TEXT AND SPEECH INTERCONVERSION

Rakshit Kulkarni
191IT245
Information Technology
National Institute of Technology Karnataka Surathkal, India 575025
Email: rakshitkulkarni.191it245@nitk.edu.in

Kiran Kumar J M
191IT126
Information Technology
National Institute of Technology Karnataka Surathkal, India 575025
Email: kirankumarjm.191it126@nitk.edu.in

Mohit Awachar
191IT231
Information Technology
National Institute of Technology Karnataka Surathkal, India 575025
Email: mohit.191it231@nitk.edu.in

Anshul Patel
191IT208
Information Technology
National Institute of Technology Karnataka Surathkal, India 575025
Email: anshupatel.191it208@nitk.edu.in

ABSTRACT:

We built two Applications (TTS & STT) for interconversion between text and speech signal. Text to speech makes the online world and beyond available to people with learning disabilities, visual impairments, and literacy challenges. While speech to text saves a lot of typing time. Two separate Graphic User Interfaces (GUI) are built of these tasks using python tkinter library. TTS applications turn text-based content into audio and enable all users to access information and communicate with others. While STT does the reverse. The SST module also allows searching converted text on the web(Google Search). We use python's pysttx3 library for TTS and python's speech_recognition library for STT respectively.Active internet connection is needed for speech to text conversion.

Keywords:

Communication, FreeTTS, FreeSST, python, pysttx3, tkinter, Speech, Text-To-Speech, Speech-To-Text , speech_recognition

INTRODUCTION:

Text-to-speech (TTS) convention transforms linguistic information stored as data or text into speech. It is widely used in audio reading devices for blind people now a days .Audiobook is one of many application of TTS.In recent times, the use of text-to-speech conversion has grown far beyond the disabled community to become a major adjunct to the rapidly growing use of digital voice storage for voicemail and voice response systems. Also developments in Speech synthesis technology for various languages have already taken place. On other hand, Speech-To-Text (SST) converts speech signal to text signal. Main applications for this conversion are online assistants(Google Assistant, Siri, etc.). SST can save a lot of energy and time required for typing. Authors can use this for writing books, journals, etc.

TTS and STT both are useful in many areas like banking-finance sector ,ATM, healthcare, learning , publish-media etc.But we have selected this topic mainly focussing on two sectors namely EDUCATION and HEALTH .We choose this problem is due to its wide application in remote areas and health sector. We build the application with limited resources available with us here we want to show its usage and expand its availability. We wanted to limit use of the internet so that it can be useful in remote areas. It can be useful to areas like hospitals where with just this application they can inform the information to each and every patient. In situations like pandemic this can play a vital role for spreading news or any notice to anywhere(not restricted to hospital,can also be in places like a civil court,police station etc) without using a piece of paper or pen as it may cause spread of the pandemic disease. Apart from this it can make life easier for people with disabilities. In India if we can implement this simple idea in any school(normal or schools for disabled) then it may help them to learn things and share their point of views without any problem and hesitations.Along with this it can also make their life easy , smoothing and self confident.TTS is very beneficial for blind students while SST is very useful for deaf students. For all this problem on ground level there is no such prominent solutions are visible in our day to day life.By introducing this idea among deaf people where they were using old methods(like sign language) or in hospitals(old pen-paper technique) it can create a great impact in their lives.We are not here ensuring that this is the only 100% effective solution but then also we have tried make it a light user friendly GUI which is not going to use internet to convert text to speech.Also the speech i.e audio file will also get stored in the system itself so that record can be maintained for audio files. On other hand speech to text requires active internet connection.

TEXT AND SPEECH INTERCONVERSION

LITERATURE SURVEY:

During this project we analysed and understood the concepts for speech to text conversion and for vice-versa.We also learned about various libraries of python which made possible to create GUI for our idea and application.For this we searched various websites,videos and papers for the useful content.
Some of the websites which made our work easier are:

- https://www.tutorialspoint.com/python/python_gui_programming.htm: This site also helps us to know how to add a widget and what are its types.
- https://docs.python.org/3/library/tkinter.html: We used this site to know about the arguments which are present in the method. Eg:-btn.place(x= ,y=) etc..
- https://www.geeksforgeeks.org/python-gui-tkinter/: We used this site to go-through how the widget works and mainly how to create buttons, how to change background color etc..
- https://www.studytonight.com/tkinter: We used this site to learn about python tkinter library and also how to create a GUI and how to add a widget and all the things we learnt from this.
- https://www.youtube.com/playlist?list=PLCC34OHNcOtoC6GglhF3ncJ5rLwQrLGnV: This is a YouTube playlist on Comedy channel for tutorial of tkinter.

AGer going through 5 research papers we liked the paper which made TTS using java and modifying it ,similarly we use python instead of java. Ifeanyi Cosmas Nwakanma ,Ikenna Oluigbo and Okpala Izunna were the authors for the paper titled TTS and got published on 5,MAY 2014 in International Journal of Research in Information Technology(IJRIT)

AUTHOR	METHODOLOGY	MERITS	LIMITATION
Ifeanyi Cosmas Nwakanma , Ikenna Oluigbo, Okpala Izunna	Object Oriented Analysis and Development Methodology (OOADM)	Discarding old system and making new system	Prerequisite knowledge about the language JSML
S.Venkateswarlu, Duvvuri B K Kamesh, Sastry Jammalamadaka	Text-to-speech device consists of two main modules, the image processing module and voice processing modules.	NO internet	Heavy application i.e. more time in processing
Partha Mukherjee, Soumen Santra, Subhajit Bhowmick, Ananya Paul, Pubali Chatterjee, Arpan Deyasi	Using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP)	Made own gui along with soGware	No web based real-time synthesis system
Chaw Su Thu Thu 1, Theingi Zin	Use of OCR system and MATLAB	Pace of the voice can altered	Installation of WIN 32 is compulsory
Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik, Supriya Agrawal	Not done as only theoretical research	provided in depth theoretical knowledge	No practical implementation

Problem Statement

Build text to speech and speech to text synthesizer separately. Save corresponding text file in computer. Implement voice change(male/female), output voice speed and other necessary features in text to speech. Implement web search of converted text in speech to text.

OBJECTIVES

The specific objectives are:

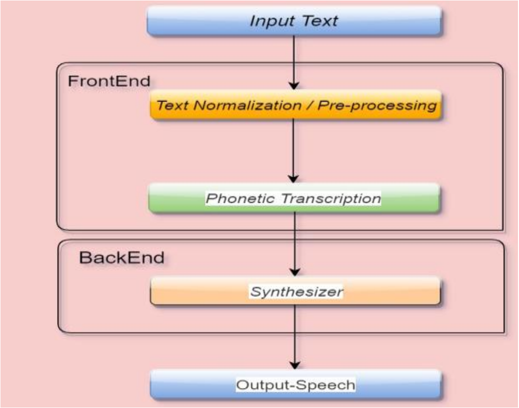
- To enable the deaf and dumb to communicate and contribute to the growth of an organization through synthesized voice.To enable the blind and elderly people enjoy a User-friendly computer interface.
- To provide an effective tool to spread information without using pen or paper in various places like hospitals,schools etc.
- Convert text to speech without the internet so that all the above objectives can be achieved

even in remote areas,also to keep record of all the audio and text files.

- Save typing time using speech to text and web search spoken words and implement Google search.

METHODOLOGY:

TTS:



Input Text: Our Text-To-Speech GUI contains mainly two types of input Options i.e the first one is the user can import the text nothing but .txt extension file from any directory of his/her device.And also the another option is user can enter anything in the given text box so that it can be converted to speech.Text language must be English.

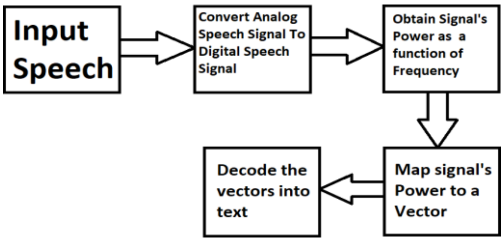
Front-End: Text Normalization / Pre Processing:The front end does two functions, the first of which is to convert raw text containing symbols, such as, number and abbreviations into equivalent words.This process is termed as normalization or pre-processing or tokenization. As part of a text-to-speech (TTS) system, the text normalization component is typically one of the first steps in the pipeline, converting raw text into a sequence of words, which can then be passed to later components of the system, including word pronunciation, prosody prediction, and ultimately waveform generation.

Phonetic Transcription : The second function is to assign phonetic transcription to each word. The front end converts each phonetic until of the text into units like phrase, clauses and sentence.This process of assigning phonetic transcription to words is called text-to-phoneme conversion. Phonetics transcription and prosody information together will form the symbolic linguistics representation.The front interfaces always engage with phonetic conversion to each unit, and divides and marks to form a speech tree or pattern tree using the speech unit which configures the tune and rhythm through phrases, clauses, and sentences. This process of transcription is known as text-to-phoneme (TTP) or grapheme-to-phoneme (GTP) conversion.

Back-End: Synthesizer : The back-end—oGen referred to as the synthesizer—then converts the symbolic linguistic representation into sound. The back-end will also include the computation of the target prosody such as pitch counter, phoneme duration which are then imposed speech output.The Text can be converted to speech using pyttsx3.It is a text-to-speech conversion library in Python.It uses native speech drivers when available and works completely offline.

Output Speech: AGer pre-processing and synthesis of text the speech is generated.The speech is the output in this Text-To-Speech GUI.

STT:


Input Speech
Convert Analog Speech Signal To Digital Speech Signal
Obtain Signal's Power as a function of Frequency
Map signal's Power to a Vector
Decode the vectors into text

Input Speech: Speech signal is inputted from the user using a microphone on the device.Allow Since the surrounding noise may vary, we allow the program a second or too to adjust the energy threshold of recording so it is adjusted according to the external noise level. This is done using

Convert Speech to Text: We use recognise_google() for this. It's working is as follows.Voice activity detectors (VADs) are used to reduce an audio signal to only the portions that are likely to contain speech. This prevents the recognizer from wasting time analyzing unnecessary parts of the signal. 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. Once digitized, the signal's power is obtained as a function of frequency. Then it is mapped to a vector of real numbers.To decode the speech into text, groups of vectors are matched to one or more phonemes—a fundamental unit of speech.

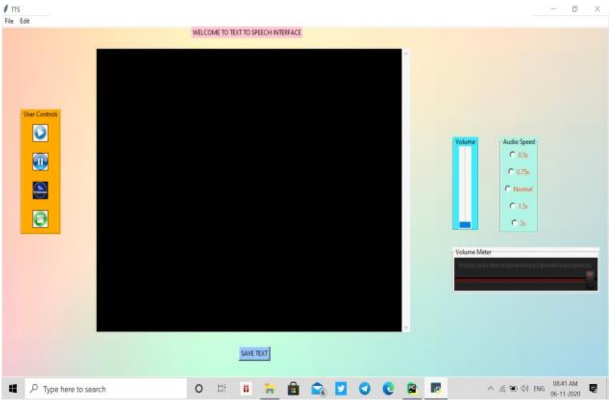
Output the text: Output is produced on text box in the application

Google search speech(if needed): If user asks to search the resluted text in internet, this is done by clicking on Google symbol which essentially gets contents in text box and search it in Google using open_new_tab() function in webbrowser library.

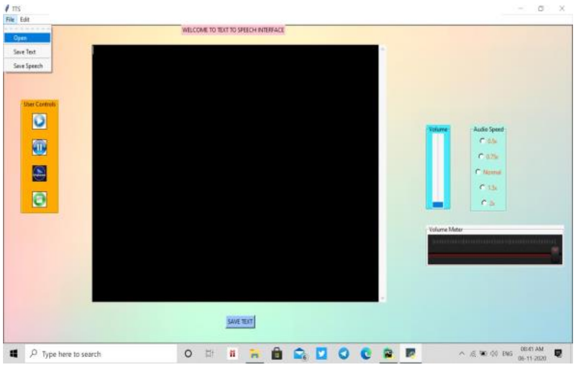
TEXT AND SPEECH INTERCONVERSION

RESULTS AND ANALYSIS:

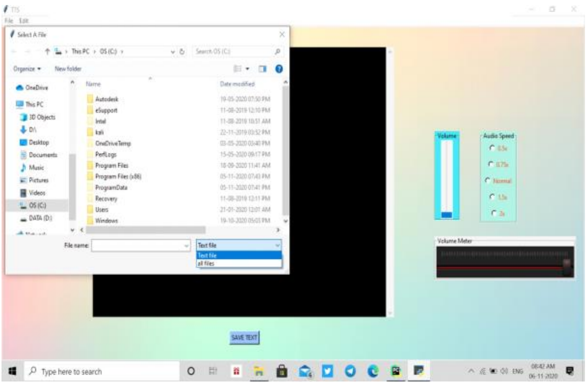
TTS



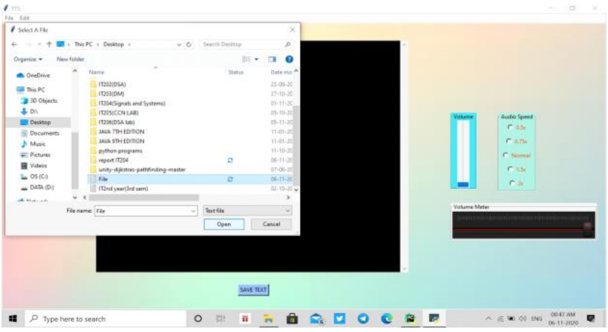
This is our Text To Speech Graphical User Interface(GUI) which consists of a widget like text field, Buttons, Slider, RadioButton and Menu Bar, Label Frame, etc. By using all the widget we made a GUI which converts Text-To-Speech.



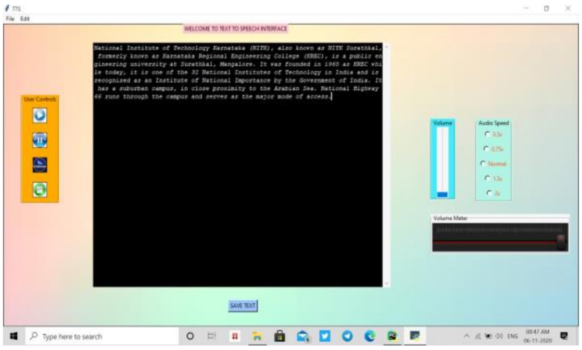
We can input Text in Text Field in two different ways.In that First way is by opening a text file(.txt) from the device.In the above Figure you can see there is a file menu bar which consists of an open command to open text(.txt) file.



After clicking on the open command you can see there is another window named Select A File is opened.Using this we can select any text file from any directory.

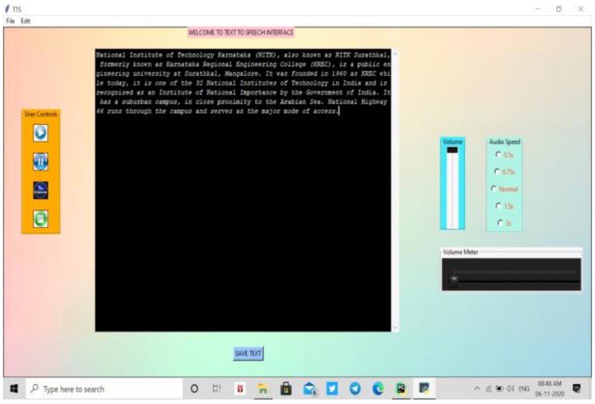


After selecting a text file(.txt) click on the open button in the Select A File named window to copy the content from the selected text file(.txt) to the Text Field which is in the TTS named window.



National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.

The Content from the selected text file(.txt) is copied into the Text field.

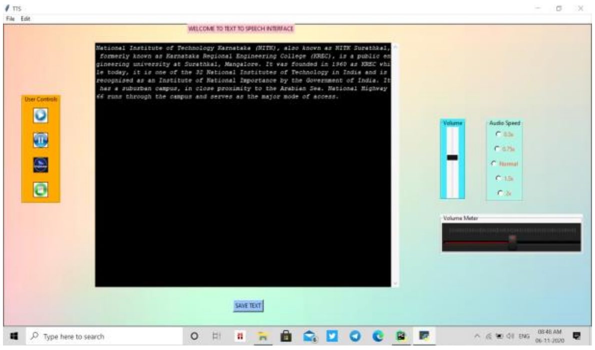


National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.

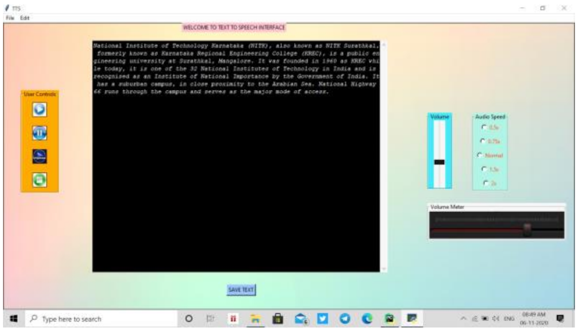
After copying the content by clicking on play button which is present in user control labelFrame to start the speech.Click on Pause button to pause the speech,unpause button to unpause the speech and stop button to stop the speech. And on the right side you can see Volume LabelFrame which is used for increasing or decreasing the volume.The volume is corresponding to the volume meter where you can see if volume is low then in the volume meter the pointer is towards the left.

TEXT AND SPEECH INTERCONVERSION

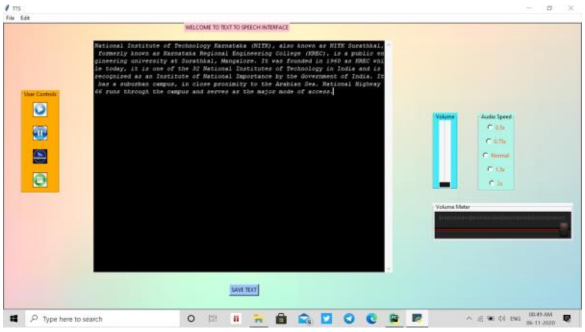
Below images show how the pointer changes according to volume.



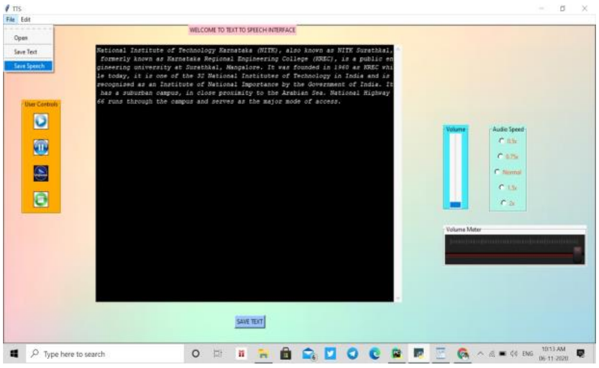
National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.



National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.

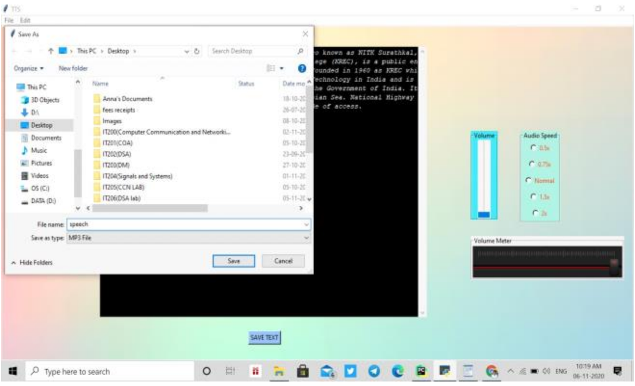


National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.



National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.

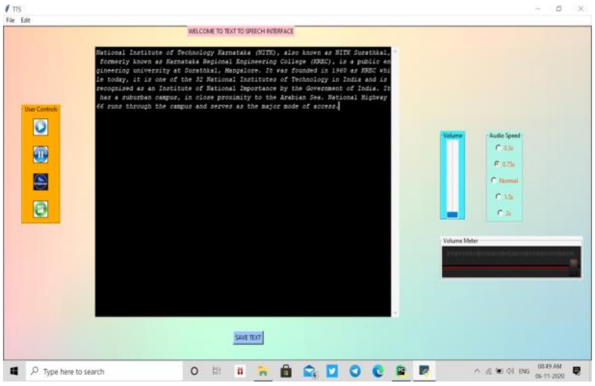
If we want to save the speech just goto File menu and click the Save Speech command.



File name: speech

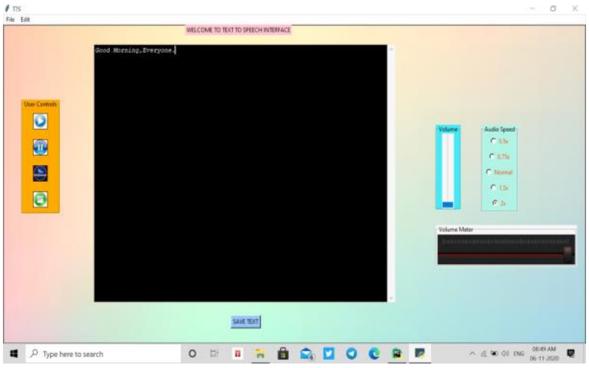
Save as type: MP3 File

AGer clicking on the Save speech button the new window will open named as Save As there you can select any directory where you want to save the audio file and give the file name and click on the save button.



National Institute of Technology Karnataka (NITK), also known as NITK Surathkal, formerly known as Karnataka Regional Engineering College (KREC), is a public engineering university at Surathkal, Mangalore. It was founded in 1960 as KREC while today, it is one of the 32 National Institutes of Technology in India and is recognised as an Institute of National Importance by the Government of India. It has a suburban campus, in close proximity to the Arabian Sea. National Highway 66 runs through the campus and serves as the major mode of access.

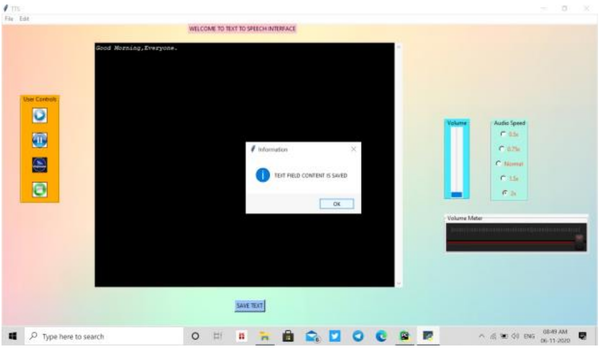
To change the Speech speed click on the options which are present in the Audio Speed LabelFrame.



Good Morning,Everyone.

The other option for filling a text Field is by writing anything into the given Text Field.

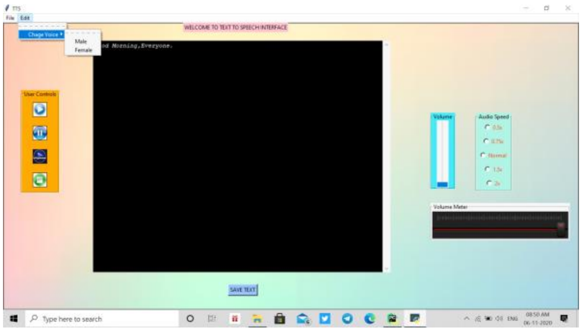
TEXT AND SPEECH INTERCONVERSION



Good Morning,Everyone.

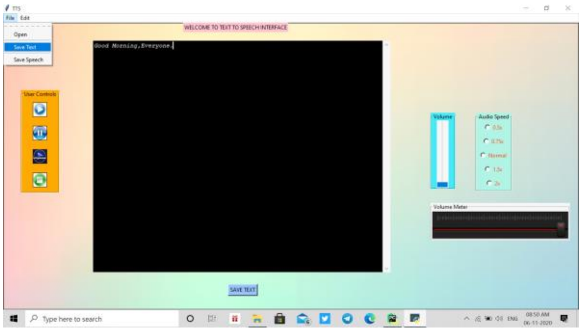
TEXT FIELD CONTENT IS SAVED

Once you stop writing in the text field click on the save text button which is just below the text field to save the content present in the text field.



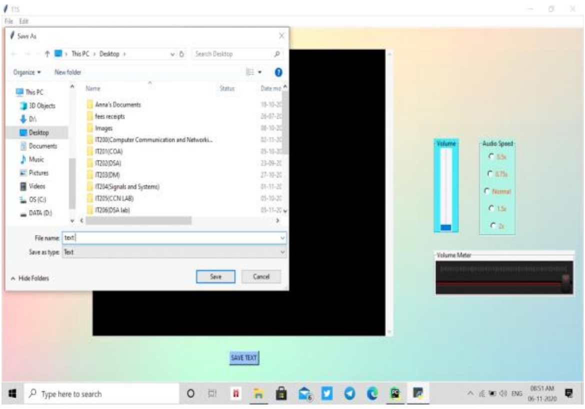
Good Morning,Everyone.

Here in the Edit menu bar we can see the change voice command which is used to change the voice to male and female.Click male to change the voice from female to male and vice-versa.



Good Morning,Everyone.

To save the content from the text field goto File menu and click on the save Text command.

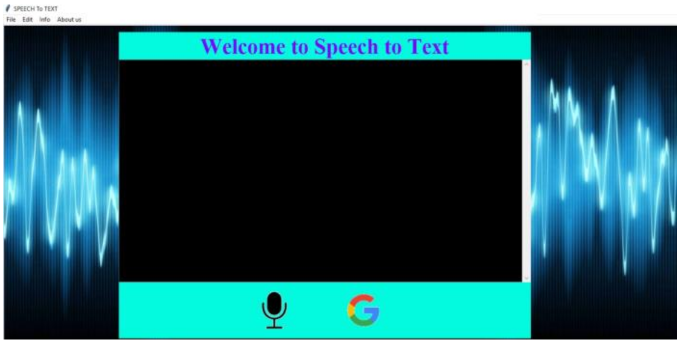


File name: text

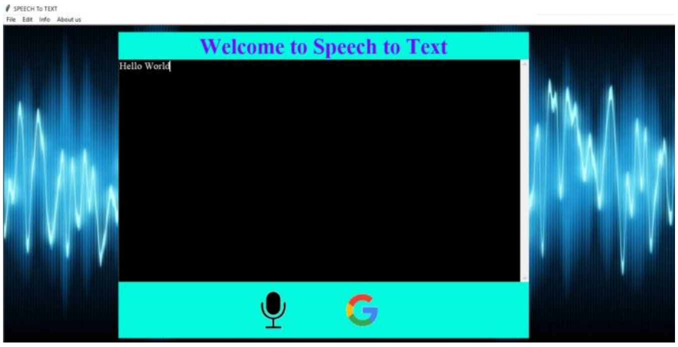
Save as type: Text

AGer clicking on the Save Text button the new window will open named as Save As there you can select any directory where you want to save the Text file and give the file name and click on the save button.

STT

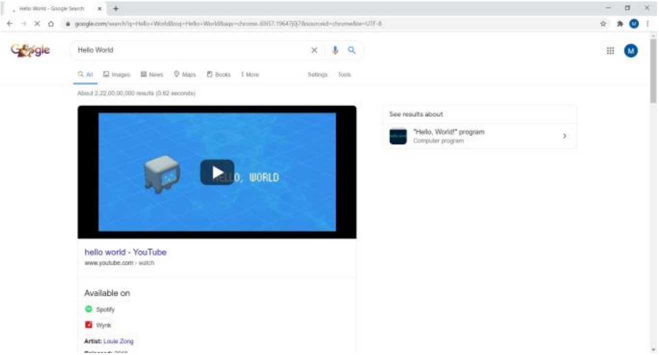


This is a gui for speech to text.It contains a text field, canvas and button widgets. It also has menu bar



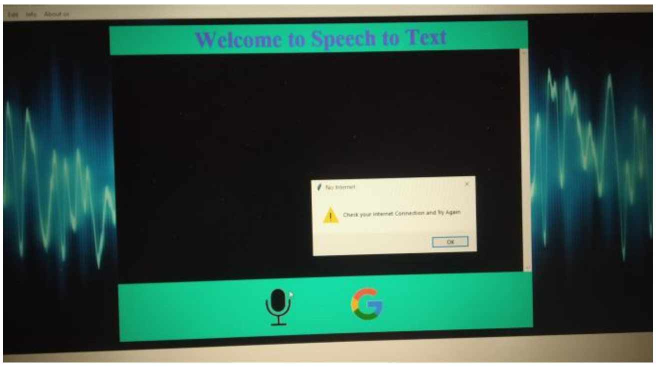
Hello World

AGer pressing the microphone button, the system will output an audio signal saying speak now. AGer this user can give speech input. This speech signal is converted to text and displayed in the Text Field.



This opens a new tab in browser which searches web for contents in our text field

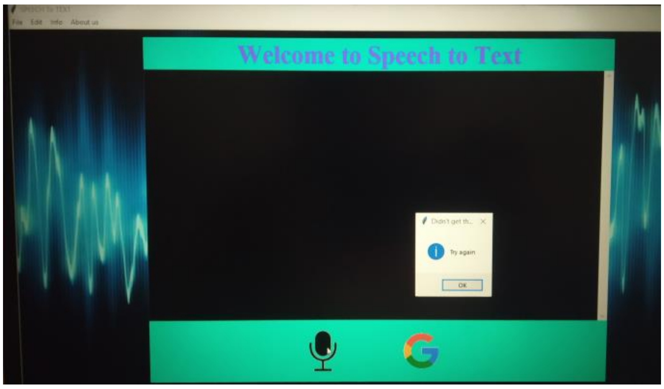
TEXT AND SPEECH INTERCONVERSION



No Internet

Check your Internet Connection and Try Again

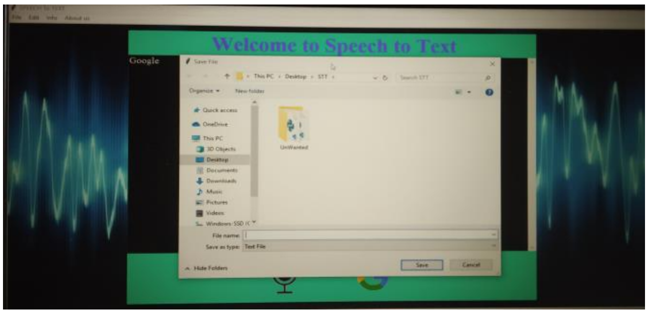
While speaking if user is not connected to internet a window pops up addressing the issue



Didn't get th...

Try again

If recognizer don't understand user's input this window pops up addressing the issue



Save and Save As options save files on the computer while Exit options closes our application. We can save our output in the text field using the save in file menu. The supported formats are .txt, .py,.html

The clear button clears contents in the text field. The info button in info menu shows guide to operate the application.The about us button gives information about developers

FEEDBACK:

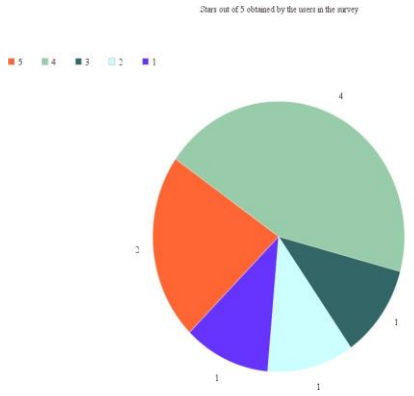
We shared our application with people and via google survey forms ,we obtained many responses including positive and negative reviews.Most of the outline of it were as follows:

POSITIVE:

- No internet use in TTS so it can be useful in rural or remote areas.
- In rural areas this concept is very new and by proper implementation ,it can make the edge in doing the easiness of their work.
- Record of audio files can be maintained in TTS
- The STT option of google is available.

NEGATIVE:

- Slow working of STT application.
- Internet usage in STT application.
- No option available to change the speed of the audio file in TTS.



Conclusion:

Text-to-Speech synthesizers have steadily developed from the last few decades to gain the current shape. It's never been so easy to use a text-to-speech program, as just one click and your computer will speak any text aloud in a clear, natural sounding voice. We have identified the various operations and processes involved in text to speech synthesis. We have also developed a very simple and attractive graphical user interface which allows the user to type in his/her text provided in the text field in the application. Our system interfaces with a text to speech engine developed for American English.We have also implemented it successfully in the marked needy areas.Also in future we will be trying it to add more and more regional and international languages so that linguistic barrier will be demolished.Also for speech to text converter we will try to make it completely offline.

TEXT AND SPEECH INTERCONVERSION

CONTRIBUTION:

Work done:

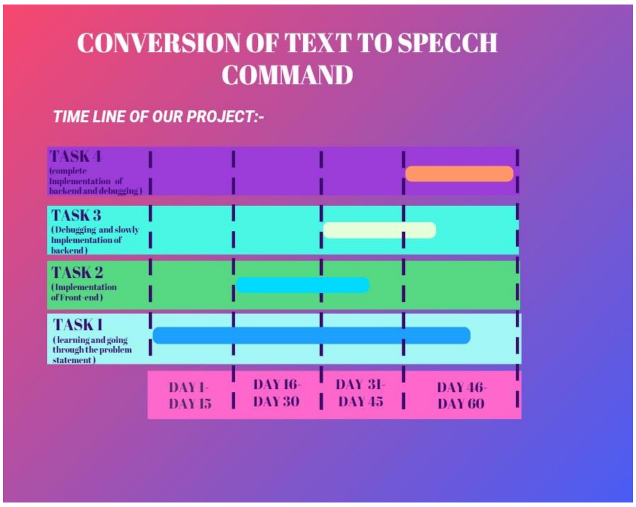
- **STT**

1. Anshul: Draw outline of gui and build start window.
2. Mohit: Implement the conversion from speech to text(core)
3. Rakshit: Build menu and 2 buttons on welcome screen and exception handling in conversion from speech to text
4. Kiran: Implement command for all options in the menu namely save, saveas, clear, exit, info and about us.

- **TTS**

First we all studied the referenced papers and we spent some time understanding the problem statement and its related things.AGer that we started making our GUI. The GUI contains FrontEnd and BackEnd.

1. **Mohit**
   - Creating window and adding menu bar(Front End)
   - He worked with Volume Slider i.e how to change the volume and corresponding to volume how volume Meter will change and also how to change the audio speed. (Back End)
2. **Anshul**
   - Creating User Control LabelFrame and adding play , pause, unpause and stop button (Front End)
   - File menu bar contains sub menu bars open,save speech and save text and Edit menu bar contain change voice command. (Back End)
3. **Kiran**
   - Creating Volume slider,Volume Meter and Audio speed controls. (Front End)
   - He worked with Text Field like taking the input text and adding it into the Text box and also saving the text content as a .txt file. (Back End)
4. **Rakshit**
   - Creating a Scrolled Text field and saving the text button. (Front End)
   - He worked with the User Control LabelFrame which consists of play pause unpause and stop button. (Back End)
   - He also did all the testing.



CONVERSION OF TEXT TO SPECCH COMMAND

TIME LINE OF OUR PROJECT:-

-