SMART CAR BY MEANS OF IoT

[1*]Nitesh P. Patel, [2]Nilesh P. Bhosle
[1]Assistant Professor, [2]Associate Professor
[1]Electronics and Telecommunication, D. Y. Patil College of Engineering, Pune, India
(*nitesh.patel@dyptc.edu.in)

Abstract: In this paper, the voice-based technology in a smart carriage with IoT is applied. The use of speech input for controlling in-vehicle tasks, examine a number of automatic system parameters, input modalities, and driver ages to evaluate the effects these variables had on driving performance, task-function usability, and driver preference/acceptance of task-function design, to develop a comprehensive set of human factors guidelines and recommendations for the use of the smart car in automotive applications. The literature review suggests that implementing speech recognition inside the automobile may improve dual-task performance and reduce the mental workload when certain types of in-vehicle tasks are performed concurrently while driving. Therefore, concurrent in-vehicle tasks that have auditory-stimulus/verbal-response characteristics may be performed using speech input with the primary driving task using IoT.The problem is not the availability of driverless cars. Therefore, voice recognition technology using the MFCC system into an artificial neural network (ANN) design parameters had a moderate impact on the usability of in-vehicle tasks using clouds. In-vehicle system designs that afford the highest level of usability must be selected for use in the automotive environment using Raspberry Pi.

Index Terms - MFCC, Raspberry pi, Voice Recognition Technology, Internet of Things (IoT).

1 .INTRODUCTION

The 2011 census showed that 1.2% of the total Indian population was physically challenged people in especially urban areas, their primary problem is transportation for their day to day life, even though there are many companies like Mercedes, BMW is developing a solution for the badly-behaved using voice recognition consequently to afford broad mechanization, so building an efficient voice recognition system for automated vehicle driving. In attendance consume numerous schemes projected through the detection of voice recognition from the early 2000s for countless submissions. Voice control system using the DSP IC mechanism can grip 256 words and it customs radio waves transmission so the SNR scheming may postponement the production or the signal may even become last, voice control systems with local speech engine require a lot of deadly works for Acoustic Model Building, as every speech recognition undoubtedly ought to not use more capitals nearby and it should have earlier reaction time. Our paper tells the usage of Google Speech API for speech to text conversion which procedures DNN-HMM for the Acoustic model as it can grip any words up to the measurement of 100 fonts and it likewise does not require user countryside in the range of speech. [1]

The advent of intelligent transportation systems (ITS) has brought new technologies inside the automotive cockpit. Drivers will be able to access greater information than is provided by current instrument panel displays and controls. Navigation, route guidance, traffic management information, collision avoidance, communication systems, and alternative methods for displaying and controlling vehicle information (speed, audio, climate control, engine status, and warning telltales) are just several examples of the new technologies proposed to improve driving performance, comfort, and convenience. Driver interaction with these systems may be different from conventional interface formats. More Information will be displayed to the driver, possibly requiring more glances inside the vehicle and for longer durations to assess or control a given system. New systems will compete for space inside the vehicle, forcing displays and controls either to become smaller or to be reconfigurable. Driver interface design will be critical for both user acceptance and performance.

Automatic speech recognition (ASR) technology has been implemented in concept vehicles to demonstrate a new method of interfacing the driver with vehicle control systems. Methods for selecting which in-vehicle systems are to be controlled through speech input are not well defined. Often, systems are arbitrarily selected for speech control due to ease of implementing the speech recognition technology, or the project engineer and/or platform customer believes this is what users should control with speech.

A detailed review of the existing scientific literature on human factors in speech recognition and speech recognition technology.

A decision tree analysis to determine which current and near-future secondary task functions are amenable to speech recognition technology.

An empirical evaluation of secondary-task functions that are determined both amenable and not amenable to speech recognition technology and the manual analogs of these secondary-task functions.

Development of design guidelines to be used by designers and engineers in determining which automotive control systems should incorporate speech

2 .LITERATURE SURVEY

After analyzing the requirements of the task to be performed, the next step is to analyze the problem and understanding its context. The first activity in the phase is studying the existing system and the other is to understand the requirements and domain of the new system. Both the

activities are similarly significant, but the primary action helps as a source of giving the purposeful conditions and then the positive design of the proposed system. Accepting the belongings and supplies of a new classification is additionally tough and involves inspired intellectual and accepting of the current consecutively system is too problematic, improper considerate of the current system can lead to change from clarification. The proposed design difficult the next investigation paper analysis.

Analysis and detection of human voice at workplace such as telecommunications, military scenarios, medical scenarios, and law enforcement is important in assessing the ability of the worker and assigning tasks accordingly. This paper represents the results from a preliminary study to recognize the speech from human voice using Mel-Frequency Cepstrum Coefficients (MFCC) features. The 16 Mel-scale warped Cepstral coefficients were used independently for reorganization of speech from two Bangla commands of our native language. Cepstral Coefficients for the utterance of 'BATI JALAO' (i.e., TURN ON LIGHT) and 'PAKHA BONDHO KORO' (i.e., TURN OFF FAN) from a particular speaker under preliminary investigation were used as features in a neural network. [2]

The Speech recognition is a major topic in speech signal processing. Speech recognition is considered as one of the most popular and reliable biometric technologies used in automatic personal identification system s. Speech recognition systems are used for variety of applications such as multimedia browsing tool, access center, security and finance. It allows people work in active environment to use computer. For a reliable and high accuracy of speech recognition, simple and efficient representation methods are required. In this paper, the zero crossing extraction and the energy level detection are applied to the recorded speech signal for voiced/unvoiced area detection. The detected voiced signals are applied for segmentation. Further, the MFCC method is applied to all of the segmented windows. The extracted MFCC data are further used as inputs for neural network training. [3]

The growth in wireless communication and mobile devices has supported the development of Speech recognition systems. So for any speech recognition system feature extraction and patter matching are two very significant terms. In this paper we have developed a simple algorithm for matching the patterns to recognize speech. We used Mel frequency cepstral coefficients (MFCCs) as the feature of the recorded speech. This algorithm is implemented simply by using the principle of correlation. All the simulation experiments were carried out using MATLAB where the method produced relatively good results. This paper gives a details introduction of recorded speech processing, design considerations and evaluation results. [4]

To utilize the robot's capabilities, it is necessary for us to communicate with them efficiently. Thus, Human Robot Interaction is attracting the attention of most of the researchers these days. In this paper a speech recognition system has been developed using different feature extraction techniques like MFCC (Mel Frequency Cepestral Coefficient), LPC (linear predictive coding) and HMM (hidden Markova Model) is used as the classifier. Less work has been done for Hindi language in this field with a vocabulary size not very large. So, work in this paper has been done for Hindi database, with a vocabulary size a bit extended. HMM has been implemented using HTK Toolkit. Afterwards the performances of both of the techniques used have been compared. The work has been done using audacity for sound recordings and Cygwin to execute the HTK commands in Linux type environment in windows platform. [5]

Speech recognition is an emerging research area having its focus on human computer interactions (HCI) and expert systems. Analyzing speech signals are often tricky for processing, due to the non-stationary nature of audio signals. The work in this paper presents a system for speaker independent speech recognition, which is tested on isolated words from three oriental languages, i.e., Urdu, Persian, and Pashto. The proposed approach combines discrete wavelet transform (DWT) and feed-forward artificial neural network (FFANN) for the purpose of speech recognition. DWT is used for feature extraction and the FFANN is utilized for the classification purpose. The task of isolated word recognition is accomplished with speech signal capturing, creating a code bank of speech samples, and then by applying pre-processing techniques. For classifying a wave sample, four layered FFANN model is used with resilient back-propagation (Rprop). The proposed system yields high accuracy for two and five classes. For db-8 level-5 DWT filter 98.40%, 95.73%, and 95.20% accuracy rate is achieved with 10, 15, and 20 classes, respectively. Haar level-5 DWT filter shows 97.20%, 94.40%, and 91% accuracy rate for 10, 15, and 20 classes, respectively. The proposed system is also compared with a baseline method where it shows better performance. The proposed system can be utilized as a communication interface to computing and mobile devices for low literacy regions. [6]

3. PROPOSED SYSTEM

Normal voice will be fed using a microphone which will be furtherly given to the Voice recognition model. This operates taking speech signal as an input process on it using Mel Frequency Cepstrum Coefficient (MFCC) in an artificial neural network provides training to the system and later on it will be fed test Raspberry Pi. These coefficients will be recorded in the cloud. As the computer is set login to the cloud via IoT.

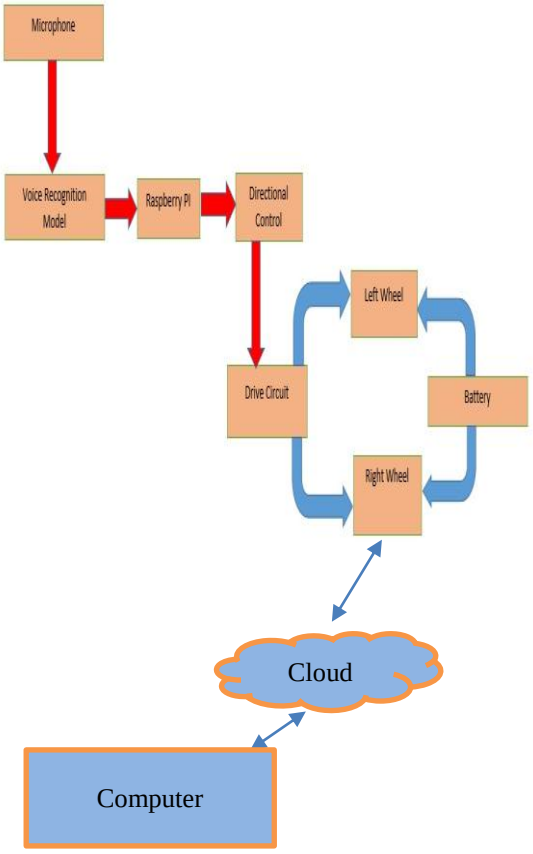

Fig 1. Smart car with IoT

Fig1. Shows the proposed system of smart car with IoT. For the successive work of a self-governing car, we are giving the voice commands via the microphone we will give the voice commands to the device. The Voice recognition module will decide whether to receive a voice command or not. This command will send to the Raspberry Pi. Raspberry Pi is the main block for this system. It contains the processer, memory& another important module for the Autonomous car. Raspberry Pi have fed signal to the direct control of the vehicle block, which basically controls the direction of the wheels. To change the direction of the wheel left or right side we use the left & right wheels & Drive circuit and light on and light off. It will also control air conditioner with Fan ON and Fan OFF. (Fan will be operating as Air Conditioner). The battery will give continuous supply to the motors' wheels when required. At other end Raspberry pi will be connect through Thingspeak cloud. Where cloud is connected to computer through internet.

The role of raspberry pi is having ARM cortex-A53 integrated with 802.11n wireless LAN and Bluetooth 4.1. It is debit card shape computer which perform number of task like camera, voice and HDMI interfaces.

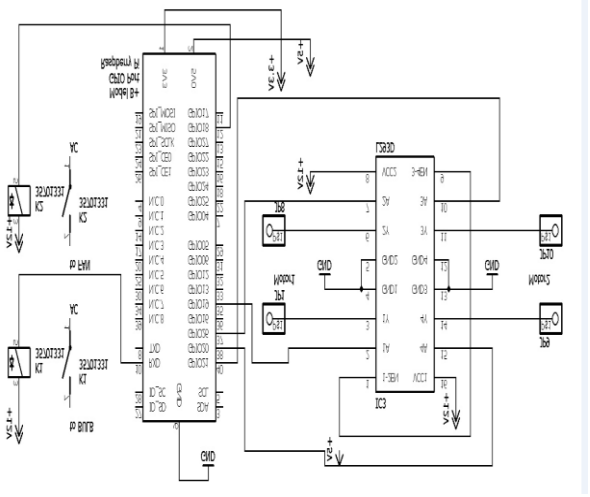

Fig 2. Circuitry Diagram of interfacing Raspberry Pi with Directional controller

Fig 2. shows the circuit diagram of the interfacing between Raspberry Pi with Directional Control. As the motor works on low voltage driver IC L893D is used as a motor driver. Here we are using two motors for the left and right side there are two wheels. This circuit of motor and motor drivers is connected to the GPIOs of the raspberry pi.

Using the audio input function of the raspberry pi we will give the voice command through the microphone depends on this commands motor will be rotated and motor rotate in which direction the vehicle is moved.

4 .SPEECH RECOGNITION USING ARTIFICIAL NEURAL NETWORK

After pre-processing, 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.

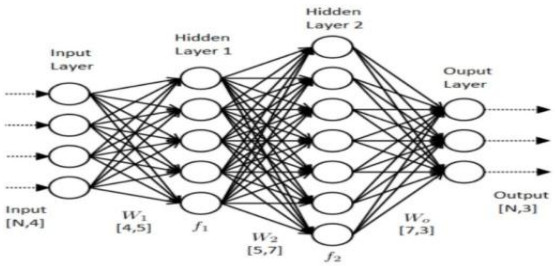

Fig 3. Artificial Neural Network [7]

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

5. Selection Criteria of Voice Recognition Model

There are different types of feature extraction LPC (Linear Prediction Coefficient Cepstral), MFCC (Mel Frequency Cepstra Coefficient), PLP (Perceptual Linear Prediction Cepstral). Due to its advantage of less complexity in the implementation of feature extraction algorithms, & good spectral smoothing MFCC's are used extensively in Automatic Speech Recognition (ASR). For better results we select MFCC.

A high level intro to the implementation steps, towards the end of into a more detailed description of how to calculate MFCCs.

1. Frame the signal into short frames.
2. For each frame calculate the period gram estimate of the power spectrum.
3. Apply the Mel filter bank to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filter bank energies.
5. Take the DCT of the log filter bank energies.
6. Keep DCT coefficients 2-13, discard the rest.

There are a few more things commonly done, sometimes the frame energy is appended to each feature vector. Delta and Delta-Delta features are usually also appended. Liftering is also commonly applied to the final features.

- **Triangular Filter-Bank design**

The necessary information in a human speech signal contained in such a frequency range, whose band shape looks like a triangle [7]. Therefore, a filter bank of 16 triangular band pass filters is designed.

- **Frame Blocking**

The input speech waveform is cropped to remove silence or acoustical interference that may be present in the beginning or end of the sound file [8]. Here sampling frequency is 8 kHz, it means 8000 samples per second. Taking 20 ms frame length, speech samples are divided into frames, each consisting of 160 samples.

- **Frame Overlapping**

50% overlapping of the frames is done to remove the disadvantage of windowing functions i.e., attenuation of the beginning and end of the signal in the calculation of the spectrum.

- **Windowing**

Each frame of speech samples is applied to the windowing block to minimize the discontinuities of the signal by tapering the beginning and end of each frame to zero. A window is shaped so that it is exactly zero at the beginning and end of the data block and has some special shape in between. This function is then multiplied with the time data block forcing the signal to be periodic. In this way, windowing reduces leakage. Here, Hanning window used due to good frequency resolution property. The Hanning window is one type of raised cosine window [9]. This cosine window is defined by:

$$w_c(\mathrm{n}) = 0.5\left(1 - \cos\frac{2\pi n}{N-1}\right) \quad for\ n = 0\ to\ n = (N-1) \qquad -(1)$$

where, n represents the sample number and N represents the width, in samples, of discrete-time, symmetrical window function w(n).

- **Fast Fourier Transform (FFT) of windowed frame**

One of the most common techniques of studying a speech signal is via the power spectrum. The power range of a speech signal defines the frequency pleased of the signal completed time. So the Discrete Fourier Transform (DFT) completed changing a finite list of equally spread out samples of a function into the list of coefficients of a finite combination of complex sinusoids, well-ordered by their occurrences, that has individuals same sample ideals. Spectral energy is calculated using 512-point DFT. It converts the sampled function from its original domain (often time or position along a line) to the frequency domain. The sequence of N complex numbers x1, x2, …., xn is transformed into an N-periodic sequence of complex numbers

X0, X1……..., XN-1 according to the DFT formula:

$$X(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi kn}{N}} \qquad -(2)$$

where, k = number of filter[10] .

A speech signal contains only real point values, so real-point Fast Fourier Transform (FFT) used for increased efficiency due to its rapid conversion ability.

The Mel Frequency Cepstrum Coefficient (MFCC) feature consumes been secondhand for scheming a text-dependent speaker identification system. The extracted speech features (MFCC's) of a speaker are quantized to a number of centroids using a vector quantization algorithm. These centroids constitute the codebook of that speaker. MFCC's are calculated in the training phase and again in the testing phase. Speakers uttered the same words once in a training session and once in a testing session later. The Euclidean distance between the MFCC's of each speaker in the training phase to the centroids of the individual speaker in the testing phase is measured and the speaker is identified according to the minimum Euclidean distance.

The withdrawal and collection of the best parametric demonstration of acoustic signals is significant task in the design of any speech recognition system; it suggestively moves the acknowledgement presentation. A solid demonstration would be on condition that by a set of Mel-Frequency Cepstrum Coefficients (MFCC), which are the outcomes of a cosine transform of the real logarithm of the short-term energy spectrum articulated on a Mel-frequency scale. The MFCCs are verified more well-organized. The intention of the MFCC includes the resulting steps.

Mel-frequency wrapping

Anthropological observation of occurrence stuffing of sounds for speech signal does not monitor a linear scale. So for each tone with a definite occurrence, f, dignified in Hz, a particular pitch is dignified on a measure called the 'Mel' scale. The name Mel comes from the word melody to

indicate that the scale is based on pitch comparisons [11]. The Mel frequency scale is a linear occurrence layout under 1000 Hz and a logarithmic spacing above 1000Hz. As a situation point, the pitch of a 1 kHz character,40dB above the perceptual range threshold, is sharp as 1000 Mels. Hence, the subsequent rough procedure to work out the Mels for a given frequency f in Hz.

$$F_{mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad -(3)$$

OR

$$F_{mel} = 1127 \log_{e}\left(1 + \frac{f}{700}\right) \quad -(4)$$

An approach to simulate the subjective spectrum is to use a filter bank, one filter for each desired Mel-frequency component. The filter bank has a triangular band pass occurrence comeback and the arrangement, as well as the bandwidth, is resolute by a continuous Mel-frequency interval.

The Mel scale filter bank is a series of 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

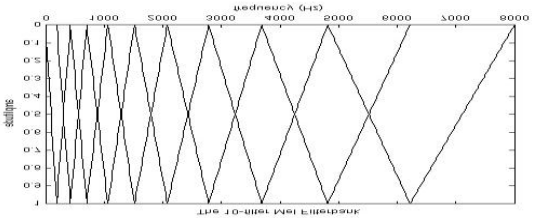

Fig 3: The ten-filter Mel filter bank [12]

bank is shown in the above figure. This corresponds to a series of band pass filters with constant bandwidth and spacing on a Mel frequency scale.

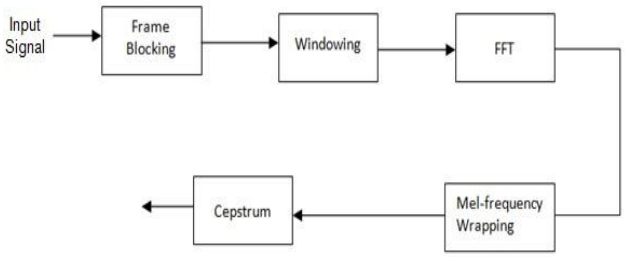

Fig 4. MFCC Feature Extraction [12]

The log Mel spectrum will convert back to time. The outcome is known as the Mel Frequency Cepstrum Coefficients (MFCC). The cepstral illustration of the speech spectrum arranges for a decent illustration of the native spectral belongings of the signal for the particular frame inquiry. Due to the Mel spectrum coefficients (and so their logarithm) are real figures, convert them to the time domain via the discrete cosine transform (DCT). Finally, log, the Mel spectrum is converted backbone to time. The outcome is called the Mel Frequency Cepstrum Coefficients (MFCC). The discrete cosine transform is done for transforming the Mel coefficients back to the time domain as per the above figure of MFCC Feature Extraction. By using artificial neural networks, At the end can be obtain further training and testing of the system where raspberry pi can be acted [12].

6. FLOWCHART OF RASPBERRY PI MODEL

The Raspberry Pi will connect if data will present then it will communicate through cloud otherwise no data present instruction will be given to voice recognition model.

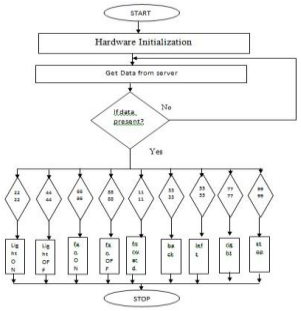

Fig 5. Communication with cloud and relevant code will be executed

7. RESULTS AND DISCUSSION

The input is a voice from the microphone is fed to the PC in terms of voice signal. The role of the PC is to provide output according to the program the commands will be sent to a raspberry pi.

- The output of raspberry pi is high then the intermediate relay module will have an output of low. Hence light will be OFF.
- The output of raspberry pi is low then the intermediate relay module will have an output of high. Hence light will be ON.
- For robot application, here DC motor is used. So a motor driver is required due to the following two reasons.
- Raspberry pi output gives max 30 to 40 mA of current but the motor needs 1Amp of current.
- Raspberry pi output gives max 3.3 V of voltage but the motor needs 12 V of voltage.

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”

Table 1. Spoken word executed with relevant code

The reason for the difference is background noise. It is due to background sound however training differs from the one whereas testing. Another reason is the technique we utter the word for the duration of testing varies with the one while training. Changed pronunciation for the same word if diverse the performance gets pretentious.

8. CONCLUSION

The recording is done for number of files by giving each and every command. Features of the .wav files are captured using MFCC. The training worked out using Artificial Neural Network (ANN) and testing of command is matched with the trained database. Once it matched with the database and it is validated then the command will be executed. In this paper, raspberry pi is used for smart cars to recognize spoken words. In addition to its Internet of Things (IoT) is used to access. The system is proven to be accurate. More accuracy can be achieved by increasing more number of iteration while training.

REFERENCES

[1] L. Li et al., **"Hybrid Deep Neural Network--Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition,"** 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, 2013

[2] 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.

[3] Chin Kim On, Paulraj M. Pandiyan, Sazali Yaacob, Azali Saudi, "**Mel-Frequency Cepstral Coefficient Analysis in Speech Recognition"** ICOCI, IEEE 2006.

[4] Arnab Pramanik, Rajorshee Raha**, "Automatic Speech Recognition using Correlation Analysis"** World Congress on Info-mation and Communication Technologies, IEEE 2012, pp.670-674.

[5] 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.

[6] Bacha Rehmam, 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.

[7] https://medium.com/coinmonks/the-artificial-neural-networks-handbook-part-1-f9ceb0e376b4

[8] Dr. Shaila D. Apte, **"Speech and Audio Processing"** Wiley India publication, ISBN :9788126534081.

[9] A. Nagoor Kani, **"Digital Signal Processing"** Tata Mc Graw Hill Publication, second edition, ISBN-13:978-0-07-008665-4, ISBN-10: 0-07-008665-6.

[10] John G. Proakis and Dimitris G. Manolakis, **"Digital Signal Processing, Principles, Algorithm, and Applications"** Pearson Publication, Fourth Edition. ISBN 978-81-317-1000-5.

[11] https://en.wikipedia.org/wiki/Mel_scale

[12] 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.