Speech Recognition using Hidden Markov Model Algorithm
Charmi N Solanki¹ Asst. Prof. Mohammed G Vayada²
¹M.E. Student ²Assistant Professor
1,2Department of Electronics & Communication Engineering
Silver Oak College of Engineering
and Technology, Ahmedabad, Gujarat

Abstract—Speech recognition applications are becoming more useful nowadays. With growth in the needs for embedded computing and the demand for emerging embedded platforms, it is required that speech recognition systems are available but speech recognition software being closed source cannot be used easily for implementation of speech recognition based devices. Aim To implement English words speech recognition system using Matlab (GUI). This work is based on Hidden Markov Model, which provides a highly reliable way for recognizing speech. Training data such as words like go up, go right, open, close etc. records in audacity open source; the system will test it with data record and display it in edit text box.

Key words: Hidden Markov Model, Mel Frequency Coefficient Cepstrum

I. INTRODUCTION
Speech recognition basically means talking to a computer, having it recognize what we are saying, and lastly, doing this in real time. This process fundamentally functions as a pipeline that converts PCM (Pulse Code Modulation) digital audio from a sound card into recognized speech.

Fig. 1: Block diagram of Speech Analysis

A. Transform the PCM digital audio into a better acoustic representation:
The input to speech recognizer is in the form of a stream of amplitudes, sampled at about 16,000 times per second. But audio in this form is not useful for the recognizer. Hence, Fast-Fourier transformations are used to produce graphs of frequency components describing the sound heard for 1/100th of a second. Any sound is then identified by matching it to its closest entry in the database of such graphs, producing a number, called the “feature number” that describes the sound.

B. Unit Matching System
Provides likelihoods of a match of all sequences of speech recognition units to the input speech.

C. Lexical Decoding:
constraints the unit matching system to follow only those search paths sequences whose speech units are present in a word dictionary.

D. Apply a "Grammar"
so the speech recognizer knows what phonemes to expect. This further places constraints on the search sequence of unit matching system.

II. BASICS HIDDEN MARKOV MODEL
A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be presented as the simplest dynamic Bayesian network. A Hidden Markov Model is a collection of states connected by transitions, as illustrated in Figure 3. It begins in a designated initial state. In each discrete time step, a transition is taken into a new state, and then one output symbol is generated in that state. The choice of transition and output symbol are both random, governed by probability distributions. The HMM can be thought of as a black box, where the sequence of output symbols generated over time is observable, but the sequence of states visited over time is hidden from view. This is why it’s called a Hidden Markov Model.

HMMs have a variety of applications. When an HMM is applied to speech recognition, the states are interpreted as acoustic models, indicating what sounds are likely to be heard during their corresponding segments of speech; while the transitions provide temporal constraints, indicating how the states may follow each other in sequence. Because speech always goes forward in time, transitions in a speech application always go forward So Hidden Markov Model is reliable algorithm.

A. Algorithms of HMM
There are three basic algorithms associated with Hidden Markov Models:
- Forward algorithm, useful for isolated word recognition;
- Viterbi algorithm, useful for continuous speech recognition; and
- Forward-backward algorithm, useful for training an HMM.

B. Limitations of HMM
- Constant observation of frames
- The Markov assumption
- Lack of formal methods for choosing a model topology
- Large amounts of training data required

III. PREVAILING FEATURE EXTRACTION METHOD
Features extraction in ASR is the computation of a sequence of feature vectors which provides a compact representation of the given speech signal. It is usually performed in three main stages. The first stage is called the speech analysis or
the acoustic front-end, which performs spectra-temporal analysis of the speech signal and generates raw features describing the envelope of the power spectrum of short speech intervals. The second stage compiles an extended feature vector composed of static and dynamic features.

A. Mel Frequency Cepstrum Coefficients (MFCC)

The most prevalent and dominant method used to extract spectral features is calculating Mel-Frequency Cepstral Coefficients (MFCC). MFCCs are one of the most popular feature extraction techniques used in speech recognition based on frequency domain using the Mel scale which is based on the human ear scale. MFCCs being considered as frequency domain features are much more accurate than time domain features Mel-Frequency Cepstral Coefficients (MFCC) is a representation of the real cepstral of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal.

Mel-Frequency Cepstrum Coefficients (MFCC) is a representation of the real cepstral of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal.

![Fig. 2: MFCC Derivation](image)

MFCC is an audio feature extraction technique which extracts parameters from the speech similar to ones that are used by humans for hearing speech. Each time frame is then windowed with Hamming window to eliminate discontinuities at the edges. The filter coefficients w(n) of a Hamming window of length n are computed according to the formula:

\[ W(n) = 0.54 - 0.46 \cos(2\pi n/N) \]

0 ≤ n ≤ N - 1

\[ w = 0, \text{otherwise} \]

Where N is total number of sample and n is current sample be established as: Frequency (Mel Scaled) = [2595log (1+f (Hz)/700)] MFCCs use Mel-scale filter bank where the higher frequency filters have greater bandwidth than the lower frequency filters, but their temporal resolutions are the same.

IV. SIMULATION ON MATLAB AND ANALYSIS OF RESULT

Matlab supports NeXT/SUN SPARC station sound files (suffix is .au), Microsoft Wave sound files (suffixes’ .wav), windows-compatible sound equipment’s, sound recording and broadcasting, as well as the audio signal of linear law and the audio signal of mu law. MATLAB can read, write, get sound information and etc. At the same time as a high-level Language integrated development environment, Matlab can create GUI (graphical user interface) programs. GUI programs are components realizing the communication between interface display and user, the user can click interactive components of mouse and keyboard to achieve a particular function, and then the output of MATLAB will display in the corresponding results area. Fig. 3 shows the GUI for English speech recognition used in experimental process. It is very convenient for speech recognition.

![Fig. 3: Final Result on Matlab](image)

In this work recognize the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficient technique. Then design and implement English command speech recognition system using Matlab (GUI), This work is based on Hidden Markov Model, which provides a highly reliable way for recognizing speech. Graphical user interface to the system which would utilize the built in MATLAB. Generate training data such as words like go up, go right, open, close etc. records in audacity/open source; the system will test and match it with data record and display it in edit text box on Graphical User Interface.

V. CONCLUSION AND FUTURE WORK

The approach is to implement for speech recognition system for English language. The MFCC is used as speech feature extractor. The algorithms are followed by VQ method for testing, helps to conclude that MFCC is more accurate feature extractor for verity speech signals. The present work was limited to phonemes of English only. The further study can be done for continuous speech recognition using MFCC Features extraction algorithm and Hidden Markov Model (HMM) for testing and modeling purpose. The very large vocabulary speech recognition (VLSR) using MFCC with PLP Features extraction algorithm and HMM combined with Artificial Neural network (ANN) for better classification.

ACKNOWLEDGMENT

This paper can be published smoothly with a lot of people especially my guide who are inseparable for their selfless help, they encouraged me and helped me to analyze my problems.

REFERENCES

Speech Recognition using Hidden Markov Model Algorithm
(IJSRD/Vol. 3/Issue 08/2015/202)

Matlab” 2014 International Conference on Reliability, Optimization and Information Technology - ICROIT 2014, India, Feb 6-8 2014


[3] XIA Lisha1, FANG Huajing1 , LI Zefang1 “HMM Parameters Estimation with Inequality Constraints” july 26-28, 2013 China


