台語關鍵詞辨識之實作與比較

Implementation and Comparison of Keyword Spotting for Taiwanese

王崇喆　Chung-Che Wang
國立清華大學資訊工程學系
Department of Computer Science
National Tsing Hua University
geniusturtle@mirlab.org

周哲玄　Che-Hsuan Chou
國立清華大學資訊工程學系
Department of Computer Science
National Tsing Hua University
stephen.chou@mirlab.org

陳亮宇　Liang-Yu Chen
國立清華大學資訊系統與應用研究所
Institute of Information Systems and Applications
National Tsing Hua University
davidson833@mirlab.org

李毓哲　Yu-Jhe Li
國立清華大學資訊工程學系
Department of Computer Science
National Tsing Hua University
liyujhe@mirlab.org

張智星　Jyh-Shing Roger Jang
國立臺灣大學資訊工程學系
Department of Computer Science and Information Engineering
National Taiwan University
jang@mirlab.org

胡訓誠 Hsun-Cheng Hu，林世鵬 Shih-Peng Lin，黃友鍊 You-Lian Huang
財團法人資訊工業策進會 智慧網通系統研究所
{johnhu, shihpeng, uln}@iii.org.tw

摘要

本論文主要探討結合語音評分與音高走勢分類器辨識，來實做台語關鍵詞辨識系統，以改進其辨識率。第一階段先分別利用了不同方法來實做台語關鍵詞辨識系統，第二階段

使用語音評分與音高走勢分類器進行驗證以改善辨識效能。首先，在第一階段使用了隱藏式馬可夫模型（hidden Markov model），以及音素匹配法（phone mismatch）。而第二階段則是將此兩種方法辨識出來的候選關鍵詞，進行語音評分和音高走勢分類器驗證，將此兩種方法分別設立門檻值，利用決策樹法進行驗證。實驗結果顯示，在兩種基礎方法後，加入語音評分做為驗證的相等錯誤率（equal error rate, EER），分別約可下降 20% 和 5%；進一步加入音高走勢分類器驗證後，約可再下降 1%，因此語音評分和音高走勢分類器對於台語關鍵詞系統的驗證是很有幫助的。

Abstract

This paper focuses on improving in the performance of a Taiwanese keyword spotting system by integrating speech assessment and pitch contour classification. In the first part of this research, we use different methods to implement a Taiwanese keyword spotting system. In second part, we improve the system by validation using speech assessment and pitch contour classification. Two methods are adopted in the first part to implement the keyword spotting system: hidden Markov model and phone mismatching method. We then perform speech assessment and pitch contour classification to validate the candidate keywords selected by these two methods to refine the results. A threshold is used for a decision tree to make the final decision. Experimental results shows that the equal error rates (ERRs) reduce about 20% and 5% after being incorporated speech assessment validation. After being incorporated with pitch contour classification, ERRs further reduce about 1%. This concludes that the validation technique using speech assessment and pitch contour classification can improve the performance of Taiwanese keyword spotting.

關鍵詞：關鍵詞辨識、隱藏式馬可夫模型、懲罰矩陣

Keywords: Keywords spotting, hidden Markov model, penalty matrix

一、緒論

本論文主要利用隱藏式馬可夫模型法與音素匹配法，實做台語關鍵詞辨識系統，再利用音高特徵與語音評分，以提升系統的辨識率。使用情境方面，我們針對 3C 產品的控制，而控制這些 3C 產品時，都是短短的指令，例如：開冷氣，開冰箱等等，故本論文針對短字詞的關鍵詞進行探討。

本論文的研究方向為實做不同方式的台語的關鍵詞辨識系統，並比較其優劣，再和語音評分和『音高走勢分類器』合併。在第一階段時我們實做的系統為一個移植性高，且可自由變換關鍵詞庫之系統，然而缺點是會犧牲辨識率，故在第二階段加入了關鍵詞的驗證來達到較好的成效。在第二階段我們以一種信心測量的方法（confidence measure, CM）進行語音評分，評判所擷取的關鍵詞語音是否足夠接近標準關鍵詞語音。而加入『音高走勢分類器』之主要原因為，使用一般關鍵字辨識系統較少使用到之語音特徵，例如音高和音量，來針對聲調做進一步的確認。經由這兩項改進，輔助原本的關鍵詞辨識系統達到較好的辨識效果。

本論文其餘部分之概要如下：第二章為相關研究；第三章為論文方法；第四章為實驗結果；第五章為結論與未來展望。

二、相關研究

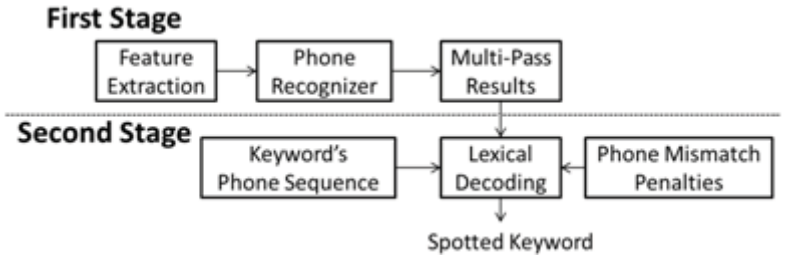
（一） 關鍵詞辨識技術

使用隱藏式馬可夫模型來做關鍵詞辨識，傳統上為針對每一個關鍵詞都個別訓練一個關鍵詞的模型。其優點為辨識率較高，但缺點為移植性不高，並且不易更換關鍵詞，一旦遇到這種情況，必須重新搜集語料和訓練關鍵詞辨識模型，故本論文所採用的方法為使用右相關雙連音素的串接形成關鍵詞模型，這樣不論在更換領域或是增加關鍵詞上面會變的較有彈性，但其缺點為辨識率較低，故必須經由更進一步的關鍵詞驗證來提升整個關鍵詞辨識系統的效果。而在做關鍵詞辨識系統時，除了關鍵詞的模型外，我們必須建立一個非關鍵詞的模型來辨識非關鍵詞的部分，我們稱之為填充模型（filler-model），本論文所採用的方法為使用聲韻母模型來當做填充模型。而在辨識過程中，通常會針對聲韻母模型加入若干的懲罰值，其原因是避免在關鍵詞辨識中其進入聲韻母的機率太高，導致關鍵詞辨識容易產生錯誤，故在實驗中我們會針對聲韻母模型做些許的調整。

（二） 填充模型選取與訓練

在本論文中我們所採用的填充模型為聲母 18 個、韻母 61 個，在進行訓練之前會先把聲母標為 fi_init，韻母標為 fi_final，之後再進行訓練程序。

（三） 利用音素匹配法實做兩階段關鍵詞辨識



圖一、音素匹配懲罰矩陣法方塊圖

利用音素匹配法實做兩階段關鍵詞辨識[1]過程如下，其第一階段仍為訓練一個基礎模型，之後將測試語料進行自由音素的解碼，其結果可以採取最佳的數個。第二階段則將這些音素序列，與關鍵詞的音素序列，利用動態規劃進行比對，其比對的方法為經由音素匹配懲罰矩陣與關鍵詞辨識，制定門檻值來找出是否含有關鍵詞。音素匹配懲罰矩陣（Phone Mismatch Penalty Matrix）主要為利用訓練語料中之發展語料（development）做為內部測試，得到一種音素與音素之間相對關係的矩陣，而音素匹配懲罰矩陣主要是利用三種會產生的錯誤進行分析，分別是替換（substitution）、插入（insertion）、刪除（deletion），對於每組音素間分別對於這三種錯誤給予不同的懲罰，乃為此矩陣之精神所在，細節可參考[1]。

1. 懲罰矩陣法

替換的懲罰矩陣公式如下，其中$\emptyset_i$，$\emptyset_j$代表不同音素，Pr〔·〕代表事件的機率，而II[α]為一

個 indicator 方程式，α_{sub}則是根據實驗與經驗而調整。取 log 原因是為了將懲罰值正規化至一個合理的數值。關於插入與刪除的懲罰矩陣公式與實做，可參考[17]。

$$PM_{sub}\ (\emptyset_i,\ \emptyset_j)=\begin{cases}-\log \Pr\left[P(x_k|\emptyset_j) < P(x_k|\emptyset_i)|p_k=\emptyset_j\right]+\alpha_{sub}\ , & \emptyset_i \neq \emptyset_j \\ 0\ , & \emptyset_i = \emptyset_j\end{cases}$$

$$\Pr\ \left[P(x_k|\emptyset_j) < P(x_k|\emptyset_i)|p_k=\emptyset_j\right] \cong \frac{\sum_{k=1}^{Np} \mathrm{II}[p_k=\emptyset_j, P(x_k|\emptyset_j)<P(x_k|\emptyset_i)]}{\sum_{k=1}^{Np} \mathrm{II}[p_k=\emptyset_j]}$$

2. 混淆矩陣法

混淆矩陣為發展語料經由自由音素解碼後的結果，在[2]中三種懲罰矩陣分別定義如下：

$$CM_{sub}(\emptyset_i,\emptyset_j)=log\{S(i,i)/S(i,j)\}\quad ,\quad CM_{ins}(\emptyset_i,\emptyset_j)=I\quad ,\quad CM_{del}(\emptyset_i,\emptyset_j)=D$$

概念上為利用辨識對的音素數目，除以辨識錯誤的音素數目來做為懲罰基準，$\emptyset_i$代表字典中第 i 個音素，I 和 D 為常數，可利用實驗調整。細節可參考[2]。

3. 距離矩陣法

在距離矩陣法中三種懲罰矩陣分別定義如下：

$$LD_{sub}(\emptyset_i,\emptyset_j)=\begin{cases}1, \emptyset_i \neq \emptyset_j\\ 0, \emptyset_i = \emptyset_j\end{cases}\quad ,\quad LD_{ins}(\emptyset_i,\emptyset_j)=1\quad ,\quad LD_{del}(\emptyset_i,\emptyset_j)=1$$

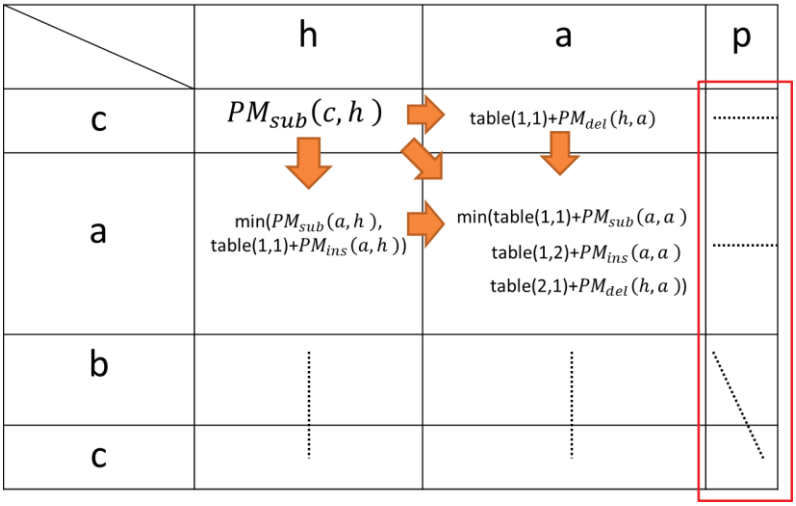
此方法為音素與音素間最基本的相互關係，而利用這種方法可以得知關鍵詞語句與測試語句的最小距離。

4. 音素匹配實做關鍵詞辨識系統

當得到三種不同的懲罰矩陣後，我們可以經由動態規劃，來得知測試音檔是否含有關鍵詞。假設$Q^{(n)}=(q_1^{(n)},q_2^{(n)},\cdots,q_{N_Q^{(n)}}^{(n)})$為一句測試音檔經由自由音素解碼的輸出結果，P = $(p_1\ ,\ p_2\ ,\ \cdots\ ,\ p_{Np})$為關鍵詞的單音素序列，$C_{i,j}^{(n)}$為從 $\left(q_i^{(n)},p_j\right)$開始計算每個點的最佳距離，而下面為關鍵詞系統的 DP 遞迴式：

$$C_{i,j}^{(n)} = \begin{cases} PM_{sub}\left(q_i^{(n)}, p_j\right), & i = j = 1 \\ C_{i,j-1}^{(n)} + PM_{del}(p_{j-1}, p_j), & i = 1, j \neq 1 \\ \min[\, PM_{sub}\left(p_{j-1}, p_j\right), C_{i-1,j}^{(n)} + PM_{ins}(q_i^{(n)}, p_j)] \,, & i \neq 1, j = 1 \\ \min\,[C_{i-1,j-1}^{(n)} + PM_{sub}\left(q_i^{(n)}, p_j\right), & \\ \quad C_{i-1,j}^{(n)} + PM_{ins}(q_i^{(n)}, p_j) & \\ \quad\quad C_{i,j-1}^{(n)} + PM_{del}(p_{i-1}, p_j)] \,, & otherwise \end{cases}$$

其中 $1 \leq i \leq N_Q^{(n)}$ and $1 \leq j \leq N_p$。我們可用以下範例圖來表示上述的方法：



圖二、音素匹配法示意圖

由於我們不知道測試音檔中關鍵詞開始的位置，故在 $i \neq 1, j = 1$ 時我們使用 $\min[\, PM_{sub}\left(p_{j-1}, p_j\right), C_{i-1,j}^{(n)} + PM_{ins}(q_i^{(n)}, p_j)]$ 而非 $C_{i-1,j}^{(n)} + PM_{ins}(q_i^{(n)}, p_j)$，來猜測關鍵詞可能的起始位置，最後我們將選取最後一行中，正規化後的最小值，做為一個測試音檔對於一個關鍵詞的懲罰值，其公式如下：

$$\mathrm{D}\,(\mathrm{P}\,, Q^{(n)}) \;=\; \min_{1 \leq i \leq N_Q^{(n)}} (C_{i,N_p}^{(n)} / l_{i,N_p}^{(n)})$$

其中各個小標意義如下：

P：關鍵詞音素字串　　Q：測試音檔音素字串

(n)：測試音檔經由自由音素解碼後第 n 名結果

C：動態規劃所填之表格　　i：第 i 個 row

N_P：最後一個 column　　l_{i,N_p}：關鍵詞音素字串長度

而當得到此懲罰值和位置後，我們利用回溯法（backtracking），找出測試音檔所對應到關鍵詞的位置。最後辨認是否有關鍵詞定義如下：

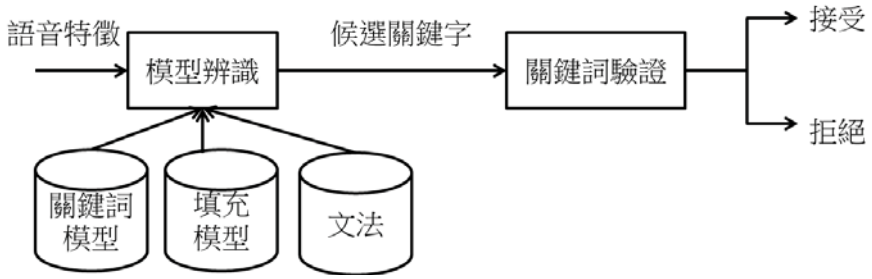
$$\min_{1\le n\le N} D(P, Q^{(n)}) \le \Upsilon$$

其中Υ為門檻值，若經由辨識得到小於或等於Υ，則表示有關鍵詞。

三、論文方法

（一） 關鍵詞驗證技術

在實做關鍵詞系統時，會分成關鍵詞的擷取和關鍵詞的驗證，以下為示意圖：

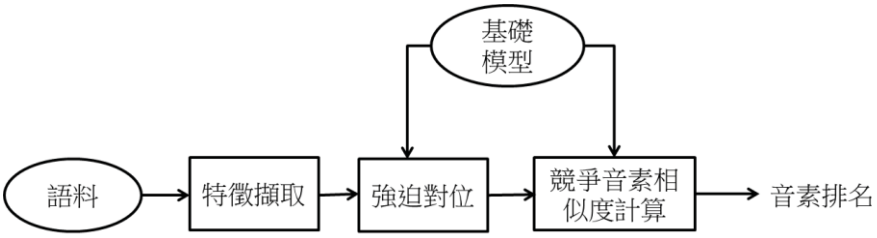


圖三、關鍵詞系統與驗證方塊圖

在實做關鍵詞辨識系統時，常常會因為模型的訓練較為簡單普遍，或是強制對位不準確的問題等等，導致關鍵詞系統辨識率不夠好，故會在第一階段關鍵詞擷取過後，加入關鍵詞驗證，來增加整個關鍵詞系統的辨識率，下面將介紹本論文所使用的驗證技術與如何結合原有的關鍵詞辨識系統。

（二） 排名評分 （rank ratio score）

排名評分(rank ratio score)[12][13]為一種對於關鍵詞的信心測量(confidence measure)，此評分方式將會定義一正確音素與其競爭音素，藉由維特比解碼計算出對數似然率進行排名，再使用此排名計算出排名分數，音素排名流程如下圖所示



圖四、語音評分計算流程

我們將待評分音素的右相關雙連音素，做為其競爭音素。接著將待評分音素與競爭音素，依據對數似然率排名，再由公式 (2)、 (3)計算出排名評分。

$$Rank\ Ratio = \frac{\text{rank}-1}{\text{p}} \qquad (2)$$

$$\text{Rank Ratio Score} = \frac{100}{1+\left(\frac{Rank\ Ratio}{a}\right)^{b}} \quad (3)$$

其中 rank 代表排名，對數似然率最高者為第一名，依此類推；p 代表競爭因素個數；而利用排名比所產生的分數是經由 a 和 b 進行調整，a 和 b 對於每一個右相關雙連音素模型會有不同的值，其主因為針對每個模型產生一個最佳的評分系統。

（三） 音高走勢分類器

音高與音量相較於梅爾倒頻係數而言，是較少被選為語音特徵的，而近年來在中文關鍵詞系統中，已有人加入聲調結合關鍵詞辨識系統來改進辨識率[11]，而在台語方面則無此方面的嘗試，故本論文多加入了音高與音量兩種特徵，來輔助原本的關鍵詞辨識系統。在台語聲調[14]方面，基本分為七種，為調 1～調 7。調 1～調 5 為非入聲字調，調 6、調 7 為入聲字調（即音節結尾為-p -t -k -h），而額外的兩種為調 8 和調 9，為變調而來的，並非單獨的字調。以下表說明台語聲調：

表一、台語聲調說明表

漢字	衫	鼻	褲	黨	人	直	血
基本頻率							
調號	1	2	3	4	5	6	7
調值	55	33	21	51	24	32	44
ForPA 拼音	sann1	pinn2	ko3	dong4	lang5	dit6	hueh7

經由上表，我們可以將台語聲調分為四個特性，平緩、上升、下降及短促急停（入聲字），本論文將針對這四種類別分別訓練一種模型，並且輔助關鍵詞系統的驗證。

1. 特徵擷取與分類器訓練

對於音高追蹤，我們採用了 UPDUDP[9]的方法。UPDUDP[9]是利用動態規劃（Dynamic programming, DP）方式，找出整句不中斷音高追蹤方法。作法是首先找出整句音檔的 AMDF[10] (average magnitude difference function)矩陣，並基於該矩陣進行 DP 之演算，取出連續不中斷的基頻軌跡曲線，使原本跳動幅度大的聲母位置的音高資訊更加完整。

音量特徵主要是作為發出不同聲調時，其音量相對大小的參考指標，而在利用時會減去平均值，觀察每個音框間的相對性。計算公式上，採用音框的每個取樣點其絕對值總和。

經由上述計算取得該音節的音高與音量向量之後，我們參考[8]的方法，取出代表該音節的特徵向量，以進行分類器的訓練。本論文利用高斯混合模型（Gaussian mixture model, GMM）來實作『音高走勢分類器』[15]。

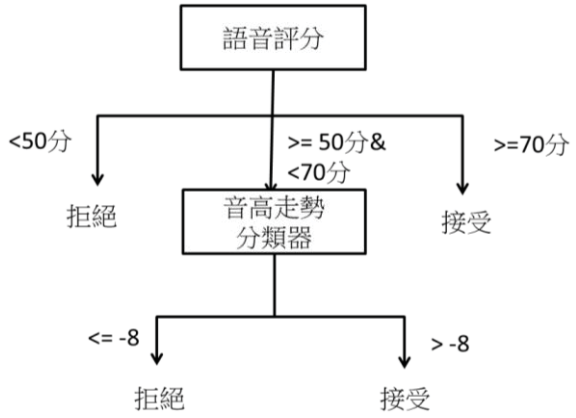
2. 音高走勢分類器驗證方法

假設我們的關鍵詞為流血（ForPA 拼音為 lau2_heh7），發音為類別 1 與類別 4，而測試

音檔經由第一階段的關鍵詞辨識可以得到候選關鍵詞，我們再將候選關鍵詞經由 GMM 得到每一個音節對於四個類別的 log-likelihood 值，之後再選取其類別 1 和類別 4 的組合相加並正規化至一個音節，此為候選關鍵詞經過『音高走勢分類器』後的分數，我們之後將利用此分數做為關鍵詞驗證的方法之一。

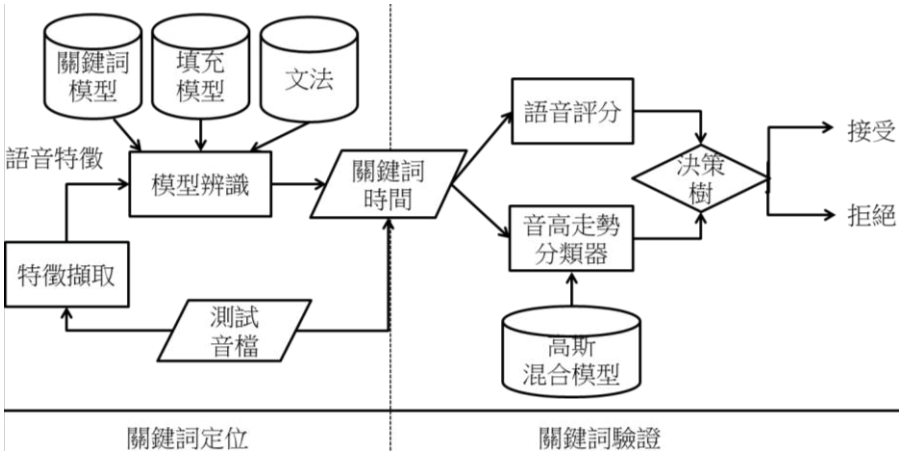
（四） 決策樹法結合語音評分與音高走勢分類器之關鍵詞辨識系統

當我們得到關鍵詞的語音評分與『音高走勢分類器』的分數後，我們將用決策樹（decision tree）決定測試音檔中是否含有關鍵詞，而因為在驗證關鍵詞時，使用語音評分比使用『音高走勢分類器』效果較佳，故在使用決策樹時採信語音評分較多，『音高走勢分類器』做為輔助來建立決策樹，以下為使用決策樹的示意圖：



圖五、關鍵詞驗證之決策樹示意圖

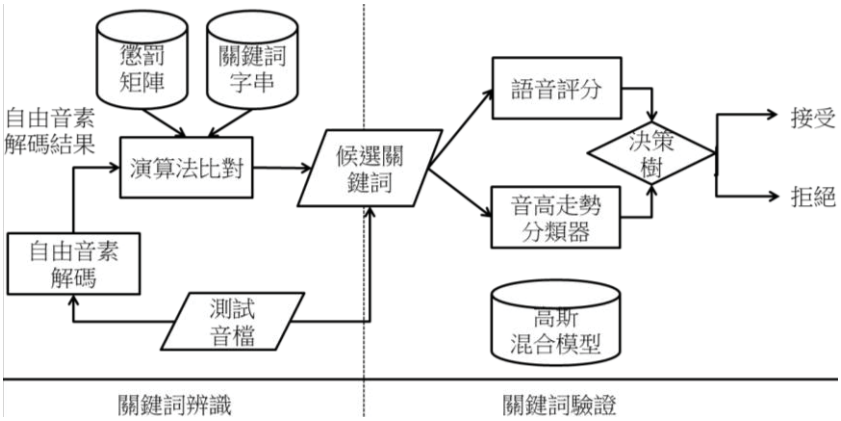
第一個選定的問題為「關鍵詞的語音評分是否大於等於 70 分或小於 50 分？」，若滿足前者則我們將接受此關鍵詞，否則拒絕此關鍵詞，而第二個問題為「『音高走勢分類器』的 log-likelihood 值是否大於-8？」。而結合前面兩種關鍵詞辨識系統——隱藏式馬可夫模型法和懲罰矩陣法，其示意圖如下：



圖六、HMM 加入關鍵詞驗證之系統方塊圖

首先我們先將第一階段關鍵詞辨識所辨識出的可能關鍵詞進行切音，得到候選關鍵詞的音節切音，之後進行語音評分和『音高走勢分類器』辨識，我們可以發現大部分的候選

關鍵詞不會高於 70 分，而正確的關鍵詞往往會高於 70 分，但仍然有可能會有例外，故又多加入『音高走勢分類器』進行驗證。



圖七、音素匹配法加入關鍵詞驗證之系統方塊圖

利用音素匹配法進行關鍵詞辨識，和隱藏式馬可夫模型大致相同，不同地方為第一階段音素匹配法是利用測試檔案的音素序列，與關鍵詞的音素序列進行比對而得，之後再利用回溯法得知關鍵詞的位置。得知候選關鍵詞的位置後我們一樣利用語音評分與『音高走勢分類器』來做關鍵詞的驗證，之後的方法與上一小節相同。

四、實驗結果與分析

（一） 訓練語料簡介

語料來源為論文[7]所蒐集的訓練語料，並把訓練語料分成訓練與發展（development）語料，分為兩種主要為實做不同的關鍵詞系統所需，下面為訓練語料數據，由數據可以知道男女生人數比例相當，並且在句數方面也相當平衡，而語料的標註是採用台語的 ForPA 拼音。

表二、訓練語料資訊

	訓練語料	發展語料	測試語料
語料名稱	TW01 和 TW02 訓練語料	TW01 和 TW02 發展語料	TW02 測試語料
錄音格式	單聲道、16kHz、16bits		
錄音者	479 人，男 253、女 226	121 人，男 64、女 57	16 人，男 8、女 8
錄音句數	93638 句	23410 句	2080 句 男 1028 句、女 1052 句
錄音時間	26 小時	6.5 小時	0.7 小時

前面有介紹到本測試語料為短字詞(一個字代表在 ForPA 標音的一個音節)的測試語料，

包含一字詞到五字詞，個數分別為 97、539、692、533、2 個。而實驗中的關鍵詞列表，及其在測試語料中的出現次數如下：

表三、關鍵詞列表

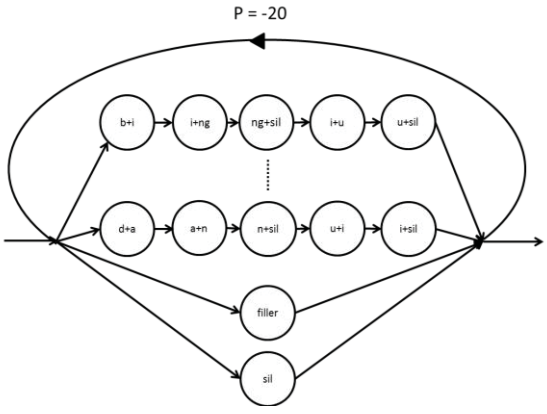
Keyword	出現次數	Keyword	出現次數	Keyword	出現次數	Keyword	出現次數
電話	15	語音	4	事件	7	單位	7
三更半暝	8	合理	4	可惡	5	政府	7
活動	8	太太	4	流血	5	運動	4
朋友	8	社會	4	程度	4	臭屁	4
台灣	7	酒醉	4	分鐘	4	火車	4

（二） 效能評估方法

本論文所採用的評估方法為錯誤拒絕率（false rejection rate ,FRR）[16]與錯誤接受率（false acceptance rate , FAR）[16]兩種，但通常一邊的錯誤率降低，另一邊的錯誤率會隨之升高，所以我們採用的為相等錯誤率（equal error rate, EER）[16]來評估關鍵詞辨識系統。

（三） 辨識網路介紹

在使用隱藏式馬可夫模型實做關鍵詞辨識系統時，所採用的辨識網路為用來分辨關鍵詞與非關鍵詞，其中關鍵詞為右相關雙連音素串接而成，如下圖所示：



圖八、關鍵詞辨識網路示意圖

另一方面，在利用音素匹配法實做關鍵詞辨識系統時，在第一階段時則使用一般的自由音素解碼。

（四） 未加入關鍵詞驗證之系統比對

1. 實驗目的

本實驗目的為找出未加入關鍵詞驗證時其兩種實做關鍵詞系統方法最佳的辨識率，利用兩種系統最佳辨識率的方法之後再進行改進，期望能得到一個最好的關鍵詞辨識系統。

2. 實驗流程與設定

在隱藏式馬可夫模型實做關鍵詞辨識系統中採取調整填充模型的懲罰值和關鍵詞模型的機率值，期望能達到一個可以接受的效果。在測試音素匹配法前，先測試自由音素解碼的準確率，之後進行音素匹配－混淆矩陣法中其插入和刪除矩陣的常數值，並測試三種不同的音素匹配法，期望得到一個最佳結果再予以改進。

3. 實驗結果：音素匹配－混淆矩陣法中不同常數之測試

在此實驗結果中我們得知音素匹配－混淆矩陣法來實做關鍵詞辨識系統其在插入與刪除矩陣常數值分別設為 3.5、4、4.5 和 5 時，以 4.5 最好，故之後我們用其與另外兩種懲罰矩陣之方法比較。

4. 實驗結果：音素匹配法與隱藏式馬可夫模型法比較

表四、HMM 與音素匹配法結果比較(未加入關鍵詞驗證)

代稱	關鍵詞系統使用方法	EER
PM	音素匹配－懲罰矩陣法	39.4%
CM	音素匹配－混淆矩陣法	34.0%
LD	音素匹配－距離矩陣法	42.2%
HMM	隱藏式馬可夫模型法	46.5%

自由因素解碼在 penalty 為-20 時結果最好，準確率為 50.32%，之後我們皆用此結果做為音素匹配法中第一階段的輸出。在單純比較利用音素匹配法實做關鍵詞辨識系統時，其音素匹配－混淆矩陣法可以得到最好的效果，其原因可能為因為在利用發展語料(development)時所建立的混淆矩陣有把音素與音素之間的關係成功的表現出來，故即使在自由音素解碼效果沒有很顯著的情況下，關鍵詞辨識系統仍有一定的效果（相等錯誤率 34%），而音素匹配－懲罰矩陣法效果較音素匹配－混淆矩陣法略為遜色一點，可能原因為雖然在製作此懲罰矩陣利用 HTK 下了許多功夫，但在音素與音素之間可能會有切音不準確之問題，導致有時 log-likelihood 值或許沒有正確的表示出音素與音素間真正的關係，所以間接的導致了關鍵詞辨識系統的正確率，最後利用距離矩陣只有單純以 0、1 來表示音素間的關係，故用其在進行關鍵詞辨識時，其效果非常不好。最後為 HMM 實做關鍵詞辨識系統時，因為填充模型較為簡單且關鍵詞模型僅為右相關雙連連音素模型串接而成，故無法達到較佳的效果是可以預期的。

（五） 加入關鍵詞驗證後系統之比對

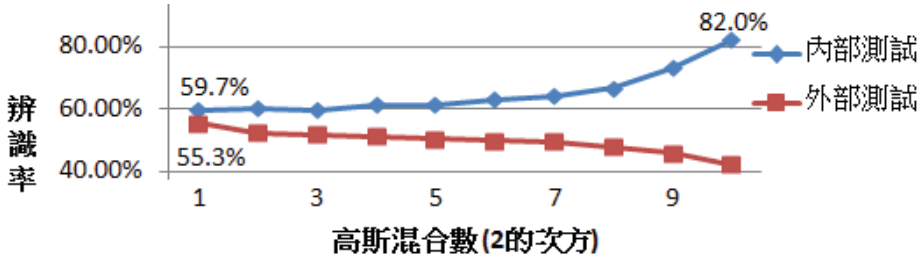
1. 實驗目的

經由上一節實驗的測試，我們得到了兩種實做關鍵詞辨識的基準，我們取一些較佳的效果來做驗證(音素匹配法取混淆矩陣法與懲罰矩陣法)，期望可以達到最佳的關鍵詞辨識系統。

2. 實驗流程與設定

一開始我們會先觀察『音高走勢分類器』的效果，得知音高分類的辨識率，之後我們先將兩種系統只加入語音評分來驗證，期望達到一個不錯的效果，之後再加入『音高走勢分類器』進行雙重驗證，觀察加入『音高走勢分類器』後是否與原本系統有互補的效果。

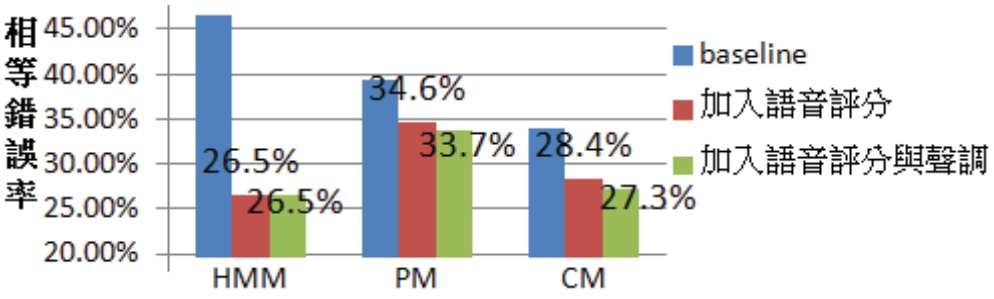
3. 實驗結果一：音高走勢分類器辨識結果



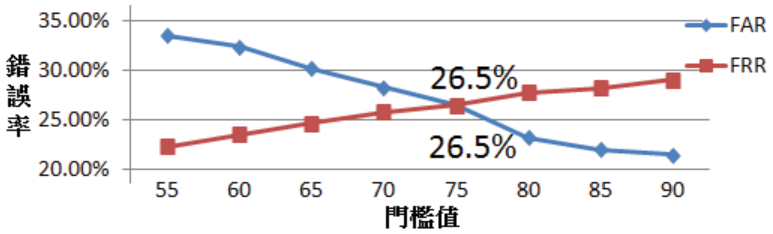
圖九、音高走勢分類器辨識結果

上圖為音高走勢分類器結果，可以得知當高斯混合數提升時，雖然內部測試的辨識率變高了，但對外部測試卻無法有幫助，故在此實驗中高斯混合數為 1 時有最好的效果，另一方面台語的入聲字（分類為 4，短促急停）辨識率大約只有 20%左右，因此大大的降低了整個音高走勢分類器其效果。

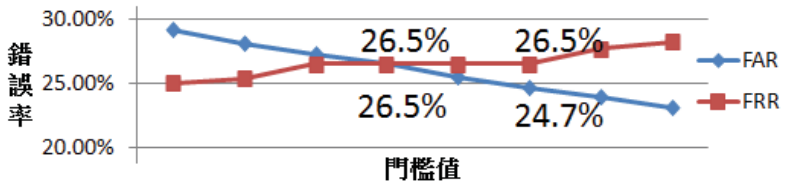
4. 實驗結果二：加入關鍵詞驗證之系統比較



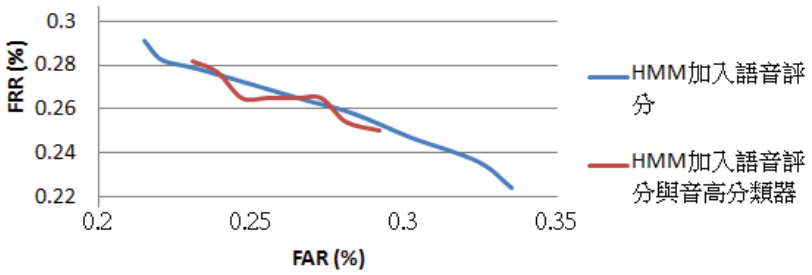
圖十、三種關鍵詞辨識系統加入關鍵詞驗證後之結果



圖十一、HMM 加入語音評分之折線圖分析



圖十二、HMM 加入語音評分和音高分類之折線圖分析



圖十三、HMM 加入語音評分與音高分類後 DET 曲線比較

經由結果我們可以知道利用語音評分來做為關鍵詞系統的驗證可以得到不錯的效果，其中隱藏式馬可夫模型法錯誤率下降了 20%，音素匹配法中之混淆矩陣法錯誤率下降了 5.6%、懲罰矩陣法下降了 4.8%，比較不符合預期的為其隱藏式馬可夫模型法其基準（baseline）較低，而音素匹配法其基準（baseline）較高，但經由語音評分後雖都有改進，但是原本效果較差的隱藏式馬可夫模型法錯誤率卻大為降低，推估原因為在進行隱藏式馬可夫模型時，其在第一階段關鍵詞辨識時是有進行切音與模型比對，此時雖然會有很高的錯誤接受率，但對於正確的關鍵詞並不會放過，而在實做語音評分時也是根據我們所建立的馬可夫模型做信心度的測量，兩階段都與我們所建立的馬可夫模型相關，故可以有較好的效果。而在使用音素匹配法時，雖然第一階段仍有使用自由音素解碼，並且利用發展語料調整音素與音素之間的相互關係並進行修正，但其內部的辨識核心仍與辨識網路無關，因此在關鍵詞驗證時無法達到我們所預期的效果。而另一方面我們加入了『音高走勢分類器』來跟語音評分做雙重的關鍵詞驗證，利用此方法在隱藏式馬可夫模型法中雖然相等錯誤率並沒有下降，但我們觀察到在相同的 FRR 中，其 FAR 降低了 1.8%，推測原因可能為因為關鍵詞的數量較少，在 FRR 中很容易達到一個飽和的情況，其錯誤率不容易在降低，但因為非關鍵詞的音檔較多，所以 FAR 的變動相對會比較大，所以在『音高走勢分類器』的部分是有助於改進 FAR。而在音素匹配法中混淆矩陣錯誤率下降了 1.1%、懲罰矩陣法下降了 0.9%，由此可以得知『音高走勢分類器』對於關鍵詞辨識系統是有幫助的。

五、結論

本論文提出了結合語音評分與音高走勢分類器的台語關鍵詞辨識系統，分別先實做了傳統實做關鍵詞系統的方法和另一種較新的方法，之後再利用語音評分做為信心度測量與『音高走勢分類器』，進行關鍵詞的驗證。而在本論文的實驗中，不論是使用隱藏式馬可夫模型法或是音素匹配法，其關鍵詞字典的擴充都很方便，前者只要更改關鍵詞的串

接模型，後者則只要更改關鍵詞的音素序列（phone sequence），故此為一個移植性高、關鍵詞擴充度高之系統，因此對於應用上可以有不錯的幫助。在隱藏式馬可夫模型法中，相等錯誤率從原本 46.5%，經由語音評分後下降至 26.5%，之後加入『音高走勢分類器』FAR 下降了 1.8%，一方面說明了利用語音評分來做為台語關鍵詞辨識系統的驗證是很有幫助的，另一方面也得知在之前未考慮到之語音的音高與音量特徵在與關鍵詞驗證做結合後是有達到互補的效果。在音素匹配法中，其混淆矩陣法與懲罰矩陣法也從相等錯誤率 34%、39.4%，經由語音評分後相等錯誤率下降為 28.4%、34.6%，而加入『音高走勢分類器』後相等錯誤率則下降到 27.3%、33.7%，音高走勢分類器辨識的改善效果也略為提高，其結果可以說是如我們預期。

六、未來研究方向

關鍵詞辨識部分，在隱藏式馬可夫模型方面，可以朝向訓練更為完善的填充模型[4]，使得關鍵詞模型與填充模型更加容易區分，讓關鍵詞驗證部分可以較為簡單。在音素匹配法部分，如何改善自由音素解碼的準確率算是首要課題，雖然可以利用發展語料調整懲罰矩陣的懲罰值，但若辨識正確而不用有懲罰值將會是最理想的。

關鍵詞驗證部分，可以試著採用其他不同的信心度測試方式來加強驗證效果，例如 LRT（likelihood ratio testing）的改良，而在『音高走勢分類器』部分可以使用 HMM 來訓練，使其達到更佳的辨識效果，或是增加像是時間長度（time duration）、停頓長度（pause duration）等來更精確的分類出入聲，或是其他聲調特徵，皆為可行之做法。

致謝

本論文經費來源由國科會計畫 NSC 99-2221-E-007 -049 -MY3 所提供

參考文獻

[1] C. W. Han, S. J. Kang, and N. S. Kim, “Estimation of phone mismatch penalty matrices for two-stage keyword spotting,” IEICE TRANSACTIONS on Information and Systems Vol.E93-D No.8 pp.2331-2335

[2] K. Audhkhasi and A. Verma, “Keyword search using modified minimum edit distance measure,” Proc. ICASSP, pp. 929-932,Apr. 2007.

[3] M. S. Barakat, C. H. Ritz, D. A. Stirling, “Keyword Spotting based on the Analysis of Template Matching Distances,” 5th International Conference on Signal Processing and Communication Systems ICSPCS,2011

[4] S.L. Zhang, Z.W. Shuang, Q. Shi, and Y. Qin, “Improved mandarin keyword spotting using confusion garbage model”, in Proc. ICPR, pp. 3700-3703, Istanbul, Turkey, 2010

[5] L.R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proc. Of the IEEE, Vol.77, No.2, pp. 257-286, Feb, 1989.

[6] Huang, X., Acero, A., and Hon, H.W., “Spoken Language Processing”, New Jersey, Prentice Hall, 2001

[7] R.-Y. Lyu, M.-S. Liang, Y.-C. Chiang, Toward Constructing A Multilingual Speech Corpus for Taiwanese (Min-nan), Hakka, and Mandarin Chinese, International Journal of Computational Linguistics & Chinese Language Processing, 2004

[8] Liao, H.-C., Chen, J.-C. , Chang, S.-C., Guan, Y.-H., Lee, C.-H., “Decision tree based tone modeling with corrective feedbacks for automatic Mandarin tone assessment”, In INTERSPEECH 2010.

[9] Chen, J.-C., Jang, J.S. R., “TRUES: Tone Recognition Using Extended Segment”, ACM Trans. Asian Lang. Inform. Process. 7, 3, Article 10, 2008.

[10] Ross, M.Shaffer, H. Cohen, A. Freudberg, R., and Manley, H., “Average Magnitude Difference Function Pitch Extractor,” IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 22,No. 5,353-362, 1974

[11] 鐘進竹，結合聲調辨認之中文關鍵詞辨認系統，國立交通大學碩士論文，民國 100 年

[12] 李俊毅，語音評分，清華大學碩士論文，民國 91 年

[13] 陳宏瑞，使用多重聲學模型以改良台語語音評分，清華大學碩士論文，民國 100 年

[14] 傅振宏，基於自動產生合成單元之台語語音合成系統，長庚大學碩士論文，民國 89 年

[15] 黃士旗，中文語音聲調辨識的改良與錯誤分析，清大碩士論文，民國 95 年。

[16] 黃冠達，應用支撐向量機於中文關鍵詞驗證之研究，國立臺灣科技大學碩士論文，民國 96 年

[17] 周哲玄，台語關鍵詞辨識之實作與比較，清大碩士論文，民國 101 年。