行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　※
※　　　　關鍵詞辨認系統中獎勵函數之設計　　　　※
※　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：ⅴ個別型計畫　　□整合型計畫
計畫編號：NSC89 － 2213 － E － 009 － 190
執行期間： 89 年 8 月 1 日至 90 年 7 月 31 日

計畫主持人：王逸如

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學電信工程系

中　華　民　國　90 年　10 月　30 日

行政院國家科學委員會專題研究計畫成果報告

關鍵詞辨認系統中獎勵函數之設計

計畫編號：NSC89－2213－E－009－190
執行期限：89 年 8 月 1 日至 90 年 7 月 30 日
主持人：王逸如　國立交通大學電信工程系
計畫參與人員： 唐大任、曹登鈞、范世明

一、中文摘要

本計畫以別於傳統的關鍵詞系統使用大量語料訓練關鍵詞模型及填充語料模型的方式，針對導致音節辨認發生錯誤的音節混淆特性，依據辨認觀測機率差值分佈予以模式化而為音節混淆量測模型。並在相鄰音節辨認分數不相關之假設下，將音節混淆模型推廣至詞組間之混淆量模型。於是計劃中所建立之關鍵詞系統將連續語音經過 TOP-N 辨認器產生辨認音節候選者，從候選音節中串接出可能的候選關鍵詞及其所對應最佳填充詞(無語言模式限制)。再引用先前所建立的音節模型分別建立所有關鍵及其相對應之填充詞組的混淆量分佈模型，而在關鍵詞錯誤率要求的條件下，依據其詞組混淆組合的模型求取個別候選關鍵詞的加分補償量(獎勵函數)，將各候選關鍵詞辨認分數加上上述獎勵函數，以期關鍵詞之辨認分數可以高於其混淆詞。如此便能依據關鍵詞系統對關鍵詞組錯誤率的要求，對各關鍵詞候選者的計算出適當的辨認分數獎勵函數。

關鍵詞：混淆量測、獎勵函式、關鍵詞辨認

Abstract

In this project, a new keyword spotting system using reward function to discriminate keywords from the background fillers was proposed. It first defined a confusion measure between Mandarin syllable-pair as the HMM recognition score difference of the correct and alternate hypothesis. And, the Gaussian distribution was used to model the confusion measure of syllable-pair. Under the assumption of independence between adjacent syllables, the confusion measure of word-pair becomes the summation of corresponding syllable-pairs. Finally, the confusing measure of word-pair was used to decide the reward function in the keyword spotting system. In the proposed system, the Top-N syllable lattice was first found by HMM syllable recognizer, and the keyword candidates can be found from syllable lattice associate with the most probable filler model. And, the reward function of each keyword candidate can be decided from confusion measure between keyword and its associated filler. By defining the desired keyword error rate, the proper reward function can be found in the keyword spotting system.

Keywords: confusion measure, reward function, keyword spotting system

二、 緣由與目的

在經電話網路語音辨認應用系統，常是目的明確的簡單對話系統。因為在許多電話語音辨認的應用中，只要關鍵詞辨認正確即可。一個成功的關鍵詞辨認系統可以讓許多語音辨認之應用成真。

三、研究方法

本計劃中利用音節間之混淆音組特性來製作關鍵詞辨認器，在此對音節間混淆量測量之資料取得、模型建立做說明。並對混淆量量測模型建立時所面對之資料量不足的問題提出利用音節合併的解決方法。其次再將音節間混淆量測推廣至詞組之混淆量測量。最後將之應用在一個關鍵字辨認系統上。

1. 國語之混淆音

國語單音節性質使辨認過程產生許多的混淆音組（confusing set**）等。由於語者發音習慣的不同或是過於簡略的發音，使得語者無法完整的發出正確的字音。一般聲母的辨認率又低於韻母，於是當混淆音發生時，多數的狀況**

為音節之韻母相同，而聲母部分發生混淆；而聲母部分正確，韻母卻發生混淆者常是因為介音或字尾鼻音。

2. N-best 音節辨認器

要在連續語音資料中觀察國語之混淆音首先我們使用一個 411 音節 TOP-N syllable lattice HMM 辨認器以便觀察每一音節之 TOP-N 輸出[1,2]。在 HMM 音節辨認結果中，若假設辨認發生錯誤時，插入及刪除發生情形遠較取代錯誤為低，則在此假設下增加 TOP-N syllable lattice 辨認候選音節個數時，正確音節出現在候選音節中的機率便會上升，此一趨勢將可以明顯改善音節辨認錯誤的問題，也更助於系統觀測混淆音節的出現及分佈。

計劃中使用之訓練語料是採用MAT2000電話語音語料庫中的DB4及DB5兩部分作為訓練語料。在已知輸入音節(音碼為i)與其TOP-N syllable lattice中其它音節（音碼為j）間之混淆量測量(confusion measure)，其可定義為：

$$C_s(S_j|S_i)=(\log(S_i|x,\Lambda_i))/L_i-\log(S_j|x,\Lambda_j))/L_j;\ \forall j\neq i$$

其中 Λ_j，L_i 分別為 i 音之 HMM 辨認模型及音長，Si 為已知音節，$\log(S_j|x,\Lambda_i)$ 為正確結果之音節辨認對數機率，因為使用[1]中之方法找尋出 TOP-N 候選音節會有長度不同的特性所以對音節長度做正規化。則此混淆量測即單音（S）在已知被辨認為 i 音及 j 音之單位音框對數觀測機率差。此混淆量測為一隨機變數，於是可由圖 1.求取音節間之混淆量。並將混淆量測假設為以高斯分佈則只需平均值及變異數兩個參數即可表示兩音節間的辨認錯誤之機率分佈狀況。

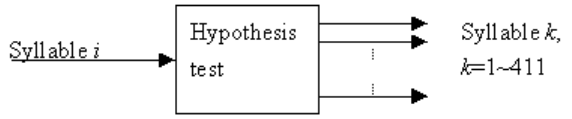

圖 1. 音節混淆量建立架構。

圖 2.為一混淆量測之示意圖。如果在辨認器某一音節 i 之辨認分數加上一個獎勵值將可降低該音節之辨認錯誤機率，當然也會增加假警報率 $P(S_j|S_j)$。若依據音節間混淆量測之分佈對辨認結果 S_i 作加分補償，我們可以由圖 2. 中找到預期錯誤率(陰影部分)與所欲加上之獎勵值之間的關係，這就是本計劃中關鍵詞辨認系統中獎勵函數設計之依據了。

若針對國語 411 音節建立與所有其他音節之混淆音組之混淆量測模型，則須建立 411x411 個機率模式，因訓練語料量不足會造成的模型可靠度降低，或無法建立模型的嚴重問題。於是在此對於語料量不足的混淆音組，將以音節間擁有相同語音性質者作合併（merge）成音群的動作，利用音節與音群的混淆量測模型以 button up 的方式來建立建立出性質相似的混淆音群，其中資料量不足音節具有相同混淆性質者歸屬為同一音群，而改以音節對音群中所有音節求取混淆機率分佈模型。下式是以各音節對其他音節之混淆向量，判斷音節間是否因性質相近並符合條件而可合併，下式就是兩音節或音群性質相似度[4]：

$$COS^{-}=\frac{x_i\cdot x_j}{\left\|x_i\right\|\times\left\|x_j\right\|},\quad i,j\in 411\ \text{syllables or class}$$

其中 x_i 向量中的第 j 的 component 之值為為 i 音節被辨認成及 j 音節之機率；$P(S_j|S_i)$。而產生新群所屬 row vector 可由合併之兩音節或音群之辨認結果向量 x_i、x_j 相加獲得。

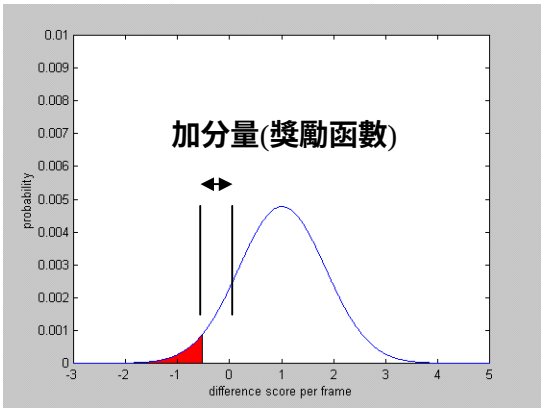

圖 2. 加分後的錯誤機率。

3. 由音節詞混淆量測獲得組之混淆量測量

接著在計劃中將把 411 音節混淆量測及利用混淆量測來預測正確音節之加分量以獲得預期辨認率之方法推廣至詞組辨認。在圖 4 中，對單詞輸入經由 411 音節 TOP-N 辨認器，詞組辨認結果及辨認分數可由 syllable lattice 辨認結果搜尋而得[3]。

在連續語音中前後音節間之混淆量測量獨立(independent)的假設下(即不考慮音節在不同詞中出現之連音效應），我們可以將混淆詞組間的混淆量測量直表示成兩詞之相對應音節間混淆量測之和，則單詞混淆量可表示為期相對各音節間混淆量測之和：

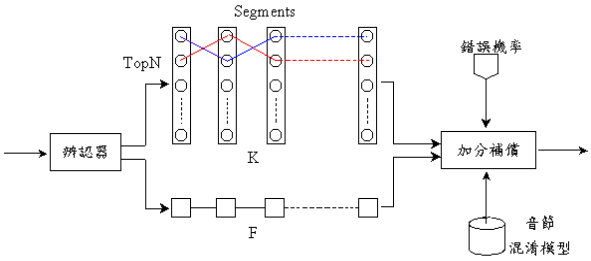

圖 4. 關鍵詞混淆量補償架構圖。

$$C_w(F|K)=\sum_{i=1}^{N}(C_s(F_i|K_i)\cdot L_i)-w\cdot(\log(P(F_1))-\sum_{i=1}^{N}P(F_i|F_{i-1}))$$
$$\cong\sum_{i}^{N}(\log P(S_{K_i}|\Lambda_{K_i})-\log P(S_{F_i}|\Lambda_{K_i}))$$

其中 $K=(K_1, .., K_N)$, $F=(F_1, \ldots,F_N)$ 為兩個混淆詞，K_i 及 F_i 為兩詞中各組成音節。

由於 $C_w(K|F)$ 之機率分佈乃是由彼此獨立的混淆音節混淆量測機率分佈所組合而成，因為單詞混淆量模型將亦為 Gaussian distribution，而它的 mean 及 variance 可直接由音節混淆量測量模型的 mean 及 variance 求得。而這也是我們選擇以高斯模型模擬音節混淆量分佈之原因。下式為詞組混淆量模型之 mean 及 variance：

$$\mu_{C_w(K,F)}\cong\sum_{k=1}^{N}(\mu_{C_s(K_k,F_k)}\cdot L_k)$$
$$\sigma^2_{C_w(K,F)}=\sum_{k=1}^{N}(\sigma^2_{C_s(K_k,F_k)}\cdot L_k^2)$$

接著我們利用 MAT2000 的語料中之單詞語料來驗證上述由音節詞混淆量測獲得詞組之混淆量測量之方法是否可行，在圖 5.的是『ㄕㄣ ㄧㄣ』與其易混淆音『ㄕㄥ ㄧㄥ』間實際混淆量測分佈與使用音節詞混淆量測獲得詞組之混淆量測模型，由圖中可發現，前述方式所建立之單詞混淆量模型，可合理的模擬出實際的詞組混淆量分佈。

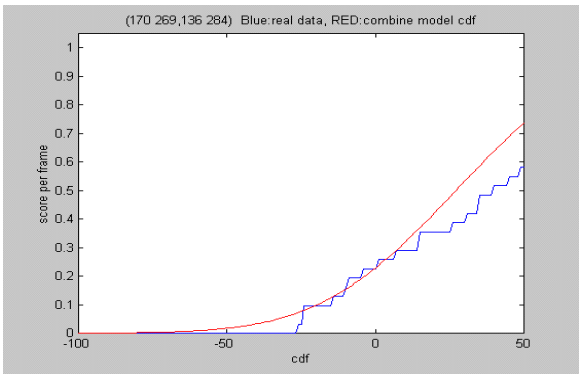

圖 5. 單詞混淆量分佈之範例(實際分佈與混淆量模型)。

於是在上述詞組混淆量測模型下，依詞組模型預測到達到特定詞組辨認率時詞組所需加上的獎勵函數，對實際單詞辨認分數差予以加分補償，圖 6.便是預測辨認率語言計辨認率與加分補償之關連。可以看出其誤差非常小。

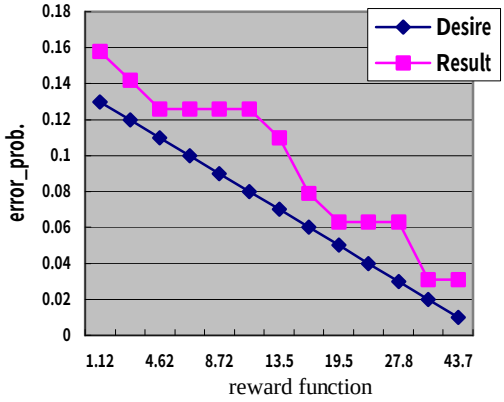

圖 6. 獎勵函數與詞錯誤率之關係圖。

4. 關鍵詞辨認系統之建立

關鍵詞辨認系統其實只是一個特殊的連續詞辨認系統，我們可以在 TOP-N syllable lattice 辨認結果中串接出候選關鍵詞及填充詞，然後再對候選關鍵詞之辨認分數加上獎勵值補償以判斷該候選關鍵詞是否出現。所以在 TOP-N syllable lattice 辨認器輸出端的結果中，我們希望能提高關鍵詞在 TOP-N syllable lattice 的 inclusion rate（涵蓋率），而經加分補償使候選關鍵詞之辨認分數大於其混淆詞，而擷取出正確的關鍵詞。當增加 N 值大小時，可以明顯的發現此趨勢與冀求的結果相符。但由於候選音節的個數增加勢必使得候選關鍵詞的個數也增加，而使系統關鍵詞辨認錯誤的機率上升。此外，放寬 syllable TOP-N 限制後，由於任意 syllable 都至少可以會在第 N+1 名的候選者出現，關鍵詞的涵蓋率是 100%。於是在此有些限制，即關鍵詞中最多只能有一個音節落入第 N+1 個候選者，這樣可以避免此 TOP-N syllables 之限制過於寬鬆，使候選關鍵詞詞大量增加而導致辨認率下降。

5. 關鍵詞辨認系統之實驗結果

計劃中電話語音的關鍵詞辨認實驗是採用 MAT2500 語料庫為訓練語料，建立了音節混淆量測模型。測試語料使用工研院關鍵詞辨認系統之語料庫，共 3715 句，1659 個關鍵詞(它們是由工研院員工電話號碼查詢系統之錄音而來)。接著我們做了下面的實驗：

實驗 A：

在不考慮假警報的出現下，由預期之關鍵詞辨認率可以由詞組混淆量模型算出所須之獎勵函數加分補償量，其關鍵詞辨認率與獎勵函數加分補償量之關係如圖 7 所示。加分量越高當然關鍵詞的辨認率會越高，當然距離預期的辨認率會有一段距離，但趨勢是對的，這是因為混淆量測一定會有誤差。

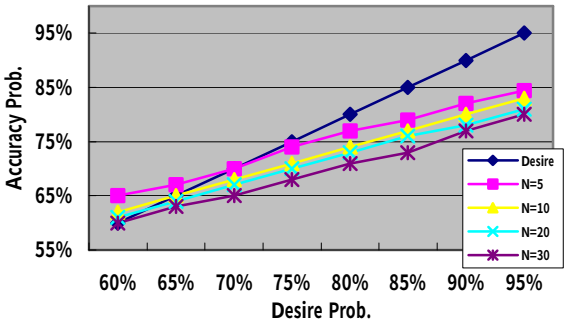

圖 7. 獎勵函數與關鍵詞辨認率分佈圖。

實驗 B：

在此實驗中，我們使用上述關鍵詞測試語料中的 350 句(包含 100 個關鍵詞)來做實驗，在考慮假警報的發生，觀察加分補償對於關鍵詞辨認率及假警報率之影響(見圖 8 及圖 9)。

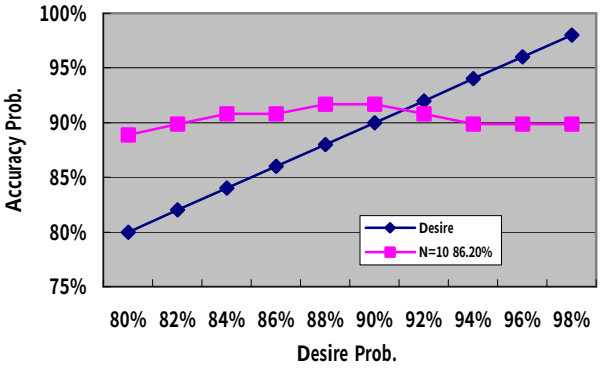

圖 8. 系統獎勵函數與關鍵詞辨認率分佈圖(考慮多關鍵詞情況)。

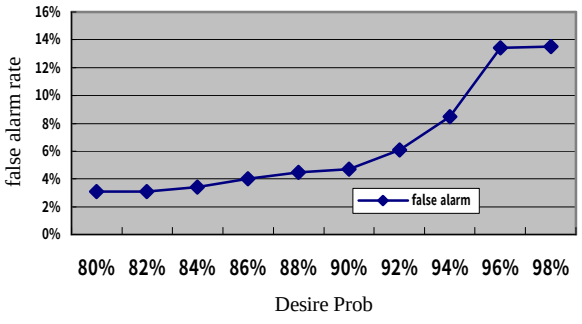

圖 9. 系統獎勵函數與假警報分佈圖。

實驗 A 中，實際辨認正確率確實有因為系統要求錯誤率下降而上升的趨勢。結果不十分符合的原因，是因為辨認錯誤中，模型預估誤差與我們所無法補償的音節插入及刪除而造成的錯誤(411 音節辨認器中音節之插入及刪除率約 3-4%)。而實驗 B 中，此測試環境下，單純不經由加分補償時，關鍵詞的正確率達 86%。而當我們對候選詞依其單詞混淆模型而加上補償量時，可以發現當系統要求正確率上升，及補償量值變大的情形下，辨認率會由 86%依照系統要求的趨勢到達 91%。但是當系統所要求的正確率過高時，其便會面臨到 false alarm 的問題反而使得辨認率下降。圖 9.中，當我們因為系統要求錯誤率下降而增加補償量時，會造成假警報的上升。不含關鍵詞之語料，經過辨認搜尋後所產生的候選關鍵詞，可能因加分補償被視為合理的關鍵詞而被擷取。所以補償量過度時，也會導致假警報的上升，使不含關鍵詞的語句因補償而產生假警報，或是句子中非關鍵詞部分反而因加上補償量而成為一個錯誤的關鍵詞。但是在 detection 問題中 miss 與 false alarm 本來就是不可兼得的。在上述系統中對於補償量調整越大，false alarm 也會加大，甚至讓 miss rate 提昇。

四、計畫成果自評

在本計劃中(1)建立了一套可靠的 411 音節混淆量測，(2)使用音節混淆量測來建立詞組間之混淆量測，(3)建立一套關鍵詞辨認系統利用詞組間之混淆量測來預估關鍵詞辨認系統中之關鍵詞獎勵函數。但在計劃中發現 selective training 在訓練語料足夠時對辨認率並無任何改善，訓練 411 音節混淆量測之語料與關鍵詞辨認系統所使用之語料之錄音環境差異會使系統效能降低，所以建立一套更 robust 的音節辨認器還是需要繼續研究的課題。

五、參考文獻

[1] Eng-Fong Huang and Hsiao-Chuan Wang, "An Efficient Algorithm for Syllable Hypothesization in Continuous Mandarin Speech Recognition", IEEE Trans., On Speech and Audio Processing, Vol2, No3, P.446-P.449, 1994.

[2] 涂家章，"使用 MAT2000 語料庫之中文語音辨認"，國立交通大學碩士論文，民國 89 年六月。

[3] 陳志豪，"利用 411 音填充模型之關鍵詞辨認系統"，國立交通大學碩士論文，民國 87 年六月。

[4] 張元貞、李琳山等，"國語語音辨認中詞群語言模型之分群方法與應用"，pp. 17-34, ROCLING-VII, 1994.