

國立交通大學

電信工程學系

碩士論文

音節間連音狀態自動標記模型研究

A Study on Auto-Labeling Model for Coarticulation
State between Syllables

研究生：陳慶安

指導教授：陳信宏 博士

中華民國九十七年八月

音節間連音狀態自動標記模型研究
A Study on Auto-Labeling Model for Coarticulation
States between Syllables

研 究 生：陳慶安

Student : Ching-An Chen

指導教授：陳信宏 博士

Advisor : Dr. Sin-Horng Chen

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

August 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年八月

音節間連音狀態自動標記模型研究

研 究 生：陳慶安

指導教授：陳信宏 博士

國立交通大學電信工程學系碩士班

中文摘要



隨著計算機之運算速度及記憶體容量的加快與增大，以大型語料庫為基礎之語音合成系統已成為目前最普遍且品質最好的語音合成系統，其方法是先輸入欲合成之文字後，再經過分析得到語言參數，接著在大型語料庫中依據所得到之語言參數找到對應之候選合成單元組，最後透過語音合成器挑選出最合適的單元組進行串接以得到合成語音。然而在單元選取時，常因為所挑選的單元與目標句之前後文字不同亦或因為合成單元本身受到連音效應影響等因素而造成合成語音在聽覺上之不適。為了改善這些缺點，在本文中我們提出利用頻譜參數來建構**音節頻譜模型**並同時標記出中文語料庫中音節間連音狀態。在模型中，我們考慮了三種影響音節頻譜之因素：基本音節類型，前、後音節類型和其音節間之連音程度，我們假設這些影響因素獨立且有加成性，在重複訓練後模型之影響因素能夠有不錯的學習效果，除此之外由模型標記之**頻譜連音狀態**在韻律參數以及語言參數上也能有合理的解釋。此方法在未來可以應用在語音合成系統中單元挑選的部份來幫助提升合成語音的品質。

A Study on Auto-Labeling Model for Coarticulation State between Syllables

Student : Ching-An Chen

Advisor : Dr. Sin-Horng Chen

Department of Communication Engineering
National Chiao Tung University

Abstract

As the computation power and the memory capacity increase, the corpus-base speech synthesis system has become the best and most popular speech synthesis system. Based on the system, the linguistic features are first derived after the text is parsed, then some appropriate units are selected as candidates. Finally, the well-pronounced speech is synthesized by concatenating the best unit sequence by the synthesizer part of system. In the unit selection process, the smooth-less places of synthesized speech usually caused by choosing the units which have different context with target units, or because the coarticulation effecting influencing. In this paper, to solve these problems, we use MFCC features to construct syllable spectral model and labeling coarticulation state between syllables in Chinese corpus at the same time. In this model, we have considered the three kinds of affecting factors with syllable spectral: the basic syllable type of current syllable, the coarticulation affecting from previous and following syllable, we assume that these three factors are independent and additive. After well-training, the affecting factor patterns could have good performance in model learning, besides the updated coarticulation states have reasonable explanation by prosody features and linguistic features. This method can improve the performance of synthesized speech by apply to unit selection process of using TTS system.

目錄

中文摘要.....	I
英文摘要.....	II
目錄.....	III
表目錄.....	V
圖目錄.....	VI
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究方向.....	1
1.3 章節概要說明.....	2
第二章 合成單元之選取.....	3
2.1 影響合成品質之因素.....	3
2.2 文獻回顧.....	3
2.2.1 依據決策樹挑選.....	4
2.2.2 依據 Cost Function 挑選.....	4
第三章 音節模型的訓練和數學說明.....	5
3.1 語料庫簡介.....	5
3.2 音節參數表示.....	5
3.3 音節模型的原理.....	6
3.4 音節類型、Final-Initial pair、連音狀態的種類設定.....	8
3.4.1 音節類型：.....	8
3.4.2 Final-Initial pair 類型：.....	8
3.4.3 連音狀態類型：.....	10
3.5 音節模型數學式.....	18

3.6 各影響因素的初始值：.....	20
3.7 各影響因素的更新值：.....	22
3.8 更新標記音節間連音狀態數學式.....	24
第四章 音節模型實驗結果與分析.....	27
4.1 初始影響因素之分析.....	28
4.1.1 初始連音狀態對頻譜參數影響分析.....	28
4.1.2 不同連音狀態種類影響之程度分析：.....	29
4.2 模型收斂後影響因素之分析.....	35
4.2.1 不同連音狀態種類影響之程度分析：.....	37
4.3 連音狀態之分析.....	41
4.3.1 韻律、頻譜連音狀態之比較：.....	41
4.3.2 Break Type 與頻譜連音狀態之分析.....	44
4.3.3 音節長度與頻譜連音狀態之分析.....	45
4.3.4 頻譜連音狀態與決策樹分析.....	45
第五章：結論與未來展望.....	52
參考文獻.....	53
附錄 音節類型與注音對照表.....	54

表目錄

表 3.1：錄音軟硬體格式表.....	5
表 3.2：音節類型分類表.....	8
表 3.3：韻母分類表，依照發音特性分成 17 類.....	9
表 3.4：聲母分類表，依照發音特性分成 7 類.....	10
表 3.5：空聲母分類表，依照發音特性分成 7 類.....	10
表 3.6：Break Type 與韻律連音狀態對照表	17
表 4.1：Break Type 與頻譜連音狀態對應表	44
表 4.2：正規化音長與頻譜連音狀態對應表.....	45



圖目錄

圖 3.1：音節 state 切割位置示意圖	6
圖 3.2：音節模型示意圖	7
圖 3.3：相關韻律資訊示意圖	11
圖 3.4：Break Type 分類決策樹示意圖	12
圖 3.5：相關 pause-duration pdfs 求取流程圖	13
圖 3.6：相關 normalized pitch jump pdfs 求取流程圖	15
圖 3.7：相關 energy-dip pdfs 求取流程圖	16
圖 3.8：漢語語音階層韻律結構圖	16
圖 3.9：音節模型之訓練流程圖	24
圖 4.1：Residual Variance 之比較圖	28
圖 4.2：Residual Variance 之比較圖	29
圖 4.3：連音狀態為 C1 時 Residual Variance V1 與 V2 前 3 states 之比較圖	30
圖 4.4：連音狀態為 C2 時 Residual Variance V1 與 V2 前 3 states 之比較圖	31
圖 4.5：連音狀態為 C3 時 Residual Variance V1 與 V2 前 3 states 之比較圖	31
圖 4.6：連音狀態為 C1 時 Residual Variance V1 與 V2 後 5 states 之比較圖	32
圖 4.7：連音狀態為 C2 時 Residual Variance V1 與 V2 後 5 states 之比較圖	33
圖 4.8：連音狀態為 C3 時 Residual Variance V1 與 V2 後 5 states 之比較圖	33
圖 4.9：log-likelihood 值 Q 之更新曲線	35
圖 4.10：Residual Variance V2 之比較圖	36
圖 4.11：連音狀態為 C1 時 Residual Variance V1 與 V2 前 3 states 之比較圖	37
圖 4.12：連音狀態為 C2 時 Residual Variance V1 與 V2 前 3 states 之比較圖	38
圖 4.13：連音狀態為 C3 時 Residual Variance V1 與 V2 前 3 states 之比較圖	38
圖 4.14：連音狀態為 C1 時 Residual Variance V1 與 V2 後 5 states 之比較圖	39

圖 4.15：連音狀態為 C2 時 Residual Variance V1 與 V2 後 5 states 之比較圖	40
圖 4.16：連音狀態為 C3 時 Residual Variance V1 與 V2 後 5 states 之比較圖	40
圖 4.17：韻律連音狀態之 Energy-dip discrete probability 分佈圖	42
圖 4.18：頻譜連音狀態之 Energy-dip discrete probability 分佈圖	42
圖 4.19：韻律連音狀態之 Pause-duration discrete probability 分佈圖	43
圖 4.20：頻譜連音狀態之 Pause-duration discrete probability 分佈圖	43
圖 4.21：決策樹分列式意圖	47
圖 4.22：連音狀態決策樹分裂圖	49



第一章 緒論

1.1 研究動機

人和人之間的溝通是透過說與聽來達到資訊的交流，而現在的電腦技術也已經朝不需透過鍵盤及滑鼠來操作的目標發展，而在現今的技術中語音合成和語音辨識的研究就是讓人和機器也可以透過最自然說與聽的方式來溝通。

文字轉語音合成 (Text-to-Speech Synthesis, TTS) 是一種普遍被使用的語音合成技術。而交通大學語音實驗室近年來已發展以大型語料庫為基礎的中文文字轉語音合成系統 (Corpus-based Mandarin TTS System)，此論文的動機便是希望能夠將所提出的方法應用在上面，來提高合成語音的品質。



1.2 研究方向

在 TTS 系統中，選取合成單元時常使用到 Cost Function 或是 Decision Tree 利用語料庫中語言參數的資訊來選取最合適的合成單元，在論文中我們提出從訊號頻譜的角度來幫助合成單元挑選，方法是利用頻譜參數(i.e., MFCC, 梅爾倒頻譜參數)來訓練音節頻譜模型，之後在給定的條件下便能夠預估音節頻譜參數，如此一來在單元選取時就能增添頻譜上的考量。

而合成單元間的連音狀態 (Coarticulation State) 也是在選取合成單元時會考量的一項重要因素。一般來說我們會用音節間的停頓長度 (Pause duration)、音節間能量最低點 (Energy-Dip Level) ... 等韻律參數 (Prosodic Feature) 來標記連音狀態，在本論文中我們希望能分析訊號頻譜與韻律連音狀態之間的關係，再進一步觀察從頻譜連音狀態與語言參數之間的關聯性，目的希望能讓 TTS 在合成時，能夠多考量了訊號頻譜上面的資訊來達到更好的合成效果。

1.3 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方向

第二章：合成單元之選取

第三章：音節模型的訓練和數學說明

第四章：音節模型實驗結果與連音狀態分析

第五章：結論與未來展望



第二章 合成單元之選取

當我們要合成一段語音的時候，需要由大型語料庫中選取適用的合成單元，此時可能會有好幾個候選合成單元，所以我們必須在所有可能的組合中選出讓合成品質最自然流暢的合成單元。

2.1 影響合成品質之因素

通常不流暢之合成聲音品質不佳的原因主要來自合成單元連接處之韻律或頻譜不連續，在 [11]一書中作者將造成所謂“不連續”的因素歸納為以下四類：

1. 前後連接音不同 (Differences in Phonetic Contexts)：

合成單元原本的前後文和欲合成之目標句前後文發音特性不同將會造成音質不連續。

2. 音節切割位置不正確 (Incorrect Segmentation)：

切割位置錯誤會造成頻譜上的不連續。

3. 聲學特性上的變動 (Acoustic Variability)：

譬如錄音時的說話快慢以及錄音環境等也是容易造成頻譜不連續的因素之一。

4. 韻律上的差異 (Different Prosody)：

合成單元連接處左右之韻律參數不連續也是讓合成音質不良的原因。

由此看來影響語音合成之品質的原因很多，所以單元選取的方法便是一項重要的課題，在 2.2 節中將提到一般單元選取主要會使用的方法。

2.2 文獻回顧

如何挑出最合適的候選單元使合成品質最自然流暢是眾多學者研究的方向之一，大舉來說有兩種主要的方法，一類是用決策樹 (Decision Tree)方式作為挑選之依據，一類是用 cost

function 來選出候選單元。

2.2.1 依據決策樹挑選

決策樹主要可分為兩種：分類樹 (Classification Tree)和迴歸樹 (Regression Tree)。分類樹的目的在於區分成一離散的類別數值；迴歸樹目的則在分析已得到的連續型參數值，至於挑選合成單元的應用上，決策樹主要依據語言參數 (Linguistic Feature)來挑選合成單元，其中語言參數包含韻律參數、詞邊界、呼吸邊界等，因此比較接近迴歸樹的應用。

2.2.2 依據 Cost Function 挑選

此類方法主要是建構一個數學式來評估不同候選單元合成時所造成的誤差值。而影響合成聲音品質的誤差主要分為兩類：

(1)單元誤差 (Unit Cost)：其差異的定義有很多方式，其中包含考慮合成單元和目標句之間語言參數的差異 [8]，或韻律參數的差異。舉例來說，語言參數包含合成單元在句子、或詞的位置，以及前後文發音特性等；韻律參數包含音節長度、音節能量、及音節平均音高 (pitch mean)等。

(2)轉移誤差 (Transition Cost)：是為合成單元與合成單元之間的差異。合成目標句的各個合成單元常常是從不同語句選出，所以在單元連接處容易發生韻律和頻譜不平滑的情況。評估轉移誤差也有很多方法，其中包含計算合成單元連接處之頻譜參數的距離 [10] (Spectral Distance)，如果頻譜之距離值愈大則轉移誤差值就會愈大。

Cost Function 的定義中還需要在以上兩種誤差之間設定合適的權重(weighting)，最後便是選擇使誤差總和最小的單元組合。

在論文中，我們所提出的音節頻譜模型功用為對某一音節來說，在給定其前、後音節類型以及音節間的連音狀態後能夠得到此音節頻譜參數之預估值，在單元選取中我們便能從許多候選詞中挑選出使頻譜參數誤差最小的單元。在第三章裡，我們將說明此音節頻譜模型所

考量會影響頻譜參數之三種因素，以及模型的訓練方式和其在數學上代表的定義。

第三章 音節模型的訓練和數學說明

在本章中將會簡單介紹訓練模型時所使用的語料以及音節參數的表示方式，接著會說明音節頻譜模型的原理以及訓練過程並列出其在數學上代表之意義。

3.1 語料庫簡介

在本論文中所使用的語料庫文字內容是取自中央研究院中文文句結構樹資料庫 1.1 版 (Sinica Treebank Version 1.1)，其中包含 429 段 Short Paragraphs，共有 56,653 個音節。音檔方面，我們請了一位專業的女性廣播員幫我們錄製，其錄音軟體設備及格式詳細如下：

表 3.1：錄音軟硬體格式表

錄音軟體	Cool Edit Pro 直接錄成聲音檔案
麥克風	單一指向性
錄音場所	普通房間
錄音情境	依照所選出文稿唸出
取樣頻率	20 kHz
發音速度	每秒約 4.6 個音節
取樣大小	16 bits
聲道	單聲道
檔案格式	Pcm

3.2 音節參數表示

在本論文中我們用 8 個狀態 (state) 來描述一個音節 HMM 模型，其中前 3 個 states 表

示音節聲母部份，後 5 個 states 則表示韻母部份。為了得到音節中每個 state 之切割位置，我們使用 HTK (Hidden Markov Model Toolkit) 軟體來訓練音節 HMM 模型並作切割音節動作，首先我們採用 HTK 說明手冊[2]中稱為 Isolated Word Model Toolkit 的訓練方法，使用的參數設定為：12 維的梅爾倒頻譜參數、1 維能量參數、frame size：20ms、frame shift：5ms。訓練語料是用經由人工校正過切割位置之 56,653 個音節。音節 HMM 模型訓練完後再對每一個音節進行至 state 位置的切割動作。最後我們以每個 state 中的所有 frame 的 12 維 MFCC 參數之平均值來表示該 state，所以每個音節的觀察值就可以用一個 $8 \times 12 = 96$ 維的 super vector 來表示。音節 state 之切割位置範例如圖所示：

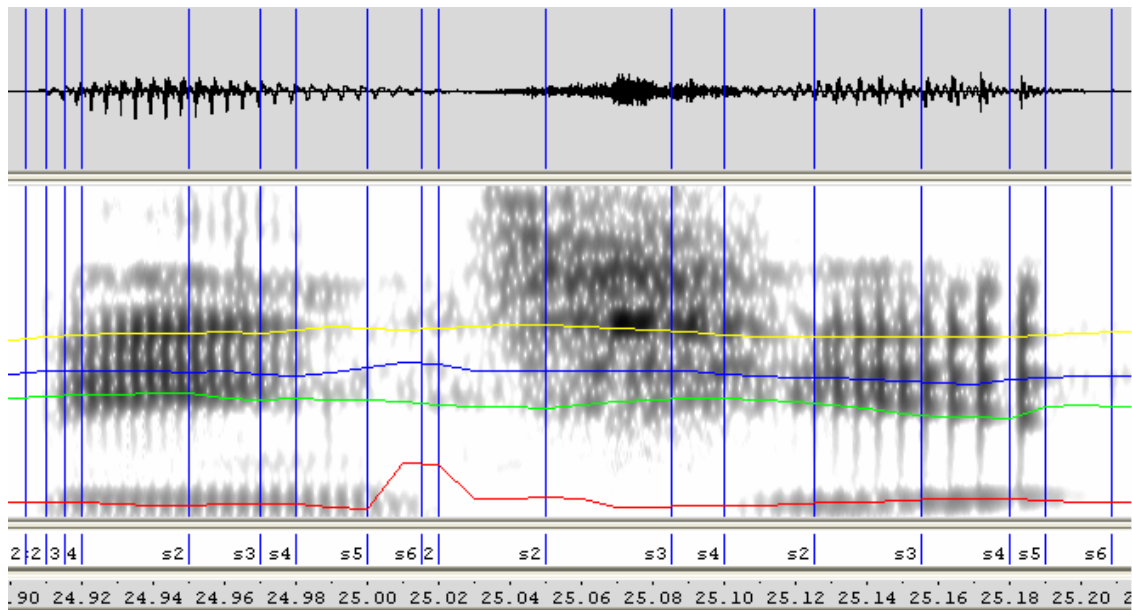


圖 3.1：音節 state 切割位置示意圖

3.3 音節模型的原理

我們考慮音節頻譜會受到本身基本音節類型影響，而且在連續語音中還容易受到前、後音節類型的影響，如此之外如果音節間連音程度愈嚴重則受前、後音節類型的影響愈大。基於以上的想法，在本論文中我們設計音節模型會受到 3 種因素影響，分別是：基本音節類型的影響因素，前、後音節的音節類型以及和前、後音節間的連音狀態的影響因素。我們假設所有的影響因素可以用累加的方式來表示，示意圖如下：

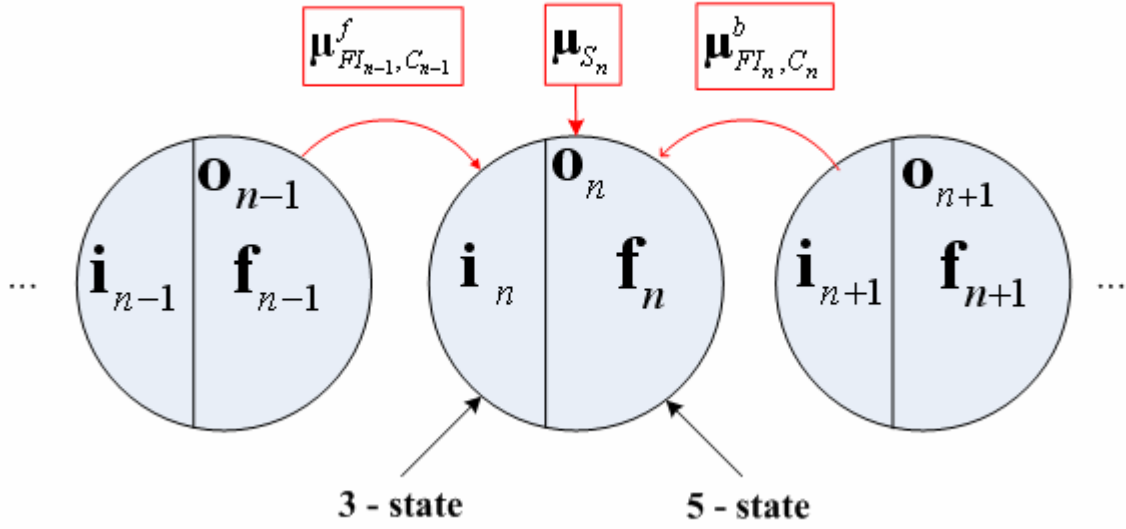


圖 3.2：音節模型示意圖

$$\mathbf{o}_n = \mathbf{r}_n + \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b \quad (3.1)$$

$\mathbf{o}_n$ ：第 n 個音節所呈現的特徵參數。

S_n ：第 n 個音節的基本音節類型。

$\boldsymbol{\mu}_{S_n}$ ：第 n 個音節受該音節類型之影響因素，以基本音節影響因素稱之。

C_n ：第 n 個音節與第 $n+1$ 個音節間的連音狀態。

FI_n ：Final-Initial pair，第 n 個音節的韻母種類與第 $n+1$ 個音節的聲母種類組合。

$\boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f$ ：與前音節之 Final-Initial pair 為 FI_{n-1} 以及連音狀態為 C_{n-1} 的情況下，第 n 個音節受前一音節之影響因素，以前音節影響因素稱之。

$\boldsymbol{\mu}_{FI_n, C_n}^b$ ：與後音節之 Final-Initial pair 為 FI_n 以及連音狀態為 C_n 的情況下，第 n 個音節受後一音節之影響因素，以後音節影響因素稱之。

$\mathbf{r}_n$ ：代表特徵參數向量 $\mathbf{o}_n$ 扣除上述三個影響因素後的殘餘值(residue)，也就是原始特徵參

數向量 $\mathbf{o}_n$ 對三個影響因素作正規化(normalized)後的值，我們可以令其機率分佈是為 zero mean ,covariance matrix $\mathbf{V}$ 的高斯分佈。

從圖 3.2 來看音節 $\mathbf{o}_n$ 受到 3 種因素影響(以框線所代表)，其中音節 $\mathbf{o}_n$ 之聲母部份受到音節 $\mathbf{o}_{n-1}$ 之韻母部份影響；音節 $\mathbf{o}_n$ 之韻母部份受到音節 $\mathbf{o}_{n+1}$ 之聲母部份影響。在 3.4 節中，我們將針對所使用的語料庫來設定模型中各影響因素的類型。

3.4 音節類型、Final-Initial pair、連音狀態的種類設定

在本節將會說明我們在音節模型中所設定之影響因素之分類數量和分類的依據。

3.4.1 音節類型：

音節類型分成 411 類，以下略舉(詳細對照表見附錄)。

表 3.2：音節類型分類表

音節類型	注音
1	ㄘ
2	ㄊ
3	ㄎ
4	ㄑ
...	...
411	ㄇㄛ

3.4.2 Final-Initial pair 類型：

我們知道在 411 種基本音節中可以將聲母部份分為 22 類、韻母部份分為 40 類，所以 Final-Initial pair 之組合有 $22 \times 40 = 880$ 種，然而對於我們所使用的語料庫來說這樣的類別數

量卻導致許多類別中的資料量太少，這樣的結果會不利於模型之訓練與分析。為了解決這個問題，我們依照發音特性將音節之 40 種韻母繼續分成 17 類、將 22 種聲母分成 7 類，其中被歸類為空聲母 (Initial NULL) 的音節因為其聲母發音特性不見得相似，舉例來說：”阿 (Y)”、“衣 (一)”、“屋 (ㄨ)”在聲母部份發音特性就不一樣。所以我們會將聲母為空聲母類型的音節再將其韻母分類 (也就是 **Final-Final** 之組合)，所以 **Final-Initial pair** 共分成 $17 \times (6 + 7) = 221$ 類，如此一來便解決大多數類別中資料量太少的問題。以下為我們使用的聲母、韻母分類表：

表 3.3：韻母分類表，依照發音特性分成 17 類

類別	韻母 (Final)	類別	韻母 (Final)
1	空韻母	10	ㄣ、一ㄣ、ㄨㄣ、ㄛㄣ
2	ㄩ、一ㄩ、ㄨㄩ	11	ㄣ、一ㄣ、ㄨㄣ、ㄛㄣ
3	ㄛ、一ㄛ、ㄨㄛ	12	ㄣ、一ㄣ、ㄨㄣ、ㄛㄣ
4	ㄣ	13	ㄣ、一ㄣ、ㄨㄣ、ㄛㄣ
5	ㄣ、一ㄣ、ㄛㄣ	14	一
6	ㄣ、一ㄣ、ㄨㄣ	15	ㄨ
7	ㄣ、ㄨㄣ	16	ㄛ
8	ㄣ、一ㄣ	17	ㄣ
9	ㄣ、一ㄣ		

表 3.4：聲母分類表，依照發音特性分成 7 類

類別	聲母 (Initial)
(鼻音_濁音) 1	ㄇ、ㄋ、ㄌ、ㄇ
(摩擦音_清音) 2	ㄈ、ㄊ、ㄑ、ㄒ、ㄣ
(爆破音_不送氣) 3	ㄅ、ㄆ、ㄇ
(塞擦音_不送氣) 4	ㄌ、ㄘ、ㄗ
(爆破音_送氣) 5	ㄆ、ㄑ、ㄒ
(塞擦音_送氣) 6	ㄌ、ㄘ、ㄗ
7	空聲母

表 3.5：空聲母分類表，依照發音特性分成 7 類

類別	空聲母 (Initial NULL)
(複韻母且介音為“一”) 1	ㄟ、ㄟㄚ、ㄟㄝ、ㄟㄟ、ㄟㄣ、ㄟㄨ、ㄟㄣ、 ㄟㄣ、ㄟㄣ、ㄟㄣ、ㄟㄣ
(複韻母且介音為“ㄨ”) 2	ㄨ、ㄨㄚ、ㄨㄟ、ㄨㄣ、ㄨㄟ、ㄨㄣ、ㄨㄣ、 ㄨㄣ、ㄨㄣ
(複韻母且介音為“ㄛ”) 3	ㄛ、ㄛㄚ、ㄛㄣ、ㄛㄣ、ㄛㄣ
(單韻母且元音為“ㄚ”) 4	ㄚ、ㄣ、ㄣ、ㄟ、ㄣ
(單韻母且元音為“ㄟ”) 5	ㄟ、ㄣ
(單韻母且元音為“ㄣ”) 6	ㄣ、ㄣ、ㄣ、ㄣ
(單韻母且元音為“ㄟ”) 7	ㄟ、ㄣ

3.4.3 連音狀態類型：

連音狀態類型數量的設定同樣需要考量到所使用之語料庫的數量，在這裡我們簡單分為三類，代表連音程度嚴重、中等、輕微。

對於模型訓練來說，連音狀態之初始設定影響著模型訓練的效果。在我們所使用的語料

庫中沒有音節間連音程度的資訊。所以為了設定一個合理的初始值我們使用 prosody 的方法預估初始之連音狀態。作法是參考文獻 [1] 中作者所提出的 Prosody Break Type 分類方法，先將語料庫中每個音節間的停頓分成 6 類之後，我們再將連音程度相近的合併，將數量降為三類。為了與模型所標記之連音狀態有所區別，初始標記之連音狀態我們以**韻律連音狀態**稱之；而模型所標記之連音狀態以**頻譜連音狀態**稱之。在 3.4.3.1 小節中將介紹文獻中作者所提出的分類方法。

3.4.3.1 Prosody Break Type 分類方法

文獻 [1] 中作者利用音節與音節間距、以及其間距中能量最低點、基頻軌跡間距 (Pitch pause) 等韻律資訊，還有音節間停頓在詞句中的位置等參數來做為 Break Type 分類決策樹之問題，相關韻律資訊以及決策樹示意圖如下。

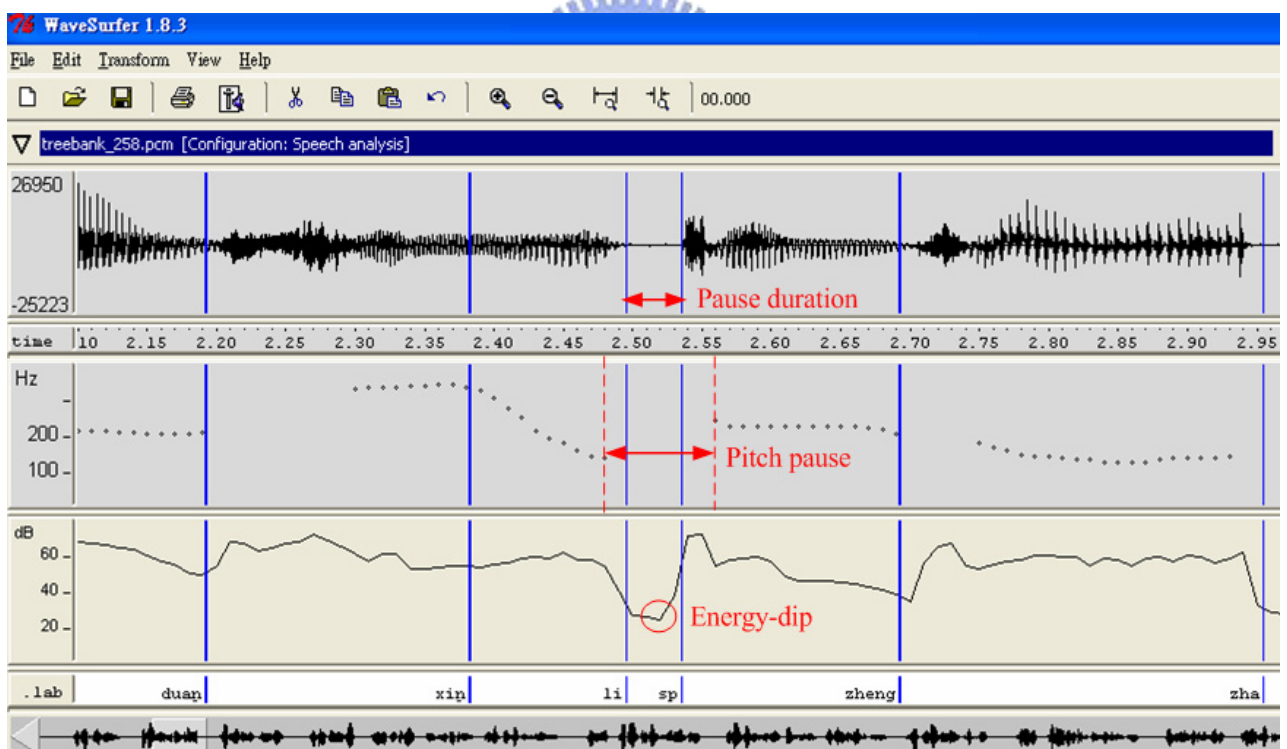


圖 3.3：相關韻律資訊示意圖

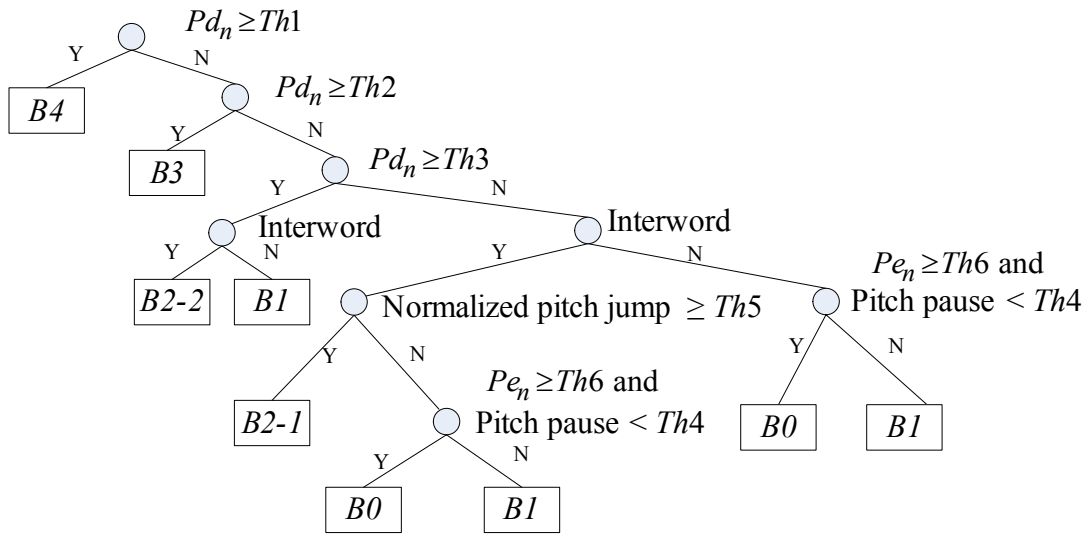


圖 3.4：Break Type 分類決策樹示意圖

(1) Th1、Th2 和 Th3 之定義

Th1、Th2 和 Th3 是分類出 B4、B3、和 B2-2/B1 之 pause-duration threshold。直觀上 juncture 位置在標點符號上時，pause-duration 會較大所以連音程度會比較輕微。我們將這類 pause-duration 做 vector quantization 分成二群，並且分別用 Gamma 分佈去 fitting，令平均值較大的分佈為 B4 之機率分佈 $f_{B4}(Pd)$ 、另一個則為 B3 之機率分佈 $f_{B3}(Pd)$ 。再者，如果 juncture 位置在 intra-word 時通常其連音程度會較嚴重，我們將這類 pause-duration 用 Gamma 分佈去 fitting，令其為 B0/B1 之機率分佈 $f_{B0/B1}(Pd)$ 。接下來我們將 juncture 位置在 inter-word 且不在標點符號上的一群加上 $f_{B3}(Pd_n) > f_{B0/B1}(Pd_n)$ 的條件令其為 B2-2，在此要加上條件原因是避免 B2-2 的 pause-duration 與 B0/B1 相似。相關 pause-duration pdfs 求取流程圖和 Th1, Th2 and Th3 之定義圖如下：

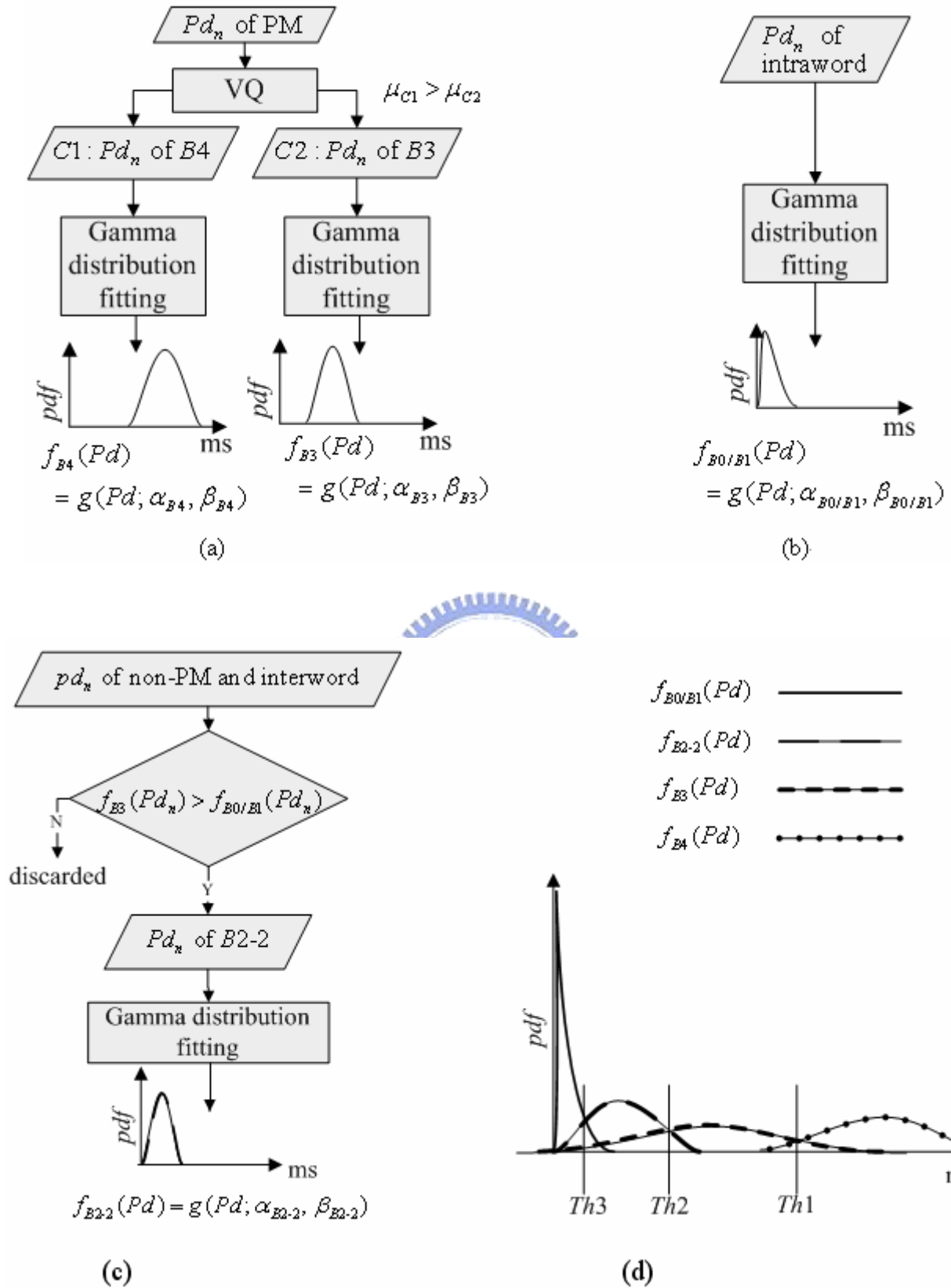


圖 3.5：相關 pause-duration pdfs 求取流程圖 (a) B3、B4； (b) B0/B1； (c) B2-2

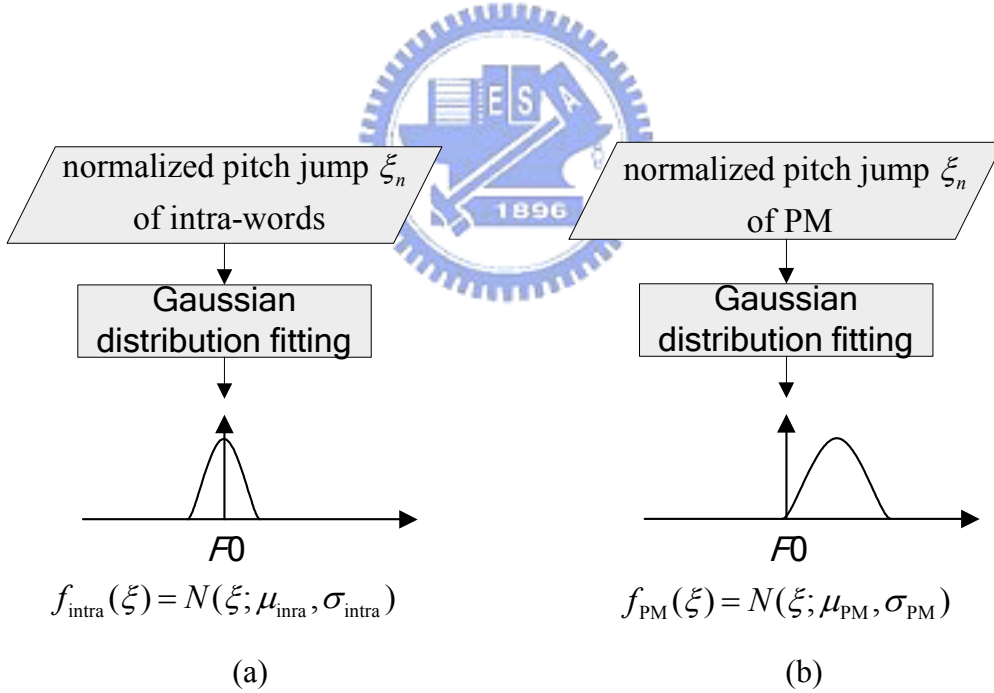
(d) Threshold $Th1$ 、 $Th2$ 、 $Th3$ 定義圖。

(2) Th5 之定義

Th5 是將 $\text{pause-duration} < Th3$ 的 juncture 再去分類出 B2-1 和 B0/B1 之 pitch jump threshold。以下是 normalized pitch jump 數學式：

$$\xi_n = (\mathbf{f}_{n+1} - \boldsymbol{\beta}_{t_{n+1}}) - (\mathbf{f}_n - \boldsymbol{\beta}_{t_n}) \quad (3.2)$$

其中 $\mathbf{f}_n$ 為第 n 個音節之 pitch 平均值； $\boldsymbol{\beta}_{t_n}$ 為第 n 個音節聲調的影響因素，其聲調影響因素為所有該聲調音節之 pitch 平均值。接著我們將位置在標點符號(PM)上之 juncture 的 normalized pitch jump 用高斯分佈 fitting，是為 $f_{\text{PM}}(\xi)$ ；intra-word juncture 之 normalized pitch jump 高斯分佈為 $f_{\text{intra}}(\xi)$ ；最後我們將 juncture 位置在 inter-word 且不在標點符號上的一群加上 $f_{\text{PM}}(\xi) > f_{\text{intra}}(\xi)$ 的條件令其為 B2-1，相關 normalized pitch jump pdfs 求取流程圖和 Th5 之定義圖如下：



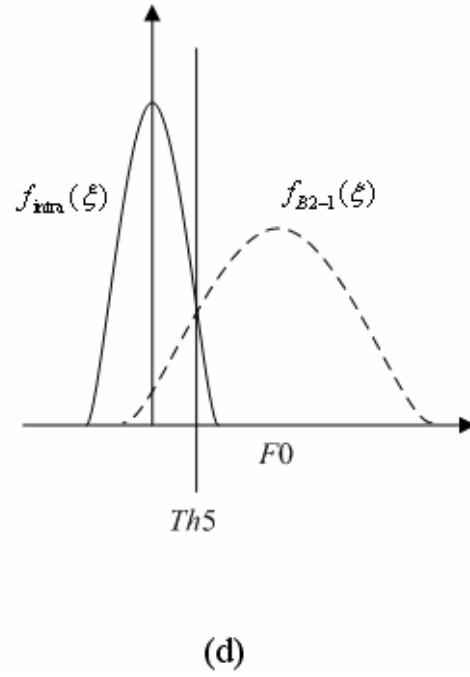
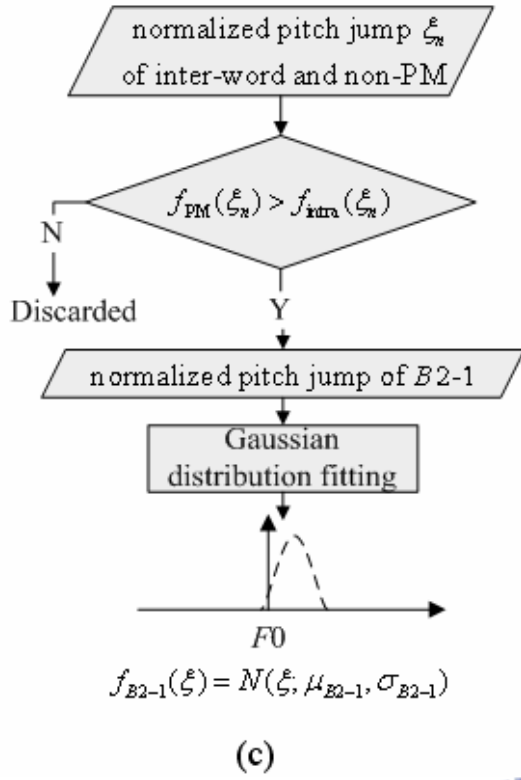


圖 3.6: 相關 normalized pitch jump pdfs 求取流程圖:(a) intra-word (b) PM (c) B2-1(d) Threshold Th5 之定義圖

(3) Th4、Th6 之定義

基本上連程度嚴重的 juncture 其 pitch pause 會很小而且 energy-dip 很大。所以我們令 Th4 為 1 個 frame 長 (=10ms)，也就是被分為 B0 類的 pitch pause 長度為 0。再來我們將剩餘未分類的 juncture 對其 energy-dip 做 VQ 分為二群並用高斯分佈 fitting，令平均值較高的那群為 B0 其 energy-dip 分佈為 $f_{B0}(Pe)$ ；另一群則為 B1，energy-dip 分佈 $f_{B1}(Pe)$ 。相關 energy-dip pdfs 求取流程和 Th6 之定義流程圖如下：

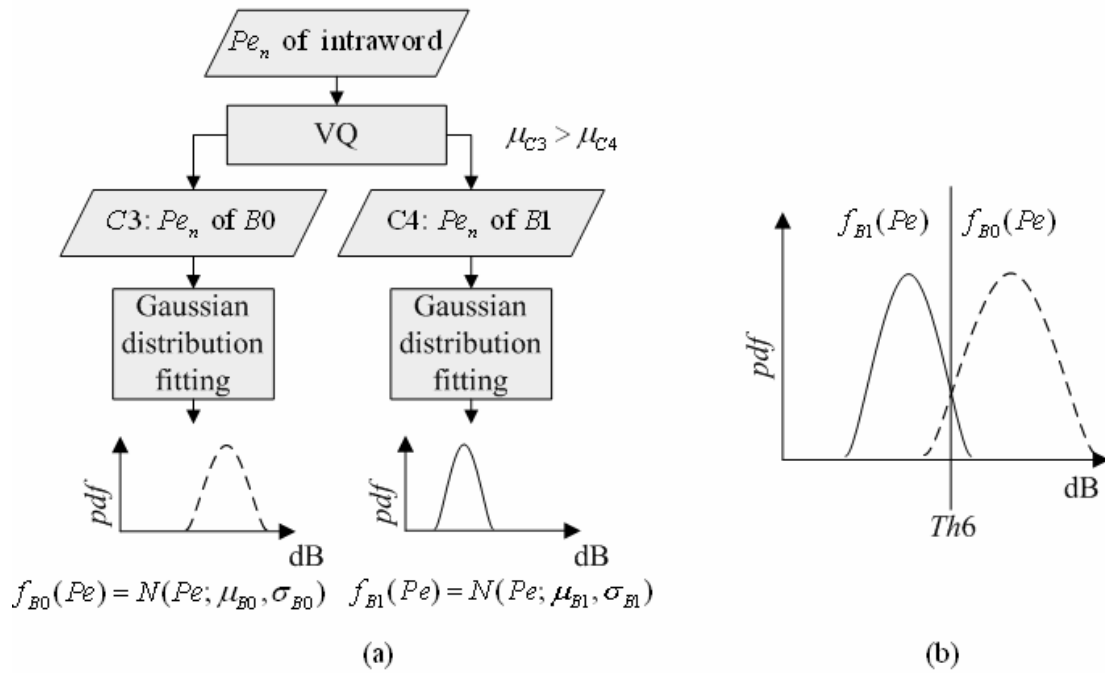


圖 3.7：相關 energy-dip pdfs 求取流程圖(a) B0、B1，(b) Threshold Th4、Th6 定義圖

3.4.3.2 Prosody Break Type 之合併

下圖為參考文獻[1]中作者所提出的漢語語音階層韻律結構圖：

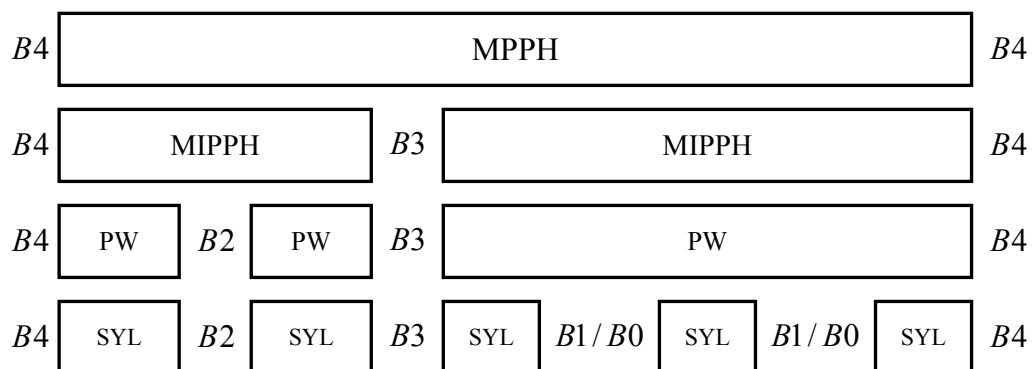


圖 3.8：漢語語音階層韻律結構圖

上圖中最低階層表示的是音節 (syllable, SYL)，此為韻律結構中最基本單元；第二階層中的韻律詞 (Prosody Word, PW)便是由一個或數個音節所構成；第三階層表示的是次要韻律片語 (Minor Prosody Phrase, MIPPH)，是由一個或數個

韻律詞所構成；最高階層是主要韻律片語 (Major Prosody Phrase, MPPH)。圖中所示的 $B0$ 、 $B1$ 、 $B2$ 、 $B3$ 和 $B4$ 為音節間的 Break Type， $B0$ 和 $B1$ 為韻律詞中音節間之停頓，其中 $B0$ 為 Tightly Couple； $B1$ 為 Normally Couple。 $B2$ 為韻律詞之間之停頓，依停頓特性又可再細分為兩類：(1) $B2-1$ ：音節間為 Minor Pitch Reset；(2) $B2-2$ ：音節間為 Short Pause。 $B3$ 為次要韻律片語之間之停頓，音節間為 Medium Pause，其通常亦伴隨著 Medium Pitch Reset。 $B4$ 為主要韻律片語之間之停頓，音節間為 Long Pause，其通常亦伴隨著 Large Pitch Reset。

接下來我們將以上 6 種 Break Type 將連音程度相近的合併成 3 類，如下表所示：

表 3.6：Break Type 與韻律連音狀態對照表

合併後之連音狀態類型	Break Type
C1 (連音程度最嚴重)	$B0$
C2 (連音程度中等)	$B1$ 、 $B2-1$
C3 (連音程度最輕微)	$B2-2$ 、 $B3$ 、 $B4$

3.5 音節模型數學式

本節描述音節模型中所有影響因素之初始和更新數學式。音節模型的機率分佈：

$$P(\mathbf{o} | C, S) \approx \prod_{n=1}^N P(\mathbf{o}_n | C, FI) \approx \prod_{n=1}^N P(\mathbf{o}_n | C_{n-1}^n, FI_{n-1}^{n+1}) \quad (3.3)$$

$\mathbf{o}$ ：觀察到的音節序列。

N ：總音節數， $N = 56653$ 。

C ：連音狀態序列。

S ：音節種類序列。

FI_n ：Final-Initial pair，第 n 個音節的韻母種類與第 $n+1$ 個音節的聲母種類組合。

一般來說殘餘值的機率分佈我們可以把它看成是 zero mean ,covariance matrix $\mathbf{V}$ 的高斯分佈，也就是 $R \sim N(\mathbf{r}_n; 0, \mathbf{V})$ 。所以我們可以將音節模型的機率分佈用下列數學式描述：

$$\begin{aligned} P(\mathbf{o}_n | C_{n-1}^n, FI_{n-1}^{n+1}) &= N(\mathbf{o}_n; \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b, \mathbf{V}) \\ &= \frac{1}{\sqrt{(2\pi)^D |\mathbf{V}|}} \exp\left(-\frac{1}{2}(\mathbf{o}_n - \mathbf{m})^T \mathbf{V}^{-1}(\mathbf{o}_n - \mathbf{m})\right) \end{aligned} \quad (3.4)$$

其中 $\mathbf{m} = \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b$ ， $D = 96$ supervector dimention。

就數學意義上來說，我們可以將音節模型的機率分佈 $\prod_{n=1}^N P(\mathbf{o}_n | C_{n-1}^n, S_{n-1}^{n+1})$ 視為概似函數 (Likelihood Function)，而我們目的就是如何對概似函數求出各個影響因素以及殘餘值的 covariance matrix $\mathbf{V}$ ，使得音節模型機率分佈有最大概似值(Maximum

Likelihood Value)。為了考量使往後的運算方便，我們計算其 log-likelihood 值。

所以我們首先對 $\prod_{n=1}^N P(\mathbf{o}_n | C_{n-1}^n, S_{n-1}^{n+1})$ 取自然對數 (Natural Logarithm)，數學式如下：

$$\begin{aligned}
 Q &= \log \left[\prod_{n=1}^N P(\mathbf{o}_n | C_{n-1}^n, FI_{n-1}^{n+1}) \right] = \sum_{n=1}^N \log \left[P(\mathbf{o}_n | C_{n-1}^n, FI_{n-1}^{n+1}) \right] \\
 &= \sum_{n=1}^N \log \left[N(\mathbf{o}_n; \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b, \mathbf{V}) \right] \\
 &= \sum_{n=1}^N \log \left[\frac{1}{\sqrt{(2\pi)^D |\mathbf{V}|}} \exp \left(-\frac{1}{2} (\mathbf{o}_n - \mathbf{m}_n)^T \mathbf{V}^{-1} (\mathbf{o}_n - \mathbf{m}_n) \right) \right] \\
 &= \sum_{n=1}^N \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| + \left(-\frac{1}{2} (\mathbf{o}_n - \mathbf{m}_n)^T \mathbf{V}^{-1} (\mathbf{o}_n - \mathbf{m}_n) \right) \right] \\
 &= -\frac{N \times D}{2} \log(2\pi) - \frac{N}{2} \log |\mathbf{V}| - \frac{1}{2} \sum_{n=1}^N (\mathbf{o}_n - \mathbf{m}_n)^T \mathbf{V}^{-1} (\mathbf{o}_n - \mathbf{m}_n)
 \end{aligned} \tag{3.5}$$

where $D = 96$, $|\mathbf{V}|$: determinant of $\mathbf{V}$, $\mathbf{m}_n = \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b$ 。

接下來，我們依序將 Q 對基本音節影響因素 $\boldsymbol{\mu}_{S_n}$ 、前音節影響因素 $\boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f$ 、後音節影響因素 $\boldsymbol{\mu}_{FI_n, C_n}^b$ 以及 covariance matrix $\mathbf{V}$ 做偏微分並使其為零，數學式如下：

(1) 基本音節影響因素：

$$\text{Let } \frac{\partial Q}{\partial \boldsymbol{\mu}_{S_n}} = 0 \Rightarrow \boldsymbol{\mu}_{S_n}^* = \frac{\sum_{n=1}^N (\mathbf{o}_n - \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f - \boldsymbol{\mu}_{FI_n, C_n}^b) \delta(S_n = S)}{\sum_{n=1}^N \delta(S_n = S)}, \quad S \in \{1, 2, \dots, 411\} \tag{3.6}$$

(2)前音節影響因素：

$$Let \frac{\partial Q}{\partial \mu_{FI_{n-1}, C_{n-1}}^f} = 0 \Rightarrow \mu_{FI_{n-1}, C_{n-1}}^{*f} = \frac{\sum_{n=1}^N (\mathbf{o}_n - \mu_{S_n} - \mu_{FI_n, C_n}^b) \delta(FI_{n-1} = \hat{f}_i, C_{n-1} = c)}{\sum_{n=1}^N \delta(FI_{n-1} = \hat{f}_i, C_{n-1} = c)}$$

where $\hat{f}_i \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ ° (3.7)

(3)後音節影響因素：

$$Let \frac{\partial Q}{\partial \mu_{FI_n, C_n}^b} = 0 \Rightarrow \mu_{FI_n, C_n}^{*b} = \frac{\sum_{n=1}^N (\mathbf{o}_n - \mu_{S_n} - \mu_{FI_{n-1}, C_{n-1}}^f) \delta(FI_n = \hat{f}_i, C_n = c)}{\sum_{n=1}^N \delta(FI_n = \hat{f}_i, C_n = c)} \quad (3.8)$$

where $\hat{f}_i \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ °

(4)Covariance Matrix $\mathbf{V}$:

$$\frac{\partial Q}{\partial \mathbf{V}} = 0 \Rightarrow \mathbf{V}^* = \frac{\sum_{n=1}^N (\mathbf{o}_n - \mathbf{m}_n)^T (\mathbf{o}_n - \mathbf{m}_n)}{N} \quad (3.9)$$

where $\mathbf{m}_n = \mu_{S_n} + \mu_{FI_{n-1}, C_{n-1}}^f + \mu_{FI_n, C_n}^b$ °

3.6 各影響因素的初始值：

在計算影響因素前我們合理的令所有影響因素皆為 0 接著代入上面推導後的數學式子，數學式如下：

(1)基本音節影響因素初始值：

$$\mu_{S_n} = \frac{\sum_{n=1}^N (\mathbf{o}_n - 0 - 0) \delta(S_n = s)}{\sum_{n=1}^N \delta(S_n = s)}, s \in \{1, 2, \dots, 41\} \quad (3.10)$$

(2)前音節影響因素初始值：

$$\mu_{FI_{n-1}, C_{n-1}}^f = \frac{\sum_{n=2}^N (\mathbf{o}_n - \mu_{S_n} - 0) \delta(FI_{n-1} = fi, C_{n-1} = c)}{\sum_{n=2}^N \delta(FI_{n-1} = fi, C_{n-1} = c)} \quad (3.11)$$

where $fi \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ 。

(3)後音節影響因素初始值：

$$\mu_{FI_n, C_n}^b = \frac{\sum_{n=1}^{N-1} (\mathbf{o}_n - \mu_{S_n} - \mu_{FI_{n-1}, C_{n-1}}^f) \delta(FI_n = fi, C_n = c)}{\sum_{n=1}^{N-1} \delta(FI_n = fi, C_n = c)} \quad (3.12)$$

where $fi \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ 。值得注意的是，在此音節模型中我們假設前音節影響因素只影響到音節的前 3 個 states；後音節影響因素只影響到音節的後 5 個 states (Initial：前 3 個 states；Final：後 5 個 states)。所以前、後音節影響因素值可以改寫為以下數學式：

(1)改寫後之前音節影響因素初始值：

$$\mu_{FI_{n-1}, C_{n-1}}^f = \frac{\sum_{n=2}^N (\mathbf{o}_n - \mu_{S_n})_1^{36} \delta(FI_{n-1} = fi, C_{n-1} = c)}{\sum_{n=2}^N \delta(FI_{n-1} = fi, C_{n-1} = c)} \quad (3.13)$$

where $fi \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ 。

(2)改寫後之後音節影響因素初始值：

$$\mu_{FI_n, C_n}^b = \frac{\sum_{n=1}^{N-1} (\mathbf{o}_n - \mu_{S_n})_{37}^{96} \delta(FI_n = fi, C_n = c)}{\sum_{n=1}^{N-1} \delta(FI_n = fi, C_n = c)} \quad (3.14)$$

where $fi \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ 。

3.7 各影響因素的更新值：

在模型的訓練過程中各影響因素不斷更新目的便是使概似值 Q 達到最大值，更新數學式如下：

(1)更新基本音節影響因素數學式：

$$\mu_{S_n}(i) = \frac{\sum_{n=1}^N (\mathbf{o}_n - \mu_{FI_{n-1}, C_{n-1}}^f(i-1) - \mu_{FI_n, C_n}^b(i-1)) \delta(S_n = s)}{\sum_{n=1}^N \delta(S_n = s)} \quad (3.15)$$

where $s \in \{1, 2, \dots, 411\}$, i : i -th iteration 。

(2)更新前音節影響因素數學式：

$$\mu_{FI_{n-1}, C_{n-1}}^f(i) = \frac{\sum_{n=2}^N (\mathbf{o}_n - \mu_{S_n}(i))_{36}^{36} \delta(FI_{n-1} = fi, C_{n-1} = c)}{\sum_{n=2}^N \delta(FI_{n-1} = fi, C_{n-1} = c)} \quad (3.16)$$

where $fi \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ 。

(3)更新後音節影響因素數學式：

$$\mu_{FI_n, C_n}^b(i) = \frac{\sum_{n=1}^{N-1} (\mathbf{o}_n - \mu_{S_n}(i))_{37}^{96} \delta(FI_n = fi, C_n = c)}{\sum_{n=1}^{N-1} \delta(FI_n = fi, C_n = c)} \quad (3.17)$$

where $fi \in \{(1,1), (2,1) \dots (16,13), (17,13)\}$, $c \in \{C1, C2, C3\}$ 。

除了影響因素的更新，在每次 iteration 時還要更新 covariance matrix $\mathbf{V}$ 後才能計算 log-likelihood 值 Q ，數學式如下：

(1)更新 covariance matrix $\mathbf{V}$ 數學式：

$$\mathbf{V}(I=i) = \frac{\sum_{n=1}^N (\mathbf{o}_n - \mathbf{m}_n(i))^T (\mathbf{o}_n - \mathbf{m}_n(i))}{N} \quad (3.18)$$

where $\mathbf{m}_n(i) = \boldsymbol{\mu}_{S_n}(i) + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f(i) + \boldsymbol{\mu}_{FI_n, C_n}^b(i)$

(2)更新 log-likelihood 值 Q 數學式：

$$Q(i) = \sum_{n=1}^N \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}(i)| - \frac{1}{2} \left((\mathbf{o}_n - \mathbf{m}_n(i))^T \mathbf{V}^{-1}(i) (\mathbf{o}_n - \mathbf{m}_n(i)) \right) \right] \quad (3.19)$$

我們必須設定收斂條件讓更新標記連音狀態的動作停止，以下為收斂條件數學式：

$$Q(i) - Q(i-1) = |Q(i-1)| \times 10^{-5}, i \geq 2 \quad (3.20)$$

以下圖 3.9 為整個音節模型之訓練流程圖，而連音狀態的更新動作將在 3.8 章說明。



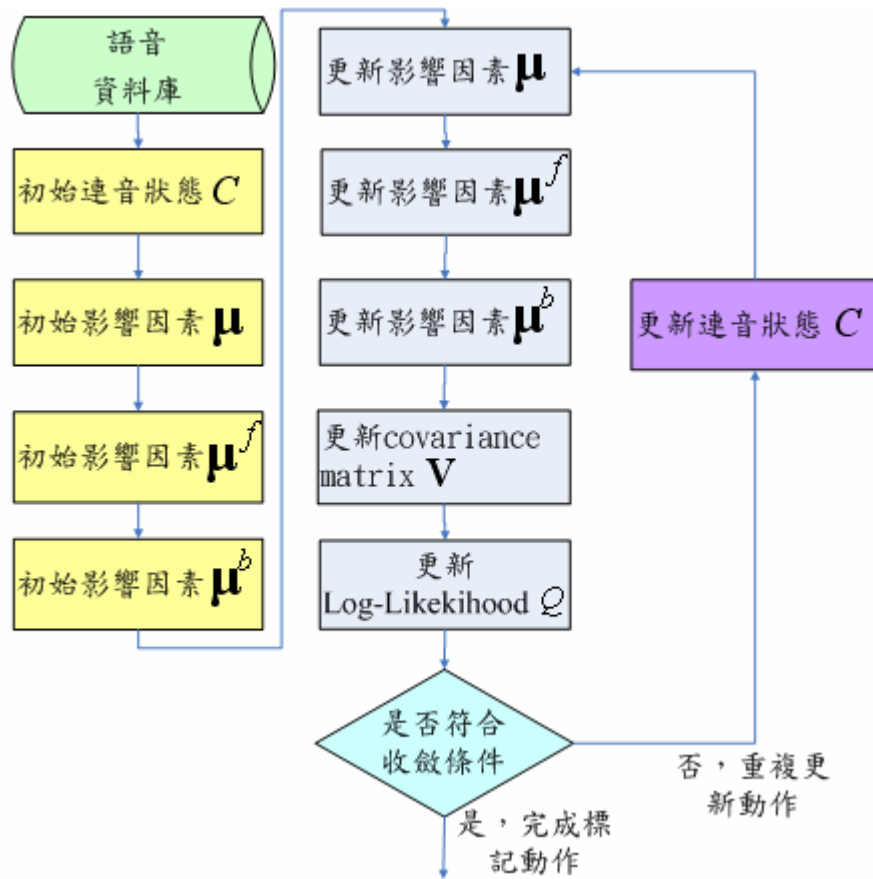


圖 3.9：音節模型之訓練流程圖

3.8 更新標記音節間連音狀態數學式

每一次在所有影響因素更新完成後必須更新連音序列才能讓音節模型之機率有最大概似值，更新連音狀態之數學式：

$$C^* = \arg \max_C P(\mathbf{O} | C, S) \quad (3.21)$$

我們可以將上面從數學式子解釋成求出使概似函數 $P(\mathbf{O} | C, S)$ 有最大概似值的連音狀態序列 C 解。換句話說就是求出連音狀態序列 C 使數學式(3.4)中的 Q 值最大，以下重述數學式 (3.4)：

$$\begin{aligned}
Q &= \sum_{n=1}^N \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \left((\mathbf{o}_n - \mathbf{m}_n)^T \mathbf{V}^{-1} (\mathbf{o}_n - \mathbf{m}_n) \right) \right] \\
&= \sum_{n=1}^N \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \left((\mathbf{o}_n - (\boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b))^T \mathbf{V}^{-1} (\mathbf{o}_n - (\boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b)) \right) \right]
\end{aligned} \tag{3.22}$$

在這裡要注意的是，因為音節模型裡每一個音節的後音節影響因素只影響其音節後 5 個 states，而其前音節影響因素只影響其音節前 3 個 states，而所有音節間 juncture 彼此是不互相影響的，所以我們可以把問題想成是求每一個 juncture 的前音節後 5 個 states 的概似值與後音節前 3 個 states 的概似值的和，我們以 Q' 稱之，那麼對於第 n 個 juncture 而言，我們的目的就是要求 Q'_n 的最大值。

對於單一音節模型 $P(\mathbf{o}_n | C_{n-1}^n, S_{n-1}^{n+1})$ 來說，我們令其單一音節之概似值為 Q_n ，其數學式敘述如下：

$$Q_n = -\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{o}_n - (\boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b))^T \mathbf{V}^{-1} (\mathbf{o}_n - (\boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b)) \tag{3.23}$$

我們利用數學式 (3.22) 將 Q'_n 用以下數學式表示：

$$\begin{aligned}
Q'_n &= \sum_{d=37}^{96} \left(-\frac{d}{2} \log(2\pi) \right) - \frac{1}{2} \log |\mathbf{V}_{(37,37)}^{(96,96)}| - \frac{1}{2} \left((\mathbf{o}_n - \mathbf{m}_n)^T \mathbf{V}^{-1} (\mathbf{o}_n - \mathbf{m}_n) \right)_{37}^{96} \\
&\quad + \sum_{d=1}^{36} \left(-\frac{d}{2} \log(2\pi) \right) - \frac{1}{2} \log |\mathbf{V}_{(1,1)}^{(36,36)}| - \frac{1}{2} \left((\mathbf{o}_{n+1} - \mathbf{m}_{n+1})^T \mathbf{V}^{-1} (\mathbf{o}_{n+1} - \mathbf{m}_{n+1}) \right)_1^{36}
\end{aligned} \tag{3.24}$$

where $\mathbf{m}_n = \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f + \boldsymbol{\mu}_{FI_n, C_n}^b$ 。在本論文中，為了增加連音狀態標記上的效果我們特別加上 Pause Acoustic Model $P(\mathbf{PD}, \mathbf{ED} | \mathbf{C})$ 的輔助，一般來說連音程度嚴重的 juncture 通常其 energy-dip 值會較高且 Pause duration 較短，反之亦然。以下為 Pause Acoustic Model $P(\mathbf{PD}, \mathbf{ED} | \mathbf{C})$ 數學式：

$$P(pd_n, ed_n | C_n) = g(pd_n; \alpha_{C_n}, \beta_{C_n}) N(ed_n; \mu_{C_n}, \sigma_{C_n}^2) \quad (3.25)$$

where

pd_n : pause duration between n-th syllable and (n+1)-th syllable

ed_n : energy-dip between n-th syllable and (n+1)-th syllable

最後我們整合數學式(3.23)並表示成以下數學式：

$$C_n^* = \arg \max_c \left(Q'_n + \log(P(pd_n, ed_n | C_n = c)) \right), c \in \{C1, C2, C3\} \quad (3.26)$$

更新連音狀態後我們便可以繼續更新影響因素，如此重複直到收斂便得到由頻譜參數標記的連音狀態序列值，我們將在下一章分析標記的結果。



第四章 音節模型實驗結果與分析

本章中實驗結果之分析主要有兩個部份：1.模型中影響因素的訓練效果、2.韻律、頻譜連音狀態之分析。

在這裡我們對全部音節考慮其影響因素後的殘餘值之變異量(Residual Variance)來觀察各影響因素所影響的程度。前面有提到過，我們把殘餘值的機率分佈看成是 zero mean ,covariance matrix $\mathbf{V}$ 的高斯分佈，也就是 $R \sim N(\mathbf{r}_n; 0, \mathbf{V})$ 。而為了往後方便表示，我們將音節殘餘值分成 3 類：

(1) $R0$ ：原始所觀察到的音節參數，也就是不扣除任何影響因素，其機率分佈如下：

$$R0 \sim N(\mathbf{r}_n; 0, V0 | \mathbf{r}_n = \mathbf{o}_n) \quad (4.1)$$

(2) $R1$ ：音節參數只考慮其基本音節影響因素後的殘餘值，其機率分佈如下：

$$R1 \sim N(\mathbf{r}_n; 0, V1 | \mathbf{r}_n = \mathbf{o}_n - \boldsymbol{\mu}_{S_n}) \quad (4.2)$$

(3) $R2$ ：音節參數考慮其基本音節影響因素和其前、後音節影響因素後的殘餘值，其機率分佈如下：

$$R2 \sim N(\mathbf{r}_n; 0, V2 | \mathbf{r}_n = \mathbf{o}_n - \boldsymbol{\mu}_{S_n} - \boldsymbol{\mu}_{FI_{n-1}, C_{n-1}}^f - \boldsymbol{\mu}_{FI_n, C_n}^b) \quad (4.3)$$

我們假設 MFCC 12 維參數的基底是互相正交的，所以我們只需觀察所有 Residual Covariance Matrix 主對角線上的值，並且為了簡潔的表示參數值在 state 上的變化程度，我們將主對角線上各屬於每一個 state 的 12 維參數值做相加的動作，以下圖 4.1 中我們觀察韻律連音狀態下之 Residual Variance：

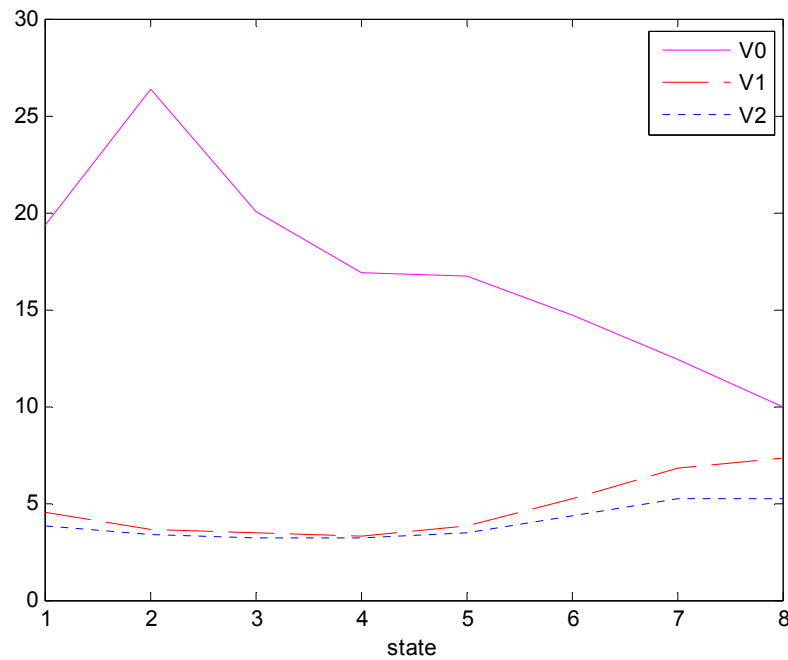


圖 4.1：Residual Variance 之比較圖

- 結果分析：上圖中 $V1$ 曲線呈現中間較低兩端較高之情況，這表示音節中愈接近前後音節之 state 所受的影響程度愈高，而 $V2$ 曲線表示音節考慮所有影響因素後 Residual Variance 能夠有效的繼續下降。

在圖 4.1 中模型在多考慮前、後音節影響因素後之曲線能夠繼續下降，但是在我們的假設中，前、後音節影響因素所影響音節頻譜的程度與音節間的連音狀態有關，在以下 4.1 節中我們進一步分析連音狀態與影響因素之關係。

4.1 初始影響因素之分析

4.1.1 初始連音狀態對頻譜參數影響分析

音節模型中，我們定義前、後音節影響因素包含連音狀態和有關係，接下來我們想知道的是從韻律參數預估的連音狀態是否真的會影響到頻譜。為了觀察連音狀態的有無對音節模型有沒有影響，我們特別定義另一個音節模型，其相關數

學式簡述如下：

$$\tilde{\mathbf{o}}_n = \mathbf{r}_n + \boldsymbol{\mu}_{S_n} + \boldsymbol{\mu}_{FI_{n-1}}^f + \boldsymbol{\mu}_{FI_n}^b \quad (4.4)$$

$\tilde{R}2$ 音節參數考慮其基本音節影響因素和其前、後音節影響因素後的殘餘值，其機率分佈如下：

$$\tilde{R}2 \sim N(\mathbf{r}_n; 0, \tilde{V}2 | \mathbf{r}_n = \mathbf{o}_n - \boldsymbol{\mu}_{S_n} - \boldsymbol{\mu}_{FI_{n-1}}^f - \boldsymbol{\mu}_{FI_n}^b) \quad (4.5)$$

我們可以發現新定義的音節模型其 Final-Initial Pair 是不考慮連音狀態的，接著我們比較兩種定義下 Residual Variance 的差異：

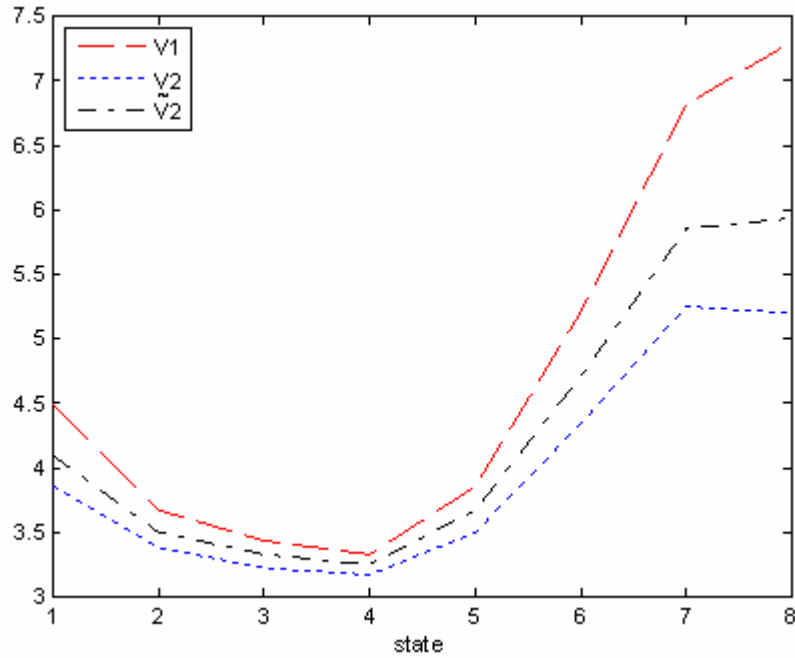


圖 4.2：Residual Variance 之比較圖

- 結果分析：從上圖我們可看出若兩者之曲線之 Residual Variance 皆與 $V1$ 比較的話， $V2$ 下降程度較 $\tilde{V}2$ 多，由此可知影響因素中韻律連音狀態的考量的確對音節模型有所貢獻。

4.1.2 不同連音狀態種類影響之程度分析：

從 4.2.1 小節我們知道韻律連音狀態會對頻譜參數造成影響，我們想再進一

步分析是否不同的程度連音狀態(C1：連音程度最嚴重、C2：連音程度次等、C3：連音程度最輕微)會對頻譜造成不同程度的影響。在連續語音中，音節的發音會受到其前後音節的種類不同有所影響，直觀上音節受到前後音節影響程度會隨著連音程度愈嚴重而愈高。作法如以下所述：

前音節影響因素之觀察：我們將所有音節依照其與前一音節間的連音狀態的種類分為3類，並分別比較其 Residual Variance $V1$ 與 Residual Variance $V2$ ，目的是為了觀察此3類音節的前音節影響因素整體的差異性，以下為實驗結果與分析：

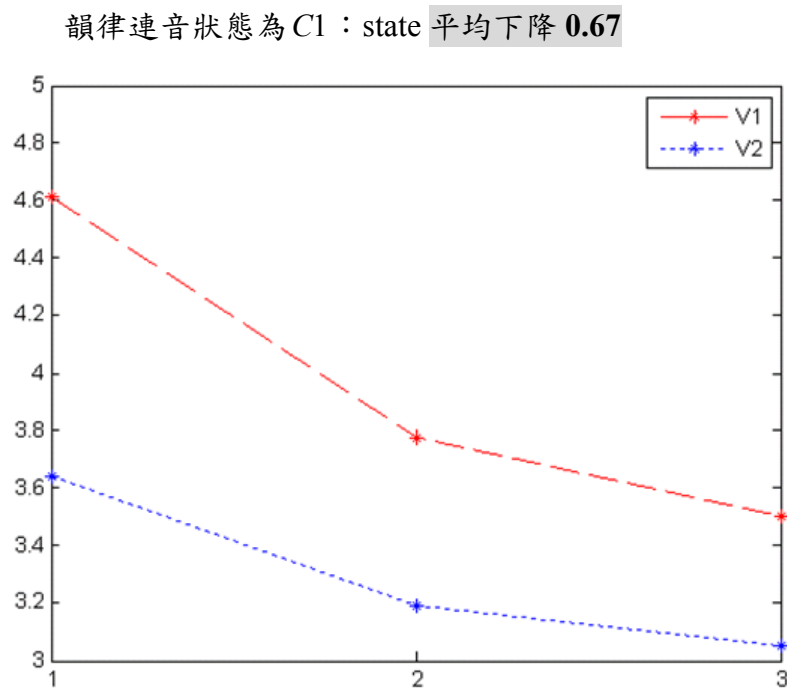


圖 4.3:連音狀態為 C1 時 Residual Variance $V1$ 與 $V2$ 前 3 states 之比較圖

韻律連音狀態為 **C2**： state 平均下降 **0.36**

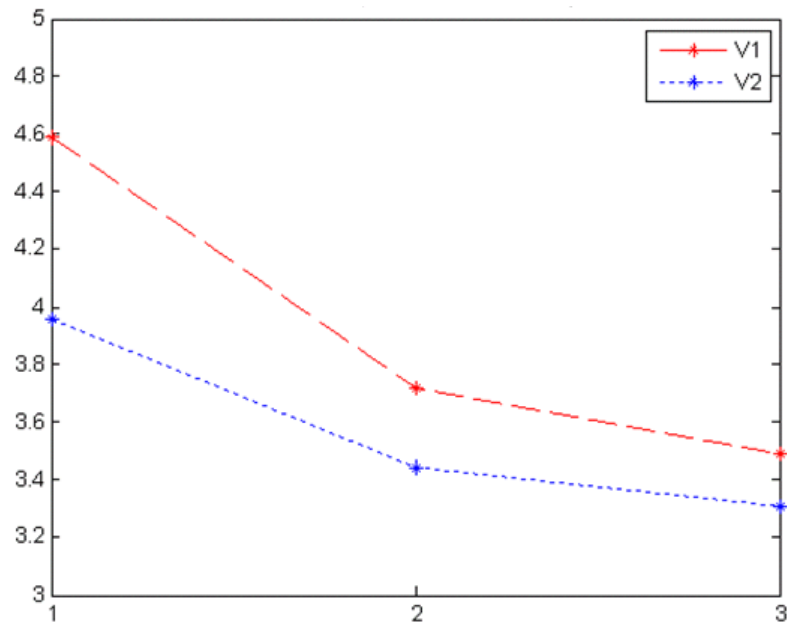


圖 4.4：連音狀態為 C2 時 Residual Variance V1 與 V2 前 3 states 之比較圖

韻律連音狀態為 **C3**： state 平均下降 **0.25**

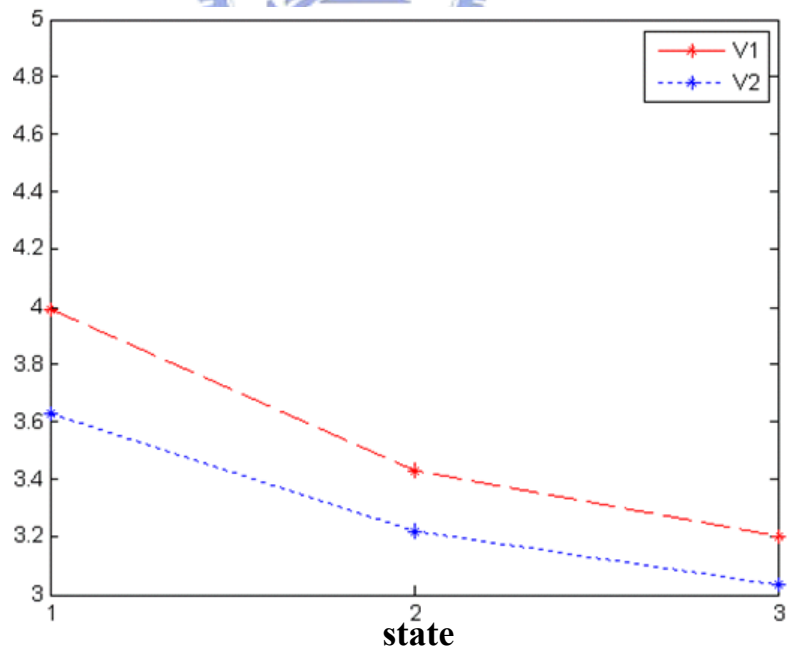


圖 4.5：連音狀態為 C3 時 Residual Variance V1 與 V2 前 3 states 之比較圖

●結果分析：從上圖我們發現前音節影響因素在連音狀態為 C1 那一類之 Residual Variance 下降幅度最多，而在連音狀態為 C3 那一類之 Residual Variance 下降幅度最少，這個結果說明連音程度越嚴重則受前音節影響愈大。

後音節影響因素之觀察：同樣的，我們也將所有音節依照其與後一音節間的連音狀態的種類分為 3 類，並分別比較其 Residual Variance $V1$ 與 Residual Variance $V2$ ，目的是為了觀察此 3 類音節的前音節影響因素整體的差異性，以下為實驗結果與分析：

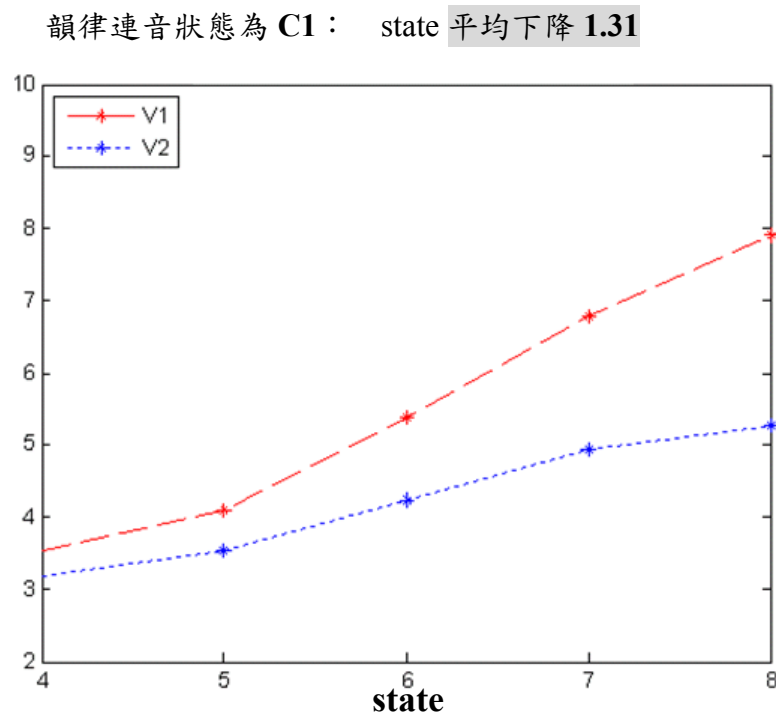


圖 4.6:連音狀態為 C1 時 Residual Variance V1 與 V2 後 5 states 之比較圖

韻律連音狀態為 **C2**：state 平均下降 **0.99**

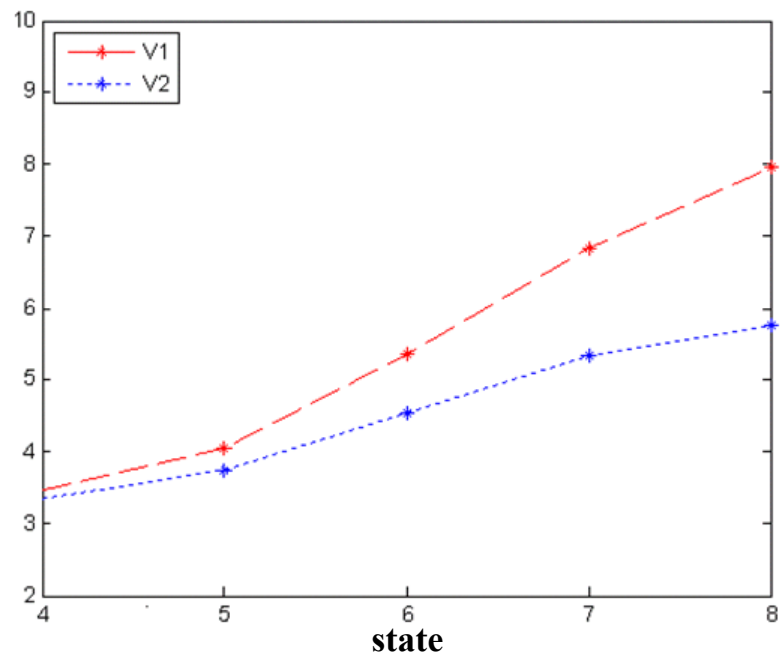


圖 4.7：連音狀態為 C2 時 Residual Variance V1 與 V2 後 5 states 之比較圖

韻律連音狀態為 **C3**：state 平均下降 **0.68**

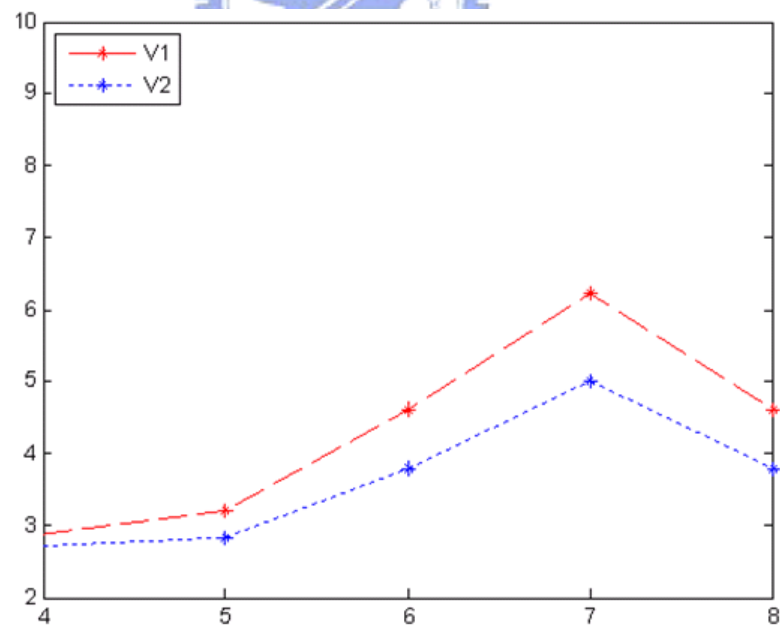


圖 4.8：連音狀態為 C3 時 Residual Variance V1 與 V2 後 5 states 之比較圖

●結果分析：從上圖我們發現後音節影響因素在連音狀態為 C1 那一類之 Residual Variance 下降幅度最多，而在連音狀態為 C3 那一類之 Residual Variance 下降幅度最少，這個結果說明連音程度越嚴重則受後音節影響愈大。

由 4.1 節中我們可以發現韻律連音狀態之定義下在模型多考慮音節影響因素後能夠使 Residual Variance 下降，如此之外不同的連音程度中也能表現出不同的結果，而這也是我們希望在模型收斂後也能夠表現的結果，以下 4.2 節中我們便分析模型收斂後之影響因素。



4.2 模型收斂後影響因素之分析

1. 目標函數概似值：

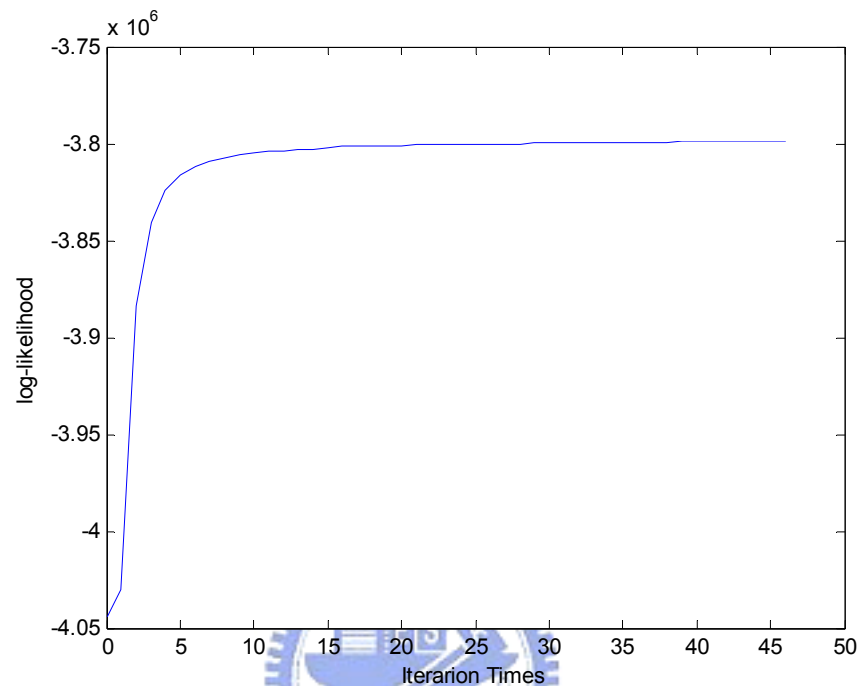


圖 4.9：log-likelihood 值 Q 之更新曲線

- 結果分析：從上圖我們發現 log-likelihood 值 Q 的確隨著重複更新遞增，這代表著模型朝著合理的趨勢在收斂。

2. Residual Variance 之比較：

我們觀察初始以及模型收斂後的 Residual Variance 曲線，如下圖 4.10。

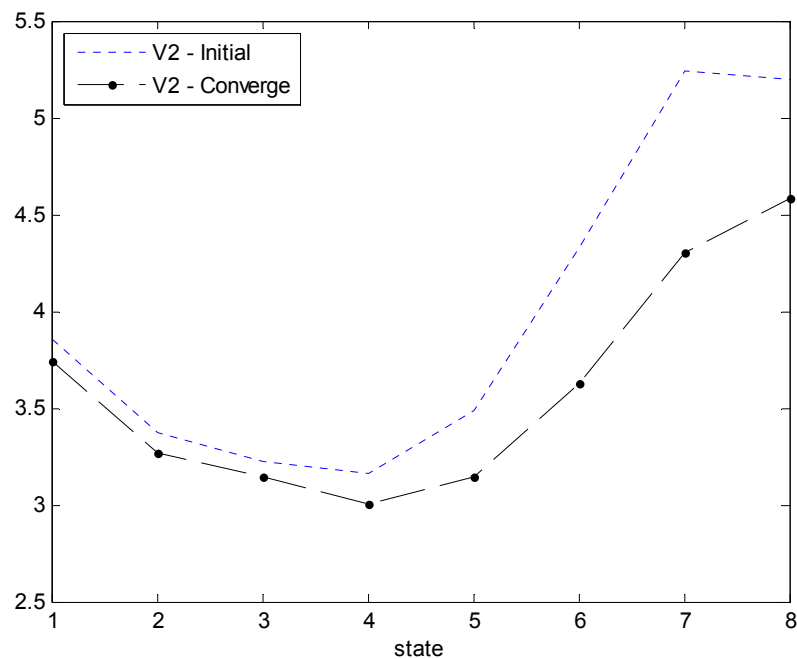


圖 4.10： Residual Variance V2 之比較圖

- 結果分析：從圖 4.10 我們可看出模型收斂後的 Residual Variance 曲線合理的比初始 Residual Variance 曲線下降，這也表示合理的收斂結果，如此之外音節韻母部份下降程度比聲母大，這也能看出在連續語音中音節受後一音節之影響程度較大。

以下 4.2.1 小節中我們利用相同方法來觀察頻譜連音狀態在不同連音程度下的表現。

4.2.1 不同連音狀態種類影響之程度分析：

頻譜連音狀態為 C1：state 平均下降 1.04

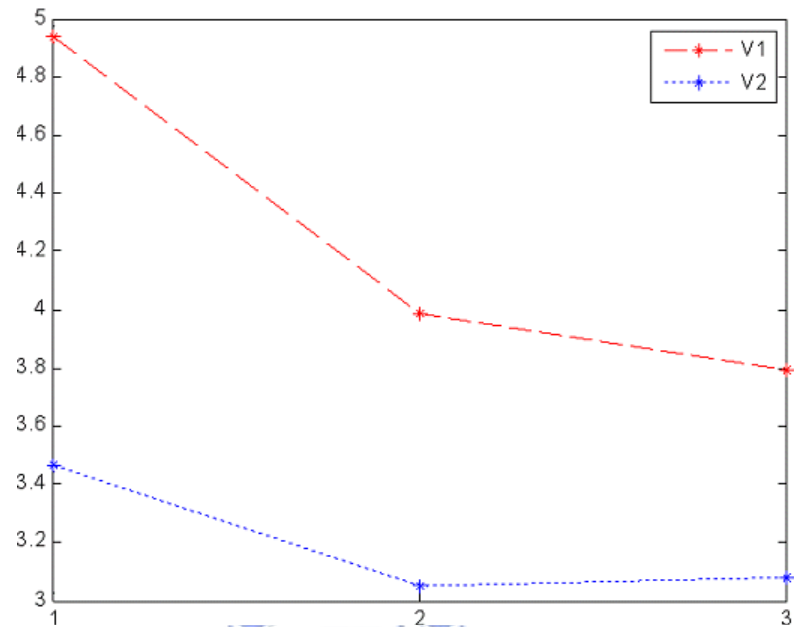


圖 4.11：連音狀態為 C1 時 Residual Variance V1 與 V2 前 3 states 之比較圖

頻譜連音狀態為 **C2**： state 平均下降 **0.55**

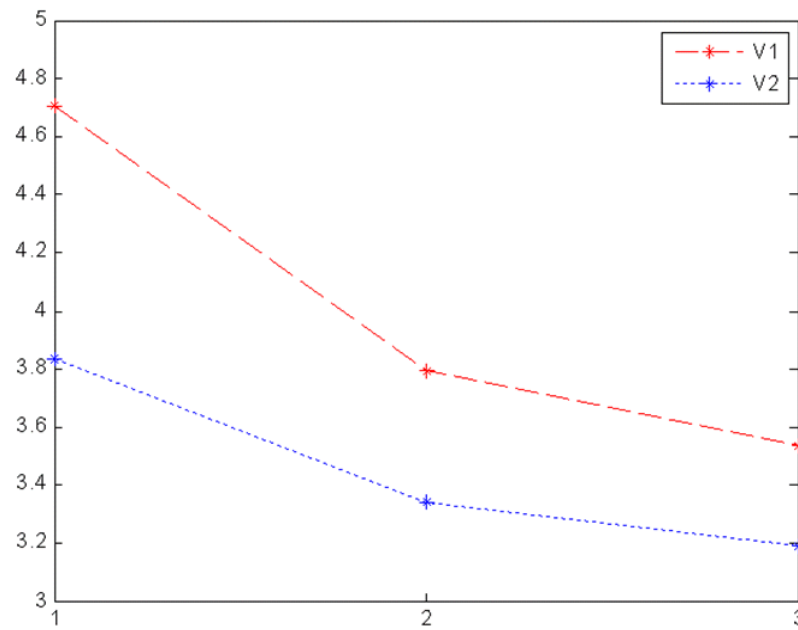


圖 4.12:連音狀態為 *C2* 時 Residual Variance V1 與 V2 前 3 states 之比較圖

頻譜連音狀態為 **C3**： state 平均下降 **0.23**

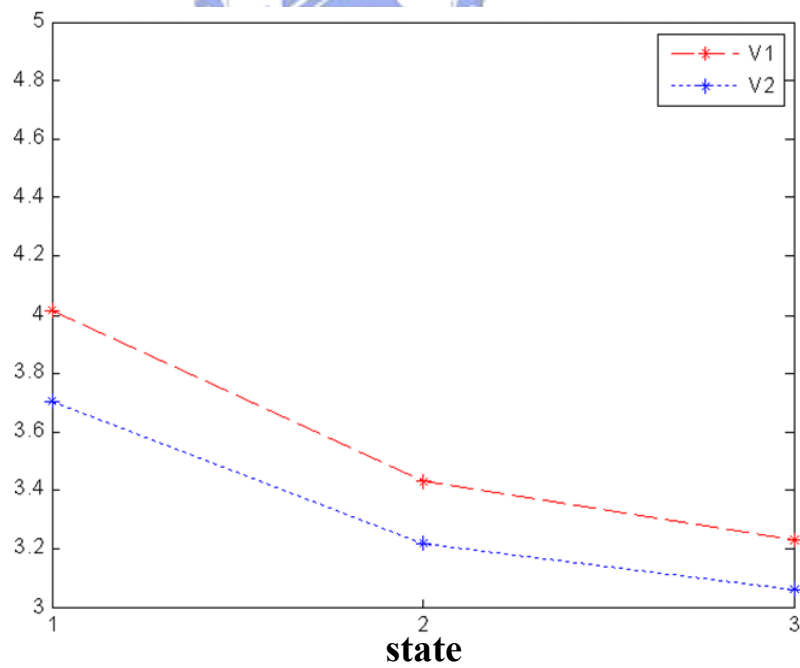


圖 4.13:連音狀態為 *C3* 時 Residual Variance V1 與 V2 前 3 states 之比較圖

- 結果分析：從上圖我們發現前音節影響因素在連音狀態為 C1 那一類之 Residual Variance 下降幅度最多，而在連音狀態為 C3 那一類之 Residual Variance 下降幅度最少，這個與我們所希望模型訓練出來的結果相同。

後音節影響因素之觀察：同樣的做法我們觀察音節之韻母部份，以下為實驗結果與分析：

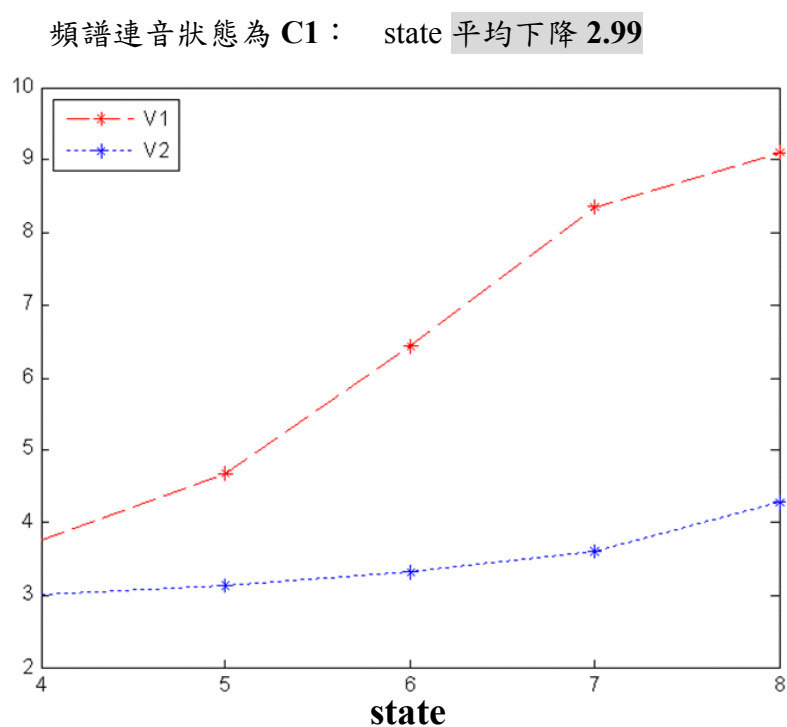


圖 4.14：連音狀態為 C1 時 Residual Variance V1 與 V2 後 5 states 之比較圖

頻譜連音狀態為 **C2**：state 平均下降 **1.82**

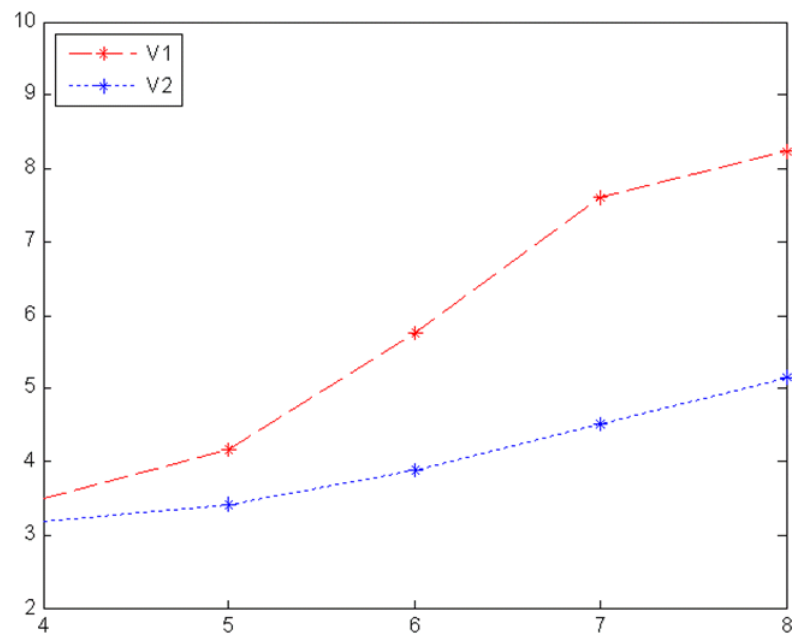


圖 4.15：連音狀態為 C2 時 Residual Variance V1 與 V2 後 5 states 之比較圖

頻譜連音狀態為 **C3**：state 平均下降 **0.64**

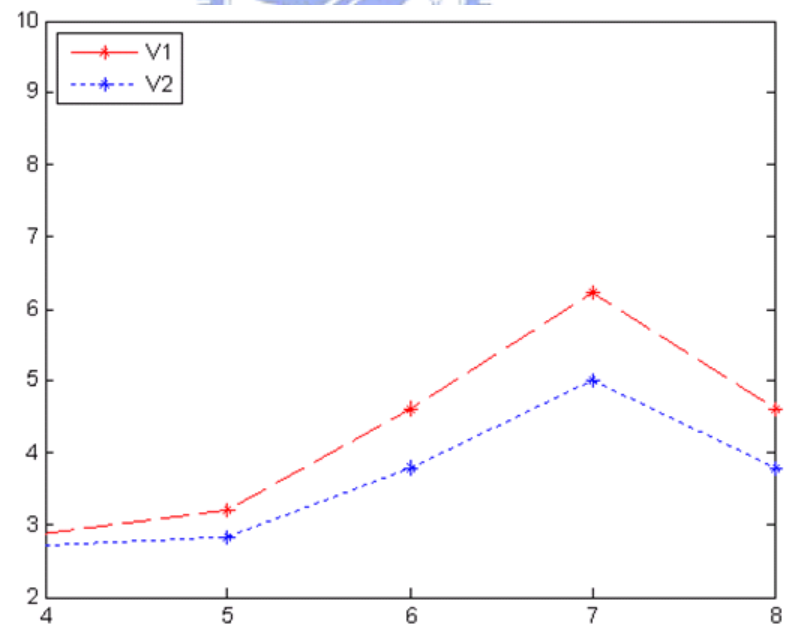


圖 4.16：連音狀態為 C3 時 Residual Variance V1 與 V2 後 5 states 之比較圖

●結果分析：從上圖我們發現後音節影響因素在連音狀態為 C1 那一類之 Residual Variance 下降幅度最多，而在連音狀態為 C3 那一類之 Residual Variance 下降幅度最少。

經過觀察發現模型收斂後影響因素之表現也能合乎我們所預期的結果，表示模型之影響因素能夠有不錯的學習效果。以下 4.3 節中將比較韻律、頻譜連音狀態與韻律參數、語言參數之間的關聯性。

4.3 連音狀態之分析

本節中我們首先透過韻律參數來比較韻律、頻譜連音狀態之間的差異，接著我們觀察頻譜連音狀態和 prosody break type 以及語言參數之間的關聯性。

4.3.1 韻律、頻譜連音狀態之比較：

1. Energy-dip 分析

在更新連音狀態時我們有加上 Pause Acoustic Model 來輔助，所以在這裡我們分別觀察韻律以及頻譜之 3 種連音狀態序列的 Energy-dip、Pause duration 之 discrete probability 分佈。



(1)韻律連音狀態序列：

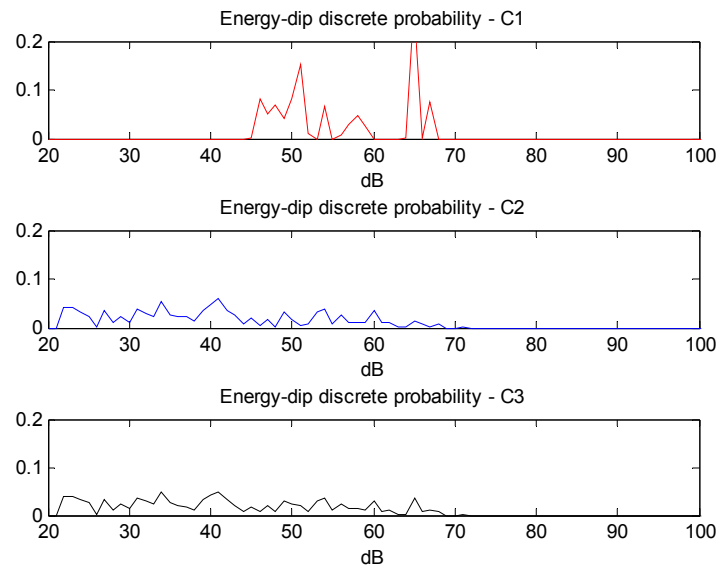


圖 4.17：韻律連音狀態之 Energy-dip discrete probability 分佈圖

(2)頻譜連音狀態序列：

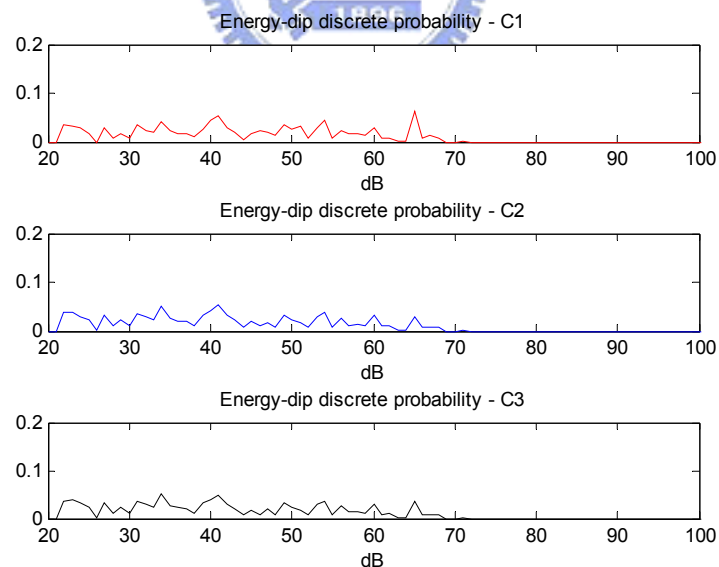


圖 4.18：頻譜連音狀態之 Energy-dip discrete probability 分佈圖

- 結果分析：圖 4.17 中 C1 的 Energy-dip 分佈較集中在 45~70dB 相較於 C2、C3 有明顯差異；但圖 4.18 中 3 種連音狀態差異性很小。這結果顯示頻譜參數的影響相較於 Pause Acoustic Model 之影響大很多，且結果卻不如我們預期的有差別性。

2. Pause duration 分析

(1) 韻律連音狀態序列：

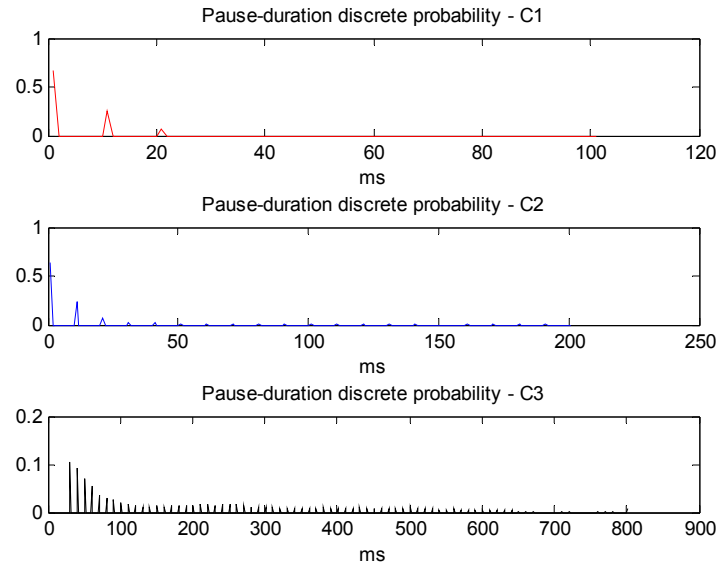


圖 4.19：韻律連音狀態之 Pause-duration discrete probability 分佈圖

(2) 頻譜連音狀態序列：

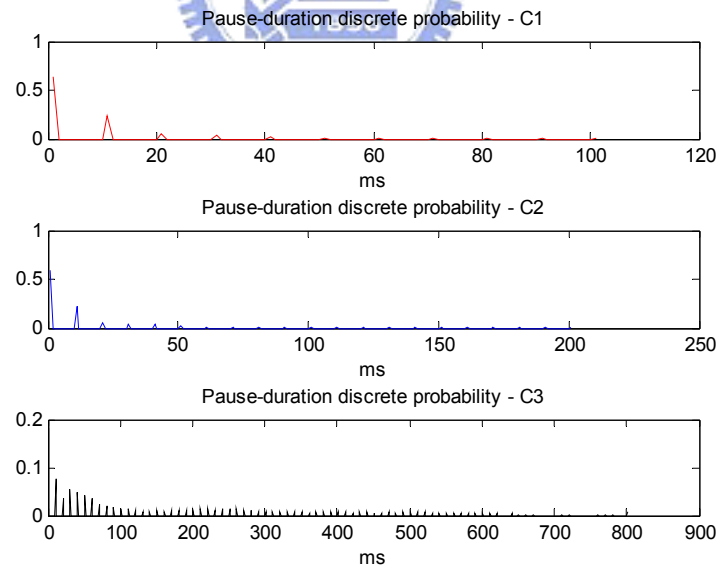


圖 4.20：頻譜連音狀態之 Pause-duration discrete probability 分佈圖

●結果分析：圖 4.19 中 C1、C2 之 Pause-duration 較集中在 0~20ms，C3 與 C1、C2 相較起來分佈很廣。而圖 4.20 之 C1、C2、C3 分佈與圖 4.19 變異不大，C3 保留其與 C1、C2 間的差異性。

4.3.2 Break type 與頻譜連音狀態之分析

我們觀察 break type 與頻譜連音狀態之間個數上的關係，以下為 Break type 與連音狀態對應圖。在第 3 章我們定義 B0 為 C1；B1、B2-1 合併為 C2；B2-2、B3、B4 則合併為 C3，下表中以**粗體線框**起。而符號“〈〉”表示數量；括號“()”表示固定一種 break type 對應於 3 種連音狀態之數量百分比。

表 4.1：BreakType 與頻譜連音狀態對應表

連音狀態 Break Type	C1	C2	C3
B0 〈4039〉	〈1894〉 (46.89%)	〈1657〉 (41.03%)	〈488〉 (12.08%)
B1 〈32623〉	〈7425〉 (22.76%)	〈20463〉 (62.73%)	〈4735〉 (14.51%)
B2-1 〈4641〉	〈724〉 (15.60%)	〈3205〉 (69.06%)	〈712〉 (15.34%)
B2-2 〈8707〉	〈520〉 (5.97%)	〈2182〉 (25.06%)	〈6005〉 (68.97%)
B3 〈3432〉	〈14〉 (0.41%)	〈93〉 (2.71%)	〈3325〉 (96.88%)
B4 〈2782〉	〈1〉 (0.04%)	〈46〉 (1.65%)	〈2735〉 (98.31%)

- 結果分析：從上圖發現每一種 break type 之數量百分比最大值(醒目提示記號處)所對應的連音狀態和我們在第 3 章所之定義 break type 合併方式相同(最大值都出現在粗體線框內)，這結果說明利用韻律參數標記之 break type 和使用頻譜參數標記之連音狀態之間的重複性偏高，這樣的結果也顯示韻律參數和頻譜參數之間有著某種程度的關聯性。此外，我們還可以看出粗體線框內中以 B3、B4 類別之數量百分比數值最大而 B4 類別之數量百分比數值最小，這或許和 break type 分類時的方式有關。

4.3.3 音節長度與頻譜連音狀態之分析

表 4.2 中內容為為正規化後之平均音長(ms)，音節考慮的影響因素為音節基本類型、音調，數學式如下：

$$sd_n = sd_n^r + \mu_{S_n} + \mu_{t_n} \quad (4.6)$$

where S_n : basic syllable type

t_n : syllable tone

表 4.2：正規化音長與頻譜連音狀態對應表

$C_{n-1} \backslash C_n$	C1	C2	C3
C1	-26.4	-12.4	22.3
C2	-18.2	-7.2	24.8
C3	-18.7	-6.1	19.6

●結果分析：表 4.2 中可以發現相對來說，音節左右連音狀態皆為 C1 的集合音長最短。此外橫列與直行的比較來看，橫列中音長隨著連音狀態不同音長之差異較直行中來的大，這表示音節受到與後一音節間之連音狀態影響較大。

4.3.4 頻譜連音狀態與決策樹分析

為了觀察我們頻譜連音狀態和語言參數之間的關係，在這裡我們透過決策樹的方法並以語言參數的資訊來當做問題集內容。

4.3.4.1 問題集

問題集中共有 18 個問題，其中 Q1～Q11 是對每一個 juncture 前音節的韻母類型以及 juncture 後音節的聲母類型來作為決策樹的問題；Q12、Q13 是對於

juncture 的在詞句中的位置來作為問題；Q14～Q18 是對於 juncture 之 break type 種類作為問題。

Q1：聲母是否為【空(Null)】？

Q2：聲母是否為【ㄅ(b)】、【ㄆ(p)】、【ㄇ(m)】？

Q3：聲母是否為【ㄈ(f)】、【ㄏ(h)】、【ㄒ(x)】、【ㄕ(k)】、【ㄓ(zh)】？

Q4：聲母是否為【ㄇ(m)】、【ㄋ(n)】、【ㄌ(l)】、【ㄋ(r)】？

Q5：聲母是否為【ㄑ(q)】、【ㄔ(ch)】、【ㄕ(c)】？

Q6：聲母是否為【ㄆ(p)】、【ㄊ(t)】、【ㄋ(k)】？

Q7：聲母是否為【ㄐ(j)】、【ㄓ(zh)】、【ㄓ(z)】？

Q8：韻母是否為單韻母(Single Vowel)：

【一(yi)】、【ㄨ(wu)】、【ㄩ(yu)】、【ㄚ(a)】、【ㄛ(o)】、【ㄜ(e)】、【ㄝ(eh)】、
【ㄦ(er)】？

Q9：韻母是否為複韻母(Compound Vowel)：

【ㄞ(ai)】、【ㄟ(ei)】、【ㄠ(ao)】、【ㄡ(ou)】、【ㄣ(ya)】、【ㄤ(ye)】、
【ㄢ(yai)】、【ㄣ(yao)】、【ㄤ(you)】、【ㄤ(yo)】、【ㄨㄚ(wa)】、
【ㄨㄛ(wo)】、【ㄨㄞ(wai)】、【ㄨㄟ(wei)】、【ㄩㄝ(yue)】？

Q10：韻母是否為鼻音結尾韻母(Nasal Ending Vowel)：

【ㄢ(an)】、【ㄣ(en)】、【ㄤ(ang)】、【ㄥ(eng)】、【ㄣ(yan)】、【ㄣ(yin)】、
【ㄤ(yang)】、【ㄥ(ying)】、【ㄨㄢ(wan)】、【ㄨㄣ(wen)】、【ㄨㄤ(wang)】、
【ㄨㄥ(weng)】、【ㄩㄢ(yuan)】、【ㄩㄣ(yun)】、【ㄩㄥ(yung)】？

Q11：韻母是否為含有介音(Medial)：

【ㄣ(ya)】、【ㄤ(ye)】、【ㄢ(yai)】、【ㄣ(yao)】、【ㄤ(you)】、
【ㄣ(yan)】、【ㄣ(yin)】、【ㄤ(yang)】、【ㄥ(ying)】、【ㄤ(yo)】、
【ㄨㄚ(wa)】、【ㄨㄛ(wo)】、【ㄨㄞ(wai)】、【ㄨㄟ(wei)】、【ㄨㄢ(wan)】、
【ㄨㄣ(wen)】、【ㄨㄤ(wang)】、【ㄨㄥ(weng)】、【ㄩㄝ(yue)】、【ㄩㄢ(yuan)】、

【ㄣㄣ(yun)】、【ㄣㄣ(yung)】？

Q12： juncture 位置是否在 inter-word 之間？

Q13： juncture 是否在標點符號位置 ？

Q14： juncture之Break type標記為 B0？

Q15： juncture之Break type標記為 B0, B1？

Q16： juncture之Break type標記為 B0, B1, B2-1？

Q17： juncture之Break type標記為 B0, B1, B2-1, B2-2？

Q18： juncture之Break type標記為 B3, B4？

4.3.4.2 決策樹分裂條件

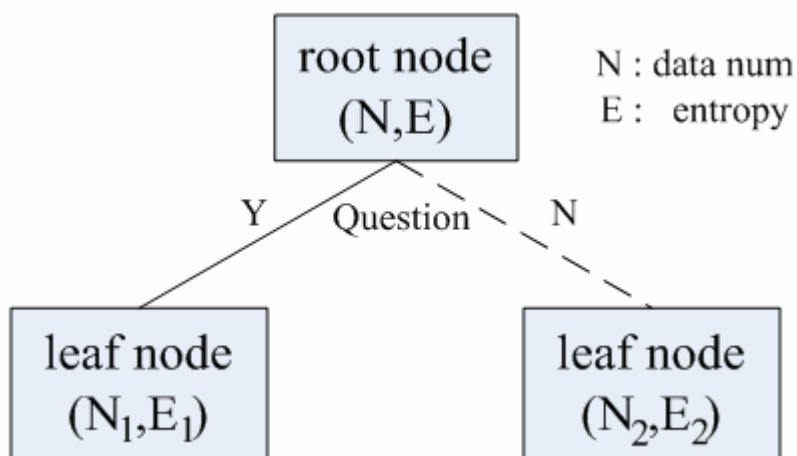


圖 4.21：決策樹分裂示意圖

上圖為決策樹之示意圖，root node 在被問過一個問題後裡面的資料會分成兩群資料，root node 必須問過所有問題集(question set)的問題後，以 information gain 最大的那一個問題來分裂，而該問題會從問題集移除。而決策樹在 2 個條件下會停止生長：

- 1.當 node 中的數量 $N < X$ 時停止分裂。
- 2.當 information gain $< Y$ 時停止分裂。

3.問題集沒有剩餘問題。

information gain 算式如下：

當決策樹在生長時我們可以依照每個 node 中的不同連音狀態 C_i 之個數計算出其 **entropy**：

$$E = -1 \times \sum_{i=1}^3 \frac{n_i}{N} \times \log\left(\frac{n_i}{N}\right), \text{ where } n_i \text{ is numbers of } C_i \text{ 。} \quad (4.7)$$
$$\text{information gain} = E - \left(\frac{N_1}{N} \times E_1 + \frac{N_2}{N} \times E_2\right) \text{ 。} \quad (4.8)$$

在這裡我們設定 $X=3000$ 、 $Y=0.007$ ，以下為決策樹分裂圖：



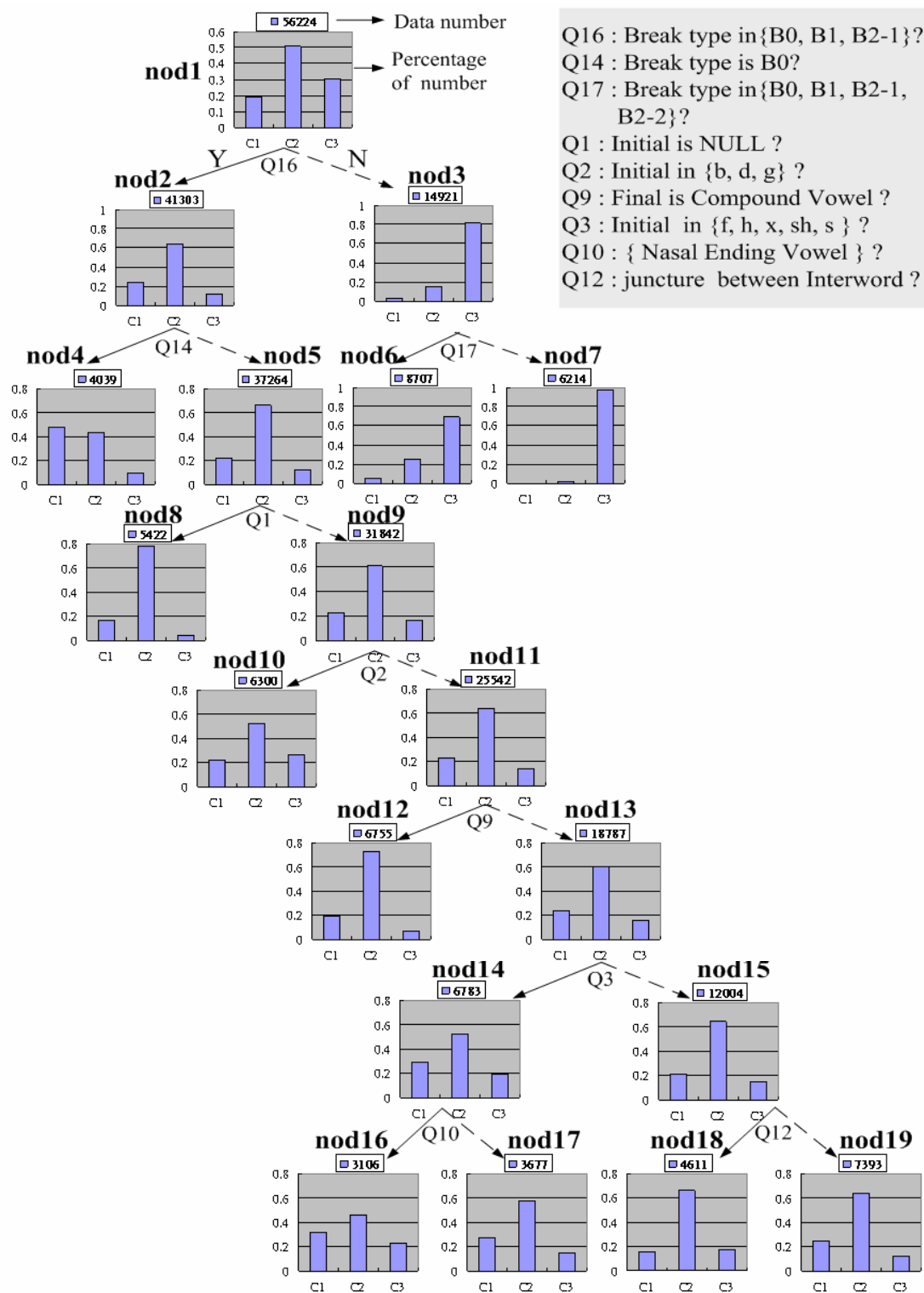


圖 4.22：連音狀態決策樹分裂圖

●**結果分析：**

以下我們分析上圖節點(nod)中連音狀態之數量百分比和語言參數之間的關聯性。

分析 1 (Q16)：

是→nod 2：Break type 標記為 B0、B1、B2-1 之 juncture，其連音狀態 C3 數量最少。

否→nod 3：Break type 標記為 B2-2、B3、B4 之 juncture，其連音狀態 C3 數量最多。

●**分析 1 小結：**我們可以發現以 nod 3 為根節點 (root nod)之子支樹 (subtree) 中的節點以 **C3** 個數最多；以 nod 2 為根節點之子支樹中的節點以 C3 個數最少 (除了 nod 10、nod 18，原因後述)。

分析 2 (Q14)：

是→nod 4：Break type 標記為 B0 之 juncture，其連音狀態 C1 數量最多。

否→nod 5：Break type 標記為 B1、B2-1 之 juncture，其連音狀態 C2 數量最多。

分析 3 (Q17)：

是→nod 6：Break type 標記為 B2-2 之 juncture，其連音狀態 C3 數量最多。

否→nod 7：Break type 標記為 B3、B4 之 juncture，其連音狀態 C3 數量最多。

●**分析 2, 3 小結：**之結果可參照表 4.1，是為合理的結果。

分析 4 (Q1)：

nod 8 是 nod 5 中聲母為空聲母之類別，而該類別聲母部份屬於有聲 (voiced)聲母，所以比起無聲 (unvoiced)聲母來說較容易對前一個音節造成影響，故 nod 8 中 C3 之數量百分比會降低。

分析 5 (Q2) :

nod 10 是 nod 9 中聲母為ㄅ、ㄆ、ㄇ為之類別，而該類別聲母發音特性為短暫閉氣之爆破音，所以相對其他類型聲母來說對前音節影響較小，故 nod 10 中 C3 之數量百分比會提高，而這也是分析 1 中 C3 個數多於 C1 的原因。

分析 6 (Q9) :

nod 12 是 nod 11 中韻母為複韻母之類別，而該類韻母發音時容易發生尾音拉長的情況而影響後一個音節，故 nod 12 中 C3 之數量百分比會降低。

分析 7 (Q3) :

nod 14 是 nod 13 中，聲母為ㄌ、ㄍ、ㄎ、ㄆ、ㄇ之摩擦音，所以相對其他類型韻母來說對後音節影響較大，故 nod 14 中 C1 之數量百分比會提高。

分析 8 (Q10) :

nod 16 是 nod 14 中韻母為鼻音結尾類別，而該類韻母發音時容易發生尾音拉長的情況而影響後一個音節，故 nod 16 中 C1 之數量百分比會提高；同理，nod 17 中 C3 之數量百分比會降低。

分析 9 (Q12) :

nod 18 是 nod 15 中 juncture 位置屬於 interword，相對於位置在 intraword 之 juncture 此類別前後音節的影響較小，故 nod 18 中 C3 之數量百分比會提高，而這也是分析 1 中 C3 個數多於 C1 的原因；同理，nod 19 中 C3 之數量百分比會降低。

第五章：結論與未來展望

由實驗結果分析可看出模型中個影響因素能學習到不錯的成果，其中前、後音節影響因素在不同連音狀態時也能在頻譜參數上看出不同的表現。如此之外，頻譜連音狀態在部份韻律參數以及語言參數的分析上都能有合理的結果。

然而因為語料庫有些音節種類之數量太少的關係，所以我們才假設一個音節的前、後音節影響因素是不互相影響並且降低聲母、韻母之類別數量，未來如果能夠將語料庫內容再擴充，那麼可以將前、後音節影響因素提升到 syllable-pair 的程度亦或更多類別之連音狀態，如此一來建構出來的音節模型應該會更合乎真實的情況。應用方面，未來如果能夠在大型語料庫合成系統中原本之單元選取方法再加上音節頻譜模型提供的頻譜方面資訊，便能夠選出更合適的合成單元以提升語音合成效果。



參考文獻

- [1] Chen-Yu Chiang, Hsiu-Min Yu, Yih-Ru Wang, and Sin-Horng Chen, “An Automatic Prosody Labeling Method for Mandarin Speech”, Interspeech, Antwerp, Belgium, August, 2007。
- [2] The HTK Book (for HTK version 3.4)
- [3] WaveSurfer Homepage : <http://speech.kth.se/wavesurfer/>
- [4] 王小川,「語音訊號處理」,全華科技圖書,民國九十三年三月。
- [5] 林銘彥,「以音節訊號特性挑選中文文句翻語音系統合成單元之研究」,國立交通大學碩士論文,民國九十六年九月。
- [6] 簡家勇,「漢語語音長度模型之研究」,國立交通大學碩士論文,民國九十六年九月。
- [7] 陳啟風,「中文單詞之韻律模式研究」,國立交通大學碩士論文,民國九十六年七月。
- [8] M.Chu, Hu Peng, “An Objective Measure for Assessment of the Concatenative TTS Segment”, Microsoft Research China , 2001。
- [9] 洪國興,「以語料庫維基礎之中文文句翻語音系統實現」,國立交通大學碩士論文,民國九十五年八月。
- [10] Jun Xu, LianHong Cai, “Spectral Continuity Measures at Mandarin Syllable Boundaries ”, Key Laboratory of Pervasive Computing (TsingHua University), P.R.China。
- [11] Xuedong Huang, Alex Acero, Hsiao-Wuen Hou, “Spoken language Processing a guide to theory, algorithm, and system development”.

附錄 音節類型與注音對照表

音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音
1	ㄘ	31	ㄘㄛ	61	ㄣㄛ	91	ㄣㄣ	121	ㄣㄣ	151	ㄣㄣ	181	ㄣㄣ	241	ㄣㄣ
2	ㄣ	32	ㄣㄣ	62	ㄣㄣ	92	ㄣㄣ	122	ㄣㄣ	152	ㄣㄣ	182	ㄣㄣ	242	ㄣㄣ
3	ㄣ	33	ㄣㄣ	63	ㄣㄣ	93	ㄣㄣ	123	ㄣㄣ	153	ㄣㄣ	183	ㄣㄣ	243	ㄣㄣ
4	ㄣ	34	ㄣㄣ	64	ㄣㄣ	94	ㄣㄣ	124	ㄣㄣ	154	ㄣㄣ	184	ㄣㄣ	244	ㄣㄣ
5	ㄣ	35	ㄣㄣ	65	ㄣㄣ	95	ㄣㄣ	125	ㄣㄣ	155	ㄣㄣ	185	ㄣㄣ	245	ㄣㄣ
6	ㄣ	36	ㄣㄣ	66	ㄣㄣ	96	ㄣㄣ	126	ㄣㄣ	156	ㄣㄣ	186	ㄣㄣ	246	ㄣㄣ
7	ㄣ	37	ㄣㄣ	67	ㄣㄣ	97	ㄣㄣ	127	ㄣㄣ	157	ㄣㄣ	187	ㄣㄣ	247	ㄣㄣ
8	ㄣ	38	ㄣㄣ	68	ㄣㄣ	98	ㄣㄣ	128	ㄣㄣ	158	ㄣㄣ	188	ㄣㄣ	248	ㄣㄣ
9	ㄣ	39	ㄣㄣ	69	ㄣㄣ	99	ㄣㄣ	129	ㄣㄣ	159	ㄣㄣ	189	ㄣㄣ	249	ㄣㄣ
10	ㄣ	40	ㄣㄣ	70	ㄣㄣ	100	ㄣㄣ	130	ㄣㄣ	160	ㄣㄣ	190	ㄣㄣ	250	ㄣㄣ
11	ㄣ	41	ㄣㄣ	71	ㄣㄣ	101	ㄣㄣ	131	ㄣㄣ	161	ㄣㄣ	191	ㄣㄣ	251	ㄣㄣ
12	ㄣ	42	ㄣㄣ	72	ㄣㄣ	102	ㄣㄣ	132	ㄣㄣ	162	ㄣㄣ	192	ㄣㄣ	252	ㄣㄣ
13	ㄣ	43	ㄣㄣ	73	ㄣㄣ	103	ㄣㄣ	133	ㄣㄣ	163	ㄣㄣ	193	ㄣㄣ	253	ㄣㄣ
14	ㄣ	44	ㄣㄣ	74	ㄣㄣ	104	ㄣㄣ	134	ㄣㄣ	164	ㄣㄣ	194	ㄣㄣ	254	ㄣㄣ
15	ㄣ	45	ㄣㄣ	75	ㄣㄣ	105	ㄣㄣ	135	ㄣㄣ	165	ㄣㄣ	195	ㄣㄣ	255	ㄣㄣ
16	ㄣ	46	ㄣㄣ	76	ㄣㄣ	106	ㄣㄣ	136	ㄣㄣ	166	ㄣㄣ	196	ㄣㄣ	256	ㄣㄣ
17	ㄣ	47	ㄣㄣ	77	ㄣㄣ	107	ㄣㄣ	137	ㄣㄣ	167	ㄣㄣ	197	ㄣㄣ	257	ㄣㄣ
18	ㄣ	48	ㄣㄣ	78	ㄣㄣ	108	ㄣㄣ	138	ㄣㄣ	168	ㄣㄣ	198	ㄣㄣ	258	ㄣㄣ
19	ㄣ	49	ㄣㄣ	79	ㄣㄣ	109	ㄣㄣ	139	ㄣㄣ	169	ㄣㄣ	199	ㄣㄣ	259	ㄣㄣ
20	ㄣ	50	ㄣㄣ	80	ㄣㄣ	110	ㄣㄣ	140	ㄣㄣ	170	ㄣㄣ	200	ㄣㄣ	260	ㄣㄣ
21	ㄣ	51	ㄣㄣ	81	ㄣㄣ	111	ㄣㄣ	141	ㄣㄣ	171	ㄣㄣ	201	ㄣㄣ	261	ㄣㄣ
22	ㄣ	52	ㄣㄣ	82	ㄣㄣ	112	ㄣㄣ	142	ㄣㄣ	172	ㄣㄣ	202	ㄣㄣ	262	ㄣㄣ
23	ㄣ	53	ㄣㄣ	83	ㄣㄣ	113	ㄣㄣ	143	ㄣㄣ	173	ㄣㄣ	203	ㄣㄣ	263	ㄣㄣ
24	ㄣ	54	ㄣㄣ	84	ㄣㄣ	114	ㄣㄣ	144	ㄣㄣ	174	ㄣㄣ	204	ㄣㄣ	264	ㄣㄣ
25	ㄣ	55	ㄣㄣ	85	ㄣㄣ	115	ㄣㄣ	145	ㄣㄣ	175	ㄣㄣ	205	ㄣㄣ	265	ㄣㄣ
26	ㄣ	56	ㄣㄣ	86	ㄣㄣ	116	ㄣㄣ	146	ㄣㄣ	176	ㄣㄣ	206	ㄣㄣ	266	ㄣㄣ
27	ㄣ	57	ㄣㄣ	87	ㄣㄣ	117	ㄣㄣ	147	ㄣㄣ	177	ㄣㄣ	207	ㄣㄣ	267	ㄣㄣ
28	ㄣ	58	ㄣㄣ	88	ㄣㄣ	118	ㄣㄣ	148	ㄣㄣ	178	ㄣㄣ	208	ㄣㄣ	268	ㄣㄣ
29	ㄣ	59	ㄣㄣ	89	ㄣㄣ	119	ㄣㄣ	149	ㄣㄣ	179	ㄣㄣ	209	ㄣㄣ	269	ㄣㄣ
30	ㄣ	60	ㄣㄣ	90	ㄣㄣ	120	ㄣㄣ	150	ㄣㄣ	180	ㄣㄣ	210	ㄣㄣ	270	ㄣㄣ

音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音	音節 類型	注 音
271	くーㄣ	301	ㄈㄨㄚˊ	331	ㄌㄨㄨㄟ	361	ㄎㄨㄟ	391	ㄋㄣㄟ
272	ㄒーㄣ	302	ㄨㄛ	332	ㄍㄨㄟ	362	ㄈㄨㄟ	392	ㄋㄣㄟ
273	ㄋーㄣ	303	ㄷㄨㄛ	333	ㄎㄨㄟ	363	ㄋㄨㄟ	393	ㄣㄟ
274	ㄋㄨㄟ	304	ㄴㄨㄛ	334	ㄈㄨㄟ	364	ㄷㄨㄟ	394	ㄣㄣㄟ
275	ㄎㄨㄟ	305	ㄹㄨㄛ	335	ㄋㄨㄟ	365	ㄋㄨㄟ	395	ㄣㄣㄟ
276	ㄷㄨㄟ	306	ㄱㄨㄛ	336	ㄷㄨㄟ	366	ㄨㄟ	396	ㄒㄣㄟ
277	ㄱㄨㄟ	307	ㄲㄨㄛ	337	ㄨㄟ	367	ㄷㄨㄟ	397	ㄋㄣㄟ
278	ㄴㄨㄟ	308	ㄴㄨㄛ	338	ㄷㄨㄟ	368	ㄴㄨㄟ	398	ㄣㄟ
279	ㄣㄨㄟ	309	ㄌㄨㄛ	339	ㄴㄨㄟ	369	ㄹㄨㄟ	399	ㄣㄣㄟ
280	ㄣㄨㄟ	310	ㄍㄨㄛ	340	ㄹㄨㄟ	370	ㄍㄨㄟ	400	ㄣㄣㄟ
281	ㄒㄨㄟ	311	ㄎㄨㄛ	341	ㄱㄨㄟ	371	ㄎㄨㄟ	401	ㄒㄣㄟ
282	ㄋㄨㄟ	312	ㄈㄨㄛ	342	ㄲㄨㄟ	372	ㄈㄨㄟ	402	ㄋㄣㄟ
283	ㄋㄨㄟ	313	ㄋㄨㄛ	343	ㄴㄨㄟ	373	ㄨㄟ	403	ㄣㄟ
284	ㄴㄨㄟ	314	ㄷㄨㄛ	344	ㄌㄨㄟ	374	ㄷㄨㄟ	404	ㄣㄣㄟ
285	ㄣㄨㄟ	315	ㄋㄨㄛ	345	ㄍㄨㄟ	375	ㄴㄨㄟ	405	ㄣㄣㄟ
286	ㄣㄨㄟ	316	ㄋㄨㄛ	346	ㄎㄨㄟ	376	ㄱㄨㄟ	406	ㄒㄣㄟ
287	ㄒㄨㄟ	317	ㄨㄟ	347	ㄈㄨㄟ	377	ㄲㄨㄟ	407	ㄴㄨㄟ
288	ㄋㄨㄟ	318	ㄷㄨㄟ	348	ㄋㄨㄟ	378	ㄴㄨㄟ	408	ㄴㄨㄟ
289	ㄷㄨㄟ	319	ㄴㄨㄟ	349	ㄷㄨㄟ	379	ㄌㄨㄟ	409	ㄌㄨㄟ
290	ㄋㄨㄟ	320	ㄹㄨㄟ	350	ㄋㄨㄟ	380	ㄍㄨㄟ	410	ㄨㄟ
291	ㄋㄨㄟ	321	ㄍㄨㄟ	351	ㄋㄨㄟ	381	ㄎㄨㄟ	411	ㄱㄨㄟ
292	ㄎㄨㄟ	322	ㄎㄨㄟ	352	ㄨㄟ	382	ㄈㄨㄟ		
293	ㄷㄨㄟ	323	ㄈㄨㄟ	353	ㄷㄨㄟ	383	ㄋㄨㄟ		
294	ㄱㄨㄟ	324	ㄨㄟ	354	ㄴㄨㄟ	384	ㄷㄨㄟ		
295	ㄨㄟ	325	ㄷㄨㄟ	355	ㄹㄨㄟ	385	ㄋㄨㄟ		
296	ㄷㄨㄟ	326	ㄴㄨㄟ	356	ㄱㄨㄟ	386	ㄋㄨㄟ		
297	ㄴㄨㄟ	327	ㄹㄨㄟ	357	ㄲㄨㄟ	387	ㄣㄟ		
298	ㄹㄨㄟ	328	ㄱㄨㄟ	358	ㄴㄨㄟ	388	ㄣㄣㄟ		
299	ㄍㄨㄟ	329	ㄲㄨㄟ	359	ㄌㄨㄟ	389	ㄣㄣㄟ		
300	ㄎㄨㄟ	330	ㄴㄨㄟ	360	ㄍㄨㄟ	390	ㄒㄣㄟ		