IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Automated Appointment Booking System via Voice Recognition

Aditi Namdev
Information Technology
SIES Graduate School of technology
Mumbai,India

Ritvik Pande
Information Technology
SIES Graduate School of technology
Mumbai,India

Praveena Viswarajan
Information Technology
SIES Graduate School of technology
Mumbai.India

***Abstract*—Appointment booking is a part of everyday life nowadays. From booking movie show tickets to booking flight seats everything is online. In recent times the pandemic has brought some tough challenges before the healthcare systems which include appointment booking. As hospitals and clinics witness an overwhelming surge of patients the healthcare workers are found to be overburdened. Oftentimes important calls made by patients to the hospitals go unanswered or result in long wait times. Speech is the primary mode of communication among human beings. Many patients would prefer to call-in and book an appointment before visiting the crowded hospitals. This paper is focused on developing an AI voice bot which works on call to communicate with patients and subsequently book appointments and/or answer questions related to that hospital. Automatic Speech Recognition (ASR) is the process of deriving the transcription of an utterance, given in the speech waveform. This paper aims at completely automating the process of booking appointments in general, all through voice recognition. Instead of having to go through contact lists and learn different UIs, patients can simply have a conversation with the bot through a familiar interface. Not only will it book an appointment but also study the patterns in the database and predict future inflow of patients and suggest constructive advice for the system. This will not only save patient's critical time, but also reduce the burden on healthcare workers and hence optimize the management of patients and their appointments.**

I. Introduction

Alternatively referred to as speech recognition, voice recognition is a computer software program or hardware device with the ability to decode the human voice. Voice recognition is commonly used to operate a device, perform commands, or write without having to use a keyboard, mouse, or press any buttons. Today, this is done on a computer with ASR software programs. Many ASR programs require the user to "train" the ASR program to recognize their voice so that it can more accurately convert the speech to text. For voice recognition to work, you must have a computer with a sound card and either a microphone or a headset. Other devices like smart phones have all of the necessary hardware built into the device. Below are some examples of where you might encounter voice recognition.

- Automated phone systems - Many companies today use phone systems that help direct the caller to the correct department. If you have ever been asked something like "Say or press number 2 for support" and you say "two," you used voice recognition.
- Google Voice - Google voice is a service that allows you to search and ask questions on your computer, tablet, and phone.
- Digital assistant - Amazon Echo, Apple's Siri, and Google Assistant use voice recognition to interact with digital assistants that helps answer questions.
- Car Bluetooth - For cars with Bluetooth or Handsfree phone pairing, you can use voice recognition to make commands, such as "call my wife" to make calls without taking your eyes off the road.

II. Need of Project

On the surface, the process of scheduling appointments and managing individual contact information seems simple enough. However, a more in-depth analysis of an organization's scheduling processes will reveal just how ineffective and costly this task can be. It is, therefore, in an organization's best interest to examine the way it currently manages its scheduling and take into account certain considerations affected by the scheduling process. These considerations include:

1) Staff resources needed for managing appointments
Scheduling appointments and reservations in the traditional manner requires manpower, whether it's the receptionist answering the phone, an administrator managing a paper appointment schedule, or a support staffer making reminder phone calls. This manual management of appointments and reservations has the potential to drive up operating costs, such as overtime pay or the need to bring on an additional employee to either manage the booking process or assist in tasks other employees are unable to complete because most of their attention is directed at the scheduling process. A quick breakdown of daily staff responsibilities and the amount of time devoted to each can reveal the level of employee focus on managing appointments and reservations.
2) Time needed for managing appointments the scheduling process at many organizations may take only a few minutes to complete, which may not seem significant in the overall scheme of things. But multiply that by the number of daily appointments or reservations—which

can total in the hundreds or thousands for larger organizations—and this quick task morphs into a time-consuming dilemma for supervisors, managers and staff. This time requirement may force organizations to spend less time on more pressing task or inflate their operating costs by requiring additional employees or overtime pay.

3) Limited hours of operation Unless an organization is open and staffed 24 hours a day, there is a limited window for individuals to schedule their appointments and reservations. This not only presents a challenge for people scheduling—as they may not have the means to contact the organizations during normal operating hours—it also equates to missed sales when they are closed for business.

4) Inconvenience for customers Not offering an online alternative to picking up the phone and calling in an appointment is an inconvenience to customers in today's Internet-connected society. Customers, clients, students and patients not only desire the ability to conduct transactions—like booking appointments or reservations—online, they also expect it. In some instances, they may look elsewhere for a competing organization that does offer it.

5) Advanced Functionality Features found in online scheduling software applications that can completely automate, streamline and improve the appointment- and reservation-booking process are simply unavailable in the traditional scheduling process

III. OBJECTIVES

The main objective of the project is to create a voice assistant to recognize the users voice via system and accordingly book appointments and alter an existing database and further generate receipts and bills. The advantages due to developed incorporated features in our project are: Offers easy assistance that is the voice bot will have a clear input-based output/response. Normal call-based appointment booking involves unnecessary phone calls with no result. It also gives way to spamming and disrupting privacy of the organization. Eliminates problems related to incoming of more than one call at once. Unnecessary liability in the name of a receptionist is diminished. Also saving the capital required. Further perform data analysis on generated data and discover patterns – Proper data analysis will be done on the generated data and new patterns will come into light along with prediction of future appointments.

A. Abbreviations and Acronyms

- ASR Automatic speech recognition
- MFCC Mel Frequency Cepstral Coefficents
- CTC Connectionist temporal classification
- RNN Recurrent Neural Network

B. Methodology

Speech processing system has mainly three tasks

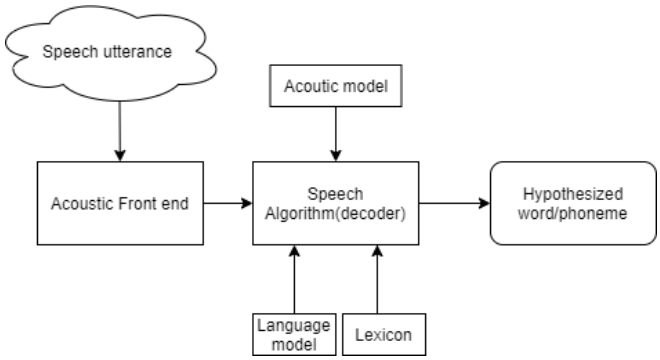

Fig. 1. Voice Recognition Architecture

- First, speech recognition that allows the machine to catch the words, phrases and sentences we speak.
- Second, natural language processing to allow the machine to understand what we speak, and
- Third, speech synthesis to allow the machine to speak.

ASR system has two important tasks—phoneme recognition and whole-word decoding. In ASR, the relationship between the speech signal and phones is established in two different steps [1]. In the first step, useful features are extracted from the speech signal on the basis of prior knowledge. This phase is known as information selection or dimensional reduction phase. In this, the dimensional of the speech signal is reduced by selecting the information based on task-specific knowledge. Highly specialized features like MFCC [2] are preferred choice in traditional ASR systems. In the second step, discriminative models estimate the likelihood of each phoneme. In the last, word sequence is recognized using discriminative programming technique. Deep learning system can map the acoustic features into the spoken phonemes directly. A sequence of the phoneme is easily generated from the frames using frame-level classification.

The three major types of end-to-end architectures for ASR are attention-based method, connectionist temporal classification (CTC), and CNN-based direct raw speech model.

C. Deep Learning Model:Convolutional Neural Network

This model consists of two stages: feature learning stage, i.e., several convolutional layers, and classifier stage, i.e., fully connected layers. Both the stages are learned jointly by minimizing a cost function based on relative entropy. In this model, the information is extracted by the filters at first convolutional layer and modeled between first and second convolutional layer. In classifier stage, learned features are classified by fully connected layers and softmax layer. This approach claims comparable or better performance than traditional cepstral feature-based system followed by ANN training for phoneme recognition on TIMIT dataset.

D. Implementation

Traditionally speech recognition models relied on classification algorithms to reach a conclusion about the distribution of

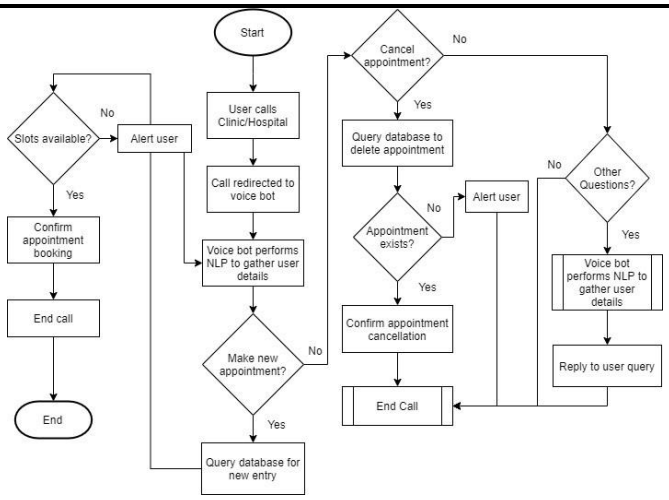

Fig. 2. Appointment Booking Process Flow Chart

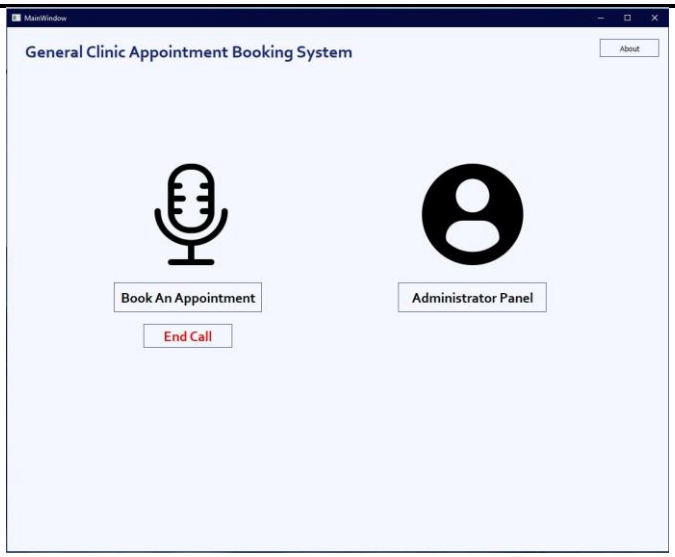

Fig. 3. Voice Input via Application

possible sounds (phonemes) for a frame. Speech recognition software uses Natural Language Processing (NLP) and deep learning neural networks to break the speech down into components that it can interpret. It converts these components into a digital state and analyzes segments of content. The software trains on a dataset of known spoken words or phrases, and makes predictions on the new sounds, forming a hypothesis about what the user is saying. It then transcribes the spoken words into text. For independent makers and entrepreneurs, it's hard to build a simple speech detector using free, open data and code. Many voice recognition datasets require preprocessing before a neural network model can be built on them. To help with this, TensorFlow recently released the Speech Commands Datasets. It includes 65,000 one-second-long utterances of 30 short words, by thousands of different people.

1) Speech to Text: We will be working with tensor flow and numpy libraries in advancement along with python to implement speech to text recognition of user. • TensorFlow allows you to build neural network models to recognize spoken words. These models typically use the Recurrent Neural Network (RNN) architecture which processes inputs organized as a sequence. TensorFlow supports all RNN variants including static RNN with a uniform length for all input sequences, dynamic RNN with the ability to have inputs of different lengths, and static bidirectional RNN • NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed.
2) Data Analytics: Python along with Google Cloud would be used for complete data analysis.This language can also be used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization.
Libraries such as Pandas,Numpy and seaborn are a few which have been used for data analytics. Complete preprocessing of the datasets is performed which includes removing duplicate values and dropping null values.Further data visualization is done via various types of graph forms including histograms and counter plots. Value ratio based on attributes like gender and neighbourhood is also useful in presenting the data in an efficient way.

The user makes a call and gets redirected to the bot. After making introduction and giving guidelines the user commands for a booking. The bot performs Natural Language processing and extracts keywords to perform query on database. Finally, the query is performed and the booking details are entered and the corresponding clinic is informed. On the other hand, if keywords fail to be extracted the booking is not made and the call is ended. If NLP fails the user is redirected towards the corresponding hospital's clinic for a human conversation. Further if a cancellation is made then the specific keyword is used as identifier and details of the booking which is to be cancelled are asked for. Then the query is run on the linked database and the booking is deleted.

The process is same as explained in the flowchart. The user that is the patient will call at the clinic then there will be a bot interaction which will be connected through a channel that is the real time call network connection. The message will be converted by the bot from speech to text through NLP following the CNN model first followed by the bot segregating the keywords. The bot as soon as it finds the appropriate keyword with date and time will forward data as query to the database. The keyword would be converted into a query and will access the database. After checking the availability i.e if there is a vacancy in the slot the bot will again get back to the user for confirmation. If not, it will ask the user to

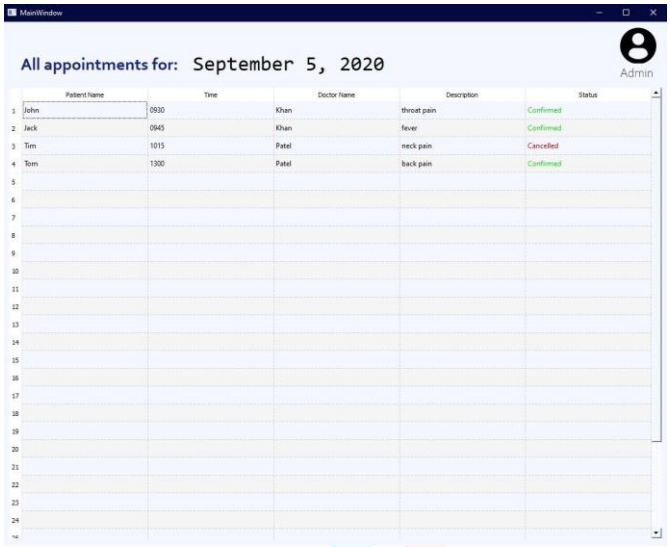

Fig. 4. Datewise Appointment Booking

Fig. 5. Male vs Female condition graph

Fig. 6. Male vs Female age ratio graph

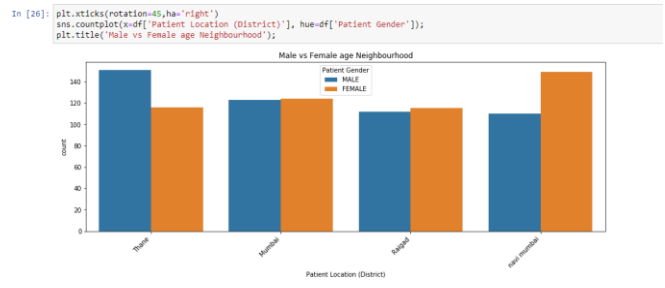

Fig. 7. Male vs Female neighbourhood graph

select another option. The process is similar for deletion and cancellation of appointments. Apart from this the keywords which were extracted from the speech can also be used for making a report which we would be analyzed using Python. The doctor, or any superior hospital staff along with the user has complete access over the booking of the appointment made.

IV. CONCLUSION

Automated and voice recognition based booking system is developed based on the method of CNN models via different packages available. The system is available as a desktop application and is easily accessible.A simple and effective UI helps in long term assistance by the system. With voice based system the need of manual management like the receptionist, hectic record keeping, human errors etc would be eliminated. Not only will the project be useful for the medical industry but can also be implemented in other industries like food delivery systems, movie show booking systems, cab booking and many such industries. Such a development in the booking system is not only helpful for technological growth but is also inevitable in the coming generations.

ACKNOWLEDGMENT

We wish to express our deep sense of gratitude to thank our project guide Prof.Samundiswary for providing timely assistance to our query and sincere guidance. We take this opportunity to thank our HOD Dr.K.Lakshmisudha and Principal Dr. Atul Kemkar for their valuable guidance and immense support in providing all the necessary facilities. We would also like to thank the entire faculty of IT Department for their valuable ideas and timely assistance in this project. Last but not the least, we would also like to thank teaching and non-teaching staff members of our college for their support, in facilitating timely completion of this project.

REFERENCES

[1] S. Negi, S. Joshi, A. K. Chalamalla and L. V. Subramaniam, ”Automatically Extracting Dialog Models from Conversation Transcripts,” 2009 Ninth IEEE International Conference on Data Mining, Miami, FL, 2009, pp. 890-895, doi: 10.1109/ICDM.2009.113.

[2] Brown, Stephen J, ”Report generation in a networked healthmonitoring system.” U.S. Patent 8,019,618, issued September 13, 2011.

[3] N. Naik, ”Connecting Google cloud system with organizational systems for effortless data analysis by anyone, anytime, anywhere,” 2016 IEEE International Symposium on Systems Engineering (ISSE), Edinburgh, 2016, pp. 1-6, doi: 10.1109/SysEng.2016.7753150

[4] S. Srivastava, S. Soman, A. Rai and P. K. Srivastava, ”Deep learning for health informatics: Recent trends and future directions,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1665-1670, doi: 10.1109/ICACCI.2017.8126082.

[5] M. Uma, V. Sneha, G. Sneha, J. Bhuvana and B. Bharathi, "Formation of SQL from Natural Language Query using NLP," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862080.

[6] K. Huang, C. Wu, Q. Hong, M. Su and Y. Chen, "Speech Emotion Recognition Using Deep Neural Network Considering Verbal and Nonverbal Speech Sounds," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 5866-5870, doi: 10.1109/ICASSP.2019.8682283.

[7] S. Pleshkova-Bekiarska and A. Bekiarski, "Building Human Mobile Robot Audio Communication Interface with Artificial Intelligence and Deep Learning," 2019 International Conference on Creative Business for Smart and Sustainable Growth (CREBUS), Sandanski, Bulgaria, 2019, pp. 1-5, doi: 10.1109/CREBUS.2019.8840078.

[8] Graham, Scott W., LiorLuker, NitzanNissim, and Brian L. Pulito. "Real-time human data collection using voice and messaging side channel." U.S. Patent 10,304,453, issued May 28, 2019.

[9] Ramachandran, Sridhar, ParitoshTyagi, SaravananMallesan, GauravKulshrestha, SohanShetty, Rohini Raman, and Medhavi Bhatia, "Method and system for routing media calls over real time packet switched connection." U.S. Patent 10,171,514, issued January 1, 2019.

[10] Companion Robots, 2017, https://romjist.ro/full-texts/paper562.pdf

[11] A Review on Automatic Speech Recognition Architecture and Approaches, April 2016, https://www.researchgate.net/