# TEXT TO SPEECH

A Mini Project Report

submitted by

## MUHAZEENA FATHIM (MES22MCA-2038)

**to**

the APJ Abdul Kalam Technological University
in partial fulfillment of the requirements for the award of the Degree

of

Master of Computer Applications



## Department of Computer Applications

MES College of Engineering
Kuttippuram, Malappuram - 679582

December 2023

# DECLARATION

I undersigned hereby declare that the project report TEXT TO SPEECH, submitted for partial fulfilment of the requirements for the award of the degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala, is a bonafide work done by me under supervision of Mr Muhammad Jabir C, Assistant Professor, Department of Computer Applications. This submission represents my ideas in my own words and where ideas or words of others have been included. I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to the ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed as the basis for the award of any degree, diploma or similar title of any other University..

Place:

Date

MUHAZEENA FATHIMA

(MES22MCA-2038)

# DEPARTMENT OF COMPUTER  APPLICATIONS
# MES COLLEGE OF ENGINEERING, KUTTIPPURAM



## CERTIFICATE

This is to certify that the report entitled **TEXT TO SPEECH** is a bona fide record of the MiniProject work during the year 2023-24 carried out by **MUHAZEENA FATHIMA (MES22MCA-2038)** submitted to the APJ Abdul Kalam Technological University, in partial fulfillment of the requirements for the award of the Master of Computer Applications, under my guidance and supervision. This report in any form has not been submitted to any other University or Institution for any purpose.

Internal Supervisor(s)                                                                    External Supervisor(s)

PG Coordinator                                                                                HEAD OF THE DEPT

# Acknowledgements

My endeavor stands incomplete without dedicating my gratitude to a few people who have contributed towards the successful completion of our project. I pay our gratitude to the Almighty for His invisible help and blessing for the fulfillment of this work. At the outset i express my heart full thanks to my Head of the Department, Prof. HYDERALI K for permitting me to do this project. I take this opportunity to express my profound gratitudeto MR. Muhammad Jabir C, my project guide for his valuable support. I also take this opportunityto thank my project coordinator MS. Febin Aziz for her timely advice and strict schedules to complete my project. I'm also grateful to all my teaching and non- teaching staff fortheir encouragement, guidance and whole-hearted support. Last but not least, I am gratefully indebted to  myfamily and friends, who gave me a precious help in doing  my project

**MUHAZEENA FATHIMA (MES22MCA-2038)**

# Abstract

Text-to-speech (TTS) technology is a cutting-edge application that converts written text into spoken words, offering a revolutionary solution for accessibility, communication, and content consumption. This transformative technology employs sophisticated algorithms to analyze and interpret textual data, generating natural-sounding speech with intonations and cadences that closely mimic human voices. TTS has profound implications for individuals with visual impairments, as it enables them to access and comprehend written information effortlessly. Beyond accessibility, TTS finds widespread use in various sectors such as navigation systems, virtual assistants, and audiobook production, enhancing user experiences with different accents and facilitating seamless interaction between humans and technology. As advancements continue, the synthesis of speech through TTS is poised to play an increasingly pivotal role in bridging communication gaps and fostering inclusivity in the digital landscape.

# Contents

*CONTENTS*

8

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Text-to-speech (TTS) technology stands at the forefront of a transformative era in human-computer interaction, revolutionizing the way we engage with written content. This innovative application harnesses the power of sophisticated algorithms to seamlessly convert written text into natural-sounding spoken words, breaking down barriers for accessibility and communication.

## 1.1   Purpose

The primary purpose of text-to-speech (TTS) technology is to facilitate the conversion written text into spoken words, thereby enhancing accessibility, communication, and user experience across various applications. TTS plays a pivotal role in making digital content more inclusive, allowing individuals with visual impairments or reading difficulties to access information effortlessly. Additionally, it contributes to the development of user-friendly interfaces in navigation systems, virtual assistants, and interactive technologies.

## 1.2 Objective

The objective of this application is:

● **User Experience Enhancement:** Improve user experiences in devices, virtual assistants, and interactive technologies by providing natural-sounding and human-like speech..

● **Accessibility:** Enable individuals with visual impairments or reading difficulties to access written information through spoken words.

● **Inclusivity:** Foster a more inclusive digital environment by breaking down barriers to information access.

● **Alternative Content Consumption:** Facilitate alternative modes of content consumption, such as audiobooks and voiceovers.

● **Advancements in Technology:** Adapt and contribute to ongoing advancements in artificial intelligence and human-computer interaction.

● **Communication Technologies:** Play a role in enhancing communication technologies through voice-activated systems and interactive applications

## 1.3 Contribution

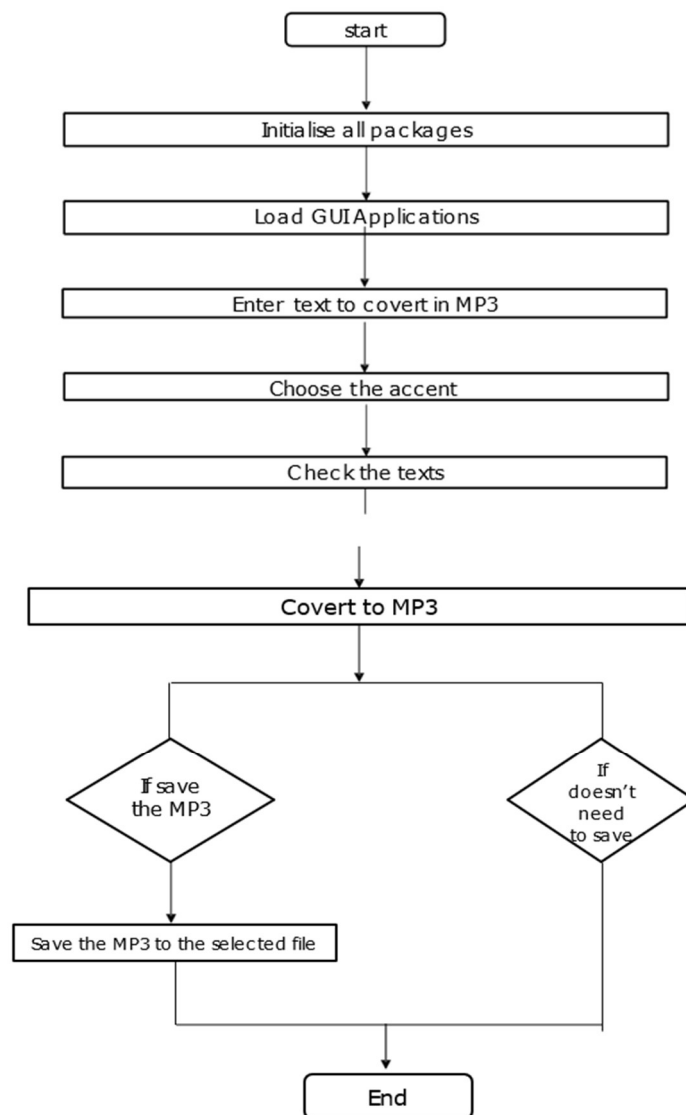The major contribution in this project are: to digital inclusivity and user experience by addressing accessibility challenges and enhancing communication across various applications.

## 1.4 Report Organization

The project report is divided into three sections. Section 2: describes the methodology used for implementing the project. Section 3: gives the results and discussions. Finally, Section 4 gives the conclusion.

*REPORT ORGANIZATION*

## Process Flow Diagram – LEVEL 0

```
                          ┌──────────────┐
                          │    start     │
                          └──────┬───────┘
                                 │
                                 ▼
              ┌──────────────────────────────────────┐
              │      Initialise all packages          │
              └──────────────────┬───────────────────┘
                                 │
                                 ▼
              ┌──────────────────────────────────────┐
              │         Load GUI Applications          │
              └──────────────────┬───────────────────┘
                                 │
                                 ▼
              ┌──────────────────────────────────────┐
              │      Enter text to covert in MP3       │
              └──────────────────┬───────────────────┘
                                 │
                                 ▼
              ┌──────────────────────────────────────┐
              │           Choose the accent            │
              └──────────────────┬───────────────────┘
                                 │
                                 ▼
              ┌──────────────────────────────────────┐
              │            Check the texts             │
              └──────────────────────────────────────┘
                                 │
                                 ▼
        ┌──────────────────────────────────────────────────┐
        │                  Covert to MP3                    │
        └──────────────────────────────────────────────────┘
```

         If save                              If doesn't
        the MP3                              need to save

   Save the MP3 to the selected file

                          End

# Chapter 2

# Methodology

## 2.1  Introduction

In the methodology section for a text-to-speech (TTS) study, the focus is on detailing the systematic approach undertaken to investigate or develop TTS technology. The introduction to the methodology provides a comprehensive overview of the research design, data sources, and the analytical techniques utilized in the study. It typically includes a justification for the chosen TTS development approach or investigation method, considerations regarding the linguistic and technological aspects, and the rationale behind selecting specific algorithms or models. Additionally, this section may delve into the evaluation metrics employed to assess the performance of the TTS system, considering factors such as naturalness, intelligibility, and efficiency. The transparency of the methodology is crucial for understanding the steps taken to achieve the research objectives and for assessing the reliability and applicability of the TTS technology developed or studied.

## 2.2 User story

A key component of agile software development is putting people first, and user-stories put actual end users at the center of the conversation. Stories use non-technical language to provide context for the development team and their efforts. After reading a user story, the team knows why they are building what they're building and what value it creates. A user story is a tool used in agile software development to capture a description of a software feature from an end-user perspective. The user story describes the type of user, what they want and why. A user story helps to create a simplified description of a requirement. User stories are one of the core components of an agile program. They help provide a user-focused framework for daily work. which drives collaboration, creativity, and a better product overall.

| UserStoryID | As a <type of user> | I want to | So that I can |
|---|---|---|---|
| 1 | user | Enter text | Enter the text and coverted into speech |
| 2 | user | Adjust the speed | Can adjust speed as slow,normal and fast |
| 4 | user | Adjust the volume and pitch | We can adjust the volume and pitch as you want |
| 3 | user | Save the speech | Can save the speech as MP3 |

## 2.4   Product backlog

A product backlog is a list of the new features, changes to existing features, bug fixes, infrastructure changes or other activities that a team may deliver to achieve  a specific outcome. The product backlog is the single authoritative source for things that a team works on. That means that nothing gets done that isn't on the product backlog. Conversely, the presence of a product backlog item on a product backlog does not guarantee that it will be delivered. It represents an option the team has for delivering a specific outcome rather than a commitment. It should be cheap and fast to add a product backlog item to the product backlog, and it should be equally as easy to remove a product backlog item that does not result in direct progress to achieving the desired outcome or enable progress toward the outcome. TheScrum Product Backlog is simply a list of all things that need to be done within the project. It replaces the traditional requirements specification artifacts. These items can have a technical nature or can be user-centric e.g., in the form of user stories.

| ID | PRIORITY | SIZE (Hours) | SPRINT | STATUS | NAME |
|----|----------|--------------|--------|--------|------|
| 1 | Medium | 8 | 1 | completed | Form Design |
| 4 | High | 23 | 2 | completed | Coding |
| 5 | Medium | 10 | 3 | completed | Testing Data |
| 6 | High | 9 | | completed | Output Generation |

## 2.5   Project Plan

A project plan that has a series of tasks laid out for the entire project, listing task durations, responsibility assignments, and dependencies. Plans are developed in this manner based on the assumption that the Project Manager, hopefully along with the team, can predict up front everything that will need to happen in the project, how long it will take, and who will be able to do it.

| ID | TASK NAME | START DATE | END DATE | PROJECT HOURS | STATUS |
|----|-----------|-----------|----------|---------------|--------|
| 1 | Sprint 1 | 15/09/2023 | 12/10/2023 | 13 | Completed |
| 2 | Sprint 2 | 13/10/2023 | 03/11/2023 | 14 | Completed |
| 3 | Sprint 3 | 08/11/2023 | 30/11/2023 | 14 | Completed |

# 2.6   Sprint backlog(plan)

The sprint backlog is a list of tasks identified by the Scrum team to be completed during the Scrum sprint. During the sprint planning meeting, the team selects a number of product backlog items, usually in the form of user stories, and identifies the tasks necessary to complete each user story. Most teams also estimate how many hours each task will take someone on the team to complete.

This project has three sprints: -

SPRINT 1 (PLAN)

| Backlog Item | Status and Completion date | Original Estimate in hours | Day 1 15/09 | Day 2 20/09 | Day 3 21/09 | Day 4 28/09 | Day 5 29/09 | Day 6 04/10 | Day 7 05/10 | Day 8 06/10 | Day 9 11/10 | Day 10 12/10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Form Design | 15/09/2023 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Coding | 06/10/2023 | 8 | 0 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 0 | 0 |
| Testing & Validation | 12/10/2023 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **Total** | | 13 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 |

Table – 1

SPRINT 2 (PLAN)

| Backlog Item | Status and Completion date | Original Estimate in hours | Day 1 13/10 | Day 2 18/10 | Day 3 19/10 | Day 4 20/10 | Day 5 25/10 | Day 6 26/10 | Day 7 27/10 | Day 8 01/11 | Day 9 02/11 | Day 10 03/11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Form Design | 13/10/2023 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Coding | 01/11/2023 | 9 | 0 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 0 | 0 |
| Testing & Validation | 03/11/2023 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| **Total** | | 14 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 |

Table -2

## 2.6. SPRINT BACKLOG(PLAN)

SPRINT 3 (PLAN)

| Backlog Item | Status and Completion date | Original Estimate in hours | Day 1 08/11 | Day 2 09/11 | Day 3 10/11 | Day 4 15/11 | Day 5 16/11 | Day 6 17/11 | Day 7 22/11 | Day 8 23/11 | Day 9 24/11 | Day 10 29/11 | Day 11 30/11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Coding | 01/11/2023 | 9 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 0 |
| Testing & Validation | 03/11/2023 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 1 |
| **Total** | | 14 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 |

Table -3

# Chapter 3

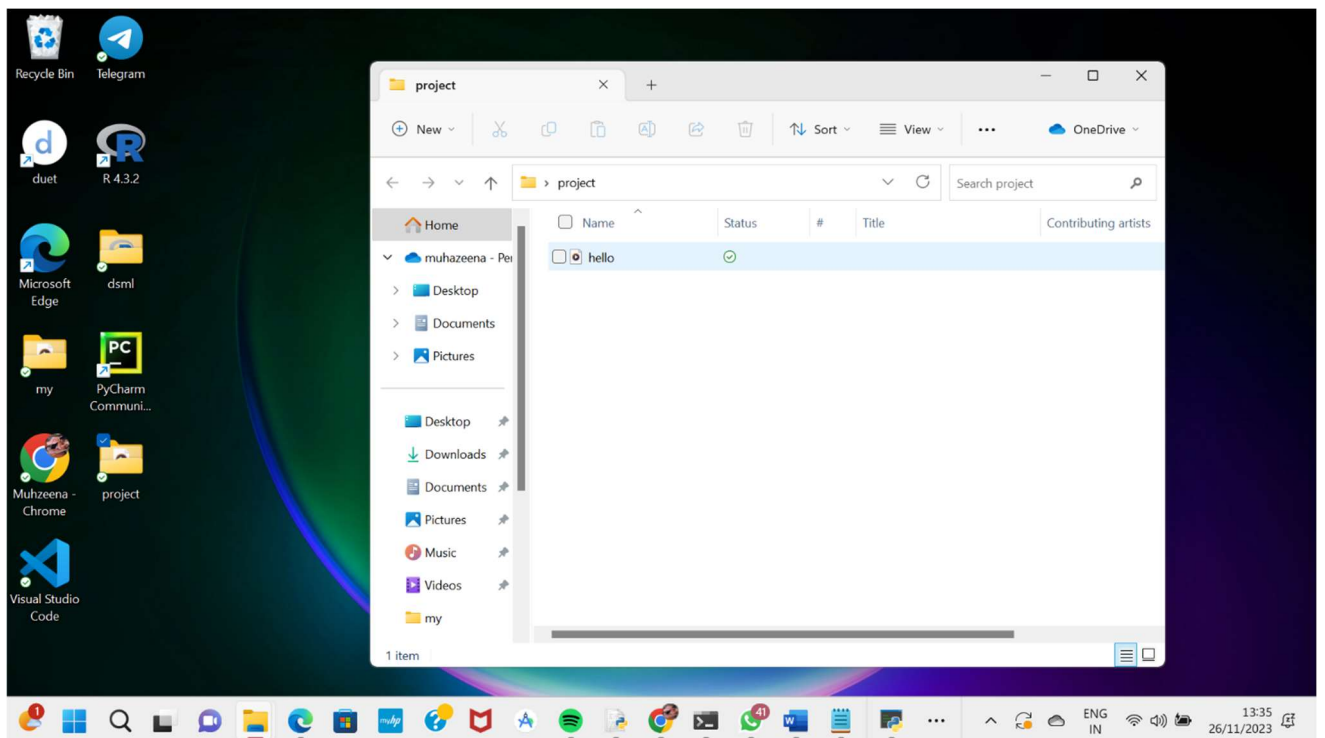# Results and Discussions

## 3.1   Results

**MAIN PAGE**
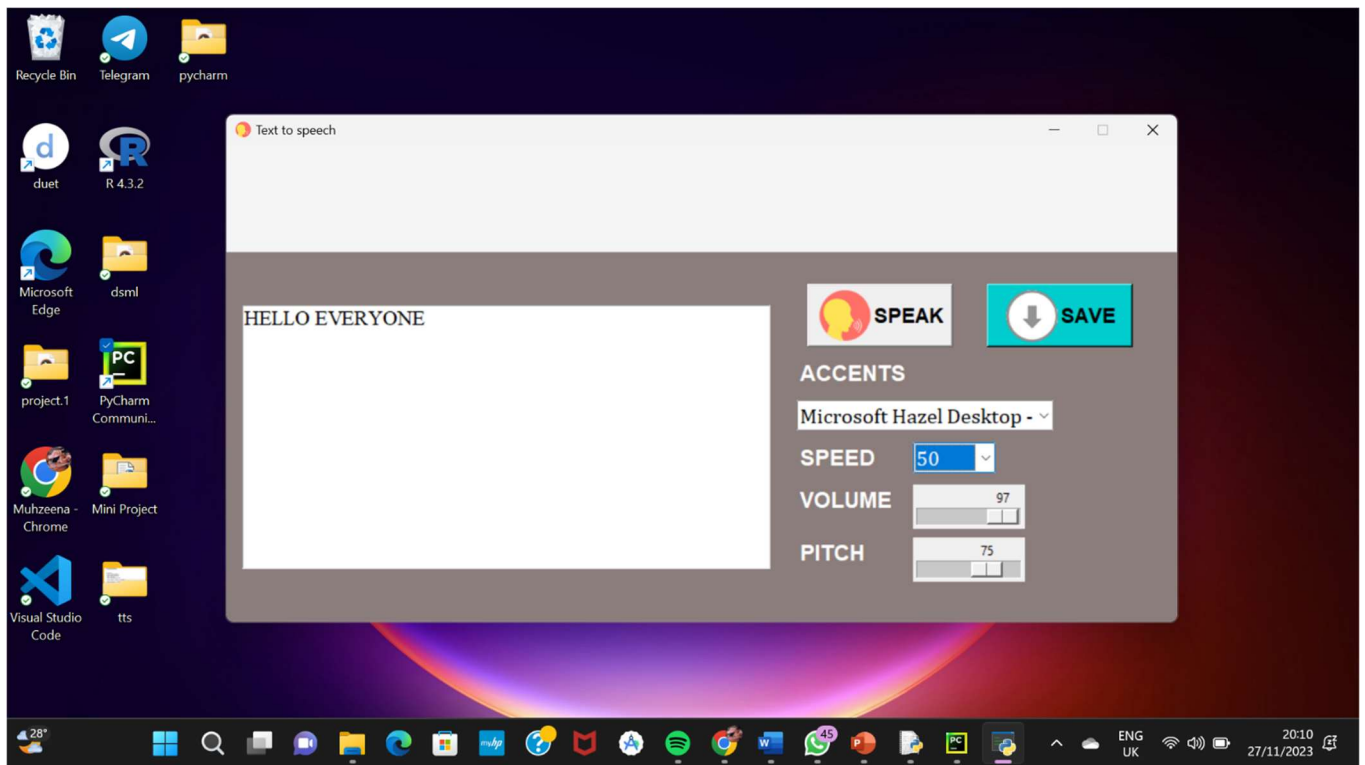
## 3.1. RESULTS

**DOWNLOAD**
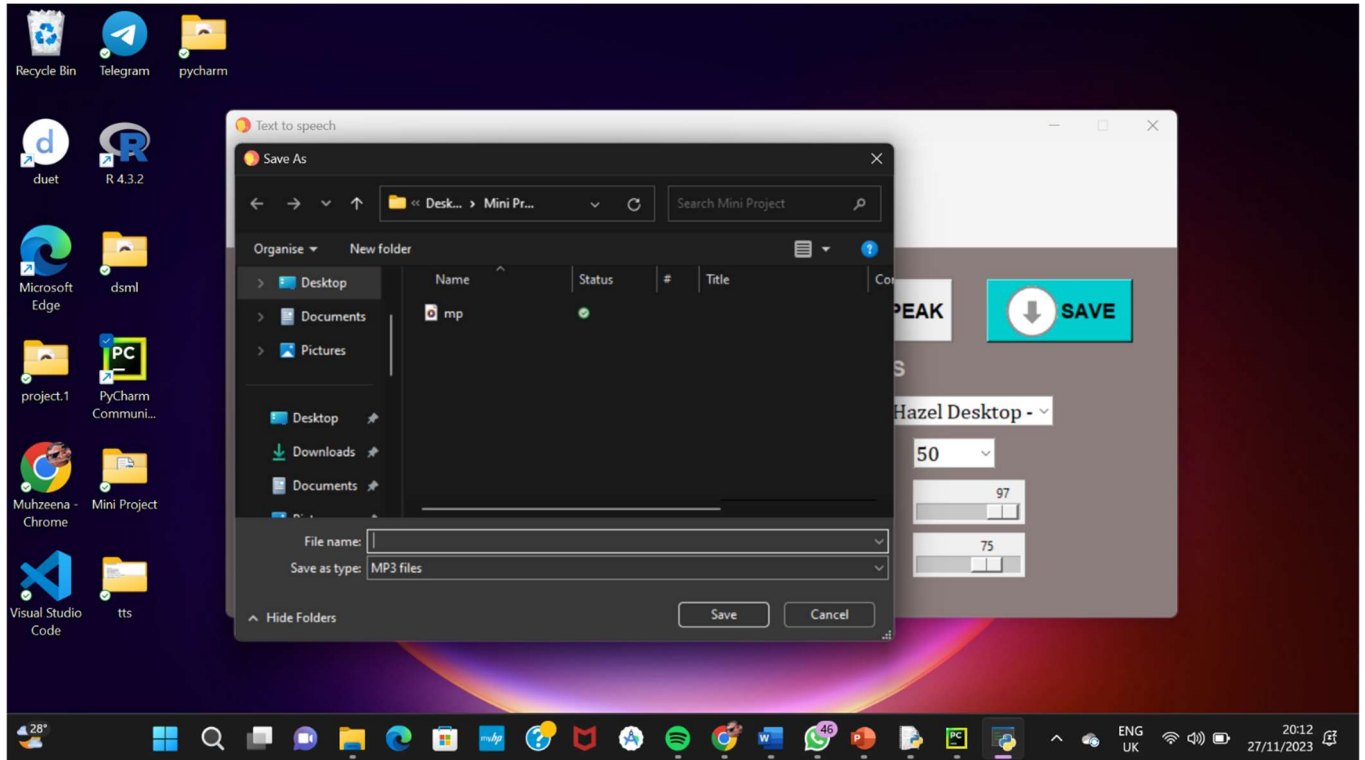


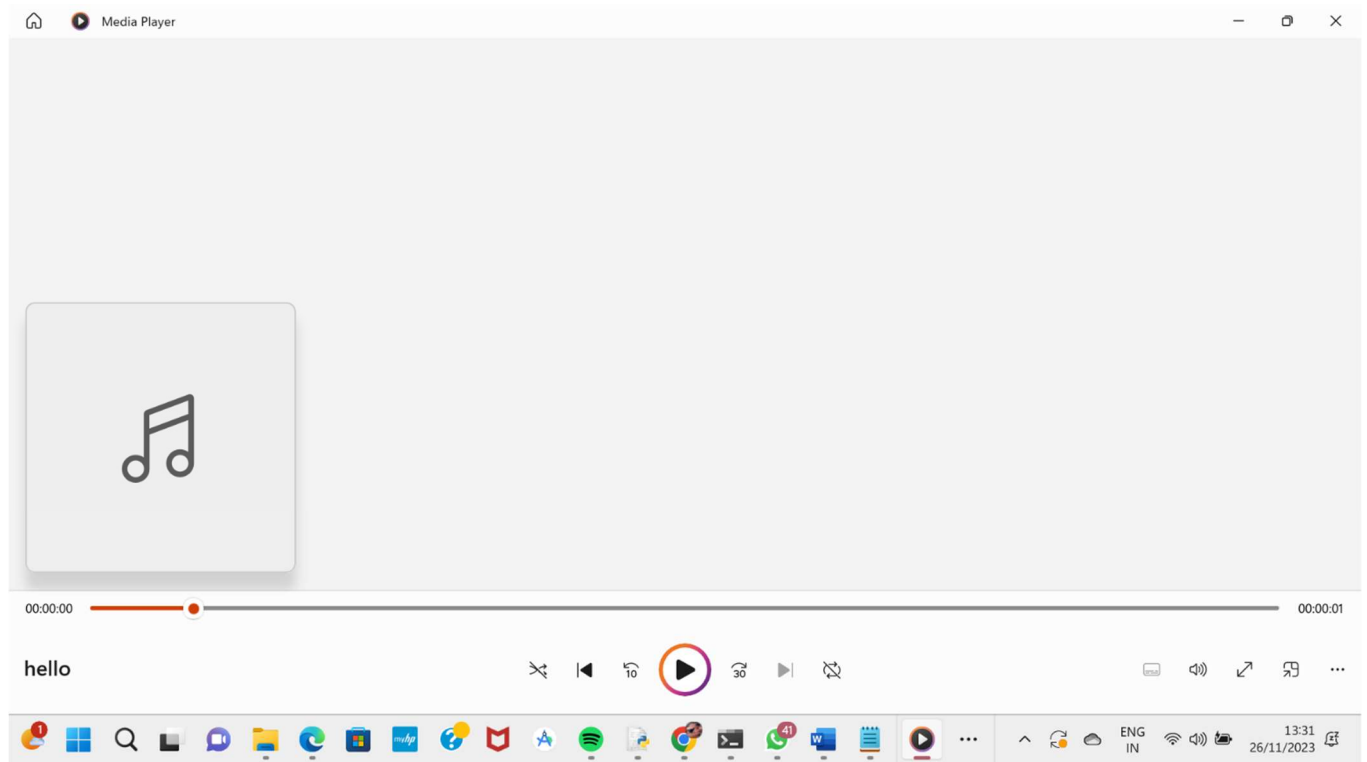**SAVED FILE**

## 3.1. RESULTS

**MP3 FILE**

# Chapter 4

# Conclusion

The  In conclusion, the Text-to-Speech (TTS) project with support for nine different accents—Korean, Chinese, Spanish, Japanese, French, and more—proves to be a versatile and inclusive solution for users worldwide. By incorporating diverse accents, the project enhances the naturalness and authenticity of the generated speech, making it more relatable and accessible to a global audience.

The inclusion of Korean, Chinese, Spanish, Japanese, French, and other accents not only broadens the project's linguistic capabilities but also caters to the cultural diversity of users. This ensures a more personalized and immersive experience, making the TTS system more adaptive and user-friendly.

Furthermore, the ability to download the generated voice as an MP3 file adds a layer of convenience for users who may want to use the synthesized speech in various applications, such as podcasts, videos, or personal projects. This feature allows users to seamlessly integrate the TTS output into their creative endeavors, expanding the utility of the project beyond just text conversion.

In essence, the Text-to-Speech project with diverse accents and downloadable MP3 functionality stands as a testament to the commitment to inclusivity, user experience, and practicality. The project not only breaks language barriers but also empowers users to engage with synthesized speech in a manner that suits their preferences and creative needs

# References

[1]      python: https://www.javatpoint.com/how-to-convert-text-to-speech-in-python

[2]      https://pypi.org/project/pyttsx3/

[3]      https://www.youtube.com/watch?v=fEL8ihL-GXg&t=965s

[4]      https://www.youtube.com/watch?v=J9Jy43oTTDc

[5]      tkinter: https://docs.python.org/3/library/tkinter.html

# Appendix

## Source Code

```
from tkinter import *
from tkinter.ttk import Combobox
from tkinter import filedialog
import pyttsx3
import os

root = Tk()
root.title("Text to speech")
root.geometry("900x450+200+100")
root.resizable(False, False)
root.configure(bg="#8B7D7B")

# Initialize pyttsx3 engine
engine = pyttsx3.init()

# Get available voices
voices = engine.getProperty('voices')

# Function to be called when the Speak button is pressed
def speaknow():
    text = text_area.get(1.0, END)
    selected_voice_index = voice_combobox.current()
    speed = speed_combobox.get()
    volume = volume_scale.get()
    pitch = pitch_scale.get()

    # Set voice properties
    engine.setProperty('voice', voices[selected_voice_index].id)
    engine.setProperty('rate', int(speed))
    engine.setProperty('volume', volume / 100.0)
    engine.setProperty('pitch', pitch / 100.0)

    # Speak the text
    engine.say(text)
    engine.runAndWait()

# Function to be called when the Download button is pressed
# Function to be called when the Download button is pressed
def download():
    text = text_area.get(1.0, END)
    selected_voice_index = voice_combobox.current()
    speed = speed_combobox.get()
    volume = volume_scale.get()
    pitch = pitch_scale.get()
```

```python
    # Set voice properties
    engine.setProperty('voice', voices[selected_voice_index].id)
    engine.setProperty('rate', int(speed))
    engine.setProperty('volume', volume / 100.0)
    engine.setProperty('pitch', pitch / 100.0)

    # Ask the user for the file name and location
    file_name = filedialog.asksaveasfilename(defaultextension=".mp3", filetypes=[("MP3 files", "*.mp3")])

    # Check if the user canceled the file dialog
    if not file_name:
        return

    # Save the text as an MP3 file
    engine.save_to_file(text, file_name)
    # Speak the text
    #engine.say(text)
    engine.runAndWait()


# Icon
image_icon = PhotoImage(file="speak.png")
root.iconphoto(False, image_icon)

# Top frame
Top_frame = Frame(root, bg="#F5F5F5", width=900, height=100)
Top_frame.place(x=0, y=0)

# Labels and Comboboxes
Label(root, text="ACCENTS", font="arial 16 bold", bg="#8B7D7B", fg="white").place(x=540, y=200)

# Text area
text_area = Text(root, font="Palatino ", bg="white", relief=GROOVE, wrap=WORD)
text_area.place(x=15, y=150, width=500, height=250)

# Add voice selection combobox
voice_combobox = Combobox(root, values=[voice.name for voice in voices], font="cambria 15", state='r', width=20)
voice_combobox.place(x=540, y=240)
voice_combobox.set(voices[0].name)

# Add speed, volume, and pitch controls
Label(root, text="SPEED", font="arial 16 bold", bg="#8B7D7B", fg="white").place(x=540, y=280)
speed_combobox = Combobox(root, values=['50', '150', '200'], font="cambria 15", state='r', width=5)
speed_combobox.place(x=650, y=280)
speed_combobox.set('150')

Label(root, text="VOLUME", font="arial 16 bold", bg="#8B7D7B", fg="white").place(x=540, y=320)
volume_scale = Scale(root, from_=0, to=100, orient=HORIZONTAL, length=100)
volume_scale.place(x=650, y=320)
volume_scale.set(50)

Label(root, text="PITCH", font="arial 16 bold", bg="#8B7D7B", fg="white").place(x=540, y=370)
pitch_scale = Scale(root, from_=0, to=100, orient=HORIZONTAL, length=100)
pitch_scale.place(x=650, y=370)
pitch_scale.set(50)
```

```
# Buttons
image_icon = PhotoImage(file="speak.png")
root.iconphoto(False, image_icon)
speakbtn = Button(root, text="SPEAK", compound=LEFT, image=image_icon, width=130, font="arial 14 bold",
command=speaknow)
speakbtn.place(x=550, y=130)

# Download button
imageicon = PhotoImage(file="download.png")
root.iconphoto(False, image_icon)
savebtn = Button(root, text="SAVE", compound=LEFT, image=imageicon, width=130, bg="#00CDCD",
font="arial 14 bold", command=download)

savebtn.place(x=720, y=130)

root.mainloop()
```