Noise Statistics Update Adaptive Beamformer With PSD Estimation for Speech Extraction in Noisy Environment

Hai Huyen Dam, Hai Quang Dam, and Sven Nordholm, *Senior Member, IEEE*

Abstract—This paper addresses the problem of extracting a desired speech source from a multispeaker environment in the presence of background noise. A new adaptive beamforming structure is proposed for this speech enhancement problem. This structure incorporates power spectral density (PSD) estimation of the speech sources together with a noise statistics update. An inactive-source detector based on minimum statistics is developed to detect the speech presence and to track the noise statistics. Performance of the proposed beamformer is investigated and compared to the minimum variance distortionless response (MVDR) beamformer with or without a postfilter in a real hands-free communication environment. Evaluations show that the proposed beamformer offers good interference and noise suppression levels while maintaining low distortion of the desired source.

Index Terms—Echo cancellation, echo suppression, microphone array, noise suppression, speech enhancement.

I. INTRODUCTION

SPEECH enhancement in hands-free communication environments such as cars has received much attention from researchers and industry due to the popularity of telephone systems [1], [2]. In a car environment, the use of mobile phones while driving is dangerous and in fact prohibited by law in many countries. One approach to overcome this problem is to mount the microphones in fixed positions in a close vicinity of the speaker. The distance between the microphones and the driver must be long enough to give some free space to the driver. Consequently, the observed signal at the microphones includes not only the driver's speech but also interferences from the loudspeaker, which can be mounted unchangeable in the car frame, and the car background noise. Generally speaking, the car background noise comes from the engine, the friction between the chassis and the air, and the friction between the tires and the

roads. The problem is to suppress the undesired sources and the background noise from the noisy observed signal. Many approaches have been suggested with mixed results [3]–[5].

Multichannel speech enhancement techniques are developed when the observed signal is being captured by several microphones. These techniques make use of both spatial and spectral properties of the observations [6]. Generally speaking, these techniques can be largely divided into three types, namely, fixed, optimum and adaptive beamforming [1]. In fixed beamforming, the beamformer weights are designed to focus into a main source direction while suppressing signals from other undesired directions [7]. This problem can be viewed as a multidimensional filter design problem. The beamformer weights are calculated based on information about the array geometry and the source localization with no statistical information about the environment or desired signals required.

Multichannel adaptive optimum filtering, on the other hand, requires statistical knowledge of the noise, the source and the environment. The beamformer coefficients are optimized such that a focussed beam is steered to a desired source direction, whilst suppressing the contributions coming from other noise directions [8]. The design requires *a priori* information about the location of the desired signal, the array geometry and the noise statistics. Based on this information, a spatial, spectral, and temporal filter can be formed to capture the desired signal. Two popular optimum multichannel techniques are multichannel Wiener filtering and linearly constrained minimum variance (LCMV) [4]. The multichannel Wiener filter provides a minimum mean square error (mmse) estimate of either the clean desired signal in one of the microphones or a given reference signal. The LCMV beamformer, on the other hand, is designed to minimize the average output power subject to some linear constraints, which preserve the desired source region.

Adaptive beamforming techniques are developed to track time-varying signal environments [9]. It maintains a beam towards the desired source while the beamformer's coefficients are updated using the information of the observed signals. In general, adaptive beamforming techniques are reported to have good interference suppression capability. Some popular adaptive beamforming techniques are the generalized sidelobe canceler (GSC) and multichannel adaptive Wiener filtering. The GSC might succumb to target signal cancellation in the presence of steering vector errors due to microphone positions and reverberation. The multichannel adaptive Wiener filtering techniques, on the other hand, exploit both spectral and spatial differences between the desired source and the interference to

H. H. Dam is with the Department of Mathematics and Statistics, Curtin University of Technology, Perth 6845, Australia (e-mail: dam@watri.org.au).

H. Q. Dam is with the University of Information Technology, Vietnam National University, Ho Chi Minh City KP6, Vietnam (e-mail:damhai@watri.org.au).

S. Nordholm is with the Western Australian Telecommunications Research Institute, Curtin University of Technology, Perth 6845, Australia (e-mail: sven@watri.org.au).

provide a clean estimate of the clean speech source signal. The speech spectral contents, however, may vary over the time and thus cannot be pre-estimated.

This paper investigates the problem with more than one speech sources corrupted by a long-term stationary background noise. This is of interest for, e.g., voice input for voice communications applications. This situation arises in many practical cases, e.g., an echo cancellation problem in a noisy car environment, where the echo is viewed as an interference corrupting a desired source signal. In general, the desired source can represent the speaker sitting in front of a microphone array and the echo signal comes from the far-end speaker through a loudspeaker. Thus, the observed signal contains the speech signals from the desired and undesired sources (e.g., echo) as well as the background noise. Given a noisy observed signal, the objective is to enhance a desired speech source while suppressing the undesired sources and the background noise.

Here, we propose a new beamformer structure developed based on the multichannel optimum Wiener filtering solution which takes into account power spectral density (PSD) estimation of the speech sources. Due to the short-term stationary property of the speech sources, these information change frequently and thus are required to be estimated at each time instance. To track the noise statistics, an inactive-source detector is developed to detect the presence of the noise and to allow the noise correlation matrix to be updated accordingly. Consequently, the beamformer weights of the proposed structure are updated based on the PSD estimations in conjunction with the noise statistics update. Performance of the beamformer is tested and compared in various cases in a hands-free communication environment. As a baseline for the comparison, a minimum variance distortionless response (MVDR) beamformer with or without a postfilter has been employed.

The organization of the paper is given as follows. Signal model is presented in Section II. The spatial correlation matrices of the speech sources are introduced in Section III. An inactive-source detector and noise statistics update is developed in Section IV. The PSD estimation of the speech sources is proposed in Section V. An adaptive beamformer structure with noise statistics and PSD updates is proposed in Section VI. Efficient implementation for coefficient update of the proposed beamformer is given in Section VII while the MVDR beamformer is given in Section VIII. Finally, simulation results in a real car hands-free environment are presented in Section IX, and conclusions are given in Section X.

II. SIGNAL MODEL

Consider a typical microphone array with L elements. Assume that the observed signal at the time index n includes the contributions from I speech sources $\mathbf{s}_i(n)$, $1 \leq i \leq I$, and a background noise $\mathbf{v}(n)$. Thus, the $L \times 1$ vector observed signal $\mathbf{x}(n)$ can be expressed as

$$\mathbf{x}(n) = \sum_{i=1}^{I} \mathbf{s}_i(n) + \mathbf{v}(n). \tag{1}$$

Here, the I sources and the noise are also assumed to be uncorrelated.
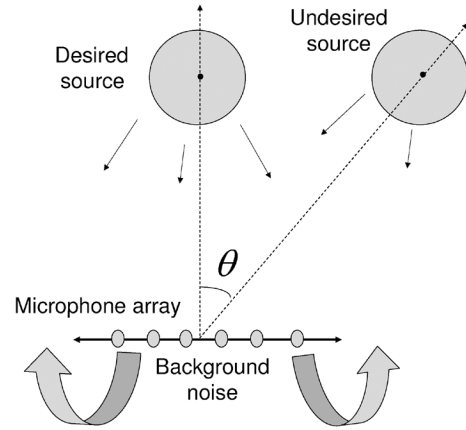

Fig. 1. Configuration example of an echo-cancellation problem with a linear microphone array, a desired source and an interference in a noisy environment.

The processing is performed in the frequency domain by using a uniform or nonuniform filter bank [10]. Here, we employ a uniformly modulated filter bank. The received signal is decomposed into M subbands by using an analysis filter bank [11]. The benefit of using an uniform filter bank is that the analysis and synthesis prototype filters can be easily designed to minimize the inband aliasing, the aliased energy between different subbands, the overall distortions as well as having low delay [11]. The filtering and processing operations are performed for each frequency bin. For a frequency ω, the observed signal $\mathbf{x}(\omega, k)$ at the time index k can be expressed as

$$\mathbf{x}(\omega, k) = \sum_{i=1}^{I} \mathbf{s}_i(\omega, k) + \mathbf{v}(\omega, k) \tag{2}$$

where $\mathbf{s}_i(\omega, k)$, $1 \leq i \leq I$, and $\mathbf{v}(\omega, k)$ are the frequency-domain contributions from the I sources and the noise. As an example for an echo cancellation problem, the number of sources is two, i.e., $I = 2$. Hence, the first source is viewed as the desired source while the second is the undesired source, see Fig. 1.

In general, the objective is to suppress $I-1$ undesired sources and the noise while maintaining the quality and timbre of the desired source. By assuming that the sources and the noise are uncorrelated, the correlation matrix $\mathbf{R}_x(\omega, k)$ of the observed signal in (2) can be expressed as

$$\mathbf{R}_x(\omega, k) = \sum_{i=1}^{I} \mathbf{R}_i(\omega, k) + \mathbf{R}_v(\omega, k) \tag{3}$$

where $\mathbf{R}_i(\omega, k)$, $1 \leq i \leq I$, and $\mathbf{R}_v(\omega, k)$ are the correlation matrices for the I sources and the noise component, respectively. In the following, the spatial correlation matrices of speech sources will be outlined.

III. SPATIAL CORRELATION MATRICES OF SPEECH SOURCES

In [12] and [13], a calibration method is employed with training samples of the sources recorded prior the beamforming process. This method is used to estimate the statistical information of the sources which includes unknown signal path

information. By doing so, all the information on the array geometry and source localization will be reflected in the solution [14].

During the calibration period, each speaker is active for a short period of time while other speakers are silent. Denote by $[K_{1,i}, K_{2,i}]$ the active time of the ith source, $1 \le i \le I$, and $\hat{\mathbf{R}}_{i,cal}(\omega)$ the correlation matrix for ith source estimated during the calibration period. This correlation matrix can be obtained as

$$\hat{\mathbf{R}}_{i,cal}(\omega) = \frac{1}{K_{2,i} - K_{1,i} + 1} \sum_{k=K_{1,i}}^{K_{2,i}} \mathbf{x}(\omega,k)\mathbf{x}^H(\omega,k). \quad (4)$$

Moreover, denote by $\hat{\mathbf{d}}_{i,cal}(\omega)$ the spatial cross-correlation vector with respect to an ℓth prechosen reference microphone, $1 \le \ell \le L$. The vector $\hat{\mathbf{d}}_{i,cal}(\omega)$ is estimated as

$$\hat{\mathbf{d}}_{i,cal}(\omega) = \frac{1}{K_{2,i} - K_{1,i} + 1} \sum_{k=K_{1,i}}^{K_{2,i}} \mathbf{x}(\omega,k)x^*(\omega,k,\ell) \quad (5)$$

where $x(\omega,k,\ell)$ is the received signal at the ℓth microphone. The spatial correlation matrix $\bar{\mathbf{R}}_i(\omega)$ and the spatial cross correlation vector $\bar{\mathbf{d}}_i(\omega)$ for the ith source can be estimated as

$$\bar{\mathbf{R}}_i(\omega) = \frac{\hat{\mathbf{R}}_{i,cal}(\omega)}{\hat{R}_{i,cal}(\omega,\ell,\ell)} \quad (6)$$

and

$$\bar{\mathbf{d}}_i(\omega) = \frac{\hat{\mathbf{d}}_{i,cal}(\omega)}{\hat{d}_{i,cal}(\omega,\ell)} \quad (7)$$

where $\hat{R}_{i,cal}(\omega,\ell,\ell)$ is the (ℓ,ℓ) element of the matrix $\hat{\mathbf{R}}_{i,cal}(\omega)$, and $\hat{d}_{i,cal}(\omega,\ell)$ is the ℓth element of the vector $\hat{\mathbf{d}}_{i,cal}(\omega)$.

Alternatively, the spatial correlation matrices $\bar{\mathbf{R}}_i(\omega)$ and the spatial correlation vectors $\bar{\mathbf{d}}_i(\omega)$, $1 \le i \le I$, can be estimated by using a soft-constrained model [15]. More specifically, each source is modeled as a distributed source within a constrained area [16]. Consequently, the spatial correlation matrices and the spatial correlation vectors are estimated based on the assumed constrained areas [17].

Denote by $\hat{\mathbf{R}}_x(\omega,k)$ and $\hat{\mathbf{R}}_v(\omega,k)$, respectively, the instantaneous correlation matrix and instantaneous noise correlation matrix and at time instant k and frequency bin ω. Similarly, denote by $\hat{p}_i(\omega,k)$ the instantaneous PSD of the ith source. Following from (3), we have

$$\hat{\mathbf{R}}_x(\omega,k) = \sum_{i=1}^I \hat{p}_i(\omega,k)\bar{\mathbf{R}}_i(\omega) + \hat{\mathbf{R}}_v(\omega,k). \quad (8)$$

The instantaneous correlation matrix $\hat{\mathbf{R}}_x(\omega,k)$ can be calculated based on the observed signal $\mathbf{x}(\omega,k)$ by using K samples as

$$\hat{\mathbf{R}}_x(\omega,k) = \frac{1}{K} \sum_{n=k-K+1}^k \mathbf{x}(\omega,n)\mathbf{x}^H(\omega,n) \quad (9)$$

where K is a prechosen number of samples.

Since $\bar{\mathbf{R}}_i(\omega), 1 \le i \le I$, is a spatial correlation matrix, it can be decomposed as

$$\bar{\mathbf{R}}_i(\omega) = \mathbf{V}_i(\omega)\mathbf{\Lambda}_i(\omega)\mathbf{V}_i^H(\omega) \quad (10)$$

where $\mathbf{V}_i(\omega)$ is an orthonormal matrix containing the eigenvectors, and $\mathbf{\Lambda}_i(\omega)$ is a diagonal matrix consisting of nonnegative eigenvalues. Denote by $\mathbf{v}_i(\omega)$ the eigenvector corresponding to the largest eigenvalue $\lambda_i(\omega)$ of the matrix $\bar{\mathbf{R}}_i(\omega)$. We can obtain a value $z(\omega,k)$ based on $\hat{\mathbf{R}}_x(\omega,k)$ and the eigenvectors $\mathbf{v}_i(\omega)$, $1 \le i \le I$, as

$$z(\omega,k) = \sum_{i=1}^I \mathbf{v}_i^H(\omega)\hat{\mathbf{R}}_x(\omega,k)\mathbf{v}_i(\omega). \quad (11)$$

We have the following property.

Property III.1: The lower bound on $z(\omega,k)$ depends on the instantaneous noise correlation matrix $\hat{\mathbf{R}}_v(\omega,k)$

$$z(\omega,k) \ge \sum_{i=1}^I \mathbf{v}_i^H(\omega)\hat{\mathbf{R}}_v(\omega,k)\mathbf{v}_i(\omega). \quad (12)$$

Equality in (12) occurs if and only if there is no-active speech signal or $\hat{p}_i(\omega,k) = 0$ for all i.

Proof: See Appendix A. □

Next, an inactive-source detector will be proposed and a noise statistics update for the noise correlation matrix based on the detector will be presented.

IV. INACTIVE-SOURCE DETECTOR AND NOISE STATISTICS UPDATE

In the following, it is assumed that the noise is long-term stationary while the speech sources are short-term stationary and have inactive periods of time.

A. Inactive-Source Detector

Consider a two-hypothesis test for the inactive-source detector. The null hypothesis H_0 represents the case where the speech signals are inactive and the alternative hypothesis H_1 is for the case where there is at least one active speech source. More specifically, we have

$$\begin{cases} H_0, & \hat{p}_i(\omega,k) = 0 \, \forall \, 1 \le i \le I \\ H_1, & \text{otherwise.} \end{cases} \quad (13)$$

The test for the two-hypothesis in (13) is developed by utilizing the idea of minimum statistics. Simply put, minimum statistics is based on the observation that a short-time subband power of a noisy speech signal exhibits peaks and valleys [18]. The peaks refer to speech activity periods while the valleys are representative values of the noise power level. By tracking the minimum power within a finite window large enough to bridge high power speech periods, the minimum statistics can be obtained.

Denote by N a number that the noise statistics of N consecutive samples can assume to be the same. To determine (13) at the time index k, we obtain the minimum value of (11) for N consecutive samples with index $[k-N+1,k]$

$$\xi(\omega,k) = \min_{k_1 \in [k-N+1,k]} z(\omega,k_1). \quad (14)$$

Following from Property 1, the decision for the hypothesis testing in (13) can be obtained by comparing the summation

$$A_1(k) = \sum_{\omega \in \Omega} z(\omega, k)$$

with a summation threshold

$$A_2(k) = \sum_{\omega \in \Omega} \xi(\omega, k)(1 + \varepsilon)$$

where Ω is the frequency range in which speech signal has the most energy, and ε is a small positive number. Here, Ω is chosen as [500, 3500] Hz and ε is chosen as 0.1. Note that the value ε can also be chosen related to the probability of a false-alarm [19]. Consequently, the hypothesis H_0 is accepted if $A_1(k) < A_2(k)$ and rejected otherwise.

B. Noise Statistics Update

The noise correlation matrix is updated during the inactive-source periods to track the noise statistics. During these periods, the instantaneous noise correlation matrix $\hat{\mathbf{R}}_v(\omega, k)$ is updated from the correlation matrix $\hat{\mathbf{R}}_x(\omega, k)$ through a first-order smoothing function. More specifically, if H_0 is accepted then $\hat{\mathbf{R}}_v(\omega, k)$ is obtained from $\hat{\mathbf{R}}_v(\omega, k-1)$ and $\hat{\mathbf{R}}_x(\omega, k)$ as

$$\hat{\mathbf{R}}_v(\omega, k) = (1 - \alpha)\hat{\mathbf{R}}_v(\omega, k-1) + \alpha\hat{\mathbf{R}}_x(\omega, k), \ \forall \, \omega \quad (15)$$

where α is a constant smoothing factor. During periods where at least one speech source is detected, i.e., H_1 is accepted, the matrix $\hat{\mathbf{R}}_v(\omega, k)$ remains the same as $\hat{\mathbf{R}}_v(\omega, k-1)$

$$\hat{\mathbf{R}}_v(\omega, k) = \hat{\mathbf{R}}_v(\omega, k-1), \ \forall \, \omega. \quad (16)$$

To reduce the computational complexity, the inactive-source detector in Section IV-A can be tested for every N_1 samples where $N_1 < N$. Consequently, the matrix $\hat{\mathbf{R}}_v(\omega, k)$ is updated once for every N_1 samples according to (15) and (16). In the sequel, instantaneous PSD estimation of speech sources will be obtained based on the estimated instantaneous noise correlation matrix.

V. INSTANTANEOUS PSD ESTIMATION OF SPEECH SOURCES

Due to the short-term stationary property of the speech sources, the PSD information for the speech sources is required to be estimated at each time instance. Thus, the objective is to estimate the PSD of the speech sources, $\hat{p}_i(\omega, k), 1 \le i \le I$, from the observed signal by exploiting the spatial difference between the speech sources and the estimated instantaneous noise correlation matrix. More specifically, the PSDs are estimated based on $\hat{\mathbf{R}}_x(\omega, k)$ in (9), the spatial correlation matrices $\bar{\mathbf{R}}_i(\omega), 1 \le i \le I$, and $\hat{\mathbf{R}}_v(\omega, k)$ obtained from (15) and (16).

Here, the instantaneous PSDs are estimated by minimizing the Frobenious norm of the correlation model mismatch error

matrix. In other words, these PSDs are obtained by solving the following constrained optimization problem:

$$\begin{cases} \min_{\hat{p}_i(\omega, k)} \left\| \hat{\mathbf{R}}_x(\omega, k) - \sum_{i=1}^I \hat{p}_i(\omega, k)\bar{\mathbf{R}}_i(\omega) - \hat{\mathbf{R}}_v(\omega, k) \right\|_F^2 \\ \hat{p}_i(\omega, k) \ge 0, \ \forall 1 \le i \le I. \end{cases} \quad (17)$$

Denote by $\hat{\mathbf{p}}(\omega, k)$ the vector containing the PSD at frequency ω and time instant k

$$\hat{\mathbf{p}}(\omega, k) = [\hat{p}_1(\omega, k), \cdots, \hat{p}_I(\omega, k)]^T.$$

Moreover, denote by $\bar{\mathbf{r}}(\omega, l, j), 1 \le l, j \le I$, an $I \times 1$ vector containing the (l, j)th element of the I spatial correlation matrices

$$\bar{\mathbf{r}}(\omega, l, j) = \left[\bar{\mathbf{R}}_1(\omega, l, j), \cdots, \bar{\mathbf{R}}_I(\omega, l, j) \right].$$

In addition, let

$$\hat{\mathbf{R}}(\omega, k) = \hat{\mathbf{R}}_x(\omega, k) - \hat{\mathbf{R}}_v(\omega, k).$$

The optimization problem (17) can be given in the following form:

$$\begin{cases} \min_{\hat{\mathbf{p}}(\omega, k)} \sum_{l,j=1}^I \left\| \hat{r}(\omega, k, l, j) - \hat{\mathbf{p}}^T(\omega, k)\bar{\mathbf{r}}(\omega, l, j) \right\|^2 \\ \hat{\mathbf{p}}(\omega, k) \ge \mathbf{0} \end{cases} \quad (18)$$

where $\hat{r}(\omega, k, l, j)$ is the (l, j)th element of the matrix $\hat{\mathbf{R}}(\omega, k)$ and $\mathbf{0}$ is an $I \times 1$ zero vector. By denoting

$$\mathbf{A}(\omega) = \sum_{l,j=1}^I \mathcal{R}\left\{ \bar{\mathbf{r}}(\omega, l, j)\bar{\mathbf{r}}^H(\omega, l, j) \right\} \quad (19)$$

and

$$\mathbf{b}(\omega, k) = -\sum_{l,j=1}^I \mathcal{R}\left\{ \hat{r}(\omega, k, l, j)\bar{\mathbf{r}}^H(\omega, l, j) \right\}$$

with $\mathcal{R}\{\cdot\}$ denotes the real part of $\{\cdot\}$. The optimization problem (18) can be reduced to a real quadratic optimization problem

$$\begin{cases} \min_{\hat{\mathbf{p}}(\omega, k)} \frac{1}{2}\hat{\mathbf{p}}^T(\omega, k)\mathbf{A}(\omega)\hat{\mathbf{p}}(\omega, k) + \mathbf{b}^T(\omega, k)\hat{\mathbf{p}}(\omega, k) \\ \hat{\mathbf{p}}(\omega, k) \ge \mathbf{0}. \end{cases} \quad (20)$$

The matrix $\mathbf{A}(\omega)$ in (19) depends only on the spatial correlation matrices and can be calculated only once. Also from (19), $\mathbf{A}(\omega)$ is positive definite. Thus, the optimization problem (20) is convex and hence the solution to (20) is unique. Consequently, (20) can be solved by using any fixed or adaptive quadratic optimization technique using, e.g., the active set method. The steps for solving (20) for a special case with $I = 2$ is given in Appendix B. This situation occurs in various practical cases including an echo cancellation problem.

In the following, the NSUAB structure employing the PSD estimation in conjunction with the noise statistics update will be developed.

VI. NOISE STATISTICS UPDATE ADAPTIVE BEAMFORMER (NSUAB) WITH PSD ESTIMATION

A NSUAB based on the optimum Wiener solution is proposed for the speech extraction problem in noisy environment. The inactive-source detector and noise correlation matrix update in Section IV in conjunction with the instantaneous PSD estimation of speech sources in Section V will be incorporated in the estimation of adaptive beamformer coefficients.

For the signal model in (3), the optimum Wiener solution $\mathbf{w}_w(\omega, k)$ for each frequency ω and time index k is expressed as

$$\mathbf{w}_w(\omega, k) = \left[\sum_{i=1}^{I} \mathbf{R}_i(\omega, k) + \mathbf{R}_v(\omega, k)\right]^{-1} \mathbf{d}_s(\omega, k) \quad (21)$$

where $\mathbf{d}_s(\omega, k)$ is the cross-correlation vector of the sth desired source with reference to the ℓth microphone.

For an adaptive system, the beamformer coefficients $\mathbf{w}_a(\omega, k)$ of the NSUAB with PSD and noise statistics updates are obtained from the beamformer coefficients $\mathbf{w}_a(\omega, k-1)$ based on (21) as

$$\mathbf{w}_a(\omega, k) = \eta \mathbf{w}_a(\omega, k-1) + (1 - \eta)$$
$$\times \left[\sum_{i=1}^{I} \hat{p}_i(\omega, k)\bar{\mathbf{R}}_i(\omega) + \hat{\mathbf{R}}_v(\omega, k)\right]^{-1} \hat{p}_s(\omega, k)\bar{\mathbf{d}}_s(\omega) \quad (22)$$

where η, $0 < \eta < 1$ is a weight smoothing factor employed to reduce the weight fluctuations from one time instance to another, and $\hat{p}_s(\omega, k)$ is the instantaneous PSD estimation for the desired source.

Finally, the output of the NSUAB with PSD updates for the frequency ω is given by

$$y(\omega, k) = \mathbf{w}_a^H(\omega, k)\mathbf{x}(\omega, k). \quad (23)$$

This output is passed through a synthesis filter bank to obtain the received signal in the time domain.

VII. EFFICIENT IMPLEMENTATION FOR COEFFICIENT UPDATE OF THE NSUAB

Depending on how wide the source area spans, the spatial correlation matrices $\bar{\mathbf{R}}_i(\omega)$ of the I sources can have one or a few dominating eigenvalues. As such, a fast updating scheme is developed for calculating the adaptive beamformer weight $\mathbf{w}_a(\omega, k)$ in (22) based on one or a few eigenvectors corresponding to one or a few dominant eigenvalues of the spatial correlation matrices. The *Matrix Inversion Lemma* [9] is employed to efficiently obtain the matrix inverse in (22) based on one or a few dominant eigenvectors of the spatial correlation matrices to reduce the computational burden. In addition, as the matrix $\hat{\mathbf{R}}_v^{-1}(\omega, k)$ is updated only once for every N_1 samples, the inverse matrix $\hat{\mathbf{R}}_v^{-1}(\omega, k)$ is required to be updated only once for every N_1 samples.

VIII. MVDR BEAMFORMER

For comparison, performance of the proposed beamformer is compared with the MVDR beamformer with or without a postfilter. The MVDR beamformer is a special case of the linearly constrained minimum variance (LCMV) beamformer [20] where a constraint is imposed to maintain a distortionless response from the desired source direction. It is noted that other types of beamformers can also be used [21]–[23]. However, there will be a tradeoff in performance between the distortion and the suppression levels.

For the model in (3), the MVDR beamformer weight $\mathbf{w}_m(\omega, k)$ is obtained as the solution to the following optimization problem

$$\begin{cases} \min_{\mathbf{w}_m(\omega, k)} & \mathbf{w}_m^H(\omega, k)\mathbf{R}_x(\omega, k)\mathbf{w}_m(\omega, k) \\ \text{subject to} & \mathbf{w}_m^H(\omega, k)\bar{\mathbf{d}}_s(\omega) = 1. \end{cases} \quad (24)$$

where $\bar{\mathbf{d}}_s(\omega)$ is the spatial cross correlation vector for the sth desired source, given as in Section III. The exact solution to this optimization problem can be expressed as

$$\mathbf{w}_m(\omega, k) = \frac{\mathbf{R}_x^{-1}(\omega, k)\bar{\mathbf{d}}_s(\omega)}{\bar{\mathbf{d}}_s^H(\omega)\mathbf{R}_x^{-1}(\omega, k)\bar{\mathbf{d}}_s(\omega)}. \quad (25)$$

For an adaptive system, the beamformer coefficients $\mathbf{w}_{m,a}(\omega, k)$ of the MVDR structure are obtained from the coefficients $\mathbf{w}_{m,a}(\omega, k-1)$ as

$$\mathbf{w}_{m,a}(\omega, k) = \eta \mathbf{w}_{m,a}(\omega, k-1) + (1 - \eta)\mathbf{w}_{m,est}(\omega, k) \quad (26)$$

where η, $0 < \eta < 1$ is a weight smoothing factor and $\mathbf{w}_{m,est}(\omega, k)$ is obtained from $\hat{\mathbf{R}}_x^{-1}(\omega, k)$ as

$$\mathbf{w}_{m,est}(\omega, k) = \frac{\hat{\mathbf{R}}_x^{-1}(\omega, k)\bar{\mathbf{d}}_s(\omega)}{\bar{\mathbf{d}}_s^H(\omega)\hat{\mathbf{R}}_x^{-1}(\omega, k)\bar{\mathbf{d}}_s(\omega)}. \quad (27)$$

The noise component in the output of the MVDR beamformer is reduced further by passing the output through a minimum mean-square error log-spectral amplitude estimator [24]. The short-time spectral amplitude (STSA) estimator minimizes the mean-square error of the log-spectra between the original STSA and its estimator. For this estimator, a "decision-directed" method is employed for the estimation of *a priori* SNR for the spectral component [25]. Minimum statistics [18] is incorporated the method to detect inactive source periods and hence the estimation of the noise power.

IX. SIMULATION RESULTS

Evaluations were performed in a real car hands-free situation for a Toyota Landcruiser 4WD driven on sealed roads, driving at a speed of 60 km/h. An six-element linear array with 40-mm spacing was mounted on the dashboard in front of the passenger seat. Data were gathered on a multichannel DAT-recorder with a sampling rate of 16 kHz and bandlimited to 200–3400 Hz.

Here, we concentrate on the double-talk case. The non double-talk case can be dealt with similarly. The two speech sources are one male and one female speech sources having the same power and operating at the same time. The environment

TABLE I
CASE 1: ONE DOMINANT EIGENVECTOR CORRESPONDING TO THE LARGEST EIGENVALUE IS USED FOR CALCULATING $z(\omega, k)$ IN (11). NOISE SUPPRESSION (NS), INTERFERENCE SUPPRESSION (IS), AND THE DESIRED SOURCE DISTORTION (SD) LEVELS FOR THE MVDR BEAMFORMER AND THE NSUAB FOR DIFFERENT CASES. THE RESULT IS THE AVERAGE OF TEN SEQUENCES WITH EACH OF LENGTH 8 s IN A REAL DOUBLE-TALK HANDS-FREE ENVIRONMENT

Beamformers			NS (dB)	IS (dB)	SD (dB)
NSUAB	Without a post-filter	$N_1 = 2$	15.09	14.25	-37.44
		$N_1 = 4$	14.98	14.00	-37.59
		$N_1 = 8$	14.87	13.69	-37.75
		$N_1 = 16$	14.79	13.41	-37.92
		$N_1 = 32$	14.79	12.90	-38.16
	With a post-filter	$N_1 = 2$	15.92	14.51	-37.30
		$N_1 = 4$	15.80	14.26	-37.45
		$N_1 = 8$	15.73	13.97	-37.60
		$N_1 = 16$	15.73	13.74	-37.72
		$N_1 = 32$	15.78	13.30	-37.90
MVDR beamformer			8.52	11.15	-37.92
MDVR followed by a post-filter			12.78	12.06	-37.81

TABLE II
CASE 2: TWO DOMINANT EIGENVECTORS CORRESPONDING TO THE TWO LARGEST EIGENVALUES ARE USED FOR CALCULATING $z(\omega, k)$ IN (11). NOISE SUPPRESSION (NS), INTERFERENCE SUPPRESSION (IS), AND THE DESIRED SOURCE DISTORTION (SD) LEVELS FOR THE NSUAB FOR DIFFERENT CASES. THE RESULT IS THE AVERAGE OF TEN SEQUENCES WITH EACH OF LENGTH 8 s IN A REAL DOUBLE-TALK HANDS-FREE ENVIRONMENT

Beamformers			NS (dB)	IS (dB)	SD (dB)
NSUAB	Without a post-filter	$N_1 = 2$	15.11	14.27	-37.43
		$N_1 = 4$	15.00	14.01	-37.59
		$N_1 = 8$	14.89	13.68	-37.75
		$N_1 = 16$	14.79	13.40	-37.92
		$N_1 = 32$	14.78	12.92	-38.16
	With a post-filter	$N_1 = 2$	15.94	14.53	-37.29
		$N_1 = 4$	15.82	14.28	-37.44
		$N_1 = 8$	15.74	13.96	-37.60
		$N_1 = 16$	15.73	13.73	-37.71
		$N_1 = 32$	15.77	13.33	-37.90

where C_d is a constant employed to normalize the performance measure

$$C_d = \frac{\int_{-\pi}^{\pi} \hat{P}_{\mathrm{in},s}(\omega) d\omega}{\int_{-\pi}^{\pi} \hat{P}_{\mathrm{out},s}(\omega) d\omega}. \tag{29}$$

Performance is also given in terms of a source distortion (SD) measure, defined as

$$\mathrm{SD} = 10 \log_{10} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| (1/C_d) \hat{P}_{\mathrm{in},s}(\omega) - \hat{P}_{\mathrm{out},s}(\omega) \right| d\omega \right) \tag{30}$$

and an interference suppression (IS) measure, defined as

$$\mathrm{IS} = 10 \log_{10} \left(\frac{\int_{-\pi}^{\pi} \hat{P}_{\mathrm{in},i}(\omega) d\omega}{\int_{-\pi}^{\pi} \hat{P}_{\mathrm{out},i}(\omega) d\omega} \right) - 10 \log_{10}(C_d). \tag{31}$$

Case 1: Consider the case using only one dominant eigenvector corresponding to the largest eigenvalue of the spatial correlation matrix for calculating $z(\omega, k)$ in (11). Table I shows the NS, IS, and SD measures for the NSUAB structure and the MVDR beamformers with or without the mmse log-spectral amplitude estimator for different cases.

The table shows the performance of the NSUAB for different values of N_1, $N_1 < N$ where N_1 is the number of samples in which the inactive-source detector and the noise statistics are updated only once. It can be seen that the NSUAB performance is slightly reduced when N_1 increases from 2 to 32. Obviously, the computational complexity is reduced for higher value of N_1 as the inactive-source detector is operated only once for every N_1 samples and the noise correlation matrix is also updated only once for every N_1 samples.

has a signal-to-noise ratio (SNR) of -6 dB and a source to interference ratio (SIR) of 0 dB. The results are obtained based on the average of ten recorded sequences with each having 8 s. Calibration data is recorded before running the algorithms for a period of 4 s.

The observed signals are transformed into the frequency domain using a $M = 64$ point uniformly over-sampled DFT filter bank with a decimation factor $M/2$ to reduce the aliasing effects between the adjacent subbands. The analysis and synthesis filter bank can be designed to minimize the transformation and reconstruction aliasing effects [11]. The value N is chosen as 44 [corresponding to a period of 1/12(s)]. The reference microphone is chosen as the middle microphone in the array, e.g., $\ell = 4$.

Performance measures of the proposed beamformer is obtained by assuming the access of the desired source, interference, and the noise separately. Denote by $\hat{P}_{\mathrm{in},s}(\omega)$ and $\hat{P}_{\mathrm{out},s}(\omega)$ the PSD of desired source input and output, respectively, with reference to a microphone. Similarly, denote by $\hat{P}_{\mathrm{in},n}(\omega)$, $\hat{P}_{\mathrm{out},n}(\omega)$, $\hat{P}_{\mathrm{in},i}(\omega)$, and $\hat{P}_{\mathrm{out},i}(\omega)$ the PSD of the input and output for the noise and interference, respectively. Performance of the beamformer is quantified in terms of a noise suppression (NS) measure, defined as

$$\mathrm{NS} = 10 \log_{10} \left(\frac{\int_{-\pi}^{\pi} \hat{P}_{\mathrm{in},n}(\omega) d\omega}{\int_{-\pi}^{\pi} \hat{P}_{\mathrm{out},n}(\omega) d\omega} \right) - 10 \log_{10}(C_d) \tag{28}$$
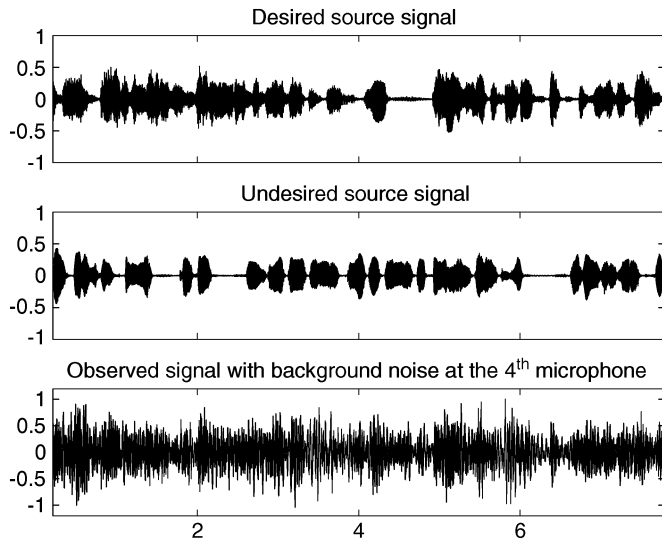
Fig. 2.   Double-talk situation. The desired, undesired sources, and the noisy observed signal at the fourth microphone.

Fig. 4.   Double-talk situation. Plots of the output for (a) the MVDR beamformer, (b) the MVDR beamformer and postfilter, and (c) the NSUAB.
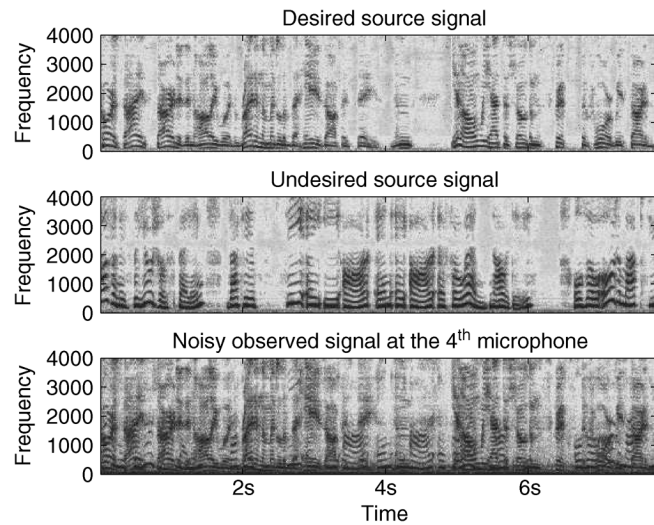

Fig. 3.   Double-talk situation. Spectrograms of the desired, undesired sources, and the noisy observed signal at the fourth microphone.

Fig. 5.   Double-talk situation. Spectrograms of the output for (a) the MVDR beamformer, (b) the MVDR beamformer and postfilter, and (c) the NSUAB.

For the MVDR beamformer, the postfilter using the mmse log-spectral amplitude estimator significantly improves the noise and interference suppressions. More specifically, the NS is increases by 4.25 dB while the IS is increased by 0.9 dB.

The NSUAB structure improves further both the NS and IS over the MVDR beamformer with mmse log-spectral amplitude estimator. More specifically, the NSUAB output when passing through the postfilter using the mmse log-spectral amplitude estimator improves up to 3.1-dB NS and 2.5-dB IS over the MVDR beamformer with the same postfilter while having approximately the same SD. For the NSUAB structure, the use of a postfilter only slightly increases the filter performance as the spectral information has already been included from the PSD estimations and the noise statistics are updated in the scheme.

Case 2: We now consider the case using two dominant eigenvectors corresponding to the two largest eigenvalues of the spatial correlation matrices for calculating $z(\omega, k)$ in (11). Table II shows the NS, IS, and SD measures for the NSUAB structure
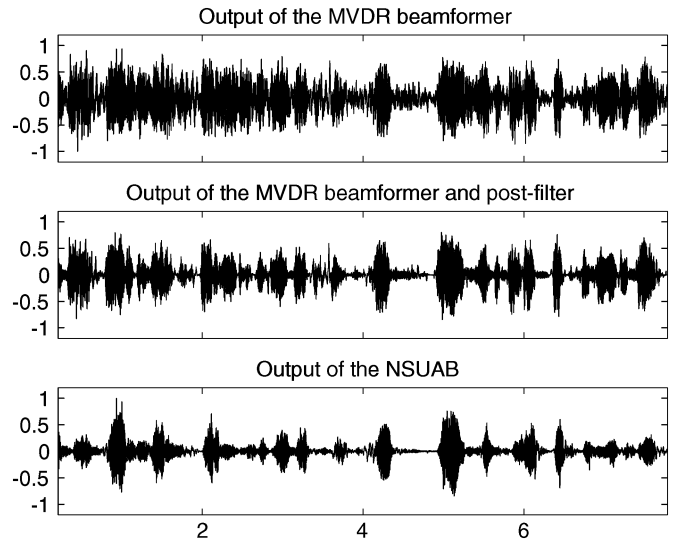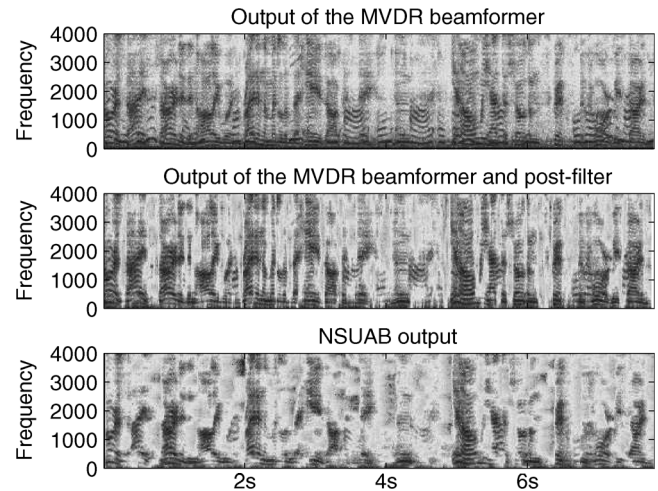
for different cases. It can be seen from Tables I and II that the use of two dominant eigenvectors gives only a slight increase in the performance of the beamformers.

Figs. 2 and 3 show, respectively, an example of the time-domain and the spectrogram plots for the desired, undesired sources and the observed signal at the fourth microphone for one of the sequences in a double-talk situation. The observed signal is noisy with both speakers speak at the same time.

Figs. 4 and 5 show, respectively, the time-domain and the spectrogram plots for the output of 1) MVDR beamformer, 2) MVDR beamformer with postfilter, and 3) the NSUAB for one of the sequences. Informal listening shows good timbre of the output signal with the undesired signal and the background noise are significantly suppressed.

X. CONCLUSION

In this paper, a new NSUAB structure with source PSD estimation and noise statistics updates has been proposed for the speech enhancement problem with more than one speech

sources corrupted by long-term stationary background noise. The NSUAB has good interference and noise suppression levels when compared with the MVDR beamformer with or without a postfilter. In addition, the beamformer has the ability to suppress undesired speech sources and a long-term stationary background noise with low desired source distortion.

APPENDIX A
PROOF FOR PROPERTY III.1.

It follows from (8) that (11) can be rewritten as

$$z(\omega,k) = \sum_{i=1}^{I} \hat{p}_i(\omega,k)\lambda_i(\omega)$$
$$+ \sum_{i=1}^{I}\sum_{l=1,l\neq i}^{I} \hat{p}_l(\omega,k)\mathbf{v}_i^H(\omega)\bar{\mathbf{R}}_l(\omega)\mathbf{v}_i(\omega)$$
$$+ \sum_{i=1}^{I} \mathbf{v}_i^H(\omega)\hat{\mathbf{R}}_v(\omega,k)\mathbf{v}_i(\omega). \tag{32}$$

Since the instantaneous power spectral density $\hat{p}_i(\omega,k)$ for the ith source at the time instance k and frequency ω is positive, i.e., $\hat{p}_i(\omega,k) \geq 0$ and the matrices $\bar{\mathbf{R}}_l(\omega)$, $1 \leq l \leq I$, are Hermitian and positive semi-definite, we have

$$\sum_{i=1}^{I} \hat{p}_i(\omega,k)\lambda_i(\omega) \geq 0 \tag{33}$$

and

$$\sum_{i=1}^{I}\sum_{l=1,l\neq i}^{I} \hat{p}_l(\omega,k)\mathbf{v}_i^H(\omega)\bar{\mathbf{R}}_l(\omega)\mathbf{v}_i(\omega) \geq 0. \tag{34}$$

Thus, following from (32)–(34), we have

$$z(\omega,k) \geq \sum_{i=1}^{I} \mathbf{v}_i^H(\omega)\hat{\mathbf{R}}_v(\omega,k)\mathbf{v}_i(\omega). \tag{35}$$

Clearly, the equality occurs if and only if there is no-active speech signal or $\hat{p}_i(\omega,k) = 0$ for all i.

APPENDIX B
PSD ESTIMATION FOR A SPECIAL CASE WITH $I = 2$

For $I = 2$, the solution for the optimization problem (20) can be obtained by using the following steps [26].

- Step 1: Calculate the unconstrained solution of (20) as

$$\tilde{\mathbf{p}}(\omega) = \mathbf{A}^{-1}(\omega)\mathbf{b}(\omega,k). \tag{36}$$

Note that the matrix $\mathbf{A}(\omega)$ does not depend on the time index k and thus $\mathbf{A}^{-1}(\omega)$ can be calculated only once.
- Step 2: If the elements $\tilde{p}_1(\omega)$ and $\tilde{p}_2(\omega)$ of $\tilde{\mathbf{p}}(\omega)$ are non-negative then $\tilde{\mathbf{p}}(\omega)$ is the solution of (20) with $\hat{p}_1(\omega,k) = \tilde{\mathbf{p}}_1(\omega)$ and $\hat{p}_2(\omega,k) = \tilde{\mathbf{p}}_2(\omega)$. Otherwise, go to Step 3.
- Step 3: We have the following four cases:

— If $b_1(\omega,k) \leq 0$ and $b_2(\omega,k) \leq 0$ where $b_1(\omega,k)$ and $b_2(\omega,k)$ are two elements of $\mathbf{b}(\omega,k)$ then the values of $\hat{p}_1(\omega,k)$ and $\hat{p}_2(\omega,k)$ are set as

$$\hat{p}_1(\omega,k) = 0 \text{ and } \hat{p}_2(\omega,k) = 0.$$

— If $b_1(\omega,k) > 0$ and $b_2(\omega,k) \leq 0$ then

$$\hat{p}_1(\omega,k) = \frac{b_1(\omega,k)}{a_{1,1}(\omega)} \text{ and } \hat{p}_2(\omega,k) = 0$$

where $a_{1,1}(\omega)$ is the (1,1)th element of $\mathbf{A}(\omega)$.
— If $b_1(\omega,k) \leq 0$ and $b_2(\omega,k) > 0$ then

$$\hat{p}_1(\omega,k) = 0 \text{ and } \hat{p}_2(\omega,k) = \frac{b_2(\omega,k)}{a_{2,2}(\omega)}.$$

— Otherwise, we have $b_1(\omega,k) > 0$ and $b_2(\omega,k) > 0$. There are two cases:
  - If

$$\frac{b_2^2(\omega,k)}{a_{2,2}(\omega)} \leq \frac{b_1^2(\omega,k)}{a_{1,1}(\omega)}$$

where $a_{2,2}(\omega)$ is the (2,2)th element of $\mathbf{A}(\omega)$ then

$$\hat{p}_1(\omega,k) = \frac{b_1(\omega,k)}{a_{1,1}(\omega)} \text{ and } \hat{p}_2(\omega,k) = 0.$$

  - Otherwise the two values are given as

$$\hat{p}_1(\omega,k) = 0 \text{ and } \hat{p}_2(\omega,k) = \frac{b_2(\omega,k)}{a_{2,2}(\omega)}.$$

$\square$

REFERENCES

[1] *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds.. New York: Springer-Verlag, 2001.
[2] *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds.. New York: Springer-Verlag, 2005.
[3] Y. Grenier, "A microphone array for car environment," *Speech Commun.*, vol. 12, pp. 25–39, Dec. 1993.
[4] J. Meyer and K. U. Simmer, "Multichannel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1997, vol. 2, pp. 1167–1170.
[5] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of hands-free speaker input in cars," *IEEE Trans. Vehicular Technol.*, vol. 42, no. 4, pp. 514–518, Nov. 1993.
[6] W. Kellermann, "A self-steering digital microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1991, vol. 5, pp. 3581–3584.
[7] D. Ward, R. Kenedy, and R. Williamson, "Chapter constant directivity beamforming," in *Microphone Arrays: Signal Processing Techniques and Applications*. New York: Springer-Verlag, 2001, pp. 3–17.
[8] J. Bitzer and K. U. Simmer, "Chapter superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*. New York: Springer-Verlag, 2001, pp. 19–38.
[9] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
[10] S. Nordholm, I. Claesson, and N. Grbić, "Performance limits in subband beamforming," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 193–203, May 2003.

[11] S. Nordholm, H. H. Dam, and A. Cantoni, "Uniform FIR filterbank optimization with group delay specifications," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4249–4260, Nov. 2005.

[12] S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: Analytical evaluation," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 3, pp. 241–252, May 1999.

[13] G. Fudge and D. Linebarger, "A calibrated generalized sidelobe canceller for wideband beamforming," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2871–2875, Oct. 1994.

[14] J. Sachar, H. Silverman, and W. Patterson, III, "Position calibration of large-aperture microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2002, vol. 2, pp. 1797–1800.

[15] N. Grbić and S. Nordholm, "Soft constrained subband beamforming for handsfree speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2002, vol. 1, pp. 885–888.

[16] H. Q. Dam, S. Nordholm, H. H. Dam, and S. Y. Low, "Maximum likelihood estimation and Cramer–Rao lower bounds for the multi-channel spectral evaluation in hands-free communication," in *Proc. IEEE Asia–Pacific Conf. Commun.*, Oct. 2005, pp. 51–54.

[17] H. Q. Dam, S. Nordholm, H. H. Dam, and S. Y. Low, "Post-filtering with multi-channel power spectral estimation using joint diagonalization in multi-speaker environments," in *Proc. IEEE Asia–Pacific Conf. Commun.*, Aug. 2006, pp. 1–5.

[18] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. Eur. Signal Process. Conf.*, Sep. 1994, pp. 1182–1185.

[19] A. Davis, S. Nordholm, and R. Togneri, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 14, no. 2, pp. 412–424, Mar. 2006.

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

[21] 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.

[22] 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.

[23] S. Shahbazpanahi, A. B. Gershman, Z. Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.

[24] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–446, Apr. 1985.

[25] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[26] C. Mohan and M. Kannan, *Optimization: Theory and Practice*. Harrow, U.K.: Alpha Science Int., Ltd., 2004.

Hai Huyen Dam received the B.S. degree (first class honors) and the Ph.D. degree (with distinction) from the Curtin University of Technology, Perth, Australia, in 1996 and 2001, respectively.

From 1999 to 2000, she spent one year at the Blekinge Institute of Technology, Ronneby, Sweden, as a Visiting Research Associate. From 2001 to 2007, she was with the Western Australian Telecommunications Research Institute (WATRI), Perth, as a Research Fellow and Senior Research Fellow. Since 2006, she has been a Senior Lecturer with the Department of Mathematics and Statistics, Curtin University of Technology, Perth. Her research interests are adaptive array processing, optimization, equalization, and filter design.

Hai Quang Dam received the B.S. degree (with distinction) from the State University of Minsk, Monsk, Belarus, in 1985 and the Ph.D. degree from the Curtin University of Technology, Perth, Australia, in 2006.

From 1995 to 2001, he was a Lecturer in the University of Natural Sciences, Vietnam National University, Ho Chi Minh City, Vietnam. In 2002, he joined Western Australian Telecommunications Research Institute (WATRI) as a Visiting Researcher. Currently, he is a Lecturer with the University of Information Technology, Vietnam National University, Ho Chi Minh City. His research interests include speech enhancement, array processing, data communication networks, and wireless communications.

Sven Nordholm (SM'04) was born in Tollarp, Sweden, 1960. He received the M.Sc.EE (Civilingenjör) degree in 1983, the Licentiate of engineering in 1989, and the Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1992.

He is a Professor and Research Director of the Signal Processing Laboratory, Western Australian Telecommunication Research Institute, A joint institute between the University of Western Australia and Curtin University of Technology. He was one of the founders of the Department of Signal Processing, Blekinge Institute of Technology (BTH), Ronneby, Sweden, in 1990. At BTH, he held positions as Lecturer, Senior Lecturer, Associate Professor, and Professor. Since 1999, he has been with the Curtin University of Technology, Perth, Australia. From 1999 to 2002, he was Director of ATRI and Professor at Curtin University of Technology. He is also Chief Technology Officer and cofounder of a start-up company, Sensear. He is an Associate Editor for the *EURASIP Journal on Advances in Signal Processing*. His main research efforts have been spent in the fields of speech enhancement, adaptive and optimum microphone arrays, acoustic echo cancellation, adaptive signal processing, subband adaptive filtering, and filter design.