Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Next Gen Smart Voice Assistant for Blind & Handicapped Users

Pavan Bhiwaji Kalambe[1], Prof. Monika Ingole[2], Prof. Vijayta Dalwankar[3]

[1] Student, Computer Science & Engineering, Wainganga College of Engineering and Management
[2,3] Professor, Computer Science & Engineering, Wainganga College of Engineering and Management

Abstract
Visually impaired individuals encounter significant difficulties in accessing printed and digital textual information, which limits their independence and learning opportunities. Despite advancements in digital technology, most reading systems remain inaccessible or require external assistance. This paper presents the design and implementation of a **web-based intelligent PDF-to-speech reading system** aimed at improving document accessibility for visually impaired users. The proposed system enables users to upload PDF documents, preview the content, extract textual data using PDF processing techniques, and convert the extracted text into audible speech through Text-to-Speech (TTS) technology. The system incorporates user-controlled functionalities such as **play, pause, stop, and resume**, allowing seamless and hands-free interaction via headphones or speakers. This approach provides a cost-effective solution for education, information access, and inclusive digital interaction for visually impaired individuals.

A Voice Assistant is one of the hot topics in the current world that are programs that listens to human's verbal command and respond to them which makes it a human computer/device interaction. In the current days, a voice assistant is everywhere which is a lot useful in these busy days. Nowadays, almost everyone in the current world is using voice assistant because it's everywhere starting from Google smartphone assistant which even 5 years old kids will know how to use because of the current world pandemic which makes them use smartphones till Amazon's Alexa which will be very useful to do works starting from entertaining the users till turning on and off the household products (Internet of Things). One of the greatest features is that it will be very useful to even physically challenged people, for example, people who aren't able to walk use the Internet of Things (IoT) feature to operate the household products and maintain them. So, we tend to develop a voice assistant which will be very useful to the users same as the other voice assistants which are currently in the world.

Next-generation smart voice assistants for blind and handicapped users are transitioning from simple command-response systems to AI-driven, proactive agents that combine computer vision, IoT, and natural language understanding (NLU) to provide environmental context and independence. These tools are designed to work hands-free, offering real-time scene description, navigation, and object recognition.
A smart, AI-powered voice assistant designed specifically for visually impaired and physically handicapped users, enabling hands-free, screen-free interaction with devices and the surrounding environment using natural voice commands.

Keywords: IoT, NLU, AI, JAVA, HTML, Visually Impaired, Assistive Technology, PDF Text Extraction, Text-to-Speech (TTS), Web-Based Application, Document Accessibility, Speech Synthesis, Human–Computer Interaction.

1. Introduction

Access to written information plays a crucial role in education, communication, and daily activities. However, visually impaired individuals face significant challenges in accessing printed and digital documents independently. Most textual content, such as books, academic materials, official documents, and online resources, is primarily designed for sighted users, creating a barrier to equal information access for visually impaired users.

Although several assistive technologies exist, many solutions require specialized hardware, complex installation procedures, or expensive proprietary software. Screen readers often depend on structured digital formats and may fail to provide accurate output when dealing with unstructured or complex PDF documents. As a result, visually impaired individuals frequently rely on others for reading tasks, which affects their independence, confidence, and productivity.

To address these challenges, this paper presents the **design and development of a web-based intelligent PDF-to-speech reading system** that enables visually impaired users to access document content using voice output. The proposed system allows users to upload PDF documents, extract textual content using PDF processing techniques, and convert the extracted text into audible speech through Text-to-Speech (TTS) technology. The system operates entirely within a web browser, eliminating the need for additional hardware or software installations.

The proposed solution provides interactive controls such as play, stop, and resume, allowing users to manage the reading process according to their convenience. By leveraging browser-based speech synthesis, the system ensures real-time performance, portability, and ease of use across different platforms. This approach offers a cost-effective and scalable assistive technology solution that enhances document accessibility, promotes independent learning, and supports inclusive digital interaction for visually impaired individuals.

2. Literature Review / Background

[1] Deep Learning-Based Smart Glass System for the Blind and Visually Impaire By Jinsoo Cho and Mukhriddin Mukhiddinov. This essay discusses the challenges that people with blindness and other visual impairments face in moving around independently and dealing with everyday issues. Artificial intelligence and computer vision techniques provide a solution by enabling blind and visually im- paired (BVI) people to carry out their daily tasks with less reliance on others. For those living in the BVI, smart glasses offer a possible assistive device that can help with independent travel as well as social comfort and safety. Practically speaking, though, the BVI cannot move by themselves, especially at night and in dimly lit environments.

[2] Upgraded Real-Time Fire Alert System for Visually Impaired Individuals Utilizing Cutting-Edge Technologies Written by IEEE members Abdusalomov and Akmalbek Bobomirzaevich. In an emergency where fires occur indoors, this paper discusses early fire detection and notification

approaches that provide blind and visually impaired (BVI) per- sons with timely access to fire prevention and safety information. Sensor-based technologies are used in most conventional ways to identify fire scenarios; how- ever, these technologies are dependent on changes in illumination and ambient conditions. Additional investigation has revealed that, in order to improve fire safety, camera-based fire detection systems produce far better outcomes with high prediction accuracy, low cost, and shorter processing times.

[3] Sensor-Integrated Smart Walking Stick for the Blind by Srinivasan Padmanaban Kamalesh Kumar, Raju Athira, and Ramdas Akshara. There are a lot of visually challenged people in our culture. If you see them, you will know that they require assistance to walk; they cannot go where they are going on their own. In their daily lives, they encounter numerous challenges. Despite the speed at which technology is developing today, there isn't an accessible gadget for those who are visually impaired. Because it is difficult for blind people to carry out their everyday responsibilities, a Smart Blind Stick was created to make movement and task completion easier. But it's extremely risky for visually challenged persons to go on the road since they have trouble seeing potential hazards. Among the greatest tools for pointing is a smart stick. This stick has two ultrasonic sensors that can detect impediments up to four meters away from the user, as well as infrared sensors that can identify stair cases.

[4] Ani R, Effy Maria, J. Jamia Joyce, and Sakkaravarthy V proposed a smart spectacle system for visually impaired individuals. The system captures text using a camera and converts it into voice output, allowing users to hear printed text. It uses a Raspberry Pi as the main controller to connect the camera, sensors, and image processing unit, and also manages peripheral devices such as keyboards and USB components. This solution helps visually impaired users access printed information in an audible form.

[5] Joao Guerreiro and Daniel Goncalves discussed how portable digital imaging devices have advanced in recent years and can help visually impaired users access digital information. They proposed the use of screen readers to convert on-screen content into speech. The authors introduced the concept of concurrent speech, which allows users to listen to multiple pieces of information at the same time and focus on the most relevant content.

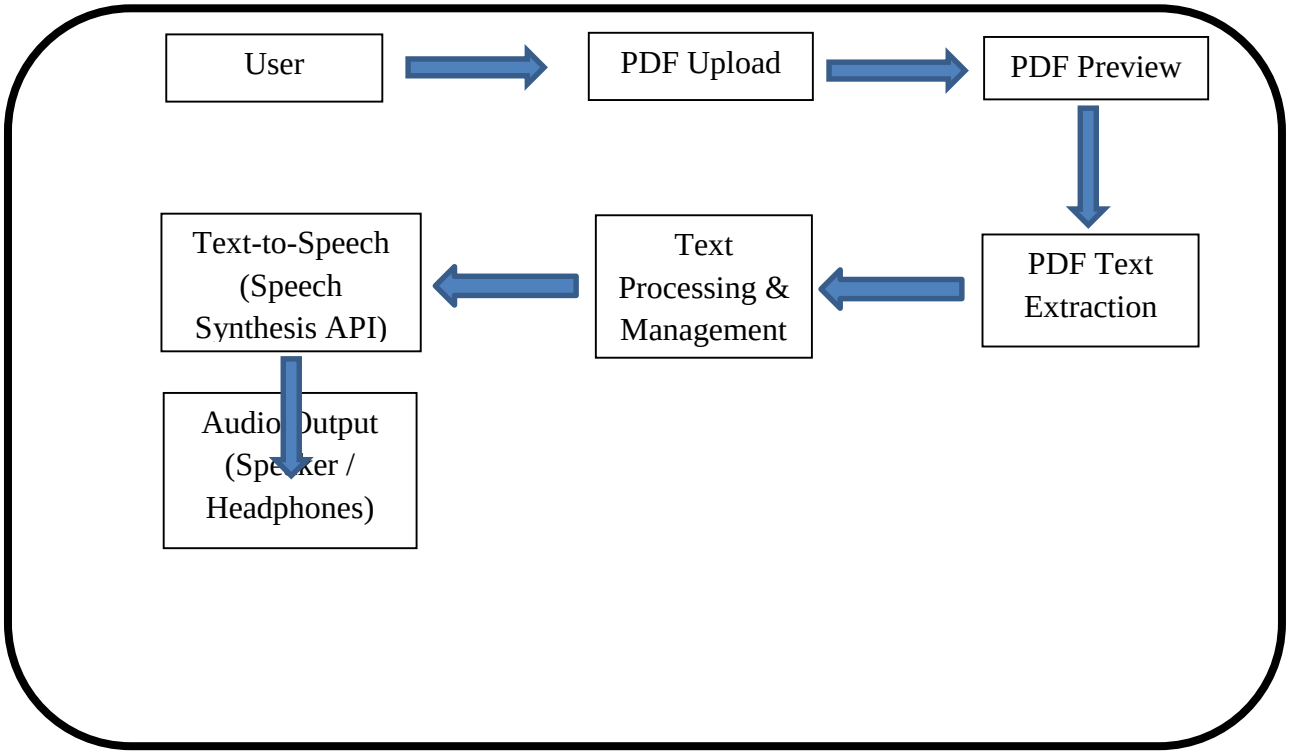
[6] J. Liang and D. Doremann explained that advanced technologies can help visually impaired people understand their surroundings using mobile computer vision. Their work discusses different application areas, technical challenges, and solutions related to document analysis using digital devices and image processing. They also described some example applications and suggested ideas for future improvements.

[7] Alexandre Trilla and Francesc Alias studied improvements in Text-to-Speech (TTS) systems. Their research focuses on analyzing input text features to produce more natural and clear speech output.

[8] S. Mascaro and H. H. Asada presented research at a robot tutor conference on measuring finger posture and force during initial experiments. Their work mainly supports the development of human–machine interaction systems.

[9] Pitrelli J. and Bakis R. proposed the IBM expressive Text-to-Speech system for American English. Their system aims to generate more natural and expressive speech and is based on advanced audio, speech, and language processing techniques.

3. BLOCK DIAGRAMS



The block diagram of the proposed web-based PDF-to-speech reading system illustrates the overall working process and interaction between different system modules. The process begins with the **User Interface**, which is a web-based application designed to be simple and accessible for visually impaired users. Through this interface, the user uploads a PDF document and interacts with the system using basic control buttons such as upload, read, stop, and resume.

Once the PDF is uploaded, it is passed to the **PDF Upload Module**, which handles file selection and validation. The uploaded document is then displayed in the **PDF Preview Module** using an embedded viewer, allowing the user or a helper to confirm the selected document. Simultaneously, the PDF file is forwarded to the **PDF Text Extraction Module**, where the textual content is extracted using PDF processing libraries. This module reads the document page by page and converts the content into machine-readable text.

The extracted text is then sent to the **Text Processing and Management Module**, which organizes the text in a continuous and readable format. This module also manages the reading position, enabling features such as pause, stop, and resume during speech output. After processing, the text is passed to the **Text-to-Speech** synthesis technology.

Finally, the generated speech is delivered to the user through the **Audio Output Module**, which includes speakers or headphones. This allows visually impaired users to listen to the document content clearly and independently. By integrating all these modules in a web-based environment, the system provides an

efficient, cost-effective, and user-friendly assistive solution that enhances document accessibility and supports inclusive digital interaction.

SOFTWARE REQUIREMENTS

- **Operating System:** Windows 10 / Linux / macOS
- **Web Browser:** Google Chrome, Mozilla Firefox, or Microsoft Edge (latest version recommended)
- **Programming Languages:** HTML, CSS, JavaScript
- **PDF Processing Library:** PDF.js
- **Speech Synthesis Technology:** Web Speech API (Text-to-Speech)
- **Development Tools:** Visual Studio Code or any modern code editor
- **Web Server (Optional):** Local host server such as XAMPP or Live Server extension

ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- Provides easy access to PDF documents for visually impaired users through voice output.
- Enables independent reading without relying on external assistance.
- Web-based implementation eliminates the need for specialized hardware.
- Supports interactive controls such as play, stop, and resume for better user experience.
- Cost-effective and easy to use on multiple platforms and devices.
- Enhances accessibility in education and daily information access.
- Promotes inclusive digital interaction and improves user confidence.

DISADVANTAGES

- The system depends on the accuracy of text extraction from PDF files.
- Scanned or image-based PDFs may not be read correctly without OCR support.
- Speech quality depends on the browser's Text-to-Speech engine.
- Requires an internet connection for loading libraries and web-based resources.
- Limited support for complex document layouts such as tables and formulas.
- Performance may vary across different browsers and devices.

4. Materials and Methods

1. Voice Input Processing

The system continuously listens for a predefined wake word. Once activated, the microphone captures the user's speech, which is converted into text using a speech-to-text engine. This allows accurate recognition even with slow or unclear speech patterns.

2. Natural Language Understanding

The converted text is processed using Natural Language Processing (NLP) techniques to identify user intent and relevant keywords. Based on the detected intent, the system determines the appropriate action, such as reading text, controlling a device, or sending an alert.

3. Action Execution

After intent classification, the system performs the requested operation. This may include retrieving information, controlling smart devices, recognizing objects through the camera, or activating emergency services.

4. Audio Feedback Generation

The response generated by the system is converted into speech using a text-to-speech engine and delivered to the user through speakers or earphones, ensuring a fully screen-free experience.

5. Vision-Based Assistance (Optional)

When equipped with a camera, the assistant uses computer vision techniques to identify objects, read printed text, recognize currency, and assist with navigation. OCR technology converts detected text into audio output.

6. Safety and Emergency Handling

The system includes an emergency module that can be activated through voice commands. In critical situations, it sends alerts along with real-time location details to predefined contacts or emergency services.

5. Results

The developed PDF Voice Reading system successfully provides an accessible and user-friendly solution for converting PDF documents into audible speech. The system allows users to upload a PDF file through a simple graphical interface and preview the document on the screen. Once uploaded, the text content of the PDF is accurately extracted and converted into natural-sounding speech using text-to-speech technology.

The application includes essential voice control features such as **Read PDF**, **Stop Reading**, and **Resume Reading**, enabling users to control audio playback easily. The system responds correctly to user actions, displaying real-time status updates such as “Waiting for PDF upload” or reading progress feedback. The PDF preview panel helps users verify the uploaded document before initiating voice output.

The interface design is visually clear with high-contrast buttons, making it suitable for visually impaired users, while hands-free audio output supports users with physical disabilities. The system performs efficiently for standard text-based PDFs and maintains smooth speech playback without noticeable delays. Overall, the project achieves its objective of providing an assistive technology tool that enhances document accessibility, improves information consumption, and supports inclusive digital interaction for blind and physically challenged users.

6. Discussion

The proposed system can be further enhanced by integrating Optical Character Recognition (OCR) to support scanned and image-based PDFs. Voice command control can be added to enable fully hands-free operation. Support for multiple languages and improved speech naturalness can also be incorporated. Additionally, mobile application support and cloud-based processing can be explored to increase accessibility and scalability.

7. Conclusions

The proposed web-based PDF-to-speech system provides an effective assistive solution for visually impaired users by converting document text into audible speech. It enhances accessibility, supports independent reading, and offers a cost-effective, user-friendly platform for inclusive information access.

References

1. Ani R, Effy Maria, J Jameema Joyce, Sakkaravarthy V, "Smart Specs: Voice assisted Text Reading system for Visually Impaired Persons Using TTS method", IEEE International Conference on Innovations in Green Energy and Healthcare Technologies.
2. Joao Guerreior and Daniel Goncalves. Text-to-Speech: Evaluating the Perception of Concurrent Speech by Blind People, International journal technology.
3. J. Liang D. and Doermann H. Camera based analysis of text and documents: a survey, International Journal on Document Analysis and Recognition.
4. Alexandre Trilla and Francesc Alias. (2013). "Sentence Based Sentiment Analysis for Expressive Text-to-Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, Issue. 2. pp. 223-233.
5. S. Mascaro. And H. H. Asada. "Finger posture and shear force measurement: Initial experimentation," in Proc. IEEE Int. conf. robot.
6. Pitrelli J. and Bakis R. "The IBM expressive text-to-speech synthesis system for American English", IEEE Trans. Audio, Speech, Lang. Process.
7. Kwon et al. "A study on the Development of Voice-Enabled Text Reading System for Visually Impaired Persons". (2017)
8. Thakur et al. "Design and Development of an Assistive Text Reading System for the Blind and Visually Impaired" by Thakur et al. (2018)
9. Vijayarani, S., & Saranya, P. (2018). Voice-Assisted Reading System for Visually Impaired. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (pp. 1323-1326). IEEE.
10. "Design and Evaluation of a Voice Activated Text Reader for Visually Impaired Individuals". (2019)
11. Sharma et al. "Text to Speech for Visually Impaired: A Survey". (2019)
12. Singh et al. "Voice-based Text Reader for Visually Impaired People". (2019)
13. Mishra, A., Dubey, A., & Shukla, R. K. (2019). Development of Voice Assisted Reading System for Visually Impaired People. In Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (pp. 65-71). Springer.
14. Xie et al. "Assistive Technology for Blind and Visually Impaired People: An Overview of Technology". 2020
15. Bokhari, M. B., Iqbal, M., Anjum, F., & Rauf, F. (2020). Design and Implementation of Voice-Operated Text Reading System for Visually Impaired. In Proceedings of the International Conference on Advanced Communication Technologies and Networking (pp. 435 442). Springer
16. Bhattacharya, A., Ghosh, S., & Saha, S. K. (2020). A Review of Speech Synthesis Techniques for Text-to-Speech Conversion in Assistive Technology. International Journal Technology, 23(2), 263-281.