International Journal of Research (IJR) Vol-1, Issue-8, September 2014 **ISSN 2348-6848**

Speaker Identification using Mel-frequency Cepstrum Coefficient

Lalit G. Patil

Department of Electrical Engineering

The M.S.University of Baroda, Gujarat, India

lalit.gecgnr@gmail.com

ABSTRACT

Now a days, Speaker identification is most popular technology and there is a much little research on it for practical application. When a human speaks any word, it contains certain frequency and amplitude. Speaker identification is the technology in which developed algorithm recognizes a word spoken by user, and it deals with identification of a person. Its applications are device control like laptop, door opening, Voice based dialing, and other security purpose also. This technology deals with signal processing. Different features are extracted from spoken words and this is the most important stage with which a programmer have to deal.

Key words- *Speaker identification, Signal Processing, DCT, FFT, MFCC*

1. INTRODUCTION

Speech or word given by user depends on various factors like as we know Speech is Time-varying signal, Well-structured communication process, Depends on known physical movements, distinct units, is different for every speaker, May be fast, slow, or varying in speed, May have high pitch, low pitch, or be whispered, Has widely-varying types of environmental noise, May not have distinct boundaries between units, has an unlimited number of words. Linear predictive coding is a one of the technique to recognize the speech.

A Human can speak at maximum of 10 kHz frequency and minimum at 70Hz. These are the minimum and maximum values and it changes for every person. The magnitude of sound is expressed in db. A normal human speech is between the interval of 100Hz to 3200Hz and its magnitude is in 30 db-90 db. Human can hear a frequency range between 16Hz and 20k Hz.

2. PROPOSED ALGORITHM

There are two stages from which a speech or algorithm have to pass. One stage is training phase in which samples of speech of user are stored, and different feature sets are extracted from them and stored for future use. Then second stage is testing phase in which input speech of single user is taken, then again its features are extracted and it is compared with training feature data sets. According to mean squared error, a person is selected and if it

matches then only access of device is given to that user otherwise access will be denied.

FEATURE EXTRACTION

Obtaining the characteristics of the speech signal is referred to as Feature Extraction. Feature Extraction is used in both training and testing phases. It has following Algorithm.

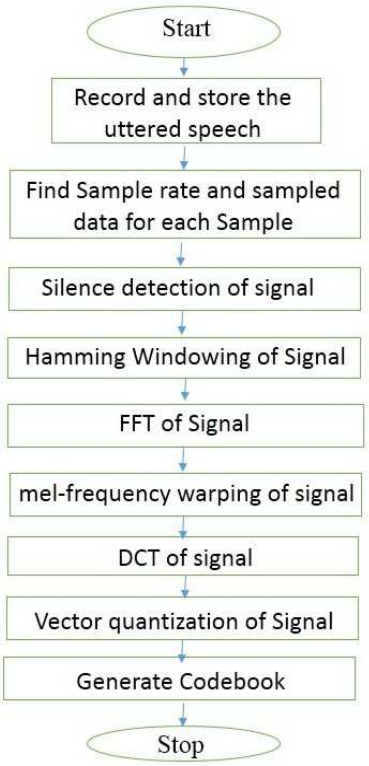

Figure 1: Algorithm for Speaker identification

The important amount of sound if frequency. The sounds are discriminated from each other based on frequency. When the frequency of a sound increases, the sound gets high-pitched and irritating. When the frequency of a sound decreases, the sound gets deepen. Sound waves are the waves that occur from vibration of the materials. In start of this algorithm a word spoken by user is recorded by recording machine. That word contains certain amount of frequency and amplitude as shown below in figure:

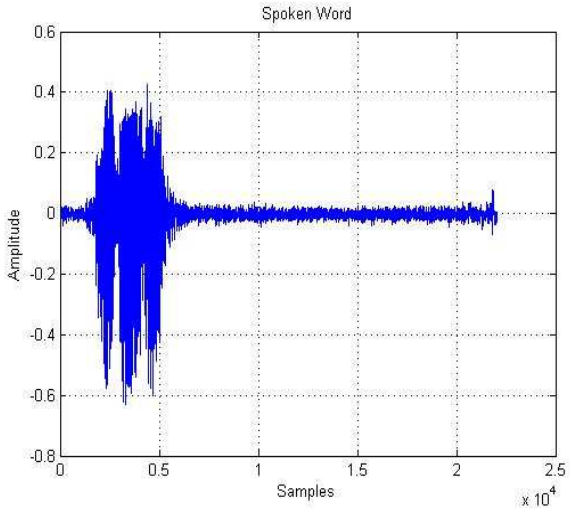

Figure 2 : Speech Signal of Word “Baroda”

SILENCE DETECTION

We have taken our audio word to be sampled at 11025 Hz for 2samples. It means device will allow to user to speak only during 2 Seconds. We take a speech signal as 11025 Samples but after observing the signal we decide that for particular word it takes approximately 300 Samples after removing static portions from uttered speech. So after using a simple thresholding we extracted the actual uttered speech.

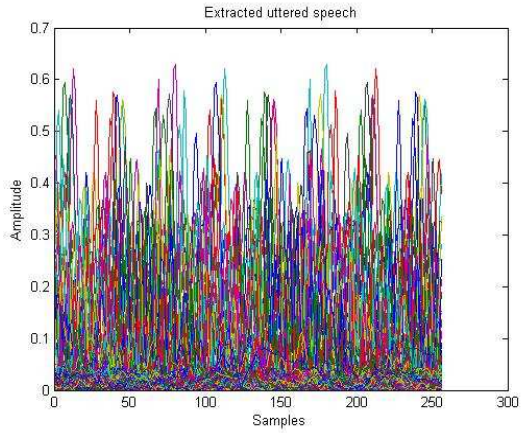

Figure 3: Speech Signal of Word "Baroda" after silence detection

HAMMING WINDOWING

$$w(n) = 0.54 - 0.46 * \cos\left(2\pi \frac{n}{N}\right),$$

$$0 \leq n \leq N$$

The Window length is L = N + 1.

Hamming window is to smooth the speech signal.

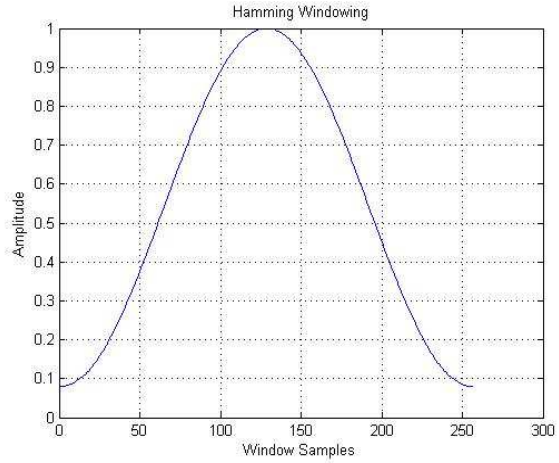

Figure 4: Speech Signal of Word "Baroda" after hamming windowing

FAST FOURIER TRANSFORM

It transform signal to frequency domain

$X(k) = \sum_{j=1}^{N} x(j)\, w_N^{(j-1)(k-1)}$

$X(k) = \left(\frac{1}{N}\right) \sum_{k=1}^{N} x(k)\, w_N^{-(j-1)(k-1)}$

Where,

$w_N = e^{\frac{(-2\pi i)}{N}}$ Is an N[th] root of unity.

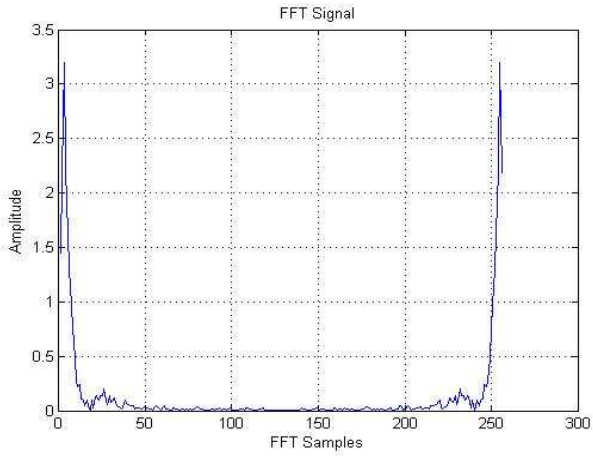

Figure 5: Speech Signal of Word "Baroda" after FFT

MEL-FREQUENCY WARPING

As we know that a human ear can perceive a sound of logarithmic nature. So, frequency is a linear scale, and mel is the logarithmic scale. So, we taken a mel-frequency into consideration and Linear to Mel frequency scale conversion and equation for it will be:

mel (f) = 2595 * $\log_{10}$ (1 + f/100)

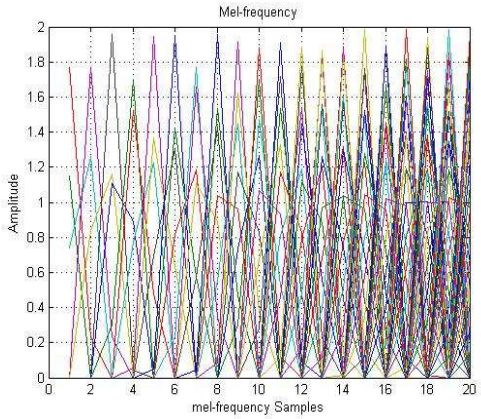

Figure 6: Speech Signal of Word "Baroda" after mel-warping

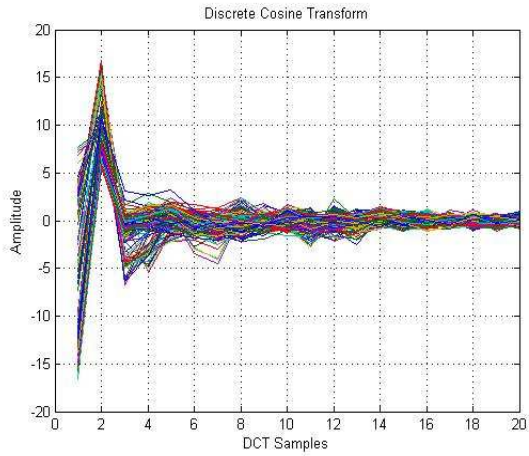

Figure 7: Speech Signal of Word "Baroda" after DCT

DISCRETE COSINE TRANSFORM

It Determine DCT of Log of the spectrum energies.

y(k) = w(k)

$$\sum_{n=1}^{N} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right)$$

$$k = 1,2,3, \dots N$$

Where,

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}} & k = 1 \\ \sqrt{\frac{2}{N}} & 2 \le k \le N \end{cases}$$

VECTOR QUANTIZATION

Now vector book is required to find a minimum distance between trained word database and test word database. There is a well-known algorithm, namely LBG algorithm [Linde, Buzo and Gray, 1980], for clustering a set of L training vectors into a set of M codebook vectors.

To generate this code book we need to Cluster the training vector. So, its algorithm follows the sequence as below:

1. Design a one vector code book, means centroid of training data set by finding a mean. Here we developed a code for 16centroids.

2. Doubling the size of each code book by splitting each current code book.

r+ = r*(1+e)

r- = r*(1-e)

r varies from 1 to the current size of the codebook, and e is a splitting parameter

Here e = 0.0003

3. Searching for Nearest-Neighbor Search: for each training vector.

4. Centroid Update t = t + x(q); x = Distance for each training data vectors

5. Repeat steps 3 and 4 until the average distance falls below a pre-set Threshold

6. Repeat steps 2, 3 and 4 until a codebook is designed

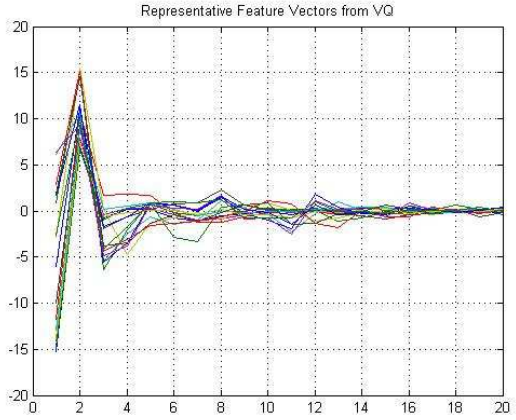

Figure 8: Vector Quantization generated using LBG Algorithm

3. EXPERIMENTAL RESULT

Two code books are generated, one for training data set and one for test data set. And determine the Mean squared error. Based on it user is selected and "Access Granted" is output if within threshold or "Access Denied" is output in case of a failure. So a correct person gets recognized or identified.

The output of a matched person is sent to the COM port of computer or at output port of DSP Processor, and accordingly application is controlled. This algorithm is itself fast enough, but for higher speeds one can go for higher processor.

Here speaker 2 is found the authenticated user, its user id is sent by MATLAB to PLC and the changes of value on PLC is shown below.

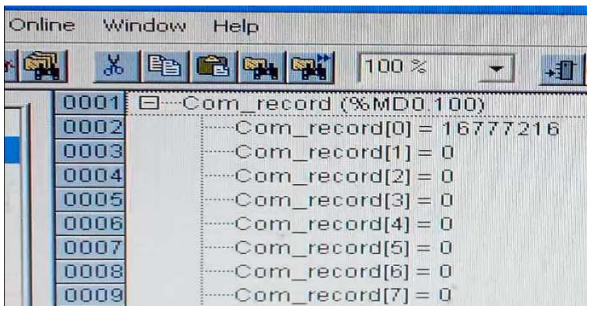

Figure 9: Data received

And according to the data sent the output on PLC is shown below. From the image provided below a one can see that at right column user two is detected and at first column accordingly device eight is being controlled.

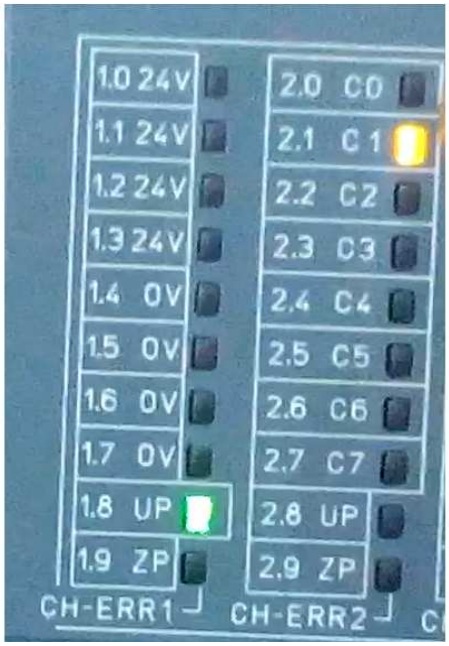

Figure 10: Output on PLC

This is DC 523 which is used as data out port whatever data is sent to the plc according to the program it sends the output on it and a LED shows the status of each pin.

4. CONCLUSION

For speaker identification, the proposed method is perfectly accurate. By much trial and error its threshold and tracking accuracy has been determined accurately for particular users. This algorithm works much fast enough and error rate is much low. And this algorithm works with practical implementation of hardware device control.

5. APPLICATIONS

Speaker identification system has many applications in almost all range of areas related with engineering and technology. Some of the applications of speaker identification systems are:

- ✓ Time and Attendance Systems
- ✓ Access Control Systems
- ✓ Telephone-Banking/Broking
- ✓ Biometric Login to telephone aided shopping systems
- ✓ Information and Reservation Services
- ✓ Security control for confidential information
- ✓ Forensic purposes

6. FUTURE SCOPE

The proposed focus is on Isolated Word recognition. This research can be extended for continuous speech recognition for numerous applications.

7. REFERENCES

[1] Balwant A. Sonkamble1* D. D. Doye2, Speech Recognition Using Vector Quantization through Modified K-means LBG Algorithm, Computer Engineering and Intelligent Systems ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol 3, No.7, 2012

[2] Davis, S.; Mermelstein, P.: "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 4 (1980).

[3] DR. H. B. Kekre , VaishaliKulkarni, Ph.D. Research Scholar , Performance Comparison of Speaker Recognition using Vector Quantization by LBG and KFCG International Journal of Computer Applications (0975 – 8887)Volume 3 – No.10, July 2010 32

[4] Hakkani-Tur, D.; Oflazer, K.; Tur, G: "Statistical Morphological Disambiguation for Agglutinative Languages", Technical Report, Bilkent University, (2000).

[5] Lawrence Rabiner, Biing-Hwang Juang – "Fundamentals of Speech Recognition"

[6] Md. Rashidul Hasan, Mustafa Jamil, Md. GolamRabbani Md. SaifurRahman , Speaker identification using mel-frequency cepstrul coefficients, 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh ISBN 984-32-1804-4 565

[7] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy and Kong-Pang Pun – " An Efficient MFCC Extraction Method in Speech Recognition", Department of Electronic Engineering, The Chinese University of Hong Kong, Hong, IEEE – ISCAS, 2006

[8] Ursin, M.: "Triphone Clustering in Continuous Speech Recognition", MS Thesis, Helsinki University of Technology, Department of Computer Science, (2002).