INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 3)

Available online at: www.ijariit.com

MFCC feature extraction for speech recognition with hybrid application

Megha Ganeshrao Kadam
meghak2290@gmail.com
Rajarshi Shahu College of Engineering, Pimpri-Chinchwad, Maharashtra

Sakshi A. Paithane
sakshi.paithane@gmail.com
Rajarshi Shahu College of Engineering, Pimpri-Chinchwad, Maharashtra

ABSTRACT

To analyze and detect human voice in different applications such as military area, medical field, and telecommunication for assigning tasks according to it. In this Human voice recognized using MFCC features with a network in such a way that it recognizes only specific person speech commands with exit the program for another one. This paper represents with a wide range of feature extraction algorithm available, MFCC is a leading approach for speech feature extraction and our current research aims to apply it on real-time hybrid based applications i.e. home automation and robotic application. The ANN has been trained for commands LIGHTS ON, LIGHTS OFF, FAN ON and FAN OFF for home automation as well as LEFT, RIGHT, FORWARD, BACK, STOP for our robotics application. ThingSpeak IOT cloud has been used as a server to send/receive commands between two clients, the PC/laptop from where the speech command is sent and the Raspberry Pi where the command will be used to control the robot and relays for home automation. The best part of the proposed system is that controlling the devices is independent of the location of the speaker. The result shows the proposed method has achieved an accuracy of 96.64% for robotic application and 94.63% for home automation speech commands.

Keywords: *Raspberry Pi 3, Thingspeak cloud server, DC motor driver, Artificial neural network, MFCC, Speech recognition.*

1. INTRODUCTION

The Speech is important to communicate with Humans, it is the Ability to express thoughts, information, and feelings by articulate sounds. In computer-based speech recognition system, a computer simply attempts to transmit the speech into the textual representation, rather than understanding it. Speech is the best way to train a machine or to communicate with a machine. In this work, we can be used as a biometric feature technology for verifying the identity of a person like banking by telephone and voice mail. Speaker recognition is classified into two parts speaker verification and speaker identification. The automatic speaker recognition is used to extract, characterize and recognize the information which is used to identity speaker with identifying a speaker by his or her voice. Speaker Recognition approaches can be divided into two approaches: text-dependent and text-independent approaches. Hidden Markov models are based on recognition system is effective under specific circumstances, but there are a major limitation in it like applicability of ASR technology in real-world environments. Artificial Neural Networks (ANN) and specifically Multi-Layer Perception's (MLP) appeared to be an effective alternative to replace or help HMM in the classification mode.

Speech Recognition system is used as most useful in the field of Robotics to make the robot capable to follow the different commands through voice only. Because voice is easily available and cheapest biometric tool, in this robot, could follow the all voice commands in all language that language will choose upon by commands which uttered by a human. Also, we can analyze a human-human interaction or human-robot interaction.

Firstly we used microphone for capturing the signal then Algorithms are applied on it to reduce noise interference with silence suppression. To make the signal free from above interference we used MFCC feature extraction technique which processed to extract the features. MFCC used as an input to ANN systems and results are obtained for speech and speaker recognition. The parameters such as performance, training state & validation are evolved from this ANN tool. The MFCC extracted features are given to ANN. It is transferred to raspberry pi through cloud that means computer to raspberry pi. All these peripheral controlled by it. Rasberry Pi is connected with bulb ,fan and dc motors and implemented as a home automation and robotic applications.

2. METHODOLOGY

Block Diagram:

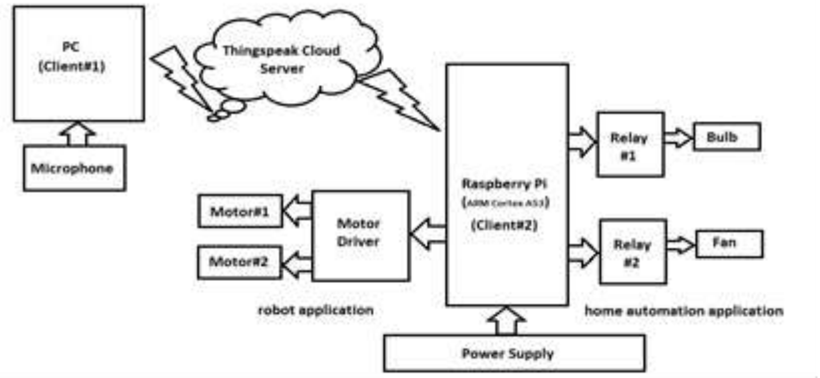

Fig 2.1 Block Diagram

2.1.1 Microphone:

Microphone firstly used early with telephones after that radio transmitters. In 1876, Emile Berliner invented the first microphone used as a telephone voice transmitter. Dynamic mics use Magnetics, are less expensive and more rugged while condenser microphones are powered by electricity and are smaller and more sensitive. A loudspeaker is one type of transducer which converts an electrical signal into sound waves, which is opposite in a Microphone function.

The microphone is the type of a transducer which converts one form of energy into another like convert acoustical energy into electrical energy. All type of microphone having different paths to convert energy but they share one thing i.e The Diaphragm is common in it.

2.1.2 PC (Client 1):

Client 1 is having the entire software platform to execute the electrical voice signal as input and commands as output.

2.1.3 Python Idle 2.7:

IDLE is Python's Integrated Development and Learning Environment.

IDLE has the following features:

● coded in 100% pure Python, using the tkinter GUI toolkit

● cross-platform: works mostly the same on Windows, Unix, and Mac OS X

● Python shell window (interactive interpreter) with colorizing of code input, output, and error messages

● multi-window text editor with multiple undo, Python colorizing, smart indent, call tips, auto-completion, and other features

● search within any window, replace within editor windows and search through multiple files (grep)

2.1.4 Mel-Frequency Cepstral Coefficients:

Mel-frequency cepstral coefficients (MFCCs) are coefficients that bunched together as an MFC. A cepstral representation of the audio clip derived from it. MFC is the main difference between the cepstrum and the mel-freq cepstrum, in that mel scale has equally spaced frequency bands, in which frequency bands are approximated the human auditory systems response closer than the linearly-spaced used in the normal cepstrum. To allow better representation of sound this frequency warping used like in audio compression.

MFCCs are calculated by following methods:

● Take the Fourier transform 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.

2.1.5 Artificial Neural Network:

It is a collection of processing elements, units or Nodes, which are connected to each other and organized in layers. The Fig. 2.1.5 shows the structure of An Artificial Neural Network. The network has the processing ability to store the inter-unit connections or weights which are used to learn the process.

Learning process is a set of training patterns given to the network and to minimize the error between outputs of net and true target values we sets the weights according to requirement. This algorithm of the weights is called as Back-propagation.

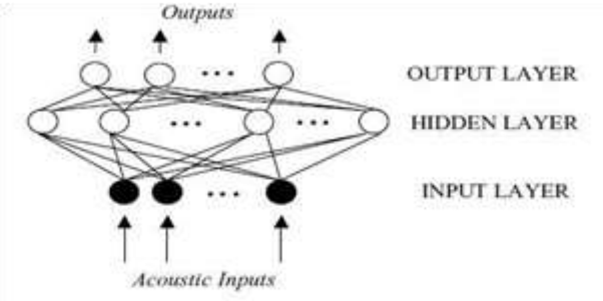

Fig 2.1.5 Artificial Neural Network

2.1.6 ThingSpeak Cloud Server:

ThingSpeak is an open source Internet of Things (IoT) application and API to store and retrieve data from things using the HTTP protocol over the Internet or via a Local Area Network. ThingSpeak enables the creation of sensor logging applications, location tracking applications, and a social network of things with status updates. ThingSpeak was introduced in 2010 by ioBridge to support IoT applications. MATLAB software gave support to ThingSpeak which is used for numerical computing and It also analyze, visualize uploaded data using Matlab without license from Mathworks.

2.1.7 Raspberry Pi (Client 2):

Raspberry Pi 3 has many features i.e A 1.2GHz 64-bit quad-core ARM Cortex-A53 CPU Integrated 802.11n wireless LAN and Bluetooth 4.1. Raspberry Pi has the Raspbian operating system which is a version of Linux. In Raspberry Pi an SD card inserted into slot on the board which acts as hard drive.It used in a traditional RCA TV set, a many modern monitor or even a TV using the HDMI port.

2.1.8 Motor Driver:

It is Integrated Circuit utilize for two purpose. First purpose to provide required current for DC motor is high of 1 Ampere but the raspberry output is of 30-40 mili ampere. So to boosting the current it is used. The next use is to provide adequate voltage to motor i.e from raspberry output is of 3.3 V and motor needs of 12 volts. This is the function of Motor Driver IC. Here L293D is used.

2.1.9 DC Motor:

The DC motor is a machine that transforms electric energy into mechanical energy in the form of rotation. DC motors have inductors inside, which produce the magnetic field used to generate movement.

A motor is an electrical machine which converts electrical energy into mechanical energy. The principle of working of a DC motor is that "whenever a current carrying conductor is placed in a magnetic field, it experiences a mechanical force". The direction of this force is given by Fleming’s left hand rule.and its magnitude is given by F=BIL.

3. PROPOSED ALGORITHM

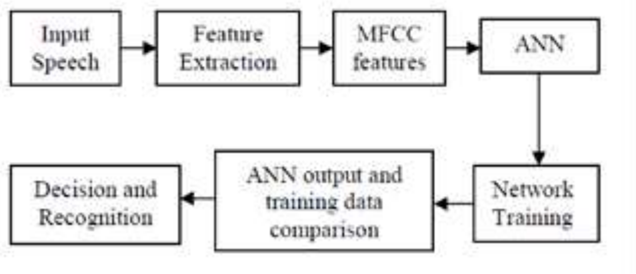

Fig 3. Process flow of recognition phase

In this algorithm, the voice is taken as input the feature extraction will perform several mathematical is carried out to get MFCC features. MFCC features are extracted from each recorded voice. Hence acoustic voice signal is converted to a set of numerical values. After that, training data-table is created using MFCC feature and target data-table also created as back-propagation neural network was used. Built-in Artificial Neural Network (ANN) is trained with these. Finally, when the test data is given, the ANN compares the test data with train data-table. If speaker is recognized user, it recognizes him and the specific command by finding the maximum possible matches within a given tolerance level. Otherwise, it terminates program.

3.1.1 Speech Recognition using ANN:

After preprocessing, the next important step is to recognize the speech using Artificial Neural Networks. In this we propose a Multilayer Mapping Network. The processing ability of the network is stored in the inter-unit connections, or weights, which are tuned in the learning process. In the learning process, a set of training patterns is presented to the network, and the weights are adjusted to minimize the error between the outputs of the net and the true target values. This update algorithm of the weights is called back-propagation. The advantage of this model is that is its flexibility & expandability of hidden layers for recognition.

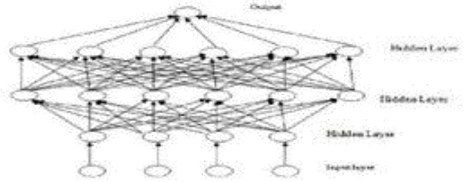

Fig. 3.1.1 Multilayer Pattern Mapping Network

3.2 MEL FREQUENCY CEPSTRAL Coefficients (MFCC):

To design a text-dependent identification of speaker in the features of MFCC used .The MFCC extracted features of a speaker are used vector quantization algorithm to quantize it into number of centroids. All Centroids made codebook of that speaker. MFCC's features are derived in two phases' i.e training phase and testing phase. The speakers vocalizes the repeated words once in a training session as well as in testing session later.In training phase of each speaker to the centroids of each speaker in the testing phase is derived and identified by using minimum Euclidean distance. The acoustic or noisy signals is important part to design any speech recognition system using extraction and selection of representation of good parameters. It affects the performance of speech recognition system. The MFCCs has been proved more efficient to speech recognition. Following are the steps to calculate the MFCC.

3.2.1 MEL-FREQUENCY SCALE:

The sound of frequency for speech is not follow a linear scale in the perception of human so each tone of actual frequency f, measured in Hz, & a subjective pitch measured on a mel-scale. It has a linear & logarithmic spacing below 1000 Hz and above 1000 Hz respectively.

$$\mathbf{Mel(\mathit{f}) = 2595*log_{10}(1 + \mathit{f}/700)}$$

A filter bank approach is used to simulate the subjective type of spectrum in it, for each required mel-freq component as one filter. Each filter bank in frequency response is having triangular bandpass in shape also spacing and bandwidth is evaluated by a constant mel-frequency interval.

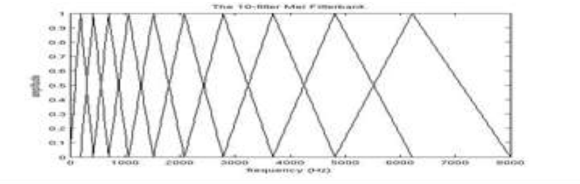

Fig 3.2.1 The ten-filter Mel filter bank

The Mel scale filter bank is a series of l triangular band pass filters that have been designed to simulate the band pass filtering believed to occur in the auditory system. Similarly 10 filter Mel filter bank is shown in Fig 3.2.1.

3.2.2. CEPSTRUM:

The result of log mel spectrum which is converted back into time called the Mel Frequency Cepstrum Coefficients (MFCC). A good representation of Frame analysis for local spectral properties of the signal by the cepstral representation of speech spectrum. We can convert them into time domain using discrete cosine transform due to MFC are real numbers. At the end, log mel-spectrum converted back into time this result is called the Mel Freq Cepstrum Coefficients.

4. RESULTS

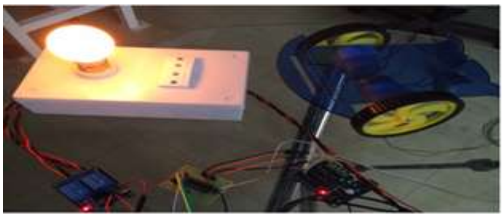

Fig 4. Experimental Setup

In this, on hardware part, the input is a voice from the microphone is fed to PC in terms of the voice signal. Role of PC is to provide output according to program the commands will be sent to the raspberry pi.

● The output of raspberry pi is high then the intermediate relay module will have an output of low. Hence bulb will be OFF.

● The output of raspberry pi is low then the intermediate relay module will have an output of high. Hence bulb will be ON.

For robot application, here DC motor is used. So the motor driver is required due following two reasons.

● Raspberry pi output gives max 30 to 40 mA of current but motor needs 1Amp of current.

● Raspberry pi output gives max 3.3 V of voltage but motor needs 12 V of voltage.

4.1 RESULT TABLE:

Table 4.1 Speech recognition result table

Speech	Actual Result	Test Result/Command Sent to Cloud
Lights on	The command is “2222”	“2222”
Lights off	The command is “4444”	“4444”
Fan on	The command is “6666”	“6666”
Fan off	The command is “8888”	“8888”
forward	The command is “1111”	“1111”
Back	The command is “3333”	“3333”
Left	The command is “5555”	“5555”
Right	The command is “7777”	“7777”
Stop	The command is “9999”	“9999”

5. CONCLUSION

Training and testing is carried out in planned manner; it has more benefits and is therefore worth considering. This is achieved by implementing MFCC with ANN. Testing and training is much faster and accurate.In this project raspberry pi is used to construct home automation and robot application to recognize spoken word. Also to access in the same Internet of Thing we used.

6. REFERENCES

[1] Pialy Barua,Kanij Ahmad,Ainul Anam Shahjamal Khan,Muhammad Sanaullah, “Neural Network-Based Recognition of Speech Using MFCC Features” 3rd International Conference on Informatics, Electronics & Vision ,IEEE ,2014.
[2] Shweta Tripathy, Neha Baranwal, G.C.Nandi, “A MFCC based Hindi Speech Recognition Technique using HTK Toolkit” Second International Conference on Image Information Processing (ICIIP-2013), IEEE, Proceedings of the 2013.pp 539-544.
[3] Arnab Pramanik, Rajorshee Raha, “Automatic Speech Recognition using Correlation Analysis” World Congress on Information and Communication Technologies, IEEE 2012, pp.670-674.
[4] Chin Kim On, Paulraj M. Pandiyan, Sazali Yaacob, Azali Saudi, “Mel-Frequency Cepstral Coefficient Analysis in Speech Recognition” ICOCI, IEEE 2006.
[5] Bacha Rehman, Zahid Halim, Ghulam Abbas, Tufail Muhammad, “Artificial Neural Network-based Speech Recognition using DWT analysis applied on isolated words from oriental languages” Malaysian Journal of Computer Science. Vol.28, 2015 pp 242-262.
[6] Umarani J. Suryawanshi, Prof. Dr. S. R. Ganorkar,“Hardware Implementation of Speech Recognition Using MFCC and Euclidean Distance” IJAREEIE, Vol. 3, Issue 8, August 2014, pp 11248-11254.
[7] Safdar Tanweer, Abdul Mobin, Afshar Alam, “Analysis of Combined Use of NN and MFCC for Speech Recognition” World Academy of Science, Engineering and Technology, IJCEACIE, Vol:8, No:9, 2014 pp. 1736-1739.
[8] Divyesh S.Mistry, Prof.A.V.Kulkarni, “Overview: Speech Recognition Technology, Mel-frequency Cepstral Coefficients (MFCC) , Artificial Neural Network (ANN)” IJERT, Vol. 2 Issue 10, October – 2013 pp.1994-2002.
[9] Sabeur Masmoudi, Mondher Frikha, Mohamed Chtourou, Ahmed Ben Hamida. “Efficient MLP Constructive training algorithm using a neuron recruiting approach for Isolated Word Recognition system” International Journal Speech technology,Springer,March2011, Volume14, Issue 1, pp 1–10.