ITIS 6617 SYSTEM INTEGRATION

FINAL PROJECT

Shanmukh Gorle
ID: 801311172

USER GUIDE

FOR

TEXT-TO-SPEECH API

(MICROSOFT AZURE)

Version: 1.0
Final Draft

University of North Carolina, Charlotte
2023

Contents

1. Introduction

1.1. Overview of Azure’s Text-to-Speech API

The Text-to-Speech API is a powerful tool that transforms written text into audible speech, leveraging the advanced capabilities of Azure AI services. This technology is incredibly versatile, catering to various applications such as reading digital content aloud, assisting visually impaired users, aiding in language learning, and more. Its primary aim is to make digital content more accessible and interactive for a diverse audience.

This user guide is designed to help you use Azure's Text-to-Speech API. Whether you're a developer or just interested in turning text into speech, we've got you covered.

Technology:
The service uses Microsoft Cognitive Services Speech SDK. This sophisticated technology ensures high-quality voice synthesis, providing a seamless and natural listening experience.

1.2. Key Features

Diverse Voice Options:
One of the standout features of our API is the extensive range of voice options available. Users can choose from various accents and languages, creating a personalized and engaging experience. This diversity in voice options opens possibilities for a more inclusive and global use of our service.

Ease of Use:
We have designed our API with user-friendliness in mind. It offers a simple and intuitive web interface for casual users while providing straightforward API endpoints for developers. This dual approach ensures that our service is accessible to a broad spectrum of users with varying levels of technical expertise.

Real-Time Conversion:
Our API excels in providing real-time text-to-speech conversion. This efficiency is crucial for applications requiring immediate vocalization of text, making our service practical and time-efficient for various use cases.

1.3. Accessibility

Web Interface:
The service is easily accessible through a user-friendly web interface. This interface is designed to be intuitive and straightforward, making it perfect for non-technical users who need a quick and hassle-free way to convert text to speech.

API Access:
For developers and more technical users, our service is also accessible through API endpoints. This feature is ideal for integrating text-to-speech capabilities into their applications or services.

1.4. Limitations

Rate Limiting:
Time Window: The application enforces a rate limit over a window of 5 seconds.
Request Limit: Up to 5 requests within this 5-second window.
Response to Exceeding Limit: If the mentioned limit is exceeded, our system will respond with a message indicating that multiple requests have been received and a 5-second timeout will be applied.
Error Message: In such cases, a 400 HTTP status code with the message: "Multiple requests received. 5 seconds timeout" will be returned.

Character Limit:
The API currently allows a maximum of 1000 characters per request. This limit ensures efficient processing and response time, providing a seamless experience for all users.

1.5. Deployment and Availability

Hosted on Digital Ocean:
Our API is reliably hosted on a Digital Ocean droplet. This hosting ensures the service is scalable and dependable, providing users with uninterrupted access.

No Authentication Required:
We have streamlined access to our API by removing the need for authentication keys. This decision was made to make the service accessible, allowing anyone to use it without complex setup processes.

2. Getting Started with the Text-to-Speech API

2.1. Access Points

Text-to-speech service is designed to be flexible and accessible, offering two primary methods of access to cater to different user needs and preferences. Below are the two methods to access the API.

Web Interface:
Ideal for users who prefer a graphical user interface. Visit the provided URL using any standard web browser.

API Endpoints:
Suitable for developers or technical users who want to integrate our service into their applications. These endpoints can be accessed using tools like Postman or through custom code.

2.2. Requirements

Internet Connectivity: A stable internet connection is required to access the web interface and the API.

Web Browser: Any modern web browser like Chrome, Firefox, or Safari will work for the web interface.

API Tools: Tools like Postman for testing API endpoints or a coding environment if you plan to integrate the API into your application.

Postman is an API platform for building and using APIs. Postman simplifies each step of the API lifecycle and streamlines collaboration so you can create better APIs - faster. Refer to https://www.postman.com/ to install Postman on your computer.

2.3. Using the Web Interface

Access the web interface by visiting http://198.199.83.188:3000/. This URL leads to the homepage of the Text to Speech service. The homepage should look something like this:

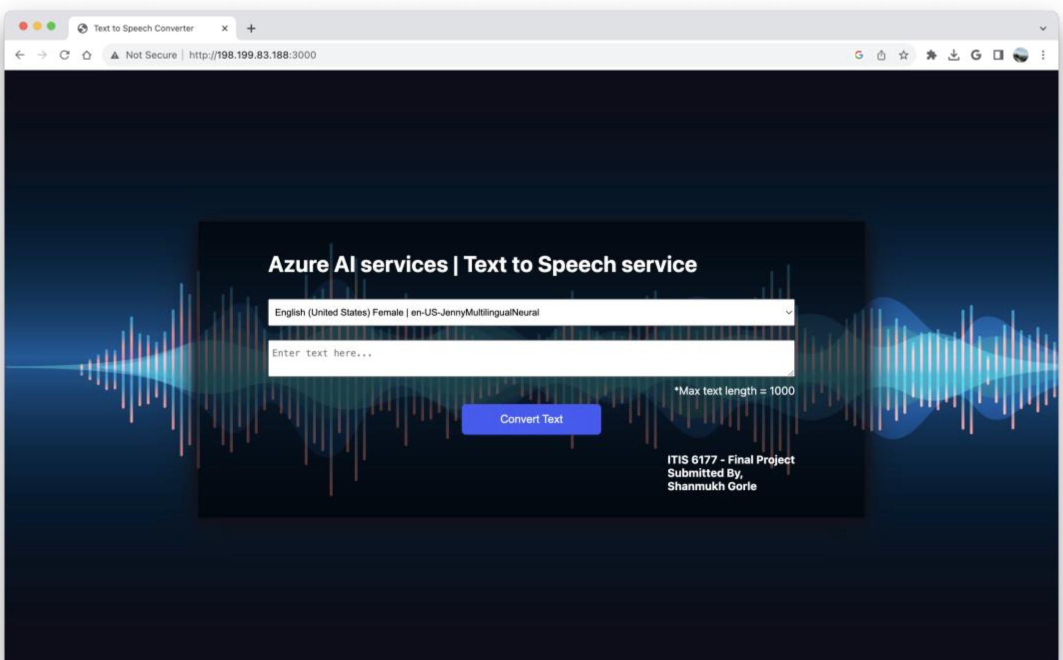
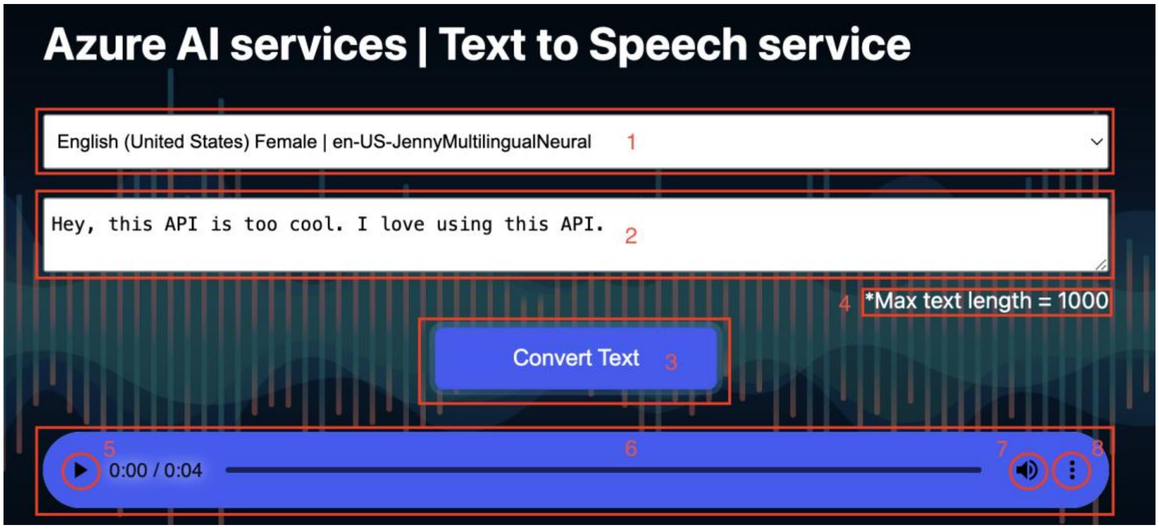


Figure 1: Home page of Text-to-speech API

Interface Overview



The web interface consists of:

- **Voice Selection Dropdown (1):** A dropdown menu to select the desired voice for speech synthesis. Select a suitable voice from the dropdown menu. Each voice may have a unique accent or language.
- **Text Input Box (2):** This is where you input the text you want to convert to speech. Type or paste your desired text into the text input box.
- **Convert Text Button (3):** A text-to-speech conversion button. Click the 'Convert Text' button to synthesize the speech.
- **Max Text Length (4):** Currently, it's the maximum character limit set (1000 characters).
- **Audio Player (6):** Appears after conversion, allowing you to play and listen to the synthesized speech.
    - **Play Audio (5):** Plays the audio.
    - **Volume Controls (7):** Controls the audio.
    - **Playback and Download Audio Options (8):** Control the playback speed and download the audio from here.

2.4. Using an API tool

Postman is a tool that can be utilized for testing and experimentation of the API. Follow the steps below to test API using Postman once it is installed.

Making a Request

Base URL: http://198.199.83.188:3000/

Swagger URL: http://198.199.83.188:3000/api-docs/

Required HTTP Headers for API Requests

When interacting with our API, setting the appropriate HTTP headers is important. These headers help ensure that your requests are handled correctly by the server.

Below is an example of the HTTP headers that can be set in Postman.

```
{
    "Cache-Control": "no-cache",
    "Content-Type": "application/json; charset=utf-8",
    "Accept": "*/*",
    "Accept-Encoding": "gzip, deflate, br",
    "Connection": "keep-alive",
     "User-Agent": "PostmanRuntime/7.35.0"
}
```

Endpoint Description

- **/voices:** Retrieves a list of available voices.
    - **Method Type:** `GET`
    - **Request Parameters:** No Parameters Required
    - **Response:** The response is in the below format.

      Example Response:

```
[{  "Name": "Microsoft Server Speech Text to Speech Voice (af-ZA, AdriNeural)",
    "DisplayName": "Adri",
    "LocalName": "Adri",
    "ShortName": "af-ZA-AdriNeural",
    "Gender": "Female",
    "Locale": "af-ZA",
    "LocaleName": "Afrikaans (South Africa)",
    "SampleRateHertz": "48000",
    "VoiceType": "Neural",
    "Status": "GA",
    "WordsPerMinute": "147"
}]
```

    - **Status Code:**

Code	Details
200	List of available voices.
500	Error in fetching the list of voices.

cURL code snippet: `curl --location '198.199.83.188:3000/voices'`

- **/generate: Converts provided text to speech.**
  - **Method Type:** `POST`
  - **Request Parameters:** Include parameters such as text (required field, the text to be converted) and voice (optional field, specifies the voice).
    Example: { text": "Hello world", "voice": "en-US-JennyNeural" }
    The voice field is optional. A voice list can be obtained from /voices endpoint. If the voice field is not specified, the default voice is en-US-JennyNeural. Please note that the maximum number of characters allowed in the text field now is 300.
  - **Response:** An audio file of the spoken text.
  - **Status Codes:**

Code	Details
200	Audio file generated successfully.
400	Bad request, invalid input data.
500	Server error or issue with Azure AI services.

cURL code snippet.

```
curl --location '198.199.83.188:3000/generate' \
--header 'Content-Type: application/json; charset=utf-8' \
--data '{
    "text": "This is a really cool API. I love using it.",
    "voice": "en-US-JennyNeural"
}'
```

Screenshots From Postman:

Figure 2 is for getting all the voices (/voices) from the API through Postman.

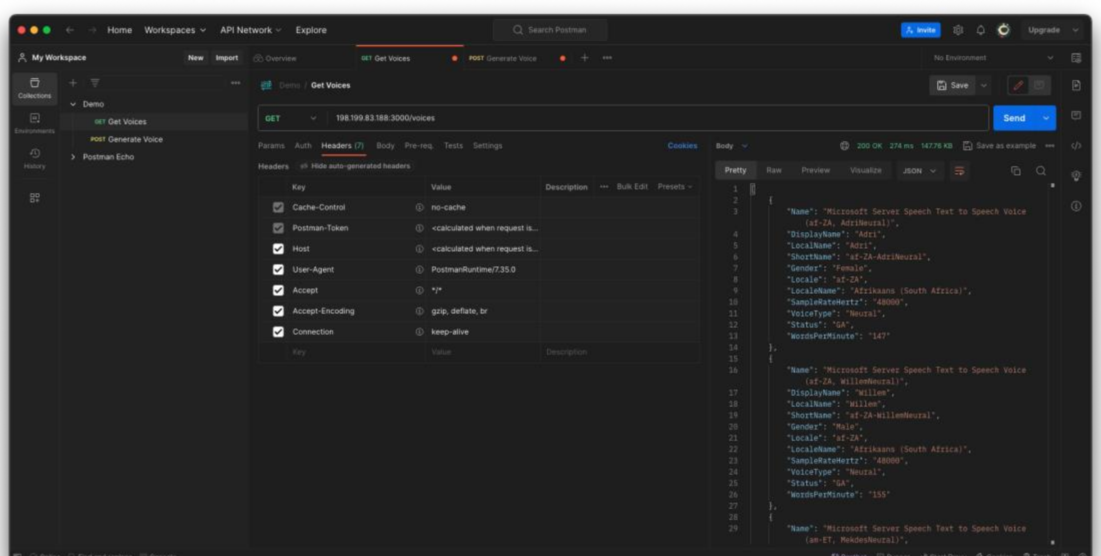


Figure 2: Postman snip for voices endpoint

Figure 3 converts text to audio format (/generate) from the API through Postman.

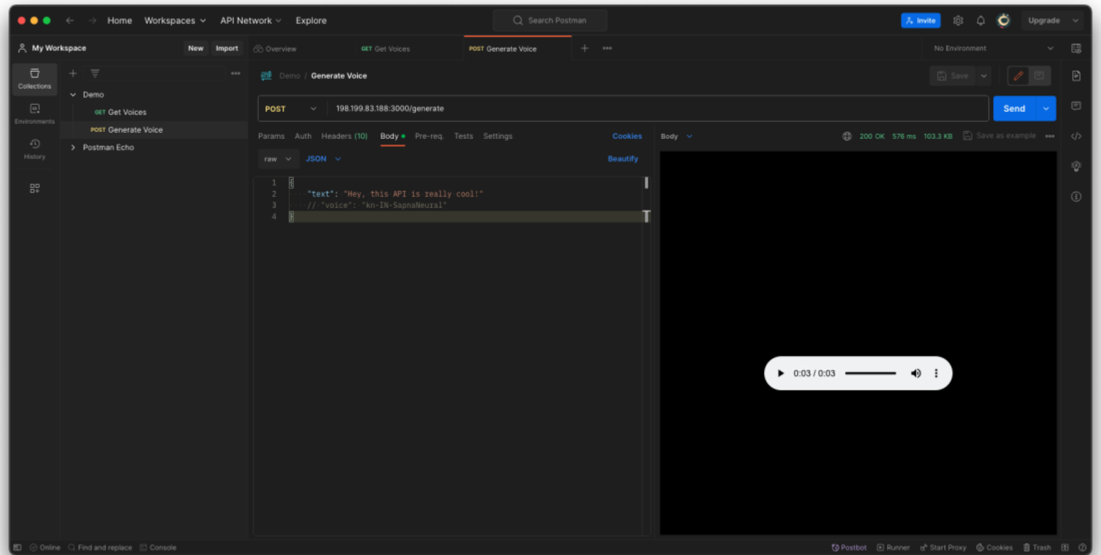


Figure 3: Postman snip for generating voice endpoint.

2.5. Alternative Testing with Swagger UI

In addition to using Postman, you can directly test the Text-to-Speech API through the Swagger UI at http://198.199.83.188:3000/api-docs. This provides a web-based interface to explore and interact with the API endpoints. It also includes all details on how to use the API effectively.

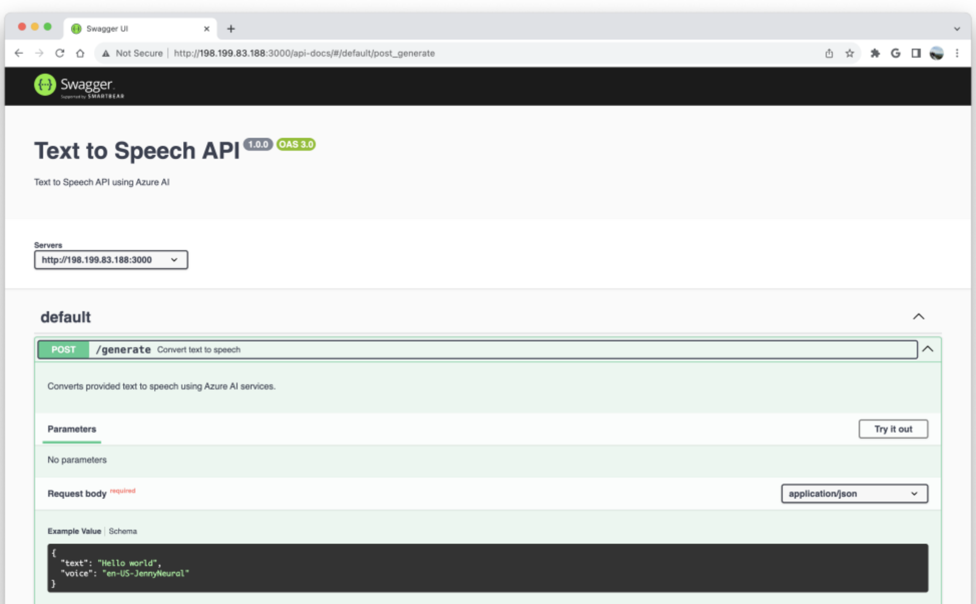


Figure 4: Swagger snip

3. Troubleshooting and Support

Here are some common issues that users may encounter and their potential solutions.

Issue: API returns a 400 Bad Request error
Solution: Ensure that all required fields are filled out correctly in the request body and that the JSON format is valid. Check headers as well.

Issue: API returns a 500 Internal Server Error
Solution: This is a server-side issue. Try the request again after some time. If the problem persists, contact support.

Issue: No sound is produced when playing the audio file
Solution: Verify that the volume is turned up and audio file is not corrupt.

Contact Support Email: sgorle@uncc.com

4. FAQs

Q: Can this API be used for both personal and commercial use?
A: Yes, the API is available for personal and commercial use.

Q: Is authentication required?
A: No, our API does not require authentication currently.

Q: Are there any limits on the number of requests that can be made?
A: A rate limit is currently in place to ensure fair usage. Please refer to the rate limiting section for details.

Q: What languages and voices that can be used?
A: A wide range of languages and voices are available. Use the /voices endpoint to retrieve the current list.

Q: Is there any character limit?
A: As of now there is a 1000-character limit to the text that can be converted to audio simultaneously.

5. Conclusion

This guide has provided all the information you need to use our Text-to-Speech API successfully. Whether you want to integrate this service into your application or convert some text to speech directly from the web interface, we encourage you to try it out and experience the benefits. For any further assistance or feedback, please don't hesitate to contact our support team. Happy synthesizing!