International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 4 (2015), pp. 73-88
http://dx.doi.org/10.14257/ijsip.2015.8.4.07

Spectral Analysis of Audio Signals with Noise Assisted Empirical Mode Decomposition

Poly Rani Ghosh[1], Keikichi Hirose[2] and Md. Khademul Islam Molla[2,3]

[1]*Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, Bangladesh*
[2]*Department of Information and Communication Eng., University of Tokyo, Tokyo, Japan*
[3]*Department of Computer Science and Eng., University of Rajshahi, Rajshahi, Bangladesh*
polyghosh@jkkniu.edu.bd, molla@gavo.t.u-tokyo.ac.jp

Abstract

A data adaptive approach to spectral analysis of audio signals is implemented in this paper. The audio signals are non-stationary as well as non-linear in nature and the traditional Fourier based spectral representation is not effective. The Hilbert spectral analysis implemented by noise assisted bivariate empirical mode decomposition (NA-BEMD) is introduced here as an efficient spectral representation scheme of audio signals. In BEMD, the fractional Gaussian noise (fGn) and analyzing speech signal are used as two separate variables. Both signals are decomposed together yielding a finite set of intrinsic mode functions (IMFs) for individual variables (signals). The use of fGn implements BEMD with dyadic filterbank characteristics. The instantaneous frequencies of individual IMFs are computed by applying Hilbert transform and then the time-frequency representation is achieved by arranging the energy with respect to time and frequency simultaneously. Such representation is called Hilbert spectrum (HS) which is analogous to spectrogram. The marginal HS derived from HS corresponds the total energy at each frequency component. The experimental results show that the Hilbert spectral analysis provides better representation of audio signal contents compared to the Fourier based approach.

Keywords: *Empirical mode decomposition, fractional Gaussian noise, Hilbert transform, spectral analysis, time-frequency representation*

1. Introduction

Spectral analysis of audio signals places them in the frequency domain to observe how the signals respond to all the various frequencies of a given bandwidth. Frequency analysis is important because it is crucial to know how audio analysis can be performed at certain frequencies [1]. The frequency domain representation of audio signals appears advantageous for two standpoints: (i) the natural frequency concept permits concise description of analyzing audio signals, (ii) the selection of frequency components suitable for specific application [2]. Presumably, features salient in frequency domain are important in production, perception, and consequently hold promise for other efficient analysis [3]. With the rapid growth of the applications of audio signals different types of spectral analysis/representation are required to apply on such system [4]. The application includes audio coding, synthesis, production, enhancement, localization etc. which are mostly performed in the spectral domain [3, 5]. The effectiveness of the mentioned applications performed in frequency domain depends on the efficiency of spectral representation of the analyzing audio signals. According to the accepted mathematical

ISSN: 2005-4254 IJSIP

approach, spectral analysis methods are all based on an a priori defined basis, which is a collection of linearly independent vectors so that the data can be rendered as a linear combination of the basis. To be of practical use, the expansion of the data in terms of this basis is established as convergent, complete, and unique. The traditional Fourier spectral analysis uses the sine and cosine functions as the basis, which satisfies all the aforementioned requirements. As a matter of fact, the word spectrum is almost used as a synonym for the Fourier spectrum [6]. The audio signals are almost invariably the result of non-stationary and nonlinear process. With such signal, the Fourier spectral analysis can only offer little in the way of a physically meaningful insight. There are various methods to circumvent this difficulty. For example, the short time Fourier transform (spectrogram), the wavelet analysis, and the Wigner-Ville spectrum are the most popular methods designed to accommodate the non-stationary nature of the data. These methods, however, are all Fourier-based or of Fourier type, that is, through convolutional computation with respect to a priori selected basis [7]. Hence, an effective data adaptive approach to frequency domain analysis is required for better fit with the non-stationary and non-linear audio signals.

In this paper, the empirical mode decomposition (EMD) [8] based Hilbert spectrum is used in spectral representation of audio signals. The key part of this approach is the EMD method with which any data set can be decomposed into a finite number of intrinsic mode functions (IMFs). An IMF is defined as any function having the same number of zero crossings and extrema and also having symmetric envelopes defined by the local maxima and minima. The recent development of EMD is focused on the use of ensemble EMD (EEMD) [9], complex EMD (CEMD) [10] and bivariate EMD (BEMD) [11]. The key advantage of the newly developed EMD methods is to achieve the accurate decomposition of the analyzing signal. The EEMD approach consists of sifting an ensemble of white noise-added signal and treats the mean as the final true result. The effect of the added white noise is to provide a uniform reference frame in the time-frequency space; therefore, the added noise collates the portion of the signal of comparable scale in one IMF. The traditional EMD is prone to mode-mixing and is designed for univariate data. A noise-assisted approach in conjunction with BEMD is implemented here for spectral representation of audio signals in order to produce localized frequency estimates at the accuracy level of instantaneous frequency. Such noise assisted BEMD (NA-BEMD) approach utilizes the dyadic filter bank property of the BEMD providing the solution of to the problem of standard EMD [12]. The obtained IMFs admit well-behaved Hilbert transformation. This decomposition method is adaptive, and, therefore, highly efficient. Since the decomposition is based on the local characteristic timescale of the data and computes instantaneous frequency through the Hilbert transform. It can reveal the intra-wave frequency modulations as functions of time and thus give sharp identifications of imbedded structures. The final presentation of the results is an energy-frequency-time as well as energy– frequency distribution, designated as the Hilbert spectrum and marginal Hilbert spectrum respectively. Both are mainly based on the EMD – a fully data adaptive decomposition technique. It is equally applicable to nonlinear and non-stationary processes and gives a physically meaningful interpretation of the data. The rest of the paper is organized as follows. The concept of different types of EMDs and the corresponding spectral representation are described in Section 2, the simulation results are illustrated in Section 3 and Section 4 includes some concluding remarks.

2. Data Adaptive Spectral Representation

It is required to interpret the underlying process through the analysis method of the data. To better understand the physical mechanisms hidden in data, the dual complication of non-stationarity and nonlinearity should be properly dealt with. A

more suitable approach to revealing nonlinearity and non-stationarity in data is not to let the analyzer impose irrelevant mathematical rules. The method of analysis should be adaptive to the nature of the data. For the methods that involve decomposition of data, the adaptive requirement calls for an adaptive basis which is to be based on and derived from the data [13]. Unfortunately, most currently available data decomposition methods have a priori basis (such as trigonometric functions in Fourier analysis), and they are not data adaptive. Once the basis is determined, the analysis is reduced to a convolution computation. This well-established paradigm is mathematically sound and rigorous. However, the ultimate goal for data analysis is not to find the mathematical properties of data; rather, it is to unearth the physical insights and implications hidden in the data. There is no a priori reason to believe that a basis arbitrarily selected is able to represent the variety of underlying physical processes [13]. Therefore, the results produced, though mathematically correct, might not be informative. The combination of the well-known Hilbert transform and the recently developed empirical mode decomposition (EMD) [8], termed as the Hilbert-Huang transform (HHT), indeed, represents such a paradigm of data analysis methodology. The HHT is designed specifically for analyzing nonlinear and non-stationary data. The key part of HHT is EMD and different types of EMDs are described in the following subsections.

2.1. Traditional EMD

The essence of EMD is to identify the oscillatory modes by their characteristic timescales in the data empirically and then decompose the data accordingly. The method is described in great detail in [8, 14]. As in [8], the time lapse between successive extrema is adopted as the definition of the timescale for a specific oscillatory mode. The decomposition method can simply use the envelopes defined by the local maxima and minima, separately. Once the extrema are identified, all the local maxima are connected by a signal reconstruction method, such as cubic spline to generate the upper envelope. The same procedure is repeated for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them. Their mean is designated as m_1, and the difference between the data $x(t)$ and m_1 is the first component, h_1, *i.e.*, $x(t) - m_1 = h_1$.

Representing a simple oscillatory motion, the IMF is the counterpart to the simple harmonic function. But the IMF is more general, it can have both amplitude and frequency modulations. From the construction of h_1 described previously, it should have been an IMF. In reality, however, overshoots and undershoots are common, which can also generate new extrema and shift or exaggerate the existing ones. Even if the fitting is perfect, a gentle hump on a slope can be amplified in further steps to become a local extremum, for, when we perform the operation $x(t) - m_1 = h_1$, we have effectively changed the local reference zero line in rectangular coordinate to m_1, which becomes a curvilinear coordinate system. After the first round of processing (which we term sifting, due to the nature of removing components of varying size), a hump could become a local maximum in h_1. Basically, the sifting process serves two purposes: to eliminate riding waves and to make the wave profiles more symmetric. Toward these ends, the sifting process has to be repeated more times. In the second sifting process, $h1$ is treated as the data, then $h_1 - m_{11} = h_{11}$.

We can repeat the sifting procedure k times, until h_{1k} is an IMF. It is then designated as the first IMF component from the data: $h_{1k} = c_1$. It is much more symmetric than h_1. As described previously, the process is indeed like sifting: to separate the finest local mode from the data first. The sifting process, however, should be applied with care, for carrying the process to an extreme could make the resulting IMF a pure frequency-modulated signal of constant amplitude. To guarantee that the IMF components retain enough physical sense of both amplitude and frequency modulations, we have to determine a

criterion for the sifting process to stop. This can be accomplished simply by limiting the repetition until three consecutive siftings all give the same numbers of zero crossings and extrema as discussed in [14]. In general, c_1 should contain the finest scale or the shortest period component of the signal. We can separate c_1 from the rest of the data by $x(t) - c_1 = r_1$. (4) Since the residue, r_1, still contains information of longer period components, it is now treated as the new data and subjected to the same sifting process as described previously. This procedure can be repeated to obtain all the subsequent r_k values, and the final result is $r_1 - c_2 = r_2$, to $r_{K-1} - c_K = r_K$.

The sifting process can be stopped by any of the following predetermined criteria: either when the component, c_K, or the residue, r_K, becomes so small that it is less than the predetermined value of substantial consequence, or when the residue, r_K, becomes a monotonic function from which no IMF can be extracted. Even for data with zero mean, the final residue still can be different from zero; thus for data with a trend, the final residue should be that trend. The complete decomposition can be represented as:

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) \qquad (1)$$

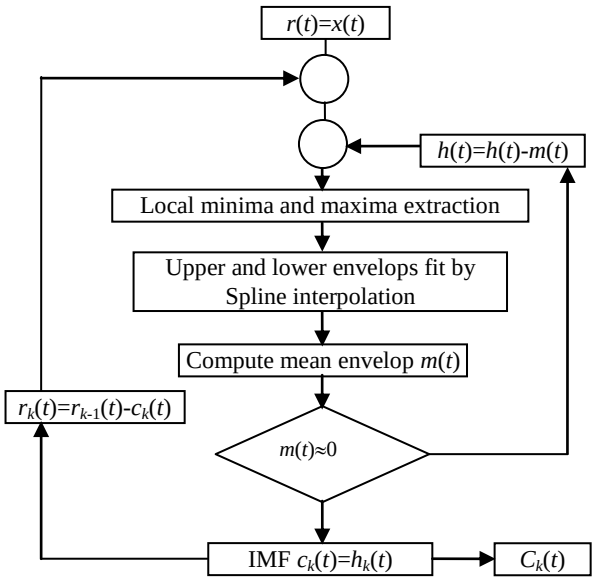

Figure 1. The Sifting Process to Compute EMD

The sifting process of EMD is illustrated in Figure 1. The decomposition of the data $x(t)$ is achieved into K empirical modes (IMFs), and a residue, r_K, which can be either the mean trend or a constant. The speech signal and its IMFs obtained by EMD are shown in Figure 2. As discussed here, to apply the EMD method, a mean or zero reference is not required. The zero reference for each component will be generated by the sifting process. Without the need of the zero reference, EMD eliminates the troublesome large DC term in data with nonzero mean values, an unexpected benefit. Thus, we have successfully defined a set of basic functions for this data. Any change in the data will result in a corresponding change of the basis function set. Therefore, this method is totally adaptive. The only basis that can represent the physics of a nonlinear and non-stationary process has to be adaptive [15]. This adaptiveness has to be so detailed that it will have to include intra-wave frequency modulation. Only by using an adaptive basis, can one fully accommodate the physics of changes in the processes.

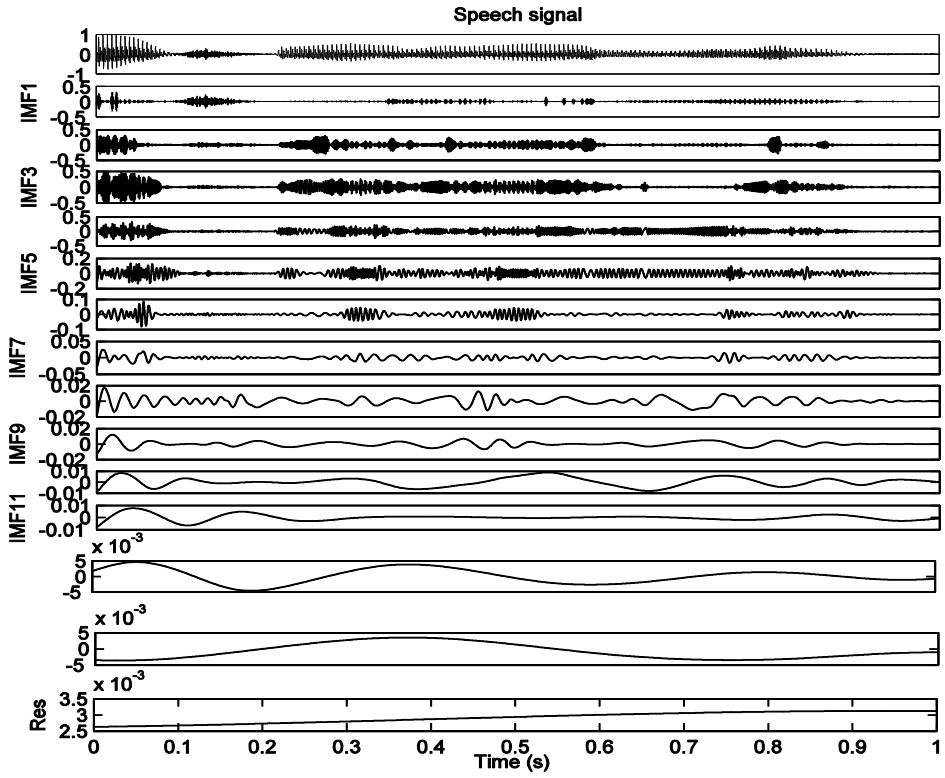

Figure 2. The Speech Signal and its Different IMFs Including Residual Signal Obtained by Traditional EMD

2.2. Instantaneous Frequency

The instantaneous frequency (IF) represents the signal's frequency at every time instance. IF is defined as the rate of change of the phase angle at the analysis time instant of the analytic version of the signal. Each IMF is a real valued signal. The analytic signal method [16] is used to calculate the instantaneous frequency of the IMFs. The analytic (complex) signal corresponding to k^{th} IMF $c_k(t)$ is defined as:

$$z_k(t) = c_k(t) + j\hbar[c_k(t)] = a_k(t)e^{j\theta_k(t)} \qquad (2)$$

where $\hbar[]$ is the Hilbert transform operator, $a_k(t)$ and $\theta_k(t)$ are instantaneous amplitude and phase respectively of the k^{th} IMF. The Hilbert transform provides a phase-shift of $\pm\pi/2$ to all frequency components, whilst leaving the magnitudes unchanged [17]. The Hilbert transform of any arbitrary time-series $s(t)$ can be defined as:

$$\hbar[s(t)] = \frac{1}{\pi}\sum_{\delta=-\infty,\delta\neq t}^{\infty}\frac{s(\delta)}{t-\delta} \qquad (3)$$

With the definition, $s(t)$ and $\hbar[s(t)]$ together form a complex conjugate yielding the analytic signal $s(t) + j\hbar[s(t)]$. The analytic signal is advantageous to determine the instantaneous quantities such as energy, phase and frequency. So, the corresponding instantaneous frequency of the k^{th} IMF can easily be derived as:

$$\omega_k(t) = \frac{\partial\tilde{\theta}_k(t)}{\partial t} \qquad (4)$$

where $\tilde{\theta}_k(t)$ represents the unwrapped version of the phase vector $\theta_k(t)$. Using Eq. (2) and (4), the analytic signal associated with each of the IMFs and thus the instantaneous frequency (IF) of each of them is calculated. The IF values of the IMFs illustrated in Figure 2 are shown in Figure 3. It is noticed that IF values of

individual IMFs are completely disjoint at any temporal position. The overall effect of IF of all IMFs is used in time-frequency (TF) representation of the time domain signal.

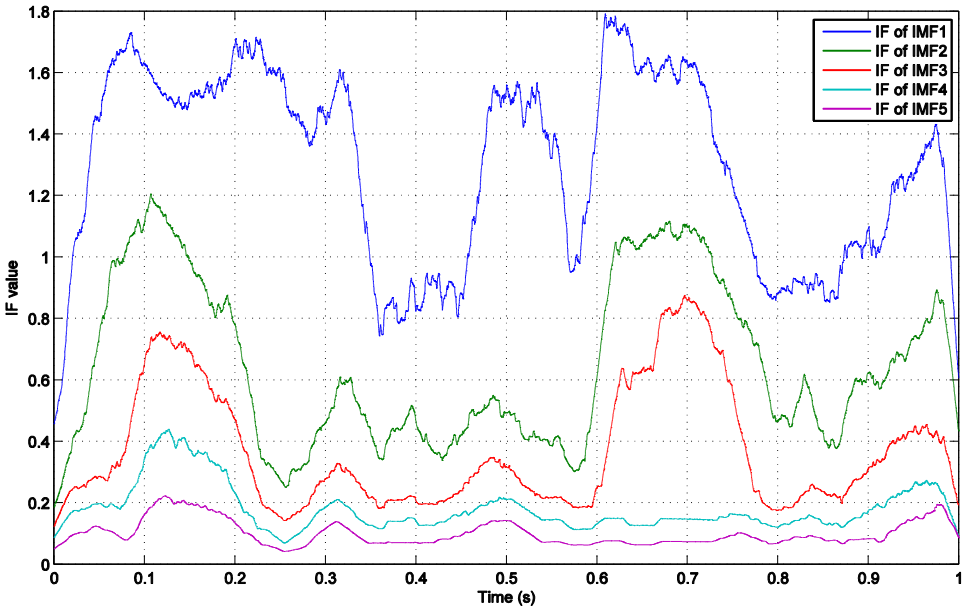

Figure 3. The Instantaneous Frequency (IF) Values of the First Five IMFs of Figure 2

2.3. Hilbert Spectrum

Hilbert Spectrum represents the distribution of the signal energy as a function of time and frequency. It is also designated as Hilbert amplitude spectrum $H(f,t)$ or simply Hilbert spectrum(HS). After performing the Hilbert transform on each IMF, the signal can be expressed as:

$$x(t) = \Re\left(\sum_{k=1}^{K} a_k(t) e^{j\int \omega_k(t)dt} \right) \qquad (5)$$

where $\Re(.)$ represents the real part of the complex number and only K IMFs are taken into consideration leaving the residue [8]. This expression enables to represent the amplitude and IF as a function of time. The instantaneous frequencies are first normalized to reflect the Nyquist properties of the frequency domain representation. The overall HS is expressed as the superposition of the individual IMFs' HSs defined as:

$$H(f,t) = \sum_{k=1}^{K} H^{(k)}(f,t) \qquad (6)$$

where $H^{(k)}(f,t)$ is the HS of the k^{th} IMF. Hence, each element of the overall HS is defined as the weighted sum of the instantaneous amplitudes of all IMFs at f^{th} frequency bin.

$$H(f,t) = \sum_{k=1}^{K} a_k(t) w_k^{(f)}(t) \qquad (7)$$

where the weight factor $w_k^{(f)}(t)$ takes 1 if the corresponding IF value falls within f^{th} band, otherwise is 0. After computing the elements over the frequency bins, H represents the instantaneous signal spectrum in TF space as a 2D table. There are various forms to represent the Hilbert spectrum. If amplitude squared is more desirable commonly to

represent energy density, then the squared values of amplitude can be substituted to produce the Hilbert energy spectrum just as well. When the visualization is more preferable than the further analytic processing of the Hilbert spectrum, it can be presented as the smoothed version using some image filtering. It is noted that the time resolution of H is equal to the sampling rate and the frequency resolution can be chosen up to Nyquist limit.

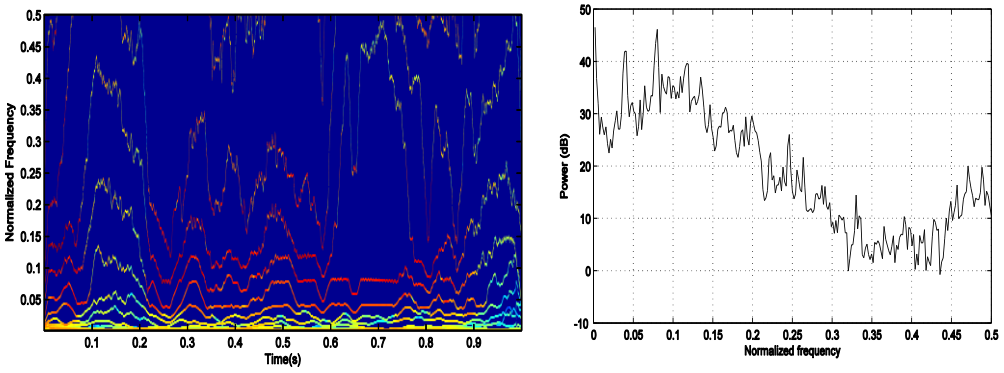

Figure 4. The Hilbert Spectrum (Left) and Marginal Hilbert Spectrum (Right) Corresponding to the Speech Signal Shown in Figure 1

The marginal spectrum defines a measure of total energy contribution from each frequency value. It represents the cumulated amplitude over the entire data length in a probabilistic sense. As we have already derived the Hilbert spectrum $H(f,t)$, the marginal spectrum $h(f)$ can be easily defined as:

$$h(f) = \int_t H(f,t)\,dt \qquad (8)$$

It is found that the marginal Hilbert spectra play a different interpretation rather than Fourier spectra [18]. The HS and the marginal HS corresponding to the speech signal shown in Figure 2 are illustrated in Figure 4. In the Fourier spectra, the existence of energy at a frequency, f, means a component of a sine or a cosine wave persisted through the time span of the data. The Fourier energy spectrum clearly represents a stack of harmonics. Whereas, the existence of energy in marginal Hilbert spectrum at the frequency, f, means only that, in the whole time span of the data, there is a higher likelihood for such a wave to have appeared locally.

2.4. Bivariate EMD (BEMD)

The univariate EMD is only suitable for univariate (real valued) signals. The complex empirical mode decomposition (CEMD) is an extension of the basic EMD suitable for dealing with complex signals [10]. The motivation to extend EMD is that a large number of signal processing applications have complex signals. In addition, this extension is applied on both the real and imaginary parts simultaneously because complex signals have a mutual dependence between the real and imaginary parts. Thus, if the decomposition is done separately, the mutual dependency will be lost. The bivariate empirical mode decomposition (BEMD) is more generalized extension of EMD. The main difference between BEMD and CEMD is that the latter uses the basic EMD to decompose complex signals, whereas BEMD adapts the rationale underlying the EMD to a bivariate framework [11, 19]. In BEMD two variables are decomposed simultaneously based on their rotating properties. The algorithm of BEMD, as proposed in [11], is as follows:

1) For $1< q < Q$,

a) Project $\hat{x}(t)$ on direction ϕ_q: $p_{\phi_q}(t) = \mathrm{Re}(e^{-j\phi_q}\hat{x}(t))$

b) Extract the maxima of $p_{\phi_q}(t) : (t_i^q, p_i^q)$

c) Interpolate the set of points $(t_i^q, e^{j\phi_q} p_i^q)$ to obtain the partial envelope curve in direction ϕ_q named $e_{\phi_q}(t)$

2) Compute the mean of all tangents:

$$e(t) = \frac{2}{Q}\sum\nolimits_q e_{\phi_q}(t) \tag{9}$$

3) Subtract the mean to obtain $\hat{c}(t) = \hat{x}(t) - e(t)$

4) Test if $\hat{c}(t)$ is an IMF:

- If yes, repeat the procedure from the step 1 on the residual signal.
- If not, replace $\hat{x}(t)$ with $\hat{c}(t)$ and repeat the procedure from step 1.

In bivariate EMD, $\hat{x}(t)$ is modeled as a complex variable $\hat{x}(t) = s(t) + j\eta(t)$; where $s(t)$ and $\eta(t)$ represents two observed real valued analyzing signals. The BEMD produces complex IMFs as well as residue. The real and imaginary part of any IMF represent the IMFs corresponding to the signals $s(t)$ and $\eta(t)$ respectively.

2.5. Noise assisted BEMD (NA-BEMD)

The ensemble EMD makes use of the dyadic filter bank property of EMD when applied to white Gaussian noise (wGn); subsequent averaging over the noise ensemble benefits from the so induced large number of extrema, and yields more localized inherent modes present in the data, in addition to the decomposition which is almost free from mode-mixing [11]. However, a consequence of adding noise directly to the data is that a trace of residual noise is likely to remain in the IMFs. The amplitude of this residuum depends on the number of realizations averaged (size of ensemble), thus, compromising the "completeness" of the retained signal.

To address the above issues, the noise-assisted BEMD (NA-BEMD) is employed here that has been originally designed for signals containing two data channels and has shown significant potential in not-stationary data analysis [20]. The NA-BEMD operates by first creating a signal consisting of one input data channel and adjacent independent realizations of fractional Gaussian noise (fGn) [21] in separate channel. The resulting bivariate signal, comprising data and noise channels, is processed using the BEMD method. The IMFs corresponding to the original data are reconstructed to yield the desired decomposition [11]. In this way, unlike EEMD, the physically disjointed input and noise subspaces within NA-BEMD prevents direct noise artifacts. It makes use of the dyadic filter bank structure of EMD for fGn for improved performance of the standard univariate EMD [20]. The fGn is a generalization of ordinary white noise. It is a versatile model of homogeneously spreading broadband noise without any dominant frequency band [22]. The statistical properties of fGn are entirely determined by its second-order structure, which depends solely on one single scalar parameter, the Hurst exponent (H) [21]. In discrete time, the fGn corresponds to a time series $[\eta_H(t), t = \ldots, -1, 0, 1, \ldots]$ indexed by a real-valued parameter $0 < H < 1$.

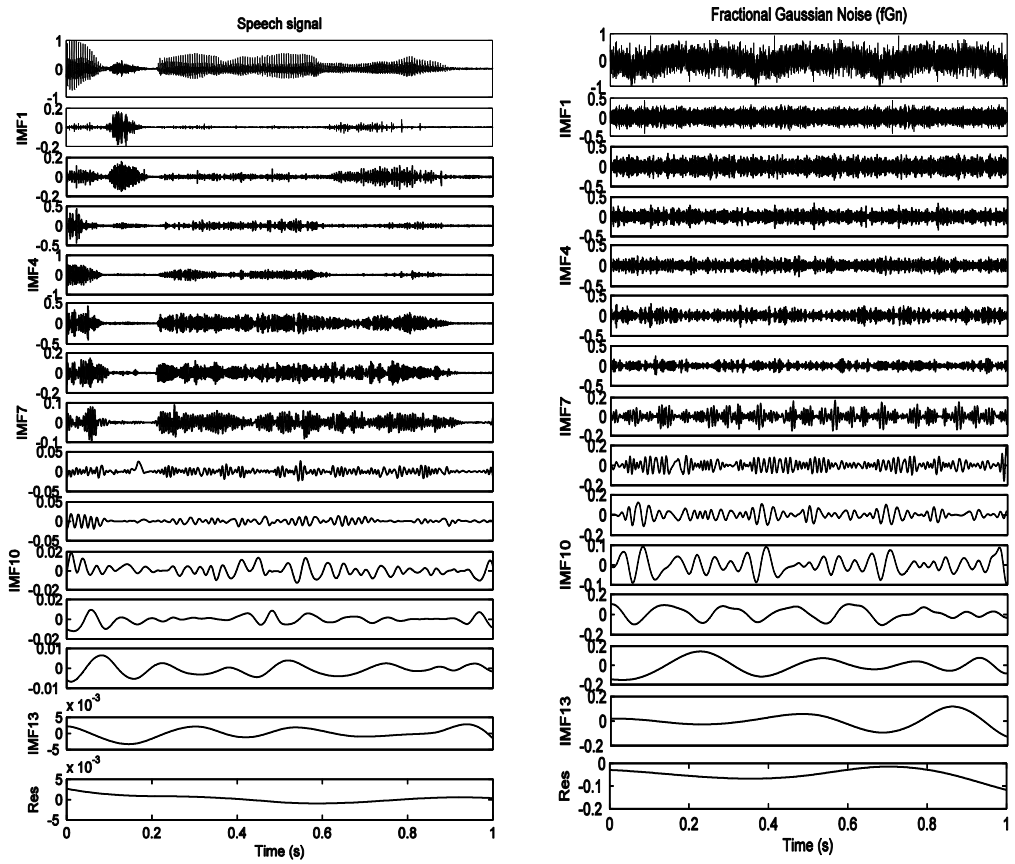

Figure 5. The IMFs of Speech Signal (Left Panel) and fGn (Right Panel) Obtained by NA-BEMD

The BEMD algorithm acts as a dyadic filter bank on the analyzing signal when applied together with fGn, exhibiting greatly enhanced alignment of the corresponding IMFs of signal across the same frequency range compared to EMD. The noise assisted BEMD further alleviates the mode mixing problem. Notice that in this way the noise is never mixed with the useful data channel, as it resides in a different subspace, and is used to enforce a filterbank structure, and thus alleviate the problem of mode mixing and provide much better definition of frequency bands inherent to the data. A set of IMFs corresponding to only the original input signal is kept by discarding the IMF subspace associated with the noise. To make decomposition with BEMD, the speech signal $s(t)$ is combined with fGn, $\eta(t)$, producing the complex signal $\hat{x}(t) = s(t) + j\eta(t)$. Both the variables (speech and fGn) are decomposed simultaneously without losing mutual dependency by using BEMD. After completion of BEMD, $\hat{x}(t)$ can be expressed as

$$\hat{x}(t) = \sum_{k=1}^{K} \hat{c}_k(t) + \hat{r}_K(t) \qquad (10)$$

where K is the total number of IMFs; complex valued $\hat{c}_k(t)$ and $\hat{r}_K(t)$ represent the k^{th} IMF and final residue, respectively. The real part $\alpha(t) = \text{Re}\langle\hat{c}(t)\rangle$ represents the IMFs of the speech signal $s(t)$ and the imaginary part $\beta(t) = \text{Im}\langle\hat{c}(t)\rangle$ corresponds to the IMFs of $\eta(t)$. Hence, the individual signals can be represented as

$$s(t) = \sum_{k=1}^{K} \alpha_k(t) + \text{Re}\langle\hat{r}_K(t)\rangle;\ \eta(t) = \sum_{k=1}^{K} \beta_k(t) + \text{Im}\langle\hat{r}_K(t)\rangle \qquad (11)$$

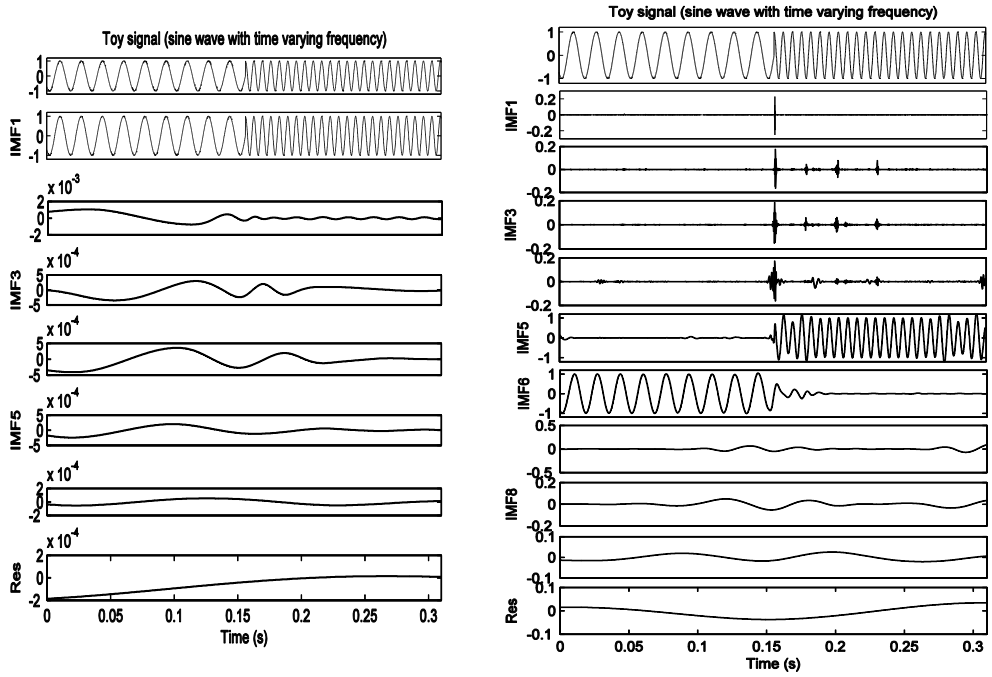

Figure 6. The IMFs of Toy Signal (Sine Wave with Time Varying Frequency) Obtained by Traditional EMD (Left Panel) and NA-BEMD (Right Panel). Higher Number of IMFs is Obtained by NA-BEMD. The 5th and 6th IMFs (Right) Represent the Individual Sinusoids

The lowest order IMF captures the highest frequency oscillation contained by the signals. The local frequency of any IMF is lower than that of just extracted before. The IMFs of speech and fGn obtained by noise assisted BEMD are shown in Figure 5. It is noticed that the number of its IMFs are higher than that of the traditional EMD yielding higher degree of frequency disjoint.

3. Simulation Results

We have obtained good results and new insights in spectral domain by applying the combination of the EMD and Hilbert spectral analysis methods to audio signals. The audio signals are non-stationary and non-linear in nature. Hence, the Hilbert spectrum based analysis is better fitted with the audio signals. To explore the efficiency of the noise assisted BEMD a toy signal – sine wave with time varying frequency is taken into consideration. The results of traditional EMD and NA-BEMD are presented in Figure 6. It is noticed that only one IMF is extracted with EMD to represent two frequency components of the toy signal. On the other hand, two separate IMFs are generated to represent the target frequency components. Higher number of IMFs is generated using NA-BEMD. Hence, the NA-BEMD based model is better to represent the frequency component of the signal. The Hilbert spectrum of the toy signal obtained by NA-BEMD and its spectrogram using short-time Fourier transform (STFT) are illustrated in Figure 7. It is noticed that the frequency components are finely localized in the HS with NA-BEMD, whereas, STFT includes a noticeable amount of cross-spectral energy. The marginal Hilbert spectrum (mHS) represents the contribution of energy at individual frequency component. The Fourier spectrum defines uniform harmonic components globally. It requires additional harmonic components to simulate non-stationary data that are non-uniform globally yielding the spread of energy for a wide range of frequency. The Figure 8 illustrates the Fourier spectrum together with mHS of the toy signal obtained by NA-

BEMD methods. It is observed that the Fourier spectrum includes the energy other than the actual frequency components of the signal.

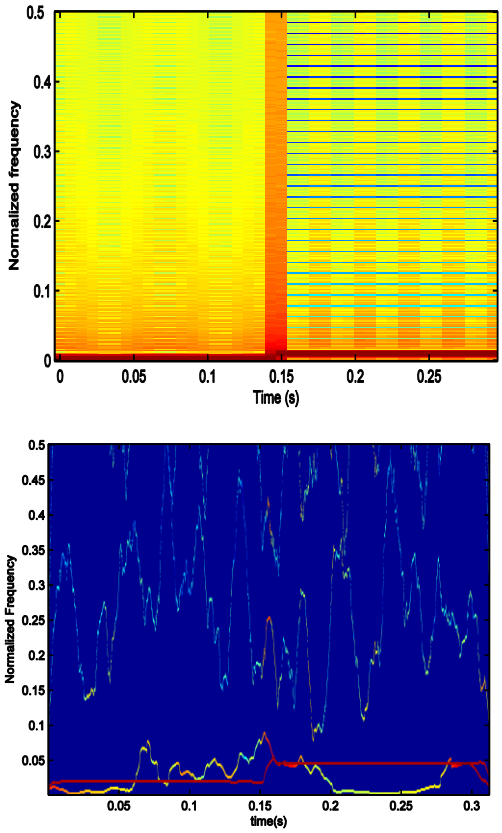

Figure 7. The Spectrogram (Left) and Hilbert Spectrum (Right) of the Toy Signal (Sine Wave with Time Varying Frequency) Obtained by NA-BEMD. The Spectrogram Includes Noticeable Amount of Cross-spectral Energy and Spurious Harmonics

The EMD on fGn acts as dyadic filter-bank [12, 15]. The dyadic property of any filter-bank structure refers that the bandwidth of any subband is the half of its just previous (high frequency) subband. When the speech signal is decomposed together with fGn using NA-BEMD, the overall decomposition acts like a dyadic filterbank. The Fourier log spectra of the IMFs of fGn and speech signal (as illustrated in Figure 5) are shown in Figure 9(a) and 9(b) respectively. It is observed that the spectra of fGn's IMF represent the dyadic characteristics. The IMFs' spectra of speech signal also represent the dyadic nature.

The time-frequency representation is an efficient way to observe the energy of the signal with respect to time and frequency simultaneously. The higher resolution of both time and frequency illustrates better represenation of the energy distribution. In STFT, it is not possible to extend the resolution of both time and frequency to the desired scale but an uncertainity.

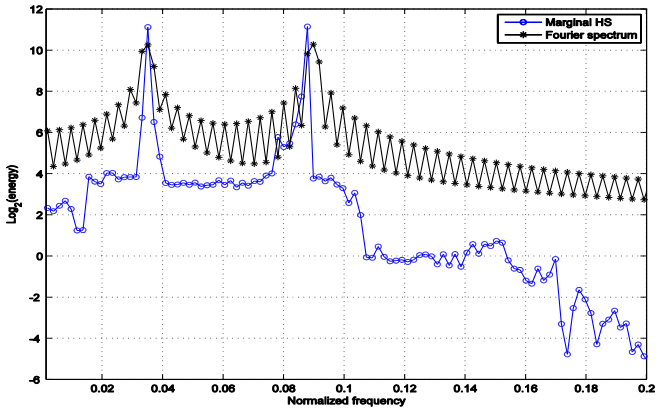

Figure 8. The Fourier Spectrum as Well as Marginal Hilbert Spectrum (mHS) of the Toy Signal (Sine Wave with Time Varying Frequency) Obtained by NA-BEMD. The Fourier Spectrum Represents the Energy as a Stack of Harmonics as Well as Higher Cross-spectral Energy

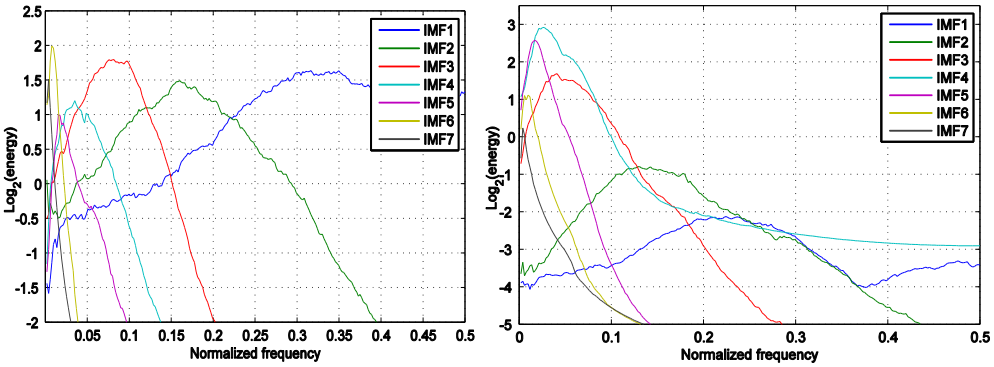

Figure 9. Fourier Log Spectrum of First 7 IMFs of fGn (Left) and Speech Signal (Right). The Energy Peak of Any fGn's IMF is at the Half of the Frequency of Just Prevous One which Illustrates the Characteristics of Dyadic Filterbank. The Spectra of Speech's IMFs Illustrae the Similar Property

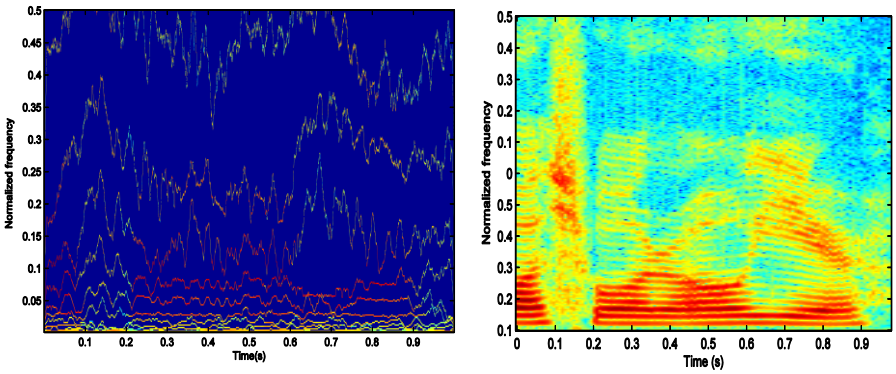

Figure 10. The Hilbert Spectrum (Left) and Spectrogram (Right) of Analyzing Speech Signal

On the other hand, there is no such limitation with HS based time-frequency representation. The speech signals used here are collected from TIMIT database, sampled with 16kHz frequency and 16-bits amplitude resolution. The 25ms Hamming window

with 15ms ovelapping is used with 512 point fast Fourier trnasform (FFT) to implement the STFT (spectrogram). The Hilbert spectrum (HS) obtained by NA-BEMD and spectrogram of the speech signal are illustrated in Figure 10. It is noticed that the energies are shaprly localized in both time and frequency scale in HS, whereas, the energies are not perfectly localized in time-frequency domain in spectrogram. The marginal HS obtained by NA-BEMD and Fourier spectrum of the speech signal are shown in Figure 11. The Fourier spectrum illustrates wide spread spectrum than that of the marginal HS.

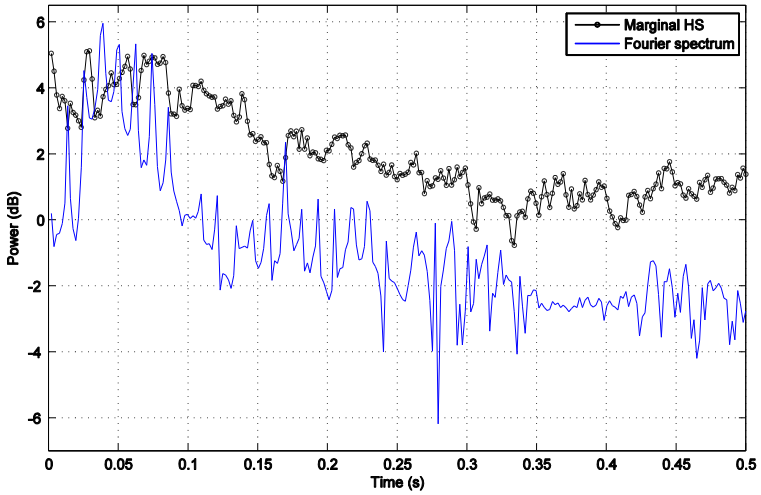

Figure 11. The Marginal Hilbert Spectrum and Fourier Spectrum of Spectrogram of the Speech Signal

Being a harmonic analysis technique, STFT spreads energy to the high frequency range as the harmonics. Conventionally, these harmonics are viewed as a matter of fact, but the HS reveals that Fourier expansion is a mathematical approximation to a nonlinear process, in which the true physical meaning is beyond the reach of Fourier-based analysis. The Fourier spectrum offers a nice mathematical presentation, yet lacks physical meaning. The energy of the signal is distributed over the predefined harmonics. Even with windowed Fourier transform, any change of signal characteristics shorter than the selected window will be obscured. Due to the overlapping of the window function the STFT also includes the cross-spectral energy between the adjacent time frames.

In HS, it is possible to present the spectral characteristics of the signal at each sampling point but more data points are required to compensate the end effects of IF calculation [23]. It does not include any noticeable amount of spectral cross-term between the time frames. In all fairness, it should be noted that the Fourier spectrum of a non-stationary signal does not make sense. That is why, to explore the event of cross-spectral term, the marginal Hilbert and STFT spectra of two sine waves are illustrated in Figure 8. The HS represents a sharper frequency definition than that shown by the STFT based spectrogram. The spectrogram illustrates the stacks of energy consisting of spurious harmonics and also includes a remarkable amount of cross-spectral terms. The marginal spectrum calculated from the HS represents the proper frequency localization of the tones with sharp energy bands at the specific frequency.

4. Conclusions

This study demonstrates the spectral analysis method of audio signals with fully data adaptive approach. The spectral representation is implemented by Hilbert spectral analysis based on EMD, a signal decomposition method suitable for non-linear and non-stationary process. The traditional Fourier based method is not effective for not stationary signal like

speech. The comparison between Hilbert spectral analysis and Fourier based spectral representation is presented in this paper. The EMD is implemented based on the direct extraction of the energy associated with various intrinsic time scales, the most important parameters of the system. Expressed in IMFs, they have well-behaved Hilbert transforms, from which the instantaneous frequencies are calculated. Thus, any event is localized on the time as well as the frequency axis. The decomposition is also viewed as an expansion of the data in terms of the IMFs. Then, these IMFs, based on and derived from the data, serve as the basis of that expansion which can be linear or nonlinear as dictated by the data, and it is complete and almost orthogonal. Most important of all, it is adaptive to the analyzing signal. The local energy and the instantaneous frequency derived from the IMFs through the Hilbert transform gives us a full energy-frequency-time distribution of the data. Such a representation is designated as the Hilbert spectrum (HS); it would be ideal for nonlinear and non-stationary data analysis. The marginal HS is analogous to Fourier spectrum in representing frequency characteristics of the signal. The Fourier spectrum spreads energy over a wide range of frequencies, whereas, the energy in mHS is sharply confined to the respective frequency component. The spectral analysis for multichannel audio signals with multivariate NA-EMD based approach is considered for future extension of this study.

References

[1] S. Lee, J. Kim and I. Lee, "Speech/audio signal classification using spectral flux pattern recognition", IEEE Workshop on Signal Processing Systems, **(2012)**.
[2] G. L. Rosen and J. D. Johnston, "Informed spectral analysis for isolated audio source parameters estimation", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, **(2011)**.
[3] J. Herre, E. Allamanche and O. Hellmuth, "Robust matching of audio signals using spectral flatness features," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, **(2001)**.
[4] S. Pfeiffer, S. Fischer and W. Effelsberg, "Automatic audio content analysis", Proceedings of the fourth ACM international conference on Multimedia, **(1997)**.
[5] C. Verron, P.-A. Gauthier, J. Langlois and C. Guastavino, "Spectral and Spatial Multichannel Analysis/Synthesis of Interior Aircraft Sounds", IEEE Transaction on Audio, Speech and Language Processing, vol. 21, no. 7, **(2013)**.
[6] N. E. Huang, C. C. Chern, K. Huang, L. W. Salvino, S. R. Long and K. L. Fan, "A New Spectral Representation of Earthquake Data: Hilbert Spectral Analysis of Station TCU129, Chi-Chi, Taiwan, 21 September 1999", Bulletin of the Seismological Society of America, vol. 91, no. 5, **(2001)**, pp. 1310–1338.
[7] M. B. Priestley, "Nonlinear and Non-stationary Time Series Analysis", Academic Press, London, **(1988)**.
[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N-C. Yen, C. C. Tung and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", Proc. R. Soc. London Series A, vol. 454, **(1998)**, pp. 903–995.
[9] Z. Wu and N. E. Huang, "Ensemble emipirical mode decomposition: A noise-assisted data analysis method", Adv. Adapt. Data Analysis, vol. 1, Issue 1, **(2009)**.
[10] T. Tanaka and D. P. Mandic, "Complex empirical mode decomposition", IEEE Signal Processing Letters, vol. 14, no. 2, **(2007)**, pp. 101-104.
[11] G. Rilling, P. Flandrin, P. Goncalves and J. M. Lilly, "Bivariate empirical mode decomposition," IEEE Signal Processing Letters, vol. 14, no. 12, **(2007)**, pp. 936–939.
[12] P. Flandrin, G. Rilling and P. Goncalves, "Empirical mode decomposition as a filter bank," IEEE Signal Processing Letters, vol. 11, no. 2, **(2004)**, pp. 112-114.
[13] N. E. Huang and Z. Wu, "A review on Hilbert-Huang transform: method and its applications to geophysical studies", Reviews of Geophysics, vol. 46, **(2008)**.
[14] N. E. Huang, Z. Shen and S. R. Long, "A new view of nonlinear water waves: the Hilbert Spectrum", Annual Review of Fluid Mechanics, vol. 31, **(1999)**, pp. 417–457.
[15] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method", Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 460, no. 2046, **(2004)**, pp. 1597-1611.
[16] M. K. I Molla, K. Hirose and N. Minematsu, "Separation of mixed audio signals by decomposing Hilbert spectrum with modified EMD", IEICE Trans. Fund., vol. E89-A, no. 3, **(2006)**, pp. 727-734.
[17] M. Cooke, "Modeling auditory processing and organization", Cambridge University Press, **(1993)**.
[18] N. E. Huang, *et al.*, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis", Proc. R. Soc. Lond. A, vol. 459, **(2003)**, pp. 2317-2345.

[19] M. U. Altaf, T. Gautama, T. Tanaka and D. P. Mandic, "Rotation invariant complex empirical mode decomposition", Proc. of IEEE Int. Conf on Acoust. Speech and Signal Proc. (ICASSP), vol. III, **(2007)**, pp. 1009-1012.
[20] N. U. Rehman, C. Park, N. E. Huang and D. P. Mandic, "EMD via MEMD: Multivariate noise-aided computation of standard EMD", Adv. Adapt. Data Analysis, vol. 5, no. 2, **(2013)**, pp. 1-25.
[21] H. Qian, "Fractional Brownian Motion and Fractional Gaussian Noise", Lecture Notes in Physics, vol. 621, **(2003)**, pp: 22-33.
[22] D. Koutsoyiannis, "The Hurst phenomenon and fractional Gaussian noise made easy", Hydrological Sciences Journal, vol. 47, no. 4, **(2002)**, pp. 573-596.
[23] M. C. Ivan and R. G. Baraniuk, "Empirical mode decomposition based frequency attributes", Proc. of the 69th SEG Meeting, Texas, USA, **(1999)**.

Authors

Poly Rani Ghosh, obtained B.Sc. and M.Sc. degrees in Computer Science and Engineering from the University of Rajshahi, Rajshahi, Bangladesh in 2009 and 2011 respectively. Then she was working as a guest researcher in Signal Processing and Computation Neuroscience (SiPCoN) Laboratory in the same department. She served as a Lecturer in the Department of Computer Science and Engineering, University of Development Alternative, Dhaka, Bangladesh from Sep 2013 to Mar 2014. Presently, she is working as a Lecturer in the Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, Bangladesh. Her research interest includes acoustic signal processing, time series analysis and brain computer interfacing (BCI).

Keikichi Hirose, received his B.E. degree in electrical engineering in 1972, and his Ph.D. degree in electronic engineering in 1977, respectively, from the University of Tokyo, Tokyo, Japan. In 1977, he joined the University of Tokyo as a Lecturer in the Department of Electrical Engineering, and, in 1994, became a Professor in the Dept. of Electronic Engineering. From 1996, he was a Professor at the Graduate School of Engineering, Department of Information and Com-munication Engineering, the University of Tokyo. On April 1, 1999, he moved to the University's Graduate School of Frontier Sciences (Department of Frontier Informatics), and again moved to Graduate School of Information Science and Technology (Department of Information and Communication Engineering) on October 1, 2004. From March 1987 to January 1988, he was a Visiting Scientist of the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, U.S.A. His research interests cover widely spoken language information processing. He led a project "Realization of advanced spoken language information processing from prosodic features," Scientific Research on Priority Areas, Grant in Aid on Scientific Research, Ministry of Education, Culture, Sports, Science and Technology, Japanese Government. He is a member of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, the International Speech Communication Association, the Institute of Electronics, Information and Communication Engineers (Fellow), the Acoustical Society of

Japan, and other professional organizations. In 2007, 2012, and 2013, he received the Best Paper Awards from the Research Institute of Signal Processing, Japan (RISP).

Khademul Islam Molla, received his B.Sc. and M.Sc. degrees in electronics and computer science from Shahjalal University of Science and Technology, Sylhet, Bangladesh in 1995 and 1997 respectively. He joined at the same department as a Lecturer in 1997. He obtained his Ph.D. degree from the Department of Frontier Informatics under the Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan in 2006. He was working as Lecturer and Assistant Professor in the Department of Computer Science and Engineering of the University of Rajshahi, Bangladesh up to August, 2006. After completing his Ph.D., he joined in the same department as Associate Professor in August 2006, and he has been a Professor since May 2012. From September 2006 to September 2008, he was working as JSPS Postdoctoral Research Fellow in the Department of Information and Communication Engineering, The University of Tokyo, Tokyo, Japan. He was a research fellow at the University of Alberta, Canada, from Nov. 2010 to Oct. 2011. He visited several universities in Japan as guest researcher. Presently, he is a visiting scientist at the University of Tokyo, Japan. His research interests include audio signal processing, blind source separation, brain computer interface (BCI), biomedical signal and image processing. He is a member of the Institute of Electrical and Electronics Engineers (IEEE). In 2007, he received the Best Paper Award from the Research Institute of Signal Processing, Japan (RISP).