A Speech Enhancement Approach Based on Improved HHT

Xueqing Xu

Oil Recovery Plant No. 3, Daqing Oilfield Company Ltd, Daqing 163113, China

xqxu@petrochina.com

Abstract. Researching on the basic theory of HHT and its application in terms of speech enhancement, with the problem of poor adaptability and flexibility of the hard threshold de-noising algorithm, we present an improved soft threshold HHT speech enhancement algorithm. Through MATLB software simulation to explore the enhanced situation of noisy speech signal in different SNR, the simulation results proved that the definition and speech intelligibility of enhanced speech signal which enhanced by soft threshold HHT is higher. The new method has strong applicability in the field of speech enhancement.

Keywords: HHT, EMD, Soft Threshold, Speech Enhancement, SNR.

1. Introduction

The main purpose of the speech enhancement is extracted the original speech signal from the noisy speech signal as pure as possible, that is remove the noise and interference, improve the quality of noisy speech signal. In current there are many typical time-frequency analysis methods such as short-time Fourier transform and Wigner-Ville transform, wavelet transform and so on, but most are based on Fourier transform, so it's difficult to get rid of the disadvantages of Fourier transform. [1]In this paper, we research on HHT method and proposed a new soft threshold HHT method. Using the MATLAB software to simulate the noisy speech signal in different SNR, this method possesses higher resolution and speech intelligibility, realize the speech enhancement very well.

2. HHT Theory

HHT algorithm includes two basic processes: empirical mode decomposition (EMD) and the Hilbert spectrum analysis. The basic process is using EMD to decompose the signal into a finite number of IMF function, to make HHT of each IMF function and to solve the instantaneous frequency can get time-frequency energy distribution in the plane of the signal, that is the Hilbert spectrum, then the marginal spectrum can be obtained.[2]

The process of EMD is as follows:

(1)Determine the all local extremum points of the signal $s(t)$, using cubic spline interpolation method fitting form upper envelope and lower envelope and calculate the mean curve up $m_1(t)$;

(2)Calculation $h_1(t) = s(t) - m_1(t)$. If $h_1(t)$ satisfy the two conditions of IMF, then $h_1(t)$ is an IMF component;

(3)If $h_1(t)$ don't satisfy the IMF's conditions, the $h_1(t)$ as original data, repeat the previous two steps, until $h_{1k}(t)$ satisfy the IMF's conditions, remark as $c_1(t)$;

(4)Separate $c_1(t)$ from $s(t)$, to obtain $r_1(t) = s(t) - c_1(t)$;

(5)Set $r_1(t)$ as original signal and repeat above process, until $r_n(t)$ meet decomposition termination conditions, get the entire IMF component $\{c_i(t)\}$ and an allowance $r_n(t)$ of signal $s(t)$.

Huang defined the termination conditions of entire process of decomposition were when an allowance $r_n(t)$ into a monotonic function or amplitude was less than a preset threshold decomposition termination.

$s(t)$ Is decomposed into N layers of IMF and allowance:

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

$$z_i(t) = c_i(t) + j\tilde{c}_i(t) = a_i(t) e^{j\theta_i(t)} \tag{2}$$

Definition of instantaneous amplitude and instantaneous phase are;

$$a_i(t) = \left[c_i^2(t) + c_i^{-2}(t) \right]^{1/2} \tag{3}$$

$$\theta_i(t) = \arctan\left(c_i(t) / c_i(t) \right) \tag{4}$$

Definition of instantaneous frequency is;

$$\omega_i(t) = \frac{d\theta_i(t)}{dt} \tag{5}$$

The original signal can be expressed in instantaneous frequency:

$$s(t) = \mathrm{Re} \sum_{i=1}^{N} a_i(t) e^{j\int \omega_i(t)dt} \tag{6}$$

$\mathrm{Re}(\bullet)$ Is real component, here omitted allowance function. Referred the right side of formula (6) Hilbert spectrum, as:

$$H_H(\omega, t) = \mathrm{Re} \sum_{i=1}^{N} a_i(t) e^{j\int \omega_i(t)dt} \tag{7}$$

To further define the marginal spectrum is:

$$H_h(\omega) = \int_{-\infty}^{\infty} H(\omega, t) dt \tag{8}$$

In experience sieve method there are two kinds of HHT termination conditions: one is component termination conditions; the other is decomposition termination conditions.

3. Speech Enhancement with HHT Method

HHT is unique method to analysis unsteady signal. Compared with the common signal analysis method it has better adaptability.[3] Speech signal is a typical unsteady signal, so the speech enhancement is realized by using HHT transform method is as follows:

Assuming that noise for speech signal

$$y(t) = s(t) + n(t) \tag{9}$$

Where $s(t)$ is pure speech signal, $n(t)$ is the Gaussian white noise which variance is σ^2.

(1) Long speech segmentation. To segment the noise speech signal $y(t)$ and there are no overlap between two paragraphs.

(2) EMD. EMD decomposition in each period of speech signal and obtain I IMF components;

(3) Speech detection. To frame of each IMF component, calculate each frame variance, judging each frame is speech frame or non-speech frame;

(4) Weighted filtering. Each IMF with different weights, the choice of weights according to the characteristics of each IMF;

(5) Soft threshold de-noising processing. For each IMF component set a threshold, soft threshold quantization processing respectively. Threshold is adaptive, related to the layer signal, and related to the layer number;

(6) Reconstruct signal. Reconstructed signal with the IMF components to obtain the enhanced signal.

4. Soft Threshold Setting

We use adaptive local threshold which according to the different scales of the EMD decomposition chose different threshold. According to the characteristics and flexible of each frame of each IMF component to decide whether the frame is a "noise" or "signal", a threshold of adaptive soft threshold quantization process, to achieve the balance between signal de-noising and retention.

Threshold is related to the layer of IMF component and the variance of layer number. Set parameter para as the energy distribution of the EMD decomposition of pure noise, and $para(i)$ is energy density which c_i in all the of the IMF component.

Due to first IMF component of white noise variance is proportional to the actual noise variance in different SNR speech, we use variance or energy of the first IMF component to estimate the noise and estimate the noise variance is:

$$\sigma^2 = v(1)/para(1) \tag{10}$$

Set the threshold changing with the scale and noise characteristics

$$T(i) = \lambda \bullet \sigma \bullet para(i) \bullet \log_{10}\left(\frac{1}{v(i)\sqrt{i}}\right), i = 1,2\cdots,I \tag{11}$$

We use formula (11) to calculate the threshold of each IMF component. The sampling points which less than the threshold value caused by noise, not participate in reconstruction and delete them. We set different threshold parameters between noise and speech. To attenuate the IMF component values of noise frame $(\lambda > 1)$. Set smaller threshold value of the IMF with the speech component to other signal $(\lambda < 1)$. Set $\lambda = 1$ to the IMF component of no main speech signal. Threshold $T(i)$ is changing with scale and choose different threshold in different scale to achieve the purpose of enhancement.
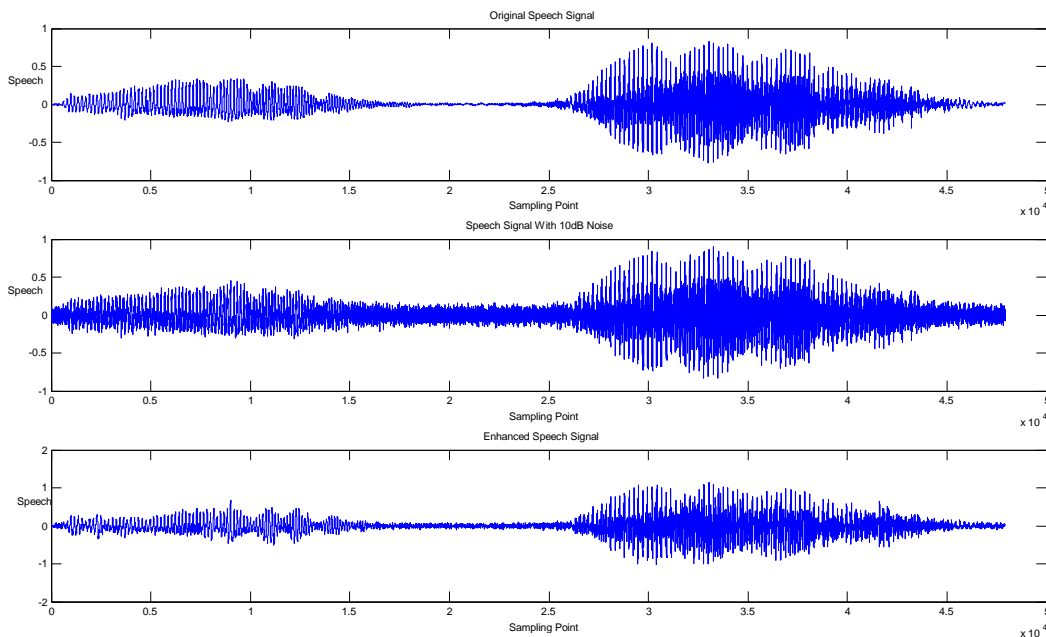
5. Experiment and Result

Fig.1 The enhancement of 10dB noisy speech signal

The speech signal is collected by Cool Edit Pro software and the speech samples using 8 kHz sampling frequency and 8bit quantification, 16 bit sampling precision. The content of the speech signal is Chinese word "ni" and "hao", that because the Chinese word "ni" is begin with voiced consonants and the Chinese word "hao" is begin with voiceless consonant. The recording

environment is quiet, signal can be considered for pure speech signal. We add different SNR Gaussian noise to test the effect of enhancement. The result of experiment is shown as below.

From the figure, we find that the effect of enhancement of noisy speech signal is perfect. When noisy signal SNR=10dB, the enhanced result is SNR=20.6 dB. When noisy signal SNR=0dB,the enhanced result is SNR=10.4 dB. All the SNR has increased about 10 dB. We actual listen to enhanced signal. The speech distortion is small, and naturalness of speech is good and speech intelligibility is well. That's because HHT focus on the local signal characteristics, structural basis function directly by the signal itself, and obtain the different scales of decomposition. This method has good local adaptability and intuitive. Compare reconstruction signal and the original signal, the difference is not big and reduce the amplitude of noise waveform obviously. The algorithm has good effect of speech enhancement.
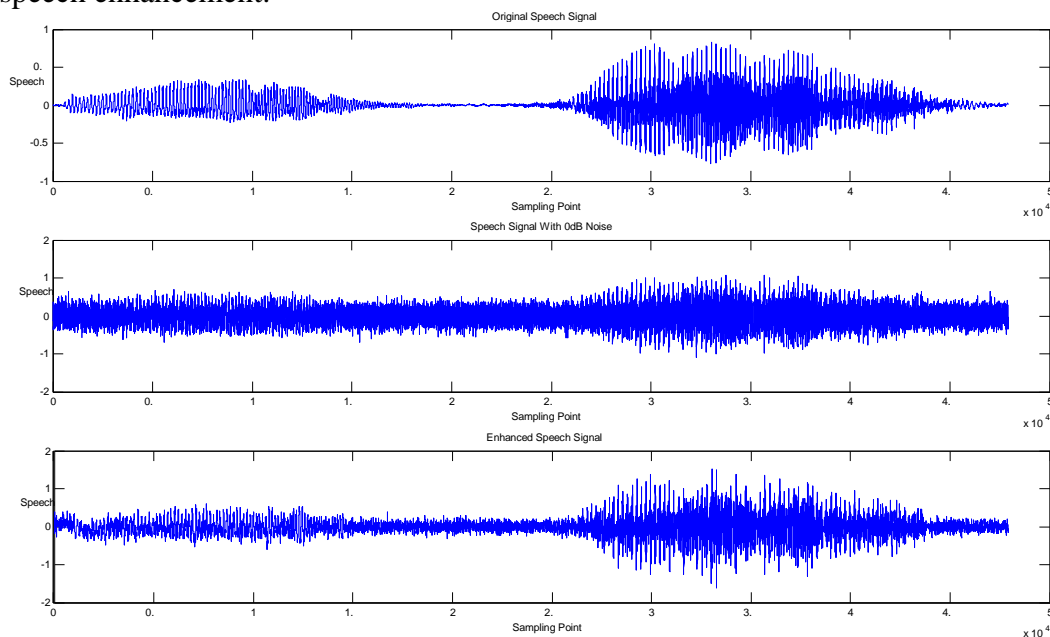

Fig.2 The enhancement of 0dB noisy speech signal

6. Conclusion

This paper we proposed the method soft threshold HHT to enhance the noisy speech signal, this method can separate principal components of IMF of noise signal and obtain the main component of speech signal. Our experiments show that the proposed method can enhance the noisy speech signal and speech distortion is small, and naturalness of speech is good and speech intelligibility is well.

References

[1] Yuan Wenhao, Lin Jiajun, Wang Yu and Chen Ning, A speech enhancement approach based on noise classification, Journal of East China University of Science and Technology, Vol.40 No.2, 2014-04.
[2] Ru Zhen, Li Xin, Chen Fei, Yang Lijian, and Li Xiang, Research and simulation of speech enhancement algorithm base on HHT, Computer Technology and Development, Vol.20 No.8, Aug 2010.
[3] Liu Liwei, Li Jinbao, Zhao Kongxin and Ding Tiefu. The application of Hilbert-Huang transform in speech enhancement, Microcomputer Information, vol.23, No.10-3, 2007.