

# **AUTO-NARRATED AUDIOBOOK FOR VISUALLY IMPAIRED AND UNEDUCATED**

## **A SOCIALLY RELEVANT PROJECT REPORT**

Submitted in partial fulfillment of the requirements for the  
award of the degree of  
Bachelor of Technology

in

**COMPUTER SCIENCE AND ENGINEERING**

**BY**

**BATCH- 7B**

**K. S.V RAVI SASTRY**

**(20331A0588)**

**K. SARATH KUMAR**

**(20331A0584)**

**N.V.S.R SRUJAN**

**(20331A05C6)**

**I.DEENA**

**(20331A0573)**

Under The Supervision

**B. SUJATA**

**(Assistant Professor)**



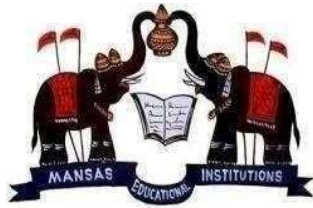
**DEPARTMENT OF COMPUTER SCIENCE AND  
ENGINEERING**

***MVGR COLLEGE OF ENGINEERING (Autonomous)***

***VIZIANAGARAM-535005, AP (INDIA)***

**(Accredited by NBA, NAAC, and Permanently Affiliated to Jawaharlal Nehru  
Technological University Kakinada)**

## CERTIFICATE



This is to certify that the project report entitled “**Auto-narrated Audiobook for visually impaired and uneducated.**”being submitted by

**K.S.V RAVI SASTRY, N.V.S.R SRUJAN, K.SARATH KUMAR, I.DEENA** bearing registered numbers **20331A0588, 20331A05C6, 20331A0584, 20331A0573** respectively, in partial fulfillment for the award of the degree of “**Bachelor of Technology**” in **Computer Science and Engineering** is a record of Bonafede work done by them under my supervision during the academic year 2022-2023.

**Dr P. Ravi Kiran Varma**  
**Head of the department**  
**Dept. Of CSE**

**B. SUJATA**  
**Assistant Professor**  
**Dept. Of CSE**

## **ABSTRACT**

The Internet is a rich source of knowledge and information without which very few works can be done. It is the most important mode of communication and connection. Visually impaired people are dependent solely on Braille books & audio recordings provided by NGOs. Owing to many constraints in the above approaches blind people cannot have a book of their choice. Not only blind people, but the illiterate people who can understand the language, but cannot read printed books are suffering because of a lack of resources.

The presented work will provide them an opportunity to have an audiobook of their choice in English or the local languages of any printed book having English, local languages like Telugu Printed text from textbooks having English will be taken as input in the form of an image which will be converted into plain editable text with the help of Optical Character Recognition (OCR). This plain text will be then fed to a Text to Speech (TTS) converter which will generate the audio output file in English and Telugu.

## TABLE OF CONTENTS

<b>CHAPTER</b>	<b>Pg-No</b>
1. INTRODUCTION	5
2. REQUIREMENTS	6
3. IMPLEMENTATION	7
4. OUTPUTS	10
5. CONCLUSION	18
6. PROJECT OUTCOME	19
REFERENCES	20

# 1. INTRODUCTION

Visually impaired people cannot learn as we people normally do. The only way to learn them is through Braille script which is a tactile writing system used by the blind & visually impaired. An A4 size page of a normal textbook is equivalent to 4 pages in Braille script. Due to this fact, Braille script books are large, heavy, take more time to read, and are not easily available as compared to normal textbooks.

Therefore, Braille books are not produced for higher studies where the book size is considerably large. Owing to these constraints many social organizations voluntarily come forward and convert normal textbooks into audiobooks by manually recording which requires a significant amount of manpower & time investment. So, the proposed converter is intended to do the same task which will eliminate the cumbersome procedure of printing large-size Braille script books & manual recording of normal textbooks.

Not only visually impaired people but also many illiterate people who know and understand the language, but they cannot read from books. Many illiterate people are interested in reading the content in books unfortunately they do not know how to pronounce words in those books. As they cannot directly read the content from books, we prepared a converter. This converter will open new doors of Knowledge Sea for the visually impaired and illiterates as they can listen to any book of their choice in English & Telugu.

## GOALS

- ✓ To develop audiobooks for visually impaired and uneducated persons.
- ✓ Easing the process of gaining knowledge from books for the visually impaired.
- ✓ To develop a handy system for people in need with ease of accessibility.
- ✓ To extract text from images using OCR accurately.
- ✓ To translate audio into local languages for the uneducated.

## **2. REQUIREMENTS**

### **SOFTWARE:**

1. Python
2. PyCharm
3. VSCode
4. Pyttsx3
5. Tesseract OCR
6. Other Modules Needed:
  - a. NumPy
  - b. Pdf2image

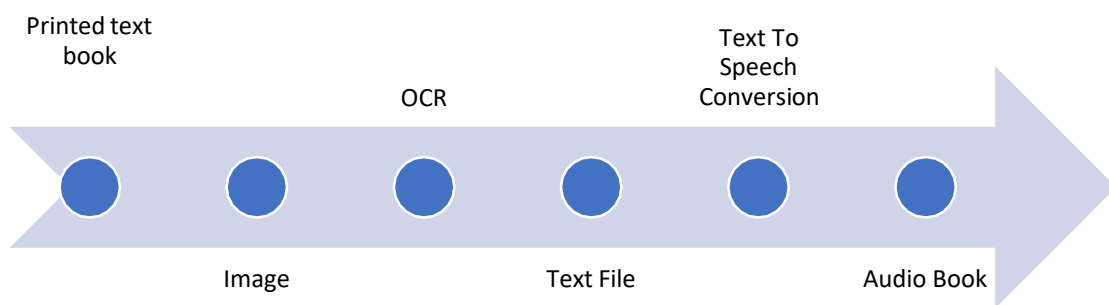
### **HARDWARE:**

1. PROCESSOR: Intel Core (I3 processor)
2. RAM: 4GB
3. HARD DISK: 500GB

### 3. IMPLEMENTATION

We can convert a printed book to audio files using **OCR** and **TTS**.

Here the image of printed a book is converted to text using **OCR (optical character recognition)** And then the text is converted to audio using **TTS (Text to speech)**. The following flow chart represents how a printed book can be converted to audiobook:



#### ➤ **OPTICAL CHARACTER RECOGNITION:**

Optical character recognition (OCR) uses a scanner to process the physical form of a document. Once all pages are copied, OCR software converts the document into a two-color or black-and-white version. The scanned-in image or bitmap is analyzed for light and dark areas, and the dark areas are identified as characters that need to be recognized, while light areas are identified as background.

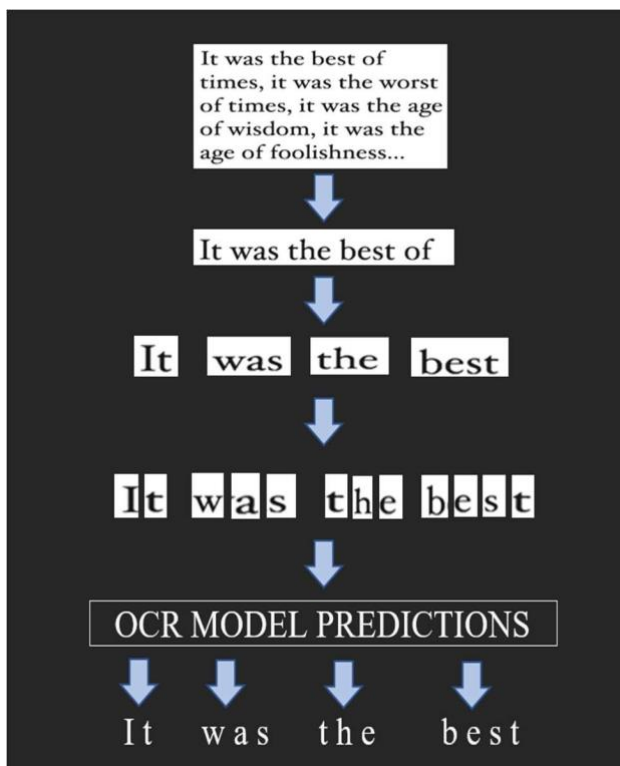
The dark areas are then processed to find alphabetic letters or numeric digits. This stage typically involves targeting one character, word, or block of text at a time. Characters are then identified using one of two algorithms pattern recognition or feature recognition.

Pattern recognition is used when the OCR program is fed examples of text in various fonts and formats to compare and recognize characters in the scanned document or image file. Feature detection occurs when the OCR applies rules regarding the features of a specific letter or number to recognize characters in the scanned document.

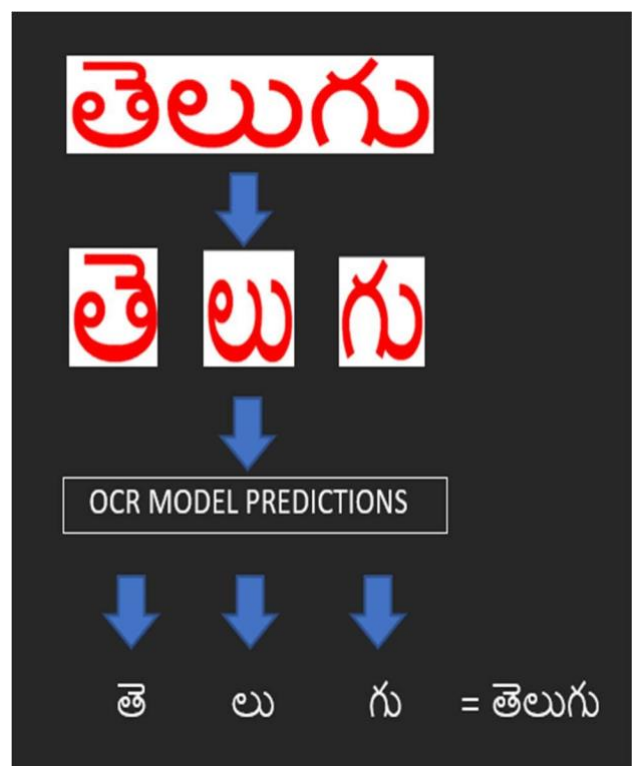
Features include the number of angled lines, crossed lines, or curves in a character. For example, the capital letter “A” is stored as two diagonal lines that meet with a horizontal line across the middle. When a character is identified, it is converted into an ASCII code (American Standard Code for Information Interchange) that computer systems use to handle further manipulations.

An OCR program also analyses the structure of a document image. It divides the page into elements such as blocks of text, tables, or images. The lines are divided into words and then into characters. Once the characters have been singled out, the program compares them with a set of pattern images. After processing all likely matches, the program presents you with the recognized text.

For example, below we have an image, now this is converted to text as



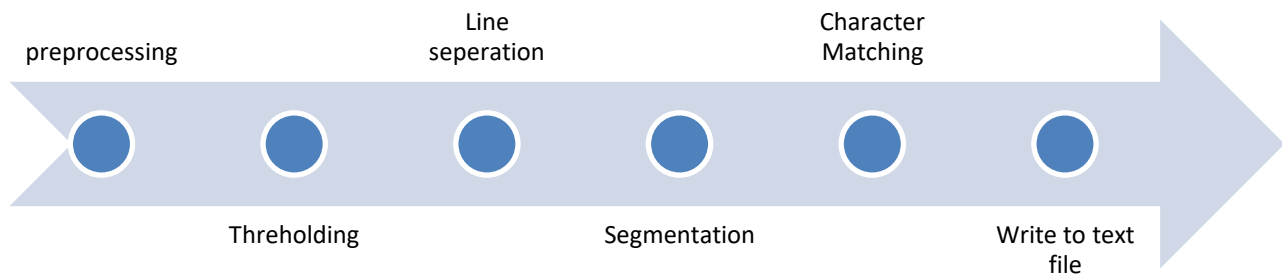
1) OCR converting English text  
English text.



2) OCR converting Telugu text image to  
image to Telugu text.



## ➤ OCR FLOW CHART



### Process:

- First it differentiates **word contours** associated with images using **OpenCV contours** and **image cropping libraries**.
- Then it differentiates the **letter contours** associated with **word contour images** using **OpenCV contour dilation** and **image cropping**.
- Now it pre-processes the letter images according to the trained OCR input using **KerasFramework in detecting** and **PIL library in image processing**.
- Finally, it converts the predictions associated OCR model to text using the **PIL library** and **python**.

## ➤ TEXT TO SPEECH(TTS):

### • English TTS Algorithm:

Text-to-Speech is implemented by using the pyttsx3 library and the audio file is played using pygame. Pyttsx3 is a text-to-speech conversion library in Python. The pyttsx3 module supports two voices first is female and the second is male which is provided by “sapi5” for windows. It supports three TTS engines:

- *sapi5* – SAPI5 on Windows
- *nsss* – NSSpeechSynthesizer on Mac OS X
- *espeak* – eSpeak on every other platform

The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. It is shipped as part of the Windows OS itself. Therefore, it can be easily used on any genuine Microsoft Windows without having any copyright infringement issues. The main motto behind using Microsoft SAPI is its excellent quality. This saves the hustles of building a completely new TTS system from scratch.

### ❖ Alternative Options:

Google Text-to-Speech API:

It works wonderfully but it had issues with it and it does not work offline.

### ➤ **Telugu TTS Algorithm:**

A speech synthesis system is divided into two categories. The front part accepts the text information and gives a symbolic representation as output. The second part takes the output from the front part as input and converts that into a synthesized waveform and gives it as output. The output of a speech synthesis system usually refers to how much the output sounds like the speech of a real person.

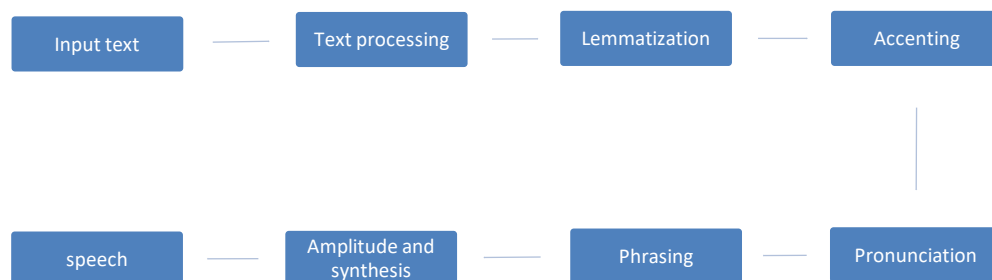
#### **I. Analysis of Text**

It is the first stage of the conversion process. Here in this phase, it identifies the words from the given text. Here first it identifies the tokens by considering whitespaces, abbreviations, hyphens, and punctuation marks. Once the first step is completed then, normalization of text must be applied.

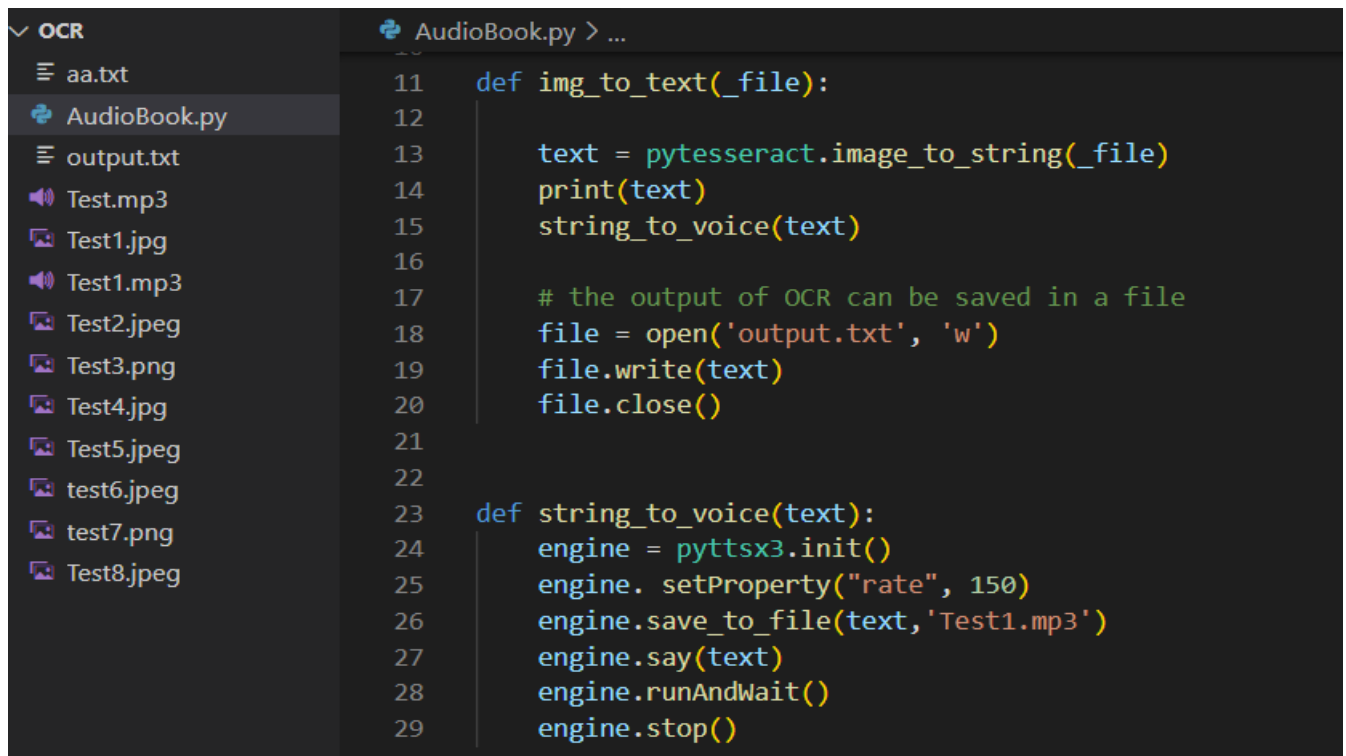
#### **II. Linguistic Analysis**

After the text analysis phase, the second phase is linguistic analysis. The work of this phase is parsing, semantic analysis, and word pronunciation assignment.

#### **❖ Process:**



## 4. OUTPUTS



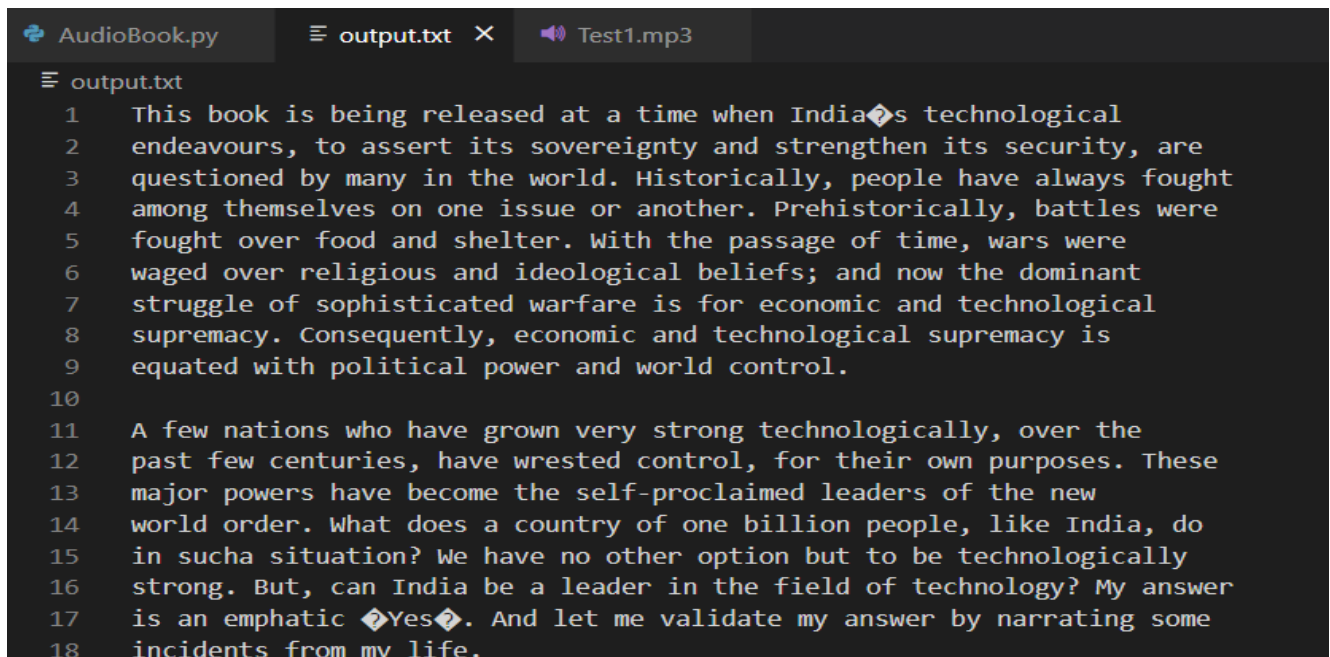
```
11 def img_to_text(_file):
12
13     text = pytesseract.image_to_string(_file)
14     print(text)
15     string_to_voice(text)
16
17     # the output of OCR can be saved in a file
18     file = open('output.txt', 'w')
19     file.write(text)
20     file.close()
21
22
23 def string_to_voice(text):
24     engine = pyttsx3.init()
25     engine.setProperty("rate", 150)
26     engine.save_to_file(text, 'Test1.mp3')
27     engine.say(text)
28     engine.runAndWait()
29     engine.stop()
```

- The above code snippet converts printed book to text and from text, it converts into audio
- **py./AudioBook.py** - run this command in the terminal to directly convert into audio.
- Test case-1

This book is being released at a time when India's technological endeavours, to assert its sovereignty and strengthen its security, are questioned by many in the world. Historically, people have always fought among themselves on one issue or another. Prehistorically, battles were fought over food and shelter. With the passage of time, wars were waged over religious and ideological beliefs; and now the dominant struggle of sophisticated warfare is for economic and technological supremacy. Consequently, economic and technological supremacy is equated with political power and world control.

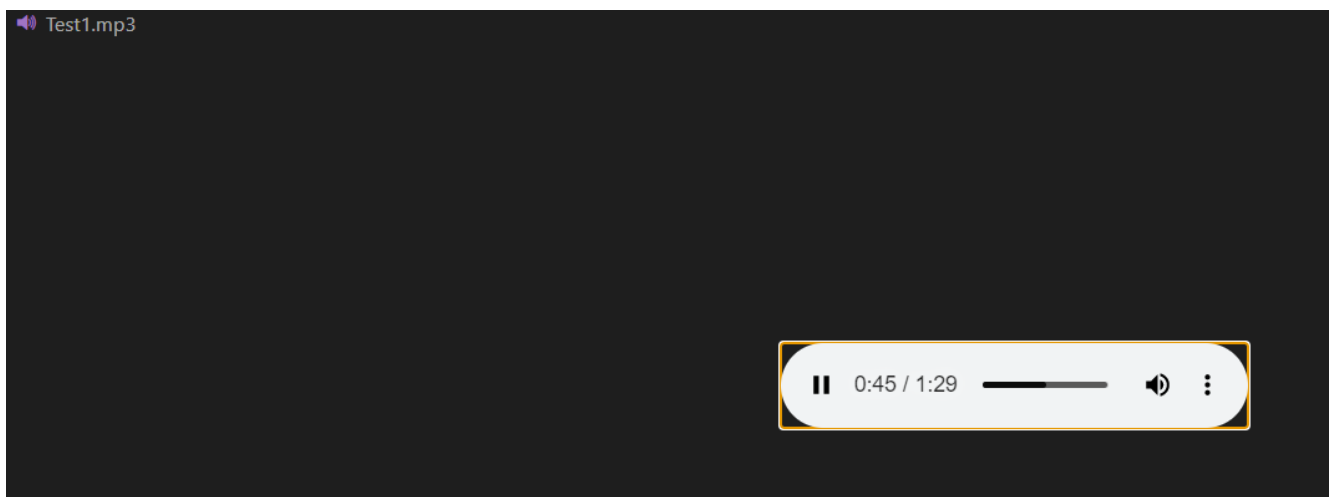
A few nations who have grown very strong technologically, over the past few centuries, have wrested control, for their own purposes. These major powers have become the self-proclaimed leaders of the new world order. What does a country of one billion people, like India, do in such a situation? We have no other option but to be technologically strong. But, can India be a leader in the field of technology? My answer is an emphatic 'Yes'. And let me validate my answer by narrating some incidents from my life.

- The above image is taken as input.



```
AudioBook.py  output.txt  Test1.mp3
output.txt
1  This book is being released at a time when India's technological
2  endeavours, to assert its sovereignty and strengthen its security, are
3  questioned by many in the world. Historically, people have always fought
4  among themselves on one issue or another. Prehistorically, battles were
5  fought over food and shelter. With the passage of time, wars were
6  waged over religious and ideological beliefs; and now the dominant
7  struggle of sophisticated warfare is for economic and technological
8  supremacy. Consequently, economic and technological supremacy is
9  equated with political power and world control.
10
11 A few nations who have grown very strong technologically, over the
12 past few centuries, have wrested control, for their own purposes. These
13 major powers have become the self-proclaimed leaders of the new
14 world order. What does a country of one billion people, like India, do
15 in such a situation? We have no other option but to be technologically
16 strong. But, can India be a leader in the field of technology? My answer
17 is an emphatic Yes. And let me validate my answer by narrating some
18 incidents from my life.
```

- The above image shows that OCR is performed and extracted text is saved as a text document.



- The above image shows that the text is converted to audio and saved as an Audio file (mp3).

➤ Test Case -2:

Operating systems are an essential part of any computer system. Similarly, a course on operating systems is an essential part of any computer science education. This field is undergoing rapid change, as computers are now prevalent in virtually every arena of day-to-day life—from embedded devices in automobiles through the most sophisticated planning tools for governments and multinational firms. Yet the fundamental concepts remain fairly clear, and it is on these that we base this book.

We wrote this book as a text for an introductory course in operating systems at the junior or senior undergraduate level or at the first-year graduate level. We hope that practitioners will also find it useful. It provides a clear description of the *concepts* that underlie operating systems. As prerequisites, we assume that the reader is familiar with basic data structures, computer organization, and a high-level language, such as C or Java. The hardware topics required for an understanding of operating systems are covered in Chapter 1. In that chapter, we also include an overview of the fundamental data structures that are prevalent in most operating systems. For code examples, we use predominantly C, with some Java, but the reader can still understand the algorithms without a thorough knowledge of these languages.

➤ Now we have given the above image as input and the results are as follows

```
≡ output.txt
1  Operating systems are an essential part of any computer system. Similarly,
2  a course on operating systems is an essential part of any computer science
3  education. This field is undergoing rapid change, as computers are now
4  prevalent in virtually every arena of day-to-day life from embedded devices
5  in automobiles through the most sophisticated planning tools for governments
6  and multinational firms. Yet the fundamental concepts remain fairly clear, and
7  it is on these that we base this book.
8
9  We wrote this book as a text for an introductory course in operating systems
10 at the junior or senior undergraduate level or at the first-year graduate level. We
11 hope that practitioners will also find it useful. It provides a clear description
12 of the concepts that underlie operating systems. As prerequisites, we assume
13 that the reader is familiar with basic data structures, computer organization,
14 and a high-level language, such as C or Java. The hardware topics required
15 for an understanding of operating systems are covered in Chapter 1. In that
16 chapter, we also include an overview of the fundamental data structures that are
17 prevalent in most operating systems. For code examples, we use predominantly
18 C, with some Java, but the reader can still understand the algorithms without
19 a thorough knowledge of these languages
```

So, we can take any image containing text as input and the respective audio will be generated.

There will be no output if the uploaded image does not contain any text.

➤ **Test-case 3**



- Here we have uploaded an image that doesn't have any text and the output is as follows

```
PS C:\Users\RAVI\OneDrive\Desktop\OCR> py ./AudioBook.py  
enter relative path of file:c:\Users\RAVI\OneDrive\Desktop\OCR\Test1.jpg  
PS C:\Users\RAVI\OneDrive\Desktop\OCR> █
```

- **As the image does not contain any text the output file is nil.**

**Conditions while the user giving the input:**

- ✓ Image should be straight and clear.
- ✓ The image must contain the text.

**Results:**

### Playing the Audio File:

In python to open .wav files, Pygame is used. To play music/audio files in pygame, pygame.mixer module is used (pygame module for loading and playing sounds). This module contains classes for loading Sound objects and controlling playback.

The whole work can be converted into C language code. So that it can be easily dumped into any DSP processor which will serve as a stand-alone device for converting a printed book into an audiobook. In such stand-alone devices, either a scanner or High-Definition camera can be interfaced with a DSP processor which will acquire images of printed book pages.





## **5. Conclusion**

A printed book-to-audiobook converter has been successfully implemented for English and Telugu. A new algorithm for English & Telugu has been implemented successfully which is almost 100

% Accurate. Such a system will greatly help illiterate and visually impaired people and people working along with them. OCR systems can't have a perfect 100 percent accuracy. So, different algorithms which can improve the accuracy of OCR for all three scripts can be employed in the future. In the presented work, OCR for Telugu is implemented. In the future, more Indian regional languages can be included.

Improvements in the output voice can be made by changing the prosody rules & Indian accent database can be used which will make the English TTS system sound more natural having an Indian accent. In the future, both these TTS systems for English & Telugu languages can be made standalone systems that are independent of any third-party TTS platform

## **6. PROJECT OUTCOME**

- 1.** From this project, we have learned how to improve our presentation skills, how to make an effective presentation, and even learned how to present our thoughts and implementation through this.
- 2.** From this project, team management skills improvised a lot, each one of us enthusiastically took each part of theirs and worked on it. We learned how coordination can help make a bigger problem into a smaller problem and each person was the leader of the team in their way.
- 3.** From this project, communication skills have got better, and how to represent ourselves in front of a crowd improved, communicating among ourselves also helped in making our work easier.
- 4.** From this project, we have learned a new software called Flask framework which is used to implement web applications using a python programming language.
- 5.** Finally, this project helped us in learning new skills, helped us to think out of the box, improve our thoughts towards society and how to take up a problem from our daily life and put some work on it to give a minimal solution that would positively benefit the society.

## References

- [1] Bernard Conan Pobiak - ADJUSTABLE ACCESS ELECTRONIC BOOKS, The Reader Project of Washington, D.C., 1992 IEEE
- [2] <http://finereader.abbyy.com/>
- [3] [http://www.irislink.com/c2-2115-189/Readiris-14--OCR-Software--Scan--Convert--Manage-your-Documents-.aspx /](http://www.irislink.com/c2-2115-189/Readiris-14--OCR-Software--Scan--Convert--Manage-your-Documents-.aspx/)
- [4] <http://www.naturalreaders.com/>
- [5] <http://www.naturalvoices.att.com/>
- [6] <http://www.knfbreader.com/>
- [7] OCR with OpenCV, Tesseract, and Python - Intro to OCR by Adrian Rose Brock
- [8] Speech and Language Processing by James H. Martin | Dan Jurafsky