PIC16

Assignment 4W

Li, Chak Yin (Anders)

704950093

The ***speechTranslation*** function combines two Google Cloud Platform APIs, Speech and Translation. The function takes in three parameters, the file name of the audio file (*file\_name)*, the format of it *(audioFormat)*, and the language spoken in the audio (*lang)*. This function will turn audio files in any language into an English text.

The function will first open the audio file in binary form. It will read the file, encode it into base64, and store the content into a variable. The content will be saved as the ‘*content’* field of the *audio* of the request body by decoding it into UTF-8. As for the *config* of the request body, *audioFormat* and *lang* are corresponded to the ‘*encoding’* and ‘*languageCode’* fields respectively. Config and audio will be combined into a dictionary and converted to JSON format. Then, the function will post a request to <https://speech.googleapis.com/v1/speech:recognize> with all the parameters and the user’s API key. After posting the request and receiving a response from it, the JSON format of the response body will be converted into a dictionary. The function then extracts the *transcript*.

Since the user can pass in audio files in any language, the function will automatically detect the language spoken in the audio file by using the *detect* method of the Google Cloud Translation API. The *transcript* obtained by the Speech API will be passed into the request body of the Translation API corresponding to the field *‘q’* and the detected language will be extracted from the response.

The last step of the function call is to pass the *transcript*, detected language, target language into the request body of the *translate* method. Finally, the *translatedText* is extracted from the response and printed to the console.

This function is particularly useful when you are messaging a bilingual friend (using voice message). It is common that a bilingual friend sometimes accidentally speaks to you in a different language (maybe because he/she is speaking in that language with someone else). This is where the function comes into play. All you need to do is to download the audio file, specify the format and language, and put them into the function. Then you can get a text of what your friend is talking about in English. No more confusion!

In completing this homework, I had some troubles in understanding what base does the *‘content’* take in. I had other troubles in encoding the audio file into base64 and decoding it so that it fits the JSON format. After overcoming these troubles, the function can be written to pass in the original audio file, which simplifies the step that the users have to encode the file themselves.