Proc. of the Second Intl. Conf. on Advances in Electronics, Electrical and Computer Engineering -- EEC 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-6935-2

Wirelessly Controlled Voice Operated Robot

[Vasundhara Anand, Pranav Bhasin, Ankit Sharma, Meenakshi Sood]

Abstract— Speech is the most common and efficient mode of communication and has large potential of being a mode of interaction with machine/computer. To communicate with the machine via speech we require an interface which replicates the human auditory system. An Automatic Speech Recognition (ASR) system is required which is able to "hear," understand," and "act upon" spoken information. This paper focuses on developing the speech recognition module and integrating it with robotics wirelessly. Our goal is to develop an interactive robot which performs actions in accordance with the voice commands communicated wirelessly to the robot after being processed in the ASR module which is speaker dependent and has a pre-defined command set.

The system would support many valuable applications like replacing a human assistant with a robot which can perform actions according to voice command in more efficient manner.

Keywords—Speech Recognition, Feature Extraction, MFCC, Robotics, Xbee, Zigbee.

I. Introduction

Automatic Speech Recognition (ASR) is a technology that allows a computer to identify the words that a user speaks. It allows a computer to recognize all words that are intelligibly spoken by any person, in real-time and noisy environment. [1]. For the system to identify and process the input voice command from the user, short time feature vectors are obtained from the input speech data. The extracted feature vectors are required to have distinct characteristics that can best represent the speech data and of the proper size that can be processed efficiently. In real-time, these characteristics are compared with a pre recorded pattern and the best match is judged as the given command. Different features like mean, pitch, Cepstral coefficients can be extracted using different techniques like correlation, covariance, MFCC[2] and further recognized using DTW (Dynamic Time Wrapping)[3], LPC(Linear Predictive Coding)[4], HMM(Hidden Markov Model) [5], neural network[6], thus implementing the complete ASR system.

Vasundhara Anand (Student), Pranav Bhasin (Student), Ankit Sharma (Student)
JAYPEE University Of Information Technology, Solan
Pin: 173215 (H.P.) India

Mrs. Meenakshi Sood (Faculty)
JAYPEE University Of Information Technology, Solan
Pin: 173215 (H.P.) India

The recognition and thereafter processing of the command is done at the processing unit. The optimal signals are communicated to the robot after taking the appropriate decision. For integrating ASR with robotics, the actions that can be performed by the robot are identified and accordingly the voice command set is defined. A database of voice commands is maintained, which is used to recognize the real-time input provided by the user. The communication between the robot and the processing unit can be wired or wireless. The latest wireless technologies as Wi-Fi, Cellular Network, XBee, RF Antenna's can be chosen.

The paper is organized as follows: Section II deals with speech recognition module including MFCC Technique. Section III describes the Robotic Module comprising of the wireless and the microcontroller sub-modules. In Section IV the proposed system design and features are explained with the achieved results.

II. Speech Recognition

The voice command uttered by the user after being converted into the digital format cannot be efficiently analyzed directly; moreover any two different commands given by the same speaker cannot be differentiated effectively. Analytically, the same command repeated again and again by the same user has high variations, which further increases the computational complexity. So, it is required to extract certain features from the uttered voice which help to identify distinguishing characteristics between two different commands and simultaneously remain similar for the same command.

A. *MFCC Feature Extraction Technique*

MFCC (Mel Frequency Cepstral Coefficient) is a technique that helps to identify the features which can be used to distinguish different commands. In Mel-frequency Cepstrum, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal spectrum. This system found to be more accurate under low varying environment but fails to recognize speech under highly varying environment. [7]

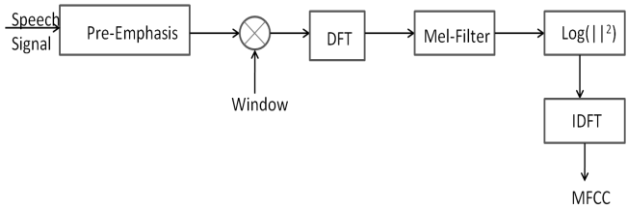

Fig.1. MFCC block diagram

Fig.1 [4] depicts the process followed to obtain the MFCC coefficients. The words uttered by the user contain peaks produced by the resonance of the human vocal tract.

SEEK DIGITAL LIBRARY

Proc. of the Second Intl. Conf. on Advances in Electronics, Electrical and Computer Engineering -- EEC 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-6935-2

The frequencies at which these peaks are produced are referred to as formants. The High frequency formants although possessing relevant information have smaller amplitude with respect to low frequency formants .A Pre-Emphasis of high frequencies is therefore required to obtain similar amplitude for all formants. Such processing is usually obtained by filtering the speech signal with a first order filter.

The voice signal thus achieved, is highly varying signal, so the traditional methods for spectral evaluation are not reliable in its processing. For voice, the statistical characteristics are less variant within the short time interval, during which a short time analysis can be performed by "windowing" a signal x'(n) into a succession of windowed sequences $x_t(n)$ t = 1,2,... ,T, called frames, which are individually processed.

The frames are passed through the chosen window, and the obtained signal is analyzed in Frequency Domain using Fast Fourier Transform algorithm (FFT) algorithm. The phase information of the DFT samples of each frame is discarded while estimating the characteristics of the vocal tract. This is consistent with the fact that phase does not carry useful information. Perceptual experiments have proven that the perception of the signal reconstructed with the random phases is almost indistinguishable from the original, if the phase continuity between successive frames is preserved [7]. The spectral analysis reveals that speech signal features are mainly due to the shape of the vocal tract.

The signal obtained is passed through the Mel Frequency Cepstrum Computation Filter which uses the frequency response of the non-uniformly spaced triangular bandpass filters that plays the role of smoothing the spectrum, performing a processing that is similar to that executed by the human ear. The filters which are linearly spaced on the Mel scale, when converted into frequency domain results in uniformly spaced filters before 1 kHz and logarithmically spaced after 1 kHz as depicted in the Fig. 2 [8].

After passing the signal through the filter the logarithm of the square magnitude of the coefficients is calculated. The magnitude discards the useless phase information while a logarithm performs a dynamic compression, making feature extraction less sensitive to dynamic variations.

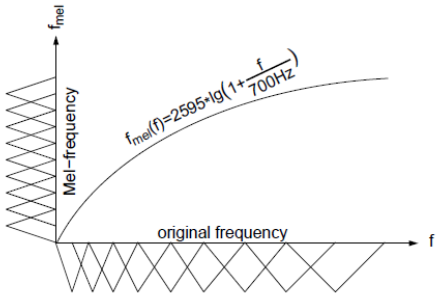

Fig.2. Original frequency to Mel-frequency conversion plot [8].

Finally, the inverse DFT of the logarithm of the filter bank output is calculated to achieve the Mel frequency Cepstrum coefficients (MFCC). These coefficients are features that are used for comparison between different voice commands.

III. Robotic Module

A. *Wireless Module Using XBee*

To develop wireless end-point connectivity to embedded devices, RF technology based Xbee can be used. XBee works in 2.4GHz ISM band [9] and can be directly connected to the serial port (at 3.3V level) of the microcontroller. It works on the Zigbee protocol which is based on the IEEE 802.15.4 networking protocol providing fast point to point, point-to-multipoint or peer-to-peer networking. This module supports data rates of up to 115kbps. XBee can be used in transparent mode, where the modules act as a serial line replacement.

B. *Robot Using ATMega16 (L) Microcontroller*

ATMega16 (L) microcontroller is used to control the task performed by the robot. To enable to communicate and send instructions to the microcontroller, USART (Universal synchronous and Asynchronous serial receiver and Transmitter), a serial communication device is inbuilt in ATMega 16(L) series [10]. The USART has to be initialized before starting communication

The robot DC motors are interfaced with the microcontroller using a H-bridge IC, L293D. It converts the input logic to power supply. By configuring the microcontroller, we can control the logic input to motor driver IC, which controls the directions and speed of the DC motors.

IV. Proposed System Design

To integrate the speech recognition with robotics we define an outline of the system that can be used as a generic approach. The speech recognition module is available on the base station, which receives and interprets voice command .The robot performs the action in accordance with the given instructions, transmitted wirelessly by the processing base station. We have used four motors for movement of the robot. The complete system broadly has two modules--The Speech recognition Module and the Robotic Module.

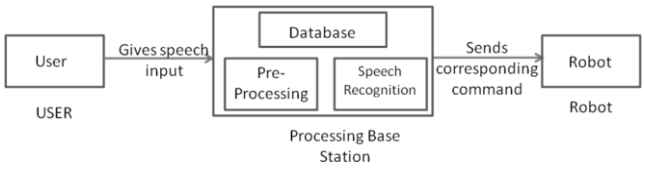

Fig.3. Speech recognition system

A. *Speech Recognition Module*

The first order filter has been used that has pre-emphasis parameter ‘a’=0.95, which gives rise to more than 20 dB amplification of the high frequency spectrum. The Fig.4 shows the plot for ‘forward’ command before pre-emphasis and after the pre-emphasis in time domain.

SEEK DIGITAL LIBRARY

Proc. of the Second Intl. Conf. on Advances in Electronics, Electrical and Computer Engineering -- EEC 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-6935-2

The speech signal is divided into frames of size 256 samples and overlapping of 101 samples is done between two consecutive frames. These frames are then passed through the Hamming window of length 256. For frequency analysis 256 point FFT is calculated as shown in Fig.5 for 'forward' command. To calculate Cepstral coefficients, 22 triangular filters placed linearly on the Mel-scale are implemented and mapped on to the frequency scale using the relation:

$$f = 700(e^{m/1127} - 1) \quad (1)$$

Where m=the frequency in Mel-Scale.
Fig.6 shows Mel filters where each filter is linearly spaced below 1kHz and non linearly spaced above 1khz.

The logarithm of the signal is taken after passing through the filters. The Direct Cosine Transform (DCT) is then taken to achieve the MFCC for the commands [5]. The MFCC achieved for "Forward" and "Stop" command are shown in the Fig.7 and Fig.8 respectively.

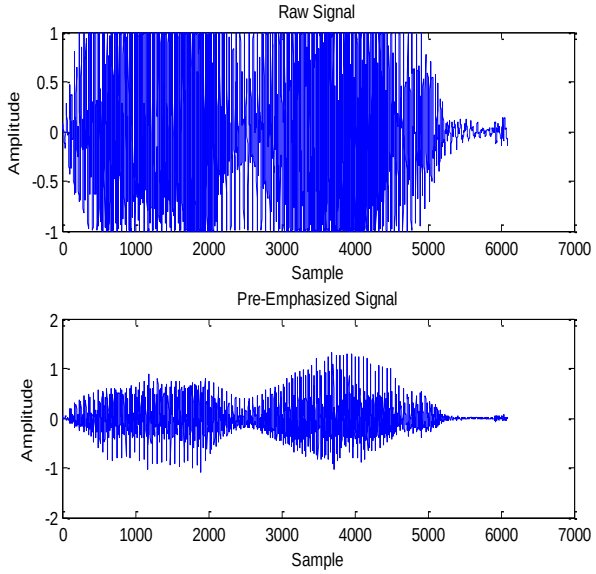

Fig.4. (a) Raw Signal (b) Pre-Emphasized Signal

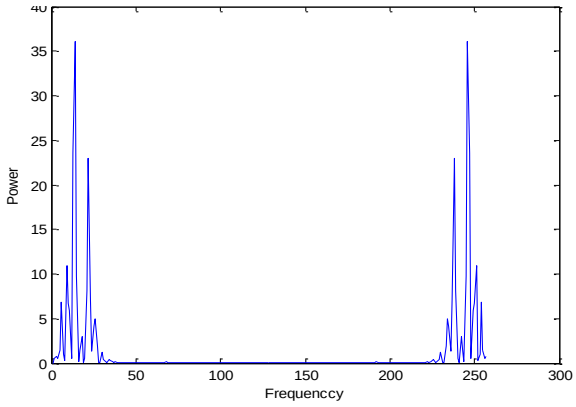

Fig.5. Power spectrum

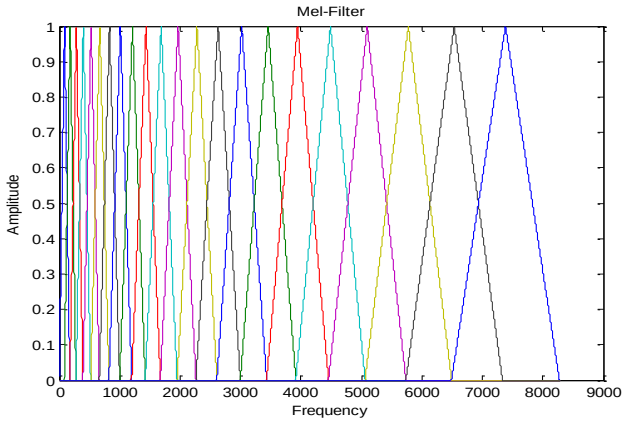

Fig.6. Mel scale filter bank

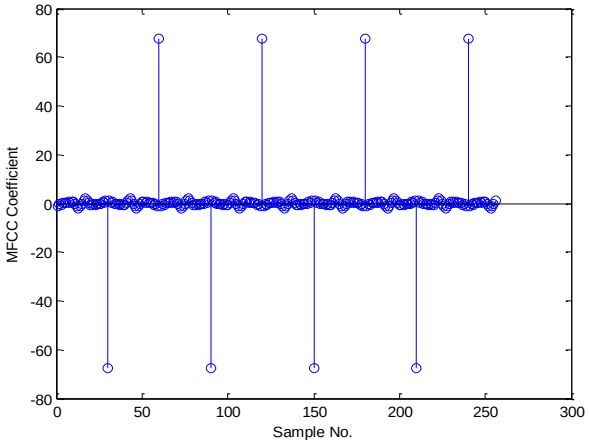

Fig.7. MFCC coefficients of 'Forward' command

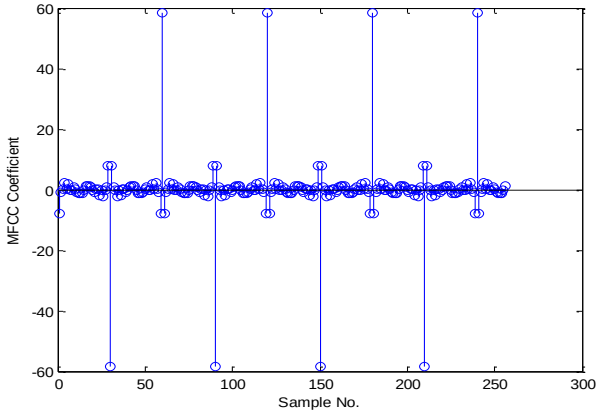

Fig.8. MFCC coefficients of 'Stop' command

When a real-time input is given to the system, the system identifies the command. MFCC coefficients of the real-time input are compared with the MFCC's of the commands stored in the database. This is done by taking the difference of individual MFCC's of real-time commands with database and recording the maximum difference with each command. The command with minimum difference is identified as the command uttered.

With the proposed system we have used two commands uttered by one user and were able to achieve 80% accuracy through this.

SEEK DIGITAL LIBRARY

Proc. of the Second Intl. Conf. on Advances in Electronics, Electrical and Computer Engineering -- EEC 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-6935-2

B. *Robotic Module*

XBee Series 2 has been used for the wireless communication which has a chip antenna and wire antenna, providing a range of 100 meters for indoor environment .One Xbee is made as Coordinator in AT(transparent) mode and the other as router/end device AT. Xbee is used to establish a network and add more end devices in the network if required.

The Microcontroller USART is initialized in Asynchronous mode with 9600 bps baud rate by setting UMSEL, U2X, UBRRL and UBRRH bits. The frame format setting is done by setting parity bit as none, one bit as stop bit and character size as 8 bits with RX pin enabled.

References

[1] Docsoft Inc.Whitepaper – "What is Automatic Speech Recognition," June 2009.

[2] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique," International Journal of Computer Applications (0975 – 8887), Vol.10, no.3, November 2010.

[3] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi , "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," Journal Of Computing, Vol. 2, Issue 3, March 2010.

[4] Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go, and Chungyong Lee, "Optimizing Feature Extraction for Speech Recognition," IEEE Trans. On Speech And Audio Processing, Vol. 11, no. 1, January 2003.

[5] Ashish Jain,Hohn Harris, "Speaker identification using MFCC and HMM based techniques," University Of Florida, April 25,2004.

[6] N.Uma, Maheswari,A.P.Kabilan, R.Venkatesh, "Speaker independent speech recognition system based on phoneme identification," Proceedings of the 2008 International Conference on Computing, Communication and Networking (ICCCN 2008).

[7] Claudio Becchetti and Lucio Prina Ricotti, "Speech Recognition Theory and C++ implementation" (Book).

[8] Sirko Molau, Michael Pitz, Ralf Schl¨uter, and Hermann Ney "Computing Mel-Frequency Cepstral Coefficients On The Power Spectrum," Lehrstuhl f¨ur Informatik VI, Computer Science Department, RWTH Aachen – University of Technology, 52056 Aachen, Germany.

[9] XBEE Datasheet-Digi.

[10] ATMega16L Datasheet-Atmel.

About Author (s):

Ms. Vasundhara Anand currently doing B.Tech 4th year from ECE department, acquired CGPA 9.5 till 7th semester. JAYPEE University Of Information Technology, Solan.

Mr. Pranav Bhasin currently doing B.Tech 4th year from ECE department, acquired CGPA 8.2 till 7th semester. JAYPEE University Of Information Technology, Solan.

Mr. Ankit Sharma currently doing B.Tech 4th year from ECE department, acquired CGPA 6.9 till 7th semester. JAYPEE University Of Information Technology, Solan.

Mrs. Meenakshi Sood is the Sr. Lecturer of department of ECE, JAYPEE University Of Information Technology, Solan, Himachal Pradesh, India. She has more than 12 years of teaching experience. Pursuing Ph.D in Biomedical Signal Processing. She is Gold Medalist and has been awarded Academic Award for her performance in Master of Engineering (Hons.) from Panjab Univeristy, Chandigarh

SEEK DIGITAL LIBRARY