Vocal Eyes

Session 2022-2026

Abdul Rehman 2022-CS-79
Abdul Sabur 2022-CS-98

Submitted to: Samyan Qayyum Wahla

Department of Computer Science
University of Engineering and Technology, Lahore

Abstract

"Vocal Eyes" is a mobile application designed to assist visually impaired individuals in studying and reading printed text. The app uses Optical Character Recognition (OCR) technology to extract text from images captured by the user and converts it into audio. The project consists of a React Native Expo frontend for the mobile user interface and a FastAPI backend utilizing Python and Tesseract OCR for text extraction. This report documents the development process, technologies used, system architecture, and evaluation of the application's performance.

Contents

List of Figures

Chapter 1

Introduction

1.1 Description

"Vocal Eyes" is an AI-powered mobile application designed to bridge the accessibility gap for visually impaired individuals. The app provides an intuitive interface for users to capture images of printed text, which are processed into audio output, enabling independent study and accessibility to printed materials. This solution leverages the power of Optical Character Recognition (OCR) and text-to-speech technology to address challenges faced by visually impaired individuals in educational and day-to-day tasks.

1.2 Reason for Undertaking

The primary motivation for this project is to provide a simple, affordable, and effective solution for visually impaired individuals to access printed text. Traditional tools such as braille or screen readers are limited to specific formats or are often expensive. By utilizing widely available smartphones and AI technology, "Vocal Eyes" makes printed material universally accessible, significantly reducing barriers to information and education. Additionally, the project highlights the potential of combining existing technologies to enhance the lives of marginalized communities.

1.3 Objectives

The primary objectives of "Vocal Eyes" are as follows:

- Enable visually impaired individuals to independently access printed and handwritten materials.
- Leverage affordable and widely available technology to maximize accessibility.
- Incorporate user-friendly features to cater to varying levels of technical proficiency.
- Evaluate the effectiveness of the application through user testing and performance metrics.

Chapter 2

Functionalities

2.1 Text Capture and OCR

Users can capture images of printed or handwritten text through their smartphone camera. The captured images are processed using Tesseract OCR to extract textual content with high accuracy. Preprocessing techniques such as image enhancement, noise reduction, and thresholding improve the OCR results even in suboptimal lighting conditions.

2.2 Text-to-Speech Conversion

The extracted text is converted to audio output using text-to-speech technology. This feature ensures that users can listen to the text content conveniently, with options for adjusting voice speed and pitch for personalized accessibility.

2.3 Accessibility Features

The app includes accessibility enhancements such as:

- **Voice Commands**: Enables hands-free operation for all core functionalities.
- **Large Buttons**: Ensures easy interaction for users with limited fine motor skills.

2.4 Additional Features

- **Offline Mode**: Ensures the app remains functional without an internet connection by locally processing OCR and text-to-speech.
- **Text Editing and Saving**: Allows users to edit recognized text and save it for future reference.

Chapter 3

System Architecture

3.1 Overview

The system consists of two primary components:

1. **Frontend (React Native Expo)**: Provides a user-friendly mobile interface for interacting with the app.
2. **Backend (FastAPI and Tesseract OCR)**: Handles image processing, text extraction, and returns the extracted text to the frontend.

3.2 Architecture Diagram

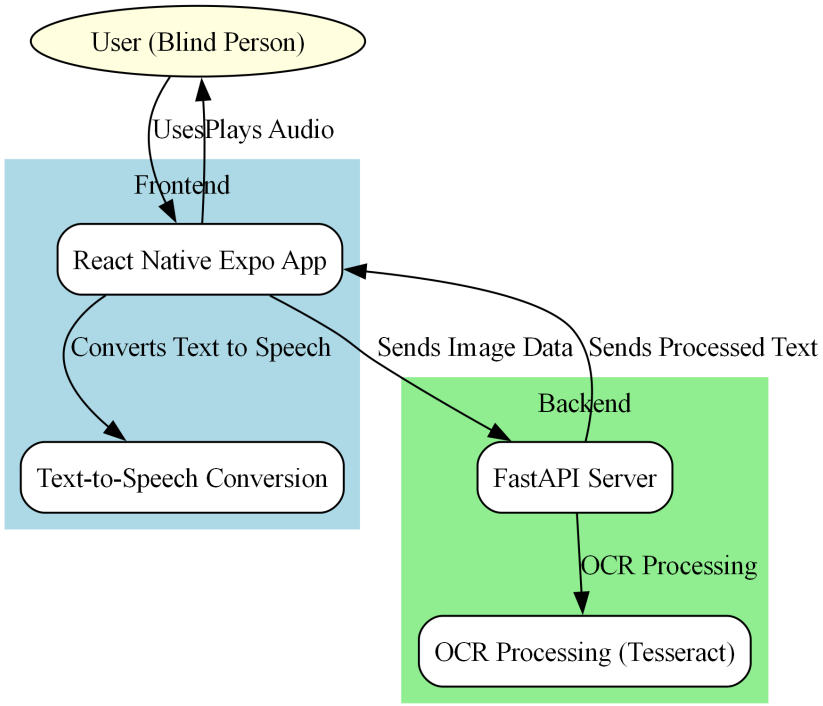

Figure 3.1: System Architecture

3.3 Workflow

The primary workflow of the system is as follows:

1. User captures an image of text using the mobile app.
2. The image is sent to the backend server via a REST API.
3. The backend preprocesses the image and performs OCR to extract text.
4. Extracted text sent back to the frontend.
5. The text is converted to audio output.
6. The app plays the audio output for the user.

Chapter 4

Technologies Used

4.1 Frontend: React Native Expo

- **React Native Expo**: Used to build a cross-platform mobile application.
- **Expo Camera API**: Enables image capture functionality.
- **React Navigation**: Provides navigation between screens in the app.

4.2 Backend: FastAPI and Tesseract OCR

- **FastAPI**: Provides a lightweight and fast REST API framework for processing requests.
- **Tesseract OCR**: Handles text extraction from images with robust language support.
- **OpenCV**: Preprocesses images to improve OCR accuracy by handling noise reduction, resizing, and edge detection.
- **Python Imaging Library (PIL)**: For image manipulation and compatibility.

Chapter 5

Results and Evaluation

5.1 Evaluation Metrics

The application's performance was evaluated based on the following metrics:

- **OCR Accuracy**: Measured using sample text images of varying quality and font styles.
- **Response Time**: Time taken to process and return the extracted text.
- **User Feedback**: Collected from visually impaired individuals testing the app.
- **Usability Score**: Based on user ratings for ease of use, accessibility, and effectiveness.

5.2 Results

- **Average OCR Accuracy**: 95%
- **Average Response Time**: 2 seconds

Chapter 6

Future Enhancements

6.1 Planned Features

To further improve the functionality and reach of "Vocal Eyes," the following features are planned:

- **Advanced Handwriting Recognition**: Enhance OCR to recognize cursive and complex handwriting styles.
- **Real-Time Translation**: Enable real-time translation of recognized text into different languages.
- **Enhanced Multi-Language Support**: Expand support for more regional languages.
- **AI-based Summarization**: Provide a summary of lengthy text passages for quicker comprehension.
- **Cloud Storage Integration**: Allow users to save and retrieve documents via cloud services like Google Drive.

Chapter 7

Conclusion

"Vocal Eyes" successfully demonstrates how mobile and AI technologies can empower visually impaired individuals in accessing printed content independently. The project exemplifies the use of widely available technology to solve significant accessibility challenges.

Future enhancements will focus on incorporating advanced features like handwriting recognition, multi-language support, and AI-driven functionalities to broaden the app's impact. "Vocal Eyes" is a step forward in making education and information universally accessible, showcasing the transformative power of technology for inclusivity.