ITIS 6177 SYSTEM INTEGRATION

YASH CHATURVEDI: 801317460

FINAL PROJECT

MICROSOFT AZURE COGNITIVE SERVICES : TEXT TO SPEECH CONVERSION

Release version : 1.0.0

Document Status : Final Draft

University of North Carolina, Charlotte
2022

Introduction:

This project uses NodeJs as a server to provide APIs that can communicate with the Microsoft Azure service. The API takes input as text from the user. Passes those parameters to the the azure service and returns a final audio file to the end user.

What is Text to Speech Conversion

Azure Text to Speech is a powerful speech synthesis capability of Azure Cognitive Services, enables developers to convert text to life like speech using AI.

Speech-to-text, also known as speech recognition, enables real-time or offline transcription of audio streams into text. The Azure speech-to-text service analyzes audio in real-time or batch to transcribe the spoken word into text. Out of the box, speech to text utilizes a Universal Language Model as a base model that is trained with Microsoft-owned data and reflects commonly used spoken language. This base model is pre-trained with dialects and phonetics representing a variety of common domains. The base model works well in most scenarios.

API Details

Input :

The Api is a HTTP Post request that accepts one parameter :

Name : text
Required : true

Example :

text: "Hi How are you doing today "

Select body , then select x-www-form-urlencoded radio button and enter your desired text.

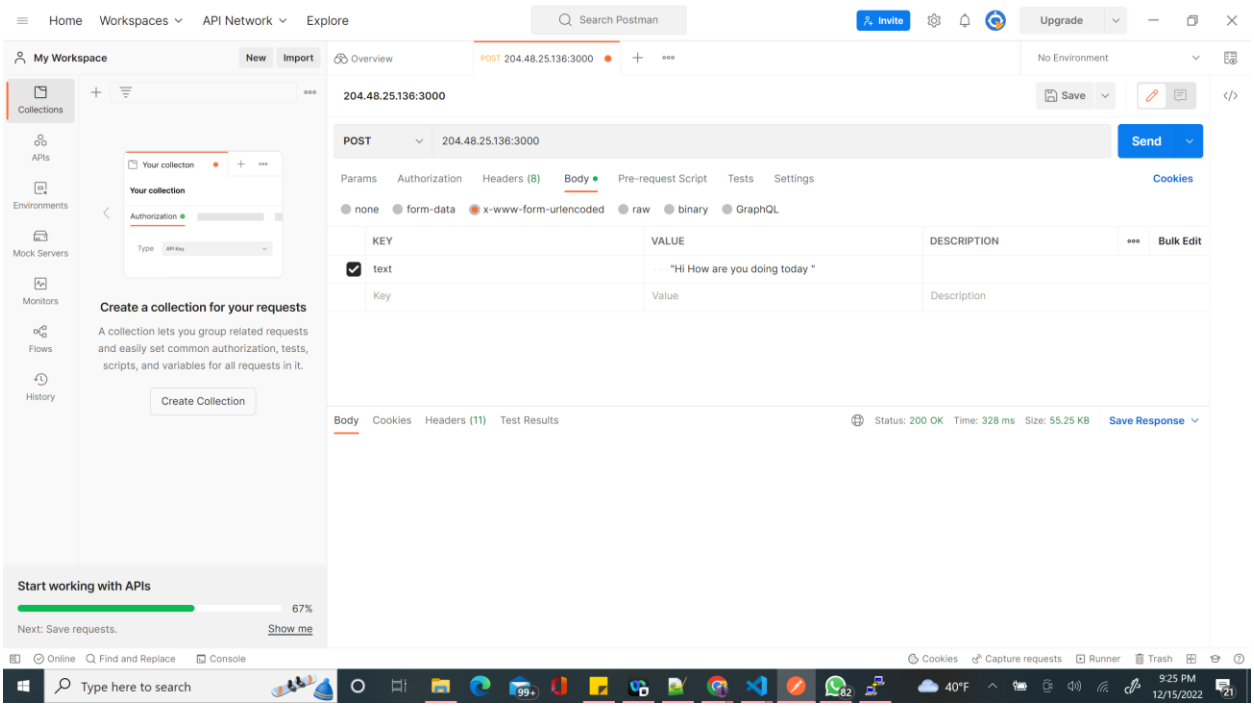
Note : click on send and download by clicking on the drop down button near send.

Please see below screenshot for your understanding.

Request Headers :

Headers | Hide auto-generated headers

KEY	VALUE
Postman-Token	<calculated when request is sent>
Content-Type	application/x-www-form-urlencoded
Content-Length	<calculated when request is sent>
Host	<calculated when request is sent>
User-Agent	PostmanRuntime/7.30.0
Accept	*/*
Accept-Encoding	gzip, deflate, br
Connection	keep-alive
text	hello kaisa hai



Response :

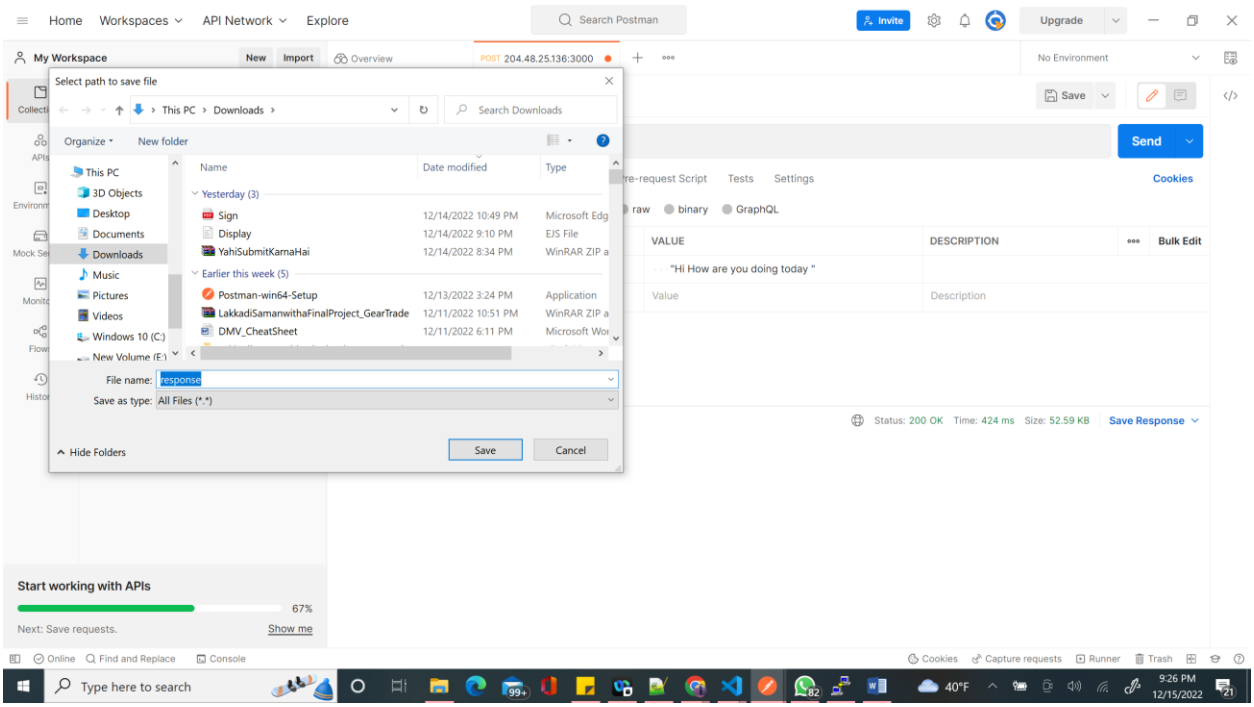
The Response is a audio file with the input text converted to speech. Please see below screenshot.

Response Headers :

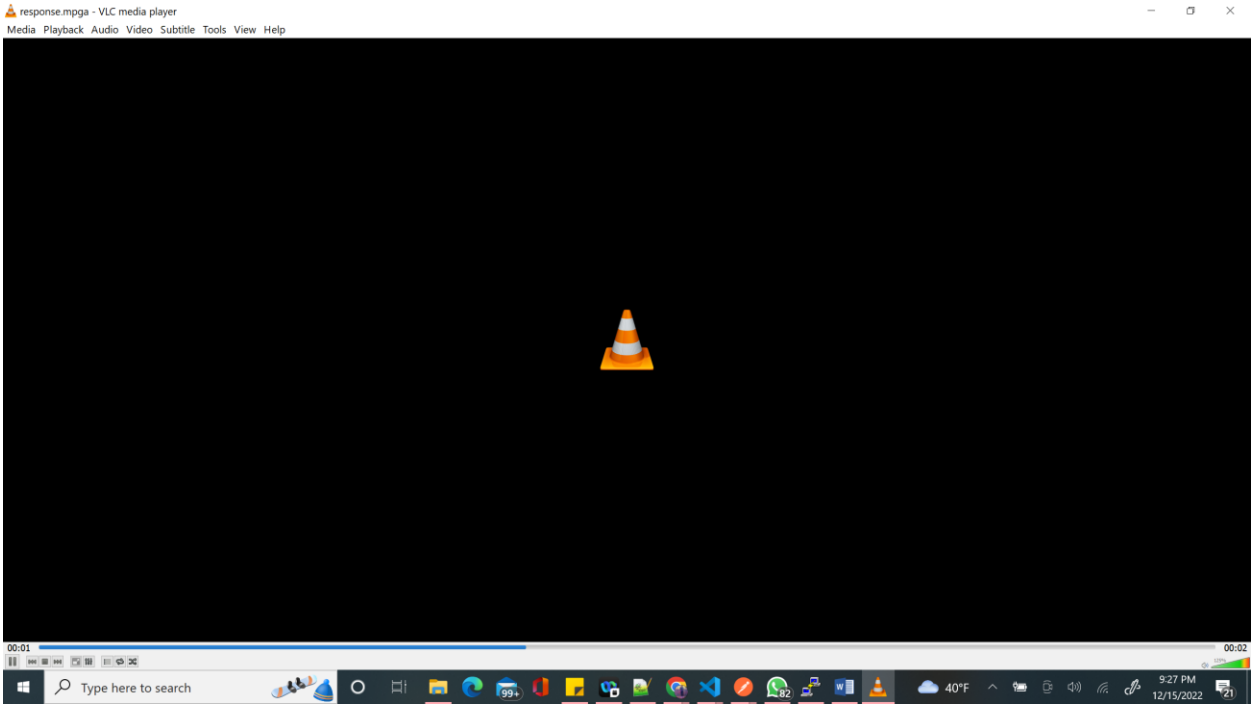
Body Cookies Headers (10) Test Results Status: 200 OK Tim

KEY	VALUE
X-Powered-By	Express
Access-Control-Allow-Origin	*
Accept-Ranges	bytes
Cache-Control	public, max-age=0
Last-Modified	Fri, 16 Dec 2022 04:30:58 GMT
ETag	W/"12dbc-18519331c84"
Content-Type	audio/mpeg
Content-Length	77244
Date	Fri, 16 Dec 2022 04:30:58 GMT
Connection	keep-alive

For better understanding.



You can play the audio file with appropriate player :



Swagger :

For better usability swagger is also used for enhancing user experience: