

Image Text to Audio Conversion Using Raspberry Pi

Gowri Ch1, Manikanta Y, Lohitha Y, Nagur Babu Sk, Arun Kumar P

Assistant Professor,ECE Department, Godavari Institute Of Engineering and Technology(A), Rajahmundry,India
Students,ECE Department , Godavari Institute Of Engineering and Technology(A), Rajahmundry, India

ABSTRACT— Basically, text files are converted into voice or audio form using text-to-speech (TTS) technology. The TTS technology is intended to assist individuals with disabilities. Ninety-one percent of individuals worldwide are visually impaired. Thus, we must assist the blind. To that end, this study proposes the use of Raspberry Pi for text-to-text-to-speech conversion. It has a webcam to take pictures for text input, which the TTS unit uses. The TS unit is mounted on a Raspberry Pi, and its output is first amplified by an audio amplifier before being sent to a speaker. The hardware components used in the design are a Raspberry Pi 3, a webcam, a 10000 mAh battery bank, and Bluetooth earphones. The TS unit is mounted on a Raspberry Pi, and its output is first amplified by an audio amplifier before being sent to a speaker. By offering a battery backup, the Raspberry Pi makes it possible for it to be carried or relocated with ease. As a result, the user can use the component whenever and wherever they choose. An internet connection is not necessary for this modification. The blind person can use this equipment independently because it is very simple to operate.

Keywords— Python programming, open-source CV libraries, OCR, and TTS services.

INTRODUCTION

Everyday tasks for those who are visually impaired can be extremely difficult, such as reading and obtaining information from printed texts. As a result, technologies that can translate images into voice must be created to facilitate information access for the blind. One potential technique that helps people with visual impairments connect with visual content is image-to-speech conversion. Nevertheless, poor quality or fuzzy photos can reduce this technology's efficacy. A study on picture-to-voice conversion for people with blur detection impairments is presented in this article. This study aims to examine how well various image processing methods identify and manage fuzzy images during the image-to-speech conversion process. The goal of the research is to find image processing methods that can distinguish blur in pictures. Examine how well various blur detection algorithms perform and evaluate how blur identification affects the accuracy of image-to-voice translation. This paper shows a system that uses a Raspberry Pi 3 B, a 5MP Raspberry Pi camera, a speaker, and an amplifier to translate images to voice. To guarantee that the text is readable and clear, the system is built to identify blur in an image before translating it to speech. To begin a new paragraph, use the enter key. The proper indent and spacing are applied automatically. The creative concept and inexpensive method that is utilized to read the text image's contents aloud. It integrates with the Raspberry Pi the concepts of text-to-speech synthesizer (TTS) and optical character recognition (OCR).

LITERATURE REVIEW

[1] Kalyani Shimane [2020] presented a technique that uses a Raspberry Pi to turn text in the paper into speech. It has a camera to take pictures, which are subsequently sent to the TTS unit as input text. Installed in a Raspberry Pi, this TTS unit amplifies its output through the use of an audio amplifier before being sent to the speaker. An internet connection is not necessary.

[2] For visually handicapped people, Jarlin James [2018] suggested a system that would take text from a paper and turn it into audio. Users can hear the text image material by using this device. In this study, they integrated the concepts of Text-to-speech synthesis with OCR (Optical Character Recognition). This method makes it simple for those who are blind or visually impaired to read text from documents, translate it into the language of their choice, and convert edited text to audio by utilizing Google Speech. Long distance can be used to expand this system.

[3] Mahalakshmi [2018] suggested a Raspberry Pi system that included a Bluetooth headset and an HD camera. The text image is captured by the camera, which then turns it into audio and outputs it through a headset. OCR will be used to identify every word, and TTS will be used to hear it. Books and papers with printed content can also be read by it.

[4] In 2017, Ms. Geetha suggested essentially In this project, pressing a button will cause the Raspberry Pi processor to switch on or the USB camera to snap a picture anytime you wish to hear the text in the image. Once a few seconds, the camera focuses, takes a picture, and sends it to the Raspberry Pi. The Raspberry Pi uses open CV libraries to convert Text To Speech once the synthesizer has separated the text from the image and OCR has been applied to identify the characters in the text.

[5] Nikhil Mishra [2017] put forth a system for converting images to speech that may be used to read text and labels on banners of various sizes. A Raspberry Pi and camera modules will be used to help a camera or a Wi-Fi or Bluetooth-based device take pictures. to decipher the text from the picture. Following picture extraction, text is then constrained. The Text to Speech device is fullyportable and can be attached to eyeglasses to quickly record photos and translate text to speech. This allows blind individuals to comprehend text and banners.

[6] In 2015, Pooja Sharmista suggested Being blind is the state of not being able to see objects because of neurological or physiological issues. A straightforward, affordable, and user- friendly virtual eye will be created

and put into use in this suggested study by Pooja Sharma, Mrs. Shimi S. L., and Dr. S. Chatterji to increase the mobility of blind and visually impaired individuals in a particular area. To enable the blind individual to navigate

securely on their own and avoid any obstacles—whether stationary or moving—the proposed work incorporates wearable technology in the form of headgear, a miniature walking stick, and shoes.

[7] The conversion of text to speech was proposed by Chaw Su Thu[2014] et al. This study examines computer-based systems that are capable of reading any text, regardless of how it was scanned into the computer and fed into an OCR system. There are numerous systems available for turning text to speech. For character recognition, the OCR system is employed. The recognized character is saved as text in a notepad file. Next, the computer receives this text file straight as an input. Then use MATLAB to speak via that. But tiny letters are not detectable by this technology.

[8] In today's world, T. Rubesh Kumar's [2014] suggested reading is unquestionably crucial. There is printed language everywhere: bank statements, reports, and receipts. While some systems already exist that show promise for portability, they are unable to handle product labeling. However, a significant drawback is that blind users have a very difficult time locating the bar code's location and pointing the bar code reader in the right direction. A camera-based assistive text reading framework has been presented by T. Rubesh Kumar and C. Purnima to assist blind people in reading text labels separating packaging from portable items they use on a daily basis.

PROPOSED SYSTEM:

The proposed system uses Raspberry Pi, HDMI cable, web camera, mouse, laptop, adopter, and speakers. There are TTS and OCR libraries to convert text to speech and separate optical characters. The main principle is web camera captures text images from any material and converts them into an audio process. The output of the result is delivered by the speaker. The Raspberry Pi board is a credit card-sized mini computer that turns a monitor, TV, and mouse into a full-fledged PC.

- Inbuilt Wi-Fi Module and it doesn't require any internet.
- Control electronic components and explore the IOT.

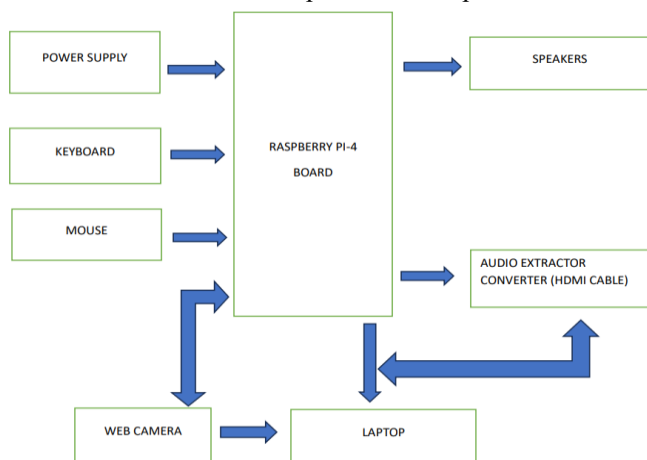


Fig. 1. BLOCK DIAGRAM

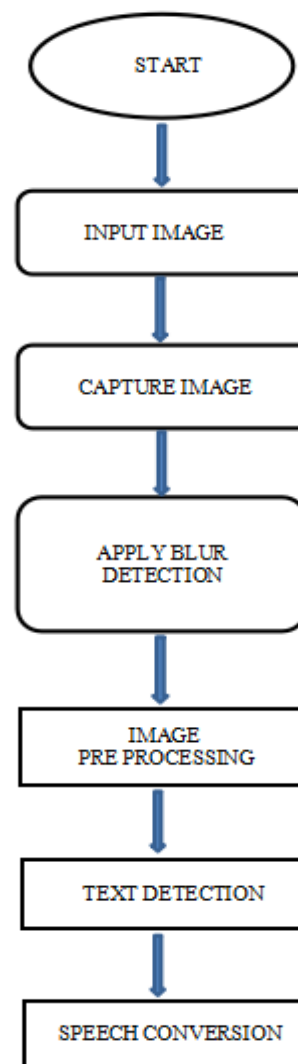


Fig.2.FLOW CHART

OVERVIEW AND TECHNIQUES

The suggested Raspberry Pi image-to-speech conversion system offers a flexible and fluently attainable option because it is made to serve without internet access. The system combines many styles and rudiments to grease the effective processing and restatement of textbooks from picture to audible voice. The camera module of the Raspberry Pi is used by the system to take film land. To ameliorate the quality of the prints that are captured, introductory image pre-processing ways are used. A robust open-source OCR machine called Tesseract OCR is used to directly prize textbooks from prints. The OCR module is designed to work with a variety of languages, typefaces, and image quality situations. The capability to do OCR locally guarantees freedom from online OCR providers. The captured textbook is turned into speech using the open-source Festival TTS technology. To ameliorate the synthetic speech's

lightheartedness and comprehensibility, TTS revision possibilities are delved. Offline functionality is assured by the original TTS module, which functions without taking an internet connection. The Raspberry Pi is equipped with an easy-to-use interface designed to promote stoner involvement. The UI makes it simple for druggies to take film land and start the image- to speech restatement process. Performance criteria for the system include OCR delicacy, conversion speed, and Raspberry Pi resource efficiency. Measures including processing speed, memory consumption, and recognition delicacy are examined to assess how successful the suggested system enterprise. This covers productivity software like LibreOffice and the Chromium web cybersurfer, as well as programming tools like Python, Scratch, and knot-RED. A sizable and vibrant community of Raspberry Pi inventors and suckers supports Raspbian. A wealth of tutorials, forums, and attestation are available for druggies to partake in their systems and ask questions.

COMMUNICATION PROTOCOLS

To enable communication between the various parts of the proposed Raspberry Pi image-to-speech conversion system, a communication protocol is necessary. The picture capture module, OCR module, TTS module, and user interface coordinate and share data according to the communication protocol. An outline of the communication protocols is provided below. The OCR module receives the image that the Raspberry Pi's camera module took to extract the text. For compatibility, a standardized picture format (such as JPEG or PNG) may be utilized. A specified data structure holding picture information may be used, in the data transmission, or a straight file transfer may be used. Following text extraction from the image by the OCR module, the text data is sent to the TTS module for speech synthesis. Text data is communicated by the OCR module in a structured format, either plain text or a particular document format. The user interface receives the TTS module's synthesized voice output and plays it back. The speech data is supplied by the TTS module in a format that can be played back by the user interface's audio playback capabilities. Users can take pictures and start the converting process with the user interface.

REQUIRED COMPONENTS

A) SOFTWARE USED:

RASPBIAN OS:

The Raspberry Pi Foundation is the developer of the Raspberry Pi line of tiny single-board computers, and Raspbian is the operating system (OS) optimized for these devices. Specifically designed to work well on Raspberry Pi hardware is Raspbian. It consists of setups and optimizations to guarantee stable operation on the comparatively low resources of Raspberry Pi boards. Numerous preinstalled applications are included with Raspbian to facilitate a range of activities and projects.

IMAGE PROCESSING LIBRARY:

Software tools or fabrics known as image processing libraries offer several functions and ways of working with digital images. These libraries can be used for a wide range of tasks, from easy bones like re-sizing and cropping to more complex bones like object discovery and pattern recognition. One of the most well-liked and frequently used image-processing libraries are Open CV. It offers a wide range of open-source algorithms for the study of images and videos, identical to object shadowing, point discovery, image filtering, and machine literacy- grounded image recognition. With support for multiple programming languages, alike to C, Python, Java, and MATLAB, Open CV is usable by a broad diapason of inventors. Operations including facial recognition, stoked reality, robotics, and medical imaging make expansive use of it.

SPEECH SYNTHESIS SOFTWARE:

Software that translates written text into spoken language is called speech synthesis software, or text-to-speech (TTS) software. It is frequently utilized in many different applications to enable natural language interfaces, increase accessibility, and enhance user experience. Written material can be entered into speech synthesis software in several formats, including plain text, HTML, and markup languages like SSML (Speech Synthesis Markup Language). The program reads the text input, processes phonetic and linguistic data, and outputs the appropriate speech signals. Neural text-to-speech (NTTS) or deep learning-based TTS are terms used to describe some sophisticated speech synthesis systems that employ machine learning techniques to produce more expressive and natural voices.

B) HARDWARE USED:

RASPBERRY PI MODEL:

The Raspberry Pi Foundation is the developer of the Raspberry Pi line of compact, reasonably priced single-board computers. The Broad Cam System on a Chip (Soc) with an ARM processor that powers Raspberry Pi boards has several onboard components, including RAM, USB ports, HDMI output, Ethernet interface, GPIO pins, and connectors for cameras and displays. Numerous operating systems are supported by Raspberry Pi, including Raspbian (formerly called Raspberry Pi OS), a Debian- based Linux distribution tailored specifically for Raspberry Pi. This covers software libraries and frameworks for a variety of uses, including web development, artificial intelligence, and electronics prototyping. Programming languages included in this are Python, C/C++, and Java. Because it offers an accessible and reasonably priced platform for teaching computer science, electronics, and programming in classrooms and community settings, Raspberry Pi has had a tremendous impact on education. Numerous projects have been made with it, such as robotics projects, media centers, weather stations, vintage game consoles, home automation systems, and more.



Fig. 3. RASPBERRY_PI WEB

CAMERA:

A camera sensor, a lens, and additional parts required for taking pictures and videos make up a web camera. They are frequently small and made to be quickly installed on a tripod, a laptop screen, or computer display. Additionally, a lot of webcams come with built-in microphones for recording sounds. Certain-time text- containing image capture is possible with this webcam. This makes it easy for users to enter printed materials like documents, signage, or other materials into the system so that text may be extracted and turned into audio.

Additionally, users can continuously feed documents or other items into the system for instantaneous audio conversion. This might be especially helpful for people who need immediate access to material that is written on paper or in educational environments. Considerably, adding a webcam to the Raspberry Pi Image Text to Audio Conversion.



Fig. 4. WEB CAMERA

RESULTS:

This system converts the spoken output from text input after processing it. Speakers or headphones that are attached to the Raspberry Pi can play this output. Users can contribute text or text-containing images to the system to engage with it. Following this input, the system ought to process it and provide spoken output appropriately.

Additionally, the system might provide customization choices including the ability to choose from a variety of voices, modify speech characteristics like pitch and velocity of speech, and support several languages.



Fig.5.ASSEMBLING COMPONENTS

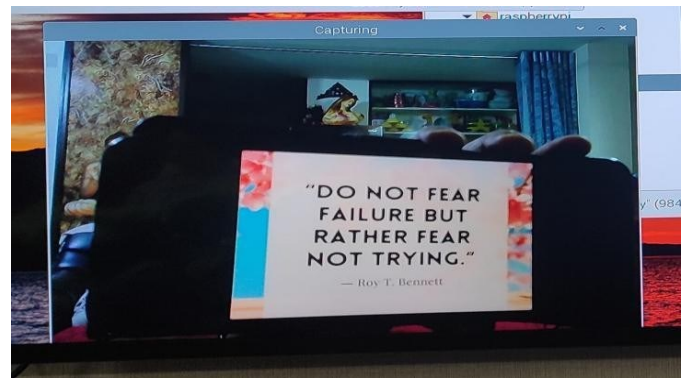


Fig.6.CAPTURING IMAGE FROM MOBILE

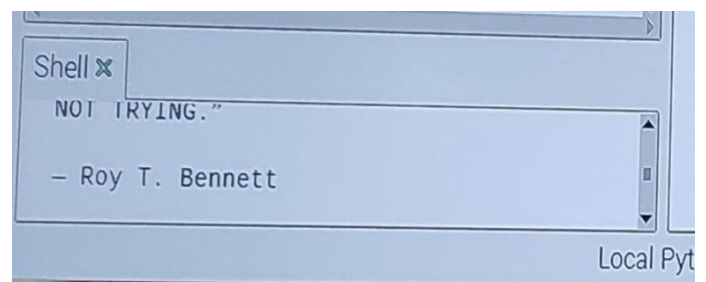
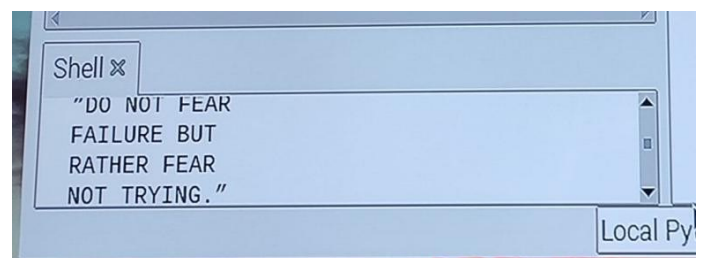
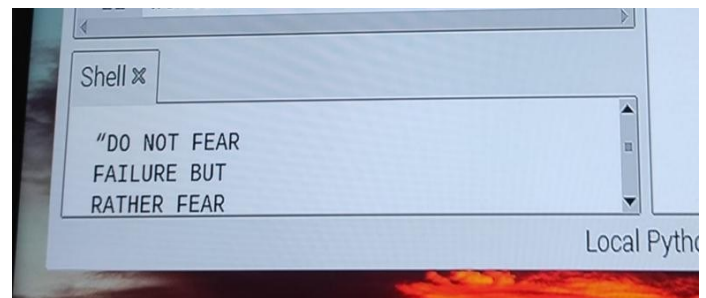


Fig. 7. OUTPUT

CONCLUSION

Thus, this technology will effectively serve the goal of assisting the blind. This gadget aids the blind and visually impaired not just when reading but also when shopping, helping them to read product labels and navigate indoor spaces at home and at college where instructions are written on boards at various locations. This technology is more dependable and accurate and has a moderate total cost of about five thousand rupees. Because it lowers the cost of producing Braille books, this method saves them a significant amount of money. The device compact design and Bluetooth earpiece make it incredibly convenient for blind users to carry about and operate. We can use deep learning to develop the system further by implementing numerous other applications, such as image captioning, however, the Raspberry Pi cannot be used for this purpose due to its High processing speed requirements and power consumption.

REFERENCES

- [1] M. Alave, S. Gore, R. Kamble, V. Phutak, and R. R. Kulkarni (2019). Raspberry Pi Text-to-Speech Conversion. International Journal of CuttingEdge Research and Science Technology.
- [2] M.S. Panse, S. Varde, and A.A. Panchal (May 20– 21, 2016). Character Recognition and Detection System for People with Visual Impairments. Presentation given at the IEEE International Conference on Electronics and Information Communication Technology, held in India recently.
- [3] Department of Information Technology, MNM Jain Engineering college, Chennai, India, "Image Text Speech Conversion in The Desired Language by Translating with Raspberry Pi," H. Rithika and B. Nithya Santhoshi. Reference: 10.1109/ICCIC.2016.7919526.
- [4] "Smart Reader for Visually Impaired Using Raspberry Pi," written by S. Sarkar, G. Pansare, B. Patel, A. Gupta, A. Chauhan, R. Yadav, and N. Battula, in International Conference on Innovations in Mechanical Sciences (ICIMS'21), IOP Conf. Series: Materials Science and Engineering 1132 (2021) 012032, IOP Publishing, 2021. From <https://iopscience.iop.org/article/10.1088/1757-899X/1132/1/012032/pdf>, with permission.
- [5] Picture to Speech Translation for Blind People, A.G. Hagargund, S.V. Thota, M. Bira, and E.F. Shaik (2017). International Journal of Engineering Research Updates and Science (IJLRET), 03(06), 09–15.2454- 5031 ISSN.taken from <https://www.ijlret.com/Papers/Vol-3-issue-6/2- B2017160.pdf>.
- [6] In 2020, A. Siby, A.P. Emmanuel, C. Lawrance, J.M. Jayan, & K. Sebastian. Converting Text to Speech for People with Visual Impairments. International Journal of Innovative Science and Research Technology, ISSN: 2456-2165, IJSRT20APR1045, 1253-1260. Individuals who translate text to speech using visual assistance.
- [7] L.L. Montague, T. Varma, S.S. Madari, and R. S. Poojary (2021). Text extraction from images and conversion of text to speech. 2278-0181 is the ISSN for the International Journal of Engineering Research and Technology (IJERT). Posted by www.ijert.org.
- [8] R.P. Medina, M.C.R. Aragon, and R.A. Pagduan. I Blur Detect: Comparison and Assessment of Image Blur Detection Methods. Quezon City, Philippines: Information Technology Department, Technological Institute of the Philippines. 2020/1030 77/103077 in the papers section of scitepress.org. Innovations in Mechanical Sciences (ICIMS'21), IOP Conf. Series: Materials Science and Engineering 1132(2021)012032, IOP Publishing, 2021. Retrieved from <https://iopscience.iop.org/article/10.1088/1757-899X/1132/1/012032/pdf>.