A Comparative Study of EMD and EEMD Based Adaptive Thresholding Method for Speech Enhancement

Nayana C.G[1], Shruthi S Prabhu[2], Parameshachari B.D[3], Rudresh M.D[4]

Assistant Professor[1,2], Associate Professor & HOD[3], Associate Professor[4]

[1,2,3] Department of TCE, [4]Department of ECE

[1,2,3]GSSS Institute of Engineering and Technology for Women, Mysuru,

[4]Kalpataru Institute of Technology, Tiptur, Tumkur, Karnataka

Corresponding Authors: *nayanacg22@gmail.com[1], shruthisprabhu@gmail.com[2], pbdgsss@gmail.com,[3] rudresh_md@rediffmail.com[4]*

Abstract

From last few decades many frequency domain speech denoising algorithms have been proposed for the speech enhancement. Purpose of speech enhancement is to improve the quality of speech which is most prominent in telecommunication applications. The main objective of this paper is to develop a time domain speech denoising algorithm which offers superior performance compared to previous approaches. This paper demonstrates the time domain methods of speech denoising called Empirical Mode Decomposition (EMD). EMD decomposes speech signal corrupted by noise into band limiting signals called intrinsic mode functions. Each of the IMF is thresholded directly; on the other hand the IMF's are divided into frames then thresholded for noise removal. One of the drawbacks of EMD is mode mixing, which is often a consequence of signal intermittency. The mode mixing can be elevated by Ensemble Empirical Mode Decomposition (EEMD) which is the modification of EMD. The performance of the EMD and EEMD algorithm are evaluated for wide range of input SNR's and good efficiency is derived for EEMD based framed thresholding for speech enhancement.

Keywords: *Adaptive thresholding Ensemble empirical mode decomposition, intrinsic mode function, Mode mixing, speech enhancement*

I. INTRODUCTION

Data analysis is an essential part in pure research and practical applications. Linear and stationary signals are easy to analyze, but the real world signals like speech are non-linear and non-stationary [4]. Analysis of such time varying signals is not an easy process. Fourier spectral analysis is an easy method for examining the global energy-frequency distributions. As a result, the term spectrum has become almost synonymous with the Fourier transform of the data. The spectrum gives us the frequencies that exist over the entire duration of the data set. However, the main idea of time-frequency analysis is to understand and describe where the frequency content of the data is changing in time.

The time-frequency (TF) representation, a two-dimensional function which indicates the energy content of a signal as a function of both time and frequency, is a powerful tool for time-varying signals. Therefore, TF representation provides temporal and spectral information simultaneously. There exists a numerous number of TF representation methods of time domain signals, such as short-time Fourier transform (STFT), wavelet transform, Wigner-Ville distribution, evolutionary spectrum, empirical orthogonal function expression. Inside those, STFT and wavelet have dominated the time-frequency analysis in signal processing.

The STFT represents the short time, snapshot like spectral representation which is nothing but a limited time window-width Fourier spectral analysis. It is simply obtained by sliding a selected size window along the time axis and applying Fourier transform in each segment. However since it relies on the traditional Fourier spectral analysis, one has to assume the data to be piecewise stationary. Therefore, in case of non-stationary signals, the STFT has limited usage. Currently, the most famous time-frequency analysis method is wavelet transform. Wavelet transform expands the signal in terms of wavelet functions which are localized in both time and frequency.

The most commonly used wavelet is Morlet, defined as Gaussian enveloped sine and cosine wave groups. The problem with Morlet wavelet is the leakage generated by the limited length of the basic wavelet function, which makes the quantitative definition of the energy-frequency-time distribution difficult. Once the basic wavelet is selected, one has to apply it for the whole data. Moreover, since Morlet wavelet is Fourier based, it also suffers from many shortcomings of Fourier spectral analysis [1].

The most recently introduced technique for analysing non-linear and non-stationary signals is the Hilbert Huang Transform (HHT)[10], which is a combination of Empirical Mode Decomposition (EMD), recently pioneered by Huang et.al. and Hilbert transform (HT). The key ingredient in HHT is the EMD which decomposes the signal in to many modes with different frequency characteristics, called the intrinsic mode functions (IMFs), and thus also alleviates the problem of sharp frequency change in the original signal. IMFs are free of riding waves; hence they give sharp identifications of the instantaneous frequencies. Therefore they are highly suitable for Hilbert transformation.

Once these IMFs are obtained, HT is applied on each IMF in order to obtain the time-frequency representation. Since EMD is specifically introduced for nonlinear and non-stationary signals, it has attracted the attention of the researchers from many areas soon after its introduction and has been implemented in numerous kinds of data, often proving its efficiency and superiority.

One of the major drawbacks of EMD is the mode mixing problem. Mode mixing is defined as a single Intrinsic Mode Function (IMF) either consisting of signals of different scales, or a signal of a similar scale remains in different IMF components. Mode mixing is often a consequence of signal intermittency. The effect of intermittence is serious aliasing in the time–frequency distribution and also the individual IMFs are unclear.

In order to overcome from the problem of mode mixing Huang et al.[11] proposed the intermittence test. But, the approach has its own problems: first, the intermittence test is based on a manually selected scale. With this manual intervention, the EMD ceases to be totally adaptive. Second, the manual selection of scales works if there are clearly separable and definable timescales in the data.

In order to overcome from the scale separation problem in intermittence test, a new noise-assisted data analysis (NADA) method is proposed, the Ensemble EMD (EEMD).In EEMD [11] the true IMF components are the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. This new approach is based on the statistical properties of white noise [12], which showed that the EMD is effectively an adaptive dyadic filter bank when applied to white noise. With this ensemble approach, we can clearly separate the scale naturally without any priori manual time scale selection.

II. EMPIRICAL MODE DECOMPOSITION

Recently developed method to decompose any non-stationary and non-linear signal into oscillating components [3] is Empirical mode decomposition (EMD).EMD obeys some basic properties, called Intrinsic Mode Functions (IMFs).

The principle of EMD technique is to decompose any signal x(t) into a set of band-limited functions Cn(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.

By these two conditions 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. A typical IMF can be observed in Figure 1. 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.

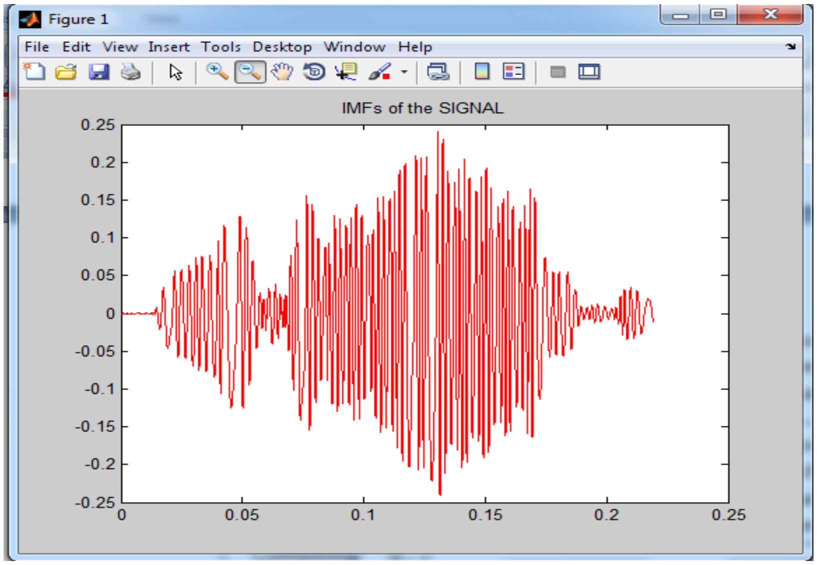


Fig 1. An IMF with same numbers of zero crossings and extreme, of envelopes with respect to zero mean.

One of the simplest methods of EMD for deriving IMF's is the sifting process. The sifting process is simple and elegant. It includes the following steps:

1. Identify the extremas (both maxima and minima of x(t))

2. Generate the upper and lower envelopes (upp(t) and low(t)) by connecting the maxima and minima points by cubic spline interpolation

3. Determine the local mean m1(t)=[upp(t)+low(t)]/2

4. Since IMF should have zero local mean, subtract out m1(t) from x(t) to get h1(t)

5. Check whether h1(t) is an IMF1 or not.

6. If not, use h1(t) as the new data and repeat steps 1 to 6 until ending up with an IMF.

Once the first IMF h1(t) is derived, it is defined as C1(t)=h1(t), which is the smallest temporal scale in x(t). To compute the remaining IMFs, C1(t) is subtracted from the original data to get the residue signal r1(t): r1(t)= x(t)- C1(t). The residue now contains the information about the components of longer periods. The sifting process will be continued until the final residue is a constant, a monotonic function, or a function with only one maxima and one minima from which no more IMF can be derived. The subsequent IMFs and the residues are computed as:

$r_1(t) - C_2(t) = r_2(t), \ldots\ldots\ldots, r_{n-1}(t) - C_n(t) = r_n(t)$ (1)

At the end of the decomposition, the data s(t) will be represented as a sum of n IMF signals plus a residue signal,

$s(t) = \sum_{i=1}^{n} C_I(t) + r_n(t)$ (2)

III. ENSEMBLE EMPIRICAL MODE DECOMPOSITION

In order to overcome from the scale separation problem in intermittence test, a new noise-assisted data analysis (NADA) method is proposed, the Ensemble EMD (EEMD).In EEMD the true IMF components are the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. This new approach is based on the statistical properties of white noise [1], which showed that the EMD is effectively an adaptive dyadic filter bank when applied to white noise. With this ensemble approach, we can clearly separate the scale naturally without any priori manual time scale selection.

The principle of the EEMD:

1. The white noise will be added which populate the whole time–frequency space uniformly with the constituting components of different scales.

2. When signal is added to this uniformly distributed white background, the bits of signal of different scales are automatically projected onto proper scales of reference established by the white noise in the background.

3. A collection of white noise cancels each other out in a time–space ensemble mean; therefore, only the signal can survive and persist in the final noise added signal ensemble mean.

4. Finite amplitude white noise is necessary to force the ensemble to

exhaust all possible solutions; the finite magnitude noise makes the different scale signals reside in the corresponding IMF, dictated by the dyadic filter banks, and render the resulting ensemble mean more meaningful.

5. The true and physically meaningful answer to the EMD is not the one without noise; it is designated to be the ensemble mean of a large number of trials consisting of the noise-added signal.

This EEMD proposed here has utilized many important statistical characteristics of noise. The EEMD utilizes the scale separation capability of the EMD, and enables the EMD method to be a truly dyadic filter bank for any data. By adding finite amplitude white noise, the EEMD eliminated largely the mode mixing problem and preserve physical uniqueness of decomposition. Therefore, the EEMD represents a major improvement of the EMD method.

“Mode mixing” is defined as any IMF consisting of oscillations of dramatically disparate scales, as a result of intermittent signals. When mode mixing occurs, an IMF can cease to have physical meaning by itself. The aliasing at each transition from one scale to another would irrecoverably damage the clean separation of scales. Such a drawback was first illustrated by Huang et al.2 in which the modelled data was a mixture of intermittent high frequency oscillations riding on a continuous low frequency sinusoidal signal.

IV. EEMD BASED SPEECH ENHANCEMENT TECHNIQUES

The idea of finding the IMFs relies on subtracting the highest oscillating components from the data, called the sifting process. Therefore the IMFs have different frequency characteristics; the upper the IMF, the higher its frequency content. The IMFs may have frequency overlaps but at any time instant the instantaneous frequencies represented by each IMF is different, the upper one having the higher frequency. With these powerful characteristics, recent studies have shown that it is possible to successfully identify and remove a significant amount of the noise components from the IMFs of a noisy speech.

Although all IMFs contain energy from both the original speech and the noise, the amount of the energy distribution is different. Since speech signals are mainly concentrated in the low and mid frequency bands, the high frequency noise components dominate the first IMFs. For instance, in case of white noise, most of the noise components are centered on the first three IMFs.

A. EEMD based Soft Thresholding

Soft thresholding strategy proposed is a powerful technique for removing the noise components from the noisy signal while paying attention on the original speech. Since the signal dominant frames are not threshold, the algorithm enables even signals with high SNRs to be processed effectively, where most reported methods even fail to hold on to the input SNR.

Let Y refer the coefficients of the noisy mixture signal and T be the threshold value for the denoising strategy. The Soft thresholding sets any coefficient with absolute value less than or equal to the threshold to zero and subtracts the threshold value from the other coefficients.

$$Y_{Th} = sgn(Y)(|Y| - |T|) \quad \text{if } |Y| > T \qquad (3)$$

$$= 0 \quad \text{if } |Y| < T$$

It can be observed that the soft thresholding algorithm removes more noise components than the hard thresholding algorithm. However in soft thresholding the amount of the signal degradation is also higher. Therefore, the thresholding strategy should be selected depending on the subjective and objective perspectives.

B. EEMD based Hard Thresholding

Hard thresholding sets any coefficient whose absolute values is less than or equal to the threshold to zero;

$$\hat{Y}_{Th} = Y \quad \text{if } |Y| > T \qquad (4)$$

$$= 0 \quad \text{if } |Y| < T$$

The value of *T* is related with the estimated standard deviation of the noise signal σ and may change depending on the proposed algorithm. Donoho has suggested the following formula for its value;

$$T = \sigma \sqrt{2\log(n)} \qquad (5)$$

C. IMFs Thresholding Method for Speech Enhancement

The Block Diagram of IMFs thresholding for speech enhancement is shown in fig 2. In this method noisy Speech is firstly passed

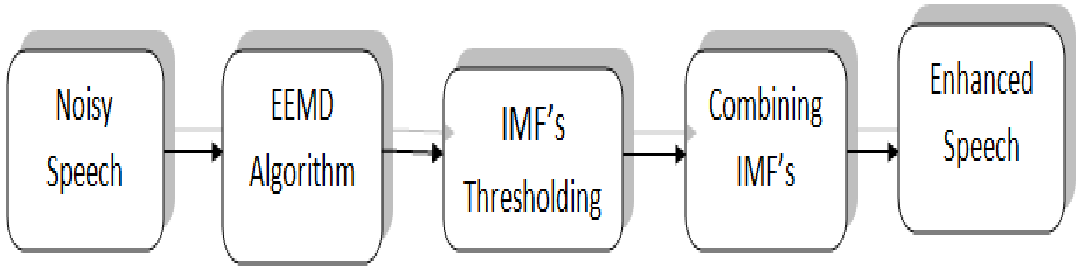


Fig 2: EEMD based IMF's thresholding speech enhancement system

through the EEMD Algorithm. The outputs of EEMD are various intrinsic time scales of the signal, called intrinsic mode functions (IMFs), these IMFs are denoised by using hard threshold techniques or soft threshold techniques by calculating threshold value by using equation (3) and (4) and then these denoised IMFs are combined to get denoised Speech.

In this method soft thresholding method gives better results than hard threshold techniques but draw back associated with both the methods is both methods eliminate noise completely but along with noise, speech data can also get eliminated this will degrade the quality of speech so we need to eliminate only noise by retaining the speech useful information this can be achieved by framing the IMFs and then thresholding this can be discussed by next method explained below.

D. IMFs Frame Thresholding Method for Speech Enhancement

In the IMFs Framed Thresholding Method is shown figure 3. First of all, EEMD is applied to the noisy speech in order to obtain the IMFs of the signal. In the IMFs Framed Thresholding Method [3], the noisy speech signal is a decomposition of the time domain signals into IMFs by EEMD, which are also time domain signals. Therefore, the obtained IMFs are divided into fixed length frames, 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 a noise dominant frame. The frequency bins are categorized as either signal or noise dominant depending on its speech and noise energy distribution.

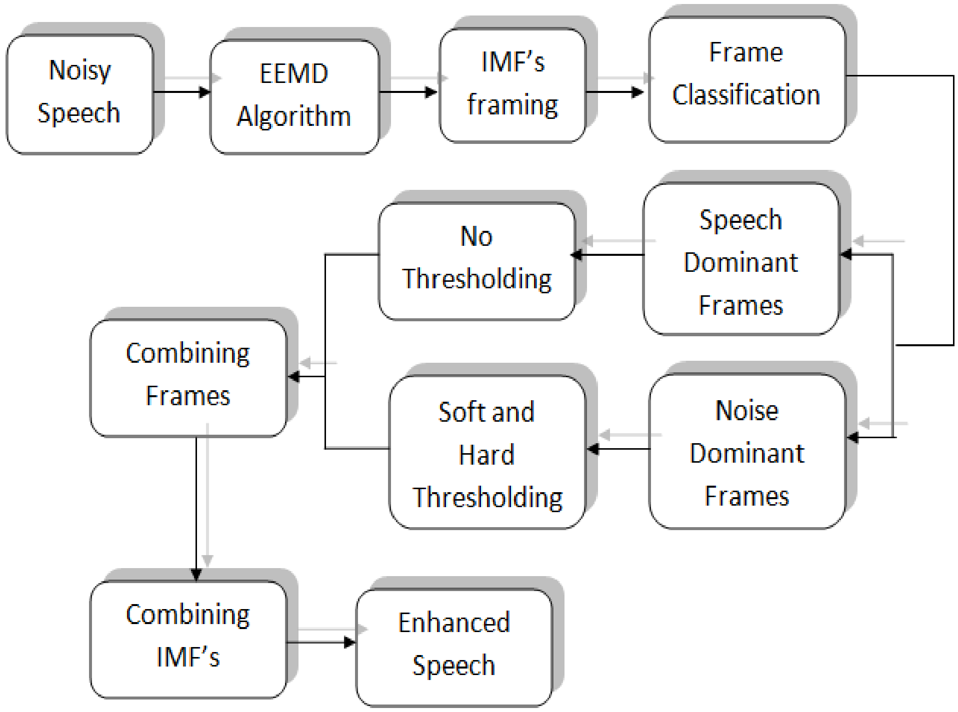


Fig 3.EEMD based IMF's framed thresholding speech enhancement system.

A novel boundary relies on the idea that a frame can be defined as a noise-dominant, if the noise power is higher than the power of the observed signal within that frame. The boundary is set to the case where the noise and speech variances are equal. Hence, generally for any frame, we can write

$$var^2=var^2_{s+n} \quad (5)$$

$$var^2= var^2_s+ var^2_n+2.cov(s,n) \quad (6)$$

Where, var^2_s and var^2_n denote to the speech and noise variance of a frame. Since, speech and noise are independent, the covariance between the two will be zero and thus we have,

$$var^2=var^2_s+ var^2_n \quad (7)$$

To properly classify the frames as speech dominant and noise dominant, the threshold point is selected at which the speech and noise variances are equal. Then the signal variance (at threshold point) can be written as:

$$var^2= 2var^2_n \quad (8)$$

Therefore, in case of equal noise and speech power and with the assumption of independency, the variance of a frame is equal to twice the noise variance of that frame. The classification condition of r^{th} frame of i^{th} IMF defined as:

$$\sigma_i^r \geq 2\, var^2_{n,I} \qquad (9)$$

The average energy of frame r of i^{th} IMF is calculated as:

$$E_i^r = \frac{1}{Q}\sum_{q=1}^{Q}(Y_{q,i}^{(r)})^2 \qquad (10)$$

Where Q is the sample length of the frame.

$Y_{q,i}^{(r)}$ denotes the samples of r^{th} frame of the i^{th} IMF, and $var^2_{n,i}$ denotes the globally estimated noise variance of that IMF. Then the proposed classification condition for rth frame of ith IMF is expressed as:

if

$$E_i^{(r)} \geq 2\,\sigma^2_{n,i}, \qquad (11)$$

$$X_{(i)}^{(r)} = X_{i(s)}^{(r)}(t)$$

Else

$$X_{(i)}^{(r)} = X_{i(n)}^{(r)}(t) \qquad (12)$$

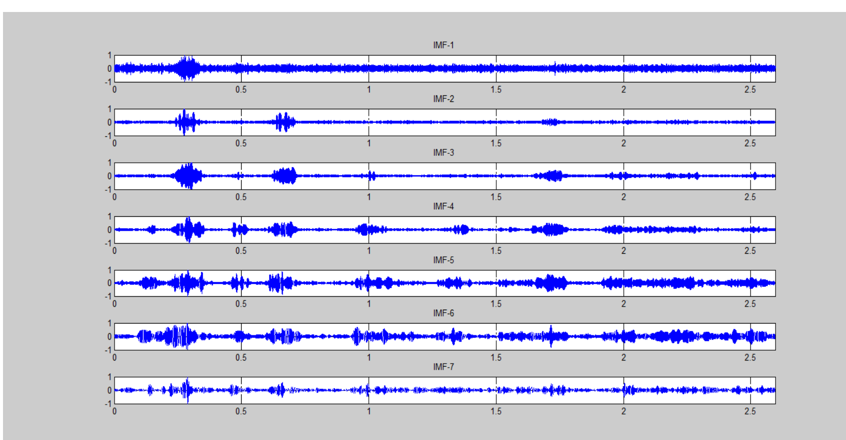
Where $X_{i(s)}^{(r)}(t)$ and $X_{i\,(n)}^{(r)}(t)$ are the classified speech and noise dominant frames, respectively.

The speech dominant frames are not thresholded.

V. SIMULATION RESULTS

A. IMF's obtained by EMD algorithm

A noisy speech signal and some selected IMF components are shown in Figure.4.



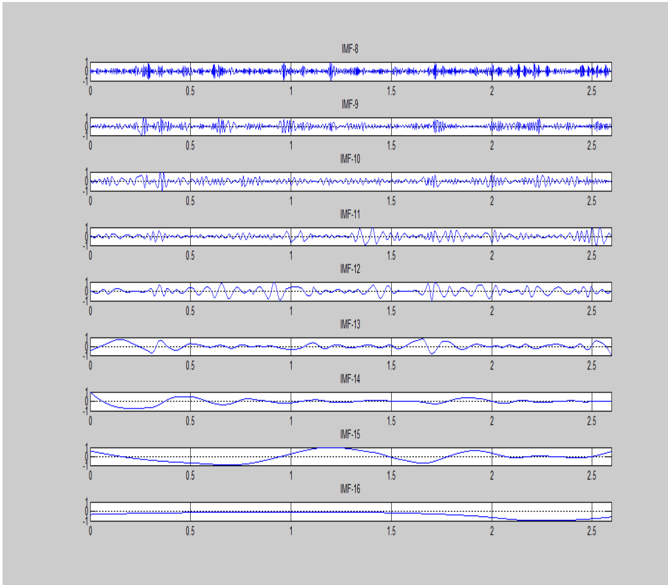


Fig.4 The illustration of EMD . A noisy speech signal at 5 dB SNR and its 15 IMFs, plus a residue signal which can be observed to be close to a constant.

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.

VI. EMD AND EEMD 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.

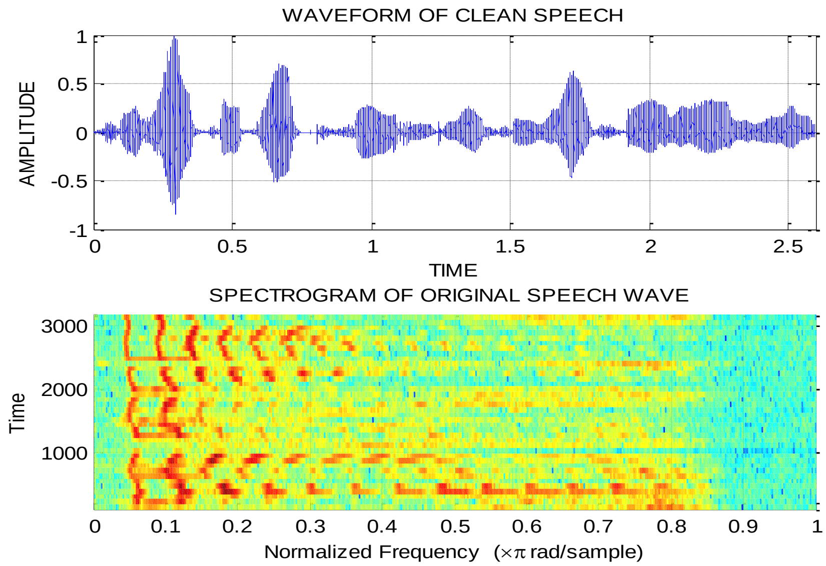


Fig.5 Plot of original speech and its spectrogram

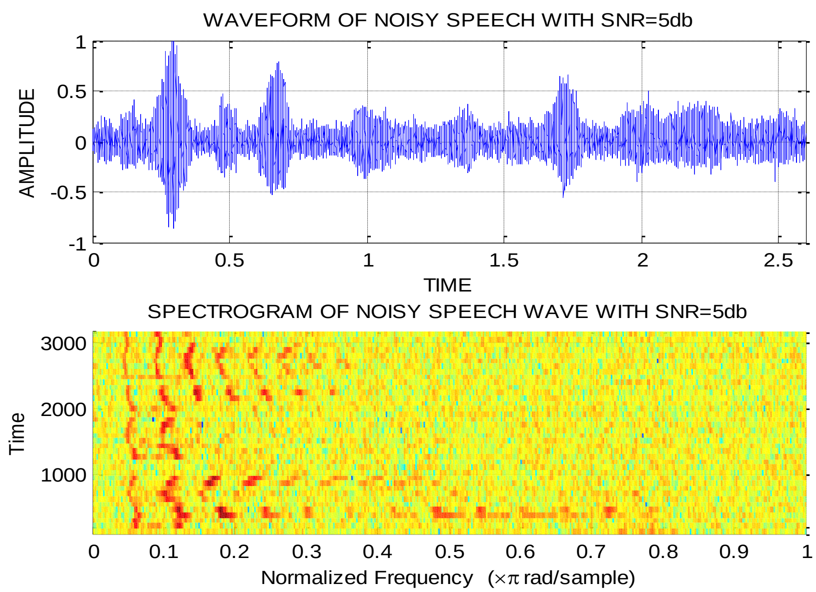


Figure.6 Plot of original speech with additive white Gaussian noise of SNR=5dB and its spectrogram

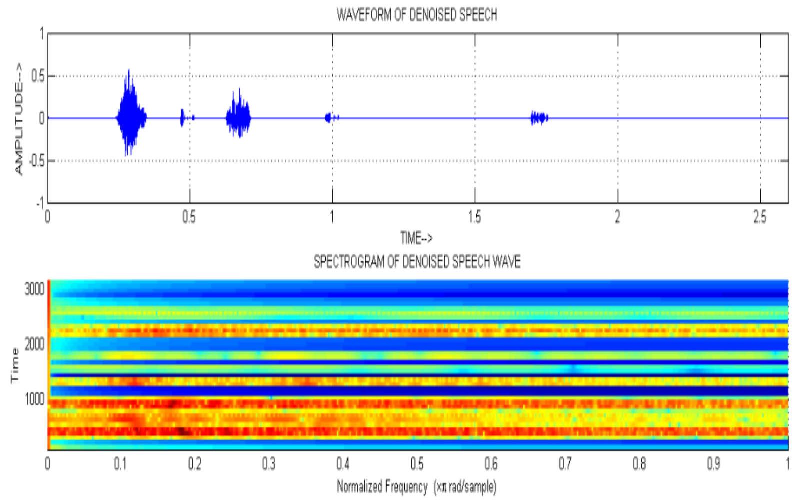


Figure.7 Plot of enhanced speech by denoising using EMD based IMFs thresholding for SNR=5dB

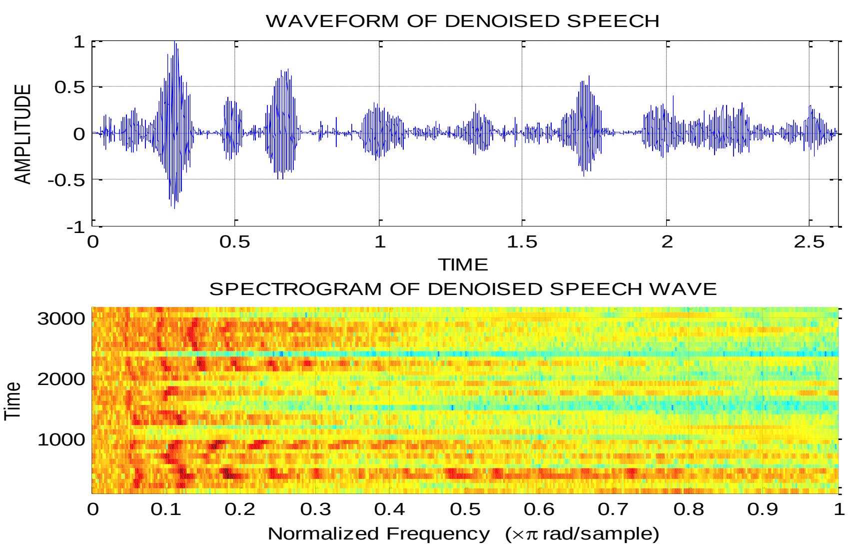


Figure.8 Plot of enhanced speech by denoising using EMD based IMFs frame thresholding for SNR=5dB

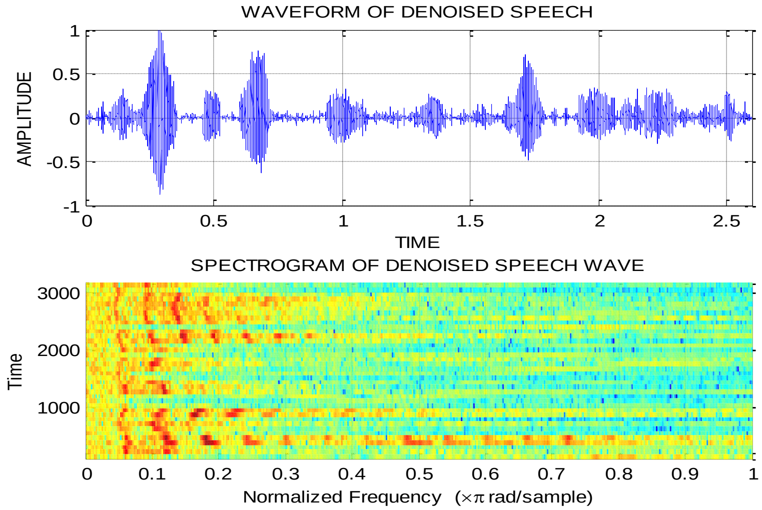


Figure.9 Plot of enhanced speech by denoising using EEMD based IMFs frame thresholding for SNR=5dB

Signal to noise ratio (SNR)

SNR is the ratio of the power of the wanted signal to the noise power.

SNR_{dB}=10 log10 (Psignal / P_{noise})

Table 1 shows comparison of output SNR of EMD and EEMD based IMF's framed thresholding method and direct IMF's thresholding method for wide range of input SNR's.

Table 1 Comparison of output SNR of EMD and EEMD based IMF's framed thresholding method and direct IMF's thresholding method corrupted by white noise.

Input SNR(dB)	**Output SNR(dB) EMD (With framing)**	**Output SNR(dB) EMD (Without framing)**	**Output SNR(dB) EEMD (With framing)**	**Output SNR(dB) EEMD (Without framing)**
5	8.3128	1.1857	1.2348	9.3128

10	11.2287	2.0603	2.3126	11.7652
15	12.862	2.7960	2.9181	13.2134
20	13.6679	3.1676	3.4123	13.8567
25	13.8172	3.3766	3.6122	14.1290

In the above table the output SNR of EEMD with IMFs framing is good compared to EMD with IMFs framing method. Hence the proposed method is able to remove the noise at all levels of speech signal and is good at low input SNR.

V. CONCLUSION

In this paper, two time domain methods called EMD based IMFs frame thresholding and EEMD based IMFs frame thresholding algorithms for speech enhancement has been proposed. 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 EEMD based IMs frame thresholding technique. 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.

The time domain methods called EMD based IMFs frame thresholding and EEMD based IMFs frame thresholding are widely useful for speech enhancement. 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 a mobile radio communication system, speech recognition system, robotics etc.

REFERENCES

[1] Yannis Kopsinis and Stephen McLaughlin “Development of EMD-Based Denoising Methods inspired by Wavelet Thresholding” IEEE transactions on signal processing, vol. 57, no. 4, april 2009.

[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] 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.

[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] Erhan Deger, "Noise-Thresholding with Empirical Mode Decomposition for Low Distortion Speech Enhancement," in the university of Tokyo, February 2008.

[6] Bouzid, A. and Ellouze, N. s.l. "Empirical Mode Decomposition of Voiced Speech Signal" IEEE, 2004. First Internation Symposium on Control, Communication and Signal Processing. pp.603-606..

[7] Barry Commins, "Signal Subspace Speech Enhancement with Adaptive Noise Estimation" Submitted for the degree of M. Eng. Sc. to the National University of Ireland, GalwaySeptember 2005 .

[8] Molla, M. K. I., Hirose, K. and Minematsu, N. s.l. "Separation of mixed audio sources by decomposing Hilbert spectrum with modified EMD" IEICE Transaction on Fundamentals of Electronics, Communication and Computer Sciences, March, 2006.

[9] Yannis, K. and Stephen, M. s.l "Enhanced empirical mode decomposition using a novel sifting-based interpolation points detection". IEEE/SP Workshop on Statistical Signal Processing, 2007. pp. 725-729.

[10] Norden E. Huang "Introduction to the Hilbert Huang Transform and its related mathematical problems.

[11] Bhawna Sharma, Sukhvinder Kaur,"Distinction Between EMD & EEMD Algorithm for Pitch Detection in Speech Processing" International Journal of Engineering Trends and Technology (IJETT) – Volume 7 Number 3- Jan 2014.

[12] Zhaohua Wu nd Norden E Huang , Empirical mode Decomposition : A noise Assisted data analysis method , Advances in adaptive data analysis Vol .1, No.1(2009) 1-41, c , World scientific publishing company.