Analysis of Different Speech Recognition Toolkit for EMR

Manu S Mohan, Shan George Mathew, Surya Paul,
V.A Nandaraj and Archana Aniyan

August 17, 2021

ANALYSIS OF DIFFERENT SPEECH RECOGNITION TOOLKIT FOR EMR

Manu S Mohan
UG Scholar, Dept. of Electronics and Communication Engineering
Adi Shankara Institute of Engineering and technology
Ernakulam, India
manusmohan@gmail.com

Shan George Mathew
UG Scholar, Dept. of Electronics and Communication Engineering
Adi Shankara Institute of Engineering and technology
Ernakulam, India
shangeorgemathew@gmail.com

Surya Paul
UG Scholar, Dept. of Electronics and Communication Engineering
Adi Shankara Institute of Engineering and technology
Ernakulam, India
suryapaul6996@gmail.com

Nandaraj V A
UG Scholar, Dept. of Electronics and Communication Engineering
Adi Shankara Institute of Engineering and technology
Ernakulam, India
nanduraj.anil@gmail.com

Archana Aniyan
Assistant Professor, Dept. of Electronics and Communication Engineering Adi Shankara Institute of Engineering and Technology
Ernakulam, India
Archana.ec@adishankara.ac.in

Abstract— Healthcare is the most complex and fast moving industry that exists. New technologies are constantly being developed, all with the potential to support the clinical practice by bringing many advantages into the health care sector. Even after these many technological advancement most of the hospitals still resides on traditional paper medical records. Not only it takes up lots of physical space, the interpretation of these medical records is itself a tedious task. Additional to this, odds of losing data is also a major concern. Hence the speech recognition electronic medical record system was introduced. As there lots of speech to text platforms available the need of better recognition system with better speech parameters is a factor. In order to that this project helps in recognizing better speech to text conversion platforms.

Keywords—EMR, Speech Recognition system, Deep Speech

I. Introduction

An electronic medical record includes information about a patient's health history, such as diagnoses, medicines, tests, allergies, immunizations, and treatment plans. Electronic medical records can be seen by all healthcare providers who are taking care of a patient and can be used by them to help make recommendations about the patient's care. Medical record can also be used as communication tool between physicians. Most of the hospital uses paper based medical record system. The advantages of paper-based medical record are it does not need extra money to build infrastructure like network, computer hardware and software, does not require someone to operate the computer, it does not need lot of electricity, easy and cheap. But, the paper-based medical record need very big room to save the medical record; hard to retrieve, sort, and access; one paper at one place; easily lost, damaged, dirty or wet.

Electronic Medical Record (EMR) or computer-based medical record is digitized version of paper-based medical record. However, paper medical records were not steadily used until 1900-1920[2]. Medical record, medical chart, and health record are different terms used to describe the documentation of a patient's medical history. The use of EMRs has not only made patient's medical information easier to read and available from almost any location in the world, but also changed the format of health records, and thus changed health care. Unlimited Storage, No loss of Data, Easy interpretation of the medical records were a few of the major breakthroughs of EMR.

The emergence of Speech Recognition EMR helped in creating the medical records much easier as instead of typing, but for creating a medical record using speech, a dedicated room is required with a noiseless environment. This in turn increases the workflow of the doctors as they have to stay in after office hours to complete the work. These are some of the main problems that we would like to solve using our project. Our EMR Model aims on having a speech to text conversion feature while at the same time making the system portable such that the doctor can create a medical record at the time of patient consultation itself.

II. LITERATURE SURVEY

A. Current system

The majority of the focus related to the modernization of medical records is placed on developed countries [1]. However, developing countries are also progressing from paper-based medical records to electronic medical records. The requirements of their systems can be dramatically different from those of the developed countries. Electronic medical records are systems that serve as an electronic version of the patient charts in a healthcare provides medical history of the patient, allowing healthcare providers to track data such as blood pressure or vaccinations over time, identify which patients are due for checkups, and improve the overall quality of care within the practice [6]. It can efficiently keep records for doctor notes, staff assessments, lab results, etc.

Electronic medical record is works based on the step-by-step procedure, the first step is the doctor's voice captured by the microphone then the gathered speech is recognized by a speech recognizer. The recognized voice is converted to text with the help of speech recognition toolkit.

.

III. OBJECTIVE

Missing of records, large space for storage, Unable to completely understand the prescription and overflow of work for the doctors are increasing by degrees in most parts of the world's healthcare system [1]. In order to overcome some of these complications, the methods used as of now which includes shifting from paper records to digital computerized records are currently not that effective, the average typing speed of a normal person is typically 38-40 wpm(words per minute) which is extremely slow and is time consuming. Replacing typing with speech recognition not only increases the benefits of creating a record much easier, it saves up a loads of time so that the Doctor can devote more time towards their patients rather than worrying about the record in itself.

Introducing speech recognition system to EMR tackle these problems regarding the portability and the reduction of workflow, by introducing the best among the available speech recognition toolkit which ensures that the Medical Records are created while the doctor is interacting with the patient making sure that the patient gets better undivided attention from the doctor. Additional to this, making the model mobile boost its efficiency. One of the key features of this model is categorizing the converted text using keywords to make the interpretation of the Medical Record much easier.

IV. SIGNIFICANCE OF SPEECH RECOGNITION

Using speech recognition in health care transcription services is now well accepted. The compelling need for the technology has been clearly documented in radiology. Radiologists often need to handle high volumes of studies daily, particularly in chest exams. Hand-typing the exam report take a large amount of time and slows down the turnaround time for getting the reports to the referring physicians [2]. In healthcare services, a speech application enables the doctor to enter the clinical findings through voice instead of using a keyboard towards the improvement of patient's care.

The use of speech recognition in healthcare would improve the effectiveness of communication between doctor and patient. It also speeds up the data capture of clinical findings and eliminates errors. By eliminating the need to type clinical findings into EMR system, the duration of doctor-patient encounter could be lessened and healthcare services would be improved

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format [3]. Rudimentary speech recognition software has a limited vocabulary of words and phrases, and it may only identify these if they are spoken very clearly. Research on automatic speech recognition began in the early fifties with the attempt to extract the significant features from acoustic data and to classify and recognize them, by using methodologies developed in the area of pattern recognition. Later, in the seventies, the artificial intelligence technologies were applied for the design of speech understanding systems. An important motivation for the research in this area is the attractive perspective of gathering a deeper understanding about the complex mechanism underlying human perception of spoken language and the characterization of speech sounds in terms of physically detectable features in the brain. The research oriented towards this goal is of great utility in psychology and in the field of development of hearing aids for the handicapped. Moreover research in speech understanding offers a good opportunity for investigating complex parallel processing systems capable of modeling human perception.

V. SPEECH RECOGNITION SYSTEM

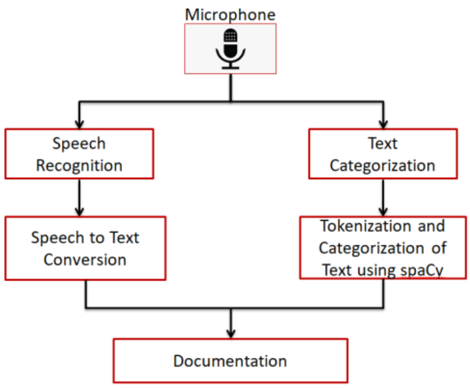

Figure 1: Flow diagram of the proposed framework.

A. Using Google cloud

So first of all we imported speech recognition as sr because it is the library that helps in performing speech recognition in python. It supports several APIs like Google Cloud, IBM Watson etc.

Before importing the speech recognition, PyAudio was installed as it helps the user to use Python to play and record audio on Windows. We initialize the recognizer to recognize our voice. The microphone is given as the source so that we real time transcription is possible.r.listen (source) command is used so that it listens to the source and stores

the audio but there are exceptions in which the audio might not be recognized properly so try and except blocks were used to overcome this limitation. For the API, we used Google cloud. Using Google cloud platform 5 hours of continuous speech examination were done so as to understand the accuracy as well as other parameters for this particular speech recognition API. After getting the results, the 4 major speech parameters like Word Error Rate (WER), Word Information Lost (WIL), Match Error Rate (MER) and finally accuracy. For clean speech Google cloud performed quite well, with WER 21.4%. After mixing five different level of environmental noise, the Google cloud showed the same WER at around 19-22% .The 5 hours of conversion result shows significantly different accuracies as well as different dependencies on the specific input provided. Approximately 87% accuracy has been observed.
Based on our study about Google cloud, some of the difficulties that occurred during the efficiency test was that the Google Cloud consumed more time to transcribe, but the project demands less time consumption.

B. Using Deep speech

Mozilla Deep speech came out for minimize that. It is an open source embedded speech-to-text engine which can run in real time .For clean speech Mozilla deep learning presents the exciting opportunity to improve speech recognition systems continually with increases in data and computation. After the 5 hours test using evaluation parameters WIL (0.33), WER (14.2%), MER (20%) and accuracy (85%) show negligible difference compared to the Google cloud. But the transcription time is far better than Google cloud

C. Using IBM Watson

The next real time platform is IBM Watson .It is also faster and accurate speech to text technology .Its automatically transcribe audio from different languages in real-time even from lower quality audio and verity of audio format. IBM Watson gave similar WER (14.29% and 14.81%).These WER are actually very good already. After mixing five different levels of environmental noise IBM Watson failed at certain point. Its global WER is 29.63%, with a word-information-lost rate at 43% (0.43) which is unfortunately high. In highly noisy environment, the WER from transcription by IBM Watson can be more than 100%. While other would be at worst less than 50%.

V. RESULT AND DISCUSSION

Initially to get good speech to text conversion it is important to convert the speech to text with good accurate and good speech parameters. Therefore by comparing all the three speech to text platform we made a comparison using all the speech parameters.
Out of the three speech recognition APIs examined, Google Cloud stands superior as compared with IBM Watson and deepspeech. Due to the fact that after testing the three APIs, Google Cloud performed the best with minimum WER and WER of 14.2 and also with an Accuracy of 87% even in noisy environment as compared to IBM Watson and deepspeech which had a WER of 21 and 29.63 with the accuracy of 85 and 87 respectively.

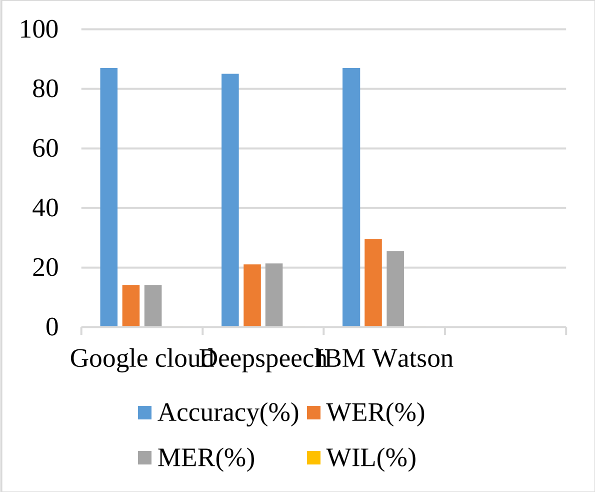

Fig 2: Graph showing different speech parameters comparisons.

Speech Parameters	Google cloud(%)	Deep Speech(%)	IBM Watson(%)
WER	14.2	21	29.63
MER	14.2	21.4	25.4
WIL	0.25	0.3	0.24
Accuracy	87	85	87

Table 1: Different speech parameters value.

VI. CONCLUSION

We live in an era where technology is advancing at an unprecedented rate. Technology plays an important role in every field. To make a system that helps to reduce the workloads of medical officers is required. From the comparison of different speech to text platform we are able to choose better platform for automatic speech recognition. Different speech recognition parameters are evaluated using appropriate mathematical formula.

ACKNOWLEDGMENT

The authors express deep sense of gratitude to department of Electronics and Communication Engineering, Adi Shankara Institute of Engineering and Technology, Kalady,for all the support, assistance and constant encouragements to carry out this work.

REFERENCE

1. M. K. Abd Ghani and I. N. Dewi, "Comparing speech recognition and text writing in recording patient health records," 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, 2012, pp. 365-370,doi: 10.1109/IECBES.2012.6498100.
2. Y. Zhao, "Speech-recognition technology in health care and special-needs assistance [Life Sciences]," in IEEE Signal Processing Magazine, vol. 26, no. 3, pp. 87-90, May 2009, doi: 10.1109/MSP.2009.931993.
3. B. Jada and R. Chania, "An Art of Speech Recognition: A Review," 2019 2nd International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 2019,pp.31-35,doe: 10.1109/ICSPC46172.2019.8976733.
4. D. O'Shaughnessy, "Acoustic Analysis for Automatic Speech Recognition," in *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1038-1053, May 2013, doe: 10.1109/JPROC.2013.2251592.
5. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res. 2010 Aug 6;10:231. doi: 10.1186/1472-6963-10-231. PMID: 20691097; PMCID:PMC2924334
6. S. Hu *et al.*, "The recommender system for a cloud-based electronic medical record system for regional clinics and health centers in China," *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, 2017, pp. 1021-1024, doi: 10.1109/BIBM.2017.8217797
7. P. Blunt and B. Haskins, "A Model for Incorporating an Automatic Speech Recognition System in a Noisy Educational Environment," *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, Vanderbijlpark, South Africa, 2019, pp. 1-7, doi: 10.1109/IMITEC45504.2019.9015907