Speech To Text Conversion For Desktop Application

Priyanka Belhekar[1], Aparna Padannakara[2], Yogita Navgire[3] and Prof. Mrs. Kavitha S[4]

[1]School of Computer Engineering & Technology,MIT AOE,Alandi(D),Pune, India
[2]School of Computer Engineering & Technology,MIT AOE,Alandi(D),Pune, India
[3]School of Computer Engineering & Technology,MIT AOE,Alandi(D),Pune, India
[4]School of Computer Engineering & Technology,MIT AOE,Alandi(D),Pune, India
[1]priyankapbelhekar26@gmail.com, [2]padannakaraaparna@gmail.com, [3]yogita.navgire555@gmail.com, [4]kavithas@comp.maepune.ac.in

* Corresponding Author: Priyanka Belhekar
Manuscript Received: Manuscript Accepted:

Abstract

Speech recognition is a fast growing engineering technology with potential benefits that are provided by the different applications of speech recognition in different areas. Nearly twenty percent people of the planet are suffering from varied disabilities. Several of them are blind or unable to use their hands effectively. The speech recognition systems in those specific cases offer a significant facilities to them, so that they can share the information with people by operating their computer through voice input. The aim of this project is that the people with disabilities can easily access computers through the voice input. This project converts the speech into text and perform the proper operations. At the initial level effort is created to perform easier and basic operations, however the software system will be updated and enhanced accordingly to perform lot of other operations. The project consist of hardware which is connected to the laptop or computers so it can start the computer by voice. Model is made up of Raspberry pi. In this paper Keyword spotting (KWS) is used with the help of Automatic Speech Recognition System (ASR) which help in focusing to decrease the false rate(FA).
Keywords: Feature Extraction, Speech Recognition, Keyword Spotting.

I. INTRODUCTION

Speech Recognition or Speech to text is nothing but simply the conversion of human audio into the text which is been said by the human[1]. It is commonly used to operate a device, perform an action or to write in a system without using keyboard, mouse or press any button. It is more useful for the blind and other disable people who can operate the computer by simply using their voice and using such application the time is also saved. The microphone is used by the system to recognize the sound which is spoken by the human or user and after recognizing the words, the system then convert that audio or voice input into text format such a technology is called as speech recognition system[2]. The output is calculated by the system with the help of recognizer in which the recognizer processes the input data or spoken words. Speech Recognition system consist of various steps. All words spoken by the human or user is consider by the speech recognition engine to give a proper output or perform a proper action which involves variety of things[6]. It is an ideal scenario in the method of speech recognition. Vocabularies, multiple users and the noisy surrounding area units are the foremost factors that are counted in by a speech recognition engine[6]. In this project, the desktop application is been developed in which the user control computer function through their voice for operating particular operations. For eg: if the user is saying open chrome then the system will take it as a command and perform the required operation which result in opening the chrome, likely other command such as open notepad, save document, delete document etc. can also be performed. But regardless of the application, it is desirable to minimize the cost of the system, as well as to make the most out of the system that we have installed. In this paper, the KWS system is described. KWS system runs on the device and implement the query by sending it to the server[11]. KWS system runs on the server side along with different models which results into more accurate result of query[10]. Once the query is accepted by the KWS system it results into string matching and if the query is correct then it take the necessary action otherwise the query is suppressed[11].

II. RELATED WORK

After going through the background of this study we have reported to the several work.

Andrew Ng [1] stated that the programmed Speech Recognition is translating of spoken words or sentences into content dialect. It is still a difficult task because of high inconstancy of voice signals. This paper gives us the data about the profound learning calculations including Deep Relief Networks(DRN) and Deep Belief Network(DBN). These systems are utilized for executing the speech recognition. There are three frameworks that are: GMM-HMM, DND-HMM and DBN. From this paper it is inferred that DBN based speech recognition framework is superior to other two discourse acknowledgment frameworks that the author has mentioned.

Michaely et al. [4]In this paper, the author has described the various researches carried out in the areas of ASR(Automatic Speech Recognition) System and the advances made in the researches. This paper has also mentioned about the types of ASR system, different challenges faced in ASR and its applications. This paper gives us the overview of the progress made in some years back till date.

Khilari et al. [2] in her study the technical idea of speech to text conversion is explained. It gives the information of technical process in each steps. As the different stages have different techniques the study of that different techniques are done according to the stages. This paper gives the idea of the different types of speech: Isolated Word System, Connected Word System, Continuous Speech, Spontaneous Speech. Speech recognition system can be developed by two ways which is, Speaker Dependant Models Speaker Independent Model As well as in applications of speech recognition is also explained.

Mon et al. [3] explained how Speech to Text conversion is helping deaf and dumb students in the educational field. This paper presents the speaker independent system. In this paper Mel-frequency Cepstral Coefficients(MFCC) method is used. It is based on the characteristics of human hearing which uses nonlinear frequency unit to simulate the human auditory system. The main goal of their project is to apply deep learning algorithm to speech recognition and compare the speech recognition performance with GMM-HMM based speech recognition method.

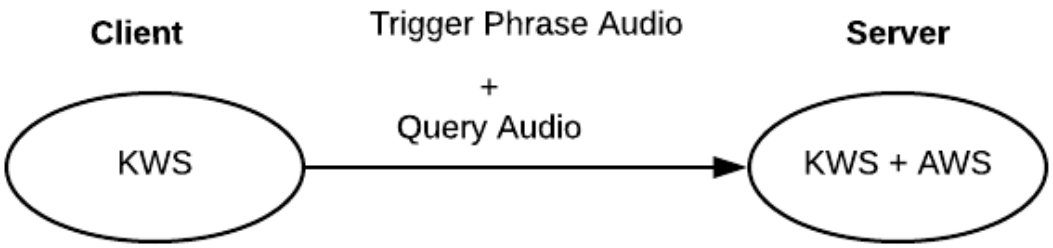
Saksamudre et al. [4] proposed the idea about different approaches for feature extraction in speech recognition system, the different types of proposed models for Automatic Speech Recognition,its advantages and disadvantages of that models and approaches which helps us to select the best model for Automatic Speech Recognition. For recognizing the speech of different speaker various approaches are given in this paper. This paper also explains the different types of speech recognition system based on utterances which include isolated words, connected words, continuous words, spontaneous speech etc[7][8]. The speaker dependent, speaker independent, speaker adaptive are the types of Automatic Speech Recognition based speaker model. Pre-processing/digital processing, feature extraction, acoustic modeling, language modeling, pattern classification are the functions of

speech recognition system. This paper also include different feature extraction techniques like Principle component analysis, Linear Discriminate Analysis, Independent Component Analysis, Linear Predictive Coding, Filter Bank Analysis, Wavelet, RASTA Filtering. The author has also proposed various approaches for pattern matching speech recognition are Template-Based Approach, Knowledge-Based Approach, Neural Network-based approach, Dynamic Time Wrapping Based Approach, Statistical-Based Approach, Hidden Markov Model Based Approach[16].

From the referred paper it is observed that the accuracy of the proposed model is less and False Rate(FR) is high. In this paper, the proposed model is with high accuracy and less false rate with some additional features included in it.

III. CLIENT-SEVER KWS SYSTEM MODEL

The paper focuses about the Keyword Spotting System (KWS) on both client side and server side.



The KWS system on the client side continuously accept the input or audio query from the user and sent it to the server side KWS system along with trigger phrase. The trigger phrase is nothing but the command like "ok google" etc which help the device on the client side to recognise the query. The query with the trigger phrase is accepted by the server side KWS system and it processed that query with the help of ASR system. If the query is not along with the trigger phrase then that query is rejected.

IV. LANGUAGE MODEL(LM) ADAPTION

Language Model is used for the detection of words from the larger text data like paragraphs, stories etc. The 2 pass LM with smaller first pass and larger second pass is used. These LMs are trained for some sentences which are without trigger phrase most of the time.

V. ENDPOINT SYSTEM

In our final application one hardware will be connected to the computer from which we can access the computer through our voice. Along with the co mmand we need to give a unique trigger phrase for eg."ok google" etc. after saying the triggered phrase along with command it get accepted by the client KWS system and that system passes it to the server KWS system. Then Server KWS system along with ASR system processes that command using different algorithms and models like string matching. After processing, if that command is valid then operation is executed otherwise that command is suppressed.

CONCLUSION

Using this project user can perform different operations like open, save, edit, shutdown etc. through voice input and simultaneously we can see the given voice input in the text format.

REFERENCES

1) Su Myat Mon,Hla Myo Tun "Speech To Text Conversion System Using Hidden Markov Model"June 2015,ISSU 06,VOLUME 4
2) Prachi Khilari1, Prof. Bhope V. P.2 "Implementation of Speech to Text Conversion International Journal of Innovative Research in Science, Engineering and Technology" Vol. 4, Issue 7, July 2015
3)] Suman K. Saksamudre,P.P. Shrishrimal,R.R. Deshmukh "International Journal of Computer Applications (0975 8887) A Review on Different Approaches for Speech Recognition System" Volume 115 No. 22, April 2015
4) Assaf Hurwitz Michaely,XuedongZhang,Gabar Simko,Carolina Parada,peter Aleksic "Keyword Spotting for Google Assistent Using Contextual Speech Recognition"
5) Iqbaldeep Kaur, 2Navneet Kaur, 3Amandeep Ummat, 4Jaspreet Kaur, 5Navjot Kaur, ISSN : 0976-8491 (Online) — ISSN : 2229-4333 (Print),Automatic Speech Recognition: A Review,Vol. 7, Iss ue 4, Oct - Dec 2016
6) Suma Swamy1 and K.V Ramakrishnan,Computer Science Engineering: An International Journal (CSEIJ),AN EFFICIENT SPEECH RECOGNITION SYSTEM,Vol. 3, No. 4, August 2013
7) Pahini A. Trivedi, 2014 IJEDR,Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks,Volume 2, Issue 4
8) Rupali.S.Chavan,Ganesh.S.Sable "An Overview Of Speech Recognition Using HMM" June 2013,Issue 6,Volume 2.
9) Prachi Khillari,Bhope.V.P "A Review On Speech To Text Conversion Methods" July 2015,Volume 4 ,Issue 7.
10) Santosh K.Gaikwad,Bharti W.Gawali,Pravin Yannawar "A Review On Speech Recognition Techniques" Nov 2010,Volume 10-No 03
11) Nnamdi Okomba S,Adegboye Mutiu Adesina Candidus O.Okwor "Survey of Technical progress in Speech Recognition By Machine over Few Years Of Research" July-Aug 2015,Volume 10,Issue 4.
12) Rashmi C R "Review Of Algorithms And Applications in Speech recognition System" 2014,Volume
13) A.Ghoshal,p.Swietojanski,s.Renals,"Multilingual Training Of Deep Neural Networks",2013

14) K.M.Knill,M.J.F.Gales,S.P.Rath,P.C.Woodland,C.Zhang and S-X.Zhang,"Investigation of Multilingual Deep Neural Networks for Spoken Term Detection",2013
15) J.G.Fiscus,J.Ajot,J.S.Garofolo,and G.Doddingtion,"Results of the 2006 Spoken Term Detection Evaluation",2007.
16) W.Stefan and H.Reinhold,"Approaches to iterative speech feature enhancement and recognition",IEEE Transaction on Audio ,Speech and Language Processing,Volume.No.5,July 2009.
17) L.Rabiner and B.H.Jaung,"Fundamentals of Speech Recognition",1993
18) F.Sadaoki,"50 years of progress in speech and speaker recognition research",Volume.1.No.2,November 2005
19) H.Sakoe and S.Chiba,"Dynamic Programming Algorithm Optimization for Spoken Word Recognition",1997
20) M. Afify, F. Liu, and H. Jiang, A new verificationbased fast-match for large vocabulary continuous speech recognition, Vol. 133 No.4 July 2005