International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 04, Volume 6 (April 2019) www.ijirae.com

HMM APPLICATION IN ISOLATED WORD SPEECH RECOGNITION

Sonali Rawat*
Department of Computer Science and Engineering
IMS Engineering College, Ghaziabad, Uttar Pradesh, India
rawatsonali3003@gmail.com

Shalvika Shrotriya
Department of Computer Science and Engineering
IMS Engineering College, Ghaziabad, Uttar Pradesh, India
shalvikashrotriya@gmail.com

Juhi Chaudhary
Department of Computer Science and Engineering
IMS Engineering College, Ghaziabad, Uttar Pradesh, India
juhi.chaudhary@imsec.ac.in

Manuscript History
Number: **IJIRAE/RS/Vol.06/Issue04/APAE10081**
Received: 02, April 2019
Final Correction: 05, April 2019
Final Accepted: 08, April 2019
Published: **April 2019**
Citation: Rawat, S., Shrotriya, S. & Chaudhary, J. (2019). HMM Application in Isolated Eord Speech Recognition. IJIRAE::International Journal of Innovative Research in Advanced Engineering, Volume VI, 273-277.
doi://10.26562/IJIRAE.2019.APAE10081
Editor: Dr.A.Arul L.S, Chief Editor, IJIRAE, AM Publications, India

Abstract —Speech recognition is always being an all-time trendy topic for discussion and also for researches and we see a major application in our life. This paper provides the work done on the application of Hidden Markov model to implement isolated word speech recognition on MATLAB and to develop and train the system for set of self-selective words for specific user (user dependent) to get maximum efficiency in word recognition system. Which uses the forward and Baum-welch algorithm and fitting Gaussian of the Baum-welch algorithm for all the iteration perform. We use a sample of 7 alphabets which are recorded in 15 different ways giving total of 105 word to use for training with each word with 15 variations. This system can be used in real world in system security using voice security system and mainly for children and impaired people.

Keyword: Hidden Markov Model; Isolated word recognition; Baum-welch; Gaussian Fitting;

I. INTRODUCTION

Sound or speech is always the part of every human being and is associated till life. Speech generation can be natural or by computer called speech synthesis. Vocalization causing speech and is important form of human conversation and plays an important role in human life, while speech recognition in making computer interactions easier. Speech recognition, made major steps in the past era, and it has various advance application towards several commercial systems and is embedded in customer call centres, google assistant, and many other voice-activated routing systems which are currently available and this isolated word speech training is found most efficient in term of dealing with noisy data as most of the systems lacs the robustness[1] and it can filter noise and can opt for multiple acoustic conditions including person's speech rate and frequency.

Speech recognition is a technology procedure to extract the feature of speech and then process it which allows computer to understand isolated word by humans inputted by hardware or in audio format then the system is trained for a single user results to train in a specific voice model by strictly using a single stochastic model Hidden Markov Model (HMM), using forward and backward algorithm along with Baum-Welch algorithm, Which resulting in creating trainer dependent or independent system using supervised along with label training in which training one model for individual word then further using this in various applications, to convert the voice for text or understanding commands, data entry and most importantly for impaired people, real time vehicle control[2] home security system etc.

II. ANALYSIS OF TYPES OF SPEECH RECOGNITION SYSTEM

Speech recognition system either work on continuous speech in which there is no or minimum pause in a fused manner and isolated speech recognition system which include a maximum pause in-between each word or spoken separately. System can be further divided on the basis of speaker, dependent or independent. In speaker independent the system is trained in a way to recognize any speech irrespective of the speaker which also make it very adaptive and robust[3]. Which have major application in voice to text for any valid speaker. In dependent system is restricted to recognize speech only from specific selected speaker as it needs the recoding of a speech and the system is trained under the selective speaker. Its major application is in command based security system or voice based security in real word.

III. SYSTEM WORK FLOW DEVELOPMENT

In our system we are using a pre-recorded voice of 7 fruits names in which each fruit is having 15 variations in speech by a single speaker in all the clips. The system is trained by using the total of 105 words which can be increases according to the need. The workflow is divided into 3 major phases, plotting of frequency time graph for each of the voice clip followed by loading into the HMM model and final is Gaussian fitting this is done on a self-testing basis in which the system recognizes and compare the generated word from the phenomes with the original word.

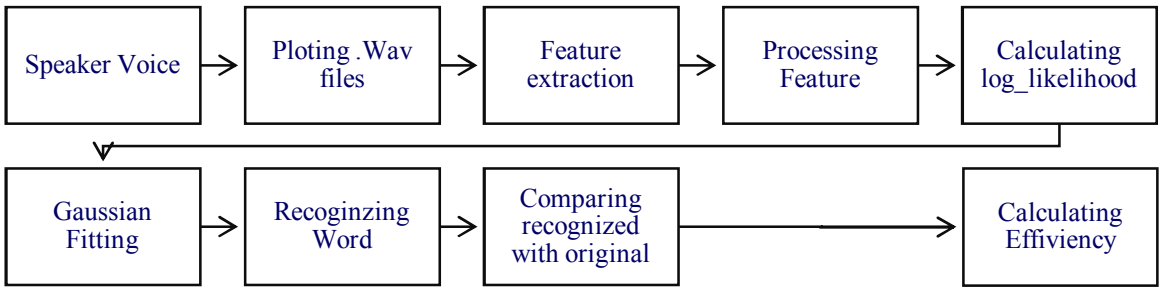

Fig.1 The workflow diagram of system implemented on MATLAB

IV. HIDDEN MARKOV MODEL

We are assuming to have a basic knowledge of Hidden Markov model and we focus to describe the two important algorithms that we used in this project. Multivariate Gaussian distributed is used for the generation of hidden states which create a mean vector and second a covariance vector. A homogenous HMM is used to which makes state transitional probabilities independent of time[4]. Total number of N states are used. An element $a_{ss}$is used in the transition probability matrix denoted by **A** which further denotes the transition probability from state *s* to state *s'*, and $\boldsymbol{\pi_s}$ is the probability for the chain to start in state *s*. The mean vector is $\boldsymbol{\mu_s}$ and covariance matrix is $\boldsymbol{\Sigma_s}$ for the multivariate Gaussian distribution modelling the observable output from state *s*. Here we are using collection of parameters to define our HMM model as $\lambda = \{\mathbf{A},\pi,\mu,\mathbf{\Sigma}\}$.

A. The Forward Algorithm

This is used to select the voice which is most likely to happen or the log likelihood and the probability density from observation o_1

$$f(o_1,\dots\dots,o_T;\lambda) = \sum_{sT} f(o_1,\dots\dots,o_T,s_T;\lambda) \quad (1)$$
$$= \sum_{sT} f(o_T|o_1,\dots\dots,o_{T-1},s_T;\lambda) f(o_1,\dots\dots,o_T,s_T;\lambda) \quad (2)$$
$$= \sum_{sT} b_{sT}(o_T) \sum_{sT-1} f(o_T|o_1,\dots\dots,o_{T-1},s_{T-1};\lambda) \quad (3)$$
$$= \sum_{sT} b_{sT}(o_T) \sum_{sT-1} f(s_T|o_1,\dots\dots,o_{T-1},s_{T-1};\lambda)\ f(o_T|o_1,\dots\dots,o_{T-1},s_{T-1};\lambda) \quad (4)$$
$$= \sum_{sT} b_{sT}(o_T) \sum_{sT-1} f(s_T|o_1,\dots\dots,o_{T-1},s_{T-1};\lambda) \quad (5)$$

The recursive structure is revealed as we reduced the problem from needing $f(o_1,\dots,o_T,s_{T;\lambda})$ for all s_T to needing $f(o_1,\dots,o_{T-1},s_{T-1};\lambda)$ for all s_{T-1}. Now using forward variable.

$$\propto_1 (s) \equiv f(o_1, S_1 = s; \lambda) \quad (6)$$
$$= b_s(o_1)\pi_s \quad (7)$$
$$\propto_t (s) \equiv f(o_1, S_t = s; \lambda \quad (8)$$
$$= b_s(o_t) \sum_{s'} a_{s's} \propto_{t-1} (S') \quad (9)$$

B. The Baum-Welch Algorithm

Maximization of the log likelihood of observation is done using this algorithm along with the training of the samples of hidden Markov model. The Baum-Welch algorithm is an iterative expectation-maximization (EM) algorithm that converges to a locally optimal solution from the initialization values.

$$\pi_s = \frac{expected\ number\ of\ times\ in\ state\ s\ at\ t = 1}{expected\ number\ of\ times\ at\ t = 1} \quad (11)$$
$$a_{ss'} = \frac{expected\ number\ of\ transitions\ from\ s\ to\ s'}{expected\ number\ of\ transitions\ from\ s} \quad (12)$$
$$\mu_S = expected\ observation\ when\ in\ state\ s \quad (13)$$
$$\Sigma s = observation\ covariance\ when\ in\ state\ s \quad (14)$$

Indicator functions and linearity of expectation are used for calculating the expected values. To calculate the probabilities, we use the backward variable similar to the forward variable. Works in same way but just in opposite manner. The iteration of algorithm done until the results are satisfactory.

V. DESIGNING THE SYSTEM

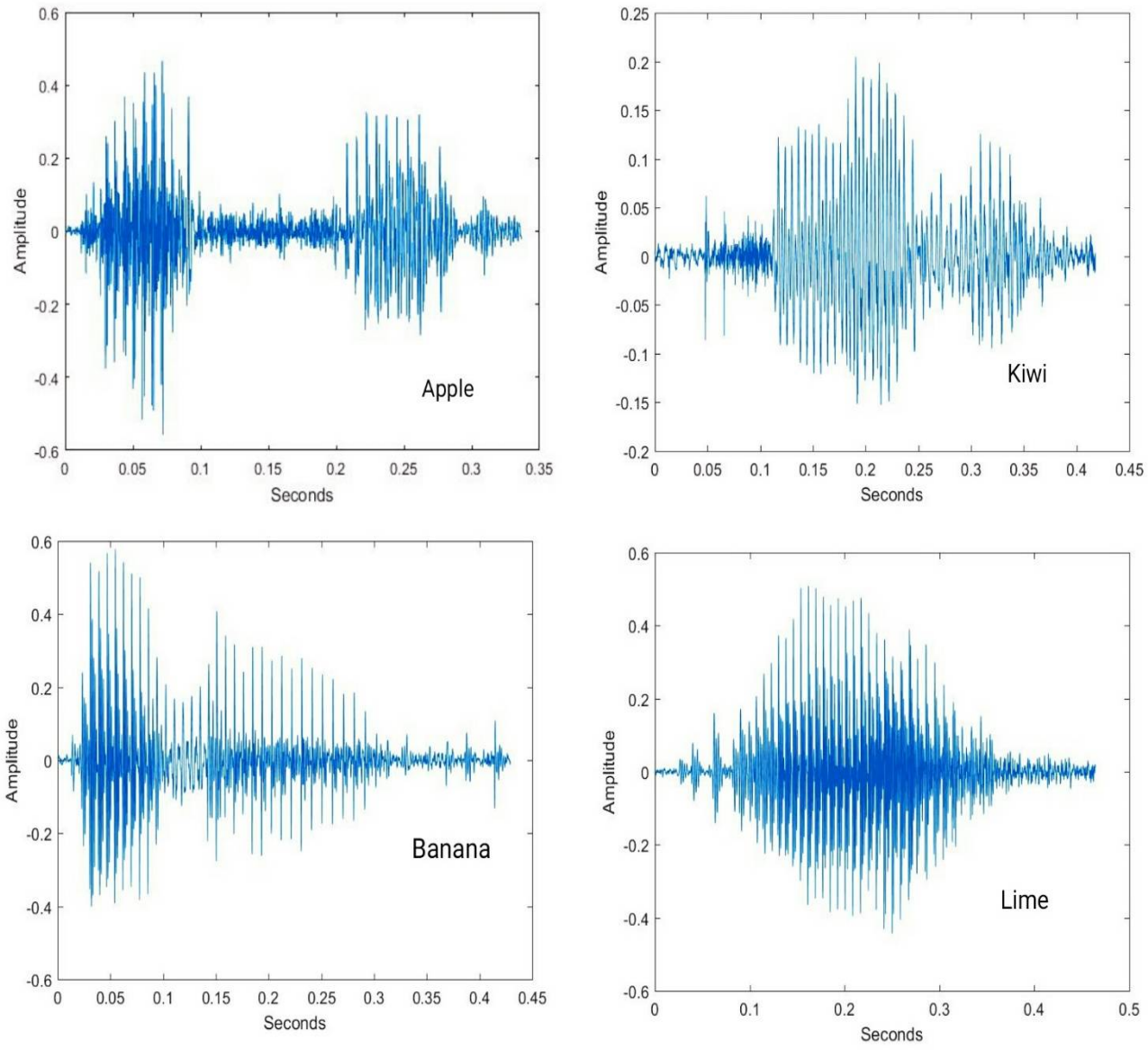

Fig. 2 Showing Amplitude Vs Seconds plot of few sample data.

A. Acquiring Speech and feature extraction

The sound is recoded using the microphone us then converted into a frequency of 8000 Hz by using the MATLAB function

Fs = 8000; framesize = 80; overlap = 20; D = 6; y = wavrecord(Duration*Fs,Fs);

Recording and saving each of the 7 fruits sounds with each fruit 15 times. Then plotting the frequency time for each of the sound for feature extraction.

B. Training

As we are using label based learning which implies that both supervised as well as unsupervised learning will be there for the given dataset. We train one state for each speech signals and each state is nothing but the phenome of the sound. Gaussian clustering is unsupervised and is based on the initial values of the Baum-Welch algorithm. We randomly generate the initial values of matrix **A** and π which strictly follows the statistical properties. Σ_s, which is the diagonal covariance matrix as defined in all the iterations performed. For initial we found 15 iterations are effective for Baum-Welch algorithm

VI. SETUP AND RESULT

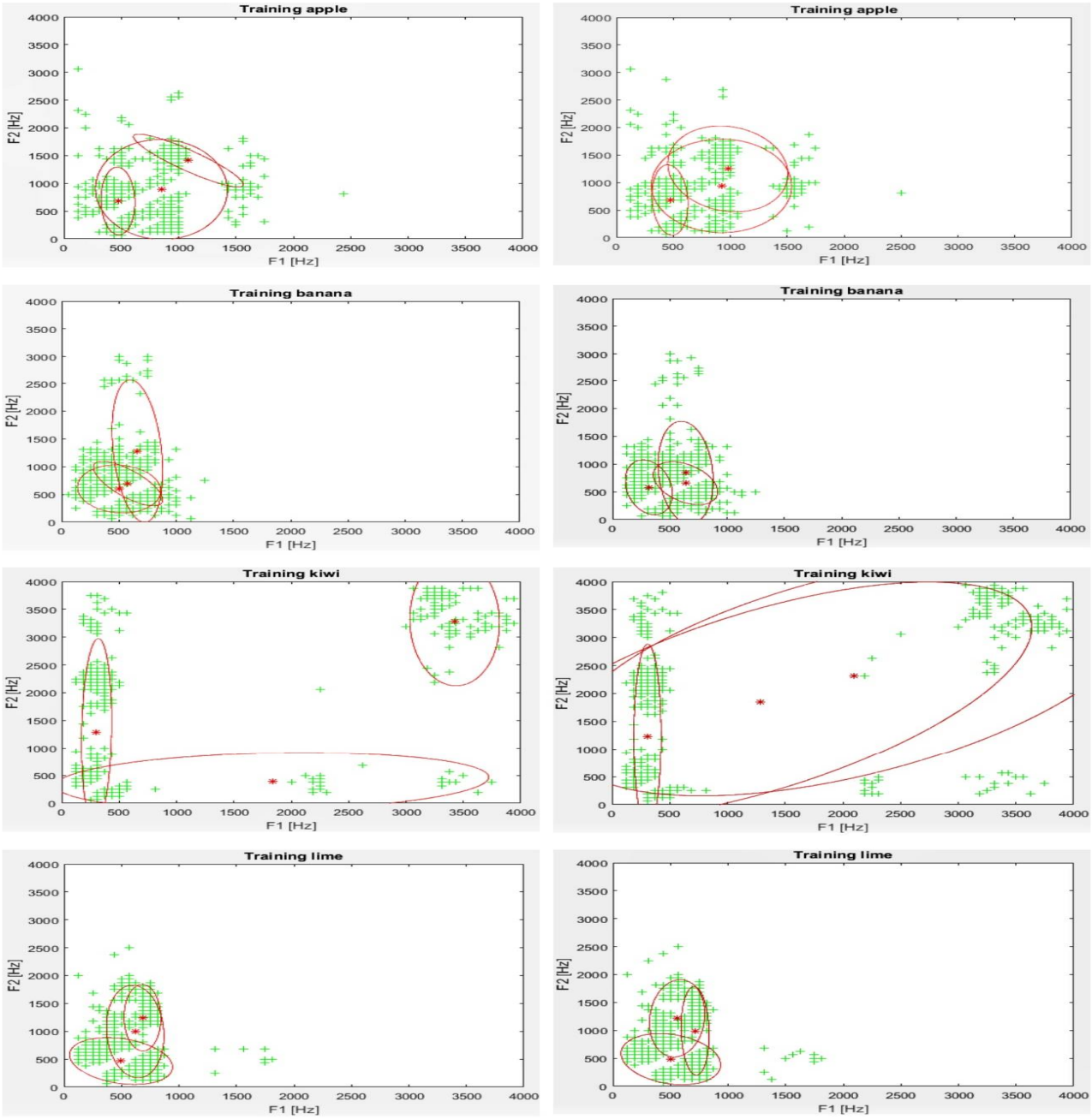

Fig.3 Comparison between the training Apple

For set of each word *i* we denote parameter as λ_i having observation state from o_1,.....o_T, from which the section of word is done using arg max $f(o_1,.....o_T; \lambda_i)$ which is given by the forward algorithm. Fitted Gaussian after iteration from the first fitted plot vs the final iteration plot can been seen in the figure 3and the variation is very observable along with at each iteration the efficiency increases from 15% to 95% efficiency is calculated by using the mcr (misses / length (word labels)) *100 which we see vary from 92% to 94%. For the complete set used.

VII. CONCLUSIONS

We finally able to create a system which is able to train the isolated speech word using HMM and supervised and unsupervised learning. The results generated are valid for only single speakers. The system could make robust by using multiple users and using continues speech instead of individual word. This results for limited word set of 105 samples and efficiency can vary for different amount of dataset.

REFERENCES

1. S.A.R. Al-Haddad, S.A. Samad, A. Hussain, K.A. Ishak and A.O.A. Noor, Robust Speech Recognition Using Fusion Techniques and Adaptive Filtering American Journal of Applied Sciences 6 (2): 290-295, 2009.
2. Shi-Huang Chen, YuRu Wei, A Study on Speech-Controlled Real-TimeRemote Vehicle On-Board Diagnostic SystemProceeding of the International multiconference on Engineers and Computer Scientists 2010, IMECS 2010, March, 7-19,2010,vol I.
3. Fadhilah Rosdi, Raja N. Ainon Isolated Malay Speech Recognition Using Hidden Markov Models, Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, May, 13-15,2008.
4. L. R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognitionProc. IEEE, Feb. 1989, vol. 77, no. 2, pp. 257–286.

