

Real-time Speech to Text to Speech : Using OpenAI TTS

Khoi Duong
Prof. Chang
CS589
3/16/2024

Table of contents

- Step 1: Implement and run Real-time Speech to Text to Speech : Building Your AI-Based Alexa - using Google's gTTS for Text-to-Speech
- Step 2: Modify Putting Everything Together by replacing Google's gTTS with OpenAI's Text-to-Speech
- Step 3: Re-run the code

6. Real-time Speech to Text to Speech : Building Your AI-Based Alexa - using OpenAI's TTS for Text-to-Speech

- Step 1: Implement and run Real-time Speech to Text to Speech : Building Your AI-Based Alexa - using Google's gTTS for Text-to-Speech
- Step 2: Modify Putting Everything Together by replacing Google's gTTS with OpenAI's Text-to-Speech
- Step 3: Re-run the code?

Step 1: Implement and run Real-time Speech to Text to Speech

For step 1, please refer to the document about [Real-time Speech to Text to Speech : Building Your AI-Based Alexa](#) and follow the steps to complete

```
(venv) PS D:\VS CODE\Python\CS589\TextSpeech> python app.py --model base --english --energy 300 --pause 0.8 --dynamic_energy --wake_word "hey computer" --verbose
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\pydub\utils.py:170: RuntimeWarning: Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work
  warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Listening...
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\whisper\transcribe.py:115: UserWarning: FP16 is not supported on CPU; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
You said the wake word.. Processing ...
You said: who is the founder of OpenAI
The answer content:, a research institute that is dedicated to developing friendly AI. He is also a co-founder of Tesla and Neuralink, companies that are focused on developing advanced technologies in the fields of electric cars and brain-computer interfaces, respectively.

Musk was born in South Africa in 1971 and showed an early interest in computers and technology. He moved to Canada at the age of 17 to attend Queen's University and then transferred to the University of Pennsylvania, where he received dual bachelor's degrees in economics and physics. He then went on to pursue a PhD in energy physics at Stanford University, but dropped out after two days to pursue entrepreneurial ventures.

In 1995, Musk co-founded Zip2, a company that provided online content publishing software for newspapers
Transform the answer to mp3 ... Result will be in reply.mp3!
```

Step 2: Replacing Google's gTTS with OpenAI's Text-to-Speech

To implement OpenAI's TTS, we need to change the code in the `reply()` function to apply the new OpenAI TTS based on the guide from [Text to speech - OpenAI API](#)

```
1  from openai import OpenAI
2
3  client = OpenAI()
4
5  response = client.audio.speech.create(
6      model="tts-1",
7      voice="alloy",
8      input="Hello world! This is a streaming test.",
9  )
10
11 response.stream_to_file("output.mp3")
```

```

from pathlib import Path
from openai import OpenAI
client = OpenAI()

def reply(result_queue, verbose):
    while True:
        question = result_queue.get()
        # We use the following format for the prompt: "Q: ?\nA:"
        prompt = "Q: {}?\nA:".format(question)

        data = openai.Completion.create(
            model="gpt-3.5-turbo-instruct",
            prompt=prompt,
            temperature=0.5,
            max_tokens=100,
            n=1,
            stop=["\n"]
        )

        # We catch the exception in case there is no answer
        try:
            answer = data["choices"][0]["text"]
            print("The answer content:" + answer)
            print("Transform the answer to mp3... Result will be in reply.mp3!")
            # Replace "answer" with the actual text you want to convert to speech
            text_to_speak = answer

            # Set the output filename (modify as needed)
            speech_file_path = Path(__file__).parent / "speech.mp3"

            # Use OpenAI TTS API with desired voice and language
            response = client.audio.speech.create(
                model="tts-1", # Adjust model if needed (check OpenAI documentation)
                voice="nova", # Choose a voice from available options
                input=text_to_speak,
                lang="en", # Specify language explicitly (optional)
            )
            response.stream_to_file(speech_file_path)
        except Exception as e:

```

```

except Exception as e:
    choices = [
        "I'm sorry, I don't know the answer to that",
        "I'm not sure I understand",
        "I'm not sure I can answer that",
        "Please repeat the question in a different way"
    ]
    answer = choices[np.random.randint(0, len(choices))]
    print("The answer content:" + answer)
    print("Transform the answer to mp3... Result will be in reply.mp3!")
    # Replace "answer" with the actual text you want to convert to speech
    text_to_speak = answer

    # Set the output filename (modify as needed)
    speech_file_path = Path(__file__).parent / "speech.mp3"

    # Use OpenAI TTS API with desired voice and language
    response = client.audio.speech.create(
        model="tts-1", # Adjust model if needed (check OpenAI documentation)
        voice="nova", # Choose a voice from available options
        input=text_to_speak,
        lang="en", # Specify language explicitly (optional)
    )

    response.stream_to_file(speech_file_path)
    if verbose:
        print(e)

# In both cases, we play the audio
audio = AudioSegment.from_file(speech_file_path)
play(audio)

main()

```

Step 3: Rerun the code

When executing the program, we have the error below:

```
(venv) PS D:\VS CODE\Python\CS589\TextSpeech> python app_openaiTTS.py --model base --english --energy 300 --pause 0.8 --dynamic_energy --wake_word "hey computer" --verbose
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\pydub\utils.py:170: RuntimeWarning: Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work
  warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Traceback (most recent call last):
  File "app_openaiTTS.py", line 89, in <module>
    from openai import OpenAI
  ImportError: cannot import name 'OpenAI' from 'openai' (D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\openai\__init__.py)
(venv) PS D:\VS CODE\Python\CS589\TextSpeech> |
```

The error happens because we use the version 0.28 for the previous assignments (CustomerEmail, Agents, DocumentLoad, etc.)

Refer to the document below to fix the problem.

[Cannot import name 'OpenAI' from 'openai' - API](#)


```
(venv) PS D:\VS CODE\Python\CS589\TextSpeech> pip install openai --upgrade
Requirement already satisfied: openai in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (0.28.0)
Collecting openai
  Downloading openai-1.14.1-py3-none-any.whl (257 kB)
    | 257 kB 1.6 MB/s
Requirement already satisfied: sniffio in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (1.3.1)
Requirement already satisfied: pydantic<3,>=1.9.0 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (2.6.4)
Requirement already satisfied: tqdm>4 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (4.66.2)
Requirement already satisfied: typing-extensions<5,>=4.7 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (4.10.0)
Requirement already satisfied: distro<2,>=1.7.0 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (1.9.0)
Requirement already satisfied: httpx<1,>=0.23.0 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (0.27.0)
Requirement already satisfied: anyio<5,>=3.5.0 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from openai) (4.3.0)
Requirement already satisfied: exceptiongroup>=1.0.2 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from anyio<5,>=3.5.0->openai) (1.2.0)
Requirement already satisfied: idna>=2.8 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from anyio<5,>=3.5.0->openai) (3.6)
Requirement already satisfied: certifi in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from httpx<1,>=0.23.0->openai) (2024.2.2)
Requirement already satisfied: httpcore==1.* in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from httpx<1,>=0.23.0->openai) (1.0.4)
Requirement already satisfied: h11<0.15,>=0.13 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from httpcore==1.*->httpx<1,>=0.23.0->openai) (0.14.0)
Requirement already satisfied: pydantic-core==2.16.3 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from pydantic<3,>=1.9.0->openai) (2.16.3)
Requirement already satisfied: annotated-types>=0.4.0 in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from pydantic<3,>=1.9.0->openai) (0.6.0)
Requirement already satisfied: colorama in d:\vs code\python\cs589\textspeech\venv\lib\site-packages (from tqdm>4->openai) (0.4.6)
Installing collected packages: openai
  Attempting uninstall: openai
    Found existing installation: openai 0.28.0
    Uninstalling openai-0.28.0:
      Successfully uninstalled openai-0.28.0
  Successfully installed openai-1.14.1
WARNING: You are using pip version 21.1.1; however, version 24.0 is available.
You should consider upgrading via the 'd:\vs code\python\cs589\textspeech\venv\scripts\python.exe -m pip install --upgrade pip' command.
(venv) PS D:\VS CODE\Python\CS589\TextSpeech>
```

Execute the program

Retry again with the new OpenAI version, we experience an old error related to ***openai.Completion.create*** that is no longer supported in the new version (1.14) of OpenAI

(The image of error is in the next slide)

For now, please refer to the blog about [openai.Completion.create degradation](#) to solve the issue.



j Regular

```
'''example completion with openai > 1.1'''  
from openai import OpenAI  
client = OpenAI()  
prompt = "True or false: a banana is smaller than a lemon.\n\n"  
  
response = client.completions.create(  
    prompt=prompt,  
    model="gpt-3.5-turbo-instruct",  
    top_p=0.5, max_tokens=50,  
    stream=True)  
  
for part in response:  
    print(part.choices[0].text or "")
```

We just need to change ***openai.Completion.create*** into ***client.completions.create*** and try to run the code again.


```

(venv) PS D:\VS CODE\Python\CS589\TextSpeech> python app_openaiTTS.py --model base --english --energy 300 --pause 0.8 --dynamic_energy --wake_word "hey computer" --verbose
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\pydub\utils.py:170: RuntimeWarning: Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work
  warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Listening...
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\whisper\transcribe.py:115: UserWarning: FP16 is not supported on CPU; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
You did not say the wake word.. Ignoring
You said the wake word.. Processing ...
You said: where is the San Francisco Bay University
Exception in thread Thread-3:
Traceback (most recent call last):
  File "C:\Program Files\WindowsApps\PythonSoftwareFoundation.Python.3.8_3.8.2800.0_x64__qbz5n2kfra8p0\lib\threading.py", line 932, in _bootstrap_inner
    self.run()
  File "C:\Program Files\WindowsApps\PythonSoftwareFoundation.Python.3.8_3.8.2800.0_x64__qbz5n2kfra8p0\lib\threading.py", line 870, in run
    self._target(*self._args, **self._kwargs)
  File "app_openaiTTS.py", line 97, in reply
    data = openai.Completion.create(
  File "D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\openai\lib\_old_api.py", line 39, in __call__
    raise APIRemovedInV1(symbol=self._symbol)
openai.lib._old_api.APIRemovedInV1:

```

You tried to access `openai.Completion`, but this is no longer supported in `openai>=1.0.0` - see the README at <https://github.com/openai/openai-python> for the API.

You can run ``openai migrate`` to automatically upgrade your codebase to use the 1.0.0 interface.

Alternatively, you can pin your installation to the old version, e.g. ``pip install openai==0.28``

A detailed migration guide is available here: <https://github.com/openai/openai-python/discussions/742>

You did not say the wake word.. Ignoring
You did not say the wake word.. Ignoring

'Completion' object is not subscriptable

```
○ (venv) PS D:\VS CODE\Python\CS589\TextSpeech> python app_openaiTTS.py --model base --english --energy 300
--pause 0.8 --dynamic_energy --wake_word "hey computer" --verbose
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\pydub\utils.py:170: RuntimeWarning: Couldn't find
ffmpeg or avconv - defaulting to ffmpeg, but may not work
  warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Listening...
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\whisper\transcribe.py:115: UserWarning: FP16 is
not supported on CPU; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
You did not say the wake word.. Ignoring
You did not say the wake word.. Ignoring
You said the wake word.. Processing ...
You said: who is the founder of OpenAI
The answer content:I'm sorry, I don't know the answer to that
Transform the answer to mp3... Result will be in speech.mp3!
app_openaiTTS.py:148: DeprecationWarning: Due to a bug, this method doesn't actually stream the response
content, `.with_streaming_response.method()` should be used instead
  response.stream_to_file(speech_file_path)
Exception msg: 'Completion' object is not subscriptable
Verbose here
```

Solution

Please refer to the document [Open AI TypeError: 'Completion' object is not subscriptable - Stack Overflow](#) to read about the error.

```
try:
    answer = data["choices"][0]["text"]
    print("The answer content:" + answer)
    print("Transform the answer to mp3... Result will be in reply.mp3!")
    # Replace "answer" with the actual text you want to convert to speech
    text_to_speak = answer
```

We need to change the line into:

answer = data.choices[0].text

Final result:

Retry again and now we have the result we wanted:

```
○ (venv) PS D:\VS CODE\Python\CS589\TextSpeech> python app_openaiTTS.py --model base --english --energy 300
--pause 0.8 --dynamic_energy --wake_word "hey computer" --verbose
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\pydub\utils.py:170: RuntimeWarning: Couldn't find
ffmpeg or avconv - defaulting to ffmpeg, but may not work
  warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Listening...
D:\VS CODE\Python\CS589\TextSpeech\venv\lib\site-packages\whisper\transcribe.py:115: UserWarning: FP16 is
not supported on CPU; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
You did not say the wake word.. Ignoring
You said the wake word.. Processing ...
● You said: who is the founder of OpenAI
The answer content: The founders of OpenAI are Elon Musk, Sam Altman, Greg Brockman, Ilya Sutskever, Wojciech Zaremba, John Schulman, and Chris Olah.
Transform the answer to mp3... Result will be in speech.mp3!
app_openaiTTS.py:123: DeprecationWarning: Due to a bug, this method doesn't actually stream the response
content, `.with_streaming_response.method()` should be used instead
  response.stream_to_file(speech_file_path)
```

References

- [Real-time Speech to Text to Speech : Building Your AI-Based Alexa](#)
- [Text to speech - OpenAI API](#)

Source code: <https://github.com/MynameisKoi/CS589/tree/main/TextSpeech>