SPEECH SIGNAL DENOISING WITH WAVELETS

IMAN KHALIL ALAK

APRIL 2018

ÇANKAYA UNIVERSITY

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

MASTER THESIS

SPEECH SIGNAL DENOISING WITH WAVELETS

IMAN KHALIL ALAK

APRIL 2018

Title of the Thesis : **SPEECH SIGNAL DENOISING WITH WAVELETS**

Submitted by **Iman Khalil Alak**

Approval of the Graduate School of Natural and Applied Sciences, Çankaya University

Prof. Dr. Can ÇOGUN
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Dr. Öğr.Üyesi Özgür ERGÜL
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Dr.Öğr.Üye. Selma ÖZAYDIN
Supervisor

Examination Date : 16.05.2018

Examining Committee Members

Dr. Öğr. Üye. Selma ÖZAYDIN

Doç. Dr. Nursel AKÇAM

Dr.Öğr.Üye. Özgür ERGÜL

STATEMENT OF NON-PLAGIARISM

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as require by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : **IMAN KHALIL**

Signature : iman

Date : 21-05-2018

ABSTRACT

IMAN, Khalil

M.Sc., Department of Electronic and communication Engineering

Supervisor: Asst.Prof. Dr. Selma ÖZAYDIN

April 2018, 82 pages

This study aims to examine the performance of wavelet transform for denoising of a speech signal. Wavelets are widely used in digital speech processing, especially in coding, enhancement or noise removing of a speech signal. In many conditions, recognizing natural speech is a challenging task due to the background noise in it. The goal of a speech denoising algorithm is to recover original speech signal by removing noise with a minimum distortion. There are various methods to help restore speech from noisy distortions. Many of the used deniosing algorithms perform this procedure in frequency domain where the power spectral density (PSD) function of the noisy signal can be examined in a short time frame. Then, the short-time spectral frequency and amplitude of clean speech is estimated for per frame of the noisy signal. As a result, estimation errors are introduced by the limitations of methods. Various spectral estimation techniques have been investigated for decades to reduce the estimation errors.

In this study, discrete wavelet transform technique is used for denosing of an input noisy speech signal. The performance of discrete wavelet transform is evaluated by using different wavelet filters such as Daubechies, Symlets or Coiflets. The analysis was performed on MATLAB software. As an input noisy speech signal, different types of environmental background noises were analyzed such as babble noise (crowd of people) or noisy speeches with different type of background vehicle noises (cars, train, plane

etc.). They were filtered from the speech signal by wavelet analysis. The input noisy speech signal was decomposed by applying four different threshold selection to the wavelet coefficient: sgtwolog, heursure, rigrsure, and minimaxi thresholding, with hard or soft thresholding techniques. Reconstructed speech was compared with the original speech signal by measuring the signal-to noise ratio (SNR) and MSE values between noisy and output signals. Contributions include detailed analysis of comparison of different wavelet family performances against different background noise types and the discovering an effective method (Maximal overlap DWT-MODWT) for denoising of noisy speech signals.

Keywords: Wavelet transform, wavelet families, thresholding method, speech processing.

ÖZET

IMAN, Khalil

Yüksek Lisans, Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Tez Yöneticisi: Dr.Öğr.Üye Selma ÖZAYDIN

Nisan 2018, 80 sayfa

Bu çalışma konuşma sinyalinden gürültünün arındırılması için dalgacık dönüşümünün performansını incelemeyi amaçlamaktadır. Dalgacıklar sayısal konuşma işlemede özellikle kodlama, iyileştirme veya gürültü temizlemede yaygın olarak kullanılırlar. Pekçok koşulda, doğal konuşma sinyalini anlama arkaplan gürültüsü nedeniyle zorlu bir iştir. Konuşma gürültüsü temizleme algoritmasının amacı gürültüyü minimum bozulmayla temizleyerek orjinal konuşma sinyalini kurtarmaktır. Konuşma sinyalini gürültüden temizlemede kullanılacak değişik metotlar mevcuttur. Kullanılan gürültü temizleme algoritmalarının pekçoğu bu işlemi, gürültü sinyalinin güç spectral yoğunluğunun kısa pencere aralıklarında incelenebildiği frekans düzleminde gerçekleştirir. Daha sonra, gürültülü sesin herbir pencere aralığı için temiz sesin spectral frekans ve genliği tahmin edilir. Sonuç olarak, metotlara bağlı olarak tahmin hataları ortaya çıkar. Tahmin hatalarını minimuma indirmek için yıllardır değişik spectral tahmin teknikleri araştırılmıştır.

Bu çalışmada, gürültülü konuşma sinyalini temizlemede kesikli dalgacık dönüşüm tekniği kullanılmıştır. Kesikli dalgacık dönüşümünün performansı Daubechies, Symlets veya Coiflets gibi dalgacık filtreler kullanılarak değerlendirilmiştir. Analiz MATLAB yazılımı üzerinde gerçekleştirilmiştir. Gürültülü konuşma sinyali olarak babble gürültü (kalabalık insan grubu) veya farklı tipte arkaplan araç gürültüleri (arabalar, tren, uçak vs) gibi çevresel arkaplan gürültüleri içeren konuşmalar analiz edilmiştir. Bunlar konuşma sinyalinden dalgacık analizle temizlenmiştir. Gürültülü konuşma sinyali, soft ve hard eşikleme teknikleri içeren Sgtwolog, Heursure, Rigrsure ve Minimaxi eşikleme teknikleri olarak dört farklı eşik metodu kullanarak alt parçalara bölünmüştür. Tekrar oluşturulan konuşma sinyali ve gürültülü sinyal karşılaştırılarak sinyal-gürültü oranı (SNR) ve hatanın ortalama karekökü (MSE) hesaplanarak ölçülmüştür. Çalışmanın katkıları, farklı wavelet ailelerinin farklı arkaplan gürültülerine karşı performans kıyaslamalarının detaylı analizi ve gürültülü konuşma sinyalinden gürültü temizleme için etkin bir metodun (Maximal overlap DWT-MODWT) ortaya konmasıdır.

Anahtar Kelimeler: Dalgacık dönüşümü, dalgacık ailesi, eşikleme metodu, konuşma işleme.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Prof. Dr. Selma ÖZAYDIN for his supervision, special guidance, suggestions, and encouragement through the development of this thesis.

It is a pleasure to express my special thanks to my mother for her valuable support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DFT	Discrete Wavelet Transform
DWT	Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
FT	Fourier Transform
STFT	Short Time Fourier Transform
WT	Wavelet Transform
HPF	High Pass Filter
LPF	Low Pass Filter
FIR	Finite Impulse Response
AWGN	Additive White Gaussian Noise
SNR	Signal to Noise Ratio
MSE	Mean Square Error
Db	Daubechies Wavelets
MODWT	Maximal Overlap Discrete Wavelet Transform

CHAPTER I

INTRODUCTION AND PREVIOUS WORK

1.1. Introduction

The Wavelets transform is a helpful instrument for multi-resolution signal analysis. Therefore, it has been used in a wide scope of applications in field of science and engineering including the electrical engineering (i.e., signal processing, speech processing, image processing, information extraction, and data compression), mathematics (especially for harmonic analysis and operator theory) and physics (i.e., quantum theory). In the case of digital signal processing, separate wavelet transform is gained by first sampling of the signal and then filtering with low-pass and high-pass Finite Impulse Response (FIR) filters. Signal de-noising and perfect reconstruction can be realized easily for orthonormal wavelets. Yves Meyer [1] has proposed a very popular orthonormal wavelet method. Since then, a lot of scientists studied on the theory and become well known due to the specific characteristics of the method. Wavelet is a modern technology method to progress signals. Wavelet analysis is a mathematical model that transforms the original signal (especially in time domain) into time-frequency domain for analysis and processing. Wavelet analysis is also very appropriate method for non-stationary data analysis.

It is usually possible to describe a wave as “an oscillating function of time or space”, such as a sinusoid. If we look at Fourier analysis, it is considered as a wave analysis. It extends time domain signals in terms of sinusoidal signals which are very valuable in mathematics, science and engineering fields. They have special advantages for periodic, time-invariant, stationary or non-stationary (time-varying) signals.

A wavelet is just a "small wave" that's a power accumulated with time and gives instrument for the analysis of transient or non-stationary signal. It's not merely the oscillating wave characteristic but even offers the capacity to allow simultaneous time and frequency analysis with a versatile mathematical formulation [13].

When we look through it historically, the theory of “wavelets” came from the study of time-frequency signal analysis, wave propagation, and sampling theory. Among the typical reasons of using wavelet transforms is that the Fourier transform analysis is not in a position to concern to the local information of signals. So, the Fourier transform is not used for analyzing signals in a joint time -frequency domain of localized signals. The idea of wavelets for the analysis of non-stationary signals by using translation and dilation of a single function was first introduced by Jean Morlet in 1982 and it was known as mother wavelet. Since then, this new concept can be viewed as the synthesis of various ideas. These ideas include so many disciplines like mathematics, physics, and engineering [14].

According to Morlet’s analysis, signals are consisted with many properties in time and frequency, but their high-frequency components have shorter time duration than their low-frequency components. So as to obtain better time resolution for the high-frequency transients and good frequency resolution for the low-frequency elements, Morlet et al. (1982a,b) investigated the concept of wavelets as a family of activity known as mother wavelet. The mother wavelets are defined by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

Where ‘a’ is a scaling parameter that measures the degree of compression or scale, and ‘b’ is a translation parameter that causes the time location of the wavelet. When $|a| < 1$, then the wavelet is a compressed version (smaller support in time-domain) of the mother wavelet and it has a close similarity mainly to higher frequencies. On the other hand, when $|a| > 1$; $\Psi_{a,b}(t)$ has a greater time-width than previous one and it corresponds to lower frequencies. So, wavelets have time-widths adjusted to their frequencies. That is why the Morlet wavelets are successful in signal processing and time–frequency signal analysis [13].

Since Wavelets are characterized by scale and position, they are very useful in analyzing variations in signals and images in terms of scale and position. As its size varies, wavelet has many advantages over the classical signal processing transformations to process data at the same time in time and frequency. Wavelets are compressed at low scale and they fit to fast-changing details to a high frequency. At high scale, the wavelets are stretched and they correspond to slow changing properties to a low frequency. Wavelets can be scaled and this scaling is known as "daughter wavelets" of a finite-length or fast decaying oscillating wave form and this is known as the "mother wavelet".

When we compare Wavelet transforms and traditional Fourier transforms, we can get more advantages of the wavelets over the Fourier transform for the signals. Functions of traditional Fourier transforms have discontinuities and sharp peaks. They are used for correctly deconstructing and reconstructing finite, non-periodic and/or non-stationary signals. It is possible to classify Wavelet transforms into two groups. These two classifications are *discrete* wavelet transforms (DWT) and *continuous* wavelet transform (CWT). CWT stands for continuous-time (analog) signals. It operates over every possible scale and translation. DWT is used for digital signals. It is used for a specific subset of scale and translation values or representation grid [8].

The scale of the functions of Wavelet transform varies in its frequency. The importance of wavelets, in fact, it is used to analyze a function as a sum of time-shifted and scaled representations of itself which is known as mother wavelet. This is to represent the function as a sum of weighted delta functions as in the case of time domain, or as a sum of weighted sinusoids which is known as the frequency domain.

The main advantage of wavelet transforms is that, wavelet analysis is used in a longer area for low-frequency information and in a shorter one for high-frequency information [15].

Audio denoising is used to attenuate noise signal. What we must bear mind is that noise can be in white noise, pink noise, realistic noise or any other kind of noises. Many non-stationary signal processing applications try to remove these types of realistic noises, but they use methods that create major damages to the non-stationary signals. There are special capabilities of the wavelet transform.

In a wavelet transform, an irregular signal is decomposed at a predefined level by first selecting a wavelet type. This process is known as scaling. The magnitude of the input audio signal is partitioned into smaller parts. In addition to this, decomposition process gives various approximations. Moreover it provides detail coefficients. But these coefficients still contain noise residuals. So, to reduce the noise level, hard or soft thresholding techniques must be used. By doing so, the produced audio signal is reconstructed using reconstruction process. [16]

The fulfillment of denoising, can be calculated by using a signal to noise ratio (SNR) method. Here, denoising is process of removal of noise from an audio signal using wavelet transform. The quality of denoised signal is evaluated using an SNR calculation formula.

One-directional wavelet transform is one of the operative transforms and it’s used for processing audio signals by providing outstanding results especially for noisy audio signals. Such type of wavelet transform allows the signal to concentrate more efficiently during the signal processing. Fourier transform, in the contrary, supplies poor form of spectrum information over the time. Moreover, due to the insufficient capacity of Fourier transform to process non-periodic signals, short time Fourier transform (STFT) is used. However, the STFT has some drawbacks due to the limit in its time-frequency resolution capability (the uncertainty principle). In STFT, it is hardly difficult to localize on short pulses due to the fixed window size. Wavelet supplies variable window sizes which are used to analyze different frequency components in an audio signal. Moreover, there are more qualities in wavelet transform that give opportunity to prefer it to Fourier transform and STFT [16].

Wavelet transform is prefered for signal denoising. Because it has different advantages for application. For example its time frequency localization and multi resolution analysis are among its advantages. Denoising is done by soft or hard thresholding of DWT coefficients. In wavelet analysis low frequency coefficients particularly stands for signal. Its high frequency coefficients represent noise. It is possible to get Denoising by selecting a threshold and removing the noise from these high frequency coefficients according to the selected threshold [12].

1.2. Literature Review on Wavelet Speech Denoising

G. Renisha and T. Jayasree [20], discusses the enhancement of speech signals in a noisy environment depending on Wavelet and Adaptive filtering (WAF). The paper elaborates that the speech signals polluted by noise are manipulated with WAF. The adaptive algorithm updates the weighting factors of the filter by using least squares to make the output exactly fit with the output which is the signal at a minimum level. This is necessary to avoid the unnecessary noise and interferences. The performance of the suggested technique is evaluated by computing the Signal-to-Noise Ratio (SNR), Peak Signal-to-Noise Ratio (PSNR), Root-Mean-Square-Error (RMSE), Percentage Root Mean Square Difference (PRD) after denoising. Such a noise was taken from the AURORA database. The signals are sampled at 8 kHz.

This paper talks about the speech signals improvements of Wavelet that works depending by WAF. The outcomes of denoising method are compared with the other traditional methods. The application out comes have showed that the process of WAF method supplies better denoising consequence and good error resolving in the presence of any noise.

M.A. Abd El-Fattah, M.I. Dessouky [19] suggests an adaptive Wiener filtering method for speech enhancement. This technique relies on the changing of the filter transfer function from sample to sample based speech signal statistics which are the local mean and the local variance. It is applied in the time domain instead of the frequency domain to lodge for the time-varying property of the speech signals. The suggested technique is compared with the traditional frequency domain Wiener filtering, spectral subtraction and wavelet denoising methods using various speech quality metrics. The simulation results show the superiority of the proposed Wiener filtering method in the case of

Additive White Gaussian Noise (AWGN) as well as colored noise. For the evaluation purpose, they have used a speech signal for the statement “We were away year ago” for a male and for a female. They use speech quality metrics such as the SNR, segmental SNR (SNRseg), Log-Likelihood Ratio (LLR) and Spectral Distortion (SD).

E.Seke, M. H. Durak, et al.[21] , the suggested a method separating the speech data into magnitude and phase where in fact the magnitude part is further separated into common and difference parts by using Common Vector Analysis (CVA). It's stated that the noise largely resided on various part and therefore it was denoised by way of a known algorithm. The speech data is reconstructed by merging three things. These are difference, common, and phase parts. By using Linear Minimum Mean Square Error Estimation algorithm on the difference part; excellent denoising details are found. These details are measured with the state of the art speech quality measures. Tests were conducted on 30 sentences spoken by 6 people who have added 4 quantities of 8 structured background noise recordings (total of 5760 recordings per method per quality measure). They concluded that the proposed CVA method was higher-ranking against the other 5 methods. However, for the tests using white noise, CVA wasn't the most effective (total of 720 recordings per method per quality measure). But, in the majority of the experiments that CVA wasn't the most effective, it’s scores were almost the best ones.

B. JaiShankar and K. Duraiswamy [22]], an audio denoising technique based on wavelet transformation is proposed. The paper presents an audio denoising technique based on block matching technique. Denoising is performed in the transformation domain and the improvement in denoising is achieved by a process of grouping closer blocks. The technique exposes each and every finest details contributed by the set of blocks and also it protects the vital features of every individual block. The blocks are filtered and replaced in their original positions. The grouped blocks overlap each other and thus a much different estimation is obtained for every element. The technique based on this denoising strategy and its efficient implementation is presented in full detail. The implementation results reveal that the proposed technique achieves a state-of-the-art denoising performance in terms of both signal-to-noise ratio and audible quality. The technique is based on the denoising strategy and its efficient implementation is presented in full detail. The implementation results reveal that the proposed technique achieves a state-of-the-art denoising performance in terms of both signal-to-noise ratio and audible quality. This paper presented an audio denoising technique based on block matching technique. The technique was based on the denoising strategy and its efficient implementation was presented in full detail. The implementation results have revealed that the process of block matching has achieved a state-of-threat denoising performance in terms of both peak signal-to-noise ratio and subjective improvement in the audible quality of the audio signal.

Gathering of the comparable blocks enhanced the productive activity of the strategy. The blocks were separated and supplanted in their unique positions from where they were withdrawn. The grouping blocks were covering each other and consequently for each component a very different estimation was acquired that were joined to expel clamor from the input signal. The lessening in the noise level deciphers that the system has secured the essential exceptional highlights of every individual block even if the finest details were contributed by gathered blocks. Also the strategy can be changed for different other audio signals and in addition for different issues that can be advantage from profoundly linear signal portrayals.

A.Sumithra,B. Thanushkodi [23] , in this work, speech improvement is proficient using distinctive thresholding on time versatile discrete Daubechies wavelet transform co-efficients. However, the delicate thresholding is best in diminishing noise yet most exceedingly bad in protecting edges, and strong thresholding is best in safeguarding edges yet most noticeably bad in de-noising. Spurred by finding a more broad case that consolidates the delicate and hard thresholding to accomplish a tradeoff between the two strategies, the trimmed thresholding technique is proposed in this paper to intensify the discourse from foundation noise. The performance of distinctive thresholding strategies are assessed by enhancing the speech debased by different noises. At last, the objective and subjective trial comes about demonstrate that the proposed scheme with trimmed thresholding is better in denoising as compared than hard and soft thresholding strategies. It also shows that the proposed technique gives better mean square error (MSE) execution than other wavelet thresholding techniques.

A period versatile wavelet with trimmed thresholding process for de-noising speech from various uproarious conditions has been introduced. Upgrade comes about exhibit that the proposed conspire indicates preferred execution over hard and soft thresholding techniques to de-commotion the signal.

Milind U., Satish K. [24], discusses the single channel speech enhancement procedures in light of Spectral Subtraction (SS), Wavelet Transform (WT) and Adaptive Wiener Filtering (AWF) in this paper. Here quantitative execution of this speech upgrade procedures is analyzed and the parameters utilized for correlation are Mean Square Error (MSE).

The MSE, Signal to Noise Ratio, Peak Signal to Noise Ratio and Average Absolute Distortion. The outcomes got demonstrate the speech upgrading capacity of the personal communication method. The noise and echo-interference can debase the first original speech. From the outcomes we presume that the execution of single channel speech improvement that depends on WT is superior to AWF and SS procedures.

CHAPTER II

WAVELET FILTERS AND WAVELET TRANSFORMS

2.1 Rule of Wavelet Transform

The wavelet transform was created as an elective way to deal with the short time Fourier transform (STFT) to solve the determination issue which is found in Fourier Transform. In the wavelet investigation the signal is increased with a window function. The change is figured independently for various portions of the time- domain signal. There are two principle contrasts between the STFT and the Wavelet change: the Fourier transforms of the windowed signals which are not taken, and hence single pinnacle is seen relating to a sinusoid. The other one is that the width of the window. This one is changed as the transform is processed transform which is presumably the most important quality for the wavelet transform. The term wavelet implies a little wave. This alludes to an oscillatory window function of limited length. The primary window function is known as mother (original) wavelet. The term mother infers that the functions with various locale utilized in the transformation process are gotten from this mother wavelet. As it were, the mother wavelet is a model for producing the other window functions.

2.2 The Continuous Wavelet Transform Theory

The procedure of Fourier analysis is indicated by the Fourier transform (FT) formula as follow:

$$F(w) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi \mathrm{ft}} dt \tag{2}$$

Where Fourier Transform disintegrates the f (t) signal into complex exponential functions of various frequencies

Here, the frequency component "f" shows up at time t1 or t2 has equal impact in the outcome on the combination. The Fourier transform tells just whether a specific recurrence segment exists or not and doesn't give a data about the time that this segment shows up. This is the reason Fourier transform isn't appropriate for the signals having time fluctuating recurrence.

For the region where the signal can be thought to be stationary, we can look at that signal with limited windows and the signals which are found in these windows are expected as stationary. This form of the Fourier transform is called as The Short Time Fourier Transform (STFT).

There is just a minor contrast amongst STFT and FT. in the case of STFT, the signal is divided into little fragments with a windowing function and the signal is thought to be stationary of these fragments. The width of this window is equivalent to the portion of the signal. The window work and the signal are then duplicated. The short coming of STFT is that one can't know the correct time- intervals portrayal of a signal, and can't comprehend what ghastly segments exist at what occurrences of times. One can just know the time interims in certain band of frequencies which is a resolution issue. Another short coming of the STFT is the determination of width of the window function.

The Continuous wavelet transform (CWT) was produced as an elective way to deal with the determination issue. It is defined as the whole finished sum resulted from the signal duplicated by scaling, moving adaptations wavelet function. There are numerous wavelet coefficients that are functions of dimension and location. Increasing every coefficient by a suitably scaled and moved wavelet results in the component wavelets of the original signal. A CWT gives properly itemized portrayal signal as far as both time and recurrence [2].

The aftereffect of the Fourier transform is sinusoidal parts of the original signal as shown in Figure 1.

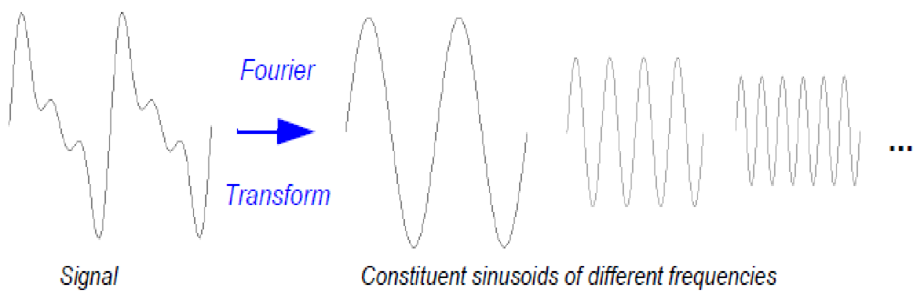

Figure 1. Fourier transform [2].

Similarly, the continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function ψ as can be seen in figure 2.

We can consider the endless CWT as the ground total of the signal increased by scaled. It shifts variants of the wavelet function ψ as the equation shows.

$$C(scale, position) = \int_{-\infty}^{\infty} f(t)\psi(scale, position, t)dt \tag{3}$$

The product of the CWT are lots of wavelet coefficients (C) as output of position and scale. Increasing these wavelet coefficients by the properly shifted and scaled wavelets gives the original signal constituent wavelets.

CWT, as an outcome, is numerous wavelet coefficients (C) as an element of scale and position. Increasing these wavelet coefficients by the fittingly scaled and moved wavelets results in the component wavelets of the original signal.

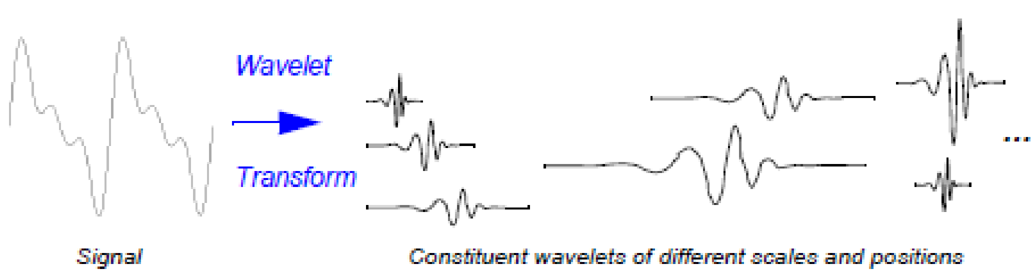

Figure 2. Wavelet transform [2].

When we scale a wavelet, it is stretched or compressed. Similarly, shifting a wavelet means delaying it [2].

Scaling a wavelet implies extending or packing it. Also, moving a wavelet implies deferring it. The constant wavelet transform (CWT) of an f(t) function is represented by the equation(4):

$$CWT_{\psi}f(a,b) = W_f(b,a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{4}$$

Where, (a) is the scaling parameter corresponding to frequency information and (b) is the translation parameter corresponding to the time interval.

CWTs are particularly helpful in tackling problems involving signal identification and detection of hidden transients (hard to detect, short-lived elements of a signal).

CWTs are especially useful in handling issues including signal distinguishing proof and recognition of hidden transients which are hard to identify, fleeting components of a signal [3].

2.3 The Discrete wavelet transform theory

There are so many useful properties in discrete wavelet transform which are useful in the time arrangement and advanced information examination. To change over a period arrangement signal, discrete wavelet transform (DWT) utilizes an arrangement of premise work originated from the wavelets. The objective is generally to lessen the commotion. In DWT, the time arrangement signal was first applied on an arrangement of a scientific capacity of wavelets to separate the signal into various constituents. At that point, discrete wavelet transform isolates these constituents into an alternate recurrence at different scales. It is important to ensure the detailed reproduced type of the first signal x (t) in light of how to test the scaling coefficients (a). In any case, there are diverse level of decay in view of various techniques for wavelets [16].

Despite the fact that the discredited continuous wavelet transform empowers the calculation of the continuous wavelet transform by computers. It isn't an advisable genuine method for discrete wavelet transform. The discrete wavelet transform (DWT), then again, gives adequate data both to examination and union of the original signal, with a noteworthy decrease in the calculation time. [10]Dissimilar to the ceaseless wavelet transform, which can work on each scale, the discrete wavelet transform (DWT) picks a subset of scales and positions to ascertain. [6]

Discrete Wavelet Transform is an intense instrument to use in an extensive variety of uses. Wavelet performs multi resolution analysis of a signal with localization in both time and frequency. DWT produces non-excess outcomes. [12].The values of f (a, b) a and b are calculated over a discrete grid [5],

$$a = 2^{-j}, b = k.2^{-j}j, k \in Z \tag{5}$$

m,n ∈Z, and Z is the set of positive integers discrete wavelet transform 0f the function f(t) is：

$$DWT_{\psi}f(m,n) = \int_{-\infty}^{\infty} f(t)\psi_{m,n}^{*}(t)\, dt \tag{6}$$

where

$$\psi_{m,n}(t) = 2^{-m}\psi(2^{m}\, t - n) \tag{7}$$

DWT utilizes multi determination channel banks and wavelet channels to break down and reproduce the original speech signal. It gives adequate data and diminishes calculation time for analysis and synthesis [4].

The DWT first uses a down sampling and afterward decays the signal using an arrangement of low and high pass filters. The yields are estimation and detail coefficients. DWT is basically a sampled version of CWT [5].

2.4 Wavelet Family

There are distinctive wavelet families with various scales like Haar, symlet (sym), Coiflet (coif), Daubechies (db), and so on to dissect and incorporate a signal. The choice of wavelet decides the last waveform form. The form (type) of the wavelet can likewise portray the acoustic characteristics and an ideal sort for a given setting can be discovered experimentally [4].

The Wavelet types can be explained by orthogonality, symmetry, vanishing moment condition, and compact support help. Wavelets are orthogonal if the inner products of a mother wavelet and it is shifted and scaled version of wavelet are zero, which empowers decomposing a signal into non-overlapping sub-frequency bands. Wavelets families are used to simulate some information, for example, frequency localization, linear phase characteristic, and introduction of discourse signals. The properties of wavelet families, such as orthogonality, symmetry, vanishing moment states of a wavelet figures out which wavelet families better depict a speech signal [5].

Diverse wavelet scales relate to various time-recurrence resolutions. The bigger wavelet scales implies smaller wavelet width and give better time determination. Essentially littler wavelet scales implies more extensive wavelet width and give better frequency resolution.

2.4.1 Haar

Haar wavelet is the first and least difficult type of wavelet group. Haar wavelet is intermittent, and looks like a step function. It speaks to an indistinguishable wavelet from Daubechies (db1) [2].

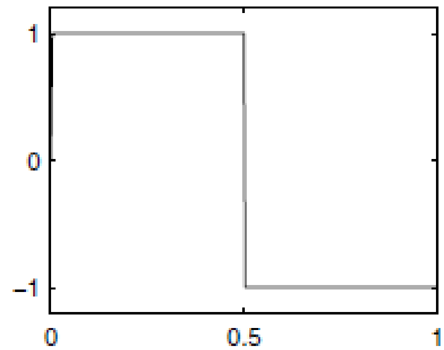

Figure 3. Haar wavelet [2]

Wavelet form of a Haar wavelet is made using the rescaled sequence of square shaped functions. Haar mother wavelet (t) and scaling function $\emptyset$ (t) are described as follows;

Wavelet type of a Haar wavelet is made by using the rescaled grouping of square formed functions. Haar mother wavelet (t) and scaling function $\emptyset$ (t) are explained as follow:

$$\varphi(t) = \begin{cases} 1 & 0 \leq t \leq \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & elesewhere \end{cases} , \ \phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & otherwise \end{cases} \tag{8}$$

2.4.2 Daubechies

Ingrid Daubechies (I.Daubechies,1988) invented what are known as minimalistically upheld orthonormal wavelets. Her strategy made discrete wavelet investigation practicable. The Daubechies wavelet transforms are characterized similarly as the Haar wavelet transform. In Daubechies technique the running midpoints and differences are figured by means of scalar items with scaling signals and wavelets. The distinction amongst Haar and Daubechies method is in the meaning of scaling signals and wavelets [2]. The Daubechies wavelet is more confounded than the Haar wavelet. Moreover, Daubechies wavelets are consistent and hence they are all the more computationally costly to utilize when contrasted with the Haar wavelet [7].

The Daubechies family wavelets' names are composed as 'dbN', where N is the order, and 'db' is the "surname" of the wavelet. The 'db1' wavelet is the same as Haar wavelet.

Figure stands for the wavelet functions psi of the following nine members of the group [2].

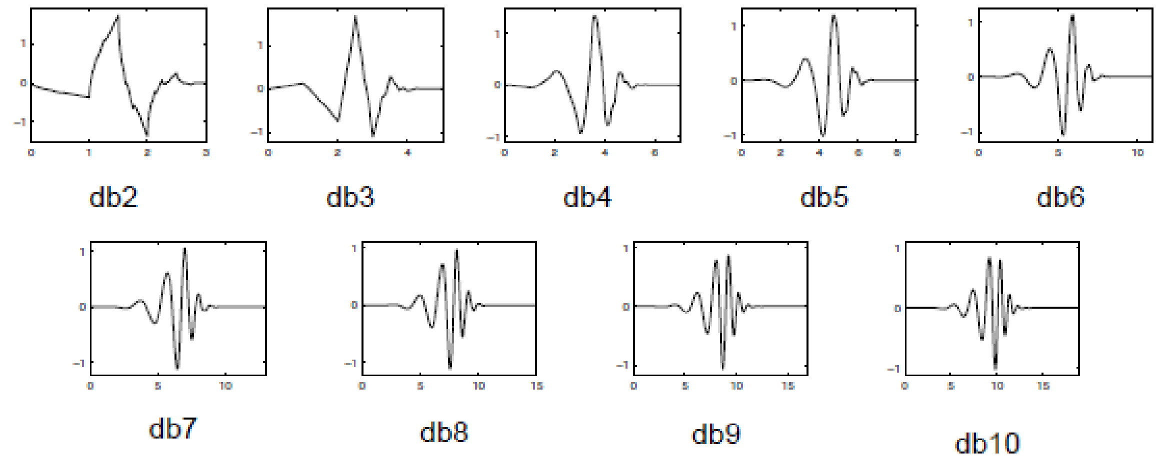

Figure 4. Daubechies wavelet with different scales [2]

Daubechies wavelets are a group of orthogonal wavelets characterizing a discrete wavelet transform. They are described by a maximal number of vanishing minutes for some given help. With every wavelet kind of Daubechies family, a scaling function produces an orthogonal multi-determination analysis. The Daubechies wavelets have the highest number 'A' of vanishing moments for given help width N=2A. Daubechies wavelets are generally used as a part of taking care of a wide scope of so many problems, for example, self-closeness properties of a signal or fractal problems, signal discontinuities, and so forth [8].

2.4.3 Coiflets

At the demand of Ronald Coifman (RC), Coiflets are discrete wavelets outlined by Ingrid Daubechies to have scaling capacities with vanishing moments worked by I. Daubechies in demand of RC. The wavelet work has 2N moments equivalent to 0 and the scaling function owns 2N-1 instant equivalent to 0. These functions possess a help of length 6N-1 [2].

As a rule, it has N number of fading movements for both scaling functions and wavelet. There are different levels in Coiflet utilizing that the signal analysis is led. For example, coif1, coif 2, coif3, coif 4, coif5 etc , this graphical portrayal can be found in the following picture [16].

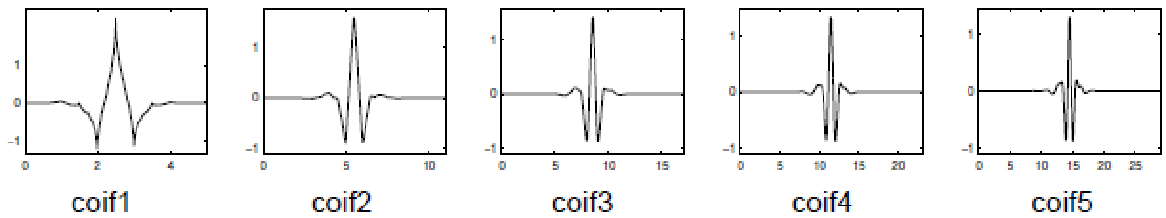

Figure 5. Coiflets wavelet with different scale [2]

2.4.4 Symlets

The symlets are about symmetrical wavelets suggested by Daubechies as changes to the db group. This two wavelet families have the same characteristic [2].

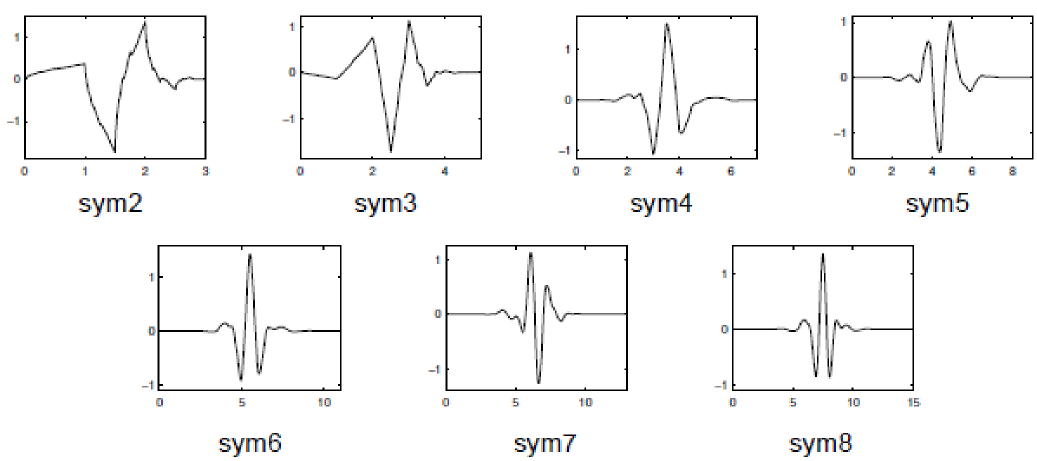

Figure 6. Symlets wavelet with different scale [2]

The above wavelet families have been used in this thesis.

Other wavelet families

2.4.5 Biorthogonal

Biorthogonal wavelet show characteristics of linear phase, that is required to reconstruction for image and signal. There are two fascinating properties which are determined by using two wavelets: one for decomposition and the other for reconstruction instead of the same single one.

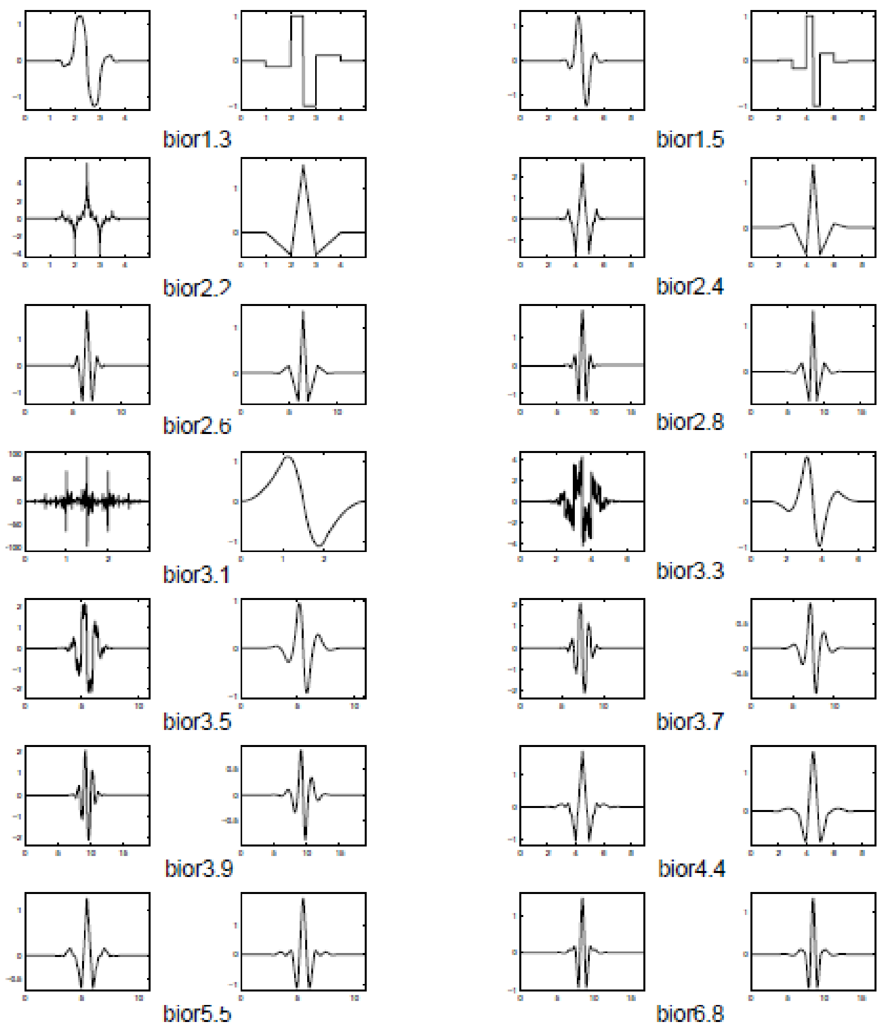

Figure 7. Biorthogonal wavelet with different scale [2]

2.4.6 Morlet

This wavelet has no scaling function, yet it is express precisely.

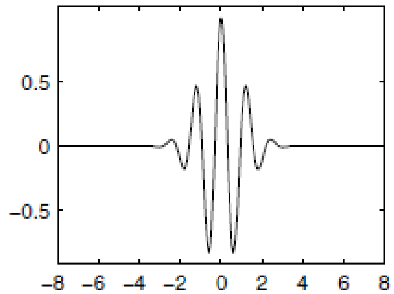

Figure 8. Morlet wavelet [2]

2.4.7 Mexican Hat

This type of wavelet does not contain scaling function. Moreover, it is derivative from a function that is corresponding to the second derivative function of the Gaussian probability density function.

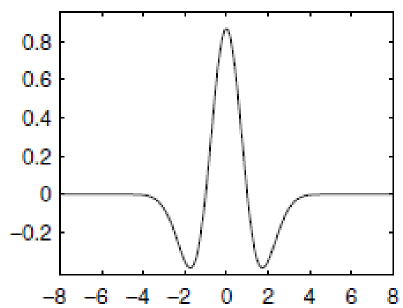

Figure 9. Mexican hat wavelet [2]

2.4.8 Meyer

The Meyer wavelet and scaling functions are expressed in the frequency domain.

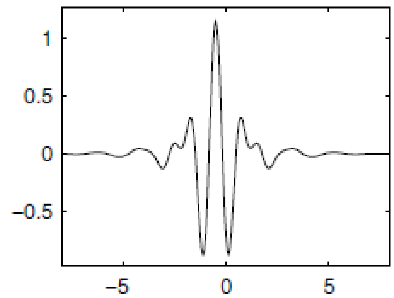

Figure 10. Meyer wavelet [2]

2.5 Thresholding Methods

Wavelet thresholding is the signal estimation system that endeavors the capacities of signal de-noising. Execution of thresholding is absolutely relies upon the sort of thresholding technique and thresholding principle used for the given application. Apply thresholding to the nitty detailed coefficients instead of to the estimation coefficients, in light of the fact that the definite coefficients contain vital parts of the signal. Therefore, the assessed wavelet coefficients are found [4].

Donoho and Johnston developed a universal thresholding rule which can effectively remove the Gaussian random noise. As wavelet examination has its premise copying the front-end auditory periphery (Mallat 1998), endeavors have been made to exploit this signal-processing apparatus for speech upgrade. The most utilized approach depends on the non-linear thresholding of the wavelet coefficients (Donoho 1995), which connects the multi-resolution analysis and non-linear filtering [6].

The approach gains by the way that a suitable transform, which is wavelet transform, extends the signal onto the transformed domain where the signal energy is amassed in a few coefficients while the noise is uniformly disseminated over the transformed domain. There are generally two ways of thresholding: one is known as hard thresholding, and the other is called soft thresholding.

2.5.1 Hard thresholding

In Hard Thresholding, all Wavelets detail coefficients whose total esteems are not as much as the threshold are set to be zero and different wavelets detail coefficients are kept [4].

It is expressed mathematically as,

$$Y = T(X, Y) = X; for\ |X| \quad > \lambda \tag{10}$$
$$O; FOR\ |X| \ \leq\ \lambda$$

The hard thresholding method evacuates the noise by thresholding just the wavelet coefficients of the detailed sub groups, while keeping the low- resolution coefficients unaltered. [7]

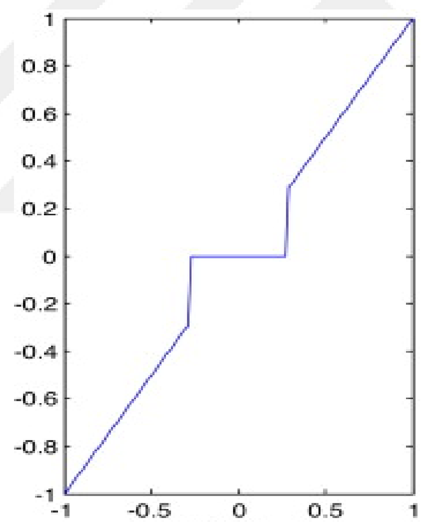

Figure 11. Hard threshold [7]

2.5.2 Soft thresholding

Soft thresholding is an extended form of hard thresholidng. It sets all wavelets detail coefficients to zero whose total values are not as much as the threshold same as hard thresholding and shrinks the non-zero coefficients towards zero. It is characterized as, [4]

$$Y = T(X, Y) = sign\{X\}(|X| - 1); for|X| > \ \lambda \tag{11}$$

$$= 0 \qquad ; for\ |X| \leq \lambda$$

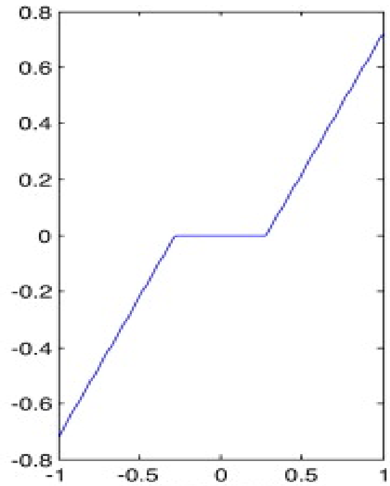

Figure 12. Soft threshold [7]

2.5.3 Threshold level estimation methods

There four main types of soft Thresholding for the threshold selection methods of wavelet threshold denoising method.

2.5.3.1.Sqtwolog method

It is possible to obtain threshold by increasing threshold σ of least value in greatest fluctuation and a coefficient. This threshold form gets a naturally good denoising impact in the process of soft-threshold. [9]

Such a method uses square root of logarithm to assess the threshold rates (Ts)

$$T_s = \sigma_j \sqrt{2\log(N)} \tag{12}$$

where

T= the threshold value

N = the length of the coefficients of noisy signal σ=MAD/0.6745(standard deviation)

With MAD represents median absolute deviation of the wavelet coefficients.

2.5.3.2.Rigrsure method

The Rigrsure, which is known as Steins unbiased risk estimator (SURE) is an adaptive thresholding method which was investigated by Donoho and Jonstone [9] and it depends on Stein's impartial probability estimation guideline. This technique computes probability estimation first by using the given threshold t, and after that minimizing the non-like hood. Then after, the threshold is acquired. [9]

The diverse thresholds are picked in various scales, and the wavelet coefficients of comparing scales are downsized.

2.5.3.3. Heursure method

It is the mix of sqtwolog form and rigrsure form. The chosen threshold is the best predicable variable threshold. On the off chance that the SNR is little, the form sqtwolog is use to choose the threshold; if SNR is large, rigrsure form will be utilized.[16]

Threshold equation (Th) can be represented mathematically as:-

$$T_h = \begin{cases} T_s & a > b \\ \min\left(T_{s,}\, T_r\right) & a \geq b \end{cases} \tag{13}$$

Where

(Ts) = the threshold achieved from sqtwolog

Tr = threshold achieved from Rigrsure.

a = s−n / N b =(log2N)$^{3/2}$ √N

s = squared wavelet coefficients

2.5.3.4. Minimaxi method

The Minimaxi selection discovers threshold by using Minimax principle. It uses a settled threshold to yield Minimax performance for mean square error against a perfect strategy. The Minimax rule is used as a part of statistics to plan estimators. Since the de-noised signal can be absorbed to unknown regression function, the Minimax estimator is the choice that realizes the minimum, over a given arrangement of functions of the greatest Mean Square Error (MSE). This strategy finds ideal thresholds [12]

In this strategy, the threshold value will be chosen by acquiring a minimum error between wavelet coefficient of noise signal and original signal.

The minimax provides maximum mean square error (MSE) developed by using Minimax rule. In the procedure of configuration evaluation, the minimax will be used to understand the minimum value of the threshold. The threshold value (Tm) is given

$$T_m = \begin{cases} \sigma_j (0.3936 + 0.1829 \log_2 N & for\ N > 32 \\ 0 & for\ N < 32 \end{cases} \tag{14}$$

Where σj=median (j0.6745) and N represents a dimension of the signal. Correspondingly, j signifies coefficient vector at scale equals to j. [16]

CHAPTER III

WAVELET ANALYSIS OF SIGNAL

In many signals, the low-frequency content of signals, are the most important one. On the other hand, the high-frequency content, gives flavor or nunance. Let’s assume a human voice. In the event that you remove the high-frequency segments, the voice sounds will different however you can still hear what's being said. On the other side, if you remove enough of the low-frequency parts, you hear nonsense sounds. In wavelet examination, we frequently talk about approximations and detail elements. The approximations are the high-scale and low-frequency segments of the signal the details are the low-scale and high- frequency components. [2]

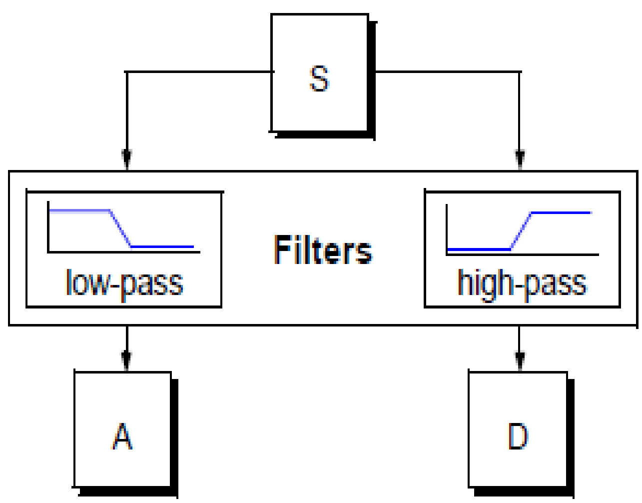

Figure 13. Approximation and detail coefficient [2]

3.1 Discrete Wavelet Decomposition

We have figured out how the discrete wavelet transform is utilized to be break down or decay signals. Such type of process is known as decomposition or analysis.

The primary phase of the DWT algorithm breaks down the signal into two sets of coefficients. They are the estimation coefficients cA (low recurrence data) and the detail coefficients cD (high recurrence data). The coefficient vectors are gotten by convolving (s) with the low-pass filter for estimation and with the high-pass filter for details.

Since the LPF and HPF are accomplished by taking the average and difference respectively, the yields are known as approximation coefficients when taking the average and detail coefficients by taking the difference. [5]

There is still another more natural approach to depict why the coefficients are known as the estimate and detail coefficients. The low-frequency signal has a more extended period than the high-frequency signal. This single represents a close approximation of the original signal. But the high frequency signal stands for defining detailed information which is related to the original signal.

The technique which is known as down- sampling is connected after each filtering so as to keep the output of a discrete WT an indistinguishable size from the original signal. For example, the first signal S contains 1000 samples of information. At that point where subsequent signals can contain thousands of samples each sum of two thousands. These signals (A and D) are very interesting, yet it is possible to get two thousand values rather than one thousand. It is possible again to have another unobtrusive approach so as to play out the putrefaction by operating wavelets. By taking care of calculation, it is possible to keep 1 point per 2 in each one of the double 2000-extent samples so as to have total data. The process is considered as down-sampling thought. It is possible to create 2 arrangements. They are approximation coefficient and detail coefficient. The procedure which incorporates down- sampling produces DWT coefficients. [2]

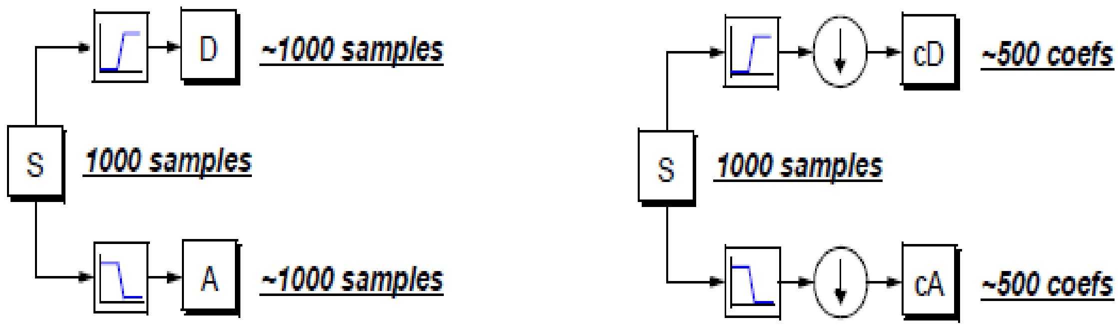

Figure 14. Decomposition of wavelet [2]

The decomposition procedure is repeated in progressive approximations being deteriorated thusly, as a result 1 signal will be separated into numerous lower resolution components. This is known as the wavelet decomposition tree.

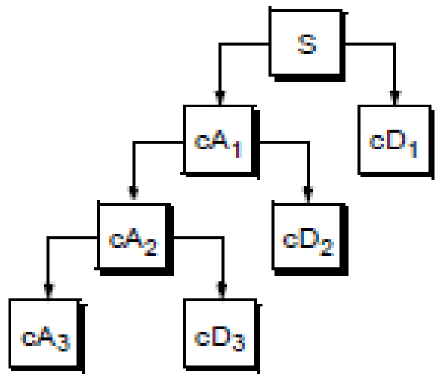

Figure 15. Wavelet decomposition tree

The above Figure shows the process of performing and the numerous deterioration. The tree is known as wavelet disintegration tree.

Levels' number: because of investigation procedure is repetitive, in principle it is proceeded with uncertainly. Realistically, the disintegration can continue just until the point that the individual subtle details comprise of a solitary sample or pixel. In addition, the processes incorporate choosing a reasonable number of levels in view of the nature of the signal, or on an appropriate property, like entropy. [8]

3.2 Discrete wavelet reconstruction

It is possible figured out how the separate wavelet transform is utilized for the analysis or signals break down purpose. This procedure is known as analysis. Another portion of the description is the way segments are gathered once again into the basic signal with no loss of any necessary data. The procedure can be called as reconstruction to mean recreation. The operational manipulation which impacts synthesis is known as inverse discrete wavelet transforms (IDWT).

By using the IDWT, it is possible to reconstruct or synthesized the original signal. The synthesis begins approximation and detail coefficients. These coefficients are reproduced by up-sampling and shifting with low pass and high-pass filters [2].

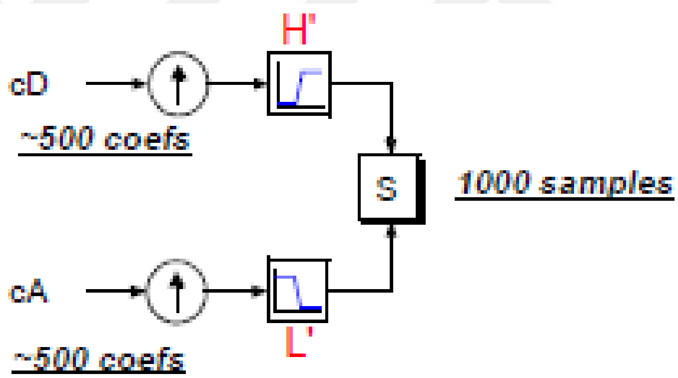

Figure 16. Reconstruction of wavelet [2]

The filtering part of the reconstruction process also bears some discussion, because it is the choice of filters that is crucial in achieving perfect reconstruction of the original signal.

Since the reconstruction process filtering part is the choice of filters, it bears some discussion. It is crucial in achieving perfect reconstruction of the original signal.

The down-sampling of the signal parts are applied amid the disintegration phase to introduces a distortion process known as aliasing. Then it turns out those precisely picked filters for the decomposition and reconstruction phases that are firmly related.

The low- pass decomposition and high-pass decomposition filters (L and H), with related recreation filters of them (L' and H')

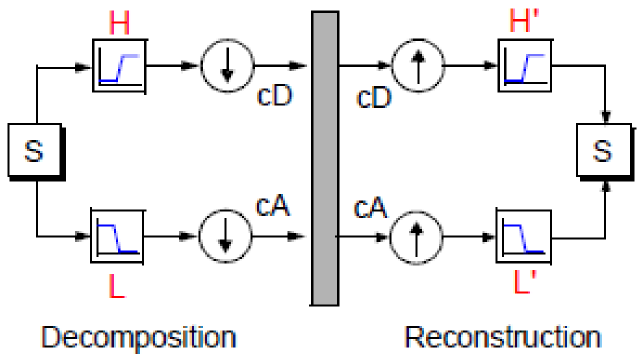

Figure 17. Decomposition and Reconstruction of wavelet [2]

As can be seen from figure 17 that the original signal can be rebuilded from the coefficients of the guessing and descriptions.

3.3 The Maximal Overlap Discrete Wavelet Transform

The maximal overlap discrete wavelet transform (MODWT) is a linear filtering operation that transforms a series into coefficients identified with varieties over an arrangement of scales. It is like the DWT in that both are linear filtering operations delivering an arrangement of time-dependent wavelet and scaling coefficients. [11]

The MODWT holds down-sampled values at each level of the decomposition that would be generally disposed of by the DWT. MODWT used to all the while assess the scale-dependent signal behaviors.

MODWT is an undecimated alteration to DWT which is a time shift invariant. In other words, it is an interpretation in the signal will bring about an interpretation of wavelet coefficients by a similar sum. In this way, MODWT adjusts wavelet coefficients at each time point with the original data index. It is possible to analyze localized signal variation with respect to scale and time and the fleeting connection to occasions. This is especially valuable as a disconnected analysis which can roughly evaluate the occasion onsets relating to the occasion records. Thus, it can appropriate a fitting occasion window size for local classification. [24]

The Maximum Overlap Discrete Wavelet Transform (MODWT) was used to assess the scale-subordinate signal behaviors simultaneously. MODWT is an undecimated alteration to DWT which is in this manner time shift invariant. That is, an interpretation in the signal will bring about an interpretation of wavelet coefficients by a same amount. In this way, MODWT adjusts wavelet coefficients at each time point with the first information list. So, it is possible to analyze and localized signal variation as for scale and time, and temporal relation to events. [17].

Let's consider time domain is the original domain of the MODWT and discrete wavelet transform (DWT) case. In spite of the fact that, these models of transformation process from time domain to time domain, these procedures are known as signal decomposition on the grounds that a given signal is decomposed into a lot of different signals with deferent levels of determination. These procedures permit recuperating the original time domain signal without losing any data. MODWT has turn around process which is known as the converse MODWT of signal reconstruction. The MODWT and DWT are executed by using a multi resolution pyramidal deterioration strategy. A recorded digitized time signal S can be broken down into detailed cD and smoothed (approximations) cA signals using high-pass and low-pass filter respectively. High pass filter has a band-pass reaction. Thus, the filter signal cD is a detailed coefficient of S and consists of higher frequency components. The approximation signal cA has a low-pass frequencies filter reaction. The decomposition of S into cA and cD is the principal scale decomposition. It is possible to perform the original signal, inversely, from the approximations and details coefficients. [18]

For the most part, the MODWT were assessed by using expansion conditions as shown below:-

$$\phi(t) = \sqrt{2}\sum_k l_k \phi(2t-k), \qquad \phi(t) = \sqrt{2}\sum_k h_k \phi(2t-k), \tag{15}$$

Father and mother wavelets were expressed by the last two equations where φ (2t − k) represents the father wavelet, and ψ (t) represents the mother wavelet. Father wavelet defines the lower pass filter coefficients (h_k) and high pass filters coefficients (l_k) are defined as:

$$l_k = \sqrt{2}\int_{-\infty}^{\infty} \emptyset\,(t)\emptyset(2t-k)dt,\ h_k = \sqrt{2}\int_{-\infty}^{\infty} \psi(t)\psi(2t-k)dt \quad (16)$$

CHAPTER IV

SPEECH ENHANCEMENT AND SIGNAL DENOISING

4.1 Speech Enhancement Methods of a Noisy Speech Signal

Speech improvement is an extensive research territory in speech signal preparing. The objective of numerous upgrade algorithms is to suppress the noise in a speech signal. Generally, noise can be added substance, multiplicative, or convolution, narrow band or broad band, and stationary or nonstationary. [25] Speech upgrade algorithms have numerous applications in speech signal processing. Signal upgrade can be invaluable to hearing debilitated people on the grounds that the capacity to create clean signals is basic to their cognizance of speech. Enhancement algorithms are likewise used as a part of conjunction with speech recognizers and speech coders as front end processing.

Enhancing the noisy speech signal before running the signal through a recognizer can increase the recognition rate. As a result it creates a more vigorous recognizer. Likewise, front end upgrading to speech coding is appeared to diminish the quantity of bits important to code the signal. [26] Speech enhancement progresses the nature of speech signal by using different algorithms. The primary target of upgrade is improvement in understandability and the general perceptual nature of corrupted speech signal using audio signal processing methods. Improving of speech signal which is corrupted by noise, or noise diminishment, is the most vital field of speech upgrade. The algorithms of speech upgrade for noise diminishment can be arranged into three classes. They are spectral restoration, filtering techniques and model-based methods [25].

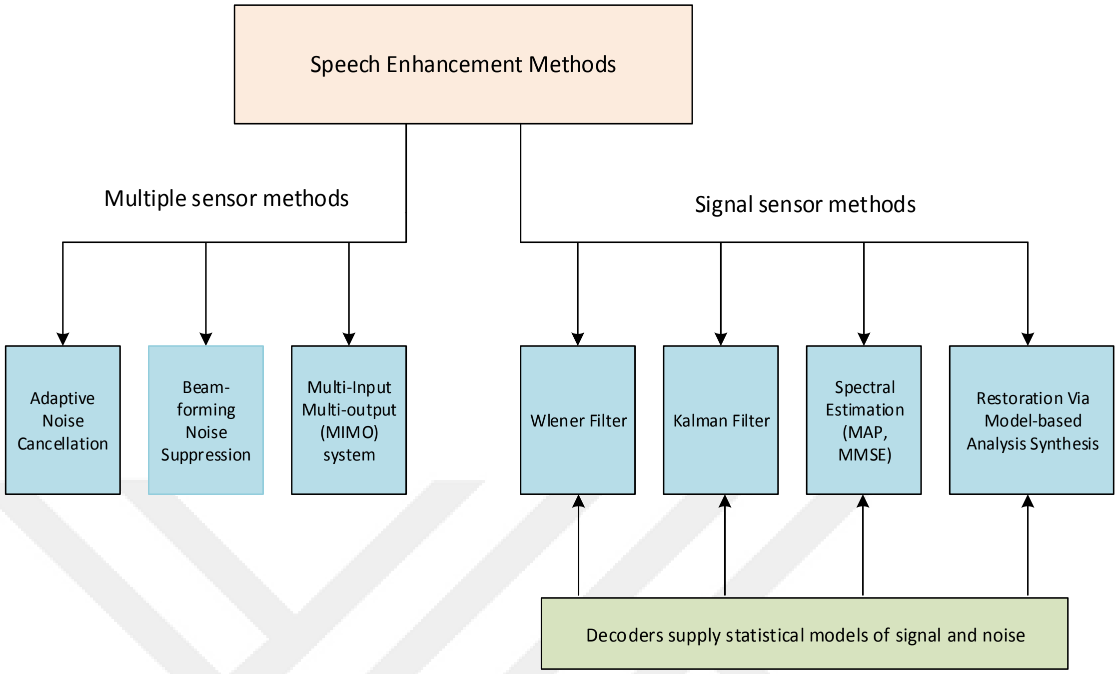

Figure 18. Speech enhancement methods [25]

Broadly used algorithm in speech improvement explore is the Wiener filter. In the event that both the signal and the noise estimates are precisely valid, this algorithm will yield the ideal gauge of the clean signal. Through limiting the mean squared error between the evaluated and clean speech signals.

The Wiener filter is produced and given by equation bellow:

$$H(\omega) = \left[\!\left[\frac{|\hat{s}(\omega)|^2}{|\hat{s}(\omega)|^2 + |N(\omega)|^2}\right]\!\right] \tag{17}$$

Where H is the Wiener filter, and S and N are the noise corrupted speech and noise spectra, respectively.

The Kalman channel is a Bayesian filter in that it utilizes the earlier probability appropriations of the signal and noise processes. The signal is thought to be a zero mean Gaussian-Markov process. The noise is thought to be zero-mean independent identically distributed (IID) Gaussian process. The filter is likewise expected that the parameters of the models of signal and noise. Filter mutilation are known from the earlier [27].

When we consider MMSE estimation of the short time spectral amplitude (STSA), its structure is similar to that of otherworldly subtraction. However as opposed to the Wiener filtering inspiration of phantom subtraction. It advances the gauge of the genuine instead of complex spectral amplitudes. Fundamental to their methods is the estimate of SNR in every frequency receptacle for which they proposed two algorithms. The first one is the maximum probability approach and the other is a decision directed approach. They are performed for better performance. The Reclamation Model-based speech enhancement utilizes earlier knowledge as an unequivocal stochastic model of speech as well as interfering noise.

Various distinctive speech models are accessible including some blend of autoregressive (AR) models, coefficient models, Hidden Markov models and pitch track models. Enhancement techniques in view of an AR model of speech for the most part put no requirement other than solidness on the evaluated set of AR coefficients [25].

The Speech is sectioned into overlapping frames of N tests. It is then changed to the frequency domain by means of discrete Fourier transform (DFT). The DFT can be defined as a particular sort of discrete transform. It is used as a part of Fourier investigation. It transforms from one function to another. This is known as the frequency domain representation or just the DFT of the first function which is the time domain [27].

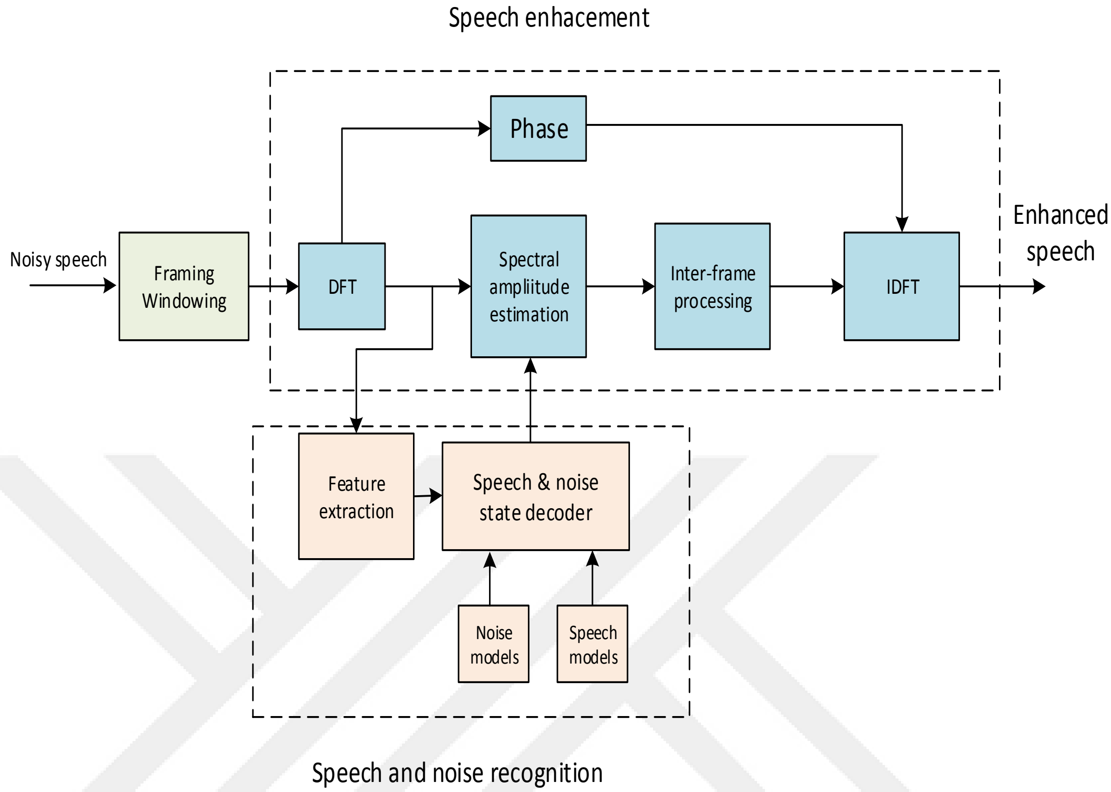

Figure 19. Block diagram of single input speech enhancement system [27]

Johnstone and Donoho were the first to proposed the wavelet based speech signal improvement method. This method depends on thresholding, the wavelet coefficients of l noisy speech signal. The major thought behind wavelets are to break down as according to scale. The wavelet examination technique is used to adopt a wavelet model function. This function is known as analyzing wavelet or mother wavelet. [4]

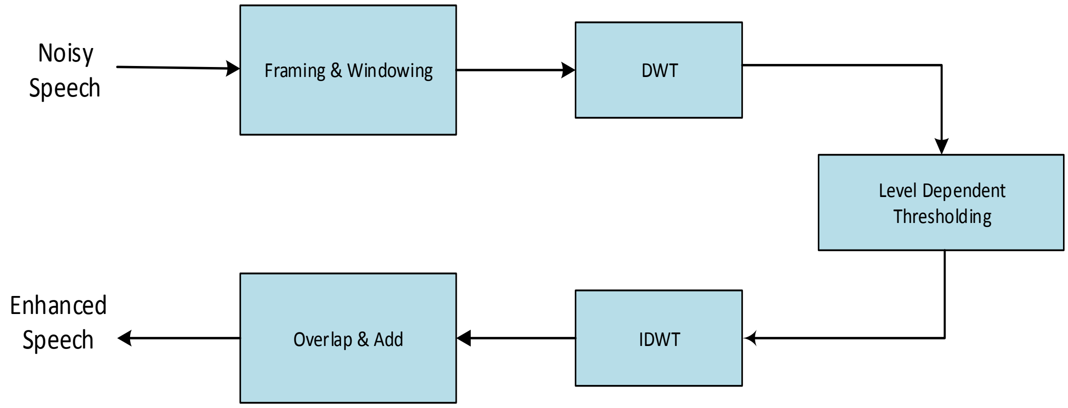

Figure 20. Block diagram of wavelet de-nosing of speech signal [4]

4.2 Noise Signal in a Speech

There are numerous types of noise. A standout amongst the most widely recognized wellsprings of noise is background noise which is constantly present at any area. The other type noise known as channel. This channel noise influences both analog and digital transmission. The other one is known as quantization noise. Such type of noise comes about because of over pressure of speech signals. We can mention again noises like multi talker babble, reverberation noise or delayed version of noise which are found in a few circumstances. The additive background noise is irregular in nature and furthermore uncorrelated with speech. It is found in different environment scenarios like workplaces, autos, city streets fans, factories, planes and so forth. The additive background noise assumed that it is made of three developing enhancement strategies. These strategies are

1) Speech and clamor signals are uncorrelated in any event finished a brief timeframe premise.

2) Noise is stationary or gradually fluctuating more than a few casing of discourse and

3) Noise can be spoke to as zero mean irregular procedures. Regarding to reverberation, reflections of speech from various objects will be mixed with the speech in a convoluted fashion [15].

We can consider noise as an undesirable signal that ends the estimation of the original message. Additionally, the noise will contain some wellspring of undesirable information relying upon the environment around it.

Actually, when we look at the nature of noise, we can get numerous sorts of noise.

There are numerous kinds and sources of noise or distortions in a signal. By looking at the nature of noise, we can categorize it in to the following types:-

1. Electronic noise:- like thermal noise and shot noise,

2. Acoustic noise:- which is emanated from moving, vibrating or colliding sources like revolving machines, moving vehicles, automobiles, spinning engine, keyboard clicks, wind and rain. Indeed these noises might come from striking sources or vibrating,

3. Electromagnetic noise that can interfere with the transmission and reception of voice, image and data over the radio-frequency spectrum. It can occur over radio frequency spectrum during transmission and reception of speech, Damage of data packets due to network blocking are caused because of the quantization noise.

4. Electrostatic noise generated by the presence of a voltage.

5. Communication channel distortion and fading.

6. Quantization noise and lost data packets due to network congestion.

Moreover, we can also classify noise in to white noise, narrow band noise, color noise, impulsive noise and band limited white noise. Electrostatic noise is one which is generated because of high voltage [16].

Once again noise can be characterized as an undesirable signal which blocks with the quantification of other signal. A noise information-producing signal. It passes on data with respect to the sources of the noise and the area in which it spreads.

4.2.1. White noise

As it is shown in the following figure, white noise is a noise which is an uncorrelated random arbitrary noise process with equal power at all frequencies. The Random noise has a similar power at all frequencies in the range of ∞. This power is fundamentally limitless power. This means such power is considered as a theoretical power or the concept is roughly theoretical concept. However, a band-limited noise process with a level range covering. The frequency scope of a band-limited communication system is practically considered a white noise process [8].

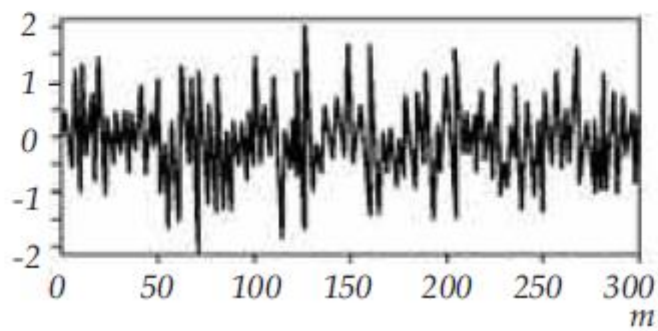

Figure 21. White noise time-domain signal [8]

4.2.2 Additive White Gaussian Noise Model (AWGN)

The noise is assumed as stationary additive white Gaussian (AWGN) process in classical communication theory.

Despite the fact that for a few issues this is a substantial suspicion and leads to mathematically helpful and useful solutions. Practically speaking, the noise is regularly time varying, associated and non-Gaussian. This is especially valid for indiscreet sort noise and for acoustic noise which is non-stationary and non-Gaussian. Additionally, henceforth it cannot be demonstrated utilizing the AWGN assumption.

4.3 Speech denoising with Wavelet Transform

Speech is an extremely fundamental route for people to pass on data from sender to receiver using routers with the emotion of a human voice. Individuals always use speech to convey messages. Such speeches can be displayed as channel following up on excitation waveform. The vocal tract shape makes certain frequencies in the excitation. This is to be amplified and attenuates other frequencies. The excitation appears as semi occasional puffs of air which makes the yield speech to appear periodic. Speech can be separated into voiced and unvoiced speeches. Voiced speech has a range with energy gathered at discrete frequencies. This means, it is concentrated at the key recurrence of the vocal folds and its multiples (harmonics). Around 33% of speech is totally an intermittent (unvoiced) coming about because of an irregular excitation that looks like white noise, via air quickly going through a thin choking in vocal tract. [15]

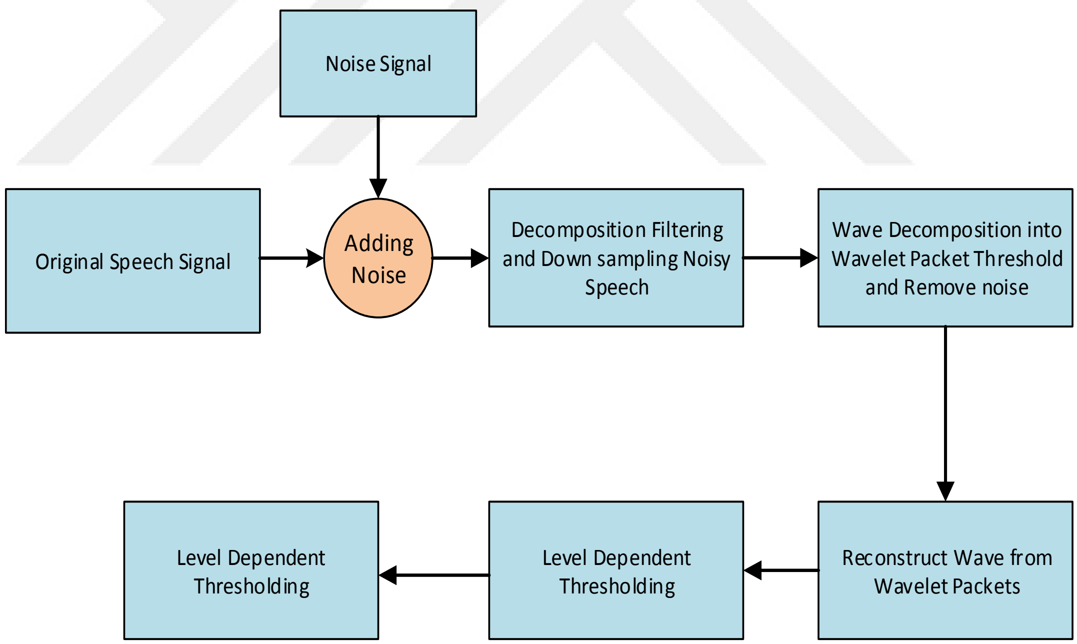

Figure 22. . Block diagram of speech signal denosing with wavelet [15]

As wavelet investigation has its premise imitating the front-end sound-related fringe, endeavors have been made to take advantage this signal-processing instrument for speech enhancement.

The most frequently used approach depends on the non- linear thresholding of the wavelet coefficients, which connects the multi resolution analysis and non- linear filtering. The Thresholding procedure is a denoising procedure. The wavelet transform disintegrates the loud speech signal into two segment coefficients, approximation or low pass coefficients and points of interest or high-pass coefficients. Every of the approximation or the detail parts has half the length of the original speech signal. Speech signal energy is concentrated in the approximation component in this way, the impact of noise on the estimate segment is little and on the detail segment is substantial. On the off chance that a thresholding procedure is performed on the detail. It decreases the noise altogether, leaving the signal energy unaffected [19].

4.4 Wavelet Transform and Fourier Transform Comparison

Fourier transforms (FT) permit the portrayal of the frequency data of the speech signal. Wavelet transforms were motivated by the inadequacies of the Fourier transform.

At the point when the FT represents a signal in the frequency domain, it can't tell where those frequency parts are available in time. Cutting the signal at a specific minute in time and transforming that into the frequency domain to get an applicable time succession of frequency data is identical to convolving the signal and the cutting window, resulting in conceivable spreading of frequency components along the frequency pivot [5].

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

The Wavelet Transform (WT) is essentially needed to analyze non-stationary signals. In other words, it is frequency reaction differs with time. Fourier Transform (FT) isn't appropriate for non-stationary signals.

A wavelet is considered as a waveform of adequately restricted duration. It is average value is always zero.

If we contrast wavelets and signal waves, we can get that wavelets are fundamentals of Fourier analysis. Sinusoids don't contain constrained term — they reach out from negative to positive infinity. If sinusoids are predictable and smooth, wavelets have a tendency of being unpredictable [2].

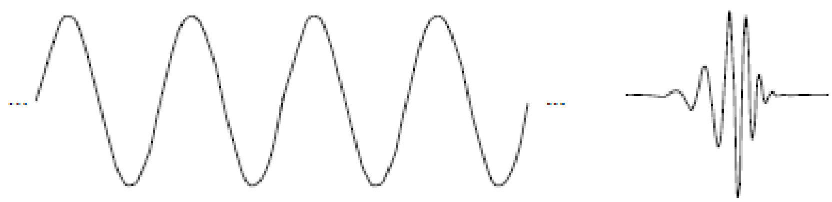

Figure 23. Signal of sin wav and wavelet [2]

Fourier analysis comprises of separating a signal into sine waves of different frequencies. Likewise, wavelet analysis is the separating of a signal into shifted and scaled forms of the original or mother wavelet.

Fourier analysis encompasses a genuine short coming in re-modeling to the frequency domain. When this happens, time information is lost. Once viewing at a Fourier transform of a signal. It's tough to inform once a selected occasion occurred.

.

Be that as it may, most intriguing signals contain various non-stationary or passing characteristics. These characteristics are float, patterns, sudden alters, and starting and finishes of actions. The attributes are the most imperative piece of the signal. The Fourier investigation isn't appropriate to distinguishing them [2].

The wavelet transform has built up a reputation for being an apparatus for signal analysis. It has frequency-resolution determination (low time- resolution) for the low frequency content of the signal while having low frequency-resolution (and high time-resolution) for the high frequency content of the signal [6].

At their disposal, Signal analysts already got a formidable arsenal of tools. Perhaps the foremost sure as shooting understood of those is Fourier analysis, that separates a signal into constituent sinusoids of assorted frequencies. Another approach to think about Fourier analysis is as a mathematical method for transforming our perspective of the signal from time-based to frequency-based.

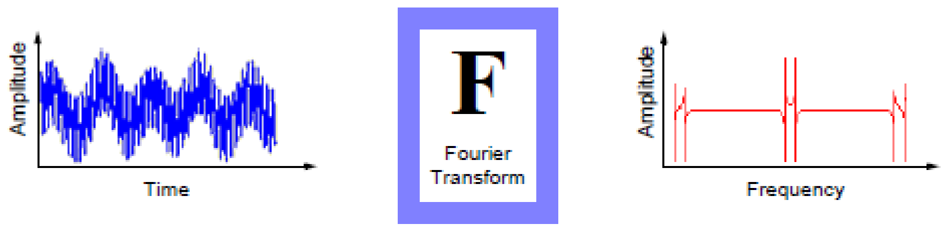

Figure 24. Fourier Transform [2]

Fourier analysis very important for many signals because the signal's frequency content is of awesome significance. So we require other systems, like wavelet analysis.

In the Fourier analysis there is a genuine disadvantage. In transforming to the frequency domain, time data is lost. When taking a gander at a Fourier transform of a signal, it is impossible to tell when a specific occasion occurred. In the event that the signal characteristics don't change considerably after some time. If it is a stationary signal, this drawback isn't essential. Nonetheless, most intriguing signals contain various non-stationary or fleeting attributes. These characteristics are drift, trends, abrupt changes, and beginnings and ends of events. These attributes are frequently the most vital part of the signal. And Fourier analysis isn't suited to distinguishing them.

With an end goal to revise this lack, Dennis Gabor (1946) adjusted the Fourier transform to analyze just a small segment of the flag at once — a technique called windowing the signal. Gabor's adaptation, which is known as the Short-Time Fourier Transform (STFT), maps a signal into a two-dimensional function of time and frequency.

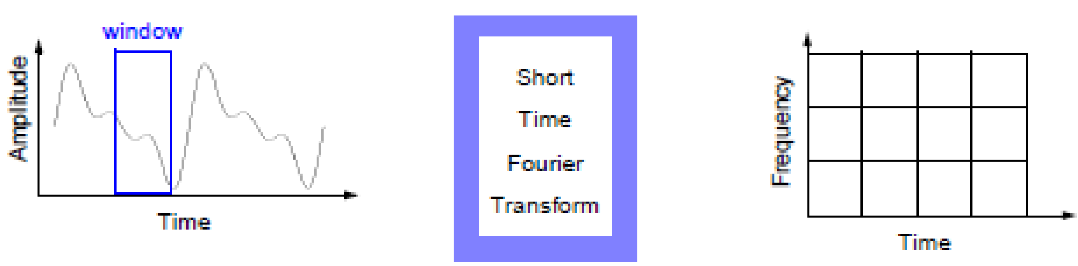

Figure 25. Short time fourier transform [2]

The STFT represents a sort of compromise between the time-and frequency based perspectives of a signal. It gives some data about both when and at what frequencies a signal occasion happens. Nonetheless, it is possible to acquire this data with restricted accuracy. The precision is dictated by the extent of the window. While the STFT trade off amongst time and frequency data can be helpful. The downside is that once we can pick a specific size for the time window, that window is the same for all frequencies. Numerous signals need a more adaptable approach. We can vary the window size to decide all the more precisely either time or frequency.

The next logical step is represented by Wavelet analysis. A windowing technique with variable-sized regions. Wavelet analysis permits the use of long time interims where we need more exact low- frequency data and shorter districts where we need high-frequency data.

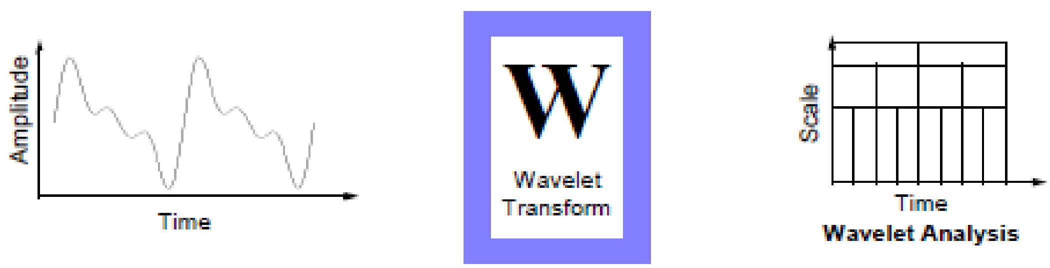

Figure 26. Wavelet transform [2]

It is possible to notice that wavelet analysis does not use a time-frequency region, but instead a time-scale region. For more data about the idea of scale and the connection amongst scale and frequency [2].

Whereas, wavelet transform provides both real and complex values as output .Fourier spectrum analysis is the leading systematic instrument for frequency domain analysis.

But, Fourier transform provides very poor form of information with spectrum variation with respect to time. In contrast, wavelet transform provides the best time-frequency information with respect to time [16].

The Fourier transform has the feature of producing an information structure equivalent to log2 n segments similar to discrete wavelet transform. In addition, both premise functions of discrete wavelet transform and Fourier transform are confined in frequency, which indeed makes some mathematical tools to take part in spectrum analysis. Likewise, both wavelet transform and Fourier transform are represented using integral function. However, Fourier customs a relationship process with exponential function (e−it) and wavelet transform customs a connection process with the transformation of any analysing wavelet φ.

Also, both Fourier transform and wavelet transform can devour genuine and complex valued function. However, the output of Fourier transform limits points to complex as it were. While, wavelet transform gives both genuine and complex values as yield. Fourier spectrum investigation is the main methodical instrument for frequency domain analysis. In any case, Fourier transform furnishes exceptionally poor type of data with information variety regarding to time. On the contrary, wavelet transform provides the best time-frequency information with respect to time [16].

Besides, neither Fourier transform nor short time Fourier transforms has capacity to manage the non-stationary signals. Then again, wavelet transform has adequate ability to perform with the non- periodic audio signals with various transient. Fourier transform is restricted to just single function and it scales a single function. On the other hand, it is conceivable to shift and scale the capacity at the same time using wavelet transform.

CHAPTER V

EXPERIMANTAL WORK AND RESULTS

5.1 Experimental Work

In this thesis study, we applied wavelet method for denoising of speech signals. For the noisy input speech signal, a noisy speech corpus (NOIZEUS) [30] is used. In this thesis study, a wavelet transform based de-noising algorithm was written and applied to the noisy input signal. Hard and soft thresholding techniques were used in two different scales. The used denoising function in Matlab is ‘wden’ which returns the denoised version sd of the input noisy signal s.

sd = wden(s,tptr,sorh,scal,n,wav)

The parameters of wden function is briefed below;

Tptr : threshold selection rule.

Sorh: specifies the thresholding of details coefficients of the decomposition at level n of signal s by the wavelet called wav (soft or hard thresholding)

Wav : wavelet model

n : level

Scal : threshold's rescaling methods ('one' is basic model, 'sln' is basic model with unscaled noise and performs threshold rescaling using a single estimation of level noise based on the first-level coefficients, ‘mln' is basic model with nonwhite noise and performs threshold rescaling using a level-dependent estimation of the level noise.)

The clean speech signal is used as reference for evaluation of the performance of denoising procedure. The clean speech signal and its spectrogram can be seen in Figure-27 and Figure28.

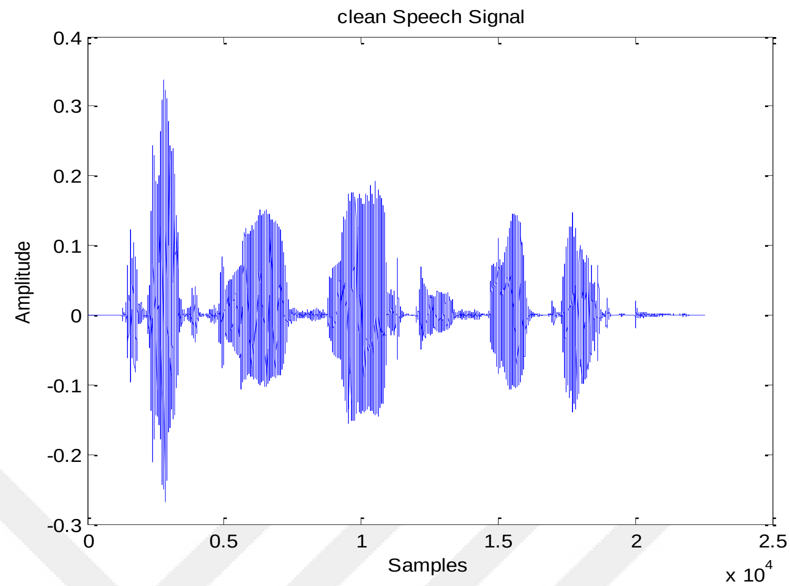

Figure27. Clean speech signal

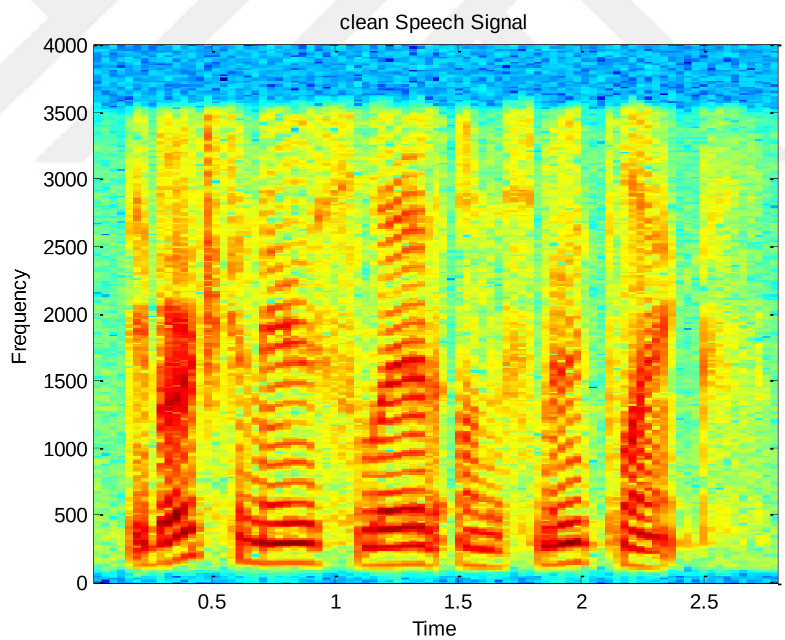

Figure28. Spectrogram of Clean speech signal

5.2 Objective Measure Methods

5.2.1 Signal to noise ratio

The signal-to-noise ratio (SNR) is the most widely used objective measure method to evaluate speech quality. The SNR values are determined by the ratio of square of clean speech to the square of the difference between the clean speech and the enhanced speech. If the summation is performed over the whole signal length, the operation is called as global SNR. The SNR measure formula is usually calculated in terms of decibel (dB) and it is given in equation (18),

$$SNR = 10\log_{10}\frac{\sum_n s^2(n)}{\sum_n[s(n)-\hat{s}(n)]^2} \tag{18}$$

Where s(n) represents the clean speech and $\hat{s}(n)$ is noisy signal.

5.2.2 Mean square error (MSE)

Minimizing mean square error (MSE) is a distance measure between the processed speech and the clean speech and it is commonly used technique in the filtering algorithms. It is computed as in equation (19),

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}\big(s(n) - s(n)\big)^2 \tag{19}$$

5.3. Noisy Speech Corpus

A noisy speech corpus (NOIZEUS) [30] was developed to use in speech enhancement algorithms. It contains 30 speech sentences produced by 3 male and 3 female speakers and corrupted by eight different natural noises at different SNRs including car, train, restaurant, airport, street, babble, train-station and exhibition noises. It can be downloaded by researchers for research studies. These noise types are artificially added to the speech signal. SNR levels of noisy input speech signals are 0dB, 5dB, 10dB, and 15dB. This corpus is described in [29].

5.4. Speech Denoising Using MATLAB

During the experimental work, one-dimensional discrete wavelet transform was used for denoising a speech signal and experimental results were obtained. The software algorithm has been written in MATLAB (Vers. R2016). Various types of mother wavelets in different scenarios were used in MATLAB programming and experimental results are obtained for denoising speech signal. SNRs and MSE measurements under all these scenarios were calculated. During all the scenarios, daubechies5 wavelet was selected at level 5, and this condition was hold constant for all different scenario types. Then, MATLAB program was run for different types of scenarios as can be seen in Table 1. The results also drawn in Figure 29. The noisy signal used in Table-1 includes car noise, train noise, babble noise, street noise, exhibition noise, restaurant noise, station noise and white noise at SNR:5dB.The input noisy speech signal was decomposed by applying four different threshold selection to the wavelet coefficient : sgtwolog, heursure, rigrsure , and minimaxi thresholding.

In Table 1, SsS, SsM, ShS, ShM can be explained as follows. The first letter represents the name of the thresholding method [Sqtwolog (S), Modwtsqtwolog (D), Heirsure (H), Rigrsure (R) and Minimax (M)] ,the second one represents soft(s)/hard(h) threshold, and the last letter represents sln(S)/mln(M) scaling (i.e. RsM means rigrisure-softthresohld-Mln scale, HhS means heirsure-hardthreshol-Sln scale)

Table 1. Results of SNR for 5dB noisy input speech signal

Thresold type	Scenario Type	car	airport	babble	exhibition	restaurant	station	street	train	whitenoise
Sqtwolog (S)	SsS	6.599453	6.599284	6.038119	6.054357	6.569429	6.395572	5.996394	4.625853	3.832965
	SsM	0.827104	0.814066	0.979351	1.815115	1.07446	0.807469	1.497182	1.019166	2.006832
	ShS	5.174613	4.659435	4.654143	5.620194	4.77907	6.06793	5.357301	6.452877	6.550975
	ShM	2.262341	2.121267	2.053034	3.341008	2.517587	2.232352	3.112913	2.650479	4.046997
Modwtsqtwolog (D)	DsM	4.566519	4.609105	4.673867	5.87466	5.061075	4.603288	5.272652	5.030505	6.063724
	DhM	8.446362	8.13396	6.71972	7.431758	7.513789	8.57759	7.31659	8.908714	10.195995
Heirsure (H)	HsS	4.889845	4.568035	4.599193	5.891204	4.693976	5.06769	5.163509	7.203949	9.41544
	HsM	5.164751	5.079769	6.259527	6.752426	6.730659	6.897959	6.758369	7.115178	8.497333
	HhS	4.430483	4.357516	4.364076	4.465338	4.373811	4.620351	4.400491	5.862836	7.788144
	HhM	5.564955	5.60122	4.93193	4.840682	4.844311	5.670954	4.846422	6.470235	8.654882
Rigrsure (R)	RsS	4.889845	4.568035	4.599193	5.891204	4.693976	5.219849	5.163509	7.161943	9.580408
	RsM	6.735506	6.437465	6.259527	6.752426	6.730659	7.132055	6.758369	7.074018	8.480397
	RhS	4.430483	4.357516	4.364076	4.465338	4.373811	4.562203	4.400491	4.893766	7.089057
	RhM	5.597294	5.428461	4.93193	4.840682	4.844311	5.597028	4.846422	5.372702	8.086111
Minimax (M)	MsS	6.937342	6.428099	6.034069	7.047137	6.556032	7.408714	6.609635	6.237572	5.599136
	MsM	2.125373	2.101124	2.245213	3.275954	2.483061	2.109691	2.962915	2.446835	3.592945
	MhS	4.817294	4.506366	4.54746	5.209974	4.611494	5.400822	5.030778	6.589164	8.118678
	MhM	4.212545	4.160218	3.789047	4.504452	4.31733	4.393178	4.598494	4.811363	6.304121

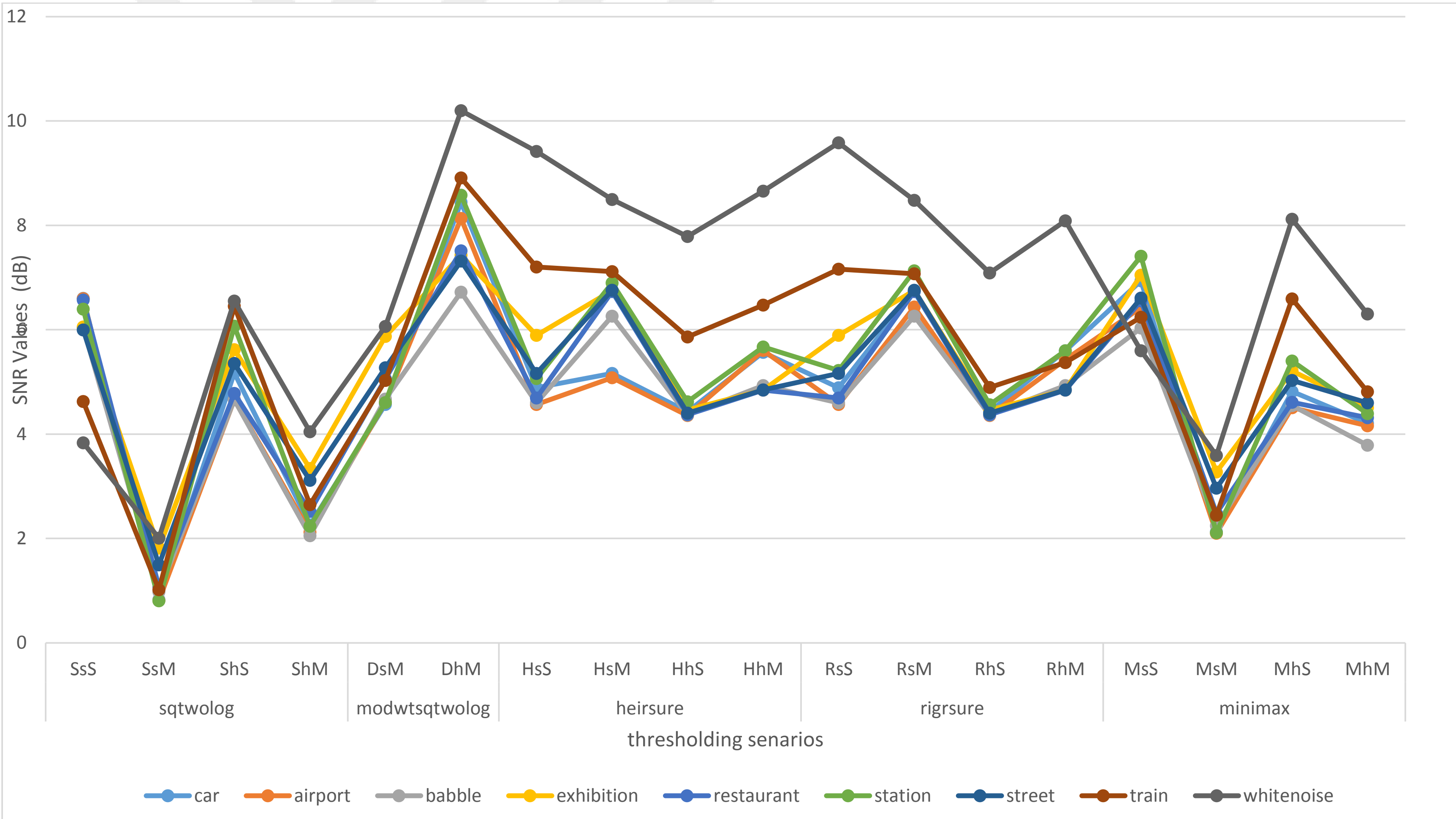

Figure 29. Comparison of wave SNR

Figure 29 Shows the graph of all the values in Table 1. As shown in Figure 1. The highest point of SNR was obtained in modwtsqtwolog and hard thresolding method with mln rescaling (scenario DhM). Input noisy signal is AWGN at 5dB.It can be easily seen from the figure 29 that AWGN noise in DhM method gives the best SNR result when compared to other scenarios and other noise types.

Table 2 shows the MSE results in different scenarios and noise types. As can be show in this table we get good performance of result for MSE in scenario (DhM) gives less MSE than other scenarios. Therefore, in this scenario (DhM) for all type of noisy speech signal we get less MSE and higher SNR value

Table 2. Results of MSE for speech signal

sp01_5db noise		car	airport	babble	exhibition	restaurant	station	street	train	whitenoise
sqtwolog	SsS	0.000256	0.000256	0.000291	0.00029	0.000258	0.000268	0.000294	0.000403	0.000484
	SsM	0.000966	0.000969	0.000933	0.00077	0.000913	0.000971	0.000828	0.000924	0.000736
	ShS	0.000355	0.0004	0.0004	0.00032	0.000389	0.000289	0.00034	0.000265	0.000259
	ShM	0.000694	0.000717	0.000729	0.000542	0.000655	0.000699	0.000571	0.000635	0.00046
modwtsqtwolog	DsM	0.000408	0.000404	0.000398	0.000302	0.000364	0.000162	0.000347	0.000367	0.000289
	DhM	0.000167	0.00018	0.000249	0.000211	0.000207	0.000405	0.000217	0.00015	0.000112
heirsure	HsS	0.000379	0.000408	0.000405	0.000301	0.000397	0.000364	0.000356	0.000223	0.000134
	HsM	0.000356	0.000363	0.000277	0.000247	0.000248	0.000239	0.000247	0.000227	0.000165
	HhS	0.000421	0.000429	0.000428	0.000418	0.000427	0.000403	0.000424	0.000303	0.000195
	HhM	0.000325	0.000322	0.000375	0.000383	0.000383	0.000317	0.000383	0.000263	0.000159
rigrsure	RsS	0.000379	0.000408	0.000405	0.000301	0.000397	0.000351	0.000356	0.000225	0.000129
	RsM	0.000248	0.000265	0.000277	0.000247	0.000248	0.000226	0.000247	0.000229	0.000166
	RhS	0.000421	0.000429	0.000428		0.000427	0.000409	0.000424	0.000379	0.000228
	RhM	0.000322	0.000335	0.000375	0.000383	0.000383	0.000322	0.000383	0.000339	0.000182
minimax	MsS	0.000237	0.000266	0.000291	0.000231	0.000258	0.000212	0.000255	0.000278	0.000322
	MsM	0.000717	0.000721	0.000697	0.00055	0.00066	0.000719	0.000591	0.000665	0.000511
	MhS	0.000386	0.000414	0.00041	0.000352	0.000404	0.000337	0.000367	0.000256	0.00018
	MhM	0.000443	0.000448	0.000489	0.000414	0.000433	0.000425	0.000405	0.000386	0.000274

5.4.1. Evaluation of SNR results of wavelet denoising with different noisy speech data

In Table 1, the SNR results of nine different input noisy signals can be seen. For these input noisy signals, wavelet denoising procedure was applied under different scenarios. All these signal enhancements have been done by first using daubechies5 (db5) wavelet at level 5. (When we selected different levels of decomposition, it didn't effected the results very much. Therefore, we used db5 at level5 for all the test conditions.)

As can be seen from the results in Table 1, when we selected MODWT method and sqtwolog threshold, we obtained higher SNR than other methods. We obtained the maximum performance for SNR for the condition modwtswtwolog and DhM (when noise type is AWGN (Additive White Gaussian Noise) and method is MODWT with hard thresholding condition). In soft thresholding condition SNR is seen as less than hard thresholding. From the results in Table 1, it can be said that MODWT with hard thresholding and mln-scaling (scenario DhM) is best technique for all kinds of noisy speech signals when compared to other scenarios because of less MSE and higher SNR value.

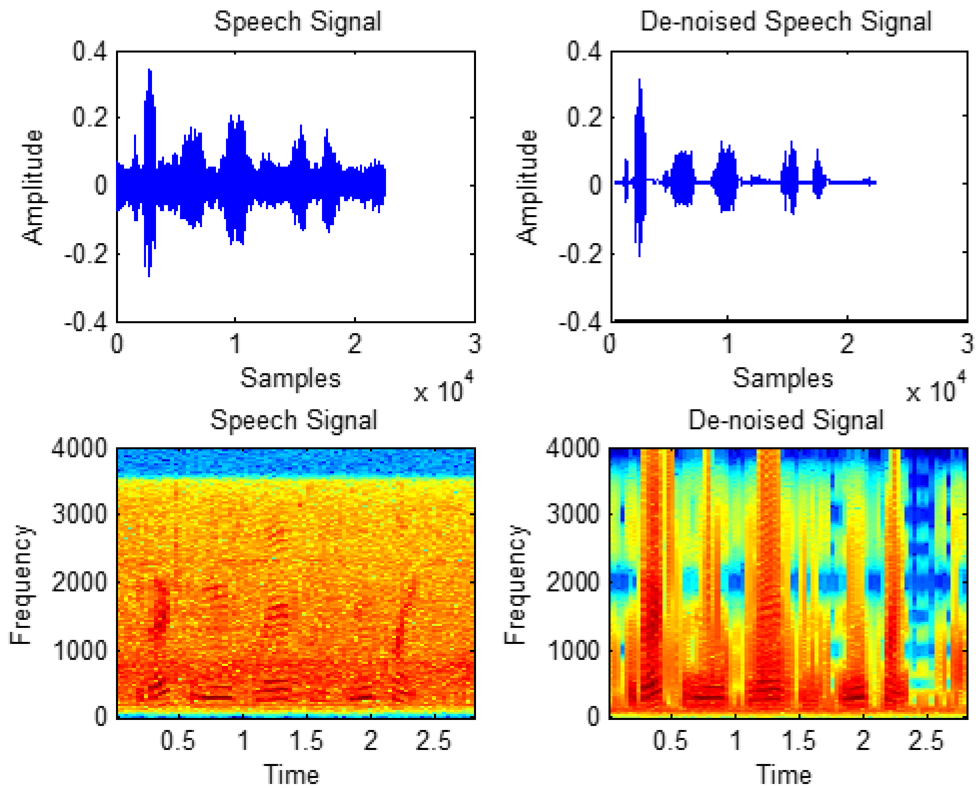

Figure 30. 5dB car noise speech signal and its denoised form during the scenario DsM

In Figure 30, the input signal with a car noise at SNR=5dB and its denoised form can be seen. During this experimental work, the input signal was decomposed using MODWT (modwtsqtwolog) and selecting the scenario (DsM). In this test condition, SNR=4.566519 dB was found. (as mentioned in section 5.4, the noise types of speech files were artificially added to the clean speech signal. Therefore, the clean speech signal in figure 1 was used as reference during all different scenarios).

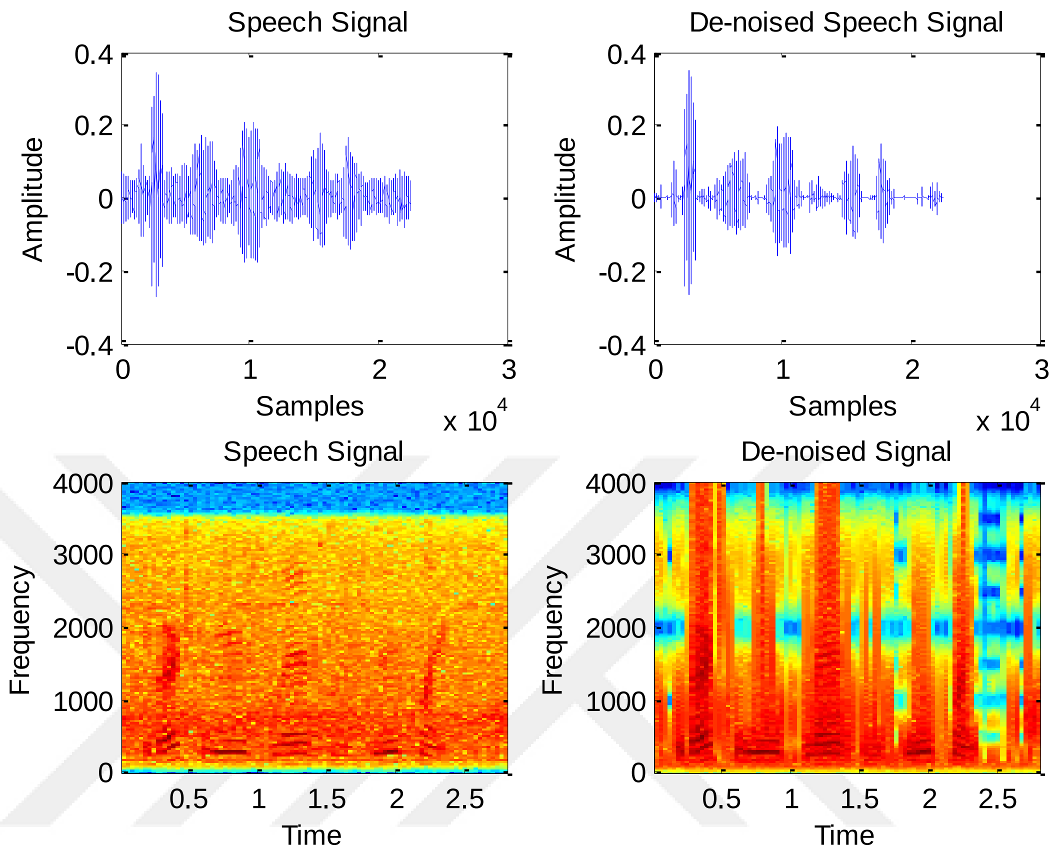

Figure 31. 5dB car noise speech signal and its denoised form during the scenario DhM.

In Figure 31, the input signal with a car noise at SNR=5dB and its denoised form can be seen. Here, the input signal was decomposed using MODWT (modwtsqtwolog) and selecting the scenario (DhM). In this test condition, SNR=8.446362dB was found (it is higher than the SNR in the scenario (DsM)).

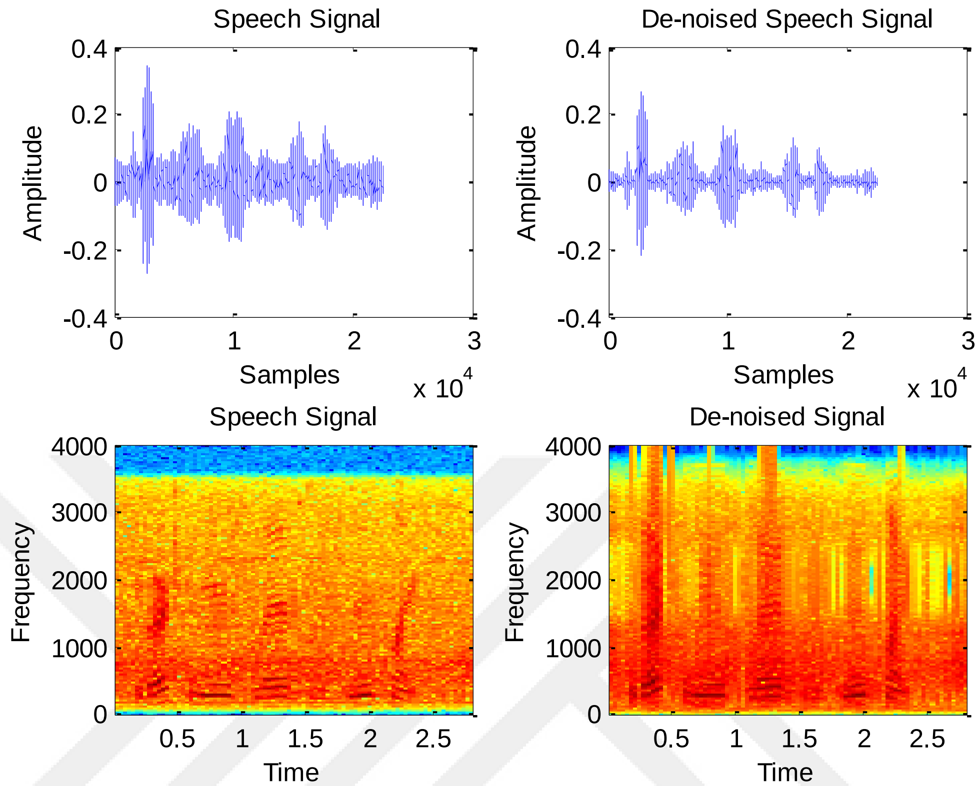

Figure 32. 5dB car noise speech signal and its denoised form during the scenario SsS

In Figure 32, the input signal with a car noise at SNR=5dB and its denoised form is seen. Here, the input signal was decomposed using sqtwolog (without MODWT) and selecting the scenario (SsS). In this test condition, SNR=6.599453dB was found.

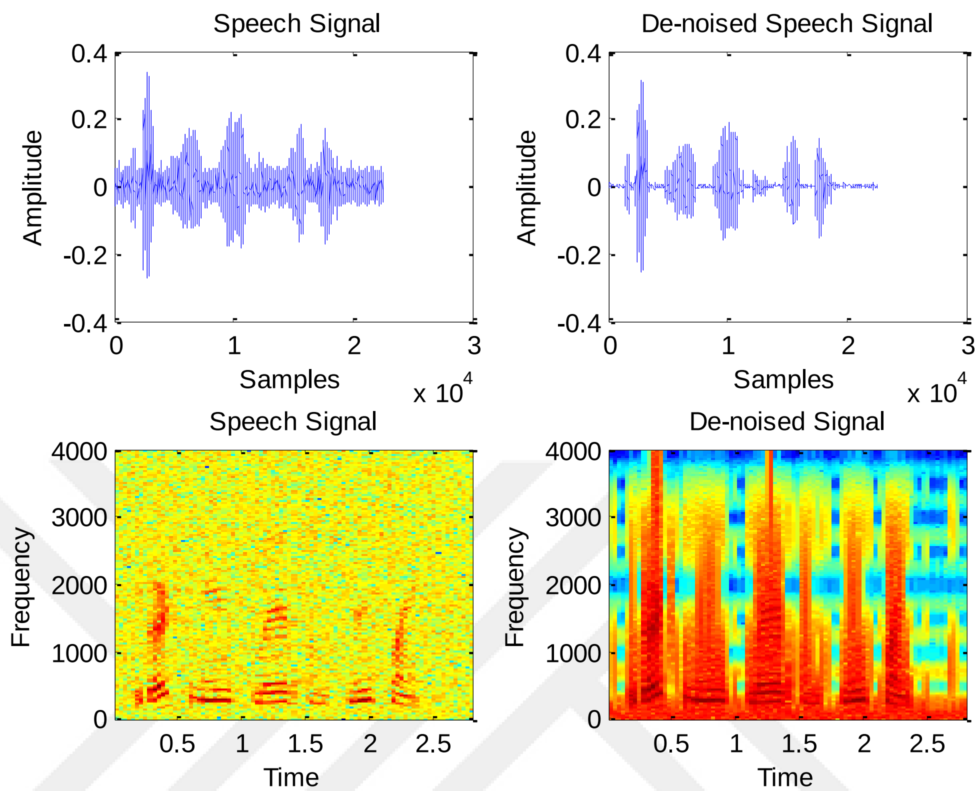

Figure 33. 5dB AWGN noisy speech signal and its denoised form during the scenario DhM.

In Figure 33. Clean speech signal (in Figure 27) was corrupted with a 5dB white gaussian noise (AWGN). De-noising is done usenig db5 wavelet work a level 5, applying threshold technique MODWT and hard thresholding. Here, the noisy input signal was decomposed using MODWT (modwtsqtwolog) and selecting the scenario (DhM). In this test condition, SNR=10.195995dB was found.

5.4.2. SNR results of wavelet denoising for different wavelet families

From the evaluation of the results table 1, it is concluded that modwtsqtwolog (scenario DhM) gives the best SNR result for all different kind of noisy speech signals. Therefore, this scenario was selected and a MATLAB program was designed to evaluate the performance of scenario DhM according to the different wavelet families. Then, by using the same clean speech signal in Figure 1, the noisy speech signals were selected at 0dB , 5dB , 10dB ,and 15dB for input noisy signal. (noise types are car noise , airport noise , babble noise , restaurant noise ,and train noise). The used wavelet functions during the tests can be shown in the Table 3.

Table 3. Wavelet family comparison at 0 dB SNR

wavelet	WaveSNR				
	airport_snr0	car_snr0	babble_0	train_snr0	restaurant_snr0
Haar	3.921514	4.953874	3.89721	4.663745	2.926489
db5	4.051875	5.293733	3.844242	4.69209	2.515858
db10	3.963526	5.293786	3.690797	4.666482	2.421438
db15	3.937564	5.210047	3.573668	4.544382	2.335742
sym5	4.224198	5.418761	3.989513	4.783298	2.528579
sym10	4.056841	5.338404	3.812885	4.781361	2.430903
sym15	4.224187	5.408234	3.800507	4.737344	2.374583
coif3	4.106695	5.371772	3.921855	4.765497	2.507153
coif4	4.086557	5.386845	3.890359	4.745099	2.448227
coif5	4.090118	5.366035	3.838619	4.749305	2.435685

From this table it can be seen that, sym5 wavelet gives better results for SNR at all type of noisy speech signals for input SNR=0db when we compared to the other wavelets family except one type of noisy speech signal (restaurant) gives higher SNR at Haar wavelet. We can also see the results from figure 34. The higher point of SNR is obtained at sym5 wavelet.

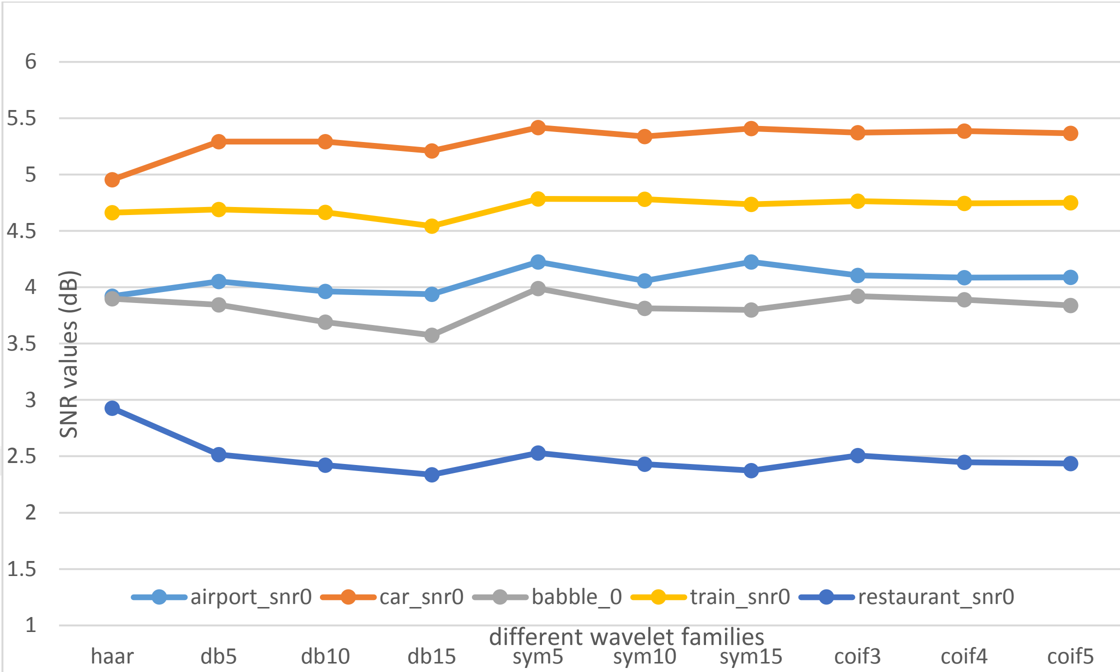

Figure 34. Comparison of wavelet family for 0 dB SNR of speech signal

Table 4. Wavelet family comparison at 5dB SNR

wavelet	wave SNR				
	airport_snr5	car_snr5	babble-snr5	train_snr5	restaurant_snr5
haar	7.378904	7.696322	6.339718	8.297602	7.430855
db5	8.13396	8.446362	6.721453	8.908714	7.513789
db10	8.11928	8.490163	6.682319	8.956868	7.425254
db15	8.040169	8.473918	6.608608	8.906647	7.352999
sym5	8.205666	8.482623	6.744722	8.968946	7.632089
sym10	8.137401	8.46651	6.705735	8.959106	7.493683
sym15	8.173276	8.503411	6.652598	8.98189	7.439504
coif3	8.210693	8.522588	6.736789	8.943089	7.571337
coif4	8.201464	8.521521	6.717993	8.935498	7.535606
coif5	8.182087	8.509159	6.699481	8.945622	7.504453

The results of different wavelets families for different types of noisy speech signals at SNR=5dB can be seen in Table 4. We obtained better SNR results at coif3 wavelet for airport, car and restaurant noisy speech signals. But, for babble and train noisy speech signals we got higher SNR value at sym5. Figure 35 shows the results in Table 4.

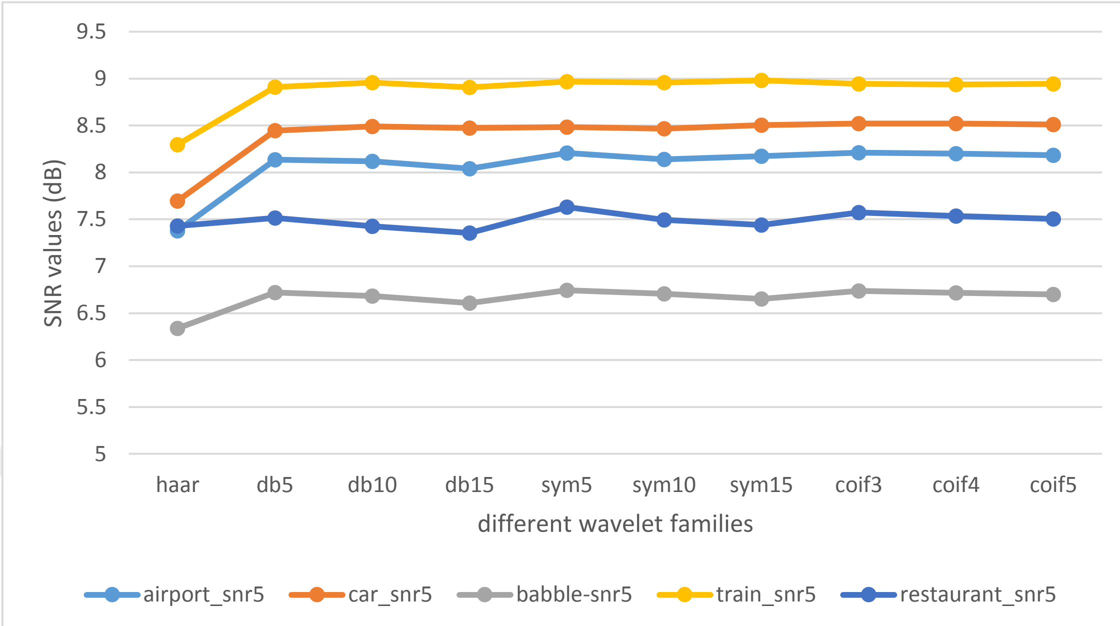

Figure 35.Comparison of wavelet family for 5 dB SNR of speech signal

Table 5. Wavelet family comparison at 10 dB SNR

wavelet	wave SNR				
	airport_snr10	car_snr10	babble_snr10	train_10	restaurant_snr10
Haar	10.790927	10.822377	10.874318	11.584533	10.734209
db5	12.014205	11.593715	11.980205	12.693516	11.560045
db10	11.886973	11.544245	11.835829	12.832093	11.528413
db15	11.897426	11.565462	11.715849	12.853611	11.461991
sym5	11.974079	11.59391	11.921995	12.706739	11.521217
sym10	12.032354	11.617111	11.938741	12.914808	11.518183
sym15	12.007936	11.699244	11.838824	12.99832	11.508912
coif3	12.014915	11.610166	11.92487	12.766572	11.531545
coif4	12.050653	11.623474	11.908661	12.857187	11.553395
coif5	12.029903	11.633294	11.882987	12.913901	11.544564

In table 5. The results of different wavelets families for different types of noisy speech signals at SNR=10dB can be seen. As can be shown from the Table 5 and also from Figure 36, airport and restaurant noisy speech signals gives better SNR performance at coif4 wavelet than other wavelets family. For speech signal type babble, db5 gives better result than other wavelets family. For the car and train noisy speech types, higher SNR value is obtained for sym15 wavelet.

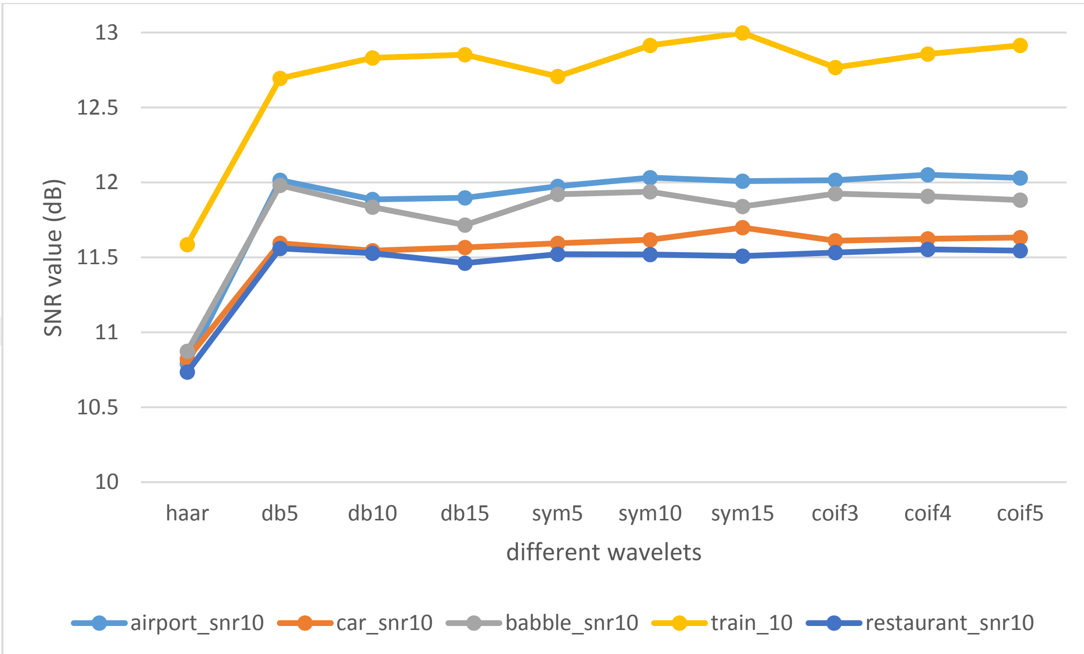

Figure 36. Comparison of wavelet family for 10 dB SNR of speech signal

Table 6. Wavelet family comparison at 15 dB SNR

wavelet	wave SNR			
	airport_snr15	car_snr15	babble_snr15	restaurant_snr15
haar	14.45143	14.273113	14.421139	14.271985
db5	15.991195	15.67425	15.85968	15.588729
db10	15.887608	15.671038	15.774993	15.660427
db15	15.675922	15.641824	15.67872	15.53728
sym5	15.899046	15.653432	15.70727	15.546851
sym10	15.894567	15.733487	15.863074	15.662126
sym15	15.802554	15.787181	15.814468	15.613053
coif3	15.905901	15.639428	15.778893	15.601235
coif4	15.907968	15.734413	15.820347	15.604301
coif5	15.893818	15.734954	15.853104	15.640554

Table 6 gives the results of different wavelets families for different types of noisy speech signals at SNR=15dB. It can easily be seen that, airport noisy speech signal gives better SNR result for db5 wavelet than other wavelets family. The interesting point here that, car noisy speech signal gave better SNR result at sym15 wavelet like in Table 5. For noisy speech signals babble and restaurant type, we got better performance at sym10 wavelet than other wavelet family. The results also can be seen in Figure37.

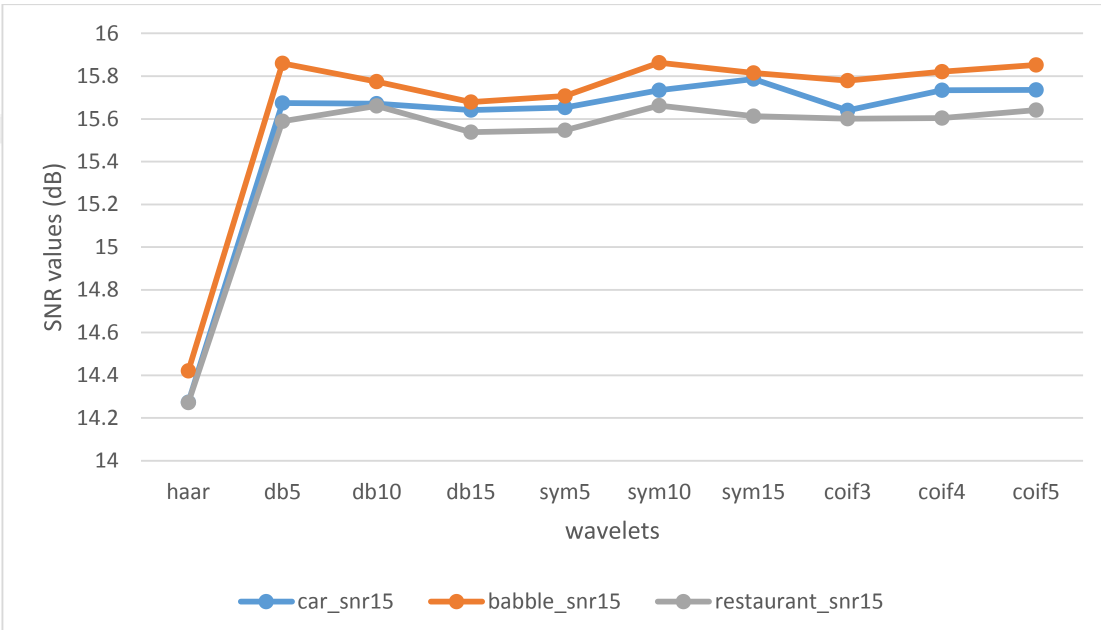

Figure 37. Comparison of wavelet family for 15 dB SNR of speech signal

CONCLUSION

Wavelet transform is a modern technology method for speech signal enhancement. In this study, we measured the performance of many kind of wavelet families according to different noisy speech signals. The aim of this work is to remove background noise by using different threshold methods by analyzing the input speech signals. In this work we measured the performance by calculating SNR and MSE values. Then, we tried different thresholding methods for input noisy signals. When we selected MODWT thresholding method together with sqtwolog threshold, we obtained higher SNR than other methods. From the evaluation of the results, it is concluded that modwtsqtwolog (scenario DhM) gives the best SNR result for all different kinds of noisy speech signals. The application of MODWT method to all these noisy speech signals to measure performance is a new technique that has not been found in any previous work in the literature. This study proved the success of this method on different types of noisy speech signals by trying many kind of scenario during tests. From the obtained results, it can be said that MODWT method can be used to obtain high quality speech enhancement of noisy speech signals.

By selecting MODWT method, we evaluated the performance of scenario DhM at different wavelet families (haar, db, sym, coif). We found that sym5 wavelet gave better result for noisy speech signal at input SNR=0db. When we used the input SNR=5db for noisy speech signal, we saw that the majority of the speech signal types gave better SNR value at coif3 wavelet. For the noisy speech signals at SNR=10db, the wavelets families giving better result are (coif5, db5 and sym15 wavelets) in different noise types. For noisy speech signal at SNR=15db, we obtained better SNR results for sym wavelet at different scales (sym10, sym15).

From all obtained results, we evaluated the performance of wavelet families for different types of noisy speech signals. It can be said that even if some wavelet families gives better results in different noise types, their performance are found close to the other wavelet families in the tests. All wavelet families give good performance against input noisy signals. When SNR level of input noisy speech is high such as SNR=0dB and SNR=5dB, the denoising performance of wavelets are higher when we compared to SNR=10dB and SNR=15dB results.

From all the test results in this study, it can be said that the usage of thresholding method for scenario (DhM) for all different wavelets family and different type of noisy speech signal is an advantage for speech signal enhancement.

For a feature work, the MODWT method can be tried on noisy image and video signals to evaluate its signal enhancement performance on these types of signals.

REFERENCES

[1] Meyer, Y. (1992). *Wavelets and operators* (Vol. 1). Cambridge university press.

[2] Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. M. (1996). Wavelet toolbox. *The MathWorks Inc., Natick, MA*, *15*, 21.

[3] Walker, J. S. (2008). *A primer on wavelets and their scientific applications*. CRC press.

[4] Venkateswarlu, S. C., Reddy, A. S., & Prasad, K. S. Speech Enhancement in terms of Objective Quality Measures Based on Wavelet Hybrid Thresholding the Multitaper Spectrum.

[5] Kim, E. (2016). *A Wavelet Transform Module for a Speech Recognition Virtual Machine* (Doctoral dissertation, Minnesota State University, Mankato).

[6] Yuan, X. (2003). *Auditory model-based bionic wavelet transform for speech enhancement* (Doctoral dissertation, Marquette University).

[7] Kumari, V. S. R., & Devarakonda, D. K. (2013). A wavelet based denoising of speech signal. *International Journal of Engineering Trends and Technology (IJETT)–Volume*, *5*, 107-115.

[8] Villanueva-Luna, A. E., Jaramillo-Nuñez, A., Sanchez-Lucero, D., Ortiz-Lima, C. M., Aguilar-Soto, J. G., Flores-Gil, A., & May-Alarcon, M. (2011). De-noising audio signals using MATLAB wavelets toolbox. In *Engineering education and research using MATLAB*. InTech.

[9] You, F., & Zhang, Y. (2013). Research of an Improved Wavelet Threshold Denoising Method for Transformer Partial Discharge Signal. *Journal of multimedia*, *8*(1).

[10] Polikar, R. (1996). The wavelet tutorial.

[11] Cornish, C. R., Bretherton, C. S., & Percival, D. B. (2006). Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Boundary-Layer Meteorology*, *119*(2), 339-374.

[12] Verma, N., & Verma, A. K. (2012). Performance analysis of wavelet thresholding methods in denoising of audio signals of some Indian Musical Instruments. *International Journal of Engineering Science and Technology*, *4*(5), 2040-2045.

[13] Burrus, C. S., Gopinath, R. A., Guo, H., Odegard, J. E., & Selesnick, I. W. (1998). *Introduction to wavelets and wavelet transforms: a primer* (Vol. 1). New Jersey: Prentice hall.

[14] Debnath, L., & Shah, F. A. (2002). *Wavelet transforms and their applications* (pp. 12-14). Boston: Birkhäuser.

[15] Chavan, M. S., & Mastorakis, N. (2010). Studies on implementation of Harr and Daubechies wavelet for denoising of speech signal. *international journal of circuits, systems and signal processing*, *4*(3), 83-96.

[16] Munegowda, B. K. (2016). *Performance and Comparative Analysis of Wavelet Transform in Denoising Audio Signal from Various Realistic Noise* (Doctoral dissertation, Napier University, Edinburgh, Scotland, United Kingdom).

[17] Wang, X. (2011). *Event detection for post lung transplant based on home monitoring of spirometry and symptoms* (Doctoral dissertation, University of Minnesota).

[18] Al Wadi, S., Hamarsheh, A., & Alwadi, H. (2013). Maximum overlapping discrete wavelet transform in forecasting banking sector. *Applied Mathematical Sciences*, *7*(80), 3995-4002.

[19] El-Fattah, M. A. A., Dessouky, M. I., Abbas, A. M., Diab, S. M., El-Rabaie, E. S. M., Al-Nuaimy, W., ... & El-Samie, F. E. A. (2014). Speech enhancement with an adaptive Wiener filter. *International Journal of Speech Technology*, *17*(1), 53-64.

[20] Renisha, G., & Jayasree, T. (2015). Enhancement of Speech Signals in a Noisy Environment based on Wavelet based Adaptive Filtering. *CInternational Journal of Signal Processing, Image Processing and Pattern Recognition*, *8*(9), 69-76.

[21] Seke, E., Durak, M. H., & Özkan, K. (2016). Speech Denoising using Common Vector Analysis in Frequency Domain. *International Journal of Applied Mathematics, Electronics and Computers*, *4*(Special Issue-1), 224-229.

[22] JaiShankar, B., & Duraiswamy, K. (2012). Audio denoising using wavelet transform. *International Journal of Advances in Engineering & Technology*, *2*(1), 419.

[23] Sumithra, A., & Thanushkodi, B. (2009). Performance evaluation of different thresholding methods in time adaptive wavelet based speech enhancement. *IACSIT Int. J. Eng. Technol*, *1*(5), 42-51.

[24] Nemade, M., & Shah, S. K. (2013). Performance comparison of single channel speech enhancement techniques for personal communication. *For International Journal of Innovative Research in Computer and Communicaiton Engineering*, *1*(1).

[25] Soundarya, S. S., & Rajani, A. (2016, August). Robust estimation of non stationary noise power spectrum using E-DATE and Kalman filter. In Inventive Computation Technologies (ICICT), International Conference on (Vol. 1, pp. 1-5). IEEE.

[26] Shaila D., Speech Processing Applications in Speech and Audio Processing

[27] So, S., & Paliwal, K. K. (2011). Modulation-domain Kalman filtering for single-channel speech enhancement. *Speech Communication*, *53*(6), 818-829.

[28] Paliwal, K., Wójcicki, K., & Schwerin, B. (2010). Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. *Speech communication*, *52*(5), 450-475.

[29] Hirsch, H. G., & Pearce, D. (2000). The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*.

[30] Loizou, P. (2017). NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms. *Speech Commun*, *49*, 588-601.