Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 52 (2010) 427–439
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

On widely linear Wiener and tradeoff filters for noise reduction

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

[a] INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, Quebec, Canada H5A 1K6
[b] WeVoice, Inc., 9 Sylvan Dr., Bridgewater, NJ 08807, USA

Received 13 August 2009; received in revised form 25 November 2009; accepted 3 February 2010

Abstract

Noise reduction is often formulated as a linear filtering problem in the frequency domain. With this formulation, the core issue of noise reduction becomes how to design an optimal frequency-domain filter that can significantly suppress noise without introducing perceptually noticeable speech distortion. While higher-order information can be used, most existing approaches use only second-order statistics to design the noise-reduction filter because they are relatively easier to estimate and are more reliable. When we transform nonstationary speech signals into the frequency domain and work with the short-time discrete Fourier transform coefficients, there are two types of second-order statistics, i.e., the variance and the so-called pseudo-variance due to the noncircularity of the signal. So far, only the variance information has been exploited in designing different noise-reduction filters while the pseudo-variance has been neglected. In this paper, we attempt to shed some light on how to use noncircularity in the context of noise reduction. We will discuss the design of optimal and suboptimal noise reduction filters using both the variance and pseudo-variance and answer the basic question whether noncircularity can be used to improve the noise-reduction performance.

Keywords: Noise reduction; Wiener filter; Widely linear Wiener filter; Circularity; Noncircularity

1. Introduction

Noise reduction, which aims at estimating the desired clean speech signal from noisy observations, is a very important problem and has attracted a significant amount of research and engineering attention over the past few decades (Benesty et al., 2005, 2009; Loizou, 2007; Vary and Martin, 2006; Huang et al., 2006). Typically, the noise-reduction process is formulated as a filtering problem where the clean speech estimate is obtained by passing the noisy speech through a noise-reduction filter. With such a formulation, the core issue of noise reduction becomes how to design an optimal filter that can fully exploit the speech and noise statistics to achieve maximum noise suppression without introducing perceptually noticeable speech distortion. While good filters can be designed in the time domain, most widely used approaches so far work in the frequency domain. The reason for working in the frequency domain are multiple, including (but not limited to): (1) the filtering process can be implemented very efficiently thanks to the fast Fourier transform; (2) the filters at different frequencies (or frequency bands) can be designed and handled independently of each other, which offers tremendous flexibility in dealing with colored noise; and (3) most of our knowledge and understanding of speech production and perception is related to frequencies, so in the frequency domain, our knowledge can be easily used to help optimize noise-reduction performance.

When we work in the frequency domain, we generally deal with complex random variables even though the original time-domain signals are real in the context of speech applications. The main concern, then, is how to design the optimal noise-reduction filters that can fully exploit

* Corresponding author. Address: 9 Iroquois Trail, Branchburg, NJ 08876, USA. Tel.: +1 908 722 5945.
E-mail addresses: benesty@emt.inrs.ca (J. Benesty), jingdongchen@ieee.org (J. Chen), ardenhuang@gmail.com (Y. (Arden) Huang).

0167-6393/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2010.02.003

the different statistics of the complex components obtained via the short-time Fourier transform (STFT). Theoretically, all the different orders of statistics should be considered during the design of the optimal noise-reduction filter. In practice, however, higher-order (higher than 2) statistics are in general difficult to estimate, and as a result, most of today's noise-reduction algorithms consider only second-order statistics. For a zero-mean complex random variable, there are two basic types of second-order statistics depending on whether the random variable is circular or noncircular.

A complex random variable A is said to be circular if its probability density function (PDF) is the same as the PDF of Ae^{jr} (Amblard et al., 1996a, Amblard et al., 1996b), where j and r are the imaginary unit ($j^2 = -1$) and any real number, respectively. This is equivalent to saying that the PDF of a circular complex random variable (CCRV) is a function of the product AA^* only (Amblard et al., 1996a), where $*$ denotes complex conjugation. An important consequence of this is that the only nonnull moments and cumulants of a CCRV are the moments and cumulants constructed with the same power in A and A^* (Amblard et al., 1996a). Now let us confine our discussion and study to the second-order issues. With the general definition of circularity, we can readily define the second-order circularity: a zero-mean complex random variable A is said to be second-order circular if its pseudo-variance is equal to zero, i.e., $E(A^2) = 0$, where $E(\cdot)$ denotes mathematical expectation and $E(AA^*) = E(|A|^2) \neq 0$. This indicates that the second-order behavior of a CCRV is well described by its variance. Note that the Fourier components of stationary signals are CCRVs (Picinbono et al., 1994). Another powerful aspect of the second-order CCRV is that the classical linear estimation theory for real random variables can easily be applied to CCRVs. As a matter of fact, most of the existing frequency-domain noise-reduction filters are derived based on the classical mean-squared estimation approach and use only the variance information while assuming that $E(A^2) = 0$.

However, the STFT coefficients of a nonstationary signal like speech are not circular variables. To illustrate this, we take a speech signal that is recorded from a female speaker with an 8-kHz sampling rate and a 16-bit quantization and partition it into overlapping frames. The overlapping factor is 75% and the frame length is 8 ms. Each frame is then transformed into the frequency domain using a 64-point FFT. For each frequency band (except the 1st and 33rd bands where the coefficients are real), we treat the coefficients as a complex random variable (for ease of exposition, let us use A to denote this random variable) and estimate its variance and pseudo-variance. Because speech is nonstationary, we cannot simply replace the mathematical expectation with a sample average. Instead, we use the recursive estimator given in Eq. (88) of (Chen et al., 2006) to estimate both the variance and pseudo-variance (more discussion on how to estimate the variance and pseudo-variance parameters will be given in Section 7).

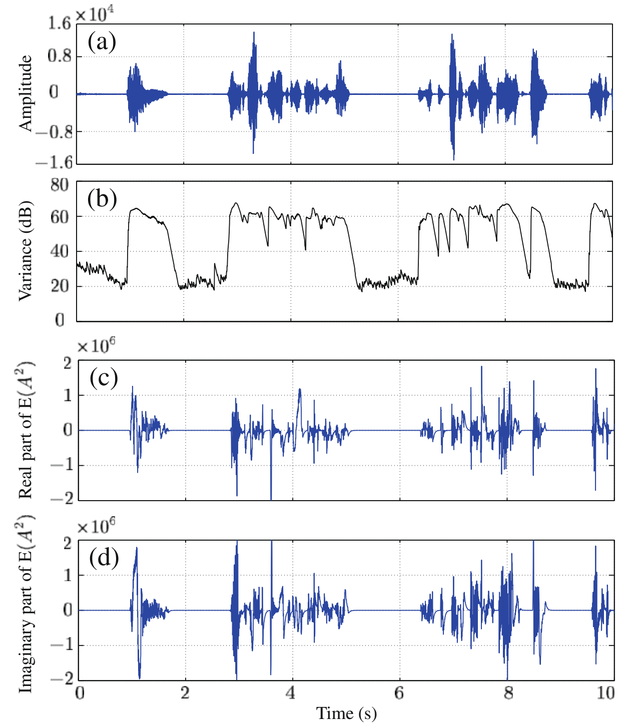

Fig. 1. Illustration of the noncircularity of the STFT coefficients of a speech signal: (a) a speech signal; (b) the $E(|A|^2)$ estimate; (c) the real part of the $E(A^2)$ estimate; and (d) the imaginary part of the $E(A^2)$ estimate.

Fig. 1 plots the estimation results for the 2nd frequency band. It is clearly seen that the pseudo-variance $E(A^2)$ of the STFT coefficients of the speech signal are not zero, so STFT coefficients of speech signals are noncircular random variables. Many natural questions then arise: is the noncircularity useful for noise reduction? If so, how do we use the noncircularity? How much it can improve noise-reduction performance? This paper attempts to answer these questions. We will study and show how to fully exploit the second-order statistics of a noncircular complex random variable (see Neeser and Massey, 1993; Schreier and Scharf, 2003 for a complete description of the second-order behavior of a complex noncircular random variable) for noise reduction. We will investigate the use of the so-called widely linear (WL) mean-squared estimation theory (Picinbono et al., 1995; Eriksson et al., 2009; Mandic and Goh, 2009; Ollila, 2008) to formulate noise-reduction algorithms in the frequency domain and explain the benefits that can be achieved with this new formulation.

The rest of this paper is organized as follows. In Section 2, we formulate the single-channel noise reduction problem in the STFT domain and give some useful definitions and explanations that will be of great help for the rest of the paper. Section 3 explains the different performance measures for noise reduction with WL estimation. In Section 4, we write the WL mean-squared error (MSE), which is a simple and powerful tool for deriving the different optimal WL filters. In Section 5, we derive the WL Wiener filter and explains its differences from the classical Wiener filter. Section 6 deals with the WL and classical tradeoff filters.