Image-to-Audio Conversion for Accessible Reading Utilizing MATLAB OCR and TTS

Shruthi Ramprasad
Student at Dept. Electrical and Electronics Engg.
Manipal Institue of Technology, Manipal
Karnataka, India
220906308

Harikishanthini K
Student at Dept. Electrical and Electronics Engg.
Manipal Institue of Technology, Manipal
Karnataka, India
220906286

***Abstract*—This report presents a sophisticated image-to-audio conversion system aimed at enhancing accessibility for visually impaired individuals by transforming printed text into audio format. Leveraging the Optical Character Recognition (OCR) and Text-to-Speech (TTS) functionalities available in MATLAB, the proposed system captures textual information from images and converts it into a synthesized audio file. The primary objective is to facilitate an inclusive reading experience for visually impaired users, effectively addressing the literacy barriers associated with conventional printed materials. This document provides an in-depth examination of the system architecture, methodology, results, and potential enhancements, thereby offering a comprehensive overview of the development process for this innovative solution.**

I. Introduction

The pervasive nature of physical text presents significant barriers to access for individuals with visual impairments. While digital formats such as e-books and screen readers have emerged as partial solutions, they remain limited in scope compared to the vast array of printed content available. Consequently, there is a critical need for a universally accessible system capable of converting physical text into audio format instantaneously.

Recent advancements in image processing, Optical Character Recognition (OCR), and Text-to-Speech (TTS) technologies have opened new avenues for addressing these challenges. OCR enables machines to interpret characters from scanned or photographed text, while TTS technology synthesizes these characters into audible speech. This project integrates these technologies within the MATLAB environment to develop a compact, real-time image-to-audio conversion system. Such a solution empowers visually impaired users to "read" physical texts—including books, labels, and other printed materials—thereby improving their access to information and fostering greater inclusivity.

II. Literature Review

- Wang and Li (2022) developed a MATLAB-based TTS system aimed at enhancing accessibility for blind readers through template matching techniques for character segmentation and a seamless OCR-TTS integration pipeline. Their work provides foundational methodologies that inform the design of OCR-TTS systems by emphasizing efficient processing techniques for fragmented text.
- Isewon et al. (2014) designed a TTS synthesizer specifically for visually impaired users, underscoring the significance of Natural Language Processing (NLP) and Digital Signal Processing (DSP) in achieving high intelligibility in speech synthesis. Their findings align closely with the objectives of this project.
- Lemmetty (1999) conducted a comprehensive review of advancements in speech synthesis technology, focusing on NLP and prosody adjustments necessary for generating natural-sounding audio output. This research supports the TTS component of our project by ensuring clarity across diverse content types.
- Swathi et al. (2013) explored language-specific TTS adaptations, particularly focusing on non-English languages. Their insights are instrumental in considering future expansions of this project to include multilingual OCR and TTS functionalities.
- Ngugi et al. (2005) contributed valuable insights into the development of a Swahili TTS system, addressing challenges related to language-specific phonetic processing—an essential consideration for producing accurate audio output across various languages and dialects.

III. Methodology

The development of the image-to-audio conversion system encompasses several key stages:

A. Overview of Text-to-Speech Technology

Text-to-Speech (TTS) technology converts written text into spoken words using various linguistic and acoustic models. The process begins with analyzing the input text to identify its linguistic structure, including phonemes—the smallest units of sound that distinguish one word from another—and prosodic features like intonation and rhythm that contribute to natural-sounding speech. The TTS system typically follows these steps:

- Text Analysis: The input text is parsed to identify words, phrases, and sentences.

- Phonetic Transcription: The recognized words are converted into phonetic representations based on predefined pronunciation rules or dictionaries.
- Prosody Generation: The system determines appropriate prosodic features such as pitch, duration, and stress patterns to enhance naturalness.
- Speech Synthesis: Finally, the phonetic and prosodic information is fed into a synthesizer that generates audible speech through waveform generation techniques.

This multi-step process ensures that the synthesized speech is intelligible and closely resembles human voice characteristics.

B. Image Acquisition

The system initiates by loading an image file—either a scanned page or photograph—utilizing MATLAB's imread() function as the input source for subsequent OCR processing.

C. Preprocessing

To enhance text extraction efficacy, the image undergoes grayscale conversion followed by binarization. This step improves contrast between text and background, which is critical for achieving accurate OCR results. Adaptive thresholding techniques are employed to accommodate varying lighting conditions and backgrounds.

```
grayImage = rgb2gray(img);
bwImage = imbinarize(grayImage, 'adaptive',
'FgndPolarity','dark', 'Sensitivity', 0.4);
```

D. OCR Processing

The preprocessed binary image is analyzed using MATLAB's ocr() function to convert recognized text regions into machine-readable characters, which are subsequently stored for TTS conversion.

```
ocrResults = ocr(bwImage);
recognizedText = ocrResults.Text;
```

E. Text-to-Speech Conversion

The recognized textual data is transformed into audio using the `System.Speech.Synthesis.SpeechSynthesizer` object available in `.NET` Framework. The resulting audio output is saved as a `.wav` file format for user accessibility.

```
NET.addAssembly('System.Speech');
obj = System.Speech.Synthesis.SpeechSynth;
obj.SetOutputToWaveFile('Op_audio.wav');
obj.Speak(recognizedText);
```

F. Output and Storage

The generated audio file is stored systematically to ensure playback accessibility across various devices, thereby enhancing the usability of printed text information.

G. Full code used

```
% Step 1: Load Image
imageFile = 'img1.jpg';
img = imread(imageFile);
imshow(img)
% Step 2: Preprocessing
grayImage = rgb2gray(img);
bwImage = imbinarize(grayImage, 'adaptive',
'ForegroundPolarity',
'dark', 'Sensitivity', 0.4);

% Step 3: OCR (Optical Character Recognition)
ocrResults = ocr(bwImage);
recognizedText = ocrResults.Text;

% Step 4: Display Recognized Text
disp(recognizedText);

% Step 5: Text-to-Speech Conversion
recognizedText = char(recognizedText);
NET.addAssembly('System.Speech');
obj = System.Speech.Synthesis.SpeechSynth;
obj.Volume = 100;
outputFileName = 'Output_audio.wav';
obj.SetOutputToWaveFile(outputFileName);
obj.Speak(recognizedText);
obj.SetOutputToDefaultAudioDevice();
disp(['Audio saved to ', outputFileName]);
```

IV. Problem Formulation & Specifications

The principal aim of this project is to devise a real-time image-to-audio conversion tool utilizing OCR and TTS technologies with specific specifications:

- Image Processing Sensitivity: Implementation of adaptive thresholding guarantees reliable text extraction under diverse environmental conditions.
- Audio Quality: Optimization of TTS parameters ensures clarity and volume appropriateness for enhanced intelligibility.
- Output Format: The system generates .wav files compatible with most media playback devices, ensuring flexibility for end-users.

This initiative seeks to significantly improve accessibility for visually impaired individuals by providing an efficient and portable solution for converting printed text into an audio format.

V. Results

Text Recognition Accuracy: The OCR system demonstrates high efficacy in extracting text from high-quality images with minimal recognition errors; however, performance declines under low-light or noisy conditions.

Audio Output Quality: The TTS synthesizer produces clear audio outputs that accurately reflect sentence structures and intonation patterns; tests indicate high intelligibility even with complex text layouts.

Performance Insights: The system operates effectively in real-time with negligible processing delays; future enhancements may include refining binarization settings to better handle complex backgrounds and extending support to additional languages.

VI. Conclusion

This project successfully implements an integrated OCR-to-TTS system using MATLAB, providing an innovative reading solution for visually impaired users. By capturing textual information from images and converting it into audio format, the project meets its accessibility objectives effectively. Future enhancements could involve expanding language support capabilities, improving preprocessing techniques for low-quality images, and optimizing processing times further. This approach paves the way for more inclusive access to printed information, empowering visually impaired individuals in their daily interactions with textual content.

References

[1] Wang, Y., & Li, W. (2022). Matlab-Based Reading System for the Blind. Academic Journal of Computing & Information Science, 5(4), 56-60. DOI: 10.25236/AJCIS.2022.050410.

[2] Isewon, I., Oyelade, J., & Oladipupo, O. (2014). Design and Implementation of Text To Speech Conversion for Visually Impaired People. International Journal of Applied Information Systems, 7(2), 25-30. DOI: 10.5120/ijais14-451143.

[3] Lemmetty, S. (1999). Review of Speech Synthesis Technology. Master's Dissertation, Helsinki University of Technology.

[4] Swathi, G., Mai, C.K., & Babu, B.R.(2013). Speech Synthesis System for Telugu Language.International Journal of Computer Applications, 81(5).

[5] Ngugi, K., Okelo-Odongo, W., & Wagacha P.W.(2005). Swahili Text-To-Speech System.African Journal of Science and Technology, 6(1), 80–89.