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(1/3)

計畫類別：個別型計畫
計畫編號：NSC91-2614-E-009-001-
執行期間：91 年 08 月 01 日至 92 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

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 92 年 5 月 26 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(1/3)

Speech Training Aids for Hearing-Impaired Mandarin Speakers

計畫編號：NSC 91-2614-E-009-001

執行期限：91 年 8 月 1 日至 92 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

共同主持人：王德譽　朝陽科技大學資訊工程系 助理教授

一、中文摘要

(關鍵詞：聽障口語，語音轉換，諧波正弦分析。)

本計畫為三年期的第一年度計劃，結合語音轉換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構音矯正機制。其關鍵在於運用個別音節所屬音框組合間的相關性，有效降低對應訓練所需的語料量及其收斂時間。我們已建立的完整聽障語料庫，進行廣泛的構音聲學分析，利用諧波正弦模型分析所得的倒頻譜係數作要素分析，再據以設計一正弦振幅對映的高斯混合模型。針對國語 68 個單音節所作的主觀聽覺測試，顯示整體辨識率由 60.88%提升至矯正後的 78.38%，且有效地提昇聽障語者的口語清晰度。此外，我們提出一種依據動態時間校準的時變音長調整機制，經證實可改善合成音質使其符合常人口語之說話速度。

英文摘要

(Keywords: hearing-impaired speech, voice conversion, harmonic sinusoidal model)

The speech of the hearing-impaired suffers from misarticulations and prosodic deviations due to the lack of auditory feedback. This project aims to devise a voice converter that modifies the impaired speech to be perceived as if it was uttered by a normal speaker. We applied the spectral conversion on speech signals analyzed by the harmonic sinusoidal model, in which sine-wave amplitudes and phases are chosen to be harmonic samples of the magnitude and phase spectra. In light of the low-redundancy spectral representation, we perform principal-component analyses on cepstral coefficients derived from sine-wave amplitudes. After that, the principal components are characterized under the form of a Gaussian mixture model with parameters converted using an EM-trained mapping function that minimizes the spectral distance between the impaired and normal speech. Also proposed is a time-varying approach to time-scale modification that uses the dynamic time warping to match the rate of articulation of the hearing-impaired speaker with that of the normal speaker.

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，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。在聽障口語教學的相關研究[1,2]，曾深入探討造成聽障者發聲缺陷的口語特徵，嘗試以人工修正並評量其對整體口語清晰度的相對影響力，其分析結果有助於聽障口語之自動矯正研究。另一項值得深入探討的研究課題是聽障者無法準確掌握超音段韻律訊息，嚴重影響其整體口語的自然流利與可理解度。這正突顯出具聲調特質的中文口語與英語有發音層次上的差異。有鑑於此，我們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。

本年度計劃的重點是針對不同音類屬性之國語單音節進行口語聲學分析，擷取一組適合頻譜轉換的語音特徵參數，再依事先訓練的對映函數[3,4]作調整，以期能還原出近似正常口語的語音頻譜包絡線。其關鍵在於考慮個別音節所屬的音框組合間存在明顯的相關特性，若能充分運用此相關性，可以有效降低語音特徵向量的維度，進而減少訓練對映函數所需的語料量及其收斂時間。一具體可行的方案是先進行語音的諧波正弦模型分析[5]，利用弦波振幅計算其倒頻譜係數再作要素分析[6]之處理。至於聽障者說話速度之矯正，我們利用動態時間校準[7]技術比對聽障與正常兩語者的發音長度，據以設計一時變的音長比例調整機制[8]，使其合成音質更為自然流利。

三、研究方法與成果

本年度計劃的研究應用語音分析合成與轉換技術，並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矯正與音長比例調整兩項機制，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，針對研究方法及各項進行步驟詳細說明如下：

(1) 國語語料庫之建立

不同音類所屬的頻譜差異會影響語音轉換效能之優劣，因此依個別音類收集足量語料作為構音聲學分析以及轉換效能評估有其必要性。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。經過篩選已錄製一男一女聽障語者足量的語料，皆為新竹師院特殊教育系學生。另外，亦錄製二位聽力正常且國語無特殊腔調的師院學生，男女各一，對相同的題材進行錄音，以作為語音轉換之參考標準。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、ㄉ、ㄊ、ㄍ、ㄎ)、擦音(ㄈ、ㄏ、

ㄒ、ㄙ、ㄕ、ㄖ)、塞擦音(ㄐ、ㄑ、ㄓ、ㄔ、ㄗ、ㄘ)、鼻音(ㄇ、ㄋ)及流音(ㄌ)與國語的八個單韻母(ㄧ、ㄨ、ㄩ、ㄚ、ㄛ、ㄜ、ㄝ、空韻)組合成國語音韻許可的 75 個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，進行發聲結構的聲門激發源與聲道共振腔之解析，主要是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，還有在超音段韻律訊息的缺失是源自於不協調的聲門控制。定義聲門激發訊號及聲道濾波器之頻率響應為

$$e(t)=\sum_{k=1}^{K(t)} a_k(t)\cdot\cos[w_k t+\Omega_k(t)]$$

$$H(w;t)=M(w;t)\cdot\exp\left[j\cdot(w;t)\right],$$

則語音由 $K(t)$ 個正弦諧波組合如下：$s(t)=\sum_{k=1}^{K(t)} a_k(t)M(w_k;t)\cdot\cos[\cdot(w_k;t)+\Omega_k(t)+w_k t]$ 其中

$$=\sum_{k=1}^{K(t)} A_k(t)\cdot\cos[\bar{}_k(t)+w_k t]$$

$A_k(t)=a_k(t)M(w_k;t)$ 與 $\bar{}_k(t)=\cdot(w_k;t)+\Omega_k(t)$ 分別為合成訊號的振幅和相位。藉由原始語音的頻譜分析，我們可以準確估算正弦模型之各項參數，並利用弦波振幅計算該音框的倒頻譜係數 $\mathbf{c}(t)=[c_1(t),c_2(t),\ldots,c_J(t)]$。

(3) 特徵向量的要素分析

運用音框組合間的相關性，可以有效降低語音特徵向量的維度，進而減少訓練對映函數所需的語料量及其收斂時間。先搜集一特定音節所屬的音框序列 $\mathbf{C}=\{\mathbf{c}(1),\mathbf{c}(2),\ldots,\mathbf{c}(T)\}$，其中 $\mathbf{c}(t)$ 表示第 t 個音框的倒頻譜係數向量，同時計算其共分散矩陣 **Ó**和均值向量 $\mathbf{m}=[m_1,m_2,\ldots,m_J]$。進一步針對共分散矩陣作統計分析，取得 J 個固定向量(eigenvectors)，再根據所對應固定值(eigenvalue)的高低依序排列得 $\mathbf{e}_i=[e_{i1},e_{i2},\ldots,e_{iJ}]$，$i=1,2,\ldots,J$。取前 D 個固定向量用以計算倒頻譜 $\mathbf{c}(t)$ 的要素向量 $\mathbf{x}(t)=[x_1(t),x_2(t),\ldots,x_D(t)]$，其中第 i 個要素值為 $x_i(t)=\sum_{j=1}^{J} e_{ij}c_j(t)$。在反向運算部分，低維度要素向量 $\mathbf{x}(t)$ 無法還原出原始的倒頻譜，只能取得其近似向量 $\hat{\mathbf{c}}(t)=[\hat{c}_1(t),\hat{c}_2(t),\ldots,\hat{c}_J(t)]$，其中 $\hat{c}_j(t)=\sum_{i=1}^{D} e_{ij}x_i(t)+\sum_{i=D+1}^{J} e_{ij}\sum_{k=1}^{J} e_{ik}m_k$，$j=1,2,\ldots,J$。

(4) 頻譜轉換機制

不同語者的發音特性具體反映在說話速度及倒頻譜係數，因而造成聽障者與正常聽力者錄音所訓練的語音模型存在著嚴重的不匹配問題。為了克服不同說話速度之差異，我們針對聽障者與聽力正常者的錄音，分別擷取其特徵參數再作動態時間校準，得到相同時間長度的正規化參數向量序列 $\mathbf{X}=\{\mathbf{x}(1),\mathbf{x}(2),\ldots,\mathbf{x}(T)\}$ 和

$\mathbf{Y}=\{\mathbf{y}(1),\mathbf{y}(2),...,\mathbf{y}(T)\}$ 。至於語音頻譜包絡線的轉換，我們利用 expectation-maximization 演算法設計一高斯混合模型組成的最佳對映函數 F，使得經轉換後的要素向量 $\hat{\mathbf{x}}=F(\mathbf{x})$ 與 $\mathbf{y}$ 之間可產生最小的均方誤差 $E=[\|\mathbf{y}-F(\mathbf{x})\|^2]$。若令 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，則推導所得的最佳對映函數為 $F(\mathbf{x})=\sum_{i=1}^{I}P(B_i|\mathbf{x})\cdot\left[\mathbf{m}_y^{(i)}+\mathbf{C}_{yx}^{(i)}\cdot\mathbf{C}_x^{(i)^{-1}}\cdot(\mathbf{x}-\mathbf{m}_x^{(i)})\right]$ 其中 $\mathbf{m}_x^{(i)}$、$\mathbf{C}_x^{(i)}$、$P(B_i|\mathbf{x})$ 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。

(5) 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，在以音框 Q 為單位之弦波合成系統中，依比例 γ 改變合成音框長度為 $Q'=\gamma Q$。而決定每個音框所應調整的比例是利用動態時間校準法則，藉由比對兩語者所屬參數在時間軸上的對應關係，據以調整來源語者的發聲長度使其與目標語者一致。利用音長調整後所合成的訊號表示為：

$s'(n)=\sum_{k=1}^{K}A_k'\cos(w_k n'+\theta_k'),\ 1\le n\le Q'$ 。音長調整過程中不能改變聲道濾波器的參數，但需依比例延長或縮短聲門激發源訊號的長度。在聲門激發振幅與聲道振幅特性不變的情況下，合成訊號的振幅在調整前後維持不變，$A_k'=a_k'M_k'=a_kM_k=A_k$。但合成的相位隨音框調整會影響鄰近音框之間原有的連續特性，因此要修正聲門激發源的相位 Ω_k'，維持音框之間的連續性而確保其自然流暢的音質。

(6) 實驗結果與分析

透過本計劃購置的語言障礙察覺分析軟體，我們得知國語不同音類的單音節在頻譜分佈、音長、音量及音高曲線上所存在的差異，亦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。舉例而言，不同韻母間的差異源自於第一及第二共振峰頻率結構，而聽障語者的頻譜存在歪斜及不連續的偏差。此外，聽障生口語中子音的清晰度普遍低於母音，而其構音可辨識度又依其所屬音類而存在明顯的差異。聲學分析結果顯示，多數塞音的辨識清晰度頗高，流音和鼻音兩有聲子音亦有較正確的發音，但多數的塞擦音和擦音會被塞音或類似塞音的錯誤構音所取代。在頻譜轉換部分，要素分析之處理可降低語音特徵向量的維度，當維度自 19 降至 15 時，倒頻譜參數之重建失真為 1.30%；降至 10 的重建失真亦僅為 2.67%，而這樣的失真對於聲音之合成品質影響甚微。實驗結果顯示，我們所提出的構音矯正及音長比例調整兩項技術，有效地改善聽障口語的音質。整體而言，在國語 68 個單音節的主觀聽覺測試中，平均辨識率由 60.88%提升至矯正後的 78.38%。測試結果紀錄於表一，其中最明顯的改善來自於聽障生最難掌握的塞擦音，其辨識率由 30%提昇到 92.5%。表二所示為倒頻譜距離的量測結果，轉換後語音之特徵參數與正常語者之參數兩者間的距離大為縮短。

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基本音節的構音矯正機制。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構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構音及韻律的矯正研究。在頻譜轉換部分，要素分析之處理確實可降低語音特徵向量的維度，大量減少訓練對映函數所需的語料量及其收斂時間。主觀的聽覺測試證實了構音矯正及音長比例調整的重要性，聽障者 68 個單音節的辨識率由 60.88%提升至矯正後的 78.38%。在下年度的研究，將依預定之規劃先進行聽障口語的韻律分析，與正常語者比較其國語四聲所對應的音高範圍及變化曲線，再據以設計中文口語的韻律矯正機制。

五、參考文獻

[1] B. Massen and D. Provel, “The effect of correcting fundamental frequency on the intelligibility of deaf speech and its interaction with temporal aspects,” *J. Acoust. Soc. Am.*, 76, pp. 1673-1681, 1984.

[2] B. Massen and D. Provel, “The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech,” *J. Acoust. Soc. Am.*, 78, pp. 877-886, 1985.

[3] Y. Stylianou, O. Cappe, and E. Moulines, “Continuous probabilistic transform for voice conversion,” *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 131-142, March 1998.

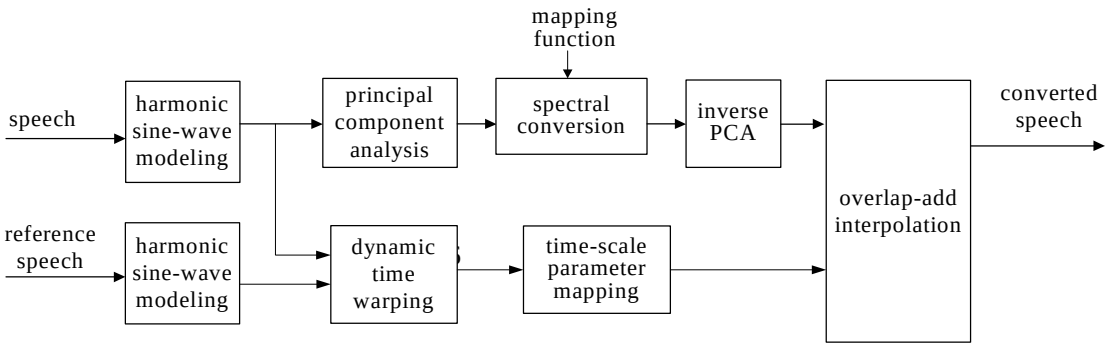
[4] C. L. Lee, W. W. Chang, Y. C. Chiang, and H. I. Hsu, “Voice conversion for enhancing the hearing-impared speech,” *International Symposium on Communication*, Tainan, Taiwan, 2001.

[5] R. J. McAulay and T. F. Quatieri, “Speech analysis-synthesis based on a sinusoidal representation,” *IEEE Trans. Signal Processing,* vol. Sp-34, pp. 744-754, Aug. 1986.

[6] I.T. Jolliffe, “Principal Component Analysis, ” *Springer-Verlag*, 1986.

[7] H. Sskoe and S. Chiba, “Dynamic programming optimization for spoken word recognition,” *IEEE Trans. Acoustic, Speech, Signal Proc.*, ASSP-26(1): 43-49, February 1978.

[8] T. F. Quatieri and R. J. McAulay, “Shape Invariant Time-scale and Pitch Modification of speech,” *IEEE Trans. Signal Processing*, vol. 40, pp.497-510, March 1992.



圖一 語音矯正系統圖

表一 矯正前後國語單音節辨識率之主觀聽覺測試結果

Syllable	impaired speech	converted speech
affricate-vowel	30%	92.5%
fricative-vowel	55.83%	63.33%
stop-vowel	86.36%	87.27%
nasal-vowel	78.89%	85.56%
liquid-vowel	76%	80%
total average	60.88%	78.38%

表二 矯正前後聽障與正常語者之間倒頻譜距離

syllable	before conversion	after conversion
Bu	12.72	3.82
Pu	10.52	4.12
Du	9.86	4.46
Tu	10.34	4.40
Gu	11.74	4.50
Ku	11.25	4.46
zhu	13.93	6.58
chu	14.49	5.65
Zu	15.24	6.38

Cu	13.32	5.96