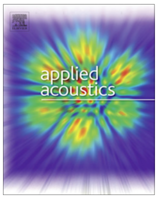
Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust



On single-channel noise reduction with rank-deficient noise correlation matrix [☆]

Ningning Pan [a], Jacob Benesty [b], Jingdong Chen [a]

[a] Center of Intelligent Acoustics and Immersive Communications and School of Marine Science and Technology, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, Shaanxi 710072, China
[b] INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, QC H5A 1K6, Canada

ARTICLE INFO

Article history:
Received 7 October 2016
Received in revised form 3 March 2017
Accepted 5 May 2017
Available online 13 May 2017

Keywords:
Noise reduction
Rank deficient
Subspace
Eigenvalue decomposition
Optimal linear filtering
Maximum SNR filter
Tradeoff filter
Wiener filter

ABSTRACT

The widely studied subspace and linear filtering methods for noise reduction require the noise correlation matrix to be invertible. In certain application scenarios, however, this matrix is either rank deficient or very ill conditioned, so this requirement cannot be fulfilled. In this paper, we investigate possible solutions to this important problem based on subspace techniques for single-channel time-domain noise reduction. The eigenvalue decomposition is applied to both the speech and noise correlation matrices to separate the null and nonnull subspaces. Then, a set of optimal and suboptimal filters are derived from the nullspace of the noise signal. Through simulations, we observe that the proposed filters are able to significantly reduce noise without introducing much distortion to the desired signal. In comparison with the conventional Wiener approach, the developed filters perform significantly better in improving both the signal-to-noise ratio (SNR) and the perceptual evaluation of speech quality (PESQ) score when the noise correlation matrix is rank deficient.

1. Introduction

Noise reduction, which is often also referred to as speech enhancement, is a problem of recovering a clean speech signal of interest from its microphone observations corrupted by additive noise [1–3]. The goal of noise reduction may vary from one application to another but, generally, it is to improve either the perceptual quality or the intelligibility or both of the noisy speech signal. This has long been a challenge in many important real-world applications, such as mobile speech communication, hearing aids, robotics, audio conferencing, and robust speech recognition, to name a few. Extensive work has been done to address this problem in the literature [1–7] and many different methods have been developed, including optimal filtering [8,9], spectral subtraction type of techniques [6,10–15], statistical approach [16–20], subspace methods [21–28], deep neural networks (DNNs) [29–32], and multichannel filtering [5,3,33].

Every of the aforementioned methods has its own pros and cons. For example, the optimal filtering and subspace methods work in the time domain. They require the estimation of the noise correlation matrix which has to be well conditioned so that its inverse can be computed reliably. Furthermore, these methods are relatively expensive in computation as matrix inversion is involved. In comparison, spectral subtraction type of techniques are computationally very efficient thanks to the use of the fast Fourier transform (FFT). However, speech distortion with this method is large, which can only be controlled by sacrificing the amount of noise reduction. The statistical approach generally assumes some *a priori* knowledge about the speech and noise distributions or even the knowledge of the joint probability distribution of the clean speech and noise signals, so that the conditional expected value of the clean speech (or its spectrum) can be evaluated given the noisy signal. If the assumed distribution does not model well the noise in real applications, which happens often, the method may suffer from dramatic performance degradation. Unlike the statistical method, the DNNs based approach does not assume any *a priori* knowledge about the statistics and distributions of the speech and noise signals; it learns all the needed information from the training data. If the signal and noise characteristics in real applications are similar to those in the training set, this method may work well but, otherwise, its performance

[☆] This work was sponsored in part by the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005 and the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University.

E-mail addresses: nnpan@mail.nwpu.edu.cn (N. Pan), benesty@emt.inrs.ca (J. Benesty), jingdongchen@ieee.org (J. Chen)

http://dx.doi.org/10.1016/j.apacoust.2017.05.010
0003-682X/© 2017 Elsevier Ltd. All rights reserved.

can be problematic. Nevertheless, the aforementioned methods are successful to a certain degree, but none of those can claim victory in dealing with the complicated noise reduction problem. Further effort in this area is indispensable.

This paper deals with the problem of single-channel noise reduction in the time domain. We focus on the scenario where the noise correlation matrix is rank deficient. This happens often in many applications where there is narrowband or harmonic interference or transit and bandlimited noise (such as door slamming, keyboard typing, etc). Unlike white and colored noises that have been intensively studied in the literature, there is not much work so far to address the noise reduction problem with a rank-deficient noise correlation matrix. The approach we take here is based on the principles of both subspace decomposition and optimal filtering. First, the eigenvalue decomposition is applied to the desired speech and noise correlation matrices. The nullspace (formed from the eigenvectors corresponding to the zero eigenvalues) of the noise correlation matrix is then used to design a set of optimal linear noise reduction filters. Using the entire nullspace of the noise signal, we can design a maximum signal-to-noise ratio (SNR) filter, which gives a high output SNR but with large speech distortion. Manipulating the dimension of this nullspace leads to a set of tradeoff filters, which can make a compromise between the output SNR and the amount of speech distortion for better perceptual speech quality.

The rest of this paper is organized as follows. In Section 2, we present the formulation of the noise reduction problem and some basic background information about the eigenvalue decomposition in the context of noise reduction. We then discuss how to design different filters including the Wiener, maximum SNR, and tradeoff filters in Section 3. Simulations in harmonic noise, keyboard typing noise, and mixture of these noises with white Gaussian noise are presented in Section 4 to demonstrate the properties of the developed filters. Finally, some conclusions are given in Section 5.

2. Noise reduction problem

The problem considered in this paper is one of recovering a clean speech signal of interest from its noisy observation (sensor signal) [8,3]:

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

where $x(k)$ is the zero-mean desired speech signal, k is the discrete-time index, $v(k)$ is the unwanted zero-mean additive noise, which can be narrowband but is assumed to be uncorrelated with $x(k)$.

With the signal model in (1), we define the input SNR as

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

where $\sigma_x^2 \triangleq E[x^2(k)]$ and $\sigma_v^2 \triangleq E[v^2(k)]$ are the variances of $x(k)$ and $v(k)$, respectively.

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

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k), \tag{3}$$

where

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

is a vector of length L, the 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 (4). Since $x(k)$ and $v(k)$ are uncorrelated by assumption, the correlation matrix (of size $L \times L$) of the noisy signal can be written as

$$\mathbf{R_y} \triangleq E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right] = \mathbf{R_x} + \mathbf{R_v}, \tag{5}$$

where $E[\cdot]$ denotes mathematical expectation, and $\mathbf{R_x} \triangleq E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ and $\mathbf{R_v} \triangleq E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively.

In the context of noise reduction, the desired signal correlation matrix, $\mathbf{R_x}$, is generally not full rank. Without loss of generality, we assume in this paper that the rank of $\mathbf{R_x}$ is equal to $P \leqslant L$. In the literature, the noise correlation matrix, $\mathbf{R_v}$, is generally assumed to be full rank and well conditioned. However, in many applications, this matrix can be rank deficient. Here, we deal with this particular case. Let us assume that the rank of $\mathbf{R_v}$ is equal to $Q < L$. Then, the objective of noise reduction (or speech enhancement) is to estimate the desired signal sample, $x(k)$, from the observation signal vector, $\mathbf{y}(k)$. It should be noticed that neither the joint diagonalization [22,25] nor the prewhitening approach can be applied to this problem [34] since they require the noise correlation matrix to be full rank.

Using the well-known eigenvalue decomposition [35], the noise correlation matrix can be diagonalized as

$$\mathbf{U_v}^T\mathbf{R_v}\mathbf{U_v} = \mathbf{\Lambda_v}, \tag{6}$$

where

$$\mathbf{U_v} = [\, \mathbf{u}_{\mathbf{v},1} \quad \mathbf{u}_{\mathbf{v},2} \quad \cdots \quad \mathbf{u}_{\mathbf{v},L}\,] \tag{7}$$

is an orthogonal matrix, i.e., $\mathbf{U_v}^T\mathbf{U_v} = \mathbf{U_v}\mathbf{U_v}^T = \mathbf{I}_L$, with $\mathbf{I}_L$ being the $L \times L$ identity matrix, and

$$\mathbf{\Lambda_v} = \mathrm{diag}(\lambda_{\mathbf{v},1}, \lambda_{\mathbf{v},2}, \ldots, \lambda_{\mathbf{v},L}) \tag{8}$$

is a diagonal matrix. The orthonormal vectors $\mathbf{u}_{\mathbf{v},1}, \mathbf{u}_{\mathbf{v},2}, \ldots, \mathbf{u}_{\mathbf{v},L}$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_{\mathbf{v},1}, \lambda_{\mathbf{v},2}, \ldots, \lambda_{\mathbf{v},L}$ of the matrix $\mathbf{R_v}$, where $\lambda_{\mathbf{v},1} \geqslant \lambda_{\mathbf{v},2} \geqslant \cdots \geqslant \lambda_{\mathbf{v},Q} > \lambda_{\mathbf{v},Q+1} = \lambda_{\mathbf{v},Q+2} = \cdots = \lambda_{\mathbf{v},L} = 0$.

In the same way, the desired speech correlation matrix can be diagonalized as

$$\mathbf{U_x}^T\mathbf{R_x}\mathbf{U_x} = \mathbf{\Lambda_x}, \tag{9}$$

where the orthogonal and diagonal matrices $\mathbf{U_x}$ and $\mathbf{\Lambda_x}$ are defined in a similar way to $\mathbf{U_v}$ and $\mathbf{\Lambda_v}$, respectively, with $\lambda_{\mathbf{x},1} \geqslant \lambda_{\mathbf{x},2} \geqslant \cdots \geqslant \lambda_{\mathbf{x},P} > \lambda_{\mathbf{x},P+1} = \lambda_{\mathbf{x},P+2} = \cdots = \lambda_{\mathbf{x},L} = 0$. The above two decompositions will be used in the rest of this paper for the purpose of deriving new optimal linear filters.

3. Filter design

3.1. Linear filter model

The most straightforward and practical way to perform noise reduction in the time domain is to apply a linear filter to the observation signal vector, $\mathbf{y}(k)$, i.e.,

$$\begin{aligned} z(k) &= \mathbf{h}^T\mathbf{y}(k) \\ &= \mathbf{h}^T[\mathbf{x}(k) + \mathbf{v}(k)] \\ &= x_{\mathrm{fd}}(k) + v_{\mathrm{rn}}(k), \end{aligned} \tag{10}$$

where $z(k)$ is the estimate of $x(k)$,

$$\mathbf{h} = [\, h_1 \quad h_2 \quad \cdots \quad h_L\,]^T \tag{11}$$

is a real-valued linear filter of length L,

$$x_{\mathrm{fd}}(k) \triangleq \mathbf{h}^T\mathbf{x}(k) \tag{12}$$

is the filtered desired signal, and

$$v_{\mathrm{rn}}(k) \triangleq \mathbf{h}^T\mathbf{v}(k) \tag{13}$$

is the residual noise.

From (10), we find that the output SNR is

$$\mathrm{oSNR}(\mathbf{h}) \triangleq \frac{\mathbf{h}^T\mathbf{R}_\mathbf{x}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_\mathbf{v}\mathbf{h}}. \tag{14}$$

Following the study in [9], we also define the speech distortion index:

$$\begin{aligned}\upsilon_{\mathrm{sd}}(\mathbf{h}) &\triangleq \frac{E\left\{\left[x_{\mathrm{fd}}(k) - x(k)\right]^2\right\}}{\sigma_x^2}\\ &= \frac{E\left\{\left[\mathbf{h}^T\mathbf{x}(k) - x(k)\right]^2\right\}}{\sigma_x^2}.\end{aligned} \tag{15}$$

The output SNR and speech distortion index will be used in this study to evaluate the performance of the noise reduction algorithms.

Now, we define the error signal between the estimated and desired signals as

$$\begin{aligned}e(k) &\triangleq z(k) - x(k)\\ &= \mathbf{h}^T\mathbf{y}(k) - x(k)\\ &= (\mathbf{h} - \mathbf{i}_1)^T\mathbf{x}(k) + \mathbf{h}^T\mathbf{v}(k)\\ &= e_{\mathrm{ds}}(k) + e_{\mathrm{rn}}(k),\end{aligned} \tag{16}$$

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

$$e_{\mathrm{ds}}(k) \triangleq (\mathbf{h} - \mathbf{i}_1)^T\mathbf{x}(k) \tag{17}$$

is the signal distortion due to the linear filter, and

$$e_{\mathrm{rn}}(k) \triangleq \mathbf{h}^T\mathbf{v}(k) \tag{18}$$

represents the residual noise. The mean-squared error (MSE) is then written as

$$\begin{aligned}J(\mathbf{h}) &\triangleq E\left[e^2(k)\right]\\ &= \sigma_x^2 - 2\mathbf{h}^T\mathbf{R}_\mathbf{x}\mathbf{i}_1 + \mathbf{h}^T\mathbf{R}_\mathbf{y}\mathbf{h}.\end{aligned} \tag{19}$$

Using the fact that $E[e_{\mathrm{ds}}(k)e_{\mathrm{rn}}(k)] = 0, J(\mathbf{h})$ can also be expressed as the sum of two MSEs, i.e.,

$$\begin{aligned}J(\mathbf{h}) &\triangleq E\left[e_{\mathrm{ds}}^2(k)\right] + E\left[e_{\mathrm{rn}}^2(k)\right]\\ &= J_{\mathrm{ds}}(\mathbf{h}) + J_{\mathrm{rn}}(\mathbf{h}),\end{aligned} \tag{20}$$

where

$$\begin{aligned}J_{\mathrm{ds}}(\mathbf{h}) &\triangleq E\left\{\left[(\mathbf{h} - \mathbf{i}_1)^T\mathbf{x}(k)\right]^2\right\}\\ &= (\mathbf{h} - \mathbf{i}_1)^T\mathbf{R}_\mathbf{x}(\mathbf{h} - \mathbf{i}_1)\end{aligned} \tag{21}$$

is the MSE of the signal distortion and

$$\begin{aligned}J_{\mathrm{rn}}(\mathbf{h}) &\triangleq E\left\{\left[\mathbf{h}^T\mathbf{v}(k)\right]^2\right\}\\ &= \mathbf{h}^T\mathbf{R}_\mathbf{v}\mathbf{h}\end{aligned} \tag{22}$$

is the MSE of the residual noise. We will show how these MSEs are used in the next subsections.

3.2. Wiener

Minimizing $J(\mathbf{h})$ in (19) with respect to $\mathbf{h}$ yields

$$\mathbf{R}_\mathbf{y}\mathbf{h} = \mathbf{R}_\mathbf{x}\mathbf{i}_1. \tag{23}$$

When $\mathbf{R}_\mathbf{y}$ is full rank, then we get the classical Wiener filter, i.e., $\mathbf{h}_\mathrm{W} = \mathbf{R}_\mathbf{y}^{-1}\mathbf{R}_\mathbf{x}\mathbf{i}_1$. A necessary condition for this filter to exist is that $P + Q \geqslant L$. Note that even if this condition is fulfilled, $\mathbf{R}_\mathbf{y}$ may be highly ill conditioned in practice. In this case, this matrix should be properly regularized in order to have a reliable estimate of the desired signal.

3.3. Maximum SNR

The maximum SNR filter has been widely studied in the literature when the noise correlation matrix is full rank. In this scenario, the maximum SNR filter is simply the eigenvector corresponding to the maximum eigenvalue of the matrix $\mathbf{R}_\mathbf{v}^{-1}\mathbf{R}_\mathbf{x}$ [36]. However, this traditional filter does not exist if $\mathbf{R}_\mathbf{v}$ is rank deficient. In this section, we show how to exploit the nullspace of the noise correlation matrix to derive a noise reduction filter that can maximize the output SNR.

Let

$$\mathbf{T}_{\mathbf{v},Q} \triangleq \left[\, \mathbf{u}_{\mathbf{v},Q+1} \quad \mathbf{u}_{\mathbf{v},Q+2} \quad \cdots \quad \mathbf{u}_{\mathbf{v},L} \,\right] \tag{24}$$

be the matrix of size $L \times (L - Q)$ composed by the eigenvectors corresponding to the null eigenvalues of $\mathbf{R}_\mathbf{v}$. We are interested in the linear filters of the form:

$$\mathbf{h} = \mathbf{T}_{\mathbf{v},Q}\boldsymbol{\alpha}, \tag{25}$$

where

$$\boldsymbol{\alpha} = \left[\, \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_{L-Q} \,\right]^T \neq \mathbf{0} \tag{26}$$

is a vector of length $L - Q$. Since $\mathbf{R}_\mathbf{v}\mathbf{T}_{\mathbf{v},Q} = \mathbf{0}$ and assuming that $\mathbf{R}_\mathbf{x}\mathbf{T}_{\mathbf{v},Q} \neq \mathbf{0}$, which is reasonable since $\mathbf{R}_\mathbf{x}$ and $\mathbf{R}_\mathbf{v}$ cannot be diagonalized by the same orthogonal matrix unless one of them is white, we have

$$\mathrm{oSNR}(\mathbf{h}) = \mathrm{oSNR}(\mathbf{T}_{\mathbf{v},Q}\boldsymbol{\alpha}) = \infty. \tag{27}$$

As a consequence, the estimate of $x(k)$ is

$$\begin{aligned}z(k) &= \mathbf{h}^T\mathbf{y}(k)\\ &= \boldsymbol{\alpha}^T\mathbf{T}_{\mathbf{v},Q}^T\mathbf{x}(k) + \boldsymbol{\alpha}^T\mathbf{T}_{\mathbf{v},Q}^T\mathbf{v}(k)\\ &= \boldsymbol{\alpha}^T\mathbf{T}_{\mathbf{v},Q}^T\mathbf{x}(k).\end{aligned} \tag{28}$$

We observe from the previous expression that this approach is able to completely cancel the noise. The noise reduction problem then boils down to finding the optimal vector $\boldsymbol{\alpha}$. The best way to find it is by minimizing the distortion of the estimated desired signal.

From the distortion based MSE in (21), we have

$$J_{\mathrm{ds}}(\boldsymbol{\alpha}) = (\mathbf{T}_{\mathbf{v},Q}\boldsymbol{\alpha} - \mathbf{i}_1)^T\mathbf{R}_\mathbf{x}(\mathbf{T}_{\mathbf{v},Q}\boldsymbol{\alpha} - \mathbf{i}_1). \tag{29}$$

The optimal value of $\boldsymbol{\alpha}$ is found by minimizing $J_{\mathrm{ds}}(\boldsymbol{\alpha})$. The solution can be divided into three cases: $P > L - Q, P = L - Q$, and $P < L - Q$.

First, let us consider the case where $P \geqslant L - Q$. In this situation, we find that

$$\begin{aligned}\boldsymbol{\alpha}_\mathrm{o} &= \left(\mathbf{T}_{\mathbf{v},Q}^T\mathbf{R}_\mathbf{x}\mathbf{T}_{\mathbf{v},Q}\right)^{-1}\mathbf{T}_{\mathbf{v},Q}^T\mathbf{R}_\mathbf{x}\mathbf{i}_1\\ &= \left(\mathbf{T}_{\mathbf{v},Q}^T\mathbf{U}_{\mathbf{x},P}\boldsymbol{\Lambda}_{\mathbf{x},P}\mathbf{U}_{\mathbf{x},P}^T\mathbf{T}_{\mathbf{v},Q}\right)^{-1}\mathbf{T}_{\mathbf{v},Q}^T\mathbf{U}_{\mathbf{x},P}\boldsymbol{\Lambda}_{\mathbf{x},P}\mathbf{U}_{\mathbf{x},P}^T\mathbf{i}_1,\end{aligned} \tag{30}$$

where

$$\mathbf{U}_{\mathbf{x},P} = \left[\, \mathbf{u}_{\mathbf{x},1} \quad \mathbf{u}_{\mathbf{x},2} \quad \cdots \quad \mathbf{u}_{\mathbf{x},P} \,\right] \tag{31}$$

is the matrix of size $L \times P$ containing the eigenvectors corresponding to the nonnull eigenvalues of $\mathbf{R}_\mathbf{x}$ and

$$\boldsymbol{\Lambda}_{\mathbf{x},P} = \mathrm{diag}(\lambda_{\mathbf{x},1}, \quad \lambda_{\mathbf{x},2}, \quad \ldots, \quad \lambda_{\mathbf{x},P}). \tag{32}$$

Substituting (30) into (25), we finally find the maximum SNR filter with minimum distortion:

$$\begin{aligned}\mathbf{h}_{\max} =& \mathbf{T}_{\mathbf{v},Q}\left(\mathbf{T}_{\mathbf{v},Q}^T\mathbf{U}_{\mathbf{x},P}\boldsymbol{\Lambda}_{\mathbf{x},P}\mathbf{U}_{\mathbf{x},P}^T\mathbf{T}_{\mathbf{v},Q}\right)^{-1}\\ &\times \mathbf{T}_{\mathbf{v},Q}^T\mathbf{U}_{\mathbf{x},P}\boldsymbol{\Lambda}_{\mathbf{x},P}\mathbf{U}_{\mathbf{x},P}^T\mathbf{i}_1.\end{aligned} \tag{33}$$

It is worth noticing that the larger is the dimension of the nullspace of $\mathbf{R}_\mathbf{v}$, the larger is the dimension of $\boldsymbol{\alpha}$ to minimize distortion. Consequently, there will be less distortion added into the desired signal.

The worst case is when $Q = L - 1$. In this situation, $\boldsymbol{\alpha}$ degenerates to a scalar, which can still significantly increase the SNR but with no control on speech distortion.

Next, we consider the case: $P = L - Q$. In this scenario, we can always express the estimate of $x(k)$ as

$$\begin{aligned} z(k) &= \boldsymbol{\alpha}_{\text{o}}^T \mathbf{T}_{\mathbf{v},Q}^T \mathbf{x}(k) \\ &= \boldsymbol{\alpha}_{\text{o}}^T \mathbf{T}_{\mathbf{v},Q}^T \mathbf{U}_{\mathbf{x},P} \mathbf{U}_{\mathbf{x},P}^T \mathbf{x}(k). \end{aligned} \tag{34}$$

Inspecting (34), we see that, in order to recover the desired signal, $x(k)$, we must have

$$\boldsymbol{\alpha}_{\text{o}}^T \mathbf{T}_{\mathbf{v},Q}^T \mathbf{U}_{\mathbf{x},P} = \mathbf{i}_1^T \mathbf{U}_{\mathbf{x},P}, \tag{35}$$

or, equivalently,

$$\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q} \boldsymbol{\alpha}_{\text{o}} = \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1. \tag{36}$$

Since $\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q}$ is a square invertible matrix, we have a unique solution for (36), which is

$$\boldsymbol{\alpha}_{\text{o},1} = \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q}\right)^{-1} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1. \tag{37}$$

We deduce that the optimal filter for this particular case is

$$\mathbf{h}_{\max,1} = \mathbf{T}_{\mathbf{v},Q} \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q}\right)^{-1} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1. \tag{38}$$

This filter perfectly recovers the desired signal (i.e., it completely cancels the noise without any distortion). Indeed, it can be verified that

$$\begin{aligned} z(k) &= \mathbf{h}_{\max,1}^T \mathbf{y}(k) \\ &= \mathbf{i}_1^T \mathbf{U}_{\mathbf{x},P} \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q}\right)^{-T} \mathbf{T}_{\mathbf{v},Q}^T [\mathbf{x}(k) + \mathbf{v}(k)] \\ &= x(k). \end{aligned} \tag{39}$$

This result can also be found directly from $\mathbf{h}_{\max}$ in (33) by taking $P = L - Q$. Therefore, (33) gives the maximum SNR filter with minimum distortion for cases with $P \geqslant L - Q$.

If $P < L - Q$, one reasonable approach is to take the minimum-norm solution of (36), i.e.,

$$\boldsymbol{\alpha}_{\text{o},2} = \mathbf{T}_{\mathbf{v},Q}^T \mathbf{U}_{\mathbf{x},P} \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q} \mathbf{T}_{\mathbf{v},Q}^T \mathbf{U}_{\mathbf{x},P}\right)^{-1} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1. \tag{40}$$

Therefore, we get another maximum SNR filter for this particular case:

$$\mathbf{h}_{\max,2} = \mathbf{T}_{\mathbf{v},Q} \mathbf{T}_{\mathbf{v},Q}^T \mathbf{U}_{\mathbf{x},P} \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q} \mathbf{T}_{\mathbf{v},Q}^T \mathbf{U}_{\mathbf{x},P}\right)^{-1} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1. \tag{41}$$

This filter is able to reduce noise but some distortion of the desired signal is expected.

3.4. Tradeoff

In practice, too much noise reduction is not necessary, since it is often achieved at the expense of high speech distortion. The most straightforward way to compromise between the amount of noise reduction and the degree of signal distortion is by considering the $L - Q'$ $(0 \leqslant Q' \leqslant Q)$ smallest eigenvalues of $\mathbf{R}_{\mathbf{v}}$. In this way, we increase the dimension of $\boldsymbol{\alpha}$ for less distortion at the cost of less noise reduction. The eigenvector matrix [of size $L \times (L - Q')$] corresponding to these smallest eigenvalues is

$$\mathbf{T}_{\mathbf{v},Q'} = \begin{bmatrix} \mathbf{u}_{\mathbf{v},Q'+1} & \mathbf{u}_{\mathbf{v},Q'+2} & \cdots & \mathbf{u}_{\mathbf{v},L} \end{bmatrix}. \tag{42}$$

Following a similar derivation as in the previous subsection, we deduce the following tradeoff filter:

$$\begin{aligned} \mathbf{h}_{\text{T},Q'} = & \mathbf{T}_{\mathbf{v},Q'} \left(\mathbf{T}_{\mathbf{v},Q'}^T \mathbf{U}_{\mathbf{x},P} \boldsymbol{\Lambda}_{\mathbf{x},P} \mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q'}\right)^{-1} \\ & \times \mathbf{T}_{\mathbf{v},Q'}^T \mathbf{U}_{\mathbf{x},P} \boldsymbol{\Lambda}_{\mathbf{x},P} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1. \end{aligned} \tag{43}$$

This filter exists only if $P \geqslant L - Q'$. For $Q' = 0$ (in this case, P must be equal to L), the tradeoff filter degenerates to the identity filter, i.e.,

$$\mathbf{h}_{\text{T},0} = \mathbf{i}_1, \tag{44}$$

for which there is neither noise reduction nor desired signal distortion.

In general, we should have

$$\begin{aligned} \text{iSNR} = \text{oSNR}(\mathbf{h}_{\text{T},0}) \leqslant \text{oSNR}(\mathbf{h}_{\text{T},1}) \leqslant \cdots \leqslant \text{oSNR}(\mathbf{h}_{\text{T},Q}) \\ = \infty \end{aligned} \tag{45}$$

and

$$0 = \upsilon_{\text{ds}}(\mathbf{h}_{\text{T},0}) \leqslant \upsilon_{\text{ds}}(\mathbf{h}_{\text{T},1}) \leqslant \cdots \leqslant \upsilon_{\text{ds}}(\mathbf{h}_{\text{T},Q}) \leq 1. \tag{46}$$

The tradeoff filter can obviously be used when $\mathbf{R}_{\mathbf{v}}$ is full rank. We have two particular cases: $P = L - Q'$ and $P < L - Q'$. For the former, we get

$$\mathbf{h}_{\text{T},1,Q'} = \mathbf{T}_{\mathbf{v},Q'} \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q'}\right)^{-1} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1, \tag{47}$$

which is a distortionless filter but some residual noise remains. For the latter case, we have

$$\mathbf{h}_{\text{T},2,Q'} = \mathbf{T}_{\mathbf{v},Q'} \mathbf{T}_{\mathbf{v},Q'}^T \mathbf{U}_{\mathbf{x},P} \left(\mathbf{U}_{\mathbf{x},P}^T \mathbf{T}_{\mathbf{v},Q'} \mathbf{T}_{\mathbf{v},Q'}^T \mathbf{U}_{\mathbf{x},P}\right)^{-1} \mathbf{U}_{\mathbf{x},P}^T \mathbf{i}_1, \tag{48}$$

which allows us to make a compromise between the amount of noise reduction and the amount of signal distortion.

4. Simulations

Having derived different optimal noise reduction filters, we study their performance via simulations in this section. We mainly use the output SNR defined in (14), the speech distortion index defined in (15), and the perceptual evaluation of speech quality (PESQ) score [37,38] as the performance measures. However, Per the reviewers' suggestion, the short-time objective intelligibility (STOI) score [39] and the log-spectral distortion (LSD) [40] are also adopted in some simulations as the performance metrics for comparison.

We consider two types of narrowband noises: synthetic harmonic signals and recorded keyboard typing noise. Moreover, in order to see the performance of the designed filters dealing with both narrowband and broadband noises, we also consider the case where the noise is a mixture of white Gaussian noise and the two types of narrowband noises.

The clean speech signals used are taken from the TIMIT database [41]. They consist of 200 sentences from 10 male and 10 female speakers. Note that we focus on the applications of narrowband voice communication, so the signals are downsampled from the original sampling rate of 16 kHz to 8 kHz.

In order to implement the filters derived in the previous sections, we need to know the correlation matrices $\mathbf{R}_{\mathbf{y}}$ and $\mathbf{R}_{\mathbf{v}}$. The $\mathbf{R}_{\mathbf{y}}$ matrix can be directly computed from the noisy signal $y(t)$. However, a noise estimator is needed to estimate the $\mathbf{R}_{\mathbf{v}}$ matrix. A number of algorithms have been developed in the literature to estimate the noise or its spectra, such as the voice activity detection (VAD) based method, the minimum statistics based algorithm [11], the improved-minima-controlled-recursive-averaging (IMCRA) method [15], etc. Those methods work reasonably well if the noise is relatively stationary. However, if the noise is highly nonstationary like the keyboard typing noise that is going to be dealt with in this work, none of those algorithms can produce reliable estimates. Currently, we are investigating a deep neural network (DNN) based noise estimation method, which seems promising in dealing with highly nonstationary noises and the results will be reported in a separate work. In the rest part of this section, however, we set aside the noise estimation problem and focus on illustrating the performance of the developed filters. So, we compute the $\mathbf{R_y}$ and $\mathbf{R_v}$ matrices using the following recursions [3]:

$$\widehat{\mathbf{R}}_{\mathbf{y}}(t) = \alpha_y \widehat{\mathbf{R}}_{\mathbf{y}}(t-1) + (1-\alpha_y)\mathbf{y}(t)\mathbf{y}^T(t), \tag{49}$$

$$\widehat{\mathbf{R}}_{\mathbf{v}}(t) = \alpha_v \widehat{\mathbf{R}}_{\mathbf{v}}(t-1) + (1-\alpha_v)\mathbf{v}(t)\mathbf{v}^T(t), \tag{50}$$

where $\alpha_y \in (0,1)$ and $\alpha_v \in (0,1)$ are two forgetting factors, which control the influence of the previous data samples on the current estimate (the initial estimate is obtained from the first 1600 signal samples with a long-time average). After obtaining $\widehat{\mathbf{R}}_{\mathbf{y}}(t)$ and $\widehat{\mathbf{R}}_{\mathbf{v}}(t)$, the clean speech signal correlation matrix is computed as $\widehat{\mathbf{R}}_{\mathbf{x}}(t) = \widehat{\mathbf{R}}_{\mathbf{y}}(t) - \widehat{\mathbf{R}}_{\mathbf{v}}(t)$. To ensure that the $\widehat{\mathbf{R}}_{\mathbf{x}}(t)$ matrix is positive semidefinite, we apply the eigenvalue decomposition and force all the negative eigenvalues to zero. These estimated correlation matrices are substituted into (6) and (9) to compute the eigenvalue decomposition for the optimal noise reduction filters.

4.1. Harmonic and white noise

In the first set of simulations, we consider the case with narrowband harmonic noise and broadband white noise. The harmonic noise is generated using the following model [42]:

$$v(k) = \sum_{m=1}^{M} A_m \cos(2\pi m k f_0/f_s + \phi_m), \tag{51}$$

where M denotes the model order, $A_m > 0$ and $\phi_m \in [0, 2\pi)$ are the amplitude and phase of the m-th harmonic, $f_0 \in [0, f_s/(2M)]$ is the fundamental frequency, and f_s is the sampling frequency. The rank of the noise signal correlation matrix, $\mathbf{R_v}$, is then $Q = 2M$. In our simulations, we choose $M = 7, A_m = 1, \phi_m = 0$, and $f_0 = 160$Hz for all $m = 1, 2, \ldots, M$. The noisy speech is obtained by adding the generated harmonic noise to the clean speech at a specified SNR level.

In the first simulation, we consider the case with only harmonic noise. The input SNR is 10 dB. To visualize the maximum SNR filter, we take a segment of speech with only 800 samples. Based on this segment, we computed the correlation matrices of the noise and noisy signals using a short-time average. The maximum SNR filter with a length of 110 is computed according to (38). Fig. 1 plots the spectrum of the synthetic noise and also the frequency response of the resulting maximum SNR filter. It is seen that the maximum SNR filter in the given harmonic noise is akin to a comb filter. It has rather small gains at harmonic frequencies where the noise consists of much energy while large gains at frequencies with not much noise energy. The noisy, clean, and enhanced speech signals of this segment are plotted in Fig. 2. It is seen that the enhanced signal is close to the desired clean signal.

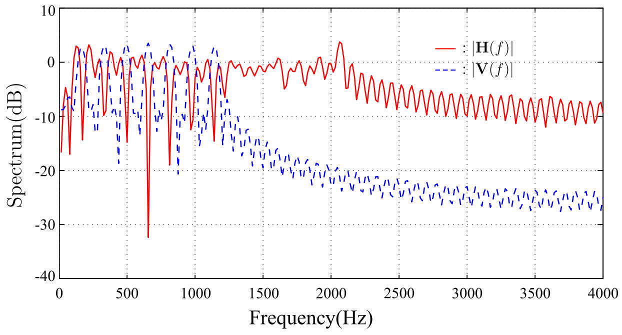


Fig. 1. Spectrum of the synthetic noise vector and the corresponding filter.

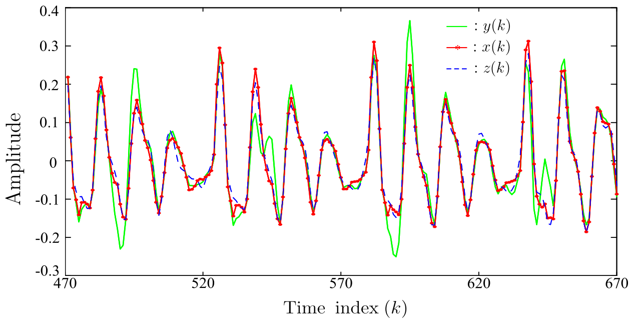


Fig. 2. A segment of the observation noisy signal, $y(k)$, the clean speech signal, $x(k)$, and the enhanced signal, $z(k)$.

Now, we use the recursive method in (49) and (50) to estimate the noisy and noise correlation matrices, based on which we compute the noise reduction filters of length 30. Fig. 3 plots the performance results of the tradeoff filter with three different values of Q' as a function of the forgetting factors (here we assume $\alpha_y = \alpha_v$ for simplicity). It is seen that the output SNR decreases monotonically with the forgetting factor for all the three values of Q'. For $Q' = 1$, the speech distortion index decreases with the forgetting factor. But the value of this index first decreases and then increases for the two cases of $Q' = 3$ and $Q' = 5$. The PESQ score also monoton-

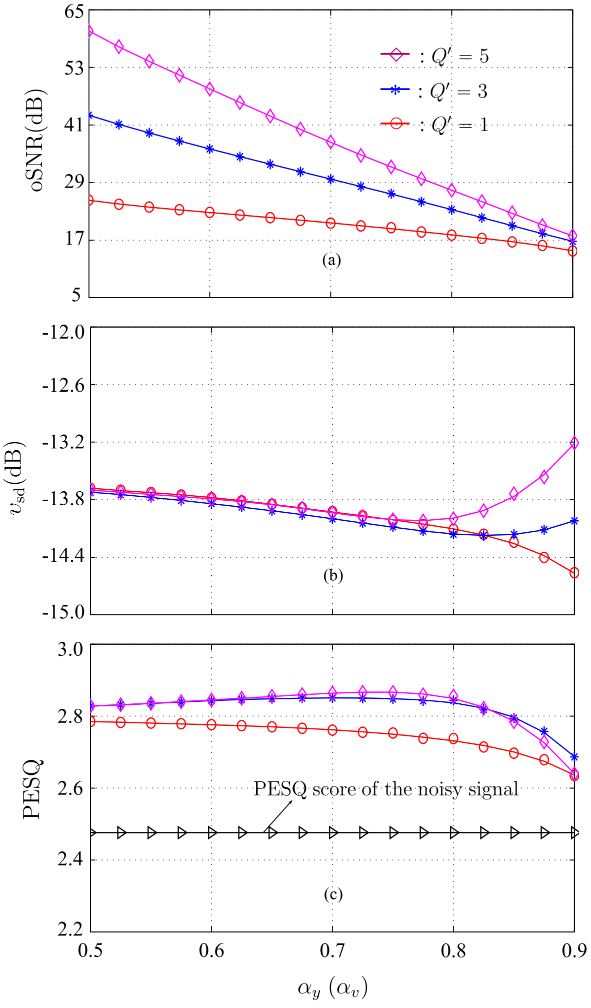


Fig. 3. Performance of the tradeoff filter as a function of the forgetting factor, $\alpha_y (= \alpha_v)$, in the synthetic harmonic noise. (a) Output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $L = 30$, iSNR = 10dB, $Q' = 1, 3, 5$, and the PESQ score of the noisy signal is 2.4813.

ically decreases with the forgetting factor with $Q' = 1$. But for $Q' = 3$ and $Q' = 5$, it first increases and then decreases. The maximum PESQ score is achieved when α_y and α_v are approximately equal to 0.8. So in the following simulations, the value of the two forgetting factors is set to 0.8.

In the second simulation, we investigate the performance of the tradeoff filter with different values of Q' $(0 \leqslant Q' \leqslant Q)$. The noise is the same as in the previous simulation. The filter length, L, is again set to 30 and both α_y and α_v are set to 0.8 according to the previous simulation. Fig. 4 plots the results as a function of Q' in three different input SNR conditions, i.e., iSNR $= 0, 5, 10$ dB. One can see clearly that the output SNR and the speech distortion index increase monotonically with Q' in all the three input SNR conditions, i.e., the larger the value of Q', the more is the noise reduction and so is the speech distortion. This is consistent with the theoretical analysis in SubSection 3.4. The PESQ score first increases with Q', then decreases (because the speech distortion is increasing), indicating that the value of Q' should be properly chosen for optimal perceptual quality.

In practical situations, noise generally contains both narrowband and broadband components. In this simulation, we consider the noise, which is a mixture of a harmonic signal (same as in the previous simulations) and white Gaussian noise. The ratio between the harmonic and white noise is 10 dB. This mixed noise is then added to the clean speech with a 10-dB input SNR to obtain the noisy signal. The results as a function of the filter length for three different values of Q' are plotted in Fig. 5. It is seen that for a given value of Q', the output SNR increases with the filter length. The speech distortion index first decreases. But if the filter length is larger than 30, the speech distortion index does not longer vary much. In contrast, the PESQ score first increases. If the filter length is larger than 30, the PESQ score varies only slightly, which is almost negligible. It is seen that with our simulation setup, the value of L between 20 and 30 is sufficient to achieve good performance.

4.2. Keyboard typing noise

In our daily life, there are various kinds of noise that have rank-deficient correlation matrices, such as door slamming noise, keyboard typing noise, etc. In this section, we recorded some keyboard typing noise with a sampling frequency of 8 kHz. We examine the performance of the optimal filters derived in Section 3 with this noise.

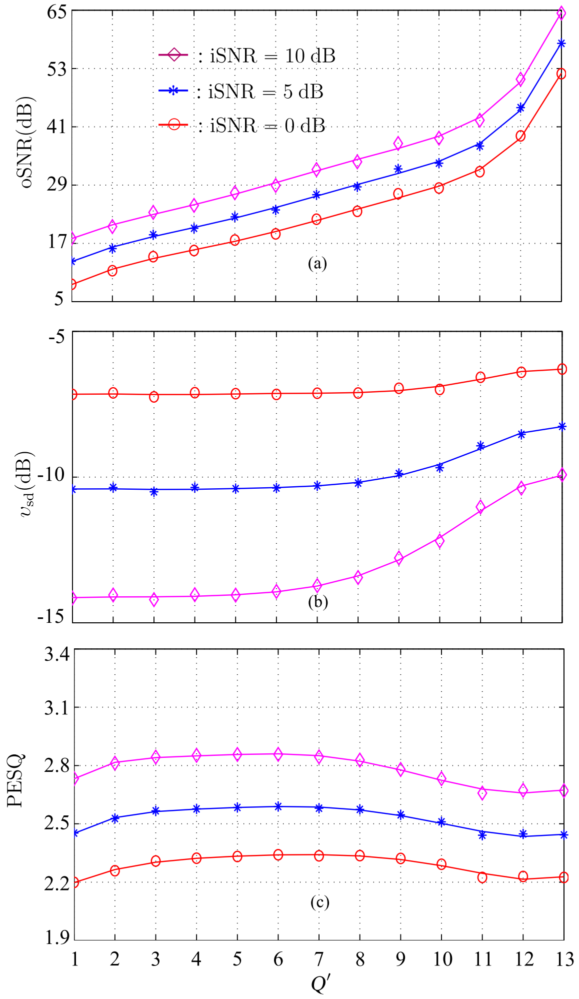


Fig. 4. Performance of the tradeoff filter as a function of Q' in the synthetic harmonic noise. (a) Output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $L = 30, \alpha_y = \alpha_v = 0.8$, iSNR $= 0, 5, 10$dB, and the PESQ scores of the noisy signals in these three iSNR conditions are, respectively, 2.0393, 2.2267, and 2.4813.

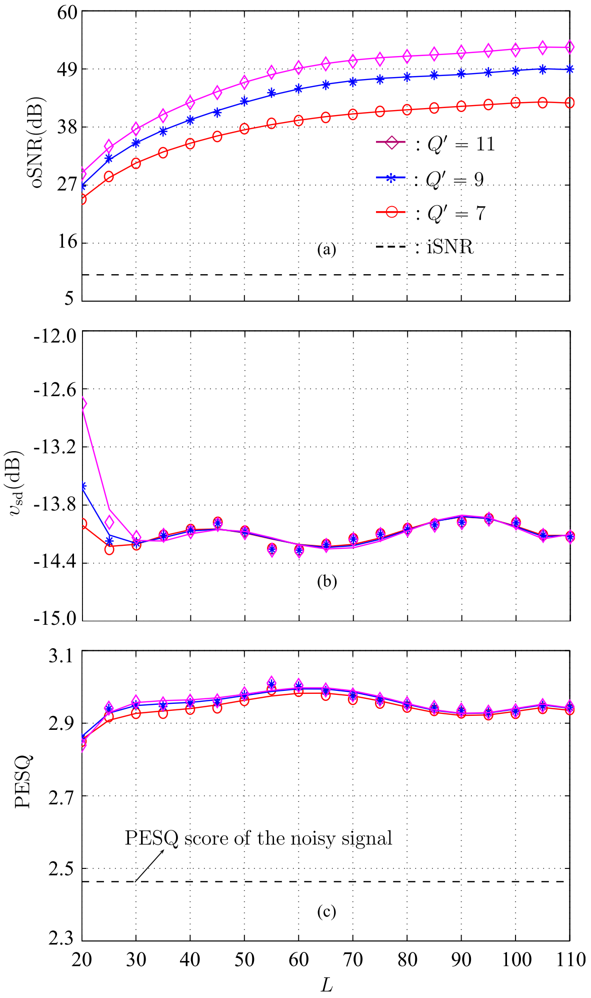


Fig. 5. Performance of the tradeoff filter as a function of the filter length, L, in the synthetic harmonic plus white Gaussian noise (at a ratio of 10 dB). (a) Output SNR, (b) speech distortion index, and (c) PESQ scores. Simulation conditions: $\alpha_y = \alpha_v = 0.8$, iSNR $= 10$dB, $Q' = 7, 9, 11$, and the PESQ score of the noisy signal is 2.4634.

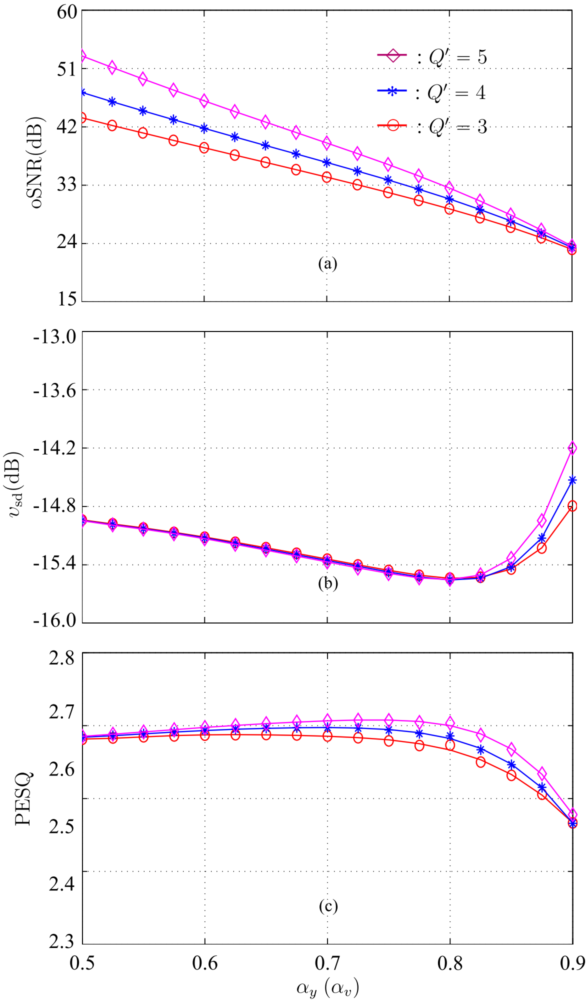


Fig. 6. Performance of the tradeoff filter as a function of the forgetting factor, $\alpha_y (= \alpha_v)$, in the keyboard typing noise. (a) Output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $L = 30$, iSNR = 10dB, $Q' = 3, 4, 5$, and the PESQ score of the noisy signal is 2.0311.

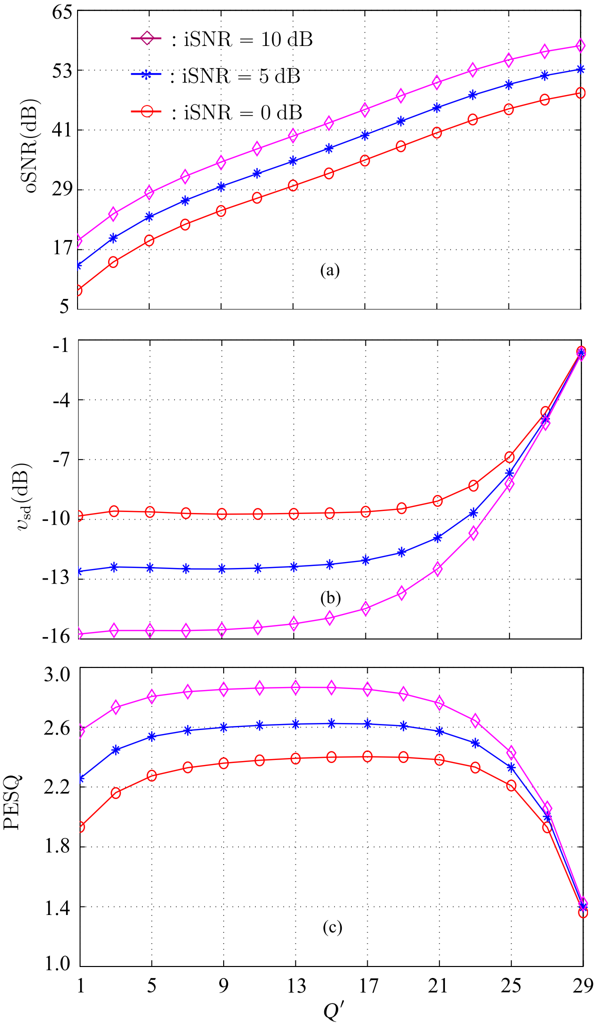


Fig. 7. Performance of the tradeoff filter as a function of Q' in the keyboard typing noise. (a) Output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $L = 30$, $\alpha_y = \alpha_v = 0.8$, iSNR = 0, 5, 10 dB, and the PESQ scores of the noisy signals in the three iSNR conditions are, respectively, 1.1959, 1.5992, and 2.0311.

Fig. 6 plots the performance of the tradeoff filters with three different values of Q' as a function of the forgetting factors (again, it is assumed that $\alpha_y = \alpha_v$). It is seen that significant improvement in terms of SNR and PESQ score is achieved. The trend of the performance as a function of the forgetting factor is similar to that in Fig. 3. Again, the optimal performance is achieved when the value of the forgetting factor is around 0.8.

The performance of the tradeoff filter as a function of the value of Q' is shown in Fig. 7. It is observed that, in all three iSNR conditions, the output SNR and the speech distortion index increase monotonically with the value of Q'. This is reasonable since as the value of Q' is getting larger, the tradeoff filter approaches the maximum SNR filter, which maximizes the output SNR, but the speech distortion with this filter is very large too. As for the PESQ score, it first increases then decreases when the value of Q' increases in the range between 5 and 19. It decreases due to the increasing speech distortion. It is worth mentioning that the tradeoff filter achieves a large gain in the PESQ score, which is greater than 0.9 in all three iSNR conditions. This shows that the designed filter can increase the speech quality significantly.

The results of STOI and LSD for different values of Q' is shown in Table 1. As seen, the developed filter achieves an improvement of the STOI score by more than 0.1, indicating that this filter is able to improve the speech intelligibility as well. Comparing the STOI results with the PESQ scores in Fig. 7(c), one can see that the change of STOI with respect to the value of Q' has a similar trend as PESQ. Similarly, LSD increases when the value of Q', which is similar to the case of the speech distortion index shown in Fig. 7 (b). The LSD results corroborate that the developed filter is capable of making a compromise between noise reduction and speech distortion.

Now, we consider the case where the noise is a mixture of the keyboard typing noise and the white noise. The ratio between the former and the latter is 10 dB. The results are shown in Fig. 8. It is obvious that good performance is achieved when the value of the filter length, L, is in the range between 20 and 30. With $Q' = 8$, the tradeoff filter improves the PESQ score from 2.0 to 2.9, indicating that the developed filter works well even when both narrowband and broadband noises coexist.

In this simulation, we evaluate the performance versus the input SNR in the presence of both keyboard typing and white Gaussion noise. The results are shown in Fig. 9. As seen, both the output SNR and the PESQ score increase while the speech distortion index decreases with the input SNR. The improvement in PESQ decreases as the input SNR increases. This is reasonable as there is less noise

Table 1
STOI score and LSD (in dB) of the tradeoff filter as a function of Q' in the keyboard typing noise. Simulation conditions: $L = 30, \alpha_y = \alpha_v = 0.8$, iSNR = 10 dB. The STOI score and LSD of the noisy signal is 0.8388 and 21.5132 dB respectively.

Q'	1	3	5	7	9
STOI	0.9140	0.9353	0.9411	0.9427	0.9429
LSD	12.3995	13.5356	13.9911	14.1867	14.2895
Q'	11	13	15	17	19
STOI	0.9423	0.9413	0.9394	0.9361	0.9299
LSD	14.3783	14.4789	14.6303	14.8129	15.0880
Q'	21	23	25	27	29
STOI	0.9190	0.8984	0.8608	0.7926	0.6604
LSD	15.4541	16.0362	17.0759	19.1762	23.7667

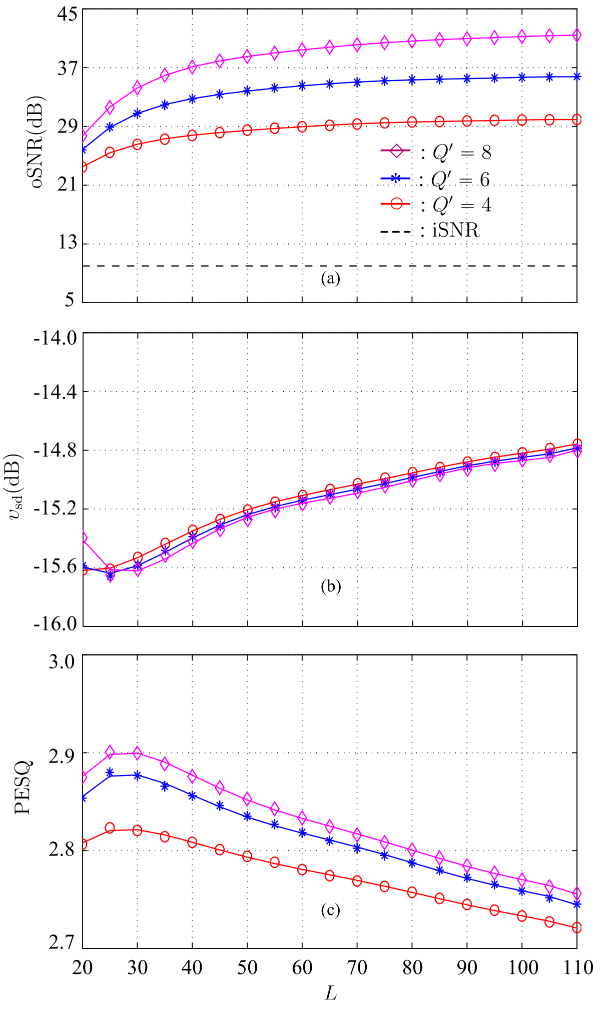


Fig. 8. Performance of the tradeoff filter as a function of the filter length in the keyboard typing plus white Gaussian noise (at a ratio of 10 dB). (a) Output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $\alpha_y = \alpha_v = 0.8$, iSNR = 10 dB, $Q' = 4, 6, 8$, and the PESQ score of the noisy signal is 2.0599.

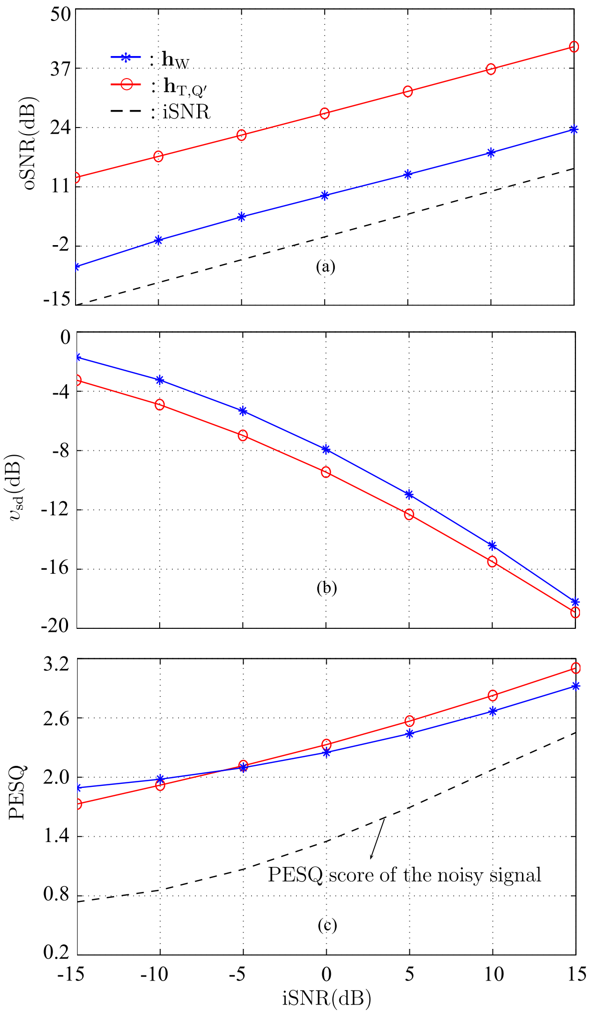


Fig. 9. Performance of the tradeoff filter and Wiener filter as a function of input SNR in the keyboard typing plus white Gaussian noise (at a ratio of 10 dB). (a) Output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $\alpha_y = \alpha_v = 0.8$ and 0.5 for tradeoff filter and Wiener filter respectively, $L = 30$, and $Q' = 10$.

to be reduced as the input SNR increases. When the input SNR is low, the gain in PESQ with the tradeoff filter is close to or even more than 1, which is significant. This, again, proves the effectiveness of the tradeoff filter. We also compared the tradeoff filter with the Wiener filter in (23). The results in Fig. 9 show that the tradeoff filter outperforms the Wiener filter in oSNR, speech distortion and PESQ score when all the parameters of the tradeoff filter are properly chosen.

To visualize the performance, we plot in Fig. 10 the spectrograms of a segment of the clean speech signal, the noisy speech signal, and the enhanced signals with the tradeoff filter and two different values of Q' ($Q' = 1$ and $Q' = 10$). It is clearly seen that both the white noise and the keyboard typing noise are attenuated significantly with the trade filter [as shown in the black boxes in Fig. 10(b), (c), and (d)]. Comparing the spectrograms, one can see that the tradeoff filter with $Q' = 10$ clearly achieves more noise reduction than the filter with $Q' = 1$. Most of the keyboard typing

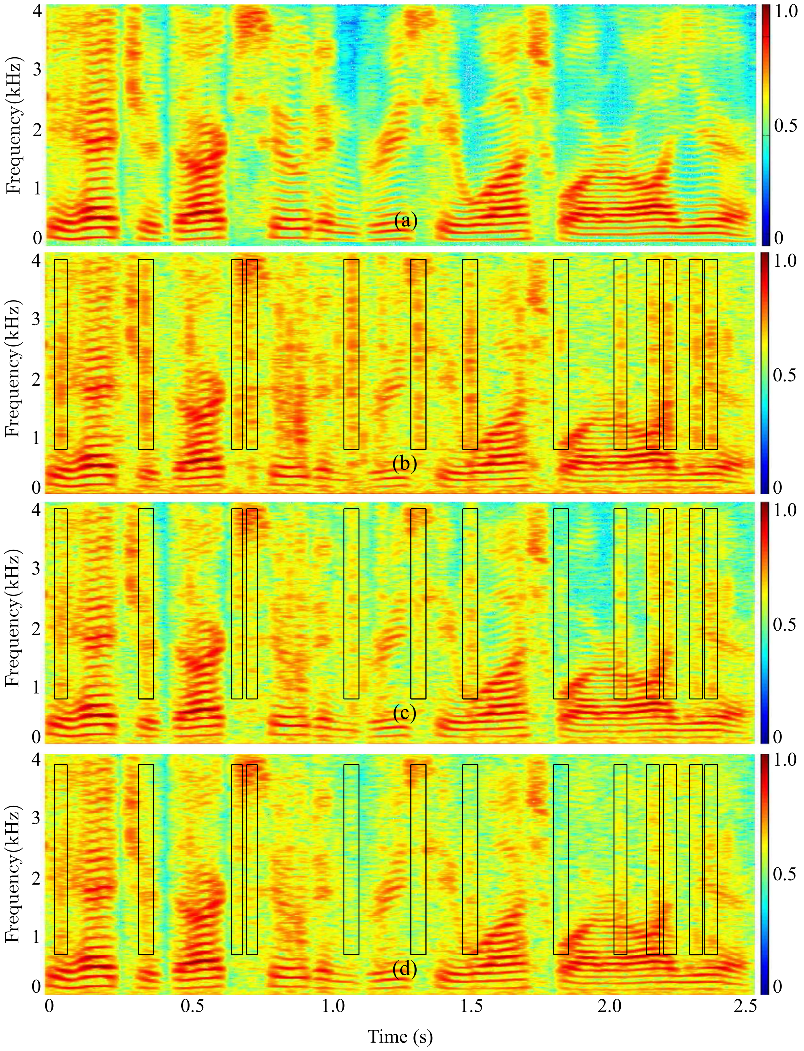


Fig. 10. Spectrograms of a segment of the clean speech, noisy speech signal, and enhanced signals with the tradeoff filter and two different values of Q'. (a) Clean speech signal, (b) noisy speech signal, (c) enhanced signal with $Q' = 1$, and (d) enhanced signal with $Q' = 10$. Simulation conditions: the input SNR is 10 dB, $\alpha_y = \alpha_v = 0.8, L = 30$, and the noise is a mixture of the keyboard typing noise and white Gaussian noise at a ratio of 10 dB.

noise is reduced with $Q' = 10$ while there is still a good amount of residual keyboard typing noise with $Q' = 1$ as shown in the black boxes in Fig. 10(c) and (d). This coincides very well with the theoretical study in Section 3.4 that the output SNR should monotonically increase with the value of Q'.

5. Conclusions

Noise reduction is a challenging problem in acoustic signal processing and voice communications. Various noise reduction algorithms have been developed over the past several decades. However, most of those cannot deal with the case where the noise correlation matrix is rank deficient. In this paper, we presented an approach that combines the principles of the linear filtering and subspace methods to deal with this problem. Specifically, eigenvalue decomposition is applied to the speech and noise correlation matrices. Then, the nullspace of the noise signal is used to design the noise reduction filters. In particular, we discussed the design of the maximum SNR and tradeoff filters. Simulations were carried out to examine the performance of the deduced filters in synthetic harmonic noise, harmonic plus white noise, keyboard typing noise, and keyboard typing noise plus white noise. Significant improvement in both SNR and PESQ was observed, which justifies the effectiveness of the developed approach.

References

[1] Benesty J, Makino S, Chen J. Speech Enhancement. Berlin, Germany: Springer-Verlag; 2005. http://dx.doi.org/10.1007/3-540-27489-8.

[2] Loizou PC. Speech Enhancement: Theory and Practice. Boca Raton Florida: CRC Press; 2007.

[3] Benesty J, Chen J, Huang Y, Cohen I. Noise Reduction in Speech Processing. Berlin, Germany: Springer-Verlag; 2009. http://dx.doi.org/10.1007/978-3-642-00296-0.

[4] Lim JS, Oppenheim AV. Enhancement and bandwidth compression of noisy speech. Proc. IEEE 1979;67(12):1586–604. http://dx.doi.org/10.1109/PROC.1979.11540.

[5] Brandstein M, Ward EDB. Microphone Arrays: Signal Processing Techniques and Applications. Berlin, Germany: Springer-Verlag; 2001. http://dx.doi.org/10.1007/978-3-662-04619-7.

[6] Boll SF. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust., Speech, Signal Process. 1979;27(2):113–20. http://dx.doi.org/10.1109/TASSP.1979.1163209.

[7] Chen J, Benesty J, Huang Y, Diethorn EJ. Fundamentals of noise reduction. In: Benesty J, Sondhi MM, Huang Y, editors. Springer Handbook on Speech Processing and Speech Communication. Berlin, Germany: Springer-Verlag; 2008. p. 843–71. http://dx.doi.org/10.1007/978-3-540-49127-9.

[8] Benesty J, Chen J. Optimal Time-domain Noise Reduction Filters–A Theoretical Study. Berlin, Germany: Springer-Verlag; 2011.

[9] Chen J, Benesty J, Huang Y, Doclo S. New insights into the noise reduction wiener filter. IEEE Trans. Acoust., Speech, Signal Process. 2006;14(4):1218–34. http://dx.doi.org/10.1109/TSA.2005.860851.
[10] Vary P. Noise suppression by spectral magnitude estimation–mechanism and theoretical limits. Signal Process. 1985;8(4):387–400. http://dx.doi.org/10.1109/InertialSensors.2014.7049411.
[11] Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Acoust., Speech, Signal Process. 2001;9(5):504–12. http://dx.doi.org/10.1109/89.928915.
[12] Inoue T, Saruwatari H, Takahashi Y, Shikano K, Kondo K. Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics. IEEE Trans. Acoust., Speech, Signal Process. 2001;19(6):1770–9. http://dx.doi.org/10.1109/TASL.2010.2098871.
[13] Miyazaki R, Saruwatari H, Inoue T, Takahashi Y, Shikano K, Kondo K. Musical-noise-free speech enhancement based on optimized iterative spectral subtraction. IEEE Trans. Acoust., Speech, Signal Process. 2012;20(7):2080–94. http://dx.doi.org/10.1109/TASL.2012.2196513.
[14] Hu K, Wang D. Unvoiced speech segregation from nonspeech interference via casa and spectral subtraction. IEEE Trans. Acoust., Speech, Signal Process. 2011;19(6):1600–9. http://dx.doi.org/10.1109/TASL.2010.2093893.
[15] Cohen I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans. Acoust., Speech, Signal Process. 2003;11(5):466–75. http://dx.doi.org/10.1109/TSA.2003.811544.
[16] McAulay RJ, Malpass ML. Speech enhancement using a soft-decision noise suppression filter. IEEE Trans. Acoust., Speech, Signal Process. 1980;28(2):137–45. http://dx.doi.org/10.1109/TASSP.1980.1163394.
[17] Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust., Speech, Signal Process. 1984;32(6):1109–21. http://dx.doi.org/10.1109/TASSP.1984.1164453.
[18] Ephraim Y. A bayesian estimation approach for speech enhancement using hidden markov models. IEEE Trans. Acoust., Speech, Signal Process. 1992;40(4):725–35. http://dx.doi.org/10.1109/78.127947.
[19] Sameti H, Sheikhzadeh H, Deng L, Brennan RL. Hmm-based strategies for enhancement of speech signals embedded in nonstationary noise. IEEE Trans. Acoust., Speech, Signal Process. 1998;6(5):445–55. http://dx.doi.org/10.1109/89.709670.
[20] Srinivasan S, Samuelsson J, Kleijn WB. Codebook-based bayesian speech enhancement for nonstationary environments. IEEE Trans. Acoust., Speech, Signal Process. 2007;15(2):441–52. http://dx.doi.org/10.1109/TASL.2006.881696.
[21] Ephraima Y, Trees HLV. A signal subspace approach for speech enhancement. IEEE Trans. Acoust., Speech, Signal Process. 2005;3(4):251–66. http://dx.doi.org/10.1109/89.397090.
[22] Jensen SH, Hansen PC, Hansen SD, Sørensen JA. Reduction of broad-band noise in speech by truncated qsvd. IEEE Trans. Acoust., Speech, Signal Process. 1995;3(6):439–48. http://dx.doi.org/10.1109/89.482211.
[23] Mittal U, Phamdo N. Signal/noise klt based approach for enhancing speech degraded by colored noise. IEEE Trans. Acoust., Speech, Signal Process. 2000;8(2):159–67. http://dx.doi.org/10.1109/ICASSP.2000.862115.
[24] Rezayee A, Gazor S. An adaptive klt approach for speech enhancement. IEEE Trans. Acoust., Speech, Signal Process. 2001;9(2):87–95. http://dx.doi.org/10.1109/89.902276.
[25] Hu Y, Loizou PC. A subspace approach for enhancing speech corrupted by colored noise. IEEE Signal Process. Lett. 2002;9(7):204–6. http://dx.doi.org/10.1109/LSP.2002.801721.
[26] Hu Y, Loizou PC. A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans. Acoust., Speech, Signal Process. 2003;11(4):334–41. http://dx.doi.org/10.1109/TSA.2003.814458.
[27] Chen J, Benesty J, Huang Y. Study of the noise-reduction problem in the karhunen-loève expansion domain. IEEE Trans. Acoust., Speech, Signal Process. 2009;17(4):787–802. http://dx.doi.org/10.1109/TASL.2009.2014793.
[28] Benesty J, Jensen JR, Christensen MG, Chen J. Speech Enhancement: A Signal Subspace Perspective. Academic Press; 2014.
[29] Xu Y, Du J, Dai L-R, Lee C-H. An experimental study on speech enhancement based on deeep neural networks. IEEE Signal Process. Lett. 2014;21(1):65–8. http://dx.doi.org/10.1109/LSP.2013.2291240.
[30] Xu Y, Du J, Dai L-R, Lee C-H. A regression approach to speech enhancement based on deep neural networks. IEEE Trans. Acoust., Speech, Signal Process. 2015;23(1):7–19. http://dx.doi.org/10.1109/TASLP.2014.2364452.
[31] Du J, Tu Y, Dai L-R, Lee C-H. A regression approach to single-channel speech separation via high-resolution deep neural networks. IEEE Trans. Acoust., Speech, Signal Process. 2016;24(8):1424–37. http://dx.doi.org/10.1109/TASLP.2016.2558822.
[32] Zhang X-L, Wang D. A deep ensemble learning method for nonaural speech separation. IEEE Trans. Acoust., Speech, Signal Process. 2016;24(5):967–77. http://dx.doi.org/10.1109/TASLP.2016.2536478.
[33] Chen J, Benesty J, Huang Y. A minimum distortion noise reduction algorithm with multiple microphones. IEEE Trans. Acoust., Speech, Signal Process. 2008;16(3):481–93. http://dx.doi.org/10.1109/TASL.2007.914969.
[34] Hansen PC, Jensen SH. Prewhitening for rank-deficient noise in subspace methods for noise reduction. IEEE Trans. Acoust., Speech, Signal Process. 2005;53(10):3718–26. http://dx.doi.org/10.1109/TSP.2005.855110.
[35] Golub GH, Loan CFV. Matrix Computations. third ed. The Johns Hopkins University Press; 1996.
[36] Huang G, Benesty J, Long T, Chen J. A family of maximum snr filters for noise reduction. IEEE Trans. Acoust., Speech, Signal Process. 2014;22(12):2034–47. http://dx.doi.org/10.1109/TASLP.2014.2360643.
[37] Mapping function for transforming raw result scores to MOS-LQO, ITU-T Rec. P.862.1, 2003.
[38] Hu Y, Loizou P. Evaluation of objective quality measures for speech enhancement. IEEE Trans. Acoust., Speech, Signal Process. 2008;16(1):229–38. http://dx.doi.org/10.1109/TASL.2007.911054.
[39] Taal CH, Hendriks RC, Heusdens R, Jensen J. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. IEEE Trans. Acoust., Speech, Signal Process. 2011;19(7):2125–36. http://dx.doi.org/10.1109/TASL.2011.2114881.
[40] Rabiner LR, Juang B-H. Fundamentals of Speech Recognition. Englewood Cliffs, NJ: Prentice-Hall; 1993.
[41] J.S. Garofolo, L.F. Lamel, W.M. Fisher, J.G. Fiscus, D.S. Pallett, N.L. Dahlgren, Darpa timit acoustic phonetic continuous speech corpus, http://www.ldc.upenn.edu/Catalog/LDC93S1.html.
[42] Jensen J, Hansen JHL. Speech enhancement using a constrained iterative sinusoidal model. IEEE Trans. Acoust., Speech, Signal Process. 2001;9(7):731–40. http://dx.doi.org/10.1109/89.952491.