Audiobooks that converts Text, Image, PDF-Audio & Speech-Text

for physically challenged & improving fluency

Kurra Santhi Sri
Associate Professor, Information Technology
*Vignan's Foundation for Science*Technology and Research
Vadlamudi, India
srisanthi @gmail.com

Chennupati Mounika, [3]Kolluru Yamini
Information Technology
*Vignan's Foundation for Science*Technology and Research
Vadlamudi, India
mounikachennupati82@gmail.com
yaminikolluru123@gmail.com

2022 International Conference on Inventive Computation Technologies (ICICT) | 978-1-6654-0837-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICT54344.2022.9850872

Abstract—The primary goal of our technology is to create a speech recognition system for physically handicapped persons. Nowadays Many independent gadgets are increasingly being used for communication reasons. One of the beneficial trends is the invention of a person's ability. This program allows for the conversion of text-to-voice, image-to-Speech & text, PDF-to-voice, speech-to-text. Voice, or another style of voice in a speech file, can be transformed into text. As a result, instead of reading, you can listen to the book, or instead of writing you can speak, and also you can extract the text from the image. This application is beneficial to persons who have physical limitations such as being deaf, blind, or having different abilities and Users who find typing difficult, painful, or impossible, as well as those who can understand what others are saying. This project uses Visual Studio to display user-friendly Python Code, and this file contains a GUI/Voice command.

Keywords— Speech Recognition, Visual Studio Code, Python, Different Abilities, Graphical User Interface/Voice command

I. INTRODUCTION

The primary aim is to develop an Audiobook System. The Audiobook system is nothing but reading out the text, PDF file, or the words in the image. This system involves the conversion of text into voice, speech into text, PDF into speech, and picture to voice. For speech to text, we will take input as a speech and get the output in text format , whereas for PDF to Speech, we can take input as a PDF file and we will get output in speech format, or we can take input as a picture then every word in the picture is converted into voice format, or we can take input as text and get the output in speech format. This was useful for deaf and blind people, people with less literacy, people who can read but cannot speak, and this will also free up your time to accomplish other tasks. People will be more enthralled if They're talking about being able to recognize their own voice. It is available to anybody can be viewed from any location. This application makes converting it into multiple forms simple. It's probable that in the future, this will be the finest and easy-to-use online application.

Modules:

- Text-to-Speech (TTS): TTS module was primarily used to translates the given inputted text into speech. It will help to determine how to pronounce text if the user is unable to pronounce the given text. It is helpful for the people who cannot speak.
- Pdf-to-Speech: To read a pdf-formatted book, utilize the PDF-to-Speech module.
- Picture-to-Speech (p2S): In this module, the user provides input in the form of an image, and the module turns the image's text into audio.
- Speech-to-Text: This function turns the provided speech into txt

AIVA-It is a google tool suggested for developing a voice-controlled individual collaborator capable of doing a variety of tasks.

II. LITERATURE SURVEY

Before beginning the project, we conducted a poll to determine How many people are interested in using audiobooks and how many are not. By this we determined that majority of people are interested in audiobooks. So, we did research on why people prefer audiobooks, as well as the reasons for their disinterest in audiobooks.

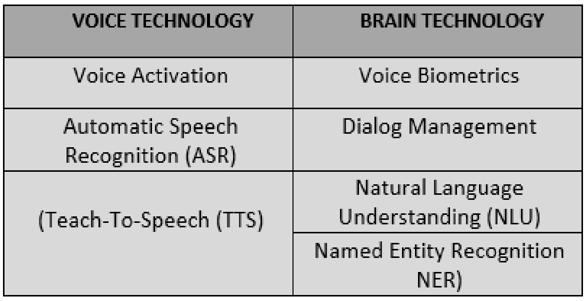
Nikhat Parveen, et.al says that the majority of the people feel that it is more convenient to listen than toread.It is especially beneficial for folks who comprehends but cannot read the language or for the people who can't speak. This Audiobook System is also excellent for deaf and blind persons, who can grasp or feel the text by listening to it rather than reading it. It is also beneficial for teachers and students. Devidutta Dash, et.al says that Teachers advance from Chalk and Talk to Touch and Teach Now a days students developed the habit of learning on their own. Some people dislike audiobooks because they believe they will reduce their reading skills.

Some people feel that they should read a book rather than listen to a book to improve their language skills. Hasan U. Zaman,et.al Says While some people are pleased that they are unable to read due to a lack of literacy, disabilities, or that they can multitask by listening to a book while doing other tasks. Others claim that they do not need to keep a book with them at all times if they become unexpectedly interested. We learned from the survey responses that the audiobook method has more advantages than disadvantages.

III. METHODOLOGY

Python is used throughout the project. We have done in visual studio. Our project is entirely online. There is a total of 4 modules such as text into voice, speech into text, PDF into speech, and picture to voice and text. The tasks include reading a book, listening to a recorded text, spelling a word, and writing a phrase.

TABLE 1.



VOICE TECHNOLOGY	BRAIN TECHNOLOGY
Voice Activation	Voice Biometrics
Automatic Speech Recognition (ASR)	Dialog Management
(Teach-To-Speech (TTS)	Natural Language Understanding (NLU)
	Named Entity Recognition NER)

Technologies for building intelligent systems that communicate with humans using natural language.

A Text to Speech (TTS)

Text to Speech can assist in reading a given text.It can transform words into audio files by pressing the submit button on a computer or other digital device.

Text is converted to phonemic representation using Text to speech, and each word is assigned a phonetic transcription. Then, as Output, this may be transformed into sound waves. The output sounds like comprehensible human speech.

To create audio output, input text is routed via many blocks. As first step breaking sentence into small units(words) and splitting words as phonemes based on the pronunciation. TTS first tokenizes the provided sentence into words, then checks the language default, which is English, before allowingus to select the speed of reading text.

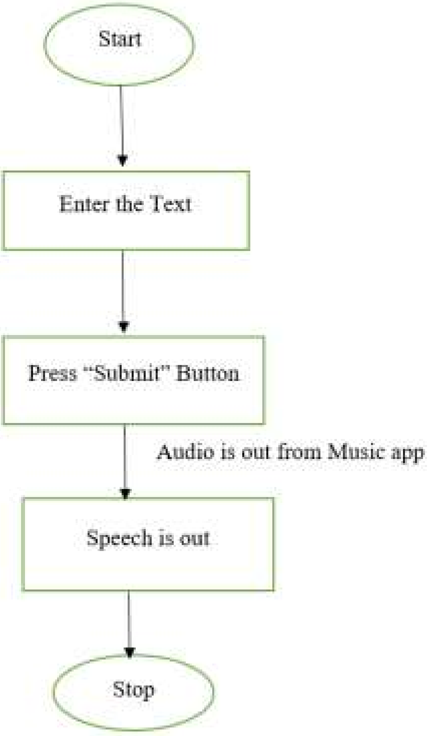


Fig. 1. Text to Speech

B.PDF to Speech

Pdf to Speech When someone prefers to listen to somethingrather than read, pdf to speech can be useful.

At the end of the execution we have a display screen with a file upload option where it accepts only PDF oriented documents. Here, we need select PDF documents by clicking on file option and then upload the required PDF file. By clicking on submit button the PDF file is converted into Audio format. In case if the uploaded is document is not a PDF format then it shows error message as "Please select the Pdf file" and we also have an option of quitting from the page.

- Import the Python modules PyPDF2 and Pyttx3.
- Open the PDF file.
- To read the PDF, use PdfFileReader().
- We simply need to provide the PDF's path as the argument.

To select the page to be read, use the getPage() method. Using extractText, extract the text from the page (). Creates a pyttx3 object. To voice out the text, use the say() and runwait()functions.

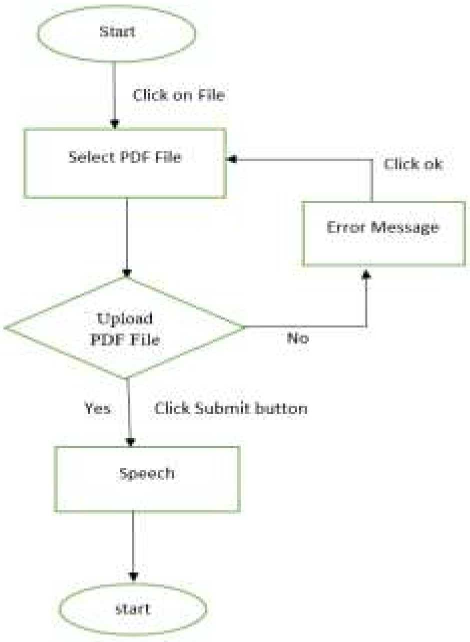


Fig. 2. PDF to Speech

C. *Image to Speech*

Our goal is to turn a text image into a string of text, save it to a file, and listen to what's written in the image using audio. For this, we'll need to import numerous libraries.

At the end of the execution we have a option of performing Three operations. The first operations is conversion of Image to Speech where the words in the image is recognized and converted into Text. The second operation is conversion of Image to Speech where the words in the image are recognized and converted into Speech. The third operation is conversion of Text to Speech and fourth option is exit. Therefore based on the users selection it performs the operations and gives the result.

A type of Python-tesseract is Pytesseract (Python-tesseract). It's a Python-based optical character recognition (OCR) tool funded by Google. pyttsx3: It's a text-to-speech library that works on any platform and may be used offline. Python Imaging Library (PIL): It improves your Python interpreter's ability to process images. Google trans: It's a free Python module that uses the Google Translate API.

We can translate the text into any language you like. Japanese, Russian, and Hindi are only a few examples. The only stipulation is that GoogleTrans understands the destination language. Additionally, pyttsx3 will only speak the languages that it recognizes. It's a free Python module that uses the Google Translate API.

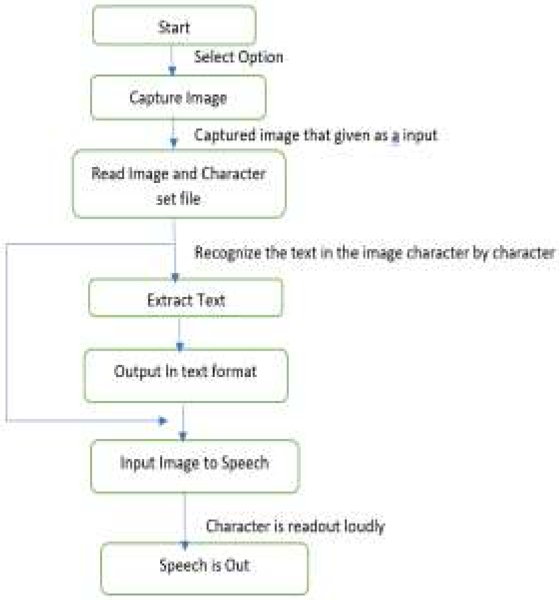

Start
Select Option
Capture Image
Captured image that given as a input
Read Image and Character set file
Recognize the text in the image character by character
Extract Text
Output In text format
Input Image to Speech
Character is readout loudly
Speech is Out

Fig. 3. Image to Text and Speech

D. *Speech To Text*

This is performed by utilizing Google Speech Recognition. There are Some offline Recognition systems, like Pocket Sphinx, Have a long installation process that necessitates the installation of several dependencies. Google Speech Recognition is one of the most user-friendly.

In this module At the end of the execution we get the result of a command appeared on the screen as "say now". Then we need to give a voice command, Later it converts Voice command into Text format. If the voice command is not provided or not recognized it shows the message of "Unknown error occurred".

To initialize the library, we must first import it and then use the init() function. Two arguments can be passed to this function. We'll utilize the say() function after initialization to have the software speak the text. Finally, run and wait is used to deliver the speech None of the say() messages will be said unless the interpreter encounters runAndwait().

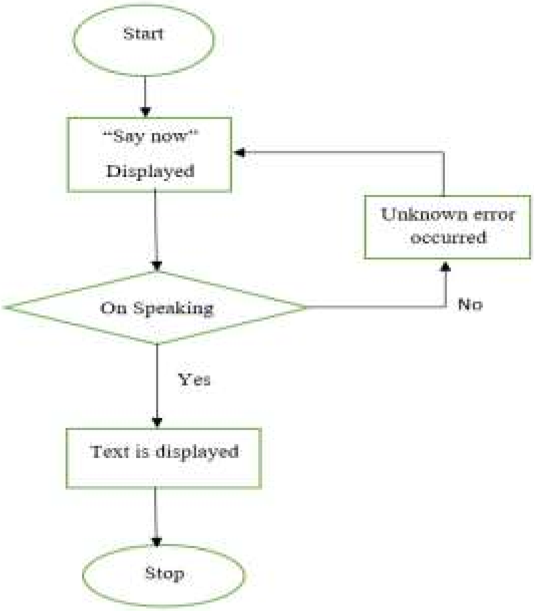


Fig. 4. Speech to Text

Workflow:-

A programming interface enables two programs to interact with each other. In the end, a Programming interface acts as a messenger that sends your request to sender you are referring then conveys the result back as a response. Setting Extraction is the process of naturally distinguishing structured data from unstructured or potentially semi-structured machine-intelligible information.

In the vast majority of circumstances, this process entails preparing human language messages using normal language handling procedures(NLP).Ongoing mixed media record reparation activities such as programmed explanation and material extraction from images/sound/video might be considered as establishing extraction test results.

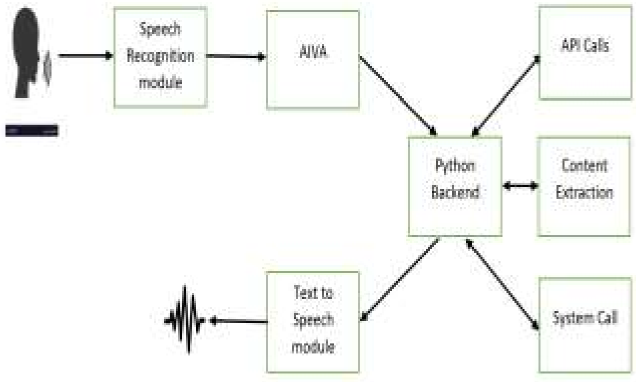


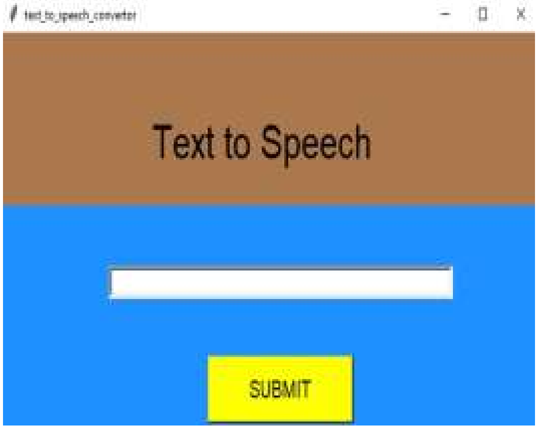
Fig. 5. Workflow

.

IV. RESULT AND DISCUSSION

These are the modules of our system. In the text to speech module, we will upload text to the following web page, On pressing the 'submit' button we will get entered text in the speech format. In the PDF to speech module, we will click the 'file' button then we can able upload any pdf file, On pressing 'submit' we will get the output in audio format. The image to text module has 4 options like Image to Speech, Image to Text, Text to Speech, and Quit. On selecting any of those options we will get output in either text or speech format. In Speech to Text, we get the command as 'say now' By speaking we get output as text format, If we didn't say anything we get 'unknown error occurred'.

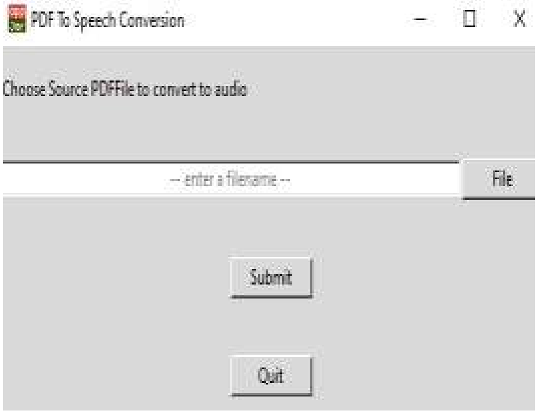
Text to Speech:-


text_to_speech_convertor

Text to Speech

Fig. 6. Text to Speech

PDF to Speech:-

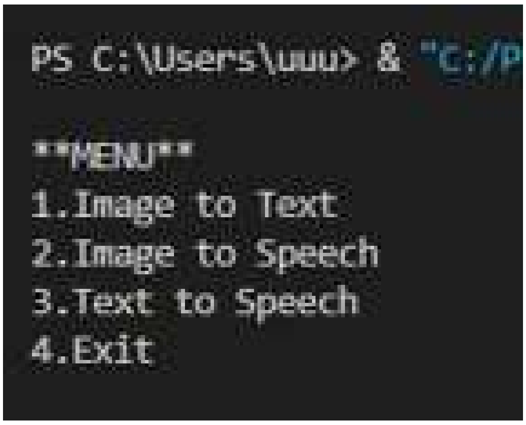

PDF To Speech Conversion

Choose Source PDFFile to convert to audio

-- enter a filename --

Fig. 7. PDF to Speech

Image to Speech:-

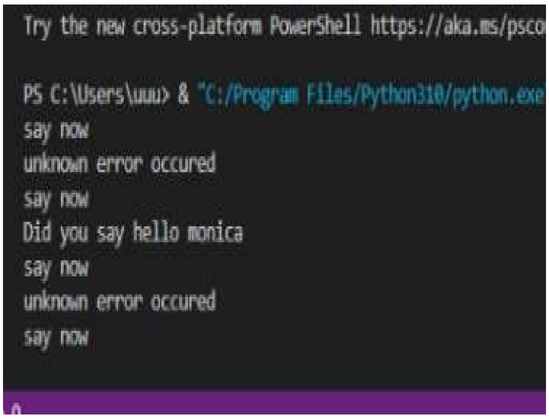


```
PS C:\Users\uuu> & "C:/P

**MENU**
1.Image to Text
2.Image to Speech
3.Text to Speech
4.Exit
```

Fig. 8. Image to Speech

Speech to Text:-



```
Try the new cross-platform PowerShell https://aka.ms/psco

PS C:\Users\uuu> & "C:/Program Files/Python310/python.exe
say now
unknown error occured
say now
Did you say hello monica
say now
unknown error occured
say now
```

Fig. 9. Speech to Text

V. FUTURE SCOPE

According to our current system, we declare that our system is beneficial for the people who are physically challenged like blind and deaf, people with a lack of education, people who are unable to speak, and people who will do multiple tasks at a time.

At present we have four modules such as speech to text , text to audio, pdf to audio conversion, and picture to speech. It is helpful for listening to books and reading the text. We also plan to upgrade this project by adding some other features that will allow future generations to acquire knowledge. As we can see from Alexa and other examples of current technology, this kind of conversion have a greater impact on the generation. On the basis of this life is made easier. In the future, we will upgrade the picture to speech module by not only converting text in the image by also describing the picture, and also we will add some more modules like audio file to text. In the future, we will bring our vision to reality by creating a Virtual Library System. We intend to create more modules that allow users to upload any type of book and receive audio files as output. It will allow us to learn knowledge in the most efficient manner possible.

VI. CONCLUSION

We started our audiobook system mainly for physically challenged people and to improve fluency for that we have developed four modules they are Text to Voice, Speech to Text, image to audio, and PDF to audio. Every module has its own features. text to speech module will read out the entered text, which will help you to spell the word correctly and mute people. Speech to text module helps to convert the speech into text. It is helpful for the deaf and people can improve their reading skills. Image to Speech this module will convert the text in the image into speech as well as text formate.PDF to Speech will convert the entire PDF file into Speech. This will also be helpful for the physically impaired, people with less education, and multitasking people. Our proposed solution will be accessible as a multilingual application and addition of some more modules and features in the next few days, allowing customers to use the program in their native language with ease. As a result, instead of reading, they can listen to a book, and also in place of writing they can speak[2] and it also extracts text from the Image. This system is extremely useful in everyday life.

REFERENCES

[1] Library Audiobook System Using Speech Recognition. Nikhat Parveen, Priyanka CH.Ruchitha Y.Geeteeka Y.VarniPriya

[2] Speech Recognition using Android. Bhushan Mokal, Sahil Patil, Aniket Kale, Prof. Archana Arudkar in 2020 https://www.irjet.net/archives/V7/i2/IRJET-V7I2628.pdf

[3] Ayushi Trivedi,Navya Pant, Pinal Shah, Simran Sonik and SupriyaAgrawal Department of ComputerScience,NMIMS University, Mumbai, India. Corresponding Author: Navya Pant. Speech to text and text to speech recognition systems-A review Artificialintelligence(AI),sometimes_called_machine intelligence_https://www.iosrjournals.or g/iosr-jce/papers/Vol20-issue2/Version-1/E2002013643.pdf.

[4] Huang,J.,Zhou,M.andYang,D.,2007,January.ExtractingChatbo Knowledge from Online Discussion Forums.In IJCAI(Vol-7,pp.423-428

[5] CMUSphnix Basic concepts of speech - Speech Recognition process". http://cmusphinx.sourceforge.netlwiki/tutorialconcepts Hasan U. Zaman, Saif Mahmood, Sadat Hossain, Iftekharul Islam Shovon, Python Based Portable Virtual Text Reader

[6] Dept. Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

[7] B.Marr,The Amazing Ways Google Uses Deep Learning AI. CortanaI ntelligence.GoogleAssistant, AppleSiri

[8] Varish, N., Parveen, N, et.al, Image Retrieval Scheme Using Quantized Bins of Color Image Components and Adaptive Tetrolet Transform, IEEE Access 2020, 8, pp. 117639-117665, 9121956

[9] Parveen, N., Roy, A., Sai Sandesh, D., Sai Srinivasulu, J.Y.P.R., Srikanth, N., Human computer interaction through hand gesture recognition technology, International Journal of Scientific and Technology Research, 2020, 9(4), pp. 505-513

[10] Gayathri .S , Porkodi Venkatesh , PushpapriyaPremkumar on Voice Assistant for Visually Impaired in 2019, https://ijesc.org/upload/40664e91149af2618afd09aaf1fca8f8.Voice%20Assistant%20for%20Visual ly%20Impaired%20(2).pdf

[11] Speech Recognition Bhuvan Taneja, Jones C J, Rohan Tanwar HMRITM (GGSIP University), Delhi, India,2021

[12] Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning &Technology.

[13] J Kiran, N Parveen, Holistic Review of Software Testing and Challenges, International Journal Of Innovative Technology and Exploring Engineering (IJITEE) , Volume.8, Issue 7, Page No pp.1506- 1521, June 201.

[14] Sreeram, G., Pradeep, S., Rao, K.S., Raju, B.D., Nikhat, P. Moving ridge neuronal espionage network simulation for reticulum invasion sensing, International Journal of Pervasive Computing and Communications, 2020.

[15] Rayan spring, RyujiTabuchi, Assessing the practicality of using an automatic speech recognition tool to teach English pronunciation online,2021.

[16] Dash, Devidutta, Arun Agarwal, Kabita Agarwal, and Gourav Misra. "Post Catastrophe Fallouts and Challenges to Swim to Safety." Journal of Information Technology 3, no. 01 (2021): 12-17.

978-1-6654-0837-0/22/$31.00 ©2022 IEEE