INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 6)
Available online at: www.ijariit.com

Automatic speech recognition system

Ishrat Sultana
isultana2012@gmail.com
Adesh Institute of Technology, Gharuan, Punjab

Nirmaljeet Kaur Pannu
pannu_abc@yahoo.com
Adesh Institute of Technology, Gharuan, Punjab

ABSTRACT

The Speech recognition systems are difficult and challenging as the same word can be spelled in many ways also change in accent has a huge impact on the accuracy of speech recognition. Most of the ASR systems in use today are designed to recognize speech in English. The major objective of this research is to design an ASR system which recognizes discrete words from Hindi language & controls some action depending on the provided input. The input voice is captured using a microphone which is then preprocessed using several algorithms like Dynamic Time Wrapping (DTW), Hidden Markov Model (HMM) etc. This paper aims at developing a simplified technique for recognition of speech spoken in the Hindi language by first modeling the system on computer-based design and then deploying it on an embedded system.

Keywords— *Word analysis, Feature extraction, Case study*

1. INTRODUCTION

A speech recognition system consists of a microphone, for the person to speak into; speech recognition software; a computer to take and interpret the speech; a good quality soundcard for input and/or output; a proper and good pronunciation. The task of this paper is to get a computer to understand spoken language. By "understand" we mean to react appropriately and convert the input speech into another medium e.g. text [1]. The applications of speech processing are mostly useful in day to day life of people. Some of the speech processing applications are Speech Coding, Text-to-Speech Synthesis, Speech Recognition, Speaker Recognition and Verification, Speech Enhancement, Speech Segmentation and Labeling (Transcription), Language Identification, Prosody, Attitude and Emotion recognition, Audio-Visual Signal Processing and Spoken Dialog Systems [2].

2. PROPOSED APPROACH

The major purpose of this system is to develop an Automatic speech recognition system which takes a voice as input in the Hindi language. The primary hardware requirements are a good microphone, a processor running at 160 MHz figure1 shows a typical ASR block diagram which consists of sound recorder, word boundary detection, feature extraction, recognition component and Acoustic Model [4].

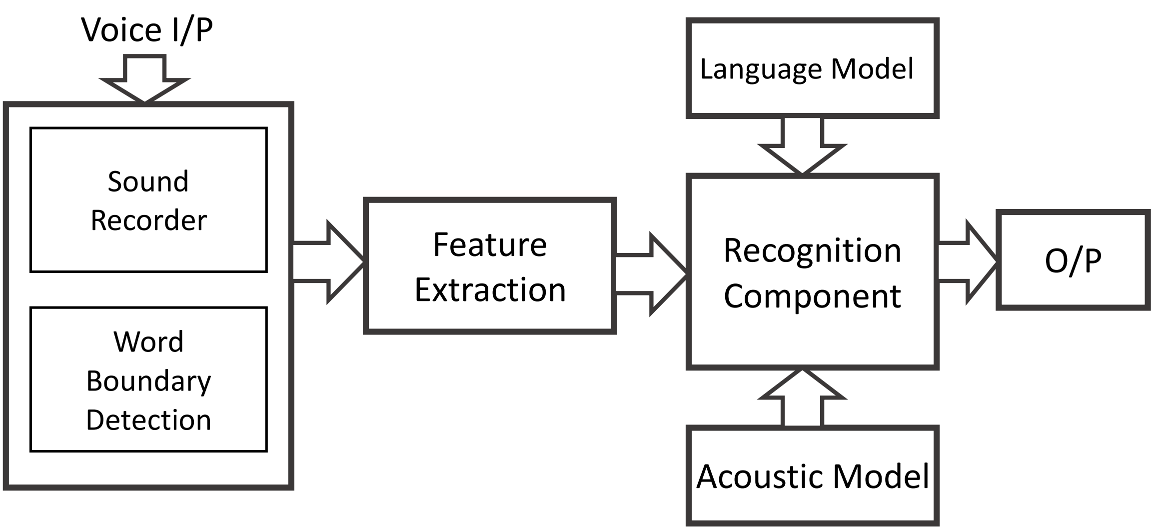

Fig. 1: ASR block diagram

2.1.1 Word Boundary Detection: This block differentiates the words to be recognized from rest of the sounds i.e. noise and determines the start and end points of the spoken word for further processing.

2.1.2 Feature Extraction: This block extracts parameters from the obtained speech signal that are unique to that signal and which differentiates the captured word from the rest of the words. It enables a compact representation of the speech waveform. In speaker-independent speech recognition, a premium is placed on extracting features that are somewhat invariant to changes in the speaker. In this step, we will use the Mel Frequency Cepstral Coefficient (MFCC) [3] method which will generate acoustic vector corresponding to the input utterance.

2.1.3 Recognition Component: It is a statistical method of matching the features extracted from sample speech to the features of predefined speech models, to find the best match and recognize the word spoken. We can use algorithms like Hidden Markov Model (HMM) or Dynamic Time Warping (DTW) for the recognition purpose. The input to this step will be the acoustic vector generated by the feature extraction step [4].

2.1.4 Acoustic Model: The acoustic information and phonetics are established here. Acoustic Model has a representation of how a word sound. Recognition system makes use of this model while recognizing the sound signal [5].

2.2 Hindi word samples analysis

The samples for different Hindi words from the same user were recorded and analyzed using MATLAB tool. The samples were recorded for the duration of 1.5 seconds with a sampling frequency of 44.1 kHz producing an array of 66150 samples. The amplitudes of the samples were plotted against time to obtain a time domain plot. The results are as follows:

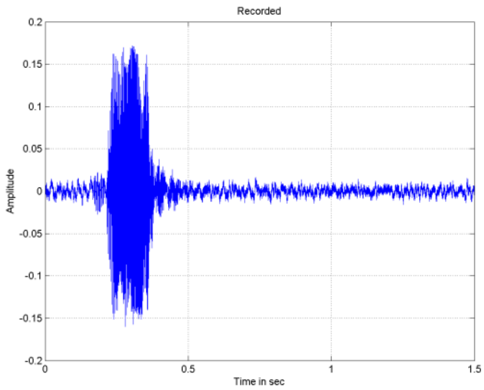

Fig. 2: Time domain plot for the word "Do" from sample 1

It can be observed that the time domain plot for the same words is almost similar. The amplitudes of samples are high for the duration for which the speaker utters the word, this indicates a high energy part of the voice sample. The values of amplitudes for the silent part of voice sample are non-zero due to background noise. This noise may interfere with the correct word recognition, hence must be eliminated.

2.3 Word boundary detection

Word boundary detection is an important step before feature extraction and further analysis of voice signal. The start and end of the words in a discrete speech are detected and words are then separated from the original array for further analysis [3]. Thus, the array size is reduced by eliminating silences in speech. The word boundary detection of the Hindi word voice samples recorded was done using MATLAB tool. The step in the operation is:

(a) Obtaining time domain plot for a voice sample
(b) Taking the absolute value of array obtained
(c) Assuming silence for first few ms, taking the average value of the first 10 samples and subtracting this average value from samples with amplitude levels less than average value repeatedly.
(d) Eliminating noise, obtaining word start and end array indices
(e) Extracting word from an original array using these indices.

The stepwise results for Hindi voice sample of the word "Do" are as shown in the figure.

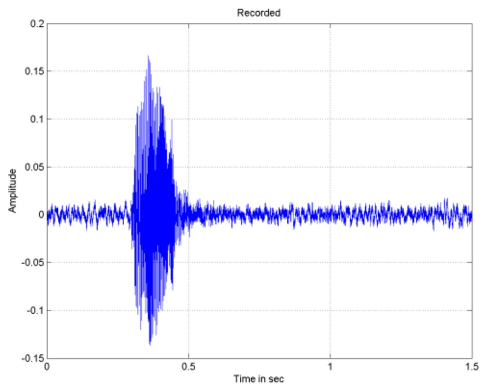

Fig. 3: Result of step a for Hindi word "Do"

Different voice samples were used and analyzed using DTW and MFCC algorithms and each sample was compared for effective contribution in the analysis.

3. CONCLUSIONS

In this paper, we conclude that the ASR system is supposed to be user independent and the same word uttered by different speakers can have considerable variations in amplitudes. Also, the varying gap between the utterances has a considerable impact on the system's performance. One of the ways to make system user independent is by using methods like Vocal Tract Normalization. Actual word can be distinguished from background noise by using techniques like Word Boundary Detection. Matching two signals in the time domain is a very difficult task. To address this problem, frequency domain analysis is used. After MFCC, various methods like Dynamic Time Warping (DTW), Hidden Markov Model (HMM) can be used for actual speech recognition.

4. REFERENCES

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

[2] Karpagavalli and Chandra, A Review on Automatic Speech Recognition Architecture and Approaches, International Journal of Signal Processing, Image Processing, and Pattern Recognition Vol.9, No.4, (2016), pp.393-404

[3] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, March 2010, ISSN 2151-9617

[4] Anjali Bala, Abhijeet Kumar, Nidhika Birla, Voice command recognition system based on MFCC AND DTW, International Journal of Engineering Science and Technology, Vol. 2 (12), 2010, 7335-7342

[5] Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit, Isolated speech recognition using MFCC and DTW, ISSN (Print): 2320 – 3765, ISSN (Online): 2278 – 8875

[6] Sadaoki Furui, Automatic speech recognition and its application to information extraction, 2-12-1, Ookayama, Meguro-Ku, Tokyo, 152-8552 Japan

