On the Time-Domain Widely Linear LCMV Filter for Noise Reduction With a Stereo System

Jingdong Chen, *Senior Member, IEEE*, and Jacob Benesty

***Abstract*—This paper deals with the problem of noise reduction in stereo sound systems where the objective is not only to reduce noise, but also to preserve the spatial information of both the desired speech and noise sources so that the listener can still localize the speech and noise sources by listening to the enhanced binaural outputs. To achieve this objective, we use the widely linear (WL) framework developed previously and convert the problem of binaural noise reduction into one of monaural filtering with complex signals. We then present a way to decompose both the complex speech and noise signal vectors into two orthogonal components: one correlated and the other uncorrelated with the corresponding current signal sample. With this decomposition, the problem of noise reduction with preservation of the spatial information of speech and noise sources is formulated as an optimization problem with two constraints: one on the desired speech and the other on the preservation of the noise signal. We then derive a WL linearly constrained minimum variance (LCMV) filter, which can take advantage of the statistics and noncircularity of the complex speech signal to achieve noise reduction. In contrast to the WL Wiener and minimum variance distortionless response (MVDR) filters developed previously that can only preserve the characteristics and spatial information of the desired sound source, this new WL LCMV filter has the potential to reduce noise while preserving the characteristics and spatial information of both the desired and noise sources at the same time. Experimental results are provided to justify the claimed merits of the proposed WL LCMV filter.**

***Index Terms*—Binaural noise reduction, linearly constrained minimum variance (LCMV) filter, noncircularity, speech enhancement, stereo sound system, time domain, widely linear (WL) estimation.**

I. Introduction

NOISE reduction with a stereophonic (or simply stereo) setup has emerged as a very important problem as stereo sound systems and devices are being more and more deployed in modern voice communications. The basic problem is to process the stereo input signals such as to mitigate the noise effect, thereby producing two (binaural) outputs with less amount of noise. But the mitigation process is required to preserve the spatial information of the sound sources so that, after noise reduction, the listener will still be able to localize the sound sources thanks to his/her binaural hearing mechanism. Apparently, applying a traditional single-channel noise reduction technique to each one of the stereo channels may not result in satisfactory performance as the spatial effects are generally destroyed. As a result, a great deal of efforts have been devoted to developing new algorithms that can achieve noise reduction and preserve sound spatial information at the same time. Broadly, these approaches can be classified into three main categories.

The first category treats the problem as a particular case of microphone arrays. The basic idea is to use the two inputs to form two beams at the same time with one generating an output for the left channel and the other producing an output for the right channel [1]–[9]. The sound spatial effect can be retained by forcing the time difference of arrival (TDOA) between the two beamformers' outputs to be the same as that of the desired signals at the two stereo input channels. While it is a viable approach, beamforming with two inputs usually can only produce a small amount of noise reduction, particularly in reverberant and noisy environments where beamforming has to perform denoising and dereverberation simultaneously.

The second class was influenced by the single-channel noise reduction technique but constraints were introduced to confine the noise reduction filter to retain sound spatial cues. The earliest such effort can be found in the hearing-aids area [10]. The basic idea in [10] is similar to the widely-known spectral subtraction [11] or parametric Wiener filter [12]–[14] but it poses a constraint on the suppression of each frequency band to preserve the spatial information of the desired sound source. If the band has the interaural time and level differences characteristic of the desired source, this band is kept unchanged; otherwise, the band is suppressed. This method was refined in [15] and then extended to a Wiener filter framework with the use of head-related transfer functions (HRTFs) for noise estimation [16]. While it can possibly obtain more noise reduction than beamforming, this second category of approaches generally add distortion to the desired speech. Moreover, it requires the *a priori* knowledge about the interaural time and level differences of the source signal, which is difficult if not impossible to acquire in many applications such as teleconferencing.

The third category is through the use of multichannel noise reduction principles such as the transfer function based generalized sidelobe canceller (TF-GSC) [17], the multichannel Wiener filter [18], and the spatio-temporal prediction method [19]. Since the multichannel noise reduction techniques are formulated to estimate the desired signals observed at the microphones, the spatial information should be naturally preserved. As a result, this class of approaches seems to be more advantageous than the other two in terms of practical usage. In an earlier contribution [20], we developed a novel framework, which basically has the same theoretical foundation as the third category of methods. Briefly, we first combine the stereo inputs

Manuscript received July 09, 2012; revised December 23, 2012, February 14, 2013; accepted February 19, 2013. Date of publication March 20, 2013; date of current version March 22, 2013. The work of J. Chen was supported in part by the Anhui Science and Technology Project (11010202191). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Nobutaka Ono.

J. Chen is with Northwestern Polytechnical University, Xi'an 710072, China (e-mail: jingdongchen@ieee.org).

J. Benesty is with INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

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.2013.2248719

together to form a complex signal: its real part corresponds to one input channel and its imaginary part corresponds to the other input channel of the stereo system. Similarly, we combine the two expected output channels into a complex signal. By doing so, the binaural noise reduction problem is converted to a monaural one. We then apply the so-called widely linear (WL) estimation theory to derive a number of optimal binaural noise reduction filters such as the Wiener, minimum variance distortionless response (MVDR), and tradeoff filters. It has been shown that these WL filters can achieve effective noise reduction while preserving the sound spatial information. In comparison with the traditional algorithms in the third category, this new framework offers a number of theoretical as well as practical advantages, including: 1) it converts the original two-input two-output problem into one of (complex) single-channel noise reduction, thereby providing a more convenient and compact way in defining the cost function and deriving the optimal filters; 2) the speech and noise statistics of both input channels can be jointly estimated with one estimator, and therefore the spatial information can be more naturally preserved; and 3) efficient algorithms for implementation can be developed.

However, all the aforementioned techniques suffer from two common limitations. First, the noise characteristics are generally modified by the noise reduction filter and, as a result, the residual noise remained in the enhanced signal would sound much different from the original noise. Second, the spatial information of the noise sources will be gone. In many applications such as teleconferencing and hearing aids, it is important that we preserve some degree of the characteristics and spatial information of the noise while performing noise reduction in order that: 1) the residual noise would be perceived as the same type of the original noise so that the listener can judge the noise environment where the signal is captured; and 2) the listener can still localize the noise source after noise reduction. For instance, in car noise environments, we would expect that after noise reduction the car noise can still be perceived as car noise and the listener can still localize where the car comes from to avoid accidents.

Although it is a very important problem, how to achieve noise reduction and preserving noise characteristics and spatial information at the same time has not been systematically addressed in the literature. This paper is dedicated to this problem. The fundamental goal consists of three aspects: 1) achieving noise reduction, thereby improving the signal-to-noise ratio (SNR) and speech quality; 2) recovering the spatial information of the desired sound source so that the listener can localize the desired sound source by listening to the enhanced binaural outputs; and 3) preserving some degree of the characteristics and spatial information of the noise signal so that the listener can still judge the original noise environment and localize the noise source from where the noise originates by listening to the residual noise. To achieve this goal, we adopt the framework developed previously in [20], i.e., we merge the two input channels into a complex signal (one channel being the real part and the other being the imaginary part) and also combine the two expected output channels into a complex signal. Then, binaural noise reduction is achieved on a sample-by-sample basis where the complex clean speech sample at every time instant is estimated by processing a vector of the complex noisy speech signal through widely linear (WL) filtering. With this formulation, our problem at hand becomes how to design an optimal WL noise reduction filter. In order to find such an optimal filter, we present a way to decompose the complex clean speech vector at every time instant into two orthogonal components: one correlated and the other uncorrelated with the complex speech sample at the current time instant. Similarly, but with a different purpose, we decompose the complex noise vector into two orthogonal components. With these two orthogonal decompositions, the problem of noise reduction with preservation of the spatial information of speech and noise sources can be formulated as an optimization problem with two constraints: one on the desired speech and the other on the preservation of the noise signal. We then derive a WL linearly constrained minimum variance (LCMV) filter, which can take advantage of the statistics and noncircularity of the complex speech signal to achieve noise reduction. In comparison with the work in [20], the major contribution of this work is threefold. 1) It formulates the noise reduction problem as not only recovering the desired speech signal and its spatial information, but also preserving the noise characteristics and the spatial information of the noise sources. 2) It introduces a way how to achieve the goal of preserving noise characteristics and the spatial information of the noise sources in noise reduction: i.e., we propose an orthogonal decomposition to decompose the noise signal vector into two components: one is correlated with the noise sample at the current time instant and consists of the noise characteristics and spatial information; and the other is the uncorrelated noise. This way, we can preserve the noise characteristics and spatial information by preserving the first component while the uncorrelated part should be minimized in noise reduction. 3) It is the first time in the field of noise reduction that an algorithm (LCMV) is derived that can achieve noise reduction and preserve the characteristics and spatial information of both the desired and noise sources at the same time.

The reminder of this paper is organized as follows. In Section II, we formulate the binaural noise reduction problem in stereo systems. We then briefly review the WL estimation theory and show how this theory can be applied to the binaural noise reduction problem in Section III. Section IV presents a way to decompose the desired signal and noise vectors into two orthogonal components. In Section V, we derive the WL LCMV filter. Section VI presents some experiments to validate the theoretical derivations. Finally, we give our conclusions in Section VII.

II. Signal Model and Problem Formulation

In this paper, we consider the signal model in which two microphones (that we refer to as right and left) capture a source signal convolved with acoustic impulse responses in some noise field. The signals received at the right and left microphones, at the discrete-time index k, are then expressed as

$$y_{\mathrm{R}}(k) = g_{\mathrm{R}}(k) * s(k) + v_{\mathrm{R}}(k) = x_{\mathrm{R}}(k) + v_{\mathrm{R}}(k), \quad (1a)$$

$$y_{\mathrm{L}}(k) = g_{\mathrm{L}}(k) * s(k) + v_{\mathrm{L}}(k) = x_{\mathrm{L}}(k) + v_{\mathrm{L}}(k), \quad (1b)$$

where $g_{\mathrm{R}}(k)$ [resp. $g_{\mathrm{L}}(k)$] is the impulse response from the unknown speech source $s(k)$ to the microphone on the right (resp. left), $*$ stands for linear convolution, and $v_{\mathrm{R}}(k)$ [resp. $v_{\mathrm{L}}(k)$] is the additive noise at the microphone on the right (resp. left). We assume that all the signals $x_{\rm R}(k), x_{\rm L}(k), v_{\rm R}(k)$, and $v_{\rm L}(k)$ are zero mean, and $x_{\rm R}(k)$ and $x_{\rm L}(k)$ are uncorrelated with $v_{\rm R}(k)$ and $v_{\rm L}(k)$. The two noise signals $v_{\rm R}(k)$ and $v_{\rm L}(k)$ can be either uncorrelated or correlated (e.g., from a same point source) but they are assumed to be non-speech and stationary so that their statistics can be estimated with the help of a voice activity detector (VAD) during silences.

The problem tackled in this paper is one of recovering the signals $x_{\rm R}(k)$ and $x_{\rm L}(k)$ given the observations $y_{\rm R}(k)$ and $y_{\rm L}(k)$. It is clear that our objective is to attenuate the contribution of the noise terms $v_{\rm R}(k)$ and $v_{\rm L}(k)$ as much as possible, and meanwhile preserve $x_{\rm R}(k)$ and $x_{\rm L}(k)$ with their spatial information so that with the enhanced signals, along with our binaural hearing process, we will still be able to localize the source $s(k)$. Furthermore, it is desirable to retain some degree of the noise characteristics, which is very important for some applications like hearing aids where the noise spectral and spatial information need to be preserved for safety reasons. We have stereo signals in model (1) but it is more convenient, as shown in [20], to work in the complex domain in order that the original (binaural) problem is transformed to a monaural noise reduction processing. Indeed, from the two real microphone signals given in (1a) and (1b), we can form the complex microphone signal as

$$y(k) = y_{\rm R}(k) + jy_{\rm L}(k) = x(k) + v(k), \tag{2}$$

where $j = \sqrt{-1}, x(k) = x_{\rm R}(k) + jx_{\rm L}(k)$ is the complex desired signal, and $v(k) = v_{\rm R}(k) + jv_{\rm L}(k)$ is the complex additive noise. Now, our problem can be described as follows: given the complex microphone signal, $y(k)$, which is a mixture of two uncorrelated complex signals $x(k)$ and $v(k)$, we attempt to minimize the effect of $v(k)$ while preserving $x(k)$ (i.e., our desired signal). This can be achieved by filtering the complex microphone signal. Therefore, the core issue of our problem is to find an optimal complex noise reduction filter. However, since we deal with complex signals that are noncircular, the classical linear filtering techniques [21]–[23] that are developed to estimate the optimal noise reduction filters for real signals cannot be directly applied. Instead, we need to use the so-called WL estimation theory, which will be discussed in the following sections.

III. Widely Linear Filtering

As can be noticed from the signal model given in (2), we deal with complex random variables. A very important statistical characteristic of a complex random variable (CRV) is the so-called circularity property or lack of it (noncircularity) [24], [25]. A zero-mean CRV, z, is circular if and only if the only nonnull moments and cumulants are the moments and cumulants constructed with the same power in z and z^* [26], [27], where the superscript * denotes complex conjugation. In particular, z is said to be a second-order circular CRV (CCRV) if its so-called pseudo-variance [24] is equal to zero, i.e., $E(z^2) = 0$, while its variance is nonnull, i.e., $E(|z|^2) \neq 0$. This means that the second-order behavior of a CCRV is well described by its variance. If the pseudo-variance $E(z^2)$ is not equal to 0, the CRV z is then noncircular. A good measure of the second-order circularity is the circularity quotient [24] defined as the ratio between the pseudo-variance and the variance, i.e.,

$$\gamma_z \triangleq \frac{E(z^2)}{E(|z|^2)}. \tag{3}$$

It is easy to show that $0 \leq |\gamma_z| \leq 1$. If $\gamma_z = 0, z$ is a second-order CCRV; otherwise, z is noncircular. A larger value of $|\gamma_z|$ indicates that the CRV z is more noncircular.

From the circularity quotient defined in (3), one can check that [20]

$$\begin{aligned}\gamma_x &= \frac{E[x^2(k)]}{E[|x(k)|^2]} \\ &= \frac{E\left[x_{\rm R}^2(k)\right] - E\left[x_{\rm L}^2(k)\right] + 2jE[x_{\rm R}(k)x_{\rm L}(k)]}{\sigma_x^2},\end{aligned} \tag{4}$$

where $\sigma_x^2 = E[|x(k)|^2]$ is the variance of the signal $x(k)$. Now, the complex random variable $x(k)$ is second-order circular (i.e., $\gamma_x = 0$) if and only if

$$E\left[x_{\rm R}^2(k)\right] = E\left[x_{\rm L}^2(k)\right] \quad \text{and} \quad E[x_{\rm R}(k)x_{\rm L}(k)] = 0. \tag{5}$$

Since the signals $x_{\rm R}(k)$ and $x_{\rm L}(k)$ come from the same source, they are in general correlated. As a result, the second condition in (5) should not be true. Therefore, we can safely state that the complex desired signal, $x(k)$, is noncircular. Similarly, we can check the complex microphone signal, $y(k)$, is noncircular either. However, the noise term $v(k)$ can be either circular or noncircular. If we assume that the noise components at the two microphones are uncorrelated and have the same power then $\gamma_v = 0$ and $v(k)$ is a second-order CCRV; otherwise, $v(k)$ is also noncircular.

Since we deal with noncircular CRVs, the classical linear estimation technique [21]–[23], which is developed for processing real signals or CCRVs, cannot be directly applied to recover $x(k)$. Instead, an estimate of $x(k)$ should be obtained using the WL estimation theory as [25], [28]

$$\hat{x}(k) = \mathbf{h}^H\mathbf{y}(k) + \mathbf{h}'^H\mathbf{y}^*(k) = \tilde{\mathbf{h}}^H\tilde{\mathbf{y}}(k), \tag{6}$$

where

$$\mathbf{y}(k) \triangleq [y(k)\; y(k-1)\cdots y(k-L+1)]^T = \mathbf{x}(k) + \mathbf{v}(k) \tag{7}$$

is a vector consisting of L successive noisy signal samples, superscripts ^{H} and ^{T} denote transpose-conjugate and transpose, respectively, $\mathbf{x}(k)$ and $\mathbf{v}(k)$ are defined in a similar way to $\mathbf{y}(k), \mathbf{h}$ and $\mathbf{h}'$ are two complex finite-impulse-response (FIR) filters of length L, and

$$\tilde{\mathbf{h}} \triangleq \begin{bmatrix}\mathbf{h} \\ \mathbf{h}'\end{bmatrix} \tag{8}$$

$$\tilde{\mathbf{y}}(k) \triangleq \begin{bmatrix}\mathbf{y}(k) \\ \mathbf{y}^*(k)\end{bmatrix} \tag{9}$$

are the augmented WL filter and observation vector, respectively, both of length $2L$.

With the WL filtering model given in (6), our binaural noise reduction problem becomes one of finding an optimal filter $\tilde{\mathbf{h}}$.

But before addressing how to estimate such a filter, we first discuss how to decompose the speech and noise signal vectors $\mathbf{x}(k)$ and $\mathbf{v}(k)$ to properly form the optimization cost function.

IV. Decomposition of the Speech and Noise Signal Vectors

A. Orthogonal Decomposition of the Speech Signal Vector

We can rewrite (6) as

$$\hat{x}(k) = \tilde{\mathbf{h}}^H[\tilde{\mathbf{x}}(k) + \tilde{\mathbf{v}}(k)] = \tilde{\mathbf{h}}^H\tilde{\mathbf{x}}(k) + \tilde{\mathbf{h}}^H\tilde{\mathbf{v}}(k), \tag{10}$$

where $\tilde{\mathbf{x}}(k)$ and $\tilde{\mathbf{v}}(k)$ are defined in a similar way to $\tilde{\mathbf{y}}(k)$. The first term on the right-hand side of (10), i.e., $\tilde{\mathbf{h}}^H\tilde{\mathbf{x}}(k)$, is a filtered version of the desired signal vector and its conjugate. This term can be partitioned into two components: one correlated and the other uncorrelated with the desired signal sample $x(k)$. To see this clearly, let us first decompose $\mathbf{x}(k)$ as

$$\mathbf{x}(k) = x(k)\boldsymbol{\rho}_x + \mathbf{x}'(k), \tag{11}$$

where

$$\begin{aligned}\boldsymbol{\rho}_x &\triangleq \frac{E\left[\mathbf{x}(k)x^*(k)\right]}{\sigma_x^2}\\ &= \left[\rho_{x,0}\ \rho_{x,1}\ \cdots\ \rho_{x,L-1}\right]^T\end{aligned} \tag{12}$$

is the (normalized) correlation vector (of length L) between $\mathbf{x}(k)$ and $x(k)$,

$$\rho_{x,l} \triangleq \frac{E[x(k-l)x^*(k)]}{\sigma_x^2} \tag{13}$$

is the correlation coefficient between $x(k-l)$ and $x(k)$ with $|\rho_{x,l}| \leq 1$, and

$$\mathbf{x}'(k) = \mathbf{x}(k) - x(k)\boldsymbol{\rho}_x \tag{14}$$

is the interference signal vector. Obviously, $x(k)\boldsymbol{\rho}_x$ is correlated with $x(k)$ and

$$E[\mathbf{x}'(k)x^*(k)] = \mathbf{0}, \tag{15}$$

so $\mathbf{x}'(k)$ is uncorrelated with $x(k)$.

Similarly, we have

$$\mathbf{x}^*(k) = x(k)\boldsymbol{\gamma}_x^* + \mathbf{x}''(k), \tag{16}$$

where

$$\begin{aligned}\boldsymbol{\gamma}_x &\triangleq \frac{E[\mathbf{x}(k)\,x(k)]}{\sigma_x^2}\\ &= \left[\gamma_{x,0}\ \gamma_{x,1}\ \cdots\ \gamma_{x,L-1}\right]^T\end{aligned} \tag{17}$$

is the (normalized) correlation vector (of length L) between $\mathbf{x}(k)$ and $x^*(k)$,

$$\gamma_{x,l} \triangleq \frac{E[x(k-l)x(k)]}{\sigma_x^2} \tag{18}$$

is the correlation coefficient[1] between $x(k-l)$ and $x^*(k)$ with $|\gamma_{x,l}| \leq 1$, and

$$\mathbf{x}''(k) = \mathbf{x}^*(k) - x(k)\boldsymbol{\gamma}_x^* \tag{19}$$

is the interference signal vector. Clearly, $x(k)\boldsymbol{\gamma}_x^*$ is correlated with $x(k)$, while $\mathbf{x}''(k)$ and $x(k)$ are uncorrelated since

$$E[\mathbf{x}''(k)x^*(k)] = \mathbf{0}. \tag{20}$$

Combining (11) and (16), we get

$$\tilde{\mathbf{x}}(k) = x(k)\mathbf{d}_x + \tilde{\mathbf{x}}'(k) = \tilde{\mathbf{x}}_{\rm d}(k) + \tilde{\mathbf{x}}'(k), \tag{21}$$

where

$$\mathbf{d}_x \triangleq \begin{bmatrix}\boldsymbol{\rho}_x\\ \boldsymbol{\gamma}_x^*\end{bmatrix}, \tag{22}$$

$$\tilde{\mathbf{x}}_{\rm d}(k) \triangleq x(k)\mathbf{d}_x \tag{23}$$

is correlated with the desired signal, $x(k)$, and will contribute to its estimation, so we call it the desired signal vector, and

$$\tilde{\mathbf{x}}'(k) \triangleq \begin{bmatrix}\mathbf{x}'(k)\\ \mathbf{x}''(k)\end{bmatrix} \tag{24}$$

is uncorrelated with $x(k)$, and will interfere with the estimation, so we call it the interference signal vector. Note that the three vectors $\mathbf{d}_x, \boldsymbol{\rho}_x$, and $\boldsymbol{\gamma}_x$ are generally time varying since speech signals are nonstationary. However, we write these vectors without using a time index to make the notation compact.

B. Orthogonal Decomposition of the Noise Signal Vector

The second term on the right-hand side of (10), which is uncorrelated with the desired speech signal, is the residual noise. Following the same line of ideas in decomposing the signal vector $\tilde{\mathbf{x}}(k)$, we can decompose the noise signal vector, $\tilde{\mathbf{v}}(k)$, into two orthogonal vectors as

$$\tilde{\mathbf{v}}(k) = v(k)\mathbf{d}_v + \tilde{\mathbf{v}}'(k) = \tilde{\mathbf{v}}_{\rm c}(k) + \tilde{\mathbf{v}}'(k), \tag{25}$$

where $\mathbf{d}_v$ and $\tilde{\mathbf{v}}'(k)$ are defined in a similar way to $\mathbf{d}_x$ and $\tilde{\mathbf{x}}'(k)$, and $\tilde{\mathbf{v}}_{\rm c}(k) \triangleq v(k)\mathbf{d}_v$ denotes the noise component that is correlated with the current noise sample $v(k)$.

C. Decomposition of the Signal Estimate

Substituting (21) and (25) into (10), we obtain

$$\begin{aligned}\hat{x}(k) &= \tilde{\mathbf{h}}^H[\tilde{\mathbf{x}}_{\rm d}(k) + \tilde{\mathbf{x}}'(k) + \tilde{\mathbf{v}}_{\rm c}(k) + \tilde{\mathbf{v}}'(k)]\\ &= \tilde{\mathbf{h}}^H[x(k)\mathbf{d}_x + \tilde{\mathbf{x}}'(k) + v(k)\mathbf{d}_v + \tilde{\mathbf{v}}'(k)]\\ &= x_{\rm fd}(k) + x'_{\rm ri}(k) + v_{\rm rnc}(k) + v'_{\rm rnuc}(k),\end{aligned} \tag{26}$$

where $x_{\rm fd}(k) \triangleq \tilde{\mathbf{h}}^H\ \tilde{\mathbf{x}}_{\rm d}(k) = x(k)\tilde{\mathbf{h}}^H\mathbf{d}_x$ is the filtered desired signal, $x'_{\rm ri}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{x}}'(k)$ is the residual interference, $v_{\rm rnc}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{v}}_{\rm d}(k) = v(k)\tilde{\mathbf{h}}^H\mathbf{d}_v$ is the residual noise component that is correlated with $v(k)$, and $v'_{\rm rnuc}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{v}}'(k)$ is the uncorrelated residual noise part. Since the four terms on the right-hand side of (26) are mutually uncorrelated, the variance of $\hat{x}(k)$ is

$$\sigma_{\hat{x}}^2 = \sigma_{x_{\rm fd}}^2 + \sigma_{x'_{\rm ri}}^2 + \sigma_{v_{\rm rnc}}^2 + \sigma_{v'_{\rm rnuc}}^2, \tag{27}$$

where

$$\sigma_{x_{\rm fd}}^2 = \sigma_x^2|\tilde{\mathbf{h}}^H\mathbf{d}_x|^2 = \tilde{\mathbf{h}}^H\mathbf{R}_{\tilde{\mathbf{x}}_{\rm d}}\tilde{\mathbf{h}}, \tag{28}$$

$$\sigma_{x'_{\rm ri}}^2 = \mathbf{h}^H\mathbf{R}_{\tilde{\mathbf{x}}'}\tilde{\mathbf{h}} = \tilde{\mathbf{h}}^H\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{h}} - \sigma_x^2|\tilde{\mathbf{h}}^H\mathbf{d}_x|^2, \tag{29}$$

[1] Note that $\gamma_{x,0} = \gamma_x$ is the circularity quotient of the complex signal $x(k)$.

$$\sigma^2_{v_{\text{rnc}}} = \sigma_v^2 |\tilde{\mathbf{h}}^H \mathbf{d}_v|^2 = \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{v}}_{\text{c}}} \tilde{\mathbf{h}}, \quad (30)$$

$$\sigma^2_{v'_{\text{rnuc}}} = \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{v}}'} \tilde{\mathbf{h}} = \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{v}}} \tilde{\mathbf{h}} - \sigma_v^2 |\tilde{\mathbf{h}}^H \mathbf{d}_v|^2, \quad (31)$$

$\mathbf{R}_{\tilde{\mathbf{x}}_{\text{d}}} = \sigma_x^2 \mathbf{d}_x \mathbf{d}_x^H$ is the correlation matrix (whose rank is equal to 1) of $\tilde{\mathbf{x}}_{\text{d}}(k)$, $\mathbf{R}_{\tilde{\mathbf{v}}_{\text{c}}} = \sigma_v^2 \mathbf{d}_v \mathbf{d}_v^H$ is the correlation matrix of $\tilde{\mathbf{v}}_{\text{c}}(k), \sigma_v^2$ is the variance of $v(k)$, and $\mathbf{R}_{\tilde{\mathbf{x}}'} = E[\tilde{\mathbf{x}}'(k)\tilde{\mathbf{x}}'^H(k)]$, $\mathbf{R}_{\tilde{\mathbf{x}}} = E[\tilde{\mathbf{x}}(k)\tilde{\mathbf{x}}^H(k)]$, $\mathbf{R}_{\tilde{\mathbf{v}}'} = E[\tilde{\mathbf{v}}'(k)\tilde{\mathbf{v}}'^H(k)]$, and $\mathbf{R}_{\tilde{\mathbf{v}}} = E[\tilde{\mathbf{v}}(k)\tilde{\mathbf{v}}^H(k)]$ are the correlation matrices of $\tilde{\mathbf{x}}'(k), \tilde{\mathbf{x}}(k)$, $\tilde{\mathbf{v}}'(k)$, and $\tilde{\mathbf{v}}(k)$, respectively.

V. Linearly Constrained Minimum Variance Filter

To derive the WL LCMV filter, we need to derive first the mean-square error (MSE) criterion. We define the error signal between the estimated and desired signals as

$$e(k) \stackrel{\triangle}{=} \hat{x}(k) - x(k) = \tilde{\mathbf{h}}^H \tilde{\mathbf{y}}(k) - x(k). \quad (32)$$

Substituting (26) into (32) gives

$$e(k) = e_{\text{d}}(k) + e_{\text{c}}(k) + e_{\text{r}}(k), \quad (33)$$

where

$$e_{\text{d}}(k) \stackrel{\triangle}{=} x_{\text{fd}}(k) - x(k) = x(k)(\tilde{\mathbf{h}}^H \mathbf{d}_x - 1) \quad (34)$$

is the signal distortion due to the WL filter,

$$e_{\text{c}}(k) \stackrel{\triangle}{=} v_{\text{rnc}}(k) = v(k)(\tilde{\mathbf{h}}^H \mathbf{d}_v) \quad (35)$$

is the correlated residual noise, and

$$e_{\text{r}}(k) \stackrel{\triangle}{=} x'_{\text{ri}}(k) + v_{\text{rnuc}}(k) \quad (36)$$

represents the residual interference-plus-(uncorrelated) noise.

The MSE is then

$$J(\tilde{\mathbf{h}}) \stackrel{\triangle}{=} E[|e(k)|^2] = \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{y}}} \tilde{\mathbf{h}} = J_{\text{d}}(\tilde{\mathbf{h}}) + J_{\text{c}}(\tilde{\mathbf{h}}) + J_{\text{r}}(\tilde{\mathbf{h}}), \quad (37)$$

where $\mathbf{R}_{\tilde{\mathbf{y}}} = E[\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}^H(k)]$ is the noisy correlation matrix,

$$J_{\text{d}}(\tilde{\mathbf{h}}) \stackrel{\triangle}{=} E[|e_{\text{d}}(k)|^2] = \sigma_x^2 |\tilde{\mathbf{h}}^H \mathbf{d}_x - 1|^2, \quad (38)$$

$$J_{\text{c}}(\tilde{\mathbf{h}}) \stackrel{\triangle}{=} E[|e_{\text{c}}(k)|^2] = \sigma_v^2 |\tilde{\mathbf{h}}^H \mathbf{d}_v|^2, \quad (39)$$

$$J_{\text{r}}(\tilde{\mathbf{h}}) \stackrel{\triangle}{=} E[|e_{\text{r}}(k)|^2] = \sigma^2_{x'_{\text{ri}}} + \sigma^2_{v_{\text{rn}}} = \tilde{\mathbf{h}}^H \mathbf{R}_{\text{in}} \tilde{\mathbf{h}}, \quad (40)$$

and

$$\mathbf{R}_{\text{in}} = \mathbf{R}_{\tilde{\mathbf{x}}'} + \mathbf{R}_{\tilde{\mathbf{v}}'} \quad (41)$$

is the interference-plus-(uncorrelated) noise correlation matrix.

With the previously defined MSEs, we now derive an optimal filter that can: 1) perfectly recover our desired signal, $x(k)$; 2) preserve some degree of the noise characteristics so that the listener can still localize the noise sources; and 3) minimize the effect of $x'_{\text{ri}}(k) + v'_{\text{rnuc}}(k)$. In order to satisfy the first condition, we need to have the constraint that $\tilde{\mathbf{h}}^H \mathbf{d}_x = 1$, which can be easily seen from (34) or (38). Similarly, if we want to preserve some degree of the noise characteristics, we can force $\tilde{\mathbf{h}}^H \mathbf{d}_v = \alpha$, where $\alpha \in [0, 1]$ is a constant. Putting these two constraints together in a matrix form, we get

$$\mathbf{D}^H \tilde{\mathbf{h}} = \mathbf{i}_\alpha, \quad (42)$$

where

$$\mathbf{D} = [\, \mathbf{d}_x \quad \mathbf{d}_v \,] \quad (43)$$

is our constraint matrix of size $2L \times 2$ and

$$\mathbf{i}_\alpha = [\, 1 \quad \alpha \,]^T. \quad (44)$$

Then, the optimal filter that satisfies all the three conditions can be obtained by minimizing the MSE, $J(\tilde{\mathbf{h}})$, subject to the constraint in (42), i.e.,

$$\tilde{\mathbf{h}}_{\text{LCMV},\alpha} = \arg\min_{\tilde{\mathbf{h}}} \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{y}}} \tilde{\mathbf{h}} \quad \text{subject to} \quad \mathbf{D}^H \tilde{\mathbf{h}} = \mathbf{i}_\alpha. \quad (45)$$

Using a Lagrange multiplier to adjoin the constraint to the objective function, we deduce the solution to (45) as

$$\tilde{\mathbf{h}}_{\text{LCMV},\alpha} = \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{D} \left[\mathbf{D}^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{D} \right]^{-1} \mathbf{i}_\alpha, \quad (46)$$

where we have assumed that the product matrix $\mathbf{D}^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{D}$ is invertible. Depending on the value of the constant α, we have the following three scenarios.

1) $\alpha = 1$.
   In this case, we have $\tilde{\mathbf{h}}_{\text{LCMV},1} = \tilde{\mathbf{i}}$, where $\tilde{\mathbf{i}}$ is the first column of the identity matrix $\mathbf{I}_{2L}$ of size $2L \times 2L$. It is easy to check that with this (identity) filter there is neither speech distortion nor noise reduction.
2) $\alpha = 0$.
   In this case, the LCMV filter perfectly recovers the desired signal, $x(k)$, and completely removes the correlated components, $v(k)\mathbf{d}_v$. With some mathematical manipulation of (46), we can rewrite this LCMV filter into the following form:

$$\tilde{\mathbf{h}}_{\text{LCMV},0} = \frac{1}{1 - |\rho|^2} \tilde{\mathbf{h}}_{\text{MVDR}} - \frac{|\rho|^2}{1 - |\rho|^2} \mathbf{t}, \quad (47)$$

   where

$$|\rho|^2 = \frac{\left| \mathbf{d}_x^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_v \right|^2}{\left(\mathbf{d}_x^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_x \right) \left(\mathbf{d}_v^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_v \right)}, \quad (48)$$

   with $0 \leq |\rho|^2 \leq 1$, $\tilde{\mathbf{h}}_{\text{MVDR}}$ is the minimum distortionless filter given in (71) of [20], and

$$\mathbf{t} = \frac{\mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_v}{\mathbf{d}_x^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_v}, \quad (49)$$

   which can be viewed as a correlated noise estimator. With this decomposition, the physical meaning of the LCMV filter can be interpreted as follows: it first applies an MVDR filter that would attenuate the noise as much as possible while keeping speech from being distorted and meanwhile uses a $\mathbf{t}$ filter to estimate the correlated component of the noise signal. The residual noise in the MVDR output consists of both correlated as well as uncorrelated components. Now, by weighting the $\mathbf{t}$ filter output and then subtracting it from the weighted output of the MVDR filter, the correlated noise component will be cancelled. The amount of the correlated noise to be subtracted from the MVDR output depends on the value of $|\rho|^2$. There are two extreme cases. The first one is when $|\rho|^2 = 0$. In this case, the LCMV filter degenerates to the MVDR filter. The second case happens when $\mathbf{d}_x = \mathbf{d}_v$. In this situation, the LCMV filter does not exist since the product matrix $\mathbf{D}^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{D}$ is singular.

3) $0 < \alpha < 1$.
In this case, the LCMV filter will perfectly recover the desired signal and meanwhile achieves a compromise between the amount of noise reduction and the preservation of the noise characteristics.

Note that the LCMV filter can also be obtained by minimizing $J_{\mathrm{r}}(\tilde{\mathbf{h}})$ subject to the constraint in (42). This time, the problem can be rewritten in the following form:

$$\tilde{\mathbf{h}}_{\mathrm{LCMV},\alpha} = \arg\min_{\tilde{\mathbf{h}}} \tilde{\mathbf{h}}^H \mathbf{R}_{\mathrm{in}} \tilde{\mathbf{h}} \quad \text{subject to} \quad \mathbf{D}^H \tilde{\mathbf{h}} = \mathbf{i}_\alpha, \tag{50}$$

for which the solution is

$$\tilde{\mathbf{h}}_{\mathrm{LCMV},\alpha} = \mathbf{R}_{\mathrm{in}}^{-1} \mathbf{D} \left[\mathbf{D}^H \mathbf{R}_{\mathrm{in}}^{-1} \mathbf{D} \right]^{-1} \mathbf{i}_\alpha. \tag{51}$$

It can be checked that the two forms of the LCMV filter in (46) and (51) are theoretically identical. In practice, however, (51) may offer some implementation advantages as the $\mathbf{R}_{\mathrm{in}}$ matrix is generally better conditioned than the $\mathbf{R}_{\tilde{\mathbf{y}}}$ matrix.

Before leaving this section and moving to the next one about experiments, we would like say a few words about the estimation of the matrix $\mathbf{R}_{\mathrm{in}}$ and the correlation vector $\mathbf{d}_x$, which are needed to implement the LCMV filter. Since the noisy signal $y(k)$ is accessible, it is straightforward to compute the correlation matrix $\mathbf{R}_{\tilde{\mathbf{y}}}$, the variance σ_y^2, and the correlation vector $\mathbf{d}_y$. Let us assume that we have a noise estimator based on a VAD. As a result, the variance σ_v^2 and the correlation vector $\mathbf{d}_v$ can be estimated. We can then obtain an estimate of σ_x^2 based on the relationship $\sigma_y^2 = \sigma_x^2 + \sigma_v^2$. With the assumption that the desired speech and noise are uncorrelated, one can easily verify that

$$\sigma_y^2 \mathbf{d}_y = \sigma_x^2 \mathbf{d}_x + \sigma_v^2 \mathbf{d}_v. \tag{52}$$

Substituting the estimates of $\sigma_y^2, \mathbf{d}_y, \sigma_v^2, \mathbf{d}_v$, and σ_x^2 into (52), one can obtain an estimate of the correlation vector $\mathbf{d}_x$. It follows immediately how an estimate of $\mathbf{R}_{\mathrm{in}}$ can be achieved according to (41).

VI. Experimental Results

In this section, we study the performance of the LCMV filter through experiments. The experiments were conducted with impulse responses measured in the varechoic chamber at Bell Labs [35], [36] and also some signals recorded in a real car environment. The Bell Labs chamber is a rectangular room, which measures 6700 mm long by 6100 mm wide by 2900 mm high and is equipped with 368 electronically controlled panels. Each panel consists of two perforated sheets whose holes, if aligned, expose sound absorbing material (fiberglass) behind, but if shifted to misalign, form a highly reflective surface. Each panel can be individually controlled so that the holes on a particular panel are either fully open (absorbing) or fully closed (reflective). As a result, a total of 2^{368} different room characteristics can be generated by varying the binary states of the 368 panels in different combinations.

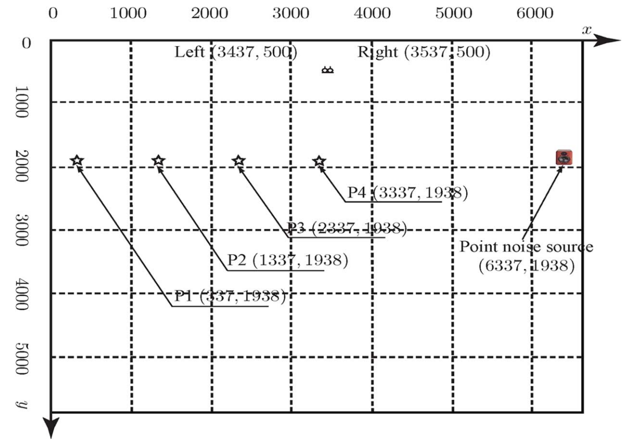

Fig. 1. Floor layout of the experimental setup in the varechoic chamber (coordinate values measured in millimeters). The two microphones are located at (3437, 500) and (3537, 500) respectively (with $z = 1400$). A speech source is placed at one of the four positions from P1 to P4 to simulate a moving talker (with $z = 1600$). A loudspeaker is placed at (6337, 1938) to play back a prerecorded car noise signal to simulate a point noise source.

A diagram of the floor layout of the experimental setup is illustrated in Fig. 1. For convenience of exposition, positions in the floor plan are designated by (x, y) coordinates with reference to the northwest corner and corresponding to millimeters along the (north, west) walls. A stereo system with two microphones is configured. The two microphones are located respectively at (3437, 500) and (3537, 500). A loudspeaker, which plays back a speech signal pre-recorded from a female talker (of 30 seconds length), is used to simulate a moving speech source and it moves back and forth from positions P1 to P4 as shown in Fig. 1. The four positions are uniformly spaced along the line $y = 1938$ with the first position P1 at (337, 1938) and the last position P4 at (3337, 1938). Another loudspeaker, playing back a pre-recorded car noise signal, is placed at (6337, 1938) to simulate a point noise source. To make the experiments repeatable, the acoustic channel impulse responses were measured from all the source positions to the two microphones. The measurement was carried out when 89% of the chamber panels were open and the corresponding reverberation time T_{60} is approximately 0.24 s. The original impulse responses were measured with a sampling rate of 48 kHz [35] and we downsampled them to 8 kHz.

A. Experiment With a Point Noise Source

In the first experiment, we consider a scenario where both the desired speech and noise are from point sources. To simulate a moving source, we changed the source position (i.e., used a new set of impulse responses) every 3.75 s first from P1 to P4 and then back. Each time the movement was restricted to the position immediately next to the current one. The noise source is a car noise signal recorded in a Volvo sedan running

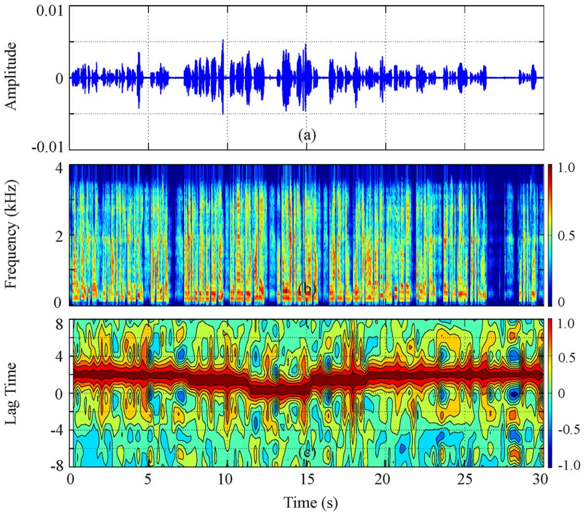

Fig. 2. Clean speech received at the microphones: (a) waveform of the left-channel signal, (b) spectrogram of the left-channel signal, and (c) contour of the cross-correlation function between the left and right channels.

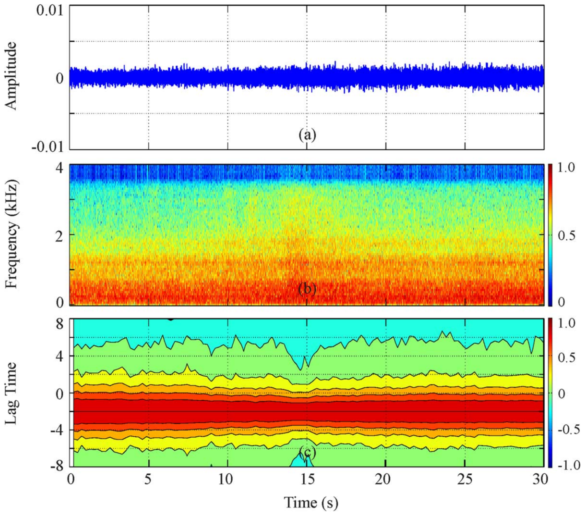

Fig. 3. Noise received at the microphones: (a) waveform of the left-channel signal, (b) spectrogram of the left-channel signal, and (c) contour of the cross-correlation function between the left and right channels.

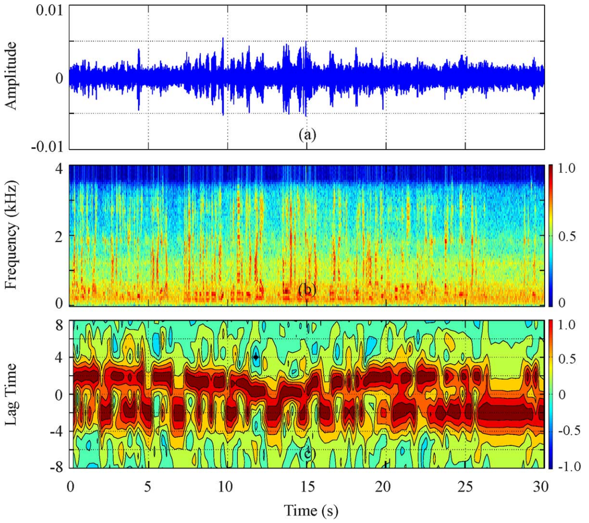

Fig. 4. Noisy speech observed at the microphones: (a) waveform of the left-channel signal, (b) spectrogram of the left-channel signal, and (c) contour of the cross-correlation function between the left and right channels. The input SNR is 5 dB.

at 55 mph on a highway with all its windows closed. Both the speech and noise sources are first convolved with the corresponding channel impulse responses and then mixed together at a signal-to-noise ratio (SNR) of 5 dB. Figs. 2 and 3 plot the waveforms and spectrograms of the (clean) speech and noise received at one of the two microphones and the noisy speech is shown in Fig. 4. To visualize the spatial sound effect, we computed the cross-correlation function between the two channels every 256 ms using a short-time average with a frame size of 256 ms. The contours of the cross-correlation functions are also plotted in Figs. 2–4, where the maxima of the cross-correlation function at each time corresponds to the current source position.

As we pointed out previously, we need to know the two matrices $\mathbf{R}_{\tilde{\mathbf{y}}}$ and $\mathbf{R}_{\mathrm{in}}$ and the two correlation vectors $\mathbf{d}_x$ and $\mathbf{d}_v$ to implement the LCMV filter derived in the previous section. Computation of the $\mathbf{R}_{\tilde{\mathbf{y}}}$ matrix is relatively straightforward because the noisy signal vector $\tilde{\mathbf{y}}(k)$ is accessible. But we need a noise estimator or a VAD in practice to compute the two correlation vectors $\mathbf{d}_x$ and $\mathbf{d}_v$. While it is a very important issue (see [37] and references therein), how to effectively estimate the noise or its statistics in a stereo system is not the main thrust of this paper. So, we will set aside this issue and compute the noisy correlation matrix and the two correlation vectors directly from the corresponding signals in most experiments (except the last one). Specifically, at each time-instant k, an estimate of the matrix $\mathbf{R}_{\tilde{\mathbf{y}}}$ is computed using a short-time average from the most recent 640 samples (40-ms long) of the noisy signal. Similarly, the correlation vectors $\mathbf{d}_x$ and $\mathbf{d}_v$ are computed directly from the clean speech and noise signals.

To evaluate the performance of a noise reduction filter, we generally need to examine both the amount of speech distortion and the degree of noise reduction due to the filter. However, since the LCMV filter does not introduce speech distortion, it is only necessary to evaluate the noise reduction part. For this purpose, we examine the input and output SNRs of the LCMV filter. The input SNR is defined as

$$\mathrm{iSNR} \triangleq \frac{\sigma_x^2}{\sigma_v^2}. \tag{53}$$

After applying the WL LCMV filter, $\tilde{\mathbf{h}}_{\mathrm{LCMV},\alpha}$, the output SNR is defined as the ratio of the variance of the filtered desired signal over the variance of the residual interference-plus-noise, i.e.,

$$\mathrm{oSNR}(\tilde{\mathbf{h}}_{\mathrm{LCMV},\alpha}) \triangleq \frac{\sigma_{x_{\mathrm{fd}}}^2}{\sigma_{x'_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rnc}}}^2 + \sigma_{v'_{\mathrm{rnuc}}}^2}. \tag{54}$$

To compute the output SNR, we first substitute the estimates of the correlation matrix $\mathbf{R}_{\tilde{\mathbf{y}}}$ and the correlation vectors $\mathbf{d}_x$ and $\mathbf{d}_v$ to (46) to obtain the LCMV filter at every time-instant k. With this filter and using (26), we get the four signals $x_{\mathrm{fd}}(k), x'_{\mathrm{ri}}(k)$, $v_{\mathrm{rnc}}(k)$, and $v'_{\mathrm{rnuc}}(k)$. We then compute the output SNR using a long-time average.

With the above conditions, the output SNR of the LCMV filter depends on two important parameters: the filter length L and the constant α. The value of L affects both the noise reduction performance and the complexity of the LCMV filter. The experimental study carried out in [20] showed that good noise reduction performance with reasonable complexity can be

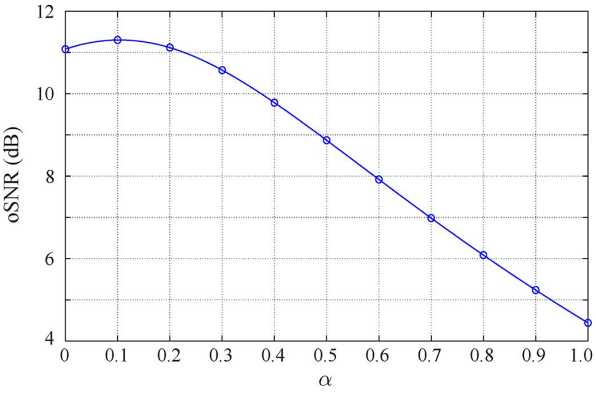

Fig. 5. The output SNR of the LCMV filter as a function of the parameter α. The input SNR is 5 dB.

achieved for different WL filters when L is between 20 and 60. Following the study in [20], we simply set L to 20 in our experiments and investigate the impact of α on noise reduction performance using experiments, i.e., we study the performance by changing the value of α from 0 to 1. Fig. 5 depicts the output SNR as a function of the parameter α. In general, the output SNR decreases as the value of α increases. The exceptional case is when α is very small where we see that the output SNR slightly increases with α. The reason for this can be explained as follows. As it was discussed previously, the noise vector is composed of two components: one correlated and the other uncorrelated with the current noise sample. For the car noise from a point source, the correlated part is dominant. As we choose a smaller value for the constant α, we get more correlated noise removed. Therefore, we see that the output SNR increases as the value of α decreases, which agrees very well with the theoretical analysis. However, as the value of α approaches to 0 and the filter focuses too much on removing the correlated noise, we would achieve less attenuation of the uncorrelated noise part. As a result, there is less overall noise attention. We will discuss more on this in the next experiment. We notice that when $\alpha = 1$, the output SNR is slightly lower than the input SNR. Theoretically, this should not happen as the LCMV filter degenerates to the identity filter and the input and output SNRs should be the same in this case. We attribute this small difference to the errors in the estimation of the correlation matrices and vectors.

To illustrate the performance of noise reduction versus the preservation of the noise characteristics, we plot in Figs. 6 and 7 the outputs of the LCMV filter for $\alpha = 0$ and $\alpha = 0.6$, respectively. It is clearly seen that the enhanced speech for $\alpha = 0$ is significantly less noisy than that for the case of $\alpha = 0.6$. However, when $\alpha = 0$, the noise spatial information is completely removed, as seen in Fig. 6(c). While there is less noise reduction for $\alpha = 0.6$, one can see that the noise spatial information is partially preserved. So, in practice, we can achieve a compromise between the amount of noise reduction and the degree of preserving noise spatial effects using the LCMV filter by setting an appropriate value of the parameter α.

B. Experiment With Both Directional and Diffuse Noise

In the second experiment, we consider a more generic case where there are both directional and diffuse noises. The directional noise is still the car noise, same as in the previous experiment. The signal-to-car-noise ratio is also 5 dB. Besides the car noise, some white Gaussian noise is added into the observation signals. As we pointed out previously, the car noise components received at the two microphones are correlated as they are from a same point source. However, the white noise components at the two channels are uncorrelated. We consider two cases: in the first one, the signal-to-white-noise ratio is 10 dB and the overall SNR (including both the car and white noises) is 3.8 dB; in the second case, the signal-to-white-noise ratio is 15 dB and the overall SNR is 4.6 dB. All the other conditions are the same as in the previous experiment. Again, we compute the noisy correlation matrix and the two correlation vectors directly from the corresponding signals. Note, however, it is difficult to distinguish between the car noise and the white Gaussian noise in practice. So, we compute the noise correlation vector from the mixture of the car and white Gaussian noises. The results for this experiment are plotted in Fig. 8. In both cases, it is seen that the output SNR first increases and then decreases as the

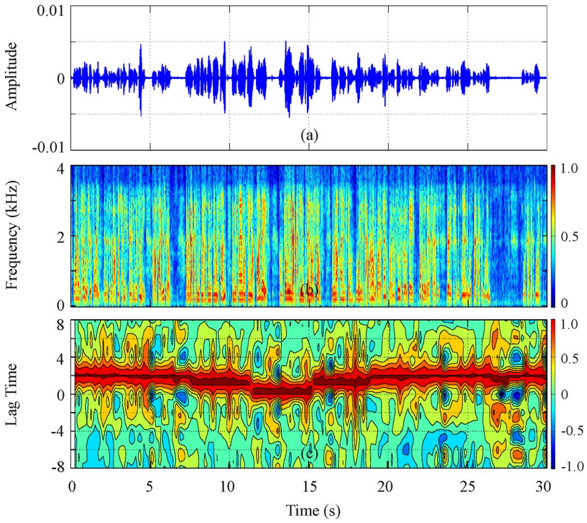

Fig. 6. The enhanced signal by the LCMV filter for $\alpha = 0$: (a) waveform of the left-channel signal, (b) spectrogram of the left-channel signal, and (c) contour of the cross-correlation function between the left and right channels. The input SNR is 5 dB and the output SNR is 11.1 dB.

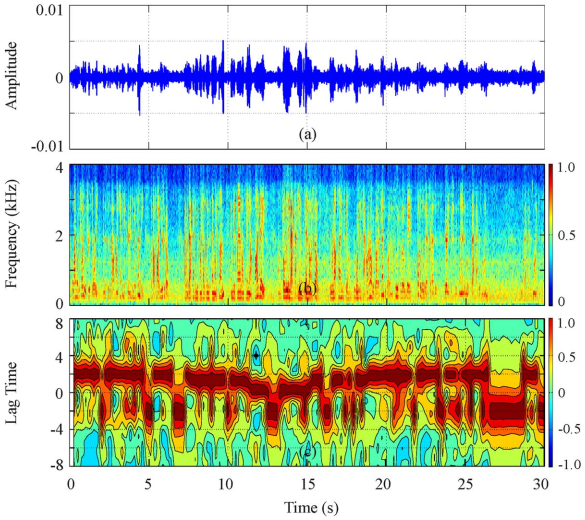

Fig. 7. The enhanced signal by the LCMV filter for $\alpha = 0.6$: (a) waveform of the left-channel signal, (b) spectrogram of the left-channel signal, and (c) contour of the cross-correlation function between the left and right channels. The input SNR is 5 dB and the output SNR is 7.9 dB.

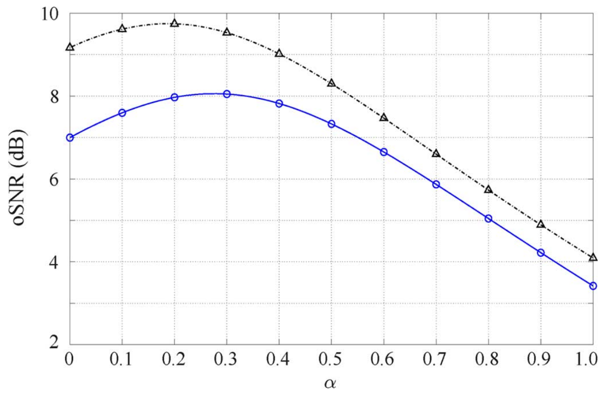

Fig. 8. The output SNR of the LCMV filter as a function of the parameter α. △: the signal-to-white-Gaussian-noise ratio is 15 dB; the signal-to-car-noise ratio is 5 dB; the overall SNR is 4.6 dB. ○: the signal-to-white-Gaussian-noise ratio is 10 dB; the signal-to-car-noise ratio is 5 dB; the overall SNR is 3.8 dB.

parameter α increases. Again, the reason for this is due to the fact that the noise consists of both correlated and uncorrelated parts. While the car noise component (from a point source) is mostly correlated, the white noise part is uncorrelated. By decreasing the value of α from 1 to 0, we get more correlated car noise reduced. When $\alpha = 0$, all the correlated noise is removed. However, as the filter focuses too much on removing correlated car noise, less amount of uncorrelated white noise is attenuated. As a result, we achieve less amount of overall noise reduction. Through experiments, we find that the highest output SNR occurs when α is approximately equal to the ratio between the intensity of the uncorrelated noise and that of the correlated noise. Therefore, a good compromise between the correlated and uncorrelated noise reduction can be achieved in practice if we can know or estimate the amount of correlated noise relative to the level of the uncorrelated noise that are in the noisy signal.

C. Comparison With WL Wiener and MVDR Filters

The WL Wiener and MVDR filters were developed in [20]. These two filters are given as follows:

$$\tilde{\mathbf{h}}_{\mathrm{W}} = \sigma_x^2 \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_x \tag{55}$$

and

$$\tilde{\mathbf{h}}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_x}{\mathbf{d}_x^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_x}. \tag{56}$$

In this experiment, we compare the performance of the WL Wiener, MVDR and LCMV filters in different SNR conditions. Again, we consider the situation where there are both directional car noise and diffuse white Gaussian noise. The signal-to-car-noise ratio is set to 10 dB. We vary the signal-to-white-noise ratio from 5 dB to 30 dB. We use the output SNR to assess the amount of noise reduction and the speech distortion index [20] to evaluate the degree of speech distortion. The speech distortion index is given by

$$\upsilon_{\mathrm{sd}}(\tilde{\mathbf{h}}) \triangleq \frac{E[|x_{\mathrm{fd}}(k) - x(k)|^2]}{\sigma_x^2}. \tag{57}$$

To compute this index, we first estimated the signal component $x_{\mathrm{fd}}(k)$ at each time instant using the corresponding optimal filter. A long-time average was then used to replace the expectation operation in (57). The results as a function of the signal-to-white-noise ratio are plotted in Fig. 9. The overall input SNR (car noise plus white noise) is also shown in the figure.

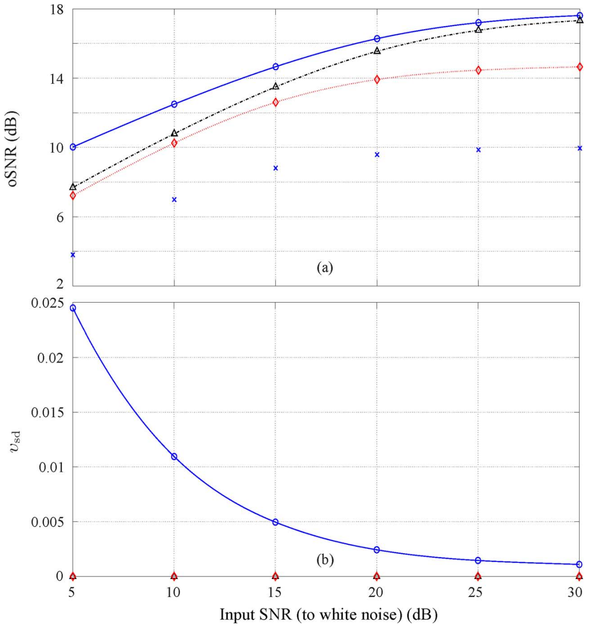

Fig. 9. The performance of the WL Wiener (○), MVDR (◇), and LCMV (◇) filters in different input SNR (to white noise) conditions: (a) the output SNR and (b) speech distortion index. For the LCMV filter, the parameter α is set to 0.6. The signal-to-car-noise ratio is 10 dB and the overall SNR is marked by × in (a).

It is seen from Fig. 9 that the Wiener filter achieves the highest output SNR; however, it adds distortion to the speech signal as shown in Fig. 9(b) where the speech distortion index for the Wiener filter is larger than 0. One can see that the value of the speech distortion index for the Wiener filter decreases as the input SNR increases. This is understandable since as the input SNR increases, there will be less amount of noise to attenuate, leading to smaller speech distortion.

In comparison, both the MVDR and LCMV filters do not introduce speech distortion as indicated by Fig. 9(b) where the value of the speech distortion index for both filters is always zero, regardless of the input SNR level. But their SNR improvement is less than that of the Wiener filter. While both the MVDR and LCMV filters do not add speech distortion and can preserve the spatial information of the desired speech signal while reducing noise, one can see that the LCMV filter yielded less SNR improvement than the MVDR filter. This is due to the fact that the LCMV filter, with $\alpha = 0.6$, needs to preserve part of the characteristics and spatial information of the noise signal. To illustrate this, Fig. 10 plots the contours of the cross-correlation functions between the left and right channels of the noisy signal and the enhanced signals by the WL Wiener, MVDR, and LCMV $(\alpha = 0.6)$ filters for the case where the signal-to-white-noise ratio is 20 dB. It is clearly seen that the noise spatial information is removed by the Wiener and MVDR filters while such information is largely preserved with the LCMV filter.

We notice from Fig. 9 that the difference in SNR improvement by the MVDR and LCMV filters becomes more significant as the signal-to-white-noise ratio increases. The underlying reason can be explained as follows. The MVDR filter is in general more effective in reducing directional noise than diffuse

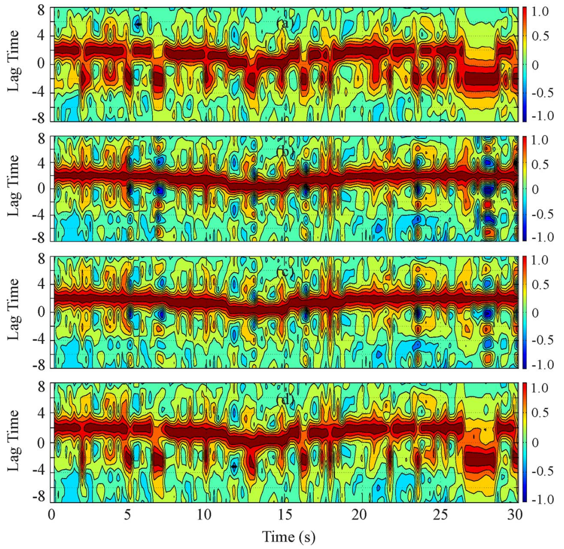

Fig. 10. The contour of the cross-correlation function between the left and right channels of (a) the noisy signal, (b) the enhanced signal by the WL Wiener filter, (c) the enhanced signal by the MVDR filter, and (d) the enhanced signal by the LCMV filter. The signal-to-white-noise ratio is 20 dB, and the signal-to-car-noise ratio is 10 dB. For the LCMV filter, α is set to 0.6.

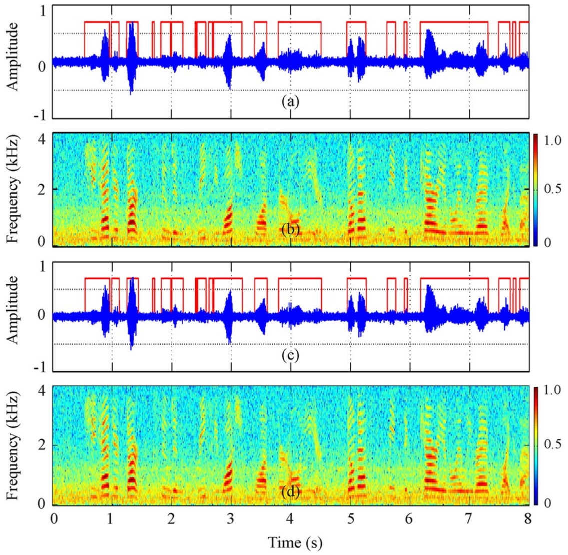

Fig. 11. Noisy signals received at the two microphones: (a) waveform of the left-channel signal, (b) spectrogram of the left-channel signal, (c) waveform of the right-channel signal, (b) spectrogram of the right-channel signal. A short-time energy based VAD is applied to detect the presence and absence of speech and the detection results are illustrated in (a) and (c).

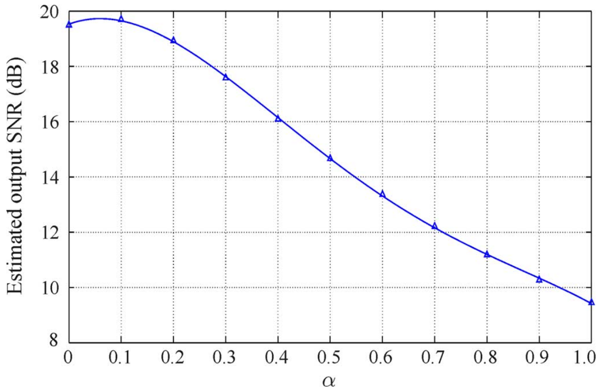

Fig. 12. The estimated output SNR of LCMV filter as a function of the parameter α. The estimated input SNR is approximately 9.1 dB.

noise. In our experimental setup, there is both correlated and diffuse noise. As the signal-to-white-noise ratio increases from 5 to 30 dB, the directional car noise becomes more dominant, and therefore the SNR improvement of the MVDR filter increases. In comparison, the LCMV filter with $\alpha = 0.6$ needs to preserve a large portion of the characteristics and spatial information of the noise. As a result, the amount of SNR gain by the LCMV filter is less dependent on the input SNR.

D. Experiment With Signals Recorded in a Real Car Environment

In this experiment, we assess the LCMV filter for its performance in a real car noise environment. Two omnidirectional microphones are mounted on the dashboard of a sedan car (in front of the front passenger seat). The spacing between the two microphones is 20 centimeters. The car is parked in a parking lot with engine on and all windows closed. A male talker seats in the front passenger seat and his voice is recorded with the two microphones. The length of the recording is approximately 35 seconds. Fig. 11 displays the first 8-second signals received by the two microphones and their spectrograms.

In the previous experiments, we directly computed the signal statistics from the corresponding signals. However, in this experiment, both the clean and noise signals are not accessible. In order to estimate the noise correlation matrix $\mathbf{R}_{\tilde{\mathbf{v}}}$, we first implemented a short-time energy based VAD. The VAD results are also shown in Fig. 11. An estimate of the $\mathbf{R}_{\tilde{\mathbf{v}}}$ matrix is computed using a short-time average with the most recent 40-ms noise samples during the absence of speech. In the presence of speech, an estimate of the clean speech correlation matrix is computed by subtracting the most recent estimate of the noise correlation matrix from that of the noisy one (the estimate of the noise correlation matrix is not updated during the presence of speech till the next silence period). The rest parameters that are needed to implement the LCMV filter can then be obtained straightforwardly. Note that in our experimental setup, the noise is stationary; so one can assume that the noise statistics during the presence of speech remain the same as the immediately previous period where the speech is absent. If the noise is nonstationary, we have to find a way to estimate the noise correlation matrix in the presence of speech. This estimation issue in stereo systems is very challenging and we will leave it for our future study.

Since both the clean and noise signals are not accessible, we cannot compute the performance metrics as we did previously. Instead, we compute the input SNR based on the VAD results. Specifically, we compute the variance of the noise signal, i.e., σ_v^2, during the absence of speech using a long-time average and the variance of the noisy signal, i.e., σ_y^2, during the presence of speech also with a long-time average. An estimate of variance of the clean speech is then obtained by subtracting the noisy variance from the noise one. The estimated clean and noise varainces are substituted into (53) to estimate the input SNR. With this method, the estimated input SNR for our recorded signals is approximately 9.1 dB. Due to the unavailability of both the clean and noise signals, one cannot compute the output SNR according to (54) after applying the WL LCMV filter, $\check{\mathbf{h}}_{\mathrm{LCMV},\alpha}$ to the noisy signal. Instead, we estimate the output SNR from enhanced signal using VAD in a similar way as we estimate

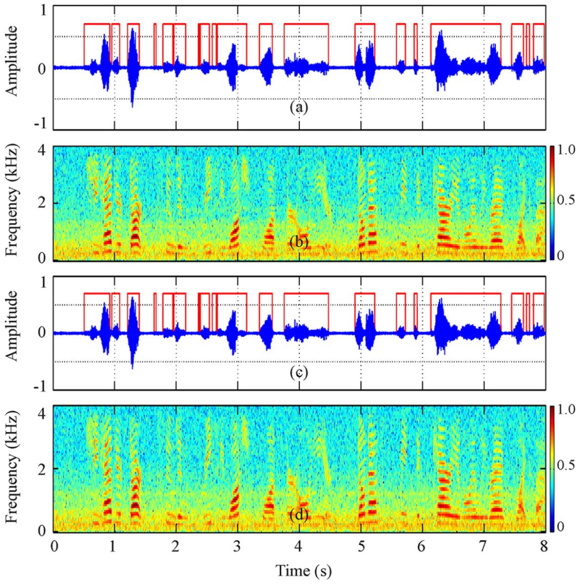

Fig. 13. Enhanced signals for the LCMV filter with $\alpha = 0.2$: (a) the enhanced signal for the left channel, (b) spectrogram of the left-channel enhanced signal, (c) the enhanced signal for the right channel, (b) spectrogram of the right-channel enhanced signal.

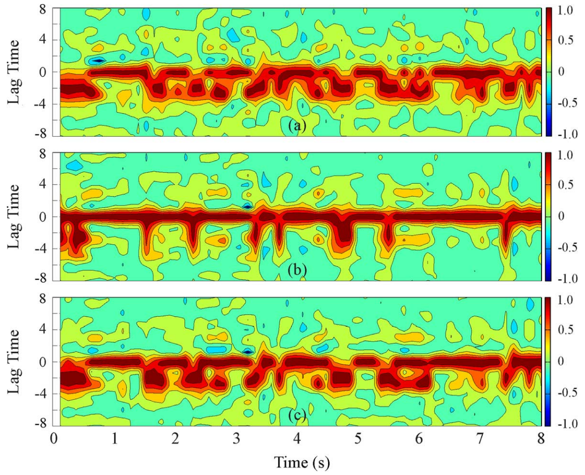

Fig. 14. The contour of the cross-correlation function between the left and right channels of (a) the noisy signal, (b) the enhanced signals by the LCMV filter with $\alpha = 0.2$, and (c) the enhanced signals by the LCMV filter with $\alpha = 0.8$.

the input SNR. The results are plotted in Fig. 12. It is seen that the LCMV filter can significantly improve the SNR. Comparing Figs. 12 and 5, one may notice that a larger SNR improvement is achieved in this experiment with real recorded signals. The underlying reasons for this are multiples. The major one is due to the difference in computing the output SNR. Specifically, in Fig. 5, the residual interference is considered a part of noise; but it is treated as part of the speech in Fig. 12 since we cannot estimate the interference component in real applications where the clean speech is not accessible. For the reader who is interested in difference between the traditional output SNR and the SNR defined in (54), please see [38].

Fig. 13 plots the enhanced signals for both the left and right channels. Comparing this figure with Fig. 11, one can see that the enhanced binaural signals are much less noisy than that of the stereo input signals.

Fig. 14 plots the contours of the cross-correlation functions between the left and right channels of the noisy signals and the enhanced signals by the LCMV filter with $\alpha = 0.2$ and $\alpha = 0.8$ respectively. It is clearly seen that the noise spatial information is better preserved with a larger value of α. Of course, with a larger value of α, there will be less noise reduction.

VII. Conclusion

This paper focused on the noise reduction problem in stereo systems that have two inputs and two outputs. By merging the two real input signals into one complex signal and also combing the two expected real output signals into a complex signal, we formulated the problem into a single-channel WL filtering framework. We then discussed a way to decompose both the complex clean speech and noise signal vectors into two orthogonal components: one correlated and the other uncorrelated with the respective signal samples at the current time instant. With this decomposition, we deduced a WL LCMV filter. Depending on how the constraint in the LCMV filter is chosen, we can make this filter either completely remove the correlated noise component or reduce part of the noise while preserving the noise characteristics and spatial information as well.

References

[1] J. E. Greenberg and P. M. Zurek, “Evaluation of an adaptive beamforming method for hearing aids,” *J. Acoustic. Soc. Amer.*, vol. 93, pp. 1662–1676, 1992.

[2] M. Kompis and N. Dillier, “Noise reduction for hearing aids: Combining directional microphones with an adpative beamformer,” *J. Acoustic. Soc. Amer.*, vol. 96, pp. 1910–1913, 1994.

[3] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, “Microphone-array hearing aids with binaural output—Part I: Fixed-processing systems,” *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

[4] D. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek, “Microphone-array hearing aids with binaural output—Part II: Two-microphone adaptive system,” *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 543–551, Nov. 1997.

[5] J. V. Berghe and J. Wouters, “An adaptive noise canceller for hearing aids using two nearby microphones,” *J. Acoust. Soc. Amer.*, vol. 103, pp. 3621–3626, 1998.

[6] Y. Suzuki, S. Tsukui, F. Asano, R. Nishimura, and T. Sone, “New design method of a binaural microphone array using multiple constraints,” *IEICE Trans. Fundamentals*, vol. E82-A, pp. 588–596, 1999.

[7] T. Lotter, B. Sauert, and P. Vary, “A stereo input-output superdrective beamformer for dual channel noise reduction,” in *Proc. EUROSPEECH*, 2005, pp. 2285–2288.

[8] K. U. Simmer, J. Bitzer, and C. Marro, , M. S. Brandstein and D. B. Ward, Eds., “Post-filtering techniques,” in *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer, 2001, ch. 3, pp. 39–60.

[9] T. Lotter and P. Vary, “Dual-channel speech enhancement by superdirective beamforming,” in *EURASIP J. Appl. Signal Process.*, Jan. 2006, vol. 2006, pp. 175–175.

[10] B. Kollmeier, J. Peissig, and V. Hohmann, “Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain,” *Scand. Audiol. Suppl.*, vol. 38, pp. 28–38, 1993.

[11] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[12] P. Vary, “Noise suppression by spectral magnitude estimation—Mechanism and theoretical limits,” *Signal Process.*, vol. 8, pp. 387–400, Jul. 1985.

[13] W. Etter and G. S. Moschytz, “Noise reduction by noise-adaptive spectral magnitude expansion,” *J. Audio Eng. Soc.*, vol. 42, pp. 341–349, May 1994.

[14] J. Chen, J. Benesty, Y. Huang, and E. J. Diethorn, “Fundamentals of noise reduction,” in *Springer Handbook on Speech Processing and Speech Communication*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2007.

[15] H. Nakashima, Y. Chisaki, T. Usagawa, and M. Ebata, “Frequency domain binaural model based on interaural phase and level differences,” *Acoustic. Sci. Tech.*, vol. 24, pp. 172–178, 2003.

[16] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki, “Two-stage binaural speech enhancement with Wiener filter for high-quality speech communication,” *Speech Commun.*, vol. 53, no. 5, pp. 677–689, 2011.

[17] 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.

[18] S. Doclo, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel Wiener filter with ITD cues for noise reduction in binaural hearing aids," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2005, pp. 70–73.

[19] J. Chen, J. Benesty, and Y. Huang, "A minimum distortion noise reduction algorithm with multiple microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 481–493, Mar. 2008.

[20] J. Benesty, J. Chen, and Y. Huang, "Binaural noise reduction in the time domain with a stereo setup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 2260–2272, Nov. 2011.

[21] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.

[22] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL, USA: CRC, 2007.

[23] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Chichester, U.K.: Wiley, 2006.

[24] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008.

[25] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York, NY, USA: Wiley, 2009.

[26] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for complex variables and signals—Part I: Variables," *Elsevier Signal Process.*, vol. 53, pp. 1–13, 1996.

[27] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for complex variables and signals—Part II: Signals," *Elsevier Signal Process.*, vol. 53, pp. 15–25, 1996.

[28] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[29] J. Benesty, J. Chen, Y. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Berlin, Germany: Springer-Verlag, 2005, ch. 2, pp. 9–41.

[30] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 1218–1234, Jul. 2006.

[31] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[32] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, Aug. 1971.

[33] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 1, pp. 926–935, Jan. 1972.

[34] M. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 6, pp. 1378–1393, Dec. 1983.

[35] A. Härmä, "Acoustic measurement data from the varechoic chamber," Technical Memorandum, Agere Systems, Nov. 2001.

[36] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, "The new Varechoic chamber at AT&T Bell Labs," in *Proc. Wallance Clement Sabine Centennial Symp.*, 1994.

[37] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 4, pp. 1141–1154, Apr. 2011.

[38] J. Benesty, J. Chen, Y. Huang, and T. Gaensler, "Time-domain noise reduction based on an orthogonal decomposition for desired signal extraction," *J. Acoust. Soc. Amer.*, vol. 132, pp. 452–446, Jul. 2012.

Jingdong Chen (M'99–SM'09) received the Ph.D. degree in pattern recognition and intelligence control from the Chinese Academy of Sciences in 1998.

From 1998 to 1999, he was with ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan, where he conducted research on speech synthesis, speech analysis, as well as objective measurements for evaluating speech synthesis. He then joined the Griffith University, Brisbane, Australia, where he engaged in research on robust speech recognition and signal processing. From 2000 to 2001, he worked at ATR Spoken Language Translation Research Laboratories on robust speech recognition and speech enhancement. From 2001 to 2009, he was a Member of Technical Staff at Bell Laboratories, Murray Hill, New Jersey, working on acoustic signal processing for telecommunications. He subsequently joined WeVoice Inc. in New Jersey, serving as the Chief Scientist. He is currently a professor at the Northwestern Polytechnical University in Xi'an, China. His research interests include acoustic signal processing, adaptive signal processing, speech enhancement, adaptive noise/echo control, microphone array signal processing, signal separation, and speech communication. Dr. Chen is currently an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, a member of the IEEE Audio and Electroacoustics Technical Committee, and a member of the editorial advisory board of the Open Signal Processing Journal. He was the Technical Program Co-Chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) and the Technical Program Chair of IEEE TENCON 2013, and helped organize many other conferences. He co-authored the books *Study and Design of Differential Microphone Arrays* (Springer-Verlag, 2013), *Speech Enhancement in the STFT Domain* (Springer-Verlag, 2011), *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study* (Springer-Verlag, 2011), *Speech Enhancement in the Karhunen-Loève Expansion Domain* (Morgan&Claypool, 2011), *Noise Reduction in Speech Processing* (Springer-Verlag, 2009), *Microphone Array Signal Processing* (Springer-Verlag, 2008), and *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006). He is also a co-editor/co-author of the book *Speech Enhancement* (Berlin, Germany: Springer-Verlag, 2005) and a section co-editor of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, Berlin, 2007).

Dr. Chen received the 2008 Best Paper Award from the IEEE Signal Processing Society (with Benesty, Huang, and Doclo), the best paper award from the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2011 (with Benesty), the Bell Labs Role Model Teamwork Award twice, respectively, in 2009 and 2007, the NASA Tech Brief Award twice, respectively, in 2010 and 2009, the Japan Trust International Research Grant from the Japan Key Technology Center in 1998, the Young Author Best Paper Award from the 5th National Conference on Man-Machine Speech Communications in 1998, and the CAS (Chinese Academy of Sciences) President's Award in 1998.

Jacob Benesty was born in 1963. He received a Master degree in microwaves from Pierre & Marie Curie University, France, in 1987, and a Ph.D. degree in control and signal processing from Orsay University, France, in April 1991.

During his Ph.D. (from Nov. 1989 to Apr. 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecomunications (CNET), Paris, France. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ, USA. In May 2003, he joined the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, as a Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He is the inventor of many important technologies. In particular, he was the lead researcher at Bell Labs who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. Also, he conceived and designed the world-first PC-based multi-party hands-free full-duplex stereo conferencing system over IP networks.

He was the co-chair of the 1999 International Workshop on Acoustic Echo and Noise Control and the general co-chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He is the recipient, with Morgan and Sondhi, of the IEEE Signal Processing Society 2001 Best Paper Award. He is the recipient, with Chen, Huang, and Doclo, of the IEEE Signal Processing Society 2008 Best Paper Award. He is also the co-author of a paper for which Huang received the IEEE Signal Processing Society 2002 Young Author Best Paper Award. In 2010, he received the "Gheorghe Cartianu Award" from the Romanian Academy. In 2011, he received the Best Paper Award from the IEEE WASPAA for a paper that he co-authored with Chen.