Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 49 (2007) 305–316
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

On the optimal linear filtering techniques for noise reduction

Jingdong Chen [a,*], Jacob Benesty [b], Yiteng (Arden) Huang [a]

[a] *Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Room 2D-534, Murray Hill, NJ 07974, USA*
[b] *Université du Québec, INRS-EMT, 800 de la Gauchetière Ouest, Suite 6900, Montréal, Que., Canada H5A 1K6*

Received 23 May 2006; received in revised form 4 December 2006; accepted 4 February 2007

Abstract

Noise reduction, which aims at extracting the clean speech from noisy observations, has plenty of applications. It has attracted a considerable amount of research attention over the past several decades. Although many methods have been developed, the most widely used one, by far, is the optimal linear filtering technique, which achieves clean speech estimate by passing the noisy observation through an optimal linear filter/transformation. The representative algorithms of this include Wiener filtering, spectral restoration, subspace method, etc. Many experiments have been carried out, from various points of view, to show that the optimal filtering technique can reduce the level of noise that is present in the speech signal and improve the corresponding signal-to-noise ratio (SNR). However, there is not much theoretical justification so far for the noise reduction and SNR improvement. This paper attempts to provide a theoretical analysis on the performance (including noise reduction, speech distortion, and SNR improvement) of the optimal filtering noise-reduction techniques including the time-domain causal Wiener filter, the subspace method, and the frequency-domain subband Wiener filter. We show that the optimal linear filter, regardless of how we delineate it, can indeed reduce the level of noise (but at a price of attenuating the desired speech signal). Most importantly, we prove that the *a posteriori* SNR (defined after the optimal filtering) is always greater than, or at least equal to the *a priori* SNR, which reveals that the optimal linear filtering technique is indeed able to make noisy speech signals cleaner. We will also discuss the bounds for noise reduction, speech distortion, and SNR improvement.

1. Introduction

Since we live in a natural environment where noise is inevitable and ubiquitous, speech signals can seldom be recorded in pure form and are generally contaminated by acoustic background noise. As a result, the microphone signals have to be "cleaned up" with digital signal processing tools before they are stored, transmitted, or played out.

The cleaning process, which is often referred to as either noise reduction or speech enhancement, can be achieved in many different ways, such as beamforming, adaptive cancellation, temporal filtering, spatial–temporal filtering, etc. The most widely used technique thus far, however, is the single-channel optimal linear filtering approach, which achieves clean speech estimate by passing the noisy observation through an optimal linear filter/transformation. A variety of such algorithms have been developed. They principally fall into one of the following four categories: *Wiener filter*, *spectral restoration*, *subspace method*, and *parametric method*.

Wiener filter: This method restores the desired speech signal by passing the noisy speech through a finite impulse response (FIR) filter whose coefficients are estimated by minimizing the mean square error (MSE) between the clean speech and its estimate (Widrow and Stearns, 1985). The Wiener filter can also be delineated in the frequency domain, resulting in various derivative techniques such as spectral subtraction (Boll, 1979; McAulay and Malpass, 1980; Lim, 1983; Lim and Oppenheim, 1979), parametric Wiener filter (Lim and Oppenheim, 1979; Vary, 1985; Etter and Moschytz, 1994; Chen et al., 2003; Diethorn, 2004), etc.

* Corresponding author. Tel.: +1 908 582 5044; fax: +1 908 582 7308.
E-mail addresses: jingdong@research.bell-labs.com (J. Chen), benesty@emt.inrs.ca (J. Benesty), arden@research.bell-labs.com (Yiteng (Arden) Huang).

0167-6393/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2007.02.002

Spectral restoration: In the frequency domain, a speech signal can be factorized into spectral amplitude and phase components. From perceptual point of view, the former is considerably more important than the latter (Lim and Oppenheim, 1979; Vary, 1985; Wang and Lim, 1982). Therefore the spectral-restoration technique recovers only the spectral amplitude (or spectral envelope) of the clean speech from that of the corrupted speech while neglecting the phase corruption (Ephraim and Malah, 1984, 1985; Virag, 1999; Chang and O'Shaughnessy, 1991).

Signal subspace: This method decomposes the vector space of the noisy speech into two orthogonal subspaces using the Karhunen–Loève transform (KLT): one is composed of both speech and noise and the other consists of noise component only. This is possible because it has been proven that the clean speech can be described with a low-rank model. After decomposition, the speech signal is estimated by removing the noise subspace, and cleaning the speech-plus-noise subspace (Ephraim and Van Trees, 1995; Dendrinos et al., 1991; Hansen, 1997; Lev-Ari and Ephraim, 2003; Rezayee and Gazor, 2001; Mittal and Phamdo, 2000; Hu and Loizou, 2003).

Parametric method: It is well known that a speech signal can be modelled as an autoregressive (AR) process. Therefore, noise reduction can be formulated as a parameter estimation problem with its objective to estimate the AR model parameters of the clean speech from the noisy observations (Paliwal and Basu, 1987; Gibson et al., 1991; Gannot et al., 1998).

Although so many optimal filtering algorithms have been developed for noise reduction, there has been remarkably little (if any) theoretical analysis of their performance. The reason may be attributed to the difficulty in quantizing the combinatorial effect between noise reduction and speech distortion. Most existing performance studies have been experimental, including: (1) ranking the mean opinion scores, (2) examining the SNR improvements, (3) inspecting the speech spectrograms, and (4) comparing the noise levels before and after the application of an algorithm. While the results are very helpful for us to understand how the algorithms behave in the specified conditions, the experimental evaluation alone is not enough to justify the algorithms. A more thorough theoretical analysis is important and imperative. Recently, we performed some analysis of the time-domain Wiener filter and proved that, as long as we have an accurate estimate of the statistics of the noisy speech and the noise signal, SNR improvement is guaranteed, no matter whether the noise is white or colored (Chen et al., 2006; Benesty et al., 2005). This paper presents our continued efforts on this topic. The main contribution of this paper is a theoretical analysis on the performance of the optimal [from the minimum-mean-square error (MMSE) sense] filtering techniques including the time-domain causal, the frequency-domain noncausal, and the constrained (subspace) Wiener filters. We show that the optimal filter, regardless of how we delineate it, can indeed reduce the level of noise. Most importantly, we prove that the *a posteriori* SNR is always greater than, or at least equal to the *a priori* SNR, provided that the statistics of the noisy speech and noise signals are accurately estimated. Also discussed are the lower and upper bounds for noise reduction, speech distortion, and SNR improvement.

2. Signal model and problem formulation

The noise-reduction problem considered in this paper is to recover a speech signal of interest $x(n)$ from the noisy observation

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

where $v(n)$ is the unwanted additive noise, which is assumed to be a zero-mean random process (white or colored) and uncorrelated with $x(n)$. This signal model can also be formulated in other forms. For example, in vector/matrix form, it is written as

$$\boldsymbol{y}(n) = \boldsymbol{x}(n) + \boldsymbol{v}(n), \tag{2}$$

where

$$\boldsymbol{y}(n) = [\, y(n) \quad y(n-1) \quad \cdots \quad y(n-L+1)\,]^{\mathrm{T}}$$

is a vector consisting of the L most recent samples of the noisy speech signal, superscript T denotes transpose of a vector or a matrix, and $\boldsymbol{x}(n)$ and $\boldsymbol{v}(n)$ are defined in a similar way to $\boldsymbol{y}(n)$. In this case, the noise-reduction problem is formulated as one of estimating $\boldsymbol{x}(n)$ from the observation $\boldsymbol{y}(n)$.

If applying the L-point discrete Fourier transform (DFT) to both sides of (2), we have the following relationship in the frequency domain:

$$Y(n, j\omega_k) = X(n, j\omega_k) + V(n, j\omega_k), \tag{3}$$

where

$$Y(n, j\omega_k) = \sum_{l=0}^{L-1} w(l) y(n-L+l+1) \mathrm{e}^{-j\omega_k l}$$

is the short-time DFT of the noisy speech at time instant n, $\omega_k = 2\pi k/L$, $k = 0, 1, \ldots, L-1$, $w(l)$ is a window function (e.g. Hamming window, Hann window) applied to the frame signal for better spectral estimation, and $X(n,j\omega_k)$ and $V(n,j\omega_k)$ are the short-time DFTs for the clean speech and the noise signal, defined in a similar way to $Y(n,j\omega_k)$. Based on this relationship, the noise-reduction problem can be expressed in the frequency domain as one of estimating $X(n,j\omega_k)$ from $Y(n,j\omega_k)$.

3. Time-domain causal Wiener filter and its performance

The Wiener filter is one of the most fundamental approaches for noise reduction, which can be formulated either in the time or in the frequency domains. In the time-domain Wiener filter, an estimate of the clean speech