EXAMPLES OF OPTIMAL NOISE REDUCTION FILTERS DERIVED FROM THE SQUARED PEARSON CORRELATION COEFFICIENT

Jiaolong Yu[1], Jacob Benesty[2], Gongping Huang[1], and Jingdong Chen[1]

[1]: Northwestern Polytechnical University
127 Youyi West Road
Xi'an, Shaanxi 710072, China

[2]: INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada

ABSTRACT

This paper studies the problem of single-channel noise reduction in the time domain. Based on some orthogonal decomposition developed recently and the squared Pearson correlation coefficient (SPCC), several noise reduction filters are derived. We will show that the optimization of the SPCC leads to the Wiener, minimum variance distortionless response (MVDR), minimum noise (MN), minimum uncorrelated speech and noise (MUSN), and linearly constrained minimum variance (LCMV) filters. We also compare the Wiener and MVDR filters derived from the SPCC to their counterparts derived from the mean-square error (MSE) criterion. Simulations are provided to illustrate the performance of all the deduced noise reduction filters.

Index Terms—Noise reduction, speech enhancement, squared Pearson correlation coefficient (SPCC), optimal filters.

1. INTRODUCTION

Noise is ubiquitous. In all voice and speech related applications, the signal of interest (i.e., speech) that is picked up by a microphone is generally contaminated by noise, which can significantly degrade its quality and intelligibility. To mitigate the effect of the noise in a recorded speech signal, noise reduction (or speech enhancement) with some signal processing techniques are generally needed, which basically attempt to clean up the noisy signal, thereby obtaining an estimate of the desired clean speech. This problem has been intensively studied over the last four decades [1]–[18]; but it remains an unsolved one due to its extreme difficulty.

Typically, noise reduction is achieved by passing the noisy signal through a linear filter. So, the core issue of noise reduction is to find an optimal filter such that, after the filtering process, the signal-to-noise ratio (SNR) is improved, implying that the microphone signal becomes cleaner. However, since the filtering operation does not only attenuate the noise, but also affects the speech signal, careful attention has to be paid to the speech distortion during the design process. Currently, most existing approaches, such as the Wiener filter and the subspace method, are derived from the mean-square error (MSE) criterion. Although these approaches have already achieved a certain degree of success, further efforts are indispensable to find either a better understanding of the traditional methods from different perspectives or new filters that can help better deal with the problem.

Recently, the so-called squared Pearson correlation coefficient (SPCC) was introduced as a cost function for analyzing noise reduction filters [16]. It was shown that several traditional filters that were derived from the MSE criterion can be found with the SPCC. One major advantage of the SPCC as compared to the MSE is that we can have many new insights of why, how, and to what degree the traditional filters work. In this paper, we continue to explore the potential of the SPCC for noise reduction. The major difference, or

novelty, of this paper as compared to the work in [16] is that we now combine the SPCC with a recently introduced orthogonal signal decomposition [17], [18] and propose other ways to derive new filters. Indeed, we show that many optimal noise reduction filters can be derived by either maximizing or minimizing the SPCC, some of which are the same as the traditional ones while others are new. For example, by maximizing the SPCC, we derive the well-known Wiener and minimum variance distortionless response (MVDR) filters that can directly achieve an estimate of the clean speech. On the other hand, by minimizing the SPCC, we can derive many filters such as the minimum noise (MN), the minimum uncorrelated speech and noise (MUSN), and the linearly constrained minimum variance (LCMV) filters that can achieve an estimate of the noise signal and then the desired signal. These filters may have some good potential in dealing with colored noise or noise exhibiting strong self correlation.

2. SIGNAL MODEL AND PROBLEM FORMULATION

The noise reduction problem considered in this paper is one of recovering the desired signal (or clean speech) $x(k)$, k being the discrete-time index, of zero mean from the noisy observation (microphone signal) [1]–[4]:

$$y(k) = x(k) + v(k), \tag{1}$$

where the zero-mean random process $v(k)$ is the unwanted additive noise, which is assumed to be uncorrelated with $x(k)$. All signals are considered to be real and broadband.

The signal model given in (1) can be put into a vector form by considering the L most recent successive time samples, i.e.,

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k), \tag{2}$$

where

$$\mathbf{y}(k) \triangleq \begin{bmatrix} y(k) & y(k-1) & \cdots & y(k-L+1) \end{bmatrix}^T \tag{3}$$

is a vector of length L, superscript ^{T} denotes the transpose of a vector or a matrix, and $\mathbf{x}(k)$ and $\mathbf{v}(k)$ are defined in a similar way to $\mathbf{y}(k)$. We can decompose the speech signal vector into two orthogonal components:

$$\mathbf{x}(k) = \boldsymbol{\rho}_{\mathbf{x}x} x(k) + \mathbf{x}_{\mathrm{u}}(k) = \mathbf{x}_1(k) + \mathbf{x}_{\mathrm{u}}(k), \tag{4}$$

where $\mathbf{x}_1(k) \triangleq \boldsymbol{\rho}_{\mathbf{x}x} x(k)$,

$$\boldsymbol{\rho}_{\mathbf{x}x} \triangleq \frac{E\left[\mathbf{x}(k) x(k)\right]}{E\left[x^2(k)\right]} = \frac{E\left[\mathbf{x}(k) x(k)\right]}{\sigma_x^2} \tag{5}$$

is the normalized correlation vector (of length L) between $\mathbf{x}(k)$ and $x(k)$, with $E[\cdot]$ denoting mathematical expectation, $\sigma_x^2 \triangleq$

$E\left[x^2(k)\right]$ is the variance of $x(k)$, and $\mathbf{x}_{\mathrm{u}}(k)$ is the uncorrelated speech signal vector, i.e., $E\left[\mathbf{x}_{\mathrm{u}}(k)x(k)\right] = \mathbf{0}$. In the same way, we can decompose the noise signal vector as

$$\mathbf{v}(k) = \boldsymbol{\rho}_{\mathbf{v}v}v(k) + \mathbf{v}_{\mathrm{u}}(k) = \mathbf{v}_1(k) + \mathbf{v}_{\mathrm{u}}(k), \qquad (6)$$

where $\mathbf{v}_1(k) \triangleq \boldsymbol{\rho}_{\mathbf{v}v}v(k)$, $\boldsymbol{\rho}_{\mathbf{v}v}$ and $\mathbf{v}_{\mathrm{u}}(k)$ are defined in a similar way to $\boldsymbol{\rho}_{\mathbf{x}x}$ and $\mathbf{x}_{\mathrm{u}}(k)$, respectively. Then, we can rewrite (2) into

$$\mathbf{y}(k) = \mathbf{x}_1(k) + \mathbf{x}_{\mathrm{u}}(k) + \mathbf{v}_1(k) + \mathbf{v}_{\mathrm{u}}(k). \qquad (7)$$

Since the four terms on the right-hand side of (7) are mutually uncorrelated, the correlation matrix of $\mathbf{y}(k)$ is

$$\mathbf{R}_{\mathbf{y}} \triangleq E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right] = \mathbf{R}_{\mathbf{x}_1} + \mathbf{R}_{\mathbf{x}_{\mathrm{u}}} + \mathbf{R}_{\mathbf{v}_1} + \mathbf{R}_{\mathbf{v}_{\mathrm{u}}}, \quad (8)$$

where $\mathbf{R}_{\mathbf{x}_1} = \sigma_x^2 \boldsymbol{\rho}_{\mathbf{x}x}\boldsymbol{\rho}_{\mathbf{x}x}^T$ is a matrix of rank 1, $\mathbf{R}_{\mathbf{x}_{\mathrm{u}}}$ is the correlation matrix of $\mathbf{x}_{\mathrm{u}}(k)$, $\mathbf{R}_{\mathbf{v}_1} = \sigma_v^2 \boldsymbol{\rho}_{\mathbf{v}v}\boldsymbol{\rho}_{\mathbf{v}v}^T$, and $\mathbf{R}_{\mathbf{v}_{\mathrm{u}}}$ is the correlation matrix of $\mathbf{v}_{\mathrm{u}}(k)$.

With the above formulation, the objective of single-channel noise reduction (or speech enhancement) in the time domain is then to find from the observation vector, $\mathbf{y}(k)$, a "good" estimate of the sample, $x(k)$, in the sense that the additive noise is significantly reduced while the desired signal is not much distorted.

3. LINEAR FILTERING AND THE SQUARED PEARSON CORRELATION COEFFICIENT

In this paper, we estimate the desired signal sample, $x(k)$, or the noise signal sample, $v(k)$, by applying a real-valued filter, $\mathbf{h}$, of length L, to the observation signal vector, $\mathbf{y}(k)$, i.e.,

$$z(k) = \mathbf{h}^T\mathbf{y}(k), \qquad (9)$$

where $z(k)$ can be either an estimate of $x(k)$ or $v(k)$. If $z(k)$ is an estimate of $v(k)$, then an estimate of $x(k)$ can be obtained as $\widehat{x}(k) = y(k) - z(k)$.

It is of great importance to know how good is the estimate $z(k)$. One of the best second-order statistics based measures to evaluate this is via the squared Pearson correlation coefficient (SPCC) [16]. We define the SPCC between $z(k)$ and $x(k)$ as

$$\rho_{zx}^2(\mathbf{h}) \triangleq \frac{E^2\left[z(k)x(k)\right]}{E\left[z^2(k)\right]E\left[x^2(k)\right]} \qquad (10)$$
$$= \frac{\sigma_x^2\left(\mathbf{h}^T\boldsymbol{\rho}_{\mathbf{x}x}\right)^2}{\mathbf{h}^T\mathbf{R}_{\mathbf{y}}\mathbf{h}} = \frac{\mathbf{h}^T\mathbf{R}_{\mathbf{x}_1}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{\mathbf{y}}\mathbf{h}}.$$

We see that the SPCC defined above depends explicitly on the filter, $\mathbf{h}$, which suggests that we can use the SPCC as a criterion to derive optimal filters. Note that many different forms of the SPCC can be used [16], but this paper focuses only on $\rho_{zx}^2(\mathbf{h})$.

4. EXAMPLES OF OPTIMAL FILTERS

Intuitively, it makes sense to maximize or minimize the SPCC in order to find an estimate of $x(k)$ or $v(k)$. It is clear that the maximization (resp. minimization) of $\rho_{zx}^2(\mathbf{h})$ will give a good estimate of $x(k)$ [resp. $v(k)$], since in this case the SPCC between $z(k)$ and $x(k)$ will be maximal [resp. minimal], implying that $z(k)$ is close to $x(k)$ [resp. $v(k)$].

4.1. Maximization of the SPCC

It is obvious that the maximization of (10) leads to an estimate of the desired signal. In (10), we recognize the generalized Rayleigh quotient [19]. It is well known that this quotient is maximized with the maximum eigenvector, $\mathbf{a}_1$, of the matrix $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}_1}$. Let us denote by $\lambda_{\mathbf{a}_1}$ the maximum eigenvalue corresponding to $\mathbf{a}_1$. Since the rank of $\mathbf{R}_{\mathbf{x}_1}$ is equal to 1, it can be checked that

$$\mathbf{a}_1 = \frac{\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\sqrt{\boldsymbol{\rho}_{\mathbf{x}x}^T\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}}, \qquad (11)$$

$$\lambda_{\mathbf{a}_1} = \sigma_x^2\boldsymbol{\rho}_{\mathbf{x}x}^T\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}, \qquad (12)$$

and the maximum of the SPCC is

$$\rho_{zx,\mathrm{max}}^2 = \lambda_{\mathbf{a}_1}. \qquad (13)$$

As a result, the optimal filter is

$$\mathbf{h}_{\alpha_1} = \alpha_1\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}, \qquad (14)$$

where $\alpha_1 \neq 0$ is an arbitrary real number, whose value is important in practice when we deal with nonstationary signals like speech[1]. Hence, the estimate of $x(k)$ is

$$\widehat{x}_{\alpha_1}(k) = \mathbf{h}_{\alpha_1}^T\mathbf{y}(k). \qquad (15)$$

Now, we need to determine α_1. There are at least two interesting ways to find this parameter. The first one is from examining the MSE between $x(k)$ and $\widehat{x}_{\alpha_1}(k)$, i.e.,

$$J\left(\mathbf{h}_{\alpha_1}\right) \triangleq E\left\{\left[x(k) - \mathbf{h}_{\alpha_1}^T\mathbf{y}(k)\right]^2\right\}. \qquad (16)$$

The second possibility is to use the distortion-based MSE, i.e.,

$$J_{\mathrm{ds}}\left(\mathbf{h}_{\alpha_1}\right) \triangleq E\left\{\left[x(k) - \mathbf{h}_{\alpha_1}^T\mathbf{x}_1(k)\right]^2\right\} \qquad (17)$$
$$= \sigma_x^2\left(1 - \mathbf{h}_{\alpha_1}^T\boldsymbol{\rho}_{\mathbf{x}x}\right)^2.$$

The minimization of $J\left(\mathbf{h}_{\alpha_1}\right)$ with respect to α_1 leads to

$$\alpha_1 = \sigma_x^2. \qquad (18)$$

Substituting (18) into (14), we get the conventional Wiener filter [4]:

$$\mathbf{h}_{\mathrm{W}} = \sigma_x^2\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}\mathbf{i}_1$$
$$= \left(\mathbf{I}_L - \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{v}}\right)\mathbf{i}_1, \qquad (19)$$

where $\mathbf{i}_1$ is the first column of the $L \times L$ identity matrix, $\mathbf{I}_L$.

By minimizing $J_{\mathrm{ds}}\left(\mathbf{h}_{\alpha_1}\right)$ with respect to α_1, we obtain

$$\alpha_1 = \frac{1}{\boldsymbol{\rho}_{\mathbf{x}x}^T\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}. \qquad (20)$$

Substituting the previous result into (14) gives the MVDR filter derived in [18]:

$$\mathbf{h}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\boldsymbol{\rho}_{\mathbf{x}x}^T\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}. \qquad (21)$$

---

[1]Obviously, for stationary signals, the value of α_1 does not matter at all as long as its value is different from zero.

4.2. Minimization of the SPCC

Another approach to find optimal filters is by minimizing (10). Therefore, the filter output will be an estimate of $v(k)$. The product matrix $\mathbf{R_y}^{-1} \mathbf{R_{x_1}}$ has $L-1$ eigenvalues equal to 0, since its rank is equal to 1. Let $\mathbf{a}_2, \mathbf{a}_3, \ldots, \mathbf{a}_L$ be the corresponding eigenvectors. The filter:

$$\mathbf{h}_{\alpha_{2:L}} = \sum_{i=2}^{L} \alpha_i \mathbf{a}_i, \tag{22}$$

where $\alpha_i, \; i = 2, 3, \ldots, L$ are arbitrary real numbers with at least one of them different from 0, minimizes (10). In this case, we have $\mathbf{h}_{\alpha_{2:L}}^T \boldsymbol{\rho}_{\mathbf{x}x} = 0$ and

$$\rho_{zx}^2 (\mathbf{h}_{\alpha_{2:L}}) = \rho_{zx,\min}^2 = 0. \tag{23}$$

We can rewrite (22) as

$$\mathbf{h}_{\alpha_{2:L}} = \mathbf{A}_{2:L} \boldsymbol{\alpha}_{2:L}, \tag{24}$$

where

$$\mathbf{A}_{2:L} \triangleq \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_3 & \cdots & \mathbf{a}_L \end{bmatrix} \tag{25}$$

is a matrix of size $L \times (L-1)$ and

$$\boldsymbol{\alpha}_{2:L} \triangleq \begin{bmatrix} \alpha_2 & \alpha_3 & \cdots & \alpha_L \end{bmatrix}^T \neq \mathbf{0} \tag{26}$$

is a vector of length $L-1$. Then, the estimates of $v(k)$ and $x(k)$ are, respectively,

$$\widehat{v}_{\alpha_{2:L}}(k) = \mathbf{h}_{\alpha_{2:L}}^T \mathbf{y}(k) = \mathbf{h}_{\alpha_{2:L}}^T \mathbf{x}_\mathrm{u}(k) + \mathbf{h}_{\alpha_{2:L}}^T \mathbf{v}(k) \tag{27}$$

and

$$\widehat{x}_{\alpha_{2:L}}(k) = y(k) - \widehat{v}_{\alpha_{2:L}}(k) = \mathbf{h}_{\alpha_{2:L}}'^T \mathbf{y}(k), \tag{28}$$

where

$$\mathbf{h}_{\alpha_{2:L}}' = \mathbf{i}_1 - \mathbf{h}_{\alpha_{2:L}} \tag{29}$$

is the equivalent filter for the estimation of $x(k)$.

The MSE between $v(k)$ and $\widehat{v}_{\alpha_{2:L}}(k)$ is

$$J(\mathbf{h}_{\alpha_{2:L}}) \triangleq E\left\{ \left[v(k) - \mathbf{h}_{\alpha_{2:L}}^T \mathbf{y}(k) \right]^2 \right\} \tag{30}$$

$$= J_\mathrm{rus}(\mathbf{h}_{\alpha_{2:L}}) + J_\mathrm{rn}(\mathbf{h}_{\alpha_{2:L}}),$$

where

$$J_\mathrm{rus}(\mathbf{h}_{\alpha_{2:L}}) \triangleq \mathbf{h}_{\alpha_{2:L}}^T \mathbf{R}_{\mathbf{x}_\mathrm{u}} \mathbf{h}_{\alpha_{2:L}}, \tag{31}$$

$$J_\mathrm{rn}(\mathbf{h}_{\alpha_{2:L}}) \triangleq E\left\{ \left[v(k) - \mathbf{h}_{\alpha_{2:L}}^T \mathbf{v}(k) \right]^2 \right\}, \tag{32}$$

are the MSEs of the residual uncorrelated speech and residual noise, respectively. Equivalently, we can write $J(\mathbf{h}_{\alpha_{2:L}})$ as

$$J(\mathbf{h}_{\alpha_{2:L}}) = J_\mathrm{rus}(\mathbf{h}_{\alpha_{2:L}}) + J_\mathrm{run}(\mathbf{h}_{\alpha_{2:L}}) + J_\mathrm{rcn}(\mathbf{h}_{\alpha_{2:L}}),$$

where

$$J_\mathrm{run}(\mathbf{h}_{\alpha_{2:L}}) \triangleq \mathbf{h}_{\alpha_{2:L}}^T \mathbf{R}_{\mathbf{v}_\mathrm{u}} \mathbf{h}_{\alpha_{2:L}}, \tag{33}$$

$$J_\mathrm{rcn}(\mathbf{h}_{\alpha_{2:L}}) \triangleq \sigma_v^2 \left(1 - \mathbf{h}_{\alpha_{2:L}}^T \boldsymbol{\rho}_{\mathbf{v}v} \right)^2, \tag{34}$$

are the MSEs of the residual uncorrelated and correlated noise, respectively. From (30) and (33), we see that there are at least three interesting ways to find $\boldsymbol{\alpha}_{2:L}$.

Now, we can design different filters. The first possibility is through minimizing the power of the residual noise, $J_\mathrm{rn}(\mathbf{h}_{\alpha_{2:L}})$. The minimization of $J_\mathrm{rn}(\mathbf{h}_{\alpha_{2:L}})$ with respect to $\boldsymbol{\alpha}_{2:L}$ gives

$$\boldsymbol{\alpha}_{2:L} = \left(\mathbf{A}_{2:L}^T \mathbf{R_v} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \mathbf{R_v} \mathbf{i}_1. \tag{35}$$

As a result,

$$\mathbf{h}_{\alpha_{2:L}} = \mathbf{A}_{2:L} \left(\mathbf{A}_{2:L}^T \mathbf{R_v} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \mathbf{R_v} \mathbf{i}_1 \tag{36}$$

and the minimum noise (MN) filter for the estimation of $x(k)$ is

$$\mathbf{h}_\mathrm{MN}' = \left[\mathbf{I}_L - \mathbf{A}_{2:L} \left(\mathbf{A}_{2:L}^T \mathbf{R_v} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \mathbf{R_v} \right] \mathbf{i}_1. \tag{37}$$

This filter is new and has never been studied in the literature.

By minimizing the MSE $J(\mathbf{h}_{\alpha_{2:L}})$, we find the minimum uncorrelated speech and noise (MUSN) filter:

$$\mathbf{h}_\mathrm{MUSN}' = \left[\mathbf{I}_L - \mathbf{A}_{2:L} \left(\mathbf{A}_{2:L}^T \mathbf{R_y} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \mathbf{R_v} \right] \mathbf{i}_1. \tag{38}$$

This second option gives also a new filter, which may reduce less noise than $\mathbf{h}_\mathrm{MN}'$ but at the same time it will reduce the level of the uncorrelated speech that can be considered as a distortion to the desired signal.

Finally, the last possibility is to minimize $J_\mathrm{rus}(\mathbf{h}_{\alpha_{2:L}}) + J_\mathrm{run}(\mathbf{h}_{\alpha_{2:L}})$ subject to $\mathbf{h}_{\alpha_{2:L}}^T \boldsymbol{\rho}_{\mathbf{v}v} = 1$. We find that

$$\boldsymbol{\alpha}_{2:L} = \frac{\left(\mathbf{A}_{2:L}^T \mathbf{R}_\mathrm{usn} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \boldsymbol{\rho}_{\mathbf{v}v}}{\boldsymbol{\rho}_{\mathbf{v}v}^T \mathbf{A}_{2:L} \left(\mathbf{A}_{2:L}^T \mathbf{R}_\mathrm{usn} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \boldsymbol{\rho}_{\mathbf{v}v}}, \tag{39}$$

where $\mathbf{R}_\mathrm{usn} = \mathbf{R}_{\mathbf{x}_\mathrm{u}} + \mathbf{R}_{\mathbf{v}_\mathrm{u}}$, whose rank is assumed to be at least equal to $L-1$. As a consequence, we get the linearly constrained minimum variance (LCMV) filter:

$$\mathbf{h}_\mathrm{LCMV}' = \mathbf{i}_1 - \frac{\mathbf{A}_{2:L} \left(\mathbf{A}_{2:L}^T \mathbf{R}_\mathrm{usn} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \boldsymbol{\rho}_{\mathbf{v}v}}{\boldsymbol{\rho}_{\mathbf{v}v}^T \mathbf{A}_{2:L} \left(\mathbf{A}_{2:L}^T \mathbf{R}_\mathrm{usn} \mathbf{A}_{2:L} \right)^{-1} \mathbf{A}_{2:L}^T \boldsymbol{\rho}_{\mathbf{v}v}}. \tag{40}$$

5. SIMULATIONS

In this section, we briefly study the different filters and compare their noise reduction performance through simulations. The clean speech signal used was recorded from a male talker in a quiet office room. It was sampled at 8 kHz. The overall length of the signal is 30 seconds. The noisy signal is obtained by adding some noise into the clean speech with an input SNR of 10 dB. The noise is a mixture of a Gaussian random process and a sine wave of 1 kHz. The ratio between the intensity of the sine wave and that of the white noise is 6 dB.

To implement the optimal filters derived in section 4, it requires the estimation of the correlation matrices and vectors. The estimate of the correlation matrices $\mathbf{R_y}$ and $\mathbf{R_v}$ is computed directly from the corresponding signals using a short-time average (the most recent 400 samples). An estimate of the correlation matrix of the clean speech, $\mathbf{R_x}$, is computed according to $\widehat{\mathbf{R}}_\mathbf{x}(k) = \widehat{\mathbf{R}}_\mathbf{y}(k) - \widehat{\mathbf{R}}_\mathbf{v}(k)$. The other correlation matrices and vectors are computed according to their definitions in Section 4.

We use both the output SNR and speech distortion index as performance measures. The former quantifies the amount of noise reduction while the latter quantifies the amount of speech distortion. It
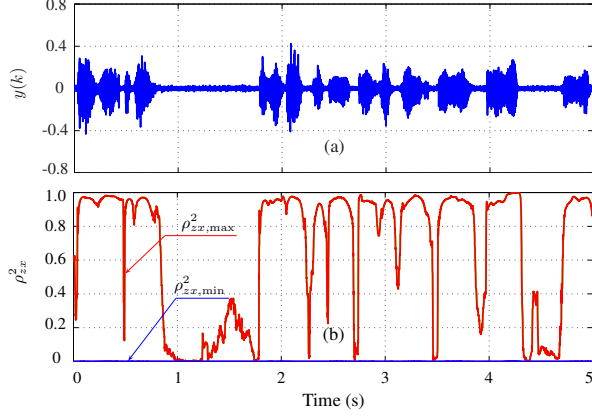
Fig. 1. Behavior of $\rho_{zx,\max}^2$ and $\rho_{zx,\min}^2$: (a) noisy signal, (b) $\rho_{zx,\max}^2$ computed with the Wiener and MVDR filters, and $\rho_{zx,\min}^2$ computed with the MN, MUSN, and LCMV filters. iSNR $= 10$ dB and $L = 8$.

is seen in Section 2 that the interference part of the clean speech signal vector is uncorrelated with the desired clean speech sample. So, this part, after filtering, should not be treated as part of the desired signal [18]. As a result, we can define the output SNR after noise reduction as

$$\text{oSNR} \triangleq \frac{E\left[x_{\text{fd}}^2(k)\right]}{E\left[x_{\text{ri}}^2(k)\right] + E\left[v_{\text{rn}}^2(k)\right]}, \tag{41}$$

where $x_{\text{fd}}(k)$ is the filtered desired signal, $x_{\text{ri}}(k)$ is the residual interference, and $v_{\text{rn}}(k)$ is the residual noise. The speech distortion index is defined as [18]

$$v_{\text{sd}} = \frac{E\left\{[x_{\text{fd}} - x(k)]^2\right\}}{E\left[x^2(k)\right]}. \tag{42}$$

The filters derived in Section 4 are divided into two classes. The first one, consisting of the Wiener and MVDR filters, attempts to maximum the SPCC, $\rho_{\mathbf{x}x}$, and the resulting filters' outputs are estimates of the desired signal, $x(k)$. The second class, consisting of the MN, MUSN, and LCMV filters, attempts to minimize the SPCC, $\rho_{\mathbf{x}x}$, and the resulting filters' outputs are estimates of the noise signal, $v(k)$.

Figure 1 illustrates the SPCC between $z(k)$ and $x(k)$ obtained with different optimal filters (for the first 5 seconds), where $\rho_{zx,\max}^2$ is obtained using the Wiener and MVDR filters (both of them yielded similar $\rho_{zx,\max}^2$) and $\rho_{zx,\min}^2$ is obtained using the MN, MUSN, and LCMV filters (all of them produced similar $\rho_{zx,\min}^2$). The value of $\rho_{zx,\max}^2$ is close to 1 during the presence of speech. This indicates that the outputs of the Wiener and MVDR filters are good estimates of $x(k)$. In comparison, the value of $\rho_{zx,\min}^2$ is close to 0. This suggests that the outputs of the MN, MUSN, and LCMV filters are good estimates of $v(k)$.

Figure 2 plots the output SNR and speech distortion index as a function of the filter length, L. It is seen that the Wiener filter has the largest SNR improvement; but it also introduces more speech distortion than other studied filters as indicated by the speech distortion index. Compared with the Wiener filter, the MVDR one has a lower output SNR but with no speech distortion. In fact, the only difference between the Wiener and MVDR filters is a scaling factor [α_1 in (14)]. But this factor is time varying due to the nonstationarity of speech. By adjusting it, the MVDR filter manages to keep the desired speech undistorted. It is also seen from Fig. 2(b) that
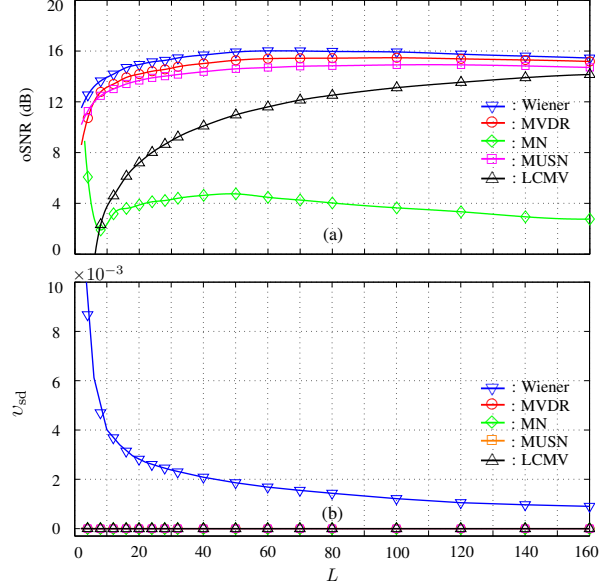

Fig. 2. Performance of the optimal filters (Wiener, MVDR, MN, MUSN, and LCMV) as a function of the filter length, L, in white Gaussian plus sinusoidal noise: (a) output SNR and (b) speech distortion index. iSNR $= 10$ dB.

all filters in the second class do not introduce any speech distortion; this is consistent with the theoretical analysis in Section 4. Both the MUSN and LCMV filters can improve the SNR if the filter length, L, is properly chosen; but for the MN filter, the output SNR is smaller than the input SNR for the studied noise condition. The underlying reason is that the MN filter only attempts to minimize the residual noise; but it causes some amplification of the residual interference. More efforts are underway to study the advantages and limits of all the deduced filters.

6. SUMMARY

In this paper, we studied the single-channel noise reduction problem in the time domain. The SPCC between the filter output and desired signal is used as a criterion. By maximizing and minimizing this SPCC, we showed how to derive two classes of noise reduction filters. The first class attempts to achieve directly an estimate of the desired speech signal by maximizing the SPCC criterion. The filters in this class are equivalent to the traditional ones derived from the MSE criterion. The second class attempts to obtain an estimate of the noise by minimizing the SPCC. Different new filters were then derived. Simulation results showed that the filters can have very different noise reduction performance. Their potential in dealing with different types of noise is under investigation.

7. RELATION TO PRIOR WORK

Noise reduction is a fundamental problem, which has a broad range of applications [1]–[18]. Typically, the problem is formulated as a filtering technique. So, the paramount issue with noise reduction is to design an optimal filter that can significantly reduce the level of the noise while keeping the distortion of the desired speech signal as low as possible [1]–[6]. Traditionally, most optimal filters are derived from the MSE criterion. In this paper, we adopted the SPCC [16] as the cost function. With the use of SPCC, we showed how to derive different noise reduction filters. Some of the filters are the same as those derived with the MSE criterion, while others are new.

8. REFERENCES

[1] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters—A Theoretical Study*. Springer Briefs in Electrical and Computer Engineering, 2011.

[2] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*. Berlin, Germany: Springer-Verlag, 2005.

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

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

[5] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech, Audio Process.*, vol. 3, pp. 439–448, Nov. 1995

[6] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, pp. 2230–244, Sept. 2002.

[7] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 113–120, Apr. 1979.

[8] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, pp. 1586–1604, Dec. 1979.

[9] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 3, pp. 251–266, July 1995.

[10] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, pp. 1109–1121, Dec. 1984.

[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp. 443–445, Apr. 1985.

[12] H. Lev-Ari and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise," *IEEE Trans. Speech, Audio Process.*, vol. 10, pp. 104–106, Apr. 2003.

[13] A. Rezayee and S. Gazor, "An adpative KLT approach for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 9, pp. 87–95, Feb. 2001.

[14] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech, Audio Process.*, vol. 8, pp. 159–167, Mar. 2000.

[15] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, pp. 137–145, Apr. 1980.

[16] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, pp. 757–765, May 2008.

[17] J. Chen, J. Benesty, Y. Huang, and T. Gaensler, "On single-channel noise reduction in the time domain," in *Proc. IEEE ICASSP*, 2011, pp. 277–280.

[18] 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. Am.*, vol. 132, pp. 452–464, July 2012.

[19] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1968.