國立交通大學

電信工程學系

碩士論文

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

A Study on Unit Selection of Using Signal Characteristic for Corpus-based Mandarin TTS System

研 究 生：林銘彥

指導教授：陳信宏　教授

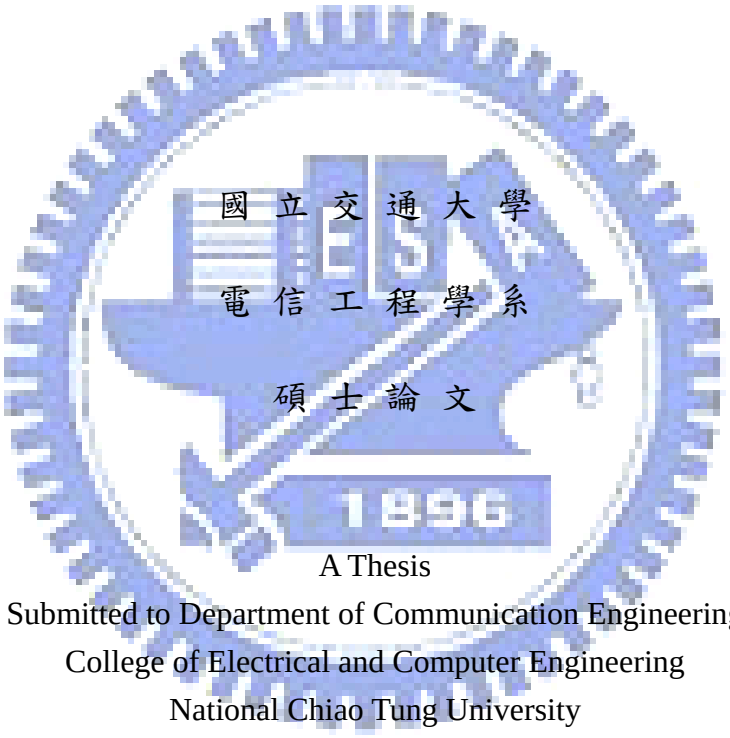
中 華 民 國 九十六 年 九 月

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

A Study on Unit Selection of Using Signal Characteristic
for Corpus-based Mandarin TTS System

研 究 生：林銘彥　　Student：Ming-Yan Lin

指導教授：陳信宏　　Advisor：Dr. Sin-Horing Chen



國 立 交 通 大 學

電 信 工 程 學 系

碩 士 論 文

A Thesis
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
in Electrical Engineering
September 2007
Hsinchu, Taiwan, Republic of China

中華民國九十六年九月

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

研究生：林銘彥　　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

本論文首先進行音節精確切割的處理，將輸入語音先經 HMM-based forced-alignment 做初步切割後，再使用 segmental k-mean algorithm 調整音節邊界，以獲得較準確的切割。接著將每一種音節分成數群，每一群取一個樣本作為未來 TTS 合成的音節波形樣本之用，做法是將每一音節先切割成數個 HMM 狀態，每一狀態抽取平均的特徵參數，再由同一種音節的狀態平均特徵參數構成一個特徵矩陣，經特徵值分解降維後，再使用向量量化分成數群，取出每一群的中心樣本。最後嘗試使用決策樹來由語言參數進行樣本的預測，以選取適當的樣本做語音的合成。實驗結果證實，將常用的 10 種音節各分成 5 群，其音節波形樣本都各自具有獨特的特徵，而由文字去預測音節合成樣本的準確度大約在 60%。

A Study on Unit Selection of Using Signal Characteristic for Corpus-based Mandarin TTS System

Student：Ming-Yan Lin　　Advisor：Dr. Sin-Horing Chen

Department of Communication Engineering

National Chiao Tung University

Abstract

In this thesis, the issue of precision speech segmentation is first addressed. A segmental k-mean algorithm is used to refine all syllable boundaries segmented by the HMM-based forced-alignment method. Then each syllable is segmented into several HMM states. A characteristic matrix is then constructed for each type of syllable by collecting the state-averaged feature vectors of all syllable samples. Dimension reduction is then performed by the singular value decomposition (SVD) method. Vector quantization is then applied to divide all samples of each syllable type into several clusters. A representative sample near the cluster center is then extracted. Lastly, a decision tree is constructed for each syllable type to build the mapping from text to these representative samples. Experimental results confirmed that all representative samples had their own distinctive properties as we extract 5 representative samples for each of 10 types of most frequent Mandarin syllables. The accuracy of representative sample prediction was around 60%.

誌謝

時間過的很快，一轉眼碩士的生涯即將尾聲，在這兩年多的求學生活，我首先要感謝的是陳信宏老師與王逸如老師，感謝你們的指導，讓我體會到思考的方法和做事的態度的重要性。同時也感謝江振宇學長如一座知識與技術的寶庫，讓我做起事來更是事倍功半。

接著還要感謝什麼事都能聊的巴金叔叔，愛講故事的輝哥，總是超神祕的希群，還有最愛強的阿德學長，及在程式上啟蒙我的東毅、見偟、國興學長，當然不能忘的，還有研究所的同學，包括熱血棒球的宏宇，總說自己為小小的獻文大大，高鐵榮譽會員啟風，愛凶的友駿，常插話的小鄧，看跑車的胤賢，常打球的小傅，及在神圖界和咨商界都堪稱大師的迷彩，還有一起走完最後一段的勇仔和士帆，希望大家未來都能闖出屬於自己的一片天。

同時還要感謝一位一直默默的守護著我的女友，在我面臨困難時不斷的給予打氣和加油，以實際的行動靜靜的叮寧和呵護，總在第一時間給予精神上的支持和心靈上的補給，最後還要感謝一路走來始終支持我的爸爸媽媽及家人，沒有你們的支持我也無法完成學業，謝謝你們，都在我生命中成為不能抹滅的一部分，我將用我所得不斷創造自己，盡我所能感恩及報答。

目錄

圖目錄

表目錄

第一章 緒論

1.1 研究動機

隨著科技的蓬勃發展，人們越來越仰賴電腦處理身邊的各項事務，電腦科技的發展也已從原本的運算能力導向變為以溝通與訊息交換為主要研究，從早期的工業電腦到現在可隨身攜帶的 NOTE BOOK，人與電腦之間的溝通方式顯得日趨重要。

觀察人類最自然的溝通方式，不外乎說與聽，能夠說出想表達的話(合成)，了解使用者的正確訊息(辨識)，為了讓人機之間能夠同樣使用如此溝通方式，語音合成和語音辨識技術的研究顯的格外重要。

現今主流的語音合成方法，是以載字句錄製音節以及錄製一個大型語音資料庫為基礎之合成方法，載字句錄音法把要錄製的合成單元鑲在一個句子中一起錄音，最後再將它切出來，這種合成單元本身具有連續語音特性，因此合成單元間的差異，可直接影響到合成語句的自然度和流暢度。

1.2 研究方向

語音合成系統中，在合成單元的選取上，常見的做法包括依據 Cost Function 挑選及使用 Decision Tree 挑選，方法中雖引入多種參數，卻鮮少對音節自身信號做研究和分析。

本論文提出對音節(語音訊號)抽取特徵參數，建立音節特徵矩陣，同時將音節做分群與標記，探討不同群組與合成語音品質的關連。首先將音節邊界找出，做為抽取音節特徵參數的區域，接著對每一音節建立其特徵矩陣，確立特徵矩陣與韻律參數的相關性後，使用 VQ 將音節分群與標記，同時使用 Decision Tree 做進一步驗證，最後以特徵矩陣搭配 Decision Tree 做一聽覺實驗，驗證音節分群屬性與合成語音主觀評量之關係。

1.3 章節概要

本論文共分為五章：

第一章　緒論：介紹本論文之研究動機與方向。

第二章　語料庫音節邊界之擷取：描述本論文如何修正 HMM 之音節邊界。

第三章　音節特徵矩陣之建立和分類：介紹如何建立 MFCC、Energy、Duration 特徵矩陣，同時觀察特徵矩陣和韻律參數之關係，分類音節與波型之關係。

第四章　實驗設計結果與分析：使用 Decision Tree 驗證第三章的觀察，同時以一聽覺實驗證實特徵矩陣與合成語句主觀評量之關係。

第五章　結論與未來展望。

第二章 語料庫音節邊界之擷取

採用大型語料庫為基礎之 TTS 系統(Corpus-base TTS)，首要工作在建立一個足夠大型的語音資料庫，以供未來語音合成單元的挑選，同時語料庫中每個可能構成合成單元之音節，其音節邊界的決定對 Corpus-base TTS 系統所合成出的語音品質有相當重要的影響，音節邊界不夠準確將會增加合成單元在串接時彼此的不匹配程度，嚴重時甚至影響輸出的語意，因此本章將提出如何提升語料庫音節邊界準確度的作法，以做為往後研究之基礎。

2.1 語料庫之說明

本論文所使用的語料為自己錄製的 Treebank 語料庫，語料文字內容為來自中央研究院中文文句結構樹資料庫 1.1 版(Sinica Treebank Version 1.1)，我們請了一位專業的女性廣播員幫我們錄音，其錄音軟硬體設備及格式詳細如下：

表格 2-1：錄音軟硬體設備規格表

錄音軟體	Cool Edit Pro 直接錄成聲音檔案
麥克風	單一指向性(uni-directional)
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(sampling rate)	20kHz
發音速度	每秒約 4.6 個音節
取樣大小	16 bits(位元)
聲道	單聲(mono)
檔案格式	pcm

2.2 基本HMM切割之建立

2.2.1 Acoustic model 建立

實驗之初，我們以手工切割 treebank 語料庫前 110 句，排除第 10 句以後，個位數編號 1,6,9 的音檔，其餘語料(80 句共 10,217 syllables)作為 HMM 之訓練語料，而排除之語料(30 句共 3,824 syllables)作為往後所需之測試語料。

以下為所使用參數及參數設定：

- 12 維 MFCC 參數。
- 1 維能量參數
- frame size：32ms。
- frame shift：10ms。

接著使用英國劍橋大學所開發之 HMM ToolKit(HTK)訓練 HMM 模型。

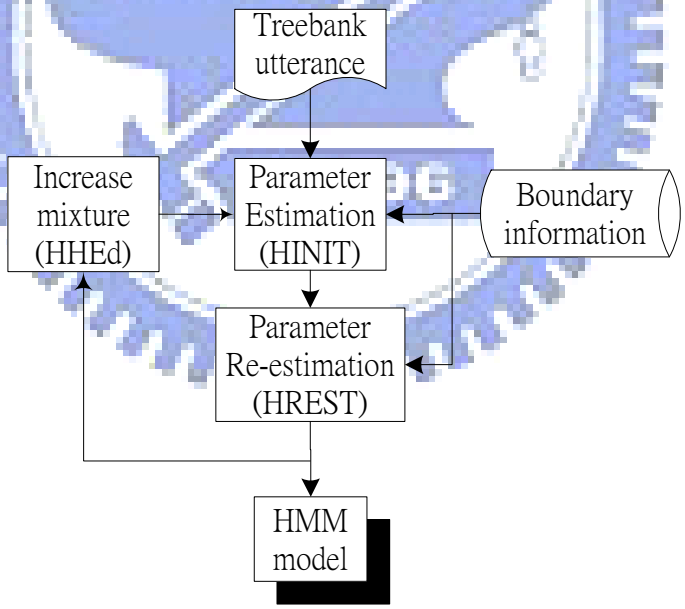


圖 2-1 已知切割位置之 HMM 模型訓練

實驗結果

下表為將所訓練得來之HMM執行forced – alignment切割位置和以手工標記做為標準答案之切割位置評比。

表格 2-2 已知切割位置之 HMM 模型 forced – alignment 之切割位置評量

	Tolerance	5ms	10ms	15ms	20ms	25ms	30ms	35ms	40ms
Inside test	Syllable begin	24.4	48.4	67.9	80.8	88.5	91.5	93.7	95.1
	Syllable end	25.9	46.8	62.9	74.3	80.9	85.2	88.3	90.6
Outside test	Syllable begin	24.5	48.8	68.5	81.2	88.7	91.7	94.0	95.1
	Syllable end	25.0	45.5	62.2	74.3	81.3	85.6	88.7	91.0

我們發現單純使用基本 HMM 模型所得之結果不盡理想，若直接拿此切割位置做為往後語音合成系統所需之切割位置，合成單元間的不匹配(Transition Cost)將會影響合成語句的品質，因此下一小節將提出如何進一步改進現有之切割位置。

2.3 音節切割位置修正

2.3.1 Segmental K-mean algorithm

語音訊號為時變訊號，受到語者/錄音環境等各種影響，因此若能利用盡量訊號本地的特性，結合預先訓練好的聲學模型，可改善原本單純利用預先訓練好的聲學模型，無視當地語音訊號特性所得之切割位置。

隨著人們發出不同的聲音，語音訊號跟著轉變，其特性(如 MFCC/能量)也有所不同，故若能適當的將本地語音訊號做片段性的分類，也可達到 forced – alignment 的作用，同時也可減少在事前訓練聲學模型時，需要大量的手工標記音檔所花的時間。

因此我們試著以 Segmental K-mean algorithm 不斷利用 K-means 將資料以片段分群，同時利用 Viterbi algorithm 找到 optimum segmental state sequence，反覆重新估算每個片段之模型直至 likelihood 收斂為止，進而改進原本使用基本 HMM 模型所得之切割位置。

實驗步驟

1. 使用手工標記的切割位置以音框長度為 20ms ，音框位移為 5ms，求出各種音素之 mean $\mathbf{u}_g$ 和 variance $\boldsymbol{\sigma}_g^2$ ，在此稱為 global model。其中 $\mathbf{u}_g = [u_1, u_2, ..., u_{12}, u_e]$; 為 12 維 MFCC 參數和 1 維能量參數之平均值，$\boldsymbol{\sigma}_g^2 = [\sigma_1^2, \sigma_2^2, ..., \sigma_{12}^2, \sigma_e^2]$; 為所對應之變異數。而相對於對多個手工標記音檔所得之 global model ($\boldsymbol{\Phi}_g \sim N(\mathbf{u}_g, \boldsymbol{\sigma}_g^2)$)，往後我們將稱對單一音檔所求之模型為 local model。
2. 搜集待修正切割位置音檔(單一音檔)之 silence 和 short pause (sp)的 12 維 MFCC 參數和 1 維能量參數，使用 VQ algorithm 將資料分成兩群，將能量大的設為 local breathe model，能量小的為 local sp model。
3. silence refinement：

   我們假設語音訊號在一個 syllable 結束進入下一個 syllable 的狀態轉移可為三種路徑。

   (1) 離開前一 syllable 的 finial model 後，直接進入下一 syllable 的 initial model。

   (2) 離開前一 syllable 的 finial model 後進入 local sp model，接著進入下一 syllable 的 initial model。

   (3) 離開前一 syllable 的 finial model 後，依序經過 local sp model、local breathe model、local sp model，進入下一 syllable 的 initial model。

   因此我們得到在區間前一 syllable(finial 部分)至後一 syllable(initial 部分)之狀態轉移如下圖。

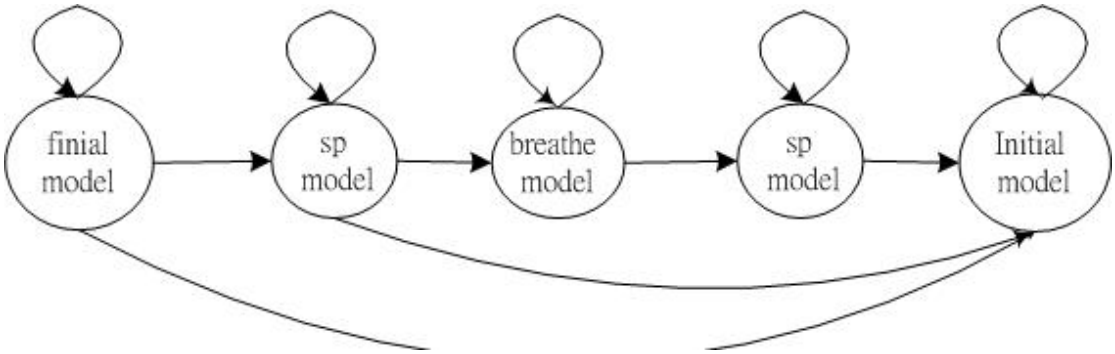


圖 2-2 silence refinement 之狀態轉移

其中 finial model 和 initial model 的初始狀態同 globe model。接著執行 Segmental K-mean algorithm 隨後在每次的 iteration 中，由更新的切割位置訓練新的 finial model 和 initial model，反覆執行直至 likelihood 收斂為止。

4. sp2 detection：

因步驟 3 所使用的 sp model 和 breathe model 皆由原始切割位置所得，並且限制從 finial model 到 breath model、breath model 到 initial mode 皆需通過 sp model，若一 finial 區域後段含有較嚴重的雜訊，如短暫的喉音或錄音環境之迴音部分(在此稱 sp2，如下圖)，則仍需剔除。

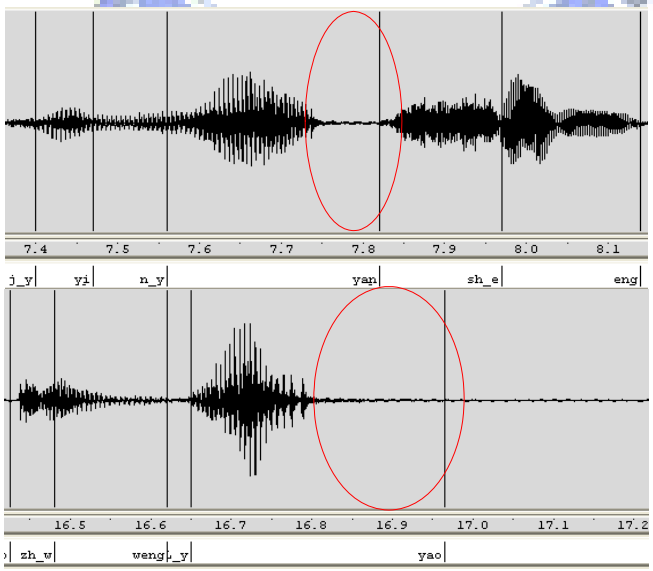


圖 2-3 silence refinement 後之切割位置

為了找出 sp2，我們採取的方法有兩步驟：

a. 搜尋可能為 sp2 之區間

直觀上可能為 sp2 的部分，皆屬 finial 區間後段能量較小的部分，固我們再次始用 Segmental K-mean algorithm，只使用 1 維能量參數，將 finial 區間分為兩片段，同時為了強化 Segmental K-mean algorithm 分出前段能量較大，後段能量較小的兩區間，初始值分別給定為前 1/2 之能量最大值和後 1/2 之能量最小值，執行完後將第二區間視為 sp2 可能的候選區間。

b. 判斷是否為 sp2

Defind likelihood：$p(\mathbf{o}_i \mid \mathbf{\Phi}_j) = \frac{1}{2\pi \left|\widehat{\sum}_j\right|^{1/2}} \exp[-\frac{1}{2}(\mathbf{o}_i - \hat{\boldsymbol{\mu}}_j)^t \times \widehat{\sum}_i^{-1} (\mathbf{o}_i - \hat{\boldsymbol{\mu}}_j)]$ (2.1)

使用 13 維參數，將候選區間所有的 frame $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_N)$ 分別帶入 globe finial model ($\mathbf{\Phi}_{gf}$) 和 globe sp2 model ($\mathbf{\Phi}_{sp2} \sim N(\mathbf{u}_{sp_2}, \boldsymbol{\sigma}^2_{sp_2})$)，計算區間中若干個 frame，likelihood sp2 大於 likelihood finial，將 finial-sp 之邊界向前位移 k 個 frame shift (如下)。

```
k = 0;
for i = 1,2,⋯,N
    if (p(o_i | Φ_sp2) > p(o_i | Φ_gf))
        k++;
    end
end
```
(2.2)

修正後的結果如下：

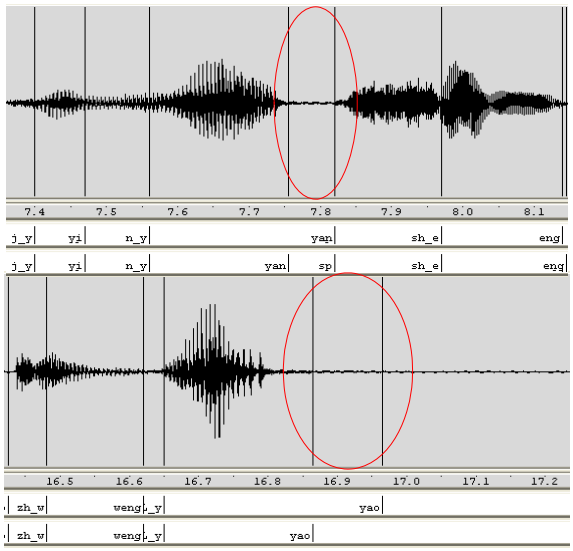


圖 2-4 經過 sp2 detection 後之結果

其中 global sp2 model 為由少許人工標記片段所求而得。

5. 以上 4 步驟得到較準確的切割位置。

以下為流程圖：

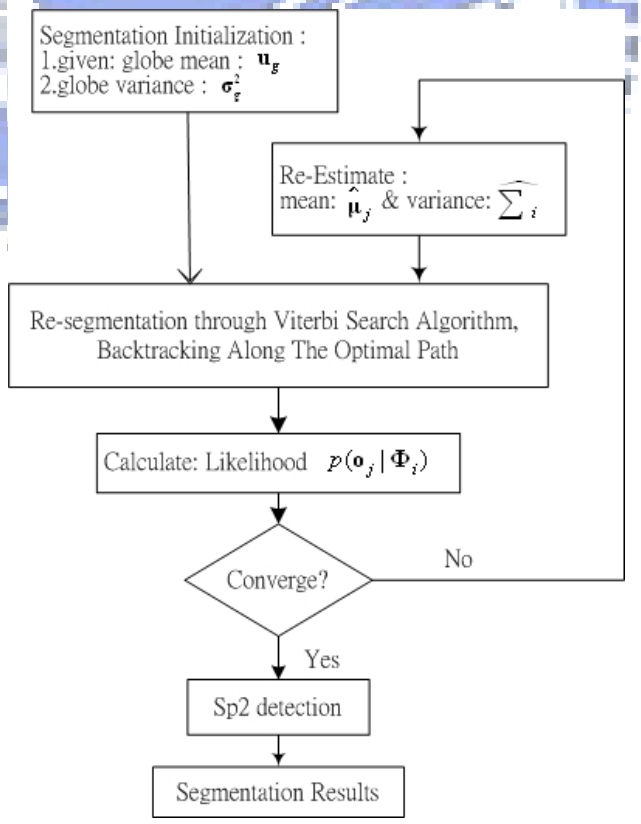


圖 2-5 修改音節切割位置流程圖

實驗結果

表格 2-3 切割位置修改前後之比較(內部測試)

	Tolerance	5ms	10ms	15ms	20ms	25ms	30ms	35ms	40ms
修正前	Syllable begin	24.4	48.4	67.9	80.8	88.5	91.5	93.7	95.1
	Syllable end	25.9	46.8	62.9	74.3	80.9	85.2	88.3	90.6
修正後	Syllable begin	35.2	64.2	79.8	88.0	92.4	94.6	96.2	97.2
	Syllable end	28.1	49.9	66.1	76.8	83.6	88.0	91.1	93.4

表格 2-4 切割位置修改前後之比較(外部測試)

	Tolerance	5ms	10ms	15ms	20ms	25ms	30ms	35ms	40ms
修正前	Syllable begin	24.5	48.8	68.5	81.2	88.7	91.7	94.0	95.1
	Syllable end	25.0	45.5	62.2	74.3	81.3	85.6	88.7	91.0
修正後	Syllable begin	36.8	64.4	78.5	86.0	90.6	93.0	94.8	96.1
	Syllable end	26.2	48.3	63.2	74.0	80.4	85.8	89.0	92.4

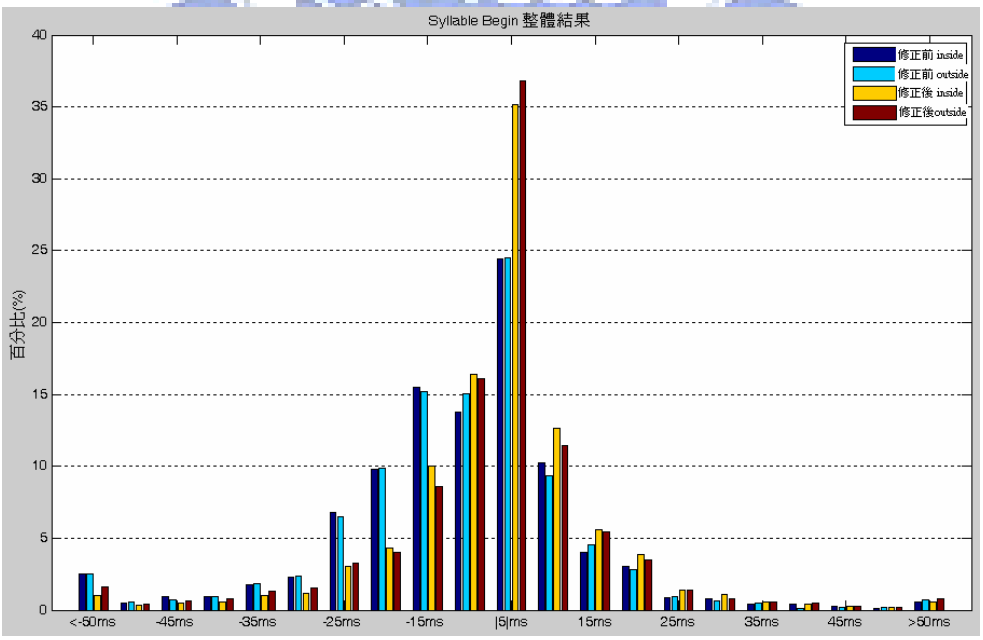


圖 2-6 Syllable Begin 切割位置修改前後比較

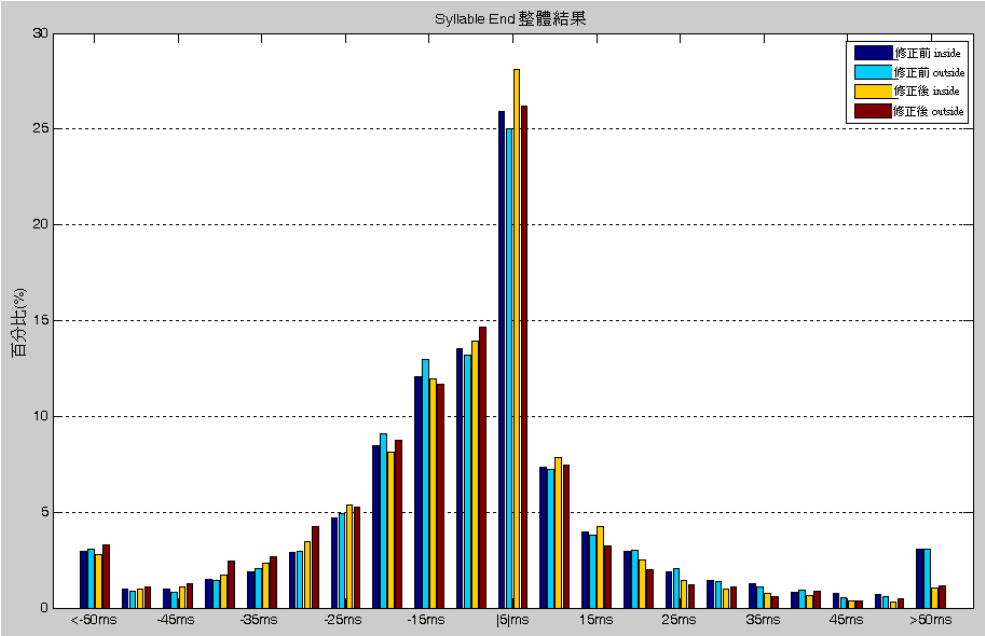


圖 2-7　Syllable End 切割位置修改前後比較

整體來說切割位置的確獲得改善，syllable begin 較為明顯，syllable end 則是略微往前一 syllable 移動，我們將在下一小節進一步分析改善後的切割位置。

2.4 修正後的切割位置分析

我們考慮三種不同的 short pause 長度，分開觀察切割位置 B 和切割位置 C 的差異。

a. 原手工標記 sp duration < 5ms

(inside test 佔 56.5% / outside test 佔 56.7%)。

b. 原手工標記 sp duration 介於 5ms ~ 50ms

(inside test 佔 21.5% / outside test 佔 20.6%)。

c. 原手工標記 sp duration > 50ms。

(inside test 佔 22.0% / outside test 佔 22.7%)

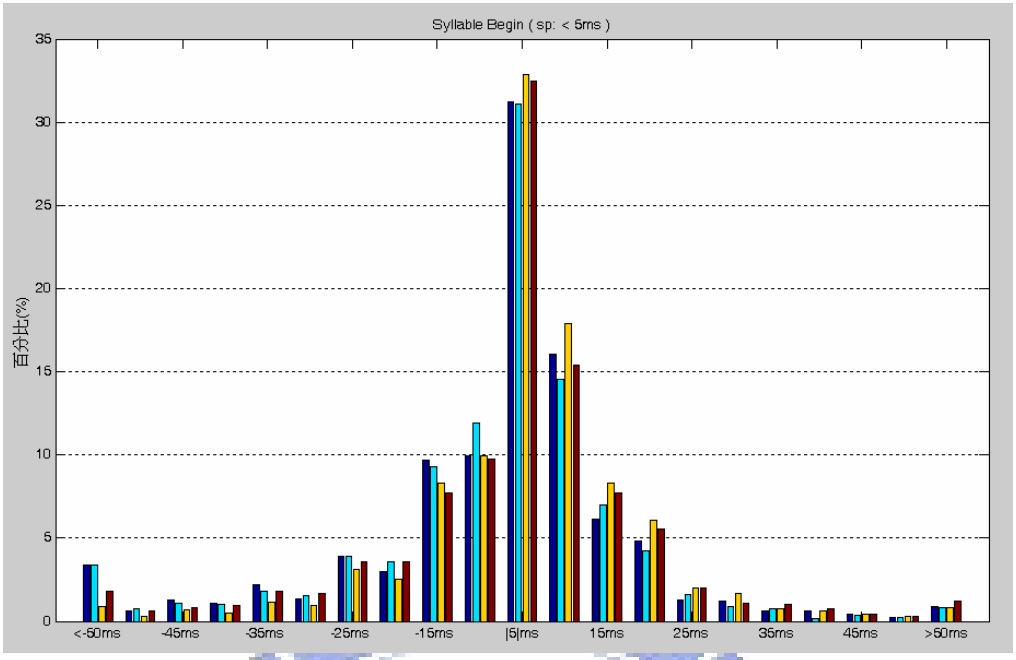


圖 2-8　Sp < 5ms Syllable Begin 切割位置修改前後比較

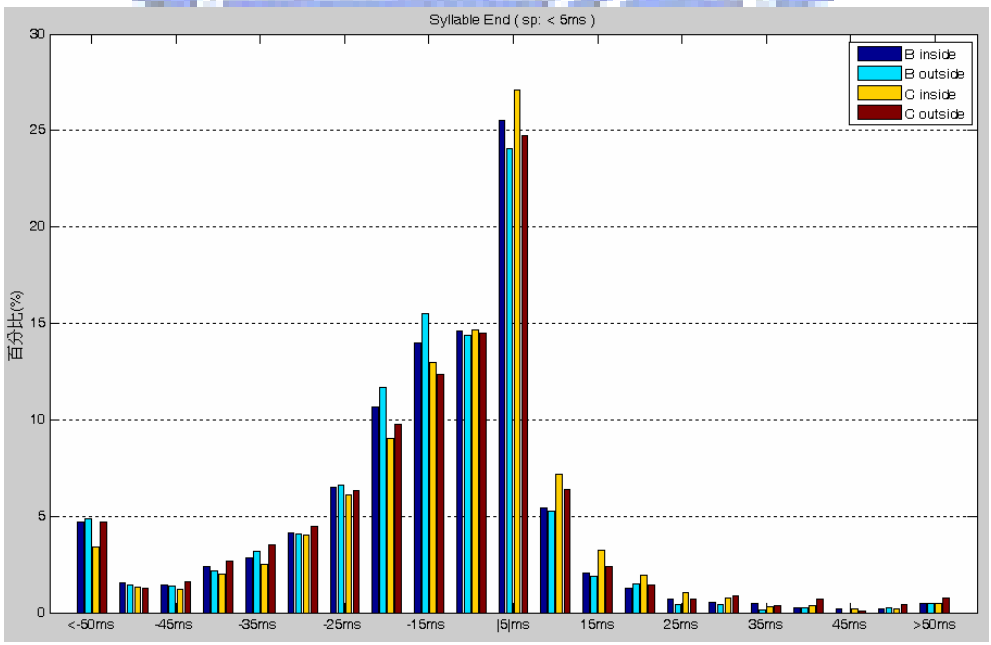


圖 2-9　Sp < 5ms Syllable End 切割位置修改前後比較

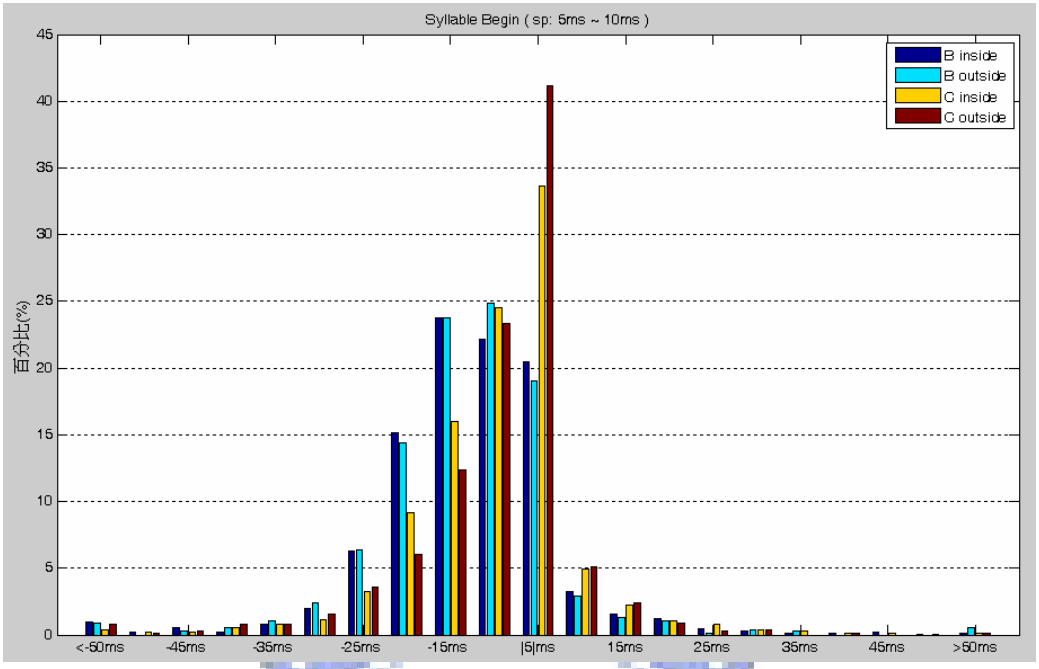


圖 2-10　Sp：5~50 ms Syllable Begin 切割位置修改前後比較

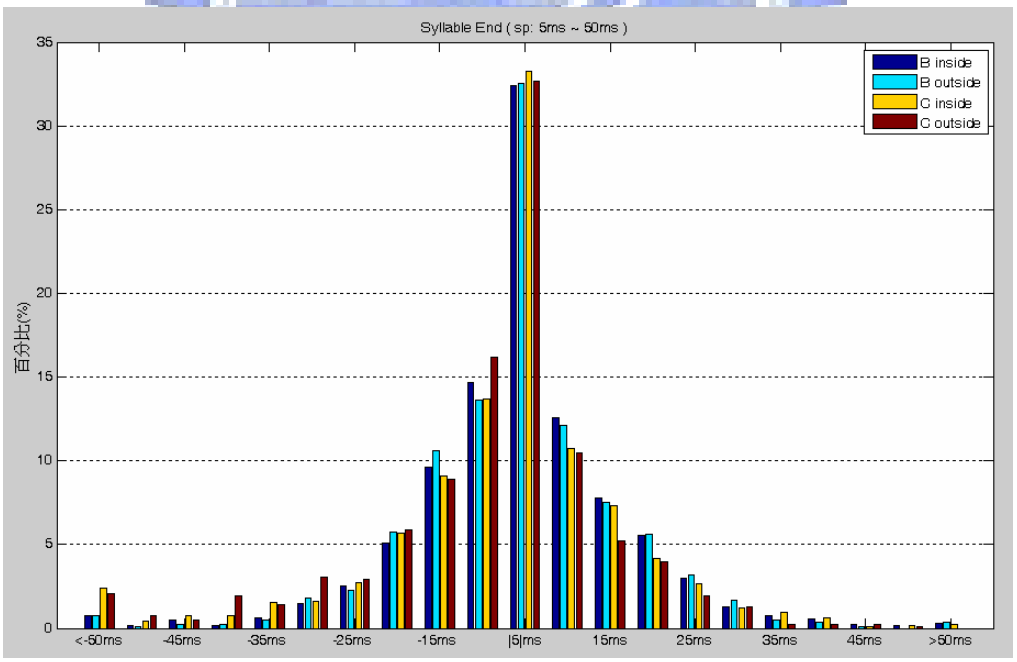


圖 2-11　Sp：5~50ms Syllable End 切割位置修改前後比較

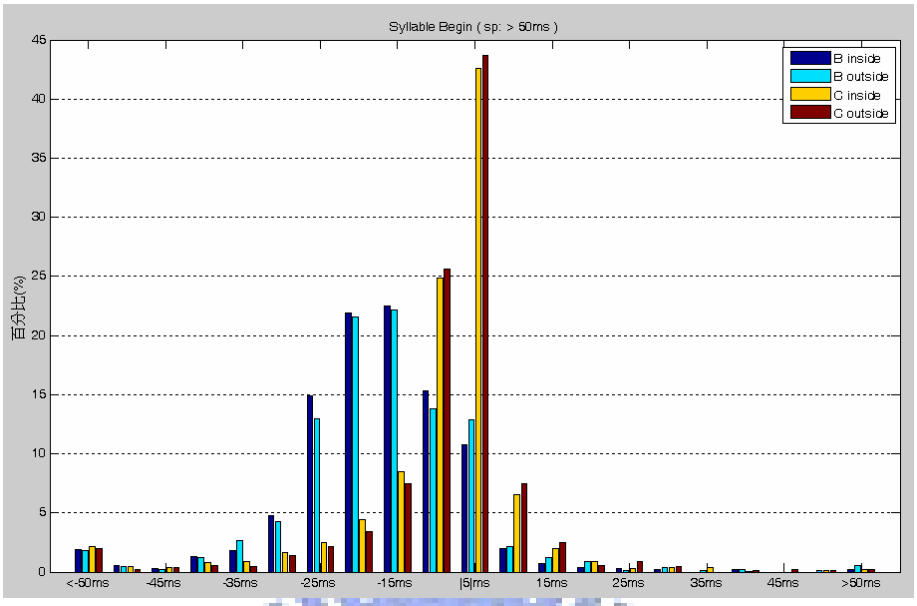


圖 2-12　Sp > 50ms Syllable Begin 切割位置修改前後比較

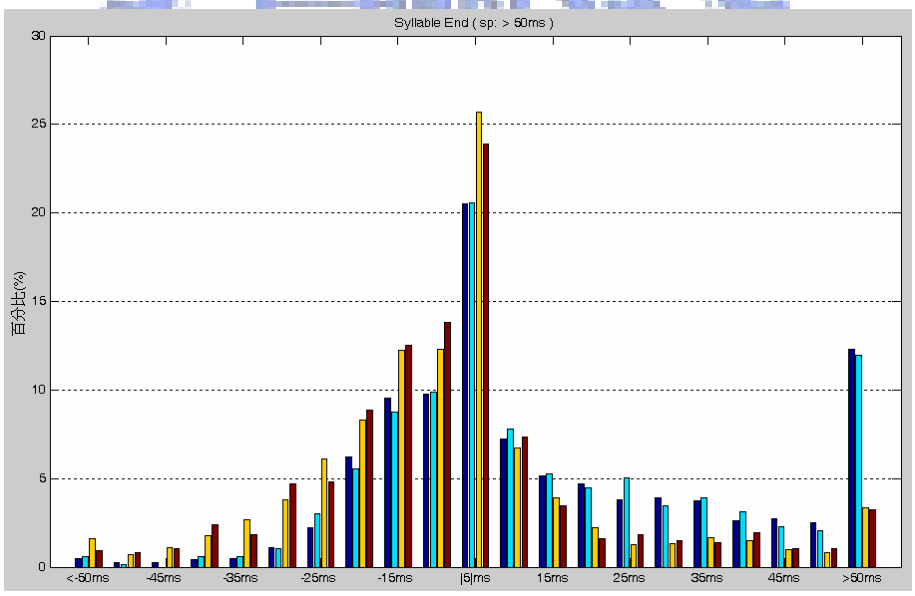


圖 2-13　Sp > 50ms Syllable End 切割位置修改前後比較

由觀察可得到下面幾點結論：

a. 在原手工標記 sp duration < 5ms 時，修改前後改變有限。

b. 在原手工標記 sp duration 介於 5ms ~ 50ms 的情況下，syllable begin 在切割位

置誤差在 20ms~30ms 時，都能有不錯的修正能力，而 syllable end 整體上則是都會往前一 syllable 一側移動。

c. 在原手工標記 sp duration > 50ms 時，syllable begin 在切割位置誤差在 20ms ~ 30ms 時，和 sp duration 介於 5ms ~ 50ms 的結果一致，都能有不錯的修正能力，而這也是基本 HMM 切割容易有誤差的區間，而 syllable end 整體上則是會改善原本過長的 short pause 長度，同時往前一 syllable 移動。

d. 我們發現在含有 short pause 中，只要原先的切割位置不偏離太遠(約 30ms) Segmental K-mean algorithm 能帶來良好的改善，而如果原先切割位置就偏離太遠，則此方法會因為本來 local 的資料不好，而無力改善。

小結

由觀察中發現 Segmental K-mean algorithm 受到原先切割位置的影響，若原始切割位置即有較大的偏離則難以調回，亦或雖然原先切割位置沒有較大的偏移量，但因原先切割位置兩側 finial phoneme 區間和 initial phoneme 區間所求出的 mean 和 variance 已經被鄰近的較為強烈之信號所影響(如下圖)則也無法調回期待之位置。

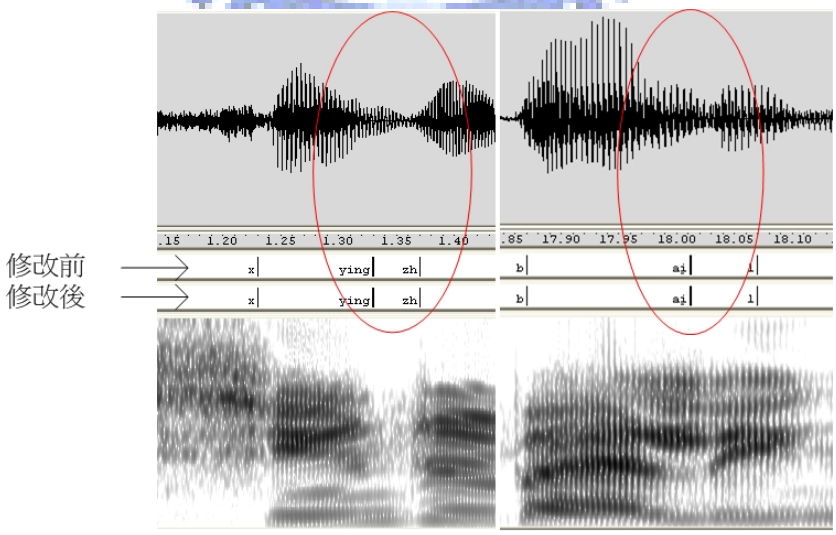


圖 2-14 尚未更正之例子

雖然修改後的切割位置並沒有達到百分百的完美，但已有相當的改善，往後我們將使用此較好之切割位置做進一步之研究。

第三章 音節特徵矩陣之建立和分類

一個良好的 Corpus-base TTS 系統，除了要求音節邊界的準確度外，如何挑選出合適的合成單元亦是重要的課題，因語音訊號為時變訊號，即使是同一語者講述同一的音節，此一音節之特性亦會受到此前後音節、語法結構、文法角色、和語者當時的心情、健康狀態等影響，而導致每一音節其語音訊號都不全然相同。

在大型語料庫中，常會遇到的課題便是同一音節在語料庫中出現極多次數，即使加上許多文法上的限制在挑選合成單元時，同一音節還是會有不少選項(出現頻率高之音節尤其明顯)，因此此一章節我們試著將語料庫中出現頻率前十名之聲調音節，分別建立特徵矩陣，分類並予以標記觀察其特性。

3.1 音節特徵矩陣之建立

3.1.1 音節內部之分割

為了建立符合一音節特徵的矩陣，必須取出能代表一音節特徵的參數，而語音訊號的特色為有時序的特性，每一音節能夠視為多個片段，多個不同特色之訊號所串接成。因此在固定住音節的邊界後，我們設定音框長度為 32ms，音框位移為 10ms ，使用 12 維 MFCC 參數訓練出 Syllable 之模型，並使用此模型和 Viterbi algorithm 在 Syllable 內部分割出數個特性相似的狀態，而音節發音含空韻母(FNULL)/空聲母(INULL)及聲母為ㄅㄉㄍ三類的 Syllable，其轉換狀態為 6 個 State，其餘音節為 8 個 State，狀態和狀態之間不可跳躍。其中訓練之聲調音節分別為：de5、yi4、bu4、ren2、shi4、you3、zai4、wo3、yi3、ta1。

表格 3-1 各音節、內部狀態平均長度

		音節	state1	state2	state3	state4	state5	state6	state7	state8
de5(ㄉㄜ˙)	平均長度(ms)	111.2	16.4	23.1	25.4	18.4	13.7	13.6		
	標準差	35.9	8	17.6	24	18.9	9.4	11.5		
yi4(ㄧㄚˋ)	平均長度(ms)	151.3	32.7	39.2	30	18.6	15.9	14.4		
	標準差	47.1	30.8	37.1	28.1	13.9	10.5	9.3		
bu4(ㄅㄨㄚˋ)	平均長度(ms)	113.6	20.6	21.4	26.8	18	13.6	13.2		
	標準差	40.9	11.1	15.9	23.1	19.9	7.9	7.4		
ren2(ㄖㄣˊ)	平均長度(ms)	213.2	30.1	20.9	31.8	34.4	23.5	22.2	31.5	17.4
	標準差	55.6	20.8	11.4	22.5	27.4	16.7	16.5	24	13.3
shi4(ㄕㄚˋ)	平均長度(ms)	216.1	29.6	65.5	21.4	41.4	37.2	22.2		
	標準差	54.3	22.8	32.4	10.3	30.2	25.6	15.9		
you3(ㄧㄡˇ)	平均長度(ms)	182.3	38.1	25.2	34.7	36.5	25.5	21.9		
	標準差	228	122	10.6	22.7	24	21.6	102		
zai4(ㄗㄞˋ)	平均長度(ms)	213.1	33.7	13.8	18.6	38.1	31.6	29	26.7	21.5
	標準差	43.3	22.8	5	8.6	30.8	27.4	18.6	16.8	19.8
wo3(ㄨㄛˇ)	平均長度(ms)	126.9	23.9	10	10	10	49.1	25		
	標準差	44.7	22.2	0	0	0	41	26.8		
yi3(ㄧˇ)	平均長度(ms)	145.3	28.1	36.4	28.2	21.5	17.2	14.1		
	標準差	44.8	25.8	35.4	23.3	13.9	12	9.6		
ta1(ㄊㄚ)	平均長度(ms)	163.2	20.5	25.3	15	23.6	31	16.8	15.4	16.6
	標準差	49.4	16.8	16.8	8.4	21.7	27.1	10.8	10.3	12.7

3.1.2 MFCC、Energy、Duration 特徵矩陣之建立

在得到語料庫中每個音節內部的邊界後，我們對上述之 tonal syllable 抽取特徵參數，令音框長度為 20ms，音框位移為 5ms，分別建立 MFCC、Energy、Duration 特徵矩陣。

- **MFCC 特徵矩陣**

在每個狀態中抽取 12 維 MFCC 參數之平均值，排成一行向量 $\mathbf{\Phi}_i$，並且將相同 tonal syllable 之特徵行向量並排成一個 super vector $\mathbf{\Phi}$。即：

$$\boldsymbol{\Phi}_{MFCC}=[\boldsymbol{\Phi}_{1,MFCC}\quad \boldsymbol{\Phi}_{2,MFCC}\quad \cdots \quad \boldsymbol{\Phi}_{N,MFCC}]=\begin{bmatrix} \mathbf{V}_{1,1}^{s} & \mathbf{V}_{2,1}^{s} & \cdots & \mathbf{V}_{N,1}^{s} \\ \mathbf{V}_{1,2}^{s} & \mathbf{V}_{2,2}^{s} & \cdots & \mathbf{V}_{N,2}^{s} \\ \vdots & \vdots & & \vdots \\ \mathbf{V}_{1,Q}^{s} & \mathbf{V}_{2,Q}^{s} & & \mathbf{V}_{N,Q}^{s} \end{bmatrix} \tag{3.1}$$

其中 $\mathbf{V}_{n,q}^{s}$ 為每個狀態中抽取 12 維 MFCC 參數之平均值，數學式表示如下：

$$\mathbf{V}_{n,q}^{s}=\frac{\sum_{t=1}^{T}\mathbf{o}_{n,t}^{s}\delta(q_{n,t}^{s}=q)}{\sum_{t=1}^{T}\delta(q_{n,t}^{s}=q)} \tag{3.2}$$

$\mathbf{o}_{n,t}^{s}$ 為 tonal syllable s 第 n 個 sample 中的第 t 個音框之 12 維 MFCC 參數向量；$q_{n,t}^{s}$ 為 $\mathbf{o}_{n,t}^{s}$ 對應的 HMM state index。

- **Energy 特徵矩陣**

$$\boldsymbol{\Phi}_{E}=[\boldsymbol{\Phi}_{1,E}\quad \boldsymbol{\Phi}_{2,E}\quad \cdots \quad \boldsymbol{\Phi}_{N,E}]=\begin{bmatrix} E_{1,1}^{s} & E_{2,1}^{s} & \cdots & E_{N,1}^{s} \\ E_{1,2}^{s} & E_{2,2}^{s} & \cdots & E_{N,2}^{s} \\ \vdots & \vdots & & \vdots \\ E_{1,Q}^{s} & E_{2,Q}^{s} & & E_{N,Q}^{s} \\ pe_{b,1}^{s} & pe_{b,2}^{s} & & pe_{b,N}^{s} \\ pe_{e,1}^{s} & pe_{e,2}^{s} & & pe_{e,N}^{s} \end{bmatrix} \tag{3.3}$$

其中 $E_{n,q}^{s}$ 為每個狀態中能量參數之平均值，數學式表示如下：

$$E_{n,q}^{s}=\frac{\sum_{t=1}^{T}oe_{n,t}^{s}\delta(q_{n,t}^{s}=q)}{\sum_{t=1}^{T}\delta(q_{n,t}^{s}=q)} \tag{3.4}$$

$oe_{n,t}^{s}$ 為 tonal syllable s 第 n 個 sample 中的第 t 個音框之能量參數；$q_{n,t}^{s}$ 為 $o_{n,t}^{s}$ 對應

的 HMM state index； $pe^{s}_{b,n}$ 、 $pe^{s}_{e,n}$ 分別為第 n 個 sample 與前、後一個音節間 short pause (sp)中能量的最小值(energy dip)。

- **Duration 特徵矩陣**

$$\mathbf{\Phi}_D = [\mathbf{\Phi}_{1,D} \quad \mathbf{\Phi}_{2,D} \quad \cdots \quad \mathbf{\Phi}_{N,D}] = \begin{bmatrix} Dur^{s}_{1,s} & Dur^{s}_{2,s} & & Dur^{s}_{N,s} \\ Dur^{s}_{1,i} & Dur^{s}_{2,i} & & Dur^{s}_{N,i} \\ Dur^{s}_{1,f} & Dur^{s}_{2,f} & & Dur^{s}_{N,f} \\ D^{s}_{1,1} & D^{s}_{2,1} & & D^{s}_{N,1} \\ D^{s}_{1,2} & D^{s}_{2,2} & & D^{s}_{N,2} \\ \vdots & \vdots & & \vdots \\ D^{s}_{1,Q} & D^{s}_{2,Q} & & D^{s}_{N,Q} \\ pd^{s}_{b,1} & pd^{s}_{b,2} & & pd^{s}_{b,N} \\ pd^{s}_{e,1} & pd^{s}_{e,2} & & pd^{s}_{e,N} \end{bmatrix} \quad (3.5)$$

其中 $Dur^{s}_{1,s}$ 為 tonal syllable s 第 n 個 sample 之長度(ms)， $Dur^{s}_{1,i}$、$Dur^{s}_{1,f}$ 為 tonal syllable s 第 n 個 sample 之 initial、finial 在 $Dur^{s}_{1,s}$ 所佔的比率(0~1)， $D^{s}_{n,q}$ 為每個狀態之 duration 長度在整個音節長度中所佔的比率(0~1)， $pd^{s}_{b,n}$ 、 $pd^{s}_{e,n}$ 分別為第 n 個 sample 與前、後一個音節間的 short pause (sp) 長度。

3.2 使用主成分分析法(Principal Component Analysis, PCA)音節分類

在得到後一 tonal syllable 的特徵矩陣 $\mathbf{\Phi}$ 後，我們先將每一變數減去其平均值，同時為了方便觀察和降低運算量，我們將減去平均值的 super vector $\mathbf{\Phi}'$ 做奇異值分解(singular value decomposition，SVD)。

$$\mathbf{\Phi}' = [\mathbf{\Phi}_1' \quad \mathbf{\Phi}_2' \quad \cdots \quad \mathbf{\Phi}_N']_{M\times N} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_N]_{M\times R} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_R \end{bmatrix}_{R\times R} [\mathbf{\Psi}]_{R\times N}$$

where $R = \min(M, N)$ (3.6)

其中：$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_R$ 。

為了進一步確定我們所建構的特徵矩陣和韻律參數有一定的相關性，我們將前幾維特徵向量 $\mathbf{u}_i$ 和語料中每一個 tonal syllable 減去平均值的特徵參數 $\mathbf{\Phi}_i'$ 和做內積，得到每一個 $\mathbf{\Phi}_i'$ 在主軸 $\mathbf{u}_i$ 上的投影量

$$\mathrm{Proj}_{\mathbf{u}_i}(\mathbf{\Phi}') = \mathbf{\Phi}'^{t} \cdot \mathbf{u}_i \quad (3.7)$$

求出此每個投影量和相對應音節之能量平均值(Energy mean)、音節長度(Syllable duration)、Pitch 相關係數，觀察其特性：

表格 3-2 MFCC 特徵矩陣之主軸和 Pitch/Energy/Duration 相關係數

		軸 1	軸 2	軸 3	軸 4	軸 5	軸 6	軸 7	軸 8	軸 9
de5	Pitch	0.11	0.17	-0.11	0.34	0.12	0.32	0.2	-0.1	0.21
	Energy	-0.13	-0.19	-0.02	0.39	0.11	0.16	0.23	-0.14	0.22
	Duration	-0.45	-0.27	0.16	0.05	-0.08	-0.07	-0.04	0.02	0.03
yi4	Pitch	0.6	0.23	0.05	-0.21	-0.15	-0.2	0.08	-0.25	-0.05
	Energy	0.55	-0.07	0.17	-0.08	-0.08	-0.38	0.18	-0.29	0.03
	Duration	-0.46	0.04	0.07	0.05	-0.13	-0.05	0.12	0.02	0.01
bu4	Pitch	0.11	0.37	-0.27	-0.25	-0.11	-0.2	0.06	-0.04	0.17
	Energy	0.12	0.12	0	-0.19	0.09	-0.03	-0.13	-0.03	0.27
	Duration	-0.48	-0.18	0.2	-0.08	0.25	0.25	0.03	-0.07	-0.09
ren2	Pitch	0.28	0.53	-0.4	-0.06	0.14	0.24	-0.04	-0.16	0.03
	Energy	0.3	0.46	-0.41	-0.2	0.21	0.18	0.13	-0.09	0.05
	Duration	-0.1	-0.18	0.06	-0.12	-0.35	0.27	-0.07	-0.01	-0.05
shi4	Pitch	-0.44	0.17	-0.27	-0.05	-0.14	-0.08	0.06	-0.1	0.07
	Energy	-0.21	0.12	-0.17	0.1	-0.16	0.11	0.08	-0.08	0.08
	Duration	0.38	-0.36	0.09	-0.09	0.08	-0.12	-0.1	0.03	0

you3	Pitch	0.13	0.27	0.54	-0.04	0.12	0.02	-0.14	0.36	0.02
	Energy	0.29	0.29	0.35	0.19	-0.02	0.13	-0.25	0.17	0.03
	Duration	-0.12	0.01	-0.01	-0.05	0.06	-0.02	0.03	-0.03	-0.01
zai4	Pitch	-0.03	0.08	-0.71	0.06	-0.13	0.05	-0.03	-0.33	0.11
	Energy	0.2	0.27	-0.25	-0.14	-0.02	-0.27	0.03	-0.23	-0.08
	Duration	-0.18	-0.24	0.03	0.09	-0.38	0.04	0.03	0	-0.09
wo3	Pitch	0.21	0.3	0.36	0.08	0.15	-0.2	-0.24	-0.02	0.04
	Energy	0.22	0.46	0.4	0.02	-0.04	-0.22	-0.17	0.11	0
	Duration	-0.35	-0.53	0.04	-0.14	0	0.31	-0.11	-0.06	-0.09
yi3	Pitch	-0.26	-0.21	0.18	0.31	-0.36	-0.04	0.22	0.05	0.04
	Energy	-0.44	0.01	0.15	0.41	-0.43	0.1	0.1	0.03	0.06
	Duration	0.16	0.1	-0.07	0.1	-0.18	0.01	-0.11	-0.08	-0.04
ta1	Pitch	0.06	-0.26	0.21	0.15	-0.35	-0.43	0.15	-0.13	0.15
	Energy	0.08	-0.27	0.08	0.07	-0.32	-0.08	-0.01	-0.07	0.15
	Duration	-0.24	0.36	0.16	0	-0.03	0.3	0.05	-0.05	0.12

由觀察可發現，MFCC 特徵矩陣做 SVD 後主軸的確和韻律參數相關，不同的音節有不同的相關性，而且隨著主軸排序的增加，和韻律參數的相關性也逐漸遞減。

表格 3-3 Energy 特徵矩陣之主軸和 Pitch/Duration 相關係數

		軸 1	軸 2	軸 3	軸 4	軸 5	軸 6
de5	Pitch	-0.48	0.3	0.02	-0.06	0.12	-0.02
	Duration	0.15	0.19	-0.25	0.19	0.33	-0.16
yi4	Pitch	-0.1	-0.58	-0.27	0	0.06	0.13
	Duration	-0.03	0.41	0.06	-0.06	0.04	0.27
bu4	Pitch	0.32	0.19	-0.03	0.12	-0.08	-0.06
	Duration	-0.45	-0.04	-0.24	-0.15	0.02	-0.09
ren2	Pitch	-0.69	0.1	-0.27	0.09	-0.01	0.03
	Duration	0.39	0.3	-0.3	0.08	-0.07	0.04
shi4	Pitch	0.33	-0.34	-0.12	-0.04	-0.13	0.05
	Duration	-0.32	0.19	0.19	0.39	0.16	0.03
you3	Pitch	-0.28	-0.46	-0.21	0.21	-0.09	-0.07
	Duration	0.04	0.1	-0.01	-0.01	0.01	-0.05
zai4	Pitch	0.38	0.05	0.02	-0.09	0.18	0.04
	Duration	-0.24	0.29	-0.08	-0.13	0.04	0.18

wo3	Pitch	0.27	-0.49	0.09	-0.05	-0.02	-0.23
	Duration	-0.53	0.3	-0.09	0.17	0.02	-0.2
yi3	Pitch	-0.16	-0.41	-0.41	0.11	-0.07	0.09
	Duration	-0.18	0.17	-0.13	0.11	-0.02	0.2
ta1	Pitch	0.22	-0.14	-0.21	0.02	0.11	0.21
	Duration	-0.59	-0.1	-0.01	0.01	0.02	-0.14

表格 3-4 Duration 特徵矩陣之主軸和 Pitch/Energy 相關係數

		軸 1	軸 2	軸 3	軸 4	軸 5	軸 6
de5	Pitch	0.22	-0.02	0.09	-0.07	0.1	0.06
	Energy	-0.13	-0.27	0.15	0.01	-0.01	0.09
yi4	Pitch	0.37	0.12	-0.12	0.18	0.07	0.17
	Energy	0.37	0.13	-0.04	0.32	0	0.17
bu4	Pitch	0.24	0.2	-0.05	0.03	0.01	0.09
	Energy	0.43	-0.12	-0.09	0.15	-0.04	0.03
ren2	Pitch	0.37	-0.13	0.1	0.02	-0.12	-0.12
	Energy	0.42	-0.11	0.08	0.03	-0.16	-0.26
shi4	Pitch	0.52	-0.03	-0.01	-0.11	0.01	-0.05
	Energy	0.32	-0.06	-0.07	0	-0.09	-0.19
you3	Pitch	-0.02	-0.05	0.19	0.06	-0.17	0.08
	Energy	0.07	-0.27	0.13	0	-0.23	0.06
zai4	Pitch	-0.26	0.19	-0.07	-0.05	-0.06	-0.03
	Energy	-0.37	0.11	0.24	0.31	0	0.08
wo3	Pitch	0.38	0.19	-0.09	-0.01	-0.06	-0.05
	Energy	0.35	0.17	0.02	-0.23	-0.03	0.02
yi3	Pitch	0.19	0	-0.19	0.15	-0.18	-0.14
	Energy	0.17	-0.02	-0.1	0.25	-0.19	-0.14
ta1	Pitch	0.14	0.15	-0.02	0.12	0.29	-0.1
	Energy	0.36	-0.04	0.01	0.38	0.1	0.04

同樣的，我們發現使用 Energy 特徵矩陣、Duration 特徵矩陣亦和韻律參數相關，隨著主軸排序的增加，相關性逐漸遞減。

在此定義降低維度至 K 維後，所能解釋的變異比例 R^2 ：

$$R^2 = \frac{\sum_{i=1}^{K} \lambda_i^2}{\sum_{i=1}^{R} \lambda_i^2} \times 100\% \tag{3.8}$$

在觀察各個特徵矩陣之奇異值後，令 MFCC 特徵矩陣 $K = 10$ ，Energy、Duration 特徵矩陣 $K = 5$ ，做為接下來研究所考慮的主軸，以下為各個音節特徵矩陣之解釋的變異比例 R^2 ：

表格 3-5 降低維度後，其 K 維主軸所能解釋之變異比例

	MFCC R^2	Energy R^2	Duration R^2
de5	68.51%	93.82%	98.77%
yi4	61.92%	96.45%	97.12%
bu4	67.13%	94.23%	99.62%
ren2	60.20%	95.72%	97.78%
shi4	60.05%	93.76%	98.90%
you3	63.15%	97.09%	99.09%
zai4	55.45%	90.14%	98.94%
wo3	69.56%	98.21%	99.62%
yi3	60.84%	97.37%	98.05%
ta1	57.03%	91.08%	99.19%

我們將第一維主軸 $\mathbf{u}_1$ 和減去其平均值後之特徵參數 $\mathbf{\Phi}'$ 做內積，使用 VQ 將音節分成 12 類，觀察其極端類別的特性，同樣的亦將所降維後的 K 維主軸 $\mathbf{u}$ 和減去其平均值後之特徵參數 $\mathbf{\Phi}'$ 做內積：

$$\mathbf{P} = \mathbf{\Phi}'^{t} \cdot \mathbf{u} = [P_1 \quad P_2 \quad \cdots \quad P_K];$$

其中 $\mathbf{u} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_K]$; 使用 VQ 將音節分成 5 類，並在下一節針對 MFCC 特徵矩陣之分類結果觀察其特性。

3.3 分類音節之特性觀察

3.3.1 第一主軸分類之極端例子

在觀察以 MFCC 特徵矩陣第一主軸所分出的極端的例子中(第一類和第十二類)，我們發現到主要的差異可由波型上觀察，如波型上是否完整或破碎，亦或有連音現像(單一側/兩側)，又或者因某些因素(如前後音節等)導致發音方式不同，語調不一致。以下有三個例子：

例 1： 下圖中四個波型 a、b、c、d 依順分別為：zai4、ta1、zai4、ta1

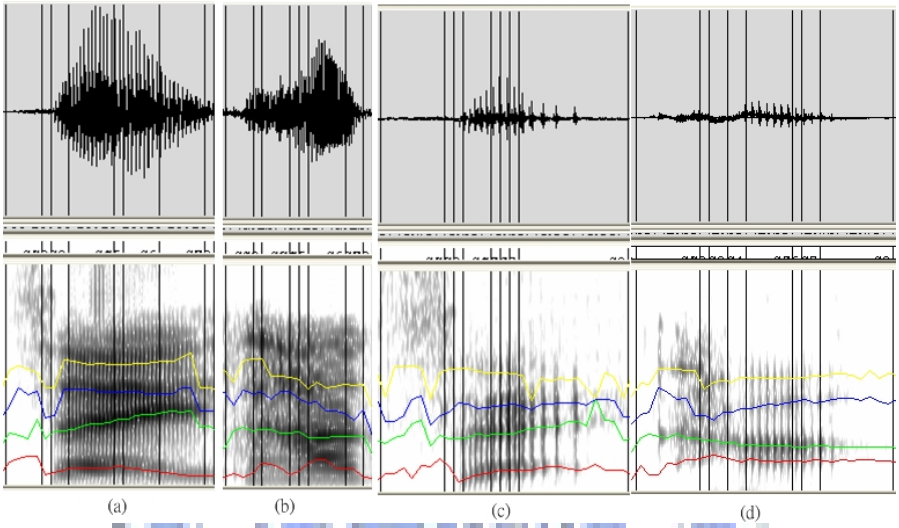


圖 3-1 波型完整與破碎之例子

由圖中可看出，後兩個波型比前兩個波型來得破碎，而且有喉音的現像。

例 2：下圖中四個波型 a、b、c、d 依順分別為：you3、yi3、you3、yi3

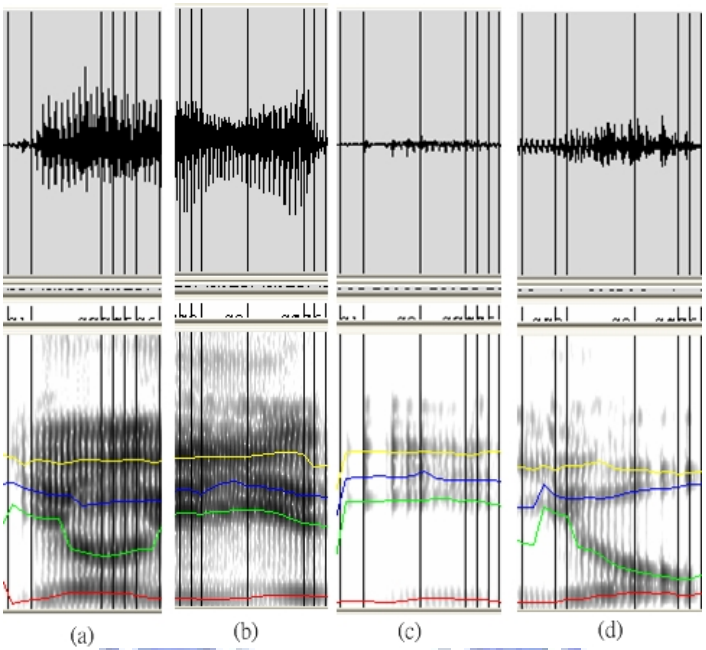


圖 3-2 連音差異之例子

由圖中可看出，前兩個波型分別和後、前音節有連音，而後兩波型則不明顯。

例 3：下圖中四個波型 a、b、c、d 依順分別為：bu4、bu4、ren2、ren2

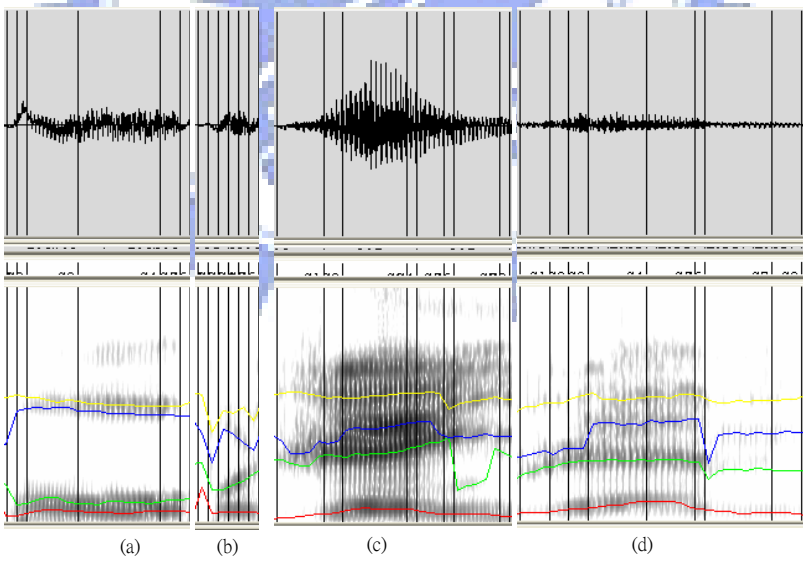


圖 3-3 波型明顯不同之例子

由圖中可看出前後兩音節明顯不同。

3.3.2 K 維主軸之分類群組之特性

在上一小節中，我們確信特徵矩陣之主軸的確可將音節依照某些特性分群，這一小節中，我們將原本考慮的第一維主軸增加為 K 維主軸，以提高可解釋的變異比例(R^2)，同時我們將音節分為 5 類。

在將所有音節之特徵矩陣降至 K 維分類後，相較於不同類，被歸為相同一類之音節必有其相似處，且愈靠近群組中心之例子愈能突顯出群組之特性，因此我們觀察每一類最接近群組中心的數個例子，找出特定群組之特性。

我們發現到，群組中心最容易有的特色為波型上是否和前後兩音節有明顯之連音。以下有兩個例子：

例 1：yi3 的第二類，其群組中心皆和上一音節連音，下圖明亮區域為 yi3。

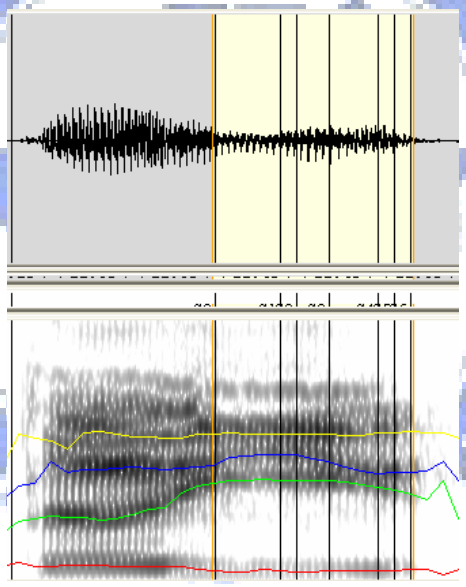


圖 3-4　yi3 群組中心之例子

例 2：de5 的第五類其群組中心皆和下一音節無連音，下圖明亮區域為 de5

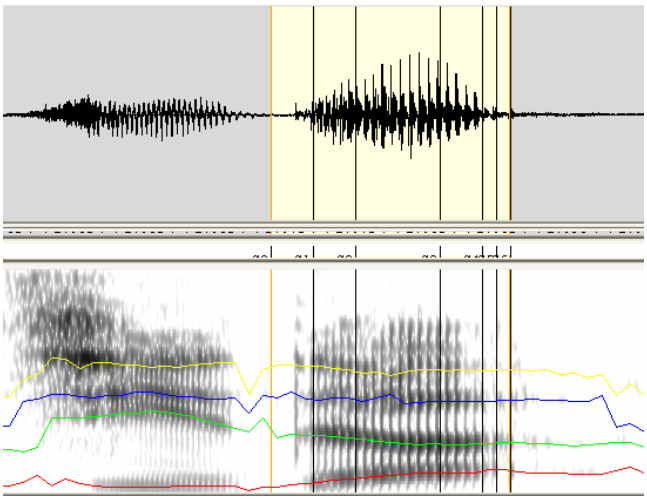


圖 3-5 de5 群組中心之例子

使用 K 維主軸分群，同一群組中之音節的確在波型上有其共通性，但並非每一群組之特性有其突出顯而易見之性質，因此我們在下一章中，引進 Decision Tree 協助我們觀察使用三種特徵矩陣主軸分群後之特性。

第四章 實驗設計結果與分析

在這一章中，我們使用 Decision Tree 進一步驗證上一章中所觀察到的結果，同時使用 Decision Tree 將音節分類之結果，和韻律參數、語言參數做連結，最後將 MFCC、Energy、Duration 特徵矩陣合併，使用 Decision Tree 找出特性相似和不相似的 node 做聽覺 AB test 進一步驗證。

4.1 Decision Tree 之建立

在此定義 Decision Tree 所使用之公式：

$$H_t(Y) = -\sum_{i=1}^{T} P(y_i \mid t) \log P(y_i \mid t) \tag{4.1}$$

$$P(y_i \mid t) = \frac{\sum_{n=1}^{N_t} \delta(c_n = i \mid t)}{\sum_{j=1}^{T} \sum_{n=i}^{N_t} \delta(c_n = j \mid t)} \tag{4.2}$$

$$\Delta H_t(Y) = H_t(Y) - (H_{t,l}(Y) + H_{t,r}(Y)) \tag{4.3}$$

其中 $H_t(Y)$、$H_{t,l}(Y)$、$H_{t,r}(Y)$ 分別為 node t 之 entropy 和 node t 之左、右兩個子節點之 entropy，Y 為 classification decision 之隨機變數，$P(y_i \mid t)$ 為在 node t 之中 class i 之機率，c_n 為 node t 中之第 n 個 tonal syllable 之 class，N_t 為 node t 中之 sample 數，T 為 tonal syllable 之分類數，在下一小節的實驗中 $T = 5$，且 stop criterion 為 $N_t < 20$ 或 $\Delta H_t(Y) < 0.01 H_t(Y)$。

- 以下為 Decision Tree 所需之問題集：

問題一：此聲調音節是否在句首？

問題二：此聲調音節是否在句尾？

問題三：此聲調音節是否為一字詞？

問題四：此聲調音節是否在詞首？

問題五：此聲調音節是否在詞中？

問題六：此聲調音節是否在詞尾？

問題七：下一音節之聲母是否為/空聲母(Null)/？

問題八：下一音節之聲母是否為/ㄅ(b)/、/ㄉ(d)/、/ㄍ(g)/？

問題九：下一音節之聲母是否為/ㄈ(f)/、/ㄏ(h)/、/ㄒ(x)/、/ㄕ(sh)/、/ㄙ(s)/？

問題十：下一音節之聲母是否為/ㄇ(m)/、/ㄋ(n)/、/ㄌ(l)/、/ㄖ(r)/？

問題十一：下一音節之聲母是否為 /ㄑ(q)/、/ㄔ(ch)/、/ㄘ(c)/？

問題十二：下一音節之聲母是否為 /ㄆ(p)/、/ㄊ(t)/、/ㄎ(k)/？

問題十三：下一音節之聲母是否為 /ㄐ(j)/、/ㄓ(zh)/、/ㄗ(z)/？

問題十四：上一音節之韻母是否為鼻音結尾韻母(Nasal Ending Vowel)？

/ㄢ(an)/、/ㄣ(en)/、/ㄤ(ang)/、/ㄥ(eng)/、/ㄧㄢ(yan)/、/ㄧㄣ(yin)/、/ㄧㄤ(yang)/、/ㄧㄥ(ying)/、/ㄨㄢ(wan)/、/ㄨㄣ(wen)/、/ㄨㄤ(wang)/、/ㄨㄥ(weng)/、/ㄩㄢ(yuan)/、/ㄩㄣ(yun)/、/ㄩㄥ(yung)/

問題十五：上一音節之韻母是否為開口(Open)韻母

/ㄚ(a)/、/ㄛ(o)/、/ㄜ(e)/、/ㄝ(eh)/、/ㄞ(ai)/、/ㄟ(ei)/、/ㄠ(ao)/、/ㄡ(ou)/、/ㄦ(er)/ /ㄧㄚ(ya)/、/ㄧㄝ(ye)/、/ㄧㄞ(yai)/、/ㄧㄠ(yao)/、/ㄧㄡ(you)/、/、/ㄧㄛ(yo)/、/ㄨㄚ(wa)/、/ㄨㄛ(wo)/、/ㄨㄞ(wai)/、/ㄨㄟ(wei)/、/ㄩㄝ(yue)/?

問題十六：上一音節之韻母是否為單韻母(Single Vowel)

/ㄧ(yi)/、/ㄨ(wu)/、/ㄩ(yu)/、/ㄚ(a)/、/ㄛ(o)/、/ㄜ(e)/、/ㄝ(eh)/、/ㄦ(er)/？

問題十七：上一音節之韻母是否為複韻母(Compound Vowel)

/ㄞ(ai)/、/ㄟ(ei)/、/ㄠ(ao)/、/ㄡ(ou)/、/ㄧㄚ(ya)/、/ㄧㄝ(ye)/、/ㄧㄞ(yai)/、/ㄧㄠ(yao)/、/ㄧㄡ(you)/、/ㄧㄛ(yo)/、/ㄨㄚ(wa)/、/ㄨㄛ(wo)/、/ㄨㄞ(wai)/、/ㄨㄟ(wei)/、/ㄩㄝ(yue)/？

問題十八：上一音節之韻母是否為空韻母 /FNULL1/、/FNULL2/？

問題十九：此聲調音節與其前音節之 Prosodic Break Type 為 B0？

問題二十：此聲調音節與其前音節之 Prosodic Break Type 為 B0 或 B1？

問題二十一：此聲調音節與其前音節之 Prosodic Break Type 為 B0、B1 或 B2-1？

問題二十二：此聲調音節與其前音節之 Prosodic Break Type 為 B0、B1、B2-1 或 B2-2？

問題二十三：此聲調音節與其前音節之 Prosodic Break Type 為 B3 或 B4?

問題二十四：此聲調音節與其後音節之 Prosodic Break Type 為 B0？

問題二十五：此聲調音節與其後音節之 Prosodic Break Type 為 B0 或 B1？

問題二十六：此聲調音節與其後音節之 Prosodic Break Type 為 B0、B1 或 B2-1？

問題二十七：此聲調音節與其後音節之 Prosodic Break Type 為 B0、B1、B2-1 或 B2-2？

問題二十八：此聲調音節與其後音節之 Prosodic Break Type 為 B3 或 B4?

4.2 分類音節之Decision Tree特性及辨識結果

4.2.1 分類音節之特性

首先使用 Decision Tree 驗證在上一章中所做出的兩個觀察：

- yi3 之 Decision Tree

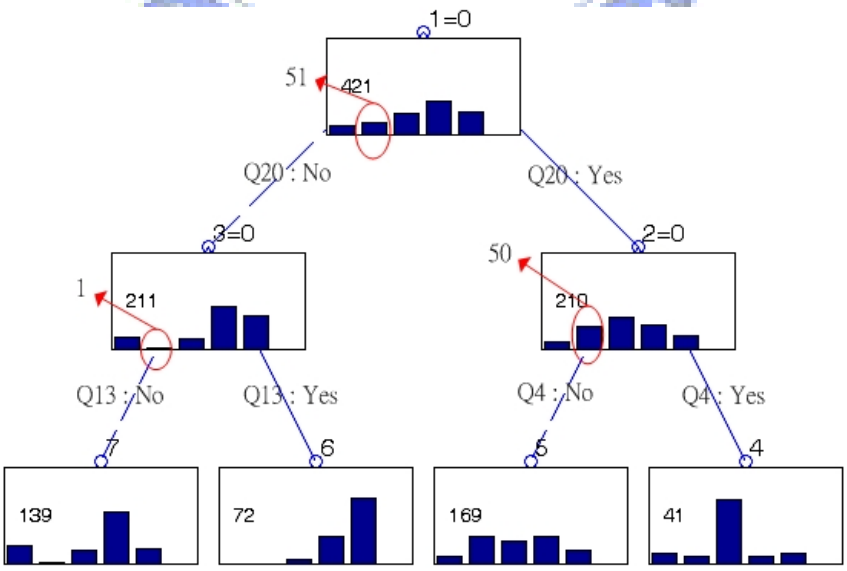


圖 4-1　yi3 之 Decision Tree

圖中每個小圖中的數字為此 node 中，五個音節類別總數，其中 node 下方為所被問之問題和答案，而前三個 node 的類別二被紅圈標記出來。

Root node 中類別二的 51 次，有 50 次是和前音節之 Prosodic Break Type 為 B0 或 B1，顯示出和上一章觀察的結果一致，與前音節有明顯的連音現象。

- de5 之 Decision Tree

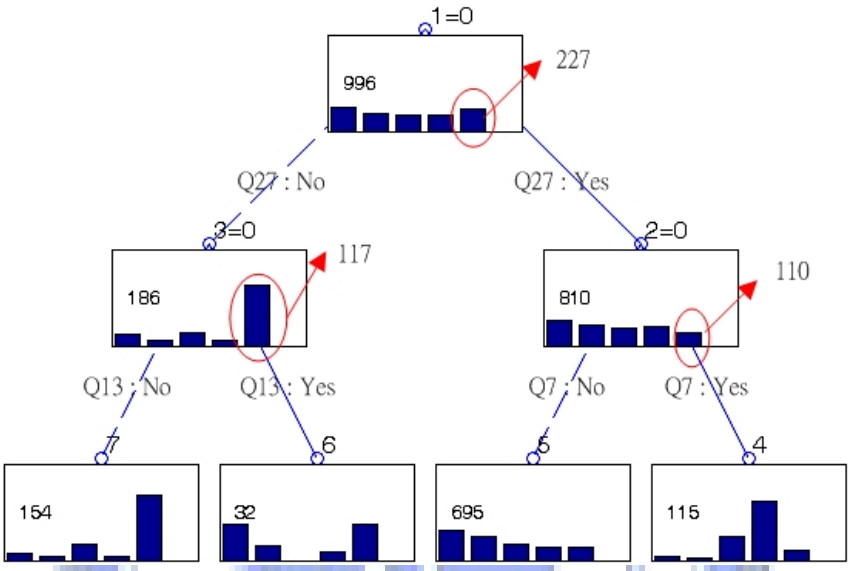


圖 4-2　de5 之 Decision Tree

類別五的 227 次中，有 117 中是後音節之 Prosodic Break Type 為 B3 或 B4，只佔類別五之 51.6%，但由圖中亦可發現到在 node3 中，類別五的次數明顯高出其他類別，甚致佔 node3 中 62.9%的數量，亦顯示出類別五為五種類別中和後方連音最不緊密的一類，因此也符合上一章的觀察。

4.2.2 進一步之分析

使用特徵矩陣之主軸所分群的特性，隨著不同的音節，不同的特徵矩陣而各有所不同，但我們亦想對三個不同的特徵矩陣多一些了解，因此我們統計出十個音節中，在三個不同的特徵矩陣其 Decision Tree 所提出的前五個不同的問題。

表格 4-1 十個音節在 Decision Tree 中被問的前 5 個不同的問題之統計

特徵矩陣 / 出現次數	MFCC	Energy	Duration
9 次			
8 次	Q27	Q21	
7 次		Q27	Q22/Q24/Q27
6 次		Q25	
5 次	Q25		Q21
4 次	Q22	Q9/Q19	Q28
3 次	Q13/Q20/Q28	Q8	Q1/Q30
2 次	Q7	Q2/ Q4/Q17/Q22/ Q24/ Q26/Q30	Q29/Q15/Q26/Q20/Q8
1 次	Q2/ Q3/Q9/ Q14/ Q16/Q19/Q29/	Q7/ Q10 /Q14/Q20	Q3/Q14/Q23/Q9

我們觀察出現四次以上的問題可以發現，不論是 MFCC、Energy、Duration 特徵矩陣，所出現的前五個不同的問題中，其中和韻律參數有較緊密的關聯，意即特徵矩能區分不同的 Break type。

而在語言參數方面，則包含下一音節之聲母為/ㄐ、ㄓ、ㄕ、空聲母、ㄇ、ㄋ、ㄌ、ㄖ/，或上一音節之韻母是否為鼻音結尾韻母(Nasal Ending Vowel)等特性容易區分，但緊密程度則不如韻律參數相關之問題。

4.2.3 Decision Tree之類別正確率

我們將每個 leaf node 貼上類別標籤，將 leaf node 中出現次數最多的類別視為此一 node 之標籤，做內部測試觀察每棵 Decision Tree 整體類別所判別之正確率。

表格 4-2 三種特徵矩陣之正確率

	MFCC(%)	Energy(%)	Duration(%)
de5	47.59	49.5	63.35
yi4	50.56	66.61	93.58
bu4	52.08	58.11	88.11
ren2	59.27	64.51	77.45
shi4	53.15	66.33	85.92
you3	49.69	68.92	88.75
zai4	50.19	64.4	75.49
wo3	58.6	60.64	82.22
yi3	52.73	63.9	85.99
ta1	56.03	62.5	82.37

MFCC、Energy 特徵矩陣正確率大多維持在 60%上下並不算理想，而 Duration 特徵矩陣雖然正確率較高，但所分出來的類別也不平均(附錄表 3)，表示我們企圖給 leaf node 貼上相同特性之標籤，同時將每個音節固定分出五類，並不能充份表示每個音節特性變異量所需的類別數，因此接下來的研究，將直接使用特徵矩陣之特性，而不對音節做分類。

4.3 合併特徵矩陣之Decision Tree特性

4.3.1 合併特徵矩陣

我們將上述的三個特徵矩陣(MFCC、Energy、Duration)合併成另一特徵矩陣 $\mathbf{\Phi}_{super}$，同樣使用 SVD 將特徵矩陣之維度降至 20 維並建立 Decision Tree。

$$\mathbf{\Phi}_{super}=[\mathbf{\Phi}_{1,super}\quad \mathbf{\Phi}_{2,super}\quad \cdots \quad \mathbf{\Phi}_{N,super}]=\begin{bmatrix} \mathbf{\Phi}_{1,MFCC} & \mathbf{\Phi}_{2,MFCC} & \cdots & \mathbf{\Phi}_{N,MFCC} \\ \mathbf{\Phi}_{1,E} & \mathbf{\Phi}_{2,E} & \cdots & \mathbf{\Phi}_{N,E} \\ \mathbf{\Phi}_{1,D} & \mathbf{\Phi}_{2,D} & \cdots & \mathbf{\Phi}_{N,D} \end{bmatrix} \tag{4.4}$$

表格 4-3 降低維度後，其主軸$K=20$所能解釋之變異比例

音節	*R2 (%)*	音節	*R2 (%)*
de5	99.93	you3	99.97
yi4	99.98	zai4	99.98
bu4	99.96	wo3	99.98
ren2	99.97	yi3	99.98
shi4	99.98	ta1	99.98

4.3.2 Decision Tree之建立與KL distance

在此定義 Decision Tree 所使用之公式：

$$\Delta L_t^q = -(\log \lambda_{t,\text{cov}}^q + \log \lambda_{t,mean}^q) \quad (4.5)$$

$$\lambda_{t,mean}^q = (1 + \frac{n_{t,l}^q \times n_{t,r}^q}{n_t^2}(\mu_{t,l}^q - \mu_{t,r}^q)^t W^{-1}(\mu_{t,l}^q - \mu_{t,r}^q))^{\frac{-n}{2}} \quad (4.6)$$

$$\lambda_{t,\text{cov}}^q = (\frac{\left|\Sigma_{t,l}^q\right|^{\alpha} \times \left|\Sigma_{t,r}^q\right|^{(1-\alpha)}}{\left|W_t^q\right|})^{\frac{n}{2}} \quad (4.7)$$

$$W_t^q = \frac{n_{t,l}^q}{n_t}\Sigma_{t,l}^q + \frac{n_{t,r}^q}{n_t}\Sigma_{t,r}^q \quad (4.8)$$

$$\alpha_t^q = \frac{n_{t,l}^q}{n_t} \quad (4.9)$$

$$n_t = n_{t,l} + n_{t,r} \quad (4.10)$$

其中，$\lambda_{t,mean}^q$、$\lambda_{t,\text{cov}}^q$ 分別為 $\mathbf{\Phi}_{\text{super}}$ SVD 後之參數在 node t 中，被詢問 question q 後所求出的平均向量相似度變化值之和、變異量矩陣所求出的相似度變化值，$\Sigma_{t,l}^q$、$\Sigma_{t,r}^q$、W_t^q 分別為 node t 中被詢問 question q 後所分裂出左右兩 node 之變異量矩陣及考慮權重之變異量矩陣，n_t、$n_{t,l}$、$n_{t,r}$ 分別為 node t 中之 sample 數和分裂出左右兩 node 之 sample，而 stop criterion 則訂為 $N_t > 30$ 或 $\frac{n_{t,i}}{n_t} > 0.1$，$i = l$ or r 或 $\Delta L_t^q > 50$。

同時 node t 中，計算每個問題q所分裂出的兩個 node 其 K-L distance d_t^q：

$$d_t^q = \frac{1}{2}\left(\int f_{t,l}^q(x)\log\left(\frac{f_{t,l}^q(x)}{f_{t,r}^q(x)}\right)dx + \int f_{t,r}^q(x)\log\left(\frac{f_{t,r}^q(x)}{f_{t,l}^q(x)}\right)dx\right) \quad (4.11)$$

其中 $f_{t,l}^q(x)$、$f_{t,r}^q(x)$ 分別為所問問題分裂出的兩個 node 亦為高斯分佈，可將 d_t^q 推導為：

$$d_t^q = (\Sigma_{t,l}^{-1} + \Sigma_{t,r}^{-1})(\mu_{t,l} - \mu_{t,r})^t(\mu_{t,l} - \mu_{t,r}) + \Sigma_{t,l}\Sigma_{t,r}^{-1} + \Sigma_{t,r}\Sigma_{t,l}^{-1} - 2I \quad (4.12)$$

4.3.3 聽覺實驗之進一步驗證

我們相信所訓練出來的 Decision Tree 的確有階層性的特性，越上層的問題越具代表性，越能夠把音節的特性區分開來，如此才能得到離越遠的 leaf node 其特性差距越大之結論。如 A、B、C 三個 sample 其在問題集的答案大都相同，A 與 B、A 與 C 各只有一個問題不同，A 與 B 不同的問題為 ΔL 最大的問題而 A 與 C 不同的問題為 ΔL 最小的問題，若將 A 所在的句子截出，將此句子之 A 替換成 B 或 C，其替換為 B 所帶來的失真將比替換為 C 所帶來的失真來得多，同時此一差異必需和聽覺上的差異是一致的。

為了驗證此一想法，我們選定五個音節依照在 root node 中 ΔL 和 KL distance 由大到小排列出每個問題之重要性，設計以下實驗：

1. 將 root node 中 ΔL 最大的問題視為最具影響力之問題 QA，能夠將音節特性做明顯的區隔，而同時考慮 ΔL 和 KL distance 最小值的問題視為最不具影響力的問題 QB。所選定之問題如下：

表格 4-4 五個測試音節選定之問題

	ta1	ren2	yi3	shi4	de5
QA	Q22	Q27	Q21	Q21	Q27
QB	Q15	Q3	Q16	Q16	Q17

2. 在資料量足夠的情況下，儘量找出只有 QA 和 QB 答案不同，而其他問題都相同的例子(依 ΔL 和 KL distance 由大到小之順序考慮)。
3. QA 和 QB 都為 YES 的群組中，隨機選出三個音節及所對應句子，一個句子做為往後替換之載字句 O，同時另音節替換至載字句 O，得到 A1、A2 兩測試句。
4. QA 為 YES 而 QB 為 NO 的群組中，隨機選出兩個音節，替換至載字句 O，得到 B1、B2 兩測試句。
5. QA 為 NO 而 QB 為 YES 的群組中，隨機選出兩個音節，替換至載字句 O，得到 C1、C2 兩測試句。
6. 將 A1、A2、B1、B2、C1、C2 隨機排列，請 12 位同學先聽過原始句子，再聽六個合成語句，依照合成語句本身流暢程度和原始語音相似度，對六個合成語句做排序，音質差異最小的排為 1，差異最大的排為 6。
7. 將 A1、A2 語句標記為原始群組，B1、B2 標記為 QB 群組，C1、C2 標記為 QA 群組，而在 12 位同學所排序的順序中，第 1、2 句主觀品質標記為甲，第 3、4 句主觀品質標記為乙，第 5、6 句主觀品質標記為丙，計算不同群組在主觀品質標記之百分比。
8. 實驗結果如下：

表格 4-5 合成語句之主觀品質標記

	主觀品質標記 (%)		
	甲	乙	丙
原始群組	57.5	30	12.5
QB 群組	38.3	52.5	9.2
QA 群組	4.2	17.5	78.3

實驗結果證實，同一群組中之音節，彼此替換所帶來的失真及不流暢程度，的確比使用不同群組替換後的句子來的少，而使用 QA 差異之群組，因為其群組特性明顯與原始群組不同，因此主觀品質標記有 78.3%被評為最差，而使用 QB 差異之群組，因其群組特性與原始群組特性差異較小，因此主觀品質標記和原始群組之差距不如 QA 群組來的明顯，但如此結果的確證實，我們使用特徵矩陣建立的 Decision Tree 其差異的確與聽覺上的差異是一致的。

第五章 結論與未來展望

一套以大型語料庫為基礎的中文語音合成系統，常會遇到的困難為合成單元音段切割位置不正確，和合成單元間的連音效應嚴重影響合成語句的品質，因此在本文中，我們先使用一套有系統的方法對音節的切割位置做修改，接著提出以音節為單位建立特徵矩陣的方式，將音節做分群與標記，使用 Decision Tree 得到每個音節中使用不同問題，所能分出群組間不同的特性及距離，證實了在同一群組的情況下或群組距離較近的情況下，所合成的語句能夠有較好的合成音質。

然而，語音辨認的技術已向跨語言的方向進展，因此語音合成也邁向多語言的研究領域，能否發展出一套方法能無關於語言之合成技術，為未來可深入研究的課題。另一方面，現今的語音系統尚無公正客觀的效能評估方式，多半採用平均鑑定分數(Mean Opinion Score, MOS)，或者如本論文所使用的方式所評鑑，如此方式費時費力，為了明確瞭解不同作法對合成系統效能帶來的影響，一套制式且客觀的評估方式將會對研究語音合成有所助益。

參考文獻

[1] The HTK Book (for HTK version 3.2.1)

[2] WaveSurfer Homepage：http://speech.kth.se/wavesurfer/

[3] X.D. Huang, A. Acero, and H.-W. Hon. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development, page 177-189. Prentice Hall, 2001.

[4] N. Campbell and A. Black. Prosody and the selection of source units for concatenative synthesis. In J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, editors, Progress in Speech Synthesis. Springer Verlag, 1995.

[5] A. Hunt and A. Black. Unit selection in a concatenative speech synthesis system using a large speech database. In *ICASSP-96*, volume 1, pages 373–376, Atlanta, Georgia, 1996.

[6] Ljolje, A., J. Hirschberg and J. P. H. van Santen, "Automatic Speech Segmentation for Concatenative Inventory Selection," *Proceedings of ESCA/IEEE Workshop on speech synthesis,* 1994, pp. 93-96.

[7] 王小川，「語音訊號處理」，全華科技圖書，民國九十三年三月。

[8] 陳科旭，「使用右文相關聲韻母模式之國語關鍵詞辨認」，國立交通大學碩士論文，民國八十九年六月。

[9] 謝寶華，「使用前後文相關HMM模型之國語連續語音辨認」，國立交通大學碩士論文，民國九十年六月。

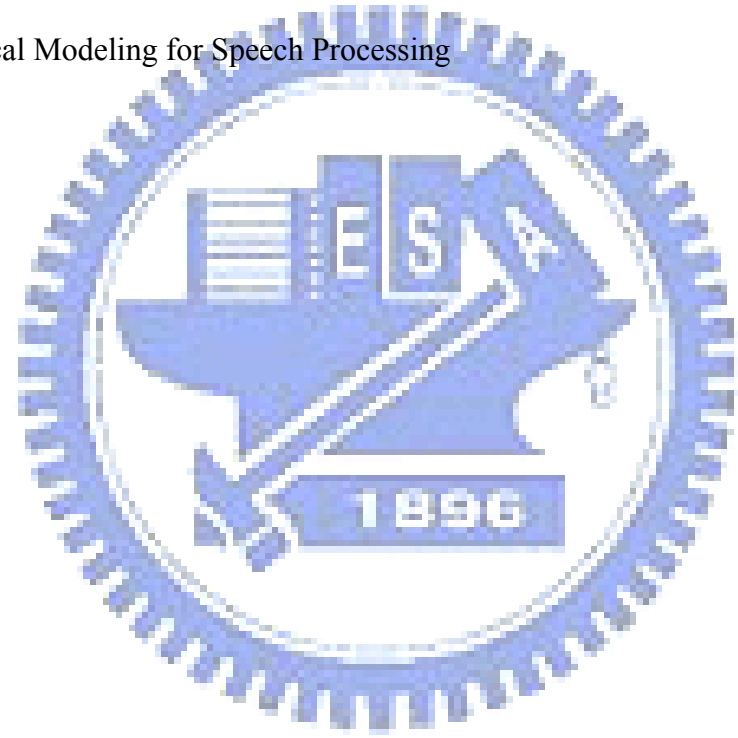
[10] 吳佩穎，「以語料庫為基礎之中文文句翻語音系統中合成單元之選取」，國立交通大學碩士論文，民國九十四年七月。

[11] 洪國興，「以語料庫為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。

[12] Y. Zhao, L.J. Wang, M. Chu, F.K. Soong, and Z.G. Cao, “Refining phoneme segmentations using speaker-adaptive context dependent boundary models,” Proc. EUROSPEECH-2005, Lisbon, Portugal,Sept. 2005.

[13] A. Sethy and S. Narayanam, “Refined speech segmentation for concatenative speech synthesis,” Proc. ICSLP-2002, pp.145–148, Denver,CO, Sept. 2002.

[14] Lijuan WANG, Yong ZHAO, Min CHU, Frank K. SOONG, Jianlai ZHOU, *Nonmembers*, *and* Zhigang CAO, *Member* “Context-Dependent Boundary Model for Refining Boundaries Segmentation of TTS Units” Special Section on Statistical Modeling for Speech Processing



附錄一

附錄表 1：MFCC 特徵矩陣之類別 Confusion Matrix

Confusion Matrix								Confusion Matrix							
de5		類 1	類 2	類 3	類 4	類 5	total	you3		類 1	類 2	類 3	類 4	類 5	total
	類 1	53	6.9	26.2	4	9.9	202		類 1	33	34	6	14	13	100
	類 2	21.6	31.1	27.5	9.6	10.2	167		類 2	8.9	62.6	7.3	2.4	18.7	123
	類 3	27.2	8.2	48.6	3.9	12.1	257		類 3	11.4	18.6	44.3	7.1	18.6	70
	類 4	11.3	6.8	23.7	50.3	7.9	177		類 4	12	22.7	1.3	40	24	75
	類 5	14.5	9.8	20.7	2.6	52.3	193		類 5	9.1	14	7.4	9.9	59.5	121
yi4		類 1	類 2	類 3	類 4	類 5	total	zai4		類 1	類 2	類 3	類 4	類 5	total
	類 1	42.3	12.6	16.2	7.2	21.6	111		類 1	0	13.7	35.3	17.6	33.3	51
	類 2	14.3	43.9	16.3	2	23.5	98		類 2	0	28.1	33.3	5.3	33.3	57
	類 3	10.1	17.3	36.7	16.5	19.4	139		類 3	0	5.6	58	3.7	32.7	162
	類 4	11.6	11.6	14.7	53.7	8.4	95		類 4	0	7	5.6	32.4	54.9	71
	類 5	8.9	7.2	10	5.6	68.3	180		類 5	0	3.5	15.6	8.7	72.3	173
bu4		類 1	類 2	類 3	類 4	類 5	total	wo3		類 1	類 2	類 3	類 4	類 5	total
	類 1	32.7	14.5	0	41.8	10.9	110		類 1	25	25	15.6	0	34.4	32
	類 2	15.9	51.2	0	26.8	6.1	82		類 2	26.7	33.3	0	0	40	30
	類 3	19	9.5	52.4	6	13.1	84		類 3	8.3	0	45.2	7.1	39.3	84
	類 4	9.6	4.2	1.2	75.9	9	166		類 4	2.1	2.1	43.8	41.7	10.4	48
	類 5	14.8	6.8	2.3	44.3	31.8	88		類 5	3.4	1.3	9.4	2	83.9	149
ren2		類 1	類 2	類 3	類 4	類 5	total	yi3		類 1	類 2	類 3	類 4	類 5	total
	類 1	50	27.5	15	7.5	0	80		類 1	40	0	22.9	34.3	2.9	35
	類 2	4.2	62	21.8	0	12	142		類 2	6	31.3	23.9	25.4	13.4	67
	類 3	18.1	14.7	67.2	0	0	116		類 3	6.1	6.1	58.5	22	7.3	82
	類 4	9.1	18.2	13.6	59.1	0	66		類 4	4	6.5	10.5	68.5	10.5	124
	類 5	4.8	21.4	17.3	0.6	56	168		類 5	8.8	6.2	7.1	30.1	47.8	113
shi4		類 1	類 2	類 3	類 4	類 5	total	ta1		類 1	類 2	類 3	類 4	類 5	total
	類 1	59.3	20.4	12.3	8	0	162		類 1	32.8	19.7	4.9	31.1	11.5	61
	類 2	17.1	54.5	18.7	9.8	0	246		類 2	1.4	34.2	4.1	46.6	13.7	73
	類 3	1.9	14.3	79.5	4.2	0	259		類 3	1.1	5.3	62.8	13.8	17	94
	類 4	14.2	44.2	11.7	30	0	120		類 4	2.8	16	2.8	72.6	5.7	106
	類 5	21.8	25.7	50.5	2	0	101		類 5	1.8	10.5	5.3	21.1	61.4	114

附錄表 2：Energy 特徵矩陣之類別 Confusion Matrix

Confusion Matrix							
de5		類 1	類 2	類 3	類 4	類 5	total
	類 1	42.4	20.7	10.3	24.5	2.2	184
	類 2	5.1	45.2	13.4	26.3	10.1	217
	類 3	3.8	6.5	64.3	2.7	22.7	185
	類 4	13.8	23.3	5.2	49.6	8.2	232
	類 5	0.6	5.6	46.1	1.1	46.6	178

Confusion Matrix							
you3		類 1	類 2	類 3	類 4	類 5	total
	類 1	48.4	25.8	21	4.8	0	62
	類 2	16.9	66.2	11.7	5.2	0	77
	類 3	0	0	90.3	0	9.7	206
	類 4	3	36.4	0	60.6	0	33
	類 5	0	0	54.1	0.9	45	111

yi4		類 1	類 2	類 3	類 4	類 5	total
	類 1	62.5	3.6	24.1	4.5	5.4	112
	類 2	10.2	69.5	7.6	11.9	0.8	118
	類 3	13.2	1.9	64.8	7.5	12.6	159
	類 4	1	14	2	79	4	100
	類 5	0.7	3	22.4	13.4	60.4	134

zai4		類 1	類 2	類 3	類 4	類 5	total
	類 1	65.2	0	1.5	28.8	4.5	66
	類 2	4.7	57.8	32.8	0	4.7	64
	類 3	0.7	15.2	76.2	0	7.9	151
	類 4	19	0	0.8	66.1	14	121
	類 5	0	8.9	39.3	1.8	50	112

bu4		類 1	類 2	類 3	類 4	類 5	total
	類 1	73.3	5.2	10.3	10.3	0.9	116
	類 2	43.8	51	3.1	2.1	0	96
	類 3	21.6	1.7	68.1	8.6	0	116
	類 4	20.4	8	9.7	52.2	9.7	113
	類 5	12.4	3.4	23.6	20.2	40.4	89

wo3		類 1	類 2	類 3	類 4	類 5	total
	類 1	51.4	14.3	5.7	28.6	0	35
	類 2	5	23.3	10	61.7	0	60
	類 3	6.7	1.7	75	0	16.7	60
	類 4	13.6	5.8	0	80.6	0	103
	類 5	0	0	36.5	7.1	56.5	85

ren2		類 1	類 2	類 3	類 4	類 5	total
	類 1	70	1.1	3.3	25.6	0	90
	類 2	0	61.8	25.2	7.6	5.3	131
	類 3	0.6	28.7	60	6.9	3.8	160
	類 4	16.2	11.7	9.7	62.3	0	154
	類 5	0	2.7	5.4	2.7	89.2	37

yi3		類 1	類 2	類 3	類 4	類 5	total
	類 1	56.3	31	4.2	8.5	0	71
	類 2	22.1	55.8	0	22.1	0	86
	類 3	4.5	0	80	0	15.5	110
	類 4	12.7	7.9	6.3	73	0	63
	類 5	12.1	1.1	35.2	0	51.6	91

shi4		類 1	類 2	類 3	類 4	類 5	total
	類 1	54.8	0	2.4	40.5	2.4	42
	類 2	2.6	19	16.4	0	62.1	116
	類 3	0	1.5	67.1	0	31.5	343
	類 4	13.1	0	2.3	83.1	1.5	130
	類 5	0.4	1.9	17.5	0	80.2	257

ta1		類 1	類 2	類 3	類 4	類 5	total
	類 1	44.3	34.3	15.7	1.4	4.3	70
	類 2	17.5	51.5	21.6	7.2	2.1	97
	類 3	6.8	3.4	80.7	6.8	2.3	88
	類 4	0	2.2	4.4	65.9	27.5	91
	類 5	2.9	1	3.9	25.5	66.7	102

附錄表 3：Duration 特徵矩陣之類別 Confusion Matrix

Confusion Matrix								Confusion Matrix							
de5		類 1	類 2	類 3	類 4	類 5	total	you3		類 1	類 2	類 3	類 4	類 5	total
	類 1	95.9	1.7	0	2.4	0	536		類 1	92.3	0	0	7.7	0	26
	類 2	76.5	10.1	0	12.3	1.1	277		類 2	37.9	0	0	62.1	0	29
	類 3	0	12.1	63.6	21.2	3	33		類 3	0	0	100	0	0	360
	類 4	46.2	10	0	40.8	3.1	130		類 4	0	0	15.3	84.7	0	59
	類 5	0	0	25	0	75	20		類 5	0	0	100	0	0	15
yi4		類 1	類 2	類 3	類 4	類 5	total	zai4		類 1	類 2	類 3	類 4	類 5	total
	類 1	92.3	7.7	0	0	0	13		類 1	74	25.4	0.6	0	0	181
	類 2	34.6	46.2	0	19.2	0	26		類 2	24	72.7	3.3	0	0	150
	類 3	0	0	97	0	3	67		類 3	0	3.2	83.3	0	13.5	126
	類 4	2.2	20	0	44.4	33.3	45		類 4	0	93.3	6.7	0	0	15
	類 5	0	0	1.7	0.4	97.9	472		類 5	0	0	4.8	0	95.2	42
bu4		類 1	類 2	類 3	類 4	類 5	total	wo3		類 1	類 2	類 3	類 4	類 5	total
	類 1	100	0	0	0	0	16		類 1	0	77.8	0	22.2	0	9
	類 2	46.7	0	0	53.3	0	30		類 2	0	71.1	2.2	26.7	0	45
	類 3	0	0	100	0	0	36		類 3	0	0	88.4	0	11.6	69
	類 4	0	0	0	64.5	35.5	62		類 4	0	35.3	23.5	41.2	0	51
	類 5	0	0	2.8	0	97.2	386		類 5	0	0	0.6	0	99.4	169
ren2		類 1	類 2	類 3	類 4	類 5	total	yi3		類 1	類 2	類 3	類 4	類 5	total
	類 1	96.8	1.8	0	1.4	0	281		類 1	0	100	0	0	0	16
	類 2	57.9	30.3	4.8	3.4	3.4	145		類 2	0	64.1	0	35.9	0	39
	類 3	0	7.4	82.1	0	10.5	95		類 3	0	0	93.5	0	6.5	31
	類 4	0	0	0	100	0	17		類 4	0	2.8	0	76.9	20.4	108
	類 5	0	0	5.9	0	94.1	34		類 5	0	0	0.9	0	99.1	227
shi4		類 1	類 2	類 3	類 4	類 5	total	ta1		類 1	類 2	類 3	類 4	類 5	total
	類 1	64.1	30.8	1.3	3.8	0	156		類 1	0	100	0	0	0	29
	類 2	6.1	91.3	0	2.6	0	459		類 2	0	99.2	0	0.8	0	120
	類 3	0.9	0	84.4	0	14.7	109		類 3	0	0	100	0	0	19
	類 4	0	1.1	0	98.9	0	93		類 4	0	32.9	0	56.2	11	73
	類 5	0	0	15.5	0	84.5	71		類 5	0	1	6.3	1	91.8	207

附錄二

附錄表 4：Root node 中各問題所得之 delta likelihood (ΔL^q)

	de5	yi4	bu4	ren2	shi4	you3	zai4	wo3	yi3	ta1
Q1	--	--	155.3	--	--	229.6	168.8	181.2	127.5	239.8
Q2	--	--	--	200.4	239.5	--	--	--	--	--
Q3	--	20.59	36.61	34.89	125.7	45.67	61.82	137.8	106.3	137.8
Q4	--	140.4	84.72	163	144.2	99.85	0	151.9	100.2	185.1
Q5	--	94.42	44.36	--	--	--	--	--	--	--
Q6	--	201.1	166.8	88.36	121.6	125.1	56.58	--	133.7	--
Q7	478.1	105	--	137.2	144.5	100.6	117	--	113.7	142
Q8	229.3	84.99	59.33	62.54	27.89	116.9	60.52	104.8	37	105.8
Q9	448.6	111.7	84.51	141	78.24	68.5	65.91	32.99	67.69	98
Q10	410.6	--	211.6	171.4	--	136.6	89.02	197.5	57.57	148.6
Q11	--	--	--	--	--	--	--	--	--	--
Q12	--	47.66	73.61	--	--	--	43.29	--	--	--
Q13	156.2	95.28	129.7	52.67	91.4	43.81	40.13	53.31	100	58.51
Q14	233.2	75.55	27.37	57.01	57.65	60.34	52	46.03	29.7	35.83
Q15	166.5	90.86	55.73	52.81	110.8	81.11	94.48	49.27	106.5	61.7
Q16	107.6	37.17	29.32	38.23	45.4	21.8	15.14	21.46	49.19	52.29
Q17	66.52	60.49	34.23	43.59	62.82	79.91	71.02	42.46	72.27	42.65
Q18	--	--	--	--	--	38.49	33.22	39.82	--	--
Q19	--	265.6	--	158.8	--	297.7	--	101.6	252.7	--
Q20	--	278.2	158.3	--	380.9	320.8	302.8	205.9	260.1	304.9
Q21	--	315.9	182.2	--	633.3	360.9	322.9	280.6	296.5	342
Q22	--	301	265.6	--	517	324.6	328.6	264.3	257	350.8
Q23	--	--	--	--	--	--	94.35	112.8	82.08	144.4
Q24	323.6	206	180.5	217.6	208.3	114.6	101.2	116.6	91.99	158.1
Q25	321.3	300.2	180.2	276.5	367.3	75.77	86.88	86.51	121.3	109.1
Q26	479.5	370.8	--	385.8	518.7	91.8	95.33	110.3	134.9	--
Q27	--	301.5	--	280.8	415.2	--	--	--	--	--
Q28	--	--	--	--	--	--	--	--	--	--

-- 表示此一問題不符合 split criterion

附錄表 5：Root node 中各問題所得之 KL distance (d^q)

	de5	yi4	bu4	ren2	shi4	you3	zai4	wo3	yi3	ta1
Q1	--	--	101.7	--	--	129.4	46.01	64.58	58.34	49
Q2	--	--	--	108.4	127.9	--	--	--	--	--
Q3	--	8.54	11.76	8.56	8.38	9.25	26.99	85.18	20.43	38.87
Q4	--	18.74	17.47	35.62	26.84	26.07	0	97.19	32.22	61.18
Q5	--	27.01	27.39	--	--	--	--	--	--	--
Q6	--	39.64	46.98	17.53	22.23	37.24	46.23	0	46.93	0
Q7	31.36	35.55	0	37.08	15.81	27.48	29.05	0	41.31	67.02
Q8	26.03	29.17	34.67	26.02	13.89	33.72	17.49	50.43	41.26	28.08
Q9	17.39	17.89	26.12	25.88	9.98	21.41	14.74	36.85	18.28	32.45
Q10	24.52	0	39.6	35.76	0	34.83	32.86	88.55	81.65	31.33
Q11	--	--	--	--	--	--	--	--	--	--
Q12	--	37.85	48.84	--	--	--	47.33	--	--	--
Q13	21.81	27.94	38.69	42.01	13.66	34.87	36.6	80.43	38.61	36.28
Q14	5.61	10.24	6.78	9.64	6.97	16.49	9.28	46.62	15.59	22.54
Q15	3.98	9.21	10.6	8.7	7.5	17.65	12.22	20.87	18.79	13.55
Q16	4.69	7.33	7.46	9.11	5.74	16.08	12.2	38.63	10.77	17.07
Q17	2.96	9.66	9.75	10.7	6.74	19.94	12.65	21.25	20.12	14.51
Q18	--	--	--	--	--	48.03	32.12	37.81	--	--
Q19	--	32.08	--	44.67	--	64.61	--	67.93	60.22	--
Q20	--	34.95	28.73	--	38.33	75.66	35.49	79.53	54.06	54.49
Q21	--	41.42	28.68	--	61.28	100.3	37.4	99.84	62.48	59.44
Q22	--	70.97	55.98	--	75.47	121.2	45.02	87.89	61.81	57.27
Q23	--	--	--	--	--	--	79.47	59.35	107.4	49.3
Q24	34.02	32.7	36.82	41.42	27.61	25.95	32.19	47.74	36.8	41.16
Q25	14.44	49.97	65.18	36.65	29.82	15.55	12.56	38.99	22.65	22.22
Q26	27.19	75.26	--	64.16	50.14	31.86	40.48	258.6	37.37	--
Q27	--	163	--	102.9	132.4	--	--	--	--	--
Q28	--	--	--	--	--	--	--	--	--	--

-- 表示此一問題不符合 split criterion

附錄三

語音合成品質問卷

以下一共有五題，每題分別有七個句子，第一個句子為原始語音，後六句(A、B、C、D、E、F)皆在某一特定音節被另一相同音節替換，請依照和原始語音的相似度，及合成語句本身流暢程度對六個合成語句做排序，1~6 分別表示由優至劣之排序。

- 第一題‘ㄊㄚ’　原始　A　B　C　D　E　F

  排序：　1　2　3　4　5　6

- 第二題‘ㄖㄣˊ’　原始　A　B　C　D　E　F

  排序：　1　2　3　4　5　6

- 第三題‘一ˇ’　原始　A　B　C　D　E　F

  排序：　1　2　3　4　5　6

- 第四題‘ㄕㄟˋ’　原始　A　B　C　D　E　F

  排序：　1　2　3　4　5　6

- 第五題‘ㄉㄜ˙’　原始　A　B　C　D　E　F

  排序：　1　2　3　4　5　6

附錄表 6：12 位同學語音合成品質問卷之統計結果

ta1	主觀品質標記 (%)		
	甲	乙	丙
原始群組	14	7	3
QB 群組	10	13	1
QA 群組	0	4	20

ren2	主觀品質標記 (%)		
	甲	乙	丙
原始群組	13	7	4
QB 群組	8	15	1
QA 群組	3	2	19

yi3	主觀品質標記 (%)		
	甲	乙	丙
原始群組	11	8	5
QB 群組	12	10	2
QA 群組	1	6	17

shi4	主觀品質標記 (%)		
	甲	乙	丙
原始群組	13	9	2
QB 群組	10	9	5
QA 群組	1	6	17

de5	主觀品質標記 (%)		
	甲	乙	丙
原始群組	18	5	1
QB 群組	6	16	2
QA 群組	0	3	21