Text to Speech Converter Using Python

[1]Sanjeev Kumar S, [2]Preksha C and [3]Pooja M
[1,2,3] Department of Electronics and Communication Engineering, AMC Engineering College, Bangalore, India.
[1]sanju9449806909@gmail.com, [2]preksha.c2001@gmail.com, [3]pooja9886451280@gmail.com

Article Info
Jenitta J and Swetha Rani L (eds.), *International Conference on VLSI, Communications and Computer Communication*, Advances in Intelligent Systems and Technologies,
Doi: https://doi.org/10.53759/aist/978-9914-9946-1-2_24
©2023 The Authors. Published by AnaPub Publications.

Abstract – In this paper, an innovative, efficient, and cost-effective technique was presented that allows users to hear the content of text images instead of reading them. This type of system helps visually impaired people to interact effectively with computers through a speech interface. Extracting text from color images is a challenging task in computer vision. Text-to-speech conversion is a method that scans and reads aloud English letters and numbers. Reading scientific literature is essential for researchers and clinicians. With the overabundance of medical and dental journals, it is important to develop a method to select and read the correct articles. This article describes the design, implementation, and experimental results of the device.

Keywords – Speech, Python, TTS, Email.

I. INTRODUCTION

This Text-to-speech synthesis - TTS - is the automatic conversion of a text into speech that is as similar as possible to a native speaker reading that text. [1] Text-to-Speech Synthesizer (TTS) is the technology that lets the computer talk to you. The TTS system receives the text as input and then a computer algorithm called the TTS engine analyzes the text, pre-processes it, and synthesizes the speech using some mathematical models. The TTS engine usually produces sound data in an audio format as output.

The text-to-speech (TTS) synthesis process consists of two main phases. The first is text analysis, where the input text is transcribed into a phonetic or other linguistic representation, and the second is the generation of speech waveforms, where the output is generated from this phonetic and prosodic information. These two phases are usually referred to as high-level and low-level synthesis [2]. A simplified version of this process shown in Figure 1 below. The input text can be, for example, data from a word processor, the standard ASCII from an email, a mobile text message, or scanned text from a newspaper. The string is then preprocessed and converted into a phonetic representation, which is usually a string of phonemes with some additional information about correct intonation, duration and stress. [3] Finally, the speech sound is generated with the low-level synthesizer using the information from the high-level synthesizer. The artificial generation of speech-like sounds has a long history, with documented mechanical experiments dating back to the eighteenth century.

II. EASE OF USE

Objective

Outline a logical and orderly approach to reading a scientific manuscript. [4] By breaking the task down into smaller, step-by-step components, you should be able to gain the skills to read a scientific article with ease.

Procedure

The reader should first read the title, summary, and conclusions. If one decides to read the entire article, one can go through the main elements of the article systematically, effectively, and efficiently. A coherent and organized method for reading articles published in scientific journals is presented.

III. PYTHON PROGRAMMING

Python is a popular programming language. [5] It was developed by Guido van Rossum and published in 1991. Python is a powerful general-purpose programming language. It is used in web development, data science, software prototyping and so on. Fortunately for beginners, Python has a simple, easy-to-use syntax. This makes Python an excellent language for beginners to learn to program with.

Python is a cross-platform programming language, meaning it can run on multiple platforms such as Windows, macOS, and Linux, and has even been ported to the Java and. NET virtual machines. It is free and open source. Even though most of today's Linux and Mac systems come with Python pre-installed, the version might be outdated. So, it is always a good idea to install the latest version. [6] A Python program can be run or implemented on platforms called IDE.

The most popular IDE's are PyCharm, Visual Studio Code, Anaconda, Python Idle, Jupyter notebook etc. A library is a previously combined set of codes that can be used iteratively, saving time. As the term implies, it is comparable to a physical library that contains reusable resources. Python has established several open-source libraries based on the fact that each library has a root source.

IV. LIBRARIES

A Python library is a collection of interrelated modules. [7] It contains bundles of code that can be used repeatedly in different programs. It makes Python programming easier and more convenient for the programmer. This is because we do not have to write the same code over and over again for different programs.

Python libraries play a very important role in the fields of machine learning, data science, data visualization, etc. As mentioned earlier, a Python library is simply a collection of codes or code modules that we can use in a program for specific operations. [8] We use libraries so that we do not have to rewrite the code that already exists in our program. But how does it work? In the Windows environment MS, the library files have a DLL extension (Dynamic Load Libraries).

When we link a library to our program and run this program, the linker automatically looks for this library. It extracts the functionalities of this library and interprets the program accordingly. In this way, we use the methods of a library in our program.

PYTTSX3

This is a library that helps convert text to speech. This library is the main library I used in my project.

REQUESTS

The requests library is the de facto standard for making HTTP requests in Python. It abstracts the complexity of requests behind a nice, simple API so you can focus on interacting with services and using data in your application.

BEAUTIFUL SOUP

Beautiful Soup is a library that allows you to easily extract information from web pages. It relies on an HTML or XML parser and provides Pythonic idioms for iterating, searching and modifying the parse tree.

V. TEXT TO SPEECH CONVERTER

Text-to-speech synthesis is performed in several steps. [9] The TTS systems receives a text as input, which it must first analyses and then convert into a phonetic description. In a further step, it then generates the prosody. From the information now available, it can generate a speech signal. The structure of the text-to-speech synthesizer can be divided into main modules:

Natural Language Processing (NLP) module: it generates a phonetic transcription of the read text, together with the prosody.

Digital Signal Processing module (DSP): It converts the symbolic information it receives from NLP into audible and intelligible speech.

The main operations of the NLP module are as follows

Text analysis: first, it segments the text into tokens. [10] By converting tokens into words, the orthographic form of the token is created. For the token "Mr." the orthographic form "Mister" is formed by expansion, the token "12" is given the orthographic form "twelve" and "1997" is converted to "nineteen ninety-seven".

Application of pronunciation rules: After the text analysis is complete, the pronunciation rules can be applied. Letters cannot be converted 1:1 into phonemes because the correspondence is not always parallel. In certain environments, a single letter may correspond to either no phoneme (e.g., "h" in "caught") or multiple phonemes ("m" in "maximum"). In addition, multiple letters may correspond to a single phoneme ("ch" in "rich").

There are two strategies for determining pronunciation:

In the dictionary-based solution with morphological components, as many morphemes (words) as possible are stored in a dictionary. The full forms are generated using inflexion, derivation, and composition rules. Alternatively, a full-form dictionary is used in which all possible word forms are stored. Pronunciation rules determine the pronunciation of words that are not in the dictionary.

In a rule-based solution, the pronunciation rules are generated from the phonological knowledge of the dictionaries. Only words whose pronunciation is a complete exception are included in the dictionary. The two applications differ significantly in the size of their dictionaries. The dictionary-based solution is several times larger than the exception dictionary of the rule-based solution. However, dictionary-based solutions can be more accurate than rule-based solutions if they have a sufficiently large phonetic dictionary.

Prosody generation: after the pronunciation is determined, the prosody is generated. The degree of naturalness of a TTS system depends on prosodic factors such as intonation modelling (phrasing and accentuation), amplitude modelling, and duration modelling (including the duration of the tone and the duration of the pauses, which determine the length of the syllable and the tempo of the speech).

The output of the NLP module is passed to the DSP module. This is where the actual synthesis of the speech signal takes place. In concatenative synthesis, the selection and linking of speech segments takes place. For individual sounds, the best option (if several suitable options are available) is selected from a database and concatenated.

Methodology

It is easy to use this application. To use this application, the user must follow certain steps as show Fig 1 and Fig 2:

* Since this is still a program version, the user must access the IDE page where the code is written and run the code.
* Select any URL from the web and enter it as input to the code.
* Wait for the program to read the input and get what you want.
* Relax, sit back and listen.

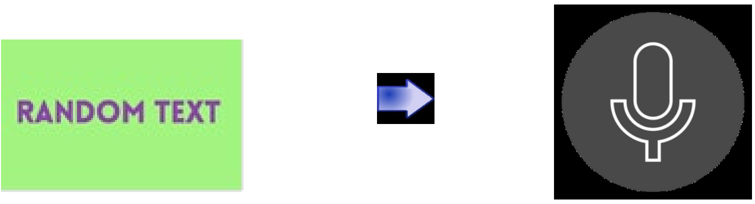


Fig 1. Methodology

VI. DATAFLOW AND IMPLEMENTATION

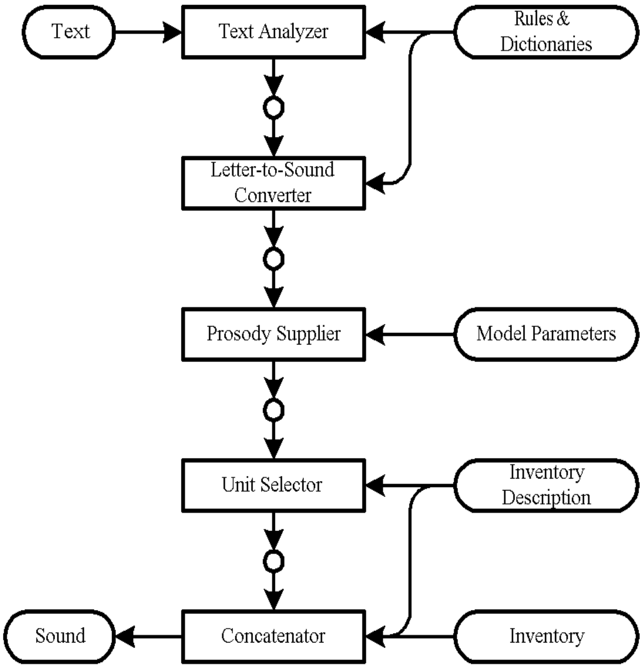


Fig 2. Implementation of methodology

Applications

The main motive of this software is to convert text into speech. Some of the main applications are:

* For visually impaired people
* There are a variety of applications in the medical field.

Advantages and Disadvantages

The advantages are:

* Saves students time in reading.
* Improves listening comprehension.

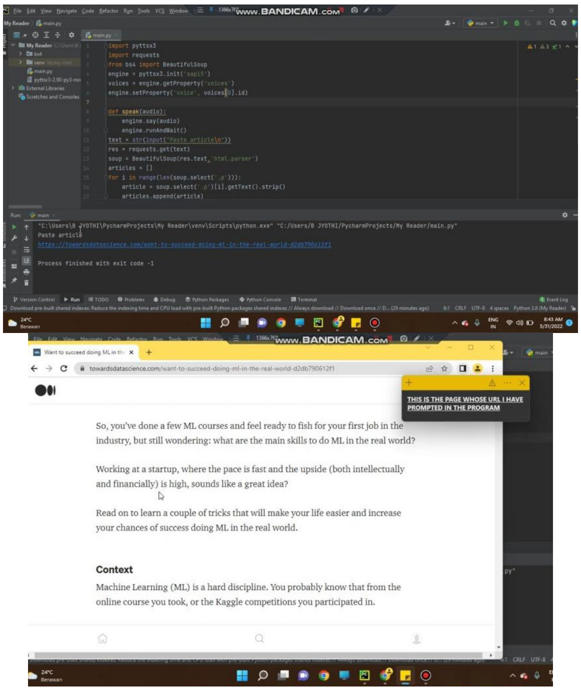
* Avoids distractions while learning.
* It is better to listen to lectures than to read them.
* It is especially for blind or visually impaired people.

The disadvantages are:

* So, there are no disadvantages of my article reader, yet, it lowers the reading level of the user.

VII. CONCLUSION

You can see the images of the text- to-speech conversion below:



```python
import pyttsx3
import requests
from bs4 import BeautifulSoup
engine = pyttsx3.init('sapi5')
voices = engine.getProperty('voices')
engine.setProperty('voice', voices[0].id)

def speak(audio):
    engine.say(audio)
    engine.runAndWait()
text = str(input("Paste article\n"))
res = requests.get(text)
soup = BeautifulSoup(res.text,'html.parser')
articles = []
for i in range(len(soup.select('.p'))):
    article = soup.select('.p')[i].getText().strip()
    articles.append(article)
```

```
"C:\Users\B JYOTHI\PycharmProjects\My Reader\venv\Scripts\python.exe" "C:/Users/B JYOTHI/PycharmProjects/My Reader/main.py"
Paste article
https://towardsdatascience.com/want-to-succeed-doing-ml-in-the-real-world-d2db790612f1

Process finished with exit code -1
```

Download pre-built shared indexes: Reduce the indexing time and CPU load with pre-built Python packages shared indexes // Always download // Download once // D... (29 minutes ago)

towardsdatascience.com/want-to-succeed-doing-ml-in-the-real-world-d2db790612f1

THIS IS THE PAGE WHOSE URL I HAVE PROMPTED IN THE PROGRAM

So, you've done a few ML courses and feel ready to fish for your first job in the industry, but still wondering: what are the main skills to do ML in the real world?

Working at a startup, where the pace is fast and the upside (both intellectually and financially) is high, sounds like a great idea?

Read on to learn a couple of tricks that will make your life easier and increase your chances of success doing ML in the real world.

Context

Machine Learning (ML) is a hard discipline. You probably know that from the online course you took, or the Kaggle competitions you participated in.

Fig 3. Converted text to speech.

The above Fig 3, text is converted to speech with the text-to-speech converter

Future Enhancement - this system can be developed to convert text to speech and vice versa, and furthermore, with the help of Artificial Intelligence, can be converted into a hologram explanation

Reference

[1] Python Crash Course: A Hands-On, Project-Based Introduction to Programming (2nd Edition)

[2] Learn Python the Hard Way: 3rd Edition

[3] Paul Taylor, "Text to speech synthesis" .

[4] Caleb Hattingh, "20 Python Libraries You Aren't Using"

[5] M. Zadka, "Text Manipulation," DevOps in Python, pp. 97–110, 2022, doi: 10.1007/978-1-4842-7996-0_6.

[6] D. Sarkar, "Text Analytics with Python," 2016, doi: 10.1007/978-1-4842-2388-8.

[7] Mamta Mittal, Gopi Battineni, Bhimavarapu Usharani, and Lalit Mohan Goyal, "Text Classification in Python," Text Analysis with Python: A Research Oriented Guide, pp. 78–120, Aug. 2022, doi: 10.2174/9789815049602122010006.

[8] N. K. Manaswi, "Speech to Text and Vice Versa," Deep Learning with Applications Using Python, pp. 127–144, 2018, doi: 10.1007/978-1-4842-3516-4_10.

[9] S. Mukhopadhyay and P. Samanta, "Analytics at Scale," Advanced Data Analytics Using Python, pp. 185–241, Nov. 2022, doi: 10.1007/978-1-4842-8005-8_7.

[10] M. Mittal, Gopi Battineni, Bhimavarapu Usharani, and L. M. Goyal, "Introduction to Python," Text Analysis with Python: A Research Oriented Guide, pp. 18–34, Aug. 2022, doi: 10.2174/9789815049602122010003.