Voice Command Based Matlab GUI for Microcontroller

Md. Saiful Islam[1], Md. Shajid-Ul-Mahmud[2]
Md. Akhtaruzzaman[3]
[123]Aeronautical Engineering Department
Military Institute of Science and Technology
Email:[1]tarekislm39@gmail.com,[2]sumahmud19@gmail.com
[3]showrov.aero@gmail.com

Md Hossam-E-Haider[4]
[4]Electrical, Electronic and Communication Engineering
Department
Military Institute of Science and Technology
Dhaka, Bangladesh
Email: [4]haider8400@yahoo.com

***Abstract*—Voice command based applications use speech to make people's life more easy and flexible. It is one of the most interesting topics in recent research spheres and still there are lots of scope for improvement. The paper represents the method of controlling electrical components by Matlab graphical user interface (GUI) using voice command. The idea of developing voice controlled application can be established through speech identification and speech verification. Speech identification is performed by feature extracting of speech signal utilizing Mel Frequency Cepstrum Coefficient (MFCC) technique in Matlab. Then vector quantization (VQ) method has been proposed for verification. The recognized speech will be compared with prerecorded database and according to those results the desired commands are sent from application to microcontroller to control the devices.**

***Index Terms*—Speech identification; speech verification; Mel Frequency Cepstrum Coefficient (MFCC); Feature Extraction; Feature Matching; Vector Quantization (VQ)**

I. Introduction

The speech is very useful way for communication between human and machine. Voice command based application refers to utilize human voice. It is the process of recognizing individual speech after comparing it from previous stored data. The overall system is established in Matlab environment and the speeches are gathered through microphone. The system comprises of two components that are speaker identification and speaker verification. In speaker identification, firstly the speech has to be gotten as input and using various techniques each speech has its own features. Many methods work for these either cepstral or spectral domain of the signal.

At first voice have to be converted into discrete time step. It is called feature extraction. MFCC works the task and stores the samples in Matlab database. MFCC is performed for both the training phase and testing phase. For speaker verification stage feature matching is done using Vector Quantization (VQ). This method consists of extracting feature vectors and generation of codebook. Then the system is tried to find the distortion of the input voice and codebook stored in the training phase. Based on the result the decision is made whether it is acceptable or not. If the speech signal is matched according to any of the recorded signals, the predesigned command should send to microcontroller. Using UART communication from PC to microcontroller, the command will be converted to the serial data. The microconroller unit thus controls the electrical devices.

This paper proposes a methodology by which human voice is recognized using MATLAB and follows particular commands to control electrical devices such as stepper motor.

Remaining parts of the paper are structured as follows: Section (II) Literature Review (III) System Overview (IV) Voice Recognition (V) Simulation (VI) Hardware Section (VII) Tests and Results (VIII) Conclusion and Future Work

II. LITERATURE REVIEW

The literatures for human machine interaction using speech recognition are described. The methods for implementation of an automatic speech recognition system are discussed elaborately[1], [2], [4].The methodologies and algorithms that are divided into several categories: Hidden Markov Model (HMM), Dynamic Time Warping (DTW), Vector Quantization (VQ), Artificial Neural Networks (ANN). Artificial Neural Networks (ANN) is very much similar to Markov Models. Markov Models normally use the probability on the other hand connection strengths and functions are used for ANN. But two of these also have disadvantages. While ANN has challenges to find appropriate weights and HMM has the problems of finding required transitions. Hidden Markov Models (HMM) is used highly in speech recognition system because of its reliability [5]. But it is very complex to implement. ANN has also great accuracy[5], [3]. For easiest and simplest recognition system MFCC and VQ approach is very efficient [6], [8]. However only MFCC cannot provide the desired level of efficiency but in combination of VQ approach the method is highly appreciated. The combination of the MFCC and VQ algorithms in speech recognition system is simpler enough while HMM and DTW have complexity[6], [7]. For speech recognition system MFCC calculates feature extractions and for feature matching VQ gives the desired result.

III. SYSTEM OVERVIEW

The overall system is divided into two sections, one is software and another is hardware. For software section Matlab and for hardware section microcontroller unit have been worked.The project aims to design a speech recognition system for controlling stepper motor using voice commands. Speech is recorded using microphone and after analysis from Matlab

978-1-5386-2175-2/17/$31.00 ©2017 IEEE

simulator microcontroller unit is used to operate the devices. The command is received by the microcontroller unit through Matlab serial communication.That work in both the testing and training sections. Block diagram of the entire system is shown in fig 1.

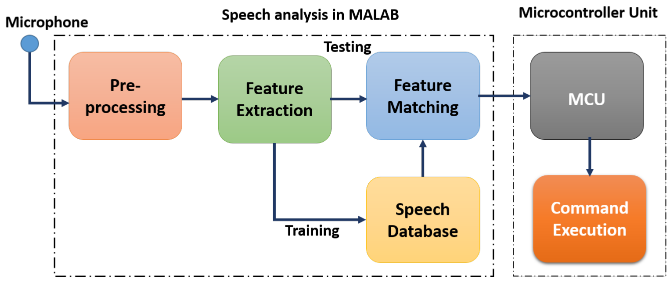


Fig. 1. System Block Diagram

IV. VOICE RECOGNITION

As previously mentioned speaker recognition have two basic phases that are training and testing. Testing phase matches the incoming speech signal with the reference training phase. Voice recognition have two parts including feature extraction and feature matching. Fig 2. shows the block diagram of voice recognition system.

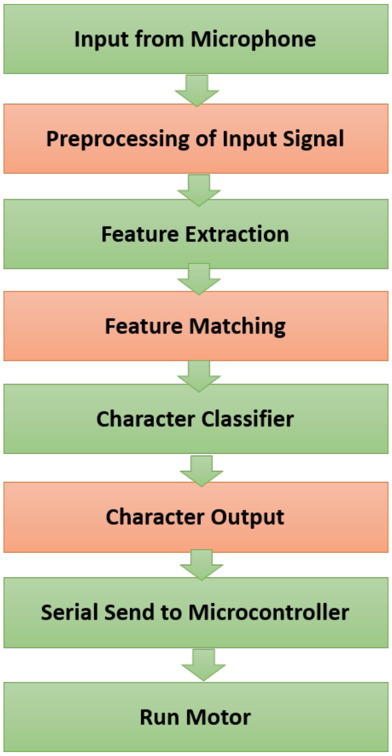


Fig. 2. Flowchart of Speech Recognition

A. FEATURE EXTRACTION

The aim of feature extraction is to analysis speech signal for recognition that is done by simplification of acoustic properties of the individual speech. Feature extraction process converts the digitized signal into numerical sequence. It is often called as reduction process. It reduces the vast amount of data by taking only the required characteristics of each signals. Among various techniques MFCC is most reknown. It works based on perception of hearing. Two types of filters are worked under this method. One is for low frequency having linearly spaced and logarithmic spaced for high frequency operation[6], [7], [8]. Block diagram of steps for the MFCC are shown in Fig 3.

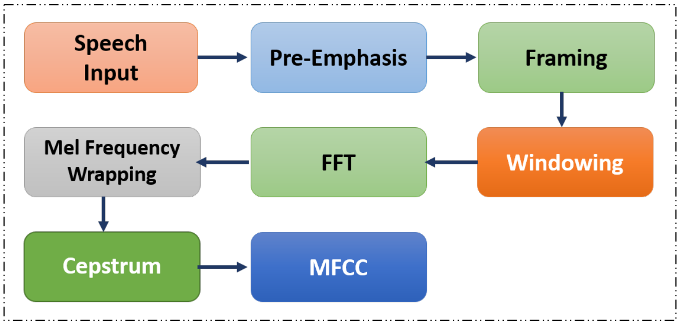


Fig. 3. Working Block of Feature Extraction

1) PRE-EMPHASIS: This section passes the input signal through a filter to increase the energy of signal. By this technique the SNR of the signal will be improved.

$$y(n) = x(n) - 0.9x(n-1) \tag{1}$$

The main purpose of the pre-emphasis is to make up the high frequency section that was lost during sound generation.

2) FRAMING: For speech analysis, short time interval is necessary because it is assumed that speech signal is not stationary for long term. The speech signal operates uniformly for 10-30 ms intervals or frames. In frames of P samples for a speech signal, the adjacent frame being Q. So the second frame will be overlapped by P-Q. The process continues and third frame will be P-2Q sample. Typically the frames P and Q are 256 and 100.

3) WINDOWING: For minimizing the discontinuities of the signal from beginning to end of every frames, all the frames should be gone through hamming window. Spectral distortion is also be reduced. For different alpha values the Hamming windows can be worked. In figure 4, hamming window of different alpha values have been shown. But typically the alpha value is normally 0.46.

$$S(t) = alpha - (1 - alpha)cos(\frac{2\pi t}{N-1}) \tag{2}$$

$$P[t] = R[t] * S[t] \tag{3}$$

Here,
N = number of samples in each frame, $0 < t < N - 1$
P[t] = Output of the signal
R(t) = input of the signal
S[t] = coefficient of hamming window

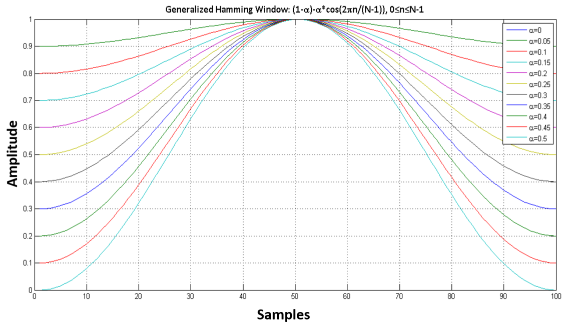


Fig. 4. Representation of Hamming Window

4) Fast Fourier Transform (FFT): FFT reconstructs this signal from time domain to frequency domain for each of the N samples. From this output it is called as spectrum.

$$X_p = \sum_{n=0}^{N-1} x_n e^{\frac{-j2\pi kn}{N}} \tag{4}$$

X_p are complex numbers and p= 0,1,2....N-1

5) MEL-FREQUENCY WRAPPING: Human frequency does not contain linear scale. For this a special frequency scale mel scale have been introduced.It Has the frequency spacing from below 1000 Hz to 1000 Hz greater than the logarithmic spacing. Using triangular band pass filters mel scale filter bank is designed.

$$Mel(f) = 2595 * log_{10}(1 + \frac{f}{700}) \tag{5}$$

DISCRETE COSINE TRANSFORM OR DCT: Finally the mel spectrum is changed to time domain. It is known as mel frequency Cepstrum coefficients (MFCC). Using DCT Discrete Cosine Transform (DCT) the conversion is made. The formula for this conversion is given below,

$$C_n = \sum_{j=1}^{j} (logP_j)[n(j - \frac{1}{2})\frac{\pi}{j}] \tag{6}$$

here, n=1,2....j and P_j is FFT Coefficients

B. FEATURE MATCHING

Speaker verification is the process for comparing an unknown speech with a set of prerecorded speeches from the database to find the best possible matching. The method of VQ comprises of a few number of feature vectors. But storing each vector that originates from the training section is very hard. From the recorded data, feature vectors are clustered and create a codebook. When the tested section appears the input signal will find the similarity to the codebook of every speech and finally calculate the difference between them. If the difference is in specific limit the decision is positive. For implementation the method two basic components are needed. One is K-means Algorithm and Euclidean Distance.

1) K-MEANS ALGORITHM: The K-means algorithm is a dynamic approach used for clustering the feature vectors[10]. Here, K is pre-designed and represents the required number of clusters those are needed. From the recorded speech the code vectors have to compute and it can start with an arbitrary estimation of code vectors until satisfied the criterion.[14] The K algorithm makes the partition of M= M1,M2,M3,,,,MT. feature vectors into specific N= N1,N2,,,,,NT centroids. That is called codewords. The set of codewords are codebooks. After choosing N cluster centroids each feature is then assigned to near most centroid. So new centroids are measured. This process continues until specific criterion is received.[13], [12] This is called mean square error among the feature vectors.

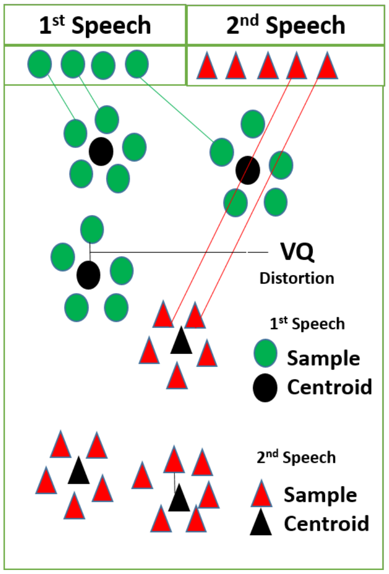


Fig. 5. Formation of VQ Codebook

In Fig 5. represents two speech signal codebooks that are green and red color. Here green color circle is code vector and black is centroid for first speech. And red color triangle is code vector and black triangle is centroid for second speech. The distance between the vectors and centroids is VQ distortion.

2) EUCLIDEAN DISTANCE: The feature vectors of speech signal is a sequence like $x1, x2, , , , , xi$ and it will then compare to the codebooks. according to verify the speech, the difference of two vectors sets should be minimum. It can be done by using Euclidean distance. The Euclidean distance is referred as the distance between two given points using Pythagorean Theorem. [15]The formula for the Euclidean distance is:

$$\sqrt{(p_1 - q_1)^2 + ...(p_t - q_t)^2} = \sqrt{\sum_{i=1}^{t} (p_i - q_i)^2}. \tag{7}$$

Where p=p_1, p_2, p_3, p_4 and q=q_1, q_2, q_3, q_4 are two points. The speech which is the lowest distortion distance is said to be the recognized speech.

V. SIMULATION

This segment exhibits the reproduction techniques of stepper engine control utilizing the Proteus and MATLAB programming interfaced with Virtual Serial Port. Keeping in mind the end goal to control the stepper engine, the MATLAB interfacing with Proteus programming is accomplished by the blend of GUI, microcontroller and virtual port correspondence amongst MATALB and Proteus. At the point when choices are chosen from the GUI the allocated ASCII code will exchange to the microcontroller by means of the COM port. When the voice recognition is successfully made the previously assigned character will be come to the microcontroller. Here using virtual com port connection with Matlab serial interface is created using Elitma virtual serial port software.

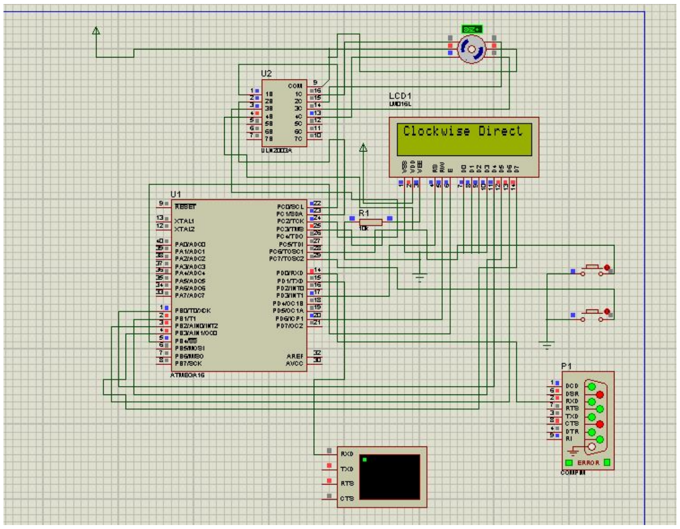
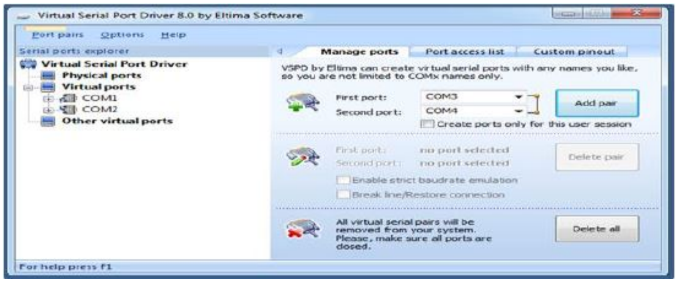


Fig. 6. Proteus Design of Hardware Part


Virtual Serial Port Driver 8.0 by Eltima Software

Port pairs Options Help

Serial ports explorer

- Virtual Serial Port Driver
  - Physical ports
  - Virtual ports
    - COM1
    - COM2
  - Other virtual ports

Manage ports | Port access list | Custom pinout

VSPD by Eltima can create virtual serial ports with any names you like, so you are not limited to COMx names only.

First port: COM3
Second port: COM4
Add pair
□ Create ports only for this user session

First port: no port selected
Second port: no port selected
Delete pair
□ Enable strict baudrate emulation
□ Break line/Restore connection

All virtual serial pairs will be removed from your system. Please, make sure all ports are closed.
Delete all

For help press F1

Fig. 7. Virtual COMPORT using ”Elitma Software”

VI. HARDWARE SECTION

For implementation of the project Atmega microcontroller, Stepper motor, stepper motor driver uln2003a have been used. To interface the USB port of the computer with the system developed, the USB-to-RS232 convertor is used. The USB port as a serial port protocol is initialized by the driver of USB-to-RS232 convertor. By transforming the RS232 levels back to 0 and 5 Volts with the use of common level connector that is MAX232 the CMOS level is converted into Transistor Logic (TTL). The receiver transmitter pin of MAX 232 is connected to receiver transmitter pin of microcontroller. The TX and RX of the USB serial converter is connected microcontroller TX, RX to the section. When voice recognition option is chosen, for On, Off, Up, Down four commands send following O, F, U, D characters.

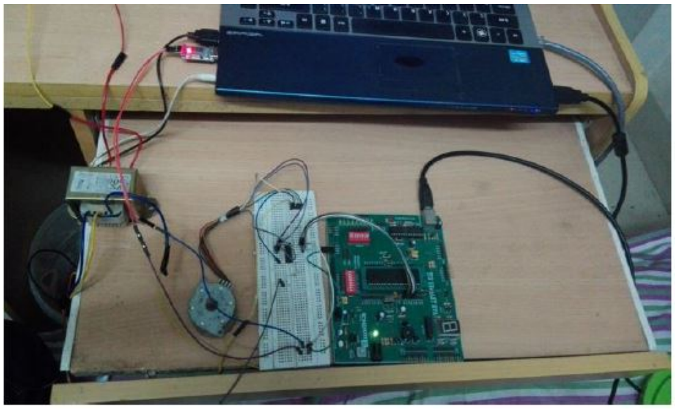


Fig. 8. Hardware Establishment

A. Tests and Results

Matlab GUI has been utilized for controlling the motor.

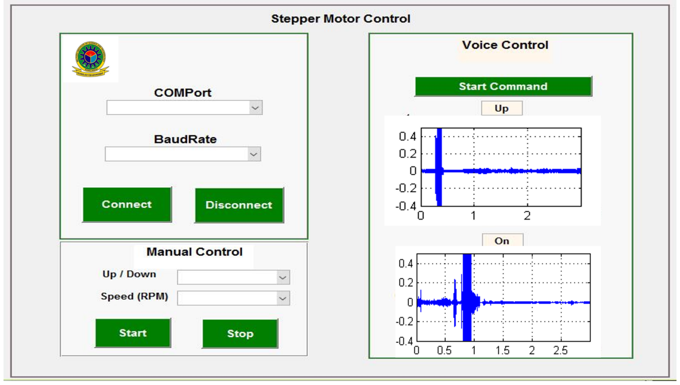


Fig. 9. Matlab GUI to control the Motor

In this experiment four speech On, Off, Up and Down have been stored in the train section. Using the feature extraction every signals framing, windowing, power spectrum, energy spectrum, mel frequency, acoustic vector, quantization vector have been observed. Here for speech signal On, the observations are shown from Fig. 11 to Fig. 13 .

Here, s2(n) = s(n) - a*s(n-1), s2(n) is the output signal and the value of a is normally between 0.9 and 1.0. In our experiment the value of a is 0.95.

While taking a frame it should be constructed in variable size to maintain always integer number of fundamental periods for windowing purpose. At the time of performing the FFT for each frames, it is assumed that the speech signal within

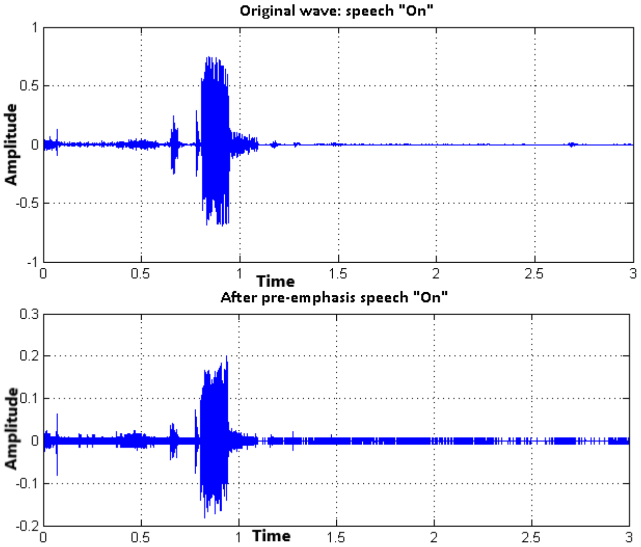


Fig. 10. Before and after Pre-emphasis of ”On” speech

this frame is continuous and periodic. But if the thing is not happened when wrapping around can be influenced on frequency response. So for this Hamming window is introduced. It multiply each of the frames to improve the continuity for the first and last points. In the above figure the frequency response for the speech On using Hamming window is much sharper.

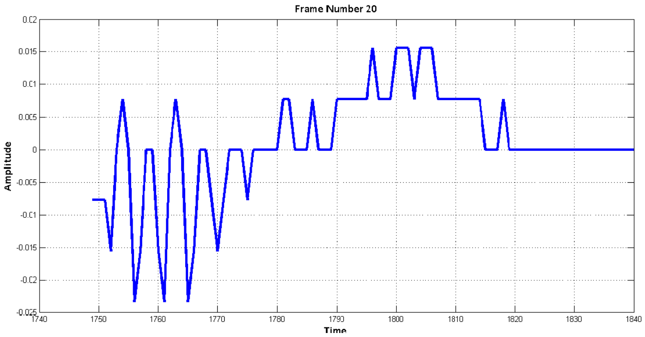


Fig. 11. Framing of Speech ”On”

Using FFT, Representation of speech for frequency domain is established in Fig. 12. Power spectrum of speech ”On” has been also presented in Fig. 13.

The Mel cepstrum is as like as real cepstrum but it has been warped for the following mel scale in Fig. 14. Fig. 15 and 16 are shown the acoustic vector and vector quantization.

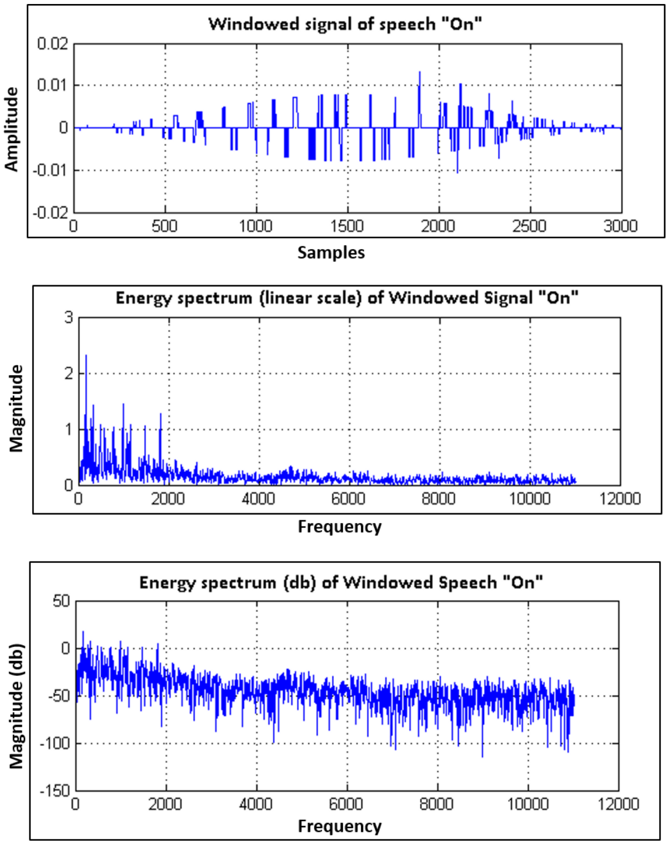


Fig. 12. Windowed and FFT Spectrum of Speech ”On”

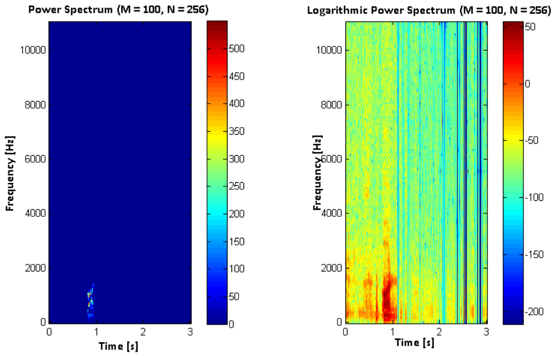


Fig. 13. Power Spectrum and Logarithmic Power Spectrum of Speech ”On”

VII. Conclusion and Future Work

The voice recognition system using Matlab have been successfully implemented. The manual command and voice recognized command are displayed in Matlab GUI. The specified data is sent from Matlab to MCU. This speech recognized technique can be very useful for blind people. The experiment of this study was tested in noiseless and noisy environment. Both times, the results give accurate performance. It has been observed that increasing numbers of centroids of the VQ provides satisfactory result. The training session has to be repeated as to update the codebooks in the database for better performance. The system may be distracted from its capability of recognition for specific voice as human speech changes over two or three years. This application is very easy to control and cost effective. It can also use for security purposes. In future it can be enhanced to design and operate for real time applications like robotics, home automation.

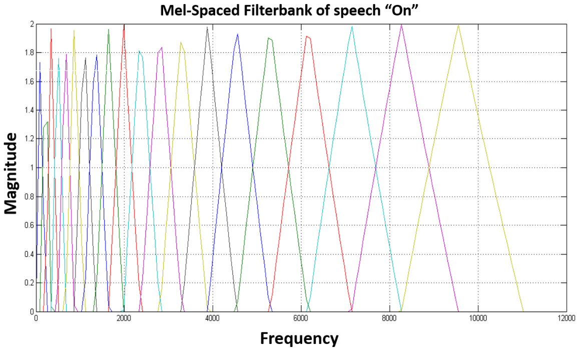


Fig. 14. Mel Frequency of the signal ”On”

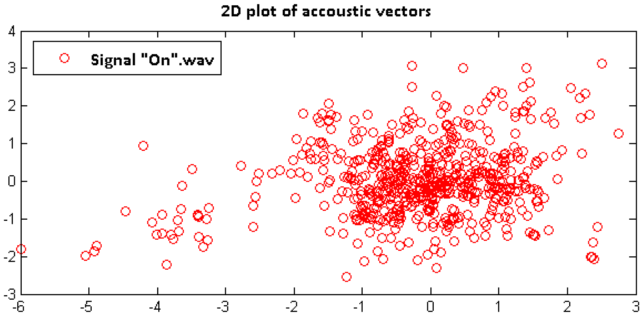


Fig. 15. Acoustic Vectors Representation in VQ

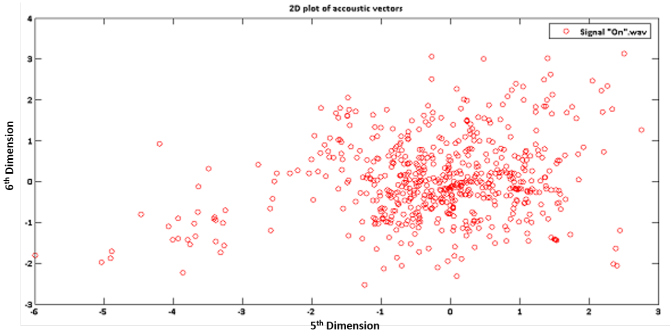


Fig. 16. Quantization Vectors of Signal ”On”

VIII. Acknowledgement

Authors are very much thankful to Mahima Garg and his group members for their voice recognition toolbox which helped a great extent in the entire program.

References

[1] Krishan Kant Lavania, Shachi Sharma and Krishna Kumar Sharma, Reviewing Human-Machine Interaction through Speech Recognition approaches and Analyzing an approach for Designing an Efficient System International Journal of Computer Applications, Volume 38, 2012.

[2] Aarti V. Jadhav and Rupali V. Pawar, Review of Various Approaches towards Speech Recognition, International Conference on Biomedical Engineering, pp: 99 103, IEEE, 2011.

[3] M A Muqeet, Speech Recognition using Digital Signal Proccessor, unpublished.

[4] Katti and Anusuya, Speech Recognition by Machine: A Review, International Journal of Computer Science and Information Security, Volume: 64, pp: 501 - 531 IEEE, 2009.

[5] L. Rabiner, A tutorial on Hidden Marcov Model and selected applications in Speech Recognition, Proceedings of the IEEE, Vol.77, No.2, 1989, pp. 257-286.

[6] umbharana, Chandresh K., 2007, Speech Pattern Recognition for Speech to Text Conversion, thesis PhD, Saurashtra University.

[7] Linde Y., Buzo A. and Gray A. M., An algorithm for Vector Quantization, IEEE Transactions on Communication, vol 28., No. 1, 1980.

[8] Y.EphirimA.Dembo L.R.Rabinar,A minimum discrimination approach for hidden markov models,IEEE Transactions on Information Theory vol.35 September 1989.

[9] S. F. Surui, Speaker independent isolated word recognizer using dynamic features of speech spectrum, IEEE Trans.ASSP,volume-34,1986.

[10] Manjot Kaur Gill, Reetkamal Kaur and Jagdev Kaur Vector Quantization based Speaker Identification, International Journal of Computer Applications (0975 8887), Volume 4 No.2, July 2010

[11] A. Bala, Abhijit kumar, Nidhika Birla, Voice Command Recognition System Based On MFCC And DTW, International Journal of Engineering Science and Technology, Vol. 2, No. 12, 2010, pp.7335- 7342.

[12] G. Welch and G. Bishop, An Introduction to the Kalman Filter, UNC-Chapel Hill, TR 95-041, July 24, 2006

[13] Soong, F., A.E., A. R., Juang, B.-H., and Rabiner, L., A vector quantization approach to speaker recognition. AT and T Technical Journal 66 (1987), 1426.3

[14] T. Kinnunen and P. Franti, Speaker Discriminative Weighting Method for VQ-Based Speaker Identification, Proc. 3rd International Conference on audio and video-based biometric person authentication (AVBPA), Halmstad, Sweden, 2001.

[15] wikipedia.org,https://en.wikipedia.org/wiki/Euclidean_distance, December 14, 2016.