國 立 交 通 大 學

電信工程研究所

碩 士 論 文

考慮語速影響與詞綴構詞之中文語音辨認系統

A Mandarin Speech Recognition System Incorporating with Speaking Rate Modeling and Word Construction

研 究 生：林俊翰

指導教授：陳信宏　教授

中 華 民 國 一百零一 年 七 月

考慮語速影響與詞綴構詞之中文語音辨認系統

A Mandarin Speech Recognition System Incorporating with Speaking Rate Modeling and Word Construction

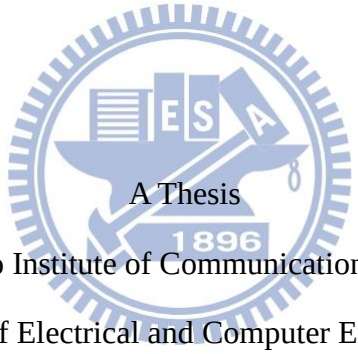
研 究 生： 林俊翰　　Student：Jun-Han Lin

指導教授： 陳信宏 博士　　Advisor：Dr. Sin-Horng Chen

國 立 交 通 大 學

電信工程研究所

碩 士 論 文



A Thesis

Submitted to Institute of Communication Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Communication Engineering

July 2012

Hsinchu, Republic of China

中 華 民 國　一百零一 年　七 月

考慮語速影響與詞綴構詞之中文語音辨認系統

研 究 生：林俊翰　　指導教授：陳信宏 博士

國立交通大學電信工程研究所碩士班

中文摘要

本研究提出一個新的中文大詞彙連續語音辨認方法來考慮綴詞的辨認及語速對辨認的影響，首先針對綴詞，從構詞學的角度出發，利用綴詞具有的規則特性將它們拆解成 sub-word 單元，再建構出一個詞群語言模型來描述它們和其他詞的關係，研究目標在於藉由增加 word lattice 的正確詞涵蓋率來降低 OOV(out-of-vocabulary)的影響，實驗結果顯示可以降低詞(word)、字(character)及基本音節(base-syllable)的絕對錯誤率分別達到 0.37%、0.27%及 0.26% (或是降低相對錯誤率達到 2.64%、2.56%及 3.38%)；其次，本論文探討語速對語音辨認的影響，做法是藉由建立一個語速控制的階層式韻律模型來描述語速對語音韻律聲學參數的影響，並將其用來協助語音辨認。實驗結果顯示所提出的考慮語速的語音辨認方法可以降低詞、字及基本音節的絕對錯誤率分別達到 1.67%、1.45%及 1.02% (或是降低相對錯誤率達到 12.25%、14.09%及 13.55%)，因此這是一個不錯的方法。

A Mandarin Speech Recognition System Incorporating with Speaking Rate Modeling and Word Construction

Student：Jun-Han Lin Advisor：Dr. Sin-Horng Chen

Institute of Communication Engineering
National Chiao Tung University

Abstract

The thesis presents a new Mandarin-speech recognition approach to considering the recognition of affix-words and the effect of speaking rate. First, the recognition of affix-words is realized via decomposing them into sub-word units. A class-based language model is then employed to describe their relations with other words. The study aims at decreasing the effect of out-of-vocabulary (OOV) words by increasing the coverage of the word lattice generated by a lexicon with size limited to 60,000. Experimental results showed the reductions of word, character, and base-syllable error rates by 0.37%, 0.27% and 0.26% absolutely (or 2.64%, 2.56%, and 3.38% relatively). Then, the effect of speaking rate on speech recognition is discussed. A speaking rate-dependent hierarchical prosody model which describes the influences of speaking rate on prosodic-acoustic features are constructed and used to assist in speech recognition. Experimental results showed that the approach of considering speaking rate in ASR leads to the reductions of word, character, and base-syllable error rates by 1.67%, 1.45% and 1.02% absolutely (or 12.25%, 14.09%, and 13.55% relatively). So, the proposed approach is very promising.

致謝

就讀研所兩年期間，非常感謝陳信宏老師時時刻刻關心我的研究進度，在研究上不斷地給與指導，引領我的研究方向，使我能夠順利畢業。感謝王逸如老師，碩一時教導我語音的基本觀念與做研究的態度與方法，使我從一個懵懵懂懂的大學生領悟到如何成為一位真正的研究生。

感謝振宇學長、智合學長、阿德學長與希群學長在研究上給予我許多幫助，一路指導我的研究，在我迷惘有困難時，願意花時間與我討論，提供建議與意見，使我能夠走向正確的道路上；感謝文良、大胖、小蝦、智障、豆腐、進竹、冠譯與銘傑學長，在研究上給予建議，生活上也非常照顧我們這屆學弟妹，由其感謝銘傑學長，感謝你一步一步地教導我，使我能夠迅速地了解我的研究；感謝人生勝利組的小邱、研究與籃球一把罩的 kiwi 以及又帥又白的睿詮，在研究上或生活上，都受到你們很大的照顧，感謝學識淵博的企鵝，觀念成熟的昂星，樂觀開朗的昌祐，好相處的雅婷以及多才多藝的秘書靖觀，感謝你們兩年來對我的包容；感謝下一屆學弟妹，很古意的阿龐，又聰明又會講話的子睿，排球超強、身材超好的奕勳，高手等級的良基與個性活潑的婉君，期許你們明年順利畢業。

最後感謝我的父母親，謝謝您們一路上的支持，使我在人生道路上走得更加堅定，一路堅持往前邁進，謝謝您們。

目錄

中文摘要 I

Abstract II

目錄 III

表目錄 VII

圖目錄 IX

第一章 緒論 1

1.1 研究動機 1

1.2 文獻回顧 1

1.3 研究方向 2

1.4 章節概要說明 3

第二章 階層式語言模型 4

2.1 語料庫簡介 4

2.2 聲學模型與語言模型之架構及建立 5

2.2.1 聲學模型之建立 5

2.2.2 文字資料庫介紹 6

2.2.3 辨識詞典選詞方式 7

2.2.4 語言模型的建立 9

2.3 綴詞語言模型的訓練 10

2.3.1 綴詞的選擇與拆解 10

2.3.2 綴詞語言模型之建立 10

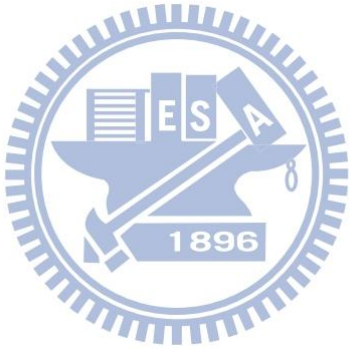
2.4 階層式語言模型辨認器 13

2.5 結果分析 13

2.5.1 語言模型評估 14

2.5.2 綴詞於 word lattice 上之涵蓋率 14
2.5.3 辨識效能結果 15
2.5.4 辨識結果之細部剖析 16
第三章 加入語速影響之韻律模型於中文大辭彙語音辨認系統 17
3.1 中文語音韻律模型階層式架構 17
3.2 階層式韻律模型之修正及設計 19
3.2.1 Break Syntax Model 21
3.2.2 韻律狀態模型 22
3.2.3 音節韻律模型 22
3.2.4 停頓聲學模型 24
3.3 加入韻律訊息於 two-stage 語音辨認系統 25
3.3.1 Joint Syntax Model 之架構與建立 26
3.3.2 特徵參數正規化 28
3.3.2.1 音節長度之語速正規化 28
3.3.2.2 停頓長度之語速正規化 29
3.3.2.3 音節音高軌跡之語速正規化 29
3.3.2.4 音節能量強度之語速正規化 30
3.3.3 The Second Stage 之實作 31
3.3.3.1 第一階段：加入多種語言資訊 31
3.3.3.2 第二階段：加入韻律邊界停頓資訊 32
3.3.3.3 第三階段：加入音節韻律狀態資訊 33
3.4 鑑別式模型組合 34
第四章 實驗結果與分析 36
4.1 階層式韻律模型之訓練 36
4.1.1 Break Syntax Model 36
4.1.2 停頓聲學模型 38
4.2 加入韻律訊息實現中文語音辨認 40

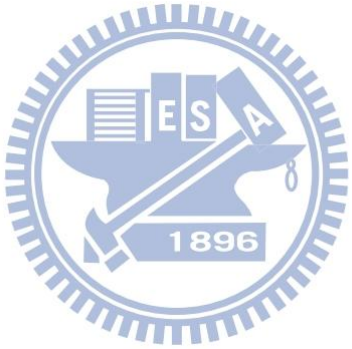
4.2.1 詞性(POS)辨認率算法.....42
4.2.2 標點符號(PM)辨認率算法.....42
4.3 辨識結果分析與比較.....42
4.3.1 各級辨認結果之比較.....43
4.3.2 傳統韻律模型與考慮語速影響之韻律模型之比較.....45
第五章 結論與未來展望.....46
5.1 結論.....46
5.2 未來展望.....46
參考文獻.....48
附錄：決策樹之問題集.....50



表目錄

表 2.1：TCC-300 語料庫統計表 5
表 2.2：MFCC 參數抽取設定檔 6
表 2.3：TF-IDF 方法剔除的詞條例 8
表 2.4：經由 TF-IDF 方法加入的詞條例 8
表 2.5：混淆度(Perplexity) 8
表 2.6：綴詞拆解範例 10
表 2.7：詞幹詞性範例 11
表 2.8：詞典涵蓋率 13
表 2.9：TCC300 測試語料的詞類統計 14
表 2.10：語言模型混淆度 14
表 2.11：word lattice 上綴詞涵蓋率 15
表 2.12：搭配語言模型之詞辨認率 15
表 2.13：搭配語言模型之字元辨認率 15
表 2.14：搭配語言模型之音節辨認率 15
表 2.15：word lattice 之詞、字及音節涵蓋率 16
表 2.16：綴詞辨識情況 16
表 3.1：韻律結構之停頓標記 19
表 3.2：韻律標記、聲學參數以及語言參數之數學符號 20
表 3.3：文本處理 27
表 4.1：詞(word)辨認率 40
表 4.2：字(character)辨認率 40
表 4.3：音節(syllable)辨認率 41
表 4.4：詞性(POS)辨認率 41

表 4.5：標點符號(PM)辨認率 41
表 4.6：搶詞狀況的改善 43
表 4.7：一字詞辨認的改善 44
表 4.8：聲調的修正 44
表 4.9：加入語速影響的結果改善 45



圖目錄

圖 2.1：語言模型訓練流程 6
圖 2.2：文本前處理流程 7
圖 2.3：新辨識路徑 11
圖 2.4：階層式系統架構 13
圖 3.1：中文語音韻律之階層式架構 [9] 17
圖 3.2：本研究所採用的階層式韻律架構 18
圖 3.3：以 two-stage 方式之韻律輔助中文語音辨認系統流程圖 25
圖 3.4：factored POS model 的 backoff 路徑 26
圖 3.5：factored PM model 的 backoff 路徑 27
圖 3.6：factored model 訓練架構流程圖 27
圖 3.7：辨認器第二級三階段實作流程圖 31
圖 4.1：break syntax model 的決策樹架構 37
圖 4.2：interword subtree 架構更深層部分 38
圖 4.3：(a)音節停頓時長，(b)正規化音節延長因子 1，(c)正規化音節延長因子 2，(d)音節間能量低點，(e)正規化基頻跳躍值之決策樹根節點之機率分佈圖 39

第一章 緒論

1.1 研究動機

近年來，電子產品不斷推陳出新，人們愈來愈注重其產品實用性與使用的便利性，而語音是人類最直接且最便利的溝通方式，以語音取代複雜的鍵盤輸入做為與機器溝通的媒介已經成為一種近代的發展趨勢。

針對中文語音辨識而言，中文詞彙不同於其它拼音語系，詞彙的邊界相當模糊、數量繁多且變化複雜，詞與詞之間也可再組合成新的詞彙，這些多元的變化造成詞彙定義上的困難，然而在傳統大詞彙語音辨認中，受限於詞典詞條數的限制，無法收錄所有中文詞彙可能的組合，使得詞典的涵蓋率不足，在進行語音辨認時，落在詞典外的詞彙(out-of-vocabulary, OOV)將無法被辨識出來。因此本研究將探討中文特性，針對若干具有較明顯的構詞規則的詞彙進行分析討論，藉此增進詞典的涵蓋率，進而解決 OOV 問題。

韻律訊息在口語中扮演著很重要的角色，能幫助人們辨認每一個詞以及整段句子的結構。其中說話速度更是一個很重要的韻律參數，它可以影響許多語音的現象，像是音節長度、停頓長度、基頻軌跡等等...，更有相關的研究指出，慢速或快速語音辨認時會造成較大的詞錯誤率(word error rate, WER)，故本研究將考慮語速對中文語音韻律的影響。

1.2 文獻回顧

在大詞彙語音辨認系統(large vocabulary continuous speech recognition, LVCSR)中，N 連語言模型(n-gram language model)最常被使用，其原理為藉由統計各種詞彙出現在文本的次數來描述詞與詞之間相接的機率，但隨著 N 值的升高，訓練語言模型時會發生資料稀疏(Data Sparseness)的問題，導致機率預估的不準確。而後為增進辨識系統效能，有許多學者提出方法來加強語言模型。其中 1992 年 Brown[1]等人將詞彙依據其特性進行分類，提出類別式 N 連語

言模型(Class-based n-gram language model)，透過加入類別資訊來訓練語言模型，則資料的預估由詞彙組合數降低為類別的組合數，進而改善資料稀疏的問題。

中文詞彙數量繁多且變化複雜，辨識詞典無法收錄所有中文詞彙可能的組合，Jou [2]、Yang [3]在其論文中提出階層式辨識系統，以構詞學的角度出發，針對人名、定量複合詞與詞綴三個類別依照其特性將之拆解進而提升詞的涵蓋率。

在自動語音辨識系統(Automatic Speech Recognition, ASR)上已有相關研究在探討語速對於辨識效能的影響：Siegler [4]等人發現快語速的語音會提高其辨識上的錯誤，並提出三種方法改進快語速語音的辨識準確率，分別是 Baum-Welch codebook 的調適、HMM 狀態轉移機率的調適以及藉由加入複合詞(Compound Words)以及利用規則法改善發音詞典，實驗結果顯示第二種方法降低了 4-6%的相對錯誤率。Martinez [5]等人討論語速與聲學參數的相關性，並將語速加入辨識考量中，實驗結果顯示慢語速的部分詞錯誤率可以有效地降低。Pfau [6]提出語速正規化的方法，其原理為透過動態調整音素(phone)長度，來消除語速對辨識系統的影響。

1.3 研究方向

本研究首先在語言模型層進行改良，更改選詞方式並考慮中文詞彙中若干具有規則特性的詞彙，藉由語法資訊進行拆解，以便提升詞典涵蓋率、降低 OOV 的問題。第一步將產生一組高涵蓋率的詞圖(word lattice)，第二步在此涵蓋率高的 word lattice 上將符合規則的較小單元重新進行構詞，並利用更精細的語言模型重新配置語言模型的分數，藉此壓抑不合理的路徑，加強可靠的路徑的機率值。

本研究將使用韻律模型，利用 Jiang [7]提出的非監督式中文語音韻律標記及韻律模式(Unsupervised Joint Prosody Labeling and Modeling, PLM)演算法，從大量未經標記的語料中訓練各種韻律模型，描述韻律邊界停頓、音節韻律狀態這兩種韻律標記與語言參數及韻律聲學參數三者之間的關係，並加入語速變數將其引入韻律模型當中，使其與語速相關，最後應用到語音辨認中，期許能進一步提升辨識效能，並解碼出詞(Word)序列、詞性(POS)序列、標點符號序列(PM)等多種語言參數序列及代表韻律架構的兩種韻律標記資訊。

1.4 章節概要說明

本論文一共分為五章，其各章節內容分配如下：

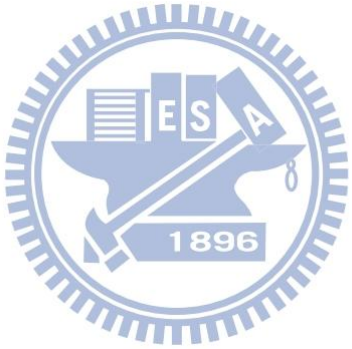
第一章：緒論。

第二章：階層式語言模型。

第三章：加入語速影響之韻律模型於中文大詞彙語音辨認系統

第四章：實驗結果及分析。

第五章：結論與未來展望。



第二章　階層式語言模型

基本語音辨認器中包含聲學模型與語言模型，聲學模型是以隱藏式馬可夫模型(HMM, Hidden Markov Model)呈現，透過該機率模型，描述發音過程之狀態轉移現象和輸出結果，語言模型則經由大量文字資料訓練出一個涵蓋範圍廣泛、適用於各種領域的語言模型，期許在加入語言模型幫助下提升語音辨識率。

然而在傳統語音辨識上，辨識率一直受限於詞典大小，而中文詞彙數量繁多且變化複雜，詞與詞之間也可再組合成新的詞彙，在這其中許多詞類為 open set，諸如數詞(Neu)、專有名詞(Nb)、綴詞(MD)…等，由於詞典有其詞彙量的限制下，無法收錄所有中文詞彙可能的組合，使得詞典的涵蓋率降低，落在詞典外的詞彙將無法被辨識出來，語音辨識效能因此成長有限。

為突破此困境，本研究針對 open set 中具有較明顯的構詞規則的若干詞彙進行處理，綴詞富有其規則特性，可視為「詞幹(stem)」與「詞綴」的組合，故將其拆解成較小的單位，以收錄較少的數量來降低 OOV 的問題，在此我們將每一種詞綴各自視為不同的類別，訓練出詞綴類別與前後詞之間的機率，加上對各自類別內的單元組合訓練出語言模型，最後整合這兩個語言模型經由重新評分得出最佳的辨識結果並進一步減少一字詞的錯誤率。

2.1 節中將對實驗所使用之語料庫做基本的介紹；2.2 節中將簡介聲學模型與語言模型之架構及建立；2.3 節將介紹如何針對綴詞訓練出語言模型；2.4 節將介紹辨認系統的整體架構；2.5 節將分析此系統的辨識結果。

2.1 語料庫簡介

本研究是使用 TCC-300 麥克風語音資料庫，它由國立台灣大學、國立成功大學及國立交通大學所共同錄製，此語料庫屬於麥克風朗讀語音，檔案統計資料如表 2.1 所示。每個學校之語句取樣頻率皆為 16,000 赫茲(Hertz)，取樣位元數為 16 位元。音檔檔頭為 4096 位元組(byte)，副檔名為*.vat。

表 2.1：TCC-300 語料庫統計表

學校名稱	文章屬性	語者總數		總音節數		音檔總數	
台灣大學	短文	男	50	男	27541	男	3425
		女	50	女	24677	女	3084
		總數	100	總數	52218	總數	6590
交通大學	長文	男	50	男	75059	男	622
		女	50	女	73555	女	616
		總數	100	總數	148614	總數	1238
成功大學	長文	男	50	男	63127	男	588
		女	50	女	68749	女	582
		總數	100	總數	131876	總數	1170

依據表 2.1，本研究會對上述內容以長句為主的語料庫分為訓練語料及測試語料，其中訓練語料的部分大約占 90%，共包含 274 位語者，長度一共約 23 小時，音節總數量為 300,836；測試語料的部分大約占 10%，共包含 19 位語者，長度一共約 2 小時，詞總數量為 15,461，音節總數量則為 26,357，此外本研究所使用的韻律模型，是從訓練語料中挑選 164 位語者，音檔長度 8.3 小時，音節總數量為 106,955 的語料來進行訓練。

2.2 聲學模型與語言模型之架構及建立

2.2.1 聲學模型之建立

由於語音訊號在頻譜上具有短時間穩定的特性及考慮到人耳聽覺效應的補償作用，本研究使用的參數為 MFCC (Mel-Frequency Cepstral Coefficients, 梅爾倒頻譜參數)，以 32 毫秒之漢明窗(Hamming Window)及每次位移 10 毫秒取出一筆資料，其成分包含 12 維 MFCC 加上 1 維能量共 13 維，取其一階變化量(delta term)和二階變化量(delta-delta term)，最後扣掉能量參數，參數一共 38 維做為本研究之發音聲學參數。其系統相關設定如表 2.2 所示。此外，聲學模型

為 411 個音節，每一個音節使用 8 個狀態的隱藏式馬可夫模型(HMM)，使用 MMI 鑑別式訓練得到。

表 2.2：MFCC 參數抽取設定檔

音框長度	32ms
音框平移	10ms
Filter bank 個數	24
取樣頻率	16kHz
Pre-emphasis Filter	First order with coefficient 0.97

2.2.2 文字資料庫介紹

辨認系統之語言模型，通常必須先具備大量的文字資料庫，利用大量的文字資料訓練出一個涵蓋範圍廣泛、適用於各個領域的語言模型，本研究使用的文字資料庫共有下述四種來源：

1.)光華雜誌(Sinorama)：內容為一般雜誌的文章，蒐集的年代範圍介於 1976 年到 2000 年之間。

2.) NTCIR：為一個建立資訊檢索系統的標竿測試集，其內容由數種不同學科領域文章構成。

3.)中研院平衡語料庫(Sinica)：它是一套由中研院錄製，內容包含多種主題，以語言分析研究為目的的資料庫。

4.) Chinese Gigaword：由 Linguistic Data Consortium (LDC)整合發行，內容包含台灣中央社、北京新華社等國際新聞。

在訓練語言模型之前，須先對語料庫的文章進行前處理，將文章中會影響辨認效能的內容移除或修改，經由文本前處理後，得到詞的總數量為 382,921,251 個，之後再以統計方式選擇詞典，這裡一共納入了 60,000 個常見詞彙，將常出現、較重要的詞收錄在詞典內以便訓練出語言模型，圖 2.1 為語言模型訓練流程：

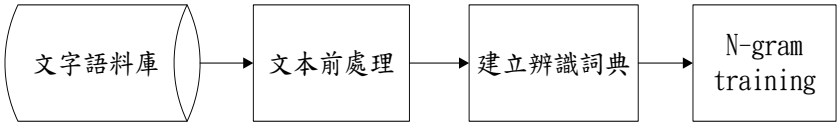


圖 2.1：語言模型訓練流程

其中文本前處理的步驟又可以再細分為以下數個步驟：

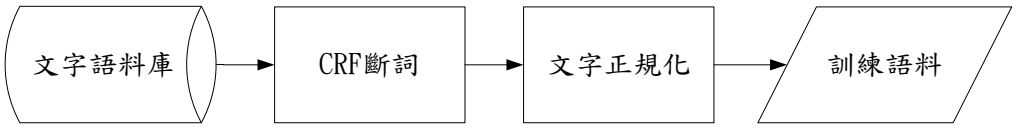


圖 2.2：文本前處理流程

2.2.3 辨識詞典選詞方式

對於建立一個完善的語言模型而言，有一項重要的關鍵在於詞典的選擇，由於受限於記憶體的大小，有別於傳統方式為直接收錄語料庫中高詞頻的六萬筆詞條，但某些詞條可能只出現在特定文章中，在此為了讓收錄的詞條為一般常見的詞語，也就是說：必須找到廣泛出現在各個文章中的詞，因此我們使用 TF-IDF (term frequency–inverse document frequency)方法來幫助我們進行選詞。

TF-IDF 是一種用於資訊檢索(IR - Information Retrieval)的常用加權技術。它是一種統計方法，用於評估一個詞對於一個文件集或一個語料庫中的其中一份文件的重要程度。TF 表示該詞條在語料庫中出現的頻率，代表其詞條的重要性隨著在語料庫出現的總次數成正比增加；IDF 則表示一個詞條普遍重要性的度量，代表其詞條類別區分能力隨著在語料庫各文章中出現的頻率成反比下降。

我們可以使用(2.1)式算出每個詞條對應的 IDF 值：

$$idf_i = \log \frac{|D|}{|\{d : d \ni t_i\}|} \tag{2.1}$$

(2.1)式中 D 表所有文件的集合，分子 $|D|$ 表示語料庫中的文件總數，d 表文件，t_i 表正在處理的詞條，分母則表示包含該詞條 t_i 的文件數目。

在進行大詞彙辨認時，我們欲收錄的詞是一般常見的詞語，觀察(2.1)式中，由於“出現的文章數”在分母項，因此我們將挑選 IDF 分數低的詞收錄進詞典中。

比較藉由 TF-IDF 方法重新計算收錄的優先順序與直接收錄高詞頻的傳統方式，發現使用

TF-IDF 選詞而更動的詞數大多為人名，這些人名因為本身出現次數高而被收錄進原先高詞頻的詞典中，但因僅僅出現在少數的文件中而遭到 TF-IDF 方法剔除，如下表 2.3 與表 2.4 所示，由此可知，TF-IDF 選詞方式能使詞典收錄到較廣泛的詞條。表 2.5 並比較直接收錄高詞頻的傳統方式與經由 TF-IDF 選詞方式來計算其混淆度(Perplexity)，計算的對象為 TCC-300 的測試語料，發現經由 TF-IDF 選詞方式能夠有較低的混淆度。

表 2.3：TF-IDF 方法剔除的詞條例

被剔除的詞	詞頻	出現文章數	IDF 值
黃乃宣	996	63	2.124
韋殿剛	1013	69	2.085
吳憶樺	1081	56	2.175

表 2.4：經由 TF-IDF 方法加入的詞條例

取代的新詞	詞頻	出現文章數	IDF 值
協商會	147	121	1.841
酒氣	147	121	1.841
崇洋	147	118	1.852

表 2.5：混淆度(Perplexity)

Lexicon	Order	ppl	ppl1
傳統方式	3	109.712	117.745
TF-IDF 方法	3	109.311	117.301

2.2.4 語言模型的建立

所有的語言都有其文法規則，利用文法規則所建立出的機率模型稱為語言模型。在大詞彙連續語音辨認時，會利用語言模型，考慮前後詞彙的關連性，期望能使輸入的語音辨認出合理且有意義的詞串。在本研究中使用了日前運用廣泛的 *n*-gram 語言模型，此模型假設任一個詞在詞串中只受到前 *n-1* 個詞的影響。

假設有一詞串共有 N 個詞，也就是 $W = w_1 w_2 ... w_N$ ，其中「w_i」代表句子中的第 i 個詞，則產生這個句子所對應的機率，可以拆解成以下一連串的條件機率之連乘：

$$P(W) = P(w_1) \cdot P(w_2 \mid w_1) ... P(w_i \mid w_{i-n+1} ... w_{i-1}) ... P(w_N \mid w_{N-n+1} ... w_{N-1}) \tag{2.2}$$

where

$$P(w_i \mid w_{i-n+1}, ..., w_{i-1}) = \frac{Count(w_{i-n+1}, ..., w_i)}{Count(w_{i-n+1}, ..., w_{i-1})} \tag{2.3}$$

由於 *n*-gram 語言模型是統計式的模型，如果訓練語料中沒出現該詞語組合，就無法預估其機率，且隨著 n 值上升，所需的訓練語料也呈指數成長。為了解決這些問題，我們以後撤平滑化(back-off smoothing)來調整模型的機率分佈。機率預估式改寫如下：

$$P(w_i \mid w_{i-n+1}, ..., w_{i-1}) = \begin{cases} a(w_{i-n+1}, ..., w_{i-1}) P(w_i \mid w_{i-n+2}, ..., w_{i-1}), & Count(w_{i-n+1}, ..., w_i) = 0 \\ d_a \cdot \dfrac{Count(w_{i-n+1}, ..., w_i)}{Count(w_{i-n+1}, ..., w_{i-1})}, & 1 \le Count(w_{i-n+1}, ..., w_i) \le k \\ \dfrac{Count(w_{i-n+1}, ..., w_i)}{Count(w_{i-n+1}, ..., w_{i-1})}, & Count(w_{i-n+1}, ..., w_i) > k \end{cases} \tag{2.4}$$

(2.4)式中後撤加權值 $a(w_{i-n+1}, ..., w_{i-1})$ 需經過正規化(normalization)處理，且滿足以下條件式：

$$\sum_{w \in V} P(w_i = w \mid w_{i-n+1}, ..., w_{i-1}) = 1 \tag{2.5}$$

另外，當 $Count(\cdot)$ 的數值很小時，可能造成預估的不準確，則將原始的 *n*-gram 機率乘上一個小於 1 的值 d_a (Discount Coefficient Factor)來進行平滑化，d_a 依據 Good-Turning discounting 計算得出，並會將扣除的機率值再平分給詞串沒有出現的 *n*-gram 機率使用。

2.3 綴詞語言模型的訓練

2.3.1 綴詞的選擇與拆解

綴詞的組合結構可拆解為「詞幹」與「詞綴」的組合，拆解方式如表 2.6 所示，依中研院所統計的詞綴數量太過龐大，在此我們參照中文資訊處理分詞規範中，僅收錄常出現的衍生詞綴、語法詞綴與名詞性接尾詞，總計共 148 個後詞綴。

為避免辨認詞典收錄過多的短詞，反而犧牲掉原先未經拆解即可收錄之高詞頻一般詞的空間，故將詞頻高的詞彙保留其長詞型式。在此綴詞將依照詞頻高低以兩種型式收錄於詞典當中：第一種為出現次數於前 50,000 詞內的高詞頻綴詞直接收錄於詞典；第二種為出現次數於 50,000 詞之外的綴詞均拆解為詞幹與綴詞收錄在詞典裡。

表 2.6：綴詞拆解範例

綴詞	詞幹	詞綴
靈敏度	靈敏	度
視覺系	視覺	系
拋棄式	拋棄	式

經由上述拆詞過程之後，在原先出現次數 50,000 到 60,000 的詞彙中，一共拆解了 957 個綴詞，其中 919 個詞幹已收錄在前 50,000 詞，屬於高詞頻一般詞部分，故在此綴詞 subword 只新增了 38 個，而其詞典剩餘空間則收錄原先不在詞典中其餘高詞頻一般詞，直至到達詞典容量上限為止，藉此提高詞典涵蓋率，降低 OOV 的數量。

2.3.2 綴詞語言模型之建立

綴詞的分群原則是以詞綴做為區分，相異的綴詞各有其不同的涵義，與詞綴相關聯的詞幹也有其相關詞性(POS)，其詞性如表 2.7 所示，故針對目前所收錄的 148 個常見詞綴各自視為

不同的類別比視為一個大類別來的適當，以其訓練詞彙間的機率來的更加可靠。

表 2.7：詞幹詞性範例

綴詞類別	詞幹詞性(46 類)	範例
們	Na	親友們、情侶們
記、物	Na、VA、VB、VC、VE、VH、VK	復仇記、漂遊記
賽、會、制、式	Na、VA、VC、VE、VG	邀請賽、責任制

本研究中會先對目前所收錄的 146 個常見詞綴建構出一個綴詞表，之後在 word lattice 上透過查表方式(綴詞表與詞性表)判斷哪些詞彙可以重新構回綴詞。如圖 2.3 觀察到當 lattice 上的「鄉親」與「們」經由綴詞表發現可組合成「鄉親們」並且「鄉親」符合該詞綴類別的詞性表，此時將會產生新的辨識路徑，而後將重新計算該詞與前後詞彙間的機率，即統計圖 2.3 中新路徑實線的機率分佈。

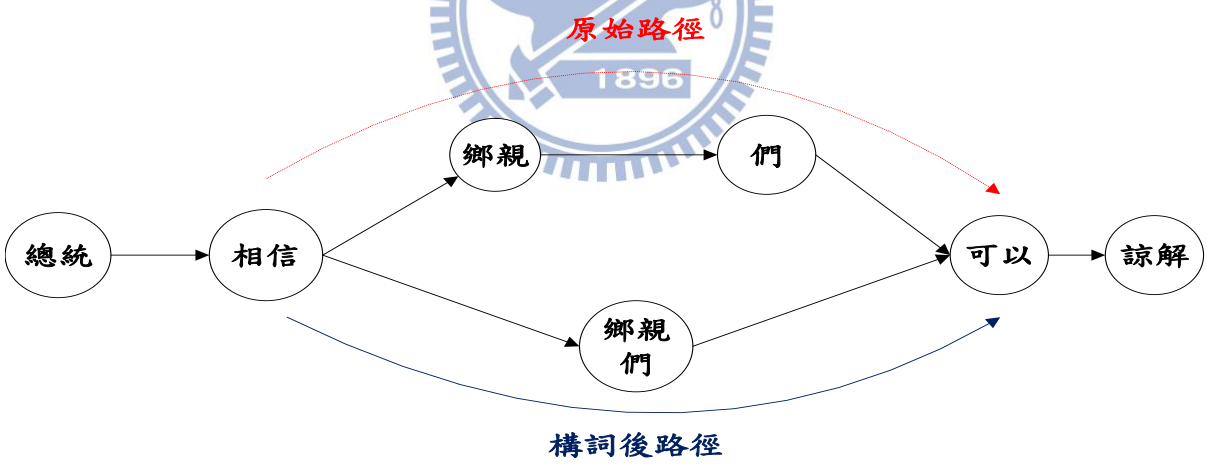


圖 2.3：新辨識路徑

我們依 148 個常見詞綴各自分為 148 個類別，將具有相同詞綴的新詞放置同一類別中，綴詞則依據其詞綴代表的類別做取代，最後透過一般詞與類別間的機率與在其類別內出現新詞之機率，藉由這兩種外部機率(Inter-word probability)和內部機率(Intra-word probability)來重新配置語言模型分數。原先 tri-gram 的機率預估式如下式(2.6)所示，bi-gram 與 uni-gram 類推之：

$$P(W) = P(w_1) \cdot P(w_2 \mid w_1) \cdot \prod_{i=3}^{N} P(w_i \mid w_{i-1}, w_{i-2}) \tag{2.6}$$

利用下式(2.7)將重新計算綴詞語言模型分數：

$$P(W_n \mid W_{n-2}W_{n-1}) = \begin{cases} P(C_n \mid W_{n-2}W_{n-1}) \cdot P(W_n \mid C_n) & \text{if } W_n \in MD \\ P(W_n \mid W_{n-2}W_{n-1}) \cdot 1 & \text{otherwise} \end{cases} \tag{2.7}$$

以下我們將深入探討這兩種外部機率和內部機率的計算方式：

1.) 外部機率(Inter-word probability)的預估： $P(C_n \mid W_{n-2}W_{n-1})$

我們將綴詞依據其詞綴代表的類別標記，透過統計類別與前後詞彙的關聯性，採用 tri-gram 模型預估之，可得到所有詞彙的機率值，在此會產生 8 種不同的情況，包括如：詞與類別之間、類別與類別之間、詞與詞之間的 class tri-gram 機率，如下式(2.8)所示。

$$P(W_n \mid W_{n-2}W_{n-1}) = \begin{cases} P(C_n \mid C_{n-2}C_{n-1}) & \text{if } W_n, W_{n-1}, W_{n-2} \in MD \\ P(C_n \mid C_{n-2}W_{n-1}) & \text{if } W_n, W_{n-2} \in MD \\ P(C_n \mid W_{n-2}C_{n-1}) & \text{if } W_n, W_{n-1} \in MD \\ P(C_n \mid W_{n-2}W_{n-1}) & \text{if } W_n \in MD \\ P(W_n \mid C_{n-2}C_{n-1}) & \text{if } W_{n-1}, W_{n-2} \in MD \\ P(W_n \mid C_{n-2}W_{n-1}) & \text{if } W_{n-2} \in MD \\ P(W_n \mid W_{n-2}C_{n-1}) & \text{if } W_{n-1} \in MD \\ P(W_n \mid W_{n-2}W_{n-1}) & \text{otherwise} \end{cases} \tag{2.8}$$

2.) 內部機率(Intra-word probability)的預估： $P(W_n \mid C_n)$

統計綴詞在該所屬類別內出現的機率，在此採用 uni-gram 模型預估之，搭配 good-turing smoothing 作為此內部機率，若為高詞頻的綴詞與一般詞情況下，則此內部機率為 1，如下式(2.9)所示。

$$P(W_n \mid C_n) = \begin{cases} P(W_n \mid C_n) & \text{if } W_n \in MD \\ 1 & \text{otherwise} \end{cases} \tag{2.9}$$

2.4 階層式語言模型辨認器

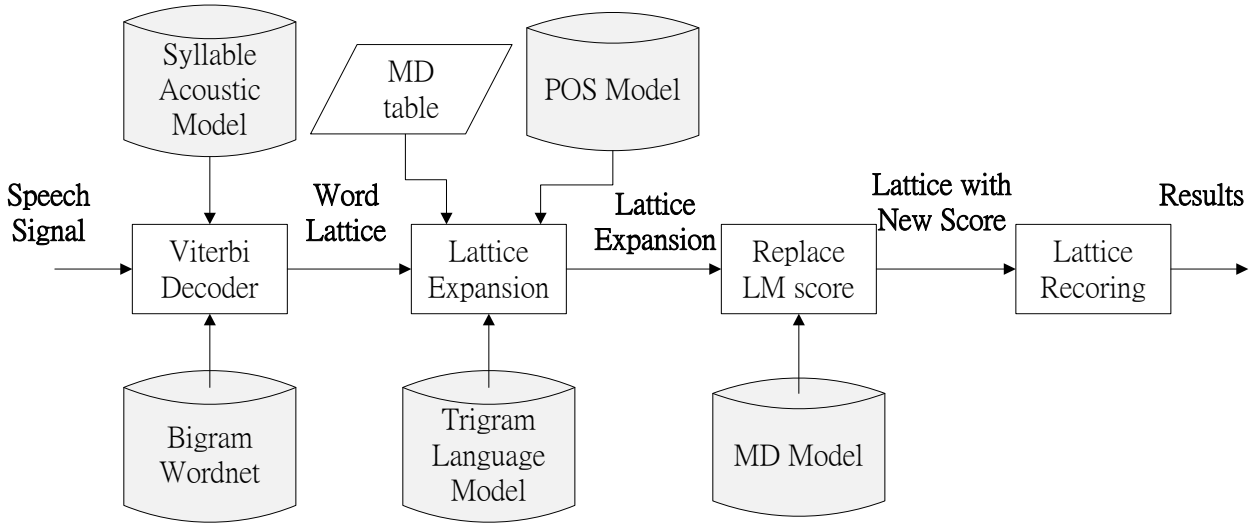


圖 2.4：階層式系統架構

圖 2.4 為此階層式系統的整體架構，首先第一級中先利用文字規則特性找出綴詞的文字規則並依據此規則針對綴詞進行拆解以便提高詞典涵蓋率，如表 2.8 所示，之後在 word lattice 上透過查詞綴表方式與判斷詞幹本身所屬的詞性查詢哪些詞彙可再構回成綴詞，避免產生「過份構詞」導致錯誤辨識結果的發生性。

表 2.8：詞典涵蓋率

拆詞前	97.13%
拆詞後	97.17%

第二級中本研究首先對綴詞建立一個更精細的語言模型，採用 class-based approach 的構想，對綴詞進行分類，依據不同詞綴建立不同的類別，將綴詞分類細緻化，同類別內的詞彙共用相同的外部機率，解決了部分詞彙出現次數稀疏的問題，最後在 word lattice 上，將分數替換成此語言模型的分數，進行重新辨識找出最佳的辨識結果。

2.5 結果分析

在這小節中，我們將以三種方式來評估傳統式語言模型(經由 TF-IDF 選詞方式選出六萬詞，

但未對綴詞進行處理)與本研究所提出的階層式語言模型對於綴詞在辨認上的影響並進行說明。

首先，本研究針對 TCC300 測試語料中綴詞數量與其所占的比例，由此可以先行知道測試語料中綴詞的正確辨識數量及上限。

表 2.9：TCC300 測試語料的詞類統計

TCC300 測試語料(226 個音檔)		
	詞條數	所佔比例
綴詞	309	2%
total words	15461	100%

2.5.1 語言模型評估

我們以混淆度(Perplexity, PPL)來評估傳統式語言模型與第一級語言模型(拆詞過後)的複雜程度，藉此判斷語言模型的好壞。語言模型帶有大量詞彙和詞彙相接的機率資訊，藉由透過這些機率資訊來預估下一個詞彙，PPL 值高代表語言模型需要較多的預估次數才會命中，較不易找到正確答案；反之，PPL 值低代表其語言模型較為單純，不需要過多的預估次數即可找出正確答案，故於 PPL 值越低時進行語音辨認，語言模型可能會呈現較好的辨識效能。表 2.10 中計算的對象為 TCC-300 的測試語料，可以發現到第一級語言模型擁有較低的混淆度。

表 2.10：語言模型混淆度

語言模型	Order	ppl	ppl1
傳統式語言模型	3	109.311	117.301
第一級語言模型	3	109.126	117.099

2.5.2 綴詞於 word lattice 上之涵蓋率

針對傳統式語言模型與第一級語言模型產生的 word lattice 進行分析，統計在理想最佳路

徑上有多少數量可以辨識回原先的綴詞，並加以計算其涵蓋率。

表 2.11：word lattice 上綴詞涵蓋率

語言模型	TCC300	最佳路徑	涵蓋率
傳統式語言模型	309	298	96.44%
第一級語言模型	309	305	98.71%

2.5.3 辨識效能結果

以下我們比較傳統式語言模型與階層式語言模型之辨識效能，分別以詞、字元、音節為辨識單元來評估其辨識效能，並由第一級語言模型產生出來的 word lattice 觀察詞、字及音節之涵蓋率。

表 2.12：搭配語言模型之詞辨認率

	Deletion	Substitution	Insertion	Accuracy (%)
傳統式語言模型	269	1531	364	86.00%
階層式語言模型	273	1467	368	86.37%

表 2.13：搭配語言模型之字元辨認率

	Deletion	Substitution	Insertion	Accuracy (%)
傳統式語言模型	206	2453	138	89.44%
階層式語言模型	199	2384	142	89.71%

表 2.14：搭配語言模型之音節辨認率

	Deletion	Substitution	Insertion	Accuracy (%)
傳統式語言模型	212	1705	145	92.21%
階層式語言模型	205	1641	148	92.47%

表 2.15：word lattice 之詞、字及音節涵蓋率

詞(word)	93.72%
字(character)	93.72%
音節(syllable)	94.54%

2.5.4 辨識結果之細部剖析

在辨識結果中對綴詞進行分析，觀察傳統式做法與本研究所提出之方法對於綴詞在辨認上的影響並進行說明，在此將對辨認結果分成三種不同情況來進行討論：

1.) Case A：正確答案應為綴詞，但辨識結果錯誤的部分。

2.) Case B：正確答案不為綴詞，但辨識結果判斷成綴詞的部分。

3.) Case C：辨識答案辨認出正確答案的部分。

表 2.16：綴詞辨識情況

	Case A	Case B	Case C
傳統式語言模型	27	6	282
階層式語言模型	6	11	303

觀察表 2.16 可得知，階層式語言模型比起傳統式語言模型多增加 21 個綴詞被正確地辨識出來，雖然階層式語言模型也增加了 5 個判斷成錯誤綴詞的情況(Case B)，但整體而言，綴詞的辨認率(i.e.,（Case C－Case B)/(綴詞總數 309))為 94.5%，比起使用傳統式語言模型的 89.3%增加許多，由此可知，為綴詞建立其語言模型是有用的，不僅僅幫助綴詞本身之辨識而其鄰近詞也更容易辨識正確。

第三章加入語速影響之韻律模型於中文大辭彙語音辨認系統

所謂韻律就是連續語音中一種跨區段(supra-segmental)的特徵，其主要表現在語速(speaking rate)、停頓時長(pause duration)、音高軌跡(pitch contour)、音量大小(energy level)等因素上，本研究考慮語速變數並且引入韻律模型當中，使其與語速相關，因此在本章節中，將會探討以下幾點: 3.1 節介紹中文語音韻律階層式架構；3.2 節介紹本研究所使用的韻律模型；3.3 節將會說明如何整合韻律模型以 two-stage 的方式加入到語音辨認中；3.4 節說明本研究如何使用鑑別式組合(Discriminative Model Combination)處理重新評分過程中多個模型之權重問題。

3.1 中文語音韻律模型階層式架構

依據語言學家的研究發現[8,9]語音的韻律結構呈現階層式架構，[9]提出一個 5 階層的中文語音韻律結構，如圖 3.1 所示：

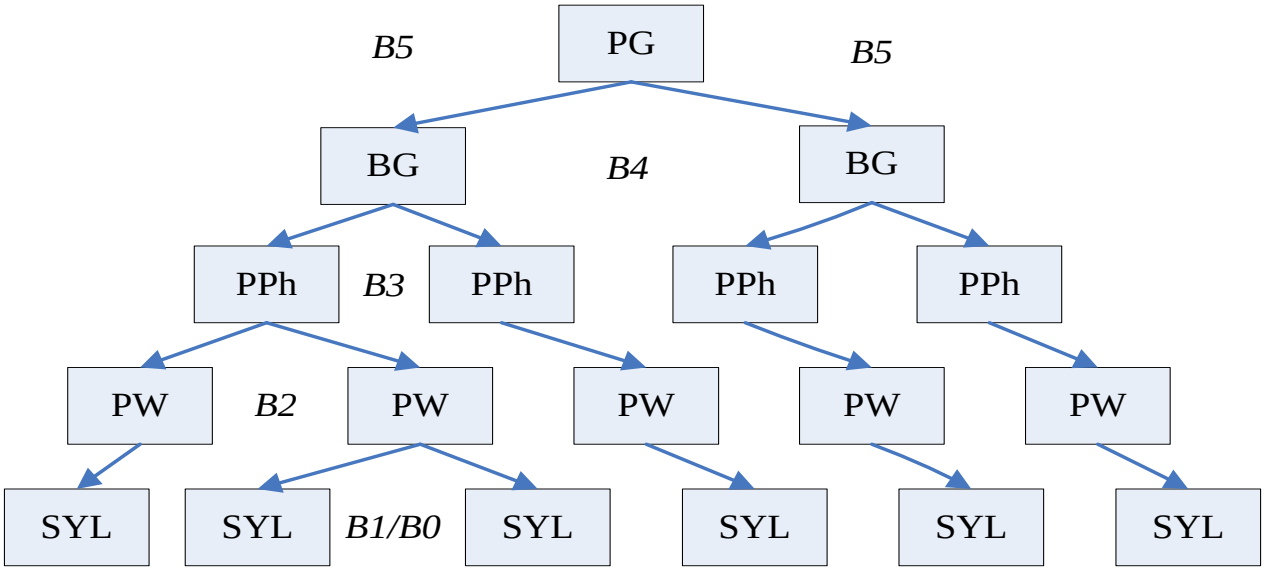


圖 3.1：中文語音韻律之階層式架構 [9]

圖中最底層為音節層次(syllable layer, SYL)，由於中文特性為一個音節一個字，故最底層的韻律單元為音節；向上發展依序為韻律詞層次(prosodic word layer, PW)，由雙音節或多音節所構成的詞組，此詞組通常在句法和語意上關係緊密；韻律短語層次(prosodic phrase layer, PPh)，由一個或多個韻律詞所組成，結尾常會帶有不明顯但可察覺之停頓；呼吸組層次(breath group, BG)，代表一個有音高及音長明顯變化的段落；最上層為韻律組句(prosodic phrase group)，由連續的呼吸組構成。這整體架構統稱「階層式多短語韻律句群(Hierarchical Prosodic Phrase Grouping，HPG)」架構[9]。

本研究以 HPG 架構為基礎，再進一步對其做修改，使用如圖 3.2 的 4 層結構，並使用兩種韻律標記來代表這階層式的韻律架構，第一種是韻律邊界停頓標記，它是用來區分階層式韻律架構中的各層韻律組成份子，如上圖 3.2 所示，本研究將 B1 細分為 B1-1、B1-2，其中 B1-1 代表 normal syllable boundary，不具有明顯停頓，B1-2 則代表詞內的音節邊界有較明顯的停頓。B2 細分為 B2-1、B2-2、B2-3，其中 B2-1、B2-2、B2-3 分別代表明顯音高位置(pitch reset)之韻律詞邊界、短停頓(short pause)之韻律詞邊界以及含有音節拉長效應(duration lengthening)後的韻律詞邊界。接著將 B4、B5 合併為 B4，因為其描述的韻律特性相當相似，於是整個架構從圖 3.1 的 5 層結構變回圖 3.2 的 4 層結構。

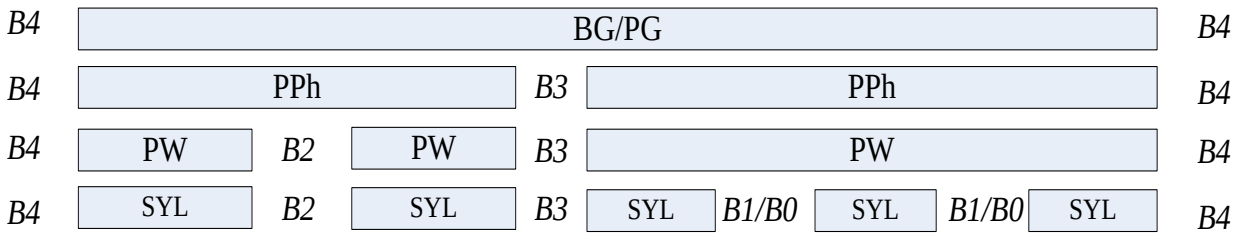


圖 3.2：本研究所採用的階層式韻律架構

本研究採用這 8 種韻律邊界停頓(break type) **B**={B0, B1-1,B1-2, B2-1, B2-2, B2-3, B3, B4} 來標記四種韻律單元：音節(SYL)、韻律詞(PW)、韻律短語(PPh)、呼吸組/韻律句組(BG/PG)，其對應關係如表 3.1 所示。

表 3.1：韻律結構之停頓標記

韻律結構	停頓標記	意義
韻律群(PG)	B3	長停頓
或呼吸群(BG)	B4	長停頓且含有明顯的基頻跳躍
	B2-1	相鄰兩音節具有明顯的基頻跳躍
韻律詞(PW)	B2-2	短停頓
	B2-3	前一音節發生音節拉長
	B0	音節邊界相鄰兩音節是緊密連接(tightly coupling)
音節(SYL)	B1-1	音節邊界相鄰兩音節是普通連接(tightly coupling)
	B1-2	詞內的音節邊界有較明顯的停頓

至於另一種韻律標記是韻律狀態，針對音節音高、音長及能量共分成三種類型，用以描述韻律架構中高層次組成份子對於音節韻律資訊帶來的影響。

3.2 階層式韻律模型之修正及設計

在本研究中，要以能幫助於語音辨認的前提之下設計韻律模型，主要任務是在給定聲學參數 $\Lambda_a=\{\mathbf{X}_a,\mathbf{X}_p\}$ 的條件下，找出最佳的語言參數序列 $\Lambda_l=\{\mathbf{W},\mathbf{POS},\mathbf{PM}\}$、韻律標記 $\Lambda_p=\{\mathbf{B},\mathbf{P}\}$ 及 acoustic segmentation Υ_s，在數學式中可視為一種求取最佳參數解的過程，如下式(3.1)所示：

$$\begin{aligned}\Lambda_l^*,\Lambda_p^*,\Upsilon_s^* &= \arg\max_{\Lambda_l,\Lambda_p,\Upsilon_s} P(\mathbf{W},\mathbf{POS},\mathbf{PM},\mathbf{B},\mathbf{P},\Upsilon_s\mid\mathbf{X}_a,\mathbf{X}_p)\\ &= \arg\max_{\Lambda_l,\Lambda_p,\Upsilon_s} P(\mathbf{W},\mathbf{POS},\mathbf{PM},\mathbf{B},\mathbf{P},\Upsilon_s,\mathbf{X}_a,\mathbf{X}_p)\end{aligned} \tag{3.1}$$

(3.1) 式中 $\mathbf{W}=\{w_1^M\}$ 是代表詞序列；$\mathbf{POS}=\{pos_1^M\}$ 是詞所對應到的詞性序列；至於 $\mathbf{PM}=\{pm_1^M\}$ 是代表標點符號序列；M 代表詞的全部數量；$\mathbf{B}=\{B_1^N\}$ 則是韻律邊界停頓標記序列，它包含八種韻律邊界停頓標記：$B_n\in\{B0, B1\text{-}1, B1\text{-}2, B2\text{-}1, B2\text{-}2, B2\text{-}3, B3, B4\}$，用來表示階層式韻律架構中的各層韻律組成份子的邊界；$\mathbf{P}=\{\mathbf{p},\mathbf{q},\mathbf{r}\}$ 則是音節韻律狀態序列，其代表的意義分別為音節音高軌跡 $\mathbf{p}=\{p_1^N\}$、音節長度 $\mathbf{q}=\{q_1^N\}$ 及音節能量強度 $\mathbf{r}=\{r_1^N\}$；N 代表音節的

全部數量；$\mathbf{X}_a$ 代表一個 frame-based 頻譜參數序列(i.e., MFCCs 及它們的一階和二階 derivatives)；$\mathbf{X}_p = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$則是一個韻律聲學參數序列，其中 **X** 代表音節韻律聲學參數，包含了音節音高軌跡(**sp**)、音節能量強度(**se**)及音節長度(**sd**)；**Y** 代表音節邊界參數，包含了音節間的停頓長度(**pd**)及音節間的能量低點(**ed**)；Z 代表差分韻律參數(differential prosodic feature)，包含了正規化的音節內基頻差(**pj**)及兩種正規化長度拉長因子 (**dl** and **df**)。其完整符號表整理於表 3.2。

表 3.2：韻律標記、聲學參數以及語言參數之數學符號

T : prosodic tag	**B** : break type={**B0, B1-1,B1-2 ,B2-1, B2-2, B2-3, B3, B4**}	
	PS : prosodic state	**p** : pitch prosodic state **q** : duration prosodic state **r** : energy prosodic state
A : prosodic feature	**X** : syllable prosodic feature	**sp** : syllable pitch contour **sd** : syllable duration **se** : syllable energy level
	Y : inter-syllabic prosodic feature	**pd** : pause duration **ed** : energy-dip level
	Z : differential prosodic features	**pj** : normalized pitch jump **dl** : normalized duration lengthening factor 1 **df** :normalized duration lengthening factor 2
	SR : speaking rate	
L : linguistic feature	**l**: reduced linguistic feature set **t**: syllable tone sequence **s**: base-syllable type **f**: final type	

以下本研究將會提出下列五種假設，以方便設計韻律模型：

假設一：如同傳統的聲學模型，頻譜參數序列 $\mathbf{X}_a$ 只會相依於詞序列 **W**。

假設二：韻律聲學參數序列 $\mathbf{X}_p$ 與韻律標記序列 Λ_p 及語言參數序列 Λ_l 相依。

假設三：音節韻律聲學參數序列 **X** 與音節邊界韻律參數序列 **Y** 及差分韻律參數序列 **Z** 彼此間相互獨立。

假設四：韻律邊界停頓標記序列 **B** 相依於鄰近相關的語言參數序列 Λ_l。

假設五：音節韻律狀態序列 **P** 相依於鄰近的韻律邊界停頓標記 **B**。

經由上述五種假設後，(3.1)式將會簡化成以下形式：

$$\Lambda_l^*, \Lambda_p^*, \Upsilon_s^* \approx \arg\max_{\Lambda_l, \Lambda_p, \Upsilon_s} \{P(\mathbf{X}_a, \Upsilon_s|\mathbf{W})P(\mathbf{W}, \mathbf{POS}, \mathbf{PM}) \cdot P(\mathbf{B}|\Lambda_l, \mathbf{SR})P(\mathbf{P}|\mathbf{B}, \mathbf{SR})P(\mathbf{X}|\Upsilon_s, \Lambda_p, \Lambda_l)P(\mathbf{Y}, \mathbf{Z}|\Upsilon_s, \Lambda_p, \Lambda_l)\} \tag{3.2}$$

(3.2)式中 $P(\mathbf{X}_a, \Upsilon_s|\mathbf{W})$ 代表聲學模型**(AM)**；$P(\mathbf{W}, \mathbf{POS}, \mathbf{PM})$ 則是 joint syntax model，它描述了 Word、POS 及 PM 彼此之間的關係；$P(\mathbf{B}|\Lambda_l)$ 是代表 break syntax model，它是利用語言參數 $\mathrm{L}=\{\mathbf{W}, \mathbf{POS}, \mathbf{PM}\}$ 來預估隱含著階層結構資訊的韻律邊界停頓 **B** 的模式，$P(\mathbf{P}|\mathbf{B})$ 稱為韻律狀態模型，用來描述韻律狀態 P 是如何受到韻律邊界停頓 **B** 影響下發生轉移變化；$P(\mathbf{X}|\Upsilon_s, \Lambda_p, \Lambda_l)$ 稱為音節韻律模型，用來說明音節韻律參數受到 B、P 和 L 的影響而產生的變化；$P(\mathbf{Y}, \mathbf{Z}|\Upsilon_s, \Lambda_p, \Lambda_l)$ 稱為停頓聲學模型，用來說明在各個不同的韻律邊界停頓和語言參數之下，音節內的聲學特性。以下我們將針對這四種韻律模型做更深入的探討。

3.2.1 Break Syntax Model

在相同語言參數之下，不同語速所產生的韻律邊界停頓 **B** 應會有所不同，故必須修正 Break Syntax Model，使其與語速相關，其數學式改寫如(3.3)式所示：

$$P(\mathbf{B}|\Lambda_l) \rightarrow P(\mathbf{B}|\Lambda_l, \mathbf{SR}) = \prod_{n=1}^{N-1} P(B_n|\mathrm{L}_n, SR_n) \tag{3.3}$$

(3.3)式中 $P(B_n|\mathrm{L}_n, SR_n)$ 是一個用來描述音節韻律邊界停頓與其相關的語言參數及語速資訊之間關係的模型，其模型由兩個步驟建構而成，第一步：依據語言參數搭配問題集使用分類樹與決策樹(CART)演算法訓練出一顆決策樹；第二步：對建構出來的決策樹之每一個終止節點(leaf

node)其帶有的八種停頓標記，藉由線性迴歸的方式來模擬各種停頓標記的出現頻率與語速兩者之間的相關性，其數學式如(3.4)式所示：

$$P(B_n = k \mid \mathrm{L}_n, SR_n) = \frac{P(B_n = k \mid \mathrm{L}_n, SR_n)}{\sum_{b=1}^{\text{Break Type\#}} P(B_n = b \mid \mathrm{L}_n, SR_n)} \approx \frac{C_{k,j} SR_n + D_{k,j}}{\sum_{b=1}^{\text{Break Type\#}} C_{b,j} SR_n + D_{b,j}} \tag{3.4}$$

3.2.2 韻律狀態模型

假設目前的韻律狀態只和前一個韻律狀態及前一個停頓標記有關，並以語速分 bin 來區分語速不同時造成韻律狀態轉移機率不同的情況，故韻律狀態模型 $P(\mathbf{P} \mid \mathbf{B})$ 可以改寫並分解成三個子模型，其數學式改寫如(3.5)式所示：

$$\begin{aligned} & P(\mathbf{P} \mid \mathbf{B}) \rightarrow P(\mathbf{PS} \mid \mathbf{B}, \mathbf{SR}) \\ & = P(\mathbf{p} \mid \mathbf{B}, \mathbf{SR}) P(\mathbf{q} \mid \mathbf{B}, \mathbf{SR}) P(\mathbf{r} \mid \mathbf{B}, \mathbf{SR}) \approx P(p_1 \mid bin(SR_1)) P(q_1 \mid bin(SR_1)) P(r_1 \mid bin(SR_1)) \\ & \cdot \left[\prod_{n=2}^{N} P(p_n \mid p_{n-1}, B_{n-1}, bin(SR_n)) P(q_n \mid q_{n-1}, B_{n-1}, bin(SR_n)) P(r_n \mid r_{n-1}, B_{n-1}, bin(SR_n)) \right] \end{aligned} \tag{3.5}$$

(3.5)式中 $P(p_n \mid p_{n-1}, B_{n-1}, \mathrm{bin}(SR_n))$、$P(q_n \mid q_{n-1}, B_{n-1}, bin(SR_n))$ 與 $P(r_n \mid r_{n-1}, B_{n-1}, bin(SR_n))$ 分別表示各個不同韻律狀態，在給定音節邊界停頓 B_{n-1} 及語速 SR_n 的情況下，從第 n-1 個音節的韻律狀態到第 n 個音節韻律狀態的轉移機率。

3.2.3 音節韻律模型

音節韻律模型 $P(\mathbf{X} \mid \Upsilon_s, \Lambda_p, \Lambda_l)$ 可以進一步分解成三個子模型，分別模擬音節音高軌跡序列(**sp**)、音長序列(**sd**)以及音節能量序列(**se**)，如(3.6)式所示：

$$\begin{aligned} P(\mathbf{X} \mid \Upsilon_s, \Lambda_p, \Lambda_l) & \approx P(\mathbf{sp} \mid \Upsilon_s, \mathbf{B}, \mathbf{p}, \mathbf{t}) P(\mathbf{sd} \mid \Upsilon_s, \mathbf{B}, \mathbf{q}, \mathbf{t}, \mathbf{s}) P(\mathbf{se} \mid \Upsilon_s, \mathbf{B}, \mathbf{r}, \mathbf{t}, \mathbf{f}) \\ & \approx \prod_{n=1}^{N} P(\mathbf{sp}_n \mid p_n, B_{n-1}^{n}, t_{n-1}^{n+1}) P(sd_n \mid q_n, s_n, t_n) P(se_n \mid r_n, f_n, t_n) \end{aligned} \tag{3.6}$$

在第一個子模型 $P(\mathbf{sp}_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1})$ 中，我們假設 $\mathbf{sp}_n$ 會受到下列因素影響：分別是目前的音高韻

律狀態 p_n、目前的聲調t_n以及在給定韻律邊界停頓 B_{n-1}和 B_n 時，前後各一個音節聲調t_{n-1}和t_{n+1}造成的連音影響，此處表示 $B_{n-1}^{n}=(B_{n-1},B_n)$，$t_{n-1}^{n+1}=(t_{n-1},t_n,t_{n+1})$。而 $\mathbf{sp}_n$ 則為第 n 個音節音高軌跡，是將音節音高軌跡進行正交展開(orthogonal expansion)，投影到四個 Legendre 多項式基底所得到的四維正交參數[10]，其表示法如下所示：

$$\mathbf{sp}_n=\mathbf{sp}_n^r+\boldsymbol{\beta}_{t_n}+\boldsymbol{\beta}_{p_n}+\boldsymbol{\beta}^f_{B_{n-1},t^n_{n-1}}+\boldsymbol{\beta}^b_{B_n,t_n^{n+1}}+\boldsymbol{\mu}_{sp} \tag{3.7}$$

在(3.7)式中，$\boldsymbol{\beta}_{t_n}$及$\boldsymbol{\beta}_{p_n}$分別是目前音節音調t_n及目前音節韻律狀態 p_n 的影響因子(Affecting Patterns, APs)；$\boldsymbol{\beta}^f_{B_{n-1},t^n_{n-1}}$及$\boldsymbol{\beta}^b_{B_n,t_n^{n+1}}$分別是第 n-1 個和第 n+1 個音節所貢獻的前後音節影響效應的APs；$\boldsymbol{\mu}_{sp}$是音高向量的總體平均值(global mean)；$\mathbf{sp}_n^r$是正規化後的 $\mathbf{sp}_n$，即為 $\mathbf{sp}_n$ 扣除$\boldsymbol{\beta}_{t_n}$、$\boldsymbol{\beta}_{p_n}$、$\boldsymbol{\beta}^f_{B_{n-1},t^n_{n-1}}$、$\boldsymbol{\beta}^b_{B_n,t_n^{n+1}}$和$\boldsymbol{\mu}_{sp}$的殘餘值(residual)。藉由假設$\mathbf{sp}_n^r$是一平均值為零的高斯隨機分布(normal distribution)，即 $N(\mathbf{sp}_n^r;0,R_{sp})$，則 $P(\mathbf{sp}_n|B_{n-1}^n,p_n,t_{n-1}^{n+1})$ 可化解成(3.8)式：

$$P(\mathbf{sp}_n\mid p_n,B_{n-1}^n,t_{n-1}^{n+1})=N(\mathbf{sp}_n;\beta_{t_n}+\beta_{p_n}+\beta^f_{B_{n-1},t^n_{n-1}}+\beta^b_{B_n,t_n^{n+1}}+\mu_{sp},R_{sp}) \tag{3.8}$$

在建構第二個子模型 $P(sd_n\mid q_n,s_n,t_n)$ 時，我們假設 sd_n 會受到下列因素影響：分別是音節韻律狀態、基本音節類型及音節聲調，因此我們可以將觀察到的音節長度 sd_n 表示成：

$$sd_n=sd_n^r+\gamma_{t_n}+\gamma_{s_n}+\gamma_{q_n}+\mu_{sd} \tag{3.9}$$

在(3.9)式中 sd_n^r 是正規化後的 sd_n；γ_{t_n}、γ_{s_n}及γ_{q_n}分別是目前音節音調、基本音節類型及目前音節韻律狀態影響效應的 APs；μ_{sd}是音節音長的總體平均值(global mean)。藉由假設sd_n^r是一平均值為零的高斯隨機分布(normal distribution)，即 $N(sd_n^r;0,R_{sd})$，則 $P(sd_n\mid q_n,s_n,t_n)$ 可化解成(3.10)式：

$$P(sd_n\mid q_n,s_n,t_n)=N(sd_n;\gamma_{t_n}+\gamma_{s_n}+\gamma_{q_n}+\mu_{sd},R_{sd}) \tag{3.10}$$

最後在建構第三個子模型 $P(se_n\mid r_n,f_n,t_n)$ 時，我們假設 se_n 會受到下列因素影響：分別是音

節韻律狀態、音節韻母類型及音節聲調，因此我們可以將觀察到的音節能量 se_n 表示成：

$$se_n = se_n^r + \omega_{t_n} + \omega_{f_n} + \omega_{r_n} + \mu_{se} \tag{3.11}$$

在(3.11)式中 se_n^r 是正規化後的 se_n；ω_{t_n}、ω_{s_n} 及 ω_{q_n} 分別是目前音節音調、目前音節韻母類型及目前音節韻律狀態影響效應的 APs；μ_{se} 是音節能量的總體平均值(global mean)。藉由假設 se_n^r 是一平均值為零的高斯隨機分布(normal distribution)，即 $N(se_n^r;0,R_{se})$，則 $P(se_n \mid r_n, f_n, t_n)$ 可化解成(3.12)式：

$$P(se_n \mid r_n, f_n, t_n) = N(se_n; \omega_{t_n} + \omega_{f_n} + \omega_{r_n} + \mu_{se}, R_{se}) \tag{3.12}$$

3.2.4 停頓聲學模型

在此使用了音節邊界參數及差分韻律參數**{Y, Z}={pd, ed, pj, dl, df}**來描述韻律邊界的聲學特性，並且假設五種韻律聲學參數間彼此互相獨立，則停頓聲學模型可以做進一步的分解，如下所示：

$$\begin{aligned} & P(\mathbf{Y},\mathbf{Z}|\Upsilon_s,\Lambda_p,\Lambda_l) \\ & \approx P(\mathbf{pd},\mathbf{ed},\mathbf{pj},\mathbf{dl},\mathbf{df}|\Upsilon_s,\Lambda_p,\Lambda_l) \\ & \approx \prod_{n=1}^{N-1}\Big\{ g(pd_n;\alpha_{B_n,\Lambda_{l,n}},\beta_{B_n,\Lambda_{l,n}}) N(ed_n;\mu_{ed,B_n,\Lambda_{l,n}},\sigma^2_{ed,B_n,\Lambda_{l,n}}) \\ & \quad \cdot N(pj_n;\mu_{pj,B_n,\Lambda_{l,n}},\sigma^2_{pj,B_n,\Lambda_{l,n}}) N(dl_n;\mu_{dl,B_n,\Lambda_{l,n}},\sigma^2_{dl,B_n,\Lambda_{l,n}}) \\ & \quad \cdot N(df_n;\mu_{df,B_n,\Lambda_{l,n}},\sigma^2_{df,B_n,\Lambda_{l,n}})\Big\} \end{aligned} \tag{3.13}$$

在(3.13)式中 pd_n 為音節邊界的停頓時長(pause duration)在這裡以伽瑪隨機分布(gamma distribution)來模擬之；ed_n 為音節間的能量低點(energy dip)，在這裡以高斯隨機分布(normal distribution)來模擬之；pj_n 為正規化的音節內基頻差序列(pitch jump)，定義如下：

$$pj_n = (\mathbf{sp}_{n+1}(1) - \boldsymbol{\beta}_{t_{n+1}}(1)) - (\mathbf{sp}_n(1) - \boldsymbol{\beta}_{t_n}(1)) \tag{3.14}$$

在(3.14)式中，$\mathbf{sp}_n(1)$定義為第一維度的音節音高軌跡；$\boldsymbol{\beta}_{t_n}(1)$則定義為第一維度的聲調影響因子。同樣的，在這裡以高斯隨機分布(normal distribution)來模擬之。最後，還有兩種正規化的音節長度拉長因子 ***dl*** 和 ***df*** 定義為：

$$dl_n = (sd_n - \gamma_{t_n} - \gamma_{s_n}) - (sd_{n-1} - \gamma_{t_{n-1}} - \gamma_{s_{n-1}}) \tag{3.15}$$

$$df_n = (sd_n - \gamma_{t_n} - \gamma_{s_n}) - (sd_{n+1} - \gamma_{t_{n+1}} - \gamma_{s_{n+1}}) \tag{3.16}$$

這裡兩種因子我們都使用高斯隨機分布(normal distribution)來模擬。在實作過程中，$P(\mathbf{pd},\mathbf{ed},\mathbf{pj},\mathbf{dl},\mathbf{df}|\Upsilon_s,\Lambda_p,\Lambda_l)$是經由語言參數搭配問題集(附錄一)使用分類數與決策樹(CART)推導出來，其節點的分類標準是依據最大概似函數增益(maximum likelihood gain)，依據不同的韻律邊界停頓同時將所有音節邊界的 pd_n、ed_n、pj_n、dl_n、df_n 做好分類，並於決策樹的每個終止節點(leaf node)統計其參數分佈。

3.3 加入韻律訊息於 two-stage 語音辨認系統

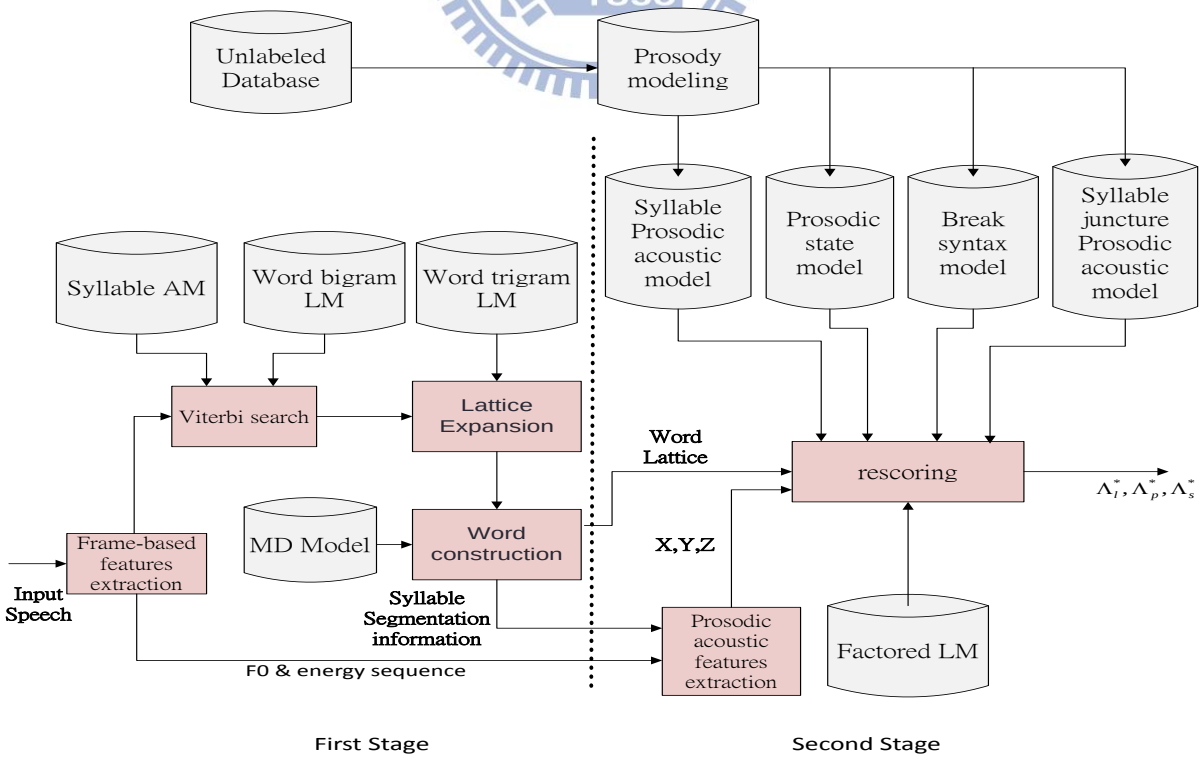


圖 3.3：以 two-stage 方式之韻律輔助中文語音辨認系統流程圖

上圖 3.3 為本研究之系統流程圖，以下將對此系統的 second stage 做詳細介紹：

3.3.1 Joint Syntax Model 之架構與建立

由 3.2 小節可以發現到，韻律模型的建立需要給定語言參數資訊，其中包含詞性(POS)以及標點符號(PM)這兩種語言參數資訊，本研究所使用的 joint syntax model 包含一個 trigram LM、一個 factored POS model 與一個 factored PM model，在這裡我們使用 FLM approach[11]建構 factored POS model 和 factored PM model 其主要概念是利用其他相關資訊(factor)的輔助來預估目標，所以這裡將充分利用到所有相關語言知識來提升預估 POS 或 PM 的準確性。

當使用多種語言資訊做預估時難免會面臨到資料量不足的問題，因此會採取退化(back 0ff)的架構，在本研究中，factored POS model 其退化路徑如下圖 3.4 所示。在最上層的部分，使用目前的詞W_i、前一個POS_{i-1}及前前一個POS_{i-2}的語言資訊來預估目前的POS_i，若此機率的組合沒有出現，則先丟棄一個 factor POS_{i-2}，退化到只由目前的詞W_i、前一個POS_{i-1}的語言資訊來預估目前的POS_i，若仍是沒有出現的話則再丟棄一個 factor POS_{i-1}，若仍是沒有出現的話，就退化到最底層的狀態，此時就一定有此機率。

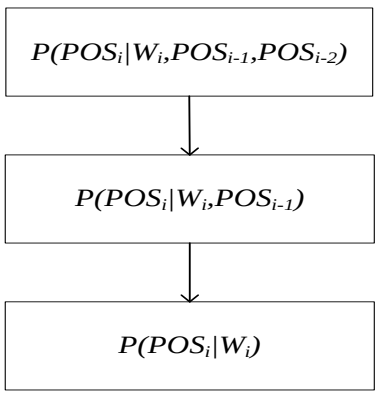


圖 3.4：factored POS model 的 backoff 路徑

依此類推，factored PM model 亦是如此，其退化路徑如下圖 3.5 所示。在最上層的部分，使用前一個詞W_i、前一個POS_{i-1}及目前POS_i的語言資訊，來預估前一個 PM 的機率，若此機

率的組合沒有出現，依序丟掉 POS_i，接著是 POS_{i-1}，然後 W_{i-1}，最終退化到 $P(PM_{i-1})$，此時就一定有此機率。

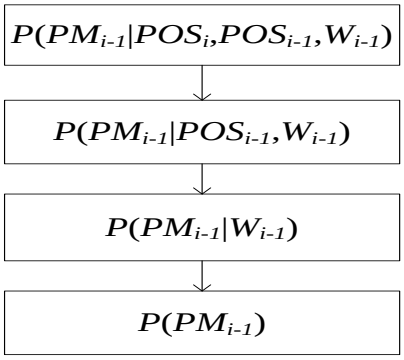


圖 3.5：factored PM model 的 backoff 路徑

其完整訓練流程圖如下圖 3.6 所示，需先對訓練文本做前處理，使其文本內所有的詞都帶有詞性(POS)與標點符號(PM)的標記如下表 3.3 所示，接著使用 SRILM toolkit[12]及利用 Witten-Bell smoothing 的方式來訓練模型，其中 flm 檔案即為設定 factored model 內每一層 backoff 結構中所需要考慮的 factor。

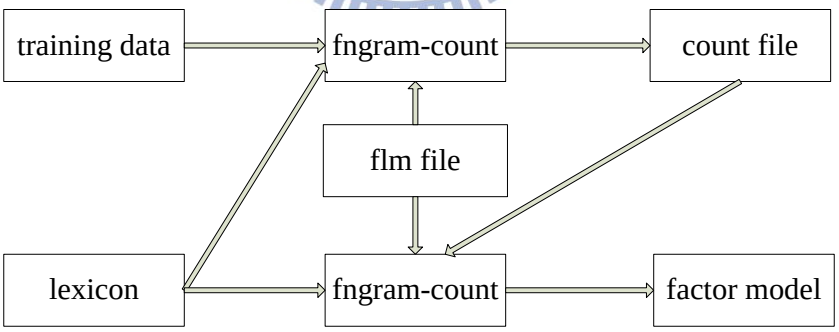


圖 3.6：factored model 訓練架構流程圖

表 3.3：文本處理

Original_data：
目前為改善交通秩序，立體停車場已陸續興建，原狹窄街道亦逐步擴建。

FLM_data：
W-目前:P-Nd:M-NONE W-為:P-P:M-NONE W-改善:P-VC:M-NONE W-交通:P-Na:M-NONE W-秩序:P-Na:M-COM W-立體:P-VH:M-NONE W-停車廠:P-Nc:M-NONE W-已:P-D:M-NONE W-陸續:P-D:M-NONE W-興建:P-VC:M-COM W-原:P-D:M-NONE W-狹窄:P-VH:M-NONE W-街道:P-Na:M-NONE W-亦:P-D:M-NONE W-逐步:P-D:M-NONE W-擴建:P-VC:M-OTH

3.3.2 特徵參數正規化

針對圖 3.3 中第二級辨認器開始，工作過程中如有牽涉到音節能量、音節長度或音節音高等韻律聲學參數時，必須要先經過正規化的步驟，藉此消除語速對於韻律聲學參數的影響，以下我們將針對這四種韻律聲學參數介紹其語速正規化的方式。

3.3.2.1 音節長度之語速正規化

中文中音節長度近似於高斯隨機分佈(normal distribution)，故針對音節長度做高斯正規化：對於音節長度的平均值使用一階多項式曲線來模擬，而對於音節長度的標準差則使用二階多項式曲線來模擬，其正規化數學式如(3.17)式所示：

$$sd'_n = \left(sd_n - \mu^{sd}(SR)\right) \Big/ \tilde{\sigma}^{sd}(SR(k)) \times \sigma_g^{sd} + \mu_g^{sd} \tag{3.17}$$

其中 sd_n 表示原始音節長度；sd'_n 表示語速正規化後的音節長度；$\mu^{sd}(SR) = SR$ 為此語句的音節長度平均值，我們將其當作此語句的語速 SR ；

$$\tilde{\sigma}^{sd}(SR) = a_1 \left(SR\right)^2 + b_1 \cdot SR + c_1 \tag{3.18}$$

為平滑化後的標準差；μ_g^{sd} 和 σ_g^{sd} 分別為訓練韻律模型時所統計出的結果，代表所有語料庫音節長度之總體平均值與平均標準差。

3.3.2.2 停頓長度之語速正規化

停頓長度較近似於伽瑪隨機分布(gamma distribution)，對於停頓長度的平均值與標準差皆使用二階多項式曲線來模擬，其正規化數學式如(3.19)式所示：

$$pd' = G^{-1}(\, G(pd, \tilde{\alpha}^{pd}(SR(k)), \tilde{\beta}^{pd}(SR(k))), \alpha_g^{pd}, \beta_g^{pd}) \tag{3.19}$$

其中 $G(pd, \alpha, \beta)$ 為伽碼分佈的累積密度函數(Cumulative Density Function, CDF)；

$$\tilde{\alpha}^{pd}(SR(k)) = (\tilde{\mu}^{pd}(SR(k)))^2 / (\tilde{\sigma}^{pd}(SR(k)))^2 \tag{3.20}$$

及

$$\tilde{\beta}^{pd}(SR(k)) = (\tilde{\sigma}^{pd}(SR(k)))^2 / \tilde{\mu}^{pd}(SR(k)) \tag{3.21}$$

為平滑化後的停頓長度伽碼分佈的參數；

$$\tilde{\mu}^{pd}(SR) = a_2 (SR)^2 + b_2 \cdot SR + c_2 \tag{3.22}$$

及

$$\tilde{\sigma}^{pd}(SR) = a_3 (SR)^2 + b_3 \cdot SR + c_3 \tag{3.23}$$

為語速控制的停頓長度平滑平均值和標準差；α_g^{pd} 和 β_g^{pd} 分別為訓練韻律模型時所統計出的停頓長度伽碼分佈的參數平均值。

3.3.2.3 音節音高軌跡之語速正規化

首先針對不同語者做 frame-based normalization 以消除不同語者先天發音上的差異，其正規化數學式如(3.24)式所示：

$$fsp_n'(k) = \frac{fsp_n(k) - \mu^{fsp}(k)}{\sigma^{fsp}(k)} \times \sigma_g^{fsp} + \mu_g^{fsp} \tag{3.24}$$

其中 k 為 speaker index，fsp_n 表示第 n 個音框的原始基頻對數值；fsp_n' 表示第 n 個音框做完 frame-based normalization 的基頻對數值；μ_g^{fsp} 和 σ_g^{fsp} 分別為訓練韻律模型時所統計出的結果，

代表所有語料庫的基頻對數值之總體平均值與平均標準差，μ^{fsp} 和 σ^{fsp} 分別為第 k 個語者基頻對數值的平均值和標準差。

接著再針對音節基頻軌跡進行正交展開(orthogonal expansion)，投影到四個 Legendre 多項式基底，用所得到的四維正交參數 $\mathbf{sp}_n = \left[a_n^0 \; a_n^1 \; a_n^2 \; a_n^3 \right]^T$ 表示其基頻軌跡，第一維正交參數代表此基頻軌跡的平均值，後三維正交參數則用來描述此基頻軌跡分佈。由於基頻軌跡分佈相對於聲調不同而有相異性，故我們在此將基頻軌跡依據不同聲調做正規化動作，其正規化數學式如(3.25)式所示：

$$sp_n'(i) = \frac{sp_n(i) - \tilde{\mu}^{sp}(SR(k), t_n, i)}{\tilde{\sigma}^{sp}(SR(k), t_n, i)} \times \sigma_g^{sp}(t_n, i) + \mu_g^{sp}(t_n, i) \tag{3.25}$$

其中 $t = 1 \sim 5$ 表示聲調類型； $i = 1 \sim 4$ 表示維度； $\mathbf{sp}_n'$ 表示語速正規化後的第 i 維基頻軌跡參數；

$$\tilde{\mu}^{sp}(SR(k), t, i) = b_4(t, i) \cdot SR + c_4(t, i) \tag{3.26}$$

和

$$\tilde{\sigma}^{sp}(SR(k), t, i) = b_5(t, i) \cdot SR + c_5(t, i) \tag{3.27}$$

為語速控制的聲調及維度相依的基頻軌跡參數的平滑平均值和標準差；μ_g^{sp} 和 σ_g^{sp} 分別為訓練韻律模型時所統計出的結果，代表所有語料庫中四維正交參數之總體平均值與平均標準差。

3.3.2.4 音節能量強度之語速正規化

音節能量強度與背景環境及錄音條件具有強大的相關性，而受語速的影響並不大。因此，音節能量強度採用 Utterance-based normalization scheme，其正規化數學式如(3.28)式所示：

$$se_n'(k) = \frac{se_n(k) - \mu^{se}(k)}{\sigma^{se}(k)} \times \sigma_g^{se} + \mu_g^{se} \tag{3.28}$$

其中 k 為 utterance index；se_n' 表示 normalized 後的音節能量強度數值；μ_g^{se} 和 σ_g^{se} 分別為訓練韻

律模型時所統計出的結果，代表所有語料庫的音節能量之總體平均值與平均標準差，μ^{se} 和 σ^{se} 分別為第 k 個音檔音節能量強度的平均值和標準差。

3.3.3 The Second Stage 之實作

在第二個 stage 開始之後，加入的韻律模型及 joint syntax model 種類高達 16 種之多，為了瞭解每個模型對於辨識系統的影響力，本研究將針對 the second stage 再細分成三個小階段，逐次加入模型資訊並觀察實驗結果，其詳細流程圖如下所示：

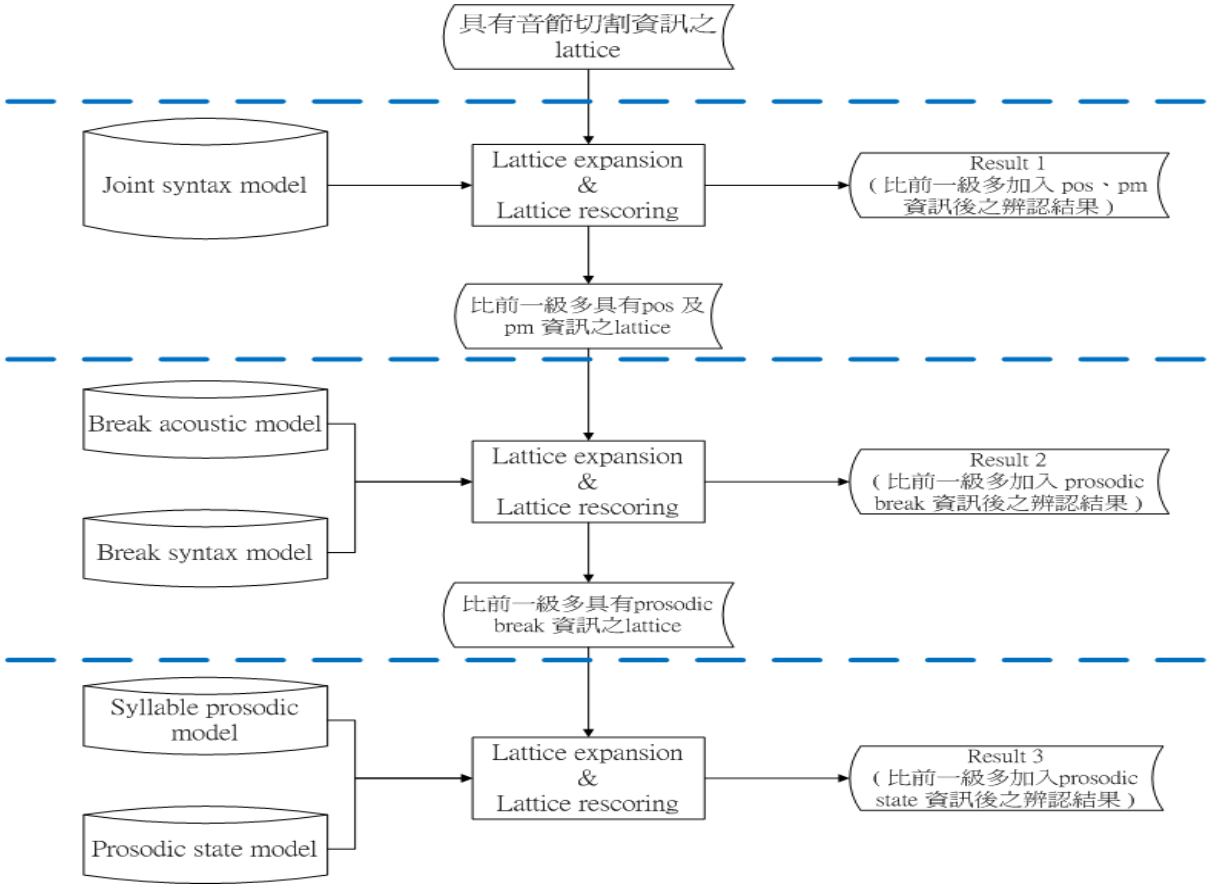


圖 3.7：辨認器第二級三階段實作流程圖

3.3.3.1 第一階段：加入多種語言資訊

如圖 3.7 所示，辨認器第二級第一階段是引入多種語言參數資訊(POS 及 PM)，在這裡需要加入的模型為 joint syntax model，此時在 word lattice 上每個 node 所帶有的 word 資訊會根據

factored POS model 與 factored PM model 找出相對應的詞性與標點符號做展開動作。

在此階段辨識結果將會解碼出多種語言參數，包含詞(word)、詞性(POS)及標點符號(PM)。

(1).node expansion :

在 word lattice 上每個 node 所帶有的 word 資訊會根據 factored POS model 找出其最佳對應的詞性(POS)，搭配 4 種標點符號(PM)；無標點符號(NONE)、頓號(DOT)、逗號(COM)、句點或驚嘆號等具有句子結束意義的標點符號(OTH)，將各個 node 展開至 1*4 倍。

(2).arc expansion :

針對原始 lattice 中各個 arc 所帶有的 word 資訊，找出相對應的 POS 數目(1)及 PM 數目(4)，並對上一個 arc 中所帶有的 word 資訊，找出相對應的 POS 數目(1)及 PM 數目(4)，再將各個 arc 展開至 1*4*1*4 倍。

3.3.3.2 第二階段：加入韻律邊界停頓資訊

第二個階段我們主要是引入韻律邊界停頓的資訊，要加入的韻律模型分別是 break syntax model 及停頓聲學模型，觀察數學式(3.14)、(3.15)及(3.16)，由於韻律基本單元為音節，針對 intraword syllable 的部分，可由程式內部處理，不需要將 lattice 作展開的動作，而 interword syllable 部份，針對 word lattice 上各個 node，觀察其 arc 中所帶有的 word 資訊中最後一個音節長度做展開，至於 word 中第一個音節長度資訊部分，因為此階段做重新評分時，其方式採用 backward viterbi search 的緣故，利用資訊傳遞的特性，已經存入下一個 word 中第一個音節長度資訊，故不需對此做展開的動作。

在此階段最後除了會解碼出多種語言參數外，同時也會將每個詞中各個音節後所接的韻律邊界停頓一併解碼出。

(1).node expansion :

針對原始 lattice 中各個 node，觀察其來源 arc 中所帶有的 word 資訊中最後一個音節長度，現假設來源 arc 中最後一個音節的長度共有 M 種，則原始 node 將展開至 M 倍。

(2).arc expansion :

針對原始 lattice 中各個 arc，觀察其 start node 的特性，假設從第(1)步中，已知 start node 將被展至 M 倍，則原始 arc 也將展開至 M 倍。

3.3.3.3 第三階段：加入音節韻律狀態資訊

最後一個階段是引入音節韻律狀態資訊，要加入的韻律模型分別是音節韻律模型及韻律狀態模型，觀察數學式(3.5)及(3.6)，現在每一個音節的 prosodic state score 會相依於上一個音節的韻律狀態，不同於第二階段中計算 intraword syllable 的 prosodic break score 不用考慮到上一個音節的韻律邊界停頓類型，在此如同採用 viterbi search 的方式，在程式內部中，我們將每一個 intraword syllable 的部分獨立處理，針對音節中的每一種 prosodic state 都找到最佳的分數，再從這個音節中 16 個 prosodic state 裡選出一個累積分數最多的 state，最後將此音節所累積的最佳資訊傳遞到下一個 node 中。

針對 interword syllable 的部分，發現到當計算某一音節的 prosodic score，必須得到前後各一個音節資訊，在此只根據後一個音節資訊作展開，也就是 lattice 將根據每個 word 中第一個音節資訊作展開，而參照以上數學式(3.5)及(3.6)，word 中第一個音節資訊正好就是音節所對應的聲調資訊，至於 word 中前一個音節資訊部分，因為此階段做重新評分時，其方式採用 forward viterbi search 的緣故，利用資訊傳遞的特性，已經存入上一個 word 中最後一個音節資訊，故不需對此做展開的動作。

在這一階段最後除了會解碼出如同上一階段的語言參數及韻律邊界停頓外，詞中各個音節所屬的三種韻律狀態：音節音高韻律狀態、音節音長韻律狀態及音節能量韻律狀態也將一併解碼出。

(1).node expansion :

針對原始 lattice 中各個 node，觀察其分出的 arc 中所帶有的 word 資訊中第一個音節的聲調，現假設分出的 arc 中第一個音節的聲調共有 M 種，則原始 node 將展開至 M 倍。

(2).arc expansion :

針對原始 lattice 中各個 arc，觀察其 end node 的特性，假設從第(1)步中，已知 end node 將被展至 M 倍，則原始 arc 也將展開至 M 倍。

3.4 鑑別式模型組合

在本研究裡，第二個 stage 中拿來作重新評分的模型高達 16 種之多，因此如何找出一組權重使這 16 個模型作結合後能夠得到最小的詞錯誤率便是非常重要之課題，如果使用 trail-and-error 的方式來決定 16 個模型的權重將非常耗費時間且缺乏效率，所以本研究使用鑑別式模型組合(Discriminative Model Combination, DMC) [13]的方法來決定權重。

DMC 的方法是先定義一個 decision error rate 的鑑別式函數(discriminant function)如(3.29)式，目標是找到一組權重使此函數的 decision error rate 最佳化。

$$\begin{aligned} g(x_1^T, w_1^S, w'^{S'}_1) &= \log P(w_1^S \mid x_1^T) - \log P(w'^{S'}_1 \mid x_1^T) \\ &= \log[P(w_1^S)P(x_1^T \mid w_1^S)] - \log[P(w'^{S'}_1)P(x_1^T \mid w'^{S'}_1)] \end{aligned} \tag{3.29}$$

(3.29)式中 $w_1^S = (w_1, ...w_S)$ 代表詞串；$x_1^S = (x_1, ...x_T)$ 代表特徵參數向量；$P(w_1^S \mid x_1^T)$ 代表在給定特徵參數條件下得到**正確詞串**的分數；而 $P(w'^{S'}_1 \mid x_1^T)$ 則代表在給定同樣特徵參數條件下得到**辨認結果詞串**的分數。當 $P(w'^{S'}_1 \mid x_1^T)$ 分數愈接近 $P(w_1^S \mid x_1^T)$ 愈好，代表其 likelihood 為最高，但分數最接近者不代表詞錯誤率(WER)會是最小。現在假如 $P(w_1^S \mid x_1^T)$ 將拆解成 M 個不同模型，其線性對數(log-linearly)組合如下：

$$P^{\Pi}_{\{\Lambda\}}(x_1^T \mid w_1^S) = \exp\{\log C(\Lambda) + \sum_{j=1}^{M} \lambda_j \log P_j(x_1^T \mid w_1^S)\} \tag{3.30}$$

(3.30)式中 $\Lambda = (\lambda_1, ...\lambda_M)^{tr}$ 代表各種模型 P_j 分數組合時的權重；$C(\Lambda)$ 代表正規化因子(normalization factor)。依據 decision error rate，我們要從 discriminant function 找出一組最佳的權重 $\boldsymbol{\Lambda}$，而且 $P(w_1^S \mid x_1^T)$ 可拆成 M 個不同模型，故(3.29)式將改寫成(3.31)式：

$$g(x_1^T, w_1^S, w'^{S'}_1) = \sum_{j=1}^{M} \lambda_j (\log P_j(w_1^S \mid x_1^T) - \log P_j(w'^{S'}_1 \mid x_1^T)) \tag{3.31}$$

最後將定義一個 smooth misclassification function $\ell(x_n, k_{n0}, \Lambda)$，搭配 Generalized Probabilistic Descent (GDP) algorithm[13]求出各種模型的權重值 $\boldsymbol{\Lambda}$，首先針對符號作些定義：

定義一：詞串 w_1^S 表示為 class k；而每個句子 x_1^T 表示為特徵參數向量 x。

定義二：訓練資料表示為 $(x_n, k_{nr}), n=1,...,N, r=0,...,K,$ 其中 N 代表句子數目；k_{n0} 代表特徵參數向量 x_n 的標準答案；$k_{nr}, r=1,...,K$ 互為彼此的競爭者，意即 K-best 序列。

定義三：$LD(k_{nr}, k_{n0})$ 代表 Levenshtein-distance，意即 hypothesis k_{nr} 的錯誤數量，錯誤包含插入性、刪除性、取代性等。

定義四：訓練語料(或 held-out data)的 smoothed empirical error rate $L(\Lambda)$ 為：

$$L(\Lambda) = \frac{1}{N}\sum_{n=1}^{N} \ell(x_n, k_{n0}, \Lambda) \tag{3.32}$$

其中

$$\ell(x_n, k_{n0}, \Lambda)^{-1} = 1 + A \cdot \left(\frac{1}{K}\sum_{r=1}^{K} e^{\left\{ -\eta LD(k_{nr}, k_{n0}) \log \frac{p^{\Pi}_{\{\Lambda\}}(k_{n0}|x_n)}{p^{\Pi}_{\{\Lambda\}}(k_{nr}|x_n)} \right\}} \right)^{-\frac{B}{\eta}} \tag{3.33}$$

$A > 0,\ B > 0,$ and $\eta > 0$.

最後利用下面的遞迴架構即可求出權重值 λ_j：

For j=1,...,M

$$\lambda_j^{(0)} = 1$$

$$\lambda_j^{(I+1)} = \lambda_j^{(I)} + \varepsilon \sum_{n=1}^{N} \ell(x_n, k_{n0}, \Lambda^{(I)}) \left(1 - \ell(x_n, k_{n0}, \Lambda^{(I)})\right) \cdot \frac{\sum_{r=1}^{K} LD(k_{nr}, k_{n0}) \log\left(\frac{p_j(k_{n0}|x_n)}{p_j(k_{nr}|x_n)}\right) \left[p^{\Pi}_{\{\Lambda^{(I)}\}}(k_{nr}|x_n) \right]^{\eta LD(k_{nr}, k_{n0})}}{\sum_{r=1}^{K} \left[p^{\Pi}_{\{\Lambda^{(I)}\}}(k_{nr}|x_n) \right]^{\eta LD(k_{nr}, k_{n0})}} \tag{3.34}$$

$$\Lambda^{(I+1)} = (\lambda_1^{(I+1)}, ..., \lambda_M^{(I+1)})^{tr} \tag{3.35}$$

在(3.34)式中 ε 代表 step size，可以發現到 λ_j 在多次遞迴中是決定於鑑別式函數 $\log\left(\frac{p_j(k_{n0}|x_n)}{p_j(k_{nr}|x_n)}\right)$ 的權重和。

第四章 實驗結果與分析

本章將介紹本研究所做的實驗結果並進一步分析結果，4.1 節中將介紹階層式韻律模型的訓練；4.2 節將列出加入韻律資訊於語音辨認系統後重新評分的結果；4.3 節將對此系統的辨識結果做分析討論。

4.1 階層式韻律模型之訓練

本研究在 TCC300 語料庫中的韻律邊界停頓的標記系統如同 3.1 小節所敘述，將韻律邊界停頓資訊分為 *B*0、*B*1-1、*B*1-2、*B*2-1、*B*2-2、*B*2-3、*B*3 及 *B*4，以這八種停頓標記用來表達音節(SYL)、韻律詞(PW)、韻律短語(PPh)、呼吸組/韻律句組(BG/PG)等邊界，並以此來訓練階層式韻律模型，在這裡將對 Break Syntax Model 和停頓聲學模型(break acoustic model)說明訓練方式。

4.1.1 Break Syntax Model

Break syntax model 是用來描述音節韻律邊界停頓與其相關的語言參數及語速資訊之間關係的模型，其模型由兩個步驟建構而成，第一步，依據語言參數搭配問題集使用分類樹與決策樹(CART)演算法訓練，將不同的韻律邊界停頓做分類得到一顆決策樹，如下圖 4.1，決策樹中的每一個終止節點(leaf node)將存入每一類韻律邊界停頓的機率值，也可以藉由觀察中間非終止節點(nonterminal node)所問到的問題來加以分析其問題的重要程度。

以下將介紹語料庫與 CART 演算法的兩個分裂停止條件設定，透過這兩個條件設定適時控制決策樹的深度，避免生長過深：

1. 採用TCC300中164位語者共約106955個音節來訓練本研究的韻律模型
2. 決策樹中各個終止節點(leaf node)其最小樣本數(音節數量)必須大於700。
3. 決策樹的訓練過程中，其相對相似度增益(relative likelihood gain)必須大於0.001。

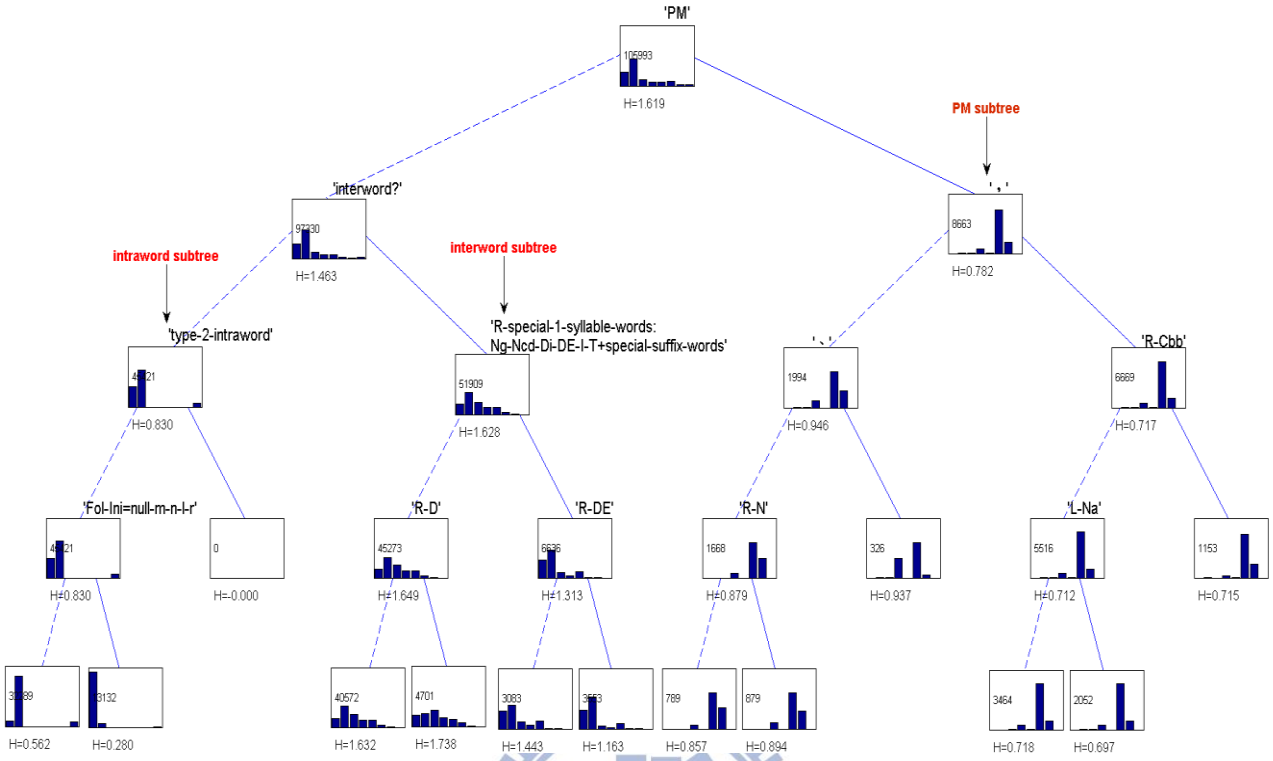


圖 4.1：break syntax model 的決策樹架構

上圖4.1中，每一個節點都帶有其對應的問題，實線表示針對其父節點的問題回答正確之走向，虛線則表示回答錯誤之走向，另外節點內代表了8種韻律邊界停頓類型之分佈圖，從左到右分別為*B*0、*B*1-1、*B*2-1、*B*2-2、*B*2-3、*B*3、*B*4、*B*1-2，節點中的數值為其樣本數(音節數量)，H則代表Shannon entropy，用來評估韻律邊界停頓類型分布之不確定性。

觀察圖 4.1 的決策樹，將此決策樹再細分為三顆子樹：PM subtree、interword subtree 及 intraword subtree，PM subtree 其根節點的韻律邊界停頓之機率分布大多集中在 *B*3、*B*4 等長停頓，由此顯示大多數 *B*3 與 *B*4 常發生在有標點符號的地方；intraword subtree 中則其韻律邊界停頓之機率分布大多集中於 *B*0、*B*1-1、*B*1-2；而 interword subtree 本身結構相較於其他兩顆子樹要複雜得多，也因此為了要使預估更精確，所考慮的語言資訊就愈多，必須向下問更多重要問題，此時像是「右邊或左邊之特殊一字詞」就是一個很重要的問題，與本研究想解決中文語音辨認上的一字詞易混淆及搶詞錯誤有很大的關連性，圖 4.2 就是針對 interword subtree 的更深層結構。

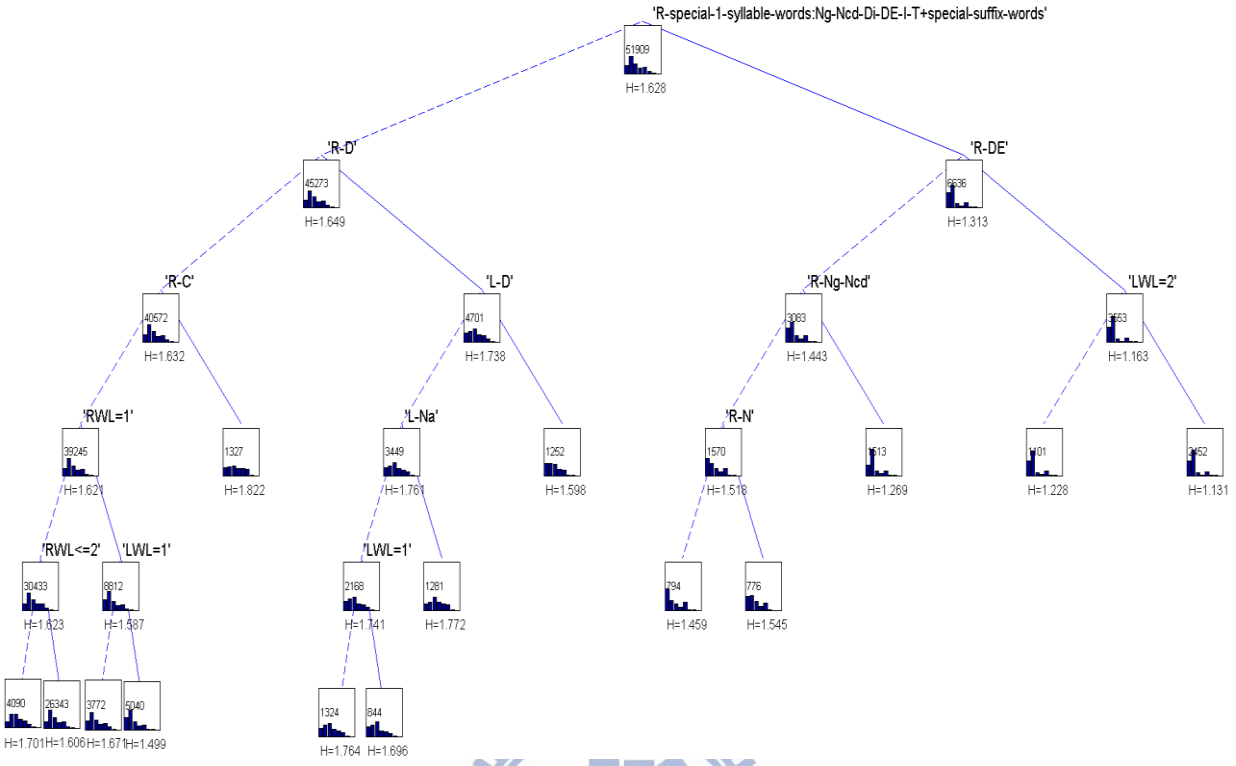


圖 4.2：interword subtree 架構更深層部分

第二步，針對決策樹中每一個終止節點(leaf node)其帶有的八種停頓標記，使用一階多項式曲線來模擬各種停頓標記的出現頻率與語速兩者之間的相關性，藉此得到 $P(\mathbf{B}_n \mid \Lambda_l, \mathbf{SR}_n)$。

4.1.2 停頓聲學模型

停頓聲學模型依據語言參數搭配問題集使用分類樹與決策樹(CART)演算法訓練建構而成，藉此描述八種停頓標記、語言參數、音節邊界參數及差分韻律參數彼此之間的關係。一般而言，在韻律階層結構中，用來區分愈高階層韻律組成份子的韻律邊界停頓通常會具有較長的停頓時長(pause duration)、較低的音節間能量低點(energy-dip)、較大的正規化基頻跳躍值(normalized pitch-level jump)、及較大的音節長度影響因子(duration lengthening factors)。圖 4.3 顯示在不同停頓標記之下，決策樹根節點中各項韻律參數的機率密度函數，可以發現到 *B*0、*B*1-1 的停頓時長非常的短、擁有較小的音節長度拉長因子、較高的能量低點以及基頻差較不明顯；*B*3、*B*4 則擁有較長的停頓時長、較大的音節長度拉長因子、較低的能量低點以及較明顯的基頻差；

*B*1-2 及 *B*2-2 則有中等的停頓時長；*B*2-1 擁有較明顯的基頻跳躍，*B*2-3 則是音節拉長因子較為明顯。這部分的實驗都相當符合上述所說的韻律邊界停頓資訊。

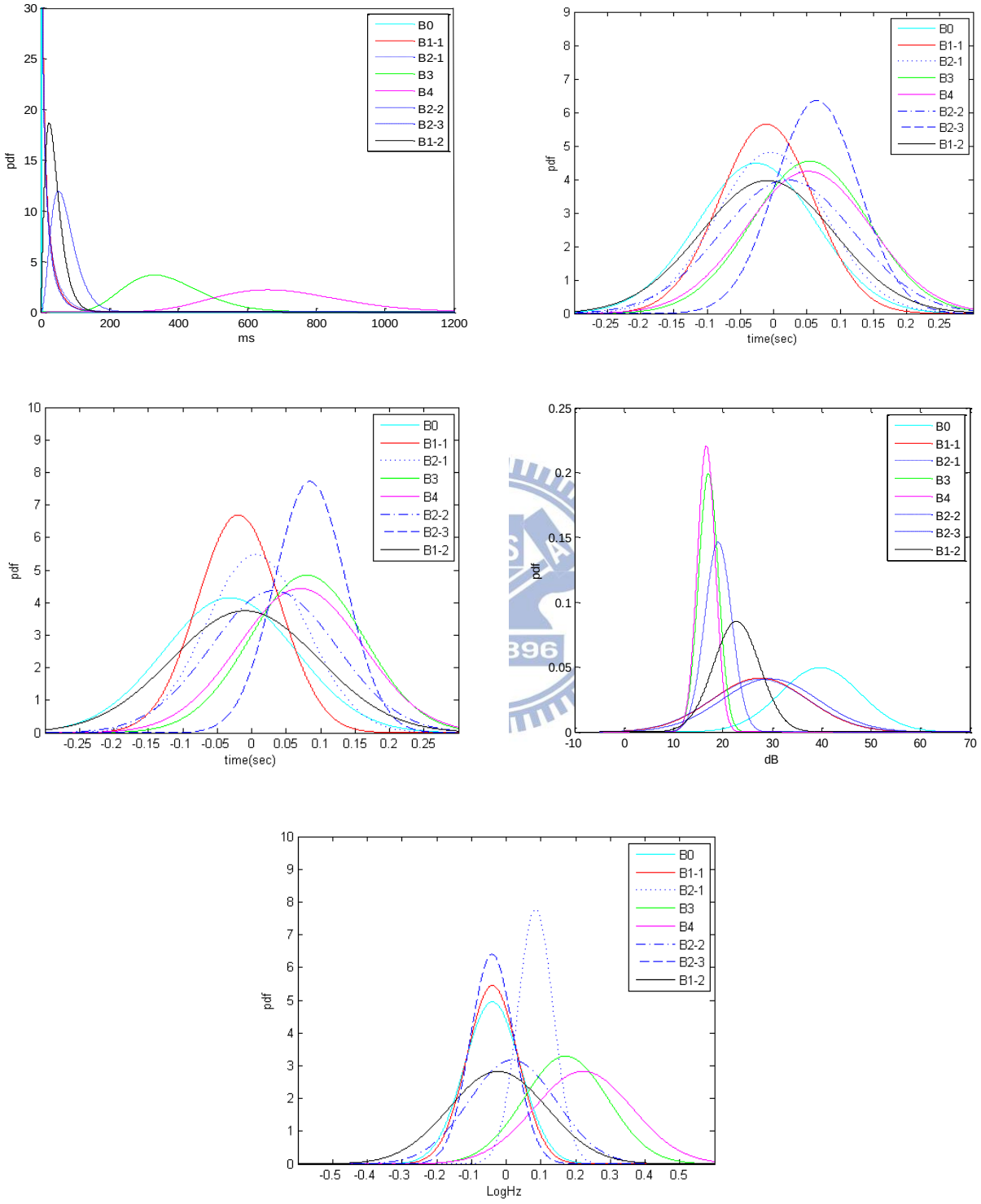


圖 4.3：(a)音節停頓時長，(b)正規化音節延長因子 1，(c)正規化音節延長因子 2，(d)音節間能量低點，(e)正規化基頻跳躍值之決策樹根節點之機率分佈圖

4.2 加入韻律訊息實現中文語音辨認

第二階段的辨識結果除了辨識出詞、字元、音節三種辨識單元，另外將會解碼出詞性(POS)及標點符號(PM)。根據圖 3.7 所示，在這裡我們將比較考慮語速影響的韻律模型(SR Model)與未加入語速影響的韻律模型(Original Model)其辨識效能的差異，以下數據中表格第一列數據為 Baseline 系統，代表在原先第一階段完成後加入 joint syntax model 進行辨識考量；第二列數據為加入 break syntax model 及停頓聲學模型進行辨識考量；第三列數據為加入音節韻律模型及韻律狀態模型進行辨識考量：

表 4.1：詞(word)辨認率

	Original_Model	SR_Model
baseline	86.37%	86.37%
+Prosodic break	87.39%	87.71%
+Prosodic break +Prosodic state	87.93%	88.04%

表 4.2：字(character)辨認率

	Original_Model	SR_Model
baseline	89.65%	89.65%
+Prosodic break	90.39%	90.69%
+Prosodic break +Prosodic state	91.12%	91.16%

表 4.3：音節(syllable)辨認率

	Original_Model	SR_Model
baseline	92.33%	92.33%
+Prosodic break	92.96%	93.20%
+Prosodic break +Prosodic state	93.47%	93.49%

表 4.4 與表 4.5 為加入考慮語速影響的韻律模型進行辨識時，各級中詞性(POS)及標點符號(PM)辨認率的計算，其辨認率計算採用 F-measure 的方式，以下將對此進行詳細說明。

表 4.4：詞性(POS)辨認率

	Precision	**Recall**	**F-Measure**
Baseline	94.36%	83.27%	88.47%
+Prosodic break	94.36%	84.24%	89.02%
+Prosodic break +Prosodic state	94.29%	84.88%	89.34%

表 4.5：標點符號(PM)辨認率

	Precision	**Recall**	**F-Measure**
Baseline	84.90%	74.93%	79.61%
+Prosodic break	88.37%	78.90%	83.37%
+Prosodic break +Prosodic state	83.98%	75.59%	79.56%

4.2.1 詞性(POS)辨認率算法

計算 POS 辨認率並不是直接針對辨認結果以及標準答案來計算，因為當 word 辨認正確時，其辨認出來的 POS 才有意義，所以我們使用另一種計算方式(F-measure)，就是先統計出在 word 辨認正確的條件之下 POS 辨認正確的數量 ***H***，以及在 word 辨認正確的條件之下 POS 總數 ***N***，最後則是 POS 答案中的總數量 ***R***。

有了以上統計結果，接下來則分別計算 POS 的 Recall (***H*** / ***R***) 及在 word 辨認正確的條件之下，POS 的 Precision (***H*** / ***N***)，最後有了 Precision 及 Recall 就能算出 F-measure score，公式如下：

$$F-measure=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{4.1}$$

4.2.2 標點符號(PM)辨認率算法

計算 PM 辨認率並不是直接針對辨認結果以及標準答案來計算，因為當 word 辨認正確時，其辨認出來的 PM 才有意義，所以我們使用另一種計算方式(F-measure)，就是先統計出在 word 辨認正確的條件之下 PM 辨認正確的數量 ***H***，以及在 word 辨認正確的條件之下 PM 總數 ***N***，最後則是 PM 答案中的總數量 ***R***。

有了以上統計結果，接下來則分別計算 PM 的 Recall (***H*** / ***R***) 及在 word 辨認正確的條件之下，PM 的 Precision (***H*** / ***N***)，最後有了 Precision 及 Recall 並依據(4.1)式就能算出 F-measure score。

4.3 辨識結果分析與比較

在此小節中，我們以兩種方式來評估加入語速影響之韻律模型是否能提升辨識系統的效能，並觀察辨識結果改善的部分：

4.3.1 各級辨認結果之比較

在 4.2 小節中，可以發現到每加入一種韻律資訊時，其辨識效能都有小幅的上升，證明這些韻律資訊都是可以幫助辨識的改善。

表 4.6、表 4.7 與表 4.8 列出經由加入韻律模型(加入韻律邊界停頓資訊和音節韻律狀態資訊)改善詞邊界判斷錯誤的搶詞問題、改善一字詞辨識不佳以及聲調修正的情況，以下數據中表格第一欄為正確答案，第二欄為 baseline 系統(加入 joint syntax model)的辨認結果，第三欄則是加入韻律模型(加入 prosodic break 及 prosodic state 資訊)後的辨認結果，並且將解碼出的韻律邊界停頓標示在最右邊。

表 4.6：搶詞狀況的改善

NCKU_f070304_0			
而	NULL	而	B2-1
四	二十	四	B0
歲	歲	歲	B0
以上	以上	以上	B1-1,B1-1
則	則	則	B2-1
有	有	有	B1-1
NCTU_m050107_0			
依據	依據	依據	B1-1,B3
甲等	家長	甲等	B1-1,B1-1
一	依約	一	B1-1
月	NULL	月	B3
乙等	以	乙等	B1-1,B2-2
半	鋼板	半	B1-1
月	NULL	月	B3

表 4.7：一字詞辨認的改善

NCTU_f010413_0			
並 選出 正 副會長	並 指出 會 將	並 選出 正 副會長	B1-1 B1-1,B2-2 B0 B1-1,B1-1,B3
NCTU_f040411_0			
一 口 不 甚 正確	NULL 提出 不 甚 正確	一 處 不 甚 正確	B1-1 B0 B2-1 B1-1 B1-1,B0

表 4.8：聲調的修正

NCTU_f010402_0			
重大 的 影響 因素	重大 的 影響 應 屬	重大 的 影響 因素	B1-2,B0 B2-1 B1-1,B1-1 B1-1,B3
NCKU_m090904_1			
北京市 是	北京市 十	北京市 是	B1-2,B1-2,B2-1 B2-3

五	**五**	**五**	B2-3
個	個	個	B2-1
申辦	申辦	申辦	B1-1,B2-3
兩千年	兩千年	兩千年	B1-1,B1-2,B2-3
奧運	奧運	奧運	B0,B2-1
城市	城市	城市	B0,B1-1
中	中	中	Bend

4.3.2 傳統韻律模型與考慮語速影響之韻律模型之比較

表 4.9 為加入語速影響之韻律模型改善傳統韻律模型的情況，針對一般語速和慢語速的部分各列舉一個例子進行討論，以下數據中表格第一欄為正確答案，第二欄為傳統韻律模型的辨認結果，第三欄為加入語速影響之韻律模型的辨認結果：

表 4.9：加入語速影響的結果改善

NCTU_f010406_0(慢語速)		
黃大洲	黃大洲	黃大洲
體恤	體恤	體恤
市銀行	**市立(B1-1,B3)** **銀行(B1-1,B4)**	**市銀行(B1-2,B1-1,B2-1)**
無法	無法	無法
配合	配合	配合
第二	第二	第二
批	批	批
彩卷	彩卷	彩卷

第五章 結論與未來展望

5.1 結論

中文語音辨識有兩個常見問題：詞邊界判斷錯誤造成搶詞的問題以及一字詞辨識率不佳，在此本研究採用兩階段式的辨認系統，第一階段對詞綴進行處理，建構出一個階層式的語言模型，一方面減少中文 OOV 的問題，另一方面也可降低詞錯誤率；第二階段加入語速影響之韻律模型，考慮語速影響語音的許多現象，經由修正後的韻律模型更清楚地描述兩種韻律標記、文本中的語言參數與語音信號中韻律聲學參數三者之間的關係，並且解碼出多種語言參數。

實驗結果顯示經由本研究提出的階層式中文語音辨識系統，詞(word)、字(character)及音節(syllable)的辨識率分別為 88.04%、91.16%及 93.49%，與第一階段的辨識結果比較，改善了 1.67%、1.45%及 1.02%的絕對錯誤率(12.25%、14.09%及 13.55%的相對錯誤率)，搶詞的問題以及一字詞的錯誤也大幅減少。

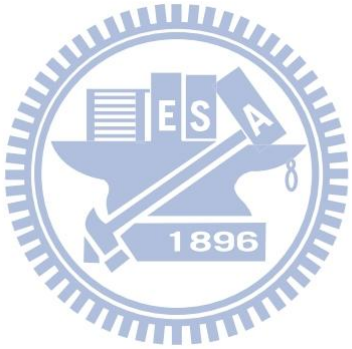
另外針對韻律模型而言，考慮語速影響的韻律模型與未加語速影響的韻律模型，其整體的改善幅度並不大，仔細探討 TCC300 語料庫，發現慢語速的音檔有部分是因為語者本身對於文本的不熟悉性，導致某些字詞長度在語者朗讀時過份被拉長，非語者本身朗讀時的速度快慢，同一個音檔其語者本身說話速度前後也並不一致。整體而言，TCC300 語料庫對於本研究所探討的語速其關係性並不顯著，因本研究是以一段語句的平均音長部分來衡量該語句的語速，但實際上仍需考慮其他的相關韻律參數加以考量，使得語速的測量結果更為可靠，這或許是辨識效能增加不顯著的主要原因之一。

5.2 未來展望

本研究從第一階段詞綴構詞開始到第二階段加入韻律模型其 word lattice 愈長愈龐大，對測試語料重新評分其所需時間也就愈長，未來必須盡可能限制 word lattice 的大小，縮短辨識所需時間。

從本研究可以延伸出三項議題值得未來進一步探討。第一，目前辨識時間過長，未來必須

大幅縮短辨識時間，盡量縮小辨認時的搜尋路徑；第二，從辨識結果可以發現，OOVs 大部分屬於人名的部分，未來在第一階段中的語言模型需考慮人名這項因素，解決人名造成的錯誤率；第三，目前本研究採用 TCC300 語料庫，屬於麥克風朗讀語音，未來若能將此辨識系統延展到更貼近生活化的自發性語音，相信語音辨認可以更廣泛地應用於生活中。



參考文獻

【1】 Peter F. Brown, Vincent J. DellaPietra, Peter V. deSouza, Jennifer C. Lai, and Robert L. Mercer. “Class-based N-gram models of natural language,” Computational Linguistics, vol. 18, no. 4, pp. 467–479, 1992.

【2】 Chien-Pang Chou, “Improvement on Language Modeling for Large-Vocabulary Mandarin Speech Recognition,” NCTU Speech Processing Lab, 2009

【3】 Yun-Shu Yang, “Large-Vocabulary Mandarin Speech Recognition using Hierarchical Language Model,” NCTU Speech Processing Lab, 2010

【4】 Matthew A. Siegler and Richard M. Stern “On The Effects of Speech Rate in Large Vocabulary Speech Recognition Systems”

【5】 F. Martinez, D. Tapias and J. Alvarez “Towards Speech Rate Independence in Large Vocabulary Continuous Speech Recognition”

【6】 T. Pfau, R.Faltlhauser, and G. Ruske “A Combination of Speaker Normalization and Speech Rate Normalization for Automatic Speech Recognition”

【7】 C.-Y. Chiang, S.-H. Chen, H.-M. Yu, and Y.-R. Wang, “Unsupervised joint prosody labeling and modeling for Mandarin speech,” *Journal of the Acoustic Society of America*, vol. 125, no. 2, pp.1164-1183, Feb. 2009.

【8】 Z. Sheng, J.-H. Tao, and D.-L. Jiang,“Chinese prosodic phrasing with extended features,”Proceedings of the IEEE ICASSP 2003, Vol. 1, pp.492-495, 2008

【9】 C.-Y. Tseng, S.-H. Pin, Y.-L. Lee. H.-M. Wang, and Y.-C Chen, “Fluent speech prosody:Framwork and modeling,”Speech Commun. Special issue on quantitive prosody modeling for natural speech description and generation, 46, 284-309 2005

【10】 S.-H. Chen and Y.-R. Wang, “Vector quantization of pitch information in Mandarin speech,” *IEEE Transactions on Communications*, vol. 38, no. 9, pp. 1317-1320, September 1990.

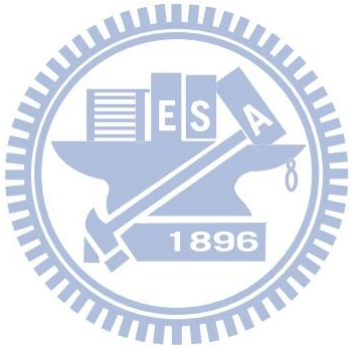
【11】 J. A. Bilmes and K. Kirchhoff, “Factor language models and generalized parallel backoff,” in *Proc. of HLT/NACCL*, 2003, pp. 4-6.

【12】 A. Stolcke, “SRILM – An extensible language modeling toolkit,” in *Proc.* ICSLP, 2002.

【13】 P. Beyerlein, “Discriminative model combination,” in *Proc*. *ICASSP* 1998, pp. 481-484.

【14】 Ming-Chieh Liu, “An Implementation of Prosody-Assisted Mandarin Speech Recognition System,” NCTU Speech Processing Lab, 2011

【15】 C.-Y. Chiang, J.-H. Yang, M.-C. Liu, Y.-R. Wang, Y.-F. Liao, and S.-H. Chen, “A New Model-based Mandarin-speech Coding System,” Proc. of INTERSPEECH-2011, Florence, Italy, pp. 2561-2564, Aug., 2011.



附錄：決策樹之問題集

The question set used to construct the decision trees for building the break syntax model $P(B_n|\mathbf{l}_n)$ and $P(pd_n, ed_n, pj_n, dl_n, df_n \mid B_n, \mathbf{l}_n)$ is listed below:

' Is the inter-syllable location an utterance boundary?'

' Is the inter-syllable location an interword?'

' Does a PM exist at the inter-syllable location'

' Does a Major PM exist at the inter-syllable location '

' Does a。exist at the inter-syllable location '

' Does a，exist at the inter-syllable location '

' Does a、exist at the inter-syllable location '

' Does a．exist at the inter-syllable location '

' Does a；exist at the inter-syllable location '

' Does a：exist at the inter-syllable location '

' Does a？exist at the inter-syllable location '

' Does a！exist at the inter-syllable location '

' Does a（exist at the inter-syllable location '

' Does a）exist at the inter-syllable location '

' Is the the preceding special prefix words + special 1-syllable words: Ng, Ncd, Di, DE, I, T'

' Is the POS of the preceding word A'

' Is the POS of the preceding word C'

' Is the POS of the preceding word D'

' Is the POS of the preceding word N'

' Is the POS of the preceding word I or T'

' Is the POS of the preceding word P'

' Is the POS of the preceding word V'

' Is the POS of the preceding word DE'

' Is the POS of the preceding word SHI'

' Is the POS of the preceding word FW'

' Is the POS of the preceding word DM'

' Is the POS of the preceding word Da Di Dk D'

' Is the POS of the preceding word Dfa'

' Is the POS of the preceding word Dfb'

' Is the POS of the preceding word Na Nb Nc Nv'

' Is the POS of the preceding word Nd'

' Is the POS of the preceding word Neu Nes Nep Neqa Neqb Nf'

' Is the POS of the preceding word Ng Ncd'

' Is the POS of the preceding word Nh'

' Is the POS of the preceding word VA VAC VG'

' Is the POS of the preceding word VB VC VCL VD VE VF VJ VK VL'

' Is the POS of the preceding word VH VHC VI'

' Is the POS of the preceding word V_2'

' Is the POS of the preceding word Caa'

' Is the POS of the preceding word Cab'

' Is the POS of the preceding word Cba'

' Is the POS of the preceding word Cbb'

' Is the POS of the preceding word Da'

' Is the POS of the preceding word Di'

' Is the POS of the preceding word Dk'

' Is the POS of the preceding word D'

' Is the POS of the preceding word Na'

' Is the POS of the preceding word Nb'

' Is the POS of the preceding word Nc'

' Is the POS of the preceding word Ncd'

' Is the POS of the preceding word Neu'

' Is the POS of the preceding word Nes'

' Is the POS of the preceding word Nep'

' Is the POS of the preceding word Neqa'

' Is the POS of the preceding word Neqb'

' Is the POS of the preceding word Nf'

' Is the POS of the preceding word Ng'

' Is the POS of the preceding word Nv'

' Is the POS of the preceding word I'

' Is the POS of the preceding word T'

' Is the POS of the preceding word VA'

' Is the POS of the preceding word VAC'

' Is the POS of the preceding word VB'

' Is the POS of the preceding word VC'

' Is the POS of the preceding word VCL'

' Is the POS of the preceding word VD'

' Is the POS of the preceding word VE'

' Is the POS of the preceding word VF'

' Is the POS of the preceding word VG'

' Is the POS of the preceding word VH'

' Is the POS of the preceding word VHC'

' Is the POS of the preceding word VI'

' Is the POS of the preceding word VJ'

' Is the POS of the preceding word VK'

' Is the POS of the preceding word VL'

' Is the length of the preceding word 1'

' Is the length of the preceding word 2'

' Is the length of the preceding word 3'

' Is the length of the preceding word 4'

' Is the length of the preceding word 5'

' Is the length of the preceding word 6'

' Is the length of the preceding word less than 2'

' Is the length of the preceding word less than 3'

' Is the length of the preceding word less than 4'

' Is the length of the preceding word less than 5'

' Is the length of the preceding word less than 6'

' Is the following special 1-syllable words: Ng, Ncd, Di, DE, I, T + special suffix words'

' Is the POS of the following word A'

' Is the POS of the following word C'

' Is the POS of the following word D'

' Is the POS of the following word N'

' Is the POS of the following word I or T'

' Is the POS of the following word P'

' Is the POS of the following word V'

' Is the POS of the following word DE'

' Is the POS of the following word SHI'

' Is the POS of the following word FW'

' Is the POS of the following word DM'

' Is the POS of the following word Da Di Dk D'

' Is the POS of the following word Dfa'

' Is the POS of the following word Dfb'

' Is the POS of the following word Na Nb Nc Nv'

' Is the POS of the following word Nd'

' Is the POS of the following word Neu Nes Nep Neqa Neqb Nf'

' Is the POS of the following word Ng Ncd'

' Is the POS of the following word Nh'

' Is the POS of the following word VA VAC VG'

' Is the POS of the following word VB VC VCL VD VE VF VJ VK VL'

' Is the POS of the following word VH VHC VI'

' Is the POS of the following word V_2'

' Is the POS of the following word Caa'

' Is the POS of the following word Cab'

' Is the POS of the following word Cba'

' Is the POS of the following word Cbb'

' Is the POS of the following word Da'

' Is the POS of the following word Di'

' Is the POS of the following word Dk'

' Is the POS of the following word D'

' Is the POS of the following word Na'

' Is the POS of the following word Nb'

' Is the POS of the following word Nc'

' Is the POS of the following word Ncd'

' Is the POS of the following word Neu'

' Is the POS of the following word Nes'

' Is the POS of the following word Nep'

' Is the POS of the following word Neqa'

' Is the POS of the following word Neqb'

' Is the POS of the following word Nf'

' Is the POS of the following word Ng'

' Is the POS of the following word Nv'

' Is the POS of the following word I'

' Is the POS of the following word T'

' Is the POS of the following word VA'

' Is the POS of the following word VAC'

' Is the POS of the following word VB'

' Is the POS of the following word VC'

' Is the POS of the following word VCL'

' Is the POS of the following word VD'

' Is the POS of the following word VE'

' Is the POS of the following word VF'

' Is the POS of the following word VG'

' Is the POS of the following word VH'

' Is the POS of the following word VHC'

' Is the POS of the following word VI'

' Is the POS of the following word VJ'

' Is the POS of the following word VK'

' Is the POS of the following word VL'

' Is the length of the following word 1'

' Is the length of the following word 2'

' Is the length of the following word 3'

' Is the length of the following word 4'

' Is the length of the following word 5'

' Is the length of the following word 6'

' Is the length of the following word less than 2'

' Is the length of the following word less than 3'

' Is the length of the following word less than 4'

' Is the length of the following word less than 5'

' Is the length of the following word less than 6'

Is the initial of the following syllable a null one or in { m, n, l, r}?

Is the initial of the following syllable a null one or in { b, d, g}?

Is the initial of the following syllable a null one or in { f, s, sh, h}?

Is the initial of the following syllable a null one or in { c, ch, q}?

Is the initial of the following syllable a null one or in { p, t, k}?

Is the initial of the following syllable a null one or in { z, zh, j}?