



OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

REVIEW ON SMALL VOCABULARY AUTOMATIC SPEECH RECOGNITION SYSTEM (ASR) FOR MARATHI

Rajashri G. Kanke¹, Maheshwari A. Ambewadikar², Manasi R. Baheti³

Research Scholar, Department of Computer Science & Information Technology, Aurnagabad, Dr.Babasaheb Ambedkar Marathwada University Aurangabad, Maharashtra, India¹,

Research Scholar, Department of Computer Science & Information Technology, Aurnagabad, Dr.Babasaheb Ambedkar Marathwada University Aurangabad, Maharashtra, India²,

Associate Professor, Department of Computer Science & Information Technology, Aurnagabad, Dr.Babasaheb Ambedkar Marathwada University Aurangabad, Maharashtra, India³,

kankerajashri@gmail.com¹, mahiambewadikar@gmail.com², manasibaheti@gmail.com³

Abstract: *Speech interfaces make available people an easy and comfortable means to interact with human and Laptop Systems or PC. This paper is an attempt to study the Small Vocabulary Speech Recognition System for Marathi Using Automatic Speech Recognition System (ASR). Speech Recognition (SR) is a main element of Speech interfaces which identifies human speech in a specific language. This review presents a speech independent Speech Recognition System, for words of Marathi language. The aim of this paper is to review the systems or interfaces developed for some specific tasks in any native language; but with focus on Marathi. The aim is to review such systems in order to develop such interfaces in Marathi language that can be embedded in some application.*

Keywords: *Speech Recognition Technology, Natural Language Understanding, Automatic Speech Recognition (ASR), Human Computer Interface (HCI), Small Vocabulary*

I INTRODUCTION

There are many applications of the Speech Recognition, including various interfaces. This work aims to develop the Speech Recognition System in Indian languages using ASR system, will contain Small Vocabulary Speech in Marathi Language. There is need of various simple, interactive devices in Marathi language. A phone dialer, a device that can be used for hands-free applications like talking to your mobile phone, computer system for daily transactions. This will be very helpful for the illiterate, disable persons. Our aim is to develop a Small Vocabulary Speech Recognition System in Marathi language that can be used to embed into one of such devices mentioned above. This review paper on more techniques, methods and algorithm to use for speech

recognition system for natural languages. Main focus on Marathi language because very few work for Marathi language in automatic speech recognition system. We refer the 8 to 10 years old paper in ASR system, maximum 30 to 20 paper refer on automatic speech recognition system. The formatter will need to create these components, incorporating the applicable criteria that follow.

II AUTOMATIC SPEECH RECOGNITION (ASR)

The process of developing the transcript in word classification of an utterance, given the speaking waveform is called Automatic Speech Recognition System (ASR). Automatic Speech Recognition (ASR) is the main challenge to the natural human to machine communication develop technology.

III SMALL VOCABULARY SPEECH

Small Vocabulary Speech is also know isolated words. It is mean that the isolated word recognizes attain usually require each utterance to have quiet on both side of sample windows. It accepts single words or single utterance at a time. Small Vocabulary Speech is very useful to every one person. Isolated words might be improved label of this class.

IV LITERATURE SURVEY

This review on Small Vocabulary Speech Recognition System in Marathi Language. There are many methods, and

different procedures and algorithm to use Automatic Speech Recognition system and advance technique to interact between computer system and human in different Indian languages. Some applications are very useful for disable peoples and physically not stay able. They cannot properly used computer system and mobile phone, so it can help to disable peoples and physically not stay able. Following table 1 shows the details of the review done for this study. It shows the methods/techniques applied, spoken language, category of token that is, isolated, continuous etc. along with some observations.

Sr. No.	Authors	Paper Name/ Year	Methods/ Techniques	Spoken Language/Cat egory of Tokens	Accuracy/ Observations
1	Maheshwari A. Ambewadik ar, Manasi R. Baheti	Review On Speech Recognition System For Disabled People Using Automatic Speech Recognition (ASR), (2019)	Automatic Speech Recognition (ASR), HCI (Human Computer Interaction), Cerebral Palsy (CP), Mel Frequency Cepstral Coefficient (MFCC). Dynamic Time Warping (DTW)	Marathi language, Marathi Words	Speech Recognition System are variable in nature and in a paper, the authors M. A. Ambewadik ar and Manasi R. Baheti reviewed the ASR systems in Marathi language. It was observed that many Speech Recognition System are available but scope is limited to English language.
2	Suhas R. Mache, Manasi R. Baheti , C. Namrata Mahender	Review on Text To Speech Synthesizer, August 2015	Text Processing, Text To Speech (TTS) synthesizer, Speech Enhancement, Linguistic analysis, Text normalization, speech synthesis techniques, Smoothing Method.	Many Regional Languages Like Marathi, Hindi, Telugu, Punjabi, Kannada etc. Sentence	In this paper discuss the text to speech synthesizer develop the system. It can used the Speech synthesis technique for the text to speech converted to the natural languages in India, smoothing method also used in this application.
3	Manasi Ram Baheti, Bharti W. Gawali, S.C. Mehrotra	Automatic Speech Recognition For Task Oriented IVRS In Marathi, (Mar to Apr 2016)	Automatic Speech Recognition (ASR), Dynamic Time Warping (DTW), HCI, Interactive Voice Response (IVRS), Mel Frequency Cepstrum Coefficient (MFCC), computerized Speech Laboratory (CSL).	Ten Sentences In Marathi Language.	This research paper develop the task oriented IVRS in natural language of Marathi .There are two technique are used in this experiment MFCC and DTW. MFCC is use a feature extraction of speech and DWF is used recognize sentence and matching the sentence. This system most useful to farmers and small town peoples to get the information in own language.

4.	Anjana Vakil, Alexis Palmer	Cross Language Mapping For Small Vocabulary ASR In Under Resourced Languages Investigating The Impact Of Source Language Choice, (2017)	Salaam algorithm, lexicon building, phoneme mapping, pre-existing recognition engine.	Under-Resourced Languages (French, English), 25 words	This research work to use the salaam method use the pronunciation lexicons mapping for under resourced language. In this paper used a salaam method. The word recognition accuracy rate greater than 95% for real time application.
5	Ravindra Parshuram Bachate, Ashok Sharma	Automatic Speech Recognition Systems for Regional Languages in India (July 2019)	ASR (Automatic Speech Recognition), Arbitrary Strength Unit(ASU), Deep Neural Network(DNN),MFCC, Noise Reduction, Language Modal, Acoustic Modal, Pronunciation Modal	Indian language, Words, Sentence	Main aim of this research paper Automatic Speech Recognition System using many feature extraction methods and techniques in English language like MFCC but not done Indian language to work in Automatic Speech Recognition System and many methods and algorithm to explain in this research paper
6	Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, Yoshua Bengio	End To End Attention Based Large Vocabulary Speech Recognition (14 Mar 2016)	Automatic Speech Recognition (ASR), Recurrent Neural Networks (RNN), Hidden Markov Models (HMMs), Gaussian Mixture Models (GMMs), N-Gram Language Models, Large Vocabulary Continuous Speech Recognition System (LVCSR),	20K Sentences, Trigram Language Model,	This paper present the HMM method replace the RNN method, RNN method is learned automatically by attention mechanism build. To use two method BiRNN layer to less the length of encoded and 2 nd method is use Decoder Network for output character. Best result (2014) 13.0% vs. 14.1%, Worst result then 1, (2015) 11.3% vs. 9.0% used in same language model.
7	Mr. Himanshu N Patel, Dr. P.V. Virparia	A Small Vocabulary Speech Recognition for Gujarati (Jan-Feb 2011)	Fast Bootstrapping , Align Gujarati Speech by Using an English Speech Recognition, Acoustic Models for the Phones	Gujarati Language, Ten Words in Gujarati Language	In this research paper main focus on Gujarati language and using a Fast Bootstrapping technique for initial phone models. Total 31 people to collect the speech data. overall recognition accuracy 88.71% recognition accuracy of 19 male and 85.28% recognition accuracy of 12 female.

8	Bradley J. Betts, Kim Binsted, Charles Jorgensen	Small Vocabulary Speech Recognition Using Surface Electromyography, (10 October 2006)	Surface Electromyography (EMG) , Neural Network Classifier , Signal Processing, HMM,	15 words in English ,data collect the 150 examples per word	In this work to Electromyography in English to develop the system total 15 words to use in system and accuracy below the 95% in a total accuracy in this experiment 71% to 77%.
9	Nikolaos Doukas, Nikolaos G. Bardis	Current Trends In Small Vocabulary Speech Recognition For Equipment Control, (06 September 2017)	Speech Recognition (SR) Frequency Cepstral Coefficients (MFCC), Dynamic Time Warping (DTW), Features Extraction Techniques, Voice Activity Detection Principle (VAD),	Sentences in Speech Processing and also Isolated words	In this work to develop the design the strong and correct speech recognition. MFCC and DTW used commonly to recognize the small vocabulary speech recognition system are present.
10	Siddharth S. More, Prashantkumar L. Borde, Sunil S. Nimbhore,	A Review on Automatic Speech Recognition System in Indian Regional Languages, (July 2018)	Automatic Speech Recognition, Speech Processing, Signal Processing, Pattern Recognition, Speech Feature Extraction Techniques, HTK toolkit ,HMM, word Error rate (WER)	Indian Regional Languages, Word, Sentences, long Sentences	This research paper on the Automatic Speech Recognition using the HTK toolkit method. This paper represent the HTK toolkit method accuracy rate is high than compare other techniques. And main focus on Indian languages.
11	Potale Shubham, Khaepude Pratik , Patil Rahul, Ajay Kumar Gupta	Design and Development of Word Recognition for Marathi Language, (2016)	Speech Recognition , MFCC, LPCC, Sphinx4 tool, JAVA, HMM,	Isolated Words in Marathi Language.	It is work on isolated words in Marathi language using a Matlab software, it can MFCC and LPCC technique used in this work to voice recognition application interact between human and internet, computer system.

The above table gives the focus on techniques, here few simple and popular techniques are mentioned in brief.

Mel Frequency Cepstral Coefficient (MFCC) Technique

MFCC stand for the Mel Frequency Cepstral Coefficient .it is most important and most useful for feature extraction for the speaker or speech signals. Mel Frequency Cepstral Coefficient estimates people Speech Signals more correctly compare to any other method or technique. It is the features form signal frames based on small term analysis. Mel Frequency Cepstral Coefficient (MFCC) are features widely used in Automatic Speech and Speaker Recognition System.

Hidden Markov Model (HMM)

This model generally purpose of Speech Recognition Systems are based on Hidden Markov Model (HMM) technique. These are numerical models that output a

classification of symbols or numbers. Hidden Markov Model (HMM) are mostly used in speech recognition system since a speech motion can be viewed as a piecewise stationary signal or a very short time stationary signal. Hidden Markov Model (HMM) are current used because they can be trained automatically and are very simple to use. This technique mostly used in Automatic speech Recognition System (ASR).

Dynamic Time Warping Method (DTW)

Dynamic Time Warping (DTW) is method that was normally used for Speech Recognition System (SP). Dynamic time warping (DTW) is a process for parallel between two classifications that may vary in period or speed. Dynamic Time Warping method has been useful to video capture, audio recording, and also useful graphics design truly. And any data or files that can be turned into a direct representation can be discovered with Dynamic time Warping (DTW).

Neural networks (NN)

Neural Networks (NN) developed as an acoustic modeling method in Automatic Speech Recognition System (ASR) in 1980. Neural Network (NN) have been used in many features of Speech Recognition (SP) such as phoneme classification. Neural Networks (NN) method very useful to Speech Recognition System most of research work to use a neural network method. Phoneme classification through Isolated Word Speech Recognition, audiovisual speaker recognition and speaker version.

After reviewing some existing work, it is observed that various techniques are used by different authors. It depends upon the application that one wants to develop. Speech is not stationary in nature hence, Speech Recognition is task oriented domain. Depending on these task, respective techniques are used. For small datasets, the techniques like MFCC, DTW etc are heavily used. For larger and standardized corpus, Neural Network based techniques are preferred. But MFCC and DTW are popular, simple and robust techniques. Depending upon the aim of the proposed system, type of speech considered, size of the database, these techniques are implemented.

V CONCLUSION

In this review paper many methods and algorithm used like Dynamic Time Warping (DTW), Mel Frequency Cepstrum Coefficient (MFCC), Speech Feature Extraction Techniques, Gaussian Mixture Models (GMM), neural networks, Salaam method are used to speech recognition system or classification of speech . In this literature review on Automatic Speech Recognition System many Indian languages used but very few work to developed in Marathi language so my main aim to many application to developed in Marathi languages

ACKNOWLEDGMENT

We thank would like to acknowledge the support of Department Of Computer Science And Information Technology, providing me facilities. Further I would like to especially thank to my guide Dr. Manasi R. Baheti.

REFERENCES

1] M. A. Ambewadikar and M. R. Baheti, Review on Speech Recognition System for Disabled People Using Automatic Speech Recognition (ASR), 2020 International Conference on Smart Innovations in Design, Environment,

Management, Planning and Computing (ICSIDEMPC), AURANGABAD, IEEE (2020), pp. 31-34, doi:10.1109/ICSIDEMPC49020.2020.9299615.

- 2] Suhas R. Mache, Manasi R. Baheti ,C. Namrata Mahender, “Review on Text To Speech Synthesizer”, JARCCCE(2015), Vol. 4, Issue 8, August 2015
- 3] Manasi Ram Baheti, Bharti W. Gawali, S.C. Mehrotra, “Automatic Speech Recognition For Task Oriented IVRS In Marathi”, IOSR Journal Of Computer Engineering (IOSR JCE), www.Iosrjournals.Org, Mar-Apr. 2016
- 4] Anjana Vakil, Alexis Palmer, Cross Language Mapping For Small Vocabulary ASR In Under Resourced Languages Investigating The Impact Of Source Language Choice, (2017) <http://lex4all.github.io/lex4all>
- 5] Ravindra Parshuram Bachate, Ashok Sharma, Automatic Speech Recognition Systems for Regional Languages in India, International Journal of Recent Technology and Engineering (IJRTE), Published: Blue Eyes Intelligence Engineering & Sciences Publication, July 2019
- 6] Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, Yoshua Bengio, End to end attention based large vocabulary speech recognition , www.arXiv:1508.04395 ,14 mar 2016
- 7] Mr. Himanshu N Patel, Dr. P.V. Virparia, “A Small Vocabulary Speech Recognition for Gujarati”, International Journal of Advanced Research in Computer Science(IJOARCS), volume 2,No.1, www.ijarcs.info ,Jan-Feb 2011
- 8] Bradley J. Betts, Kim Binsted, Charles Jorgensen, Small vocabulary speech recognition using surface electromyography, Elsevier, www.sciencedirect.com , www.elsevier.com/locate/intcom, 10 October 2006
- 9] Nikolaos Doukas, Nikolaos G. Bardis, Current Trends in Small vocabulary speech recognition for equipment control, AIP publishing, <https://doi.org/10.1063/1.4996686> , 06 September 2017
- 10] Siddharth S. More, Prashantkumar L. Borde, Sunil S. Nimbhore, “A Review on Automatic Speech Recognition System in Indian Regional Languages”, IJOCA (International journal of computer applications) vol. 181, No.4, July 2018
- 11] Potale Shubham, Khaepude Pratik , Patil Rahul, Ajay Kumar Gupta, Design and Development of Word Recognition for Marathi Language, Vol.7 (IJCSIT), www.ijcsit.com, 2016