I.HEAR.YOU: A WEB-BASED IMAGE-TO-SPEECH CONVERTER APPLICATION USING OPTICAL CHARACTER RECOGNITION AND SPEECH SYNTHESIS

Mulya Fajar Ningsih Alwi (001202000101)

Advisor: Rosalina, S.Kom., M.Kom

Informatics | 2023

President University

CONTENTS

Q **BACKGROUND**

Q **PROBLEM**

Q **OBJECTIVES**

Q **METHODOLOGY**

Q **APPLICATION UI**

Q **CONCLUSION**

Q **FUTURE WORKS**

BACKGROUND

Technology has become immensely helpful in addressing daily human challenges, expanding beyond limitations. Image-to-speech technology, enabling recognition of image content and audio playback, exemplifies this progress. Advancements in optical character recognition and speech synthesis techniques have significantly improved the accuracy and efficiency of this technology. However, the complexity of visual information remains a significant obstacle, requiring the integration of optical character recognition algorithms and speech synthesis techniques to analyze visual content and generate meaningful audio descriptions accurately. Despite challenges, image-to-speech technology has the potential to revolutionize digital content interaction and consumption. Indonesia faces significant literacy issues, ranking low in global literacy rates. To address this, the proposed project aims to develop a technology leveraging machine learning and artificial intelligence to help illiterate individuals comprehend information from images. By inputting image files, the machine will recognize and convert them into audio files for easy understanding and accessibility.

PROBLEM

Individuals with visual impairments or even illiterate people have faced numerous challenges in accessing visual information, including images on the internet or even in their surroundings. While screen readers are commonly used to convert text-to-speech, there is a lack of technology available to convert images-to-speech in a meaningful way or in the easiest way for impaired people.

This final project aims to solve the problem of how to help visually impaired individuals or even illiterate people to be able to find out the information that contains in the image around them by hearing aloud the audio from the image in the easiest way using a user-friendly application.
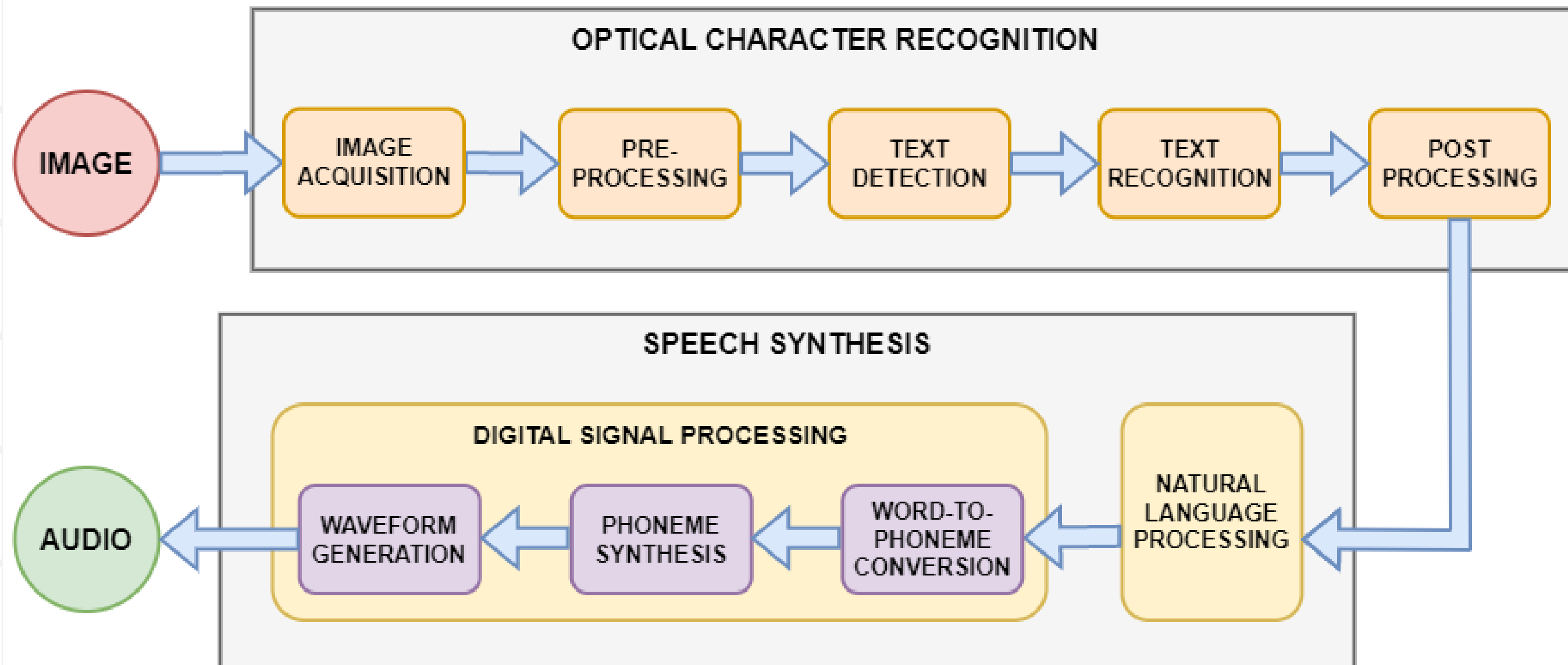
OBJECTIVES

This final project aims to develop an image-to-speech web-based application that will enable visually impaired individuals or illiterate people to access information contained in the image from several sources in the easiest way using a user-friendly application.

The application will utilize the Optical Character Recognition (OCR) technique to analyze the image's content and generate natural language descriptions that can be read aloud to the user in a human language with different languages supported using a Text-to-Speech (TTS) Synthesizer technique by using its two main components, such as Natural Language Processing (NLP) and Digital Signal Processing (DSP). The application also has a feature that can translate the detected text into various languages supported by the application and its audio file.

By addressing this gap in accessibility technology, the application will enhance the independence and quality of life for visually impaired and illiterate individuals.

APPLICATION UI

HOME PAGE UI

ABOUT PAGE UI

Image-to-Speech Converter Web App | Informatics | 2023 | President University

APPLICATION UI

UPLOAD PAGE UI

IMAGE DECODED PAGE UI
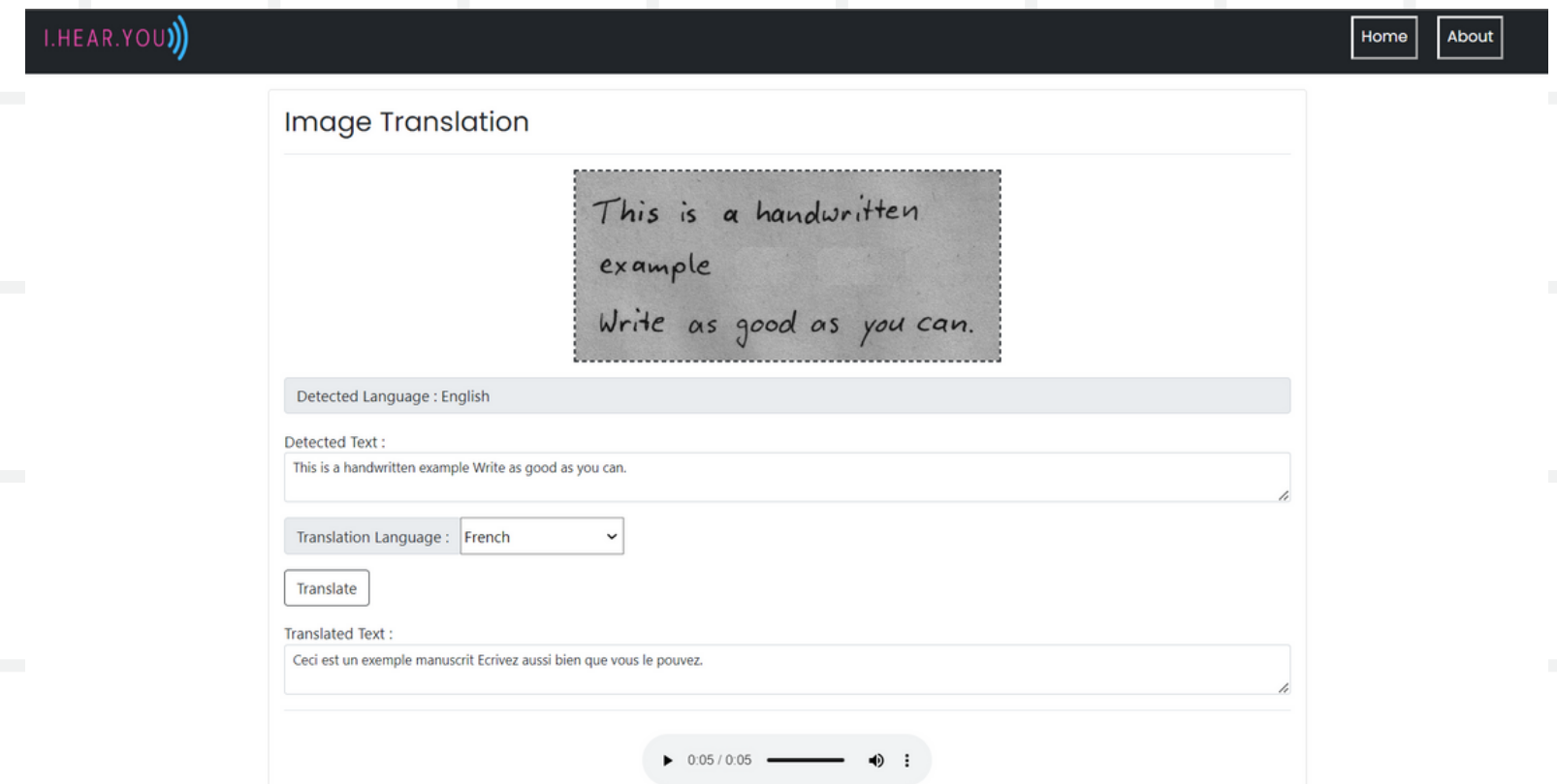
Image-to-Speech Converter Web App | Informatics | 2023 | President University

CONCLUSION

In conclusion, the use of Optical Character Recognition (OCR) and Speech Synthesis technology works as intended in the web application and gives the desired output from the beginning to the end. From the testing experiments focused on calculating the accuracy of OCR performance on the sample images, the web application achieved an average accuracy rate of 0.976, an average precision of 0.982, and an average recall of 0.970, showcasing its effectiveness in accurately converting images to text. The experiments highlight the successful implementation of OCR techniques and provide valuable insights for improving the web application's performance in real-world scenarios. Hopefully, the user of this web application can get an unforgettable and enjoyable experience when using the web application by utilizing all features that are available in the web application, especially in enabling visually impaired individuals or illiterate people to access information that is contained in the image form around them in the easiest way possible using this user-friendly web application.

FUTURE WORKS

The web application works as intended but there are some aspects that can be improved to make the web application even better such as improving the web application into a mobile-based application that can be used online, allowing to upload and process multiple images, adding a feature that allows image detection by taking images from the camera, adding more languages to be used in a translation feature, increasing the accuracy and the processing speed of the OCR engine in recognizing the image's character, especially for the image that contains a non-English language or a more complex character, and improving the user interface into a more simple and easy-to-use user interface in order to give the best user experience for the visually impaired people when using the web application.

THANK YOU

Presentation by Mulya Fajar Ningsih Alwi

President University