

# 塔米爾文語音辨識系統之設計研究

陳志堅

國立中山大學電機系副教授  
chenc@mail.ee.nsysu.edu.tw

林威廷

國立中山大學電機系碩士生  
chenc@mail.ee.nsysu.edu.tw

## 摘要

本論文針對南印度、斯里蘭卡及新加坡等地常用之塔米爾語，作語音辨識之研究，根據塔米爾語之發音方式，找出 149 類常用單音節，以每輪每次錄製兩個一四聲不同聲調之相同單音，錄製五輪共十次之策略，作為單音訓練之方式，建立塔米爾語資料庫中，共 149 類的聲紋特徵。本系統萃取梅爾頻率倒頻譜係數與線性預估倒頻譜係數作為語音之特徵，並利用隱藏式馬可夫模型及音位結構學比對，建立辨識系統。在 CPU 時脈 2.5 GHz 的 Intel Core 2 Quad 個人電腦與 Ubuntu 10.04 作業系統環境下，本論文實作了 3,500 筆塔米爾文常用語詞辨識系統，系統之正確率可達到 88.74%，平均辨識時間約在 1.5 秒以內。而系統之總訓練時間約為 1.5 小時。

**關鍵詞：**塔米爾文語音辨識系統、梅爾頻率倒頻譜係數、線性預估倒頻譜係數、隱藏式馬可夫模型、音位結構學。

## Abstract

Tamil language is widely used in southern India, Sri Lanka and Singapore. This paper investigates the design and implementation strategies for a Tamil speech recognition system. It utilizes the speech features of the 149 common Tamil mono-syllables as the major training and recognition methodology. A training database of 10 utterances per mono-syllable is established by applying Tamil pronunciation rules. These 10 utterances are collected through reading five rounds of the same mono-syllables twice with different tones. The first pronounced pattern has high pitch of tone one, while the second one has falling pitch of tone four. Mel-frequency cepstral coefficients, linear predicted cepstral coefficients, and hidden Markov model are used as the two feature models and the syllable recognition model respectively. The recognized syllable strings are then refined by phonotactical

rules to obtain the optimal result. Under the Intel Core 2 Quad 2.5 GHz personal computer and Ubuntu 10.04 operating system environment, a correct phrase recognition rate of 88.74% can be reached for a 3,500 Tamil phrase database. The average computation time for the system is within 1.5 seconds, and the training time for the systems is about one and half hours.

**Keywords:** Tamil speech recognition system, Mel-frequency cepstral coefficients, Linear predictive cepstral coefficients, Hidden Markov model, Phonotactics.

## 1. 前言

塔米爾語(Tamil)是印度前五大使用語言之一，亦是印度南部塔米爾納德邦、朋迪榭里聯邦、斯里蘭卡及新加坡等地的官方語言，根據 2009 年美國桑莫語言學院(Summer Institute of Linguistics, SIL)統計[9]，全世界約有 6,500 萬人使用塔米爾語作母語，使用人口排名全球第 18 名，與韓語 6,600 萬使用人口相去不遠。

2010 年涵蓋巴西、俄羅斯、印度、中國與南非的金磚五國(BRICS)正式形成[2]。印度作為世界五個主要新興市場的重要成員，其人口和國土面積在全球佔有重要份額，且是世界經濟增長的主要動力之一。在印度，除使用人口數排名第一，常用於北印度的印地語(Hindi)外，塔米爾語是印度南方塔米爾納德邦，人民生活溝通及書寫的主要溝通語言，境內充滿印度教的建築，亦被稱為「寺廟之鄉」，朝聖者絡繹不絕，形成了另一種特殊的南印度風情。除了塔米爾納德邦外，斯里蘭卡是塔米爾語另一個主要的使用地區，該國以農業為主要的經濟產業，尤其以錫蘭紅茶為最重要的出口產品，與印度及肯亞並列世界三大紅茶產國；基於旅遊經貿、工商開拓及語言學習的需要，吾人希冀建立一套塔米爾文語音辨識系統，達到查詢與學習的效果，並增進吾人對印度及斯里蘭卡之多元文化與歷史的了解。

## 2. 塔米爾文語音學簡介

塔米爾文共有 247 個原生字母[1,6,10]，其中包含 12 個母音、18 個子音以及子音與母音組合而成的字母符號 216 個，字母符號概念與日文較為相近，即以單一符號表示子音搭配母音之組合發音，另有一原生字母「ஃ」，不單獨使用，用於與子音 ஃ 組合發 [f] 音，共計 247 種不同表示符號，詳見附錄 1。

塔米爾文字為「表音文字」，亦即見符號即可讀，但其複雜的文字組合，常使人敬而遠之，塔米爾語的 12 個母音，有長音、短音及雙母音的區別，分別是(அ、இ、உ、எ、ஔ)等五個短母音及與其相互對應的五個長母音(ஆ、ஈ、ஊ、ஏ、ஐ)，另外有兩個雙母音(ஔ、ஔா)，其發音以國際音標(IPA)表示，如下表 1 所示：

表 1 塔米爾母音發音表

字母	அ	ஆ	இ	ஈ	உ	ஊ
發音	[a]	[a:]	[ɪ]	[i]	[u]	[u:]
字母	எ	ஏ	ஐ	ஔ	ஔ	ஔா
發音	[ɛ]	[e]	[ai]	[ɔ]	[o]	[au]

塔米爾語的子音共有 18 個，包含了五個爆破音(க், ச், ட், த், ப்)、六個鼻音、兩個邊音(ல், ள்)、三個擦音及兩個滑音(半母音)，其中五個爆破音較為特殊，會根據所處位置不同而產生變音情況，以下介紹塔米爾語之子音發音及變音情形[1,6,10]：

表 2 塔米爾子音發音表

字母	க்	ச்	ட்	த்	ப்	ற்
發音	[k]	[tʃ]	[t]	[t]	[p]	[r]
字母	ங்	ஞ்	ண்	ந்	ம்	ன்
發音	[ŋ]	[ɲ]	[ɳ]	[n:]	[m]	[n]
字母	ய்	ர்	ல்	வ்	ழ்	ள்
發音	[j]	[r]	[l]	[v]	[ɻ]	[l]

上述子音中有 6 個較為特殊的發音，以國際音標(IPA)來看，[t]發音近似發[t]音後捲舌收尾，[ɲ]和[ɳ]發音近似[nj]及[ng]，[r]發[zw]音，[ɻ]發音較接近中文注音中的「ㄖ」，[l]是將[l]音的卷舌程度加重並延長發音。

上表 2 中的前五個子音為爆破音，又稱閉塞子音，是指在發音前先將氣流通路閉塞，再突然放開使氣流一次爆出，在短時間內釋放高能量，故稱爆破音；塔米爾語中的爆破音，會依據前後字母發音情形而產生變音，有下列幾種情形：鼻子音(ங்、ஞ்、ண்、ந்、ம்、ன்)後、兩母音間、ர் / ய் 之後、及位於詞首時，其中[d]發音，為將[d]發捲舌之音，下表 3 為爆破音之變化規則表[1,6,10]：

表 3 塔米爾爆破音變音規則

	鼻音後	母音間	ர்/ய்後	詞首	其他
க்	[g]	[h]	[h]	[k]	[k]
ச்	[dʒ]	[s]		[tʃ]	[tʃ]
ட்	[d]	[d]		[t]	[t]
த்	[d]	[d]		[t]	[t]
ப்	[b]	[b]		[p]	[p]

除母音與子音外，塔米爾語中還存在一種字母，是以 க், ச், ... , ள் 為基底，將母音轉化為一種附標文字，結合形成塔米爾的子音及其餘字母(附錄 1)，此類附標文字其發音概念與日文文字相近，字母可自成一音節，母音型態變化見下表 4：

表 4 塔米爾母音變化

字母	子音	ஆ	இ	ஈ	உ	ஊ
附標文字	ஃ	ஃா	ஃி	ஃீ	ஃு	ஃூ
字母	எ	ஏ	ஐ	ஔ	ஔ	ஔா
附標文字	ஃெ	ஃே	ஃை	ஃொ	ஃோ	ஃௌ

塔米爾文中的附標文字發音為子音+母音的發音方式，書寫上是以子音搭配母音附標文字後形成新的文字[1,6,10,14]，如下表 5 所示：

表 5 塔米爾附標文字結合方式

母音 子音	அ [a]	ஆ [a:]	இ [i]	...
க [k]	க [ka]	கா [ka:]	கி [ki]	
ச [ç]	ச [ça]	சா [ça:]	சி [çi]	
⋮				

綜括之，就塔米爾語中的字母發音來看，塔米爾語母音有長短之分，且五個爆破子音之發音，會因其位置之不同，產生變音現象。另外，需注意的是：塔米爾文字，並無標示重音之習慣，但重音一般落在第一音節；複數名詞是由原始詞彙後加上特定綴詞於字尾構成；塔米爾語名詞有陽性與陰性之分，不可混用。

### 3. 語音辨識系統的流程

本論文之語音辨識系統架構，如右圖 1 所示。語音錄音之規格為取樣率每秒 11,025 個樣本、位元數是 16，音框長度為 256 點，前後音框重疊 128 點。辨識系統架構在 CPU 時脈為 2.5 GHz 的 Intel Core 2 Quad 之個人電腦與 Ubuntu 10.04 之作業系統環境下。待測語音之信號，經由音節分析判斷後，運用含時間一次微分的 26 維梅爾頻率倒頻譜係數 (Mel-frequency cepstral coefficients, MFCC) 來模擬人耳聽覺對於不同頻率之變化情形，及運用 20 維線性預估倒頻譜係數 (Linear predictive cepstral coefficients, LPCC) [3,4,5] 來模擬發聲腔道之變化，進行單音節之雙特徵參數萃取，再與隱藏式馬可夫模型 (Hidden Markov model, HMM) [3,4,5] 所訓練出之塔米爾語單音模型，運用維特比演算法 (Viterbi algorithm) [3,4,5]，來得到最接近的候選單音群。最後再經由塔米爾語詞資料庫，以音位結構學進行交叉比對，篩選出最合適的辨識結果。

#### 3.1 梅爾頻率倒頻譜係數特徵

梅爾倒頻譜係數是倒頻譜特徵 (Cepstrum feature) 的一種加強的分析方法。它除了具有倒頻譜係數，分離聲道特徵 (Vocal tract feature) 與聲音刺激源 (Voice excitation source) 之特性外，亦增加了一些特別的心理聲學感知性能 (Psycho-acoustically perceived feature)。根據這方面的研究，發現人類的聽覺對於頻率的感知能力，與實際的頻率並不是呈現線性的關係，

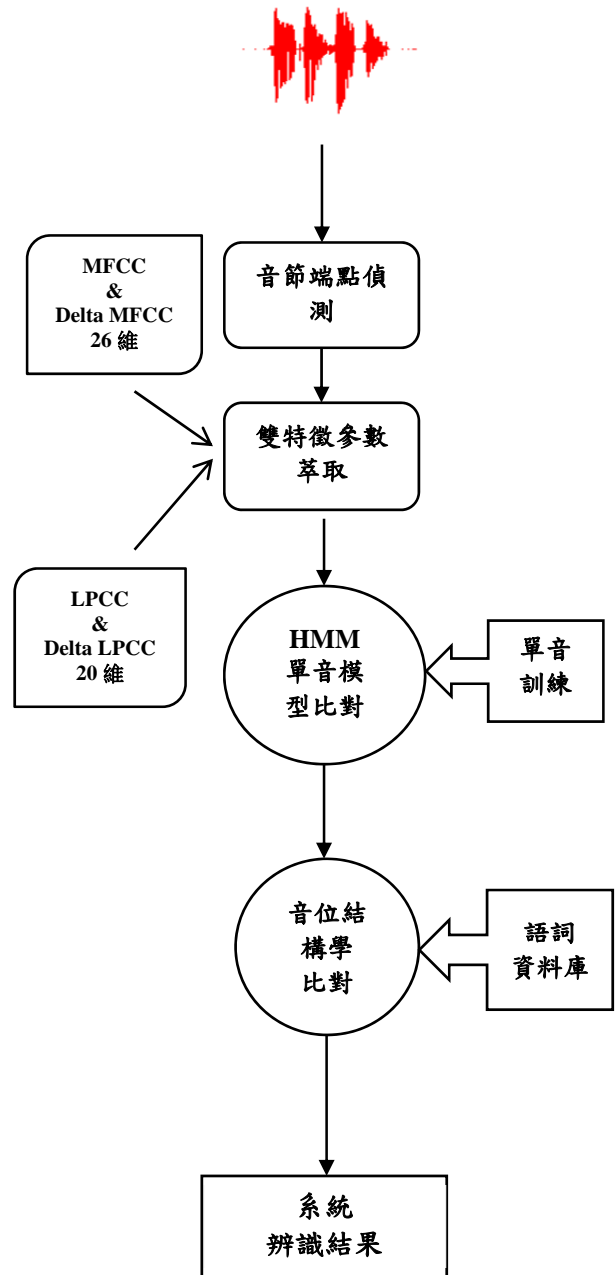


圖 1 辨識系統流程圖

而是呈現對數 (Logarithmic) 關係，亦即低頻感知能力較佳，高頻感知能力較遲頓。這關係可以如下之數學式子來近似：

$$M(f) = 2595 \times \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1)$$

其中  $M(f)$  為梅爾頻率， $f$  為實際頻率。

我們若將輸入的語音訊號不管高低頻而一視同仁的處理，就變得不是很恰當，換句話說，系統應該是對語音的低頻部份較為重視，而高頻訊號部份則可以有些許忽略，或者是說

從低頻部份中萃取出較多的特徵，而從高頻部份中萃取出較少的特徵。如此一來，語音辨識系統之特徵設計，便會更符合人類聽覺系統的真實狀況，而可獲得較優良之辨識結果。為了模擬人類的聽覺特性，我們導入了臨界頻帶濾波器組(Critical band based filter bank)，來達成這樣的目的。在這個濾波器組(Filter bank)中，每個濾波器的頻率響應皆設為三角形，且其間距(Spacing)與頻寬(Bandwidth)，投影到梅爾頻域(Mel-frequency domain)後，其梅爾間距(Mel-frequency spacing)和梅爾頻寬皆相同。這種濾波器組，其實際的頻域(Real frequency domain)分佈情況，如下圖所示：

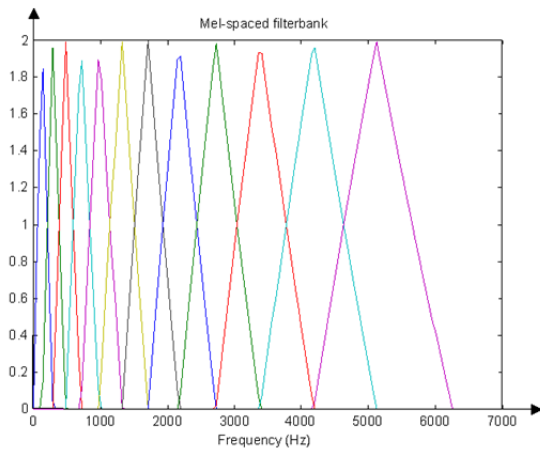


圖 2 濾波器群組示意圖

而其帶通濾波函數之數學式為：

$$B_m(k) \begin{cases} 0 & , k < f_{m-1} \\ \frac{k - f_{m-1}}{f_m - f_{m-1}} & , f_{m-1} \leq k \leq f_m \\ \frac{f_{m+1} - k}{f_{m+1} - f_m} & , f_m \leq k \leq f_{m+1} \\ 0 & , f_{m+1} < k \end{cases} \quad (2)$$

其中  $B_m(k)$  為第  $m$  個三角形濾波器中第  $k$  點之數值， $f_m$  為第  $m$  個三角濾波器之中心頻率，本系統採用 26 組頻帶之梅爾頻率濾波器。吾人先以 FFT 求取每個音框的頻譜平方值，再乘上每個三角濾波器的各頻帶增益，作內積累加，之後取對數，得到各個梅爾頻帶之對數能量：

$$Y(m) = \log \left\{ \sum_{k=f_{m-1}}^{f_{m+1}} |X(k)|^2 B_m(k) \right\} \quad (3)$$

其中  $|X(k)|^2$  為信號經傅式轉換後的能量頻譜。

因傅立葉頻譜函數  $|X(k)|$  為對稱半音框長度  $N/2$  之偶對稱函數， $Y(m)$  亦必為偶對稱，故其反傅立葉轉換之正弦部分為零。因此，吾人僅需以反餘弦轉換，代替反傅利葉轉換，即可得到梅爾倒頻譜係數 (Mel-frequency cepstral coefficients, MFCC)，其公式如下：

$$mfcc(i) = \sqrt{\frac{2}{M}} \beta(i) \sum_{m=1}^M Y(m) \cos \left( \frac{\pi i (2m + 1)}{2M} \right) \quad (4)$$

$, 0 \leq i \leq N - 1$

$$\beta(i) = \begin{cases} \frac{1}{\sqrt{2}} & , i = 0 \\ 1 & , i = 1, 2, \dots, N - 1 \end{cases} \quad (5)$$

上述公式中， $\beta(i)$  為加權量，用以對 MFCC 作正規化， $N$  為 MFCC 維度， $M$  為梅爾濾波器個數。

本系統採用  $N=13$ ， $M=26$  之設計。另外，為獲取更多有用之語音特徵，吾人將以 13 個梅爾頻率倒頻譜係數，做進一步時間差量之運算，讓特徵維度提升為 26，以更多係數去描述語音的時間變化特性。

### 3.2 線性預估倒頻譜係數特徵

線性預估模型 (Linear predicted model, LP model) 是描述自然界中時間或空間相關性信號 (Temporally or spatially correlated signals) 相當有效的利器。人類語音信號前後相鄰點之聲量大小關係，可以下列數學式表達：

$$S'_n(m) = \sum_{k=1}^P a_k S_n(m - k) \quad (6)$$

式中  $S'_n(m)$  即為線性預估值，其中  $S_n(m - k)$  為第  $n$  個音框中第  $m - k$  個數值， $a_k$  為線性預估參數 (Linear predictive coefficient, LPC)， $P$  為線性預估模型的階數， $m = P, P+1, \dots, N-1$ 。 $N$  為音框總點數。吾人可運用已知真實答案的前面  $P$  個點，經最佳化  $\{a_k, k = 1, 2, \dots, P\}$  的線性推估，得到近似真實值的下一個時間點之預估值。最佳線性預估係數  $a_k$ ，是經下列最小平方差的原理推導求得：

$$E_n = \sum_{m=p}^{N-1} [S_n(m) - S'_n(m)]^2 \quad (7)$$

將上式中之  $S'_n$ ，由(6)式代入，並對  $a_k$  偏微分，以求取誤差量  $E_n$  之極小值：

$$\frac{\partial E_n}{\partial a_i} = 0, i = 1, 2, \dots, P \quad (8)$$

$$\begin{aligned} & \sum_{m=p}^{N-1} S_n(m)S_n(m-i) \\ &= \sum_{k=1}^P a_k \sum_{m=p}^{N-1} S_n(m-k)S_n(m-i) \end{aligned} \quad (9)$$

令自相關函數  $R_n$  為

$$R_n(i) = \sum_{m=p}^{N-1} S_n(m)S_n(m-i) \quad (10)$$

則(9)式可簡化為

$$R_n(i) = \sum_{k=1}^P a_k R_n(i-k), 1 \leq i \leq P \quad (11)$$

將(11)式轉為矩陣形式，可得  $R \cdot \alpha = r$  (12)

$$\begin{bmatrix} R_n(0) & \dots & R_n(P-1) \\ R_n(1) & \dots & R_n(P-2) \\ R_n(2) & \dots & R_n(P-3) \\ \vdots & \ddots & \vdots \\ R_n(P-1) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \vdots \\ R_n(P) \end{bmatrix}$$

上式經由反矩陣運算或經由 Levinson-Durbin 快速演算法，可得到最佳線性預估參數  $a_k$ ，其中  $k=1 \sim P$ 。再運用如下之倒頻譜公式加以計算，可得到線性預估倒頻譜係數 (Linear predictive cepstral coefficients, LPCC)

$$C_0 = \ln G \quad (13)$$

$$C_n = a_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) C_k a_{n-k}, 1 \leq n \leq P \quad (14)$$

$$C_n = \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) C_k a_{n-k}, n > P \quad (15)$$

其中  $C_n$  為第  $n$  個 LPCC， $a_n$  為第  $n$  個 LPC， $G$  為音量增益。線性預估倒頻譜係數最大優點在於其計算量甚小，且系統使用極少量係數，即可萃取出足以代表語音訊號的特徵參數

### 3.3 隱藏式馬可夫模型

隱藏式馬可夫模型 (HMM) 是一種描述物理與生物現象，相當有效的機率統計模型。HMM 是一種雙重隨機過程 (Doubly stochastic process)，透過模型狀態與觀察值不是決定性 (一對一對應)，而是隨機性 (一對多對應) 的假說，適切合理地詮釋自然界的的不確定性。HMM 常被用來描述語音發聲“音長不定”與字詞子母音間出現“時程不拘”的隨機過程。隱藏式馬可夫模型，可以下列數學式來表示：

$$\Phi = (A, B, \pi) \quad (16)$$

其中  $\Phi$  為隱藏式馬可夫模型，含有  $A$ 、 $B$  與  $\pi$  三個元素。 $A = \{a_{ij}\}$  為  $N \times N$  階狀態轉移機率矩陣， $a_{ij}$  說明狀態  $i$  轉移到狀態  $j$  的機率， $N$  為描述語音發聲過程所使用的總狀態數； $B = \{b_i(o_t)\}$  為  $M \times N$  階與狀態相關之輸出機率矩陣， $o_t$  為時間  $t$  之觀察向量， $M$  為描述語音發聲過程中，在各個狀態下所使用的混合高斯分布個數； $\pi = \{\pi_i\}$  為  $N \times 1$  階起始狀態機率分佈向量。 $A$ 、 $B$  與  $\pi$  之數學表示式如下：

$$A = \begin{cases} \frac{N_{state} - 1}{N}, i = j \\ \frac{1}{N_{state}}, j - i = 1 \\ 0, otherwise \end{cases} \quad (17)$$

$$B = \{b_i(o_t)\} = \sum_{k=1}^M p_{ik} N(o_t, \mu_{ik}, \Sigma_{ik}) \quad (18)$$

$$\pi = \{\pi_i\} = \{1, 0, 0, \dots, 0\} \quad (19)$$

(18)式中

$$\begin{aligned} & N(o_t, \mu_{ik}, \Sigma_{ik}) \\ &= \frac{1}{\sqrt{(2\pi)^D |\Sigma_{ik}|}} \exp \left[ -\frac{1}{2} (o_t - \mu_{ik})^T \Sigma_{ik}^{-1} (o_t - \mu_{ik}) \right] \end{aligned} \quad (20)$$

$$\mu_{ik} = \sum_{n=1}^{NC_{ik}} \left( \frac{X_{ik,n}}{NC_{ik}} \right) \quad (21)$$

$$\Sigma_{ik} = \sum_{n=1}^{NC_{ik}} \left[ \frac{(X_{ik,n} - \mu_{ik})(X_{ik,n} - \mu_{ik})^T}{NC_{ik} - 1} \right] \quad (22)$$

上式中， $D$  為特徵向量維度， $NC_{ik}$  為在第  $i$  狀態下屬於第  $k$  個高斯的音框總數， $X_{ik,n}$  為屬於第  $i$  狀態下第  $k$  個高斯的第  $n$  個音框之特徵向量， $\mu_{ik}$  與  $\Sigma_{ik}$  分別表示在第  $i$  狀態下第  $k$  個高斯分量之平均向量與互變數矩陣。

吾人運用平均向量  $\mu_{ik}$  與對角式互變異量矩陣 (Diagonal covariance matrix)  $\Sigma_{ik}$ ，來表達本論文中與狀態相關之輸出機率矩陣  $B$  的內涵，而系統之起始狀態機率分佈向量  $\pi$ ，設為  $(1, 0, \dots, 0)$  的  $N$  維向量。系統狀態轉移機率矩陣  $A$ ，平均向量  $\mu_{ik}$ ，與對角式互變異量矩陣  $\Sigma_{ik}$  三組參數，是經由 Baum-Welch algorithm，又稱為期望值最大化之演算法 (Expectation maximization algorithm, EM algorithm) [3,4,5]，藉反覆重估語音訓練模型中，相同類別塔米爾語單音節之特徵向量而獲得 [7,8,11~13]。

#### 4. 單音訓練策略與系統辨識率

在本論文中，吾人針對塔米爾語，蒐集了食衣住行等常用詞彙共 3,500 筆，如表 5 所示。由此 3,500 筆資料庫，可歸納出 149 類塔米爾常用單音節 (附錄 2)；塔米爾語屬非聲調語言，其音節中「重音」及「連音音位」，對音節訓練之完整性，具有相當的影響，故本研究採一輪錄製兩個不同聲調之相同單音，作為單音訓練之策略，並探討訓練輪數對辨識率之影響。

表 5 塔米爾文常用語詞數量統計表

語詞數 \ 資料量	模擬語詞數量
2 音節	600 筆
3 音節	1,100 筆
4 音節	900 筆
5 音節	900 筆
合計	3,500 筆

##### 4.1 每輪每次錄兩個單音的條件下，不同訓練輪數對辨識率的影響

在訓練單音確定後，根據對於塔米爾語之研究，需考量音節「重音」及語詞前後音節之「連音音位」關聯性 [15]，故採以陰平一聲與去聲四聲連續錄製兩個不同聲調單音之訓練方式，每輪 149 組單音，錄製若干輪，觀察單音訓練輪數與辨識率之關係，如下表 6 所示：

表 6 訓練次數與辨識率的關係

輪數	2 音節辨識率 (600)	3 音節辨識率 (1,100)	4 音節辨識率 (900)	5 音節辨識率 (900)	平均辨識率 (3,500)
2 輪	80.50%	82.64%	88.11%	91.00%	85.83%
3 輪	81.50%	84.18%	89.00%	92.11%	87.00%
4 輪	82.50%	85.45%	89.67%	92.89%	87.94%
5 輪	83.33%	86.36%	90.22%	93.78%	<b>88.74%</b>
6 輪	83.67%	86.64%	90.33%	94.00%	88.97%

由表 6 中，吾人得知當單音訓練輪數變多時，辨識率會跟著上升；但當增到第五輪後，辨識率提升之幅度，便趨飽和。由訓練五輪增到六輪，訓練時間增加 20%，系統辨識率僅能增加 0.23%。因此，本辨識系統以一次錄製兩個不同聲調單音，共訓練五輪，每類單音獲取十次訓練語料之方式，建立塔米爾語之 149 類單音聲紋模型，系統平均辨識率可達 88.74%。

#### 5. 塔米爾文語音辨識系統實作效能

吾人針對 3,500 筆塔米爾常用語詞資料庫做測試，以上述陰平一聲及去聲四聲連音錄製之方式，錄製 149 組單音共五輪，每單音共可得十次訓練語料，其系統辨識率如下表所示：

表 7 塔米爾文常用語詞之系統正確辨識率

字詞數 \ 資料量	模擬語詞數量	辨識率
2 音節	600 筆	83.33 %
3 音節	1,100 筆	86.36 %
4 音節	900 筆	90.22 %
5 音節	900 筆	93.78 %
合計	3500 筆	88.74 %

#### 6. 結論

本論文針對塔米爾語之發音特性，以 149 類常用單音節，作為訓練聲紋模型之依據，並以梅爾頻率倒頻譜係數與線性預估倒頻譜係數，來萃取單音節之聲紋特徵，利用隱藏式馬可夫模型，來描述聲紋參數之變化，最後經由音位結構學的比對，找出最合適的語詞，使本系統達到可靠的辨識效能。

在研究單音訓練策略的過程中，可以發現單音訓練輪數與辨識率成正比關係，訓練樣本數越多，系統辨識率越高；但當訓練樣本數，到達一個臨界值後，辨識率的增加，便會趨於飽和。其次經考量語詞音節間有其「連音音位」與「音調變化」之關係後，兩連音與一四聲之訓練策略，為最理想之機制。因此，本論文之辨識系統運用五輪十次之聲紋訓練機制，針對149類之塔米爾語常用單音節，採取陰平一聲與去聲四聲的錄製方式，來獲取塔米爾文語音辨識系統之訓練語料。在CPU時脈2.5 GHz的Intel Core 2 Quad個人電腦與Ubuntu 10.04作業系統環境下，本系統針對3,500筆塔米爾語資料庫作辨識效能之評析，其正確辨識率可達88.74%。辨識時間約在1.5秒以內，而系統所需的訓練時間約為1.5小時。

### 參考文獻

- [1] 張國強, *泰米爾語教程*, 中國傳媒大學出版社, 2009
- [2] 維基百科  
<http://en.wikipedia.org/wiki/BRICS>
- [3] Chu, W.C., *Speech Coding Algorithms*, Wiley Publisher, 2003
- [4] Huang, X.D., Acero, A. and Hon, H.W., *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*, Pearson Education, Taiwan 2005
- [5] Jurafsky, D. and Martin, J.H., *Speech and Language Processing*, Prentice Hall, Taiwan, 2009
- [6] Keane, E., *Prominence in Tamil*, Journal of the International Phonetic Association, 36(1), 1-20, 2006.
- [7] Krishnamurthy, E. V. "Automatic phonetic transcription of Tamil in Roman script," *Proceedings Mathematical Science*, 86.6, pp. 503-512, 1977.
- [8] Kumar, C.S. & Wei, F.S., "A bilingual speech recognition system for English and Tamil," *Proceedings of Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia*, Vol. 3, pp.1641-1644, 2003.
- [9] Lewis, M. P., *Ethnologue: Languages of the World*, Sixteenth edition, SIL International, 2009.
- [http://www.ethnologue.com/ethno\\_docs/distribution.asp?by=size](http://www.ethnologue.com/ethno_docs/distribution.asp?by=size)
- [10] Omniglot, <http://www.omniglot.com/>
- [11] Saraswathi, S., Geetha, T. V., and Saravanan, K., "Integrating language independent segmentation and language dependent phoneme based modeling for Tamil speech recognition system," *Asian Journal of Information Technology*, Vol. 5, No. 1, pp.38-43, 2006
- [12] Saraswathi, S., and Geetha, T.V. "Morpheme based language model for Tamil speech recognition system," *The International Arab Journal of Information Technology*, Vol.4, No.3, pp.214-219, 2007.
- [13] Schultz, T., and Waibel, A., "Multilingual and cross-lingual speech recognition," *Proceedings of DARPA Workshop on Broadcast News Transcription and Understanding*, pp. 259-262, 1998.
- [14] Vinton, J.E., *Phonetic emphasis in Tamil*, 1994.
- [15] Wang, L., Ambikairajah, E., and Choi, E.H.C., "Automatic language recognition with tonal and non-tonal language pre-classification," in the *15th European Signal Processing Conference, EUSIPCO 2007*, pp.2375-2379, 2007.

附錄 1：塔米爾語字母表

	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஓ	ஔ	ஔ
	[a]	[a:]	[ɪ]	[i]	[u]	[u:]	[ɛ]	[e]	[ai]	[ɔ]	[o]	[au]
க்	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கௌ
[k]	[ka]	[ka:]	[kɪ]	[ki]	[ku]	[ku:]	[kɛ]	[ke]	[kai]	[kɔ]	[ko]	[kau]
ச்	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சௌ
[tʃ]	[tʃa]	[tʃa:]	[tʃɪ]	[tʃi]	[tʃu]	[tʃu:]	[tʃɛ]	[tʃe]	[tʃai]	[tʃɔ]	[tʃo]	[tʃau]
ட்	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டௌ
[t]	[ta]	[ta:]	[tɪ]	[ti]	[tu]	[tu:]	[tɛ]	[te]	[tai]	[tɔ]	[to]	[tau]
த்	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தௌ
[t]	[ta]	[ta:]	[tɪ]	[ti]	[tu]	[tu:]	[tɛ]	[te]	[tai]	[tɔ]	[to]	[tau]
ப்	ப	பா	பி	பீ	பு	பூ	பெ	பே	பை	பொ	போ	பௌ
[p]	[pa]	[pa:]	[pi]	[pi]	[pu]	[pu:]	[pɛ]	[pe]	[pai]	[pɔ]	[po]	[pau]
ற்	ற	றா	றி	றீ	று	றூ	றெ	றே	றை	றொ	றோ	றௌ
[r]	[ra]	[ra:]	[rɪ]	[ri]	[ru]	[ru:]	[rɛ]	[re]	[rai]	[rɔ]	[ro]	[rau]
ங்	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙௌ
[ŋ]	[ŋa]	[ŋa:]	[ŋɪ]	[ŋi]	[ŋu]	[ŋu:]	[ŋɛ]	[ŋe]	[ŋai]	[ŋɔ]	[ŋo]	[ŋau]
ஞ்	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞௌ
[ɲ]	[ɲa]	[ɲa:]	[ɲɪ]	[ɲi]	[ɲu]	[ɲu:]	[ɲɛ]	[ɲe]	[ɲai]	[ɲɔ]	[ɲo]	[ɲau]
ண்	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணௌ
[ɳ]	[ɳa]	[ɳa:]	[ɳɪ]	[ɳi]	[ɳu]	[ɳu:]	[ɳɛ]	[ɳe]	[ɳai]	[ɳɔ]	[ɳo]	[ɳau]
ந்	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நௌ
[n:]	[na]	[na:]	[nɪ]	[ni]	[nu]	[nu:]	[nɛ]	[ne]	[nai]	[nɔ]	[no]	[nau]
ம்	ம	மா	மி	மீ	மு	மூ	மெ	மே	மை	மொ	மோ	மௌ
[m]	[ma]	[ma:]	[mi]	[mi]	[mu]	[mu:]	[mɛ]	[me]	[mai]	[mɔ]	[mo]	[mau]
ன்	ன	னா	னி	னீ	னு	னூ	னெ	னே	னை	னொ	னோ	னௌ
[n]	[na]	[na:]	[ni]	[ni]	[nu]	[nu:]	[nɛ]	[ne]	[nai]	[nɔ]	[no]	[nau]
ய்	ய	யா	யி	யீ	யு	யூ	யெ	யே	யை	யொ	யோ	யௌ
[j]	[ja]	[ja:]	[ji]	[ji]	[ju]	[ju:]	[jɛ]	[je]	[jai]	[jɔ]	[jo]	[jau]
ர்	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ	ரௌ
[r]	[ra]	[ra:]	[ri]	[ri]	[ru]	[ru:]	[rɛ]	[re]	[rai]	[rɔ]	[ro]	[rau]
ல்	ல	லா	லி	லீ	லு	லூ	லெ	லே	லை	லொ	லோ	லௌ
[l]	[la]	[la:]	[li]	[li]	[lu]	[lu:]	[lɛ]	[le]	[lai]	[lɔ]	[lo]	[lau]
வ்	வ	வா	வி	வீ	வு	வூ	வெ	வே	வை	வொ	வோ	வௌ
[v]	[va]	[va:]	[vi]	[vi]	[vu]	[vu:]	[vɛ]	[ve]	[vai]	[vɔ]	[vo]	[vau]
ழ்	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழௌ
[ɻ]	[ɻa]	[ɻa:]	[ɻɪ]	[ɻi]	[ɻu]	[ɻu:]	[ɻɛ]	[ɻe]	[ɻai]	[ɻɔ]	[ɻo]	[ɻau]
ள்	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளை	ளொ	ளோ	ளௌ
[ɻ]	[ɻa]	[ɻa:]	[ɻɪ]	[ɻi]	[ɻu]	[ɻu:]	[ɻɛ]	[ɻe]	[ɻai]	[ɻɔ]	[ɻo]	[ɻau]
ஃ												



附錄2：塔米爾語常用單音表

編號	塔米爾文	羅馬拼音	音標
0001	அ	a	[ a ]
0002	அன்	an	[ an ]
0003	ப	ba	[ ba ]
0004	பெ	be	[ bε ]
0005	பி	bi	[ bɪ ]
0006	பொ	bo	[ bɔ ]
0007	பு	bu	[ bu ]
0008	ச	cha	[ tʃa ]
0009	சை	chai	[ tʃaɪ ]
0010	சல்	chal	[ tʃal ]
0011	சன்	chan	[ tʃan ]
0012	சற்	char	[ tʃar ]
0013	செ	che	[ tʃε ]
0014	சி	chi	[ tʃɪ ]
0015	சொ	cho	[ tʃɔ ]
0016	சு	chu	[ tʃu ]
0017	த	da	[ da ]
0018	தை	dai	[ dai ]
0019	தல்	dal	[ dal ]
0020	தன்	dan	[ dan ]
0021	தற்	dar	[ dar ]
0022	தி	di	[ di ]
0023	திற	dir	[ dir ]
0024	தொ	do	[ dɔ ]
0025	து	du	[ du ]
0026	ஏ	e	[ e ]
0027	ஏன்	en	[ εn ]
0028	க	ga	[ ga ]
0029	கல்	gal	[ gal ]
0030	கி	gi	[ gi ]
0031	கு	gu	[ gu ]
0032	க	ha	[ ha ]
0033	கை	hai	[ hai ]
0034	கல்	hal	[ hal ]
0035	கன்	han	[ han ]
0036	கற்	har	[ har ]
0037	கி	hi	[ hi ]
0038	கொ	ho	[ ho ]
0039	கு	hu	[ hu ]
0040	இ	i	[ i ]

編號	塔米爾文	羅馬拼音	音標
0041	இல்	il	[ il ]
0042	இற்	ir	[ ir ]
0043	ச	ja	[ dʒa ]
0044	சி	ji	[ dʒɪ ]
0045	சு	ju	[ dʒu ]
0046	க	ka	[ ka ]
0047	கை	kai	[ kai ]
0048	கல்	kal	[ kal ]
0049	கன்	kan	[ kan ]
0050	கற்	kar	[ kar ]
0051	கெ	ke	[ kε ]
0052	கி	ki	[ ki ]
0053	கொ	ko	[ ko ]
0054	கொல்	kol	[ kol ]
0055	கு	ku	[ ku ]
0056	குன்	kun	[ kun ]
0057	ல	la	[ la ]
0058	லை	lai	[ lai ]
0059	லன்	lan	[ lan ]
0060	லற்	lar	[ lar ]
0061	லெ	le	[ lε ]
0062	லி	li	[ li ]
0063	லொ	lo	[ lo ]
0064	லு	lu	[ lu ]
0065	லுன்	lun	[ lun ]
0066	ம	ma	[ ma ]
0067	மை	mai	[ mai ]
0068	மன்	man	[ man ]
0069	மெ	me	[ mε ]
0070	மி	mi	[ mi ]
0071	மின்	min	[ min ]
0072	மொ	mo	[ mo ]
0073	மு	mu	[ mu ]
0074	முன்	mun	[ mun ]
0075	ன	na	[ na ]
0076	னை	nai	[ nai ]
0077	னால்	nal	[ nal ]
0078	னன்	nan	[ nan ]
0079	னற்	nar	[ nar ]
0080	னெ	ne	[ nε ]

