Applied Acoustics 96 (2015) 11–19

Contents lists available at ScienceDirect

Applied Acoustics

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

A speech enhancement approach based on noise classification

Wenhao Yuan *, Bin Xia

College of Computer Science and Technology, Shandong University of Technology, Shandong 255049, China

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form 16 December 2014
Accepted 11 March 2015
Available online 25 March 2015

Keywords:
Speech enhancement
Noise classification
Improved minima controlled recursive averaging

ABSTRACT

For speech enhancement, most existing approaches do not consider the differences, between various types of noise, which significantly affect the performance of speech enhancement. In this paper, we propose a novel speech enhancement approach by taking into account the different characteristic statistical properties of various noise on the basis of noise classification. To classify noise, an effective noise classification method is firstly developed by exploiting the features of noise energy distribution in the Bark domain. Then, based on the noise types, the speech enhancement approach is obtained by forming the optimal parameter combinations for the optimally modified log-spectral amplitude (OM-LSA) speech estimator with the improved minima controlled recursive averaging (IMCRA) noise estimator, where the parameter combinations consisting of the smoothing parameters for smoothing the noisy power spectrum and the recursive averaging in the noise spectrum estimation as well as the weighting factor for the *a priori* SNR estimation, are built through the enhancement of noisy speech samples. Finally, extensive experiments are carried out in terms of objective evaluation under various noise conditions, and the experimental results show that the proposed approach yields better performance compared with the conventional OM-LSA with IMCRA in speech enhancement.

1. Introduction

Single channel speech enhancement has been one of the most widely used approaches for the enhancement of noisy speech which is a crucial component of speech signal processing in noisy environments [1–6]. The spectral subtraction method proposed by Boll in [7] is a popular single channel speech enhancement technique, which substantially reduces the noise level in the noisy speech. According to the basic principle of spectral subtraction method, two major components generally should be considered in a practical speech enhancement system: the estimation of speech, and the estimation of noise power spectrum [8,9].

As for estimating the speech, a commonly used approach is the minimum mean-square error (MMSE) short-time spectral amplitude (STSA) estimator, which is derived by Ephraim and Malah in [10]. By utilizing decision directed approach to smooth the *a priori* SNR recursively, the MMSE estimator successfully conquers the main drawback of the conventional spectral subtraction method that it may introduce an annoying distortion called musical noise into the enhanced speech. Subsequently, Ephraim and Malah derived a MMSE log-spectral amplitude (LSA) estimator in literature [11] which minimizes the mean-square error of the log-spectra. Further, by modifying the gain function of the LSA estimator based on two hypotheses associated with the speech presence uncertainty, Cohen presented an optimally modified LSA (OM-LSA) speech estimator [8], which shows significant superiority in speech enhancement.

Considering the noise power spectrum estimation, Martin proposed a noise PSD estimation algorithm based on minimum statistics (MS) [12], which tracks the minima values of the smoothed spectrum of the noisy speech over a finite window, and then multiplies the result by a bias factor to achieve the unbiased estimate of noise spectrum. Another successful noise PSD estimation approach, known as the minima controlled recursive averaging (MCRA) algorithm [13], is to search the local minimum similarly to MS, and then compare the ratio of the noisy speech to the local minimum against a threshold to find the noise-only regions. The noise PSD estimate is updated by tracking the noise-only regions of the noisy speech spectrum. In [14], Cohen presented the improved MCRA (IMCRA), which uses a different method to track the noise-only regions based on the estimated speech presence probability. In [15], Rangachari and Loizou updated the noise PSD estimate continuously in every frame using the speech presence probability which was obtained by comparing the ratio of noisy speech power spectrum to its local minimum against a frequency-dependent threshold. The more recent work on noise PSD estimation is the MMSE-based algorithm with bias

* Corresponding author.
E-mail addresses: why_sdut@126.com (W. Yuan), xiabin@sdut.edu.cn (B. Xia).

http://dx.doi.org/10.1016/j.apacoust.2015.03.005
0003-682X/

compensation (MMSE-BC) proposed by Hendriks and Heusdens [16–18], which employs a limited maximum likelihood estimate of the *a priori* SNR to obtain an MMSE estimate of the noise periodogram. The MMSE-BC estimator reduces the computational complexity greatly without degrading the performance of noise tracking. In [19], Gerkmann and Hendriks further improved the MMSE-BC estimator by making use of a soft speech presence probability with fixed priors, and presented an unbiased MMSE-based noise PSD estimator, which is of an even lower complexity than the MMSE-BC.

To distinguish speech from noise, both the estimation of speech and the estimation of noise power spectrum have fully considered the differences between speech and noise. For example, the estimation of noise PSD is generally based on the assumption that the noise power is varying more slowly than the speech power. Besides, the differences are also made between different types of noise in the characteristic statistical properties. However, the design of the speech enhancement algorithms usually does not take the differences into consideration, which causes these algorithms to be not always optimal for various noise environments. As to improve the performance of speech enhancement, these algorithms should be adjusted to adapt to different types of noise and deal with them respectively. In other words, we can improve the performance of these speech enhancement algorithms by incorporating the noise classification into them.

For noise classification, a variety of features have been proposed, including time domain features [20], spectral domain features [21] and the features derived from linear predictive coding (LPC) and wavelet transforms [22,23], of which the mel-frequency cepstral coefficients (MFCC) features are most widely used. Much work has been done to recognize the nonspeech audio based on the MFCC features. Ma et al. described an acoustic environment classifier using a 39-dimensional MFCC feature vector [24]. In [25], to yield higher recognition accuracy for environmental sounds, the matching pursuit algorithm is used to obtain effective time–frequency features as the supplement of the MFCC features. Gopalakrishna et al. utilized the MFCC + Δ MFCC features to classify the background noise environment in real time for automatic tuning of noise suppression algorithms for cochlear implant applications [26]. The classification accuracy for the above studies varied from 80% to 95% under different databases, which implies that there is still work to do to extract more effective features for noise classification.

In this paper, on the basis of the speech enhancement scheme based on the IMCRA noise PSD estimator and the OM-LSA speech estimator, we propose a speech enhancement approach using noise classification of noisy speech. Firstly, we define a parameter combination related to the noise types, which includes some principal parameters in the OM-LSA with IMCRA, such as the smoothing parameters for the smoothing of the noisy power spectrum and the recursive averaging in the noise spectrum estimation, as well as the weighting factor for the *a priori* SNR estimation. Through the enhancement of noisy speech samples, by identifying the optimal parameter combinations for the speech enhancement scheme based on the IMCRA and the OM-LSA under specific noise environments, we obtain the optimal parametric OM-LSA with IMCRA. Secondly, to recognize the noise type of the noisy speech, we propose a support vector machine-based noise classification method, which exploits the features of noise energy distribution in the Bark domain. Thirdly, by choosing the the optimal parameter combination for the speech enhancement scheme based on the OM-LSA and the IMCRA according to the recognized noise type, we implement the noise PSD estimation and calculate the enhanced speech using the optimal parametric OM-LSA with IMCRA. Objective quality tests are performed to evaluate the proposed approach under various noise environments, which validate the superior performance of the proposed approach to the conventional speech enhancement scheme based on the OM-LSA and the IMCRA.

The rest of the paper is organized as follows. Section 2 briefly reviews the speech enhancement scheme based on the IMCRA and the OM-LSA, and Section 3 presents the optimal parametric OM-LSA with IMCRA for various noise. Section 4 introduces the support vector machine-based noise classification method. In Section 5, we describe the proposed noise classification-based speech enhancement approach. The performance of the proposed noise classification method and speech enhancement approach is evaluated in Section 6. Finally, conclusions are given in Section 7.

2. Review of OM-LSA with IMCRA

Let y denote an observed noisy signal in the time domain, which is the sum of a clean speech x and an uncorrelated additive noise d. By applying the short-time Fourier transform (STFT), we have

$$Y(k,l) = X(k,l) + D(k,l) \tag{1}$$

in the time–frequency domain, where k represents the frequency bin index, and l is the frame index.

In the IMCRA, the noise PSD is estimated by recursively averaging past spectral power values of the noisy measurement during periods of speech absence and holding the estimate during speech presence [14]. Under speech presence uncertainty, the conditional speech presence probability is employed, and the recursive averaging can be obtained by

$$\bar{\lambda}_d(k,l+1) = \tilde{\alpha}_d(k,l)\bar{\lambda}_d(k,l) + [1 - \tilde{\alpha}_d(k,l)]|Y(k,l)|^2 \tag{2}$$

where

$$\tilde{\alpha}_d(k,l) \triangleq \alpha_d + (1-\alpha_d)p(k,l) \tag{3}$$

is a time-varying frequency-dependent smoothing parameter. $\alpha_d (0 < \alpha_d < 1)$ denotes a smoothing parameter, and $p(k,l)$ is the conditional speech presence probability. Through introducing a bias compensation factor β, the noise PSD estimate is given by

$$\hat{\lambda}_d(k,l+1) = \beta \cdot \bar{\lambda}_d(k,l+1) \tag{4}$$

The estimation of the speech presence probability is based on a Gaussian statistical model in the IMCRA, and is obtained by

$$p(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp(-\upsilon(k,l))\right\}^{-1} \tag{5}$$

where $q(k,l)$ is the *a priori* probability for speech absence, γ and ξ represent the *a posteriori* and the *a priori* SNRs respectively, and $\upsilon \triangleq \gamma\xi/(1+\xi)$.

In order to calculate the *a priori* speech absence probability $q(k,l)$, two iterations of smoothing and minimum tracking are carried out. Let $S(k,l)$ denote the smoothed periodogram of the noisy measurement, then the time smoothing in the first iteration is performed by a first-order recursive averaging

$$S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s)S_f(k,l) \tag{6}$$

where $\alpha_s (0 < \alpha_s < 1)$ is a smoothing parameter, and $S_f(k,l)$ is obtained by the frequency smoothing of the noisy power spectrum

$$S_f(k,l) = \sum_{i=-\omega}^{\omega} b(i)|Y(k-i,l)|^2 \tag{7}$$

where b denotes a normalized window function of length $2\omega + 1$. The time smoothing in the second iteration is similar to that in the first iteration, and utilizes the same smoothing parameter.