International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04635
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Implementation of the Image Text to Speech Conversion in the Desired Language by Translating with Raspberry Pi

N. MALLISHWARI, B.Tech III- ECE (226F1A0453),

Department of Electronics and Communication Engineering , Pallavi Engineering College, Survey No.209, Swathi Residency Road KUNTLOOR, Hayathnagar, Kuntloor Village, Hayathnagar_Khalsa, Hyderabad, Telangana 501505.

Dr. K. NAVEEN KUMAR, M.Tech., Ph.D., MISTE., MIAEng., MIETE

Professor & HOD of Department of Electronics and Communication Engineering, , Pallavi Engineering College, Survey No.209, Swathi Residency Road KUNTLOOR, Hayathnagar, Kuntloor Village, Hayathnagar_Khalsa, Hyderabad, Telangana 501505.

Mr.B. LAXMAN, M.Tech

Professor & HOD of Electronics and Communication Engineering, Pallavi Engineering College, Survey No.209, Swathi Residency Road KUNTLOOR, Hayathnagar, Kuntloor Village, Hayathnagar_Khalsa, Hyderabad, Telangana 501505.

B. ARTHISHA, B. Tech III- ECE (226F1A0408),

Department of Electronics and Communication Engineering, Pallavi Engineering College, Survey No.209, Swathi Residency Road KUNTLOOR, Hayathnagar, Kuntloor Village, Hayathnagar_Khalsa, Hyderabad, Telangana 501505.

CH. BHARATH, B. Tech III- ECE (226F1A0412),

Department of Electronics and Communication Engineering, Pallavi Engineering College, Survey No.209, Swathi Residency Road KUNTLOOR, Hayathnagar, Kuntloor Village, Hayathnagar_Khalsa, Hyderabad, Telangana 501505.

M.RAJU, B. Tech III- ECE (236F1A404),

Department of Electronics and Communication Engineering, Pallavi Engineering College, Survey No.209, Swathi Residency Road KUNTLOOR, Hayathnagar, Kuntloor Village, Hayathnagar_Khalsa, Hyderabad, Telangana 501505.

ABSTRACT:

The main problem in communication is language bias between the communicators. This device basically can be used by people who do not know English and want it to be translated to their native language. The novelty component of this research work is the speech output which is available in 53 different languages translated from English. This paper is based on a prototype which helps user to hear the contents of the text images in the desired language. It involves extraction of text from the image and converting the text to translated speech in the user desired language. This is done with Raspberry Pi and a camera module by using the concepts of Tesseract OCR [optical character recognition] engine, Google Speech API [application program interface] which is the Text to speech engine and the Microsoft translator. This relieves the travelers as they can use this device to hear the English text in their own desired language. It can also be used by the visually impaired. This device helps users to hear the images being read in their desired language.

Image Text to Speech (ITTS) conversion is an assistive technology that bridges the gap between visual and auditory information, making printed text accessible to visually impaired individuals. This project focuses on developing a system that captures text from images using a camera module, processes the image to extract text using Optical Character Recognition (OCR), and then converts the recognized text into audible speech using Text-to-Speech (TTS) synthesis. The implementation utilizes a Raspberry Pi as the core processing unit, integrated with a webcam for image capture and a speaker for audio output. The Python-based system employs libraries such as Tesseract OCR for text extraction and pyttsx3 or gTTS for speech generation.

Key Words: Raspberry Pi (any model, e.g., Raspberry Pi **,** Camera Module / USB Webcam**,** Speaker / Audio Output**,** MicroSD Card**,**GPIO (if using external controls), Python, Bash / Shell Script, Linux (Raspberry Pi OS),Open Source Libraries, Automation , Script Crontab (for scheduling tasks)

INTRODUCTION: There are already many systems which read images and give voice output. But this system gives voice output in any language desired by the user. This is done by capturing the image which is to be read using a raspberry pi camera module. Raspberry pi is a credit card sized single board computer. The operating system used is Raspbian. A 15 cm ribbon cable is used to attach the camera module to the raspberry pi. The coding is done using python language. The Optical character recognition engine converts the images of text into machine encoded text and saves it in a text file. Tesseract is the OCR engine which is used for extracting the English text from the image and storing it in a text file. The text to speech engine converts text to speech output. Speak is a speech synthesizer which can easily be used in raspberry pi for speech output in English. For translating it to other languages Google text to speech engine and Microsoft translator is used. Google text to speech is a screen reader which speaks the text on the screen.

The core components of an ITTS system include image acquisition, Optical Character Recognition (OCR), and Text-to-Speech (TTS) synthesis. With the help of a camera or webcam, the system captures images of documents, books, signs, or any printed material. The image is then analyzed using OCR techniques—such as Tesseract OCR—to detect and extract readable text. Finally, the extracted text is converted into audio using TTS engines like pyttsx3 or Google Text-to-Speech (gTTS), enabling the system to "read" the content aloud.

This project leverages a **Raspberry Pi**, a compact and affordable single-board computer, to build a portable and low-power ITTS device. The Raspberry Pi handles the entire workflow—from image capture and processing to audio playback—making it suitable for real-world applications, especially in the domain of assistive technologies.

The development of such systems contributes significantly to digital inclusion, allowing visually impaired users to independently access printed content, thus improving their quality of life and educational opportunities. Additionally, ITTS systems can be useful in multilingual environments, enabling real-time translation and narration of text in various languages.

This project distinguishes itself from existing models that depend on RFID, Bluetooth, or GSM by offering a hybrid solution using GPS, cloud-based control, and mobile integration. The proposed method not only supports real-time decision-making but is also cost-effective, scalable, and suitable for deployment in existing smart city infrastructure.

Image Acquisition:
The process begins with capturing an image that contains textual information. This could be anything from a book page, a signboard, a printed document, or even handwritten notes. A simple USB webcam or a Pi Camera Module attached to a **Raspberry Pi** can serve as the image capture device. The image must be clear and well-lit to ensure optimal text recognition. **Optical Character Recognition (OCR)**:
After the image is captured, OCR technology is applied to extract the text. **Tesseract OCR**, an open-source OCR engine developed by Google, is commonly used for this purpose. It scans the image pixel by pixel, identifies character patterns, and converts them into machine-encoded text. OCR can process multiple languages and various fonts, and recent versions are also capable of handling skewed or distorted text to some extent.

Text-to-Speech (TTS) Conversion:
Once the text is extracted, it is converted into audio using a Text-to-Speech engine such as **pyttsx3**, **gTTS**, or **espeak**. The engine processes the string of text and synthesizes it into a natural-sounding voice that is output through a speaker or headphone. This allows users to listen to the content, providing an auditory alternative to reading.

Using a **Raspberry Pi** as the core processing unit makes the system highly cost-effective, compact, and suitable for portable or embedded use. The Raspberry Pi can run a lightweight Linux operating system, and with Python scripting, it can automate the entire workflow. Since the Raspberry Pi also supports various input/output peripherals, it can be connected with buttons for user control or even integrated with Braille displays or mobile apps.

This project has a wide range of real-world applications:

- **Assistive devices for the visually impaired**
- **Educational tools for children or people with reading difficulties**

Existing and Proposed Methodology:

Existing Methodology: The concept of converting text from images to speech is not entirely new. Several systems, tools, and mobile applications already exist that use a similar methodology involving image capture, Optical Character Recognition (OCR), and Text-to-Speech (TTS) conversion. The existing solutions vary in complexity, cost, platform, and accuracy. Here's an overview of the commonly used methodologies and systems:
Mobile-Based Applications

Many smartphone applications use built-in cameras and cloud-based OCR and TTS services to read text aloud. Examples include:

- **Seeing AI (Microsoft)**:
  A free app designed for visually impaired users that can read printed text, recognize faces, describe scenes, and identify products using barcode scanning.
- **Google Lookout**:
  Available on Android devices, it uses AI to read documents, recognize currency, and provide information about objects in the surroundings.
- **Voice Dream Scanner**:
  This app captures and reads out scanned documents with high accuracy and multilingual support.

2. Desktop OCR and TTS Software

PC-based software systems like **Kurzweil 1000**, **ABBYY FineReader**, or **NaturalReader** allow users to scan documents using scanners and convert them to speech.

Limitations:

- High cost
- Not portable
- Often overkill for simple, daily use

3. Embedded Devices and Assistive Tools

Devices like **OrCam MyEye** are wearable cameras for the blind that read out printed text from any surface, recognize faces, and identify objects.

Limitations:

- Extremely expensive (costs over $3000 USD)
- Limited customization
- Proprietary hardware and software

Proposed Methodology: The proposed system is a low-cost, portable, and offline-capable device that performs image-based text recognition and audio output using the Raspberry Pi platform. It is designed to assist visually impaired individuals and others who struggle with reading printed or handwritten text. The methodology involves the integration of image processing, optical character recognition (OCR), and text-to-speech (TTS) technologies. The entire system is built using open-source tools to ensure accessibility, affordability, and flexibility..

1. Image Capture

- A USB webcam or Raspberry Pi camera module is used to capture images of documents or text-based materials.
- A physical button or automatic trigger may initiate the capture process.
- The captured image is saved in a standard format (e.g., .jpg or .png).

2. Image Preprocessing

- To improve OCR accuracy, preprocessing steps are applied using **OpenCV** or **Pillow** libraries:
  - Grayscale conversion
  - Noise removal
  - Binarization (thresholding)
  - Skew correction (optional)

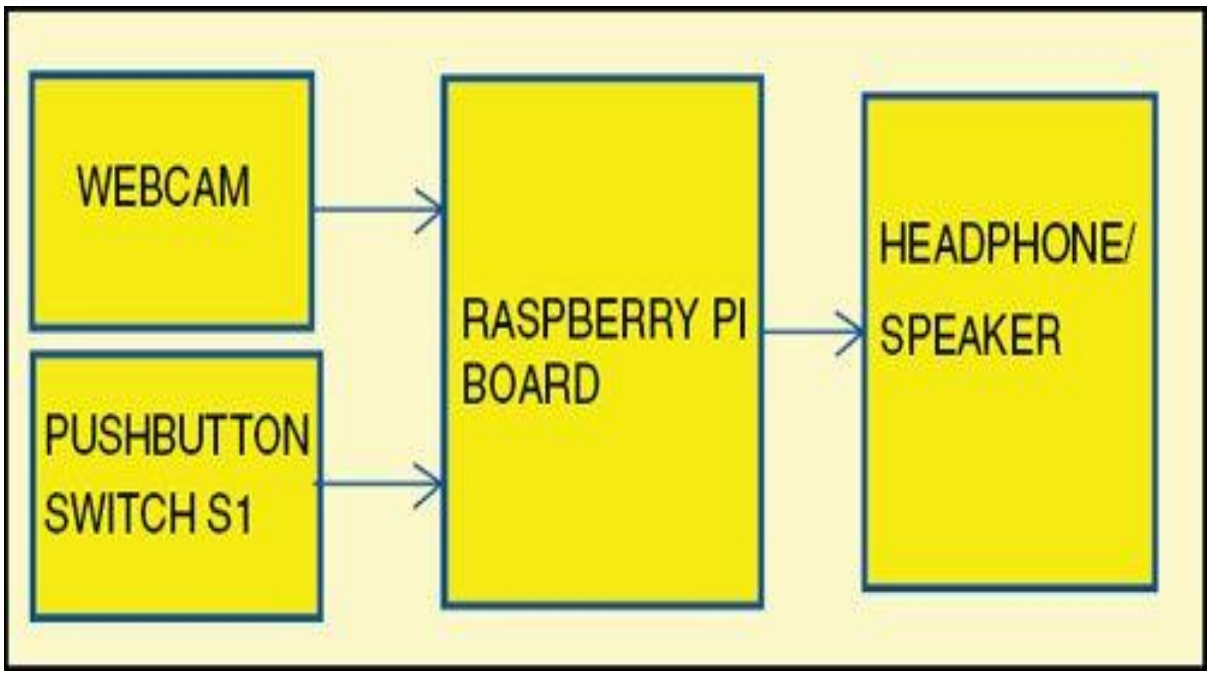
. Text Extraction using OCR

- The preprocessed image is passed to **Tesseract OCR**, an open-source engine capable of recognizing printed and handwritten text.
- The recognized text is converted into a string format and temporarily stored for the next phase.
- Multilingual support can be added by training Tesseract with language-specific models.

4. Text-to-Speech (TTS) Conversion

- The extracted text is converted into speech using a TTS engine:
  - **pyttsx3**: Works offline and supports multiple voices.

Block Diagram :



Applications:

1. Assistive Technology for the Visually Impaired

- Enables blind or low-vision individuals to read books, newspapers, signs, and documents through audio.
- Helps them become more independent in daily tasks such as reading instructions, menus, or labels.
- Reduces the need for human assistance in reading.

2. Educational Tools

- Supports students with **dyslexia**, **learning disabilities**, or **reading difficulties** by allowing them to listen to text content.
- Can be used in inclusive classrooms to help students understand printed material more easily.

3. Smart Reading Devices

- Can be integrated into **e-readers**, **portable scanners**, or **pen-like devices** for hands-free reading.
- Useful in libraries and educational institutes for scanning and reading reference material.

4. Public Information Access Systems

- Can be installed in **public places** (bus stands, railway stations, ATMs) to read out important notices or ticketing information.
- Increases accessibility for all users, including those who cannot read the local language.

5. Multilingual Communication

- The OCR system can be trained to recognize multiple languages, and TTS can read the translated text aloud.
- Useful for **tourists** or **migrants** to understand signage or documents in unfamiliar languages.

6. Industrial and Office Automation

- In document processing systems, it can be used to scan printed text and read it aloud for quick reviews.
- Can be added to machines or kiosks for automated reading of instructions, error messages, or data logs

7. Healthcare and Medical Settings

- Helps elderly patients or visually impaired individuals to understand medication instructions, prescriptions, or reports.
- Reduces the risk of medication errors due to misreading.

8. Accessibility in Digital Libraries

- Enhances digital library systems by making scanned or archived text accessible to all users through speech.

This system's wide applicability in both assistive and mainstream technologies makes it a valuable solution for promoting **inclusion**, **ease of access**, and **digital literacy**. Whether in schools, homes, public spaces, or workplaces, Image Text to Speech Conversion can greatly enhance the way people interact with written information.

Advantages:

1. Accessibility:

- **Visual Impairment:**

The system allows visually impaired individuals to access written content by converting images of text into audio.

- **Language Barriers:**

By translating text from the original language into the user's desired language, it overcomes communication barriers.

- **Educational Resources:**

It can make learning materials accessible to individuals with different learning styles or those who may struggle with reading traditional formats.

2. Convenience:

- **Portability:**

The Raspberry Pi is a compact, low-cost computer, making it easy to transport and use in various settings.

- **Real-time Conversion:**

The system can convert images of text into audio in real-time, allowing for immediate access to information.

- **Versatility:**

It can be used to process various image formats and can be adapted for different translation needs.

3. Cost-Effectiveness:

- **Affordable:**

The Raspberry Pi is relatively inexpensive, making the system affordable for individuals and organizations.

- **Open-Source Technology:**

Many of the software and tools used in this system are open-source, further reducing costs.

4. Other Benefits:

- **Educational and Research Purposes:**

The system can be used to digitize books, records, and other resources, making them easier to search and analyze.

- **Enhanced User Experience:**

By providing auditory feedback, the system can engage users more effectively, leading to higher

Disadvantages: While the Image Text to Speech (ITTS) system using Raspberry Pi provides several benefits, it also has some limitations that can affect its overall efficiency, user experience, or scalability. Recognizing these drawbacks is essential for planning future improvements and deployments.

1. OCR Accuracy Depends on Image Quality

- The system heavily relies on the quality of the input image.
- Poor lighting, low resolution, skewed angles, or blurry images can lead to **incorrect or incomplete text recognition**.
- Handwritten or stylized fonts may be **poorly recognized** or ignored entirely.

2. Limited Processing Power

- The Raspberry Pi, although powerful for its size, has **limited CPU and memory** compared to a standard PC.
- Complex images or long documents may lead to **slower processing speeds**, affecting real-time performance.

3. Noisy or Robotic Voice Output

- Offline TTS engines like pyttsx3 or espeak may produce **unnatural, robotic-sounding voices**.
- For more natural speech, cloud-based TTS (like Google TTS) is better, but it **requires internet access**.

4. Limited Multilingual and Regional Language Support (Offline)

- While Tesseract supports multiple languages, **regional or less-common languages may not perform well offline**.
- Text-to-speech engines also have **limited voice options** for certain languages, especially in offline mode

5. User Interface Limitations

- A basic system may lack a graphical interface or interactive controls, which can limit usability for non-technical users.
- Without proper audio navigation or tactile feedback, **visually impaired users may struggle to interact** with the device independently.

6. Not Suitable for Continuous Reading

- The system works well for short passages or single pages.
- Reading **entire books or large documents** continuously may not be practical due to

Future Scope:

The current Image Text to Speech (ITTS) system serves as a foundational prototype for assisting visually impaired individuals and others who face difficulty reading printed text. However, with advances in hardware, machine learning, and user-interface design, the system can be expanded and enhanced significantly. The following points outline the potential future developments and improvements of the project:

1. Integration with Artificial Intelligence (AI) and Machine Learning

- Implement AI-based OCR systems (like Google Vision AI or EasyOCR) for **more accurate text recognition**, even under poor lighting or noisy backgrounds.
- Use machine learning to **improve recognition of handwritten, cursive, or stylized text**.
- AI can also help in **scene understanding** — recognizing context, object labels, or priority content in an image.

2. Multilingual Translation and Speech Output

- Incorporate **real-time language translation** so the system can read text in one language and speak it in another.
- Useful for multilingual regions, travelers, or education in language learning.
- Can be extended to support **regional Indian languages** like Telugu, Tamil, Kannada, Hindi, etc.
- **3. Voice Command and Interaction**
- Implement **voice-activated controls** using speech recognition (e.g., "read this page", "translate to Hindi").
- Makes the system more **hands-free and user-friendly**, especially for visually impaired users.

References:

[1] Dr. KOMMU NAVEEN, Dr. R.M.S PARVATHI, on "A Review of Deep Learning Applications for Speech Processing Improvement & Applications" "Journal of Telecommunication, Switching Systems and Networks (JOTSSN)" | ISSN: 2454-637, p-ISSN: 2455-638| www.jotssn.com | Impact Factor: 7.569| || Volume 9, Issue 2, September 2022 || | DOI: DOI (Journal): 10.37591/JoTSSN/|http://engineeringjournals.stmjournals.in/index.php/JoTSSN/index

[2] Dr. KOMMU NAVEEN, Dr. R.M.S Parvathi, on "A Comprehensive Review On Machine Learning Applications of Convolutional Neural Networks to Medical Image Analysis" at 'International Conference on Robotics and Communication Technology-ICRCT-2022)' Page No: 50-57, ISSN : 2708-1079; IC Value: 45.98; Impact Factor: 7.379; VOLUME-03, ISSUE-1; 4th January, 2022. URL:https://doi.org/10.46379/jscer.2022.030104

[3] Hatice Cinar Akakin and Metin N. Gurcan, "Content Based Microscopic Image Retrieval System for Multi-Image Queries", IEEE Transactions On Information Technology In Biomedicine, VOL. 16, NO. 4, JULY 2012

[4] Yong-Hwan Lee and Sang-Burm Rhee, Bonam Kim, "Content- based Image Retrieval Using Wavelet Spatial-Color and Gabor Normalized Texture in Multi-resolution Database", 978-0- 7695- 4684-1/12 $26.00 © 2012 IEEE, DOI 10.1109/IMIS.2012.98

[5] Dr. KOMMU NAVEEN, Dr. R.M.S Parvathi, on "Contract and Feature Extraction of CBIR Method Using Soft Computing Techniques in Machine Learning" at 'Journal of Instrumentation Technology and Innovations)' Page No: 17-27; ISSN (PRINT): 2249-4731; ISSN (ONLINE): 2347-7261 VOLUME-12; ISSUE-01; April-May -2022. DOI(journal):10.37591/JoITI, URL: http:www.engineeringjournals.stmjournals.in/index.php/JoITI/index.

[6] S. Mangijao Singh , K. Hemachandran , "Content-Based Image Retrieval using Color Moment and Gabor Texture Feature", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012; ISSN (Online): 1694- 0814

[7] Dr. K. Naveen Kumar, Dr. M. B. Raju, Prof. A. Sridhar, IJRAR August 2024, Volume 11, Issue 3 www.ijrar.org (E-ISSN 2348-1269, P- ISSN 2349-5138) IJRAR24C2120 , International Journal of Research and Analytical Reviews (IJRAR) 2024 IJRAR August 2024, Volume 11, Issue 3 www.ijrar.org (E-ISSN 2348-1269, P-ISSN 2349-5138).

[8] Kanchan Saxena, Vineet Richaria, Vijay Trivedi,"A Survey on Content Based Image Retrieval using BDIP, BVLC AND DCD", Journal of Global Research in Computer Science , Vol.3, No. 9, September 2012 ,ISSN-2229-371X.

[9] Dr. K. Naveen Kumar, Dr. M. B. Raju, Prof. A. Sridhar on International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue VIII Aug 2024- Available at www.ijraset.com.

10) Dr. KOMMU NAVEEN, Dr. R.M.S Parvathi, on "A Review On Content Based Image Retrieval Systems Features derived by Deep Learning Models" at 'International Journal For Research In Applied Science And Engineering Technology (IJRASET)' Page No: 42-57, ISSN : 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429; VOLUME-09, ISSUE-XII; November- 2021. URL:www.ijraset.com/index.html/doi.org.