CSO 211 Project Report

Title: "Development of AI-Based Visual Aid with Integrated Reading Assistant for the Completely Blind"

Group Members:

1.Name: Shayon Dasgupta
Roll Number: 23074026
2.Name: Soumadip Majumder
Roll Number: 23074027

Institute: Indian Institute Of Technology(BHU), Varanasi

Department: Computer Science And Engineering(CSE), Integrated Dual Degree, Second Year

Course: Computer System Organization (CSO-211)

Date Of Submission: 1st Of December, 2024

Introduction:

Background Information:

The development of AI-based visual aids integrated with reading assistants is a transformative solution for the totally blind. An increasingly visually-dependent world makes it rather difficult to access text-based information, like books, signs, or documents, for the visually impaired. Advanced technologies in artificial intelligence, such as computer vision and natural language processing, provide the means to bridge this gap. With such systems converting the text from visuals into audio output or another form of interactive content, the persons are better empowered about independence in education, at work, or in daily living. Such innovations would make it easier towards an accessible society where people are blind would easily get themselves involved within and around their environments.

Problem Statement:

Completely blind people lack access to most visual information-whether printed out or even displayed digitally-including texts and other documents, which also limits their autonomy in the learning process and everyday circumstances. Therefore, there should be AI-based support for vision with an integrated reader assistant offering audio explanation of the visual content.

Methodology:

Research Design/Approach:

This project employs a modular, implementation-driven approach using React Native for cross-platform app development, integrating AI models and advanced libraries for audio processing, image recognition, and natural language interaction. The system architecture is divided into three core functionalities: audio recording and transcription, AI-based interaction, and image processing, all orchestrated through seamless API integrations.

This process has been divided into the following parts:

- ## Audio Recording and Transcription:

1) Audio Recording:

(a)Users can record their instructions via a microphone using the expo-av library in .wav format.
(b)The recording process is initialized, controlled, and stopped via user interactions.

2) Speech-to-Text Transcription:

(a) The recorded audio is sent to the OpenAI Whisper API using a FormData object.
(b) The API processes and returns the transcribed text, which is stored for further processing.

- ## ChatGPT Integration for AI Responses:

1) Processing Instructions:

(a)The transcribed text is sent to OpenAI GPT-4 via an API call with a custom system prompt for context.
(b)GPT-4 generates a response based on the input, which is displayed in the app.

2) Text-to-Speech (TTS):

(a) The response text is converted into audio using expo-speech.
(b) Users can listen to the AI-generated response in real-time.

- ## Image Capture And Analysis:

1) Image Capture:

(a) The app requests camera permissions and allows users to capture images using expo-image-picker.
(b) Captured images are saved temporarily in the device’s storage.

2) Artificial Intelligence (AI) Analysis:

(a) The Image URL is sent to GPT-4, which analyses and provides a description or answer based on the image content.

▪ Key Functions And Their Roles:

1) Audio Recording:

(a) **getMicrophonePermission**: Requests and ensures microphone permissions.
(b) **startRecording**: Starts capturing audio with predefined options.
(c) **stopRecording**: Stops recording and processes the audio file.
(d) **sendAudioToWhisper**: Sends recorded audio to OpenAI's **Whisper API** for transcription.

2) Image Handling:

(a) **openCamera**: Opens the Camera and captures an Image.
(b) sendImageToChatGPT: Sends the image URL to ChatGPT for image description.

3) Text Processing:

(a) sendToGpt: Submits user or transcribed text to GPT-4 for generating contextual responses.
(b) speakText: Converts the GPT-generated response to speech.

▪ Tools And Technologies:

1) React Native Framework:

(a) **Enables cross -platform development.**

2) Expo Libraries:

(a)expo-av: For audio recording.

(b)expo-speech: For text-to-speech conversion.
(c)expo-image-picker: For capturing images.
(d)expo-file-system: For handling file storage.

3) APIs:

(a)OpenAI Whisper API for speech-to-text transcription.
(b)GPT-4 for natural language processing and image analysis.

▪ Potential Issues And Fixes:

1) Permission Errors:

(a) Proper handling of both microphone and camera permissions must be ensured.
(b) Detailed alerts to guide users if permissions are denied must be provided.

2) AUDIO File Format Mismatch:

(a) It must be ensured that the Whisper API supports the audio format (.wav) generated by the recording setup.

3) Environment Variable Mismatch:

(a) It must be verified that the EXPO_PUBLIC_OPENAI_API_KEY environment variable is correctly set and accessible in the project.

4) File Handling In sendAudioToWhisper():

(a) FileSystem.getInfoAsync is used to validate the presence of audio files before uploading.

• Limitations:

i. **Transcription Accuracy:** The quality of transcription depends on the clarity of recorded audio. Background noise may affect the results.

ii. **Image Quality:** The accuracy of image analysis relies on the clarity and resolution of the captured image.

iii. **API Costs:** High usage of Whisper and GPT-4 APIs may lead to increased operational costs.

iv. **Limited Voice Customization:** Text-to-speech options are currently limited to default voices provided by expo-speech.

Results:

• Findings:

The project successfully developed a cross-platform mobile application that serves as an AI-powered visual aid integrated with a reading assistant for the completely blind. Key findings include:

(1) Audio Recording and Transcription:

- The app allows users to record audio instructions using a microphone and processes the recordings into text via OpenAI Whisper API.
- The transcription was found to be accurate in clear, noise-free environments, with minor inconsistencies in the presence of background noise.

(2) ChatGPT Integration:

- The transcribed instructions are sent to GPT-4 for generating contextually relevant and accurate responses.
- The app converts these responses into audio output using text-to-speech (TTS), which provides a clear and understandable auditory experience for users.

(3) Image Capture and Analysis:

- Users can capture images using the app's camera functionality, and these images are analysed by GPT-4 to provide descriptive or interpretative feedback.
- The image analysis performed well for clear and high-resolution images but was less effective for low-quality or blurry images.

(4) System Performance:

- The app demonstrated smooth operation across platforms, showcasing effective integration of APIs and libraries for audio, image, and natural language processing.
- The processing time for audio and image inputs was efficient, although optimization is required for handling larger files or continuous usage.

(5) User Interface:

- The interface was intuitive and accessible, ensuring ease of use for visually impaired individuals. Basic permission prompts and error alerts were effective in guiding users through the functionalities.

Limitations:

- **Transcription and Image Accuracy**: Dependent on input quality; background noise and low-resolution images affected performance.
- **API Costs**: Heavy usage of Whisper and GPT-4 APIs could lead to increased operational expenses.
- **Voice Customization**: Limited voice options in TTS reduce personalization for users.

The app demonstrates a scalable and practical solution for assisting visually impaired individuals by making visual information more accessible. Further work can expand its functionality and optimize the system for broader application.

Conclusions:

• Summary Of Findings:

The project successfully achieved its goal of developing an AI-based visual aid with an integrated reading assistant for the completely blind. Key accomplishments include:

- Seamless integration of advanced AI technologies, such as OpenAI Whisper API for speech-to-text transcription and GPT-4 for natural language understanding and image analysis.
- Effective conversion of audio instructions into actionable outputs, enabling visually impaired individuals to interact with their environment and access information independently.
- A user-friendly mobile application built with React Native, ensuring accessibility and compatibility across multiple platforms.
- The modular design of the app allows for efficient handling of audio recording, text-to-speech conversion, and image recognition tasks.
- However, the findings also highlighted areas for improvement, such as the dependency on high-quality input (audio and images), limited voice customization, and the potential cost of API usage for large-scale deployment.

• Recommendations:

1.Enhancing Accuracy:

- Implement advanced noise-reduction techniques to improve transcription accuracy in noisy environments.
- Use more sophisticated image preprocessing algorithms to handle low-quality or blurry images effectively.

2.Reducing Costs:

- Explore open-source or cost-effective alternatives to APIs to minimize operational expenses.

3.Improving Personalization:

- Incorporate customizable text-to-speech options to better cater to user preferences.

4.User Feedback Integration:

- Conduct usability studies with visually impaired users to identify areas of improvement and further refine the app's functionality and interface.

5.Expanding Features:

- Introduce additional capabilities such as multi-language support, real-time object recognition, or navigation assistance to broaden the app's utility.

The project lays a solid foundation for creating accessible and inclusive technologies for visually impaired individuals. With further enhancements, the app has the potential to significantly improve the quality of life for its users and promote a more inclusive society.