Performance Analysis of Different Wavelet Families in Recognizing Speech

Sonia Sunny[1], David Peter S[2], K Poulose Jacob[3]

[1]Dept. of Computer Science, Cochin University of Science & Technology, Kochi, India
[3]Dept. of Computer Science, Cochin University of Science & Technology, Kochi, India
[2]School of Engineering, Cochin University of Science & Technology
Kochi, India

***Abstract*— Automatic Speech Recognition (ASR) is one of the challenging areas of research in digital signal processing and engineering due to its wide range of applications. In this paper, a speech recognition system is developed for recognizing speaker independent spoken isolated words in Malayalam. Voice signals are sampled directly from the microphone and the features are extracted using Discrete Wavelet Transforms (DWT). Different types of wavelet families are available for speech processing and mathematical analysis. Since DWT uses wavelets, the main issue here is to find out the optimal wavelets for speech recognition. This paper investigates the performance of different wavelet families like Haar, Daubechies, Symlets, Coiflets etc. A multi-layer neural network trained with back propagation algorithm is used for classification. The proposed method is implemented for 1000 speakers uttering 10 isolated words each. The experimental results show different recognition accuracies for different wavelet families and the best result of 90.2% is obtained using Daubechies wavelet families with order 4.**

***Keywords*— Speech Recognition, Feature Extraction, Wavelet Families, Discrete Wavelet Transforms, Classification, Artificial Neural Networks.**

I. INTRODUCTION

Speech is the most prominent, natural and primary means of communication among human beings and speech recognition technology has been around since 1960s. Speech recognition has tremendous growth in the last few years due to the advances in signal processing, algorithms, new architectures and hardware [1]. But a hundred percent perfect speech recognition system is unable to be developed. Many factors affect the recognition accuracy and performance of a speech recognition system. Some of the factors include gender, emotional state of the speaker, differences in terms of their accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed. Moreover, speech patterns are also distorted by background noise and echoes, as well as electrical characteristics [2]. Thus designing a perfect speech recognition system is a very complex problem.

In this work, the speech recognition system is designed for recognizing speaker independent isolated words. Speaker independent speech recognition systems can handle a wide range of speakers. In isolated word recognition, there is a pause between each word and only single words are identified at a time. The performance of a speech recognition system is measurable. The most widely used measurement is the recognition accuracy. The recognition accuracy of a speech recognition system mostly depends on the features extracted from the speech signals. In this work, we have used Discrete Wavelet Transforms for feature extraction since it produces better results than many other methods [3]. The multi-resolution analysis property of the wavelet transform is suitable for extracting feature vectors [4]. Wavelets are now used in many areas like signal synthesis and analysis, de-noising and compression, pattern recognition and signal and image processing [5]. Since there are different mother wavelets belonging to different wavelet families available, the choice of the wavelet family, mother wavelet and its order greatly affect the accuracy of the analysis. This paper deals with the problem of choosing the most suitable wavelet family and the most suitable mother wavelet for feature extraction. Since neural networks are good in pattern recognition, they are used for classification here.

The organization of the paper is as follows. In section 2, a brief description of the objective of our work and the methodology used is explained. The isolated spoken words database used is given in section 3. Data preprocessing using wavelet denoising is described in section 4. A brief summary of different wavelet families, mother wavelets and their characteristics are explained in section 5. The theory of feature extraction and a review of the basic principles of the discrete wavelet transforms used during this stage are illustrated in section 6. The classification stage using artificial neural networks (ANN) is discussed in section 7. Section 8 presents the detailed analysis of the experiments done and the results obtained. Last section contains the conclusions and future work.

II. STATEMENT OF THE PROBLEM AND METHODOLOGY

Nowadays, wavelets are very much used for feature extraction. The main advantage of using wavelets is their fundamental vanishing moment property and the ability of the scaling functions to reproduce polynomials [6]. There are different types of wavelet families and mother wavelets available. So, when using wavelet transforms for feature extraction, an important question which arises is regarding the choice of the suitable wavelet family and hence mother wavelet that should be used for the analysis of speech signals. Selection of the wavelet family and mother wavelet plays an important role in the performance of the wavelet transforms. Discrete wavelet transforms are used for feature extraction here. The main objective of this work is to evaluate the performance of different wavelet families and mother wavelets that can be used along with DWT.

This work is divided into four modules. First, the Malayalam words database is created since there is no built-in database available in Malayalam. These recorded speech signals are often degraded by background noise. So this noise should be removed by pre-processing techniques before extracting the features. So wavelet denoising technique based on soft thresholding is used for noise suppression. The third stage is the feature extraction stage where the feature vector set is derived using DWT. Finally during classification stage, a supervised neural network classifier is used in order to obtain the classification performance of the proposed features.

III. DATA PREPARATION

Creation of the database is the first stage of almost all recognizer development work. Here, ten isolated words from Malayalam language are chosen to create the database since there is no built-in database available in Malayalam. Thousand speakers are selected to record the words. Each speaker utters 10 words. Thus the database consists of a total of 10000 utterances of the spoken words. We have used 400 male speakers, 400 female speakers and 200 children for creating the database. Male, female and voice of children differ in pitch, frequency, phonetics and many other factors due to the difference in physiological as well as psychological factors. The speech samples are taken from speakers of different ages between 6 and 70. The samples stored in the database are recorded by using a high quality studio-recording microphone at a sampling rate of 8 KHz (4 KHz band limited). The spoken words, words in English, their International Phonetic Alphabet (IPA) format and translation in English are shown in Table 1.

TABLE 1.
WORDS STORED IN THE DATABASE AND THEIR IPA FORMAT

Words in Malayalam	Words in English	IPA format	English translation
ഓണം	Onam	/əunʌm/	Onam
ചിരി	Chiri	/tʃiri/	Smile
വീട്	Veedu	/vi:də/	House
കുട്ടി	Kutti	/kuţi/	Child
മരം	Maram	/mʌrəm/	Tree
മയിൽ	Mayil	/mʌjil/	Peacock
ലോകം	Lokam	/ləukʌm/	World
മൗനം	Mounam	/maunəm/	Silence
വെള്ളം	Vellam	/ve!!ʌm/	Water
അമ്മ	Amma	/ʌmmʌ/	Mother

IV. DATA PREPROCESSING

In this work, we have used wavelet denoising method for removing noise. The two popular thresholding based wavelet denoising method are the hard and the soft thresholding functions [7]. Hard thresholding sets to zero any element whose absolute value is lower than the threshold. Soft thresholding is an extension of hard thresholding. The elements whose absolute values are lower than the threshold are first set to zero and then shrinks the nonzero coefficients towards 0. Hard and soft thresholding can be defined as

$$X_{Hard} = \begin{cases} X & if \quad |X| > \tau \\ 0 & if \quad |X| \leq \tau \end{cases} \qquad (1)$$

$$X_{Soft} = \begin{cases} sign(X)\ (|X| - |\tau|) & if \quad |X| > \tau \\ 0 & if \quad |X| \leq \tau \end{cases} \qquad (2)$$

Where X represents the wavelet coefficients and ι is the threshold value. In this paper, soft thresholding is used for wavelet denoising. Threshold value can be obtained using different methods. In this paper, the threshold used is the universal threshold developed by Donoho and Jonstone [8] which is defined as

$$\iota = \sigma\sqrt{2\log(N)} \qquad (3)$$

where σ is the standard deviation and N is the length of the signal. Standard deviation σ can be calculated as σ = MAD/0.6745, where MAD is the median of the absolute value of the wavelet coefficients. The outline of the denoising algorithm used is

a) Compute the discrete wavelet transform of the noisy signal up to 8 levels.
b) Use soft thresholding technique to shrink some detail wavelet coefficients.
c) Compute the inverse discrete wavelet transform.

The denoised signals obtained after pre-processing are then applied to DWT for extracting the features.

V. Wavelet Families

Wavelets means small waves which are functions that satisfy certain mathematical requirements and were developed to overcome the shortcomings of Fourier Transforms. The main characteristics of wavelets are their multi-resolutional, multi-scale analysis, ability to have various time-frequency distributions, fundamental vanishing moment property, ability of the scaling functions to reproduce polynomials etc. The mathematical foundations of wavelets and multiresolution analysis were laid down by Mallat [9] and Meyer [10].

Different wavelet families can be used in Wavelet analysis. There are different types of wavelet families [9] such as Haar, Daubechies, Biorthogonal, Symlets, Coiflets, Morlet, Mexican Hat, Meyer etc. Wavelets differ in the length of support of the mother wavelet, the number of vanishing moments, the symmetry or the regularity, the existence of a corresponding scaling function etc. Since all the translations and scaling are through the mother wavelet, the selection of the mother wavelet plays an important role in obtaining good recognition accuracy in speech recognition. Moreover, a wavelet family which works well for a particular application may not be good for some other application. So we cannot say that one wavelet is superior to another because its performance depends on the application. In this work, the performance of different wavelet families in recognizing isolated words in Malayalam is carried out. Each family consists of wavelet functions or mother wavelets but with different order. A review of these wavelet functions and their properties are given below.

A. Haar wavelet

The Haar wavelet or db1 is the simplest and oldest type of wavelets and has the shortest support among all orthogonal wavelets. It is a sequence of rescaled, square-shaped functions which together form a wavelet family. They are well suited for analysis of signals with sudden transitions although they are not continuous and differentiable. Fig. 1 shows the plot of the haar wavelet.

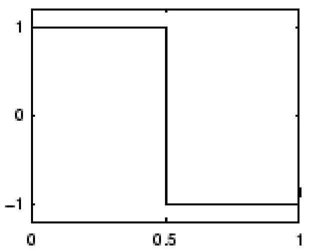

Fig. 1 Plot of Haar Wavelet

B. Daubechies Wavelets

These are the most popular wavelets that represent foundations of Digital Signal Processing. Many researches in speech recognition are based on this wavelet because they represent a collection of orthogonal mother wavelets which are characterized by a maximal number of vanishing moments for some given length of the support. Daubechies wavelets are not symmetric in nature and use overlapping windows and are represented as 'db'. Due to this property, the high frequency coefficient spectrum reflects all high frequency changes. So they are suitable for denoising signals and for compression [11]. Fig. 2 shows the plots of this family.

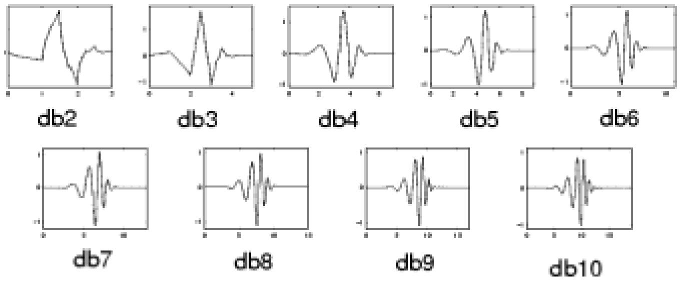

Fig. 2 Plots of different Daubechies wavelets

C. Coiflets Family

It has the highest number of vanishing moments. They are compactly supported and orthogonal but are near from symmetry. It has the surname of 'coif and has five members in this family starting from 'coif1' to 'coif5'. Fig. 3 shows the members of this family.

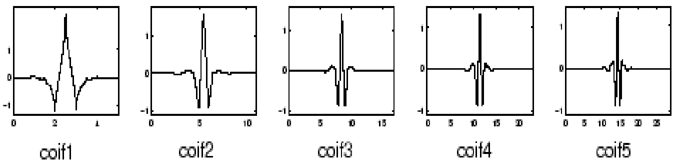

Fig. 3 Plots of different Coiflets Family Wavelets

D. Symlets Family

These are symmetrical wavelets which were invented to modify the symmetry property of the 'db' family. It contains seven members starting from 'sym2' to 'sym8'. This family is almost symmetrical, orthogonal, compactly supported in time and has N vanishing moments. The members of this family are shown in fig 4.

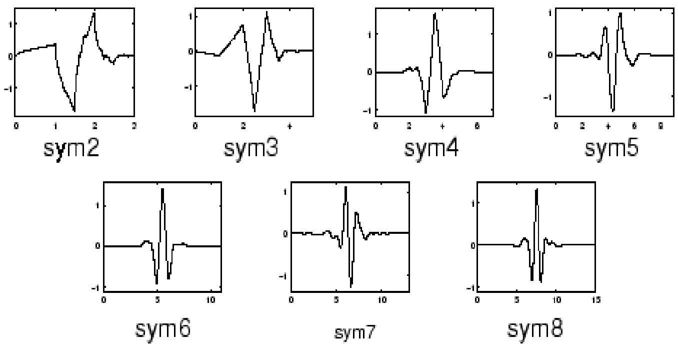

Fig. 4 Plots of different Symlet Wavelets

VI. Feature Extraction using Discrete Wavelet Transforms

A wavelet transform is used to transform the signal under investigation into a more useful representation. Wavelet transforms are well suited for non-stationary signals like speech because of its time frequency localizations [12]. In discrete wavelet transform, the scale and position are changed in discrete steps [13]. The main advantage of the wavelet transforms is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges [14].

The Discrete Wavelet Transform is defined by the following equation

$$W(j,K)=\sum_{j}\sum_{k}X(k)2^{-j/2}\psi(2^{-j}n-k) \quad (4)$$

Where Ψ(t) is the basic analyzing function called the mother wavelet. In DWT, the original signal passes through two complementary filters, namely low-pass and high-pass filters, and emerges as two signals, called approximation coefficients and detail coefficients [15]. In speech signals, low frequency components known as the approximations h[n] are of more importance than high frequency signals known as the details g[n] as the low frequency components characterize a signal more than its high frequency components [7]. The successive high pass and low pass filtering of the signal can be obtained by the following equations.

$$Y_{high}[k]=\sum_{n}x[n]g[2k-n] \quad (5)$$

$$Y_{low}[k]=\sum_{n}x[n]h[2k-n] \quad (6)$$

All Where Y_{high} (detail coefficients) and Y_{low} (approximation coefficients) are the outputs of the high pass and low pass filters obtained by sub sampling by 2. The filtering is continued until the desired level is reached according to Mallat algorithm [17]. The discrete wavelet decomposition tree over 3 levels is shown in figure 5.

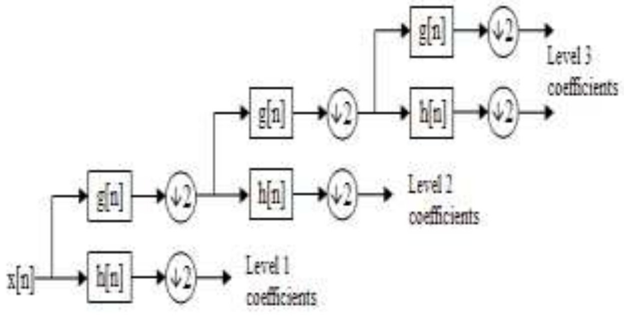

Fig. 5 DWT Decomposition over 3 levels

VII. Classification using ANN

Classification techniques are used to group the data instances into proper classes. ANNs have been applied to an increasing number of real-world problems of considerable complexity. Neural Networks have become a very important method for pattern recognition because of their ability to deal with uncertain, fuzzy, or insufficient data. Algorithms based on Neural Networks are well suitable for addressing speech recognition tasks. Inspired by the human brain, neural network models attempt to use some organizational principles such as learning, generalization, adaptivity, fault tolerance etc. [18].

Here a Multi Layer Perceptron (MLP) structure is used which consists of an input layer, a hidden layer and an output layer. In this work, we have used a supervised error back propagation algorithm. There are two phases in supervised pattern recognition, namely training and testing. The process of extraction of features relevant for classification is common in both phases. During the traning phase, the parameters of the classification model are estimated using training data. During testing or recognition phase, the features of test pattern is matched with the trained model of each and every class. In training mode, target vectors are given which represents the classes. But during testing, no target vector is provided. The test pattern is declared to belong to that whose model matches the test pattern best.

VIII. Experiments and Results

In this paper, the performance of different wavelet families is evaluated. So different wavelets are used for feature extraction using DWT. The feature vectors obtained in each case is given to an MLP structure. The number of feature vectors obtained is 12 and the best results were obtained from decomposition level 8. The feature vectors obtained at 8^{th} level of word Onam using db4 wavelet is shown below.

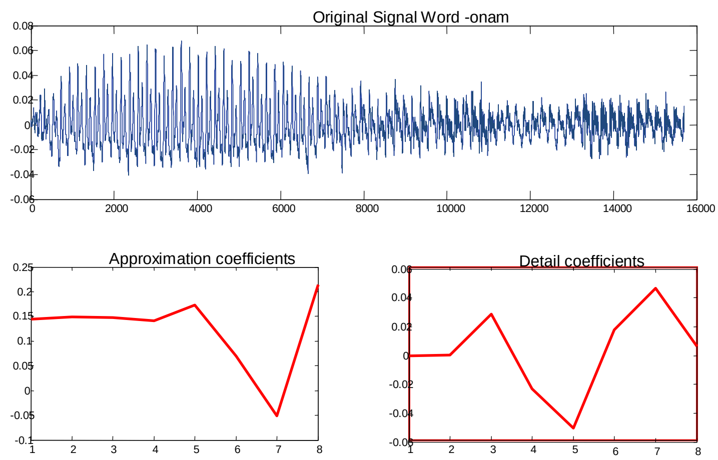

Fig. 6 Decomposition of word Onam at 8^{th} level using DWT

The feature vector set obtained is divided into 3 sets. 70% data is used for training, 15% for validation and 15% for testing. The results obtained using different wavelet families are given below. The graph representing the various wavelet families and the corresponding recognition accuracies are given in figure 7.

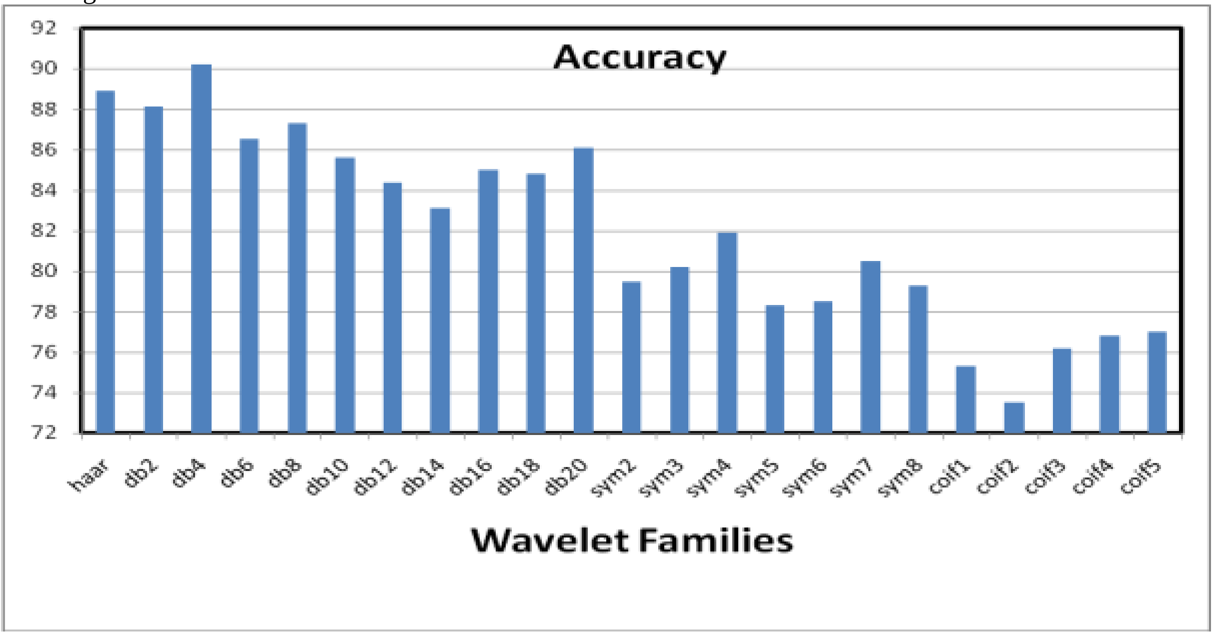

Figure 7. Graph representing wavelet families and recognition accuracy percentage.

A. Results from Haar and Daubechies wavelets

The results obtained using different wavelets in the daubechies wavelet family are given below. The best result obtained is for db4 among db family.

TABLE 2
PERFORMANCE EVALUATION OF DB FAMILY WAVELETS

Wavelet	**Recognition accuracy %**
Haar or db1	88.9
Db2	88.1
Db4	90.2
Db6	86.5
Db8	87.3
Db10	85.6
Db12	84.4
Db14	83.1
Db16	85.0
Db18	84.8
Db20	86.1

B. Experiments using Symlet wavelets

The results obtained using different wavelets in the symlet wavelet family are given below. Optimal results were obtained using sym4.

TABLE 3.
PERFORMANCE EVALUATION OF SYMLET FAMILY WAVELETS

Wavelet	Recognition accuracy %
Sym2	79.5
Sym3	80.2
Sym4	81.9
Sym5	78.3
Sym6	78.5
Sym7	80.5
Sym8	79.3

C. Experiments using Coiflets wavelets

The results obtained using different wavelets in the coiflet wavelet family are given below. Optimal results were obtained using coif4.

TABLE 4
PERFORMANCE EVALUATION OF COIF FAMILY WAVELETS

Wavelet	Recognition accuracy %
Coif1	75.3
Coif2	73.5
Coif3	76.2
Coif4	76.8
Coif5	74.0

From the results obtained, the best results were obtained using db4 wavelets with a recognition accuracy of 90.2%. Haar wavelets also produced good results with an accuracy of 88.9%.

IX. Conclusion

In this paper, an automatic speech recognition system is designed for isolated spoken words in Malayalam using DWT and ANN. Different wavelet families are used for extracting features and a comparative study of these is performed here. A better performance of identification with high recognition accuracy of 90.2% is obtained for almost all the wavelets from this study. But daubechies wavelets are found to be more efficient than the other wavelets. The optimal recognition rate is obtained for db4 type of wavelet. Thus a Daubechies wavelet is an elegant tool for the analysis of non-stationary signals like speech. The experiment results show that this hybrid architecture using discrete wavelet transforms using db4 wavelet and neural networks could effectively extract the features from the speech signal for automatic speech recognition. Alternate classifiers like Support Vector Machines, Genetic algorithms, Fuzzy set approaches can be used as an extension of this work.

Acknowledgment

I wish to express my sincere gratitude to all who have contributed throughout the course of this work.

References

[1] R. Lawrence ,"Applications of Speech Recognition in the Area of Telecommunications," in *Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding*, 1997, p. 501-510.

[2] Recognition [Online]. Available: http://www.learnartificialneuralnetworks.com/speechrecognition.html

[3] Santosh K.Gaikwad, Bharti W.Gawali and Pravin Yannawar, "A Review on Speech Recognition Technique," *International Journal of Computer Applications*, Vol. 10, No.3, pp. 16-24, Nov. 2010.

[4] Wei-Bing Hu, Kai-Cheng Li and Dang-Jun Zhao," A Novel Probabilistic Neural Network System for Power Quality Classification Based On Different Wavelet Transform," in *Proc. of the International Conference on Wavelet Analysis and Pattern Recognition*, 2007, p. 746-750.

[5] H. T. Yang and C. C. Liao, "A de-noising scheme forenhancing wavelet-based power quality monitoring system," *IEEE Trans. Power Delivery*, Vol. 16, No. 7, pp. 353-360, July 2001.

[6] Cédric Vonesch, Thierry Blu and Michael Unser, "Generalized Daubechies Wavelet Families," *IEEE Trans. On Signal Processing*, Vol. 55, No. 9, pp. 4415-4429, September 2007.

[7] S. Kadambe, P. Srinivasan."Application of adaptive wavelets for speech," *Optical Engineering*, Vol. 3(7), pp. 2204-2211, 1994.

[8] D.L. Donoho, "De-noising by soft thresholding," *IEEE transactions on information theory*, Vol. 41, No. 3, pp. 613-627, 1995.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.

[10] Y. Meyer, *Wavelets and Operators*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[11] Mohamed I. Mahmoud, Moawad I. M. Dessouky, Salah Deyab, and Fatma H. Elfouly, "Comparison between Haar and Daubechies Wavelet Transformations on FPGA Technology," *World Academy of Science, Engineering and Technology*, pp. 68-72, 2007.

[12] S. Mallat, *A wavelet Tour of Signal Processing*, Academic Press, San Diego: 1999.

[13] K. P Soman, K.I Ramachandran, N.G Resmi, *Insight into Wavelets From Theory to Practice*, PHI Learning Private Ltd, New Delhi: 2010.

[14] Elif Derya Ubeyil, "Combined Neural Network Model Employing Wavelet Coefficients for ECG Signals Classification," *Digital signal Processing*, Vol. 19, pp. 297-308, 2009.

[15] S. Chan Woo, C.Peng Lin, R. Osman, " Development of a Speaker Recognition System using Wavelets and Artificial Neural networks, " *in Proc. of 2001 Int. Symposium on Intelligent Multimedia, Video and Speech processing*, 2001, pp. 413-416.

[16] S .G. Mallat, " A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis And Machine Intelligence*, Vol.11, pp. 674-693, 1989.

[17] Economou K., Lymberopoulos D, "A New Perspective in Learning Pattern Generation for Teaching Neural Networks," *Neural Networks*, Vol. 12, Issue 4-5, pp. 767-775, 1999.