

# EE250 Final Project: Speech-to-Text converter

## Team members:

- Jiansong Li
- Robert Zhang

## Link to our demo video:

[https://drive.google.com/file/d/1myDE-PqRci2Lu9yxwQPe\\_3Xbz\\_2OvKVm/view?usp=share\\_link](https://drive.google.com/file/d/1myDE-PqRci2Lu9yxwQPe_3Xbz_2OvKVm/view?usp=share_link)

## Block Diagram :

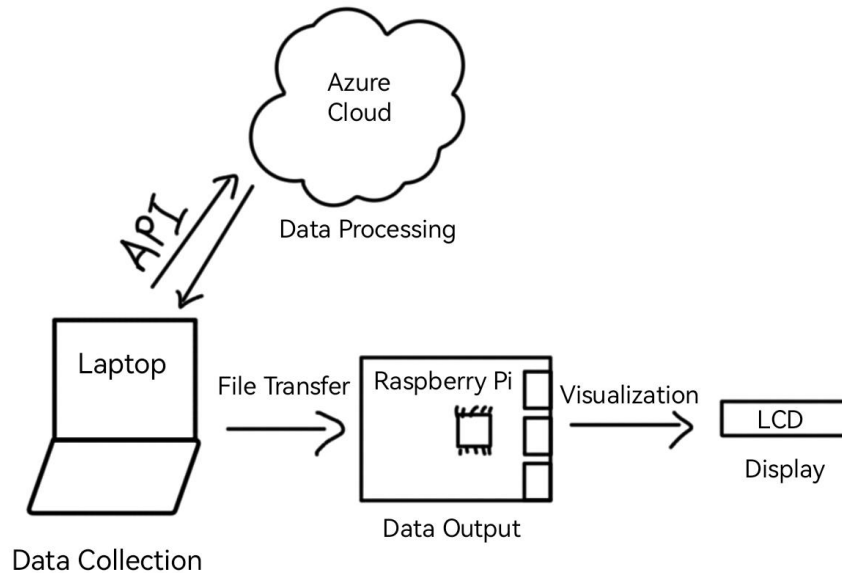


Figure 1: Block Diagram

## Description of our final project:

In this project, we aim to make a speech-to-text converter. The main functionality of this converter is to provide a visual presentation of our speech. It basically works as a subtitle generator. People are able to detect what you have said visually. I think this could be useful on occasion for public speaking, presentation, or communication with people who are deaf or hard

of hearing. Moreover, I believe you must have watched the video with subtitles, so our Final project is basically performing the process of converting audio to text.

As illustrated in the block diagram, we use the laptop to record our audio and save it as a wav file by accessing the python package of wavio and sounddevice. Then, in order to process the audio data we recorded, we utilize Azure Speech Recognition API as the platform to connect to the Azure cloud and convert the wav file to a text file. The API will take the responsibility of detecting our speech and converting it to text output. Next, we use the TCP Socket protocol to transfer our text file output from the laptop to Raspberry Pi. In this transferring process, our laptop serves to provide text data that contains our recorded speech. Then, the RPi works as a client to send a request to the server and receive the speech text data. Finally, to display the recorded speech text or the subtitles, we utilize a GrovePi LCD screen to print out the text data as a visualization of our speech.

### **Reflection and possible improvement:**

One of the limitations of our project is the conversion quality. It won't perfectly convert the audio to text. For example, the background noise and accent affect the speech conversion. We recommend having a better microphone or recording in a quiet room. Another limitation of our project is the LCD display. The LCD screen provided by GrovePi can only display English text. To resolve this problem, we have to change the display component displays the unique characters. In addition, due to the size of the LCD screen, it is difficult to print a long sentence once at a time.

The second limitation is the problem of synchronization. Our initial plan is to create a synchronous speech-to-text converter or a subtitle generator. However, due to the time limitation, we found it hard to build a stable connection between the node recording audio and the node of displaying speech text based on our knowledge of IoT and skill of coding at the current stage. Therefore, we just give up on making it synchronous, and we switch to working on converting recorded speech. To solve this problem and make a better tool for converting speech to text, we will delve into improving it to be synchronous in the future.