Android TTS OCR Converter System for people with Visual Disability

Dr. Anita TN[1], Akshatha S[2], Rachana B[3], Rahul Jangra[4], Venkatesh Rayudu[5]
[1,2,3,4,5] *Dept. of Computer Science and Engineering, Sir M. Visvesvaraya Insitute of Technology Bengaluru, KA, India*

***Abstract*—In the world of rapid digitalization, people with visual impairment fall behind, the visually presented information in the physical world remains out of reach from these people, to help them able to receive the written information from screens, boards, charts, etc. while also bridging the gap between the language barriers, this app will use audio as the medium for the information sharing. Using OCR TTS tech, the focus is the provide a simple and easy to use UI for the User.**

Keywords—Assistive technology, OCR, Text-to-Speech, Android, Visual impairment, Accessibility.

I. INTRODUCTION

Accessing printed and visually displayed information is still a challenge for lakhs of people with visual impairments, majority of public use information is visually displayed, making it a barrier even in a technology enabled age.

The text-to-speech (TTS) and optical character recognition (OCR) technologies have existed in widely available formats through internet and phone apps, there have been developments on applications of these, but they lack finesse to be used every day and provide a variety for language support for regional cases, focusing on daily ease of use and audible medium as alternative to visual, through mobile apps. Aiming for minimized cost, easy accessibility and future development possibilities the approach followed in this paper is android app using Kotlin.

II. LITRATURE REVIEW

A. Existing technologies

Several visually assisting tools are available, these solutions primarily rely on OCR, While these applications have made significant strides in accessibility, they often rely on cloud-based AI processing, require an active internet connection, or have complex user interfaces that may not be suitable for all users.

B. Existing Assistive Technologies for the Visually Impaired

Applications like JAWS (Job Access with Speech) and NVDA (Non Visual Desktop Access) provide text-to-speech functionalities but are primarily limited to digital content.

Hardware devices such as OrCam MyEye and KNFB Reader allow users to scan and listen to printed text but are often expensive and lack accessibility for all users.

Apps like Seeing AI (by Microsoft) and Google Lookout use OCR and AI to extract and vocalize text from images, making them accessible and portable solutions

While these applications have made significant strides in accessibility, they often rely on cloud-based AI processing, require an active internet connection, or have complex user interfaces that may not be suitable for all users.

C. Research Studies on OCR and TTS-Based Assistive Systems

• Smith et al. (2019), "Enhancing Accessibility: A Smartphone-Based OCR System for Visually Impaired Users" - Highlights the importance of real-time OCR processing on mobile devices and compares cloud-based and offline OCR models, emphasizing that cloud-based OCR (e.g., Google Cloud Vision API) offers higher accuracy but depends on internet connectivity.

• John Williams, Emily Brown (2020), "Optical Character Recognition for Assistive Technologies" - Explores the use of OCR for text extraction, finding it effective for printed text but struggling with handwritten and low-light images.

• Michael Lee, Sophia Chen (2021), "Integrating OCR and TTS for Visually Impaired Individuals" - Focuses on developing a mobile-based OCR-TTS system, showing that real-time text-to-speech conversion improves accessibility but manual navigation remains a limitation.

• James Anderson, Olivia White (2022), "Gesture-Based Controls in Accessibility Applications" - Examines gesture-based controls for assistive tools, concluding that hands-free interaction enhances usability but requires training for effective use

III. PROBLEM IDENTIFICATION

Visually impaired individuals face significant challenges in interacting with their surroundings due to their inability to read printed text. This limitation can hinder their independence, mobility, and access to crucial information, affecting their quality of life.

OCR (Optical Character Recognition) and TTS (Text-to-Speech) solutions, have emerged as technical tools to help in accessibility of visual information, as software solutions. However, many existing solutions have limitations such as high cost, internet dependency, and complex user interfaces, making them inaccessible to a broader audience.

This project aims to overcome these challenges by developing a cost-effective, user-friendly, and offline-capable Android application that enables visually impaired individuals to read printed text.

Significance of the Work

1. Enhancing Accessibility and Independence –
   By providing a hands-free, voice-controlled interface, the application enables visually impaired individuals to access printed text without requiring external assistance.
2. Providing a Cost-Effective Solution -
   High-end assistive devices such as OrCam MyEye are prohibitively expensive, limiting their availability to only a small section of users.
3. Eliminating Internet Dependency -
   Unlike cloud-based OCR solutions that require constant internet access, this system aims to function without internet connectivity, making it reliable even in areas with limited or no network coverage.
4. User-Friendly and Hands-Free Operation
   Traditional OCR-based applications often require users to navigate through touch-based menus, which can be challenging for visually impaired individuals.
5. Real-World Impact
   The application will enable users to read signs in public places, identify product labels in stores, access printed instructions, and read documents independently.

IV. OBJECTIVE & SIGNIFICANCE

The primary objective of this project is to design an Android-based assisting app that helps visually impaired individuals to access physically printed information in real-life through audio as a medium. The system removes gap between the physically displayed information to digital world.

The key objectives and the significance of the proposed work are outlined below:

- To develop a smartphone application capable of extracting text from images using OCR. The application uses the camera to capture text from various sources such as printed documents, signs, and labels, providing a portable and accessible solution.
- To convert the extracted text into speech using TTS for real-time audio feedback. The integration of a Text-to-Speech (TTS) engine enables users to listen to the recognized content immediately, eliminating the need for visual confirmation.
- To support offline functionality using ML Kit as a fallback when cloud OCR fails or is unavailable.
  This ensures reliability in areas with limited or no internet connectivity.
- To implement gesture-based controls (volume buttons) for user-friendly, hands-free operation. Visually impaired users can trigger the capture function without relying on touchscreen gestures, improving accessibility and ease of use.
- To ensure simple and intuitive interface with optional on-screen buttons for diverse user preferences.
  While gesture control is the primary mode of interaction, on-screen controls are included to accommodate different levels of visual impairment or personal comfort.

The significance of this work lies in its potential to empower visually impaired individuals by providing them with greater independence.

• To develop a smartphone application capable of extracting text from images using
OCR.

• To convert the extracted text into speech using TTS for real-time audio feedback.

• To support offline functionality using ML Kit as a fallback when cloud OCR fails or is unavailable.

• To implement gesture-based controls (volume buttons) for user-friendly, hands-free operation.

• To ensure a simple and intuitive interface with optional on-screen buttons for diverse user preferences.

V. IMPLEMENTATION AND FEATURES

1. Buttons Triggered Image controls
   The complete control of the app is distributed over the volume keys of the phone, divided into 4 options of single and double press (2 types) for each of the volume up and down key (2 buttons), covering four options to capture, repeat, stop and exit.
2. Online and offline functionality
   The OCR can be performed over internet as well as offline, for optional take aways between availability of internet and speed of results.
3. Multi Language Support
   Over the availability of internet the app is able to read through different languages covered by the Google OCR library

Tool	Google Cloud Vision	ML OCR kit	TTS engine
Languages	50+	limited	40+
Indian Languages	Strong	Weak	Strong
Offline	No	Yes	Yes
Customizable	Yes	No	Yes

Table 1. Language support details

4. Dynamic OCR Mode Switching

The app automatically switches between Google Cloud Vision API and ML Kit based on the availability of internet, ensuring uninterrupted service.

5. Real-Time Voice Feedback

The app provides immediate voice feedback to users as soon as text is detected, helping them interact with their surroundings effortlessly

6. Label Detection

The app can detect labels (objects/items) during text processing. This feature is still in debugging but will soon enhance the app by announcing labels like "laptop" or "chair"

7. Salient features

- Vibration on App Open and Close
- Autofocus
- Automatic volume increase
- Announce Top 2 Labels Before Captured Text
- Customizable TTS

VI. EXAMPLE USAGE

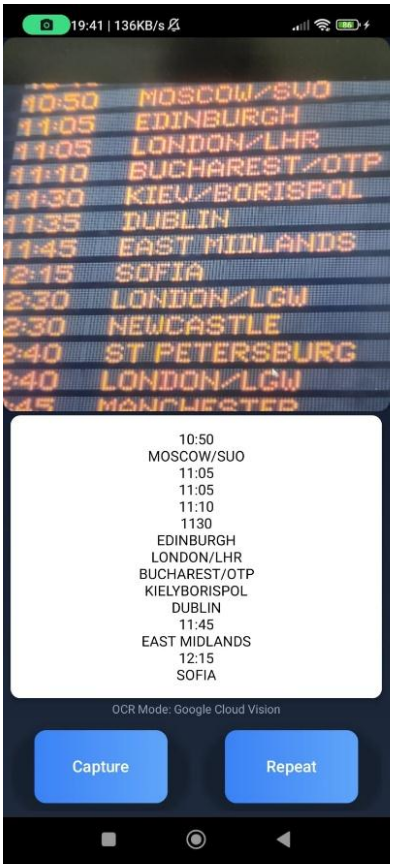

Image 1. Station board

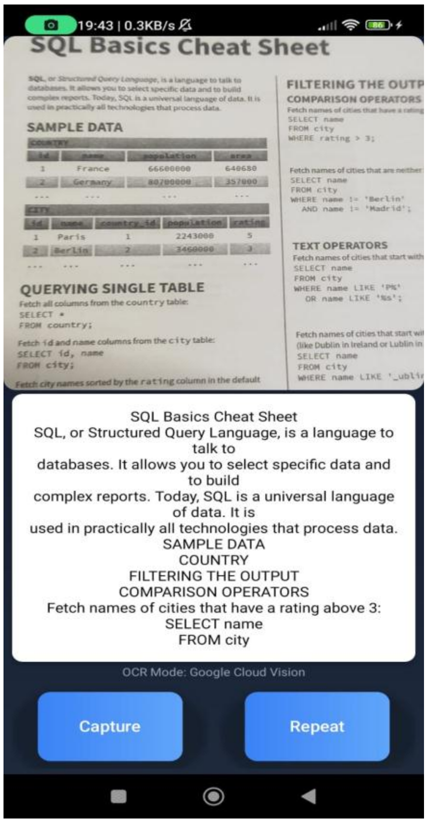

Image 2. Printed document

Image 1 : The primary usage of the app is to capture real world fast changing information

Image 2 : The result show a printed text that is parsed by the app accurately, and this'll help user to go through document.

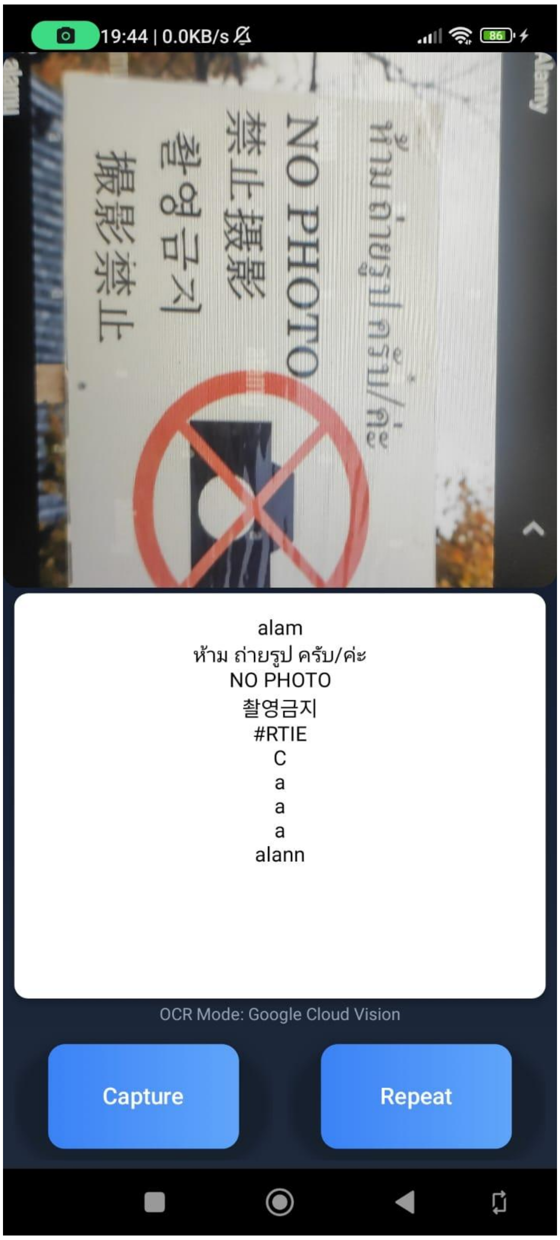

Image 3:
This image shows signboard on the road, also goes a step further to read the local language, updates on this will be a translation feature for the languages.

VII. METHODOLOGY AND WORKFLOW

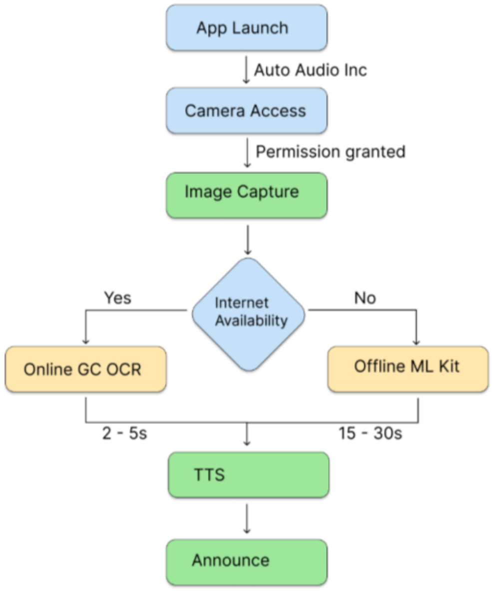

Image 1. Flow Diagram

1. Initialization
- App Launches: Mobile application initializes.
- Camera Access: Requests and acquires permission to access the phone camera.

2. Image Acquisition
- Capture Input:
  - User captures an image through the in-app camera interface.
  - Alternatively, user can load an existing image from the device storage.

3. OCR Processing Path
- Routing Decision: Check for internet availability to determine which OCR path to follow:
- Online Mode (Google Vision Cloud API):
  - Image is sent to Google Vision Cloud API.
  - Extracted text is returned from the cloud.
  - * Average Latency: 2.5 – 4 seconds.*
- Offline Mode (ML Kit On-device):
  - Image is processed locally using ML Kit.
  - Text is extracted on the device.
  - * Average Processing Time: 15 – 30 seconds.*

4. TTS (Text-to-Speech) Feedback
- Extracted text is sent to the TTS Engine:
  - Synthesized into audio format.
  - Audio is played to the user for auditory feedback.

- *□ Estimated TTS Response Time: 0.3 – 0.8 seconds.*

5. User Feedback Loop

- Option to repeat the audio.
- Option to re-capture or select another image.

Path	Online	Offline
OCR Type	Google API	ML Kit
Avg Time (s)	2.5 – 4	15 – 30
TTS Time (s)	0.3 – 0.8	0.3 – 0.8
Total Est. Time	2.8 – 4.8	15.3 – 30.8

Table 2. Time Taken for each workflows

VIII. REQUIREMENTS

The app should work efficiently on devices running Android 8.0 (Oreo) and above and should function across different screen sizes and hardware configurations,

Camera: Minimum 5 MP resolution for accurate OCR processing,

Processor: Quad-core or higher for efficient image processing,

RAM: At least 2GB RAM for smooth performance

REFERENCES

[1] Smith, R., et al. (2007). An Overview of the Tesseract OCR Engine. https://tesseractocr.github.io/

[2] Raman, T. V. (1996). Emacspeak—Enabling Access for Blind Users. http://emacspeak.sourceforge.net/

[3] Harper, S., & Yesilada, Y. (2018). Web Accessibility: A Foundation for Research. https://www.springer.com/gp/book/9781447174407

[4] Chen, D., & Odobez, J. M. (2005). Text Recognition in Natural Images. https://www.sciencedirect.com/science/article/pii/S0031320305002502

[5] Google Cloud Vision API Documentation. https://cloud.google.com/vision/docs

[6] Android Developers Documentation – CameraX and TTS API. https://developer.android.com/

[7] World Health Organization (WHO) Reports on Visual Impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-vision-impairment

[8] W3C Accessibility Guidelines. https://www.w3.org/WAI/S. Liu, " Wi-Fi Energy Detection Testbed (12MTC)," 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC