TEXT TO SPEECH FOR VISUALLY IMPAIRED PEOPLE

Main Project Report

Submitted by

MOHAMMED SHEKEEB.K

Reg No: FIT22MCA-2090

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

Master of Computer Applications

Of

A P J Abdul Kalam Technological University

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®

ANGAMALY-683577, ERNAKULAM(DIST)

APRIL 2024

DECLARATION

I, MOHAMMED SHEKEEB.K hereby declare that the report of this project work, submitted to the Department of Computer Applications, Federal Institute of Science and Technology (FISAT), Angamaly in partial fulfillment of the award of the degree of Master of Computer Application is an authentic record of my original work.

The report has not been submitted for the award of any degree of this university or any other university.

Date: 23/04/24

Place: Angamaly

Shekeeb

Mohammed Shekeeb K

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®
(An ISO 9001:2015 Certified, NAAC ('A+') Accredited Institution with NBA Accredited Programmes)
Approved by AICTE - Affiliated to APJ Abdul Kalam Technological University, Kerala)
Owned and Managed by Federal Bank Officers' Association Educational Society
HORMIS NAGAR, MOOKKANNOOR P.O., ANGAMALY - 683 577, ERNAKULAM, KERALA, INDIA.
Tel : (0) 0484 - 2725272 E-Mail : mail@fisat.ac.in Website : www.fisat.ac.in

Focus on Excellence

23rd April 2024

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr/Ms. MOHAMMED SHEKEEB.K (Reg. No. FIT22MCA-2090) has successfully completed his/her Main Project with the title "Text to speech for visually impaired people" under the guidance of MS. ROSE MARY MATHEW in the Department of Computer Applications, FISAT, during the period from 05.02.2024 to 23.04.2024.

Dr. DEEPA MARY MATHEWS
HEAD OF THE DEPARTMENT

FEDERAL INSTITUTE OF SCIENCE & TECHNOLOGY (FISAT) ANGAMALY
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS

FEDERAL INSTITUTE OF SCIENCE ANDTECHNOLOGY

(FISAT)®

ANGAMALY, ERNAKULAM-683577

DEPARTMENT OF COMPUTER APPLICATIONS

FISAT

Focus on Excellence

CERTIFICATE

*This is to certify that the project report titled "**TEXT TO SPEECH FOR VISUALLY IMPAIRED PEOPLE**" submitted by **MOHAMMED SHEKEEB.K [Reg No: FIT22MCA-2090]** towards partial fulfillment of the requirements for the award of the degree of Master of Computer Applications is a record of bonafide work carried out by him during the year 2024.*

Project Guide

Ms. Rose Mary Mathew

FEDERAL INSTITUTE OF SCIENCE & TECHNOLOGY (FISAT) ANGAMALY
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS

Head of the Department

Dr. Deepa Mary Mathews

Submitted for the viva-voice held on ………………………

Examiner:

ACKNOWLEDGEMENT

Gratitude is a sentiment that transcends words and surpasses silence. In completing this project, I sought and received invaluable guidance, assistance, and cooperation from several individuals, all of which flowed abundantly with the grace of God.

I extend my profound gratitude to **Dr. Jacob Thomas V**, Principal for granting me access to the college facilities. A special thanks to **Dr. Deepa Mary Mathews**, Head of the Department, MCA, whose expertise, and guidance were invaluable throughout the journey.

Throughout the project duration, I received unwavering support from **Ms. Rose Mary Mathew**, my project guide, and **Dr. Shahna K U**, my scrum master. I express my heartfelt gratitude to them with immense pleasure. Their dedication and assistance were indispensable in navigating through the complexities of the project.

I would like to convey my sincere thanks to all the faculty members in our department for their consistent encouragement and enduring support. The guidance provided by the lab faculty members is also acknowledged with boundless gratitude.

Lastly, my deepest thanks go to my parents, friends, and well-wishers who contributed in various ways to the preparation of my project. Above all, I am grateful to God for orchestrating everything to perfection.

ABSTRACT

Text-to-Speech (TTS) technology has emerged as a pivotal tool in the realm of artificial intelligence and human-computer interaction. This paper presents a comprehensive overview of the advancements in TTS technology, encompassing its historical development, underlying principles, and contemporary applications. Beginning with a historical perspective, we trace the evolution of TTS from its rudimentary beginnings to its current state-of-the-art capabilities. We delve into the fundamental components of TTS systems, elucidating the intricate processes involved in converting textual input into natural-sounding speech output. Moreover, we explore the various techniques and methodologies employed in TTS synthesis, including concatenative synthesis, formant synthesis, and statistical parametric synthesis.

Finally, we address the existing challenges and future directions in TTS research, including improving the robustness of TTS systems across different languages and dialects, enhancing the expressiveness and emotional realism of synthesized speech, and mitigating ethical concerns related to synthetic voice manipulation. We conclude with reflections on the transformative potential of TTS technology in reshaping human-computer interaction and advancing the frontiers of artificial intelligence.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In a world dominated by visual information, individuals with visual impairments often face substantial barriers in accessing written content, limiting their educational and professional opportunities as well as their social inclusion. Traditional methods like braille have been valuable but have not fully addressed the dynamic and diverse range of digital content available today. Thus, there is a pressing need for innovative solutions to enhance accessibility for the visually impaired community. This project aims to tackle this challenge head-on by developing a state-of-the-art text-to-speech (TTS) system tailored specifically to the needs and preferences of visually impaired individuals.

Introducing a text-to-speech (TTS) project tailored for visually impaired individuals represents a groundbreaking endeavor in accessibility technology. At its core, this initiative aims to bridge the gap between the digital world and those with visual impairments, offering them greater autonomy and access to information. Through the innovative integration of advanced algorithms and speech synthesis techniques, this project endeavors to convert written text into spoken words, thereby enabling visually impaired individuals to perceive and engage with textual content effortlessly.

One of the key features of this TTS project is its versatility and adaptability across various digital platforms and content formats. Whether it's accessing websites, books, documents, or emails, visually impaired individuals can rely on the TTS system to convert written text into spoken words in real-time. Furthermore, the project team remains committed to enhancing the compatibility and interoperability of the work of TTS system with assistive technologies and mainstream devices, ensuring seamless integration into the daily lives of visually impaired individuals.

Beyond its immediate utility, the text-to-speech project holds the potential to catalyze broader societal change by promoting digital inclusion and accessibility. By advocating for the widespread adoption of inclusive design principles and accessibility standards, the project seeks to foster a more inclusive digital landscape where individuals of all abilities can fully participate and contribute. Ultimately, through its relentless pursuit of innovation and inclusivity, the text-to-speech project endeavors to empower visually impaired individuals to navigate the digital world with confidence and independence.

Text-to-speech technology is a form of assistive technology that converts written text into spoken words. It utilizes natural language processing algorithms and synthesized speech to vocalize electronic text, enabling individuals with visual impairments to access written content through auditory means. TTS systems can be integrated into various devices and platforms, including computers, smartphones, e-book readers, and navigation systems, offering versatility and flexibility in usage.

The applications of TTS technology for visually impaired individuals are manifold, encompassing education, employment, social interaction, and everyday tasks. In education, TTS facilitates access to textbooks, academic materials, and online resources, enabling visually impaired students to participate fully in learning activities. In the workplace, TTS assists in accessing documents, emails, and other digital content, promoting employment opportunities and productivity for visually impaired individuals. Moreover, TTS enhances social inclusion by enabling access to digital communication platforms, online news, and social media, fostering connections with peers and communities.

The benefits of TTS technology extend beyond accessibility to empowerment, independence, and autonomy for visually impaired individuals. By providing access to a wide range of information and resources, TTS empowers individuals to pursue education, employment, and personal interests on equal footing with their sighted peers. TTS also promotes independence by enabling individuals to navigate digital environments, perform tasks independently, and access information in real-time without reliance on sighted assistance. Furthermore, TTS fosters autonomy by allowing individuals to control the pace, volume, and customization of speech output according to their preferences and needs.

Despite its significant benefits, TTS technology faces several challenges and limitations in meeting the diverse needs of visually impaired individuals. These include limitations in accuracy, naturalness, and language support, as well as barriers related to affordability, availability, and compatibility with different platforms and devices. Addressing these challenges requires ongoing research and development efforts to improve the quality, functionality, and accessibility of TTS systems.

In conclusion, text-to-speech technology holds immense potential in enhancing accessibility and empowerment for visually impaired individuals. By providing access to written content through auditory means, TTS promotes inclusion, independence, and autonomy across various domains of life. However, addressing the challenges and advancing the capabilities of TTS technology is essential to maximize its benefits and ensure equitable access to information for all individuals, regardless of visual ability.

1.2 SCOPE

The scope of text-to-speech (TTS) technology encompasses a broad range of applications and potential impacts across various sectors.

1. Accessibility: One of the primary scopes of TTS technology is to enhance accessibility for individuals with visual impairments. By converting written text into spoken words, TTS systems enable visually impaired individuals to access a wide array of digital content, including websites, documents, e-books, and more. This accessibility extends beyond just the visually impaired to include individuals with dyslexia, learning disabilities, or those who prefer auditory learning.

2. Education: TTS has significant implications for the field of education. It can assist students in comprehending and engaging with educational materials, regardless of their reading ability. TTS can be integrated into e-learning platforms, textbooks, and educational software to provide audio support for students during reading and

studying sessions. It can also facilitate language learning by providing pronunciation assistance and enabling learners to hear correct word usage.

3. Assistive Technology: TTS technology plays a crucial role in the development of assistive technologies for individuals with disabilities. It can be integrated into devices such as smartphones, tablets, and computers to provide real-time spoken feedback and navigation assistance. TTS also enables the development of innovative solutions such as navigation apps for the visually impaired, voice-controlled interfaces, and communication aids for individuals with speech impairments.

4. Multimedia Content Accessibility: TTS expands the accessibility of multimedia content by providing audio descriptions for videos, graphics, and other visual elements. This ensures that individuals who are blind or visually impaired can fully engage with multimedia content, including online videos, presentations, and instructional materials.

5. Customer Service and Communication: In the realm of customer service and communication, TTS technology enables businesses and organizations to provide inclusive services to customers with visual impairments. Automated TTS systems can read aloud text-based information on websites, mobile apps, and interactive voice response (IVR) systems, ensuring that all customers can access important information.

CHAPTER 2

PROOF OF WORK

Visual impairment presents significant challenges for accessing printed material, with Braille often being the primary alternative. However, limitations exist when printed material is not available in Braille, and existing electronic solutions can be prohibitively expensive. To address this, we propose an affordable mobile application designed for visually impaired individuals. Leveraging image-to-text conversion through Optical Character Recognition (OCR) and text-to-speech conversion (TTS) frameworks, the application allows users to capture images of printed material using a mobile camera and convert them into speech format. This paper outlines the development and functionality of the application, highlighting its user-friendly design and the inclusion of alert sounds for enhanced usability.

Visual impairment affects approximately 2.3 billion people globally, impeding their ability to access textual information. While Braille is commonly used, its limitations and the high cost of electronic alternatives underscore the need for accessible and affordable solutions. This paper presents a website aimed at bridging this gap by enabling visually impaired individuals to access contents of a book through text-to-speech conversion.

Visual impairment encompasses a range of conditions, with distance visual acuity serving as a key measure. Individuals with visual acuity worse than 3/60 face significant challenges in reading printed material. Braille has traditionally been relied upon for reading and writing, utilizing a system of raised dots to represent letters and characters. However, Braille has limitations, particularly when material is not available in this format.

The proposed website offers a practical solution by leveraging OCR and Text-Speech frameworks to convert pdf format of particular material into spoken text. Users can simply upload pdf format of articles, e-books, newspaper, and the application processes the text, which is then read aloud. The interface is designed to be intuitive and user-friendly, with audio provided to guide users through the process.

Key features of the mobile application include text conversion, and speech output. The OCR framework accurately identifies text from images, while the TTS framework ensures clear and natural speech synthesis. Alert sounds are incorporated to provide feedback and aid navigation within the application, enhancing overall usability for visually impaired users.

In conclusion, the mobile application represents a significant advancement in accessibility for visually impaired individuals by providing a cost-effective solution for accessing printed material. Through the integration of image-to-text and text-to-speech conversion technologies, users can independently engage with a diverse range of textual content. This not only promotes inclusivity but also enhances the overall quality of life for visually impaired individuals by granting them greater autonomy in accessing information.

Looking ahead, future iterations of the application could focus on further refining usability and expanding functionality to better align with the evolving needs of the visually impaired community. By continually improving the user experience and incorporating feedback from users, the application can remain a valuable tool in enhancing accessibility and empowering visually impaired individuals in their daily lives.

2.1 EXISTING SYSTEM

•Screen Readers: These software programs utilize TTS engines to read aloud the text displayed on a computer screen. They are crucial for visually impaired individuals to interact with digital content.

•Standalone TTS Applications: Mobile apps and dedicated devices can convert digital text files (eBooks, documents) into speech for on-the-go information access.

Current Text-to-Speech (TTS) systems for visually impaired people typically rely on two key technologies:

- Optical Character Recognition (OCR)
- Text-to-Speech Engines

1. Optical Character Recognition (OCR): This software extracts text from physical documents or digital images. Common OCR tools like Tesseract can convert scanned documents, signs, or product labels into digital text.

2. Text-to-Speech Engines: These engines convert the extracted text into synthetic speech. Popular options include Apple's VoiceOver, MicrosoftNarrator, and Google Text-to-Speech. These engines offer various voices, allowing users to customize the experience.

2.2 PROPOSED SYSTEM

- Context-Aware TTS: Systems that understand the context of the text (e.g., news articles) could adjust speech style and emphasis for better comprehension.

- Integration with Assistive Technologies: Seamless integration with Braille displays or refreshable Braille lines could offer users a choice between audio and tactile reading experiences.

- Emotional Intelligence in TTS: Voices that convey emotions could make the listening experience more engaging and improve comprehension, especially for complex content.

By focusing on these advancements, i am created more user-friendly, versatile, and natural-sounding TTS systems, ultimately empowering visually impaired individuals to access information more efficiently and independently.

2.2 OBJECTIVES

1. Accessibility and User Empowerment

- Enhance access to information: The primary aim of enhancing access to information through Text-to-Speech (TTS) technology is to eliminate barriers faced by visually impaired individuals when accessing written content. By converting text into high-quality audio output that is clear and easily understandable, TTS technology ensures that visually impaired individuals can perceive and comprehend written information without relying on sight. This process involves sophisticated algorithms that analyze textual content and synthesize speech output that closely resembles natural human speech, ensuring clarity and articulation. By providing real-time access to information across various formats and platforms, TTS empowers visually impaired individuals to stay informed, engaged, and independent in educational, professional, and social contexts, thereby promoting inclusivity and equality in the digital age.

- Promote independent information access: The system aimed at promoting independent information access for visually impaired individuals strives to empower them to engage with information autonomously, fostering greater autonomy and participation in various facets of life. By harnessing technologies such as Text-to-Speech (TTS) and screen readers, visually impaired individuals gain the ability to interact with digital content independently, navigating documents, websites, emails, and more without relying on sighted assistance. This autonomy is further enhanced through customizable features, allowing users to adjust speech settings and navigation controls according to their preferences. By facilitating independent access to information, this system not only boosts confidence and self-reliance but also promotes inclusivity and equal participation in education, employment, communication, and leisure activities. In essence, it enables visually impaired individuals to lead more fulfilling and empowered lives, breaking down barriers to information and fostering a more equitable society.

2. User-Centric Design and Customization

- Optimize the listening experience: The system aims to optimize the listening experience for visually impaired individuals by offering customizable features that cater to individual preferences, ultimately enhancing comprehension and usability. Through features such as adjustable voice speed and voice gender selection, users can tailor the auditory output to suit their unique needs and preferences. This customization allows individuals to control the pace and style of speech, ensuring that the synthesized audio is comfortable to listen to and conducive to comprehension. By providing options for voice speed and gender, the system accommodates diverse preferences and enhances the overall user experience, ultimately promoting greater engagement and accessibility for visually impaired individuals.

- Seamless integration: The TTS system prioritizes seamless integration with existing assistive technologies, such as screen readers, to ensure a smooth workflow and effortless conversion of on-screen text into speech for visually impaired individuals. By seamlessly integrating with these assistive tools, the TTS system enables users to access and interact with digital content seamlessly, without disruption or compatibility issues. This integration streamlines the process of converting on-screen text into speech, providing visually impaired individuals with a cohesive and efficient user experience. As a result, users can navigate through various applications, documents, and web pages with ease, enhancing their ability to access information and participate fully in digital environments.

3. Advance TTS Technology for the Future

- Stay ahead of the curve: The project is committed to staying ahead of the curve by continuously monitoring advancements in Text-to-Speech (TTS) technology and incorporating cutting-edge features to enhance the listening experience for visually impaired individuals. By remaining updated on the latest developments, the project aims to integrate innovative capabilities such as emotional intelligence and context-awareness into the TTS system. These advancements will enable the system to generate speech output that not only accurately conveys information but also reflects nuanced emotional tones and adapts to contextual cues. By incorporating features that mimic human speech patterns more closely, the project endeavors to create a more natural and engaging listening experience for users, ultimately improving comprehension and usability. This proactive approach ensures that the TTS system remains at the forefront of accessibility technology, continuously evolving to meet the changing needs and expectations of visually impaired individuals.

- Expand functionalities: Future goals for the project involve expanding functionalities to encompass real-time text recognition and speech conversion, enabling visually impaired users to access information from their surroundings seamlessly. By implementing these advanced capabilities, the system will be equipped to recognize printed or digital text in real-time using image recognition technology and instantly convert it into speech output. This functionality will empower users to access information from various sources, including signs, labels, menus, and other printed materials, enhancing their independence and autonomy in navigating the physical world. By expanding its functionalities to include real-time text recognition and speech conversion, the project aims to further bridge the accessibility gap for visually impaired individuals, enabling them to engage with their environment more effectively and confidently.

CHAPTER 3
SYSTEM ANALYSIS AND DESIGN

3.1 SYSTEM ANALYSIS

The text-to-speech system designed for visually impaired individuals undergoes thorough system analysis to ensure optimal functionality and usability. This analysis involves assessing various aspects such as accuracy in converting text to speech, naturalness of the synthesized speech, compatibility with different input formats, and accessibility features for visually impaired users. Additionally, the system is evaluated for its ability to handle different languages and dialects, ensuring inclusivity for users worldwide. Usability testing plays a crucial role in the analysis, focusing on user interface design, ease of navigation, and integration with other assistive technologies. Through comprehensive system analysis, the text-to-speech system can be refined and tailored to meet the specific needs of visually impaired individuals, ultimately enhancing their reading experience and promoting greater independence and accessibility in accessing textual information.

3.1.1 INTRODUCTION

- Requirement Gathering: It involves identifying key stakeholders, including both the visually impaired users themselves and accessibility experts. Through user interviews, focus groups, and surveys, the project team can gather data on user needs, preferred functionalities, and challenges faced with current TTS systems. These needs are then categorized as functional requirements, defining core functionalities, and non-functional requirements, addressing usability, performance, and security.

- Functional Analysis: It focuses on input mechanisms like handling digital text, screen reader capture, and potentially even real-time text recognition. Text processing examines grammar, punctuation, and special characters for accurate pronunciation and phrasing. The system must integrate seamlessly with a high-quality TTS engine offering a variety of natural-sounding voices. User control over speech rate, pitch, and volume is crucial, along with voice selection and background noise reduction for a personalized experience. Finally, the system should generate clear and compatible audio output for various assistive devices. By thoroughly analyzing these functionalities, developers can ensure the TTS system effectively converts text to speech in a way that empowers visually impaired users to access information with ease.

- Non-functional Analysis: It examines user experience and system characteristics. Accessibility is paramount, ensuring the interface is usable for visually impaired users, potentially through text-to-speech prompts or integration with screen readers. Performance is crucial, with minimal latency in speech generation for a smooth listening experience. The system should be adaptable to handle diverse text formats, languages, and future advancements in TTS technology (scalability). Finally, if the system handles sensitive information, robust security measures are essential to protect user privacy. By considering these non-functional aspects, developers can create a TTS system that is not only functional but also user-friendly, reliable, and future proof, ultimately empowering visually impaired people to access information with greater confidence.

- Data Analysis: It is critical for refining this TTS system based on user needs. Information gathered during interviews and surveys is analyzed to identify patterns in user behavior and challenges. This analysis can reveal how often people use TTS, in what situations, and which features are most desired (e.g., voice selection, noise control, real-time text recognition). It can also pinpoint difficulties users face with existing systems, such as unnatural voices, slow processing, or compatibility issues. By analyzing this data, the project team can prioritize functionalities based on user needs, identify areas for improvement, and ultimately develop a system that effectively addresses the challenges visually impaired people experience when accessing information through text-to-speech technology.

- Performance Evaluation: Performance evaluation of the text-to-speech (TTS) system for visually impaired individuals involves assessing its accuracy, efficiency, and reliability in converting text to speech. This evaluation encompasses various aspects, including the system's ability to accurately interpret text from different input sources, the naturalness and intelligibility of the synthesized speech output, and its responsiveness and compatibility with different devices and platforms. Additionally, performance evaluation may involve benchmarking the TTS system against industry standards and competing solutions to gauge its effectiveness and identify areas for improvement. Through rigorous testing and analysis, developers can ensure that the TTS system meets the needs of visually impaired users and delivers a high-quality and seamless reading experience.

3.1.2 METHODOLOGY

The methodology for developing the text-to-speech system for visually impaired individuals involves comprehensive research to understand user needs, followed by defining system requirements. Technology selection, prototype development, and iterative refinement through user testing ensue, focusing on optimizing user interface and accessibility. Rigorous testing assesses system performance, including accuracy, reliability, and usability across various contexts. Validation ensures the system effectively meets user needs, resulting in a robust solution that enhances accessibility and promotes independence in accessing textual information for visually impaired individuals.

1. Text Processing:

➢ Preprocessing: During the preprocessing stage, the input text undergoes initial cleaning and preparation to ensure it is in a suitable format for further processing. This involves various tasks aimed at standardizing and refining the text. One crucial aspect of preprocessing is normalization, which involves ensuring consistent capitalization, punctuation, and spacing throughout the text. By standardizing these elements, normalization helps create uniformity and readability, making the text more conducive to subsequent processing tasks. Additionally, preprocessing may involve tasks such as removing irrelevant characters or symbols, handling special cases, and addressing any formatting inconsistencies. Overall, preprocessing lays the groundwork for effective text analysis and speech generation by refining the input text and preparing it for more advanced linguistic processing. It involve:

- Normalization: Normalization is a crucial step in text preprocessing, aimed at ensuring consistency in capitalization, punctuation, and spacing throughout the text. This process helps standardize the text, making it more uniform and easier to process. Consistent capitalization ensures that words are formatted in a standardized manner, such as converting all letters to lowercase or uppercase, or applying title case rules. This prevents inconsistencies that could affect text analysis and speech synthesis.

Punctuation normalization involves ensuring that punctuation marks are used consistently and appropriately. This may include standardizing the use of punctuation marks such as periods, commas, and quotation marks, and handling special cases like hyphens and dashes. Spacing normalization focuses on maintaining consistent spacing between words and sentences. It ensures that there are no extra spaces or missing spaces that could affect the interpretation of the text.

By normalizing capitalization, punctuation, and spacing, the text becomes more consistent and structured, facilitating subsequent processing tasks such as phoneme conversion and speech generation. This improves the overall quality and readability of the synthesized speech output.

- Text Analysis: In the text analysis stage, the focus shifts towards a more in-depth examination of the text in preparation for speech generation. This phase involves breaking down the textual content into its fundamental components and analyzing various linguistic aspects to ensure accurate and natural-sounding speech output.

  During text analysis, several key tasks are performed to refine and prepare the text for synthesis into spoke

  In the text analysis stage, the focus shifts towards a more in-depth examination of the text in preparation for speech generation. This phase involves breaking down the textual content into its fundamental components and analyzing various linguistic aspects to ensure accurate and natural-sounding speech output.

  During text analysis, several key tasks are performed to refine and prepare the text for synthesis into spoken words.

  These tasks may include:

- Phoneme Conversion: Each word in the text is converted into a sequence of phonemes, which are the smallest units of sound in a language. This process involves mapping each word to its corresponding phonetic representation, taking into account pronunciation variations and phonological rules.

- Prosodic Analysis: Prosody refers to the patterns of stress, intonation, and rhythm in speech. In this stage, the text is analyzed to identify and incorporate appropriate prosodic features, such as pitch contours, emphasis on certain words or phrases, and pauses, to ensure that the synthesized speech sounds natural and expressive.

- Language Modeling: Language modeling involves analyzing the text for grammatical structure, syntax, and semantics. By understanding the linguistic context of the text, the system can generate speech that is coherent and contextually appropriate. This may involve incorporating knowledge of grammar rules, word relationships, and semantic meaning into the speech synthesis process.

- Error Handling: Text analysis also includes identifying and handling any errors or ambiguities in the input text. This may involve resolving spelling mistakes, correcting grammatical errors, or disambiguating homophones and other lexical ambiguities to ensure accurate speech synthesis.

Overall, text analysis is a critical stage in the speech synthesis process, as it lays the foundation for generating high-quality and natural-sounding speech output. By delving deeper into the linguistic structure of the text and incorporating phonetic, prosodic, and semantic information, the system can produce synthesized speech that is clear, intelligible, and contextually appropriate.

2. Speech Generation:

In the speech generation phase, the Text-to-Speech (TTS) system transforms the processed text into spoken words.
This phase involves two main processes:

- waveform generation
- post-processing

- Waveform Generation:

  In this step, the TTS engine utilizes the phoneme sequence and prosody information obtained during text analysis to produce a digital representation of the speech waveform. The phoneme sequence represents the sequence of basic speech sounds corresponding to the words in the input text. Prosody information includes details about intonation, stress, rhythm, and other aspects of speech that contribute to naturalness and expressiveness. Using this information, the TTS engine generates a digital waveform that represents the speech signal. This waveform typically consists of a sequence of digital samples that encode the amplitude of the sound wave at different points in time. The waveform represents the acoustic properties of the synthesized speech, including pitch, duration, and intensity.

- Post-Processing:

- After waveform generation, the raw synthesized speech may undergo additional refinements in the post-processing stage to improve naturalness and overall quality. This may involve applying various techniques to enhance the clarity, smoothness, and intelligibility of the speech output. Post-processing techniques may include:
- Smoothing transitions between phonemes and words to reduce abruptness and improve fluency.
- Applying pitch modification to adjust the intonation and pitch contour of the speech, ensuring that it sounds more natural and expressive.
- Removing or reducing noise and distortion from the synthesized speech signal to improve clarity and intelligibility.
- Adding pauses and other prosodic features to enhance the rhythm and pacing of the speech.
- The goal of post-processing is to refine the synthesized speech to make it sound more human-like, natural, and pleasant to listen to. By applying these final refinements, the TTS system can produce high-quality speech output that closely resembles natural human speech.

In summary, the speech generation phase of the TTS process involves converting processed text into a digital waveform representation of speech, followed by post-processing to refine and enhance the quality and naturalness of the synthesized speech output. These processes collectively enable the TTS system to generate clear, intelligible, and expressive speech that meets the needs and expectations of users.

3. Additional Considerations:

- Voice Selection: Voice selection is a feature commonly found in Text-to-Speech (TTS) systems that allows users to choose from a variety of virtual voices, each with distinct characteristics such as gender, accent, and speaking style. This functionality provides users with the flexibility to customize the synthesized speech output to better suit their preferences, needs, and context.

- Gender: One of the primary characteristics that users can select when choosing a virtual voice is gender. TTS systems typically offer options for both male and female voices, allowing users to choose the gender that they prefer or find most suitable for their purposes. This choice can be based on personal preference, cultural considerations, or the specific context in which the synthesized speech will be used.

- Speed: Users can tailor the speaking speed to their preferences, selecting a pace that feels comfortable and natural to them. Some users may prefer a faster speaking speed for efficiency and rapid consumption of information, while others may prefer a slower speed for enhanced clarity and comprehension.

Overall, voice selection in TTS systems offers users the opportunity to customize the synthesized speech output to their preferences and needs. By providing options for different voices with varying characteristics such as gender, accent, and speaking style, TTS systems enhance the usability, accessibility, and personalization of synthesized speech for a wide range of users and applications.

3.1.3 HARDWARE REQUIREMENTS

- Processor: A standard CPU (Central Processing Unit) on a computer or mobile device should suffice for most software-based TTS systems.

- Memory (RAM): Enough RAM to run the TTS software smoothly. The exact amount depends on the specific software but wouldn't require high-end specifications.

- Storage: Storage space to store the TTS software and potentially pre-recorded voice data (if using concatenative synthesis). Again, the amount depends on the software but wouldn't be a major factor.

- Audio Output Device: Headphones or speakers with decent audio quality.

.

3.1.4 SOFTWARE REQUIREMENTS

- Operating System (OS): The choice of operating system depends on the target platform. Popular options include Windows, macOS, Android, and iOS. The system should be compatible with the chosen OS and its accessibility features.

- Text Processing Software: Software libraries or tools capable of handling tasks like text normalization, abbreviation expansion, and part-of-speech tagging. Open-source libraries like FreeText or commercial options can be used.

- Text-to-Speech Engine: The core component responsible for speech generation. Open-source TTS engines like eSpeak or commercially licensed options with advanced features like voice selection and emotional intelligence can be considered.

3.2 SYSTEM DESIGN

The system design for the text-to-speech solution tailored to visually impaired individuals entails a meticulous process focused on usability and accessibility. It begins with conceptualizing the user interface, emphasizing intuitive navigation and integration of auditory cues to guide users seamlessly through the application. The architecture incorporates image-to-text conversion via Optical Character Recognition (OCR) and subsequent text-to-speech synthesis, ensuring accurate interpretation and natural-sounding speech output. Accessibility features such as high contrast, large fonts, and voice commands are integrated to accommodate diverse user needs. Additionally, the system is designed to be compatible with various mobile platforms and devices, enhancing its versatility and usability for visually impaired individuals across different contexts. Iterative refinement based on user feedback ensures that the system design remains responsive to evolving user requirements, ultimately delivering a robust and user-friendly solution for accessing textual information.

3.2.1 INTRODUCTION

Imagine a world brimming with information, yet vast portions remain inaccessible to those who are visually impaired. Text-to-Speech (TTS) technology bridges this gap, transforming written content into clear and natural-sounding speech. This project delves into the design of a TTS system specifically tailored to the needs of visually impaired individuals, empowering them to independently access and engage with a wealth of information.

Our design goes beyond simply converting text to speech. We understand that reading preferences and styles vary greatly. The system will offer a user-centric approach, allowing users to personalize their experience. Imagine customizing speech rate, selecting the perfect voice for different contexts, or even filtering out background noise for a focused listening experience.

Furthermore, the seamless integration of the text-to-speech (TTS) system with existing assistive technologies such as screen readers is paramount. This integration facilitates effortless conversion of on-screen text into concise audio, greatly enhancing daily interactions with computers and digital devices. For instance, envision a student effortlessly accessing textbooks and online resources or a professional seamlessly navigating through documents and emails, all empowered by the innovative capabilities of this TTS system.

The impact of this project extends beyond individual users to broader societal realms. By enabling independent access to information, this TTS system fosters greater social inclusion. It empowers visually impaired individuals to actively participate in education, employment, and various facets of modern life. Through this design, the system lays the groundwork for a more inclusive and information-rich world, benefiting not only visually impaired individuals but society as a whole.

Additionally, the versatility of the TTS system extends its impact to various domains, including education, employment, and leisure. Students can seamlessly access textbooks and online resources, enhancing their learning experience, while professionals can efficiently navigate documents and emails, improving productivity and efficiency. By facilitating access to information in multiple contexts, the TTS system enriches the lives of visually impaired individuals, enabling them to actively engage in diverse activities and pursuits.

In conclusion, the innovative design and functionalities of the TTS system offer a transformative solution for visually impaired individuals, revolutionizing their access to information and promoting greater independence and inclusion. By fostering seamless integration with existing assistive technologies, empowering users across diverse contexts, and extending its impact to various domains, the TTS system signifies a significant step towards creating a more inclusive and equitable society for all.

3.2.2 MODULE DESCRIPTION

The module description outlines the distinct components and functionalities of the text-to-speech (TTS) system designed for visually impaired individuals. It encompasses several key modules, including image-to-text conversion through Optical Character Recognition (OCR), text preprocessing, speech synthesis, and user interface design. The OCR module accurately interprets printed text from images captured by a mobile camera, while text preprocessing ensures optimal formatting and clarity for speech synthesis. The speech synthesis module utilizes advanced algorithms to generate natural-sounding speech output, enhancing the user experience. Additionally, the user interface module focuses on designing an intuitive and accessible interface, incorporating features such as high contrast, large fonts, and voice commands to cater to diverse user needs. Through meticulous module description, the TTS system is tailored to meet the specific requirements of visually impaired individuals, enabling seamless access to textual information and promoting greater independence and inclusivity.

- **TEXT INPUT MODULE**

  The Text Input Module serves as the entry point for the Text-to-Speech (TTS) system, responsible for processing text sourced from various digital sources. One of the primary functions of this module is handling digital text files, which may include documents.

- Digital Text Files: Processes text from documents.

- Screen Reader Capture: Grabs text displayed on the computer screen for narration.

Overall, the Text Input Module plays a crucial role in facilitating the input of text data into the TTS system, enabling the subsequent processing and synthesis of speech output. By supporting a diverse range of digital text sources and formats, it enhances the versatility and applicability of the TTS system across different use cases and domains.

- **TEXT PREPROCESSING MODULE**

  It is a crucial component of the Text-to-Speech (TTS) system responsible for preparing the input text for further processing and synthesis. This module performs several key tasks to enhance the quality and accuracy of the synthesized speech output.

- Normalization:

  The normalization process ensures consistent formatting of the text, including aspects such as capitalization, punctuation, and spacing. By standardizing these elements, normalization improves the readability and consistency of the text, facilitating subsequent processing stages.

  For example, normalization may convert inconsistent capitalization (e.g., "tExt" to "text"), ensure proper punctuation usage, and address spacing inconsistencies.

- Abbreviation and Acronym Expansion:

  Abbreviations and acronyms are common in textual content but may pose challenges for accurate pronunciation in synthesized speech.

  The abbreviation and acronym expansion task involves converting abbreviated forms into their full forms to ensure accurate pronunciation by the TTS system.

By performing normalization and abbreviation/acronym expansion, the Text Preprocessing Module improves the quality and accuracy of the input text, ultimately enhancing the clarity and intelligibility of the synthesized speech output. These preprocessing tasks lay the foundation for more effective text analysis and speech synthesis in subsequent stages of the TTS pipeline.

➢ TEXT ANALYSIS MODULE

The Text Analysis Module plays a pivotal role in the Text-to-Speech (TTS) system by delving deeper into the input text to prepare it for accurate and natural speech synthesis. This module encompasses two essential tasks:

- Phoneme Conversion:

- Phonemes are the basic units of speech sounds, representing distinct sounds in a language.

- The phoneme conversion task involves converting each word in the input text into a sequence of phonemes.

- By breaking down words into their constituent phonemes, the TTS system can accurately synthesize the corresponding speech sounds, ensuring correct pronunciation and naturalness in the synthesized speech output.

- For example, the word "cat" may be converted into the phoneme sequence /k/ /æ/ /t/.

- Prosody Prediction:

- Prosody refers to the patterns of stress, intonation, rhythm, and pitch variation in speech.

- The prosody prediction task analyzes various factors such as punctuation, sentence structure, and context to determine prosodic features that influence speech characteristics.

- This analysis includes predicting factors such as voice speed (rate of speech), gender selection (male or female voice), volume (loudness), and intonation patterns.

- For instance, the presence of a question mark at the end of a sentence may indicate a rising intonation pattern associated with questions.

- By accurately predicting prosodic features, the TTS system can produce speech output that sounds natural, expressive, and contextually appropriate.

➢ SPEECH GENERATION MODULE

The Speech Generation Module is a critical component of the Text-to-Speech (TTS) system responsible for transforming processed text into synthesized speech. One of the primary tasks of this module is waveform generation, which involves creating a digital representation of the speech waveform using various synthesis techniques, such as concatenative or parametric synthesis.

- Waveform Generation:
- In waveform generation, the module creates a digital representation of the speech waveform, which comprises the amplitude of the sound wave over time.
- This waveform represents the acoustic properties of synthesized speech, including characteristics such as pitch, duration, and intensity.
- Waveform generation techniques can vary depending on the specific synthesis approach employed by the TTS system. Two common techniques are concatenative synthesis and parametric synthesis.

Through waveform generation, the Speech Generation Module converts the linguistic information derived from the input text into acoustic signals, resulting in synthesized speech that is intelligible, expressive, and contextually appropriate. This module is integral to the overall functionality and effectiveness of the TTS system in facilitating communication and accessibility for users.

➢ AUDIO OUTPUT MODULE

This module encompasses two essential functionalities:

- Screen Reader Integration:
  - Screen reader integration enables seamless collaboration between the TTS system and existing screen reader software.
  - Screen readers are assistive technologies that convert on-screen text into speech or Braille output, primarily used by individuals with visual impairments or reading difficulties.
  - By integrating with screen readers, the Audio Output Module ensures that the synthesized speech is seamlessly integrated into the overall screen reading experience.
  - This integration allows visually impaired users to access and navigate digital content more effectively, as the synthesized speech provides auditory feedback for on-screen text content.

- Headphone/Speaker Compatibility:
  - The Audio Output Module is designed to output the synthesized speech audio to headphones or speakers connected to the user's device.
  - This compatibility ensures that users can listen to the synthesized speech output privately through headphones or share it with others using external speakers.
  - By supporting headphone/speaker compatibility, the TTS system accommodates diverse user preferences and environments, enabling users to access synthesized speech output in various contexts, such as in public spaces, private settings, or while on-the-go.

By offering screen reader integration and headphone/speaker compatibility, the Audio Output Module enhances the accessibility, usability, and versatility of the TTS system for users with diverse needs and preferences. Whether integrated with existing assistive technologies or delivering speech output through headphones or speakers, this module ensures that synthesized speech is accessible and readily available to users in a format that suits their individual requirements.

3.2.3 SYSTEM ARCHITECTURE

The designed text-to-speech (TTS) system for visually impaired people leverages a modular architecture for efficient text processing and user interaction. This modular approach breaks down the system into distinct functional blocks, each handling a specific task. The first module tackles input, accepting text from various sources. This includes digital files, on-screen text captured by screen readers.

Following text input, a dedicated processing module takes center stage. Here, the system cleans and prepares the text, ensuring consistent formatting and expanding abbreviations for accurate pronunciation. It then delves deeper, converting words into their basic building blocks of sound – phonemes – and predicting speech patterns like rhythm and pitch based on sentence structure and punctuation. This processed information is then passed on to the core of the system – the Text-to-Speech engine. Here, magic happens. The engine utilizes the provided phonemes and prosody information to generate a digital representation of the speech waveform. This can involve techniques like stitching together pre-recorded sounds or building the waveform based on parameters (parametric synthesis). In some systems, users may even have the option to select their preferred voice for the synthesized speech.

The final stage involves refining the raw generated speech to ensure a natural and pleasant listening experience. A post-processing module eliminates any abrupt transitions between sounds and adjusts volume and noise levels for clarity. Finally, the polished audio reaches the user through their chosen output method, be it headphones, speakers, or seamless integration with existing screen readers for on-screen text narration. An optional user interface layer can further enhance the experience by providing text-to-speech prompts to guide users and offering customization options for speech rate, pitch, and volume. This modular architecture offers several advantages. It allows for independent development and improvement of individual modules, making the system adaptable and scalable for future advancements. More importantly, it ensures a robust and user-centric system, empowering visually impaired people to access information through high-quality synthesized speech.

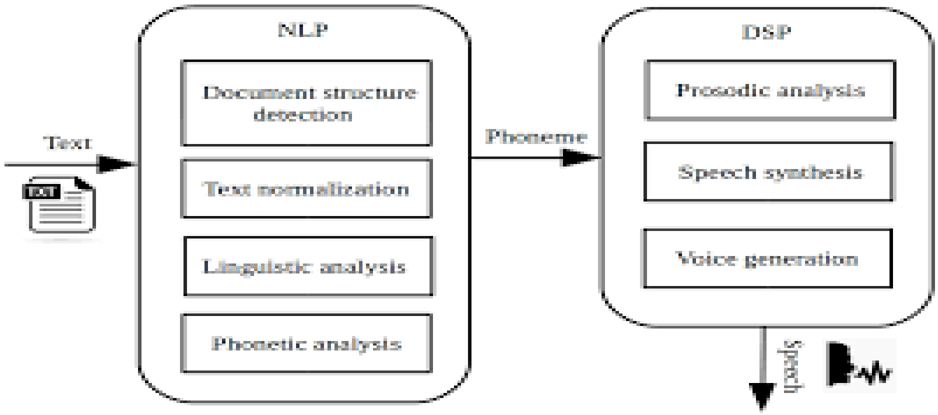

Figure 1 Architecture Diagram

3.3 RESULT AND DISCUSSION

The designed text-to-speech (TTS) system promises significant results in empowering visually impaired users to access information independently. This system acts as a key that unlocks a vast world of knowledge, previously inaccessible due to visual limitations. By converting diverse text formats into clear, natural-sounding speech, the system removes barriers. Digital files on-screen text captured by screen readers are all transformed into a readily consumable audio format. This empowers users to access ebooks, educational materials, and even everyday documents with newfound independence, fostering a sense of self-reliance in information acquisition.

Furthermore, the system prioritizes user experience by offering a personalized listening environment. Users can create a comfortable and engaging experience by customizing speech rate, pitch, and potentially even selecting their preferred voice. This caters to individual learning styles and preferences, ensuring everyone can benefit from the system's functionalities. Additionally, seamless integration with screen readers eliminates the need for users to switch between applications. They can effortlessly convert on-screen text into concise audio, creating a smooth workflow for interacting with digital content. Ultimately, this user-centric approach translates to greater social inclusion and participation. By promoting independent access to information, the TTS system empowers visually impaired people to actively engage in education, employment, and everyday activities, fostering a more inclusive society where everyone has equal access to knowledge.

CHAPTER 4

SUMMARY

4.1 CONCLUSION

In the digital age, access to information is fundamental, yet visually impaired individuals often face significant barriers when it comes to accessing printed or digital text. The introduction of a Text-to-Speech (TTS) system marks a transformative step in bridging this gap, offering a solution that converts diverse text formats into high-quality audio speech. This system holds immense potential to empower visually impaired individuals by granting them access to a wealth of knowledge that was previously inaccessible.

The TTS system's ability to process various text formats, including physical documents through Optical Character Recognition (OCR), opens up new avenues for accessing information. By converting text into clear and intelligible audio, the system dismantles barriers and empowers users with newfound independence. No longer bound by the limitations of printed text, visually impaired individuals can navigate through documents, eBooks, web pages, and more, with ease and confidence.

One of the key strengths of the designed TTS system lies in its focus on user experience. Customizable speech settings, such as speed and potentially voice selection, allow users to tailor the listening experience to their preferences. This level of customization caters to individual learning styles and ensures a comfortable and engaging environment for all users. Moreover, seamless integration with screen readers streamlines the workflow, eliminating the need to switch between applications and enhancing overall usability.

Looking ahead, the modular architecture of the TTS system paves the way for continuous improvement and innovation. Future advancements could include real-time text recognition for on-the-fly speech generation from physical surroundings, further enhancing accessibility and usability. Additionally, incorporating emotional intelligence into the TTS engine could result in more natural-sounding and engaging speech output, fostering deeper engagement with the content.

In conclusion, the text-to-speech (TTS) system stands as a pivotal tool in empowering visually impaired individuals, facilitating their active engagement in an inclusive society. By placing user needs at the forefront and prioritizing accessibility, the system serves as a bridge to overcome the information gap, ensuring equitable access to information as a fundamental right for all individuals. Through its blend of advanced technology and user-centered design, the TTS system has the transformative potential to enrich lives, providing avenues to knowledge, opportunities, and fostering a more inclusive societal landscape.

Embracing ongoing development and refinement, the TTS system continues to evolve, bolstering its impact and relevance in the lives of visually impaired individuals. By remaining responsive to emerging needs and technological advancements, the system adapts to meet the evolving challenges and aspirations of its users. This commitment to continuous improvement underscores its significance as a catalyst for positive change, reinforcing the principles of inclusivity and empowerment.

Ultimately, the TTS system embodies more than just a technological advancement; it embodies a vision of a world where every individual, regardless of visual impairment, can fully participate and contribute. Through its transformative capabilities, the system not only opens doors to knowledge and opportunities but also fosters a culture of empathy, understanding, and equality. In this way, the TTS system paves the way towards a future where inclusivity is not just an aspiration but a tangible reality for all.

4.2 FUTURE ENHANCEMENT

The designed Text-to-Speech (TTS) system stands as a beacon of empowerment for visually impaired individuals, offering unprecedented access to information. Its core functionalities already mark significant improvements, but the modular architecture hints at exciting future advancements that promise to revolutionize the user experience.

One area ripe for exploration is real-time text recognition. Envision a scenario where users can point their device's camera at physical documents or signs, instantly converting captured text into spoken language. This breakthrough would redefine daily activities, enabling seamless information access in real-world environments.

Furthermore, imbuing the TTS engine with emotional intelligence presents another intriguing possibility. By analyzing sentiment and context, the system could generate speech that not only conveys words but also the underlying emotions. This nuanced approach would create a more engaging listening experience, fostering deeper comprehension and connection with the content.

Looking ahead, personalization emerges as a key frontier. Picture a system that learns individual user preferences over time, automatically adjusting speech parameters based on context or previous interactions. This level of customization would elevate user comfort and create a truly tailored listening experience.

By actively pursuing these advancements, the TTS system transcends its role as a mere accessibility tool. It becomes a dynamic companion, empowering visually impaired individuals to not just access information but engage with it on a deeper level. This future holds immense promise for fostering greater inclusion and participation across all aspects of life.

CHAPTER 5

APPENDIX

5.1 SOURCE CODE

5.1.1 BACKEND

```
import pyttsx3
import PyPDF2

engine=pyttsx3.init()
speaker=pyttsx3.init()

def speaknow():
    print("speaking")
    text = text_area.get(1.0, END)
    gender = gender_combobox.get()
    speed = speed_combobox.get()
    voices = engine.getProperty('voices')

    def setvoice():
        if gender == 'Male':
            engine.setProperty('voice', voices[0].id)
        else:
            engine.setProperty('voice', voices[1].id)

        if speed == 'Fast':
```

```
            engine.setProperty('rate', 250)
        elif speed == 'Normal':
            engine.setProperty('rate', 150)
        else:
            engine.setProperty('rate', 50)

        engine.say(text)
        engine.runAndWait()

    if text.strip(): # Check if text is not empty
        setvoice()
    else:
        print("No text to speak")

def download():
    text=text_area.get(1.0,END)
    gender=gender_combobox.get()
    speed=speed_combobox.get()
    voices = engine.getProperty('voices')

    def setvoice():
        if (gender=='Male'):
            engine.setProperty('voice',voices[0].id)
            path=filedialog.askdirectory()
            os.chdir(path)
            engine.save_to_file(text,'text.mp3')
            engine.runAndWait()
        else:
            engine.setProperty('voice',voices[1].id)
            path=filedialog.askdirectory()
            os.chdir(path)
            engine.save_to_file(text,'text.mp3')
            engine.runAndWait()
```

```
    if (text):
        if (speed=='Fast'):
            engine.setProperty('rate',250)
            setvoice()
        elif (speed=='Normal'):
            engine.setProperty('rate',150)
            setvoice()
        else:
            engine.setProperty('rate',50)
            setvoice()

selected_file_path = ""

def fileDialog():
    global selected_file_path
    path = filedialog.askopenfilename()
    selected_file_path = path
    print("Selected file path:", selected_file_path)
    book = open(path, 'rb')
    pdfReader = PyPDF2.PdfReader(book)
    pages = len(pdfReader.pages)

    try:
        start_page_number = int(start_page_number_entry.get())
        end_page_number = int(ending_page_number_entry.get())
    except ValueError:
        print("Please enter valid page numbers (integers).")
        return

    # Ensure end page number is within document boundaries
    if end_page_number > pages:
```

```
        print(f"Error: Ending page number ({end_page_number}) exceeds total pages ({pages}).")
        return

    # Clear existing text in the text area
    text_area.delete(1.0, END)

    for num in range(start_page_number - 1, end_page_number): # Adjust for zero-based indexing
        page = pdfReader.pages[num]
        txt = page.extract_text()

        # Append the text to the text area
        text_area.insert(END, txt + "\n\n")

        # Read aloud the text
        speaker.say(txt)
        speaker.runAndWait()

    book.close()

# Update the "Choose File" button command to call the modified fileDialog() function
choose_file_button = Button(page_number_frame, text="Choose File", font="arial 12", command=fileDialog)
choose_file_button.pack(side=RIGHT)



def speech():
    speaknow()
    if button_clicked:
        fileDialog()
```

```
# Add a global variable to track whether the "speech" button is clicked
button_clicked = False

def on_speech_button_click():
    global button_clicked
    button_clicked = True
    speech()
```

5.1.2 FRONTEND

```
import tkinter as tk
from tkinter import *
from tkinter import filedialog
from tkinter import filedialog
from tkinter.ttk import Combobox
import os
root = Tk()

#title
root.title("Text To Speech")
root.geometry("900x450+250+150")
root.resizable(False,False)
root.configure(bg="#305065")

#icon
icon_image = PhotoImage(file="speak.png")
root.iconphoto(True, icon_image)

#TopFrame
Top_frame=Frame(root,bg="white",width=900,height=100)
Top_frame.place(x=0,y=0)

logo=PhotoImage(file="speaker logo.png")
Label(Top_frame,image=logo,bg="white").place(x=10,y=1)
Label(Top_frame,text="TEXT TO SPEECH",font="arial 20
bold",bg="white",fg="black").place(x=100,y=35)

text_area=Text(root,font="Robote 20",bg="white",relief=GROOVE,wrap=WORD)
text_area.place(x=10,y=150,width=500,height=150)
```

```
# Page number and choose file area
page_number_frame = Frame(root, bg="#305065", highlightbackground="white")
page_number_frame.place(x=10, y=310, width=500, height=40)

label_page = Label(page_number_frame, text="Start Page Number:", font="arial 12",
bg="#305065", fg="white").pack(side=LEFT)
start_page_number_entry = Entry(page_number_frame, font="arial 12", width=5)
start_page_number_entry.pack(side=LEFT)

label_page1 = Label(page_number_frame, text="End Page Number:", font="arial 12",
bg="#305065", fg="white").pack(side=LEFT)
ending_page_number_entry = Entry(page_number_frame, font="arial 12", width=5)
ending_page_number_entry.pack(side=LEFT,  padx=10)   # Place after start page
number

#voice and speed dropdown
Label(root,text="VOICE",font="arial                    15
bold",bg="#305065",fg="white").place(x=580,y=160)
Label(root,text="SPEED",font="arial                    15
bold",bg="#305065",fg="white").place(x=760,y=160)

gender_combobox=Combobox(root,values=['Male','Female'],font="arial
14",state='r',width=10)
gender_combobox.place(x=550,y=200)
gender_combobox.set('Male')

speed_combobox=Combobox(root,values=['Slow','Normal','Fast'],font="arial
14",state='r',width=10)
speed_combobox.place(x=730,y=200)
speed_combobox.set('Normal')

imageicon1=PhotoImage(file="speak.png")
```

```
btn=Button(root,text="Speak",compound=LEFT,image=imageicon1,width=125,font=
"arial 14 bold",command=speech)
btn.place(x=550,y=290)

imageicon2=PhotoImage(file="download.png")
save=Button(root,text="Save",compound=LEFT,image=imageicon2,width=125,font=
"arial 14 bold",command=download)
save.place(x=730,y=290)
```

5.1.3 SCREENSHOTS

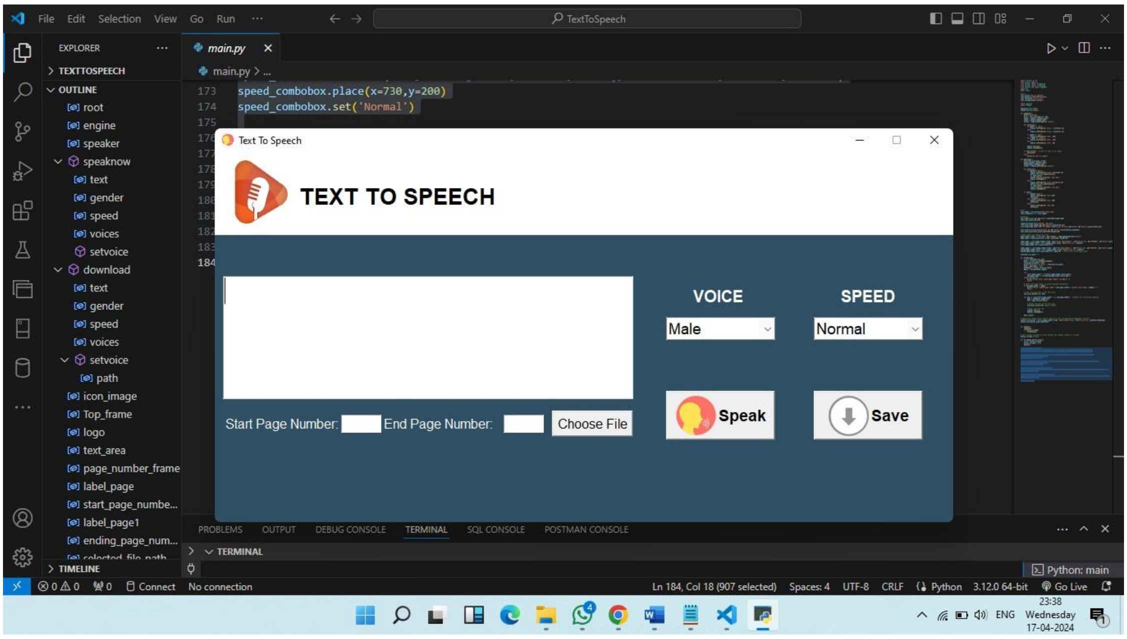

Fig 5.3.1 Basic Structure

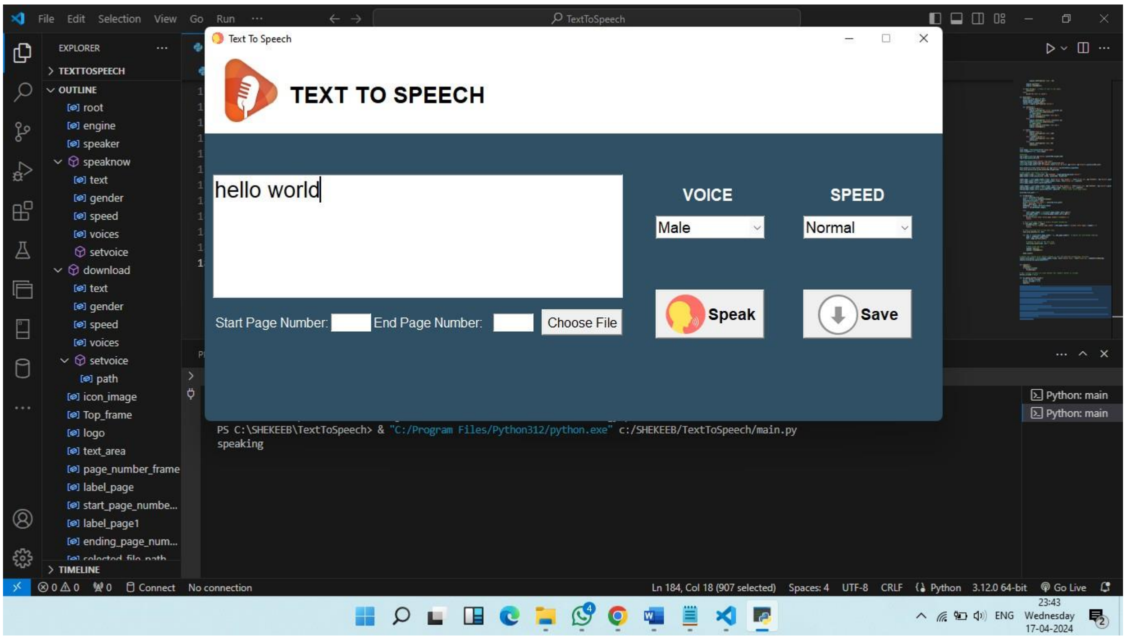

Fig 5.3.2 Reading text from textbox

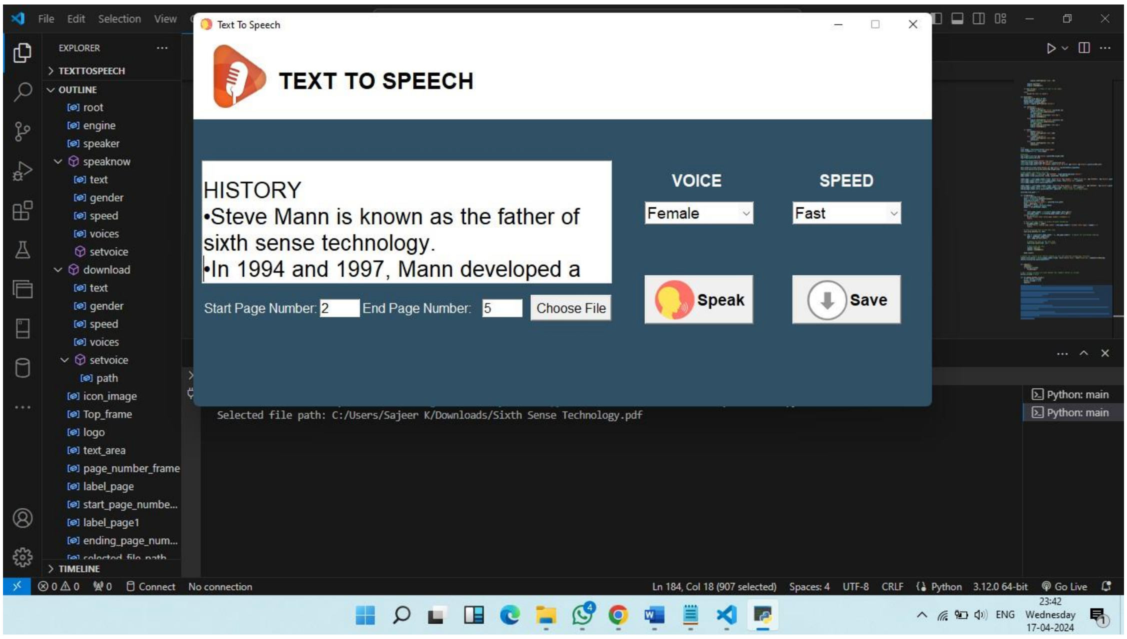

Fig 5.3.3 Reading text from a PDF file

CHAPTER 6

REFERENCES

1. https://data-flair.training/blogs/python-text-to-speech/

2. https://nevonprojects.com/text-to-speech-project/

3. https://paperswithcode.com/task/text-to-speech-synthesis

4. https://ijrpr.com/uploads/V3ISSUE5/IJRPR4449.pdf

5. https://www.ijert.org/research/implementation-of-text-to-speech-conversion-IJERTV3IS030548.pdf