# **專題成果報告-語者分離技術輔助學習障礙診斷**

國立中央大學資訊工程學系 王昱翔

1. **摘要**

學習障礙乃是指兒童智力正常，但在 聽、說、讀、寫、算，有一個或多個的困難，及早診斷學習障礙，對孩童的發展相當重要。現今的診斷方法是讓受試者接受測驗並錄音，再讓醫生針對錄音判斷受試者的狀況，然而現行的學習障礙診斷中並沒有統一錄音環境，造成語音大小差異大和環境噪音容易過大，再加上能取得的診斷錄音過於稀少，使學習障礙的診斷難以透過現成的機器學習模型作為輔助。

本專題**針對錄音品質不佳的狀況**，透過語者分離 (Speaker Separation) 技術**分離受試者的語音訊號**，再透過語音辨識模型計算有助於診斷的特徵，進而協助醫生直接分析受試者的語音，免除噪音或是其他語音的影響，幫助醫生進行診斷，**節省醫生診斷的時間，**同時**本專題已取得台北護理健康大學暨萬芳醫院教授翁仕明醫師同意，將支援資料收集與領域知識之協助指導**。

1. **研究動機與研究問題**

依我國《身心障礙及資賦優異學生鑑定辦法》定義**，**學習障礙屬於身心障礙類別中的一種，無關智力，因某種功能異常，使患者在注意、記憶、理解、推理、表達或動作協調等能力異於常人，導致患者在生活中的聽、說、讀、寫與計算上，遭遇一項或多項的困難，進一步影響在學校的學習狀況，衍伸信心不足、學校適應不良、人際關係不佳等情況，及早診斷出學習障礙對兒童發展相當重要。

目前**學習障礙的診斷主要有三個問題**:

1. 診斷流程耗費時間與精力。由於醫生需要親自寫錄音的逐字稿，再藉逐字稿判斷受試者是否有理解能力、推理能力和表達能力上的問題，全程都倚靠人力。
2. 逐字稿會喪失有用資訊。由於語句間停頓時間還有說話速度都可能和受試者的表達能力有關係，但是單純的逐字稿無法表現出這種特徵。
3. 診斷缺乏統一客觀標準。實務情況下，無法斷定「有某些特徵的人」，就一定有學習障礙，導致診斷時醫師往往得依靠個人的直覺。

若想加速學習障礙診斷，自動化的語者分離 (Speaker Separation) 將不可或缺。現今，雖然機器學習在語者分離任務的表現已經優良許多，但是應用在學習障礙診斷的錄音上仍有問題，像是診斷錄音稀少而不足以訓練一個模型，受試者的音量過小而被當作噪音，噪音過大導致無法分辨出語音，或是因為模型不曾以孩童的語音訓練，使**模型在分離受試者語音時效果不佳**。

本專題將透過研究由**台北護理健康大學暨萬芳醫院教授翁仕明醫師收集的資料**，並以**語者分離**為核心技術，**設計一套流程與方法，解決引入機器學習至學習障礙診斷所面臨的首要問題**，如語音的音量過小，環境噪音過大以及孩童的聲紋特性，造成模型表現不佳，以加速深度學習輔助學習障礙診斷的進程 。

1. **研究方法**
   1. **資料分析**

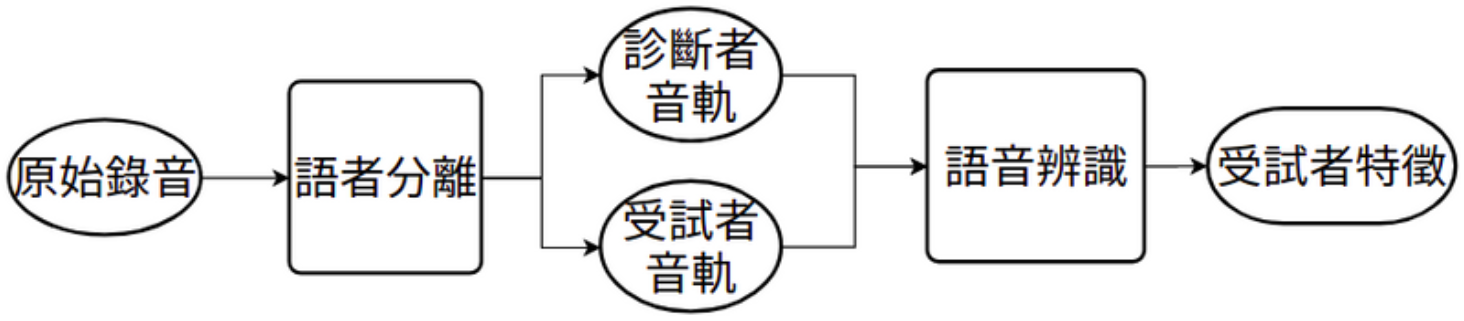
本專題所使用的診斷錄音，**皆由台北護理健康大學暨萬芳醫院教授翁仕明醫師所提供**，其內容為受試者接受各種測驗的錄音，因此主要包含受試者與診斷者兩人的語音，然而現行的學習障礙診斷中並沒有統一錄音環境，導致診斷錄音品質參差，可能會出現環境噪音過大或是受試者聲音過小，甚至是出現他人語音等問題，造成建立診斷輔助工具的主要障礙，診斷錄音資料集的狀況如表一所示。

表一、 資料集概況

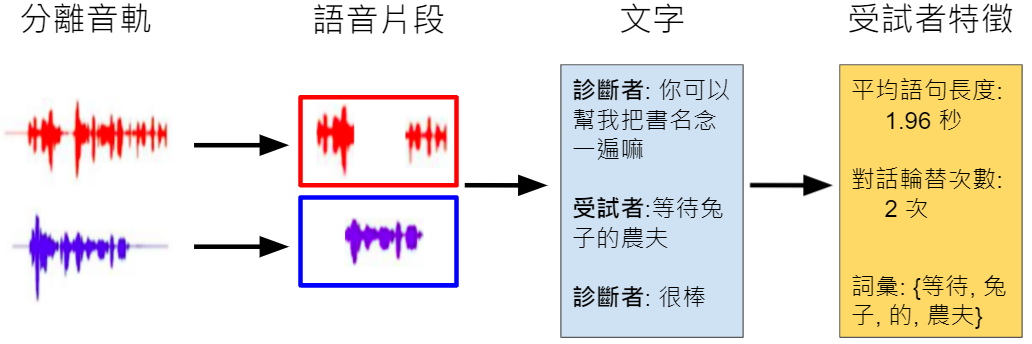
|  |  |  |  |
| --- | --- | --- | --- |
| 診斷錄音資訊 | **首批** | **單側聽損** | **萬芳** |
| 筆數 | 10 | 14 | 8 |
| 總長 | 3 小時 | 5.5 小時 | 6 小時 |
| 主要狀況 | 人聲大小適中  幾乎無噪音 | 兩人以上人聲  受試者聲音過小  有小噪音 | 強烈噪音干擾人聲 |

* 1. **系統設計**

本專題建構的系統運行流程如圖一所示。系統是由語者分離與語音辨識模組所組成，前者的輸入為混雜語音和噪音的原始診斷錄音，利用我們**自行訓練的 Conv-TasNet** 分離出僅有單人語音的的診斷者與受試者音軌，後者利用 WEBRTC VAD偵測分離音軌中的語音片段，再使用Google Speech Recognition API 將語音片段轉為文字，最後透過 CKIP Transformers 中的斷詞服務對文字進行斷詞，得到與學習障礙相關的特徵，如平均語句長度、對話輪替次數以及詞彙等有助於診斷的特徵。語音辨識模組的功能如圖二所示。



圖一、 系統流程圖



圖二、 語音辨識模組運行概念圖

* 1. **資料集建構**

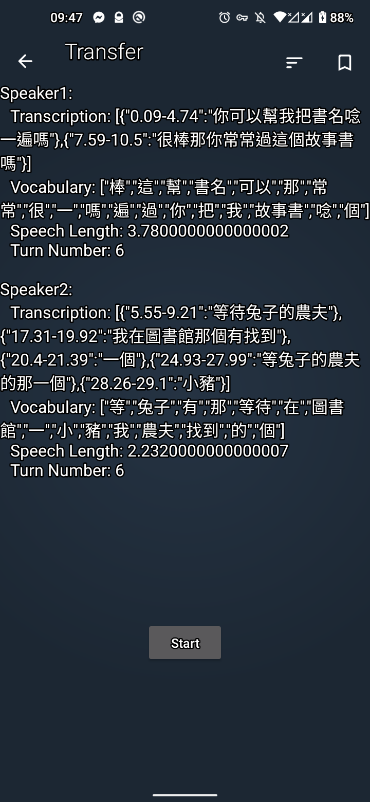
由於診斷錄音稀少造成語者數量受限，加上錄音環境常有噪音，我們除了使用首批、單側聽損以及萬芳三批診斷錄音建構語者分離資料集以外，也加入 LibriSpeech ASR 資料集以增加語者多樣性，並混入WHAM 噪音資料集以進行資料增強。為了使用診斷錄音建構語者分離資料集，必須先標記診斷錄音中受試者與診斷者語音的時間戳記，以得到可以用來製作訓練資料的語音片段。藉由隨機混合兩個語音片段以及一段噪音作為模型的輸入音訊，原先的語音片段作為模型的預期輸出，生成語者分離資料集後，再隨機分為70%的訓練資料和各15%的驗證資料以及測試資料。

* 1. **建構模型**

我們選擇Conv-TasNet作為系統中的語者分離模型，因為其分離效果不錯，也在許多研究論文中作為 baseline。透過使用開源的 Asteroid 語者分離套件建構Conv-TasNet 模型，超參數設定皆為 librimix 訓練腳本的預設值。由於資料集內的預期輸出可能帶有噪音，僅使用語者分離指標 (SI-SDR) 可能無法準確評估模型分離成效，所以我們也另外用字符錯誤率 (CER) 評估語音辨識的結果，以間接評估語者分離模型的成效 。

1. **成果與展望**

經過測試發現超參數中的語音長度會影響模型表現，設定在一至二秒時會導致模型無法有效分離語音，而長度設為三秒時表現最佳，因此最後設定語音長度為三秒。目前系統於測試資料集達到SI-SDR 7.8，CER 33%， 而且語者分離模型能在多數情況下抑制噪音，但是未能處理語者聲音過小以及多語者的情況。與此同時，我們也建立了Demo 用 APP展示系統輸出，為讓醫師使用此系統做準備。Demo用APP的運行結果範例如圖二所示。



圖三、 Demo 用APP 語者特徵截圖

未來的首要目標是建立診斷輔助服務以供醫師使用，以根據醫師回饋調整系統，而系統可能的改進方向有，改用一次只分離一個語者的模型，以解決多語者以及語者聲音過小的問題，或是跟換語音辨識模型，以提高字詞辨識率。等系統成熟時，則可以增加自然語言處理和異常檢測功能至系統，分析語音中有助於診斷的特徵，以及預測語者為學習障礙患者的機率，向醫師提供更多種分析方式的結果，進一步加速診斷的速度。