

Design and Development of Word Recognition for Marathi Language

Potale shubham^{#1}, Kharpude Pratik^{#2}, Patil rahul^{#3}
¹Ajay kumar gupta

Abstract - Speech is the most prominent form of communication among human beings. English is an official language at many institutes all over the world and maximum research work is carried out for English language. As far concern with Indian languages, small amount of work is carried out. The work carried out for Indian languages namely Hindi, Bengali, Punjabi, Telugu, Malayalam and Assamese. But very little amount of work is carried out for Marathi language.

It is said that spoken Marathi Language changes at every 14 miles and has different types of phonetics and intonation. On the other hand Marathi is used as official language at very few places. There are very few institutes which have authority to comment on correctness of language. The authenticity of correctness of language is a major problem in regional languages.

Our project is capable to recognize the isolated Marathi word. At the initial level effort is made to provide help for basic operations. Most of the research is carried out on MATLAB but we go for Sphinx 4 platform as it concerns with JAVA.

Key Words: Speech Recognition System, Sphinx4, MATLAB etc.

1. Introduction

Speech is most prominent way of communication in world wide. Speech is the heart of human activity because it helps human to interact each other in more natural and effective way. Speech recognition is the process to identify words or phrase from spoken language and convert into machine readable format. Speech recognition is the process by which a computer (or other type of machine) identifies spoken words. Basically, it means talking to your computer, AND having it correctly recognizes what you are saying. Most of the communication in Maharashtra is done in Marathi language .While going through papers we studied that for Marathi language very small amount of work and research is done and is not enough to use in practical application. This project can be extended to other languages in Maharashtra were each and every person can use voice application to command, to search, to interact with internet computer etc.

1.1 Literature Survey

Speech recognition came into existence during 1920. The first machine i.e. Radio Rex ,a toy to recognize voice was manufactured. Bell Labs developed a speech synthesis machine at the World fair in New York. But later on they discarded efforts based on an incorrect conclusion that the AI is ultimately required for success. In order to develop systems for ASR, attempts were made in 1950s where researchers studied the fundamental concepts of phonetic-acoustic. Most of the systems in 1950 for recognizing speech examine the vowels spectral resonances of each utterance. At Bell Labs Davis, Biddulph and Balashek(1952) premeditated a isolated digit recognition system for a single speaker using formant frequencies estimated during vowel regions of each digit. At RCA Labs, Olson and Belar (1950) built 10 syllables recognizer of a single speaker and Forge and Forge built a speaker-independent 10-vowel recognizer at MIT Lincoln Lab, by measuring spectral resonances for vowels. Fry and Denes (1959) tried to build a phoneme recognizer to recognize four vowels and nine consonants at University College in England by using a spectrum analyser and a pattern matcher to make the recognition decision. Japanese labs entering recognition field in 1960-70. As computers are not fast enough, they designed special purpose H/W as a part of their system. In Tokyo, Nagata et.al described a system of the Radio Research lab, was a H/W vowel recognizer. Another effort was the work of Sakai and Doshita in 1962, of Kyoto University who built a H/W phoneme recognizer. In 1963, Nagata and co-workers at NEC Labs built a the digit recognizer. This led to a long productive research program. In 1970, the key focus of research was on isolated word recognition. IBM researchers studied in large vocabulary speech recognition. At AT&T Bell Labs, researchers began speaker independent speech recognition experiments. A large number of clustering algorithms were used to find the number of distinct patterns required to represent words

A model based spectral estimation algorithm has been developed. In 2000, a variational Bayesian estimation technique was developed. It is based on posterior distribution of parameters.

Giuseppe Richardi has developed a technique to solve the problem of adaptive learning in ASR. In 2005, some improvements have been made on Large Vocabulary Continuous Speech recognition system for performance improvement. A 5-year national project Corpus of Spontaneous Japanese (CSJ) was conducted in Japan. It consists of 7 million of words approximately, corresponding to speech of 700 hours.

2. TYPES OF SPEECH RECOGNITION SYSTEM

There are three types of Speech are as follows:

2.1 Isolated Words

Isolated word recognizers usually require each utterance to have quiet (lack of an audio signal) on BOTH sides of the sample window. Often, these systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances (usually doing processing during the pauses). Isolated Utterance might be a better name for this class.

2.2 Connected Words

Connect word systems (or more correctly 'connected utterances') are similar to Isolated words, but allow separate utterances to be 'run-together' with a minimal pause between them.

2.3 Continuous Speech

Continuous recognition is the next step. Recognizers with continuous speech capabilities are some of the most difficult to create because they must utilize special methods to determine utterance boundaries. Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Basically, it's computer dictation.

3. PROPOSED SYSTEM

Sphinx is a simple feasible and open source tool used in speech recognition engine to communicate with smart devices. We are using sphinx tools with java language to identify isolated word for Marathi language. We are using Small vocabulary initially and then we will expand the vocabulary to large amount. Here we are going to create a Database of 50 Speakers (25 Female & 25 Male). And starting from 10 isolated words we will take 5 utterances of each word from thoughts 50 speakers.

4. TOOLS FOR SPEECH

RECOGNITION

Following are the various tools used for ASR

PRAAT: It is free software with latest version 5.3.04 which can run on wide range of OS platforms and meant for recording and analysis of human speech in mono or stereo

AUDACITY: It is free, open source software available with latest version of 1.3.14(Beta) which can run on wide range of OS platforms and meant for recording and editing sounds.

CSL: Computerised Speech Lab is a highly advanced speech and signal processing workstation (software and hardware). It possesses robust hardware for data acquisition and a versatile suite of software for speech analysis.

SPHINX: Sphinx 4 is a latest version of Sphinx series of speech recognizer tools, written completely in Java programming language. It provides a more flexible framework for research in speech recognition.

SCARF: It is a software toolkit designed for doing speech recognition with the help of segmental conditional random fields.

MICROPHONES: They are being used by researchers for recording speech database. Sony and I-ball has developed some microphones which are unidirectional and noiseless.

5. SUMMARY

RESEARCH IN SPEECH RECOGNITION HAS BEEN CARRIED OUT INTENSIVELY FOR THE LAST 60 YEARS.

6. PERFORMANCE OF SPEECH RECOGNITION SYSTEM

The performance of speech recognition is specified in terms of accuracy and speed. Accuracy is measured in terms of performance accuracy which is known as word error rate (WER) whereas speed is measured with the real time factor.

Word Error Rate It is a common metric of the speech recognition performance. As recognized word sequence have a different length from the reference word sequence, there is difficulty in measuring performance.

By using sphinx 4 tool it gives accuracy about app. 60% of word recognition of isolated words. which is more as compare to other tools.

As this project is run using Sphinx 4 and Speech Recognition domain. For speech recognition domain there is no provision for Test case diagram. This project is not a software engineering project. In this project only implementation and testing will done using sphinx 4 tool..

7. Conclusions

Hence we studied work done for isolated Marathi language and we come to a conclusion that for

isolated word all the work is done in MFCC and LPCC or in Matlab No another work is done using Sphinx tools

8.ACKNOWLEDGMENT

We thank the anonymous referees for helpful comments and suggestions. We also thank College to provide us a useful guidance and to encourage us for this project.

References

- [1] Project (Outing Unfortunate Characteristics of HMMs). CiteSeerX: 10.1.1.395.7249.
- [2] Kincaid, Jason. "The Power Of Voice: A Conversation With The Head Of Google's Speech Technology". Tech Crunch. Retrieved 21 July 2015.
- [3] Froomkin, Dan. "THE COMPUTERS ARE LISTENING". The Intercept. Retrieved 20 June 2015.
- [4] Deng, L.; Hinton, G.; Kingsbury, B. (2013). "New types of deep neural network learning for speech recognition and related applications: An overview".doi:10.1109/ICASSP.2013.6639344.
- [5] Markoff, John (November 23, 2012). "Scientists See Promise in Deep-Learning Programs". New York Times. Retrieved 20 January 2015.
- [6] "Speech Recognition for Learning". National Center for Technology Innovation. 2010. Retrieved 26 March 2014.
- [7] Mohri, M. (2002). "Edit-Distance of Weighted Automata: General Definitions and Algorithms" (PDF). International Journal of Foundations of Computer Science. **14** (6): 957–982. doi:10.1142/S0129054103002114. Retrieved 2011-03-28
- [8] Sadaoki Furui, November 2005, 50 years of Progress in speech and Speaker Recognition Research , ECTI Transactions on Computer and Information Technology, Vol.1. No.2.
- [9] K.H.Davis, R.Biddulph, and S.Balashek, 1952, Automatic Recognition of spoken Digits, J.Acoust.Soc.Am.,24(6):637-642.
- [10] H.F.Olson and H.Belar, 1956, Phonetic Typewriter , J.Acoust.Soc.Am.,28(6):1072-1081.
- [11] J.W.Forgie and C.D.Forgie, 1959, Results obtained from a vowel recognition computer program , J.Acoust.Soc.Am., 31(11),pp.1480-1489.
- [12] D.B.Fry, 1959, Theoretical Aspects of Mechanical speech Recognition , and P.Denes, The design and Operation of the Mechanical Speech Recognizer at

Universtiy College London, J.British Inst. Radio Engr., 19:4,211-299.

- [13] K.Nagata, Y.Kato, and S.Chiba, 1963, Spoken Digit Recognizer for Japanese Language , NEC Res.Develop., No.6.
- [14] T.Sakai and S.Doshita, 1962 The phonetic typewriter, information processing 1962 , Proc.IFIP Congress.