Asian Journal of Convergence in Technology Volume III, Issue III
ISSN No.:2350-1146, I.F-2.71

OPTIMIZE APPROACH TO VOICE RECOGNITION

Mr.Nitesh Purushottam Patel
Electronics and Telecommunication
Modern COE,
Pune, Indian
niteshpatel.dyp@gmail.com

Prof. Aparna P. Laturkar
Electronics and Telecommunication
Modern COE,
Pune, Indian
aparna.laturkar@moderncoe.edu.in

Abstract— **Speech is the most efficient way to train a machine or communicate with a machine. This work focuses on the objective to recognize the word or the phase spoken by human, keywords in high speed. Recognition systems based on hidden Markov models are effective under particular circumstances, but do suffer from some major limitations that limit applicability of ASR technology in real-world environments. However, over the last few years, several attempts have been undergone to evaluate the HMM deficiencies. Artificial Neural Networks (ANN) and more specifically multilayer Perceptron's (MLP) appeared to be a promising alternative in this respect to replace or help HMM in the classification mode. But ANNs were unsuccessful in dealing with long time sequences of speech signals. So taking the advantages of both the systems into consideration it was proposed to combine HMM and ANN within a single, hybrid architecture. The goal in hybrid systems for ASR is to take advantage from the properties of both HMM and ANNs, improving flexibility and ASR performance.**

Keywords— *Recognition, Hidden Markov Models (HMM), Artificial Neural Networks (ANN) and more specifically multilayer Perceptron's (MLP).*

Introduction

The process is initiated by capturing the signal using microphone. Algorithms are applied to reduce the noise interference and silence suppression .the signal free from above interference is then processed to extract the features. MFCC is used as a feature extraction technique. These features are used as an input to ANN & HMM systems. The results are obtained for speech & speaker recognition in Matlab. For ANN ' NNTOOL' proves to be more efficient and accurate for training & testing of speech samples. The parameters such as performance ,training state & Validation are evolved from this tool. The performance plot is the graph of average Mean Square Error (MSE) Vs No of Epochs which shows the behavior of Training & Testing data . furthermore the training state validation is also plotted. The features extracted from MFCC are now given to HMM for speaker recognition. In Hidden Markov Model, the speech samples are trained & compared with the test sample. The test speech sample is compared with all the trained stored speech samples and the best match is found. Finally speaker is recognized using HMM. The main advantage of this system is that the input signals need not be preprocessed again for HMM. The preprocessing steps done for ANN can be directly applied to HMM thus reducing the complexity. HMM works on the principal of Markov process which depends on probability of States & generates an Observation Sequence. This model consist of underlying states (Hidden States), Transition Probabilities(probabilities from one state to another).

I. HYBRID SYSTEM

A. Overview of Hybrid System

A brief Although HMM is effective approaches to the problem of acoustic modeling in ASR, allowing for good recognition performance under many circumstances, it also suffers from some limitations. In order to overcome this limitations of HMM in late 1980s, many researchers began to use artificial neural networks (ANNs) for ASR. ANN was expected to carry out the recognition task. In spite of their ability to classify short-time acoustic phonetic units, such as individual phonemes, ANNs failed as a general framework for ASR, especially with long sequences of acoustic observations like those required in order to represent words from a dictionary or whole sentences. This is mainly due to the lack of ability to model long-term dependencies in ANNs. In order to make the recognition more efficient & Accurate led to the idea of combining HMM and ANNs within a single, novel model, known as hybrid HMM/ANN.

B. Implementation of Hybrid System

Connectionist modeling of speech is the youngest development in Automatic Speech Recognition. This approach focuses on the representation of knowledge and integration of knowledge sources [11]. The Artificial Neural Network (ANN) Models are used for connectionist speech recognition but with limited success. ANN has a good discriminative power which provides the discrimination between the classes which helps in classification of phonemes [8]. In ANN the assumptions about the underlying data distribution need not be made. An Artificial Neural Network is a collection of simple processing elements, called units or nodes, which are connected to each other and organized in layers. Its functionality is based on the biological neuron. Even though ANN is a Parametric model, no assumptions about the underlying data distribution have to be made, which is a Contrast to HMM. In HMM the changes in the distribution can be modelled by an underlying process, which moves between different states .These states represent different output distributions [12,13,14]. The process that moves from state to state can be modelled with an ordinary Markov chain. Another drawback of HMM is that the assumptions about the underlying data distributions have to be made. The hybrid HMM/ANN system is an attempt to combine the strengths of both HMMs and ANNs, the temporal structure modeling of speech with HMM sequences and the discriminative abilities of the ANNs. Initially speech signal is captured and pre-processed. After acquisition spectrogram is computed & analyzed. The most common format is to carry out STFT (Short Time Fourier Transform) which is a graph with two geometric dimension as horizontal axis represents time, vertical axis is frequency & third axis represents amplitude. MFCC's are calculated which provides as an input for further processing [15,18] Generally 15 to 20 features are extracted from MFCC & this features are provided as an input to the ANN and HMM separately. The main advantage is that the preprocessed signal used for ANN can be used for HMM also. In this way making the system a hybrid architecture separate training and testing of data can be done with any of the two methods .

C. Process Flow

Here we propose a methodology to identify speaker and detection of speech. The Fig. 5.1. Demonstrates the process flow. In this approach the input speech signal is acquired at first. Most of the electronic recording equipment has an effect of noise on the recorded sound signal. But the captured sound signals are varies speaker to speaker by age, sex, anatomic variation and emotion. This added emotions & noise has to be neutralized otherwise it makes the system unstable in recognition of speech. When speech recognition is carried out, the minimum noise also can weigh down the process of neural network during training and processing. The signal is then neutralized which reduces the emotional effects of the speech signals with noises. After preprocessing the features are extracted with the help of feature extraction technique MFCC (Mel frequency cepstral coefficients). There are different types of feature extraction LPC (Linear Prediction Coefficient Cepstra), MFCC (Mel Frequency Cepstra Coefficient), PLP (Perceptual Linear Prediction Cepstra). Due to its advantage of less complexity in implementation of feature extraction algorithm, & good spectral smoothing MFCC's are used extensively in Automatic Speech Recognition (ASR)[17]. MFCC features are derived from the FFT magnitude spectrum by applying a filter bank which has filters evenly spaced on a warped frequency scale. The Mel-scale used in this work is to map between linear frequency scale of speech signal to logarithmic scale for frequencies higher than 1 kHz. This makes the spectral frequency characteristics of signal closely corresponding to the human auditory Perception [16]

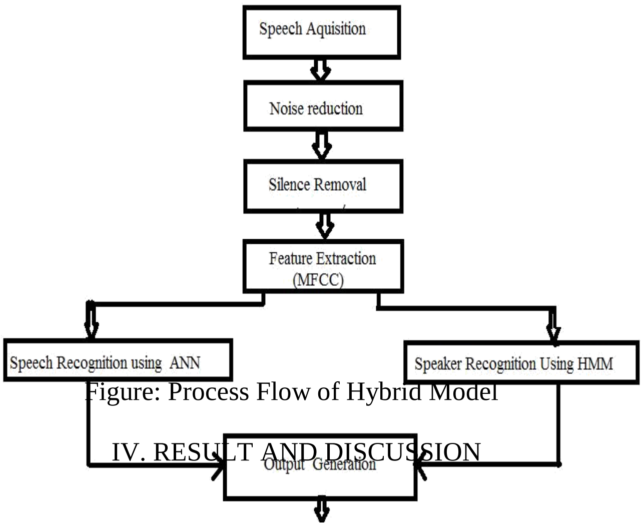

Speech Aquisition
Noise reduction
Silence Removal
Feature Extraction (MFCC)
Speech Recognition using ANN
Speaker Recognition Using HMM
Output Generation

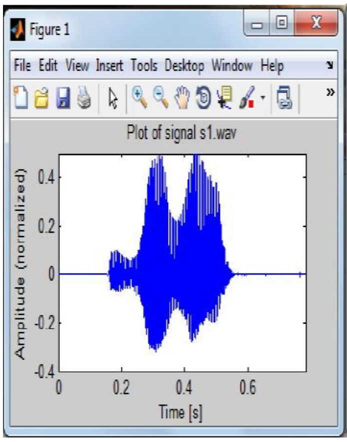
Figure: Process Flow of Hybrid Model

IV. RESULT AND DISCUSSION

The proposed method discussed earlier is simulated using MATLAB 7.0 Version. The results are obtained in two Approaches. At first the simulation results using programming in Matlab is implemented. Later on the same process is implemented using Simulink tool in Matlab. The tool named 'NNTOOL' is used for simulation and the performance results are obtained. After comparing both the approaches the results obtained from 'NNTOOL 'are more accurate .The simulation results for both the approaches are shown below

Approach using Matlab Programming -
In this Approach we have tried to recognize speech of users by storing the voice samples in database as well as accepting real time voice samples as an input to the system. One input can be considered at a time, this input is pre-processed & given to ANN. Sufficient no of samples for particular speaker are stored in the database. This is done for storing samples with different pitch, emotions etc. Once the samples are stored in database, they are trained and features are extracted using MFCC. Initial steps of pre-processing of the speech signal includes time domain representation, pre-emphasis as shown in Fig 6.1(A& B) respectively. After framing the signal is windowed with the help of Hamming window. Hamming window is mostly preferred over other windows because it has very wide main lobe & does not have end point discontinuities. The Fig 6.2 shows speech signal before and after windowing. Fast Fourier Transform is then computed to

convert time domain signal into frequency domain. The Fig 6.3 shows FFT plot.



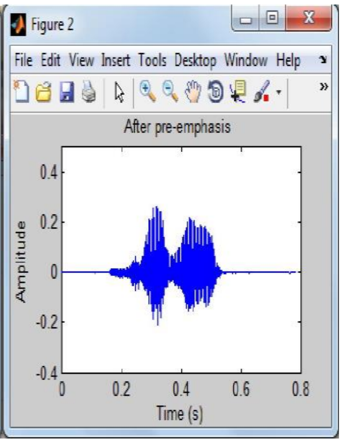
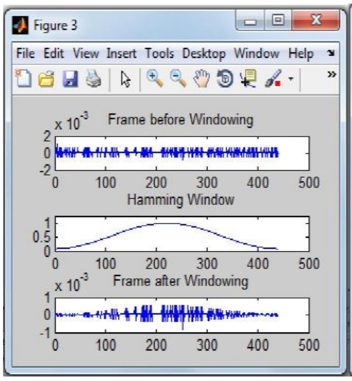


Fig 6.1(A) Time domain **Fig 6.1(B) pre-emphasis**



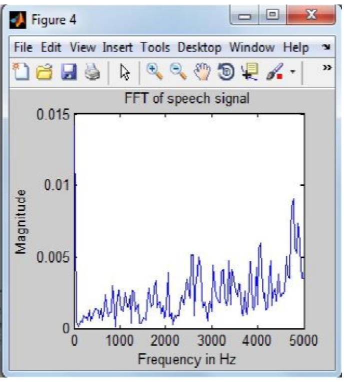


Fig 6.2 Windowing **Fig 6.3 FFT**

In Recognition Phase the test speech sample is taken as an input..After converting the signal in frequency domain MFCC is used as an feature Extraction process. The MFCC plot is shown in Fig 6.4 .The features extracted for stored speech samples are then compared with the features of speech input sample. During the comparison if input sample matches with the samples in the database then the system provides access to that particular speaker. On the contrary if input sample does not match with the samples stored in the database then access to that particular speaker will be denied by the system. The spectrogram is also plotted for the speech sample used in Recognition. The Fig. 6.5 (A&B) shows this illustration. The same procedure is repeated

for real time speech processing by recording the samples in real time with the help of microphone. This is how speech is recognized using ANN.

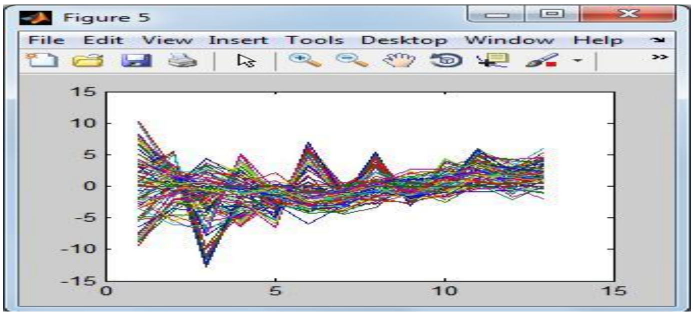
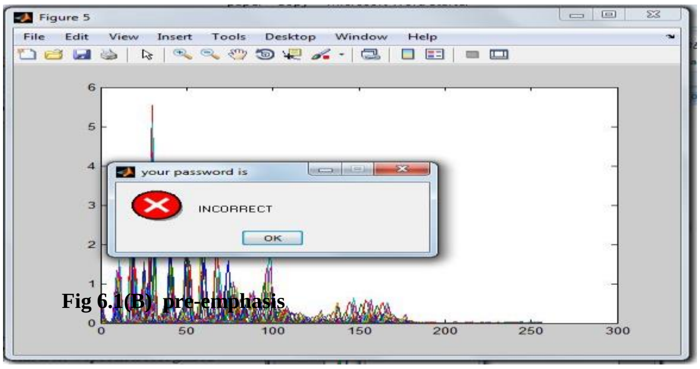


Fig 6.4 Mel frequency cepstral Coefficients(MFCC)



your password is

INCORRECT

OK

Fig 6.1(B) pre-emphasis

Fig 6.5(B) GUI for Access Denied

False Rejection Ratio (FRR) : It is one of important factor deciding the accuracy of any Biometric system .

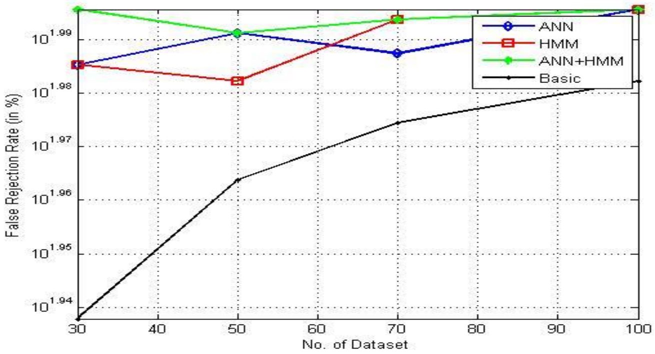


Fig 6.12 (False Rejection ratio) FRR

Mathematically it is defined as the ratio of No of Successful matches of Speech Samples to the Total no of Samples in the database. FRR is calculated for all the methods ANN, HMM, Hybrid(ANN/HMM) & Basic simultaneously and the results are obtained for different values of Datasets as shown in Table . fig shows the graphical representation of FRR versus No of samples in database

False Acceptance Ratio (FAR): This Parameter decides number of times the access to unintended user is provided to the system. Mathematically it is defined as the ratio of Difference between Total no of Samples & No of Successful matches of Speech Samples to the Total no of

Samples in the database. FAR is calculated for all the methods ANN, HMM, Hybrid(ANN/HMM) & Basic simultaneously and the results are obtained for different values of Datasets as shown in Table . fig shows the graphical representation of FRR versus No of samples in database

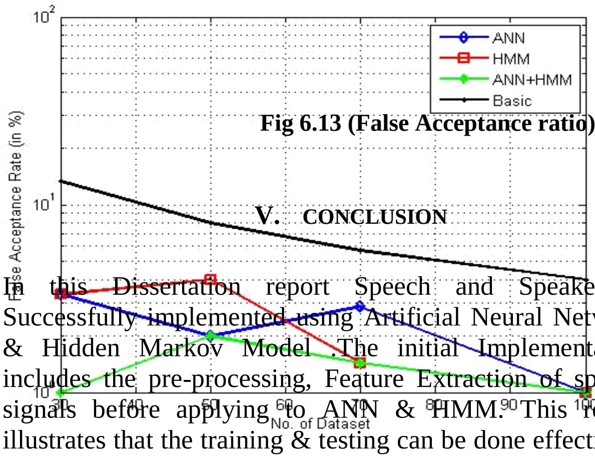


Fig 6.13 (False Acceptance ratio) FAR

V. CONCLUSION

In this Dissertation report Speech and Speaker is Successfully implemented using Artificial Neural Network & Hidden Markov Model .The initial Implementation includes the pre-processing, Feature Extraction of speech signals before applying to ANN & HMM. This report illustrates that the training & testing can be done effectively with the help of ANN , HMM & using Hybrid (ANN/HMM) system. The results obtained in table and graph for FRR & FAR clearly indicates that Hybrid system (ANN/HMM) proves to be Efficient & Accurate as Compared with other methods. The Accuracy of the system increases as the no of samples in the database increases.

REFERENCES

[1] Davis, K., Biddulph, R., and Balashek, S., "Automatic Recognition of Spoken Digit," J. Acoust. Soc. Am. 24: Nov 1952, p. 637.

[2] Wiqas Ghai Navdeep Singh " Literature review on Automatic Speech Recognition" International Journal of Computer Applications (0975- 8887)Volume 41- No.8,March 2012.

[3] H.F. Ong and A.M. Ahmad " Malay Language Speech Recogniser with Hybrid Hidden Markov Model & Artificial Neural Network (HMM/ANN)" International Journal of Information and Education Technology, Vol. 1, No.2, June 2011..

[4] Dominique Genoud, Dan Ellis and Nelson Morgan , "Combined Speech and Speaker Recognition with speaker-adapted connectionist models" International Computer Science Institute, 1947 Center St, Berkeley, CA 94704.

[5] Mondher Frikha, Ahmed Ben Hamida ," A Comparitive Survey of ANN and Hybrid HMM/ANN Architectures for Robust Speech Recognition" American Journal of Intelligent Systems 2012

[6] Jaume padrell-sendra, Dario Martin-Iglesias and Fernando Diaz-de-Maria ," Support Vector Machine for Continuous SpeechRecognition " Florence ,Italy.

[7] Zhong-Hua Quan, De-Shuang Huang, Kun-Hong Liu, Kwok-Wing Chau , "A Hybrid HMM/ANN Based Approach for Online Signature Verification " International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007

[8] Sabeur Masmoudi, Mondher Frikha, Mohamed Chtourou,Ahmed Ben Hamida, " Efficient MLP constructive training algorithm using a neuron recruiting approach for Isolated Word Recognition system" International Journal Speech Technology (2011).

[9] Xian Tan " Hybrid Hidden Markov Model and Artificial Neural Network for Automatic Speech Recognition" Pacific-Asia Conference on Circuits,Communications and System 2009 .

[10] Niladri Sekhar Dey, Ramakanta Mohanty,K.L. Chugh " Speech & Speaker Recognition System using Artificial Neural Networks and Hidden Markov Model" International Conference on Communication System and Network Technologies (2012).

[11] Nelson Morgan and Herve Bourlard ," An Introduction to Hybrid HMM/Connectionist Continuous Speech Recognition" Berkley, USA.

[12] Lawrence R. Rabiner , Fellow, IEEE, "A Tutorial on Hidden Markov model and Selected Applications in Speech Recognition"

[13] Mondher Frikha, Ahmed Ben Hamida and Mongi Lahiani ,"Hidden Markov models isolated word recognizer with optimization of acoustical analysis and modeling techniques" International Journal of the Physical Sciences Vol. 6(22), pp. 5064-5074, 2 October, 2011 .

[14] zica valsan, inge gavat and bogdan Sabac ," Statistical and Hybrid Methods for Speech Recognition in Romanian" International Journal of Speech Technology 5, 259–268, 2002.

[15] Anjali Bala" Voice Command Recognition System Based on MFCC and DTW" et al. / International Journal of Engineering Science and Technology Vol. 2 (12), 2010, 7335-7342.

[16] Ben J. Shannon, Kuldip K. Paliwal, " A Comparative Study of Filter Bank Spacing for Speech Recognition" Microelectronic engineering research conference 2003,

[17] " Speech Recognition using MFCC" Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012 Pattaya (Thailand)

[18] Vibha Tiwari, "MFCC & its Application is Speaker recognition International Journal on Emerging Technologies" 1 (1): 19-22(2010).

[19] R. Rojas " Back Propagation Neural Networks" Springer Verlag, Berlin ,1996

[20] "Speech recognition using Back Propagation Algorithm" 1991 TECNON IEEE Region10 International Conference on Energy ,Comminication and Control.

[21] D.B Paul,"Speech recognition using HMM" Volume 3 ,1990 lincoln Laboratory Journal

[22] "Levenberg–Marquardt Training" Hao Yu Auburn University Bogdan M. Wilamowski Auburn University