A Cloud based Medical Transcription using Speech Recognition Technologies

Sushmita Kulkarni[1], Dattaprasad A. Torse[2], Deepak Kulkarni[3]

[1]M.Tech, Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Belagavi, India
[2,3]Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Belagavi, India

---***---

Abstract: *Digital Healthcare has become the most prominent and trending platform for treatment now a days. One such initiative is to build a doctor-friendly digital system. This system will allow doctors to store their patient details, consultations, surgeries performed and many more related information about each and every patient unlike the traditional methods. To build a prototype showcasing digital medical transcription platform which will help surgeons and physicians to document their patients consultations and summary of surgeries performed by recording with a click of a button. Some open source network technologies like uniMRCP, open source EPBX scalable FreeSwitch, standard protocols of Voice Over IP, (i.e. signalling - SIP and audio media - RTP), Speech Recognition engines supporting uniMRCP as Google Speech Recognition or CMU's PocketSphinx are used. The main idea behind is to transform voice recording to a text document to be presented as a part of Electronic Medical Record system using Speech Recognition and Synthesis technologies.*

Key Words: *Cloud, Freeswitch, uniMRCP, Google Speech Recognition plugin, PocketSphinx.*

I. INTRODUCTION

The surgeons are using traditional methods of documenting procedures/surgeries carried out and they are either using paper documentation or typing the procedure/surgery details performed on a patient in a text editor of their choice. Some of these documentations are not completely integrated with patient electronic record systems. In smaller hospitals, the doctors are still using traditional methods of paper documentation for each and every recording of patient history, surgeries performed and other clinical details. The doctors and associated medical staff spend lot of time in documenting and recording patient details during the clinical visits and after the surgery performed by the doctor. Some of these manual and time consuming documentation activities can be digitized using voice activated and cloud based recording and transcription technologies presented in this paper. These technologies work together and help the healthcare providers in digitizing the various documentations seamlessly. As time evolved the methods of handling, the records kept on evolving. Modern times bring to the stage of handling patient records in a digital way from typing long patient-data to Transcribing. Transcription involves technologies of speech to text conversions using Speech recognition technologies.

There are less different solutions reviewed and presented for electronic medical record transcription using networking tools. Although there are many works displaying different solutions for speech to text, conversions using machine learning and AI technologies.

E-Healthcare systems can come to rescue of people during pandemics, natural disasters or at times when patient cannot reach hospitals. The main focus is to enable and evolve digital healthcare platform, in turn bridging the gaps between doctor's community and patients. Thereby becoming doctor's partner in enabling and enhancing the clinical OPD and online consultations using newer technologies.

II. LITERATURE

Transcribing paper-based archives into digital form was helpful step for educators, clinical researchers and people capturing data on fields for research purposes, where records were kept in spreadsheets on regular basis. Usually when it comes to medical records involving confidential patient data, needs to be handled very sensitively. According to Mohammad M. Ghassemi et al. [1] proposes a tool based on machine learning, crowd intelligence, optical character recognition, image segmentation and crowd sourcing. This involved protection of personal information using images of paper-based spreadsheet transcription into digital form. The steps followed in the algorithm were, (1)cell-level image extraction, (2) recognition of digits within the cell using machine learning, (3) uncertain machine transcription content correction by humans, (4) human transcribed content results as a feedback to improve machine classification performance. The limitations here are prolonged process of transcribing, collection of spreadsheet data on regular basis and high-ended machine learning algorithms.

Over a span of 20 years Health Information Technology's (HIT) healthcare awareness and hazards for safety risk analysis has been introduced in medical and healthcare organizations. The main idea was to link events to type of hazards with efficient engineer centric solutions, for safety during adverse situations. Healthcare hazards means losing sensitive patient records if there is less awareness within healthcare. This may influence patient safety corresponding to erroneous output of Medical Information System (MIS) like the Electronic Health Record. Richard W. Jones et al. [2] highlights prevalence of indirect hazards and regulatory standard measures implemented during deployment of addressed problem. This research not only focuses on identification and removal of the risk but also indirect ones after deployment. The problem is addressed in three ways.(1)modified MIL-STD-882E addresses currently existing deficiencies when user makes executable decisions. It defines risks associated with erroneous Information System (IS) system failure modes, Software Control Categories (SCC) hazard severity table to get Software Criticality Index (SwCI), whose outputs are given to Level of Rigor (LOR). (2)Health applications (mHealth App) risk assessment. (3)Generic 8-Step IS safety management Process adapted to applications. It is very important to have a hazard-free transcribing system; hence, there is always a priority for patient data safety.

Voice over Internet Protocol (VoIP) and Electronic Private Branch Exchange (EPBX) are cost effective methods unlike the traditional. Asterisk is an Open Source is a Linux based server and Private Branch Exchange framework that allows a user to have a phone-system of one's choice because of it flexibility to customize modules. Mohammed Abdul Qadeer et al. [3] implements an Asterisk server within a local Wi-Fi network and Public Switch Telephony Network (PSTN) for registered devices within University usage. The application architecture model involved Asterisk server, a Client and PSTN Exchange for placing a voice and video based call over a private Wi-Fi cloud.

A distributed application Real-time online interactive application (ROIA)is a Cloud environment emerging in large scale. Additionally have issues like scalability and network latency. Previous researchers tried and focused on mixed deployment of ROIA and extension of ROIA. LIU Dong, et al. [4] proposes a system in which a new technology MRCP is deployed that overcomes network latency issues, scalability of ROIA in cloud computing. The solution focuses on MRCP architecture and external balance strategy to overcome fluctuations of concurrent users and network latency requirements. The MRCP architecture has ROIA Servers (RS) and one MRCP Local Controller (MLC) for each data centre distributed across the world. MLC and RS responsibility is load balancing, storage, zoning and instancing.

FreeSwitch acts as a PBX (Private Branch Exchange) server, open source scalable soft-switch. It follows client-server architecture. Abdullah Mohammad Ansari et al. [5] implemented Interactive Voice Response System (IVRS) model for Session Initiation Protocol (SIP) based phones. Backbone of this application is scalable FreeSwitch connects to SIP-based soft phones either desktop or mobile client as a FreeSwitch server. The SIP registers themselves as client to FreeSwitch servers, which in turn has information of all registered clients, and other connected FreeSwitch servers. The idea of accessing information in web browser of phone while on call could save time and make this a more reliable approach.

Process of making a computer system understand what we speak is nothing but computer speech recognition or interpretation of voice in the form of text. There are many such recognition of speech software for appropriate speech to text conversions. Aditya Amberkar et al. [6] proposes a Speech to Text using Speech Recognition and Recurrent Neural Networks (RNN) based speech recognition model for prediction. Initially the speech, which is an analog signal is digitized or sampled by Nyquist theorem and pre-processing of the signal to 20-millisecond chunks is done. This pre-processed data is fed to RNN. The application of RNN increases performance accuracy in much speech to text conversion engines like Java, Python based snowboy hot word detection, C, CMU pocket-sphinx. Amazon's Alexa and Google's STT are online speech to text engines whereas CMU pocket-sphinx is offline conversion engine but training the dataset is done online. Although training the RNN algorithm is complex, it results as best algorithm for speech processing and voice controlled technologies.

Worldwide, commercial applications are having high demand for Automatic Speech recognition (ASR) but in India, it is still evolving. Chadalavada Sai Manasa et al. [7] developed acoustical model for the speech recognition in Hindi using CMU's PocketSphinx with a database of 177 words and dictionary of cross language adapted for speech recognition such as English.ASR model is based on Gaussian Mixture Hidden Markov model(GMM-HMM) based acoustic modelling using LPC and Mel Frequency cepstral coefficients (MFCC) for feature extraction. PocketSphinx is lightweight, free and real-time continuous medium vocabulary Speech Recognition system developed for hand held devices.

Freeswitch is a highly scalable engine for routing, interconnecting communication protocols for any type of media namely audio, video or texts and is a cross platform telephone exchange that bridges business solution gaps. It uses embedded languages like Lua or JavaScript that makes it more flexible. Wei Tang, et al. [8] introduces a soft switch solution i.e. FreeSwitch for efficient communication dispatching and accuracy in information using IMS architecture and SG-UAP based application.

Sila Chunwijitra, et al. [9], propose a cloud based framework for speech recognition in Thai language. They also deploy Docker (lightweight Linux container) platform to migrate baseline Distributed speech recognition (DSR) system. The main idea here is to improve response time in real time using cloud computing. Furthermore, the workflow is modified by paralleling running multiple Speech Recognition (SR) Engines with help of utterance decoding. Then on Word Error Rate(WER)is computed and results seem to be scalable and reliable with no significant difference between proposed and baseline approaches. Hence overall performance is boosted with cloud computing benefits and improved response time in terms of real-time factor (RTF).

Resource Sharing is the benefit of using cloud-based web services. Sila Chunwijitra, et al. [10] focuses on distributing and sharing resources for Automatic Speech Recognition(ASR) applications. In case of Transcription, ASR needs more resources as many utterances must be handled in real time computing. For this key solution is scaling ASR by multithread processing, exploiting multiplexing and demultiplexing technique to network socket or distributing ASR in real-time streaming or distributing engines (load balance). This proposed work reduces RTF by 15% of the improved framework when compared to the baseline system architecture and shares lesser resources like working memory.

"Google Cloud Speech API" is a Speech-to-Text and Text-to-speech converting Google service, whose speech recognition accuracy is high due to its deep learning neural network algorithms. The algorithms do not require high performance processors because everything is processed in cloud. Gustavo Boza-Quispe, et al. [11] proposes an user friendly speech interface to access tourist semantic information based on Google Cloud Platform. The flow has stages like Text-to-Speech(TTS) and Speech-to-Text(STT) Converter, Web Interface, SPARQL Generator and Semantic Representation.

Due to increased adoption of smart phones and other consumer devices speech has become one of the modes of interaction. Yanzhang He, et al. [12] focuses beyond acoustic (AM), pronunciation (PM), and language (LM) models) satisfying computational and memory constraints improved in earlier large vocabulary continuous speech recognition (LVCSR)systems of ASR. Their model throws 20% improved WER over a embedded baseline system because of the E2E speech recognizer based on the RNN-T model. This model runs double as fast as Google Pixel phone.

III. PROPOSED SYSTEM

The SIP client is integrated with an Android APP that initiates voice activated conversation with the doctor or surgeon intended to document the procedure or surgery performed on a patient. Typically, the doctor will start the conversation with patient ID and then starts speaking as he or she normally does over the regular phone call. The FreeSWITCH IVR system records the speech of the doctor which is essentially the document that is supposed to be typed in a traditional way. The recorded speech or audio is then processed by the server application that communicates over SIP and MRCP signaling protocols. The voice is transmitted between FreeSWITCH and MRCP server via RTP packets. The MRCP server uses either Google Speech Recognition (GSR) cloud based API or PockSphinx module for transcribing the doctors speech to a text. The transcribed text is returned by the GSR or PocketSphinx is stored in the respective patient record database for presentation as a part of patient Electronic Medical Record flows during clinical visits or reviews conducted by the doctor or surgeons in subsequent follow-ups. The transcribed documentation could be made available and viewable as a plain text at any time once the dictated document has been transcribed in near real-time basis. The proposed system is as shown in Fig-1. Steps involved in the back-end

1. Call IVR-Internal extension (eg:1000)
2. Announce patient ID
3. Start audio recording or voice mail option
4. Store audio file(as patient_id.wav)
5. notify/send command to Uni-MRCP Server
6. Initiate Speech to Text / enable speech-to-text API on Google Cloud Platform
7. Store Text file

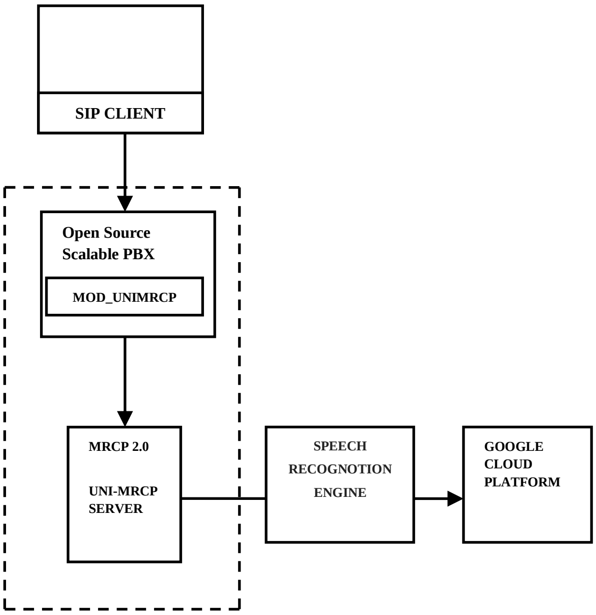

Fig-1: Block Diagram of Cloud Based Medical Transcription

IV. CONCLUSION

In this paper, emerging technologies that offer flexible resources(cloud), Open source cross platform telephone performing multiple functions(Asterisk, freeswitch), and speech engines (ASR, Google Speech API, Pocketspinx)are reviewed. FreeSwitch is an open source EPBX like the Asterisk but is more at its flexibility and abundance to scale and add modules as per the choice of the user, which makes it a reliable and convenient platform summarized as one machine doing multiple tasks. Therefore, our proposed model uses technologies like uniMRCP, open source EPBX FreeSwitch, Speech Recognition engines supposing uniMRCP as Google Speech Recognition or CMU's PocketSphinx for more accurate results. Here networking tools and emerging open source technologies are utilized. The doctor will save the time in manually documenting the patient information after the surgery by dictating the notes using android application and transcription platform. Hence, the proposed system is cost effective, reliable and most importantly can be implemented on cloud and need not use any system resources for storage. As the proposed solution is cloud based, doctor can access from anywhere, through any device having internet, and store unlimited patient data.

REFERENCES

[1] Ghassemi, Mohammad M., et al. "An open-source tool for the transcription of paper-spreadsheet data: Code and supplemental materials available online: Https://github. com/deskool/images to spreadsheets." 2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017.

[2] Jones, Richard W., and James E. Mateer. "Indirect risk related failures of Medical Information Systems." 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2019.

[3] Qadeer, Mohammed Abdul, Kanika Shah, and Utkarsh Goel. "Voice-video communication on mobile phones and PCs' using asterisk EPBX." 2012 International Conference on Communication Systems and Network Technologies. IEEE, 2012.

[4] Liu, Dong, and Yue-Long Zhao. "A new approach to scalable ROIA in cloud." 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies. IEEE, 2013.

[5] Ansari, Abdullah Mohammad, Md Faisal Nehal, and Mohammed Abdul Qadeer. "SIP-based interactive voice response system using freeswitch epbx." 2013 Tenth International Conference on Wireless and Optical Communications Networks (WOCN). IEEE, 2013.

[6] Amberkar, Aditya, et al. "Speech Recognition using Recurrent Neural Networks." 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018.

[7] Manasa, Chadalavada Sai, K. Jeeva Priya, and Deepa Gupta. "Comparison of acoustical models of GMM-HMM based for speech recognition in Hindi using PocketSphinx." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.

[8] Wei Tang, Wei, et al. "Design and implementation of information and communication dispatching system based on FreeSwitch platform." Journal of Physics: Conference Series. Vol. 1449. No. 1. IOP Publishing, 2020.

[9] Chunwijitra, Sila, et al. "A cloud-based framework for Thai large vocabulary speech recognition." 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE, 2016.

[10] Chunwijitra, Sila, et al. "Distributing and Sharing Resources for Automatic Speech Recognition Applications." 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA). IEEE, 2019.

[11] Boza-Quispe, Gustavo, et al. "A friendly speech user interface based on Google cloud platform to access a tourism semantic website." 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON). IEEE, 2017.

[12] He, Yanzhang, et al. "Streaming end-to-end speech recognition for mobile devices." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.