Report On

Image to Voice Converter

v

Submitted in partial fulfillment of the requirements of the Course project in Semester IV of Second Year Computer Engineering

by

Vaibhav Tatkare (Roll No. 49)

Kunal Ushinkar (Roll No. 55)

Supervisor

Prof. Sneha Mhatre

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**



**(2023-24)**

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**

**CERTIFICATE**

This is to certify that the project entitled “Image to voice converter” is a Bonafide work of Vaibhav Tatkare (Roll No. 49) Kunal Ushinkar (Roll No. 55) submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in semester IV of Second Year Computer Engineering.

**Supervisor**

Prof. Sneha Mhatre

Internal Examiner External Examiner

Dr Megha Trivedi

Head of Department

Dr. H.V. Vankudre

Principal

**Index**

1. Abstract
2. Problem Statement
3. Module Description
4. Block Diagram

5.Code

6.Output

7.Conclusion

8.Reference

9.Code

**Abstract:**

The Image to Voice Converter project aims to develop a Python application that can convert images into spoken words. This project will utilize the Google Cloud Vision API for image recognition and the Google Text-to-Speech API for converting recognized text into speech. The user will be able to upload an image through a user-friendly interface, and the application will process the image to extract any text present. Once the text is extracted, it will be converted into speech using the Text-to-Speech API, and the spoken words will be played back to the user. This project demonstrates the integration of multiple APIs to create a useful and innovative application for converting visual information into audible format.

1

**Problem Statement:**

Develop an Image to Voice Converter application using Python that allows users to convert images containing text into spoken words. The application should provide a user-friendly interface for uploading images and should utilize the Google Cloud Vision API for text extraction from images and the Google Text-to-Speech API for converting the extracted text into speech. The goal is to create a tool that can assist individuals with visual impairments or those who prefer auditory information in accessing textual content from images. The application should be efficient, accurate, and easy to use, providing a seamless conversion process from image to voice.

**Module Description:**

1. Image Input Module:

- Description: This module handles the input of images containing text.

- Functionality: Allows users to upload an image or provide a file path to an image.

- Output: Returns the image data for further processing.

2. Text Extraction Module:

- Description: This module extracts text from the input image using Pytesseract.

- Functionality: Utilizes Pytesseract to perform OCR on the image and extract text.

- Output: Returns the extracted text.

3. Text Preprocessing Module:

- Description: This module preprocesses the extracted text for better readability and accuracy.

- Functionality: Removes unnecessary characters, corrects spelling mistakes, and formats the text.

- Output: Returns the preprocessed text.

4. Text-to-Speech Conversion Module:

- Description: This module converts the preprocessed text into speech.

- Functionality: Uses a text-to-speech (TTS) engine, such as the `gTTS` library, to generate an audio file.

- Output: Returns the audio file containing the spoken words.

5. Audio Output Module:

- Description: This module handles the output of the audio file.

- Functionality: Plays the audio file for the user to listen to the converted text.

- Output: Audio playback for the user.

3

6. User Interface Module:

- Description: This module provides a user-friendly interface for interacting with the converter.

- Functionality: Allows users to upload an image, initiate the conversion process, and listen to the audio output.

- Output: Visual feedback and audio playback controls for the user.

7. Error Handling Module:

- Description: This module manages errors that may occur during the conversion process.

- Functionality: Provides feedback to the user in case of errors and suggests corrective actions.

- Output: Error messages and instructions for the user to resolve issues.

8. Main Controller Module:

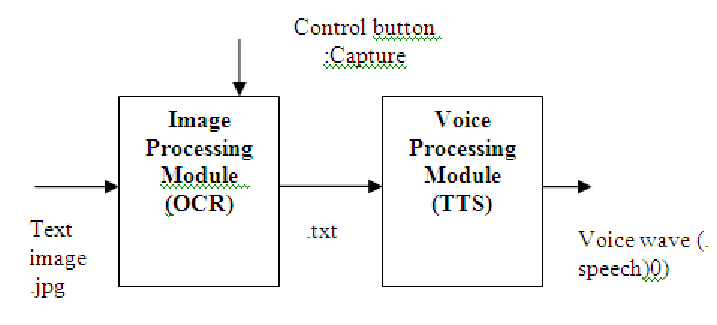
- Description: This module acts as the main controller, orchestrating the flow of the conversion process.

- Functionality: Coordinates the interactions between the various modules to convert image text to speech.

- Output: Controls the overall conversion process and manages the flow of data between modules.

4

**Block Diagram**



Block Diagram: Image To Voice

Code:-

from PIL import Image

from gtts import gTTS

import pytesseract

from pytesseract import image\_to\_string

pytesseract.pytesseract.tesseract\_cmd = r'C:\Program Files\Tesseract-OCR\tesseract.exe'

def image\_to\_sound(path\_to\_image):

"""

Function for converting an image to sound

"""

try:

loaded\_image = Image.open(path\_to\_image)

decoded\_text = image\_to\_string(loaded\_image)

cleaned\_text = " ".join(decoded\_text.split("\n"))

print(cleaned\_text)

sound = gTTS(cleaned\_text, lang="en")

sound.save("newsound.mp3")

return True

except Exception as bug:

print("The bug thrown while excuting the code\n", bug)

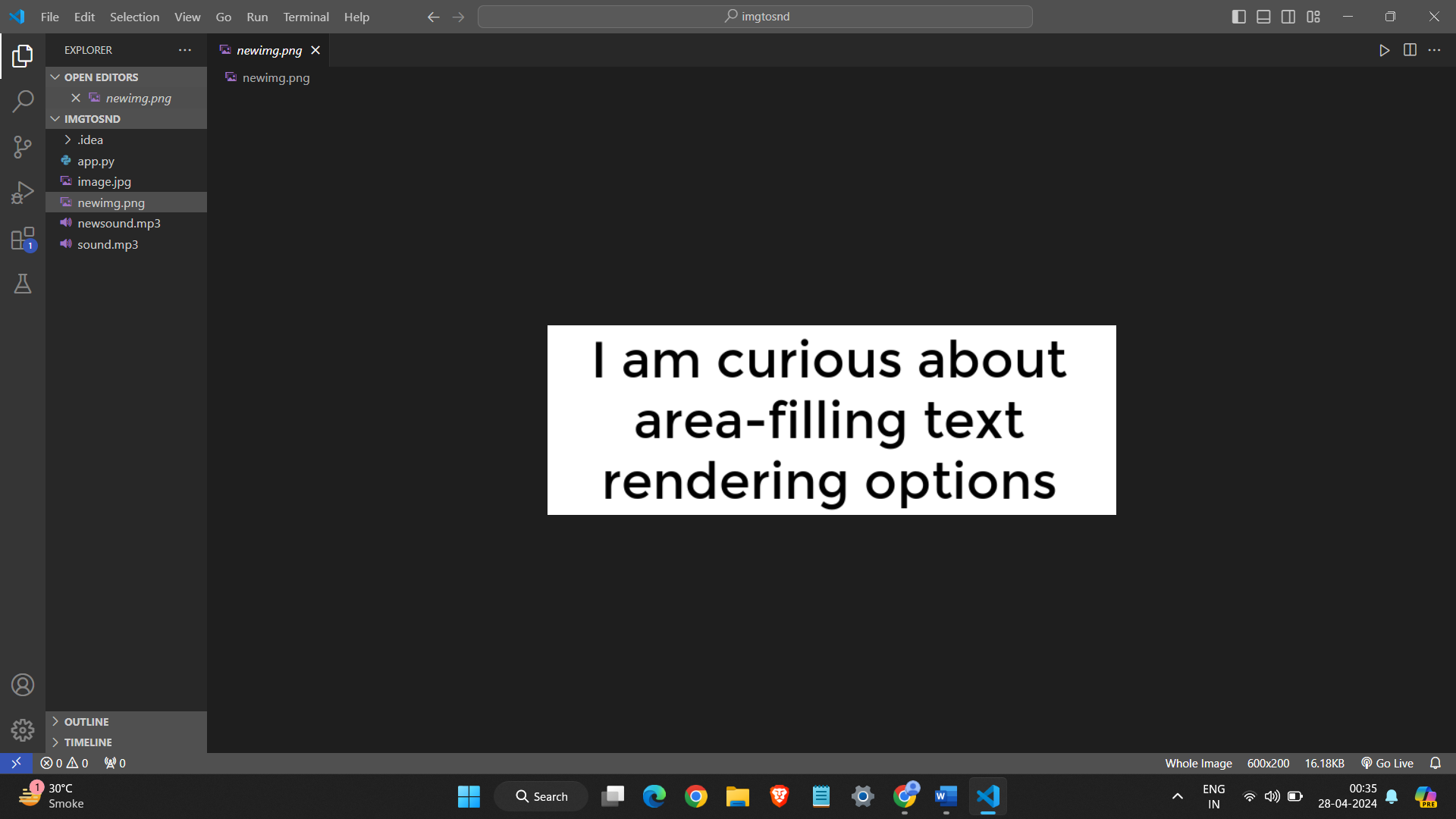
return

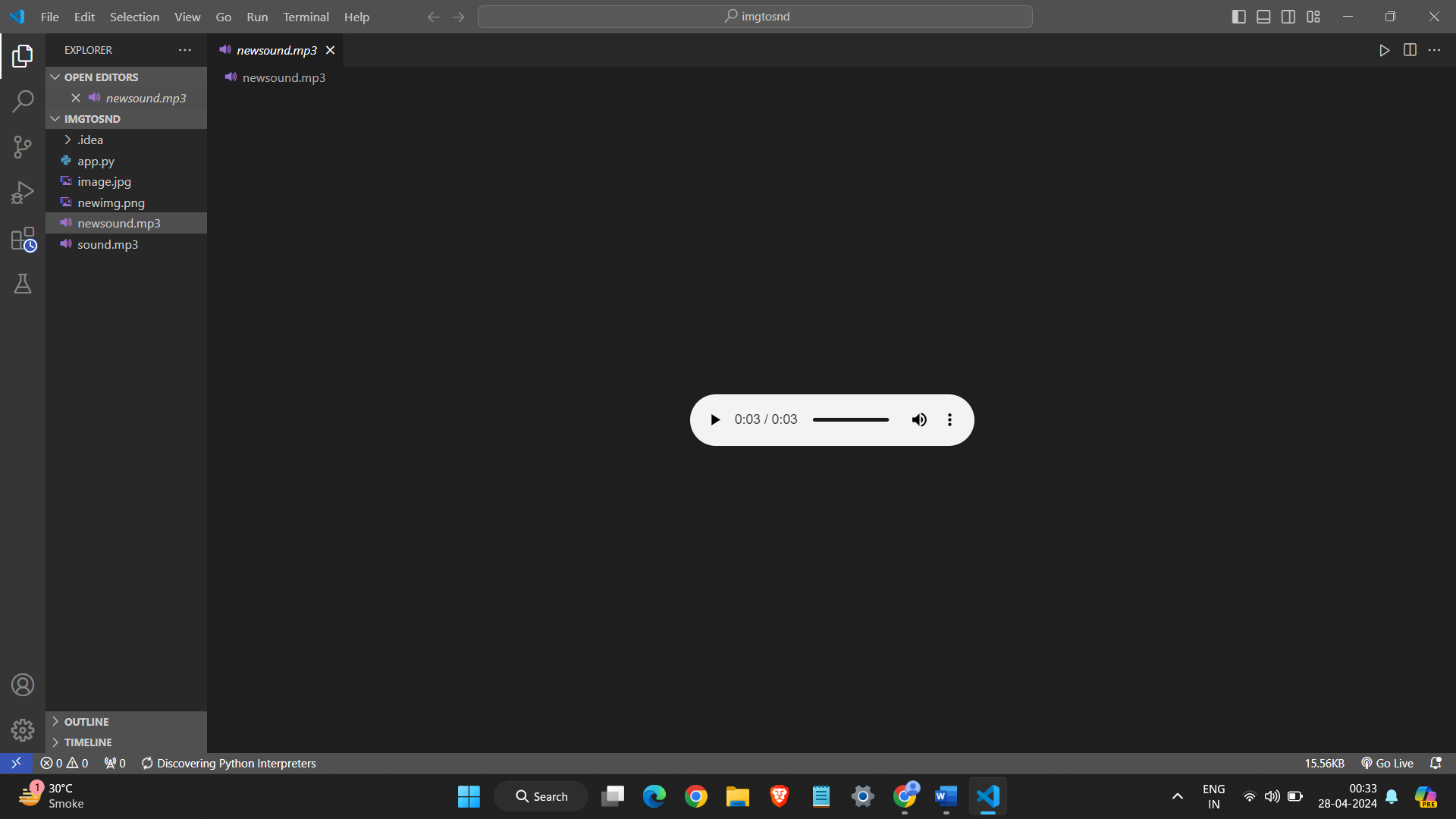
if \_\_name\_\_ == "\_\_main\_\_":

image\_to\_sound("newimg.png")

input()

Results:-





**Conclusion:**

The Image to Voice Converter project using Pytesseract in Python provides a valuable tool for converting text in images into spoken words. By leveraging the power of OCR and TTS technologies, the project enhances accessibility for visually impaired individuals and provides a convenient way to convert text in images into an audio format for easy consumption.

The modular design of the project ensures flexibility and scalability, allowing for easy integration of additional features and improvements. The user-friendly interface makes it simple for users to upload an image, initiate the conversion process, and listen to the audio output.

Overall, the project demonstrates the practical applications of OCR and TTS technologies in enhancing accessibility and usability for a wide range of users. It serves as a foundation for further development and exploration in the field of image-to-speech conversion and accessibility solutions.

**References:**

Stackoverflow, tutorialpoints , wikepidiea, Openai