ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

# HINDI SPEECH RECOGNITION SYSTEM USING MFCC AND HTK TOOLKIT

**Nikita Dhanvijay *, Prof. P. R. Badadapure**
Department of Electronics and Telecommunication
JSPM'S imperial college of engineering
Wagholi, pune - 412207, India

**DOI**: 10.5281/zenodo.212079

## ABSTRACT

This paper presents the approach for Hindi fruit name recognizer system. Every person has its uniqueness in his speech. So in this approach the database speech samples are collected from different 20 speakers with two iterations. These recordings are used to train by acoustic model. This model is trained on 20 speaker database having vocabulary size is 45 words. HTK toolkit is used to train the input data and evaluation of the results. The proposed system gives a recognition rate of 94.28% for sentence and 98.09 for word level.

**KEYWORDS**: HMM (Hidden Markov Model), ASR (Automatic Speech Recognition), Speech Recognition (SR). MFCC (Mel Frequency Cepstral Coefficient)

## INTRODUCTION

Today computer is the device which connects the world. To communicate between human and computer, the input devices such as mouse keyboard is needed. In a developing country like India, computer illiteracy is more. They can't handle a computer by using input devices. So there is a need of such a technique, so that every person can communicate with world without prior knowledge of computer.

The barrier between computer and human can be minimized by using speech as an input. Speech is one of the powerful and natural tools for the communication. Existing speech recognition systems are working well for European language like English. It can be possible to implement such a system for India's most speak able language 'Hindi'.

The primary aim of this system is to implement an ASR system for Hindi language. The system takes the Hindi audio as an input to the system along with its textual labels. Speech recognizer converts the spoken word into text.

There are some software tools available for automatic speech recognition: Carnegie Mellon University in 2011, developed software called "SPHINX", HTK toolkit and LVCSR.

The organization of this paper as follows: In section 2, the detailed survey of the previous speech recognition system has been studied. The section 3 explains the proposed system based on HTK toolkit. The results are explained in section 4 and conclusion in the last section of the paper.

## LITERATURE REVIEW

This section explains reported techniques for automatic speech recognition system.

Annu Choudhary et al. [1] implemented an isolated words and connected words ASR for Hindi language. The process of converting an acoustic waveform into the text is called speech recognition. The system was developed by HTK toolkit based on HMM (Hidden Markov Model). The MFCC feature extraction technique is use by ASR System. The Hindi language is supported for isolated and connected words. The results show the accuracy for isolated words as 95% and for connected words as 90%.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Bhadragiri Jagan Mohan et al. [2] implemented a Speech recognition system using MFCC (Mel-Frequency Cepstral Coefficients) and DTW (Dynamic Time Wrapping) algorithm in MATLAB environment. MFCC algorithm is use for feature extraction and DTW is use for pattern matching. Results are obtained by one time training and continuous testing phases. The results shows that saving ten templates for each word in training phase give good results compared with five templates. The efficiency in detecting isolated words is 100 % for two syllable words compared with one syllable word.

Babita Saxena et al. [3] present digit speech recognition system for different speaker for different environment. The noisy environment like vehicles Horn noise, home noises such as opening door sound, silence in the room etc. The training model is trained using 8 speakers. For recognition of digit speech HTK toolkit is used.

Vidwath R. Hebse [4] et al. implemented an Automatic Speech Recognition System using MFCC and HMM. Preprocessing, Feature Extraction Technique, MFCC and HMM are used for speech recognition system. The HMM is used to develop a voice based user machine interface system. This user machine system can be using various applications and it wills advantages as real interface.

M. A. Anusuya [5] explains Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition. Paper shows that, the feature extracted with wavelet transforms has the highest accuracy for recognition of words. This has been proved in both the conditions for clean and noisy speech signals. It is shown that the accuracy of the system get increase by use of DWT and wavelet packets for noisy as well as clean signals. It is showed that, using wavelets the recognition accuracy can be increased for both clean and noisy speech signals. , the paper shows that, any feature extraction method with wavelet yields, good recognition accuracy of the speech signals.

Tarun Pruthi [6] et al. describes the implementation of Swaranjali, an experimental, speaker-dependent, real-time, isolated word recognizer for Hindi. The paper explains the implementation of the system. An experimental isolated word recognition system for Hindi language was implemented. The scheme proposed uses a standard implementation, with some modifications to the noise detection/elimination algorithm and the HMM training algorithm. The training set for the vector quantizer was obtained by recording utterances of a set of Hindi words encompassing the Hindi alphabet. The results were found to be satisfactory for a vocabulary of Hindi digits. Further improvement can be obtained by a better VQ codebook design, with the training set including utterances from a large number of speakers with variation in ages and accents.

Dimitris Spiliotopoulos [7] et al. explains human-robot spoken dialogue interaction in the context of Hygeiorobot to build a mobile robotic assistant for hospitals. The main focus on dialogue management issues and on particular issues that needs to be addressed in human robot interaction and the considerations that influenced design of Hygeiorobot's dialogue manager. The system has been tested on robot controller simulator. The spoken word is more suitable as robot doesn't carry Mouse and Keyboard and is intended to be used by people with little or no computing experience.

Preeti Saini et al. [8] aim to build a speech recognition system for Hindi language. HTK toolkit is used to develop the system. It recognizes the isolated words using acoustic word model. The system is trained for 113 Hindi words. Training data has been collected from nine speakers. Automated Speech Recognition (ASR) is the ability of a machine or program to recognize the voice commands or take dictation which involves the ability to match a voice pattern against a provided or acquired vocabulary. The experimental results show that the overall accuracy of the presented system with 10 states in HMM topology is 96.6% and 95.49%.

## PROPOSED SYSTEM

The architecture of proposed speech recognition system is shown in figure 1. The system processes through training and testing module.

The different phases of automatic speech recognition system are explained below:

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

### 1.1. Data Preparation for Hindi Language

In India, there are more than 200 languages being spoken. Among them 22 languages are official languages. Most of the languages among them were evolved from the ancient Brahmi script and having same phonetic structure.

Most of the scripts in India were written in the Devanagari. It includes Marathi, Hindi, Nepali, Bengali etc. The character set of Hindi language is shown in Table 1:

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vowels | अ | आ | इ | ई | उ | ऊ | ऋ | ए | ऐ | ओ | औ | अं | अः |
| | a | ā | i | ī | u | ū | r | e | ai | o | au | aṁ | aḥ |
| Gutturals (कवर्ग) | क | ख | ग | घ | ङ | | | | | | | | |
| | ka | kha | ga | gha | ṅa | | | | | | | | |
| Palatals (चवर्ग) | च | छ | ज | झ | ञ | | | | | | | | |
| | ca | cha | ja | jha | ña | | | | | | | | |
| Cerebrals (टवर्ग) | ट | ठ | ड | ढ | ण | | | | | | | | |
| | ṭa | ṭha | da | dha | ṇa | | | | | | | | |
| Dentals (तवर्ग) | त | थ | द | ध | न | | | | | | | | |
| | ta | tha | da | dha | na | | | | | | | | |
| Labials (पवर्ग) | प | फ | ब | भ | म | | | | | | | | |
| | pa | pha | ba | bha | ma | | | | | | | | |
| Semi-Vowels | य | र | ल | व | | | | | | | | | |
| | ya | ra | la | va | | | | | | | | | |
| Sibilants | श | ष | स | | | | | | | | | | |
| | sa | pa | sa | | | | | | | | | | |
| Aspirate | ह | | | | | | | | | | | | |
| | Ha | | | | | | | | | | | | |

Table 1: Hindi Character Set

### 1.2. Preprocessing

Speech signals are analog signal which cannot be processed through digital filters. The main task of preprocessing is to convert the speech signal into the format which can process by the recognizer. The analog signal first convert into digital signal by using first order filter to spectrally flatten the signal. In next step these signals convert into sequence of frame with specific timestamp.

### 1.3. Feature Extraction

Feature extraction is process of computing the acoustic correlation between the speech signals which represent the audio waveform. It keeps the useful information of the audio signal and discards the other. The extracted features are stored in an array called as feature vector.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

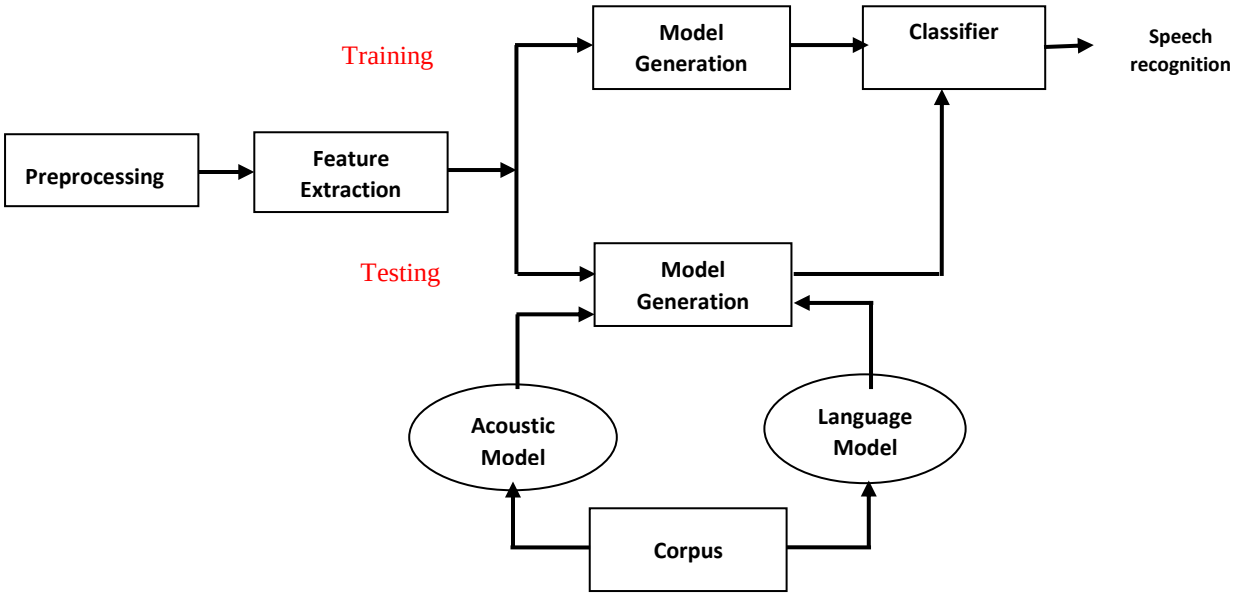


**Figure 1: Architecture of ASR system**

Feature extraction for this module having two steps: in first step non uniform filters on Fourier spectrum of speech signals is applied and cepstrum coefficients are extracted while in second step try to reducing the features vectors into minimum feature space.

In the proposed system the features are extracted using MFCC technique. in this approach, the speech signals are separated into frames. Frame are preprocess through the pre-emphasis filter to get amplified higher frequencies. Hamming window and Fourier spectrum is applied to compute the windowed frame signal. The Mel Spaced filter bank is applied on the signal and converts these signals to the cepstral coefficient by DCT.

The working flow of MFCC algorithm is as shown in figure 2.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

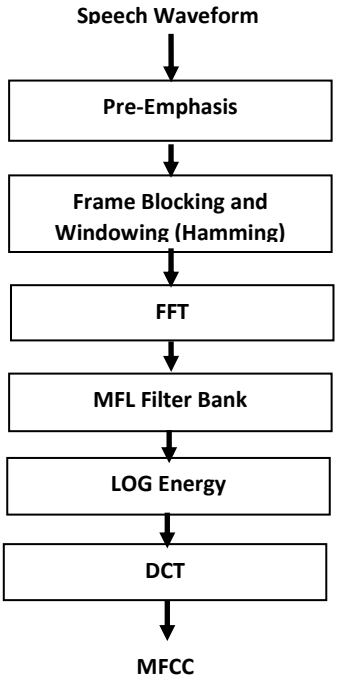


**Figure 2: MFCC feature extraction**

Above Figure shows the overall process to extract the MFFC vectors from the speech signal. It gives special importance to the process of MFCC extraction is applied over each frame of speech signal independently. The MFCC vectors will be obtained from each speech frame after the pre-emphasis and the frame blocking and windowing stage. The 1st step of the MFCC extraction process is to compute the Fast Fourier Transform (FFT) of each frame and obtain its magnitude. At the end of the extraction process of MFCC is to apply the modified DCT to the log-spectral-energy vector, obtained as input by the mail filter bank, resulting in the desired set of coefficients called MFCC

### 1.4. Model Generation using Gaussian Mixture HMM

Model generation is important for classification. Different approaches for model generation are HMM, SVM, Artificial Neural Network, and DBN. Some hybrid approaches i.e. combination of two or more classifiers are also used. As speech recognition system is concern, Human Markov Model (HMM) gives better results. In the current system, the HMM is used as a model. Gaussian mixtures are used to calculate the likelihood of observation vectors.

#### a. Phonetic representation

To representation of audio signal in the form of linguistic unit called label, In this approach we select small vocabulary size of 45 isolated words.

#### b. HMM

Hidden Markov Model (HMM) is used as a statistical model. Each phonetics and sub-word is realized by HMM. The HMM have three problems: evaluation of probability, best sequence calculation and estimation of parameters. The first problem of probability estimation is solving by forward algorithm such as Viterbi search. The best sequence is calculated by decoding process and parameters are solved by MLE algorithm.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

## RESULTS AND DISCUSSION

The input speech was recorded and each corpus was made by AUDOCITY software. The recordings were sampled at 16 KHz frequency and 16 bit per sample.

The experimentation is performing on the database 45 fruit name in Hindi language recorded by 20 people with 2 repetitions. Database contains total (45x20x2=1800) samples. At front end MFCC extract the features and at the back end MLE is used. Corpus is made manually.

Performance evaluation of the system is evaluated by recognition rate. The recognition rate is calculated as

$$Recognition\ Rate = \frac{Successfully\ detected\ words}{Total\ no.of\ words\ in\ test\ dataset} \qquad (1)$$

| Method | H | S | I | N | Correct rate |
|---|---|---|---|---|---|
| Sentence | 1697 | 103 | - | 1800 | 94.28% |
| Word | 5297 | 103 | | 5400 | 98.09% |

**Table 2: Correct rate analysis of proposed system**

## CONCLUSION

Human speech recognition is difficult because of variation in the frequency of person to person. In this paper we have proposed a speech recognition system for HINDI language. 45 fruit names are taken as input of this system. The features are extracted using MFCC algorithm and extracted features were classified using Hidden Markov Model with Gaussian mixture to generate acoustic models.

The proposed system is implemented on HTK Toolkit. It gives recognition rate of 94.28% for sentence level and 98.09% for word level.

## REFERENCES

1. A. N. Kandpal and M. Rao, "Implementation of PCA and ICA for Voice Recognition and Seperation of Speech," in proc. of IEEE International Conference on Advanced Management Science (ICAMS), vol. 3, pp. 536-538, 2010.
2. M. A. Anusuya and S. K. Katti, "Mel Frequency Discrete Wavelet Coefficients for Kannada Speech Recognition using PCA," in Proc. of Int. Conf. on Advances in Computer Science, 2010.
3. Tarun Pruthi, Sameer Saksena and Pradip K Das, "Swaranjali: Isolated word recognition for hindi language using VQ and HMM," in proc. Of International conference on multimedia processing and systems, Aug. 2000.
4. D. Spiliotopoulos, I. Androutsopoulos, and C. D. Spyropoulos, "Human- Robot Interaction based on Spoken Natural Language Dialogue", in proc. Of the european workshop on service and humanoid robots.
5. A. A M Abushariah, T. S. Gunawan, O. O. Khalifa, and M. A. M. Abushariah, "English Digits SR System based on Hidden Markov Models," in Proc. IEEE International conference on computer and communication engineering, pp. 1-5, May 2010.
6. K. Kumar and R. K. Agarawal, "Hindi speech recognition using HTK," in International Journal of Computing and Business Research, vol. 2, May, 2011.
7. S. Young, et al., The HTK Book. December, 1995.
8. Gaikwad, S.K. and Gawali, B.A. A review on speech recognition technique. In International Journal of Computer Applications, volume 10, November, 2010.
9. H.-J. Bohme, T. Wilhelm, and J. Key. An approach to multi-modal human-machine interaction for intelligent service robots. Robotics and Autonomous Systems, elsevier science, 44:83–96 December2004.