IACSIT International Journal of Engineering and Technology Vol.1,No.5,December,2009
ISSN: 1793-8236

# Performance Evaluation of Different Thresholding Methods in Time Adaptive Wavelet Based Speech Enhancement

A. Sumithra M G, *Member, IACSIT*  B. Thanushkodi K

***Abstract*— Speech quality and intelligibility might significantly deteriorate in the presence of background noise, especially when the speech signal is subject to subsequent processing. Speech enhancement algorithms have therefore attracted a great deal of interest in the past two decades. Wavelets provide a powerful tool for non-linear filtering of signals contaminated by noise and wavelet thresholding de-noising techniques provide a new way to reduce noise in signal. In this work speech enhancement is accomplished through the use of different thresholding on time adaptive discrete Daubechies wavelet transform co-efficients. However, the soft thresholding is best in reducing noise but worst in preserving edges, and hard thresholding is best in preserving edges but worst in de-noising. Motivated by finding a more general case that incorporates the soft and hard thresholding to achieve a compromise between the two methods, the trimmed thresholding method is proposed in this paper to enhance the speech from background noise. The performance of different thresholding methods are evaluated by enhancing the speech corrupted by various noises. Finally, the objective and subjective experimental results show that the proposed scheme with trimmed thresholding is superior in denoising as compared to hard and soft thresholding methods. It also indicates that the proposed method gives better mean square error (MSE) performance than other wavelet thresholding methods.**

***Index Terms*— Speech enhancement, Time adaptive Daubechies wavelet transform, Time adaptation factor, Thresholding.**

## I. Introduction

In communication systems, speech signals can be contaminated by environmental noise and, as a result, the communication quality can be affected making the speech less intelligible. Voice quality and intelligibility are always important for communication systems, either wired or wireless, either in human-to-human or human-to-machine interactions. In order to obtain near-transparent speech communications, for example via mobile phones, speech enhancement techniques have been employed to improve the quality and intelligibility of the noise corrupted speech and/or the speech recognition performance. The problem of de-noising consists of removing noise from corrupted signal without altering it. The corrupting noise sources are usually classified into additive and convolutional. The former very often dominates in real world applications, and the spectral subtraction (SS) approach has been a very popular example solution for it [1],[2],[3]. To subtract the noise components from the input noisy speech, the SS algorithm has to estimate the statistics of the additive noise in frequency domain. Under low signal-to-noise ratio (SNR) conditions, a spectral flooring process is usually taken to prevent the over-subtraction situation occurred. However, all such processes very often produce some unnatural residual noise in the enhanced speech, the so-called musical noise, due to the inevitable random tone peaks generated in the time-frequency spectrogram. Previous studies have pointed out that this perceivable residual noise can be effectively alleviated by considering the masking effect in human auditory system [4],[5] i.e., the residual noise will not be perceived if it is under the masking thresholds in human auditory functions.

In recent years, several alternative approaches such as signal subspace methods [6], have been proposed for enhancing the degraded speech. In subspace method the estimation of signal subspace dimension is difficult for unvoiced period and transitional regions. Existing approaches to this task include traditional methods such as spectral subtraction and Ephraim Malah filtering [7], a drawback of this technique is the necessity to estimate the noise or the signal to noise ratio. This can be a strong limitation when recording with non-stationary noise and for situations where the noise can not be estimated.

Fourier domain was long been the method of choice to suppress noise. Recently however, methods based on the wavelet transformation have become increasingly popular. Wavelets provide a powerful tool for non-linear filtering of signals contaminated by noise. Mallat and Hwang [8] have shown that effective noise suppression may be achieved by transforming the noisy signal into the wavelet domain, and preserving only the local maxima of the transform. Alternatively, a reconstruction that uses only the large-magnitude coefficients has been shown to approximate well the uncorrupted signal. In other words, noise suppression is achieved by thresholding the wavelet transform of the contaminated signal. The method of wavelet threshold de-noising is based on the principle of the

Manuscript submitted July 3, 2009 for review. F. A. Author is with the Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India (phone: 09865816671).

S. B. Author, is with Akshya College of Engineering and Technology , Coimbatore, Tamil Nadu. India. (phone: 09843394451).

multiresolution analysis. The discrete detail coefficients and the discrete approximation coefficients can be obtained by a multi-level wavelet decomposition. Wavelet-based techniques using coefficient thresholding [8], using adaptive thresholding [9] approaches have also been applied to speech enhancement. Donoho introduced wavelet thresholding (shrinking) as a powerful tool in denoising signals degraded by additive white noise and more recently a number of attempts have been made to use perceptually motivated wavelet decompositions coupled with various thresholding and estimation methods. Although the application of wavelet shrinking for speech enhancement has been reported in literature [10]-[13], there are many problems yet to be resolved for a successful application of the method to speech signals degraded by real environmental noise types. The most known thresholding methods in the literature are the soft and hard thresholding. We can expect that the technique of soft thresholding would introduce more error or bias than hard thresholding does [6]. But on the other hand, soft thresholding is more efficient in de-noising. To achieve a compromise between the two methods, the trimmed thresholding method is proposed in this paper for noisy speech co-efficient to reduce the noise. In addition in this paper, trimmed thresholding technique is used for de-noising as well as good in preserving the edges.

The main objective of the proposed method is to improve on existing single-microphone schemes for an extended range of noise types and noise levels, thereby making this method more suitable for mobile speech communication applications than the existing. This algorithm introduces a speech enhancement system based on a time adaptive discrete wavelet denoising using trimmed thresholding. The performance of the proposed method was evaluated on several speakers and under various noise conditions including white noise, pink noise, F16 cockpit noise, babble noise, high frequency channel noise and car interior noise. Subjective experiments by means of a listening test shows that the system based on this method has significant improvement over the wavelet based approach using hard and soft thresholding and the state-of-the-art speech enhancement system. The results of the proposed method shows that it is well suited for adverse noise conditions and yields better spectral performance. It is very important characteristic for speech recognition or speaker verification.

This paper is organized as follows: Section I represents a survey on the related works. Section II presents principle of wavelet and discrete wavelet computation. Section III presents the proposed scheme for speech signal enhancement. In section IV de-noising by thresholding in the wavelet domain introduced and the hard ,soft thresholding are reviewed and a new trimmed thresholding method is proposed for speech enhancement purpose that has better performance. Section V shows the steps for implementation and the experimental results were discussed, which validate the proposed thresholding algorithm. Finally section VI summarizes the presented research work

## II. Background – Wavelet Analysis

Wavelet transform has been intensively used in various fields of signal processing. It has the advantage of using variable size time-windows for different frequency bands. This results in a high frequency-resolution (and low time-resolution) in low bands and low frequency-resolution in high bands. Consequently, wavelet transform is a powerful tool for modeling non-stationary signals like speech that exhibit slow temporal variations in low frequency and abrupt temporal changes in high frequency. Moreover, when one is restricted to use only one (noisy) signal (as in single-microphone speech enhancement), generally the use of the subband processing can result in a better performance. Therefore, wavelet transform can provide an appropriate model for speech signal denoising applications. In the present work, the computation of Discrete Wavelet Transform (DWT) providing sufficient information for both analysis and synthesis of the original signal, with a significant reduction in the computation time. The DWT is considerably easier to implement without needing to perform numerical integration as like Continuous wavelet transform (CWT).

### *A. DWT Computation*

The DWT analyze the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information shown in

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k] g[n-k]$$

Fig.1.The DWT of a signal x(n) is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response *g* resulting in a convolution of the two as in (1)

The signal is also decomposed simultaneously using a high-pass filter h. The output giving the detail coefficients from the high-pass filter and approximation coefficients from the low-pass filter. It is important that the two filters are related to each other and they are known as a quadrature mirror filter (2). However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then sub sampled by 2.

$$\begin{aligned} y_{\mathrm{high}}[n] &= \sum_{k=-\infty}^{\infty} x[k] h[2n-k] \\ y_{\mathrm{low}}[n] &= \sum_{k=-\infty}^{\infty} x[k] g[2n-k] \end{aligned} \tag{2}$$

This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled.

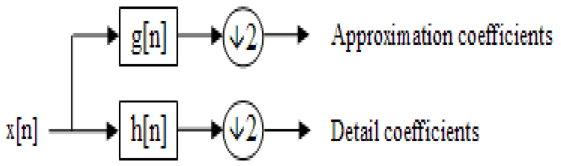



Fig.1 Block diagram of filter analysis

This decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled. This is represented as a binary tree with nodes representing a sub-space with a different time-frequency localization. The tree is known as a filter bank.

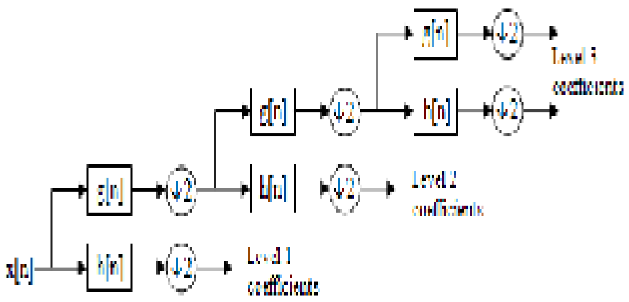



Fig.2 A three level filter bank

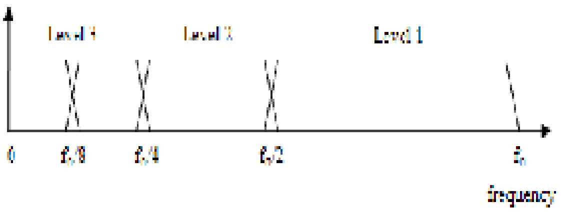



Fig.3. Frequency domain representation of the DWT

At each level in the above diagram the signal is decomposed into low and high frequencies shown in Fig.2 and Fig.3. Due to the decomposition process the input signal must be a multiple of $2^n$ where $n$ is the number of levels. This proposed method of analysis has gone up to $4^{th}$ level decomposition using Daubechies wavelet-4. The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling function which generates an orthogonal multi resolution analysis. It is one of the brightest wavelet on research which is compactly supported orthonormal wavelets [14].

## III. Proposed Scheme using Time Adaptive Daubechies Wavelet Transform

Wavelet decomposition transforms signal from time domain to time-scale domain, and it can describe the local feature well in both time domain and frequency domain. Because the amplitude of the discrete detail coefficients of the noise decreases with the level increasing, we can select a threshold, modify and process all of the discrete detail coefficients at all scale by hard thresholding or soft thresholding or trimmed thresholding so as to remove noise [10]. A noisy speech signal can be modeled as the sum of clean speech and additive background noise. If the signal includes ambient noise, the result is an additive signal model given by,

$$x = y + n \tag{3}$$

where x is noisy signal, y is clean speech and n is additive noise component . So that

$$X=Y+N \tag{4}$$

where X = Wx, Y = Wy, N = Wn in wavelet domain [14]. The matrix notation represents the coefficients across each scale and time. Block diagram of the proposed approach is shown in Fig.4. First DWT of the noisy speech is taken then the time adaptive nature is captured by time varying linear factor $T(a,t)$ calculation for each scale ($a = 2^m$) and time ($\tau = n\,2^m$) using (5). This factor only affects the duration of amplitude envelope of wavelet, but not affects the frequency.

$$T(a, t+\Delta t) = \frac{1}{\left(1 - \dfrac{C_s}{C_s + \left|X_{TADWT}(a,t)\right|}\right)\left(1 + \left|\dfrac{\partial}{\partial t} X_{TADWT}(a,t)\right|\right)} \tag{5}$$

For implementation based on Yao and Zhang's work [14] for cochlear implant coding, coefficients at 22 scales, m = 7,8,…28 are calculated using numerical integration of the CWT. These 22 scales corresponds to center frequencies logarithmically spaced from 225 Hz to 5300Hz have considered in this method. $C_s = 0.8$ is a constant representing non linear saturation effects in the cochlear model [15]. Since, the primary adaptation mechanism involves variation of the wavelet time support, the impact of initial time support was done by turning off adaptation mechanism ($T(a,t) = 1$). The resulting time adaptive wavelet transform coefficients $X_{TADWT}(a,t)$ are calculated from the product of DWT coefficients $X_{DWT}(a,t)$ with a time constant $K(a,t)$ and the same is substituted in (5) for time adaptation mechanism. From the reported analysis [8], [11],

$$X_{TADWT} = K(a,t) * X_{DWT}(a,t)$$
$$K(a,t) = \frac{\sqrt{\Pi}}{C}\,\frac{1}{\sqrt{1+T^2(a,t)}} \tag{6}$$

where $C_0 + C_1 + C_2 + C_3 = C = 2$ (normalizing constant). The normality is obtained by (7).

$$C_0 = \frac{(1+\sqrt{3})}{4}, \quad C_1 = \frac{(3+\sqrt{3})}{4}, \quad C_2 = \frac{(3-\sqrt{3})}{4}, \quad C_3 = \frac{(1-\sqrt{3})}{4} \tag{7}$$

Then $h(k) = C_k / \sqrt{2}$ , co-efficients for db4 wavelets are as follows [16],

$$h(0) = \frac{\left(1+\sqrt{3}\right)}{4\sqrt{2}} \;,\quad h(1) = \frac{\left(3+\sqrt{3}\right)}{4\sqrt{2}}$$
$$h(2) = \frac{\left(1-\sqrt{3}\right)}{4\sqrt{2}} \;,\quad h(3) = \frac{\left(3-\sqrt{3}\right)}{4\sqrt{2}} \tag{8}$$

Since, it is a discrete wavelet this computational method requires no integration and is more efficient. Removing noise components by thresholding the wavelet coefficients is based on the observation that in many signals (like speech), energy is mostly concentrated in a small number of wavelet dimensions. The coefficients of these dimensions are relatively large compared to other dimensions or to any other signal (specifically noise) that has its energy spread over a large number of coefficients. Hence, by setting smaller coefficients to zero, one can nearly optimally eliminate noise while preserving the important information of the original signal.

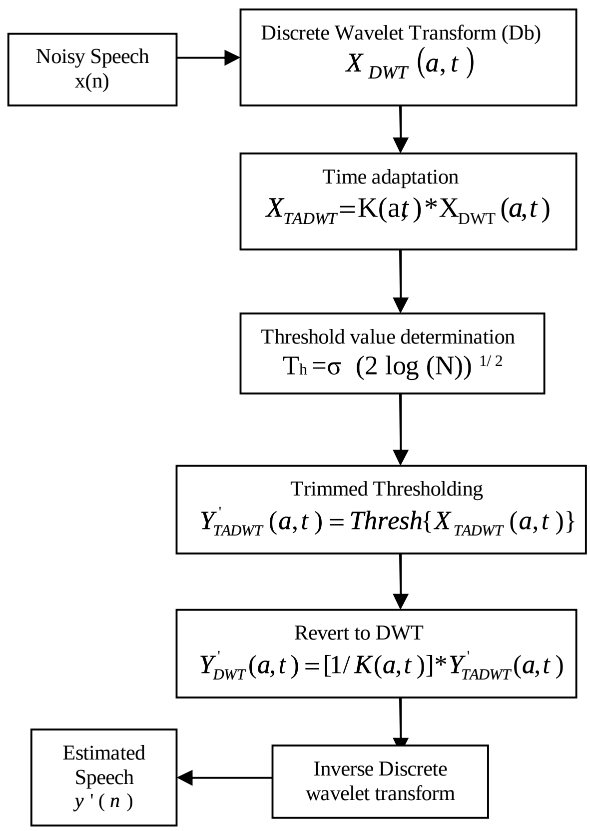



Fig.4 Proposed Scheme

In wavelet representation noise characteristics will tend to be characterized by smaller coefficients across time and scale while signal energy will be concentrated in larger coefficients. This offers the possibility of using threshold to separate the signal from the noise. Generally, the selected threshold has to be multiplied by the median value of the detail coefficients at some level which is called threshold processing. To find the comparatively the best method of threshold processing, we study how they influence the performance of wavelet de-noising.

Specifically, in each wavelet band, we calculate the variance of the noisy coefficients. Using (9), each variance then can then be employed to set the threshold to a value based on the noise energy in the band. In practical situations, one is often encountered with colored rather than white noises. Assuming zero-mean Gaussian noise, the coefficients will be Gaussian random variables of zero mean and variance $\sigma^2$, the standard deviation σ is thus estimated by (9),

$$\sigma = (1 / 0.6745) \text{ Median } (\,|\,c_i\,|\,) \tag{9}$$

where $c_i$ represents high frequency wavelet coefficients which are used to identify the noise components at first level decomposition. The set of standard deviation values can now be used as the "noise profile" for threshold setting [15]. This noise profile estimation enables the algorithm to cope with colored noises. Threshold value [10] can be determined by (10),

$$T_h = \sigma\,(2 \log (N \log_2 N)^{1/2} \tag{10}$$

where $T_h$ is the threshold value. N is the length of the noisy signal. In threshold selection, we should not ignore the detail coefficients in every level that probably influence the robustness of the threshold estimating. So we have to rescale a selected threshold in some level. In this paper, the threshold is dependent on the detail coefficients at every level.

## IV. DENOISING BY WAVELET THRESHOLDING

The wavelet de-noising technique is called thresholding, it is a non linear algorithm. It can be decomposed in tree steps. The first one consists in computing the coefficients of the wavelet transform (WT) which is a linear operation. The second one consists in thresholding these coefficients. The last step is the inversion of the thresholded coefficients by applying the inverse wavelet transform, which leads to the de-noised signal. This technique is simple and efficient. However it relies heavily on the choice of the threshold, which in its turn depends on the noise distribution. In the wavelet thresholdimg de-noising, we should first select a threshold and process the components of wavelet transform of the noisy signal in order to improve signal-to-noise ratio (SNR).

### A. *Soft and Hard thresholding*

In the literature there are two types of thresholding techniques applicable to speech processing which are Hard Thresholding and Soft thresholding. Hard thresholding can be described as the usual process of setting to zero the elements whose absolute values are lower than the threshold. Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards 0. Let $T_h$ denote the given threshold. The hard thresholding is defined by (11),

$$Y'_{TADWT} = \begin{cases} X_{TADWT} \;, & |X_{TADWT}| \geq T_h \\ 0 \;, & |X_{TADWT}| < T_h \end{cases} \tag{11}$$

ISSN: 1793-8236

The soft thresholding is defined by (12),

$$Y'_{TADWT} = \begin{cases} \text{Sign}\ (|X_{TADWT}|)\cdot(|X_{TADWT}| - T_h),\ |X_{TADWT}| \ge T_h \\ 0 \qquad\qquad ,\ |X_{TADWT}| < T_h \end{cases} \qquad (12)$$

where $Y'_{TADWT}$ is thresholded time adaptive wavelet co-efficient of estimated speech signal and $X_{TADWT}$ is time adaptive wavelet co-efficient of the noisy speech signal .

### *B. Trimmed thresholding*

Motivated by finding a more general case that incorporates the soft and hard thresholding, we proposed the following thresholding rule as in (13),

$$Y'_{TADWT} = \begin{cases} X_{TADWT} * \dfrac{|X_{TADWT}|^{a} - |T_h|^{a}}{|X_{TADWT}|^{a}},\ |X_{TADWT}| \ge T_h \\ 0 \qquad\qquad ,\ |X_{TADWT}| < T_h \end{cases} \qquad (13)$$

$T_h$ is chosen as an estimate of noise level. When α = 1, it is equivalent to soft thresholding; when α →∞, it is equivalent to hard thresholding. Fig.5 graphically shows its relation with soft and hard thresholding. It can be clearly seen that trimmed thresholding is something between hard and soft thresholding. With careful tuning of parameter α for a particular signal, one can achieve best de-noising effect within thresholding framework.

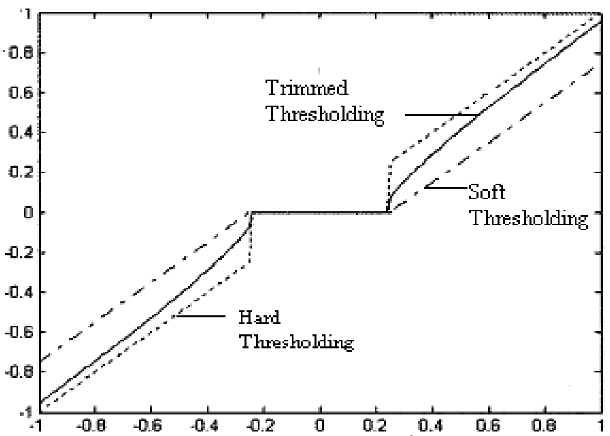



Fig.5. Hard, Soft and Trimmed thresholding function

## V. Implementation and Evaluation

In this section, the implementation steps of the proposed method were discussed. Furthermore, the performance of the best thresholding method is evaluated and compared it with other methods of thresholding. A clean speech sentences from the TIMIT database [17] is corrupted by different noises for various SNR ranging from -10dB to +10 dB are considered as noisy input speech. Objective and subjective tests were conducted to evaluate the quality of the proposed method. Objective Quality measures provide a measure based on a mathematical comparison of the original and processed speech signals. The quality of speech signal is a subjective measure which reflects the way the signal is perceived by listeners. It can be expressed in terms of how pleasant the signal sounds are or how much effort is required to understand the message. To evaluate the effectiveness of using this proposed method for denoising of speech signals, we compared it to other two thresholding on this task. Two sets of additive noise experiments were implemented on this data. In the first, white Gaussian noise was added to the sentences from the TIMIT database at SNR levels of -10,-5, 0, +5, +10 db. In the second, specific noise characteristics including pink noise, car noise, babble noise, F-16 cockpit noise and HF channel noise was added at 6dB SNR level to evaluate how well the methods work with non-white and relatively non-stationary noise sources. Signal-to-noise ratio (SNR), Itakuro-Saito (IS) distance and MMSE are used as objective measurements criteria for both set of experiments.

### *A. Implementation steps:*

Step:1 Computation of the discrete wavelet transform for noisy speech.

Step:2 Computation of time adaptation factor and multiply with discrete wavelet coefficients using (6).

Step:3 Estimate the noise using (9) and determine the threshold value using (10) then apply different thresholding techniques for the time adaptive wavelet co-efficients using (11), (12) and (13).

Step:4 Inverse Time Adaptive Discrete Wavelet transform is taken through dividing the co-efficients by that adaptation factor, which yields DWT coefficients.

Step:5 Taking Inverse Discrete wavelet Transform ( IDWT) the enhanced speech with reduced noise components is obtained while applying trimmed thresholding.

But in case of hard and soft thresholding post filtering is done to achieve the comparable results.

### *B. Objective Measure Evaluation*

Objective Quality measures provide a measure based on a mathematical comparison of the original and processed speech signals that can be easily implemented and reliably reproduced.

#### *Signal to noise ratio*

The global SNR values are determined by the following equation,

$$SNR_{dB} = 10\log_{10}\left(\frac{\sum_n s^2(n)}{\sum_n [s(n) - \hat{s}(n)]^2}\right) \qquad (14)$$

where $s(n)$ = clean speech and $\hat{S}(n)$ = estimated speech. If the summation is performed over the whole signal length, the operation is called as global SNR.

#### *Minimum Mean Square Error*

Mean Square Error (MSE) is defined as to be the average power of the difference between the enhanced speech and clean one. It can be obtained by

$$\hat{r} = E\left\{[S(n) - \hat{S}(n)]^2\right\} \qquad (15)$$

The objective of any speech enhancement system is to minimize this MSE.

*Itakura-Saito (IS) distance*

IS distance is a meaningful measure of performance when the two waveforms differ in their phase spectra.

$$d(a,b)=\frac{(a-b)^T\ R(a-b)}{a^T R(a)} \tag{16}$$

where 'a' is the vector for the prediction coefficients of the clean speech signal, vector R is the (Toeplitz) autocorrelation matrix of the clean speech signal and vector 'b' is the prediction coefficients of the enhanced signal. Many reported experiments confirmed that two spectra would be perceptually nearly identical if the distance is from 1 to 10, with lower values indicating lesser distance and better speech quality.

### C. *Subjective Measure Evaluation*

This provide a broad measure of performance since a large difference in quality is necessary to make it distinguishable to the listener. The mean opinion score (MOS) provides a numerical measure of the quality of human speech. The scheme uses subjective tests (opinionated scores) that are mathematically averaged to obtain a quantitative indicator of the system performance. To determine MOS, a number of listeners rate the quality of test sentences by hearing test. Based on the perceived quality of enhanced speech listener gives a rating for each sentence as follows: (1) Bad (2) Poor (3) Fair (4) Good (5) Excellent. The MOS is the arithmetic mean of all the individual scores, and can range from 1 (worst) to 5 (best).

Average MOS is computed by having a group of 20 listeners to rate the quality of the enhanced speech on a five point scale, then averaging the results. For all measures, results are averaged, giving a single evaluation metric for each method for input speech corrupted by various noises shown in Fig. 6. and it shows that the obtained average MOS of Trimmed thresholding is better as compared with the obtained results for Hard and Soft thresholding. But in the case of babble and HF channel noise the performance of the proposed thresholding method is similar to soft thresholding.

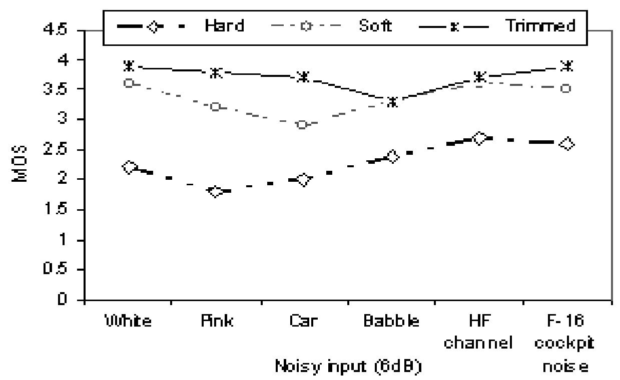



Fig.6 MOS comparisons for different noises

In addition Itakura-Saito (IS) distance and MMSE are computed for the estimated speech for three different thresholding methods at different noise conditions are shown in Fig.7 and Fig.8. By referring Fig 7& 8 it is clear that the trimmed thresholding method outperforms in speech signal de-noising as compared to other two methods in all aspects.

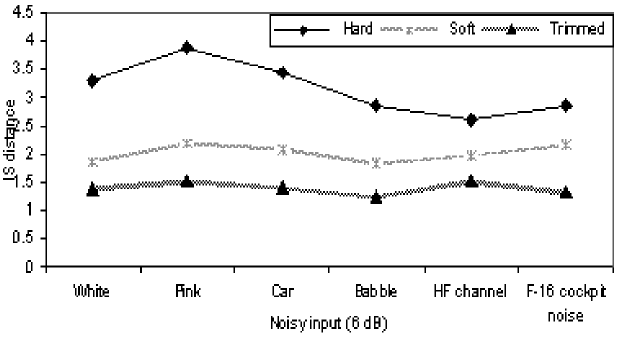



Fig.7 IS distance measure comparisons for different noises

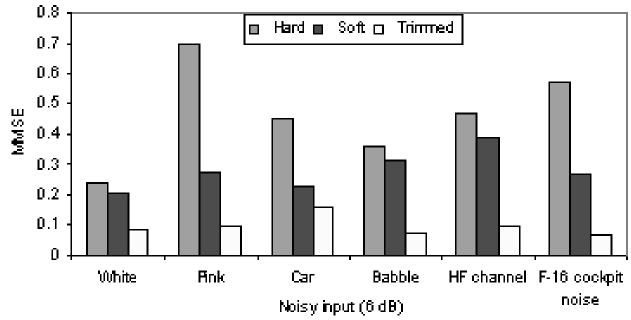



Fig.8 MMSE comparisons for different noises

Output SNR results for the white noise condition across range of SNR values are shown in Fig.9. Thresholding methods compared include Hard, Soft and proposed trimmed method. From fig.9 the proposed thresholding method clearly have the best performance for this white noise condition. The output SNR of the proposed method, Soft thresholding method shows the linear SNR improvement as compared with hard thresholding.

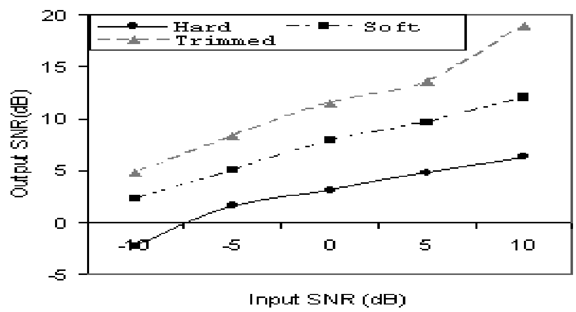



Fig.9. Output SNR results for white noise case at -10,-5, 0,+5, +10 dB input SNR

Proposed method gave about 15 dB improvement at the lower input SNRs decreasing to about 9dB improvement at the higher input SNRs. Similarly, soft thresholding gave about 12 dB improvement at the lower input SNRs decreasing to about 3dB improvement and hard thresholding gave about 7dB improvement at the lower input SNRs decreasing to about -3dB improvement at the higher input SNRs respectively.

SNR improvement across varying realistic noise conditions at 0 dB SNR are shown in Fig.4. Here the results are given

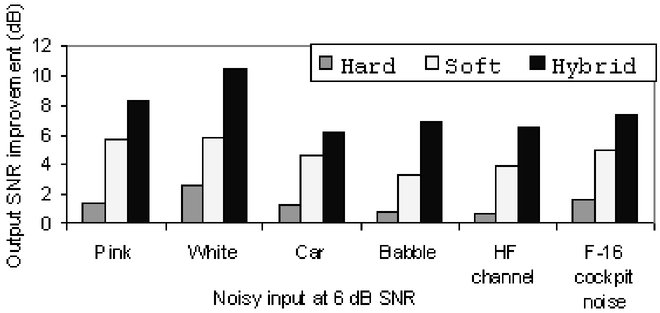
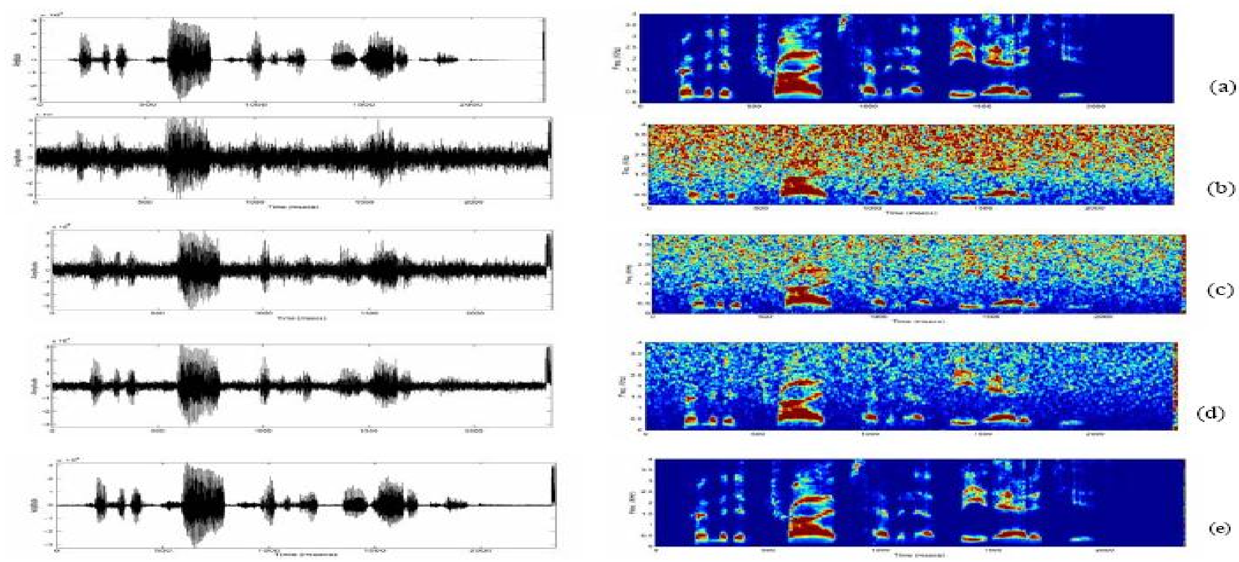



Fig.10 SNR comparisons for varying noise conditions at 0dB SNR

as net improvement, so that relative effectiveness can be seen for all six noise conditions as a function of thresholding method. The proposed method substantially outperforms the other methods in all cases, but relatively less performance in car noise conditions.Fig.11, 12 and 13 shows the time domain representation and spectrogram (shows the energy in a signal at each frequency and at each time) of the noisy, clean and enhanced speech using different thresholding methods. From results it is clear that in time domain the estimated speech using trimmed thresholding is more identical to the clean reference speech as compared with other two thresholding methods. In case of spectrogram the enhanced speech using trimmed is more comparable with spectrogram of clean speech as compared to the obtained results of other two methods. The intensity variations in enhanced speech using different thresholding algorithm are due to decrease in signal strength.

Fig.11 Time domain representation and Spectrogram of (a).Clean Speech (b) Noisy speech (white noise 0 dB)
(c) Enhanced speech using Hard thresholding (d) Enhanced speech using Soft thresholding



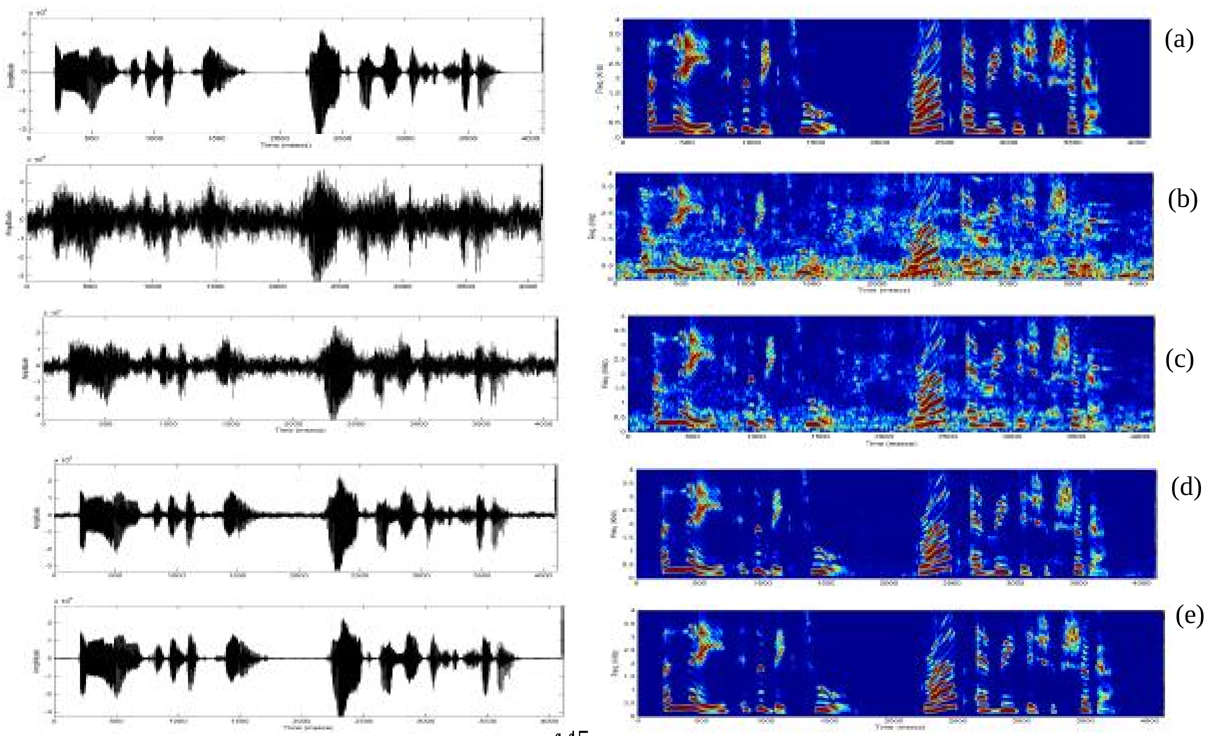



Fig.12 Time domain representation and Spectrogram of (a).Clean Speech (b) Noisy speech (Babble Noise 6 dB)
(c) Enhanced speech using Hard thresholding (d) Enhanced speech using Soft thresholding

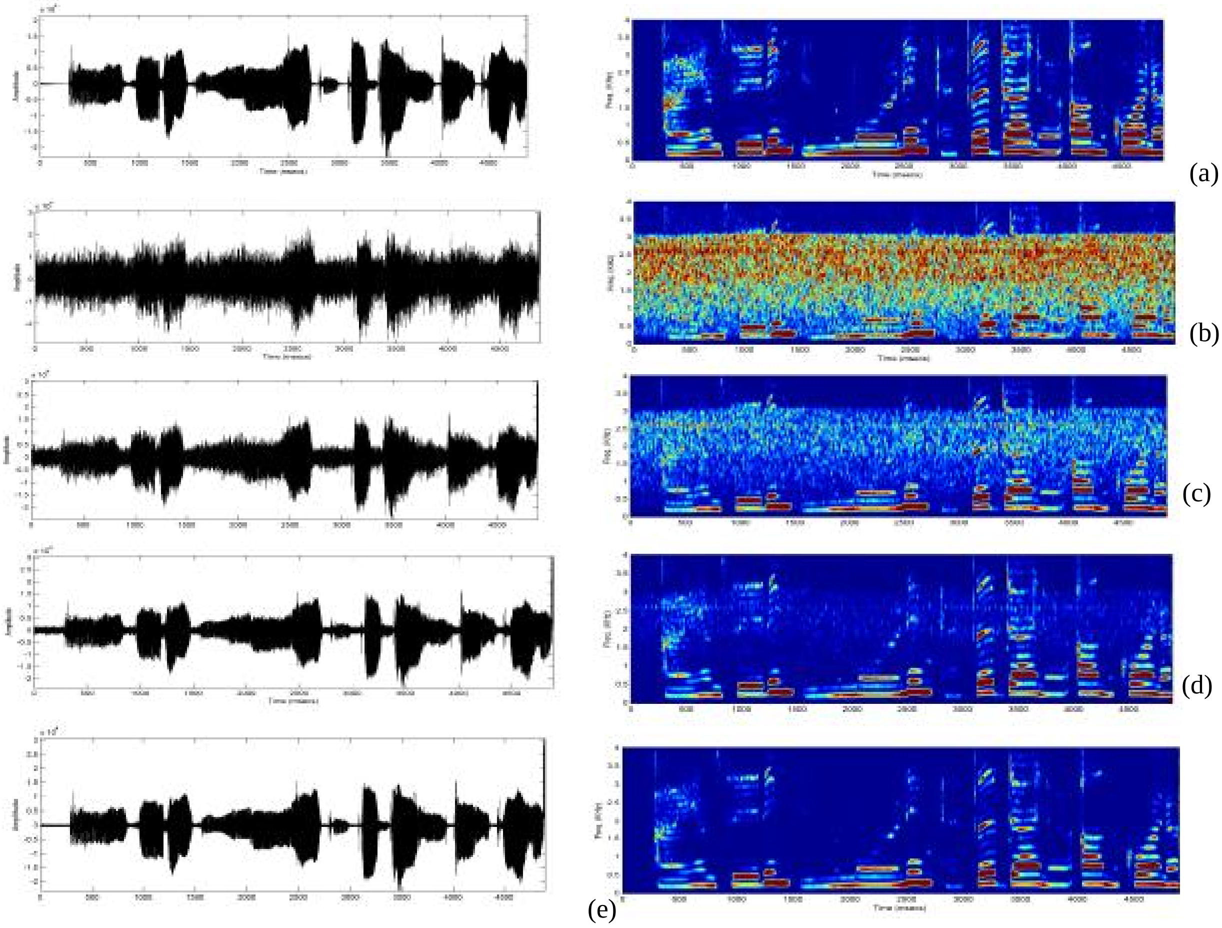



Fig.13 Time domain representation and Spectrogram of (a).Clean Speech (b) Noisy speech (HF channel Noise 6 dB)
(c) Enhanced speech using Hard thresholding (d) Enhanced speech using Soft thresholding

## VI. Conclusion

A time adaptive wavelet with trimmed thresholding process for de-noising speech from different noisy conditions has been presented. Enhancement results demonstrate that the proposed scheme shows better performance than hard and soft thresholding methods to de-noise the signal. It has shown that it can considerably enhance the noisy speech corrupted by white and colored noises. In case of signal de-noising using hard and soft thresholding a post processing band pass filter is used to reduce the noise effectively [11].But while using the trimmed thresholding method without post filtering a very good results were obtained. The competency of the proposed system to extract a clear and intelligible speech from various adverse noisy environments in comparison with other well-known thresholding methods has been demonstrated through both objective and subjective measurements. The quality and intelligibility tests were proved that the enhanced speech and clean speech have better similarities on time and frequency domain analysis. In spite of the powerful performance for additive white noise case, the proposed method produces better performance in real time noisy environment like F-16 cockpit, babble, HF channel noise. The proposed method was well suited to enhance the speech even for very strong noise condition since it has produced better performance than the existing algorithms. The limitation of this proposed scheme is the proper tuning of the parameter α for each noise conditions. Future work on this approach will include the adaptation of the parameter α and modified thresholding techniques for other noisy cases like street, helicopter, train noise and industrial noises etc,. Further this algorithm can be implemented in FPGA for enhancing speech in digital hearing aids.

## Acknowledgment

We gratefully acknowledge the cooperation of the people who participated in the subjective test. We would like to express our sincere gratitude to the management of Bannari

Amman Institute of Technology, Sathymangalam, India who provided the facilities to do our research.

## REFERENCES

[1] S.F.Boll, " Suppression of acoustic noise In speech using spectral subtraction", IEEE Trans. Acoustics. Speech. Signal processing, vol. 27,pp. 13-120, April 1979.

[2] Djigan, V.I., Sovka, P. and Cmejla, R., "Modified Spectral Subtraction based Speech Enhancement", Proc. of the 1999 IEEE Workshop on Acoustics Echo and Noise Control – IWAENC'99, Sept. 1999, Pensylvania, USA, pp. 64-67.

[3] Martin, R., "Spectral subtraction based on minimum statistics", in Proc. EUSIPCO, pp 1182-1185, September 1994.

[4] D.E.Tsoukalas , J.N.Mourjopoulos & Kokkinakis "Speech enhancement based on audible noise suppression" , IEEE Trans. Speech Audio . Proc., vol.5,pp. 479–514,Nov.1997.

[5] N.Virag,"Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech Audio Processing, vol. 7, pp. 126–137, Mar. 1999.

[6] Y. Ephraim & H.L. V Trees, "A signal subspace Approach for Speech Enhancement ", IEEE Trans. Speech and Audio Processing vol.3, no.4 pp.251-265, Sep 1995.

[7] Ephraim. Y., Malah. D., " Speech Enhancement Using a minimum mean-square error log-spectral amplitude estimator", IEEE Trans. Acoust. Speech Signal Processing ASSP-32(6), 1109-1121,1984.

[8] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," IEEE Trans. on Infomation Theory, vol. IT-38, pp. 617-643, 1992.

[9] D.L. Donoho ," De-noising by soft thresholding", IEEE Trans. on Information Theory, vol. 41 no. 3, 613-627, May 1995.

[10] Michael T.Johnson , Xiaolong Yuan , Yao Ren, " Speech signal enhancement through adaptive wavelet thresholding", Speech Communication, 49 (10),pp. 123-133, 2007.

[11] Sumithra M G , Thanuskodi K, Anitha M R, " Modified Time Adaptive Wavelet Based Approach for Enhancing Speech from Adverse Noisy Environment", ICGST International Journal on Digital Signal Processing, vol.9, issue 1,pp.33-40, April 2009.

[12] W.Seok and K.S.Bae, "Speech enhancement with reduction of noise components in the wavelet domain", in Proceedings of the ICASSP, pp.1323- 1326, 1997.

[13] Yasser Ghanbari, Mohammad Reza Karami Mollaei, " A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets", Speech Communication 48 , 927-940, 2006.

[14] Yao,J.,Zhang,Y.T., "The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations", IEEE trans.Biomed.Eng.49(11), 1299-1309,2002.

[15] Yao,J., "An Active model for otoacoustic emissions and its application to time frequency signal processing", Ph.D. thesis, The Chinese University of Hong Kong, Hong Kong, 2001.

[16] K.P. Soman, K.I. Ramachandran, " Insight in to wavelets - From theory to practice", Prentice - Hall of India Private Ltd, 2nd edition, 2006, pp 81-83 .

[17] J.Sgarofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database", NIST speech disc 1-1.1,oct 1990.

**Mrs.M.G.Sumithra**, born in Salem District, TamilNadu State, India in1973, received B.E. in Electronics and Communication Engineering from Govt. College of Engineering , Salem, India in 1994 and received M.E in Medical Electronics from College of Engineering, Guindy, Anna University, Chennai, India in 2001. She is currently Professor in Department of ECE, Bannari Amman Inst. of Technology in, Sathyamangalam, Tamilnadu, India and pursuing her research in Speech Processing. Her areas of interest are Signal Processing and Biomedical Engineering. She has published 20 technical papers in National and International conferences and one technical paper in National and two technical papers in International Journals.

**Dr. K. Thanushkodi**, born in Theni District, TamilNadu State,India in1948, received the BE in Electrical and Electronics Engineering from Madras University, Chennai. MSc (Engg) from Madras University, Chennai and PhD in Electrical and Electronics Engineering from Bharathiar University, Coimbatore in 1972, 1976 and 1991 respectively. He is currently Principal in Akshya College of Engineering and Technology , Coimbatore, Tamil Nadu. India. His research interests lie in the area of Computer Modeling and Simulation, Computer Networking, Signal Processing and Power System. He has published 40 technical papers in National and International Journals.