應用轉換函式於具發音變異及台灣口音之英文語音辨識

沈涵平　吳宗憲　蔡佩珊　黎煥中*
國立成功大學資訊工程研究所
Email:{happy, chwu, pshan}@csie.ncku.edu.tw
*財團法人資訊工業策進會-創新應用服務研究所
Email: wiselyli@iii.org.tw

摘要- 本論文提出一具發音變異及口音腔調之英文聲學模型建立之方法。本論文首先藉由蒐集之英文語音資料庫(EAT)來偵測具台灣口音之英文發音所可能產生之發音變異及口音腔調，再估計出正常發音與變異發音之間的線性關係，以建立一轉換函式，再藉由此變異發音與發音特徵推測出其他同類型發音特徵之發音的變異模型，藉此達成提升辨識帶有發音變異或口音腔調的語音資料。本論文提出之方法產生之聲學模型用於辨識EAT之帶有發音變異與腔調口音之語料可得到50.91%之字正確率，比起只使用正常發音聲學模型對該語料辨識之23.26%字正確率有著顯著的提升，此技術可應用於辨識一般民眾在講述英語時容易出現發音變異或口音腔調的問題，提供更精確與便利的語音人機互動。

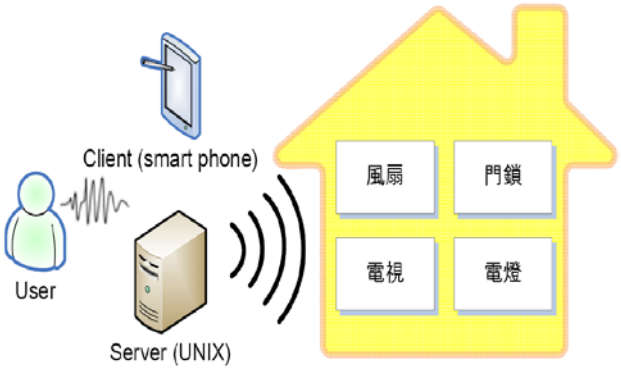
關鍵詞：語音辨識、發音變異、腔調、人機互動

Abstract- This paper presents an approach to acoustic model construction for the recognition of Taiwanese-accented English speech with pronunciation variation. This study fist analyzes the collected English speech database (EAT) to detect possible variations, and estimate the linear relation between the correct pronunciation and pronunciation variation. According to the relation, a transformation function is constructed to obtain other pronunciation variation models with similar pronunciation characteristics. This system can be applied to a speech recognition system to recognize the non-native speaker's speech with pronunciation variation or accent. The models generated by the proposed method could achieve 50.91% word recognition accuracy for the speech data with pronunciation variation and accent. Compared to the system which uses the general acoustic models and obtains 23.26% word accuracy, the proposed method can achieve significant improvement. This approach can be applied to the systems in human-machine interaction via speech.

Keywords: Speech recognition, pronunciation variation, accent , human computer interaction.

一、簡介

在人機互動中，語音為最直覺又最方便的方式。因此，語音辨識可普遍地應用於日常生活中。如圖(一)，語音辨識系統可用於數位生活上，利用語音達成電視、電燈、風扇等電器的開啟與關閉，可以提升居家生活的品質與便利，而語音辨識系統也可用於公眾場合當中，比方說車站之語音購票系統、餐廳之語音點餐系統等等，皆可提供民眾更親切與更快速服務管道，亦可減少人力資源上的浪費，所以開發一個具有良好的辨識率與富有強健性的語音辨識器能夠幫助民眾在人機互動上之便利性。



圖(一) 語音辨識與數位生活之人機互動概念圖

然而，對於使用在人機互動之語音辨識系統，使用者在口說時所產生之發音變異或腔調口音往往成為降低語音辨識器辨識率之最大因素之一。如何克服發音變異與腔調口音成為使辨識器更富強健性最重要的一個議題。而又在全球化的影響之下，學習母語以外的語言

已經成為全球的趨勢，當下尤其以英語為全球民眾爭相學習的第二語言，而民眾口說非母語之語言時往往會挾帶著母語之腔調，產生發音變異。在台灣，〝中式英文〞即為語調產生發音變異之實例。對於人機互動而言，如何去做到互動的親切與便利是非常重要的課題，倘若因為發音變異的影響導致語音辨識器不斷的出現錯誤，原本期望其所帶來的便利性反而會因為辨識率的降低成為人機互動的阻礙。

本論文之目的在設計一帶有發音變異與腔調口音資訊之英文聲學模型建立方法，整合正常發音模型與所建立之發音變異模型於語音辨識器上，使其可辨識具有發音變異與腔調之語句，以增進辨識率。本論文會根據所收集到的語料，配合正常發音之聲學模型，來找出可能的發音變異，並根據正常發音與變異發音之間的關係來估算出之間的線性轉換函式，再將此轉換函式套用於同類型發音特徵的正常發音聲學模型上，以產生同類型正常發音之發音變異模型，最後再根據鑑別性函式來檢測模型是否具鑑別性以決定哪些模型適合用於辨識，最後將正常發音與發音變異之模型一起用於語音辨識器上進行辨識。本系統期望讓民眾能與一具抗發音變異與腔調之語音辨識系統進行互動，能透過此富強健性之互動系統來提高一般民眾對語音操作的方便性，免去 語音辨識錯誤所造成的不便性。根據以上的目標，本論文提出下列研究重點:

1. 提出一語料中未出現之發音變異模型之建立方法。

2. 運用鑑別性函式以決定富鑑別力之發音變異模型。

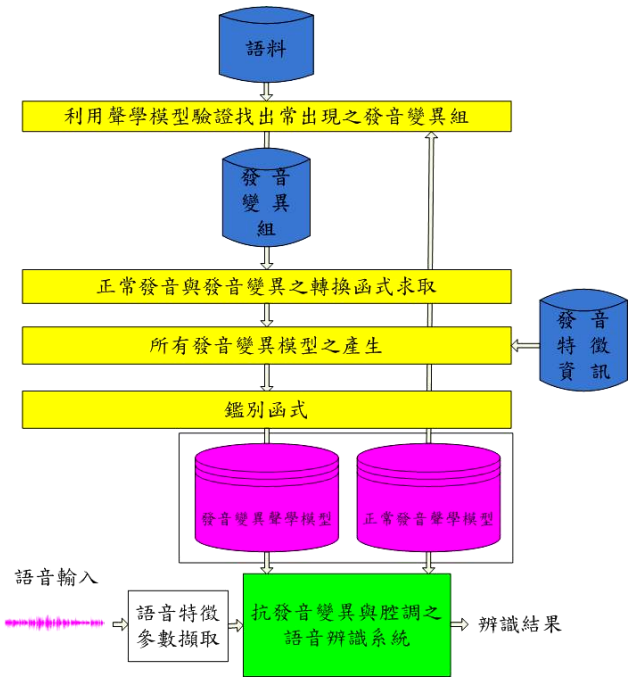
3. 整合以上方法，建立一具抗發音變異與腔調之語音辨識器。

二、有關本論文之研究情況

有關語音辨識，目前國內外主要的研究方法為 1)利用特徵擷取之技術：對於語音擷取梅爾倒頻譜(MFCC)參數，構成一向量，做為該語音之特徵，再將該特徵與各個次音節之模型做比對，找出最相似之次音節作為該語音之結果。2)隱藏式馬可夫模型(HMM)之建立：對於定義出的各個次音節利用預先蒐集好的語料建立各個次音節的 HMM 聲學模型，HMM 模型的結構可以適當的用在語音辨識上[3-5][11-12][15-18][22]。(1)(2)的方法目前被廣泛應用在各國語言的語音辨識器上，而對於具抗發音變異之語音辨識器也是一重要研究主題[1-2][6-10][13-14][19-21][23-25]。然而，目前所提的方法，主要是收集發音變異與正常發音語料一起訓練為一個模型，此法雖可增加對變異性容忍的程度，但同時也模糊了模型之精確性與鑑別力。另外此法也只能模擬出語料中所出現之變異情形，語料中未出現的部份則無法模擬到。另外也有部分研究是將發音變異模擬在語言模型或發音辭典上，而此類作法和本論文所提出之作法則可以相輔相成，在辨識效能上會有加成之作用。

三、研究方法

本論文擬於透過所建立之發音變異聲學模型，配合正常發音之聲學模型，共同辨識具有發音變異或腔調口音之語音以提高語音辨識率。圖(二)為系統架構圖



圖(二) 系統架構圖

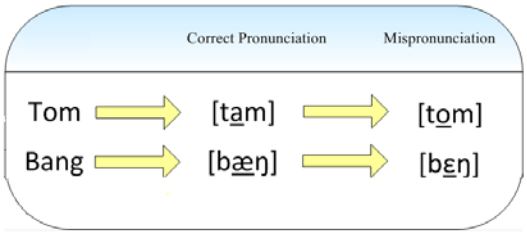
本系統首先會利用正常發音模型，去對帶有發音變異與腔調口音之語料作聲學模型驗證，找出發音變異組，並且對發音變異組找出正常發音與變異發音之間的轉換函式，再配合發音特徵推論並估算出其他發音的變異模型，最後再使用鑑別函式鑑定模型鑑別力，留下有鑑別力的變異模型配合正常發音模型做

辨識。

因此，底下將詳列出重要的幾個研究步驟：

(一) 正常發音與發音變異之聲學驗證

我們透過聲學驗證程序，來找出語料中可能之發音變異，可能之發音變異實例如圖(三)，為英語之發音變異實例



圖(三) 英語發音變異之實例

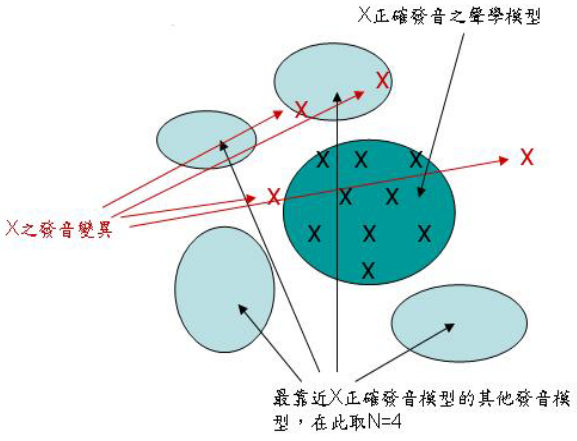
本論文提出式(1)與式(2)於聲音資料之變異程度之驗證

$$P_{verification}(x) = \log g(x \mid \lambda_{correct}) - \log g(x \mid \lambda_{anti-model}) \quad (1)$$

$$g(x \mid \lambda_{anti-model}) = \frac{1}{N}\sum_{n=1}^{N} g(x \mid \lambda_n) \quad (2)$$

其中 x 為發音特徵，Pverification(x)為發音 x 發音正確之信心值，函式 g 為辨識計分函式，

λcorrect 為 x 正確發音模型，λanti-model 為與 x 正確發音模型最相近的發音模型集，N 為與 x 正確發音模型最相似的 N 個發音模型，圖(四)為模型驗證之示意圖



圖(四) N=4 時之模型驗證

當發音 x 與 x 之正確發音模型相符，式(1)的第一個項數值會較大，第二個項數值會較小，如此 Pverification(x)的數值會較大，亦即沒有發音變異的發生；反之則是有變異情形產生，再去將此數值與一閥值做比對，即可找到發音變異組。

(二) 轉換函式之估測

在找到語料中的發音變異組之後，我們要假設此一個音與其發音變異之間的關係為一個線性轉換關係 Y=AX+R，X 為正常發音，Y 則為變異發音，我們要去估測出一個 λ 使得 P(X,Y|λ)的值為最大，當中 λ 為 X 與 Y 隱藏式馬可夫模型中的參數與轉換函式之 A 與 R，因為 X 與 Y 皆使用隱藏式馬可夫模型所以 P(X,Y|λ)可改寫為式(3)

$$P(\mathbf{X},\mathbf{Y}\mid\lambda) = \sum_{\forall q} P(\mathbf{X},\mathbf{Y},q\mid\lambda) = \sum_{\forall q} \pi_{q_0}\prod_{t=1}^{M} a_{q_{t-1}q_t} b_{q_t}(\mathbf{x}_t,\mathbf{y}_t) \quad (3)$$

其中

$$b_j(\mathbf{x}_t,\mathbf{y}_t) = b_j(\mathbf{y}_t\mid\mathbf{x}_t)b_j(\mathbf{x}_t) \quad (4)$$

$$b_j(\mathbf{y}_t\mid\mathbf{x}_t) = \Sigma(\mathbf{y}_t;\mathbf{A}_j\mathbf{x}_t+\mathbf{R}_j,\ {}_j) \quad (5)$$

$$b_j(\mathbf{x}_t) = N(\mathbf{x}_t;\ {}_j,\ {}_j) \quad (6)$$

其中 π 為初始機率，a 為狀態轉移機率，b 為狀態觀測機率，q 為狀態變數，J 為狀態指標，t 為時間指標，Σ 為變異數。

接著我們可以使用 EM 演算法去估計出 A 與 R。此 EM 演算法可分成兩個步驟:

E-step: Q 函式:

$$Q(\lambda'\mid\lambda) = \sum_{q} P(q\mid\mathbf{O},\lambda)\log P(\mathbf{O},q\mid\lambda') \quad (7)$$

$$\log P(\mathbf{O},q\mid\lambda') = \log \pi'_{q_1} + \sum_{t=1}^{T}\log a'_{q_{t-1}q_t} + \sum_{t=1}^{T}\log b'_{q_t}(\mathbf{O}_t) \quad (8)$$

$$Q(\lambda'\mid\lambda) = Q_\pi(\lambda'\mid\lambda) + Q_a(\lambda'\mid\lambda) + Q_b(\lambda'\mid\lambda) \quad (9)$$

其中 O={X,Y}={x1,y1…,xT,yT}

M-step:最大化 Q 函式

$$\hat{\lambda}' = \arg\max_{\lambda'} Q(\lambda'\mid\lambda) \quad (10)$$

利用 Lagrange multiplier 最佳化的方法求

出 A 與 R

$$\mathbf{A}_j' = \left(\sum_{t=1}^{T} r_t(j)(\mathbf{y}_t - \mathbf{R}_j)\mathbf{x}_t^T \right) \left(\sum_{t=1}^{T} r_t(j)\mathbf{x}_t\mathbf{x}_t^T \right)^{-1} \quad (11)$$

$$\mathbf{R}_j' = \frac{\sum_{t=1}^{T} r_t(j)(\mathbf{y}_t - \mathbf{A}_j'\mathbf{x}_t)}{\sum_{t=1}^{T} r_t(j)} \quad (12)$$

(三) 發音變異模型之產生

估計出某個音的正常發音與變異發音之轉換函式後，我們可將此轉換函式套至其他同類型發音特徵的正常發音模型上，轉換出同類型正常發音的發音變異模型，表(一)為英語之發音特徵

表(一) 英語之音素發音特徵

Broad Class	English
Voiced plosive	b, d, g
Unvoiced plosive	p, t, k
Fricatives	f, v, th, dh, s, sh, hh
Affricatives	ch, jh, z, zh
Nasals	m, n, ng
Liquids	r, l
Glides	w, y
Front vowels	ih, eh, ae, iy, ey
Central vowels	ah, uh, er
Back rounded vowels	Ao
Back unrounded vowels	aa, uw, ow, ay, oy, aw

舉塞擦音(Affricatives)為例，倘若我們根據語料求出了 z 正常發音與 z 變異發音之間的轉換函式，根據發音特徵我們可將此轉換函式套至 ch, jh 與 zh 的正常發音模型上，使其產生相對應之發音變異模型。

(四) 鑑別性函式篩選模型

在將所有發音變異模型使用在語音辨識器上之前，我們還會去檢測變異模型是否適合使用於辨識之上，式(13)為一鑑別性函式

$$d_i = -g(Y_i, \lambda_{Y_i}) + \max_{Y_m \neq Y_i, m=1,2...,N} g_{Y_m}(Y_i, \lambda_{Y_m}) \quad (13)$$

當中 di 為鑑別函式之鑑別度數值，g 為辨識計分函式，Yi 為第 i 種發音的變異發音，λYi 為第 i 種發音的變異模型，λYm 為第 m 種發音的發音模型。式(13)中第一項代表變異語料與變異語料模型之計分結果，第二項則為變異語料與其他語料模型計分結果之最大值，當 di 值大於一個閥值時，表示該模型易與其他模型混淆，鑑別性不高，不使用該模型於辨識；反之則表示模型鑑別性高，將其用於最後之辨識上。

(五) 語音辨識器之建立

倘若我們希望能夠了解一個說話者所說話的內容是什麼，我們就必須要有一個語音辨識器來辨識說話者說話的內容。要建立起這樣一個語音辨識系統，需要以下個幾個步驟:

(a)語音特徵擷取：根據說話者所說出的語句，做語句特徵的擷取，此處所擷取的特徵參數為梅爾倒頻譜系數(MFCC)，MFCC 主要是模擬人耳的聽覺過程，相對於其它參數它對語音波形的變化不敏感，更加穩定，並且有較低的計算量與儲存量。其計算流程如下

1) 計算頻譜能量值：

首先我們假設輸入訊號 $x(n)$，則計算方法如下：

$$\tilde{x}(k) = \sum_{n=0}^{N_w-1} x(n)W(n)e^{-j2\pi nk/N_w}; \quad 0 \le k \le N_w \quad (14)$$

其中，N_ω 為音框(frame)大小。$W(n)$ 漢明視窗(Hamming window)計算方法如下：

$$W(n) = \beta_\omega(0.54 - 0.46 \times \cos(\frac{2\pi n}{N_\omega - 1})); \quad 0 \le n \le N_\omega \quad (15)$$

其中 β_ω 是指正規劃因素(normalization factor)。則頻譜能量給定如下：

$$X_k = |\tilde{x}(k)|^2; \qquad 0 \le k < K \quad (16)$$

其中，K 等於 $N_\omega/2$ 因為頻譜對稱所以只考慮一半的頻譜即可

2) 計算每一個頻帶的能量：

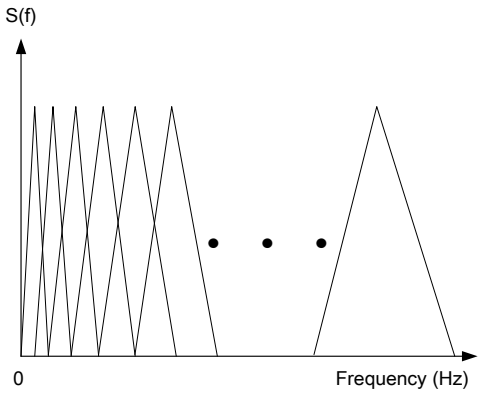
首先我們

$$E_j = \sum_{k=0}^{K-1} \phi_j(k)X_k; \qquad 0 \le j < J \quad (17)$$

其中，J 等於 18 是 ϕ_j 三角濾波器(triangular filters)實驗值最佳之個數，並且有下列限制：

$$\sum_{k=0}^{K-1}\phi_j(k)=1;\quad \forall j \tag{18}$$

三角濾波器的設計是為符合人類的聽覺模型，如圖(五)所示：



圖(五) 三角濾波器對應於頻帶之示意

3) 計算 MFCC：

$$c_m=\beta_c\sum_{j=0}^{J-1}\cos(m\frac{\pi}{J}(j+0.5))\log_{10}(E_j) \tag{19}$$

最後 MFCC 的計算可以視為對數能量(log energy)和一權重向量 V_m 相乘。權重向量 V_m 計算如下：

$$V_m=\{\cos(m\frac{\pi}{J}(j+0.5))\mid 0\le j<J\} \tag{20}$$

因此我們可以簡化公式為：

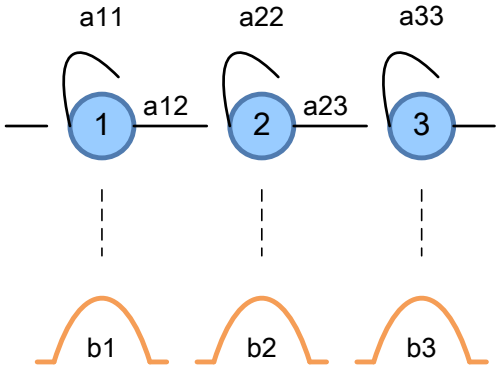
$$c_m=\beta_c\sum_{j=0}^{J-1}V_{m,j}\log_{10}(E_j) \tag{21}$$

其中，β_c 為振幅因子(amplification factor)設定與 β_ω 有關。

歸納特徵參數的擷取程序為訊號取樣及音框處理、高頻濾波器處理、漢明窗處理、快速傅立葉轉換與梅爾倒頻譜參數求取。

(b) 聲學模型：本論文中會對各個音節建立隱藏式馬可夫模型(HMM)作為其聲學模型，馬可夫模型的概念是一個離散時域有限狀態自動機，隱藏式馬可夫模型是指一馬可夫模型的內部狀態外界不可見，外界只能看到各個階段的輸出值。用 HMM 來描述語音信號需做兩個假設，一是內部狀態的轉移只與上一個狀態有關，另一個是輸出值只與當前狀態有關，這兩個假設大大降低了模型複雜度。我們使用的 HMM 是由左向右、帶自環、帶跨越的拓樸結構來對辨識單元建構模型，一個音素的 HMM 含有三個狀態，一個詞就是構成詞的多個音素之 HMM 串行起來構成的 HMM，而整個模型就是詞與靜音(silence)組合的 HMM。

圖(六)為三個狀態的 HMM 示意圖，其中 λ 為這組模型所要估算的參數，表示這組 HMM 的方式為 $\lambda=\{\pi,A,B\}$，每個 HMM 有若干狀態，每個狀態包含一組狀態轉移機率（State Transition Probability），$A=\{a_{ij}\}$，用以決定狀態 i 轉移至狀態 j 的機率，如 a_{11} 表示狀態 1 自轉的機率，a_{12} 表示狀態 1 轉至狀態 2 的機率，同時每個狀態有觀測的機率分佈（Observation Probability Distribution），$B=\{b_j(o_t)\}$，用以決定觀測對象 o_t 出現在狀態 j 的機率值，另外初始狀態為 $\pi=\{\pi_i\}$，用來表示模型是從狀態 i 開始的機率。



圖(六) HMM 的狀態示意圖

HMM 利用大量的訓練語料，訓練出一組可以描述各種語音特性的聲學模型，並且已經成功的應用在語音辨識上。在訓練 HMM 的過程中，通常利用 Viterbi 演算法（Viterbi Algorithm）來解碼，由給定的模型 λ 和觀察

樣本序列$\mathbf{O}$，選擇最佳狀態的轉移過程，可以找到在時間t時，這個音框最可能落在哪個狀態之中。我們針對 Viterbi 演算法做完回溯運算處理後的切割出來的邊界，去收集每個狀態裡的語音框，記錄下音框總數，建立對應的區間模型。當 Viterbi 演算法執行完畢後，可以得到哪些音框對應於哪些狀態的資訊。每一個狀態再利用這些音框數量資訊，建立對應的區間模型。

假設狀態序列為$\mathbf{q}=\{q_1,q_2,q_3,....,q_T\}$，觀察樣本序列$\mathbf{O}=\{o_1,o_2,o_3,....,o_T\}$，$\delta_t(i)$記錄到第$t$時間為止，累積至第$i$個狀態的最高機率值，$\psi_t(i)$為記錄最大值的陣列，演算法過程如下所示：

----------Viterbi algorithm ----------

初始化

$$\delta_1(i)=\pi_i b_i(o_1) \qquad 1\le i\le N$$
$$\psi_1(i)=0$$

遞迴計算

$$\delta_t(j)=\max_{1\le i\le N}\left[\delta_{t-1}(i)a_{ij}\right]b_j(o_t) \qquad 2\le t\le T$$
$$\psi_t(j)=\arg\max_{1\le i\le N}\left[\delta_{t-1}(i)a_{ij}\right] \qquad 1\le j\le N$$

遞迴終止

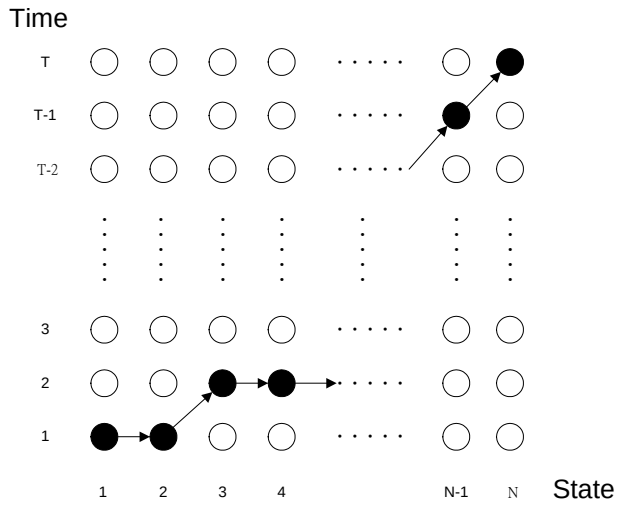
$$p^*=\max_{1\le i\le N}\left[\delta_T(i)\right]$$
$$q_t^*=\arg\max_{1\le i\le N}\left[\delta_T(i)\right]$$

回溯運算狀態路徑

$$q_t^*=\psi_{t+1}(q_{t+1}^*), \qquad t=T-1,T-2,\ldots,1$$

----------Viterbi algorithm ----------

當 Viterbi 演算法執行完畢後，可以得到哪些音框對應於哪些狀態的資訊。每一個狀態再利用這些音框數量資訊，建立對應的區間模型。圖(七)為 Viterbi 演算法示意圖：



圖(七)　Viterbi 演算法示意圖

最後在辨識上我們會使用正常發音之聲學模型與先前所建立之發音變異聲學模型一起用於辨識上。

四、系統設定及評估

本論文使用 EAT(English Across Taiwan)-台灣口音英語語料庫來做語音辨識器之模型訓練與發音變異模型之偵測，訓練與測試的語料沒有重複，該語料中有由主修英語之學生發音之語句，和非主修英語之學生發音之語句，實驗用的語料為語料庫中女性部份的語料，用來訓練正常發音的部分為主修英語的學生所發音之語句，總共取了 955 句來訓練正常發音模型，我們另外也分別使用 230 句與 660 句非主修英語學生的發音來找出發音變異組並產生發音變異模型，而用來測試的語料為非主修英語學生發音之語句，總共辨識 100 句，由麥克風錄製而成，語音資料格式為 16 kHz，16 bits，使用 HTK(Hidden Markov Model Toolkit，由英國劍橋大學所開發)做為我們實驗中的模型訓練與辨識元件，MFCC 參數取 39 維，音框大小為 32ms，而音框每次移動距離為 2/3 個音框，辭典大小為 1359 字。

我們首先使用利用訓練語料訓練出之正常發音聲學模型去辨識測試語料，以此系統做為基線系統，可以得到 23.26%的字辨識率，根據我們的推測辨識率如此低落的主因即是因為非主修英語學生的發音不正確所導致，這個結果也顯示出發音變異與腔調口音對於辨

識器是有負面影響的。

我們接著利用 660 句非主修英語語者的資料找出語料中所有之發音變異，對每個英語正常發音聲學模型皆產生一對應之變異模型，將此 79 個模型(39 個正常發音模型加上 39 個變異發音模型加上 1 靜音模型)一起用於辨識，可以得到字辨識正確率為 52.09%，由此一結果可以得知配合發音變異模型在辨識帶有發音變異與腔調口音之語音可以提升辨識效能(稱此系統為抗發音變異辨識系統 A)。

接著我們利用 230 句語料(約前一實驗語料之三分之一)，經過發音變異模型驗證、發音變異模型產生與發音變異鑑別力檢測之後，最後我們總共會產生出 27 個發音變異模型，產生之變異模型顯示於表(二)，其中之英文符號代表該變異模型對應之正常發音模型。

表(二) 產生之發音變異模型

Broad Class	Variation Model
Voiced plosive	b', d', g'
Unvoiced plosive	p', t', k'
Fricatives	f', v', th', dh', s', sh', hh'
Nasals	m', n', ng'
Liquids	r', l'
Central vowels	ah', uh', er'
Back unrounded vowels	aa', uw', ow', ay', oy', aw'

將此變異模型加上原有之正常發音模形去辨識測試語料，可以得到 50.91%的字辨識率，雖然辨識率沒有利用大量變異語料得到的發音變異模型去辨識所得到的來的好，但效果其實是差不多的，並且在模型數量上是有減少的，可以節省辨識時間，並且在蒐集帶有發音變異與腔調口音語料的部份省下大量的人力與時間成本。(稱此系統為抗發音變異辨識系統 B)，實驗結果統整於表(三)。

表(三) 實驗結果

系統	基線系統	抗發音變異辨識系統 A	抗發音變異辨識系統 B
字辨識率	23.26%	52.09%	50.91%

實驗結果證明我們所提出的方法，是可以抗發音變異與腔調口音，增加系統強健性與辨識效率。

五、結論

本論文主要提出方法來建立一帶有發音變異及台灣口音之英文語音辨識之聲學模型，並整合該模型於辨識器建立一具抗發音變異與腔調之語音辨識系統，主要挑戰在於產生語料中未出現之發音變異之模型。雖然，目前已有許多學者研究抗發音變異之語音辨識器，但是 目前可見之方法仍以建構語料中可能之發音變異為主，倘若能夠推測出語料中不會出現之發音變異，那我們可以省去很多在收集發音變異語料之人力。

本論文結合變異模型驗證、轉換函式之求取、發音變異模型之產生與模型鑑別力檢測來完成一系統化之具抗發音變異與腔調口音之英文語音辨識聲學模型建立方法。根據實驗結果顯示，我們所建構出的語音辨識系統比起沒有針對發音變異與腔調處理的語音辨識系統辨識率是有所提升的，倘若能將此系統作民眾與電腦之人機互動介面，對於民眾在使用電腦上的便利性將會有所提升。

六、誌謝

本研究依經濟部補助財團法人資訊工業策進會「 98 年度資訊應用與整合技術開發第二期計畫(1/4) 」辦理。

七、參考文獻

[1] Chien-Lin Huang, Chung-Hsien Wu, Yi Chen, Chin-Shun Hsu, Kuei-Ming Lee, "Unsupervised pronunciation grammar generation for non-native speech recognition", TENCON 2007 – 2007 IEEE Region 10 Conference Oct. 30 2007-Nov. 2 2007.

[2] Che-Kuang Lin, Lin-Shan Lee, "Pronunciation Modeling for Spontaneous Speech Recognition using Latent Pronunciation Analysis and Prior Knowledge", In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) , Volume 4, 15-20 April, 2007.

[3] D. Povey, "Discriminative Training for Large Vocabulary Speech Recognition," Ph.D Dissertation, Peterhouse, University of Cambridge, July 2004.

[4] F. Jelinek, "Statistical Methods for Speech Recognition," the MIT press, 1999.

[5] Fung. P, Liu Yi, "Triphone model

reconstruction for Mandarin pronunciation variations", In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Vol. 1, 6-10, 2003.

[6] Hamalainen. A, Bosch. L, Boves. L,"Modelling Pronunciation Variation using Multi-Path HMMS for Syllables", In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Vol. 9, 15-20, 2007.

[7] Jian Yang, Peishan Wu, Dan Xu, "Mandarin Speech Recognition for Nonnative Speakers Based on Pronunciation Dictionary Adaptation", Chinese Spoken Language Processing, 2008, ISCSLP '08, 6th International Symposium on 16-19 Dec. 2008.

[8] Kanokphara. S, Tesprasit. V, Thongprasirt. R, "Pronunciation variation speech recognition without dictionary modification on sparse database", In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Vol. 1, 6-10, 2003.

[9] Kim M, Yoo Rhee Oh, Hong Kook Kim, "Non-native pronunciation variation modeling using an indirect data driven method", In: Proc. IEEE Int. Conf. ASRU, 2007.

[10] Kulkami. K, Sengupta. S, Ramasubramanian. V, Bauer. J.G, Stemmer. G, "Accented Indian English ASR: Some early results", Spoken Language Technology Workshop, 2008, SLT 2008, IEEE, 15-19, Dec. 2008.

[11] Long Nguyen, Bing Xiang, Mohamed Afify, Sherif Abdou, Spyros Matsoukas, Richard Schwartz, and John Makhoul," The BBN RT04 English Broadcast News Transcription System," in Proc. INTERSPEECH, 2005.

[12] L. R. Bahl, F. Jelinek and R. L. Mercer, "A Maximum Likelihood Approach to Continuous Speech Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No.2, pp.179-190, 1983.

[13] Lan Wang, Xin Feng, Meng. H. M, "Mispronunciation detection based on cross-language phonological comparisons", Audio, Language and Image Processing, 2008, ICALIP 2008, International Conference on 7-9 July, 2008.

[14] Min-Siong Liang, Ren-Yuan Lyu, Yuang-Chin Chiang, "Phonetic Transcription using Speech Recognition Technique Considering Variations in Pronunciation", In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) , Volume 4, 15-20, April, 2007.

[15] Nouza, J., et al., 2004."Very Large Vocabulary Speech Recognition System for Automatic Transcription of Czech Broadcast Programs." In: Proc. Int. Conf. on Spoken Language Processing (ICSLP), Jeju, Korea, 2004.

[16] Ney, H., Ortmanns, S., "Dynamic Programming Search for Continuous Speech Recognition," IEEE Signal Processing Magazine, vol. 16, no. 5, 1999, pp. 64-83.

[17] R. Bayeh, S. Lin, G Chollet, C. Mokbel, "Towards multilingual speech recognition using data driven source/target acoustical units association", ICASSP'04, vol. I, pp. 521-524, Montreal, Canada, May 2004.

[18] S. Ortmanns, H. Ney, X. Aubert, "A Word Graph Algorithm for Large Vocabulary Continuous Speech Recognition," Computer Speech and Language, Vol. 11, pp.11-72, 1997.

[19] Sakti. S, Markov. K, Nakamura. S, " Probabilistic Pronunciation Variation Model Based on Baysian Network for Conversational Speech Recognition", Universal Communication, 2008, ISUC '08. Second International Symposium on. 15-16 Dec. 2008.

[20] Tsai. M-Y, Chou. F-C, Lee. L-S, "Pronunciation Modeling With Reduced Confusion for Mandarin Chinese Using a Tree-Stage Framework", Audio, Speech, and Language Processing, IEEE Transactions on, Volume 15, Issue 2, Feb. 2007.

[21] Wu Peishan, Yang Jian, "Acoustic modeling in mandarin speech recognition of minority accent in Yunnan", Control Conference, 2008. CCC 2008. 27th Chinese 16-18 July 2008.

[22] X. Aubert, "An Overview of Decoding Techniques for Large Vocabulary Continuous Speech Recognition,"

Computer Speech and Language, Vol. 16, pp. 89-114, 2002.

[23] Xiaodong He, Yunxin Zhao, “Prior knowledge guided MEL based model selection and adaptation for nonnative speech recognition”, In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) , Volume 1, 17-21 May. 2004.

[24] Yoo Rhee Oh, Jae Sam Yoon, Hong Kook Km, “Acoustic Model Adaptation based on Pronunciation Variability Analysis for Non-Native Speech Recognition,” In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) , 2006.

[25] Yuan-Yuan Pu, Jian Yang, Hong Wei, Dan Xu, “A study on Yunnan dialectal Chinese speech recognition”, Machine Learning and Cybernetics, 2008 International Conference on, Volume 5, 12-15 July 2008.