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

Multilingual Speech Recognition Using Radial Basis Function (RBF) Neural Network

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 - *"Automatic Speech Recognition" of audio signal is very useful now a days, for telecommunication, language recognition and speaker verification process it can be used. Speech Recognition can be applied to automation of houses, offices and telecommunication services. In this paper Speech Recognition & Language Recognition have done for Bengali, Chhattisgarhi, English and Hindi speech signals. The Bengali, Chhattisgarhi, English, Hindi speech signals are "Ekhone Tumi Jao", "Ae Bar Teha Ja", "Now This Time You Go" and "Ab Is Bar tum Jao" respectively. This method is mainly applied in two stages , in the first stage Speech Recognition and Language Recognition have done with Radial Basis Function neural Network (RBFNN) and in the second stage Speech Recognition and Language Recognition have done with the combination of the Particle Swarm Optimization (PSO) technique and RBFNN. For the feature extraction Mel Frequency Cepstral Coefficients (MFCC) & Linear Predictive Coding (LPC) is used. The system accuracy and performance is measured on the basis of "Recognition Rate" and amount of error. Multilingual Speech Recognition and Language Recognition with PSO feature selection technique gives the better Recognition Rate as compare to the without PSO feature selection technique*.

Key Words: **Multilingual Speech Recognition, Language Recognition, Linear Predictive Coding, Mel Frequency Cepstrum Coefficients, Artificial Neural Network, Radial Basis Function Neural Network, Automatic Speech Recognition, Particle Swarm Optimization**

1. INTRODUCTION

1.1 Speech Recognition

Speech recognition (SR) is the method in which the speech signals are analyzed by the system on the basis of knowledge base, and then recognize by the soft computing methods. In the soft computing method mainly ANN is used. For the speech recognition preprocessing, feature extraction, ANN training and ANN testing is the necessary process. It is also known as "automatic speech recognition" (ASR).

1.2 Multilingual Speech Recognition

Speech recognition of more than two languages can be done with soft computing technique like ANN, Fuzzy Logic etc this process is called "Multilingual Speech Recognition." Two or more languages are first analyzed by the system which is depending on system training. After analyze the signal it recognizes by the system in the testing process. The basic process in Multilingual Speech Recognition is feature extraction, training and testing of the speech signals, training and testing is done by ANN.

1.3 Artificial Neural Network (ANN)

ANN is the learning model, used for function approximation. The ANN is consists of many nodes which is also called the processing elements. The nodes are interred connected with weighted connection. The simple multilayer ANN is consists of input layer, hidden layer and output layer. The diagram of ANN is shown in Fig-1.

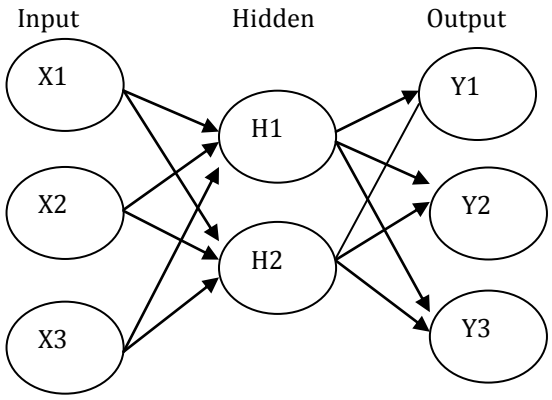

Fig-1: Artificial Neural Network

Various methods have been adopted for the speech recognition in literature; similarly for language identification many methods are adopted. Back Propagation Neural Network [1] with LPC [1] feature is used for English Alphabet recognition, Spiking Recurrent Self Organizing [2] used for the speech recognition with MFCC [2, 11] features. Time Delay Neural Network (TDNN) [3], hybrid ANN/HMM [4] model is also used for speech recognition which is the combination of Artificial Neural Network (ANN) and Hidden Markov Model (HMM) , in this both method MFCC features is used. 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.

This paper proposed Multilingual Speech Recognition and Language Recognition with RBFNN, PSO feature selection & without PSO feature selection. The paper is organized as follows methodology is given in section 2, section 3 is result and discussion, section 4 is conclusion and future scope.

2. METHODOLOGY

Multilingual Speech Recognition has wide range of application in many types of field; so that research work should be increase by applying some new methods. In past the speech recognition has done for English language, Turkish and Spanish languages, English alphabets, English digits and number etc. The applied methodology of this research work split up into two stages and comparison between these two stages. First stage is speech recognition and Language Recognition of Bengali, Chhattisgarhi, English and Hindi speech signal with Artificial Neural Network and the second stage is speech recognition and Language Recognition of Bengali, Chhattisgarhi, English and Hindi speech signal with the combination of Particle Swarm Optimization (PSO) technique and Artificial Neural Network. After applying these two methods the comparison has done based on recognition rate and error. In this proposed work Radial Basis Function Neural Network (RBFNN) is used for both speech and language recognition.

The flow chart of the methodology in given in Fig2, the methodology is divided into two stages. In first stage the signal is first read by the system then preprocessing is done, MFCC and LPC feature extraction are obtained after preprocessing, further the training and recognition is done by the RBFNN by training and test samples. Second stage is similar to the first stage except that in second stage after MFCC and LPC feature extraction Particle Swarm Optimization (PSO) feature selection technique is applied, rest all the processes are same in first and second stage. At last the comparison is done between these two stages on the basis of Recognition Rate and error. The complete methodology of this research work is given in Fig-2.

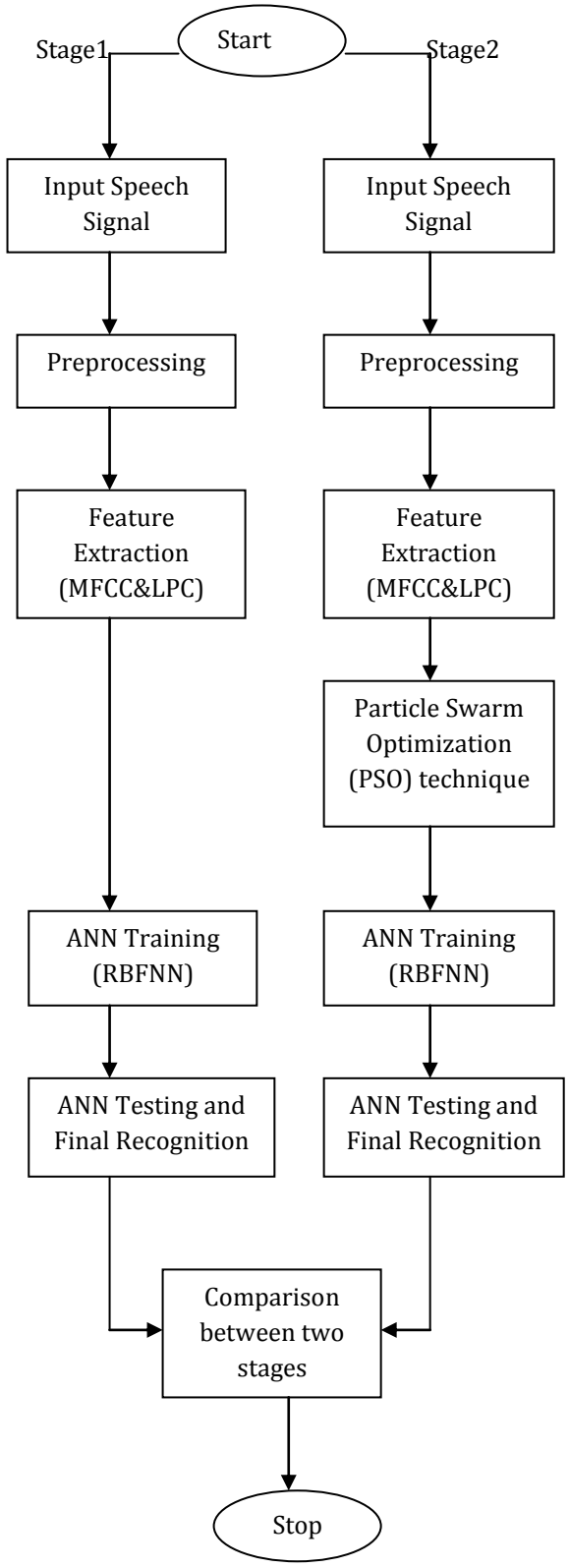

Fig-2: Steps of the Methodology

2.1 MFCC Feature Extraction

MFCC features are most widely used technique in the speech recognition process. After MFCC technique the audio signals are converted in the form of coefficients. The flow chart of MFCC feature extraction is given in Fig-3.

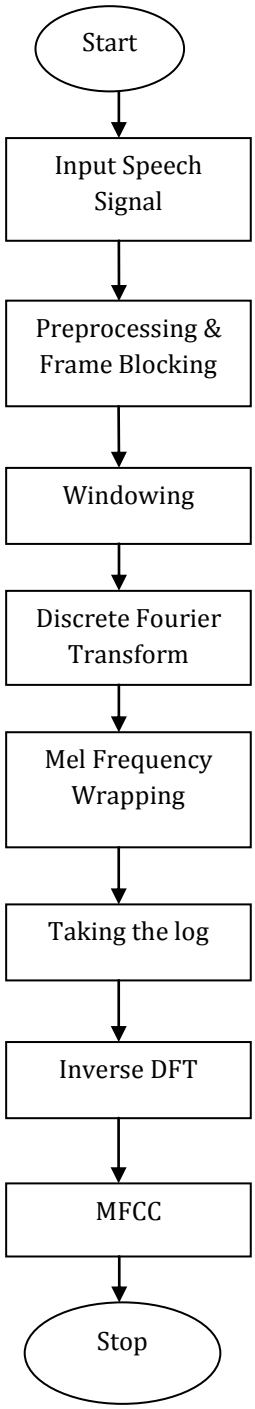

Fig-3: Flow Chart of MFCC Analysis

2.1.1 Preprocessing & Frame Blocking- When the input signals are read by the system then preprocessing of the signal is done at sampling frequency of 8KHz,filter is also used in this method. In frame blocking the speech signals are cropped to avoid the silence and the overlapping.

2.1.2 Windowing- After frame blocking the input signals are applied to the windowing to avoid the discontinuities of the speech signals. For this process Hanning window [11] is used, the expression is given by:

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

Where n is the number of samples, N is the length of the filter and W (n) is the Hamming window function.

2.1.3 Discrete Fourier Transform- It is the most common technique used in MFCC analysis, it converts the time domain signal into Frequency domain signal. The sequence of N complex numbers x1, x2…xn is transformed into an N-periodic sequence of complex numbers X0, X1…XN-1. DFT is given by

$$Xk = \sum_{n=0}^{N-1} xn * \varepsilon \frac{-j2\Pi Kn}{N}$$

Where K is the length of the filter and n is the number of samples.

2.1.4 Mel Frequency Wrapping- The spectrum which is obtained when DFT technique is wrapped according to the Mel Scale. Human perception of the frequency contents of sound does not obey the linear scale rule. So that for every tone with a frequency, F, measured in Hz, a subjective pitch is measured in to a scale called the "Mel" scale. The Mel frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Mel frequency is given by the following formula

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

2.1.5 Mel spectrum - The mel spectrum is obtained after mel frequency wrapping, after the log and Inverse DFT it converted back into time domain. MFCC feature extraction is the compact view of the signals, it is the numerical value of the applied input signal, it is very important process in speech recognition

2.2 LPC Feature Extraction

LPC analysis is considers as a strong Feature Extraction process of the input signal analysis to compute the main parameters of speech signals. LPC analysis consists of Pre-emphasis; frame blocking, Hamming Window, Auto Correlation analysis. 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. The block diagram for the LPC analysis is given in Fig-4.

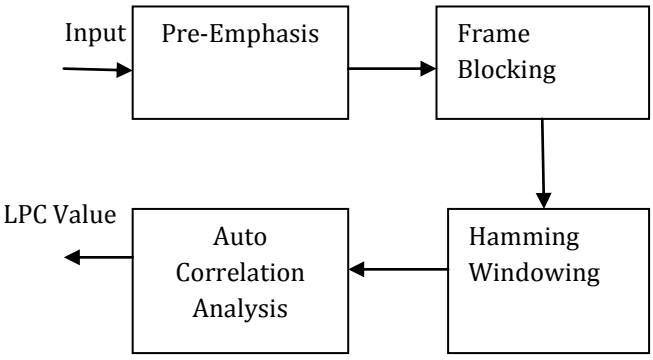

Fig-4: Block Diagram of LPC Analysis

2.2.1 Pre- Emphasis- Pre emphasis process is implementing on the speech signal before the feature extraction phase. In Pre emphasis the high frequency formants considered with lower amplitudes than low frequency formants. The main aim of pre emphasis is decrease the dynamic range of high spectra's.

2.2.2 Frame Blocking- The frame blocking is the process of segmentation of the speech signal in the frames. The speech signals over the segmented frames are assumed to be stationary with constant applied math properties. The continuity of the speech signal is preserved by overlapping the various frames. The amount of overlap samples controls the change in parameters from frame to frame to ensure a high correlation between LPC estimated coefficients.

2.2.3 Windowing- After the frame blocking, windowing is performed to suppress the energy of the frames at the edges. For preventing the rapid changes at the end points frame blocking reduce the discontinuities of each frame at the edges. Hamming widow [1] is widely used window with given function.

$$W(n) = .54 - .46\cos\frac{2\Pi n}{N-1}; 0 \le n \le N-1$$

2.2.4 Linear Predictive Coding (LPC) Analysis- In LPC the coefficients are derived by using the Levinson-Durbin algorithm. In Levinson-Durbin autocorrelation computation is used to calculate the highest autocorrelation value of the input data. The calculated highest autocorrelation value P, is termed as the order of the LPC analysis. LPC feature extraction is a very useful method for the speech recognition of the speech signals.

2.3 Training, Testing and Final Recognition by RBFNN

For the training and testing of input speech signals Back Propagation Artificial Neural Network (BPANN) is used. For training there are total 1020 samples for speech recognition there are total 800 samples for Language Recognition, the total numbers of testing samples are 340 for both process. RBF is also a multilayer feed forward neural network and it is widely used for the function approximation and pattern recognition. It used Gaussian Activation function to calculate the output response of the neurons. The basic architecture of the Radial Basis Function Network shown in Fig-5 , it made up of three layers namely input layer, hidden layer and output layer. X1, X2 and X3 are the input layer neurons, and Z1, Z2 and Z3 are the hidden layer neurons, and Y1, Y2 and Y3 are the output layer neurons. Between the input and hidden layers hypothetical connection is used and weighted connection is used between the hidden and output layer. The number of neurons can be increase in each layer as per requirement of the system; the selection of the neurons in the hidden layer is the complex work. In stage1 after the feature extraction training and testing of the input signals are done by the RBFNN, similarly in stage2 after the PSO technique the RBFNN is used for the training and the recognition procedure.

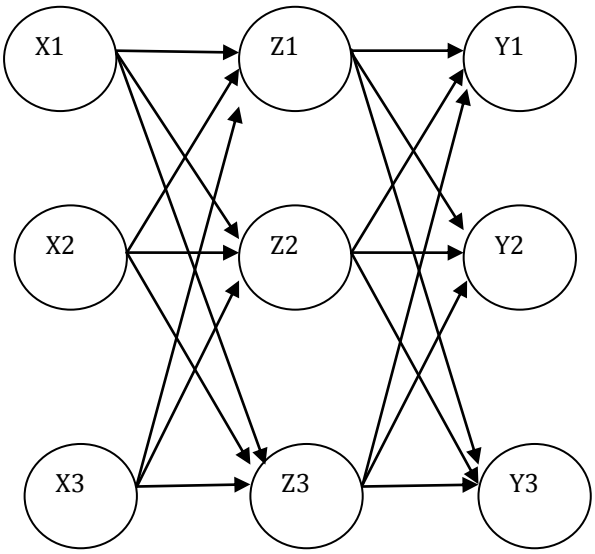

Fig-5: RBFNN Architecture

In the flow chart of the methodology we can see that Stage1 and stage2 is similar except that in stage2 Particle swarm Optimization (PSO) feature selection process is also used. After the PSO feature selection the training and testing is done in stage2.

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 is very much similar to the Genetic Algorithm (GA), PSO has some advantages over GA which are it is easy to implement and in PSO there are very few parameters to adjust. PSO has been applied in function optimization, ANN training, fuzzy system control etc. RBF neural network is used after the PSO technique for the training and the testing of the speech signals. RBF neural network is a feed forward type neural network and it is widely used for the speech recognition and the function approximation. RBF neural network gives the

good Recognition Rate and it reduces the error. The flow chart for PSO technique is given in Fig-6 which consists of many steps:

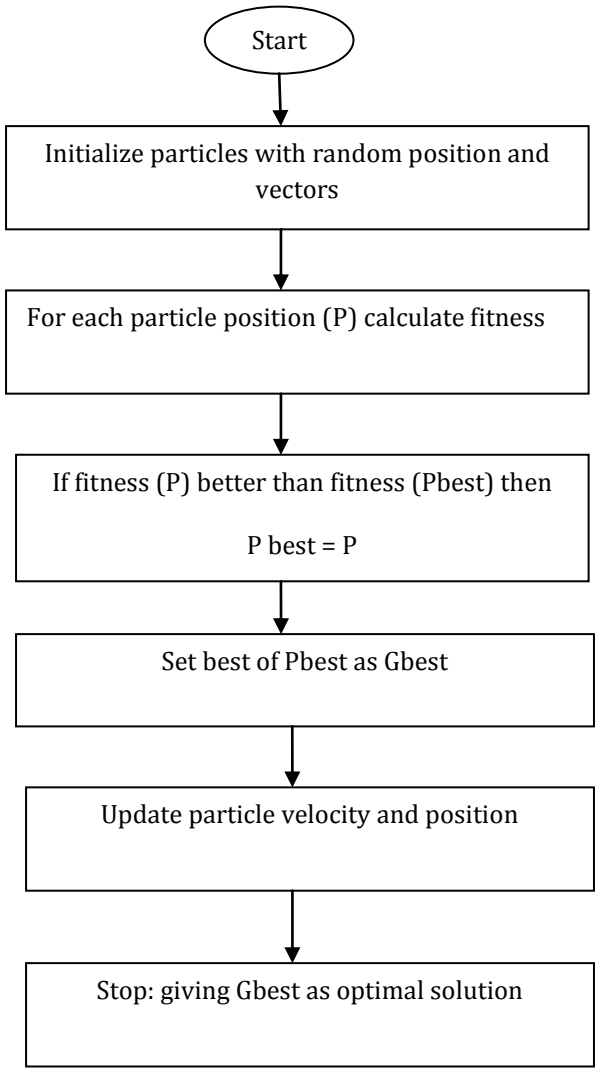

Fig-6: Flow Chart of PSO technique

3. RESULT & DISCUSSION

3.1 Experimental Setup

This research work requires the Database of the Bengali, Chhattisgarhi, English and Hindi language. These database or speech signals have recorded with the microphone at the sampling frequency of 44.1 KHz. Database has collected of 20 persons for these four languages. The sentence which is recorded by the each person's is "Ekhone Tumi Jao", "Ae Bar Teha Ja", "Now This Time You Go" and "Ab Is Bar Tum Jao", these sentences are of Bengali, Chhattisgarhi, English and Hindi languages respectively. After that each word of these sentences 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.

3.2 Multilingual Speech Recognition Result

The performance of various methods can be evaluated by considering the "Recognition Rate" and the "Percentage of error" of different Speech signals. 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.

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. For a good speech recognition system the recognition rate should be high and the percentage of error should be very less.

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 without PSO (Phase1)	93%	5 to 7%
2. Multilingual Speech Recognition with PSO (Phase2)	95%	3 to 5%

3.3 Language Recognition (LR) 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. 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.

Table-2: Comparison of both methods of Language Recognition

Methods	Recognition Rate	Percentage of error
Language Recognition (LR) without PSO (stage1)	92%	5 to 8%
Language Recognition (LR) with PSO (stage2)	94%	4 to 6 %

4. CONCLUSIONS

As we discussed above in this research work Multilingual Speech Recognition and Language Recognition has done with and without PSO technique. Each of this research work has two stages First stage is speech recognition of Bengali, Chhattisgarhi, English and Hindi speech signal with Artificial Neural Network and the second stage is speech recognition of Bengali, Chhattisgarhi, English and Hindi speech signal with the combination of Particle Swarm Optimization (PSO) feature selection and Artificial Neural Network. Speech recognition gives recognition rate of 93% without PSO technique and Speech recognition gives recognition rate of 95% with PSO technique. The percentage of error varies from 5 to 7%.

Language Identification is also performed for Bengali, Chhattisgarhi, English and Hindi speech signal. LID gives recognition rate of 92% without PSO technique and LID gives recognition rate of 94% with PSO technique. The percentage of error varies from 6 to 8%.

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