IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 4, Issue 8, August 2016

Study of Hidden Markov Model for Isolated Word Recognition

Vedanti V. Tungikar[1], Jayashree Mokashi[2]

MKSSS's Cummins College of Engineering, Pune[1, 2]

Abstract: Speech processing is a very wide field and Speech recognition covers big part of it. It is a measure of converting an acoustic signal taken from microphone to a set of words. Speech recognition is a very important aspect for voice driven service portals, speech interface in automotive navigation and guidance system or speech driven application, and isolated word recognition is a base of speech recognition. In this project, the isolated word recognition system is designed which is based on Hidden Markov Model. The recognition process is divided into four stages. It starts with Analysis followed by Feature extraction. In feature extraction important features are extracted such as pitch, pitch duration etc. In next stage acoustic model is formed and recognition stages comes followed by acoustic model. Recognition process makes use of HMM with 3 hidden states. HMM used here is context-dependent that is monophone based. In HMM three algorithms are used, forward algorithms for finding out probability distribution, Viterbi algorithm is used to find hidden states and Baum-Welch algorithm is used to fit the model. Word is recognized with higher accuracy at last stage.

Keywords: Speech recognition, Hidden Markov Model, Acoustic model.

I. INTRODUCTION

Speech recognition is a very broad and emerging technology in the field of signal processing. Speech recognition can be useful in applications such as security for criminal identification, automotive navigation and guidance, in voice driven portals. Speech recognition is nothing but converting the input signal into its textual form with the help of some algorithms [1]. The most useful algorithm for automatic speech recognition is a statistical model called as Hidden Markov Model. The speech recognition systems which are HMM based converts input signal into the smallest unit that is phoneme. HMM models are based on Markov chain. The base of the speech recognition is isolated word recognition. Isolated word is confined of single word or an utterance. Recently for isolated word recognition also widely used algorithm is HMM. Before HMM, some algorithms were used which were based on dynamic programming and distortion measures for finding the best match between unknown words.

In 1970, Baker and Bakis were the first to work on HMM based isolated word recognition system, then this research was continued by IBM. After that in 1983, Levinson et al studied the algorithm consideration and applications of HMM. Rabiner et al described their experimental study about HMM in talker-independent recognition for isolated digits [1].

Series of experiments were carried several years back in which HMM were used with multivariate Gaussian output densities in the isolated word recognition problem. Initially input data was served by feature vector (phonemes which are output of feature extraction process.) consisting of log LPC error and eight log area ratio [1] [2]. For recognition process 100 talkers speech were used as database and error for 100 talker was 0.9%. The database from 100 talker was recorded in two phases. The data from first phase was used for training purpose and the data from second phase was used for testing purpose. Some precautions were taken to avoid use of same talker in both the phases.

II. OVERVIEW OF SPEECH RECOGNITION

Speech recognition is a very far-reaching aspect of ongoing research of signal processing. It is very important part of human- computer future relation.

Speech recognition is nothing but recognizing anybody's voice, that's why automatic speech recognition has many applications in voice driven portals and person identification and verification.

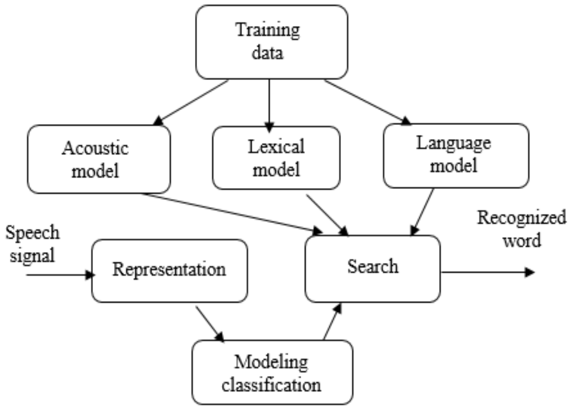

Fig 1. Speech recognition system

DOI 10.17148/IJIREEICE.2016.4821

III. VARIOUS CLASSES OF RECOGNITION SYSTEM

Depending on speakers and type of speech, the speech recognition system broadly classified into two types. Along with change in speaker and type of speech, there are many parameters which affects the system.

A. Based on speech

- Isolated word recognition: Isolated word recognition stands for recognition of a single word. In this type boundary conditions are not important. It uses single utterance at a time.
- Connected word recognition: The working of connected word recognition is same as that of isolated word recognition it just takes several utterances at a time instead of single utterance.
- Continuous speech recognition: In this type, the complete sentences are used instead of single utterance and it is more difficult as in this type a series of words are taken for recognition.

B. Based on speaker

- Speaker dependent: These types of system depends on the speaker's voice. That is these systems demands the speaker must train the software by giving their voice to it. These systems are more accurate as compared to speaker independent systems. This system does not provide flexibility.
- Speaker independent: This type of system works for any type of speaker talking in a particular type of language. These systems are more flexible but less accurate.

IV. DATABASE GENERATION

Database is created of seven fruit names namely Apple, Banana, Kiwi, Pineapple, Orange, Peach, and Lime. For each fruit name 15 utterances are recorded by a single speaker whose age I from 18-22. There are total 105 utterances which are involved in vocabulary. The most important part is to choose no of hidden states for HMM.

V. ARCHITECTURE OF SPEECH RECOGNITION SYSTEM

Basically there are four stages of speech recognition system.

A. Analysis:
The first step is to input the signal into the system. This input speech can a single utterance or a word. It can be any acoustic speech. The single utterance or word will be given to system as an input in case of isolated word recognition. In case of continuous word recognition series of utterances or sentences will be given to system as input.

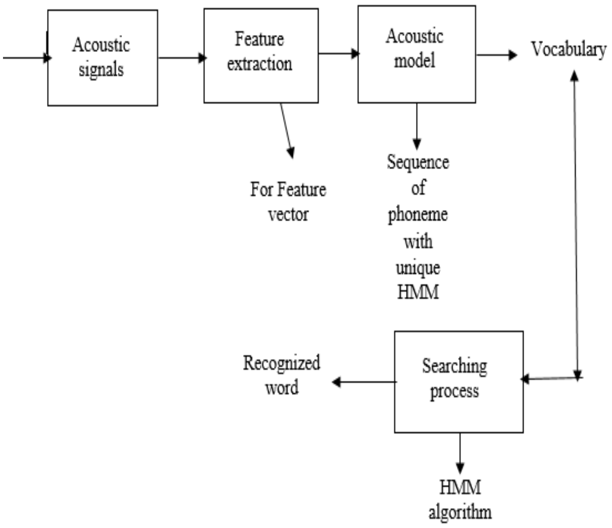

Fig 2. Architecture of speech recognition system

B. Feature extraction:
The main purpose of feature extraction is to convert input speech signal into some representation for further processing. In feature extraction, samples of input signals are taken at every 10-15ms. These input signals are segmented into the lowest form which is phoneme for speech. The output of feature extraction is phonemes which is also known as feature vectors. Generally MFCC and LPC these two methods are used for feature extraction, but in this system fast Fourier transform is used. The source sample signal is sampled at 8000 Hz and quantized with 16 bits. At every 10ms samples are taken. The samples are taken at every 10ms because the signal coming from human throat is stationary during this period. The extracted features are taken from frequency domain but before the feature extraction is carried out by fast Fourier transform and for that speech signal is multiplied with Hamming window.

C. Acoustic modeling
The smallest unit of speech signal is phoneme because it delivers discrimination between meanings of words that why the input speech signal is portioned into phonemes during feature extraction process. In acoustic modelling, a unique Hidden Markov Model is assigned to every phoneme. Every phoneme HMM consists of 3 hidden states, i.e alpha, beta, and gamma.

Acoustic model is nothing but the vocabulary. Each word in this vocabulary has distinct HMM. When unknown word comes, it is scored against all HMM models, and the HMM with maximum score is considered as recognized word. In other words, the HMM model which has maximum probability distribution is considered as recognized word. The output of the acoustic model is sequence of phonemes. In many cases there are many words with same phonemes in that case we need to go for language structure. Key and khakee these words are the example of it.

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified
Vol. 4, Issue 8, August 2016

D. Searching algorithm

1. Hidden Markov Model
The search algorithm used in this isolated word recognition process is Hidden Markov Model. HMM is a statistical model and very rich in mathematical structure. Hidden Markov model is a stochastic finite set of states which in concerned with probability distribution [2] [3].

The reason why HMM can be used as a search algorithm for speech recognition because a speech signal can be considered as a stationary signal when samples are taken over 10ms. And during this period, this speech signal can act as a Markov model. And another reason is that HMM can be trained automatically and very easily. And HMM also provides flexibility. Another advantage of HMM is that it can be used in variety of applications [3].

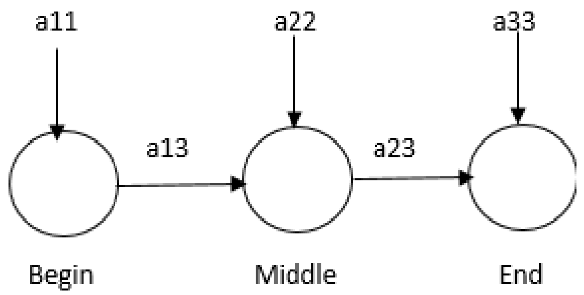

Fig 3: Hidden Markov Model for single phoneme

For Hidden Markov Model there are three concerns. First is evaluation i.e. calculating probability distribution? Another is determining hidden states and last one is learning which is also known as training in this recognition system.

2. Forward algorithm
The probability distribution of HMM is evaluated with the help of forward algorithm. The simple way of calculating probability distribution is to enumerate all the possible states sequence of length T [4].

Let's say q_1 produces o_1 observations and q_2 produces o_2 observations and so on. Then chain rule is applied on those sequences but this algorithm is very inefficient. No of calculations required for chain rule is N^T. That's why forward algorithm is used for evaluation of p (o/λ). It uses partial observations so the no of calculations are less and it is N^2 T.

3. Baum-Welch algorithm
Most difficult part of HMM is to adjust the model parameters so as to get maximum probability distribution. Training of HMM is done by Baum-Welch algorithm, which creates the best model [4]. Need of adjusting the model parameters is to find maximum likelihood between them.

VI. RESULTS

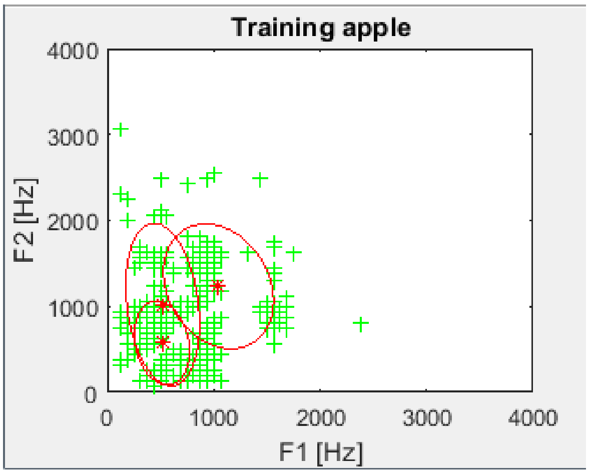

Figure 4(a): Training Apple

Figure 4 (a), (b), (c), (d) shows the sequence of phoneme which comes out after recognition of words. In figure, ellipse shows the 80% confidence region where phoneme of word lies. Green positive signs show the frequencies of the words. And star shows the states i.e. recognized words.

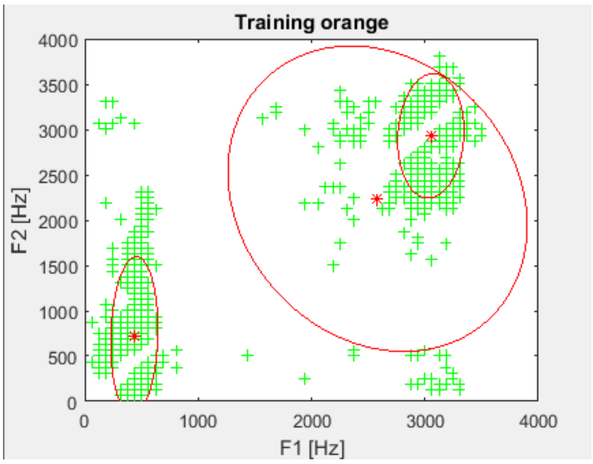

Figure 4(b): Training Orange

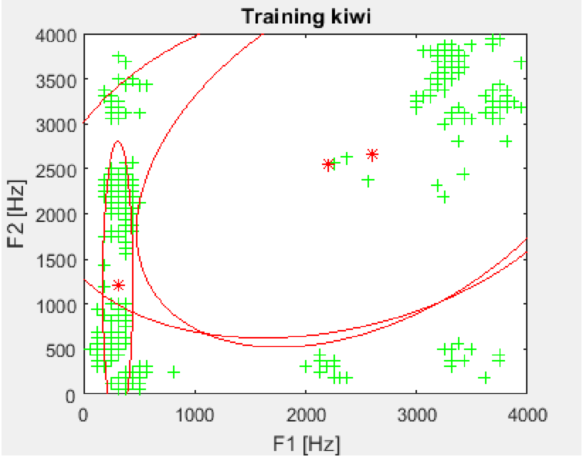

Figure 4(c): Training Kiwi

DOI 10.17148/IJIREEICE.2016.4821

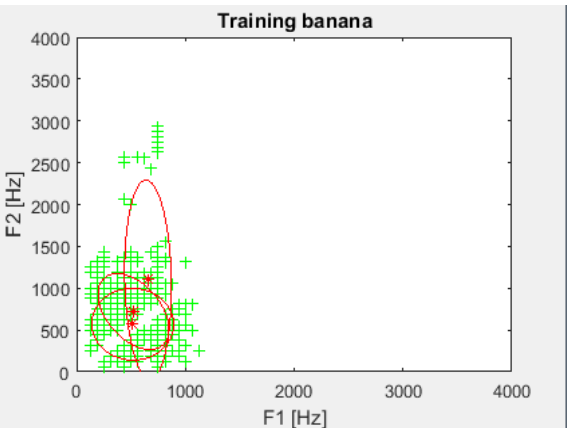

Figure 4(d): Training Banana

Table 1: Recognition parameters

MCR	0.035
NO of Miss	0
Accuracy (%)	99.65%

VII. CONCLUSION

This isolated word recognition system recognizes words with very high accuracy. The accuracy is high because this system is trained with a single speaker’s voice. Is the recognition system is trained with more than one speaker’s voice, it will not give desired results. For future work, triphone based HMM can be used as a searching algorithm which can be helpful for continuous speech recognition system.

REFERENCES

[1] ALAN B. PORITZ and ALAN G. RICHTER, “On Hidden Markov Models in Isolated Word Recognition”, ICASSP 86 IEEE Conference of Acoustic speech and signal processing 1986.

[2] Aymen, Mbarki, Ammari Abdelaziz, Sghaier Halim, and Hasen Maaref. "Hidden Markov Models for automatic speech recognition", International Conference on Communications Computing and Control Applications (CCCA), 2011.

[3] Rabiner, L.R., “A tutorial on Hidden Markov Models and selected applications in speech recognition,” Proceedings of the IEEE, 77, No.2, 1989, pp.257-286.

[4] Saptarshi Boruah, Subhash Basishtha, “A Study on HMM based Speech Recognition System”, Department of Information Technology, 978-1-4799-1597-2/13/$31.00 ©2013 IEEE.

[5] Rabiner L, Juang B. An introduction to hidden Markov Models. IEEE ASSP Magazine, Jan., 1986, pp.4-16.

[6] Preeti Saini, Parneet Kaur,” Automatic Speech Recognition: A Review”, International Journal of Engineering Trends and Technology- Volume4 Issue2- 2013.

DOI 10.17148/IJIREEICE.2016.4821