Dictation Module Using Automatic Speech Recognition in Machine Learning

Vaishnavi Kocheta[1], Shubhangi Shinde[2], Kiran Nadkar[3], Snehal Jambhulkar[4],
Prof Suchita Wankhede[5]

Student, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India[1,2,3,4]

Guide, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India[5]

Abstract: Using Artificial Intelligence and Machine Learning our project proposes a system named Dictation Module. It is a device that can help society to ace their work and help people in multitasking. As it is rightly said that "If a person cannot perform multiple tasks at the same time, he cannot achieve everything he wants". Multitasking is the basic asset of every individual in the society. Unlike others, the socially active businessmen or lawyers are forced to be dependent on this aspect. So we have proposed a device that helps to do multiple tasks anywhere you want. Imagine you are driving and you realize you forgot to mail someone something very important, how will you do it in the traditional way is to reach out to the destination and then start typing your mail and send it, which will cost you lots of time as well as some mishap can occur due to this process. But using our module you can dictate the mail to the machine and tell it to send it to the desired person while driving the car simultaneously. We have also included an additional thing to our basic idea that now we are going to store all the files in an online storage i.e. in Telegram using a Telegram Bot so that user can access it from anywhere even if he doesn't have his local storage with him/her.

Keywords: Artificial Intelligence, Machine Learning, PySimpleGUI ,Voice Recognition, Speech Recognition, Text to Speech , Speech to Text conversion, Yagmail ,Telegram.

INTRODUCTION:

Keeping up with the evolving world that wants you to multitask every now and then you need to combat yourself with emerging technologies . So here's presenting a smart system that requires only your voice to work with . As the name suggests "DICTATION MODULE " your requirement only for the completion of your work is you need to " DICTATE " and using the capabilities of our system that is "INTERACTION" with users, work becomes fun and we get the feel that we are chatting with any human like system. Using the appropriate Graphical User Interface we can read the speech that we are dictating.

LITERATURE SURVEY

1."Implementation of Text to Speech Conversion" This paper includes implementation of Optical Character Recognition (OCR) in MATLAB. An image containing some characters is submitted to OCR and the text written into it is extracted to a text file. A feedforward neural network is used for classification resulting into recognition of text. The text is then converted to speech using Win 32 SAPI (Speech Application Programming Interface).

2."A Study of Speech Recognition" Various Speech Recognition approaches (The acoustic-phonetic approach, Dynamic time warping (DTW), Neural Network based approach, etc.) are discussed in the given paper. Also some speech feature extraction techniques such as Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficients (MFCC), Signal Subspace Method are stated.

3."Text to Speech Conversion" Text-To-Speech Synthesizer is developed based on Raspberry Pi V2. The process involves image processing (where Tesseract OCR converts .jpg to .txt form) and voice processing (Festival software is used to convert the .txt to speech).

4."Design and implementation of Text to Speech for visually impaired people" This paper talks about implementation of Text-To-Speech Synthesizer using Natural Language Processing (NLP) Module (which converts text into phonetic transcription along with prosody) and Digital Signal Processing (DSP) Module (which transforms symbolic information or phonetic transcription received from NLP into audible and intelligible speech).

5."VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT" In this paper, for Speech-To-Text conversion they use Mel Frequency Cepstral Coefficients (MFCC) for feature extraction, Vector Quantization (VQ) for feature mapping, Hidden Markov Model (HMM) to create models for each letter and Viterbi Algorithm for feature testing of the dataset. To show desired text output MATLAB interface is used.
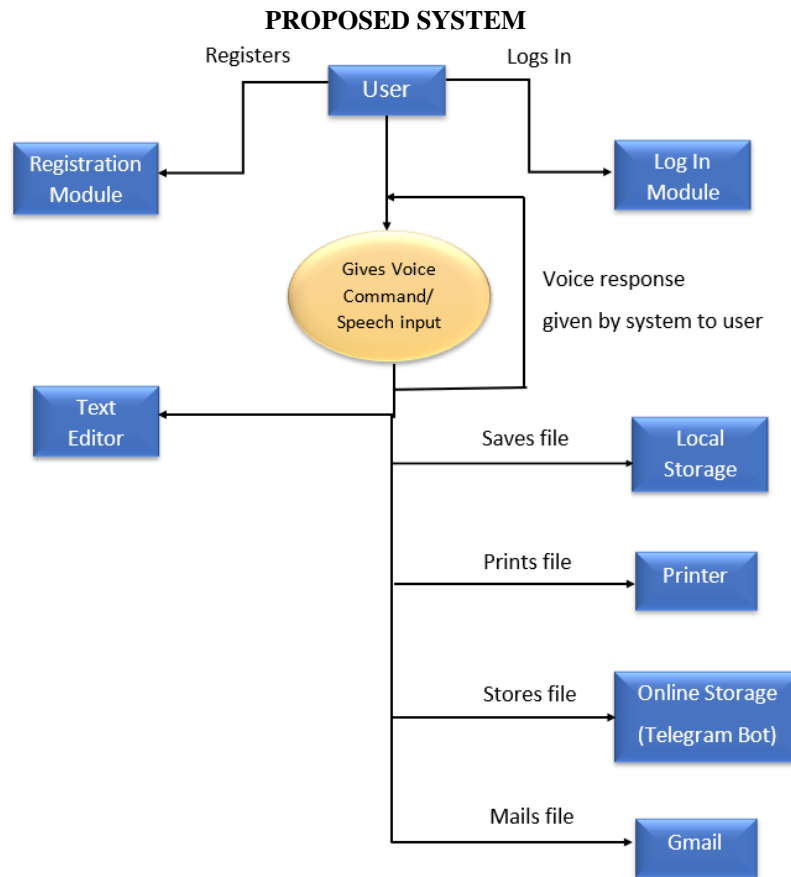
PROPOSED SYSTEM


Fig.1.System architecture

As discussed in our previous paper our main motive was to create an user friendly and effective system for multitasking, so here's the walk through of our "Dictation Module".

System Walk Through :
As per our Dictation Module one can dictate any document he wants to create with the use of his voice hassle free because of Interactive System which supports voice commands as well.
1. Dictation Module starts with Authenticating users.
User will have to Register himself/herself in order to use the system. User name and Password can be of his/her choice so as for them to get back. No duplication entertained.
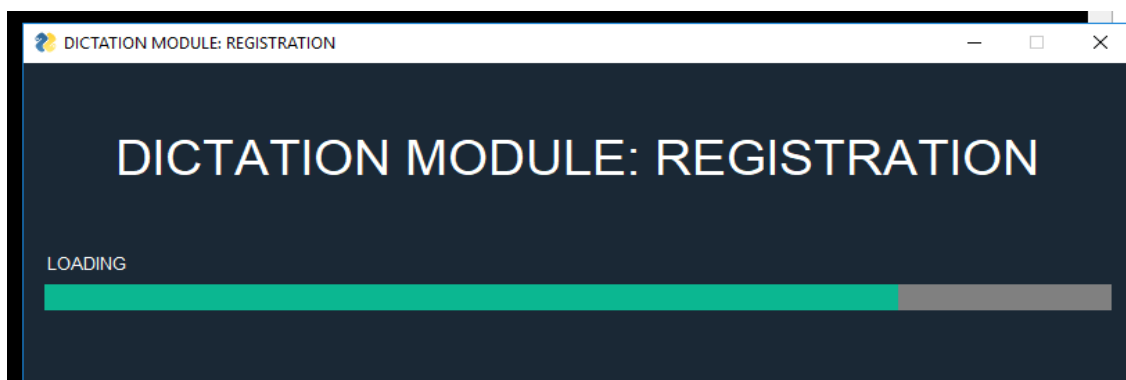

Separate Document of User's Username and Password exists where the number of users and their credentials are saved.
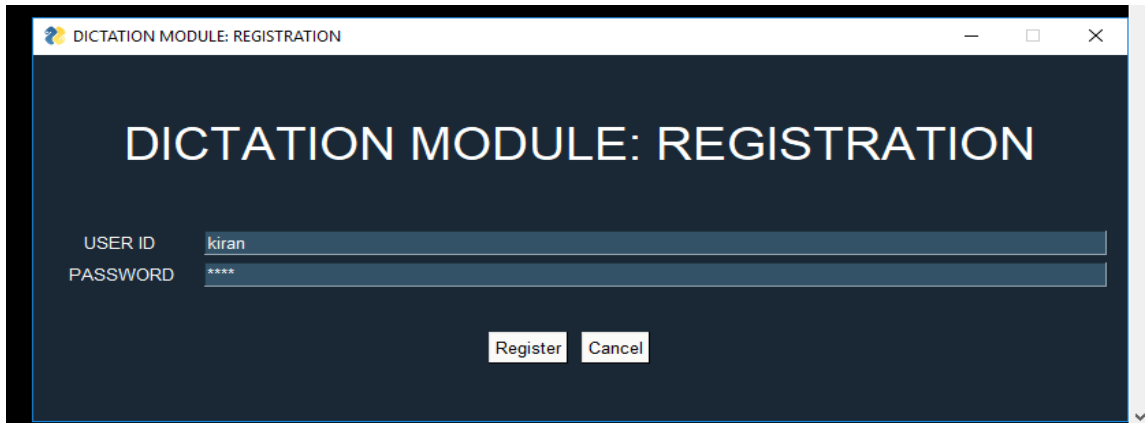
If the user logins the system speaks out as
"ACCESS GRANTED "
Or else,
"ACCESS DENIED"
"NAME NOT FOUND. PLEASE REGISTER."
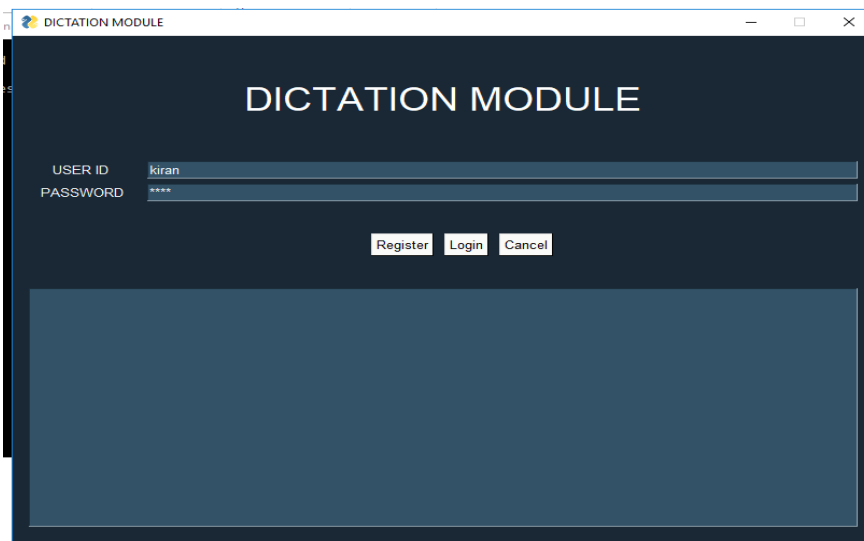
2. Once the system recognizes the voice of the speaker it progresses by saying
"START SPEAKING "
however if it fails in recognizing the speech it says
"SORRY COULDN'T RECOGNISE"
by which the user can understand whether the system is taking its voice or not .
As we developed an interactive system , system asks the user to start speaking and as the users speaks it keep on collecting the data received via user and displays it on the GUI console.
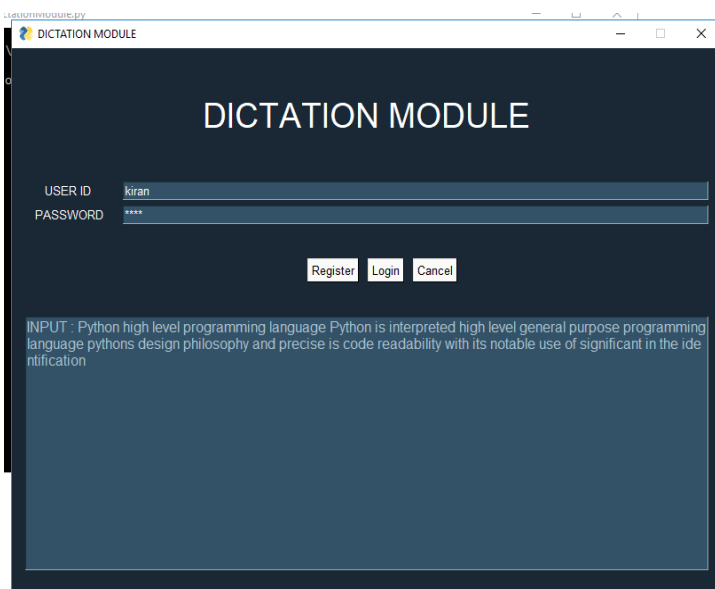

As user stops speaking, interaction is done with user via system.
3. Next, system asks
"DO YOU WANT TO SAVE THIS FILE?"
Then we speak out as "PLEASE SAVE" and the filename as well. Again if it does not recognize the command it says
"SORRY COULDN'T RECOGNISE"
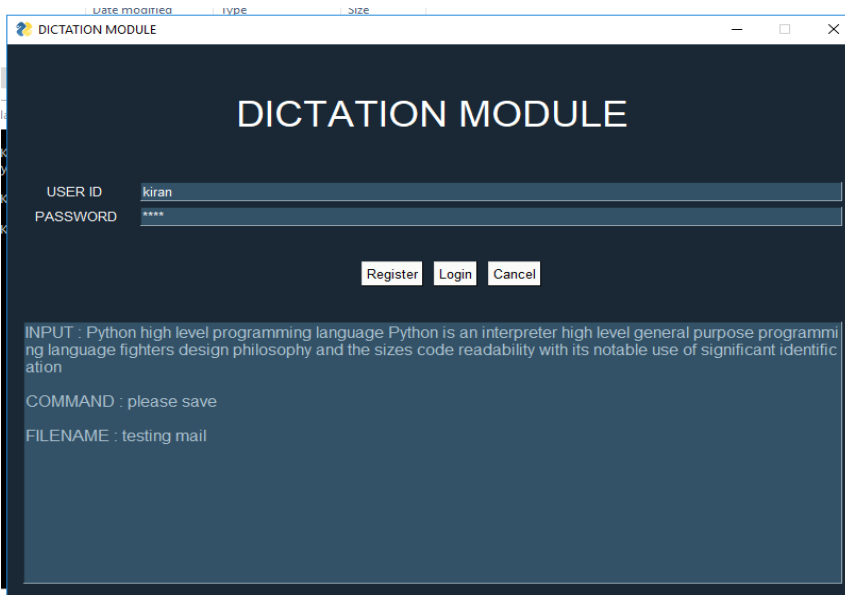
However on receiving the filename and on saving it successfully it says "FILE SAVED SUCCESSFULLY"
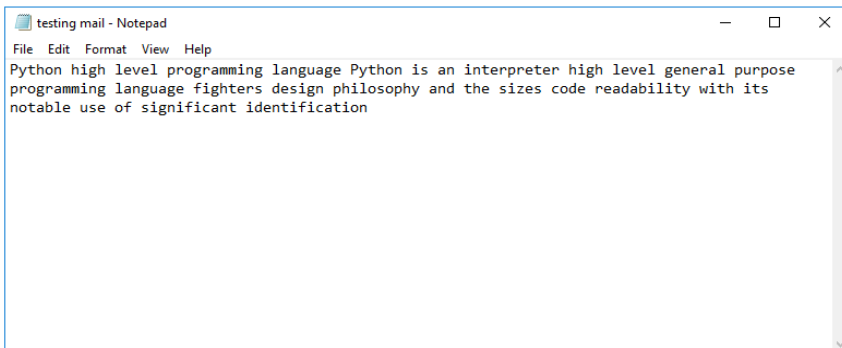
4. Likewise it has the interaction for further part that is printing the file and sharing with Telegram Bot for storing online.

System asks,

DO YOU WANT TO PRINT THIS FILE?

and then,

DO YOU WANT TO SEND THIS FILE ON TELEGRAM?
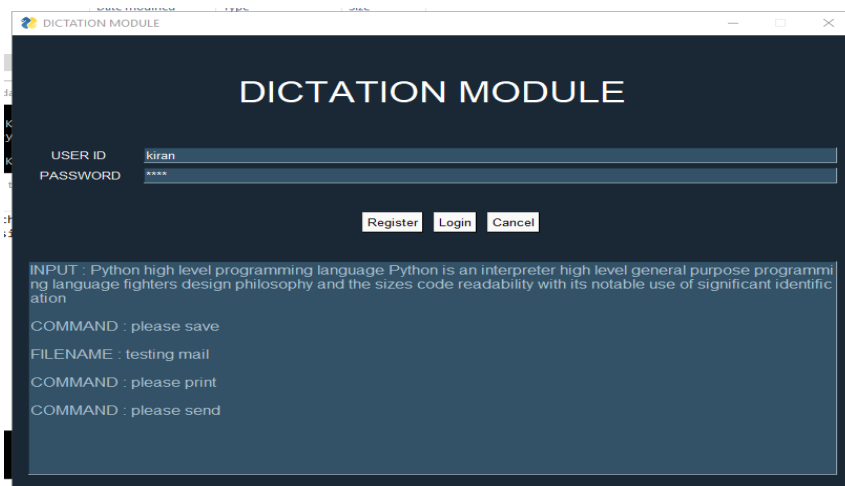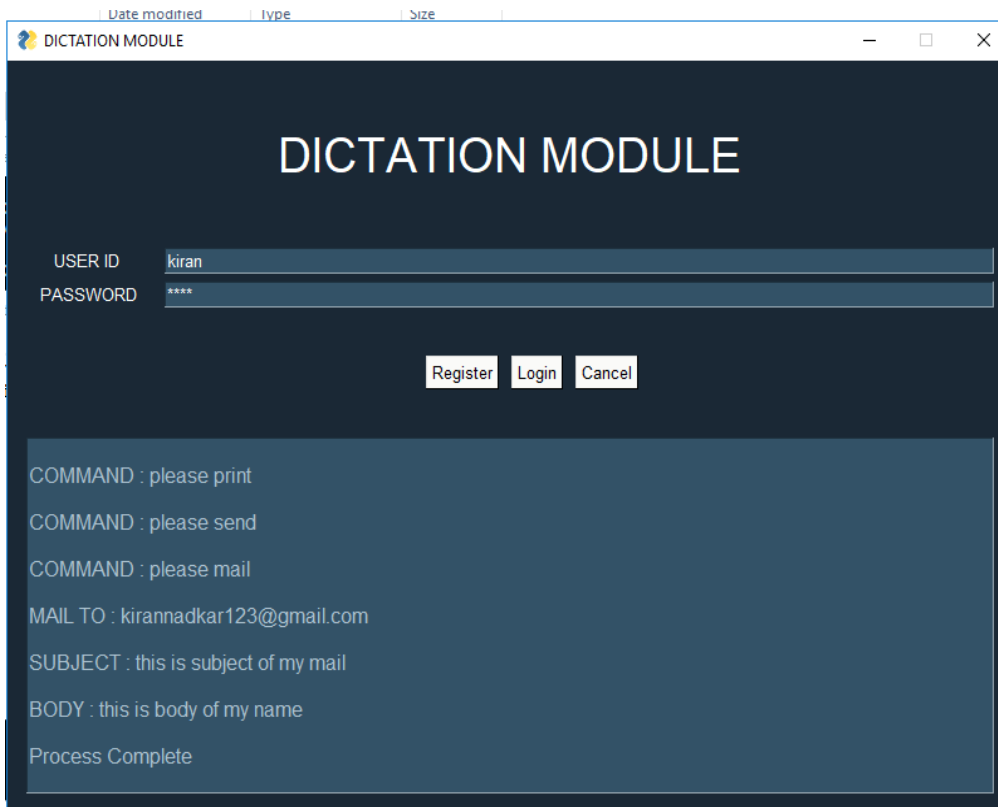
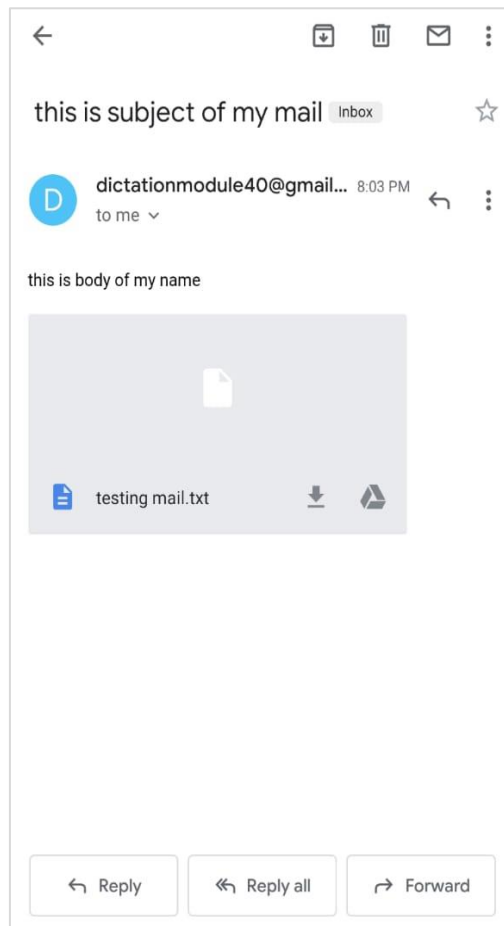If user says yes, system does the job

5. Next it follows the flow given below to mail the file,

DO YOU WANT TO MAIL THIS FILE?

ENTER LOGIN DETAILS   (Sender's mail id and password)

ENTER MAIL ID
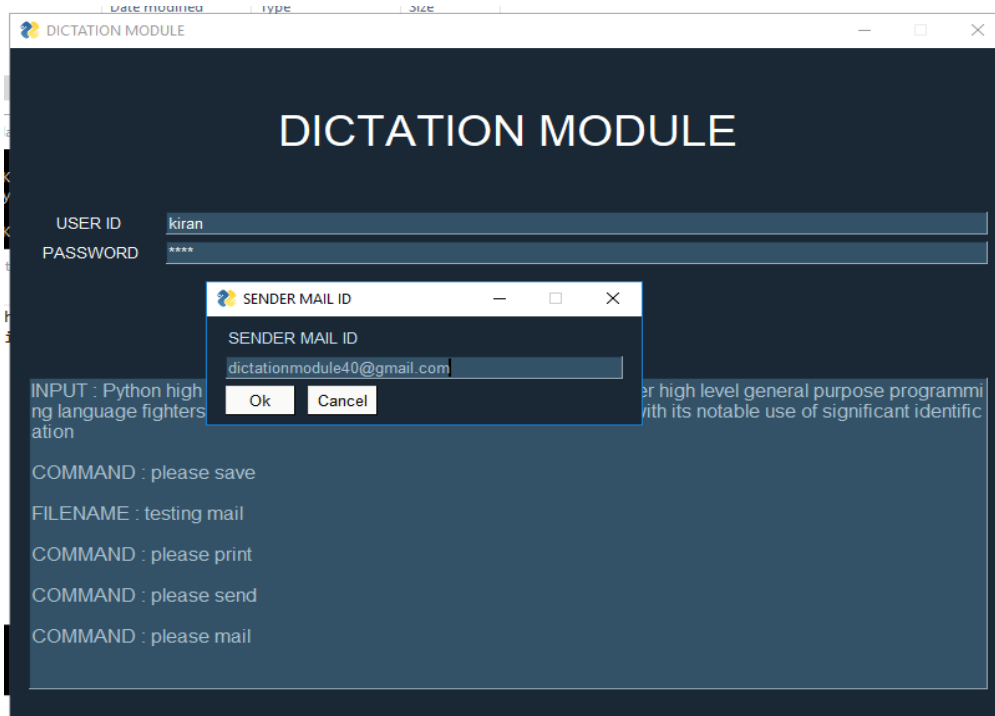
MAIL TO   (Receiver mail id)

ENTER SUBJECT   (Subject for the mail)

ENTER BODY   (Body for the mail)

The mail sent by system –

Important components of our Module :

Speech Recognition :
We used this library for recognising speech and to convert it into text. It acts as a wrapper for several popular speech APIs and it is extremely flexible.

pyttsx3 :
We used this library for text to speech conversion. It works offline. It supports "SAPI5" TTS engine.

YAGMAIL:
It is a simple GMAIL/SMTP client that is created to remove the hassle out of sending emails.
With this Python library, you can write a few lines of code to include email sending
capabilities to your application and sidestep using the cumbersome traditional approaches.

os.startfile(filename,"print") Function :

We've used this function which belongs to os library to print the file directly from the system.

PySimpleGUI:
PySimpleGUI is a new Python GUI library.

PySimpleGUI is a python package that enables python programmers of all levels to create GUIs.

PySimpleGUI helps us to make an event loop by creating an infinite while loop that reads events from the window object.

You specify your GUI window using a "layout" which contains widgets.

Widgets are nothing but the elements in PySimpleGUI.

Telegram Bot :

It is an additional thing in our project.It is used for the online storage purpose.

The Telegram bot API allows third-parties to create bots that use the Telegram app as the main interface.They run inside the Telegram environment and do not require any further installation.

We will access this Telegram API using one of its python wrapper: python-telegram-bot (link).
File to Bot allows the users to save files on the cloud with unlimited storage.

You can send files in the chat and they will get stored category wise.Uploaded files could be checked and downloaded easily through it.There is no limitation on file size and all file types are supported by the Telegram bot.

➢   **Hardware Resources Required :**

1.   CPU Speed 2GHz
2.   RAM 4 GB
3.   Good Quality Microphone
4.   Printer

➢   **Software Resources Required :**

1.   Operating System : Windows
2.   Programming Language : Python
3.   Telegram
4.   Gmail

CONCLUSION:
This paper gives a descent approach for Speech To Text conversion after studying various researches done by multiple researchers in this field. We have given brief idea about how our system is going to work. Our Proposed System aims to be very useful for people to interact with other people of the society as well as lawyers and businessmen for whom dictation technology is really necessary to speed up their work. It is also very user friendly and can be proven time saving method for intended users.

FUTURE SCOPE:

We look forward to do more research about this field and try and implement this system for more than one language and with whatever additional functionalities we could add also we would like to make the system handy to carry it wherever you wish to.

As voice based systems are in demand nowadays, our system has a lot of scope in future where we can make many custom modifications using trending technologies.

REFERENCES:

[1] Chaw Su Thu Thu , Theingi Zin *"Implementation of Text to Speech Conversion"* International Journal of Engineering Research & Technology (IJERT) IJERT ISSN: 2278-0181 Vol. 3 Issue 3, March – 2014

[2]] Kaladharan N*, "A study of Speech Recognition"* International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 9, September-2015

[3] S.Venkateswarlu, D.B.k. Kamesh, J.K.R Sastry, and Radhika Rani, *"Text to speech conversion"* Indian Journal of Science and Technology, Vol 9(38), DOI: 10.1.7485/ijst/2016/v9i38/102967, October 2016

[4] Itunuoluwa Isewon, Jelili Oyelade, Olufunke *Oladipupo "Design and Implementation of Text To Speech Conversion for Visually Impaired People"* International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 7– No. 2, April 2014

[5] Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad *"VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT"* Journal of Applied and Fundamental Sciences

[6]PySimpleGUI Inc, *Python GUIs for Humans. https://pysimplegui.readthedocs.io/en/latest/*

[7] David Amos. (2012). *The Ultimate Guide To Speech Recognition With Python. https://realpython.com/python-speech-recognition/*

[8] Joska de Langen. (2012). *Sending Emails With Python. https://realpython.com/python-send-email/*

[9] Vikas Jha. (Sep 19, 2020). *Sending Images, File or attachment on Telegram Using Python Program. https://youtu.be/NYT1KFE1X2o*

[10] adityasaini70. (29 May, 2021). *How to print all files within a directory using Python?*
https://www.geeksforgeeks.org/how-to-print-all-files-within-a-directory-using-python/