聽障者多模中文口語訓練模型與分析
(Multimodal Speech Training for the Hard of Hearing in Mandarine)

何慶祥
龍華科技大學
ee071@gm.lhu.edu.tw

曾坤川
和春技術學院
kc.tseng@gmail.com

雷曼菁
和春技術學院
mc.lei@gmail.com

摘要

本研究為協助聽障人士提升口語發聲能力和溝通品質，提出將語音的聲學模型轉換以視覺形式呈現，輔助聽障人士進行發聲訓練，並建構口語發聲練習平台，提高聽障人士發聲練習的自主性，為聽障人士實現更加明確的聽與更清晰的說的目標。

本研究將韻母、聲母與聲調等語音元素的聲音訊號，藉聲學模型轉換為可對應發聲原理的頻譜分布與動態變化。學習者可看見韻母發聲過程的軌跡，是否符合口腔與舌位的變化，聲母的發聲動作與出氣方式是否有正確的隨時間變化，聲調的掌握是最困難的，本研究採用以 3 個狀態過程來描述 4 種聲調。經過聽覺訊息視覺化以及視覺與觸覺的多模回饋，聽障者可以藉以自主練習提升發音的清晰度。

Abstract

This study aims to assist the hearing-impaired in improving their vocal ability and communication quality and proposes to convert the acoustic model of speech into visual form to assist the hearing-impaired in self-vocalization training. Establish a platform for oral vocalization practice to assist the hearing-impaired to learn independently. The core goal of this platform is to enable hearing-impaired people to hear more clearly and speak more clearly.

In this study, the sound signals of speech elements such as vowel, consonant, and tone were converted into spectral distribution and dynamic changes that could correspond to the principle of sound pronunciation through the advanced model. Learners can see the trajectory of the vowel vocalization process, whether it meets the changes in the oral cavity and tongue position, whether the consonant's vocal action and exhalation mode change correctly with time, and the mastery of tone is the most difficult, this study uses 3 state processes to describe 4 tones. Through the visualization of auditory information and multimodal feedback of sight and touch, the hearing impaired can improve their pronunciation through self-practice.

關鍵字：聽障、聲學模型、視覺形式、多模
Keywords: hard of hearing, acoustic model, visual form, multimodal

1 簡介

患有嚴重聽覺障礙者與正常聽人的溝通過程中，如何讓交談的對象，能理解聽障者欲表達的內容，對聽障者非常重要也最具挑戰。

聽障者若自幼即嚴重聽力損失，如何在語言學習黃金期有效的接受語言教學指導，學習正確有效地發聲，對口語能力養成非常重要；即便是已具語言能力才面臨聽力損失障礙，仍須要接受口語教育，方能維持較正常的發聲能力。然而，無論是先天或後天造成聽損，聽障礙人士的語言表達，會因無法有效的接收聽覺回饋，影響發聲的控制甚至語言的使用，並逐漸惡化，讓聽障者更不願意以口語進行互動(林珮瑜等，2006)。

改善聽障者發聲能力及品質，可以從三個面向來達成(林珮瑜等，2006)，分別是(1)適性的發聲指導，例如：具專業能力及經驗的語言教學或語言治療專家，提供適性的專業指導；(2)聲音知覺的強化，例如：使用助聽器、人工電子耳等，同時強化對殘存聽覺認知的訓練；(3)說話訊息的回饋，例如：輔具、聲音訊號的解析，或說話者的動作或臉部表情。

然而，因醫療及語言專業人力有限，若能藉由電腦軟硬體及資通訊技術的協助，提供更有效的訊息接收及發送方式協助聽障者，方能使溝通過程更方便及正確的完成訊息傳遞，協助聽障者在友善的環境，聽得更清楚，說得更清晰(Thida et al., 2020; Virkkunen et al., 2019; Lukkarila, 2017)。

2 文獻探討

AssistiveWare(2023)指出圖板或觸控式螢幕為最簡易的輔助構通(Augmentative and alternative communication, AAC)裝置，可以用來取代或輔助口語溝通。提供圖片或符號用於特定項目與活動，可供大部分日常生活需要。這些裝置可藉鍵盤、觸控螢幕或使用者特定話語，來傳遞意圖；使用顯示面板朝上的文字顯示器，讓兩人容易面對面交流訊息；運用拼字軟體加快訊息輸入速度；或將文字圖片轉換為話語，甚至可選擇聲音，例如男女、大人小孩或不同口音。

Snips (Coucke et al., 2018)是一個以 AI 技術為基礎且及時聯網的語音平台，提供裝置間的互動運作以及客製化的語音經驗，Snips 是採用個人專用設計(Private by Design)技術的語音輔助系統，運作於 edge 平台。蘋果的 iOS 裝置 iPhone、iPad 或 iPod touch 可藉著即時聆聽功能，將聲音傳送到助聽裝置，進行聲音串流、來電回應、調整設定等，同時可協助使用者，在噪音環境中進行對談，或是聽到不同房間的談話。

Jain et al. (2016)提出以口腔型狀的 2D 動態呈現，作為語音發音訓練的輔助。學齡前的聽力障礙兒童由於缺乏聲音回饋，接收語言訊息有困難，若經由口語輔助訓練，特別是對關鍵發聲動作的視覺回饋，對比學習者與教師或參考語者的發聲動作，可獲取正確的回饋資訊。系統提供口腔形狀的動態呈現，及產生語音的視覺回饋，可呈現聲波、聲音強度、頻譜圖、基頻以及口腔形狀圖，做為語音訓練使用。

Dudy (2016)提出自動發音分析系統(automatic pronunciation analysis system)，指出學前與學齡兒童約有 10%受到音韻障礙的影響，使得人際互動與溝通以及學業表現不佳。有效的發聲訓練通常需要較長時間的練習與互動，多數兒童不易獲得口語語言病理專家的協助。因此使用電腦輔助發聲訓練，包括：從大量兒童口語資料庫進行聲學模型訓練；訓練目標族群的錯誤發聲模型；訓練錯誤發聲音素的正確聲學模型等(McKechnie et al., 2018; Lee er al., 2015)。

Virkkunen et al. (2019)提出聽障者對話輔助系統，解決聽障者參與對談時的困擾，包括多方談話時的交互影響，或來自環境的聲學干擾。研究亦針對聽力障礙者，對基於自動語音辨識技術的個人談話助理的喜好。建構了兩個原型平台提供聽障者使用，其中一組使用行動裝置，採用擴增實境技術，讓聽障者可同時觀察說話者的動作及嘴型，同時有 ASR 即時翻譯的文字(Usha and Alex, 2023; Qu et al., 2017; Jian et al., 2015)。

國內學者(賴俞靜、劉惠美，2014；李芃娟，1999；張小芬等，2014；張蓓莉，2000)使用語音聽力檢測系統，輔助進行聽障兒童發音教學實驗，分析聽障兒童的音標學習、聽音仿說以及口語教學成效等的相關性。結果顯示輔助系統可以有效地進行教學成效的檢測，也有助於提升部分學生聽音仿說的能力。亦針對知覺障礙學生說話清晰度做知覺分析研究，發現平均而言聽障生的語詞清晰度為 30.74%，聲調清晰度為 53.92%，短句清晰度為 49.83%，且中重度聽障學生的表現優於重度聽障學生。語詞、聲調、短句之清晰度彼此有相關性，以發音部位為舌面後音的正確率最高，舌尖音發音正確率最低。以發音方法分析，發音正確率最高的是邊音，正確率最低的是塞擦送氣音。

3 多模口語訓練

3.1 聲學模型

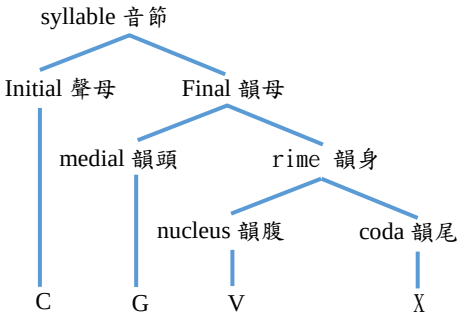

圖 1. Initial-Final 音節樹狀結構模型

中文音節可以使用 Initial-Final 模型來表示，其結構可以用樹狀圖表示，如圖 1。其中 C=聲母(consonant)、G=滑音(glide)、V=韻母(vowel)、X=聲母或韻母。韻母分韻頭與韻身，韻身可分韻腹與韻尾，韻腹又稱為主要韻母(main vowel)為必要單元；韻頭為銜接聲母與韻腹的發聲過程，有ㄧ、ㄨ、ㄩ三個介音；韻尾為複韻母的結尾，如ㄞ的韻腹為ㄚ韻尾為ㄧ，或為鼻韻母的結尾鼻腔音，如ㄢ包含韻腹ㄚ與ㄋ相同的韻尾(Triskova, 2011)。

每一個中文字為一個音節，音節的發聲包括 3 個單元，分別是聲母(Initial 或 consonent)、韻母(final 或 vowel)及聲調(tone)。聲母在前韻母在後，有些音節不具有聲母，但韻母是必要的單元，每個音節具有一個聲調。在 Initial-Final 音節模型中，每音節由 Initial 與 Final 兩個發音單元組成，如「在」由聲母ㄗ及韻母ㄞ組成，聲調為 4 聲。Initial 代表位於音節開始的發音單元，但不是必要單元，通常為聲母；final 為音節結束的單元，為必要單元，通常為韻母；聲調則有四種變化分別是 1~4 聲，輕聲則是一種短促發聲過程。音節結構亦可以用堆疊架構表示如圖 2(Triskova, 2011)。

<table>
<tr><td rowspan="4">Initial
(聲母)</td><td colspan="3">Tone(聲調)</td></tr>
<tr><td colspan="3">Final(韻母)</td></tr>
<tr><td rowspan="2">Medial
(韻頭)</td><td colspan="2">Rime(韻身)</td></tr>
<tr><td>Nucleus/
main Vowel
(韻腹)</td><td>Coda/Ending
(韻尾)</td></tr>
</table>

圖 2 中文音節的堆疊架構模型.

另外，音節亦可採用狀態圖來描述，如圖 3，其中 Initial(聲母)單元包含狀態 I，Final(韻母)單元中 M 表示 Medial(韻頭)、N 表示 Nucleus(韻腹)、C 表示 Coda(韻尾)、T 表示 Tone(聲調)。國語發音的音節一定具有韻腹，故圖 3 中的狀態 N 不能被跳過。每一個結束單元(Final)會對應到一種聲調，因此聲調會跨越多個 Final 單元的節點。

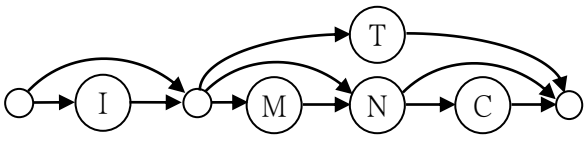

圖 3. 國語音節狀態圖結構模型

3.2 韻母模型

依據單韻母共振峰位置與發聲時的口型及舌位的關聯，在複韻母發聲過程中，共振峰座標移動路徑的描繪，可作為視覺形式的發聲修正參考。正常發聲語者之 F1-F2 及 F1-F2-F3 座標分別為圖 4 與圖 5(Jain et al., 2016)。

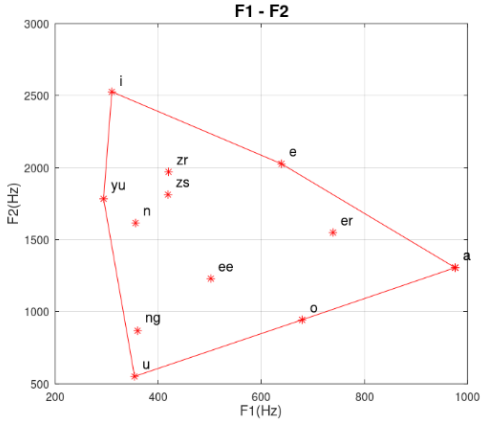

圖 4 韻母 F1-F2 位置圖

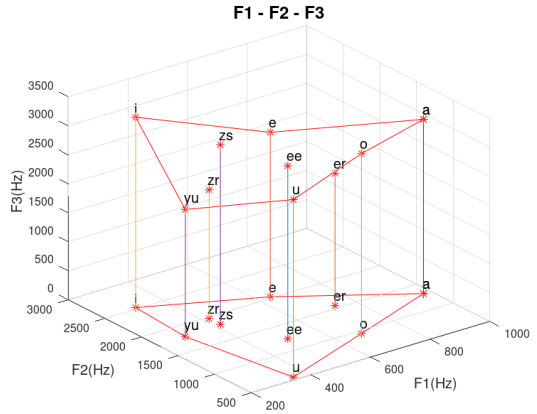

圖 5 韻母 F1-F2-F3 位置圖

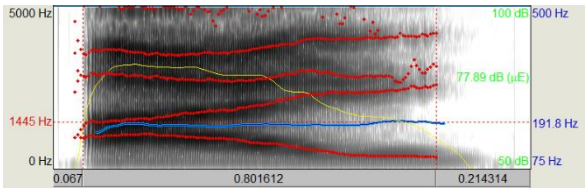

圖 6 音素 ai 的頻譜圖與聲學參數

藉韻母的 F1-F2 位置圖作為發聲的視覺形式回饋，舉例來說，F1-F2 圖中極端的 6 個韻母可圍成一個封閉區域稱為韻母空間(vowel space)，此空間的大小與形狀可作為評估韻母發聲是否清晰的指標。舉音素 ai 例，頻譜圖如圖 6，其中包含聲學參數 4 個共振峰 F1~F4、能量與基頻(F0)的軌跡圖。若將音素 ai 發聲過程的 F1 與 F2 變化描繪於 F1-F2 平面，可以形

成<F1(t), F2(t)>座標軌跡與韻母空間的對應關係，並據此與發聲狀態變化相關聯，如圖 7。

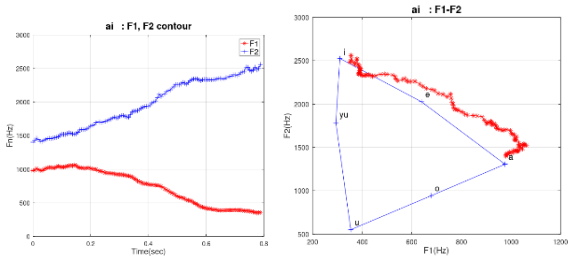

圖 7 音素 ai 的共振峰 F1, F2 變化

聲學韻母空間所圍成的區域可作為辨別發聲狀態的視覺化工具，以 7 組基本韻母ㄧ、ㄨ、ㄝ、ㄚ、ㄛ、ㄩ、ㄜ為例，包括 6 組韻母空間的轉角韻母(corner vowels)，以及一組中央韻母(central vowel)ㄜ，可參考區域面積的大小及形狀差異，執行自我發聲訓練。雙韻母(diphthongs)、複合韻母(compound vowels)、鼻音(nasal vowel)以及滑音(vowels with glide)均可依據韻母 F1, F2 座標變化的位置與樣態，分辨發聲狀態是否需要調整，以及如何進行調整。韻母練習的操作程序說明如表 1(Patil and Shah, 2015; Dudy et al., 2018)。

項目	操作方式
單韻母	作為口型與舌位定位訓練。分為舌尖音ㄧ、ㄨ、ㄝ及舌面音ㄚ、ㄛ、ㄩ、ㄜ。
複韻母	發聲過程需要變換嘴型及舌位，藉兩個基本音素正向、反向交互發音。如ㄞ：ㄚ+ㄧ/ㄧ+ㄚ; ㄟ：ㄝ+ㄧ/ㄧ+ㄝ；ㄠ：ㄚ+ㄨ/ㄨ+ㄚ; ㄡ：ㄛ+ㄨ/ㄨ+ㄛ。
鼻韻母	發聲過程由一個基本韻母開始，以空韻母ㄣ’(ㄋ嘴型與舌位)或ㄥ’(ㄍ嘴型與舌位)結束，嘴型及舌位同複韻母的發聲方式，然需控制咽的開合，改變氣流通道切換口腔與鼻腔作為共鳴腔，韻尾為鼻腔共鳴。如ㄢ：ㄚ+ㄜ+ㄣ’；ㄣ：ㄜ+ㄣ’；ㄤ：ㄚ+ㄥ’；ㄥ：ㄛ+ㄥ’
介音+韻母	含ㄧ、ㄨ、ㄩ三種介音，可與單、複或鼻韻母組成發音如下說明。 1. ㄧ加單韻：ㄧㄚ、ㄧㄝ、ㄧㄛ；加複韻：ㄧㄞ、ㄧㄠ、ㄧㄡ；加鼻韻：ㄧㄢ、ㄧㄥ、ㄧㄤ。 2. ㄨ加單韻：ㄨㄚ、ㄨㄛ；加複韻：ㄨㄞ、ㄨㄟ；加鼻韻：ㄨㄢ、ㄨㄣ、ㄨㄤ、ㄨㄥ。 3. ㄩ加單韻：ㄩㄝ；加鼻韻：ㄩㄢ、ㄩㄣ、ㄩㄥ。

表 1 各組韻母練習操作方式

各類韻母發聲練習，可藉多模訊息回饋，包括聽覺、視覺以及觸覺，進行發聲狀態的調整，如表 2。

項目	多模回饋發聲練習
單韻母	聽覺：依據聲譜分析結果轉化為 F1、F2 共振峰位置的視覺資訊，並以語者的聲學空間確認韻母在空間中的相對位置。 視覺：口腔形狀可以藉由即時影像的回饋及比對，反覆訓練形成發音習慣。 觸覺：藉舌頭觸感確認發音正確，藉旋律感、收舌跟貼下齒與舌用力，建立 7 個基本音的舌頭觸感及習慣。
複韻母	聽覺：依據聲譜分析結果轉化為 F1-F2 動態圖曲線走勢的視覺資訊。 視覺：口腔形狀藉由嘴型到位，反覆訓練會形成肌肉記憶，養成發音習慣。 觸覺：藉舌頭觸覺感覺回饋構音位置是否正確。
複韻母	聽覺：依據聲譜分析結果轉化為聲音頻譜分析共振峰(F1, F2)的動態走勢及曲線圖的視覺資訊。 視覺：空韻母ㄥ’的下顎位置較低，空韻母ㄣ’的下顎位置較高。 觸覺；藉由雙手方法，一個手觸碰頭頂或胸前；另一個手接近鼻腔位置，發音時候由口腔至鼻腔感知氣流位置變換。
介音+韻母	聽覺：依據聲譜分析結果轉化為聲音頻譜分析共振峰(F1, F2)的動態走勢及曲線圖的視覺資訊。 視覺：臉頰肌肉控制嘴型的動態變化。 觸覺：藉舌頭觸感確認構音位置。

表 2 韻母多模回饋發聲練習

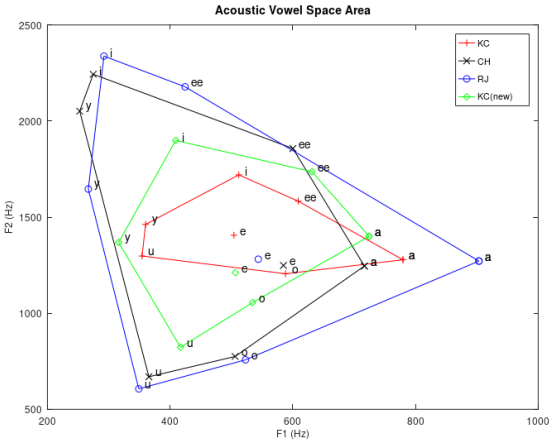

圖 8 聲學韻母空間(曾坤川，2021)

圖 8 為韻母空間之比較，黑色線為正常語者 CH，藍色線為正常語者 RJ，紅色線為聽障語者接受訓練前的韻母空間，綠色線為聽障語

者接受訓練後的韻母空間，圖中可見韻母空間形狀由扁平轉變為往四周擴張，面積也有顯著的增加(Crap, 2019)。

3.3 聲母模型

將圖 3 音節狀態結構圖中 Initial 模型修正如圖 9，在聲母前加上靜音(Sil)狀態，可使 Initial 模型更準確的代表塞音與塞擦音特性，圖 10 至圖 16 為各類聲母+ㄚ(或ㄧㄚ)的聲波信號與頻譜，及共振峰等參數的估測情形(Fang et al,. 2019)。

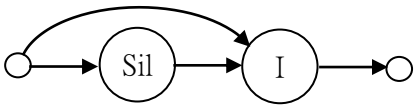

圖 9 Initial 模型

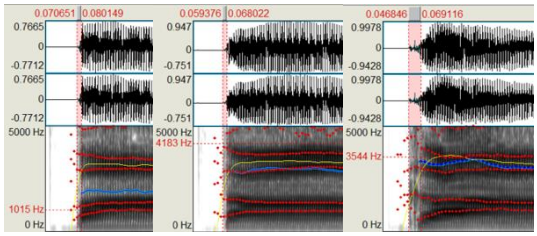

圖 10 塞音-不送氣(左)ㄅ (中)ㄉ (右)ㄍ

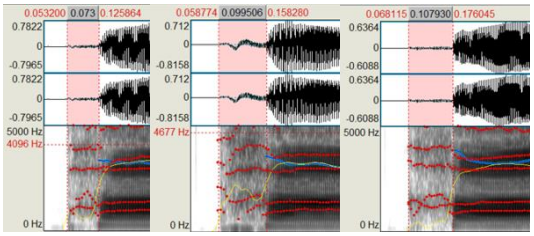

圖 11 塞音-送氣(左)ㄆ(中)ㄊ(右)ㄎ

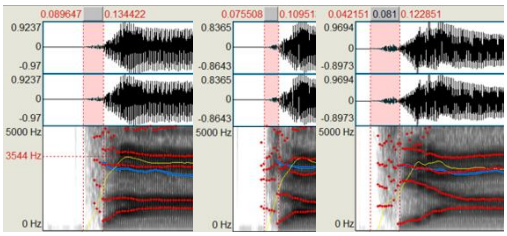

圖 12 塞擦音-不送氣 (左)ㄗ(中)ㄓ(右)ㄐㄧ

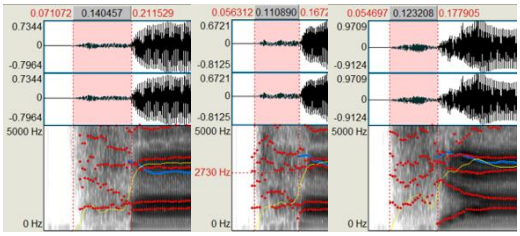

圖 13 塞擦音-送氣(左)ㄘ(中)ㄔ(右)ㄑㄧ

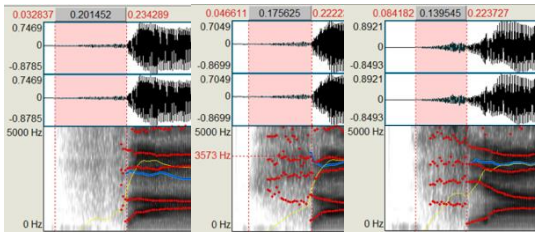

圖 14 擦音(左)ㄙ (中)ㄕ (右)ㄒㄧ

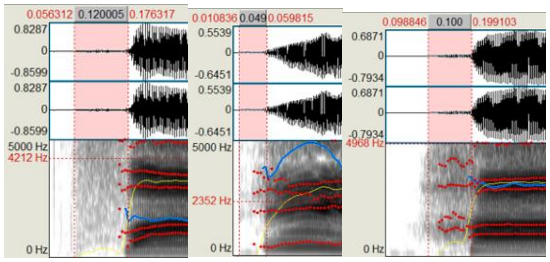

圖 15 擦音 (左)ㄈ (中) ㄖˋ(濁) (右)ㄏ

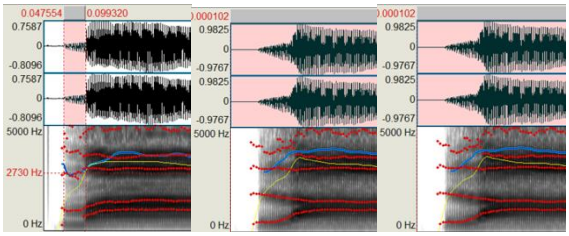

圖 16 濁音(右) ㄇ (左) ㄋ (右)ㄌ

聲母發聲練習操作方式及多模回饋發聲練習，說明如下：

項目	操作方式
成阻位置	6 種成阻位置一種使用雙唇，另五種使用不同舌位分別與齒、齦、顎形成阻斷，以塞音或塞擦音方式除阻，並於除阻後送氣或不送氣。操作方式如下： **不送氣**：ㄅㄚ、ㄉㄚ、ㄍㄚ為塞音，ㄗㄚ、ㄓㄚ、ㄐㄧㄚ為塞擦音。 **送氣**：ㄆㄚ、ㄊㄚ、ㄎㄚ為塞音，ㄘㄚ、ㄔㄚ、ㄑㄧㄚ為塞擦音。
發聲起始時間	塞擦音(不送氣)、塞擦音(送氣)以及擦音於除阻後會有不同強度及時長的氣流通過狹窄通道產生摩擦音。操作方式如下： **塞擦音(不送氣)**：ㄗㄚ、ㄘㄚ、ㄙㄚ **塞擦音(送氣)**：ㄓㄚ、ㄔㄚ、ㄕㄚ **擦音**：ㄐㄧㄚ、ㄑㄧㄚ、ㄒㄧㄚ
清音音源位置	清音音源位置由外而內，音源後聲道長度由短而長，F1 逐漸降低。操作方式如下： ㄈㄚ、ㄙㄚ、ㄕㄚ、ㄒㄧㄚ、ㄏㄚ
送氣/不送氣	3 組塞音與 3 組塞擦音共 6 組，分別以除阻後送氣與不送氣對照發聲。操作方式如下： 雙唇：ㄅㄚ/ㄆㄚ

	舌尖前：ㄅㄚ/ㄊㄚ、 舌根：ㄍㄚ/ㄎㄚ 舌尖中：ㄗㄚ/ㄘㄚ 舌尖後(捲舌)：ㄓㄚ/ㄔㄚ 舌面：ㄐㄧㄚ/ㄑㄧㄚ
濁音(鼻音/邊音/擦音)	舌位擺放好後，藉聲帶振動產生聲音，並搭配ㄚ韻母發聲。操作方式如下： ㄇㄚ、ㄋㄚ、ㄌㄚ、ㄖㄚ
靜默時長	將聲母+ㄚ發聲練習的音節前加上韻母ㄚ，可進行靜默時長(持阻期)的測量。操作方式如下： ㄚ+聲母+ㄚ

表 3 聲母發聲練習操作方式

項目	多模回饋發聲練習
成阻位置	聽覺：聲音波形與頻譜分析除阻時點、VOT 頻譜分布及共振峰轉折。 視覺：感測並顯示口腔出氣的情形。 觸覺：口腔出氣強度。
發聲起始時間	聽覺：聲音波形及頻譜分析 VOT 長度及除阻瞬間能量。 觸覺：口腔出氣強度。
清音音源位置	聽覺：聲音頻譜分析高頻能量分布情形。 視覺：嘴型變化。
送氣/不送氣	聽覺：聲音頻譜分析 VOT 除阻瞬間能量及接續高頻雜訊持續情形。 視覺：嘴型變化過程及時序。 觸覺：口腔出氣強度。
濁音(鼻音/邊音/擦音)	聽覺：聲音頻譜分析基頻變化以及共振峰分佈。 視覺：嘴型變化。 觸覺：舌位變化、頭頂鼻音產生的共振。
靜默時長	聽覺：聲音波形及頻譜分持阻期的時間長度。 視覺：嘴型變化過程及時序。

表 4 聲母多模回饋發聲練習

單位：ms			全部聲母		塞/塞擦音		擦音		滑音邊音	
			正常	聽障	正常	聽障	正常	聽障	正常	聽障
CV	VOT	平均值	112.8	283.4	80.8	179	189.1	532.9	88.4	202
		標準差	20.1	37.9	12.4	31.9	31.8	49.6	27.3	38.5
VCV	Sil	平均值	67.2	223.6	117.7	233.8	0	248.8	0	132
		標準差	16.3	79.7	28.5	79.3	0	90.7	0	59.5
	VOT	平均值	144.6	208.6	90.1	143.4	238	367.2	175.7	153
		標準差	19.2	67.6	13.1	48.6	30	117.1	22	44.7
	Sil+VOT	平均值	211.8	432.2	207.8	377.2	238	615.9	175.7	285
		標準差	28.8	97.6	29.9	76	30	155.9	22	67.1

表 5 聲母參數平均值分析(曾坤川，2021)

單音節發音多為聲母加韻母形式，成阻所需時長無法由單一個單音節發聲測得，故須藉由 VCV 的發聲組合，如ㄚ+ㄅ+ㄚ，分析聲母發聲的長短，進行時間控制的訓練。比較正常語者與聽障語者的聲母發聲情形，分析 CV 與 VCV 的 VOT 與發聲前的靜音，如表 5。其中，聽障語者發聲平均長度是正常語者的 2 倍，發擦音前正常語者無靜音區間，而聽障語者卻有顯著的靜音。聽障語者的 VOT 時間明顯較長，變異量亦較顯著，需要強化發聲器官的時間控制，進行發聲流暢度的改善。

3.4 聲調模型

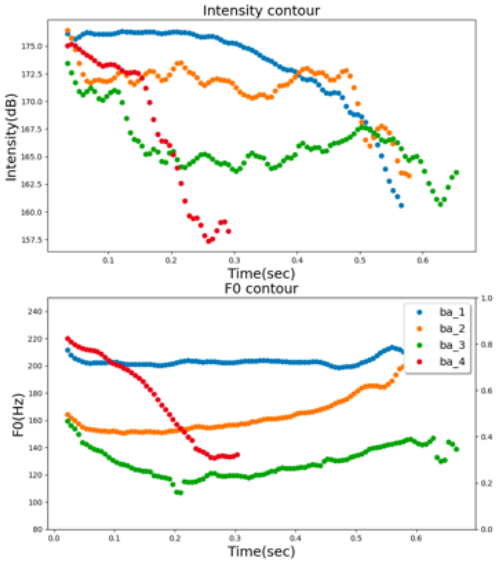

圖 17 正常語者發 ba 的四組聲調

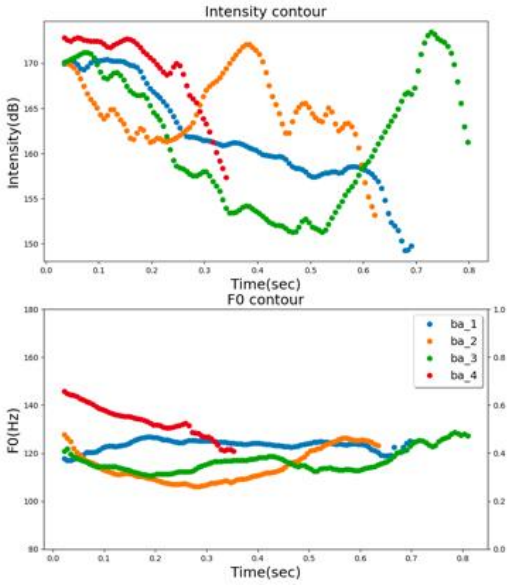

圖 18 聽障語者發 ba 的四組聲調

比較聽障語者與正常語者發ㄅㄚ的四聲，其聲調及強度變化如圖 17 與圖 18(曾坤川，2021)，正常語者聲調的起伏明確，高低差異清晰，經訓練後聽障語者聲調起伏有顯著的改善，然對聲調高低的掌握能力仍待加強。

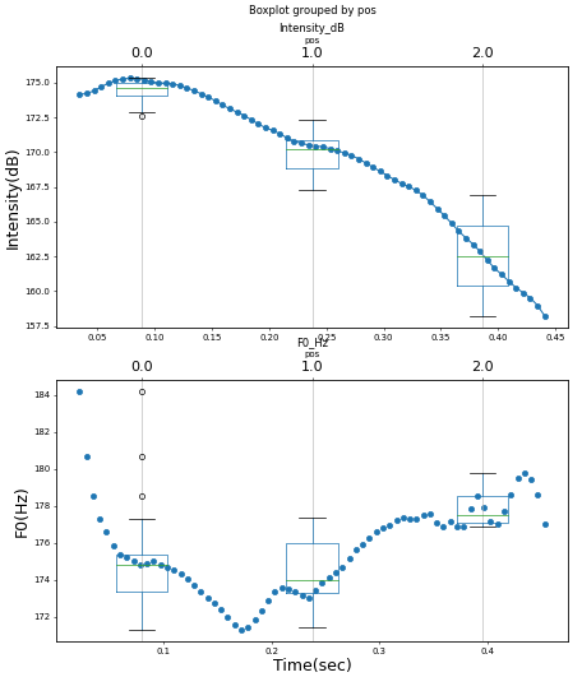

圖 19 基頻 F0(B)-F0(M)-F0(E)模型

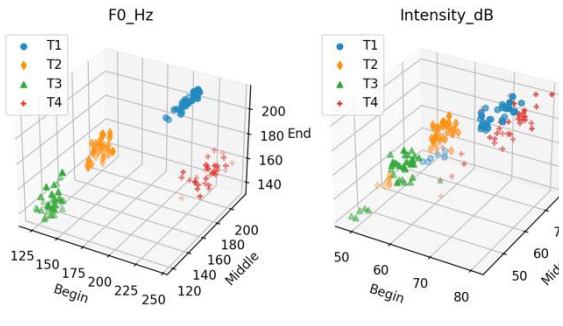

圖 20 正常語者的三段式模型分布圖

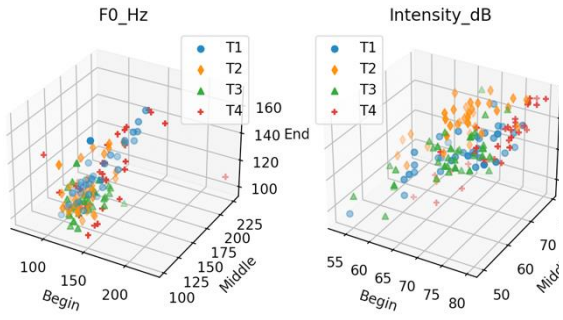

圖 21 聽障語者的三段式模型分布圖

如圖 19，將圖 3 中聲調模型 T，修正為前中後(Begin-Middle-End)三段式模型，分別定義為 T(B)、T(M)與 T(E)。每段模型包括基頻(F0)與強度(I)，基頻的三段參數模型為 F0(B)、F0(M)與 F0(E)，強度為 I(B)、I(M)與 I(E)。各段頻率與強度參數的分布組成聲調參數模型，每一組聲調以<F0(B), F0(M), F0(E)>及<I(B), I(M), I(E)>，標定於三維空間，如圖 20 與圖 21。正常語者與聽障語者分別發 6 組音節ㄅㄚ、ㄆㄚ、ㄇㄚ、ㄊㄚ、ㄍㄚ、ㄎㄚ的 4 種聲調，並進行參數標定與分析。將各組音節聲調對應至 F0(B)-F0(M)-F0(E)空間，可明顯看出正常語者的基頻與強度的離散分布均顯著；而聽障語者 4 組聲調的分布，其中聲調高低差異不顯著，但可見前中後段的參數變化，四組聲調的相對位置仍呈現較為明顯的差異。因此，聲調表現以前中後三段模型轉換為視覺形式呈現，可作為調整與改善聲調發聲的參考。

4 結論

本研究提出將國語發音以視覺形式呈現，包括韻母的韻母聲學空間對照圖、聲母的 Sil+VOT 狀態圖以及聲調的 Begin-Middle-End 模型，讓聽覺障礙的語者，可跟據發聲原理及圖形化的回饋，進行自我的發聲練習，未來可加入視覺與觸覺回饋，應可達成更有效的自我發聲練習效果。

致謝

國科會計畫 109-2637-E-268 -001 及 111-2637-E-262 -003 補助。

參考文獻

李芃娟，1999，聽覺障礙兒童國語塞擦音聲學特質分析研究，特殊教育與復健學報 7，頁 79-112。

林珮瑜、何恬、李芳宜、林香均、李沛群、蔡昆憲(譯)，2006，言語科學－理論與臨床應用，心理出版社。(Carole T. Ferrand, 2006)

張小芬、古鴻炎、吳俊欣，2004，聽障學生國語語詞聲調人耳評分與電腦分析之初探，特殊教育研究學刊，26 期，p.221~245。

張蓓莉，2000，聽覺障礙學生說話清晰度知覺分析研究，特殊教育研究學刊，18 期，53-78 頁，民 89。

曾坤川，2021，聽語障人士電腦輔助口語訓練之研究，碩士論文，和春技術學院電機系碩士班。

鄭靜宜，2011，語音聲學：說話的科學，心理出版社。

賴俞靜、劉惠美，2014，電腦輔助教學系統對提高國中聽覺障礙學生聽辨能力及語詞清晰度之成效，2014 年兩岸溝通障礙學術研討會，pp 21-33。

AssistiveWare. 2023. What is AAC? (https://www.assistiveware.com/learn-aac/what-is-aac).

A. Coucke, A Saade, A Ball, T Bluche, A.Caulier, D. Leroy, C. Doumouro, T. Gisselbrecht, F. Caltagirone, T. Lavril, M. Primet, & J. Dureau. 2018. Snips Voice Platform: An embedded Spoken Language Understanding system for private-by-design voice interfaces. ArXiv. /abs/1805.10190.

L´szló Czap. 2019. *Automated Speech Production Assessment of Hard of Hearing Children*. IEEE Journal of Selected Topics in Signal Processing (Early Access), 2019.

Shiran Dudy, Steven Bedrick, Meysam Asgari, Alexander Kain. 2018. *Automatic analysis of pronunciations for children with speech sound disorders*. CSL, 2018, pp 62-84.

S.-H. Fang, C.-T. Wang, J.-Y. Chen, Y. Tsao and F.-C. Lin. 2019. *Combining Acoustic Signals and Medical Records to Improve Pathological Voice Classification*. APSIPA, 2019, pp 1-11.

Rahul Jain, K. S. Nataraj, and Prem C. Pandey. 2016. *Dynamic Display of Vocal Tract Shape for Speech Training*. 22nd NCC, pp. 1-6.

D. Jiang, W. Zou, S. Zhao, G. Yang and X. Li. 2018. *An Analysis of Decoding for Attention-Based End-to-End Mandarin Speech Recognition*. 11th ISCSLP, 2018, pp. 384-388.

S. J. Lee, B. O. Kang, H. Chung and J. G. Park. 2015. *A useful feature-engineering approach for a LVCSR system based on CD-DNN-HMM algorithm*. 23rd EUSIPCO, 2015, pp. 1421-1425.

Juri Lukkarila. 2017. *Developing a Conversation Assistant for the Hearing Impaired Using Automatic Speech Recognition*. Master Thesis, Aalto University, 2017.

J. McKechnie, B. Ahmed, R. Gutierrez-Osuna, P. Monroe, P. McCabe & K. J. Ballard. 2018. *Automated speech analysis tools for children's speech production: A systematic literature review*. International Journal of SLP, 2018, pp 583-593.

A. S. Patil and M. S. Shah. 2015. *Comparison of vocal tract shape estimation techniques based on formant frequencies, autocorrelation, covariance and lattice*. ICNTE, 2015, pp. 1-6.

Z. Qu, P. Haghani, E. Weinstein and P. Moreno. 2017. *Syllable-based acoustic modeling with CTC-SMBR-LSTM*. IEEE ASRU, 2017, pp. 173-177.

A. Thida, N.N. Han, S.T. Oo, S Li, and C. Ding. 2020. *VOIS: The First Speech Therapy App Specifically Designed for Myanmar Hearing-Impaired Children*, O-COCOSDA, 2020, pp. 151-154.

Hana Triskova. 2011. *The Structure of the Mandarin Syllable: Why, When and How to Teach it*. ORIENTAL ARCHIVE Vol. 79:1, 2011, pp. 99-134.

G.P. Usha and J.S.R Alex. 2023. *Speech assessment tool methods for speech impaired children: a systematic literature review on the state-of-the-art in Speech impairment analysis.* Multimed Tools Appl. https://doi.org/10.1007/s11042-023-14913-0

A. Virkkunen, J. Lukkarila, K. Palomki, andM. Kurimo. 2019. *A user study to compare two conversational assistants designed for people with hearing impairments*, 8th SLPAT, 2019, pp. 1-8.

H. K. Vorperian and R. D. Kent. 2007. *Vowel acoustic space development in children: a synthesis of acoustic and anatomic data*. JSLHR, 50(6), 2007, pp. 1510–1545.

P. Wu and M. Wang, 2020. *Large Vocabulary Continuous Speech Recognition with Deep Recurrent Network*. ICSIP, 2020, pp. 794-798.

H. Yan, Q. He and W. Xie, 2020. *Crnn-Ctc Based Mandarin Keywords Spotting*. ICASSP 2020, pp. 7489-7493.