Applied Acoustics 126 (2017) 26–35

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

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

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

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

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

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.,

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

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,

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

is the filtered desired signal, and

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

is the residual noise.

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