Noise Reduction from the speech signal using WP coefficients and Modified Thresholding

Krunal S. Panchal
M. Tech Student
Electronics & Communication Engineering Department
Sagar Institute of Research & Technology, Bhopal, M. P. , India

Abstract

Performance of the thresholding based speech enhancement methods largely depend on the estimate of the exact threshold value as well as on the choice of the thresholding function. A speech enhancement method is presented, in which a custom thresholding function is used and employed upon the Wavelet Packet (WP) coefficients of the noisy speech. The thresholding function is capable of switching between modified hard, soft and supersoft thresholding functions depending on a parameter that decides the signal characteristics under consideration. The threshold is determined based on the statistical modeling of the Teager energy operated WP coefficients of the noisy speech. This custom thresholding function is very effective in reduction of the white noise from the noisy speech thus resulting in an enhanced speech with better quality and intelligibility.
Keywords: Noise Reduction, Wavelet coefficients, Thresholding function, Speech Enhancement

I. INTRODUCTION

The aim of speech enhancement, is to improve the intelligibility and/or quality of speech in order to facilitate better communication in noisy environments. Conventional speech enhancement techniques try to remove noise from a noisy speech signal with minimal impact on the speech itself. Another approach to speech enhancement, which has not received as much attention, is to modify the speech signal itself, by emphasizing certain acoustical cues, in order to make it more intelligible in noisy environments.

Speech has always been the most dominant and common way of communication. The information contained in the spoken word is conveyed by the speech signal. In order to analyze speech transmission and processing, we need to understand the basic structure of the speech signal and its production models.

A simple engineering model, referred to as the source/filter model, can thus be built based on this production mechanism. If we assume that the vocal tract is a linear time-invariant system with a periodic impulse-like input, then the pressure output at the lips is the convolution of the impulse-like train with the vocal tract impulse response, and therefore is itself periodic. This is a simple model of a steady-state vowel. The speech utterance consists of a string of vowel and consonant phonemes whose temporal and spectral characteristics change with time, corresponding to a changing excitation source and vocal tract system [15].

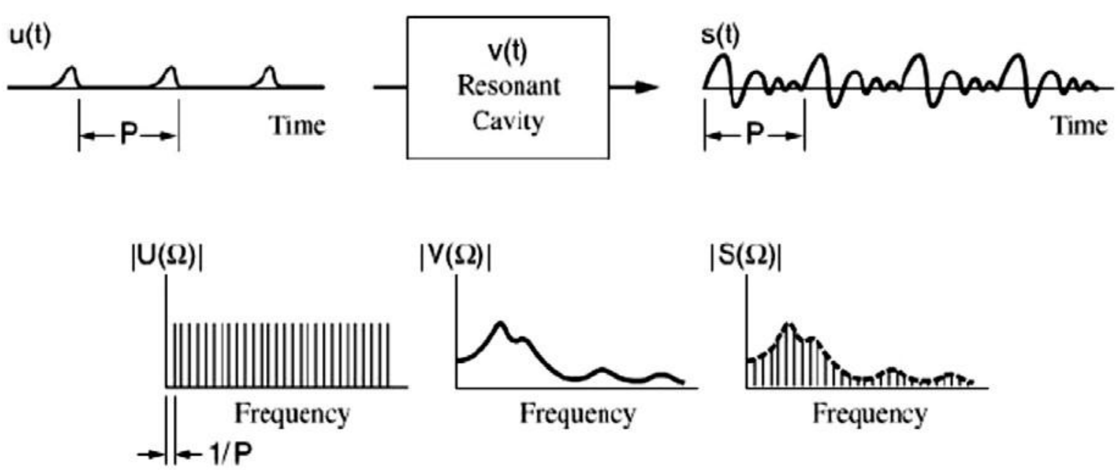


Fig. I: Speech production mechanism in time and frequency domain[15]

II. SPEECH ENHANCEMENT

The review on the nature of speech is used to describe parts of speech that constitute transient speech the speech component that we are developing an algorithm to extract. The aim of speech enhancement techniques is to improve the quality and/or intelligibility of speech. Conventional speech enhancement techniques try to improve the quality of speech by reducing the amount of noise in a noisy speech signal. An alternative approach to speech enhancement, which has not received as much attention, is to modify the speech signal itself before it is corrupted by noise to make it more intelligible in the presence of background noise.

In a communication system where either the speaker or listener is in a noisy environment, or the transmission channel is noisy, the intelligibility and quality of speech may be severely degraded, making communication difficult if not impossible. This may also fatigue the communicators as the speaker may have to raise his/her voice while the listener may have to concentrate more. The aim of speech enhancement systems is to improve the quality and/or intelligibility of speech and to reduce communicator fatigue in order to facilitate better communication in noisy environments.

Examples of applications where speech enhancement has provided substantial benefits include wireless communication, aircraft-control tower communication, within aircraft communication, speech recognition, and speech coding. There are two basic approaches to speech enhancement: noise reduction and speech modification.

A. Noise Reduction

This Section reviews several noise reduction techniques, including spectral subtraction, Wiener filtering and the minimum mean-square error short-time spectral amplitude estimator, based in part on the review by Lim and Oppenheim.

Spectral subtraction is a noise reduction technique that tries to estimate the short-time spectrum of an additive noise that is corrupting a speech signal. The estimated short-time spectrum of noise is subtracted from the short-time spectrum of noisy speech to obtain an estimate of the short-time spectrum of original speech, and the estimated spectrum is combined with the phase of noisy speech to estimate the original speech. These operations can be viewed as an attempt to enhance the speech-to-noise ratio by attenuating the short-time spectrum when the speech-to-noise ratio is relatively low and not attenuating the short-time spectrum when the speech-to-noise ratio is relatively high. Various spectral subtraction methods differ on how the estimate of the short-time spectrum of the additive noise is obtained.

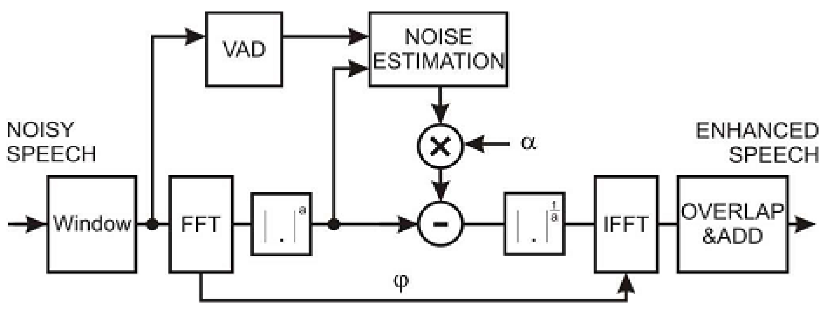


Fig. II : Spectrum Subtraction[11]

B. Speech Perception and Transient Speech

As mentioned earlier, an alternative approach to speech enhancement is to modify the speech signal itself before it is corrupted by noise to make it more intelligible in the presence of background noise. Efforts in the speech community to enhance speech in this manner are reviewed after the discussion on the acoustical cues that influence speech perception.
These studies relate the place of articulation of stop consonants to the patterns in transitions of formants observed on spectrograms. It was noted, however, that spectrographic patterns for a particular phoneme typically look very different in different contexts. For example, Liberman noted that /d/ in the syllable /di/ has a transition that rises into the second formant of /i/, while /d/ in /du/ has a transition that falls into the second formant of /u/ . The most important cues are sometimes among the least prominent parts of the acoustic signal. The studies cited above also accentuate the importance of formant transitions as acoustic cues for identifying and distinguishing some phonemes. Although these studies were conducted in noise- free environments, we expect the same acoustic cues to be important for identifying and differentiating phonemes in noisy environments.

C. Speech Enhancement by Speech Modification

Modified speech that emphasizes speech transitions can be created by applying time-, frequency- or time-frequency-domain processes to original speech. Thomas and Niederjohn processed speech by high pass filtering with a cutoff of 1100 Hz and an asymptotic attenuation slope of +12 dB/octave followed by infinite amplitude clipping . The clipper output waveform was strictly binary and its axis crossings represented those of the filtered signal in timing and polarity. Psycho-acoustic testing showed that their filtered/clipped speech was more intelligible in band-limited (frequency range of 250 to 6800 Hz) white noise than unmodified speech when both speech signals are presented at the same SNR. The filter cutoff frequency and asymptotic attenuation slope were determined by psycho-acoustic testing of a range of values.

D. Time Domain versus Transform Domain

Speech enhancement can be performed in both time and frequency domains. Time domain techniques include those utilizing Finite Impulse Response (FIR) filters and Infinite Impulse Response (IIR) filters, Linear Predictive Coefficients (LPC), Kalman filtering, Hidden Markov Models (HMM), etc. Transform domain techniques are techniques in which a transformation is first performed on the noisy speech before filtering, followed by the corresponding inverse transformation in order to restore the original speech.

The main advantage of performing the noise filtering or reduction process in the transform domain lies in the relative ease of distinguishing and removing noise from speech. For example, the energy of white noise is uniformly spread throughout the entire frequency spectrum, but the energy of speech, especially voiced speech, is concentrated in certain frequencies. The transform-domain approach is equivalent to the signal subspace approach coined by Ephraim in. As the speech signal only occupies certain bins, hence subspace, and the remaining bins are considered to be purely noise, noise removal/reduction could be achieved with ease to a certain extent. Hence ideally, the transform coefficients should be fully decorrelated and independent of each other. The transform should also be fast and non-computationally intensive for real time applications. Last but not least important is that the transform must be reversible.

E. Time Domain Methods

There are several time domain methods which directly work on the noisy speech signal without converting it into transform domain.

1) *Time Domain Noise Model*
2) *Comb Filtering*
3) *LPC Based Filtering*
4) *Kalman Filtering*
5) *HMM Filtering*
6) *Neural Networks*

F. Transform Domain Methods

The analog input speech is sampled at 8 kHz and quantized to 16 bits. The digitized speech is then partitioned into overlapping frames. The commonly used amount of overlap is 50% or 75%. Hanning window is then applied to each frame. The windowed frame then passes through the DFT transform stage. The output complex coefficients are then separated into magnitude and phase. The phase is left untouched while the noisy magnitude is filtered. After filtering the magnitude component is recombined with the phase, and the inverse transform operation is carried out. Following which, the overlap and add technique is applied to reconstruct the output speech signal.

1) *Psychoacoustics*
2) *Noise Variance Estimation*

III. WAVELETS

A. Continuous Time Wavelets

Consider a real or complex-valued continuous time function ψ(t) with the following properties

(1) The function integrates to zero

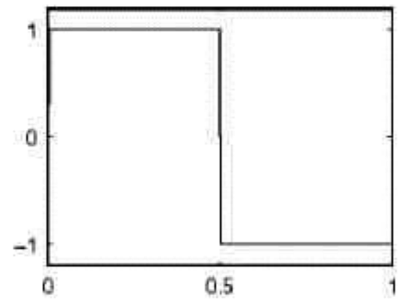
$$\int_{-\infty}^{\infty} \psi(t) \cdot dt = 0 \qquad (3.1)$$

(2) It is square integrable or, equivalently, has finite energy

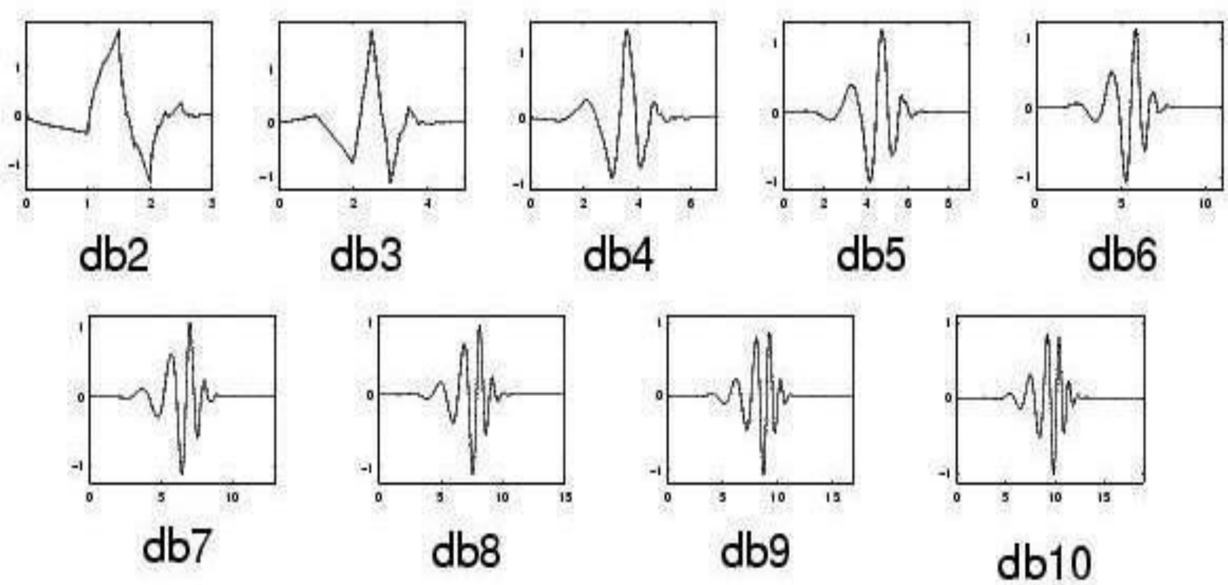
$$\int_{-\infty}^{\infty} |\psi(t)|^2 \cdot dt < \infty \qquad (3.2)$$

A function is called “mother wavelet” if it satisfies these two properties. There is an infinity of functions that satisfy these properties and thus qualify to be mother wavelet. The simplest of them is the ‘Haar wavelet’. Some other wavelets are ‘Mexican hat’, ‘Morlet’.

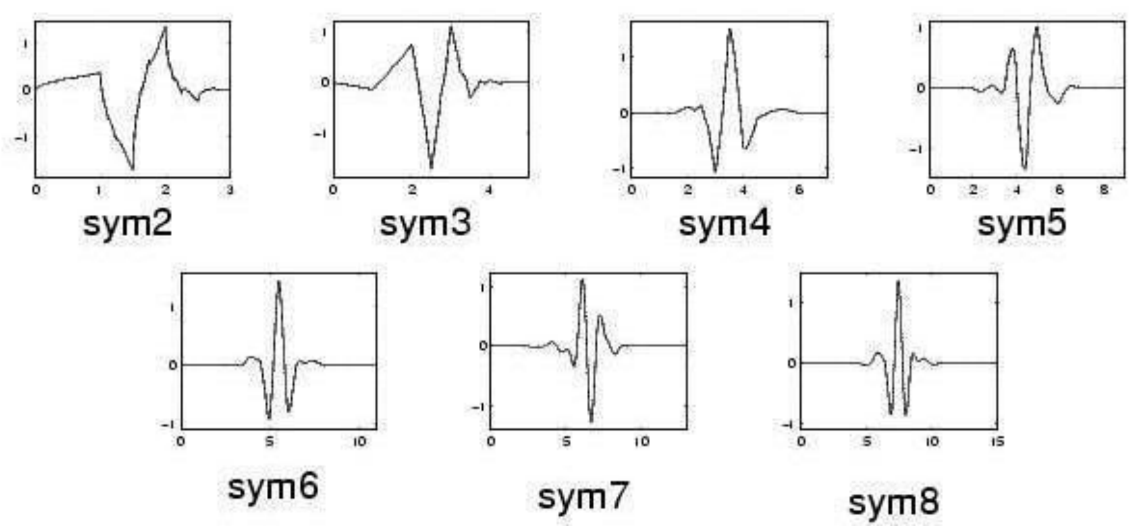
Apart from this, there are various families of wavelets. Some of the families are 'Daubechies' family, 'Symlet' family, 'Coiflet' family etc. In this project, the main stress is given on the Daubechies family, which has db1 to db10 wavelets.



(a) Haar Wavelet



(b) Daubechies family



(c) Symlet family

Fig. III: (a),(b),(c) : Some Wavelet Functions[10]

B. The Continuous Wavelet Transform (CWT)

Consider the following figure, a sinusoid and a wavelet.

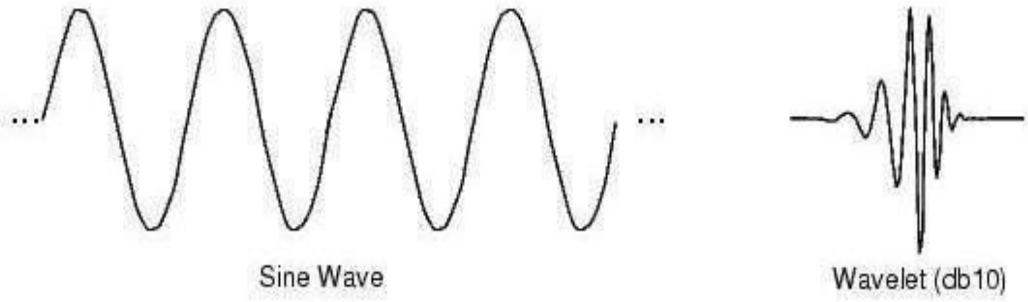


Fig. 4: Comparing Sine Wave And Wavelet[10]

We know that wavelet is a waveform of effectively limited duration that has an average value of zero.

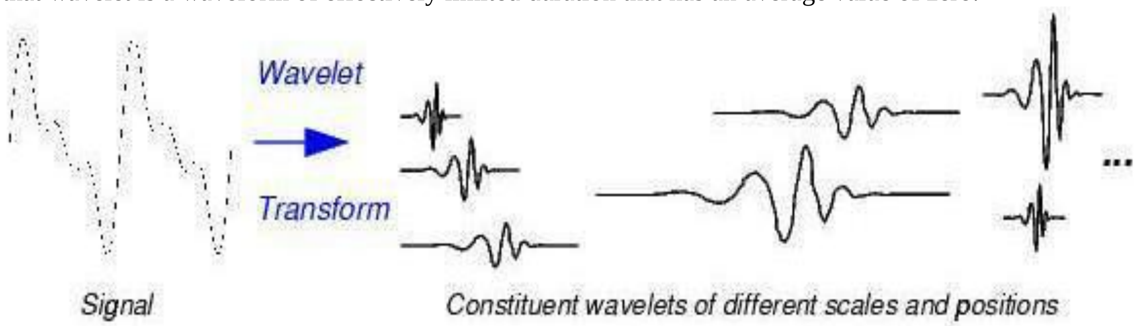


Fig 5: Decomposition of Signal into Wavelets[10]

The above diagram suggests the existence of a synthesis equation to represent the original signal as a linear combination of wavelets which are the basis function for wavelet analysis (recollect that in Fourier analysis, the basis functions are sines and cosines). This is indeed the case. The wavelets in the synthesis equation are multiplied by scalars. To obtain these scalars, analysis equation needed, just as in the Fourier case.

Here are the five steps of an easy recipe for creating a CWT.

Step. 1 : Take a wavelet and compare it to a section at the start of the original signal.

Step. 2 : Calculate a number, C, that represents how closely correlated the wavelet is with this section of the signal. The higher C is, the more the similarity. More precisely, if the signal energy & the wavelet energy are equal to one, C may be interpreted as correlation coefficient. The results will depend on the shape of the wavelet you choose.

Step. 3 : Shift the wavelet to the right and repeat steps 1 and 2 until you've covered the whole signal.

Step. 4 : Scale (stretch) the wavelet and repeat steps 1 through 3.

Step. 5 : Repeat steps 1 through 4 for all scales.

When you are done, you'll have the coefficients produced at different scales by different sections of the signal. The coefficients constitute the results of a regression of the original signal performed on the wavelets. And you can make a plot on which the x-axis represents position along the signal (time), the y-axis represents scale, and the color at each x-y point represents the magnitude of the wavelet coefficient C.

IV. THRESHOLD SELECTION ALGORITHMS

There are many formulas for obtaining threshold value. In this section we review some of the most popular of them. In all these formulas λ is the threshold value.

A. Universal method[4]

Donoho and Johnstone derived a general optimal universal threshold for the white Gaussian noise under a mean square error criterion and its side condition that with high probability, the enhanced signal f̂ is at least as smooth as the clean signal f In this method threshold is selected as:

$$\lambda = \hat{\sigma}\sqrt{2\log_e(n)} \quad (4.1)$$

In this formula n is number of samples in the noisy signal and σ is the standard deviation of noise that is estimated by the relation

$$\hat{\sigma} = \left[\frac{median(|Y_{ij}|)}{0.6745} \right] \quad (4.2)$$

In which Y_{ij} is the first level detail coefficients of wavelet transform of noisy speech.

B. Minimax method[2]

In this method that is also proposed by Donoho and Johnstone, it supposed that, Y=N(μ,1) is the observation, then λ is selected such that minimizes the following relation:

$$\Lambda_n^* = \inf_{\lambda} \sup_{\mu} \left\{ \frac{E(\sigma_\lambda(Y) - \mu)^2}{n^{-1} + \min(\mu^2, 1)} \right\} \quad (4.3)$$

Where Λ_n^* is the shrink function or thresholding algorithm and n is number of signal samples.

C. SURE method[2]

SURE or Stein Unbiased Risk is also introduced by Donoho and Johnstone for wavelet de-noising, this method denoises wavelet coefficients so that the mean squared error is minimized, where MSE is estimated by Stein's unbiased risk estimator based on the variance of the coefficients.

D. Teager Energy Operator[1]

The Teager Energy Operator (TEO) is a powerful nonlinear operator proposed by Kaiser (1993), capable to extract the signal energy based on mechanical and physical considerations.

$$y[n] = y[n] - y[n+1]y[n-1] \quad (4.4)$$

E. Other Thresholding Algorithms

There are many thresholding algorithm for removing noise from the noisy signal, some of them are as follows.

1) Hard Thresholding[6]

Hard thresholding can be described as the usual process of setting to zero the elements whose absolute values are lower than the threshold. The hard threshold signal is x if x > thr and is 0 if x < thr, where thr is a threshold value.

$$T_{Hard}(x) = \begin{cases} x & |x| \geq thr \\ 0 & |x| < thr \end{cases} \quad (4.5)$$

2) Soft Thresholding[6]

Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards 0. If x thr, soft threshold signal is (sign (x). (x - thr)) and if xthr , soft threshold signal is 0. Hard thresholding is the simplest method but soft thresholding has nice mathematical properties and gives better denoising performance.

$$T_{Soft}(x) = \begin{cases} Sign(x).(x - thr) & x \geq thr \\ 0 & -thr \leq x < thr \\ Sign(x).(x + thr) & x < -thr \end{cases} \quad (4.6)$$

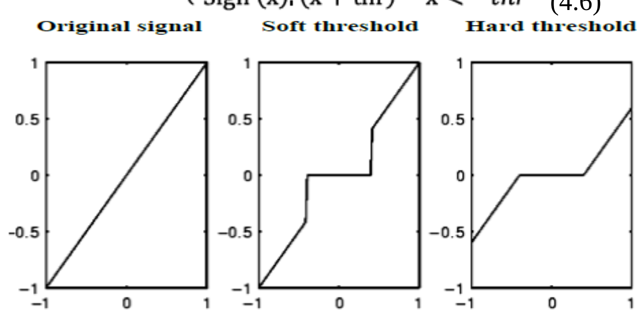


Fig. 6: Hard and Soft Thresholding [5]

3) Super Soft Threshold [2]

In the Super-Soft thresholding algorithm instead of setting some wavelet coefficients to zero, the algorithm attenuates the coefficients depending on their distance from the threshold. This idea is based on the fact that forcing some wavelet coefficients to zero causes observable sharp time-frequency discontinuities in the speech spectrogram that can decrease the quality of the enhanced speech signal.

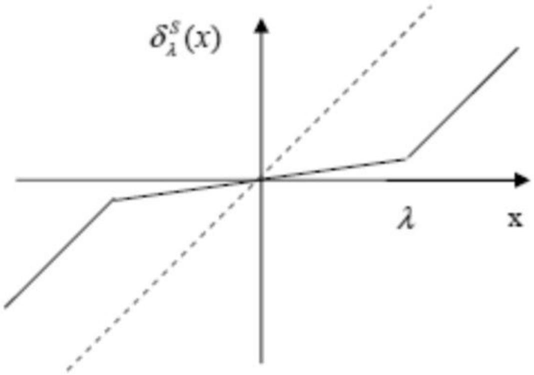


Fig. 7: Super Soft Thresholding

V. IMPLEMENTATION

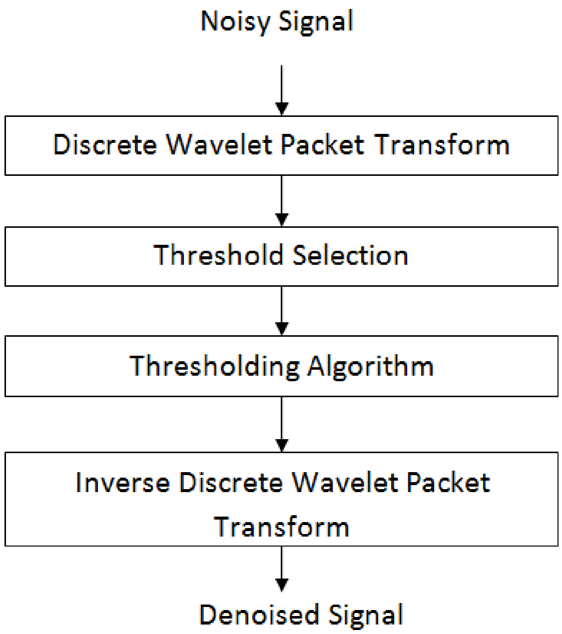


Fig. 8 : Basic Wavelet packet Thresholding[5]

Implementation of Basic Noise removal algorithm using wavelet packet transform is shown in flowchart. A male speech signal has been taken as original signal. The sampling frequency is 8kHz and 40000 samples of the signal are used. White Gaussian noise (WGN) is used to model the background noise.

Mean Square Error:

Mean square error (MSE) used for comparison between original signal (without noise) and recovered signal using wavelet technique.

$$\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{N}}\left(\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}}(\boldsymbol{x}(\boldsymbol{n}) - \hat{\boldsymbol{x}}(\boldsymbol{n}))^{\mathbf{2}}\right) \qquad (6.1)$$

Where, x(n) is the original speech signal and x~(n) is estimated speech signal obtained by the proposed method and N is number of samples in the signal.

Here super soft thresholding and level three decomposition is used for noise remove from speech signal. The noisy speech signal used is with SNR 20dB and we get mse 0.004683.

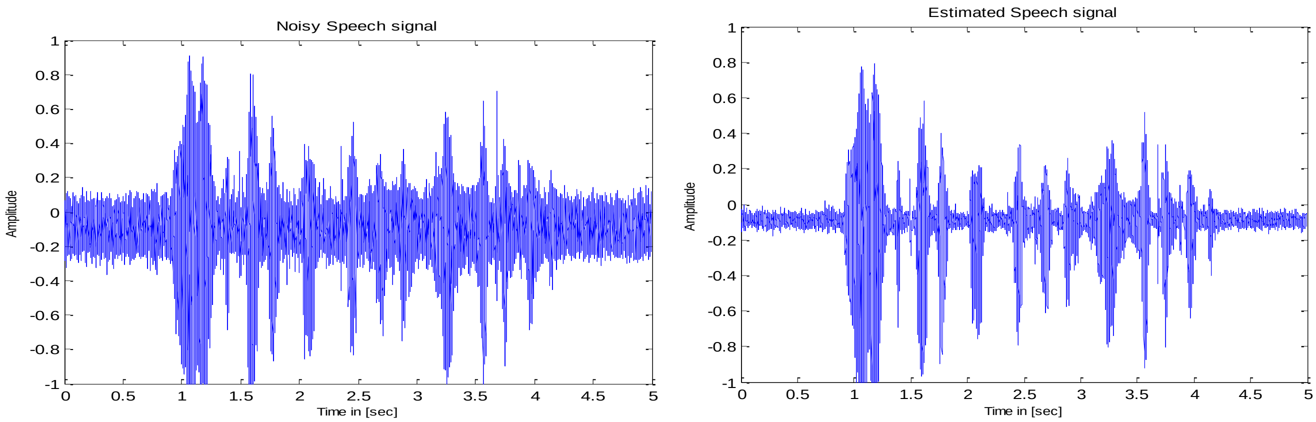


Fig. 9: Noisy and Estimated Speech Signals

Table - 1
MSE Comparison for Different Thresholding

Sr. No.	SNR (in dB)	Mean Square Error		
		Soft Thresholding	Super Soft Thresholding a = 0.08	Super Soft Thresholding a = 0.008
1	5	0.039626	0.039124	0.038236
2	10	0.017139	0.017021	0.016799
3	15	0.008793	0.008692	0.008629
4	20	0.004910	0.004819	0.004683
5	25	0.002947	0.002852	0.00279

VI. CONCLUSION

An improved wavelet thresholding speech enhancement system is used, which uses the Super-Soft thresholding algorithm to improve the noisy speech using wavelet coefficients in a way that avoids sharp time-frequency discontinuities in the speech spectrogram that can decrease the quality of the enhanced speech signal. During different analysis it is found that super soft thresholding is better than hard and soft thresholding in acoustic sense. Higher threshold removes noise well, but the part of original signal is also removed with the noise. So it is the trade of between threshold value and speech signal quality.

REFERENCES

[1] Tahsina Farah Sanam, Celia Shahnaz, "An Approach for Noisy Speech Enhancement through soft thresholding function by Employing the Teager Energy Operation on Wavelet Packet" in TENCON - IEEE Region 10 Conference, Dhaka, Bangladesh, November 2011.
[2] Saeed Ayat, "A New Method for Threshold Selection in Speech Enhancement by Wavelet Thresholding" in ICCCM-2011.
[3] S. Li, Y. Tian, G. Lu, Y. Zhang, H. Xue, J. Wang , and X. Jing, "a new kind of non-acoustic speech acquisition method based on millimeter wave radar", In Electromagnetic Research, Vol. 130, 17–40, 2012.
[4] Yujie Tian and Li Wang, "An Improved Denoising Method for Speech Signal" in IEEE 2nd International Conferance AIMSEC – August 2011.
[5] Rajeev Aggarwal, Jay Karan Singh, Vijay Kr. Gupta and Dr. Anubhuti Khare "Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold" in International Journal of Computer Applications, Volume-20, No.5, April 2011.
[6] Hui Ou, John S. Allen and Vassilis L. Syrmos, "Frame-Based Time-Scale Filters for Underwater Acoustic Noise Reduction" in IEEE Journal of oceanic engineering, April 2011.
[7] Z. A. Jaffery , K. Ahmad and Pankaj Sharma, "Estimation of Speech Signal in the presence of White Noise using Wavelet Transform" in Int. Conf. on Control, Communication and Power Engineering ACEEE 2010.

[8] Y. Ghanbari, M. R. Karami and S. Y. Mortazavi "A New Speech Enhancement System Based on the Adaptive Thresholding of Wavelet packets" in ICEE 2005.
[9] Zhang Jun-chang, Zhang Yi and Ye Zhen "Adaptive Speech Enhancement Based on Classification of Voiced/Unvoiced Signal" in IEEE International Conference on Multimedia and Signal Processing, 2011.
[10] Stephane Mallat, "A Wavelet Tour of Signal Processing", *3rd ed.*Academic Press, dec. 2008.
[11] Ekaterina Verteletskaya and Boris Simak , "Noise Reduction Based on Modified Spectral Subtraction Method", in IJCS 2011.
[12] B. Widrow and S. D. Stearns, "Adaptive noise canceling: Principles and applications, Practise-Hall Signal Processing Series , New Jourcy , 1985
[13] Simon Haykin, "Adaptive Filter Theory", Kindersley Publication, 1995.
[14] S. M. Kuo and D. R. Morgan, "Active Noise Control Systems-Algorithms and DSP Implementations", New York: Wiley, 1996
[15] Thomas F. Quatieri, "Discrete-Time Speech Signal Processing: Principles and Practice" New Jersey: Prentice Hall, 2001.