

# Robotic Control using Speech Recognition and Android

Gaurav Chauhan, Prasad Chaudhari

Dept. of E & TC Engg., MIT Academy of Engg., [gaurav\\_chauhan15@outlook.com](mailto:gaurav_chauhan15@outlook.com), (M) 9595502999

**Abstract**— Speech processing is becoming more and more popular these days providing immense security. Also, many of the projects undertaken by engineers are based on various platforms neglecting security and authentication. The MFCC method used for speech processing is practically acclaimed and provides better results than its other counterparts namely HMM, LPC, WT etc. Furthermore, Android, a largely popular platform providing powerful capabilities and an open architecture is commonly used to have control over a device. The Development of Radio Frequency transmission has led to carving a new technology “Bluetooth”. Bluetooth converges with android to provide a far better controlling platform. This paper aims to brief and use the practical approach of robotics through a popular platform android and the speech recognition method Mel Frequency Cepstral Coefficients (MFCC). Also, it gives the industry an optimized method for basking in information regarding temperature, humidity, gas leakage in challenging surroundings and provides security with voice authentication.

**Keywords**— MFCC, Android, Bluetooth, Cepstrum, Smartphone, RF module, Sensors, Speech Recognition, Linde-Buzo-Gray, Fourier Transform

## INTRODUCTION

A robot is a mechanical or may be virtually artificial envoy, mostly an electro-mechanical machine that is influenced by a computer program and an electronic circuitry. Robots have replaced human activities in the support of performing those repetitive and dangerous tasks which humans sometimes choose not to do, or are incapable to do due to some inhibitions and size conditions, or even those such as in industries where humans could not survive the extreme environments that may be produced. For such requirements of the industry, this project has aimed to withstand the atmosphere and complete the tasks given by the means of simple control using speech and smartphone. Speech recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. Recognition technique makes it possible to the speaker's voice to be used in verifying their identity and control access to services. The most popular spectral based parameter used in recognition approach is the Mel Frequency Cepstral Coefficients called MFCC [2, 3]. The speech input is processed using MFCC. Commands are assigned using MFCC. Android smartphones are undoubtedly the most popular gadget these days. You will find various applications on the internet that exploit inbuilt hardware in this mobile phone such as Bluetooth, Infrared, NFC and Wi-Fi, to control and manipulate other devices. Presented here is an assignment applying technology to control a robot by using application running on android smartphone. The control commands are dispatched from Bluetooth of the smartphone. The controlling device of the whole system is a microcontroller, Bluetooth module and a pair of DC motors that are interfaced to the microcontroller. The data collected by this Bluetooth module from the Android smartphone is fed as input to the microcontroller. The Microcontroller acts accordingly on the DC motors of the robot. [5] The robot assembly in this venture can be made to maneuver in all four directions using the android smartphone. [6]

## WORKING

The working of the whole system can be divided into two parts –

(A) Control Unit      (B) Robot Unit

These two units consist of the main working of the project and are divided based on the main function carried out.

### Control Unit

At first, an input of speech is taken through the microphone on the computer/laptop. This input is then processed through computing software. [2] A programming code is written to assign a command to the taken input speech signal. These signals which have been assigned commands are then exported from the PC to a wireless RF module (in this case a Zigbee module) using a RS 232 to TTL converting IC (MAX 232). The signals that are in analog nature are converted to digital nature so as to be compatible with the RF

module. This RF module is used to have a wireless control; containing a transmitter on one end and a receiver to the other. The main function of this unit is to have a control over the Robot Unit.

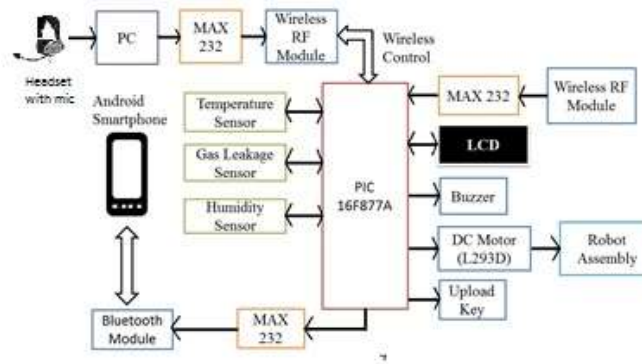


Figure 1: Block Diagram

### Robot Unit

The Robot Unit consists of a main device – The Microcontroller (in this case PIC 16F877A). The main function of this unit here is to drive the robot assembly. The secondary function is to acquire information through the sensors and upload it. Sensors are also interfaced on one side of the microcontroller as shown in figure 1.

The other components interfaced are LCD, RF module (Receiver), Bluetooth module, Robot assembly, buzzer. The signals are received at the RF module which is interfaced to the Microcontroller. According to the signal (command), the sensors work. There are three sensors namely (1) Humidity sensor, (2) Temperature sensor and (3) Gas Leakage sensor. The humidity sensor is used to acquire the information regarding the humidity in atmosphere. The temperature sensor gives the temperature of the surrounding. The gas leakage sensor is used in gas leakage detecting and is suitable for detecting of LPG, iso-butane, propane, LNG, to avoid the noise of alcohol and cooking fumes and cigarette smoke. It alerts if there is any gas leakage through a buzzer which is interfaced to the microcontroller on the other side. The Bluetooth module interfaced to the microcontroller is used to transfer and receive data to/from the smartphone. For android smartphone to have control over the robot, Bluetooth module is used. An android application can be used to control the robot on the smartphone like 'Blueterm' or an application can be programmed using android for a specific use. The info can be uploaded to the PC and to the smartphone by using a switch key called 'Upload Key' in the figure which is again interfaced to the microcontroller. The data is uploaded by the working of Bluetooth and RF module. The signals are given to the motor driver IC that drives the DC motor. The DC motor is used as the legs of the robot. In short, the robot assembly is driven by the motor driver IC. The LCD displays information acquired by the sensors.

### Study of MFCC

The study of MFCC was necessary to start the initialization of the project. Mel-frequency Cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC (Mel-Frequency Cepstrum). They are derived from a type of Cepstral representation of an audio clip. Cepstral representation is a type of representation of a signal in which the spectrum of a signal is obtained. First, the Fourier transform (FFT) of this spectrum is obtained. Second, its logarithm is calculated which then finally results in calculating direct cosine transform (DCT) of this logarithm. The Cepstrum is then acquired in the form of coefficients from the calculated DCT. The difference between the Cepstrum and the Mel-frequency Cepstrum is that, the frequency bands are uniformly spaced on the Mel scale, which approximates the human auricular system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. These MFCCs are then used in programming for further representation. MFCC is an optimized technique for speech processing than its less efficient counterparts like HMM, DWT, LPC. In speech processing we generally use the real cepstrum, which is obtained by applying an inverse Fourier Transform of the log spectrum of the signal. In fact, the name "cepstrum" comes from inverting the first syllable of the word "spectrum". It can be shown that the real cepstrum is the even part of the complex cepstrum [1]. In digital signals, we replace the Fourier Transform by the Discrete Fourier Transform. MFCCs are derived as follows:

1. Take the Fourier Transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the Mel-Scale, using triangular windows which overlap.
3. Take the log of the powers at each of the Mel frequencies.
4. Take the Direct Cosine Transform (DCT) of the list of Mel log powers, assuming it were a signal.

5. The MFCCs are the amplitudes of the resulting spectrum.[4]

Consider a sample speech signal. We represent the 'Spectrogram' of this signal. A spectrogram of a signal is the Time-Frequency representation of a signal. We take a sample speech spectrum which we have to record and play shown in fig 2. From this signal, we have to remove the silent part (including noise) which is considered to be the error from the signal.

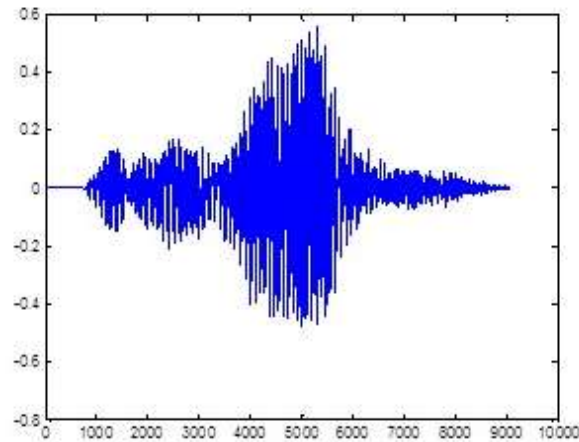


Fig 2: Spectrum of recorded Signal

Our goal is to separate the spectral envelope and the spectral details from the spectrum such that the sum of the former and the latter one is the silence part; an example of this removed error is shown in fig 3. To achieve this separation we use FFT. An FFT on spectrum referred to as Inverse FFT (IFFT). We are dealing with spectrum in log domain. IFFT of log spectrum would represent the signal in pseudo frequency axis. We've captured the spectral envelope. Yet, perceptual experiments have said that a human ear concentrates on certain regions rather than using whole of the spectral envelope.

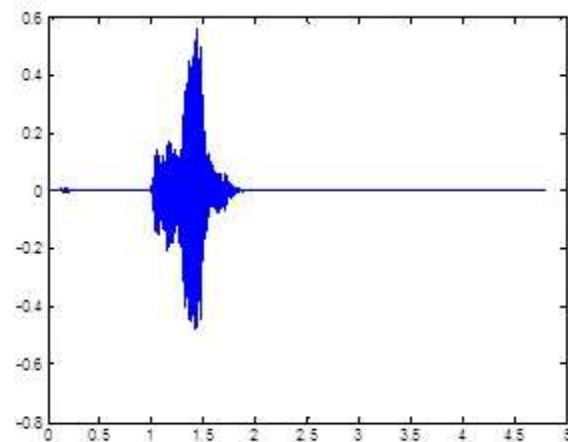


Figure 3: Spectrum of recorded signal w/o silence (noise)

Mel-Frequency analysis of speech is based on human perception experiments. Mel-Frequency Analysis is more closely concentrated on the human auditory system. It is observed that human ear acts as a filter. It focuses on only some particular frequency components. These filters are unevenly spaced on the frequency axis with higher number of filters in the low frequency area and vice-versa. Cepstral coefficients obtained for Mel spectrum are referred to as Mel-Frequency Cepstral Coefficients often denoted by MFCC. MFCC are mostly used features in state-of-art speech recognition system.

### Noise Sensitivity

MFCC values are not very robust in the presence of additive noise, and so it is common to normalize their values in speech recognition systems to lessen the influence of noise. Some researchers propose modifications to the basic MFCC algorithm to improve robustness, such as by increasing the log-Mel-amplitudes to a suitable power (around 2 or 3) before accounting the DCT, which reduces the leverage of low-energy components.

### PREPARATION OF DATABASE

Here, in this project, we have to prepare a database. This database is for the voice signals. The idea behind voice recognition is that firstly, we prepare a database of voice signals in a (.wav) format. Then, finally when we record the voice signals for recognition, they are compared with the database produced. For example, we first record a voice command through a mic in the PC and save them in (.wav) format. While coding, we prepare a loop which will continue comparing the signals recorded with the signals in the database till the distance is approximately met.

Following is the code used to compare the signals:

```
“fopen(comp);  
ifstrcmp(nm,'forward.wav')  
fprintf (comp,'F')  
end  
ifstrcmp(nm,'reverse.wav')  
fprintf(comp,'B')  
end  
ifstrcmp(nm,'left.wav')  
fprintf(comp,'L')  
end  
ifstrcmp(nm,'right.wav')  
fprintf(comp,'R')  
end  
ifstrcmp(nm,'stop.wav')  
fprintf(comp,'S')  
end  
ifstrcmp(nm,'temperature.wav')  
fprintf(comp,'T')  
end  
ifstrcmp(nm,'humidity.wav')  
fprintf(comp,'H')  
end  
ifstrcmp(nm,'mode.wav')  
fprintf(comp,'M')  
end”
```

When the comparing approximates a value nearer to the voice signal saved in the database, it'll round it off and then make a conclusion of the signal recorded. Example, if a “Forward” voice command is saved in the database, the code written will compare the “Forward” signal from user with the signal in database. After it compares, the value of distance is approximated and then it recognizes that the signal is “Forward”. This signal is now coded in a short English alphabet which can be received by the robot with the help of Wi-Fi or Bluetooth. The robot recognizes this with the alphabet sent.

### SPEECH RECOGNITION

Recognition System has two algorithms namely:

- (1) Feature Extraction
- (2) Feature Matching

### Feature Extraction Algorithm

The process of Feature Extraction Algorithm can be stated as follows:

1. First, we block the speech signal into frames of N samples, with adjacent frames having a separation of M (M<N).
2. Second, is to windowing each individual frame resulting in minimization of signal discontinuities i.e. spectral distortion.
3. Third, convert these frames of N samples from Time domain to Frequency domain using FFT.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N}, k = 0, 1, 2 \dots N-1$$

4. Fourth, use a filter bank of triangular band pass frequency response to subjectively simulate the linear scale into mel-scale.
5. Finally, we convert the log mel-spectrum back into time domain resulting in acquiring of MFCCs. These MFCCs are collectively called Mel Frequency Cepstrum.

$$C_n = \sum_{k=1}^K (\log S_k) \cos \left[ n \left( k - \frac{1}{2} \right) \frac{\pi}{K} \right]$$

### Mathematical Representation

Suppose the spectrum of the signal is denoted as  $x[k]$ . The Spectral envelope as  $h[k]$  and the spectral details as  $e[k]$ .

Our Goal is to obtain the separation of the spectral envelope and spectral details such that,  $\log X[k] = \log H[k] + \log E[k]$ .

To achieve this separation, we take the FFT of the spectrum. An FFT of a spectrum referred to as Inverse FFT (IFFT). We are representing the spectrum in the log domain so as to simplify the process.

Now, the IFFT of the log spectrum can be represented as in the pseudo- frequency axis.

On this axis we consider two low and high frequency regions. And these spectrums are now represented as a peak lines on the axis giving a result of what we have desired.

So, summing up all,  $X[k] = H[k] E[k]$ .

$$|X[k]| = |H[k]| |E[k]|$$

Where,  $| \cdot |$  - denotes the magnitude of the expression.

Taking log on both sides, we get,

$$\log |X[k]| = \log (|H[k]| + |E[k]|)$$

Also, taking IFFT now, we get,  $x[k] = h[k] + e[k]$ .

For Mel-Frequency Analysis, Spectrum when implies Mel- Filters, we have Mel- Spectrum.

Now say,  $\log X[k] = \log (\text{Mel-Spectrum})$

We perform Cepstral analysis on  $\log X[k]$ , and obtain

$$x[k] = h[k] + e[k] \text{ after taking IFFT.}$$

Cepstral Coefficients  $h[k]$  calculated for Mel-Spectrum are referred to as Mel-Frequency Cepstral Coefficients often denoted by MFCC. [9]

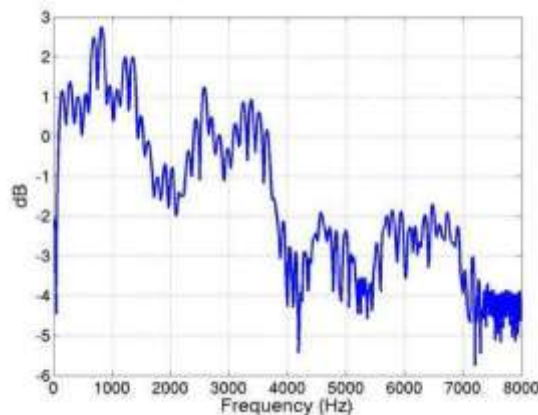


Figure 3: Mel-Filters (Filters in frequency region)

### Feature Matching

Feature matching is the technique of recognition like some of the popular methods – Dynamic Time Warping (DTW), Hidden Markov modeling (HMM), and Vector Quantization (VQ). Here, we're using the VQ method for matching purpose. As, we recall that a

database has been prepared for the need of comparison in order to completely recognize the speech. VQ is a process of mapping vectors from an expanded, large space of vectors to a finite number of regions in that space. This particular region is individually called as a “cluster” and can be represented by its center known as a “codeword”, and so, the collection of these codewords is called a “Codebook”. This region may also be called as a Voronoi region, and it is stated by:

$$V_i = [x \in R^k : \|x - y_i\| \leq \|x - y_j\|, \text{for all } j \neq i]$$

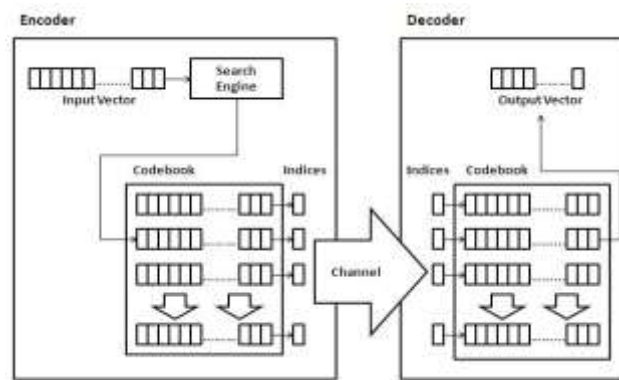


Fig 4: Schematic of Vector Quantizer (Encoder as in a PC and the Decoder as in the Microcontroller)

The size of the codebook is  $K$ , input vector which is of dimension  $L$ . In order to notify the decoder of which code vector is been selected, we use  $\lceil \log_2 K \rceil / L$  (each code vector will contain the reconstruction value of  $L$  source samples, the number of bits per sample.) i.e. 8 bits to represent 256 code vectors.

LBG (Linde-Buzo-Gray Algorithm) is a vector quantization algorithm used to derive a good codebook. The steps are as follows:

1. Determine the number of codewords i.e.  $N$ , or size of the codebook.
2. Select  $N$  codewords at random (from the set of input vectors), and let that be the initial codebook.
3. Apply the Euclidean distance formula to calculate the distance between the input vector in the cluster and each codeword.
4. Calculate new set of codewords by obtaining the average of each cluster.

$$y_i = \frac{1}{m} \sum_{j=1}^m x_{ij}$$

Where,  $i$  is the component of each vector (in  $x, y, z, \dots, n$  directions),  $m$  is the number of vectors in the cluster.

5. Repeat the steps 2 and 3 until one of the two happens – (a) codewords have not changed or (b) the change is them is infinitesimal. [8]

## ANDROID APPLICATION

The wireless-networking standard technology called Bluetooth has subtly become an innovative way to control a robot and a technology to replace the cables. Using an Android device to control a robot over Bluetooth is another step forward in remote robotics control by sending commands with the flick of a wrist. With an opened architecture and powerful proficiency, Android has become popular operating system among intense hobbyists able to build remote control applications with small development resources. They use smartphones or tablets that run Android OS and build applications feasible of developing remote controlled robots by sending some sort of signals wirelessly and at simple movements of the device or touching the screen. Based on the Java programming language, a built-in Bluetooth module, and a series of useful sensors already integrated and having permanent Internet connectivity, almost any Android device is categorized as a perfect tool for remote robotics control over Bluetooth. The idea of this paper is to use an Android application that allows you to communicate with a robot over the Bluetooth technology. The robot can respond to button, and swipes on the touch screen. In this way, you can control the robot to transport from one place to the other using commands forward, reverse, left and right.

## Bluetooth Technology

Every technology is bounded by some imperfections, and the Bluetooth technology is feasibly the best way for remote control as long as the robot is in the range of the Android device. The wireless communication is between multiple devices. One device runs the

Android OS, while the second device is the robot with a Bluetooth module. On the Android device, the control system is simple and uses an application to control the Bluetooth service on Serial Port Profile (SPP) connection. The application has to have error-free data transmission using Bluetooth module according to the sensors, actuators, UIs, touchscreen, and the traits of the application. On the robot side, you have to add a Bluetooth module connected to the robot controller. The Bluetooth module is a mini device designed for data transfer between peripheral devices. Moreover, we can say this mini device is able to synchronize the I/O data between the robot and the Android device.

### **Android OS**

Android is a mobile operating system (OS) based on the Linux kernel and currently developed by Google. With a user interface based on open architecture and having full independence over development, Android is designed primarily for touchscreen mobile devices such as smartphones and tablets. As of 2015, Android has the largest installed base of any mobile OS. It is a great platform for a robotic system control because it's much cheaper than any other ARM-based processing unit. Android platform is the widest used in the world and runs the largest number of smartphones worldwide. This is the reason why here we have used android as a platform to control the movements of the robot.

### **ACKNOWLEDGMENT**

The successful completion of our research work within the stipulated time frame is a result of collective efforts of our group as well as all the people who provided the continuous support. Here, we would like to thank all those people for their timely guidance. Our Head of Department for his encouragement and providing the required resources. We would also like to thank our Project Guide for her constant guidance without which our research paper was impossible. Also we would like to thank our college, MIT AOE, Alandi (D), Pune for providing us the platform to present our knowledge in terms of this research paper.

### **CONCLUSION**

This paper successfully explained the working of speech recognition using MFCC. It showed a unique feature extraction method for performing speech recognition. This speech based control had problems for recognizing due to noise and inadequate sound pitch level but, it is truly secure for controlling robots and is an excellent method in modern robotics and Speech processing.

It was also seen that android is a great platform to establish control over robots. It is also simple to use. The Bluetooth module helped to have a smooth connection between the robot and the smartphone. Information about the environment was sent to the Robot through RF module and the transmission was observed to be without glitches, error free and fast. Also, the collected data was stored and sent to the user mobile using Bluetooth module.

### **REFEREENCES:**

- Chadawan Ittichaicharon, Siwat, Thaweesak, "Speech recognition using MFCC", International Conference on Computer Graphics, Simulation and Modelling July 28-29, 2012/ Pattaya(Thailand).
- Sonam Kumari, Kavita Arya, Komal Saxena (GBTU), "Controlling of Device through Voice Recognition using Matlab", International Journal of Advanced Technology and Engineering Research(IJATER).
- Ahmed Q. Al – thahab,"Control of Mobile Robot using Speech Recognition", Journal of Babylon University, Pure and Applied Sciences, No.(3), Volume 19 : 2011.
- Nidhi Desai, Prof. Kinnal Dhameliya, Prof. Vijendra Desai, "Recognition Voice Command for Robot using MFCC and DTW", International Journal of Advanced Research in Computer and Communication Engineering, Volume 3: Issue 5 May 2014.
- Zaid El Omari, Samer Khamiseh, Lyad Abu Doush, Eslam Al Maghayreh, Yarmouk University, Jordan,"Using Mobile Phone to Control Movable Lego Robot Supported by Simple Robotic Arm", ICIT 2013, The 6<sup>th</sup> International Conference on Information Technology.
- Sujaya Bhattacharjee, C. Yashuwanth, "An Intelligent Agriculture Environment Monitoring System using Autonomous Mobile Robot", Information Technology, SRM University, Kattankulathur, India.
- Kim, "Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation", in Educational Technology&Society, vol.9, June 2006, pp. 322-334.
- Balwant A.Sonkamble,D.D.Doye,"Speech Recognition Using Vector Quantization through Modified K-mean LBG Algorithm",in Computer Engineering and Intelligent system,ISSN 2222-1719,Vol 3,No7,2012
- Mr.Kashyap Patel,Dr.R.K.Prasad,"Speech Recognition and Verification using MFCC ",International Journal Of Advanced Research in Computer science And Software Engineering,Vol 3, Issue 5, May 2013