International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Learning Vector Quantization (LVQ) Neural Network Approach for Multilingual Speech Recognition

Rajat Haldar[1], Dr. Pankaj Kumar Mishra[2]

1Electronics & Telecommunication Engineering Department

RCET, Bhilai (C.G.) India

haldarrajat12@gmail.com

2Electronics & Telecommunication Engineering Department

RCET, Bhilai (C.G.) India

pmishra1974@yahoo.co.in

Abstract – *Speech is the most popular and efficient way to communicate and interact with each other. It is also a useful medium to connect with the machine. This paper presents a technique for Speech Recognition & Language Identification of Indian Languages. Speech Recognition & Language Identification have performed for Bengali, Chhattisgarhi, English and Hindi language using Learning Vector Quantization (LVQ) Neural Network and Particle Swarm Optimization (PSO) technique. This method is applied into two phases, in the first phase MFCC and LPC feature extraction technique, Learning Vector Quantization (LVQ) neural network used for Multilingual Speech Recognition and Language Identification. Second phase is approximately similar to the first stage except that PSO technique is used in the second phase for the both procedure. MFCC and LPC are most useful feature extraction technique for speech recognition. On the basis of "Recognition Rate" the system performance is measured. Multilingual Speech Recognition and Language Identification using LVQ neural network and PSO technique gives slightly better Recognition Rate as compare to the without PSO technique***.***

***Key Words*: Artificial Neural Network, Multilingual Speech Recognition, Learning Vector Quantization, Language Identification, Mel Frequency Cepstrum Coefficients, Linear Predictive Coding, Automatic Speech Recognition, Particle Swarm Optimization.**

1. INTRODUCTION

Speech recognition of more than two languages can be performed with neural networks, in this process at first system analyze the input speech signals on the knowledge base and then it perform the recognition, it is called "Multilingual Speech Recognition." The basic steps applied in Multilingual Speech Recognition are feature extraction, training and testing of the speech signals. In this paper LVQ neural network used as a classifier for training and testing of the speech signals. Language Identification is also performed in this paper by using LVQ neural network and PSO technique. Particle Swarm Optimization (PSO) technique is very useful technique for optimizes the data and selects the better value. In car system the speech recognition can be very useful, in car system the door open after the recognition of the human voice. Similarly for other security purpose like home security, mobile security purpose "Speech Recognition" can be use, in Military for air traffic control, to giving command in a secure way, in the field of communication for example telephonic conversation it is useful.

1.1 Artificial Neural Network (ANN)

The ANN is consists of many nodes which is also called the processing elements. The nodes are interred connected to each other with weighted connection. ANN is useful for pattern recognition and function approximation. The simple multilayer ANN is consists of input layer, hidden layer and output layer. The diagram of ANN is shown in Fig-1 where X1, X2 and X3 are the input layer neurons, H1, H2 and H2 are hidden layer neurons, Y1, Y2 and Y3 are the output layer neurons.

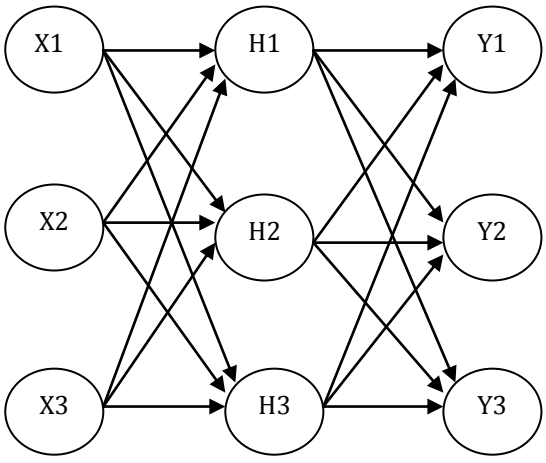

Fig-1: Artificial Neural Network

Several methodologies have performed for the speech recognition in literature. Back Propagation Neural Network [1] approach with LPC [1] feature has been performed for English Alphabet recognition, Hierarchical Spiking Recurrent Self Organizing [2] with MFCC [2, 11] features method performed for speech recognition. Time Delay Neural Network (TDNN) [3], hybrid ANN/HMM [4] approach is also used for speech recognition. Hybrid ANN/HMM model is the combination of Artificial Neural Network (ANN) and Hidden Markov Model (HMM), TDNN and ANN/HMM is applied with MFCC features for speech recognition. Convolution Neural network [5] and Perceptron network [7] is also used for speech recognition. For speech recognition and speaker identification ANN [11] is used. Language Identification [12] has been done for the Bosque context by applying hybrid ANN/HMM model, Perceptron Network [7, 12] and self organizing map [12].

This paper proposed Multilingual Speech Recognition Language Identification using LVQ neural network and PSO technique. This paper organized as follows methodology is given in section 2, section 3 is result and discussion, section 4 is conclusion and future scope.

2. METHODOLOGY

The methodology of this research work divided into two phases and comparison between these two phases. In the first phase MFCC and LPC feature extraction technique, Learning Vector Quantization (LVQ) neural network used for Multilingual Speech Recognition and Language Identification. Second phase is approximately similar to the first stage except that PSO technique is used in the second phase for the both procedure. After applying these two steps comparison is performed between these two phases. LVQ neural network used as a classifier for the training and testing procedure. In this proposed project MFCC and LPC are used for the feature extraction technique. In speech recognition it is the key step so that it is necessary to keep more attention to feature extraction procedure.

The flow chart of the methodology in given in Fig-2, the methodology is divided into two phases. In first phase PSO technique is not performed and in the second phase PSO technique has been adopted for feature selection. In the first phase at first the signal is loaded to MATLAB for preprocessing, and then it is followed by MFCC and LPC feature extraction further training and testing by LVQ neural network. Second stage is similar to the first stage except that in second stage MFCC and LPC feature extraction is followed by Particle Swarm Optimization (PSO) feature selection technique. The training and testing procedure is performed by LVQ neural network for both stages, rest all the procedures are same in first and second stage. At last the comparison has done between these two phases on the basis of Recognition Rate and error. Signal preprocessing, feature extraction, training and testing are the mandatory steps for any speech signal recognition. LVQ neural network is same as Kohonen Self Organizing Map except that in LVQ supervised learning is used.

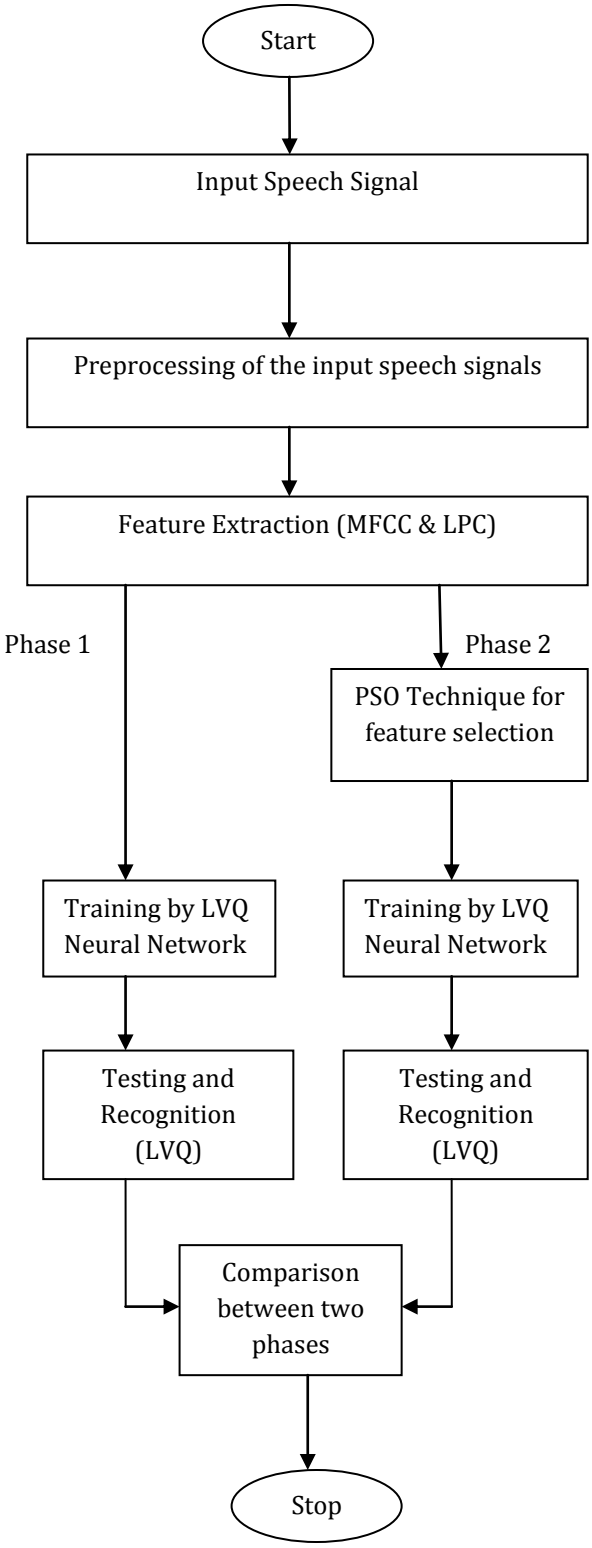

Fig-2: Steps of the Methodology

2.1Feature Extraction

In machine learning, speech and pattern recognition feature extraction plays a key role. In speech recognition it is the key step so that it is necessary to keep more attention to feature extraction procedure. Feature Extraction reduces the amount of resources to show a large set of the input data and gives the compact view of the input speech signal. It also converts the audio signal into numerical forms, these features values are applied to the Learning Vector Quantization (LVQ) neural network for further procedure. In this proposed project MFCC [11] and LPC [1] are used for the feature extraction technique.

2.1.1. MFCC Feature Extraction

Mel frequency cepstral coefficients (MFCC) collectively make up an MFC. They are derived from a kind of cepstral illustration of the audio clip. The distinction between the cepstrum and also the mel frequency cepstrum is that within the MFC, the frequency bands are unit equally spaced on the mel scale, which approximates the human additive system's response more closely than the linearly-spaced frequency bands utilized in the traditional cepstrum. MFCC [11] features are obtained by applying following steps:

1. At first the input signals are read by the system then it send for preprocessing at 8 KHz frequency.
2. In frame blocking procedure signals are divided into frames, overlapping of frames is totally avoided.
3. For the windowing procedure Hanning window is used, it avoids the discontinuities of the signals.For this process Hanning window [11] is used, the expression is given by:

$$W(n) = .5(1 - \cos\frac{2\pi n}{N-1})$$

4. Taking the Fast Fourier Transform of the signal is the next process. In the mel frequency wrapping procedure the FFT spectrum wrapped according to the mel scale. Mel frequency is calculated by following formula

$$Fmel = 2595 * \log 10(1 + \frac{F}{700})$$

5. Taking the log of the signal is the next process.
6. After the Inverse DFT we can obtain the MFCC value. MFCC feature extraction are the compact view of the signals, it is the numerical value of the applied input signal. These MFCC values are applied for the further training and testing process.

The flow chart of MFCC analysis is given in Fig-3.

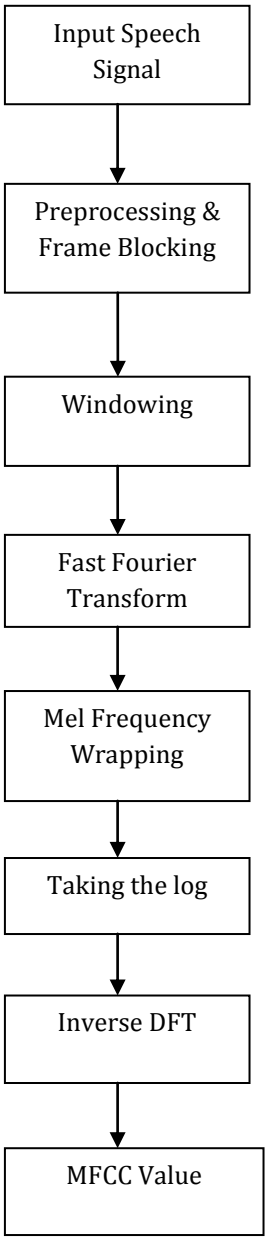

Fig-3: Flow Chart of MFCC Analysis

2.1.2 LPC Feature Extraction

LPC analysis consists of Pre-emphasis; frame blocking, Hamming Window, Auto Correlation analysis. In frame blocking procedure input signals are divided into frames and frame overlapping is avoided in this part. On the basis of best auto correlation value the LPC coefficient are selected. LPC feature extraction is also a mostly used feature extraction technique for the speech recognition system, when LPC feature extraction technique applied on the input speech signals then the signal converted into the numerical value. For windowing process Hamming window is used. The block diagram for the LPC analysis is given in Fig-4.

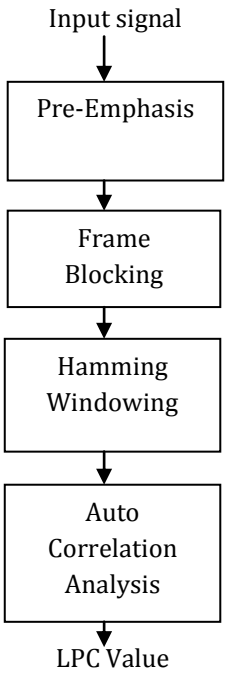

Fig-4: Block Diagram of LPC Analysis

2.3 Learning Vector Quantization (LVQ) Neural Network for Training & Testing

The architecture of the LVQ neural network is same as the Kohonen Self Organizing Map except that supervised learning is used in the LVQ networks. LVQ networks is a competitive neural network and in LVQ output is known so that the winner neuron output is always compared with the desired output and depending upon the comparison the weight modification is done. It is similar to the single layer feed forward network. The architecture is shown in Fig-5

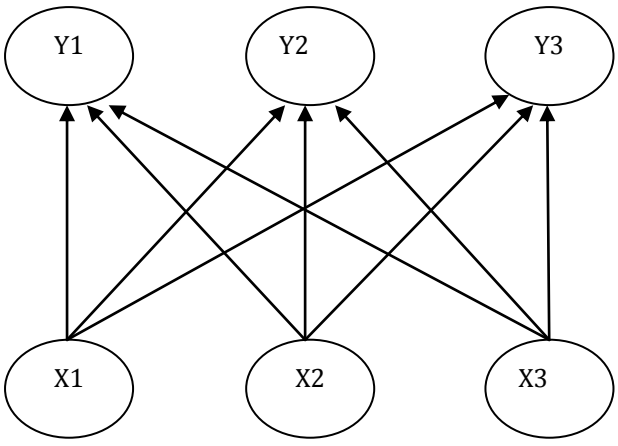

Fig-5: LVQ Neural Network Architecture

As shown in Fig- 2 the speech recognition and language identification is performed without PSO technique in phase1 but in the phase2 PSO technique is used after the MFCC and LPC feature extraction process to select the best iteration value. Phase1 and Phase2 are quite similar to each other except that in Phase2 PSO technique is used, PSO is very advance optimization technique and it very much similar to the Genetic Algorithm (GA).

2.4 Particle Swarm Optimization (PSO) techniques

Particle Swarm Optimization (PSO) is a stochastic optimization technique which is developed in 1995, this technique is inspired by bird flocking. This technique is developed by Dr. Ebehart and Dr. Kennedy. PSO technique shown in Fig- 6 which consists of three steps:

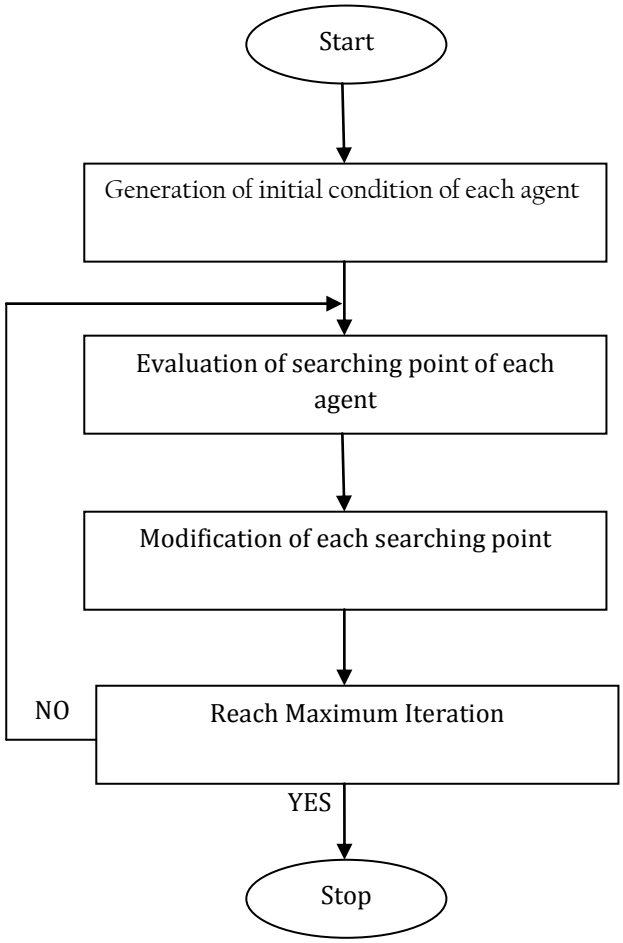

Fig- 6: Particle Swarm Optimization (PSO) flow chart

3. RESULT & DISCUSSION

3.1 Experimental Settings

For this proposed research work Database of the Bengali, Chhattisgarhi, English and Hindi languages are necessary. These databases are recorded with the recorder; the sampling frequency of recording is 44.1 KHz. The samples of 20 persons have taken for the observation, the recorded

sentences are "Ekhone Tumi Jao", "Ae Bar Teha Ja", "Now This Time You Go" and "Ab Is Bar Tum Jao", and these sentences are from Bengali, Chhattisgarhi, English and Hindi languages respectively. After the recording each word of the sentence has separated with the help of "AUDACITY" tool, now the each word is ".wav" file and it can be easily loaded to the MATLAB for further processing. For the training and testing of input speech signals Radial Basis Function Neural Network (RBFNN) is used. For training there are total 1020 samples for speech recognition there are total 800 samples for Language Identification, the total numbers of testing samples are 340 for both process.

3.2 Multilingual Speech Recognition Result

The performance of various methods can be evaluated by considering the following two points.

1. Recognition Rate (RR): Recognition Rate is the ratio of total numbers of recognized signals to the total numbers of applied signals for speech recognition. It can be given by the following expression-

$$RR = \frac{\text{Number of recognized signals} * (100)}{\text{Total Number of signals}}$$

2. Percentage of error (PE): If the actual output is different from the desired output then the error occurs. The Performance graph of Epochs value is shown in Fig-7:

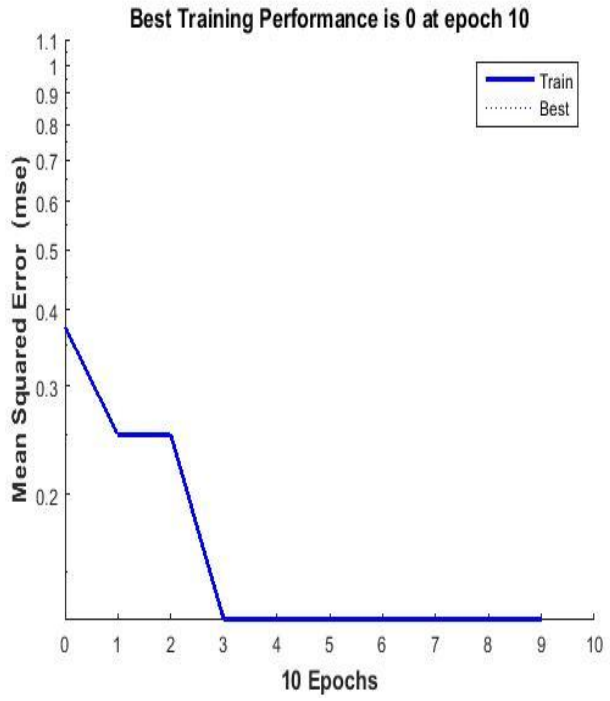

Fig- 7: Performance Graph of epochs value

On the basis of "Recognition Rate" and "Percentage of error" the result of the both phases is given in Table 1.

Table-1: Comparison of both methods of Speech Recognition

Methods	Recognition Rate	Percentage of error
1.Multilingual Speech Recognition with LVQ neural network (Phase1)	90%	8 to 10%
2. Multilingual Speech Recognition with LVQ neural network and PSO technique (Phase2)	92%	6 to 8%

3.3 Language Identification (LID) Result

On the basis of "Recognition Rate" and "Percentage of error" the result of the Language Recognition of both phases is given in Table 2. The recognition rate is reaches up to sufficient level. The Performance graph of Epochs value is shown in Fig-8:

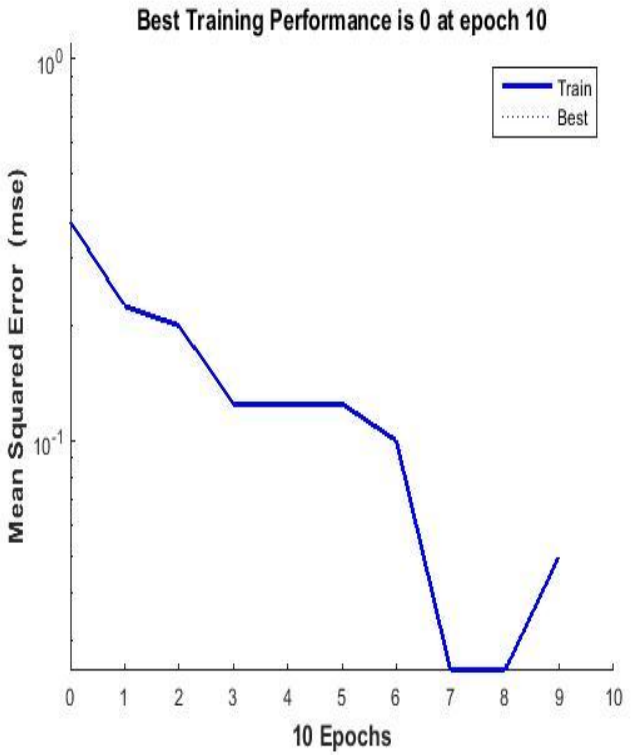

Fig- 8: Performance Graph of epochs value

Table-2: Comparison of both methods of Language Identification

Methods	Recognition Rate	Percentage of error
Language Identification with LVQ neural network (phase1)	88%	10 to 12%
Language Identification with LVQ neural network and PSO (phase2)	90%	8 to 10%

4. CONCLUSIONS

Multilingual Speech recognition gives recognition rate of 90% without PSO technique and Speech recognition gives recognition rate of 92% with PSO technique. The percentage of error varies from 8 to 10%.

Language Identification is also performed for Bengali, Chhattisgarhi, English and Hindi speech signal. LID gives recognition rate of 88% without PSO technique and LID gives recognition rate of 90% with PSO technique. The percentage of error varies from 8 to 10%.

So we can conclude that Multilingual Speech Recognition with PSO gives good result as compare to without PSO, similarly in Language Recognition with PSO gives good result as compare to without PSO.

REFERENCES

[1] Neelima Rajput and S.K. Verma, "Back Propagation Feed forward neural network approach for Speech Recognition", Department of C.S.E, GBPEC, Pauri Gharwal, Uttrakhand, India, IEEE 2014

[2] Behi Tarek, Arous Najet, Ellouze Noureddine , "Hierarchical Speech Recognition system using MFCC feature extraction and dynamic speaking RSOM", Laboratory of Signal, Image and Information Technologies National Engineering school of tunis, Enit Université Tunis El Manar, Tunisia, IEEE 2014

[3] Burcu Can, Harun Artuner, "A Syllable-Based Turkish Speech Recognition System by Using Time Delay Neural Networks (TDNNs)", Burcu Can, Harun Artuner Department of Computer Engineering Hacettepe University Ankara, Turkey, IEEE 2013

[4] Mohamed ETT AOUIL Mohamed LAZAAR Zakariae EN-NAIMANI, "A hybrid ANN/HMM models for arabic speech recognition using optimal codebook", Modelling and Scientific Computing Laboratory, Faculty of Science and Technology, University Sidi Mohammed ben Abdella Fez, MOROCCO, IEEE2013

[5] Ossama Abdel-Hamid Abdel-rahman Mohamed Hui Jiang Gerald Penn, "Applying Convolutional Neural Networks Concepts To Hybrid NN-HMM Model For Speech Recognition", Department of Computer Science and Engineering, York University, Toronto, Canada, IEEE 2012

[6] Anup Kumar Paul ,Dipankar Das, Md. Mustafa Kamal, "Bangla Speech Recognition System using LPC and ANN", Dhaka City College, Dhaka, Bangladesh, IEEE 2009

[7] Md Sah Bin Hj Salam, Dzulkifli Mohamad, Sheikh Hussain Shaikh Salleh, "Temporal Speech Normalization Methods Comparison in Speech Recognition Using Neural Network.", Comp. Science and Info. System University Technology Malaysia 81300 Skudai, Johor, Malaysia, IEEE 2009

[8] Purva Kulkarni, Saili Kulkarni, Sucheta Mulange, Aneri Dand, Alice N Cheeran, "Speech Recognition using Wavelet Packets, Neural Networks and Support Vector Machines.", Department of Electrical Engineering Veermata Jijabai Technological Institute Mumbai, India, IEEE 2014

[9] Javier Gonzalez-Dominguez, David Eustis, Ignacio Lopez-Moreno, Francoise Beaufays, and Pedro J. Moreno ,"A Real-Time End-to-End Multilingual Speech Recognition Architecture.", IEEE 2014

[10] Niladri Sekhar Dey, Ramakanta Mohanty, K. L. chugh et al [10] proposed "Speech and Speaker Recognition System using Artificial Neural Networks and Hidden Markov Model.", IEEE 2014

[11] Pialy Barua, Kanij Ahmad, Ainul Anam Shahjamal Khan, Muhammad Sanaullah "Neural Network Based Recognition of Speech Using MFCC Features.", 2Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh, IEEE 2014

[12] Oscal T.-C. Chen, Chih-Yung Chen ,"A Multi-lingual Speech Recognition System Using a Neural Network Approach.", Computer & Communication Research Laboratories, Industrial Technology Research Institute,Hsinchu, Taiwan, R.O.C., IEEE 1996

[13] G. Rigoll, c. Neukirchen "A New Approach to Hybrid HMM/ANN Speech Recognition Using Mutual Information Neural Networks.", Gerhard-Mercator-University Duisburg Faculty of Electrical Engineering, Department of Computer Science Bismarckstr. 90, Duisburg, Germany, IEEE 1995