A MINIMUM VARIANCE PARTIALLY DISTORTIONLESS RESPONSE FILTER FOR SINGLE-CHANNEL NOISE REDUCTION

Xianghui Wang[1]*, Jingdong Chen*[1]*, and Jacob Benesty*[2]

[1]CIAIC and School of Marine Science and Technology
Northwestern Polytechnical University
Xi'an, Shaanxi 710072, China

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

ABSTRACT

This paper deals with the problem of single-channel noise reduction. Thanks to the eigenvalue decomposition, we arrange the eigenvalues of the speech correlation matrix in such a way that all the spectral mode signal-to-noise ratios (SNRs) of the noisy speech are ordered in a descending manner. By maintaining no speech distortion in the spectral modes with high input SNRs while allowing some degree of speech distortion in the modes with low input SNRs, we develop a minimum variance partially distortionless response (MVPDR) filter. We first formulate the problem and derive this filter within the general filtering framework. Then, the MVPDR filter is applied to the single-channel noise reduction problem in both the time and time-frequency domains. In comparison with the minimum variance distortionless response (MVDR) filter based on the subspace decomposition, the developed MVPDR filter can provide much more freedom for controlling the compromise between noise reduction and speech distortion to achieve higher speech quality. Simulations are conducted and preliminary results justify the advantages of the deduced MVPDR filter.

***Index Terms*—**Noise reduction, speech enhancement, single-channel, optimal linear filtering, minimum variance partially distortionless response filter.

1. INTRODUCTION

The existence of noise can cause significant degradation of speech quality and impairment of speech intelligibility, thereby affecting the normal functions of speech communication and processing systems. To mitigate the effect of noise, many noise reduction algorithms have been developed over the past few decades [1–3] including spectral subtractive algorithms [4–6], optimal filtering techniques [7–9], statistical-model-based methods [10–13], subspace approaches [14–16], and deep neural networks (DNNs) based methods [17–19]. While those methods have achieved some degree of success, noise reduction remains a challenging problem due to its extreme difficulty.

This paper studies the problem of noise reduction. Following the principle of the subspace method [14, 15], we apply the eigenvalue decomposition to the speech correlation matrix and then rearrange the corresponding eigenvalues according to the spectral mode signal-to-noise ratios (SNRs) of the noisy speech. By maintaining no speech distortion in the spectral modes with high input SNRs while allowing some degree of speech distortion in the modes with low input SNRs, we develop a minimum variance partially distortionless response (MVPDR) filter. This filter is first developed within the general filtering framework. It is then applied to deal with the problem of single-channel noise reduction in both the time and time-frequency domains. This new filter can be viewed as an extension of the minimum variance distortionless response (MVDR) filter developed recently for single-channel noise reduction [7], which can be viewed as a unification of the subspace and linear filtering methods.

This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005.

2. SIGNAL MODEL AND PROBLEM FORMULATION

We consider the general signal model of an observation signal vector of length L [7]:

$$\begin{aligned}\mathbf{y} &= \left[\begin{array}{cccc} y_1 & y_2 & \cdots & y_L \end{array} \right]^T \\ &= \mathbf{x} + \mathbf{v}, \end{aligned} \tag{1}$$

where the superscript ^{T} is the transpose operator, and $\mathbf{x}$ and $\mathbf{v}$ are the speech and additive noise signal vectors, respectively, which are defined similarly to the noisy signal vector, $\mathbf{y}$. We assume that the components of the two vectors $\mathbf{x}$ and $\mathbf{v}$ are zero mean, stationary, and circular. We further assume that these two vectors are uncorrelated, i.e., $E\left(\mathbf{x}\mathbf{v}^H\right) = E\left(\mathbf{v}\mathbf{x}^H\right) = \mathbf{0}_{L\times L}$, where $E(\cdot)$ denotes mathematical expectation, the superscript ^{H} is the conjugate-transpose operator, and $\mathbf{0}_{L\times L}$ is a matrix of size $L \times L$ with all its elements equal to 0. In this context, the correlation matrix (of size $L \times L$) of the observations is

$$\mathbf{\Phi}_{\mathbf{y}} = E\left(\mathbf{y}\mathbf{y}^H\right) = \mathbf{\Phi}_{\mathbf{x}} + \mathbf{\Phi}_{\mathbf{v}}, \tag{2}$$

where $\mathbf{\Phi}_{\mathbf{x}} = E\left(\mathbf{x}\mathbf{x}^H\right)$ and $\mathbf{\Phi}_{\mathbf{v}} = E\left(\mathbf{v}\mathbf{v}^H\right)$ are the correlation matrices of $\mathbf{x}$ and $\mathbf{v}$, respectively. In the rest of this paper, it is assumed that $\mathrm{rank}\left(\mathbf{\Phi}_{\mathbf{x}}\right) \leq L$ while $\mathrm{rank}\left(\mathbf{\Phi}_{\mathbf{v}}\right) = L$.

One of the most important measures in noise reduction is the SNR; it gives a pretty accurate idea on how noisy the observations are. With the signal model given in (1), the input SNR is defined as

$$\mathrm{iSNR} = \frac{\mathrm{tr}\left(\mathbf{\Phi}_{\mathbf{x}}\right)}{\mathrm{tr}\left(\mathbf{\Phi}_{\mathbf{v}}\right)}, \tag{3}$$

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

Here, what we consider as our desired signal is the first entry of $\mathbf{x}$, i.e., x_1. Then, the objective of noise reduction is to estimate x_1 from $\mathbf{y}$. This should be done in such a way that the noise is reduced as much as possible with no or little distortion to the desired signal sample [2, 8, 9, 20].

3. LINEAR FILTERING

Our objective is to estimate x_1 from $\mathbf{y}$. From the conventional linear filtering theory, we know that the desired signal is estimated as

$$z = \mathbf{h}^H\mathbf{y} = \mathbf{h}^H\left(\mathbf{x} + \mathbf{v}\right) = x_{\mathrm{fd}} + v_{\mathrm{rn}}, \tag{4}$$

where z is the estimate of x_1, $\mathbf{h}$ is a complex-valued filter of length L, $x_{\mathrm{fd}} = \mathbf{h}^H\mathbf{x}$ is the filtered desired signal, and $v_{\mathrm{rn}} = \mathbf{h}^H\mathbf{v}$ is the residual noise. We deduce that the variance of z is

$$\phi_z = E\left(|z|^2\right) = \phi_{x_{\mathrm{fd}}} + \phi_{v_{\mathrm{rn}}}, \tag{5}$$

where $\phi_{x_{\mathrm{fd}}} = \mathbf{h}^H \mathbf{\Phi}_{\mathbf{x}} \mathbf{h}$ and $\phi_{v_{\mathrm{rn}}} = \mathbf{h}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{h}$.

The output SNR, obtained from (5), helps quantify the SNR after filtering. It is given by

$$\mathrm{oSNR}\left(\mathbf{h}\right) = \frac{\phi_{x_{\mathrm{fd}}}}{\phi_{v_{\mathrm{rn}}}} = \frac{\mathbf{h}^H \mathbf{\Phi}_{\mathbf{x}} \mathbf{h}}{\mathbf{h}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{h}}. \tag{6}$$

Then, the main objective of noise reduction is to find an appropriate $\mathbf{h}$ that makes the output SNR greater than the input SNR, i.e., $\mathrm{oSNR}\left(\mathbf{h}\right) > \mathrm{iSNR}$. Consequently, the quality of the noisy signal may be enhanced.

4. MINIMUM VARIANCE PARTIALLY DISTORTIONLESS RESPONSE FILTER

To derive the new filter, we need to exploit the spectrum of $\mathbf{\Phi}_{\mathbf{x}}$. Using the well-known eigenvalue decomposition [21], the speech correlation matrix can be diagonalized as

$$\mathbf{Q}^H \mathbf{\Phi}_{\mathbf{x}} \mathbf{Q} = \mathbf{\Lambda}, \tag{7}$$

where

$$\mathbf{Q} = \left[\begin{array}{cccc} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_L \end{array} \right] \tag{8}$$

is a unitary matrix, i.e., $\mathbf{Q}^H \mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}_L$, with $\mathbf{I}_L$ being the $L \times L$ identity matrix, and

$$\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_L\right) \tag{9}$$

is a diagonal matrix. The orthogonal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_L$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_L$ of the matrix $\mathbf{\Phi}_{\mathbf{x}}$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$.

Using (7), we can express the input SNR as

$$\begin{aligned} \mathrm{iSNR} &= \frac{\mathrm{tr}\left(\mathbf{Q}^H \mathbf{\Phi}_{\mathbf{x}} \mathbf{Q}\right)}{\mathrm{tr}\left(\mathbf{Q}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{Q}\right)} \\ &= \frac{\sum_{l=1}^{L} \lambda_l}{\sum_{l=1}^{L} \mathbf{q}_l^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{q}_l}, \end{aligned} \tag{10}$$

from which we deduce the lth spectral mode input SNR:

$$\mathrm{iSNR}_l = \frac{\lambda_l}{\mathbf{q}_l^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{q}_l},\ l = 1, 2, \ldots, L. \tag{11}$$

Let us denote by $\lambda_{(l)},\ l = 1, 2, \ldots, L$ (with corresponding eigenvectors $\mathbf{q}_{(l)},\ l = 1, 2, \ldots, L$), the eigenvalues of $\mathbf{\Phi}_{\mathbf{x}}$ ordered in such a way that the L spectral mode input SNRs are ordered from the largest to the smallest, i.e.,

$$\mathrm{iSNR}_{(1)} \geq \mathrm{iSNR}_{(2)} \geq \cdots \geq \mathrm{iSNR}_{(L)}, \tag{12}$$

where

$$\mathrm{iSNR}_{(l)} = \frac{\lambda_{(l)}}{\mathbf{q}_{(l)}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{q}_{(l)}},\ l = 1, 2, \ldots, L. \tag{13}$$

Our goal in this paper is to find a noise reduction filter that does not distort the P spectral modes of the desired signal corresponding to the P largest spectral mode input SNRs. We call this filter MVPDR.

Given the required ordering, we can reformulate the diagonalization of $\mathbf{\Phi}_{\mathbf{x}}$ as

$$\mathbf{Q}_{(\cdot)}^H \mathbf{\Phi}_{\mathbf{x}} \mathbf{Q}_{(\cdot)} = \mathbf{\Lambda}_{(\cdot)}, \tag{14}$$

where

$$\begin{aligned} \mathbf{Q}_{(\cdot)} &= \left[\begin{array}{ccccccc} \mathbf{q}_{(1)} & \cdots & \mathbf{q}_{(P)} & \mathbf{q}_{(P+1)} & \cdots & \mathbf{q}_{(L)} \end{array} \right] \\ &= \left[\begin{array}{cc} \mathbf{Q}_{(P)} & \mathbf{Q}'_{(L-P)} \end{array} \right] \end{aligned} \tag{15}$$

and

$$\mathbf{\Lambda}_{(\cdot)} = \mathrm{diag}\left(\lambda_{(1)}, \lambda_{(2)}, \ldots, \lambda_{(L)}\right). \tag{16}$$

It is clear that

$$\mathbf{I}_L = \mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H + \mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)}. \tag{17}$$

The matrices $\mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H$ and $\mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)}$ are two orthogonal projectors of rank P and $L - P$, respectively. Hence, $\mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H$ is the orthogonal projector onto the speech subspace of $\mathbf{\Phi}_{\mathbf{x}}$ with the P largest spectral mode input SNRs and $\mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)}$ is the orthogonal projector onto the speech subspace of $\mathbf{\Phi}_{\mathbf{x}}$ with the $L - P$ smallest spectral mode input SNRs.

Using (17), we can write the speech vector as

$$\begin{aligned} \mathbf{x} &= \mathbf{Q}_{(\cdot)} \mathbf{Q}_{(\cdot)}^H \mathbf{x} \\ &= \mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H \mathbf{x} + \mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)} \mathbf{x} \\ &= \mathbf{x}_{\mathrm{u}} + \mathbf{x}_{\mathrm{d}}. \end{aligned} \tag{18}$$

As a result, the desired signal sample is decomposed as

$$x_1 = \mathbf{i}^T \mathbf{x} = x_{\mathrm{u},1} + x_{\mathrm{d},1}, \tag{19}$$

where $\mathbf{i}$ is the first column of $\mathbf{I}_L$, $x_{\mathrm{u},1} = \mathbf{i}^T \mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H \mathbf{x}$ is the desired signal component that we want undistorted, and $x_{\mathrm{d},1} = \mathbf{i}^T \mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)} \mathbf{x}$ is the desired signal component that we can afford to distort. From (18), we observe that the distortionless constraint for the P spectral modes of interest of the desired signal is

$$\mathbf{h}^H \mathbf{Q}_{(P)} = \mathbf{i}^T \mathbf{Q}_{(P)}. \tag{20}$$

Indeed, left-multiplying (18) by $\mathbf{h}^H$ and applying (20), we get the estimate of the desired signal:

$$\begin{aligned} z = \mathbf{h}^H \mathbf{x} &= \mathbf{h}^H \mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H \mathbf{x} + \mathbf{h}^H \mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)} \mathbf{x} \\ &= \mathbf{i}^T \mathbf{Q}_{(P)} \mathbf{Q}_{(P)}^H \mathbf{x} + \mathbf{h}^H \mathbf{Q}'_{(L-P)} \mathbf{Q}'^H_{(L-P)} \mathbf{x} \\ &= x_{\mathrm{u},1} + \mathbf{h}^H \mathbf{x}_{\mathrm{d}}. \end{aligned} \tag{21}$$

It is clear from the previous expression [with the constraint given in (20)] that the P spectral modes of the desired signal corresponding to the P largest spectral mode input SNRs are not distorted at all since they are not affected by the filtering operation.

Now, the MVPDR can be derived by minimizing the residual noise, i.e., $\phi_{v_{\mathrm{rn}}}$, subject to the distortionless constraint given in (20). Mathematically, this is equivalent to

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{h} \quad \text{subject to} \quad \mathbf{h}^H \mathbf{Q}_{(P)} = \mathbf{i}^T \mathbf{Q}_{(P)}. \tag{22}$$

Solving the above constrained optimization problem, we find the MVPDR filter:

$$\mathbf{h}_{\mathrm{MVPDR},P} = \mathbf{\Phi}_{\mathbf{v}}^{-1} \mathbf{Q}_{(P)} \left(\mathbf{Q}_{(P)}^H \mathbf{\Phi}_{\mathbf{v}}^{-1} \mathbf{Q}_{(P)} \right)^{-1} \mathbf{Q}_{(P)}^H \mathbf{i}. \tag{23}$$

Alternatively, we can express (23) as

$$\mathbf{h}_{\mathrm{MVPDR},P} = \mathbf{\Phi}_{\mathbf{y}}^{-1} \mathbf{Q}_{(P)} \left(\mathbf{Q}_{(P)}^H \mathbf{\Phi}_{\mathbf{y}}^{-1} \mathbf{Q}_{(P)} \right)^{-1} \mathbf{Q}_{(P)}^H \mathbf{i}. \tag{24}$$

There are two interesting particular cases of this filter. For $P = L$, we get the identity vector, i.e., $\mathbf{h}_{\mathrm{MVPDR},L} = \mathbf{i}$, which does not affect the observations; as a consequence, there is neither noise reduction nor speech distortion. The second case is when $P = \mathrm{rank}\left(\mathbf{\Phi}_{\mathbf{x}}\right) \leq L$, which changes $\mathbf{h}_{\mathrm{MVPDR},P}$ to the MVDR filter [7] since $x_{\mathrm{d},1} = 0$.

As far as the output SNR is concerned, we should always have

$$\begin{aligned} \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVPDR},1}\right) &\geq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVPDR},2}\right) \geq \cdots \\ &\geq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVPDR},L}\right) = \mathrm{iSNR}. \end{aligned} \tag{25}$$

Therefore, the smaller P, the better the output SNR.

We remember that the MVPDR filter does not affect the compo-

nent $x_{\mathrm{u},1}$ but does affect the component $x_{\mathrm{d},1}$. So we can evaluate the overall distortion of the desired signal with the speech distortion index [8]:

$$\upsilon\left(\mathbf{h}_{\mathrm{MVPDR},P}\right) = \frac{E\left(\left|x_1 - \mathbf{h}_{\mathrm{MVPDR},P}^H \mathbf{x}\right|^2\right)}{\phi_{x_1}} = \frac{E\left(\left|x_{\mathrm{d},1} - \mathbf{h}_{\mathrm{MVPDR},P}^H \mathbf{x}_{\mathrm{d}}\right|^2\right)}{\phi_{x_1}}, \quad (26)$$

where $\phi_{x_1} = E\left(|x_1|^2\right)$ is the variance of x_1. The smaller the value of $\upsilon\left(\mathbf{h}_{\mathrm{MVPDR},P}\right)$, the smaller the distortion of the component $x_{\mathrm{d},1}$. We should also have

$$\upsilon\left(\mathbf{h}_{\mathrm{MVPDR},1}\right) \geq \upsilon\left(\mathbf{h}_{\mathrm{MVPDR},2}\right) \geq \cdots \geq \upsilon\left(\mathbf{h}_{\mathrm{MVPDR},L}\right) = 0. \quad (27)$$

In the next two sections, we show how to apply this idea to the single-channel noise reduction problem in the time and time-frequency domains.

5. SINGLE-CHANNEL NOISE REDUCTION IN THE TIME DOMAIN

The single-channel noise reduction problem in the time domain consists of recovering the real-valued desired signal (or clean speech) $x(t)$, t being the discrete-time index, of zero mean from the noisy observation (microphone signal) [8]:

$$y(t) = x(t) + v(t), \quad (28)$$

where $v(t)$, assumed to be a zero-mean real random process, is the unwanted additive noise that can be either white or colored but is uncorrelated with $x(t)$.

The signal model given in (28) can be put into a vector form by considering the L most recent successive time samples, i.e.,

$$\mathbf{y}(t) = \begin{bmatrix} y(t) & y(t-1) & \cdots & y(t-L+1) \end{bmatrix}^T = \mathbf{x}(t) + \mathbf{v}(t), \quad (29)$$

where $\mathbf{x}(t)$ and $\mathbf{v}(t)$ are defined in a similar way to $\mathbf{y}(t)$. Then, the goal here is to estimate $x(t)$ from $\mathbf{y}(t)$.

The estimate of $x(t)$ with the proposed filter is

$$z(t) = \mathbf{h}_{\mathrm{MVPDR},P}^T \mathbf{y}(t), \quad (30)$$

where

$$\mathbf{h}_{\mathrm{MVPDR},P} = \mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Q}_{(P)}\left(\mathbf{Q}_{(P)}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Q}_{(P)}\right)^{-1}\mathbf{Q}_{(P)}^T\mathbf{i}, \quad (31)$$

$\mathbf{R}_{\mathbf{v}} = E\left[\mathbf{v}(t)\mathbf{v}^T(t)\right]$ is the correlation matrix of $\mathbf{v}(t)$, and $\mathbf{Q}_{(P)}$ is derived from the correlation matrix, $\mathbf{R}_{\mathbf{x}} = E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right]$, of $\mathbf{x}(t)$ as explained in Section 4.

6. SINGLE-CHANNEL NOISE REDUCTION IN THE TIME-FREQUENCY DOMAIN

Using the short-time Fourier transform (STFT), (28) can be rewritten in the time-frequency domain as [9, 22]

$$Y(k,n) = X(k,n) + V(k,n), \quad (32)$$

where the zero-mean complex random variables $Y(k,n)$, $X(k,n)$, and $V(k,n)$ are the STFTs of $y(t)$, $x(t)$, and $v(t)$, respectively, at frequency bin $k \in \{0, 1, \ldots, K-1\}$ and time frame n. Here, the sample $X(k,n)$ is the desired signal.

In the conventional approach, $X(k,n)$ is estimated by simply applying a positive (but less than 1) gain to $Y(k,n)$. However, the noise reduction performance may be limited.

A more general approach to estimate the desired signal is by filtering the observation signal vector of length L [22]:

$$\mathbf{y}(k,n) = \begin{bmatrix} Y(k,n) & Y(k,n-1) & \cdots & Y(k,n-L+1) \end{bmatrix}^T = \mathbf{x}(k,n) + \mathbf{v}(k,n), \quad (33)$$

where $\mathbf{x}(k,n)$ and $\mathbf{v}(k,n)$ resemble $\mathbf{y}(k,n)$. Thanks to this method, the non-negligible interframe correlation is taken into account, which is not the case when just a gain is used. As a consequence, we can better compromise between noise reduction and speech distortion.

Then, with the proposed approach, the estimate of $X(k,n)$ is

$$Z(k,n) = \mathbf{h}_{\mathrm{MVPDR},P}^H(k,n)\mathbf{y}(k,n), \quad (34)$$

where

$$\mathbf{h}_{\mathrm{MVPDR},P}(k,n) = \mathbf{\Phi}_{\mathbf{v}}^{-1}(k,n)\mathbf{Q}_{(P)}(k,n)\left[\mathbf{Q}_{(P)}^H(k,n)\mathbf{\Phi}_{\mathbf{v}}^{-1}(k,n)\times\mathbf{Q}_{(P)}(k,n)\right]^{-1}\mathbf{Q}_{(P)}^H(k,n)\mathbf{i}, \quad (35)$$

$\mathbf{\Phi}_{\mathbf{v}}(k,n) = E\left[\mathbf{v}(k,n)\mathbf{v}^H(k,n)\right]$ is the correlation matrix of $\mathbf{v}(k,n)$, and $\mathbf{Q}_{(P)}(k,n)$ is derived from the correlation matrix, $\mathbf{\Phi}_{\mathbf{x}}(k,n) = E\left[\mathbf{x}(k,n)\mathbf{x}^H(k,n)\right]$, of $\mathbf{x}(k,n)$ as explained in Section 4.

7. SIMULATIONS

In this section, simulations are carried out to evaluate the performance of the developed MVPDR noise reduction filter. Due to space limitation, only results of the time-frequency domain filter are presented. The clean speech signals are taken from the TIMIT database [23]. All speech signals from the speakers MSDS0 (male speaker) and FKSR0 (female speaker) are used. We downsampled all the signals from the original sampling rate of 16 kHz to 8 kHz. To obtain the noisy speech, a car noise recorded in a sedan running at 50 MPH on a highway is added to the clean speech. The noise signal is properly scaled to control the input SNR level.

The realization process of the noise reduction filter is the same as that in [24]. In the implementation, the correlation matrices of the clean speech and noise signals are needed to be known. In this paper, estimates of the two matrices $\mathbf{\Phi}_{\mathbf{y}}(k,n)$ and $\mathbf{\Phi}_{\mathbf{v}}(k,n)$ are computed directly from the noisy and noise signals, respectively, in the time-frequency domain using a recursive method [9]. Then, the estimate of the correlation matrix of the clean speech is computed as $\hat{\mathbf{\Phi}}_{\mathbf{x}}(k,n) = \hat{\mathbf{\Phi}}_{\mathbf{y}}(k,n) - \hat{\mathbf{\Phi}}_{\mathbf{v}}(k,n)$.

We use the output SNR defined in (6) and the speech distortion index defined in (26) as the performance measures. They are computed in the time domain, i.e., all the time-frequency domain signals are converted back to the time domain and the performance measures are then computed by replacing the mathematical expectation in their definitions by a long time average. Moreover, the perceptual evaluation of speech quality (PESQ) [25] standard is also used to evaluate the overall quality of the processed speech signals by the developed MVPDR filter. Note that the raw PESQ mean opinion score (MOS) is mapped to the PESQ MOS-LQO (listening quality objective) score in our simulations [26].

In the first simulation, we study the performance of the MVPDR filter as a function of P with different filter lengths (L = 2, 4, 6 and 8). The input SNR is 10 dB. The results are plotted in Fig. 1. For the purpose of comparison, the results of the single-channel MVDR filer with P = rank($\mathbf{\Phi}_x$) [7, 22, 24, 27, 28] are also plotted in the figure. As seen in Fig. 1, both the output SNR and speech distortion index decrease with the value of P monotonically for all the studied filter lengths, which coincides with the theory analysis. In contrast, the PESQ score decreases with P if the value of L is s-

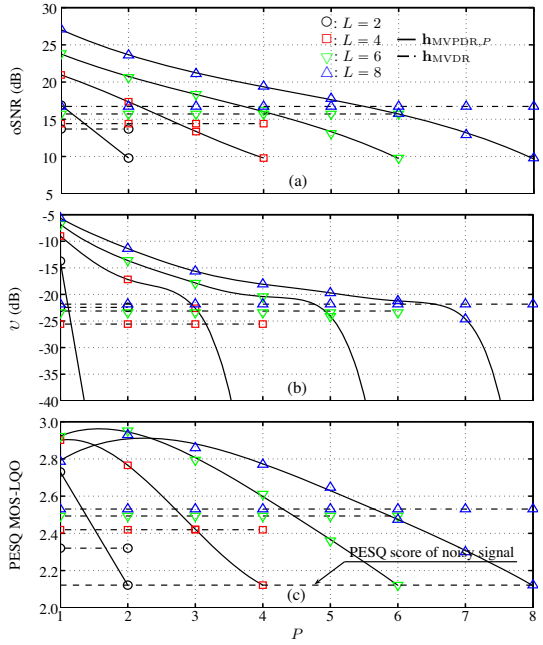

Fig. 1. Performance of $\mathbf{h}_{\mathrm{MVPDR},P}$, as a function of P, and the single-channel MVDR filter in a car noise environment with different filter lengths: (a) oSNR, (b) speech distortion, and (c) PESQ score. The input SNR is 10 dB.

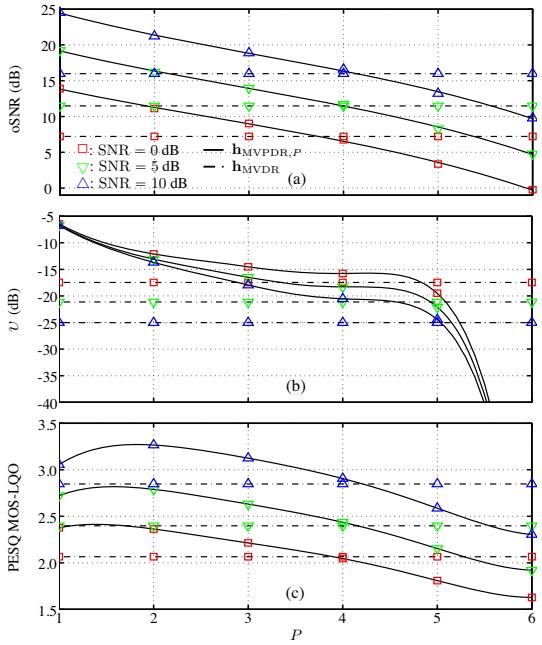

Fig. 2. Performance of $\mathbf{h}_{\mathrm{MVPDR},P}$, as a function of P, and the single-channel MVDR filter in a car noise environment in different iSNR conditions: (a) oSNR, (b) speech distortion, and (c) PESQ score. The filter length $L = 6$ and $T_{60} \approx 240$ ms.

mall; but it first increases with the value of P and then decreases if the filter length is not too small (> 4). In theory, better performance should be achieved with a longer filter. In practice, however, the choice of the filter length is a tradeoff between the noise reduction performance and computational complexity. There will be even some performance degradation if the filter is too long. This is due to: 1) there is not much correlation between the current frame and the far-distance ones; 2) the estimation error of the correlation matrices grows with the increase of the filter length.

As seen, the MVPDR filter achieves a higher a PESQ score than the MVDR filter if the value of P is properly chosen.

In the second simulation, we consider a more practical acoustic environment where there is reverberation. We first measured the acoustic impulse response of a reverberant room. The reverberation time T_{60} of this room is approximately 240 ms. Then, we convolved the speech signals from the speakers MSDS0 and FKSR0 in the TIMIT database with the measured impulse response and this convolved speech was used as the desired, target clean speech. The car noise was subsequently added to control the input SNR. In this simulation, the impact of the value of P on noise reduction is studied with different input SNR levels. The filter length is set to be 6. Figure 2 plots the output SNR, the speech distortion index, and the PESQ score, all as a function of P. Again, results of both the MVPDR and the single-channel MVDR filters are presented. As seen, the trends of both the output SNR and the speech distortion index are similar to those in the previous simulation, which, again, coincides with our theoretical analysis.

Again, the PESQ score obtained by the MVPDR filter is higher than that of the MVDR filter when P is appropriately selected.

8. CONCLUSIONS

Through eigenvalue analysis of the speech correlation matrix and arrangement of the spectral mode input SNRs, a minimum variance partially distortionless response (MVPDR) filter was developed, which maintains no speech distortion for spectral modes with high input SNRs and allows some speech distortion for spectral modes with low input SNRs. We showed how to apply this MVPDR filter to the noise reduction problem in both the time and time-frequency domains. Simulations showed that this MVPDR filter can yield significant improvement in PESQ score and outperforms the recently developed MVDR filter for single-channel noise reduction if the parameter is properly chosen.

9. RELATION TO PRIOR WORK

Noise reduction, which aims at improving either speech quality or speech intelligibility or both at the same time, has been an active area of research in speech and signal processing for decades. Researchers and engineers have attempted to attack this problem from different perspectives and many algorithms have been developed such as spectral subtraction algorithms [4–6], optimal filtering techniques [7–9], statistical-model-based methods [10–13], subspace approaches [14–16], and deep neural networks (DNNs) based methods [17–19]. However, due to the complicated and unknown nature of noise, noise reduction remains a challenging problem. Recently, a framework that unifies the widely studied subspace and linear filtering methods was investigated and a single-channel MVDR filter was developed [7, 22, 24, 27, 28]. This filter was shown to be able to improve performance of the traditional subspace and linear filtering methods. In this paper, the idea of the single-channel MVDR filter is further generalized and an MVPDR filter was developed, which maintains no speech distortion for spectral modes with high input SNRs and allows some distortion for spectral modes with low input SNRs. In comparison with the single-channel MVDR filter, this MVPDR filter can provide more freedom for controlling the compromise between noise reduction and speech distortion to achieve higher speech quality, which is corroborated with simulations.

10. REFERENCES

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

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

[3] J. Chen, J. Benesty, Y. Huang, and E. J. Diethorn, "Fundamentals of noise reduction," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., Berlin, Germany: Springer-Verlag, 2008, pp. 843–871.

[4] M. R. Schroeder, "Apparatus for suppressing noise and distortion in communication signals," Apr. 27, 1965. US Patent 3,180,936.

[5] M. R. Schroeder, "Processing of communications signals to reduce effects of noise," Sept. 24 1968. US Patent 3,403,224.

[6] 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.

[7] J. Benesty, M. G. Christensen, J. R. Jensen, and J. Chen, "A brief overview of speech enhancement with linear filtering," *EURASIP J. Adv. Signal Process.*, vol. 2014, 10 pages, Nov. 2014.

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

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

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

[11] 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. ASSP–32, pp. 1109–1121, Dec. 1984.

[12] 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.

[13] P. J. Wolfe and S. J. Godsill, "Simple alternatives to the Ephraim and Malah suppression rule for speech ehancemnet," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2001, pp. 496–499.

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

[15] P. C. Hansen and S. H. Jensen, "Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: survey and analysis," *EURASIP J. Adv. Signal Process.*, vol. 2007, p. 24, Jun. 2007.

[16] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, pp. 2230–2244, Sept. 2002.

[17] Y. Xu, J. Du, L. Dai, and C. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, pp. 7–19, Jan. 2014.

[18] J. Du, Y. Tu, L. Dai, and C. Lee, "A regression approach to single-channel speech separation via high-resolution deep neural networks," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, pp. 1424–1437, Aug. 2016.

[19] X. Zhang and D. Wang, "A deep ensemble learning method for monaural speech separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, pp. 967–977, May 2016.

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

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

[22] J. Benesty, J. Chen, and E. Habets, *Speech Enhancement in the STFT Domain*. Berlin, Germany: Springer Briefs in Electrical and Computer Engineering, 2011.

[23] J. W. Lyons, "DARPA TIMIT acoustic-phonetic continuous speech corpus," *Technical Report NISTIR 4930, National Institute of Standards and Technology*, 1993.

[24] X. Wang, J. Benesty, and J. Chen, "A single-channel noise cancelation filter in the short-time-fourier-transform domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2016, pp. 5235–5239.

[25] *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*, ITU-T Rec. P.862, 2001.

[26] *Mapping Function for Transforming Raw Result Scores to MOS-LQO*, ITU-T Rec. P.862.1, 2003.

[27] J. Benesty and J. Chen, *Optimal Time-domain Noise Reduction Filters–A Theoretical Study*. Berlin, Germany: Springer Briefs in Electrical and Computer Engineering, 2011.

[28] S. M. Nørholm, J. Benesty, J. R. Jensen, and M. G. Christensen, "Single-channel noise reduction using unified joint diagonalization and optimal filtering," *EURASIP J. Adv. Signal Process.*, vol. 2014, 11 pages, Dec. 2014.