ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

A Modified Speech Enhancement Algorithm Based on the Continuous Wavelet Transform

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

P.G. Student, Department of Electronics and Communication Engineering, NIT Manipur, Imphal, Manipur, India[1]

Assistant Professor, Department of Electronics and Communication Engineering, NIT Manipur, Imphal, Manipur, India[2]

ABSTRACT: Speech is a common medium of human interaction. Normally Speech signals are corrupted with noises that reduce the quality and intelligibility of the speech signals. In such noisy environment listening task is very difficult at the end user. In order to make speech signals effective the noisy speech signals are need to be enhanced. Speech enhancement is used many times for pre processing of speech for computer speech recognition system. The development and widespread deployment of digital communications systems during the last twenty years have brought increased attention to the role of speech enhancement in speech processing problems. For this purpose, this paper present a single channel speech enhancement technique that uses continuous wavelet transform**s** (CWT)**.**The CWT coefficients were obtained by convolving the noisy signal with the Morlet wavelet scaled at various frequencies. The CWT coefficients were then threshold by using a known thresholding algorithm. Finally, the inverse CWT of the threshold CWT coefficients was obtained by convolving with the Morlet wavelet scaled at the same frequencies as it was done for the forward CWT computation and performing a weighted summation across scales and thereby producing a enhanced speech signal. Signal to noise ratio (SNR) and Mean square error (MSE) is computed to evaluate the performance of the proposed method. The simulation is performed in MATLAB 2013 by considering different performance measures. The results obtained are compared against the known traditional Wavelet packet transform (WPT) method.

KEYWORDS: Intelligibility; Speech Enhancement; Threshold Function; SNR; MSE

I. INTRODUCTION

The speech enhancement is required in many situations in which the signal is to be communicated or stored. The enhancement of speech is an important problem within the field of speech and signal processing, with impact on many computer based speech recognition, coding and communication applications. The quality of speech is judged by two aspects: perceptibility and intelligibility .Perceptual quality, i.e., how pleasant it sounds to a human listener, depends in terms of its 'goodness'. On the other hand, intelligibility, how much information can be extracted from a speech is a measure of accuracy that depends on the correct identification of syllables, words or sentences employed for testing. Since Speech signals are nonlinear and non stationary in nature, the performance of related studies significantly dependent on the analysis method. The underlying goal of speech enhancement is to improve the quality and intelligibility of the signal, as perceived by the human listeners.

The Bionic Wavelet transform (BWT) is based on an auditory model of the human cochlea [1].It is a time-frequency method that captures the non-linearties present in the basilar membrane and translating those into adaptive time-scale transformations of the underlying mother wavelet.BWT extends one-dimensional –adjustable-resolution along the frequency axis in traditional wavelet transform to two-dimension-adjustable-resolution.It has been proposed as a method for wavelet decomposition of speech signals (Yao and Zhang, 2001, 2002) [1] [2]. The CWT computation was employed with a Morlet Wavelet with a base fundamental frequency "ω_o" of the unscaled mother wavelet of 15165.4 Hz for the human auditory system, per Yao and Zhang's original work [1]. In [1], CWT coefficients are computed at 22

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

scales (for m=7-28). The centre frequency at each scale of the morlet wavelet is given by $\omega_m=\omega_o/(1.1623)^m$ where m=7,8,...28.The most distinguishing characteristic of BWT is that its resolution in the time-frequency domain can be adaptively adjusted not only by the signal frequency but also by the signal instantaneous amplitude and its first order differentialThe details are dercribed in [1].The wavelet based techniques in speech enhancement performs thresholding to the wavelet coefficients (D.L. Donoho, 1994) [3].A good representational capability of the Wavelet transform on speech signals could lead to a better separation of signal and noise components within the wavelet coefficients and therefore better enhancement results. The proposed work is inspired from [4] [9].

The paper is organized as follows. In Section II describes Speech enhancement technique based on the Wavelet transform .The steps of the algorithm are explained. Section III introduces the Continuous Wavelet transform method using Morlet wavelet and outlines the experimental method, including the flowchart. The results of these experiments are presented and discussed in Section IV, followed by overall conclusions in Section V.

II. RELATED WORK

The extension of the Wavelet transform is the Wavelet packet transform (WPT) described in [7].In this case the filtering process is iterated on both high and low frequency components,

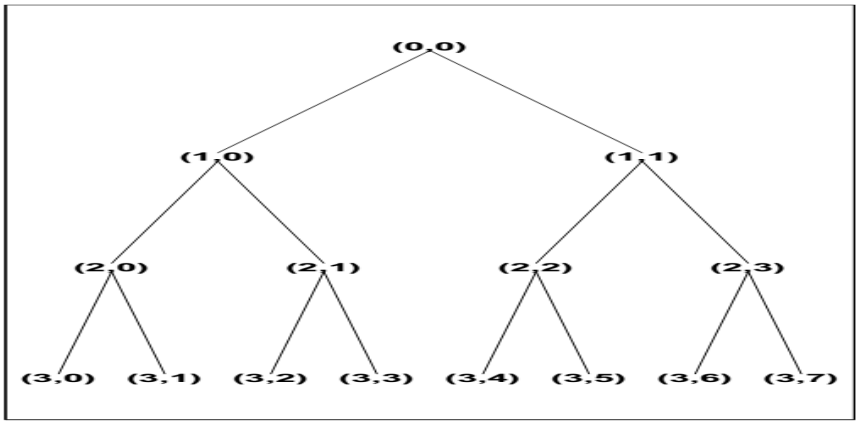

Figure 1(a): WPT Tree

The third level decomposition of the wavelet packet transform is shown in the above figure 1 (a) in which high and low frequencies are divided in each levels of decomposition.High pass filter removes the low frequency components of the signal and corresponding filter parameters becomes the detailing part of the wavelet coefficients.Low pass filter removes the high frequency components of the signal and the corresponding filter parameters become the smoothing part of the wavelets coefficients In Wavelet domain, thresholding is used to smooth out or to remove some of the generated coefficients by fixing a particular threshold value. This reduces the noise content of the signal under the non-stationary environment.

There are two common ways to threshold the resulting wavelet coefficients. They are Hard thresholding and Soft thresholding. In Hard thresholding, the coefficients values are set to zero whose absolute value is below a threshold value. In Soft thresholding, it goes one step further and reduces the magnitude of the remaining coefficients by the threshold value.

The Wavelet packet transform (WPT) is implemented using following steps.

1. Consider a clean speech signal x(n).
2. Generate an Additive White Gaussian noise w(n) which is random in nature and add it to the original signal x(n).Mathematically it can be expressed as

$$y(n)=x(n)+w(n)$$

3. Compute the wavelet packet transform (using Haar as a Wavelet) of the noisy speech signal y(n).

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

4. Choose a threshold value for thresholding.The optimal threshold value is chosen by universal threshold method.
5. Apply the Hard thresholding or the Soft thresholding or a known thresholding algorithm to the Wavelet Packet coefficients.
6. The original speech signal is reconstructed by taking the inverse wavelet packet transform.

After applying all these 6 steps, an enhanced speech signal is obtained.

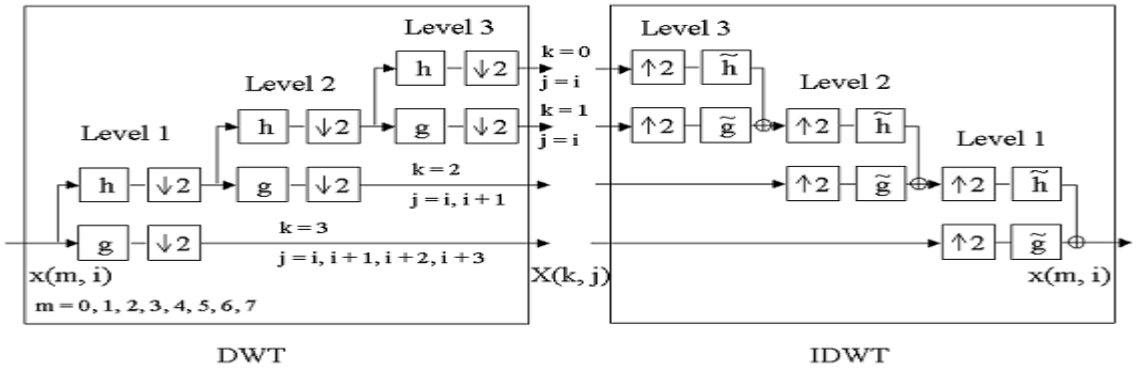

Figure 1(b): Three level Wavelet packet transform

The above figure 1(b) shows the third level decomposition and reconstruction of the wavelet packet transform in which high and low frequencies are divided into sub groups.

The proposed method performs Continuos wavelet transform using Morlet as a Wavelet instead of a Haar wavelet.The CWT (Continuos wavelet transform) is often expressed as

$$X_{\mathrm{CWT}}(a,\tau) = \langle x(t), \varphi_{a,\tau}(t) \rangle = \frac{1}{\sqrt{a}} \int x(t) \varphi^* \left(\frac{t-\tau}{a} \right) \mathrm{d}t$$

The CWT is the convolution of the shifted and dilated versions of the mother wavelet with the original signal.Idealy long windows are employed on low frequency parts of the speech signal for good frequency resolution and short windows are employed on high frequency components of the speech signal.It succeeds in adjusting time and frequency resolution without defeating the uncertainty principle.The Morlet Wavelet does not have a scaling function and therefore does not support the fast discrete wavelet transform.The inner product of the CWT is calculated normally by resorting it to numerical integration using computers. There are some fast existing algorithms for the continuos wavelet transform such as algorithm a'trous(Holschneider 1989),chirp-z transform(Jones 1991),Mellin transform(Bertrand 1990).The CWT can operate on every scale whereas the discrete wavelet transform chooses a subset of scales and positions to calculate.

The basic idea of the BWT is to make the envelope of the mother wavelet time varying according to the characteristics of the target signal by introducing a time varying function. The total no. of scale in the BWT as mentioned earlier operates on 22 scales and hence the speech signal was decomposed to 22 different channels with different centre frequencies ranging from 225 Hz to 5291 Hz.The properties of BWT are (i) It is a nonlinear transform that has the high sensitivity and frequency selectivity.(ii)BWT represents the signal with a concentrated energy distribution.(iii)The inverse BWT can reconstruct the original signal from its time frequency representation. The details are described in [1],[2],[5] and [6].The propose work is inspired from [4] [9] [1] and [2].

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

III. PROPOSED METHOD

The proposed work suggest the wavelet denoising methods by thresholding the continuous wavelet transform(CWT) coefficients and then computing the inverse wavelet transform of the thresholded coefficients. The degree to which a particular set of wavelet coefficients form a useful or compact representation of a signal is a function of how well the mother wavelets matches the underlying signal characteristics, as well as the time and scale selected. The CWT computation was employed with a Morlet Wavelet with a base fundamental frequency "ω_o" of the unscaled mother wavelet of 15165.4 Hz for the human auditory system, per Yao and Zhang's original work [1] [2].The Morlet wavelet does not have scaling functions and therefore does not support the fast discrete wavelet transform. The support length of the Morlet wavelet is chosen as [-4, 4] with different centre frequencies for each scale. The centre frequency at each scale of the logarithmic spacing is given by $\omega_m=\omega_o/(1.1623)^m$,m=0,1,2,... .These frequencies were spaced logarithmically in the desired frequency range. In this paper, the CWT coefficients are computed at 24 scales (for m=6-29).The center frequencies corresponding to these scales varies from 193.47 Hz to 6150.98 Hz which are logarithmatically spaced. The CWT coefficients were obtained by convolving the noisy signal with the Morlet wavelet scale to the above mentioned frequencies. The CWT coefficients generated at the required number of scales were then thresholded by using different thresholding algorithms. The overview of the proposed method is shown in the following figure 2.

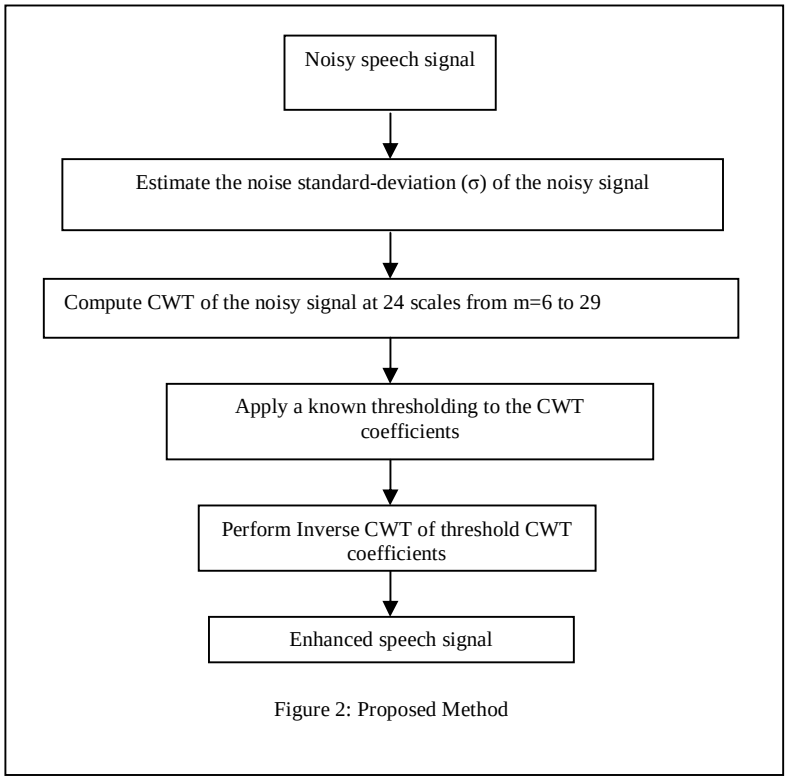

Figure 2: Proposed Method

The above figure 2 shows the flow diagram of the proposed work.A noisy speech signal is taken as an input.From the input noisy signal "x_{noisy}", the standard deviation or the Median absolute deviation can be calculated by the formula $\sigma=std(x_{noisy})$ or $\sigma=mad(x_{noisy})/0.6745$ respectively which gives an approximate estimation of the noise level of the noisy signal "x_{noisy}".In this paper, Median absolute deviation is calculated for evaluation of the threshold value. One of the first methods for selection of threshold was developed by Donoho and Johnstone [3][13] and it is called as universal threshold. It is given by Threshold, $T=\sigma \times \sqrt{2log2\ (N)}$ where N denotes the no. of samples of the noisy signal.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

Thresholding is an important step in removing noise from a noisy signal. Hard thresholding and Soft thresholding are the two common and most popular methods. In this paper a different thresholding algorithm is used described in [8].The CWT coefficients, "C" of the noisy signal is shrinked by the thresholding algorithm [8] given below.

$$\text{New Coefficients} = C - \text{sign}(C) \times T \qquad \text{IF} \quad |C|^3 > T$$

$$= 0 \qquad \text{OTHERWISE}$$

In this thresholding algorithm, the CWT coefficients, "C" to the power three which are smaller than threshold value, "T" are set to zero otherwise coefficients are shrunk in magnitude according to the above formula. The CWT coefficients computed are available in a matrix form. The number of columns in the matrix is equal to the signal length while the number of rows is 24, since the computation is done for scales from m=6 to 29.Thus in this work, the above thresholding algorithm were applied to each element of the CWT coefficient matrix. Finally the inverse CWT of the thresholded CWT matrix was computed by convolving with the Morlet wavelet (as was done for forward CWT computation) at same scales and performing a weighted summation across scales.

IV EXPERIMENTAL RESULTS

The proposed method is simulated on a male sentence taken from the NOIZEUS database by using the MATLAB 2013 software, and it is compared with traditional DWT method. The clean speech material is the short sentence "A good book informs of what we ought to know" which is spoken by a male speaker. The sampling frequency is 8 kHz. The noisy speech signal is generated by adding White Gaussian noise and also four kinds of real world noises taken from the NOIZEUS data base.

The Signal to Noise ratio (SNR) and Mean Square Error (MSE) are the two main parameter for evaluation of the output.SNR is calculated as the ratio of signal power to the noise power in decibels.MSE is the mean square difference between the original clean signal and the estimated signal. The denoising is successful if output MSE is lower than the input MSE. The lower the MSE, the better the enhancement quality. The SNR and the MSE are expressed below mathematically.

$$\text{SNR}=10\log_{10}\left[\frac{\sum_{n=0}^{N-1}x(n)^2}{\sum_{n=0}^{N-1}(x(n)-\hat{x}(n)^2}\right]$$ where x(n) is the original signal, $\hat{x}(n)$ is the noisy signal and N=no. of samples

$$\text{MSE}=\frac{1}{N}\sum(xi-yi)2$$ where x_i is the original signal and y_i is the enhanced signal.

The results have been demonstrated in the form of comparison tables and bar charts. The results of the proposed method are compared with the traditional WPT method on the basis of SNR and MSE. The additive White Gaussian noise is added to the original speech signal to introduce distortions, with SNR 0dB thereby creating a noisy signal.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

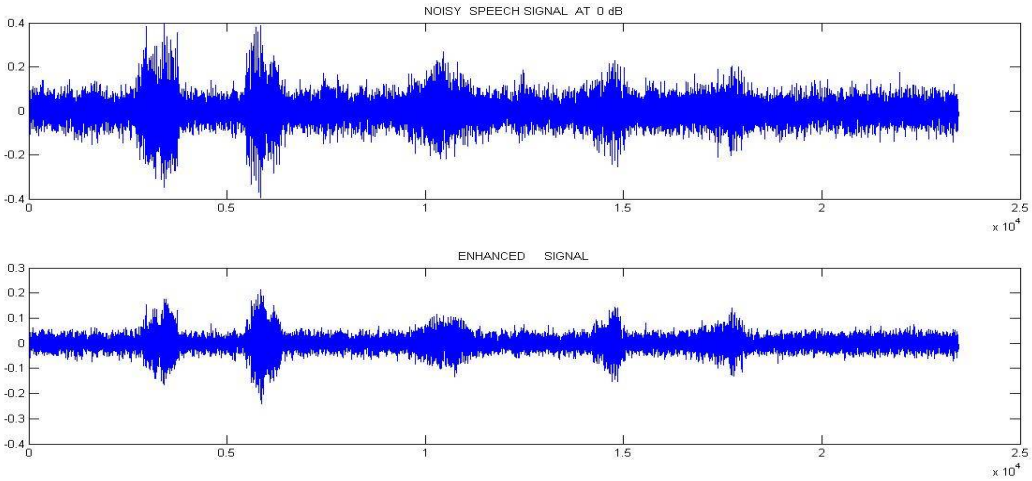

Figure 3: Noisy Signal and Enhanced Speech Signal using proposed method

The above figure 3 shows the waveform of the white Gaussian noise added noisy signal at 0 dB SNR and the enhanced signal obtained by using the proposed method.

Input Noises at 0 dB	WPT Method				Proposed Method			
	Pre SNR	Post SNR	I/P MSE $\times 10^{-7}$	O/P MSE $\times 10^{-7}$	Pre SNR	Post SNR	I/P MSE $\times 10^{-7}$	O/P MSE $\times 10^{-7}$
Airport	0 dB	1.6074 dB	0.31965	0.62545	0 dB	4.8426 dB	0.31965	0.32754
Babble	0 dB	1.7584 dB	1.1106	0.75191	0 dB	4.1476 dB	1.1106	0.98884
Car	0 dB	2.7350 dB	0.03318	0.030868	0 dB	5.7078 dB	0.03318	0.023885
Station	0 dB	2.8532 dB	0.63673	0.18600	0 dB	4.7702 dB	0.63673	0.00954
AWGN	0 dB	3.3357 dB	3.1841	0.053523	0 dB	3.4583 dB	3.1841	1.6643

Figure 4: Comparison Table

The comparison table are shown above in the figure 4.The above figure shows the different types of noises at 0 dB SNRs .The I/P MSE is shown in the table.The post SNRs and the Output MSE is also shown for the Wavelet Packet transform method and the proposed method.

To evaluate the performance of the proposed method, SNR and MSE is computed for various types of noises including additive white Gaussian noise at 0 dB's SNRs. The results obtained, from the simulation, compared in terms of bar chart for SNR and MSE are shown in figure 5 and figure 6 respectively.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

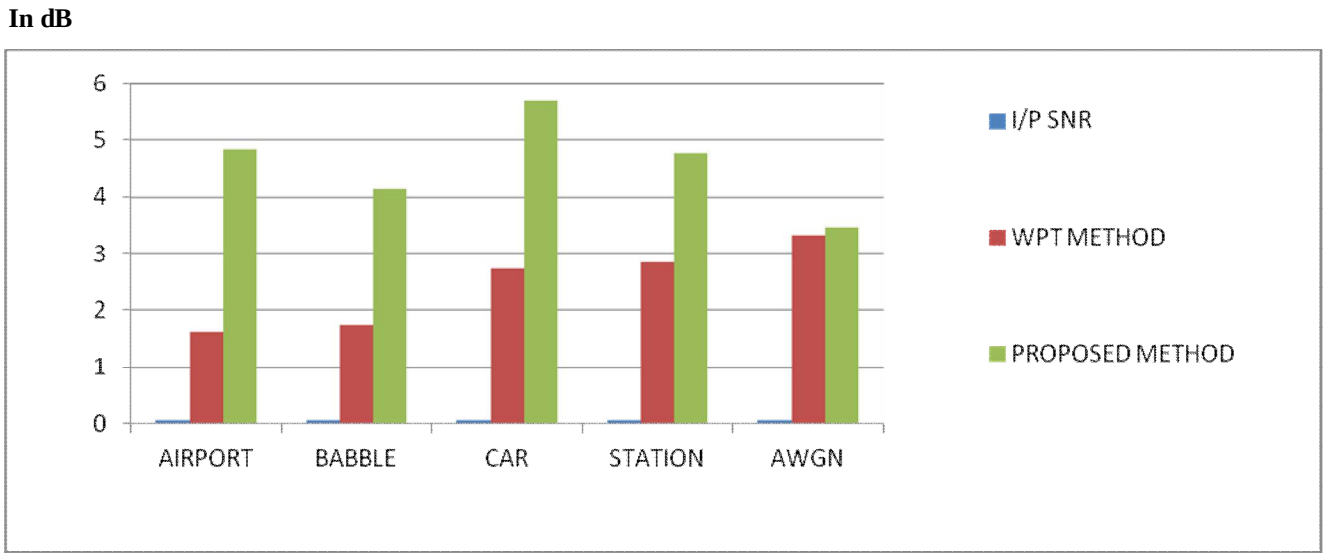

Figure 5: Comparison in terms of SNR

In terms of SNR, the proposed method is better than the WPT method .It is observed that the proposed method gives the best SNR values for all the different types of noises at 0dB.In order to be visible in the bar chart,I/P SNRs is taken as 0.05 dB.

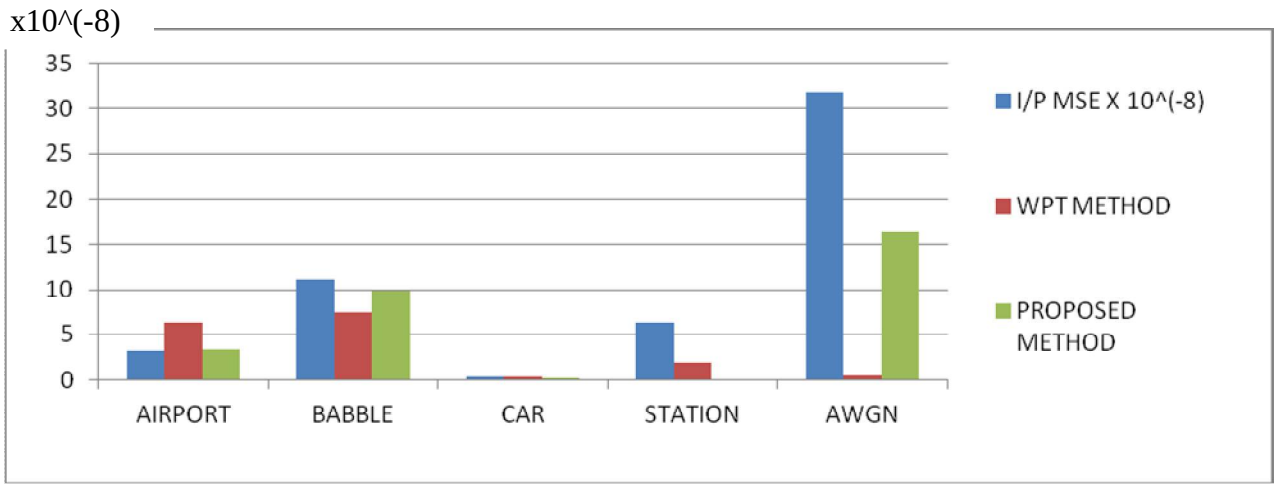

Figure 6: Comparison in terms of MSE

In terms of MSE, the proposed method is better for Airport noise, Car noise and Station noise etc. The lower the MSE, the better the signal quality. Lower MSE means the closer match between two signals.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

V. CONCLUSION

This paper presented a speech enhancement method based on the Continuous Wavelet transform using Morlet wavelet. It was observed that best results were obtained for the Input SNR's that lies between 0dB and 5dB.The proposed algorithm provided better results as compared to traditional DWT algorithm especially for high level noise (i.e. at 0 dB I/P SNR's).The present work is limited for Input SNR's (0 dB to 5 dB) and also the algorithm need to change the range of scale for different I/P SNR's, here, in this case the range of scale is fixed for different I/Ps SNRs. Future work on this approach will include the development of time-varying thresholding techniques based on tracking a priori SNR, such as those used in the Ephraim Malah comparative method and in some recent perceptual wavelet denoising techniques.

REFERENCES

[1]Jun Yao ,and Yuan-Ting Zhang ,"Bionic Wavelet Transform New Time Frequency Method Based on an Auditory Model , "IEEE Trans. Biomed. Eng .Vol no.48,No 8,pp. 856-863,August 2001

[2]Jun Yao,and Yuan-Ting 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,November 2002

[3]D.L. Donoho ,"Denoising by soft thresholding," IEEE Trans. Inform. Theory,Vol. 41,No. 3,pp. 613-627,1995

[4]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" in Speech communication (Elsevier),Vol. No. 70 ,pp. 1-12, 2015

[5]Michael T. Johnson,Xiaolong Yuan,Yao Ren "Speech signal enhancement through adaptive Wavelet thresholding" in Speech Communication (Elsevier) ,Vol. No. 49,pp. 123-133,2007

[6]Xiaolong. Yuan,"Auditory Model-based Bionic Wavelet Transform for speech enhancement," M.Sc. thesis,Milwaukee,Wisconsin,May 2003

[7]Milind Kansara, and prof. Neeta Chapatwala,"Noise Reduction from the Speech Signal using Wavelet Packet Transform" in IJECSE: International Journal of Electronics and Computer Science Engineering,ISSN 2277-1956,V2N2 ,pp 745-751

[8]R. Dhivya, and Judith Justin , "A Novel Speech Enhancement Technique" in IJRET: International Journal of Research in Engineering and Technology ,ISSN 2319-1163,Vol 3,pp. 98-102

[9]Rupali Sharma, Preety D. Swami, "Enhancement of Speech Signal by Adaptation of Scales and thresholds of Bionic Wavelet Transform Coefficients" in International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol.3,Special Issue 3,April 2014

[10]Jun Yao,and Y. T. Zhang, "From Otoacoustic Emission Modelling to Bionic Wavelet Transform, "Proceedings of the 22nd Annual EMBS International Conference, ,Chicago IL ,pp no. 314-316,July 23-28,2000.

[11]Rajeev Aggarwal, Jay Karan Singh, Vijay Kr. Gupta and Dr. Annubhuti Khare,"Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold" in IJCA 2011,Vol 20-No. 5,pp no.14-19,April 2011

[12]Kanu patel,Vishnu patel,Dipal patel and Hardik Patel, "A Modified Speech Enhancement Algorithm Based on the Discrete Wavelet transform" in Indian Journal of Applied Research,Vol 5,pp no.87-90,Jan,2015

[13]D.L. Donoho and I.M. Johnstone,"Threshold selection for wavelet shrinkage of noisy data",proc. 16th Annual conference of the IEEE Engineering in Medicine and Biology society,Marland,USA,1994

[14]Ephraim, Y., Malah, D., 1984,"Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust.Speech Signal Process.Vol. ASSP-32 No. 6,pp. 1109-1121,December 1994

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