Applied Acoustics 76 (2014) 249–261

Contents lists available at ScienceDirect

Applied Acoustics

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

Noise reduction using three-step gain factor and iterative-directional-median filter

Ching-Ta Lu *

Department of Information Communication, Asia University, Taichung 41354, Taiwan, ROC

ARTICLE INFO

Article history:
Received 5 October 2012
Received in revised form 17 August 2013
Accepted 20 August 2013
Available online 16 September 2013

Keywords:
Speech enhancement
Post processing
Musical residual noise
Spectral subtraction
Median filter

ABSTRACT

Musical residual noise is a major problem for a speech enhancement system. This noise is very annoying to the human ear and can significantly deteriorate the perception quality of enhanced speech. In this study, we aim at reducing the quantity of musical residual noise by a two-stage speech enhancement approach. In the first stage a preprocessor enhances noisy speech using an algorithm which combines the two-step-decision-directed and the Virag methods. In the second stage the enhanced speech signal is post-processed by an iterative-directional-median filter to significantly reduce the quantity of residual noise, while maintaining the harmonic spectra. Experimental results show that the proposed approach can significantly improve the performance of a speech enhancement system by reducing the quantity of residual noise.

1. Introduction

Speech enhancement is useful in many applications such as voice communication and automatic speech recognition. Recently, many novel schemes [1–17] have proposed enhancing a speech signal which is corrupted by additive noise. Although their improvement is presented in terms of background noise reduction, the main drawback is the appearance of annoying musical residual noise which is caused by randomly spaced spectral peaks that come and go in successive frames, and occur at random frequencies. Some novel schemes attempt to reduce the effect of musical residual noise by utilizing the masking effects of the human auditory system [6,8–11,13,16–20]. In these algorithms only the audible noise components are suppressed, this results in the reduction of speech distortion.

Ding et al. [4] used a hybrid Wiener spectrogram filter for noise reduction, followed by a multi-blade post-processor which exploited the two-dimensional features of the spectrogram to preserve speech quality and to further reduce residual noise. Yu et al. [14] proposed using a non-diagonal audio denoising algorithm through adaptive time–frequency block thresholding to enhance an audio signal. This algorithm can adjust system parameters to the signal's property by minimizing a Stein estimation of the risk. Experimental results showed that this method can improve the quality of an audio signal. Plapous et al. [12] proposed a two-step-decision-directed (TSDD) algorithm to improve the estimate of the a priori SNR for a decision-directed approach. Experimental results showed that the performance of the decision-directed approach could be significantly improved by their novel method. Zhang et al. [15] proposed using a non-linear high frequency technique to re-synthesize the upper-band signal based on the lower-bands speech for vowel speech, yielding the harmonic spectra being reconstructed. Ghanbari and Karami-Mollaei [5] proposed using an adaptive threshold on a modified hard thresholding function. Although many novel speech enhancement systems can efficiently remove background noise, the musical residual noise is still apparent and annoying to the human ear. This has lead to many studies [4,6,8,10–12,20–27] proposing new techniques to reduce the effect of musical residual noise. For example, Lu [23] derived a smoothing factor as a second stage to reduce the effect of musical residual noise. An accurate estimate of the a priori SNR is critical for eliminating musical noise. Hence, Lu et al. [16] proposed using a three-step-decision gain factor to efficiently remove background noise in noisy speech. Since musical residual noise still exists in enhanced speech, they proposed to reduce it by using a directional median filter [24]. Moreover, they [25,26] proposed incorporating a block-median filter to serve as a post processor. Experimental results have shown that musical residual noise can be efficiently removed, at the expense of introducing a greater amount of speech distortion. Esch and Vary [27] proposed using a post-filter, for the spectral weighting gains, that is capable of reducing musical residual noise. This method includes a detector for speech pauses and low SNR conditions. They adaptively smooth

* Address: Department of Information Communication, Asia University, 500, Lioufeng Rd., Wufeng, Taichung City 41354, Taiwan, ROC. Tel.: +886 4 23323456x1869; fax: +886 4 23305824.

E-mail addresses: lucas1@ms26.hinet.net, lucas@asia.edu.tw

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.08.015

the weighting gains over frequency for those two conditions to reduce the musical residual noise. Recently, many studies have proposed using a median filter to remove impulse noise in image denoising applications [28–32]. We found that the properties of musical tones in the spectrogram are similar to those of impulse noise in the spatial domain. Consequently, these techniques can be modified and applied to remove musical tones in the spectrogram.

Based on the above findings, utilizing the masking properties of the human auditory system allows algorithms to enhance speech with lower levels of distortion. However, residual noise remains reducing the perceived quality of the enhanced speech. In this paper we employ a three-step-decision gain factor [16] to preprocess noisy speech; enabling background noise to be efficiently removed. This method can improve the performance of the Virag method [20] in the frequency domain by providing better estimates of the noise-masking threshold using the two-step-decision-directed algorithm [12]. This algorithm [16] forms the preprocessing stage. A directional median filter [24] is then modified and added as a post-processor to further reduce the residual noise. The spectral variation for each spectral bin of a frame is sequentially analyzed in the spectrogram. If a spectral bin is classified as a musical tone, it is iteratively modified by the median spectrum in an optimum direction of a window. Experimental results show that the proposed approach can significantly improve the performance of the Virag [20] method by reducing the quantity of residual noise, while the speech distortion can also be maintained at an acceptable level. In addition, the proposed approach also outperforms the two-step-decision-directed (TSDD) speech enhancement algorithm [12] in most experiments.

The rest of this paper is organized as follows. Section 2 describes the proposed speech enhancement system. Section 3 briefly introduces a preprocessor. Section 4 describes the proposed post processor. Section 5 demonstrates the experimental results. Conclusions are finally drawn in Section 6.

2. Proposed speech enhancement system

The block diagram of the proposed speech enhancement system is shown in Fig. 1. Firstly, a noisy speech signal is transformed into the frequency domain by fast Fourier transform, where the magnitude of the transformed coefficient is called a spectrum bin hereafter. The two-step-decision-directed (TSDD) algorithm [12] is employed to enhance noisy speech. This enhanced speech is only used to find the noise-masking threshold (NMT) which is applied in the Virag method to adapt the generalized spectral subtraction algorithm, rather than to be the output signal in the first stage. Noisy speech is filtered by the Virag method. The output signal is called preprocessed speech. Owing to the preprocessed speech consisting of many musical tones, a post processing system is required for their removal.

In this study, we use an iterative-directional-median filter (IDMF) to refine the preprocessed signal. This filter is a two-dimensional speech enhancement in spectrograms, i.e., a spectrum bin of each frame is iteratively modified by the median value in an optimum direction. Firstly, a N $\times$ N (5 $\times$ 5 in the experiments) window is employed to analyze the spectral variation property for the central spectrum bin. If the spectral variation is large over the neighbors, the reference spectrum is classified as a musical tone. The median filter is performed in the optimum direction, which has the smallest spectral distance over the neighbors. If the spectral variation is small in the optimum direction, the central spectrum bin may be a harmonic spectrum. This spectrum bin is kept unchanged to maintain speech quality. In addition, a directional median filter is iteratively performed to further reduce residual noise. We also modify the threshold to decide whether the central spectrum bin is a musical tone or not. This threshold is updated for each iteration step and is adapted by the SNR of a subband.

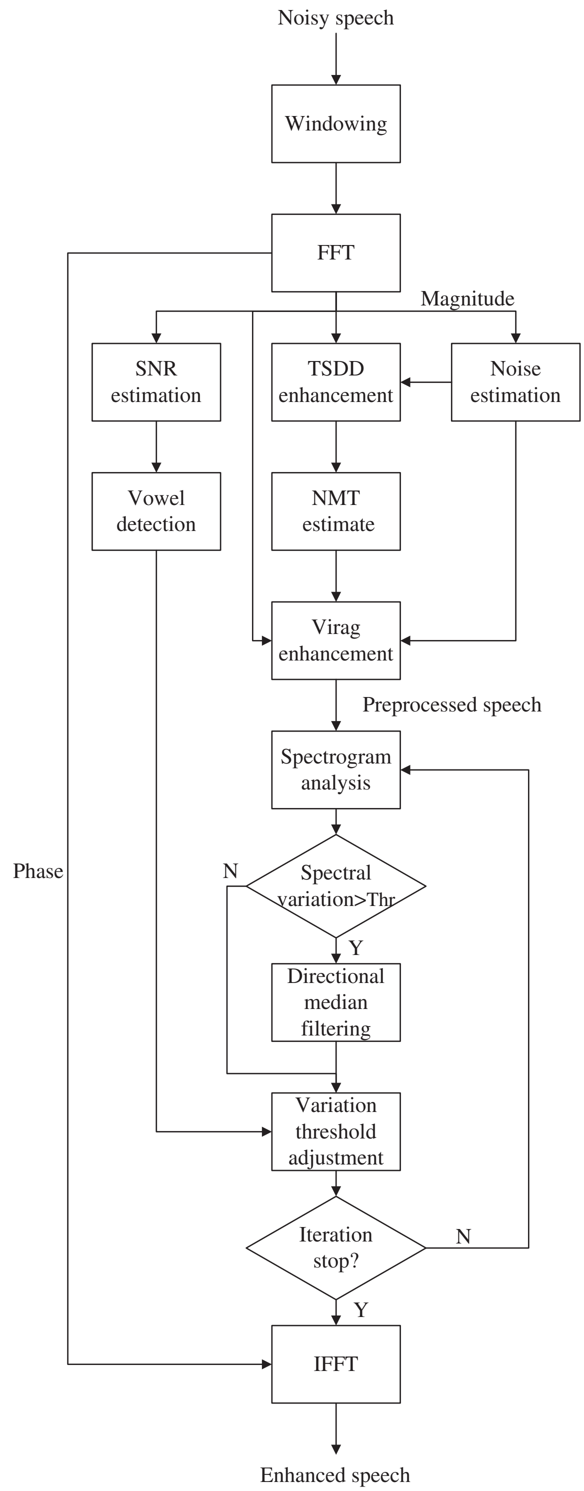

Fig. 1. Block diagram of proposed speech enhancement system.

A noisy speech signal $y(m,n)$ can be modeled as the sum of clean speech $s(m,n)$ and additive noise $d(m,n)$ in the frame m of the time domain, i.e.,

$$y(m,n) = s(m,n) + d(m,n) \tag{1}$$

The spectral estimate of the speech signal $\tilde{S}(m,\omega)$ is obtained by multiplying a gain factor $g(m,\omega)$ with the noisy spectrum $Y(m,\omega)$ in a subband, i.e.,