Speech Recognition System

Om Thakkar (201501109)
Shivam Raval (201501088)

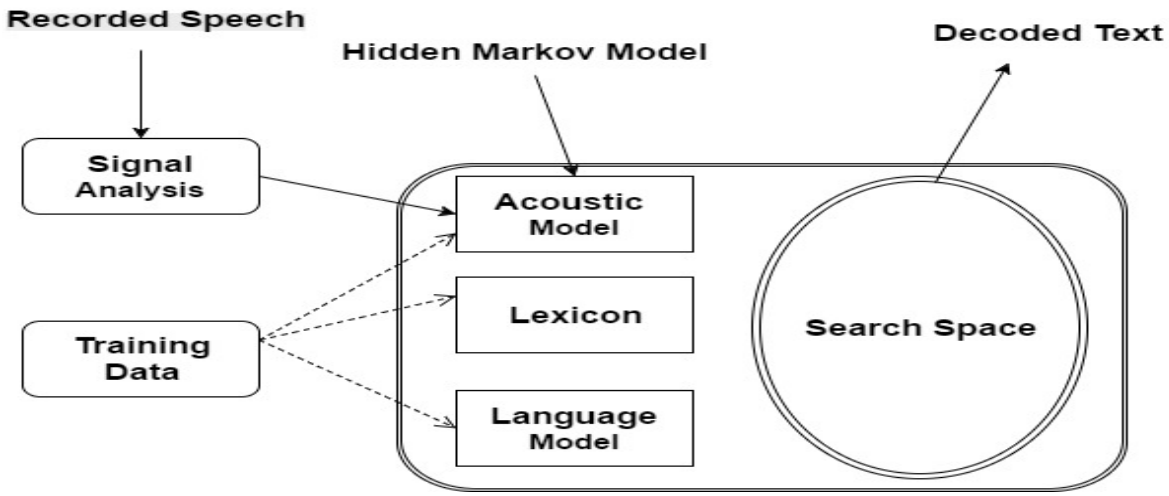
School of Engineering and Applied Sciences Ahmedabad University
MA202-Probability and Random Processes
Guided by : Dr. Dhaval Patel

May 9, 2017

Background and Motivation

- Speech Recognition System (SRS) - A software technology that identifies the words spoken by the user.
- Early SRS applied grammatical rules to identify speech.
- **Limitations** : flaws due to different densities of voice, varied accents.
- Today's SRS uses Hidden Markov Model and Neural Networks to determine likelihood.
- **Motivation for developing SRS** : its high use in software technologies and for people with disabilities.

Block Diagram



(a) Block Diagram of Speech Recognition System

Probabilistic Model Used : Hidden Markov Model

- **Assumption** : System being modeled is a Markov process with unobserved (hidden) states.
- A finite set of states, each of which is associated with a (generally multidimensional) probability distribution.
- Transitions among the states are governed by a set of probabilities called transition probabilities.
- Observable output from a hidden state : Multivariate Gaussian Distribution.
- Parameter set, $\lambda = \{A, \pi, \mu, \sum\}$.

PsuedoCode/Algorithm

- Calculate forward probabilities with forward algorithm (calculates probability of a state at a certain time 't', given history of observations from 1 to 't').
- Calculate backward probabilities with backward algorithm (calculates probability of observed states starting from last step and then moving backwards in time.).
- Calculate contributions of the current sequence to the transition and emission probabilities of the model.
- Calculate new model parameters (start, transition, emission probabilities).

PsuedoCode/Algorithm (Continued)

- Calculate new log likelihood of the model.
- Stop when a maximum number of iterations is achieved.

- **Classification** : The selection of a word is done based on the results of the following equation :

$$\text{predictedword} = \arg\max_i f(o_1, o_2, ..., o_T; \lambda_i)$$

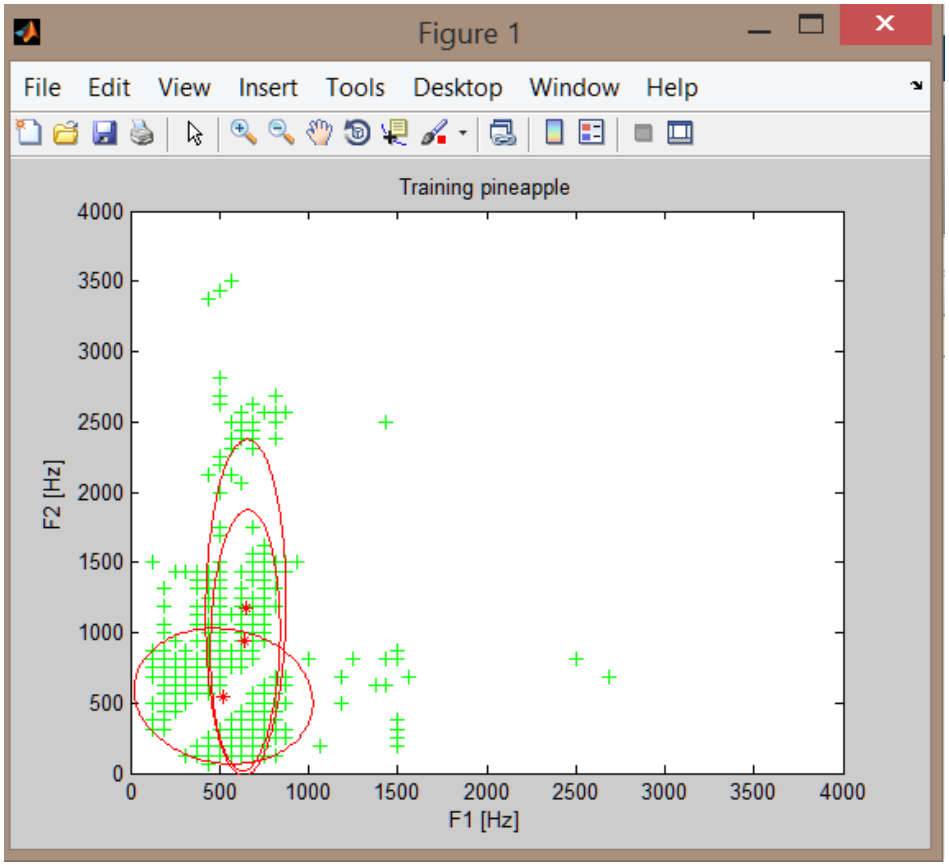
where, λ_i denotes parameter set for word i and o_i denotes observation.here, λ_i denotes parameter set for word i and o_i denotes observation.

Results / Inferences

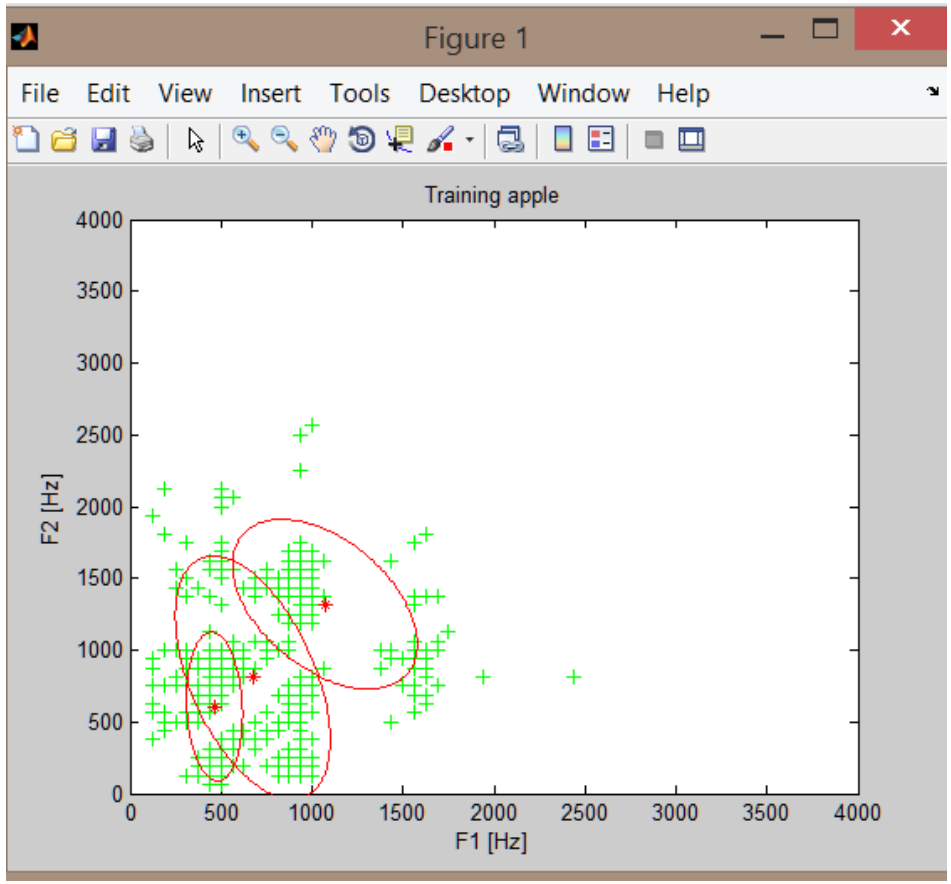
- Number of words : 7
- Number of utterances per word : 15
- Total samples : 105
- System does not perform well for different speakers.
- Increasing number of iterations for Baum Welch Algorithm did not show significant improvements because most time is spent during feature extraction, not in training.
- Result obtained for Number of states, N = 3:

$$\boxed{\text{Misclassification rate} = 5.7\%}$$

Experimental Results



(b) Pineapple



(c) Apple

Future Scope

- SRS for isolated-word speech recognition was implemented and tested.
- Three possible extensions :
  - System support for several speakers
    - use more robust features
  - Support for continuous speech
    - detect word boundaries and then proceed with isolated - word recognizer.
  - Support for recognizing Homophones.

Thank You!