International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.01, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

# Hindi Speech Recognition

**Ms. Kajal Jewani Asst Prof.[1], Shreesh Rao[2], Prashant Dombale[3], Ronit Dhonde[4], Mrudali Birla[5]**

[1,2,3,4,5]*Department of computer science, V.E.S.I.T., Mumbai University*
*Sindhi Society, Chembur, Maharashtra, (India)*

## ABSTRACT

*Speech is one of the most common and widely accepted mode of communication between humans. It has been working efficiently well and a number of attempts have been made to use it for human computer interaction. Most of these system work with English language. This paper has tried to describe plausible techniques to recognize speech in Hindi. It has limned some of the previously existing speech recognition systems briefly along with the techniques used and has compared them in terms of flexibility, accuracy and celerity. Another purpose of this paper is to open the door for research in speech recognition using other vernacular languages as well.*

***Keywords— Speech recognition, Hidden Markov Model, Vector quantization, MFCC.***

## I.INTRODUCTION

Speech is the fundamental method of modern day communication. Various ideas and thoughts in the minds of different individuals are shared with the help of different languages. A person's speech may comprise of various words, phrases and sentences. In India, since Hindi is the primary language, people exchange their opinions and thoughts by making use of Hindi language.

Hindi language belongs to the Devanagiri script and hence there is a one-to-one correspondence between the sound and the syllable that represents that sound.

Speech recognition is the process of analyzing the input speech from the user, followed by identifying the words and phrases present within the input and then converting them into a machine-readable format. Comparison between the input speech is done with the in-built vocabulary present within the system. The task is to make the computer understand the Hindi speech and then react appropriately by converting the speech into another medium. A Microphone, computer and a good quality sound card are some of the rudimentary apparatus for a speech recognition system.

## II.SYSTEM CLASSIFICATION

Speaker dependent models are developed for individual users. A particular user can operate this model with his voice more accurately than other users. It is accurate but not as flexible as speaker independent system. So, if any other user tries to give input with his voice, the system will not respond in the same way. Hence, as this model is for an individual speaker, it is not adaptive to any other user. But, this type of models is more accurate

as there is no variation in the voice and overall pattern remains the same. Speaker independent model can take voice of variety of users. Also, they are very adaptive and flexible as different users can operate on it. But, as it takes voice of different users the overall accuracy reduces due to the changing pattern in voices. They are more expensive as compared to speaker dependent models and also more difficult to develop as they are complex.

There are different types of speech which can be classified in different ways. Isolated word takes single utterances which are not connected to each other. It can analyze individual words at a time. In Connected word, there is combination of two or more isolated words and allows the utterances which are different to run with a slight pause. When the speaker speaks continuously without any gap, it is known as continuous speech. A special method is needed in order to capture such a kind of speech. Lastly, there is also a kind of speech which is not practised or rehearsed, called as spontaneous speech. It allows variety of words and speeches to be taken.

There are different types of word matching techniques in our system.

Whole word matching: In this type of matching technique, we take the entire sentence and match it with a prerecorded template of words. It requires a lot of memory and takes less processing as compared to sub word matching technique. The template is stored in form of recognition vocabulary in the system and from there words are recognized.

Sub-word matching: In this type of matching technique, sub-words are taken and then pattern is recognized on basis of these words. It usually takes more processing as compared to whole word matching

## III.LITERATURE SURVEY

A recognition system basically follows an algorithm which tries to match the input syllable with the corresponding alphabet present in the dictionary that has been already present in the computer database. A variety of methods can also be used to perform the analysis procedure, but among all of them Mel-Frequency Ceptrum (MFC) turns out to be the most efficient and effective method as it comprises of a strong property where it can compute power spectrum by performing Fourier Analysis. As far as the modelling techniques are concerned the Hidden Markov Model (HMM) proves out to be the most feasible option, where the system being modelled is assumed to be a Markov process with unobserved states.

A speech recognition system is always followed by a speaker recognition system, as it increases the accuracy and efficiency of the software. As discussed earlier, the MFC can extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Identification of a word with the help of HMM on Quantized feature vectors is done in order to maximize the log likelihood values for the spoken word. Formulation of the techniques mentioned above can be executed by the coding the techniques in MATLAB. For best performance and accurate results, MFC and distance minimum algorithm can be combined. According to the experiments performed with this combination the overall efficiency level has turned out to be of 95 percent.

There are a few systems available in the market for speech recognition in Hindi. Shrutlekhan-Rajbhasha is one of the Hindi speech recognition software application available in the market. It is developed by C-DAC and IBM. One drawback of this software is that if mixed English–Hindi dictation is given, it can recognize Hindi words but it is not successful to recognize English words. Also, another variant of this software is available in

the market that is Vachantar-Rajbhasha, which takes English sound as input, converts it to English text and then translates it to Hindi. It uses MANTRA-Rajbhasha translation engine. The major task done by both "Shrutlekhan-Rajbhasha" and its variant "Vachantar-Rajbhasha" is conversion of Hindi to a particular type of script and translation from Hindi to English but it does not perform other tasks and its use is limited.

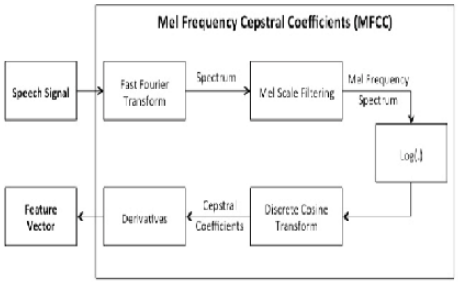
Some of the existing systems very popular in market like Microsoft's cortana, Apple's Siri, Google Now and Amazon Alexa. These can be operated using English language only. Few of them support languages like Hindi to some extent but they show very little accuracy and have various problems like accent of the user and its interpretation by the system. Even though these systems have great features and show high accuracy but are not very accessible to the people who do not know English. Our system mainly aims to cater needs of such population in India. Google has a Hindi Input app which lets Android users type in Hindi. It works like a keyboard extension. Once you install the app, you can type in Hindi. You can compose text messages, type emails and create word documents in Hindi. The basic testing of app has shown less degree of accuracy. Also, one should be able to write to access the app whereas in our system a person not able to write can also access it.

Baidu known as "Google of China" is the country's biggest search engine and at 96 percent. Its voice recognition is better than most humans at identifying spoken words. The system understands both English and Mandarin. In China, voice commands are more popular as it takes up a lot of their time for typing in Mandarin. Similarly, our system will be useful in India to take voice commands using Hindi.

## IV.METHODOLOGY USED

1. Feature Extraction: Feature extraction is the most fundamental step of the speech recognition system. Recording of various speech samples of each word from the vocabulary is done by different speakers. The samples so far collected are then transformed from analog to digital form with the help of sampling frequency of about 16Hz. Filtering of background noise is crucial so that the input speech signal can be processed correctly and thus producing error free outputs. Quantization is a process which can assist in the filtering step. The incoming sound is transformed into an internal representation by feature extraction and also reconstruction of the original signal is possible. MFCC, PLP, RAST, LPCC are some of the techniques which assist to extract features, however MFCC is widely used.

2. Mel Frequency Cepstral Coefficients (MFCC): The design and formulation of MFCC is done in accordance with the human auditory system and hence can be used in every state of speech recognition system or art speech. MFCC includes 5 basic steps which perform the feature extraction task. Framing, Windowing, DFTH, Mel filter bank algorithm and computing the inverse of DFT are the MFCC steps.



3. Hidden Markov Model(HMM): The model comprises of two stochastic inter-related processes similar to Markov Chain. The only difference is that, output symbol and the transitions are probabilistic. Every HMM state comprises of output symbols set known as output probabilities with finite number of states Q= {q1, q2…., qn}. One process is related to the transitions among the states which are controlled by a set of probabilities named as transition probabilities which are used to model the temporal variability of speech. Other process is concerned with the state output observations O= {o1, o2, ..., on} regulated by Gaussian mixture distributions bj(ot) where 1<=j<=N, to simulate the spectral variability of speech. Another possibility that every sequence of states that has the same length as the symbol sequence is also evident. The word 'Hidden' in the model name signifies the denial of viewership of the sequence of states from the observer. The Markov nature of the HMM i.e. the probability of being in a state is dependent only on the previous sate, admits use of the Viterbi algorithm to generate the given sequence symbols, without having to search all possible sequences. At a particular instance, one process is assumed to be in some state while the observation is formed by the other process representing the same current state. aij = P [Qt+1=j| Qt=i] represents the underlying Markov chain that alters states in accordance to its transition from state I to state j.

4. Dynamic time warping(DTW): It is a master technique to find an optimal alignment between two given time dependent sequences under certain restricted conditions. Matching of the two sequences is performing in a non-linear fashion by warping them intuitively. Normally different speech patterns are compared with the help of DTW. In fields such as data mining and information retrieval, DTW has been successfully applied to automatically cope with time deformations and different speeds associated with time-dependent data.

The objective of DTW is to compare two (time-dependent) sequences X: = (x1, x2..., xN) of length N ∈ N and Y: = (y1, y2..., yM) of length M ∈ N. These sequences may be discrete signals (time-series) or, more generally, feature sequences sampled at equidistant points in time. In the following, we fix a feature space denoted by F. Then xn, ym ∈ F for n ∈ [1: N] and m ∈ [1: M]. To compare two different features x, y ∈ F, one needs a local cost measure, sometimes also referred to as local distance measure, which is defined to be a function.

One of the most primitive approaches in an isolated speech recognition system was to compare the incoming speech with the prototypical version of the word present within the vocabulary, and considering the closest match for the next step. A number of issues arise on factors considering the form of templates and how are they compared to the incoming signals. In order to avoid these problems, feature vectors can be used. The problem with this approach is that if a constant window spacing is used, the lengths of the input and stored sequences is unlikely to be the same. Moreover, within a word, there will be variation in the length of individual phonemes: Cassimy might be uttered with a long /A/ and short final /i/ or with a short /A/ and long /i/. The matching process needs to compensate for length differences and take account of the non-linear nature of the length differences within the words.

The Dynamic Time warping algorithm eliminates this issue by finding an optimal match between two sequences of feature vectors which allows compressed and stretched sections of the sequence. Vector quantization technique performs the process of modelling probability density functions by distribution of prototype vectors.

## V.APPLICATIONS

1. Agricultural development: The rural population of India can search for better solutions for improving the growth of crops and can also get information about new and efficacious techniques for growing particular crops. They can easily search for information using voice recognition which makes it much easier and user-friendly for them.

2. Source of learning for the blind people: People who cannot see can use voice recognition feature for performing their tasks and for searching relevant information, all this in Hindi. It will be convenient for them to operate the device using voice and language will not be a barrier for Indians as Hindi language is known to most of the people in our country.

3. Travel: If a person who is comfortable with Hindi and not articulate with his diction in English is travelling to another country and wants to communicate with the aborigine of that place or read a board written in vernacular language, he can use our system and get his answer in Hindi.

4. Source of communication: People can communicate with each other using voice SMSs or calls as directed to the system. They can convert voice-to-text and even email the messages. It will be beneficial for people who cannot read or write to communicate vocally.

5. Education: This system can also be used in schools located in remote rural areas for educational purposes. Both teachers and students can benefit from this system.

## VI.CONCLUSIONS

In this paper, development of an effective Speech to text conversion system was limned along with the techniques used for development. This system particularly focuses on Hindi language. The main techniques in our system are chosen after an exhaustive research and comparison with other available methods and algorithm. Taking into consideration every known technique that are used in currently existing systems, we came up with the most effective ones for our system which include HMM, MFCC and DTW. Speech recognition is one of the most integrating areas of machine learning. Abridging the language barrier and increasing the human-computer interaction for people with less or no knowledge of English language will be our main goal.

## VII.ACKNOWLEDGEMENT

We are thankful to our college, Vivekanand Education Society's Institute of Technology, for considering our project and providing us help at every stage of the research. A special thanks to Ms. Kajal Jewani for mentoring and guiding us in the right direction. We would like to thank Dr.(Mrs.) Nupur Giri and our principal Dr. (Mrs.) J.M. Nair for giving this valuable opportunity. We are immensely grateful to everyone for their help with the research without which it would have been difficult to get our job done. Also, thanks to our families for the moral support and encouragement. It is a great pleasure to acknowledge the help we received from the Department of Computer Engineering.

**REFERENCES**

[1] Preeti Saini, Parneet Kaur, Automatic Speech Recognition: A Review, in International Journal of Engineering Trends and Technology- Volume4Issue2- 2013

[2] Suma Swamy, K.V. Ramakrishnan, An Efficient Speech Recognition System, in An International Journal (CSEIJ), Vol. 3, No. 4, August 2013.