Hybrid Approach for Real Time Tricky Gujarati Word Recognition

Shruti R. Kandoriya[1], Shital Patel[2]
[1]Student L.C. Institute of Technology, Gujarat, India, [2]Assistant Prof. L.C. Institute of Technology, Gujarat, India
[1] Computer Engineering Department
[1] L.C. Institute of Technology, Gujarat, India,

Abstract: *Many researchers 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 recognizes Gujarati spoken by human and convert it into text. The aim is recognition of the Gujarati tricky words. In this dissertation Work, I have proposed a method extracts by speech using Mel recurrence Cepstral Coefficient (MFCC) feature extraction technique and HMM model. The recognizer is working in abundance of three essential structure squares to be specific Feature extraction, Training and Testing (Recognition). The proposed here executes the Mel recurrence Cepstral Coefficient (MFCC) with a specific end goal to figure the otherworldly components of the discourse signal Utilizing Support Vector Machine (SVM) to perceive discourse test to give fantastic results for secluded words. It comprises of detached words that are isolated by quiets. This proposed system provides high accuracy for Gujarati language.*

Keywords: *Artificial Intelligence, Pre-processing, MFCC, HMM, ANN.*

I. INTRODUCTION

AUTOMATIC SPEECH RECOGNITION (ASR)

Automatic speech recognition (ASR) can be defined as the independent, computer - driven transcription of spoken language into readable text in real time. ASR is technology that allows a computer to identify the words that a person speaks into a microphone or telephone and convert it to written text. Having a machine to understand fluently spoken speech has driven speech research for more than 50 years. Although ASR technology is not yet at the point where machines understand all speech, in any acoustic environment, or by any person, it is used on a day - to - day basis in a number of applications and services. The ultimate goal of ASR research is to allow a computer to recognize in real - time, with 100% accuracy, all words that are intelligibly spoken by any person, independent of vocabulary size, noise, speaker characteristics or accent. Today, if the system is trained to learn an individual speaker's voice, then much larger vocabularies are possible and accuracy can be greater than 90%.

HISTORY OF ASR

The earliest attempts to devise systems for automatic speech recognition by machine were made in the 1950s. Much of the early research leading to the development of speech activation and recognition technology was funded by the National Science Foundation (NSF) and the Defence Department's Defence Advanced Research Projects Agency (DARPA). Much of the initial research, performed with NSA and NSF funding, was conducted in the 1980s. (Source: Global Security.Org) Speech recognition technology was designed initially for individuals in the disability community. For example, voice recognition can help people with musculoskeletal disabilities caused by multiple sclerosis, cerebral palsy, or arthritis achieves maximum productivity on computers. During the early 1990s, tremendous market opportunities emerged for speech recognition computer technology. The early versions of these products were clunky and hard to use. The early language recognition systems had to make compromises: they were "Tuned" to be dependent on a particular speaker, or had small vocabulary, or used a very stylized and rigid syntax. However, in the computer industry, nothing stays the same for very long and by the end of the 1990s there was a whole new crop of commercial speech recognition software packages that were easier to use and more effective than their predecessors. In recent years, speech recognition technology has advanced to the point where it is used by millions of individuals to 3 automatically create documents from dictation. Medical transcriptionists listen to dictated recordings made by physicians and other health care professionals and transcribe them into medical reports, Correspondence, and other administrative material. An increasingly popular method utilizes speech recognition technology, which electronically translates sound into text and creates transcripts and drafts of reports. Transcripts and reports are then formatted; edited for mistakes in translation, punctuation, or grammar; and checked for consistency and any possible errors. Transcriptionists working in areas with standardized terminology, such as radiology or pathology, are more likely to encounter speech recognition technology. Use of speech recognition technology will become more widespread as the technology becomes more sophisticated. Some voice writers produce a transcript in real time, using computer speech recognition technology. Speech recognition-enabled voice writers pursue not only court reporting careers, but also careers as closed captioners and Internet streaming text providers or caption providers.

ASR WORKING

The goal of an ASR system is to accurately and efficiently convert a speech signal into a text message transcription of the spoken words independent of the speaker, environment or the device used to record the speech (i.e. the microphone). This process begins when a speaker decides what to say and actually speaks a sentence. (This is a sequence of words possibly with pauses, uh's, and um's.) The software then produces a speech wave form, which embodies the words of the sentence as well as the extraneous sounds and pauses in the spoken input. Next, the software attempts to decode the speech into the best estimate of the sentence. First it converts the speech signal into a sequence of vectors which are measured throughout the duration of the speech signal. Then, using a syntactic decoder it generates a valid sequence of representations.

ASR BENEFITS

There are fundamentally three major reasons why so much research and effort has gone into the problem of trying to teach machines to recognize and understand speech:

- Accessibility for the deaf and hard of hearing
- Cost reduction through automation
- Searchable text capability

DEVELOPMENTS IN ASR

Aside from the scientists and technicians who are engaged in ASR research and development, most people who think about ASR underestimate its complexity. It is more than automatic text-to-speech, ASR requires fast computers with lots of data capacity and memory--a necessary condition for complex recognition tasks, and the involvement of speech scientists, linguists, computer scientists, mathematicians, and engineers. The search is on for ASR systems that incorporate three features: large vocabularies, continuous speech capabilities, and speaker independence. Today, there are numerous systems which incorporate these combinations.

II. RELATED WORK

Sr. No.	Paper Title	Authors	Methods /Algorithms Used	Drawbacks And/or Future Scope
1	[1] Automatic Speech Recognition of Isolated Words in Hindi Language	Priyanka Wani, U. G. Patil, D.S. Bormane, S.D. Shirbahadurkar	MFCC for feature extraction and KNN with GMM for isolated word recognition.	Result is not effective its satisfactory Accuracy is 94.31.
2	[2] Development of Speech recognition technique for Marathi numerals using MFCC & LFZI algorithm	Deepali Malewadi, Gouri Ghule	MFCC and LFZI used for feature extraction and SVM covers the back-end for pattern classification.	Recognize only Marathi Numerical word, future work can be done in text.
3	[3] Isolated word Automatic Speech Recognition (ASR) System using MFCC, DTW & KNN	Muhammad Atif Imtiaz, Gulistan Raja	MFCC used for feature extraction where DTW is applied for speech feature matching. KNN is employed as a classifier.	This system is provide solution for Speaker dependent data whereas further work can be done for speaker independent

4	[4] Voice and speech recognition in Tamil language	Kiran R., Nivedha K., Pavithra Devi S., Subha T	HMM is used to create voice and speech recognition system in smart phone.	The use of these applications are limited due to language barriers that is there is no Flexibility for native users. And only used in smartphone
5	[5] MFCC based noise reduction in ASR using Kalman filtering	Anuradha P Nair, Shoba Krishnan, Zia Saquib	MFCC over LPC for speech recognition. LPC with Kalman filters have been elaborated and MFCC.	-
6	[6] Implementation and performance evaluation of continuous Hindi speech recognition	Shweta Singhal, Ajay Kumar Garg, Dr. Rajesh Kumar Dubey	MFCC in front-end while HMM backend. Once process completed MFCC replaced by PLP in front-end and same HMM replaced by GMM in back-end.	This paper tested solution for 100 words, for future increase vocabulary size.

Table 1: Related Work

III. PROPOSED SYSTEM

Pre-processing:

Pre-processing include pre-emphasis, Segmentation, Filtering, Noise Removal, Active Voice Detection.

Feature Extraction

The feature extraction module parameterizes the speech into a parsimonious sequence of feature vectors that contain the relevant information about the sounds. The most efficient parameters so far are MFCC because they approximate the human auditory system. The pattern matching algorithms are applied on extracted features to map unknown speech with known speech signals used during training phase.

Feature Matching / Pattern Recognition (Classification) Approach

Pattern recognition is the process of computing the unknown test pattern with each sound class reference pattern and computing a measure of similarity between them. After computing training of the system at the time of testing pattern are classified to recognize the speech.

Types of pattern recognition

1. Template based approach
2. Stochastic approach

Template base approach to recognition have provide family of techniques. Usually template for entire words are constructed.

- A database is created for Guajarati words having 100 words from 5 speakers for training. For testing purpose 5 observations has been carried out for five Guajarati speech sentences consisting of same words which are there in the isolated word training database.
- The observations are done for speaker dependent data as same speakers are used who have uttered for training purpose.
- In this work the system's performance for pattern matching algorithms (HMM) is compared and applied on Guajarati languages.

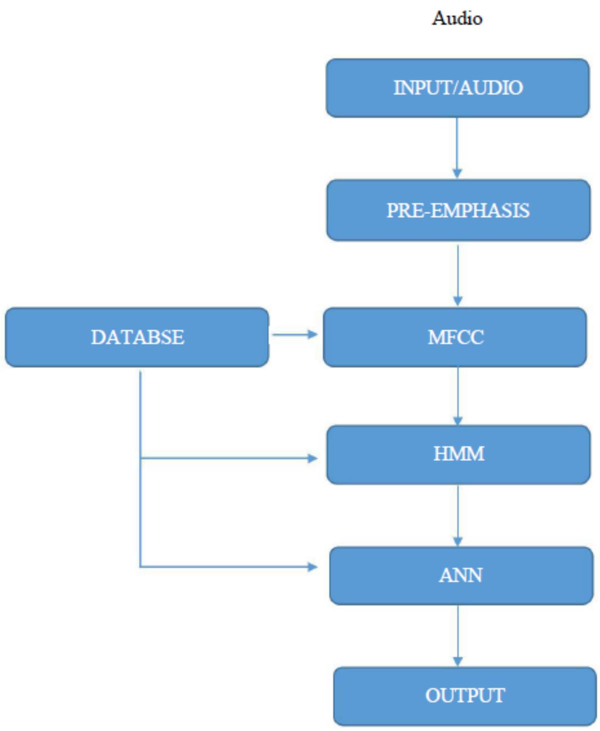

Figure 1: Proposed System Block Diagram

IV. SIMULATION ENVIROMENT AND RESULT ANALYSIS

Step: 1 - Recording input as audio file

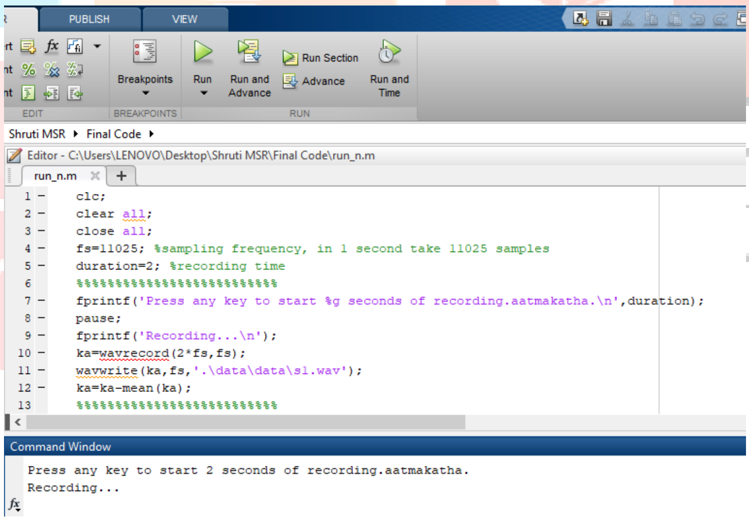

Figure 2: Screenshot of Record input data

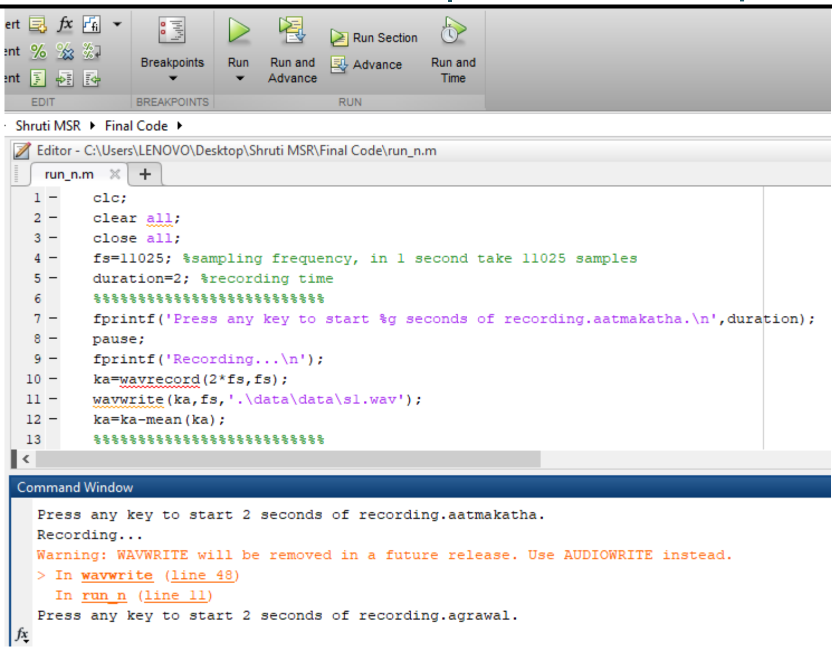

Figure 3: Screenshot of saved input data in current directory

Step: 2 - Feature Extraction will run in input of audio file. For testing data enter user specific voice

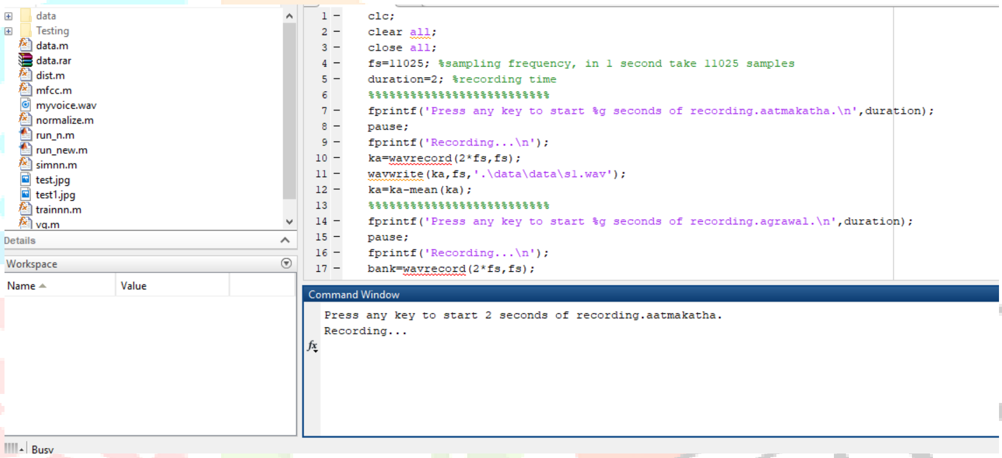

Figure 4: Screenshot of Feature Extraction

Step: 3 - Compare with all training sample and allow sample at a time on basis of MFCC. It will compare with all sample and allow a minimum distance base voice signal (minimum frequency difference). And show match wave with minimum distance.

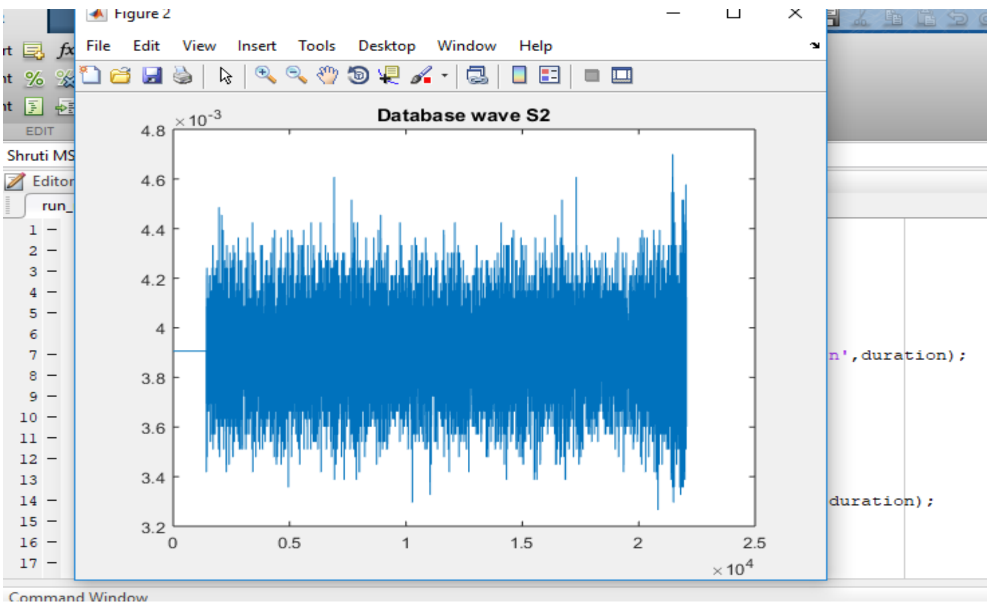

Figure 5: Screenshot of Sample Comparison

Step: 4 - For text data it will call java file and open and display Guajarati word.

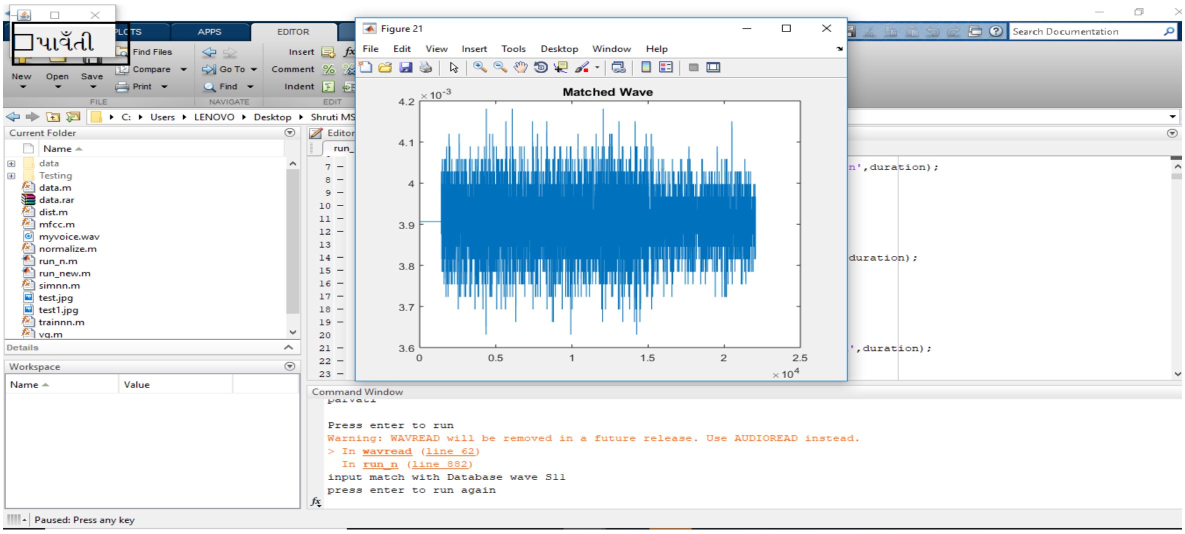

Figure 5: Screenshot of Match Frequency

V. CONCLUSION

Being a Computer Engineering student I always try to make and/or develop a system which can be useful to the society in general. Till date the Automatic Speech Recognition was limited to English and few other Indian Languages. I saw a research gap for the Automatic Speech Recognition exclusively for the Gujarati Language hence selected this topic. The work described in this report and the work I am about to do is my humble effort to be useful to the society by the application of Computers Systems and would like it to be appreciated.

VI. REFERENCES

[1] “Automatic Speech Recognition of Isolated Words in Hindi Language”, By Priyanka Wani ; U. G. Patil ; D.S. Bormane ; S.D. Shirbahadurkar, Computing Communication Control and automation (ICCUBEA), 2016 International Conference on – 2017

[2] "Development of Speech recognition technique for Marathi numerals using MFCC & LFZI algorithm", Deepali Malewadi , Gouri Ghule, - Computing Communication Control and automation (ICCUBEA), 2016 International Conference on -2017

[3] "Isolated word Automatic Speech Recognition (ASR) System using MFCC, DTW & KNN" Muhammad Atif Imtiaz, Gulistan Raja - Multimedia and Broadcasting (APMediaCast), 2016 Asia Pacific Conference on - 2017

[4] "Voice and speech recognition in Tamil language", Kiran R., Nivedha K., Pavithra Devi S., Subha T, - Computing and Communications Technologies (ICCCT), 2017 2nd International Conference on – 2017

[5] "MFCC Based Noise Reduction in ASR Using Kalman Filtering", Anuradha P Nair, Shoba Krishnan, Zia Saquib - Advances in Signal Processing (CASP), Conference on - 2016

[6] "Implementation and performance evaluation of continuous Hindi speech recognition", Ankit Kuamr, Mohit Dua, Arun Choudhary - Electronics and Communication Systems (ICECS), 2014 International Conference on – 2014

[7] “Automatic Speech Recognition for Connected Words using DTW HMM for English Hindi Languages ” by Shweta Singhal, Ajay Kumar Garg, Dr. Rajesh Kumar Dubey, International Conference on Communication, Control and Intelligent Systems (CCIS) – 2016

[8] https://en.wikipedia.org/wiki/Gujarati_language#cite_note-68

[9] https://analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/

[10] https://en.wikipedia.org/wiki/Types_of_artificial_neural_networks

[11] http://www.practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[12] https://link.springer.com/content/pdf/10.1007/0-387-29295-0_24.pdf

[13] https://in.mathworks.com/company/newsletters/articles/developing-an-isolated-word-recognition-system-in-matlab.html

[14] https://www.scribd.com/document/218588019/Isolated-Digit-Recognition-System

[15] http://www.dhananjaih.com/isolated-word-recognition.html