An Automatic Voice Prescription Using Viterbi Algorithm

Swapna A[1] ,Vedavathi K[2] ,Reddy Sai Pavan Keerthi Teja[3]
[1, 2, 3] Dept of Computer Science and Engineering
[1, 2, 3] Mahatma Gandhi Institute of Technology, JNTUH

Abstract- *Drug side effects are the leading problem of death in the world, and drug or prescription mistakes cause tens of thousands of mortality worldwide each year. Most of these mistakes are made by caregivers taking the incorrect medication or dosage due to obscure handwriting, drug interactions, confusing medicine names, etc. This work proposes a framework that signifies use of voice recognition techniques for medicine prescription. The goal is to avoid wrong medication for normal ailments viz. fever, cough, cold, body pain etc. by designing virtual application on voice-based medicine prescription.In here a doctor can fill the details of a form like problem, medication, gender, weight etc using voice. Then he can sign the form digitally for authenticity and fill in the hospital details.Now the form filled is converted to PDF format and send to the respective patient email for his future reference. This method can save wealth and life in medical centers throughout the world, particularly in developing countries where the practice process is generally cumbersome and paper based. In India, inadequate drugs kill thousands of people and normal illness causes severe disability.The system proposed is targeted to those doctors and clinics that are still using paper-based handwritten prescriptions and cannot afford the existing Electronic Health Record systems available.*

Keywords- Hidden Markov Model (HMM), Viterbi algorithm, Automatic Speech Recognition (ASR), mel-frequency cepstrum

I. INTRODUCTION

A major issue in India is that generally prescriptions are still written by hand and the readability of such handwritten prescriptions is very limited. Several cases have come to light, where an error in understanding the prescription by a chemist has led to the wrong medication, which caused severe health issues to the patient[1]. But usage of the traditional Electronic Health Record System(EHR) in generating an electronic prescription is tedious and a bit costly for a doctor in India. A proper infrastructure would be needed for setting up such a system. Also, the operating cost and time for the system would also add up. Either the doctor would operate the system on his own consuming time or an operator if employed would cost money. Furthermore, patient record addition and retrieval consume 49 percent of a doctor's time in using an EHR system[2].

The requirement to enhance healthcare and provide effective healthcare has exceedingly improved the use of technology solutions in the healthcare sector. The Internet has received significant attention in recent years, but the voice is still a common convenient and direct way to communicate person-to-
person or person-to-computer. Google Speech API technology enables users to communicate with the Internet utilizing speech
recognition technology simultaneously with voice commands. Google's API is a new standard for developing major voice and
pitch control applications. This technique decreases the expenses and develops writing efficiency. Voice assist systems are incorporated in various fields, including data providers such as warehouses, air travel, financial institutions [3], customer service [4], e-learning [5], and various areas.Security of the prescription and doctors notes is possible by directly delivering the prescription to the user's email. Having discovered somespecifications and objectives for the system, in the following section we shall discuss the modeling and design of the voice-based prescription

II. METHODOLOGY

A. System Architecture:

Fig 1: data flow of voice prescription

This is implemented in Visual studio code editor. In here first the doctor given input to our applicayion using microphone and the signals from the input are given to the hidden morkov model using google API call.The signals are then trasformed using Mel Frequency Cepstral Coefficents,which then converts our signals into spectogram and then acoustic model,languagemodel,pronounciation model are applied to it to convert into text.Then the text is filled into GUI fields and then hospital,Doctor details are added to the form .Then the from is digitally signed and coverted into PDF format for sending the PDF to the patient gmail for future reference.

B. Graphical User Interface and Speech Recognition Module :

In here the GUI for the prescription is built using tkinter and the GUI contain fields ,buttons and text box etc.The labels in our GUI are Name, Phone Number, Age, Gender, Weight, Problem, Medication.The user can manually fill this text box or by using his voice.To fill the necessary input box or text box a user needs to use a microphone.In here we are using Google cloud speech API to send the input audio to HMM model present in google cloud .

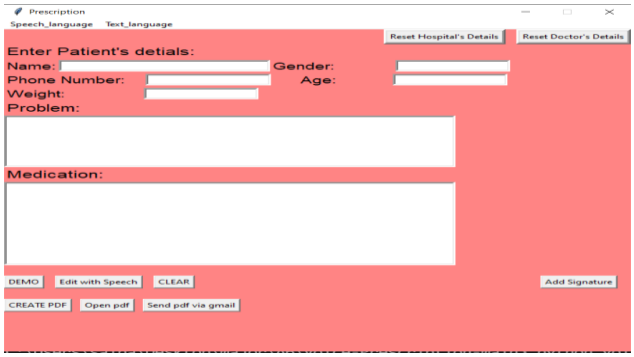

Fig 2: Graphical User Interface for Prescription using tkinterlibrary

The following process takes place in converting from speech to text:

Let's get a high-level overview first. The diagram below is a high-level architecture for speech recognition that links HMM (Hidden Markov Model) with speech recognition.

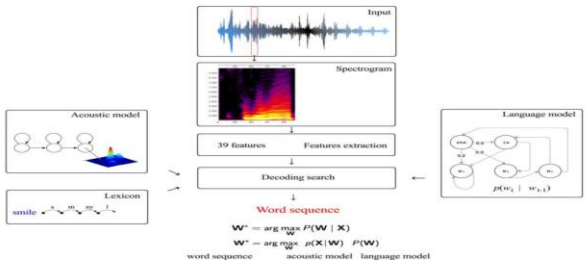

Fig 3 :High-level architecture for speech recognition that links HMM with speech recognition.

The first component of speech recognition is, of course, speech. Speech must be converted from physical sound to an electrical signal with a microphone, and then to digital data with an analog-to-digital converter.

In HMM, the speech signal is divided into 10-millisecond fragments. The power spectrum of each fragment, which is essentially a plot of the signal's power as a function of frequency, is mapped to a vector of real numbers known as cepstral coefficients. The dimension of this vector is usually small—sometimes as low as 10, although more accurate systems may have dimension 32 or more. For This mel-frequency cepstrum technique is used.

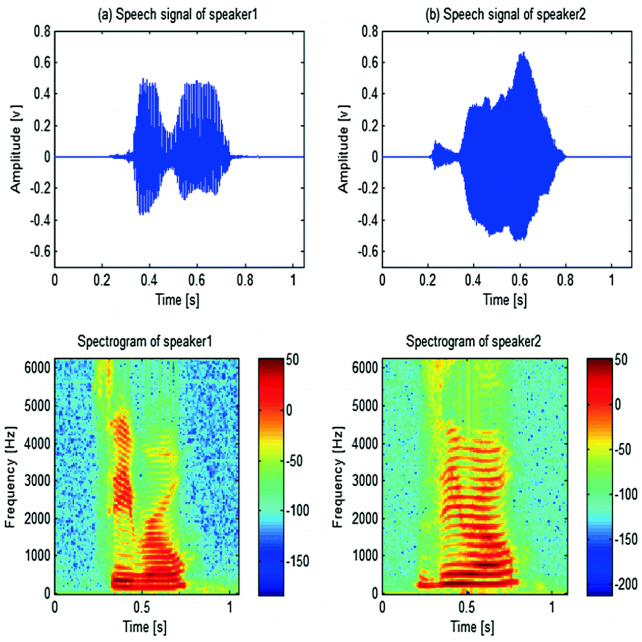

Fig 4: Converting voice signal to spectrogram using MFCC

A pronunciation model can use tables to convert words to phones, or a corpus is already transcribed with phonemes already. The acoustic model is about modeling a sequence of feature vectors given a sequence of phones instead of words.The language model is about the likelihood of the word sequence. For example, “I watch a movie” will be more likely than “I you movie watch” or “I watch an apple”. It predicts the next word given the previous words. If we approximate it with a first-order Markov chain, the next word will depend on the current word only. We can estimate it by counting the occurrence of word pairs in a corpus.

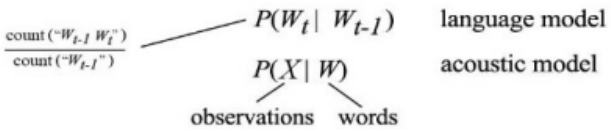

Fig 5: Formula for predicting next word given the previous word

The distribution of features for a phone can be modeled with a Gaussian Mixture Model (GMM). We will learn it with training data. The transition between phones and the corresponding observable can be modeled with the Hidden Markov Model (HMM). So if we can find an optimal way to search the phone sequence efficiently, this may not sound bad after all.

An HMM model composes of hidden variables and observables. The top nodes below represent the phones and the bottom nodes represent the corresponding observables (the audio features). The horizontal arrows demonstrate the transition in the phone sequence for the true label “she just …”.

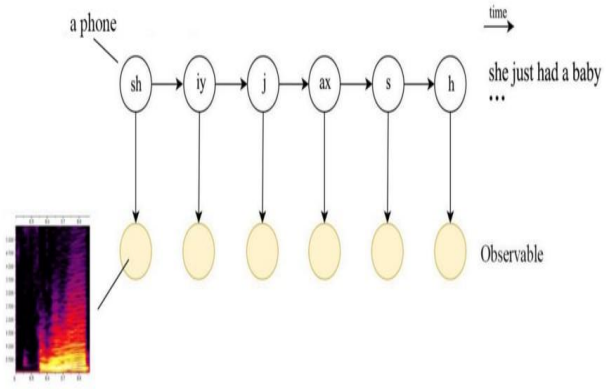

Fig 6: Diagram showing the Transition in Phone sequence of a sentence

With the Viterbi algorithm we can find the optimal sequence in polynomial time for finding all the combinations of phonemes

Viterbi algorithm:

we want to express our components recursively. Given the state is j at time t, vt(j) is the joint probability of the observation sequence with the optimal state sequence.So not only it can be done, the equation is similar to the forward algorithm except the summation is replaced by the maximum function. Instead of summing over all possible state sequence in the forward algorithm, Viterbi algorithm takes the most likely path.Finding the internal states that maximize the likelihood of observations is similar to the likelihood method. We just replace the summation with the maximum function.In this algorithm, we also record the maximum path that leads to each node at time t (the red arrow above), i.e. we backtrace the optimal path for each node.

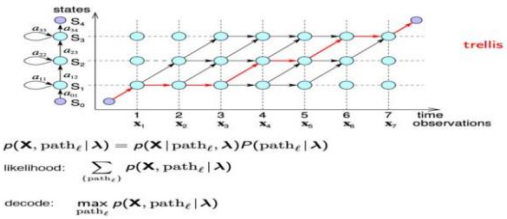

Fig 7:States vs Time Observation graph where x-axis represents time observations and y-axis showing states

C. Custom PDF and Authenticity module :

In this module a custom pdf is designed according for our needs and all the neccessary fields like name, age, gender, weight, problem, medication are constructed and before creating the PDF it aslo checks whether all the fields in the GUI are filled by the user .If the fields are not filled it will inform the user to fill the fields which are neccessary.When all the requirements are satisfied then the information is taken fromo the user interface and then it is converted into PDF.

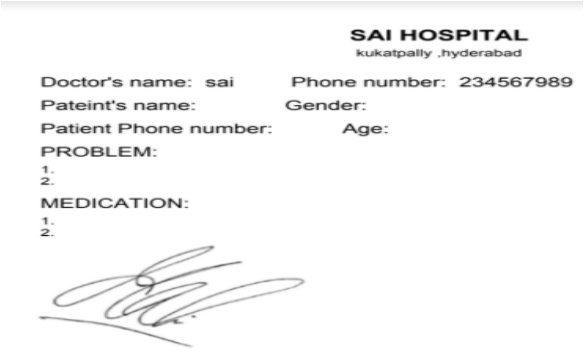

Fig 8:GUI for authentication and Custom PDF

D. Sending Mail module :

In this module the user can send the PDF to the patient using smptlib which is simple mail transport protocol.In order to send binary files to an email server that is designed to work with textual data, they need to be encoded before transport. This is most commonly done using base64, which encodes binary data into printable ASCII characters.In here The MIME multipart() message accepts parameters in the form of RFC5233-style key/value pairs, which are stored in a dictionary and passed to the .add_header method of the Message base class.

After doing thisyou will have to adjust your Gmail account's security settings to allow access from your Python codeTurn Allow less secure apps to ON

III. RESULTS

In here we tested the how accurately our system captures the voice of the doctor (Note: The test is conducted in a natural environment which is commomly referred as noisy environment) and sends the PDF to the patient. For this the application is given input several times and checked how good it is doing. The below figures shows the snapshots of our input for Name field which includes both accurate and error occured when taken input from user.

Fig 9 : Correct and Incorrect examples for voice recognition

TABLE I Accuracy Obtained after tesing

Total Tests conducted	Correctly Recognized the voice	Incorrectly Recognized the Voice	percentage of accuracy
50	42	8	84

The below figure hows the output generated when the form is converted into the PDF

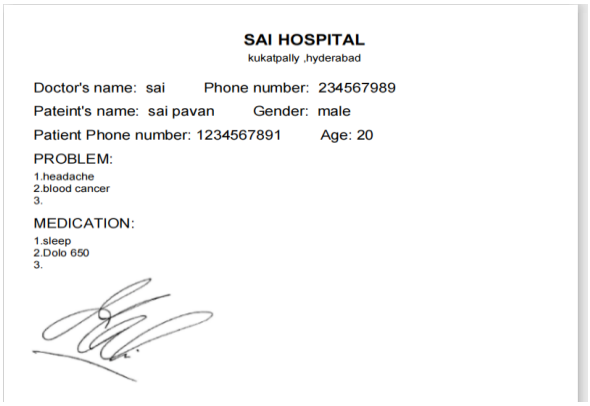

Fig 10 :Example of GUI that is converted into PDF

The below figure shows the result for successfully sending the email

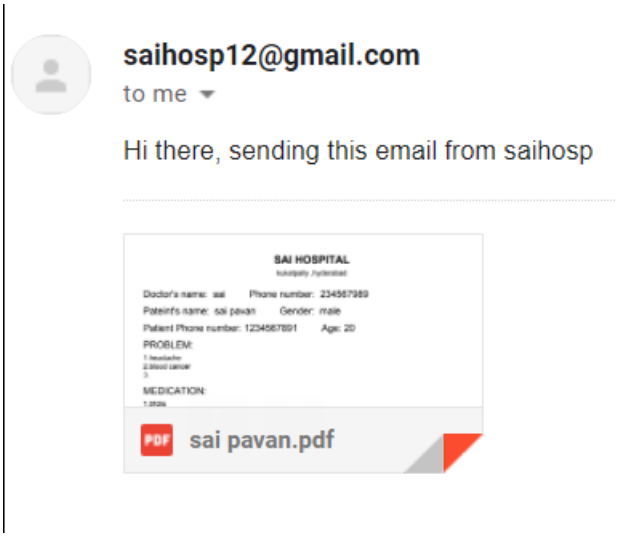

Fig 11 :Conformation for sending email to the patient

IV. CONCLUSION

The proposed and implemented system reduced the amount of time consumed in creating. This system is available to all the doctors of India through their computers. We have implemented an innovative solution to solve the problem of illegible handwritten prescriptions. Voice-based e-prescription needs a minimal change in the workflow of a doctor but in the long run, it will create a huge impact in developing a digital ecosystem for patients. The implemented system will reduce the patient record access time and maintain high security and privacy of patient data.

V. FUTURE WORK

For future work,we are trying to include different language for taking input(So that it can be used by users who can speak in there mother tongue providing the comfort for them and also the choice of choicing different languages from

the choices given) and also develop a android app for easy access through smart phones. we are planning to integrate and use the system in the realhospital ecosystem to test and validate the implementation and to analyze the impact it will create in the healthcare domain.

REFERENCES

[1] C. Sinsky et al., ―Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties,‖ Ann. Intern. Med., vol. 165, no. 11, p. 753, Dec. 2016

[2] Atayero A. A, Ayo C. K. Ikhu-Omoregbe N. A,Azeta, A. A. (2009), "Implementation of 'ASR4CRM': An Automated SpeechEnabled Customer Care Service System", IEEE Xplore. 2009. Eurocon'09 IEEE International Conference, Saint-Petersburg, Russia. May 18 – 23 2009. pp. 1712-1715.

[3] Azeta A. A., N. A. Ikhu-Omoregbe, C. K. Ayo & A.A. Atayero (2008), "Development and Deployment of VoiceXML-Based Banking Applications", Journal of Computer Science & its Application, An Internationa l Journal of the Nigeria Computer Society (NCS), Vol. 15 No.1 June Edition, pp. 59-72

[4] D. K. Sokol and S. Hettige, ―Poor handwriting remains a significant problem in medicine,‖ J. R. Soc. Med., vol. 99, no. 12, pp. 645– 646, Dec. 2006

[5] Akhil Patel, "A Study of Speech Recognition" - International Journal of Creative Research Thoughts

[6] Vedavathi K and A Swapna" An Approach to Identify Human based upon Teeth Recognition using ANN"Vol.12 No.9 (2021), 674 –677,Apr.2021