Robust Adaptive Beamformer for Speech Enhancement Using the Second-Order Extended H_∞ Filter

Jwu-Sheng Hu, *Member, IEEE*, Ming-Tang Lee, and Chia-Hsin Yang, *Student Member, IEEE*

***Abstract*—This paper presents a novel approach to implement the robust minimum variance distortionless response (MVDR) beamformer. The robust MVDR beamformer is based on the optimization of worst-case performance and provides an excellent robustness against an arbitrary but norm-bounded desired signal steering vector mismatch. For real-time consideration, the beamformer was formulated into state-space observer form and the second-order extended (SOE) Kalman filter was derived. However, the SOE Kalman filter assumes an accurate system dynamic and statistics of the noise signals. These assumptions limit the performance under uncertainties. This paper develops the SOE H_∞ filter for the implementation of the robust MVDR beamformer. The estimation criterion in the SOE H_∞ filter design is to minimize the worst possible effects of the disturbance signals on the signal estimation errors without a prior knowledge of the disturbance signals statistics. Experimental results demonstrate the performance of the proposed algorithm in a noisy and reverberant environment and show its superiority of the robustness against mismatches over the robust MVDR beamformer based on the SOE Kalman filter.**

***Index Terms*—Beamformer, beamforming, H_∞ filter, Kalman filter, robust MVDR beamformer.**

I. Introduction

SPEECH enhancement algorithms have attracted a great deal of interest in the past three decades since the desired speech signal is usually contaminated by background noise and influenced by reverberation (see [1] and references therein). Among several existing speech enhancement algorithms, microphone array beamformers are commonly used for hands-free speech communication or recognition. To cope with environmental changes, various adaptive beamformers were proposed to improve the performance. One of the key issues in adaptive beamformers is the sensitivity due to the mismatch between the actual steering vector of the desired signal and presumed one [2], [3]. The mismatch can be induced by signal point errors [4], imperfect array calibration [5], or the channel effect (e.g., near-far problem [6], environment heterogeneity [7] and source local scattering [8]). In the presence of these effects, an adaptive beamformer can easily mix up the desired signal and interference components; that is, it suppresses the desired signal instead of maintaining the desired response with minimal distortion. This phenomenon is commonly referred as *signal self-nulling* [9]. As a result, several efforts have been devoted to design robust adaptive beamformers [2].

Modifications to classic beamformer techniques were extensively studied to enhance the robustness. The linear constrained minimum variance (LCMV) beamformer was proposed in [10] to minimize the array output power under a look-direction constraint. Another popular technique is the generalized sidelobe canceler (GSC) algorithm which essentially transforms the LCMV constrained minimization problem into an unconstrained one [11]. In the last decade, several techniques addressing the problem of the mismatch of the steering vector in the LCMV or GSC structure were developed [12]–[15]. Further, some *ad hoc* approaches were discussed to overcome the arbitrary desired signal mismatches, such as the diagonal loading of the sample covariance matrix [16], [17] and the eigenspace-based beamformer [18], [19]. An alternative research direction to mitigate the problem of mismatch is to abandon the delay-only propagation assumption and explicitly model the acoustic signal propagation from the source to the sensors by transfer functions (TF) [20], [21]. Affes *et al.* replace the steering vector of simple delay with unknown finite-impulse response (FIR) filters, whose coefficients are determined by an adaptive principal eigenvector tracking algorithm [22]. Instead of estimating the TF's, Gannot *et al.* considered the TF ratio between each sensor pair, and several adaptive beamformer algorithms using the GSC structure have been proposed [23]–[25]. The major concern for both TF's and TF ratios is the need of a pre-training procedure. This might limit their applications, especially under a dynamic environment.

Most of the early methods of robust adaptive beamformers are rather *ad hoc* in that the choice of parameters or the structural modifications is not directly related to the uncertainty of the steering vector [2]. Recently, more rigorous approaches were proposed to cope with unknown mismatches via worst-case optimization [3], [26]. Unlike the earlier methods, they make explicit use of the uncertainty set of the steering vector. The work in [3] minimizes the output interference-plus-noise power while maintaining a distortionless response for the worst-case steering vector mismatch. The robust minimum variance distortionless response (MVDR) problem was formulated as a

Manuscript received July 16, 2011; revised December 23, 2011; accepted August 03, 2012. Date of publication August 15, 2012; date of current version October 18, 2012. This work was supported in part by the National Science Council, Taiwan, under Grant # NSC 99-2622-E-009-005-CC2. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sharon Gannot.

The authors are with the Department of Electrical Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, (e-mail: jshu@cn.nctu.edu.tw; lhoney.ece97g@g2.nctu.edu.tw; chyang.ece92g@nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2012.2213246

second-order cone program and solved in polynomial time via the interior point method. A number of extensions of the robust MVDR beamformer of [3] have been considered [27]–[30]. However, the main shortcoming of these extensions is that they do not have a computationally efficient online implementation. It was shown in [49] that the MVDR problem can be formulated into a state-space form where both the performance index and constraint are considered as the outputs. Standard Kalman filtering technique was applied to obtain a recursive algorithm [49]. Following the similar idea, El-Keyi *et al.* [31] developed the state-space model of the robust MVDR beamformer and derived the second-order extended (SOE) Kalman filter for recursive implementation.

The SOE Kalman filter assumes that the dynamics of the signal generating processes are known, so are the statistical properties of noise signals (i.e., uncorrelated and zero-mean Gaussian with known covariance) [32]. However, these assumptions limit the performance since the complex acoustic dynamics is difficult to model and the uncorrelated zero-mean Gaussian noise assumption is quite stringent considering the variety of environmental interferences. To relax these assumptions, this paper proposes the SOE H_∞ filter for the MVDR beamformer of [3] that requires no prior knowledge of the noise statistics but bounded energy. Several studies on the linear and nonlinear H_∞ filter or mixed Kalman/ H_∞ filter have been presented [33]–[42]. Despite these efforts to expand the use of H_∞ filter to different domains for robustness, there are still very few works using H_∞ filter to cope with the model uncertainty in adaptive beamformer [44].

In this paper, the SOE H_∞ filter is used to implement the robust MVDR beamformer [3] in the frequency domain. The estimation criterion in the SOE H_∞ filter design is to minimize the worst possible effects of the unknown disturbance signals (initial condition error, process noise and measurement noise) on the signal estimation errors. This estimation criterion makes the SOE H_∞ filter more suitable for speech enhancement in the cases of unknown noise statistics, steering vector uncertainty and microphone mismatch. To derive the SOE H_∞ filter, the second-order Taylor series expansion is used to approximate the nonlinear function involved in the beamformer. However, the quadratic terms appear in the series expansion are too complex to make the solution tractable. In this work, they are approximated by the estimation error sample covariance matrix which effectively simplifies the problem. The proposed robust MVDR beamformer based on the SOE H_∞ filter is implemented in frequency domain and applied to the acoustic environment. The experiments are performed in a noisy and reverberant environment and the experimental results reveal the proposed beamformer's superiority of the robustness against mismatches for speech enhancement over the robust MVDR beamformer based on the SOE Kalman filter.

The paper is organized as follows. The speech enhancement problem and background information on adaptive beamformer and robust MVDR beamformer of [3] are presented in Section II. In Section III, the SOE Kalman filter for the implementation of the robust MVDR beamformer of [3] is briefly reviewed and followed by the introduction of the proposed SOE H_∞ filter. In Section IV, the selection of the weighting matrices and performance bounds are studied. In Section V, the performance of the SOE Kalman filter and the SOE H_∞ filter for speech enhancement is evaluated. Finally, conclusions are drawn in Section VI.

II. Problem Formulation

Consider P speech sources and M microphones in the reverberant and noisy environment ($M > P$). The received signal of the m-th microphone can be written as:

$$x_m(t) = \sum_{p=1}^{P} a_{mp}(t) \otimes s_p(t) + n_m(t) \tag{1}$$

where t is the discrete-time index and each symbol in (1) represents:

$\otimes$	convolution operation;
$a_{mp}(t)$	the transfer function from the p-th sound source to the m-th microphone;
$s_1(t)$	the desired speech signal;
$s_2(t) \sim s_P(t)$	the nonstationary interfering speech signals (competing speech signals);
$n_m(t)$	the (directional or omni-directional) stationary noise of the m-th microphone.

Typically, the impulse response function $a_{mp}(t)$ is assumed to be time-invariant over the observation period. In this paper, the competing speech signals, $s_2(t) \sim s_P(t)$, are regarded as interference signals. Applying the short time Fourier transform (STFT) operation to (1) yields:

$$X_m(k,\omega) = \sum_{p=1}^{P} A_{mp}(\omega) S_p(k,\omega) + N_m(k,\omega) \tag{2}$$

where k is the frame number and ω is the frequency index. $X_m(k,\omega)$, $S_p(k,\omega)$ and $N_m(k,\omega)$ are the STFT of the respective signals, which are complex-valued. $A_{mp}(\omega)$ is the approximation of the STFT of the impulse response $a_{mp}(t)$ since the length of the impulse response is generally infinite. The beamformer output is given by

$$Y_{\mathrm{MV}}(k,\omega) = \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega) \boldsymbol{X}(k,\omega) \tag{3}$$

where $^{\mathrm{H}}$ denotes conjugation transpose; $\boldsymbol{X}(k,\omega) = [X_1(k,\omega), \ldots, X_M(k,\omega)]^{\mathrm{T}}$ and $\boldsymbol{w}_{\mathrm{MV}}(\omega)$ is the beamformer weights. The well-known MVDR beamformer minimizes the output power of interference-signals-plus-stationary-noise while maintaining a distortionless response to the desired signal. The frequency domain MVDR problem is given by

$$\min_{\boldsymbol{w}_{\mathrm{MV}}} \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega) \mathbf{R}_{xx}(\omega) \boldsymbol{w}_{\mathrm{MV}}(\omega) \text{ subject to } \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega) \tilde{\boldsymbol{A}}(\omega) = 1 \tag{4}$$

where

$$\mathbf{R}_{xx}(\omega) = E\left\{ \boldsymbol{X}(k,\omega) \boldsymbol{X}^{\mathrm{H}}(k,\omega) \right\}. \tag{5}$$

$\mathbf{R}_{xx}(\omega)$ is the $M \times M$ correlation matrix and $\tilde{\boldsymbol{A}}(\omega) \in \mathbb{C}^{M\times 1}$ is the presumed steering vector. The solution of the MVDR problem is given by [43],

$$\boldsymbol{w}_{\mathrm{MV}}(\omega) = \frac{\mathbf{R}_{xx}^{-1}(\omega)\tilde{\boldsymbol{A}}(\omega)}{\tilde{\boldsymbol{A}}^{\mathrm{H}}(\omega)\mathbf{R}_{xx}^{-1}(\omega)\tilde{\boldsymbol{A}}(\omega)}. \tag{6}$$

In practice, the correlation matrix is unavailable and is usually approximated by

$$\hat{\mathbf{R}}_{xx}(\omega) = \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{X}(k,\omega)\boldsymbol{X}^{\mathrm{H}}(k,\omega) \tag{7}$$

where N is the number of frames available. The sample correlation matrix of (7) is used in (6) to replace the true correlation matrix and the resulting solution is commonly referred to the sample matrix inversion (SMI) algorithm [42]. If the desired signal is present in the training procedure, the SMI algorithm degrades dramatically [3]. The other disadvantage of the SMI algorithm is that it does not provide the sufficient robustness against a mismatch between the presumed steering vector and the actual one. The norm of the mismatch can be bounded by some known constant $\varepsilon > 0$ and the actual steering vector belongs to the set

$$\Lambda(\omega) \equiv \left\{ \boldsymbol{C}(\omega) |\, \boldsymbol{C}(\omega) = \tilde{\boldsymbol{A}}(\omega) + \boldsymbol{e}(\omega), \|\boldsymbol{e}(\omega)\| \le \varepsilon \right\}. \tag{8}$$

Based on this uncertainty description, Vorobyov *et al.* [3] formulated the robust MVDR beamforming problem as

$$\begin{aligned} &\min_{\boldsymbol{w}_{\mathrm{MV}}} \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\hat{\mathbf{R}}_{xx}(\omega)\boldsymbol{w}_{\mathrm{MV}}(\omega) \\ &\qquad\qquad \text{for all } \boldsymbol{C}(\omega)\in\Lambda(\omega) \\ &\text{subject to } \left|\boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\boldsymbol{C}(\omega)\right| \ge 1. \end{aligned} \tag{9}$$

The semi-infinite non-convex constraint in (9) was reformulated as a single constraint considering the worst-case. This leads to the following.

$$\begin{aligned} &\min_{\boldsymbol{w}_{\mathrm{MV}}} \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\hat{\mathbf{R}}_{xx}(\omega)\boldsymbol{w}_{\mathrm{MV}}(\omega) \\ &\text{subject to } \min_{\boldsymbol{C}(\omega)\in\Lambda(\omega)} \left|\boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\boldsymbol{C}(\omega)\right| \ge 1. \end{aligned} \tag{10}$$

It can be proven that the inequality constraint in (10) can be replaced by an equality one as [3],

$$\begin{aligned} &\min_{\boldsymbol{w}_{\mathrm{MV}}} \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\hat{\mathbf{R}}_{xx}(\omega)\boldsymbol{w}_{\mathrm{MV}}(\omega) \\ &\text{subject to } \left|\boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\tilde{\boldsymbol{A}}(\omega) - 1\right|^2 = \varepsilon^2\boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\boldsymbol{w}_{\mathrm{MV}}(\omega). \end{aligned} \tag{11}$$

The problem in (11) has been solved in [3] using second-order cone (SOC) programming. Moreover, several extensions of the robust MVDR beamformer have been considered. For example, a Newton-type iterative method was proposed for this problem and its modification [27], [28]. Re-formulating (11) into a state-space observer form facilitates the application of the SOE Kalman filter [31]. In the following, we briefly review the SOE Kalman filter solution and present a new approach based on the SOE H_∞ filter.

III. Robust MVDR Beamformer Based on the SOE Kalman and the SOE H_∞ Filter

For the convenience of analysis, the mean square error (MSE) between the zero signal and the filtered output is introduced as,

$$E\left[\left|0 - \boldsymbol{X}^{\mathrm{H}}(k,\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega)\right|^2\right] = \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(\omega)\mathbf{R}_{xx}(\omega)\boldsymbol{w}_{\mathrm{MV}}(\omega) \tag{12}$$

where $E(\cdot)$ denotes the expectation operation. The constraint in (11) can be rewritten as

$$g_2(\boldsymbol{w}_{\mathrm{MV}}(k,\omega)) = 1 \tag{13}$$

where

$$\begin{aligned} g_2(\boldsymbol{w}_{\mathrm{MV}}(k,\omega)) = {} & \varepsilon^2\boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(k,\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega) \\ & - \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(k,\omega)\tilde{\boldsymbol{A}}(\omega)\tilde{\boldsymbol{A}}^{\mathrm{H}}(\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega) \\ & + \boldsymbol{w}_{\mathrm{MV}}^{\mathrm{H}}(k,\omega)\tilde{\boldsymbol{A}}(\omega) + \tilde{\boldsymbol{A}}^{\mathrm{H}}(\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega). \end{aligned} \tag{14}$$

Therefore, the robust MVDR beamformer problem can be formulated as

$$\begin{aligned} &\min_{\boldsymbol{w}_{\mathrm{MV}}} E\left[\left|0 - \boldsymbol{X}^{\mathrm{H}}(k,\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega)\right|^2\right] \\ &\text{subject to } g_2(\boldsymbol{w}_{\mathrm{MV}}(k,\omega)) = 1. \end{aligned} \tag{15}$$

The state-space model of the constraint minimization problem in (15) is:

State equation

$$\boldsymbol{w}_{\mathrm{MV}}(k+1,\omega) = \boldsymbol{w}_{\mathrm{MV}}(k,\omega) + \boldsymbol{v}_s(k,\omega). \tag{16}$$

Measurement equation

$$\begin{aligned} \overline{\boldsymbol{y}} &= \begin{bmatrix} \boldsymbol{X}^{\mathrm{H}}(k,\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega) \\ g_2(\boldsymbol{w}_{\mathrm{MV}}(k,\omega)) \end{bmatrix} + \boldsymbol{v}_m(k,\omega) \\ &= \boldsymbol{g}(\boldsymbol{w}_{\mathrm{MV}}(k,\omega)) + \boldsymbol{v}_m(k,\omega) \end{aligned} \tag{17}$$

where $\boldsymbol{v}_s(k,\omega)$ and $\boldsymbol{v}_m(k,\omega)$ are the process and measurement noise respectively, and the measurement vector $\overline{\boldsymbol{y}} = \begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathrm{T}}$ is chosen for the MVDR problem.

A. Second-Order Extended Kalman Filter Algorithm

To apply the SOE Kalman filter, the noise processes $\boldsymbol{v}_s(k,\omega)$ and $\boldsymbol{v}_m(k,\omega)$ are assumed to be white, zero-mean, uncorrelated, and have known covariance matrices $\widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{R}}$ respectively.

$$\begin{aligned} E\left[\boldsymbol{v}_s(k,\omega)\boldsymbol{v}_s^{\mathrm{H}}(k,\omega)\right] &= \widetilde{\mathbf{Q}} \\ E\left[\boldsymbol{v}_m(k,\omega)\boldsymbol{v}_m^{\mathrm{H}}(k,\omega)\right] &= \widetilde{\mathbf{R}} \\ E\left[\boldsymbol{v}_s(k,\omega)\boldsymbol{v}_m^{\mathrm{H}}(k,\omega)\right] &= 0. \end{aligned} \tag{18}$$

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

$$\mathrm{MSE} = E\left[\left|\boldsymbol{w}_{\mathrm{MV}}(k,\omega) - \hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega)\right|^2\right]. \tag{19}$$

To present the SOE Kalman filter solution, we start by evaluating the first derivative of $g(\boldsymbol{w}_{\mathrm{MV}}(k,\omega))$ which is denoted as $\mathbf{G}_{\boldsymbol{w}}(k,\omega)$ and the second derivative of $\boldsymbol{g}(\boldsymbol{w}_{\mathrm{MV}}(k,\omega))$ which is denoted as $\mathbf{G}^{(1)}_{\boldsymbol{ww}}(\omega)$ and $\mathbf{G}^{(2)}_{\boldsymbol{ww}}(\omega)$:

$$\begin{aligned}&\mathbf{G}_{\boldsymbol{w}}(k,\omega)\\&=\left[\nabla_{\boldsymbol{w}_{\mathrm{MV}}}\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{w}_{\mathrm{MV}}(k,\omega))\right]^{\mathrm{T}}\\&=\begin{bmatrix}\boldsymbol{X}^{\mathrm{H}}(k,\omega)\\ \varepsilon^2\boldsymbol{w}^{\mathrm{H}}_{\mathrm{MV}}(k,\omega)-\left(\tilde{\boldsymbol{A}}(\omega)\tilde{\boldsymbol{A}}^{\mathrm{H}}(\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega)\right)^{\mathrm{H}}+\tilde{\boldsymbol{A}}^{\mathrm{H}}(\omega)\end{bmatrix}\end{aligned}\tag{20}$$

$$\begin{aligned}&\mathbf{G}^{(1)}_{\boldsymbol{ww}}(\omega)\\&=\nabla_{\boldsymbol{w}_{\mathrm{MV}}}\nabla^{\mathrm{H}}_{\boldsymbol{w}_{\mathrm{MV}}}\left\{\boldsymbol{X}^{\mathrm{H}}(k,\omega)\boldsymbol{w}_{\mathrm{MV}}(k,\omega)\right\}=0\end{aligned}\tag{21}$$

$$\begin{aligned}&\mathbf{G}^{(2)}_{\boldsymbol{ww}}(\omega)\\&=\nabla_{\boldsymbol{w}_{\mathrm{MV}}}\nabla^{\mathrm{H}}_{\boldsymbol{w}_{\mathrm{MV}}}\left\{g_2(\boldsymbol{w}_{\mathrm{MV}}(k,\omega))\right\}=\varepsilon^2\mathbf{I}-\tilde{\boldsymbol{A}}(\omega)\tilde{\boldsymbol{A}}^{\mathrm{H}}(\omega)\end{aligned}\tag{22}$$

where $\mathbf{I}$ is the identity matrix. For the state space model (16) and (17), the SOE Kalman filter solution is given by [32]

$$\begin{aligned}\hat{\boldsymbol{w}}_{\mathrm{MV}}(k+1,\omega)&=\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega)\\&\quad+\tilde{\mathbf{K}}(k,\omega)\left[\overline{\boldsymbol{y}}-\boldsymbol{g}(\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega))\right]+\boldsymbol{\pi}(k,\omega)\end{aligned}\tag{23}$$

where

$$\boldsymbol{\pi}(k,\omega)=\frac{1}{2}\tilde{\mathbf{K}}(k,\omega)\cdot\begin{bmatrix}0\\1\end{bmatrix}\cdot\mathrm{tr}\left\{\mathbf{G}^{(2)}_{\boldsymbol{ww}}(\omega)\tilde{\mathbf{P}}^{-}(k,\omega)\right\}\tag{24}$$

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

$$\begin{aligned}\tilde{\mathbf{K}}(k,\omega)&=\tilde{\mathbf{P}}^{-}(k,\omega)\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\Big[\mathbf{G}_{\boldsymbol{w}}(k,\omega)\tilde{\mathbf{P}}^{-}(k,\omega)\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\\&\quad+\boldsymbol{\Lambda}(k,\omega)+\tilde{\mathbf{R}}\Big]^{-1}\end{aligned}\tag{25}$$

$$\tilde{\mathbf{P}}^{-}(k,\omega)=\tilde{\mathbf{P}}^{+}(k-1,\omega)+\tilde{\mathbf{Q}}\tag{26}$$

$$\tilde{\mathbf{P}}^{+}(k,\omega)=\left[\mathbf{I}-\tilde{\mathbf{K}}(k,\omega)\mathbf{G}_{\boldsymbol{w}}(k,\omega)\right]\tilde{\mathbf{P}}^{-}(k,\omega)\tag{27}$$

where

$$\boldsymbol{\Lambda}(k,\omega)=\frac{1}{2}\begin{bmatrix}0&0\\0&1\end{bmatrix}\mathrm{tr}\Big\{\mathbf{G}^{(2)}_{\boldsymbol{ww}}(\omega)\tilde{\mathbf{P}}^{-}(k,\omega)\mathbf{G}^{(2)}_{\boldsymbol{ww}}(\omega)\tilde{\mathbf{P}}^{-}(k,\omega)\Big\}\tag{28}$$

is the correction term introduced according to $\boldsymbol{\pi}(k,\omega)$; $\tilde{\mathbf{K}}(k,\omega)$ is the Kalman gain; $\tilde{\mathbf{P}}^{-}(k,\omega)$ is the *a priori* error covariance matrix and $\tilde{\mathbf{P}}^{+}(k,\omega)$ is the *a posteriori* error covariance matrix. After some algebra operations [32], the Kalman gain and covariance matrices $\tilde{\mathbf{P}}^{-}(k,\omega)$ and $\tilde{\mathbf{P}}^{+}(k,\omega)$ can be rewritten as follows,

$$\begin{aligned}&\tilde{\mathbf{K}}(k,\omega)\\&=\tilde{\mathbf{P}}^{-}(k,\omega)\Big[\mathbf{I}+\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\Big(\boldsymbol{\Lambda}(k,\omega)+\tilde{\mathbf{R}}\Big)^{-1}\\&\quad\times\mathbf{G}_{\boldsymbol{w}}(k,\omega)\tilde{\mathbf{P}}^{-}(k,\omega)\Big]^{-1}\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\Big(\boldsymbol{\Lambda}(k,\omega)+\tilde{\mathbf{R}}\Big)^{-1}\end{aligned}\tag{29}$$

$$\begin{aligned}&\tilde{\mathbf{P}}^{-}(k+1,\omega)\\&=\tilde{\mathbf{P}}^{-}(k,\omega)\Big[\mathbf{I}+\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\Big(\boldsymbol{\Lambda}(k,\omega)+\tilde{\mathbf{R}}\Big)^{-1}\\&\quad\times\mathbf{G}_{\boldsymbol{w}}(k,\omega)\tilde{\mathbf{P}}^{-}(k,\omega)\Big]^{-1}+\tilde{\mathbf{Q}}.\end{aligned}\tag{30}$$

B. Second-Order Extended H_∞ Filter Algorithm

In contrast to minimizing the expected value of the estimation error variance like the SOE Kalman filter, another strategy is to minimize the worst possible effects of the disturbances on the signal estimation errors. This corresponds to minimize the infinity norm of the input-output relation. In this case, no assumptions on the noise statistics are required (such as (18)) but the bounds of the noise energy. Considering the state space model (16) and (17), and the estimation of a linear combination of $\boldsymbol{w}_{\mathrm{MV}}(k,\omega)$, i.e.,

$$\boldsymbol{z}(k,\omega)=\mathbf{C}\boldsymbol{w}_{\mathrm{MV}}(k,\omega)\tag{31}$$

where $\mathbf{C}$ is a user-defined matrix. It can be set as an identity matrix ($\mathbf{C}=\mathbf{I}$) to estimate $\boldsymbol{z}(k,\omega)$ (as in the Kalman filter). The estimate of $\boldsymbol{z}(k,\omega)$ is denoted by $\hat{\boldsymbol{z}}(k,\omega)$ and the estimate of initial state $\boldsymbol{w}_{\mathrm{MV}}(0,\omega)$ is denoted by $\hat{\boldsymbol{w}}_{\mathrm{MV}}(0,\omega)$. The design criterion of the SOE H_∞ filter is to find $\hat{\boldsymbol{z}}(k,\omega)$ that minimizes its mean square error under arbitrary but bounded $\boldsymbol{v}_s(k,\omega)$, $\boldsymbol{v}_m(k,\omega)$ and $\boldsymbol{w}_{\mathrm{MV}}(0,\omega)$. The performance index J can be defined as:

$$\begin{aligned}J=&\sum_{k=0}^{N-1}\left|\boldsymbol{z}(k,\omega)-\hat{\boldsymbol{z}}(k,\omega)\right|^2_{\mathbf{S}(k,\omega)}\\&\times\Bigg[\left|\boldsymbol{w}_{\mathrm{MV}}(0,\omega)-\hat{\boldsymbol{w}}_{\mathrm{MV}}(0,\omega)\right|^2_{\mathbf{P}^{-1}(0,\omega)}\\&+\sum_{k=0}^{N-1}\left(\left|\boldsymbol{v}_s(k,\omega)\right|^2_{\mathbf{Q}^{-1}(k,\omega)}+\left|\boldsymbol{v}_m(k,\omega)\right|^2_{\mathbf{R}^{-1}(k,\omega)}\right)\Bigg]^{-1}.\end{aligned}\tag{32}$$

The notation $|\boldsymbol{x}(k,\omega)|^2_{\mathbf{S}(k,\omega)}$ is defined as the square of the weighted (by $\mathbf{S}(k,\omega)$) L_2 norm of $\boldsymbol{x}(k,\omega)$, i.e., $|\boldsymbol{x}(k,\omega)|^2_{\mathbf{S}(k,\omega)}=\boldsymbol{x}^{\mathrm{H}}(k,\omega)\mathbf{S}(k,\omega)\boldsymbol{x}(k,\omega)$. The matrices $\mathbf{P}(0,\omega)$, $\mathbf{Q}(k,\omega)$, $\mathbf{R}(k,\omega)$ and $\mathbf{S}(k,\omega)$ are symmetric positive definite matrices chosen by the user based on the specific problem. To simplify the analysis, we assume the weighting matrices $\mathbf{Q}(k,\omega)$, $\mathbf{R}(k,\omega)$ and $\mathbf{S}(k,\omega)$ are independent of frame and frequency. Furthermore, we set the weighting matrix $\mathbf{S}(k,\omega)=\mathbf{I}$ for the sake of comparing with the estimation-error variances minimization problem of the Kalman filter. Hence, (32) can be reformulated as

$$J=\frac{\displaystyle\sum_{k=0}^{N-1}\left|\boldsymbol{z}(k,\omega)-\hat{\boldsymbol{z}}(k,\omega)\right|^2_{\mathbf{S}}}{\left|\boldsymbol{w}_{\mathrm{MV}}(0,\omega)-\hat{\boldsymbol{w}}_{\mathrm{MV}}(0,\omega)\right|^2_{\mathbf{P}^{-1}(0,\omega)}+\displaystyle\sum_{k=0}^{N-1}\left(\left|\boldsymbol{v}_s(k,\omega)\right|^2_{\mathbf{Q}^{-1}}+\left|\boldsymbol{v}_m(k,\omega)\right|^2_{\mathbf{R}^{-1}}\right)}.\tag{33}$$

To solve the problem, a performance bound γ is selected and $\hat{\boldsymbol{z}}(k,\omega)$ is computed to satisfy

$$\sup J<\gamma\tag{34}$$

where sup represents supremum. The purpose of the SOE H_∞ filter is to minimize sup J, that is, the infinity norm of the criteria. Therefore, it can be interpreted as a *minmax* problem where the estimator strategy $\hat{\boldsymbol{z}}(k,\omega)$ plays against the exogenous inputs $\boldsymbol{v}_s(k,\omega)$, $\boldsymbol{v}_m(k,\omega)$ and the uncertainty of the initial state $\boldsymbol{w}_{\mathrm{MV}}(0,\omega)$ [39]. Hence, the performance criterion in (34) is equivalent to

$$\min_{\hat{\boldsymbol{z}}} \max_{\boldsymbol{v}_s,\boldsymbol{v}_m,\boldsymbol{w}_{\mathrm{MV}}(0,\omega)} J=-\gamma\left|\boldsymbol{w}_{\mathrm{MV}}(0,\omega)-\hat{\boldsymbol{w}}_{\mathrm{MV}}(0,\omega)\right|^2_{\mathbf{P}^{-1}(0,\omega)} + \sum_{k=0}^{N-1}\left[\left|\boldsymbol{z}(k,\omega)-\hat{\boldsymbol{z}}(k,\omega)\right|^2_{\mathbf{S}}-\gamma\left(|\boldsymbol{v}_s(k,\omega)|^2_{\mathbf{Q}^{-1}}+|\boldsymbol{v}_m(k,\omega)|^2_{\mathbf{R}^{-1}}\right)\right]. \quad (35)$$

Since $\boldsymbol{v}_m(k,\omega)=\boldsymbol{y}-\boldsymbol{g}(\boldsymbol{w}_{\mathrm{MV}}(k,\omega))$, $\boldsymbol{z}(k,\omega)=\mathbf{C}\boldsymbol{w}_{\mathrm{MV}}(k,\omega)$, and $\hat{\boldsymbol{z}}(k,\omega)=\mathbf{C}\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega)$, (35) can be rewritten as:

$$\min_{\hat{\boldsymbol{w}}_{\mathrm{MV}}} \max_{\boldsymbol{v}_s,\boldsymbol{v}_m,\boldsymbol{w}_{\mathrm{MV}}(0,\omega)} J=-\gamma|\boldsymbol{w}_{\mathrm{MV}}(0,\omega)-\hat{\boldsymbol{w}}_{\mathrm{MV}}(0,\omega)|^2_{\mathbf{P}^{-1}(0,\omega)} + \sum_{k=0}^{N-1}\Bigg[|\boldsymbol{w}_{\mathrm{MV}}(k,\omega)-\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega)|^2_{\overline{\mathbf{S}}} - \gamma\left(|\boldsymbol{v}_s(k,\omega)|^2_{\mathbf{Q}^{-1}}+|\overline{\boldsymbol{y}}-\boldsymbol{g}(\boldsymbol{w}_{\mathrm{MV}}(k,\omega))|^2_{\mathbf{R}^{-1}}\right)\Bigg] \quad (36)$$

where $\overline{\mathbf{S}}=\mathbf{C}^{\mathrm{H}}\mathbf{S}\mathbf{C}$.

Considering a second-order approximation of the nonlinearity in (36), the solution of (36) leads to the SOE H_∞ filter. The solution of the SOE H_∞ filter for a class of discrete-time nonlinear system has been derived [44] and hence the solution of SOE H_∞ filter for (16) and (17) is given as,

$$\hat{\boldsymbol{w}}_{\mathrm{MV}}(k+1,\omega)=\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega)+\mathbf{K}(k,\omega)\left[\overline{\boldsymbol{y}}-\hat{\boldsymbol{y}}_h(k,\omega)\right] \quad (37)$$

$$\mathbf{K}(k,\omega)=\mathbf{P}(k,\omega)\Bigg[\mathbf{I}-\frac{1}{\gamma}\overline{\mathbf{S}}\mathbf{P}(k,\omega)+\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\mathbf{R}^{-1} \times\mathbf{G}_{\boldsymbol{w}}(k,\omega)\mathbf{P}(k,\omega)\Bigg]^{-1}\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\mathbf{R}^{-1} \quad (38)$$

$$\mathbf{P}(k+1,\omega)=\mathbf{P}(k,\omega)\Bigg[\mathbf{I}-\frac{1}{\gamma}\overline{\mathbf{S}}\mathbf{P}(k,\omega)+\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\mathbf{R}^{-1} \times\mathbf{G}_{\boldsymbol{w}}(k,\omega)\mathbf{P}(k,\omega)\Bigg]^{-1}+\mathbf{Q} \quad (39)$$

$$\lambda(k+1,\omega)=\Bigg[\mathbf{I}-\frac{1}{\gamma}\overline{\mathbf{S}}\mathbf{P}(k,\omega) +\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\mathbf{R}^{-1}\mathbf{G}_{\boldsymbol{w}}(k,\omega)\mathbf{P}(k,\omega)\Bigg]\lambda(k,\omega) -\mathbf{G}^{\mathrm{H}}_{\boldsymbol{w}}(k,\omega)\mathbf{R}^{-1}\left(\overline{\boldsymbol{y}}-\hat{\boldsymbol{y}}_h(k,\omega)\right) \quad (40)$$

$$\overline{\mathbf{P}}(k+1,\omega)=\eta\overline{\mathbf{P}}(k,\omega)+(1-\eta) \times\mathbf{P}(k,\omega)\lambda(k,\omega)\lambda^{\mathrm{H}}(k,\omega)\mathbf{P}^{\mathrm{H}}(k,\omega) \quad (41)$$

where $0<\eta\leq 1$ (we set $\eta=0.9$); $\overline{\boldsymbol{y}}-\hat{\boldsymbol{y}}_h(k,\omega)$ is the error term and the predicted measurement is obtained by

$$\hat{\boldsymbol{y}}_h(k,\omega)=\begin{bmatrix}\mathbf{X}^{\mathrm{H}}(k,\omega)\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega)\\ g_2(\hat{\boldsymbol{w}}_{\mathrm{MV}}(k,\omega))+0.5\cdot\mathrm{tr}\left\{\mathbf{G}^{(2)}_{ww}(\omega)\overline{\mathbf{P}}(k,\omega)\right\}\end{bmatrix}. \quad (42)$$

In this paper, we let the weighting matrices $\widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{R}}$ have the same structure as $\mathbf{Q}$ and $\mathbf{R}$ for comparison,

$$\mathbf{Q}(\omega)=\sigma_s^2\mathbf{I},\mathbf{R}(\omega)=\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}=\sigma_1^2\begin{bmatrix}1 & 0\\ 0 & \xi\end{bmatrix} \quad (43)$$

where the parameters σ_s^2, σ_1^2, σ_2^2, and ξ are discussed in Section IV.

IV. Simulation

In this section, the selection of the weighting matrices and the performance bounds are studied. Here we use the same structure for the weighting matrices in the SOE Kalman filter [31], the first-order extended (FOE) H_∞ filter [32], and the SOE H_∞ filter, as defined in (43). Firstly, the selection of the parameters σ_s^2 and ξ are discussed. Then, the selection of the pair (ρ,ε) is analyzed under different number of sensors and mismatch conditions. Subsequently, the effect of the performance bound γ in the SOE H_∞ filter is studied and compared with the SOE Kalman filter.

The input data are rearranged into frames and frequencies by short-time Fourier transform (STFT) processing. The size of fast Fourier transform (FFT) is set as 256, the shift size is set as 128, and the rectangular window is used for STFT processing. To ensure the spatial characteristics, the simulations 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, $\mathrm{c}\fallingdotseq 346\ \mathrm{m/s}$). Thereafter, the performance of the wideband case for speech enhancement is measured in the next section.

In our simulations, a uniform linear array (ULA) with $M=6$ omnidirectional microphones placed $d=5$ cm apart is used. The ULA is steered toward the direction $\theta=0^\circ$ with steering vector mismatch. Two interferences are produced to impinge toward the array from the directions $\theta=45^\circ$ and $\theta=-45^\circ$, both with 30 dB interference-to-noise ratio (INR). All the aforementioned signals are generated from a Gaussian random generator. To highlight the robustness effect of different selections of the weighting matrices and the performance bounds, the desired signal is assumed to be always present in the training cell with $\mathrm{SNR}=0$ dB (i.e., the same scale as the noise power which results in about -30 dB input SINR). The training size $N=100$ is used for the simulations.

First, the selection of parameters σ_s^2 and ξ are discussed. In weighting matrix $\mathbf{Q}$, σ_s^2 controls the variance of the random walk for the weighting update. Since the testing environment is assumed to be stationary, σ_s^2 is set to zero. Next, we consider the parameter ξ. Recall from (17), σ_1^2 and σ_2^2 represent the variance of filtered output error and constraint error respectively. In [31], 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 $\|\mathbf{w}_{\mathrm{MV}}\|^2(M\sigma^2+\sigma_{\mathrm{n}}^2)$, where σ^2 and σ_{n}^2 are the desired signal and sensor noise powers, respectively. Hence, we define

$$\rho=\frac{\sigma_1^2}{(M\sigma^2+\sigma_{\mathrm{n}}^2)}. \quad (44)$$

And the characteristic of choosing the parameter σ_1^2 is discussed by ρ in the subsequent examples. The latter parameter σ_2^2 should

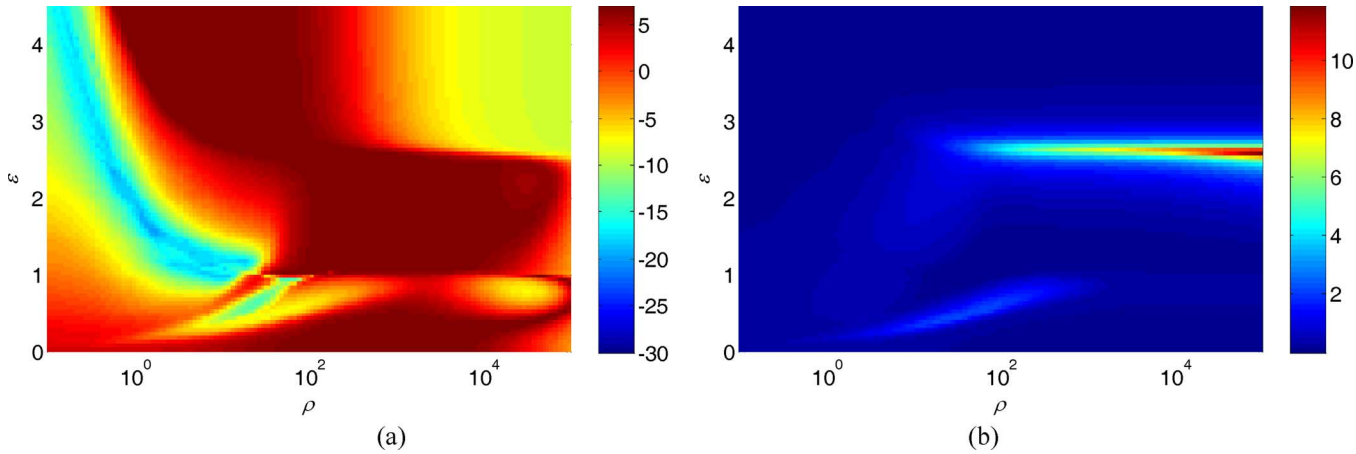

Fig. 1. (a) Output SINR, (b) Squared norm of the weight vector, versus (ρ, ϵ) with $M = 6, \Delta\theta = 2^\circ$. (SOE Kalman filter).

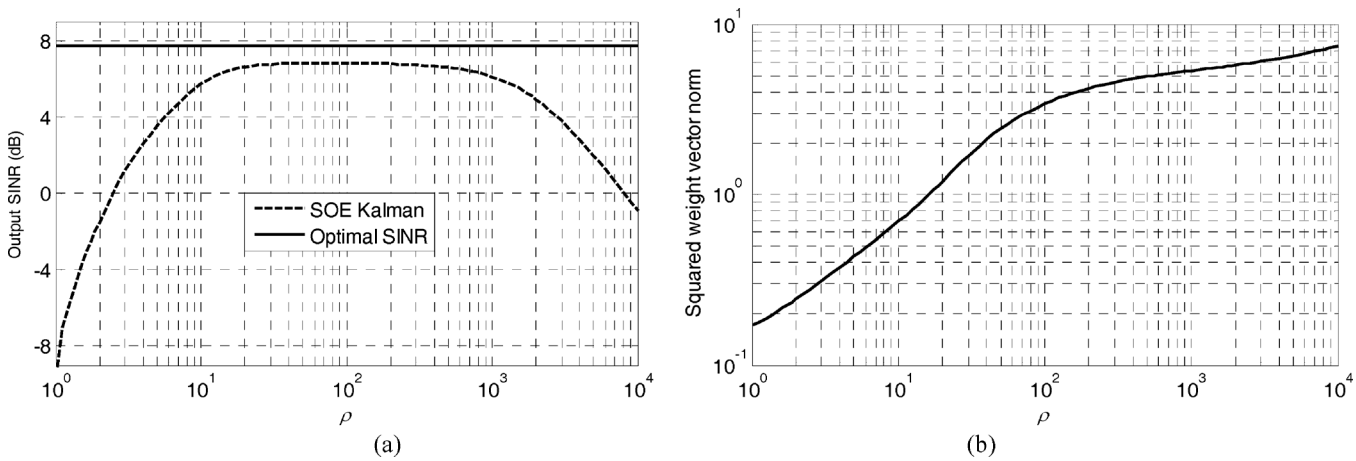

Fig. 2. (a) Output SINR, (b) Squared norm of the weight vector, versus ρ with $\varepsilon = 2.55, M = 6, \Delta\theta = 2^\circ$. (SOE Kalman filter).

be chosen very small to satisfy the robustness constraint with a high accuracy. For the latter one, rather than setting the parameter σ_2^2, we extract the ratio $\xi = \sigma_2^2/\sigma_1^2$ as in (43) for the following numerical and mathematical reasons:

1) ξ determines the condition number of the weighting matrix $\mathbf{R}$.
2) The constrained Kalman algorithm converges to optimum MVDR filter if ξ is small enough [49].

Hence, $\xi < 10^{-8}$ is recommended to fulfill the requirements above (note that ξ should not be set too small in order to keep the weighting matrix $\mathbf{R}$ well-conditioned). The above reasons also give physical meanings to the selection of ξ for the SOE H_∞ filter. But care must be exercised to select this parameter in the SOE H_∞ filter. It should consider the effect of the correct term $\mathbf{\Lambda}(k, \omega)$. The effect of the correct term on the parameter ξ can be ignored only when the state is close to the stationary point.

Second, the choices of the performance bound ε and the parameter ρ $(= \sigma_1^2/(M\sigma^2 + \sigma_n^2))$ of the SOE Kalman filter are studied. In the first example, the true source impinges toward the array from the directions $\theta = 2^\circ$ (direction mismatch $\Delta\theta = 2^\circ$). The parameters of the SOE Kalman filter were selected as $\sigma_s^2 = 0$ and $\xi = 10^{-10}$. Fig. 1 shows the output SINR and the weight vector norm versus different choices of pair (ρ, ε) with $M = 6$. From Fig. 1(a) we can see that there are many choices of pair (ρ, ε) which can achieve high output SINR. However, if we want to keep the output SINR high while maintaining the output power level, then the selection of the performance bound ε becomes critical, as shown in Fig. 1(b). It can be also seen from Fig. 1 that most choices of pair (ρ, ε) with high output SINRs correspond to small vector norms. In Fig. 2, with a proper selection of, $\varepsilon = 2.55$, the output SINR remains close to the optimal SINR for a range about 20 to 1000 of the values ρ, and the weight vector's norm can be controlled between 1 and 6 $(= M)$.

The number of sensors M changes the effect the pair (ρ, ε) on the output SINR and the weighting vector's norm. Besides, as the number of sensors M decreases, the choices of (ρ, ε) to achieve a high output SINR become limited. Moreover, the loss of output SINR or output power is inevitable when the number of sensors M becomes too small to handle the robustness constraint. Fig. 3 gives an example when the number of sensors M reduces to 4.

Fig. 4 shows an example when the mismatch becomes severe with $M = 6$ omnidirectional microphones. Comparing to Fig. 1, it can be observed that the characteristics of the squared norm is almost unchanged; however, the output SINR obviously degraded due to a more severe mismatch. Note that there are still some selections of the pairs (ρ, ε) which can yield high output

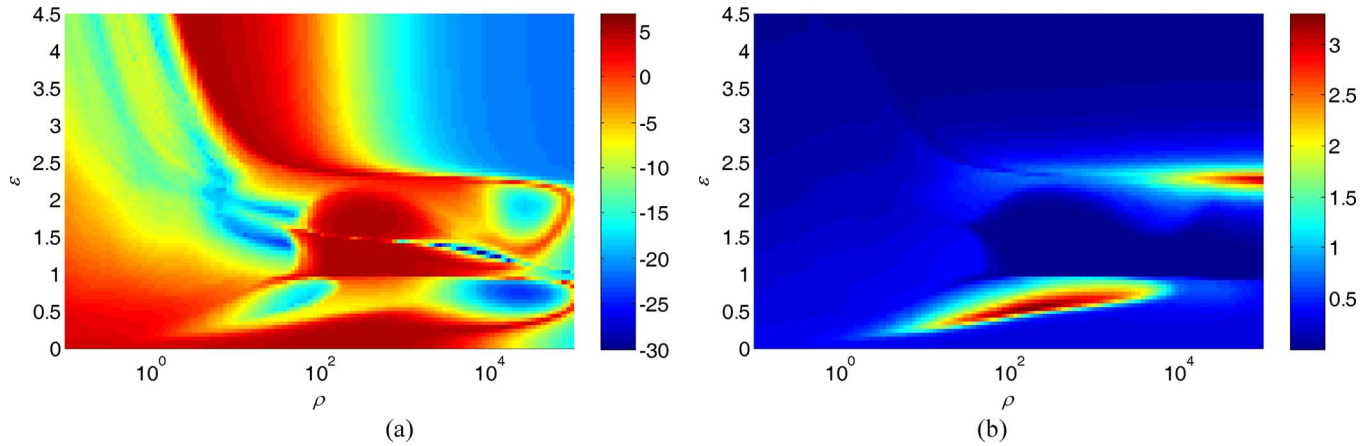

Fig. 3. (a) Output SINR, (b) Squared norm of the weight vector, versus (ρ, ε) with $M = 4$, $\Delta\theta = 2°$. (SOE Kalman filter).

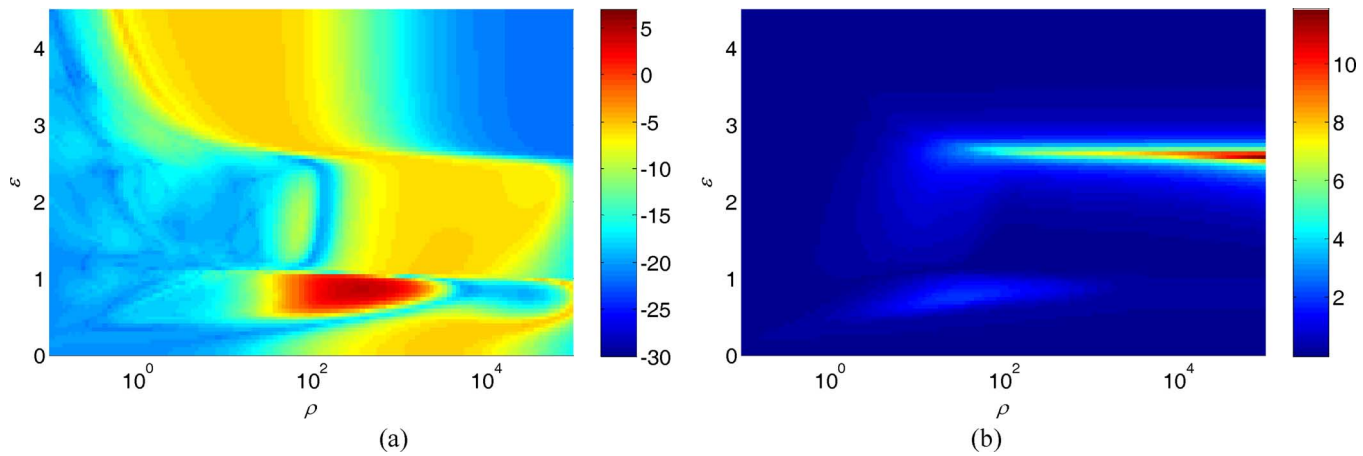

Fig. 4. (a) Output SINR, (b) Squared norm of the weight vector, versus (ρ, ε) with $M = 6$, $\Delta\theta = 16°$. (SOE Kalman filter).

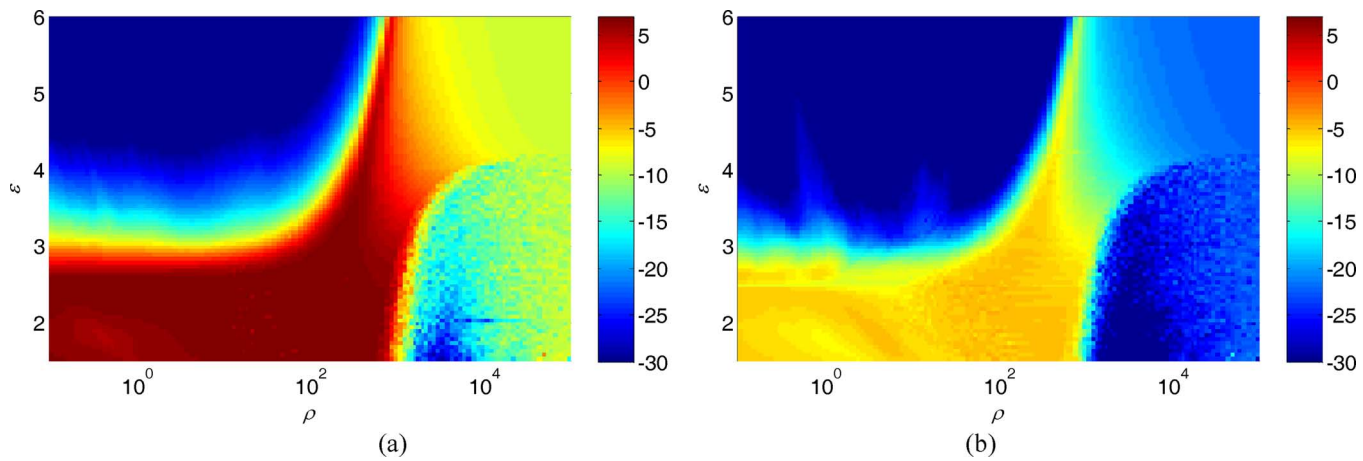

Fig. 5. Output SINR versus (ρ, ε), (a) $\Delta\theta = 2°$, (b) $\Delta\theta = 16°$, with $M = 6$. (FOE H_∞ filter).

SINRs, but these selections are not the same as the cases under slight mismatch, as shown in Fig. 1(a).

In the third example, we compare the sensitivity to the direction mismatch conditions between the SOE Kalman filter, the FOE H_∞ filter, and the SOE H_∞ filter. The previous examples show the selections of the pairs (ρ, ε) of the SOE Kalman filter that yield high output SINRs change with different number of microphones and different mismatch conditions. In this example, we show that with a proper selection of the performance bound γ, the SOE H_∞ filter can be designed comparatively insensitive to the mismatch conditions; however, the FOE H_∞ filter is unable to achieve the performance under the same settings. The selection of the performance bound γ is critical; here we suggest setting $\gamma = 1$ for this example. Note that the setting of initial conditions is another issue for the SOE H_∞ filter. It is important to fulfill the first- and second-order necessary conditions to allow the existence of an optimal solution for the estimated state $\hat{\boldsymbol{w}}_{\mathrm{MV}}$. Besides, the performance bound γ can not be set too small to make the matrix inversion in (38) and (39) ill-conditioned. In Figs. 5 and 6, we analyze two direction mis-

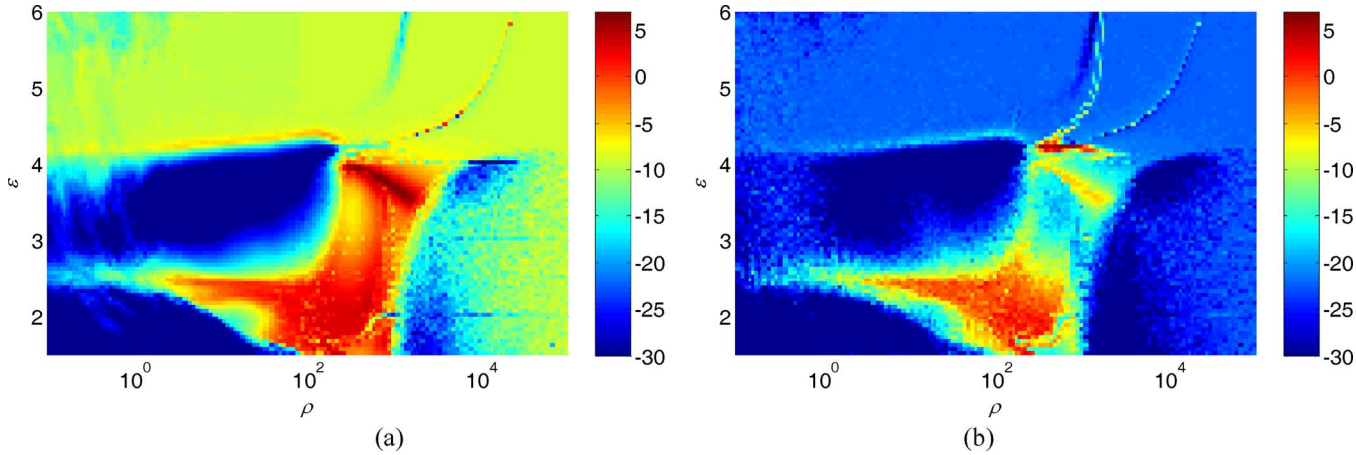

Fig. 6. Output SINR versus (ρ, ε), (a) $\Delta\theta = 2°$, (b) $\Delta\theta = 16°$, with $M = 6$. (SOE H_∞ filter).

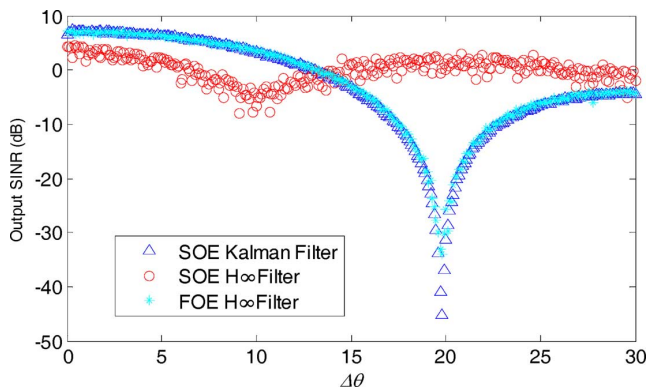

Fig. 7. Comparison between the SOE Kalman filter, FOE H_∞ filter, and the SOE H_∞ filter. Output SINR versus $\Delta\theta$ with $M = 6$.

match conditions in accordance with the SOE Kalman filter case (i.e., $\Delta\theta = 2°$ and $\Delta\theta = 16°$). The parameters of the FOE H_∞ filter and the SOE H_∞ filter were selected as $\sigma_s^2 = 0$, $\xi = 10^{-10}$, and $\gamma = 1$. For the FOE H_∞ filter , it can be seen from Fig. 5 that even the shape of the output SINR does not change with different mismatch conditions, but the value of the output SINR still degrades with larger mismatch. However, for the SOE H_∞ filter case (refer to Fig. 6), it can be observed that the region around (200, 2) maintain good output SINR in both mismatch conditions (which can also be compared with the SOE Kalman filter case in Figs. 1(a) and 4(a)). Fig. 7 gives a comparison between the SOE Kalman filter, the FOE H_∞ filter, and the SOE H_∞ filter. The parameters $\sigma_s^2 = 0$, $\xi = 10^{-10}$, $(\rho, \varepsilon) = (200, 2)$ were set for all the tested algorithms. The performance bound γ of the FOE H_∞ filter and the SOE H_∞ filter were set to 1. In Fig. 7, it can be seen that the output SINRs of both the FOE H_∞ filter and the SOE Kalman filter decrease when the mismatch condition gets worse, while the output SINR of the SOE H_∞ filter can be maintained in a relatively good range. When the mismatch is small, the performance of both the FOE H_∞ filter and the SOE Kalman filter is better than the SOE H_∞ filter. The virtue of the SOE H_∞ filter is the robustness against different mismatch conditions.

For the application in speech enhancement, since the mismatch conditions differ from frequencies, the proposed SOE H_∞ filter can have a better performance in average.

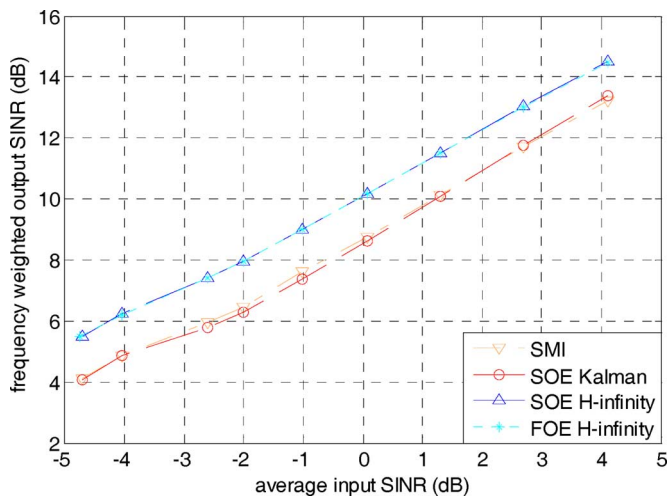

Fig. 8. Frequency weighted output SINR results.

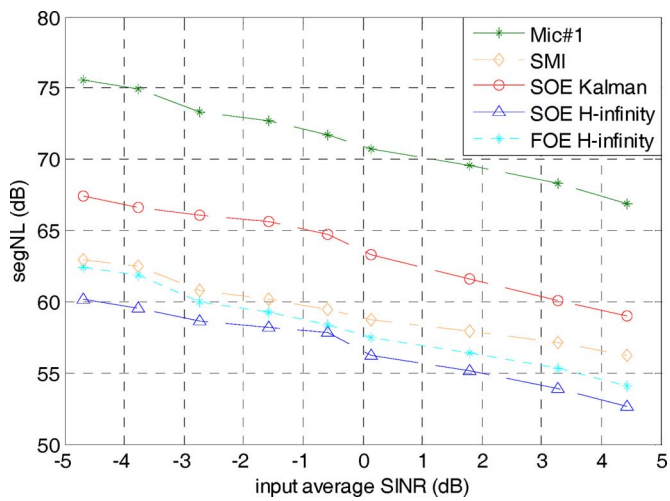

Fig. 9. segNL results.

V. Experimental Results

In this section, experiments in a real room are presented for comparison. Mismatches can easily happen for speech enhancement applications in real environment. For example, the reverberation is one of the common reasons that cause mismatches.

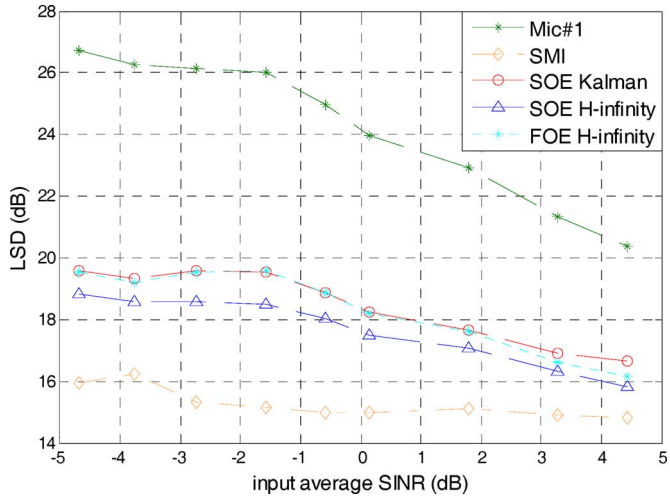

Fig. 10. LSD results.

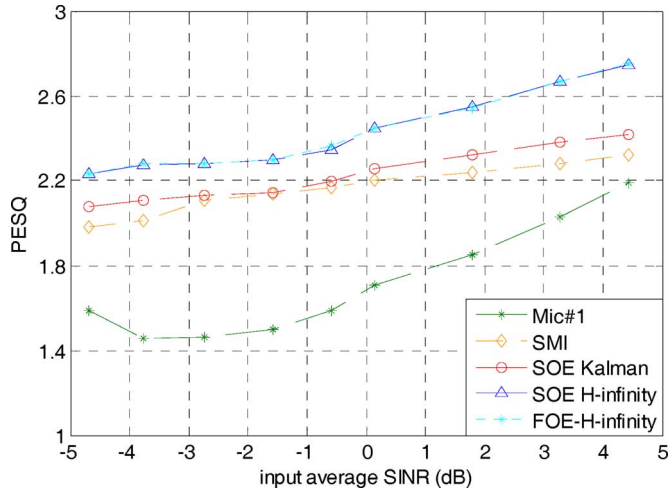

Fig. 11. PESQ results.

Second, the violation of the spatial Nyquist theorem at low frequencies degrades the performance and limits the effective observations in a finite time. Moreover, the speech signals are non-stationary and non-zero-mean, which makes the assumption of the SOE Kalman filter invalid. The mismatch conditions vary among frequencies (including ρ and ε). That means a general selection of pair (ρ, ε) to keep high output SINR across all frequency bands may not exist. According to the simulation results in the previous section, there is a selection of the performance bound γ that makes the SOE H_∞ filter comparatively robust to the variation of the mismatches than the SOE Kalman filter. In this case, the performance bound γ should be increased to handle the mismatch conditions and to guarantee the matrix inversion in (38) and (39) to be well-conditioned (here we set $\gamma = 100$). In the following, we show the superiority of the SOE H_∞ filter for speech enhancement using several performance criteria of speech enhancement.

The room dimension is 10 m × 6 m × 3.6 m and the reverberation time at 1000 Hz is 0.52 second. A uniform linear array (ULA) with $M = 4$ omnidirectional microphones was placed on a table at a distance of 2 m from the wall. The sampling rate, STFT process are the same as the simulation. According to the investigation of room acoustics [46], the number of eigen-frequencies can be obtained by the following equation:

$$\overline{Q} = \frac{4\pi}{3} B \left(\frac{f_s}{2c} \right)^3 \tag{45}$$

where B represents the geometrical volume, f_s denotes the sampling frequency, and c means the sound velocity. This equation indicates that the number of poles is very large when the room volume is high, and that the transient response occurs in almost any processing duration. In this experimental environment, the number of poles is about 1.398×10^6 when the sampling frequency is 8 kHz and the room volume is 216 m^3. Accordingly, the STFT window length ($= 256$) in this experiment is shorter than the channel response duration and is likely to create channel modeling error. The desired speech signal at 0° consists of sentences from TCC-300 database [47] spoken by 150 males and 150 females. The interference signal is speech signal spoken by a male and the noise signal at 45° is the white noise. The amplified microphone signals were digitized by 16-bit AD converters.

Four performance indices are used to measure the waveform property directly. The first is the frequency weighted output signal-to-interference-plus-noise ratio (SINR) defined as

$$\mathrm{SINR}_{avg} = 10 \cdot \log_{10} \frac{\sum_k \sum_{\omega=\omega_1}^{\omega_2} G(\omega) {Y_s}^2(k,\omega)}{\sum_k \sum_{\omega=\omega_1}^{\omega_2} G(\omega) {Y_\mathrm{n}}^2(k,\omega)} \tag{46}$$

where Y_s is the filtered output of the desired signal, and Y_n is the filtered output of interferences plus noise. (ω_1, ω_2) is the frequency bands to be evaluated. $G(\omega)$ is a frequency weighing function, which can be selected to emphasize on the ear's critical bands. Because the robust MVDR beamformer does not consider the de-reverberation effect, the SINR is used to compare the desired signal energy to the interference-plus-noise energy. The second quality measure is segmental noise level (segNL) defined as

$$\mathrm{segNL(dB)} = \frac{1}{K} \sum_{k=1}^{K} \left(10 \cdot \log_{10} \left(\sum_{i=1}^{I} x^2(i + kI) \right) \right) \tag{47}$$

which is evaluated when the desired speech is inactive.

The third quality measure is log spectral distortion (LSD) defined as

$$\mathrm{LSD} = \frac{1}{K} \times \sum_{k=1}^{K} \sqrt{\frac{1}{W} \sum_{\omega=1}^{W} \left(20 \cdot \log_{10} |A_{11}(\omega) S_1(k,\omega)| - 20 \cdot \log_{10} |Y(k,\omega)| \right)^2} \tag{48}$$

where $Y(k,\omega)$ is the STFT of the filtered output. The LSD measures the distance between the desired speech recorded by the first microphone and the algorithm output. Note that a lower LSD level corresponds to a better performance. The last one is the PESQ (Perceptual Evaluation of Speech Quality) [50]. It is a widely accepted industry standard for objective voice quality

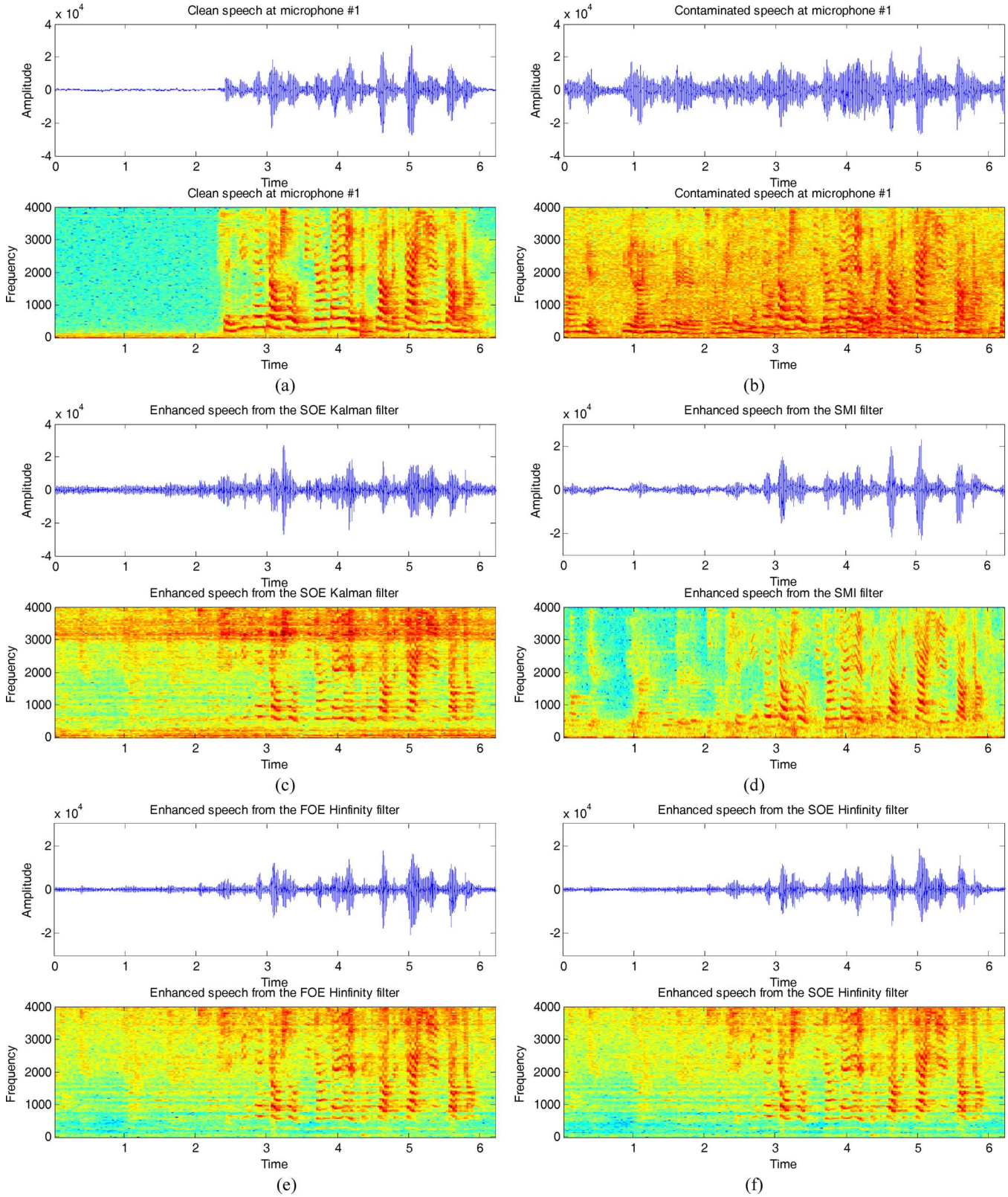

Fig. 12. Waveforms and spectrograms: (a) the clean speech recorded by the first microphone; (b) the contaminated speech at the first microphone (Input $\mathrm{SINR} = 1.7$ dB); (c) the enhanced speech obtained by the SOE Kalman filter; (d) the enhanced speech obtained by the SMI beamformer; (e) the enhanced speech obtained by the FOE H_∞ filter; (f) the enhanced speech obtained by the SOE H_∞ filter.

testing and is used here to access the overall speech enhancement performance.

Since the proposed SOE H_∞ filter is based the MVDR formulation, the classical MVDR beamformer (also known as sample matrix inversion, SMI) [43], the SOE Kalman filter [31], and the first-order extended (FOE) H_∞ filter [32] are implemented for comparison. The noise covariance matrix of the SMI beamformer is estimated during desired source absence and the SMI

beamformer is implemented using (6) and (7). Note that the SMI beamformer does not consider the steering error. The experimental results are shown in Figs. 8–11 according to different input average SINR, and Mic#1 represents the contaminated speech recorded by the first microphone. As can be seen, the proposed SOE H_∞ filter outperforms the SOE Kalman filter and the SMI beamformer in SINR, SegNL and PESQ measures. However, both SOE H_∞ and Kalman filters result in a higher distortion of the original speech waveform than the SMI beamformer as both filters are more aggressive in reducing interferences. This can be observed from the waveforms and spectrograms in Fig. 12. Reducing the interferences, particularly in the low frequency range, also destroys the desired speech signals (Fig. 12(c) and (d)). It is due to high spatial coherence among microphones for small microphone spacing. Note that the SMI beamformer also has the similar effect (blurring of the speech features in spectrogram at low frequency range in Fig. 12(e)). Nonetheless, the proposed SOE H_∞ filter yields the best speech enhancement performance in terms of PESQ. As a final note, the SOE Kalman filter performs poorly in PESQ when the input SINR is low (Fig. 11). This indicates that the algorithm is greatly influenced by model uncertainties (room reverberation and microphone mismatch) under real environment. It shows that the proposed SOE H_∞ filter is more robust in these speech quality measures. The FOE H_∞ filter has similar performance in SINR and PESQ as the SOE H_∞ filter. The main difference can be seen from their LSD (Fig. 10). It shows that SOE H_∞ filter results in less distortion than FOE one under mismatch in real practice.

VI. Conclusion

The SOE H_∞ filter-based robust MVDR beamformer for the acoustic environment has been proposed and the detail derivation of the SOE H_∞ filter has also been given in this paper. For the derivation of the SOE H_∞ filter, the second-order Taylor series expansion is used to approximate the nonlinear function and the second-order term is approximated by the estimation error sample covariance matrix. The SOE H_∞ filter provides a rigorous method for dealing with systems that have model uncertainty and this theoretical advantage has been confirmed by the speech enhancement experiments. Experimental results show that the proposed SOE H_∞ filter-based robust MVDR beamformer outperforms the SOE Kalman filter-based robust MVDR beamformer. The algorithm developed in this work is based on the sub-band decomposition using STFT. It would be interesting to derive its counterpart in wide-band approach similar to the treatment in [48]. This is left as a future research topic.

References

[1] I. Cohen, J. Benesty, and S. Gannot, *Speech Processing in Modern Communication: Challenges and Perspectives*, 1st ed. New York: Springer, 2010, vol. 3, Topics in Signal Processing.

[2] J. Li and P. Stoica, *Robust Adaptive Beamforming*. New York: Wiley, 2006.

[3] S. A. Vorobyov, A. B. Gershman, and Z. Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[4] J. W. Kim and C. K. Un, "An adaptive array robust to beam pointing error," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1582–1584, Jun. 1992.

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

[6] Y. J. Hong, C. C. Yeh, and D. R. Ucci, "The effect of finite-distance signal source on a far-field steering Applebaum array—Two dimensional array case," *IEEE Trans. Antennas Propag.*, vol. 36, no. 4, pp. 468–475, Apr. 1988.

[7] J. Ringelstein, A. B. Gershman, and J. F. Böhme, "Direction finding in random inhomogeneous media in the presence of multiplicative noise," *IEEE Signal Process. Lett.*, vol. 7, no. 10, pp. 269–272, Oct. 2000.

[8] D. Astely and B. Ottersten, "The effects of local scattering on direction of arrival estimation with MUSIC," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3220–3234, Dec. 1999.

[9] A. B. Gershman, "Robustness issues in adaptive beamforming and high-resolution direction finding," in *High-Resolution and Robust Signal Processing*, Y. Hua, A. B. Gershman, and Q. Cheng, Eds. New York: Marcel Dekker, 2003, pp. 63–110.

[10] O. L. Frost, "An algorithm for linear constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[11] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.

[12] H. Cox, R. M. Zeskind, and M. H. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[13] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[14] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process*, vol. 48, no. 2, pp. 386–398, Feb. 2000.

[15] S. Doclo, A. Spriet, J. Wouters, and M. Moone, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Commun.*, vol. 49, no. 7–8, pp. 636–656, 2007.

[16] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.

[17] M. H. Er and T. Cantoni, "An alternative formulation for an optimum beamformer with robustness capability," *Proc. Inst. Elect. Eng. Radar, Sonar, Navig.*, pp. 447–460, Oct. 1985.

[18] L. Chang and C. C. Yeh, "Performance of DMI and eigenspace-based beamformers," *IEEE Trans. Antennas Propagat.*, vol. 40, no. 11, pp. 1336–1347, Nov. 1992.

[19] D. D. Feldman and L. J. Griffiths, "A projection approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 867–876, Apr. 1994.

[20] D. Rabinkin, R. Renomeron, J. Flanagan, and D. F. Macomber, "Optimal truncation time for matched filter array Processing," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, May 1998, pp. 3629–3632.

[21] M. Dahl and I. Claesson, "Acoustic noise and echo canceling with microphone array," *IEEE Trans. Veh. Technol.*, vol. 48, no. 5, pp. 1518–1526, Sep. 1999.

[22] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 425–437, Sep. 1997.

[23] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[24] G. Reuven, S. Gannot, and I. Cohen, "Dual-source transfer-function generalized sidelobe canceller," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 711–727, May 2008.

[25] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[26] Z. L. Yu, W. Ser, M. H. Er, Z. Gu, and Y. Li, "Robust adaptive beamformers based on worst-case optimization and constraints on magnitude response," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2615–2628, Jul. 2009.

[27] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[28] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[29] S. Vorobyov, A. B. Gershman, Z.-Q. Luo, and N. Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 108–111, Feb. 2004.
[30] K. Zarifi, S. Shahbazpanahi, A. B. Gershman, and Z.-Q. Luo, "Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 295–305, Jan. 2005.
[31] A. El-Keyi, T. Kirubarajan, and A. B. Gershman, "Robust adaptive beamforming based on the Kalman filter," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3032–3041, Aug. 2005.
[32] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. New York: Wiley, 2006.
[33] B. Hassibi and T. Kailath, "H_∞ adaptive filtering," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 1995, pp. 949–952.
[34] G. A. Einicke and L. B. White, "Robust extended Kalman filtering," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2596–2599, Sep. 1999.
[35] A. G. Kallapur, I. R. Petersen, and S. G. Anavatti, "A discrete-time robust extended Kalman filter for uncertain systems with sum quadratic constraints," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 850–854, Apr. 2009.
[36] U. Shaked and N. Berman, "H_∞ nonlinear filtering of discrete-time processes," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2205–2209, Sep. 1995.
[37] F. Yang, Z. Wang, Y. S. Hung, and H. Shu, "Mixed H_2/H_∞ filtering for uncertain systems with regional pole assignment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 438–448, Feb. 2005.
[38] W. Li and Y. Jia, "H_∞ filtering for a class of nonlinear discrete-time systems based on unscented transform," *Signal Process.*, vol. 19, no. 12, pp. 3301–3307, 2010.
[39] X. M. Shen and L. Deng, "Game theory approach to discrete H_∞ filter design," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 1092–1095, Apr. 1997.
[40] X. Shen and L. Deng, "A dynamic system approach to speech enhancement using the H_∞ filtering algorithm," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 4, pp. 391–399, Jul. 1999.
[41] J. S. Hu, W. H. Liu, and C. C. Cheng, "Robust beamforming of microphone array using H_∞ adaptive filtering technique," *IEICE Trans. Fundam. of Electron., Commun. Comput. Sci.*, vol. E89-A, no. 3, pp. 708–715, Mar. 2006.
[42] D. Labarre, E. Grivel, M. Najim, and N. Christov, "Dual H_∞ algorithms for signal processing — Application to speech enhancement," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5195–5208, Nov. 2007.
[43] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.
[44] J. S. Hu and C. H. Yang, "Second-order extended H_∞ filter for nonlinear discrete-time systems using quadratic error matrix approximation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3110–3119, Jul. 2011.
[45] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1978.
[46] H. Kuttruf, *Room Acoustics*. London: Elsevier, 1991, ch. 3, p. 56.
[47] "The Association for Computational Linguistics and Chinese Language Processing," [Online]. Available: http://www.aclclp.org.tw/corp.php
[48] A. El-Keyi, T. Kirubarajan, and A. B. Gershman, "Adaptive wideband beamforming with robustness against steering errors," in *Proc. IEEE Workshop Sens. Arrays Multi-Channel Process. (SAM'06)*, Waltham, MA, Jul. 2006, pp. 11–15.
[49] Y. H. Chen and C. T. Chiang, "Adaptive beamforming using the constrained Kalman filter," *IEEE Trans. Antennas Propag.*, vol. 41, no. 11, pp. 1576–1580, Nov. 1993.
[50] Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs 2001, ITU-T P.862.

Jwu-Sheng Hu (M'94) received the B.S. degree from the Department of Mechanical Engineering, National Taiwan University, Taiwan, in 1984, and the M.S. and Ph.D. degrees from the Department of Mechanical Engineering, University of California at Berkeley, in 1988 and 1990, respectively.

From 1991–1993, he was an assistant professor in the Department of Mechanical Engineering, Wayne State University, Detroit, MI where he received the Research Initiation Award from National Science Foundation. Since 1993, he joined the Department of Electrical Engineering, National Chiao Tung University, Taiwan, and became a full professor in 1998. He served as the vice-chairman of the Department since 2006. From 2008; he works in-part at the Industrial Technology Research Institute (ITRI) of Taiwan where he serves as the advisor for the Intelligent Robotics program and principle investigator of several join research projects on robotics across academia, industry and research institutions. He also serves as a research faculty at the National Chip Implementation Center (CIC) of Taiwan for embedded system design applications. He is now the chairman of the Taipei Chapter, IEEE RA Society. His current research interests include robotics, microphone array, mechatronics, and embedded systems.

Ming-Tang Lee was born in Taoyuan, Taiwan, in 1984. He received the B.S. degree and the M.S. degree in Electrical and Control Engineering from National Chiao Tung University, Taiwan in 2007 and 2008 respectively. He is currently a Ph.D. candidate in Department of Electrical and Control Engineering at National Chiao Tung University, Taiwan. His research interests include array signal processing, adaptive signal processing, detection and estimation, and speech enhancement.

Chia-Hsin Yang (S'06) was born in Taipei, Taiwan, in 1981. He received the B.S. degree from the Department of Electrical Engineering of National Chiao Tung University, Taiwan in 2003 and the M.S. and Ph.D. degrees from the Institute of Electrical and Control Engineering of National Chiao Tung University, Taiwan in 2005 and 2010 respectively. He is currently working at MediaTek Inc., Taiwan, as a design engineer for speech signal processing. His research interests include sound source localization, speech enhancement, microphone array signal processing, and adaptive signal processing.