*International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*
**Web Site: www.ijettcs.org Email: editor@ijettcs.org**
**Volume 4, Issue 3, May-June 2015** **ISSN 2278-6856**

# Government Policies Search using Marathi Speech Recognition System– Based on MFCC with Gammatone Filter

**[1,]Mugdha Parande, Prof. Shanthi Therese[2] , Prof. Vinayak Shinde[3]**

[1]Mumbai University, Shri L.R. Tiwari College of Engineering,
Kanakia Park,Mira Road(East).Thane,.Maharashtra, India

[2] Thadomal Shahani Engineering College P. G. Kher Marg, (32nd Road), Linking Road.
Bandra (West), Mumbai - 400 050.

[3] Shri L.R. Tiwari College of Engineering, Kanakia Park,Mira Road(East).Thane,India

**Abstract**
***Speech is a natural mode of communication for people. Yet people are so comfortable with verbal skills rather than writing. Today's digital world needs to interact with the computers via speech which can be easier and fast. Outstanding work in speech recognition and feature extracting has produced the commercial speech recognition systems for speech driven computing and word-processing systems. This paper deals with an application to make an easy search to retrieve Government Policies from system database by using the techniques of speech recognition. For developing application we use hybrid feature extraction technique i.e. MFCC with Gammatone Filter Bank and for recognition Vector Quantization is used.***
**Keywords:** MFCC, Gammatone Filter Bank

## 1. INTRODUCTION

There are various ways to communicate with each other such as writing, speaking etc. Speech is the most desirable medium of communication. As we know Government of India facilitates tones of policies and scheme for development and growth but every person cannot take benefits from it because lack of policy reaches. This is the main huddle which resists our India to reach at the top of the world. This paper proposes recognition application to interact with system to get Government of India's policies and schemes. In recent decades, speech recognition systems have significantly improved. Nevertheless,
obtaining good performance in noisy environments still remains a very challenging task. Therefore, in proposed application Gammatone Filter used as a pre-process for reducing noise distortion and then Mel Frequency Cepstral Coefficient (MFCC) is used for extracting features. At the end Vector Quantization is one of the techniques for recognition.

### 1.1 What is speech Generation?

Speech recognition system which recognizes whatever the user said. But for recognizing speech one needs to understand how speech generates and what features can make more accurate perception of speech.
Some messages or ideas come to one's mind and then it converts into language code for example you know English and Marathi then first you decide which language you want to use. Now you convert a message into phonemes sequence of corresponding language. For example in English there are 45 phonemes. Now muscular like tongue, lips and vocal tract produce an acoustic waveform. Most of time, same word of acoustic wave may differ because of duration, loudness and pitch of sound. When this acoustic waveform comes to ear then ear's basilar membrane works like a Spectrum Analyzer. Different points on basilar membrane respond to different frequencies. In acoustic waveform, different features are found like pitch period, loudness, and formant frequencies. Depending on extracted features, your mind decides phoneme in a sequence corresponding to acoustic waveform reaching one's ear. From those phonemes, word and sentences are generated and then one's mind studies that message. [8]

## 2. SPEECH RECOGNITION PROCESS

Process of Speech Recognition: Figure [1] shows Block diagram of speech recognition. Speech is the input to Automatic Speech Recognition (ASR) system. Speech is divided into sequence of 10- msec frames for faster processing. In many cases, the math assumes that the signal is periodic. However, when it take a rectangular window to extract an observation at one frame, It

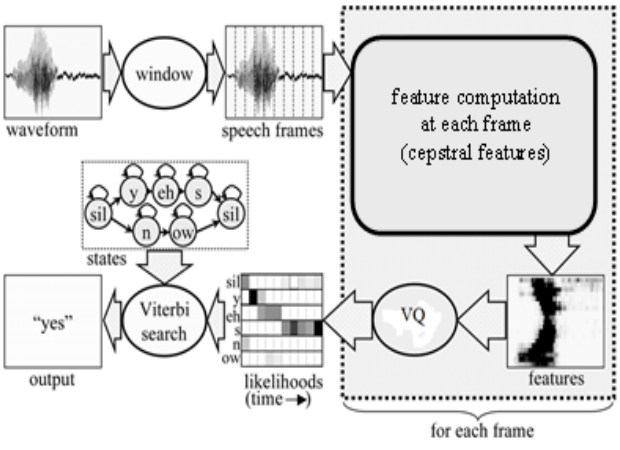


**Figure [1]** Block diagram of Speech Recognition

discontinuities in the signal at the end. So window of signal can be of other shapes, making the signal closer to zero at the end. Window size does not have equal frame size.

Output of pre-emphasis phase is given to feature extractor for feature computation at each frame. There are many feature extraction algorithms such as PLP, LPC [11], MFCC, and PNCC [12]. Those algorithms generate (features) cepstral coefficient values. Vector Quantization (VQ) is a method of automatically partitioning a feature space into different clusters based on training data. A "codebook" lists central locations of each cluster, and gives each cluster a name. This can be used for data reduction or for probability estimation. At the end Viterbi search algorithm which gives recognized words from the given speech.

## 3. Proposed Speech Recognition Approach

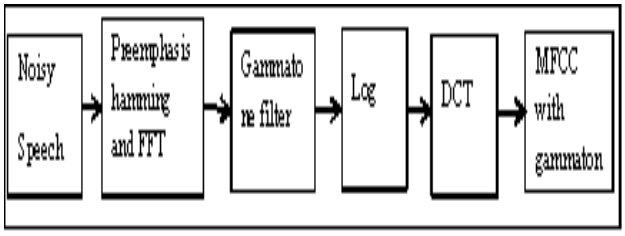


**Figure [2]** Proposed feature extraction approach

The figure shown proposes approach of speech recognition. It differs from traditional MFCC technique for feature extraction that involves Gammatone Filter. It is a linear filter, described by an impulse response. Which is supports the product of a gamma distribution and sinusoidal tone. It's mainly used in proposed system for preprocessing phase which helps to discard distortion occurred by noise.

An impact of Gammatone filter shown in figure [1] and [2], both figures represents wav files. In fig [1] "whatsup" is a word spoken by a person that wav file has noise distortion because of environment therefore, it gives output as "wh@at$%sup". The same word spoken by the same person, that wav file represents in fig [2] but it gives clean wav file compared with fig [1] because of Gammatone filter which reduces noise distortion.

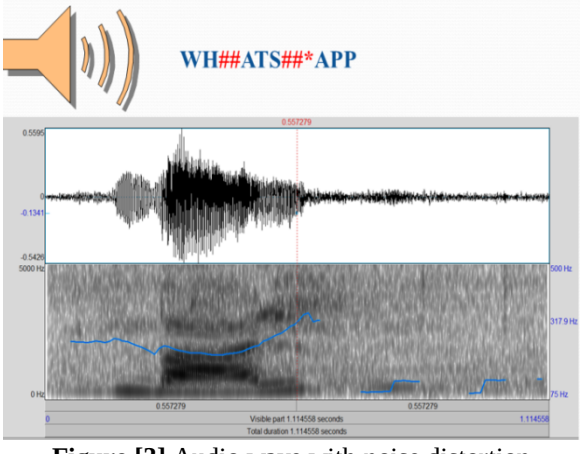


**Figure [3]** Audio wave with noise distortion

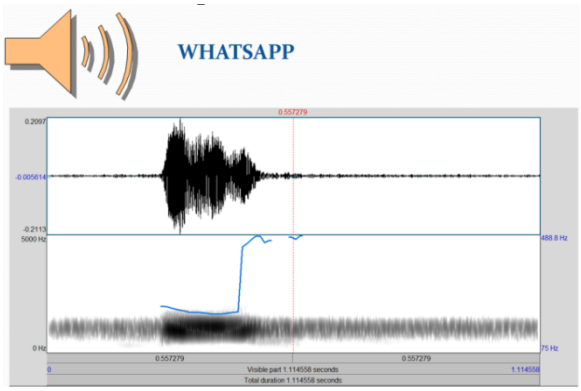


**Figure [4]** Audio Wave with Gammatone Filter

Preprocess of Gammatone Filter and Feature Extraction of MFCC is composed of the following steps:

1. Pre-emphasis

$$s'_n(m) = s_n(m) - 0.97 \cdot s_n(m-1)$$

2. The impulse response of a Gammatone filter centered at frequency f is given by

$$g(t)=at^{n-1}e^{-2\pi bt}\cos(2\pi ft+\emptyset)$$

3. Hamming Window

$$h(n) = 0.54 - 0.46\cos(\frac{2\pi\, n}{N-1})$$

4. Power Spectrum (not db scale)

$$S = (Xr2 + Xi2)$$

5. Mel Scale Filter Banks (Triangular Filter)

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

6. Compute log Spectrum from filter banks

$$10\log 10(S)$$

7. Convert log energies from filter banks to cepstral coefficients

$$c_i = \sum_{j=1}^{N} m_j \cos(\frac{\pi\, i}{N}(j-0.5)) \qquad m_j = \log\text{energy values}, \quad N = \text{number of filter banks}$$

8. Weight cepstral coefficients

$$c'_n = \exp(n\cdot k)c_n \qquad k = 0.6$$

## 4. Proposed Application of Recognition

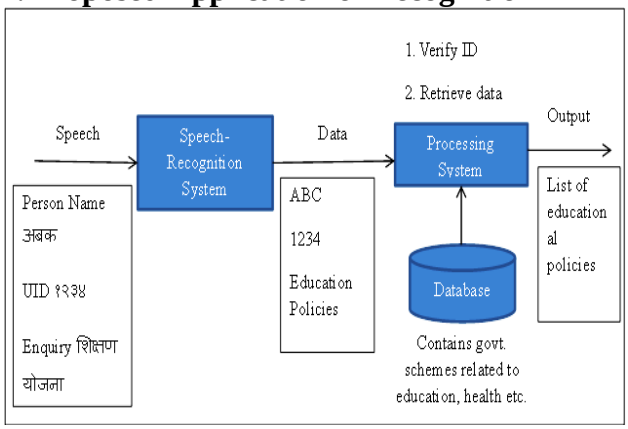


**Figure [5]** Proposed application for speech recognition

Figure shows flow of application. There are two modules
1) Speech Recognition system to recognize sound data and convert it into text data which uses proposed speech recognition approach as mentioned above.
2) Processing System to perform Retrieval function which retrieves Government of India's policies and schemes from database. Finally displaying the output as per the user requires.

## 5. Experimental Results

The proposed approach gives more accurate result than traditional MFCC because of Gammatone Filter Bank. Experimental Results are shown in the figure [1] and [2]. Those results generated by MATLAB program which distinguish the impact of pre-process with gammatone filter bank and without gammatone filter bank. Figure[1] has more dark blue color than Figure[2] it means figure[1] has more noise distortion than the other. This was visual comparison used before implementing speech recognition application with Gammatone Filter as pre-process.

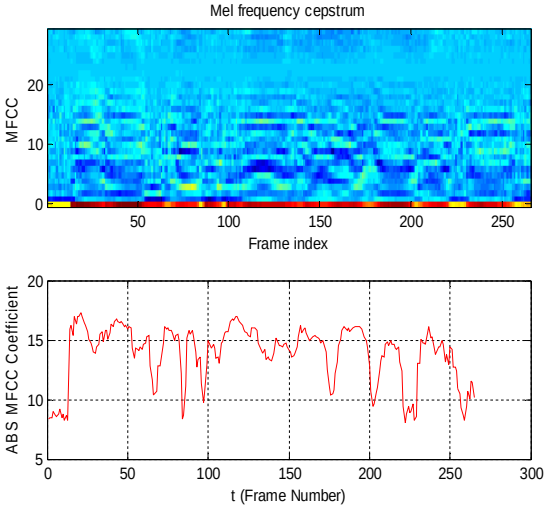


**Figure [6]** Result of Mel Frequency cepstral Coefficient

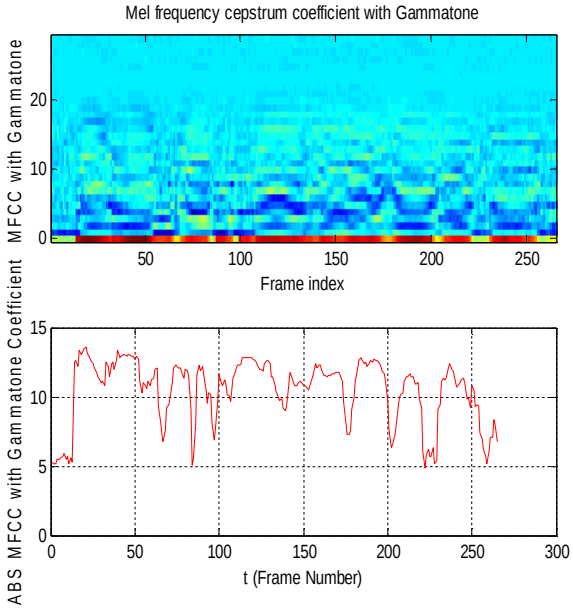


**Figure [7]** Result of Mel Frequency Cepstral Coefficient With gammatone filter Bank

**Table [1]** Recognition accuracy with white noise

| Word | Accuracy with Gammatone Filter Bank (%) | Accuracy without Gammatone (%) |
|---|---|---|
| India | 97 | 90 |
| Bharat | 96 | 89 |
| Shikshan | 96 | 90 |
| Sarkar | 97 | 90 |
| Swacha | 96 | 89 |
| Krushi | 96 | 90 |
| Mahila | 94 | 88 |
| Vikas | 97 | 90 |
| Yojna | 96 | 89 |
| Adhar | 94 | 90 |
| Digital | 95 | 92 |
| Gram | 97 | 90 |
| Jyoti | 96 | 89 |
| Bal | 96 | 90 |
| Adhar | 94 | 90 |
| Sukanya | 95 | 92 |
| Beti | 97 | 90 |
| Bachav | 94 | 92 |

The Speech Recognition Application composes the following steps
1. Need to Train dataset.
2. Creates dictionary for mapping codebook.
3. After training of dataset, Vector Quantization generate text file which contain features.
4. Start the application with Login Username and Password (To administer for security purpose).
5. Search Indian Government policies and schemes with one's voice input.
6. Retrieve Indian Government policies and schemes.

Speech Recognition Application can train Marathi language to generate VQ codebook file. This application now mainly focuses on rural citizen in Maharashtra therefore application converts Marathi language into English and then retrieve function call for displaying Indian Government policies and schemes. In Training Phase, each word needs to train; every training dataset contains 20 samples. Those samples consist of 10 male samples and 10 female samples for multiple user purpose. In the application, Dictionary is needed for mapping purpose. This contains 50 words in Marathi language. After login to application any one can take benefits from application through simple input as voice. Click 'Record Button' and speak name of the policy or the scheme. Automatically spoken voice converts into text and displays into the text area. Click 'Retrieve Button' to display the policy or the scheme file on the application screen.

Snapshot of Application

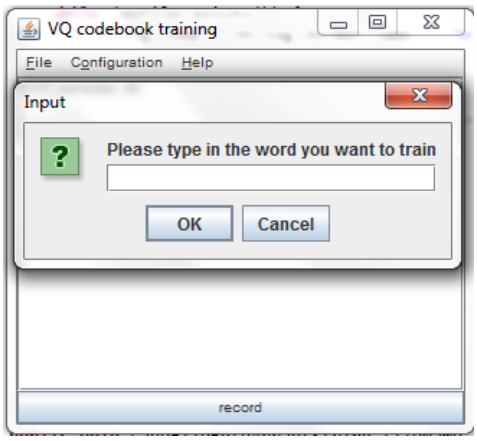


**Figure [8]** Training dataset dialogue box

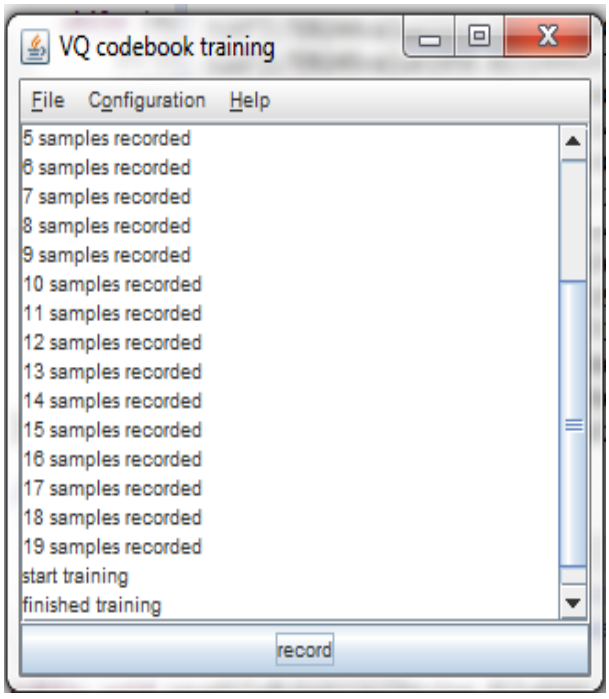


**Figure [9]** Finishing Training dataset

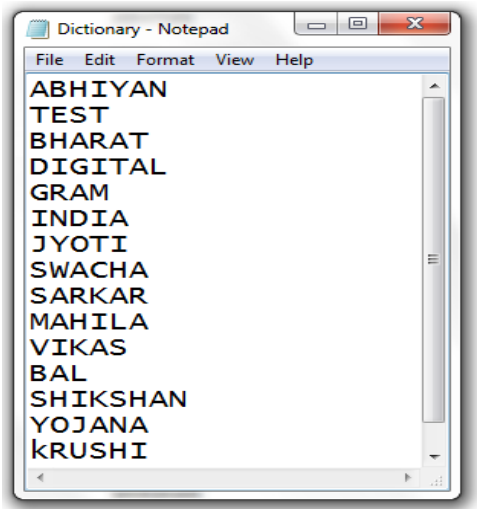


**Figure [10]** Screen of dictionary


**Figure [11]** Login Page of application

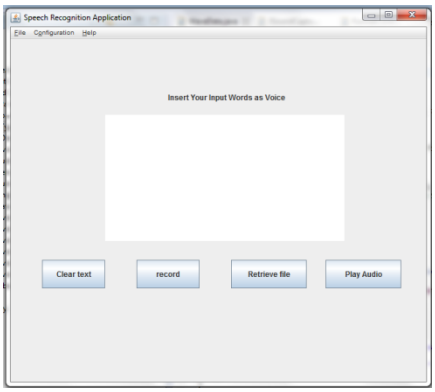


**Figure [12]** Screen of recognition in application

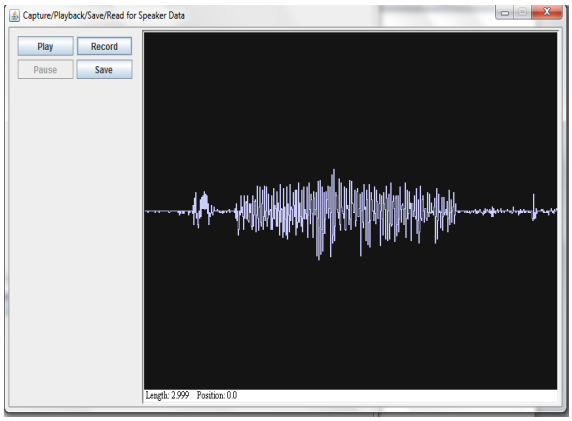


**Figure [13]** Screen of recording in application

## 6. CONCLUSION

The Speech recognition application can effectively be used in public sector. This application is very useful to illiterate people in respect to being familiar with Indian Government policies and schemes. Implementation of application based on MFCC with Gammatone Filter Bank as pre-process which helps to improve accuracy of recognition. Accuracy of recognition in white noise is more while using Gammatone Filter Bank with MFCC. There are some drawbacks like need to wide search area and to improve more accurate probability of getting search estimation.

## References

[1]. N. S. Uchat, "Hidden Markov Model and Speech Recognition," Department of Computer Science and Engineering, IIT Bombay, Mumbai, 2012.

[2]. P. M. a. T. F. Gellert Sarosi Mihaly Mozsary,"Comparison of Feature Extraction Methods for Speech Recognition in Noise Traffic Noise Environment,"Dept. of Telecommunications and Media Informatics Budapest University of Technology and Economics Budapest, Hungary, vol. 3, no. March 2008, pp. 735-743, 2008

[3]. R. K. A. a. M. Dave, "Using Gaussian Mixtures for Hindi Speech Recognition System," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, no. 4, pp. 50-64, December 2012

[4]. Shanthi Therese S. Chelpa Lingam, "Review of Feature Extraction Techniques in Automatic Speech Recognition", International Journal of Scientific Engineering and Technology, vol. 02, no. 06, pp. 479-484, 2013.

[5]. "Marathi Isolated Word Recognition System using," BhartiW.Gawali,SantoshGaikwad,PravinYannawar,SureshC.Mehrotr, vol. 01, no. 01, pp. 21-26, 2011.

[6]. M. Brookes. VOICEBOX: Speech Processing Toolbox for MATLAB. Software, available [Mar, 2011] from, www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[7]. E. Ambikairajah, J. Epps, L. Lin. Wideband speech and audio coding using gammatone filter banks. Proc. ICASSP'01, Salt Lake City, USA, May 2001, vol.2, pp.773-776.

[8]. "Fundamentals of Speech Recognition" Lawrence Robiner book published by Prentice Hall Signal Processing Series.

[9]. http://ocvolume.sourceforge.net

[10]. Octavian Cheng, Waleed Abdulla, Zoran Salcic "Performance Evaluation of Front-endProcessing for Speech Recognition Systems" in Electrical and Computer Engineering Department, The University of Auckland in 2005.

[11]. Namrata Dave1," Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition",International Journal For Advance Research In Engineering And Technology, vol.1, issue VI , July 2013.

[12]. Chanwoo Kim and Richard M. Stern "Power-Normalized Cepstral Coefficients (PNCC) For Robust Speech Recognition" In IEEE, 2013

[13]. R.Schl uter, I. Bezrukov, H. Wagner, H. Ney, "Gammatone Features and Feature Combination for Large Vocabulary Speech Recognition" In Biology Department RWTH Aachen University, Aachen, Germany

[14]. Vimala.C, Radha.V,"Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words" in International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 378-383

## AUTHOR

**Mugdha Parande** received the B.E in DKTE College of Engineering in Information Technology and M.E.(pursuing) degree in Computer Engineering from Shree L.R. Tiwari College of Engineering.

**Prof. Shanthi Therese** working in Thadomal College of Engineering as a Associate Professor of department of Information technology. Pursuing Ph.D. Area of specialization are Data Mining Techniques & Algorithms , Image Processing , Speech processing

**Prof. Vinayak Shinde** working in Shree L.R. Tiwari College of Engineering as a Assistant Professor and H.O.D of department of computer science. Pursuing Ph.D.
Area Of Specialization are Microprocessor And Microcontrollers Coa And Netwoking