Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 78 (2016) 73–83
www.elsevier.com/locate/specom

Single-channel noise reduction via semi-orthogonal transformations and reduced-rank filtering[☆]

Wei Zhang[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*

Received 10 January 2015; received in revised form 22 November 2015; accepted 2 December 2015
Available online 9 February 2016

Abstract

This paper investigates the problem of single-channel noise reduction in the time domain. The objective is to find a lower dimensional filter that can yield a noise reduction performance as close as possible to or even better than that obtained by the full-rank solution. This is achieved in three steps. First, we transform the observation signal vector sequence, through a semi-orthogonal matrix, into a sequence of transformed signal vectors with a reduced dimension. Second, a reduced-rank filter is applied to get an estimate of the clean speech in the transformed domain. Third, the estimate of the clean speech in the time domain is obtained by an inverse semi-orthogonal transformation. The focus of this paper is on the derivation of semi-orthogonal transformations under certain estimation criteria in the first step and the design of the reduced-rank optimal filters that can be used in the second step. We show how noise reduction using the principle of rank reduction can be cast as an optimal filtering problem, and how different semi-orthogonal transformations affect the noise reduction performance. Simulations are performed under various conditions to validate the deduced filters for noise reduction.

Keywords: Noise reduction; Speech enhancement; Semi-orthogonal transformation; Wiener filter; MVDR filter; Reduced-rank filter.

1. Introduction

The problem of single-channel noise reduction is to recover a clean speech signal of interest from its microphone observations (Benesty and Chen, 2011; Benesty et al., 2009; Loizou, 2007). Due to the importance and broad range of applications, a great deal of efforts have been devoted to this problem over the last decades and many algorithms have been developed e.g., Wiener (1949), Boll (1979), Berouti et al. (1979), Lim and Oppenheim (1979), Ephraim and Malah (1984), Trees and Harry (2001). However, these algorithms achieve noise reduction generally by paying a price of adding speech distortion. One exceptional case is the reduced-rank or subspace method, which has the potential to introduce less distortion if the desired signal correlation matrix is rank deficient and this rank is correctly estimated. This paper is, therefore, devoted to the reduced-rank filtering methods.

The idea of "reduced rank" was first developed in the field of signal estimation (Huffel, 1993; Moor, 1993; Scharf, 1991; Scharf and Tufts, 1987; Tufts and Kumaresan, 1982a, 1982b). It was then applied to the noise reduction problem in the so-called subspace approach (Dendrinos et al., 1991), where the singular value decomposition (SVD) of the noisy data matrix was used to estimate and remove the noise subspace and the estimate of the clean signal was then obtained from the remaining subspace. This approach gained more popularity when Ephraim and Van Trees proposed to decompose the covariance matrix of the noisy observation vector (Ephraim and Trees, 1995). The subspace method was found better than the widely used spectral subtraction (Boll, 1979) for noise reduction in the sense that it has less speech distortion with little music residual noise. Today, the principle has been

☆ This work was supported in part by the NSFC under Grant No. 61425005.

* Corresponding author. Tel.: +86 02988491079.
E-mail addresses: weizhang@mail.nwpu.edu.cn (W. Zhang), benesty@emt.inrs.ca (J. Benesty), jingdongchen@ieee.org (J. Chen).

http://dx.doi.org/10.1016/j.specom.2015.12.007
0167-6393/

studied to deal with not only white (Ephraim and Trees, 1995) but also colored noise (Hu and Loizou, 2003; Huang and Zhao, 1998; 2000; Mittal and Phamdo, 2000; Rezayee and Gazor, 2001). Besides the SVD (Moor, 1993; Scharf, 1991; Scharf and Tufts, 1987; Tufts and Kumaresan, 1982a, 1982b) and the eigenvalue decomposition (EVD) (Ephraim and Trees, 1995; Hu and Loizou, 2003), truncated (Q)SVD (Hansen and Jensen, 1998; Jensen et al., 1995) and triangular decompositions (Hansen and Jensen, 2007) were also investigated in the subspace approach. More recent works on reduced-rank filtering can be found in Hansen and Jensen (2013), Nørholm et al. (2014), Zhang et al. (2014).

This paper is also concerned with the application of reduced-rank principle to noise reduction. But unlike most existing work (e.g., Dendrinos et al., 1991; Ephraim and Trees, 1995; Goldstein et al., 1999; 1998; Scharf, 1991; Scharf and Tufts, 1987), which exploits the structure of either the signal data or covariance matrix to find the signal and noise subspaces, this paper develops a more flexible framework. We choose a semi-orthogonal matrix to do data transformation instead of directly decomposing the subspaces. The semi-orthogonal matrix is not unique, and it can be derived under different criteria. The resulting semi-orthogonal matrices represent the characteristic of both the signal and noise, and thus might be used in various conditions. Another contribution of the paper is the derivation of the optimal filters under the reduced-rank framework.

In this framework, noise reduction is achieved in three steps. We first prefilter the full-length observed vector by a semi-orthogonal matrix, resulting in a reduced-dimension vector. In other words, we apply a linear transformation that transforms the observed data vector to a new coordinate system where the basis are defined by the columns of the semi-orthogonal matrix. This is workable because the dimension of the signal subspace is smaller than that of the observed noisy signal space. The second step is to design an optimal reduced-rank filter and apply this filter to get an estimate of the clean speech in the transformed domain. Note that the optimal filter is matrix-valued and the noisy signal is processed by a vector-by-vector basis. The estimate of the clean speech in the time domain is finally obtained by an inverse semi-orthogonal transformation. We will discuss how to derive different semi-orthogonal transformations under certain estimation criteria and how to design different reduced-rank optimal filters. We will also illustrate the flexibility of this new framework in controlling the compromise between noise reduction and speech distortion.

The rest of the paper is organized as follows. In Section 2, the signal model and problem formulation are presented. Section 3 gives the definition of the semi-orthogonal transformation. Then in Section 4, the principle of linear filtering with a rectangular matrix is discussed. Section 5 presents some performance measures for evaluation and analysis of noise reduction. In Section 6, different optimal filters are derived under a given semi-orthogonal transformation. Different semi-orthogonal transformations are discussed in Section 7. Some simulations are presented in Section 8. Finally, conclusions are drawn in Section 9.

2. Signal model and problem formulation

The noise reduction problem considered in this paper is one of recovering the desired speech signal $x(k)$, k being the discrete-time index, of zero mean from the noisy observation (sensor signal) (Benesty and Chen, 2011; Benesty et al., 2009):

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

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

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

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

where

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

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

$$\mathbf{R}_\mathbf{y} \triangleq E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right] = \mathbf{R}_\mathbf{x} + \mathbf{R}_\mathbf{v}, \tag{4}$$

where $E[\cdot]$ denotes mathematical expectation, and $\mathbf{R}_\mathbf{x} \triangleq E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ and $\mathbf{R}_\mathbf{v} \triangleq E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively. The noise correlation matrix, $\mathbf{R}_\mathbf{v}$, is assumed to be full rank, i.e., equal to L. Then, the objective of noise reduction in this paper is to find a "good" estimate of the vector $\mathbf{x}(k)$ from the observation signal vector $\mathbf{y}(k)$ in the sense that the additive noise is significantly reduced while the desired signal is not much distorted.

3. Semi-orthogonal transformation

We recall that $\mathbf{x}(k)$ is the desired signal vector that we want to estimate from the observation signal vector, $\mathbf{y}(k)$.

Let

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_0 & \mathbf{t}_1 & \cdots & \mathbf{t}_{P-1} \end{bmatrix} \tag{5}$$

be a semi-orthogonal matrix of size $L \times P$, i.e., $\mathbf{T}^T\mathbf{T} = \mathbf{I}_P$, where $\mathbf{I}_P$ is the $P \times P$ identity matrix and $P \leq L$. We define the transformed desired signal vector of length P as

$$\begin{aligned} \mathbf{x}'(k) &\triangleq \mathbf{T}^T\mathbf{x}(k) \\ &= \begin{bmatrix} x'_0(k) & x'_1(k) & \cdots & x'_{P-1}(k) \end{bmatrix}^T, \end{aligned} \tag{6}$$