Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 35 (2016) 93–115

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

New features for speech enhancement using bivariate shrinkage based on redundant wavelet filter-banks

Hamid Reza Tohidypour [a,*], Seyed Mohammad Ahadi [b]

[a] Electrical and Computer Eng. Dept., University of British Columbia, Vancouver, BC, Canada
[b] Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran

Received 24 February 2014; received in revised form 20 March 2015; accepted 1 June 2015
Available online 24 June 2015

Abstract

In most of the wavelet based speech enhancement methods, it is assumed that the wavelet coefficients are independent of each other. However, investigating the joint histogram of the wavelet coefficients reveals some dependencies among them. In this regard, Sendur proposed a probability density function (pdf) that models the relation between a wavelet coefficient of image signal and its parent. Then, this pdf is utilized to propose a bivariate shrinkage function which uses the dependencies between the child–parent wavelet coefficients of Image signals to enhance the noisy images. In this paper, we intend to find wavelet structures which are more suitable for speech enhancement based on bivariate shrinkage. We show that the dependencies between the child–parent wavelet coefficients can only be modeled rather easily up to two stages of two-channel discrete wavelet transform using the Sendur's pdf. However, the bivariate shrinkage function works better in three-channel redundant wavelet filter-bank with dilation 2, since it has a joint distribution which is similar to the Sendur's pdf up to the fourth stage of decomposition for speech signals. Furthermore, we show that three-channel higher density wavelet obtained by eliminating the downsampling part of the third channel is more suitable for the bivariate shrinkage function when it is utilized for speech enhancement. Then, appropriate filter values for three-channel higher density wavelet filter-bank are found. Moreover, we propose four-channel double density discrete wavelet filter-bank which leads to some improvement in speech enhancement results. Since the probability of speech presence is higher in lower frequencies, we suggest level-dependent bivariate shrinkage. Finally, Sendur bivariate shrinkage is optimized for speech enhancement and new methods are proposed by combining former successful methods with the bivariate shrinkage function.
© 2015 Elsevier Ltd. All rights reserved.

Keywords: Wavelet transform; Speech enhancement; Redundant filter-banks; Four channel double density discrete wavelet; Three-channel higher density discrete wavelet; Bivariate wavelet shrinkage; Zero moments

1. Introduction

The applications of speech signal processing such as speech recognition, speech enhancement and speech coding have been increased recently. In traditional methods, features are obtained by Fourier transform. However, these methods have shortcomings in analyzing non-stationary signals (Wu and Lin, 2009). Short Time Fourier Transform

* Corresponding author. Present address: Department of Electrical and Computer Engineering, University of British Columbia, 2332 Main Mall, Vancouver, BC, Canada V6T 1Z4. Tel.: +1 604 822 2872; fax: +1 604 822 5949.
E-mail address: htohidyp@ece.ubc.ca (H.R. Tohidypour).

http://dx.doi.org/10.1016/j.csl.2015.06.004
0885-2308/© 2015 Elsevier Ltd. All rights reserved.

has uniform time and frequency resolutions (de Andrade Bresolin et al., 2008). Hence, it is difficult to detect sudden changes and transient parts of the signals (de Andrade Bresolin et al., 2008; Farooq and Datta, 2001).

In the past two decades, the wavelet transform has become a strong tool in speech processing (Favero, 1994). This transform has the advantage of alterable window length in time domain. The wavelet transform has higher frequency resolution in low-frequency parts of the signal, while it has higher time resolution in high frequency parts of the signal (Xueying and Zhiping, 2004). On the other hand, different basic functions can be used in wavelet transform. These basic functions are called mother wavelets. The mother wavelets give us ability to design wavelets based on our needs. In addition, wavelet transform has multi-resolution property (Tufekci et al., 2006).

One of the major application of the wavelet transform is signal denoising. In 1995, Donoho proposed a method for wavelet denoising based on the thresholding approach (mentioned in Ghanbari and Karami, 2006). Here, a threshold was used to separate components including speech from those including noise only (Jafer and Mahdi, 2003; Sheikhzadeh and Abutalebi, 2001; Wang et al., 2006). In most algorithms based on Donoho's thresholding, the wavelet coefficients were assumed to be independent (Sendur and Selesnick, 2002). In Bahoura and Rouat (2006), researchers proposed time-scale adaption method to improve speech denoising based on Donoho's approach. In their method (Bahoura and Rouat, 2006), wavelet packet transform was utilized to analyze speech signals. In order to improve Donoho's thresholding, a level-dependent threshold was used. Then a new method was proposed to find noise samples in each stage of decomposition (Bahoura and Rouat, 2006). In this method, first, speech samples were categorized into two categories, namely, the speech samples and the noisy samples. Then, appropriate threshold were suggested for speech samples and noisy samples. In this study, lower threshold values were considered for speech samples in comparison with noisy samples (Bahoura and Rouat, 2006).

In most wavelet-based speech enhancement methods (i.e. Donoho's thresholding), it is assumed that the wavelet coefficients are independent. Therefore, each of the enhanced coefficients is calculated independently. However, Sendur and Selesnick (2002) have shown that there are significant dependencies among wavelet coefficients of the image signal. In their studies, first the dependencies between a wavelet (child) coefficient and its parent are modeled by a non Gaussian probability distribution function (pdf). By using this pdf, they find a new shrinkage function known as bivariate shrinkage function. This shrinkage function uses the child and parent wavelet coefficients to eliminate the noisy parts of the wavelet coefficient. It is reported that this assumption improves the performance of the wavelet-based methods for image denoising (Sendur and Selesnick, 2002).

In order to make appropriate wavelet based denoising methods, in addition to wavelet shrinkage functions, the properties of different wavelet structures are needed to be considered. For instance, down-sampling operator of the discrete wavelet makes this type of wavelet a time variant transform (Farooq and Datta, 2001; Weickert et al., 2008; Abdelnour, 2005; Abdelnour and Selesnick, 2005; Selesnick and Abdelnour, 2004; Selesnick, 2001, 2006; Tohidypour et al., 2010a,b). Most wavelet transforms, which have been used for speech processing in the past, are based on the critically sampled filter-banks (Tohidypour et al., 2012; Lebrun and Selesnick, 2004). In these filter-banks, the number of input samples remains the same in transform domain. Therefore, there is no redundancy in these systems. But these filters are not time-invariant (Tohidypour et al., 2010a,b), because of the down-sampling factors. This means that, for example, if a signal is down-sampled by 2, a sample of signal among two samples is chosen. Therefore, with a small signal displacement, the wavelet coefficients may be changed. This time-variant property decreases the ability of noise estimation. Previous studies have shown that the shift-variant property of critically sampled wavelet filter-banks has some disadvantages in the case of noisy speech with periodic noise (Weickert et al., 2008). Therefore, nearly-shift-invariant wavelet structures are considered for denoising based on wavelet transform. New wavelet transform based on M-channel filter-banks gives us more degrees of freedom in designing nearly-shift-invariant wavelet structures. In this regard, Selesnick and Abdelnour suggest appropriate filters for some redundant M-channel wavelet filter-banks, which are nearly-shift-invariant (Abdelnour, 2005; Abdelnour and Selesnick, 2005; Selesnick and Abdelnour, 2004; Selesnick, 2001, 2006). Moreover, two types of the four channel redundant wavelet filter-banks are proposed by Selesnick and Tohidypour (Abdelnour and Selesnick, 2005; Tohidypour et al., 2012). The last one was called four-channel higher density wavelet filter-bank (Tohidypour et al., 2012). Note, the redundant transform is a transform that expands C point input signal to D point output, while $D > C$ (Selesnick, 2006).

Review of the previous works reveal that in most wavelet based speech enhancement methods, critically sampled discrete wavelet transforms or critically sampled wavelet packets were used. Also, in most of the previous studies, Donoho's threshold was used for noise estimation (Ghanbari and Karami, 2006; Jafer and Mahdi, 2003; Sheikhzadeh and Abutalebi, 2001; Wang et al., 2006; Bahoura and Rouat, 2006).

Download English Version:

https://daneshyari.com/en/article/558248

Download Persian Version:

https://daneshyari.com/article/558248

Daneshyari.com