INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

WAVELET ENTROPY AND NEURAL NETWORK FOR TEXT-DEPENDENT SPEAKER IDENTIFICATION

Ms.M.D.Pawar[1], Ms.S.C.Saraf[2] , Ms.P.P.Patil[3]

[1] Dept. Of Electronics and Telecommunication Engineering, M .I.T, Aurangabad(M.S), INDIA,
. [2] Dept. Of Electronics and Telecommunication Engineering,M.I.T,Aurangabad(M.S), INDIA,
. [3] Dept. Of Electronics and Telecommunication Engineering,M.I.T,Aurangabad(M.S), INDIA
[1]Manjupawar2583@gmail.com, [2]saraf.suchita0@gmail.com, [3]priti.patil87@gmail.com.

Abstract: In this Present study, the technique of wavelet transform and neural network were developed for speech based text-dependent and text0independent speaker identification. 390 feature were fed to feed-forward back propagation neural network for classification The function of feature extraction and classification are performed using wavelet and neural network system. The declared result shows that the proposed method can make an effective analysis with average identification rate reaching 98%. The best recognition rate selection obtained was for FFBPNN (Feed Forward Back Propagation Neural Network).

Keywords: Wavelet Transform, Neural Network, FFBPNN, Feature Extraction, Database.

I. INTRODUCTION

The Speaker Identification is the process of utterance speaker verification, on the other hand, is the technology of accepting or rejecting the identity claim of a speaker .Over the last four decays many solutions of speaker recognition have been appeared in literature. The algorithm for pattern classification , was motivatated by Patterson's and Womack's and wee's proof's that the mean Square Error (MSE) Solution of the Pattern classification solution gives a minimum mean square error approximation to Baye's discrimination weighted by the probability density function of the sample. All audio techniques start by converting the raw speech signal into sequences of acoustic feature vector carrying distinct information about the signal. This feature extraction is also called 'front-end' in the literature. The most commonly used acoustic vectors are Mel Frequency Coefficients (MFCC), Linear Prediction Cepstral Coefficient (LPCC) Coefficient. All these features are based on the Spectral information derived from a short time windowed segment of speech signal.

One of the most Common short-term spectral measurements currently used are Linear Predictive coding (LPC) derived Coefficients and their regression Coefficients. A Spectral envelope reconstructed from a truncated set of cepstral coefficients is much smoother than one reconstructed from LPC coefficients. Therefore it provide a stabler

representation from one repletion to another speakers utterances' As for the regression coefficients, typically the first and second order coefficients are extracted at every frame period to represent the spectral dynamics.
These coefficients are derivatives of the cepstral coefficients and are respectively called the delta and delta-delta cepstral coefficients. Text dependent methods are usually based on template matching techniques. In this approach , the input utterances is represented by a sequences of feature vectors, generally short-term spectral feature vectors. The time axis of the input utterance and each reference template or reference model of the registered speakers are aligned using a dynamic time warping (DTW) Algorithm and the degree of similarity between them , accumulated from the beginning to the end of the utterance, is calculated. The Hidden Markov Model (HMM) can efficiently model statistical variation in spectral features. Therefore HMM based methods were introduced as extensions of the DTW-based methods, and have achieved significantly better recognition accuracies, One of the most successful text-independent recognition methods based on vector quantization (VQ) , In this methods , VQ Codebooks consists of a small number of representative feature vectors are used as an efficient means of characterizing. Speaker-specific features. A Speaker-specific codebook is generated by clustering the training feature vectors to each speaker. In the recognition stage, an input utterance is vector-quantized using the codebook of each reference speaker and the VQ distortion accumulated over the entire input utterance is used to make the recognition decision. Temporal variation in speech signal parameter over the long term can be represented by stochastic Markova transition between states. Therefore, methods using an ergodic HMM, Where all possible transition between states are allowed, have been proposed. Speech segments are classified into one of the board phonetic categories corresponding to the HMM states. After the classification, appropriate features are selected.
It has been shown that a common ergodic HMM method is far superior to a discrete ergodic HMM method and that a continuous ergodic HMM method is as robust as VQ based method when enough training data is available.
A method using statistical dynamic features has recently been proposed. In this method, a multivariate auto-regression (MAR) Model is applied to the time series of cepstral vectors and used to characterize speakers [26].
In this paper , the wavelet Transform based speaker recognition system is proposed. This system is divided into two main blocks, signal enhancement by feature extracting and identification. In the first block use Adaline as neural network to enhance each sub-signal that produced by the DWT. This system depends on DWT generates the desired sub-signals to the neural net, This means multiple input will be applied to the neural net depends on selected level. The same process is applied for noisy signal, The output of the neural net will be back to the Inverse DWT to be reconstructed enhanced signal. The aim of this method is to filter the speech signal from the noise in selected sud-signals of distinct frequency sub-band. This assist greatly in eliminating special frequencies and can preserve other frequencies that are essential for speech recognition. In speaker identification two blocks are applied: the first is WGD block, where three continuous wavelet transform (CWT) sub-signals are used of different pass band of frequency, high average and low . Then stastical functions are used to extract gender features. Sharp threshold between male and female is achieved. The second step is speaker identification using the enhanced sub-signal of suitable level that must be on speakers own features frequency, depending on his anatomical structure of his own vocal tract and other working parts through speaking process is used foe feature extraction. And finally neural network is used for classify the features.

2 .PROPOSED METHOD

In My paper wavelet transform based identification system is presented. This particular system based on two main blocks shown in fig.1. In this first block speech signal is enhanced by Continuous Wavelet Transform method. The Second block contains the feature extraction method by Discrete wavelet transform and classification using FFBPNN(Feed Forward Back Propagation Neural network). Feature extraction method in the second block is divided into two steps: Wavelet Gender Discrimination and Feature Extraction, which assists in discrimination the speech signal into two classes (male and female) that makes the feature extraction by wavelet and is more efficient.

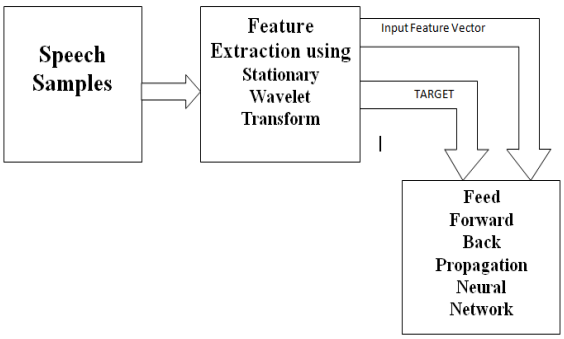

Fig.1 Block diagram. Of Proposed Method

A. Speech Signal

For the evaluation of the speaker identification system, our own in-house speech corpus In-House Dataset.In-House dataset, compiled in the laboratory in the institute contains the speech samples of institute's students. This contains 10 speech samples of 40 speakers. Sound-fourge software was used for recording the speech and converted them into .wav file. Matlab7.9 was used to record speaker's voice

B Wavelet transform

The WT is designed to address the problem of nonstationary signals. It involves representing a time function in terms of simple, fixed building blocks, termed wavelets. These building blocks are actually a family of functions which are derived from a single generating function called the mother wavelet by translation and dilation operations. Dilation, also known as scaling, compresses or stretches the mother wavelet and translation shifts it along the time axis.
The WT can be categorized into continuous and discrete. Continuous wavelet transform (CWT) is defined by,

$$\mathrm{CWT}(a,b) = \int_{-\infty}^{+\infty} x(t)\psi^{*}_{a,b}(t)\,dt, \qquad \text{.... (1)}$$

Where $x(t)$ represents the analyzed signal, a and b represent the scaling factor (dilatation/compression coefficient)
And translation along the time axis (shifting coefficient), respectively, and the superscript asterisk denotes the complex conjugation. ψa,b(·) is obtained by scaling the wavelet at time *b* and scale *a*:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \qquad \text{.....(2)}$$

where $\psi(t)$ represents the wavelet Continuous, in the context of the WT implies that the scaling and translation parameters *a* and *b* change continuously. However , calculating wavelet coefficients for every possible scale can represent a considerable effort and result in a vast amount of data.

C. Stationary Wavelet Transform

As the classical DWT suffers a drawback the DWT is not a time invariant transform. This means that, even with periodic signal extension, the DWT of a translated version of a signal *X* is not, in general, the translated version of the DWT of *X*. How to restore the translation invariance, which is a desirable property lost by the classical DWT.

The idea is to average some slightly different DWT, called e-decimated DWT, to define the stationary wavelet transform (SWT). This property is useful for several applications such as breakdown points detection.

***D.* Features used in the proposed system**

Wavelet coefficient provides information about the energy distribution of the signal in time and frequency. The following features have been used in this thesis work.

• The mean of the absolute value of the approximate and detail coefficients at each level. These features provide frequency distribution information of the signal.

• The standard deviation of the approximate and detail coefficients at each level. These features provide information about the amount of change of the frequency distribution.

3. Feature Matching

Feature matching is the important stage in speaker identification system and several techniques exist that can be used to model speakers based on the features extracted from speech samples. In this Project a standard classifier is used, That is Artificial Neural Network (ANN) having discriminative-training power,

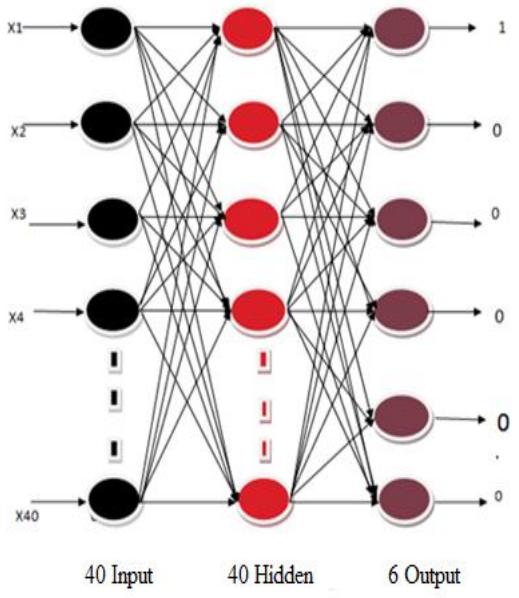

Figure 2 Architecture of Feed Forward Back-propagation Neural Network

4. Back-propagation Neural Network.

Back propagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities, Firstly the feature vectors of all speakers have been collected in a matrix then a target matrix is designed so that all feature vectors belonging to one speaker is labelled as “one” and the components for the remaining feature vectors are labelled as “zero” in the target matrix. In the experiments a three layer fully connected network is used with 20 neurons as input. To implement FFBNN can use Matlab neural network toolbox by function newff, tansig , Purlin transfer function and trainlm back

propagation training function , newff commend builds a network of three layers: 20 neurons input layer, 20 neurons hidden layer and 5 neurons output layer as shown in fig.2,. After training with the target by train, Cross- validation technique has been applied for finding the appropriate architecture of the network. The back propagation algorithm has been used for training the network. To control the weight modifications, a learning rate constant was used. Optimal value of learning rate constant is important because if the value is very low, learning takes forever and if the value is big, the learning disrupts the previous knowledge.

Newff=(minmax(nb),[40 6],,{'tansig','purelin'},'trainlm');.....(3)

The transfer function in the first layer will be tan-sigmoid, and the output layer transfer function will be linear. The values for the first element of the input vector will range between -1 and +1, and the training function will be trainlm.

Table 1: Performance of Back-propagation Neural Network Classification

Female – 20	
Target Output	**Actual Output**
0	0.09
0	0.007
1	0.968
0	0.02
1	1.02
0	0.002

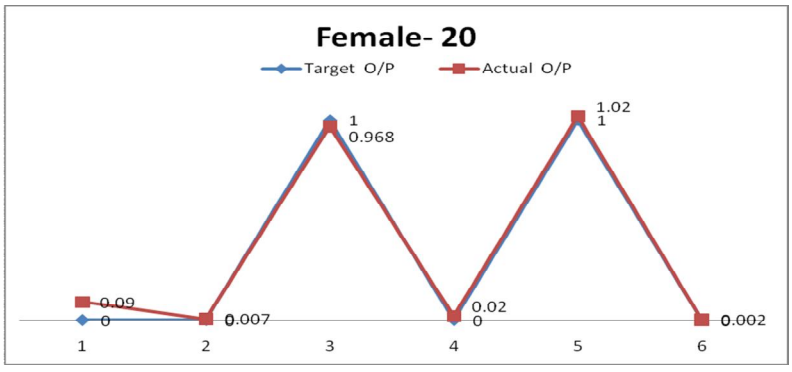

Figure.3 Performance of Neural Network Classification Female-20

5 Conclusion and Discussion

The aim of this project was to implement an application that would verify a speaker's identity by using the speaker's voice print. In this Present Work Wavelet Transform and Neural network based approach has been used for speaker identification system. Stationary Wavelet Transform is successfully used to extract feature in order to build robust speaker identification system. Neural Networks (ANN) are a branch of the field known as "Artificial Intelligence" ANN are based on the basic model of the human brain with capability of generalization and learning. The purpose of this simulation to the simple model of human neural cell is to acquire the intelligent features of these cells. The term "artificial" means that neural nets are implemented in computer programs that are able to handle the large number of necessary calculations during the learning process.ANN have been introduced in solving a lot of problems related to prediction, classification, control and identification. This is due to their high ability to learn from experience in order to improve their performance and to adapt themselves to changes in the environment in addition to their ability to deal with incomplete information or noisy data and can be very effective especially in situations where it is not possible to define the rules or steps that lead to the solution of a problem. A neural network is a powerful data modelling tool that is able to capture and represent complex input/output relationships. The purpose of the neural network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown.ANN have shown high efficiency as predictive tool by looking at the present information's. The SWT algorithm is very simple one. More precisely, for level 1, all the decimated DWT for a given signal can be obtained by convolving the signal with the appropriate filters as in the DWT case but without down sampling. Then the approximation and detail coefficients at level 1 are both of size *N*, which is the signal length. The main advantage of the SWT is de-noising.

Feature extraction was the main task for speaker identification system. This involved speaker specific features extraction from speech signal in Stationary wavelet transform domain. The system's accuracy depends upon the feature vector which was used to create the speaker's model for classification purpose.The design and development of speaker identification system has been presented in this work. The speech dataset used for experiment contains ten samples of twenty male and twenty female voice profiles. The experiment results indicate that Stationary wavelet transform produce best performance. The proposed design of the speaker identification system used fourty feature vectors containing forty features of fourty people in that 20 are male and 20 are female speakers. The speech corpus achieved 98% accuracy using 40 (In-House Database) speakers

REFERENCES

[1] *K.Drqrouq, T.Abu Hilal, M Sharif,s.and A-Al Qawasmi,'Speaker identification using wavelet and neural network', IEEE Transctions on computers,July2009.*
[2] *M.a.Al-Aloaoui, 'A new Weighted Generalized Algorithm for Pattern Recognition." IEEE Transactions on computers, vol. C-5,no 10,October 1977.*
[3] *D. Gabor, 'Theory of communication' Journal of IEEE, 1993.*
[4] *J.M.Naik, L.P.Nestsh and G.R. Doddingont , 'speaker Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing , Glasgow, Scotland, May 1989, pages(524-527).*
[5] *B Abdel-Rahman Al-Qawasmi and Khaled Daqrouq; 'Discrete Wavelet Transform with Enhancement Filter for ECG signal'.*
[6] *M.AMAL-Alaoui Some Applications of Pattern Recognition , Ph.D. Thesis electric engineering department, Georgia Institute of December, 1974.*
[7] *AN EFFICIENT FEATURE SELECTION METHOD FOR SPEAKER RECOGNITION, Hanwu sun, Bin Ma and Haizhou Li,Insititute of infocomm research..*
[8] *Institute for Infocomm Research agency for Science, Technology and research, Singapore.*
[9] *T. . Matsui and S. Furui. Concatenated phoneme models for IEEE Proceedings of the 1993 International*

conferences acoustic speech and signal and speech signal processing , April 1993.
[10] *Chakroborty, S., Roy, A. and Saha, G., "Improved Independent speaker identification by combining MFCC with vidence from Flipped Filter Banks',*
[11] *T. Matsui and S. Furui. Comparison of text-independent speaker recognition methods using VQ-distortion and discreet continuous HMM, IN IEEE Procedeeing of 1999.by R.B.Polikar,volume III AND VOL.IV,on computers,*
[14] *. Liu C. H., Chen O. T. C. A Text independence speaker Identification system using PARCOR and AR model Vol 3-335-336,2002.*
[15] The *GABOR 'Theory of Communication, Journel of I.E.E. 93- 97 PP ,(429-19).*
[16] Introduction *to speech recognition Kimberlee A. Kemble, program manager,IBM Coporation,20008.*