HMM based Automatic Speech Recognition Analysis

Anita1, Preeti Saini2
1Student, Electronics & Communication Engineering, Institute of Technology and Science, Bhiwani Haryana INDIA
2Faculty, Electronics & Communication Engineering, Institute of Technology and Science, Bhiwani Haryana, INDIA

ABSTRACT
This project’s ‘HMM Based Automatic Speech Recognition Analysis main motive is just to generate an Automatic speech recognition which is clear an accurate using Hidden Markov Model (HMM) to get accurate results at number of frequency ranges related to human voice. Here is a record of 12 different words which is recorded by using a number of different speakers that includes male and female both (especially female). Thereafter, the speech recognitions result reports are compared with different feature extraction methods in this project instead of one method. Because, an Earlier research work on this project thesis only one feature extraction method has been used and also using a recognition of seven small vocal sounds using HMM (Hidden Markov Model). This speech recognition system mainly divided into two major blocks here in this project. First Block includes the recording data base and feature extraction of all recorded signals. Here we use Mel frequency cepstral coefficients (MFCC), linear cepstral coefficients and fundamental frequency as feature extraction methods instead of one extraction method which were used earlier. For obtaining a Mel frequency cepstral coefficients (MFCC), a signal is passing through following parameters named as pre emphasis, framing, applying window function, Fast Fourier transform, filter bank and then discrete cosine transform, where as a linear frequency cepstral coefficients does not use Mel frequency. Now the Second part includes the description of HMM used for modeling and recognizing the spoken words. All the raining samples are clustered using K-means algorithm. Gaussian mixture containing mean, variance and weight are modeling parameters. Here is also a role of Baum Welch algorithm. it is used for training the samples and re-estimate the parameters. Finally in the thesis, the Viterbi algorithm recognizes best sequence that exactly matches for given sequence there is given during speech vocal sounds which has to be recognized. Here all the simulations are done by using the MATLAB tool.

Keywords---- MATLAB, Rule Viewer, Operating System window 7, HMM, MFCC, Window Techniques, Feature Extraction Methods.

I. INTRODUCTION

The Speech Recognition is one of the main factor through which we take all reports here further. Here are five divisions of Speech Recognition named as isolated spoken word recognition, continuous speech recognition, text dependent spoken word recognition, speaker independent and speaker dependent recognition. In the isolated spoken word recognition, there are different different speakers who speak different different words during training and anytesting signal (in other case, the same word spoken by any other or same person) and then it should be recognised. The second one, Continuous Speech Recognition can be further divided into connected speech word recognition and conversational speech recognition. Here is a little difference in between the Isolated Spoken and Continuous Speech Recognition i.e word recognises is limited in number of words (vocabulary) in Isolated System but in the case of continuous speech recognition, no limit on words and it may focuses on understanding the sentences. Now its about Spoken word recognition, it can also be speaker-dependant (in which case the acoustic features have to be varied for every time speaker changes). There are large numner of application for speaker independent recognition. When we talk about the commercial applications then it is better than others but have some complexity during process. Therefore, above mentioned complexity of Speaker Independent Recognition used during an unique type of acoustic features for each persons. Therfore, as per reports, only Isolated Spoken Word Recognition is generally implemented in speaker independent systems. Here we have recorded different words which are single and complete both from different speakers like each person records in their own Vocal Sound like digits “One, Two, Three……Nine, Ten”. These are completely trained and all over under testing any one among above words. These words have been spoken by any one of the above speakers
and then recognition/result is obtained. Along with the above mentioned data the second data base used were vocal sound like “banana, pencil, hello………movement etc”. Here is a recognition of single words and a complete word respectively. As we mentioned here that these words has been tested by saying after the different different speakers. Let have a result further.

II. LITERATURE REVIEW

When we talk about the Literature Review of this Topic then the Speech word recognition area has a wide research area in which there is not limit to work and researches. There were so many Researchers who have worked Over Speech Recognition Many of the researchers by using different different Methodologies to solve speech recognitions problems by mostly using these Methods named as Hidden Markov Models (HMM), Dynamic Time warping (DTW) and many others. In this Literature review, we finally concludes that among all the methodologies which we mentioned here which have been used for speech recognition problem. The Hidden Markov models are one the best statistical models which is more accurate and successful in modelling of speech parameters. Even in this Hidden Markov Model, number of Researchers have been make a number of changes by using number of methodologies to get best Speech Recognition. So, Most of them were used Extraction Methods to get some features of speech like fundamental frequency, energy coefficients and many others. When we talk about the Review on other feature extraction method like linear prediction coding then these extracted features method are good for speech processing only (speech coding) but not speech recognition.

So, the Complete study of a Review on Hidden Markov Model and it application in Speech recognition (especially on isolated spoken word recognition). Here are two types of Hidden Markov Model like discrete and continuous Hidden Marov Model which were studied. I have also referred here a tool kit of Hidden Markov Method which have been by using a MATLAB programming language.

III. FEATURE EXTRACTION METHOD

INTRODUCTION

This Feature Extraction chapter will be going to introduce the significance, meaning and methods of Feature Extraction. There are two important speech parameters named as pitch frequency and cepstral coefficient. For the purpose of speech recognition, speaker verification and speech synthesis, etc. One of the Researcher must extract the features of the speech segment such as fundamental frequency forms, linear predictive coefficient (LPC), Mel frequency cepstral coefficient (MFCC), cepstral coefficients, line spectral pairs, 2-D and 3-D spectrogram, etc. There are also some time domain features and frequency domain features named as frequency domain, cepstral domain, wavelet domain, discrete cosine transform (DCT) domain and many more. Here, We concentrate discuss fundamental frequency and cepstral coefficients for speech recognition purpose.

FUNDAMENTAL FREQUENCY

Always a speech signal consists of different frequencies which are harmonically in nature and they are related to each other in the form a series. The lowest frequency of this harmonic series is called as the fundamental frequency or pitch frequency. But the Pitch frequency is the fundamental frequency of the vocal cords. There are so many different techniques which are available in market to get the fundamental frequency (f0) like auto correlation method and FFT based extraction. There are some steps which are followed during pitch extraction are given below -

1. Take the spoken word or vocal sound.
2. Take FFT of the above signal with certain point number.
3. Track the first peak in the FFT output to find the fundamental frequency. Frequency resolution is decided by FFT point number.
4. The fundamental frequency is the FFT point number multiplied by the frequency resolution.

MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCCs)

Speech Feature Extraction is a term which is defined as the fundamental requirement of any speech recognition system. It is represented by the mathematic representation of the speech file. In a human speech recognition system, the main aim is to classify the source files using a reliable representation that reflects the difference between vocal sounds.

FOURIER TRANSFORM

The Fourier transform is defined as an operation that transforms one complex-valued function of a real variable into another domain. There are wide applications of Fourier Transform. Here are such applications like as signal processing and speech processing, the signal before applying FFT will be time domain. However, the FFT converts time domain signal to frequency domain.

For a continuous function of one variable f(t), the Fourier Transform F(f) will be defined as:

\[ F(f) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi ft} dt \]

and the Inverse Transform as

\[ f(t) = \int_{-\infty}^{\infty} F(f) e^{j2\pi ft} df \]

where j is the square root of -i and e denoted the natural exponent.

IV. HIDDEN MARKOV MODEL (HMM) DEFINITION
Hidden Markov Model is defined as the statistical finite state machine in which the system being modelled and hence assumed to be in Markov process. Hidden Markov Model consists of states to model the sequence of observation data. But here the states to model are not visible but the output data is completely depends on the states input. It generally called as statistical Baysian dynamic network with Markov rule considering into account.

**MARKOV PROCESS**

The Hidden Markov process states that the transition to next state depends only on the current state and its output probability of current state but not on all the previous states. Therefore, an above process is called as first order Markov process. Suppose, if transition to next state depends on k previous states then that process is called as kth order markov process.

**HIDDEN MARKOV MODEL MOTIVATING EXAMPLE BELOW**

Let us consider three containers or called then urns which are named as U1, U2 and U3 which contains 100 balls each. And these each 100 balls are of three different colours Red, Green and Blue in some proportion as shown below :-

Container 1 :- U1

Red=60 Red=10 Red=40

Container 2 :- U2

Green=40 Green=10 Green=50

Container 3 :- U3

Blue=30 Blue=40

Figure 4.1: HMM example with containers and balls

**V. EXPERIMENTAL RESULTS**

There are number of experiments have performed with different number of speakers by getting different spoken vocal sounds are taken into consideration while obtaining the overall recognition rate using Hidden Markov Model (HMM). Here in this chapter, results have been obtained for speaker independent, small vocabulary isolated spoken word recognition are discussed. The main objective or aim of the thesis is to recognise or to detect set of ten spoken vocal sounds or words. We considered two set of words consists of : { one, two, three, four, five, six, seven, eight, nine and ten}, { apple, mango, hello, hi, calm, weight, vain, hello, poem, key, move}. These words recorded with microphone using audacity software through which recorded signals are directly represented in ‘.wav’ format or they are in the format ‘.wav’. Along with this, Many different models were taken into consideration like frame length variation, order of Mel frequency cepstral coefficients and linear frequency cepstral coefficients and their delta coefficients.

Feature extraction : MFCC and LFCC and their delta coefficients
Order features : 13 ( 12 MFCC/LFCC + Log energy)
Filter banks : Triangular (both Mel and linear)
Frame size : 256

Confusion matrix is used for comparing results of finite machine classification methods like Hidden Markov Model used for word classification.
Figure 1: Percentage of recognition of each word using MFCC for first data set

VI. CONCLUSIONS

All over, here is a conclusion of this thesis work and that is just an extension of the work which was done by LAWRENCE RABINIER, a developer of HMM tool kit, Naval post graduate school and so many. The main objective behind thesis is to get good speech recognition rate at high end of the human voice generally female voice. Here we get increase in number of spoken words and hence finally to improve the overall speech recognition rate of isolated spoken vocal sounds. The above mentioned and displayed results conclude that the work in this thesis project improves the speech recognition especially at the high end of the human voice. All the objectives or aims proposed in this thesis project were successfully achieved by implementing MFCC, LFCC and discrete hidden markov models in MATLAB programming language.

REFERENCES