An EMD based Speech Enhancement Using Adaptive Thresholding Techniques

Nayana C.G[#1]
M tech (DEC), 4th Sem
Srinivas School of Engineering
Mukka,Surathkal,, Manglore
[1]nayanacg22@gmail.com

Mr.Nagaraja N.S[#2]
Assistant Professor, E&CE Dept...
Srinivas School of Engineering
Mukka,Surathkal, Manglore
[2]nagaraj_n_s@yahoo.com

***Abstract*— This paper presents a new speech enhancement algorithm using data adaptive soft thresolding technique. The noisy speech signal is decomposed into a finite set of band limited signals called intrinsic mode functions (IMFs) using empirical mode decomposition (EMD). Each IMF is divided into fixed length frames. On the basis of noise contamination, the frames are classified into two groups – noise dominant frames and speech dominant frames. Only the noise dominant subframes are thresholded for noise suppression. A data adaptive threshold function is computed for individual IMF on the basis of its variance. We propose a function for optimum adaptation factor for adaptive thresholding which was previously prepared by the least squares method using the estimated input signal to noise ratio (SNR) and calculated adaptation factor to obtain maximum output SNR. Moreover, good efficiency of the algorithm is achieved by an appropriate subframe processing. After noise suppression, all the IMFs (including the residue) are summed up to reconstruct the enhanced speech signal. The Hilbert Huang transform(HHT) of the speech is constructed which is useful in order to obtain the instantaneous time and frequency of speech. The comparison between direct EMD based thresholding and adaptive EMD based thresholding are done in terms of subjective measure, spectrogram and waveforms.**

Keywords—*Adaptive thresholding, adaptation factor, empirical mode decomposition, intrinsic mode function, speech enhancement, Hilbert Huang transform (HHT)*

I. INTRODUCTION

In many speech related systems, the desired signal is not available directly; rather it is mostly contaminated with some interference sources. These background noise signals degrade the quality and intelligibility of the original speech, resulting in a severe drop in the performance of the applications. There are different types of noise signals which affect the quality of the original speech. It may be a wide-band noise in the form of a white or colored noise, a periodic signal such as in hum noise, room reverberations, or it can take the form of fading noise. It is also possible that the speech signal may be simultaneously attacked by more than one noise source. The most common type of noise in time series analysis and signal processing is the white noise. That is why this project is mainly concerned in this kind of noise.

The degradation of the speech signal due to the background noise is a severe problem in speech related systems and therefore should be eliminated through speech enhancement algorithms. Speech enhancement aims at improving the perceptual quality and intelligibility of a speech signal in noisy environments, mainly through noise reduction algorithms. Such types of processes may be applied to a mobile radio communication system, a speech recognition system, a set of low quality recordings, or to improve the performance of hearing aids. Speech Enhancement is a classical problem in signal Processing, Particularly in the case of additive white Gaussian noise where different noise reduction methods have been proposed. When noise estimation is available, then filtering gives accurate results. However, these methods are not so effective when noise is difficult to estimate. Linear methods such as Wiener filtering are used because linear filters are easy to implement and design. These linear methods are not so effective for signals presenting sharp edges or impulses of short duration. Furthermore, real signals are often nonstationary. In order to overcome these shortcomings, nonlinear methods have been proposed and especially those based on wavelet thresholding [2-3].

The idea of wavelet thresholding relies on the assumption that signal magnitudes dominate the magnitudes of noise in a wavelet representation so that wavelet coefficients can be set to zero if their magnitudes are less than a predetermined threshold. A limit of the wavelet approach is that basis functions are fixed, and, thus, do not necessarily match all real signals. To avoid this problem, time-frequency atomic signal decomposition can be used. As for wavelet packets, if the dictionary is very large and rich with a collection of atomic waveforms which are located on a much finer grid in time-frequency space than wavelet and cosine packet tables, then it should be possible to represent a large class of real signals; but, in spite of this, the basic functions must be Specified. Recently, a new data-driven technique, referred to as empirical mode decomposition (EMD) has been introduced by Huang [4] for analysing data from nonstationary and nonlinear processes. The major advantage of the EMD is that basis functions are derived from the signal itself. Hence, the analysis is adaptive in contrast to the traditional methods where basis functions are fixed. The EMD is based on the sequential extraction of energy associated with various intrinsic time scales of the signal, called intrinsic mode functions (IMFs)[7-8], starting from finer temporal scales (high-frequency IMFs) to coarser ones (low frequency IMFs). The total sum of the IMFs matches the signal very well and therefore ensures

completeness. We have shown that the EMD can be used for signals denoising or filtering. The denoising method reconstructs the signal with all the thresholded IMFs.

II. EMD BASED SPEECH ENHANCEMENT

The proposed speech enhancement system is based on the IMFs Framed Thresholding Method, which is adaptive in nature. The noisy speech signal is decomposed in to time domain signals called IMFs by EMD, which are also time domain signals. Therefore, the obtained IMFs are divided into 4 ms frames, thus each having 64 data samples for a 16 kHz sampling frequency ,it is more appropriate to use the definition 'frame' to refer to the time frames. These frames are characterized as either a signal dominant or noise dominant frames. The frames are categorized as either signal or noise dominant depending on its speech and noise energy distribution. Fig.1 shows the proposed EMD based adaptive thresholding speech enhancement system.

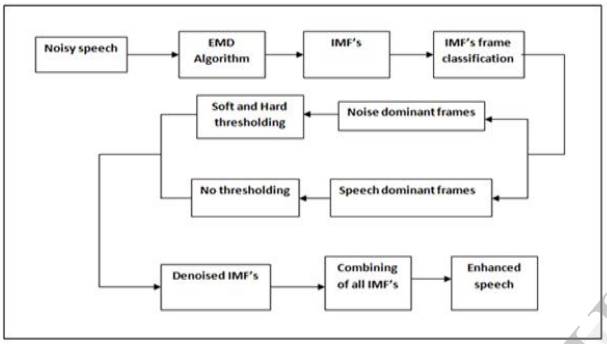

Fig. 1 EMD based adaptive thresholding speech enhancement system

The proposed method involves the following steps.viz...

- Applying EMD algorithm to the noisy speech, which decomposes input signal in to IMF's.
- IMF's segmentation in to frames.
- Classification of frames in to noise and speech dominant frames.
- Adaptive Thresholding (Soft and Hard).
- Combing of denoised IMF's.

A.EMD Algorithm

Empirical mode decomposition (EMD) was recently developed by Huang et al[4,10]. to decompose any non-stationary and nonlinear signal into oscillating components obeying some basic properties, called Intrinsic Mode Functions (IMFs).

The principle of EMD technique is to decompose any signal s(t) into a set of band-limited functions $C_n(t)$, which are zero mean oscillating components, simply called the IMFs. Each IMF satisfies two basic conditions:

1. In the whole data set the number of extrema and the number of zero crossings must be same or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The first condition is similar to the narrow-band requirement for a Gaussian process and the second condition is a local requirement induced from the global one, and is necessary to ensure that the instantaneous frequency will not have redundant fluctuations as induced by asymmetric waveforms. The name intrinsic mode function is adopted because it represents the oscillation mode in the data. With this definition, the IMF in each cycle, defined by the zero crossings, involves only one mode of oscillation, no complex riding waves are allowed. IMF is not restricted to a narrow-band signal; it can be both amplitude and frequency modulated, in fact it can be non-stationary. The idea of finding the IMFs relies on subtracting the highest oscillating components from the data with a step by step process, which is called the sifting process [9].

Although a mathematical model has not been developed yet, different methods for computing EMD have been proposed after its introduction. The very first algorithm is called the sifting process. The sifting process is simple and elegant. It includes the following steps: Consider a signal s(t) for which

- Identify the extrema (both maxima and minima of s(t)).
- Generate the upper and lower envelopes (u(t) and l(t)) by connecting the maxima and minima points by cubic spline interpolation.
- Determine the local mean $m_1(t)=[u(t)+l(t)]/2$
- Since IMF should have zero local mean, subtract out $m_1(t)$ from s(t) to get $h_1(t)$.
- Check whether $h_1(t)$ is an IMF or not.
- If not, use $h_1(t)$ as the new data and repeat steps 1 to 6 until ending up with an IMF.

B. Frame Classification

The categorization of the frame is one of the key points of the soft or hard thresholding algorithm. The main purpose in this categorization is to make it possible to eliminate the noise signals without degrading the original speech components. This makes the soft thresholding algorithm to be applicable for a wide range of SNR values. However, applying this algorithm directly to the IMFs of the noisy speech signal will fail for two reasons. First, IMFs will have different noise and speech energy distribution, which suggests that each IMF will have a different noise and speech variance. Second, due to the decomposition, the variance of the IMF frames will have more fluctuations than that of the noisy speech frames. Therefore the noise variance of each IMF should be defined separately and the limit for frame categorization should have a larger value. In order to guarantee that all the noisy frames are thresholded. A novel limit relies on the idea that a frame can be defined as a noise dominant frame, if the noise power within that frame is greater than the speech power. Therefore, the limit should be set to the case where the noise and speech variance are same. therefore,

in case of equal noise and speech power, with the assumption of independency, the variance of a frame is equal to twice the noise variance. That is why; the limit for the Categorization of frames should be set to two times of the globally estimated noise variance.

C. Variance of the IMFs

The estimation of the variance of each IMF plays an important role in the performance of the proposed EMD domain adaptive thresholding algorithm. In order to estimate the variance, the IMFs are divided into 4ms frames and the variance of each frame is stored in a variance array. The variance array is sorted in ascending order. Since the speechless parts will mostly have the lowest variance, the noise variance of the IMFs can be estimated from these speechless parts of the array.

D. Denoising by soft thresholding

The drawbacks of traditional soft-thresholding algorithms are significantly reduced by the proposed EMD based adaptive thresholding technique. The frame classification criteria described in [5-7] is modified here. The soft thresholding is applied on each IMF. It is known that the thresholding function is dependent on the signal (speech) and noise variances of each IMF. The signal and noise variances are computed for individual IMF.Then soft thresholding technique is applied on each subframe of each IMF on the basis of computed variances. The threshold function is computed for individual IMF and hence such thresholding technique is termed as adaptive thresholding. We calculate the noise variance of speech from its silent part of the observed speech signals. For that, each IMFs is divided into frames of duration of 20 ms. The global noise variance is calculated from the silent part of the *i*th IMF. In order to remove noise from the *i*th IMF, each frame is further subdivided into subframes of duration of 4ms. Then the subframes are classified as either a speech dominant or a noise dominant based on the noise variance. The proposed adaptive thresholding technique provides an effective boundary for the subframe classification. The soft thresholding is carried out on each subframe of each IMF adaptively. After properly suppression of noise using soft-thresholding, all the IMFs are summed up to get the enhanced speech signal.

III. SIMULATION RESULTS

A. IMF's obtained by EMD algorithm

A noisy speech signal and some selected IMF components are shown in Figure.2. It can be observed that higher order IMFs contain lower frequency oscillations than that of lower order IMFs. This is reasonable, since sifting process is based on the idea of subtracting the component with the longest period from the data till an IMF is obtained. Therefore the first IMF will have the highest oscillating components; the components with the highest frequencies. Consequently, the higher the order of the IMF, the lower its frequency content will be. However, the IMFs may have frequency overlaps but at any time instant the instantaneous frequencies represented by each IMF are different. The EMD is not band pass filtering, but is an effective decomposition of non-linear and non-stationary signals in terms of their local frequency characteristics.

The estimation of the variance of each IMF plays an important role in the performance of the proposed EMD domain adaptive thresholding algorithm. In order to estimate the variance, the IMFs are divided into 4ms frames and the variance of each frame is stored in a variance array. The variance array is sorted in ascending order. Since the speechless parts will mostly have the lowest variance, the noise variance of the IMFs can be estimated from these speechless parts of the array. Figure.3 shows a plot of the variance of the frames for the first 6 IMFs of a noisy speech signal at 10dB.

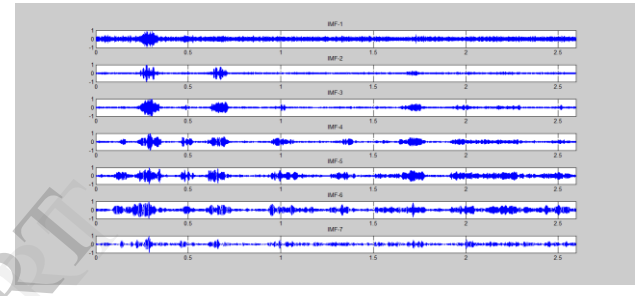

Fig.2 The illustration of EMD. A noisy speech signal at 5 dB SNR and its 15 IMFs, plus a residue signal which can be observed to be close to a constant. The speech signal considered: "She had your dark suit in greasy wash water all year".

B. Sorted variance of frames

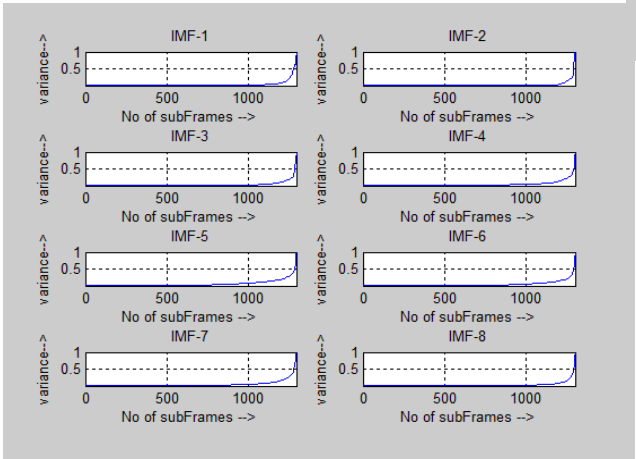

Figure.3 Sorted variance of 4ms frames for the first 8 IMFs of a noisy speech at 15dB SNR

C. Instantaneous Frequency

Instantaneous frequency (IF) represents signal's frequency at an instance and it is defined as the rate of change of the phase angle at the instant of the analytic

(complex) version of the signal. Every IMF is a real valued signal. Analytic signal method is used to compute the instantaneous frequency of the IMF components. The analytical signal corresponding to the m[th] IMF $C_m(t)$ is defined as,

$$Z_m(t)= c_m(t) +j\,h[.][c_m(t)]= \alpha_m(t)e^{j\theta m(t)} \quad (1)$$

Where h[.] refers to the Hilbert transform operator, $\alpha_m(t)$ and $\theta_m(t)$ are the instantaneous amplitude and phase respectively of the *m*th IMF and *j* is the notation of complex term.The Hilbert transform provides a phase-shift of $\pm\pi/2$ to all frequency components, whilst leaving the real parts unchanged. The analytic signal is advantageous in determining the instantaneous quantities such as energy, phase and frequency. Then the IF of the *m*th IMF can easily be derived as,

$$\omega_m(t) = d\,\theta_m(t)/dt \quad (2)$$

Where $\theta_m(t)$ is the unwrapped version of the instantaneous phase $\theta_m(t)$. The concept of IF is physically meaningful only when applied to mono component signals, which have been loosely defined as narrow band. To apply the concept of IF to arbitrary signals it is necessary to first decompose the signal into a series of mono-component IMFs, the derivation of IF on each component provides the meaningful physical information.

D. *Hilbert Spectrum*

Hilbert Spectrum (HS) represents the distribution of the signal energy as a function of time and frequency. Having obtained the IMFs, to construct HS of the signal, the Hilbert transform is applied to each IMF and the instantaneous frequency is computed according to equation (2).Here the residue is left out, since it is either a monotonic function or a constant. It can be observed that the Hilbert Huang Transform representation of the data gives both amplitude and the frequency as function of time. This makes HHT highly efficient for analysing non-stationary signals. Figure 4 illustrates the Hilbert Huang spectrum of the noisy mixture signal shown in Figure 2.

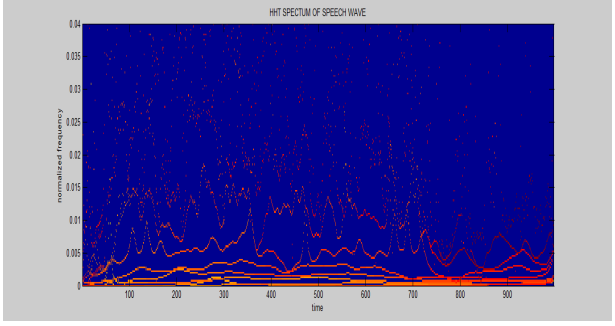

Figure 4. Hilbert Huang Spectrum of the noisy speech signal

IV. EMD BASED ADAPTIVE THRESHOLDING SPEECH ENHANCEMENT TECHNIQUE APPLIED TO NOISY SPEECH

To illustrate the effectiveness of the proposed algorithm, extensive computer simulations were conducted for the selected TIMIT database. In order to observe the performance for a wide range of SNRs waited white noise were added to the clean speech signal to obtain the noisy signal at different SNRs.

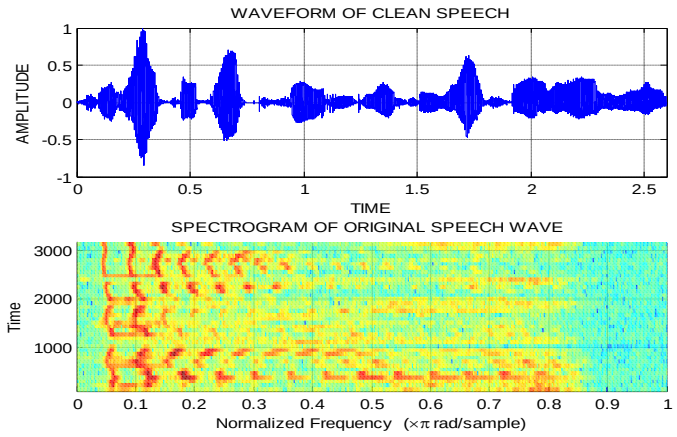

Fig.5 Plot of original speech and its spectrogram

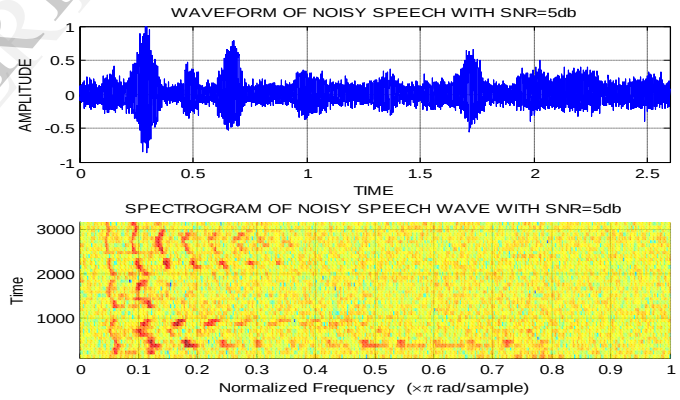

Figure.6 Plot of original speech with additive white Gaussian noise of SNR=5dB and its spectrogram

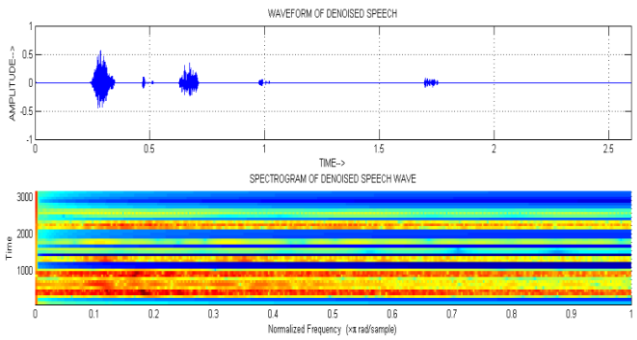

Figure.7 Plot of enhanced speech by denoising using direct EMD based soft thresholding for SNR=5dB

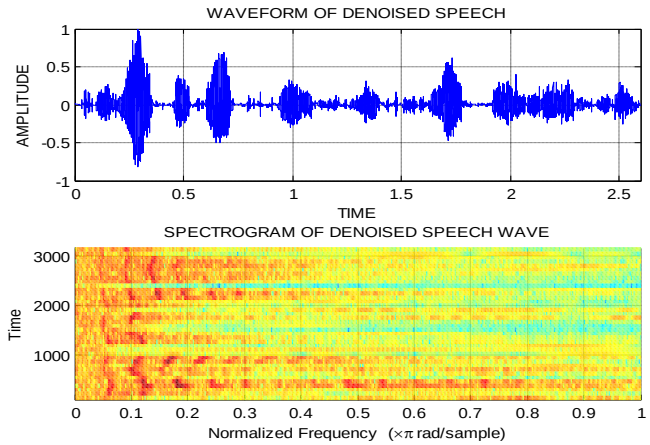

Figure.8 Plot of enhanced speech by denoising using EMD based adaptive hard thresholding with framing for SNR=5dB

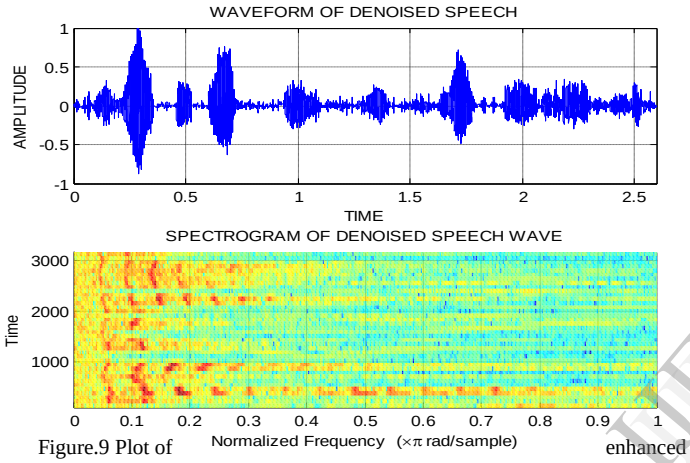

Figure.9 Plot of enhanced speech by denoising using EMD based adaptive hard thresholding with framing for SNR=5dB

V.CONCLUSION AND FUTURE WORK

In this project we have implemented the time methods called direct EMD based thresholding algorithms and adaptive EMD based thresholding algorithms for speech enhancement.A novel data adaptive algorithm is presented to effectively suppress the noise components in all frequency levels of noisy speech signal. The improvement of SNR of noise contaminated speech is achieved by removing noise using EMD based adaptive thresholding technique. An adaptation factor is introduced in the adaptive threshold function. The optimal value of adaptation factor is computed on the basis of estimated input SNR. The experimental result shows that the proposed speech enhancement algorithm works most efficiently for a wide range of input SNR. The performance of this algorithm (in terms of subjective measure, spectrogram and waveforms) is tested with the speech contaminated with white noise. Threshold value to get maximum output SNR is required to be computed.However, the EMD based algorithm suffers from computational complexity and the empirical process takes long time and it is not suitable to apply for real time processing. The further research can be conducted to decrease the computational cost of EMD based methods. Speech enhancement aims at improving the perceptual quality and intelligibility of a noisy speech signal mainly through noise reduction. Speech enhancement may be applied to mobile radio communication system, speech recognition system, robotics etc.

REFERENCES

1. Md. Ekramul Hamid, Somlal Das, Keikichi Hirose and Md. Khademul Islam Molla, "*Speech Enhancement Using EMD Based Adaptive Soft-Thresholding (EMD-ADT)*," International Journal of Signal Processing, Image Processing and Pattern Recognitiong, Vol. 5, No. 2, June, 2012.
2. Navin Chatlani and John J. Soraghan",EMD-Based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement", *IEEE* transactions on audio, speech, and language processing, vol. 20, no. 4, may 2012.
3. Yannis Kopsinis, *Member* and Stephen McLaughlin, "Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding", IEEE transactions on signal processing, vol. 57, no. 4, april 2009.
4. N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-Ch. Yen, C.C. Tung, and H.H. Liu, "*The Empirical Mode Decomposition and The Hilbert Spectrum for Non-Linear and Non-Stationary Time Series Analysis*," in Proc. R. Soc., Lond. A 454, pp.903-995, 1998.
5. S. Salahuddin, et. al., "Soft thresholding for DCT speech enhancement", Electronics Letters, vol. 38, **(2002)**.
6. Hadhami Issaoui, Aicha Bouzid and Noureddine Ellouze,"Comparison *between Soft and Hard Thresholding on Selected Intrinsic Mode Selection*", 2012 6th International Conference on Sciences of Electronics,Technologies of Information and Telecommunications (SETIT).
7. *Denoising by soft thresholding*. Donoho, D. L. 3, s.l.: *IEEE* Trans. on Information Theory, 1995, Vol. 41, pp. 613-627.
8. *Empirical Mode Decomposition of Voiced Speech Signal*. Bouzid, A. and Ellouze, N. s.l. : *IEEE*, 2004. First Internation Symposium on Control, Communication and Signal Processing. pp.603-606..
9. *Enhanced empirical mode decomposition using a novel sifting-based interpolation points detection*. Yannis, K. and Stephen, M. s.l.: *IEEE/SP* Workshop on Statistical Signal Processing, 2007. pp. 725-729.
10. Huang, N. E. and Nii, O. A. O. *The Hilbert Huang Transform in Engineering*. s.l. : CRC Press.