

音声認識プログラム UDP

このプログラムは、M5StickC Plus のマイクで拾った音声をニューラルネットワークで処理して、特定の単語を認識すると Wi-Fi でつながっている PC に UDP 通信でメッセージを送信します。デフォルトでは「Yes」と「No」しか認識しませんが、自分の声を録音してニューラルネットワークを訓練することによって、他の単語を認識させるようにすることができます。



音声認識プログラムを動かす

まずは、デフォルトの「Yes」と「No」を認識するプログラムを動かしてみましょう。このプログラムを動かすには以下が必要です。

- M5StickC Plus
- USB ケーブル
- ドライバーのインストール
- PlatformIO (Visual Studio Code の拡張機能) のインストール

ドライバーのインストール

M5StickC Plus と PC を USB ケーブルでつないだら、デバイスマネージャーを開いてドライバーをインストールします。ドライバーのダウンロードページはこちら¹です。詳細は公式サイト²の「Driver Installation」に従ってください。(それ以外の Arduino-IDE のインストールなどは今回は必要ありません。)

PlatformIO のインストール

プログラムをビルドしてデバイスにアップロードするために Visual Studio Code の拡張機能である PlatformIO を使用します。VS Code の拡張機能のタブで「PlatformIO IDE」と検索してイン

¹ <https://ftdichip.com/drivers/vcp-drivers/>

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

ストールしてください。

指示に従えばインストールできるはずですが、環境によってはアンチウイルスソフトによってインストールできなかったり、インストールできてもその後実行できなかったりする場合があります。その時は、「C:\Users\{username}\.platformio\penv\Scripts\」などにある、「platformio.exe」や「pio.exe」が実行できるようにアンチウイルスソフトの設定を見直してください。

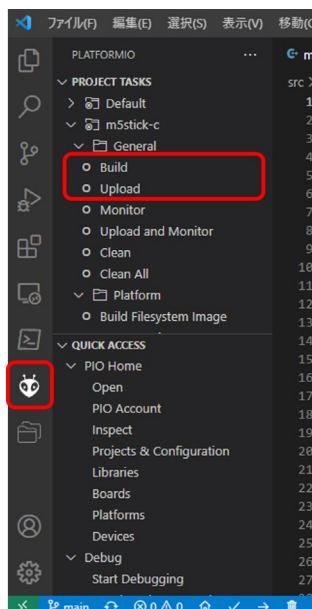
プログラムを実行する

配布されている「speech-recognition-udp.zip」を解凍して、そのフォルダを VS Code で開きます。そして、「./include/wifi_settings.h」を開いて、ssid, password の欄に Wi-Fi の設定, send_to_ip, send_to_port の欄に PC 側の IP アドレスやポート番号を記入します。

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

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

その後、PlatformIO のタブを開いて、「Build」を押してプログラムをコンパイルします。それが終了したら、M5StickC Plus が PC とつながっていて電源が入っていることを確認してから「Upload」を押して、プログラムをデバイスにアップロードします。アップロードが終わると自動でプログラムが開始します。（開始しない場合は電源ボタンを押してください。）アップロードが済んだらケーブル接続は外しても構いません。



側面
Side
↑
電源ボタン
Power Button

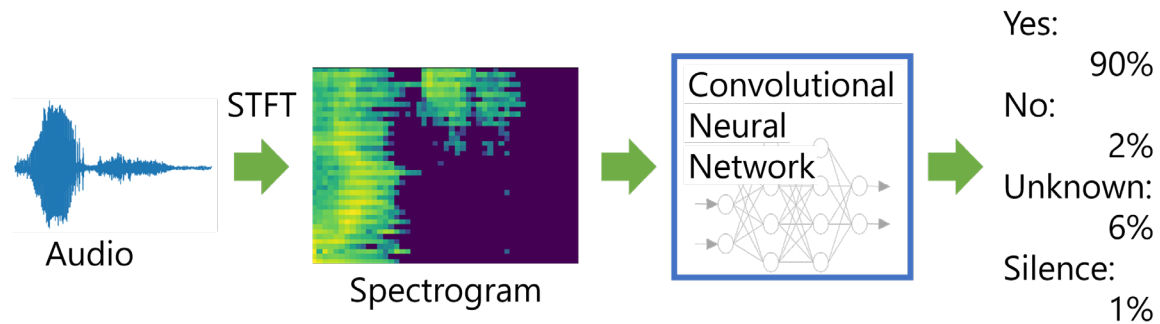
「yes」や「no」と発声すると単語を認識して、メッセージが指定したアドレス・ポートに送信されます。任意の UDP 通信アプリを使用して確認することができます。（メッセージが受信できない場合は、アドレス・ポート番号やファイアウォールの設定を確認してください。）

電源ボタンを押すとプログラムを再起動することができます。（長時間起動していると正常に動作

しなくなるため、電源ボタンを押してリセットしてください。) 電源ボタンを長押しすると電源を切ることができます。

音声認識プログラムの仕組み

ここでは、このプログラムの大まかな流れを説明します。



まず、マイクで拾われた音声波形は、短時間フーリエ変換 (STFT) という演算をすることによってスペクトログラムに変換されます。スペクトログラムは 2D 画像のように表されます。横軸が時間、縦軸が周波数で、色がその時間、その周波数帯の音の強さを表します。スペクトログラムに変換することにより、音の情報をコンパクトにすることができます。そして、スペクトログラムがニューラルネットワークに入力されて、どの単語であるかの確率が出力されます。このプログラムでは、さらに複数回の予測の平均を取ってそれぞれの単語の確率を求めて、ある閾値を超えたら単語を認識します。