# Sound Script

## Abhishek V S[1], Gagan Raj R P[2], Roopa T[3]

Sri Siddhartha Institute Of Technology (A Constituent College of Sri Siddhartha Academy of HigherEducation)

Maralur, Tumkur-572105[1,2]

Guide, Assistant Professor, ISE Dept, SSIT, Tumakuru[3]

**Abstract:** Audiobooks are ideal for anyone who, like the majority of us, likes to listen rather than read. It is just not possible to purchase and store them in your home on a bookshelf. Audiobooks are also a good method to relax your eyes and take a break from the continual stimulation of digital devices. Others can take advantage of them to save time. For example, keep up with books while doing different tasks at same time. It has the potential to not only alleviate problems for millennials, but also to be a highly useful tool for visually impaired people. The ability to transform any material into an audiobook is a true gift to civilization. Our technology can be put to use in the development of such tools. Text-to-speech and other read-aloud programs are widely utilized to assist students in developing their reading comprehension abilities. A PDF to audio system is a screen reader program that has been designed and developed specifically for the purpose of effective audio communication. International Organization for Standardization (ISO) established PDFs as an open standard document format for the aim of displaying and transmitting information securely (ISO). One of the most convenient formats for electronic communication and information transmission is the document format. It's critical if we want to improve accessibility for screen readers by including audio into our material. Among the features PDF documents provide are text links and buttons as well as audio and video files. Many languages may be supported by the PDF to audio technology, which will allow users to hear text being read aloud (spoken).

**Keywords:** Audiobooks, Communication, Text-to-Speech

## I. INTRODUCTION

Audiobooks are recorded versions of a book's text that we listen to rather than read. They can be literal word-for-word transcriptions of novels or shortened versions that exclude unnecessary language. For instance, short stories or novels can be converted into audiobooks to allow scholars to rapidly obtain a description of the work. Text-to-speech (TTS) related audio reading tools are widely accepted across the world, implemented to encourage students' listening skills. They help children listen to short stories at bedtime, as parents can now convert any story or light novel into an audiobook.

Audiobooks have gained significant popularity due to their convenience and versatility. They allow individuals to consume literature and educational content while performing other tasks, such as driving, exercising, or doing household chores. This multitasking capability makes audiobooks an essential tool in today's fast-paced world, where time management is crucial. Furthermore, audiobooks cater to various learning styles, benefiting auditory learners who may retain information better when heard rather than read.

The integration of text-to-speech (TTS) technology has further enhanced the accessibility and production of audiobooks. TTS systems convert written text into spoken words using sophisticated algorithms that mimic natural speech patterns. This technology is not only valuable for creating audiobooks but also for assisting individuals with visual impairments, dyslexia, and other reading difficulties.

The ability to listen to books and documents empowers these individuals, providing them with greater access to information and knowledge. PDFs, or Portable Document Formats, are a widely used digital document format maintained by the International Organization for Standardization (ISO). They are preferred for electronic communication due to their ability to preserve the formatting and layout of the original document, regardless of the device or software used to view them. PDFs can contain links, buttons, forms, and multimedia elements, making them highly versatile.

The conversion of PDFs to audiobooks is a significant advancement, combining the robustness of PDFs with the convenience of audiobooks. This process involves using Optical Character Recognition (OCR) technology to extract text from images within PDF files. OCR is a widely used method for recognizing text embedded in images, such as scanned documents and photos.

Once the text is extracted, TTS systems can convert it into spoken words, creating an audiobook. This technology is particularly useful for creating accessible content for the visually impaired, as well as for people who prefer auditory learning or need to consume information while multitasking.

The benefits of converting PDFs to audiobooks extend beyond accessibility and convenience. This technology can significantly enhance the educational experience by providing students with an alternative way to engage with their study materials. For example, students can listen to textbooks and academic papers during their commute, making productive use of their travel time. Additionally, the use of audiobooks can improve language skills, such as pronunciation, intonation, and listening comprehension.

The project described in this paper aims to develop a PDF to audiobook converter using Python libraries such as PDFMiner and Google Text-to-Speech (gTTS), and employing Tinker for the graphical user interface (GUI). PDFMiner is a robust tool for extracting text from PDF files, while gTTS provides a simple and effective way to convert text into speech. The combination of these technologies allows for the creation of audiobooks from PDF documents, making information more accessible and convenient to consume.

## II. RELATED WORK

R. M. K. Sinha's Ph.D. thesis introduced a groundbreaking Syntactic Pattern Analysis System for Devanagari script recognition, revolutionizing the field of character recognition and laying the foundation for further research in the domain of Indian language processing.

The collaboration between R. M. K. Sinha and H. Mahabala resulted in a seminal paper published in IEEE Transactions on Systems, Man, and Cybernetics in August 1979. Their work significantly advanced the field of machine recognition of Devanagari script by proposing innovative methodologies and algorithms.

Bernard Conan Pobiak's research on "Adjustable Access Electronic Books" within The Reader Project of Washington, D.C. in 1992, sponsored by IEEE, aimed to enhance accessibility and usability of electronic books, pioneering efforts in adaptive reading technologies.

S. Palit and B. B. Chaudhuri's feature-based scheme for machine recognition of printed Devanagari script, presented in their paper published in 1995, contributed significantly to pattern recognition, image processing, and computer vision research, particularly in the context of Indian language text analysis.

The programming system developed by P. V. S. Rao and R. B. Thosar in 1974 enabled groundbreaking studies in speech synthesis, facilitating advancements in artificial intelligence and human-computer interaction by providing a robust framework for experimentation and development in this field.

P. Bhaskararao and S. Mathew's phonemic transcription rules for text-to-speech synthesis of Hindi, published in 1992 as part of the proceedings of "Computer Processing of Asian Languages-2," played a crucial role in the development of natural and intelligible Hindi speech synthesis systems, contributing to the accessibility of information in Indian languages.

T. S. Agrawal's discussion on synthesizing Hindi speech using Klsyn and Hlsyn at the 2003 Workshop on Spoken Language Processing at TIFR, Mumbai, highlighted advancements in speech synthesis technology tailored for Indian languages, paving the way for more natural and expressive voice interfaces.

## III. SYSTEM DESIGN

The translation of PDF text to audio signals is in high demand due to PDF's extensive use as the world's most popular document format. In addition to instructional purposes, they may be used for a variety of other tasks, such as automobille steering, train station announcements, telecommunications response services, and reading e-mail. PDF files cannot be viewed or read by those with vision damages, which is a big disadvantage. This study examines the problems of turning PDF text into spoken word.

One of the ways to make PDF-based text seem more accurate when played over an audio system is to make synthetic speech sound more natural. Because PDF is the most widely used document format on the planet, there is a demand for PDF text to be converted to audio signals.

In addition to instructional objectives, they may be used for a variety of other tasks, such as automobile navigation, train station notifications, telecommunications reply services, and reading e-mail.

PDF files are not accessible to those with poor vision, which is a severe drawback. In this research, the difficulties of translating PDF text to voice are examined. In PDF-based writing, one challenge is to make the audio output of artificial speech sound more truthful.

You'll notice when you look at our Audiobook converter that it converts PDF text (or pictures) into voice by making slight alterations inside the same PDF document.
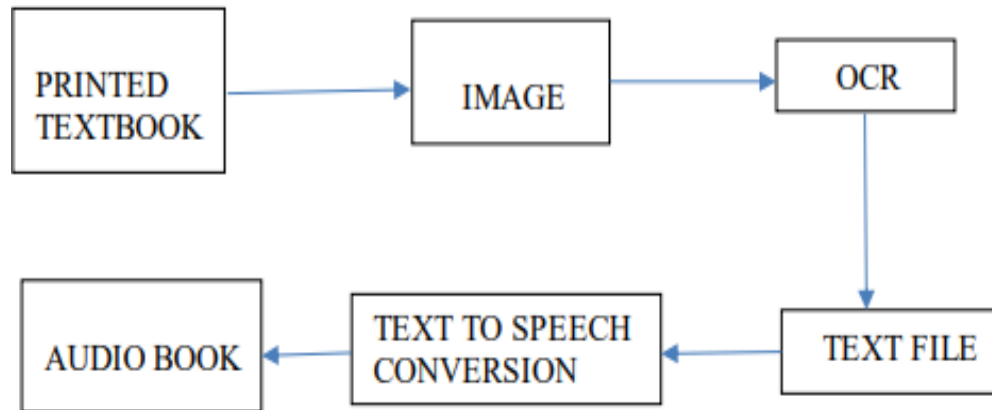


Fig1. System Architecture

Fig.1 describes the flow of converting a.pdf file into a.mp3 file as a core functionality. Two libraries, pyttsx3 and pdf miner, were used to accomplish this.

One of the most popular programs for mining data from PDF files is PDFMiner. It's the only PDF application that concentrates only on text data mining and analysis. If you use PDFMiner, you'll be able to see where text appears on a page, along with font and line information.

A Python library for converting text to voice is available in the form of pyttsx3. If you don't have access to the Internet this library can still be used. Extraction of data, page-by-page document splitting, merging, and more are all possible The PyPDF2, pdfminer, and pyttsx3 modules will be imported.

Using pdfminer, the text is then mined from the pdf file and saved as a text file. It will be transformed to an mp3 file using Google's text-to-speech module, gtts. Windows player will be used to play the output.

## IV.    CONCLUSION

When it comes to reading basic PDF text files, this strategy performs remarkably well. When a user selects a PDF, they should be able to convert it to audio, and then read the text to see what was read. Contribution by pupils who have trouble reading should be permitted.

That this study was a success is no surprise, given how widely audiobooks are used in literateness and library programs around the country. Students who struggle with reading might benefit from using audio books as a reading strategy, according to these findings. Research on the use of audiobooks with reading understanding younger and older than those who participated in the  study would be of interest to us, as would discussion of the use of audio books by English Language Learners.

Code does not currently have a stop option, but I intend to design one and include more advanced machine learning techniques into the audiobook. The ability to add features that recognize a user's voice and perform the function exactly how the user wants may be attained with machine learning. Those with incapacities, such as the blind and the disabled, will be the primary beneficiaries of this new feature.

## REFERENCES

[1]. R. M. K. Sinha, "A Syntactic pattern analysis system and its application to Devanagari script recognition," Ph.D.Thesis, Dept. Elect. Eng., Indian Institute of Technology, Kanpur, India, 1973.

[2]. R. M. K. Sinha and H. Mahabala, "Machine recognition of Devanagari script" IEEE Trans. Syst. ManCybern., vol. 9,no. 8, pp. 435–441, Aug.1979.

[3]. Bernard Conan Pobiak - ADJUSTABLE ACCESS ELECTRONIC BOOKS, The Reader Project ofWashington, D.C.,1992 IEEE

[4]. S. Palit and B. B. Chaudhuri, "A feature-based scheme for the machine recognition of printed Devanagari script" in Pattern Recognition, Image Processing and Computer Vision, P. P. Das and B. N. Chatterjee, Eds. New Delhi, India: Narosa Publishing House, 1995, pp. 163–168.

[5]. P. V. S. Rao and R. B. Thosar, "A programming system for studies in speech synthesis," IEEE Trans. Acoust.,Speech and Signal Processing, vol. 22, no. 3, pp. 217–225, 1974.

[6]. P. Bhaskararao and S. Mathew, "Phonemic transcription rules for text-to-speech synthesis of Hindi," in In: Sinha,R.M.K. (ed) Proc. Computer Processing of Asian Languages-2. New Delhi: Tata McGraw Hill., 1992.

[7]. S. S. Agrawal, "Synthesizing Hindi speech using Klsyn and Hlsyn for natural sounding," in Workshop on Spoken Language Processing, TIFR, Mumbai, 2003