Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 35 (2016) 58–72

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

Speech enhancement based on wavelet packet of an improved principal component analysis☆

Mohamed anouar Ben messaoud*, Aïcha Bouzid, Noureddine Ellouze

Electrical Engineering Department, National School of Engineers of Tunis, University of Tunis El Manar, Tunisia

Received 25 June 2014; received in revised form 7 March 2015; accepted 1 June 2015
Available online 19 June 2015

Abstract

In this paper, we propose a single-channel speech enhancement method, based on the combination of the wavelet packet transform and an improved version of the principal component analysis (PCA). Our method integrates ability of PCA to de-correlate the coefficients by extracting a linear relationship with what of wavelet packet analysis to derive feature vectors used for speech enhancement. This allows us to operate with a convenient shrinkage function on these new coefficients, removing the noise without degrading the speech. Then, the enhanced speech obtained by the inverse wavelet packet transform is decomposed into three subspaces: low rank, sparse, and the remainder noise components. Finally, we calculate the components as a segregation problem. The performance evaluation shows that our method provides a higher noise reduction and a lower signal distortion even in highly noisy conditions without introducing artifacts.
© 2015 Elsevier Ltd. All rights reserved.

Keywords: Speech enhancement; Wavelet packet analysis; Principal components analysis; Sparse matrix

1. Introduction

The goal of speech enhancement in a single-channel recording varies according to specific applications, such as to reduce listener fatigue, to improve the overall speech signal quality, to enhance intelligibility, and to increase the efficiency of the voice communication device (Loizou, 2007).

In general, speech denoising algorithms can be categorized into four broad classes: spectral subtractive algorithms, statistical-model-based algorithms, wavelet transform, and subspace algorithms.

Spectral subtraction (SS) (Boll, 1979) is one of the first algorithms applied to the problem of speech enhancement. It is simple to implement but it distorts the speech signal and introduces additional annoying noise known as "musical noise". Many algorithms have been proposed to remove this phenomenon including perceptually motivated techniques by Petrovsky et al. (2004) and the aspects of the human auditory system (Lu and Loizou, 2008), but their optimality in a sense of linear estimation is not clear.

☆ This paper has been recommended for acceptance by Tom Quatieri.

* Corresponding author. Tel.: +216 22634563.
E-mail addresses: anouar.benmessaoud@yahoo.fr (M.A. Ben messaoud), bouzidacha@yahoo.fr (A. Bouzid), n.ellouze@enit.rnu.tn (N. Ellouze).

http://dx.doi.org/10.1016/j.csl.2015.06.001
0885-2308/© 2015 Elsevier Ltd. All rights reserved.

Statistical model-based algorithms are one of the most commonly used classes of speech denoising methods. Recovery of the clean speech transform coefficients, or their magnitudes, is treated as a Bayesian estimation problem with known speech and noise statistics. Many estimators have been derived under different assumptions for the noise and speech distributions. The Wiener filter algorithms consist in estimating an optimal filter from the noisy speech spectrum by minimizing the Mean Square Error (MMSE) (Gui and Kwan, 2005). The MMSE estimator is used to evaluate the short-time spectral amplitude (STSA) based on a priori signal-to-noise ratio (SNR) estimation and Gaussian statistics. The prior SNR estimation was performed using "decision estimator" proposed by Ephraim and Malah (1984).

Wavelet transforms (WT) have been applied to various speech applications. The basic principle of speech enhancement in wavelets analysis is based on the thresholding of the discrete wavelet coefficients (DWC) to segregate the components corresponding to the target speech from those of the noise. However, when we apply a fixed threshold for the DWC of speech, some unvoiced speech frames can be eliminated with the additional ranges noise thus degrading the intelligibility of the enhanced speech. To solve this problem, the thresholding must be modified over time. Therefore, diverse adaptive wavelet thresholding procedures are proposed as the universal threshold proposed by Donoho and Johnstone (1995), Stein's Unbiased Risk Estimate (SURE) strategy described in Hu and Loizou (2004), Bayes Shrink exploiting a Bayesian estimate is described in Leporini and Pesquet (2001). The main drawback of the WT is the restricted number of frequency bands. Also, the unvoiced frames of noisy speech have proven to be problematic in terms of the wavelet shrinking. Since the unvoiced parts of speech include many noise-like high frequency components, removing them in the wavelet domain can degrade the quality and intelligibility of the enhanced speech.

The speech subspace algorithms consist in projecting the noisy speech segments onto orthogonal subspaces. The speech subspace is composed of high-energy vectors in the segment's principal component (PC) basis. The first algorithm was introduced by Dendrinos et al. (1991) who proposed to use the Singular Value Decomposition (SVD) technique to eliminate the noise subspace for speech denoising. Therefore, Ephraim et al. (1996) have used the fast Fourier transform (FFT) to approximate PC basis. These methods have removed the "musical noise" artifacts but the subspace approach improves perceived speech quality without increasing speech intelligibility. A well-known method proposed by Hu and Loizou (2003) is based on a joint digitalization of the noise and clean speech covariance matrices leading to the optimal estimators. Unfortunately, an efficient implementation of the subspace based approaches with an optimal choice of parameters is a challenging task and mainly in the case of colored and babble noise. To overcome the limitation, many approaches based on the speech segregation for enhancing the perceptual intelligibility of the speech degraded by additive background noise have been presented such as K-SVD by Sigg et al. (2010) and non-negative matrix factorization (NMF) by Mohammadiha et al. (2011), and Mysore and Smaragdis (2011). However, these methods always demand prior training for supervised segregation, empirical parameters, or particular features.

So, a number of studies using principal component analysis (PCA) are proposed. The essential object is to obtain a set of orthogonal factors that describe the variance of the observations and track the new factors considered to determine the necessary features without prior training. Speech processing by PCA (Joliffe, 2002) is extensively applied as a classical multivariate speech processing tool. For speech segregation, a robust extension of classical PCA by generalizing an eigenvalue decomposition of a pair of covariance matrices is proposed by Benabderrahmane et al. (2010). It may be used in speech denoising by Shinde et al. (2012), speech identification by Abolhassani et al. (2007), and speech recognition by Takiguchi and Ariki (2007).

In this paper, we propose a speech enhancement method based on the combination of the wavelet packet transform (WPT) and an improved version of the principal component analysis (IPCA). In fact, the orthogonal basis functions produced by the WPT, provide satisfactory results at low and high frequencies (Donoho et al., 1995; Ghanbari and Karami, 2006). Projection of a noisy speech onto these basis functions may be realized efficiently by passing the noisy speech through a tree-structured conjugate quadrature filter bank. In the proposed method, we apply the WPT, which allows to obtain appropriate time-frequency resolution, and to adaptively select relative frequency band based on the type of the speech to be estimated. To further improve denoising character of WPT, the paper describes a method, which combines the ability of the PCA tool to decorrelate the variables by extracting a linear relationship with that of wavelet packet analysis to enhance the speech signal. Our basic idea is to construct powerful filters by applying the PCA in the wavelet packet domain, thus getting a compaction of the speech energy into a few principal components (PC's), while the noise is spread over all the transformed coefficients. This allows us to operate with a convenient shrinkage function on these new coefficients, removing the noise without degrading the speech.

Then, we apply our improved version of the PCA at the enhanced speech obtained by the inverse wavelet packet transform. Our extension of PCA technique exploits the benefit of sparse PCA in the context of classification. Based

Download English Version:

https://daneshyari.com/en/article/558246

Download Persian Version:

https://daneshyari.com/article/558246

Daneshyari.com