International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 6

ISSN: 2321-8169
4154 – 4157

# Speech Recognition Using HMM/ANN Hybrid Model

Supriya S. Surwade
M.E.(Purs.),JSPM'S BSIOTR ,Wagholi, Pune
*supriya77surwade@gmail.com*

Dr. Y.S. Angal
Phd., Shri Guru Gobind Singhji Institute Of Engineering and Technology, Nanded
*yogeshangal*@yahoo.co.in

***Abstract***- By the analysis on the principle of speech recognition system, a speech recognition system was designed by using LPC2148 as the hardware platform and MATLAB 2012 as the software platform. Speech recognition is an important component of biological identification which is an integrated technology of acoustics, signal processing and artificial intelligence. 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. Attempts were made to overcome these limitations with the adoption of artificial neural networks as an alternative paradigm for ASR, but ANNs were unsuccessful in dealing with long time sequences of speech signals. So taking the limitations and advantages of both the systems 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 For Speech recognition features from speech sample are extracted & mapping is done using Artificial Neural Networks. Multilayer pattern mapping neural network, which works on the principle of back propagation algorithm is proposed. Finally Speaker Recognition is done using Hidden Markov Model (HMM). The specialty of this model is the flexible and expandable hidden layer for recognition.

***Index Terms*** *– Hidden markov model (HMM), artificial neural network (ANN), Speech processing and recognition,ARM7,MATLAB 2012.*

*****

## I. INTRODUCTION

The diagram of the system is as follows. Under the control of ARM7, LPC2148 processor the appointed speech password said by the being tested person is pre-emphasized, windowing and feature extracted, and then is compared with the training model parameters stored in the system database. The extracted features are then optimized by using HMM. Here the HMM is trained by using the baumwelch algorithm and for the training of HMM viterbi decoding algorithm is used The output of HMM is given to ANN for further classification of the closest user. Here the feed forward ANN is used for classification.

The block diagram of speech recognition access control system using HMM/ANN is shown in figure-1.Here the input signal is given to the system where it is preprocessed by using first order high pass filter. Then as the speech signal is discontinuous, its end point detection is done .It is the product of the average high energy of the vowel and the high zero crossing rate of the consonant n is opened or closed .In preprocessing, first the analog input signal is converted to digital using the sampling technique .Here the sampling is done by using the hamming window. Windowing is done to to select the specific frequency of the speech signal. The feature extraction stage is followed by the preprocessing Here the features of the given input signal is extracted by using MFCC that is mel frequency cepstral coefficient algorithm.

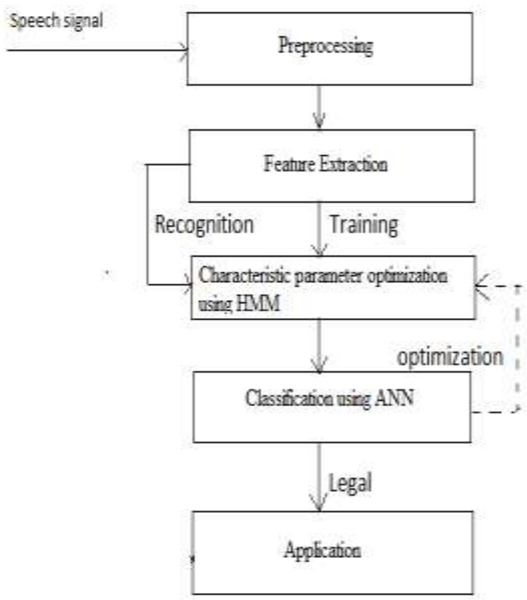


Fig.1.Block Diagram of speech recognition access control system using HMM/ANN.

This is the testing phase. The testing phase is followed by the training phase. After that the comparison is made between the feature extracted in the database and that of the real signal by using HMM/ANN hybrid model .on the basis of the comparison the signal having highest closing rate is find out. The first n characters of the string are compared and the final recognition result is given. On the basis of that the application is opened or closed.

ISSN: 2321-8169
4154 – 4157

## II. BASIC IDEA OF SPEECH RECOGNITION USING HMM RECOGNITION

Hidden markov model is a probability model where the characteristic of the signal are determined by the stotachastic process of observation symbol.

HMM need to solve the three basic problem

1) Using forward backward algorithm calculate the output probability.
2) Using the viterbi decoding algorithm select the best state chain in order to explain the observation symbol sequence.
3) Using Baum Welch algorithm find the transmission matrix and emission matrix.

Stages of HMM recognition are as follows

1) Generating a Test Sequence
2) Estimating the State Sequence
3) Estimating Transition and Emission Matrices
4) Estimating Posterior State Probabilities
5) Changing the Initial State Distribution

Hidden Markov Model is acclimated to admit the accent afterwards accent is accustomed application ANN. HMM is an statistical apparatus for clay abundant sequences to accomplish Observable arrangement if characterized by an basal process.the ascribe to the HMM is the abstracts alone by ANN. The adding of absolute amount is pre-decided & the adding with the ascribe abstracts will be done in the processing phase. HMM assuredly decides the accession & Rejection of new ambit in hidden layers. During processing,the abstracts set will be justified with the absolute value. If not alone again the assay of frequency, time, and amplitude will be performed and abstraction of new appearance will be calculated.

## III. SPEECH RECOGNITION USING ANN

After Normalization,the next important footfall is to admit the accent application Artificial Neural Networks . In this we adduce a Multilayer Mapping Arrangement .The Fig. 2 Shows the Multilayer Pattern Mapping Neural Network. This Multilayer Mapping arrangement works on the assumption of BackPropagation algorithm. The advantage of this archetypal is that is its adaptability & expandability of hidden layers for recognition. Neural networks accept abounding similarities with Markov models. Both are statistical models which are represented as graphs. Where Markov models use probabilities for state transitions, neural networks use affiliation strengths and functions. A key aberration is that neural networks are fundamentally alongside while Markov chains are serial. Frequencies in speech, action in parallel, while affricate alternation and words are about serial. This agency that both techniques are actual able in a altered context. As in the neural network, the claiming is to set the appropriate weights of the connection, the Markov model challenge is award the adapted alteration and observation probabilities. In abounding accent recognition systems, both techniques are implemented calm and work in a accommodating relationship. Neural networks perform very able-bodied at acquirements phoneme anticipation from highly parallel audio input, while Markov models can use the phoneme ascertainment probabilities that neural networks provide to aftermath the likeliest phoneme arrangement or word. This is at the amount of a amalgam access to accustomed language understanding.

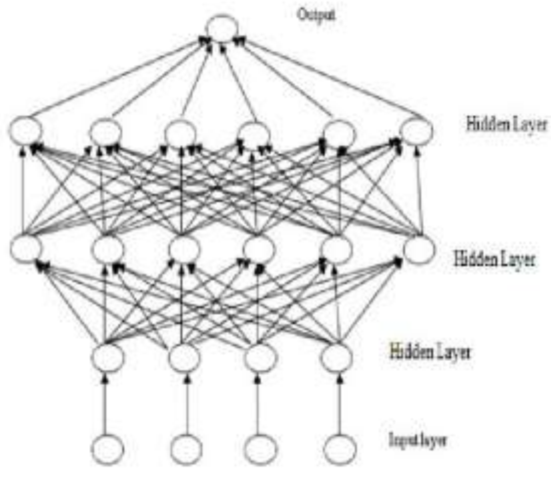


Fig.2. Multilayer pattern mapping network

## IV. MFCC FEATURE EXTRACTION

We apperceive that animal ears, for frequencies lower than 1 kHz, apprehend tones with a beeline calibration instead of logarithmic scale for the frequencies college that 1 kHz. The mel frequency scale is a beeline abundance agreement beneath 1000 Hz and a logarithmic agreement aloft 1000 Hz. The voice signals accept a lot of of their activity in the low frequencies. It is as well absolute accustomed to use a mel-spaced clarify coffer showing the aloft characteristics. The afterward almost blueprint is acclimated to compute the mels for a accustomed abundance in Hz as follows

ISSN: 2321-8169
4154 – 4157

$$m = 2595 \log(1 + f/700) \qquad (1.1)$$

For anniversary accent with an absolute abundance f (in Hz), a subjective angle is abstinent on a calibration alleged the ‘mel’ scale. The angle of a 1 kHz tone, 40 dB aloft the perceptual hearing beginning is authentic as 1000 mels. The cepstrum is the advanced Fourier transform of a spectrum. It is appropriately the spectrum of a spectrum, and has certain backdrop that accomplish it advantageous in abounding types of signal analysis. One of its added able attributes is the actuality that any periodicities, or again patterns, in a spectrum will be sensed as one or two specific apparatus in the cepstrum. If a spectrum contains several sets of sidebands or harmonic series, they can be ambagious because of overlap. But in the cepstrum, they will be afar in a way agnate to the way the spectrum separates repetitive time patterns in the waveform.

## V. RESULTS AND DISCUSSION

In this paper 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 preprocessed & it is optimized by HMM and 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 ,frequency etc. The Fig.3 shows the GUI model for ANN ..once the samples are stored in database ,they are trained and features are extracted. In Recognition Phase the a single speech sample is taken as an input. The Fig. 4 shows the spectrogram obtained for feature extraction The Fig. 5.shows the speech recognition GUI model. The features extracted for stored speech samples are then compared with the features of speech input sample. ANN uses multilayer mapping network which uses Back propagation algorithm for training the samples to find the best match in the comparison. During the comparison if input sample matches with the samples in the database then the system provides access to that particular speaker. The same procedure is repeated for real timespeech processing by recording the samples in real time with the help of microphone. This is how speech is recognized using ANN. The data now rejected by ANN is used as an input to the HMM system.

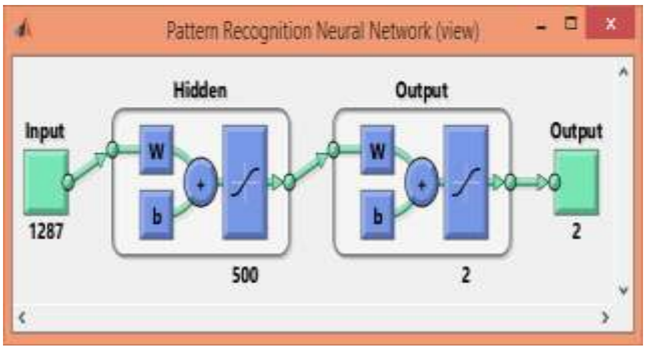


Fig.3 GUI Model For ANN

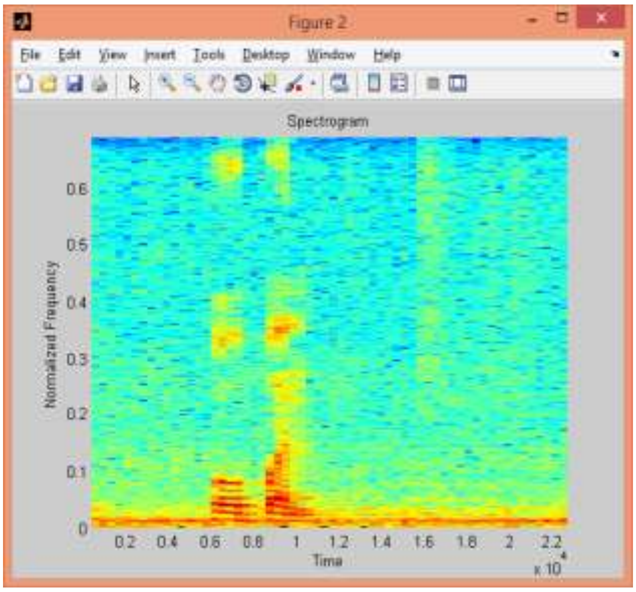


Fig. 4 GUI For Spectrogram

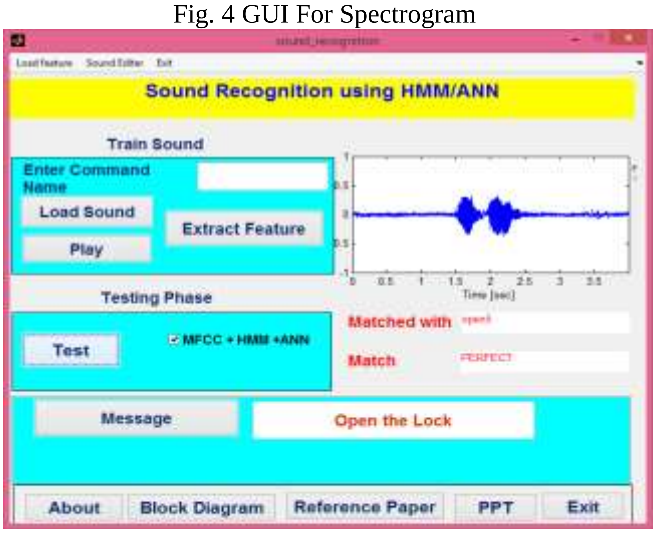


Fig.5 GUI For Speech recognition output

## VI. CONCLUSION

In conclusions, this paper has proposed the use of HMM/ANN hybrid model for speaker dependent isolated word speech recognition access control system. The speech recognition control access system is implemented using MATLAB 2012 software .The work was initially focused on feature extraction ,feature optimization using HMM and classification of extracted feature is done using ANN.The experimental results showed that the recognition rate for the isolated words is improved by using hybrid HMM/ANN technology for recognition. For future work the training data size can be increased and modified HMM/ANN hybrid model can be used to increase the recognition rate. Viterbi decoding algorithm is used to find the best state path.

 ISSN: 2321-8169
4154 – 4157

## VII. ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their useful comments to improve this paper.

## REFERENCES

[1] Li Bo, Wang Dong-xia, Zou De-jun, Hu Tie-sen, , “On Speech Recognition Access Control System Based on HMM/ANN” Proc. IEEE, vol. no. 978, Feb. 2010.

[2] Rabier L.R.,”A Tutorial on Hidden Markov Model And Selected Applications In Speech Recognition”, proc.IEEE,Vol.no.77,1989.

[3] Davis And biddulph and Balashek, “Automatic Recognitn of spoken Digit”,J.Acoust.Soc.Am.24:nov1952.p.637.

[4] Wei HAN,Cheong-Fat CHAN,Chiu-sing CHOY and Kong-Pang PUN, “An Efficient MFCC Extraction Method In Speech Recognition”,IEEE 2006.

[5] Ben J. Shannon,Kuldip K. Paliwal,”A Comparative Study Of Filter Bank spacing for Speech recognition”, Vol No.3,Feb. 2010.

[6] Anjali Bala,”Voice Command Recognition Based on MFCC And DTW”,Vol.no.2(12),p.7335-73422010.

[7] Abdulrahman Alarifi, issa Alkurtass and abdul MalikS., “Arabic text-dependent speaker verification for mobile devices using artificial neural networks”,vol. 7(7),pp.1073-1082,9 feburary2012.

[8] H.F.Ong and A.M.Ahmad,“Malay language speech recognizer with hybrid hidden markov model and artificial neural network”,vol. 1,no.2,pp.114-119,june 2011.

[9] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, “Neural Networks used for Speech Recognition”,vol.20,pp.1-7,2010.

[10] Xinguang li,Minfeng yao and jianeng yang,”speech recognition approach based on speech feature clustering and HMM”,Vol.7,No.9.september 2012.

[11] Fadhil H.T.Al-dulaimy,WANG zuoying“Peripheral nonlinear time spectrum features algorithm for large vocabulary mandarin automatic Speech Recognition”,vol.10,No.2,pp.74-182,april 2005.

[12] LIU Hong,QIAN Yanmin,LIU Jia”English speech recognition system on chip”,Vol.16,No.1,Feburary 2011.

[13] Mark s. Hawley,stuart p. Cunningham,phil D green,pam enderby“A voice input voice output communication aid for people with severe speech impairment”, IEEE vol.21, Janurary 2013.

[14] Kuldip paliwal,Benjamin Shannon,james lyons and kamil wojcicki,”speech signal based frequency wraping”,Vol.16,No.4,april 2009.

[15] Yoshua bengio,renato de mori,Giovanni flammia“Global optimization of aneural network hidden markov model hybrid”,vol.3,No.2,march 1992.

[16] Parameshachari B D,sawan kumar gopy,Gooneshwaree hurry,Tulsirai T. Gopaul”,A study on smart home control system through speech”,Vol.69,No.19,May 2013.