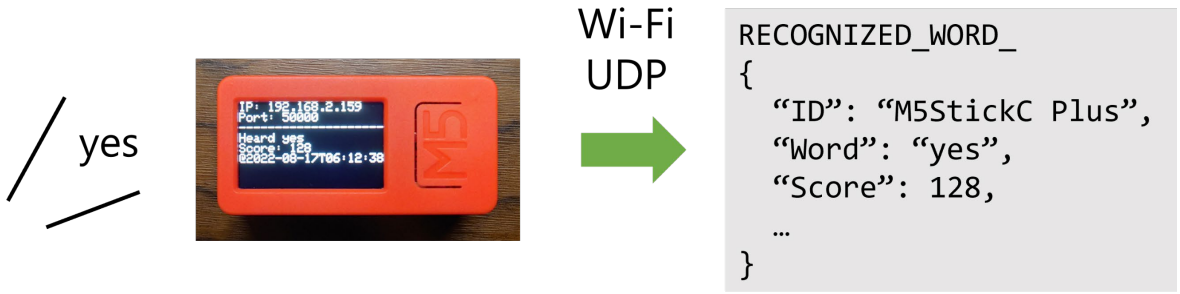
Speech Recognition and UDP Communication

This program uses a neural network to process the audio picked up by the M5StickC Plus microphone, and when it recognizes a specific word, it sends a message via UDP communication to a PC connected via Wi-Fi. By default, it only recognizes "Yes" and "No". You can record your own voice and train the neural network to recognize other words.



Quick Start

Let's try out the default program, which recognizes only "Yes" and "No". We will need the followings to run the program.

- M5StickC Plus
- USB cable
- Installing the driver
- Installing PlatformIO (Visual Studio Code extension)

Installing the Driver

Connect M5StickC Plus and your PC with a USB cable. Open Device Manager and install the driver. You can download the driver from this page.[1] For more details, refer to the official setup guide[2] and follow the instruction on "Driver Installation." (You can ignore the other sections, such as installing Arduino-IDE.)

Installing PlatformIO

We will use PlatformIO, a Visual Studio Code extension, to compile and upload the program to device. Search for "PlatformIO IDE" in the VS Code extension tab and install.

---

[1] https://ftdichip.com/drivers/vcp-drivers/

[2] https://docs.m5stack.com/en/quick_start/m5stickc_plus/arduino

Follow the instruction to complete the installation. However, anti-virus software might hinder you from installing or running PlatformIO afterwards. In that case, re-configure your anti-virus settings and make sure “platformio.exe” and “pio.exe”, which should be in “C:¥Users¥{username}¥.platformio¥penv¥Scripts¥” or somewhere similar, to be able to run the programs.

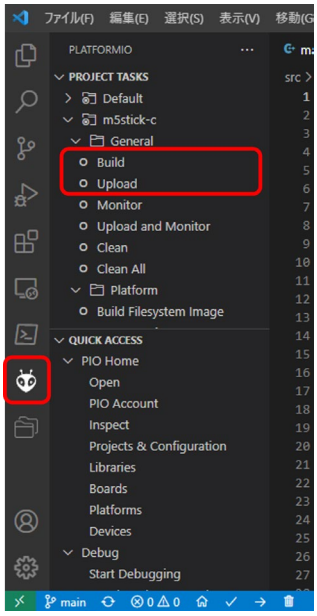
Run the Program

Unzip “speech-recognition-udp.zip” and open the folder in VS Code. Open “./include/wifi_settings.h” and fill up your Wi-Fi settings (`ssid`, `password`) and IP address and port number for your PC (`send_to_ip`, `send_to_port`).

```
constexpr char ssid[] = "SSID";
constexpr char password[] = "PASSWORD";

constexpr char send_to_ip[] = "192.168.XXX.XXX";
constexpr int send_to_port = 50001;
```

After that, open PlatformIO tab and press “Build” to compile the program. Then, make sure M5StickC Plus is connected to your PC and the power is on and press “Upload” to upload the program to device. When the upload is done, the program should start automatically. (If it doesn’t start, press the power button.) You can now unplug the cable after the upload.

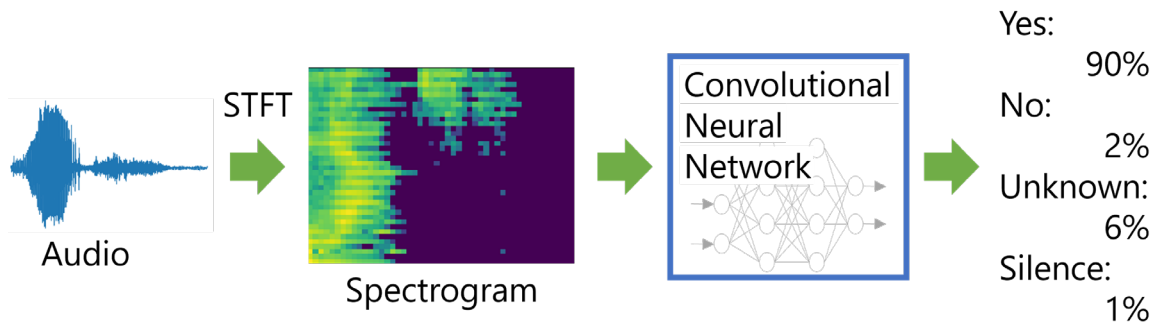


Say “yes” or “no” aloud and it will recognize the word and send messages to the specified address and port. You can check the messages using any UDP communication application. (If you don’t receive any messages, check the address and port number or your firewall settings.)

Press the power button to restart. (When the program is run for long time, it will stop working properly. Press the power button to reset.) Long press the power button to turn off the device.

How It Works

Here is the general flow of this speech recognition program.



First, the audio waveform picked up by the microphone is converted into a spectrogram by an operation called short-time Fourier transform (STFT). A spectrogram is represented as a 2D image. The horizontal axis represents time, the vertical axis represents frequency, and the color represents the intensity of the sound at that time and frequency. By converting to a spectrogram, sound information can be downsized. Then the spectrogram is input to the neural network, which outputs the probability of which word it is. This program averages over multiple predictions over time to compute the probability for each word, and it recognizes the word when it exceeds a given threshold.