HMM Based Gujarati Tricky Words Recognition

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

ABSTRACT

Many researcher are working to make computer to understand naturally spoken language. For international language like English this technology has grown to a matured level. we present a model which recognize Gujarati spoken by human and convert it into text. The aim is recognition Guajarati tricky words. In this dissertation Work, proposed method extracts by speech using MFCC feature extraction technique and HMM model. The recognizer is work in abundance of three essential structure squares to be specific Feature extraction, Training and Testing(Recognition).The plan proposed here executes the Mel recurrence Cepstral Coefficient(MFCC) with a specific end goal to figure the otherworldly components of the discourse signal. Utilizing HMM (Hidden Markov Model) to perceive discourse test to give fantastic results for secluded words. It comprises of detached words that are isolated by quiets. This proposed system provide high accuracy for Guajarati word.

Keywords : *Acoustic Model, Hidden Markov Model, Gujarati, Speech-To-Text*

---

1. INTRODUCTION

Speech basic terms of human communication[4]. Speech recognition is a software engineering terms and it is otherwise called programmed discourse recognition(ASR).It is convert speech into text[1].Speech is most natural and efficient form of exchanging information among human[2].Speech recognition is multileveled pattern recognition technique. In speech recognition acoustical signal are examined and structured into a hierarchical form of words, phonemes, phrases and sentences[3].In speech recognition each level may provide additional temporal constraints, for example word pronunciations or legal word sequences, which can compensate for errors or uncertainties of lower levels[3].It is recognition process by computer what a human said[4].Speech-to-text technique is new research idea to help handicap people with the voice promoted writing tools[5].Speech recognition is most general form for acoustic waveform to a written equivalent of the message information[5]. It is helpful framework for individuals who may experience the ill effects of handicaps that influence their composition capacity however can utilize their discourse to make content on computer[1].It is overall advantage is manage the time[1].Some people speech faster then write or typing making fewer mistek[1]. Keybord is a prominent medium for writing however it is not exceptionally convenient,as it requires a specific measure of expertise for powerful use [4].A mouse require good eye and hand co-ordination[4].Difficult to use that physically challenge people[4].All the constrains have deleted.Speech interface solve these problem[4].Speech recognition software only identified those words or phrases which is spoken clearly or noiseless environment[5].

2. BLOCK DIAGRAM OF STT

A typical Speech-To-Text system incorporates different phases that are speech pre-processing, feature extraction, acoustic analysis, modeling, decoding and filtering as shown in Fig. 1. The speech is used as system input which passes through phases and converts it into text format.

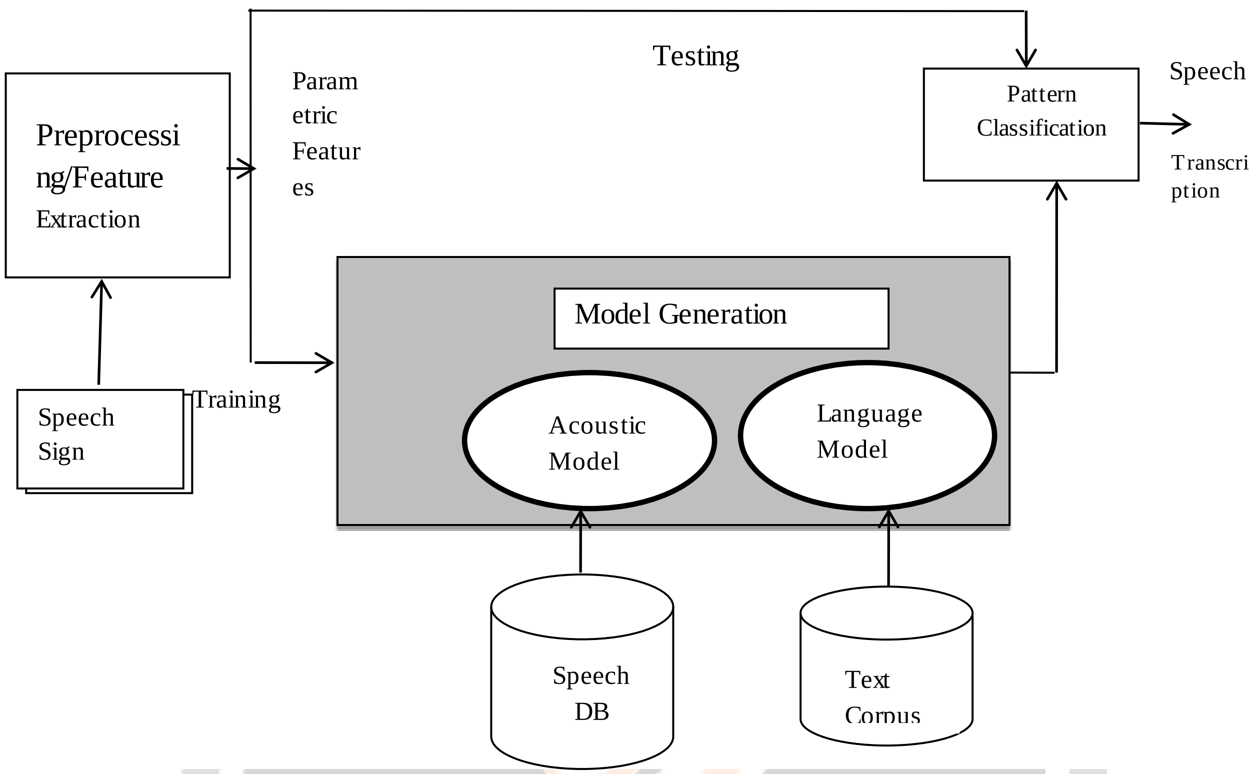


Fig.1.1 System Architecture of for Automatic Speech Recognition System

After decoding phase, filtering phase extract recognized word from the output transcription and represent in text format

ARCHITECTURE OF STT

A detailed STT architecture as shown in include eight phases that are
1) Speech Pre-Processing

2) Acoustic Analysis – feature extraction
3) Acoustic Model Generation – Re-Estimation

4) Language Model – Parsing

5) Pronunciation Model

6) Decoding

7) Filtering Methods

Speech Pre-Processing
Through the recording tool input speech data is stored in .wav format. Quality of recorded data is depending upon recording device and speed. Segmentation is used to divide raw data speech signal into evenly spaced frames. The frame size is selected in smaller size to capture rapid transitions and achieve sufficient resolution in frequency domain. The frame size can vary from 10 milliseconds to 25 milliseconds. Segmented frames are analyzed to produce feature vector which is useful for speech sound classification. The required measurement parameters for feature analysis are recording channel, environmental noise and speaker variability.
Feature Extraction
Feature extraction step finds the set of parameters of articulations that have acoustic relationship with speech signals and these parameters are computed through indulgence of the acoustic waveform. These parameters are recognized

as features. The main focal point of feature extractor is to keep the appropriate information and remove inappropriate one. To execute this process, feature extractor divides the acoustic signal into 10-25 ms. Data acquire in these frames is multiply by window task. There are various types of window task that can be utilized such as "hamming Rectangular, Blackman, Welch or Gaussian etc". Thusly includes have been separated from each fram. There are few techniques for feature extraction such as "Mel-Frequency Cepstral Coefficient (MFCC),Linear Predictive Cepstral Coefficient (LPCC), Perceptual Linear Prediction (PLP), wavelet and RASTA-PLP (Relative Spectral Transform)" Processing and so forth.[5]

Acoustic Model Generation – Re-estimation

Acoustic Modeling is the major a portion of ASR framework. In acoustic displaying, the association between the acoustic data and phonetics is set up. Acoustic model plays imperative part in execution of the system and in charge of computational burden. Preparing sets up co-connection between the fundamental speech units and the acoustic perceptions. Preparing of the framework requires making an pattern delegate for the components of class utilizing one or more pattern that relate to speech of the same class. Numerous models are accessible for acoustic modeling out of them Hidden Markov Model (HMM) is broadly utilized and accepted as it is efficient algorithm for training and recognition[12].

Language Modeling

A language model contains the structural accessible in the language to produce the probabilities of event. These instigates the likelihood of a word event after a word arrangement. For the most part Speech recognition frameworks utilizes bi-gram, tri-gram, n-gram dialect models for finding right word arrangement by anticipating the probability of the nth word, utilizing the n-1 prior words. In discourse acknowledgment, the PC framework matches sounds with word succession. The "language model" distinguish word and expression that has comparable sound. For instance, in American English, the expressions similar to recognize speech and wreck a nice beach have similar elocution other than denote extremely dissimilar things. These ambiguity are simpler to determine when confirmation as of the language model is integrated with the pronunciation model and the acoustic model[5]

Pronunciation Model

Pronunciation model is used to develop correspondence between different HMMs to form model for each input sentence. The concurrency is checked between HMM name and variable specified in language model.

Decoding – in pronunciation model word models are created by concatenating sub word models which are composed into a decoding networkDecoding network generates transcription file after processing HMM definition and task network files.

Filtering Methods

Filtering method use to extract recognized word from output transcription file and convert into relevant text format.

PERFORMANCE PARAMETERS

Word Recognition Rate (WR) = (N – D – S / N) / 100

WER is the common measure of speech recognition system performance. Through the power low correlation is measured between recognized words with spoken words from transcription file.

WER = (S + D + I / N) * 100

Where S = Number of Substitutions

D = Number of Deletion

I = Number of Insertion

N = No of Words in Reference

RESULT

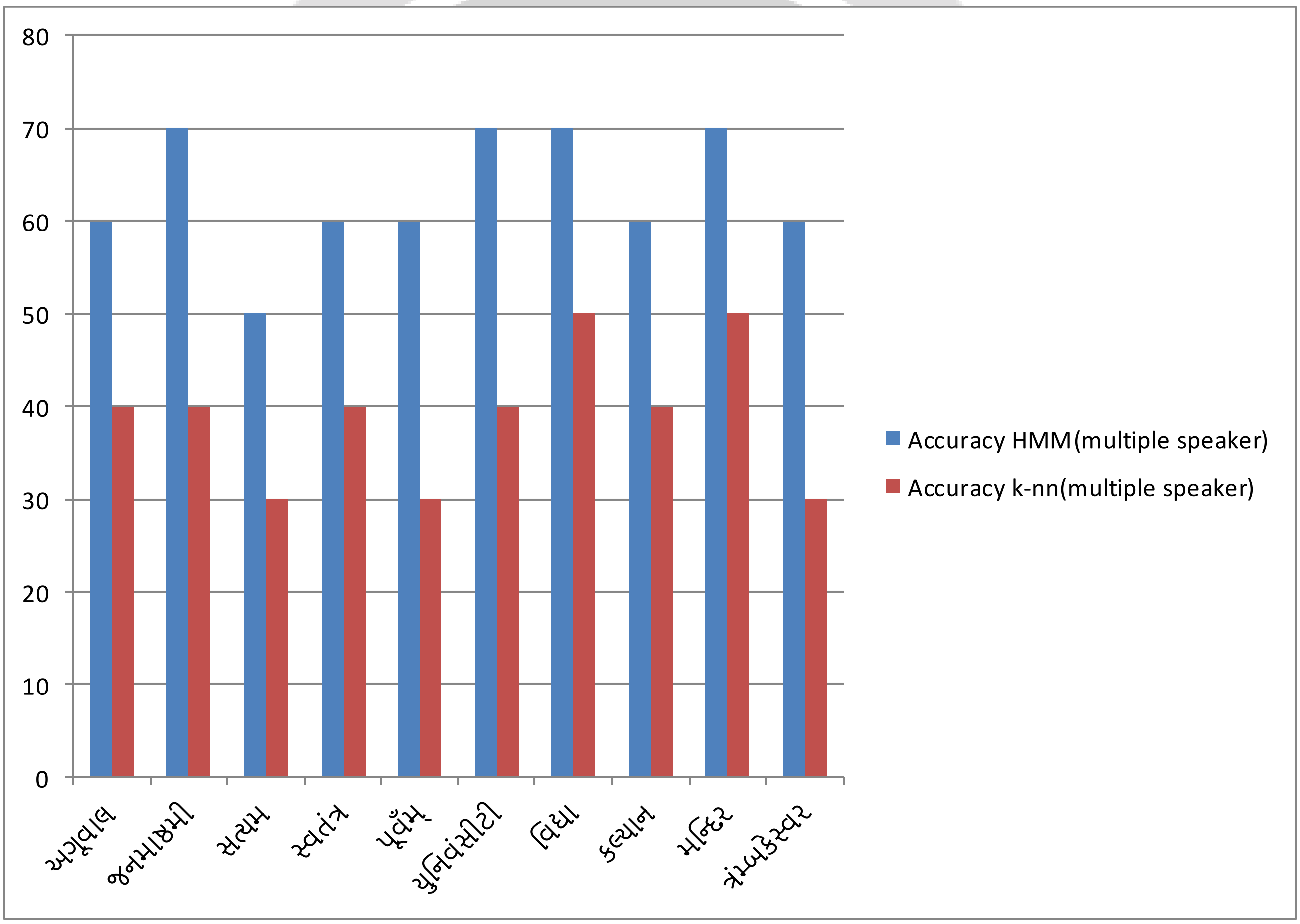
experiment is to find tricky gujarati words accuracy by test deta set multiple time using HMM and k-nn classifier.

Table 1:Experiment

	Accuracy HMM(multiple speaker)	Accuracy k-nn(multiple speaker)
અગૂવાલ	60	40

જનમાષ્ઠમી	70	40
સત્યમ	50	30
સ્વતંત્ર	60	40
પૂર્વમ્	60	30
યુનિવંસીટી	70	40
વિદ્યા	70	50
કલ્યાન	60	40
મન્દિર	70	50
ત્રંમ્બકેસ્વર	60	30

Chart of measures for Accuracy single and multiple and speaker for tricky gujarati words



Conclusion

The existing model accepts gujarati numeral,words and tricky words spoken by speaker and then that spoken numeral,words,tricky words are converted into editable text by MATLAB.In Existing System use of MFCC as a feature extraction technique and K-NN(K-Nearest Neighbor) as a Classifier.HMM, MFCC is use in proposed model.Implimenting K-NN and HMM classifier and HMM give better accuracy than K-NN.

References

[1] AnchalKatyal, AmanpreetKaur, Jasmeen Gill, " Automatic Speech Recognition: A Review", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-3, Issue-3, February 2014.

[2] SantoshK.Gaikwad,BharatiW.Gawali,PravinYannawar, "A Review On Speech Recognition Technique", International Journal of Computer Applications(0975-8887) Volume 10-No3,November 2010.

[3] V.Vaidhehi1, Anusha J, Anand P, "The Comparative Study of Speech Recognition Models--Hidden Markov and Dynamic Time Wrapping Model",International Research Journal of Innovative Engineering,Volume1, Issue 3 of March 2015.

[4] Ms.Vrinda,Mr. ChanderShekhar," SPEECH RECOGNITION SYSTEM FOR ENGLISH LANGUAGE", International Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 1, January 2013.

[5] Bharat C. Patel, Apurva A. Desai,"Recognition of Spoken Gujarati Numeral and Its Conversion into Electronic Form", International Journal of Engineering Research &Technology ,Vol. 3 Issue 9, September-2014.

[6] SumanK.Saksamudre,P.P.Shrishrimal,R.R.Deshmukh, "A Review on Different Approaches for Speech Recognition System",International Journal of Computer Application,Volume 115-No.22,April 2015

[7] IshanBhardwaj, Narendra D Londhe," Hidden Markov Model Based Isolated Hindi Word Recognition", 2012 2nd International Conference on Power, Control and Embedded Systems.