Hybrid Approach for Real Time Tricky Gujarati Word Recognition: Review Paper

Shruti R. Kandoriya[1],Shital Patel[2]

[1]Student, Computer Engineering Department, L.C. Institute of Technology, Gujarat, India
[2] Assistant Professor, Computer Engineering Department, L.C. Institute of Technology, Gujarat, India

ABSTRACT

Many researchers are working to make computer to understand naturally spoken language. Forinternationallanguage like English this technology has grown to a matured level. We presenta 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 Frequency Cepstral Coefficient (MFCC) feature extraction technique and HMM model. The recognizer isworking in abundance of three essential structure squares to be specific Feature extraction,Training and Testing (Recognition). The proposed here executes the Mel Frequency Cepstral Coefficient (MFCC) with a specific end goal to figure the otherworldly components of thediscourse signal Utilizing Support Vector Machine (SVM) to perceive discourse test to give fantastic results forsecluded words. It comprises of detached words that are isolated by quiets. This proposedsystem provides high accuracy for Gujarati language.

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

---

1. INTRODUCTION:

2.1 What is ASR?

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

2.2 History of ASR

The earliest attempts to devise systems for automatic speech recognition by machine were made in the1950s. Much of the early research leading to the development of speech activation and recognitiontechnology was funded by the National Science Foundation (NSF) and the Defense Department'sDefense Advanced Research Projects Agency (DARPA). Much of the initial research, performed withNSA and NSF funding, was conducted in the 1980s. (Source: Global Security.Org) Speech recognitiontechnology was designed initially for individuals in the disability community. For example, voicerecognition can help people with musculoskeletal disabilities caused by multiple sclerosis, cerebralpalsy, or arthritis achieves maximum productivity on computers. During the early 1990s, tremendousmarket opportunities emerged for speech recognition computer technology. The early versions of theseproducts were clunky and hard to use. The early language recognition systems had to makecompromises: they were
"Tuned" to be dependent on a particular speaker, or had small vocabulary, orused a very stylized and rigid syntax. However, in the computer industry, nothing stays the same forvery long and by the end of the 1990s there was a whole new crop of commercial speech recognitionsoftware 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 to3automatically create documents from dictation. Medical transcriptionists listen to dictated recordingsmade by physicians and other health care professionals and transcribe them into medical reports,Correspondence, and other administrative material. An increasingly popular method utilizes speechrecognition technology, which electronically translates sound into text and creates transcripts and draftsof reports. Transcripts and reports are then formatted; edited for mistakes in translation, punctuation, orgrammar; and checked for consistency and any possible errors. Transcriptionists working in areas withstandardized terminology, such as radiology or pathology, are more likely to encounter speechrecognition technology. Use of speech recognition technology will become more widespread as thetechnology becomes more sophisticated. Some voice writers produce a transcript in real time, usingcomputer speech recognition technology. Speech recognition-enabled voice writers pursue not onlycourt reporting careers, but also careers as closed captioners and Internet streaming text providers orcaption providers.

2.3 How Does ASR Work?

The goal of an ASR system is to accurately and efficiently convert a speech signal into a text messagetranscription of the spoken words independent of the speaker, environment or the device used to recordthe speech (i.e. the microphone). This process begins when a speaker decides what to say and actuallyspeaks a sentence. (This is a sequence of words possibly with pauses, uh’s, and um’s.) The softwarethen produces a speech wave form, which embodies the words of the sentence as well as the extraneoussounds and pauses in the spoken input. Next, the software attempts to decode the speech into the bestestimate of the sentence. First it converts the speech signal into a sequence of vectors which aremeasured throughout the duration of the speech signal. Then, using a syntactic decoder it generates avalid sequence of representations.

2.4 Benefit of ASR

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

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

2.5 Developments in ASR

Aside from the scientists and technicians who are engaged in ASR research and development, mostpeople 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 complexrecognition tasks, and the involvement of speech scientists, linguists, computer scientists,mathematicians, and engineers. The search is on for ASR systems that incorporate three features: largevocabularies, continuous speech capabilities, and speaker independence. Today, there are numeroussystems which incorporate these combinations.

2. LITERATURE REVIEW:

2.1"Automatic Speech Recognition of Isolated Words inHindi Language" [1]

Overview:

Speech recognition is a broad subject as speech is natural way of communication. The acousticand language model for this system are available but mostly in English language. In India thereare so many peoples who can’t understand or speak English. So the speech recognition systemin English language is of no use for these people. Here we presented Isolated Hindi wordsrecognition system which is a part of Automatic Speech Recognition (ASR) system. AutomaticSpeech Recognition (ASR) is also called as computer speech recognition. The main goal ofASR system understands a voice by computer or microphone and converts it into the text toperform required task. The performance accuracy of speech recognition is highly depends onfeature extraction and pattern recognitiontechnique. In this paper, we are using MFCC asfeature extraction technique, K-Nearest Neighbour (KNN) with GMM (Gaussian MixtureModel) for recognition of Hindi isolated words. For practical analysis we will prepare the Hindiwords speech dataset of different males and females speakers.

Issues / Limitations:

- In India there are lots of people that they are not able to speak or understand the Englishlanguage. Existing ASR systems are available only in few languages and have not beenused in any of the Indian languages. Thus this hinders the native users to make use ofthe technical advancement of ASR systems.

Solution Offered:

- MFCC for feature extraction and KNN with GMM for isolated word recognition.

Conclusion:

- Result is not effective its satisfactory Accuracy is 94.31.

2.2 "Development of Speech recognition technique forMarathi numerals using MFCC & LFZI algorithm"[2]

Overview:

India is a multilingual country, so a very less amount of work is done on the Regional languages. Marathi is one of the important regional language of Maharashtra. New interfacesystem with high precision is in progress of development by researchers. This paper presentsSpeech recognition by collection of database of isolated Marathi numerals ranging from zero(Shunya) to nine (Nau) which is implemented using two feature extraction techniques MelFrequency Cepstral Coefficient (MFCC) and Low pass Filter Zero Interpolation (LFZI). Theuttered speech samples of Marathi numerals are recorded of both male and female for about 1sec of duration. A database of total 1000 samples is collected and pre-processing is done oneach sample and further implementation that is feature extraction and classification is carriedout by the MATLAB. LFZI is one of the discrete wavelet transform method. It has numerousapplications such as bank chequeprocessing, pass port number, postal zip code, for physicallyimpaired people.

Issues / Limitations:

- India is a multilingual country, so a very less amount of work is done on the Regionallanguages.

Solution Offered:

- MFCC for feature extraction and KNN with GMM for isolated word recognition.
- The uttered speech samples of Marathi numerals are recorded of both male and femalefor about 1 sec of duration. A database of total 1000 samples is collected and pre9processing is done on each sample and further implementation that is feature extractionand classification is carried out by the MATLAB.
- State vector machine (SVM) covers the back-end for pattern classification.

Conclusion:

- Recognize only Marathi Numerical word, future work can be done in text.

2.3"Isolated word Automatic Speech Recognition (ASR)System using MFCC, DTW & KNN" [4]

Overview:

Automatic Speech Recognition (ASR) System is defined as transformation of acoustic speechsignals to string of words. This paper presents an approach of ASR system based on isolatedword structure using Mel-Frequency Cepstral Coefficients (MFCC's), Dynamic TimeWrapping (DTW) and K-Nearest Neighbour (KNN) techniques. The Mel-Frequency scale usedto capture the significant characteristics of the speech signals; features of speech are extractedusing MFCC's. DTW is applied for speech feature matching. KNN is employed as a classifier.The experimental setup includes words of English language collected from five speakers. Thesewords were spoken in an acoustically balanced, noise free environment. The experimentalresults of proposed ASR system are obtained in the form of matrix called confusion matrix.The recognition accuracy achieved in this research is 98.4 %.

Issues / Limitations:

Speaker independent language is not able work in noisy and robust environed.

Solution Offered:

The Mel-Frequency scale used to capture the significant characteristics of the speechsignals;

- Features of speech are extracted using MFCC's.
- DTW is applied for speech feature matching. KNN is employed as a classifier.

Conclusion:

- This system is provide solution for Speaker dependent data whereas further work canbe done for speaker independent

2.4 "Voice and speech recognition in Tamil language"[4]

Overview:

In our project, our intention is to create a voice and speech recognition system in smart phonesthat recognizes voice and captures the speech data in Tamil and stores and converts the capturedspeech as text in Tamil language itself. This can be used in voice dialing, sending SMS bysaying out the message and the captured message is sent to the recipient in Tamil. There hasnot been much consideration for Tamil language to be used in voice and speech recognition insmart phones. For native users Tamil voice recognitions and speech would provide moreflexibility in smart phone experience. Also people who have only been used to their nativelanguage Tamil, would feel easier to use the speech recognition system in their smart phonesif provided in Tamil. There will be no more difficulty in usage of phones for local users andthere is no need for any learning to use the smart phone. Automatic Speech Recognition (ASR)system have achieved a great success in many applications. Among them, Template Matchingtechniques like Dynamic Time Warping (DTW), Statistical Pattern Matching techniques suchas Hidden Markov Model (HMM) and Gaussian Mixture Models (GMM), Machine Learningtechniques such as Neural Networks (NN), Support Vector Machine (SVM), and DecisionTrees (DT) are most popular. For this system, highest word recognition accuracy is achievedwith HMM technique. It offered 100% accuracy during training process and approximately98% for testing process.

Issues / Limitations:

- Existing systems are available only in few languages and have not been used in any ofthe Indian languages. Thus this hinders the native users to make use of the technicaladvancement of ASR systems.

Solution Offered:

- HMM (Hidden Markov Model) are used to create voice and speech recognition systemin smart phone.

Conclusion:

- The use of these applications is limited due to language barriers that is there is noflexibility for native users.
- Lack of development in speech recognition systems in local languages has hinderedIndian smart phone users to make use of this technological advancement who feeldifficulty to use these systems in a foreign language rather than their own language.

2.5 "MFCC based noise reduction in ASR using Kalmanfiltering" [6]

Overview:

Speech enhancement using Kalman filter is an extensively researched area. The vast majorityof work done in this area uses linear predictive coding (LPC) for modeling speech signal. Afew important studies have revealed the superiority of Mel Frequency Cepstral Coefficients(MFCC) over LPC for speech recognition. With this paper, the shortcomings of speechenhancement using LPC with Kalman filters have been elaborated and MFCC, a much morefavored technique is used along with Kalman filter to ascertain proficient parameters from anoisy signal, which can be used for Automatic speech recognition (ASR).

Issues / Limitations:

- The majority of work done in this area uses linear predictive coding (LPC) formodeling speech signal.

Solution Offered:

- MFCC over LPC for speech recognition. With this paper, the shortcomings of speechenhancement using LPC with Kalman filters have been elaborated and MFCC.

Conclusion:

- Performance of MFCC with alternative forms of Kalman filter and its comparativestudy.

2.6"Implementation and performance evaluation ofcontinuous Hindi speech recognition,"[6]

Overview:
Speech to Text recognition is the ability of a machine to recognize the human speech andconvert in to text sequence. In this paper, we compare the performance of isolated word,connected word, and continuous speech recognition system with different vocabulary sizes.Hidden Markov Model toolkit HTK 3.4.1 is used to develop the system. For feature extraction,Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Prediction (PLP) both areused in this paper. The aim of this paper is to build a high performance speech recognitionsystem for Hindi language. Hidden Markov Model (HMM) and Gaussian Mixture Model(GMM) are used at the back-end of our proposed system. The system is trained for 100Hindi words and each word 10 utterances have been recorded for training of the ASR system.The experimental result shows that the overall accuracy of proposed system with 100 worddictionary size is 95.40%, when we use the combination of MFCC and GMM for automaticspeech recognition (ASR) system.

Issues / Limitations:

- As the size of dictionary increase the performance of system is decrease because ofthere are no trainee for short pause.

Solution Offered:

- Two algorithm are use MFCC in front-end while HMM backend. Once processcompleted MFCC replaced by PLP in front-end and same HMM replaced by GMM inbackend.

Conclusion:

- This paper tested solution for 100 words, for future increase vocabulary size.

2.7"Automatic speech recognition for connected words usingDTW/HMM for English/ Hindi languages,"[7]
Overview:
This work presents an automatic speech recognition (ASR) system for connected words. Aconnected ASR system has been implemented by extending an isolated word recognizer forspeaker dependent data. The work has been applied for English as well as Hindi language. Thetraditional approach of Mel frequency cepsral coefficient (MFCC) is used as features of thespeech signal. Hidden markov model (HMM) and dynamic time warping (DTW) are used atback-end for feature mapping of unknown utterances. A database of isolated English/Hindiwords is created for training phase while sentences are used for testing phase. The results areexpressed in terms of percentage word error rate (WER). The performance of system for twofeature extraction techniques (HMM, DTW) is compared.

Issues / Limitations:

- Most of the ASR build in English language. Hindi is being third most spoken languagein the worlds. For Hindi there are ASR are not available while English has lots oftranslator.
- Speaker Dependent.

Solution Offered:

- MFCC is used as features of the speech signal. HMM and DTW are used at back-endfor feature mapping of unknown utterances.
- A database of isolated English/Hindi words is created for training phase while sentences are used for testing phase.
- The results are expressed in terms of percentage word error rate (WER). The performance of system for two feature extraction techniques (HMM, DTW) is compared.

Conclusion:

- Further the work can extended for speaker independent data by improving theefficiency of isolated word recognizer.
- The real time system can be made by extending the database.

3. COMPARATIVE TABLE:

Table-1 Table ofComparison

SR. NO	TITLE	METHOD/ALGORITHM	DRAWBACK AND FUTURE SCOPE
1	Automatic Speech Recognition of Isolated Words inHindi Language	MFCC for feature extraction and KNN with GMM for isolated word recognition.	Result is not effective its satisfactory Accuracy is 94.31.
2	Development of Speech recognition technique for Marathi numerals using MFCC & LFZI algorithm	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	Isolated word Automatic Speech Recognition (ASR)System using MFCC, DTW & KNN	MFCC used for feature extraction where DTW is applied for speech feature matching. KNN isemployed as a classifier.	This system is provide solution for Speaker dependent data whereas further work can be done for speaker independent
4	Voice and speech recognition in Tamil language	HMM is used to create voice and speech recognition systemin 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	MFCC based noise reduction in ASR using Kalmanfiltering	MFCC over LPC for speech recognition. LPC with Kalman filters have been elaborated and MFCC.	
6	Implementation and performance evaluation of continuous Hindi speech recognition	MFCC in front-end while HMM backend. Once processcompleted MFCC replaced by PLP in front-end and same HMM replaced by GMM inback-end.	This paper tested solution for 100 words, for future increase vocabulary size.

4. CONCLUSION:

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

5. REFERENCES:

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

[2]"Development of Speech recognition technique for Marathi numerals using MFCC & LFZI algorithm", DeepaliMalewadi ,GouriGhule, - Computing Communication Control and automation (ICCUBEA), 2016

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

[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

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

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