

行政院國家科學委員會專題研究計畫 期中進度報告

適合聽障者國語構音及韻律矯正之訓練輔具研究(2/3)

計畫類別：個別型計畫

計畫編號：NSC92-2218-E-009-009-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學電信工程學系

計畫主持人：張文輝

共同主持人：江源泉

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 93 年 5 月 28 日

行政院國家科學委員會專題研究計畫期中報告

適合聽障者國語構音及韻律矯正之訓練輔具研究(2/3)

Speech Training Aids for Hearing-Impaired Mandarin Speakers

計畫編號：NSC 92-2218-E-009-009

執行期限：92年8月1日至93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，韻律轉換，諧波正弦轉換，音高曲線。)

本計畫為三年期的第二年度計劃，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的韻律矯正機制。其中整合音長及音高調整兩項技術，用以改善聽障語者在韻律層次上的發音偏差。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。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。藉由聽覺評估得知，整合音高及音長的韻律轉換系統可以有效提升聽障中文口語的音質。

英文摘要

(Keywords: hearing-impaired speech, prosodic conversion, harmonic sinusoidal model, pitch contour.)

This research studies the application of pitch and time-scale modification to hearing-impaired speech enhancement. The basic strategy is a sinusoidal analysis-synthesis system that models and modifies the amplitude and phase contributions of the vocal cord excitation and vocal tract. The rate of articulation was changed by time-scale modification in accordance with the dynamic time warping. We also proposed a VQ-based approach to modeling the modification of prosodic features extracted from the pitch contour by orthogonal polynomial transform. Results of perceptual evaluation indicate that the proposed system can improve the quality and intelligibility 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無法經由完全的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。在早期已有多項研究在探討構音參數語音轉換的處理[3,4]。而本計劃將進一步探討聽障者無法準確掌握超音段韻律訊息，以致於嚴重影響其整體口語的自然流利與可理解度之因素，以及結合中文聲學上特有音調性質的相關知識。進行廣泛深入的口語聲學及韻律分析，再據以提出有效運用於補償其異常口語的具體方案。

本年度計劃的重點包括收集聽障語者與正常語者各類韻律變化之語料，以進行中文口語超音段聲學特性分析。並設計韻律矯正模型以修正聽障語者不容易掌握的音長及音高變化之現象。矯正系統的處理單元包括：正弦語音分析模型、音長調變模組、音高調變模組以及弦波合成器。而系統設計之考量必須建立在發聲語者構音特性維持不變的條件下，提供一個調變象徵音長以及音高特性的語音轉換模型。其作法是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，透過聲門激發訊號中音長與音高參數的調變，達成韻律矯正之目的。然而，由於韻律訊息屬於超音段層次的問題，不同於構音矯正以音框為處理單位，系統設計須含跨數個音框的音長與音高變化參數，方能全面的呈現韻律矯正之作用。音長調整係利用不同語者音節中，構音特徵參數進行動態時間校準(dynamic time warping)的比對，以決定聲音縮短或增

長所需的因子。而音高曲線則藉由事先訓練的向量量化對應機制，決定矯正訊號所屬新的音高曲線。透過此系統將適當地達成聽障語者韻律矯正，修正為自然流利的聲音。

三、研究方法與成果

本年度計劃的研究配合聽障口語聲學分析的相關知識，順利開發出中文口語韻律矯正系統，其中包含音長與音高比例調整兩項機制，有效地改善聽障口語的理解清晰度。針對研究方法及各項進行步驟詳細說明如下：

(1) 國語四聲音調語料庫之建立

本計劃為了超音段的聲學分析與韻律矯正系統製作與測試，已針對中文不同音類的 45 個單音節及其所屬的四種音調變化錄製了足量的語料。工作內容包括錄音題材的規劃、聽障語者素質之篩選以及錄音環境與設備的考量。語者限定其純音聽力平均值大於 70 dB 的先天性感官神經性患者，以確保所蒐集的語料在音段與超音段的特徵上，都與常人所錄語料在聽覺上有足夠的差異。經過篩選已錄製二男二女聽障語者足量的語料，為新竹師院特殊教育系學生與新竹培英國中學生男女各一。另外，亦針對四位聽力正常且國語無特殊腔調的學生，男女各半，對相同的題材進行錄音，以作為韻律轉換之參考標準。至於錄音題材的規劃，以語料量最小，而其內容符合研究目的原則下，以國語可出現在音節首之 21 個聲母：塞音(ㄅ、ㄆ、ㄇ、ㄏ、ㄎ、ㄌ)、擦音(ㄉ、ㄊ、ㄒ、ㄍ、ㄆ、ㄑ、ㄒ、ㄓ、ㄔ、ㄗ、ㄘ、ㄙ)、塞擦音(ㄐ、ㄑ、ㄒ、ㄓ、ㄔ、ㄗ、ㄘ、ㄙ)、鼻音(ㄇ、ㄋ)及流音(ㄌ)與國語的三個單韻母(ㄚ、ㄨ、ㄩ)組合成國語音韻許可的單音節與其在四聲聲調變化。錄音時，以朗讀方式唸讀，每音節各唸十次，錄音環境噪音量不超過 50 dBA，麥克風與說話者口部的距離則為 10 公分。

(2) 諧波正弦模型製作

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[5]，定義第 m 個音框弦

波訊號表示為

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

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。

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

$$n_0(m) = n_0(m-1) + J_m P(m), \quad (2)$$

進而可求得激發源相位

$$\Omega_k(m) = -[mQ(m) - n_0(m)] w_k(m), \quad (3)$$

其中 $w_k(m) = k w_0(m)$ 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。另外，透過對訊號分析可得到聲道系統的振幅頻率響應 $\hat{M}(w; m)$ 與相位頻率響應 $\hat{\Phi}(w; m)$ ，而再以 $w_k(m)$ 為大小對頻率軸取樣可求出所需的系統振幅 $M_k(m) = \hat{M}(w_k; m)$ 以及系統相位 $\Phi_k(m) = \hat{\Phi}(w_k; m)$ 。而韻律轉換系統即透過激發源與聲道系統參數的調變，達成所預期韻律矯正的目的。

(3) 音長比例調整

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，其原理是在以音框大小 $Q(m)$ 之弦波分析模型中，依比例改變合成音框長度為 $Q'(m) = \rho(m) Q(m)$ 。當 $\rho(m) > 1$ 表示聲音訊號會被延長，而 $\rho(m) < 1$ 則使得訊號被縮短。所以透過 $Q(m)$ 長度訊號分析之語音參數，我們將調整合成音框的長度為 $Q'(m)$ ，在維持聲門系統參數 $M_k(m)$ 、 $\Phi_k(m)$ 和 $w_0(m)$ 不變的情況下，找出新的激發啟始時間 $n'_0(m)$ 以及激發源訊號調變後新的相位 $\Omega'_k(m) = -[mQ'(m) - n'_0(m)] w_k(m)$ ，其中 $n'_0(m) = n'_0(m-1) + J'_m P(m)$ 。

透過上述音長調變技術，在決定了每個音框所應調整的比例後，我們即可調整至所要的效果。首先，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巨。因此，調整比例的決定先利用訊號分析標記音節中聲母和韻母的位置。然後聲母部分的調整藉由與參考語者之間的比對，以線性比例進行調變；韻母部分則利用動態時間校準法則[6]建立其與參考語者間最佳的調整比例。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。

(4) 音高曲線調整

以往的語音轉換研究鮮少觸及語者韻律特性的處理，原因在於韻律變化的特性較為複雜，且沒有令人滿意的數學模型提供進行分析。而藉由單音節聲學分析得知中文四聲聲調的變化與音高曲線的高低起伏存在著某種關連性，所以我們將運用音高曲線調變機制達成聲調修正之目的。

音高調變技術基於諧波正弦分析模型，其目的在改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 $n'_0(m) = n'_0(m-1) + J'_m P'(m)$ 以及新的激發源相位 $\Omega'_k(m) = -[mQ(m) - n'_0(m)]w'_k(m)$ 。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，則以調變後的激發源輸入聲道系統對於不同正弦波成分之振幅與相位響應為 $M'_k(m) = \hat{M}(w'_k; m)$ 與 $\Phi'_k(m) = \hat{\Phi}(w'_k; m)$ 。

藉由語音分析可以觀察每個音節發聲過程基頻變化的紀錄 $\{w_0(0), w_0(1), \dots, w_0(M)\}$ 。基於聲調變化為超音段訊息，我們設計以音節為單位觀察並抽取所屬音高變化曲線，進行音高曲線轉換機制的開發。首先，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，將利用一組在長度上作正規化的正交 Legendre 多項式 $\Psi_j(\cdot)$ 作為基底函數展延音高曲線 [7] 為 $\mathbf{a} = (a_0, a_1, a_2, a_3)^T$ ，其轉換式為

$$a_j = \frac{1}{M+1} \sum_{m=0}^M w_0(m) \cdot \Psi_j\left(\frac{m}{M}\right), \quad j = 0, 1, 2, 3. \quad (4)$$

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，我們設計一個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 $\mathbf{C}^{(t)}$ 。接著利用兩語者成對的音高參數進行碼書之間所有碼字相對應的狀況統計。最後，利用統計量作為權值線性組合 $\mathbf{C}^{(t)}$ 內每個碼書，則產生我們所要的音高參數對應碼書 $\mathbf{C} = \{\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{N_c-1}\}$ 。所以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，並輸出其索引值 i ，則可得到最佳的轉換參數 $\hat{\mathbf{a}} = \mathbf{c}_i$ ，並產生一調變後的音高曲線

$$w'_0(m) = \sum_{j=0}^3 \hat{a}_j \cdot \Psi_j\left(\frac{m}{M}\right), \quad 0 \leq m \leq M. \quad (5)$$

(5) 實驗結果與分析

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，以作為韻律矯正系統設計之想法。整合音長與音高曲線調變技術，目的是要得到更近似參考語者發聲行為超音段層次的特性，同時保證合成訊號流暢的音質。藉由實驗測試，證實此系統用於聽障語者中文語音的改善與加強有很大的作用。

音長比例調整之結果利用頻譜圖分析確認調變技術能在維持聲音訊號原有的頻譜結構下，同時達成音長比例修正的目的。以圖一所顯示三個頻譜圖分別為聽障語音、經過線性插補技術(linear interpolation)之語音以及經過音長比例調整的語音，可以清楚的瞭解音長比例調整之作用有效的保有原語音頻譜-能量分佈的結構，即調變之語音在聽覺感官上與原語者相同。另外，基於向量量化的音高曲線轉換機制利用語音分析軟體觀察其轉換語音與聽障和正常語者間，基頻變化的差異，如圖二所示為單音節'去'的四種聲調變化，經音高曲線調變的處理，由於轉換過程保持平均音高的大小所以音高的位置維持在 200Hz 附近，而其改變的形式會因為轉換而近似於正常語者之變化。此外，結合音長比例與音高曲線調整的韻律矯正處理，以五名聽力正常之研究生進行聽覺測試，評量聽障語音轉換前後對於單音節四種聲調變化可理解性的辨認度，測試結果列於表一。就整體的平均而言，轉換

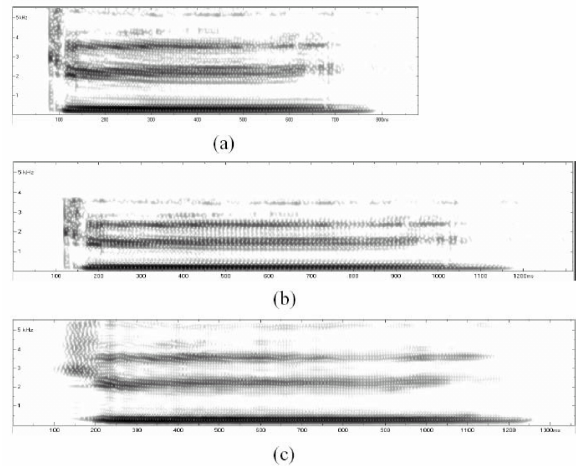
前後辨認度由 69% 增加至 86.25% ,而在較容易混淆的二聲與三聲在轉換後都有明顯的改善。

四、結論與討論

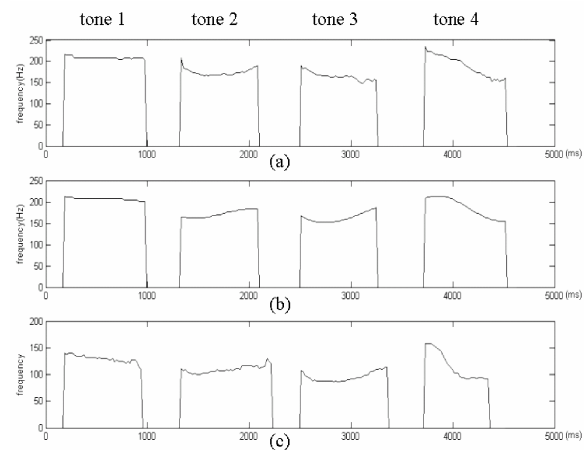
本年度計劃之具體成果為開發並製作一國語單音節的韻律矯正機制。先收集完整的語料庫以提供研究所需之參數訓練與相關測試，藉由語言障礙察覺分析軟體，進一步觀察聽障語者在發音過程所運用之發音技巧與正常語者之間的異同。這些分析資料有助於中文口語韻律的矯正研究。透過系統中對各音節進行可調適性的音長比例調整可修正聽障語者在發聲延遲上不準確的情形。而針對中文四聲語調所進行的音高曲線轉換機制可有效改正聲音音高變化的缺失。進一步整合音長與音高調變技術於矯正系統，藉由聽覺評量顯示對聲調辨認的正確性提供大幅度的改善。

五、參考文獻

- [1] Mosen, R., "Toward measuring how well hearing-impaired children speak", *Journal of Speech and Hearing Research*, 21:197-219, 1978.
- [2] Massen, B. and Provel, D., "The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech", *J. Acoust. Soc. Am.*, 78:877-886, 1985.
- [3] Abe, M., Nakamura, S., Shikano, K., and Kuwabara, H., "Voice conversion through vector quantization", *Proc. ICASSP*, vol. 1:655-658, 1988.
- [4] Stylianou, Y., Cappe, O., and Moulines, E., "Continuous probabilistic transform for voice conversion", *IEEE Trans. Speech and Audio Processing*, vol. 6:131-142, March 1998.
- [5] McAulay, R. J. and Quatieri, T. F., "Speech analysis-synthesis based on a sinusoidal representation", *IEEE Trans. Acoust., Speech Sig. Process.*, vol. 34:744-754, 1986.
- [6] Rabiner, L. and Juang, B. H., *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [7] Chen, S. H. and Wang, Y. R., "Vector quantization of pitch information in Mandarin speech", *IEEE Trans. Communications*, vol. 38,:1317-1320, 1990.



圖一：音長調變頻譜圖。(a)聽障語音,(b)運用線性插補技術的調變語音,(c)音長比例調變語音。



圖二：單音節‘去’四種聲調之音高曲線。(a)聽障語音,(b)調變後的語音,(c)參考語者的語音。

Perceived as	Stimulus			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	98	32	16	0
Tone 2	0	43	35	1
Tone 3	2	23	45	9
Tone 4	0	2	4	90

(a)

Perceived as	Stimulus			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	97	0	1	0
Tone 2	3	81	26	1
Tone 3	0	19	73	5
Tone 4	0	0	0	94

(b)

表一：四聲聲調聽覺辨認結果。(a)聽障語音,(b)轉換後語音。