Speech Communication 86 (2017) 64–74

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Rayleigh modeling of teager energy operated perceptual wavelet packet coefficients for enhancing noisy speech

Md Tauhidul Islam[a], Celia Shahnaz[b,*], Wei-Ping Zhu[c], M. Omair Ahmad[c]

[a] Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas 77843, United States
[b] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[c] Department of Electrical and Computer Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada

ARTICLE INFO

Article history:
Received 25 January 2015
Revised 3 November 2016
Accepted 8 November 2016
Available online 16 November 2016

Keywords:
Speech enhancement
Perceptual wavelet packet transform
Teager energy
Rayleigh
Kullback–Liebler divergence

ABSTRACT

For enhancing noise corrupted speech, Rayleigh modeling of Teager energy (TE) operated perceptual wavelet packet (PWP) coefficients of the noisy speech is proposed in this paper. In order to obtain an enhanced speech, a threshold is derived and applied on the PWP coefficients by employing a custom thresholding function, which is designed based on a combination of μ-law and semisoft thresholding functions. The effectiveness of the proposed method is evaluated for car and multi-talker babble noise corrupted speech signals through performing extensive simulations using the NOIZEUS database. The proposed method is found to outperform some of the state-of-the-art speech enhancement methods not only at high but also at low levels of SNR in terms of standard objective measures and subjective evaluations including formal listening tests.

1. Introduction

Determination of a signal that is corrupted by additive or multiplicative noise has been of interest because of its importance in both theoretical and practical applications. The main interest is to recover the real signal from the noise-mixed data received from microphone, ECG machine, radar, mobile phone or any other sound devices. Our aim is to make the recreated signal close to the original one. The use of such operation has application in broad area of speech communication applications, such as mobile telephony, speech coding and recognition, and hearing aid devices (Deller et al., 2000; Loizou, 2007; O'Shaughnessy, 2000).

Over decades, several methods have been developed for noise reduction and speech enhancement. We can divide these methods in mainly three categories based on their domains of operation, namely time domain, frequency domain and time-frequency domain. Time domain methods include the subspace approach (Ephraim and Trees, 1995; Hu and Loizou, 2003), frequency domain methods include methods based on discrete cosine transform (Chang, 2005), spectral subtraction (Boll, 1979; Yamashita and Shimamura, 2005), minimum mean square error (MMSE) estimator (Ephraim and Trees, 1995; Loizou, 2007), Wiener filtering (Almajai and Milner, 2011; Jebara, 2006) and time frequency-domain methods involve the employment of the family of wavelets (Bahoura and Rouat, 2001; Donoho, 1995; Ghanbari and Mollaei, 2006; Johnson et al., 2007; Shao and Chang, 2007). All these methods have their advantages and disadvantages. The methods which use thresholding as the means of removing noise, universal thresholding (Donoho, 1995), Stein's unbiased risk estimator (SURE) (Luisier et al., 2007), wavelet packets filter (WPF) based thresholding (Bahoura and Rouat, 2001) and BayesShrink (Chang et al., 2000) are the prominent ones. In universal thresholding method, a common threshold derived from noise power is used to threshold the wavelet coefficients. SURE applies Steins uncertainty and BayesShrink applies Bayes principle to determine the threshold. WPF is a modified version of universal threshold method with speech and silent frame detection ability.

In this paper, instead of direct employment of the TE operator on the noisy speech, we apply the TE operator on the PWP coefficients of the noisy speech. The main contribution of this paper lies in the fitting of a Rayleigh pdf suitable for modeling the TE operated PWP coefficients of noisy speech and then using it to analytically determine an appropriate adaptive threshold. Designing a custom thresholding function based on the idea of combining μ-law and semisoft thresholding functions is another contribution of the paper. The threshold thus obtained from the statistical modeling is applied on the PWP coefficients of noisy speech by employing the custom thresholding function in order to obtain an enhanced speech.

* Corresponding author.
E-mail addresses: tauhid@tamu.edu (M.T. Islam), celia.shahnaz@gmail.com (C. Shahnaz), weiping@ece.concordia.ca (W.-P. Zhu), omair@ece.concordia.ca (M.O. Ahmad).

http://dx.doi.org/10.1016/j.specom.2016.11.002
0167-6393/

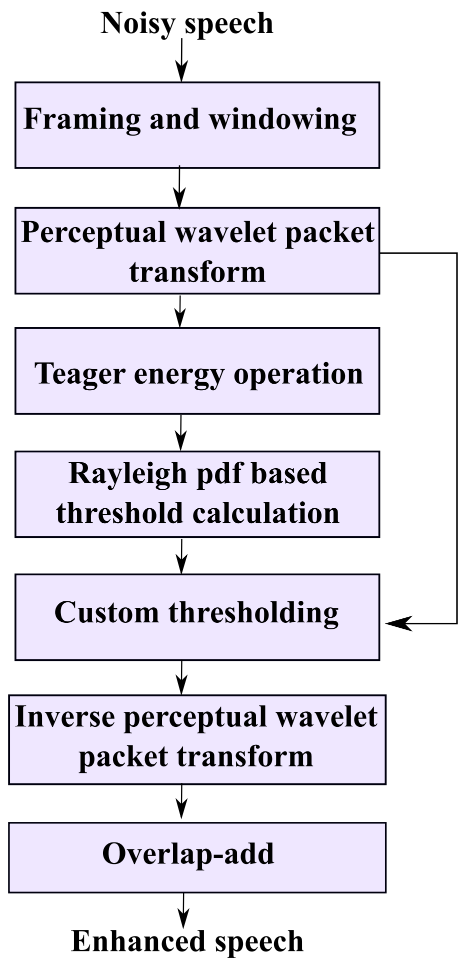

Fig. 1. Block diagram for the proposed method.

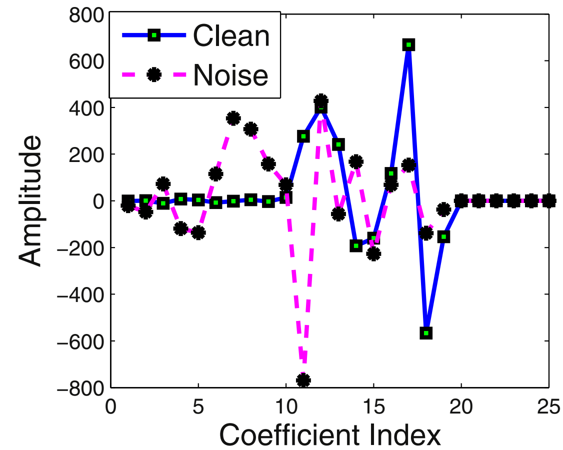

Fig. 2. PWP Coefficients of a noisy speech subband at an SNR of 5dB.

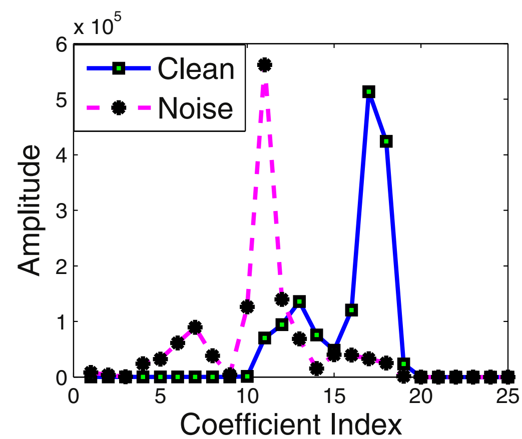

Fig. 3. TE Operated PWP Coefficients of a noisy speech subband at an SNR of 5dB.

The paper is organized as follows. Section 2 presents the proposed method. Section 3 describes results. Concluding remarks are presented in Section 4.

2. Proposed method

The block diagram for the proposed method is shown in Fig. 1. It is seen from Fig. 1 that PWP transform is first applied to each input speech frame. Then, the PWP coefficients are subject to Teager energy approximation with a view to determine a threshold value for performing thresholding operation in the PWP domain. On using a custom thresholding function, an enhanced speech frame is obtained via inverse perceptual wavelet packet (IPWP) transform.

2.1. Perceptual wavelet packet transform

The key element of the PWP transform is the use of the mel warping function to determine the wavelet packet decomposition structure (Sarikaya et al., 1998). The main motivation behind using this transform is its ability to decompose the signal according to human auditory system. At lower frequencies, where human auditory system can differentiate the pitches precisely, PWP transform decomposes the signal in finer bands. On the other hand, at higher frequencies, PWP transform creates less number of bands as the human cochlea can not differentiate among small differences in higher frequencies. The perceptual mel scale is a scale of pitches judged by listeners to be equal in distance from one another. The conversion of frequency to mel is given in Sarikaya et al. (1998).

The clean and noise PWP coefficients in a subband of a noisy speech frame at an SNR of 5 dB is plotted in Fig. 2. It is seen from this figure that for most of the coefficient indices, clean and noise PWP coefficients are not separable. Based on similar analysis performed on many speech signals corrupted by different noises, it is found that the time and frequency resolution provided by PWP transform is not sufficient to separate PWP coefficients of clean speech from that of noise even at a high SNR of 5 dB. Since TE operator has better time and frequency resolution (Kaiser, 1993), it can be very useful in handling noise. Therefore, we apply the discrete time TE operator on the PWP coefficients.

2.2. Teager energy operator

Letting $W_{k,\,m}$ as the mth PWP coefficient in the kth subband, the mth TE operated coefficient $t_{k,\,m}$ corresponding to the kth subband of the PWP transform is given by

$$t_{k,m} = T[W_{k,m}]. \tag{1}$$

Fig. 3 presents the clean, noise and noisy TE operated PWP coefficients in a subband of a noisy speech frame at the same SNR as used in Fig. 2. It is seen from this figure that at the indices where TE operated PWP coefficients of clean speech have higher values, the TE operated PWP coefficients of noise show lower values. As a result, thresholding operation on the noisy PWP coefficients needs a low threshold value to remove the noise leaving the speech undistorted. On the contrary, at the indices, where TE operated PWP coefficients of clean speech have lower values, the