



Speech enhancement using adaptive thresholding based on gamma distribution of Teager energy operated intrinsic mode functions

Özkan ARSLAN¹, Erkan Zeki ENGİN^{2,*}

¹Department of Electronics and Communication Engineering, Faculty of Engineering,
Tekirdağ Namık Kemal University, Tekirdağ, Turkey

²Department of Electrical and Electronics Engineering, Faculty of Engineering, Ege University, İzmir, Turkey

Received: 04.04.2018

Accepted/Published Online: 08.01.2019

Final Version: 22.03.2019

Abstract: This paper introduces a new speech enhancement algorithm based on the adaptive threshold of intrinsic mode functions (IMFs) of noisy signal frames extracted by empirical mode decomposition. Adaptive threshold values are estimated by using the gamma statistical model of Teager energy operated IMFs of noisy speech and estimated noise based on symmetric Kullback–Leibler divergence. The enhanced speech signal is obtained by a semisoft thresholding function, which is utilized by threshold IMF coefficients of noisy speech. The method is tested on the NOIZEUS speech database and the proposed method is compared with wavelet-shrinkage and EMD-shrinkage methods in terms of segmental SNR improvement (SegSNR), weighted spectral slope (WSS), and perceptual evaluation of speech quality (PESQ). Experimental results show that the proposed method provides a higher SegSNR improvement in dB, lower WSS distance, and higher PESQ scores than wavelet-shrinkage and EMD-shrinkage methods. The proposed method shows better performance than traditional threshold-based speech enhancement approaches from high to low SNR levels.

Key words: Speech enhancement, empirical mode decomposition, gamma distribution, Teager energy, Kullback–Leibler divergence

1. Introduction

The corruption of speech signals by environmental noise negatively affects the quality and intelligibility of speech, resulting in a severe decrease in the performance of the applications. Speech enhancement is an important research field with applications in speech coding, automatic speech recognition systems, digital hearing aids, and mobile communication systems. Speech enhancement algorithms must be developed to minimize speech distortion while maximizing noise reduction. In recent years, various methods have been developed to reduce noise while maintaining the quality and intelligibility in speech enhancement problems. These methods are divided into three basic groups based on the time, frequency, and time-frequency domains. Time domain methods include the subspace enhancement method [1] and frequency domain methods include spectral subtraction [2, 3], MMSE estimators [4], and Wiener filtering [5, 6]. The time-frequency methods include wavelet transform [7, 8] and empirical mode decomposition (EMD) methods [9–12]. In speech enhancement, time domain methods such as the subspace approach give rise to speech distortion and residual noise and it is very difficult to use them in real time in terms of computational load. The spectral subtraction approach is historically one of the first algorithms based on a simple principle that is recommended for noise reduction. In this method, the speech signals are degraded by unwanted disturbing sound known as “musical noise”. Many

*Correspondence: erkan.zeki.engin@ege.edu.tr

algorithms aim to eliminate these undesired sounds [3]. The frequency domain methods, which are statistical model-based speech enhancement algorithms such as MMSE and log-MMSE estimators, focus on the nonlinear estimation of the amplitude from the discrete Fourier coefficients of the signal complexity using various statistical models and optimization criteria [13]. One of the major problems of these methods is that the variance of the spectral coefficients is quite large. Nonlinear time-frequency domain methods such as wavelet transform and empirical mode decomposition are proposed to overcome these shortcomings in speech enhancement algorithms.

In the methods based on wavelet transform [14, 15], noisy speech is decomposed and then the enhanced speech can be obtained by using threshold values, which are the decision criteria. A fixed threshold value is applied to all wavelet coefficients and this may cause elimination of speech components. Thus, adaptive threshold techniques are preferred in noisy speech enhancement. However, there are some disadvantages of speech enhancement by wavelet transform. These disadvantages are the use of the fixed mother wavelet and the variation of the number of subbands used in perceptual wavelet packet transform. Therefore, EMD, a new temporal signal decomposition method in recent years, has been proposed to analyze nonstationary and nonlinear time series. In this adaptive method, the signal decomposes into oscillatory components that are called intrinsic mode functions (IMFs). It is based on the principle of decomposing a signal into the sum of high and low frequency components. The greatest advantage of the method is that the basis functions are extracted from the signal itself. Therefore, the decomposition is adaptive, unlike the existing state-of-the-art algorithms in which basis functions are fixed.

In speech enhancement problems, filtering based on EMD [9, 16] or threshold approaches [12, 17] are proposed to reduce noise. However, in the EMD method, setting a fixed threshold value can cause elimination of the component of the speech besides the noise. Therefore, the adaptive threshold values from the IMF coefficients instead of using a fixed threshold value give better improvement for noisy speech [18]. The adaptive threshold values can be determined using the Teager energy operator (TEO) as well as an enhancement technique based on wavelet transform [19–21].

In this study, we have developed a new approach to the adaptive threshold of IMF coefficients to obtain enhanced speech. In our proposed method, a suitable statistical model of the energy of the IMF coefficients of speech and nonspeech regions is derived unlike other EMD-based speech enhancement methods. By using the statistical distribution, the adaptive threshold values are determined based on the possibility of speech or the possibility of noise for noisy speech signals. In this proposed approach, the noisy speech signals are decomposed to IMFs and residual signal by the EMD method and the TEO is applied to these IMFs. The probability density function of noisy speech and noise are obtained by the gamma distribution of Teager energy (TE) operated IMFs for noisy speech. Symmetric Kullback–Leibler (SKL) divergence is used to identify the suitable threshold values for each mode. Unlike traditional hard or soft thresholding functions, a semisoft thresholding function is utilized to get enhanced speech.

The paper is structured as follows. In Section 2, the algorithm for determining the adaptive threshold values is explained. In Section 3, the speech quality and intelligibility measures are briefly explained. In Section 4, the results obtained via the proposed method are presented and compared with traditional shrinkage methods. Finally, Section 5 contains some concluding remarks.

2. Proposed method

The steps of the proposed speech enhancement approach are given in Figure 1. As shown in Figure 1, noisy input speech is separated into frames. Since the noise characteristic plays an important role in improving the

noisy speech signals, noise is first obtained by a noise estimation algorithm from the noisy speech signal. Then the EMD method is applied to each frame of noisy speech and the estimated noise signal. The TEO is applied to noisy speech and noise and the adaptive threshold values are calculated based on distributions of the TEO for each mode. After this process, IMF coefficients of noisy signals are thresholded. On the threshold of the IMF coefficients, the semisoft thresholding function, which is frequently used in recent years, has been used. Finally, an enhanced signal is achieved with inverse empirical mode decomposition and the overlap-and-add method.

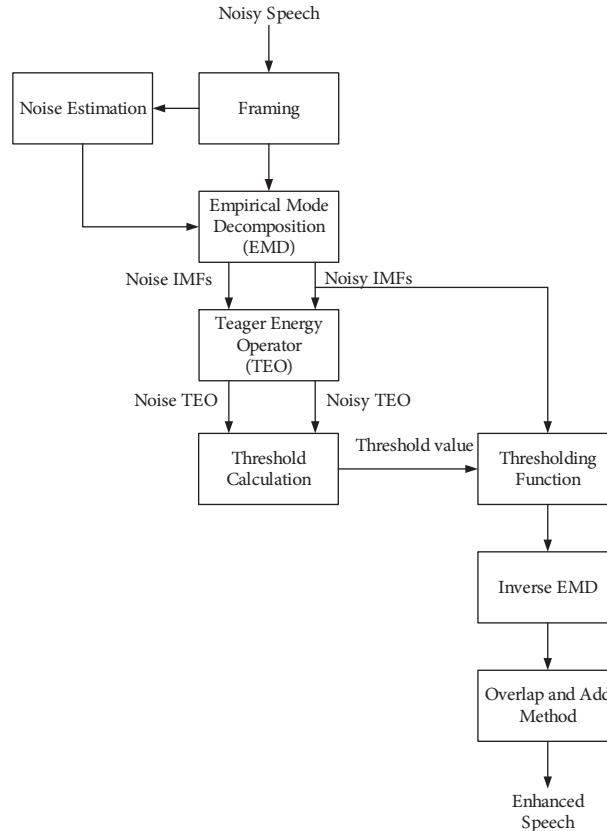


Figure 1. The steps of the proposed algorithm.

2.1. Noise estimation

In speech enhancement, determining the characteristic of the noise is important in improving the performance of enhanced speech. Recently, the optimally modified log-spectral amplitude estimator (OMLSA) algorithm [22] has been frequently used as a noise estimation method in speech enhancement studies [23, 24]. The traditional speech enhancement approaches seek to obtain the spectrum of clean speech under the speech presence hypothesis. However, a modified estimator is determined the characteristic of noise under a speech absence probability and this approach performs better when the enhanced speech is obtained [25, 26]. This noise estimator, developed by Eprahim and Mallah [22], is obtained by multiplying the spectral gain in the frequency domain and the estimated probability of speech presence.

2.2. Empirical mode decomposition

EMD, a fundamental part of the Hilbert–Huang transform (HHT), is an adaptable algorithm for nonstationary signals. The EMD method decomposes the signal into a sum of oscillating components. It is an adaptable

method unlike Fourier transform (FT) or Wavelet transform (WT). EMD is preferred over other decomposition methods due to the basis functions used for signal decomposition being obtained from the signal itself. Therefore, unlike other methods that use fixed basis functions, this method is also adaptable to signal analysis and decomposition [27]. EMD can be viewed as a type of decomposition that is used to separate the original signal from the high-frequency components into the low-frequency components. Any complex signal can be decomposed by EMD into IMFs that are used as a set of basis functions to express the signals.

The sifting process separates the signal into IMFs via the following steps:

- 1) Identify all extrema (local minima and maxima) of the original signal, $x(t)$.
- 2) The extrema are calculated by cubic spline interpolation to obtain the upper $e_{max}(t)$ and lower $e_{min}(t)$ envelopes.
- 3) The average value of the upper and lower envelopes ($m_1(t) = (e_{max}(t) + e_{min}(t))/2$) is calculated and subtracted from the original signal to obtain $h_1(t) = x(t) - m_1(t)$.
- 4) $h_1(t)$ is evaluated as it provides two IMF conditions, which are given below:
 - a) The difference between the number of extrema and zero crossings is no more than one.
 - b) The average value of the upper and the lower envelopes is zero. If $h_1(t)$ satisfies the condition of being an IMF, the first $IMF_1(t)$ is accepted as $IMF_1(t) = h_1(t)$.
- 5) If $h_1(t)$ cannot meet the above conditions, it is considered as a new signal and it is repeated over Steps 1–4 on $h_1(t)$ to form $h_2(t)$. The sifting process is repeated until $h_2(t)$ is considered as an IMF. If $h_2(t)$ cannot meet two conditions, the stopping criterion is calculated to complete the sifting process. The stopping criterion can be defined as follows:

$$SD(i) = \sum_{t=0}^N \frac{|h_{i-1}(t) - h_i(t)|^2}{h_{i-1}^2(t)}. \quad (1)$$

Generally, the SD range is between 0.2 and 0.3. If $h_2(t)$ satisfies SD, then it is accepted as $IMF_1(t) = h_2(t)$. If $h_2(t)$ does not satisfy the criterion, operations are repeated over $h_2(t)$ and it is calculated as $IMF_1(t) = h_i(t)$.

- 6) The residual signal, $r_1(t) = x(t) - IMF_1(t)$, is obtained. Then the original signal is separated into IMF components, such as $IMF_1(t), IMF_2(t), \dots, IMF_n(t)$, and residual sequence $r_n(t)$. Finally, the original signal is defined as:

$$x(t) = \sum_{i=1}^N IMF_i(t) + r_n(t). \quad (2)$$

Figure 2 shows the first five IMFs and residual signal obtained from decomposing a signal with different frequency components. As can be seen from Figure 2, the EMD method decomposes the signal from high frequencies to low frequencies.

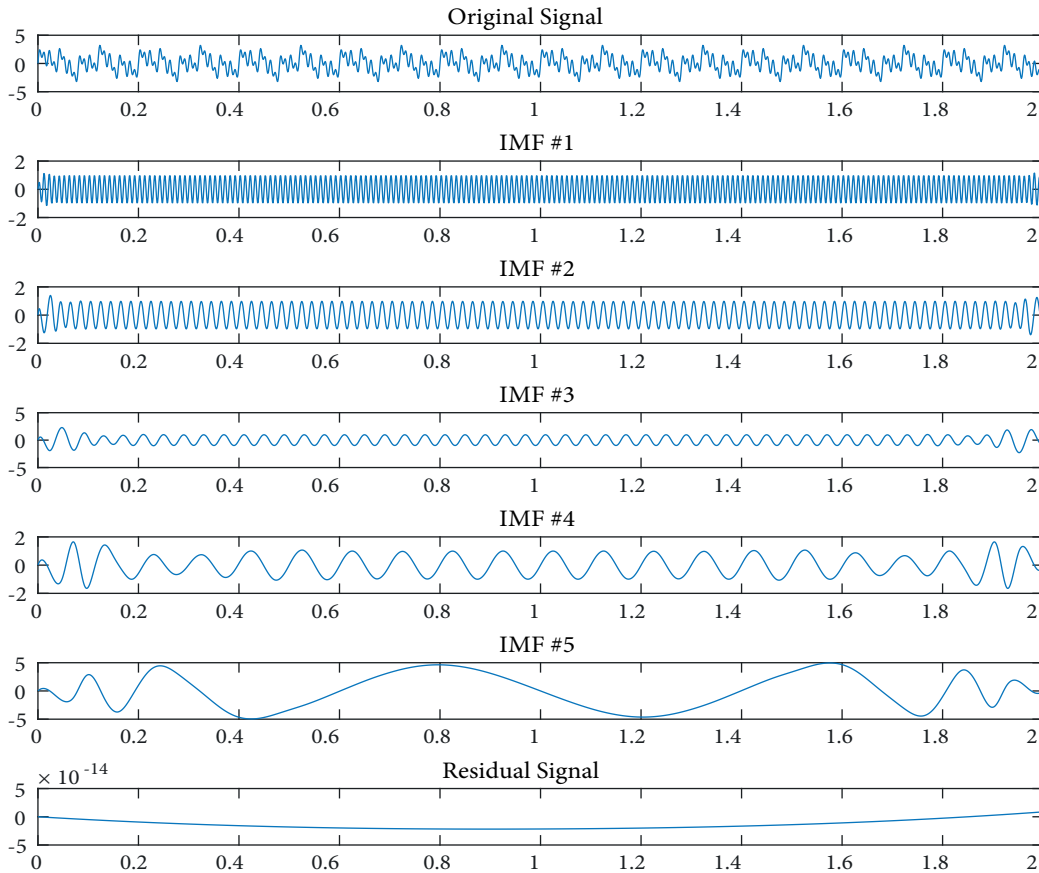


Figure 2. The first five IMFs and residual signal of a signal with different frequency components.

2.3. Teager energy operator

The TEO is an effective and powerful operator for nonstationary signals and it has been widely used in various speech applications [28, 29]. The discrete-time TEO for signal $y[n]$ is expressed as:

$$\Psi_d(y[n]) = y[n]^2 - y[n+1]y[n-1]. \quad (3)$$

The TEO is almost instantaneous because it just requires three samples of a signal. The TEO can be used to determine the activity of speech or nonspeech since it is able to predict the energy of the nonstationary signals. Then it provides a suitable threshold value for speech or nonspeech regions. However, when the TEO is used directly in noisy speech, unwanted disturbance effects (artifacts) and amplified noises can occur. Hence, it is more plausible to apply the TEO of the IMFs instead of applying the TEO to the noisy speech [18]. The determination of threshold values using the TEO provides greater potential for enhanced noisy speech [30]. The noisy speech signals are first separated into IMFs and residual signals by the EMD method and then the TEO is applied to each mode as follows:

$$E_{k,m} = \Psi[IMF_{k,m}(t)], \quad (4)$$

where k and m represent each mode and frame, respectively.

2.4. Generalized gamma statistical model

The performance of threshold-based speech enhancement methods depends primarily on the threshold value and thresholding function. It is not sensible to use a fixed threshold value for all EMD modes in the improvement of noisy speech. The threshold value for each mode must be adaptable in order to increase the speech quality and ensure intelligibility. Considering the probability distribution function (pdf) of the signal energy $E_k(t)$, an appropriate similarity measure should be used to find a more accurate threshold value [21, 31]. Due to the nonstationary structure of speech, the correct pdf estimate of speech or its energy is difficult. Therefore, instead of extracting the pdf of the speech energy directly, it would be more appropriate to extract the appropriate probability distribution over the empirical histogram of the speech energy. The determination of the most appropriate probability distribution function to fit the empirical histogram of the speech energy is important in determining the threshold values. The Gaussian, exponential, Rayleigh, and gamma distributions are investigated in order to determine the optimal distribution to fit the empirical histogram in white, car, and babble noise conditions at seven different SNR levels ranging from -15 dB to 15 dB. Figure 3 shows the empirical histogram, Gaussian, and gamma distributions of the TEO applied to IMF coefficients at -10 , 0 , and 10 dB SNR levels for the car noise condition. Similarly, Figure 4 shows the distributions of the TEO-applied IMF coefficients of the estimated noise signal for the same noise and SNR levels. These figures show that the gamma distribution clearly fits the empirical histogram of the TE of noisy speech and noise. Therefore, the gamma distribution function is used to determine the threshold values that are adaptable for each mode.

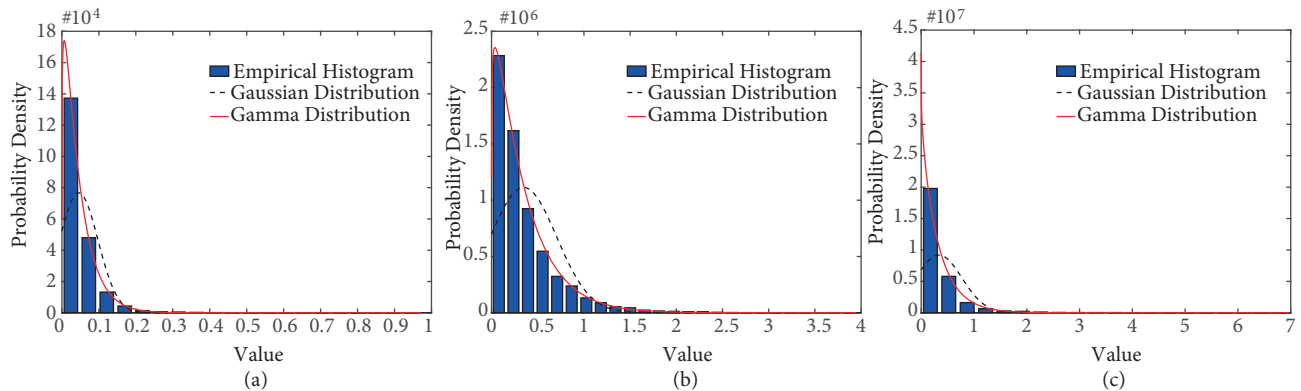


Figure 3. Empirical histogram, Gaussian, and gamma distributions for noisy speech IMFs energy at (a) -10 dB, (b) 0 dB, and (c) 10 dB SNR levels.

2.5. Proposed adaptive threshold calculation

It is proposed to determine the threshold values by using the probability distribution of the energy of the IMF coefficients instead of directly using the Teager energy of these coefficients of each mode separated by EMD. It is expected that the energy distribution of noisy speech and noise will be similar at a certain range. Also, outside this range, it is expected that the probability distribution of clean speech and the energy of noisy speech will be equal. Thus, the threshold value is correctly obtained by using an appropriate model mapping scheme or similarity level among the probability distributions. The widely used Kullback–Leibler divergence similarity criterion is used to determine the distance between two distributions [32, 33]. The K-L divergence is zero if the distributions of the noise and noisy speech are exactly the same. Otherwise, it must be positive. In this study,

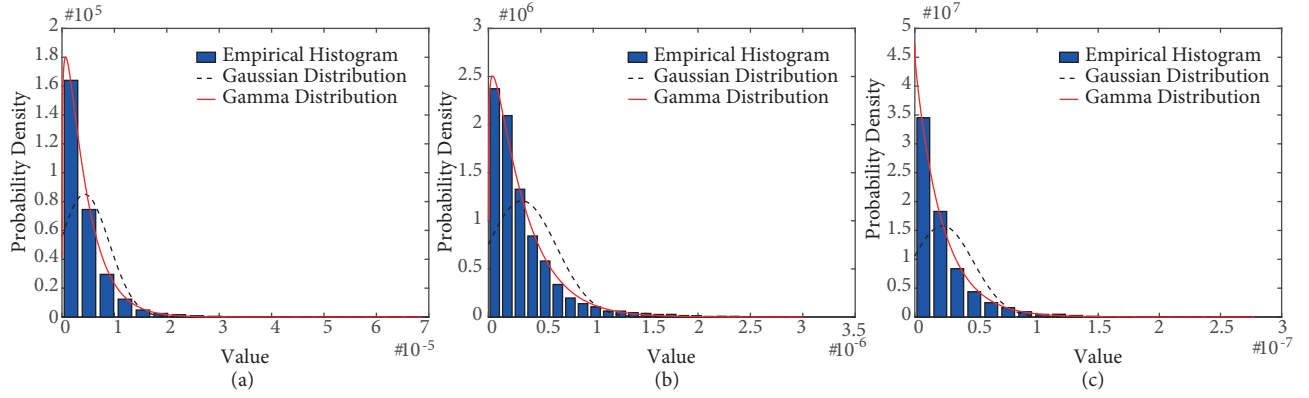


Figure 4. Empirical histogram, Gaussian, and gamma distributions for estimated noise IMFs energy at (a) -10 dB, (b) 0 dB, and (c) 10 dB SNR levels.

SKL divergence is used and it can be expressed as:

$$SKL(p, q) = \frac{KL(p, q) + KL(q, p)}{2}, \quad (5)$$

where p and q denote the gamma probability density functions of the energy of noisy speech and noise calculated from related histograms with N number of bins. $KL(p, q)$ is calculated as:

$$KL(p, q) = \sum_{i=1}^N p_i(E_{k,m}(t)) \ln \frac{p_i(E_{k,m}(t))}{q_i(E_{k,m}(t))}. \quad (6)$$

$p_i(E_{k,m}(t))$ and $q_i(E_{k,m}(t))$ are the approximate gamma probability distribution function of the energy of the noisy speech and noise $E_{k,m}(t)$, respectively. For adaptive threshold values λ , the SKL divergence between $p_i(E_{k,m}(t))$ and $q_i(E_{k,m}(t))$ is approximately zero:

$$SKL(p_i(E_{k,m}(t)), q_i(E_{k,m}(t))) \approx 0. \quad (7)$$

The generalized gamma probability distribution function is expressed as [34, 35]:

$$p(x) = \frac{\gamma \beta^v}{\Gamma(v)} \left| \frac{x}{\sigma_s} \right|^{\gamma v - 1} \exp\left(-\beta \left| \frac{x}{\sigma_s} \right|^\gamma\right). \quad (8)$$

$\Gamma(\cdot)$ represents the gamma function, and β and v represent the scale parameter associated with the a priori SNR and shape parameter of the generalized gamma function, respectively. In Eq. (8), the γ , β , and v parameters significantly affect the probability distribution of the IMF coefficients in each mode. The γ parameter is usually chosen to be 1 or 2 [35]. In this study, the γ and β parameters were chosen as 1 and the v parameter was chosen as 2. In this case, the gamma probability distribution function for $p_i(E_{k,m}(t))$ and $q_i(E_{k,m}(t))$ can be expressed as:

$$p_i(E_{k,m}(t)) = \frac{x}{\sigma_s} \exp\left(-\frac{x}{\sigma_s}\right) = \frac{x}{\sqrt{\sigma_r^2 + \sigma_n^2}} \exp\left(-\frac{x}{\sqrt{\sigma_r^2 + \sigma_n^2}}\right), \quad (9)$$

$$q_i(E_{k,m}(t)) = \frac{x}{\sigma_n} \exp\left(-\frac{x}{\sigma_n}\right). \quad (10)$$

σ_r^2 and σ_n^2 represent the power of clean signal and noise, respectively, and $\sigma_s^2 = \sigma_r^2 + \sigma_n^2$ denotes the power of the noisy signal.

The threshold value λ is obtained by Eqs.(7), (9), and (10):

$$\int_0^\lambda \left[\frac{x}{\sigma_s} \exp\left(-\frac{x}{\sigma_s}\right) - \frac{x}{\sigma_n} \exp\left(-\frac{x}{\sigma_n}\right) \right] I_1 dx = 0. \quad (11)$$

Given that $I_1 = \ln\left(\frac{\sigma_s}{\sigma_n}\right) \exp\left(-\frac{x}{\sigma_s} + \frac{x}{\sigma_n}\right)$, by solving Eq. (11), the λ threshold values for each mode are calculated as:

$$\lambda(k) = \frac{\sigma_n(k) \sqrt{1 + \gamma(k)}}{\ln \sqrt{1 + \gamma(k)}}, \quad (12)$$

where $\gamma(k)$ represents the segmental SNR values of the k th EMD mode and can be defined as:

$$\gamma(k) = \frac{\sigma_r^2(k)}{\sigma_n^2(k)}. \quad (13)$$

The adaptive threshold values obtained by the proposed method for the first 10 IMF coefficients of the noisy speech signal that is corrupted by babble noise are given in Figure 5. Threshold values show differences in terms of noise type, SNR level, and mode number. As the amount of noise increases, the segmental SNR ($\gamma(k)$) value decreases and threshold values increase accordingly as given in Eq.(12). Also, since the noise is reduced after the 10th mode, the threshold values are decreasing and almost all the components of the noisy speech pass through the threshold.

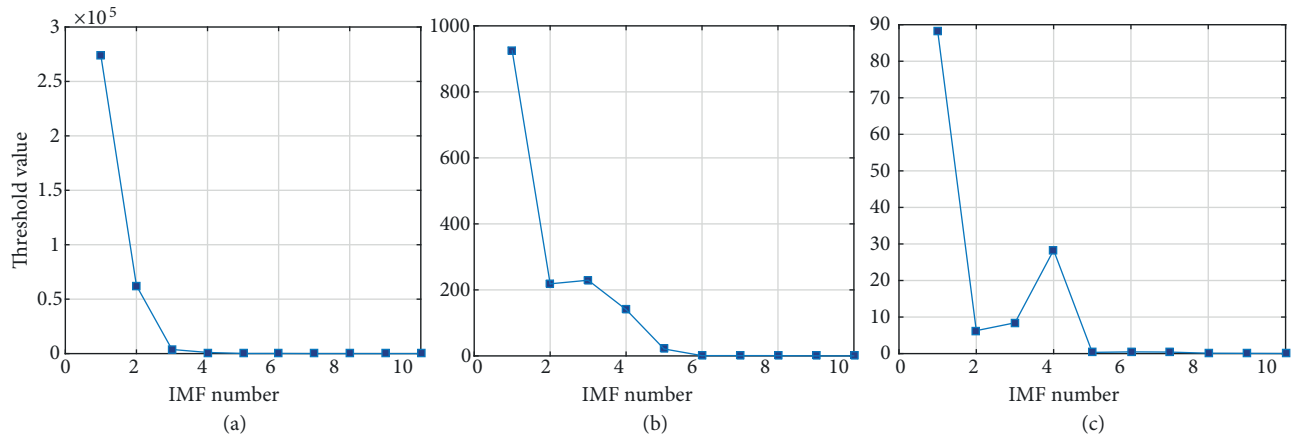


Figure 5. Adaptive threshold values in babble noise conditions for different SNR levels: (a) -10 dB, (b) 0 dB, and (c) 10 dB.

2.6. Thresholding function

The soft thresholding strategy is a powerful technique for removing the noise components from the noisy signal in many speech enhancement applications [15, 30, 36, 37]. The soft thresholding function, defined as the threshold λ_1 and determined by Eq. (12), is applied to the m th frame of the k th mode of $IMF_{k,m}(t)$. It has been proved that the soft thresholding function removes time-frequency discontinuities. However, $IMF_{k,m}(t)$ coefficients are

shifted by $\lambda_1(k)$ when using soft thresholding. A semisoft thresholding function is used to overcome the shift problem. In this case, the shift of the coefficients according to the amount of the threshold value is prevented. In this study, we propose to use a semisoft function for thresholding the IMF coefficients.

Letting $\lambda_2(k) = \sqrt{2}(\lambda_1(k))$, the semisoft thresholding function is defined as:

$$\tilde{f}_{k,m}(t)_{ss} = \begin{cases} 0 & \text{if } |IMF_{k,m}(t)| \leq \lambda_1(k) \\ IMF_{k,m}(t) & \text{if } |IMF_{k,m}(t)| > \lambda_2(k) \\ \text{sgn}(IMF_{k,m}(t)) \cdot G & \text{, otherwise,} \end{cases} \quad (14)$$

$$G = \left[\frac{\lambda_2(k)|IMF_{k,m}(t)| - \lambda_1(k)}{\lambda_2(k) - \lambda_1(k)} \right], \quad (15)$$

where $\tilde{f}_{k,m}(t)$ denotes the $IMF_{k,m}(t)$ coefficients of the semisoft thresholded EMD modes. Figure 6 shows the relationship between input IMF coefficients and output thresholded IMF coefficients for hard, soft, and semisoft thresholding functions.

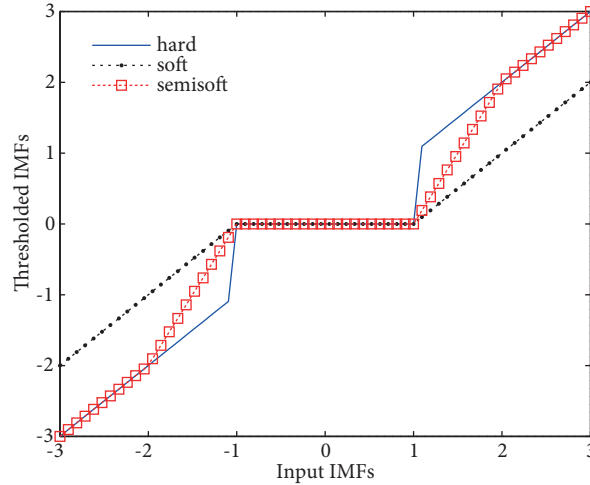


Figure 6. Input-output relationship for hard, soft, and semisoft thresholding functions.

2.7. Inverse empirical mode decomposition

The enhanced speech $\hat{s}[n]$ is obtained by the inverse EMD of frame-based thresholded IMFs. The standard overlap-and-add method is used to reconstruct the frames. Then the enhanced speech frames are reconstructed and the enhancement process is completed. The enhanced speech is defined as:

$$\hat{s}[n] = EMD^{-1}(IMF_{k,m}(t))_{ss}. \quad (16)$$

3. Objective and subjective speech quality measures

This section presents the objective quality and intelligibility measures accepted for the evaluation of speech enhancement methods. In addition, spectrogram representations are examined for subjective evaluation of speech enhancement methods. While the speech quality is evaluated by segmental SNR (SegSNR) and weighted spectral slope (WSS), the intelligibility is evaluated by the perceptual evaluation of speech quality (PESQ) measure.

The segmental SNR is calculated by using the time or frequency domain. Since the calculation of the SegSNR in the time domain is simple, it is widely used. The energy of the speech signal changes with time and in the majority of these time periods the signal has large energy and the noise is relatively inaudible. Therefore, the SNR is calculated in a short period of time and the average is called the SegSNR [38]. The SegSNR improvement in dB is taken into account in the objective evaluation of speech enhancement methods.

The WSS, which is another objective quality measure, is calculated between spectral slopes of clean and enhanced speech signals. These spectral slopes are extracted from each frequency band. The fact that the WSS value is low means that the speech enhancement method is successful at noise reduction.

The PESQ measure is used to objectively assess the intelligibility of speech enhancement methods. The PESQ is a standard measure recommended by ITU-T for speech intelligibility evaluation [39]. The PESQ is a measure that can be used to estimate the mean opinion score (MOS) of clean speech and its degraded speech. The PESQ scores provide information on the intelligibility of enhanced speech and a higher PESQ score indicates better intelligibility for enhanced speech.

4. Experimental results

The proposed method was tested on English sentences taken from the NOIZEUS database [40]. The sampling frequency of the speech signals was 8 kHz. The clean speech signals were corrupted by white, car, and babble noises from the NOISEX92 [41] database with SNR levels ranging from -15 dB to 15 dB. The performance of the proposed method was compared with two speech enhancement algorithms, including wavelet-shrinkage [15] and EMD-shrinkage [9, 27]. The Hamming window function was used to get frames and frame length was set to 64 ms with 32 ms overlapping. The noisy speech signals and estimated noise were decomposed into 7 IMFs by the EMD method. In order to obtain the gamma probability density functions of TE-operated IMFs, γ and β parameters were chosen as 1 and the ν shape parameter was chosen as 2. In the wavelet-shrinkage method, db8 was used as the basis function and the signals were decomposed into 8 levels. In order to obtain the enhanced speech, the semisoft thresholding function was used for the proposed method and the soft thresholding function was used for wavelet-shrinkage and EMD-shrinkage methods.

4.1. Objective evaluation

4.1.1. Results for white noise-corrupted speech

The proposed method has been compared with traditional shrinkage methods according to SegSNR improvement in dB, WSS, PESQ scores, and spectrogram representation. The overall output segmental SNR improvement of the proposed and other methods for white noise is shown in Figure 7.

In Figure 7, the performance of the proposed method is superior to the other methods in terms of SegSNR improvement and WSS values. As the SNR level increases for all methods, the segmental SNR improvement decreases. Figure 7 shows the WSS values of the proposed and other methods for speech signals degraded by white noise at different SNR levels. It can be seen from this figure that the proposed method has much lower WSS values than the other methods at all SNR levels for white noise.

The PESQ scores are given in Table 1 for speech signals that are corrupted at different SNR levels with white noise. As seen in Table 1, higher PESQ results are obtained with the proposed method at all SNR levels. As a result of the PESQ scores, the proposed method is more effective and superior to other methods in providing intelligibility of speech.

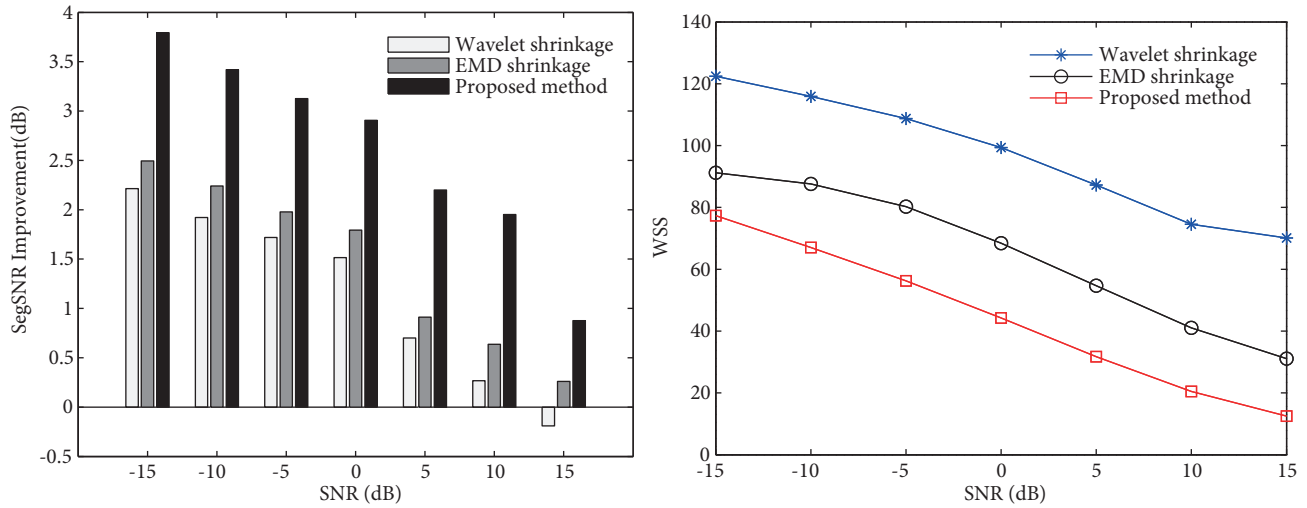


Figure 7. SegSNR improvement and WSS values in white noise for enhancement methods.

Table 1. PESQ scores for different methods in white noise.

SNR	Wavelet-shrinkage	EMD-shrinkage	Proposed method
-15	1.06	1.16	1.38
-10	1.13	1.30	1.40
-5	1.28	1.33	1.73
0	1.48	1.54	1.91
5	1.71	1.91	2.53
10	1.95	2.25	2.97
15	2.15	2.56	3.11

4.1.2. Results for car noise-corrupted speech

The SegSNR improvement in dB, WSS, and PESQ of the proposed and other methods is shown in Figure 8 and Table 2 for car noise. In Figure 8, the evaluation results of the proposed method are compared with the other methods in the car noise condition for all SNR levels in terms of SegSNR improvement in dB. It is seen in Figure 8 that the proposed method provides the best SegSNR improvement for different SNR levels. The segmental SNR improvement increases as SNR decreases for all methods.

Figure 8 demonstrates the WSS values of the proposed and other methods. As shown in this figure, the proposed method using the semisoft thresholding function gives the lowest WSS value. According to these results, the proposed method is more effective than the other methods as improved speech is assessed by the WSS objective quality measure. Also, the WSS value increases as SNR decreases for all methods.

The PESQ results of speech enhancement methods are given in Table 2 for speech degraded by car noise at different SNR levels. In Table 2, it can be seen that higher PESQ results are obtained with the proposed method. These PESQ scores show that the proposed method is more effective and superior to other methods in providing intelligibility of speech.

4.1.3. Results for babble noise-corrupted speech

The SegSNR improvement in dB, WSS values, and PESQ scores of the proposed and other methods is shown in Figure 9 and Table 3 for babble noise. In Figure 9, the evaluation results of the proposed method are compared

Table 2. PESQ scores for different methods in car noise.

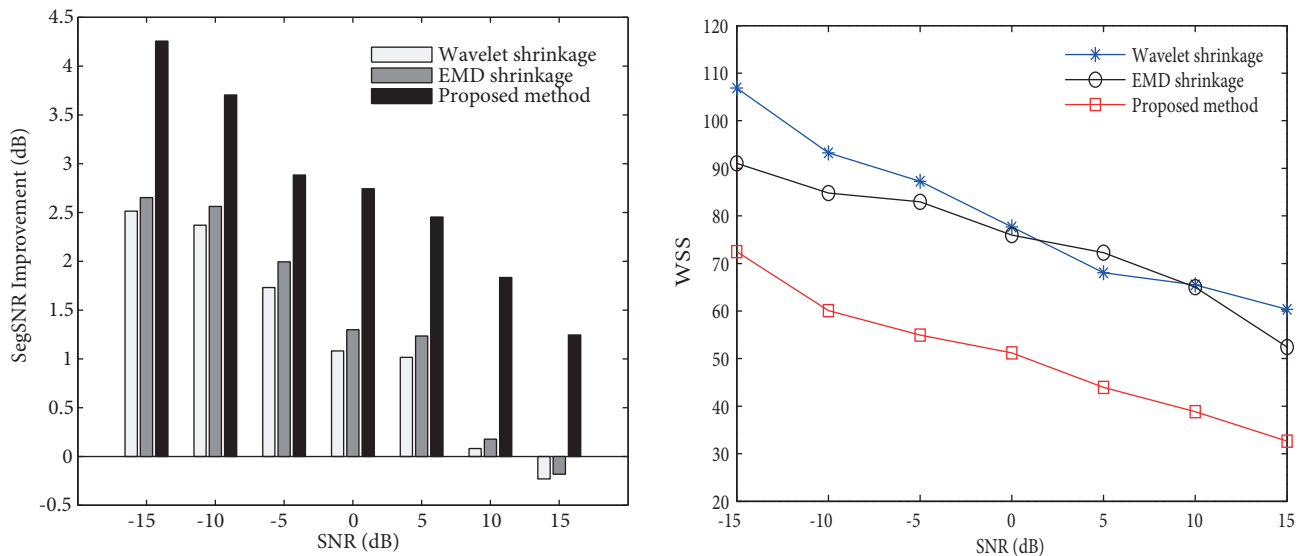
SNR	Wavelet-shrinkage	EMD-shrinkage	Proposed method
-15	0.36	0.79	1.48
-10	0.69	1.07	1.64
-5	0.89	1.41	1.79
0	1.25	1.73	1.93
5	1.61	2.11	2.54
10	1.87	2.42	2.83
15	2.12	2.69	2.94

with the other methods in babble noise conditions for all SNR levels in terms of SegSNR improvement in dB and WSS values. It is seen in Figure 9 that the proposed method provides higher SegSNR improvement and lower WSS values for different SNR levels. The segmental SNR improvement and WSS values increase as SNR decreases for all methods.

The PESQ results of speech enhancement methods are given in Table 3 for speech degraded by babble noise at different SNR levels. In Table 3, it can be seen that higher PESQ results are obtained with the proposed method. These PESQ scores show that the proposed method is more effective and superior to other methods in providing intelligibility of speech.

4.2. Subjective evaluation

In order to obtain a subjective quality evaluation of improved speech signals, the spectrogram representation is used. In Figure 10, the spectrogram representations of enhanced speech are given for all methods. As can be seen in this figure, the high-frequency components of speech are also thresholded with noise in the wavelet-shrinkage method. The high-frequency components of the speech are eliminated with noise and this elimination negatively affects the intelligibility of speech as shown in Table 2. In the EMD-shrinkage approach, which is

**Figure 8.** SegSNR improvement and WSS values in car noise for enhancement methods.

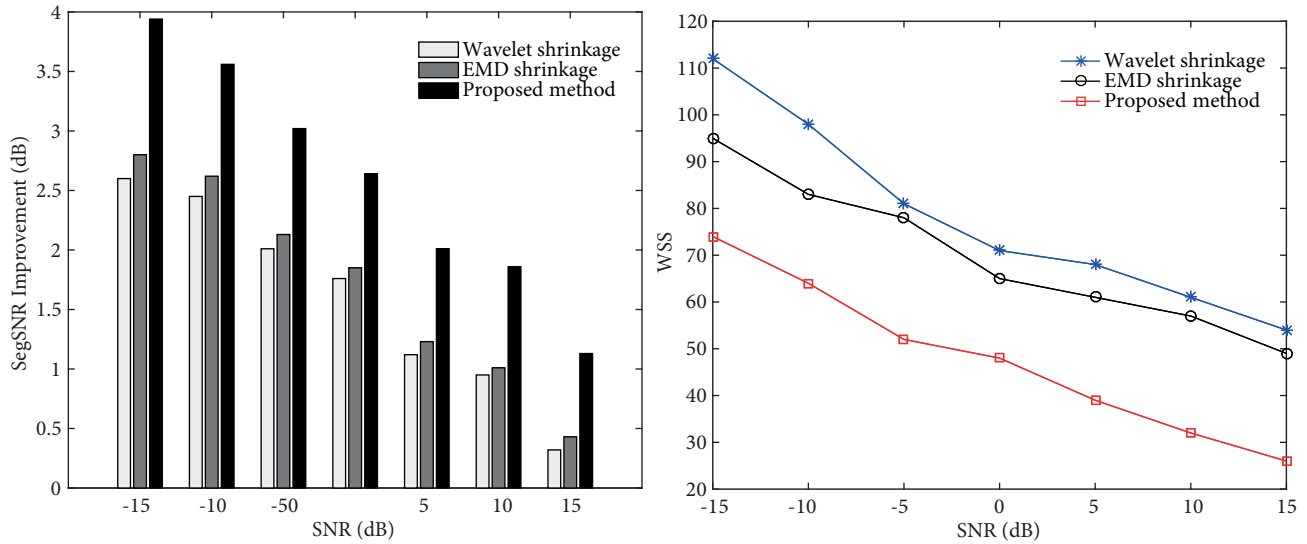


Figure 9. SegSNR improvement and WSS values in babble noise for enhancement methods.

Table 3. PESQ scores for different methods in babble noise.

SNR	Wavelet-shrinkage	EMD-shrinkage	Proposed method
-15	0.43	0.81	1.53
-10	0.74	1.15	1.79
-5	0.99	1.56	1.92
0	1.33	1.84	2.17
5	1.66	2.21	2.68
10	1.95	2.55	2.86
15	2.02	2.60	3.02

another method that is used, it is observed that all components of speech are protected as much as possible while noise is not sufficiently thresholded. It is clear from Figure 10 that the high-frequency components of speech are not eliminated and also the noise is significantly reduced in the proposed method.

5. Conclusions

This study has proposed a novel speech enhancement algorithm based on the gamma statistical model of TE-operated IMFs using the EMD signal decomposition method. The gamma statistical model is adopted to identify the adaptive threshold values by using TE-operated IMFs of noisy speech and estimated noise. The threshold values here are set according to speech and nonspeech regions instead of a unique threshold value. Then the IMFs of noisy speech are thresholded by a semisoft thresholding function. The proposed method improved segmental SNR, WSS, and PESQ, which are highly correlated with speech quality and intelligibility. Simulation results show that the proposed methods are effective and outperformed in terms of higher segmental SNR, lower WSS values, and higher PESQ scores compared with the wavelet-shrinkage and EMD-shrinkage methods for different SNR levels. In order to confirm the results obtained, spectrogram representations of enhanced speech were evaluated for all methods and it is clear that the high-frequency components of speech are not eliminated and also the noise is significantly reduced in the proposed method.

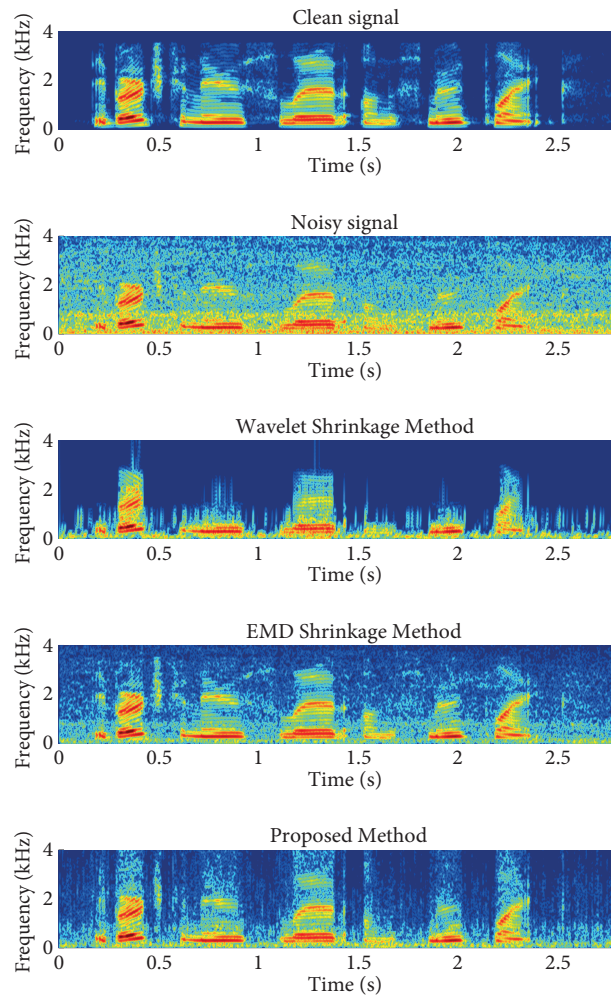


Figure 10. Spectrogram representations of the proposed and other methods for car noise condition at 5 dB SNR level.

References

- [1] Loizou PC, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. *J Acoust Soc Am* 2005; 5: 2791-2793.
- [2] Paliwal KK, Schwerin B, Wójcicki KK. Modulation domain spectral subtraction for speech enhancement. In: Tenth Annual Conference of the International Speech Communication Association; 6–10 September 2009; Brighton, UK. pp. 1327-1330.
- [3] Lu Y, Loizou PC. A geometric approach to spectral subtraction. *Speech Commun* 2008; 6: 453-466.
- [4] Hu Y, Loizou PC. A comparative intelligibility study of single-microphone noise reduction algorithms. *J Acoust Soc Am* 2007; 3: 1777-1786.
- [5] Kasap C, Arslan M. A unified approach to speech enhancement and voice activity detection. *Turk J Elec & Eng and Comp Sci* 2014; 21: 527-547.
- [6] Savoji MH, Chehrehsa S. Speech enhancement using Gaussian mixture models, explicit Bayesian estimation and Wiener filtering. *Iranian Journal of Electrical and Electronic Eng* 2014; 10: 168-175.
- [7] Ghanbari Y, Karami-Mollaei MR. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Commun* 2006; 48: 927-940.

- [8] Cohen I. Enhancement of speech using bark-scaled wavelet packet decomposition. In: 7th European Conference on Speech Communication and Technology; 3–7 September 2001; Aalborg, Denmark. pp. 1933-1936.
- [9] Khaldi K, Boudraa A, Bouchikhi A, Alouane M. Speech enhancement via EMD. *EURASIP J Adv Signal Process* 2008; 1: 1-8.
- [10] Kopsinis Y, McLaughlin S. Development of EMD-based denoising methods inspired by wavelet thresholding. *IEEE T Signal Process* 2009; 57: 1351-1362.
- [11] Hamid ME, Das S, Hirose K, Molla M. Speech enhancement using EMD based adaptive soft-thresholding (EMD-ADT). *International Journal of Signal Processing, Image Processing and Pattern Recognition* 2012; 5: 1-16.
- [12] Hamid ME, Molla MKI, Dang X, Nakai T. Single channel speech enhancement using adaptive soft-thresholding with bivariate EMD. *ISRN Signal Process* 2013; 20: 1-9.
- [13] Ephraim Y. Statistical-model-based speech enhancement systems. *P IEEE* 1992; 80: 1526-1555.
- [14] Donoho DL, Johnstone JM. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 1994; 81: 425-455.
- [15] Donoho DL. De-noising by soft-thresholding. *IEEE T Inf Theory* 1995; 41: 613-627.
- [16] Sharma R, Prasanna SM. A better decomposition of speech obtained using modified empirical mode decomposition. *Digit Signal Process* 2016; 58: 26-39.
- [17] Kemiha M. Empirical mode decomposition and normal shrink thresholding for speech denoising. *Int J Inf Theory* 2014; 3: 27-35.
- [18] Khaldi K, Boudraa AO, Komaty A. Speech enhancement using empirical mode decomposition and the Teager–Kaiser energy operator. *J Acoust Soc Am* 2014; 135: 451-459.
- [19] Sanam TF, Shahnaz C. A combination of semisoft and μ -law thresholding functions for enhancing noisy speech in wavelet packet domain. In: 7th International Conference on Electrical & Computer Engineering (ICECE); 20–22 December 2012; Dhaka, Bangladesh. pp. 884-887.
- [20] Sanam TF, Shahnaz C. Teager energy operation on wavelet packet coefficients for enhancing noisy speech using a hard thresholding function. *Signal Processing: An International Journal* 2012; 6: 22-35.
- [21] Islam MT, Shahnaz C, Zhu WP, Ahmad MO. Speech enhancement based on student t modeling of teager energy operated perceptual wavelet packet coefficients and a custom thresholding function. *IEEE/ACM T Audio Speech Lang Process* 2015; 23: 1800-1811.
- [22] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE/ACM T Audio Speech Lang Process* 1985; 33: 443-445.
- [23] Zao L, Coelho R, Flandrin P. Speech enhancement with emd and hurst-based mode selection. *IEEE/ACM T Audio Speech Lang Process* 2014; 22: 899-911.
- [24] Nsabimana FX, Subbaraman V, Zölzer U. A single channel speech enhancement technique exploiting human auditory masking properties. *Adv Radio Sci* 2010; 8: 95-99.
- [25] Soon Y, Koh SN, Yeo CK. Improved noise suppression filter using self-adaptive estimator of probability of speech absence. *Signal Process* 1999; 75: 151-159.
- [26] Kim NS, Chang JH. Spectral enhancement based on global soft decision. *IEEE Signal Process Lett* 2000; 7: 108-110.
- [27] Khaldi K. Processing and analysis of sounds signals by Huang transform (empirical mode decomposition: EMD). PhD, Télécom Bretagne, Université de Bretagne Occidentale, Brest, France, 2012.
- [28] Kaiser JF. Some useful properties of Teager’s energy operators. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*; 27–30 April 1993; Minneapolis, MN, USA. pp. 149-152.
- [29] Maragos P, Quatieri TF, Kaiser JF. Speech nonlinearities, modulations, and energy operators. In: *International Conference on Acoustics, Speech, and Signal Processing*; 14–17 April 1991; Toronto, Canada. pp. 421-424.

- [30] Bahoura M, Rouat J. Wavelet speech enhancement based on the Teager energy operator. *IEEE Signal Process Lett* 2001; 8: 10-12.
- [31] Islam MT, Shahnaz C, Zhu WP, Ahmad MO. Rayleigh modeling of Teager energy operated perceptual wavelet packet coefficients for enhancing noisy speech. *Speech Commun* 2017; 86: 64-74.
- [32] Joyce JM. Kullback-Leibler divergence. In: *International Encyclopedia of Statistical Science*. Heidelberg, Germany: Springer, 2011. pp. 720-722.
- [33] Hershey JR, Olsen PA. Approximating the Kullback Leibler divergence between Gaussian mixture models. In: *2007 IEEE International Conference on Acoustics, Speech, and Signal Processing*; 15–20 April 2007; Honolulu, HI, USA. pp. 317-320.
- [34] Erkelens JS, Hendriks RC, Heusdens R, Jensen J. Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors. *IEEE T Audio Speech Lang Process* 2007; 15: 1741-1752.
- [35] Sun P, Qin J. Wavelet packet transform based speech enhancement via two-dimensional SPP estimator with generalized gamma priors. *Arch Acoust* 2016; 41: 579-590.
- [36] Rudresh MD, Rajeshwari KM, Sujatha S, Suresh M. EMD based speech enhancement using soft and hard threshold techniques. *Int J Res Eng Technol* 2016; 5: 534-541.
- [37] Deger E, Molla MKI, Hirose K, Minematsu N, Hasan MK. Speech enhancement using soft thresholding with DCT-EMD based hybrid algorithm. In: *15th European Signal Processing Conference*; 3–7 September 2007; Poznan, Poland. pp. 75-79.
- [38] Loizou PC. *Speech Enhancement: Theory and Practice*. Boca Raton, FL, USA: CRC Press, 2007. pp. 503-505.
- [39] ITU. *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs*. Rec. ITU-T P. 862. Geneva, Switzerland: ITU, 2001.
- [40] Hu Y, Loizou PC. Evaluation of objective quality measures for speech enhancement. *IEEE T Audio Speech Lang Process* 2008; 16: 229-238.
- [41] Varga A, Steeneken HJ. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Commun* 1993; 12: 247-251.