

行政院國家科學委員會專題研究計畫 成果報告

適合聽障者國語構音及韻律矯正之訓練輔具研究(3/3)

計畫類別：個別型計畫

計畫編號：NSC93-2213-E-009-123-

執行期間：93年08月01日至94年07月31日

執行單位：國立交通大學電信工程學系(所)

計畫主持人：張文輝

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 94 年 10 月 31 日

行政院國家科學委員會專題研究計畫報告

適合聽障者國語構音及韻律矯正之訓練輔具研究(3/3)

Speech Training Aids for Hearing-Impaired Mandarin Speakers

計畫編號：NSC 93-2213-E-009-018

執行期限：93年8月1日至94年7月31日

主持人：張文輝 交通大學電信工程系 教授

共同主持人：江源泉 新竹師範學院特殊教育系 副教授

一、中文摘要

(關鍵詞：聽障口語，構音轉換，韻律轉換，諧波正弦轉換，訓練輔具。)

本計畫為三年期的研究，結合正弦語音分析與語言障礙察覺分析的相關知識，開發一種適用於聽障中文口語練習的電腦輔具。其中整合構音及韻律調整兩項技術，用以改善聽障語者在音段及超音段層次上的發音偏差。相關的調整技術將架構於正弦語音分析模型，先解析語音訊號中聲門激發訊號以及聲道系統的振幅和相位，再調整聲學上的相關特徵參數，以達成音長及音高矯正之目的。構音轉換依據高斯混和模型設計最佳轉換函數，藉由轉換函數對應出具聲道特性的特徵參數。音長調整模型是運用動態時間調整演算法的結果，依適當比例調整正弦模型之參數。而音高調整則以單音節為單位，針對其音高曲線參數進行最佳化向量碼書對應轉換。整合構音與韻律轉換開發具備視覺回饋線索且適合聽障語者的發聲學習輔具。

英文摘要

(Keywords: hearing-impaired speech, spectral conversion, prosodic conversion, harmonic sinusoidal model, training aids.)

This research studies the combined use of spectral and prosodic conversions to enhance the hearing-impaired Mandarin speech. The analysis-synthesis system is based on a sinusoidal representation of the speech production mechanism. By taking advantage of the tone structure in Mandarin speech, pitch contours are orthogonally transformed and applied within the sinusoidal framework to perform pitch modification. Also proposed is a time-scale modification algorithm that finds accurate alignments between hearing-impaired and normal utterances. Using the alignments, spectral conversion is performed on subsyl-

labic acoustic units by a continuous probabilistic transform based on a Gaussian mixture model. Results of perceptual evaluation indicate that the proposed system greatly improves the intelligibility and the naturalness of hearing-impaired Mandarin speech.

二、計劃緣由與目的

聽障口語品質不佳源於語言發展的過程缺乏聽覺回饋隨時監聽並修正自己的發聲，以符合常人口語在個人聽覺上所建立的標準。一般聽障者的口語典型特徵，可分音段與超音段兩部分的異常[1,2]，前者包括母音與子音的替代與歪曲式的構音錯誤，後者則包括不正確的節律、缺乏變化的音調、較慢的說話速度以及過重的鼻音。經由整合過去之研究成果開發一種適合於聽障者使用的中文口語訓練輔具，其構想是利用語音轉換技術補償聽障者的異常口語，再針對轉換前後的語音分別進行聲學比對，並將其結果顯示於螢幕上作為受訓者視覺的學習線索達到口語練習之個人化與便利化的運用。

三、研究方法與結果

本研究應用語音分析、合成與轉換的技術，再配合聽障口語聲學分析的相關知識，順利開發出中文口語的構音與韻律調整機制，有效地改善聽障口語的理解清晰度。系統方塊圖如圖一所示，針對研究方法之說明如下：

(1) 諧波正弦模型

諧波正弦模型分析主要是利用多項正弦波元合成語音，與原音比較以決定其頻率、相位與振幅等參數的最佳組合。在語音分析端，考量聽障語者口語障礙的成因，除了造成構音偏差的聲道共振腔控制外，還有在超音段訊息的缺失是源自於不協調的聲門控制，所以將進行發聲結構的聲門激發源與聲道共振腔之解析[3]。定義第 m 個音框弦

波訊號表示為

$$s(n) = \sum_{k=1}^{L(m)} a_k(m) M_k(m) \cos[\Omega_k(m) + \Phi_k(m)]$$

其中 $L(m)$ 為弦波個數， $a_k(m)$ 和 $\Omega_k(m)$ 代表聲門激發源的振幅及相位， $M_k(m)$ 和 $\Phi_k(m)$ 則分別模擬聲道系統之振幅與相位響應。進一步將振幅響應 $M_k(m)$ 轉換為倒頻譜係數 $\{c_i(m)\}_{i=0}^{24}$ 作為頻譜轉換之特徵參數。而在合成端利用最小相位模型以轉換後的倒頻譜係數 $\{c'_i(m)\}_{i=0}^{24}$ 估計聲門激發源的振幅及相位 $\hat{M}(w; m)$ 和 $\hat{\Phi}(w; m)$ 。

計算激發源相位 $\Omega_k(m)$ ，需預估聲音訊號的基頻週期 $P(m)$ 大小，再利用鄰近音框的激發啟始時間以及基頻週期的累加找出在長度為 $Q(m)$ 的音框內，與鄰近音框激發啟始時間有一致相位之時間點，調整 J_m 找出最靠近音框中央點之處，即為此音框之激發啟始時間

$$n_0(m) = n_0(m-1) + J_m P(m),$$

進而可求得激發源相位

$$\Omega_k(m) = -[mQ(m) - n_0(m)]w_k(m),$$

其中 $w_k(m) = kw_0(m)$ 為第 k 個正弦波成分之頻率大小， $w_0(m)$ 是聲音訊號之基頻。在合成處理所需的聲道系統的振幅頻率響應與相位頻率響應，即以 $w_k(m)$ 為對頻率軸取樣可求出所需的系統振幅 $\hat{M}_k(m) = \hat{M}(w_k; m)$ 以及系統相位 $\hat{\Phi}_k(m) = \hat{\Phi}(w_k; m)$ 。而韻律轉換處理即在調整音框的基頻與合成音框大小，並重新估算激發啟始時間以匹配鄰近音框激發相位變化的平滑性，重新計算每個諧波頻率所對應的系統振幅大小，再利用調整過後之參數以諧波合成的方法產生新的聲音訊號。

(2) 頻譜轉換機制

頻譜轉換的作用是利用轉換函數調變具有語者構音特性的倒頻譜係數，同時考慮不同語者的發音速度的差異，對於使用者（聽障語者）和參考語者（正常語者）語料庫，藉由動態時間校準求得具相同時序長度的正規化參數向量序列 $\mathbf{X} = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)\}$ 和 $\mathbf{Y} = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(T)\}$ ，分別表示為使用者和參考語者的倒頻譜係數序列，用以作為模型參數訓練之用。而轉換函數 F 的設計是基

於高斯混合模型的統計架構進行最佳均方誤差估算 $\hat{\mathbf{x}} = F(\mathbf{x})$ [4]，

$$F(\mathbf{x}) = \sum_{i=1}^I P(B_i | \mathbf{x}) \cdot \left[\mathbf{m}_y^{(i)} + \mathbf{C}_{yx}^{(i)} \cdot \mathbf{C}_x^{(i)-1} \cdot (\mathbf{x} - \mathbf{m}_x^{(i)}) \right]$$

其中令 I 為高斯混合數目且 $\mathbf{C}_{yx}^{(i)}$ 為 \mathbf{x} 與 \mathbf{y} 的互分散矩陣， $\mathbf{m}_x^{(i)}$ 、 $\mathbf{m}_y^{(i)}$ 、 $\mathbf{C}_x^{(i)}$ 、 $P(B_i | \mathbf{x})$ 分別代表第 i 個高斯成分的均值向量、共分散矩陣及出現機率。而函數的訓練則利用 Expectation-Maximization 疊代演算法算出所需的轉換參數值。實際上運用頻譜轉換技術於構音矯正時，需考量中文單音節聲母與韻母在時間延遲比例相差甚鉅導致訓練語料資訊不平衡，為防止訓練語料不平衡的影響，將兩者分置於不同的高斯混合模型個別進行參數訓練。

(3) 韻律轉換機制

韻律調整處理包含了音長調整與音高調整兩部分，音長比例調整旨在改變語者的說話速度而不影響原語者發聲的特性調，依調變合成音框長度為 $Q(m)' = \rho(m)Q(m)$ 。音高調變技術基於諧波正弦分析模型，調變聲門激發訊號的基頻大小 $w'_0(m) = 1/P'(m)$ ，進一步修正激發啟始時間以及新的激發源相位。同時，為了保持原語者發聲的特性，即聲道系統之頻率響應維持不變的條件下，重新計算相對應各個正弦波成分之振幅與相位響應。

對於音長調整之準則目的在於修正原始語音與參考語者發聲的節律一致，在考量中文發音為聲母-韻母的結構，以及不同音類間其音長之標準差異甚鉅。調整比例的決定先利用邊界偵測分析標記音節中聲母和韻母的位置。聲母部分的調整藉由與參考語者之間的延遲長度比對，以線性比例進行調變；韻母部分則利用動態時間校準法則 [5] 建立其與參考語者間最佳的調整準則。據此可讓調變語音的長度與參考語音一致，並在維持音框之間訊號的連續性的同時確保其自然流暢的音質。

中文四聲聲調的變化與音高曲線的高低起伏的軌跡一致，所以建置音高曲線調變法則，用以決定每一個音框所屬基頻調變的準則。藉由語音分析可量測每個音節中基頻變化的紀錄 $\{w_0(0), w_0(1), \dots, w_0(T)\}$ 。考量音高曲線長度不一致以及要能完整捕捉音高變化的特性，利用一組在長度上作正規化

的正交 Legendre 多項式 $\Psi_j(\cdot)$ 作為基底函數展延基頻變化軌跡[6]為維度為四的特性向量 $\mathbf{a} = (a_0, a_1, a_2, a_3)^T$ ，其中 a_0 表示音高變化之平均值，其餘參數表示變化的斜率與曲率之特性。進一步，設計基於向量量化的參數對應模型，先收集聽障語者與正常語者足量的音高變化參數，並個別地訓練其所屬的向量量化碼書 $\mathbf{C}^{(s)}$ 與 $\mathbf{C}^{(t)}$ 。接著統計碼書之間所有碼字相對應的狀況，據此設計音高特性參數的對應碼書 $\mathbf{C} = \{\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{N_c-1}\}$ 。得音高曲線轉換的程序：由語音分析得此音節之音高曲線參數 \mathbf{a} 後，接著與碼書 $\mathbf{C}^{(s)}$ 比對找到最近似的量化碼字其索引值為 i ，則可得到最佳的特性轉換參數 $\hat{\mathbf{a}} = \mathbf{c}_i$ ，再以多項式 $\Psi_j(\cdot)$ 合成調變後的音高曲線 $\{w'_0(0), w'_0(1), \dots, w'_0(T)\}$ 。

四、聽障輔具製作

在聽障口語教學的相關研究，曾嘗試以人工修正並評量其對整體口語清晰度的相對影響，其分析結果有助於聽障口語矯正技術之研究。而國內目前聽障口語的臨床治療與教學對於超音段的國語四聲一般藉由手部動作高低進行訓練，這對於缺乏音調高低概念的聽障者而言，訓練效果十分有限。因此，整合語音轉換各項的技術開發構音及韻律矯正的視覺回饋系統，將可以提供聽障語者個人化的口語學習輔具。

此系統的設計依據功能區分包含了三個單元，分別為語音轉換模型的訓練、構音學習以及韻律學習。而學習的語料內容提供了擦音、塞音以及塞擦音三大音類搭配不同韻母之單音節選單，使用者在練習初始階段可依據個人學習的步調自由選擇特定的語料進行反覆的練習。在語音轉換模型的訓練過程，其主要目的是在收集並建立轉換模型參數訓練演算法中所需足量的使用者語料庫，系統會要求使用者針對所選的練習語料進行一連串的訓練語料收集，同時將提供相對應的口部發聲變化的影像作為使用者練習時所需之視覺資訊。在構音練習過程，透過頻譜轉換處理可產生一個新的修正語音，使用者依據相關提示操作，可以迅速的觀察個人原始語音和修正語音在頻譜分析圖上的差異，藉由反覆的觀察與聆聽個人與

修正訊號上的異同，達成構音學習的目標。在韻律練習過程，使用者可以搭配各個單音節進行國語四聲的語調練習，透過音高轉換處理將提供語者原始的音高變化與修正後的音高變化曲線圖作為韻律學習過程必要的視覺回饋資訊，同樣地也可以重覆聆聽修正前後聲音的變化。

五、結論與討論

經由語言障礙察覺分析，可得知國語各單音節在頻譜分佈、音長及音高曲線上所存在的差異，並觀察聽障語者在發音過程所運用之構音技巧與正常語者之間的異同，提供我們更多的資訊來瞭解聽障語者發音所存在的問題，以作為構音與韻律矯正系統設計之想法。其中透過聲母和韻母的發聲延遲統計分析顯示兩者之間發聲延遲差異很大。聽障語者和正常語者之間在不同音類音長的比例也不盡相同，由觀察顯示聽障語者慣用塞音或類似塞音的發聲模式取代擦音和塞擦音，結果使得聲音的延遲變得十分短促。因此，在設計轉換系統時，將先透過分析偵測聲母-韻母的分界，在個別的進行所對應之處理。

韻律特性表現在音高曲線的變化，我們利用音高曲線調整機制，可控制修正的語音維持原語者發聲的個人性質，僅改變音高變化的軌跡，使修正的語音具有完整的音高起伏。圖二所示分別針對聽障語者、修正語音與正常語者顯示國語發聲“應用軟體”的音高曲線變化，結果顯示修正後的語音在國語三聲和四聲的變化較為完整，可以產生近似正常語者音高變化之合成語音。另外，比較整合音長調整與構音轉換的修正語音和聽障語音之間的聽力鑑別度測試，結果如圖三所示，修正語音在擦音、塞擦音與母音上提供了較為自然可辨的合成語音。

聽障口語無法經由完全的聽覺回饋隨時監聽並修正自己的發聲，提供一個多元的視覺回饋介面，給予使用者更多發聲過程的視覺資訊，用以輔助其發聲學習是本研究要達成的目標。整合此計畫歷年所完成的矯正機制，我們提供了個人化的發聲學習輔具，此輔具提供了基礎的發聲影片教學、進階的構音和韻律學習，讓使用者彈性的選擇適合自己發聲學習表單與步調。期望藉由此軟體讓使用者獲得發聲學習上實質的幫助。

