國　立　交　通　大　學

電信工程學系碩士班

碩士論文

中文 TTS 系統語音合成之改進

An Improvement of Speech Synthesis for Mandarin TTS System

研 究 生： 林立峰

指導教授： 陳信宏　博士

中華民國九十三年七月

中文 TTS 系統語音合成之改進

An Improvement of Speech Synthesis for Mandarin TTS System

研 究 生：林立峰　　Student : Li-Feng Lin

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

國立交通大學

A Thesis
Submitted to Department of Communication Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the Degree of
Master of Science
In
Electrical Engineering
July 2004
Hsinchu, Taiwan, Republic of China

中華民國九十三年七月

中文 TTS 系統之語音合成改進

研究生：林立峰　　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

以語料庫為基礎（Corpus-based）的中文文句轉語音系統是現今中文語音合成的主流，在實作這樣的系統前，首先必須對一個大量音節的語料庫做切割，本論文嘗試建立一套處理大量語料庫的標準流程，內容包括語音自動切割與修正、基週軌跡偵測與修正、不良發音偵測，以作為未來發展中文 Corpus-based TTS 系統的基礎。另外，我們針對國立交通大學電信研究所過去所發展的一個中文文句轉語音系統的合成器和樣本音節資料庫作改進，在合成器的部分，對於音節相連時產生能量及基頻不連續的現象做處理，並且改變合成摩擦類子音的方式；在樣本音節的選取上，將原有每一音節使用一個各讀音樣本，置換成由連續語音語料庫中抽取的三種不同長度的樣本，以降低 TD-PSOLA 合成器因音長改變太大所引發聲音品質不佳的效應，經過以上的改良後，合成出的語音較為自然順暢。

關鍵詞：中文文句轉語音系統、語音自動切割、基週軌跡偵測、樣本音節資料庫

An Improvement of Speech Synthesis for Mandarin TTS System

Student：Li-feng Lin Advisor：Dr. Sin-Horng Chen

Department of Communication Engineering
National Chiao Tung University

ABSTRACT

In this thesis, a standard pre-processing procedure is established for the development of corpus-based Mandarin text-to-speech (TTS) systems. It includes automatic speech segmentation, syllable pitch contour detection, and pronunciation error detection. Besides, some improvements of the existing Mandarin TTS system, developed previously in National Chiao Tung University, are discussed. Firstly, a new acoustic inventory is constructed. Three waveform templates with different durations for each base-syllable are extracted from a large continuous-speech database to replace the single isolated waveform template which is pronounced too long in length. The degradation in syllable-duration modification of TD-PSOLA synthesizer caused by too large compression or stretching is therefore greatly reduced. Secondly, a processing to eliminate the energy and pitch discontinuities in syllable waveform concatenation is done. Lastly, the method to synthesize fricative sounds is changed from re-sampling to overlap-add. Experimental results showed that the quality of the synthesized speech was greatly improved.

Keywords: Mandarin text-to-speech, automatic speech segmentation, pitch contour detection, acoustic inventory

誌謝

這二年來，感謝陳信宏老師帶我進入了語音合成的世界裡，並且教導我做研究的態度和方法，同時也要謝謝王逸如老師，常常指正我報告時邏輯上的錯誤；謝謝廖元甫老師提供了許多很棒的主意，郭威志與傅振宏學長在口試前的鼎力相助，沒有你們，這篇論文不可能完成。

當然，還得感謝實驗室的同學們：謝謝阿德和俊良大力的協助，你們就好像我的 HTK BOOK 一樣，有問必答；矮小可愛的樹哥，當我們開你玩笑時都不生氣；祺翰的熱心幫助，許多事情多虧有你在；嘉俊提供睡覺的地方，讓我不至於流落街頭；能早日找到小太陽的諺哥，你的球技是舒解壓力的好伙伴；六年來還是這麼愛搞笑的性獸，有你在的地方總是笑聲不斷…… 謝謝你們在這些日子來所帶來的歡笑，讓實驗室的生活一點也不無聊。

謝謝國興和希群，在最後一個月被我指使來指使去地，幫助我完成語料庫的準備工作（國興你二個小時切四個音檔的紀錄我會永遠記得），也期待你們能將我未完成的部分完成。

最後，我要特別感謝猴兒和可愛的家人們，在求學階段裡，沒有你們的鼓勵與支持，是很難如此順利完成。

目錄

表目錄

圖目錄

第一章 緒論

1.1 研究動機

近幾年來，隨著科技的快速發展，人類愈來愈仰賴電腦來處理身邊各項的事務，因此人與電腦間的溝通就顯得格外重要。

觀察人類最自然的溝通方式，不外乎聽與說：聽出正確的訊息（辨識），說出要表達的話（合成），為了讓這二種方式，也能成為人類與機器間的構通模式，語音辨識和語音合成技術的研究與發展扮演了舉足輕重的地位，而本論文主要著重在合成流利語音的部分。

語音合成主要的功能便是將儲存在電腦中的文字內容轉換成語音輸出，因此我們可以稱語音合成系統為文句轉語音（Text to Speech, TTS）系統。這樣的系統可應用於許多方面，如有聲電子書、電話查尋系統等，值得一提的是，當合成系統與網路結合時，可將網路上的資訊轉換成語音，提供殘障者一個獲取廣泛資訊的來源。

1.2 研究方向

實作一套語音合成系統前，須先進行語料庫的音節切割工作，在以往語音資料量少的情況下，我們以人工切割的方式，進行音節切割的動作；但隨著儲存設備的快速發展，以大量語料庫為基礎的語音合成系統（Corpus-based TTS system），為現今的主流，為了實現這樣的系統，發展一套自動化音節切割工具變得極為重要；目前自動切割的動作，大都採用語音辨認技術中 HMM（Hidden

Markov Model）辨認器來進行語料庫的切割動作，但這樣的切割動作受限於原始 Model 的好壞，使得切割位置有所偏移，因此本論文提出一套方法，藉由語音的聲學特性，適當的調整由辨認器切割出的位置。

對於交通大學語音實驗室過去所發展的合成系統而言，語音資料庫是以中文音節(Syllable)為單元，共儲存了 411 個中文基本音節，我們稱之為樣本音節；由於樣本音節是以單音錄製，因此每個音節的長度都很長，這對於使用時域基頻同步疊加（TD-POSOLA）演算法的合成器來說，是難以合成出自然流利的語音，因此我們試著從連續語料庫中，對每個中文音節取出具有長、中、短三種不同長度的音節作為樣本音節，再加上改進舊有語音合成器合成出的聲音在音節相接處不連續的問題，期待能改善合成的品質。

1.3 章節概要說明

本論文共分為四章，各章概要說明如下：

第一章 緒論：說明本篇論文的研究動機、研究方向及章節概要。

第二章 語音資料庫的前處理過程：介紹處理大量語音資料庫的流程，以求得正確的語音韻律參數；過程中包含語音自動切割與修正、基頻軌跡的求取和調整。經過這樣的步驟後，方可掌握整個語料庫，語音的韻律變化。

第三章 語音合成技術實作與改進：以實作的方式，找出以往語音合成器的缺點後，提出改進的方法；並且將舊有的樣本音節，置換為來自自然語音的音節，使得合成出的語音品質有所提升。

第四章 結論與未來展望：對本篇論文所提出的方法下結論，並說明未來改進的方向。

第二章 語音資料庫的前處理過程

TTS 系統要能合成出自然流利的語音，關鍵在於語音的韻律變化是否自然順暢，尤其當合成單元相連接時所遇到韻律不連續現象，更是破壞整體合成品質的主要因素【9】，若能減少合成語句中合成單元相接的次數，對於提升合成的品質將有極大的幫助，為了達成這個目的，以大量語料庫為基礎（Corpus-based）的 TTS 系統成為近年來最熱門的話題；這樣的系統與過去我們所發展的合成系統最大的差別，即在於語音資料庫的大小。

對於資料量極少的語料庫而言，韻律參數的求取可以由人工求得，當考慮龐大的語料庫時，同樣方法雖然能求得精確數值卻非常耗時，為了加快處理的速度，我們建立了一套半自動的處理流程，用來降低人工處理所耗費的時間和成本；這樣的流程包含了對語料庫作語音資料的切割（Segment）及求取語音資料的基頻軌跡參數，經過這樣的處理後，才能求得這音正確的韻律變化，這些語料方可用於 TTS 語音資料庫中。

在接下來的小節中，我們將對目前所處理的語料做簡單的說明，並且詳細地介紹整個處理流程。

2.1 語料庫說明：

語料庫文字稿（Text）的部分來自於中央研究院中文句結構樹資料庫 1.1 版（Sinica Treebank Version 1.1），為中央研究院詞庫小組從中央研究院平衡語料庫（Sinica Corpus）中，抽取句子，經由電腦剖析成結構樹，並加以人工修正、檢驗後的所得的成果，詳細的內容請參考【14】。聲音的部分為女性語者，

依照文字稿所念出來的。發音速度為每秒約 4.6 個音節，共有 52192 個音節，錄音時取樣率為 20kHz。

2.2 切割資訊的求取與修正：

由於語料庫資料量過大，若用人工切割音（phone）的位置，需花費太多的時間，因此我們使用以 HMM（Hidden Markov Model）為核心的辨認器，先做初步強制切割（Forced alignment）的工作，frame rate 設定為 10ms，如圖 二-1，使用的原始模型由 TCC300 語料庫訓練而成，圖 二-2 為自動切割的結果。

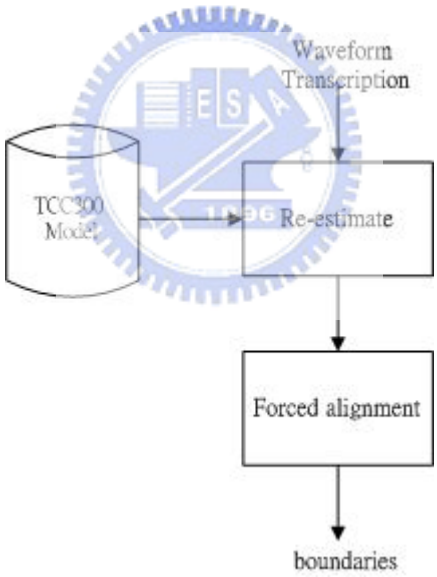


圖 二-1：Forced alignment 切割流程圖

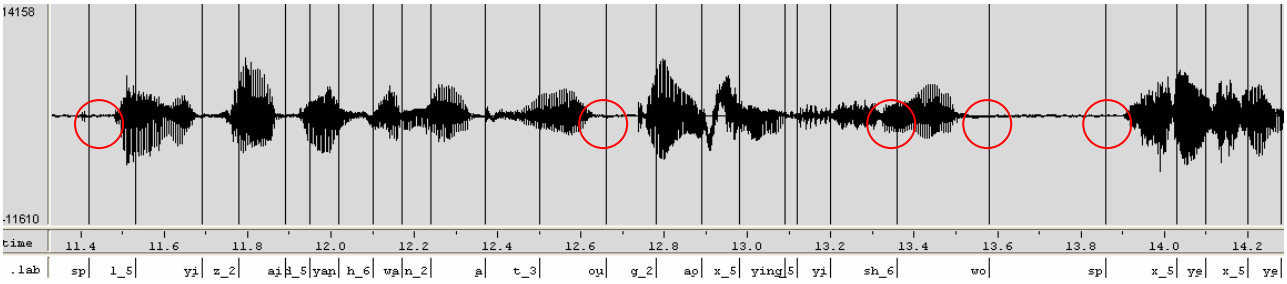


圖 二-2：HMM 自動切割結果（圈圈處為切割有偏差之處）

觀察自動切割的結果發現有二個主要問題存在，其一為音節 boundary 有偏差的現象；其二為中文音節的子音、母音交界點位置不正確。這對合成時樣本音節的挑選及合成音質造成極大的影響，因此我們試著在自動切割後加上半自動調整切割位置的動作，希望將切割位置修改至正確的邊界。

2.2.1 調整切割位置：

我們將以二種不同的調整方式，利用能量（energy）、過零率（zero crossing rate，ZCR）和 Voicing probability【5】參數，分別對音節間切割位置及音節裡子音母音交界的位置作修正，詳細的步驟如下：

2.2.1.1 求取門檻值

切割位置的修正常使用到背景雜訊的能量來決定音節結束或起始位置，本節中我們稱此能量為門檻值；門檻值的求法為：以長度大於 200ms 以上的 silence 能量作為樣本，求出樣本的平均值（Mean）和標準差（Standard deviation）後，將平均值加上兩倍的標準差所得到結果。由於語料庫中，每個音檔錄製的時間和環境不一樣，因此我們需對不同的音檔分別求取門檻值。

2.2.1.2 音節間切割位置的修正

以合成單元選取為考量因素，利用 energy、ZCR 參數，以音節與音節間 silence 的長度為主要分類，再配合子音類型的不同，調整切割位置，分類說明如下：

1. 音節間 silence 長度大於 200ms：

對於 silence 長度足夠的情形，通常表示句子已告一個段落，語者在音節結束後一小段期間內並無發音，故 silence 的能量降至與錄音的背景雜訊相當的等級，直到下個音節即將發聲之際，會有吸氣聲造成能量的上升。因此若要判斷音節結束點，只須找出能量低於背景雜訊的點，在此，我們將第一個遇到低於門檻值的點判斷為音節結束點。

接下來考量下一音節起始點：因為當吸足氣發音後，音節開始的能量和 silence 間會有一較大的能量差，因此可用尋找能量變化最大點為方法，判斷音節起始點位置，在這裡，我們挑選能量差最大的二點，當中第一點的位置作為音節起始點，圖 二-3 為這一類型的例子，實線為原始切割位置，虛線為修正後的切割位置。

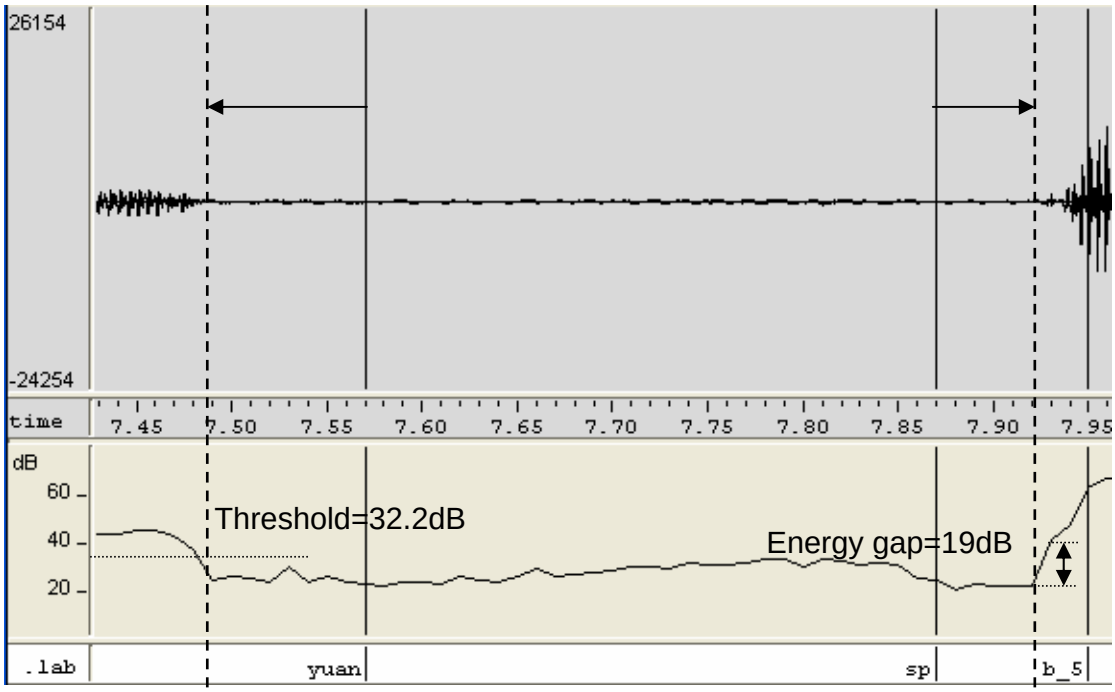


圖 二-3：silence 長度大於 200ms 時音節 boundary 的調整

2. 音節間 silence 長度大於等於 10ms 但小於 200ms：

這種長度的 silence，有一定程度的能量存在，針對此特性在尋找音節結束點時，先嘗試尋找是否有能量小於門檻值的點，若有則可使用上述的方式；若無我們將觀察範圍內能量最小值（Local Minimum）的位置視為最佳結束點，如圖 二-4 所示。

至於後一音節起始點，我們以音節的子音，分作二類來處理，第一類為具有較高 ZCR 的摩擦子音，其它則屬於第二類。對於第一類的子音，我們以 ZCR 變化最大的點當作起始點，如圖 2-4 所示，第二類則延用能量變化最大的點做為音節起始點。

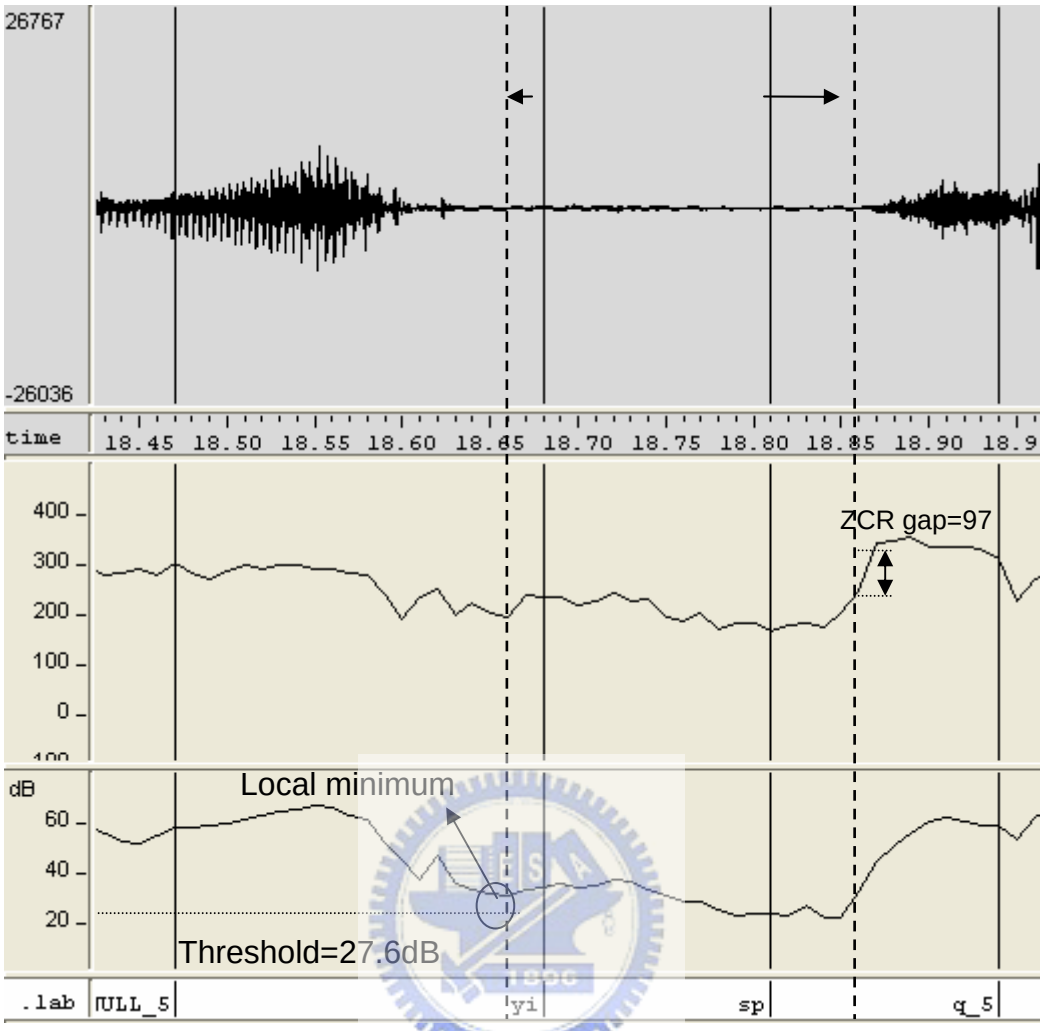


圖 二-4：silence 長度大於 20ms 小於 200ms 時音節 boundary 的調整

3. 音節間無 silence：

對於此類的處理，需將狀況分為四類：

a. 若 pitch contour 相連，energy 沒有降至與門檻值相當時則不修正音節切割位置，如圖 二-5 所示，並輸出分類標記，告知此音節在合成時須與跟下一音節一起使用。

b. 若 pitch contour 相連，energy 降至門檻值，將此音節結束位置與下個音節起始位置調整至最低能量點，如圖 二-6 所示。並輸出分類標記，告知此音節在合成時，可單獨使用。

c. 若 pitch contour 不連續，energy 沒有降至與門檻值相當，

則利用下一個音節子音分類的方式修正切割位置;若子音為摩擦音類，可利用 ZCR 變化量最大的點，求出起始和結束位置，如圖 二-7 所示；若子音為其它分類，則以能量最低點當作起始位置與結束位置。當調整結束後，輸出分類標記，告知在合成時，此音節雖可用於單音合成，但頭尾能量在連接時要特別留意。

d. 若 pitch contour 不連續，energy 降至門檻值，將此音節結束位置調整至最低能量點，而下一個音節起始位置的判斷方式與 c 大致相同，差別只在子音不為摩擦音類時，是以能量變化最大的點做為音節起始點。經過這樣的步驟，可能會在原本無 silence 的音節間產生短暫停（short pause），如圖 二-8 所示。最後輸出一標記，告知此音節在合成時，可單獨使用。

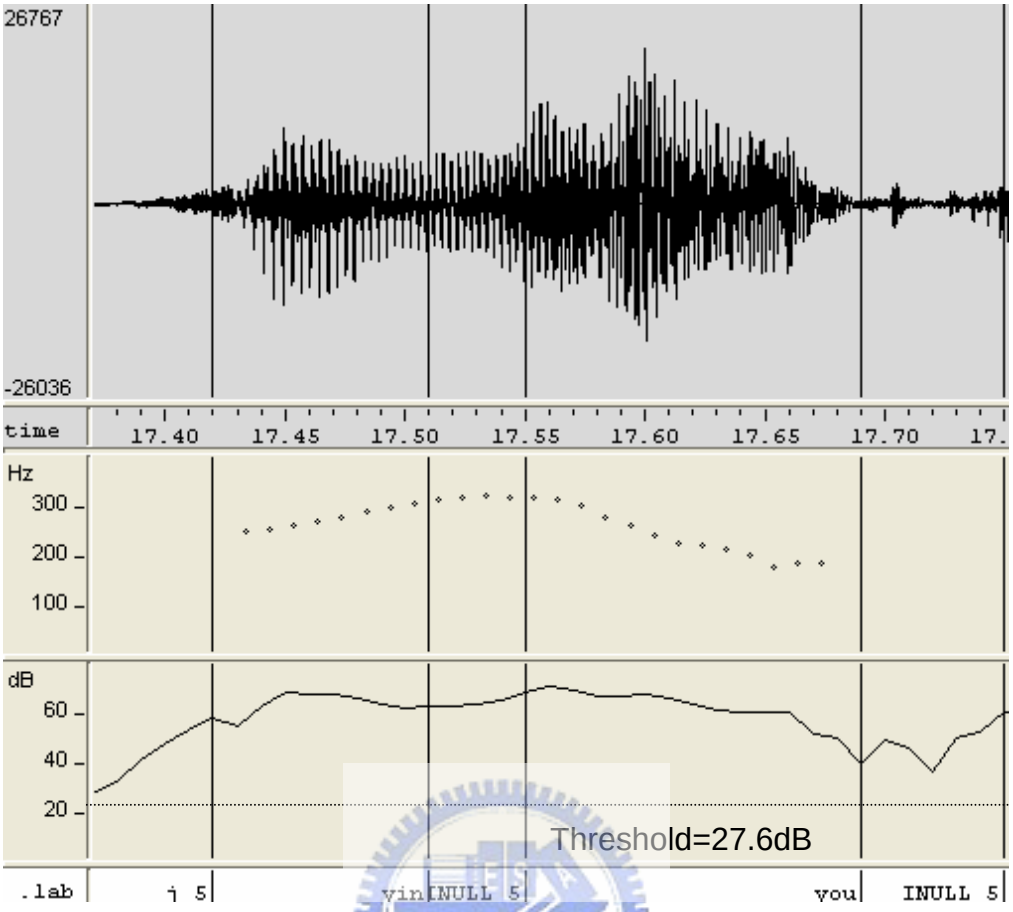


圖 二-5：音節間無 silence 且 pitch contour 連續、energy 大於門檻值時不調整音節切割位置

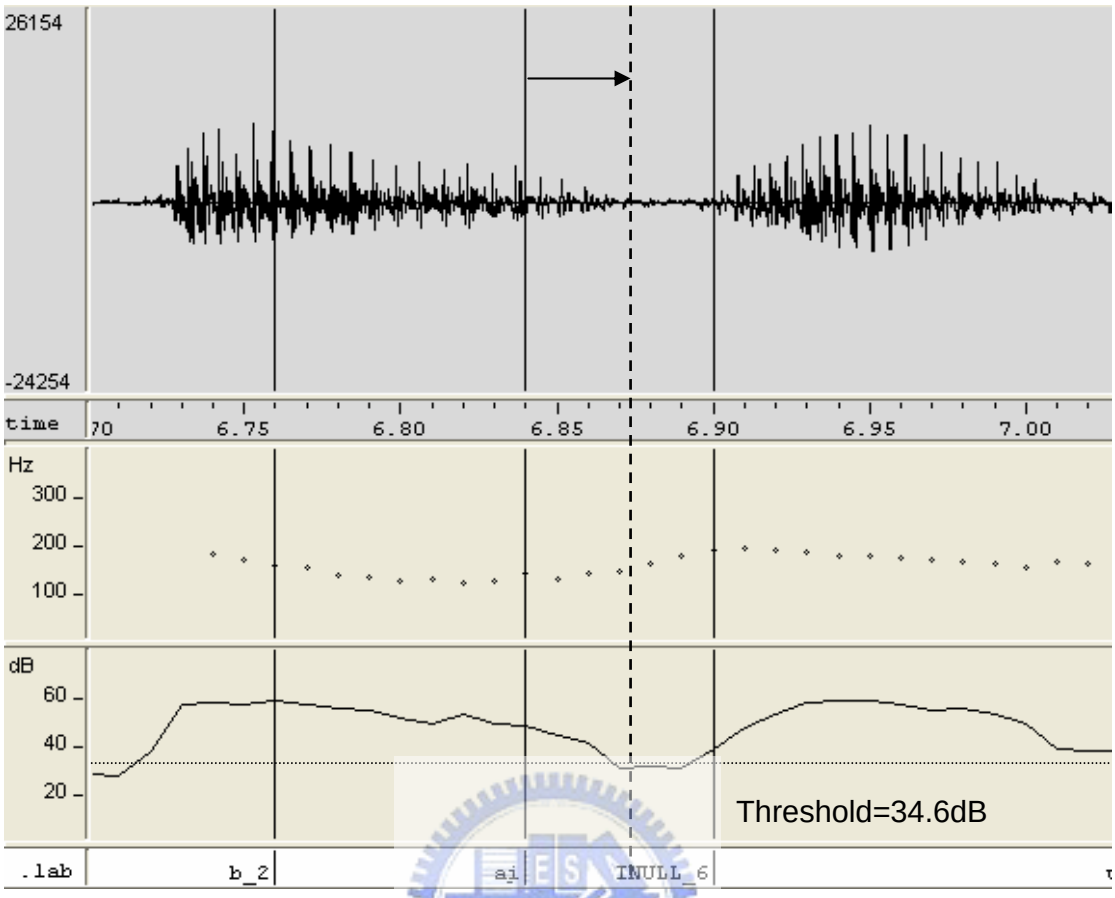


圖 二-6：音節間無 silence 且 pitch contour 連續、energy 小於門檻值時以能量最小值做為調整 boundary 依據

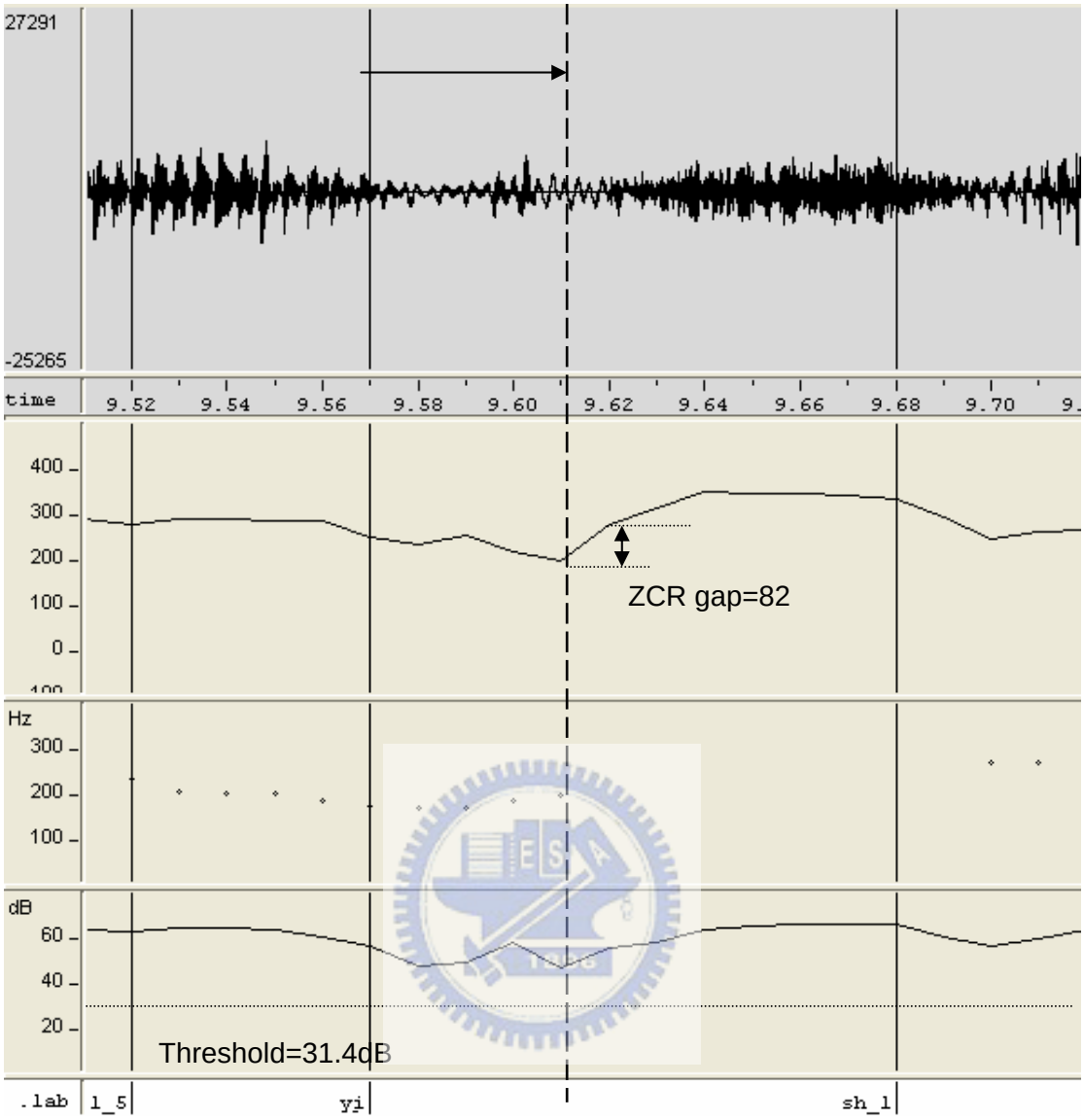
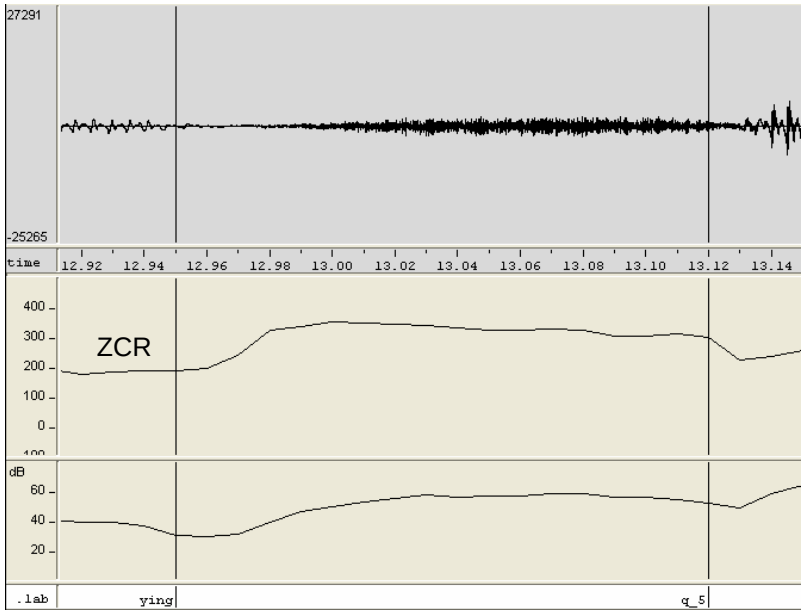


圖 二-7：音節間無 silence、pitch contour 不連續、energy 大於門檻值且後一音節為摩擦音類時以 ZCR 作為調整切割位置依據

原始切割位置



調整後切割位置

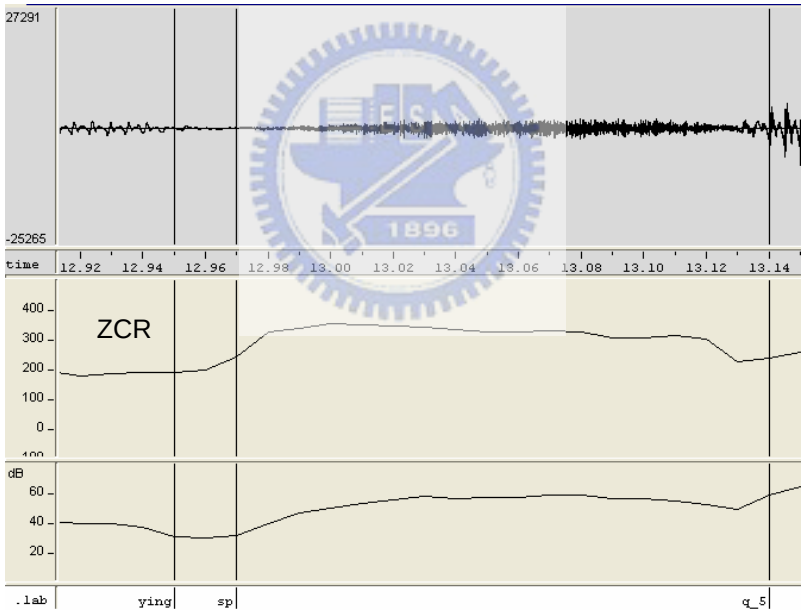


圖 二-8: 音節間無 silence、pitch contour 不連續、energy 小於門檻值時調整切割位置後出現短暫停

2.2.1.3 子音、母音交界位置之修正

利用子音母音是否有週期性的特性可將大部分的交界處正確的修正。在此，我們以 Wavesurfer 所求出語音的 voicing probability 參數為判斷標準。但若發現整個音節皆無週期性發生時，我們改用 energy 參數修正切割位置。

1. 利用 voicing probability：

當 voicing probability 等於 1 時，表示語音訊號有週期性，等於 0 時則否。如圖 二-9，以音節中第一個出現 1 的位置為分界點。

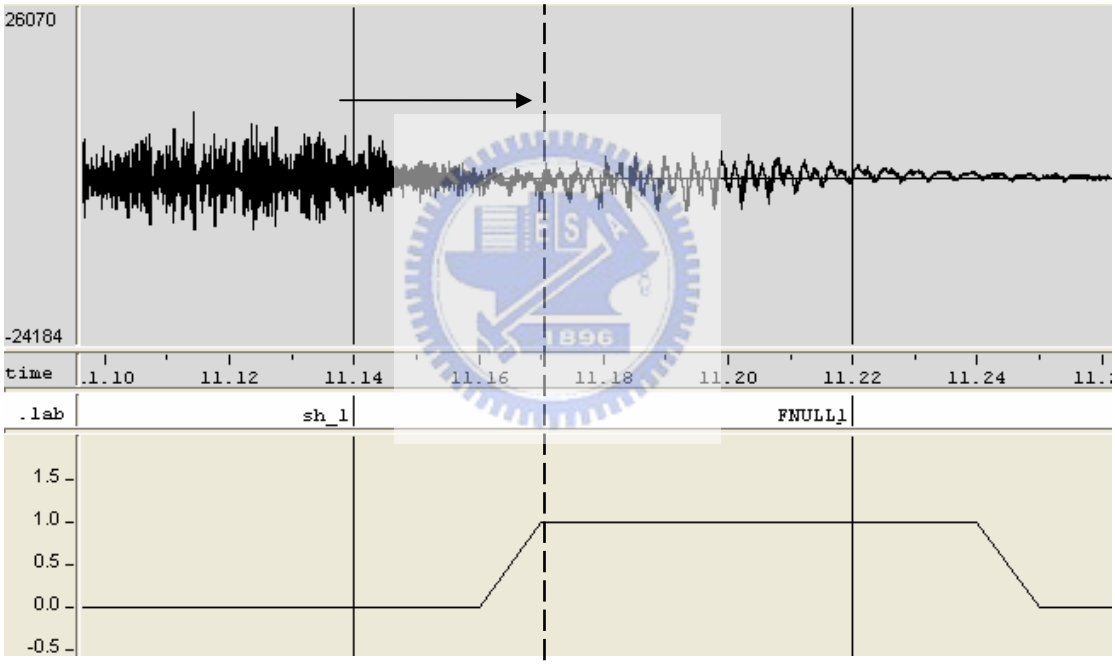


圖 二-9：利用 voicing probability 參數決定子音母音交界位置

2. 利用 energy：

子音母音交界時，通常能量會有較大變化，因此我們以能量變化最大的點做交界，如圖 二-10 所示。

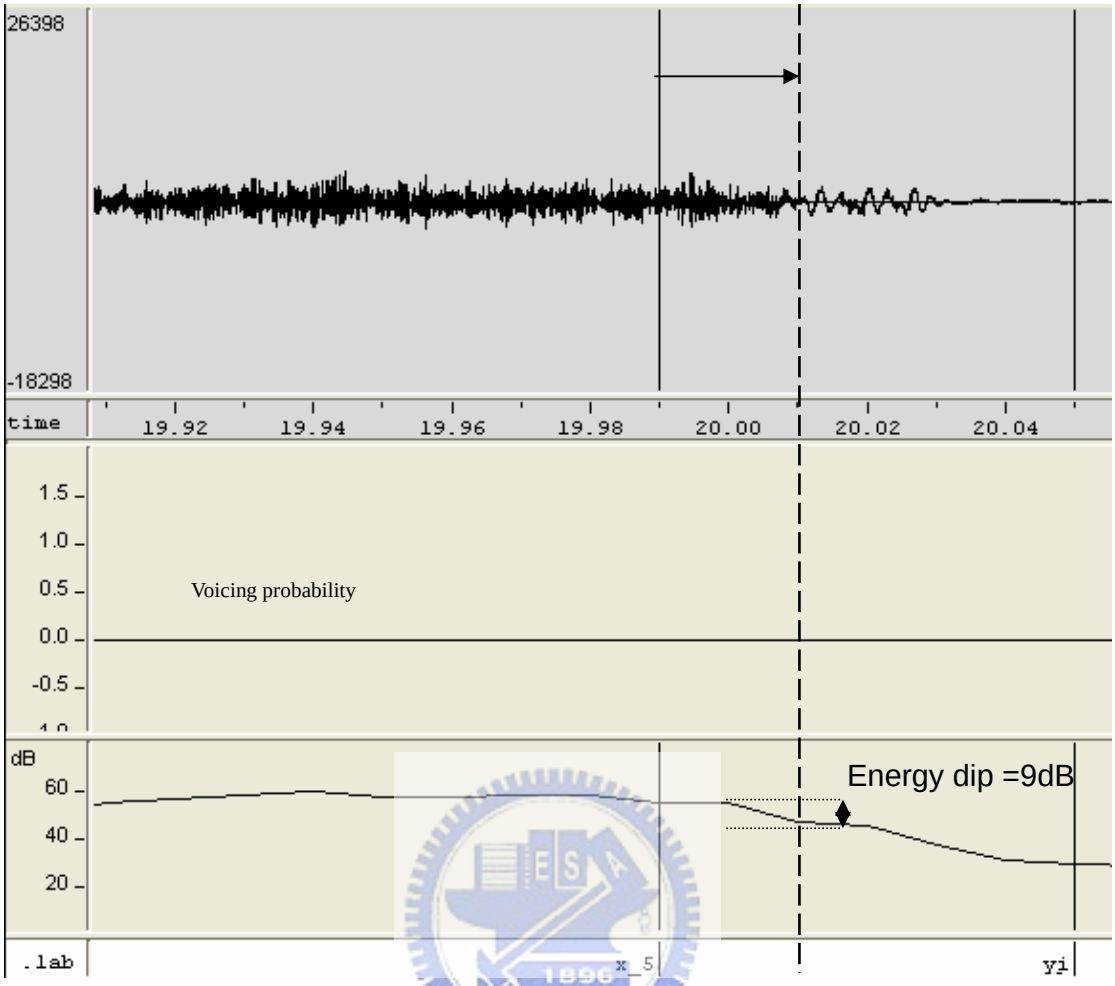


圖 二-10：利用 energy 決定子音母音交界位置

這裡有一點要特別注意，中文子音ㄇ、ㄋ、ㄌ、ㄖ具有週期性，並且能量在交界時也是連續的，因此以上的修正方法並不合適。所幸以合成的觀點而言，這一類子音的音節，在合成時並不需用到子、母音交界位置的資訊，因此碰到此類音節時，以不調整切割位置為原則。

以上的切割位置，皆以 frame 為單位，為了合成單元選取的正確，我們將切割位置轉換至以 sample 為單位。轉換的方法為在切割位置所在的 frame 裡，細分出 sub-frame（5 個 sample 點為一個 sub-frame），並找出語音能量和最低的 sub-frame，最後在 sub-frame 裡尋找語音能量最低的 sample 點即為所求。這裡的語音能量指的是每個 sample 的絕對值大小。

2.2.2 調整切割位置實驗結果與討論：

以人工實際切割幾個音檔為標準答案，分別計算自動切割和調整過後，切割位置與標準答案的偏移量，若切割位置因 pitch 連續，energy 連續而不調整時，並不納入統計之內；最後將所有切割偏移量以 5ms（80 個 sample 點）為間距，計算 0ms 至 45ms 的個數及大於 45ms 以上的個數，結果紀錄於表 二-1。

偏移長度(ms)	自動切割	調整自動切割
0~5	1700	4286
5~10	1725	2684
10~15	1852	1364
15~20	1285	1167
20~25	1131	308
25~30	721	211
30~35	539	253
35~40	438	191
40~45	327	143
45 以上	1172	283
總數	10890	10890
平均偏移量	20.820ms	10.175ms

表 二-1：切割偏移長度個數統計表

由表 二-1 可知，經過調整切割位置後，偏移長度小於 10ms 的比例，已有原始的 31%增至 64%，而偏移長度小於 20ms 的比例，也由原來的 60%增至 87%；平均偏移量也降低了約 10ms。

對於偏移量大於 20ms 的情況，經由實際觀察後可發現造成偏移過大的主因為：

1. 人工切割位置不正確：

   人工切割時，往往因子音能量過低而造成音節起始位置判斷錯誤，這類的子音通常為「ㄈ、ㄏ」；如圖 二-11 所示，(a) 為人工切割結果；(b) 為自動切割的結果，子音 f_6 的起始位置偏移了 110ms，但是自動切割的結果較為正確。當我們遇到這類子音，是以 ZCR 參數的變化量做為調整的方式，因此不會因為能量過低而發生問題，圖 二-11 (c) 為修正的結果。

2. 自動切割 silence 不準：

   我們修正切割位置的方式主要是以自動切割出的 silence 長度為分類，因此若切割出的 silence 不正確，將會影響最後的修正結果，如圖 二-12，(a) 為自動切割結果，(b) 為修正後的結果，兩種結果的偏移量均大。經過觀察多個此類錯誤後，可發現導致自動切割錯誤的原因在於音節的起頭有很強的氣音，這是沒有控制錄音品質所造成的現象。對於這一類的錯誤，我們可以在調整切割後，檢查每個 silence 的能量是否大於一個極大的值 (50dB)，若有，則以手動的方式，將切割調整回正確的位置。

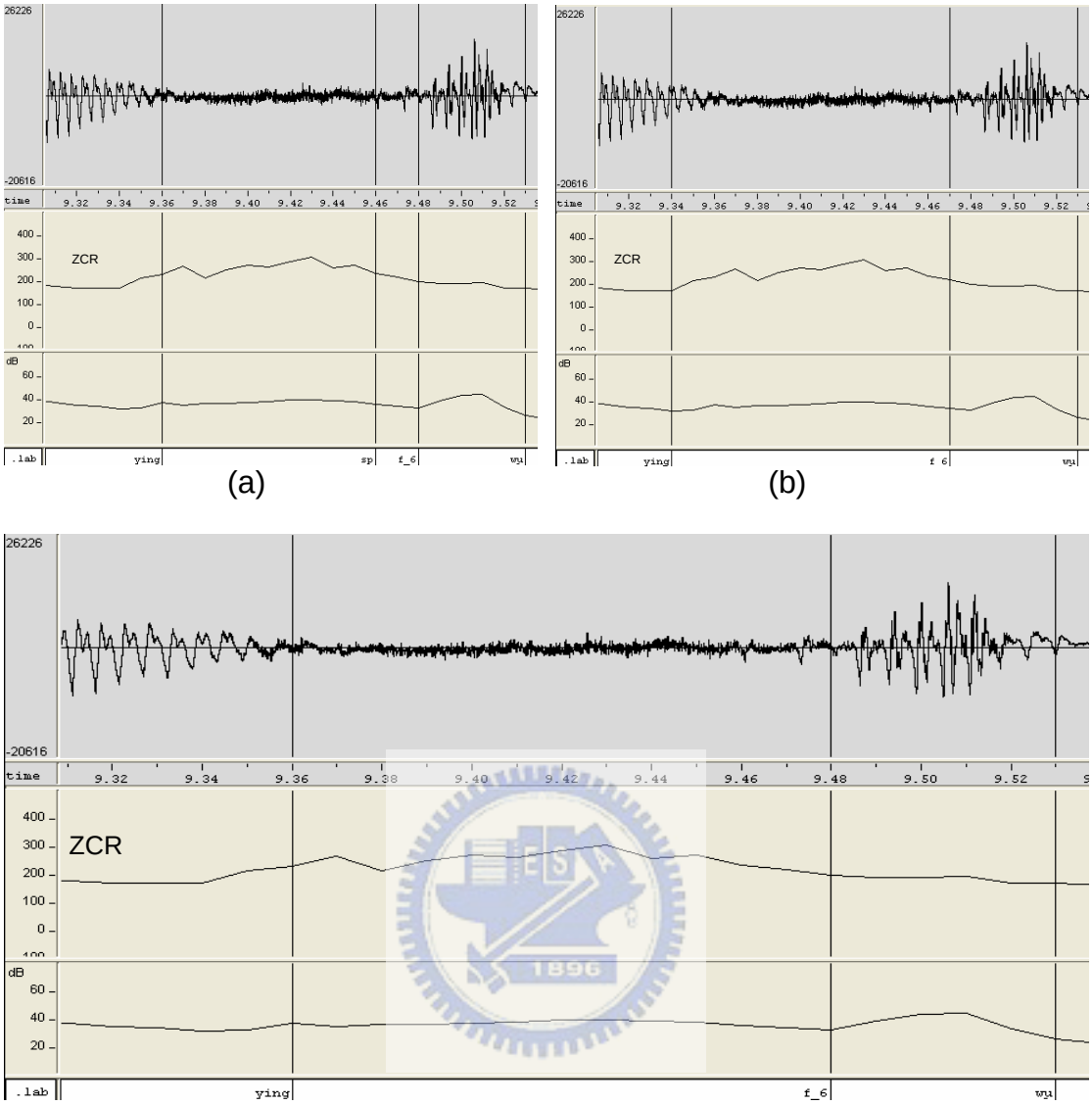
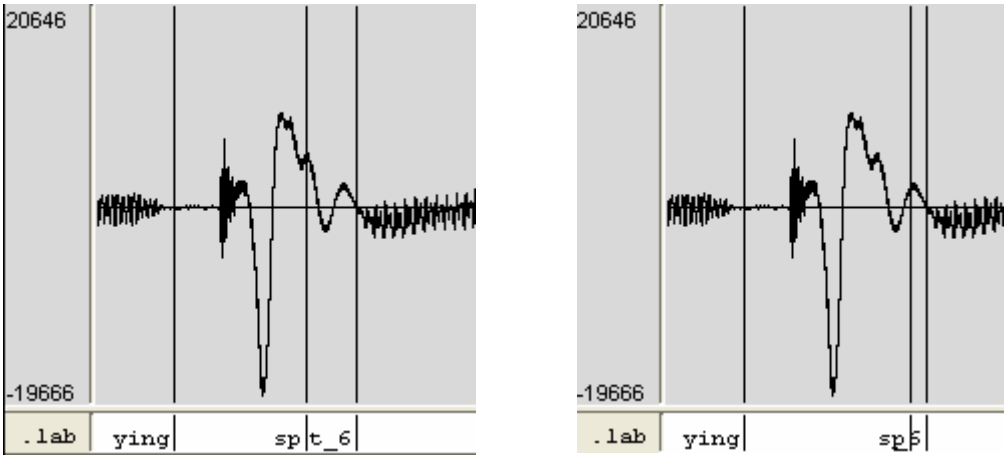


圖 二-11：能量過低導致人工切割錯誤



(a) (b)

圖 二-12：氣音所造成切割上的錯誤

2.3 基頻軌跡參數的求取：

基頻軌跡求取方式有很多種【4】【6】，目前我們是透過 Wavesurfer 這套軟體將每個音節的基頻軌跡求出；如同其它的方法，二倍基頻（Double pitch）、半倍頻（Half pitch）的錯誤會出現在求得的基頻軌跡裡，為了能正確取得基頻資訊，我們試著建立一套檢查方式，在音節切割位置正確的情況下，將錯誤的基頻位置找出，以人工修訂的方式更正之。

2.3.1 基頻軌跡檢查前處理：

由 Wavesurfer 求得之基頻值（F0）常在音節的起頭或結尾，或是音節中出現一點不規則跳動，如圖 二-13 所示。為了避免此一狀況影響接下來的工作，若此點發生在 pitch contour 的頭尾，且不與前一個或下一個音節的 pitch contour 相連，則將此點捨去，若發生在 pitch contour 中，則利用內插法將此點補回。

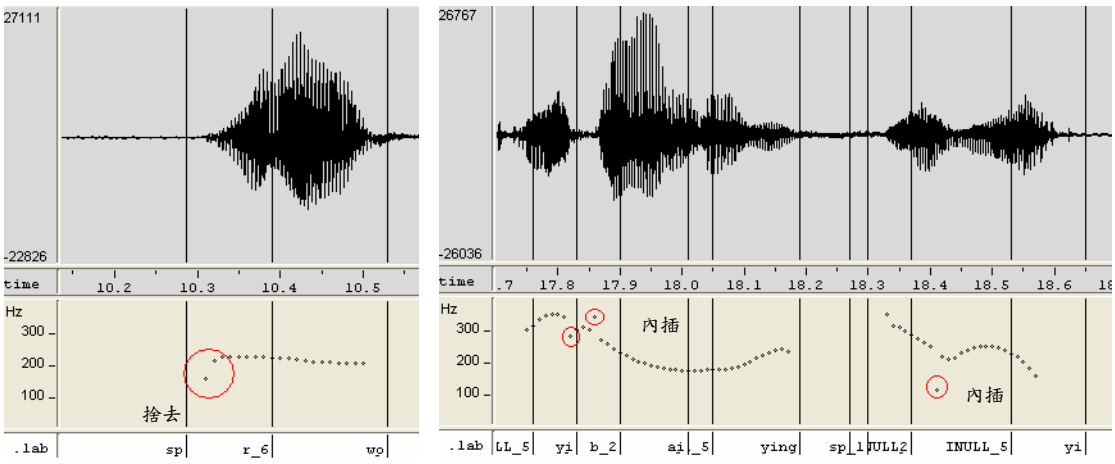


圖 二-13：由 Wavesurfer 求得之基頻軌跡圖（圈圈部分為異常值）

2.3.2 基頻軌跡檢查流程：

下圖為一音節區間內，基頻（F0）數值示意圖，虛線表 syllable boundary 位置。利用我們所提出的檢驗方式，可將音節區間內基頻有問題的地方標示出，其檢查的流程如下：

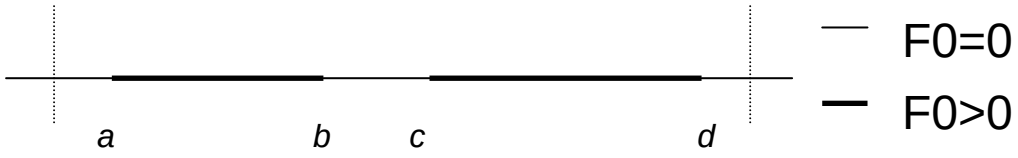


圖 二-14：音節 F0 示意圖

1. 如圖 二-14，先找出區間內第一個 F0 非零的位置 *a*，由此位置開始記錄連續非零的 F0 值，直到下一個 F0 為零的位置 *b* 為止。
2. 接著再找下一個非零值位置 *c*，若 *b*、*c* 間只有一個值為 0 的 F0，則用此位置左右各一點自動內插出此點之 F0（此動作可在前處理中發現），若有二個以上為 0 之 F0，則利用上述 1 之方法，從 *c* 點開始將此區間所有 F0 非零的區段找出。
3. 在正常的情況下，一個音節區間內必有 F0，且 F0 會連續，因此只會找出一段 F0 非零的區段。若區段數為零，表示此音節求不出 F0，若區段數大於等於二個，則表示切割位置有問題。
4. 若求得區段只有一個，為了求得基頻軌跡參數，F0 的個數要大於等於四，因此判斷區段內 F0 個數，若小於四個，則輸出錯誤訊息。
5. 若個數大於四個，求取音節 F0 的平均值 M，並依序檢查每個 F0 的變動是否超過一個範圍，依據觀察的結果，我們將此範圍設為±30Hz，若超過此範圍，則有 Double pitch 的現象發生。

6. 依步驟 1~5，求出每個音節的 M_j（並非所有音節皆可求出 M 值），比較 M_j 是否在一個範圍內（人說話時的 F0 約在 60~450Hz），若不在範圍內則視為錯誤發生。
7. 依序檢查 M_j 變動的範圍，若變化過大，則發生 pitch jump 的現象，但在 prosodic phrase boundary 時，M 值會突然變化，此現象屬於正常的變動，因此為了避免誤判，我們將連續觀察三個音節的 M 是否變動過大，才決定是否有 pitch jump 發生。

2.3.3 基頻軌跡檢查結果與討論

經由上面的檢查，我們可將錯誤分為：沒有求得 F0、音節內基頻軌跡不連續、F0 個數少於四個、double pitch、F0 超過範圍及 pitch jump 發生。除了音節內基頻軌跡不連續為切割位置錯誤造成外，其它錯誤發生原因可歸為下面幾類：

1. 音節發音時間短，或是在音節中，unvoiced 部分比 voiced 部分時間長很多，會造成求不出 F0 及 F0 個數小於四個的錯誤，尤其當子音為摩擦音類時，容易發生第二種狀況，如圖 二-15，(a) 為音節發音過短，(b) 為子音長度過長所造成求不出 F0 的現象。
2. 音節尾音拉得太長時，或音節與下個音節發生耦合現象時，會有 double pitch 錯誤發生。如圖 二-16 所示，(a) 為音尾發生 double pitch，(b) 為音節耦合後發生 double pitch 的現象。
3. 音節為輕聲，且在句尾時，常發生 pitch jump 的現象。

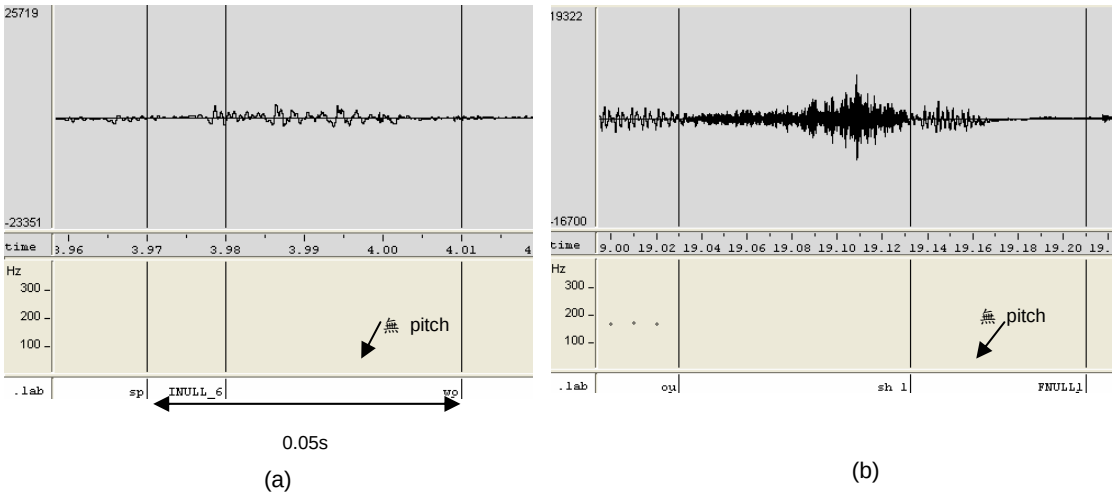


圖 二-15：音節中求不出 F0

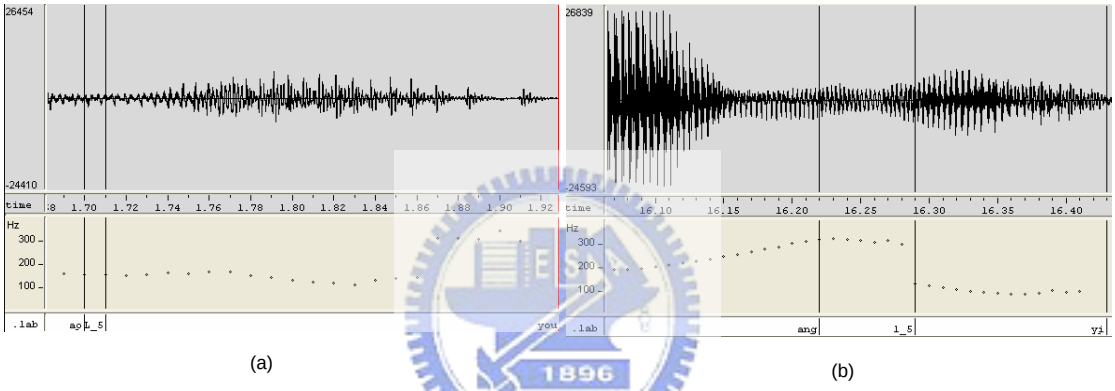


圖 二-16：基頻軌跡發生 double pitch

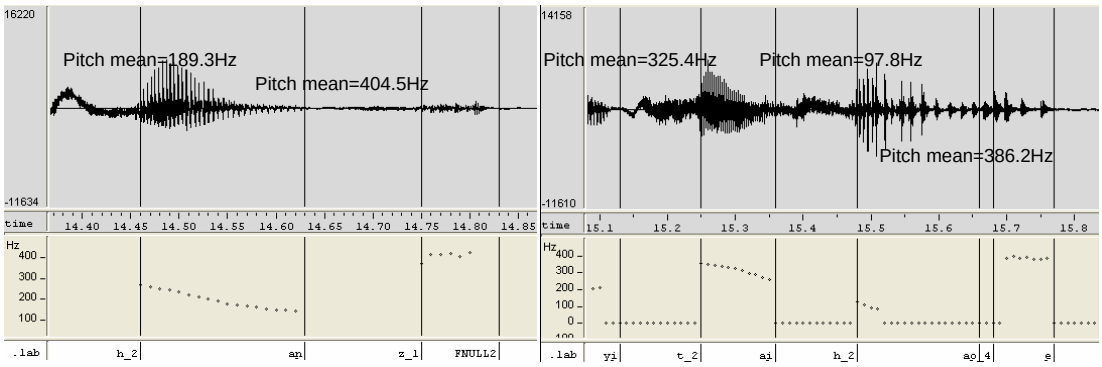


圖 二-17：基頻軌跡發生 pitch jump

經過以上的步驟，將所有錯誤找出後，我們以人工修正的方式，修訂基頻軌跡，當修正完成後，即可利用下一節介紹的方式，求得基頻軌跡參數。

2.3.4 基頻軌跡參數

語音合成的實作上，為了降低儲存所有基頻軌跡所花費的大量記憶體及運算時間，我們以基頻軌跡參數【7】代替原有基頻軌跡；基頻軌跡參數為所有基頻軌跡正交化展開之前四階係數，其詳細數學式如下：

$$a_j = \frac{1}{N+1}\sum_{i=0}^{N} Pitch(i)\cdot\Phi_j\left(\frac{i}{N}\right) \qquad (2.1)$$

其中 (a_0, a_1, a_2, a_3) 代表基頻軌跡的四個參數，$Pitch(i)$ 為原始基頻軌跡，$0 \le i \le N$ ，N+1 為基頻軌跡的長度。而 $\Phi_j(\frac{i}{N}), 0 \le j \le 3$ 為正交化函數，其定義如下：

$$\Phi_0\left(\frac{i}{N}\right) = 1$$

$$\Phi_1\left(\frac{i}{N}\right) = \left[\frac{12\cdot N}{(N+2)}\right]^{1/2}\cdot\left[\left(\frac{i}{N}\right) - \frac{1}{2}\right]$$

$$\Phi_2\left(\frac{i}{N}\right) = \left[\frac{180\cdot N^3}{(N-1)(N+2)(N+3)}\right]^{1/2}\cdot\left[\left(\frac{i}{N}\right)^2 - \left(\frac{i}{N}\right) + \frac{N-1}{6\cdot N}\right]$$

$$\Phi_3\left(\frac{i}{N}\right) = \left[\frac{2800\cdot N^5}{(N-1)(N-2)(N+2)(N+3)(N+4)}\right]^{1/2}\cdot \left[\left(\frac{i}{N}\right)^3 - \frac{3}{2}\left(\frac{i}{N}\right)^2 + \frac{6N^2-3N+2}{10\cdot N^2}\left(\frac{i}{N}\right) - \frac{(N-1)(N-2)}{20\cdot N^2}\right] \qquad (2.2)$$

若知基頻軌跡參數，可以利用下式重建基頻軌跡。

$$\hat{Pitch}(i) = \sum_{j=0}^{3} a_j\cdot\Phi_j\left(\frac{i}{N}\right), 0 \le i \le N \qquad (2.3)$$

第三章 語音合成技術實作與改進

近幾年來，國立交通大學電信工程學系語音實驗室致力於國語語音合成系統的發展，並且已成功合成出相當流利的語音【12】。其基本流程如圖 三-1 所示，由四個主要部分組成：文句分析(Text Analysis, TA)、韻律訊息產生器(Prosody Generator)、樣本音節 （Acoustic Inventory） 的產生與選取、語音合成器（Waveform Synthesizer)。

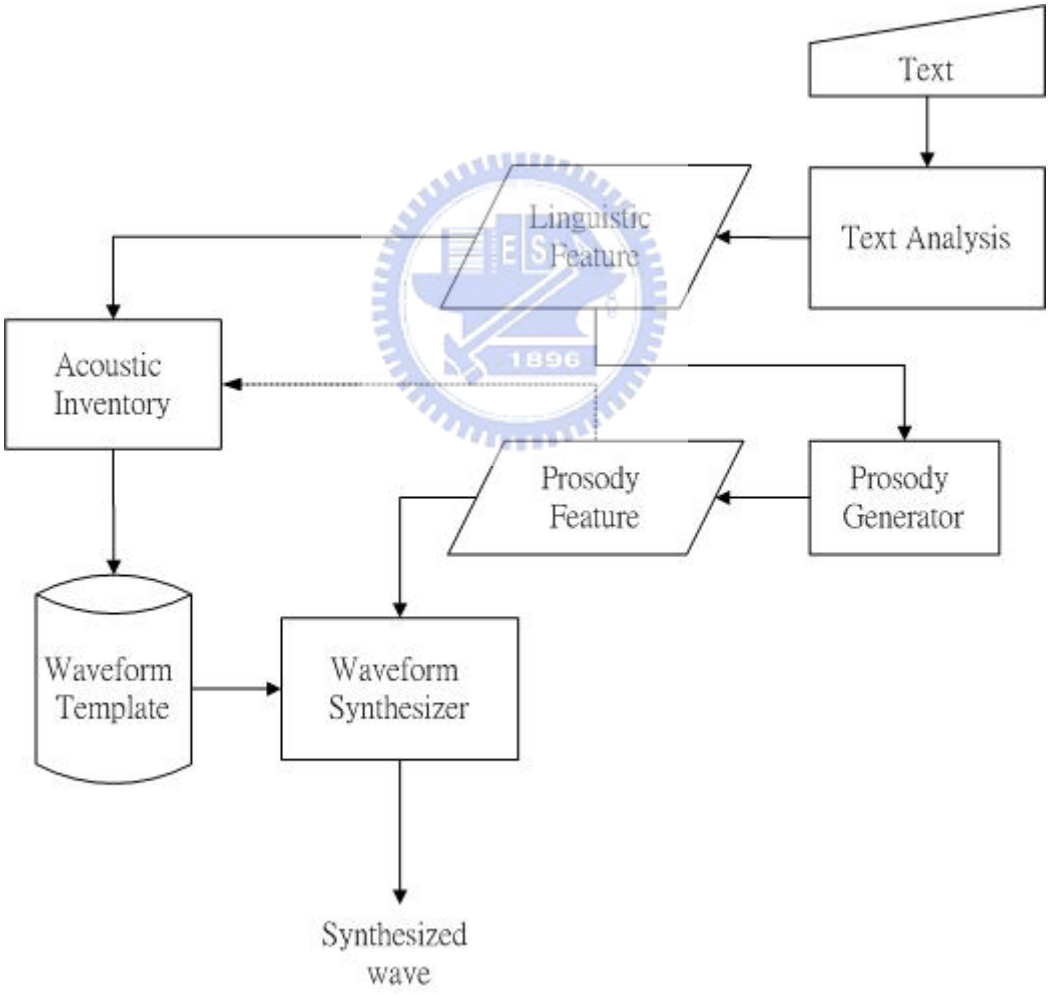


圖 三-1：語音合成系統架構圖

文字分析的主要目的，在於對輸入的文字，分析出正確的音碼串（Syllable

sequence)、詞串（Word sequence）和詞類串（Part-of-speech sequence, POS sequence）等語言參數（Linguistic features），因此又稱文字分析器為斷詞器（Tagger）。目前中文的斷詞器的斷詞精確率已達 0.875，召回率 0.793【11】，未來將針對破音字分析及詞組（word chunk）選取做更進一步的研究。

韻律產生器的功能，是對 TA 所產生的語言參數 syllable sequence、word sequence、POS sequence 分別取出詞長、詞類和音碼作為輸入，輸出每個音節所對應的韻律訊息，包含四個基頻軌跡參數、三個時間長度參數和一個能量參數。目前韻律產生器是以遞迴式類神經網路（Recursive neural network, RNN）的概念建構而成【12】，由於類神經網路可模擬人腦學習與記憶的功能，因此在長時間的訓練下可獲得不錯的效果。

以下我們將詳細介紹樣本音節的製作與選取和語音合成器製作的過程。

3.1 樣本音節的製作與選取：

初期的樣本音節是以單音節錄製而成，因此音節長度有過長的現象（平均長度 0.65s），然而使用 TD-PSOLA 合成語音，若樣本音節與合成音節的長度相差太多，會造成合成品質不佳【1】，為了改善這個問題，我們將從 Treebank 連續語料庫中，分別對每個中文 411 音節選取在音節長度上具有長、中、短性質的三個樣本音節，使得合成時能夠選出最接近合成音節長度的樣本音節，以增進合成品質。其挑選的原則如下：

1. 在音節長度最小值、平均值、最大值附近尋找適合的音節，如圖 三-2。

圖 三-2：三個樣本音節在音節長度上的分佈情形

2. 選取適合的音節：

對於一個音節是否為合適的樣本音節，我們可從此音節與上、下音節相連時的聲學特性（Acoustic features）來判斷。在此我們主要是以 pitch contour 和 energy level 兩種參數，將音節相連的情況分為四類，這四類分別為 pitch contour 相連、無明顯 energy dip，pitch contour 相連、energy dip 夠深，pitch contour 不相連、無明顯 energy dip，pitch contour 不相連、energy dip 夠深。詳細的分類原則與例子在 2.2.1 節中已說明，而我們認為除了 pitch contour 相連、無明顯 energy dip 為不合適的音節外，其餘的情況皆可作為合適的樣本音節。唯有在實際合成，有些種類的樣本音節需再經過處理。例如當 pitch contour 不相連、energy level 過高時，我們需將樣本音節的尾音以能量平滑下降的方式作處理。

3. 編修樣本音節

觀察選取出的樣本音節後，發現音節子音為摩擦音時，子音能量有過大的現象，如圖 三-3 所示，這對之後合成聲音的品質造成相當程度的影響，因此，對於這類的情況，我們在不影響音節聲音品質下，將子音能量降低。調降的詳細的過程如下：

a. 求出子音、母音的最大能量 E_u 、E_v ，並設定調降能量為 p dB。

b. 依式（3.1）、（3.2）將求出 G_u 、G_t 求出並設定 G_v 為 1，其中 y 為能量過渡的時間長度。

c. 最後依式（3.3）將聲音訊號 $x(n)$ 調整至 $x'(n)$。

$$G_u = 10^{\frac{(E_v - E_u - p)}{20}} \qquad (3.1)$$

$$G_t = \frac{(1-G_u)}{y} \tag{3.2}$$

$$x'(n) = \begin{cases} x(n) \times G_u \text{ , } n \in \text{能量調降區間} \\ x(n) \times G_t \text{ , } n \in \text{能量過度區間} \\ x(n) \times G_v \text{ , } n \in \text{能量不變區間} \end{cases} \tag{3.3}$$

依照上述的方式，嘗試調降 3dB、6dB、9dB、後，發現降 9dB 後的聲音品質仍佳，且波形在能量過度區間的改變影響不大，如圖 三-4，因此我們採用降 9dB 後的波形為樣本音節。

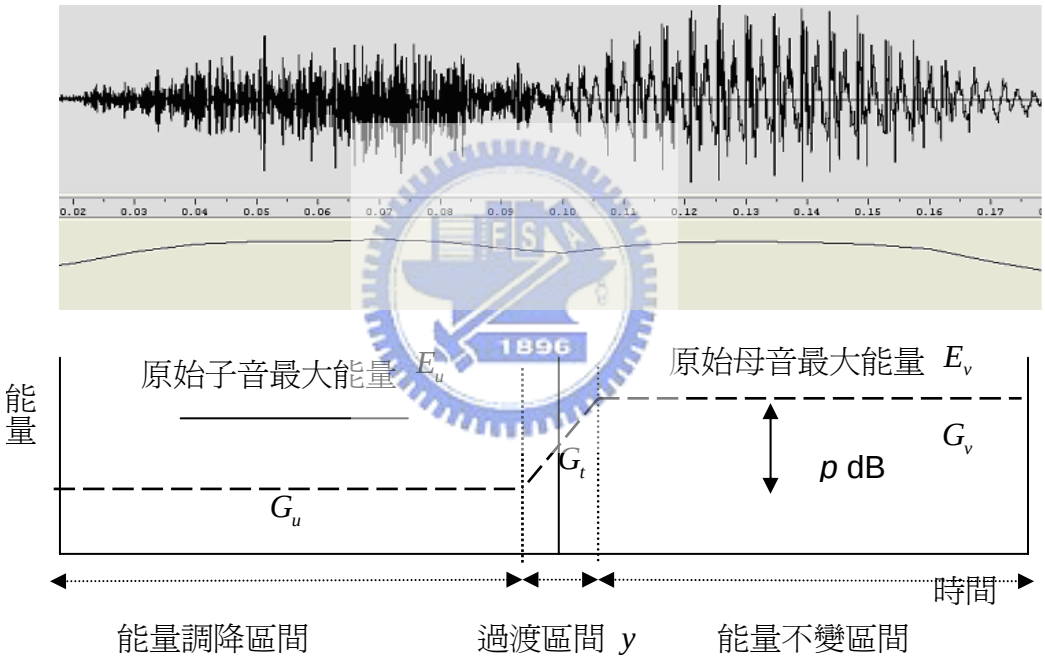


圖 三-3：處理樣本音節子音能量過大之示意圖

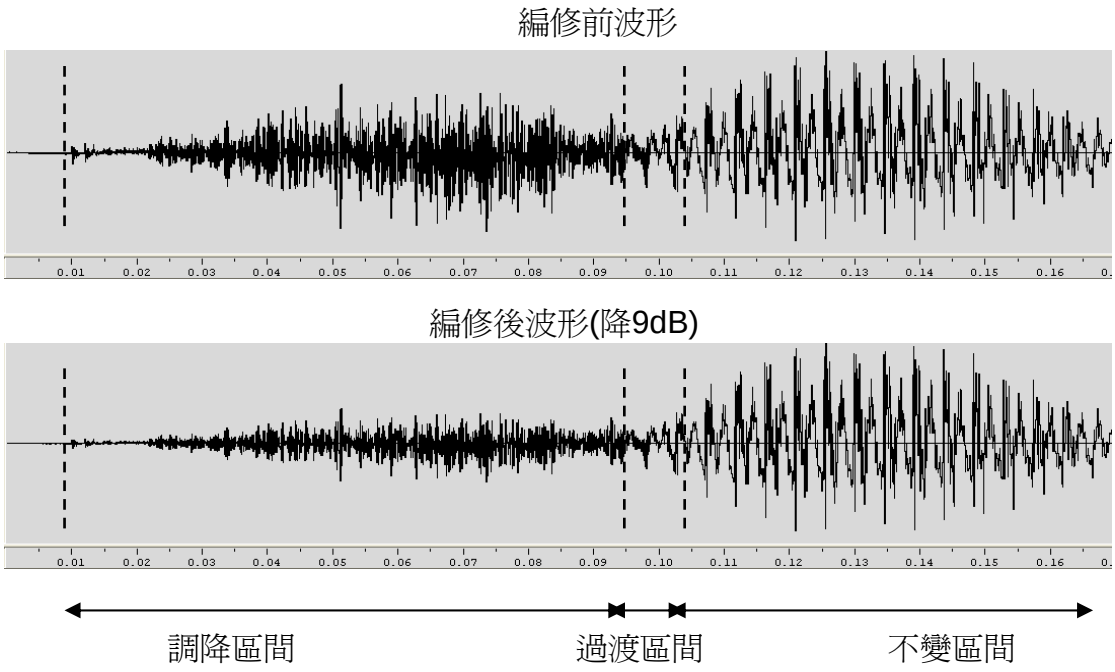


圖 三-4：編修前與編修後（能量降 9dB）的波形

4. 標示基調標記位置

當樣本音節選取後，在已知音節切割位置及 pitch contour 的情況下，我們以半自動的方式，標示出音節正確的基調標記（pitch mark）位置，標示的流程如下：

a. 搜尋音節具週期性（通常為音節母音）部分的中點 c，並由基頻軌跡中選取最接近 c 點的基頻值 $\mathrm{F0}_c$。

b. 在範圍 $\left(c-\frac{\mathrm{S}}{2\times\mathrm{F0}_c}, c+\frac{\mathrm{S}}{2\times\mathrm{F0}_c}\right)$ 裡尋找語音能量最大的 sample 點 p。其中 S 為取樣率。而 p 點則為我們所找到的第一個基頻標記位置。

c. 從 p 點往前往後找上一個和下一個基頻標記位置，尋找的範圍分別為 $\left(p_{\mathrm{pre}}-\frac{\mathrm{S}}{\mathrm{F0}_{p_{\mathrm{pre}}}}\text{-}20, p_{\mathrm{pre}}-\frac{\mathrm{S}}{\mathrm{F0}_{p_{\mathrm{pre}}}}+20\right)$、$\left(p_{\mathrm{pre}}+\frac{\mathrm{S}}{\mathrm{F0}_{p_{\mathrm{pre}}}}\text{-}20, p_{\mathrm{pre}}+\frac{\mathrm{S}}{\mathrm{F0}_{p_{\mathrm{pre}}}}+20\right)$，$p_{\mathrm{pre}}$ 為上次求得之基頻標記位置（第一次時，$p_{\mathrm{pre}}=p$）。

d. 利用 3.4、3.5 式，取出一個中心點在 p_{pre}、長度為 $\frac{\text{S}}{\text{F0}_{p_{\text{pre}}}}$ 的波形 $v_{p_{\text{pre}}}(n)$，及中心點在 p_{cand}（p_{cand} 為尋找範圍中的 sample 點）的波形 $v_{p_{\text{can}}}(n)$，其中 $\text{F0}_{p_{\text{pre}}}$ 為前一個基頻標記所在位置、$s(n)$ 為語音訊號。

e. 利用 3.6 式挑選出搜尋範圍內兩波形相關係數（Cross correlation）最大值的點 p^*_{cand}，為下一個基頻標記位置。

f. $p_{pre} = p^*_{\text{cand}}$、重覆 c、d、e 步驟往前往後依次求出每一個基頻標記的位置，直至音節邊界。

$$v_{p_{\text{pre}}}(n) = \begin{cases} s(n) & n \in (p_{\text{pre}} - L, p_{\text{pre}} + L) \\ 0 & \text{others} \end{cases} \tag{3.4}$$

$$v_{p_{\text{cand}}}(n) = \begin{cases} s(n) & n \in (p_{\text{cand}} - L, p_{\text{cand}} + L) \\ 0 & \text{others} \end{cases} \tag{3.5}$$

$$p^*_{\text{cand}} = \arg\max_{\forall\{p_{\text{cand}}\}} \left[\sum_{n=-L}^{L} v_{p_{\text{pre}}}(n) v_{p_{\text{cand}}}(n) \right]$$

$$L = \frac{\text{S}}{2 \times \text{F0}_{\text{pre}}} \tag{3.6}$$

自動標記後，為了確保正確性，我們仍須透過人工檢查的方式，將錯誤的基頻標記修正。以免因基頻位置不正確而影響合成品質。

經過以上的步驟，我們已取得樣本音節、樣本音節的 pitch mark，及所有樣本音節的聲學資訊，這些聲學參數將紀錄於波形表（wave table）裡，提供合成時，選取樣本音節的依據。

在下一小節裡，我們將介紹基頻同步疊加（PSOLA）合成法的原理，並應用時域基頻同步疊加（TD-PSOLA）演算法作為我們合成器的主要核心。

3.2 基頻同步疊加合成法

PSOLA 合成方法【2】【3】分三個步驟：首先是基頻同步分析，將原始語音信號與一序列基頻同步視窗函數相乘，可得到一序列相互重疊的短時信號；然後將這些短時信號進行適當的時域或頻域變換，得到與合成基頻軌跡同步的一序列合成短時信號；最後將合成短時信號重疊相加得到合成語音。

3.2.1 基頻同步分析

將原始語音信號 $x(n)$ 與一序列基頻同步的視窗 $h_m(n)$ 相乘得到一序列短時信號 $x_m(n)$

$$x_m(n) = h_m(t_m - n)x(n)$$

其中 t_m 是基頻同步標記的位置。視窗函數 $h_m(t_m - n)$ 的中心位於 t_m 處，一般採用漢寧窗(Hanning Window)。視窗長度通常大於一個基頻，造成相鄰的短時信號總有一部份重疊。一般而言視窗長度選取為相應位置處基頻 P 的 μ 倍，$2 \le \mu \le 4$，這樣可得到：

$$h_m(n) = h(\frac{n}{\mu P})$$

其中 $h(\bullet)$ 是具正規化長度的視窗函數。

3.2.2 基頻同步變換

將短時信號序列 $x_m(n)$ 轉換成與合成語音基頻標記 $\tilde{t}_q$ 同步的合成短時信號

序列$\tilde{x}_q(n)$，此轉換包括三個基本操作：改變短時信號個數、改變短時信號間的時間延遲及對每個短時信號的波形作變換。合成基頻標記$\tilde{t}_q$的個數取決於基頻調節係數β和時長調節係數γ，透過適當的演算法找出$\tilde{t}_q \to \tilde{t}_m$的映射關係，從而確定由哪些分析信號產生相對應的合成信號。

根據短時信號的波形變換方法的不同，可得到以下三種不同的PSOLA演算法：

1. 時域基頻同步疊加（TD-PSOLA）演算法：合成短時信號直接採用相對應的分析信號，只是將其在時間軸上移位$\delta_q = \tilde{t}_q - t_m$，即$\tilde{x}_q(n) = x_m(n-\delta_q) = x_m(n+t_m-\tilde{t}_q)$。此演算法只是刪除或重複某些短時信號，並根據時長改變和基頻改變的要求調整短時信號間的時間延遲。
2. 頻域基頻同步疊加（FD-PSOLA）演算法：合成信號是將分析信號$x_m(n-\delta_q)$進行頻域變換與反變換後得到。
3. LP-PSOLA 演算法：對多脈衝 LPC 合成中的多脈衝激勵進行時域 PSOLA 處理，以達到改變基頻的目的。雖然是將信號源與濾波器分開考慮，這種方法比 FD-PSOLA 演算法簡單。

上述三種 PSOLA 演算法，其複雜程度由難到易依次為：FD-PSOLA、LP-PSOLA 和 TD-PSOLA，從基頻和時長調節的效果來看，這三種算法效果相當，並且都比現有的 LPC 合成器和共振峰合成器好；但合成出的效果相當【12】，因此我們的合成器是以 TD-PSOLA 技術為核心。

3.2.3 基頻同步疊加合成

1. 簡單重疊相加（Allen and Rabiner, 1977）：

$$\tilde{x}(n)=\frac{\sum_{q}\alpha_q\tilde{x}_q(n)}{\sum_{q}\tilde{h}_q(\tilde{t}_q-n)} \tag{3.7}$$

其中表示$\tilde{h}_q(n)$合成視窗序列；正規化因子α_q用來補償基頻變換所造成能量變化。

2. 最小平均方重疊相加（Griffin and Lim, 1984）：

$$\tilde{x}(n)=\frac{\sum_{q}\alpha_q\tilde{x}_q(n)\tilde{h}_q(\tilde{t}_q-n)}{\sum_{q}\tilde{h}_q^{\,2}(\tilde{t}_q-n)} \tag{3.8}$$

從頻譜上來解釋，這種合成方法是使合成短時信號$\tilde{x}_q(n)$的頻譜與相對應合成信號的短時頻譜平方誤差為最小。在公式（3.7）中的分母的作用是一個正規化因子，用來補償相鄰視窗的重疊的差異所造成的能量的變化。在窄頻帶條件下，特別是當合成視窗的長度選取為合成信號基頻週期的二倍時，這個因子幾乎是常數。在這種情況下，如果假設$\alpha_q=1$，則公式簡化為：

$$\tilde{x}(n)=\sum_{q}\tilde{x}_q(n) \tag{3.9}$$

此時合成信號只是合成短時系號的線性和。

3.2.4 時長變化

時長變換可以與基頻變換同時進行，也可以單獨進行。如果只做時長變換，則不需要頻域變換，只用 TD-PSOLA 方法即可。最簡單的一種時長變換為時長變

換因子是常數的情形。這種情形下，如果要延長語音，那麼就將部分短時分析信號重複，反之，如果要縮短語音，則刪去部分分析短時信號。圖三-5 中帶箭頭直線分別為分析時間軸和合成時間軸上的基頻標記，虛線代表其間的對應關係。（a）所示為降低說話速度，有兩個分析短時信號 c1 與 c2 被重複；（b）所示為提高說話速度，有兩個分析短時信號 c3 與 c4 被刪除。

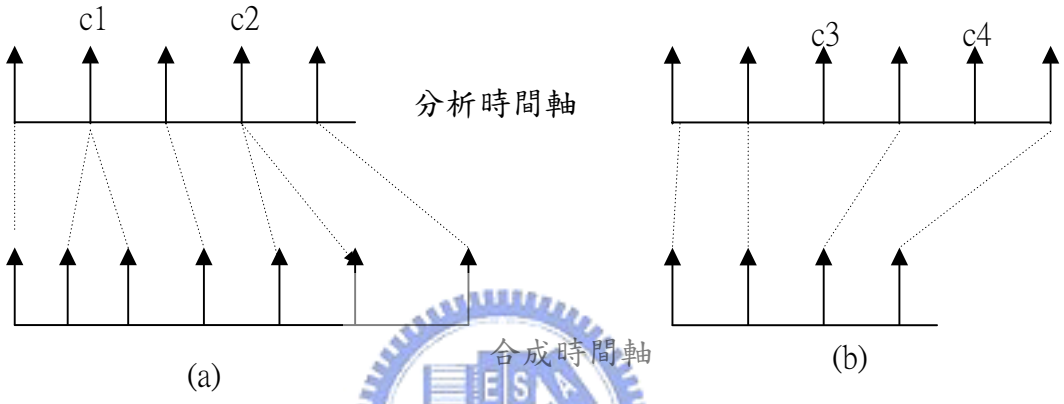


圖 三-5： 以 TD-PSOLA 改變時長

3.2.5 音高變化

進行音高變化，即為按照音高變化因子改變相鄰合成短時信號的時間延遲。此時音高變化本身隱含著時長的變化，一般情形下，基頻和時長都要變化。最簡單的一種情形是基頻變化因子與時長變化因子相等 $\beta = \gamma$ ，此時分析週期與合成週期是一對一對應的。通常這兩個因子並不相等，但我們可以將之視為先做一次時長因子與基頻因子都為 β 的變換，再做一次對時長因子為 γ / β 的變換。在實際操作裡，根據基頻因子和時長因子得出分析時間軸與合成時間軸的對應關係，從而直接找出分析信號和合成信號的對應關係，只進行一次變換較可以完成所需的基頻和時長變化。

3.3 語音合成器的實作與改進

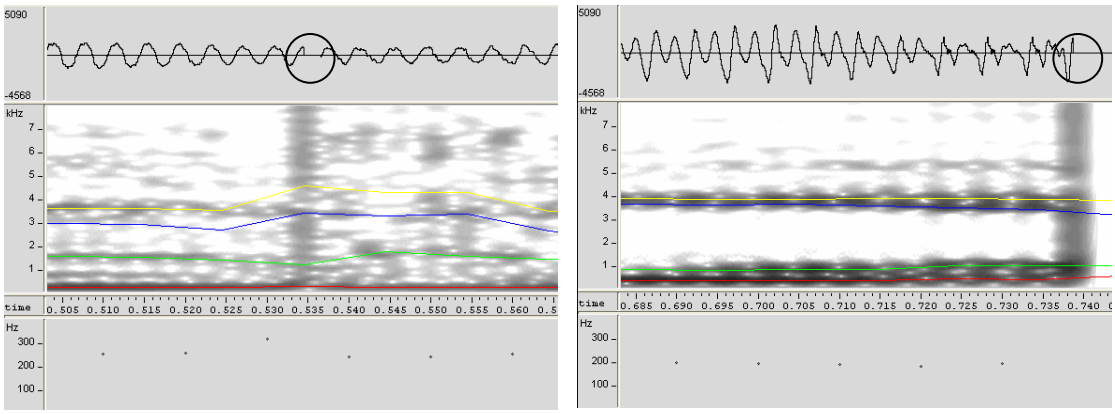
本節主要是以工程的角度，對於現有以 TD-PSOLA 技術為核心的合成器加以修正，以改善合成語音的品質。由於只針對合成器做改進，因此我們使用 TTS copy synthesizer【1】的作法，即以完全正確的韻律訊息做為輸入，避免參雜其它不必要的因素來影響合成品質。copy synthesizer 的作法如下：

1. 選定音檔後，用聽寫的方式將 Transcription 譯出（此動作在正確標示文句的語言參數，即模擬 TA 產生的結果）。
2. 標出正確的切割位置及 pitch mark，自動產生正確的韻律訊息（此動作在模擬 prosody generator）。
3. 將語言參數、韻律訊息、饋入合成器中產生合成語音。

將由以上作法求得之語言參數、韻律訊息作為原始合成器【12】的輸入，並且使用原始音節為樣本音節（除去樣本音節不佳所造成的影響），所合成出的語音在句尾及音節相接時會有「嗶嗶啵啵」的雜音，由圖 三-6 觀察波形可得知造成此現象的原因為

1. 合成語音在音節尾部能量劇降（energy discontinuity），如圖 三-6（b）所示。
2. 音節與音節間若有耦合效應時合成語音會發生相接不連續的現象，如圖 三-6（a）所示。

這二種現象會使聲音聽起來不順暢，因此接下來我們將針對這二個問題作分析，並在不改變 TD-PSOLA 構架的前提下做修正動作，以提高合成語音的流利度。



(a) (b)

圖 三-6：原始合成器合成語音波形、聲譜、基頻軌跡圖

3.3.1 音節尾部能量陡降

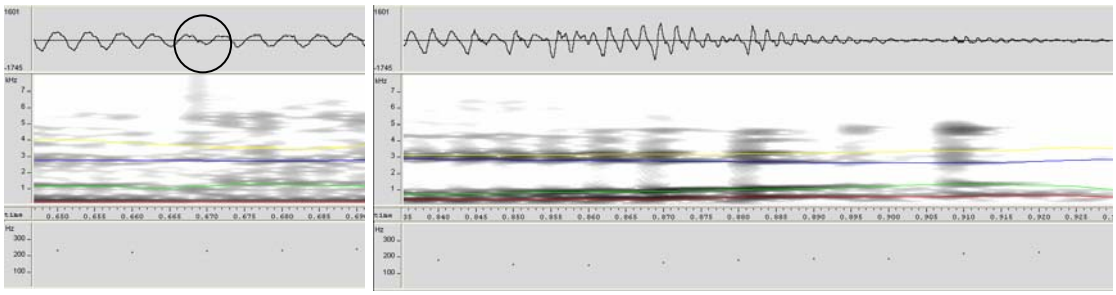
此種錯誤通常出現於詞尾或句尾，由於發聲的氣已用盡，因此音尾有無週期性的現象產生。在合成時，這種尾音往往被忽略而造成能量的不連續，針對此問題，我們試著將無週期的尾音直接接入合成出的音節尾，但需考慮到無週期尾音在樣本音節總長度中所佔的比例關係，以期最後合成出的音節中，無週期尾音與音節總長度的比例維持正確的關係。

3.3.2 音節相接不連續

音節相連時，下一音節若為空聲母，常常會發生音節的耦合效應。在語音合成時，若不考慮音節相接的耦合效應，在相接的位置上常會發生 phase discontinuity 和 pitch discontinuity 和現象，為了避免此現象影響合成品質，許多新的合成演算法被提出【1】【8】【10】。在不改變主要合成方法的原則下，我們試著找出造成這些現象的主因。由實驗中發現造成這些不協調現象的原因有二，第一在於兩音節標示的 pitch mark 相位不一致，第二為 PSOLA 使用的

window 於音節相接處並不同步。解決第一個問題只需重新標示 pitch mark 即可，而第二個問題需利用額外的 buffer 儲存此音節合成出的最後一個波形，與下個音節的起始合成波形做同步累加的動作，以達成 window 同步的結果。

圖 三-7 為改進這二個缺點後，合成波形圖；由（b）圖可看出，音節尾音能量已緩慢下降，而（a）圖顯示在音節相接時不連續的現象已改善。



(a) (b)

圖 三-7：改進後合成器合成語音波形、聲譜、基頻軌跡圖

以上所討論的問題發生在樣本音節直接從原始音檔音節中取的情形，當我們將樣本音節換為經 3.1 節介紹的方式所選取出的音節後，除了以上所描述的狀況外，我們發現合成出來的音檔在某些聲母的部分有較尖銳的發音，使得合成句子聽起來像是機器人發出的，因此我們必須重新檢視聲母合成的方式，以改善此狀況。

3.3.3 聲母合成方式的改進

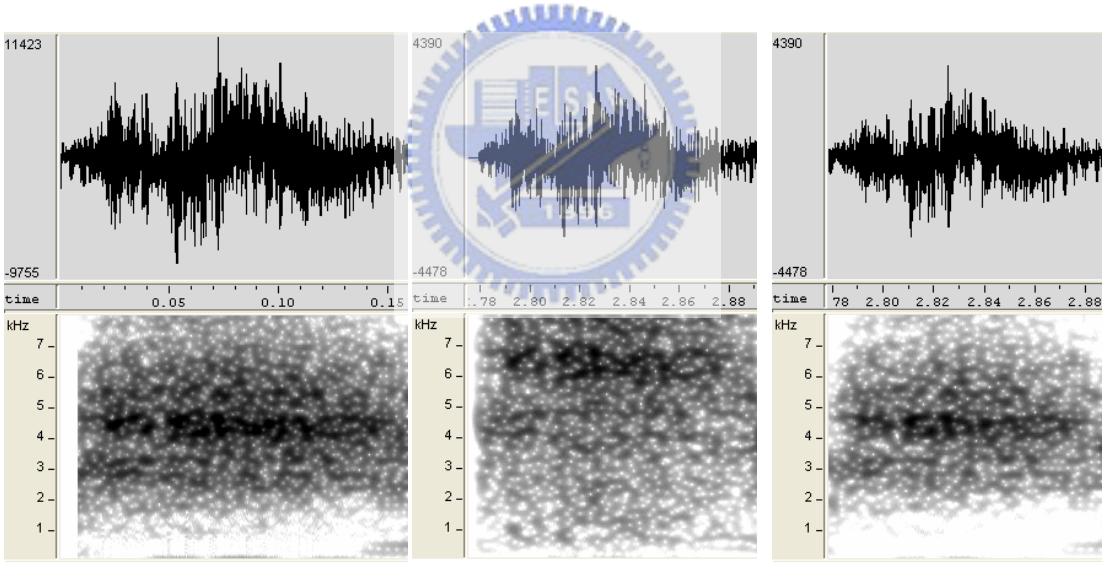
在舊有合成器中，聲母合成方式分為三大類：

1. 若聲母含有韻律性，則與其韻母視為一體，一同處理。此類的聲母包括ㄇ、ㄋ、ㄌ、ㄖ及空聲母。
2. 若聲母為爆破音，如「ㄅ、ㄆ…」則為聲母直接取出，不做任何處理。

3. 其它類的聲母，合成時對聲母做「re-sampling」的處理。

針對之前描述的問題，經由觀察可發現其原因來自於對聲母做 re-sampling 時，合成聲母的能量在高頻時會變大，如圖 三-8，(a)為原始樣本音節聲母，(b)為經過 re-sampling 後的聲母；為此，我們提出一個新的合成方式，能降低上述的效應，其作法如下：

1. 以韻母第一及第二個 pitch mark 間距為寬度，由第一個 pitch mark 位置開始，朝聲母部分做等寬度的標示 mark 動作，直至音節開頭。如圖 三-9 所示。
2. 以類似合成韻母的作法【13】，合成聲母，最後合成的結果可由圖 三-8（c）的聲譜圖看出，新作法的聲母能量分佈較接近原始能量分佈。



(a) (b) (c)

圖 三-8：摩擦音類聲母波形、聲譜圖

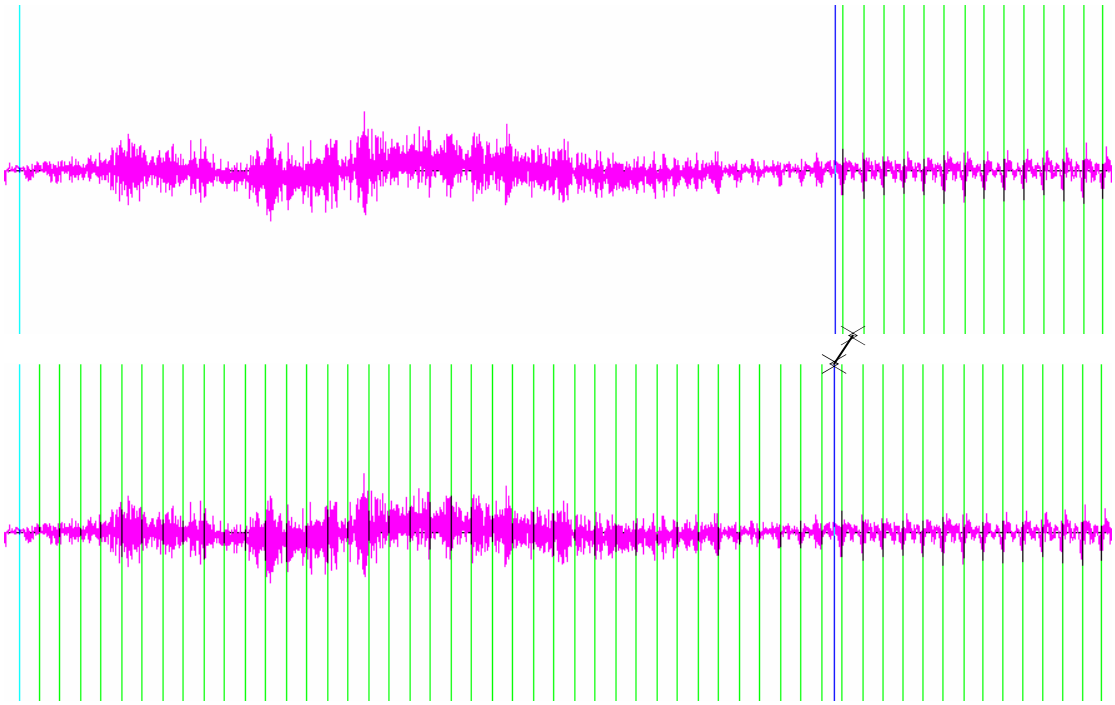


圖 三-9：聲母標示標記動作之示意圖

以新選出的音節作為樣本音節，加上改進過的語音合成器所合成出的句子，主觀上，我們認為聲音的品質有明顯的提升。

第四章 結論與未來展望

4.1 結論

本論文可分為二個部分:在第一部份裡我們建立了一套處理大量語料庫的標準流程，包含語音的自動切割、修正和基頻軌跡的求取與調整；第二部分則為針對過去發展之文句轉語音系統合成器和樣本音節資料庫作改進。經由實作後，我們可以得到下面幾點結論：

1. 我們所提出調整切割位置的方法，可將由 HMM 辨識器自動切割出音的位置，修正至較佳的切割位置，這樣的方法應用於切割大量語料庫上是可行且有效率的，並在利用正確切割位置的資訊下，基頻軌跡的錯誤偵測，也可達到一定的效果。

2. 在合成器以 TD-PSOLA 為架構的前提下，我們修正了以往影響合成品質的幾個要素：摩擦類子音合成時高頻能量過大、音節連接時語音不連續及音節尾音能量遽降等現象，使合成出的聲音品質更為流利。

3. 對於每個中文音節，我們建立了一組具有長、中、短三種不同長度的女性樣本音節，同時這樣的音節來自自然語音中，可使合成出的聲音，在品質上有明顯的提升。

4.2 未來展望

1. 我們已能對大量的語料庫做切割處理動作，並且求得語料的正確基頻軌跡，這是建立一個以大量語料庫為基礎的文句轉語音系統前所須完成的前處理動作，下一步，我們將以建立這樣的系統為目標，著手努力。

2. 在以較少語音資料量的前提下，選取長度上具有長、中、短特性的三個樣本音節的方法，已可合成出相當流利的語音，未來可對幾個較常用的字，加入更多的樣本音節，或是加入常用且易耦合的詞組於樣本中，可以減少因合成器不當的調整所帶來合成語音不順暢的現象。

3. 在挑選樣本音節的過程中，發現大多數以摩擦類起頭的音節，子音能量太強，使得合成時的句子有吵雜的現象。在實驗中，我們是以壓縮子音能量的方式改善這樣的問題，但這樣的調整，多少會改變原始音檔的結構；所以最根本的解決辦法，是在錄製語料庫時，要求音檔的品質；當有品質較佳的語料庫時，合成的聲音會更自然流暢。

4. 將新的斷詞器【11】和由本論文提出新的合成器及樣本音節，加上原始的韻律產生器，依照 Windows Speech API 的標準，包裝成一完整的 TTS 系統，可供視窗環境下使用。

參考文獻

[1] Dutoit, T., *An Introduction to Text-to-Speech Synthesis*, 1997, Kluswer Academic Publishers.

[2] E. Moulines, and F. Charpentier, "Pitch-synchronous Waveform Processing Technique for Text-to-Speech Synthesis Using Diphones," Speech Communication 9, pp.453-467, 1990.

[3] F. Charpentier and Moulines, "Pitch-synchronous Waveform Processing Technique for Text-to-Speech Synthesis Using Diphones," European Conf. On Speech Communication and Technology, pp.13-19, Paris, 1989.

[4] J. D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation," *IEEE Trans. On Audio and Electroacoustics, Vol.20*, pp.367-377, Dec.1972

[5] Kåre Sjölander and Jonas Beskow, "Wavesurfer – An open source speech tool," ICSLP 2000.

[6] L.R. Rabiner, "On the use of Autocorrelation Analysis for Pitch Detection," *IEEE Trans. on Acoustic, Speech and Signal Processing*, Vol.Assp-25, pp.24-33, Feb. 1977

[7] S.H. Hwang, S.H. Chen, and Y.R. Wang, "A Mandarin Text-to-Speech system," in Proc. ICSLP-96, pp.1421-1424, Oct.1996.

[8] Stylianou, Y. "Removing linear phase mismatches in concatenative speech synthesis" *IEEE Trans. on Speech and Audio Processing*, Volume9 , Issue 3 , March 2001 Pages:232 – 239

[9] V. Kraft, "Does the Resulting Speech Quality Improvement Make a Sophisticated Concatenation of Time-Domain Synthesis Units Worthewhile?" *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*, New Paltz, NY, pp65-68.

[10] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, *Spoken Language Processing*, Prentice-Hall, Inc.

[11] 江振宇， "中文斷詞器之改進"，國立交通大學碩士論文，民國九十三年六月。

[12] 魯弘茂，"中文語音合成技術之實作與分析"，國立交通大學碩士論文，民國九十一年六月。

[13] 盧鵬任，"中文文句翻語音系統之改進"，國立交通大學碩士論文，民國八十六年六月。

[14] 陳鳳儀, 蔡碧芳, 陳克健, 黃居仁，"中文句結構樹資料庫（Sinica Treebank）的構建"，中央研究院資訊所、中央研究院語言所。

附錄一

國語 411 個音節與 22 個聲母、39 個韻母模型對照表

音節	注音	聲母	韻母
1	ㄓ	16	1
2	ㄔ	17	1
3	ㄕ	18	1
4	ㄖ	19	1
5	ㄗ	20	1
6	ㄘ	21	1
7	ㄙ	22	1
8	ㄚ	1	2
9	ㄓㄚ	16	2
10	ㄔㄚ	17	2
11	ㄕㄚ	18	2
12	ㄗㄚ	20	2
13	ㄘㄚ	21	2
14	ㄙㄚ	22	2
15	ㄍㄚ	10	2
16	ㄎㄚ	11	2
17	ㄏㄚ	12	2
18	ㄅㄚ	6	2
19	ㄉㄚ	7	2
20	ㄋㄚ	8	2
21	ㄌㄚ	9	2
22	ㄅㄚ	2	2
23	ㄆㄚ	3	2
24	ㄇㄚ	4	2
25	ㄈㄚ	5	2
26	ㄛ	1	3
27	ㄌㄛ	9	3
28	ㄅㄛ	2	3
29	ㄆㄛ	3	3
30	ㄇㄛ	4	3
31	ㄈㄛ	5	3
32	ㄜ	1	4
33	ㄓㄜ	16	4
34	ㄔㄜ	17	4
35	ㄕㄜ	18	4
36	ㄖㄜ	19	4
37	ㄗㄜ	20	4
38	ㄘㄜ	21	4
39	ㄙㄜ	22	4
40	ㄍㄜ	10	4
41	ㄎㄜ	11	4
42	ㄏㄜ	12	4
43	ㄅㄜ	6	4
44	ㄉㄜ	7	4
45	ㄋㄜ	8	4
46	ㄌㄜ	9	4
47	ㄞ	1	6
48	ㄓㄞ	16	6
49	ㄔㄞ	17	6
50	ㄕㄞ	18	6
51	ㄗㄞ	20	6
52	ㄘㄞ	21	6
53	ㄙㄞ	22	6
54	ㄍㄞ	10	6
55	ㄎㄞ	11	6
56	ㄏㄞ	12	6
57	ㄅㄞ	6	6
58	ㄉㄞ	7	6
59	ㄋㄞ	8	6
60	ㄌㄞ	9	6
61	ㄅㄞ	2	6
62	ㄆㄞ	3	6
63	ㄇㄞ	4	6
64	ㄝ	1	5
65	ㄓㄟ	16	7
66	ㄕㄟ	18	7
67	ㄗㄟ	20	7
68	ㄙㄟ	22	7
69	ㄍㄟ	10	7
70	ㄏㄟ	12	7
71	ㄅㄟ	6	7
72	ㄋㄟ	8	7
73	ㄌㄟ	9	7
74	ㄅㄟ	2	7
75	ㄆㄟ	3	7
76	ㄇㄟ	4	7
77	ㄈㄟ	5	7
78	ㄠ	1	8
79	ㄓㄠ	16	8
80	ㄔㄠ	17	8
81	ㄕㄠ	18	8
82	ㄖㄠ	19	8
83	ㄗㄠ	20	8
84	ㄘㄠ	21	8
85	ㄙㄠ	22	8
86	ㄍㄠ	10	8
87	ㄎㄠ	11	8
88	ㄏㄠ	12	8
89	ㄅㄠ	6	8
90	ㄉㄠ	7	8
91	ㄋㄠ	8	8
92	ㄌㄠ	9	8
93	ㄅㄠ	2	8
94	ㄆㄠ	3	8
95	ㄇㄠ	4	8
96	ㄡ	1	9
97	ㄓㄡ	16	9
98	ㄔㄡ	17	9
99	ㄕㄡ	18	9
100	ㄖㄡ	19	9
101	ㄗㄡ	20	9
102	ㄘㄡ	21	9
103	ㄙㄡ	22	9
104	ㄍㄡ	10	9
105	ㄎㄡ	11	9
106	ㄏㄡ	12	9
107	ㄅㄡ	6	9
108	ㄉㄡ	7	9
109	ㄋㄡ	8	9
110	ㄌㄡ	9	9
111	ㄆㄡ	3	9
112	ㄇㄡ	4	9
113	ㄈㄡ	5	9
114	ㄢ	1	10
115	ㄓㄢ	16	10
116	ㄔㄢ	17	10
117	ㄕㄢ	18	10
118	ㄖㄢ	19	10
119	ㄗㄢ	20	10
120	ㄘㄢ	21	10
121	ㄙㄢ	22	10
122	ㄍㄢ	10	10
123	ㄎㄢ	11	10

音節	注音	聲母	韻母
124	ㄏㄢ	12	10
125	ㄉㄢ	6	10
126	ㄊㄢ	7	10
127	ㄋㄢ	8	10
128	ㄌㄢ	9	10
129	ㄅㄢ	2	10
130	ㄆㄢ	3	10
131	ㄇㄢ	4	10
132	ㄈㄢ	5	10
133	ㄣ	1	11
134	ㄓㄣ	16	11
135	ㄔㄣ	17	11
136	ㄕㄣ	18	11
137	ㄖㄣ	19	11
138	ㄗㄣ	20	11
139	ㄘㄣ	21	11
140	ㄙㄣ	22	11
141	ㄍㄣ	10	11
142	ㄎㄣ	11	11
143	ㄏㄣ	12	11
144	ㄋㄣ	8	11
145	ㄅㄣ	2	11
146	ㄆㄣ	3	11
147	ㄇㄣ	4	11
148	ㄈㄣ	5	11
149	ㄤ	1	12
150	ㄓㄤ	16	12
151	ㄔㄤ	17	12
152	ㄕㄤ	18	12
153	ㄖㄤ	19	12
154	ㄗㄤ	20	12
155	ㄘㄤ	21	12
156	ㄙㄤ	22	12
157	ㄍㄤ	10	12
158	ㄎㄤ	11	12
159	ㄏㄤ	12	12
160	ㄉㄤ	6	12
161	ㄊㄤ	7	12
162	ㄋㄤ	8	12
163	ㄌㄤ	9	12
164	ㄅㄤ	2	12
165	ㄆㄤ	3	12
166	ㄇㄤ	4	12
167	ㄈㄤ	5	12
168	ㄓㄥ	16	13
169	ㄔㄥ	17	13
170	ㄕㄥ	18	13
171	ㄖㄥ	19	13
172	ㄗㄥ	20	13
173	ㄘㄥ	21	13
174	ㄙㄥ	22	13
175	ㄍㄥ	10	13
176	ㄎㄥ	11	13
177	ㄏㄥ	12	13
178	ㄉㄥ	6	13
179	ㄊㄥ	7	13
180	ㄋㄥ	8	13
181	ㄌㄥ	9	13
182	ㄅㄥ	2	13
183	ㄆㄥ	3	13
184	ㄇㄥ	4	13
185	ㄈㄥ	5	13
186	ㄧ	1	14
187	ㄐㄧ	13	14
188	ㄑㄧ	14	14
189	ㄒㄧ	15	14
190	ㄉㄧ	6	14
191	ㄊㄧ	7	14
192	ㄋㄧ	8	14
193	ㄌㄧ	9	14
194	ㄅㄧ	2	14
195	ㄆㄧ	3	14
196	ㄇㄧ	4	14
197	ㄨ	1	15
198	ㄓㄨ	16	15
199	ㄔㄨ	17	15
200	ㄕㄨ	18	15
201	ㄖㄨ	19	15
202	ㄗㄨ	20	15
203	ㄘㄨ	21	15
204	ㄙㄨ	22	15
205	ㄍㄨ	10	15
206	ㄎㄨ	11	15
207	ㄏㄨ	12	15
208	ㄉㄨ	6	15
209	ㄊㄨ	7	15
210	ㄋㄨ	8	15
211	ㄌㄨ	9	15
212	ㄅㄨ	2	15
213	ㄆㄨ	3	15
214	ㄇㄨ	4	15
215	ㄈㄨ	5	15
216	ㄩ	1	16
217	ㄐㄩ	13	16
218	ㄑㄩ	14	16
219	ㄒㄩ	15	16
220	ㄋㄩ	8	16
221	ㄌㄩ	9	16
222	ㄧㄚ	1	17
223	ㄐㄧㄚ	13	17
224	ㄑㄧㄚ	14	17
225	ㄒㄧㄚ	15	17
226	ㄌㄧㄚ	9	17
227	ㄧㄝ	1	18
228	ㄐㄧㄝ	13	18
229	ㄑㄧㄝ	14	18
230	ㄒㄧㄝ	15	18
231	ㄉㄧㄝ	6	18
232	ㄊㄧㄝ	7	18
233	ㄋㄧㄝ	8	18
234	ㄌㄧㄝ	9	18
235	ㄅㄧㄝ	2	18
236	ㄆㄧㄝ	3	18
237	ㄇㄧㄝ	4	18
238	ㄧㄞ	1	19
239	ㄧㄠ	1	20
240	ㄐㄧㄠ	13	20
241	ㄑㄧㄠ	14	20
242	ㄒㄧㄠ	15	20
243	ㄉㄧㄠ	6	20
244	ㄊㄧㄠ	7	20
245	ㄋㄧㄠ	8	20
246	ㄌㄧㄠ	9	20
247	ㄅㄧㄠ	2	20
248	ㄆㄧㄠ	3	20
249	ㄇㄧㄠ	4	20
250	ㄧㄡ	1	21
251	ㄐㄧㄡ	13	21
252	ㄑㄧㄡ	14	21
253	ㄒㄧㄡ	15	21
254	ㄉㄧㄡ	6	21
255	ㄋㄧㄡ	8	21
256	ㄌㄧㄡ	9	21
257	ㄇㄧㄡ	4	21
258	ㄧㄢ	1	22
259	ㄐㄧㄢ	13	22
260	ㄑㄧㄢ	14	22
261	ㄒㄧㄢ	15	22

音節	注音	聲母	韻母
262	ㄅㄧㄢ	6	22
263	ㄊㄧㄢ	7	22
264	ㄋㄧㄢ	8	22
265	ㄌㄧㄢ	9	22
266	ㄍㄧㄢ	2	22
267	ㄎㄧㄢ	3	22
268	ㄇㄧㄢ	4	22
269	ㄧㄣ	1	23
270	ㄐㄧㄣ	13	23
271	ㄑㄧㄣ	14	23
272	ㄒㄧㄣ	15	23
273	ㄋㄧㄣ	8	23
274	ㄌㄧㄣ	9	23
275	ㄍㄧㄣ	2	23
276	ㄎㄧㄣ	3	23
277	ㄏㄧㄣ	5	23
278	ㄧㄤ	1	24
279	ㄐㄧㄤ	13	24
280	ㄑㄧㄤ	14	24
281	ㄒㄧㄤ	15	24
282	ㄋㄧㄤ	8	24
283	ㄌㄧㄤ	9	24
284	ㄧㄥ	1	25
285	ㄐㄧㄥ	13	25
286	ㄑㄧㄥ	14	25
287	ㄒㄧㄥ	15	25
288	ㄅㄧㄥ	6	25
289	ㄊㄧㄥ	7	25
290	ㄋㄧㄥ	8	25
291	ㄌㄧㄥ	9	25
292	ㄍㄧㄥ	2	25
293	ㄎㄧㄥ	3	25
294	ㄇㄧㄥ	4	25
295	ㄨㄚ	1	26
296	ㄓㄨㄚ	16	26
297	ㄔㄨㄚ	17	26
298	ㄕㄨㄚ	18	26
299	ㄍㄨㄚ	10	26
300	ㄎㄨㄚ	11	26
301	ㄏㄨㄚ	12	26
302	ㄨㄛ	1	27
303	ㄓㄨㄛ	16	27
304	ㄔㄨㄛ	17	27
305	ㄕㄨㄛ	18	27
306	ㄖㄨㄛ	19	27
307	ㄗㄨㄛ	20	27
308	ㄘㄨㄛ	21	27
309	ㄙㄨㄛ	22	27
310	ㄍㄨㄛ	10	27
311	ㄎㄨㄛ	11	27
312	ㄏㄨㄛ	12	27
313	ㄅㄨㄛ	6	27
314	ㄊㄨㄛ	7	27
315	ㄋㄨㄛ	8	27
316	ㄌㄨㄛ	9	27
317	ㄨㄞ	1	28
318	ㄓㄨㄞ	16	28
319	ㄔㄨㄞ	17	28
320	ㄕㄨㄞ	18	28
321	ㄍㄨㄞ	10	28
322	ㄎㄨㄞ	11	28
323	ㄏㄨㄞ	12	28
324	ㄨㄟ	1	29
325	ㄓㄨㄟ	16	29
326	ㄔㄨㄟ	17	29
327	ㄕㄨㄟ	18	29
328	ㄖㄨㄟ	19	29
329	ㄗㄨㄟ	20	29
330	ㄘㄨㄟ	21	29
331	ㄙㄨㄟ	22	29
332	ㄍㄨㄟ	10	29
333	ㄎㄨㄟ	11	29
334	ㄏㄨㄟ	12	29
335	ㄉㄨㄟ	6	29
336	ㄊㄨㄟ	7	29
337	ㄨㄢ	1	30
338	ㄓㄨㄢ	16	30
339	ㄔㄨㄢ	17	30
340	ㄕㄨㄢ	18	30
341	ㄖㄨㄢ	19	30
342	ㄗㄨㄢ	20	30
343	ㄘㄨㄢ	21	30
344	ㄙㄨㄢ	22	30
345	ㄍㄨㄢ	10	30
346	ㄎㄨㄢ	11	30
347	ㄏㄨㄢ	12	30
348	ㄉㄨㄢ	6	30
349	ㄊㄨㄢ	7	30
350	ㄋㄨㄢ	8	30
351	ㄌㄨㄢ	9	30
352	ㄨㄣ	1	31
353	ㄓㄨㄣ	16	31
354	ㄔㄨㄣ	17	31
355	ㄕㄨㄣ	18	31
356	ㄖㄨㄣ	19	31
357	ㄗㄨㄣ	20	31
358	ㄘㄨㄣ	21	31
359	ㄙㄨㄣ	22	31
360	ㄍㄨㄣ	10	31
361	ㄎㄨㄣ	11	31
362	ㄏㄨㄣ	12	31
363	ㄉㄨㄣ	6	31
364	ㄊㄨㄣ	7	31
365	ㄌㄨㄣ	9	31
366	ㄨㄤ	1	32
367	ㄓㄨㄤ	16	32
368	ㄔㄨㄤ	17	32
369	ㄕㄨㄤ	18	32
370	ㄍㄨㄤ	10	32
371	ㄎㄨㄤ	11	32
372	ㄏㄨㄤ	12	32
373	ㄨㄥ	1	33
374	ㄓㄨㄥ	16	33
375	ㄔㄨㄥ	17	33
376	ㄖㄨㄥ	19	33
377	ㄗㄨㄥ	20	33
378	ㄘㄨㄥ	21	33
379	ㄙㄨㄥ	22	33
380	ㄍㄨㄥ	10	33
381	ㄎㄨㄥ	11	33
382	ㄏㄨㄥ	12	33
383	ㄉㄨㄥ	6	33
384	ㄊㄨㄥ	7	33
385	ㄋㄨㄥ	8	33
386	ㄌㄨㄥ	9	33
387	ㄩㄝ	1	34
388	ㄐㄩㄝ	13	34
389	ㄑㄩㄝ	14	34
390	ㄒㄩㄝ	15	34
391	ㄋㄩㄝ	8	34
392	ㄌㄩㄝ	9	34
393	ㄩㄢ	1	35
394	ㄐㄩㄢ	13	35
395	ㄑㄩㄢ	14	35
396	ㄒㄩㄢ	15	35
397	ㄌㄩㄢ	9	35
398	ㄩㄣ	1	36
399	ㄐㄩㄣ	13	36

音節	注音	聲母	韻母
400	ㄑㄩㄣ	14	36
401	ㄒㄩㄣ	15	36
402	ㄌㄩㄣ	9	36
403	ㄩㄥ	1	37
404	ㄐㄩㄥ	13	37
405	ㄑㄩㄥ	14	37
406	ㄒㄩㄥ	15	37
407	ㄦ	1	38
408	ㄧㄛ	1	39
409	ㄥ	1	13
410	ㄟ	1	7
411	ㄇㄛ	4	4