A TIME-DOMAIN WIDELY LINEAR MVDR FILTER FOR BINAURAL NOISE REDUCTION

Jingdong Chen

Northwestern Polytechnical University
127 Youyi West Rd,
Xi'an, Shaanxi 710072, China
jingdongchen@ieee.org

Jacob Benesty

INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada
benesty@emt.inrs.ca

ABSTRACT

This paper deals with the problem of binaural noise reduction in the time domain with a stereophonic sound system. We first form a complex signal from the stereo inputs with one channel being its real part and the other being its imaginary part. By doing so, the binaural noise reduction problem is converted to a single-channel noise reduction problem via the widely linear (WL) model. The WL estimation theory is then used to derive the minimum variance distortionless response (MVDR) noise reduction filter that can fully take advantage of the noncircularity of the complex speech signal to achieve noise reduction while preserving the desired signal (speech) and spatial information. Experiments are provided to justify the effectiveness of this MVDR filter.

Index Terms— Binaural noise reduction, stereo sound system, noncircularity, widely linear (WL) estimation, minimum variance distortionless response (MVDR) filter, time domain.

1. 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 $x_{\mathrm{R}}(k)$, $x_{\mathrm{L}}(k)$, $v_{\mathrm{R}}(k)$, and $v_{\mathrm{L}}(k)$ are zero mean, and $x_{\mathrm{R}}(k)$ and $x_{\mathrm{L}}(k)$ are uncorrelated with $v_{\mathrm{R}}(k)$ and $v_{\mathrm{L}}(k)$. The two noise signals $v_{\mathrm{R}}(k)$ and $v_{\mathrm{L}}(k)$ 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 $x_{\mathrm{R}}(k)$ and $x_{\mathrm{L}}(k)$ given the observations $y_{\mathrm{R}}(k)$ and $y_{\mathrm{L}}(k)$. It is clear that our objective is to attenuate the contribution of the noise terms $v_{\mathrm{R}}(k)$ and $v_{\mathrm{L}}(k)$ as much as possible, and meanwhile preserve $x_{\mathrm{R}}(k)$ and $x_{\mathrm{L}}(k)$ 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 $s(k)$. We have stereo signals in model (1); but we believe that it is more convenient to work in the complex domain in order that the original (stereo) problem is transformed to a single-channel noise reduction processing. Indeed, from the two real microphone signals given in (1a) and (1b), we can form the complex microphone signal as

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

where $j = \sqrt{-1}$, $x(k) = x_{\mathrm{R}}(k) + jx_{\mathrm{L}}(k)$ is the complex desired signal, and $v(k) = v_{\mathrm{R}}(k) + jv_{\mathrm{L}}(k)$ 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, the classical linear techniques [5], [6] 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 widely linear (WL) estimation theory, which will be discussed in the following section.

2. WIDELY LINEAR MODEL

As can be noticed from the 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) [1], [2]. 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^* [3], [4], 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 [1] is equal to zero, i.e., $E\left(z^2\right) = 0$, while its variance is nonnull, i.e., $E\left(|z|^2\right) \neq 0$. This means that the second-order behavior of a CCRV is well described by its variance. If the pseudo-variance $E\left(z^2\right)$ is not equal to 0, the CRV z is then noncircular. A good measure of the second-order circularity is the circularity quotient [1] defined as the ratio between the pseudo-variance and the variance, i.e.,

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

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, and a larger value of $|\gamma_z|$ indicates that the CRV z is more noncircular.

Now, let us examine whether the complex desired signal $x(k)$ in (2) is second-order circular or not. It is easy to check that

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

978-1-4577-0693-6/11/$26.00 ©2011 IEEE

where $\sigma_x^2 = E\left[|x(k)|^2\right]$ is the variance of $x(k)$. One can check from (4) that the CRV $x(k)$ is second-order circular (i.e., $\gamma_x = 0$) if and only if

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

Since the signals $x_{\mathrm{R}}(k)$ and $x_{\mathrm{L}}(k)$ 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, and so is the complex microphone signal, $y(k)$.

Since we deal with noncircular CRVs as demonstrated above, the classical linear estimation technique [5], [6], which is developed for processing real signals or CCRVs, cannot be applied to recover $x(k)$. Instead, an estimate of $x(k)$ should be obtained using the WL estimation theory as [7], [2]

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

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

$$\tilde{\mathbf{h}} \triangleq \begin{bmatrix} \mathbf{h} \\ \mathbf{h}' \end{bmatrix}, \quad \tilde{\mathbf{y}}(k) \triangleq \begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}^*(k) \end{bmatrix}, \tag{8}$$

are the augmented WL filter and observation vector, respectively, both of length $2L$. We can rewrite (6) as

$$\hat{x}(k) = \tilde{\mathbf{h}}^H\left[\tilde{\mathbf{x}}(k) + \tilde{\mathbf{v}}(k)\right] = x_{\mathrm{f}}(k) + v_{\mathrm{rn}}(k), \tag{9}$$

where $\tilde{\mathbf{x}}(k)$ and $\tilde{\mathbf{v}}(k)$ are defined in a similar way to $\tilde{\mathbf{y}}(k)$, $x_{\mathrm{f}}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{x}}(k)$ is a filtered version of the desired signal and its conjugate of L successive time samples, and $v_{\mathrm{rn}}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{v}}(k)$ is the residual noise. From (9), we see that $\hat{x}(k)$ depends on the vector $\tilde{\mathbf{x}}(k)$; but not all the components in $\tilde{\mathbf{x}}(k)$ contribute to the estimation of the desired signal sample $x(k)$. To see this clearly, let us decompose the vector $\tilde{\mathbf{x}}(k)$ into two orthogonal vectors: one corresponding to the desired signal at time k and the other corresponding to the interference. Let us first decompose $\mathbf{x}(k)$ as

$$\mathbf{x}(k) = x(k)\boldsymbol{\rho}_x + \mathbf{x}'(k), \tag{10}$$

where

$$\boldsymbol{\rho}_x \triangleq \frac{E\left[\mathbf{x}(k)x^*(k)\right]}{\sigma_x^2} = \left[\rho_{x,0}\quad \rho_{x,1}\quad \cdots\quad \rho_{x,L-1}\right]^T \tag{11}$$

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

$$\rho_{x,l} \triangleq \frac{E\left[x(k-l)x^*(k)\right]}{\sigma_x^2} \tag{12}$$

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

$$\mathbf{x}'(k) = \mathbf{x}(k) - x(k)\boldsymbol{\rho}_x \tag{13}$$

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

$$E\left[\mathbf{x}'(k)x^*(k)\right] = \mathbf{0}, \tag{14}$$

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

Similarly, we have

$$\mathbf{x}^*(k) = x(k)\boldsymbol{\gamma}_x^* + \mathbf{x}''(k), \tag{15}$$

where

$$\boldsymbol{\gamma}_x \triangleq \frac{E\left[\mathbf{x}(k)x(k)\right]}{\sigma_x^2} = \left[\gamma_{x,0}\quad \gamma_{x,1}\quad \cdots\quad \gamma_{x,L-1}\right]^T \tag{16}$$

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

$$\gamma_{x,l} \triangleq \frac{E\left[x(k-l)x(k)\right]}{\sigma_x^2} \tag{17}$$

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

$$\mathbf{x}''(k) = \mathbf{x}^*(k) - x(k)\boldsymbol{\gamma}_x^* \tag{18}$$

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

$$E\left[\mathbf{x}''(k)x^*(k)\right] = \mathbf{0}. \tag{19}$$

Combining (10) and (15), we get

$$\tilde{\mathbf{x}}(k) = x(k)\mathbf{d}_x + \tilde{\mathbf{x}}'(k) = \tilde{\mathbf{x}}_{\mathrm{d}}(k) + \tilde{\mathbf{x}}'(k), \tag{20}$$

where

$$\mathbf{d}_x \triangleq \begin{bmatrix} \boldsymbol{\rho}_x \\ \boldsymbol{\gamma}_x^* \end{bmatrix}, \quad \tilde{\mathbf{x}}'(k) \triangleq \begin{bmatrix} \mathbf{x}'(k) \\ \mathbf{x}''(k) \end{bmatrix}, \tag{21}$$

$\tilde{\mathbf{x}}_{\mathrm{d}}(k) \triangleq x(k)\mathbf{d}_x$ is correlated with the desired signal, $x(k)$, and will contribute to its estimation, so we call it the desired signal vector. In comparison, $\tilde{\mathbf{x}}'(k)$ is uncorrelated with $x(k)$, and will interfere with the estimation, so we call it the interference signal vector.

Substituting (20) into (9), we obtain

$$\begin{aligned} \hat{x}(k) &= \tilde{\mathbf{h}}^H\left[\tilde{\mathbf{x}}_{\mathrm{d}}(k) + \tilde{\mathbf{x}}'(k) + \tilde{\mathbf{v}}(k)\right] \\ &= \tilde{\mathbf{h}}^H\left[x(k)\mathbf{d}_x + \tilde{\mathbf{x}}'(k) + \tilde{\mathbf{v}}(k)\right] \\ &= x_{\mathrm{fd}}(k) + x'_{\mathrm{ri}}(k) + v_{\mathrm{rn}}(k), \end{aligned} \tag{22}$$

where $x_{\mathrm{fd}}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{x}}_{\mathrm{d}}(k) = x(k)\tilde{\mathbf{h}}^H\mathbf{d}_x$ is the filtered desired signal and $x'_{\mathrm{ri}}(k) \triangleq \tilde{\mathbf{h}}^H\tilde{\mathbf{x}}'(k)$ is the residual interference. We observe that the estimate of the desired signal at time k is the sum of three terms that are mutually uncorrelated. Therefore, the variance of $\hat{x}(k)$ is

$$\sigma_{\hat{x}}^2 = \sigma_{x_{\mathrm{fd}}}^2 + \sigma_{x'_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2, \tag{23}$$

where

$$\sigma_{x_{\mathrm{fd}}}^2 = \sigma_x^2\left|\tilde{\mathbf{h}}^H\mathbf{d}_x\right|^2 = \tilde{\mathbf{h}}^H\mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}}\tilde{\mathbf{h}}, \tag{24}$$

$$\sigma_{x'_{\mathrm{ri}}}^2 = \mathbf{h}^H\mathbf{R}_{\tilde{\mathbf{x}}'}\tilde{\mathbf{h}} = \tilde{\mathbf{h}}^H\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{h}} - \sigma_x^2\left|\tilde{\mathbf{h}}^H\mathbf{d}_x\right|^2, \tag{25}$$

$$\sigma_{v_{\mathrm{rn}}}^2 = \tilde{\mathbf{h}}^H\mathbf{R}_{\tilde{\mathbf{v}}}\tilde{\mathbf{h}}, \tag{26}$$

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

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

It is clear from (22) that the objective of our noise reduction problem is to find optimal filters that can minimize the effect of $x'_{\mathrm{ri}}(k) + v_{\mathrm{rn}}(k)$ while preserving the desired signal, $x(k)$.

3. TIME-DOMAIN WL MVDR FILTER

To derive the WL MVDR filter, we need to derive first the mean-square error (MSE) criterion.

We define the error signal between the estimated and desired signals as

$$e(k) \triangleq \hat{x}(k) - x(k) = \tilde{\mathbf{h}}^H \tilde{\mathbf{y}}(k) - x(k), \tag{27}$$

which can be written as the sum of two uncorrelated error signals:

$$e(k) = e_{\mathrm{d}}(k) + e_{\mathrm{r}}(k), \tag{28}$$

where

$$e_{\mathrm{d}}(k) \triangleq x_{\mathrm{fd}}(k) - x(k) \tag{29}$$

is the signal distortion due to the WL filter and

$$e_{\mathrm{r}}(k) \triangleq x'_{\mathrm{ri}}(k) + v_{\mathrm{rn}}(k) \tag{30}$$

represents the residual interference-plus-noise.

The MSE is then

$$J\left(\tilde{\mathbf{h}}\right) \triangleq E\left[|e(k)|^2\right] = \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{y}}} \tilde{\mathbf{h}} = J_{\mathrm{d}}\left(\tilde{\mathbf{h}}\right) + J_{\mathrm{r}}\left(\tilde{\mathbf{h}}\right), \tag{31}$$

where

$$J_{\mathrm{d}}\left(\tilde{\mathbf{h}}\right) \triangleq E\left[|e_{\mathrm{d}}(k)|^2\right] = \sigma_x^2 \left|\tilde{\mathbf{h}}^H \mathbf{d}_x - 1\right|^2, \tag{32}$$

$$J_{\mathrm{r}}\left(\tilde{\mathbf{h}}\right) \triangleq E\left[|e_{\mathrm{r}}(k)|^2\right] = \sigma_{x'_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2 = \tilde{\mathbf{h}}^H \mathbf{R}_{\mathrm{in}} \tilde{\mathbf{h}}, \tag{33}$$

and

$$\mathbf{R}_{\mathrm{in}} = \mathbf{R}_{\tilde{\mathbf{x}}'} + \mathbf{R}_{\tilde{\mathbf{v}}} \tag{34}$$

is the interference-plus-noise covariance matrix.

With the previously defined MSEs, it is clear that the MVDR filter can be derived by minimizing $J\left(\tilde{\mathbf{h}}\right)$ subject to the constraint $J_{\mathrm{d}}\left(\tilde{\mathbf{h}}\right) = 0$. Mathematically, this can be transformed into the following optimization problem

$$\min_{\tilde{\mathbf{h}}} \tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{y}}} \tilde{\mathbf{h}} \quad \text{subject to} \quad \tilde{\mathbf{h}}^H \mathbf{d}_x = 1. \tag{35}$$

Using a Lagrange multiplier to adjoin the constraint to the cost function and then taking the gradient with respect to $\tilde{\mathbf{h}}$ and equating the result to zero, we obtain

$$\tilde{\mathbf{h}}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_x}{\mathbf{d}_x^H \mathbf{R}_{\tilde{\mathbf{y}}}^{-1} \mathbf{d}_x}. \tag{36}$$

Obviously, the MVDR filter can also be deduced by minimizing $J_{\mathrm{r}}\left(\tilde{\mathbf{h}}\right)$ subject to the constraint $\tilde{\mathbf{h}}^H \mathbf{d}_x = 1$. This time, the problem can be written into the following form

$$\min_{\tilde{\mathbf{h}}} \tilde{\mathbf{h}}^H \mathbf{R}_{\mathrm{in}} \tilde{\mathbf{h}} \quad \text{subject to} \quad \tilde{\mathbf{h}}^H \mathbf{d}_x = 1, \tag{37}$$

for which the solution is

$$\tilde{\mathbf{h}}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{d}_x}{\mathbf{d}_x^H \mathbf{R}_{\mathrm{in}}^{-1} \mathbf{d}_x}. \tag{38}$$

With the use of the following relation

$$\mathbf{R}_{\tilde{\mathbf{y}}} = \sigma_x^2 \mathbf{d}_x \mathbf{d}_x^H + \mathbf{R}_{\mathrm{in}}, \tag{39}$$

we can rewrite (38) as

$$\tilde{\mathbf{h}}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\tilde{\mathbf{y}}} - \mathbf{I}_{2L}}{\mathrm{tr}\left(\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\tilde{\mathbf{y}}}\right) - 2L} \mathbf{i}_1, \tag{40}$$

where $\mathbf{i}_1$ is the first column of the identity matrix $\mathbf{I}_{2L}$ of size $2L \times 2L$. It is easy to check that the three forms of the MVDR filter in (36), (38), and (40) are theoretically identical.

4. EVALUATION OF THE WL MVDR FILTER

4.1. Theoretical Evaluation

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 MVDR 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 signal-to-noise ratios (SNRs) of the MVDR filter. The input SNR is defined as

$$\mathrm{iSNR} \triangleq \frac{\sigma_x^2}{\sigma_v^2}, \tag{41}$$

where $\sigma_v^2 \triangleq E\left[|v(k)|^2\right]$ is the variance of the complex additive noise. After applying a WL filter $\tilde{\mathbf{h}}$, 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[2], i.e.,

$$\mathrm{oSNR}\left(\tilde{\mathbf{h}}\right) \triangleq \frac{\sigma_{x_{\mathrm{fd}}}^2}{\sigma_{x'_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2} = \frac{\sigma_x^2 \left|\tilde{\mathbf{h}}^H \mathbf{d}_x\right|^2}{\tilde{\mathbf{h}}^H \mathbf{R}_{\mathrm{in}} \tilde{\mathbf{h}}} = \frac{\tilde{\mathbf{h}}^H \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}} \tilde{\mathbf{h}}}{\tilde{\mathbf{h}}^H \mathbf{R}_{\mathrm{in}} \tilde{\mathbf{h}}}. \tag{42}$$

The objective of the noise reduction filter is to make the output SNR greater than the input SNR so that the quality of the noisy signal will be enhanced. Now, let us introduce the quantity $\mathrm{oSNR}_{\max}$, which is defined as the maximum output SNR that can be achieved through filtering so that

$$\mathrm{oSNR}\left(\tilde{\mathbf{h}}\right) \leq \mathrm{oSNR}_{\max}, \ \forall \tilde{\mathbf{h}}. \tag{43}$$

It can be checked from (42) that this quantity is equal to the maximum eigenvalue of the matrix $\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}}$, i.e.,

$$\mathrm{oSNR}_{\max} = \lambda_{\max}\left(\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}}\right). \tag{44}$$

The filter that can achieve $\mathrm{oSNR}_{\max}$ is called the maximum SNR filter and is denoted by $\tilde{\mathbf{h}}_{\max}$. It is easy to see from (44) that $\tilde{\mathbf{h}}_{\max}$ is the eigenvector corresponding to the maximum eigenvalue of $\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}}$. Clearly, we have

$$\mathrm{oSNR}_{\max} = \mathrm{oSNR}\left(\tilde{\mathbf{h}}_{\max}\right) \geq \mathrm{oSNR}\left(\mathbf{i}_1\right) = \mathrm{iSNR}. \tag{45}$$

Since the rank of the matrix $\mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}}$ is equal to 1, we also have

$$\mathrm{oSNR}_{\max} = \mathrm{tr}\left(\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{d}}}\right) = \sigma_x^2 \mathbf{d}_x^H \mathbf{R}_{\mathrm{in}}^{-1} \mathbf{d}_x, \tag{46}$$

[2] In this paper, we consider the interference as part of the noise in the definitions of the performance measures.

where tr$(\cdot)$ denotes the trace of a square matrix.

Substituting (38) into (42), we can deduce that

$$\text{oSNR}\left(\tilde{\mathbf{h}}_{\text{MVDR}}\right) = \text{oSNR}_{\max}. \qquad (47)$$

It is of great interest to observe that the MVDR filter maximizes the output SNR without introducing speech distortion. Note that the $\tilde{\mathbf{h}}_{\max}$ filter is different from $\tilde{\mathbf{h}}_{\text{MVDR}}$ by a scaling factor although both can maximize the output SNR. This scaling factor is time-varying and can cause discontinuity in both the speech and the residual noise level, which is eventually transformed into speech distortion. Therefore, it is recommended to use the MVDR filter rather than the maximum SNR one in practice.

4.2. Experimental Evaluation

Experiments were conducted with the clean speech recorded in a reverberant but quiet room. The room is 6 m long and 5 m wide. For ease of exposition, positions in the room are designated by (x, y) coordinates with reference to one corner of the room with $0 \leq x \leq 6$ and $0 \leq y \leq 5$. A stereo recording system is configured where two same omnidirectional microphones are placed respectively at (3.0, 0.5) and (3.1, 0.5). A female talker reads a story while walking along the line $y = 3$ m and her voice is recorded with a sampling rate of 8 kHz. The recorded stereo signal is treated as the clean speech. White Gaussian noise is then added into the signal so that the input SNR is equal to 10 dB. The overall length of the signal is 30 s. We set the filter length to $L = 40$.

To implement the MVDR filter given in (36), we need to know the correlation matrix $\mathbf{R}_{\tilde{\mathbf{y}}}$ and the correlation vector $\mathbf{d}_x$. In this paper, we compute the $\mathbf{R}_{\tilde{\mathbf{y}}}$ matrix from the noisy signal using a short-time average. Specifically, at each time instant k, an estimate of the matrix $\mathbf{R}_{\tilde{\mathbf{y}}}$, i.e., $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}}(k)$, is computed using the most recent 640 samples (40-ms long) of the noisy signal $y(k)$. To obtain an estimate of the $\mathbf{d}_x$ vector, we first compute the $\mathbf{R}_{\tilde{\mathbf{v}}}$ matrix directly from the noise signal (without using any VAD) with a short-time average using the most recent 1280 samples (80-ms long). Subtracting the computed $\hat{\mathbf{R}}_{\tilde{\mathbf{v}}}(k)$ matrix from $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}}(k)$, we obtain an estimate of the $\mathbf{R}_{\tilde{\mathbf{x}}}$ matrix at time k, i.e., $\hat{\mathbf{R}}_{\tilde{\mathbf{x}}}(k)$. Then the estimate of the $\mathbf{d}_x$ vector, i.e., $\hat{\mathbf{d}}_x(k)$, is the first column of $\hat{\mathbf{R}}_{\tilde{\mathbf{x}}}(k)$ normalized by its first element. Substituting $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}}(k)$ and $\hat{\mathbf{d}}_x(k)$ into (36), we obtain the MVDR filter. With this filter and using (22), we get the three signals $x_{\text{fd}}(k)$, $x'_{\text{ri}}(k)$, and $v_{\text{rn}}(k)$. We then computed the output SNR using a long-time average. The output SNR for this experiment is equal to 14.5 dB. In other words, the SNR improvement is 4.5 dB.

To illustrate the results, we plot, in Fig. 1, the spectrograms of the clean, noisy, and enhanced speech signals for the left channel (we do not show those for the right channel due to space limit). It is clearly seen that the enhanced signal is much cleaner than the noisy one.

To visualize the spatial sound effect, we computed the cross-correlation function between the two channels every 64 ms using a short-time average with a frame size of 64 ms. The location of the maximum value of this function indicates the position of the talker at time instant k. Figure 2 shows the contours of the time-varying cross-correlation function of the clean, noisy, and enhanced signals. One can notice that the noise has significantly modified the sound spatial effect. It is clearly seen that the MVDR filter has recovered the spatial effect.

5. CONCLUSIONS

This paper focused on the binaural noise reduction problem in stereo systems that have two inputs and two outputs. By merging the two real input signals into one complex signal, we formulated the problem into a WL filtering framework. Under this new framework, we derived a time-domain MVDR noise reduction filter, which was shown to be able to not only enhance the noisy speech, but also recover the spatial effects of the speech source.

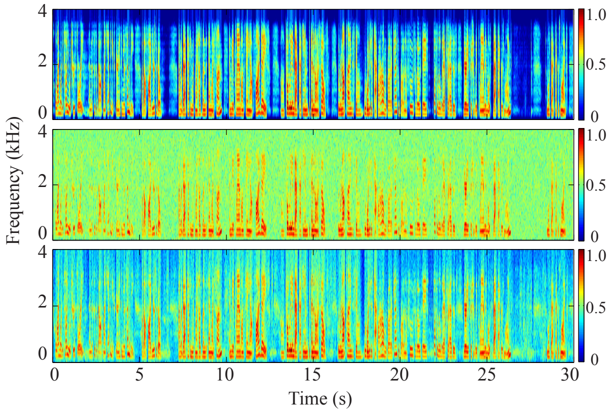

Figure 1: The spectrograms of the left channel. Upper trace: clean speech; middle trace: noisy speech at SNR = 10 dB; lower trace: noise reduced speech.

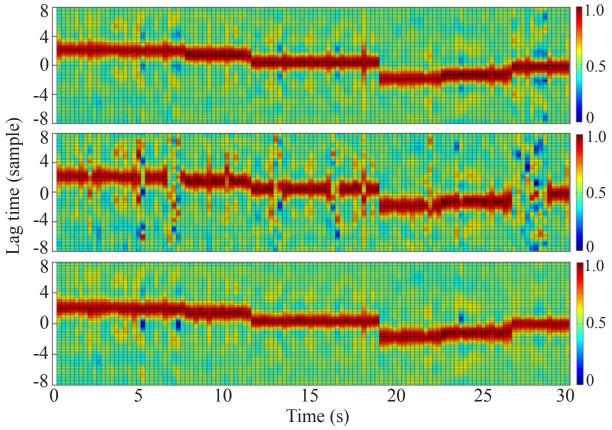

Figure 2: The contours of the short-time cross correlation coefficients between the left and right channels. Upper trace: clean speech; middle trace: noisy speech at SNR = 10 dB; lower trace: noise reduced speech.

6. REFERENCES

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

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

[3] P. O. Amblard, M. Gaeta, and J. L. Lacoume, “Statistics for complex variables and signals–Part I: variables,” *Signal Process.*, vol. 53, pp. 1–13, 1996.

[4] P. O. Amblard, M. Gaeta, and J. L. Lacoume, “Statistics for complex variables and signals–Part II: signals,” *Signal Process.*, vol. 53, pp. 15–25, 1996.

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

[6] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Chichester, England: John Wiley & Sons Ltd, 2006.

[7] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Trans. Signal Process.*, vol. 43, pp. 2030–2033, Aug. 1995.

[8] J. Chen, J. Benesty, Y. Huang, and S. Doclo, “New insights into the noise reduction Wiener filter,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, pp. 1218–1234, July 2006.