A Comparative Study on different Method of Speech Recognition

Rajat Haldar[1] and Pankaj Kumar Mishra[2]

[1]Electronics & Telecommunication Department
RCET, Bhilai
Bhilai (C.G.) India
haldarrajat12@gmail.com

[2]Electronics & Telecommunication Department
RCET, Bhilai
Bhilai (C.G.) India
pmishra1974@yahoo.co.in

Abstract: ***In this paper we will present approach or technique for the "Multilingual Speech Recognition". A lot of methods are available for Multilingual Speech Recognition like LPC(linear predictive coefficient) with back propagation feed forward neural network, SOM (Self Organizing Map) with hybrid ANN/HMM, LVQ (Linear Vector Quantization) with ANN, LPC using ANN etc. Multilingual Speech Recognition is a wide research area. It can be used in language identification, acoustic phonetic decoding, language modeling etc. The keen interest on Multilingual Speech Recognition arises on the fact that there are many language in India which is used by the different people like Hindi, English, Bengali, Sanskrit etc. So for different Language identification we can use Multilingual Speech Recognition technique. The main problem arises in this fact that the some language achieves higher recognition rate than the other languages, in this case the desired output and the actual output is different. The proper operation of the system can be measure on the bases of recognition rate. As the result outcome from the different approaches is satisfactory to the required level, The system using a hybrid ANN/HNN model and back propagation feed forward Artificial Neural Network gives the recognition rate very high.***

Keywords: *Multilingual Speech Recognition, Linear Predictive Coefficients Mel Frequency Cepstral Coefficients, Self Organizing Map, Artificial Neural Network, Linear Vector Quantization, Hidden Markov Model, Speech Recognition*

1.Introduction

Automatic Speech Recognition (ASR) is a broad research area and it is widely used by the research community. Robust speech recognition systems can be applied to automation of homes, workplace or business, observation of producing processes, automation of telephone or telecommunication services, written materials of specialized medical reports and development of aids for the disable. Continuous speech recognition systems find applications in voice process dialing wherever eyes free, hands free dialing of numbers is viable. The researches on "Multilingual Speech Recognition" Systems increases as a result of there are many languages in India for example Hindi, English, Sanskrit, Bengali etc. So for identification of the language and to use in a proper manner we have the need of the Multilingual Speech Recognition. Hence the researcher is taking keen interest in this research area. The error is distinction of original output and desired output computed on the premise of gradient descent method. The functioning of the system is evaluated on the premise of recognition rate. There are different method for Multilingual Speech Recognition like LPC(linear predictive coefficients) with back propagation feed forward neural network, SOM(Self Organizing Map) with hybrid ANN/HMM, LVQ(Linear Vector Quantization) with ANN, LPC using ANN etc.

The organization of the paper is as follows literature survey is given in section; section 3 gives conclusion of literature review. Section 4 gives an idea about problem related to Speech Recognition. Section 5 different methods related to Speech Recognition. Section 6 gives the result of various Speech Recognition techniques which are reviewed and section 7 concludes the paper.

2. Literature Review

Many works have been done in the past in the field of Speech Recognition and various methods were adopted for Speech Recognition. Neelima Rajput and S.K.Verma et. al [1] proposed "Back Propagation Feed forward neural network technique for Speech Recognition." Linear Predictive coefficients (LPC) are used for Feature extraction of input audio signals. Back propagation (BP) neural network used for training. The result shows that 94% accuracy in the recognition rate using proposed system.

Behi Tarek, Arous Najet, Ellouze Noureddine et. al [2] proposed "Hierarchical Speech Recognition system using MFCC features and dynamic speaking RSOM." In this paper, there is proposed a new type of unsupervised and competitive learning algorithms used to deal with temporal sequences. The first variant named Hierarchical Dynamic recurrent spiking self-organizing map (HD-RSSOM). The second variant is a hierarchical model which represents a multi-layer of HD-RSSOM model. Multi-layered HD-RSSOM reach good recognition rates in the range of 80 and 95% in training and test set.

Burcu Can, Harun Artuner et. al [3] proposed "A Syllable-Based Turkish Speech Recognition System by Using Time Delay Neural Networks (TDNNs)." This paper, presents a model for Turkish Speech recognition. The model is syllabus-based, where the recognition is done by syllabus as speech recognition units. In this paper accuracy is %65.6 on large vocabulary continuous speech. In addition, they define an algorithm for the automatic detection of syllabus edge which provides an accuracy of 44%.

Mohamed EttaouiL, Mohamed Lazaar, Zakariae En-naimani et. al [4] proposed "A hybrid combination of ANN/HMM models for Arabic digit recognition using optimal codebook." In this paper they propose an Arabic numbers recognition system based on hybrid combination of Artificial Neural Network and Hidden Markov Model (ANN/HMM). With a neural network of size 34, the recognition rate is equal to 84%, with 36 neurons is equal to85% and with 48 neuron rate is equal to 86%.

Osama Abdel Hamid, Abdel Rahman Mohamed, Hui Jiang, Gerald Penn et. al [5] proposed "Applying convolution Neural Networks concept to hybrid NN-HMM Model for speech recognition." In this paper, they propose to apply CNN to speech recognition within the framework of hybrid ANN/HMM Model. In this case, the recognition error is reduced from 22.57% to 20.1%.

Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal et. al [6] proposed "Bangla Speech Recognition System using LPCC and ANN." The speech processing stage are speech starting and end point detection, windowing, filtering, calculating the Linear Predictive Coefficients (LPC) and Cepstral Coefficients and finally constructing the codebook by vector quantization. The second process is pattern recognition system using Artificial Neural Network (ANN). The Bangla words have been recognized with satisfactory level of accuracy.

Md Sah Bin Hj Salam, Dzulkifli Mohamad, Sheikh Hussain Shaikh Salleh et. al [7] proposed "Temporal Speech standardization strategies comparison in Speech Recognition victimization Neural Network." These papers compares three strategies for speech temporal standardization namely the linear extended linear and nil soft normalizations on keep apart speech using different sets of learning amount on multi layer perceptron neural network with adaptional learning

Purva Kulkarni, Saili Kulkarni, Sucheta Mulange, Aneri Dand, Alice N Cheeran et. al [8] proposed "Speech Recognition using Wavelet Packets, Neural Networks and Support Vector Machines.". These feature sets are compared to the results from MFCC and within the second methodology, a feature set is obtained by concatenating completely different levels that carry significant data, obtained after wavelet packet decomposition of the signal. For feature matching Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are used as classifiers. Experimental results show that the proposed strategies improve the recognition rates.

Javier,Francoise Beaufays, and Pedro J. Moreno et. al [9] proposed "A Real-Time throughout Multilingual Speech Recognition Architecture." In this paper, present the throughout multi-language ASR structure, build up and deployed that allows users to select arbitrary combinations of spoken languages.

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." They propose a methodology to identify speaker and detection of speech. Additionally recognition of speaker using Hidden Markov Model also will be presented in this paper.

Barua, Kanij Ahmad, Ainul Anam Shahjamal Khan, Muhammad Sanaullah et. al [11] proposed "Neural Network Based Recognition of Speech Using MFCC Features." This paper represents the results from a preliminary study to acknowledge the speech from man voice using mel-frequency cepstrum coefficients (MFCC) features.

Oscal T.-C. Chen, Chih-Yung Chen et. al [12] proposed "A Multi-lingual Speech Recognition System Using a Neural Network Approach." A learning vector quantization method based on the dynamic time warp theme is planned for the speech recognition. The recognition accuracy for different users is improved by adapting the speech info victimization using the learning vector quantization method.

G. Rigoll, c. Neukirchen et. al [13] proposed "Novel Approach to Hybrid HMM/ANN Speech Recognition victimization Mutual data Neural Networks." This paper presents a brand new approach to speech recognition with hybrid HMM/ANN technology.

3. Conclusion of Literature Review

As we discussed for Speech Recognition different method can be applied. LVQ with ANN, LPC with ANN, the hybrid combination of the ANN/HMM and LPC with the back propagation feed forward Artificial Neural Network etc. The accuracy of the system is depending upon the "Recognition Rate." The Recognition Rate of the different method is varying according to the technique adopted in the system. LVQ with ANN and LPC with ANN method have good recognition rate. The recognition rate of this two method is satisfied to the desired level. If we are using hybrid combination of the AAN/HMM then the recognition rate is increase. It is increase up to 85%. The another method can be used for Multilingual Speech Recognition is the LPC with Back propagation feed forward artificial neural network. When we are applying this method then the Recognition Rate reaches up to 94%. Hence we can conclude that out of this method the Back Propagation Feed Forward Artificial Neural Network with Linear Predictive Coding technique has the better recognition rate. The BPANN (Back Propagation Artificial Neural Network) has recognition rate better than the PNN (Probabilistic Neural Network) also.

4. Problem Defination

The main problem arises in the speech recognition is that it has mainly done for the single language recognition for example English, Bengali, Turkish etc. If we apply Multilingual Speech Recognition for more than one language then result can increase. For the feature extraction either MFCC or LPC is used, no hybrid combination is used. If hybrid features extraction use then Recognition Rate may be increase up to satisfactory level. In speech recognition mainly BPNN, hybrid ANN/HMM, RSOM is used we can also apply RBF (Radial Basis Function) Neural Network for Multilingual Speech Recognition. The another problem is that for different language the "Recognition Rate" varies and for some language the recognition rate is not up to the desired level.

5. Methodologies

5.1 Speech Recognition with Back Propagation Neural Network

Speech recognition of the English Alphabet with Back propagation Neural Network is proposed. The input is in the form of LPC coefficients. In the proposed system neural network training is based on the error. The error is calculated from the difference of the desired output and actual output. The calculated error values are used to update the weight matrix of the neurons of the neural network. The input data sets used to train the neural network can be partitioned in to the independent weights. Back propagation neural network used in the system in following steps.

1. In the first step we choose and fix the architecture for the Back Propagation neural network, which comprised of input, Hidden and output layers, all units have their sigmoid functions value.
2. In the second step the weights among all the processing units are assigned.
3. Each input pattern is used in order to re-train the weights in the Back Propagation neural network.
4. In the last step the error is calculated for each weight of an input audio data, a termination condition is checked.

In Back Propagation neural network architecture the weights of input and hidden layers are adjusted according to the target output value. Back propagation neural network architecture shown in the following figure.

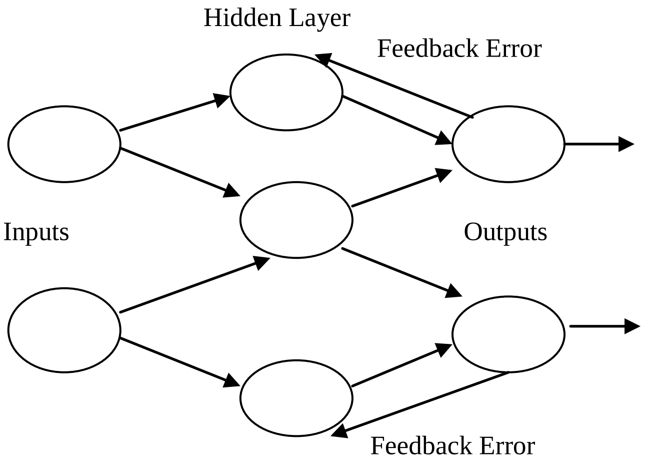


Figure. 1 Back Propagation Neural Network

In the above figure we can see that the Back Propagation Neural Network consists of three layers which is Input layer, Hidden layer and Output layer. Each layer is connected to each other with weighted connection. If the actual output is different than the desired output then the error occurs. For removing the error it is feedback to the hidden layer for the weight updating. After the weighted updating if the actual output is same to the desired output then there is no need of weight updating. The main steps of the Speech Recognition using Back Propagation Neural Network is Feature extraction of the input data, Training of the neural network and the testing of the neural network. For the feature extraction process LPC (Linear Predictive Coefficient) analysis is used. For training and testing process Back Propagation Neural Network is used which can feedback the error for updating the weight in the hidden layer. In proposed system Back propagation neural architecture contains the all above units and the basic steps of proposed system are defined as follows. The data flow diagram of the proposed system shown in Figure 2.

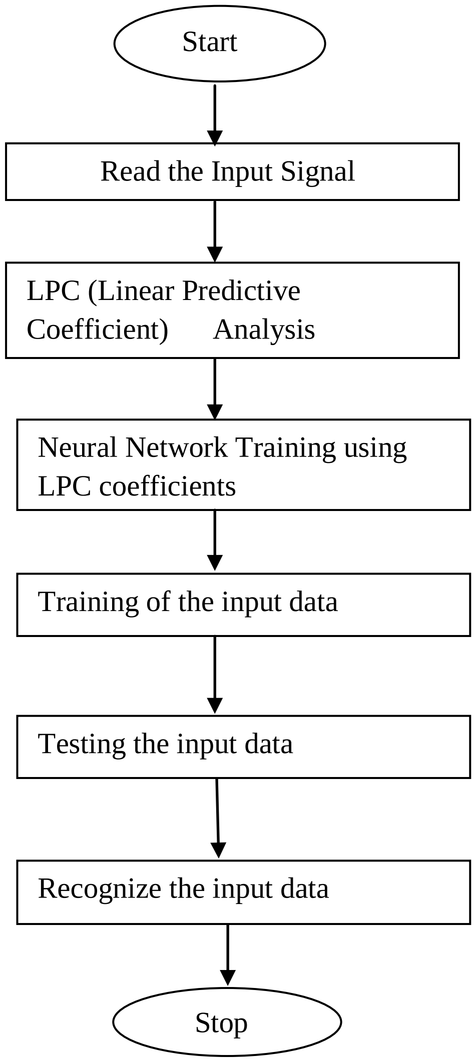
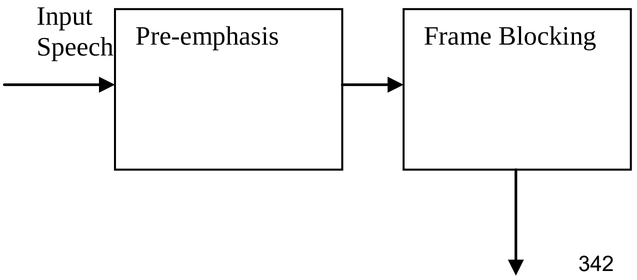

Start
Read the Input Signal
LPC (Linear Predictive Coefficient) Analysis
Neural Network Training using LPC coefficients
Training of the input data
Testing the input data
Recognize the input data
Stop

Figure. 2 Flow Diagram of proposed work

5.1.1. Linear Predictive Coefficient (LPC) Analysis

LPC analysis is considers as a strong Feature Extraction process of the input signal analysis to compute the main parameters of speech signals. It is passive feature extraction technique and it encodes speech at low bit rate and also provides the accurate estimates of speech parameters of input Speech signal. The basic idea of LPC feature extraction process is to estimate an input speech sample as a linear Combination of past speech sample. LPC analysis consists of Pre-emphasis; frame blocking, Hamming Window, Auto Correlation analysis. The block diagram of the LPC analysis shown in the Figure 3.



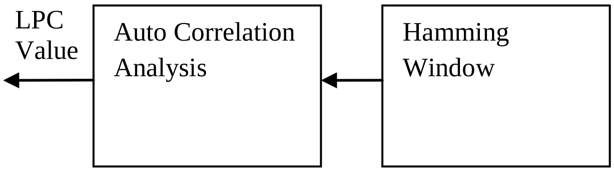


Figure. 3 LPC Analysis

5.2 Speech Recognition Using MFCC Feature Extraction and Neural Network

This method represents the results from a preliminary study to recognize the speech from voice using mel-frequency cepstrum coefficients (MFCC) features. Result of matching features in a neural network demonstrates that MFCC features work significantly to recognize speech. The MFCC features are extracted from each recorded voice. After that, training data-table is created using MFCC feature data of two persons (recognized user and any one of other users) and target data-table also created as back-propagation neural network was used. Built-in Artificial Neural Network (ANN) is trained with these. Finally, when the test data is given, the ANN compares the test data with train data. The block diagram of the recognition is shown in Figure 4.

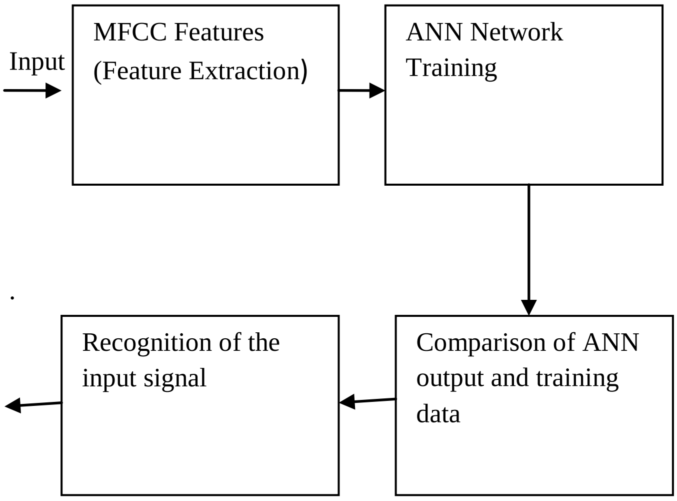

Input
MFCC Features (Feature Extraction)
ANN Network Training
Recognition of the input signal
Comparison of ANN output and training data

Figure.4 Block Diagram of the Recognition phase

5.2.1 MFCC feature extraction

Mel-frequency cepstral coefficients (MFCC) are coefficients that collectively make up an MFC. They are derived from a kind of cepstral illustration of the audio clip (a nonlinear "spectrum-of-a-spectrum"). 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. This frequency deformation gives higher illustration of sound, for example, in audio compression. MFCCs are unit ordinally derived as follows:

1. Take the Fourier transform of a signal.
2. Mapping the power spectrum obtained higher than onto the mel scale, mistreatment triangular overlapping windows.
3. Take the logs of the powers at every of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers.
5. The MFCCs are unit the amplitudes of the ensuing spectrum. The block diagram of the MFCC analysis shown in Figure 5.

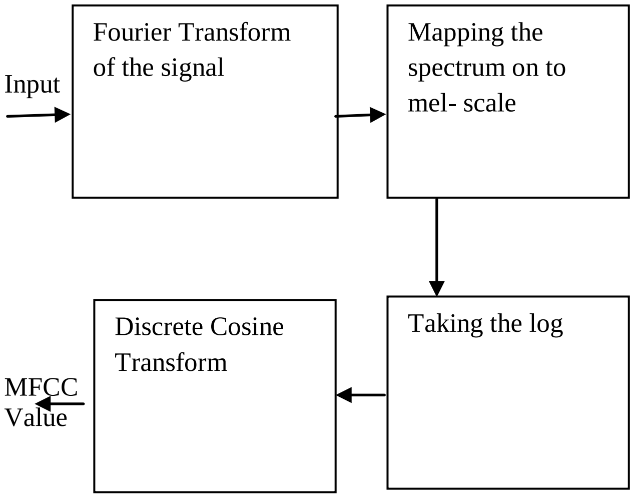

Input
Fourier Transform of the signal
Mapping the spectrum on to mel- scale
Taking the log
Discrete Cosine Transform
MFCC Value

Fig. 5 MFCC analysis

5.3 Speech Recognition using a hybrid ANN/HMM model

There is also the alternative of using Neural Networks. But another interesting framework applied in ASR is indeed the hybrid Artificial Neural Network (ANN) and Hidden Markov Model (HMM) speech recognizer that improves the accuracy of the two models. In the present work, we propose an Arabic digits recognition system based on hybrid Artificial Neural Network and Hidden Markov Model (ANN/HMM). In this proposed method MFCC feature extraction process is used.

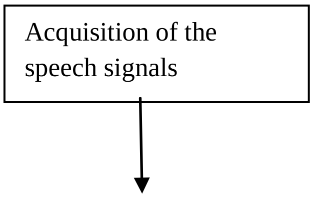
5.3.1. Hidden Markov Model (HMM)

The Hidden Markov Model could be a finite set of states, every of that is related to likelihood distribution. Transitions among the states are ruled by a group of chances called transition probabilities. A HMM is a random automaton with a random output method connected to each. A hidden Markov model (HMM) could be a applied mathematical Markov model within which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. A HMM are often given because the simplest dynamic theorem network. More precisely, an HMM is defined by:

A = (Q,O,P, B,N) ,Where

Q: Set of finite states of the model of cardinal N,
O: Set of finite observations of the model of cardinal M,
P: Matrix of state transition probability,
B: Matrix of probability to emit,
N: Initial distribution of states.

The flow chart of the proposed work is shown in Figure 6:



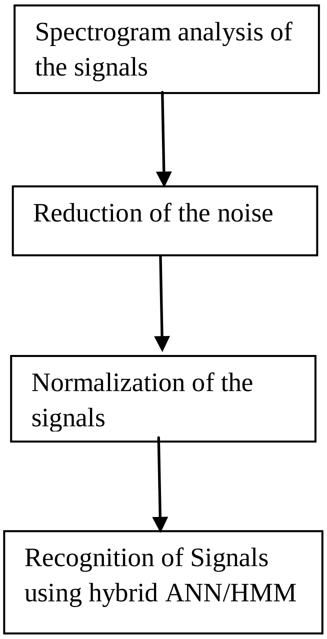

Spectrogram analysis of the signals
Reduction of the noise
Normalization of the signals
Recognition of Signals using hybrid ANN/HMM

Figure.6 Flow Chart of Recognition using ANN/HMM

6. Result

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

A. Recognition Rate(RR): It can be given by the following expression-

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

B. Percentage of error (PE): If the actual output is different from the desired output then the error occurs. Percentage of error can be defined as how much it different from the desired or target output.

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

Table.1
Comparison of different methods of Speech Recognition

Methods	Recognition Rate	Percentage of error
Speech Recognition with Back Propagation Neural Network with LPC feature extraction	93-94%	3- 4%
Speech Recognition with Artificial Neural Network with MFCC feature extraction	60-80%	20-22%
Speech Recognition with hybrid ANN/HMM model with MFCC feature extraction	84-86 %	6-7%

7. Conclusion

The Recognition Rate (RR) and Percentage of error (PE) for various methods is given by table. The RR by using Back Propagation method with LPC feature extraction is found to be 93-94%. The RR by using Artificial Neural Network (ANN) with MFCC features is 60-80%. The Recognition Rate of the hybrid ANN/HMM model with MFCC features is about 84-86%. So from the above table it is clear that the Recognition Rate with Back Propagation method and LPC features is greater than the remaining two methods. The hybrid combination of ANN/HMM gives Recognition Rate higher than the Recognition Rate of Artificial Neural Network (ANN) with MFCC features. So it can be concluded that the Recognition Rate is best in the Back Propagation method with using LPC features.

The percentage of error in the Back Propagation Method is very less. It is about 3-4 % which is less than the remaining two methods. The percentage of error in hybrid ANN/HMM model is about 6-7%. The percentage of error in Artificial Neural Network (ANN) with MFCC features method is 20-22%. So it can be concluded that the Percentage of error is least in the Back Propagation method.

Reference

[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 ArtunerDepartment 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