Research Article

Single Channel Speech Enhancement Using Adaptive Soft-Thresholding with Bivariate EMD

Md. Ekramul Hamid,[1,2] Md. Khademul Islam Molla,[2,3] Xin Dang,[1] and Takayoshi Nakai[1]

[1] *Department of Information Science and Technology, Shizuoka University, Hamamatsu-shi 432-8561, Japan*

[2] *Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh*

[3] *Department of Information and Communication Engineering, University of Tokyo, Tokyo 113-0033, Japan*

Correspondence should be addressed to Md. Ekramul Hamid; ekram_hamid@yahoo.com

This paper presents a novel data adaptive thresholding approach to single channel speech enhancement. The noisy speech signal and fractional Gaussian noise (fGn) are combined to produce the complex signal. The fGn is generated using the noise variance roughly estimated from the noisy speech signal. Bivariate empirical mode decomposition (bEMD) is employed to decompose the complex signal into a finite number of complex-valued intrinsic mode functions (IMFs). The real and imaginary parts of the IMFs represent the IMFs of observed speech and fGn, respectively. Each IMF is divided into short time frames for local processing. The variance of IMF of fGn calculated within a frame is used as the reference term to classify corresponding noisy speech frame into noise and signal dominant frames. Only the noise dominant frames are soft-thresholded to reduce the noise effects. Then, all the frames as well as IMFs of speech are combined, yielding the enhanced speech signal. The experimental results show the improved performance of the proposed algorithm compared to the recently reported methods.

1. Introduction

The research on speech enhancement is motivated by the rapidly growing market of speech communication applications, such as teleconferencing, hands-free telephony, hearing-aids, and speech recognition. In hands-free communication systems, the microphone(s) is typically placed at a certain distance from the speaker. In adverse acoustic environment, various noise sources make the speech signal corrupted. Although, the human auditory system is remarkably robust in most adverse situations, noise effects heavily affect the performance of automatic speech recognition (ASR) systems. The performance of an ASR system trained in one specific environment will drop considerably when used in another acoustic environment [1].

Several approaches have already been proposed to improve the speech enhancement results. Although the microphone array based approach exhibits better results, at the same time speech processing research community is trying to reduce the number of microphones (channels). The spectral subtraction is one of the early methods to reduce the noise effects from the observed speech signals. In this method, the noise reduction is achieved by appropriate adjustment of the set of spectral magnitudes [2]. Its basic requirement is the noise spectrum which is determined from the nonspeech segments [3]. In such single channel speech enhancement system, the residual noise is a usual issue. It decreases the speech intelligibility and hence further processing is required to reduce the residual noise. The subband approach to single channel speech enhancement is another potential method. Fourier transform and wavelet transform are dominating methods widely used in subband-based speech enhancement techniques. But the Fourier transformation is not suitable to analyze nonstationary signals like speech. There are several approaches of using wavelet transformation in subband decomposition. The decomposition results are varied with the different parameters, for example, the basis wavelet, the number of decomposition

levels, and so forth. Moreover, the selection of parameters also depends on the analyzing data. Therefore, a data adaptive tool for analyzing nonstationary and nonlinear signal is highly desirable [4].

In the previous study [5], empirical-mode-decomposition-(EMD-) based data adaptive thresholding algorithm is introduced. EMD is developed by Huang et al. [6] to decompose any nonstationary signal into a finite set of bases called intrinsic mode functions (IMFs). Instead of the speech signal, the variance of each IMF is used to determine the adaptive threshold, and hence better performance is achieved in [5]. Its main drawback is to find the speechless part to determine the noise variance. The performance of this method depends on the efficiency of voice activity detection (VAD), and hence it is not convenient to implement for practical applications. In this study, bivariate EMD (bEMD), the generalized extension of traditional EMD [7, 8], is employed to resolve the mentioned problem. The fractional Gaussian noise (fGn) has interesting characteristics with EMD [9, 10]. The EMD on fGn acts as dyadic filter banks [10]. The energies of the IMFs decrease almost linearly with increasing their order. It implies that the higher frequency IMFs contain more energies than that of the lower frequencies [11, 12]. The analyzing speech signal and fGn are decomposed together with bEMD. The noise variance is determined from the individual IMF of fGn which is used here as the reference signal.

The BEMD is applied to the complex signal which is composed of noisy speech and fGn as real and imaginary components, respectively. Each of the obtained IMF is divided into frames. The energy of speech frame is compared with that of the fGn to classify the frame as noise or speech dominant. Note that the real and imaginary parts of any complex-valued IMF correspond with the IMF of speech and noise, respectively. The soft-thresholding is applied to only the noise dominant speech frames with optimal adaptive threshold [5]. Then, the real components (i.e., IMFs of speech) of the processed IMFs are summed together to reconstruct the enhanced speech.

This paper is organized as follows: the application of bEMD on speech and noise signals are described in Section 2, the noise variance estimation process is explained in Section 3, the proposed speech enhancement method using bEMD is described in Section 4, experimental results are illustrated in Section 5, and finally, Section 6 contains some concluding remarks.

2. BEMD of Speech and Reference Signals

The univariate EMD (uEMD) decomposes any signal into a finite set of basis waveforms modulated in both amplitude and frequency (AM-FM). The idea behind is that a signal with fast oscillations superimposed on slow oscillations. The EMD is designed to define a local low frequency component as the local trend, supporting a local high frequency component as a zero-mean oscillation. The principle of the uEMD technique is to decompose a signal into a sum of the band-limited functions termed as 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 different 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.

The first condition is similar to the narrowband requirement for a stationary 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. There exist many approaches of computing uEMD [6, 9, 10].

In order to handle bivariate time series, the uEMD is extended to complex-valued EMD which is called bivariate EMD [7]. The main difference between the bEMD and the complex-EMD [8] is that the latter uses uEMD to decompose real and imaginary parts of complex signals, whereas the bEMD adapts the rationale underlying the EMD to a bivariate framework. In bEMD, two variables are decomposed simultaneously (without losing mutual dependency) based on their rotating properties. The bEMD algorithm [7] is summarized as follows:

(1) Let the complex signal is $x(t)$ with time index t.

(2) For $1 < m < M$,

(a) projections of M signal in direction of φ_m is given by $\{p\varphi_m\}_{m=1}^M$, so project the complex signal $x(t)$ by using a unit complex number $e^{-j\varphi_m}$ in the same direction is given by

$$p\varphi_m(t) = R\left(e^{-j\varphi_m}x(t)\right), \quad m = 1, 2, \ldots, M, \qquad (1)$$

where $R(\cdot)$ is the real part of the complex signal;

(b) find the locations $\{t_j^m\}_{m=1}^M$ corresponding to the maxima of $\{p\varphi_m(t)\}_{m=1}^M$;

(c) interpolate between the maxima points $[t_j^m, x(t_j^m)]$ to obtain the partial envelope curve in direction φ_m named $\{e_{\varphi_m}(t)\}_{m=1}^M$.

(3) Compute the mean of all tangents: $e(t) = 2/M \sum_m e_{\varphi_m}(t)$.

(4) Subtract the mean from input signal to obtain $c(t) = x(t) - e(t)$.

(5) Test if $c(t)$ is an IMF:

(a) if yes, repeat procedure from the Step (1) on the residual signal,

(b) if no, replace $x(t)$ with $c(t)$ and repeat the procedure from Step (1).

The BEMD of a signal $x(t)$ can be expressed as

$$x(t) = \sum_{k=1}^{K} c_k(t) + \varepsilon_K(t), \tag{2}$$

where K is total number of IMFs, $c_k(t)$ is the kth IMF, and $\varepsilon_K(t)$ is the final residue. In bivariate EMD, the input signal $x(t)$ is represented as $x(t) = s(t) + j\eta(t)$ [13], where $j = \sqrt{-1}$, $\eta(t)$ is the fGn, and $s(t)$ is the observed noisy speech. The additive noise $\rho(t)$ is combined with the clean speech $\tilde{s}(t)$, yielding $s(t)$ defined as $s(t) = \tilde{s}(t) + \rho(t)$. It is noted that the real and imaginary parts of any IMF $c_k(t)$ are denoted by $\delta_k(t) = \text{Re}\langle c_k(t)\rangle$ and $\gamma_k(t) = \text{Im}\langle c_k(t)\rangle$ representing the IMF of speech and fGn, respectively.

Fractional Gaussian noise (fGn) is a generalization of ordinary white noise. It is a versatile model of homogeneously spreading broadband noise without any dominant frequency band. Consequently, the statistical properties of (fGn) are entirely determined by its second-order structure, which depends solely on one single scalar parameter and the Hurst exponent (H). In discrete time, the (fGn) corresponds to a time series $[\eta_H(t), t = \dots, -1, 0, 1, \dots]$ indexed by a real-valued parameter $0 < H < 1$. The energies of the subbands (IMFs) of (fGn) obtained by uEMD behave almost linearly as a function of the IMF index [10].

The decomposition produces two separate sets of IMFs corresponding to individual signals as shown in Figure 1. The Fourier spectrum of the IMFs of fGn illustrated in Figure 2(a); exhibits the property of dyadic filter banks [9, 10], that is, the center frequency of any IMF is half of that of the IMF just previously extracted. The energies of the IMFs are decreasing with the increase of their order as shown in Figure 2(b). It implies that that the lower order (higher frequency) IMFs contribute higher energies in fGn. It is agreed with the assumption that speech signal contains more noise at higher frequency, and hence it is justified to use the fGn as the reference signal to determine the noise variance. Moreover, more IMFs are generated when fGn is combined with speech to apply bEMD. The amplitude of the fGn is adjusted according to the noise variance of the observed speech signal.

3. Estimation of Noise Variance

The effective estimation of noise variance of the noisy speech plays a vital role to the performance of the proposed bEMD-based speech enhancement algorithm. The speech signal is considered as a smoothly varying signal with additive Gaussian noise of zero mean, and both are uncorrelated with each other. It can be transformed through a low-order polynomial form as

$$f(s) = \sum_{n=0}^{N} a_n s^n, \tag{3}$$

where N is the degree of polynomial. The purpose of the transformation is to find out the variance of a random

variable $y = f(s)$ resulting from the propagation of the noisy speech signal s through the nonlinear polynomial function. The Taylor series expansion is used to obtain exact expression of variance of the random variable distribution [14]. The low-order polynomial terms are suppressed using filter based on a finite difference expression. So, the initial distribution can be written as

$$s = \bar{s} + z, \tag{4}$$

where $\bar{s}$ is the expectation value of s and z is a zero mean random variable. Then, the Taylor expansion for y is as

$$y = f(\bar{s}) + \sum_{n=1}^{N} \frac{(z)^n}{n!} \frac{d^n f}{ds^n}\bigg|_{s=\bar{s}}, \tag{5a}$$

and we can determine the first-order moment

$$\bar{y} = E(y) = f(\bar{s}) + \sum_{n=2}^{N} \frac{\vartheta_n}{n!} \frac{d^n f}{ds^n}\bigg|_{s=\bar{s}}, \tag{5b}$$

where ϑ_n denotes the nth-order moment of the random variable z. The variance of a random variable is its second central moment, the expected value of the squared deviation from the mean [15]. Considering again the Taylor series expansion, the noise variance can be computed as follows:

$$\begin{aligned} \sigma_y^2 &= E\left[(y - \bar{y})^2\right] \\ &= E\left[\left(f(\bar{s}) + \sum_{n=1}^{N} \frac{(z)^n}{n!} \frac{d^n f}{ds^n}\bigg|_{s=\bar{s}} \right. \right. \\ &\quad \left. \left. - f(\bar{s}) - \sum_{n=2}^{N} \frac{\vartheta_n}{n!} \frac{d^n f}{ds^n}\bigg|_{s=\bar{s}} \right)^2\right]. \end{aligned} \tag{6}$$

Thus, the obtained overall noise variance is used to adjust the energy of the fGn prior to the decomposition. Then, the derived fGn is combined with the noisy speech to form the complex signal to be decomposed by bEMD.

4. Speech Enhancement Method

In a frame by frame basis, the noise variance of $\delta_k(t)$ is compared with the variance of $\gamma_k(t)$ to classify the frames signal and noise dominants. The variance of fGn is used here as the reference noise variance. The thresholding is performed only to the noise dominant frames. This process completely overcomes the limitation of computing noise variance from silence part as required in the case of traditional EMD-based approach [16].

4.1. Frame Classification. It is required to set a threshold to classify the frames into signal and noise dominants. The boundary is set to the case where the noise and the observed speech variances within a frame are equal. In case of independence of speech and noise, the covariance between
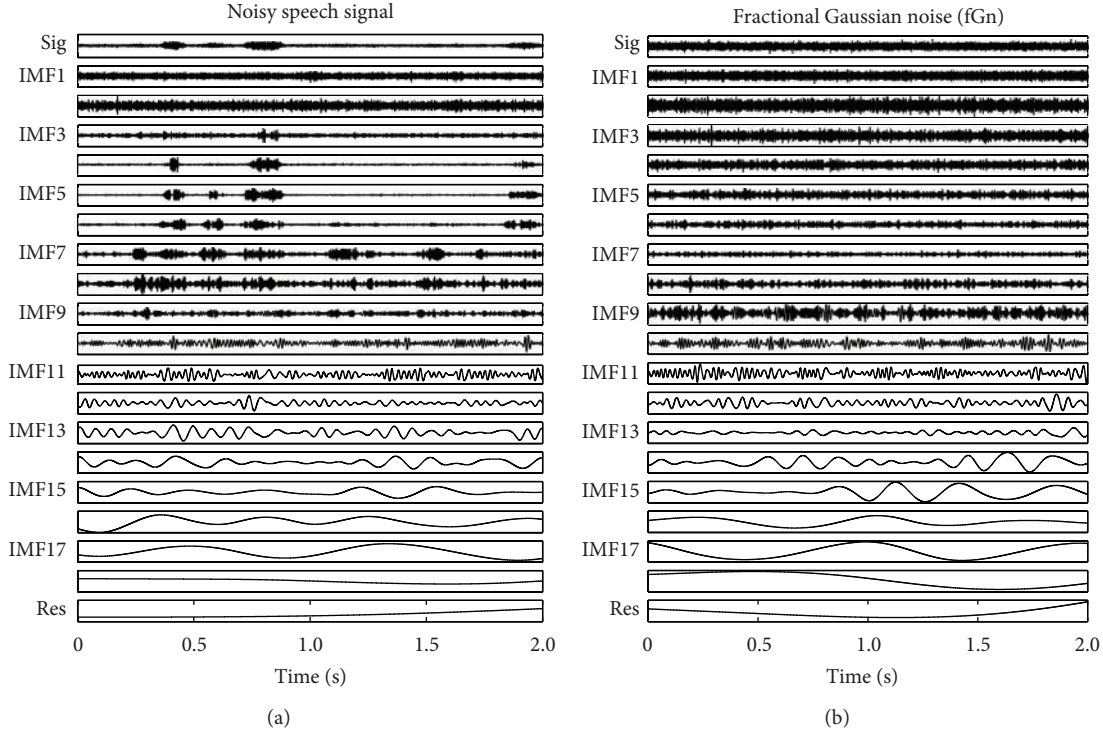
Figure 1: IMFs of noisy (0 dB SNR) speech (a) and fGn (b) obtained by applying bEMD.

the two will be zero [4]. Hence, generally for any frame, we can write

$$\sigma_s^2 = \sigma_{\bar{s}}^2 + \sigma_\rho^2, \tag{7}$$

where σ_s^2 indicates the variance of a frame of $\delta_k(t)$. For equal noise and signal power, we get

$$\sigma_s^2 = \sigma_{\bar{s}}^2 + \sigma_\rho^2 \approx 2\sigma_\rho^2 \approx 2\sigma_\eta^2, \tag{8}$$

where σ_s^2 indicates the frame variance of $\delta_k(t)$ and σ_η^2 indicates the frame variance of $\gamma_k(t)$ [4]. Therefore, in case of equal noise and signal power, with the assumption of independency, the variance of a frame is equal twice the noise variance. The comparisons of variances as a function of frame index for different IMFs are illustrated in Figure 3. The variance of signal dominant frame is always higher than that of the noise dominant. Each IMF is divided into frames (4 ms) to perform local processing in time domain. No frequency analysis has been performed with this short data rate frames, so there is no chance for end effect. Each frame of $\delta_k(t)$ is processed using the reference noise variance estimated from $\gamma_k(t)$. The soft-thresholding is applied on each noise dominant frame of $\delta_k(t)$. The classification condition of rth frame of kth real-IMF is defined as

$$\varphi_{k(r)} \geq 2\sigma_{\eta k(r)}^2, \tag{9}$$

where $\sigma_{\eta k(r)}^2$ is the variance of the rth frame of $\gamma_k(t)$ and $\varphi_{k(r)}$ is the average power of frame r of length Q of $\delta_k(t)$, and it is calculated as

$$\varphi_{k(r)} = \frac{1}{Q} \sum_{q=1}^{Q} \left| \delta_{k(r)}(q) \right|^2. \tag{10}$$

The proposed adaptive thresholding technique provides an effective boundary for the frame classification. If the condition given in (9) is satisfied, the frame is classified as speech dominant, otherwise noise dominant. It is noted that each IMF is a signal with zero local mean, and hence its average energy directly corresponds to the variance. Hence, the use of the comparison used in (9) is reasonable.

4.2. Adaptive Soft-Thresholding. It is more efficient to reduce the noise components from each IMF adaptively by using a frame-based soft-thresholding strategy [16]. Soft-thresholding strategy proposed in [17] is a powerful technique of speech enhancement for a wide range of input SNRs. It thresholded only the noise dominant frames and kept remain the same in case of the signal dominant frames. The soft-thresholding is carried out on each noise dominant frame of each $\delta_k(t)$ adaptively. After properly suppression of noise using soft-thresholding, all the real-IMFs are summed
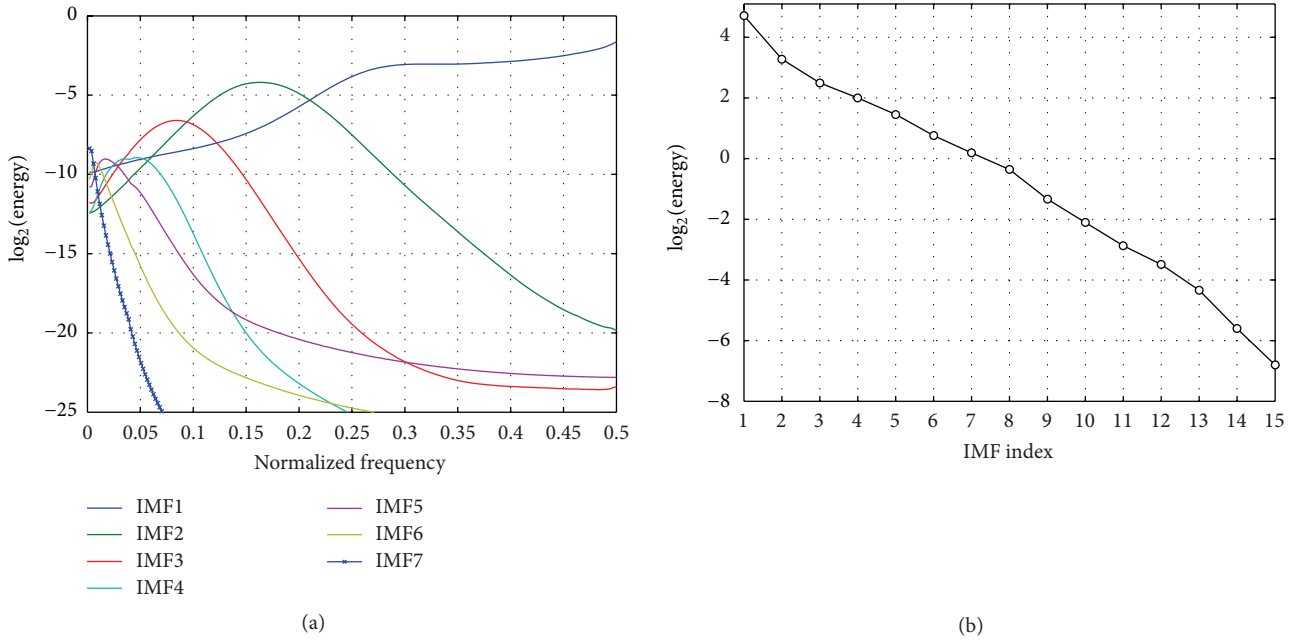
(a)

(b)

FIGURE 2: (a) Fourier log spectrum of IMFs of fGn. The energy peak of any IMF is at the half of the frequency of the previous one which illustrates the characteristics of dyadic filter bank. (b) Log energies of individual IMFs of fGn. It decreases almost linearly, and higher frequency (lower order) IMFs contain more energies.
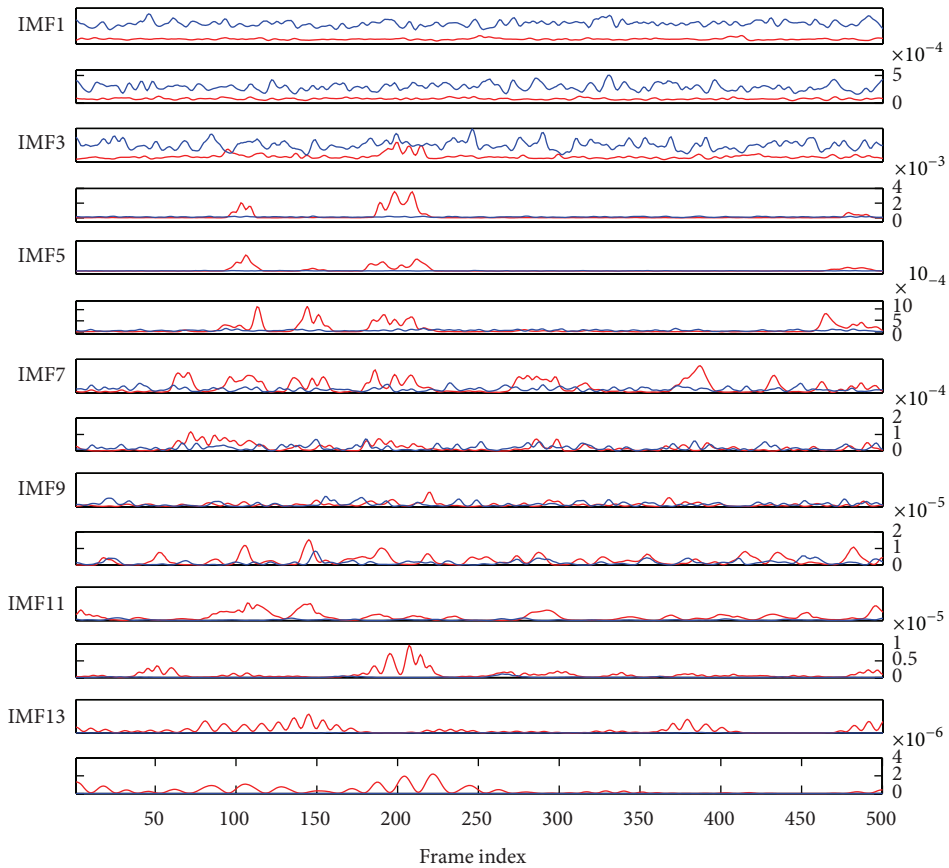

FIGURE 3: Frame-wise noise variance (blue line) and speech energy (red line) of corresponding IMFs computed from imaginary and real parts, respectively.

up to get the enhanced speech signal. For any given rth frame of $\delta_k(t)$, the thresholded coefficient is calculated as

$$
\widetilde{\delta}_{k(r)}(q) = \begin{cases} 0, & \text{if } |\delta_{k(r)}(q)| \leq \sigma_{\eta k(r)} \\ \text{sign}\left[\delta_{k(r)}(q)\right]\phi_{k(r)}(q), & \text{otherwise,} \end{cases}
$$

(11)

where $\phi_{k(r)}(q) = \max[0, \delta_{k(r)}(q) - b\lambda_k]$, $\sigma_{\eta k(r)}$ is the noise level of rth frame of $\gamma_k(t)$, and $\delta_{k(r)}(q)$ denotes the qth sample of rth frame of $\delta_k(t)$; the product $b\lambda_k$ is the adaptive threshold function, while b is being the sorted index of $|\delta_{k(r)}(q)|$. The threshold factor λ_k is varied adaptively for individual IMF based on its variance. An estimated value of λ_k can be obtained as

$$
\lambda_k = \frac{\beta \sigma_{\eta k}}{\sqrt{1/Q \sum_{q=1}^{Q} q^2}},
$$

(12)

where $\sigma_{\eta k}^2$ is the noise variance of $\gamma_k(t)$, β is the adaptation factor, and Q is the frame length (in sample). The optimum β is calculated experimentally as [5]

$$
\beta_{\text{opt}} = f(g) = a_0 + a_1 g + a_2 g^2 + a_3 g^3
$$

(13)

to be fitted to the data points (g_i, h_i), $i = 1, 2, \ldots, I$, where g_i and h_i are the input SNR and optimum value of β (to obtain the maximum output SNR), respectively, for the training speech data and I (=9) represent the maximum number data points to be fitted.

4.3. Proposed Algorithm. The proposed bEMD-based speech enhancement algorithm can be summarized as follows.

(i) The overall noise variance is approximately estimated from observed speech signal by (6), and such variance is used to adjust the amplitude of fGn.

(ii) Noisy speech signal and fGn are combined producing complex signal $x(t)$.

(iii) bEMD is used to decompose $x(t)$ into complex-valued IMFs in which real and imaginary parts correspond the IMFs of speech and fGn, respectively.

(iv) Each IMF is divided into 4 ms frames to perform very local soft-thresholding. The frame variance of fGn's IMF and energy of the corresponding frame of speech are computed.

(v) The frame variance of fGn is used as data adaptive reference energy for binary classification of the corresponding speech frame into noise or signal dominant.

(vi) Only the noise dominant frames are processed using data adaptive soft-thresholding. The optimum adaptation factor is computed using (13). The signal dominant frames are left untouched.

(vii) All the processed IMFs of speech signals are summed up to obtain the enhanced speech.

TABLE 1: Overall output SNR using proposed method and comparison with other methods [2, 3] for white noise.

Noise level	SWm [3]	HST [17]	uEMD [5]	bEMD
0 dB	7.18	7.56	8.0	8.6
5 dB	10.49	10.21	11.6	12.2
10 dB	13.52	13.14	15.1	15.7
15 dB	17.24	15.87	18.8	19.5
20 dB	20.75	20.01	22.6	23.2
25 dB	24.48	24.85	26.6	27.1
30 dB	29.08	29.24	31.2	31.3

5. Experimental Results and Discussions

The effectiveness of the proposed algorithm is tested using computer simulation with different 10 male and 10 female utterances (English sentences) randomly selected from TIMIT Database. The sampling frequency of all the speech signals is set to 16 kHz. The white noise is added to the clean speech to obtain the noisy speech signals at different noise levels. The simulation is performed over those noisy speech signals. The denoising results of the proposed method are illustrated in Figure 4. The waveforms as well as the spectrograms of the clean and noisy of 10 dB SNR are shown in Figures 4(a) and 4(b). The outputs of the uEMD [5] and proposed bEMD-based approach are presented in Figures 4(c) and 4(d), respectively. It is observed in Figure 4(c) that using the uEMD, a small amount of noise is still remaining in the enhanced speech at low frequency regions. It is obvious that a considerable amount of noise is reduced using the bEMD, a better performance when is shown.

The overall output SNR (for white noise) of the proposed method is compared with the algorithms—spectral weighting-based speech enhancement SWm [3], hard and soft-thresholding (HST) technique [17], and univariate EMD-based soft-thresholding (uEMD) [5] as illustrated in Table 1. It is observed that both SWm [3] and HST [17] algorithms suffer from speech degradation at higher SNRs (above 20 dB). We can conclude from the outcomes of Table 1 and Figure 4 that the proposed bEMD results in a high speech enhancement score and clear sound without loss of speech content. The principal limitation of uEMD is to determine the speech-less part to estimate noise variance and that is inefficient for the application of continuous speech processing.

Although overall SNR is a good measure for quantifying performance, it has a little perceptual meaning. A better measure can be achieved by calculating segmental SNR (segSNR) calculated within the frames of short duration. In this experiment, the 20 ms of frame length is used with 13.75 ms overlapping between the adjacent frames. Figure 5 shows the comparisons between the input and output segSNR for white noise as a function of frame index obtained by uEMD [5] and the proposed bEMD-based methods. It is observed in Figure 5 that the segmental output SNR is higher than that of the input SNR over all the frames of 0 dB and 5 dB noisy speech. Hence, the noise dominant frames are classified properly using the proposed method, and the noise is removed from those frames. It is noted that the
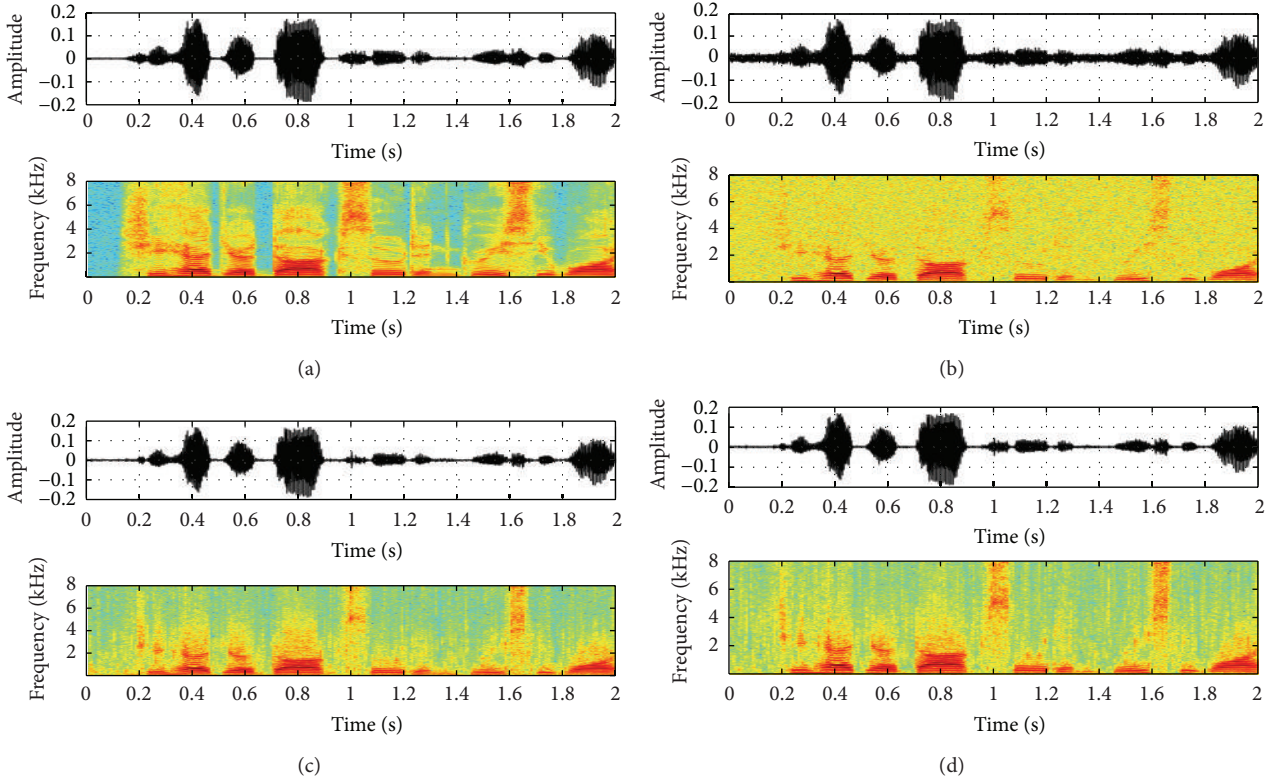
FIGURE 4: (a) Waveform (top) and spectrogram (bottom) of clean speech. (b) Waveform (top) and spectrogram (bottom) of noisy (10 dB) speech for white noise. (c) Waveform (top) and spectrogram (bottom) of enhanced speech obtained by uEMD-based method. (d) Waveform (top) and spectrogram (bottom) of enhanced speech obtained by bEMD-based method.
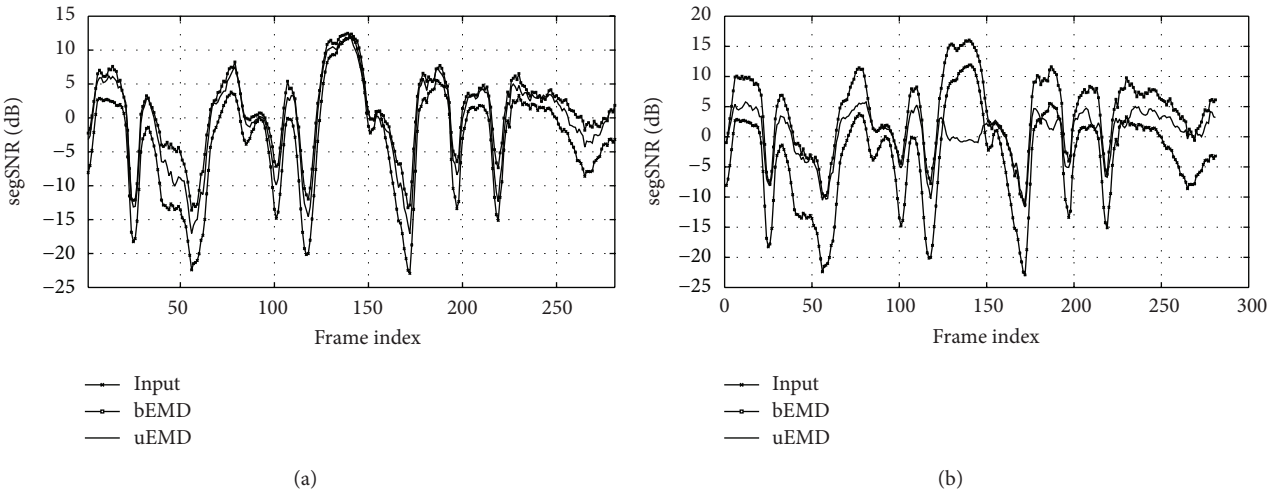

FIGURE 5: Segmental SNR using bEMD and uEMD-based denoising techniques for noisy speech corrupted by white noise—0 dB SNR (a) and 5 dB SNR (b).

classification of frames using IMFs of fGn is performed effectively over the whole speech. Since the segmental SNR provides high correlation of subjective result, the proposed bEMD-based adaptive thresholding algorithm works well in this respect.

The experiment is also carried out with these noisy signals to observe the efficiency of the algorithm in terms

of the perceptual evaluation of speech quality (PESQ) [18]. Figure 6 shows the speech enhancement performance of the proposed method and a comparison of that with uEMD in terms of PESQ. The values 4 and 0 of PESQ measurement represent highest and lowest perceptual quality of the speech, respectively. Considering the theoretical background and experimental results, it is obvious to state that the proposed
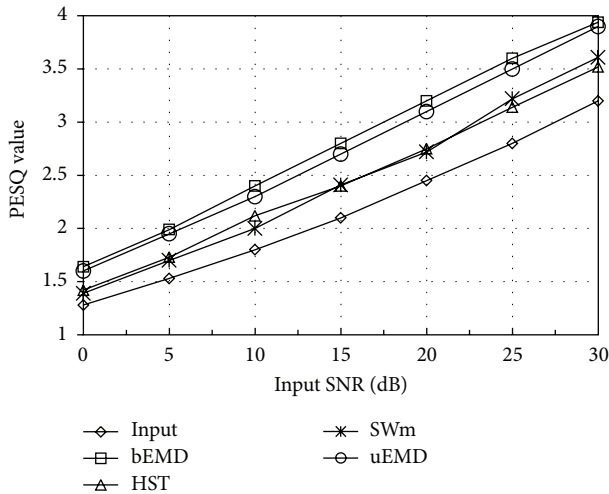
FIGURE 6: Input/output PESQ for speech contaminated with white noise at different SNRs levels using uEMD and bEMD-based methods.

bEMD-based algorithm exhibits better results in speech enhancement with the minimization of the loss of speech intelligibility.

6. Conclusions

In this paper, a data adaptive soft-thresholding algorithm is introduced to effectively suppress the background noise components from the noisy speech signal. It is more challenging to determine the noise variance as well as the adaptive threshold in single channel speech enhancement. In the previous study [5], the variance of noise is calculated from the speechless part of single channel speech based on speech/silence detector. It is obvious that the performance of thresholding approach is very much dependent on the perfectness of detection of noise variance and threshold as well. The newly developed bivariate EMD (bEMD) is applied here to resolve the mentioned problem in more challenging situation. In bEMD, the speech signal and the reference signals, that is, fGn, are considered as the real and imaginary components and decomposed simultaneously, yielding the same number of IMFs for both signals. The results decomposed by the bEMD are more legible than that by the ordinary EMD, and it performs better analysis of speech signals. An adaptation factor is used in the adaptive threshold function. The optimal value of adaptation factor is computed based on the estimated input SNR. The experimental result shows that the proposed speech enhancement algorithm works efficiently for a wide range of input SNR. The performance (both quantitative and qualitative) of this algorithm is tested with the speech contaminated with white noise and nonwhite noises. The proposed method works well for white noise. Although uEMD-based method exhibits a slight better performance for non-white noise, the major advantage of bEMD is that it does not require the detection of silence part within the speech signal. Further research is

required to improve the performance of the proposed method with different types of noises.

References

[1] S. F. Beaufays and V. Digalakis, "Training data clustering for improved speech recognition," in *Proceedings of the EUROSPEECH*, Madrid, Spain, 1995.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[4] E. Deger, M. K. I. Molla, K. Hirose, N. Minematsu, and M. K. Hasan, "Speech enhancement using soft thresholding with DCT-EMD based hybrid algorithm," in *Proceedings of the EUSIPCO*, Septembre 2007.

[5] M. E. Hamid, S. Das, K. Hirose, and M. K. I. Molla, "Speech enhancement using EMD based adaptive soft-thresholding (EMD-ADT)," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 5, no. 2, pp. 1–16, 2012.

[6] N. E. Huang, Z. Shen, S. Long et al., "The empirical mode decomposition and Hilbert spectrum for non-linear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, pp. 903–995, 1998.

[7] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 936–939, 2007.

[8] T. Tanaka and D. P. Mandic, "Complex empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 2, pp. 101–104, 2007.

[9] 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*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[10] P. Flandrin, G. Rilling, and P. Gonçalvés, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[11] M. K. I. Molla, T. Tanaka, T. M. Rutkowski, and A. Cichocki, "Separation of EOG artifacts from eeg signals using bivariate EMD," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 562–565, March 2010.

[12] M. K. I. Molla, P. R. Ghosh, and K. Hirose, "Bivariate EMD-based data adaptive approach to the analysis of climate variability," *Discrete Dynamics in Nature and Society*, vol. 2011, Article ID 935034, 21 pages, 2011.

[13] D. Looney and D. P. Mandic, "Multiscale image fusion using complex extensions of EMD," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1626–1630, 2009.

[14] M. B. Luca, S. Azou, G. Burel, and A. Serbanescu, "On exact Kalman filtering of polynomial systems," *IEEE Transactions on Circuits and Systems I*, vol. 53, no. 6, pp. 1329–1340, 2006.

[15] A. Kagan and L. A. Shepp, "Why the variance?" *Statistics and Probability Letters*, vol. 38, no. 4, pp. 329–333, 1998.

[16] C. Park, D. Looney, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition," *IEEE Transactions on Neural*

Systems and Rehabilitation Engineering, vol. 19, no. 4, pp. 366–373, 2011.

[17] S. Salahuddin, S. Z. Al Islam, M. K. Hasan, and M. R. Khan, "Soft thresholding for DCT speech enhancement," *Electronics Letters*, vol. 38, no. 24, pp. 1605–1607, 2002.

[18] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ): a new method for speech quality assessment of telephone networks and codecs," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing*, pp. 749–752, May 2001.