IJAREEIE

e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

Volume 14, Issue 4, April 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.807

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 4, April 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404014 |

Text-to-Speech Reader for Visually Impaired

Dr.S.Maheswari, S.Tejashree, V.S.Dharun, S.Udhaya Krishnan, M.Tharani Kumar,

Dr.N.Saravanakumar

Department of EEE, Kongu Engineering College, Perundurai, Tamil Nadu, India

Associate Professor, Department of ECE, Mahendra Institute of Technology, Namakkal, India

ABSTRACT: In the digital era, researchers are developing assistive devices to help visually impaired individuals access information. This paper proposes a Text-to-Speech (TTS) and Object Detection System using a Raspberry Pi. It integrates Tesseract OCR for text recognition and YOLOv8 for object detection. Google Text-to-Speech (gTTS) converts extracted text and identified objects into audible speech. The system utilizes OpenCV for image processing and is implemented in Python. With a user-friendly interface and minimal hardware, it ensures real-time processing. Designed for accessibility and portability, it enhances independence for visually impaired users. This solution promotes inclusivity, making navigation and information access easier in diverse environments.

KEYWORDS: Text-to-Speech, Optical Character Recognition, YOLOv8.

I. INTRODUCTION

Access to information is crucial for everyone, yet individuals who are blind or visually impaired often encounter significant barriers when printed materials are unavailable in Braille or require advanced literacy to interpret. This paper proposes a Text-To-Speech (TTS) system that integrates Optical Character Recognition (OCR) and Raspberry Pi technology to convert printed text into natural-sounding speech, thereby empowering visually impaired users to access literature independently. Enhancing its functionality further, the system incorporates YOLOv8 for object detection, which enables real-time identification of objects in the user's surroundings. This dual capability not only facilitates the reading of printed text but also assists users in navigating their environment by recognizing obstacles and identifying important objects. With an emphasis on real-time processing, portability, and user-friendly operation, this innovative approach helps break down the barriers to information access and promotes greater independence for individuals with visual impairments.

There are several research papers have been published about the text to speech system, Théophile K. Dagba and Charbel Boco, proposed a [2] study on creating a TTS system for the Fon language, using the Multisyn algorithm. This system focuses on converting written Fon text into natural-sounding speech. The paper explains key components like NLP for phonetic transcription and digital signal processing for generating speech. The authors describe building the system in Java and using the festival speech synthesis platform. They also discuss the challenges of developing this TTS system for a tonal language like Fon. J.P. Olive and M.Y. Liberman, proposed a [1] review of text-to-speech synthesizers, detailing the conversion of text into spoken words. They discussed key steps, such as transforming text into phonological structures and sound, along with challenges in system integration. Dongmei Li proposed a [3] study on an English text-to-speech conversion algorithm based on machine learning. This study has addressed the problem of feature recognition in speech during conversion, improved efficiency by modifying rhythm with PSOLA and handling polyphone pronunciation with the C4.5 algorithm, and it has developed a system model to evaluate performance with part-of-speech rules and HMM-based prosody prediction. The algorithm is shown to be performing well and with practical applications. Dutoit proposed a [4] comparative study of four speech models for high-quality TTS systems. The paper analyzes the models in terms of database compression, computational load, intelligibility, and segment quality. The models are the classical auto-regressive, hybrid harmonic/stochastic, the TD-PSOLA algorithm, and the multi-band re-synthesis PSOLA model. The study overcomes prosody matching and segment concatenation challenges by providing insight into the performance of each algorithm.

M. Smith and R. Adams [8] propose a framework for multilingual TTS systems by developing synthetic voices from scratch, voice development without new acoustic data, and synthesizing new languages with minimal data. It evaluates unit selection synthesis and explores techniques to build high-quality multilingual voices efficiently. Nazir and Malik proposed [7] a comprehensive review of deep learning-based end-to-end text-to-speech (TTS) synthesis systems. It includes advancements driven by deep learning, datasets that are being used for training TTS models, comparison between traditional and modern approaches, and has underlined the success of end-to-end models in attaining high

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 4, April 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404014 |

quality results with Mean Opinion Scores. Reddy, Vaishnavi, and Kumar propose [5] a comprehensive review of speech-to-text and text-to-speech systems, highlighting deep learning techniques such as CNNs, RNNs, and transformers. This review would involve applications, challenges, and future directions; thus, it would bring emphasis on advancements, multimodal integration, and personalized speech synthesis.

Text-To-Speech (TTS) readers convert written text into spoken words, enabling visually impaired individuals to access textual content effortlessly. These systems use advanced algorithms, natural language processing, and Optical Character Recognition (OCR) to read digital and printed materials. Integrating object detection with YOLOv8, modern TTS readers can now identify and describe objects in real-time, enhancing navigation and accessibility. For speech synthesis, the system utilizes Google Text-to-Speech (gTTS) to generate natural-sounding audio output. Features like customizable voice options, speed controls, and multilingual support make them versatile and user-friendly. Widely used in audiobooks, language learning, and assistive technology, TTS readers combined with YOLOv8 for object detection and gTTS for speech conversion empower visually impaired users with greater independence in both reading and daily life.

II. PROPOSED TTS MODEL

The proposed method enhances Text-to-Speech (TTS) systems for visually impaired individuals by integrating object detection alongside text recognition. A Raspberry Pi-based system with a push-button interface activates a camera to capture images of printed text and surrounding objects. Optical Character Recognition (OCR) extracts textual content, while an object detection model, utilizing YOLOv8, identifies key objects in the environment. The extracted text is converted into speech using the Google Text-to-Speech (gTTS) module, ensuring clear and natural voice output. This dual-functionality system enhances accessibility by enabling users to comprehend both written information and their surroundings more effectively.

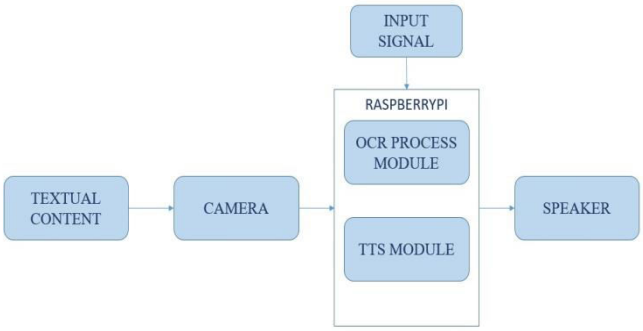


Fig.1: Proposed TTS

The working process begins when the user presses a button, triggering the camera to capture an image. The system first runs OCR to extract any textual content and then applies object detection to identify important objects. The recognized text is processed through gTTS for speech synthesis, while detected objects are named and described using pre-trained AI models. The audio output is delivered through speakers or headphones, providing real-time assistance. This approach improves independence and safety, allowing visually impaired individuals to navigate their environment while accessing both textual and contextual information.

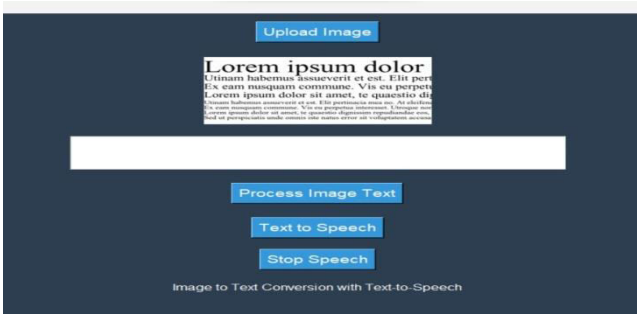


Fig.2 : Simulational output results of the proposed model

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 14, Issue 4, April 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404014 |

III. RESULTS AND DISCUSSION

The Text-To-Speech (TTS) Reader with Object Detection for the Visually Impaired enhances accessibility by enabling users to comprehend printed text and identify surrounding objects. Built on a Raspberry Pi, this portable system integrates Tesseract OCR for text recognition and YOLOv8 for real-time object detection. A camera module captures images, and the extracted text, along with identified objects, is converted into speech using Google Text-to-Speech (gTTS). With minimal user interaction—a single button press, as shown in Fig. 3(a) the system delivers audio feedback through headphones, allowing users to stay aware of their surroundings. Developed using Python, OpenCV, and TensorFlow, this solution enhances independence, making printed materials and environmental context more accessible for visually impaired individuals.

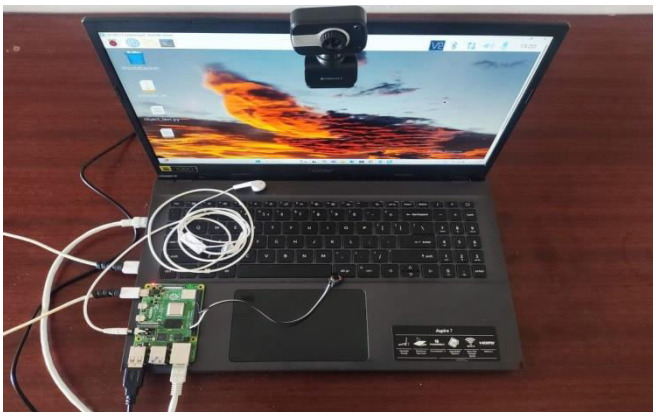


Fig.3. (a) Hardware implication

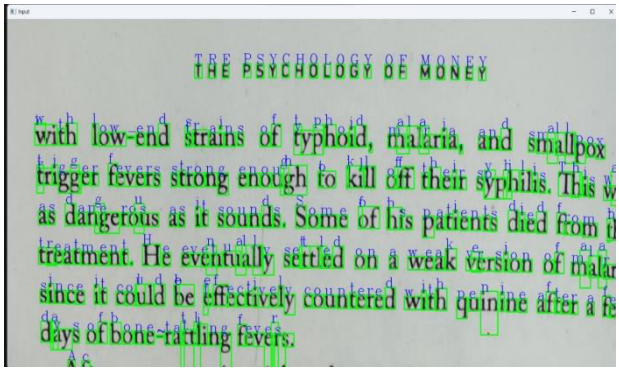


Fig.3 (b) Original text and detected image

The Graphical User Interface (GUI) was designed for simplicity to cater to visually impaired users, featuring large, accessible buttons and voice feedback to guide them through capturing text and reading it aloud. It responded quickly to user inputs, ensuring a smooth experience without delays. Accessibility features included high contrast visuals and large fonts for users with partial vision, along with customization options for the Text-To-Speech output, allowing adjustments to speech rate, volume, and pitch. The GUI seamlessly integrated with the hardware, triggering the camera and initiating OCR and TTS processes without noticeable lag, providing a streamlined interaction for users. The process is done by GUI is shown in Fig.3(b).

IV. CONCLUSION

In conclusion, the proposed enhancements to tts systems will significantly improve accessibility for visually impaired individuals by integrating advanced voice synthesis, expressive speech modulation, and expanded language support. By enhancing speech naturalness and incorporating emotional modulation, users will experience a more engaging and human-like auditory experience, reducing cognitive strain and making long reading sessions more comfortable. The expansion of language and dialect support ensures that the system is inclusive and accessible to a diverse audience,

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.807 | A Monthly Peer Reviewed & Refereed Journal |

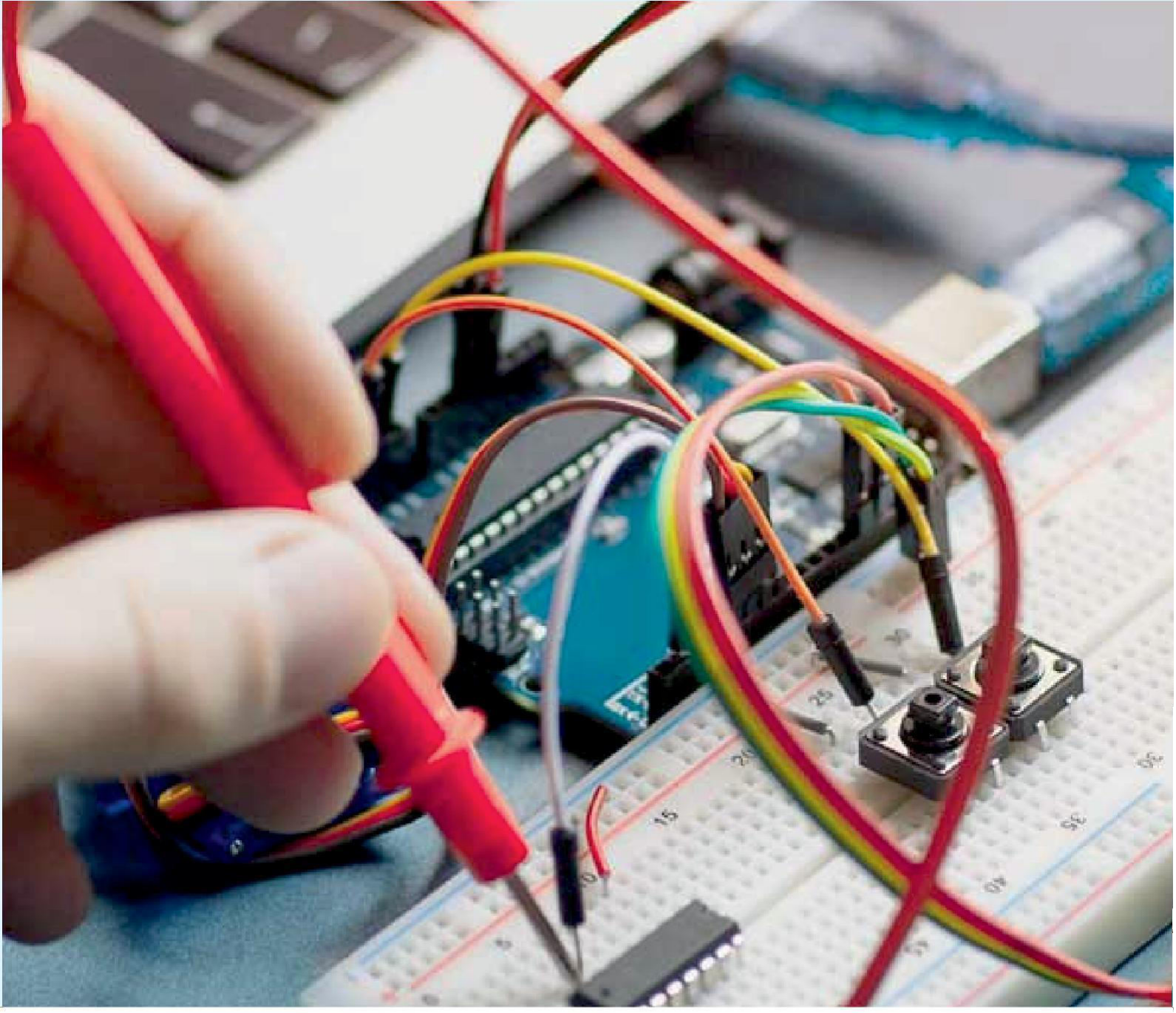
|| Volume 14, Issue 4, April 2025 ||

| DOI:10.15662/IJAREEIE.2025.1404014 |

allowing users to interact with digital content in a way that feels intuitive and familiar. By incorporating superior voice synthesis for lesser-known languages and regional accents, the system will break language barriers and provide a more personalized experience for users from different linguistic backgrounds. These advancements will empower visually impaired individuals to navigate digital content effortlessly, improving their ability to access information, engage with educational materials, and enhance their overall independence in daily life.

REFERENCES

1. Olive, Joseph P., and Mark Y. Liberman. "Text to speech—An overview." *The Journal of the Acoustical Society of America*, vol. 5, no. 9, pp. 10–14, 2015.
2. Dagba, Theophile K., and Charbel Boco. "A Text to Speech system for Fon language using Multisyn algorithm." *Procedia Computer Science*, vol. 40, no. 4, pp. 225–266, 2002.
3. Dongmei, Li. "Design of English text-to-speech conversion algorithm based on machine learning." *Journal of Intelligent & Fuzzy Systems*, vol. 22, no. 15, pp. 201–209, 2007.
4. Dutoit, Thierry. "High quality text-to-speech synthesis: A comparison of four candidate algorithms." *Proceedings of ICASSP'94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 9, no. 4, pp. 52–89, 2019.
5. V. M. Reddy, T. Vaishnavi and K. P. Kumar, "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning," *Proceedings IEEE Conference Acoustics, Speech, and Signal Processing*, pp. 1–4, 2024.
6. D. Bigioi and P. Corcoran, "Challenges for Edge-AI Implementations of Text-To-Speech Synthesis," *IEEE International Conference on Embedded Systems*, pp. 45–50, 2021.
7. Nazir, Owais, and Aruna Malik, "Deep Learning End to End *Speech* Synthesis: A Review," *IEEE International Conference on Artificial Intelligence*, pp. 120–126, 2021.
8. M. Smith and R. Adams, "Multilingual Text-to-Speech Synthesis," *Proceedings IEEE International Symp. Signal Processing*,vol.3, pp. 156–161, 2004.
9. C. Miao, Q. Zhu, M. Chen, J. Ma, S. Wang and J. Xiao, "EfficientTTS 2: Variational End-to-End Text-to-Speech Synthesis and Voice Conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1650-1661, 2024.
10. Z. Yin, "An Overview of Speech Synthesis Technology," *IEEE International Conference on Innovations in Information Technology*, pp. 95–100, 2020.
11. M. Hamed and Z. Lachiri, "Expressivity Transfer In Transformer-Based Text-To-Speech Synthesis,"*IEEE 7th International Conference on Advanced Technologies*, Signal and Image Processing,vol.1, pp. 443–448, 2024.
12. H. Kim and S. Lee, "Advances in Neural Text-To-Speech Models," *IEEE Transaction Neural Networks*, vol. 32, no. 5, pp. 1321–1332, 2023.
13. K. Singh, "Exploring Robustness in Neural TTS for Noisy Environments," *IEEE Conference Acoustics and Signal Processing*, vol. 7, no. 1, 2020.
14. C. Wang et al., "Real-Time Text-To-Speech Synthesis Using Parallel WaveGAN," *Proceedings IEEE International Conference Multimedia Expo*, pp. 389–394, 2022.
15. A. Bose and K. Jain, "Cross-Lingual Adaptation for Neural Speech Synthesis," *IEEE Transaction Speech Audio Process*., vol. 30, pp. 512–523, 2022.
16. J.Rorberts and M.Daniels, "Transformers in speech synthesis:state of the art", *IEEE TransactionMachine Learning*, vol. 29,no.3, pp. 512–518, 2022.
17. A.Nguyen and T. Le, "End-to-End Systems for High-Quality Speech Synthesis," *Proceedings IEEE Conference Artificial Intelligence Applications*, pp. 64–69, 2023.
18. M. Gupta and R. Verma, "Interactive Applications of Text-to-Speech Systems," *IEEE Conference Human-Computer Interaction*, pp. 202–208, 2021.
19. E. Harper, "Expressive Speech Synthesis Using GAN Models," I*EEE International Conference Machine Learning in Speech Processing*, vol. 1, pp. 200–205, 2023.B. Wright,
20. "Improvements in Prosody for Text-to-Speech Synthesis," *IEEE Transaction Signal Processing*, vol. 27, pp.345-352, 2023.



INNO SPACE
SJIF Scientific Journal Impact Factor

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

9940 572 462 6381 907 438 ijareeie@gmail.com

Scan to save the contact details