

第3章 母語輔助個人化發音混淆網路

系統要能偵測出語者的錯誤發音，必需要知道語者可能發生的發音錯誤規則，依據規則產生發音混淆網路，以此作為系統辨識網路來偵測語者是否有錯誤發音。系統使用的錯誤發音類型的參考來源有三：一是傳統上常見的中文母語學習者在學習英文發音時可能有的錯誤情形，另外則是本論文提出的中英文兩種不同子母音之間可能的取代對應關係，以及系統動態偵測使用者個人可能的錯誤方式。系統根據以上三種錯誤發音類型，建立個人化發音混淆網路，以此發音混淆網路作為辨識網路來偵測使用者的錯誤發音。本章節將說明三種錯誤類型的內容以及建立個人化發音混淆網路的方法和使用。

3.1 中文母語語者典型的英文發音錯誤型態

一般的發音錯誤可以歸納成以下兩種類型。

一、字轉音錯誤 (Phonetic Transcription Error)

這類型錯誤發生的原因在於學習者對英文字母拼字所對應到的音標規則不熟悉，而導致在將英文字母轉成音標時發生錯誤，而唸出不正確的音標。例如：
hibernate / haɪbəˌnet / 單字中的 i，會容易導致將 / aɪ / 唸成 / ɪ /。

表 3-1 列出幾個較常出現的錯誤情型：

表 3-1：字轉音錯誤

錯誤型態	範 例
/ a / → / o /	<u>T</u> om、J <u>o</u> hn
/ z / → / s /	day <u>s</u> 、hus <u>b</u> and
/ ɔ / → / ɑ /	wr <u>o</u> ng、c <u>o</u> rporate
/ aɪ / → / ɪ /	h <u>i</u> bernate、w <u>i</u> fi
/ æ / → / ɑ /	st <u>a</u> ff、 <u>a</u> s

二、發音錯誤 (Pronunciation Error)

這類型錯誤發生的原因是受到母語的影響，導致發音不正確。大部分的非母語者講英文時，難免都會帶著強烈的母語口音。Power 等人的研究提出[7][8][9]，以中文為母語的英語學習者，比較容易犯下列幾類發音上的錯誤。例如：/ p, t, k / 這三個無聲子音經常在字尾和句末被習慣性地省略；/ b, d, g / 這三個有聲子音在字尾和句末往往難以清晰的聽到；/ m, n, ŋ / 在字尾和句末通常是弱而短地聽到。下列表 3-2 和表 3-3 列出幾個較常出現的錯誤類型。

表 3-2：母音常見發音錯誤型態

	錯誤型態	範 例
長母音變短母音	/ɑ/ → /ɔ/	not → nought
	/ɑ/ → /u/	God → good
	/u/ → /ʊ/	fool → full
	/i/ → /ɪ/	seat → sit
長母音發音位置混淆	/æ/ → /ɑ/	cat → cart
	/æ/ → /e/	man → men
	/æ/ → /ʌ/	ran → run
	/ʌ/ → /æ/	cup → cap

表 3-3：子音常見發音錯誤型態

	錯誤型態	範例
有聲音變無聲音	/v/ → /f/	van → fan
	/z/ → /s/	rise → rice
	/b/ → /p/	bill → pill
	/d/ → /t/	made → mate
摩擦音變滑音	/v/ → /w/	vet → wet
摩擦音變爆裂音	/θ/ → /t/	thin → tin
	/ð/ → /d/	they → day
流音混淆	/l/ → /r/	light → right
摩擦音混淆	/ð/ → /z/	clothe → close
	/θ/ → /s/	thin → sin

3.2 英文音素和中文音素發音上的對應關係

語言學家提出：當新語言的子母音有和學習者的母語中相類似的音，學習者會習慣以母語的發音方式來發出新語言的音[8]。若新語言中有子母音是不存於母語中的，學習者則可能以母語的音來取代不會唸的音。於是，中文和英文的子母音之間存在著某種可能的對應關係，透過分析找出這種對應關係，我們便能從中文發音方式來加強預測中文母語學習者可能的錯誤發音。

系統中尋找中英文子母音對應關係的程式流程如圖 3-1 所示。

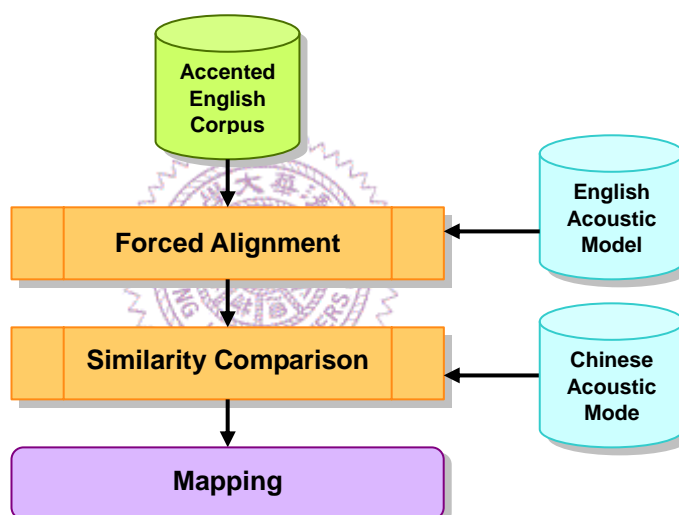


圖 3-1：尋找英文發音音素和中文發音音素對應關係流程

首先以訓練好的英文聲學模型對具有中文口音的英文語料進行音素切音，得出每個句子中每個英文音素各自的起始位置，再對句子中每個英文音素區段求算每個中文聲學模型的對數機率值，有最高機率值的中文音素則為與這個英文音素區段最相近的發音。見圖 3-2 的範例，單字 analog 經過切音後切割出英文音素 / ænəlɒg / 每個音素的位置，再計算找出各區段間最相似的中文音素為 / ㄙㄣˋ ㄣˊ ㄌㄠˊ ㄍㄨˋ ㄉㄨˋ ㄍㄨˋ /。最後，統計每個英文音素和各種中文音素的對應次數，累積得最高票數的則視其為最佳對應。

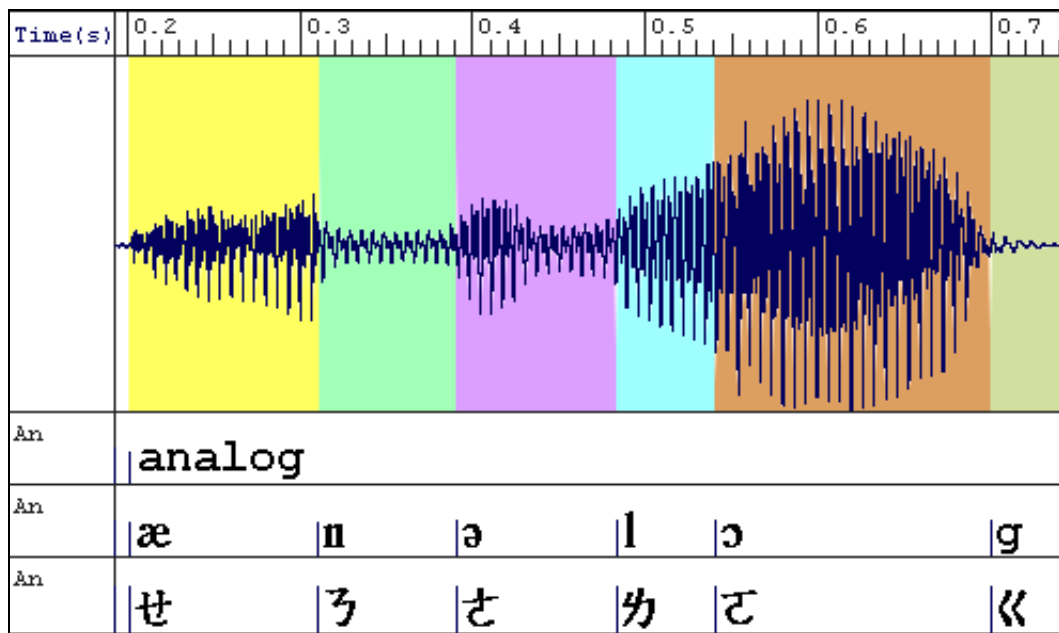


圖 3-2：尋找取代對應關係範例

3.3 偵測和建立個人化發音錯誤規則

目前多數的 CAPT 系統使用的 PCN 皆為依據事先定義的錯誤規則而建立的，譬如前兩小節描述的語言學家歸納的錯誤或是兩種語言音素在發音上的對應，依此產生 PCN。但這種 PCN 的內容固定而不彈性，能預測出的錯誤範圍有限，無法偵測出定義之外的錯誤。而每個人的發音方式和習慣必不相同，不同學習者皆可能具有各自獨特的口音腔調，於是導致個人化的錯誤混淆類型。因此，我們提出以下的方法，讓系統在疊代更新過程中，自動偵測個人化發音錯誤類型來產生個人發音錯誤規則，並用於建立個人化調適的發音混淆網路（PPCN），讓系統能更彈性地偵測出使用者任何可能的錯誤情形。

產生個人化發音錯誤規則的流程如圖 3-3。當使用者輸入語音進來，除了以先前步驟建立好的發音混淆網路（BPCN）進行辨識，偵測使用者的語音是否含有系統定義中的錯誤情形。同時，系統會對使用者的語音另外進行個人錯誤型態偵測。將輸入

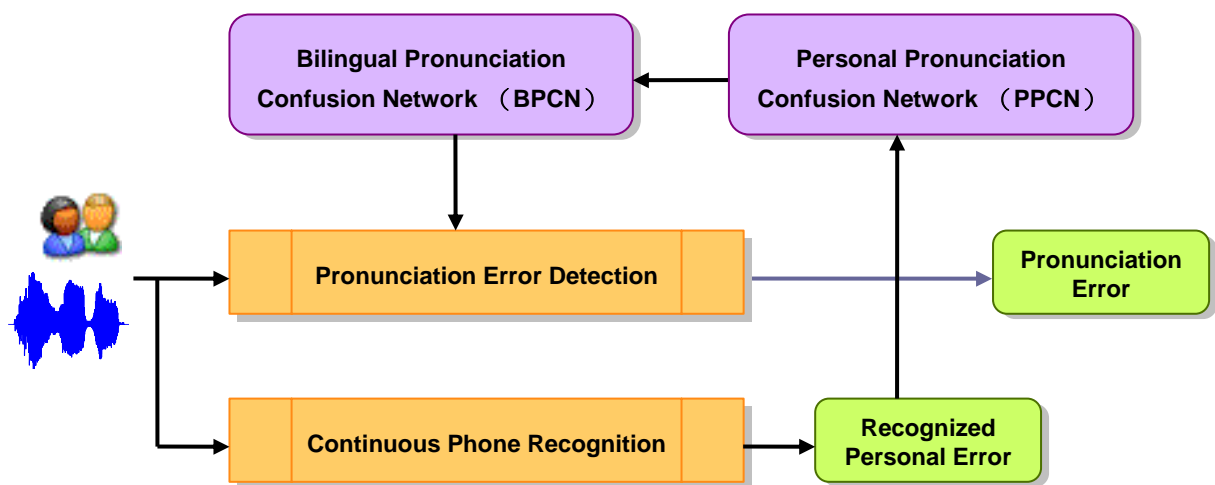


圖 3-3：建立個人化發音混淆網路（PPCN）流程圖

語音進行 Continuous Phone Recognition，並從辨識結果中觀察是否有使用者個人特殊的錯誤混淆型態，若有則將之定義成使用者的個人錯誤混淆規則，並依據規則建立個人化發音混淆網路（PPCN）。這個個人化發音混淆網路於是加入到原先系統使用的 BPCN 中，當使用者再次輸入相同文字內容的語音時，系統將以新合併的發音混淆網路進行辨識。

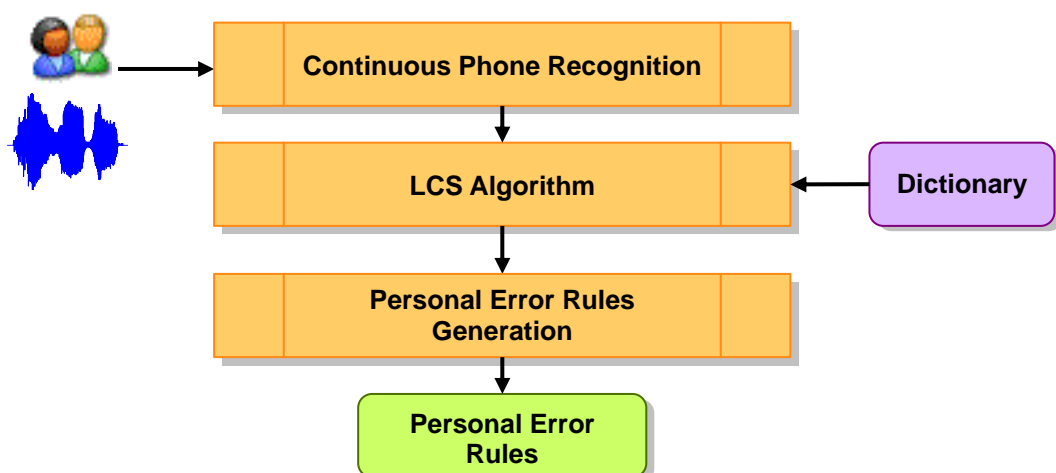


圖 3-4：建立個人化發音錯誤規則流程

圖 3-4 為建立個人化發音錯誤規則的細節流程。首先，系統對使用者的輸入語音進行連續音素辨識，接著將辨識結果與字典的正確發音內容利用 LCS (Longest Common Subsequence) 演算法進行比對，找出兩個字串中最小差異的部份，並從中偵測出語者可能的習慣錯誤混淆發音，以產生語者個人錯誤混淆規則。其中，系統是以英文字典中的單字拼音作為正確的語音發音內容，而 LCS 演算法的作用是在於找出輸入兩字串中最長的相同子字串。經由 LCS 演算法所得出的子字串，會是將原本字串中刪除零個到多個字元後，兩個輸入字串中最長的相同部份。圖 3-5 為 LCS 演算法的執行範例圖示。英文單字 postcard 的字典拼音為 / postkard /，而使用者語音經過 Continuous Phone Recognition 後得的辨識結果為 / poztəkar /，利用 LCS 演算法於是找出最大子字串為 / potkar /，於是相對得知兩個字串中有三個不同處，分別為 /s/→/z/、/ /→/ə/、/d/→/ /。由此系統可判斷出在 /s/→/z/ 處，使用者可能發生拼音取代性錯誤；在 / /→/ə/ 處，使用者可能發生拼音插入性錯誤；在 /d/→/ / 處，使用者可能發生拼音刪除性錯誤。

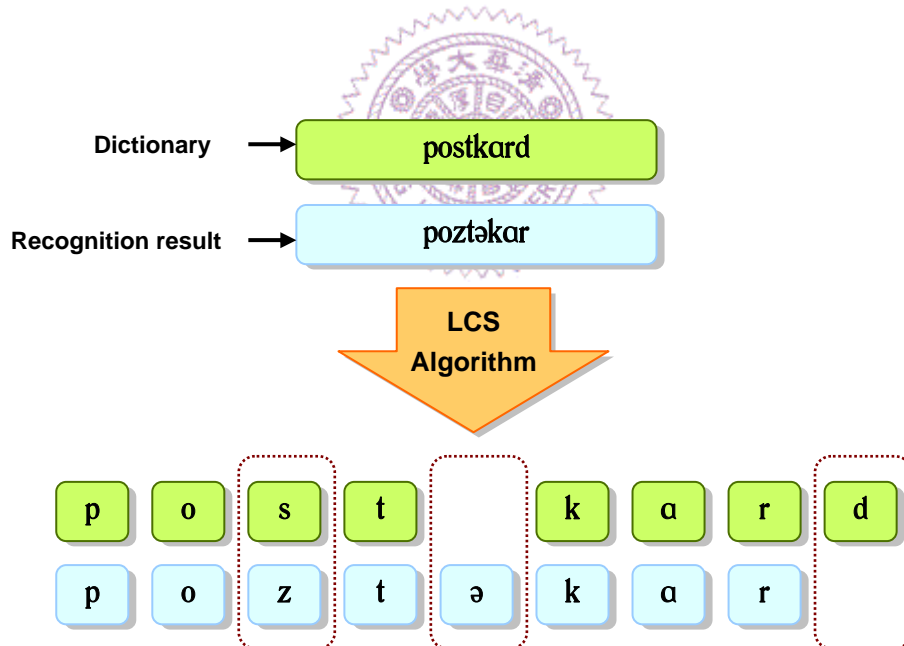


圖 3-5：LCS 演算法尋找最大子字串範例圖示

經過 LCS 演算法找出字串差異部份後，系統則能從結果中偵測找出使用者可能的個人發音錯誤習慣，建立個人化發音錯誤規則。系統偵測錯誤規則的分類方式有三種。第一種為音素在單字中的位置，分為字首、字中或字尾。第二種為音素的種類，

分為母音或子音。第三種為音素錯誤的型態，分為插入性錯誤、刪除性錯誤和取代性錯誤。若以圖 3-5 的辨識結果為範例，系統於是能建立三項個人化發音錯誤規則。第一項規則為：字中子音 /s/ 發生取代性錯誤變為子音 /z/；第二項規則為：字中子音 /t/ 後方產生插入性錯誤增加母音 /ə/；第三項規則為：字尾子音 /d/ 發生刪除性錯誤。

3.4 建立個人化調適發音混淆網路

發音錯誤的偵測主要是先針對句子內容建立對應的辨識網路，利用此辨識網路來辨識語音內容。根據前三節介紹的發音錯誤類型，我們可以將所有可能的發音建立在辨識網路中（包括正確發音與所有可能的錯誤混淆發音）。例如，表 3-4 中列出單字 “doze off” 在英文發音時傳統常見的發音錯誤與英文和中文音素的發音對應關係，圖 3-6 則是依據表 3-4 中的規則所產生的 BPCN。在圖 3-6 的 BPCN 中，中間實線路徑為依照字典正確發音的內容，上方點線路徑為傳統常見的發音錯誤，下方虛線路徑則為英文和中文音素發音對應關係的路徑。當使用者輸入語音進入系統，系統自語音中取出 39 維特徵參數後，依照此辨識網路利用 Viterbi 演算法找出一條最佳路徑產生辨識結果，再將辨識結果和人工標示的真實發音內容比對。當系統偵測出辨識結果內容和正確發音路徑內容有不相符時，即辨認出使用者於該句語音中有存在錯誤混淆發音。

表 3-4：與 “doze off” 相關的常見發音錯誤與中英文取代對應關係

Typical Error		Bilingual Map	
Phone	Error Type	Phone	Error Type
/d/ → /t/	Substitution	/o/ → ㄊ	Substitution
/z/ → /s/	Substitution	/ɔ/ → ㄙ	Substitution
/ɔ/ → /o/	Substitution	/f/ → ㄈ	Substitution
/o/ → /ɔ/	Substitution		

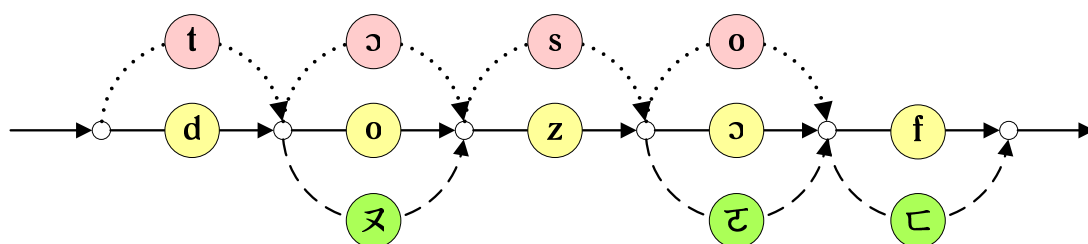


圖 3-6：“doze off”的 BPCN

當系統再經過偵測建立使用者個人發音錯誤規則後，我們得到使用者個人的錯誤混淆規則，例如表 3-5 所示。於是系統可產生包含使用者個人錯誤混淆規則的個人化發音混淆網路（PPCN）。如圖 3-7，下方的點虛線路徑即為使用者個人的錯誤規則。接著，系統便能將新增的 PPCN 路徑加至原本的 BPCN 中，建立包含三種規則來源的 BPPCN，其路徑如圖 3-8 所示。

表 3-5：與“doze off”相關的使用者個人錯誤規則

Personal Error	
Phone	Error Type
/z/ → / /	Deletion
/o/ → /ɔ/	Substitution
/f/ → /v/	Substitution

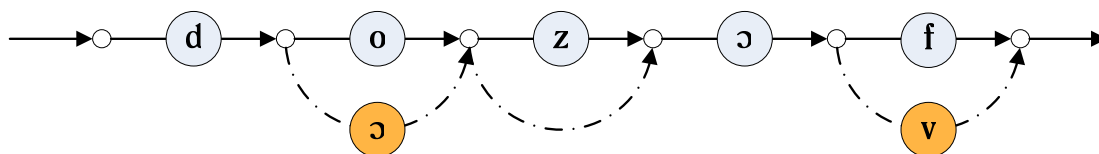


圖 3-7：“doze off”的 PPCN

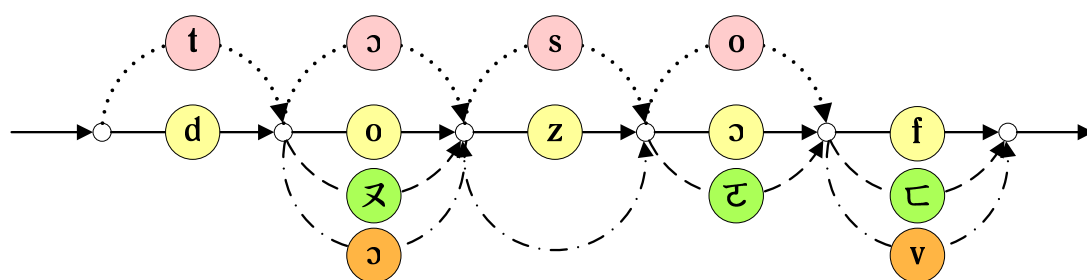


圖 3-8：“doze off”的 BPPCN

我們的系統以此方式彈性地偵測使用者任何可能的錯誤型態並新增個人化錯誤規則，動態產生包含更多種錯誤混淆型態的發音混淆網路，打破一般系統偵測範圍的限制而更有效的偵測錯誤混淆發音。

