附件B

**2024淡江大學資工週專題說明書**

**一、專題名稱：**日文發音評分系統

**二、指導教授：**陳建彰

**三、專題組員(姓名/系級)：**林聿朔/資工三A、蘇柏修/資工三A、吳天宇/資工三A、施吉益/資工三A、陳緯榛/資工三A

**四、關鍵字:** 1.發音評分 2.網頁 3. 自監督模型

**五、專題內容簡略描述:**

**專題目的**

在全球化趨勢的推動下，日語的學習對於世界各地的學習者來說變得日益重要。儘管如此，我們學校在提供精準評估學生日語發音標準和快速反饋方面還存在一定的空缺。目前缺乏一個合適的系統來協助學生在發音上自我改進，同時能減輕教師在評估學生發音時的工作壓力。本專題旨在填補這一空白，透過開發一套自動發音評估系統，以支持我們學校日語系教育的需求。

我們選擇了自監督學習模型(SSL)作為開發此系統的基礎，這種模型可以直接從原始的語音音檔中學習到豐富的語音特徵，從而免去了傳統語音處理中對特徵工程的依賴。這一創新方法不僅提高了發音評估的精準度，而且簡化了評估過程，讓反饋更加即時和準確。

此系統的開發將賦予我們學校一個方便、快速的教學輔助工具，幫助教師更有效地進行教學，並提供給學生隨時隨地都可練習的便利。透過此系統，我們期待學生能夠在課餘時間，更有效地提升自己的日語發音能力。

**使用技術**

1. 特徵提取

本專題使用經過專門在日文語音集上訓練的HuBERT模型來進行語音特徵的提取。HuBERT模型是一種自監督學習模型，它通過預測隱藏的語音單元來學習語音表示，而不需要任何手工標記的轉寫。在我們的系統中，HuBERT模型經過日語語料的預訓練後，進行了針對自動語音識別(ASR)的微調，使其能夠更精確地識別和分析從基本的聲學特徵到複雜的語義層面的各種信息。這對於後續評分模型預測評估發音的準確性至關重要。

1. 語音辨識

為了提高發音評分的準確性，我們的系統同時利用自動語音識別(ASR)技術將語音轉換為對應的文本。這一步是必要的，因為評分模型在進行發音評分時不僅需要語音的聲學特徵，還需捕捉語音內容的文本。將語音轉換為文本後，我們將這些文本信息轉化為嵌入向量，這些向量與HuBERT模型提取的聲學特徵一同作為評分模型的輸入。

1. 評分模型

評分模型使用的是雙向長短期記憶網絡(BLSTM)，這種網絡結構非常適合於處理語音評分的問題。BLSTM能夠有效地處理時間序列數據，學習語音在時間上的依賴關係，並且可以同時考慮過去和未來的上下文信息，這對於發音評估尤為重要。在我們的系統中，BLSTM用於分析從HuBERT模型提取的特徵及相對應的文本信息，並根據這些信息對發音進行評分。

1. 降噪模型：

在語音處理系統中，降噪是一個重要的挑戰，尤其是在非理想的錄音環境中。我們的系統採用了FRCRN降噪模型來改善錄音質量。FRCRN結合了卷積神經網絡(CNN)和循環神經網絡(RNN)的優點，有效地從語音錄音中去除背景噪音。卷積層負責提取語音特徵，而循環層則處理時間序列中的信息，這樣的結合不僅提高了降噪的效果，也保持了語音數據的時序完整性。這使得即使在噪聲環境下也能提供安靜穩定的音檔，使系統能準確地進行發音評估，大大提高了系統在各種使用條件下的實用性和可靠性。

1. 網站開發：

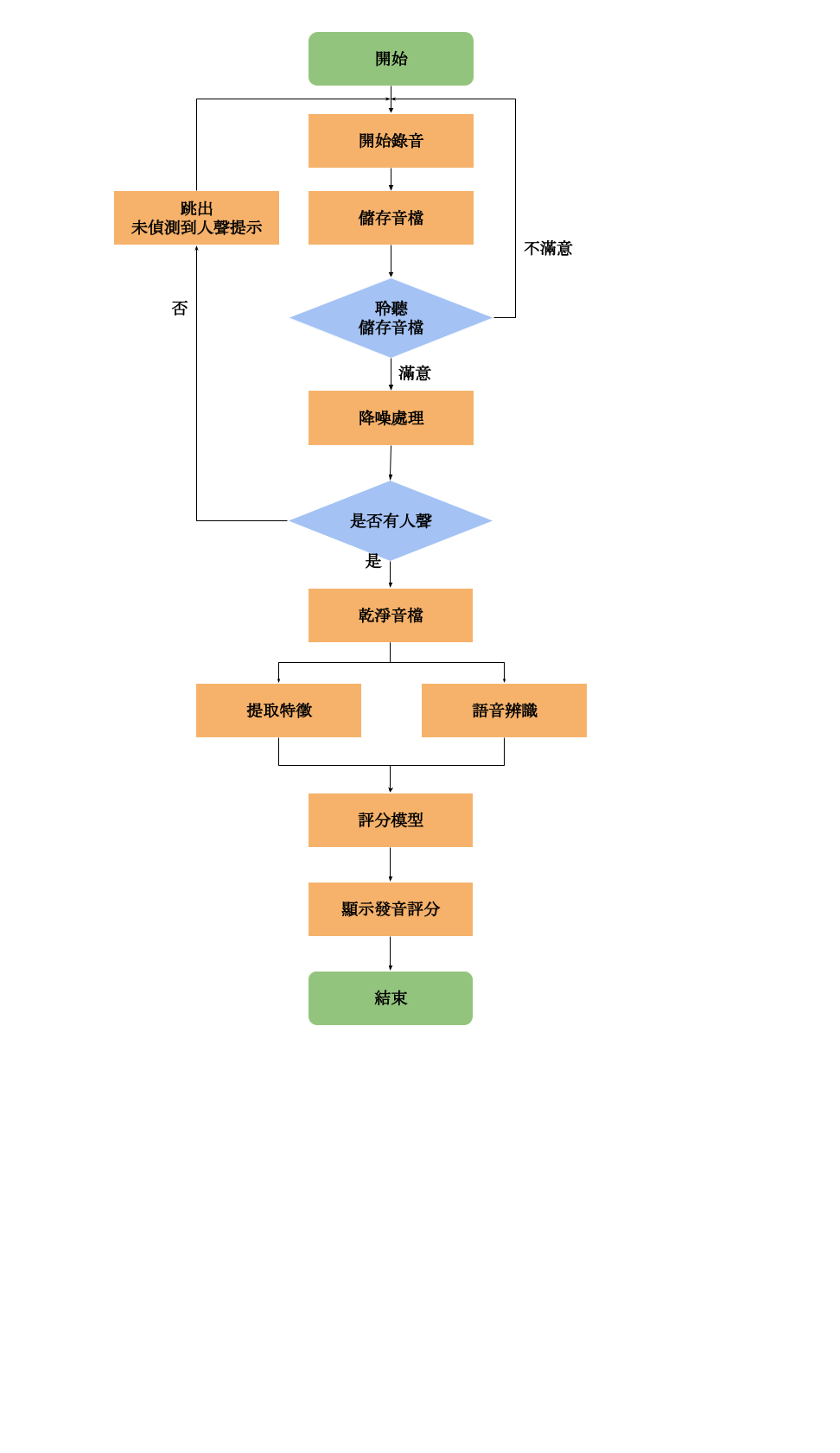
我們選擇了Flask作為我們的Web應用框架，因為它的簡潔性和靈活性特別適合快速開發並部署我們的自動發音評估系統。Flask基於Python，這與我們後端處理技術的相容性完美結合。此外，Python生態系統提供的豐富庫使得我們能夠為用戶打造一個功能全面而穩定的平台。除了Web平台，我們也將網站功能打包成APP，為使用者提供更便利的使用方式。使得我們的系統不僅限於桌面使用，而是能在更多的設備上提供相同的精準評估分析。這樣無論是在移動設備或是桌面環境中，學習者都能夠輕鬆地訪問我們的系統，並在日常生活中實踐和提高日語發音水準。

**功能描述**

我們的自動發音評估系統透過一個友好的網站介面向使用者提供即時的服務。為學習日語的人士提供一個簡便的途徑，來檢測和評估他們的語音發音是否標準。使用者只需在網站上按照圖示進行幾個簡單的步驟，錄製或上傳語音檔案，剩下的工作就交給系統來完成。

當使用者上傳語音後，系統會立即進行處理。與傳統的語音評估工具不同，我們的系統專注於提供明確且直觀的反饋—語音發音是標準還是不標準。簡潔的反饋機制讓學習者能夠快速識別出自己發音的準確性，從而在日後的學習中進行方向性的改進。

圖一為系統運作過程，

****

圖一、系統流程圖

以下會逐一介紹網頁的使用過程：

一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

自動產生的描述

**1.按允許才能開始**

一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

自動產生的描述

**2.點擊按鈕開始講話**

一張含有 文字, 螢幕擷取畫面, 軟體, 網頁 的圖片

自動產生的描述

**3.講完後按結束**

一張含有 文字, 螢幕擷取畫面, 軟體, 網頁 的圖片

自動產生的描述

**4.上傳後顯示評分結果**

**可以播放剛才錄音的音檔**

一張含有 文字, 螢幕擷取畫面, 軟體, 多媒體軟體 的圖片

自動產生的描述

**(請將專題說明書之電子檔於4月26日中午前交至)**