

基於深度學習的即時手語辨識

Realtime Sign Language Recognizer Based on Deep Learning

指導教授：王士豪

專題成員：賴佑寧、李易宗、丁琪

開發工具：Python、TensorFlow、OpenCV、
PyQt5

測試環境：Windows 11

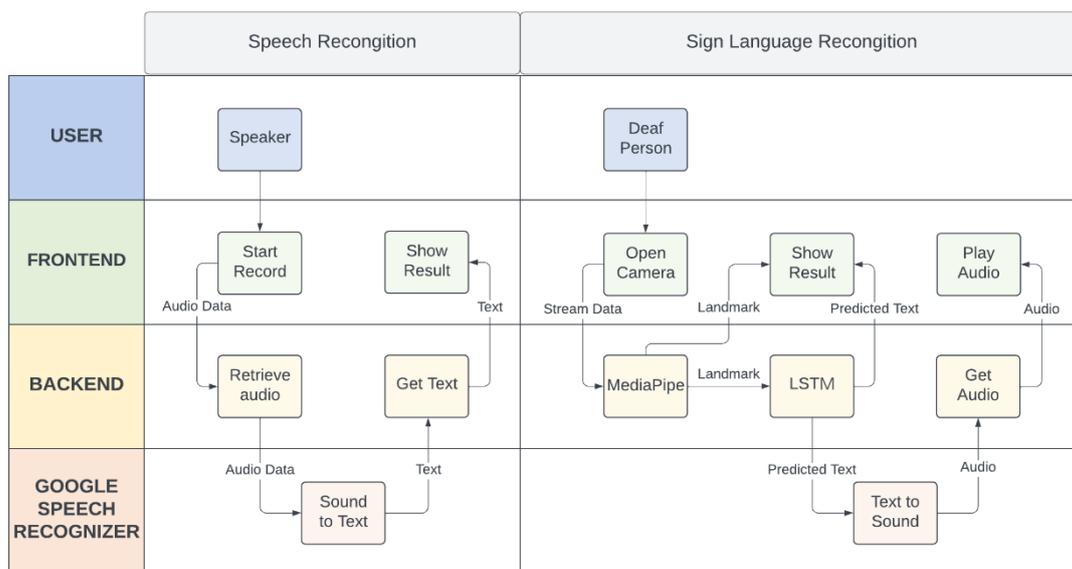
一、簡介：

經由世界衛生組織的調查結果顯示，到了2025年，預計有近 25 億人有一定程度的聽力損失，至少有 7 億人需要聽力康復。且根據發展心理學的研究，聽障人士在早期語言發展部分與一般聽人有所不同。一般的小朋友可以透過自己的母語來認識這個世界，但先天性聽障的小朋友使用的是手語來認識世界，因此手語對大多聽障人士而言就是他們最熟悉的母語。對於文字理解的部分，聽障人士在學習與使用文字方面其實較一般聽人困難許多。因此在透過說話與文字溝通都有困難的情況下，人與人之間的距離感不可避免地產生。

由於世界各國使用的手語並不相同，因此我們以臺灣使用的手語設計一套系統。系統包含即時的手語辨識與語音辨識兩個功能，目的為方便聽障人士與一般聽人溝通。在使用者界面的部分模擬聊天室的模式，透過按鈕開始錄製手語或語音，並即時將辨識的文字結果顯示在輸入框；再次按下按鈕即可結束錄製，並送出輸入框中的文字。聽障人士可以使用手語辨識功能在鏡頭前比出手語，利用深度學習將手語轉換為文字，文字會顯示於聊天室且同時以語音撥放文字內容；一般聽人則可以使用語音辨識功能，將說話內容轉為文字顯示於聊天室。如此聽障人士與一般聽人便可以藉由此系統達到雙向溝通，聽障人士可以接收到文字訊息，一般聽人則可以接收到文字與語音訊息。

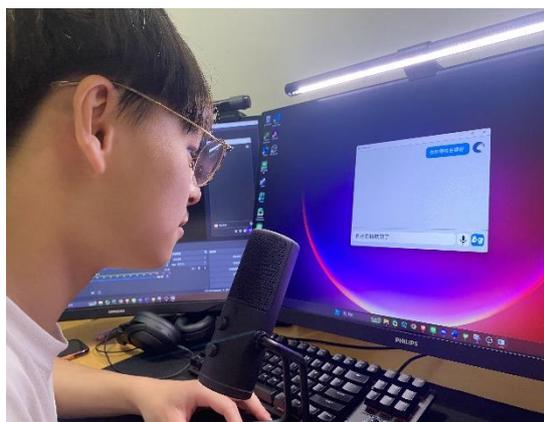
手語辨識部分是使用深度學習來實作，自行錄製的手語動作作為資料集，使用 MediaPipe 偵測人體骨架33個點的三維座標與是否可見、左右手各21個點的三維座標。單一手語動作使用30個 frame 的骨架座標作為模型輸入資料，並且以 LSTM 為基礎模型架構來訓練模型。語音辨識部分將即時錄製到的語音降噪後分段，再將分段的語音傳送到 Google API 取得語音轉文字的結果。語音合成同樣使用 Google API 取得文字轉語音的結果。

以下為系統流程圖：



二、測試結果：

1. 即時語音辨識



2. 即時手語辨識

