DISSERTATION

on

Wavelet Transforms in Speech Processing

Submitted in Partial Fulfilment of the Requirements

for the Award of Degree of

Master of Technology
in
Computer Science and Applications

Submitted By

Arun Kumar

(Regn. No. 601203004)

Supervised By

Dr. R. K. Sharma
Professor

SCHOOL OF MATHEMATICS AND COMPUTER APPLICATIONS
THAPAR UNIVERSITY
PATIALA – 147004

NOVEMBER 2014

CERTIFICATE

This is to certify that the work which is being presented in this dissertation titled, "Wavelet Transforms in Speech Processing", in partial fulfilment of the requirements for the award of the degree of Master of Technology in Computer Science and Applications submitted to school of Mathematics and Computer Applications of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. R. K. Sharma and refers other researchers' work which are duly listed in the reference section.

The matter presented in the dissertation has not been submitted for the award of any other degree of this or any other University.

Arun

(Arun Kumar)
Regn. No. 601203004

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. R. K. Sharma)
Professor
SMCA

Countersigned by

(Dr. Rajesh Kumar)
Head, SMCA
Thapar University
Patiala

(Dr. S. S. Bhatia)
Dean of Academic Affairs
Thapar University
Patiala

i

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my mentor and supervisor, Dr. R. K. Sharma for his immense help, able guidance, stimulating suggestions and constant encouragement. He always provided me the much needed motivation and enthusiastic atmosphere while working under him. It was a great pleasure to do dissertation under his scholarly supervision.

I am also thankful to Dr. Rajesh Kumar, Head, School of Mathematics and Computer Applications; and Dr. Singara Singh, P.G. Coordinator for their regular support and encouragement.

I would also like to thank all the faculty members and staff of the department who were always there at the need of the hour and provided me all the help and facilities which I required for the completion of this work.

I express my thanks to my family for their love, support and enthusiastic encouragement without which I could not have completed this work. I also want to thank all my friends for their optimism and moral support extended to me throughout the completion of this work. Finally, I thank the Almighty who gave me enough strength to complete this work.

Arun

(Arun Kumar)
Regn. No. 601203004

ii

ABSTRACT

This is an era of computers. Computers have become an integral part of human life as they are influencing our lives in every possible way, and are used by every person irrespective of age and sex. The ease with which we can exchange information between user and computer is of immense importance today. But the input devices like mouse and keyboard when used as an interface to exchange the information have some limitations also. Speech which is a natural and quick way of exchanging the information between humans, if used to communicate with computers can overcome all these limitations. Speech recognition is an area of research which has attracted many researchers across the world.

Speech is a time varying signal. The information contained in the signal is very difficult to analyze. Traditional methods of speech recognition use Mel Frequency Cepstral Coefficients (MFCC) and Short Time Fourier Transform (STFT) to extract the features out of a speech signal, and the model has been successfully implemented in many speech recognition engines. The wavelet transform has better time-frequency localization property as compared to the Short Time Fourier Transform (STFT). For better feature extraction of the speech signal we use the applications of wavelets, that is, speech recognition. Here, in the present work, the concept of wavelet transform has been used for feature extraction from the speech signals. The speech signal of short duration has been taken for the purpose. All the speech signals are phonemes (fundamental unit of speech). The processed speech signal was fed to a neural network for classification to measure the performance of the feature extraction. For the classification of speech signal vector, the self-organizing feature map (SOFM) has been used which is a type of the unsupervised learning method. The results are good and require further experimentation in future.

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

1.1 Speech Processing

Speech processing is the study of speech signals and processing methods of these signals. The signals are frequently processed in digital representations, as such this is also known as digital signal processing. There are various aspects of speech processing such as storage, manipulation, acquisition and transfer, and output of these speech signals.

It is also directly related to the Natural Language Processing (NLP), as its input should be able to come from and output should be capable of going away to the NLP applications. For example, text-to-speech synthesis might make use of a syntactic parser on its input text; and speech recognition output can be used by information extraction techniques. There are various applications of speech processing, including compression and synthesis of the human speech and recognition.

Speech processing has a number of associated areas:

- **Speech Recognition:** The speech processing, also known as voice recognition, is the analysis of the content of a speech signal of a language and the conversion of the speech signal into a computer-readable format (Daubechies, 1992).
- **Speaker Recognition:** The aim of the speaker recognition system is to analyze the identity of the speaker.
- **Speech Coding:** It is a specific approach for the data compression, and is an important step in the area of telecommunication.
- **Voice Analysis for Medical Purposes:** There are various voice analysis for medical purposes such as dis-function of the vocal cords, analysis of vocal loading.
- **Speech Synthesis:** The speech synthesis is the artificial synthesis of speech which is computer-generated speech. The advancement in the area of speech synthesis improves the computer's usability for the visually impaired.

- **Speech Enhancement:** It is the process of enhancing the quality of speech signals and/or transparency of the speech signals, like reduction of the audio noise for the audio signals.
- **Speech Compression:** It is an important step in the area of telecommunication for the most amount of information which can be stored or heard, and transferred for a given set of time and space constraints.

Speech is the most likely form for the communication of human beings. One of the most information laid signal is speech. The speech sound has a number of multi-layered temporal, spectral and rich variations that suggest intention, accent, expression, words, gender, age, style of speaking and state of health of the speaker and emotion. Speech sounds are generated by the push of the inhaled air from the lungs through the vocal tract and vocal cords and away from the nose and lips airways that are produced by the air pressure vibrations.

A language is written in the sequence of the elementary alphabets. Speech is also a series of the acoustic symbols and sounds called as phonemes that suggest the spoken structure of the language. There are about 40-60 phonemes in English language. Out of these, a huge number of spoken words can be constructed. In general, the construction of every phoneme sound is exaggerated by the neighbouring phoneme.

The speech signals communicate much more as compared to the spoken words. The information that is conveyed by the speech signal is multi-layered, and has the time and frequency modulation of carriers of the information as tone and formants. The information conveyed in the speech has the following contents:

(a) Acoustic Phonetic Symbol: These symbols are the most basic unit of the speech signals from which large number of speech units such as words and syllables are formed.

(b) Prosody: These are the rhythms of the speech signals that are mostly intonation signals conceded by the stress and the pitch curve. Prosody eliminates the ambiguity like whether a spoken word is a question or a sentence or helps to signal such information as there are boundaries between link sub phrases and different parts of the speech.

(c) Gender Information: The gender is conveyed by the size and physical properties of the vocal tract and by the pitch (that are related to the elementary frequency of the voiced sounds). The female voice has a higher pitch and the higher resonance frequencies due to the difference in the vocal anatomy.

(d) Age: The age is conveyed by the flexibility of the vocal cords, and by the effect of the size and the pitch. The pitch of voice of children may be more than 300 Hz.

(e) Speaker Identity: The speaker identity is conveyed by the physical properties of a person's vocal tract, pitch intonations and vocal folds.

Since the computers have been developed progressively, there are varieties of research activities that we are dealing in the area of computer human interface. The input devices like keyboard and mouse are even though very popular way to interact with the computer, but there are some limitations also as the keyboard require a definite quantity of skill for effective and fast usage of the computer and mouse. On the other hand, it also requires a very good hand and eye co-ordination.

Speech is a very easy and natural way for the exchange of information, if it is used as a way to interact with the computer and a medium to solve all the problems. The speech recognition system makes it possible for computers to track human voice and understand the human languages. The main aim of the speech recognition system is to create systems and techniques for the speech as an input to the machine.

During the last 60 years, many advancements and researches took place in the area of speech processing, and many systems were developed during the period, but despite all this there is still one more research challenge in the area of speech recognition, i.e., the accuracy of automatic speech recognition.

The speech recognition system performs well when tested on data that is similar to that used for training. However, lack of robustness of the speech recognition system continues to be a serious problem to the speech recognition systems. The speech recognition systems present the speech waveform as a feature vector. There is some flavour of cepstral coefficients for the extraction of feature vectors such as, Linear Predictive Coding (LPC) Cepstral coefficients, or Mel Frequency Cepstral Coefficient (MFCC). Linguistic and acoustic are the models that have been used with the features to estimate what speech waveform said (Blu, 1993).

Cepstral coefficient is a mature technique to extract feature vectors, but gives limited robustness.

Speech is a difficult signal that is produced as a result of many transformations occurring at various levels; and speech recognition is a miscellaneous field with a lot of applications in which the speech processing is an important thing (Lei and Tung, 2005). Speech recognition and processing are now a demanding area of research due to the broad variety of applications such as weather forecasting, mobile applications, video games, agriculture, transcriptions, etc.

Many techniques work well, because there is an amplitude difference between vowels and consonant, but by making a precise determination, it is not so easy to find where each signal ends. Traditionally, we use Short Time Fourier Transform (STFT) for extraction of the speech signals for the time-frequency localization. It is always difficult to extract the frequency component and to locate the beginning of a speech segment accurately at the same time.

In recent times, the wavelet transform has become an option for the Short Time Fourier Transform (STFT) for the analysis of non-stationary signals. The wavelet transform has localization in both time and frequency domain (Burrus *et al*., 1997). In this, we use a constant-Q analysis for the representation of the speech signal in the time-scale plane. The wavelet transform has proved to be a potential in the speech processing applications in many areas of interest. These applications include synthesis, speech analysis, and modification of speech and music sound, pitch detection and speech compression and speech recognition.

Traditional methods that we use for the analysis of speech signals use Mel Frequency Cepstral Coefficient (MFCC) and the Short Time Fourier Transform (STFT) to extract the feature out of the speech signals; and the model that we use has been successfully implemented in several speech recognition systems. The performance of wavelet transform varies depending on the context, and the results are good only when we use the speech signals of short durations. The wavelet transform has better time-frequency localization property than STFT (Gupta and Gilbert, 2001). Due to this reason, wavelet transform was applied for better feature extraction of speech signals.

For better representation of the speech signal, the application of wavelet analysis is considered. Speech recognition has many real time applications, and is being widely used in computers, cellular phones and other security systems. However,

these systems failed to provide correct classification of human speech into words. Speech recognizer has the extraction stage and the classification stage. Then, the parameter from the feature extraction stage is compared to the parameter extracted from the signal stored in a template or database. The parameter is then fed to the neural network for classification.

The speech recognition system carries out those applications which are based on the speech features obtained by the Fourier Transform (FT), Short Time Fourier Transform (STFT) and Linear Predictive Coding (LPC) techniques. These methods have some limitations. These methods accept stationary signals and have localized only in the time domain. There is another method that overcomes these limitations is wavelet transform. The wavelet transform has localizations in both the time and the frequency domain, and also accepts the non-stationary signals. Discrete Wavelet Transform method has been used for speech processing.

1.2 Speech Recognition

Speech recognition is the method for the conversion of the spoken input into text. With the help of speech recognition, it allows us to provide an input to an application with our voice. It is just like typing on the keyboard, click with the mouse and by pressing a key on the phone keypad gives input to an application, and the speech recognition method helps us to provide input by talking. The speech recognition, sometimes, is also known as Automatic Speech Recognition (ASR) or speech to text (STT).

Speech recognition basically means talking to a computer, having it recognized what we are saying, and lastly, doing this in real time. This process fundamentally functions as a pipeline that converts Pulse Code Modulation (PCM) digital audio from a sound card into recognized speech. The elements of speech recognizer are displayed in Figure 1.1.

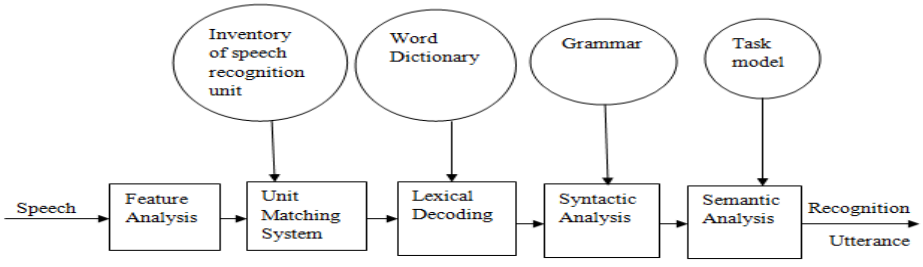


Figure 1.1: Block Diagram of a Speech Recognizer

1.3 Speech Recognition Model

The typical model of Speech Recognition is given in Figure 1.2.

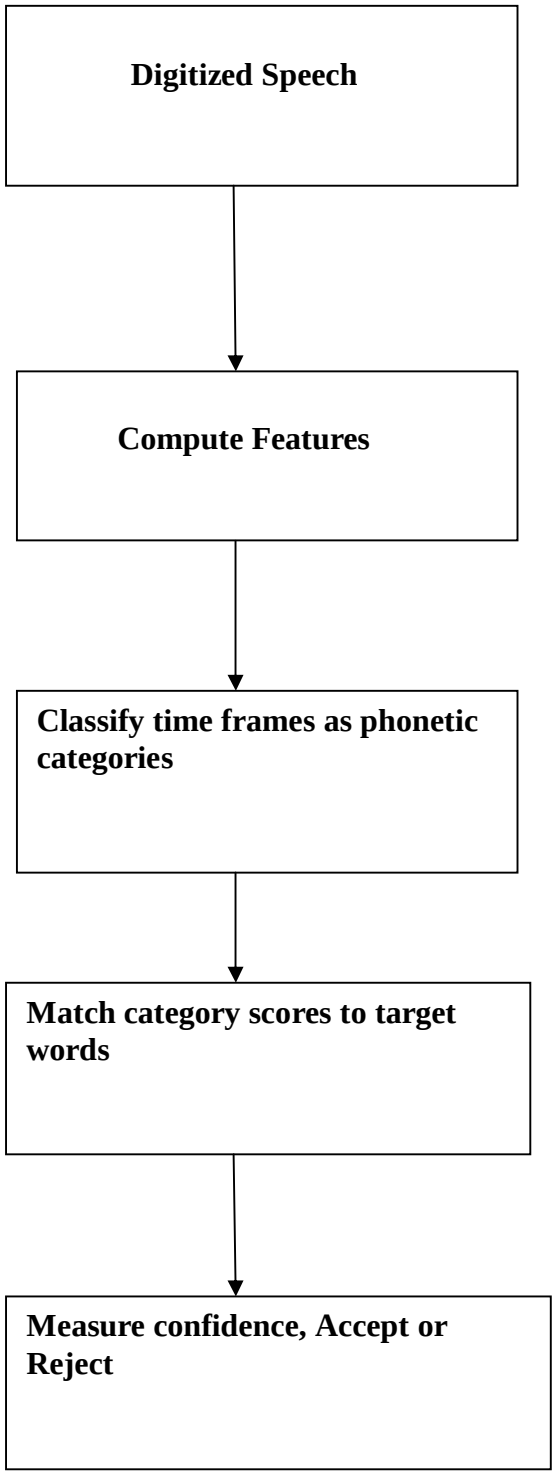


Figure 1.2: Typical Speech Recognition Model

1.4 Components of Speech Recognition System

The components of Speech Recognition System are shown in Figure 1.3.

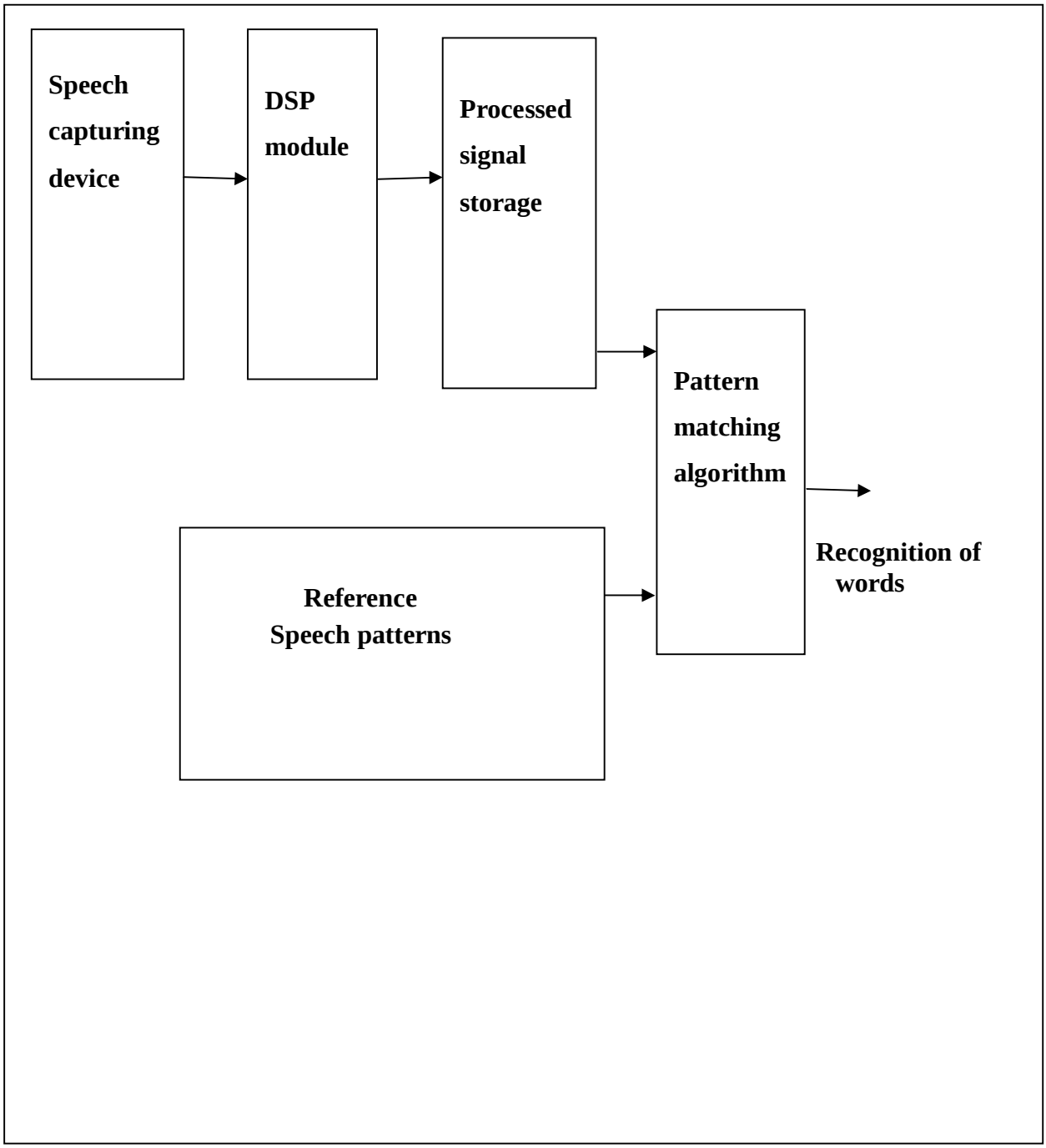


Figure 1.3: Components of Speech Recognition System

1.5 Some Basic Speech Terminologies

(a) Utterance: The utterance is the stream of speech signals between any two periods of the silence. When the users have to say something, then it is known as utterance.

(b) Phoneme: A phoneme is the unit of the phonology, which is when combined with another phoneme form a unit such as words or morphemes. A phoneme has the semantic content and is the minimum amount of the sound such as the phoneme EH versus the phoneme AE detects the difference between the two words bet and bat. The phoneme can also be described as the minimum contrastive linguistic unit which may change the meaning of the phoneme. So, the difference between the two words of the English like kiss and kill in the meaning is the result due to the exchange of the phoneme. The two words that differ in the meaning due to a contrast of a phoneme form a minimum pair. To convert a spoken input into text, the speech engine uses all types of statistical model, data and algorithm.

(c) Grammar: A grammar is a term that uses a set of rules, or an exacting syntax, to define the phrase and words that can be predictable by the speech recognition engine. In linguistics, the grammar is the set of rules overriding the composition of words, phrases and clauses in a given natural language. The term 'grammar' also refers to the study of the rules and includes the fields such as syntax, phonology, and morphology.

(d) Speaker Dependence: It explains the level to which a speech recognition system requires knowledge in the processing of the speech signals successfully of speaker's individual voice. The speech recognition engine is able to train your voice according to the manner how you speak the phrase and words. The speech recognition system needs a user to train your voice known as speaker dependent systems.

So, these are some of the basic speech terminologies which have been used frequently in this research work.

1.6 Wavelets

Wavelets are functions that assure certain requirements. The name wavelet comes from the requirement that they should integrate to zero. A wavelet is a wave with an amplitude that begins with zero. The wavelet with amplitude begins with zero, increases, then again integrates to zero. There are various types of wavelets. One can

select between the efficiently supported wavelets, even wavelets, wavelets with easy statistical expressions and wavelets with simple connected filters.

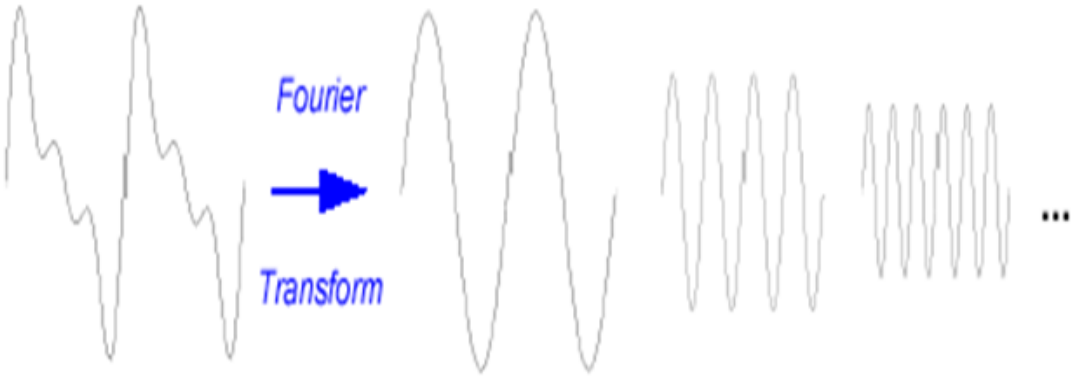
The theory of the wavelet is quite similar to that of the Fourier analysis. The wavelet uses the same approach in order to extract the important information which is to plan linearly the signal on a function base. However, while the Fourier functions are oscillating substantially, the wavelets are the functions that have localization with time.

In mathematical form, the wavelet series is an illustration of a square integrable (real or complex valued) functions with an orthonormal set of the basic functions, complete functions generated by a wavelet. The wavelet transform is one of the best methods for the time-frequency localizations. The wavelet transform is a method for monument up functions, data into components of dissimilar frequency or operators, components, allow studying each component separately. The fundamental idea of the wavelet transform is to present some arbitrary function $f(t)$ as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets can be obtained from a wavelet called the mother wavelet, translations (shift), or by dilation or contractions (scaling).

Continuous Wavelet Transforms:

- Fourier transforms:

The diagram of a Fourier Transform is given in Figure 1.4.



Signal Constituent sinusoid of different frequencies

Figure 1.4: Fourier Transform

A Continuous Wavelet Transform (CWT) can be defined as the sum over all times of the signal multiplied by scale, shifted edition of the wave function.

$$\gamma(s,\tau) = \int f(t)\ \Psi^*_{s,\tau}(t)\,dt$$

Difference between Wavelet Analysis and Fourier Analysis

There are several important differences between the wavelet analysis and the Fourier analysis. Fourier transform has localization only in the time domain but the wavelet transform has localization in both time and frequency domain. If there is a small frequency change in Fourier transforms, then there is a change everywhere in the time domain. Wavelets have localizations in both time and frequency domain (via dilations and translations). This localization has great advantage in many cases. This work uses the continuous wavelet transform (CWT) to divide a continuous-time function into wavelets. Unlike Fourier transforms, the continuous wavelet transform has the ability to build a time frequency representation of the signal and also its good localization in time and frequency domain. So, the main difference between the wavelet analysis and the Fourier analysis is that the wavelet analysis has localization in both time and frequency domain, but the Fourier analysis has localization in the time domain only, as illustrated in the function given below.

$$X_W(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\Psi^*\left(\frac{t-b}{a}\right)dt$$

where, $x_0\,(t)$ is a continuous function that is localized in both time and frequency domain called the mother wavelet and * represents the operation of complex conjugate.

1.7 Wavelet Transforms

The 'wavelet transform' is a term that is based on the wavelet named as mother wavelet. A wavelet is a wave like oscillations that has an amplitude that starts with zero, increases and then decreases back to zero. A wavelet is a statistical function with exacting characteristics such as zero mean and finite energy. For convenience, the energy is frequently normalized. Wavelets have localizations in both frequency and time domain (contradiction and dilations). Mallat in 1989 demonstrated that we can

characterize the wavelet family by scaling function that is built from the wavelet named as mother wavelet. They form a low pass filter and a high pass filter couple that can also be used for the Fast Fourier Transform (FFT). The analysis of the non-stationary signals that we have done with the Fourier Transform (FT) and the Short Time Fourier Transform (STFT) using the FT and STFT does not give any suitable results. By using the wavelet analysis, we can obtain better results. The main advantage of using the wavelet analysis is the capability to perform local analysis. The wavelet analysis as compared to the Short Time Fourier Transform (STFT) makes it possible to perform a multi-resolution analysis.

1.8 Haar Wavelet

Haar wavelets are the wavelets that were proposed initially by the researchers. In mathematical terms, the haar wavelet is the sequence of the rescaled "square-shaped" functions which collectively form a wavelet family or basis. There are some disadvantages of the haar wavelet such as their non-differentiability characteristics, but their simplicity characteristics make them more popular. The main advantage of the haar wavelet is that it is not continuous and not differentiable. The mother wavelet is the scaling and constant function on which acts as an average, and is the step function (Fig.1.5) which acts as numerical differentiation.

The typical diagram of the haar mother wavelet is produced in Figure 1.5.

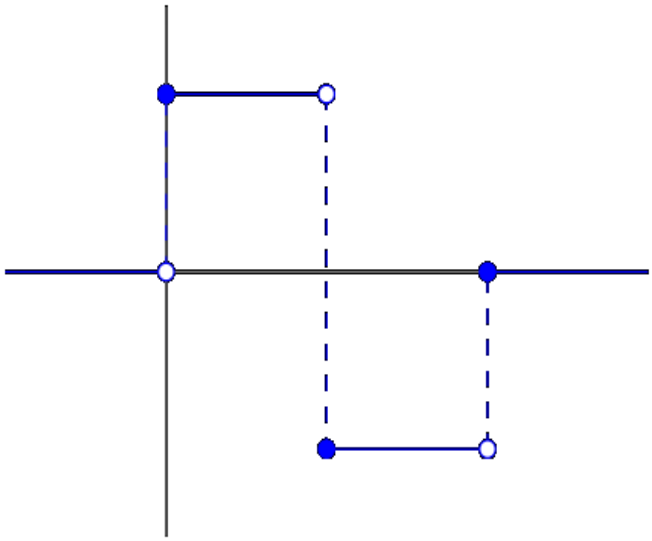


Figure 1.5: Haar Mother Wavelet

1.9 Daubechies Wavelet

The daubechies is a family of the wavelets that is discovered by Ingrid Daubechies. The concept is almost the same to that of haar wavelet. The difference between both the wavelets is that they are differing in that how scaling functions and the wavelets are defined. For the Daubechies wavelet transform, the scaling signals and the wavelets produce longer support that they generated difference and average using few more values of the signal. The daubechies wavelet is the family of the orthogonal wavelets that defines a discrete wavelet transform that are characterized by maximum number of vanishing moments and is widely used in the applications of the wavelet transform. They are efficient in order to extract the information about the speech signals. They are the family of orthogonal wavelets that has the highest number of vanishing moments and is the one whose scaling function has the external phase (Fig. 1.6). They were constructed from the Daubechies db2 wavelet recursively, which is equal to the haar wavelet. They are numbered with even numbers db2n wavelet (e.g., sdb2, db4, db6... db20) that has n vanishing moments.

The typical diagram of the daubechies wavelet is shown in Figure 1.6.

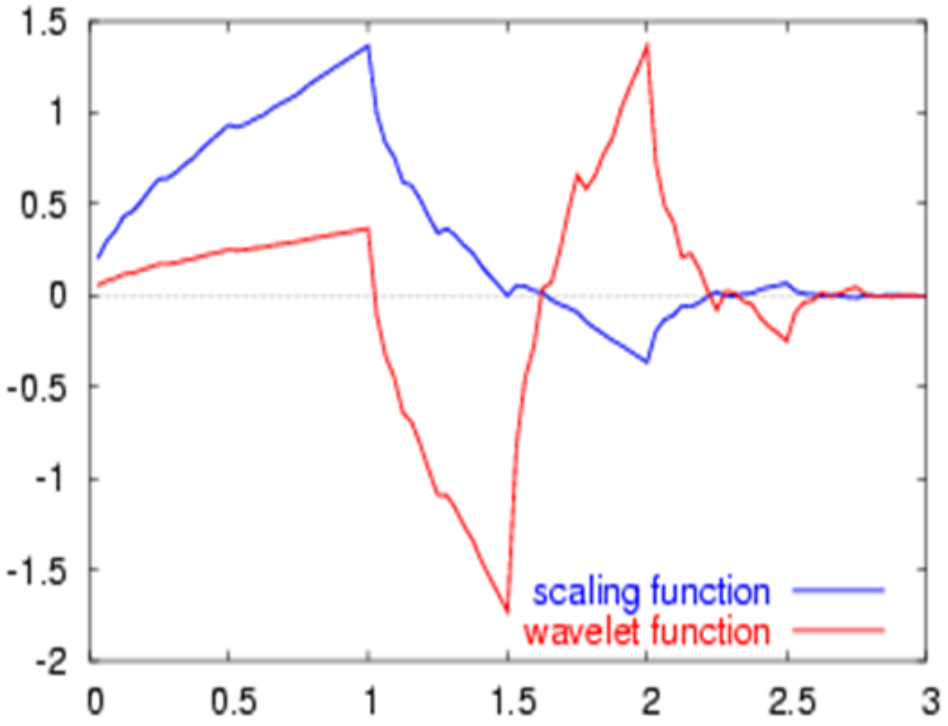


Figure 1.6: Examples of Daubechies Wavelet (Mother Wavelet and Scaling Function) db4

1.10 Continuous Wavelet Transforms

The continuous wavelet transforms are quite similar to the Fourier Transforms (FTs). Unlike Fourier transforms, the continuous wavelet transforms have localization in the only frequency domain. Since the wavelets are localized in both the frequency and time domain, the wavelets have to be shifted to transform the whole space in time. Thus, the coefficients that are used in this are defined in the frequency and space, while the coefficients of the Fourier Transform (FT) are defined on the frequency axis.

$$C(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi\left(\frac{t-b}{a}\right) y(t)dt$$

The signal can be rebuilt perfectly from these coefficients with a double integration on time and frequency axis. Thus, the coefficients are highly redundant.

1.11 Discrete Wavelet Transforms

In mathematical term, the Discrete Wavelet Transform (DWT) is any wavelet transform, for which the wavelets are discretely sampled. The discrete wavelet transform has an advantage over the Fourier transform in its temporal resolution, that is, the Fourier transform has localized only the frequency domain, but the discrete wavelet transform has localization in both the time and the frequency domain.

The diagram of the Discrete Wavelet Transform tree is displayed in Figure 1.7.

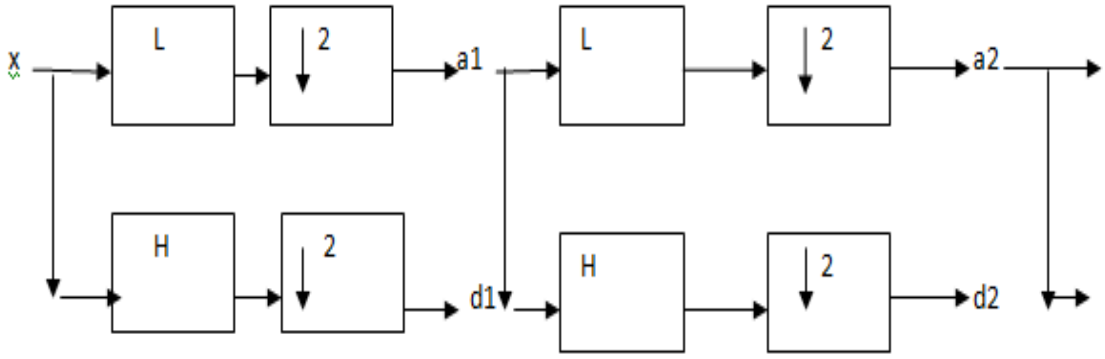


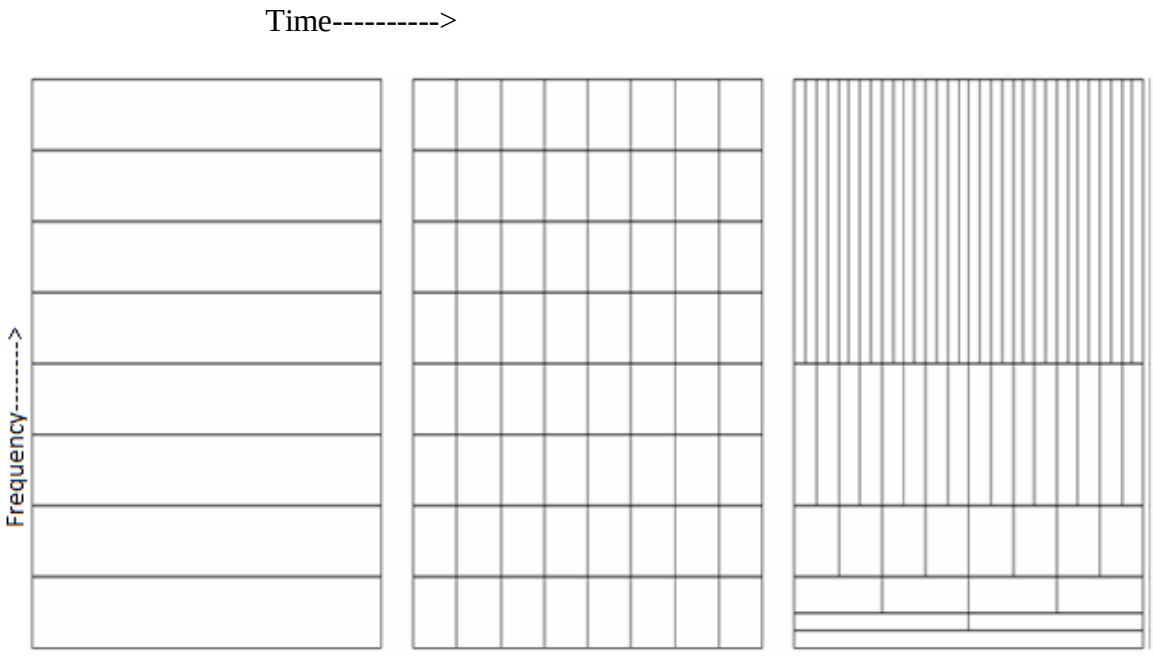
Figure 1.7: Discreet Wavelet Transform Tree

Here, H and L represent high and low-pass filters respectively, ↓ 2 represent sub-sampling. The discrete wavelet transform (DWT) is a linear transformation and is the process of transforming the data vector into a different vector numerically of the same length that operates on a data vector whose length is an integer power of two. The discrete wavelet transform is a method that separates data into dissimilar frequency components, and then studies each frequency component with a resolution matched to its scale.

The discrete wavelet transform is directly derived from the CWT. It consists of using only dyadic wavelets (contraction and dilatation of the mother wavelet by powers of 2) with a sampling of the coefficients:

$$C(j,k) = \sum\sum \boldsymbol{\psi}\left(\mathbf{2}^{-\boldsymbol{j}}\boldsymbol{n} - \boldsymbol{k}\right)\boldsymbol{y}(\boldsymbol{k}) * \mathbf{2}^{-\frac{\boldsymbol{j}}{\mathbf{2}}}$$

Figure 1.8 represents the time/frequency tiling with FFT, STFT and DWT.



Fourier Transform Windowed Fourier Transform Wavelet Transform

Figure 1.8: Time/Frequency Tiling with FFT, STFT and DWT

The discrete wavelet transform eliminates the redundancy of the continuous wavelet transform (CWT) coefficients, and has an optimal smallness (the reformation of the analog signal is possible only, if the condition for the sampling theorem is

satisfied). In this phrase, the Discrete Wavelet Transform (DWT) is very analogous to the Short Time Fourier Transform (STFT). The difference lies in that the wavelet transform uses the dilations of the function rather than tiling equivalently the frequency and time axis, the wavelet transform uses dilatation of the function so that we have high frequency resolution and low time resolution at low frequencies and opposite at the other end at low frequencies.

1.12 Wavelet Packet Transforms

A somewhat dissimilar algorithm is occasionally used in some music applications. To increase the information available in the highest part of the frequency axis (where the frequency resolution is the lowest), these sub-bands are also processed by a wavelet transform. The result is often sorted in a binary tree. Each level of the tree contains as many coefficients as the original signal.

1.13 Applications of Wavelet Transform

- Data compression
- Denoising
- Source and channel coding
- Biomedical
- Non-destructive evaluation

1.14 Limitations of Wavelet Transform

- Shift Sensitivity
- Poor Directionality
- Absence of Phase Information

In the next chapter, the literature on these topics has been reviewed briefly.

CHAPTER-2

LITERATURE SURVEY

The review of literature is all the important for a scientific study. It provides a background to the study being proposed. In this chapter, an effort has been made to identify the gap that lies in literature survey. The relevant studies undertaken for the discussion have been classified into the following sub-headings:

(i) Survey on Speech Recognition
(ii) Survey on Wavelets
(iii) Survey on Artificial Neural Network in Speech Recognition.

2.1 Survey on Speech Recognition

Gilbert (2001) intended that Automatic Speech Recognition (ASR) is an area of research, which deals with the recognition of speech by machine in several conditions. ASR performs well under restricted conditions, but performance degrades in noisy environments. This presents a brief survey on Automatic Speech Recognition on Malays Corpus and multi-modal speech recognition on other Corpus. The work on audio speech recognition was started only few decades back. After years of research and development the performance of automatic speech recognition remains one of the important research challenges (eg., variables of the context, database, and environment). The criteria for designing Speech Recognition systems are pre-processing filter, end-point detection, feature extraction techniques, speech classifiers, database, and performance evaluation. The main existing problem in Automatic Speech Recognition (ASR) is noisy environment. Therefore, it is necessary to design and develop a robust multi-modal speech recognition system.

Lie and Tung (2005) declared that speech compression is the area of digital signal processing that concentrates on reducing the bit rate of the speech signal for conduction and storage without a major loss of quality. Now - a days, a new technique, also called as a wavelet transform, has been projected for the analysis of

speech signals. This technique has been successfully implemented in many image compression applications. So far, no such consideration has been paid in the area of speech compression using wavelet. In this work, the researchers have evaluated the wavelet transformation technique based on the speech signals. Different wavelet filters have been used for the selection of the best filter that is effective for the analysis of the speech signal for providing a low bit rate. The algorithm also compared to the speech compression techniques such as Linear Predictive Coding (LPC). The Haar wavelet transform is an easy step to implement, and is also the fastest. However, because of the problem of discontinuty, it is not easy to replicate a continuous signal. As compared to Linear Predictive Coding (LPC) wavelet gives a higher signal to noise ratio (SNR) and better quality than LPC.

Lu and Wang (2003) planned an efficient and robust method to improve the performance of the speaker identification system to extract the features out of a speech signal that are also capable of operating in a noisy environment. The input speech signal is decayed into several frequency channels on the basis of multi-resolution characteristics of the wavelet transform. To capture the characteristics of a signal, the researchers calculated the Mel Frequency Cepstral Coefficient (MFCC) of the wavelet channels. For the recognition of the speech signals, as they gave good recognition for the speaker features as compared to the dynamic time wrapping (DTW), Hidden Markov Model was used. As the researchers compared the proposed method with the Mel Frequency Cepstral Coefficient (MFCC), it shows that the proposed method not only improves the recognition but also reduces the influence of noise. So, a robust and efficient feature extraction method has been presented for exploitation with the speaker identification system. On the basis of time frequency analysis of the wavelet transform, approximation changes were obtained.

Sunny (2012) proposed a method for the development of an effective and robust speaker recognition system. The pre-processing of the speech signal is an important step for speech recognition. In this work, different speech processing techniques has been used. The recognition accuracy has been achieved with the wavelet transform. Therefore, it shows that when wavelet transform is applied on to the predictable methods, the speech signal recognition increases by using the discrete wavelet transform, and also wavelet packets for noisy and clean speech signal respectively. So, the results show that by increasing the performance of the pre-processed speech

signals, the wavelet methods provide good results over conventional methods. This shows that feature extracted with the wavelet transform gives better accuracy for the recognition of words.

Blu (1993) planned a new element of feature vectors that use de-niosed wavelet coefficients for robust speech recognition. The better robustness to sudden spectrum distortion or to additive noise starts with more strong acoustic features. To get more important time and frequency features, the use of wavelet coefficients is forced by the capability of wavelet coefficients and by the human acoustic pre-processing modelling. The wavelet de-noising highlights the most important information for the speech signals, and adds more robustness. Now current speech recognition systems provide good results, when they are tested on data similar to that used for training. The speech recognition system presents the waveform of speech as a feature vector. The Cepstral coefficient is an older method for feature vectors, but it gives restricted results as seen by the complexity of the state of the art of the system for distortions and noise. So, the wavelet transform is an efficient approach for the analysis of speech signals and feature extraction of the speech signals.

Burrus (1997) proposed that the pitch detection is an efficient method for the analysis of speech processing and speech recognition. In this work, an algorithm known as pitch detection algorithm was created on the basis of second generation wavelet transform. This algorithm decreases the load of those algorithms, which are based on the classical wavelet transform. Then, the pitch detection algorithm that is developed was experienced for both synthetic speech signals and real signals. For the evaluation of accuracy of the speech signals and robustness of the algorithm, some experiments were conducted under noisy environment conditions. After evaluation, it was observed that the proposed algorithm was robust to noise and gave an accurate estimate for both the high-pitched speakers and low pitched speakers of the pitch period. Then several different wavelet filters were cosidered that were created on basis of second generation wavelet transform for the evaluation of the speech signals on the basis of proposed algorithm.Then it has been found that haar transforms give better performance as compared to the other wavelet filters. An evaluation of both the real speech signals and the synthetic speech signals has been done. The pitch periods of both synthetic and real speech signals were calculated accurately. Then the researchers used different wavelet filters such as haar, daubechies 6, daubechies 9-7 to

see the effect of different wavelet filters on the proposed pitch detection algorithm. It shows that the algorithm gives good performance with the haar transform.

Seok and Bae (1997) determine that the two most important methods of speech processing and speech recognition are speech segmentation and pitch determination. For the determination of the pitch period of the speech signals, the authors proposed a method that is based on the time based event detection method. On the basis of the discrete wavelet transform, the proposed method detects the voiced speech that is local in frequency, and also it determines the pitch periods. They further explained the pitch determination method on the basis of discrete wavelet transform and time based speech segmentation. As compared to other wavelet methods the method used by the researchers in their study for simplification of the wavelet analysis to a filtering function has less complexity. This method is also less expensive. By simulation and real speech signals computations it is clear that it is both robust to noise and accurate.

Hao and Zhu (2000) proposed a method to design and implement English digits speech recognition system using Matlab (GUI). This work was based on the Hidden Markov Model (HMM) which provides a highly reliable way for recognizing speech. The system is able to recognize the speech waveform by translating the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficients (MFCC) technique. It focuses on all English digits from zero through nine based on isolated words structure. Two modules were developed, namely the isolated words speech recognition and the continuous speech recognition. Both modules were tested in clean and noisy environments, and showed successful recognition rates. In clean environment and isolated words speech recognition module, the multi-speaker mode achieved an accuracy of 99.5% whereas the speaker- independent mode could achievee 79.5% accuracy only. However, in clean environment and continuous speech recognition module, the multi-speaker mode achieved an accuracy of 72.5% whereas the speaker-independent mode could achieve 56.25% accuracy only. Further, in noisy environment and isolated words speech recognition module, the multi-speaker mode achieved an accuracy rate of 88%, whereas the speaker-independent mode could achieve 67% accuracy only. In noisy environment and continuous speech recognition module, the multi-speaker mode achieved an accuracy of 82.5% whereas the speaker-independent mode could achieve 76.67% accuracy only. These recognition rates are relatively successful, if compared to other similar systems.

2.2 Survey on Wavelets

Farooq and Datt (1999) developed a speech recognition system by using two dissimilar feature extraction techniques; and a relative study was carried out for the recognition of the speaker independent spoken isolated words. The researchers followed the hybrid approach with the Linear Predictive Coding (LPC) and the Artificial Neural Network (ANN). The other method used by them was the combination of Discrete Wavelet Transform (DWT) and Artificial Neural Network (ANN). Voice signals were directly sampled from the microphone; and these were processed through the techniques used for extracting the features. Both the methods generated good recognition accuracy, but the discrete wavelet transform was found to be more appropriate for recognition of speech signals because of the multi-resolution properties and efficient time frequency localization. In the LPC technique, the input signals are broken into blocks or vframes. In the case of DWT, a collection of wavelets are presented for the analysis of signals.

Long and Datta (1996) characterised that the wavelet transform applications can be used for the analysis of speech signals. Some difficulties are formed in the speech signals, in the synthesis, and the analysis of signals. But the technique followed by the researchers to evaluate uses the wavelets for the analysis of speech signals and synthesis that distinguished the voiced or the unvoiced speech, and method for choosing the best possible wavelets for the speech signals. This relative observation gave results that are produced by listening to the speech signal using both the scalar or the vector quantized wavelet parameters. For the synthesis of voice, daubechie wavelet technique has been carried out. The scheme is verified using the simulation, and gives good results.

Strang (1996) believed that the dyadic wavelet transform is a useful tool for the processing of speech signals, but the poor frequency resolution (its low Q-factor) restricts the efficiency for the proceesing of oscillatory signals similar to speech or similar to the vibration measurements. For this, the researchers built up a more elastic family of the wavelet transform for which the frequency resolution varied. The wavelet transform was able to achieve higher Q-factor. It is an easily invertible constant Q-discrete wavelet transform that is to be implemented using the iterated filter banks. So, it is expected that the planned wavelet transform will provide an

effective way for the representation of a larger class of signals such as short time periodic / oscillatory signals, and other signals arising from the physical vibration phenomenon.

Kamarthi and Pittner (1997) planned a new method for extraction of feature vector consisting of Mel Frequency Discrete Wavelet Cofficients (MFDWCs). The MFDWCs are then produced by the Discrete Wavelet Transform (DWT) on the mel scaled filter banks of a speech frame. The main purpose for using the Discrete Wavelet Transform (DWT) is that, it is localized both in the time and frequency domain. The MFDWC is similar to that of multiresolution features and sub-band features. In that case, both multiresolution analysis and sub-band feature attempt to gain good time and frequency localization. The MFDWC has better time frequency localization as compared to the multiresolution features and sub-band features. The researchers also compared the performance of MFDWC with the multiresolution features and sub-band features and with the Mel Frequency Cepstral Coefficients (MFCC) and calculated the performance of new features for noisy and clean speech. So, it shows that the wavelet transforms have better time frequency localization.

Hao and Zhu (2002) proposed an approach to obtain feature vectors from the images based on their colour information, *i.e.*, content based retrieval. He used a new approach, namely, MPEG-7 with its corresponding framework MPEG-21 to capture features in the feature vectors. After obtaining the features, these were compared with the images present in the database. Principal Component Analysis, Moment Invariant technique and Wavelet Decomposition were used to capture the features as feature vectors from the image. Then the correlated images were found using those feature vectors. Different accuracies were achieved on changing the weight for the extraction coefficients and changing the threshold level. The results obtained finally proved that a good performance was achieved by the proposed system.

Lahmiri *et al.* (2007) proposed a new approach to capture features in the feature vectors using Hybrid DWT. It automatically obtained the vectors from the mammograms. This approach uses Gabor Filter along with DWT. It has a good scope in the screening system of cancer. Two-dimensional DWT was used to get

coefficients as features in the feature vector from the mammogram. After that Gabor was used to calculate the standard deviation and average from the image obtained.

Shi Yunhui *et al.* (2001) proposed a new type of continuous wavelet transform. For the processing and representation of the speech signals, there are some applications of wavelet transform for the analysis, compression, synthesis, and classification of speech signals. The most efficient feature for the analysis of speech signals is their non-statioanry character. Traditionally, we use Short Time Fourier Transform (STFT) for the processing of speech signals for a long time. But these days wavelet transform is a good alternative for the processing of speech signals. The wavelet transform has localization in both time and frequency domain as compared to traditional spectral basis. This property of wavelet transform is very useful for the processing of speech signals. There are some usual tasks for speech processing such as pitch detection, modelling of speech signals and their synthesis, extraction of feature vector in speech processing, decomposition of signal into different parts equivalent to phonemes and sounds, compressions and reconstructions of signals, and recognition of various speech signals. There are many applications that are used for the analysis of speech signals such as image processing, fingerprints, signal processing and so on. So, it is clear that wavelet transform is a good alternative for the analysis of speech signals as it has localization both in time and frequency domain . Further, the most common feature for the analysis of speech signal is their non-stationary character.

Yang *et al.* (2006) discussed an approach to detect contours and obtain feature vectors based on them for surface or shape matching. Features were obtained from the point clouds using this approach. They are nothing but the surfaces with high complexity. They contained surfaces with branch points, blends and contours that are open. The features to match the shapes were angle of rotation and the length of the arc. The analysis in the end showed that the method proposed gave optimal results in segmentation of the surfaces with huge complexity.

Shensa (1992) introduced a new concept of the wavelet transform and the representation of the Fourier transform. Such a wavelet transform uses a fully scalable modulated window, but not all possible shifts. A geometrical locus of frequency-time

point for the wavelet transform is derived, and examples are given. The locus is considered optimal for the Fourier transform, when a signal can be recovered by using only values of its wavelet transform defined on the locus. The inverse Fourier transform is also represented by the wavelet transform defined on specific points in the time-frequency plane. The concept of the wavelet transform can be extended for representation of other unitary transforms. Such an example for the Hartley transform is described, and the reconstruction formula is given.

Kadambe and Bartels (1992) addressed the difficulty of speech recognition on the basis of wavelet transform from a speech signal as a way to help match phonemes. As a means of comparison, this work uses a pattern of pre-recorded wavelet transforms. The application explained that how wavelets can be used for good accuracy of the speech recognition. In this, the wavelet transform is used to extract the coefficient and for the classification of the phonemes, it uses cross correlation. The silence is also removed from the signals. The researchers used daubechies 8 wavelet to attain the five octave of the signals. The results show that how wavelets can be used to obtain better accuracy of the phonemes.

2.3 Survey on Artificial Neural Network in Speech Recognition

Polur *et al.* (2000) gave an approach of the Artificial Neural Networks which is used as a research tool to accomplish Automated Speech Recognition of normal speech. A small size vocabulary containing the words YES and NO is chosen. Spectral features using cepstral analysis are extracted per frame and imported to a feed-forward neural network which uses a backpropagation with momentum training algorithm. The network is trained to recognize and classify the incoming words into the respective categories. The output from the neural network is loaded into a pattern search function, which matches the input sequence with a set of target word patterns. The level of variability in input speech patterns limits the vocabulary and affects the reliability of the network. The results from the first stage of this work are satisfactory and, thus, the application of artificial neural networks in conjunction with cepstral analysis in isolated word recognition holds promise.

Jain *et al.* (1996) proposed a method for speech and image processing to conserve the feature. The researchers used continuous wavelet transform (CWT) and discrete

wavelet transform along with the Artificial Neural Network (ANN) to accomplish the automatic speech recognition. Speech is a time varying signal. The most interesting thing about the speech signal is that the information enclosed in the speech signal is very difficult to analyze. Traditional methods that we use to analyse the speech signals are Mel Frequency Cepstral Cofficient (MFCC) and Short Time Fourier Transform (STFT) for the extraction of feature vectors out of speech signals; and the model is successfully implemented in several speech recognition systems. The performance of the speech signals depends on the context; and the results are good only if the speech signals are of short durations. The researchers revealed that the wavelet transform has better time frequency localization than short time frequency transform (STFT). They used the concept of wavelet transform for the extraction of feature vector out of speech signals. After that, they took speech signals of small duration and all the speech signals were fundamental unit of speech (phonemes). For the performance of feature extraction, they used the neural network for the classification of speech signals. So, wavelet transform is an efficient method for the feature extraction out of the speech signals. It has better time frequency localization as compared to short time Fourier transform which makes it good candidates for speech recognition.

Wu and Du (1996) provide a more effective method for the representation of the speech signals with the application of wavelet transform. They present an efficient and robust method for the feature extractions of the speech processing. Then the input speech signal is decayed into several frequency channels on the basis of the time frequency multi-resolution characterstics of the wavelet transform. The major problems regarding the design of this wavelet based on the speech recognition system are to choose decomposition level in the discrete wavelet transform (DWT), to choose best valuable wavelet for speech signals, and to select the feature vector from the wavelet coefficients. Now, in particular, the automatic classification of several speech signals is described using the discrete wavelet transform (DWT) and then compared with various wavelets. Then, the feature extraction technique based on the different wavelets and its performance is investigated on the isolated word recognition system. If the feature vectors are extracted properly then the features obtained on the basis of the wavelet transform show greater recognition rate. The wavelets have both weakness and strength for the identification of the feature of the speech signals. Also,

neural network classifier improves the performance of the recognition system extensively. It is clear that the wavelet transform is an important method for the extraction of feature vectors.

Chee pang Lim *et al.* (2001) proposed an approach to a new innovative voice web network. The voice web requires a voice recognition and authentication system incorporating a reliable speech recognition technique for secured information access across the Internet. In line with this requirement, the researchers investigated the applicability of artificial neural networks for speech recognition. In their experiment, a total number of 200 vowel signals from individuals with different gender and races were recorded. The filtering process was performed using the wavelet approach to de-noise and to compress the speech signals. An artificial neural network, especially the Probabilistic Neural Network model, was then employed to recognize and to classify vowel signals into the respective categories. A series of parameter settings for the PNN model were investigated, and the results obtained were analyzed and discussed.

Shing-Tai Pan *et al.* (2006) used Artificial Neural Network (ANN) to recognize speech. They applied Genetic algorithm (GA) to replace the Steepest Descent Method (SDM) for the training of Back Propagation Neural Network (BPNN). Thus, the performance of speech recognition was improved by the proposed method in this paper. The non-specific speaker recognition was trained by SDM. The recognition rate achieved in this experiment was up to 91%. This shows that if BPNN is trained by genetic algorithm, higher recognition rate can be attained**.**

Chugh *et al.* (2002) proposed a methodology for speech recognition with speaker identification based on Hidden Markov Model. Within their research acquisition of speech signal, analysis of spectrogram, neutralization, and extraction of features for recognition, mapping of speech using Artificial Neural networks is presented. In their investigation, the researchers realized a method of mapping by using back propagation rules of neural networks. This algorithm is especially suitable for huge set of input and output speech mapping**.**

Athulya (2004) discussed a novel technique for recognition of the isolated question words from Malayalam (one of the south Indian languages) speech query. Most of the research works in Information Extraction focus only on written language processing, in which a few are devoted to the study of Spoken Language Information Extraction.

They created and analyzed a database consisting of 250 isolated question words. Discrete Wavelet Transform (DWT) is used for the purpose of feature extraction; and Artificial Neural Network (ANN) is used for classification and recognition. A recognition accuracy of 80% could be achieved from this experiment.

Maaly (2006) developed a method for the recognition of Arabic sounds using Artificial Neural Networks. Most of the researches on speech recognition (SR) are based on Hidden Markov Models (HMM). Despite the fact that Arabic is a language that is spoken by millions of people, and it is the sixth spoken language in the world. There has been a scarcity of research in Arabic language recognition. Speech recognition systems will be more used as they have started to replace some of the functions normally accomplished with a keyboard. These and many other reasons encouraged the researchers to continue in this field.

Mark *et al.* (2010) presented two different artificial neural network approaches for phoneme recognition for text-to-speech applications. The first approach is Staged Backpropagation Neural Networks; and the second approach is Self- Organizing Maps. Several current commercial approaches rely on an exhaustive dictionary approach for text-to-phoneme conversion. Applying neural networks for phoneme mapping for text-to-speech conversion creates a fast distributed recognition engine. This engine not only supports the mapping of missing words on the database, but it can also mitigate contradictions related to different pronunciations for the same word. The ANNs, presented in this work were trained based on the 2000 most common words in American English. Performance metrics for the 5000, 7000 and 10000 most common words in English were also estimated to test the robustness of these neural networks.

CHAPTER-3

DATA COLLECTION & FEATURE EXTRACTION

Data collection and computation of features are two very important phases required before recognition phase in continuous speech recognition. Section 3.1 describes the method used for data collection, while Section 3.2 discusses algorithm to compute features. It may be noted that these two methods are not dependent on each other and may be planned together as the performance of the methods used in these stages affects the overall classification rate of the algorithm.

3.1 Data Collection Phase

Speech, in purely physical terms, is nothing but a longitudinal wave which is transmitted as a pressure variation between low and high pressure with the rate of pressure variation from low to high, to low again, determining the frequency. The degree of pressure variation (namely, the difference between the high and the low) determines the amplitude. So, speech is normal and best handled when stored as a vector of samples, with each individual value being a double-precision floating point number. A sampled sound can be completely specified by the sequence of these numbers plus one other item of information: the sample rate.

The data collection is the most important step in speech recognition. Only an efficient database can yield a good speech recognition system. As we know different people say words differently. This is due to the difference in the pitch, and slang pronunciation.

The total colea has been used in this work for the recording purpose. Colea requires the user to specify sampling frequency which provides the estimate that how many samples per second should be recorded, the duration of the recording, number of bits that should be used to store the sample value which can be either 8 or 16 bits depending on the requirement of storing the sample as floating point number or integer number and the filename. For the present work, recordings were done at the sampling frequency of 44100 with 16 number of bits used to store the sample value

and the recordings were saved in wave file format. We recorded 25 samples of the speech signals using the microphone of the words cat, bee, place, dip and see. If we plot these speech signals then a graph of number of samples vs amplitude is obtained which is shown in Figure 3.1 for the word cat.

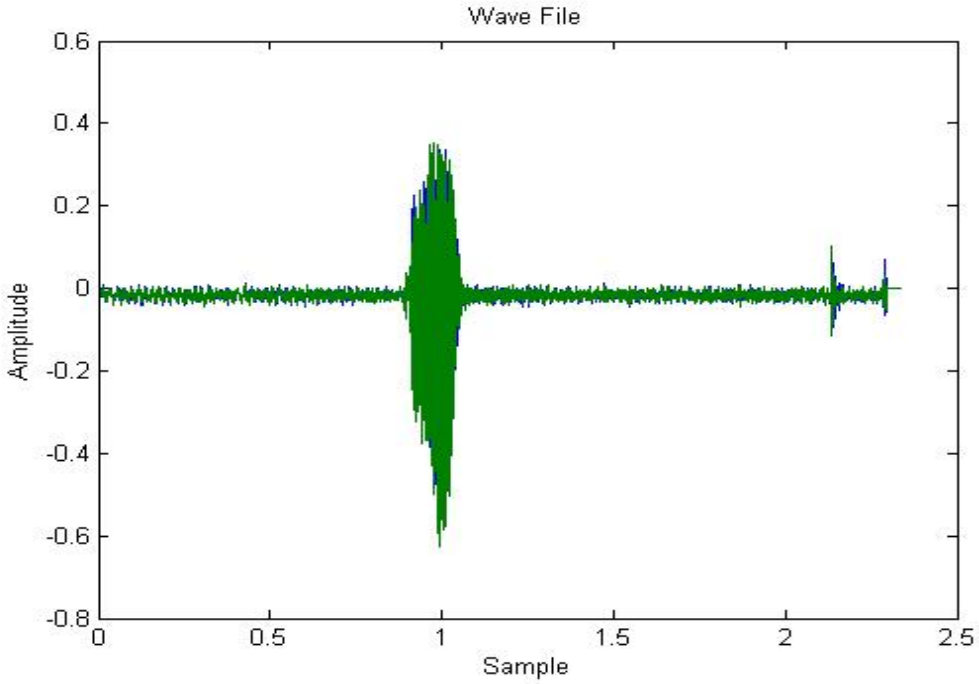


Figure 3.1: Waveform of Cat Plotted against Number of Samples and Amplitude

For this work, the sounds were recorded from males 25 times. The words were spoken 5 times by speakers using inbuilt microphone of the laptop in mild noisy conditions.

3.1.1 Decomposition of Speech Signals

The next step in the data collection is the speech decomposition. For this, we can use different techniques like Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficients (MFCC), Short Time Fourier Transform (STFT) and the wavelet transform. Over the last 10 years wavelet transform is largely used in speech recognition. Speech recognition models, generally, carry out some kind of specification/ recognition based upon the speech features which are usually obtained via time-frequency representations such as Short Time Fourier Transform (STFT) and Linear Predictive Coding (LPC) techniques. In some respects, these methods may not be suitable for representing speech; they assume signal stationary within a given time frame and may not have the ability to localize the event accurately. Furthermore, the

LPC approach assumes a particular linear model of speech production which strictly speaking is not the case.

The wavelet transform overcomes some of these limitations. It can provide a constant Q-analysis of a given signal by projection on to a set of basis functions that are scale variant with frequency. Each wavelet is a shifted scaled version of an original or mother wavelet. These families are orthogonal to each other.

The tiling of time-frequency plane via wavelet transform is shown in Figure 3.2.

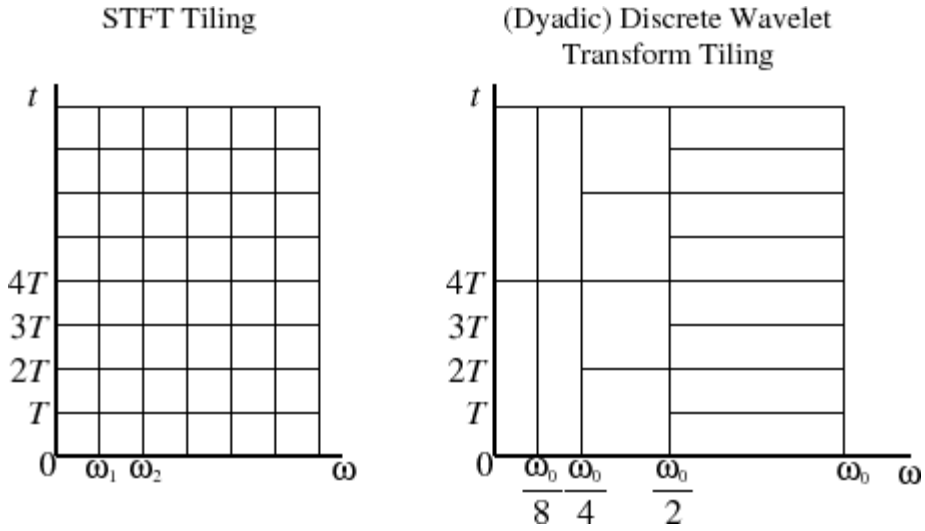


Figure 3.2: Tiling of the Time-Frequency Plane via Wavelet Transform

The wavelet transform can be applied to the non-stationary signals and has localizations to time and frequency domain. It concentrates into small portion of the signal which can be considered as stationary. It has got variable size window unlike constant size window in STFT. Wavelet transform (WT) gives us information about what bands of frequencies are there in a given interval of time.

3.2 Feature Extraction

Humans have the capacity to differentiate between different types of sounds. If we want a machine to identify the type of sound it will need some features or characteristics of speech on the basis of which it can differentiate different kinds of regions in the speech signal to identify whether it is voiced, unvoiced or silent. In feature extraction, speech is converted into a stream of feature vectors which contain only that information about the given utterance that is important for the correct

classification. These features are then stored in efficient sets of feature vectors. Features can be classified into two basic categories:

One are the temporal features which are easy to extract, simple and have easy physical interpretation like average energy, zero crossing rate, maximum amplitude, and maximum energy. These features are, generally, used during pre-processing and classification of speech signal like for silence removal, etc.

Another classification of features can be termed as spectral features. These features provide a lot of information about the spoken phone, that is why these are used for recognition of speech. For extracting these features; firstly, domain data is converted into frequency domain by applying Fourier transform on it and then spectral information or spectral features are extracted from this. MFCC, LPC, and WT power spectral analyses are the different features that appear under this category.

In tradition, to extract the feature out of the speech signal, we use methods like Short Time Fourier Transform (STFT) and Mel Frequency Cepstral Coefficients (MFCC); and the model has been successfully implemented in many speech recognition methods, but there is not so much accuracy by using these methods.

Then we came across the concept of the wavelet transform. By using the wavelet transform method, the classification became easier because it has better time frequency localization property as compared to the Short Time Fourier Transform (STFT) and the Mel Frequency Coefficients (MFCC). So, in order to get better feature extraction out of the speech signals, the wavelet transform is a good candidate for feature extractions.

In this project, neural network tool has been used for the classification of the phoneme against the various wavelets. To do this, the speech signals of small duration have been taken; and all the speech signals are phonemes (fundamental unit of speech). To measure the performance of the feature extraction, the processed speech signal vector was fed to the neural network for classification.

This project had the main problem of extracting the features from the speech signals against the various wavelets. However, this problem was solved by recording the audio samples (25 samples) using the microphone of the system, and then used the neural network tool for the phoneme classification and measured their performance against the various wavelets. The results are very interesting and require further experimentation which will be continued in future.

The figure below shows the classification of the phoneme using the Haar wavelet. Using the neural network, the classification becomes easier when the sampled data are farther from samples of different classes.

Figure 3.3 represents distribution of weights using Haar Wavelet coefficients as features.

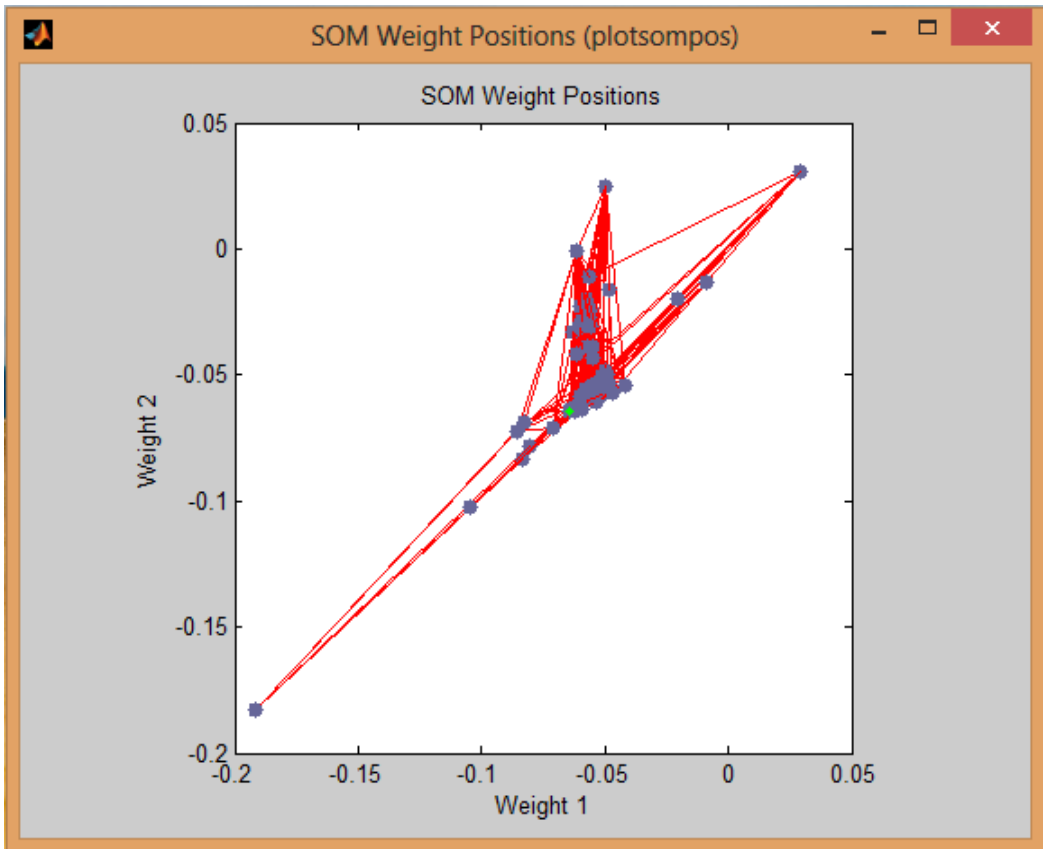


Figure 3.3: Classification with Haar Wavelet

In the case of haar wavelet, the sampled data are relatively farther from each other. So, the classification becomes easier in the case of haar wavelet. Different wavelets have been used for the classification of the phonemes. Thus it can be said that the wavelet transform is an important method for feature extraction out of the speech signals.

The artificial neural networks have been trained for the five phonemes considered in this work.

CHAPTER-4

IMPLEMENTATION & RESULTS

4.1 Experiments and Results

The very first thing that demands our attention is audio signal coming into the recognition engine. It contains not only the speech data, but also background noise. This noise can interfere with the recognition process; and the speech engine must handle the environment in which the audio is spoken.

This is the work of the speech recognition engine to convert spoken input into text. To do this, it employs all kinds of statistics, data and software algorithms. The very first job of the speech recognition engine is to process the audio signal i.e. , coming and convert it into a format that is best suited for further analysis. Then engine searches for the best match, if the speech data is in the proper format. It does this by taking into consideration the words and phrases it knows about. The knowledge of the environment is provided in the form of an acoustic model. Once it identifies the most likely match for what was said, it returns what it recognized as a text string.

The experiments conducted right from capturing the audio data till the classification for the purpose of present study can be elaborated in the following steps. MATLAB 7.0 has been used for the experiments and simulations.

1. First of all, the audio signal was captured. The microphone of the system was used to capture the audio samples. In all, 25 samples were taken which included 5 samples each for the letters /a/ as in cat, /b/ as in bee, /c/ as in place, /d/ as in dip, /e/ as in see.
2. Then started the process of normalizing the signal (used the MATLAB function normc () to normalize the vectors).
3. Then cleaned the Gaussian noise by using the thresholding and moving window approach to clean the Gaussian noise.

4. Then, Discrete Wavelet Transform was used. A 2- level decomposition was used; and then the scaling coefficients were retained for phoneme classification.
5. Before feeding the vectors to self-organizing feature maps, these were again normalized.
6. Then, the MATLB Self-Organizing Map (SOM) toolbox was used for the phoneme classification, and measured their performance against the different wavelets.

The experimental results were obtained by using the MATLAB 7.0. Firstly, the sound signal was recorded through the microphone. As many as 25 samples were recorded to get the results. Then, the audio file was converted into the .wav file. After reading the file the signal was normalized by using the normc () function. After that the white Gaussian noise was cleaned by using the thresholding and the moving window approach. Then, the 2-D level discrete wavelet transform was used for the phoneme classification, and again the signal was normalized before feeding it into the Self-Organizing Map (SOM). Then, the Self Organizing Map was used for the classification of phonemes; and their performance was checked with the different wavelets. Here, it is to be find out which wavelet is the best candidate for the feature extraction of the speech signals. So, different wavelets were used for the classification of the phonemes. Following are the five plots of the phonemes after cleaning the captured sound:

Plot for the Phoneme ‘a’

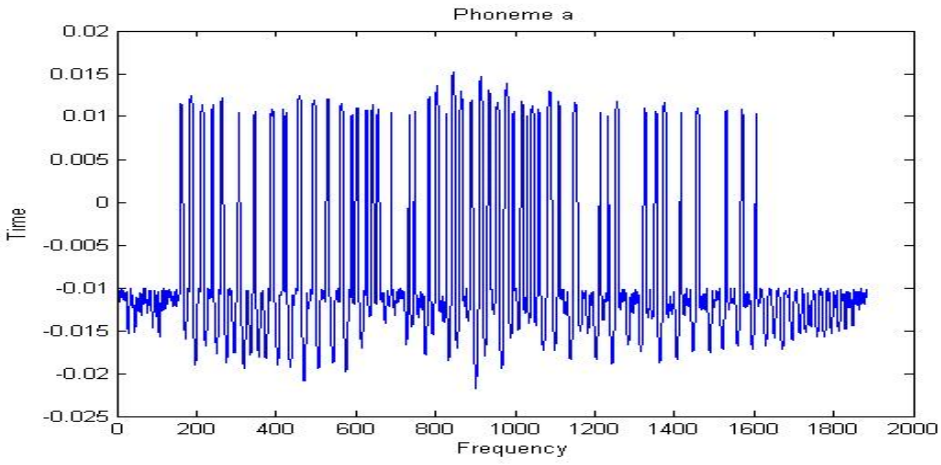


Figure 4.1: Plot of the Phoneme ‘a’ after Cleaning the Captured Sound

Figure 4.1 shows the plot for the phoneme ‘a’ after cleaning the captured sound. All the sound has been removed from the sound signal. The plot exhibited above has appeared after cleaning the white Gaussian noise.

Plot for the Phoneme ‘b’

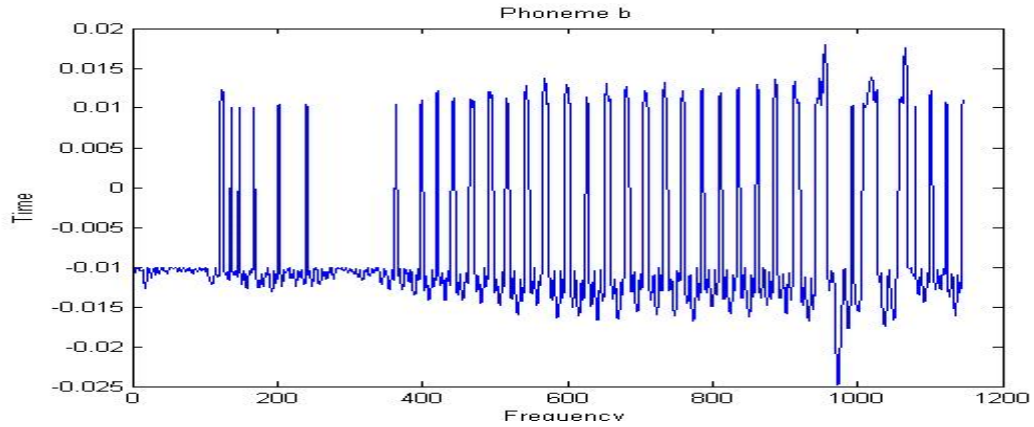


Figure 4.2: Plot for the Phoneme ‘b’ after Cleaning the Captured Sound

Figure 4.2 displays the plot for the phoneme ‘b’ after cleaning the captured sound. The plot shown above has appeared after removing the white Gaussian noise.

Plots for the Phoneme ‘c’

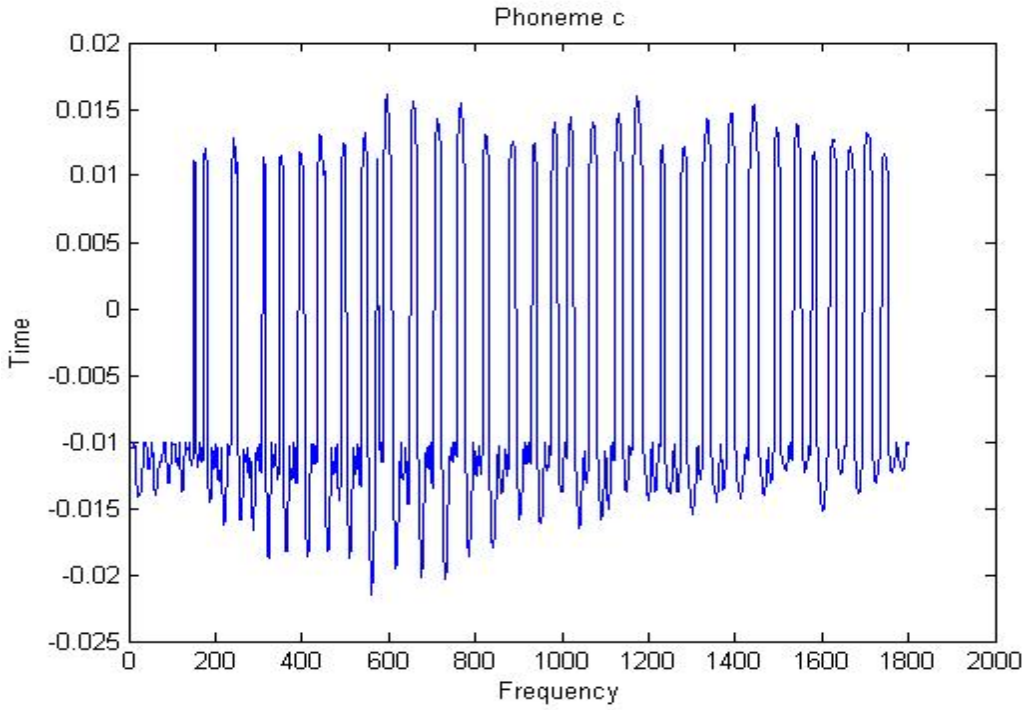


Figure 4.3: Plot for the Phoneme ‘c’ after Cleaning the Captured Sound

Figure 4.3 was presents the plot for the phoneme ‘c’ after cleaning the white Gaussian noise. All the noise was removed from the phoneme after capturing the sound signal.

Plots for the Phoneme ‘d’

The figure given below (Figure 4.4) provides the plot for the phoneme ‘d’ after removing the white Gaussian noise from the captured signal. The plot appearing in this figure has been is drawn between the frequency and the time domain. It has been drawn after completely removing the noise.

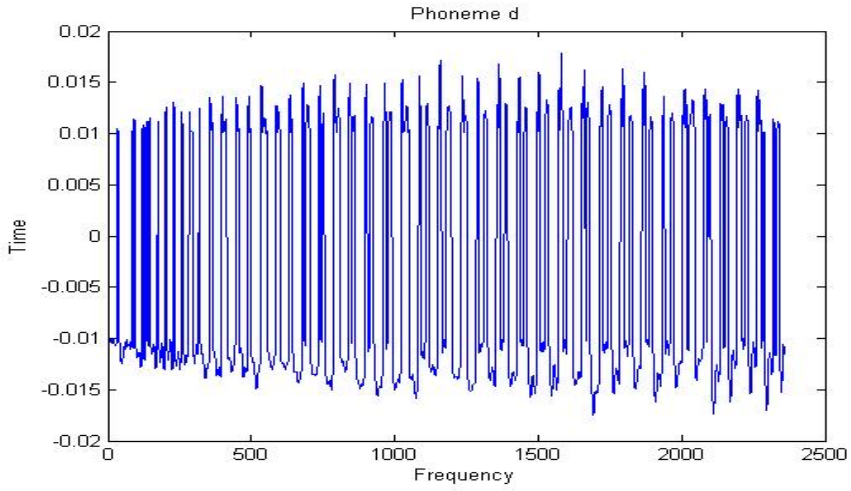


Figure 4.4: Plot for the Phoneme ‘d’ after Cleaning the Captured Sound

Plot for the Phoneme ‘e’

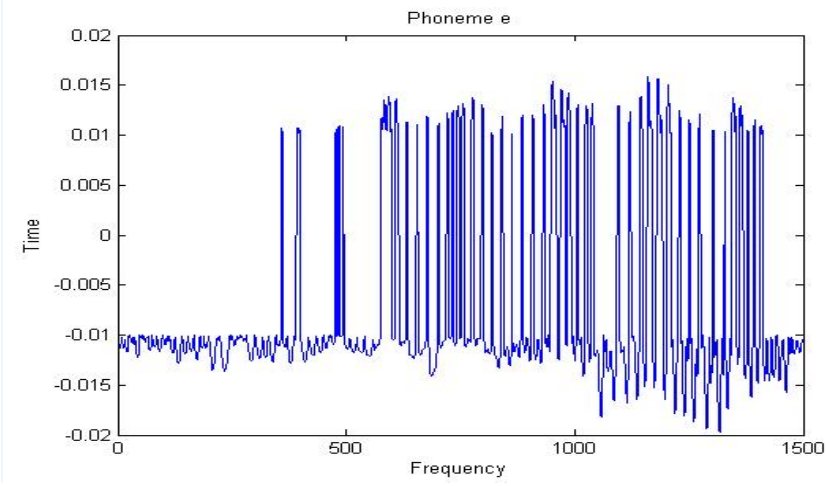


Figure 4.5: Plot for the Phoneme ‘e’ after Cleaning the Captured Sound

The figure given above carries the plot for the phoneme ‘e’ after removing the white Gaussian noise. The plot has been drawn between the frequency and the time domain after cleaning the noise from the phoneme.

Thus, it is evident that the plot for a phoneme should be drawn only after cleaning the white Gaussian noise. It helps to remove the noise from the entire phoneme.

4.2 Training Results

In this, with the various neural network topologies the sampled data was trained. There are 25 sampled data for which the network was trained for 5000 iterations to get the clustering. In this, the green dots refer to the sample data points and the blue dots refer to the neurons. MATLAB has been used for the analysis of entire data.

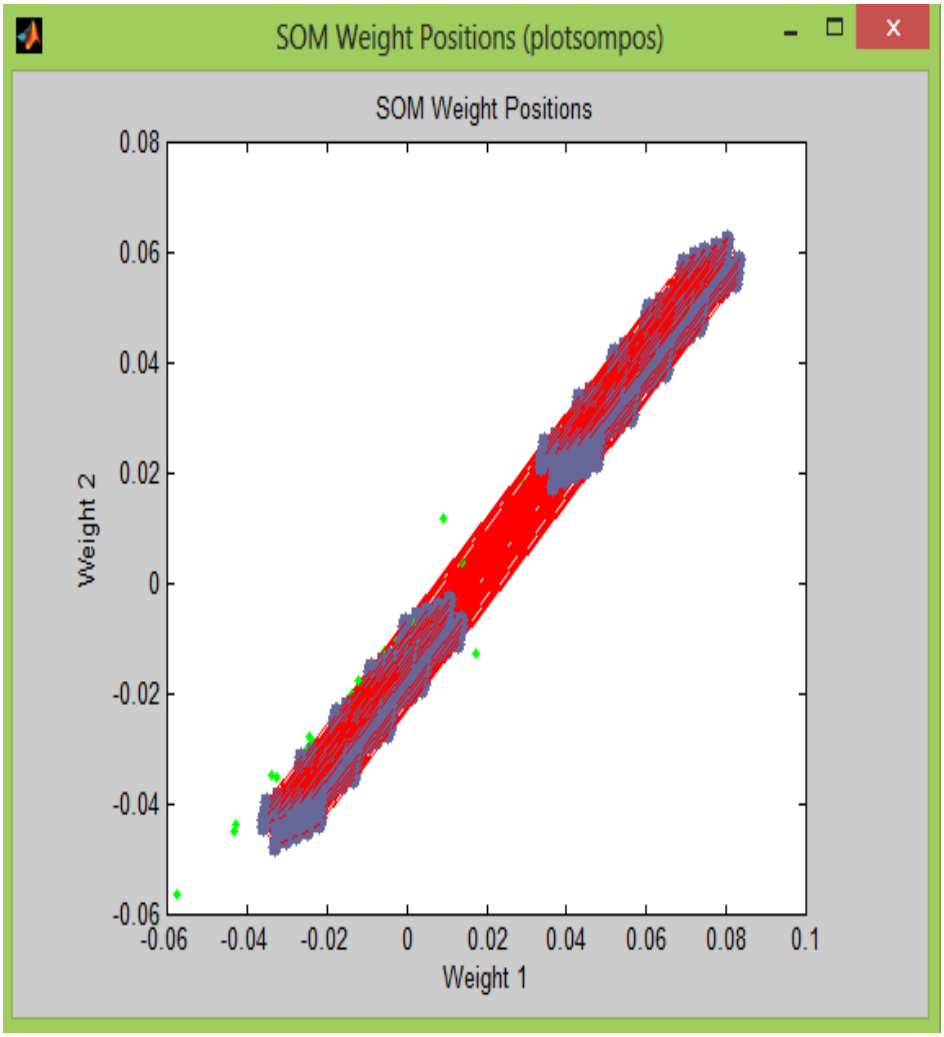


Figure 4.6: Initial Topology

Figure 4.6 highlights the initial topology of the wavelets without the classification of the neural network. In this figure, the blue dots represent the number of neurons and the green dots number of data samples captured. Thus, the graph shown above clearly presents the wavelets without the classification of neural network. Still the network topology without the classification depends on the dataset provided.

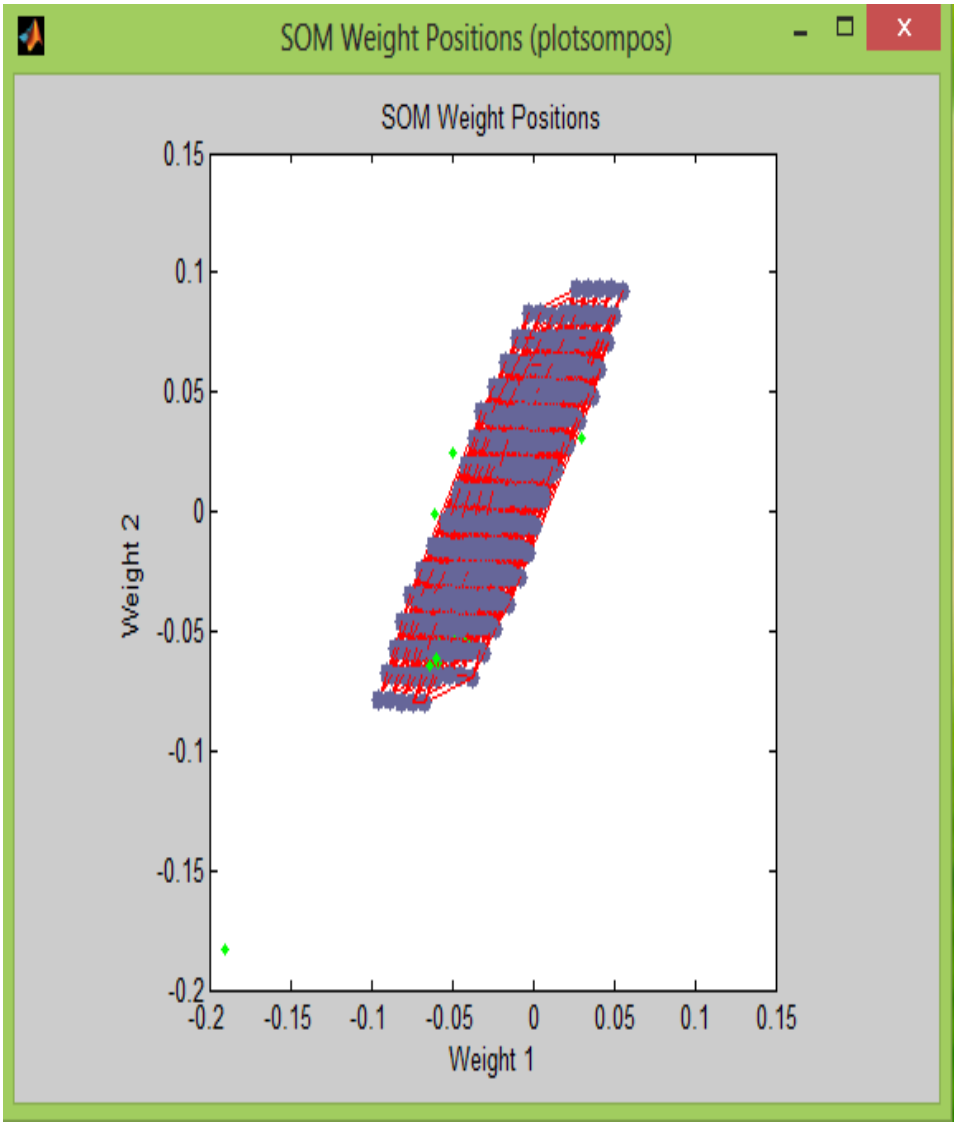


Figure 4.7: Classification with Haar Wavelets

The figure given above provides the classification of the haar wavelets. In this figure, the blue dots represent the number of neurons, while the green dots the number of data samples. It can be observed that a large number of data samples are farther from each other. It makes the classification easier for the neural network. In the case

of haar wavelets, the sample data are much farther from each other. So, the classification becomes easier in the case of haar wavelets. Thus, somewhere the haar wavelet is a good candidate for the feature extraction of the speech signal.

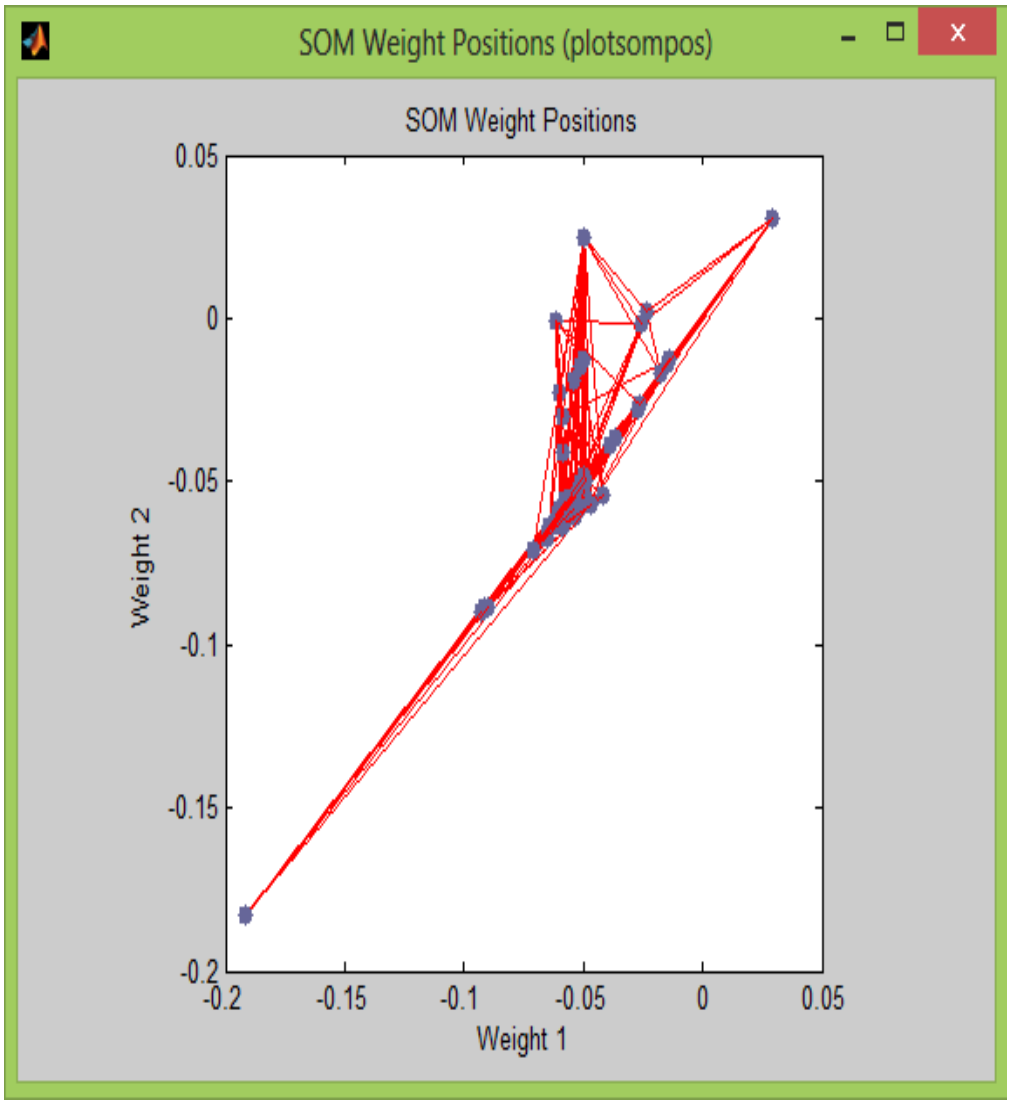


Figure 4.8: Classification of Haar Wavelet with SOM Toolbox

The figure given above presents the classification of haar wavelets with the Self-organization Map (SOM) toolbox. As already observed in the case of haar wavelets, the classification becomes easier as a number of data samples are farther from each other. The network topologies can be trained with Self-organizing Map toolbox. The SOM toolbox is used to train the sampled data that has been captured. So, the main purpose of this toolbox is to train the sampled data. Therefore, the

classification becomes easier for the neural network, if a good number of data sample are farther from each other.

4.3 Classification with db-4 Wavelets

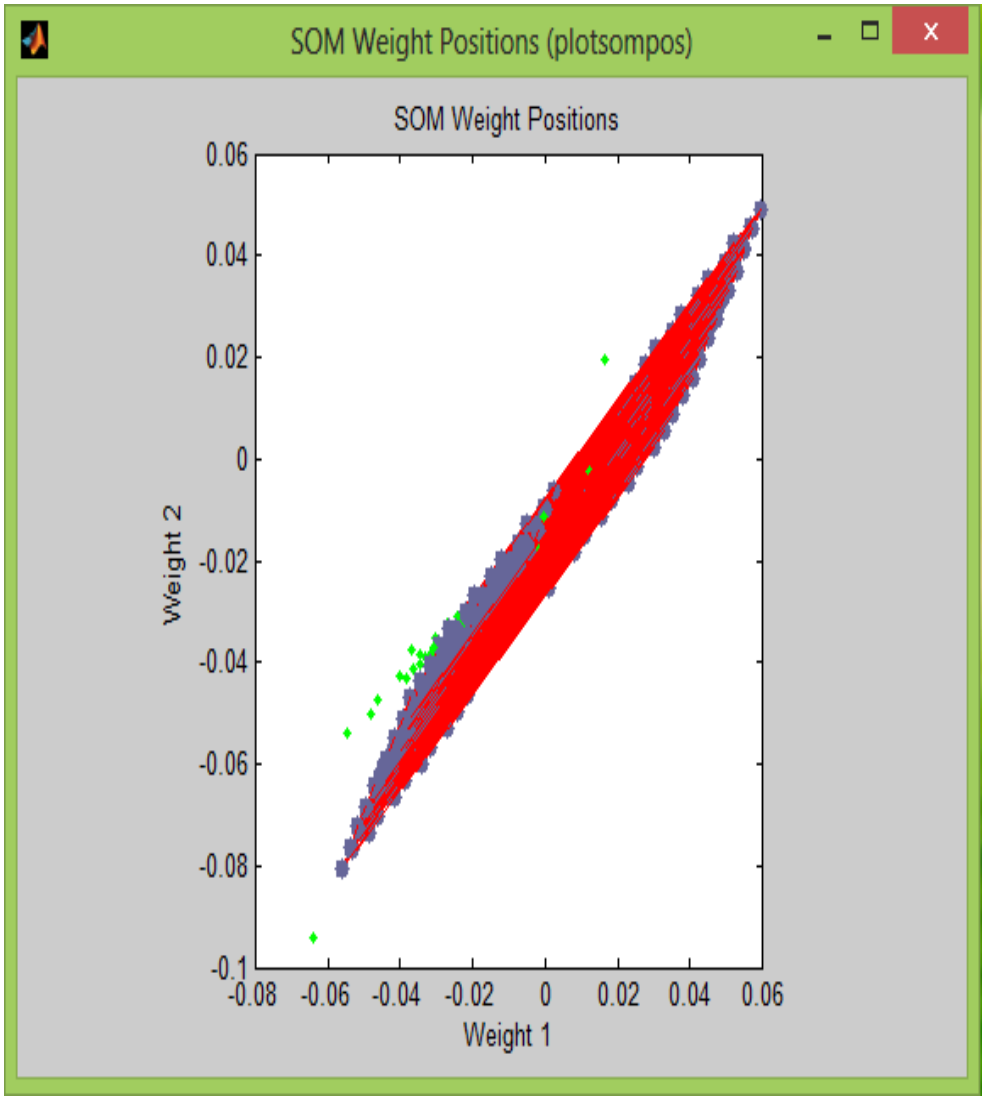


Figure 4.9: Classification with db 4 Wavelet

The figure shown above (Fig.4.9) provides the classification of daubechies 4 wavelets. We find that the sampled data are quite closer to each other. As, observed earlier also, if the number of data samples are closer to each other, then the classification is not easy for the neural network. So, the classification is not easy in the case of daubechies 4 wavelet for the neural network. Here, the green dots represent the sampled data, while the blue dots the number of neurons. In the case of

daubechies 4 wavelet, the number of sampled data is quite closer to each other. So, the classification becomes more difficult for the neural network.

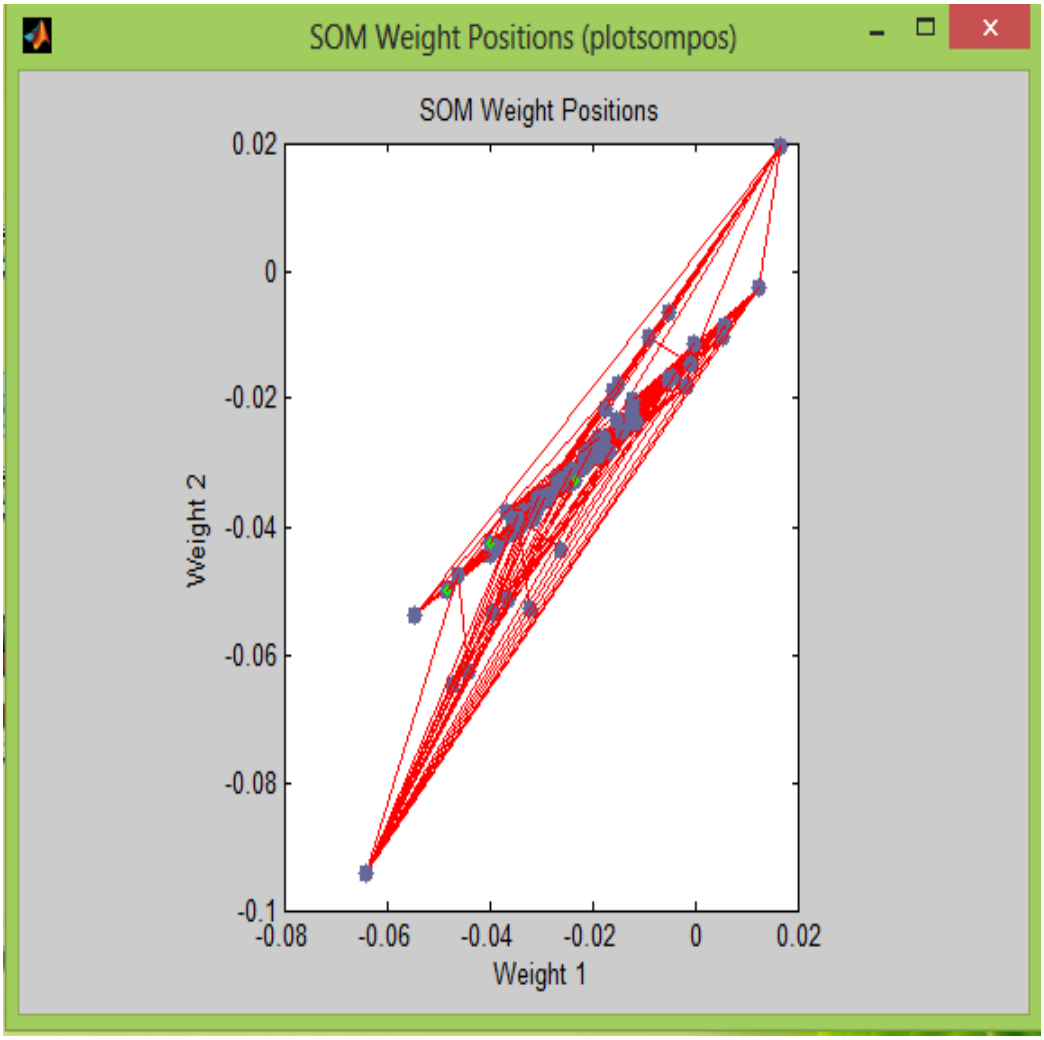


Figure 4.10: Classification of db 4 Wavelet with SOM Toolbox

The figure given above presents the classification of Daubechies 4 wavelets with the Self-organizing Map (SOM) toolbox. In this case, the classification is not easier because the sample data are closer to each other. In such a situation, the classification becomes more difficult for the neural network.

4.4 Classification with db-8 Wavelets

The figure given below presents the classification of daubechies 8 wavelets. It has been observed that a number of data samples in the case of daubechies 8 wavelets are closer to each other. Thus, in the case of the daubechies wavelets classification is not easy for the neural network.

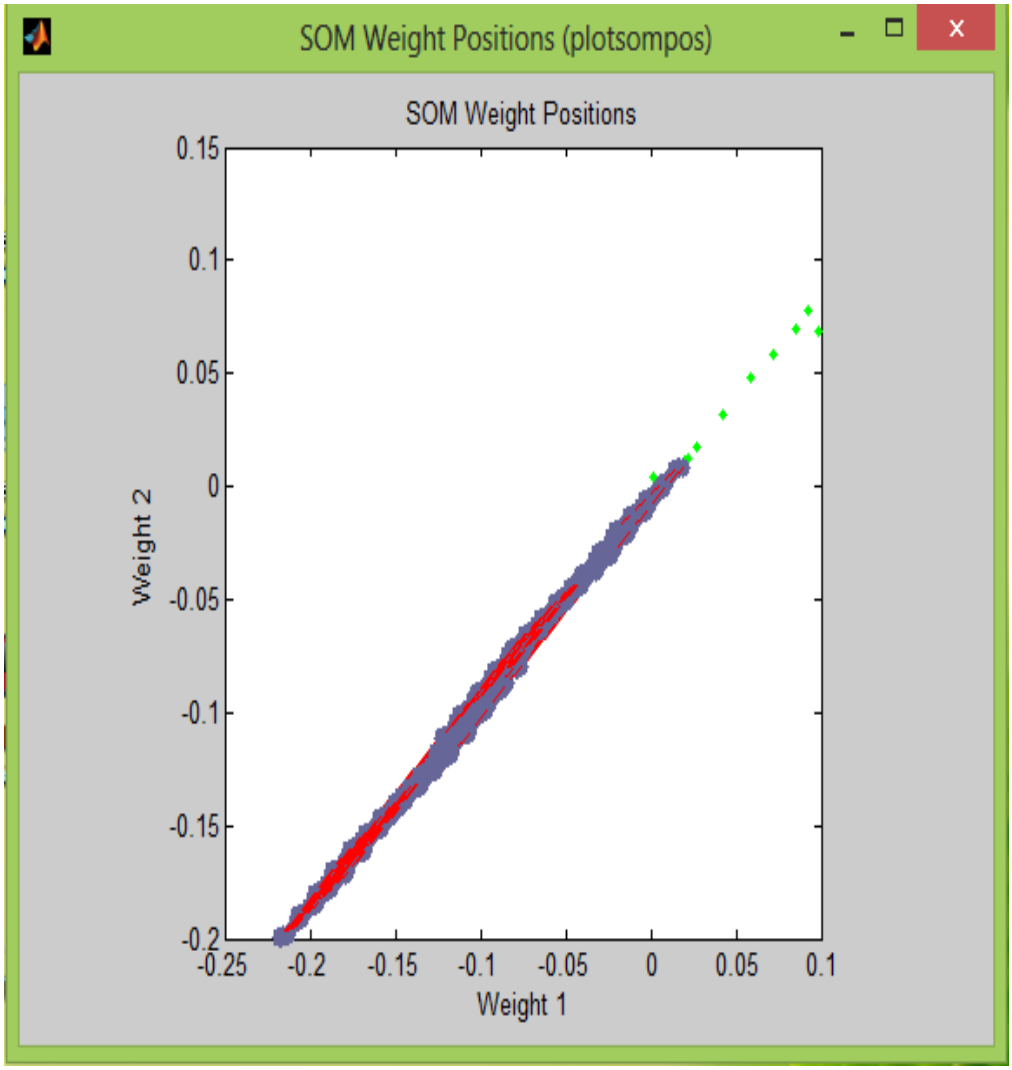


Figure 4.11: Classification with db 8 Wavelets

Figure 4.11 displays the classification of daubechies 8 wavelets. In this case, the classification is not easy because the sampled data are quite closer to each other. The green dots present the sampled data and the blue dots the number of neurons. It is clear from the figure that the numbers of data sample are quite closer to each other. It makes the classification more difficult in the case of db-8 wavelets for the neural network. Thus, the classification of daubechies 8 wavelets is not possible for the neural network.

4.5 Classification of the db-8 Wavelet with SOM Toolbox

The figure given below presents the classification of the daubechies 8 wavelets. In this case, the classifications for the neural network is not easy because of the number

of the data samples are closer to each other. The blue dots present the number of neurons and the green dots the numbers of the data samples.

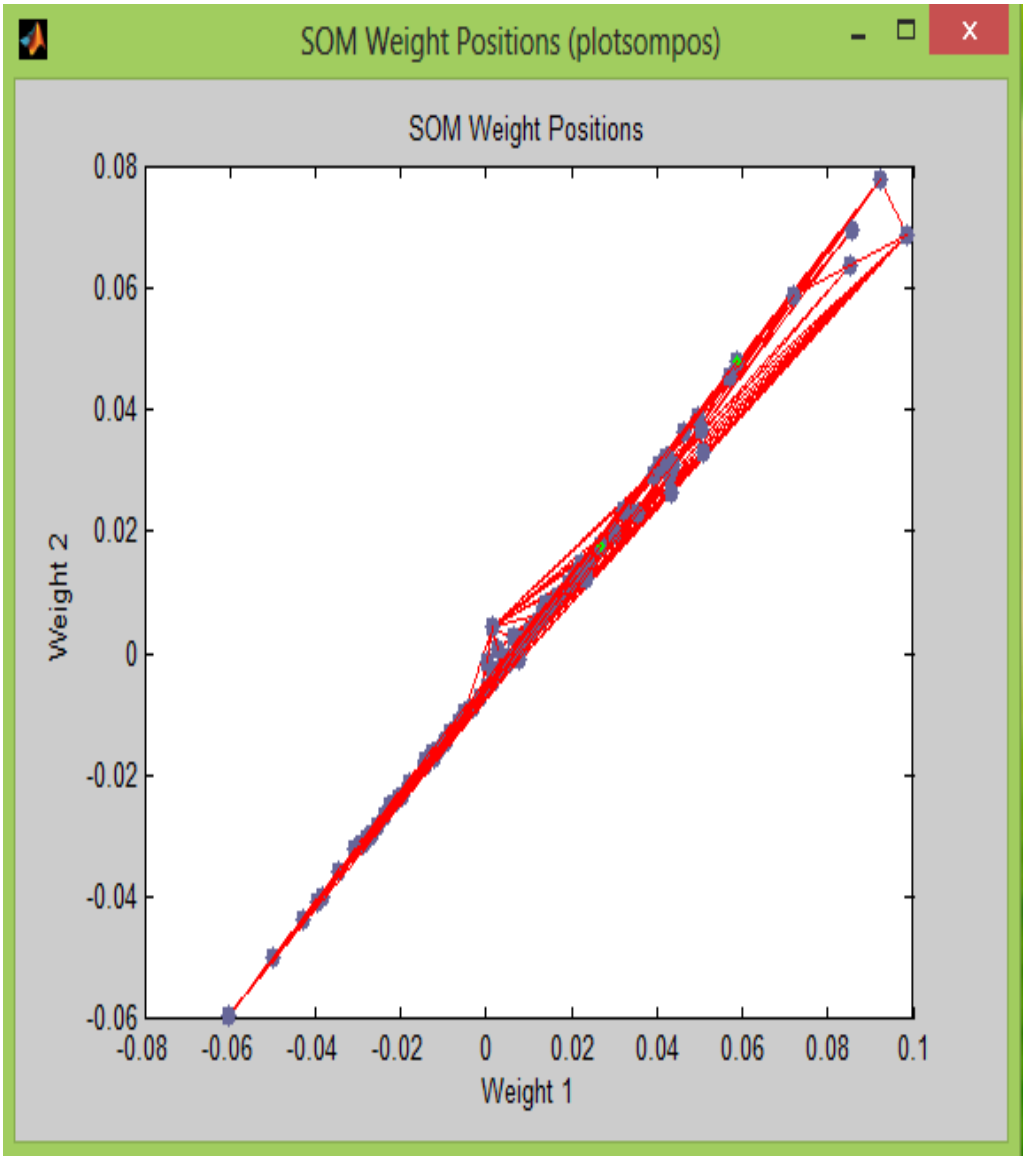


Figure 4.12: Classification of db-8 Wavelet with SOM Toolbox

In this case, the numbers of sampled data are quite closer to each other. Therefore, the classification becomes more difficult for the neural network.

4.6 Classification with bior-1.5 Wavelets

The figure given below presents the classification of the bior 1.5 wavelets. In the case of the bior-1.5 wavelet, the classification becomes somewhere easy because of the data samples are farther from each other. The blue dots represent the number of neurons and the green dots the number of data samples.

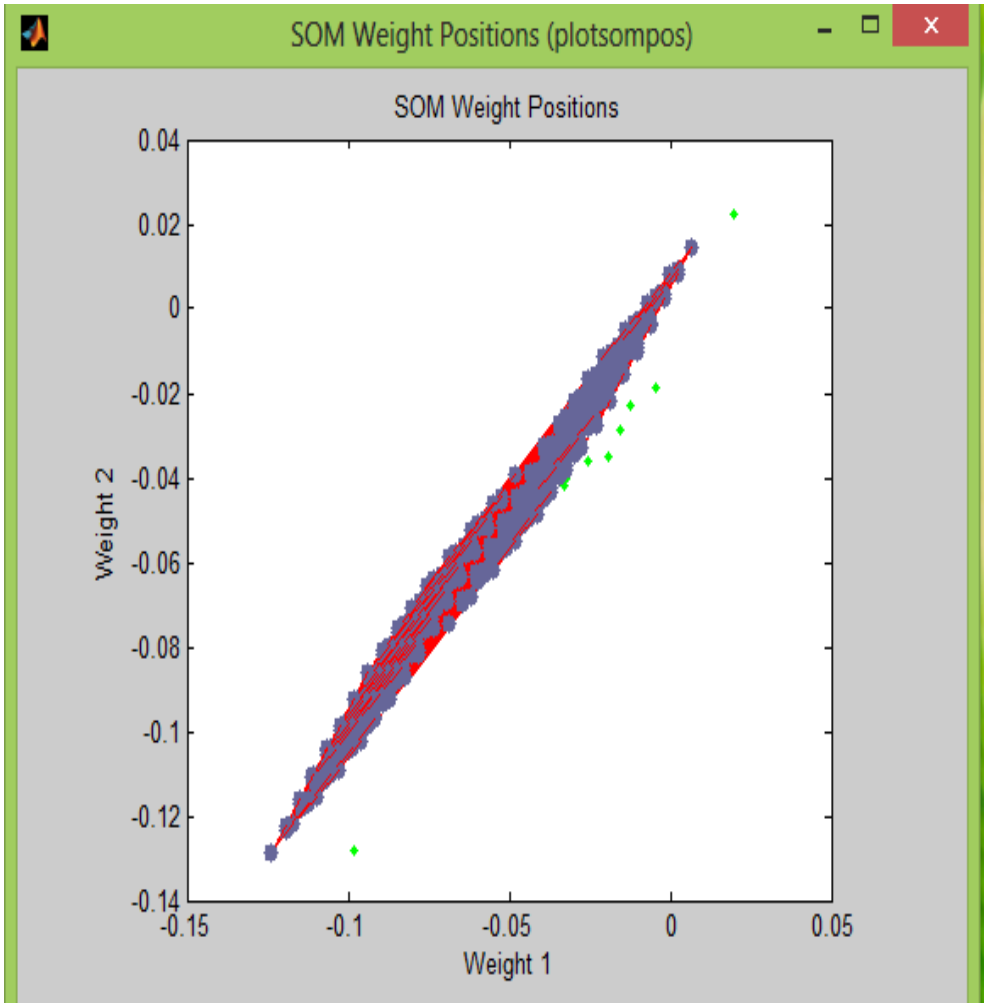


Figure 4.13: Classification of bior-1.5 Wavelets

4.7 Classification of the Bior-1.5 Wavelets with the SOM Toolbox

The figure given below presents the classification of the bior-1.5 wavelets with the Self-organizing Map (SOM) toolbox. The blue dots present the number of neurons and the green dots the number of the data samples. Thus, it is observed that if the numbers of the data samples are far from each other, then the classification becomes easier. In this case, the numbers of data samples are somewhere farther from each other. So, the classification becomes easier for the neural network.

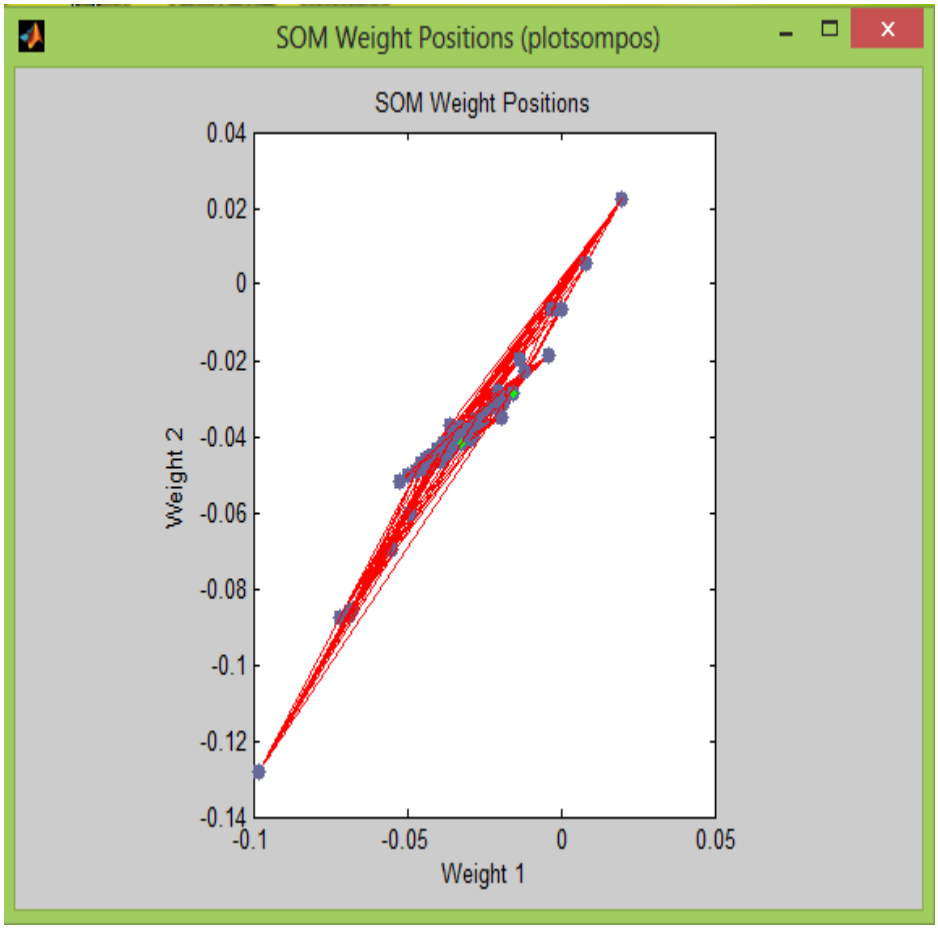


Figure 4.14: Classification of bior-1.5 Wavelet with SOM Toolbox

Thus, in all the wavelets except the Haar wavelet the sampled dataset are closer to each other. So, for the neural network the classification is not easy.

In the above figures, as the number of dataset elements are not so much I could not get proper clustering after the 5000 iterations. The performance of the neural network also varies with the scaling coefficients. As we observe that the sampled data in the case of the haar wavelet and the bior wavelet are farther from each other. When the sampled data are farther from each other than the classifications become easier for neural network and when the sampled data are closer to each other than the classification become more difficult for the neural network. This shows that the haar wavelet and the biorthogonal wavelets are a better candidate for the feature extraction of the speech signals.

CHAPTER-5

CONCLUSION & FUTURE SCOPE

5.1 CONCLUSION

Speech recognition is an area of research which has attracted a large number of researchers. Although most of them have contributed significantly in its development, yet there is enough scope for further research. Thus, the issuses like speech pre-processing, speech representation and feature extraction in the area of speech recognition demand our immediate attention.

The wavelet transform is a better method for feature extraction of the speech signals. The work undertaken in this project involves a comprehensive analysis of the feature extraction of the speech signals (phonemes) by using wavelets. The wavelet transform has better time frequency localization as compaerd to the Short Time Fourier Transform which makes it a better candidate for the feature extraction of the speech signals. A lot of research work is required to be done in the area of speech recognition. It is of utmost importance to decide which wavelet is best suitable for the feature extraction of the speech signals. It has been observed that analysis of the speech signals using different wavelets has provided us good results. Thus, the wavelet transform is a better choice for the analysis of speech signals as compared to Short Time Fourier Transform.

5.2 FUTURE SCOPE

The wavelet transform has proved itself to be the best selection for feature extraction of the speech signals. It has the localization in both the time domain and frequency domain. It is always important to decide which wavelet is best suitable for the feature extraction. A lot of research work has been carried out in the field under study, but still more needs to be done.

Like any other research, the present one has its own limitations which restrict its scope. Time and limited financial resources are the other contraints. It is obvious that this work being limited in its scope may have left room for further research in the

area. Other wavelets, namely, coiflets wavelets, symlets wavelets and morlets wavelets can further be experimented to extract features for better classifications.

APPENDIX

A.1 Neural Network

An Artificial Neural Network (ANN) is an information processing standard which is inspired by the biological nervous systems, such as the brain, process information. The main element of this paradigm is the novel structure of the information processing system. It consists of huge number of highly inter-related processing elements such as neurons working in unison to solve specific problems. An Artificial Neural Network (ANN) is configured for a definite application, such as data classification or pattern recognition through a learning process.

In order to run the complete simulations of the exacting circuits in the brain, computational neurobiologists have constructed very complex computer models of neurons. We are more interested in the specific properties of the neural network as a computer scientist, independent of how they are implemented in the brain.
The typical diagram of the artificial neuron is shown in the following figure.

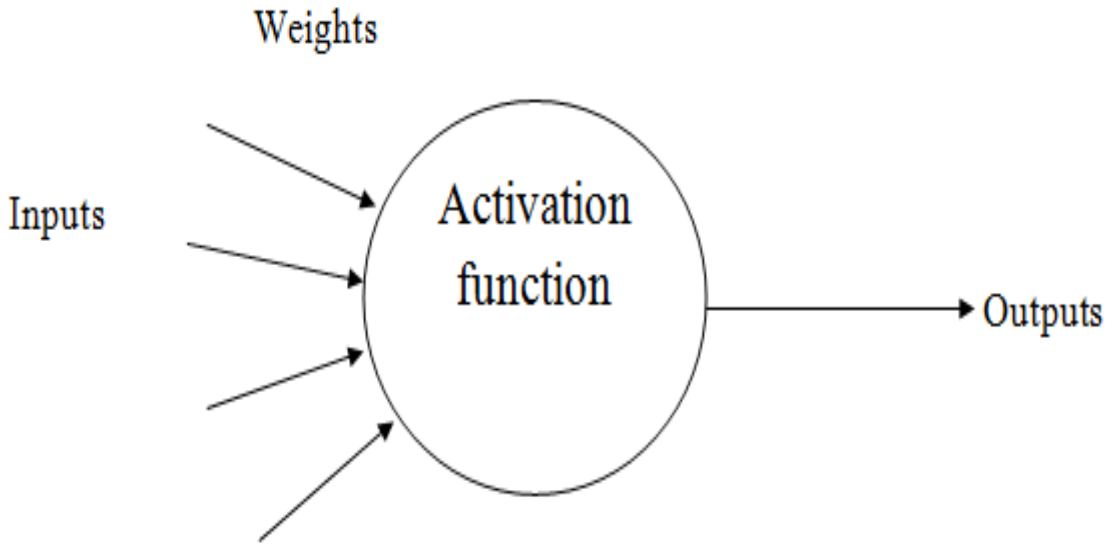


Figure A.1: Artificial Neuron

There are two types of learning in the neural network:

a) Supervised Learning (Back propagation network).
b) Unsupervised Learning (Self-organizing maps).

The diagram of the Neural Network is given in Figure A.2.

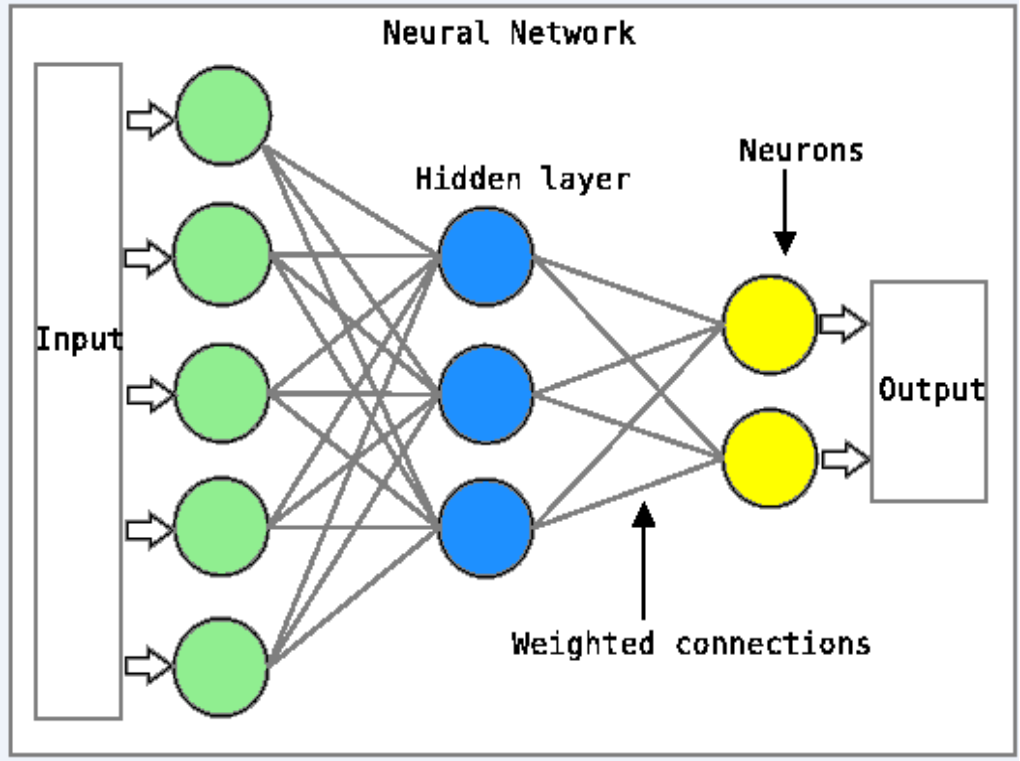


Figure A.2: Neural Network

A.2 Self-Organizing Feature Maps (SOFM or Kohonen's SOM)

The Kohonen's self-organizing map (SOM) is based on unsupervised competitive learning. Its main purpose is to transform an input pattern of arbitrary dimension in a one-dimensional or two-dimensional map in a topologically ordered fashion. The training algorithm for a SOM can be summarized in two stages: competition and cooperation. In the competition stage, a random input pattern is chosen as the similarity between this pattern and all the neurons of the network is calculated by the Euclidean distance, defined by:

id = argmin

Where, i = 1 ...k, and the index of the neuron with lowest distance is selected. In the cooperation stage, the synaptic weights that connect the winner neuron in the input pattern are updated. The weights of neurons neighboring the winner neuron are also updated by:

(t+1) = (t) + (t) h (t) (− (t)),

Where, t is the number of training iterations, (t+1) is the new weight vector, (t) is the current weight vector, h(t) is the neighborhood function and is the input pattern.

A.3 Algorithm for Kohonon's Self-Organizing Map

1. Suppose the output nodes are connected in the array (usually 1, 2 or 3 dimensional).
2. Suppose that the network is fully connected – all the nodes in the input layer are connected to all the nodes in the output layer.
3. Then, we use the competitive learning algorithm as follows:
   a) At random choose an input vector x.
   b) Then, determine the winning output node i, where w_i is the weight vector connecting the inputs to output node i.

      The above equation is equivalent to wi x >= wk x only if the weights are normalized.

$$|w_i \text{ -X}| \leq |w_k \text{ - X}| \quad \forall K$$

   c) Given the winning node i, the weight update is

$$W_K \text{ (new)} = W_K \text{ (old)} + \mu \aleph \text{ (i, k) (X-}W_K\text{)}$$

Where, $\aleph\,(i, k)$ is called the neighborhood function that has value 1 when i=k and falls off with the distance $|r_k - r_i|$ between units i and k in the output array. Thus, units close to the winner as well as the winner itself, have their weights updated appreciably. The weights associated with far away output nodes do not change significantly. It is here that the topological information is supplied. Nearby units receive similar updates, and thus, end up responding to nearby input patterns.

A.4 SOM Toolbox

The SOM (Self-Organizing Map) toolbox is a vector quantization method which places the prototype vector on a regular low-dimensional grid in an ordered fashion. This makes the SOM (Self-organizing Map) a powerful visualization tool. The SOM toolbox is an implementation of the SOM and its visualization in the Mat lab 5 computing environment. The SOM is also known as the Self-organizing Map or known as Self-Organizing Feature Map or Kohenen map; and it is neural network method based on the unsupervised learning method (Lu and Wang, 2003). This toolbox contains the functions for visualizations, creations and analysis of the Self organizing Map.

The SOM toolbox is a function package used for implementing the Self-Organizing Map (SOM) algorithm and more. Using the SOM toolbox we can do the following:

1. We can train SOM with different network topologies and learning parameters;
2. We can compute different error, quality and measures for the SOM;
3. We can visualize SOM using u-matrices, component planes, cluster color coding and color linking between the SOM and other visualization methods; and
4. We can do correlation and cluster analysis with SOM.

SOM toolbox also features other data analysis methods related to VQ, clustering, dimension reduction, and proximity preserving projections such as:

1. Data pre-processing tools.
2. K-means, K-nearest neighbor classifier and Learning Vector Quantizer (LVQ).
3. Agglomerative hierarchical clustering and dendrograms.
4. Principal component analysis (PCA).
5. Sammon's projection.
6. Curvilinear Component Analysis (CCA).

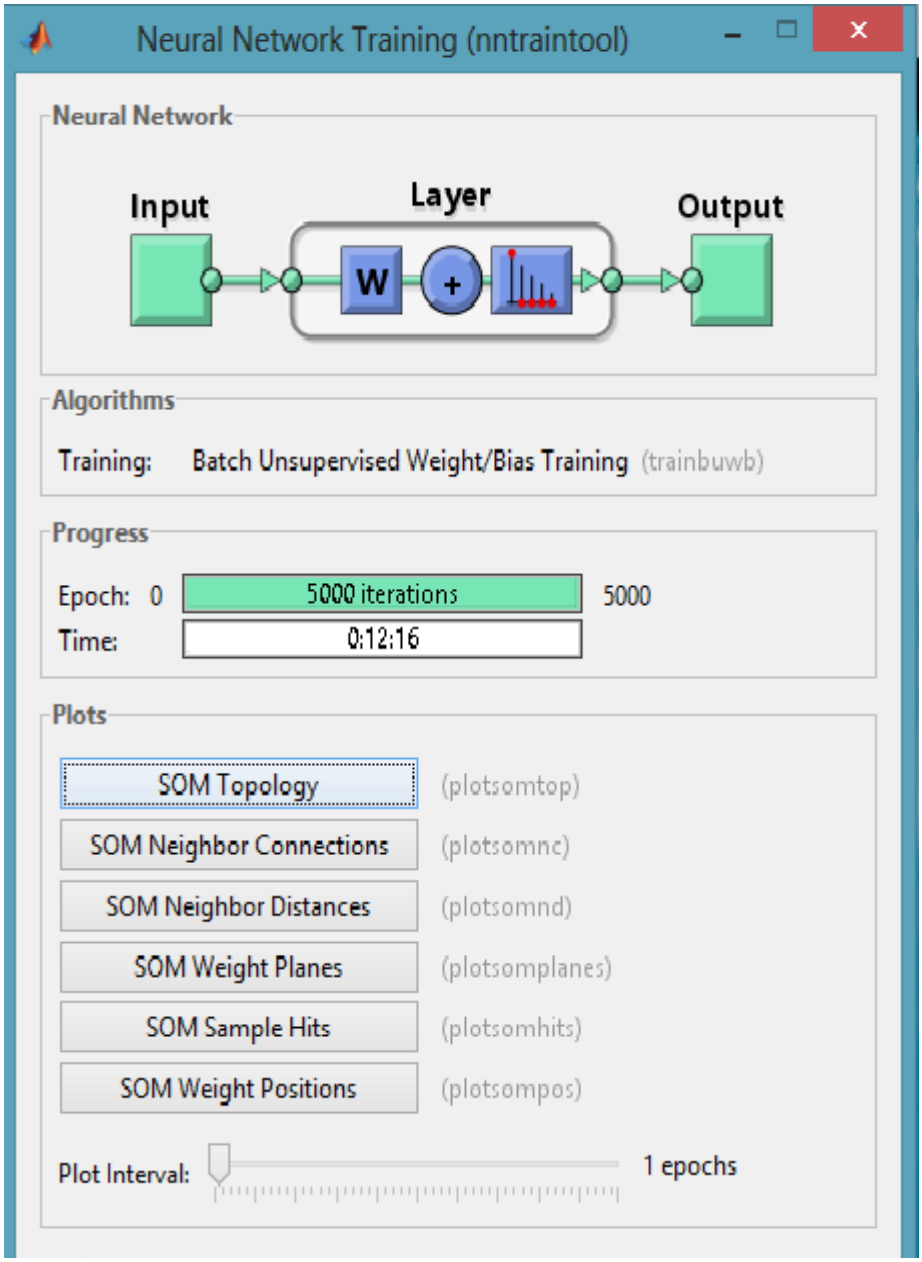


Figure A.3: SOM toolbox

The figure shown above presents the Self-organizing Map (SOM) toolbox that is used for the classification of the neural network. The SOM toolbox is a library package and is used for the implementation of the Self-organizing Map and more. Using the SOM toolbox, we can train the different network topologies and the learning parameter. In the present work, the SOM toolbox has been used for the classification of the neural network with different wavelets. Thus, this toolbox is mainly used for the classification of different network topologies.

REFERENCES

[1] Gupta, M.; and Gilbert, A. (2001), "Robust Speech Recognition using Wavelet Coefficient Features," *IEEE Proceeding of the Automatic Speech Recognition Understanding*, pp. 445-448.

[2] Lei, S. F.; and Tung, Y. K. (2005), "Speech enhancement for Non-stationary Noises by Wavelet Packet Transform and Adaptive Noise Estimation," *IEEE Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 41-44.

[3] Lu, C. T.; and Wang, H. C. (2003), "Enhancement of Single Channel Speech based on Masking Property and Wavelet Transform," *Speech Communication*, Vol. 41, No. 2-3, pp. 409-427.

[4] Sunny, S.; Peter, D.; and Jacob, K. P. (2012), "Recognition of Speech Signals: An Experimental Comparison of Linear Predictive Coding and Discrete Wavelet Transforms," *International Journal of Engineering Science*, Vol. 31. No. 4, pp. 1594-1601.

[5] Blu, T.; (1993), "Iterated Filter Banks with Rational Rate Changes-Connection with Discrete Wavelet Transforms." *IEEE Transaction on Signal Processing*, Vol. 41, pp. 3232-3244.

[6] Burrus, C. S.; Gopinath, R.; and Guo, H. (1997), "*Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice-Hall Press.

[7] Seok, J. W.; and Bae, K. S. (1997), "Speech Enhancement with Reduction of Noise Components in the Wavelet Domain," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1323-1326.

[8] Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia.

[9] Mallat, S. (2008), A *Wavelet Tour of Signal Processing*, Academic Press.

[10] Hao, Y.; and Zhu, X. (2000), "A New Feature in Speech Recognition based on Wavelet Transform," *IEEE 5th International Conference on Signal Processing*, Vol. 3. pp. 1526-1529.

[11] Farooq, O.; and Datt, S. (1999), "Wavelet Transforms for Dynamic Feature Extraction of Phonemes," *Acoustic Letter*, Vol. 23, pp. 79-82.

[12] Long, C. J.; and Datta, S. (1996), "Wavelet based Feature Extraction for Phoneme Recognition," *IEEE 4th International Conference on Speech and Language Processing*, Vol. 1, pp. 264-267.

[13] Farooq, O.; and Datta, S. (2001), "Mel filter-like Admissible Wavelet Packet for Speech Recognition," *IEEE Signal Processing Letters*, pp. 196-198.

[14] Strang, G. (1996), *Wavelets and Filter Banks*, Wellesley-Cambridge Press.

[15] Kamarthi, S. V.; and Pittner, S. (1997), "Fast Fourier and Wavelet Transform for Flank Wear Estimation – A Compression," *Mechanical Systems and Signal Processing*, Vol. 11, pp. 791-809.

[16] Kim, K.; and Youn, D. H.; and Lee, C. (1997), "Evaluation of Wavelet Filters for Speech Recognition," *IEEE International Workshop on Automatic Speech Recognition and Understanding*, Vol. 4, pp. 2891-2894.

[17] Obaidat, M. S.; and Lee, C.; and Sadoun, B.; and Neslon, D. (1999), " Estimation of Pitch Period of Speech Signal using a New Dyadic Wavelet Transform," *Journal of Information Sciences*, Vol. 119, pp. 21-39.

[18] Mallat, S. G. (1989), "Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Transaction on Pattern analysis Machine Intelligence*, Vol. 31, pp. 674-693.

[19] Joseph, S. M.; and Feroz Sha, A.; and Babu Anto, P. (2010), "Speech Compression: A Comparative Study between Discrete Wavelet and Wavelet Packet Decomposition," *International Journal of Computer and Network Security*, Vol. 2 No.7.

[20] Tan, B.T.; and Fu, M.; and Spray, A.; and Dermod, P. (1996), "The use of Wavelet Transform for Phoneme Recognition," *IEEE 4th International Conference on Spoken Language*, Vol. 4, pp. 2431-2434.

[21] Barbier, L.; and Chollet, G. (1991), "Robust Speech Parameters Extraction for Word Recognition in Noise using Neural Networks," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 145-158.

[22] Agbinya, J. I. (1996), "Discrete Wavelet Transform Techniques in Speech Processing", *IEEE Tencon Digital Signal Processing Applications*, Vol. 2, pp. 514-519.

[23] Ramli, A. R.; and Ibrahim, A.; and Syed, A. R. (2003), "Comparing Speech Compression using Wavelets with other Speech Compression Schemes," *IEEE student Conference on Research and Development*, Vol. 1, pp. 55-58.

[24] Lu, C. T.; and Wang, H. C. (2003), "Enhancement of Single Channel Speech based on Masking Property and Wavelet Transform," *Speech Communication*, Vol. 41, No. 2-3, pp. 409–427.

[25] Tan, B. T.; and Lang, R.; and Schroder, H.; and Dermody, P. (1994), "Applying Wavelet Analysis to Speech Segmentation and Classification," *Wavelet Applications in Proceeding of SPIE*, pp. 750-761.

[26] Davenport, M.; and Garudadri, H. (1991), "A Neural Network Acoustic Phonetic Feature Extractor based on Wavelets," *IEEE Pacific Rim Conference on Communication Computers and Signal Processing*, Vol. 2, pp. 449-452.

[27] Tewfik, A. H. (1992), "On the Optimal Choice of a Wavelet for Signal Representation," *IEEE Transaction on Information Theory*, Vol. 38, No. 2.

[28] Petropulu, W. C.; and A. P. (1996), "Pitch Determination and Speech Segmentation using the Discrete Wavelet Transform", *IEEE International Symposium on Circuits and Systems*, Vol. 2, pp. 45-48.

[29] Obaidat, M. S.; and Lee, C.; and Zhang, T.; and Nelson, G. (1999), "Wavelet Algorithm for the Estimation of Pitch Period of Speech Signals," *IEEE 3rd International Conference on Electronics, Circuits and Systems*, Vol. 119, pp. 471-474.

[30] Janer, L.; and Bonet, J; and Lleida, E. S. (1996), "Pitch Detection and Voiced/Unvoiced Decision Algorithm based on Wavelet Transforms", *IEEE 4th International Conference on Spoken Language*, Vol. 2, pp.1209-1212.

[31] Shensa, M. J. (1992), "The Discrete Wavelet Transform: Wedding the A Trous and Mallat Algorithms," *IEEE Transactions on Signal Processing*, Vol. 40, pp. 2464-2482.

[32] Bahl, L. R.; and Brown, P.; and Mercer, R. L. (1988), "Speech Recognition with Continuous Parameter Hidden Markov Models," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 40-43.

[33] Burton, D.; and Shore, J.; and Buck, J. (1990), "Isolated Word Speech Recognition using Multiresolution Vector Quantization Codebooks," *IEEE Transaction on Acoustics, Speech and Signal Processing*, pp. 837-849.

[34] Noue, D. L.; and Levinson, P.; and Sondhi, M. (1987), "Incorporating the Time Correlation between Successive Observations in an Acoustic-Phonetic Hidden Markov Model for Continuous Speech Recognition," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 43-48.

[35] Chang, S.; and Kwon, Y.; and Yang, S. (1998), "Speech Feature Extracted from Adaptive Wavelet for Speech Recognition," *IET on Electronics Letters*, Vol. 34, pp. 2211–2213.

[36] Hsieh, C.; and Lai, E.; and Wang, Y. (2003), "Robust Speaker Identification System based on Wavelet Transform and Gaussian Mixture Model," *Journal of Information Science and Engineering*, Vol. 11, pp. 267-282.

[37] Kadambe, S.; and Bartels, B. (1992), "Application of the Wavelet Transform for Pitch Detection of Speech Signals," *IEEE Transactions on Information Theory*, Vol.32, pp. 712-718.

[38] Gaikwad, S. K.; and Gawali, B. W. (2010), "A Review on Speech Recognition Technique," *International Journal of Computer Applications*, Vol. 10.

[39] Davis, S. B.; and Mermelstein, P. (1980), "Comparison of parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transaction on Acoustic, Speech, Signal Processing*, Vol. 4, pp. 357–366.

[40] Berouti, M.; and Schwartz, R.; and Makhoul, J. (1979), "Enhancement of Speech Corrupted by Acoustic Noise," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. 4, pp. 208-211.

[41] Quatieri, T. F.; and Wang, T. T. (2002), "2-D Processing of Speech with Application to Pitch Estimation," *In Processing of 7th International Conference on Spoken Language*, pp. 1737–1740.

[42] Rabiner, L. R. (1989), ''A tutorial on hidden Markov Models and Selected Application in Speech Recognition,'' *In Proceeding of the IEEE*, Vol. 77, No. 2, pp. 257–286.

[43] Srinivasan, P.; and L. Jamieson, H. (1998), "High Quality Audio Compression using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling," *IEEE Transactions on Signal Processing*, Vol. 46, pp. 1085-1093.

[44] Ahmadi, S.; and Spanias, A. S. (1999), "Cepstrum Based Pitch Detection using a New Statistical Voiced/Unvoiced Classification Algorithm," *IEEE Transaction on Acoustic, Speech, Signal Processing*, Vol. 7, pp. 333-338.

[45] Wu, Y.; and Du, R. (1996), "Feature Extraction and Assessment using Wavelet Packets for Monitoring of Machining Processes," *Mechanical Systems and Signal Processing*, Vol. 10, pp. 29-53.

[46] Jain, A. K.; and Mao, J.; and Mohiuddin, K. M. (1996), "Artificial Neural Network: A Tutorial," *Computer*, Vol. 29, pp. 31-44.

[47] Levinson, S. E. (1985), "Structural Methods in Automatic Speech Recognition," *IEEE Proceeding*, Vol. 73, No. 11, pp. 1625-1650.

[48] Unser, M.; and Aldroubi, A. (1996), "A Review of Wavelets in Biomedical Applications," *Proceedings of the IEEE*, Vol. 84, No. 4, pp. 626- 638.

[49] Mark, J; and Fabio, A. (2000), "Neural Network for Text-to-Speech Phoneme Recognition," *IEEE International Conference on Systems, Man and Cybernetics,* Vol. 5, pp. 3582-3587.

[50] Polur, D; and Yang, J. (2001), "Isolated Speech Recognition using Neural Network," *IEEE 23rd Annual International Conference on Engineering in Medicine and Biology Society,* Vol. 2, pp. 1731-1734.

[51] Maaly, I. A.; and Obaid, M. (2006)," Speech Recognition using Artificial Neural Network," *IEEE Conference on Information and Communication Technologies,* Vol. 1, pp. 1246-1247.

[52] Min- Lun Lan; and Shing-Tai Pan; and Chih-Chin Lai. (2006), "Using Genetic Algorithm to Improve the Performance of Speech Recognition Based on Artificial Neural Network," *IEEE Conference on Innovative Computing, Information and Control,* Vol. 2, pp. 527-530.