An Improvement in Representation of Audio Signal in Time-Frequency Plane using EMD-2TEMD Based Approach

Kazi Mahmudul Hassan[1*], Ekramul Hamid[2] and Takayoshi Nakai[3]

[1]*Department of Computer Science and Engineering, Varendra University, Rajshahi 6204, Bangladesh.*
[2]*Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh.*
[3]*Department of Electric and Electronic Engineering, Shizuoka University, Hamamatsu-shi, Japan.*

Corresponding author: munnakazi92@gmail.com

Abstract

This study proposed an enhanced time-frequency representation of audio signal using EMD-2TEMD based approach. To analyze non-stationary signal like audio, time-frequency representation is an important aspect. In case of representing or analyzing such kind of signal in time-frequency-energy distribution, hilbert spectrum is a recent approach and popular way which has several advantages over other methods like STFT, WT etc. Hilbert-Huang Transform (HHT) is a prominent method consists of Empirical Mode Decomposition (EMD) and Hilbert Spectral Analysis (HSA). An enhanced method called Turning Tangent empirical mode decomposition (2T-EMD) has recently developed to overcome some limitations of classical EMD like cubic spline problems, sifting stopping condition etc. 2T-EMD based hilbert spectrum of audio signal encountered some issues due to the generation of too many IMFs in the process where EMD produces less. To mitigate those problems, a mutual implementation of 2T-EMD & classical EMD is proposed in this paper which enhances the representation of hilbert spectrum along with significant improvements in source separation result using Independent Subspace Analysis (ISA) based clustering in case of audio signals. This refinement of hilbert spectrum not only contributes to the future work of source separation problem but also many other applications in audio signal processing.

Keywords: Time-Frequency Representation; HHT; Hilbert Spectrum; EMD; 2T-EMD; Filter Bank; Fractional Gaussian Noise, Independent Subspace Analysis;

INTRODUCTION

Audio signals are information rich nonstationary signals which play an important role in our day-to-day communication, perception of environment, and entertainment. Because of its non-stationary nature of audio signal, time or frequency only approaches are inadequate to analyzing these signals. A joint time-frequency (TF) representation approach would be a better choice which provides some temporal and spatial information simultaneously.

Due to the several limitations of existing time-frequency methods like Fourier Transform, Short-Time Fourier Transform, Wavelet Transform etc., in 1998 Huang et al proposed a new method known as Hilbert–Huang Transform (HHT) which includes Empirical Mode Decomposition (EMD) and Hilbert Transform (HT) to analyze non-linear and non-stationary signals in time, frequency and energy domain [1].The outcome of HHT process is instantaneous frequency, amplitude and phase which are used to generate Hilbert

Spectrum (HS), a new representation of a nonstationary signal like audio in 3D graph where time-frequency-energy is represented by x, y, and z axis respectively. Using EMD method, any complicated data set can be decomposed into a finite and often small number of components called Intrinsic Mode Functions (IMF). EMD process suffers from some mathematical problem like prediction problem, spline problems, optimization problems etc. [2]. Most importantly it produces less number of IMFs to make a successful decomposition of an audio signal in time-frequency domain [3].

Turning Tangent Empirical Mode Decomposition (2T-EMD) is a novel EMD algorithm which differs from other approaches of EMD like EEMD, BEMD etc. by its computational lightness and algorithmic simplicity, is developed to overcome some limitations of EMD like boundary conditions, sifting stopping condition etc. [4]. But 2T-EMD method decomposes a signal with so many redundant IMFs which makes the Hilbert Spectrum representation of an audio signal clumsy and unfruitful for relevant applications.

Besides Independent Subspace Analysis (ISA) is a prominent method for separating individual audio sources from a single-channel mixture [13]. ISA is based on Independent Component Analysis (ICA), where ISA relaxes the constraints of ICA which requires at least as many mixture observation signals as sources. Second improvement over ICA is the use of dynamic components to represent nonstationary signal. Similarity of dynamic components over small time steps is used to track the sources. Partitioning a matrix of independent component cross-entropies which is known as ixegram is used for grouping components. The ixegram measures the mutual similarities of components in an audio segment and clustering the ixegram yields the source subspace and time trajectories [13].

In this work, a mutual implementation of 2T-EMD and classical EMD method is proposed based on the analysis of filter bank properties of both methods with respect to Fractional Gaussian Noise (fGn) which is a generalization of ordinary white noise and also a versatile model of homogeneously spreading broadband noise without any dominant frequency[5]. EMD acts essentially as a dyadic filter bank is stochastic situation which is observed in numerical experiments based on broadband noise like fGn [6]. This paper work try to figure out 2T-EMD property in stochastic situation using similar experiment used in case of EMD and proposed a method based on the outcome. In this proposed method, initially a nonstationary signal like audio is decomposed with 2T-EMD method until it follows dyadic filter bank properties. A threshold point is needed to be selected (which is in between IMF index 10 to 14 in case of audio signals) based on that filter bank properties which can vary according to other applications. This decomposition generates an IMFs set and a residue signal where this residue signal is further decomposed with classical EMD method which produces another set of IMFs. Finally those two IMF set are concatenated to form final set of IMFs which later on hilbert transformed and plotted into time-frequency plane. In the last section of this paper, some experimental results of separating individual audio sources from a single channel mixture using ISA based clustering which is applied on the outcome of proposed method is showed, where each spectrogram represents the separated sources in time-frequency-energy distribution.

TURNING- TANGENT EMD BASICS

The hilbert spectrum is a relatively new joint time-frequency representation introduced in [1]. Two phases are required to generate the hilbert spectrum. In the first phase, EMD is employed, which is an adaptive decomposition method [6]. Then discrete hilbert transform (DHT) is employed in the second phase. HS is generated by the combination of EMD and DHT. EMD focuses on the level of local oscillations and decomposes the signal into a finite set of AM-FM oscillating components which are bases of the decomposition. The principle of the EMD technique is to decompose a signal s(t) into a sum of the band-limited functions $\alpha_m(t)$ or bases called intrinsic mode functions (IMFs). Each IMF satisfies two basic conditions: (i) in the whole data set, the number of extrema and the number of zero crossings must be the same or differ at most by one, (ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero [1]. To successfully decompose an audio signal, the classical EMD method produces a few number of IMFs which is insufficient for different application like source separation, frequency analysis etc.

Turning Tangent Empirical Mode Decomposition also called 2T-EMD is another approach of EMD algorithm which differs from other by its lightness of computation and simplicity [4]. The 2T-EMD algorithm redefines the signal mean envelope and offers the possibility to decompose multivariate signals without any projection. There are several improvement is done in 2T-EMD over EMD.

Firstly, a robust computation of the mean trend is preferably obtained for 2T-EMD by averaging two envelops:

A first envelop interpolates the even indexed barycenters which include signal borders, and a second envelop interpolates the odd indexed barycenters which also include signal borders. i.e. let $h_{even}(t)$ and $h_{odd}(t)$ are even and odd indexed upper envelop respectively. Similarly $l_{even}(t)$ and $l_{odd}(t)$ are even and odd indexed lower envelop respectively. So the ultimate local mean $\mu_1(t)$ will be calculated as –

$$\mu_1(t) = ((h_{even}(t) + h_{odd}(t))/2 + (l_{even}(t) + l_{odd}(t))/2)/2 \qquad (1)$$

which is actually an average of two local mean $\mu_{even}(t)$ (even indexed) and $\mu_{odd}(t)$ (odd indexed)[4].

Secondly, the interpolation is performed using cubic splines with classical boundary conditions as for the classical EMD, where the signal border is directly added to the list of estimated oscillation barycenters.

Thirdly, the sifting process is stopped using a Cauchy-like criterion. In case of audio signal, using 2T-EMD rather than EMD produces a great number of IMFs. Instead of mitigating the issues of limited number of IMFs in EMD, it produces many redundant IMFs especially in lower frequency band of any audio signal. Due to this, the hilbert spectrum representation of any audio signal using 2T-EMD method instead of classical EMD becomes quite meaningless especially in the lower frequency band [3].
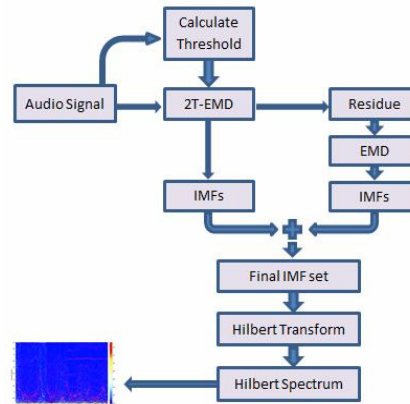
Fig. 1. Block diagram of proposed method

PROPOSED METHOD

To mitigate those problems, a method of combined implementation of 2T-EMD and classical EMD is proposed which is given below:

St. 0: Choose a *threshold* point based on filter bank properties for selecting number of IMFs from 2T-EMD method using fGn.

St. 1: Decompose the signal with 2T-EMD method to get first IMFs set and a *residue signal*.

St. 2: Decompose the *residue signal* with classical EMD method and get another IMFs set.

St. 3: Concatenate the 2T-EMD IMFs set with EMD IMFs set to make the final set of IMFs.

St. 4: Hilbert Transform the IMFs and generates Hilbert Spectrum.

The block diagram of proposed method is shown in Fig. 1.

ANALYSIS

To select a threshold point as stated in the proposed method, we need to analyze 2T-EMD methods filter bank property based on fractional Gaussian noise (fGn). In order to better understand the way 2T-EMD behaves in stochastic situation involving broadband noise, we report here on numerical experiment based on fGn. In case of EMD such an experiment emits that it acts essentially as a dyadic filter bank resembling those involved in wavelet decompositions [7].

For a given signal x(t), EMD and 2T-EMD both ends up with a representation of the form:

$$x(t) = m_k(t) + \sum_{k=1}^{K} d_k(t) \qquad (2)$$

Where $m_k(t)$ stands for a residual "trend" and the "modes" $\{d_k(t), k=1,... K\}$ are constrained to be zero-mean AM-FM waveforms [6].

The fractional Gaussian noise (fGn) is used here as a reference signal which is generalization of ordinary white noise. It is a versatile model of homogeneously spreading broadband noise without any dominant frequency band [8]. Recall that fGn is defined as the increment process of fractional Brownian motion (fBn) [9]. The statistical properties fGn are entirely determined by its second-order structure, which depends solely upon one single parameter, Hurst exponent [10]. In case of EMD it follows the filter bank property where H can vary from 0.1 to 0.9. The data length has been typically set to N=4096 and, for each value of H, 5000 individual sample paths generated via the Wood and Chan's algorithm [11]. In this experiment we set H=0.6 based on preliminary analysis.

(a) (b)

(c)
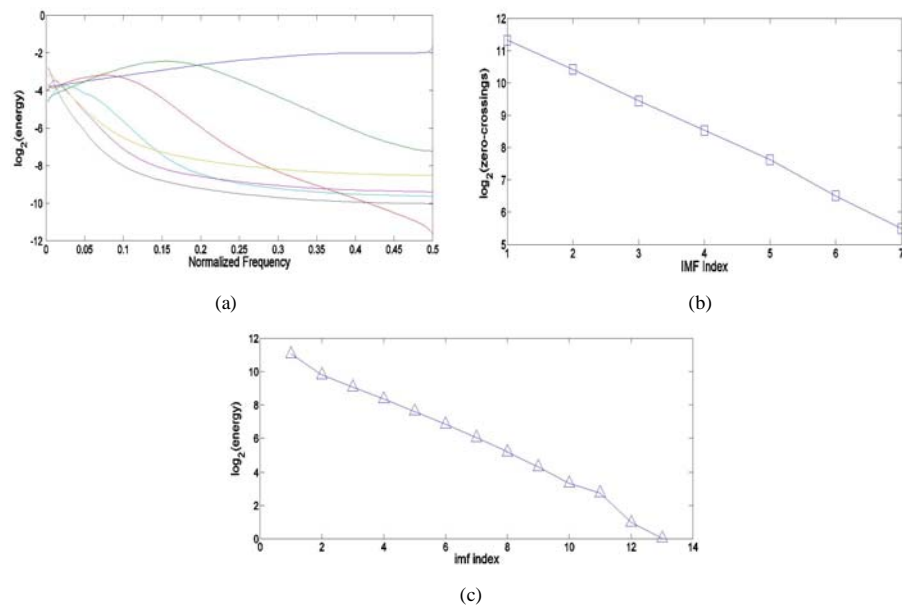
Fig. 2. Apply EMD on fGn, the average spectra of the 7 first IMFs
(a) Log_2(energy) vs. normalized frequency
(b) Log_2(zero crossing) vs. IMF index
(c) Log_2(energy) vs. IMF index

Applying EMD on each of the sample path of fGn generates separate IMFs set. In Fig. 2 (a) the average spectra of the 7 first IMFs are plotted as a function of normalized frequency. Fig. 2(b) shows the logarithm (base 2) of the average number of IMFs zero-crossing which is decreasing linearly as the increase of IMF index [6]. And finally Fig. 2(c) shows the logarithm (base 2) of the average energy of IMFs which are similarly decreasing like zero-crossing as the increase of IMF index.

Applying similar experiment using 2T-EMD on fGn with that same parameter, the following result is observed shown in Fig. 3. It shows that in case of 2T-EMD, the
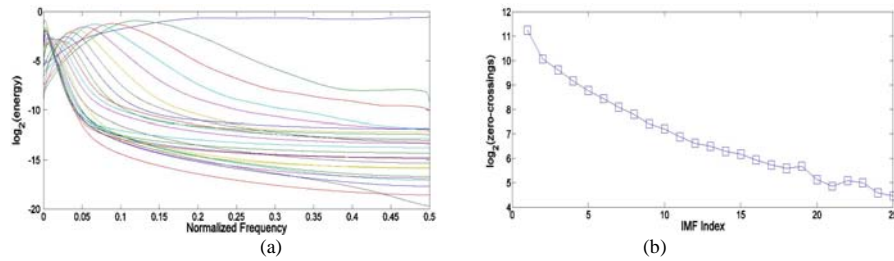
Fig. 3. Apply 2T-EMD on fGn, the average spectra of first 25 IMFs
(a) Log_2(energy) vs. normalized frequency
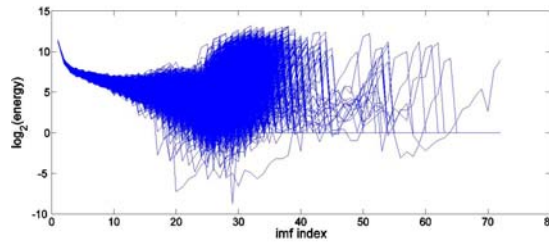(b) Log_2(zero crossing) vs. IMF index

Fig. 4. Apply 2T-EMD on fGn's each sample path, Log_2(energy) of each
IMF set vs. IMF index

relation of normalized frequency bin index and zero-crossing of IMFs with IMFs index is not linear like EMD. Also it is visible that the frequency bin overlapping issues in Fig.3(a) due to the excessive number of IMFs. Plotting individual IMF sets log (base 2) energy with respect to IMF index, we get the result (shown in Fig. 4) where a huge number of IMFs energy is increasing after first few and as the IMF index increased, this rate is increasing too which is contradictory with previous study [12].

Taking histogram of first energy increment IMF index of each sample fGn path IMFs set, the outcome is following (see Fig. 5):
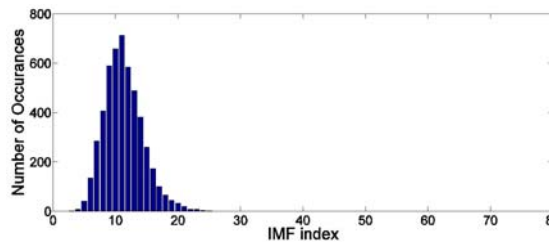

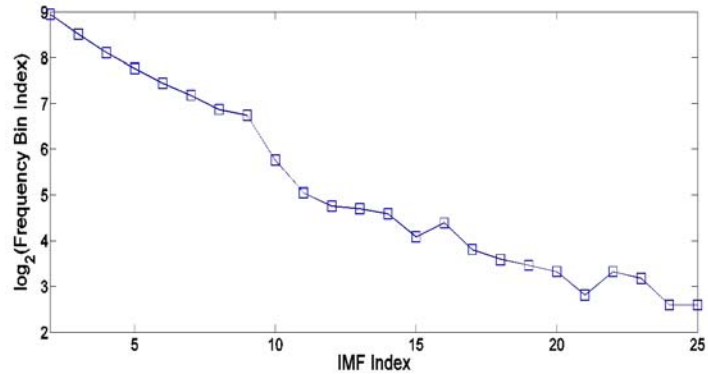Fig. 5. Histogram of fGn IMFs energy increment (1st) (Apply 2T-EMD on the fGn)

Fig. 6. IMFs Log$_2$ (norm. freq. bin index) vs. IMF index (Apply 2T-EMD on the fGn)

According to the average log$_2$ (zero-crossing) of IMFs, the rate of changes in normalized frequency bin index decreased inconsistently after first few(10 to 14) IMFs (see Fig. 6). Experiment result exhibits inconsistency of maintaining filter bank property by 2T-EMD method after first few IMFs (10 to 14 IMF index approx.) generation. A threshold point has to be selected (between IMF indexes 10 to 14 which can varies depending on application) to stop 2T-EMD methods IMF generation process.

RESULTS AND DISCUSSIONS

The efficiency of the proposed method is tested by representing real world audio signal from their mixture. The sample audio mixture length is one second which is sampled at 16 kHz sampling rate and 16-bit amplitude resolution, containing flute sound and male voice pronouncing 1, 2, 3 and 4 with some short interval between each number.

As discussed earlier, pure 2T-EMD method produces so many redundant IMFs especially in lower frequency band of audio signal which makes the hilbert spectrum representation clumsy. Experiment result shows that the rate of changes in IMF's zero-crossing decreasing slowly with the increment of IMF index which can be observed from Fig. 3(b) and due to this, higher index IMF causes frequency bin overlapping problem in hilbert spectrum (see Fig. 7(a)). Fig. 7(b) shows that proposed method mitigates frequency bin overlapping problem by reducing number of IMFs and with the increment of rate of change in zero-crossing of IMFs (see Fig. 8(a)) and also the average log$_2$ (energy) of IMFs are continuously decreasing with the increment of IMF index (see Fig. 8(b)). The only energy increment is in the concatenation point which is because of different mean definition of EMD and 2T-EMD methods. Hilbert spectrum of sample audio signal using pure 2T-EMD method and proposed method are shown in Fig. 7(a) and 7(b) respectively.
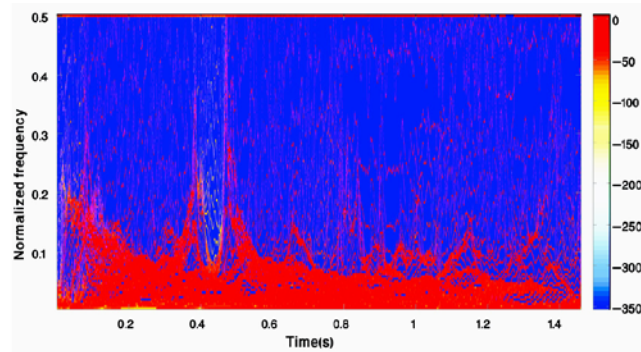
(a)
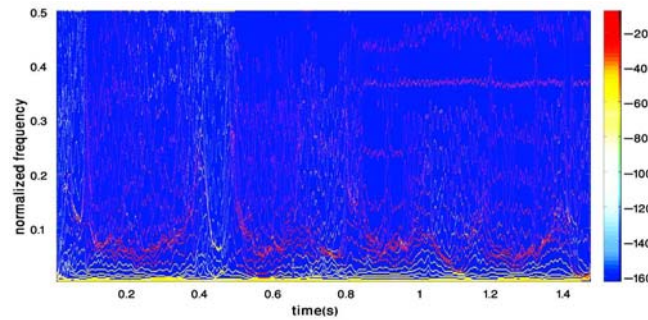
(b)
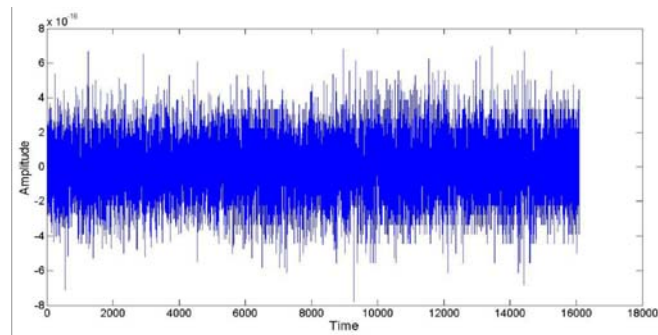
(c)

Fig. 7. Hilbert spectrum of sample audio signal using
(a) pure 2T-EMD method, (b) proposed method
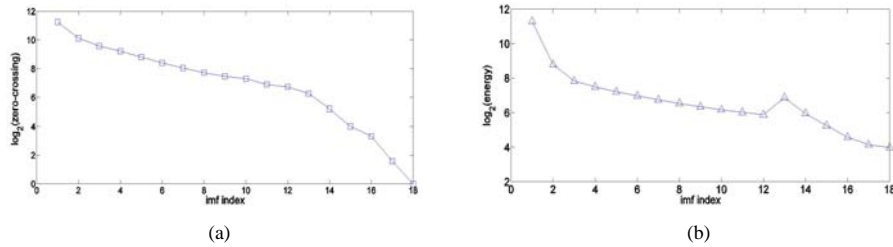and (c) reconstruction error signal

(a)   (b)

Fig. 8. Apply proposed method on fGn, the average spectra of first 18 IMFs
(a) Log_2(zero crossing) vs. IMF index   (b)) Log_2(energy) vs. IMF index

It's quite visible from both 2T-EMD and proposed method based hilbert spectrum representations that the later one has more distinguishable energy density which represents the presents of information with respect to the time and frequency domain. Fig. 7(c) also shows the reconstruction error signal that has maximum magnitude $8x10^{-16}$ which is acceptable in most applications. For separating individual audio sources from single channel mixture, independent subspace analysis based clustering method is applied on the outcome of proposed method where the parameters are instantaneous amplitude and instantaneous phase. An initial experimental results of ISA based clustering are spectrogram of separated sources (in this sample audio signal 'Flute' and 'Speech') which shown in Fig. 9. where x, y and z axis represents time, frequency bin index and energy distributions respectively.

(a)   (b)
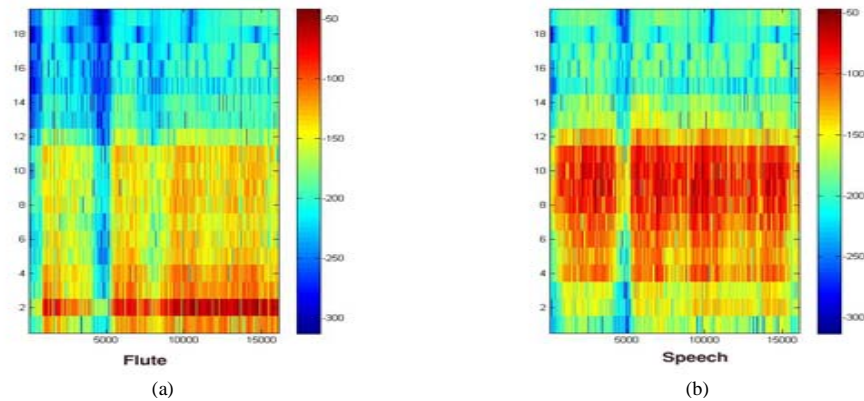
Fig. 9. Spectrogram of separated sources using ISA based clustering method from sample audio signal where (a) Flute and (b) Speech represents corresponding sources. The horizontal and vertical axis represents time slice and frequency bin index respectively

Conclusions

In this paper, a new method is introduced which is a combined approach of 2T-EMD and EMD methods, emphasizing on the improvement of pure 2T-EMD based hilbert spectrum. This proposed method can much more efficiently decompose an audio signal into sufficient number of IMFs which not only enhances hilbert spectrum representation

but also contributes several kind of audio signal processing work. This work reveals the nature of 2T-EMD method in stochastic situation which analyzes with respect to the fractional Gaussian noise. This enhancement of hilbert spectrum can now represents significant amount of information in both time and frequency domain with energy distribution which can also be used to separate the sources using different technique where ISA based clustering method is implemented as an example in this paper and an initial experimental results of separated sources are shown.

Acknowledgment

References

[1] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C.Yen, C.C. Tung and H.H. Liu, "The empirical mode decomposition and hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. Roy. Soc. London A, Vol. 454, pp. 903–995, 1998.

[2] Peel, M.C., G.E. Amirthanathan, G.G.S. Pegram, T.A. McMahon and F.H.S. Chiew "Issues with the application of empirical mode decomposition analysis", EGU 2008 - Session CL21, Peel et al

[3] Md. Ekramul Hamid, Md. Khademul Islam Molla, Md. Iqbal Aziz Khan, Takayoshi Nakai. "Speech enhancement using hilbert spectrum and wavelet packet based soft-thresholding", Science Journal of Circuits, Systems and Signal Processing. Vol. 1, No. 1, 2015, pp. 1-8. doi:10.11648/j.cssp.20150401.12

[4] Julien Fleureau, Jean-Claude Nunes, Amar Kachenoura, Laurent Albera, and Lotfi Senhadji, "Turning Tangent Empirical Mode Decomposition: a framework for mono- and multivariate signals", IEEE Transactions On Signal Processing, Vol. 59, No. 3, March 2011.

[5] G. Rilling, P. Flandrin, and P. Goncalves, "Empirical mode decomposition, fractional Gaussian noise and Hurst exponent estimation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2005, vol. 4, pp. 489–492.

[6] P. Flandrin, G. Rilling and P. Gonqalves, "Empirical mode decomposition as a filter bank", IEEE Signal Processing Letters, vol. 11, no. 2, (2004).

[7] Z. Wu and N.E. Huang, "A study of the characteristics of white noise using the Empirical Mode Decomposition method," submitted to Proc. Roy. Soc. London A, Dec. 2002.

[8] Md. Khademul Islam Molla, Md.Rabiul Islam, Toshihisa Tanaka, Tomasz M. Rutkowski,"Artifact suppression from EEG signals using data adaptive time domain filtering", Article In Neurocomputing 97:297-308 · November 2012

[9] B.B. Mandelbrot and J.W. van Ness, "Fractional Brownian motions, fractional noises and applications," SIAM Rev., Vol. 10, pp. 422–437,1968.

[10] G. Rilling, P. Flandrin, and P. Goncalves, "Empirical mode decomposition, fractional Gaussian noise and Hurst exponent estimation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2005, vol. 4, pp. 489–492.

[11] A.T. Wood and G. Chan, "Simulation of stationary processes in $[0,1]^d$," J. Comp. Graph. Stat., Vol. 3, pp. 409–432, 1994.

[12] Md. Ekramul Hamid, Md. Khademul Islam Molla, Xin Dang, Takayoshi Nakai, "Single channel speech enhancement using adaptive soft-thresholding with bivariate EMD", Hindawi Publishing Corporation, ISRN Signal Processing, Volume 2013,Article ID 724378, 9 pages.

[13] M. A Casey, A Westner, "Separation of mixed audio sources by independent subspace analysis" Proceedings of the International Computer Music Conference, Berli, August 2000