (V+V).

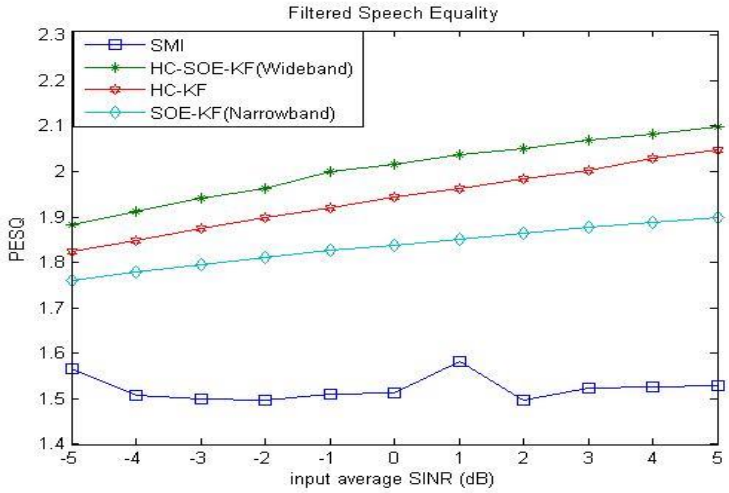
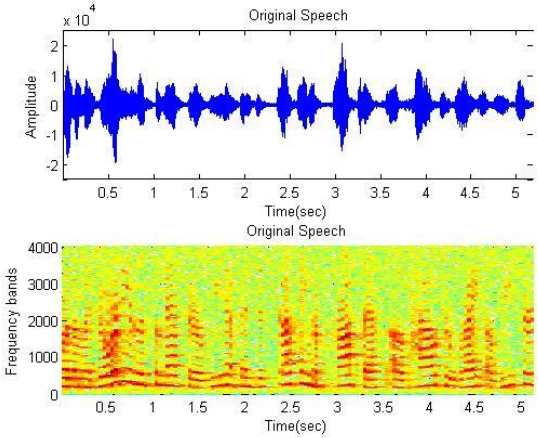
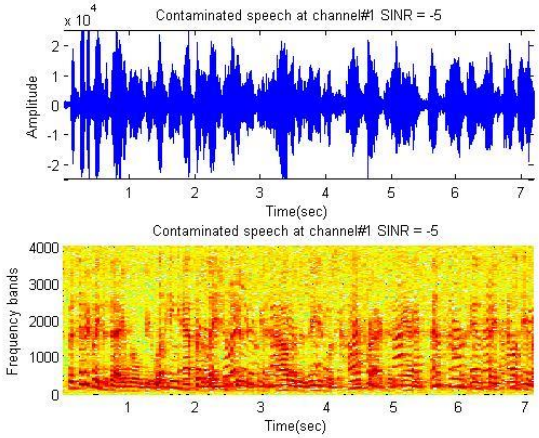


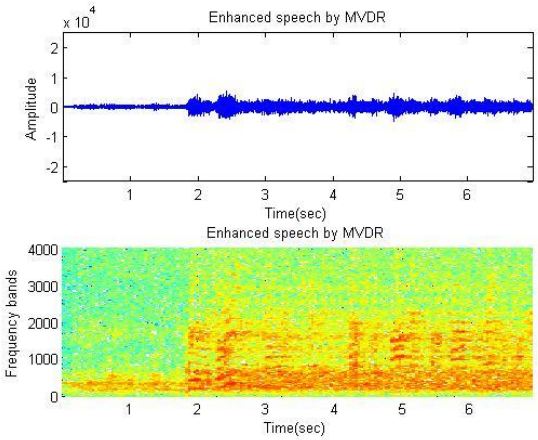
Fig. 50 PESQ Comparison of algorithms (V+V).



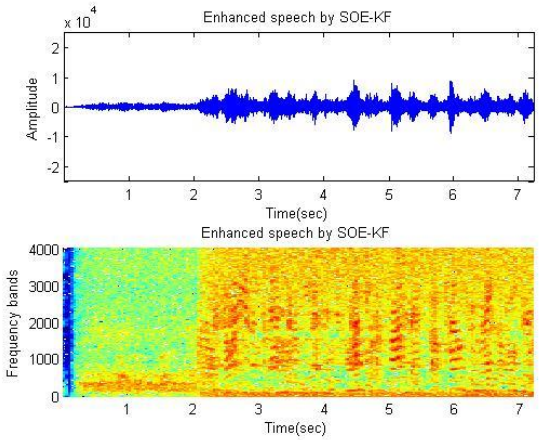
(a) Original Speech



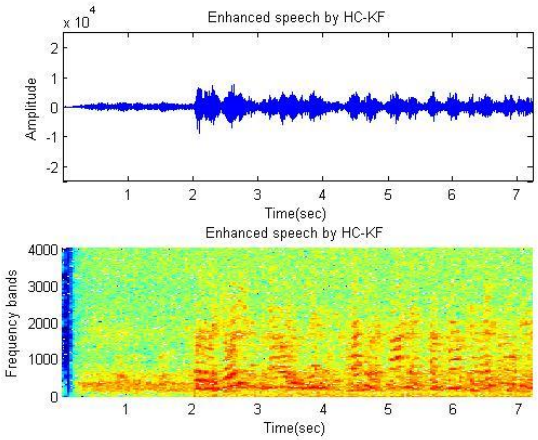
(b) Contaminated Speech



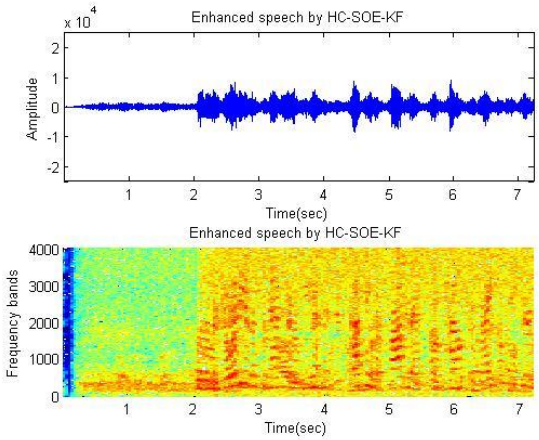
(c) SMI algorithm



(d) SOE-KF algorithm

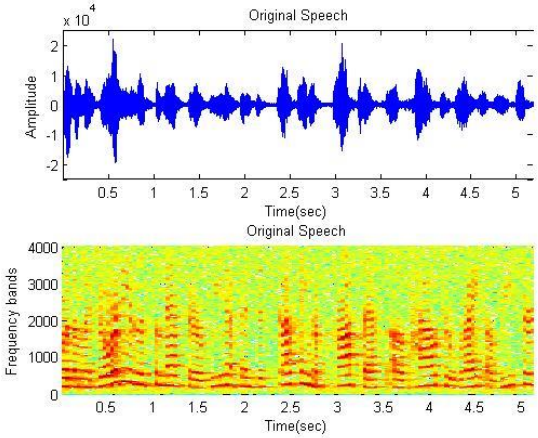


(e) HC-KF algorithm

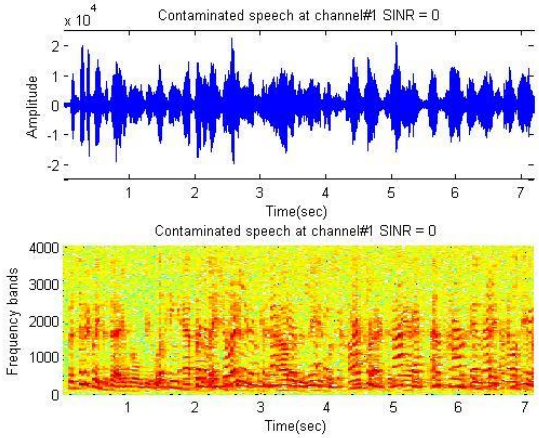


(f) HC-SOE-KF algorithm

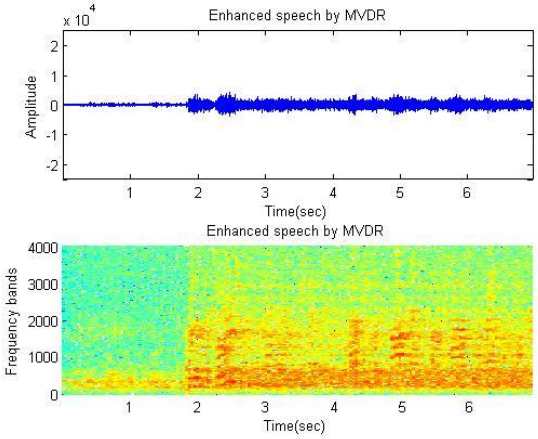
Fig. 51 Filtering results in input avgSINR -5(dB)(V+V).



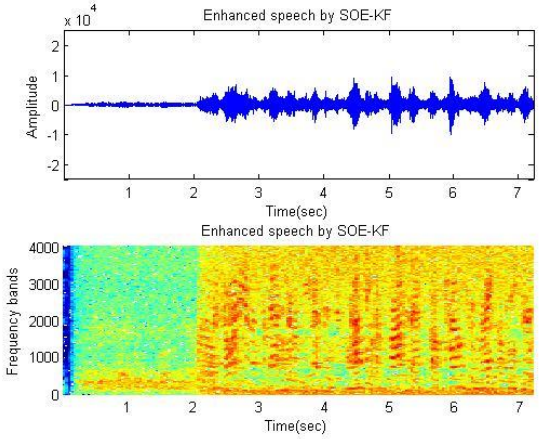
(a) Original Speech



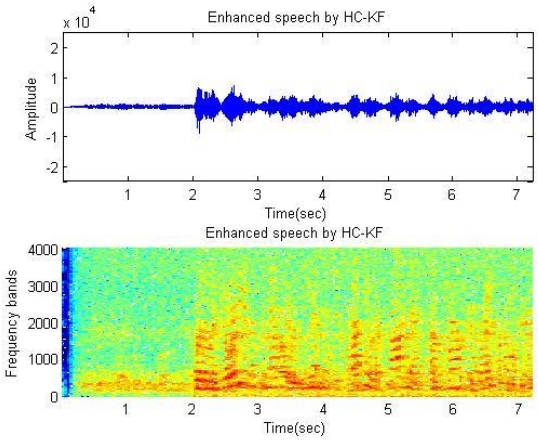
(b) Contaminated Speech



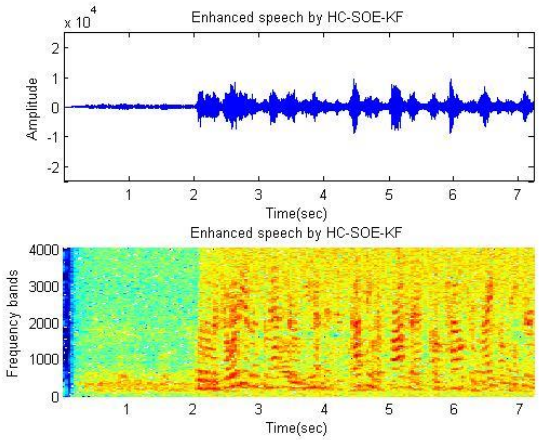
(c) SMI algorithm



(d) SOE-KF algorithm

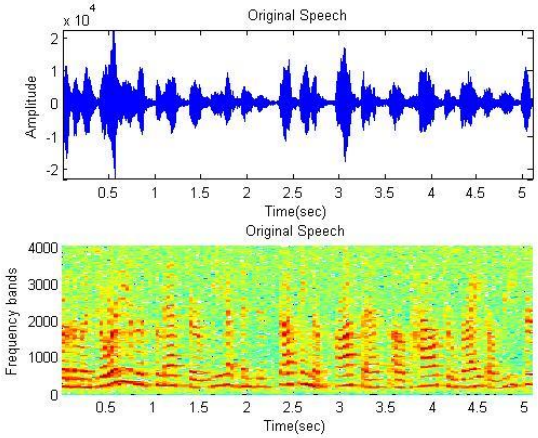


(e) HC-KF algorithm

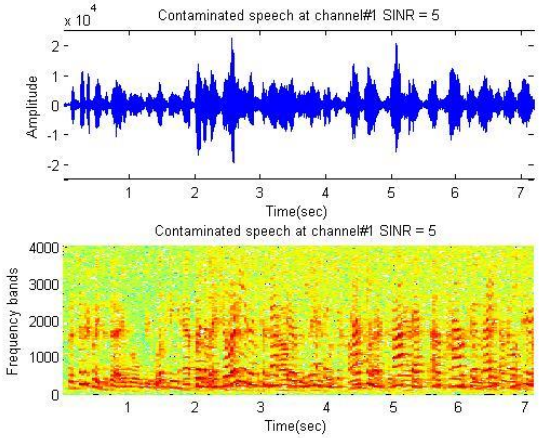


(f) HC-SOE-KF algorithm

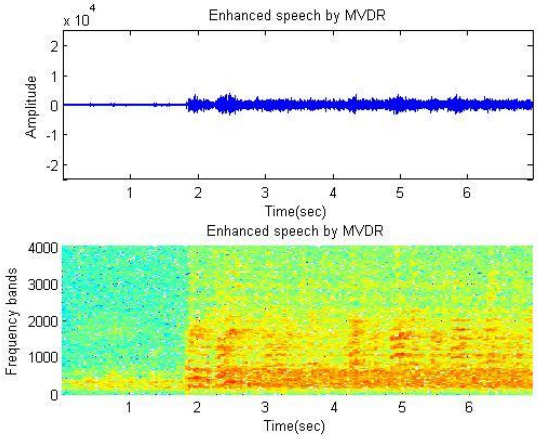
Fig. 52 Filtering results in input avgSINR 0(dB) (V+V).



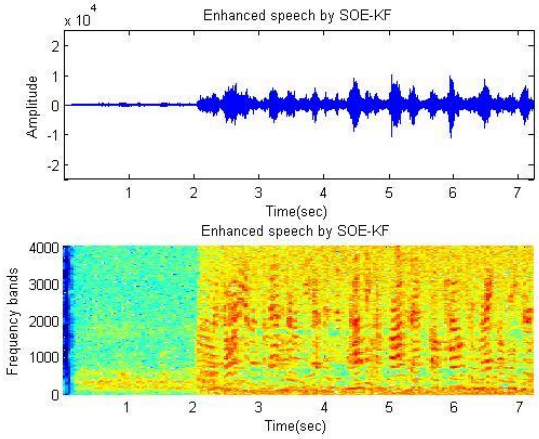
(a) Original Speech



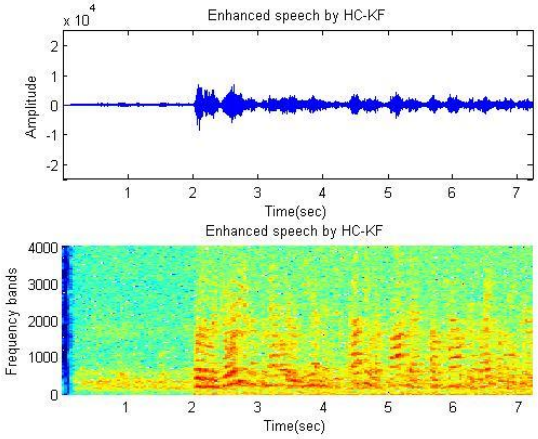
(b) Contaminated Speech



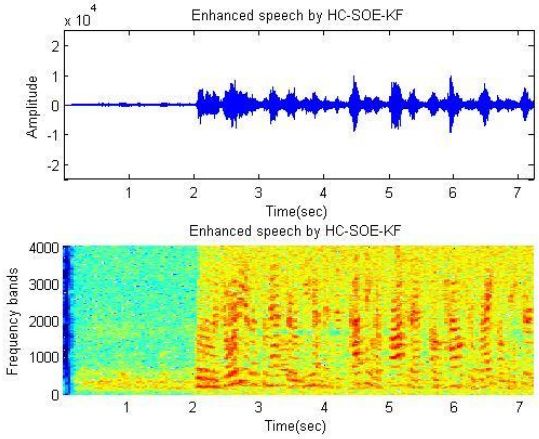
(c) SMI algorithm



(d) SOE-KF algorithm



(e) HC-KF algorithm



(f) HC-SOE-KF algorithm

Fig. 53 Filtering results in input avgSINR 5(dB) (V+V).

Chapter 5. Conclusion and Future Study

The results of performance in speech enhancement are presented in chapter 4. The proposed beamformer null-constraint could be used to improve the OutputSINR both in HC-KF and SOE-KF. Using the wideband selection of steering vector bound in constrained Kalman filter works in solving signal mismatch problem. In this thesis, the fact can be observed that the proposed method is capable of noise reduction and maintaining the signal distortionless at the same time. The robustness of beamformer against array steering error compared to other algorithms is shown in Fig. 42-44;48-50. The relationship between the noise reduction and the dereverberation is a tradeoff of a beamformer. Because the existence of desired source from different direction is unknown and the environment reverberation is non-stationary, the better noise reduction performance may suppress the desired source, particularly in low frequency-bands when the number of arrays is small. To make sure the directivity of steered response is in low frequency-bands, the regulation of steering vector bound is needed according to the characteristic as in Fig. 17-19.

There are several areas for improvement in selecting the appropriate steering vector bound and tracking the beamformer nulls for interference suppression. The inequality nonlinear beam constraint would amplify the desired source and diffused noise at meanwhile; and the null constraint would make a distortion to the desired source in low frequency-bands. Therefore, it would be interesting in noise post-filtering using broadband beamforming similar to [29][30]. This is left as a future research topic. As a conclusion, a more robust adaptive post-filter for diffused noise and more solid sources tracking (similar to superdirective beamformer [34] in low frequency-bands) will be helpful to overcome the signal mismatch problem under such environment further.

REFERENCE

[1] D. Johnson and D. Dudgeon, Array Signal Processing: “Concepts and Techniques,” Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[2] D. H. Johnson and D. E. Dudgeon, Array Signal Processing., Prentice Hall, Englewood Cliffs, NJ, 1993.

[3] B. Darren Ward and A. Rodney Kennedy and C. Robert Williamson, “ConstantDirectivity Beamforming,” *Microphone Arrays Digital Signal.*, pp. 3-17, Springer Berlin Heidelberg, 2001.

[4] B. D. Van Veen and K. M. Buckly, “Beamforming: a versatile approach to spatial filtering,” *IEEE Acoustic, Speech, Signal Processing Magazine.*, pp. 4-24, Apr. 1988.

[5] O. L. Frost, “An algorithm for linearly constrained adaptive array processing,” *Proc. IEEE.*, vol. 60, no. 8, pp. 926-935, Aug. 1972.

[6] J. Capon, “High resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE.*, vol. 57, no. 8, pp. 1408-1418, August. 1969.

[7] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. on Antennas Propagation.*, vol. AP-30, pp. 27-34, Jan. 1982.

[8] J. W. Kim and C. K. Un, “An adaptive array robust to beam pointing error,” *IEEE Trans. Signal Processing.*, vol. 40, no. 6, pp. 1582-1584, Jun. 1992.

[9] N. K. Jablon, “Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections,” *IEEE Trans. Antennas Propagation.*, vol. AP-34, no. 8, pp. 996-1012, Aug. 1986.

[10] A. B. Gershman, Y. Hua and Q. Cheng, Eds., “Robustness issues in adaptive beamforming and high-resolution direction finding, in High-Resolution and

Robust Signal Processing," pp. 63-110, Marcel Drekker, New York, 2003.

[11] M. L. R. de Campos, S. Werner, J. Antonio Apolinário, Jr., "Constrained Adaptation Algorithms Employing Householder Transformation On Signal Process," *IEEE Transactions on Signal Processing.*, vol. 50, no. 9, September 2002.

[12] Y. H. Chen, C. T. Chiang, "Adaptive beamforming using the constrained Kalman filter," *IEEE Trans. Antennas Propagation*., vol. 41, no. 11, pp. 1576–1580, Nov. 1993.

[13] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory Appl*., 2010, Vol. 4, Iss. 8, pp. 1303–1318.

[14] C. Jiang and Y. Zhang, "Some Results on Kalman Filtering with Linear Equality State Constraints," *2011 6th IEEE Conference on Industrial Electronics and Applications*.

[15] N. Gupta, "Mathematically Equivalent Approaches for Equality Constrained Kalman Filtering," *arXiv:0902.1565v1 [math.OC] 10 Feb 2009*

[16] S. Eng Nai and W. Ser and Z. Liang Yu and H. Chen, "Iterative Robust Minimum Variance Beamforming," *IEEE Transactions on Signal Processing.*, vol. 59, no. 4, April 2011.

[17] A. Sergiy Vorobyov and B. Alex Gershman and Zhi-Quan Luo, "Robust Adaptive Beamforming Using Worst-Case Performance Optimization," *IEEE Transactions on Signal Processing.*, vol. 51, no. 2, February 2003.

[18] A. El-Keyi and T. Kirubaraj and A. B. Gershman, "Robust adaptive beamforming based on the Kalman filter," *IEEE Transactions on Signal Processing.*, vol. 53, no. 8, August 2005.

[19] J. Li, P. Stoica and Z. Wang, “On Robust Capon Beamforming and Diagonal Loading,” *IEEE Transactions on Signal Processing.*, vol. 51, no. 7, July 2003.

[20] R. G. Lorenz and S. P. Boyd, “Robust minimum variance beamforming,” *IEEE Transactions on Signal Processing.*, vol. 53, no. 5, May 2005.

[21] H. Jwu-Sheng and L. Ming-Tang and Y. Chia-Hsin, “Robust Adaptive Beamformer for Speech Enhancement Using the Second-Order Extended H∞ Filter,” *IEEE Transactions on Audio, Speech, and Language Processing.*, vol. 21, no. 1, January 2013.

[22] W. Liu, S. Weiss, Wideband Beamforming Concepts and Techniques, 2rd ed., Wireless Communications and Mobile Computing vol. 17, John Wiley & Sons, 2010.

[23] H. L. V. Trees, Detection, Estimation, and Modulation Theory, Optimum Array Processing. Detection, Estimation, and Modulation Theory vol. 4, John Wiley & Sons, 2004.

[24]B. D. Carlson, “Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays,” *IEEE Transactions on Aerospace and Electronic Systems.*, vol. 24. no 4 July 1988.

[25] D. D. Feldman and L. J. Griffiths, “A projection approach for robust adaptive beamforming,” *IEEE Transactions on Signal Processing.*, vol. 42, no. 4, April 1994.

[26] N. Gupta, R. Hauser, “Kalman Filtering with Equality and Inequality State Constraints,” *arXiv:0709.2791v1 [math.OC] 18 Sep 2007*.

[27] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra2, Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs ITU-T P.862, 2001.

[28] C. Yu-Cheng, “Adaptive Beamformer for Speech Enhancement Using Kalman Filter with Reference Signal Tracking,” September 2011.

[29] M. Souden, J. Benesty, and S. Affes, “A study of the LCMV and MVDR noise reduction filters,” *IEEE Trans. Signal Process*,. vol. 58, no. 9, pp. 4925-4935, Sep. 2010.

[30] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, “New insights into the MVDR beamformer in room acoustics,” *IEEE Trans. Audio, Speech, Lang. Process*., vol. 18, no. 1, pp.158-170, Jan. 2010.

[31] E. A. P. Habets, Room impulse Response Generator Technische Universiteit Eindhoven, 2006, Tech. Rep.

[32] G. Lathoud and I.A. McCowan, “A Sector-Based Approach for Localization of Multiple Speakers with Microphone Arrays,” in *Proc*. SAPA 2004, Oct. 2004.

[33] G. Lathoud, “*Spatial-Temporal Analysis of Spontaneous Speech with Microphone Arrays*”, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne, Dec. 2006.

[34] S. Doclo, and M. Moonen, “Superdirective Beamforming Robust Against Microphone Mismatch,” *IEEE Trans. on Audio,Speech, and Language Processing*,. vol. 15, no. 2, Feb. 2007.