Detailed Study and Analysis of Audio Denoising Techniques by T-F Block Thresholding using STFT and Wavelets

Prashant S. Korke[1]
P. G. Dept.,
MBES College of Engineering, Ambajogai,
India, 431517
[1]pskorke@gmail.com

Veeresh G. Kasabegaudar[2]
P. G. Dept.,
MBES College of Engineering, Ambajogai,
India, 431517
[2]veereshgk2002@rediffmail.com

Abstract— This paper presents noise minimization by adaptive time frequency block thresholding procedure using discrete wavelet transform to achieve better SNR of the various audio signals. Removing noise from the audio signals requires no diagonal processing of time frequency coefficient to avoid musical noise. First, a block thresholding estimation procedure is introduced which adjusts all the parameters adaptively to signal property by minimizing a stein estimation of the risk. In this we are comparing performance of BT with LSA and PS on various audio signals such as Mozart, Speech, Timit with different SNR values. Later we demonstrate noise reduction by adaptive soft/hard thresholding procedure using DWT. We represent a comparative analysis of performance of various types of wavelet i.e. Haar,db10,Coinf5 and sym5 for denoising the Mozart, Speech, Timit signals in the presence of white Gaussian noise. Simulation results are performed in MATLAB 7.9.0(R2009b). Finally we compare the results of BT,soft thresholding& hard thresholding.

Keywords— *Audio denoising, Block thresholding, DWT,Softthresholding, Hard thresholding, PSNR.*

*****

I. Introduction

Background environment noise deteriorate the audio/speech signals. Diagonal time-frequency algorithm attenuates the noise by processing each window Fourier or wavelet coefficient independently with empirical wiener [1], power subtraction [2-4] and thresholding operators [5]. Ephraim and Malah [6, 7] suggested musical noise can be attenuated with non-diagonal time frequency estimators that regularizes the estimation by time frequency coefficients. First half introduces nondiagonal audio denoising algorithm through adaptive time-frequency block thresholding[8]. Block thresholding has been introduced by Cai and Silverman in mathematical statistics [9, 10]. For audio time-frequency denoising, block thresholding is effective in noise reduction. Block parameters are automatically adjustedby minimizing a stein estimator of the risk [11] which is obtained from noisy signal values. This improves the SNR and perceived quality with respect to state of the art audio denoising algorithm.

Another method DWT has become popular tool for denoising the audio signals. Wavelets are front end processor for speech recognition by exploiting the time-frequency resolution of wavelet transform. This mother wavelet is based on the Hanning window. Audio signals are often affected by white noise and it is most difficult to detect and remove. This noise is removed by discarding small coefficients. White noise can be handled either by hard and soft thresholding. Hard thresholding sets all the wavelet coefficients below given threshold value equal to zero. Soft thresholding smoothens the signal by reducing wavelet coefficient by equating threshold value and modifies the audio signal. In this section we use multiresolution wavelet filter bank such as Haar, db10, Sym5, coif5. The filter is selected based on noise level and other parameters in the signal. For getting optimum denoising result better threshold level has to be estimated.

Recently various wavelet based methods have been proposed for the purpose of speech denoising. This method is based on thresholding in the signal that each wavelet coefficient of the signal is compared to given threshold.

Section 2 discusses on the methods for audio denoising. In Section 3, audio denoising using short time Fourier transform (STFT) is covered. The concept of denoising with wavelet transform is discussed in Section 4. Results and discussions of the study are presented in Section 5. Finally, conclusions of the study are presented in Section 6.

II. METHODS FOR AUDIO DENOISING

Diagonal time-frequency audio denoising algorithms attenuate the noise by processing each window coefficient independently, with empirical Wiener, power subtraction or thresholding operators. The basic methods of audio denoising are “Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor” and “Audio signal denoising with adaptive block attenuation”. Depending upon the SNR considered, the audio denoising techniques are basically divided in to

- Diagonal Estimation Techniques
- Nondiagonal Estimation Techniques

A. Diagonal Estimation

Basic time-recurrence denoising calculations process every weakening variable just from the relating loud coefficient and are along these lines called corner to corner estimators. These calculations have a restricted execution and produce a musical commotion. In Diagonal Estimationthe Posterior SNR is considered. Back SNR is the SNR of the Audio Noisy Signal.The SNR Signal-to-Noise-Ratio(S/N) is a specialized term used to portray the nature of sign.

Corner to corner estimators of the SNR are figured from the a posteriori SNR. The constriction variable of these corner to corner estimators just relies on loud coefficients with no time-recurrence regularization. The subsequent constricted coefficients in this manner absence of time-recurrence normality. It produces disconnected time-recurrence coefficients which reestablish segregated time-recurrence structures that are seen as a musical clamor.

B. Nondiagonal Estimation

To diminish musical commotion and additionally the estimation chance, a few creators have proposed to appraise apriori SNR with a period recurrence regularization of the posteriori SNR.Resulting constriction considers in this manner rely on the information values in an entire neighborhood of and the subsequent estimator is said to be nondiagonal. This nondiagonal estimation method is known as Block Thresholding (BT).

In this paper, execution of Mozart, speech,& timit Signal with various SNR qualities is finished. " Mozart" is a musical extract that contains generally brisk notes played by a performance oboe. Mozart Signal is inspected at 11 kHz. The Mozart Signal is tainted by Gaussian repetitive sound diverse plentifulness. Brief time Fourier changes with half-covering windows utilized as a part of the investigations. These windows are the square foundation of Hanning windows of size 50 ms for "Mozart signal". The fundamental stride used to denoise the musical commotion signal appeared in the accompanying square outline.

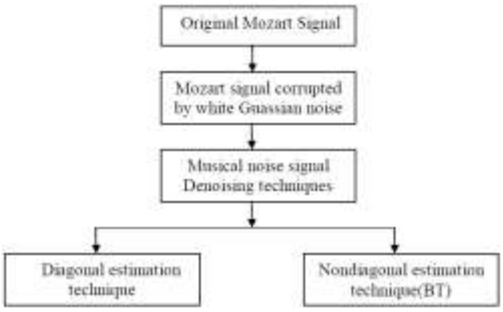

Figure1: Block diagram of denoising musical noise signal.

Algorithm for Diagonal estimation using MMSE-LSA estimator procedure [12]:

1. Estimate the initial silence period for estimating initial noise parameter.
2. Apply overlap add method.
3. Consider initial noise power spectrum mean and variance.
4. Calculate magnitude spectrum distance.
5. Determine posterior SNR.
6. Apply decision directed method for a priori SNR and determine log spectral MMSE.
7. Determine reconstructed time domain signal.
8. Obtain denoised signal.
9. Compare the SNR of the musical noise signal and the denoised signal.

Similarly, the algorithm for Nondiagonal Estimation using adaptive time-frequency Block Thresholding procedure [12]**:**

1. Determine the SNR of the musical noise signal.
2. Apply Hanning window.
3. Apply half overlapped window
4. Apply STFT.
5. Determine Block size by considering Stein Unbiased Risk Estimator.
6. Apply Block Thresholding.
7. Apply inverse STFT.
8. Obtain denoised signal.
9. Compare the SNR of the musical noise signal and the denoised signal.

III. SHORT TIME FOURIER TRANSFORM

The short-time Fourier transform (STFT) is a Fourier related change used to decide the sinusoidal recurrence and stage substance of neighborhood areas of a sign as it changes after some time. The Fourier change expect the sign is broke down over record-breaking of an unending length. This implies there can be no understanding of time in the recurrence space, thus no understanding of a recurrence changing with time. Scientifically, recurrence and time are orthogonal can't blend one with the other. However, we can without much of a stretch comprehend that some signs do have recurrence parts that change with time. A piano tune, for instance, comprises of various notes played at various times or discourse can be heard as having pitch that ascents and falls after some time. The Short Time Fourier Transform (STFT) tries to assess the way recurrence content changes with time. STFT changes over the sign from time space to time-recurrence area. It gives ghostly parts.

A.Time Frequency Audio Denoising

Time-frequency audio denoising procedures compute STFT of the noisy signal and process the resulting coefficients to reduce the noise. Numerical experiments are performed with STFT that are used in audio processing.

The audio signal f is contaminated by a noise ϵ that is modeled as zero-mean Gaussian process independent of f:

$$y[n] = f[n] + \epsilon[n], \quad n=0, 1, \ldots\ldots N-1 \qquad (1)$$

Where $y[n]$ =audio noisy signal, f[n]=audio signal, and ϵ[n]=noisy signal.

A time-frequency transform decomposes audio signal y over a family of time-frequency atoms $\{g_{l,k}\}_{l,k}$ where l and k are the time frequency indices. The resulting coefficients written as

$$Y[l,k] = \langle y, g_{l,k} \rangle = \sum_{n=0}^{N-1} y[n] g_{l,k}^{*}[n] \quad (2)$$

Where * denotes the conjugate. In this paper these time-frequency atoms define a tight frame [18, 43] there exists A>0 such that [12]

$$||y||^2 = \frac{1}{A} \sum_{l,k} |\langle y, g_{l,k} \rangle|^2 \quad (3)$$

This implies simple reconstruction formula [12]

$$y[n] = \frac{1}{A} \sum_{l,k} Y[l,k] g_{l,k}[n] \quad (4)$$

The constant A is a redundancy factor. Short- time Fourier atoms can be written: $g_{l,k}[n] = w[n-lu] \exp(\frac{i2\pi kn}{K})$ where $w[n]$ is a time window of size K . $l and k$ areinteger time and frequency indexes. A denoising algorithm gives time-frequency coeffiients by multiplying each of them by an attenuation factor $a[l,k]$ to reduce th noise component. The resulting "denoised" signal estimator is[12]:

$$\hat{f}[n] = \frac{1}{A} \sum_{l,k} \hat{F}[l,k] g_{l,k}[n] = \frac{1}{A} \sum_{l,k} a[l,k] Y[l,k] g_{l,k}[n] \quad (5)$$

The noise coefficient variance is supposed to be estimated with methods such as [5, 13, 14].

To reduce musical noise as well as estimation risk, the priori SNR $\xi[l,k]$ with time-frequency regularization of a posteriori SNR $\gamma[l,k]$. The resulting nondiagonal estimator is[12]:

$$\hat{f}[n] = \frac{1}{A} \sum_{l,k} a[l,k] Y[l,k] g_{l,k}[n] \quad (6)$$

The decision directed SNR estimator is introduced in Ephraim and Malah[6] with first order recursive time filtering[12]:

$$\hat{\xi}[l,k] = \alpha \hat{\xi}[l-1,k] + (1-\alpha)(\gamma[l,k]-1) \quad (7)$$

This decision directed SNR estimator has been applied with various attenuation rules such as empirical Wiener estimator [15], Ephraim and Malah's minimum mean square error spectral amplitude MMSE-SA [6], log spectral amplitude estimator MMSE-LSA [7] and Wolfe and Godsill's minimum mean square error spectral power estimator MMSE-SP [16] that are derived from Guassian speech model [6, 7, 17-21].

B. Hanning Window

Windowing of a simple waveform causes its Fourier transform to have non-zero values (commonly called spectral leakage) at frequencies other than ω. It tends to be worst (highest) near ω and least at frequencies farthest from ω. If there are two sinusoids, with different frequencies, leakage can interfere with the ability to distinguish them spectrally. If their frequencies are dissimilar, then the leakage interferes when one sinusoid is much smaller in amplitude than the other. That is, its spectral component can be hidden by the leakage from the larger component. That tradeoff occurs when the window function is chosen. The Hanning and Hamming windows, both of which are in the family known as "raised cosine" windows, are respectively named after Julius von Hanning and Richard Hamming. The term "Hanning window" is sometimes used to refer to the Hanning window. Hanning window smoothens the transmission of a window in speech signal. It reduces the ripples to the maximum extent. The Hanning window function is given by:

$$w(n) = 0.5\left(1 - \cos 2\pi\left(\frac{n}{N}\right)\right), 0 \leq n \leq N \quad (8)$$

The Window length is L=N+1.

The performance of Hanning window is given by the width of main lobe and magnitude of the side lobe. The width of main lobe in window spectrum is 8π/N and the maximum sidelobe magnitude in window spectrum is fixed at -41dB. In hanning window the main lobe width is same as other window techniques but the magnitude of side lobe is reduced to the maximum possible extent. So hanning window is preferable.If Fourier transform is applied to a non-stationery signal, it results in occurrence of more side lobes. To reduce the no. of side lobes "Hanning Window" is used in STFT. The main advantage of windowing is that it is reasonably straight forward to obtain minimal computational effort and number of side lobes will be reduced. A Hanning window has continuous derivatives. A normalized Hanning window has a sharper transition. It has the advantage of generating a tight frame STFT [12].

C. Thresholding Types

Thresholding gives amplitude separation. To well separate signal and noise, thresholding is used. The purpose of a filter is for frequency separation and frequency signal restoration. So for amplitude separation thresholding is used. Depending upon the type of noise present in the signal, the thresholding is determined basically in two forms is[12]

- Soft Thresholding
- Hard Thresholding

In Soft thresholding, the coefficients which are within the Threshold value are consider as zero and subtract the Threshold value from the coefficients which are above the Threshold value. Depending upon the changes in the noise signal threshold value will change in soft thresholding.

ISSN: 2321-8169
861 - 870

In Hard Thresholding, the coefficients which are within the Threshold value are consider as zero and the coefficients which are above the Threshold value remain same and are considered as actual coefficients of the signal. In hard thresholding the threshold value is fixed. The basic difference between the Soft Thresholding and Hard Thresholding is as shown below with an example.

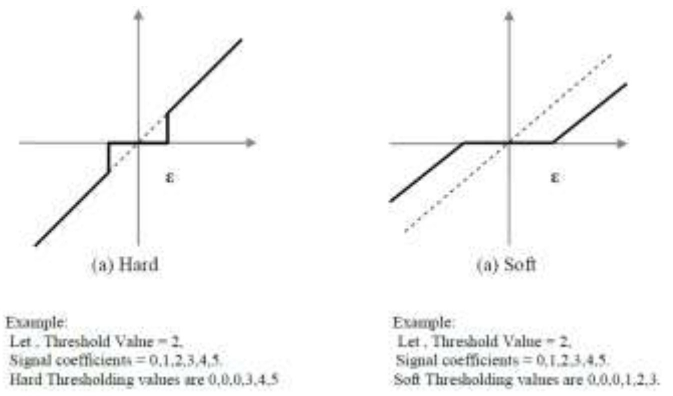

Figure 2: Thresholding techniques a) Hard b) soft thresholding[12].

D. Threshold Limits

Many methods for setting the threshold have been proposed. The most time-consuming way is to set the threshold limit on a case-by-case basis. The limit is selected such that satisfactory noise removal is achieved. Two rules are generally used for thresholding the coefficients (soft/hard thresholding). Hard thresholding sets zeros for all wavelet coefficients whose absolute value is less than the specified threshold limit. In this study, we adopt the hard thresholding method.

E. Derivation of Time -Frequency Block Thresholding

Block threholding was introduced in statistics by Cai and Silverman[9, 10, 22] and studied by Hall et al.[23-25] to obtain signal estimators. For audio signal denoising, we explain an adaptive block thresholdingnondiagonal estimator that automatically adjusts all parameters.

A time –frequency block thresholding estimator gives power subtraction by calculating a single attenuation factor over time-frequency blocks. The time-frequency plane $\{l, k\}$ is divided in I blocks B_i whose shapes chosen arbitrarily. The signal estimator $\hat{f}$ is calculated from noisy data y with constant attenuation factor a_i over each block B_i

$$\hat{f}[n] = \sum_{i=1}^{I}\sum_{(l,k)\in B_i} a_i Y[l,k]\, g_{l,k}[n]. \quad (9)$$

The resulting attenuation factor a_i is computed with power subtraction estimator

$$a_i = \left(1 - \frac{\lambda}{\bar{\xi}_i + 1}\right) \quad (10)$$

A block thresholding estimator can thus interpreted as nondiagonalstimation derived from averaged SNR estimations over blocks.

F. Block Thresholding Risk and Choice of λ

An upper bound of the risk of block thresholding estimator is computed by analyzing bias and variance terms separately. Upper bound of the oracle risk with blocks is always larger than oracle risk without blocks. If the frame is an orthogonal basis, where all the blocks B_i have same size $B^{\#}$ and the noise is Gaussian white noise with varianceσ^2,Cai [4] proved that [12]

$$r = E\left\{\left\|\hat{f} - f\right\|^2\right\} \le 2\lambda R_{bo} + 4N\sigma^2 prob\{\bar{\epsilon}^2 > \lambda\sigma^2\} \quad (11)$$

Where $prob\{\}$ is probability measure. When λ increases first term increases and variance term decreases. Similarly, when the block size $B^{\#}$ increases the oracle risk R_{bo} increases whereas the variance decreases. Adjusting λ and block size $B^{\#}$ interpreted as an optimization between the bias and the variance of block thresholding estimator. The parameter λ is depending upon by adjusting residual noise probability

$$Prob\{\bar{\epsilon}^2 > \lambda\sigma^2\} = \delta \quad (12)$$

The probability δ is perceptual parameter. We set δ=0.1%

TABLE I: THRESHOLDING LEVEL λ CALCULATED FOR DIFFERENT BLOCK SIZE [12]

$B_i^{\#}$	4	8	16	32	64	128
λ	4.7	3.5	2.5	2.0	1.8	1.5

Let $B_i^{\#}$=L_i*W_i is rectangular block size$B^{\#}$, where L_i ≥2 and $W_i \ge 2$ are repectively the block length in time and block width in frequency. Table I [12] gives values for a frequency width $W_i \ge 2$ where λ takes nearly same values for $W_i = 1$and $W_i = 2$.

The partition of macro blocks in to blocks of different sizes is as shown below:

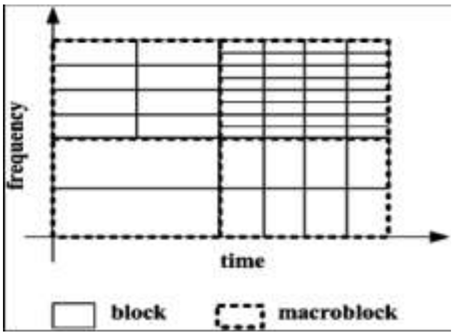

Figure 3: Partition of macroblocks into different sizes[12].

The adaptive block thresholding chooses the sizes by minimizing an estimate of the risk. The risk cannot be calculated since is unknown, but it can be estimated with Stein risk estimate. The adaptive block thresholding groups coefficients in blocks whose sizes are adjusted to minimize the Stein risk estimate and it attenuates coefficients in those blocks.

Low amplitude noise generated by an estimation procedure [2, 26] and attenuation factor is

$$\tilde{a_M}[l,k] = \max(\tilde{a}[l,k], a_0) \quad (13)$$

Where, $0 < a_0 \ll 1$ is minimum attenuation factor of the noise.

IV. WAVELET TRANSFORM

A wavelet is a small wave, which has its energy concentrated in time to give a tool for the analysis of non-stationary, or time varying phenomena [27].Wavelet transforms convert a signal into a series of wavelets and provide a way for analyzing waveforms, bounded in both frequency and duration. By using wavelet Transform, we can get the frequency information which is not possible by working in time-domain. There are many different wavelet systems that can be used effectively. It is a system which is set of building blocks to represent a signal. Wavelet transform perform analysis and calculation of the coefficient from the signal efficiently. Wavelet analysis is continuous analysis to express signal appearance. The continuous wavelet transform performs a multi-resolution analysis by contraction and dilation of the wavelet functions and discrete wavelet transform (DWT) uses filter banks for the construction of the multiresolution time-frequency plane [28]. The filter choice is depend upon the noise level and other parameters. For good denoising result, a good threshold level has to be estimated. The wavelet function and decomposition level also plays an important role in the quality of the denoise signal [28]. There are various wavelet based methods have been proposed for the purpose of denoising. The wavelet split coefficient method is a speech denoising procedure to remove noise by shrinking the wavelet coefficients [29, 30]. An improved method using adaptive thresholding is suggested in [27]. To improve SNR, a block thresholding estimation method in presence of transients & harmonics, is suggested in [31, 32]. For the denoising speech signals in wavelet we use different types of wavelets such as Haar, coif5, Db10, bior3.3, and sym5.

Wavelet function given by S. Mallat[33] is,

$$\varphi_{m,n}(k) = 2^{\frac{-m}{2}}\varphi(2^{-m}k - n) m, n \epsilon Z \quad (14)$$

The discrete wavelet transform as given by [34],

$$DWT(m,n) = \left(f, \varphi_{m,n}\right) = 2^{\frac{-m}{2}} \sum_{k=-\infty}^{\infty} f(k).\varphi(2^{-m}k - n) \quad (15)$$

Now we find threshold value that will use to remove noise from noisy signal, but also recover original signal efficiently. If the threshold value is too high, it will also remove contents of the original signal and if the threshold value is too low, denoising will not work properly.

One of the first method for selection of threshold was proposed by Donoho and Johnstome[31] and it is called universal threshold.

$$thr = \sigma n \sqrt{2 log2(N)} \quad (16)$$

where N denotes no. of samples of noise and σn standard deviation of noise. Universal threshold was proposed in [32] and modified with factor 'k' in order to obtain high quality output signal

$$thr = k.\sigma n \sqrt{2 log2(N)} \quad (17)$$

Threshold value obtained by Eq. (17) gives better results.
The soft and hard thresholding methods used to estimate wavelet coefficients in wavelet threshold denoising [28]. Hard thresholdings zeroes small coefficients and soft thresholding soften the coefficients exceeding the threshold by lowering value by threshold value. The hard threshold signal is x if $x \geq thr$ and is 0 if $x < thr$, where thr is threshold value. Soft thresholding is an extension of hard thresholding. In this first setting all the elements to zero whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towords 0. If $x \geq thr$,soft threshold signal is $(sign(x).(x - thr))$. Fig. 4 shows the comparison between hard and soft thresholding (threshold value '$thr = 0.5$').

$$T_{Hard}(x) = \begin{cases} x, & |x| \geq thr \\ 0, & |x| < thr \end{cases} \quad (18)$$

$$T_{soft}(x) = \begin{cases} sign(x).(x - thr) & x \geq thr \\ 0 & -thr \leq x < thr \\ sign(x).(x + thr) & x < -thr \end{cases} \quad (19)$$

ISSN: 2321-8169
861 - 870

Original signal   soft threshold   hard threshold

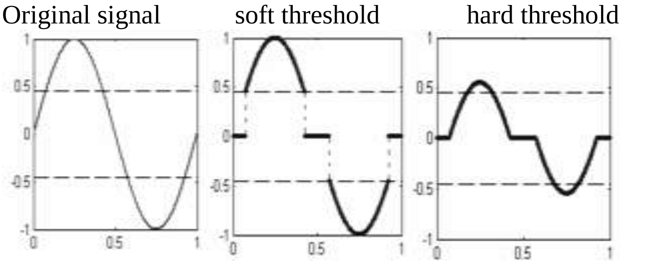

Figure 4: Original signal, hard thresholding and soft thresholding[35].

Denoising algorithm scheme is showed in Fig. 5. Inverse DWT is applied to get denoise time domain signal.

Algorithm for Adaptive time-frequency Block Thresholding using DWT procedure [35]:

1. Determine the SNR of the musical noise signal.
2. Apply Hanning window.
3. Apply half overlapped window
4. Apply DWT.
5. Determine Block size by considering Stein Unbiased Risk Estimator.
6. Apply soft Thresholding.
7. Apply inverse DWT.
8. Obtain denoised signal.
9. Compare the SNR of the musical noise signal and the denoised signal.

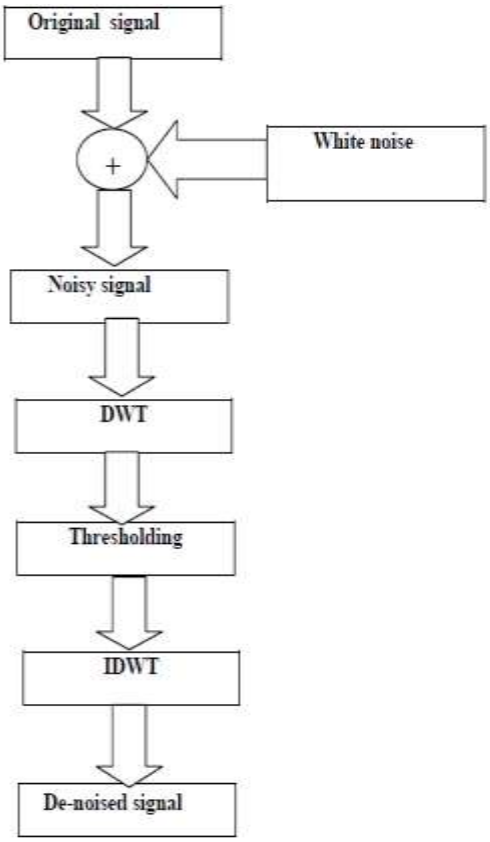

Figure 5: Flow chart of denoise algorithm [35].

V. RESULTS AND DISCUSSIONS

A. Denoising Results of BT with STFT

The simulations presented below have been performed on various types of audio signals. Piano is an example that contains a single clear stroke,Mozart is a musical excerpt that contains relatively quick notes played by a solo oboe, Timit-M and Timit-F are respectively male and female utterances taken from Timit database[36]. Timit signal is sampled at 16 KHz and others are sampled at 11 KHz. They are corrupted by Gaussian white noise of different amplitude. Short time fourier transform with half overlapping windows used in the experiments. These windows are the square root of Hanning windows of size 50ms for “Piano” and “Mozart” and 20ms for “Timit”. The value a_0 in (13) has chosen for block thresholding is $a_0 = 0.05$ for partial noise removal and $a_0 = 0$ for maximum noise removal.

The objective measures are respectively the SNR and the segmental SNR [37] defined as

$$SNR = 10\ log_{10} \frac{\sum_{n=0}^{N-1} f^2[n]}{\sum_{n=0}^{N-1} (f[n] - \tilde{f}[n])^2} \qquad (20)$$

$$SegSNR = \frac{1}{H} \sum_{l=0}^{H-1} T\left(10 log_{10} \frac{\sum_{n=0}^{s-1} f^2[(\frac{n+ls}{2})]}{\sum_{n=0}^{s-1} (f[\frac{n+ls}{2}] - \tilde{f}[\frac{n+ls}{2}])^2}\right) \qquad (21)$$

where H represents the no. of frames in the signal,s is the no. of samples per frame.Performance of Minimum Mean Square Error Log Spectral Amplitude Estimation algorithm by using Decision Direct method (MMSE-LSA-DD) and Block Thresholding (BT) algorithm in terms of SNR is presented in Table II.From these comparisonsit may be noted that the residual noise masks the musical noise.The performance of Block Thresholding of Mozart, speech,&timit Signal for different SNR values is also shown in Table II.

TABLE II: COMPARISION OF SNR OF PS, LSA AND BT WITH VARIOUS AUDIO SIGNALS

Signal & SNR	PS	LSA	BT
Mozart 5dB	7.614	7.623	14.90
Mozart 10dB	12.37	12.62	16.10
Mozart 15dB	17.43	17.727	16.35
Speech 10dB	13.24	13.60	16.13
Speech 15dB	16.63	16.92	16.32
Timit 5dB	2.11	0.35	13.90
Timit 10dB	7.746	6.643	15.73
Timit 15dB	11.13	11.10	16.21

 ISSN: 2321-8169
861 - 870

From Table II we can observe that the segmental SNR of LSA based algorithm achieved better SNR than the power subtraction method. The block thresholding SNR is better than both the segmental SNR of power subtraction and LSA.

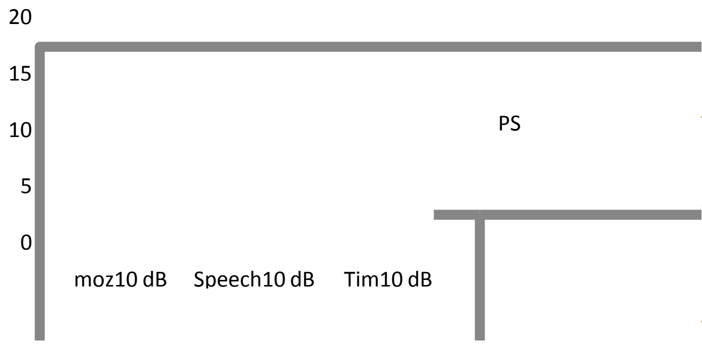

Figure 6: SNR of PS, LSA and BT.

B. Denoising Results of Soft and Hard Thresholding with DWT

The performance of various wavelets is used in denoising of various audio signals such as Mozart, speech and timit which are corrupted by additive White Gaussian noise. Thus in this section various audio signals such as Mozart, speech and timit which are corrupted by additive White Gaussian noise. Thus in this section we compare performance of four wavelets Harr, Db10, Coif5 and sym5 with audio signals. For the performance comparision and measurement of quality of denoising, the peak signal to noise ratio (PSNR) is calculated between original speech signal and denoised speech signal by [33],

$$PSNR = 10log_{10}\left(\frac{f_{max}^{2}}{MSE}\right) \quad (22)$$

Where, f_{max} is maximum value of signal and MSE is mean square error given by,

$$MSE = \frac{1}{N}\sum_{K=1}^{N}[f_d(k) - f(k)]^2 \quad (23)$$

Table III compares the performance Thresholding (BT) algorithm in terms of SNR with STFT & DWT transform. From Table III we also compare soft & hard thresholdingSNR value using wavelet transform.The performance of Block Thresholding of Mozart Signal for different SNR values is shown in the Table III.

TABLE III: COMPARISION OF BT WITH STFT AND WAVELET (SOFT AND HARD THRESHOLDING)

Signal&SNR	**BT with STFT**	**BT with soft thresholding wavelet**	**BT with hard thresholding wavelet**
Mozart 5dB	8.4143	11.57	9.84
Mozart 10dB	10.90	15.02	12.76
Mozart 15dB	13.64	18.75	15.97
Mozart 20dB	16.63	22.64	19.48
Mozart 25dB	20.16	27.43	23.62

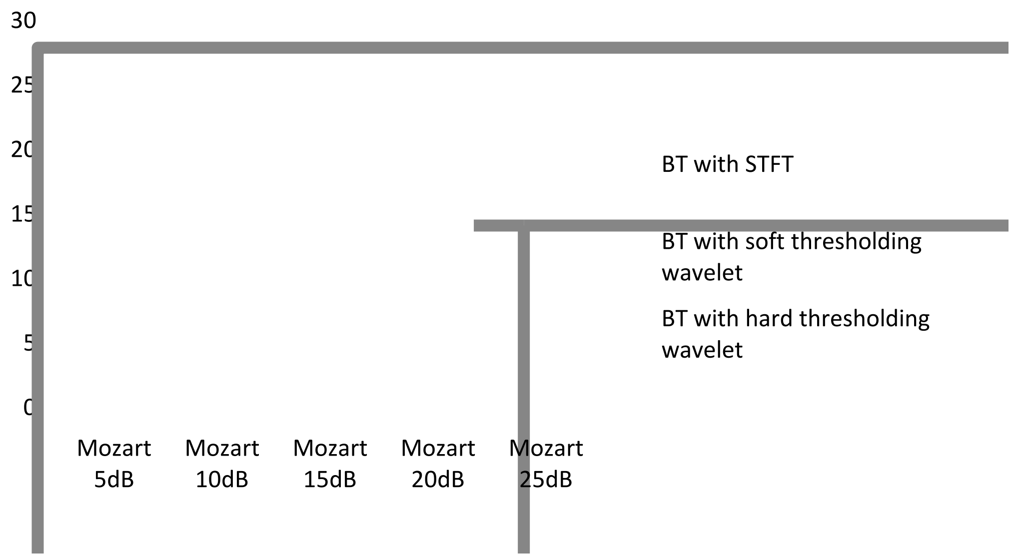

Figure 7: Comparison of BT with STFT and soft and hard thresholding wavelet.

Table IV shows the comparison of wavelets Db10, Coif5, Haar, and Sym5 with various soft thresholded audio signals.

ISSN: 2321-8169
861 - 870

TABLE IV: COMPARISION OF DIFFERENT WAVELETS ON DIFFERENT AUDIO SIGNALS

Signal & SNR	Coif5	Db10	Sym5	Haar
Mozart 5dB	11.57dB	11.09dB	10.69dB	7.39dB
Mozart 10dB	15.02dB	14.06dB	13.78dB	10.50dB
Mozart 15dB	18.75dB	17.62dB	17.48dB	14.03dB
Mozart 20dB	22.64dB	21.52dB	21.24dB	16.64dB
Mozart 25dB	27.43dB	25.53dB	25.53dB	16.74dB
Speech10DB	15.61dB	14.71dB	14.47dB	10.96dB
Timit5db	9.61dB	8.77dB	9.45dB	6.69dB
Timit10db	13.70dB	12.54dB	13.57dB	10.73dB
Timit15db	16.69dB	14.98dB	16.28dB	13.13dB

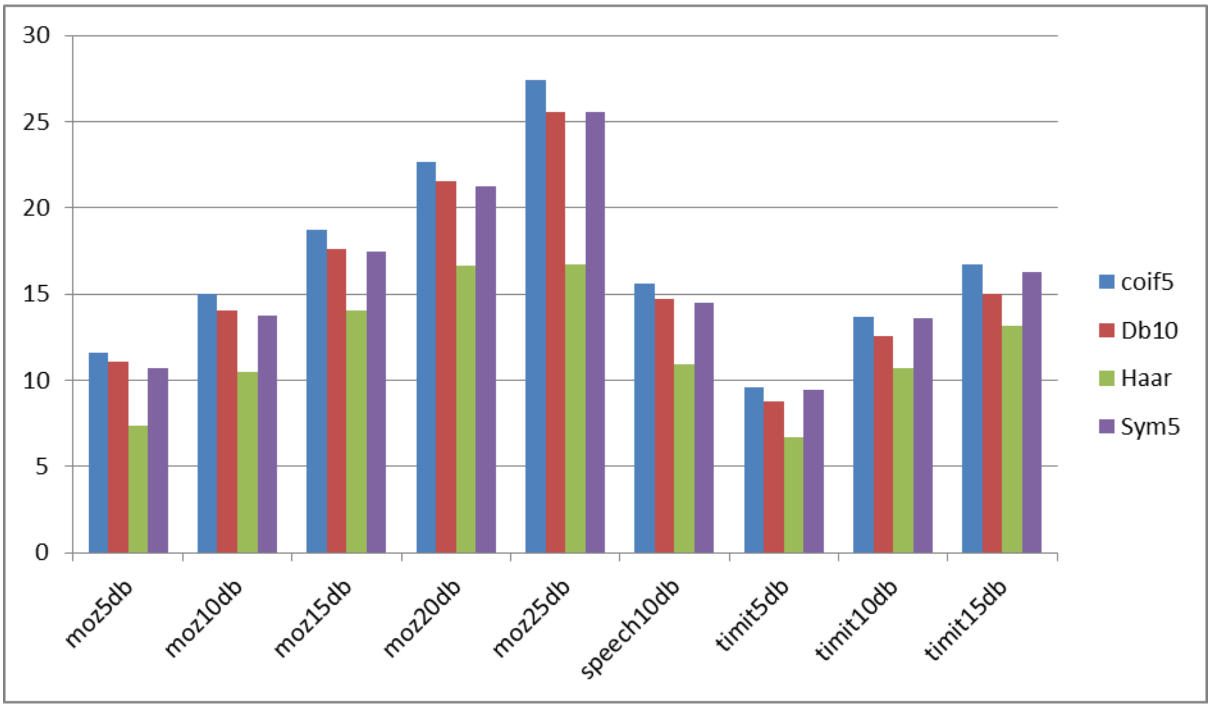

Figure 8: Comparison of various wavelets on different signals.

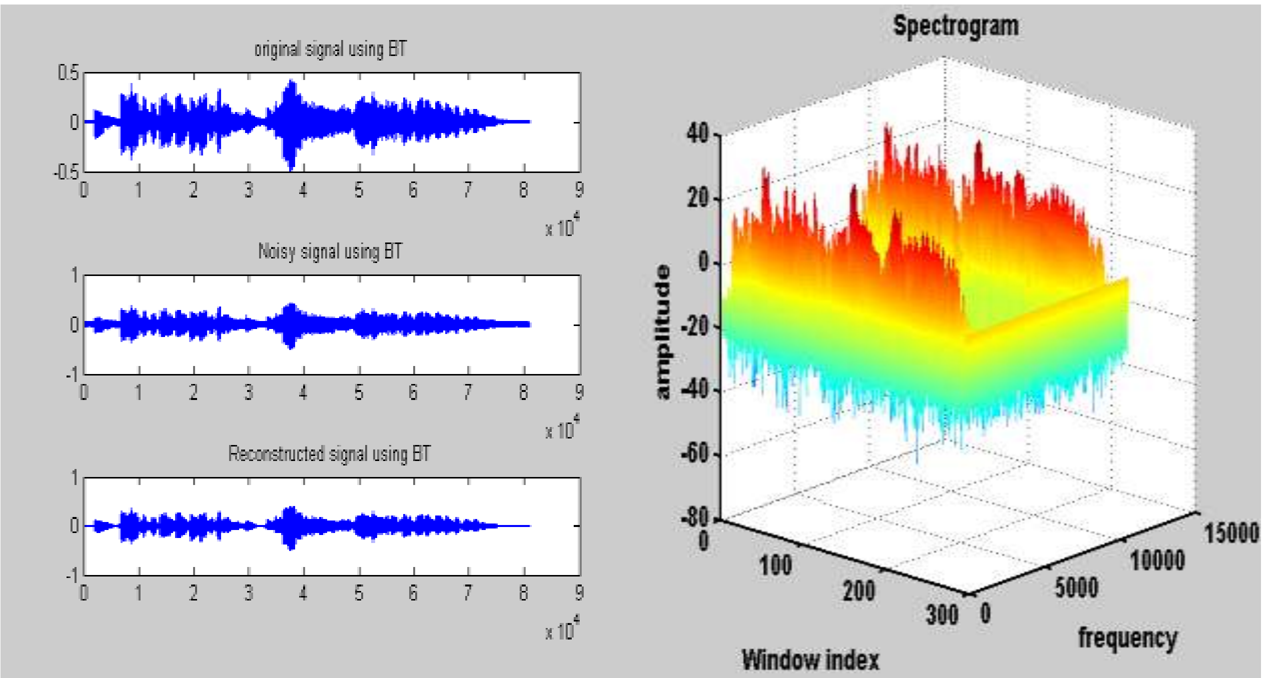

Figure 9: Graph of reconstructed signal using BT.

ISSN: 2321-8169
861 - 870

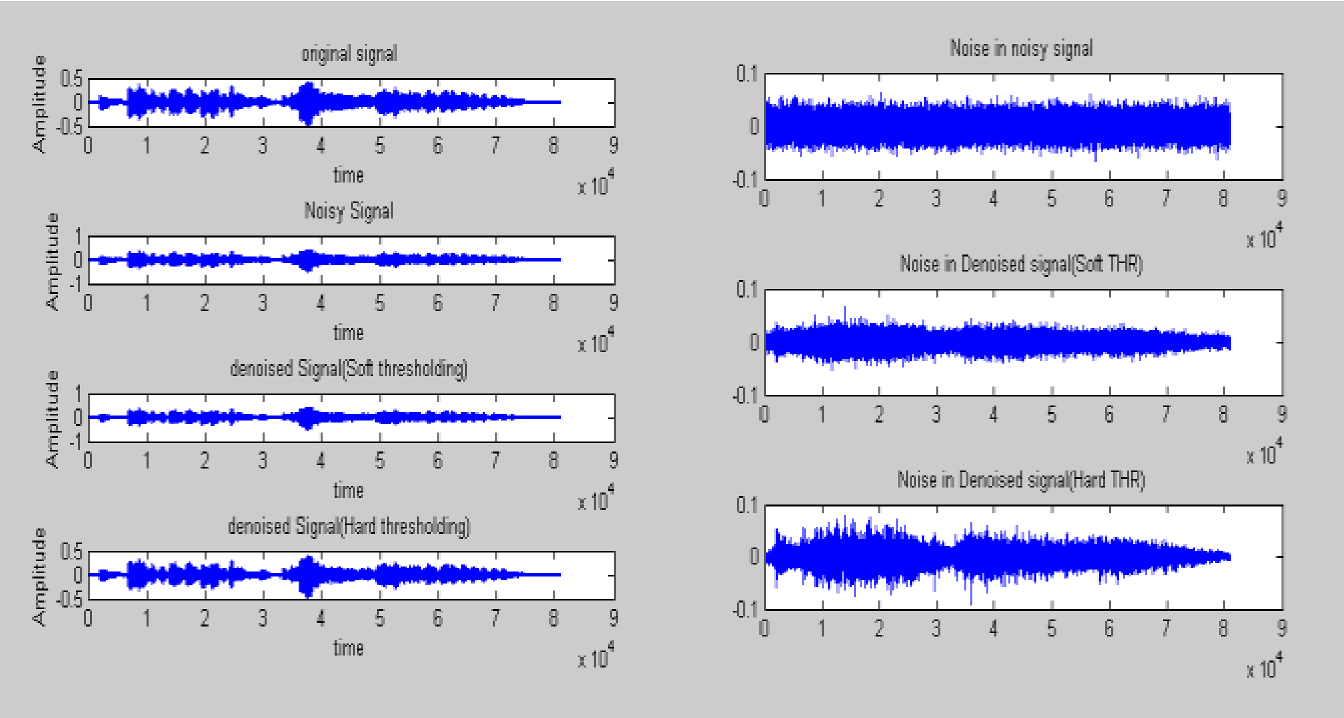

Figure 10: Graph of reconstructed signal using wavelet (soft & hard thresholding).

VI. CONCLUSIONS

In this paper we presented several techniques for audio denoising. From the study it is observed that nondiagonal time-frequency estimators found to be more effective to remove noise from audio signals. Also, adaptive audio block thresholding algorithm that adapts all parameters of audio signal and this adaptation is performed by minimizing stein unbiased risk estimator (SURE). Furthermore we presented an adaptive soft thresholding algorithm that adapts all parameters the time-frequency regularity of the audio signal. As seen from the presented results, performance of Coif5 and Db10 are better than other wavelets such as Harr and Sym5. This is mainly because Harr and Sym5 produce unwanted distortion in reconstructed noise signals. Overall from all the results presented Coif5 gives less distortion and maximum peak signal to noise ratio.

REFERENCES

[1] R. J. McAulay and M. L. Malpass,"Speech enhancement using soft decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28,no.2, pp.134-145,Apr. 1980.

[2] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoustics, Speech,Signal Processing (ICASSP)*, vol. 4, pp. 208–211,1979.

[3] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction,"*IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-27,no. 2, pp. 113–120, Apr. 1979.

[4] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *proc. IEEE*, vol. 67, Dec. 1979.

[5] D. Donoho and I. Johnstone, "Idea spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

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

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans.Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445,Apr. 1985.

[8] G. Yu, E. Bacry, and S. Mallat, "Audio signal denoising with complex wavelets and adaptive block attenuation," in *Proc. IEEE Int. Conf.Acoustics, Speech, Signal Processing (ICASSP)*, vol. 3, pp. 869–872, Apr. 2007.

[9] T. Cai, "Adaptive wavelet estimation: A block thresholding and oracle inequality approach," *Ann. Statist.*, vol. 27, pp. 898–924, 1999.

[10] T. Cai and H. Zhou, "A Data-driven block thresholding approach to wavelet estimation," Statistics Dept., Univ. of Pennsylvania, Tech.Rep., 2005.

[11] C. Stein, "Estimation of the mean of a multivariate normal distribution,"*Ann. Statist.*, vol. 9, pp. 1135–1151, 1980.

[12] Guoshen Yu, StephaneMallat and Emmanuel Barcy, "Audio denoising by time frequency block thresholding," *IEEE Trans. on signal processing*, vol.56, no.5, pp. 1830-1839, May 2008.

[13] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.

[14] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.

[15] I. Cohen, "Enhancement of speech using bark-scaled wavelet packet decomposition," in *Eurospeech*, Scandinavia, 2001.

[16] P. J. Wolfe and S. J. Godsill, "Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement," in *Proc. IEEE Workshop Statistical Signal Processing*, pp. 496–499.Aug. 2001.

[17] I. Cohen, "Relaxed statistical model for speech enhancement and *a priori* SNR estimation," *IEEE Trans. Speech Audio Process.*, vol. 13,no. 5, pp. 870–881, Sep. 2005.

[18] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Process.Lett.*, vol. 9, no. 4, pp. 113–116, Apr. 2002.

[19] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418,Nov. 2001.

[20] N. S. Kim and J. H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Process. Lett.*, vol. 7, no. 5, pp. 108–110,May 2000.

[21] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in mon-stationary noise environments," *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Phoenix, AZ, Mar. 1999.

[22] T. Cai and B. W. Silverman, "Incorporation information on neighboring coefficients into wavelet estimation," *Sankhya*, vol. 63, pp.127–148, 2001.

[23] P. Hall, G. Kerkyacharian, and D. Picard, "A note on the wavelet oracle,"*Stat. Probab. Lett.* vol. 43, pp. 415–420, 1999.

[24] P. Hall, G. Kerkyacharian, and D. Picard, "Block threshold rules for curve estimation using kernel and wavelet methods," *Ann. Statist.*, vol.26, pp. 922–942, 1998.

[25] P. Hall, G. Kerkyacharian, and D. Picard, "On the minimax optimality of block thresholded wavelet estimators," *StatisticaSinica*, vol. 9, pp.33–50, 1999.

[26] Y. Shao and C. H. Chang, "A generalized time-frequency subtraction method for robust speech enhancement based on wavelet filter bank modeling of human auditory system," *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 37, no. 4, pp. 877–889, Aug. 2007.

[27] C. Sidney Burrus, Ramesh A. Gopinath, HaitaoGuo, *Introduction to Wavelets and Wavelet Transforms*, 1st Edition, Prentice Hall, 1998.

[28] Ir. M. Steinbuch, and Ir. M.J.G. van de Molengraft, Eindhoven University of Technology, Control Systems Technology Group Eindhoven, *Wavelet Theory and Applications*, aliterature study, R.J.E. Merry, June 7 (2005).

[29] Adhemar Bultheel, *Wavelets with Applications in Signal and Image Processing*, Sept. 22 (2003).

[30] Alexandru Isar, DorinaIsar , May (2003): "Adaptive denoising of low SNR signals", Third International Conference on WAA 2003, Chongqing, P. R. China, 29-31, pp. 821-826, 2003.

[31] D.L. Donoho, Stanford University: "De-noising by Soft Thresholding", *IEEETrans. Information theory*, vol. 41, no. 3. May 1992.

[32] MatcoSaric, LukiBilicic,HrvojeDujmic, "White Noise Reduction of Audio Signalusing Wavelets Transform with Modified Universal Threshold", 2005.

[33] A.K. Verma, Neema Verma, "A comparative performance analysis of wavelets in dnoising speech signals", National conference on Advancement of technologies- Information systems & computer networks (ISCON-2012).

[34] Daubechies Ingrid, "Ten lectures on wavelets", 9e, SIAM, ISBN: 780-89871-274-2, 2006.

[35] Rajeev aggarwal, Jay Karan singh,Vijay Kr. Gupta and Dr. Anubhuti Khare, " Elimination of white noise from speech signal using wavelet transform by soft and hard thresholding," *VSRD-IJEECE*, vol.1(2), pp. 62-71,2011

[36] J. S. Garofolo, Getting Started with the DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database. National Institute of Standards and Technology (NIST), Gaithersburgh, MD, 1988.

[37] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*. New York: Prentice-Hall, 1988.