資源受限運算環境下華英混雜語音辨識系統

Mandarin/English Mixed-Lingual Speech Recognition System on Resource-Constrained Platforms

洪維廷　Wei-Tyng Hong
元智大學通訊工程學系
Department of Communications Engineering
Yuan Ze University

陳弘啓　Hong-Ci Chen
元智大學通訊工程學系
Department of Communications Engineering
Yuan Ze University

廖宜斌　I-Bin Liao
中華電信研究所
Telecommunication Laboratories, Chunghwa Telecom

王文俊　Wern-Jun Wang
中華電信研究所
Telecommunication Laboratories, Chunghwa Telecom

摘要

本論文提出結合華語、英語語音模型進行混合關鍵詞辨識，並且於資源受限制的情況下進行全整數運算。我們將所處理過後的定點化特徵參數經過 RASTA 濾波器[1]，成爲定點化 RASTA 特徵參數。而針對關鍵詞詞彙結構的相關性，我們建立一個樹枝狀的搜尋架構，並且利用光束搜尋法，減少辨識語料中音框所需的音節節點，進而減少搜尋空間，並且維持一定的辨識率。

針對辨識語料所產生的華語、英語語音模型相似度分數差異，我們提出偏差補償值應用於英語模型，並提出改變前置詞搜尋機制。增加可忽略搜尋前置詞路徑與無關詞垃圾模型路徑，以因應使用者沒有按照使用規則說出定義內前置詞的情形，藉此測試對於主詞的辨識率。

關鍵詞：　詞典樹、光束搜尋法、垃圾模型

Keywords：Lexicon Tree、Beam Search、Garbage Model

一、緒論

目前在自動語音辨識的發展課題上，大部分仍以單一語言構成的語音爲主，混合語言的語音研究仍則較爲少見。在 2006 年，國內成功大學電機工程研究所黃建霖等人[2-3]，利用聲學與文脈分析應用於多語(multilinguality)語音辨識單元的產生，並且利用融合技術(Fusion)結合聲學相似度及前後文脈分析，有效提升多語語言的辨識率。在 2008 年，國內成功大學電機工程研究所李奇峰等人[4]發展一套多語言辨識單元集技術，在原有的語音辨識系統前端，建立一個有效的混語語音模型，並且將系統實現於 PDA 之嵌入式系統裝置上。在 2004 年時工研院資通所所發表的「中英文混雜關鍵詞萃取技術」[5]，提出一個中英文辨識分數的偏差補償值應用於中英文混雜辨識，去克服中英文聲學模型間的差異性，並且在關鍵詞認證方面，則提出一個正規化的驗證機制，而從驗證分數裡找出一組最好的結果。

上述所提到的多語語音辨識即是結合多個單一語言(monolingual)模型，於前端判斷所說詞彙的語言種類，之後再進行多語語音辨識，而在本論文中的混合語音辨識研究，則是辨別華英語混合的關鍵詞詞彙，並且於辨識時灌進華語、英語語音模型，利用每個時間音框中所累積的相似度分數，去判別出最佳的辨識結果。

在於關鍵詞的萃取方面，可以分爲關鍵詞模組與無關詞模型。在 1985 年，Higgins 與 Wohlford[6]利用連續語音辨識應用於關鍵詞萃取(keyword Spotting)，並且定義填充模型(filler template)去表示無關詞部份。在 1989 年，Rohlicek 等人[7]提出在關鍵詞模組狀態中的機率密度函數加入權重值去建立無關詞模組。在 1990 年，Rose 與 Paul 提出[8]使用單音節來當作無關詞模組。近幾年來，陸續提出利用 HMM Base 的次音節關鍵詞與無關詞模組應用於關鍵詞的辨識。在無關詞模組方面，在 1996 年，Caminero 等人[9]提出加入垃圾模型(Garbage Model)，並且套用鑑別分析準則於言語判別(utterance verification)。在辨識結構上，加上垃圾模型於文法規則上，可以降低對於關鍵詞誤判的機率。

當系統的無關詞與關鍵詞模組加大時，所需要的搜尋時間將會大大增加，尤其是花費大量資源在搜尋路徑方面，這對於應用於資源受限的智慧型裝置有極大的挑戰，所以必須利用較快速有效的方法，去達成一定水準的辨識結果。在 1990 年，Frank k.Soong 等人[10]提出利用樹枝狀結構下的快速搜尋法找尋 N-Best 的辨識結果，應用於連續語音辨識。在 2004 年，Xie Lingyun 等人[11]，提出在每個時間音框點，利用動態刪減去調節搜尋範圍，並且應用於大詞彙連續語音辨識系統。

本論文模擬於資源受限的環境下做全整數運算。首先，建構前置詞(Prefix)與主詞(Main)的詞典樹(Tree Lexicon)，利用詞彙內字與字的連結關係，佈置成樹狀結構，爾後當成搜尋時的依據。當測試語料進來時，我們利用光束搜尋法(Beam Search)去找出每個語料的每個音框(Frame)可能存活的節點，並且經過排序剔除分數值較低的節點，最後找出一組最佳路徑當成辨識結果。而爲了使系統更加有延展性，我們改變前置詞的搜尋機制，加入可忽略搜尋前置詞路徑，測試使用者忘記說定意內前置詞時，可以繞過此路徑保留對於主詞的辨識效能；加入無關詞垃圾模型，測試使用者說錯定義內前置詞時，錯誤路徑可以被吸收，藉此保留對於主詞的辨識率。

二、華英語混合辨識系統

華語、英語混合辨識的系統流程圖如下圖 2-1。

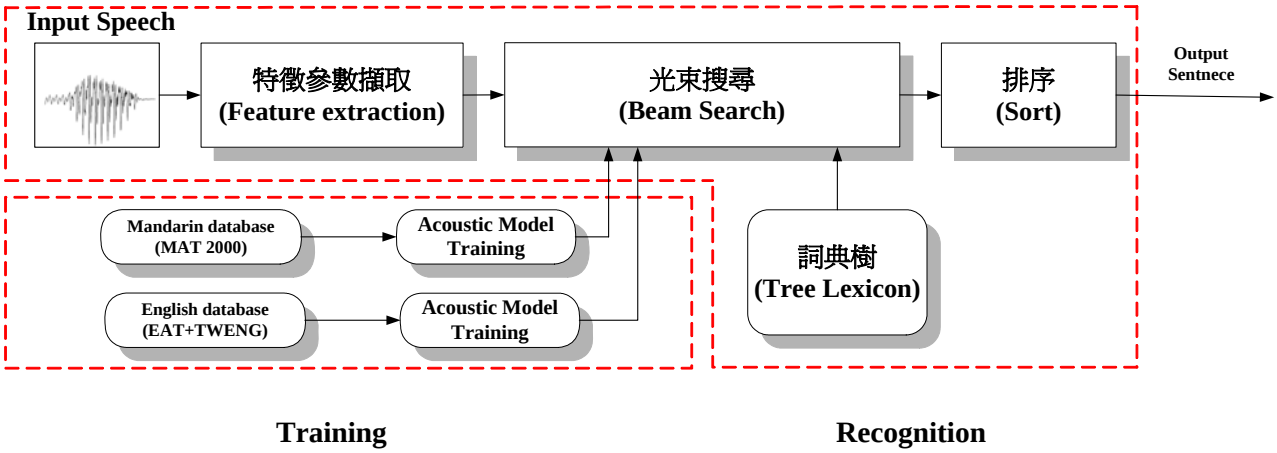

圖 2-1：華語、英語混合辨識流程圖

根據上圖的流程圖，我們的系統主要分爲兩個部份，分別爲訓練階段與辨識階段。其中訓練階段是在 Linux 下做浮點數運算，利用隱藏式馬可夫模型分別訓練出華語語音模型與英文語音模型。在測試階段，爲了符合於資源受限的環境，所以我們於 PC 上進行全整數運算。當一個測試語料進來時，我們先做語音訊號的特徵參數擷取，之後進入光束搜尋法的程序，結合詞典樹與匯入華語、英語語音模型進行關鍵詞萃取。並從所記錄的節點路徑中，經過語音模型相似度分數的累積與排序，尋找出最佳的前三名辨識結果。

2.1 隱藏式馬可夫模型

隱藏式馬可夫模型是一種以機率統計的方式來做辨識的模型，辨識語音的度量是計算從模型產生的機率值大小。一般常用於狀態觀測機率的機率密度函數爲高斯混合模型(Gauss Mixture Model)。但我們爲了因應在智慧移動裝置上面的資源限制，所以採用拉普拉斯分佈(Laplacian Distribution)做爲我們的狀態觀測機率。其 probability density function 定義如下式：

$$Lap(x,u,v)=\frac{1}{2v}\exp\left(-\frac{|x-u|}{v}\right) \tag{1}$$

其中 u 爲 location parameter，v 爲 scale parameter。假設一個維度爲 D、特徵值爲 x 的語音訊號，假設在每個特徵參數值之間是獨立的關係。我們以此狀態中所有混合數(mixture)最大的機率值來代表隱藏式馬可夫模型第 j 個狀態的觀測機率：

$$\begin{aligned} b_j(x) &= \max_m \left\{ \prod_{d=1}^{D} Lap\left(x_d; u_{j,m,d}; v_{j,m,d}\right) \right\} \\ &= \left\{ \prod_{d=1}^{D} \frac{1}{2v_{j,m,d}} \exp\left(-\frac{|x_d - u_{j,m,d}|}{v_{j,m,d}} \right) \right\} \end{aligned} \tag{2}$$

其中$u_{j,m,d}$和$v_{j,m,d}$分別爲隱藏式馬可夫模型第 j 個狀態上第 m 個混合數(mixture)之第 d 維度的 location parameter 和 scale parameter。假設 k 滿足下列式子：

$$k = \arg\max_{m} \left\{ \prod_{d=1}^{D} Lap\left(x_d; u_{j,m,d}; v_{j,m,d}\right) \right\} \tag{3}$$

$b_j(x)$的 log 式子可表示如下：

$$\log\left(b_j(x)\right) = C_{j,k} - \sum_{d=1}^{D} \frac{\left|x_d - u_{j,k,d}\right|}{v_{j,k,d}} \tag{4}$$

其中$C_{j,k}$爲和 HMM 第 j 個 state 第k 個 mixture 的參數相關的數值(和特徵值 x 無關)，可預先算出作爲另一個模型參數。接下來要進行的是降低計算量和防止溢位(overflow)：

1 經由適當的定點化技術，將$x_d, u_{j,k,d}, v_{j,k,d}, c_{j,k}$轉換成 16-bit 整數$x_d^{'}, u_{j,k,d}^{'}, v_{j,k,d}^{'}, c_{j,k}^{'}$，如此可保證$\log(b_j(x))$爲 16-bit 整數範圍；另外在 Beam Search 模組中 Viterbi Search 使用之 log-likelihood 累加陣列(accumulated array)則以 32-bit 整數宣告，這樣可降低累加陣列 overflow 的機率。

2 將除法運算拿掉，令整數$\beta_{j,k,d} = \frac{2^s}{v_{j,k,d}^{'}}$，經過適當式子重整，可以由下列的全整數運算式子逼近：

$$\log(b_j(x)) \approx C_{j,k}^{'} - \left[\sum_{d=1}^{D} \beta_{j,k,d} \times \left| x_d^{'} - u_{j,k,d}^{'} \right| \right] >> s \tag{5}$$

其中 >> 爲 bit-shift right 運算元。

2.2 詞典樹

在建構整個詞典樹的過程中，包含了前置詞(Prefix Word)與主詞(Main Word)。在本論文中，所有詞彙皆爲華語、英語或華英語混雜的語料，其中前置詞的個數爲 10 個，主詞個數爲 200 個。在此，我們舉出一個範例，例子中前置詞有〝Email〞與〝查詢〞，而主詞有〝元智大學〞、〝元智大學通訊所〞與〝元智大學通訊所辦公室〞等五個詞彙，利用樹狀結所構成的辨識詞彙集合如圖 2-2 所示。T 代表詞彙的終結。

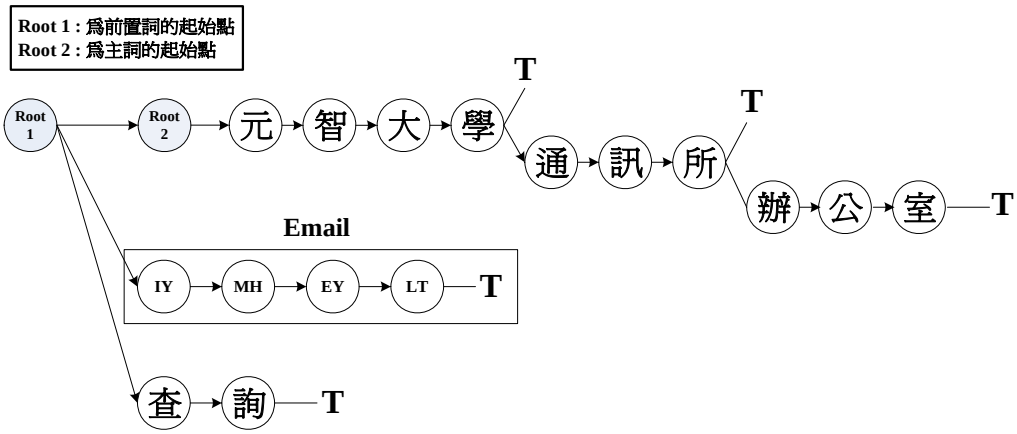

圖 2-2：樹狀結構圖(一)

依據上圖 2-2 的方式可以將關鍵詞詞庫建立成一個樹狀結構的詞典樹，其中華語音節點對應 411 個音節碼，而英語音節點則是對應 134 個右相關音素模型(Phone Model)，之後將一階動態搜尋器的搜尋空間由華語的 411 音節碼與英語的 134 個右相關音素模型展開所可以連結的音節點節點。如果爲華語節點，則有 7 個狀態，包含基本音節之 Initial 2 個與 Final 4 個狀態，跟一個可以跳過的靜音狀態來描述音與音之間可能存在/不存在的靜音，如下圖 2-3 所示。如果爲英語節點，則有 3 個狀態，包含 2 個基本音節狀態與 1 個靜音狀態，如圖 2-4 所示。在搜尋的法則上，華語部份於 Final 最後一個狀態限制只可以連結靜音狀態或者華語部分的 Initial 第一個狀態或者英語部份第一個狀態；英語部份則是在狀態 2 之後只可以連結靜音狀態或者華語部份 Initial 第一個狀態或者英語的第一個狀態。

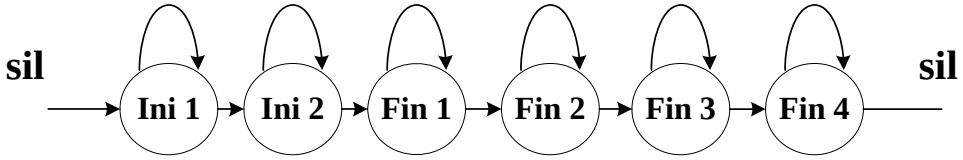

圖 2-3：華語節點中之狀態圖

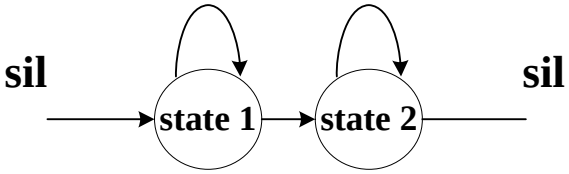

圖 2-4：英語節點中之狀態圖

2.3 光束搜尋法實做

在光束搜尋法設計實作方面我們考慮音框中語音節點會屬於靜音節點、華語節點、英語節點與垃圾模型節點四種情形。四種不同的節點會根據詞典樹與文法限制去展開可連結的節點，並且依據節點內的語音模型相似度分數大小，通過光束搜尋法的篩選，產生光束寬度(BeamWidth)。以下我們利用虛擬碼(Pseudo Code)去描述系統內一個音框資訊搜尋語音節點的演算法：

*Step*1：t=0 的音框進來。(t=音框 index)

*Step*2：依照佈於詞典樹裡所有前置詞詞彙的第一個音節去計算華語或英語語音模型的相似度分數，並且與靜音模型相似度分數進行分數上的排序。

*Step*3：t=1 的音框進來，依據 *Step*2 所產生的音節個數，重複執行 *Step*5 與 *Step*6，並進入 *Step*7 產生新的音節個數。

*Step*4：t=2 到 t=T 的音框進來，依據 *Step*7 所產生的音節個數，重複執行 *Step*5 與 *Step*6。

*Step*5：節點判別爲靜音節點，當節點內標號爲-1 時，則生長佈於詞典樹裡所有前置詞詞彙的第一個音節點；標號不爲-1 時，則依據詞典樹中此標號之後可以連結的節點進行連結。

*Step*6：節點判別爲華語節點、英語節點或垃圾模型節點，如果此節點未生成前一個語音狀態，則生成；如果非最後一個語音狀態，則產生下個語音狀態。如果已經到最後一個語音狀態，則判斷是否爲前置詞或主詞的終結點，並且可以連結到靜音節點。此時節點如果不爲主詞的終結點，則此節點可以繼續連結其他華語節點或英語節點。

*Step*7：累積 *Step*5 與 *Step*6 所產生的節點，進行靜音、華語、英語、垃圾模型節點間語音模型相似度分數間的排序，並且經過光束搜尋法的篩選，剔除分數較小的節點。

*Step*8：最後一個音框內的語音節點經過相似度分數排序過後，進入辨識程序。

在此我們也舉出一個例子，說明詞典樹與光速搜尋法如何結合應用於系統中。我們設定總共有 6 個詞彙，包含了 3 個前置詞與 3 個主詞，共有 15 個節點。依據詞彙內音節佈成的樹狀結構如下圖 2-5：

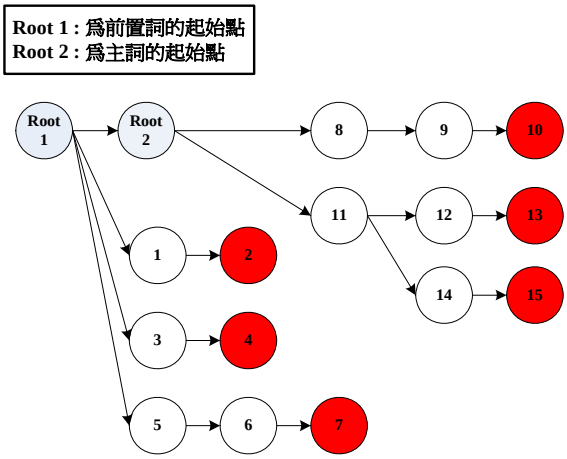

圖 2-5：樹狀結構圖(二)

其中數字 1~15 代表詞彙內的音節，紅色代表此詞彙的結束節點。下圖 2-6 以圖 2-5 的詞典樹爲依據所產生的光速搜尋法路徑。當測試語料進來時，先生長前置詞、靜音與垃圾模型節點，再生長主詞節點，並且依照節點內狀態所對應的語音模型相似度分數大小，進行排序與篩選。在此，圓圈間的連結所代表的是節點內狀態的轉移情形。而此例子中，每個節點有 3 個狀態數，爲 S0、S1、S2。而靜音與垃圾模型節點各爲一個狀態數。

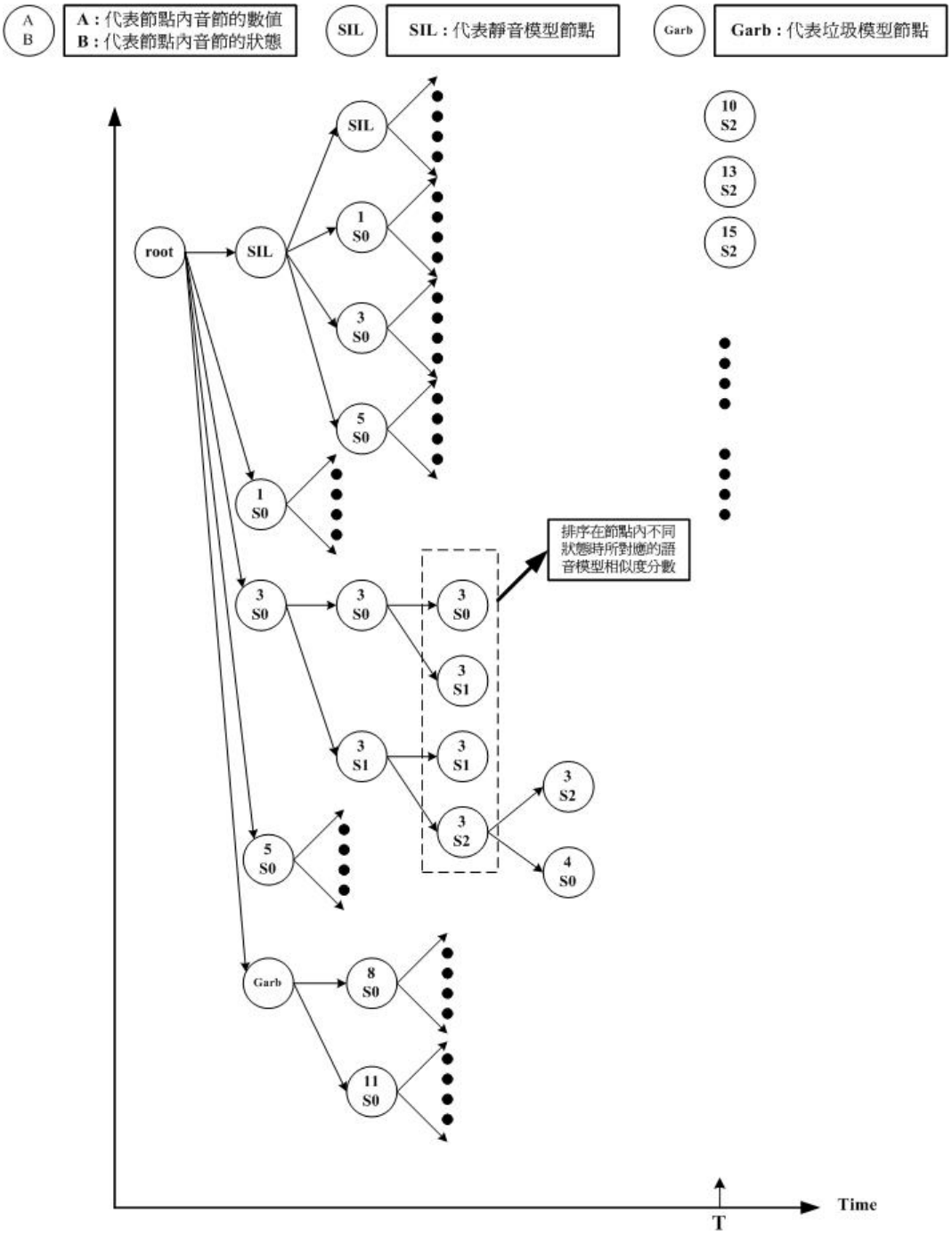

圖 2-6：光束搜尋法路徑圖

2.4 華英語混合關鍵詞萃取

所謂關鍵詞的萃取，是指在特殊辨認詞彙下，事先對此特殊詞彙選取若干個關鍵詞，在辨識的時候只要將預先定義好的關鍵詞萃取出來。這部分我們主要是介紹華英語混合的語音辨識，重點是放在關鍵詞的辨識上面，其中關鍵詞包含了前置詞與主詞的辨識，例如：〝*查詢 元智大學*〞。在一些特殊情況下，例如使用者忘記說或說錯定義內前置詞時，針對於主詞關鍵詞萃取辨識還是我們的目的，這時候我們就在搜尋機制中加入可忽略搜

尋前置詞的路徑或無關詞垃圾模型路徑，去降低對於主詞的破壞性，保留對於主詞的辨識率。

而關於搜尋機制的改變，我們主要針對前置詞搜尋機制方面做探討。首先，於前置詞詞庫中加入可忽略搜尋前置詞的路徑，此路徑並非定義內的前置詞，而是爲一個靜音模型的路徑。對此我們測試使用者忘記說定義內前置詞時對於主詞的辨識率。爲了使系統更有彈性與效能，接下來我們在前置詞詞庫中加入無關詞垃圾模型路徑，去抵抗當使用者說錯定義內前置詞的情況。而整個前置詞搜尋路徑的改變，如下圖 2-7 所示。因爲關鍵詞辨認的技術比連續語音辨識來的廣泛且穩定，所以我們將對華英語混合的關鍵詞辨識做以下更詳細的討論。

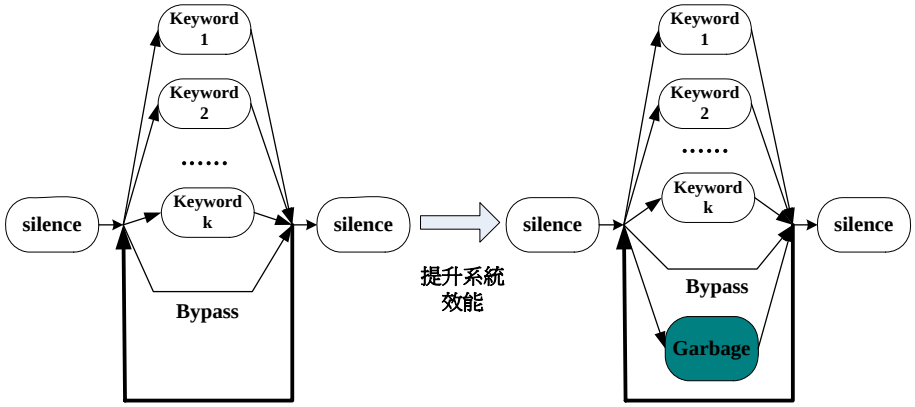

圖 2-7：前置詞搜尋機制圖

2.4.1 關鍵詞模組

關鍵詞即根據特殊辨識詞彙的不同，而去事先定義、選取的辨識標的。在前置詞方面，有尋找、打電話給、Email...等等日常生活中人與人之間會交際互動的詞彙；而在主詞方面，則以中英文人名爲主。我們採用次音節中的右相關音素模型(RCD)串連來產生關鍵詞模組，作爲聲學層次的辨識，這樣可以根據特殊辨識詞彙的不同，而前置詞或者主詞的加減，不需重新訓練模型，使的系統更有彈性。

2.4.2 無關詞垃圾模型

無關詞垃圾模型類似於是填充模型，利用填充模型的混淆來拉下無關詞的分數，進而增加辨識率。而本論文中的無關詞垃圾模型是將華語訓練語料中全部聲母部分所有音框特徵參數資訊重新訓練成一個單一新的聲母模型，只有一個狀態數且混合數爲 32，而也將韻母所有音框特徵參數資訊重新訓練成一個單一新的韻母模型，同樣也設定爲一個狀態數混合數爲 32。匯集新的聲母、韻母模型成爲我們所定義的垃圾模型，並且把它加入到所定義的前置詞詞庫路徑之中。雖然我們無法預測使用者的說話情形，但我們必定知道在每句句子之中必定含有我們要辨識的對象，即關鍵詞的存在。在辨識的過程之中我們可以利用連結字音的辨識方法將關鍵詞與無關詞結合成一個辨識單元作處理，達成我們辨識關鍵詞的目的。

2.4.3 關鍵詞萃取的排列

關鍵詞的萃取架構，在前置詞部分，可以包含關鍵詞模組或者無關詞模型；主詞部份則有關鍵詞模組。在本論文中，我們舉出二種主要出現於關鍵詞萃取中搜尋機制改變的情形。假設定義 2 個前置詞分別爲〝**查詢**〞與〝**Email**〞與定義 3 個主詞，分別爲〝**元智大學**〞、〝**台灣大學**〞與〝**交通大學**〞。第一種情況如圖 2-8，

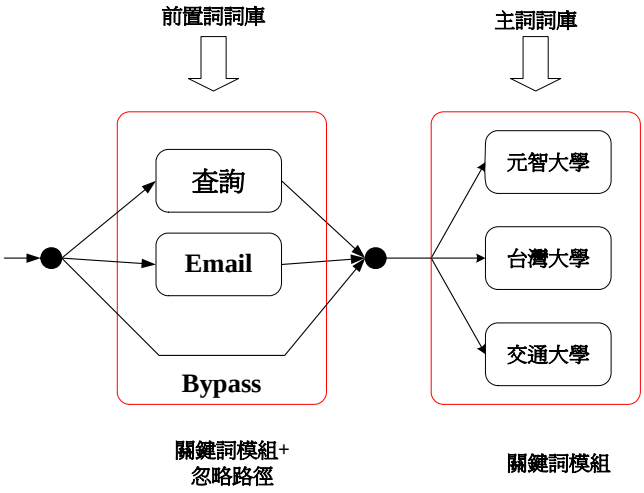

圖 2-8：於前置詞中加入可忽略搜尋前置詞路徑圖

在前置詞詞庫中加入了可忽略前置詞路徑，測試使用者忘記說前置詞時，是否可以讓前置詞搜尋機制繞過此路徑，保留住對於主詞的辨識。第二種情形如圖 2-9，

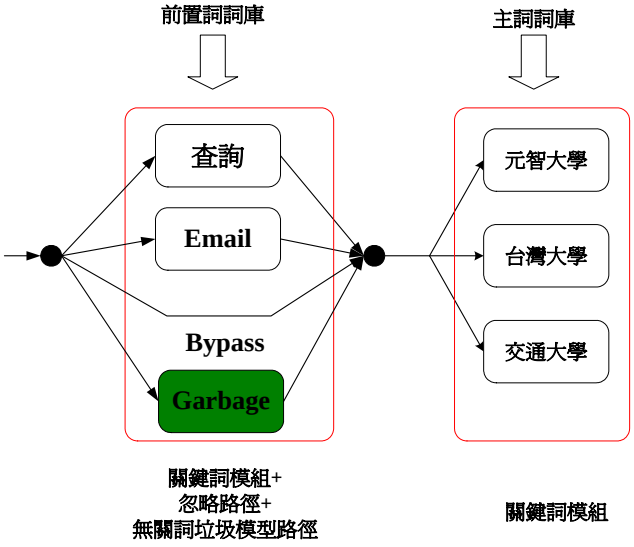

圖 2-9：於前置詞中加入無關詞垃圾模型路徑圖

於前置詞詞庫中加入可忽略前置詞路徑與無關詞垃圾模型路徑，可應用於當使用者說錯定義內前置詞或忘記說定義內前置詞情況時，不會大幅破壞對於主詞的辨識率。

2.5 語音模型的補償

華語與英語的訓練語料來自不同的錄音方式與環境，加上辨認單元和參數量不同造成模型解析度不同，所以當測試語料進來時，我們必須去觀察所產生的華語或英語模型間相似度分數的差距並且作補償的動作。在此，我們會分別對英語模型與無關詞垃圾模型做偏差值補償。

2.5.1 華英語混合關鍵詞的偏差補償

華語語音模型與英語語音模型因訓練參數量不同，所以當要辨識的語音進來時，所產生的華語、英語相似度分數在分布空間上必然不同，進而在分數上產生一定的落差。經過統計與觀察之後，兩者間的差距如圖 2-10 所示。

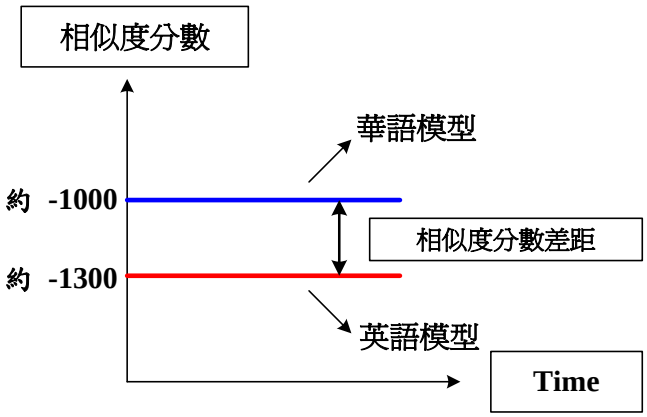

圖 2-10：華英語相似度分數差距圖

因爲相似度分數的差異所以造成在搜尋關鍵字辨識時，容易將華語語音節點與英語語音節點互相判別錯誤，所以爲了改善此判別錯誤的情形，我們在進行搜尋語音節點的時候，對每個英語模型以每個音框爲單位，乘上固定的偏差補償值，如下數學式(6)所示：

$$L_E^{'} = L_E \times 0.7896 \tag{6}$$

其中 $L_E^{'}$ 爲乘過偏差補償值後的英語模型相似度分數值。

2.5.2 無關詞垃圾模型的偏差補償

在本論文之中，華語模型與無關詞垃圾模型都是利用相同訓練語料所訓練而成，只是差別在聲母與韻母的狀態數多寡。在無關詞垃圾模型方面，我們只使用聲母、韻母各一個狀態數，所以訓練出來的模型會比較無鑑別性。而當無關詞垃圾模型與華語模型的相似度分數值進行競爭的時候，我們發現因爲無關詞垃圾模型的相似度分數值小於華語模型，兩者之間有一段差距，造成即使使用者說錯前置詞時，搜尋路徑還是會連帶影響到對於主詞的判斷，導致主詞辨識錯誤。所以我們對無關詞垃圾模型的相似度分數乘上一個靜態偏差補償值，如下數學式(7)、(8)所示：

$$L_{ini}^{'} = L_{ini}^{'} \times 0.824 \tag{7}$$

$$L_{fin}^{'} = L_{fin}^{'} \times 0.8205 \tag{8}$$

其中 L'_{ini} 與 L'_{fin} 爲經過偏差補償值後的聲母、韻母模型相似度分數值。

三、實驗語料

實驗所需的華語訓練語料爲「中華民國計算語言學會」所提供的 2000 人「國語語音資料庫」(Mandarin speech database Across Taiwan, 簡稱 MAT2000)[12]，此語料庫是透過公眾電話網路所錄製，取樣頻率和位元數分別爲 8KHz 和 16 位元。總共選取 84737 句語料當做華語模型訓練語料。而英語的訓練語料爲台灣腔英文資料庫(English Across Taiwan，EAT)[13]加上台灣腔式英語訓練語料(TWENG)。台灣腔英文資料庫分爲麥克風語料與電話語料，其中電話語料可細分爲固定式電話(PSTN)語料與行動電話(GSN)語料。取樣頻率和位元數分別爲 8kHz 和 16 位元，選取 90475 句訊練語料。台灣腔式英文少量語料，其取樣頻率和位元數也是爲 8kHz 和 16 位元，這部分我們選取 9535 句語料當成訓練語料。

實驗中測試語料爲 ME_Speech Corpus，是由 12 位男生與 12 位女生利用 10 個前置詞與 200 主詞所混合組成的關鍵詞辭彙錄製而成。總共有 4800 句語料，其中第一部份 2400 句語料爲包含了前置詞與主詞混合而成的關鍵詞，另外第二部份的 2400 句爲第一部份字句刪除前置只留下主詞的詞彙。

四、實驗分析

本系統中語音信號之取樣頻率爲 8kHz。每個輸入音框(frame)之特徵參數是由 12 維的「梅爾刻度式倒頻譜參數」(Mel-scale cepstrum)及對應之 12 維「倒頻譜差量參數」(delta Mel-scale cepstrum)加上 1 維「差量對數能量」(delta log energy)與 1 維「差差量對數能量」(delta delta log energy)所構成 26 維度的特徵參數，此參數並經過 RASTA 濾波器用於降低通道效應。辨認搜尋單元採用次音節模型，華語部分由 100 個 2 狀態的右相關聲母(initial)模型、38 個 4 狀態的韻母(final)模型與 1 個狀態的靜音(silence)模型組成。而英語部份由 134 個 2 狀態右文相關英語音素(Phone)模型與 1 個狀態的靜音(silence)模型組成。每一個狀態皆爲高斯混合模型，而協方差矩陣則是假設對角矩陣來代表，用於降低運算量。模型的高斯混合(mixture)數目則是根據訓練語料的多寡估算產生，其中華語聲母模型和韻母模型最多有 8 個高斯混合數，靜音模型則由 16 個高斯組成的高斯混合模型；而英語音素模型最多有 16 個高斯混合數，靜音模型則由 64 個高斯組成的高斯混合模型。辨認搜尋採用光束搜尋(beam search)模式之關鍵詞辨認，辨認核心皆經定點化並適合 PDA 級資源下運作。

4.1 改變搜尋機制

實驗中，我們提出兩種搜尋路徑加入系統的搜尋機制中。第一，在前置詞中加入忽略搜尋前置詞路徑，第二，在前置詞中加入無關詞垃圾模型路徑。在這兩種情況下，我們去測試系統對於主詞的辨識效能。以下我們分類爲四種搜尋機制，如下表一：

表一、搜尋機制

	英文模型偏差補償	忽略搜尋前置詞路徑	無關詞垃圾模型路徑
搜尋機制 1			
搜尋機制 2	✓		
搜尋機制 3	✓	✓	
搜尋機制 4	✓	✓	✓

4.2 辨識率比較

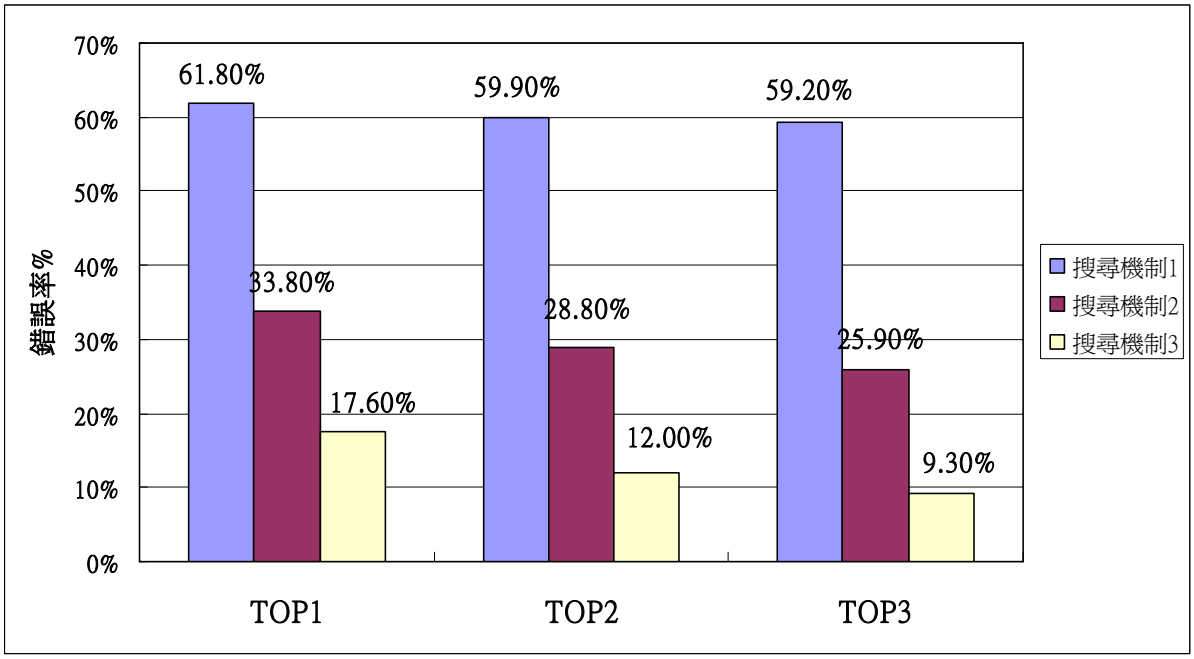

圖 4-1：加入偏差補償值與可忽略前置詞搜尋路徑的比較

首先我們先實驗測試語料對於有無英語模型偏差補償值與有無忽略搜尋前置詞路徑的效果，如上圖 4-1。有對英語模型相似度分數乘上偏差補償值的搜搜尋機制 2 與搜尋機制 3 都比搜尋機制 1 好上許多。在第一名的相對錯誤改善率方面，搜尋機制 2 比搜尋機制 1 改善了 45.2%，而搜尋機制 3 比搜尋機制 1 改善了 71.5%。而針對有加上忽略搜尋前置詞路徑，在第一名的相對錯誤改善率方面，搜尋機制 3 比搜尋機制 2 改善了 47.9%。

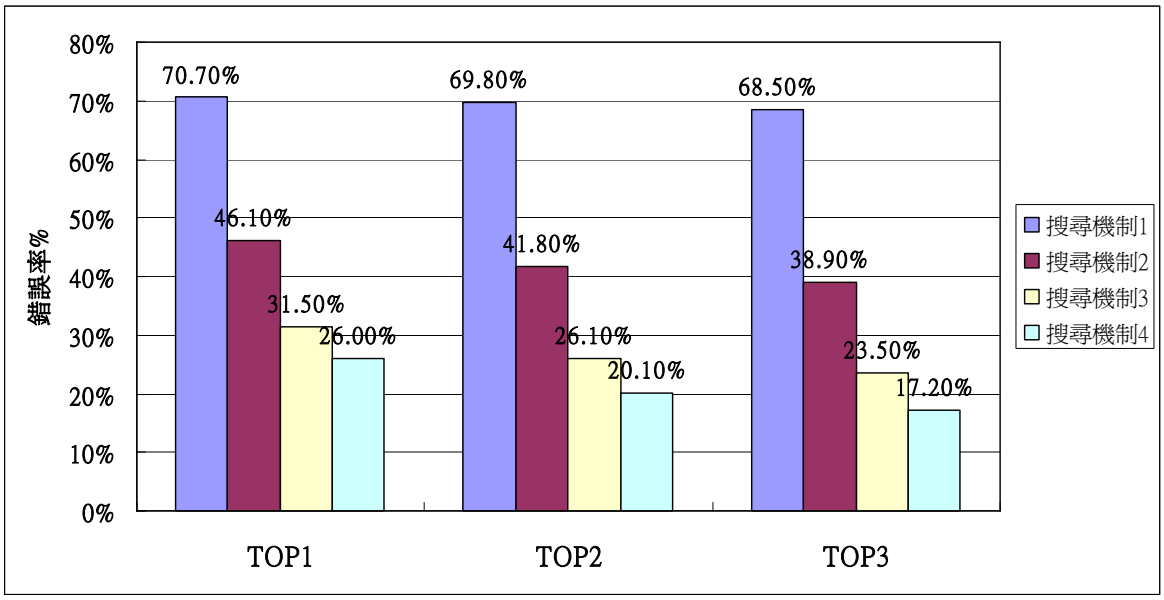

圖 4-2：加入偏差補償值、忽略前置詞搜尋路徑與無關詞垃圾模型路徑比較

接下來我們把 10 個前置詞中的 5 個華語前置詞全部刪掉，製造出語料中有說錯前置詞的情形，並且對於 4 個搜尋機制做實驗，如圖 4-2。對於有英文模型偏差補償值的搜尋機制 2 與 3，比較搜尋機制 1 在第一名的錯誤改善率爲 34.8%、55.4%，而加入忽略前置詞路徑後，搜尋機制 3 比搜尋機制 2 於第一名的錯誤率改善了 31.7%。再加上無關詞垃圾模型路徑之後，搜尋機制 4 比搜尋機制 3 在第一名的錯誤率改善了 17.5%。由此可知，當系統中搜尋機制越來越完整後，可針對使用者沒按規定說定義內前置詞的情形下，對主詞還有一定的辨識效能。

五、結論

在本論文中，我們先建立出一個詞典樹的架構，並且利用樹枝狀結構的概念，把辨識詞庫內每個辭彙的音節依序佈成樹狀。之後，我們把光束搜尋法應用於語音節點的搜尋與篩選，累積每個時間點的存活語音節點，進行相似值大小的排序。於實驗中，對測試語料所計算出的英語模型相似度分數乘上一個偏差補償值，可以拉近與華語模型產生出的相似度分數距離，並且大幅提升對於主詞辨識率。而當測試語料中含有無定義內前置詞的字句時，於前置詞搜尋機制中加入忽略搜尋前置詞路徑，讓測試語料於無前置詞的音框階段，語音節點可以進入此路徑，並且不破壞對於主詞語音節點的連結。

爲了讓系統的搜尋機制更加有彈性，我們於系統中加入無關詞垃圾模型路徑並且搭配無關詞垃圾模型的偏差補償值。改變定義內前置詞的數量，讓原本的測試語料有了說錯定義內前置詞的情形發生。系統在測試語料含有錯誤前置詞音框時，可以使所產生的錯誤節點路徑被無關詞垃圾模型路徑所吸收。經由實驗可得，有了此路徑對於說錯定義內前置詞的情形下，對於主詞的辨識率仍有一定的效果。

誌謝 本研究承中華電信研究所提供計畫費補助，謹誌謝忱。

參考文獻

[1] H. Hermansky, N. Morgan, "RASTA Processing of Speech," *IEEE Transactions SAP*, vol. 2, pp. 578-589, Oct 1994.

[2] C. L. Huang, C-H Wu, "Phone Set Generation Based on Acoustic Contextual Analysis for Multilingual Speech Recognition," *ICASSP*, vol. 4, pp. 1017-1020, 2007.

[3] C. L. Huang, C-H Wu, "Generation of Phonetic Units for Mixed-Language Speech Recognition Based on Acoustic and Contextual Analysis," *Computers, IEEE Transactions*, vol. 56, pp. 1225-1233, 2007.

[4] Po-Yi Shih, Jhing-Fa Wang, and Hsiao-Ping Lee, "Acoustic and Phoneme Modeling Based on Confusion Matrix for Ubiquitous Mixed-Language Speech Recognition," *SUTC*, pp. 500-506, June 2008.

[5] 遊山銳, 簡世傑等人, "中英文混雜關鍵詞萃取技術," *TEPS*, pp. 66-79, 2004.

[6] A. L. Higgins, R. E. Wohlford, "Keyword Recognition Using Template Concatenation," *ICASSP*, vol. 10, pp. 1233-1236, 1985.

[7] J. R. Rohilcek, W. Roukos, and H. Gish, "Continuous Hidden Markov Models for Speaker Independent Word Spotting," *ICASSP*, pp. 627-630, 1989.

[8] R. Rose, D. Paul, "A Hidden Markov Model Based Keyword Recognition System," *ICASSP*, vol. 1, pp. 129-132, 1990.

[9] J. Caminero, C. de la Torre, L. Villarrubia, C. Martin, and L. Hernandez, "On-line Garbage Modeling with Discriminant Analysis for Utterance Verification," *ICSLP*, vol. 4, pp. 2111-2114, Oct 1996.

[10] T. Svendsen, F. K. Soong, and H. Pumhagen, "Optimizing Baseforms for HMM-Base Speech Recognition," *Proceedings of EuroSpeech*, pp. 783-786, 1995.

[11] X. Lingyun, D. Limin, "Efficient Viterbi Beam Search Algorithm Using Dynamic Pruning," *ICOSP*, vol. 1, pp. 699-702, 2004.

[12] H. C. Wang, F. Seide, C. Y. Tseng, and L. S. Lee, "MAT2000 – Design, Collection, and Validation on a Mandarin 2000-speaker Telephone Speech Database," *ICSLP*, pp. 460-463, Beijing, China, 2000.

[13] http://www.aclclp.org.tw/doc/eat_brief.pdf.