Advanced Research in Electrical and Electronic Engineering
p-ISSN: 2349-5804; e-ISSN: 2349-5812 Volume 2, Issue 12 July-September (2015) pp. 5-10

http://www.krishisanskriti.org/Publication.html

Enhancement of Speech Signal by Transform Method and Various threshold Techniques: A Literature Review

Rajkumar Angamba Singh[1] and K. Pritamdas[2]

[1]PG student, NIT Manipur
[2]NIT Manipur
E-mail: [1]rsingh1412@gmail. com, [2]kpritamdas@nitmanipur. ac. in

Abstract—*This paper presents a detailed analysis on the speech enhancement algorithms in the wavelet transform domain. In transform domain approach noise attenuation is performed on transform coefficients. The weakness and strength of various transform such as Discrete Fourier Transform, Discrete Cosine Transform, Karhunen Loeve Transform and Wavelet transform are discussed. This paper mainly focus on continuous Wavelet Transform and thresholding of these coefficients for speech enhancements. The Wavelet transform scales and thresholds both are adaptive depending upon the level of the noise of the noisy speech signal. The different thresholding methods such as hard thresholding and soft thresholding are discussed and adaptive thresholding is performed on CWT coefficients. The results are measured using Signal-to-Noise(SNR) and Segmental Signal-to-noise ratio (SSNR) for additive White Gaussian noise at various inputs SNR level. The overall results indicate that the SNR and SSNR for White Gaussian noise of the ENAT-BWT method are far better than Bionic Wavelet transform and Wavelet Packet transform.*

1. INTRODUCTION

The term "Enhancement of a Speech signal" is to improve the quality of the speech signal by degrading the amount of noise present in the speech signal. The speech signal is often associated with background noise like babble noise, train noise, F16 cockpit noise, restaurant noise etc. The different methods of Speech enhancement by Fourier Transform [1], Karhunen Loeve Transform [2]-[3], Discrete Cosine transform [4] and wavelet transform [5] are discussed on the subsequent topic. The merits and demerits of various transform methods are analysed. The wavelet transform techniques reduce computational complexity and achieve better noise reduction performance. Wavelet denoising techniques [6] perform noise reduction using thresholding. This process can be summarized into three steps i. e. computing the coefficients of the wavelet transform (WT) which is a linear operation, the 2nd step is thresholding of these coefficients which is a non-linear operation and last step is to take inverse of the Wavelet transform which leads to denoised signal. In the ENAT-BWT method, the noise standard deviation "σ" of the incoming noisy signal is to be estimated first. The "σ" is computed as the median absolute deviation/0. 675 of the wavelet coefficients belonging to the diagonal sub-band coefficients. For negative SNR levels the WT of noisy signal at 22 scales, from 7 to 28 is computed and for positive SNR levels the WT of noisy signal at 22 scales, from 6 to 33 is computed. The threshold for various noise levels are calculated according to the type of the thresholding algorithm used. Finally inverse wavelet transform of the coefficients is computed and thus providing the enhanced speech signal. The results are compared with Bionic Wavelet transform and Wavelet packet transform.

This paper is organized as follows. Section 2, 3and 4 gives an overview of speech enhancement by Fourier Transform (FT), Karhunen Loeve transform (KLT) and Discrete cosine transform (DCT) respectively. Section 5 gives the merits and demerits of FT, KLT and DCT. Section 6 gives the overview of Wavelet Packet transform (PWT). Section 7 gives the overview of Bionic Wavelet transform. Section 8 introduces the ENAT-BWT method. Section 9 includes the evaluation and results of these experiments, followed by overall conclusion in Section 10.

2. SPECTRAL SUBTRACTION METHOD

The spectral subtraction method is historically one of the first algorithms proposed for noise reduction. It is very simple method and easy to implement, it is based on the principle that we can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present i. e. during 'speech pauses'. The basic assumption is noise is additive. Its spectrum doesn't change with time means noise is stationary or it's slowly time varying signal, whose spectrum doesn't change significantly between the updating periods.

Let y(n) be he noise corrupted input speech signal which is composed of the clean speech signal x(n) and the additive noise signal d(n). Mathematically, it can write in time domain and Fourier domain as given in the eqn 1 and 2 respectively.

$$y(n)=x(n)+d(n) \quad (1)$$

$$Y[\omega]=X[\omega]+D[\omega] \quad (2)$$

$Y[\omega]$ can be expressed in terms of magnitude and phase as $Y[\omega]=|Y[\omega]|e^{j\phi y}$

where $|Y[\omega]|$ is the magnitude spectrum and ϕ is the phase spectra of the corrupted noisy speech signal Noise spectrum in terms of magnitude and phase spectra as

$$D[\omega]=|D[\omega]|e^{j\phi y}$$

The magnitude of the noise spectrum $|D[\omega]|$ is unknown but can be replaced by its average value computed during non-speech activity i. e. during speech pauses. In speech enhancement phase spectra is kept constant. The clean speech signal can be estimate by simply subtracting noise spectrum from noisy speech spectrum. Mathematically, it is given as

$X[\omega]=[Y[\omega]-D[\omega]]e^{j\phi y}$. In magnitude it is given as

$X[\omega]=|Y[\omega]|-|D[\omega]|$. Similarly for power spectrum,

$X[\omega]^2=[|Y[\omega]|^2-|D[\omega]|^2]$

The enhanced speech signal is finally obtained by computing the inverse Fourier transform of the estimated clean speech $|X[\omega]|$ for magnitude.

3. KLT TRANSFORM BASED SPEECH ENHANCEMENT

The KLT transform based speech enhancement is also very useful in speech processing. The KLT is known by many names such as Hotelling transform, method of Principal components or Eigen vector transform.

Karhunen and Loeve introduce the transform for continuous series while Hotelling transform is the discrete version of the KLT. Since the transform is performed using the Eigen vectors of the autocorrelation matrix of the input data, it is also known as Eigen vector transform. For real N×1 random vector, u (n), n=0……. N-1}, the basis vector (ϕ_k)of the KL transform(ϕ) are given by the orthonormalized eigenvectors of its auto correlation matrix, R, that is

$$R\phi_k=\lambda_k\phi_k,\ 0\leq k\leq N-1$$

The KL transform of u is defined as $v=\phi^{*T}u$ and the inverse transform is $u=\phi v$. The decomposition of the vector space of the noisy signal into a signal subspace and noise subspace is performed by applying KLT to the noisy signal [2]-[3]. The linear estimation is performed by modifying the KLT components which represent the signal subspace by a gain function determined by the estimation criterion. The remaining KLT components are nulled. The enhanced signal is obtained from inverse KLT of the altered components. An example of the application of the KLT for speech enhancement is described in [2]-[3].

4. DISCRETE COSINE TRANSFORM METHOD

A clean speech is first divided into frames with 50% overlapping, and the transform is performed. The transformed coefficients are then sorted according to their magnitude and n coefficients with the lowest energy set to zero. The speech is reconstructed using the weighted overlap add technique [7]. If DCT is used a higher upper bound is possible since DFT only attempt to correct the noisy amplitude but not the phase component that actually results in an upper bound on the maximum improvement in SNR level. If the phase is replaced by random noise uniformly distributed between $-\Pi$ to $+\Pi$, a rough and completely unvoiced speech is obtained. On the other hand, if the phase is replaced by zero, the reconstructed speech sounds completely voiced and monotonous. Therefore it is not correct to view the phase as totally unimportant and especially for high levels of noise; the reconstructed speech quality will be affected. For DCT coefficients are real and can be considered to have a binary phase value. An example of the DCT for speech enhancement is described in [4]

5. MERITS AND DEMERITS

DCT provides higher energy compaction as compared to DFT. DFT doesn't consider phase component however DCT considers phase component. For a window size of N DCT has N independent spectral component while the DFT only produces N/2+1 independent spectral components, as the other component are complex conjugate. DCT results in a higher upper bound for speech enhancement using the DFT described in [4]. KLT is also optimal in energy compaction and it fully decorrelates all the coefficient. However it is not commonly used because unlike other transform those are functionally independent of the data, KLT depends on the second order statistics of the data. Since it is data dependent, it is not possible to come up with a fast transform without making assumptions of the data described in [2] and [3].

The wavelet transform has the advantage of its fast implementation as its computational complexity is of the order N. Also the no. of taps in the digital filters is normally small. More importantly, it has variable tradeoff between time resolution and frequency resolution. The basic WT provides better frequency resolution for low frequency component. On the other hand, for higher frequency components, the time resolution is better. The main disadvantage of the WT is the limited no of frequency bands. Fourier and Discrete cosine transform have only one basis function, the wavelet transform has many possible choices of basis function. MATLAB libraries recognize more than a dozen wavelets and are currently known in the literature.

6. WAVELET PACKET TRANSFORMS (PWT)

PWT is a form of wavelet packet decomposition that is tailored to match the Bark scale used in speech processing applications. PWT denoising is a combination of bark-scaled wavelet packet decomposition (BS-WPD), a soft decision gain modification and a magnitude decision-directed estimation technique. It attains speech enhancement by introducing some redundancy in the wavelet packet decomposition. The PWT is an over complete representation meant specifically for audio signals that achieve higher frequency resolution than the critical band decomposition and a higher time resolution than the the conventional wavelet packet decomposition (WPD). Expanding all the high frequency sub bands as in critical band wavelet packet decomposition (CB-WPD) results in an increase in computational complexity and at the same time reduces the perceptual quality of unvoiced sounds. Initially, the BS-WPD and CB-WPD splits the audio frequency range 0-8Khz into 21 subbands. Redundancy for speech enhancement is provided by further decomposing the CB-WPD into four no decimated sub bands by a two level over complete expansion. Thus 84 sub bands are obtained as outputs of low pass and high pass wavelet filters without downsampling. This over complete auditory representation results in PWT denoising when combined with modified Wiener filtering and the "magnitude" decision directed estimation described in [9]-[10].

7. BIONIC WAVELET TRANSFORM

BWT is an adaptive wavelet transform that models the auditory system. It has high sensitivity and selectivity with compact energy representations of signals. The resolution of BWT is adaptive in the time frequency plane in the sense that it can be adjusted by the signal frequency as well as signal instantaneous amplitude and first order differential of the signal [8]. The resolution in case of traditional wavelet transform (WT) is adjustable along both the time axis and the frequency axis. BWT combines the active biological mechanism of the auditory system with the wavelet transform and has proven very effective for speech enhancement predominantly for cochlear implants [8].

8. ENAT-BWT

This method proposes Estimated Noise and adaptive threshold Bionic Wavelet transform (ENAT-BWT) speech enhancement technique. This technique is based on Bionic Wavelet transform (BWT). A block diagram of overall approach is shown in fig 1.

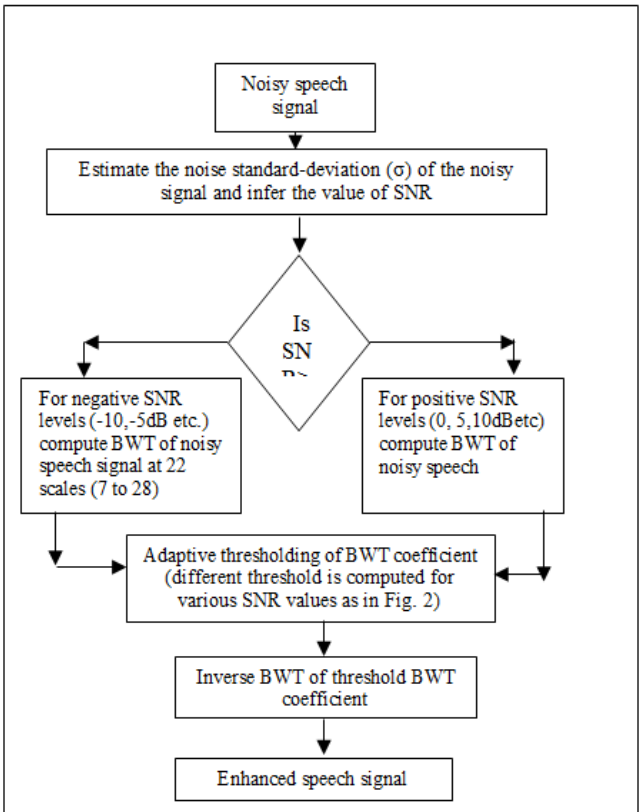


Fig. 1: Block diagram of the ENAT-BWT algorithm

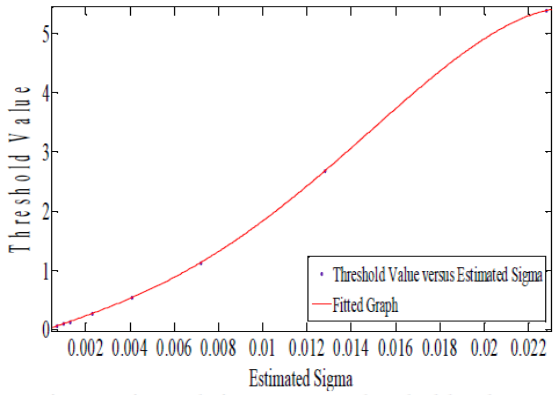


Fig.2 Estimated sigma versus threshold value graph.

The noise standard deviation "σ" of the incoming noisy signal is to be calculated first. For this Discrete wavelet transform (DWT) of the noisy speech signal is computed using Daubechies wavelet of order 5. Then, standard-deviation (σ) is computed as the median absolute deviation/0. 675 of the wavelet coefficients belonging to the diagonal sub-band. For negative SNR levels such as -10, -5 dB etc, the bionic wavelet transform (BWT) of the noisy speech signal at 22 scales, from 7 to 28 is taken. At SNR levels of 0, 5 and 10 dB i. e. for positive SNR level, the Bionic wavelet transform (BWT) of noisy speech signal at 28 scales, from 6 to 33is taken. Initially

the thresholds for noise are determined manually that provide the best signal to noise ratios (SNR). Then, using curve fitting approach a generalized model is obtained that provides the best threshold parameter for input noisy signal of any noise standard deviation. The graph obtained after curve fitting is given in Fig. 2. Thus the algorithm selects the threshold value from the graph and soft thresholding is applied to the BWT coefficients. Finally inverse bionic wavelet transform (IBWT) of thresholded BWT coefficients is computed. This provides the enhanced speech signal.

9. EVALUATION AND RESULTS

A). Criterion of evaluation

For evaluation of the ENAT-BWT technique, the results are compared to the BWT and PWT techniques. The signal to noise ratio (SNR) and Segmental Signal to noise ratio (SSNR) are the performance comparison parameters in this paper. Signal to noise ratio is given a

$$SNR(db) = 10\log_{10}\left[\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1}(x(n)-\hat{x}(n)^2}\right]$$

Where x(n) and x(n) are the original and enhanced speech signals respectively and N is the number of samples in the speech signal.

Segmental Signal to Noise Ratio is given as

$$SSNR(db) = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\left[\frac{\sum_{n=Nm}^{Nm+N-1} x(n)^2}{\sum_{n=Nm}^{Nm+N-1}(x(n)-\hat{x}(n)^2}\right]$$

Where M is the number of frames, N is the size of frame and Nm is the beginning of the m-th frame.

B). Experimental Results

This section presents the experimental results of the ENAT-BWT algorithm at SNR levels of -10, -5, 0, 5 and 10dB, and compares its performance with the Wavelet packet transform (WPT) and the Bionic wavelet transform (BWT) algorithm. Five speech signals taken from the TIMIT Acoustic-Phonetic continuous speech corpus [11], were used to evaluate the proposed algorithm. Results are averaged across 5 utterances used as examples, giving a single evaluation metric for each method. Implementation was done using the MATLAB wavelet toolbox (The Math works Inc., 2011). SNR and SSNR results for white noise conditions are analyzed.

The ENAT-BWT method shows the best performance for additive white Gaussian noise conditions. The algorithm shows the best SNR movements at -10, -5, and also at +5dB noise case as can be seen from Fig. 3 (Table I). For SSNR calculation, number of frames taken is 250 and the starting frame's sample no is 5000. This method shows the best SSNR improvements at -8. 8, -7. 1, -4. 8, -2 and 12. 4dB input SSNR levels. The SSNR results obtained for white Gaussian noise conditions are presented in fig 4 (Table II).

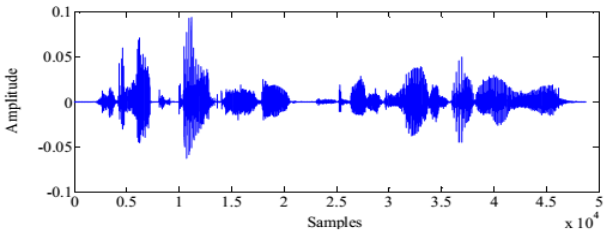
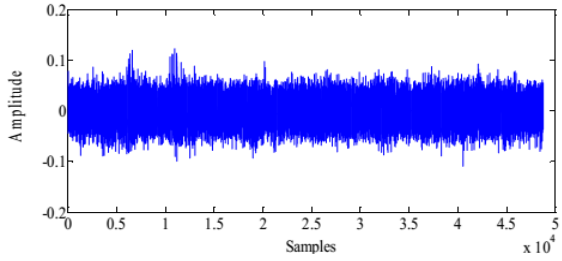
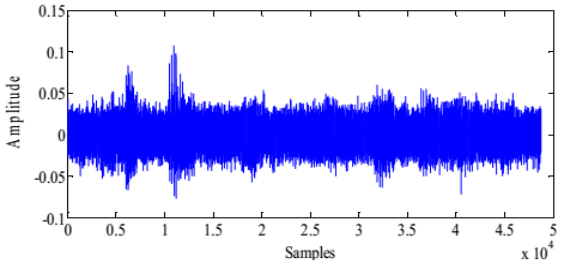
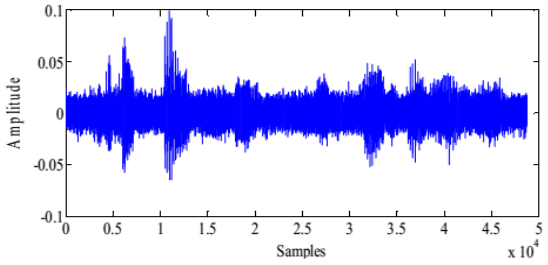
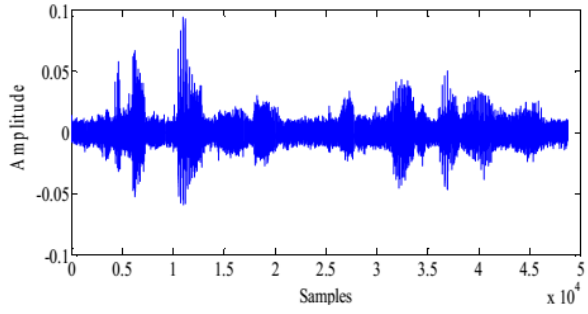


Fig.5 Original Speech Signal









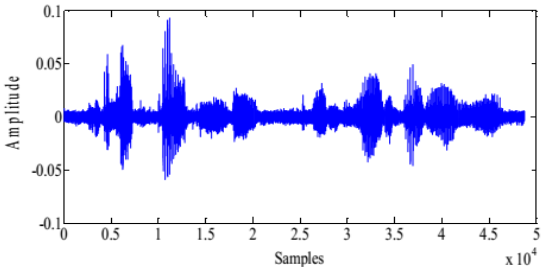


Fig.6 Noisy signal at -10,-5, 0, 5 and10dB input SNR level respectively.

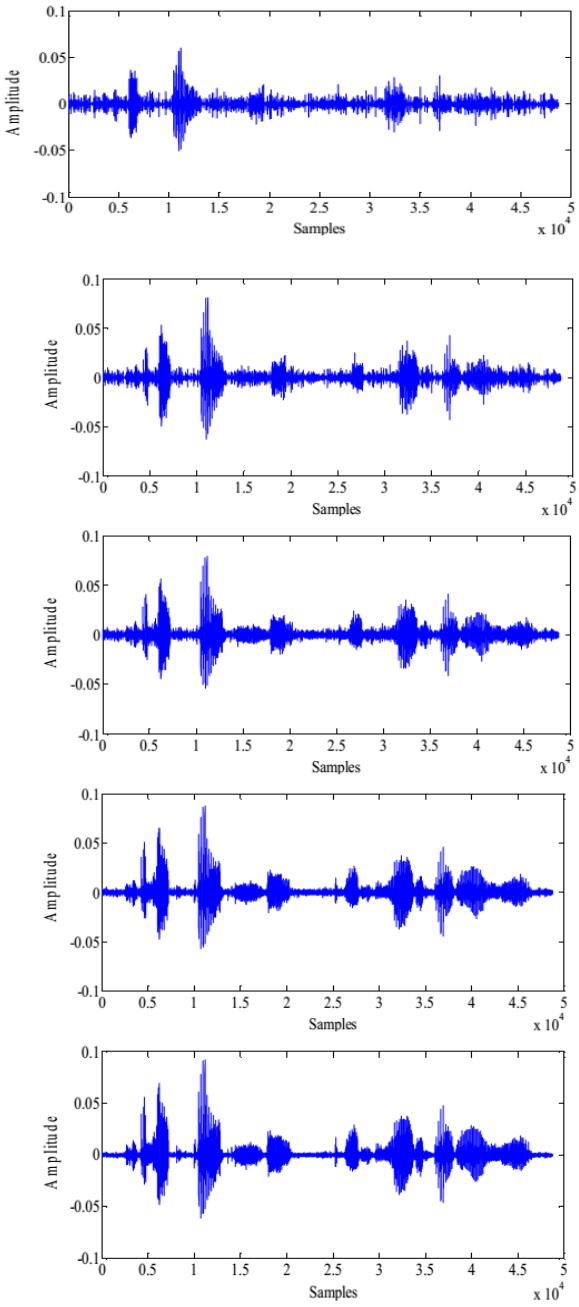


Fig.7 Enhanced signal at -10,-5, 0, 5 and10dB input SNR level respectively.

The qualitative performance of the ENAT-BWT can be seen from Fig. 5, Fig 6, and Fig. 7. Fig. 5 shows the original speech signal on which the experiments were conducted. The noisy signal and enhanced signal at -10, -5, 0, 5 and 10dB input SNR levels are shown in Fig. 6 and Fig 7 respectively.

Table I: Speech quality evaluation in terms of signal to noise ratio (SNR) for speech corrupted by white Gaussian noise at various inputs SNRs

Input SNR (dB)	Output SNR (dB)		
	BWT [9]	PWT [12]	ENAT-BWT
-10	2	1. 4	3. 19
-5	4. 9	3. 5	5. 86
0	7. 9	6	8. 16
5	11	9	11. 89
10	13. 8	13	15. 41

TABLE 2: Speech quality evaluation in terms of segmental signal to noise ratio (SSNR) for speech corrupted by white Gaussian noise at various inputs SSNRs

Input SSNR	Output SSNR (dB)		
(dB)	BWT[9]	PWT[12]	ENAT-BWT
-8. 8	-1. 3	-3. 2	0. 18
-7. 1	-1	-2. 8	1. 32
-4. 8	0. 5	-1. 4	2. 82
-2	2. 4	0. 4	4. 79
1. 24	4. 4	3	7. 12

10. CONCLUSION

The ENAT-BWT used a modified algorithm for speech signal enhancement from the Bionic Wavelet transform has been discussed. In this method, the no. of scales for computation of BWT is different for different SNR inputs. A generalized model is obtained that provides the best threshold parameter for input noisy signal of any noise standard deviation. The optimum threshold value is thus automatically selected and thresholding is applied to the BWT coefficients. Finally inverse bionic wavelet (IBWT) of threshold BWT coefficients is computed. This provides the enhanced speech signal. Experimentals evaluations were performed on speech signals from the TIMIT database, corrupted by Gaussian noise at various SNR levels. The performance was evaluated in terms of Signal to noise ratio (SNR) and Segmental SNR. Denoising results show superior performance of the proposed method as compared to WT.

REFERENCE

[1] Steven F. Boll "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", *IEEE Transactions on Acoustic, Speech, and Signal Processing, VOL. ASSP-27, No. 2, April 1979.*

[2] X. Zhen and I. Y. Soon, "Denoising using KLT, ", in proceedings of the 4th National Undergraduate Research Opportunities Program Congress, Vol. 2, pp. 760-763, 1998.

[3] Y. Ephraim and D. Malah, "A Signal Subspace Approach for Speech enhancement, ", *IEEE Trans. Speech and Audio processing, vol. 3, pp. 251-266, 1995.*

[4] Ing Yann Soon, Soo Ngee Koh, Chai Kiat Yeo, "Noisy peech Enhancement using Discrete Cosine transform "Elsevier Speech communication 24(1998) 249-257.

[5] R. M. Rao, and A. S. Bopardikar, Wavelet transforms :Introduction to theory and applications, 6th ed., Pearson Education, 2005.

[6] D. L. Donoho, "Denoising by soft thresholding, "*IEEE Trans. Inform. Theory, Vol. 41, no. 3, pp 123-133, 2007*

[7] R. E. Crochiere, "A weighted overlap-add method of short time Fourier analysis/ synthesis, "*IEEE Trans. Acoustic. Speech Signal process. ASSP-28 (1980) 99-102.*

[8] Yao, J., Zhang, Y. T., 2001, "Bionic Wavelet transform new time frequency method based on an auditory model. "*IEEE Trans. Biomed. Eng. 48(8), 856-863*

[9] Johnson, M. T., Yuan., Ren, Y., 2007"Speech signal enhancement through adaptive wavelet thresholding, "Elsevier Speech commun. 49(2), 123-133

[10] Preety D. Swami, Rupali sharma, Alok jain, Dhirendra K. swami, "speech enhancement by noise driven adaptation of perceptual scales and thresholds of continuous wavelet transform coefficients" Elsevier Speech communication 70(2015)1-12

[11] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallet, and N. Dahlgren, et al., TIMIT Acoustic-phonetic Continuous Speech corpus: Linguistic Data Consortium, 1993.

[12] I. Cohen, "Enhancement of speech using bark-scaled wavelet packet decomposition, "paper presented at the Eurospeech, Denmark, 2001.

[13] J. Yao, and Y. T. Zhang, "The application of Bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations, '"*IEEE Trans. Biomed. Engineering, vol. 49, no. 11, pp. 1299-1309, 2002*

[14] R. polikar"The wavelet tutorial by Robi Polikar, " Available: http://users. rowan. edu/~polikar/WAVELETS/W Tutorial. html, 1996.