E-Vision

Yashaswini K
Dept. of CSE
Global Academy of technology
Bengaluru, India

Sharanya Murthy
Dept. of CSE
Global Academy of technology
Bengaluru,India

Vibha K Chitravvara
Dept. of CSE
Global Academy of Technology
Bengaluru, India

Sahana N
Dept. of CSE
Global Academy of Technology
Bengaluru, India

Vedica
Dept. of CSE
Global Academy of technology
Bengaluru, India

Abstract— **The technology known as text-to-speech (TTS) has the ability to read any given text out loud, whether it is typed by a user or processed by an optical character recognition (OCR) system. In the field of text-to-speech technology, thre are several systems available that can translate typical English text into speech. The purpose of this study is to examine the use of optical character recognition in conjunction with text-to-speech technology, and to develop an image-to-speech conversion system that is easy to use and cost-effective using MATLAB. This article describes the OCR approach employed to recognize uppercase letters from A to Z and numbers from 0 to 9. The recognition process is rapid, and the identified characters are saved as text.**

Keywords: text-to-speech, TTS, computer technology, optical character recognition, OCR, speech synthesis, image-to-speech conversion, character recognition, numbers, and text.

I. INTRODUCTION

Speech synthesis is the process of creating artificial human speech. A speech synthesizer is a computer-based system that can be integrated with hardware or software to generate speech. Text-to-speech systems (TTS) translate plain text into speech, while other systems translate symbolic representations of language, such as phonetic symbols, into speech. The TTS protocol converts stored voice data or text into speech, and is commonly used in audio playback devices for people with visual impairments. However, in recent years, text-to-speech technology has extended beyond the disability community, becoming a useful addition to voice mail and answering systems that rely on digital recordings. Advances in text-to-speech technology have also been made.

II. RELATED MACHINE LEARNING APPROACHES

The proposed system consists of two main components: the image processing module and the speech processing module. The image processing module captures images using a camera and converts them into text. The speech processing module then converts the text into sound and loops it with clear quality to ensure it can be perceived. Optical Character Recognition (OCR) is a technology that recognizes characters automatically through an optical mechanism, mimicking the capabilities of human vision. The image processing module uses OCR to convert .jpg files to .txt files, which are then processed by the speech processing module to generate speech. To improve image recognition accuracy, images are first converted to binary before being processed by OCR.

A. OPTICAL CHARACTER RECOGNITION

In 2014, a paper proposed a system consisting of two main modules: an image processing module and a speech processing module. The image processing module utilizes a camera to capture an image and then converts it into text. The speech processing module then converts the text into sound with defined characteristics and plays it in a loop. Optical Character Recognition (OCR) is a technology that automatically recognizes characters through an optical mechanism, which reduces the reliance on human vision. OCR involves several steps, including optical image acquisition, localization and segmentation, pre-processing, feature extraction, classification, and post-processing, ultimately resulting in the conversion of the recognized text to the desired format, such as RTF, TXT, or PDF.

OCR technology works by analyzing the image of a document and identifying the individual characters that make up the text. OCR software then converts these characters into machine-readable text using advanced algorithms and machine learning techniques. OCR software can recognize and process different types of fonts, sizes, and styles of text, as well as handwritten text in some cases.

OCR technology has come a long way since its inception in the mid-20th century. Early OCR systems were limited in their ability to accurately recognize and process text due to limitations in computing power and image processing capabilities. However, with advancements in technology, OCR has become more accurate, faster, and capable of recognizing a wider range of characters and languages.

Today, OCR is an essential technology for businesses and organizations looking to digitize and streamline their document management processes. OCR software is available as both standalone applications and integrated solutions that can be used with other document management tools. OCR technology is also becoming more accessible to individuals,

with many mobile scanning apps now offering OCR functionality.

B. TEXT TO SPEECH

The proposal made by professors at Mandalay University of Technology in 2014 suggested that text-to-speech (TTS) technology can be utilized to enable a computer system to read any text aloud. TTS is an artificial method of creating human speech that involves the use of sophisticated algorithms to produce understandable and natural output. Natural language processing techniques are commonly used in TTS synthesis, and the first step in the process involves processing the text to be synthesized. There are several methods available for generating synthetic speech.

a)A mobile application called "Intelligent Eye" was developed in April 2020 to assist visually impaired individuals with a set of useful features, such as light detection, color detection, object recognition, and banknote recognition. The app achieved approximately 85% accuracy in its features.

b) In May 2014, a paper was published on developing a text-to-speech synthesizer in English for the disabled and hard of hearing. The focus was on creating an ideal combination of human behavior and computer applications to establish a one-way interactive medium between computers and users.

In 2020, a team of researchers from Nepal Institute of Technology, Palmar, Manipal, including Prof. Sriraksha Nayak and Prof. Chandrakala C B, proposed, implemented and validated a system [3] to assist blind people with audio-based instructions on sensitivity and specificity. The system includes email-supporting features, speech-to-text conversion, email sending and reading support by converting text to speech, and creating audio from cash transaction SMS provided as part of the Inbox messaging service.

The system uses the Text to Speech class to convert text to speech, which takes three parameters. The first parameter is the text to say, and the second is a constant that can be Queue Flush or Queue Add. Queue Flush will remove all previous entries from the play queue and replace them with new entries. In queue addition, a new entry is added to the end of the previously read queue. The Text to Speech class adopts many methods, and the method used in the project is set Language, which is used to select the language. Error methods are used to find errors.

The Stop method represents the client's stop request. When the user speaks the destination address, the location manager class will get the destination location from the GPS. This will help blind people find their own way in outdoor spaces.

In 2020, a group of computer science researchers at Universiti Malaysia Computer Science and Engineering University, including Ismail Sahak and Ong Huey Fang, found that most visually impaired people cannot recognize new objects even by touching them. This means that if the

packaging of a product they buy changes, they may not be able to identify the product without someone telling them what it is.

Moreover, blind people may have difficulty identifying objects nearby or even their surroundings or location. They rely on other senses such as hearing and smell, which can be challenging in a new environment. Respondents in the study agreed that they would prefer to have knowledge of their surroundings without having to travel.

A group of researchers from the Department of Electronic Technology at the Faculty of Engineering Minuto de Dios-UNIMINUTO in Bogota, Colombia, Ivey Business School, Colombia, and the Department of Electrical and Computer Engineering at the University of West London, Canada [5] summarized a set of observations through research. They found that people with visual impairments and disabilities used a variety of assistive technologies, including wheelchairs, prostheses, and forearms with manual or gestural control for hearing or speech impairments, and assistants in sign language with gloves or computer vision. They also used everyday life systems with fall detectors, IoT systems for home automation and device control, and genderless wellness devices and gaming apps for people with disabilities.

The researchers found that (n=85, 43.36%) of respondents used Raspberry Pi and its different versions (2, 3, 4, or zero), (n=100, 51.02%) used Arduino boards, and (n=11, 5.61%) used other devices. The distribution of reference designs according to assistive technology type is discussed in RQ2.

Table: Analysis of Methods and results

Sl. No	Paper Title	Method Used	Result
1	Intelligent Eye: A Mobile Application for Assisting Blind People	Light detection, color detection, object recognition and banknote recognition	About 85% accuracy
2	Text-To-Speech Synthesis	Text-to-speech synthesizer in English for the disabled and hard of hearing	One-way interactive medium between computers and users
3	Audio-based instructions for guiding blind people	Text to speech class	System validated with sensitivity and specificity
4	Challenges faced by visually impaired people	Survey	Difficulty recognizing new objects and navigating new environments
5	Assistive technologies for people with disabilities	Various methods and devices	Distribution of reference designs discussed in RQ2

III. RESULTS

Each article underwent a thorough research and analysis process.

•[1] The visually impaired population includes individuals with various vision-related issues such as astigmatism, age-related eye diseases, dyslexia, myopia, and gradual vision loss.

•[2] Electronic text-to-speech conversion was successfully implemented. This method converts text from documents, web pages, or e-books into synthesized speech using computer speakers. For image-to-text conversion, the first step is to convert the image to a grayscale image. Thresholding is then applied to convert grayscale images to binary images, which are then converted to text using MATLAB.

• [3] The mobile application was designed to have a user-friendly interface for the visually impaired. The project's goal is to provide real-time reading assistance to blind individuals by accurately processing images and converting them to clear speech. Iris is used to identify objects and generate descriptive text based on images taken by the camera. The overall results are satisfactory, and the proposed mobile application can assist blind individuals in daily activities, making them more independent.

•[4] The application's features can be enhanced by incorporating lower latency and face detection. With the upcoming 5G technology, applications can access advanced algorithms and image processing services in the cloud, which can provide nearly instantaneous results.

The paper presents a detailed description of the system architecture and the different components involved. The OCR engine used in the system is Tesseract, which is an open-source OCR engine with a high degree of accuracy. The system also uses Python libraries for image processing and text manipulation.The results of experiments conducted to evaluate the performance of the system.

The experiments involved using different types of images with varying degrees of complexity, such as images with different fonts, sizes, and colors. The results show that the system achieved a high degree of accuracy in recognizing characters from these images and converting them to speech.

The paper also discusses the limitations of the system and suggests possible future work to address these limitations. The limitations include the inability of the system to recognize

The text-to-speech industry offers many systems that can convert English text into speech. The purpose of this research is to study the use of OCR with text-to-speech technology and develop an affordable and use r- friendly image- to- speech conversion system using MATLAB. By combining OCR and TTS technologies,we can create a powerful tool that can make it easier for people with visual impairments or reading difficulties to access written information.

handwritten text and the need for further optimization to improve the speed and efficiency of the OCR engine.

IV. CONCLUSIONS

Individuals who have visual impairments such as astigmatism, age-related eye diseases, dyslexia, myopia, and gradual vision loss, among other vision-related problems, make up the population with partial vision. The primary objective of the application is to offer visually impaired individuals a user-friendly interface. The aim of the project is to provide immediate reading aid to the blind by quickly and accurately processing images and converting them into clear speech.

REFERENCES

[1] Milios Awad, Jad El Haddad, Edgar Khneisse. Intelligent Eye: A Mobile Application for Assisting Blind People‖ 2018 IEEE Middle

[2] Nwakanma Ifeanyi, Oluigbo Ikenna and Okpala Izunna. ―Text–To–Speech Synthesis (TTS)‖ IJRIT International Journal of Research in Information Technology Volume 2, Issue 5, May 2014, Pg: 154-163

[3] Nayak, S., & C B, C. (2020). Assistive Mobile Application for Visually Impaired People. International Journal of Interactive Mobile Technologies (iJIM), 14(16), pp. 52–69.

[4] Ismail Sahak, Ong Huey Fang, Syuhada Abdul Rahman. ―Assistive Mobile Application for the Blind‖ Vol 2760-paper4A. Sudha, P. Gayathri, "Effective analysis & predictive model of stroke disease using classification methods", IJCA(0975-8887), Vol.43-No.14, April 2012.

[5] JONATHAN ÁLVAREZ ARIZA(Member, IEEE) & JOSHUA M. PEARCE Low-Cost Assistive Technologies for Disabled People Using Open-Source Hardware andSoftware

[6] R. Shrivastava, R. K. Singh, and P. Kumar, "Development of a real-time image-to-speech conversion system using text-to-speech technology," in Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems, pp. 205-210, 2017.

[7] A. Mohan, S. Vijayakumar, and S. Sathyanarayanan, "Implementation of text-to-speech and optical character recognition systems for visually impaired people," in Proceedings of the 2018 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5, 2018.

[8] N. Pandya, S. Patel, and S. Shah, "Image to speech conversion system using OCR and TTS," in Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 116-120,