

國立交通大學

電信工程學系

碩士論文

中文自發性語音之韻律標記及韻律模式
Joint Prosody Labeling and Modeling for Mandarin
Spontaneous Speech

研究生：周裕倫

指導教授：王逸如 博士

中華民國九十八年七月

中文自發性語音之韻律標記及韻律模式
Joint Prosody Labeling and Modeling for Mandarin
Spontaneous Speech

研 究 生：周裕倫

Student：Yu-Lun Chou

指導教授：王逸如 博士

Advisor：Dr. Yih-Ru Wang

國 立 交 通 大 學

電 信 工 程 學 系

碩 士 論 文



Submitted to Department 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 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年七月

中文自發性語音之韻律標記及韻律模式

研 究 生：周裕倫

指導教授：王逸如 博士

國立交通大學電信工程學系碩士班

中文摘要

韻律於許多語音處理研究中為相當有幫助的資訊，但前提是需要大量已標記的語料庫，並利用統計式的方式達成。由於語料庫的標記相當費時費力，特別是在自發性語音，所以當今已有韻律標記之中文自發性語音語料庫仍是寥寥無幾，因此本研究針對中央研究院所提供之「現代漢語口語對話語料庫」進行韻律之標記。

本研究以自動的方式實現韻律標記，並且探討自發性語音之韻律變化。在此即利用語音信號中之韻律參數以及文字中之語言學參數，以非監督式的方法對此語料庫進行韻律標記，並訓練其韻律模型。本研究所標記的韻律標記為停頓標記及韻律狀態，其中停頓標記表示韻律單位的邊界，而韻律狀態的序列所代表的是上層韻律單位之變化。進而透過分析本研究所訓練出的模型參數，探討自發性語音中上層韻律變化的狀況，以及韻律標記、韻律參數、語言參數三者之間的關係，並且亦針對自發性語音中相較於朗讀式語音所沒有的一些特性進行分析。

本研究發現，此語料庫經過自動標記以後，在自發性語音之韻律表現和特殊現象、以及詞語修補發現許多韻律變化現象，這些發現將可幫助未來進行自發性語音之相關研究提供更多有用的資訊。

Joint Prosody Labeling and Modeling for Mandarin Spontaneous Speech

Student : Yu-Lun Chou

Advisor : Dr. Yih-Ru Wang

Department of Communication Engineering
National Chiao Tung University

Abstract

In recent years, prosodic information are widely used in spontaneous speech processing. In those previous works, prosodic features are firstly extracted from speech corpus labeled with prosody tags or boundary types, and then prosodic models are built and used in the tasks. However, to prepare a large spontaneous speech corpus with prosody tags being properly labeled is in general a difficult task, performance of human labeling can not be guaranteed even if well-experienced annotators are involved.

In this thesis, an investigation on the prosody of Mandarin spontaneous speech is conducted by using the unsupervised joint prosody labeling and modeling (PLM) method proposed previously for read speech. It labels an unlabeled spontaneous Mandarin speech corpus with two types of prosody tags, break type of inter-syllable juncture and prosodic state of syllable, and builds four prosodic models simultaneously. The break tag can be used to delimit prosodic constituents of a hierarchical prosody structure, and the prosodic state can be used to construct the prosodic feature patterns of prosodic constituents. The four prosodic models describe the relationships of acoustic prosodic features, prosody tags of utterances, and the linguistic features of the associated texts. Its effectiveness was confirmed by the experimental results on an unlabeled

dialogue corpus, MCDC. Many meaningful characteristics of spontaneous-speech prosody were explored from the parameters of the well-trained prosodic models. The patterns of high-level prosodic constituents of a prosody hierarchy were derived. An analysis of disfluencies related to the labeling results was also discussed. Those results would be very advantageous in providing rich prosodic information for ASR.



致謝

首先非常感謝陳信宏老師和王逸如老師在研究上的耐心指導，才讓我能順利走到寫致謝這一步，感謝陳老師能在研究上給我許多的指點並且幫助我完成大大小小的論文，感謝王老師指導我如何從一個大學生變成研究生以及給我許多成長磨練的機會，和兩位老師的閒聊之中也讓我學到許多生活和做事上的方法及態度。

接著要感謝一路教導我和一起奮戰的振宇學長、交大星光幫阿德學長、教我許多研究技巧的智合學長、為我每晚介紹實驗室歷史的輝哥學長、讓我了解業界和當兵情況的巴金叔、愛護地球的希群學長以及從高中就一直指導我到現在的至鴻學長，你們除了教導我研究上的專業知識，也為我點亮一窄明燈，讓我自己對未來的路更清楚，和你們一起討論研究是一段辛苦卻又快樂的時光。感謝和我一起奮戰兩年的交大宋不挑、官腔杜 QQ、一起睡實驗室奮戰的小帥哥，無論是修課或拼畢業都一起爆肝渡過。還有陪我一起去香港圓夢的妞妞、不辭辛勞幫我標記語料庫的撲馬、每晚一起看中職嘴砲的承燁、讓我嵌入式系統能順利修過的宥余、一起耍白痴的皓翔、常常搞笑的天然呆舒舒、幫我改作業的小卡，感謝學弟妹們的幫忙，你們讓實驗室充滿更多歡樂的氣氛。還有一年就偷跑去聯發科的小廣、該打個牌了吧的庚達還有黑字營營長阿宅，你們留下的奮戰精神是我撐下來的力量。我還要感謝高中和大學的夥伴們，子軒、包子、大 B、小強、溫董、阿福、翁老，你們是我的解悶良方，讓我在研究生生活之外有更多采多姿的生活，還有陪我一起走過研究所這條路上各種快樂和憂愁的思思妹。最後我要由衷感謝爸爸、媽媽及遠在美國的舅舅，和其他家人長久以來的支持，從小到大對我的教育和栽培一點也沒少，才讓我能夠順利走到完成碩士學業這一步。僅將此篇論文獻給人生旅途上的各位，接下來是該往下一個目標前進的時候了，GO!!

目錄

中文摘要.....	I
Abstract.....	II
致謝.....	IV
目錄.....	V
表目錄.....	VIII
圖目錄.....	IX
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究方向.....	2
1.3 相關研究.....	3
1.3.1 聲學模型之相關研究.....	3
1.3.2 韻律標記及模型之相關研究.....	3
1.4 章節概要說明.....	4
第二章 現代漢語口語對話語料庫介紹.....	5
2.1 語料庫簡介.....	5
2.1.1 文字轉寫.....	5
2.1.2 語音資料.....	7
2.2 自發性語音特性之簡介.....	7
2.3 語料庫相關統計.....	11
第三章 自發性語音聲學模型.....	15
3.1 訓練語料及測試語料.....	15
3.2 聲學模型之建立.....	16
3.2.1 特徵參數抽取.....	16

3.2.2 聲學模型之建立流程.....	16
3.3 實驗結果.....	18
第四章 自發性語音韻律模型.....	23
4.1 韻律模型之設計.....	23
4.1.1 中文語音韻律階層式架構.....	23
4.1.2 模型設計.....	25
4.2 韻律模型參數之萃取.....	30
4.2.1 聲學參數抽取.....	31
4.2.2 韻律相關聲學參數之萃取.....	34
4.2.3 韻律相關語言學參數之萃取.....	36
4.3 模型訓練.....	37
4.3.1 初始化.....	38
4.3.1 重覆疊代.....	47
第五章：實驗結果.....	49
5.1 音節韻律模型.....	49
5.1.1 音節層次之影響型態.....	51
5.1.2 上層韻律狀態之影響型態.....	58
5.2 停頓標記聲學模型.....	61
5.3 韻律狀態轉移模型.....	62
5.4 停頓標記語言模型.....	65
5.5 韻律標記結果之分析.....	68
5.5.1 停頓標記及韻律狀態之分析.....	68
5.5.2 上層韻律單元內之變化情況.....	68
5.5.3 韻律標記範例.....	74
5.6 韻律標記結果與詞語修補之討論.....	75
第六章：結論與未來展望.....	83

6.1 結論.....	83
6.2 未來展望.....	83
參考文獻.....	84
附錄一.....	87
附錄二.....	89
附錄三.....	105
附錄四.....	107
附錄五.....	109



表目錄

表 2.1：對話主題與語者對照表.....	6
表 2.2：各語者之基頻、音節長度以及能量統計值.....	12
表 2.3：IP 於 MCDC 語料庫中出現之機率	14
表 3.1：訓練語料統計.....	15
表 3.2：測試語料統計.....	16
表 3.3：HMM 模型之設定.....	17
表 3.4：音節辨認率.....	18
表 3.5：高錯誤率，且刪除型錯誤機率前十名之音節.....	19
表 3.6：經由 ASR 歸納之常見音節合併現象，其中粗體為被刪除之音節，syllable pair merge rate 即為此種 canonical form 的狀況下，出現此種 surface form 之機率	20
表 4.1：韻律標記、聲學參數以及語言參數之數學符號.....	26
表 5.1：基本音節中，不同組合之 AP 下音節韻律模型參數之 TRE.....	50
表 5.2：特殊音節中，不同組合之 AP 下音節韻律模型參數之 TRE.....	50
表 5.3：不同韻律單元之音節長度.....	69
表 5.4：不同組合之 AP 下之 TRE.....	74

圖目錄

圖 2.1：MCDC 語料庫中，文字以及漢語拼音轉寫之架構圖	7
圖 2.2：詞語修補之基本型態以及範例，其中「*」為 IP 點	10
圖 2.3：MCDC 語料庫中，音節聲調分布圖	12
圖 2.4：MCDC 語料庫中，詞長分布圖	13
圖 2.5：MCDC 語料庫中，POS 種類分布圖	13
圖 2.6：MCDC 語料庫中，每段切割之音節數分布圖	14
圖 3.1：基準系統之建立流程	17
圖 3.2：加入狀態轉移機率示意圖	18
圖 3.3：加入音節合併資訊之聲學模型示意圖	21
圖 3.4：加入音節合併資訊後之辨識率	22
圖 4.1：中文語音韻律之階層式架構概念	23
圖 4.2：本研究使用之中文自發性語音韻律階層式架構	24
圖 4.3：音節長度之 (a)數值分布及 (b)累積分布	25
圖 4.4：音節韻律參數及音節間韻律參數	31
圖 4.5：語音開頭時基頻不穩與半頻現象及其修正後基頻軌跡	32
圖 4.6：語者編號 01L 之 (a)基頻資料及其半頻偵測模型和 (b)修正後基頻資料	32
圖 4.7：(a)包含極端長停頓之停頓長度資料分布和 (b)修正後之資料分布及 (c) S 型函數	36
圖 4.8：分類停頓標記之決策樹示意圖	40
圖 4.9：已標記音節邊界之音節停頓長度分佈	41
圖 4.10：Th1, Th2 和 Th3 之定義方法：計算 (a) B4、B3 和 B2-2 與 (b)其他停頓標記之停頓 長度的機率分佈以及 (c) B4、B3、B2-2 和其他停頓標記之門檻值	42
圖 4.11：Th5 之定義方法：計算 (a)詞內與 (b)已標記之音節邊界正規化基頻差之機率分佈以 及 (c)門檻值之定義	43

圖 4.12：Th6 之定義方法：計算 (a)B0 和 B1 音節能量低點之機率分佈以及 (b)門檻值之定義	44
圖 4.13：Th7 之定義方法：計算 (a)詞內音節邊界和 (b)已標記之音節邊界的相鄰兩音節正規化音節延長因子之機率分佈以及 (b)門檻值之定義	45
圖 5.1：疊代次數與標的函數值	49
圖 5.2：五個中文聲調之 AP	51
圖 5.3：聲調決策樹之示意圖	53
圖 5.4：考慮連音現象一聲聲調之 AP	53
圖 5.5：考慮連音現象二聲聲調之 AP	54
圖 5.6：考慮連音現象三聲聲調之 AP	55
圖 5.7：考慮連音現象之四聲聲調 AP	56
圖 5.8：基本音節中 (a)聲調以及特殊音節中 (b)特殊音節型態基頻 AP 之第一維數值	56
圖 5.9：基本音節中 (a)聲調、(b)基本音節型態以及特殊音節中 (c)特殊音節型態之音節長度 AP	57
圖 5.10：基本音節中 (a)聲調、(b)韻母型態以及以及特殊音節中 (c)特殊音節型態之能量 AP	58
圖 5.11：基本音節中，基頻之 (a)各韻律狀態之數值及 (b)韻律狀態分布圖與 (c)正規化基頻數值分布圖之比較	59
圖 5.12：基本音節中，音節長度之 (a)韻律狀態之數值及 (b)韻律狀態分布圖與 (c)正規化音節長度分布圖之比較	59
圖 5.13：基本音節中，音節能量之 (a)各韻律狀態之數值及 (b)韻律狀態分布圖與 (c)正規化音節分布圖能量之比較	59
圖 5.14：(a)基本音節及 (b)特殊音節之基頻韻律狀態數值	60
圖 5.15：(a)基本音節及 (b)特殊音節之音節長度韻律狀態數值	60
圖 5.16：(a)基本音節及 (b)特殊音節之能量韻律狀態數值	61
圖 5.17：(a)音節停頓長度 (b)音節間能量低點 (c)正規化基頻跳躍值與 (d)正規化音節延長	

因子之分布圖，其中括號中之數值為平均值。.....	62
圖 5.18：各停頓標記下，基頻韻律狀態轉移的狀況，顏色越深的線表示此種狀態轉移的機率越大.....	63
圖 5.19：各停頓標記下，音節長度韻律狀態轉移的狀況，顏色越深的線表示此種狀態轉移的數量越多.....	64
圖 5.20：各停頓標記下，音節能量韻律狀態轉移的狀況，顏色越深的線表示此種狀態轉移的數量越多.....	65
圖 5.21：停頓標記語言模型決策樹之示意圖，其中停頓標記分布圖之橫軸由左至右分別為 $B0$ 、 $B1$ 、 $B2-1$ 、 $B2-2$ 、 $B2-3$ 、 $B3$ 以及 $B4$	66
圖 5.22：停頓標記語言模型之決策樹。.....	67
圖 5.23：詞間音節邊界停頓標記語言模型之決策樹.....	67
圖 5.24：(a) 所有音節邊界之停頓標記分布圖以及 (b)排除特殊音節邊界之停頓標記分布圖68	
圖 5.25：各種韻律單元在不同長度下之基頻影響型態，顏色越深的線表示此種長度之韻律單元數量越多.....	71
圖 5.26：各種韻律單元在不同長度下之音節長度影響型態，顏色越深的線表示此種長度之韻律單元數量越多.....	72
圖 5.27：各種韻律單元在不同長度下之能量影響型態，顏色越深的線表示此種長度之韻律單元數量越多.....	73
圖 5.28：韻律標記範例：實線代表 PPh 或 PG/BG 之邊界，虛線代表 PW 之邊界，點線代表隔離特殊韻律現象之邊界；由上至下為基頻，音節長度以及能量之韻律狀態標記，空心圓表示各參數韻律狀態加上總體平均值之數值，而實心圓為實際獲得的參數值.....	74
圖 5.29：各型態 IP 對應之停頓標記分布圖.....	75
圖 5.30：(a)語用功能之重覆與 (b)類似部分重覆之重覆型態範例：虛線代表 PW 之邊界，點線代表隔離特殊韻律現象之邊界；由上至下為基頻(LogHz)，音節長度(sec)以及能量(dB)之韻律狀態標記，空心圓表示各參數韻律狀態加上總體平均值之數值，而實心圓為實際獲得的參數值.....	76

圖 5.31：詞語重覆 (a)兩段分割、 (b)三段分割、 (c)四段分割與 (d)五段分割中，不同長度下，基頻之變化，顏色越深的線表示分割中此種長度之數量越多	78
圖 5.32：詞語重覆 (a)兩段分割、 (b)三段分割、 (c)四段分割與 (d)五段分割中，不同長度下，音節長度之變化，顏色越深的線表示分割中此種長度之數量越多	79
圖 5.33：詞語重覆 (a)兩段分割、 (b)三段分割、 (c)四段分割與 (d)五段分割中，不同長度下，能量之變化，顏色越深的線表示分割中此種長度之數量越多	80
圖 5.34：部分重覆中，兩段分割在不同長度下，基頻之變化，顏色越深的線表示分割中此種長度之數量越多	81
圖 5.35：部分重覆中，兩段分割在不同長度下，音節長度之變化，顏色越深的線表示分割中此種長度之數量越多	81
圖 5.36：部分重覆中，兩段分割在不同長度下，能量之變化，顏色越深的線表示分割中此種長度之數量越多	81
圖 5.37：詞語更正中，兩段分割在不同長度下，基頻之變化，顏色越深的線表示分割中此種長度之數量越多	82
圖 5.38：詞語更正中，兩段分割在不同長度下，音節長度之變化，顏色越深的線表示分割中此種長度之數量越多	82
圖 5.39：詞語更正中，兩段分割在不同長度下，能量之變化，顏色越深的線表示分割中此種長度之數量越多	82

第一章 緒論

1.1 研究動機

近幾年在語音辨認 (Automatic Speech Recognition, ASR) 上之研究，已漸漸由朗讀式語音 (read speech) 轉向自發性語音 (spontaneous speech)，由於朗讀式語音已有良好的辨認系統，但自發性語音之辨認率仍舊與朗讀式語音有一段差距，而造成自發性語音辨認率不如朗讀式語音之原因，主要為自發性語音有較快的語速 (speaking rate)、語者發音時未經大腦良好的規劃以及語音形態的差異 (mismatch of speech style)。

上述這些原因會造成自發性語音有一些特殊的語音現象。首先，由於自發性語音有較快的語速，人類為了節省發聲時所需的力量，會造成某些語音在發音時被省略或產生發音變異 (pronunciation variation) 以及音節合併 (syllable contraction) 等現象；此外自發性語音未經大腦良好的規劃，使語流中常會出現遲疑 (hesitation)、詞語修補 (disfluency) 等現象；然而語者發音時未經大腦良好的規劃以及語音形態的差異，使自發性語音中有不合乎文法結構之語句，以及許多感歎詞 (particle)、語者慣用的語助詞 (marker) 出現。以上都是會造成自發性語音和朗讀式語音相比辨認率大幅下降之原因。

Furui【1】在以自發性語音之理解 (spontaneous speech understanding) 為目標下，提出幾點必須要解決之課題，首先要達到語音理解就必須有語句邊界偵測 (sentence boundary detection) 和語音摘要 (speech summarization) 這些技術，但這些技術都必須經由前端 ASR 所支持。而在 ASR 中的聲學模型 (Acoustic Model, AM) 中，因為發音變異或音節合併這些現象而必須作修正；另外語言模型 (Language Model, LM) 則因語音形態和朗讀式語音之差異，以及難以大量獲得自發性語音之文字轉寫 (transcription)，所以勢必要以語言模型調適 (LM adaptation) 的方式獲得適合自發性語音之語言模型。此外在自發性語音中另一個非常特殊的現象，即語流中之詞語修補，若能偵測詞語修補中斷點 (Interrupt Point, IP) 的訊息不但可以幫助自發性語音之理解，也可以幫助 ASR 上面的辨認。

由前人的文獻可知韻律在語句邊界偵測【2】、詞語修補中斷點偵測【2-3】以及自發性語音辨認【4-6】，這一些問題上都有所助益，前人的文獻並顯示若能有一個良好的韻律標記系統，不但能利用韻律標記訊息使統計式模型之效能增加，並能使訓練模型時，不再是直接輸入一連串的韻律參數，而是標記後之韻律行為，進而降低模型之複雜度，且能實際觀察到自發性語音中各種事件所對應到之韻律現象，有效的處理上述自發性語音所產生之問題，最終幫助自發性語音之辨認。

1.2 研究方向

在本篇論文中，將以獲得一個良好的韻律標記系統為目標，而基頻軌跡、音節長度以及音節能量是韻律當中三個最重要的參數，因此首先我們必須得到音節邊界的資訊，才能進一步獲得音節內之基頻軌跡、長度以及能量。為了節省人力成本及時間，本論文使用一個自動求取音節邊界的方法，即利用聲學模型經由語音信號和文字的強迫對齊（forced alignment）後獲得音節切割資訊，因此本論文首先以建立一個自發性語音之聲學模型為目標，以獲得音節邊界資訊。

接著本研究為了能獲得一個省時且具一致性的自發性語音韻律標記，在此利用江振宇博士所提出之非監督式中文語音韻律標記及韻律模型【7】為基礎作修改，以獲得適合自發性語音之韻律標記系統。此方法以中文語音韻律階層式架構為基礎，利用語音信號中萃取之韻律聲學參數，包含上述的音節基頻軌跡、音節長度以及音節能量，和文字中萃取之語言學參數資訊，對語音中每一個音節邊界（syllable juncture）的停頓狀態以及上層韻律的變化作標記，試圖利用這些韻律標記（prosodic tags）描述自發性語音中韻律的變化，而獲得一個良好的中文自發性語音韻律標記。最後本研究利用韻律標記之結果，觀察詞語修補及其相對應之韻律標記情況並加以分析。

1.3 相關研究

1.3.1 聲學模型之相關研究

在自發性語音聲學模型中所遭遇到的問題，主要在於發音變異使聲學模型混淆以致效能下降。而目前發音變異的問題主要可以由語音辨認中的兩個層次來解決，首先是在辨認字典（lexicon）的詞中加入可能之發音變異【8】。另外則是改進底層之聲學模型；文獻【9】中提出以強迫辨認（forced recognition）以及反覆的修正語音之轉寫（iterative transcribing），以獲得較符合真正語音信號之文字轉寫，並重新訓練聲學模型，藉此得到較能描述真正發音之聲學模型；文獻【10】中則利用決策樹的方式決定額外訓練的發音變異聲學模型，接著使用 state tying 以及 mixture tying 的方式，來得到具有高鑑別能力以及低複雜度的聲學模型。

1.3.2 韻律標記及模型之相關研究

將韻律資訊利用至語音辨認主要分為三類，第一類為利用韻律參數對辨認結果所產生之詞格（word lattice）重新計算分數，直接利用韻律參數來驗證（verification）在詞格中不同路徑其對應切割位置之可靠程度【4】；第二類則為以事件為基礎（event-based）的方式增加語音辨認之效能【5】，利用韻律參數建立一個偵測事件之模型，例如：類語句邊界（sentence-like unit）或詞語修補中斷點，並利用事件及詞的序列一起建立語言模型，對辨認結果所產生之詞格重新計算分數；第三類則是利用韻律以及句法的關係建立韻律相關的語言模型（prosody dependent LM），來描述韻律以及詞之結合機率，並利用韻律邊界的資訊建立韻律相關的聲學模型（prosody dependent AM）【6】。

由以上可得知，有一個良好的韻律的標記系統，是可以有效地幫助語音辨認以及訓練韻律相關的模型，而在許多朗讀式語音資料庫中都已經有好的韻律標記，但具有韻律標記之自發性語音語料庫為數較少。在過去自發性語音的韻律標記，主要以ToBI【11】以及它的延伸為主，例如：C-ToBI【12】、X-JToBI【13】以及將ToBI運用於Switchboard和CallHome語料庫

中【14】。Batliner等人【15】利用簡單的句法分析發展韻律標記系統，將每一個詞的邊界分為三個主要的M-Class或分成五個語法類別（syntactic classes），其中三個主要的M-Class有：M0—子句內部邊界（clause internal）、M3—子句邊界（clause boundary）以及MU—需要經由人工判斷的邊界，而五個語法類別主要有：S1—特殊詞之邊界（boundary after a particle）、S2—詞組邊界（phrase boundary）、S3—子句邊界（clause boundary）、S4—語句邊界（sentence boundary）以及S0—無邊界（no boundary）。但以上這些韻律標記系統都需要人為的介入，使標記時較為費力耗時，且經由不同的人標注會使標註結果失去一致性。

1.4 章節概要說明

本論文的内容共分為五章：

第一章：緒論：介紹本論文之研究動機與研究方向。

第二章：現代漢語口語對話語料庫介紹：介紹本研究使用之自發性語音語料庫及其特性與統計分析。

第三章：自發性語音聲學模型：建構自發性語音之聲學模型。

第四章：自發性語音韻律模型：建構自發性語音之韻律模型以及其模型訓練之演算法。

第五章：實驗結果：對韻律模型參數與韻律標記之結果作分析，並觀察自發性語音中詞語修補之特性。

第六章：結論與未來展望。

第二章 現代漢語口語對話語料庫介紹

現代漢語口語對話語料庫 (Mandarin Conversational Dialogue Corpus, MCDC) 【16】是由中央研究院語言學研究所曾淑娟博士等人於2000~2002年間所錄製，其語者是由台北市民隨機抽樣，並依據16~25歲、26~35歲以及36~45歲三大年齡層，選出60位語者(37位女性、23位男性)，共錄製30段對話，但其中有轉寫的對話僅有8段對話，分別為編號01、02、03、05、09、10、25以及26，其中包含了16位語者(9位女性、7位男性)，本研究將以此8段對話作為實驗之語料。本章節將對此語料庫作簡介。在2.1節將介紹MCDC語料庫之資料格式；在2.2節將介紹此語料庫之自發性語音特性；在2.3節將介紹在MCDC語料中語言學上或聲學上之統計資料。

2.1 語料庫簡介



2.1.1 文字轉寫

MCDC語料庫為兩語者對話之語料，音檔採雙聲道的方式錄製，兩個聲道分別錄製一位語者之語料，其簡介如表2.1所示。在標註時大致上以對話中語者轉換處為一個段落作轉寫，轉寫內容主要包括：對應之音檔名稱、發音人、音檔起始及結束時間、語音之文字轉寫以及其發音相對之漢語拼音，文字以及漢語拼音的轉寫包括語言及非語言部分，如圖2.1所示，其中語言部分包含發音標記及其他語音中之特殊現象，發音標記乃採用SAMPA (Speech Assessment Methods Phonetic Alphabet) 音標系統，而其他語言中之特殊現象將在2.2節作詳細的介紹；另外在非語言部分主要是標記非人類產生之室內雜訊，以及人類所產生之副語言現象 (para-linguistic) 或非語言現象 (non-linguistic)，例如：咳嗽聲、笑聲、呼吸聲等。此外MCDC語料庫在標記時是採用類似XML (Extensible Markup Language) 資料格式的標記方式，以下為一個段落之文字轉寫範例及說明：

<segment> /*語者轉換開始處*/
 <voicefile>D:\MCDC\stereo_01\mc dc-01-01.wav /*對應之音檔名稱*/
 <speaker>MISC-07-female-29 /*發音人*/
 <start>020976 /*此段文字轉寫對應音檔之開始時間*/
 <end>025360 /*此段文字轉寫對應音檔之結束時間*/
 <translator>Fen /*文字轉寫人*/
 <chinese> /*內容文字標記*/
 O 我在一家公關公司上班 (unrecognizable non-speech sound)
 </chinese>
 <english> /*語者發音之漢語拼音標記*/
 O wo3 zai4 yi4 jia1 gong1 guan1 gong1 si1 shang4 ban1 (unrecognizable non-speech sound)
 </english>
 <comment>
 </comment>
 </segment>

表 2.1：對話主題與語者對照表

對話序號	長度 (分鐘)	發音人	聲道 (L/R)	語者編號	對話主題
mc dc-01	61	MISC-08-male-25	R	01R	工作、休閒活動 、經濟、開車
		MISC-07-female-29	L	01L	
mc dc-02	63	MISC-10-male-35	R	02R	休閒活動、經濟 、工作、性別、政治
		MISC-09-female-37	L	02L	
mc dc-03	61	MISC-12-female17	R	03R	家庭、學校、購物 、生涯規劃、明星
		MISC-11-female16	L	03L	
mc dc-05	63	MISC-15-male-40	L	05L	工作、家庭、社會階級 、保險、歷史、省籍情節
		MISC-16-female-46	R	05R	
mc dc-09	66	MISC-23-female-30	R	09R	工作、旅行、生活態度 、環保、健康
		MISC-24-female-35	L	09L	
mc dc-10	54	MISC-26-male-23	R	10R	電影、政治、軍隊 、捷運、學校、經濟
		MISC-25-male-35	L	10L	
mc dc-25	55	MISC-57-male-43	L	25L	交通、工作、小孩 、旅行、電腦、管理
		MISC-58-female-45	R	25R	
mc dc-26	46	MISC-60-male-24	R	26R	工作、求職、家庭 、車禍、學英文、婚姻、軍隊
		MISC-59-female-37	L	26L	

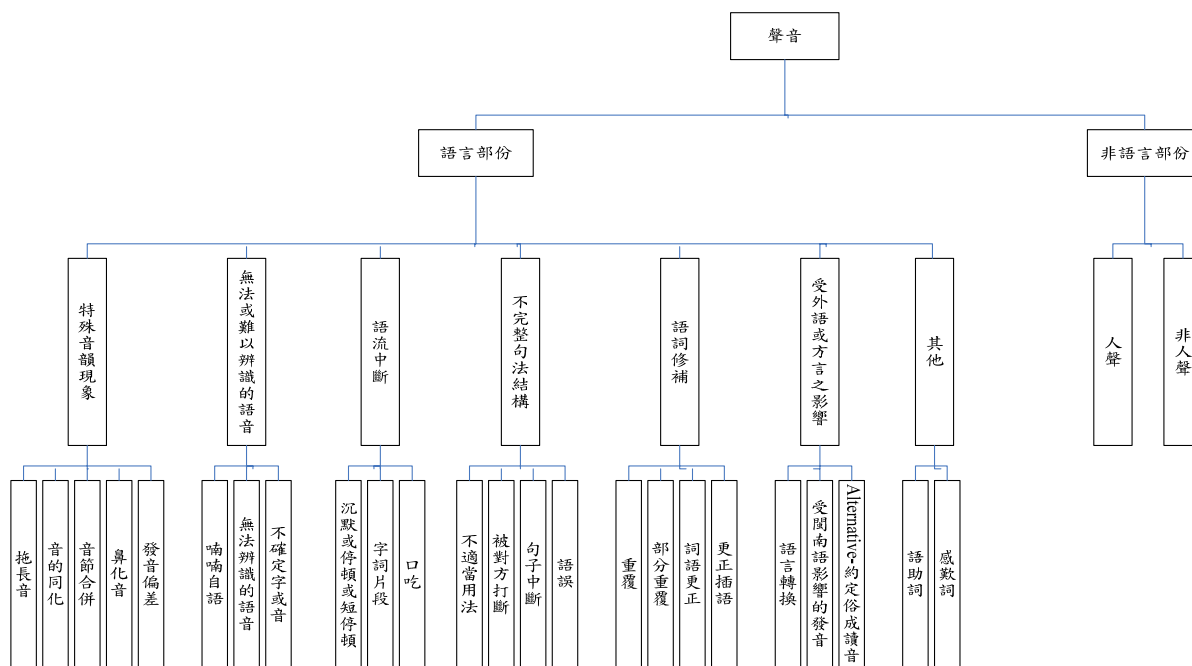


圖 2.1：MCDC 語料庫中，文字以及漢語拼音轉寫之架構圖

2.1.2 語音資料

MCDC 語料庫在收錄時以取樣率 48KHz 之音檔錄製，並且將兩位發音人之語料分別錄於左右聲道，在本研究中將每組對話語料之左右聲道抽取，並轉換為兩個單聲道之音檔，分別為對話中兩位語者之語料，並且將其取樣頻率下降至 16KHz，再利用每一段落相對應之開始及結束時間作切割，經由以上處理後產生 7,085 段音檔將作為本研究所使用之語料。

2.2 自發性語音特性之簡介

自發性語音之語句不如朗讀式語音，有一完整之句法架構及良好的發音，因此常發生在語言學中較詞層次 (word level) 更為上層之行為，例如：語句不完整或詞語修補等現象，以及出現標註系統中無法以基本音標表示之發音，因此為了完整標記自發性語音之特性以利吾人對自發性語音之完整了解，必需對自發性語音作更進一步的標記，在此將介紹 MCDC 語料庫中，曾淑娟博士針對自發性語音之特性，額外標註的重要現象做整理，在本研究中主要關注之現象為「特殊音韻現象」、「無法或難以辨識的語音」、「語流中斷」、「詞語修補」以及「其它」，如下：

➤ 特殊音韻現象

在本研究中關注之特殊音韻現象主要有音的同化(assimilation)、拖長音 (lengthening) 以及音節合併 (syllable contraction)。其中音的同化是由於字的發音受到相鄰發音部位或方法的影響而改變本身發音。拖長音常發生於句尾，或語者猶豫時，刻意托長發音以維持發話權所致。而音節合併主要是因為自發性語音中語速較快，或是語者發話時將較不帶訊息的語句迅速掠過，而發音者為了節省力氣，將相鄰之音節結構改變而產生。其標記範例如下：

音的同化：

文字轉寫：賴先生<b assimilation>呢您</b assimilation>從事什麼工作

漢語標記：lai4 xian1 shen1 [n@n2] nin2 cong2 shi4 shen2 me5 gong1 zuo4

托長音：

文字轉寫：我目前是從<b lengthening>事</b lengthening>外貿

漢語標記：wo3 mu4 qian2 shi4 cong2 shi4 wai4 mao4

音節合併：

文字轉寫：但是相對跟淡水 A<b syllable contraction>那種</b syllable contraction>什麼木
柵那邊比就少很多了

漢語標記：dan4 shi4 xiang1 dui4 gen1 dan4 shui3 A na4 zhong3 shen2 me5 mu4 zha4 ne4
bian1 bi3 jiu4 shao3 hen3 duo1 le5

➤ 無法或難以辨識的語音：

在本研究中關注之無法或難以辨識的語音主要有無法辨識的語音 (unrecognizable speech sound) 以及不確定字或音 (uncertain)。其中無法辨識的語音為標記者確定為人類所發出之語音，但標記員無法辨認何字何意何音。而不確定字或音包括：可猜測出大概的語音內容，但無法百分之百確定；以及無法根據語意猜測出對應字詞，但可清楚記錄出其發音。其標記範例如下：

無法辨識的語音：

文字轉寫：因為<b unrecognizable speech sound>@</b unrecognizable speech sound>太貴了

漢語標記：yin1 wei4 @ tai4 gui4 le5

不確定字或音：

文字轉寫：至少我對我自己的車子<b uncertain>有</b uncertain>有一個瞭解程度 BA

漢語標記：zhi4 shao3 wo3 dui4 wo3 zi4 ji3 [le5] che1 zi5 you3 you3 yi2 ge5 liao3 jie3 cheng2 du4 BA

不確定字或音：

文字轉寫：<b uncertain>[fa1]</b uncertain>因為大概離台北市區比較遠一點所以人不會那麼多

漢語標記：[fa1] yin1 wei4 da4 gai4 li2 tai2 bei3 shi4 [ju1] bi3 shao1 wei2 yuan3 yi4 dian3 suo3 yi3 ren2 bu2 hui4 ne4 me5 duo1

➤ 語流中斷：

在本研究中關注之語流中斷主要有沉默（silence）、停頓（pause）或短停頓（short break），為語者在語流中因話題銜接不上或自身所產生之沉默，其標記範例如下：

停頓：

文字轉寫：然後黃線好像是九百<b pause>@</b pause>然後有的開到一千二

漢語標記：[na2] hou4 huang2 xian4 hao3 xiang4 shi4 jiu3 bai3 @ ran2 hou4 you3 [le5] kai1 [rao4] yi4 qian1 er4

➤ 詞語修補：

在本研究中關注之詞語修補主要有重覆(repetition)、部分重覆(restart)、詞語更正(repair)以及更正插語(editing term)，其基本型態以及範例如圖 2.2 所示。重覆為完整地重覆詞語一次以上；而部分重覆是因他人插話被打斷或因說話者自身的緣故而重覆詞語的片斷，與完整的詞語重覆不同；詞語更正為語者自覺說出的話不適當，立即更正說話內容。更正插語是出現在被更正詞語（reparandum）與更正詞語（correction）之間，或是出現在完整重覆或部分重覆中，兩個重覆詞語之間。在此並定義詞語修補中斷點（IP）為被更正詞語與更正後詞語間之停頓點，或完整重覆或部分重覆中的兩個重覆詞語間之停頓點，。其標記範例如下：

重複：

文字轉寫：要處理可是<b repetition>又有< repetition-IP>又有</b repetition>ZHE GE 情理法理情

漢語標記：yao4 chu3 li3 ke3 shi4 you4 you3 @ you4 you3 ZHE GE qing2 li3 fa3 fa3 li3 qing2

部分重覆：

文字轉寫：真的是稍微動用一下就覺得<b restart>很< restart-IP>很不< restart-IP>很不夠</b restart>這樣子

漢語標記：zhen1 de5 shi4 @ shao1 wei2 dong4 yong4 yi2 xia4 jiu4 jue2 de2 hen3 @ hen3 [bu] hen3 bu2 gou4 zhe4 yang4 zi5

詞語更正：

文字轉寫：當時<b repair>我才反應到<repair-IP>我才意識到</b repair>說其實愛是需要填補的

漢語標記：dang1 shi2 wo3 cai2 fan3 ying1 dao4 wo3 cai2 yi4 shi4 dao4 shuo1 qi2 [zhe5] ai4 shi4 xu1 yao4 tian2 bu3 de5

更正插語：

文字轉寫：外貿 A 是<b repair>進口<repair-IP><b editing term>EN</b editing term>出口</b repair>嗎

漢語標記：wai4 mao4 A @ shi4 jin4 [k-ou] EN chu1 kou3 ma5

基本型態： (被更正詞語) * [更正插語] (更正詞語)

重覆範例： 今天 (王建民) * (王建民) 是先發投手

部分重複範例： 今天 (王) * (王建民) 是先發投手

詞語更正範例： 今天 (陳金鋒) * (王建民) 是先發投手

更正插語範例： 今天 (陳金鋒) * [EN] (王建民) 是先發投手

多層次範例： 今天 (陳金鋒) * [EN] (王) * (王建民) 是先發投手

圖 2.2：詞語修補之基本型態以及範例，其中「*」為 IP 點

➤ 其它：

在其他的部分主要有語助詞 (marker) 以及感嘆詞 (particle)。語助詞為語者本身在語流中慣用的插入語，這些習慣插語有其基本詞彙意義，但在語流中習慣插語已不保有其原有的完整語意，而較具語用功能。而感嘆詞為不具標準語意的感嘆詞，其為語用成份居多。其標記

範例如下：

語助詞：

文字轉寫：室內就是一小間一小間 MA <b marker>NA</b marker>露天就是大家一起 A

漢語標記：shi4 nei4 jiu4 shi4 yi2 xiao3 jian1 yi4 xiao3 jian1 MA @ NA @ lu4 tian1 jiu4 da4
[tai1] yi4 qi3 A @

感嘆詞：

文字轉寫：去什麼富基漁港<b particle>A</b particle>那些

漢語標記：qu4 [ze3] me5 fu4 ji1 yu2 gang3 A nei4 xie1

2.3 語料庫相關統計

在本節將對語料庫中韻律相關之聲學以及語言學資訊做相關的統計，在韻律當中的變化主要利用語調、節奏以及聲音的強度，來顯示語音中之抑揚頓挫，而它們所相對應到之聲學資訊分別為音節基頻軌跡、音節長度以及音節能量¹，其各語者之統計資料如表 2.2 所示，由此表可求得此語料庫之平均發音速度（Articulation Rate, AR）為 5.88（syllables/sec），可得知在自發性語音當中之語速相當快。在此基頻軌跡將以音框為單位求取基頻數值，而音節長度以及能量則以音節為單位求取，參數求取的詳細方式將於 4.2.1 節作介紹。

接著統計韻律相關之語言學資訊，主要有音節聲調、詞長（word length）以及詞類（Part of Speech, POS）。圖 2.3 顯示出了 MCDC 語料庫當中，一到五聲音節之聲調分布圖，此圖中顯示出四聲以及三聲的音節佔大多數；圖 2.4 以及圖 2.5 則顯示出 MCDC 語料庫中，詞長與詞類的分布圖，詞長以一字詞以及二字詞居多，詞類則以名詞及動詞佔大多數。圖 2.6 為 MCDC 語料庫中每段切割之音節數分布圖，由此可看出在此語料庫以短句居多。

¹在本研究中，音節能量採用每一個音節當中韻母的能量最大值，而非音節的能量軌跡或音節能量平均值。

表 2.2：各語者之基頻、音節長度以及能量統計值

speaker no.	gender	pitch (LogHz)		duration (sec)		EngLevel (dB)	
		mean	std	mean	std	mean	std
01L	F	5.24	0.18	0.19	0.09	46.93	6.99
01R	M	4.66	0.21	0.18	0.09	53.68	6.96
02L	F	5.35	0.17	0.18	0.07	55.27	6.69
02R	M	4.92	0.26	0.16	0.08	56.13	7.96
03L	F	5.38	0.16	0.17	0.09	51.89	7.34
03R	F	5.37	0.14	0.17	0.08	53.20	5.41
05L	M	4.84	0.25	0.17	0.10	37.19	7.16
05R	F	5.28	0.26	0.18	0.09	51.91	7.07
09L	F	5.25	0.15	0.18	0.08	51.37	5.39
09R	F	5.33	0.16	0.18	0.09	56.12	6.77
10L	M	4.56	0.21	0.15	0.08	48.01	7.74
10R	M	4.73	0.20	0.17	0.10	52.28	7.36
25L	M	4.54	0.23	0.16	0.08	44.16	6.58
25R	F	5.26	0.22	0.16	0.08	53.46	6.90
26L	F	5.23	0.24	0.16	0.08	51.41	8.10
26R	M	4.94	0.18	0.16	0.09	51.49	6.28
Average		5.05	0.20	0.17	0.08	50.40	6.82

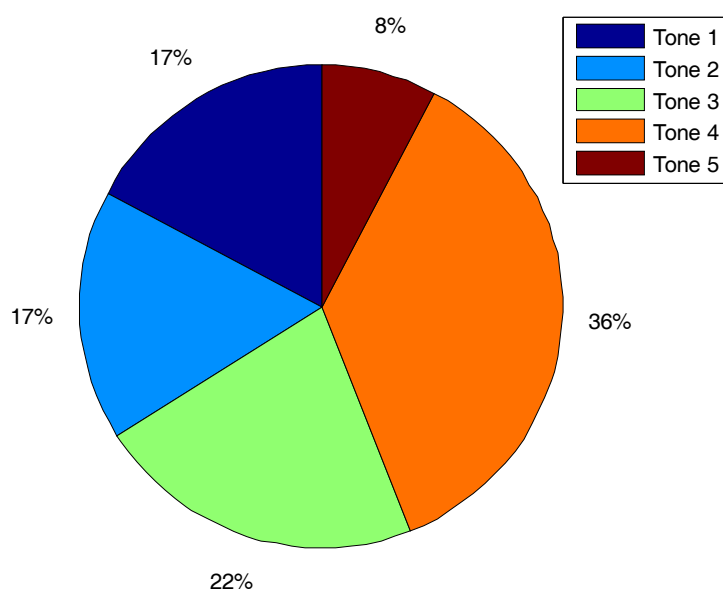


圖 2.3：MCDC 語料庫中，音節聲調分布圖

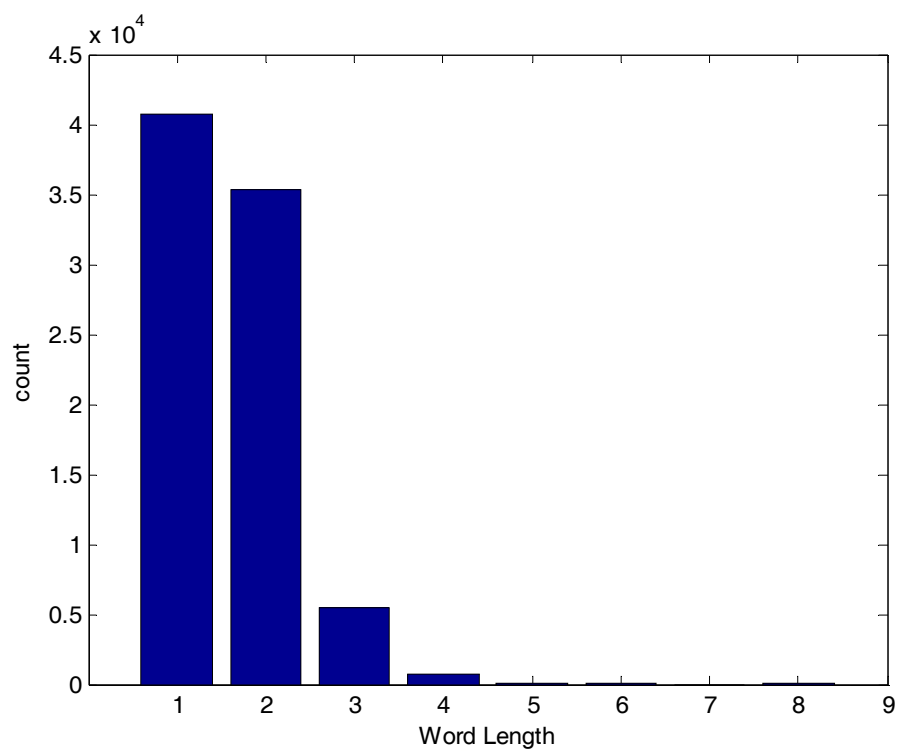


圖 2.4：MCDC 語料庫中，詞長分布圖

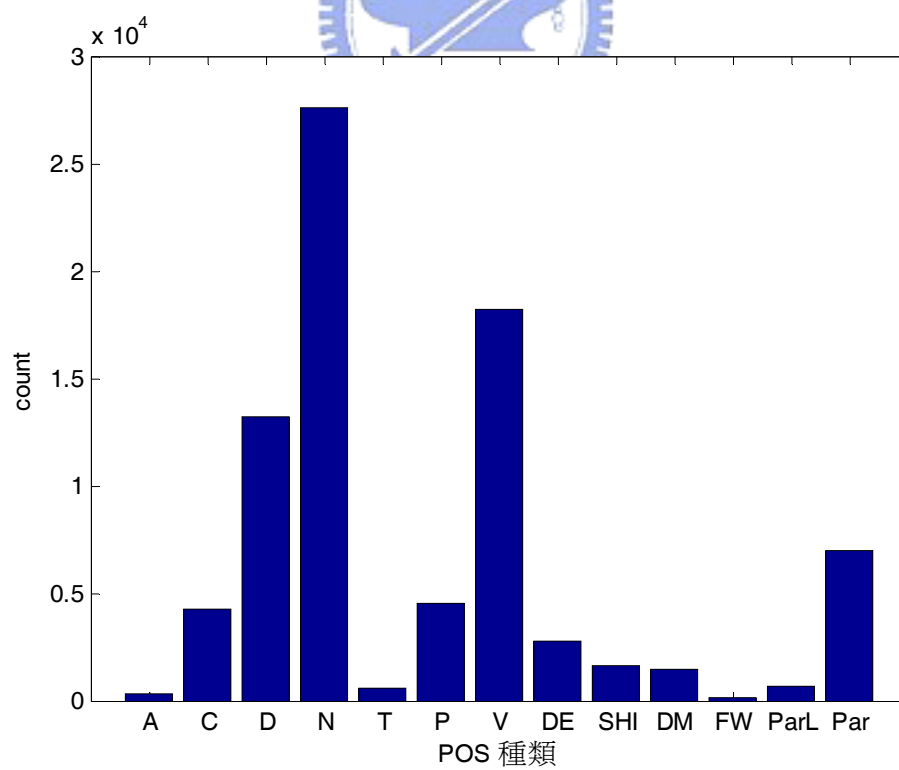


圖 2.5：MCDC 語料庫中，POS 種類分布圖

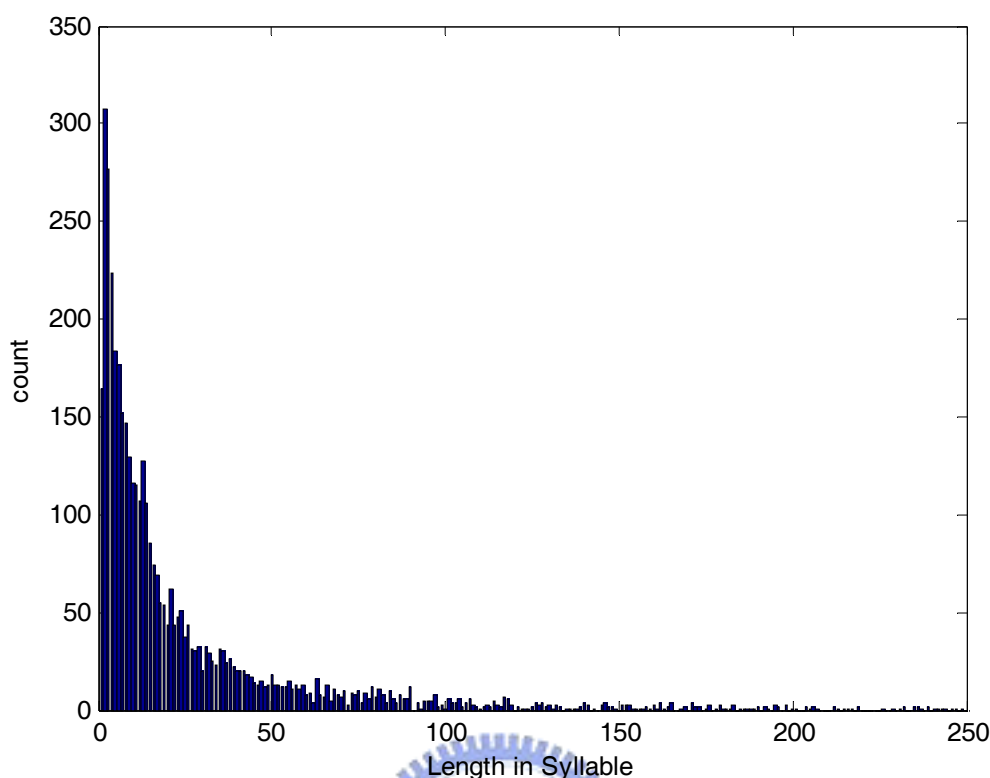


圖 2.6：MCDC 語料庫中，每段切割之音節數分布圖

由於本研究將在最後討論詞語修補之現象，因此先對詞語修補做初步的統計，在此語料庫當中標記「重覆」、「部分重複」以及「詞語更正」之總個數各為 933、333 以及 660，而所有語句中之音節邊界都會是 IP 的候選邊界 (candidate)，所以在此統計所有音節邊界 IP 出現的機率如表 2.3 所示，由此表可知在所有音節邊界出現 IP 之機率為 1.96%，而此語料庫當中尚有許多詞語修補尚未做標注，在此並不列入考慮。

表 2.3：IP 於 MCDC 語料庫中出現之機率

Syllable Juncture Type	IP			non-IP
	Repetition	Repair	Restart	
Count	1379	362	764	128737
Percentage	1.08%	0.28%	0.60%	98.04%

第三章 自發性語音聲學模型

在語音信號處理中建構聲學模型時，隱藏式馬可夫模型（Hidden Markov Model, HMM）是廣為人們所利用的，在本章節中將利用 MCDC 語料庫以及劍橋大學開發之 HTK（HMM Tool Kit）軟體【17】建立一不特定語者（speaker independent）之自發性語音聲學模型，提供本研究求取韻律參數之使用。3.1 節將介紹建構聲學模型之訓練語料，以及評估聲學模型效能之測試語料；3.2 節將介紹本研究中建立聲學模型之流程，3.3 節將分析聲學模型之實驗結果。

3.1 訓練語料及測試語料

本研究建立之系統為一不特定語者之聲學模型，因此訓練語料中包含 16 位語者各自取 9/10 之語音段落所組成，而各語者剩餘之 1/10 語料即為本研究所使用之測試語料，其詳細統計資料如表 3.1 以及 3.2 所示。值得注意的是，在 MCDC 語料當中存在著未含有 411 音節之語句段落，在此會將其排除在訓練語料以及測試語料之外。

表 3.1：訓練語料統計

	411 syllable	Particle	Paralinguistic	Uncertain	Filler	Foreign Word
音節數	104,736	9,688	11,289	3,725	1,743	156
總音節數	131,337					
總段落數	6,166					
音檔長度	8.97 (hours)					

表 3.2：測試語料統計

	411 syllable	Particle	Paralinguistic	Uncertain	Filler	Foreign Word
音節數	14,704	922	1,205	557	263	20
總音節數	17,671					
總段落數	450					
音檔長度	1.09 (hours)					

3.2 聲學模型之建立

3.2.1 特徵參數抽取

在訓練模型前，首先必須獲得足以充分描述語音特性，且參數量較原語音信號小之特徵參數，而語音處理當中，最廣泛為人使用之特徵參數為梅爾頻率倒頻譜係數 (Mel-Frequency Cepstrum Coefficient, MFCC)，本研究也將使用此特徵參數，以 32 毫秒之漢明窗 (Hamming window) 且每位移 10 毫秒為一筆資料，求取 12 維 MFCC 並加上一維能量係數，以及這 13 維係數之一階與二階變量 (delta and delta-delta) 為特徵參數，但單純的能量在參數中較缺乏鑑別性，因此去除能量係數，得到 38 維向量作為本研究語音資料之聲學特徵參數。在本研究也將利用倒頻譜平均值正規化法 (Cepstrum Mean Normalization, CMN) 藉此消除不同語音信號之通道效應。

3.2.2 聲學模型之建立流程

由於自發性語音尚無良好之聲學模型，因此本研究利用朗讀式語料庫 TCC-300 所訓練之聲學模型，協助建構自發性語音之聲學模型。首先利用朗讀式語音之聲學模型，將 MCDC 語料之漢語拼音轉寫和語音信號作強迫對齊，獲得每一個音節切割位置的資訊，並且利用此切割資訊訓練各個音節之聲學模型，其建立流程如圖 3.1 所示。值得注意的是，在自發性語音當中具有許多基本音節之外的其他音節，例如：語助詞、感歎詞、不確定字或音以及副語

言現象，在本研究中這些特殊音節將利用前人所建立之自發性語音聲學模型【18】，同樣經由強迫對齊後獲取特殊音節之切割位置，並另外建立其聲學模型，在此數量過少之其他音節將被歸類至填充模型（filler model）。

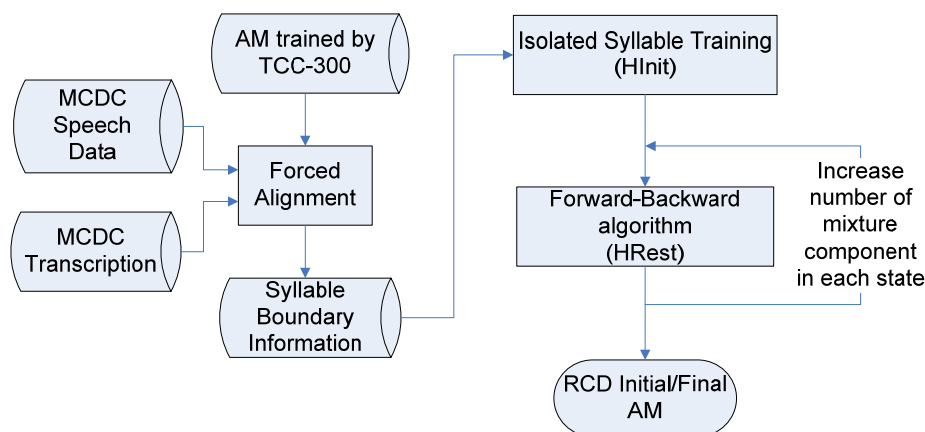


圖 3.1：基準系統之建立流程

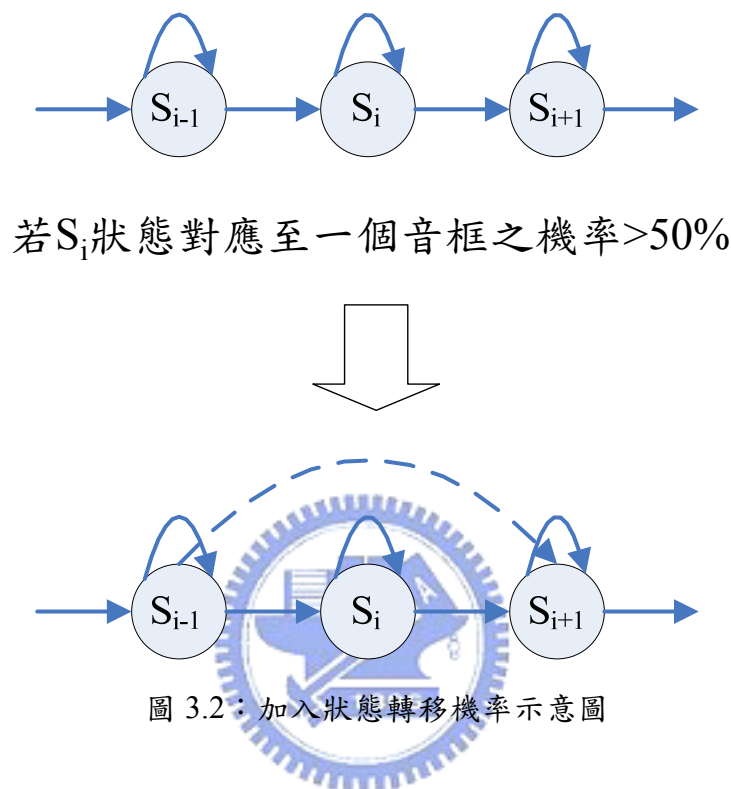
本研究使用之隱藏式馬可夫模型以聲母及韻母為單位，採用音節內右相關聲/韻母模型（Right-ocntext-dependent Initial/Final Model, RCD），每一個聲母之 HMM 模型採用 3 個由左至右（left-to-right）的狀態（state）表示，而韻母之 HMM 模型則採用 5 個狀態來表示，另外填充模型以及短靜音將以 1 個狀態來表示。其中每一個狀態以平均 64 個高斯分布之高斯混合模型（Gaussian Mixture Model, GMM）描述其特徵參數之分布，各模型之 HMM 設定如表 3.3 所示。

表 3.3：HMM 模型之設定

HMM 模型類別	狀態個數	模型數量 ²
Right-context dependent initial	3	97
Final	5	37
Paralinguistic	3	9
Particle	3	23
Uncertain	3	73
Filler	1	1
Foreign word	3	1
Silence	3	1
Short pause	1	1

²在 MCD 語料庫缺少「c_o」、「n_o」以及「s_o」之聲母模型；「eh」、「yai」以及「yo」之韻母模型。

由於自發性語音之語速較快，人類常因為節省發音力氣而省略音節內的某些發音，為了觀察那些發音是常被省略的，在此利用訓練語料之語音及文字，以強迫對齊的方式觀察那些狀態常只對齊至一個音框(=10ms)，若某狀態對齊至一個音框之機率大於 50%，則允許此狀態可被跳過，並重新對模型做訓練之收斂為止，如圖 3.2 所示。



3.3 實驗結果

一般來說聲學模型之效能是由辨認率來評估，本研究之音節辨認率如表 3.4 所示，由表中可以發現，自發性語音中刪除型錯誤數量較多。

表 3.4：音節辨認率

	Correct	Accuracy	Deletion	Substitution	Insertion	Total
Syllable	50.09%	47.57%	2,323	5,889	413	16,452

接著本研究觀察錯誤率高($>40\%$)且錯誤型態大部分為刪除型錯誤之音節，如表 3.5 所示。並發現其左右之音節皆有大量相同之音節成對出現，且成對之另一音節皆有很大的機率發生相同的取代型錯誤，如表 3.6 所示，此即為音節合併現象，其標準型態 (canonical form)

合併之發音 (surface form) 範例如下：

正確答案 (canonical form) 辨認答案 (surface form)		
zhe (這)	jiang (降)	(取代型錯誤)
yang (樣)	NULL	(刪除型錯誤)

另一方面觀察發生合併現象之成對音節，大部分都對應至一個功能詞 (function word)，此現象與曾淑娟博士在【19】中之發現相同。經由辨認結果之觀察，我們可以獲得一些相當重要的音節合併現象，因此在語音辨認時，可以利用簡單的觀察將這些音節合併之訊息加入至辨認辭典當中，而且表 3.6 中之音節合併發生之個數佔辨認中刪除型錯誤的 31.0% (719/2323)，表示若成功解決這些顯而易見的音節合併現象，將會減少一定程度的刪除型錯誤，並獲得更良好之辨識率。

表 3.5：高錯誤率，且刪除型錯誤機率前十名之音節

<i>Rank</i>	<i>Syllable</i>	<i>Deletion/ Syllable Error (%)</i>	<i>Syllable Error Rate (%)</i>
1	yi	70.43	52.16
2	de	65.80	49.80
3	ye	63.13	46.42
4	zhe	59.98	57.71
5	me	51.79	48.15
6	ran	51.09	41.26
7	ru	46.15	61.01
8	neng	45.50	43.85
9	le	45.05	61.18
10	wei	45.01	48.22

表 3.6:經由 ASR 歸納之常見音節合併現象,其中粗體為被刪除之音節, syllable pair merge rate 即為此種 canonical form 的狀況下,出現此種 surface form 之機率

<i>Rank</i>	<i>Canonical form syllable pair</i>	<i>Canonical form Syllable pair count</i>	<i>Surface form</i>	<i>Surface form count</i>	<i>Syllable pair merge rate (%)</i>
1	ne yi (那一)	36	nei	20	55.56
2	wo ye (我也)	115	wei	62	53.91
3	na yi (那一)	77	nai	40	51.95
4	ni ye (妳也)	22	nie	11	50.00
5	suo yi (所以)	479	sui	237	49.48
6	ta ye (他也)	54	tai	25	46.30
7	zhe yang (這樣)	620	jiang	266	42.90
8	zhe yi (這一)	59	zhei	24	40.68
9	bu yao (不要)	84	biao	34	40.48

在此本研究利用這些觀察到的資訊,以動態時軸規劃 (dynamic time programming) 以及反覆的修正語音之轉寫改進聲學模型,如圖 3.3 所示,其中改進聲學模型之步驟如下:

➤ **Step1: 動態時軸規劃**

首先為了找出語音中音節合併發生的位置,在此將 ASR 之辨識結果與標準型態之文字轉寫做動態時軸規劃,一般將辨認結果及標準型態之文字轉寫作動態時軸規劃時,大部分以文字的對齊為主,但在自發性語音中辨識結果之錯誤率非常高,若單純使用文字的對齊作動態時軸規劃會產生對齊不準確的狀況,因此本研究利用標準型態之文字轉寫作強迫對齊後與辨識結果之時間資訊作動態時軸規劃,將辨識結果與正確答案時間上重疊時間最大的音節做對齊,修改後動態時軸規劃之最佳路徑必須滿足下式:

$$D(i, j) = \min_k [D(i-1, k) + d(k, j)] \quad (3-1)$$

其中 $d(i, j)$ 表示正確答案中第 i 個音節與辨識結果中第 j 個音節之成本函數 (cost function) 如下:

$$d(i, j) = -\max\{0, \min(t_{EOA}(i), t_{EOR}(j)) - \max(t_{SOA}(i), t_{SOR}(j))\} \quad (3-2)$$

且 $t_{SOA}(i)$ 及 $t_{EOA}(i)$ 代表正確答案中第 i 個音節之起始時間及結束時間; $t_{SOR}(j)$ 及 $t_{EOR}(j)$ 代表辨識結果中第 j 個音節之起始時間及結束時間。

➤ **Step2：修正語音之轉寫**

利用表 3.6 作查表對測試及訓練語料做重新轉寫，若正確答案與辨識答案之對齊出現表 3.6 之音節合併現象，則作重新轉寫的動作。

➤ **Step3：重新訓練音節模型**

利用新的文字轉寫對訓練語料做模型之重新訓練。

➤ **Step4：重覆 Step1 到 Step4 至找尋到最佳音節辨識率之聲學模型**

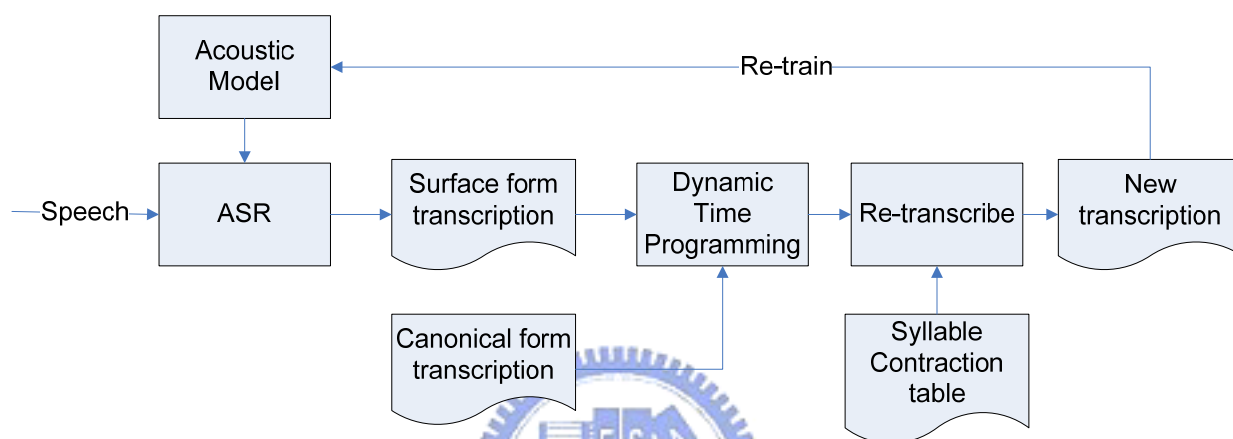


圖 3.3：加入音節合併資訊之聲學模型示意圖

在經過動態時軸規劃以及反覆的修正語音之轉寫並重新訓練聲學模型後，得到之最佳辨識率如圖 3.4 所示，辨識率提升 1.32%，刪除型錯誤也明顯的下降，但值得注意的是，由於在本研究之目標為獲得一個較符合實際語音訊號之聲學模型，但標準型態之文字轉寫並不符合實際語音訊號，因此正確答案於本研究中也作了重新轉寫的動作，可以發現聲學模型改進的原因大部分是因為修改了正確答案，而使刪除型錯誤增加以及插入型錯誤減少，進而增加辨識率，表示音節合併的資訊在聲學模型上並無有效的改進，必須於更上層之辨認字典的詞中加入音節合併發音變異之資訊，藉以增加辨識率，但本論文在聲學模型上之研究以獲得音節切割位置資訊為主，因此不再深入討論。

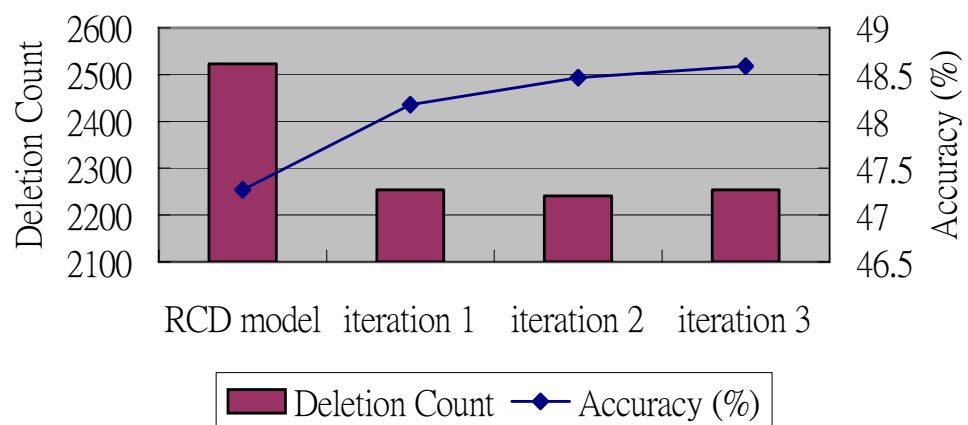


圖 3.4：加入音節合併資訊後之辨識率



第四章 自發性語音韻律模型

本論文基於【7】中所提出之方法作修改，完成適合自發性語音之韻律模型及標記，本研究利用了語音信號中之聲學參數和文字上的語言參數，以非監督式（unsupervised）的方法訓練韻律模型並預估每段語音之韻律標記。4.1 節將介紹中文語音韻律階層式架構之概念及韻律模型之設計；4.2 節將會介紹在本研究中所使用的韻律參數之抽取方法，以及為了完成非特定語者之韻律模型而使用之參數正規化方法；4.3 節將介紹模型參數訓練及更新之演算法。

4.1 韻律模型之設計

本節中將介紹中文語音韻律階層式架構之概念，並說明在本研究所使用之中文語音韻律階層式架構，以及韻律模型之設計，在此定義四個子模型來描述語音信號中聲學參數以及文字上語言參數與韻律標記之關係。

4.1.1 中文語音韻律階層式架構

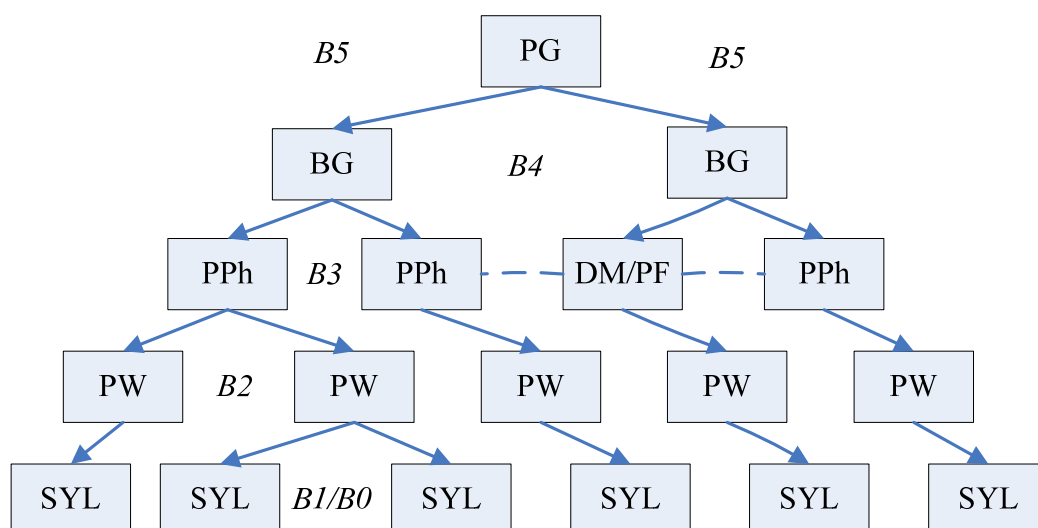


圖 4.1：中文語音韻律之階層式架構概念

在過去的研究顯示，中文的韻律結構是具有階層性的架構（hierarchy structure），傳統上

韻律階層式架構由底層至上層，主要由音節 (Syllable, SYL)、韻律詞 (Prosody Word, PW)、韻律短句 (Prosody Phrase, PPh) 以及句調 (intonation phrase) 所構成，此外鄭秋豫博士【20】提出將連續的 PPh 組合成一個呼吸群 (Breathe Group, BG) 來代表大範圍且具有基頻及音節長度高度變化之語句，藉此表示韻律更上層之貢獻，同樣地定義由連續的 BG 所組成的韻律群 (Prosody Group, PG)，值得注意的是鄭秋豫博士在流利語音的韻律架構當中，定義 PPh 間存在著某些可插入的篇章提語 (Discourse Marker, DM) 或韻律填充 (Prosody Filler, PF)，以連接鄰近的 PPh，如圖 4.1 所示。

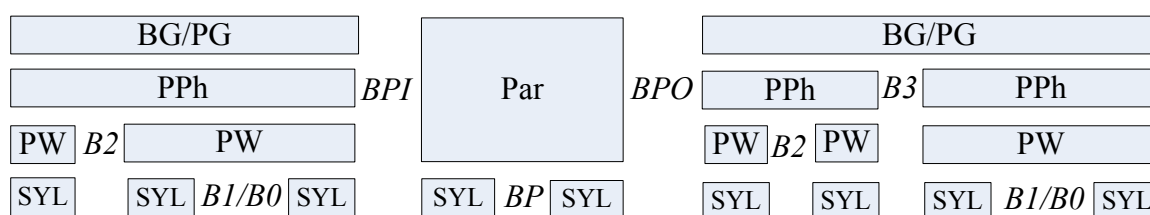


圖 4.2：本研究使用之中文自發性語音韻律階層式架構

在本研究中，基於鄭秋豫博士所提出的韻律階層式架構為基礎作修改並設計韻律模型，如圖 4.2 所示，值得注意的是，為了避免自發性語音中出現的許多特殊現象，影響其他正常語流之統計特性，進而使模型參數失去其特性，在此我們重新定義了一個特殊韻律現象 (Particular Prosody Phenomena, Par) 的單元，藉此隔離正常語流之語句，特殊韻律現象中包含：韻律特性和基本音節差距較大之語助詞或感歎詞³、無法或難以辨識的語音、受相鄰音節同化之音節以及發生嚴重拉長之音節⁴，這些嚴重拉長的音節數量雖小但長度較其他基本音節大，會嚴重影響模型之統計特性，如圖 4.3 所示，因此本研究將所有音節分為對應特殊韻律單元之特殊音節 (particular syllable) 以及對應正常語流之基本音節 (base syllable)。此外在本研究所使用的四層韻律結構中，主要由十種停頓標記 (break type) 分別為 {B0, B1, B2-1, B2-2, B2-3, B3, B4, BPI, BP, BPO} 來區分韻律結構中每一層的韻律單元。首先，B0 以及 B1 定義為 SYL 的邊界，分別表示音節邊界相鄰兩音節是緊密連接 (tightly coupling) 或普通連接 (normal coupling)；B2-1、B2-2 以及 B2-3 定義為 PW 之邊界，分別表示相鄰兩音節具有

³本研究將第二章所定義之感歎詞或語助詞中，「ZHE GE」、「NA GE」、「NE GE」、「NEI GE」以及「SHEN ME」視為兩個音節，其餘皆視為一個音節。

⁴嚴重音節拉長決定方式為一經驗法，本研究定義音節長度前 99% 的資料為嚴重音節拉長。

明顯的基頻跳躍、短停頓以及前一音節發生音節拉長 (pre-boundary lengthening)；*B3* 以及 *B4* 定義為 *PG* 和 *BG* 之邊界，具有長停頓和大的基頻跳躍；*BPI*、*BP* 以及 *BPO* 分別代表後一音節為特殊音節、相鄰之兩音節皆為特殊音節以及前一音節為特殊音節，藉此將特殊韻律現象與正常語流之語句隔離。

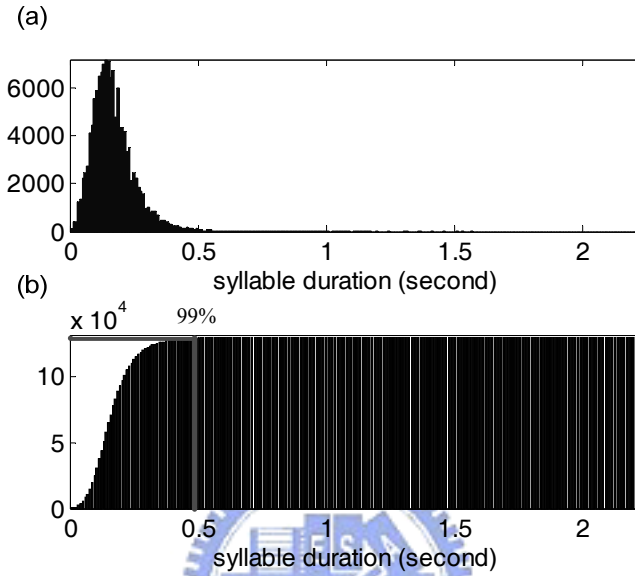


圖 4.3：音節長度之 (a)數值分布及 (b)累積分布

4.1.2 模型設計

本研究利用語音信號上的聲學參數 **A**，以及文字上語言學的參數 **L**，以模型為基礎 (model-based) 估計此語句中最有可能的韻律標記序列 **T***，因此可將其看作一個數學估計的問題，其數學式如下：

$$\mathbf{T}^* = \arg\max_{\mathbf{T}} P(\mathbf{T}|\mathbf{A}, \mathbf{L}) = \arg\max_{\mathbf{T}} P(\mathbf{T}, \mathbf{A}|\mathbf{L}) \quad (4-1)$$

在此定義兩種韻律標記，第一種為 4.1.1 節中所定義之音節停頓標記序列 **B**；第二種則是韻律狀態 (prosody state) 序列 **PS**，它是經由扣除音節及其相鄰音節對韻律之影響並量化後所得到，以描述韻律上層之變化狀況及其對韻律參數之貢獻值，在本研究中，將所有基本音節量化為 16 個韻律狀態，並且將特殊音節另外分出 4 個韻律狀態量化之。

表 4.1：韻律標記、聲學參數以及語言參數之數學符號

B: break type = { <i>B0</i> , <i>B1</i> , <i>B2-1</i> , <i>B2-2</i> , <i>B2-3</i> , <i>B3</i> , <i>B4</i> , <i>BPI</i> , <i>BP</i> , <i>BPO</i> }		
T: prosodic tag	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
		df: normalized duration lengthening factor
L: linguistic feature	l: reduced linguistic feature set	
	t: syllable tone sequence	
	s: base-syllable type	
	f: final type	

接著定義本研究中使用之韻律聲學參數 **A**，以及文字上的語言學參數 **L**，如表 4.1 所示。本研究考慮的韻律聲學參數包含兩大類，第一類是與韻律狀態有緊密關係的音節韻律參數 **X** (syllable prosodic feature) 主要有：音節基頻軌跡 **sp** (syllable pitch contour)、音節長度 **sd** (syllable duration) 以及音節能量 **se** (syllable energy level)；第二類是與停頓標記有緊密相關的特徵參數，又細分為兩類分別是音節間韻律參數 **Y** (inter-syllable prosodic feature) 以及相鄰兩音節差異之韻律參數 **Z** (differential prosodic feature)，音節間韻律參數有：音節間停頓長度 **pd** (pause duration) 以及音節間能量低點 **ed** (energy-dip level)；相鄰兩音節差異之韻律參數有：相鄰兩音節之正規化基頻跳躍值 **pj** (normalized pitch jump) 以及相鄰兩音節之正規化音節延長因子 **dl** (normalized duration lengthening factor)，這些參數將在後面的章節做更詳細的定義。另外，文字上的語言學特徵參數 **L**，主要包含語言學中音節以及詞層次上的參數。音節層次上的參數主要包含了音節聲調序列 **t** (tone sequence)、基本音節型態 **s** (base syllable type) 或韻母型態 **f** (final type)；其它語言參數 **I**，主要包含音節邊界種類⁵ (syllable juncture type)、詞長以及詞類。

⁵ 音節邊界種類主要分為：詞內音節邊界 (intra-word syllable juncture) 以及詞間音節邊界 (inter-word syllable juncture)

經由以上的定義，我們可以將 4-1 式改寫為：

$$P(T,A|L)=P(A|T,L)P(T|L)=P(X,Y,Z|B,PS,L)P(B,PS|L) \quad (4-2)$$

其中 $P(X,Y,Z|B,PS,L)$ 為廣義韻律參數模型 (general prosodic feature model)，其物理意義為下層所得到的韻律聲學參數 X 、 Y 、 Z ，是由上層的韻律標記 B 、 PS 以及語言參數 L 所控制。而 $P(B,PS|L)$ 為廣義韻律語言模型 (general prosody-syntax model)，它主要在描述韻律標記 B 、 PS 和語言參數 L 之間的關係。

由於定義之停頓標記 B 已帶有相鄰兩音節間之韻律資訊，因此在已知停頓標記的狀況下，可以假設音節韻律參數 X 與音節間韻律參數 Y 及相鄰兩音節差異之韻律參數 Z 互相獨立，因此可將廣義韻律聲學模型 $P(X,Y,Z|B,PS,L)$ 一分為二，其數學式如下：

$$P(X,Y,Z|B,PS,L)=P(X|B,PS,L)P(Y,Z|B,PS,L) \quad (4-3)$$

其中 $P(X|B,PS,L)$ 為音節韻律模型 (syllable prosodic model)，其物理意義為音節中的基頻軌跡、音節長度及音節能量，是由上層的韻律標記 B 、 PS 以及語言參數 L 所控制，其中語言參數又以音節的聲調序列 t 之影響最為嚴重。而 $P(Y,Z|B,PS,L)$ 為停頓標記聲學模型 (break-acoustic model)，它描述不同韻律標記 B 、 PS 以及語言參數 L 的狀況之下，音節間韻律參數 Y 及相鄰兩音節差異之韻律參數 Z 分布的情況。同樣地，經由假設上層的韻律標記 B 、 PS 與音節聲調 t 互相獨立，我們也可將廣義韻律語言模型 $P(B,PS|L)$ 一分為二，數學式如下：

$$P(B,PS|L) \approx P(B,PS|I) = P(PS|B,I)P(B|I) \approx P(PS|B)P(B|I) \quad (4-4)$$

其中 $P(PS|B)$ 為韻律狀態轉移模型 (prosodic state model)，描述在已知停頓標記的狀況之下，韻律狀態轉移之機率。 $P(B|I)$ 則為停頓標記語言模型 (break-syntax model)，主要描述停頓標記 B 和語言參數 I 之間的關係。因此本研究中設計了四個子模型分別為：音節韻律模型、停頓標記聲學模型、韻律狀態轉移模型以及停頓標記語言模型，來描述韻律狀態、停頓標記與聲學和語言學參數之間的關係。

本研究之音節韻律模型是假設音節基頻軌跡、音節長度以及音節能量可拆解成各個影響因子 (affecting factor) 之貢獻，這些影響因子包含：音節之韻律狀態 p_n 、包含音節左右邊界停頓標記 B_n 、 B_{n-1} 及鄰近音節聲調 t_{n+1} 、 t_{n-1} 影響之音節聲調 t_n (i.e. 音節連音現象之影響) 以及音節基本型態 s_n 或音節韻母型態 f_n ，不同的音節韻律模型將視影響程度，對應至不同影響因子之組合，將在之後做更詳細之介紹。因此可將音節韻律模型拆解成三個模型分別為音節基頻軌跡模型、音節長度模型以及音節能量模型，其數學式如下：

$$p(\mathbf{X}|\mathbf{B}, \mathbf{P}, \mathbf{S}, \mathbf{L}) \approx p(\mathbf{sp}|\mathbf{B}, \mathbf{p}, \mathbf{t}) p(\mathbf{sd}|\mathbf{B}, \mathbf{q}, \mathbf{t}, \mathbf{s}) p(\mathbf{se}|\mathbf{B}, \mathbf{r}, \mathbf{t}, \mathbf{f})$$

$$\approx \prod_{n=1}^N p(\mathbf{sp}_n | B_{n-1}^n, p_n, t_{n-1}^{n+1}) \prod_{n=1}^N p(sd_n | B_{n-1}^n, q_n, t_{n-1}^{n+1}, s_n) \prod_{n=1}^N p(se_n | B_{n-1}^n, r_n, t_{n-1}^{n+1}, f_n) \quad (4-5)$$

其中

$$\mathbf{sp}_n = \begin{cases} \mathbf{sp}_n^r + \beta_{t_{n-1}^{n+1}, B_{n-1}^n} + \beta_{p_n} + \mu & , \text{ if } nth \text{ syllable is base syllable} \\ \mathbf{sp}_n^r + \beta_{pr_n} + \beta_{p_n'} + \mu & , \text{ if } nth \text{ syllable is particular syllable} \end{cases} \quad \text{for } 1 \leq n \leq N \quad (4-6)$$

為第 n 個音節之音節基頻軌跡，是由四維正交化係數以向量的方式表示，將在後面的章節有進一步的介紹； \mathbf{sp}_n^r 為 \mathbf{sp}_n 正規化 (normalization) 後之基頻殘存值 (residual)； β_x 則為某一影響因子 x 之影響型態 (Affecting Pattern, AP)； μ 為所有 AP 之總體平均值 (global mean)，在此為了使特殊音節及正常音節之 AP 可互相比較，因此賦予它們擁有相同的總體平均值。同樣地，第 n 個音節之音節長度及音節能量可表示如下：

$$sd_n = \begin{cases} sd_n^r + \gamma_{t_{n-1}^{n+1}, B_{n-1}^n} + \gamma_{q_n} + \gamma_{s_n} + \mu_d & , \text{ if } nth \text{ syllable is base syllable} \\ sd_n^r + \gamma_{pr_n} + \gamma_{q_n'} + \mu_d & , \text{ if } nth \text{ syllable is particular syllable} \end{cases} \quad \text{for } 1 \leq n \leq N \quad (4-7)$$

$$se_n = \begin{cases} se_n^r + \alpha_{t_{n-1}^{n+1}, B_{n-1}^n} + \alpha_{r_n} + \alpha_{f_n} + \mu_e & , \text{ if } nth \text{ syllable is base syllable} \\ se_n^r + \alpha_{pr_n} + \alpha_{r_n'} + \mu_e & , \text{ if } nth \text{ syllable is particular syllable} \end{cases} \quad \text{for } 1 \leq n \leq N \quad (4-8)$$

如同以上數學式所描述，本研究將特殊音節另行賦予其特殊音節型態之影響因子 pr_n 以及特殊音節之韻律狀態 p_n' 。接著經由假設正規化後之殘存值為一平均值為零之高斯分佈，可將音節基頻軌跡模型、音節長度模型以及音節能量模型改寫為：

$$P(\mathbf{sp}_n | p_n, B_{n-1}^n, t_{n-1}^{n+1}) = \begin{cases} N(\mathbf{sp}_n; \boldsymbol{\beta}_{t_{n-1}^{n+1}, B_{n-1}^n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\mu}, \mathbf{R}) \\ \quad , \text{ if } n\text{th syllable is base syllable} \\ N(\mathbf{sp}_n; \boldsymbol{\beta}_{pr_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\mu}, \mathbf{R}') \\ \quad , \text{ if } n\text{th syllable is particular syllable} \end{cases} \quad \text{for } 1 \leq n \leq N \quad (4-9)$$

$$P(sd_n | q_n, B_{n-1}^n, t_{n-1}^{n+1}, s_n) = \begin{cases} N(sd_n; \gamma_{t_{n-1}^{n+1}, B_{n-1}^n} + \gamma_{q_n} + \gamma_{s_n} + \mu_d, R_d) \\ \quad , \text{ if } n\text{th syllable is base syllable} \\ N(sd_n; \gamma_{pr_n} + \gamma_{q_n} + \mu_d, R'_d) \\ \quad , \text{ if } n\text{th syllable is particular syllable} \end{cases} \quad \text{for } 1 \leq n \leq N \quad (4-10)$$

$$P(se_n | r_n, B_{n-1}^n, t_{n-1}^{n+1}, f_n) = \begin{cases} N(se_n; \alpha_{t_{n-1}^{n+1}, B_{n-1}^n} + \alpha_{r_n} + \gamma_{f_n} + \mu_e, R_e) \\ \quad , \text{ if } n\text{th syllable is base syllable} \\ N(se_n; \alpha_{pr_n} + \alpha_{r_n} + \mu_e, R'_e) \\ \quad , \text{ if } n\text{th syllable is particular syllable} \end{cases} \quad \text{for } 1 \leq n \leq N \quad (4-11)$$

值得注意的是，本研究將利用決策樹（decision tree）以資料驅動（data-driven）的方式，自動分類 $\boldsymbol{\beta}_{t_{n-1}^{n+1}, B_{n-1}^n}$ 、 $\gamma_{t_{n-1}^{n+1}, B_{n-1}^n}$ 以及 $\alpha_{t_{n-1}^{n+1}, B_{n-1}^n}$ 之 AP，藉此同時描述音節聲調及前後音節連音現象（coarticulation）對韻律參數之影響，將在後面的章節做更進一步的介紹。

接著經由假設音節間韻律參數 \mathbf{Y} 及相鄰兩音節差異之韻律參數 \mathbf{Z} 與韻律狀態 \mathbf{PS} 及聲調序列 \mathbf{t} 之間互相獨立，可將停頓標記聲學模型之數學式改寫如下：

$$P(\mathbf{Y}, \mathbf{Z} | \mathbf{B}, \mathbf{PS}, \mathbf{L}) \approx P(\mathbf{Y}, \mathbf{Z} | \mathbf{B}, \mathbf{I}) \approx \prod_{n=1}^{N-1} P(pd_n, ed_n, pj_n, dl_n | B_n, \mathbf{I}_n) \quad (4-12)$$

為了數學上容易處理，我們將 $P(pd_n, ed_n, pj_n, dl_n | B_n, \mathbf{I}_n)$ 簡化成停頓長度之伽瑪分布與其他參數之高斯分布相乘，其數學式如下：

$$P(pd_n, ed_n, pj_n, dl_n | B_n, \mathbf{I}_n) = g(pd_n; \alpha_{B_n, \mathbf{I}_n}, \beta_{B_n, \mathbf{I}_n}) N(ed_n; \mu_{B_n, \mathbf{I}_n}, \sigma_{B_n, \mathbf{I}_n}^2) \\ \cdot N(pj_n; \mu_{B_n, \mathbf{I}_n}, \sigma_{B_n, \mathbf{I}_n}^2) N(dl_n; \mu_{B_n, \mathbf{I}_n}, \sigma_{B_n, \mathbf{I}_n}^2) \quad (4-13)$$

在每種停頓標記狀況下，各個參數之機率分布，將以最大概似度增益為分裂準則（splitting criterion of maximum likelihood gain）之決策樹實現，其問題集（question set）將由詢問語言

參數 \mathbf{l} 之問題所產生。

此外韻律狀態轉移模型將利用馬可夫模型（Markov Model）來實現，其數學式如下：

$$P(\mathbf{p}|\mathbf{B}) \approx P(p_1) \left[\prod_{n=2}^N P(p_n | p_{n-1}, B_{n-1}) \right] \quad (4-14)$$

其中 $P(p_1)$ 為各語句中第一個音節韻律狀態之機率； $P(p_n | p_{n-1}, B_{n-1})$ 為已知音節前邊界之停頓標記 B_{n-1} 之下，前一音節韻律狀態 p_{n-1} 轉移到現在音節韻律狀態 p_n 之機率。

最後我們簡化停頓標記語言模型為：

$$P(\mathbf{B}|\mathbf{l}) = \prod_{n=1}^{N-1} P(B_n | \mathbf{l}_n) \quad (4-15)$$

並且以最大概似度增益為分裂準則之決策樹來實現它，每一節點中將產生每一種停頓標記之機率，其問題集將由詢問語言參數 \mathbf{l} 之問題所產生。

4.2 韻律模型參數之萃取

韻律的表現是受到上層的語言學資訊所影響，除此之外下層的聲學資訊也可觀察到相對應的韻律表現。如同 4.1 節所述，本論文考慮的韻律聲學參數主要包含了音節韻律參數、音節間韻律參數以及相鄰兩音節差異之韻律參數，音節韻律參數有：音節內之基頻軌跡、長度以及能量；音節間韻律參數有：音節間停頓長度以及能量低點；相鄰兩音節差異之韻律參數有：相鄰兩音節之正規化基頻跳躍值以及正規化音節延長因子，如圖 4.4 所示。要求得這些韻律聲學參數首先必須得到基頻軌跡、音節長度、停頓長度以及能量軌跡的聲學參數。在此小節當中將會介紹如何抽取這些聲學參數以及語言學參數，以估計韻律模型。4.2.1 節會介紹如何由語音訊號求取聲學參數；4.2.2 節會介紹如何利用聲學參數萃取本研究中所使用之音節韻律參數、音節間韻律參數以及相鄰兩音節差異韻律參數；4.2.3 節將會介紹如何得到語言學參數。

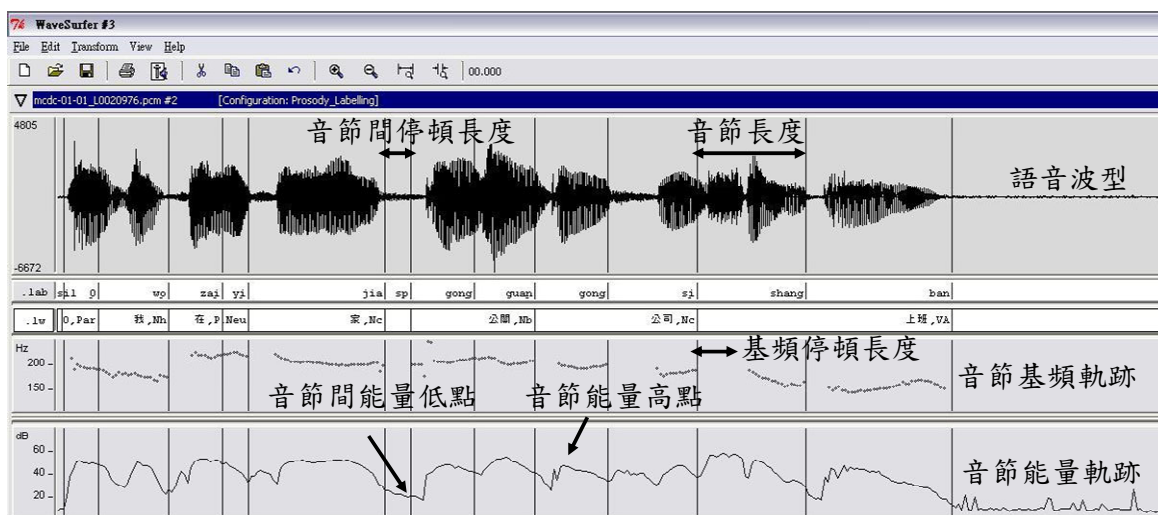


圖 4.4：音節韻律參數及音節間韻律參數

4.2.1 聲學參數抽取

要得到音節或音節間之韻律參數，我們首先必須從語音訊號中得到一些未經加工的聲學參數資料 (raw data)，主要有音節基頻軌跡、音節長度、停頓長度以及能量軌跡。

音節基頻軌跡在韻律當中扮演非常重要的角色，我們先以軟體 Wavesurfer 當中的 ESPS (Entropic Signal Processing System) 演算法【21】求取每一個音框內之基頻數值後，發現基頻數值會發生嚴重的倍頻(double pitch)、半頻(half pitch)以及語音開頭和結尾時不穩(on-set & off-set)等現象，如圖 4.5 所示。這些現象會使訓練之模型效能下降，因此本論文採用【22】中的基頻修正方法，經統計後發現未經加工的基頻數值可由一個簡單的高斯混合模型描述，如圖 4.6 (a)所示，因此本論文使用一個高斯混合模型，針對每一個語者之基頻數值偵測並修改倍頻及半頻的基頻數值，修正後之基頻軌跡及基頻數值分布，如圖 4.5 及圖 4.6 (b)所示。值得注意的是，經過倍頻及半頻修正後之基頻數值，本論文都將取對數基頻軌跡(Log-F0 contour)作處理。

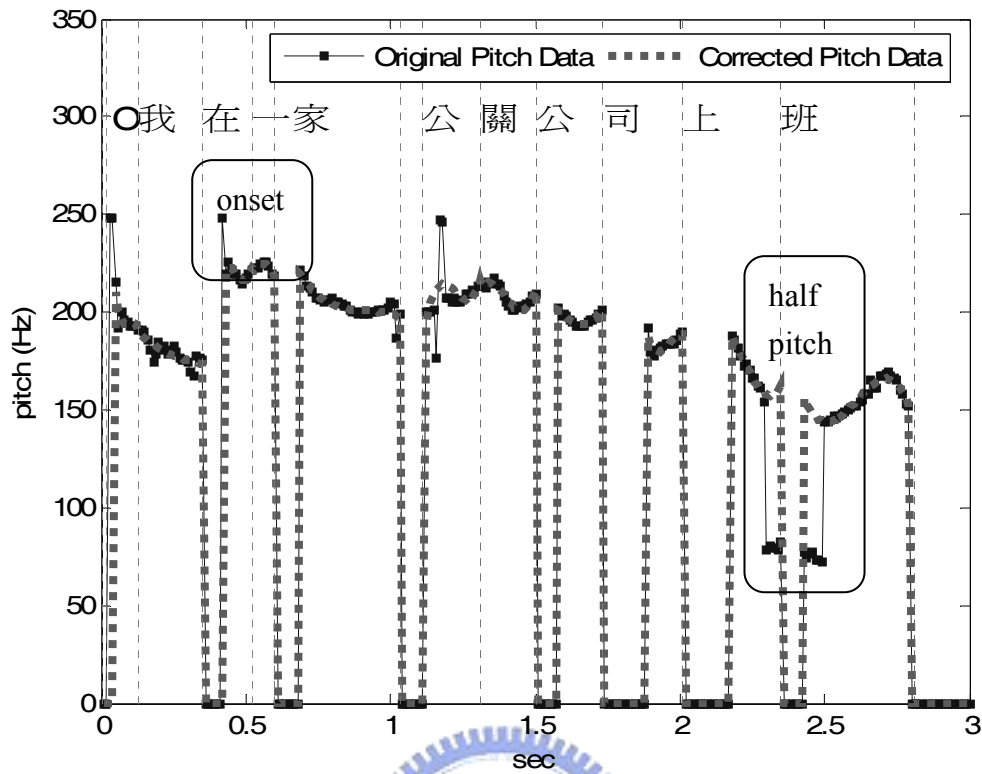


圖 4.5：語音開頭時基頻不穩與半頻現象及其修正後基頻軌跡

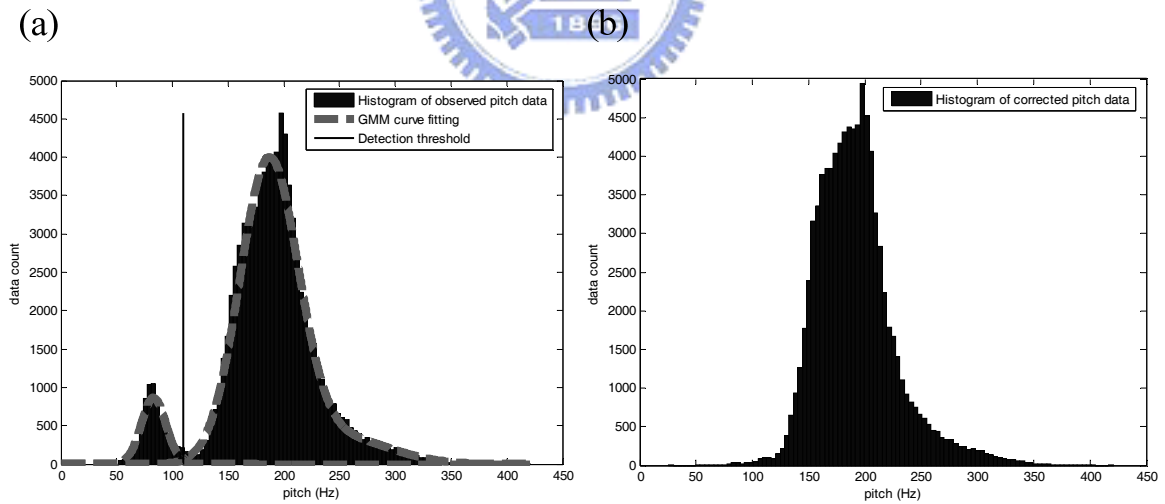


圖 4.6：語者編號 01L 之 (a)基頻資料及其半頻偵測模型和 (b)修正後基頻資料

接著，由於本研究最終目的為建立一個非特定語者之韻律模型，因此為了消除每位語者因生理上的不同，而產生先天上基頻的差異，使參數缺乏鑑別性，在此我們會對每位語者之基頻數值進行高斯正規化 (Gaussian normalization)【23】。經過高斯正規化後第 s 位語者的第 i 個音框之對數基頻數值 f_i 如下：

$$\hat{f}_i^s = \left(\frac{f_i^s - \mu_s}{\sigma_s} \right) \cdot \sigma_{average} + \mu_{global} \quad (4-16)$$

其中 μ_s 及 σ_s 分別為第 s 位語者基頻數值之平均值及標準差，以及 μ_{global} 及 $\sigma_{average}$ 分別為所有語者之總體平均值及平均標準差，其數學式分別如下：

$$\mu_s = \frac{\sum_{i=1}^{I(s)} f_i^s}{I(s)} \quad (4-17)$$

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^{I(s)} (f_i^s - \mu_s)^2}{I(s)}} \quad (4-18)$$

$$\mu_{global} = \frac{\sum_s \sum_{i=1}^{I(s)} f_i^s}{\sum_s I(s)} \quad (4-19)$$

$$\sigma_{average} = \sqrt{\frac{\sum_s \sum_{i=1}^{I(s)} (f_i^s - \mu_s)^2}{\sum_s I(s)}} \quad (4-20)$$



在此 $I(s)$ 為第 s 位語者基頻之總音框數。值得注意的是，在求取所有平均值及標準差時，由於特殊音節之特性和基本音節相差甚多，會影響韻律參數之統計值，因此本研究將不蒐集這些特殊音節計算平均值及變異數。

接著每一個音節所對應到之音節長度以及音節間停頓長度，將使用第三章獲得之聲學模型並利用 HTK 軟體，以強迫對齊的方法對每一個音檔作音節之切割，得到每一個音節的音節長度以及停頓長度。

最後經由 Wavesurfer 軟體，我們也可以求得語音中每一個音框的能量數值，而得到能量軌跡，以供往後求取韻律相關的能量參數。

4.2.2 韻律相關聲學參數之萃取

在本小節中將分別於 4.2.2.1、4.2.2.2 以及 4.2.2.3 介紹如何由聲學參數萃取音節韻律參數、音節間韻律參數以及相鄰兩音節差異之韻律參數。

4.2.2.1 音節韻律參數之萃取

在此將利用先前得到的聲學參數，求取音節韻律參數，主要有音節基頻軌跡、音節長度以及音節能量。首先先針對音節基頻軌跡的資料作處理，為了將基頻軌跡更加平滑化，以衰弱少數音框基頻數值求取不佳的影響，以及減少模型訓練時表示基頻資訊的參數量，在此使用【24】所提出，利用一個片段線性（piecewise linear）方法，對音節內之基頻軌跡做正交化展開（orthogonal expansion），投影到四個雷建德多項式（Legendre polynomial）基底，求得四維正交參數，達到將基頻軌跡平滑化以及參數化的動作。經過正交化展開後，第 n 個音節的四維正交化參數如下：

$$a_{j,n} = \frac{1}{N+1} \sum_{i=0}^N \hat{f}_{i,n} \cdot \phi_j\left(\frac{i}{N}\right), \quad \text{for } j = 0, 1, 2, 3 \quad (4-21)$$

其中 $\hat{f}_{i,n}$ 為第 n 個音節的第 i 個音框，經過修正後的基頻數值，且 $0 \leq i \leq N$ ； $N+1$ 為第 n 個音節基頻軌跡之總音框數；四個雷建德多項式基底如下：

$$\phi_0\left(\frac{i}{N}\right) = 1 \quad (4-22)$$

$$\phi_1\left(\frac{i}{N}\right) = \left[\frac{12 \cdot N}{(N+2)} \right]^{1/2} \cdot \left[\left(\frac{i}{N} \right) - \frac{1}{2} \right] \quad (4-23)$$

$$\phi_2\left(\frac{i}{N}\right) = \left[\frac{180 \cdot N^3}{(N-1)(N+2)(N+3)} \right]^{1/2} \cdot \left[\left(\frac{i}{N} \right)^2 - \left(\frac{i}{N} \right) + \frac{(N-1)}{6 \cdot N} \right] \quad (4-24)$$

$$\begin{aligned} \phi_3\left(\frac{i}{N}\right) = & \left[\frac{2800 \cdot N^5}{(N-1)(N-2)(N+2)(N+3)(N+4)} \right]^{1/2} \\ & \cdot \left[\left(\frac{i}{N} \right)^3 - \frac{3}{2} \left(\frac{i}{N} \right)^2 + \frac{6 \cdot N^2 - 3 \cdot N + 2}{10 \cdot N^2} \left(\frac{i}{N} \right) - \frac{(N-1)(N-2)}{20 \cdot N^2} \right] \end{aligned} \quad (4-25)$$

且求得四維參數後，還原之基頻軌跡為：

$$\hat{f}_{i,n} = \sum_{j=0}^3 a_{j,n} \phi_j\left(\frac{i}{N}\right) \quad (4-26)$$

由此可知，我們只需求取每一個音節的四維參數即可代表音節整段的基頻軌跡，達到參數化的目的，其中第一維代表基頻軌跡之位階(pitch level)，其餘三維則描述基頻軌跡之形狀(pitch shape)，因此我們定義第 n 個音節的基頻軌跡參數如下：

$$\mathbf{sp}_n = [a_{0,n} \ a_{1,n} \ a_{2,n} \ a_{3,n}] \quad (4-27)$$

接著定義音節長度 sd_n 即為強制對齊後，第 n 個音節之音節長度，以及音節能量 se_n 即為第 n 個音節內，韻母所對應之能量最大值。值得注意的是，同 4.2.1 節在本研究將對音節長度以及音節能量進行高斯正規化的動作，以消除每一個語者先天上語速以及能量的差異，在此計算平均值時，同樣將所有特殊音節對應到的音節長度和能量數值排除在外。

4.2.2.2 音節間韻律參數之萃取

接著利用先前得到的聲學參數，求取音節間韻律聲學參數，主要有音節間停頓長度以及音節間能量低點。首先定義音節間停頓長度 pd_n 為語音經過強制對齊後，第 n 個音節後所跟隨之停頓長度，通常音節間停頓長度越大，兩音節之間的互相影響會較弱，值得注意的是，因為 MCDC 為對話語料，在語者思考或發話權在另一語者但標記者沒有將其分割時，會切割出極端的長停頓，其分布如圖 4.7 (a) 所示，這些極端的長停頓在韻律當中的意義和普通長停頓是相同的，因此本研究設計一個 S 型函數 (sigmoid function) 將極端之長停頓壓抑，如圖 4.7 (c) 所示，其數學式如下：

$$pd'_n = 2 \cdot \left(\frac{1}{1 + \exp(-6 \cdot pd_n)} - 0.5 \right) \quad (4-28)$$

而壓縮後之音節停頓長度分布如圖 4.7 (b) 所示，在往後韻律模型的訓練將使用壓縮後之音節停頓長度 pd'_n 。

此外音節間能量低點 ed_n 為第 n 個音節後之停頓長度內的能量最低點，通常音節間能量低點數值越小，兩音節之間的互相影響較弱。

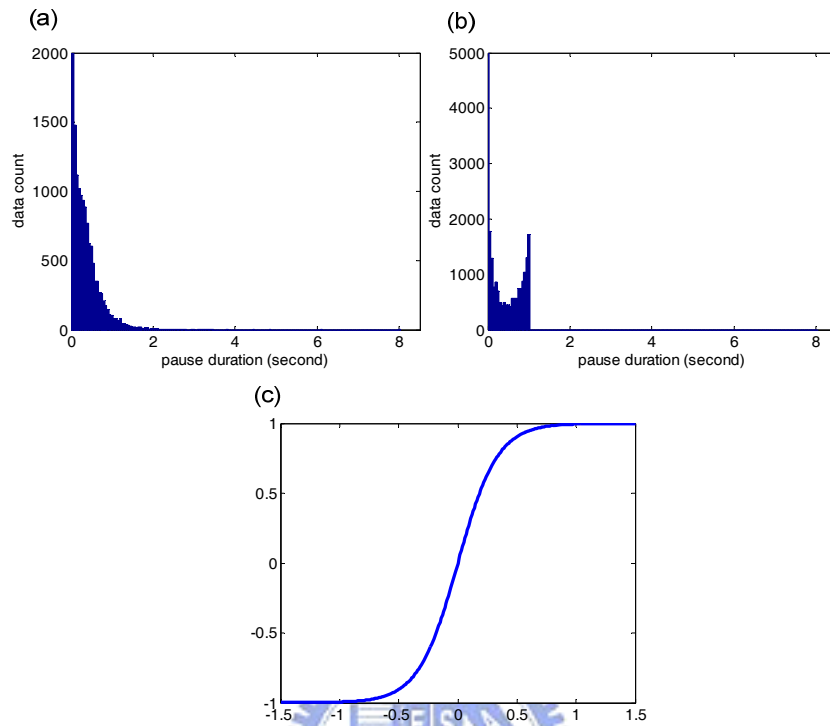


圖 4.7：(a)包含極端長停頓之停頓長度資料分布和 (b)修正後之資料分布及 (c) S 型函數

4.2.2.3 相鄰兩音節間差異韻律參數之萃取

最後利用先前得到的聲學參數，求取相鄰兩音節差異之韻律參數，主要有正規化基頻跳躍值和正規化音節延長因子。正規化基頻跳躍值為扣掉音節層次影響因素後，相鄰兩音節基頻軌跡平均值之差，通常正規化跳躍值越大，表示音節之間的連音現象越弱。正規化音節延長因子為扣掉音節層次影響因素後，相鄰兩音節之音節長度差值，來評估前一音節是否發生音節延長的現象，這些相鄰兩音節差異之韻律參數將在後面的章節有更進一步的定義。

4.2.3 韻律相關語言學參數之萃取

本研究中求取之語言學參數有音節層次以及詞層次上的訊息。音節層次的參數包含了基本音節型態或韻母型態以及音節聲調；詞層次上的參數，主要包含了音節邊界種類、詞長以及詞類。本研究參考附錄一之詞類對照表，並且利用文獻【25】中所實現之中文斷詞器，將 MCDC 資料庫當中每段分割之內容文字進行剖析，如下：

原始內容文字： O 我 在 一 家 公 關 公 司 上 班
 經剖析後文字： O (Par) 我(Nh) 在(P) 一(Neu) 家(Nc) 公關(Nb) 公司(Nc)
 上班(VA) @非語音聲(ParL)

接著再利用附錄二之音節類型、注音以及音節型態對照表，即可獲得詞中每一個音節之字元 (character)、基本音節型態、音節聲調、音節在詞中位置、詞長以及詞類。我們將資料整理如下：

/*字元*/	/*聲調+音節型態編號*/	/*詞長+音節在詞中位置*/	/*詞類編號*/
O	26448	101	60
我	3302	101	24
在	4051	101	26
一	1186	101	17
家	1223	101	14
公	1380	201	13
關	1345	202	13
公	1380	201	14
司	1007	202	14
上	4152	201	28
班	1129	202	28

此外經由詞長以及音節在詞中位置可以得到音節邊界種類資訊，例如：當詞長等於音節在詞中位置時，此音節後之邊界即為詞間音節邊界 (inter-word syllable juncture)，反之則為詞內音節邊界 (intra-word syllable juncture)。值得注意的是，在本研究當中每一個詞所對應到之 2 類、8 類、23 類以及 46 類的詞類都將被使用。

4.3 模型訓練

完成韻律相關之聲學參數以及與語言學參數萃取後，本節中將介紹建構韻律模型之演算法，本研究在訓練四個韻律子模型時，將基於最大概似度準則 (maximum likelihood criterion) 採用逐項最佳化程序 (sequential optimization procedure) 來訓練並更新模型參數，其標的函

數 (objective function) 如下：

$$Q = \left(\prod_{n=1}^N P(\mathbf{sp}_n | p_n, B_{n-1}^n, t_{n-1}^{n+1}) p(sd_n | q_n, B_{n-1}^n, t_{n-1}^{n+1}, s_n) p(se_n | r_n, B_{n-1}^n, t_{n-1}^{n+1}, f_n) \right) \left(P(p_1) P(q_1) P(r_1) \prod_{n=2}^N P(p_n | p_{n-1}, B_{n-1}) P(q_n | q_{n-1}, B_{n-1}) P(r_n | r_{n-1}, B_{n-1}) \right) \left(\prod_{n=1}^{N-1} (p(pd_n, ed_n, pj_n, dl_n | B_n, \mathbf{I}_n) P(B_n | \mathbf{I}_n)) \right) \quad (4-29)$$

在訓練及更新模型參數的過程主要分為兩個步驟，為初始化 (initialization) 以及重覆疊代 (iteration)，分別會在 4.3.1 節以及 4.3.2 節作介紹。

4.3.1 初始化

在許多數據驅動的模型訓練前，都必須有一個符合聲學或語言學特性之初始模型，來幫助我們獲得一個具有物理意義的收斂結果，並且幫助模型的訓練能快速收斂，以下將介紹模型初始化的步驟：

➤ Step1：計算總體平均值

首先將特殊韻律現象之音節排除在外，計算基頻、音節長度以及能量的總體平均值 μ 、 μ_d 以及 μ_e ，數學式如下：

$$\mu = \frac{\sum_{ns_n \in \text{base syllable}} \mathbf{sp}_n}{\sum_{ns_n \in \text{base syllable}} 1} \quad (4-30)$$

$$\mu_d = \frac{\sum_{ns_n \in \text{base syllable}} sd_n}{\sum_{ns_n \in \text{base syllable}} 1} \quad (4-31)$$

$$\mu_e = \frac{\sum_{ns_n \in \text{base syllable}} se_n}{\sum_{ns_n \in \text{base syllable}} 1} \quad (4-32)$$

➤ Step2：計算聲調影響因子之 AP

在許多研究中都認為聲調對音節中韻律表現的影響最大，所以在此我們將聲調作為第一個影響基頻、音節長度以及能量之 AP，並且利用去除總體平均值之殘存值 $\mathbf{sp1}_n = \mathbf{sp}_n - \mu$ 、

$sd1_n = sd_n - \mu_d$ 以及 $se1_n = se_n - \mu_e$ 之平均，計算聲調影響因子對基頻、音節長度以及能量產生之 AP， β_t 、 γ_t 以及 α_t ，在此同樣將特殊韻律現象所對應的音節排除在外，如下：

$$\beta_t = \frac{\sum_{n:s_n \in \text{base syllable}} sp1_n \delta(t_n = t)}{\sum_{n:s_n \in \text{base syllable}} \delta(t_n = t)}, \text{ for base syllable } s \text{ corresponding tone } t \quad (4-33)$$

$$\gamma_t = \frac{\sum_{n:s_n \in \text{base syllable}} sd1_n \delta(t_n = t)}{\sum_{n:s_n \in \text{base syllable}} \delta(t_n = t)}, \text{ for base syllable } s \text{ corresponding tone } t \quad (4-34)$$

$$\alpha_t = \frac{\sum_{n:s_n \in \text{base syllable}} se1_n \delta(t_n = t)}{\sum_{n:s_n \in \text{base syllable}} \delta(t_n = t)}, \text{ for base syllable } s \text{ corresponding tone } t \quad (4-35)$$

值得注意的是，在初始化時本研究先暫時不考慮包含連音現象之影響因子 $\beta_{n-1}^{n+1, B_{n-1}^n}$ 、 $\gamma_{n-1}^{n+1, B_{n-1}^n}$ 以及 $\alpha_{n-1}^{n+1, B_{n-1}^n}$ ，待有較準確之上層韻律狀態的貢獻值後，再於重覆疊代時利用決策樹產生包含前後音節連音現象之聲調 AP。

➤ **Step3：計算基本音節型態影響因子之 AP**

因為基本音節型態和韻母型態對音節長度以及能量也有所影響，例如：擁有介音相較於沒有介音之音節長度較長，以及韻母「a」、「i」、「u」之能量會較「an」、「en」、「ang」大，因此我們將音節型態作為第二個影響音節長度及能量之影響因子，並且利用去除總體平均值及聲調影響之殘存值 $sd2_n = sd1_n - \gamma_{t_n}$ 以及 $se2_n = se1_n - \alpha_{t_n}$ 之平均，計算音節型態影響因子對音節長度以及能量產生之 AP， γ_s 以及 α_s ，在此同樣將特殊韻律現象所對應的音節排除在外如下：

$$\gamma_s = \frac{\sum_{n:s_n \in \text{base syllable}} sd2_n \delta(s_n = s)}{\sum_{n:s_n \in \text{base syllable}} \delta(s_n = s)}, \text{ for base syllable type } s \quad (4-36)$$

$$\alpha_f = \frac{\sum_{n:f_n \in \text{base syllable}} se2_n \delta(f_n = f)}{\sum_{n:f_n \in \text{base syllable}} \delta(f_n = f)}, \text{ for final type } f \quad (4-37)$$

➤ **Step4：計算特殊韻律現象音節之 AP**

最後再針對特殊韻律現象所對應到的音節作處理，使每一類特殊韻律現象之音節都擁有基頻、音節長度以及能量的 AP，並且利用去除總體平均值影響之殘存值 $sp1_n$ 、 $sd1_n$ 以及 $se1_n$ 之平均，計算特殊韻律現象音節對基頻、音節長度以及能量產生之 AP， β_{pr} 、 γ_{pr} 以及 α_{pr} ，如下：

$$\beta_{pr} = \frac{\sum_{n:s_n \in \text{particular syllable}} sp1_n \delta(pr_n = pr)}{\sum_{n:s_n \in \text{particular syllable}} \delta(pr_n = pr)}, \text{ for particular syllable type } pr \quad (4-38)$$

$$\gamma_{pr} = \frac{\sum_{n:s_n \in \text{particular syllable}} sd1_n \delta(pr_n = pr)}{\sum_{n:s_n \in \text{particular syllable}} \delta(pr_n = pr)}, \text{ for particular syllable type } pr \quad (4-39)$$

$$\alpha_{pr} = \frac{\sum_{n:s_n \in \text{particular syllable}} se1_n \delta(pr_n = pr)}{\sum_{n:s_n \in \text{particular syllable}} \delta(pr_n = pr)}, \text{ for particular syllable type } pr \quad (4-40)$$

➤ Step5：標記初始停頓標記

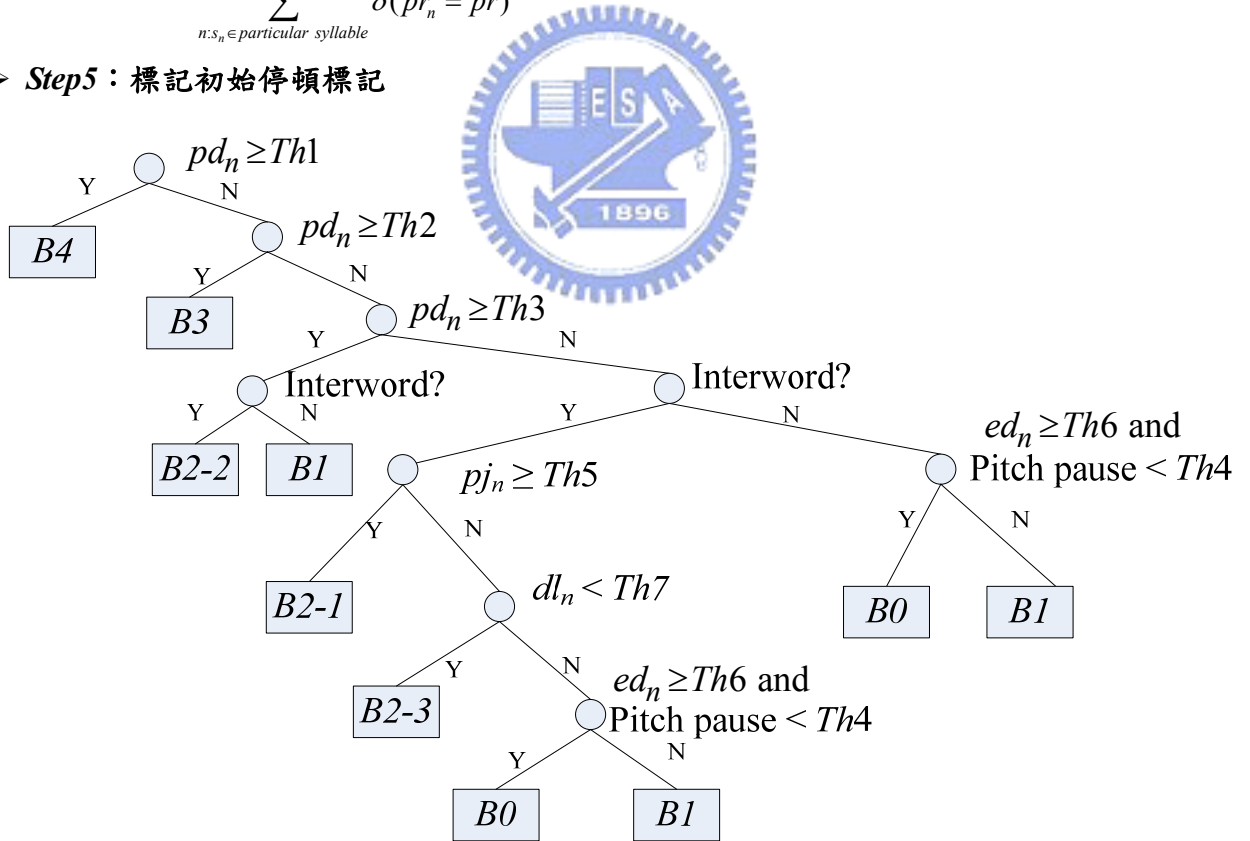


圖 4.8：分類停頓標記之決策樹示意圖

接著使用音節停頓長度、音節間能量低點、正規化基頻跳躍值以及正規化音節延長因子，並利用決策樹的方式，對所有音節邊界處標記初始的停頓型態，如圖 4.8 所示，為了自動得到

一個合理且具有物理意義的決策樹門檻值，在此我們會使用一個系統化的統計方式得到這些決策樹之門檻值。

A. $Th1$, $Th2$ 和 $Th3$ 之定義：

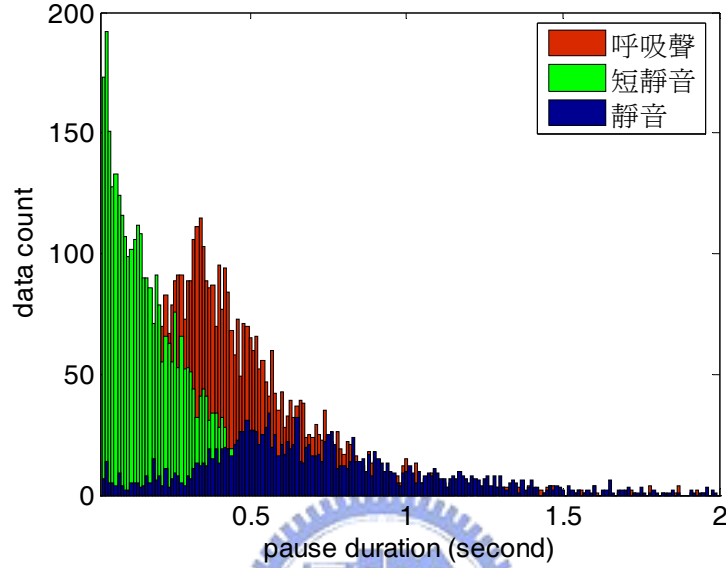


圖 4.9：已標記音節邊界之音節停頓長度分佈

$Th1$ 、 $Th2$ 和 $Th3$ 為區分 $B4$ 、 $B3$ 、 $B2-2$ 以及其它停頓標記之停頓長度門檻值，由於 MCDC 語料庫當中已具有一些語流中斷之標記 (Labeled Juncture, LJ)，例如：靜音 (silence)、短靜音 (short pause, sp) 或呼吸聲 (breathe)，表示這些標記所對應之音節邊界，都有人類可明顯觀察到的停頓現象，其資料分布如圖 4.9 所示，因此在本研究使用一個半監督式 (semi-supervised) 的方法得到 $B4$ 、 $B3$ 、 $B2-2$ 以及其它停頓標記對應之停頓長度機率分佈 $f_{B3}(pd)$ 、 $f_{B4}(pd)$ 、 $f_{B2-2}(pd)$ 和 $f_{else}(pd)$ 以決定 $Th1$ 、 $Th2$ 和 $Th3$ 這些門檻值，如圖 4.10 所示。但是由於人類的標記會有不一致性以及自動切割的方式會造成切割位置不準確的狀況，因此先計算標記為「短靜音」以及其他已標記之停頓長度平均值 μ_{sp} 以及 μ_{LJ} ，以收集較為可靠之停頓長度資料，其中詞內音節邊界、短靜音以及其他已標記之停頓長度資料，如下：

$$pd_{intra} = \{pd_n : pd_n \in \text{intra-word syllable juncture}\} \quad (4-41)$$

$$pd_{sp} = \{pd_n : pd_n \in \text{short pause}, pd_n - \mu_{sp} < pd_n - \mu_{LJ}, pd_n \geq 0.03\} \quad (4-42)$$

$$pd_{LJ} = \{pd_n : pd_n \in \text{silence or breathe}, pd_n - \mu_{LJ} < pd_n - \mu_{sp}, pd_n \geq 0.03\} \quad (4-43)$$

接著以向量量化 (Vector Quantization, VQ) 的方式將 pd_{LJ} 中的資料分為兩群， pd_{sp} 以及 pd_{intra} 的資料視為第三群和第四群，使用伽瑪分佈建構這四群資料的機率分佈，如圖 4.10 (a)和(b)所示，平均值由高至低排列之分布分別為 $f_{B4}(pd)$ 、 $f_{B3}(pd)$ 、 $f_{B2-2}(pd)$ 及 $f_{else}(pd)$ ，並且設定等機率點對應之停頓長度值即為 $Th1$ 、 $Th2$ 和 $Th3$ ，如圖 4.10 (c) 所示。

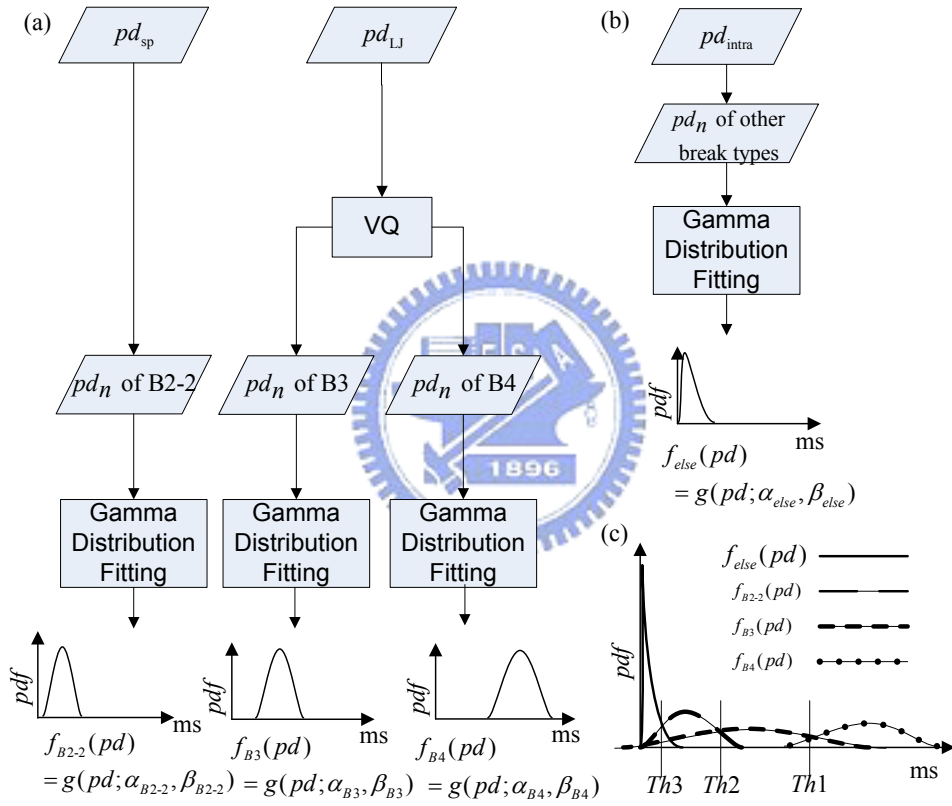


圖 4.10： $Th1$ 、 $Th2$ 和 $Th3$ 之定義方法：計算 (a) B4、B3 和 B2-2 與 (b) 其他停頓標記之停頓長度的機率分佈以及 (c) B4、B3、B2-2 和其他停頓標記之門檻值

B. $Th5$ 之定義：

$Th5$ 是將小於 $Th3$ 之詞間音節邊界，再分為 B2-1 以及 B0/B1/B2-3 之正規化基頻差值。首先定義第 n 個音節的正規化基頻跳躍值 pj_n 為扣掉總體平均值以及聲調 AP 後，基頻殘餘量之第一維差值，如下：

$$pj_n = \text{sp}2_{n+1}(1) - \text{sp}2_n(1) \quad (4-44)$$

其中扣掉總體平均值以及聲調 AP 後之基頻殘餘量如下：

$$\text{sp}2_n = \begin{cases} \text{sp}1_n - \beta_{pr_n}, & \text{if } nth \text{ syllable is particular syllable } pr \\ \text{sp}1_n - \beta_{t_n}, & \text{if } nth \text{ syllable is base syllable } s \text{ corresponding tone } i \end{cases} \quad (4-45)$$

接著將詞內音節邊界和所有已標記之音節邊界，所對應到之相鄰兩音節正規化基頻跳躍值，各自以高斯分佈建構這兩群資料之機率分布，並設定兩個高斯機率分佈的等機率點所對應到之正規化基頻差值為 $Th5$ 。

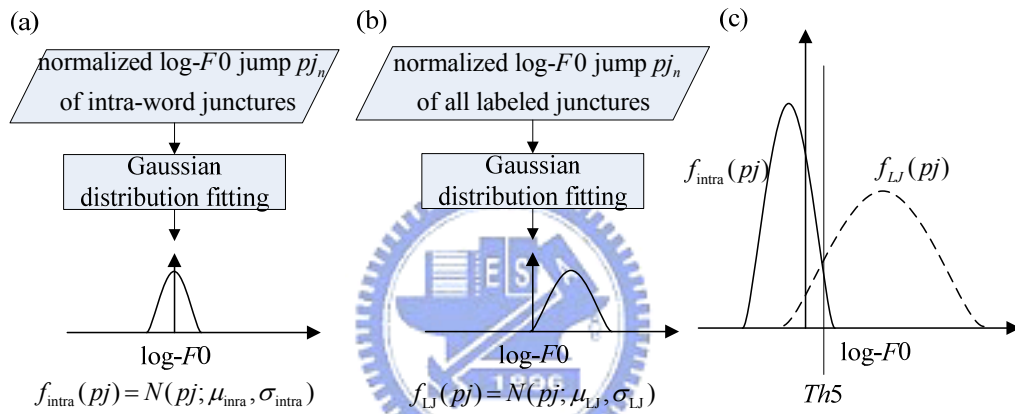


圖 4.11： $Th5$ 之定義方法：計算 (a)詞內與 (b)已標記之音節邊界正規化基頻差之機率分佈以及 (c)門檻值之定義

C. $Th4$ 和 $Th6$ 之定義：

$Th4$ 和 $Th6$ 是將小於 $Th3$ 之詞內音節邊界，再分為 $B0$ 以及 $B1$ 之基頻停頓長度及音節間能量低點，其中基頻停頓長度即為兩音節間基頻軌跡之距離，如圖 4.4 所示。一般來說 $B0$ 表示兩音節有非常緊密的連接，所以相較於 $B1$ 會有較短的基頻停頓長度，以及較大的音節能量低點，所以我們定義 $Th4$ 的值為一個音框長度 (=10ms)。接著將所有詞內音節邊界所對應之能量低點資料，以向量量化的方式分為兩群，並且用兩個高斯分佈建構這兩群的機率分布，平均值由高至低排列分別為 $f_{B0}(pd)$ 及 $f_{B1}(pd)$ ，並且設定等機率點對應之能量低點值即為 $Th6$ ，如圖 4.12 所示。

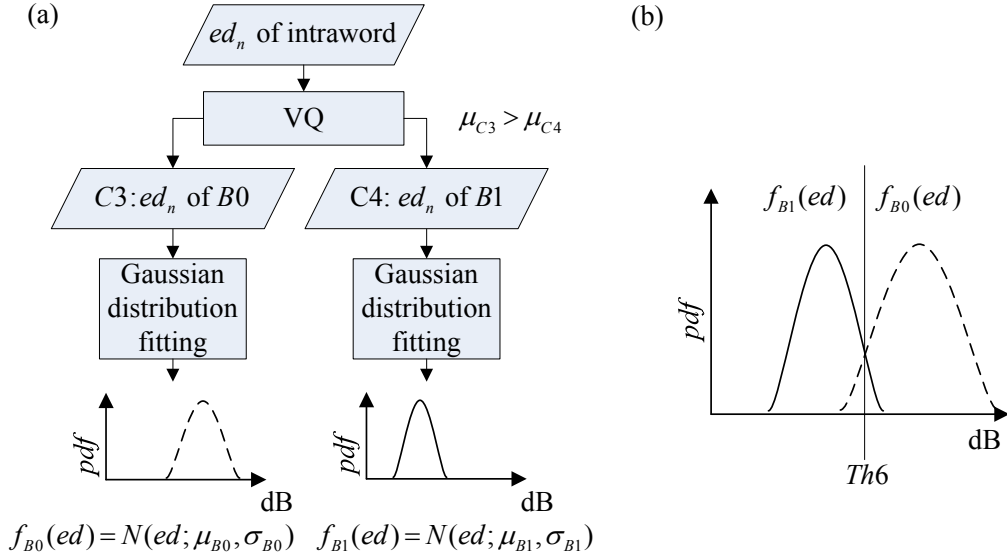


圖 4.12：Th6 之定義方法：計算 (a)B0 和 B1 音節能量低點之機率分佈以及 (b)門檻值之定義

D. Th7 之定義：

Th7 是將小於 Th3 且小於 Th5 的詞間音節邊界，再分為 B2-3 以及 B0/B1 之相鄰兩音節正規化音節延長因子門檻值。首先定義相鄰兩音節之正規化音節延長因子 dl_n ，如下：

$$dl_n = sd3_n - sd3_{n+1} \quad (4-46)$$

其中 $sd3_n$ 為第 n 個音節扣掉總體平均值、音調以及基本音節型態 AP 之音節長度殘存值，如下：

$$sd3_n = \begin{cases} sd1_n - \gamma_{pr}, & \text{if } nth \text{ syllable is particular syllable} \\ sd2_n - \gamma_s, & \text{if } nth \text{ syllable is base syllable} \end{cases} \quad (4-47)$$

本研究以正規化音節延長因子來表示在上層的韻律當中，音節邊界之前一音節是否有發生音節延長的現象，若數值越大表示前一音節延長之現象越嚴重。接著我們將詞內音節邊界和所有已標記之音節邊界，所對應到之相鄰兩音節正規化音節延長因子數值，各自以高斯機率建構這兩群之機率分佈，並設定兩個高斯機率分佈等機率點所對應到之正規化音節延長因子值為 Th7，如圖 4.13 所示。

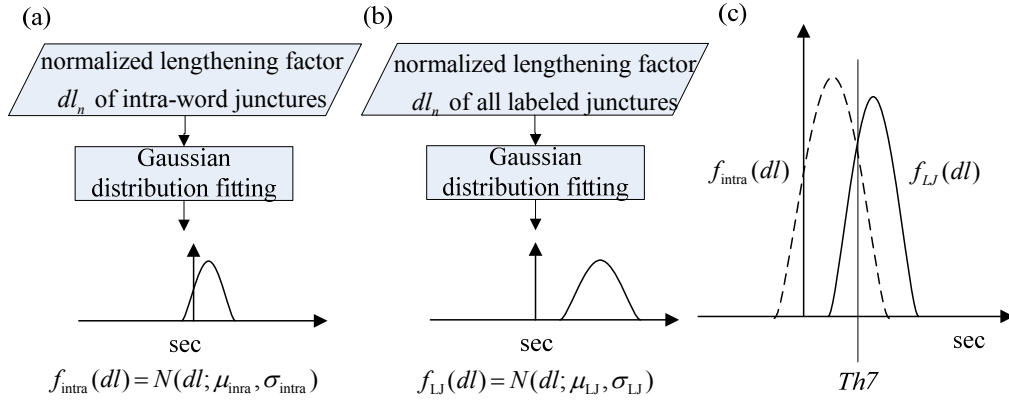


圖 4.13：Th7 之定義方法：計算 (a)詞內音節邊界和 (b)已標記之音節邊界的相鄰兩音節正規化音節延長因子之機率分佈以及 (b)門檻值之定義

➤ Step6：標記初始韻律狀態

得到初始的停頓標記之後，本研究將扣掉總體平均值、聲調以及基本音節型態 AP 之基頻、音節長度以及能量殘存值 $sp2_n$ 、 $sd3_n$ 以及 $se3_n$ ，以向量量化的方式計算基頻、音節長度以及能量之每一個韻律狀態的代表值 β_{p_n} 、 γ_{q_n} 以及 α_n ，並且針對每一個音節標記其擁有之韻律狀態，其中能量殘存值的定義如下：

$$se3_n = \begin{cases} se1_n - \alpha_{pr}, & \text{if } nth \text{ syllable is particular syllable} \\ se2_n - \alpha_s, & \text{if } nth \text{ syllable is base syllable} \end{cases} \quad (4-48)$$

值得注意的是，本研究將基本音節之韻律狀態分為 16 個狀態，而特殊韻律狀態之音節則另外再分 4 個韻律狀態。

➤ Step7：求取韻律狀態轉移模型

韻律狀態以及停頓標記完成後，將這些標記的結果以相對計數率 (relative count) 的方式計算每種停頓標記的情況之下，韻律狀態轉移之機率 $P(\mathbf{P}|\mathbf{B})$ 。

➤ Step8：求取停頓標記聲學模型及停頓標記語言模型

本研究採用 CART (Classification And Regression Trees) 演算法實現停頓標記聲學模型以及停頓標記語言模型之決策樹。停頓標記聲學模型以最大概似度增益為分裂準則區分停頓長度 pd 、音節間能量低點 ed 、正規化基頻跳躍值 pj 以及正規化音節延長因子 dl ，其問題集 Θ_l 是由詢問語言參數 \mathbf{l} 的問題所設計，如附錄三所示。在停頓標記聲學模型中，每一類停頓標記都將產生一顆決策樹，並且在每一個節點以停頓長度之伽瑪分布 $g(pd_n; \alpha_{B_n, l_n}, \beta_{B_n, l_n})$ 與其他參數之高斯分布 $N(ed_n; \mu_{B_n, l_n}, \sigma_{B_n, l_n}^2)$ 、 $N(pj_n; \mu_{B_n, l_n}, \sigma_{B_n, l_n}^2)$ 以及 $N(dl_n; \mu_{B_n, l_n}, \sigma_{B_n, l_n}^2)$ 相乘來表示。同樣

地，停頓標記語言模型是以最大概似度增益為分裂準則區分各種停頓標記，其問題集 Θ_2 同樣是由詢問語言參數 \mathbf{l} 的問題所設計，如附錄四所示。在停頓標記語言模型中，在每一個節點將使用相對計數率的方式產生每種停頓標記之機率。

➤ **Step10：求取音節韻律模型之共變異矩陣**

音節韻律模型包含了音節基頻軌跡模型、音節長度模型以及音節能量模型，是經由扣除所有影響基頻、音節長度以及能量之 AP 後，將其殘存值以一個平均值為零的高斯分布去描述它，且共變異矩陣分別如下：

$$\mathbf{R} = \frac{\sum_{n:s_n \in \text{base syllable}} \mathbf{sp}_n^r (\mathbf{sp}_n^r)^T}{N_{\text{base}}}, \text{ for all base syllable} \quad (4-49)$$

$$R_d = \frac{\sum_{n:s_n \in \text{base syllable}} sd_n^r (sd_n^r)^T}{N_{\text{base}}}, \text{ for all base syllable} \quad (4-50)$$

$$R_e = \frac{\sum_{n:s_n \in \text{base syllable}} se_n^r (se_n^r)^T}{N_{\text{base}}}, \text{ for all base syllable} \quad (4-51)$$

$$\mathbf{R}' = \frac{\sum_{n:s_n \in \text{particular syllable}} \mathbf{sp}_n^r (\mathbf{sp}_n^r)^T}{N_{\text{par}}}, \text{ for all particular syllable} \quad (4-52)$$

$$R'_d = \frac{\sum_{n:s_n \in \text{particular syllable}} sd_n^r (sd_n^r)^T}{N_{\text{par}}}, \text{ for all particular syllable} \quad (4-53)$$

$$R'_e = \frac{\sum_{n:s_n \in \text{particular syllable}} se_n^r (se_n^r)^T}{N_{\text{par}}}, \text{ for all particular syllable} \quad (4-54)$$

其中

$$\mathbf{sp}_n^r = \begin{cases} \mathbf{sp}2_n - \boldsymbol{\beta}_{p_n}, & \text{if } nth \text{ syllable is base syllable} \\ \mathbf{sp}2_n - \boldsymbol{\beta}_{p'_n}, & \text{if } nth \text{ syllable is particular syllable} \end{cases} \quad (4-55)$$

$$sd_n^r = \begin{cases} sd3_n - \gamma_{q_n}, & \text{if } nth \text{ syllable is base syllable} \\ sd3_n - \gamma_{q'_n}, & \text{if } nth \text{ syllable is particular syllable} \end{cases} \quad (4-56)$$

$$se_n^r = \begin{cases} se3_n - \alpha_{r_n}, & \text{if } nth \text{ syllable is base syllable} \\ se3_n - \alpha_{r'_n}, & \text{if } nth \text{ syllable is particular syllable} \end{cases} \quad (4-57)$$

分別為第 n 個音節扣除所有 AP 之基頻、音節長度以及能量殘存值，且 N_{base} 及 N_{par} 分別為基本音節和特殊音節之總個數。

➤ Step11：建構聲調分類器

在之前的初始化過程中並沒有將音節間的連音現象考慮在內，但連音狀態對音節基頻軌跡會造成很大的影響，因此在單純利用音節聲調之影響因子求得較為準確之韻律狀態貢獻值後，利用扣除總體平均值以及韻律狀態貢獻之基頻殘存值，以 CART 演算法及最小歐幾里得距離（minimum Euclidean distance）為分裂準則區分包含連音現象聲調之 AP，實現一個包含連音現象之聲調 AP 分類器（classifier），供重覆疊代時使用，其問題集 Θ_3 是由詢問左右音節聲調及音節前後邊界之停頓標記的問題所設計，如附錄五所示。在聲調分類器中，每一類聲調都將產生一顆決策樹，並且在每一個節點將產生一組基頻、音節長度以及能量之 AP。其中扣除基頻總體平均值以及韻律狀態基頻貢獻值之殘存值如下：

$$sp3_n = sp_n - \beta_{p_n} - \mu, \quad \text{for } nth \text{ syllable is base syllable} \quad (4-58)$$

4.3.1 重覆疊代

在有好的初始化模型後，以下將介紹模型重覆疊代的步驟：

➤ Step1：更新影響因子

使用聲調分類器計算 $\beta_{t_{n-1}^{n+1}, B_{n-1}^n}$ 、 $\gamma_{t_{n-1}^{n+1}, B_{n-1}^n}$ 以及 $\alpha_{t_{n-1}^{n+1}, B_{n-1}^n}$ 之 AP；以及利用(4-36)和(4-37)式更新 γ_{s_n} 和 α_s ，在此之前所有聲調影響因子之 AP，將被聲調分類器分類後之 AP 所取代。

➤ Step2：更新共變異矩陣

由(4-49)至(4-54)式更新共變異矩陣 \mathbf{R} 、 R_d 以及 R_e 。

➤ Step3：重新標記韻律狀態

利用維特比搜尋演算法（Viterbi search algorithm）重新標記韻律狀態，使(4-29)式之標的函數 Q 之概似度（likelihood）最大化。

➤ Step4：更新韻律狀態轉移模型及共變異矩陣

利用相對計數率更新韻律狀態轉移模型，並且由(4-49)至(4-54)式更新共變異矩陣。

➤ Step5：重新標記停頓狀態

利用維特比搜尋演算法（Viterbi search algorithm）重新標記停頓標記，使(4-29)式之標的函數 Q 之概似度最大化，值得注意的是，本研究將對維特比搜尋演算法加以限制，若停頓長度小

於 0.03 秒之音節邊界，只允許此音節邊界可能之停頓標記為 $\{B0, B1, B2-1, B2-3\}$ ，反之音節邊界可能之停頓標記為 $\{B2-2, B3, B4\}$ 。

➤ **Step6：更新韻律狀態轉移模型及共變異矩陣**

利用相對計數率更新韻律狀態轉移模型，以及由(4-49)至(4-54)式更新共變異矩陣。

➤ **Step7：更新停頓標記聲學模型以及停頓標記語言模型**

利用 CART 演算法以及問題集 Θ_1 及 Θ_2 更新停頓標記聲學模型以及停頓標記語言模型。

➤ **Step8：重覆 Step1 到 Step8 至收斂**



第五章：實驗結果

由於韻律為語音中超區段（suprasegmental）之現象，因此在訓練韻律模型時本研究將小於五個音節之語句或特殊音節過多⁶之語句排除在外，共包含 3,501 段語句共 121,242 個音節。逐項最佳化程序共重覆疊代 59 次收斂，其對應之標的函數值（total likelihood of objective function）如圖 5.1 所示。在本章節當中將會介紹四個子模型參數之觀察結果以及韻律標記之分析。5.1 節將介紹韻律模型參數之分析；5.2 節將分析韻律標記之結果；5.3 節將會針對自發性語音中之詞語修補的部份和韻律標記互相對應並討論。

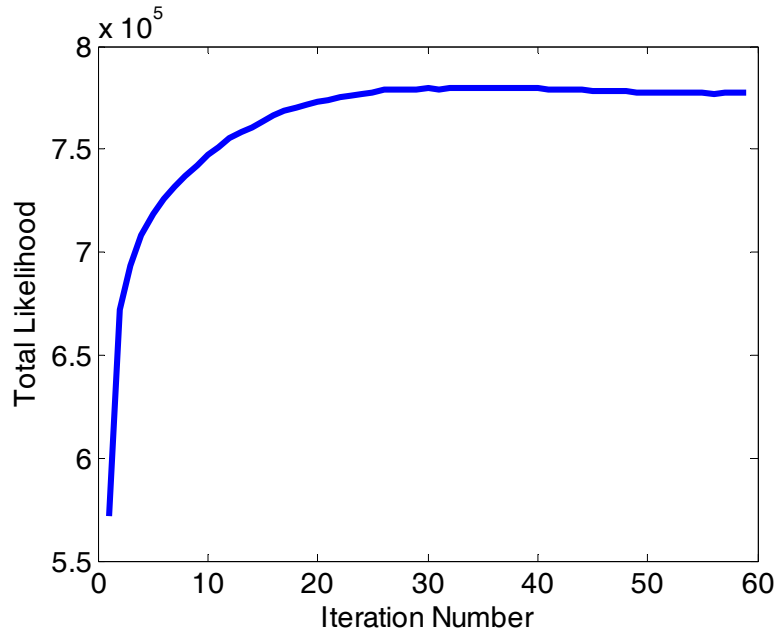


圖 5.1：疊代次數與標的函數值

5.1 音節韻律模型

音節韻律模型主要分成三個子模型分別為：音節基頻軌跡模型 $p(\text{sp}|\mathbf{B}, \mathbf{p}, \mathbf{t})$ 、音節長度模型 $p(\text{sd}|\mathbf{B}, \mathbf{q}, \mathbf{t}, \mathbf{s})$ 以及音節能量模型 $p(\text{se}|\mathbf{B}, \mathbf{r}, \mathbf{t}, \mathbf{f})$ ，它們主要描述音節中各個影響因子對音節基頻軌跡、音節長度以及音節能量之貢獻，首先觀察三個模型收斂前後之共變數矩陣，分別如

⁶ 本研究定義過多特殊韻律現象之語句為：語句中特殊韻律現象音節與普通音節比率大於 1/6。

下：

$$\mathbf{R}_{sp} = \begin{bmatrix} 341.8 & -53.4 & -10.3 & -1.1 \\ -53.4 & 35.1 & -1.4 & -3.6 \\ -10.3 & -1.4 & 11.01 & 34.2 \\ -1.1 & -3.6 & 34.2 & 4.8 \end{bmatrix} \times 10^{-4} \Rightarrow \mathbf{R}_{sp^r} = \begin{bmatrix} 10.5 & -0.4 & -0.4 & 0.1 \\ -0.4 & 30.1 & -0.9 & -3.1 \\ -0.4 & -0.9 & 10.2 & 0.3 \\ 0.1 & -3.1 & 0.3 & 4.6 \end{bmatrix} \times 10^{-4}$$

$$R_{sd} = 576.4 \times 10^{-5} \Rightarrow R_{sd^r} = 9.7 \times 10^{-5}$$

$$R_{se} = 45.0 \Rightarrow R_{se^r} = 1.1$$

觀察後可以發現，收斂前後各個音節韻律模型之共變數矩陣數值都明顯變小，由此可知，確實能使用各個影響因子來承載這些音節中韻律參數的變化。接著表 5.1 及表 5.2 中列出了基本音節及特殊音節中，扣除不同影響因子之下，各韻律參數之總殘餘誤差值（Total Residual Error, TRE），即為扣掉各種組合 AP 後殘存值之變異數與原始資料之變異數的比值，藉此觀察各個影響因子對音節中韻律參數變化之貢獻大小，由表 5.1 可以得知基本音節中基頻軌跡的變化貢獻由大至小分別為：韻律狀態以及包含連音現象之聲調；而對音節長度及音節能量的變化貢獻由大至小分別為：韻律狀態、基本音節型態以及包含連音現象之聲調，本研究將對這些 AP 分為音節層次及上層韻律狀態層次的影響，在 5.2.1 及 5.2.2 節作更進一步的探討。

表 5.1：基本音節中，不同組合之 AP 下音節韻律模型參數之 TRE

Pitch		Duration		Energy	
APs	TRE	APs	TRE	APs	TRE
+Tone with Coarticulation	90.61%	+Tone with Coarticulation	93.95%	+Tone with Coarticulation	94.27%
		+Base Syllable	83.70%	+Final	86.80%
+Prsodic State	14.11%	+Prsodic State	1.68%	+Prsodic State	2.54%

表 5.2：特殊音節中，不同組合之 AP 下音節韻律模型參數之 TRE

Pitch		Duration		Energy	
APs	TRE	APs	TRE	APs	TRE
+Particular Syllable Class	88.10%	+Particular Syllable Class	96.60%	+Particular Syllable Class	91.40%
+Prsodic State	48.30%	+Prsodic State	18.60%	+Prsodic State	44.20%

5.1.1 音節層次之影響型態

在 5.1.1.1、5.1.1.2 以及 5.1.1.3 節將觀察音節層次中基頻、音節長度以及音節能量，各影響因子 AP 的變化狀況。

5.1.1.1 音節層次中基頻之影響型態

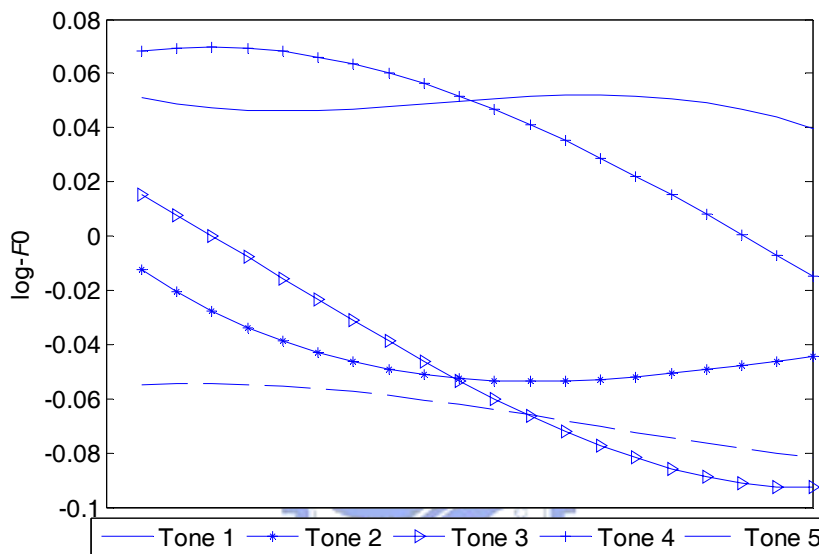


圖 5.2：五個中文聲調之 AP

首先觀察考慮連音現象之聲調影響因子，由於此影響因子經由一個聲調分類器之決策樹所產生，因此藉由觀察每一個聲調決策樹的根節點（root node）來顯示未考慮連音現象時，每一個聲調之 AP，如圖 5.2 所示，由此圖與朗讀式語音之聲調【7】相比，發現在自發性語音當中，聲調之 AP 之動態範圍（dynamic range）以及基頻軌跡之曲度都較小，在此我們推測原因有以下幾點：

- 首先為自發性語音之語速較朗讀式語音快，音節會產生緊密連接的狀況較多，因此音節基頻軌跡會受鄰近音節的嚴重干擾，以及自發性語音中發音經常不完全而使聲調之基頻軌跡相當凌亂。
- 接著又因為語速較快，人類為了節省發音所需之力氣而刻意降低音節基頻變化之幅度，因此弱化了基頻軌跡之曲度以及動態範圍。
- 最後由於本研究所得之音節切割乃利用聲學模型強迫對齊得來，而自發性語音至今仍

無良好的聲學模型，因此會造成切割位置不準確的狀況而影響音節中基頻軌跡之求取。以上幾點也是使自發性語音中各個 AP 之 TRE 值較小之原因。

接著藉由觀察決策樹之各節點，找出考慮連音現象各聲調之 AP，決策樹之示意圖如圖 5.3 所示；各個聲調之決策樹如圖 5.4 至圖 5.7 所示。首先在這四張圖中詢問音節邊界左右之停頓標記皆以「是否為 $B0$ 」或「是否為 $B1$ 」的問題為主且位於決策樹較上層之節點，而在決策樹分裂之演算法中，會將較具鑑別性之問題先詢問，這表示連音現象在音節沒有產生韻律上的停頓時較為嚴重；接著觀察「左邊音節邊界是否為 $B0$ 」或「左邊音節邊界是否為 $B1$ 」的問題皆位於決策樹之根節點或第二層之節點，這表示連音現象是受前一音節之影響較大，這些現象都與【7】中觀察到連音現象之結果相同。此外觀察各聲調節點中與根節點之標準型態（typical pattern）相差甚大之 AP，如下：

➤ 一聲之決策樹，如圖 5.4 所示：

- (1) 觀察節點 4 及節點 12，前一音節皆為三聲，且一聲之起頭較三聲之結尾高，而產生 low-to-high mismatch，使一聲之 AP 起頭下沉，其中又以節點 12 與前一音節為緊密連接而產生較嚴重之連音現象。
- (2) 觀察節點 22，後一音節皆為四聲且為緊密連接，且一聲之結尾與四聲之開頭皆為位處高的層級，但人類為了突顯四聲之基頻軌跡，而刻意提高一聲之結尾。
- (3) 另外在節點 34 中之音節之結尾也發生上揚的現象，是由於音節後邊界皆標記為 $B2-1$ ，而一聲本身之結尾已經是處在高的狀態，為了突顯基頻跳躍而拉高一聲結尾之音高。

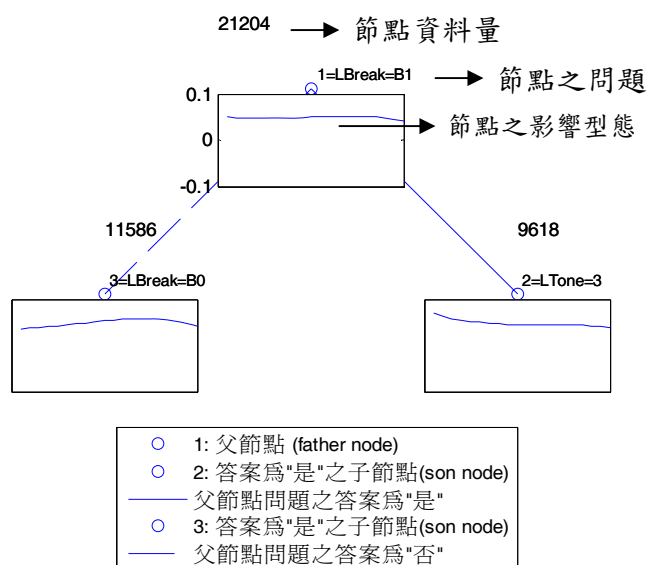


圖 5.3：聲調決策樹之示意圖

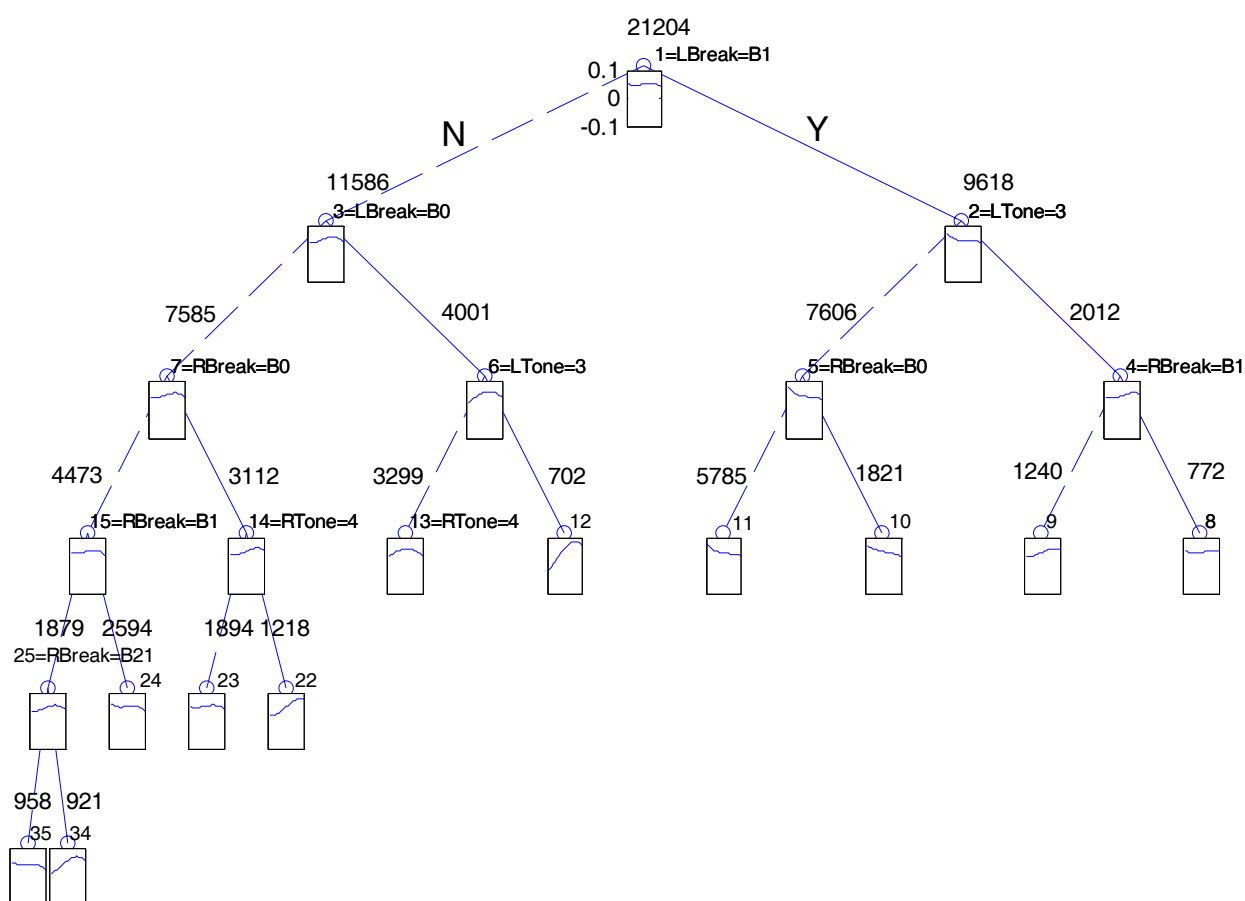


圖 5.4：考慮連音現象一聲聲調之 AP

➤ 二聲之決策樹，如圖 5.5 所示：

(1) 觀察節點 41、節點 45 以及它們的子節點 (son node) 之 AP，前一音節皆為一聲或

四聲，且二聲之起頭較一聲或四聲之結尾低，而產生 high-to-low mismatch，使二聲之 AP 起頭上揚，尤其又以節點 43 以及節點 53 與前音節為緊密連接而更加嚴重。

- (2) 觀察節點 57，後一音節皆為三聲，且二聲之結尾較後一音節三聲之起頭低，而產生 low-to-high mismatch，使二聲之 AP 結尾上揚。

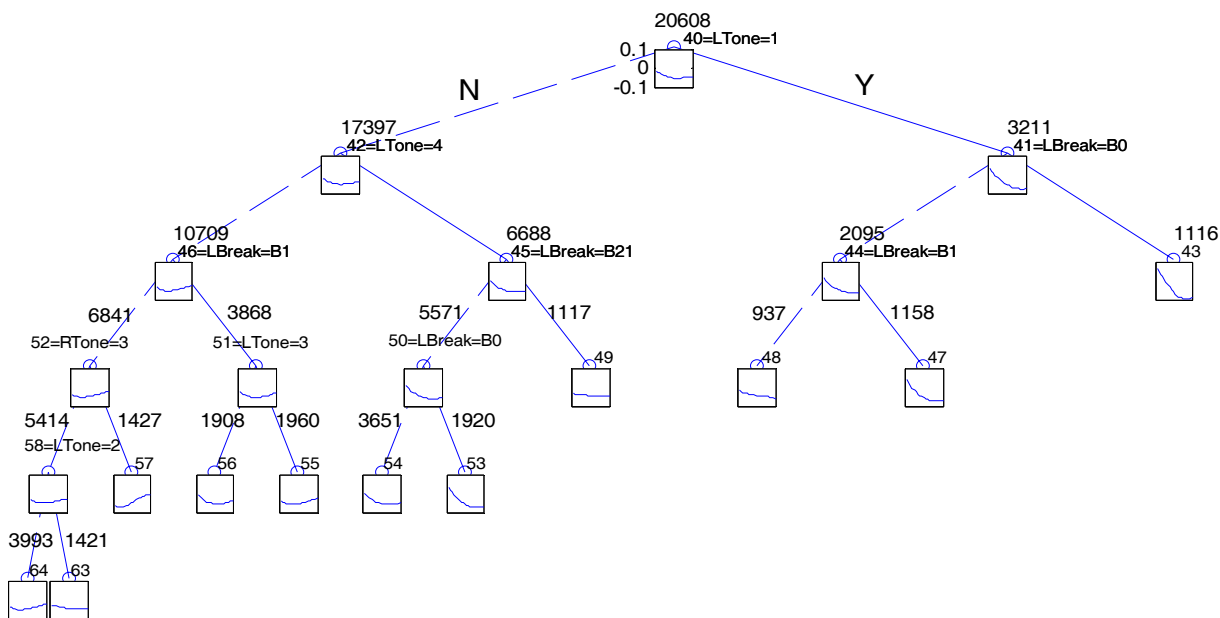


圖 5.5：考慮連音現象二聲聲調之 AP

➤ 三聲之決策樹，如圖 5.6 所示：

- (1) 觀察節點 72 及其子節點，此節點之音節左邊界皆與前一音節緊密連接，卻在音節右邊界結尾處發生嚴重下沉，表示在自發性語音當中，當三聲與前一音節緊密連接時，人類會順勢將三聲之基頻軌跡迅速下沉並且利用尾端的稍微揚起來強調三聲，而不是將三聲拉至較高的準位再下沉。
- (2) 觀察節點 82、86、96 以及 98，後一音節皆為三聲，且三聲之結尾較後一音節三聲之起頭低，而產生 low-to-high mismatch，使三聲 AP 之結尾上揚
- (3) 而節點 84 則是因為音節後邊界皆標記為 *B2-1*，為了突顯基頻跳躍而拉高三聲之結尾。

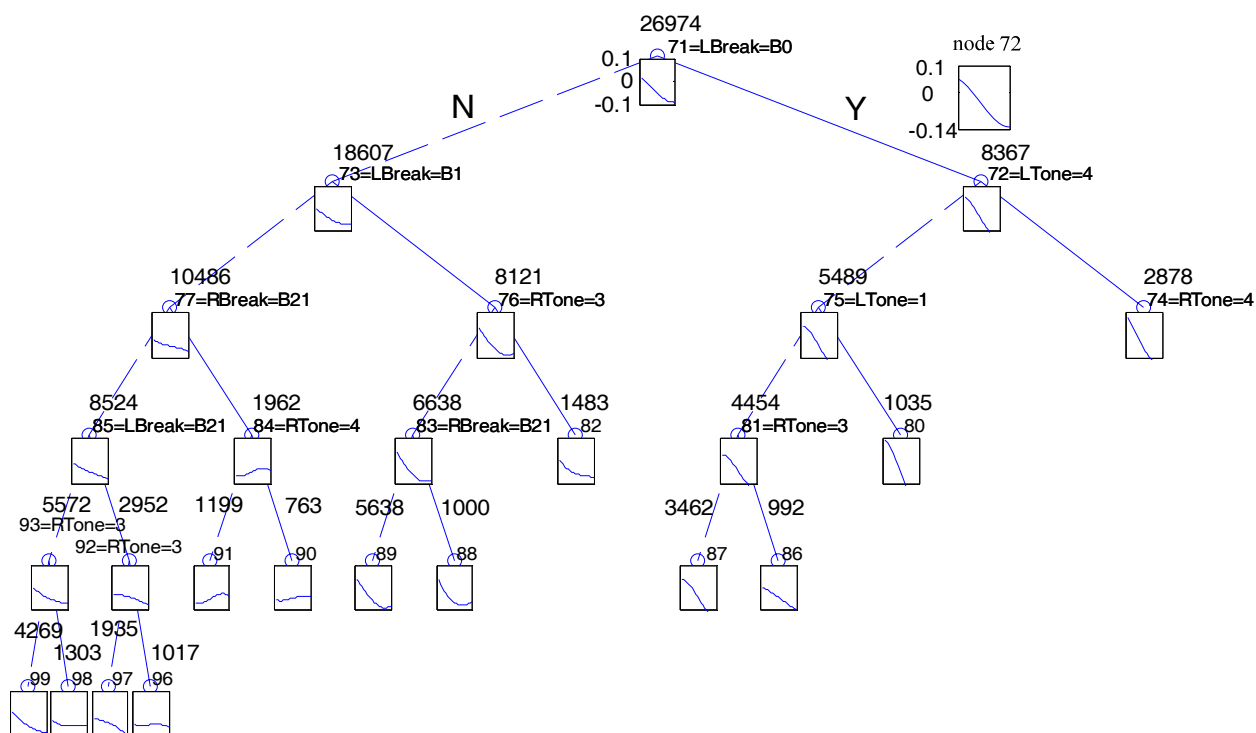


圖 5.6：考慮連音現象三聲聲調之 AP

➤ 四聲之決策樹，如圖 5.7 所示：

- (1) 觀察節點 115、133 及 163，前一音節皆為三聲，且四聲之起頭較三聲之結尾高，而產生 low-to-high mismatch，使四聲之 AP 起頭下沉。
- (2) 觀察節點 125、161 及 157，前一音節皆為一聲，且一聲之結尾與四聲之開頭皆為位處高的層級，人類為了突顯四聲之基頻軌跡，而刻意提高四聲之開頭。
- (3) 觀察節點 129、151 及 153，後一音節皆為四聲，且四聲之結尾較後一音節四聲之起頭低，而產生 low-to-high mismatch，使四聲 AP 之結尾上揚。

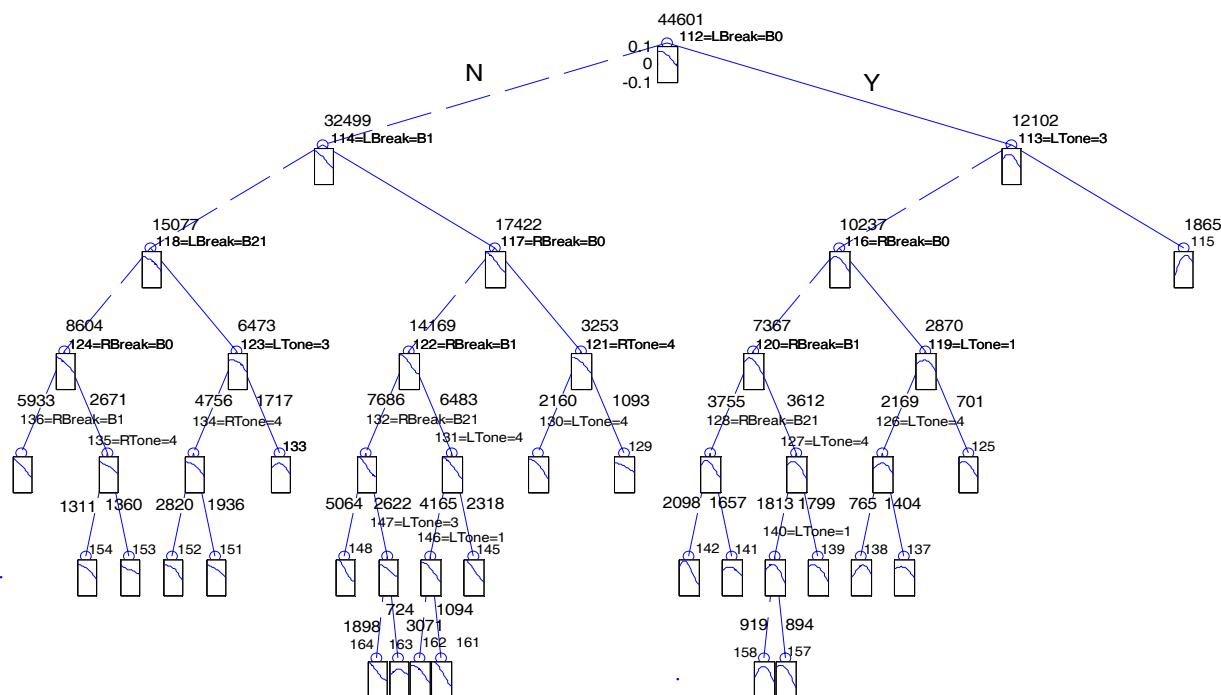


圖 5.7：考慮連音現象之四聲聲調 AP

最後觀察基本音節與特殊音節在音節層次基頻 AP 之差異，如圖 5.8 所示。在基本音節當中一聲及四聲之基頻平均值較高，但基本音節與特殊音節相比，特殊音節之基頻平均值大部分偏高。

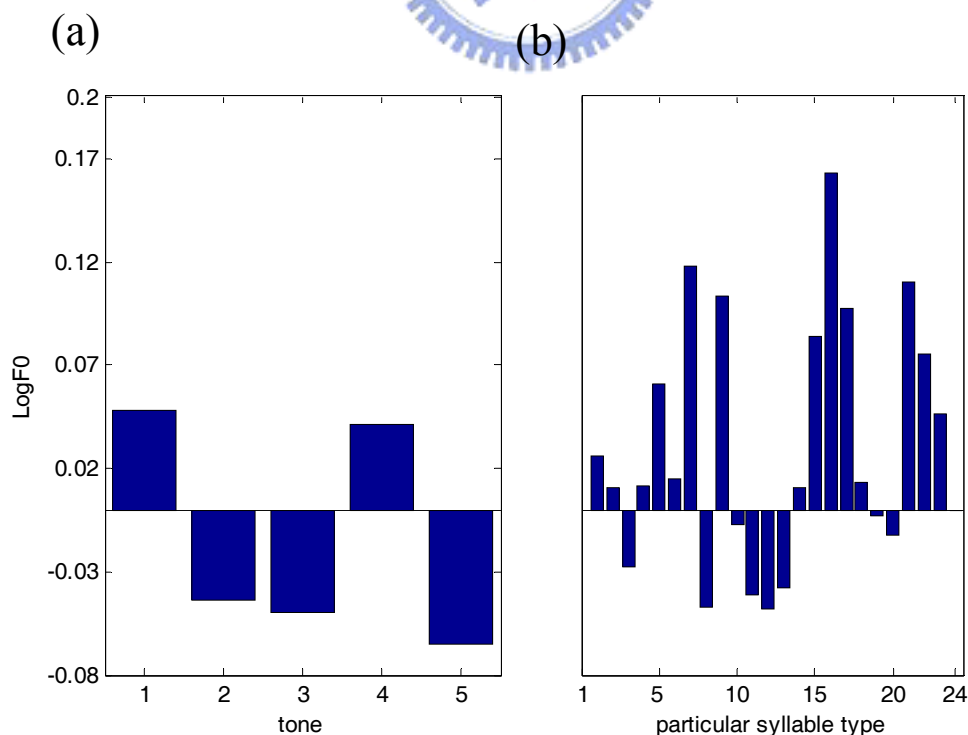


圖 5.8：基本音節中 (a)聲調以及特殊音節中 (b)特殊音節型態基頻 AP 之第一維數值

5.1.1.2 音節層次中音節長度之影響型態

接著觀察各個影響因子下音節長度之 AP，並與特殊音節做比較，如圖 5.9 所示。在基本音節中一聲及四聲之音節長度較長，五聲明顯較短，而基本音節與特殊音節相比，其動態範圍差不多，但特殊音節之長度以偏長稍多。

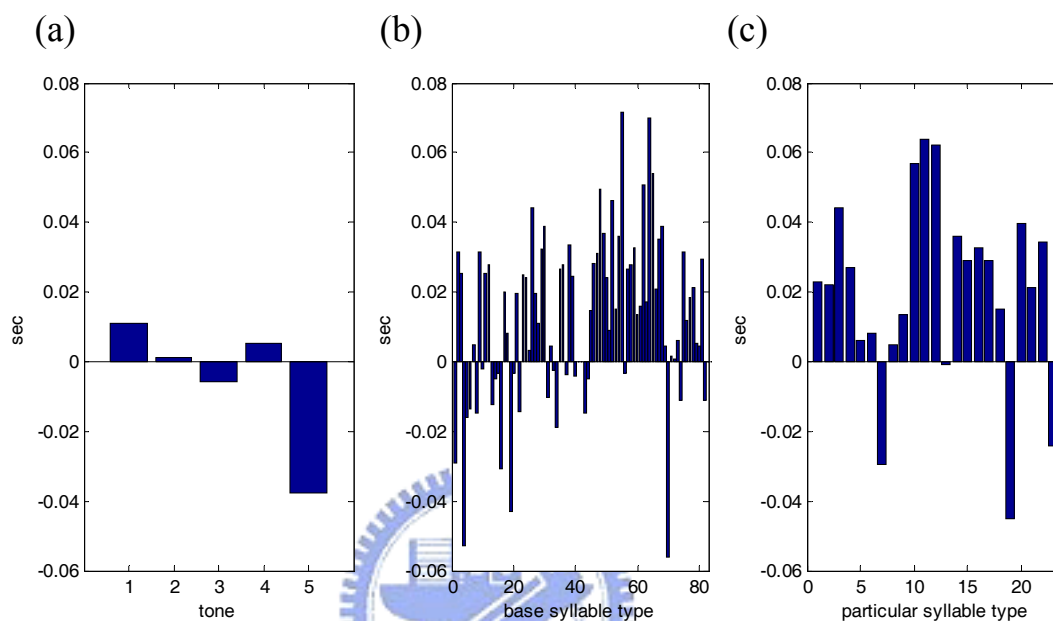


圖 5.9：基本音節中 (a)聲調、(b) 基本音節型態以及特殊音節中 (c)特殊音節型態之音節長度 AP

5.1.1.3 音節層次中音節長度之影響型態

最後觀察各個影響因子能量之 AP，並與特殊音節做比較，如圖 5.10 所示。在基本音節中一聲及四聲之能量較大，五聲能量較小。而基本音節與特殊音節相比，其動態範圍差不多，但特殊音節之能量大部分明顯偏小，由此可知語者在特殊音節發音時，因為其大多數為語音中較不帶訊息之篇章提語、感歎詞或語助詞，因此不會加強其能量，這些基頻、音節長度或能量上的現象可提供未來在偵測篇章提語、感歎詞或語助詞時，作為一個有用的訊息。

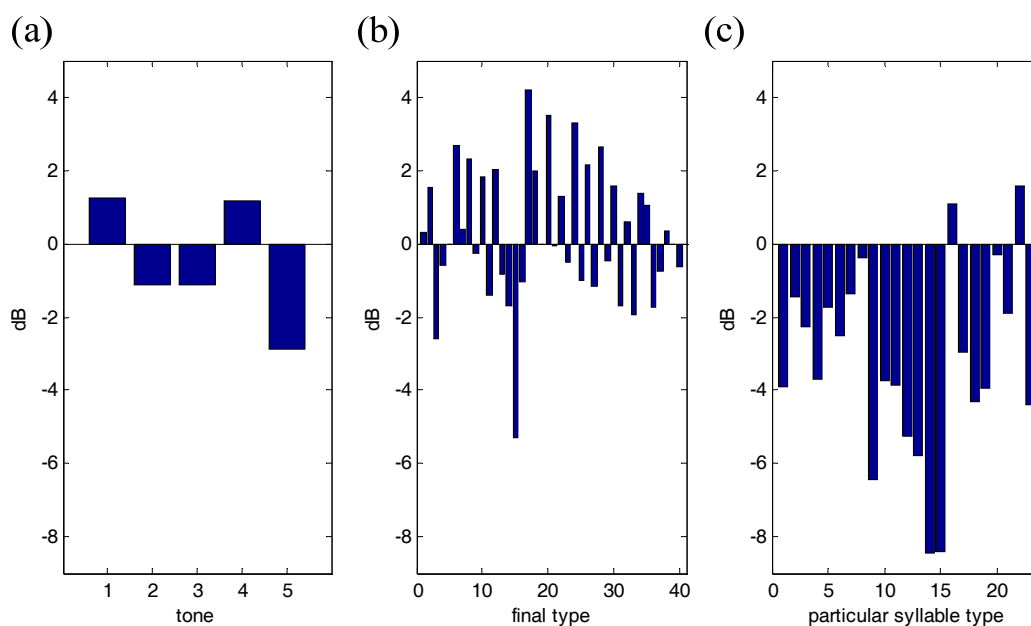


圖 5.10：基本音節中 (a)聲調、 (b) 韻母型態以及以及特殊音節中 (c)特殊音節型態之能量

5.1.2 上層韻律狀態之影響型態

扣除音節之 AP 後，韻律狀態代表韻律上層之貢獻以及變化，在此將對基本音節以及特殊音節之韻律狀態作觀察。

A. 基本音節之韻律狀態

在此觀察基本音節中韻律狀態，對音節基頻軌跡、音節長度以及音節能量之貢獻，如圖 5.11 至 5.13 所示。在此我們可以發現相較於其他影響因子，韻律狀態之動態範圍皆大於其他影響因子，再度顯示韻律之變化大部分是由音節以上的層次所貢獻。

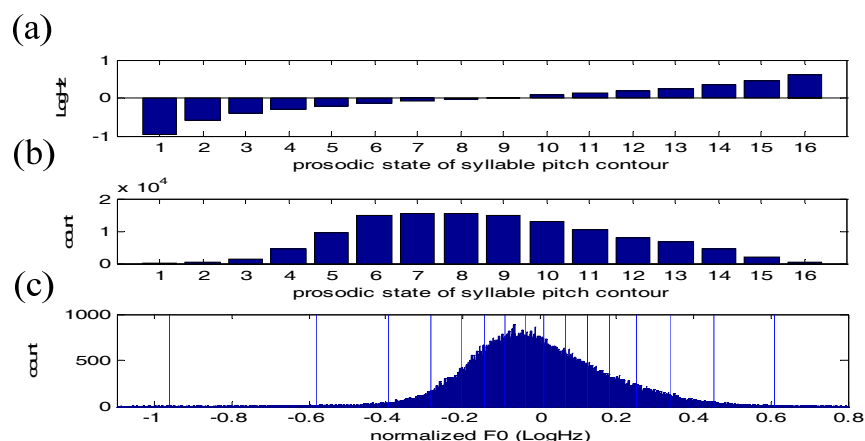


圖 5.11：基本音節中，基頻之 (a)各韻律狀態之數值及 (b)韻律狀態分布圖與 (c)正規化基頻數值分布圖之比較

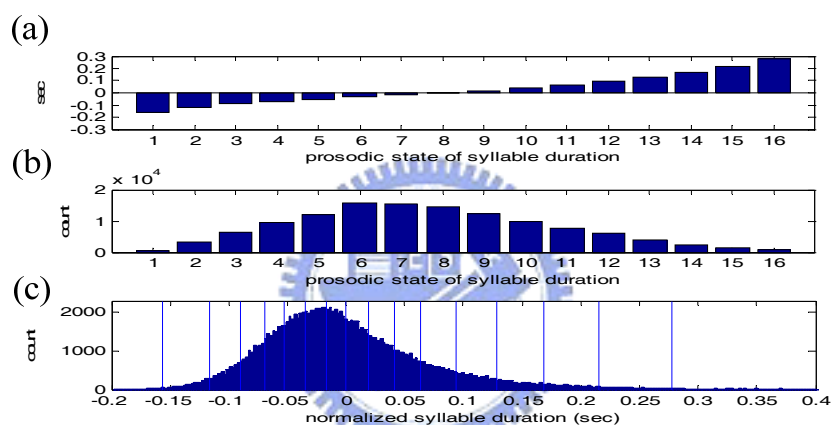


圖 5.12：基本音節中，音節長度之 (a)韻律狀態之數值及 (b)韻律狀態分布圖與 (c)正規化音節長度分布圖之比較

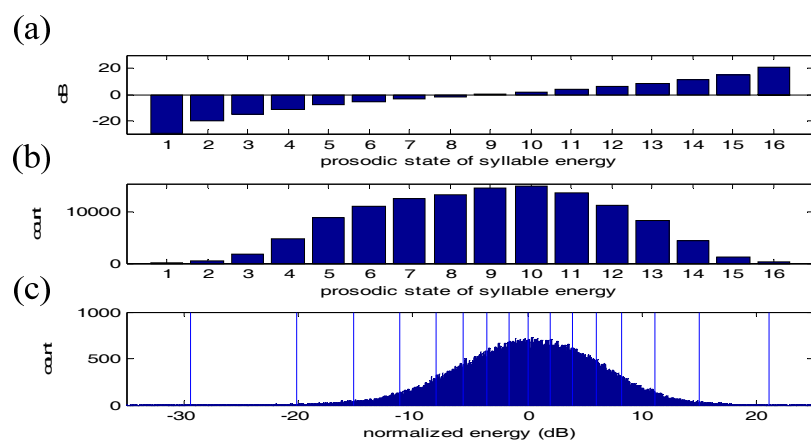


圖 5.13：基本音節中，音節能量之 (a)各韻律狀態之數值及 (b)韻律狀態分布圖與 (c)正規化音節分布圖能量之比較

B. 基本音節與特殊音節韻律狀態之比較

接著比較基本音節以及特殊音節中基頻軌跡、音節長度以及音節能量韻律狀態之數值，如圖 5.14 至 5.16 所示。由圖 5.15 可知雖然特殊音節之音節長度在音節層次之影響程度以及動態範圍和基本音節差不多，但在韻律狀態上卻明顯偏高，因此若不將特殊音節與基本音節隔離的話，將會汙染基本音節韻律狀態的表現，使上層韻律狀態之數值有偏差。

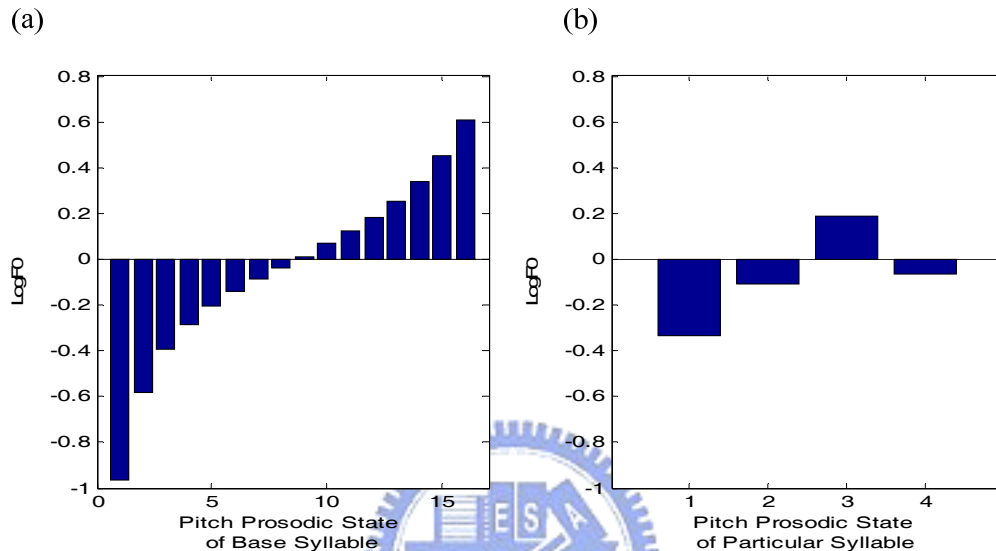


圖 5.14：(a) 基本音節及 (b)特殊音節之基頻韻律狀態數值

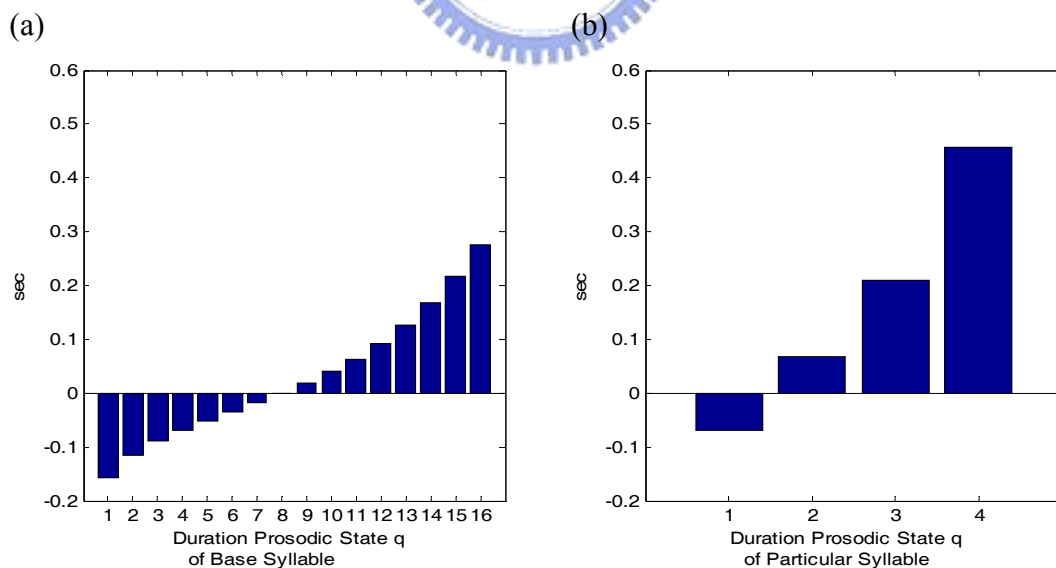


圖 5.15：(a) 基本音節及 (b)特殊音節之音節長度韻律狀態數值

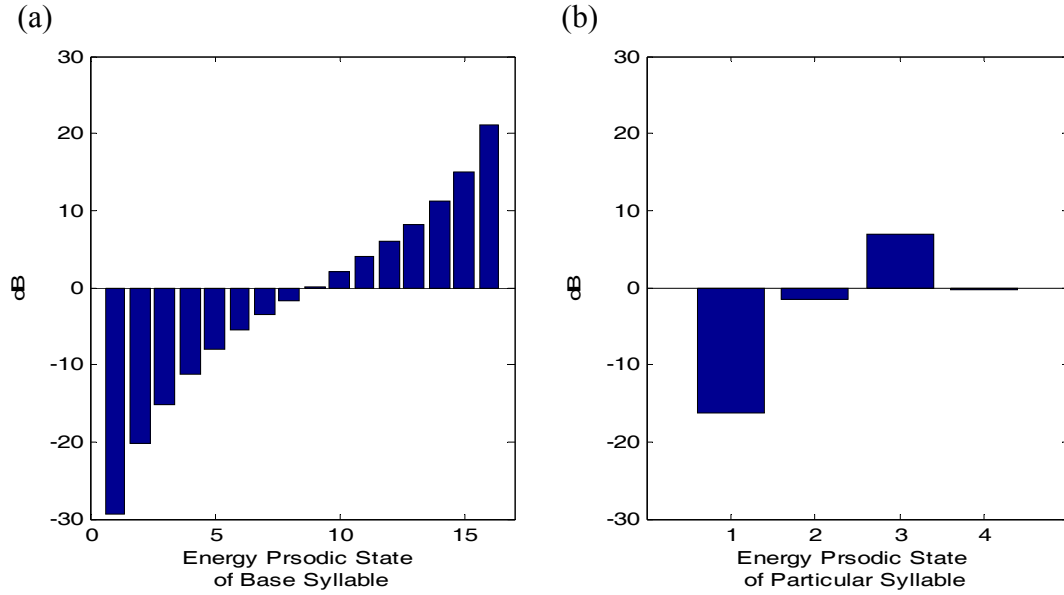


圖 5.16：(a) 基本音節及 (b)特殊音節之能量韻律狀態數值

5.2 停頓標記聲學模型

停頓標記聲學模型主要包含了四個子模型 $g(pd_n; \alpha_{B_n, I_n}, \beta_{B_n, I_n})$ 、 $N(ed_n; \mu_{B_n, I_n}, \sigma_{B_n, I_n}^2)$ 、 $N(pj_n; \mu_{B_n, I_n}, \sigma_{B_n, I_n}^2)$ 以及 $N(dl_n; \mu_{B_n, I_n}, \sigma_{B_n, I_n}^2)$ ，在各個停頓標記的情況下，皆以詢問語言參數 I 的問題產生這四個模型之決策樹，它們描述了音節邊界之停頓標記 B 、語言參數 I 與音節間韻律參數 Y 及相鄰兩音節差異之韻律參數 Z 之間的關係。圖 5.17 顯示各種停頓標記之下，決策樹根節點中各個參數之分布圖，即為不考慮語言參數 I 下，各停頓標記之參數分布。由分布圖可得知用來隔離上層韻律單元 PG/BG 以及 PPh 之停頓標記 $B3$ 、 $B4$ 會有較長的停頓長度、較低的音節間能量低點、較大的音節基頻跳躍值以及較大的正規化音節延長因子。而 $B0$ 與 $B1$ 相較之下，皆擁有非常短的停頓長度，但 $B0$ 有較大的音節間能量低點值，表示 $B0$ 音節邊界相鄰之兩音節為緊密連接。 $B2-2$ 與 $B0/B1$ 以及 $B3/B4$ 相比，其音節停頓長度居中，用以區分 PW 之邊界。而 $B2-1$ 以及 $B2-3$ 之停頓長度與 $B1$ 比較之下相差不多，但 $B2-1$ 擁有較大之基頻跳躍值、 $B2-3$ 有較大之正規化音節延長因子，表示 $B2-3$ 之前一音節都發生了音節拉長的現象。由此可知人類不但使用明顯的停頓表示 PW 邊界，也使用基頻的跳躍以及音節的

拉長來表示之。而這些參數的表現也與本研究先前定義中文語音韻律結構之停頓標記吻合。

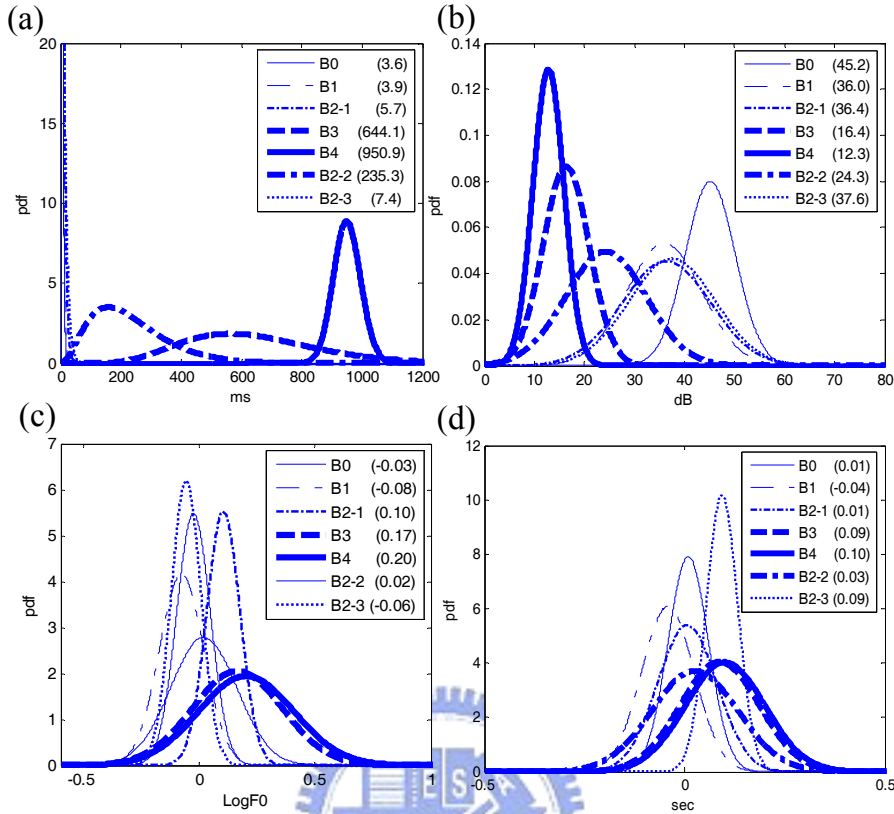


圖 5.17：(a) 音節停頓長度 (b)音節間能量低點 (c)正規化基頻跳躍值與 (d)正規化音節延長因子之分布圖，其中括號中之數值為平均值。

5.3 韻律狀態轉移模型

韻律狀態轉移模型 $P(p_n | p_{n-1}, B_{n-1})$ ，描述在已知停頓標記的狀況之下，前一音節之韻律狀態轉移至現在音節韻律狀態之轉移機率，如圖 5.18 所示。在 $B0$ 以及 $B1$ 的情況下韻律狀態之轉移皆為不變或由高至低，這表示在 PW 之內上層韻律狀態皆由高的層級往下降，但是當音節間標記為 $B0$ 時，由於語速較快且音節間為緊密連接，所以韻律狀態大部份維持不變。又 $B2-1$ 、 $B3$ 以及 $B4$ 之韻律狀態都有上升的現象，這表示人類會使用基頻跳躍來表示 PW、PPH 或 BG/PG 之邊界。

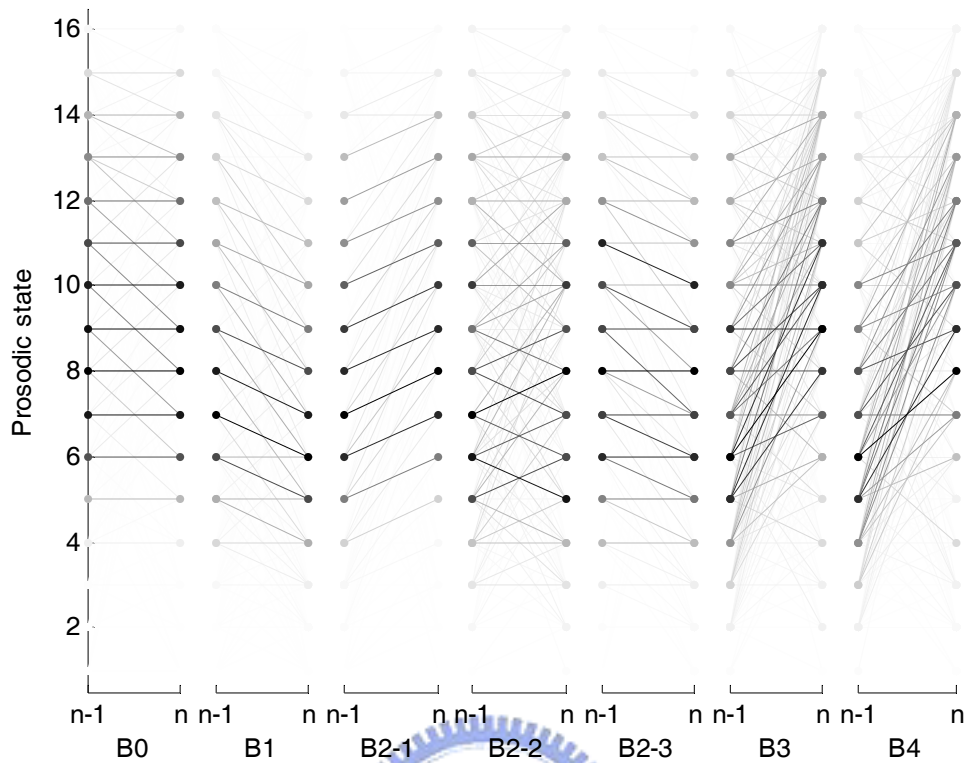


圖 5.18：各停頓標記下，基頻韻律狀態轉移的狀況，顏色越深的線表示此種狀態轉移的機率

接著音節長度韻律狀態轉移模型 $P(q_n | q_{n-1}, B_{n-1})$ 如圖 5.19 所示，在 $B0/B1$ 的情況下，音節長度韻律狀態大部分為上升，但在 $B0$ 的情況下音節長度韻律狀態卻會產生下降的狀況，這表示在自發性語音的 PW 之中，語速大致的走勢為慢至快，但當兩音節為緊密連接時人類通常會將後一個音節的長度縮短以加快發音之速度。而大的停頓標記，如 $B3$ 以及 $B4$ ，音節長度韻律狀態皆大幅的下降，這表示人類會利用一個音節的延長來區分兩段 PPh 或 BG/PG。又 $B2-3$ 之音節長度韻律狀態也明顯的下降，表示在沒有音節停頓時，人類也會用音節的拉長來表示 PW 之邊界。

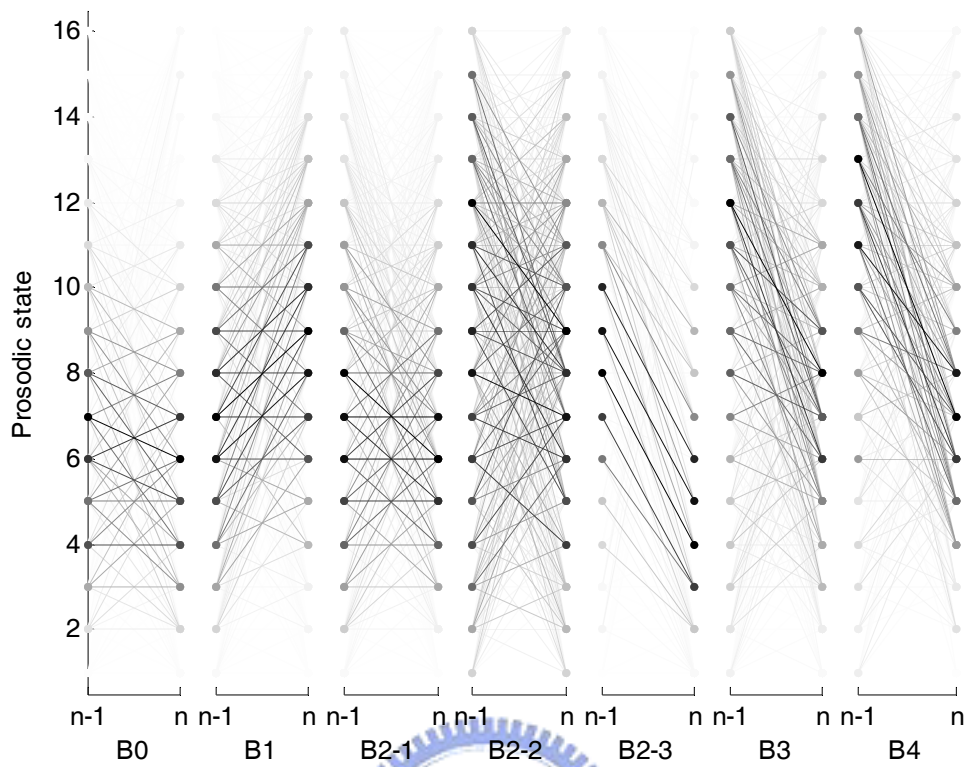


圖 5.19：各停頓標記下，音節長度韻律狀態轉移的狀況，顏色越深的線表示此種狀態轉移的數量越多

最後音節能量韻律狀態轉移模型 $P(r_n | r_{n-1}, B_{n-1})$ 如圖 5.20 所示，在大的停頓標記如 $B3$ 以及 $B4$ ，音節能量韻律狀態皆大幅的上升，這表示在一個 PPh 或 BG/PG 的單元邊界，會產生大的能量跳躍，且 PPh 或 BG/PG 中的能量走勢為由高至低。

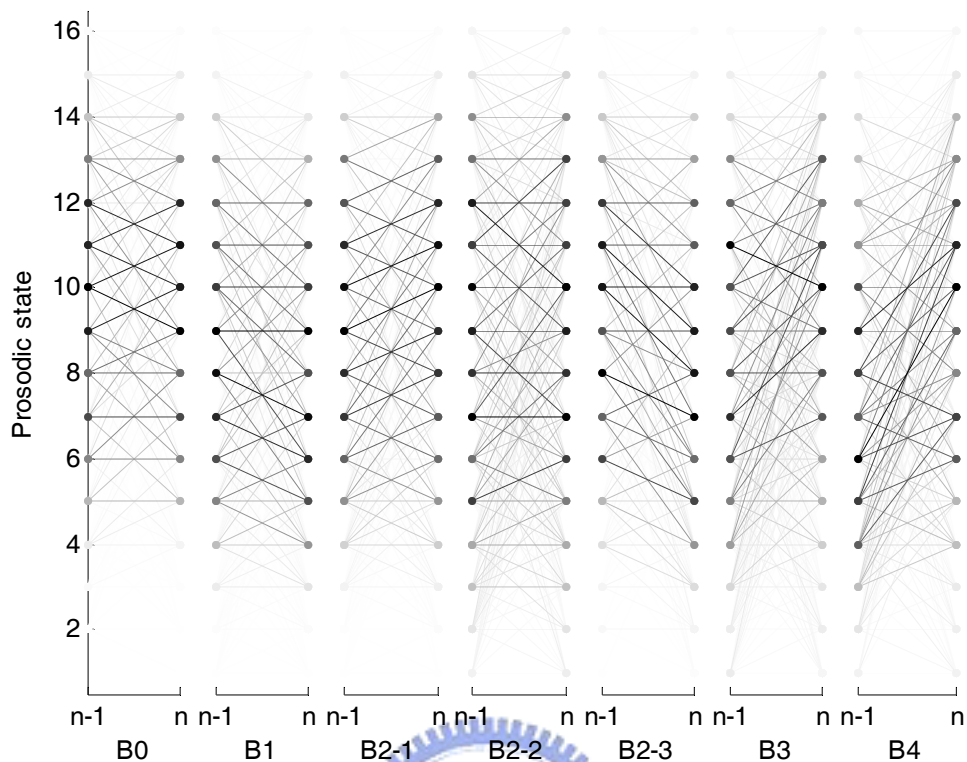


圖 5.20：各停頓標記下，音節能量韻律狀態轉移的狀況，顏色越深的線表示此種狀態轉移的數量越多

5.4 停頓標記語言模型

停頓標記語言模型描述了各種語言參數 \mathbf{l} 組合的情況之下，停頓標記 \mathbf{B} 的分布情況。本研究經由詢問語言參數之問題來實現此決策樹。決策樹之示意圖如圖 5.21 所示，其中停頓標記分布圖之橫軸由左至右分別為 $B0$ 、 $B1$ 、 $B2-1$ 、 $B2-2$ 、 $B2-3$ 、 $B3$ 以及 $B4$ ，停頓標記語言模型之決策樹如圖 5.22 所示。此決策樹節點 1 至節點 3 皆為將特殊音節與正常音節邊界隔離的問題。由節點 7 開始觀察正常音節之邊界，若為詞內音節邊界時，節點 9 及其往下之所有子節點大部分皆標示為 $B0$ 或 $B1$ ，這表示人類在唸一個詞時，幾乎都不會在詞內作一個明顯的韻律停頓。

接著觀察詞間音節邊界往下之節點，如圖 5.23 所示，首先發現節點 10 以及節點 17，停頓標記皆以 $B3$ 、 $B4$ 或 $B2-1$ 為主，且音節邊界後所跟隨的詞類皆為關係連接詞，表示在自發

性語音中，人類在利用關係連接詞連接兩段語句之前通常都會有韻律停頓的狀況。以下為一個標記後的範例：

範例一：

… 不是(Na) 興趣(Na) 不(D) 興趣(Na) 的(DE) 問題(Na) **B2-1** 而是(Cbb) 說(VE) 像(VG) 這個(DM) 東西(Na) …

範例二：

… 歡迎(VJ) 光臨(VCL) **B2-1** 紐約(Nc) 紐約(Nc) A(Par) 什麼(Nep) **B3** 可是(Cbb) 常常(D) 我(Nh) **B2-1** 就是(Cbb) 很(Dfa) 想(VE) 去(VCL) **B2-1** 那邊(Ncd) …

觀察節點 22 及 34，發現音節邊界大多為 **B2-1**，且音節邊界前一個詞或後一個詞皆為代名詞，表示在自發性語音中，人類會在代名詞的前後來強調此代名詞或代名詞後所接的語句。以下為一個標記後的範例：

範例一：

… 退伍(VH) 兩年(Nd) **B2-1** 渾渾噩噩(VH) **B2-1** 你(Nh) 看(VE) 兩年(Nd) 已(D) 經過(VCL) 了(Di) …

範例二：

… 我(Nh) **B2-1** 就(D) 覺得(VK) **B2-1** 不(D) 值得(VH) 我(Nh) **B2-1** 就(D) 不會(D) 想(VE) 去(D) 玩(VC) …

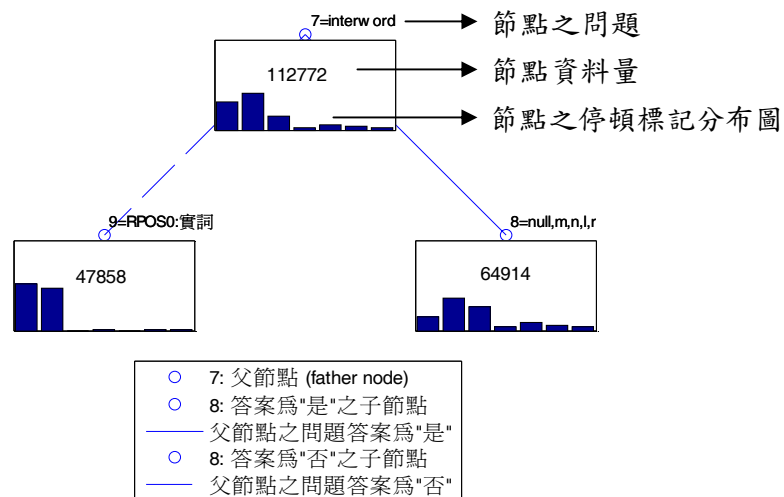


圖 5.21：停頓標記語言模型決策樹之示意圖，其中停頓標記分布圖之橫軸由左至右分別為

B0、*B1*、*B2-1*、*B2-2*、*B2-3*、*B3* 以及 *B4*

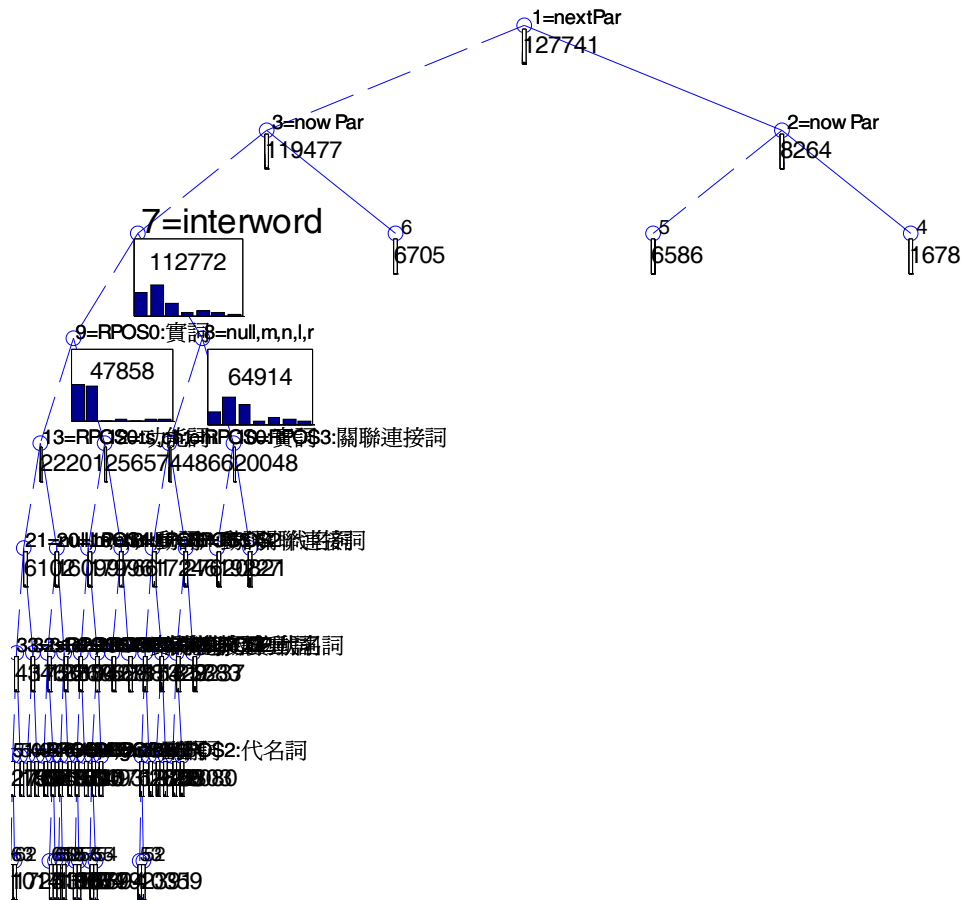


圖 5.22：停頓標記語言模型之決策樹。

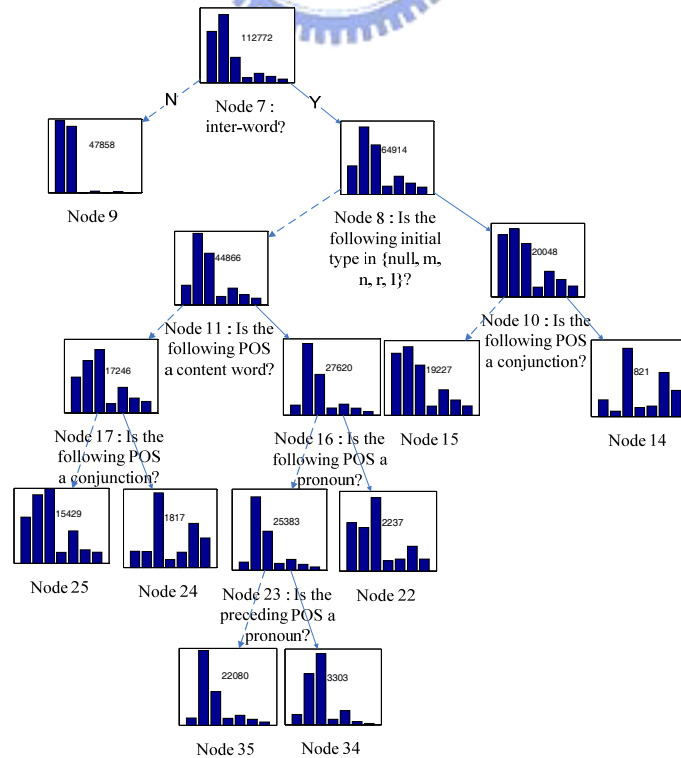


圖 5.23：詞間音節邊界停頓標記語言模型之決策樹

5.5 韻律標記結果之分析

觀察各個韻律模型參數後，接著將觀察語音中韻律標記之狀況，首先 5.5.1 將觀察停頓標記之分布情況；5.5.2 節將觀察上層韻律單元內之變化情形；5.5.3 將展示一個韻律標記之範例。

5.5.1 停頓標記及韻律狀態之分析

圖 5.24 為停頓標記之分布圖，其中圖 5.24 (a) 為所有音節邊界對應之停頓標記分佈圖，而圖 5.24 (b) 為去除用以隔離特殊音節之停頓標記後，停頓標記的分布圖，與朗讀式語音【7】相比 *B2-1*、*B2-2* 以及 *B2-3* 出現的比率差不多，但 *B2-1* 所佔之比率上升且 *B2-2* 所佔之比率下降⁷，表示人類在語速較快的自發性語音中，在 PW 邊界時，較少產生音節間的停頓，取而代之的是使用基頻的跳躍。

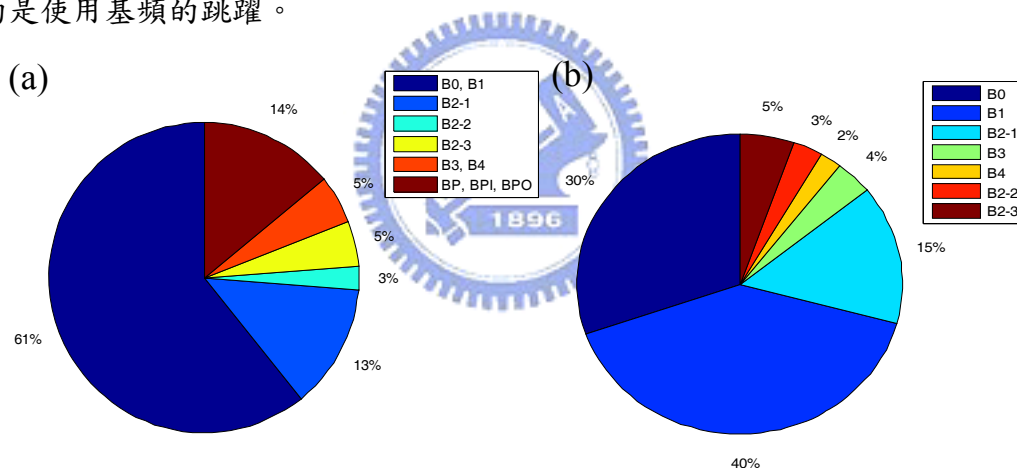


圖 5.24：(a) 所有音節邊界之停頓標記分布圖以及 (b) 排除特殊音節邊界之停頓標記分布圖

5.5.2 上層韻律單元內之變化情況

如圖 4.2 所定義之中文語音韻律階層式架構，有了停頓標記的狀況下，可將語音之韻律分為各層級韻律單元之組合疊加，也就是 PW、PPh 以及 BG/PG 對韻律之貢獻，首先表 5.3

⁷ 在此語料庫中 *B2-1*、*B2-2* 以及 *B2-3* 出現之機率為 23%，而在朗讀式語音中為 22%；此語料庫出現 *B2-1* 之比率為 15%，在朗讀式語音中為 10%；此語料庫出現 *B2-2* 之比率為 3%，在朗讀式語音中為 6%。

統計了各種韻律單元之音節長度，在此發現 PG/BG 之長度較朗讀式語音短⁸【7】，在此推測原因有二，其一為本研究設計之基本架構中包含特殊韻律單元，如圖 4.2 所示，因此若原本完整的 PG/BG 單元很可能會被 Par 所隔開，分裂為兩個 PG/BG，因此造成 PG/BG 之長度較短；其二為 MCDC 語料為一自發性對話語料，語音很有可能因為對方的打斷或語者本身的思考及遲疑產生長停頓，標示出 B3 或 B4，形成一個短的 PG/BG，使語句較短，如圖 2.6 所示。但在鄭秋豫博士之中文語音韻律階層式架構【20】中的定義，DM/PF 是扮演連接 PPh 的腳色，而本研究 Par 大部分的組成份子為 DM/PF，所以 PPh 與 PW 大部分並不會為 Par 所隔開，因此長度與朗讀式語音差不多。

表 5.3：不同韻律單元之音節長度

PW		PPh		PG/BG	
mean	std	mean	std	mean	std
2.55	1.92	5.23	5.89	6.27	8.74

接著本研究利用扣除總體平均值、聲調以及基本音節型態 AP 之基頻、音節長度以及能量殘存值 $sp2_n$ 、 $sd3_n$ 以及 $se3_n$ ，觀察 PW、PPh 以及 BG/PG 內韻律之變化，由於這些扣除音節層次 AP 之殘存值包含 PW、PPh 以及 BG/PG 這些上層韻律狀態之貢獻，因此可將 PW、PPh 以及 BG/PG 仿效音節韻律模型之作法，首先將各個韻律單元依長度分類分類為

$$PW_n=(i,j) \quad (5-1)$$

$$PPh_n=(i,j) \quad (5-2)$$

$$BG/PG_n=(i,j) \quad (5-3)$$

分別代表音節 n 位於一個具有 i 個音節之 PW、PPh 以及 BG/PG 的第 j 個位置，以表示各種長度之韻律單元，且觀察不同韻律單元且不同長度之下之 AP，定義如下：

$$sp2_n=pm_n^r+\beta_{PW_n}+\beta_{PPh_n}+\beta_{BG/PG_n} \quad (5-4)$$

$$sd3_n=dm_n^r+\gamma_{PW_n}+\gamma_{PPh_n}+\gamma_{BG/PG_n} \quad (5-5)$$

$$se3_n=em_n^r+\alpha_{PW_n}+\alpha_{PPh_n}+\alpha_{BG/PG_n} \quad (5-6)$$

⁸ 朗讀式語料中，PW 之平均長度為 2.80 個音節；PPh 之平均長度為 7.46 個音節；PG/BG 之平均長度為 16.85 個音節。

其中 \mathbf{pm}_n^r , dm_n^r 和 em_n^r 分別表示第 n 個音節扣除音節層次以及各韻律單元 AP 的基頻軌跡、音節長度以及能量之殘存值，並且根據 MMSE (Minimum Mean Square Error) 法則訓練這三個模型，使其誤差函數達到收斂，本研究定義之基頻軌跡、音節長度和音節能量之誤差函數如下：

$$E_p = \sum_{n=1}^N \left| \mathbf{sp}2_n - \beta_{PW_n} - \beta_{PPh_n} - \beta_{BG/PG_n} \right|^2 \quad (5-7)$$

$$E_d = \sum_{n=1}^N \left(sd3_n - \gamma_{PW_n} - \gamma_{PPh_n} - \gamma_{BG/PG_n} \right)^2 \quad (5-8)$$

$$E_e = \sum_{n=1}^N \left(se3_n - \alpha_{PW_n} - \alpha_{PPh_n} - \alpha_{BG/PG_n} \right)^2 \quad (5-9)$$

經由以上處理後，可獲得各種韻律單元在不同長度之下，基頻、音節長度以及能量之影響形態，分別如圖 5.25、5.26 以及 5.27 所示。

首先觀察基頻在各種韻律單元以及不同長度的狀況下之 AP，如圖 5.25 所示，發現在 PG/BG 這一層相較於朗讀式語音【7】由高至低的趨勢並不明顯，且動態範圍較朗讀式語音小，其原因可能為 PG/BG 被切割所致，造成其求取之 AP 不準確。而 PPh 以及 PW 這兩層之趨勢皆是由高至低迅速下降，並無朗讀式語音中長的 PPh 會有「升高-下降-維持-下降」(rasing-falling-sustaining-falling)或長的 PW 會有「下降-維持-下降」(falling-sustaining-falling)之趨勢，表示人類在自發性語音中，腦中構思 (speech planning) 與發音的過程 (speech production) 是同時進行，韻律的規劃是較為局部性且迅速的，相較於朗讀式語音較無長遠之韻律規劃。

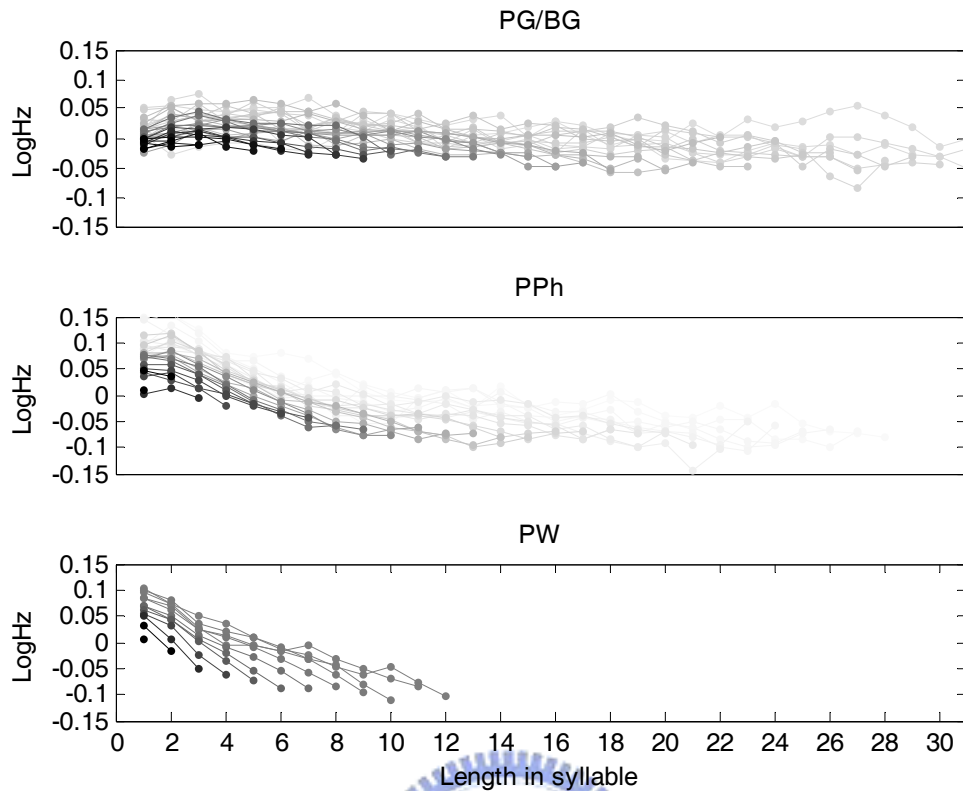


圖 5.25：各種韻律單元在不同長度下之基頻影響型態，顏色越深的線表示此種長度之韻律單元數量越多

接著觀察音節長度在各種韻律單元以及不同長度的狀況下之 AP，如圖 5.26 所示，發現在 PPh 及 PW 中，音節長度的趨勢皆為「起頭長-縮短-節尾拉長」和「起頭短-節尾拉長」，再度證實人類會以句尾的拉長來表示一個 PPh 以及 PW 的結束，且相較於朗讀式語音結尾音節拉長之現象更為嚴重。而與朗讀式語音【7】相同的是，節尾拉長處皆位於 PPh 以及 PW 之倒數第三個音節開始發生，但在自發性語音當中倒數第三個音節並無發生音節縮短的現象。

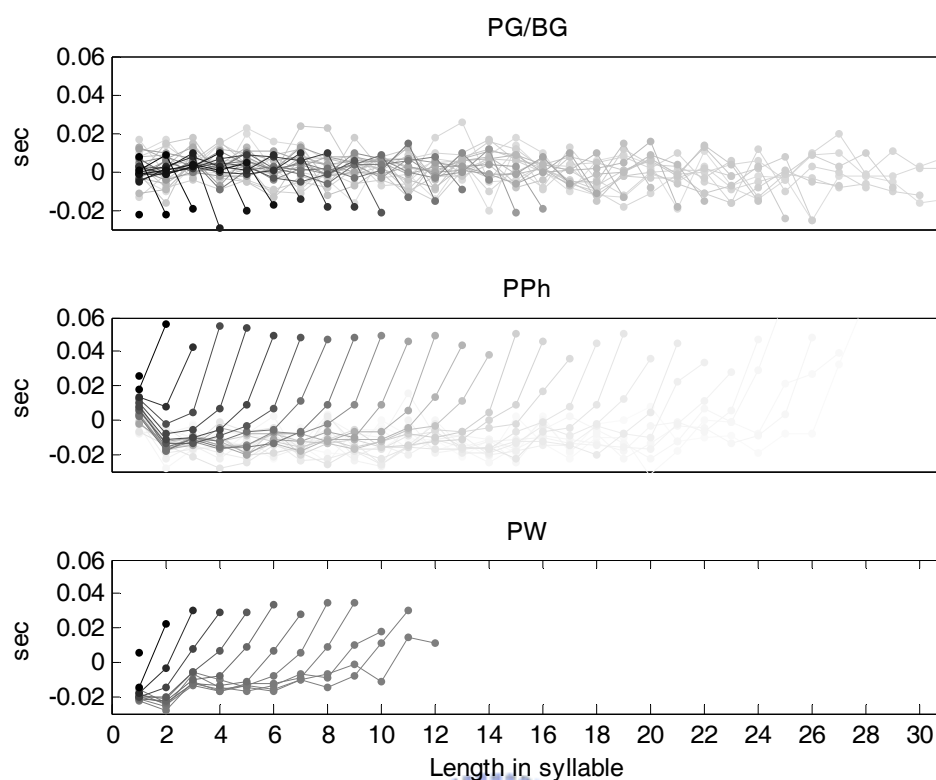


圖 5.26：各種韻律單元在不同長度下之音節長度影響型態，顏色越深的線表示此種長度之韻律單元數量越多

圖 5.27 為能量在各種韻律單元以及不同長度的狀況下之 AP，發現在長的 PPh 會有「起頭升高-下降-維持-下降」(rasing- falling-sustaining-falling) 之趨勢，和朗讀式語音【7】之「下降-維持-下降」(falling-sustaining-falling) 相比，起頭會有一個能量加強的現象。此外短的 PPh 以及 PW 之趨勢則是由高至低迅速下降，相較於朗讀式語音【7】之 PW，節尾並無能量拉高的現象。

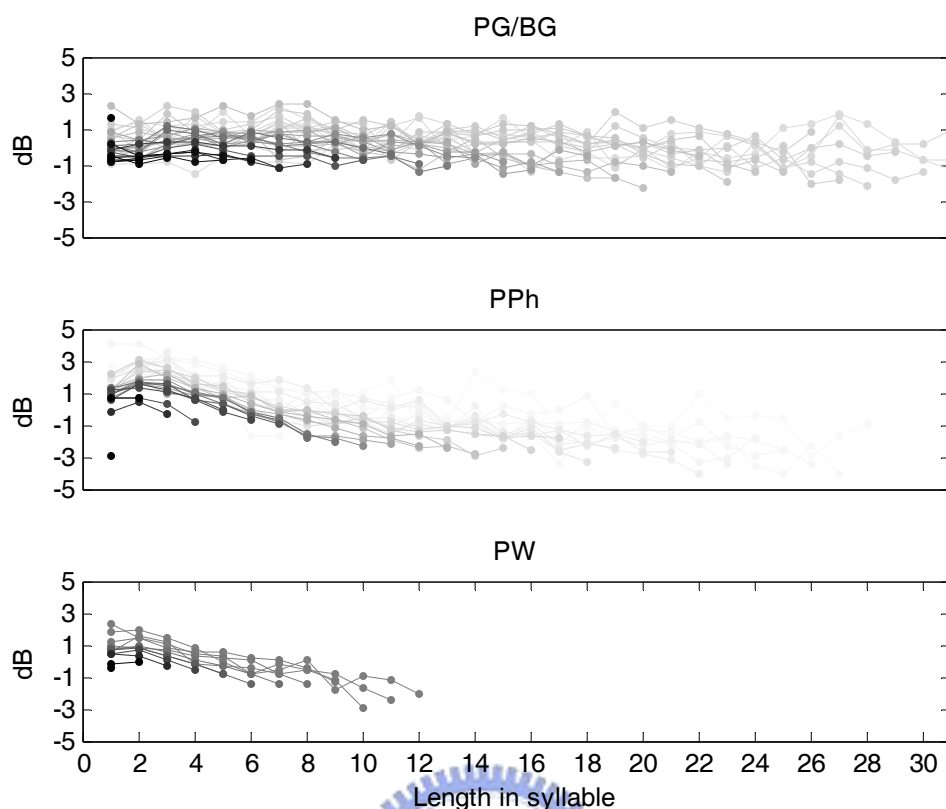


圖 5.27：各種韻律單元在不同長度下之能量影響型態，顏色越深的線表示此種長度之韻律單元數量越多

最後利用 PG/BG、PPh 以及 PW 之總殘餘誤差值，藉此觀察 PG/BG、PPh 以及 PW 與各個影響因子對音節中韻律參數變化之貢獻大小，由表 5.4 可以得知音節中基頻軌跡的變化貢獻由大至小分別為：韻律狀態、PPh、包含連音現象之聲調、PG/BG 以及 PW；而對音節長度的變化貢獻由大至小分別為：韻律狀態、基本音節型態、PW、包含連音現象之聲調、PPh 以及 PG/BG；能量的變化貢獻由大至小分別為：韻律狀態、基本音節型態、包含連音現象之聲調、PPh、PG/BG 以及 PW。由此可知相較於音節層次之貢獻，基頻由 PW、PPh 這些韻律上層之貢獻較大，而音節長度以及能量則是以音節層次之貢獻較大，在此 PG/BG 之貢獻非常小，可能是因為被 Par 切割以致無法良好描述 PG/BG 之趨勢。

表 5.4：不同組合之 AP 下之 TRE

Pitch		Duration		Energy	
APs	TRE	APs	TRE	APs	TRE
+Tone with Coarticulation	90.61%	+Tone with Coarticulation	93.95%	+Tone with Coarticulation	94.27%
		+Base Syllable	83.70%	+Final	86.80%
+PW	88.84%	+PW	74.20%	+PW	84.57%
+PPh	78.51%	+PPh	69.22%	+PPh	79.12%
+PG/BG	75.68%	+PG/BG	66.79%	+PG/BG	76.24%
+Prsodic State	14.11%	+Prsodic State	1.68%	+Prsodic State	2.54%

5.5.3 韻律標記範例

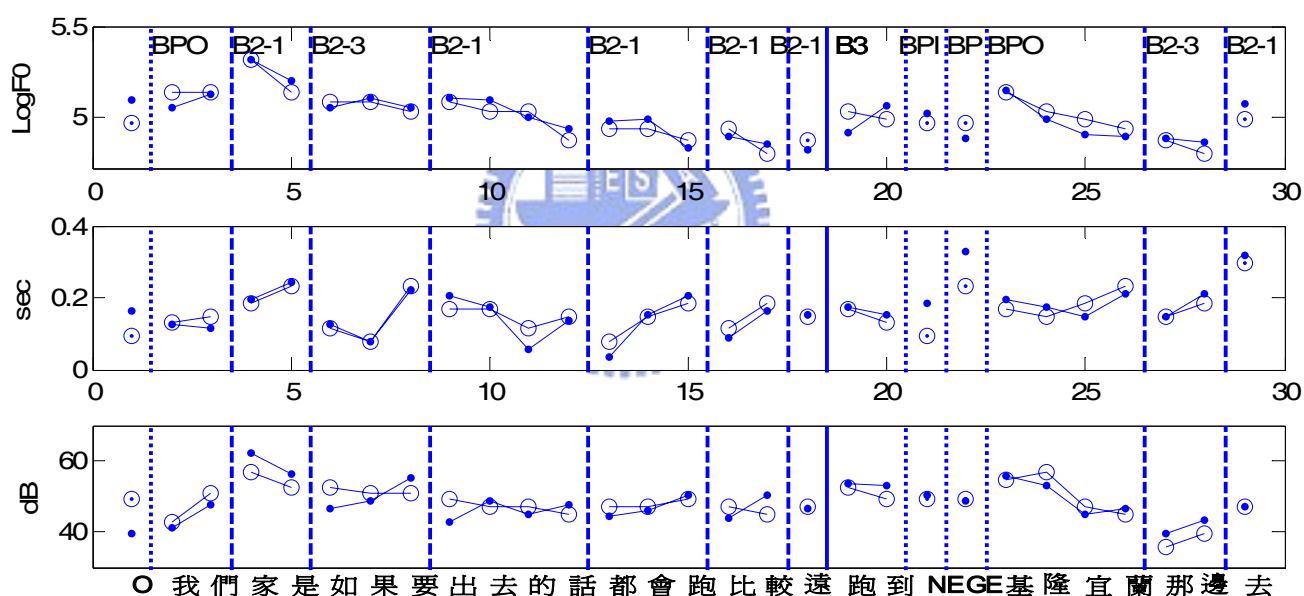


圖 5.28：韻律標記範例：實線代表 PPh 或 PG/BG 之邊界，虛線代表 PW 之邊界，點線代表隔離特殊韻律現象之邊界；由上至下為基頻，音節長度以及能量之韻律狀態標記，空心圓表示各參數韻律狀態加上總體平均值之數值，而實心圓為實際獲得的參數值

圖 5.28 為一個韻律標記之範例，首先在「我們」之後，有一個基頻跳躍的 B2-1，這與停頓標記語言模型中，在代名詞之後常出現 B2-1 之結果互相呼應；而在「家是」之後，由於語者正在遲疑，而產生一個音節延長，以維持發話權；接著語者按照固定的節奏一直到「比

較遠」，並以長的停頓結束此 PPh；在下一段 PPh 中，語者講完「跑到」之後因為思考而加入了語助詞，接著為了強調「基隆宜蘭」而產生基頻迅速下降之重音現象（pitch accent）；最後此語者在大的韻律段落皆習慣以基頻的跳躍做最後一個音節之結束。

5.6 韻律標記結果與詞語修補之討論

在本節當中將分析自發性語音中，詞語修補現象與韻律標記對應之分析，首先觀察詞語修補中非常重要的訊息 IP 相對於停頓標記之狀況，圖 5.29 顯示各種型態的詞語修補 IP 相對應之停頓標記分布圖，由此圖對應所有音節邊界停頓標記之分布(圖 5.24)可以發現，音節邊界出現韻律停頓之停頓標記 *B2-1*、*B2-2*、*B2-3*、*B3* 以及 *B4* 的機率，在 IP 發生時皆較所有音節邊界發生之機率高，尤其又以部份重覆以及詞語更正更為顯著，表示人類在發生詞語修補時，韻律上都會出現停頓的行為。而經由觀察 MCDC 語料庫，重覆之型態又可分為兩類，第一種為類似部分重覆之重覆型態，例如：「也許也許將來」；第二種為單純回應或表達狀態的語用功能之重覆型態，例如：「對對對對對」，但 MCDC 語料庫並未細分此兩類重覆，而經由非正式統計之觀察，發現未出現韻律停頓之重覆 IP，大部分發生在單純回應或表達狀態之重覆，如圖 5.30 所示，因此使重覆中之 *B0*、*B1* 數量較多。

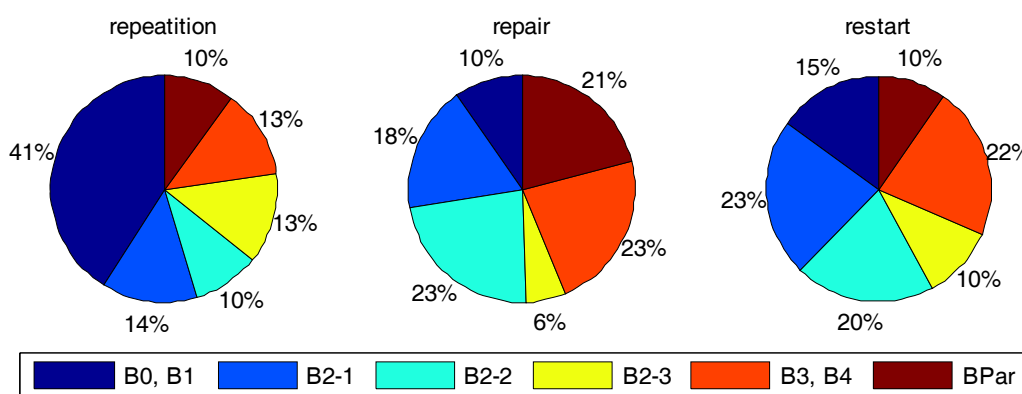


圖 5.29：各型態 IP 對應之停頓標記分布圖

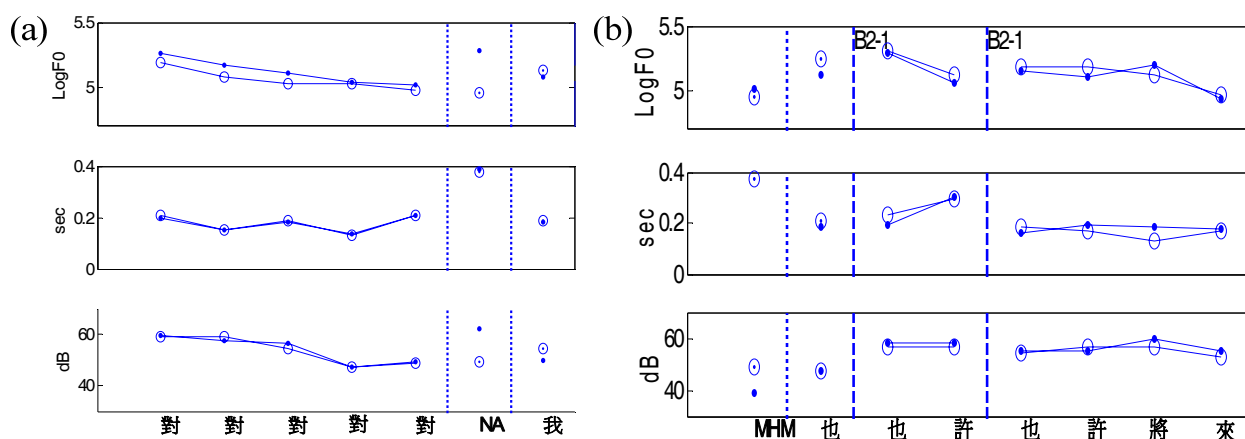


圖 5.30：(a)語用功能之重覆與 (b)類似部分重覆之重覆型態範例：虛線代表 PW 之邊界，點線代表隔離特殊韻律現象之邊界；由上至下為基頻(LogHz)，音節長度(sec)以及能量(dB)之韻律狀態標記，空心圓表示各參數韻律狀態加上總體平均值之數值，而實心圓為實際獲得的參數值

由於詞語修補為一個範圍內出現之現象，且此範圍經由 IP 切割為被更正詞語以及更正詞語，接著本研究利用扣除總體平均值、聲調以及基本音節等音節層次 AP 之基頻、音節長度以及能量殘存值 $sp2_n$ 、 $sd3_n$ 以及 $se3_n$ 之平均，觀察不同型態的詞語修補中，每一個切割範圍內不同長度下韻律上層之變化情況，如圖 5.31 至 5.39 所示。

首先觀察詞語修補中的重覆現象，其基頻、音節長度以及能量之韻律上層變化如圖 5.31 至 5.33 所示。在此分割區段大於三段之重覆，大多對應至單純回應或表達狀態之詞語重覆，觀察得知，這些單純回應或表達狀態之詞語重覆，基頻與能量之趨勢皆由高往低的位階，例如先前所舉的範例：「對對對對對」；而分割範圍為兩段且長度不為 1 之重覆，大部分為類似部分重覆之重覆型態，例如先前所舉的範例：「也許也許將來」，在這種狀況下，其每段分割中之基頻皆為下降的趨勢，但在更正詞語之開頭，基頻值會被拉回至被更正詞語開頭之準位點，且更正詞語中之音節長度及能量皆較被更正詞語小，這與曾淑娟博士在【26】中所發現之現象吻合。

接著觀察部分重覆以及詞語更正兩個型態較為相近的詞語修補，首先觀察音節長度之變化，如圖 5.35 以及圖 5.38 所示，發現這兩種型態的詞語修補中，更正詞語之音節長度皆較

被更正詞語短，也與曾淑娟博士之發現相同，並且被更正詞語中之最後一個音節大部分都發生音節延長的情況。但不同的是在更正詞語中，基頻以及能量之初始值會被拉到較被更正詞語初始值還要高的準位，以此強調更正詞語，並修正被更正詞語。經由以上這些 IP 所對應之停頓標記分布，以及詞語修補中基頻、音節長度以及能量變化之分析，可以更了解詞語修補時韻律之變化，進而將這些訊息利用到偵測詞語修補 IP 以及更正詞語修補之語音自動轉寫的應用中。



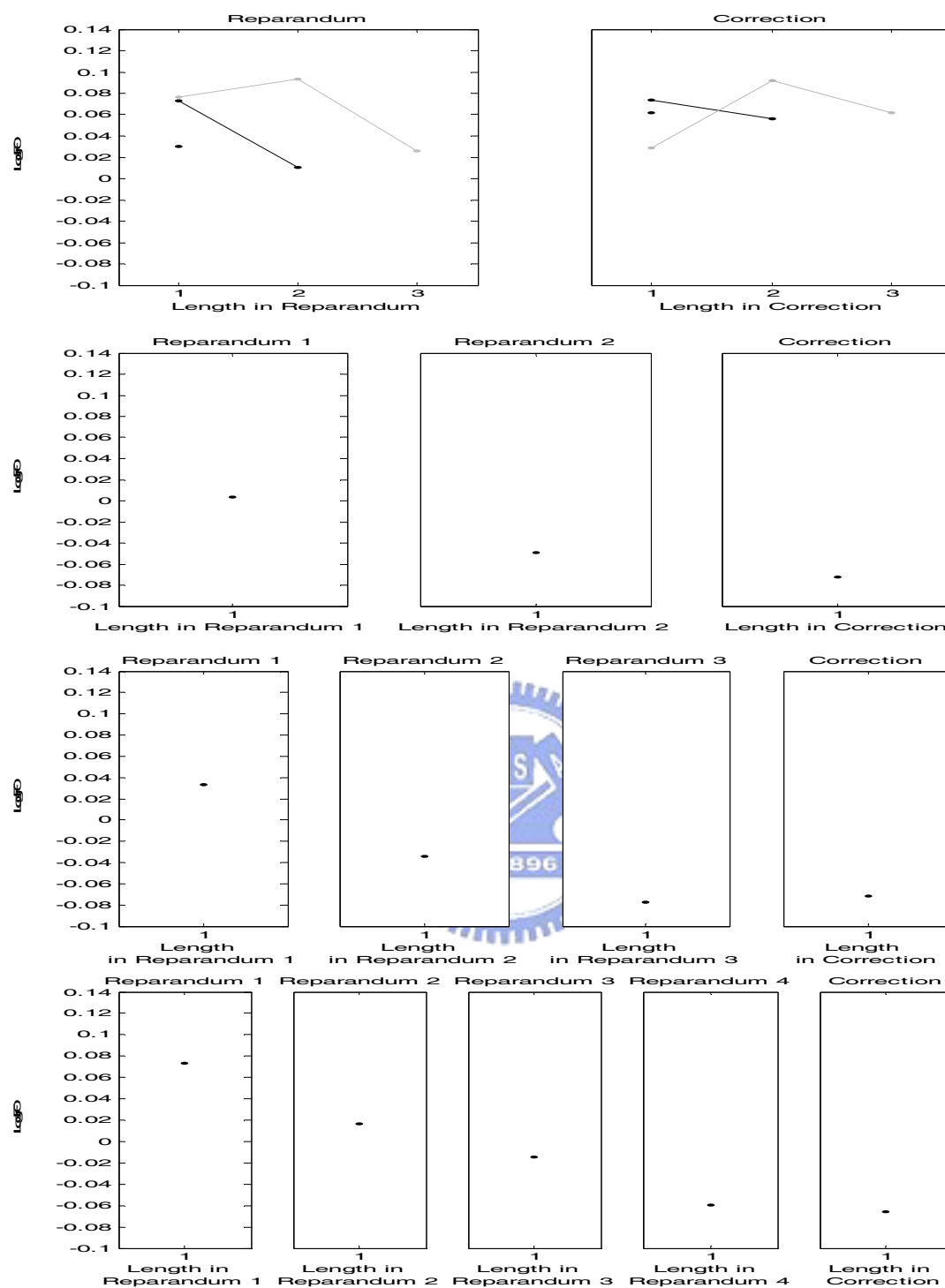


圖 5.31：詞語重覆 (a)兩段分割、(b)三段分割、(c)四段分割與 (d)五段分割中，不同長度下，基頻之變化，顏色越深的線表示分割中此種長度之數量越多

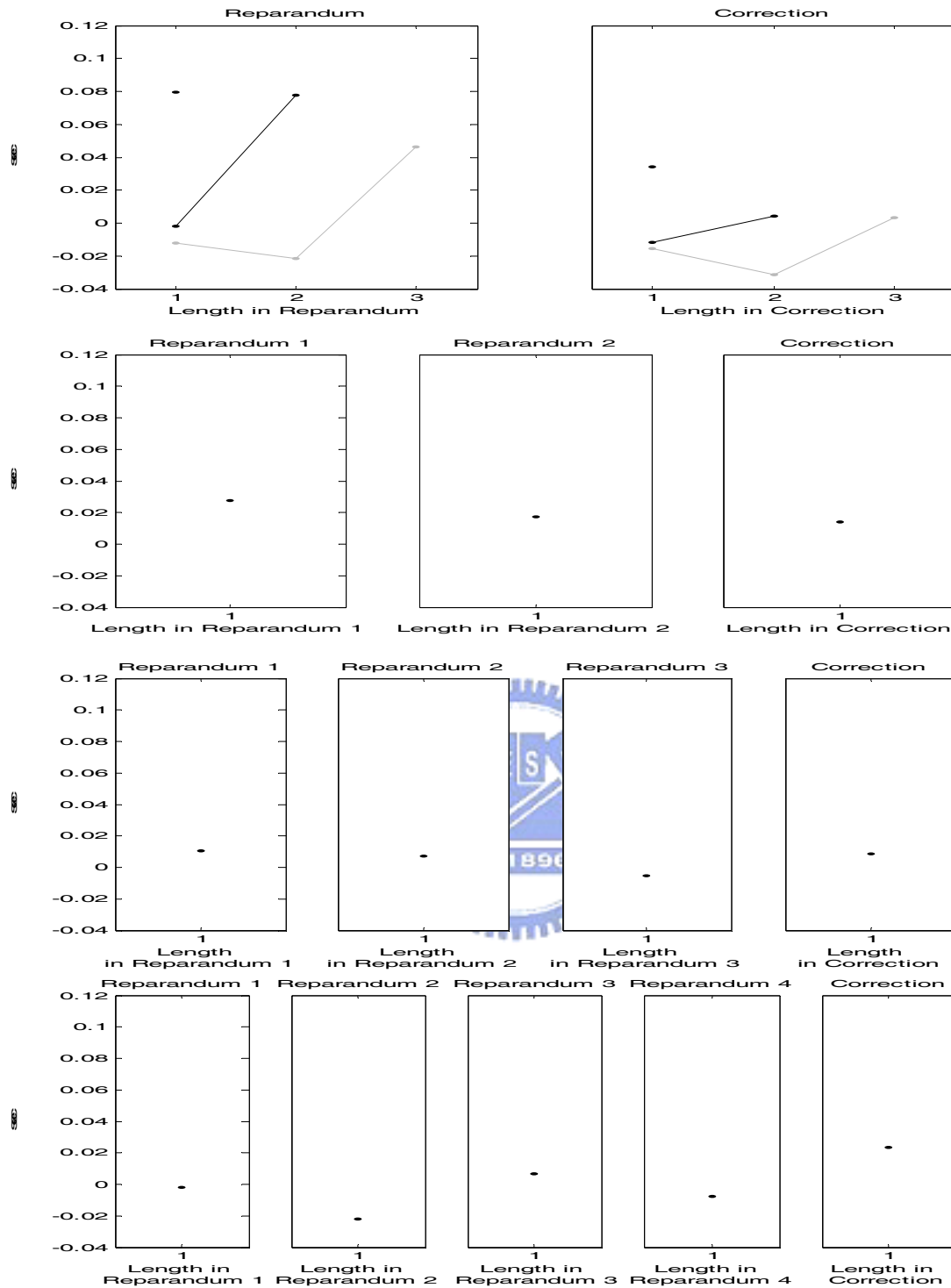


圖 5.32：詞語重覆 (a)兩段分割、(b)三段分割、(c)四段分割與 (d)五段分割中，不同長度下，音節長度之變化，顏色越深的線表示分割中此種長度之數量越多

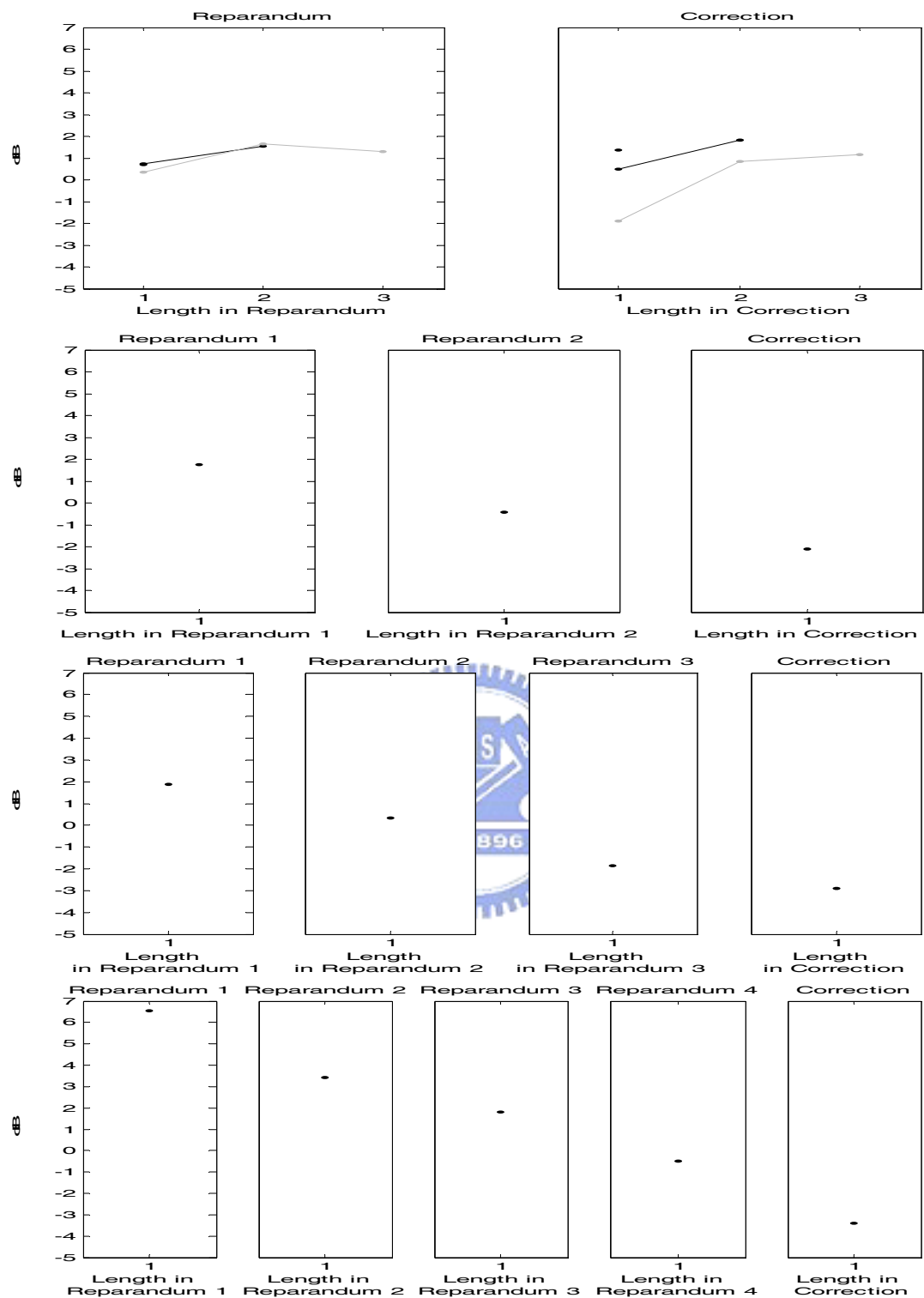


圖 5.33：詞語重覆 (a)兩段分割、(b)三段分割、(c)四段分割與 (d)五段分割中，不同長度下，能量之變化，顏色越深的線表示分割中此種長度之數量越多

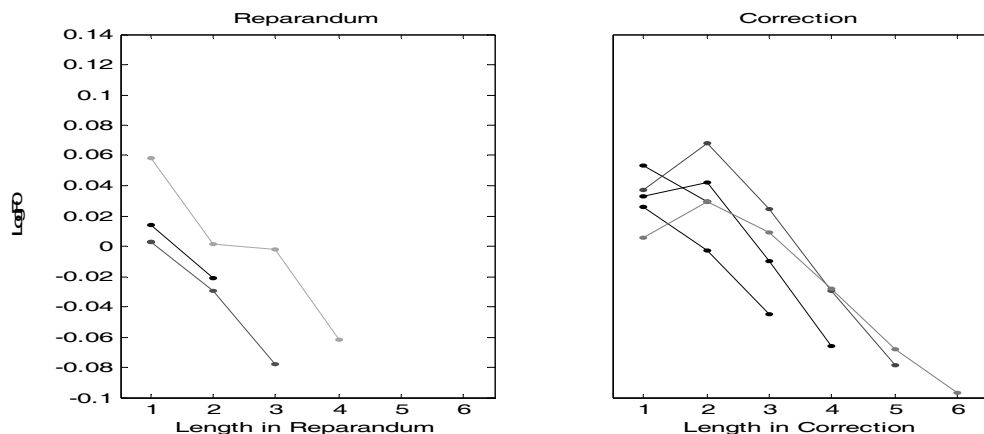


圖 5.34：部分重覆中，兩段分割在不同長度下，基頻之變化，顏色越深的線表示分割中此種長度之數量越多

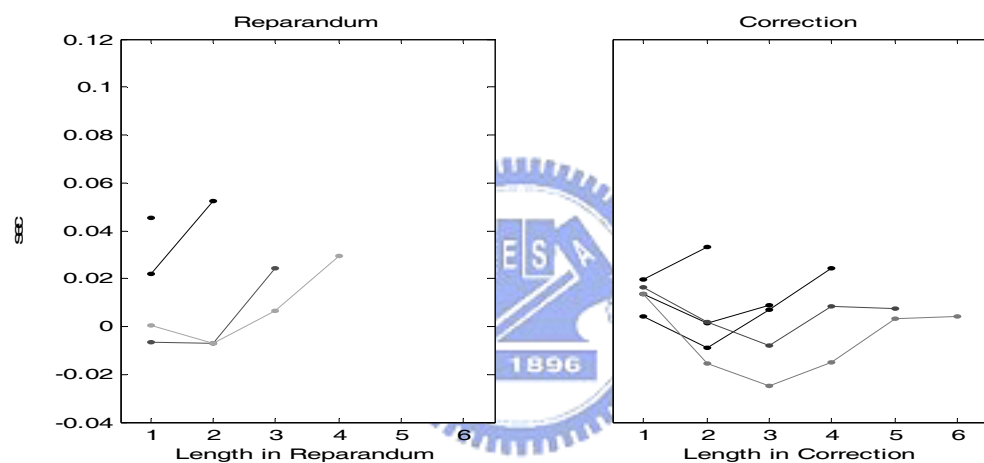


圖 5.35：部分重覆中，兩段分割在不同長度下，音節長度之變化，顏色越深的線表示分割中此種長度之數量越多

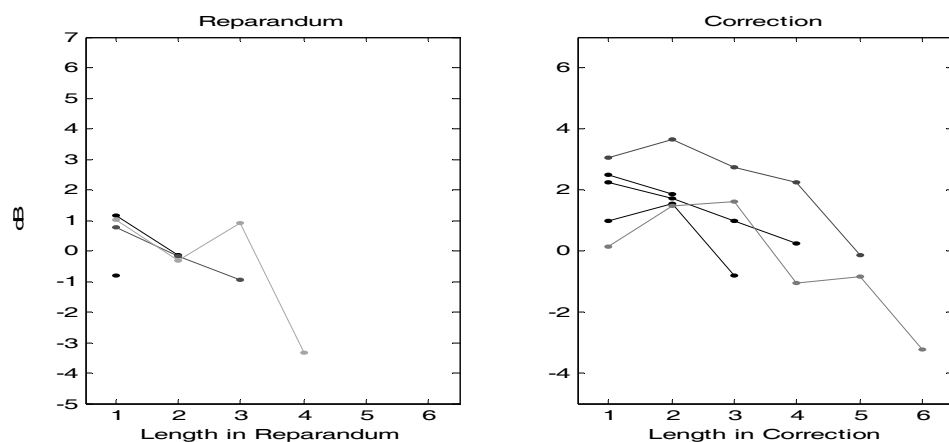


圖 5.36：部分重覆中，兩段分割在不同長度下，能量之變化，顏色越深的線表示分割中此種長度之數量越多

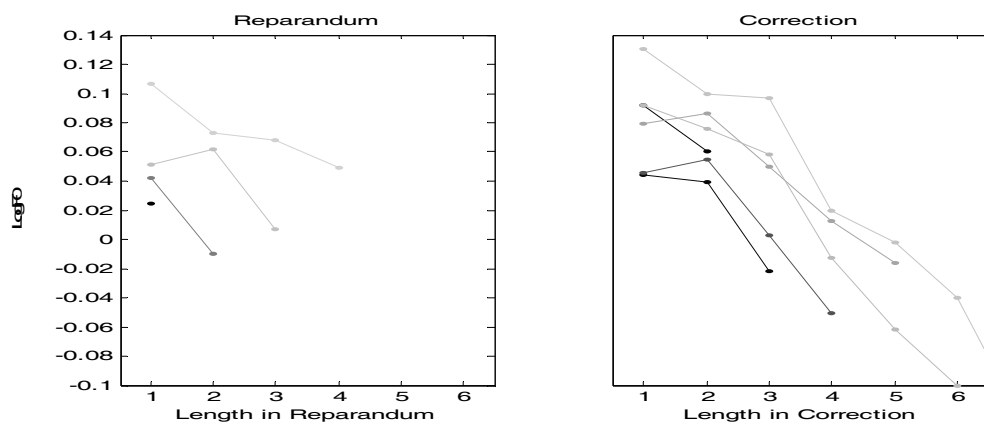


圖 5.37：詞語更正中，兩段分割在不同長度下，基頻之變化，顏色越深的線表示分割中此種長度之數量越多

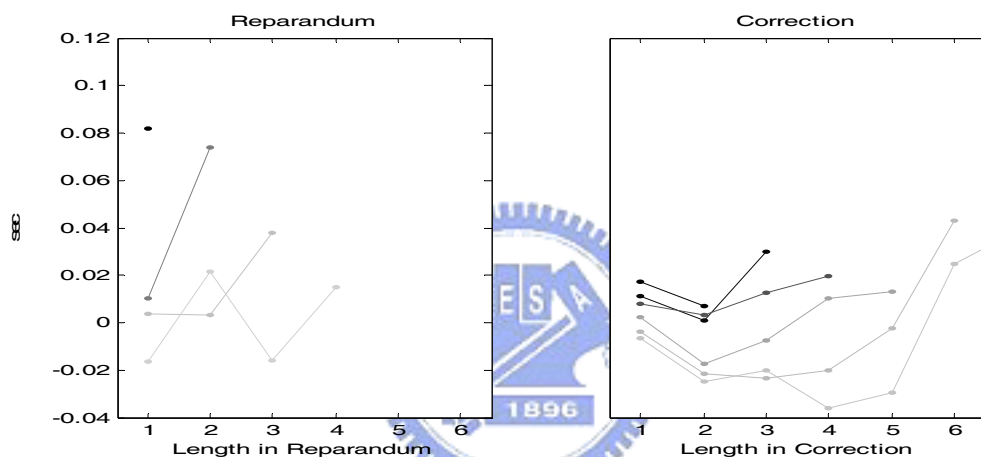


圖 5.38：詞語更正中，兩段分割在不同長度下，音節長度之變化，顏色越深的線表示分割中此種長度之數量越多

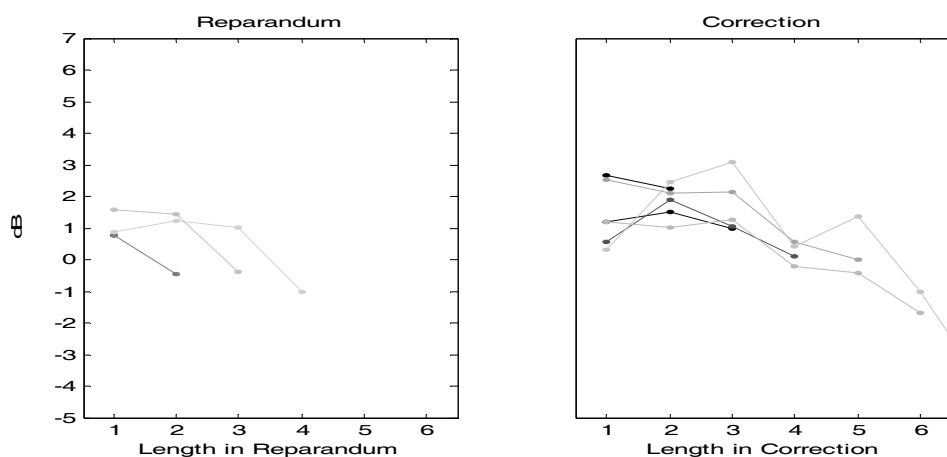


圖 5.39：詞語更正中，兩段分割在不同長度下，能量之變化，顏色越深的線表示分割中此種長度之數量越多

第六章：結論與未來展望

6.1 結論

本研究主要利用聲學參數以及語言學參數，建立一個自發性語音之韻律模型，並針對此自發性語音語料庫，以模型為基礎自動完成韻律標記。本研究主要有四個模型分別為：音節韻律模型、停頓標記聲學模型、韻律狀態轉移模型以及停頓標記語言模型，以描述語音中上層韻律狀態、停頓標記與聲學和語言學參數之間的關係，並分析各個模型參數之物理意義，藉此發現自發性語音中，上層以及音節層次之韻律變化情況，以及語言學參數與韻律的互動關係。

在此，本研究也利用模型參數觀察自發性語音中的特殊現象，例如：篇章提語和感歎詞等，與正常語音的差異；最後，並利用韻律標記結果，觀察詞語修補中韻律變化的情況，以上這些結果都顯示韻律在這些自發性語音中之特有現象，或文字上與韻律的關係，都帶有許多有用的訊息，以提供未來在自發性語音上的研究與發展更多有益的訊息，藉以提升自發性語音之語音辨認或語音理解。

6.2 未來展望

在韻律模型方面，本研究只利用了 4 個韻律狀態以及簡單的分類，描述自發性語音中特殊音節之上層以及音節層次的韻律變化，在未來可將其作更細緻化的處理，藉此觀察更細部的韻律現象，進而對偵測自發性語音中的特殊現象提供有用的訊息。

另外方面，可以利用詞語修補中所觀察到韻律變化的資訊作為詞語修補中 IP 偵測的訊息；或者利用停頓標記語言模型中，語言學參數與韻律的關係加入自發性語音辨認當中，藉此獲得更高的辨識率。最後，若完成良好的韻律模型甚至可將其利用至自發性語音之語音合成系統。

參考文獻

- 【1】S. Furui, "Recent progress in corpus-based spontaneous speech recognition," *IEICE Transactions on Information and Systems*, Vol. E88-D, No. 3, pp. 366-375, 2005.
- 【2】Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies," *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 14, No. 5, pp. 1526-1540, 2006.
- 【3】C. K. Lin, and L. S. Lee, "Improved Features and Models for Detecting Edit Disfluencies in Transcribing Spontaneous Mandarin Speech," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 7, pp. 1263-1278, 2009.
- 【4】M. Ostendorf, I. Shafran, and R. Bates, "Prosody models for conversational speech recognition," in *Proc. of the 2nd Plenary Meeting and Symposium on Prosody and Speech Processing 2003*, pp. 147-154.
- 【5】A. Stolcke, E. Shriberg, D. Hakkani-Tür, and G. Tür, "Modeling the prosody of hidden events for improved word recognition," in *Proc. of Eurospeech 1999*, pp. 311-314.
- 【6】K. Chen, M. Hasegawa-Johnson, A. Cohen, S. Borys, S.S. Kim, J. Cole, and J.Y. Choi, "Prosody dependent speech recognition on radio news corpus of American english," *IEEE Transactions on Speech, Audio, Language Processing*, Vol. 14, No. 1, pp. 232-245, 2006.
- 【7】江振宇, "非監督式中文語音韻律標記及韻律模式", 國立交通大學博士論文, 民國九十八年三月。
- 【8】M.Y. Tsai, F.C. Chou, and L.S. Lee, "Pronunciation modeling with reduced confusion for Mandarin Chinese using a three-stage framework," *IEEE Transactions on Speech and Audio Processing*, Vol. 15, No. 2, pp. 661-675, 2006.

- 【9】 M. Riley, W. Byrne, M. Finke, S. Khudanpur, and A. Ljolje, “Stochastic pronunciation modeling from hand-labelled phonetic corpora,” *Speech Communication*, Vol. 29, No. 2-4, pp. 209-224, 1999.
- 【10】 Y. Liu, and P. Fung, “State-Dependent Phonetic Tied Mixtures with Pronunciation Modeling for Spontaneous Speech Recognition,” *IEEE Transactions on Speech and Audio Processing*, Vol. 12, No. 4, pp. 351-364, 2004.
- 【11】 K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg, “TOBI: A standard for Labeling English Prosody,” In *Proc. of ICSLP 1992*, pp. 864-870.
- 【12】 Aijun Li, “Chinese prosody and prosodic labeling of spontaneous speech,” in *Proc. of Speech Prosody 2002*, pp. 39-46.
- 【13】 Maekawa, K., H. Kikuchi, Y. Igarashi and J. Venditti. “X-JToBI: An extended J_ToBI for spontaneous speech,” in *Proc. of ICSLP 2002*, pp. 1545-1548.
- 【14】 M. Ostendorf, I. Shafran, S. Shattuck-Hufnagel, L. Carmichael, and W. Byrne, “A prosodically labeled database of spontaneous speech,” in *Proc. of the ISCA Workshop on Prosody in Speech Recognition and Understanding 2001*, pp. 119-121.
- 【15】 A. Batliner, R. Kompe, A. Kießling, M. Mast, H. Niemann, and E. Nöth, “M = Syntax + Prosody: A syntactic prosodic labeling scheme for large spontaneous speech databases,” *Speech Communication*, Vol. 25, No.4, pp. 193-222, 1998.
- 【16】 S.C. Tseng, “Processing Spoken Mandarin Corpora,” *Traitement automatique des langues, Special Issue: Spoken Corpus Processing*, Vol. 45, No. 2, pp. 89-108, 2004.
- 【17】 S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book Version 3.0*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- 【18】 李柏蒼, “自發性國語語音辨識”, 國立交通大學碩士論文, 民國九十六年八月。
- 【19】 S.C. Tseng, “Monosyllabic word merger in Mandarin,” *Language Variation and Change*, Vol. 17, No.3, pp. 231-256, 2005.

- 【20】 C.Y. Tseng and Z.Y. Su, “Corpus approach to phonetic investigation - methods, quantitative evidence and findings of Mandarin speech prosody,” in *Proc. of Oriental COCOSDA Workshop 2006*, pp. 123-138.
- 【21】 K. Sjölander and J. Beskow, “Wavesurfer - an open source speech tool,” in *Proc. of ICSLP 2000*, pp. 464-467.
- 【22】 M. Sönmez, L. Heck, M. Weintraub, and E. Shriberg, “A lognormal tied mixture model of pitch for prosody-based speaker recognition,” in *Proc. of Eurospeech 1997*, pp. 1391–1394.
- 【23】 D. T. Chappell and J. H. L. Hansen, “Speaker-Specific Pitch Contour Modeling and Modification,” in *Proc. of ICSLP 1998*, pp. 885-888.
- 【24】 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, 1990.
- 【25】 江振宇, “中文段詞器之改進”, 國立交通大學碩士論文, 民國九十三年七月。
- 【26】 S.C. Tseng, “Repairs in Mandarin Conversation,” *Journal of Chinese Linguistics*, Vol. 34, No.1, pp. 80-120, 2006

附錄一

詞類分類表

2 類詞類			8 類詞類			23 類詞類			46 類詞類		
編號	中文詞類	代號	編號	中文詞類	代號	編號	中文詞類	代號	編號	中文詞類	代號
1	實詞	S	1	非謂形容詞	A	1	非謂形容詞	A	1	非謂形容詞	A
2	功能詞	F	2	連接詞	C	2	連接詞	C	2	對等連接詞	Caa
2	功能詞	F	2	連接詞	C	2	連接詞	C	3	連接詞，如： 等等	Cab
2	功能詞	F	2	連接詞	C	2	連接詞	C	4	連接詞，如： 的話	Cba
2	功能詞	F	2	連接詞	C	2	連接詞	C	5	關聯連接詞	Cbb
2	實詞	S	3	副詞	D	5	副詞	D	6	數量副詞	Da
2	實詞	S	3	副詞	D	3	動詞前程度副詞	Dfa	7	動詞前程度副詞	Dfa
2	實詞	S	3	副詞	D	4	動詞後程度副詞	Dfb	8	動詞後程度副詞	Dfb
2	實詞	S	3	副詞	D	5	副詞	D	9	時態標記	Di
2	實詞	S	3	副詞	D	5	副詞	D	10	句副詞	Dk
2	實詞	S	3	副詞	D	5	副詞	D	11	副詞	D
1	實詞	S	4	體詞	N	6	普通名詞	N	12	普通名詞	Na
1	實詞	S	4	體詞	N	6	普通名詞	N	13	專有名詞	Nb
1	實詞	S	4	體詞	N	6	普通名詞	N	14	地方詞	Nc
1	實詞	S	4	體詞	N	9	後置詞,位置詞	Ng	15	位置詞	Ncd
1	實詞	S	4	體詞	N	7	時間詞	Nd	16	時間詞	Nd
1	實詞	S	4	體詞	N	8	定詞,量詞	Ne	17	數詞定詞	Neu
1	實詞	S	4	體詞	N	8	定詞,量詞	Ne	18	特指定詞	Nes
1	實詞	S	4	體詞	N	8	定詞,量詞	Ne	19	指代定詞	Nep
1	實詞	S	4	體詞	N	8	定詞,量詞	Ne	20	數量定詞	Neqa
1	實詞	S	4	體詞	N	8	定詞,量詞	Ne	21	後置數量定詞	Neqb
1	實詞	S	4	體詞	N	8	定詞,量詞	Ne	22	量詞	Nf
1	實詞	S	4	體詞	N	9	後置詞,位置詞	Ng	23	後置詞	Ng
1	實詞	S	4	體詞	N	6	普通名詞	N	47		Nv
1	實詞	S	4	體詞	N	10	代名詞	Nh	24	代名詞	Nh

2 類詞類			8 類詞類			23 類詞類			46 類詞類		
編號	中文詞類	代號	編號	中文詞類	代號	編號	中文詞類	代號	編號	中文詞類	代號
2	功能詞	F	5	感嘆、語助詞	T	12	感嘆、語助詞	T	25	感嘆詞	I
2	功能詞	F	6	介詞	P	11	介詞	P	26	介詞	P
2	功能詞	F	5	感嘆、語助詞	T	12	感嘆、語助詞	T	27	語助詞	T
1	實詞	S	7	動詞	V	13	不及物動詞	VA	28	動作不及物動詞	VA
1	實詞	S	7	動詞	V	13	不及物動詞	VA	29	動作使動動詞	VAC
1	實詞	S	7	動詞	V	14	及物動詞	VC	30	動作類及物動詞	VB
1	實詞	S	7	動詞	V	14	及物動詞	VC	31	動作及物動詞	VC
1	實詞	S	7	動詞	V	14	及物動詞	VC	32	動作接地方賓語動詞	VCL
1	實詞	S	7	動詞	V	14	及物動詞	VC	33	雙賓動詞	VD
1	實詞	S	7	動詞	V	14	及物動詞	VC	34	動作句賓動詞	VE
1	實詞	S	7	動詞	V	14	及物動詞	VC	35	動作謂賓動詞	VF
1	實詞	S	7	動詞	V	13	不及物動詞	VA	36	分類動詞	VG
1	實詞	S	7	動詞	V	15	狀態不及物動詞	VH	37	狀態不及物動詞	VH
1	實詞	S	7	動詞	V	15	狀態不及物動詞	VH	38	狀態使動動詞	VHC
1	實詞	S	7	動詞	V	15	狀態不及物動詞	VH	39	狀態類及物動詞	VI
1	實詞	S	7	動詞	V	14	及物動詞	VC	40	狀態及物動詞	VJ
1	實詞	S	7	動詞	V	14	及物動詞	VC	41	狀態句賓動詞	VK
1	實詞	S	7	動詞	V	14	及物動詞	VC	42	狀態謂賓動詞	VL
1	實詞	S	7	動詞	V	16	有	V_2	43	有	V_2
2	功能詞	F	8	的	DE	17	的，之，得	DE	44	的，之，得	DE
2	功能詞	F	9	是	SHI	18	是	SHI	45	是	SHI
1	實詞	S	11	外文	FW	20	外文標記	FW	46	外文標記	FW
1	實詞	S	10	定量複合詞	DM	19	定量複合詞	DM	58	定量複合詞	DM
4	Paralinguistic	ParaL	13	Paralinguistic	ParaL	24	Paralinguistic	ParaL	59	Paralinguistic	ParaL
5	Particle	Par	14	Particle	Par	25	Particle	Par	60	Particle	Par
5	Particle	Par	14	Particle	Par	25	Particle	Par	61	Marker	Marker

附錄二

音節類型、注音以及音節型態對照表

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
1	ㄗ	zhi	zh	2	FNULL1	1
2	ㄘ	chi	ch	3	FNULL1	1
3	ㄙ	shi	sh	4	FNULL1	1
4	ㄖ	ri	r	5	FNULL1	1
5	ㄗ	zi	z	6	FNULL2	40
6	ㄘ	ci	c	7	FNULL2	40
7	ㄙ	si	s	8	FNULL2	40
8	ㄚ	a	INULL	1	a	2
9	ㄗㄚ	zha	zh	2	a	2
10	ㄘㄚ	cha	ch	3	a	2
11	ㄙㄚ	sha	sh	4	a	2
12	ㄗㄚ	za	z	6	a	2
13	ㄘㄚ	ca	c	7	a	2
14	ㄙㄚ	sa	s	8	a	2
15	ㄍㄚ	ga	g	9	a	2
16	ㄎㄚ	ka	k	10	a	2
17	ㄏㄚ	ha	h	11	a	2
18	ㄉㄚ	da	d	15	a	2
19	ㄊㄚ	ta	t	16	a	2
20	ㄋㄚ	na	n	17	a	2
21	ㄌㄚ	la	l	18	a	2
22	ㄅㄚ	ba	b	19	a	2
23	ㄆㄚ	pa	p	20	a	2
24	ㄇㄚ	ma	m	21	a	2
25	ㄈㄚ	fa	f	22	a	2
26	ㄛ	o	INULL	1	o	3
27	ㄌㄛ	lo	l	18	o	3
28	ㄅㄛ	bo	b	19	o	3
29	ㄆㄛ	po	p	20	o	3
30	ㄇㄛ	mo	m	21	o	3

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
31	ㄈ	fo	f	22	o	3
32	ㄜ	e	INULL	1	e	4
33	ㄗ	zhe	zh	2	e	4
34	ㄘ	che	ch	3	e	4
35	ㄙ	she	sh	4	e	4
36	ㄖ	re	r	5	e	4
37	ㄗ	ze	z	6	e	4
38	ㄘ	ce	c	7	e	4
39	ㄙ	se	s	8	e	4
40	ㄍ	ge	g	9	e	4
41	ㄎ	ke	k	10	e	4
42	ㄏ	he	h	11	e	4
43	ㄉ	de	d	15	e	4
44	ㄊ	te	t	16	e	4
45	ㄋ	ne	n	17	e	4
46	ㄌ	le	l	18	e	4
47	ㄞ	ai	INULL	1	ai	6
48	ㄗㄞ	zhai	zh	2	ai	6
49	ㄘㄞ	chai	ch	3	ai	6
50	ㄙㄞ	shai	sh	4	ai	6
51	ㄗㄞ	zai	z	6	ai	6
52	ㄘㄞ	cai	c	7	ai	6
53	ㄙㄞ	sai	s	8	ai	6
54	ㄍㄞ	gai	g	9	ai	6
55	ㄎㄞ	kai	k	10	ai	6
56	ㄏㄞ	hai	h	11	ai	6
57	ㄉㄞ	dai	d	15	ai	6
58	ㄊㄞ	tai	t	16	ai	6
59	ㄋㄞ	nai	n	17	ai	6
60	ㄌㄞ	lai	l	18	ai	6

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
61	ㄅㄢ	bai	b	19	ai	6
62	ㄆㄢ	pai	p	20	ai	6
63	ㄇㄢ	mai	m	21	ai	6
64	ㄜ	eh	INULL	1	eh	5
65	ㄗㄟ	zhei	zh	2	ei	7
66	ㄕㄟ	shei	sh	4	ei	7
67	ㄗㄟ	zei	z	6	ei	7
68	ㄙㄟ	sei	s	8	ei	7
69	ㄍㄟ	gei	g	9	ei	7
70	ㄏㄟ	hei	h	11	ei	7
71	ㄉㄟ	dei	d	15	ei	7
72	ㄋㄟ	nei	n	17	ei	7
73	ㄌㄟ	lei	l	18	ei	7
74	ㄅㄟ	bei	b	19	ei	7
75	ㄆㄟ	pei	p	20	ei	7
76	ㄇㄟ	mei	m	21	ei	7
77	ㄈㄟ	fei	f	22	ei	7
78	ㄠ	ao	INULL	1	ao	8
79	ㄗㄠ	zhao	zh	2	ao	8
80	ㄔㄠ	chao	ch	3	ao	8
81	ㄕㄠ	shao	sh	4	ao	8
82	ㄖㄠ	rao	r	5	ao	8
83	ㄗㄠ	zao	z	6	ao	8
84	ㄘㄠ	cao	c	7	ao	8
85	ㄙㄠ	sao	s	8	ao	8
86	ㄍㄠ	gao	g	9	ao	8
87	ㄎㄠ	kao	k	10	ao	8
88	ㄏㄠ	hao	h	11	ao	8
89	ㄉㄠ	dao	d	15	ao	8

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
90	ㄊㄠ	tao	t	16	ao	8
91	ㄋㄠ	nao	n	17	ao	8
92	ㄌㄠ	lao	l	18	ao	8
93	ㄅㄠ	bao	b	19	ao	8
94	ㄆㄠ	pao	p	20	ao	8
95	ㄇㄠ	mao	m	21	ao	8
96	ㄡ	ou	INULL	1	ou	9
97	ㄓㄡ	zhou	zh	2	ou	9
98	ㄔㄡ	chou	ch	3	ou	9
99	ㄕㄡ	shou	sh	4	ou	9
100	ㄖㄡ	rou	r	5	ou	9
101	ㄗㄡ	zou	z	6	ou	9
102	ㄘㄡ	cou	c	7	ou	9
103	ㄙㄡ	sou	s	8	ou	9
104	ㄍㄡ	gou	g	9	ou	9
105	ㄎㄡ	kou	k	10	ou	9
106	ㄏㄡ	hou	h	11	ou	9
107	ㄉㄡ	dou	d	15	ou	9
108	ㄊㄡ	tou	t	16	ou	9
109	ㄋㄡ	nou	n	17	ou	9
110	ㄌㄡ	lou	l	18	ou	9
111	ㄆㄡ	pou	p	20	ou	9
112	ㄇㄡ	mou	m	21	ou	9
113	ㄈㄡ	fou	f	22	ou	9
114	ㄢ	an	INULL	1	an	10
115	ㄓㄢ	zhan	zh	2	an	10
116	ㄔㄢ	chan	ch	3	an	10
117	ㄕㄢ	shan	sh	4	an	10
118	ㄖㄢ	ran	r	5	an	10
119	ㄗㄢ	zan	z	6	an	10
120	ㄘㄢ	can	c	7	an	10

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
121	ㄙㄢ	san	s	8	an	10
122	ㄍㄢ	gan	g	9	an	10
123	ㄎㄢ	kan	k	10	an	10
124	ㄏㄢ	han	h	11	an	10
125	ㄉㄢ	dan	d	15	an	10
126	ㄊㄢ	tan	t	16	an	10
127	ㄋㄢ	nan	n	17	an	10
128	ㄌㄢ	lan	l	18	an	10
129	ㄅㄢ	ban	b	19	an	10
130	ㄆㄢ	pan	p	20	an	10
131	ㄇㄢ	man	m	21	an	10
132	ㄈㄢ	fan	f	22	an	10
133	ㄣ	en	INULL	1	en	11
134	ㄗㄣ	zhen	zh	2	en	11
135	ㄘㄣ	chen	ch	3	en	11
136	ㄑㄣ	shen	sh	4	en	11
137	ㄣ	ren	r	5	en	11
138	ㄗㄣ	zen	z	6	en	11
139	ㄘㄣ	cen	c	7	en	11
140	ㄙㄣ	sen	s	8	en	11
141	ㄍㄣ	gen	g	9	en	11
142	ㄎㄣ	ken	k	10	en	11
143	ㄏㄣ	hen	h	11	en	11
144	ㄋㄣ	nen	n	17	en	11
145	ㄅㄣ	ben	b	19	en	11
146	ㄆㄣ	pen	p	20	en	11
147	ㄇㄣ	men	m	21	en	11
148	ㄈㄣ	fen	f	22	en	11
149	ㄤ	ang	INULL	1	ang	12
150	ㄗㄤ	zhang	zh	2	ang	12

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
151	ㄔㄤ	chang	ch	3	ang	12
152	ㄕㄤ	shang	sh	4	ang	12
153	ㄖㄤ	rang	r	5	ang	12
154	ㄗㄤ	zang	z	6	ang	12
155	ㄘㄤ	cang	c	7	ang	12
156	ㄙㄤ	sang	s	8	ang	12
157	ㄍㄤ	gang	g	9	ang	12
158	ㄎㄤ	kang	k	10	ang	12
159	ㄏㄤ	hang	h	11	ang	12
160	ㄉㄤ	dang	d	15	ang	12
161	ㄊㄤ	tang	t	16	ang	12
162	ㄋㄤ	nang	n	17	ang	12
163	ㄌㄤ	lang	l	18	ang	12
164	ㄅㄤ	bang	b	19	ang	12
165	ㄆㄤ	pang	p	20	ang	12
166	ㄇㄤ	mang	m	21	ang	12
167	ㄈㄤ	fang	f	22	ang	12
168	ㄓㄥ	zheng	zh	2	eng	13
169	ㄔㄥ	cheng	ch	3	eng	13
170	ㄕㄥ	sheng	sh	4	eng	13
171	ㄖㄥ	reng	r	5	eng	13
172	ㄗㄥ	zeng	z	6	eng	13
173	ㄘㄥ	ceng	c	7	eng	13
174	ㄙㄥ	seng	s	8	eng	13
175	ㄍㄥ	geng	g	9	eng	13
176	ㄎㄥ	keng	k	10	eng	13
177	ㄏㄥ	heng	h	11	eng	13
178	ㄉㄥ	deng	d	15	eng	13
179	ㄊㄥ	teng	t	16	eng	13
180	ㄋㄥ	neng	n	17	eng	13

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
181	ㄌㄥ	leng	l	18	eng	13
182	ㄅㄥ	beng	b	19	eng	13
183	ㄆㄥ	peng	p	20	eng	13
184	ㄇㄥ	meng	m	21	eng	13
185	ㄈㄥ	feng	f	22	eng	13
186	ㄧ	yi	INULL	1	yi	14
187	ㄐㄧ	ji	j	12	yi	14
188	ㄑㄧ	qi	q	13	yi	14
189	ㄒㄧ	xi	x	14	yi	14
190	ㄉㄧ	di	d	15	yi	14
191	ㄊㄧ	ti	t	16	yi	14
192	ㄋㄧ	ni	n	17	yi	14
193	ㄌㄧ	li	l	18	yi	14
194	ㄅㄧ	bi	b	19	yi	14
195	ㄆㄧ	pi	p	20	yi	14
196	ㄇㄧ	mi	m	21	yi	14
197	ㄨ	wu	INULL	1	wu	15
198	ㄓㄨ	zhu	zh	2	wu	15
199	ㄔㄨ	chu	ch	3	wu	15
200	ㄕㄨ	shu	sh	4	wu	15
201	ㄖㄨ	ru	r	5	wu	15
202	ㄗㄨ	zu	z	6	wu	15
203	ㄘㄨ	cu	c	7	wu	15
204	ㄙㄨ	su	s	8	wu	15
205	ㄍㄨ	gu	g	9	wu	15
206	ㄎㄨ	ku	k	10	wu	15
207	ㄏㄨ	hu	h	11	wu	15
208	ㄉㄨ	du	d	15	wu	15
209	ㄊㄨ	tu	t	16	wu	15
210	ㄋㄨ	nu	n	17	wu	15

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
211	ㄌㄨ	lu	l	18	wu	15
212	ㄅㄨ	bu	b	19	wu	15
213	ㄆㄨ	pu	p	20	wu	15
214	ㄇㄨ	mu	m	21	wu	15
215	ㄈㄨ	fu	f	22	wu	15
216	ㄩ	yu	INULL	1	yu	16
217	ㄐㄩ	ju	j	12	yu	16
218	ㄑㄩ	qu	q	13	yu	16
219	ㄒㄩ	xu	x	14	yu	16
220	ㄋㄩ	nyu	n	17	yu	16
221	ㄌㄩ	lyu	l	18	yu	16
222	ㄧㄚ	ya	INULL	1	ya	17
223	ㄐㄧㄚ	jia	j	12	ya	17
224	ㄑㄧㄚ	qia	q	13	ya	17
225	ㄒㄧㄚ	xia	x	14	ya	17
226	ㄌㄧㄚ	lia	l	18	ya	17
227	ㄧㄝ	ye	INULL	1	ye	18
228	ㄐㄧㄝ	jie	j	12	ye	18
229	ㄑㄧㄝ	qie	q	13	ye	18
230	ㄒㄧㄝ	xie	x	14	ye	18
231	ㄉㄧㄝ	die	d	15	ye	18
232	ㄊㄧㄝ	tie	t	16	ye	18
233	ㄋㄧㄝ	nie	n	17	ye	18
234	ㄌㄧㄝ	lie	l	18	ye	18
235	ㄅㄧㄝ	bie	b	19	ye	18
236	ㄆㄧㄝ	pie	p	20	ye	18
237	ㄇㄧㄝ	mie	m	21	ye	18
238	ㄧㄞ	yai	INULL	1	yai	19
239	ㄧㄠ	yao	INULL	1	yao	20
240	ㄐㄧㄠ	jiao	j	12	yao	20

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
241	ㄑ一ㄠ	qiao	q	13	yao	20
242	ㄒ一ㄠ	xiao	x	14	yao	20
243	ㄉ一ㄠ	diao	d	15	yao	20
244	ㄊ一ㄠ	tiao	t	16	yao	20
245	ㄋ一ㄠ	niao	n	17	yao	20
246	ㄌ一ㄠ	liao	l	18	yao	20
247	ㄅ一ㄠ	biao	b	19	yao	20
248	ㄆ一ㄠ	piao	p	20	yao	20
249	ㄇ一ㄠ	miao	m	21	yao	20
250	ㄧㄡ	you	INULL	1	you	21
251	ㄐ一ㄡ	jiu	j	12	you	21
252	ㄑ一ㄡ	qiu	q	13	you	21
253	ㄒ一ㄡ	xiu	x	14	you	21
254	ㄉ一ㄡ	diu	d	15	you	21
255	ㄋ一ㄡ	niu	n	17	you	21
256	ㄌ一ㄡ	liu	l	18	you	21
257	ㄇ一ㄡ	miu	m	21	you	21
258	ㄧㄢ	yan	INULL	1	yan	22
259	ㄐ一ㄢ	jian	j	12	yan	22
260	ㄑ一ㄢ	qian	q	13	yan	22
261	ㄒ一ㄢ	xian	x	14	yan	22
262	ㄉ一ㄢ	dian	d	15	yan	22
263	ㄊ一ㄢ	tian	t	16	yan	22
264	ㄋ一ㄢ	nian	n	17	yan	22
265	ㄌ一ㄢ	lian	l	18	yan	22
266	ㄅ一ㄢ	bian	b	19	yan	22
267	ㄆ一ㄢ	pian	p	20	yan	22
268	ㄇ一ㄢ	mian	m	21	yan	22
269	ㄧㄣ	yin	INULL	1	yin	23

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
270	ㄐㄧㄣ	jin	j	12	yin	23
271	ㄑㄧㄣ	qin	q	13	yin	23
272	ㄒㄧㄣ	xin	x	14	yin	23
273	ㄋㄧㄣ	nin	n	17	yin	23
274	ㄌㄧㄣ	lin	l	18	yin	23
275	ㄅㄧㄣ	bin	b	19	yin	23
276	ㄆㄧㄣ	pin	p	20	yin	23
277	ㄇㄧㄣ	min	m	21	yin	23
278	ㄧㄤ	yang	NULL	1	yang	24
279	ㄐㄧㄤ	jiang	j	12	yang	24
280	ㄑㄧㄤ	qiang	q	13	yang	24
281	ㄒㄧㄤ	xiang	x	14	yang	24
282	ㄋㄧㄤ	niang	n	17	yang	24
283	ㄌㄧㄤ	liang	l	18	yang	24
284	ㄧㄥ	ying	NULL	1	ying	25
285	ㄐㄧㄥ	jing	j	12	ying	25
286	ㄑㄧㄥ	qing	q	13	ying	25
287	ㄒㄧㄥ	xing	x	14	ying	25
288	ㄉㄧㄥ	ding	d	15	ying	25
289	ㄊㄧㄥ	ting	t	16	ying	25
290	ㄋㄧㄥ	ning	n	17	ying	25
291	ㄌㄧㄥ	ling	l	18	ying	25
292	ㄅㄧㄥ	bing	b	19	ying	25
293	ㄆㄧㄥ	ping	p	20	ying	25
294	ㄇㄧㄥ	ming	m	21	ying	25
295	ㄨㄚ	wa	NULL	1	wa	26
296	ㄓㄨㄚ	zhua	zh	2	wa	26
297	ㄔㄨㄚ	chua	ch	3	wa	26
298	ㄕㄨㄚ	shua	sh	4	wa	26
299	ㄍㄨㄚ	gua	g	9	wa	26
300	ㄎㄨㄚ	kua	k	10	wa	26

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
301	ㄏㄨㄚˊ	hua	h	11	wa	26
302	ㄨㄛˊ	wo	INULL	1	wo	27
303	ㄓㄨㄛˊ	zhuo	zh	2	wo	27
304	ㄔㄨㄛˊ	chuo	ch	3	wo	27
305	ㄕㄨㄛˊ	shuo	sh	4	wo	27
306	ㄖㄨㄛˊ	ruo	r	5	wo	27
307	ㄗㄨㄛˊ	zuo	z	6	wo	27
308	ㄘㄨㄛˊ	cuo	c	7	wo	27
309	ㄙㄨㄛˊ	suo	s	8	wo	27
310	ㄍㄨㄛˊ	guo	g	9	wo	27
311	ㄎㄨㄛˊ	kuo	k	10	wo	27
312	ㄏㄨㄛˊ	huo	h	11	wo	27
313	ㄉㄨㄛˊ	duo	d	15	wo	27
314	ㄊㄨㄛˊ	tuo	t	16	wo	27
315	ㄋㄨㄛˊ	nuo	n	17	wo	27
316	ㄌㄨㄛˊ	luo	l	18	wo	27
317	ㄨㄞˊ	wai	INULL	1	wai	28
318	ㄓㄨㄞˊ	zhuai	zh	2	wai	28
319	ㄔㄨㄞˊ	chuai	ch	3	wai	28
320	ㄕㄨㄞˊ	shuai	sh	4	wai	28
321	ㄍㄨㄞˊ	guai	g	9	wai	28
322	ㄎㄨㄞˊ	kuai	k	10	wai	28
323	ㄏㄨㄞˊ	huai	h	11	wai	28
324	ㄨㄟˊ	wei	INULL	1	wei	29
325	ㄓㄨㄟˊ	zhui	zh	2	wei	29
326	ㄔㄨㄟˊ	chui	ch	3	wei	29
327	ㄕㄨㄟˊ	shui	sh	4	wei	29
328	ㄖㄨㄟˊ	rui	r	5	wei	29
329	ㄗㄨㄟˊ	zui	z	6	wei	29
330	ㄘㄨㄟˊ	cui	c	7	wei	29

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
331	ㄅㄨㄟ	sui	s	8	wei	29
332	ㄍㄨㄟ	gui	g	9	wei	29
333	ㄎㄨㄟ	kui	k	10	wei	29
334	ㄏㄨㄟ	hui	h	11	wei	29
335	ㄉㄨㄟ	dui	d	15	wei	29
336	ㄊㄨㄟ	tui	t	16	wei	29
337	ㄨㄢ	wan	INULL	1	wan	30
338	ㄓㄨㄢ	zhuan	zh	2	wan	30
339	ㄔㄨㄢ	chuan	ch	3	wan	30
340	ㄕㄨㄢ	shuan	sh	4	wan	30
341	ㄖㄨㄢ	ruan	r	5	wan	30
342	ㄗㄨㄢ	zuan	z	6	wan	30
343	ㄘㄨㄢ	cuan	c	7	wan	30
344	ㄙㄨㄢ	suan	s	8	wan	30
345	ㄍㄨㄢ	guan	g	9	wan	30
346	ㄎㄨㄢ	kuan	k	10	wan	30
347	ㄏㄨㄢ	huan	h	11	wan	30
348	ㄉㄨㄢ	duan	d	15	wan	30
349	ㄊㄨㄢ	tuan	t	16	wan	30
350	ㄋㄨㄢ	nuan	n	17	wan	30
351	ㄌㄨㄢ	luan	l	18	wan	30
352	ㄨㄣ	wen	INULL	1	wen	31
353	ㄓㄨㄣ	zhun	zh	2	wen	31
354	ㄔㄨㄣ	chun	ch	3	wen	31
355	ㄕㄨㄣ	shun	sh	4	wen	31
356	ㄖㄨㄣ	run	r	5	wen	31
357	ㄗㄨㄣ	zun	z	6	wen	31
358	ㄘㄨㄣ	cun	c	7	wen	31
359	ㄙㄨㄣ	sun	s	8	wen	31

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
360	ㄍㄨㄣ	gun	g	9	wen	31
361	ㄎㄨㄣ	kun	k	10	wen	31
362	ㄏㄨㄣ	hun	h	11	wen	31
363	ㄉㄨㄣ	dun	d	15	wen	31
364	ㄊㄨㄣ	tun	t	16	wen	31
365	ㄌㄨㄣ	lun	l	18	wen	31
366	ㄨㄤ	wang	INULL	1	wang	32
367	ㄓㄨㄤ	zhuang	zh	2	wang	32
368	ㄔㄨㄤ	chuang	ch	3	wang	32
369	ㄕㄨㄤ	shuang	sh	4	wang	32
370	ㄍㄨㄤ	guang	g	9	wang	32
371	ㄎㄨㄤ	kuang	k	10	wang	32
372	ㄏㄨㄤ	huang	h	11	wang	32
373	ㄨㄥ	weng	INULL	1	weng	33
374	ㄓㄨㄥ	zhong	zh	2	weng	33
375	ㄔㄨㄥ	chong	ch	3	weng	33
376	ㄖㄨㄥ	rong	r	5	weng	33
377	ㄗㄨㄥ	zong	z	6	weng	33
378	ㄘㄨㄥ	cong	c	7	weng	33
379	ㄙㄨㄥ	song	s	8	weng	33
380	ㄍㄨㄥ	gong	g	9	weng	33
381	ㄎㄨㄥ	kong	k	10	weng	33
382	ㄏㄨㄥ	hong	h	11	weng	33
383	ㄉㄨㄥ	dong	d	15	weng	33
384	ㄊㄨㄥ	tong	t	16	weng	33
385	ㄋㄨㄥ	nong	n	17	weng	33
386	ㄌㄨㄥ	long	l	18	weng	33
387	ㄩㄝ	yue	INULL	1	yue	34
388	ㄐㄩㄝ	jue	j	12	yue	34
389	ㄑㄩㄝ	que	q	13	yue	34

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
390	ㄒㄩㄝ	xue	x	14	yue	34
391	ㄋㄩㄝ	nu:e	n	17	yue	34
392	ㄌㄩㄝ	lu:e	l	18	yue	34
393	ㄩㄢ	yuan	INULL	1	yuan	35
394	ㄐㄩㄢ	juan	j	12	yuan	35
395	ㄑㄩㄢ	quan	q	13	yuan	35
396	ㄒㄩㄢ	xuan	x	14	yuan	35
397	ㄌㄩㄢ	lu:an	l	18	yuan	35
398	ㄩㄣ	yun	INULL	1	yun	36
399	ㄐㄩㄣ	jun	j	12	yun	36
400	ㄑㄩㄣ	qun	q	13	yun	36
401	ㄒㄩㄣ	xun	x	14	yun	36
402	ㄌㄩㄣ	l:un	l	18	yun	36
403	ㄩㄥ	yong	INULL	1	yung	37
404	ㄐㄩㄥ	jiong	j	12	yung	37
405	ㄑㄩㄥ	qiong	q	13	yung	37
406	ㄒㄩㄥ	xiong	x	14	yung	37
407	ㄦ	er	INULL	1	er	38
408	ㄧㄝ	yo	INULL	1	yo	39
409	ㄥ	eng	INULL	1	eng	13
410	ㄟ	ei	INULL	1	ei	7
411	ㄇㄝ	me	m	21	e	4
414	ㄋㄨㄣ	nun	n	8	wen	31
415						
416	ㄋㄧㄚ	nia	n	8	ya	17
417						
418	ㄉㄣ	den	d	15	en	11
419	ㄉㄧㄚ	dia	d	15	ya	17
420	ㄎㄟ	kei	k	10	ei	7
430	*	A	INULL	1	a	2

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
390	ㄒㄩㄝ	xue	x	14	yue	34
391	ㄋㄨːㄝ	nu:e	n	17	yue	34
392	ㄌㄨːㄝ	lu:e	l	18	yue	34
393	ㄩㄢ	yuan	INULL	1	yuan	35
394	ㄐㄩㄢ	juan	j	12	yuan	35
395	ㄑㄩㄢ	quan	q	13	yuan	35
396	ㄒㄩㄢ	xuan	x	14	yuan	35
397	ㄌㄨːㄢ	lu:an	l	18	yuan	35
398	ㄩㄣ	yun	INULL	1	yun	36
399	ㄐㄩㄣ	jun	j	12	yun	36
400	ㄑㄩㄣ	qun	q	13	yun	36
401	ㄒㄩㄣ	xun	x	14	yun	36
402	ㄌㄨːㄣ	l:un	l	18	yun	36
403	ㄩㄥ	yong	INULL	1	yung	37
404	ㄐㄩㄥ	jiong	j	12	yung	37
405	ㄑㄩㄥ	qiong	q	13	yung	37
406	ㄒㄩㄥ	xiong	x	14	yung	37
407	ㄦ	er	INULL	1	er	38
408	ㄧㄝ	yo	INULL	1	yo	39
409	ㄥ	eng	INULL	1	eng	13
410	ㄟ	ei	INULL	1	ei	7
411	ㄇㄝ	me	m	21	e	4
414	ㄋㄨㄣ	nun	n	8	wen	31
416	ㄋㄧㄚ	nia	n	8	ya	17
418	ㄉㄣ	den	d	15	en	11
419	ㄉㄧㄚ	dia	d	15	ya	17
420	ㄎㄟ	kei	k	10	ei	7
430	*	A	INULL	1	a	2
431	*	AI	INULL	1	ai	6
432	*	BA	b	19	a	2

音碼	注音	漢語拼音	子音拼音(22 類)	子音編號(22 類)	母音拼音(40 類)	母音編號
433	*	E	INULL	1	e	4
434	*	EI	*	23	*	41
435	*	GE	g	9	e	4
436	*	HAN	*	24	*	42
437	*	HEIN	*	25	*	43
438	*	HEN	*	26	*	44
439	*	HO	*	27	*	45
440	*	LA	l	18	a	2
441	*	MA	m	21	a	2
442	*	ME	m	21	e	4
443	*	MHM	*	28	*	46
444	*	MHMHM	*	29	*	47
445	*	NA	n	17	a	2
446	*	NE	n	17	e	4
447	*	NO	n	17	ou	9
448	*	O	INULL	1	o	3
449	*	SHEN	sh	4	en	11
450	*	WA	INULL	1	wa	26
451	*	YA	INULL	1	ya	17
452	*	YOU	INULL	1	you	21
453	*	ZHE	zh	2	e	4
454	*	filler	*	30	*	48
460	*	@BREATHE	*	31	*	49
461	*	@CLEAR_THROAT	*	32	*	50
462	*	@COUGH	*	33	*	51
463	*	@LAUGH	*	34	*	52
464	*	@NOISE	*	35	*	53
465	*	@NON_SPEECH_SOUND	*	36	*	54
466	*	@SMACK	*	37	*	55
467	*	@SPEECH_SOUND	*	38	*	56
468	*	@SWALLOW	*	39	*	57

附錄三

The question set Θ_1 used to construct the decision trees for building the break-acoustics model $p(pd_n, ed_n, pj_n, dl_n, df_n | B_n, \mathbf{I}_n)$ is listed below:

1. Syllable Level

$Q_1 1.1$: Is the initial of the following syllable a null one or in $\{m, n, l, r\}$?

$Q_1 1.2$: Is the initial of the following syllable a null one?

$Q_1 1.3$: Is the initial of the following syllable in $\{b, d, g\}$?

$Q_1 1.4$: Is the initial of the following syllable in $\{f, s, sh, shi, h\}$?

$Q_1 1.5$: Is the initial of the following syllable in $\{m, n, l, r\}$?

$Q_1 1.6$: Is the initial of the following syllable in $\{ts, ch, chi\}$?

$Q_1 1.7$: Is the initial of the following syllable in $\{p, t, k\}$?

$Q_1 1.8$: Is the initial of the following syllable in $\{tz, j, ji\}$?

$Q_1 1.9$: Is the inter-syllable location an inter-word?

$Q_1 1.10$: Is the inter-syllable location a intra-word?

2. Questions related to sentence level features

All the following questions are subject to a prerequisite condition that the current inter-syllable location is an inter-word.

2.1 Word length

$Q_1 2.1.1 \sim 4$: Is the preceding word an $n \in \{1, 2, 3, 4\}$ -syllable word?

$Q_1 2.1.5 \sim 8$: Is the following word an $n \in \{1, 2, 3, 4\}$ -syllable word?

$Q_1 2.1.9$: Is the length of the preceding word in syllable greater than 4?

$Q_1 2.1.10$: Is the length of the following word in syllable greater than 4?

2.2 Level-1 POS and special tags

$Q_1 2.2.1 \sim 11$: Is the POS of the preceding word A/C/D/N/I/P/T/V/DE/SHI/DM?

$Q_1 2.2.12 \sim 22$: IS the POS of the following word A/C/D/N/I/P/T/V/DE/SHI/DM?

2.3 Level-2 POS

$Q_1 2.3.1 \sim 33$: Is the POS of the preceding word
Ca/Cb/Da/Db/Dc/Dd/Df/Dg/Dh/Di/Dj/Dk/Na/Nb/Nc/Nd/Ne/Nf/Ng/Nh/VA/VB/VC/VD/VE/VF/V
G/VH/VI/VJ/VK/VL/V_2?

$Q_1 2.3.34 \sim 66$: Is the POS of the following word
Ca/Cb/Da/Db/Dc/Dd/Df/Dg/Dh/Di/Dj/Dk/Na/Nb/Nc/Nd/Ne/Nf/Ng/Nh/VA/VB/VC/VD/VE/VF/V
G/VH/VI/VJ/VK/VL/V_2?

2.4 Level-3 POS

$Q_1 2.4.1 \sim 15$: Is the POS of the preceding word
Caa/Cab/Cba/Cbb/Dfa/Dfb/Ncd/Neu/Nes/Nep/Neq/VA2/VC1/VH16/VH22?

$Q_1 2.4.16 \sim 30$: Is the POS of the following word
Caa/Cab/Cba/Cbb/Dfa/Dfb/Ncd/Neu/Nes/Nep/Neq/VA2/VC1/VH16/VH22?

2.5 Combination of POS

$Q_1 2.5.1 \sim 7$: Does the POS of the preceding word belong to {Da, Db, Dc, Dd, Dg, Dh, Di, Dj, Dk}/{Na, Nb, Nc}/{Ncd, Ng}/{I, T}/{VA, VG}/{VB, VC, VD, VE, VF, VJ, VK, VL}/{VH, VI}?

$Q_1 2.5.8 \sim 14$: Does the POS of the following word belong to {Da, Db, Dc, Dd, Dg, Dh, Di, Dj, Dk}/{Na, Nb, Nc}/{Ncd, Ng}/{I, T}/{VA, VG}/{VB, VC, VD, VE, VF, VJ, VK, VL}/{VH, VI}?

附錄四

The question set Θ_2 used to construct the decision trees for building the break-syntax model $P(B_n | \mathbf{I}_n)$ is listed below:

1. Syllable Level

$Q_2 1.1$: Is the initial of the following syllable a null one or in $\{m, n, l, r\}$?

$Q_2 1.2$: Is the inter-syllable location an inter-word?

$Q_2 1.3$: Is the inter-syllable location a intra-word?

2. Word Level

All the following questions are subject to a prerequisite condition that the current inter-syllable location is an inter-word.

2.1 Word length

$Q_2 2.1.1 \sim 4$: Is the preceding word an $n \in \{1, 2, 3, 4\}$ -syllable word?

$Q_2 2.1.5 \sim 8$: Is the following word an $n \in \{1, 2, 3, 4\}$ -syllable word?

$Q_2 2.1.9$: Is the length of the preceding word in syllable greater than 4?

$Q_2 2.1.10$: Is the length of the following word in syllable greater than 4?

2.2 Level-1 POS and special tags

$Q_2 2.2.1 \sim 11$: Is the POS of the preceding word A/C/D/N/I/P/T/V/DE/SHI/DM?

$Q_2 2.2.12 \sim 22$: IS the POS of the following word A/C/D/N/I/P/T/V/DE/SHI/DM?

2.3 Level-2 POS

$Q_2 2.3.1 \sim 33$: Is the POS of the preceding word Ca/Cb/Da/Db/Dc/Dd/Df/Dg/Dh/Di/Dj/Dk/Na/Nb/Nc/Nd/Ne/Nf/Ng/Nh/VA/VB/VC/VD/VE/VF/VG/VH/VI/VJ/VK/VL/V_2?

$Q_2 2.3.34 \sim 66$: Is the POS of the following word Ca/Cb/Da/Db/Dc/Dd/Df/Dg/Dh/Di/Dj/Dk/Na/Nb/Nc/Nd/Ne/Nf/Ng/Nh/VA/VB/VC/VD/VE/VF/VG/VH/VI/VJ/VK/VL/V_2?

2.4 Level-3 POS

Q_2 2.4.1~15 : Is the POS of the preceding word Caa/Cab/Cba/Cbb/Dfa/Dfb/Ncd/Neu/Nes/Nep/Neq/VA2/VC1/VH16/VH22?

Q_2 2.4.16~30 : Is the POS of the following word Caa/Cab/Cba/Cbb/Dfa/Dfb/Ncd/Neu/Nes/Nep/Neq/VA2/VC1/VH16/VH22?

2.5 Combination of POS

Q_2 2.5.1~7 : Does the POS of the preceding word belong to {Da, Db, Dc, Dd, Dg, Dh, Di, Dj, Dk}/{Na, Nb, Nc}/{Ncd, Ng}/{I, T}/{VA, VG}/{VB, VC, VD, VE, VF, VJ, VK, VL}/{VH, VI}?

Q_2 2.5.8~14 : Does the POS of the following word belong to {Da, Db, Dc, Dd, Dg, Dh, Di, Dj, Dk}/{Na, Nb, Nc}/{Ncd, Ng}/{I, T}/{VA, VG}/{VB, VC, VD, VE, VF, VJ, VK, VL}/{VH, VI}?



附錄五

The question set Θ_3 used to construct the decision trees for tone classifier is listed below:

1. Break type

- $Q_31.1$: Is the following syllable juncture labeled B0?
- $Q_31.2$: Is the following syllable juncture labeled B1?
- $Q_31.3$: Is the following syllable juncture labeled B2-1?
- $Q_31.4$: Is the following syllable juncture labeled B2-2?
- $Q_31.5$: Is the following syllable juncture labeled B2-3?
- $Q_31.6$: Is the following syllable juncture labeled B3?
- $Q_31.7$: Is the following syllable juncture labeled B4?
- $Q_31.8$: Is the following syllable juncture in end of sentence?
- $Q_31.9$: Is the preceding syllable juncture labeled B0?
- $Q_31.10$: Is the preceding syllable juncture labeled B1?
- $Q_31.11$: Is the preceding syllable juncture labeled B2-1?
- $Q_31.12$: Is the preceding syllable juncture labeled B2-2?
- $Q_31.13$: Is the preceding syllable juncture labeled B2-3?
- $Q_31.14$: Is the preceding syllable juncture labeled B3?
- $Q_31.15$: Is the preceding syllable juncture labeled B4?
- $Q_31.16$: Is the preceding syllable juncture in start of sentence?

2. Adjacent Tones

- $Q_32.1$: Is the following syllable tone1?
- $Q_32.2$: Is the following syllable tone2?
- $Q_32.3$: Is the following syllable tone3?

$Q_3 2.4$: Is the following syllable tone4?

$Q_3 2.5$: Is the following syllable tone5?

$Q_3 2.6$: Is the preceding syllable tone1?

$Q_3 2.7$: Is the preceding syllable tone2?

$Q_3 2.8$: Is the preceding syllable tone3?

$Q_3 2.9$: Is the preceding syllable tone4?

$Q_3 2.10$: Is the preceding syllable tone5?

