Information Systems and Signal Processing Journal (2016) 1: 19-23
Clausius Scientific Press, Canada

Research on Speech Enhancement Algorithm Based on EMD in Noisy Environments

Cheng Huawei[1, a], Zhang Dexiang[2,b,*] , Fang Fei [3,c] and Lu Weiqing[3,d]

[1] School of Electronic and Electrical Engineering of Anhui Sanlian University, Hefei, 230601, China
[2] School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China
[3] School of Electronic and Electrical Engineering of Anhui Sanlian University, Hefei, 230601, China
[a] chw520.good@163.com, [*b]zdxdzxy@126.com(corresponding author), [c]austfei@126.com, [d]velvet0810@sina.com

Keywords: Speech Enhancement, EMD, Spectral Subtraction, Noisy Environments.

Abstract: This paper presents a new technique for speech enhancement in a noisy environment based on the empirical mode decomposition (EMD) algorithm and spectral subtraction. With the EMD, the noise speech signals can be decomposed into a sum of the band-limited function called intrinsic mode functions (IMFs), which is a zero-mean AM-FM component. Then spectral subtraction of the IMF components with noise can be used to eliminate the effect of the noise for speech enhancement. Experimental results demonstrate the advantage of using the proposed simultaneous detection and estimation approach with the proposed algorithm, which facilitate suppression of transient noise with a controlled level of speech distortion.

1. Introduction

In the field of speech enhancement, we are interested in the reduction of noise from noise corrupted speech in order to improve its intelligibility and quality. Speech enhancement is crucial for speech recognition accuracy. Commonly the frequency band of noise is wide that of original signal is limited and mainly lies in low frequency bands. How to eliminate the effect of the noise constitutes a challenging problem in speech processing. Speech enhancement algorithms have been developing considerably. Various methods have been investigated in the literature for performing speech enhancement [1].

The single-channel speech enhancement problem has received wide attention and consequently numerous algorithms have been proposed on the subject. Spectral conversion has been applied previously in the context of voice conversion, and has been shown to successfully transform spectral features with particular statistical properties into spectral features that best fit different target statistics. These can be grouped into spectral subtraction, MMSE estimation, Wiener filtering, Kalman filtering, and subspace methods. Several of these methods employ the analysis modification synthesis framework [2]. Speech enhancement systems often operate in the short-time Fourier transform domain, where the speech spectral coefficients are estimated from the spectral coefficients of the degraded signal. Concentrating on the additive noise problem, one of the most popular,

effective, and simple algorithms to implement is spectral subtraction. The method although simple is quite effective [3].

Time–frequency analysis methods may be the most popular traditional technique for speech enhancement in the signal processing methods. However, the tradition time-frequency analysis methods such as short time Fourier transform (STFT) and wavelet transform have respective limitations. In STFT analysis method, once the time-frequency window function is chosen, its size of the time–frequency window would be fixed, therefore the time and frequency resolution are same for all components that include different time scales. Wavelet transform analysis method could prove local features in both time and frequency domains. However, wavelet analysis method is an adjustable window Fourier transforms in essential aspect. Therefore, wavelet analysis is not a self-adaptive signal processing method in nature [4].

In this paper, we use a new technique called the empirical mode decomposition (EMD) has recently been introduced by Norden E. Huang et al. in 1998. The idea behind EMD method is to adaptively decompose non-stationary signal into a number of zero-mean amplitude basis function termed intrinsic mode functions. This method has quite good characters to analyze non-stationary signal and nonlinear signal. A new technique for speech enhancement in a noisy environment based on the empirical mode decomposition (EMD) algorithm and spectral subtraction is proposed.

2. Empirical Mode Decomposition

The principle of the EMD technique is to decompose a signal into a sum of the band-limited functions called intrinsic mode functions. Given a signal $x(t)$, detect the local maxima and minima of $x(t)$. Then generate the upper envelope $x_u(t)$ and the lower envelops $x_l(t)$ by connecting the maxima and minima separately with cubic spline interpolation. Then the mean of the two envelops is denoted as:

$$m_1(t) = (x_u(t) + x_l(t))/2 \tag{1}$$

Thus, the first IMF $h_1(t)$ is obtained as:

$$h_1(t) = x(t) - m_1(t) \tag{2}$$

We can repeat this sifting procedure k times, until h_{1k} is an IMF, that is

$$h_{1k} = h_{1(k-1)} - m_{1k} \tag{3}$$

Then, designate h_{1k} as c_1

$$c_1 = h_{1k} \tag{4}$$

where c_1 is the first IMF of the original signal. We can separate $c_1(t)$ from the rest of the data by:

$$r_1(t) = x(t) - c_1(t) \tag{5}$$

Note that the residue $r_1(t)$ still contains some useful information. We can therefore treat the residue as a new signal and apply the above procedure to obtain:

$$r_N(t) = r_{N-1}(t) - c_N(t) \tag{6}$$

The sifting process will be continued until the final residue $r_N(t)$ is a constant, a monotonic function, or a function with only one maxima and one minima from which no more IMF can be derived. Combining the equations in (5) and (6) yields the EMD of the original signal:

$$x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t) \tag{7}$$

The result of the EMD produces N IMF and a residue signal. It is observed that higher order IMFs contain lower frequency oscillations than that of lower order IMFs. If we interpret the EMD as a time-scale analysis method, each IMF reflects the characteristic of each scale, which shows the intrinsic mode characteristic of non-stationary and nonlinear signal.

3. Spectral Subtraction Method

Traditional analysis-modification-synthesis based speech enhancement methods modify only the magnitude spectrum while keeping the noisy phase spectrum unchanged for synthesis. The proposed speech enhancement method is based on the AMS framework commonly employed in speech processing. The AMS framework consists of three stages: 1) the analysis stage, where the input speech is processed using DSTFT analysis; 2) the modification stage, where the noisy complex spectrum undergoes some kind of modification; and 3) the synthesis stage, where the inverse discrete short-time Fourier transform (IDSTFT) operation is followed by the overlap-add (OLA) synthesis to reconstruct the output signal.

The spectral subtraction method is a simple and effective method of noise reduction [5]. In this method, an average signal spectrum and average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio (SNR) is improved. It is assumed that the signal is distorted by a wide-band, stationary, additive noise, the noise estimate is the same during the analysis and the restoration and the phase is the same in the original and restored signal.

The noisy signal y(t) is a sum of the desired signal x(t) and the noise n(t):

$$y(t) = x(t)) + n(t) \tag{8}$$

In the frequency domain, this may be denoted as:

$$Y(j\omega) = X(j\omega) + N(j\omega) \;\Rightarrow\; X(j\omega) = Y(j\omega) - N(j\omega) \tag{9}$$

where Y(jω), X(jω), N(jω) are Fourier transforms of y(t), x(t), n(t), respectively.

The statistic parameters of the noise are not known, thus the noise and the speech signal are replaced by their estimates:

$$\hat{X}(j\omega) = Y(j\omega) - \hat{N}(j\omega) \tag{10}$$

The noise spectrum estimate $\hat{N}(j\omega)$ is related to the expected noise spectrum $E[|N(j\omega)|]$ which is usually calculated using the time-averaged noise spectrum $\overline{N}(j\omega)$ taken from parts of the recording where only noise is present. The noise estimate is given by:

$$\hat{N}(j\omega) = E[|N(j\omega)|] = |\overline{N}(j\omega)| = \frac{1}{K}\sum_{i=0}^{K-1}|N_i(j\omega)| \tag{11}$$

where $|N_i(j\omega)|$ is the amplitude spectrum of the i-*th* of the *K* frames of noise. To obtain the noise estimate, the part of the recording containing only noise that precedes the part containing speech

signal should be analysed (the length of the analysed fragment should be at least 300 ms). To achieve this, additional speech detector has to be used.

4. Experiments and Results

The experiment signal is a pure speech signal is shown in Fig.1. The sampling frequency is set to 10 kHz with 16-bit amplitude resolution. The noised experiment data is the pure speech signal in Figure 1 corrupted by white noise and SNR=7.5dB. The noised speech signal is shown in Fig. 2.

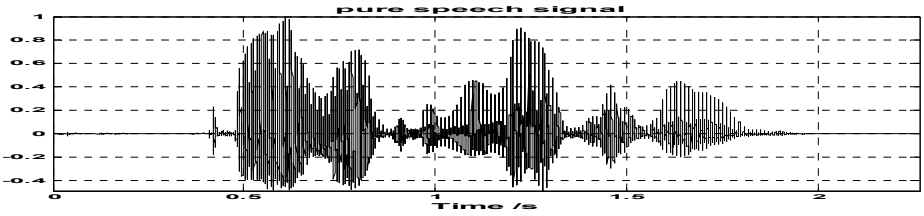

Fig. 1 The pure speech signal

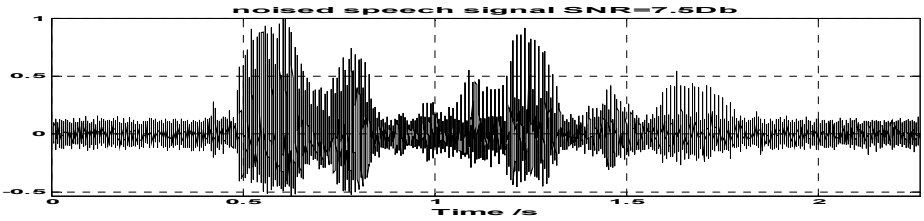

Fig. 2 The noised speech signal corrupted by white noise and SNR=7.5dB.

The noise removal procedures are as follows: (1) Decompose the noised original signal to IMFs by the EMD method; (2) Remove the IMFs whose content belongs to the noise on the base of the sudden increase of its amplitudes using spectral subtraction method; (3) Reconstruct the signal with the rest of IMFs.

Figure 3 shows the result using the spectral subtraction speech enhancement method. Figure 4 shows the result using the proposed speech enhancement method. The experimental results illustrate the newly proposed speech enhancement method is effective.

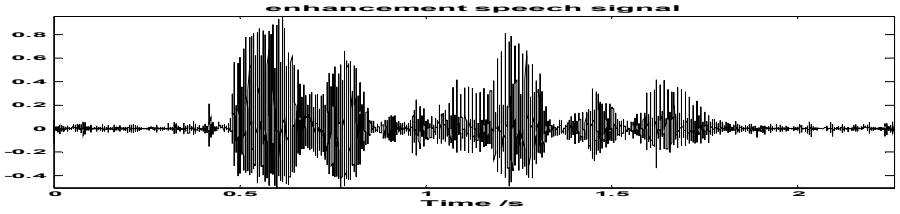

Fig.3 the result using the spectral subtraction speech enhancement method

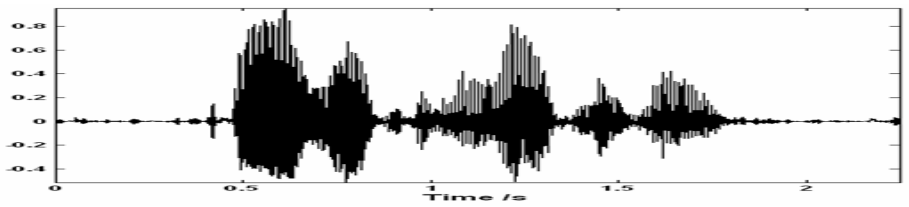

Fig.4 the result using the proposed speech enhancement method

5. Summary

In this paper, we have proposed a new methodology to enhance speech of a speech signal embedded in noise based on the empirical mode decomposition (EMD) algorithm and spectral subtraction. The EMD is well adapted to non-stationary signals and has an excellent time resolution. Our approach has been to provide initial estimates of the clean speech parameters from the noisy speech using spectral subtraction. The EMD and spectral subtraction gives good speech enhancement of a speech signal corrupted with noise. Our experiments show the proposed method can accurately extract the speech signals in differently background noises.

Acknowledgements

This work was financially supported by the Chinese National Science Foundation Grant (No.61272025), the Anhui Sanlian University Natural Science Foundation (kjzd2016002 and 2014Z014 and 2015Z017).

References

[1] G. Tanyer, H. Ozer, IEEE Trans. Speech Audio Processing, Vol. 8 (2000), p. 478

[2] Y.ephfsim, D. Malah, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 32(1984), p. 1109.

[3] N. E. Huang, Z. Shen and S. R. Long, Proceedings of the Royal Society London A, Vol. 454 (1998), p. 903

[4] S. Mallat, IEEE Trans. Pattern Anal. Machine Intell. Vol. 11 (1989), p. 674.

[5] Ramos A L L, Holm S, Gudvangen S, et al. International Journal of Electronics and Telecommunications, Vol. 59(2016), p. 93.