

# Teager Energy Operation on Wavelet Packet Coefficients for Enhancing Noisy Speech Using a Hard Thresholding Function

**Tahsina Farah Sanam**

*Institute of Appropriate Technology  
Bangladesh University of Engineering and Technology  
Dhaka, Bangladesh*

*tahsina@iat.buet.ac.bd*

**Celia Shahnaz**

*Department of Electrical and Electronic Engineering  
Bangladesh University of Engineering and Technology  
Dhaka, Bangladesh*

*celia@eee.buet.ac.bd*

---

## Abstract

In this paper a new thresholding based speech enhancement approach is presented, where the threshold is statistically determined by employing the Teager energy operation on the Wavelet Packet (WP) coefficients of noisy speech. The threshold thus obtained is applied on the WP coefficients of the noisy speech by using a hard thresholding function in order to obtain an enhanced speech. Detailed simulations are carried out in the presence of white, car, pink, and babble noises to evaluate the performance of the proposed method. Standard objective measures, spectrogram representations and subjective listening tests show that the proposed method outperforms the existing state-of-the-art thresholding based speech enhancement approaches for noisy speech from high to low levels of SNR.

**Keywords:** Teager Energy Operator, Wavelet Packet Transform, Statistical Modeling, Thresholding Function

---

## 1. INTRODUCTION

Estimating a signal that is corrupted by additive noise has been of interest to many researchers for practical as well as theoretical reasons. The problem is to recover the original signal from the noisy data. We want the recovered signal to be as close as possible to the original signal, retaining most of its important properties. There has been an increasing interest in noisy speech enhancement in a broad range of speech communication applications, such as mobile telephony, speech coding and recognition, and hearing aid devices [1]-[5]. Since the presence of noise seriously degrades the performance of the systems in such applications, the efficacy of the systems operating in a noisy environment is highly dependent on the speech enhancement techniques employed therein.

Various speech enhancement methods have been reported in the literature describing the know how to solve the problem of noise reduction in the speech enhancement methods. Speech enhancement methods can be generally divided into several categories based on their domains of operation, namely time domain, frequency domain and time-frequency domain. Time domain methods includes the subspace approach [6]-[10], frequency domain methods includes speech enhancement methods based on discrete cosine transform [11], the spectral subtraction [12]-[16], minimum mean square error (MMSE) estimator [17]-[21], Wiener filtering [22]-[25] and time frequency-domain methods involve the employment of the family of wavelet [26]-[34]. All the methods have their own advantages and drawbacks. In the subspace method [6]-[10], a mechanism to obtain a tradeoff between speech distortion and residual noise is proposed with the cost of a heavy computational load. Frequency domain methods, on the other hand, usually need less computation. In particular, although spectral subtraction method [12]-[16], is simple and provides a tradeoff between speech distortion and residual noise to some extent, it suffers from an artifact known as "musical noise" having an unnatural structure that is perceptually annoying, composed of tones at random frequencies and has an increased variance. In the MMSE estimator [17]-[21], the frequency spectrum of the noisy speech is modified to reduce the noise from noisy speech in the frequency domain. A relatively large variance of spectral coefficients is the problem of such an estimator. While adapting filter gains of the MMSE estimator, spectral outliers may emerge, that is especially difficult to avoid under noisy conditions.

One of the major problems of Wiener filter based methods [22]-[25] is the requirement of obtaining clean speech statistics necessary for their implementation. Both the MMSE and the Wiener estimators have a moderate computation load, but they offer no mechanism to control tradeoff between speech distortion and residual noise. Among the methods using time-frequency analyses, an approach of reducing different types of noise that corrupt the clean speech is the use of nonlinear techniques based on Discrete Wavelet Transform (DWT) [26]-[34], which is a superior alternative to the analyses based on Short Time Fourier Transform (STFT). The main challenge in such denoising approaches based on the thresholding of the wavelet coefficients of the noisy speech is the estimation of a threshold value that marks a difference between the wavelet coefficients of noise and that of clean speech. Then, by using the threshold, the designing of a thresholding scheme to minimize the effect of wavelet coefficients corresponding to the noise is another difficult task considering the fact that the conventional DWT based denoising approaches exhibit a satisfactory performance only at a relatively high signal-to-noise ratio (SNR).

For zero-mean, normally distributed white noise, Donoho and Johnstone proposed Universal threshold based method [30] for denoising the corrupted speech. For noisy speech, applying a unique threshold for all the wavelet or WP coefficients irrespective of the speech and silence segments may suppress noise to some extent, but it may also remove unvoiced speech segments thus degrading the quality of the enhanced speech. Statistical modeling is another approach of thresholding [28], where the threshold of WP coefficients is determined using the similarity distances between the probability distributions of the signals. Since speech is not always present in the signal, the thresholding must be adapted over time so that it is larger during portions without speech and smaller for those with speech. This will eliminate as much of the noise as possible while still maintaining speech intelligibility. However, the method in [28] requires an estimate of noise variance to distinguish speech frames from that of the noise ones with a view to set different thresholds for them. In order to decide a time adaptive threshold considering speech or silence frame, some estimate of the signal energy over time is necessary. A popular technique to estimate the required energy of the signal is Teager Energy Operator [35]. In [29], Teager energy operator (TEO) proposed by Kaiser [35] is employed to compute a time-adaptive threshold (TAT) value to threshold the WP coefficients of the noisy speech. But, TAT method suffers from an over thresholding problem if the speech signal is just contaminated by slight noises as this method uses an absolute offset parameter to distinguish speech frames from that of the noise ones.

In this paper, we develop a new thresholding method in the wavelet packet domain, where the threshold is adapted with respect to speech and silent segments. Since, TEO is a popular way to estimate the speech signal energy, instead of direct employment of the TEO on the noisy speech, we apply the TEO on the WP coefficients of the noisy speech. Unlike the approach of threshold determination directly from the WP coefficients of the noisy speech, we determine an appropriate threshold by performing the statistical modeling of the TE operated WP coefficients of noisy speech and employed a hard thresholding function for obtaining an enhanced speech.

## 2. BRIEF BACKGROUND

### 2.1. Wavelet Packet Transform

A method based on the Wavelet Packet Transform is a generalization of the Wavelet Transform based decomposition process that offers a richer range of probabilities for the analysis of signals, namely speech. In wavelet analysis, a speech signal is split into sets of approximation and detail coefficients. The set of approximation coefficients is then itself split into a second-level approximation and detail coefficients, and the process is repeated. Mallat algorithm is one of the efficient ways to construct the discrete wavelet transform (DWT) by iterating a two-channel perfect reconstruction filter bank over the low pass scaling function branch. However, this algorithm results in a logarithmic frequency resolution, which does not work well for all the signals. In order to overcome the drawback as mentioned above, it is desirable to iterate the high pass wavelet branch of the Mallat algorithm tree as well as the low pass scaling function branch. Such a wavelet decomposition produced by these arbitrary subband trees is known as wavelet packet (WP) decomposition.

In wavelet analysis, only scale space is decomposed, but wavelet space is not decomposed. By the restriction of Heisenberg's uncertainty principle, the spatial resolution and spectral resolution of high frequency band become poor thus limiting the application of wavelet transform. In particular, there are some problems with the basic wavelet thresholding method, when it is applied to the noisy speech for

the purpose of enhancement. An important shortcoming is the shrinkage of the unvoiced segments of speech which contain many noise-like speech components leading to a degraded speech quality. On the other hand, in wavelet packet analysis, the wavelet space is also decomposed thus making the higher frequency band decomposition possible. Since, both the approximation and the detail coefficients are decomposed into two parts at each level of decomposition, a complete binary tree with superior frequency localization can be achieved. This particular feature of the WP transform is indeed useful for enhancing speech in the presence of noise.

### 2.2. Teager Energy Operator

The Teager Energy Operator (TEO) is a powerful nonlinear operator proposed by Kaiser [36], capable to extract the signal energy based on mechanical and physical considerations. The continuous form of the TEO is given as,

$$\Psi_c[y(t)] = \left(\frac{d}{dt}y(t)\right)^2 - y(t)\frac{d^2}{dt^2}y(t), \tag{1}$$

where,  $\Psi_c[\cdot]$  and  $y(t)$  represent the continuous TEO and a continuous signal, respectively. For a given bandlimited discrete signal  $y[n]$ , the discrete-time TEO can be approximated by,

$$\Psi_d(y[n]) = y[n]^2 - y[n+1]y[n-1], \tag{2}$$

The discrete time TEO in (2) is nearly instantaneous since only three samples are required for the energy computation at each time instant. Due to this excellent time resolution, the output of a TEO provides us with the ability to capture the energy fluctuations and hence gives an estimate of the energy required to generate the signal. Note that, in case of speech signal, directly using the TEO on original speech may result in much undesired artifact and enhanced noises as TEO is a fixed-sized local operator.

In context of the speech enhancement by thresholding via WP analysis, the threshold must be adapted over time, since speech is not always present in the signal. It is expected that the threshold should be larger during periods without speech and smaller for those with speech. In order to obtain an idea of speech/nonspeech activity for deciding the corresponding threshold value, it is required to estimate the signal energy over time. Since TEO is a popular way to estimate the speech signal energy, instead of direct employment of the TEO on the original speech, it is reasonable to apply the TEO on the WP coefficients. In comparison to the approach of threshold determination from the WP coefficients of noisy speech, the approach intended to determine threshold from the TE operated WP coefficients has the potential to eliminate as much of the noise as possible, while still maintaining speech intelligibility in enhanced speech.

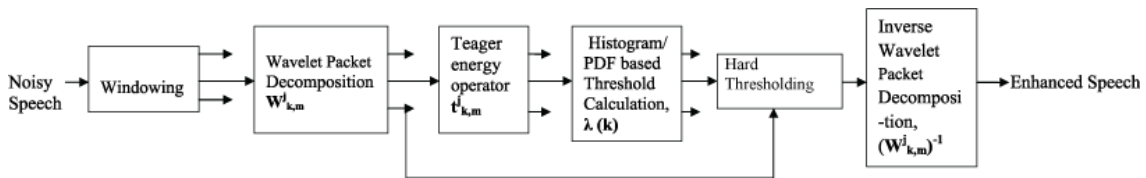


FIGURE 1: Block Diagram of the Proposed Method

## 3. PROPOSED METHOD

The block diagram of our proposed system is shown in Fig. 1. It is seen from Fig. 1 that WP transform is first applied to each input speech frame. Then, the WP coefficients are subject to Teager Energy approximation with a view to determine a threshold value for performing thresholding operation in the WP domain. On thresholding, an enhanced speech frame is obtained via inverse wavelet packet (IWP) transform.

### 3.1. Wavelet Packet Analysis

For a  $j$  level WP transform, the noisy speech signal  $y[n]$  with frame length  $N$  is decomposed into  $2^j$  subbands. The  $m$ -th WP coefficient of the  $k$ -th subband is expressed as,

$$W_{k,m}^j = WP[y[n], j], \quad n = 1, \dots, N. \tag{3}$$

where,  $m = 1, \dots, N/2^j$  and  $k = 1, \dots, 2^j$ . Since the frequency resolution provided by the full WP transform is not sufficient to separate speeches from low-frequency noises, WP coefficients of the noisy speech have some noise handling limitations. To this end, we apply the discrete time TEO on the obtained WP coefficients,  $W_{k,m}^j$ .

### 3.2. Teager Energy (TE) Approximation

The application of the discrete-time TEO on the  $W_{k,m}^j$  results in a set of TEO coefficients  $t_{k,m}^j$ . The  $m$ -th TEO coefficient corresponding to the  $k$ -th subband of the WP is given by,

$$t_{k,m}^j = \Psi_d[W_{k,m}^j], \quad k = 1, \dots, 2^j \quad (4)$$

i.e; using (2)

$$t_{k,m}^j = [W_{k,m}^j]^2 - [W_{k,(m-1)}^j][W_{k,(m+1)}^j], \quad k = 1, \dots, 2^j \quad (5)$$

In comparison to the operation of WP transform on the noisy speech, the TEO operation on the WP coefficients of the noisy speech is able to enhance the discriminability of speech coefficients among those of noise. This energy tracking operator can be understood when considering sinusoidal oscillation that occur with a simple harmonic oscillator. If TEO is applied to a discrete time signal, it causes an effect as if a single sinusoid of amplitude  $A$  and frequency  $\omega$  is passed through three adjacent points with index  $n - 1, n, n + 1$  thus yielding an output sequence that is a varying signal proportional to  $A^2 \sin^2 \omega$ , where the frequency  $\omega$  is normalized with respect to the sampling frequency. This signal in essence is a measure of the 'energy' in that signal as a function of time. It is thus important that the original discrete signal consist primarily of a single component. Since a wavelet coefficient stream in a single subband is primarily a single component signal, it is valid to apply TEO on the WP coefficients of the noisy speech.

### 3.3. Statistical Modeling of TEO Operated WP Coefficients

The outcome of a speech denoising method based on the thresholding in a transform domain depends mainly on two factors, namely the threshold value and the thresholding functions. The use of a unique threshold for all the WP subbands is not reasonable. As a crucial parameter, the threshold value in each subband is required to be adjusted very precisely so that it can prevent distortion in the enhanced speech as well as decrease annoying residual noise. In order to remove the noisy coefficients with low distortion in the enhanced speech signal, the value of threshold has to be different in the speech and silent frames. The value of the threshold in the silent frames is smaller than it in the speech frames. Also, the use of conventional thresholding functions, for example, Hard and Soft thresholding functions often results in time frequency discontinuities. In order to handle such problems, we propose a new thresholding function employing a threshold value determined for each subband of the WP by statistically modeling the TE operated WP coefficients  $t_{k,m}^j$  with a probability distribution rather than choosing a threshold value directly from the  $t_{k,m}^j$ . This idea is also exploited to determine a threshold value for each subband of an silent frame which is different from that of each subband of a speech frame.

The main issue in wavelet thresholding is estimating an appropriate threshold value  $\lambda$ . In the range of  $-\lambda$  and  $\lambda$ , the noisy speech wavelet coefficients are similar to the noise wavelet coefficients and outside of this range the wavelet coefficients of the noisy speech are similar to that of the clean speech [38]. So it is expected that in the range of  $-\lambda$  and  $\lambda$ , [30],[26],[37], the probability distribution of the noisy speech coefficients would be nearly similar to that of the noise coefficients. Furthermore, the probability distribution of the noisy speech coefficients is expected to be similar to that of the clean speech coefficients outside of this range. In a certain range, the probability distribution of the  $t_{k,m}^j$  of the noisy speech is expected to be nearly similar to those of the noise. Also, outside that range, the probability distribution of the  $t_{k,m}^j$  of the noisy speech is expected to be similar to those of the clean speech. Thus by considering the probability distributions of the  $t_{k,m}^j$  of the noisy speech, noise and clean speech, a more accurate threshold value can be obtained using a suitable scheme of pattern matching or similarity measure between the probability distributions. Since speech is a time-varying signal, it is difficult to realize the actual probability distribution function (pdf) of

speech or its  $t_{k,m}^j$ . As an alternative to formulate a pdf of the  $t_{k,m}^j$  of speech, we can easily formulate the histogram of the  $t_{k,m}^j$  and approximate the histogram by a reasonably close probability distribution function, namely Gaussian distribution [28]. In frequency domain, the use of a Gaussian statistical model is motivated by the central limit theorem since each Fourier expansion coefficient can be seen as a weighted sum of random variables resulting from the observed samples [39]. Other distributions were also proposed for the real and imaginary parts of the STFT coefficients [40], [24], the STSA coefficients [40], and the complex STFT coefficients [42], [17]. While it has been proposed that the Fourier expansion coefficients of speech signals may not be Gaussian-distributed, those assumptions are usually motivated by long-term averages of the speech signal which may not be applicable to specific short-time utterances. Moreover, many estimators using a Gaussian distribution do not have an analytical counterpart when using other distributions [41]. Therefore, many researchers consider only Gaussian distributed complex STFT coefficients in their works [43]. However, in our work, we have implemented our algorithm for 30 sentences of the NOIZEUS database and 4 different noise signals (white, car, pink, and babler noises) and it is verified that, in each subband, the pdfs of the  $t_{k,m}^j$  of the noise, clean speech and noisy speech can be sufficiently well described by the Gaussian distribution. Fig. 2, Fig. 3, and Fig. 4 shows instances of these results for clean speech, noisy speech, and noise, respectively.

### 3.4. Adaptive Threshold Calculation

Analysis on the speech signal shows that, the value of entropy has the ability to detect speech/silence frames [38], [45]. Also, the entropy of each subband of the  $t_{k,m}^j$  is found different from each other. So, an entropy measure may be chosen to select a suitable threshold value adaptive to each subband as well as adaptive to the speech/silence frames. Some popular similarity measures that are related to the entropy functions are the Variational distance, the Bhattacharyya distance, the Harmonic mean, the Kullback Leibler(K-L) divergence, and the Symmetric K-L divergence [47]. All these measures are used to estimate the similarity between two pdfs. As all of these distance measures have nonnegative values, zero flag is a very suitable distinctive for recognizing the similarity between two pdfs. Note that, if two pdfs are exactly the same, only two measures (the K-L divergence and the symmetric K-L divergence) will be equal to zero. So, in order to determine an adaptive threshold value based on the idea of entropy quantified by an appropriate similarity measure, we proceed as follows,

1. The average of the  $t_{k,m}^j$  of different segments is calculated.
2. The histograms of the averaged  $t_{k,m}^j$  in each sub-band is obtained. The number of bins in the histogram has been set equal to the square root of the number of samples divided by two.
3. Since  $t_{k,m}^j$  of clean speech, noisy speech and noise are positive quantity, there histograms in each sub-band can be approximated by the positive part of a pdf following the Gaussian distribution as shown in Fig. 2, 3 and 4.

The K-L divergences is always nonnegative and zero if and only if the approximate Gaussian distribution functions of the  $t_{k,m}^j$  of noisy speech and that of the noise or the approximate Gaussian distribution functions of the  $t_{k,m}^j$  of the noisy speech and that of the clean speech are exactly the same. In order to have a symmetric distance between the any two approximate Gaussian distribution functions as mentioned above, the Symmetric K-L divergence has been adopted in this paper. The Symmetric K-L divergence is defined as,

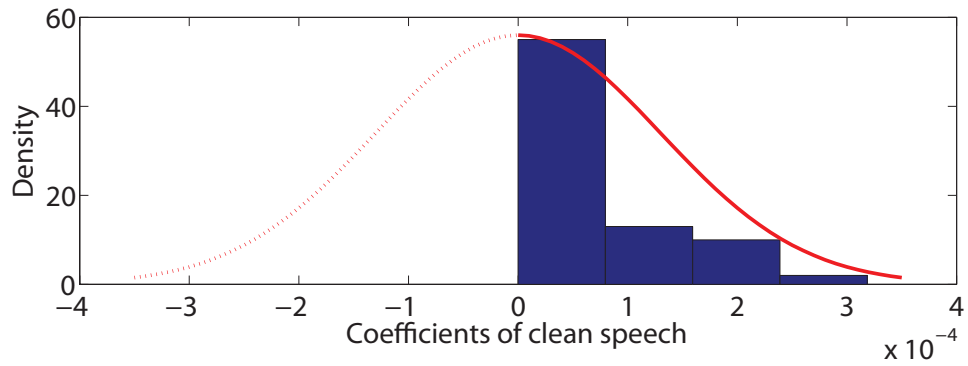
$$SKL(p, q) = \frac{KL(p, q) + KL(q, p)}{2}, \tag{6}$$

where,  $p$  and  $q$  are the two approximate Gaussian pdfs calculated from the corresponding histograms each having  $M$  number of bins and  $KL(\cdot)$  is the K-L divergence given by,

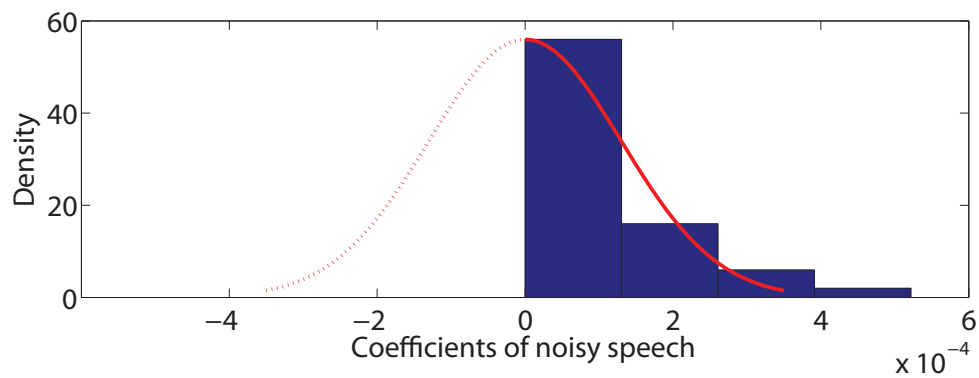
$$KL(p, q) = \sum_{i=1}^M p_i(t_{k,m}^j) \ln \frac{p_i(t_{k,m}^j)}{q_i(t_{k,m}^j)} \tag{7}$$

In (7),  $p_i(t_{k,m}^j)$  represents the approximate Gaussian pdf of the  $t_{k,m}^j$  of the noisy speech estimated by,

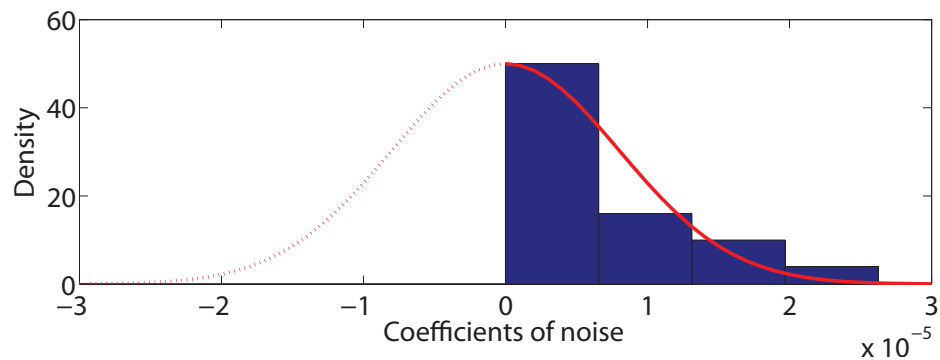
$$\hat{p}_i(t_{k,m}^j) = \frac{\text{Number of coefficients in the } i\text{th bin of histogram}}{\text{Total number of coefficients in each subband}} \tag{8}$$



**FIGURE 2:** Probability distribution of TEO coefficients of clean speech



**FIGURE 3:** Probability distribution of TEO coefficients of noisy speech



**FIGURE 4:** Probability distribution of TEO coefficients of noise

Similarly, the approximate Gaussian pdf of the  $t_{k,m}^n$  of the noise and that of the  $t_{k,m}^c$  of the clean speech can be estimated from (8) and denoted by  $\hat{q}_i(t_{k,m}^n)$  and  $\hat{q}_i(t_{k,m}^c)$ , respectively. Below a certain value of threshold  $\lambda$ , the  $t_{k,m}^n$  of the noisy speech, the Symmetric K-L divergence between  $\hat{q}_i(t_{k,m}^n)$  and  $\hat{q}_i(t_{k,m}^c)$  is approximately zero, i.e,

$$SKL(\hat{q}_i(t_{k,m}^n), \hat{q}_i(t_{k,m}^c)) \approx 0, \tag{9}$$

where the bins lie in the range  $[1, \lambda]$  in both  $\hat{p}_i(t_{k,m}^j)$  and  $\hat{q}_i(t_{k,m}^j)$ . Alternatively, above the value  $\lambda$  of the  $t_{k,m}^j$  of the noisy speech, the Symmetric K-L divergence between  $\hat{p}_i(t_{k,m}^j)$  and  $\hat{q}_i(t_{k,m}^j)$  is closely zero, i.e.,

$$SKL(\hat{p}_i(t_{k,m}^j), \hat{q}_i(t_{k,m}^j)) \approx 0, \quad (10)$$

In (10), the bins lie in the range  $[\lambda+1, M]$  in both  $\hat{p}_i(t_{k,m}^j)$  and  $\hat{q}_i(t_{k,m}^j)$ . Using (6) and (7) in evaluating (9) and (10), we get,

$$\sum_{i=1}^{\lambda} [\hat{p}_i(t_{k,m}^j) - \hat{q}_i(t_{k,m}^j)] \ln \frac{\hat{p}_i(t_{k,m}^j)}{\hat{q}_i(t_{k,m}^j)} \approx 0 \quad (11)$$

$$\sum_{i=\lambda+1}^M [\hat{p}_i(t_{k,m}^j) - \hat{q}_i(t_{k,m}^j)] \ln \frac{\hat{p}_i(t_{k,m}^j)}{\hat{q}_i(t_{k,m}^j)} \approx 0 \quad (12)$$

From (11), it is apparent that  $t_{k,m}^j$  of the noisy speech lying in the range  $[1, \lambda]$  can be marked as  $t_{k,m}^j$  of noise and needed to be removed. Similarly, (12) attests that the  $t_{k,m}^j$  of the noisy speech residing outside  $[1, \lambda]$  can be treated as similar to the  $t_{k,m}^j$  of the clean speech and considered to be preserved. For obtaining a general formula for the threshold value  $\lambda$  in each subband, we use continuous real mode in (11) and (12), thus obtain,

$$\int_1^{\lambda} \left[ \frac{\sqrt{\vartheta}}{\sqrt{2\pi}\sigma_S} \exp\left(-\frac{\vartheta x^2}{2\sigma_S^2}\right) - \frac{1}{2\pi\sigma_N} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) \right] \ln \left( (1 - \sqrt{\vartheta}) \exp\left(-\frac{\vartheta x^2}{2\sigma_S^2} + \frac{x^2}{2\sigma_N^2}\right) \right) dx \approx 0 \quad (13)$$

$$\int_{\lambda+1}^{\infty} \left[ \frac{\sqrt{\vartheta}}{\sqrt{2\pi}\sigma_S} \exp\left(-\frac{\vartheta x^2}{2\sigma_S^2}\right) - \frac{1}{2\pi\sigma_S} \exp\left(-\frac{x^2}{2\sigma_S^2}\right) \right] \ln \left( (\sqrt{\vartheta}) \exp\left(\frac{(1 - \vartheta)x^2}{2\sigma_S^2}\right) \right) dx \approx 0 \quad (14)$$

where,

$$\vartheta = \sigma_S^2 / (\sigma_N^2 + \sigma_S^2) \quad (15)$$

where,  $\vartheta = \sigma_S^2 / (\sigma_N^2 + \sigma_S^2)$ ,  $\sigma_N$  [44], [46] and  $\sigma_S$  are the variances of noise and clean speech in each subband, respectively. For computing  $\lambda$ , we first simplify the equations (13) and (14) to solve. Since the symmetric K-L is a nonnegative distance, in a specified range, its minimum value can be found to be nearly zero. Thus the value of  $t_{k,m}^j$  for which the threshold reaches its optimum value can be determined by minimizing (13) or (14). It is well known that an optimum value of a function in a given range can be calculated by setting its derivative, with respect to the variable expected to optimize the function value, to zero. Since (13) is a definite integral, the derivative of the function defined in the left hand side (L.H.S) of (13) representing the Symmetric K-L divergence between  $\hat{p}_i(t_{k,m}^j)$  and  $\hat{q}_i(t_{k,m}^j)$  is zero. On the other hand, the derivative of the function obtained in the L.H.S of (14) representing the Symmetric K-L distance between  $\hat{p}_i(t_{k,m}^j)$  and  $\hat{q}_i(t_{k,m}^j)$  is calculated and set to zero. By simplifying the either derivatives, an optimum value of  $\lambda$  can be obtained which is adaptive to each subband of a frame.

$$\lambda(k) = \sigma_N(k) \sqrt{2(\gamma_k + \gamma_k^2) \ln \left( \sqrt{1 + \frac{1}{\gamma_k}} \right)} \quad (16)$$

where,  $\sigma_N$  is the variance of noise in each subband],  $k$  is the sub-band index and  $\gamma_k$  is the segmental SNR calculated as,

$$\gamma_k = \sigma_S^2(k) / \sigma_N^2(k) \quad (17)$$

We calculate the second order derivation of the L.H.S of (14) with respect to the obtained threshold to demonstrate that the calculated threshold minimize (14). As the second order derivation of the L.H.S of (14) is nonnegative, the obtained thresholds are valid. In order to have smaller threshold for higher input SNR values, we have to adjust the threshold obtained by (16). Since the variance of the noise is inversely proportional to the input SNR, we can modify (16) as,

$$\lambda(k) = \left[ \sigma_N(k) / \sqrt{\gamma_k} \right] \sqrt{2(\gamma_k + \gamma_k^2) \ln \left( \sqrt{1 + \frac{1}{\gamma_k}} \right)} \quad (18)$$

Since in the silent segment of a noisy speech, only noise exists, a threshold value different than that used in the speech segment should be selected in order to eliminate the noise completely. The Symmetric K-L divergence between the  $t_{k,m}^j$  of the noisy speech and that of the  $t_{k,m}^j$  of the noise is nearly zero in the non speech subbands. Exploiting this idea a suitable time adaptive threshold value  $\lambda'$  can be obtained as,

$$\lambda'(k) = \begin{cases} \max(t_{k,m}^j), & \text{SKL}(\hat{p}_i(t_{k,m}^j), \hat{q}_i(t_{k,m}^j)) \approx 0 \\ \lambda(k), & \text{Otherwise.} \end{cases} \quad (19)$$

### 3.5. Denoising by Thresholding

Removing noise components by thresholding operation of the WP coefficients is based on the fact that for many signals (such as speech), the energy is mostly concentrated in a low frequency region that corresponds to a small number of lower WP coefficients. So, by thresholding the WP coefficients, we can reduce the effect of the high frequency noise components on the speech signal components. We employ hard thresholding for denoising purpose.

Hard thresholding sets zero to the noisy speech WP coefficients whose absolute value is below the threshold. Noting the threshold determined by (19) as  $\lambda_1(k)$  and using it, the hard thresholding function can be applied on the  $m$ -th WP coefficients of the  $k$ -th subband  $V_{k,m}^j$  as,

$$\hat{V}_{k,m}^j = \begin{cases} V_{k,m}^j & |V_{k,m}^j| \geq \lambda_1(k) \\ 0, & |V_{k,m}^j| < \lambda_1(k). \end{cases} \quad (20)$$

Here,  $\hat{V}_{k,m}^j$  stands for the  $m$ -th WP coefficients of the  $k$ -th subband after the hard thresholding operation.

### 3.6. Inverse Wavelet Transform

The enhanced speech frame is synthesized by performing the inverse WP transformation  $WP^{-1}$  on the resulting thresholded WP coefficients,  $\hat{V}_{k,m}^j$

$$\hat{S}[n] = WP^{-1}(\hat{V}_{k,m}^j) \quad (21)$$

where,  $\hat{S}[n]$  represents the enhanced speech frame. The final enhanced speech signal is reconstructed by using the standard overlap-and-add method.

## 4. SIMULATION

In this Section, a number of simulations is carried out to evaluate the performance of the proposed method.

### 4.1. Simulation Conditions

Real speech sentences from the *NOIZEUS* database are employed for the experiments, where the speech data is sampled at 8 KHz. To imitate a noisy environment, noise sequence is added to the clean speech samples at different signal to noise ratio (SNR) levels ranging from 15 dB to -15 dB. Four different types of noises, such as, white, car, and pink are adopted from the *NOISEX92* [20] and *NOIZEUS* databases.

In order to obtain overlapping analysis frames, hamming windowing operation is performed, where the size of each of the frame is 512 samples with 50% overlap between successive frames. A 3-level WP decomposition tree with db10 bases function is applied on the noisy speech frames and the Teager energy operation is performed on the resulting WP coefficients. By computing the threshold from (19), a hard thresholding function is developed and applied on the WP coefficients of the noisy speech using (20).



#### 4.2. Comparison Metrics

Standard Objective metrics [48], namely, overall SNR improvement in dB, Perceptual Evaluation of Speech Quality (PESQ) and Weighted Spectral Slope (WSS) are used for the evaluation of the proposed method. The proposed method is subjectively evaluated in terms of the spectrogram representations of the clean speech, noisy speech and enhanced speech. Informal listening tests are also carried out in order to find the analogy between the objective metrics and subjective sound quality. The performance of our method is compared with some of the existing thresholding based speech enhancement methods, such as, Universal [30], WTHSKL [29] and TAT [28] in both objective and subjective senses.

#### 4.3. Objective Evaluation

##### 4.3.1. Results on White Noise-corrupted Speech

The results in terms of all the objective metrics, such as, SNR improvement in dB, PESQ and WSS obtained by using the Universal, WTHSKL, TAT, and proposed methods for white noise-corrupted speech are presented in Fig. 5 through Fig. 6 and in Table 1.

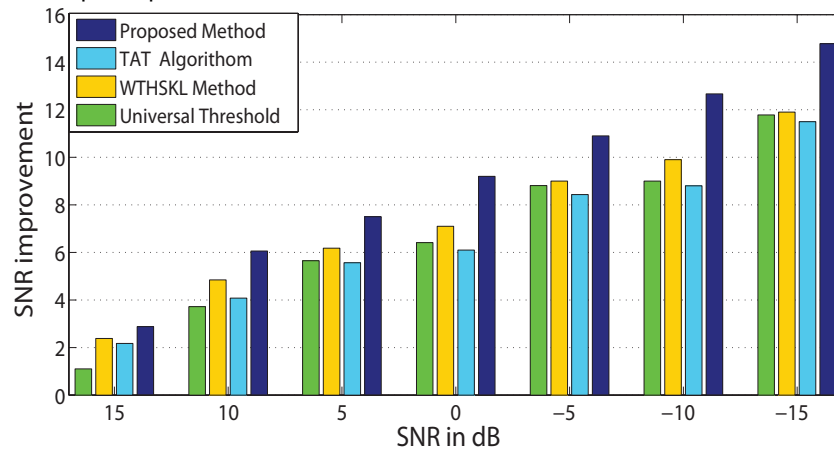
Fig. 5 shows the SNR improvement in dB obtained by using different methods employing hard thresholding function in the presence of white noise, where the SNR varies from 15 dB to -15 dB . It is seen from this figure that in the SNR range under consideration, the improvement in SNR in dB is comparable for all the comparison methods, but they show comparatively lower values relative to the proposed method at all the levels of SNR.

The PESQ scores vs SNR obtained by using different methods are portrayed in Fig. 6. This figure shows that the proposed method using the hard thresholding function is capable of producing enhanced speech with better quality as it gives larger scores of PESQ for a wide range of SNR levels whereas, the PESQ scores resulting from all other methods are comparable and relatively lower even at a high SNR of 15 dB. It is also seen from Fig. 6 that the difference in PESQ scores of the proposed method and that of the other methods increases as SNR decreases, thus indicating the effectiveness of the proposed method using hard thresholding function in enhancing speech even in a severe noisy environment.

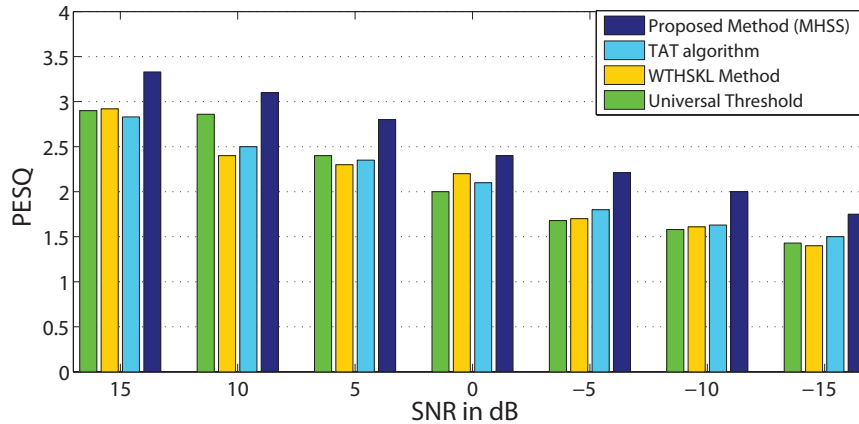
The WSS values obtained by using different methods are summarized in Table 1. for varying SNR of 15 dB to -15 dB. For a particular method in Table 1, the WSS increases as SNR decreases. At a particular SNR, such as -15 dB, the proposed method using hard function is superior in a sense that it gives the lowest WSS value, whereas the other methods produce comparatively higher values of WSS.

##### 4.3.2. Results on Car Noise-corrupted Speech

Now, we present the results in terms of all the objective metrics as mentioned above obtained by using the Universal, WTHSKL, TAT, and the proposed methods in Table 2 and in Fig. 7 through Fig. 8 for car noise-corrupted speech.



**FIGURE 5:** Performance comparison of different methods using hard thresholding function in terms of SNR Improvement in dB for white noise corrupted speech.



**FIGURE 6:** Performance comparison of different methods using hard thresholding function in terms of PESQ for white noise corrupted speech.

SNR [dB]	Universal	TAT	WTHSKL	Proposed Method
15	34.84	27.2115	27.1	22.7
10	42.1	34.5904	33.2337	29.1
5	51.3	43.2807	42.7895	37.3
0	62.5	55.1436	56.4079	46.82
-5	73.13	67.1598	66.8581	59.1
-10	90	85.43	86.1	78.1
-15	100.6	100.63	92.77	84.8

**TABLE 1:** Performane comparison of WSS for different methods in the presence of white noise.

In Table 2, the performance of the proposed method using hard thresholding function is compared with that of the other methods at different levels of SNR. For a method in Table 2, the SNR improvement in dB increases as SNR decreases. At a low SNR of -15 dB, the proposed method yields the highest SNR improvement in dB. Such larger values of SNR improvement in dB at a low level of SNR attest the capability of the proposed method in producing the enhanced speech with better quality even for car noise-corrupted speech.

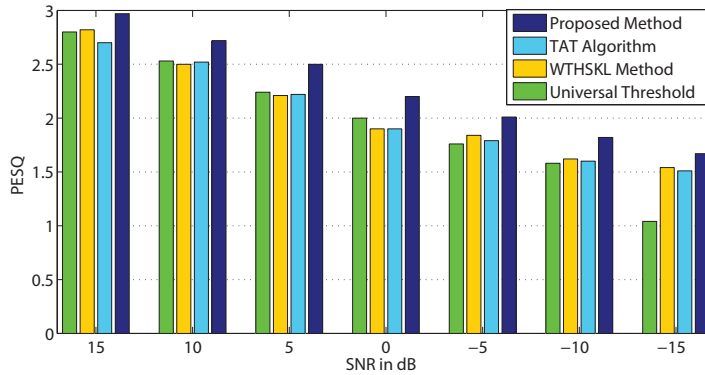
SNR [dB]	Universal	TAT	WTHSKL	Proposed Method
15	1.9	3.4	0.6	3
10	3.8	4.01	2.9	4.51
5	4.9	5	4	6.8
0	6.7	5.7	5.5	8.1
-5	7	7.23	8.1	9.14
-10	8.9	9.6	9.2	10.78
-15	10.1	10.63	11.6	12.7

**TABLE 2:** Performance comparison of SNR improvement in dB for different methods in the presence of car noise.

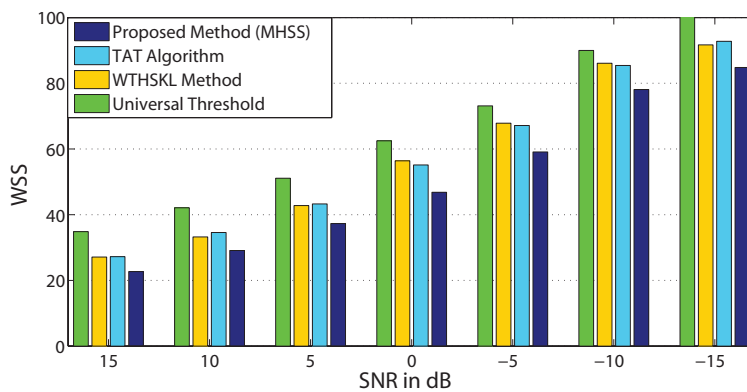
In the presence of car noise, the PESQ scores at different SNR levels resulted by using the other methods are compared with respect to the proposed method employing hard thresholding function in Fig. 7. It can be seen from the figure that at a high level of SNR, such as 15 dB, Universal, WTHSKL and TAT methods show lower values of PESQ scores, whereas the PESQ score is much higher, as expected, for the proposed method. The proposed method also yields larger PESQ scores compared to that of the other methods at lower levels of SNR. Since, at a particular SNR, a higher PESQ score

indicates a better speech quality, the proposed method is indeed better in performance even in the presence of a car noise.

Fig. 8 represents the WSS values as a function of SNR for the proposed method employing hard thresholding function and that for the other methods. As shown in the figure that the WSS values resulting from all other methods are comparable and relatively larger for a wide range of SNR levels, whereas the proposed method is capable of producing enhanced speech with better quality as it gives lower values of WSS at a low SNR of -15 dB.



**FIGURE 7:** Performance comparison of different methods using hard thresholding function in terms of PESQ scores for car noise corrupted speech.

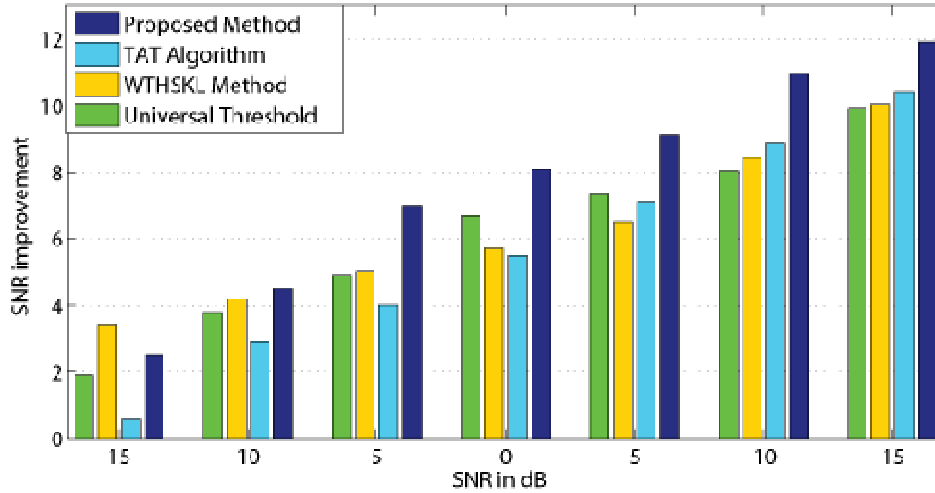


**FIGURE 8:** Performance comparison of different methods using hard thresholding function in terms of WSS for car noise corrupted speech.

#### 4.3.3. Results on Pink Noise-corrupted Signal

All the objective metrics for evaluating the performance of the proposed method relative to the other methods for pink noise-corrupted speech are computed and depicted in Fig. 9 through Fig. 10 and in Table 3.

The SNR improvement in dB resulted by using different methods are summarized in Fig. 9. It is vivid from this figure that the other methods produce comparatively lower improvement in SNR in dB in the whole SNR range, while the proposed method using hard thresholding function continues to remain superior in a sense that it gives the highest improvement in SNR in dB even at an SNR as low as -15 dB of pink noise.



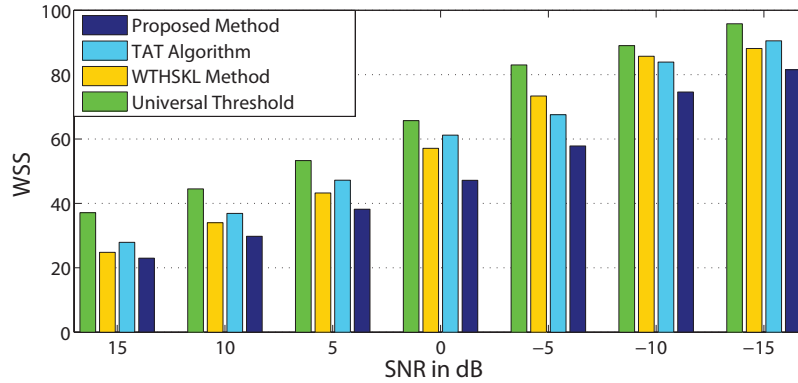
**FIGURE 9:** Performance comparison of the SNR Improvement in dB for different methods in the presence of pink noise.

The PESQ scores of the proposed method and that obtained by using different comparison methods are shown in Table 3 with respect to SNR levels varying from high (15 dB) to low (-15 dB). It is clear from the table that the other methods continue to provide lower PESQ scores, while the proposed method maintain comparatively higher PESQ scores even in the presence of severe pink noise of -15 dB.

The variation of the output WSS with respect to SNR levels for different methods and that for the proposed method using hard thresholding function is portrayed in Fig. 10. It is evident from analyzing each of these figures that, in the whole SNR range, the other methods continue to produce much higher WSS values with respect to the proposed method using hard thresholding function. Note that, the propose method performs the best in a sense that it yields the lowest WSS values almost at different SNR levels.

SNR [dB]	Universal	TAT	WTHSKL	Proposed Method
15	2.8	2.7	2.82	2.97
10	2.53	2.52	2.5	2.71
5	2.24	2.22	2.21	2.5
0	2	1.9	1.87	2.1
-5	1.55	1.65	1.67	1.9
-10	1.4	1.47	1.52	1.7
-15	1.3	1.39	1.43	1.5

**TABLE 3:** Performance comparison of PESQ scores for different methods in the presence of pink noise.



**FIGURE 10.** Performance comparison of different methods using hard thresholding function in terms of WSS values for pink noise corrupted speech.

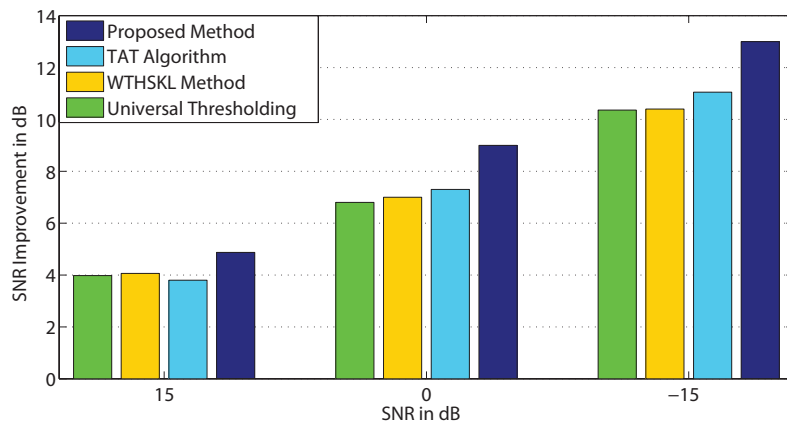
4.3.4. Results on Multi-talker Babble Noise-corrupted Speech

The results obtained from the multi-talker babble noise-corrupted speech in terms of the SNR improvement in dB, PESQ scores, and WSS values for the proposed method using hard thresholding function and that for the other methods are depicted in Fig. 11 through Fig. 13 at particular SNR levels of 15 dB, 0 dB and -15 dB. It is noticeable from these figures that the performance of all the methods degrades in the presence of multi-talker babble noise compared to that in the pink or car or white noise, but the proposed method retains its superiority with respect to all the levels of SNRs.

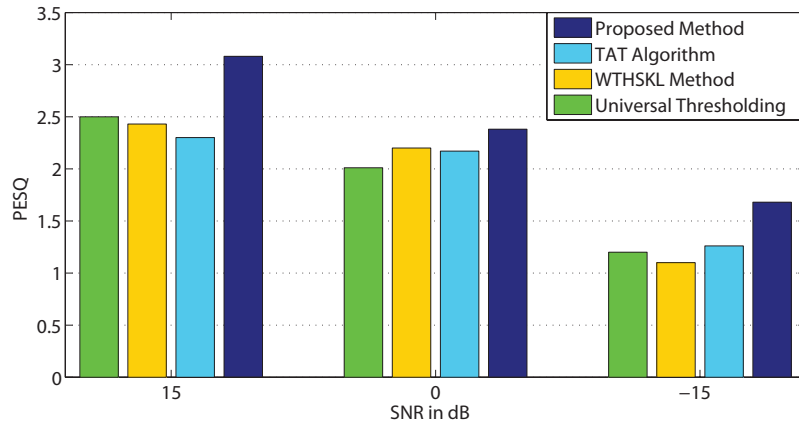
Fig. 11 provides plot for the SNR improvement in dB for all the methods for babble noise-corrupted speech. It is seen that the proposed method maintains better performance at all the SNR levels considered. Also the proposed method still remains the best thus showing higher capability of producing enhanced speech with better quality at a very low level of SNR of 0 dB or even lower than that.

In similar babble noisy condition, the PESQ scores resulting from using the speech enhancement methods under consideration are shown in Fig. 12. As seen, the proposed method continues to provide better results for the low levels of SNR, such as -15 dB.

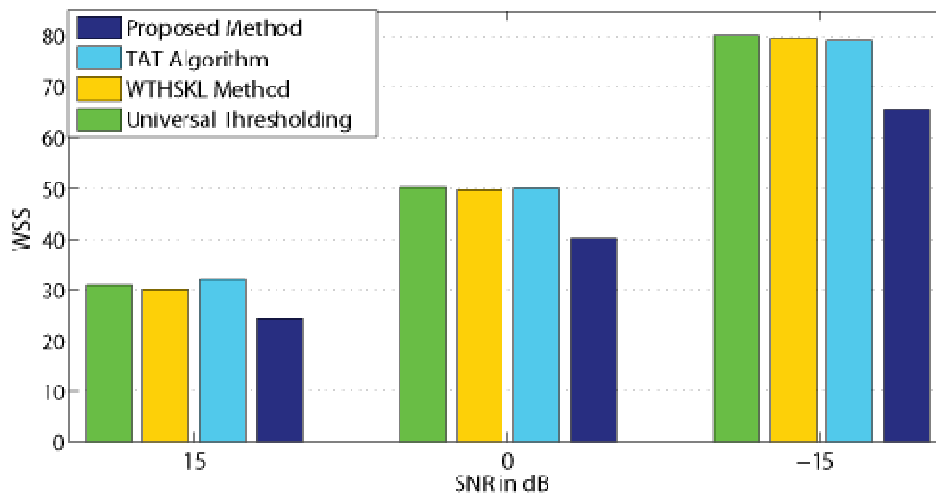
Also, the WSS values obtained from all the methods as a function of SNR are plotted in Fig. 13 for babble noise-corrupted speech. This figure illustrates that, as expected, the WSS values of the proposed method are somewhat increased in comparison to the other noisy cases, but its performance still remains better than that provided by the other methods for a wide range of SNR values from 15 dB to -15 dB.



**FIGURE 11.** Performance comparison of different methods using hard thresholding function in terms of SNR improvement in dB for babble noise corrupted speech.



**FIGURE 12.** Performance comparison of different methods using hard thresholding function in terms of PESQ scores for babble noise corrupted speech.



**FIGURE 13.** Performance comparison of different methods using hard thresholding function in terms of WSS values for babble noise corrupted speech.

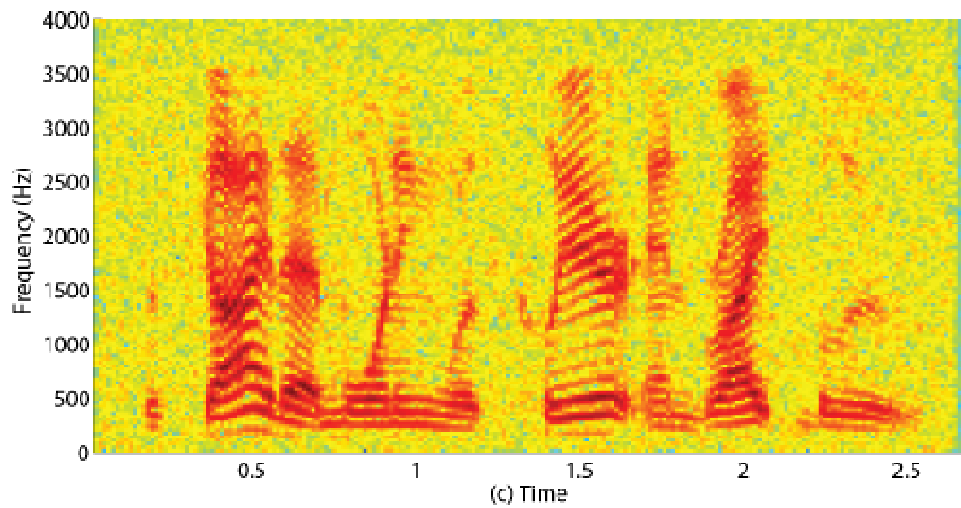
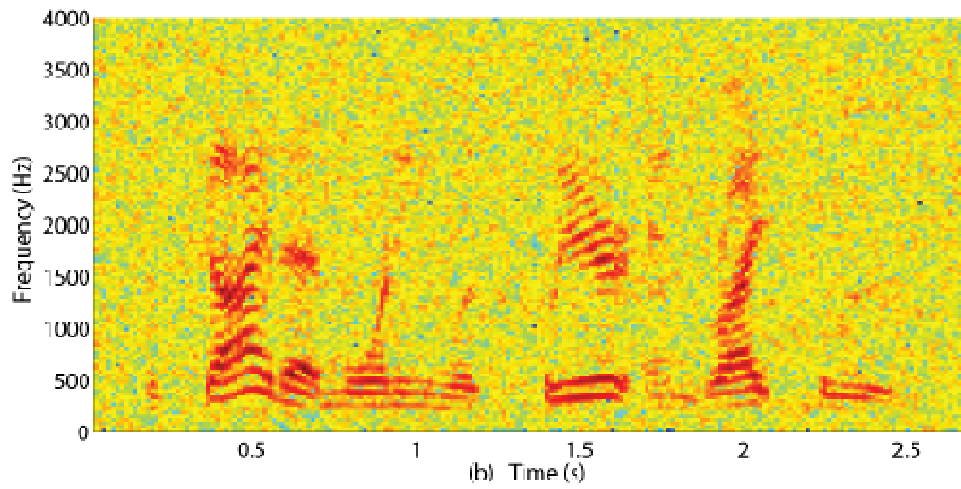
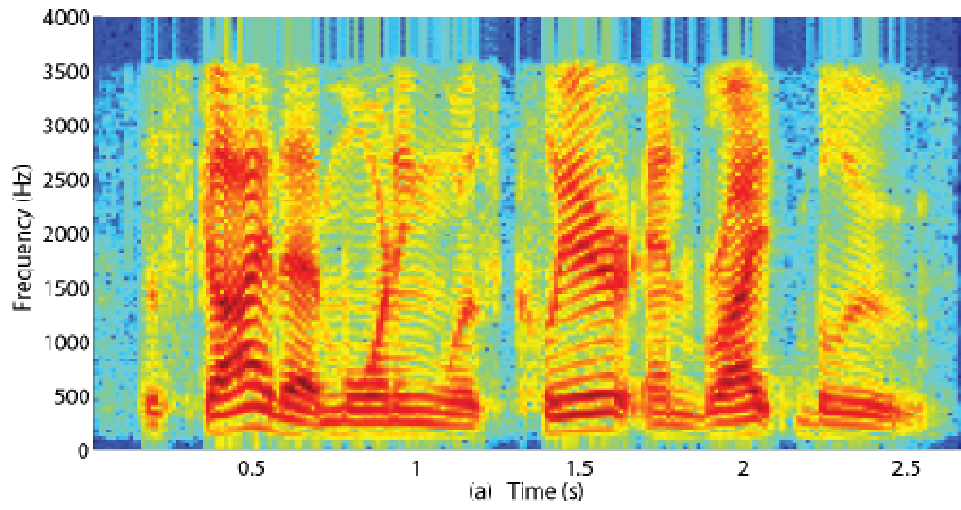
#### 4.4. Subjective Evaluation

In order to evaluate the subjective observation of the enhanced speech obtained by using the proposed method, spectrograms of the clean speech, the noisy speech, and the enhanced speech signals obtained by using all the methods are presented in Fig. 14 and 15 for white noise corrupted speech at an SNR of 5 dB and car noise corrupted speech at an SNR of -5 dB, respectively. It is evident from these figures that the harmonics are preserved and the amount of distortion is greatly reduced in the proposed method no matter the speech is corrupted by white or car noise regardless of its level. Thus, the spectrogram observations with lower distortion also validate our claim of better speech quality as obtained in our objective evaluations in terms of higher SNR improvement in dB, higher PESQ score and lower WSS in comparison to the other methods.

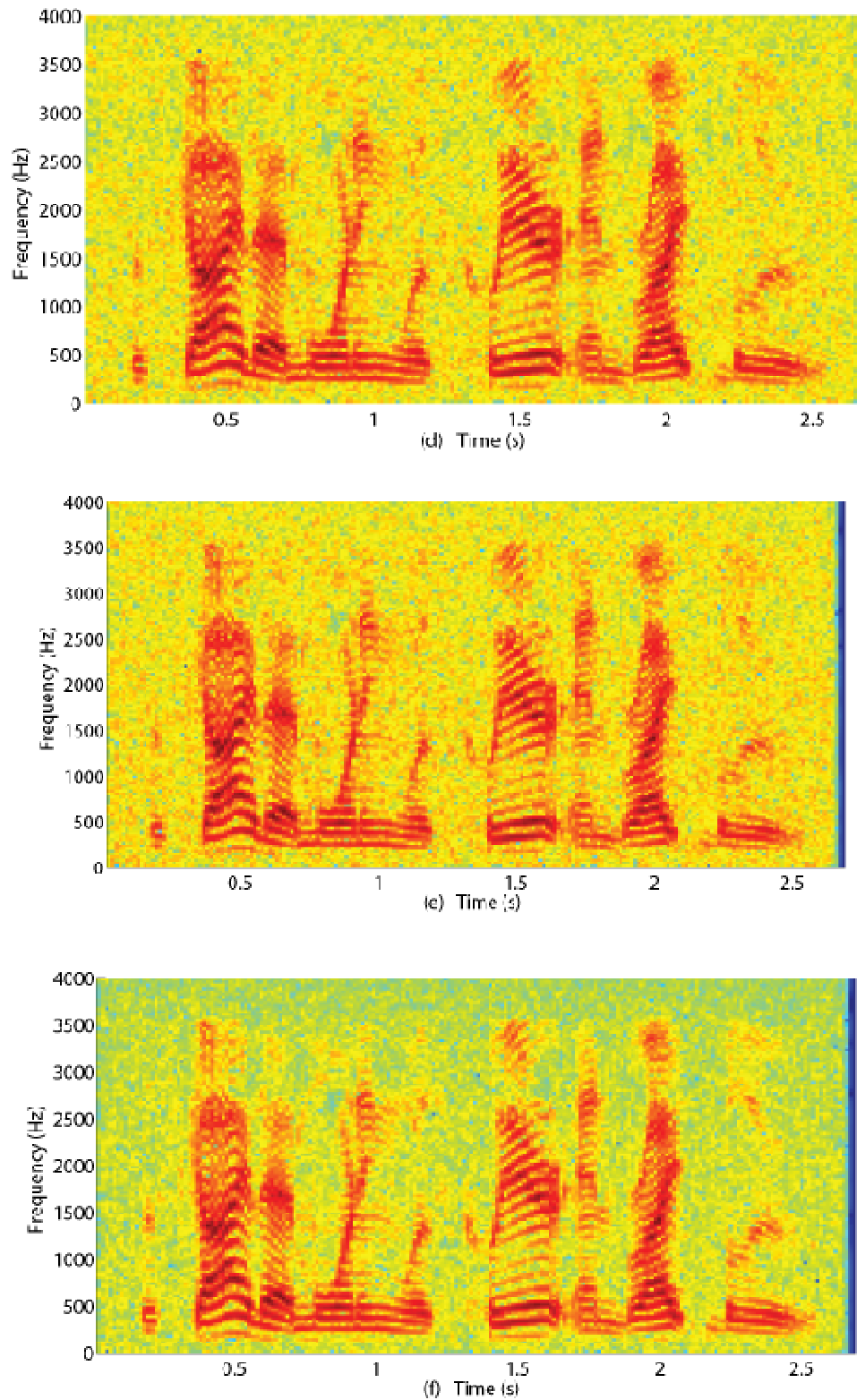
Informal listening tests are also conducted, where the listeners are allowed and arranged to perceptually evaluate the clean speech, noisy speech, and the enhanced speech signals. It is found that the subjective sound quality of the proposed method possess the highest correlation with the objective evaluation in comparison to that of the other methods in case of all the noises considered at different levels of SNR.

## **5. CONCLUSION**

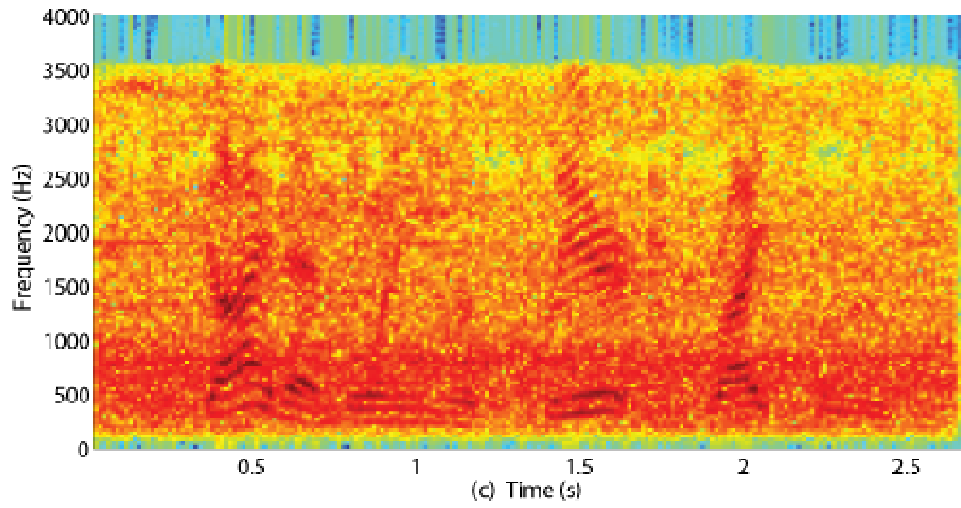
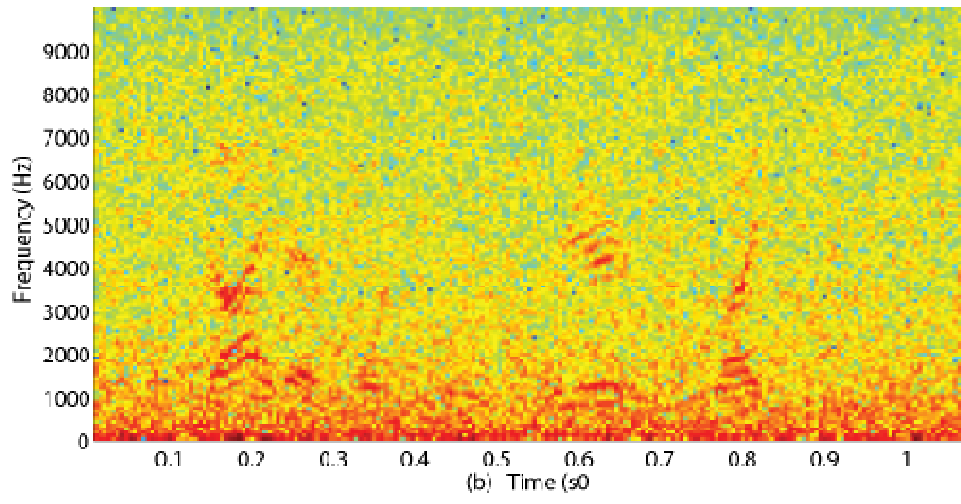
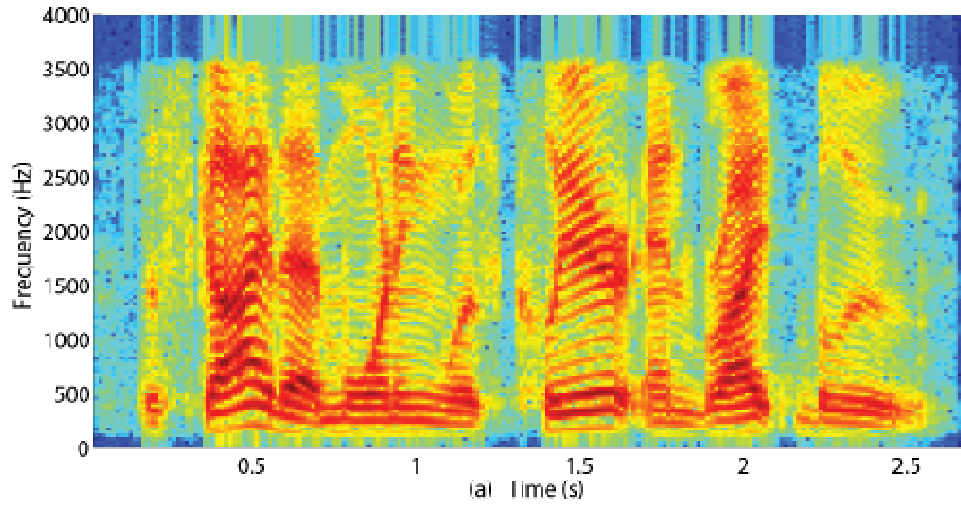
An improved wavelet-based approach to solve the problems of speech enhancement using the Probability distribution of Teager Energy Operated wavelet Packet coefficients has been presented in this paper. We incorporated a statistical model-based technique with teager energy operator of the wavelet packet coefficients to obtain a suitable threshold using symmetric K-L divergence. For solving the equation of pdf's, we choose Gaussian distribution as an acceptable pdf for noisy speech, clean speech and noise TEO coefficients in each sub-band. Unlike the unique threshold based method, the threshold value here is adapted based on the speech and silence segments. Then, by employing hard thresholding function the WP coefficients of the noisy speech are thresholded in order to obtain a cleaner speech. Simulation results show that the proposed method yields consistently better results in the sense of higher output SNR in dB, higher output PESQ, and lower WSS values than those of the existing thresholding based methods, hence results in a better enhanced speech than the existing thresholding methods.

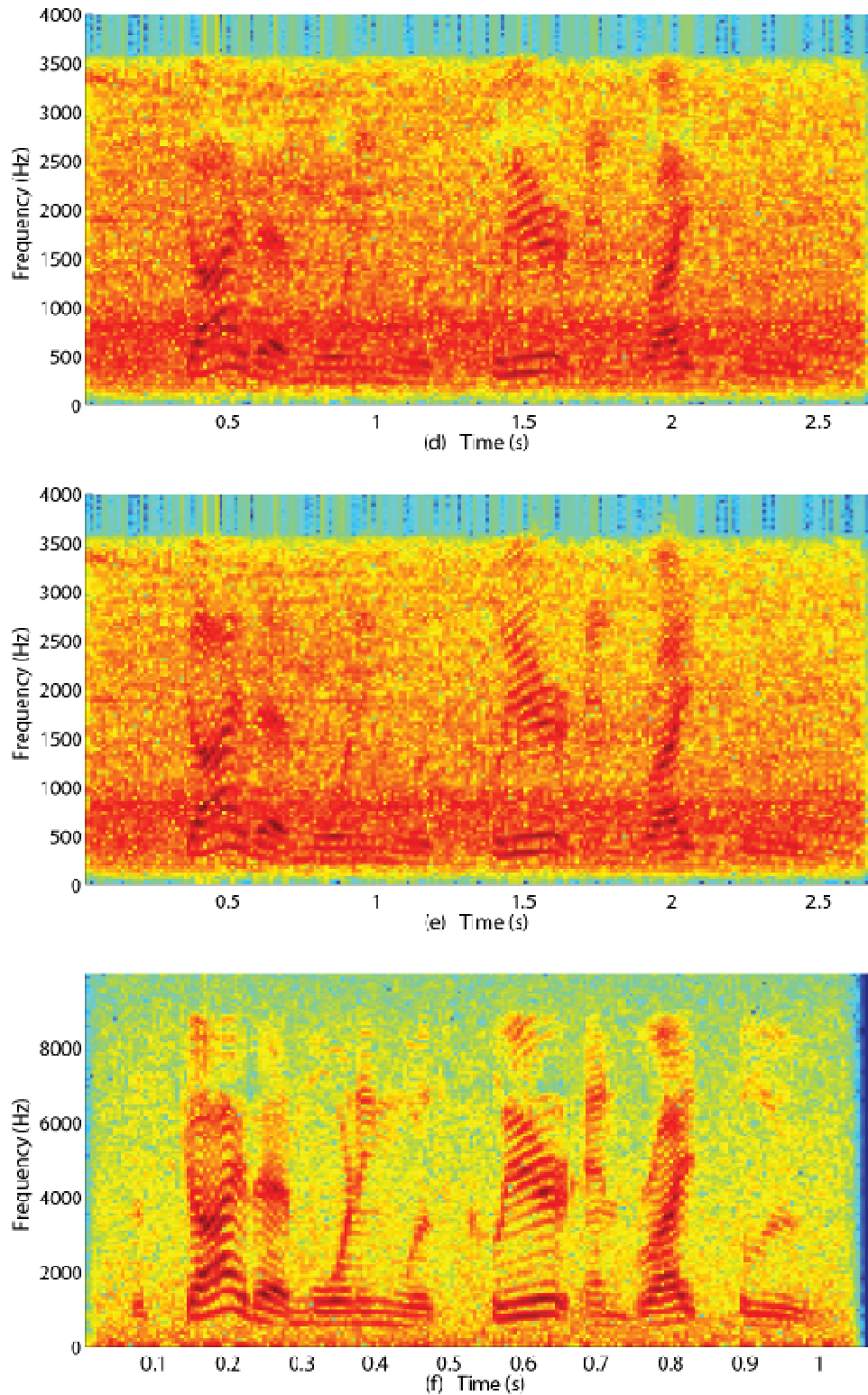






**FIGURE 14.** Spectrogram of sp10.wav utterance by a male speaker from the NOIZEUS database: (a) Clean speech, (b) Noisy speech (white noise from NOISEX92 database of SNR 5 dB), (c), (d), (e), (f)- enhanced speech signals obtained by using the Universal, TAT, WTHSKL, and the proposed methods, respectively.





**FIGURE 15.** Spectrogram of sp01.wav utterance by a male speaker from the NOIZEUS database: (a) Clean speech, (b) Noisy speech (car noise from NOIZEUS database of SNR -5 dB), (c), (d), (e), (f)-enhanced speech signals obtained by using the Universal, TAT, WTHSKL, and the proposed methods, respectively.

## REFERENCE

- [1] O'Shaughnessy, D., "Speech Communications: Human and Machine", 2nd Edition, *Wiley-IEEE Press*, 1999.
- [2] Loizou, P. C., "Speech Enhancement: Theory and Practice," *Boca Raton: CRC Press*, 2007.
- [3] Deller, J. Jr., Hansen, J. and Proakis, J., "Discrete-Time Processing of Speech Signals," *NY: IEEE Press*, 2000.
- [4] Lim, J., and Oppenheim, A., "Enhancement and bandwidth compression of noisy speech" *Proc. IEEE*, vol. 67, No. 12, pp. 221-239, Dec. 1979.
- [5] Virag, N., "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on Speech and Audio Processing*, volume 7, no. 2, pp. 126-137, Mar 1999.
- [6] Ephraim, Y. and Van Trees, H. L., "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, volume 3, pp. 251-266, 1995.
- [7] Mittal, U., and Phamdo, N., "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech Audio Processing*, volume 8, pp. 159-167, March 2000.
- [8] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise" *IEEE Trans. Speech, Audio Process*, volume. 11, pp. 334– 341, Jul. 2003.
- [9] Jabloun, F. and Champagne, B., "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, volume. 11, pp. 700-708, 2003.
- [10] You, C. H., Koh, S. N., and Rahardja, S.; , "An invertible frequency eigen domain transformation for masking-based subspace speech enhancement," *IEEE Signal Processing Letters*, volume.12, no.6, pp. 461- 464, June 2005.
- [11] Chang, J.-H., "Warped discrete cosine transform-based noisy speech enhancement," *IEEE Trans. Circuits and Systems II: Express Briefs*, volume 52, pp. 535 – 539, 2005.
- [12] Gustafsson, H., Nordholm, S.E., and Claesson, I., "Spectral subtraction using reduced delay convolution and adaptive averaging" *Speech and Audio Processing, IEEE Transactions on*, vol.9, no.8, pp.799-807, Nov 2001.
- [14] Kamath, S., and Loizou, P.; "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pp. IV-4164, 13-17, May 2002.
- [15] Yamashita, K. and Shimamura, T., "Nonstationary noise estimation using low-frequency regions for spectral subtractio," *Signal Processing Letters*, volume. 12, pp. 465-468, 2005.
- [16] Boll, S., "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, volume.27, pp. 113-120, Apr. 1979.
- [17] Chen, B. and Loizou, P. C., "A Laplacian-based (MMSE) estimator for speech enhancemen," *Speech Communication*, volume. 49, pp. 134-143, 2007.
- [18] P. C. Loizou, "Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech, Audio Process*, volume. 13, pp. 857–869, Sep. 2005.
- [19] Ephraim, Y. and Malah, D., "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume.33, no.2, pp. 443- 445, Apr 1985.

- [20] Sameti, H., and Sheikhzadeh, H., "L. Deng, R. Brennan, HMM-based strategies for enhancement of speech signals embedded in nonstationary noise," *IEEE Trans. Speech and Audio Processing*, volume 6, pp. 445–455, 1998.
- [21] Hansen, J. H. L., Radhakrishnan, V., and Arehart, K.H., "Speech Enhancement Based on Generalized Minimum Mean Square Error Estimators and Masking Properties of the Auditory System," *IEEE Transactions on Audio, Speech, and Language Processing*, volume. 14, no.6, pp. 2049-2063, Nov. 2006.
- [22] Almajai, I., and Milner, B., "Visually Derived Wiener Filters for Speech Enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, volume. 19, no.6, pp. 1642-1651, Aug. 2011.
- [23] Ben Jebara, S., "A Perceptual Approach to Reduce Musical Noise Phenomenon with Wiener Denoising Technique," *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, volume 3, pp. 14-19, May 2006.
- [24] Martin, R., "Speech Enhancement Based on Minimum Mean-Square Error Estimation and Super gaussian Priors," *IEEE Transactions on Speech and Audio Processing*, volume. 13, no.5, pp. 845- 856, Sept. 2005.
- [25] Papoulis, A. and Pillai, S. U., "Probability, Random Variables and Stochastic Processes," 4th Edition, *McGraw-Hill*, 2002.
- [26] Chang, S., Kwon, Y., Yang, S. I., and Kim, I. J., "Speech enhancement for non-stationary noise environment by adaptive wavelet packet," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume. 1, pp. 1-561 -1-564, 2002.
- [27] Yi, H. and Loizou, P.C., "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Signal Processing Letters*, volume. 12, pp. 59-67, 2004.
- [28] Tabibian, S. and Akbari, A. and Nasersharif, B., "A new wavelet thresholding method for speech enhancement based on symmetric Kullback-Leibler divergence," *14th International CSI Computer Conference (CSICC)*, pp. 495-500, 2009.
- [29] Bahoura, M. and Rouat, J., "Wavelet speech enhancement based on the Teager energy operator," *IEEE Signal Processing Letters*, volume. 8, pp. 10-12, 2001.
- [30] Donoho, D.L., "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, volume. 41, pp. 613-627, 1995.
- [31] Ghanbari, "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets," *Speech Communication*, volume 48, pp. 927 – 940, 2006.
- [32] Sheikhzadeh, H., and Abutalebi, H. R., "An improved wavelet-based speech enhancement system," *EUROSPEECH*, pp. 1855–1858, 2001.
- [33] Shao, Y., Chang, C. H., "A Generalized Time–Frequency Subtraction Method for Robust Speech Enhancement Based on Wavelet Filter Banks Modeling of Human Auditory System," *IEEE Transactions on Systems, Man, and Cybernetics*, volume. 37, no.4, pp.877-889, Aug. 2007.
- [34] Johnson, M. T., Yuan, X., and Ren, Y., "Speech signal enhancement through adaptive wavelet thresholding", *Speech Communication*, 2007.
- [35] Kaiser, J., "Some useful properties of teager's energy operators," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing, (ICASSP)*, volume 3, pp. 149 –152.
- [36] Kaiser, J., "Some useful properties of teager's energy operators," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pp.149 –152.

- [37] Lallouani, A., Gabrea, M., and Gargour, C., "Wavelet based speech enhancement using two different threshold-based denoising algorithms," *Canadian Conference on Electrical and Computer Engineering*, volume 1, pp. 315 – 318, 2004.
- [38] Malayeri, A., "Noise speech wavelet analyzing in special time ranges," *Int. Conf. Advanced Communication Technology (ICACT)*, volume 1, pp. 525 –528.
- [39] Ephraim, Y.; Malah, D.;; "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume. 32, no.6, pp. 1109- 1121, Dec 1984.
- [40] Erkelens, J.S.; Hendriks, R.C.; Heusdens, R.; Jensen, J.; , "Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients With Generalized Gamma Priors," *IEEE Transactions on Audio, Speech, and Language Processing*, volume. 15, no.6, pp.1741-1752, Aug. 2007.
- [41] Benesty, J., Sondhi, M., and Huang, Y., "Handbook of Speech Processing," *Springer*, 2008.
- [42] Gazor, S.; Wei Zhang; , "Speech enhancement employing Laplacian-Gaussian mixture," *IEEE Transactions on Speech and Audio Processing*, volume. 13, no.5, pp. 896- 904, Sept. 2005.
- [43] Plourde, E., "Bayesian short-time spectral amplitude estimators for single channel speech enhancement," *Ph.D. thesis*, Montreal, Que., Canada, Canada, 2009.
- [44] Makovoz, D., "Noise Variance Estimation In Signal Processing," *IEEE International Symposium on Signal Processing and Information Technology*, pp.364-369, Aug. 2006.
- [45] Babaie-Zadeh, M., and Jutten, C., "A general approach for mutual information minimization and its application to blind source separation," *IEEE Signal Processing Letters*, volume 85, pp. 975–995, 2005.
- [46] De Souza, P., "A statistical approach to the design of an adaptive self normalizing silence detector, Acoustics," *IEEE Transactions on Speech and Signal Processing*, volume. 31, pp. 678 – 684, 1983.
- [47] Busso, C.; Sungbok Lee; Narayanan, S.; , "Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.17, no.4, pp.582-596, May 2009.
- [48] Hu, Y., and Loizou, P., "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, volume. 16, pp. 229 –238, 2008.