Speech samples recognition based on MFCC and vector Quantization

1Ch. Ramaiah  
Department Of Computer Science and Engineering  
V.R.Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India  
ramaiah.challa@gmail.com  

2Dr. V. Srinivasa Rao  
Department Of Computer Science and Engineering  
V.R.Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India  
hodcse@vrsiddhartha.ac.in

Abstract—this paper aims to develop speech sample recognition system regardless their meaning by using well known feature extraction algorithm MFCC with the combination of Vector quantization. Because this combination is easy to apply as well as understand where HMMs are more complicate to implement. It gives good results when compare to so many techniques. Here MATLAB is used here for this implementation. Total system is carried out through two phases. First phase is training phase in this MFCC is applied former for feature extraction. Later VQ code book are generated by trained them. In the second phase testing the matched sample is returned as result. This processing is based on calculation of distance between test sample and pre generated code words.

Key words: MFCC, vector quantization, code book, MATLAB.

I. INTRODUCTION

Speech is one of the natural forms of communication. Recent developments made speech recognition system as secure system. By developing the speech recognition systems we can make the speech samples for interaction between human being and computer. The development of ASR[1] (Automatic Speech Recognition) increased gradually until the invention of Hidden Markov Models (HMM) in early 1970’s. Many of the professional’s contribution were to make use of ASR technology to what can be seen nowadays of various advancements in fields like multi-modal, multi-lingual/cross-lingual ASR using statistical techniques such as HMM, SVM, neural network.

There are so many popular algorithms for the implementation. Generally speech sample recognition system has to be designed through two phases. They are speech sample training phase and speech sample recognition phase. Feature extraction algorithms are applied to loaded speech samples to obtain features. These features are taken to be trained. Trained samples act as reference for future testing.

This speech recognition system will be useful in so many speech based applications. In many fields like Banking, Security, Scientific and Voice based applications are the applicable areas for these speech recognition systems. Speech recognition

II. METHODOLOGY

Our system methodology is illustrated as below
The methodology consists following steps
Step1: extracting the features from the input speech samples using MFCC.
Step2: training the features and generating the code book of them using Vector Quantization technique.
Step3: extract the features of test sample and measure the Euclidian distance to all trained samples in code book.
Step4: return the sample as matched sample which is near to the test sample.

III. FUNCTIONAL CONCEPTS AND ALGORITHMS

A. Speech signal
Speech signal is the basic input to our system. Basically it is an analog signal we can plot speech signal as shown below to observe its amplitudes.

![Speech sample](image)

Figure 2. Speech sample

B. Frequency spectrum
The frequency spectrum of a time-domain signal is a representation of that signal in the frequency domain. We can generate the frequency spectrum by using Fourier transform of the signal. The resulting values are generally presenting as the amplitude and phase, both plotted versus frequency.

Filtering: Filtering means eliminating unwanted components (like frequency etc) from a signal.

C. MFCC Algorithm
The first step in the implementation of any speech recognition system is extracting the features. There are so many well known algorithms for this purpose[2]. MFCC is an efficient feature extracting algorithm. Feature is an individual measurable property heuristic property of a phenomenon being observed. Mel-frequency spectrum is representation of the short term power spectrum of a speech signal.

MFCCs are generated as follows
- Take the Fourier transform of (a windowed excerpt of) a signal.
- Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
- Take the logs of the powers at each of the mel frequencies.
- Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
- The MFCCs are the amplitudes of the resulting spectrum.

The basic flowchart of MFCC algorithm is shown as below figure:

![Flowchart of MFCC](image)

Speech sample: In this system we have to read .wav file (by recording) into our space. With general sample rate is 22050.

Frame blocking: The loaded sample (continues signal) is divided into no of frames of N samples. Adjacent frames are separated by M. we have to consider M value to be lesser than N. so there is overlapping of frames.

Windowing: Here each individual frames from above step are windowed in order to minimize signal discontinuities and spectral distortion. We can use any type of window but hamming window is used to decrease the signal to zero at the beginning and end of each frame. Here we multiply each frame with the considered window function.

Hamming window function

\[ w[n] = 0.08 + 0.4(1 + \cos(2\pi n/N)); \text{ where } 0 \leq n \leq N-1; \]

The representation of hamming window

![Hamming window](image)

FFT: the next step is FFT which is an efficient algorithm to compute DFT. It converts time domain into frequency domain. The difference between FFT and DFT is FFT is faster than DFT.
Mel-filter bank analysis: It uses two types of filters names linear filter and logarithmic filters. Filter bank consists two address spaces one is for linear which for lower than 1000Hz and second for logarithmic which is for higher than 1000Hz. Mels can be calculated by using the below formula

$$\text{Mel}(f) = 2595 \times \log_{10}(1+f/700)$$

Next log function applied for mel spectrum.

DCT: In this step spectrum is converted to time domain. These are the spectrum coefficients.

D. Vector Quantization

There are so many algorithms available for training and feature matching. HMM models, Dynamic Type Warping (DTW), Vector Quantization are the well known. Here we used Vector Quantization (VQ) for out system[5]. VQ is a procedure of mapping vectors from a large vector space of MFCC output to finite number of regions in that space. Each region is known as cluster centered by codeword. Code book is simply collection of codewords. The below diagram illustrates this recognition process.

![Figure 5. Vector quantization concept](image)

After the featured vectors extracted from input speech sample provide a set of training vectors. The VQ code book is developed by using those training vectors. There is an popular algorithm LBG algorithm [ Linde, Buzo and Gray, 1980]. To cluster a set of L training vectors into a set of M codebook vectors the algorithm is follow the below steps[6].

Step1: Design a 1-vector codebook. This is the centroid of the entire set of training vectors.

Step2: Double the size of the codebook by splitting each current codebook. According the rule

$$y_n^+ = y_n(1+\epsilon)$$

$$y_n^- = y_n(1-\epsilon)$$

where n varies from 1 to the current size of the codebook, and \(\epsilon\) is a splitting parameter (we choose \(\epsilon=0.01\)).

Step3: Nearest-Neighbor Search: for each training vector, find the codeword in the current codebook that is closest

Step4: Centroid Update: update the codeword in each cell using the centroid of the training vectors assigned to that cell.

Step5: Iteration 1: repeat steps 3 and 4 until the average distance falls below a preset threshold

Step6: Iteration 2: repeat steps 2, 3 and 4 until a codebook size of M is designed.

The LBG algorithm designs an M-vector codebook in stages. It starts first by designing a 1-vector codebook, then uses a splitting technique on the codewords to initialize the search for a 2-vector codebook, and
continues the splitting process until the desired M-vector codebook is obtained.

IV. RESULTS

Considered two samples

![Waveform of sample1 with frequency 22050](image1)

![Waveform of sample2 with frequency 16000](image2)

After MFCC:

![MFCC feature vectors for sample1](image3)

![MFCC feature vectors for sample2](image4)

Figure 6. Representation of MFCC feature vectors

After generation of VQ code:
After training phase:

```matlab
>> testing('test1.wav')
test1.wav is ok.
test2.wav is ok.
test3.wav is ok.
test4.wav is ok.
test5.wav is ok.
test6.wav is ok.
test7.wav is ok.
test8.wav is ok.
```

**CONCLUSION**

In this system we tested 8 samples regarding their meaning by kept them in trained directory. In testing phase we test one sample. Our systems successfully recognize the matched sample from trained directory. It has good recognition capability how ever systems based on HMM are having highest rates they are very complicated to develop. So this is considerable system.

**REFERENCES**


VQ codes for both samples

![VQ codes for both samples](image.png)


Mahmoud I. Abdalla and Hanaa S. Ali “Wavelet-Based Mel- Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models".