ABSTRACT

COVID 19 pandemic was declared as a matter of serious concern and considering this, it was necessary to detect symptoms of COVID 19 at its early stage so that it can be cured. Hence performing COVID 19 test was essential. Thousands of new cases were reported on a daily basis within a few months. And for the Government and Healthcare workers who were already working hard to defend people from the virus spread, testing procedures became an additional challenge. Carrying this activity on such a large scale along with maintaining social distancing and minimum physical contact can be very time consuming for healthcare workers. So we considered designing such a program that would directly store an individual's information in systematic format by recognizing the details through speech. And this stored data can be used for further analysis. This would make the testing procedure less time consuming and more systematic till some extent. In this way more time can be given in looking after the patients as giving medical aid to affected individuals is also of utmost importance.

TABLE OF CONTENTS

TABLE OF FIGURES

Chapter 1

INTRODUCTION

On January 30th, 2020, the World Health Organization declared the COVID-19 pandemic, a Public Health Emergency, as a matter of serious concern. It has majorly influenced the lives of the global community, thereupon, people's lives have been threatened. India, experiencing its highest 2 peaks in its COVID- 19 cases since last year, acknowledged the fact that the testing is not only an efficient cure at individual level but also a way of contacting infectious people and enforcing isolation. Therefore, testing helps to scale down the spread of COVID-19 virus. But, one of the issues identified was lack of trained manpower, in that matter, a research suggested that training and workshops organized, at national and international level, for COVID-19 testing will be helpful for generating the necessary skilled manpower. To make the use of manpower wisely, this project aims to provide fast data storing service via speech recognition. It can provide the best alternative method compared to manual data entering. It not only saves time but also reduces manpower effort. The user gives voice input which is displayed on the screen as well as stored in the device.

1.1 Objectives

According to WHO, the primary initiative that is to be taken by every country is to control the pandemic by social distancing to slow down the transmission and reduce fatality associated with COVID-19. But, in a few cases, this seemed to be violated at COVID-19 testing centers. So, the model plays a crucial role in maintaining the following precaution:

- It ensures minimal contact between the candidates and the volunteers which promotes disciplined social distancing.

- The enlistment of the details becomes a timesaving procedure. The program makes sure that candidates will not spend time standing in queue. Thus, it helps in conducting the process fast and smoothly.
- This project intends to significantly reduce the burden of healthcare professionals, resulting in upgradation of overall procedure.

1.2 Intended audience

The job of registration specialists is to enlist the information of the patients into the computer system. Through this program, an attempt is made to reduce the work of hospital registration specialists. Thus, this program mainly focuses on the hospital registration process.

Chapter 2

LITERATURE REVIEW

Michael H. Cohen, the first machine to recognize speech to any significant degree commercially named, Radio Rex (toy) was manufactured in 1920. This miniature bulldog would jump from a doghouse upon hearing a person say the words “Radio Rex.” Although Radio Rex was quite primitive in many ways, being able to recognize only a single phrase using simple hardware, it displayed several interesting properties, including real-time speaker-independent recognition.[1]

Sadaoki Furui, the earliest attempts to devise systems for automatic speech recognition by machine were made in 1950s, when various researchers tried to exploit the fundamental ideas of acoustic phonetics. During 1950s, most of the speech recognition systems investigated spectral resonances during the vowel region of each utterance which were extracted from output signals of an analogue filter bank and logic circuits.[2]

K.H.Davis, R.Biddulph et.al, In 1952, at Bell laboratories, Davis, Biddulph, and Balashek built a system for isolated digit recognition for a single speaker. The system relied heavily on measuring spectral resonances during the vowel region of each digit. [3]

D.B.Fry, In 1959, at University College in England, Fry and Denes tried to build a phoneme recognizer to recognize four vowels and nine consonants. They used a spectrum analyzer and a pattern matcher to make the recognition decision. A novel aspect of this research was the use of statistical information about allowable sequences of phonemes in English (a rudimentary form of language syntax) to improve overall phoneme accuracy for words consisting of two or more phonemes. [4]

T.Sakai, S.Doshita et.al, However, the decade (1960-1970) started with several Japanese laboratories entering the recognition arena and building special purpose hardware as part of their systems. On early Japanese system, described by Suzuki and Nakata of the Radio Research Lab in Tokyo, was a hardware vowel recognizer. An elaborate filter bank

spectrum analyzer was used along with logic that connected the outputs of each channel of the spectrum analyzer (in a weighted manner) to a vowel decision circuit, and majority decisions logic scheme was used to choose the spoken vowel. Another hardware effort in Japan was the work of Sakai and Doshita of kyoto University in 1962, who built a hardware phoneme recognizer. A hardware speech segmented was used along with a zero crossing analysis of different regions of the spoken input to provide the recognition output.[5]

C.C.Tappert, N.R.Dixon et.al, Another great milestone of the 1970 was the beginning of a longstanding, highly successful group effort in large vocabulary speech recognition at IBM in which researchers studied three distinct tasks over a period of almost two decades, namely the New Raleigh language for simple database queries, the laser patent text language for transcribing laser patents, and the office correspondent tasks called Tangora, for dictation of simple memos. [6]

B.Lowrre, One of the demonstrations of speech understanding was achieved by CMU in 1973 there system was able to use semantic information to significantly reduce the number of alternatives considered by the recognizer. CMUs Harphy system was shown to be able to recognize speech using a vocabulary of 1,011 words with reasonable accuracy. One of the particular contributions from the Harpy system was the concept of graph search, where the speech recognition language is represented as a connected network derived from lexical representations of words. The Harpy system was the first to take advantage of a finite state network (FSN) to reduce computation and efficiently determine the closest matching string. [7]

L.R.Rabiner, Speech research in the 1980s was characterized by a shift in technology from template based approaches to statistical modeling methods especially the hidden Markov model approach. Although the methodology of hidden Markov modeling (HMM) was well known and understood in a few laboratories(Primarily IBM, Institute for Defense Analyses (IDA), and Dargon systems), it was not until widespread publication of the methods and theory of HMMs, in the mid 1980, that the technique became widely applied in virtually, every speech recognition research laboratory in the world. Today, most practical speech

recognition systems are based on the statistical frame work developed in the 1980s and their results, with significant additional improvements have been made in the 1990s. [8]

Kamlesh Sharma, In 2007, Microsoft released Windows Vista, the first version of Windows to incorporate speech recognition. Windows Speech Recognition is a new feature in Windows Vista, built using the latest Microsoft speech technologies. Windows Vista Speech Recognition provides excellent recognition accuracy that improves with each use as it adapts to the speaking style and vocabulary. Speech Recognition is available in English (U.S.), English (U.K.), German (Germany), French (France), Spanish (Spain), Japanese, Chinese (Traditional), and Chinese (Simplified).[9]

Michiel Bacchiani, Another major innovation happened to be Google's GOOG-411, a telephone-based directory service. This will serve as a foundation for the company's future Voice Search product. These include the telephony network, an application server that executes the voice application, a TTS server, a recognition server with its acoustic, language, and pronunciation models (AM, LM, PM), the Google Maps service to execute business queries, and an SMS gateway to send information to mobile users. [10] Welch, Chris, In 2011 came in with Apple's Siri, a digital personal assistant. In addition to being able to recognize speech, Siri is able to understand the meaning of what it is told and take appropriate action. Microsoft announced Cortana, a digital personal assistant similar to Siri in the year of 2014. Another milestone in 2014 was Amazon's Echo, a voice controlled speaker. The Echo is powered by Alexa, a digital personal assistant similar to Siri and Cortana. While Siri and Cortana are not the most important features of the devices on which they run, the Echo is dedicated to Alexa [11]

Chapter 3

GENERAL PROCESS OF CONVERSION FROM SPEECH TO TEXT

Few decades ago, computers could barely get past the basic function of a machine. But with the rise of natural language processing, it all changed. It is due to the powerful subfield of AI called Speech recognition, that gives the computer the power to actually understand the exact sounds of a person and make sense out of it. Speech recognition allows for the recognition and translation of Spoken words into text. Its working can be understood as - Breaks down audio into individual sounds, converting these sounds into digital format and lastly using algorithms and models to find the most suitable word fit in that particular language.

1. Analog digital converter (ADC) translates sound waves to digital data.

Human speech creates vibrations into the air. ADC converts these analog waves into binary data that the machines are aware of. Secondary function of ADC is to remove unnecessary noise and normalizing sound and speed of speech to match the prerecorded samples.

2. Spectrogram analysis frequencies of sounds.

Spectrometer separates the approaching data into different bands of frequency and is analyzed henceforth.

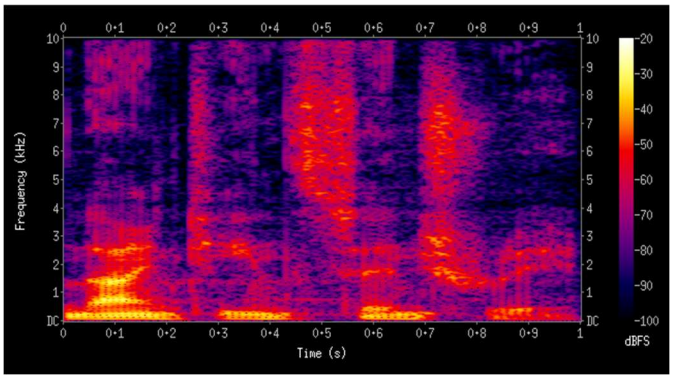


Fig.3.1: Frequency vs time graph[12]

2.1 Working of spectrogram

Device plots the graph (time vs frequency) of the incoming sounds. As all the words are made up of distinct vowel sounds that have different frequencies, they are recorded on this device showing colour variation in the plotted graph. The brighter area signifies higher frequency and darker depicts lower frequencies.

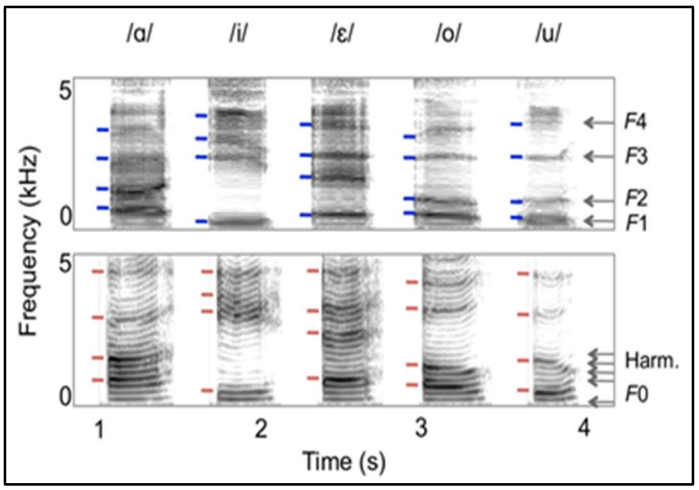


Fig.3.2: Frequencies of Vowels[13]

Fig. 4.3 shows that different vowels have different frequencies. These frequency patterns are pre-programmed into the computers, allowing it to recognize when a spoken sound matches a specific vowel sound.

3. Machine detects phonemes and connects them to dictionary words.

A phoneme is the smallest unit of sound in speech. In simpler terms, words broken down into its sounds will give the phonemes present in a particular word. To have a better understanding, some examples of phonemes are as follows:

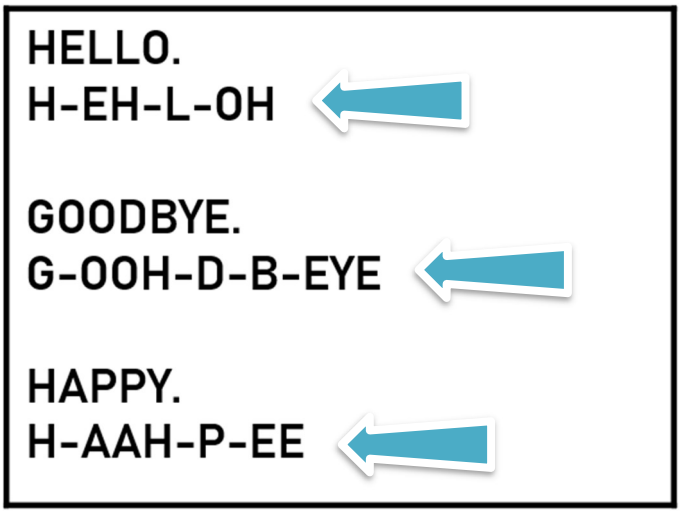


Fig.3.3: Word Phonemes

These phonemes are then compared to the words present in the inbuilt dictionary through complex models, giving the output into text. Nevertheless, it is known that human communication is complex, consisting of different accents and pronunciations which is not a part of an in-built dictionary. This gap is fulfilled by the Hidden Markov model (HMM) first introduced in the 1980s. It is described as one of the key technologies developed in the field of ASR. As explained, the speech input is thus generally converted into text.

Chapter 4

REQUIREMENTS

The following chapter mentions the system requirements that are prerequisites for the efficient functioning of the program.

3.1 Python packages

The most fundamental functions of this system are helped achieved by the packages used in this program. These libraries should be installed according to the version of the used python IDEs So, python modules that are needed to be installed are mentioned as follows:

a) Pyaudio
b) Speech Recognition
c) Pandas
d) Pyttsx

2. Microphone system

The speaker needs to speak the required details into the mic for successful enrolment by the system.

Note:

At the time of running the code for the first time, the excel file 'Registered Members' should be made before running the program. The location of this file should be the folder in which the code is being executed.
The excel file should be closed at the time of running the program.

Chapter 5

METHODOLOGY

In this chapter, starting with the flow of the program, then python packages required and lastly, one of the major aspects of our program data storage in Excel, have been explained in detail.

5.1 Flow of the program

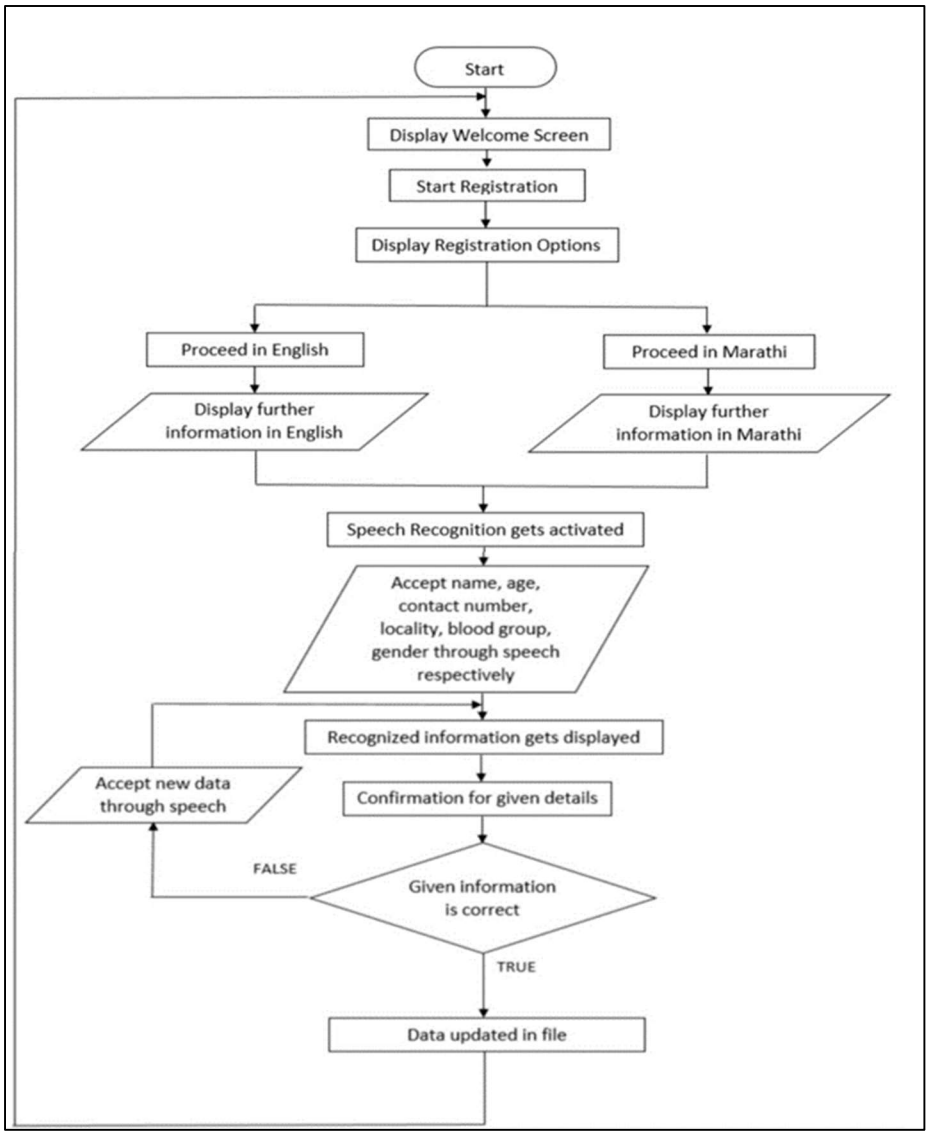


Fig.5.1: Flow of program

A flowchart precisely is a representation of flow of program in the form of a diagram. Each shape in flowchart has its own significance. Flowchart shows the path in which the program proceeds.

When the program starts, initially the system displays two options on screen. If user choose first option then the further information gets displayed in English. If user choose second option then the further information gets displayed in Marathi. Further the system ask user “Please say your name” , ”Please say your age” and other required information. If user says his name then automatically name gets displayed on the screen using speech recognition function. If some of the given information needs to be changed then system asks for the new input for a particular field. If user gives confirmation about the information given is correct then the data successfully gets stored in excel file and the program automatically start one more time for second entry.

5.2 System design

5.2.1 Packages involved in the program.

a] Pyttsx3 package:

pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 and 3. An application invokes the pyttsx3.init() factory function to get a reference to pyttsx3. Engine instance. It is convenient to use this module which converts the recognized text into speech into both male and female voice.

The main motive behind adding this module into the system is, first of all, to make the program more user friendly as listening is a more approachable way than reading. Secondly, as sound can be recognized easily from a faraway distance, it becomes trouble-free to maintain social distancing.

This module is used to let the candidate know the required information that is to be taken from them.

b] Speech Recognition package:

The main objection of the program is- registration through speech recognition. So, this library proves to be the backbone of the system.

Speech recognition, as the name suggests, refers to automatic recognition of human speech. Several speech recognition libraries have been developed in Python. However, Speech Recognition library, which is the simplest of all the libraries, is used to set up this system. This made the system more accessible and easier to use.

Its main purpose is to convert speech into text. It not only helps in saving time by speaking instead of typing but also is the most suitable way for the candidate to put in the details. In this program, once the candidate gets to know about the required information, the commands of this module get initiated and the program gets ready to take the input.

c] Date time package:

As date and time are not a datatype of its own, this module can be used to work with and include date and time in the program. The datetime module supplies classes for manipulating dates and times.

Datetime module includes now() function that can be primarily used to get current time and date. It returns the current local date and time. It has a significant role in getting data more detailed. Using this function, it becomes possible to add a column of date and time when a candidate registered for getting tested. So, in the long run, this helps record the date time and of the candidate registered and ultimately helps in graphical analysis as well.

d] Random module:

The random module is a built-in module to generate the pseudo-random variables. It is used to get a random number, obtain random values from a list of values, shuffle elements randomly, etc. This module made it possible to assign a unique ID to the registered candidate.

Once successful enrolment is done, the system would allot a passcode to the candidates at the primary stage itself. This makes the process easier and faster at upcoming steps. It also reduces the chance of inaccuracy or misinformation in the final registered list, if there was any misinterpretation by the system before.

There might be an instance, where the same candidate has to be tested twice. So, in that case, there will be 2 entries for the same person. But the passcode will help us find the latest entry by that candidate

e] Pandas package:

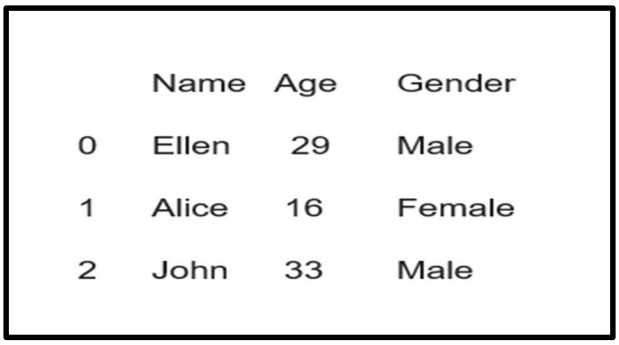
Pandas is a fast and easy to use open-source data analysis and manipulation tool used for data manipulation and analysis. It is well suited for different kinds of data such as tabular data, heterogeneously typed columns, ordered and unordered time series data or any other forms of observational and statistical datasets. One of the features of this module is dataset into excel file.

The major purpose of using pandas is to create a dataframe by the inputs given by the candidates. Basically, dataframe is a type of data structure that contains rows and columns and arithmetic operations aligned on both rows and columns.

Pandas library consists of a function- DataFrame.to_excel. This converts the dataframe into an excel file which becomes accessible easily. This excel file can then be used for the analysis purpose.

5.3 Data storage into Excel

Pandas is an open source python package that provides numerous tools for data analysis, using a few methods, one of them being Dataframes. It organizes data into rows and columns, creating a two-dimensional data structure. To create a dataframe, we use the pandas method 'DataFrame()' and pass a dictionary as the argument. The keys of the dictionary represent the column names and the values represent the data in them. Information of multiple rows can be added to at the same time by passing a list as the value.



	Name	Age	Gender
0	Ellen	29	Male
1	Alice	16	Female
2	John	33	Male

Fig.5.2: Dataframe

With pandas, we can open and read an excel file using the 'read_excel()' function and passing the name of the file as the argument. This can also be saved in a variable and be used to print the existing data on the file. Concatenation of data, which means to add one set of data to another can also be done by calling the 'concat()' function and passing the names of the dataframes in a list as the argument. To export the dataframe into an excel sheet, the function 'to_excel()' is used with the name of the excel file as the argument. Lastly, after all changes have been made, we call the '.save()' function to save the updates and close the file.

Column1	Name	Age	Gender
0	Ellen	29	Male
1	Alice	16	Female
2	John	33	Male

Fig.5.3: Final stored data

Chapter 6

EXPERIMENTATION

The working and the execution of the program has been explained with the help of images.

6.1 Implementation of the program

The following images show the outcomes of the program in both English as well as Marathi. They depict what the program yields primarily after the very first execution.

6.1.1 Execution of program in English

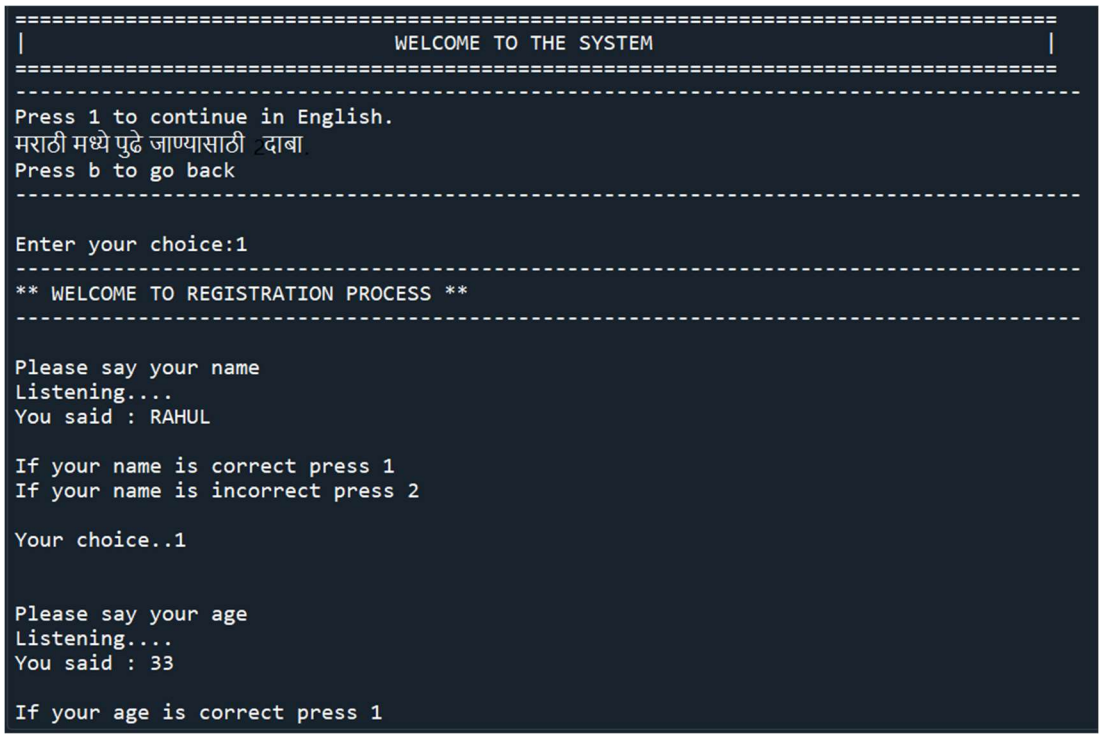


Fig. 6.1: Start of program in English

```
Your choice..1


Please say your phone number
Listening....
You said : 777777777

If your Phone Number is correct press 1
If your Phone Number is incorrect press 2

Your choice..1


Please say your Area or Locality
Listening....
You said : MG ROAD

If your Locality is correct press 1
If your locality is incorrect press 2

Your choice..1


Please say your blood group
Listening....
You said : B NEGATIVE

If your Blood Group is correct press 1
If your Blood Group is incorrect press 2
```

Fig. 6.2: Taking input through speech and converting to text

```
Please say your gender
Listening....
You said : MALE

If your Gender is correct press 1
If your Gender is incorrect press 2

Your choice..1

Date And time 2021-08-01 14:10:30.696424
pass code 3659
--------------------------------------------------------------------------------
Enrollment Successful !
--------------------------------------------------------------------------------
|| Name:RAHUL
|| Age:33
|| Contact number:777777777
|| Locality:MG ROAD
|| Blood Group:B NEGATIVE
|| Gender:MALE
|| Date and time:2021-08-01 14:10:30.696424
|| Pass Code:3659
```

Fig. 6.3: Confirmation of details

6.1.2 Execution of program in Marathi

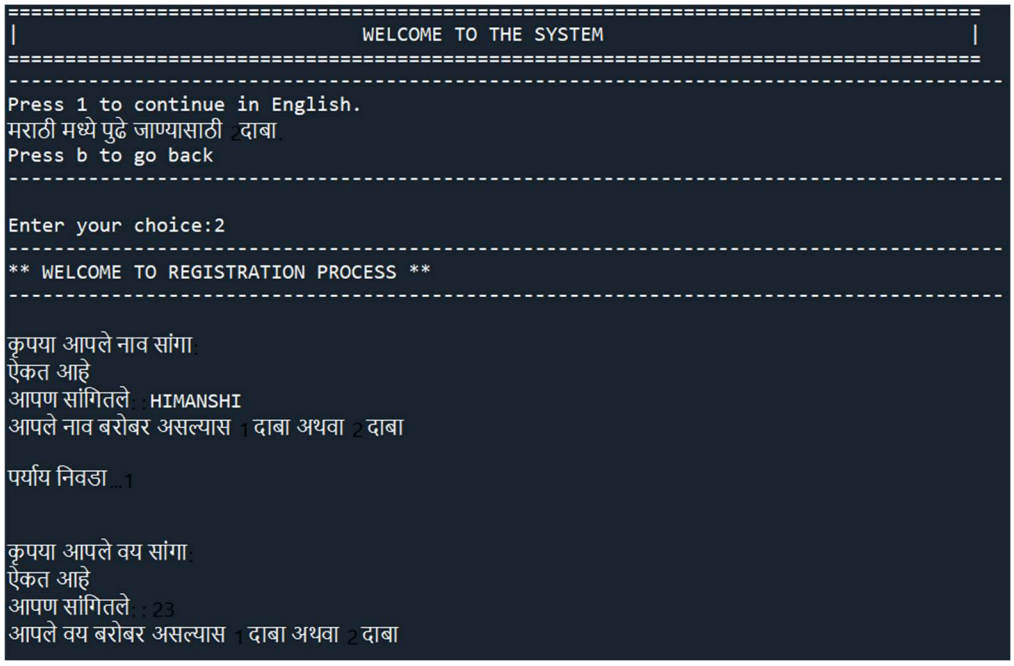


Fig. 6.4: Start of program in Marathi

Fig. 6.5: Taking input through speech and converting to text

```
पर्याय निवडा 1


कृपया आपले लिंग सांगा
ऐकत आहे
आपण सांगितले: FEMALE
आपले लिंग बरोबर असल्यास 1 दाबा अथवा 2 दाबा

पर्याय निवडा 1

Date And time 2021-08-01 14:19:14.658913
कृपया आपला पासकोड लक्षात ठेवा
pass code 2311
-----------------------------------------------------------------------------------------
Enrollment Successful !
-----------------------------------------------------------------------------------------
आपली माहिती बरोबर असल्याची खात्री करा
|| Name:HIMANSHI
|| Age:23
|| Contact number:999999999
|| Locality:RANGOLI LAWN
|| Blood Group:A POSITIVE
|| Gender:FEMALE
|| Date and time:2021-08-01 14:19:14.658913
|| Pass Code:2311
```

Fig. 6.6: Confirmation of details

6.1.3 Final display in excel

	A	B	C	D	E	F	G	H	I
1	Date	Time	Name	Age	Gender	BloodGroup	Area	ContactNumber	Passcode
2	2021-07-31 00:00:00	19:46:03	NISHTHA	53	FEMALE	A POSITIVE	VMV ROAD	1111111111	8604
3	2021-07-31 00:00:00	19:48:55	NAVEEN	42	MALE	O POSITIVE	HINDUJA PLOT	6666666666	9680
4	2021-07-31 00:00:00	19:51:30	RAHUL	33	MALE	B NEGATIVE	MG ROAD	7777777777	1548
5	2021-07-31 00:00:00	19:59:27	HIMANSHI	23	FEMALE	B POSITIVE	RANGOLI LAWN	9999999999	6749
6	2021-07-31 00:00:00	20:03:27	SHERIN	36	FEMALE	A POSITIVE	PANCHAVATI SQUARE	7777777777	7882
7	2021-07-31	20:09:35	MOHIT	48	MALE	A POSITIVE	SUNSHINE HILL	4444444444	7330

Fig. 6.7: Final display of Excel file

Chapter 7

RESULT AND DISCUSSION

Registration is the process through which an individual's identity is enrolled into the records of a hospital. For this PBL project, we strived to create a pre testing registration system with automatic speech processing to create an intelligent interface and seek the voice input of users so that these interfaces allow intelligent interaction with users and establish a natural and easy communication between the system and public. The project 'Pre-testing Registration through Speech Recognition ' is to make the registration process faster and friendly to both users and the candidates. The program takes care of all necessary requirements for the registration process and is capable of providing effective storage of information. The information thus obtained gets stored in a systematic tabular form which can be easily filtered for further analysis. Analysis can include monitoring the number of registrations from various age groups and localities. Through this project an attempt is made to use the Speech Recognition technology as a medium for collecting and storing data in a trouble freeway.

Chapter 8

CONCLUSION

Earlier studies about the registration system revealed that a manual appointment routine is a very time-consuming task for registration. With an appointment system or a customized system that would fulfil the present requirement, centers/hospitals can offer candidates/patients comfort and optimize the usage at the centers. If most of the appointment tasks are automated, it will reduce the time of the staff, allowing them to work on other tasks within the centers.

Taking the present scenario into consideration, social distancing and the need to safeguard public health is the major motivation behind this semi-automated registration system. Last few decades, in the desire to automate work with machines, research in speech recognition has gained much enthusiasm. Automatic speech recognition has greatly contributed to the development of artificial intelligence, which seeks to create very flexible methods of handling the machine, this allows the user to communicate and exchange information without using known input/output modules such as the keyboard. Voice-based input/output techniques are very useful in several areas, such as the care of disabled people, the use of cars, in particular when driving, etc. All speech processing techniques which include speech synthesis, speech processing, speaker identification, Speaker verification make it possible to create human machine interfaces or perform voice interaction. Hence, there could be no better way than using ASR as a fundamental of this system. It is our firm opinion that there are a lot of features that can be added to the current program to take it to the next level and to implement it in real-time. We also believe, once the system is ready to put into practice, only coming up with this innovative technique of registration won't be enough, encouraging the testing centers to accept this program so as to upgrade the procedures.

Future Scope

Using text to speech conversion for giving instructions and speech to text conversion for taking input, storing the obtained input in systematic form is a much quicker and easier way. With some extra features included to the system that would make it more realistic, it can be used in following environments to make the settings more upgraded:

a) The same program can be used in colleges during admissions to keep a record of students' entry and exit time along with other details.
b) It can be used to take feedback from people as they just need to speak and all that will be stored in the form of text.
c) It can be used during big competitions for quick registration of participants.
d) In buildings and apartments, the entry and exit time of people arriving at someone's flat along with other details can be done with such a system.
e) With maximum accuracy and improvements, a system can be made which on hearing an individual's symptoms would guide them with medicines and procedures to be followed and the same information can be stored in the system for future reference.

REFERENCES

[1] Michael H. Cohen, dl.acm.org/doi/10.5555/975074, *Voice User Interface Design,*2004

[2] Sadaoki Furui*, 50 years of Progress in speech and Speaker Recognition Research ,* ECTI Transactions on Computer and Information Technology, 2005.

[3] K.H.Davis, R.Biddulph, and S.Balashek*, Automatic Recognition of spoken Digits,*1952.

[4] D.B.Fry, *Theoritical Aspects of Mechanical speech Recognition , and P.Denes, The design and Operation of the Mechanical Speech Recognizer at Universtiy College London,* 1959.

[5] T.Sakai and S.Doshita, *The phonetic typewriter*, information processing 1962 , 1962.

[6] C.C.Tappert, N.R.Dixon, A.S.Rabinowitz, and W.D.Chapman, *Automatic Recognition of Continuous Speech Utilizing Dynamic Segmentation, Dual Classification, Sequential Decoding and Error Recover,*1971.

[7] B.Lowrre, *The HARPY speech understanding system , Trends in Speech Recognition,*1990.

[8] L.R.Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition ,* 1989.

[9] Kamlesh Sharma, *Exploration of Speech Enabled Systems for English,*2012.

[10]Michiel Bacchiani, *Deploying goog-411: early lessons in data, measurement, and testing,* Google.inc

[11]Welch, Chris. *"Amazon just surprised everyone with a crazy speaker that talks to you".* The Verge

[12]<*https://en.wikipedia.org/wiki/Spectrogram#/media/File:Spectrogram-19thC.png*>

[13]<*https://www.researchgate.net/figure/Spectrograms-of-the-vowels-i-o-and-u-international-phonetic-symbols_fig2_277131520*>