Single-channel noise reduction using optimal rectangular filtering matrices

Tao Long, Jingdong Chen, Jacob Benesty, and Zhenxi Zhang

ARTICLES YOU MAY BE INTERESTED IN

Flow simulations on an organ pipe foot model
The Journal of the Acoustical Society of America **133**, 1102 (2013); https://doi.org/10.1121/1.4773861

Multichannel active noise control for spatially sparse noise fields
The Journal of the Acoustical Society of America **140**, EL510 (2016); https://doi.org/10.1121/1.4971878

Noise reduction with complex bilateral filter
The Journal of the Acoustical Society of America **142**, 3486 (2017); https://doi.org/10.1121/1.5016807

Time-domain noise reduction based on an orthogonal decomposition for desired signal extraction
The Journal of the Acoustical Society of America **132**, 452 (2012); https://doi.org/10.1121/1.4726071

Active control of scattered acoustic fields: Cancellation, reproduction and cloaking
The Journal of the Acoustical Society of America **140**, 1502 (2016); https://doi.org/10.1121/1.4962284

Multi-channel Kalman filters for active noise control
The Journal of the Acoustical Society of America **133**, 2105 (2013); https://doi.org/10.1121/1.4792646

Single-channel noise reduction using optimal rectangular filtering matrices

Tao Long[a)]

Key Laboratory of Biomedical Information Engineering of Education Ministry, Institute of Biomedical Analytical Technology and Instrumentation, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an, Shaanxi 710049, China

Jingdong Chen

Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, Shaanxi 710072, China

Jacob Benesty

Institut National de la Recherche Scientifique-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, Quebec H5A 1K6, Canada

Zhenxi Zhang

Key Laboratory of Biomedical Information Engineering of Education Ministry, Institute of Biomedical Analytical Technology and Instrumentation, Life science department, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an, Shaanxi 710049, China

This paper studies the problem of single-channel noise reduction in the time domain and presents a block-based approach where a vector of the desired speech signal is recovered by filtering a frame of the noisy signal with a rectangular filtering matrix. With this formulation, the noise reduction problem becomes one of estimating an optimal filtering matrix. To achieve such estimation, a method is introduced to decompose a frame of the clean speech signal into two orthogonal components: One correlated and the other uncorrelated with the current desired speech vector to be estimated. Different optimization cost functions are then formulated from which non-causal optimal filtering matrices are derived. The relationships among these optimal filtering matrices are discussed. In comparison with the classical sample-based technique that uses only forward prediction, the block-based method presented in this paper exploits both the forward and backward prediction as well as the temporal interpolation and, therefore, can improve the noise reduction performance by fully taking advantage of the speech property of self correlation. There is also a side advantage of this block-based method as compared to the sample-based technique, i.e., it is computationally more efficient and, as a result, more suitable for practical implementation.
© 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4773269]

I. INTRODUCTION

In many applications of speech processing, an effective noise reduction algorithm is required. Over the past several decades, many algorithms have been developed and improved (Benesty *et al.*, 2009; Loizou, 2007; Chen *et al.*, 2007; Vary and Martin, 2006; Ephraim, 1992; Lim, 1983). However, it is well known that these single-microphone algorithms only achieve noise reduction at a price of modifying the desired speech signal, leading to speech distortion. In general, the more the noise is attenuated, the more the speech is distorted.

Recently, it has been shown that by decomposing the clean speech signal vector into two orthogonal components, i.e., the desired speech component that is correlated with the speech sample to be estimated and the interference component that is uncorrelated with the speech sample to be estimated, many new noise reduction filters can be deduced (Chen *et al.*, 2011; Benesty and Chen, 2011; Benesty *et al.*, 2012). Particularly, the minimum variance distortionless response (MVDR) filter can be designed thanks to this orthogonal decomposition, which can achieve noise reduction without distorting the desired speech signal in the single-channel case, which has never been seen before in the literature.

The algorithms developed in Benesty *et al.* (2012) are sample-based approaches, i.e., one sample at a time is estimated, where the speech signal at the current time instant is always predicted using the past samples through forward prediction. In a more recent effort (Li *et al.*, 2012), we extended the sample-based approach into a block-based framework and deduced some algorithms that estimate a vector of the desired speech signal through filtering a frame of the noisy signal with a rectangular filtering matrix. In comparison with the sample-based techniques, this block-based method is shown to have the potential for better noise reduction performance as it exploits both the forward and

---

[a)]Author to whom correspondence should be addressed. Electronic mail: longtao_2002@163.com

backward prediction as well as the temporal interpolation and, therefore, can fully take advantage of the speech property of correlation with neighboring (both past and future) samples. Another side advantage of the block-based framework as compared to the sample-based one is the computational complexity. As we know, the largest computational burden of a time-domain noise reduction algorithm is from the matrix inversion. In the sample-based method, a matrix inversion is needed for every sample. But in a block-based algorithm, it is only needed for each vector. So the larger the vector size, the lower the computational burden.

This paper is a continuation of the work presented in Li et al. (2012). The major contribution of this work is three-fold. (1) It extends the previous block-based approach to a more general form with the inclusion of a delay parameter. As it will be shown later, this delay parameter adds flexibility in controlling the degree of non-causality in the filtering matrices so that the output signal-to-noise ratio (SNR) can be improved and the speech distortion can be lowered. (2) It presents a more comprehensive theoretical study and thorough analysis of the block-based noise reduction approach. (3) More experimental investigation is provided; this further justifies the advantages of the block-based approach.

The rest of this paper is organized as follows. In Sec. II, we describe the signal model and the noise reduction problem that is to be tackled in this work. We then discuss how to deduce different optimal filtering matrices such as the maximum SNR, Wiener, MVDR, prediction, and tradeoff filtering matrices in Sec. III. In Sec. IV, we use experiments to evaluate the performance of the different filters with different values of the important parameters. Finally, we give our conclusions in Sec. V.

II. SIGNAL MODEL AND PROBLEM FORMULATION

In the time domain, we assume that the observed signal, $y(k)$, is an additive mixture of the clean speech, $x(k)$, and the noise, $v(k)$, i.e.,

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

where $x(k)$ and $v(k)$ are assumed to be uncorrelated and zero-mean random processes, and k is the discrete-time index. All signals are considered to be real and broadband.

The signal model given in Eq. (1) can be put into a vector form:

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

where

$$\mathbf{y}(k) \triangleq [\,y(k) \quad y(k-1) \quad \cdots \quad y(k-L+1)\,]^T \tag{3}$$

is a vector of length L, superscript T denotes transpose of a vector or a matrix, and $\mathbf{x}(k)$ and $\mathbf{v}(k)$ are defined in a similar way to $\mathbf{y}(k)$.

In this paper, we estimate more than one sample at a time. Therefore, we define the following two vectors of length M:

$$\tilde{\mathbf{x}}(k) \triangleq [\,x(k) \quad x(k-1) \quad \cdots \quad x(k-M+1)\,]^T, \tag{4}$$

$$\tilde{\mathbf{x}}_n(k) \triangleq \tilde{\mathbf{x}}(k-n), \tag{5}$$

where $M \leq L$ and n can be any integer in the interval $[0, L-M-1]$. Our objective in this work is to estimate the desired signal vector, $\tilde{\mathbf{x}}_n(k) = \tilde{\mathbf{x}}(k-n)$. (Note that the delay parameter, n, is introduced to allow some flexibility of using non-causality in the filtering process, which can help improve noise reduction performance as it will be shown later.) This can be achieved by applying a linear transformation to $\mathbf{y}(k)$ (Benesty et al., 2009; Benesty and Chen, 2011), i.e.,

$$\begin{aligned} \tilde{\mathbf{z}}(k) &= \mathbf{H}_n \mathbf{y}(k) = \mathbf{H}_n[\mathbf{x}(k) + \mathbf{v}(k)] \\ &= \tilde{\mathbf{x}}_{\mathrm{f},n}(k) + \tilde{\mathbf{v}}_{\mathrm{rn},n}(k), \end{aligned} \tag{6}$$

where the vector $\tilde{\mathbf{z}}(k)$ of length M is the estimate of $\tilde{\mathbf{x}}_n(k)$,

$$\mathbf{H}_n = [\,\mathbf{h}_{n,1} \quad \mathbf{h}_{n,2} \quad \cdots \quad \mathbf{h}_{n,M}\,]^T \tag{7}$$

is a rectangular filtering matrix of size $M \times L$,

$$\begin{aligned} \mathbf{h}_{n,m} &= [\,h_{n,m,0} \quad h_{n,m,1} \quad \cdots \quad h_{n,m,L-1}\,]^T, \\ &\quad m = 1, 2, \ldots, M \end{aligned} \tag{8}$$

are finite-impulse-response (FIR) filters of length L, and

$$\tilde{\mathbf{x}}_{\mathrm{f},n}(k) \triangleq \mathbf{H}_n \mathbf{x}(k), \tag{9}$$

$$\tilde{\mathbf{v}}_{\mathrm{rn},n}(k) \triangleq \mathbf{H}_n \mathbf{v}(k) \tag{10}$$

are the filtered speech and residual noise, respectively. Depending on the values of M and n, there are five important cases of Eq. (6) as described in the following text.

(a) $M = 1$ and $n = 0$. In this situation, $\tilde{\mathbf{z}}(k)$ becomes a scalar $z(k)$, which is the estimate of $x(k)$. The filtering matrix $\mathbf{H}_0$ degenerates to a causal FIR filter (vector) $\mathbf{h}_0^T$ of length L. This case has been studied in Chen et al., (2011); Benesty et al., (2012).

(b) $M = 1$ and $n > 0$. In this case, the vector $\tilde{\mathbf{z}}(k)$, again, degenerates to the scalar $z(k)$ as in the previous situation. But the difference is that this time, $\mathbf{H}_n$ degenerates to a non-causal FIR filter $\mathbf{h}_n^T$ of length L and $z(k)$ is the estimate of $x(k-n)$. This case has been studied in Jensen et al. (2012).

(c) $M = L$. In this situation, n has to be 0, $\tilde{\mathbf{z}}(k)$ is a vector of length L and $\mathbf{H}_0$ is a square matrix of size $L \times L$. This scenario has been widely covered in the literature such as (Ephraim and Trees, 1995; Jensen et al., 1995; Hu and Loizou, 2002; Benesty et al., 2009).

(d) $1 \leq M \leq L$ and $n = 0$. This case has been briefly studied in Li et al. (2012) and will be more comprehensively investigated in this paper.

(e) $1 \leq M \leq L$ and $0 \leq n \leq L$. This general case is the focus of this paper.

By definition, our desired signal is the vector $\tilde{\mathbf{x}}_n(k)$. Therefore we need to extract it from $\mathbf{x}(k)$. For that, we consider decomposing $\mathbf{x}(k)$ into two orthogonal components:

One that is correlated with (or is a linear transformation of) the desired signal $\tilde{\mathbf{x}}_n(k)$ and another that is orthogonal to $\tilde{\mathbf{x}}_n(k)$ and, hence, will be considered as the interference component. Specifically, the vector $\mathbf{x}(k)$ is decomposed into the following form:

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n}\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1}\tilde{\mathbf{x}}_n(k) + \mathbf{x}_{\mathrm{i},n}(k) \\ &= \mathbf{x}_{\mathrm{d},n}(k) + \mathbf{x}_{\mathrm{i},n}(k), \end{aligned} \tag{11}$$

where

$$\mathbf{x}_{\mathrm{d},n}(k) \triangleq \mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n}\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1}\tilde{\mathbf{x}}_n(k) = \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}\tilde{\mathbf{x}}_n(k) \tag{12}$$

is a linear transformation of the desired signal, $\mathbf{R}_{\tilde{\mathbf{x}}_n} \triangleq E[\tilde{\mathbf{x}}_n(k)\tilde{\mathbf{x}}_n^T(k)]$ is the correlation matrix (of size $M \times M$) of $\tilde{\mathbf{x}}_n(k)$ with $E[\cdot]$ denoting mathematical expectation, $\mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n} \triangleq E[\mathbf{x}(k)\tilde{\mathbf{x}}_n^T(k)]$ is the cross-correlation matrix (of size $L \times M$) between $\mathbf{x}(k)$ and $\tilde{\mathbf{x}}_n(k)$, $\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} \triangleq \mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n}\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1}$, and

$$\mathbf{x}_{\mathrm{i},n}(k) = \mathbf{x}(k) - \mathbf{x}_{\mathrm{d},n}(k) \tag{13}$$

is the interference signal. It is easy to see that $\mathbf{x}_{\mathrm{d},n}(k)$ and $\mathbf{x}_{\mathrm{i},n}(k)$ are orthogonal, i.e.,

$$E[\mathbf{x}_{\mathrm{d},n}(k)\mathbf{x}_{\mathrm{i},n}^T(k)] = \mathbf{0}_{L \times L}. \tag{14}$$

For the particular case $M = L$, we have $\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} = \mathbf{I}_L$, which is the identity matrix (of size $L \times L$), and $\mathbf{x}_{\mathrm{d},n}(k)$ coincides with $\mathbf{x}(k)$, which obviously makes sense. For $M = 1$, $\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}$ simplifies to the normalized correlation vector (Chen *et al.*, 2011):

$$\gamma_{\mathbf{x},n} = \frac{E[\mathbf{x}(k)x(k-n)]}{E[x^2(k-n)]}. \tag{15}$$

Substituting Eq. (11) into Eq. (6), we get

$$\begin{aligned} \tilde{\mathbf{z}}(k) &= \mathbf{H}_n[\mathbf{x}_{\mathrm{d},n}(k) + \mathbf{x}_{\mathrm{i},n}(k) + \mathbf{v}(k)] \\ &= \tilde{\mathbf{x}}_{\mathrm{fd},n}(k) + \tilde{\mathbf{x}}_{\mathrm{ri},n}(k) + \tilde{\mathbf{v}}_{\mathrm{rn},n}(k), \end{aligned} \tag{16}$$

where

$$\tilde{\mathbf{x}}_{\mathrm{fd},n}(k) \triangleq \mathbf{H}_n\mathbf{x}_{\mathrm{d},n}(k), \tag{17}$$

$$\tilde{\mathbf{x}}_{\mathrm{ri},n}(k) \triangleq \mathbf{H}_n\mathbf{x}_{\mathrm{i},n}(k) \tag{18}$$

are the filtered desired signal and the residual interference, respectively. It can be checked that the three terms $\tilde{\mathbf{x}}_{\mathrm{fd},n}(k)$, $\tilde{\mathbf{x}}_{\mathrm{ri},n}(k)$, and $\tilde{\mathbf{v}}_{\mathrm{rn},n}(k)$ are mutually uncorrelated. Therefore the correlation matrix of $\tilde{\mathbf{z}}(k)$ is

$$\mathbf{R}\tilde{\mathbf{z}} \triangleq E[\tilde{\mathbf{z}}(k)\tilde{\mathbf{z}}^T(k)] = \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{fd},n}} + \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{ri},n}} + \mathbf{R}_{\tilde{\mathbf{v}}_{\mathrm{rn},n}}, \tag{19}$$

where

$$\mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{fd},n}} = \mathbf{H}_n\mathbf{R}_{\mathbf{x}_{\mathrm{d},n}}\mathbf{H}_n^T, \tag{20}$$

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{x}}_{\mathrm{ri},n}} &= \mathbf{H}_n\mathbf{R}_{\mathbf{x}_{\mathrm{i},n}}\mathbf{H}_n^T \\ &= \mathbf{H}_n\mathbf{R}_\mathbf{x}\mathbf{H}_n^T - \mathbf{H}_n\mathbf{R}_{\mathbf{x}_{\mathrm{d},n}}\mathbf{H}_n^T, \end{aligned} \tag{21}$$

$$\mathbf{R}_{\tilde{\mathbf{v}}_{\mathrm{rn},n}} = \mathbf{H}_n\mathbf{R}_\mathbf{v}\mathbf{H}_n^T, \tag{22}$$

$\mathbf{R}_{\mathbf{x}_{\mathrm{d},n}} = \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}\mathbf{R}_{\tilde{\mathbf{x}}_n}\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T$ is the correlation matrix (the rank of which is equal to M) of $\mathbf{x}_{\mathrm{d},n}(k)$, $\mathbf{R}_{\mathbf{x}_{\mathrm{i},n}} \triangleq E[\mathbf{x}_{\mathrm{i},n}(k)\mathbf{x}_{\mathrm{i},n}^T(k)]$ is

the correlation matrix of $\mathbf{x}_{\mathrm{i},n}(k)$, $\mathbf{R}_\mathbf{x} \triangleq E[\mathbf{x}(k)\mathbf{x}^T(k)]$ and $\mathbf{R}_\mathbf{v} \triangleq E[\mathbf{v}(k)\mathbf{v}^T(k)]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively.

Now, the error signal between the estimated and desired signals can be defined as a vector of length M:

$$\tilde{\mathbf{e}}_n(k) \triangleq \tilde{\mathbf{z}}(k) - \tilde{\mathbf{x}}_n(k) = \tilde{\mathbf{e}}_{\mathrm{d},n}(k) + \tilde{\mathbf{e}}_{\mathrm{r},n}(k), \tag{23}$$

where

$$\begin{aligned} \tilde{\mathbf{e}}_{\mathrm{d},n}(k) &\triangleq \tilde{\mathbf{x}}_{\mathrm{fd},n}(k) - \tilde{\mathbf{x}}_n(k) \\ &= \tilde{\mathbf{x}}_{\mathrm{fd},n}(k) - \tilde{\mathbf{x}}(k-n) \\ &= (\mathbf{H}_n\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} - \mathbf{I}_M)\tilde{\mathbf{x}}(k-n) \end{aligned} \tag{24}$$

is the signal distortion due to the rectangular filtering matrix with $\mathbf{I}_M$ being the $M \times M$ identity matrix and

$$\begin{aligned} \tilde{\mathbf{e}}_{\mathrm{r},n}(k) &\triangleq \tilde{\mathbf{x}}_{\mathrm{ri},n}(k) + \tilde{\mathbf{v}}_{\mathrm{rn},n}(k) \\ &= \mathbf{H}_n\mathbf{x}_{\mathrm{i},n}(k) + \mathbf{H}_n\mathbf{v}(k) \end{aligned} \tag{25}$$

represents the residual interference plus noise.

Having defined the error signal, we can now write the mean-square error (MSE) criterion:

$$\begin{aligned} J(\mathbf{H}_n) &\triangleq \frac{1}{M} \cdot \mathrm{tr}\{E[\tilde{\mathbf{e}}_n(k)\tilde{\mathbf{e}}_n^T(k)]\} \\ &= \frac{1}{M}\left[\mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{x}}_n}) + \mathrm{tr}(\mathbf{H}_n\mathbf{R}_\mathbf{y}\mathbf{H}_n^T) - 2\mathrm{tr}(\mathbf{H}_n\mathbf{R}_{\mathbf{y}\tilde{\mathbf{x}}_n})\right] \\ &= \frac{1}{M}\left[\mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{x}}_n}) + \mathrm{tr}(\mathbf{H}_n\mathbf{R}_\mathbf{y}\mathbf{H}_n^T) - 2\mathrm{tr}(\mathbf{H}_n\mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n})\right], \end{aligned} \tag{26}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a square matrix, $\mathbf{R}_\mathbf{y} \triangleq E[\mathbf{y}(k)\mathbf{y}^T(k)]$ is the correlation matrix of $\mathbf{y}(k)$, and

$$\begin{aligned} \mathbf{R}_{\mathbf{y}\tilde{\mathbf{x}}_n} &\triangleq E[\mathbf{y}(k)\tilde{\mathbf{x}}_n^T(k)] \\ &= E[\mathbf{x}(k)\tilde{\mathbf{x}}_n^T(k)] \\ &= \mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n} \end{aligned} \tag{27}$$

is the cross-correlation matrix between $\mathbf{y}(k)$ and $\tilde{\mathbf{x}}_n(k)$.

Using the fact that $E[\tilde{\mathbf{e}}_{\mathrm{d},n}(k)\tilde{\mathbf{e}}_{\mathrm{r},n}^T(k)] = \mathbf{0}_{M \times M}$, $J(\mathbf{H})$ can be expressed as the sum of two other MSEs, i.e.,

$$J(\mathbf{H}_n) = J_{\mathrm{d}}(\mathbf{H}_n) + J_{\mathrm{r}}(\mathbf{H}_n), \tag{28}$$

where

$$J_{\mathrm{d}}(\mathbf{H}_n) \triangleq \frac{1}{M} \cdot \mathrm{tr}\{E[\tilde{\mathbf{e}}_{\mathrm{d},n}(k)\tilde{\mathbf{e}}_{\mathrm{d},n}^T(k)]\}, \tag{29}$$

$$J_{\mathrm{r}}(\mathbf{H}_n) \triangleq \frac{1}{M} \cdot \mathrm{tr}\{E[\tilde{\mathbf{e}}_{\mathrm{r},n}(k)\tilde{\mathbf{e}}_{\mathrm{r},n}^T(k)]\}. \tag{30}$$

With these MSE criteria, we are now ready to derive different optimal filtering matrices. However, before leaving this section and moving to the next one on optimal filtering matrices, let us first present a scheme that can jointly diagonalize the three symmetric matrices $\mathbf{R}_\mathbf{y}$, $\mathbf{R}_{\mathbf{x}_{\mathrm{d},n}}$, and $\mathbf{R}_{\mathrm{in}}$, where $\mathbf{R}_{\mathrm{in}} \triangleq \mathbf{R}_{\mathbf{x}_{\mathrm{i},n}} + \mathbf{R}_\mathbf{v}$. This diagonalization will be used later on to analyze some optimal filtering matrices.

From the signal model given in Eq. (2) and the signal decomposition in Eq. (11), it is easy to check that $\mathbf{R_y} = \mathbf{R_{x_{d,n}}} + \mathbf{R}_{in}$. The two matrices $\mathbf{R_{x_{d,n}}}$ and $\mathbf{R}_{in}$ can be jointly diagonalized as follows (Searle, 1982; Strang, 1988):

$$\mathbf{B}^T \mathbf{R_{x_{d,n}}} \mathbf{B} = \mathbf{\Lambda}, \tag{31}$$

$$\mathbf{B}^T \mathbf{R}_{in} \mathbf{B} = \mathbf{I}_L, \tag{32}$$

where $\mathbf{B}$ is a full-rank square matrix (of size $L \times L$) and $\mathbf{\Lambda}$ is a diagonal matrix the main elements of which are real and nonnegative. Furthermore, $\mathbf{\Lambda}$ and $\mathbf{B}$ are the eigenvalue and eigenvector matrices, respectively, of $\mathbf{R}_{in}^{-1} \mathbf{R_{x_{d,n}}}$, i.e.,

$$\mathbf{R}_{in}^{-1} \mathbf{R_{x_{d,n}}} \mathbf{B} = \mathbf{B}\mathbf{\Lambda}. \tag{33}$$

Because the rank of the matrix $\mathbf{R_{x_{d,n}}}$ is equal to M, the eigenvalues of $\mathbf{R}_{in}^{-1} \mathbf{R_{x_{d,n}}}$ can be ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M > \lambda_{M+1} = \cdots = \lambda_L = 0$. In other words, the last $L - M$ eigenvalues of $\mathbf{R}_{in}^{-1} \mathbf{R_{x_{d,n}}}$ are equal to zero while its first M eigenvalues are positive, with λ_1 being the maximum eigenvalue. We also denote the corresponding eigenvectors by $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_M, \mathbf{b}_{M+1}, ..., \mathbf{b}_L$. With the preceding decomposition, it is easy to check that

$$\mathbf{B}^T \mathbf{R_y} \mathbf{B} = \mathbf{\Lambda} + \mathbf{I}_L. \tag{34}$$

Therefore the three matrices are simultaneously diagonalized. Note that the preceding diagonalization was proposed and used in Jensen *et al*. (1995) and Hu and Loizou (2002) and Hu and Loizou (2003) but for the classical subspace approach (Ephraim and Trees, 1995).

III. OPTIMAL RECTANGULAR FILTERING MATRICES

In this section, we are going to derive several important filtering matrices that can help reduce the noise picked up by the microphone signal.

A. Maximum SNR

First, let us derive a filter that can maximize the output SNR. From the signal model given in Eqs. (1) or (1), the input SNR is defined as

$$iSNR \triangleq \frac{\sigma_x^2}{\sigma_v^2} = \frac{tr(\mathbf{R_x})}{tr(\mathbf{R_v})}, \tag{35}$$

where $\sigma_x^2 \triangleq E[x^2(k)]$ and $\sigma_v^2 \triangleq E[v^2(k)]$ are the variances of the desired clean speech and noise signals, respectively. Now with the use of the decomposition of the output signal given in Eq. (16), we can write the output SNR as

$$oSNR(\mathbf{H}_n) = \frac{tr(\mathbf{R_{\tilde{x}_{fd,n}}})}{tr(\mathbf{R_{\tilde{x}_{ri,n}}} + \mathbf{R_{\tilde{v}_{rn,n}}})}$$

$$= \frac{\sum_{m=1}^{M} \mathbf{h}_{n,m}^T \mathbf{R_{x_{d,n}}} \mathbf{h}_{n,m}}{\sum_{m=1}^{M} \mathbf{h}_{n,m}^T \mathbf{R}_{in} \mathbf{h}_{n,m}}. \tag{36}$$

(Note that the interference should be treated as part of the residual noise because it is uncorrelated with desired speech

component.) It is straightforward to see that the maximum SNR filter is obtained by maximizing Eq. (36). To find such a filter, let us first give the following lemma.

Lemma I. With the filtering matrix $\mathbf{H}_n$, the output SNR satisfies

$$oSNR(\mathbf{H}_n) \leq \max_m \frac{\mathbf{h}_{n,m}^T \mathbf{R_{x_{d,n}}} \mathbf{h}_{n,m}}{\mathbf{h}_{n,m}^T \mathbf{R}_{in} \mathbf{h}_{n,m}} \triangleq \chi. \tag{37}$$

Proof. Let us define the following real and positive coefficients: $a_m \triangleq \mathbf{h}_{n,m}^T \mathbf{R_{x_{d,n}}} \mathbf{h}_{n,m}$ and $b_m \triangleq \mathbf{h}_{n,m}^T \mathbf{R}_{in} \mathbf{h}_{n,m}$. Then, we have

$$\frac{\sum_{m=1}^{M} a_m}{\sum_{m=1}^{M} b_m} = \sum_{m=1}^{M} \left(\frac{a_m}{b_m} \cdot \frac{b_m}{\sum_{i=1}^{M} b_i} \right). \tag{38}$$

Now, let us define the following two vectors:

$$\mathbf{u} \triangleq \left[\frac{a_1}{b_1} \frac{a_2}{b_2} \cdots \frac{a_M}{b_M} \right]^T, \tag{39}$$

$$\mathbf{u}' \triangleq \left[\frac{b_1}{\sum_{i=1}^{M} b_i} \frac{b_2}{\sum_{i=1}^{M} b_i} \cdots \frac{b_M}{\sum_{i=1}^{M} b_i} \right]^T. \tag{40}$$

Using the Holder's inequality, we have

$$\frac{\sum_{m=1}^{M} a_m}{\sum_{m=1}^{M} b_m} = \mathbf{u}^T \mathbf{u}' \leq \|\mathbf{u}\|_\infty \|\mathbf{u}'\|_1 = \max_m \frac{a_m}{b_m}, \tag{41}$$

where $\|\cdot\|_\infty$ and $\|\cdot\|_1$ denote, respectively, the ℓ_∞ and ℓ_1 norm. Then it follows immediately that the inequality given in Eq. (37) holds, which completes the proof of the lemma. Based on the previous Lemma, we can give the following theorem.

Theorem I. The maximum SNR filtering matrix is given by

$$\mathbf{H}_{max,n} = \begin{bmatrix} \beta_1 \mathbf{b}_1^T \\ \beta_2 \mathbf{b}_1^T \\ \vdots \\ \beta_m \mathbf{b}_1^T \end{bmatrix}, \tag{42}$$

where $\beta_m, m = 1, 2, ..., M$ are real numbers with at least one of them different from 0, and $\mathbf{b}_1$ is the eigenvector of the matrix $\mathbf{R}_{in}^{-1} \mathbf{R_{x_d}}$ corresponding to the maximum eigenvalue λ_1. The corresponding output SNR is

$$oSNR(\mathbf{H}_{max,n}) = \lambda_1. \tag{43}$$

Proof. From Lemma I, we know that the output SNR is upper bounded by χ the maximum value of which is clearly λ_1. On the other hand, it can be checked from Eq. (43) that $oSNR(\mathbf{H}_{max,n}) = \lambda_1$. Because this output SNR is maximal, it

is straightforward to see that $\mathbf{H}_{\text{max}}$ is the maximum SNR filtering matrix.

Now, we can check that the maximum SNR filtering matrix has the following property.

Property I. The output SNR with the maximum SNR filtering matrix is always greater than or equal to the input SNR, i.e., $\text{oSNR}(\mathbf{H}_{\text{max},n}) \geq \text{iSNR}$.

This property can be easily verified, which will not be presented here.

From the preceding analysis, we can see that for a fixed L, increasing the value of M (from 1 to L) will, in general, increase the output SNR of the maximum SNR filtering matrix because more and more information is taken into account. However, we should expect that the amount of speech distortion due to the maximum SNR filter would also increase significantly as M is increased.

B. Wiener

If we differentiate the MSE criterion, $J(\mathbf{H}_n)$, defined in Eq. (26), with respect to $\mathbf{H}_n$ and equate the result to zero, we find the Wiener filtering matrix:

$$\mathbf{H}_{\text{W},n} = \mathbf{R}_{\mathbf{y}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1}. \tag{44}$$

Using the identity filtering matrix $\mathbf{I}_{\text{i}} = [\mathbf{I}_M \; \mathbf{0}_{M \times (L-M)}]$, we can rewrite the Wiener filtering matrix as

$$\mathbf{H}_{\text{W},n} = \mathbf{I}_{\text{i}} \mathbf{R}_{\mathbf{x}} \mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{I}_{\text{i}}(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}} \mathbf{R}_{\mathbf{y}}^{-1}). \tag{45}$$

Because

$$\mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n} = \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} \mathbf{R}_{\tilde{\mathbf{x}}_n}, \tag{46}$$

we can rewrite Eq. (44) as

$$\mathbf{H}_{\text{W},n} = \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1}. \tag{47}$$

The correlation matrix of $\mathbf{y}(k)$ can be written as

$$\mathbf{R}_{\mathbf{y}} = \mathbf{R}_{\mathbf{x}_{\text{d},n}} + \mathbf{R}_{\text{in}} = \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T + \mathbf{R}_{\text{in}}. \tag{48}$$

Determining the inverse of $\mathbf{R}_{\mathbf{y}}$ from Eq. (48) with the Woodbury's identity gives

$$\mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{R}_{\text{in}}^{-1} + \mathbf{R}_{\text{in}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} (\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1} + \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1}$$
$$\times \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1}. \tag{49}$$

Now, substituting Eq. (49) into Eq. (47), we get another interesting formulation of the Wiener filtering matrix:

$$\mathbf{H}_{\text{W},n} = (\mathbf{I}_M + \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1}$$
$$= (\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1} + \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1}. \tag{50}$$

C. MVDR

The celebrated MVDR approach, requiring no distortion to the desired signal, is usually derived in the multichannel case. Interestingly, with the new block-based framework, we can also derive the MVDR in the single-channel case, just like in Chen *et al.* (2011) and Benesty and Chen (2011). The corresponding rectangular filtering matrix is obtained by minimizing the MSE of the residual interference-plus-noise, $J_{\text{r}}(\mathbf{H}_n)$, with the constraint that the desired signal is not distorted. Mathematically, this is equivalent to

$$\mathbf{H}_{\text{MVDR},n} = \arg \min_{\mathbf{H}_n} \frac{1}{M} \cdot \text{tr}(\mathbf{H}_n \mathbf{R}_{\text{in}} \mathbf{H}_n^T)$$
$$\text{subject to } \mathbf{H}_n \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} = \mathbf{I}_M. \tag{51}$$

The solution to the preceding optimization problem is

$$\mathbf{H}_{\text{MVDR},n} = (\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\text{in}}^{-1}, \tag{52}$$

which is interesting to compare to $\mathbf{H}_{\text{W},n}$ in Eq. (50).

Obviously, by using the Woodbury's identity of $\mathbf{R}_{\mathbf{y}}^{-1}$ in Eq. (49) we can rewrite Eq. (52) as

$$\mathbf{H}_{\text{MVDR},n} = (\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1}. \tag{53}$$

From Eqs. (47) and (53), we deduce the relationship between the MVDR and Wiener filtering matrices:

$$\mathbf{H}_{\text{MVDR},n} = (\mathbf{H}_{\text{W},n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{H}_{\text{W},n}. \tag{54}$$

D. Prediction

Let $\mathbf{G}_n$ be a temporal prediction matrix of size $M \times L$ so that

$$\mathbf{x}(k) \approx \mathbf{G}_n^T \tilde{\mathbf{x}}_n(k). \tag{55}$$

A new distortionless filtering matrix for noise reduction based on preceding prediction can be derived by

$$\mathbf{H}_{\text{P},n} = \arg \min_{\mathbf{H}_n} \text{tr}(\mathbf{H}_n \mathbf{R}_{\mathbf{y}} \mathbf{H}_n^T)$$
$$\text{subject to } \mathbf{H}_n \mathbf{G}_n^T = \mathbf{I}_M, \tag{56}$$

from which we deduce the solution

$$\mathbf{H}_{\text{P},n} = (\mathbf{G}_n \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{G}_n^T)^{-1} \mathbf{G}_n \mathbf{R}_{\mathbf{y}}^{-1}. \tag{57}$$

The best way to find $\mathbf{G}_n$ is in the minimum MSE (MMSE) sense. Indeed, define the error signal vector:

$$\mathbf{e}_{\text{P},n}(k) \triangleq \mathbf{x}(k) - \mathbf{G}_n^T \tilde{\mathbf{x}}_n(k) \tag{58}$$

and form the MSE:

$$J(\mathbf{G}_n) \triangleq E[\mathbf{e}_{\text{P},n}^T(k) \mathbf{e}_{\text{P},n}(k)]. \tag{59}$$

The minimization of $J(\mathbf{G}_n)$ with respect to $\mathbf{G}_n$ leads to

$$\mathbf{G}_{n,\text{o}} = \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T, \tag{60}$$

and substituting this result into Eq. (57) gives

$$\mathbf{H}_{\mathrm{P},n} = (\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathbf{y}}^{-1}, \tag{61}$$

which is identical to the MVDR.

It is interesting to observe that the error signal vector with the optimal matrix, $\mathbf{G}_{n,\mathrm{o}}$, corresponds to the interference signal vector, i.e.,

$$\mathbf{e}_{\mathrm{P},n,\mathrm{o}}(k) = \mathbf{x}(k) - \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} \tilde{\mathbf{x}}_n(k) = \mathbf{x}_{\mathrm{i},n}(k). \tag{62}$$

This result is a consequence of the orthogonality principle.

E. Tradeoff

In the tradeoff approach, we minimize the speech distortion index with the constraint that the noise reduction factor is equal to a positive value that is greater than 1. Mathematically, this is equivalent to

$$\min_{\mathbf{H}_n} J_{\mathrm{d}}(\mathbf{H}_n) \quad \text{subject to} \quad J_{\mathrm{r}}(\mathbf{H}_n) = \beta J_{\mathrm{r}}(\mathbf{I}_{\mathrm{i}}), \tag{63}$$

where $0 < \beta < 1$ to ensure that we get some noise reduction. By using a Lagrange multiplier, $\mu > 0$, to adjoin the constraint to the cost function and assuming that the matrix $\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T + \mu \mathbf{R}_{\mathrm{in}}$ is invertible, we easily deduce the tradeoff filtering matrix:

$$\mathbf{H}_{\mathrm{T},\mu,n} = \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T (\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n} \mathbf{R}_{\tilde{\mathbf{x}}_n} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T + \mu \mathbf{R}_{\mathrm{in}})^{-1}, \tag{64}$$

which can be rewritten, thanks to the Woodbury's identity, as

$$\mathbf{H}_{\mathrm{T},\mu,n} = (\mu \mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1} + \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathrm{in}}^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n})^{-1} \mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T \mathbf{R}_{\mathrm{in}}^{-1}, \tag{65}$$

where μ satisfies $J_{\mathrm{r}}(\mathbf{H}_{\mathrm{T},\mu,n}) = \beta J_{\mathrm{r}}(\mathbf{I}_{\mathrm{i}})$. Usually, μ is chosen in a heuristic way, so that we have the following cases:

(a) When $\mu = 1$, we have $\mathbf{H}_{\mathrm{T},1,n} = \mathbf{H}_{\mathrm{W},n}$, which is the Wiener filtering matrix.
(b) When $\mu = 0$ [from Eq. (65)], we get $\mathbf{H}_{\mathrm{T},0,n} = \mathbf{H}_{\mathrm{MVDR},n}$, which is the MVDR filtering matrix.
(c) If $\mu > 1$, the tradeoff filter tends to attenuate more noise as compared to Wiener; but this is achieved at the expense of higher speech distortion.
(d) If $\mu < 1$, the tradeoff filter tends to have less noise reduction with also lower distortion as compared to Wiener.

For the tradeoff filtering matrix, we have the following property.

Property II. The output SNR with the tradeoff filtering matrix is always greater than or equal to the input SNR, i.e., $\mathrm{oSNR}(\mathbf{H}_{\mathrm{T},\mu,n}) \geq \mathrm{iSNR}, \forall \mu \geq 0$.

This property can be proved by induction, which will not be presented here.

Comparing the output SNRs of the MVDR, Wiener, and tradeoff filtering matrices, we have the following inequalities.

(a) If $\mu \geq 1$,

$$\mathrm{oSNR}(\mathbf{H}_{\mathrm{MVDR},n}) \leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{W},n}) \leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{T},\mu,n})$$
$$\leq \mathrm{oSNR}(\mathbf{H}_{\max,n}). \tag{66}$$

(b) If $\mu \leq 1$,

$$\mathrm{oSNR}(\mathbf{H}_{\mathrm{MVDR},n}) \leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{T},\mu,n}) \leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{W},n})$$
$$\leq \mathrm{oSNR}(\mathbf{H}_{\max,n}). \tag{67}$$

Using the joint diagonalization given from Eqs. (31) to (34), we can write the tradeoff filtering matrix into the so-called subspace form. Indeed, from Eq. (64), we get

$$\mathbf{H}_{\mathrm{T},\mu,n} = \mathbf{T} \begin{bmatrix} \mathbf{\Sigma}_\mu & \mathbf{0}_{M \times (L-M)} \\ \mathbf{0}_{(L-M) \times M} & \mathbf{0}_{(L-M) \times (L-M)} \end{bmatrix} \mathbf{B}^T, \tag{68}$$

where

$$\mathbf{T} = \mathbf{I}_{\mathrm{i}} \mathbf{B}^{-T} \tag{69}$$

and

$$\mathbf{\Sigma}_\mu = \mathrm{diag}\left(\frac{\lambda_1}{\lambda_1 + \mu}, \frac{\lambda_2}{\lambda_2 + \mu}, \dots, \frac{\lambda_M}{\lambda_M + \mu} \right) \tag{70}$$

is an $M \times M$ diagonal matrix. Expression (68) can also be written as

$$\mathbf{H}_{\mathrm{T},\mu,n} = \mathbf{I}_{\mathrm{i}} \mathbf{M}_{\mathrm{T},\mu}, \tag{71}$$

where

$$\mathbf{M}_{\mathrm{T},\mu} = \mathbf{B}^{-T} \begin{bmatrix} \mathbf{\Sigma}_\mu & \mathbf{0}_{M \times (L-M)} \\ \mathbf{0}_{(L-M) \times M} & \mathbf{0}_{(L-M) \times (L-M)} \end{bmatrix} \mathbf{B}^T. \tag{72}$$

So, the tradeoff filter $\mathbf{H}_{\mathrm{T},\mu,n}$ is the product of two matrices: The rectangular identity matrix and an adjustable square matrix of size $L \times L$ the rank of which is equal to M.

Note that $\mathbf{H}_{\mathrm{T},\mu,n}$ as given in Eq. (68) is not, in theory, valid for $\mu = 0$ as this expression was derived from Eq. (64), which is clearly not defined for this particular case. In practice, however, it is possible to set $\mu = 0$ in Eq. (68); but this does not lead to the MVDR filter.

F. Particular case: M = L

For $M = L$, the parameter n can only be 0 and the rectangular matrix $\mathbf{H}_n$ becomes a square matrix of size $L \times L$. To distinguish this particular scenario from the previous general case where $\mathbf{H}_n$ is a rectangular matrix, let us denote $\mathbf{H}_n = \mathbf{H}_0$ as $\mathbf{H}_{\mathrm{S}}$ in this situation. It can be verified, in this case, that $\mathbf{x}_{\mathrm{i},n}(k) = \mathbf{x}_{\mathrm{i}}(k) = \mathbf{0}_{L \times 1}$ and, as a result, $\mathbf{R}_{\mathrm{in}} = \mathbf{R}_{\mathrm{v}}$, $\mathbf{R}_{\mathbf{x}_{\mathrm{i},n}} = \mathbf{R}_{\mathbf{x}_{\mathrm{i}}} = \mathbf{0}_{L \times L}$, and $\mathbf{R}_{\mathbf{x}_{\mathrm{d},n}} = \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} = \mathbf{R}_{\mathbf{x}}$. Given these conditions, it can be checked that

$$\mathbf{H}_{\mathrm{S,max}} = \begin{bmatrix} \beta_1 \mathbf{b}_1^T \\ \beta_2 \mathbf{b}_1^T \\ \vdots \\ \beta_L \mathbf{b}_1^T \end{bmatrix}, \tag{73}$$

J. Acoust. Soc. Am., Vol. 133, No. 2, February 2013

Long *et al.*: Single-channel noise reduction   1095

$$\mathbf{H}_{\mathrm{S,W}} = \mathbf{R}_\mathbf{x}\mathbf{R}_\mathbf{y}^{-1} = \mathbf{I}_L - \mathbf{R}_\mathbf{v}\mathbf{R}_\mathbf{y}^{-1}, \tag{74}$$

$$\mathbf{H}_{\mathrm{S,MVDR}} = \mathbf{I}_L, \tag{75}$$

$$\begin{aligned}\mathbf{H}_{\mathrm{S,T},\mu} &= \mathbf{R}_\mathbf{x}(\mathbf{R}_\mathbf{x} + \mu\mathbf{R}_\mathbf{v})^{-1} \\ &= (\mathbf{R}_\mathbf{y} - \mathbf{R}_\mathbf{v})[\mathbf{R}_\mathbf{y} + (\mu - 1)\mathbf{R}_\mathbf{v}]^{-1},\end{aligned} \tag{76}$$

where $\mathbf{b}_1$, a vector of length L, is the eigenvector corresponding to the maximum eigenvalue, λ_1, of the matrix $\mathbf{R}_\mathbf{v}^{-1}\mathbf{R}_\mathbf{x}$. In this case, the MVDR is the identity matrix and all the studied optimal filtering matrices are very different. It can be shown that

(a) for $\mu \geq 1$,

$$\begin{aligned}\mathrm{iSNR} = \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,MVDR}}) &\leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,W}}) \\ &\leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,T},\mu}) \leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,max}}) = \lambda_1;\end{aligned} \tag{77}$$

(b) for $0 \leq \mu \leq 1$,

$$\begin{aligned}\mathrm{iSNR} = \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,MVDR}}) &\leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,T},\mu}) \\ &\leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,W}}) \leq \mathrm{oSNR}(\mathbf{H}_{\mathrm{S,max}}) = \lambda_1.\end{aligned} \tag{78}$$

Applying the joint diagonalization given from Eqs. (31) to (34) to Eq. (76), we get

$$\mathbf{H}_{\mathrm{S,T},\mu} = \mathbf{B}^{-T}\mathbf{\Lambda}(\mathbf{\Lambda} + \mu\mathbf{I}_L)^{-1}\mathbf{B}^T, \tag{79}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_L)$ and $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_L]$ are the two matrices that consist of, respectively, the eigenvalues and eigenvectors of the matrix $\mathbf{R}_\mathbf{v}^{-1}\mathbf{R}_\mathbf{x}$. It is believed that a speech signal can be modeled as a linear combination of a number (smaller than the dimension of the signal vector) of some linearly independent basis vectors (Ephraim and Trees, 1995; Jensen *et al.*, 1995; Dendrinos *et al.*, 1991; Hu and Loizou, 2003; Jabloun and Champagne, 2005; Hermus *et al.*, 2007). As a result, the vector space of the noisy signal can be decomposed in two subspaces: The signal-plus-noise subspace of length L_{s} and the identity subspace of length L_{n}, with $L = L_{\mathrm{s}} + L_{\mathrm{n}}$. This implies that the last L_{n} eigenvalues of the matrix $\mathbf{R}_\mathbf{v}^{-1}\mathbf{R}_\mathbf{x}$ are equal to zero. Therefore we can rewrite Eq. (79) to obtain the subspace-type filtering matrix:

$$\mathbf{H}_{\mathrm{S,T},\mu} = \mathbf{B}^{-T}\begin{bmatrix} \mathbf{\Sigma}_\mu & \mathbf{0}_{L_{\mathrm{s}} \times L_{\mathrm{n}}} \\ \mathbf{0}_{L_{\mathrm{n}} \times L_{\mathrm{s}}} & \mathbf{0}_{L_{\mathrm{n}} \times L_{\mathrm{n}}} \end{bmatrix}\mathbf{B}^T, \tag{80}$$

where now

$$\mathbf{\Sigma}_\mu = \mathrm{diag}\left(\frac{\lambda_1}{\lambda_1 + \mu}, \frac{\lambda_2}{\lambda_2 + \mu}, \ldots, \frac{\lambda_{L_{\mathrm{s}}}}{\lambda_{L_{\mathrm{s}}} + \mu}\right) \tag{81}$$

is an $L_{\mathrm{s}} \times L_{\mathrm{s}}$ diagonal matrix. This algorithm is often referred to as the generalized subspace approach. One should note, however, that there is no noise-only subspace with this formulation. Therefore noise reduction can only be achieved by modifying the speech-plus-noise subspace by setting μ to a positive number.

G. Summary

Note that it can be shown that all the optimal filtering matrices derived in this section are related to the following expression:

$$\mathbf{H}_{\mathrm{o}} = \mathbf{A}_{\mathrm{o}}\mathbf{\Gamma}_{\mathbf{x}\tilde{\mathbf{x}}_n}^T\mathbf{R}_{\mathrm{in}}^{-1}, \tag{82}$$

where $\mathbf{A}_{\mathrm{o}}$ is a square matrix of size $M \times M$. Depending on how we choose $\mathbf{A}_{\mathrm{o}}$, we can easily obtain the different optimal filtering matrices given previously. In other words, these optimal filtering matrices derived before are equivalent up to the matrix $\mathbf{A}_{\mathrm{o}}$.

IV. EXPERIMENTAL RESULTS

A. Performance evaluation with known signal statistics

The clean speech signal used in the experiments 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 s. Noise is added into the clean speech to control the input SNR.

The implementation of the noise reduction filtering matrices derived in Sec. III requires the estimation of the correlation matrices $\mathbf{R}_\mathbf{y}$, $\mathbf{R}_\mathbf{x}$, $\mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n}$, and $\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1}$. We compute all these matrices directly from the corresponding signals so that we can put our focus on illustrating the noise reduction performance of the proposed methods with different values of the parameters. Specifically, at each frame, the matrix $\mathbf{R}_\mathbf{y}$ is computed using the most recent 600 samples of the noisy speech signal with a short-time average, and the matrices $\mathbf{R}_\mathbf{x}$, $\mathbf{R}_{\mathbf{x}\tilde{\mathbf{x}}_n}$, and $\mathbf{R}_{\tilde{\mathbf{x}}_n}^{-1}$ are computed using a same short-time average but from the clean speech signal.

To evaluate the amount of noise reduction, the output SNR, as defined in Eq. (36), is adopted as the objective performance measure. The higher is the value of $\mathrm{oSNR}(\mathbf{H}_n)$, the more the noise is reduced.

We also evaluate the amount of speech distortion using the speech distortion index (Chen *et al.*, 2011; Benesty and Chen, 2011):

$$\upsilon_{\mathrm{sd}}(\mathbf{H}_n) = \frac{E[\tilde{\mathbf{e}}_{\mathrm{d},n}(k)\tilde{\mathbf{e}}_{\mathrm{d},n}^T(k)]}{\mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{x}}_n})}. \tag{83}$$

The speech distortion index is always greater than or equal to 0 and should be upper bounded by 1 for optimal filtering matrices; so the higher is the value of $\upsilon_{\mathrm{sd}}(\mathbf{H}_n)$, the more the desired signal is distorted. Both the output SNR and speech distortion index are computed based on the 30-s long processed signals using a long-time average.

The first experiment investigates the influence of the filter length L on the noise reduction performance. White Gaussian noise is used and the input SNR is 10 dB. We set the delay parameter n to 0. The results of this experiment are plotted in Fig. 1. Figure 1(a) shows that the output SNR of both the Wiener and MVDR filtering matrices first increases with L for all the studied values of M. The SNR improvement is then saturated and does not increase much with L. Therefore we need to choose an appropriately large value of L (e.g., $L \geq 30$) to achieve a reasonable amount of SNR improvement in practice; but it should not be too large because this can significantly increase the computational complexity without leading to much performance
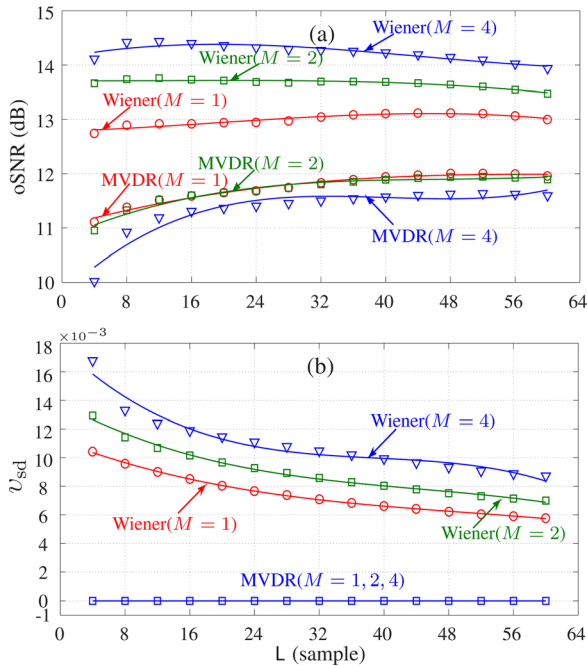
FIG. 1. (Color online) Performance of the Wiener and MVDR filtering matrices as a function of the filter length L in white Gaussian noise: (a) output SNR and (b) speech distortion index. The input SNR is 10 dB, $M = 1, 2$, and 4, and $n = 0$.

FIG. 2. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the block size M in white Gaussian noise: (a) output SNR and (b) speech distortion index. The input SNR is 10 dB, $L = 44$, and $n = 0$.

improvement. Figure 1(b) shows that the speech distortion index of the Wiener filtering matrix decreases linearly with L, while such index of the MVDR filtering matrix is approximately 0 for all the different values of M and L.

The second experiment evaluates the noise reduction performance as a function of the block size M. Again, white Gaussian noise is used, the input SNR is 10 dB, and the delay parameter n is set to 0. Based on the previous experiment, we set L to 44. The results for this experiment are plotted in Fig. 2. One can see from Fig. 2(a) that the output SNR of the Wiener and tradeoff (with $\mu = 0.5$ and $\mu = 1.5$) filtering matrices increases quickly as M increases up to 8, and then continues to increase but with a slower rate, while the output SNR of the MVDR filtering matrix decreases as M increases. The speech distortion index of the Wiener and tradeoff filtering matrices increases with M, while the speech distortion index of the MVDR filtering matrix is always approximately 0. One can see that the MVDR filtering matrix achieves less noise reduction as compared to the Wiener filtering matrix, but it does not introduce speech distortion, which is a strong advantage in practice.

The third experiment assesses the noise reduction performance of the different filtering matrices as a function of the input SNR, i.e., iSNR. Again, we only consider the case $n = 0$. Based on the previous experiments, we set $M = 4$ and $L = 44$. The results are sketched in Fig. 3. One can observe from Fig. 3(a) that the output SNR of all the studied three filtering matrices increases linearly with iSNR. In comparison, the output SNR of the MVDR filtering matrix increases faster than that of the Wiener and tradeoff filters as the input SNR increases. The speech distortion index of the Wiener and tradeoff filtering matrices decreases quickly first and slowly approaches to zero. As compared with the Wiener filtering matrix, one can clearly see that the tradeoff filter with $\mu = 0.5$
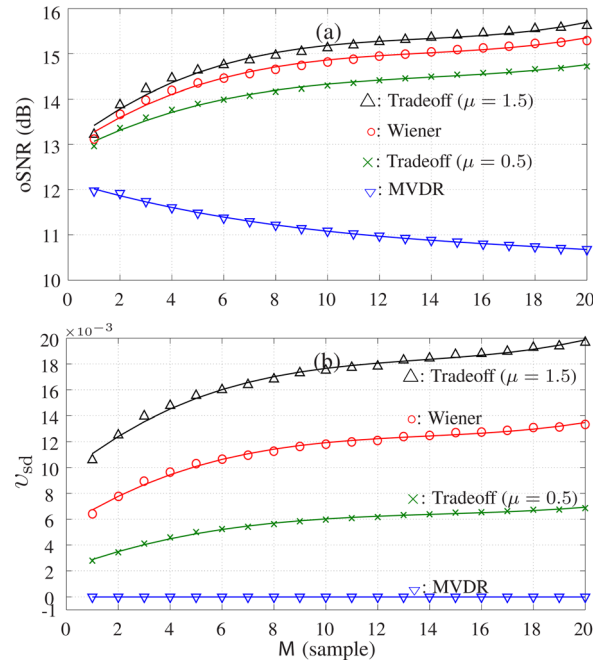
leads to a smaller SNR improvement but also introduces less speech distortion, while with $\mu = 1.5$ leads to a larger SNR improvement but with more speech distortion.

The fourth experiment studies the effect of the delay parameter n on the noise reduction performance of the different optimal filtering matrices. Figure 4 presents the results for the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering
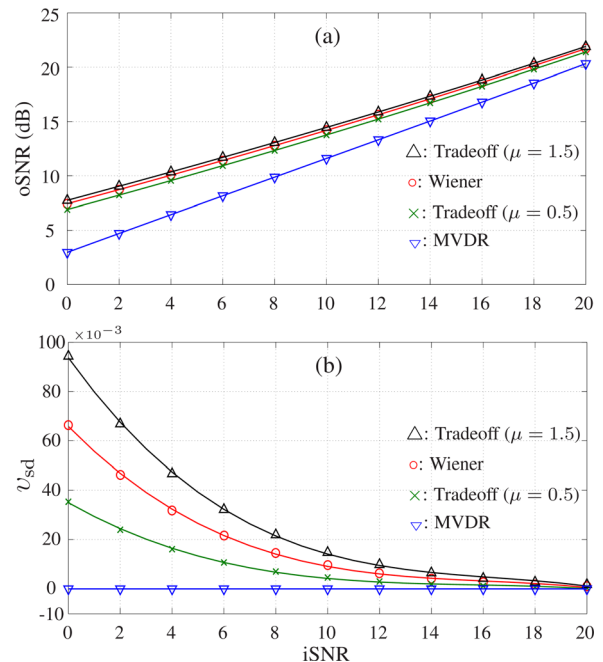

FIG. 3. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the input SNR in white Gaussian noise: (a) output SNR and (b) speech distortion index. $M = 4$, $L = 44$, and $n = 0$.
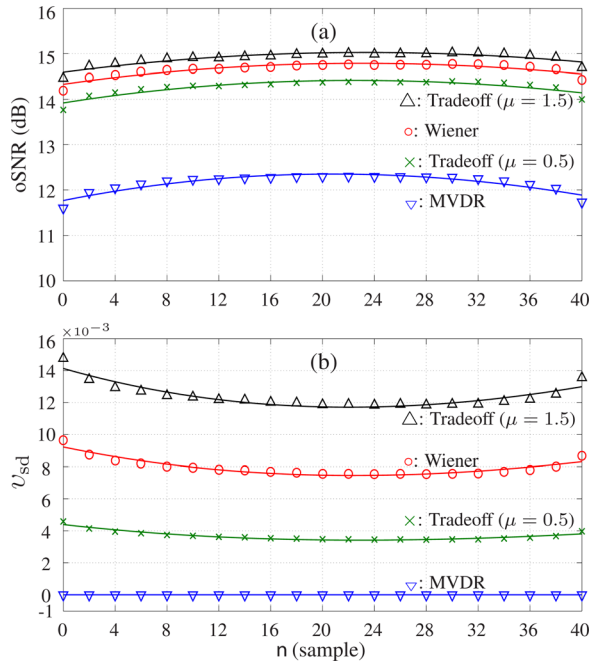
FIG. 4. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the parameter n in white Gaussian noise: (a) output SNR and (b) speech distortion index. The input SNR is 10 dB, $M = 4$, and $L = 44$.
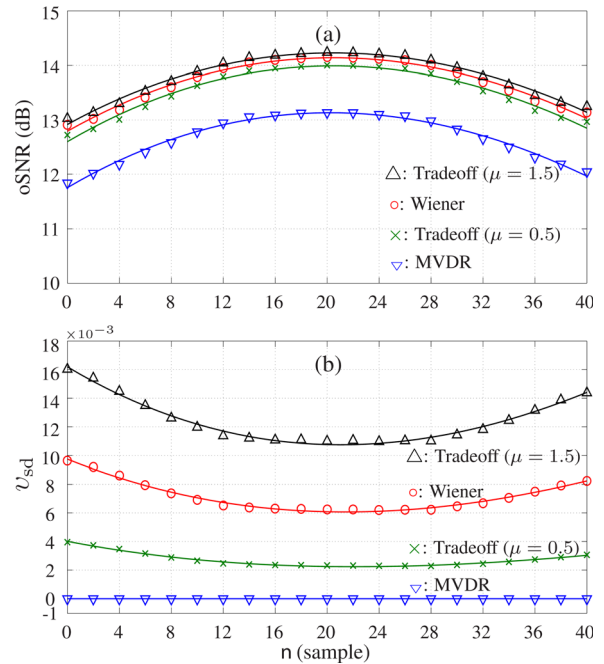

FIG. 5. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the parameter n in car noise: (a) output SNR and (b) speech distortion index. The input SNR is 10 dB, $M = 4$, and $L = 44$.

matrices with $M = 4$ and $L = 44$. It is seen from Fig. 4 that the output SNR first increases while the speech distortion decreases as the value of n is increased, after reaching an extremum, the output SNR decreases while the speech distortion increases; this indicates that allowing some flexibility on the non-causality can help improve the noise reduction performance. Interestingly, the best performance is achieved when n is approximately $(L - M)/2$. The underlying reason may be explained as follows. A speech sample is generally correlated with both its past and future samples. The correlation is stronger between closely neighboring samples than between far-distance samples. In general, the degree of such correlation seems to be approximately symmetric. In other words, given a lag time, the amount of correlation between the current and previous samples is similar to that between the current and future samples. As a result, when $n = (L - M)/2$, the speech self correlation is fully utilized during the filtering process, which leads to the best performance.

In the fifth and sixth experiments, instead of using white Gaussian noise, we use a car noise signal, which was recorded in a sedan car running at 50 miles/h on a highway. Figure 5 presents the effect of the delay parameter n on the noise reduction performance of the different optimal filtering matrices. Again, we set $M = 4$ and $L = 44$. One can see that the trend of the performance as a function of n is similar to that in Gaussian noise shown in Fig. 4, and the best performance is achieved when n is approximately $(L - M)/2$. In comparison with the performance in white Gaussian noise, we get less noise reduction with the same filtering matrix in car noise. Figure 6 presents the noise reduction results as a function of the input SNR, i.e., iSNR, where $M = 4$, $L = 44$, and $n = 20$ [according to $(L - M)/2$]. One can observe that the performance trend is similar to that shown in Fig. 3: The

output SNR of all the studied three filtering matrices increases linearly with the input SNR; the speech distortion index of the Wiener and tradeoff filtering matrices decreases quickly first and slowly approaches to zero while the distortion index for the MVDR filtering matrix is always zero.

In the seventh and eighth experiments, we consider a babble noise signal, which was recorded in a New York Stock
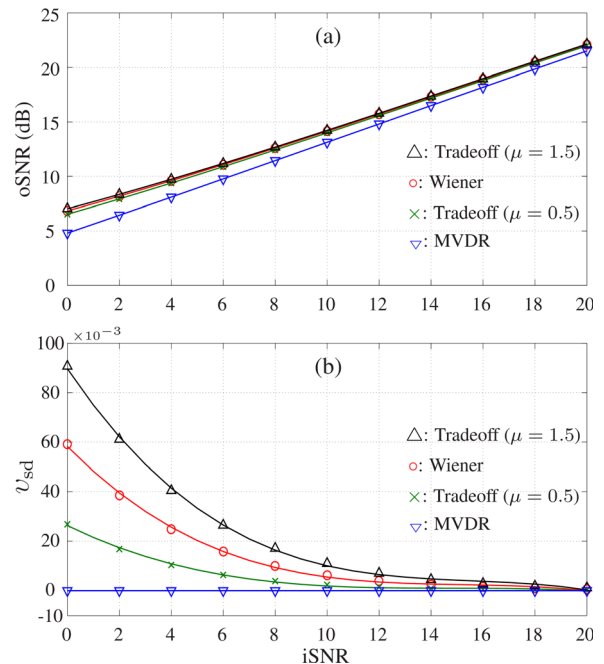

FIG. 6. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the input SNR in car noise: (a) output SNR and (b) speech distortion index. $M = 4$, $L = 44$, and $n = 20$.
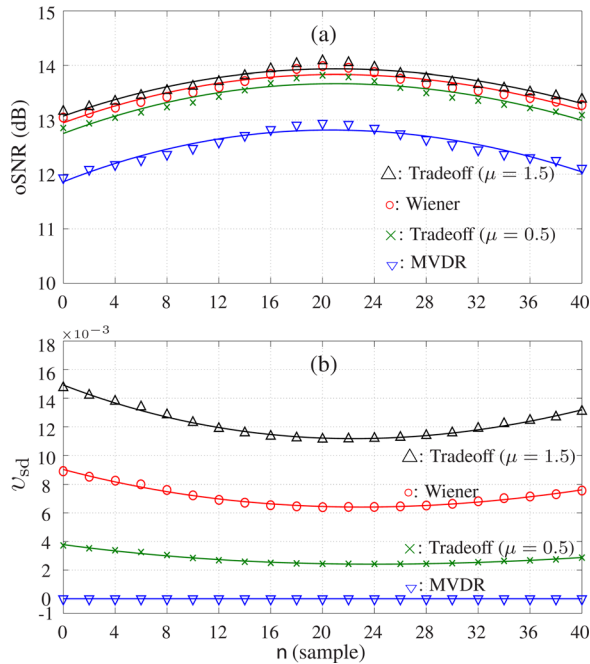
FIG. 7. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the parameter n in NYSE noise: (a) output SNR and (b) speech distortion index. The input SNR is 10 dB, $M = 4$, and $L = 44$.

Exchange (NYSE) room; it consists of sounds from various sources such as speakers, telephone rings, electric fans, etc. Figure 7 presents the effect of the delay parameter n on the noise reduction performance of different optimal filtering matrices, where $M = 4$ and $L = 44$. Once again, we observe that the performance changes with n in a similar way to that in white Gaussian and car noise conditions. Figure 8 presents the
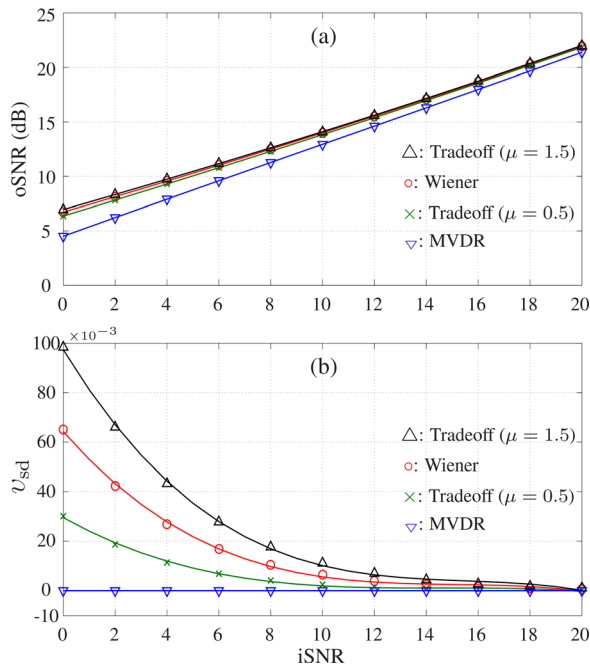

FIG. 8. (Color online) Performance of the Wiener, MVDR, and tradeoff ($\mu = 0.5$, $\mu = 1.5$) filtering matrices as a function of the input SNR in NYSE noise: (a) output SNR and (b) speech distortion index. $M = 4$, $L = 44$, and $n = 20$.

TABLE I. Computational complexity of the Wiener filter as a function of the filter length L and block size M.

Algorithm step	(Real-valued) multiplications
Estimation of $\mathbf{R_y}$ (with short time average)	$M \times L^2$
Estimation of $\mathbf{R_x} = \mathbf{R_y} - \mathbf{R_v}$ (with short time average)	$M \times L^2$
Computing $\mathbf{\Gamma_{x\tilde{x}_n}} = \mathbf{R_{x\tilde{x}_n}} \mathbf{R_{\tilde{x}_n}^{-1}}$ (Gauss-Jordan elimination)	$3 \times M^3 + M^2 \times L$
Computing $\mathbf{R_{in}} = \mathbf{R_y} - \mathbf{\Gamma_{x\tilde{x}_n}} \mathbf{R_{\tilde{x}_n}} \mathbf{\Gamma_{x\tilde{x}_n}^T}$	$M^2 \times L + L^2 \times M$
Computing $\mathbf{H_{W,n}} = (\mathbf{R_{\tilde{x}_n}^{-1}} + \mathbf{\Gamma_{x\tilde{x}_n}^T} \mathbf{R_{in}^{-1}} \mathbf{\Gamma_{x\tilde{x}_n}})^{-1}$ $\cdot \mathbf{\Gamma_{x\tilde{x}_n}^T} \mathbf{R_{in}^{-1}}$	$6 \times M^3 + 2 \times M^2 \times L$ $+ 2 \times L^2 \times M + 6 \times L^3$
Total/block:	$9 \times M^3 + 4 \times M^2 \times L$ $+ 5 \times L^2 \times M + 6 \times L^3$
Total/sample:	$9 \times M^2 + 4 \times M \times L$ $+ 5 \times L^2 + (6 \times L^3)/M$

noise reduction results as a function of the input SNR, i.e., iSNR, where $M = 4$, $L = 44$, and $n = 20$.

Note that all the experiments demonstrated that it is possible for the Wiener, MVDR, and tradeoff filtering matrices to gain more noise reduction by choosing a proper value of M that is greater than 1 as compared to the case with $M = 1$. This shows the advantage of using a filtering matrix over using a filtering vector developed in Benesty and Chen (2011) and Chen *et al.* (2011).

Another benefit of using a filtering matrix instead of a filtering vector is that the implementation can be made computationally more efficient. As a matter of fact, it can be easily checked that the complexity of the studied algorithms in this paper is a function of the filter length L and block size M. For example, the complexity of the Wiener filter in terms of the number of multiplications is summarized in Table I (the results of the MVDR and tradeoff filters are similar to that of the Wiener filter), where we assume that the correlation matrix is computed using a short time average and the matrix inversion is implemented using the Gauss–Jordan elimination method, which requires $3 \times L^3$ multiplications (Cormen *et al.*, 1990). Figure 9 plots the complexity of the
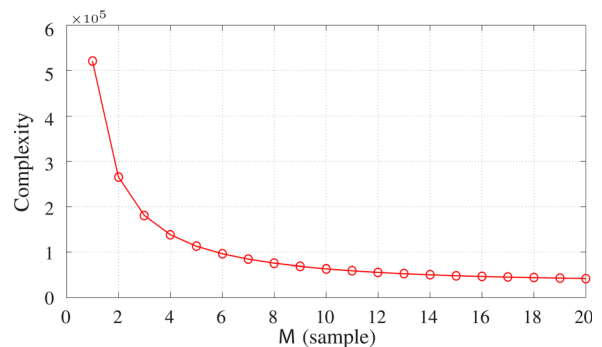

FIG. 9. (Color online) Complexity of the Wiener filter (with $L = 44$) as a function of the block size M.

Wiener filter as a function of the block size M for $L = 44$. It is clearly seen that the complexity of the Wiener filter decreases with M. This illustrates the advantage of the block-based approach ($M > 1$) over the sample-based method ($M = 1$) in terms of computational complexity.

B. Performance evaluation with estimated statistics

In the previous experiments, we assumed that the clean, noisy, and noise signals were accessible so that their correlation matrices can be directly computed from the corresponding signals. This assumption enabled us to properly study the impact of the different parameters on the noise reduction performance as well as the performance upper bounds of the different filtering matrices. In real-world applications, however, the statistics of the clean speech and noise have to be estimated because both the clean and noise signals are not accessible. In this section, we use experiments to evaluate the performance of the optimal rectangular filtering matrices deduced in Sec. III in a realistic application scenario where only the noisy signal is accessible. In this case, we typically need a voice activity detector (VAD) to detect the noise signal and compute the noise correlation matrix $\mathbf{R_v}$ from the estimated noise signal. Then an estimate of the clean speech correlation matrix $\mathbf{R_x}$ can be obtained by subtracting the noise correlation matrix from the noisy one. The matrices $\mathbf{R_{x\tilde{x}_n}}$ and $\mathbf{R_{\tilde{x}_n}}$ can be estimated in a similar way. As can be seen, the paramount issue in noise estimation in real applications is a VAD, which has been intensively studied in the literature. Many methods have been developed, and representative ones include the explicit speech/non-speech detection method, the minimum statistics approach (Martin, 2001), the histogram based method (Hirsch and Ehrlicher, 1995), the quantile-based method (Stahl et al., 2000), the statistical method with signal presence probability (Cohen, 2003), and the sequential estimation using single-pole recursion (Diethorn, 1997; Chen et al., 2006), etc. In this experiment, we adopt the VAD algorithm based on the so-called two-sided single-pole recursion developed in (Diethorn, 1997; Chen et al., 2006). This algorithm can achieve fairly accurate estimation of the noise signal and the noise statistics from the noisy speech and has been used in several practical noise reduction systems. For a detailed description of this VAD algorithm and its implementation, see Diethorn (1997); Chen et al. (2006).

As seen in the previous experiments, we have three different types of noise environments, i.e., white Gaussian, car, and NYSE noises. Due to space limitation, we only report the results in the car noise environment where the input SNR is 10 dB. Based on the previous study, we set $L = 44$ and vary the value of M from 1 to 20. The delay parameter n is chosen as $n = (L - M)/2$.

The results of this experiment are shown in Fig. 10, where besides the output SNR and speech distortion index, we also plot the scores measured with the perceptual evaluation of speech quality (PESQ) algorithm (ITU-T P.862). PESQ scores are well known to have a high correlation with subjective quality evaluation. As can be seen, all the three studied filtering matrices can improve the SNR and speech
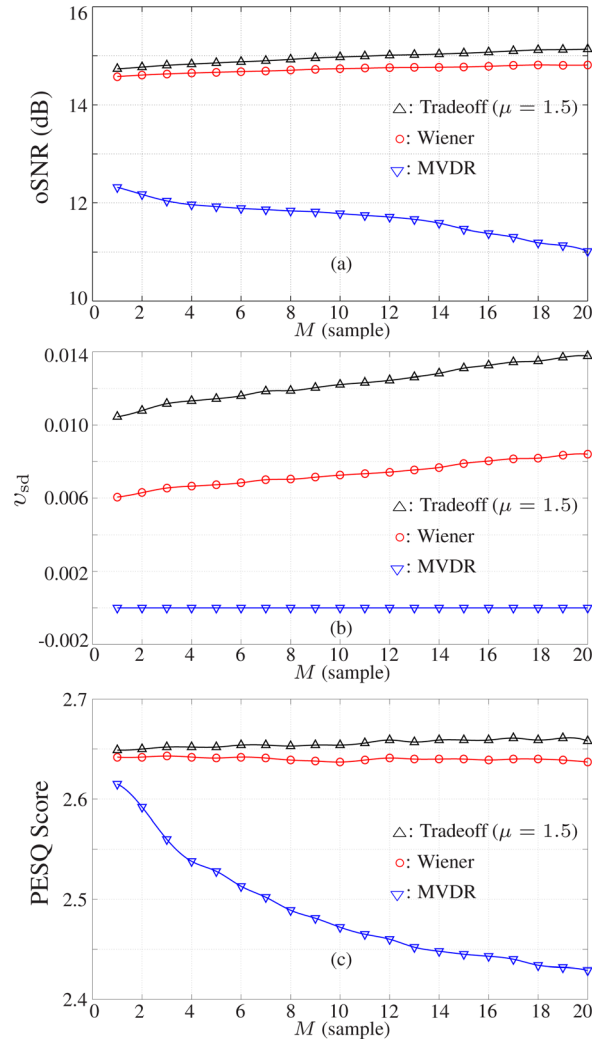

FIG. 10. (Color online) Performance of the Wiener, MVDR, tradeoff ($\mu = 0.5$), and tradeoff ($\mu = 1.5$) filtering matrices as a function of the parameter M in car noise, where the noise statistics are estimated with a VAD based on a two-sided single-pole recursion. The input SNR is 10 dB, $L = 44$, and $n = (L - M)/2$. The PESQ score between the clean and noisy speech is 2.4.

quality with the statistics estimated using a VAD because the output SNR is always greater than the input SNR, and the PESQ scores between the clean and enhanced speech is always larger than that between the clean and noisy speech. The output SNR of the Wiener and tradeoff (with $\mu = 1.5$) filters increases slightly with M, while it decreases with M for the MVDR filter. This corroborates with the results observed previously. The value of the speech distortion index of the Wiener and tradeoff filtering matrices increases with M; but it is approximately 0 for the MVDR filtering matrix. The PESQ score between the clean and enhanced speech depends on many factors such as the SNR and the amount of speech distortion. Both the Wiener and tradeoff (with $\mu = 1.5$) filters have a slightly higher output SNR with a larger value of M; but they also introduce more speech distortion in this case. As a result, their PESQ scores do not change much with M. In comparison, the MVDR filter does not introduce speech distortion; but it achieves less noise reduction as M is increased. So, its PESQ score decreases with M.

Comparing the results of this experiment with those in Fig. 5, one can see that the noise reduction performance with the use of VAD is close to (slightly lower output SNR and larger speech distortion index) that with the signal statistics being directly computed from the corresponding signals. Of course, the accuracy of VAD always plays a very important role in noise reduction performance in practical applications. This is, however, beyond the scope of this paper.

V. CONCLUSIONS

In this paper, we developed a block-based approach to noise reduction where a vector of the desired clean speech is recovered through filtering a frame of the noisy signal with a rectangular filtering matrix. In this framework, the most critical issue is how to find an optimal filtering matrix. To obtain such a matrix, we presented a scheme that decomposes the clean speech vector into two orthogonal components, i.e., the desired speech part and the interference component. Based on this orthogonal decomposition, we formed different MSE criteria and deduced several optimal filtering matrices including Wiener, tradeoff, prediction, and MVDR. Through both theoretical derivation and experiments, it was demonstrated that (1) the block-based method, when the block size is properly chosen, can yield better noise reduction performance than the sample-based technique; (2) the block-based approach can be computationally more efficient and therefore is more practical to implement than the sample-based method; (3) introducing a delay parameter to control the flexibility in the non-causality can help improve the noise reduction performance; and (4) in the single-channel case with the use of filtering matrices, it is possible to derive an MVDR that achieves noise reduction without distorting the desired clean speech.

ACKNOWLEDGEMENT

Benesty, J., and Chen, J. (**2011**). *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study* (Springer-Verlag, Berlin), pp. 1–79.

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (**2009**). *Noise Reduction in Speech Processing* (Springer-Verlag, Berlin), pp. 1–229.

Benesty, J., Chen, J., Huang, Y., and Gaensler, T. (**2012**). "Time-domain noise reduction based on an orthogonal decomposition for desired signal extraction," J. Acoust. Soc. Am. **132**(1), 452–464.

Chen, J., Benesty, J., Huang, Y., and Diethorn, E. J. (**2007**). "Fundamentals of noise reduction," in *Springer Handbook on Speech Processing and Speech Communication*, edited by J. Benesty, M. M. Sondhi, and J. Chen (Springer-Verlag, Berlin), Chap. 43, pp. 843–871.

Chen, J., Benesty, J., Huang, Y., and Doclo, S. (**2006**). "New insights into the noise reduction Wiener filter," IEEE Trans. Audio, Speech Lang. Process. **14**, 1218–1234.

Chen, J., Benesty, J., Huang, Y., and Gaensler, T. (**2011**). "On single-channel noise reduction in the time domain," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Cohen, I. (**2003**). "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Trans. Speech, Audio Process. **11**, 466–475.

Cormen, T. H., Leiserson, C. E., and Rivest, R.L. (**1990**). *Introduction to Algorithms* (MIT Press, Cambridge, MA), pp. 1–1312.

Dendrinos, M., Bakamidis, S., and Carayannis, G. (**1991**). "Speech enhancement from noise: A regenerative approach," Speech Commun. **10**, 45–57.

Diethorn, E. J. (**1997**). "A subband noise-reduction method for enhancing speech in telephony and teleconferencing," in *Proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics* (New Paltz, NY).

Ephraim, Y. (**1992**). "Statstical-model-based speech enhancement systems," Proc. IEEE **80**, 1526–1555.

Ephraim, Y., and Trees, H. L. V. (**1995**). "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Process. **3**, 251–266.

Hermus, K., Wambacq, P., and Hamme, H. V. (**2007**). "A review of signal subspace speech enhancement and its application to noise robust speech recognition," EURASIP J. Adv. Signal Process. **2007**.

Hirsch, H. G., and Ehrlicher, C. (**1995**). "Noise estimation techniques for robust speech recognition," Proc. IEEE ICASSP **1**, 153–156.

Hu, Y., and Loizou, P. C. (**2002**). "A subspace approach for enhancing speech corrupted by colored noise," IEEE Signal Process. Lett. **9**, 204–206.

Hu, Y., and Loizou, P. C. (**2003**). "A generalized subspace approach for enhancing speech corrupted by colored noise," IEEE Trans. Speech Audio Process. **11**, 334–341.

Jabloun, F., and Champagne, B. (**2005**). "Signal subspace techniques for speech enhancement," in *Speech Enhancement*, edited by J. Benesty, S. Makino, and J. Chen (Springer-Verlag, Berlin), Chap. 7, pp. 135–139.

Jensen, J. R., Benesty, J., Christensen, M. G., and Jensen, S. R. (**2012**). "Non-causal time-domain filters for single-channel noise reduction," IEEE Trans. Speech Audio Process. **20**, 1526–1541.

Jensen, S. H., Hansen, P. C., Hansen, S. D., and Sørensen, J. A. (**1995**). "Reduction of broad-band noise in speech by truncated qsvd," IEEE Trans. Speech Audio Process. **3**, 439–448.

Li, C., Benesty, J., and Chen, J. (**2012**). "Optimal rectangular filtering matrix for noise reduction in the time domain," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Lim, J. S. (**1983**). *Speech Enhancement* (Prentice-Hall, Englewood Cliffs, NJ), pp. 1–363.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Boca Raton, FL), pp. 1–585.

Martin, R. (**2001**). "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process. **9**, 504–512.

Searle, S. B. (**1982**). *Matrix Algebra Useful for Statistics* (Wiley, New York), pp. 1–438.

Stahl, V. Fischer, A., and Bippus, R. (**2000**). "Quantile based noise estimation for spectral subtraction and Wiener filtering," Proc. IEEE ICASSP **3**, 1875–1878.

Strang, G. (**1988**). *Linear Algebra and Its Applications*, 3rd ed. (Harcourt Brace Jovanonich, Orlando, FL), pp. 1–496.

Vary, P., and Martin, R. (**2006**). *Digital Speech Transmission: Enhancement, Coding and Error Concealment* (Wiley and Sons, Chichester, England), pp. 1–635.