A survey on Guajarati language recognition

[1]Ms.Puspa Machhar, [2] Mr.Dipak Agrawal
[1]ME Student,CE Department, [2] Assistant Professor,CE Department
[1,2]SIE,Vadodara,GTU,Gujarat

Abstract—Speech recognition deals with identifying the spoken words and converting it into equivalent text form. Many application use speech recognition such as direct voice input in aircraft, data entry, speech-to-text processing, voice user interfaces such as voice dialing and many more. Many hand-held devices support voice commands for operating the devices but many few of them supports the facility in local languages. The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. The objective of this review paper is to summarize and compare some of the well-known methods used in various stages of speech recognition system and identify research topics.

Keywords—Speech Recognition,Hidden MarkovModel (HMM), feature extraction.

INTRODUCTION

Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing[2]. Speech to text conversion is the ability of a machine to recognize speech sound and convert in to text sequence as close as possible[1]. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages. The main goal of speech recognition area is to develop techniques and systems for speech input to machine[2]. Speech Recognition approach is used extensively to solve real-world challenges. So many factors affect accuracy and performance of speech recognition system. Due to different grammar rules, noisy environment, and pronunciations of speaker independent speech recognition system is a challenging job[3].

Fig.1 shows basic representation of speech recognitionsystem in simple equation which contains featureextraction, database, network training and testing ordecoding. The recognition process is shown below (Fig.1)[8].

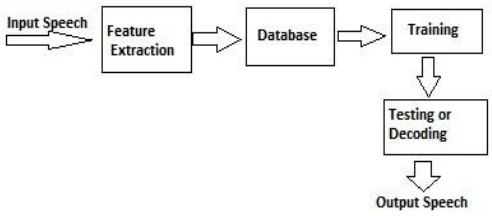

Fig.1 Speech Recognition Diagram

Types of Speech

Speech recognition system can be separated in different classes by describing what type of ullerances they can recognize [4].

a. Isolated Word

Isolated word recognizes attain usually require each utterance to have quiet on both side of sample windows. It accepts single words or single utterances at a time .This is having "Listen and Non Listen state". Isolated utterance might be better name of thisclass [4].

b. Connected Word

Connected word system are similar to isolated words but allow separate utterance to be "run together minimum pause between them[4].

c. Continuous speech

Continuous speech recognizers allows user to speak almost naturally, while the computer determine the contnt.Recognizer with continues speech capabilities are some of the most difficult to create because they utilize special method to determine utterance boundaries[4].

d. Spontaneous speech

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed .an ASR System with spontaneous speech ability should be able to handle a variety of natural speech feature such as words being runtogether[4].

FEATURE EXTRACTION METHODS

Features extraction in ASR is the computation of a sequence of feature vectors which provides a compact representation of the given speech signal. It is usually performed in three main stages. The first stage is called the speech analysis or the acoustic front-end, which performs spectra-temporal analysis of the speech signal and generates raw features describing the envelope of the power spectrum of short speech intervals. The second stage compiles an extended feature vector composed of static and dynamic features. Finally, the last stage transforms these extended feature vectors into more compact and robust vectors that are then supplied to the recognizer [5].

1. Mel-Frequency Cepstral Coefficients (MFCC)

The most prevalent and dominant method used to extract spectral features is calculating Mel-Frequency Cepstral Coefficients (MFCC). MFCCs are one of the most popular

feature extraction techniques used in speech recognition based on frequency domain using the Mel scale which is based on the human ear scale. MFCCs being considered as frequency domain features are much more accurate than time domain features[5].

The Mel-frequency Cepstrum Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. The MFCC are based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics ofspeech. The signal is divided into overlapping frames to compute MFCC coefficients. Let each frame consist of N samples and let adjacent frames be separated by M samples where M<N. Each frame is multiplied by a Hamming window where the Hamming window equation is given by:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2n\pi}{N-1}\right)$$

In the third step, the signal is converted from time domain to frequency domain by subjecting it to Fourier Transform. The Discrete Fourier Transform (DFT) of a signal is defined by the following:

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

In the next step the frequency domain signal is converted to Mel frequency scale, which is more appropriate for humanhearing and perceptions. This is done by a set of triangular filters that are used to compute a weighted sum of spectral components so that the output of the process approximates a Mel scale. Each filter's magnitude frequency response is triangular in shape and equal to unity at the centre frequency and decrease linearly to zero at centre frequency of two adjacent filters. The following equation is used to calculate the Mel for a given frequency:

$$M = 2595log10(1 + \frac{f}{700})$$

In the next step the log Mel scale spectrum is converted to time domain using Discrete Cosine Transform (DCT). DCTis defined by the following, where is a constant dependent on N:

$$X_k = \propto \sum_{i=0}^{N-1} \left(x_i \cos\left\{\frac{(2i+1)\pi k}{2N}\right\}\right)$$

The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficients is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vectors.

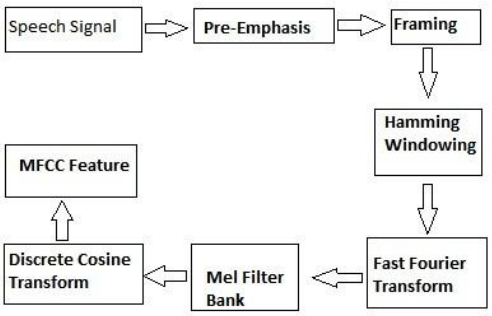

Fig.2 Block Diagram of MFCC Feature Extraction Techniques

A block diagram of the MFCC processes is shown in Figure. Block diagram of MFCC The speech waveform is cropped to remove silenceor acoustical interference that may be present in the beginning or end of the sound file. The windowing block minimizes the discontinuities of the signal by tapering the beginning and end of each frame to zero. The FFT block converts each frame from the time domain to the frequency domain. In the Mel-frequency wrapping block, the signal is plotted against the Mel spectrum to mimic human hearing. In the final step, the Cepstrum, the Mel -spectrum scale is converted back to standard frequency scale. This spectrum provides a good representation of the spectral properties of the signal which is key for representing and recognizing characteristics of the speaker[10].

2.Linear Predictive Codes (LPC)

It is desirable to compress signal for efficient transmission and storage. Digital signal is compressed before transmission for efficient utilization of channels on wireless media. For medium or low bit rate coder, LPC is most widely used. The LPC calculates a power spectrum of the signal. It is used for formant analysis. LPC is one of the most powerful speech analysis techniques and it has gained popularity as a formant estimation technique [5].

3. Relative Spectra processing (RASTA):

Hajer Rahali is used RASTA method to extract the relevant information from the audio signal. The main goal of the work is to improve the robustness of speech recognition system in additive noise and real time reverberant environment .Hynek Hermansky et. al. discussed relationship with human auditory perception system and extend the original method to the combination of additive noise and convolution noise. He used band pass filter of time trajectories of logarithmic parameter of speech[10].

4.Perceptual Linear prediction (PLP):

The Perceptual Linear Prediction PLP model developed by Hermansky. PLP models the human speech based on the concept of psychophysics of hearing [2, 9]. PLP discardsirrelevant information of the speech and thus improves speechrecognition rate. PLP is identical to LPC except that its spectral characteristics have been transformed to match characteristics of human auditory system[5].

Method	Property	Advantage	Disadvantage
Linear predictive code	10 to 16 lower sequence coefficient, Static feature extraction	Spectral analysis is done with a fixed resolution along a	Frequencies are weighted equally on a linear scale while the frequency

	method	subjective frequency scale i.e. Mel frequency scale	sensitivity of the human ear is close to the logarithmic
Mel-frequency Cepstrum Coefficients (MFCC)	Power spectrum is computed by implementing Fourier Analysis	This method is used for find our features.	MFCC values are not very robust in the presence of additive noises it is common to normalize their values in speech recognition system to reduce the influence of noise.
RASTA Filtering	For Noisy speech	It find out feature in noisy data	It increases the dependence of the data on its previous context.

Hidden Markov Models (HMM)

HMM is doubly stochastic process with an underlying stochastic process that is not observable, but can only be observed through another set of stochastic processes that produce sequence of observed symbols. The basic theory behind the Hidden Markov Models (HMM) dates back to the late 1900s when Russian statistician Andrej Markov first presented Markov chains. Baum and his colleagues introduced the Hidden Markov Model as an extension to the first-order stochastic Markov process and developed an efficient method for optimizing the HMM parameter estimation in the late 1960s and early 1970s. Baker at Carnegie Mellon University and Jelinek at IBM provided the first HMM implementations to speech processing applications in the 1970s. Proper credit should also be given to Jank ferguson at the Institute for defense Analysis for explaining the theoretical aspects of three central problems associated with HMMs, which will be further discussed in the following sections. The technique of HMM has been broadly accepted in today's modern state-or-the art ASR systems mainly for two reasons: its capability to model the non-linear dependencies of each speech unit on the adjacent units and a powerful set of analytical approaches provided for estimating model parameters[7].

Artificial neural network

Artificial neural network provides fantastic imitation of information processing analogues to human nervous system. The common choice of classification and pattern recognition is used as Multilayer Feed Forward Back propagation method in Neural Network . The Audio- Visual Speech Recognizer (AVSR) used is based On HMMs process and was trained for huge vocabulary continuous speech a Neural Network Genetic Algorithm can be used with neural network for performance improvement by optimizing parameter combination[9].

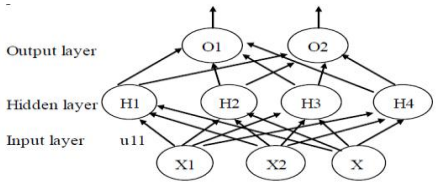

Fig.3 Artificial Neural Network

But in this paper multi-layer feed forward back propagation neural network as shown in Figure 5 with total number of features as number of input neurons in input layer for LPC, PLP and MFCC parameters respectively. As shown in Figure Neural Network consists of input layer, hidden layer and output layer. Variable number of hidden layer neurons can be tested for best results we can train network for different combinations of epochs with target as least amount error rate[9].

Conclusion

After reading literature review number of techniques and method, model available for real time speech processing and recognition. HMM, MFCC widely use and available for speech recognition for normal Gujarati word recognition but not for tricky word recognition. so using hybrid approach for real time speech processing is done and also recognize tricky word using matlab 14.And also maintain max accuracy for Gujarati word.

REFERENCES

[1].Ankit Kuamr, Mohit Dua, Tripti Choudhary," Continuous Hindi Speech Recognition Using Gaussian Mixture HMM" ,IEEE Students' Conference on Electrical, Electronics and Computer Science,2014.

[2].Miss Himanshu,Sarbjit Kaur, Vikas Chaudhary ," Literature Survey on Automatic Speech Recognition System" ,International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, July 2014.

[3].Jinal H. Tailor, Dipti B. Shah," Review on Speech Recognition System for Indian Languages", International Journal of Computer Applications (0975 – 8887) Volume 119 – No.2, June 2015.

[4].Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar," A Review on Speech Recognition Technique", International Journal of Computer Applications (0975 – 8887),Volume 10– No.3, November 2010.

[5].Namrata Dave," Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition",INTERNATIONAL

JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY, Volume 1, Issue VI, July 2013.

[6]. Bhoomika Dave, D. S. Pipalia,"An Approach to Increase Word Recognition Accuracy in Gujarati Language",International Journal of Innovative Research in Computer and Communication Engineering,Vol. 3, Issue 7, July 2015

[7].Bhupinder Singh, Neha Kapur, Puneet Kaur," Speech Recognition with Hidden Markov Model: A Review", International Journal of Advanced Research in Computer Science andSoftware Engineering,Volume 2, Issue 3, March 2012

[8].Nidhi Desai, Prof.Kinnal Dhameliya, Prof.Vijayendra Desai," Feature Extraction and Classification Techniques for Speech Recognition: A Review",International Journal of Emerging Technology and Advanced Engineering,Volume 3, Issue 12, December 2013.

[9].Dr E.Chandra, K.Manikandan, M. Sivasankar," A Proportional Study on Feature Extraction Method in Automatic Speech Recognition System",INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL,ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 1, January 2014

[10]Pratik K. Kurzekar, Ratnadeep R. Deshmukh , Vishal B. Waghmare , Pukhraj P. Shrishrimal ,"A Comparative Study of Feature Extraction Techniques for Speech Recognition System",International Journal of Innovative Research in Science,Engineering and Technology,Vol. 3, Issue 12, December 2014.