Speech Signal Enhancement Using Wavelet Threshold Methods

[1] P.Sunitha, Assoc. professor,
ECE, Pragati Engineering College,
A.P, India. sunitha4949@gmail.com

[2] T.Sekhar, Asst. Professor,
ECE ,Pragati Engineering College,
A.P, India, sekhar.tarra@gmail.com

Abstract— **This paper presents a Novel Approach for Enhancement of Speech Signal using different wavelet threshold methods. It is clearly observed that a signal which is obtained from the real world environment is often corrupted by means of noise. This noise effectively reduces the performance of the signal and it must be removed effectively for further processing of signal. This paper investigates the use of wavelet threshold methods i.e Minimax, SURE(Heuristic and Rigorous) and Square –root-log are investigated in Enhancement of speech signal corrupted by White Gaussian noise along with performance analysis of wavelets i.e. Haar,Db10, Coif5,sym5 at different levels of decomposition and performance analysis of threshold methods. This paper also investigates the use of wavelet shrinkage methods in Enhancement of speech signal corrupted by White Gaussian at five different noise levels i.e 0dB, 5dB, 10dB, 15dB, 20dB along with performance analysis of wavelets. Further comparative analysis has been made between Daubechies, Haar, Symlet, Coiflet. Biorthognal to explore the optimal wavelet selection for Enhancement of speech signal in terms of Peak Signal to Noise Ratio (PSNR). It has been found that Daubechies wavelet higher PSNR is achieved than others.**

Index terms: **Minimax, SURE (Heuristic and Rigorous) and Square –root-log.**

I. INTRODUCTION

In voice communications, speech signals can be contaminated by environmental noise and, as a result, the communication quality can be affected making the speech less intelligible. Furthermore, compression of noise speech with low bit rate vocoder may result in considerable quality degradation due to frequent estimation errors of speech production model parameters required by the vocoder. This problem can be reduced significantly by speech enhancement (or noise cancellation), which may enable more pleasant voice communication by suppressing the noise components in input signals.

Generally, it is assumed that the noisy speech signal is formed additively by speech and noise signals in which the noise is generated by environmental sources such as vehicles, street noise, babble, etc. Therefore, in real environments complete noise cancellation is not feasible as it is not possible to completely track varying noise types and characteristics that change with time. However, by assuming that the noise characteristics change slowly in the background noise levels producing more pleasant and intelligible speech quality. Speech enhancement techniques can help the speech model

parameter extraction process used in low bit – rate vocoders and hence they are becoming an integral part of low-bit rate speech coding systems.

Over the last three decades, many kinds of speech enhancement techniques adaptive filtering, model–based methods. The transform-based techniques, transforms the time –domain signal into other domains, suppress noise components, and apply the corresponding inverse transform to reconstruct enhanced speech signal. Discrete Fourier transformer (DFT), Discrete cosine transformer (DCT), Karhunen-Loeve transformer (KLT), and wavelet transformer (WT) are widely-known transformer methods. DFT – based technique have been intensively investigated based on short-time spectral amplitudes (STSA). A KLT-based technique, called signal subspace-based methods, decomposes the space into signal (or speech) and noise subspace by means of eigen decomposition, and then suppresses the noise component in the eigen values. DCT-based techniques are of lower computational complexity and higher frequency resolution than DFT-based methods. It is also possible to consider WT-based methods in order to simultaneously exploit the time and frequency characteristics of noisy speech signals. Adaptive filtering on the other hand, cancels the noise using adaptive filters such as the Kalman filter. A Kalman filter models noisy speech signals in terms of state space and observation equations, which present the speech production process and the noise addition model together with channel distortion, respectively. Kalman filters normally assume a white Gaussion noise distribution; however, Gibson *et al*. Proposed a generalization of Kalman-filtering over coloured noise signals. Finally, model-based technique classify the noise signals using an *a priori* speech model, such as hidden Markov and voiced/unvoiced models, and then conducts the enhancement depending on classified speech model. This method can be useful for improving the noise reduction performance for various kinds of speech signals. However, it requires extra training to build the model with intensive computation. In addition, it may exhibit model selection errors which cause significant speech quality degradation. Fundamentally, it is not easy to handle complicated speech signals with finite number of speech models.

Here is presented an investigation of use of wavelet threshold methods speech signal Enhancement. Unlike Fourier

Transform based spectral analysis for frequency domain analysis, Wavelet Transform (WT) is a powerful tool for signal processing to extract the relevant time-amplitude information from a signal for its multi resolution possibilities and is suitable for application to non stationary signals, whose frequency response varies in time[1,2]. Wavelet coefficients represent a measure of similarity in the frequency content between a signal and a chosen wavelet function [2]. By wavelet analysis at high scales, extract global information from a signal called approximations, whereas at low scales can extract fine information from a signal called details [3]. In signal processing, wavelets are used for many purposes [2], such as Enhancement, detecting trends, breakdown point's discontinuities in higher derivatives and self similarity in signals. Since speech signal is one dimensional and non-stationary, it is often corrupted by means of various kinds of noises present in environment like White Gaussian noise, Colored noise, Burst noise etc., In order to develop an efficient and effective technique of Enhancement of audio signal. It is necessary to separate useful contents of audio from noise by suitable method. An efficient method using wavelet shrinkage, which performs well over conventional frequency selective filter approach, is given in [4, 5]. To improve SNR, a block thresholding estimation method in presence of transients and harmonics is suggested in [6, 7]. Heuristic argument estimator removes noise effectively then Steins Unbiased Risk Estimator (SURE) by discarding purely noise coefficients in thresholding. An efficient improvement in SNR can be achieved by selec0tive smoothing at each scale of time frequency plot thus avoiding manual selection of coefficients [8]. Further to obtain minimum error and maximum signal to noise ratio wavelet with more varnishing moment is suggested in [8].

A. Basic Audio Theory

Sound is the vibration of an elastic medium, where gaseous, liquid or solid. These vibrations are a type of mechanical wave that has the capability to simulate human ear and to create a sound sensation in the brain. In air, sound is transmitted due to pressure vibrations at a rate of change that is called frequency. The difference between the extreme values of pressure represents its amplitude. Pressure vibrations in the range of 20 Hz to 20 kHz produce the sound which is audible to the human ear and this is more receptive when it is between 1kHz to 4kHz. In physical terms, the sound is a longitudinal wave that travels through the air due to vibrations of the molecules. Similar to light, sound wave can be reflected, absorbed, diffracted, or refracted.

Audio signals, which represent longitudinal variations of pressure in a medium, are converted into electrical signal by piezoelectric transducers. Transducers coverts the energy of a mechanical displacement into an electrical signal, either voltage or current. The main advantage of converting an audio signal into an electrical signal is that the signal can now be processed. An example is an audio signal obtained from the transducer that can be converted into an encoded digital data stream by using an analog-digital converter (ADC) and constitutes digital processing of analog signals. Alternatively, if a digital-analog converter (DAC) is applied to digital data stream, the audio signal transmits through an amplifier and a speaker. The process is shown symmetrically in the figure1, which identifies the important steps in digital audio signal.
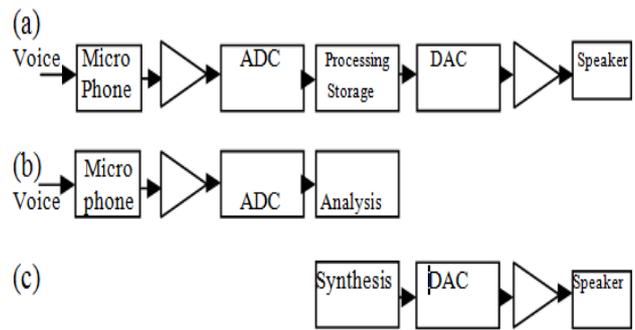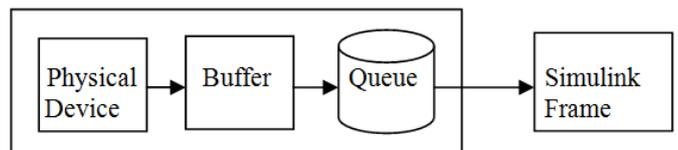

Fig. 1. Shows the process of digital processing of three types of audio signal. Part(a) represents a complete digital audio processing comprising (from left to right) a microphone, amplifier, ADC, digital processing material, DAC, amplifying section and speaker; an audio recognition system in (b), and a set of audio synthesis(c).

B. Recording audio signals in simulink/MATLAB

Once in the digital domain, these signals can be processed, transmitted or stored. We found that the Audio Device Block in Simulink enables experimentation and processing of digital signals. The from Audio Device Block reads audio Data from an Audio Device in real time.

From Audio Device Block

Fig.2. Shows the process of identifies the main steps in a digital audio processing system based in simulink software

C. Basic noise theory:

Noise is defined as an unwanted signal that interferes with the communication or measurement of another signal. A noise itself is an information-bearing signal that conveys information regarding the sources of the noise and the environment in which it propagates.

There are many types of sources of noise or distortions and they include:

1) *Electronic noise* such as thermal noise and short noise
2) *Acoustic noise* emanating from moving, vibrating or colliding sources such as revolving machines, moving vehicles, keyboard clicks, wind and rain.

3) *Electromagnetic noises* that can be interfere with the transmission and reception of voice, image and data over the radio frequency spectrum.
4) *Electrostatic noise* generated by the presence of a voltage,
5) *Communication Channel* distortion and fading and
6) *Quantization noise* and the lost data packets due to network congestion.

This paper is organized by various sections, section-II describes to brief introduction to wavelet transform, Section – III gives denoising scheme, section – IV gives Experiments and Results, section – V describes conclusion followed by References.

II. DISCRETE WAVELET TRANSFORM

Assume the observed signal
$$y(t) = s(t) + n(t)$$
Contains the original signal s(t) with additive noise n(t) as functions of time t to be sampled. Let W(.) and W^{-1}(.) denote the forward and inverse wavelet transform. Let D(., λ) denote the Enhancement operator with threshold λ. We intend to wavelet de noise y(t) in order to recover s^(t) as an estimate of S(t).

$$S = W(Y)$$
$$Z = D(Y,\lambda)$$
$$\hat{S} = W^{-1}(Z)$$

Similar to the Fourier series expansion, the DWT maps a continuous variable Ψ(t) into a sequence of coefficients, the resultant coefficients are called discrete wavelet transform of Ψ(t). Its representation involves the decomposition of the signals in wavelet basis function Ψ(t) given by

$$\Psi_{a,b}(t)= \frac{1}{\sqrt{a}}\ [(t-b)/a]\ \ a, b\ \varepsilon\ R \quad -(1)$$

Where a, b are called scale and position parameters as respectively.

The multi resolution analysis is given by S. Mallet and Mayer proves that any conjugate mirror filter characterizes a wavelet Ψ. The wavelet decomposition of a signal x(t) based a multi resolution theory can be obtained using filter [3], the filter based wavelet decomposition is shown in fig. 3.
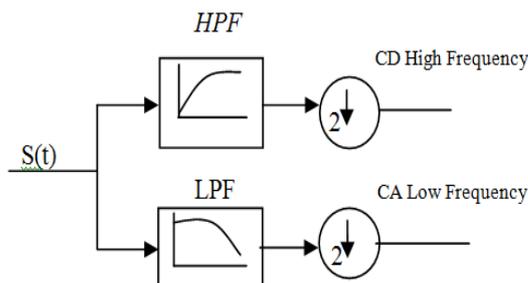

Fig 3.:single level wavelet decomposition(analysis).

The above arrangement has used two wavelet decomposition filters which are high pass and low pass respectively followed by down sampling by 2 producing half input data point of high and low frequency. The high frequency coefficients (CD) and low frequency coefficients are called approximate coefficients (CA) . The signal can be reconstructed back by inverse wavelet transform. The corresponding filter bank structure for reconstruction is shown in figure 4.
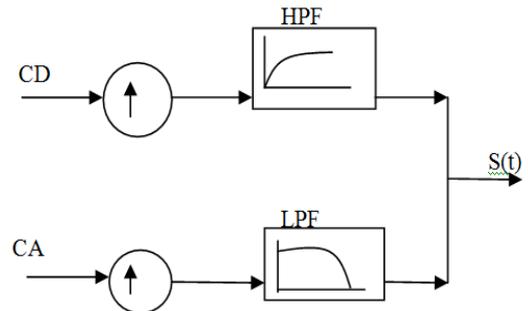

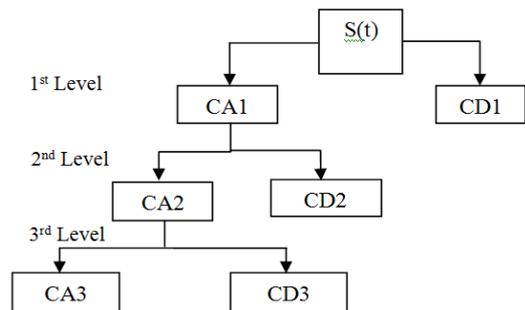Fig 4 : single level wavelet reconstruction(synthesis)

Fig 5: Three level wavelet decomposition tree.

The signal S(t) can be de composed into several levels. A three level wavelet decomposition tree is shown in figure 5.

III. WAVELET ENHANCEMENT SCHEME

Let us assume signal S(t) is corrupted by noise n(t)as y(t)= S(t) + n(t) where n(t) is white Gaussian noise . the wavelet based de noising scheme is shown in figure 6.
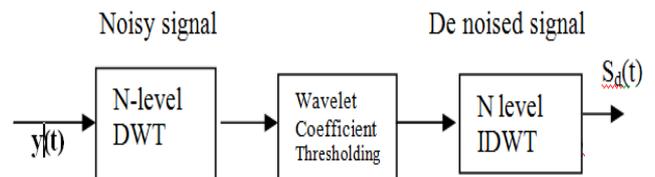

Fig 6: over all de noising scheme

As seen from figure 6 the de noising scheme involves three main steps.
(1) N-level wavelet decomposition of input noisy signal.
(2) Threshold estimation and thresholding of wavelet coefficients.
(3) N-level inverse wavelet transform for reconstruction of de noised signal.

De noising using wavelet involves thresholding, in which coefficients having less thresholding value compared to the specific threshold value (λ) are set to zero. This process helps to eliminate noise by keeping original signal characteristics. This is called hard thresholding. Another one i.e soft thresholding set the wavelet coefficients which are above the threshold.

Value as well as simply shrinks or scales other coefficients which are above the threshold value [1].
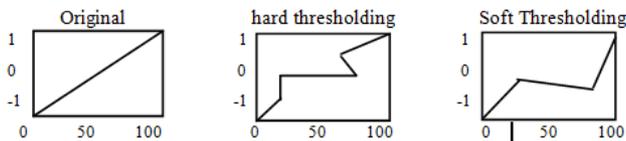

Fig 7: Soft & Hard Thresholding Methods

The hard and soft thresholding techniques are used for denoising process. The hard and soft threshold operations with threshold λ are defined using (2) and (3)

$n(d) = [(d_i)]$, $|d_i| \geq \lambda$ and $[0]$, $|d_i| \leq \lambda$ - (2)
$n(d) = [Sign(d_i).(|d_i| - \lambda)]$, $|d_i| \geq \lambda$ and $[0]$, $|d_i| \leq \lambda$ - (3)

Where d_i is the input and are noisy wavelet coefficient available in the detailed sub bands which are to be thresholded X is the threshold value and n(d) is the thresholded output which is used to estimate noiseless coefficients.

The hard can be unstable or more sensitive to small changes in the data while soft thresholding avoid discontinuities and is more stable than hard thresholding.

A. Threshold Methods

1) Minimax Criteria: this method finds threshold (λ) using Minimax principle. It uses a fixed threshold to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics to design estimators. Since the de-noised signal can be assimilated to the estimator of the unknown regression function. The minimax estimator is the option that realizes the minimum. Over a given set of functions of the maximum Mean Square Error(MSE). This procedure finds optimal thresholds. The threshold is given by

$$\lambda = \begin{cases} \sigma(0.3936 + 0.1829\,log2\,N) \\ 0 \end{cases} \quad N>32 \text{ and } N<32 \ - (4)$$

Where σ = median $\left(\dfrac{\omega}{0.6745}\right)$ and ω is the wavelet coefficient vector at unit scale and N is the length of signal vector.

2) Sqtwolog Criterion: the threshold values (λ) are calculated by universal threshold (square root log) method given by,

$$\lambda_j = \sigma_j \sqrt{2\,log(Nj)} \quad - (5)$$

Where N_j is the length of the noisy signal at j^{th} scale and σ_j is the Median Absolute Deviation (MAD) at the jth scale given by

$$\sigma_j = \frac{MADj}{0.6745} = \frac{median|\omega|}{0.6745} \quad - (6)$$

where ω represent wavelet coefficient at scale j.

3) Rigrsure: It is a soft threshold is evaluator of unbiased risk. Suppose $W = [\omega_1, \omega_2, \ldots \ldots \omega_N]$ is a vector consists of the square of wavelet coefficients from small to large. Select the minimum value $r_b(b^{th} r)$ from risk vector, wich is given as,

$$R = \{r_i\}_{I=1,2,\ldots\ldots} N = \frac{[N-2i+(n-i)\omega_i+\sum_{k=1}^{i}\omega_{ki}]}{N} \quad - (7)$$

as the risk value. The selected threshold is $\lambda = \sigma\sqrt{\omega_b}$ where, ω_b is the bth squared wavelet coefficient (coefficient at minimum risk) chosen from the vector W and σ is the standard deviation of the noisy signal.

4) Heursure: Threshold is selected using a combination of Sqtwolog and Rigrsure methods. If the signal to noise ratio is very small, the SURE method's estimation is poor. In such case, fixed form threshold of Sqtwolog method gives better threshold estimation [13]. Let threshold obtained from sqtwolog method is λ_1 and threshold obtained from Rigrsure is λ_2 then Heuristic SURE gives the threshold given by,

$$\lambda = \begin{cases} \lambda_1 & A > B \\ min(\lambda_1, \lambda_2) & A \geq B \end{cases} \quad - (8)$$

where, $A = \frac{s-N}{N}$ and $B = (\log_2 N)^{3/2}\sqrt{N}$. The N is length of wavelet coefficient vector and s is the sum of squared wavelet coefficients given as

$$s = \sum_{i=1}^{N}\omega_i^2. \quad - (9)$$

Threshold determination is an important problem. A small threshold may yield a result which may be noisy and large threshold can cut significant part of signal thus losing the important details of the signal.

IV. RESULTS AND DISCUSSIONS

This paper compares the performance analysis of wavelet threshold methods have been applied on the speech signal in English which is taken from a male speaker at a sampling frequency of 25 KHz. All the threshold methods are tested for white Gaussian noise .For performance comparison and measurement of quality denoising is calculated between speech signal S(t) and the denoised speech signal $S_d(t)$ is given by

$PSNR = 10 \log_{10} (s_{max}^2 /MSE)$ – (10)

Where s_{max} is the maximum value of signal and is given by
$Smax = max (max(s(t), max (S_d (t)))$

And MSE is Mean Square Error given by
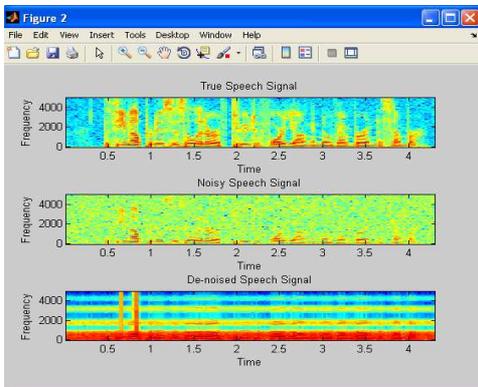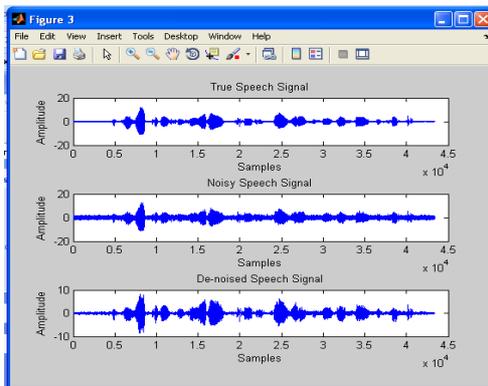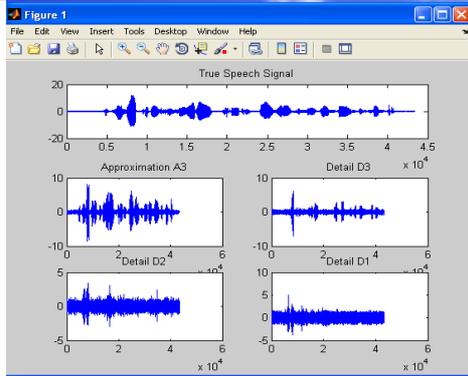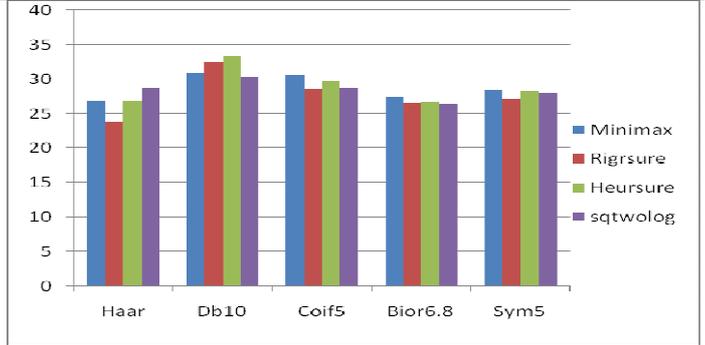$$MSE = \frac{1}{N} \sum_{K=1}^{N} [S_d(t) - S(t)]^2 \quad - (11)$$

Fig.9: Comparison of PSNR (dB) at level-1 Decomposition

TABLE II
COMPARISION OF PSNR (DB) AT LEVEL-2
DECOMPOSITION

Method	Haar	Db10	Coif5	Bior6.8	Sym5
Minimax	25.444	29.767	29.71	26.519	27.219
Rigrsure	22.644	27.643	27.29	24.459	26.029
Heursure	24.605	29.689	28.71	25.602	26.529
sqtwolog	24.802	30.065	28.02	24.609	25.592

Fig.10:.Comparision of PSNR (dB) at level -2 Decomposition

Fig. 8. MATLAB Simulation Results

TABLE I
COMPARISION OF PSNR (DB) AT LEVEL-1
DECOMPOSITION

Method	Haar	Db10	Coif5	Bior6.8	Sym5
Minimax	26.831	30.991	30.67	27.529	28.389
Rigrsure	23.846	32.536	28.62	26.672	27.184
Heursure	26.824	33.362	29.71	26.752	28.323
sqtwolog	28.826	30.382	28.73	26.506	28.042

TABLE III
COMPARISION OF PSNR (DB) AT LEVEL-3
DECOMPOSITION

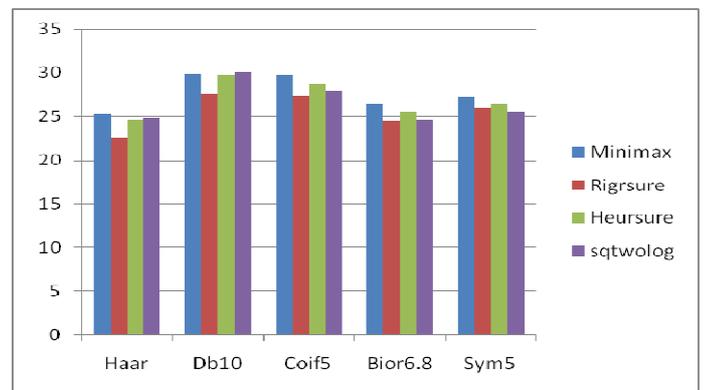| Method	Haar	Db10	Coif5	Bior6.8	Sym5
Minimax	22.492	26.062	25.77	24.514	25.389
Rigrsure	19.784	23.502	24.69	21.592	23.412
Heursure	25.605	27.034	26.82	22.689	25.029
sqtwolog	27.802	27.294	26.71	23.689	24.129

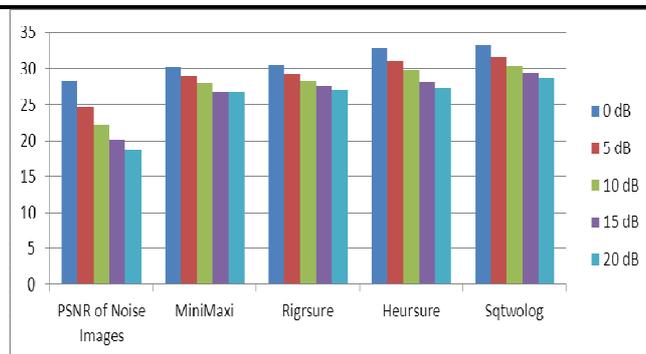Fig.11: Comparison of PSNR (dB) at level-3 Decomposition

TABLE IV
COMPARISION OF PSNR (DB) AT LEVEL-4
DECOMPOSITION

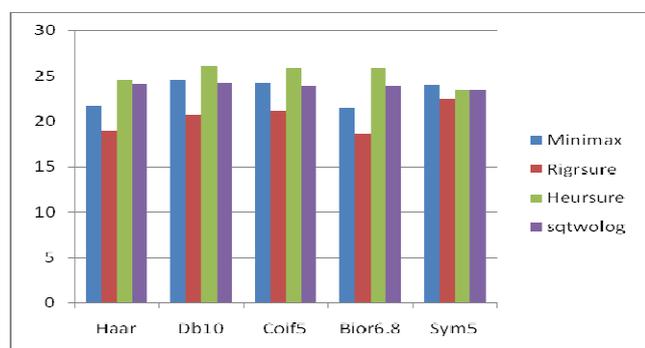| Method	Haar	Db10	Coif5	Bior6.8	Sym5
Minimax	21.782	24.519	24.17	21.523	23.972
Rigrsure	18.962	20.609	21.19	18.563	22.386
Heursure	24.502	26.062	25.82	25.864	23.409
sqtwolog	24.142	24.209	23.81	23.824	23.424

Fig.12: Comparison of PSNR (dB) at level-4 Decomposition
Haar: Haar wavelet, Db: Doubechies wavelet,
coif: Coiflet wavelet, sym: Symlet wavelet

TABLE V
COMPARATIVE ANALYSIS OF THRESHOLD
METHODS

S.No	Noise Levels	PSNR of Noise Signal	PSNR of denoised signal using different algorithm			
			MiniMaxi	Rigrsure	Heursure	Sqtwolog
1	0 dB	28.143	30.288	30.582	32.875	33.246
2	5 dB	24.599	28.857	29.182	31.104	31.607
3	10 dB	22.102	27.914	28.238	29.865	30.428
4	15 dB	20.170	26.723	27.531	28.027	29.487
5	20 dB	18.588	26.702	26.972	27.306	28.611

Fig.13. Comparative analysis of threshold methods

TABLE VI
COMPARATIVE ANALYSIS BETWEEN DIFFERENT
WAVELET FAMILIES BY USING SQTWOLOG METHOD
IN TERMS OF PSNR.

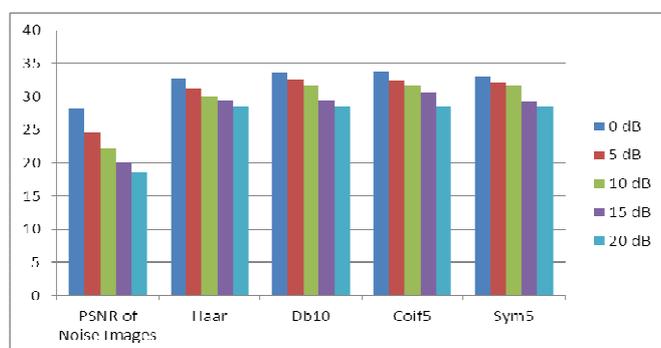| S.No	Noise Level	PSNR of Noise signal	Haar	Db10	Coif5	Sym5
1	0 dB	28.143	32.814	34.606	33.712	33.062
2	5 dB	24.599	31.204	32.625	32.354	32.204
3	10 dB	22.102	30.017	31.705	31.642	31.617
4	15 dB	20.170	29.308	30.362	30.640	29.224
5	20 dB	18.588	28.467	29.562	28.526	28.437

Fig.14. Comparative analysis between different wavelet
families by using sqtwolog

In addition, the sqtwolog method is performed by actually performing the comprehensive study of daubechies wavelets (Db -2 to Db – 10), haar wavelet, coiflet wavelets (coif-1 to coif-5),symlet wavelets(sym-2 to sym-5) but here are only the results of sym 5 wavelet.

A comparative analysis has been performed between Minimaxi/Rigrsure, Heursure and sqtwolog and the results in the form of PSNR are given in table 1,2,3,4 for 5 different wavelets. It is observed as the level of decomposition is increased from level 1 to level-4, the PSNR values of noise signals go on reducing and improved in terms of PSNR for de-noised signals. When Db10 is applied at all the decomposition levels, significant increase in PSNR is obtained in comparison

Haar, Daubechies, Coiflet, Symlet, Biorthognal. This demonstrates a significant improvement in signal quality and comparative analysis has been performed between Minimaxi/Rigrsure, Heursure and sqtwolog and the results in the form of PSNR is given in table5 for 5 different levels of white Gaussian noise for sym-5 wavelet. It is observed as the level of Gaussian noise increased from 0dB to 20dB. The PSNR values of noise signals go on reducing and improved in terms of PSNR for denoised signals. When sqtwolog is applied on all the 5 noisy signals significant increase in PSNR is obtained in comparison to Minimaxi Rigrsure and Hersure. This demonstrates a significant improvement in signal quality.

In addition, the threshold method is performed by actually performing the comprehensive study of Daubechies wavelets (Db -2 to Db – 10), Haar wavelet, Coiflet wavelets (Coif-1 to Coif-5), Symlet wavelets (sym-2 to sym-5) and Biorthognal.

V. CONCLUSION

In this paper the performance analysis of different wavelet threshold methods in di-noising is investigated along with 4 different levels of decomposition with white Gaussian noise. The result shows that as the level of decomposition increases the value of PSNR decreases. Thus the lower levels of decomposition can be preferred as seen from the result performance of db10 better than as compared to other wavelets. Haar wavelet is poorer as compared to others also it gives unwanted distortion in the reconstructed signal. Overall from the results, it is clear that db10 is best suitable giving minimum distortion and maximum PSNR.

REFERENCES

1) S.Mihov, D.Doychev, R.Ivanov practical investigation of specific types of noice signals for the purpose of their hearing aid devices, processsings of ICEST – 2009 vo12 PP 399-402, viliko Tarnovo 2009.
2) G.Openherim J.M Poggi, M.Misiti, Y.Misiti wavelet tool box themath works, Inc., Natick Massachersetts 01760, April 2001.
3) Fundamentals of Wavelets, Theory Algorithms and Applications by J.C.Goswamy, A.K.Chan
4) Domoho D.L. and Johnson I.M., "Enhancement by soft thresholding", IEEE Transaction on Information Theory, 1995,41,(3) PP.613 – 627.
5) D.L.Donoho and I.M.Johnstone "Minimax Estimation via wavelet shrinkage" Annals of Statistics, 26(3):879 921, 1998.
6) Guoshen Yu, Stephane Mallat, Emmanuel Bercy, "Audio Denoising by Time Frequency Block Thresholding", IEEE transaction on signal processing. Vol. 56, No.5, May 2008.
7) S.Sreekanth, P.Dinesh Khanna, P.Uma Mallikarjuna, "An Efficient Noise Reduction by using Diagonal and Non diagonal Estimation Techniques", Proceedings of the International Conference on Computationgal Intelligence, 2010. PP 393 – 398.
8) A. Sony John, Uday B. Desai, "signal Enhancement using the wavelet Transform and Regularization", 1997, IEEE.
9) Denoising speech signals by wavelet Transform. Annual Journal of Electronics, 2009. ISSN 1313-1814.
10) Performance Analysis of wavelet Thresholding methods in Enhancement of Audio signals of some Indian Musical Instruments. ISSN:0975 – 5462, Vol. No. 05th May,2012. PP. 2047 – 2050.

P.Sunitha received the B.Tech degree in Electronics and Communication Engineering from JNTU college of Engineering , Kakinada in 2006 and M.Tech in Digital Electronics and Communication Systems (DECS) from JNTU Kakinada. She is working towards the Ph.D degree from JNTUK.

She is currently working as an Assoc. Professor in the dept. of Electronics and Communication Engineering in Pragati Engineering College, Andhra Pradesh, India. Her Research interest includes Speech processing,Signal Processing, Image Processing and VLSI.

T.Sekhar received the B.Tech degree in Electronics and Communication Engineering from Chaitanya Institute Science & Technology, Madhavapatnam, Kainada in 2010 and M.Tech in Embedded Systems (ES) from Pragati Engineering College, Surampalem in 2013.

He is currently working as an Asst. Professor in the dept. of Electronics and Communication Engineering in Pragati Engineering College, Andhra Pradesh, India. He Research interest includes Embedded System, Signal Processing, and Image Processing.