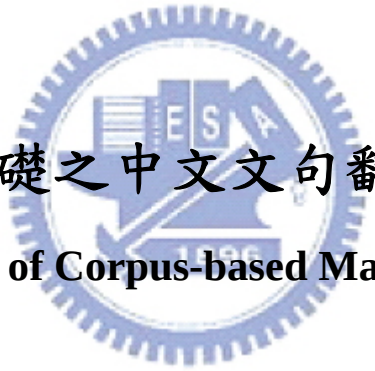
# 國 立 交 通 大 學

## 電信工程學系

## 碩士論文



# 以語料庫為基礎之中文文句翻語音系統實現

# An Implementation of Corpus-based Mandarin TTS System

研 究 生：洪國興

指導教授：陳信宏　博士

中華民國九十五年八月

# 以語料庫為基礎之中文文句翻語音系統實現

# An Implementation of Corpus-based Mandarin TTS System


研 究 生：洪國興　　Student：Kuo-Hsing Hung

指導教授：陳信宏　　Advisor：Dr. Sin-Horng Chen



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
in Electrical Engineering

August 2006
Hsinchu, Taiwan, Republic of China

中華民國九十五年八月

# 以語料庫為基礎之中文文句翻語音系統實現

研究生：洪國興　　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

文字轉語音系統中所使用的語料庫已由小量只提供有限基本音節的方式，演變為以大型語料庫為基礎的語料庫。在本論文中，實作了一套以語料庫為基礎之中文文句翻語音系統，除了整合既有合成系統中之文字分析器、韻律訊息產生器與波形合成器外，我們特地設計一個合成單元的選擇機制來解決以語料庫為基礎的合成系統會遇到的兩大問題，如何有效率地由大量語料中搜尋到可用之合成單元？以及如何挑選一個最佳的合成單元連接方式。本實作中，在不損害音質的前提下，針對搜尋效率改進了前人提出的連續相關比對法，也以更加嚴密的方式重新定義了用來挑選合成單元的 cost function。

最後，為了此系統使用上的方便性，我們設計了一套圖形化使用者介面。在此介面上，使用者直接輸入文字，然後可依賴系統自動合成語音，或者以其意願選擇合成單元。

# An Implementation of Corpus-based Mandarin TTS System

Student：Kuo-Hsing Hung Advisor：Dr. Sin-Horng Chen

Institute of Communication Engineering

National Chiao Tung University

## Abstract

Synthesis units in Mandarin Text-to-Speech system have migrated from small inventory of base-syllables to a large corpus. In this thesis, a corpus-based Mandarin Text-to-Speech system is implemented. Besides integrating the present text analyzer and the prosodic information generator, the study emphasizes on designing a unit-selection algorithm to solve the two main problems of searching all possible synthesis unit candidates in a huge corpus efficiently and selecting an appropriate unit sequence. We improve the efficiency of the continuous-correlative comparison method without decreasing the quality of the synthesis waveforms. Also we re-define the cost function used in the unit selection.

Finally, for users' convenience, we design a graphical user interface for this system. The user can directly type his input text, and get the synthesis waveform and some intermediate information automatically.

# 誌謝

在這裡要特別感謝陳信宏老師，在這兩年間不斷給予我研究及學問上的指導，以及王逸如老師對我們孜孜不倦的教導。郭威志學長帶領我進入語音的領域，林立峰學長教導我語音合成的基本概念，江振宇學長在我無法解決程式上的問題時，總是能適時的給予解答。喜歡看黑澀會的阿德學長，還有白天是國小老師晚上是研究生的輝哥，世上最愛老婆的智合兄，來無影去無蹤的希群學長也使我的在學生涯更加精彩。

具有召喚靈氣的鴻彥，每天打季賽的見偟，喜歡辣妹的世哲，到好湯猛吃蝦子的東毅，學習 HTK 的快樂世帆，喜歡說我又瘦了的振豐，喜歡讀書的阿勇，老愛唱眉飛色舞台客 Paul 以及眾多的學弟，有了你們，兩年的實驗室生活真是多采多姿。

如果人生可以重來一次，我或許可以選擇不讀電信，但不願意不與你們相識，謝謝大家，珍重再見！

# 目錄

# 表目錄

# 圖目錄

# 第一章　緒論

## 1.1 研究動機

電腦在現代化的資訊生活中，已成為日常生活的一部份，人類與機器的溝通不再侷限於特定職業或需求的人群身上，而是每天都會發生在每個人身上的事。不論是家庭主婦至金融機構辦理事務，上班族通勤時使用衛星導航系統，電腦無時無刻貢獻其功用。然而人類最自然的溝通方式，不外乎聽與說：聽出正確的訊息，說出欲表達的話語。為了建立起便利的人機介面，語音辨認與語音合成的研究，便扮演了舉足輕重的角色。

語音合成系統已發展多年，近年來發展方向由過去以小型語料庫提供基本音節的合成系統演變為以大型語料庫為基礎的語音合成系統，此轉變的趨勢所帶來的挑戰與創新是我們所感興趣的研究。

## 1.2 研究方向

以大型語料庫為基礎的語音合成系統與過去小型語料庫所不同的關鍵點在於，過去所注重的方向是如何產生正確的合成資訊，及控制合成技術所帶來的失真，而現在所在意的是，如何在一個大型語料中尋找可用來合成的資料，以及在此當中挑選出最佳的組合。過去交通大學語音實驗室已經嘗試建構了一套由大型語料庫中挑選出合成單元的機制[1]，可惜的是此機制並未發展到具有使用者介面之即時系統。現在更進一步要將系統發展為具有使用者介面的線上即時系統，更重要的是整合及改進語音合成系統中各模組的效能，以期能產生更為流利自然的合成語音。

## 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研究方向。

第二章　適合用於語音合成之大型語料庫之設計與語音合成系統資料庫建置

第三章　以語料庫為基礎之語音合成系統架構

第四章　系統設定與系統效能評估

第五章　結論與未來展望

# 第二章　適合用於語音合成之大型語料庫之設計與語音合成系統資料庫建置

決定以語料庫為基礎之語音合成系統效能關鍵之一，為此系統採用的語料庫是否適合用於語音合成，以及系統建構者是否由此語料庫中抽取出足夠的資訊以供系統使用，本章即是探討此一主題。第一節論述適合用於語音合成的語料庫條件，之後數節描述由語料庫原始資料建立用於系統中資料庫的過程。大致上來說，為了之後用於搜尋與挑選適當的合成單元，我們事先求取出語料庫中許多不同的語音特性，諸如音節位置、音高、能量之類的參數，並依據資料型式的不同，分為文字資料庫與語音參數資料庫，以供合成系統之用。

## 2.1 適合用於語音合成之大型語料庫之設計

在接下來的小節中，將介紹我們是如何建立起一套適用於語音合成系統的大型語料庫。

### 2.1.1 適用於語音合成之大型語料庫條件

一個語料庫是否適用於以語料庫為基礎的語音合成系統，癥結在於其是否擁有各種不同的合成單元。就中文而言，每個中文字對應一個音節(syllable)，音節有五種聲調(tone)的變化，常見的中文字約一萬兩千多字，但如果只以發音來區分，總共大約只有 1300 種音節，如果再去除聲調的差別，則只有 411 種基本音節。

一般認為，適合用於 TTS 系統的語料庫應具有「豐富語音」(phonetically rich)與「豐富韻律」(prosodically rich)兩個特性[2]。所謂「豐富語音」是指具有各式各樣的音節連接方式；而「豐富韻律」則是指語料庫具有多種不同的韻律變化。

雖然我們希望語料庫能夠包含所有的可能性，然而顧慮到語料庫的大小，一些取捨是必要的。

### 2.1.2 大型語料庫之資料來源

目前合成系統所使用之語料庫，其文字部分來自於「中央研究院中文文句結構樹資料庫 1.1 版」(Sinica Treebank Version 1.1)[3]，從中央研究院詞庫小組之「中央研究院現代漢語語料庫」得來。目前語料庫所使用之中文文句結構樹，共有 11,109 棵（語料庫的各項數據列舉於附錄二中），其形式如下例所示：

```
PP(Head:P43:依據|DUMMY:NP(property:NP・的(head:NP(property:Nca:
行政院|Head:Ncb:主計處)|Head:DE:的)|Head:Nad:統計))#，
(COMMACATEGORY)。
```

其中包含了語法訊息、語意角色及文字內容，上例也可表示為樹狀的結構如下圖所示。

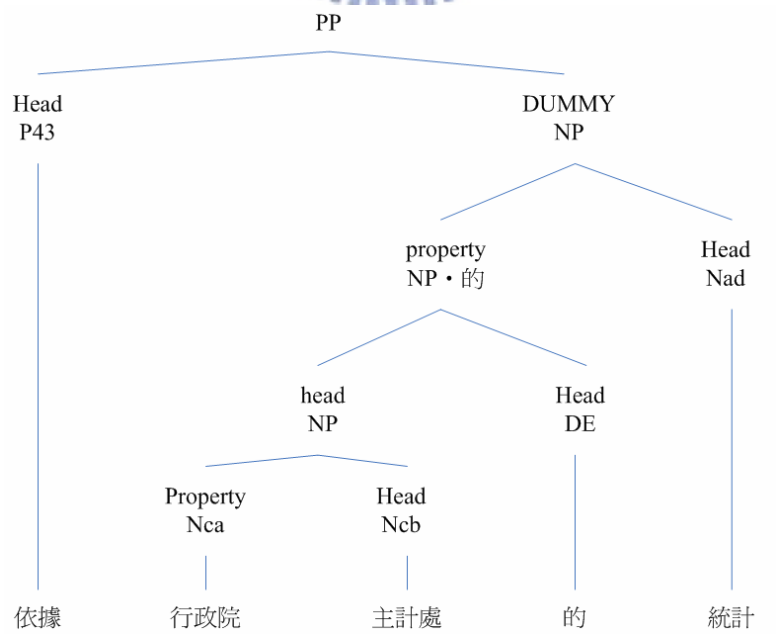



**圖 2-1-1：一個中文文句結構樹圖形表示的例子**

### 2.1.3 文字內容之萃取

我們首先由中文結構樹，將其文字內容萃取出來，上一例萃取之文字內容為

| |
|---|
| 依據行政院主計處的統計， |

並取得其相對應的詞性與詞長，並藉由文字分析器標記其發音，以聲碼表示之，聲碼與實際發音之對照關係請參照附錄一，「國語 411 基本音節總音表」，產生的文件格式如下，其中第一行為原始文字，第二行為音碼，第三行為斷詞結果、第四行為詞性：

```
依 1186  201 P43
據 4217  202 P43
行 2287  301 Nca
政 4168  302 Nca
院 4393  303 Nca
主 3198  301 Ncb
計 4187  302 Ncb
處 4199  303 Ncb
的 5043  101 DE
統 3384  201 Nad
計 4187  202 Nad
， 6001  101 PM
```

我們將所有的結構樹經過如此處理後，其文字內容部分計有 69,062 個詞，123,128 個字，平均詞長約為 1.78 個字，詳細的詞長分佈如下：

**表 2-1-1：語料庫詞長分佈表格**

| **詞長** | **個數** |
|---|---|
| 一字詞 | 25,244 |

| | |
|---|---|
| 二字詞 | 36,482 |
| 三字詞 | 5,461 |
| 四字詞 | 1,356 |
| 五字詞 | 262 |
| 六字詞 | 117 |
| 七字詞 | 71 |
| 八字詞 | 48 |
| 九字詞 | 7 |
| 十字詞 | 4 |
| 十一字詞 | 5 |
| 十二字詞 | 2 |
| 十三字詞 | 1 |
| 十四字詞 | 2 |
| 總數 | 69,062 |

詞長分佈圖

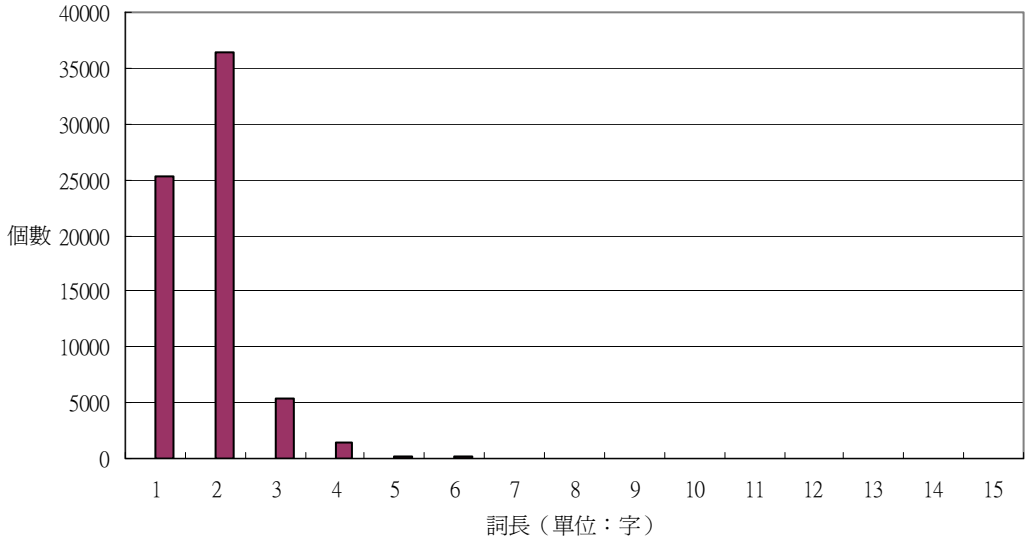



**圖 2-1-2：語料庫詞長分佈圖**

### 2.1.4 錄製音檔

產生如上所提的文字檔後，接下來要錄製音檔。我們將數個連續的結構樹文字內容湊成一篇短文，計有 1,433 篇短文，以一篇短文錄製為一個音檔的格式請錄音人員以流利的朗讀方式唸出。我們請了一位專業的女性廣播人員幫我們錄製語音，為了減少未來在合成單元切割或韻律求取上的錯誤，以得到較佳的合成單元品質，我們在錄製的過程中若有發生口吃、猶豫、或唸錯的情形，我們會請錄音人員再重新錄製該句直到正確為止，錄音軟硬體設備及格式詳如下表：

| | |
|---|---|
| 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) | 20 kHz |
| 發音速度 | 每秒約 4.6 個音節 |
| 取樣大小 | 16 bits (位元) |
| 聲道 | 單聲道(mono) |
| 檔案格式 | pcm |

## 2.2 語音合成系統之文字資料庫建置

由中研院所提供的中文結構樹中，我們可以知道一段文字是如何由不同種類的詞所構成的，然而此原始的資料並不足夠用於下一章節所要提出的合成系統，原因有三點，一是系統內部所使用的詞典可能未收錄中研院所提供的詞；二是之後的系統是以「詞」為出發點的搜尋單元，原始的資料中有許多字數較長的詞被拿來合成的機率甚低（如定量複合詞及專有名詞）；三是我們的文字分析器能標記中文姓氏、詞綴等有用的資訊，相對應的，語料庫也必須標記這些資訊。在此節中，將針對文章進一步整理，來提升文字資料庫的用途。。

### 2.2.1 補充詞典中缺乏的詞類

由中研院的平衡語料庫中，我們首先將所有的詞和所使用的詞典對照，找出詞典中所沒有收錄的部分，並以人工篩選的方式將其中的定量複合詞與專有人名移除，此兩類的詞之後要以特別的方式處理，將這些篩選後的結果加入目前所使用的詞典中，計有 238 個詞。

### 2.2.2 長詞化為短詞組合

在中研院的平衡語料庫中，存在著許多長詞，其詞長在四字以上。然而由於之後合成系統所搜尋的合成單元基本單位是一個完整的詞，所以這些詞長較長的詞被使用到的機率不大。當這些詞的部份內容也可以視作一個詞時，我們希望也能夠拿來合成。基於這個理由，我們將語料庫中詞長在四字以上的詞重新人工檢視一次，並把其中可以視作另一個詞的部分標記起來，也就是說，如果某個詞的部份內容和在詞典中記錄的詞相同時，當作這部分也是一個詞，以下舉例說明。

譬如說，在平衡語料庫中有「高爾夫球場」這個詞，我們希望它同時也可以包含「高爾夫」、「球場」、「高爾夫球」這些在詞典中的詞。原本我們記載中研院的文字資訊的文件格式如下

| | | | |
|---|---|---|---|
| 所 | 3309 | 201 | Cbca |
| 以 | 3186 | 202 | Cbca |
| 他 | 1019 | 201 | Nhaa |
| 們 | 5147 | 202 | Nhaa |
| 能 | 2180 | 101 | Dbab |
| 優 | 1250 | 201 | VH11 |
| 閒 | 2261 | 202 | VH11 |
| 地 | 4190 | 101 | DE |
| 揮 | 1334 | 201 | VA4 |
| 桿 | 3122 | 202 | VA4 |
| 高 | 1086 | 501 | Ncb |
| 爾 | 3407 | 502 | Ncb |
| 夫 | 1215 | 503 | Ncb |
| 球 | 2252 | 504 | Ncb |
| 場 | 3151 | 505 | Ncb |
| ， | 6001 | 101 | PM |

更改過後如下，

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 所 | 3309 | Cbca | 1 | 201 | | | |
| 以 | 3186 | Cbca | 1 | 202 | | | |
| 他 | 1019 | Nhaa | 1 | 201 | | | |
| 們 | 5147 | Nhaa | 1 | 202 | | | |
| 能 | 2180 | Dbab | 1 | 101 | | | |
| 優 | 1250 | VH11 | 1 | 201 | | | |
| 閒 | 2261 | VH11 | 1 | 202 | | | |
| 地 | 4190 | DE | 1 | 101 | | | |
| 揮 | 1334 | VA4 | 1 | 201 | | | |
| 桿 | 3122 | VA4 | 1 | 202 | | | |
| 高 | 1086 | Ncb | 4 | 501 | 401 | 301 | 000 |
| 爾 | 3407 | Ncb | 4 | 502 | 402 | 302 | 000 |
| 夫 | 1215 | Ncb | 4 | 503 | 403 | 303 | 000 |
| 球 | 2252 | Ncb | 4 | 504 | 404 | 000 | 201 |
| 場 | 3151 | Ncb | 4 | 505 | 000 | 000 | 202 |
| ， | 6001 | PM | 1 | 101 | | | |

和原來格式不同的是，新增了一個欄位來標示此段文字可以構成多少個詞，以「高爾夫球場」為例，第四行的 4 代表有四個詞，分別對應到一個五字詞「高爾夫球場」、一個四字詞「高爾夫球」、一個三字詞「高爾夫」、一個二字詞「球場」。整個語料庫的文字內容經過我們如此的整理之後，整體的詞數有少許上升，變動後的數據如下。

**表 2-2-1：標記短詞組合後之語料庫詞長分佈表**

| 詞長 | 原始個數 | 標記短詞組合而增加的個數 |
|---|---|---|
| 一字詞 | 25,244 | 0 |
| 二字詞 | 36,482 | 653 |
| 三字詞 | 5,461 | 151 |
| 四字詞 | 1,356 | 15 |
| 五字詞 | 262 | 0 |
| 六字詞 | 117 | 0 |
| 七字詞 | 71 | 0 |
| 八字詞 | 48 | 0 |
| 九字詞 | 7 | 0 |
| 十字詞 | 4 | 0 |
| 十一字詞 | 5 | 0 |
| 十二字詞 | 2 | 0 |
| 十三字詞 | 1 | 0 |
| 十四字詞 | 2 | 0 |
| 總數 | 69,062 | 819 |

### 2.2.3 標記詞綴

在之後的合成過程中，合成單元必須至少是一個詞。然而此種原則並不適合在帶有詞綴的詞之合成上。假設我們要合成「台灣人」(「人」是詞綴)，然而我們沒有「台灣人」這個詞，此時系統會採用其他替代方式合成，如以三個單字詞連接或者以傳統 PSOLA 方式合成。然而如果我們有「台灣」這個詞以及「大陸人」(「人」是詞綴）這個詞，此時利用「台灣」與「大陸人」的「人」此兩聲音片段來合成，會比替代方式來的好。為了此一目的，我們需要標記出語料庫中哪些詞的哪些字是詞綴。

首先，利用文字分析器中的前詞綴清單和後詞綴清單將語料庫中三字詞以上可能是帶有詞綴的詞列出，利用帶有詞綴的詞去掉詞綴之後亦是一個完整詞的特性（「台灣人」是一個詞，「台灣」也是一個詞），將不合此特性之詞去除。最後再以人工檢查，此人工檢查主要目的是去除定量複合詞，定量複合詞之後要特別處理。標記詞綴後的文件格式如下：

```
才 2052 Dd    1 101
使 3003 VL4   1 101
郝 3088 Nbc   1 101
內 4072 Nac   1 201
閣 2040 Nac   1 202
與 3216 Caa   1 101
總 3377 Nac   2 301 201
統 3384 Nac   2 302 202
派 4062 Nac   2 303 002
的 5043 DE    1 101
意 4186 Nad   1 201
圖 2209 Nad   1 202
破 4029 VH11  1 201
碎 4331 VH11  1 202
， 6001 PM    1 101
```

請看上例「總統派」的部分，第三行的 2 代表此部分有兩種組成方式，其一是一個三字詞「總統派」，另一個是一個二字詞「總統」帶一個後詞綴「派」。002 是後詞綴的代碼，而 001 是前詞綴的代碼。藉由這種方式，我們將整個語料庫標記了 283 個（26 種）前詞綴以及 2,296 個（83 種）後詞綴（附錄三、詞綴清單與統計數據），至於標記詞綴所增加的詞數及最後語料庫的詞數請參考下表。

**表 2-2-2：標記詞綴後之語料庫詞長分佈表**

| 詞長 | 原始個數 | 標記短詞組合而增加的個數 | 標記詞綴而增加的個數 | 最後個數 |
|---|---|---|---|---|
| 一字詞 | 25,244 | 0 | 0 | 25,244 |
| 二字詞 | 36,482 | 653 | 1697 | 38,832 |
| 三字詞 | 5,461 | 151 | 93 | 5,705 |
| 四字詞 | 1,356 | 15 | 12 | 1,383 |
| 五字詞 | 262 | 0 | 0 | 262 |
| 六字詞 | 117 | 0 | 0 | 117 |
| 七字詞 | 71 | 0 | 0 | 71 |
| 八字詞 | 48 | 0 | 0 | 48 |
| 九字詞 | 7 | 0 | 0 | 7 |
| 十字詞 | 4 | 0 | 0 | 4 |
| 十一字詞 | 5 | 0 | 0 | 5 |
| 十二字詞 | 2 | 0 | 0 | 2 |
| 十三字詞 | 1 | 0 | 0 | 1 |
| 十四字詞 | 2 | 0 | 0 | 2 |
| 總數 | 69,062 | 819 | 1802 | 71,683 |

### 2.2.4 標記中文姓氏

和詞綴類似的情況是中文姓氏，中文姓氏的性質類似於前詞綴。當輸入句含有一個中文姓名時，我們一樣也希望其姓氏可以由語料庫內的姓氏來合成。我們將語料庫中所有的姓氏其文章位置特別標記在一份清單中，並且以此清單的資料作為一個姓氏快速查詢表格。如此，當系統需要知道某個姓氏是否存在於語料庫中時，便可由此表格找尋到位置。在語料庫中，共發現 645 個（69 種）中文姓氏，詳情請看附錄五，中文姓氏清單與統計數據。

### 2.2.5 定量複合詞的特別處理

定量複合詞的麻煩之處在於其無限的組合方式。對之後以詞為主的搜尋法則而言，不同的 BIG5 內容的詞即為不同的搜尋目標。舉例而言，「一千二百三十四」與「一千二百三十五」僅僅只差異一個字，卻被系統認為是完全不同的詞。我們希望此兩者可以有一定程度的相關，且提高語料庫中定量複合詞的使用率，於是將數詞拆為更小的單位。分解的基準選擇為「十進位的單位」，如十、百、千、萬、億及其組合。如「一億兩千三百四十五」將被拆解為「一億」、「兩千」、「三百」、「四十」、「四十五」。因為我們不希望在拆解的過程中製造出單字詞，於是最後的「五」將會與前面的單位結合為「四十五」。除了數字部分外，定量複合詞還包括它的單位部分，單位有可能是一個字，如「人」、「位」、「個」等，也有可能是二字詞以上，如「小時」、「分鐘」、「美元」、「公分」等。對於一個字的單位，我們以對待後詞綴的方式處理，至於二字詞以上的單位，我們以短詞組合的方式處理。我們將語料庫中所有的定量複合詞，都經過以上的方式處理後，一旦輸入文句中含有定量複合詞，我們便可以以較小的片段將之合成出波形。

## 2.3 語音合成系統之語音參數資料庫建置

在建立了大型語料庫之文字資料庫與音檔後，為了之後挑選合成單元之用，我們需要更多的語音特性以供挑選機制作為挑選合成單元之依據。在此小節中，將介紹我們是如何由原始的文字檔與音檔求取出各類語音特性參數。

### 2.3.1 切割資訊的求取

在經過上述的處理之後，我們已有了最原始的語料庫資訊：文字內容與聲檔。然而實際用於合成的語料庫，需要更多的資訊，如標示每處音節所在位置的切割資訊，此節即是要說明如何產生此一訊息。

我們所使用的軟體為 HTK (Hidden Markov Model Toolkit)，而我們所採用的訓練模型方法在 HTK 說明手冊[4]中稱為“Isolated Word Style Training”，此一名詞的原始定義請見 HTK 說明手冊第二章第三節第二小節“Training Tools”。此訓練流程簡述如下：

我們先使用從 TCC300 語料庫訓練得來的模型，對所有語料做一次強制切割(forced alignment)，以此資訊作為“Isolated Word Style Training”中訓練初始模型的原始切割資訊，之後再以 HTK 所提供的訓練工具，重新訓練過，其中關於參數的設定為：38 維的參數，包含 12 階的梅爾倒頻譜參數(Mel-frequency cepstral coefficients, MFCCs)與能量對數值(log energy)，及其一階微分與二階微分，扣除原本的能量對數值後共 38 維；其音框大小(frame size)設為 32ms；音框位移(frame rate)設為 5ms。詳細的流程請見下二圖：

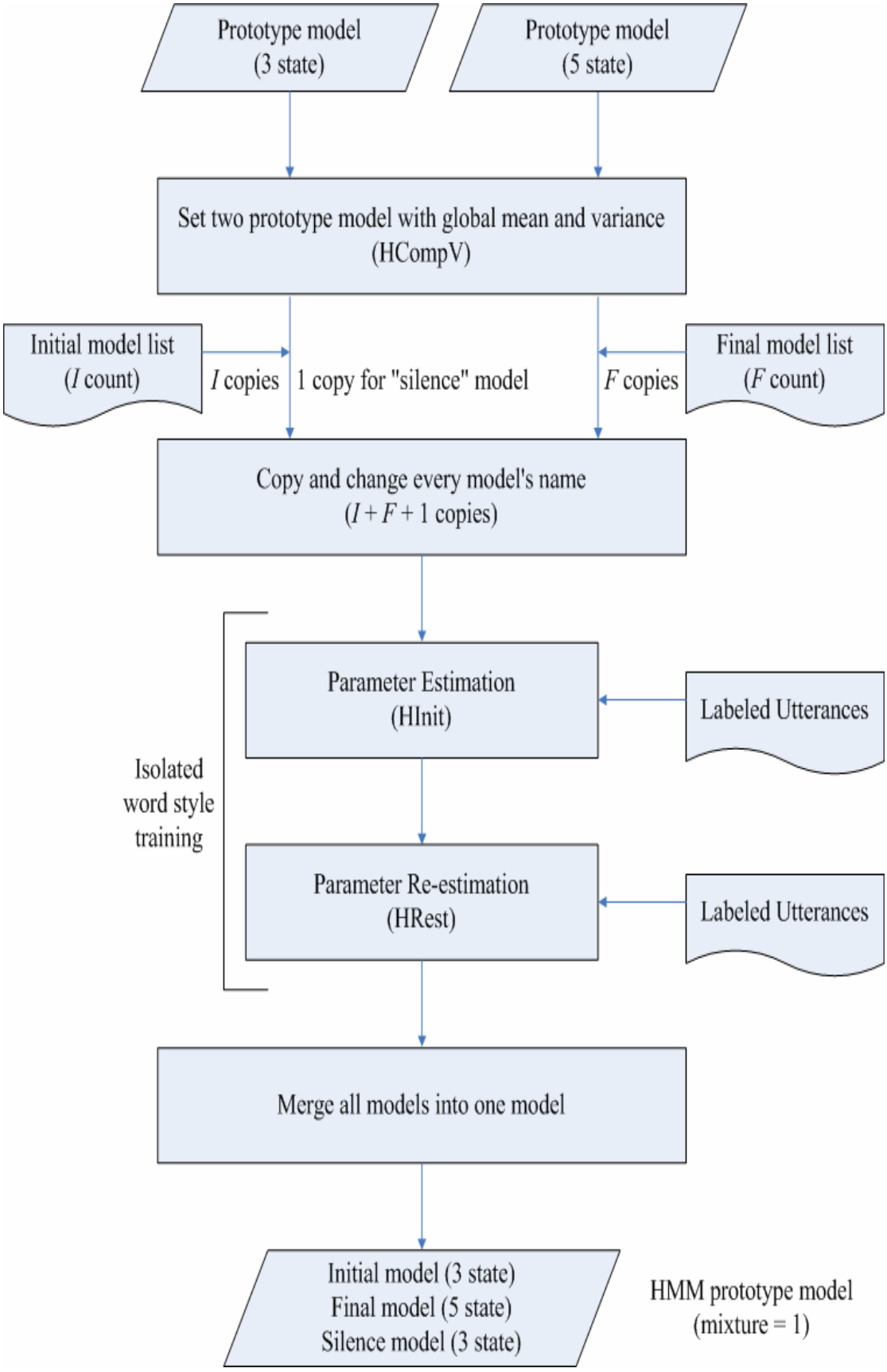



**圖 2-3-1：The flowchart of creating a HMM prototype model**

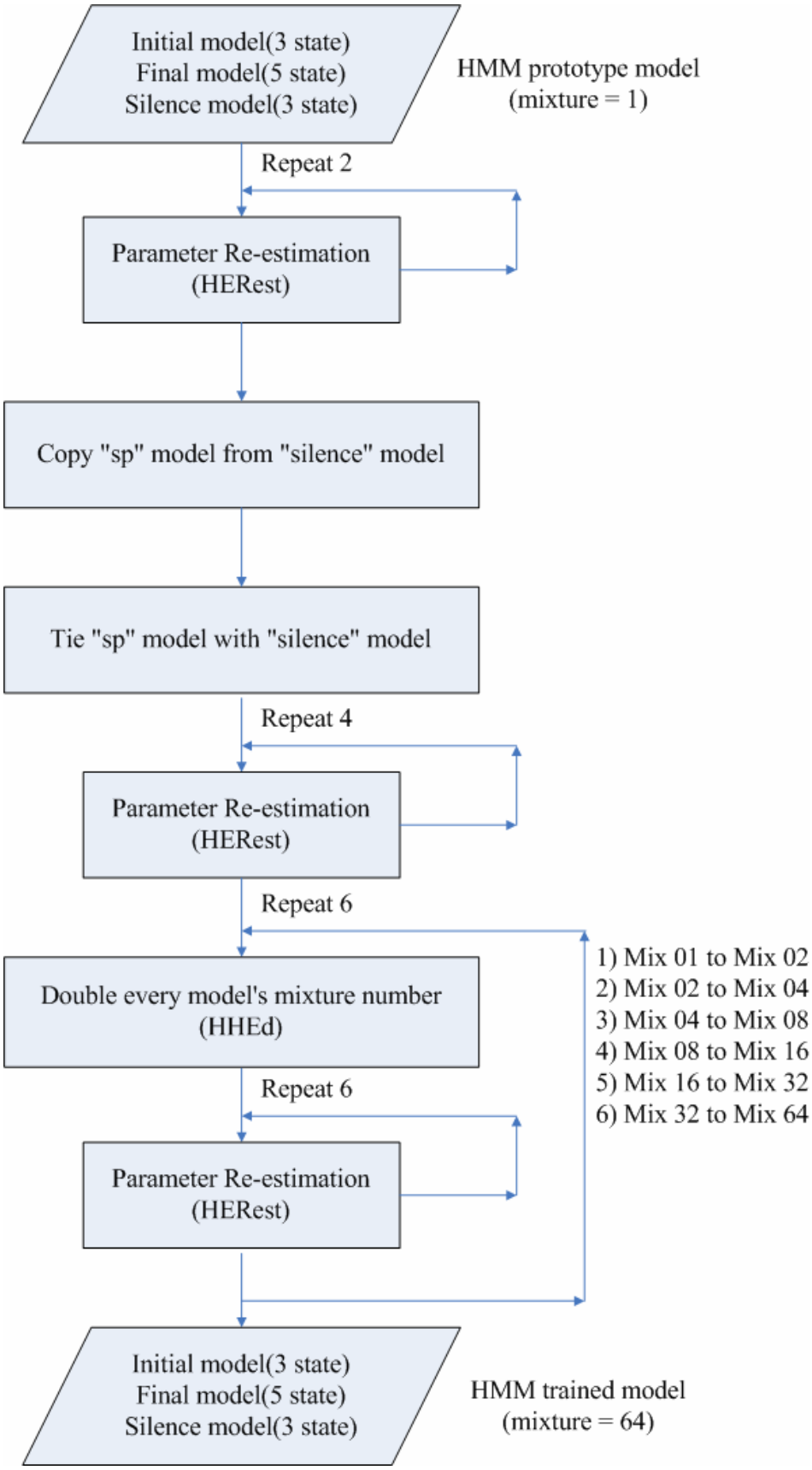



**圖 2-3-2：The flowchart of training a HMM model**

### 2.3.2 切割資訊的修正

使用 HTK 軟體所得到的音節邊界位置，不適合直接用於合成系統中，其中兩個原因是，一為音節的位置有所偏差，二為 HTK 所提供的切割位置，為以音框表示的位置，和之後合成系統以樣本表示的位置單位有所不同，接下來針對此二點對切割位置做些調整。

先討論音節位置偏差問題，請看下圖。假設粗黑實線是 HTK 所提供的切割位置，然而粗黑實線所在位置的前後兩個音框，其實有一重疊的部分。為了公平起見，此兩個音框應平分此重疊部分，如下圖粗黑虛線所示。我們將所有的切割資訊皆做此調整，發現誤差的確有所減少，可見得此調整是必須的。

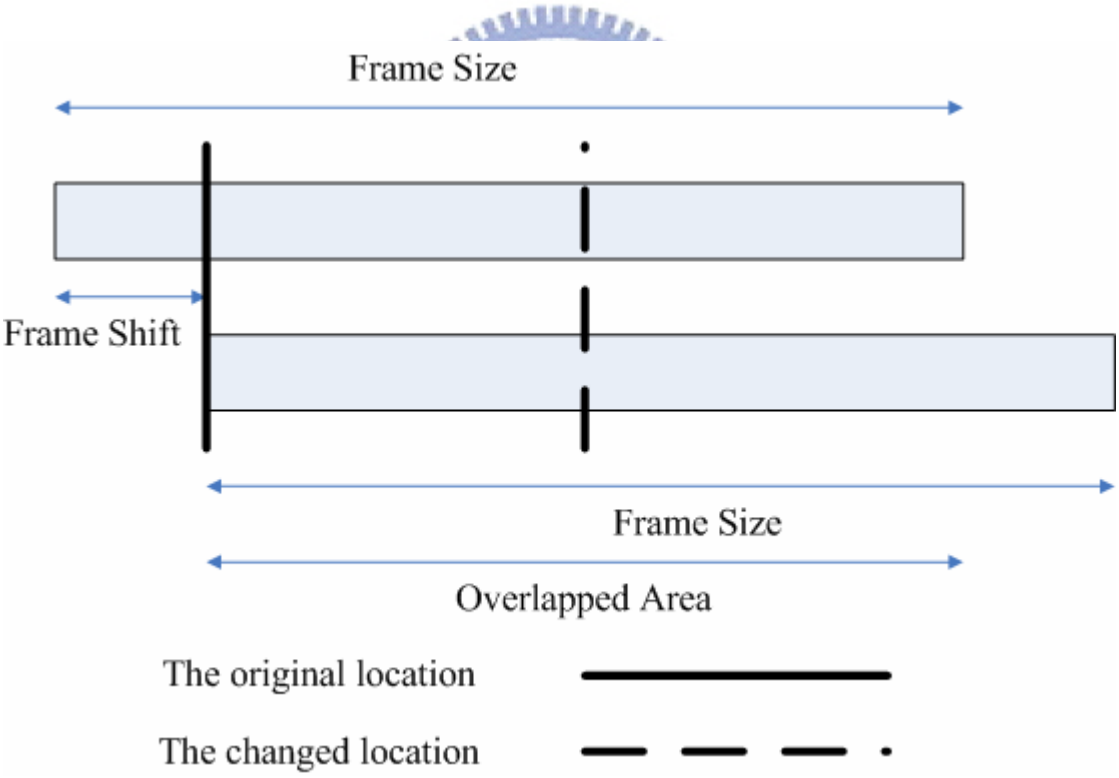



**圖 2-3-3：相鄰音框重疊部分示意圖**

之後合成系統處理聲檔是以樣本(sample)為長度單位在處理，因此我們應當以樣本為長度單位表示切割資訊。此轉換的方法在前人的論文中提過[5]，在此

簡述一次。以 HTK 所提供的切割位置為中心，前後各取一個音框位移的長度，在此兩個音框位移的長度中，以每 5 個樣本點為一個子音框(sub-frame)，計算出每一個子音框語音訊號樣本振幅絕對值總和，尋找此值最小的子音框。在此子音框中的樣本點，樣本振幅絕對值最小者即是我們最後決定切割位置的樣本點。

在整個語料庫中，總共有 123,128 個音節，其中有 52,192 個音節的切割位置曾經以人工方式標記之。我們將上述修正前和修正後的切割位置與此人工標記的位置做比較，以人工標記的位置作為標準答案，並以下式計算平均的差異量，

$$ErrorMean = \frac{\sum_{i=1}^{N} \left| L_{i_{test}} - L_{i_{ref}} \right|}{N} \qquad (2\text{-}3\text{-}1)$$

$L_{i_{test}}$ : The compared location,
represented in original HTK format ($10^{-7}$sec).
$L_{i_{ref}}$ : The reference location.
$N$: The total count for the comparison.

修正前的平均誤差為 33.74ms，修正後的平均誤差為 28.68ms，可見得此修正有助於改善切割資訊。

### 2.3.3 求取語料庫的能量資訊

在之後的合成系統中，我們需要語料庫中更多的語音特性。在這一節中，說明其能量大小的求取方法。首先，我們依照切割資訊將每段音節的波形取出，在此波形中，依照下式將每一個音框的能量求出，如果此音節有多個音框，以平均能量來代表此一音節的能量。如將一句的音節能量平均，則在我們所使用的語料庫中，此數值介於 52dB 至 66dB 之間，平均值為 60.81dB。

$$P_x(m) = 10\log_{10}\left(\frac{1}{N}\sum_{n=1}^{N}\left|w_n \times f_x(n;m)\right|^2\right)$$

$m$ : The frame index.

$N$ : The total samples of a frame. (2-3-2)

$w_n$ : The $n$-th value of the Hamming window.

$f_x(n;m)$ : The magnitude of the $n$-th sample in the $m$-th frame.

### 2.3.4 求取語料庫的音高軌跡資訊

在語音的特性參數中，韻律訊息扮演了重要的角色，其中音高(pitch)的變化是我們所關注的一項議題，此節說明此資訊的求取方式。首先，我們利用WaveSurfer[6]軟體所提供的 ESPS 演算法，求出每段音檔的音高軌跡，下圖為某個音檔的音高軌跡。

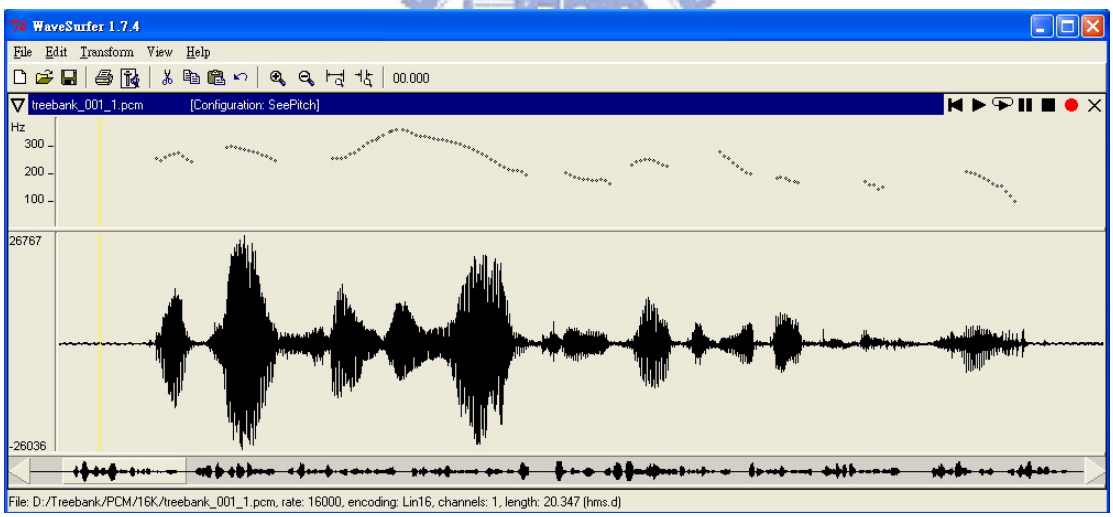



**圖 2-3-4：一個以 WaveSurfer 軟體求取音高的例子**

我們將此音高軌跡藉由 WaveSurfer 軟體提供的功能存成文字檔。然而此資料是以音框為單位，每個音框有一個音高數據。為了將此資料轉變為每段音節一組數據，我們藉由前人所提出的轉換式[7]，可將一段連續的音高軌跡轉化為四個正交參數表示，轉換方法詳述如下：

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：

$$a_j = \frac{1}{N+1}\sum_{i=0}^{N} Pitch(i)\bullet\Phi_j(\frac{i}{N})$$

$a_j$ : The orthogonal expansion parameter, $0 \le j \le 3$.

$Pitch(i)$ : The original pitch contour, $0 \le i \le N$. (2-3-3)

$N+1$ : The length of the original pitch contour.

$\Phi_j(\frac{i}{N})$ : The $j$th basis function

其中正交展開之基底函數 (basis function)的定義如下：

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

$$\Phi_1(\frac{i}{N})=\left[\frac{12N}{(N+2)}\right]^{1/2}\left[\left(\frac{i}{N}\right)-\frac{1}{2}\right] \quad (2\text{-}3\text{-}4)$$

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

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

如果我們知道的是四個正交參數，也可藉由下式重建音高軌跡，其中變數的定義與前兩式相同。

$$Pitch(i) = \sum_{j=0}^{3} a_j\bullet\Phi_j(\frac{i}{N}),\ 0 \le i \le N \quad (2\text{-}3\text{-}5)$$

藉由上述的轉換式，每段音節的音高變化皆被表示為四個正交參數，如此統一的格式就更容易被使用於合成系統中。

## 2.4 語音合成系統資料庫建置總結

在此章中，我們論述了一個作為語音合成系統語料庫所應具備的條件，並介紹了我們系統中所使用的語料庫。作為系統應用之用，語料庫除了原始的音檔之外，我們特地對語料庫的文字資料作了許多的加工，並且事先對於原始音檔建立了語音參數資料庫，以供快速使用之便。在次一章節中，將實際介紹系統架構，並且可以看出系統是如何利用此些資料庫，以達快速且高效能的語音合成系統。

# 第三章 以語料庫為基礎之語音合成系統架構

先前的語音合成系統，大多是以少數的基本音節經過 PSOLA 合成出連續語音，雖然在此技術上我們已有不錯的成果[8]，但是如果改成以語料庫為基礎的語音合成系統，在語音的自然度與韻律的流暢度上皆可大幅改善。在此章中，將試著說明發展以語料庫為基礎的合成系統之構想與系統架構，並詳細敘述每一部分的實作細節。

## 3.1 構想流程圖與系統架構

與先前以少量基本音節為資料庫的合成系統所不同的是，以大量語料庫為基礎的語音合成系統其重點有三點：

1. 如何設計一個適合用於語音合成系統的大型語料庫。
2. 如何在此大型語料庫中搜尋到可用來合成的語音片段。
3. 如何在可用來合成的語音片段當中，挑選出一個最佳的組合來合成。

關於第一點語料庫的設計問題，已在第二章中介紹過。在此小節中，我們將以第二點與第三點作為本章系統架構的出發點。

### 3.1.1 以語料庫為基礎之語音合成系統構想流程圖

以語料庫為基礎的語音合成系統流程圖如下圖所示，輸入為欲合成的文句，輸出為合成的語音，中間每塊方塊代表合成系統中的一個模組，右方的圓柱形代表不同階段會使用到的資料庫。整個流程圖，在此簡述一遍。當使用者輸入文句

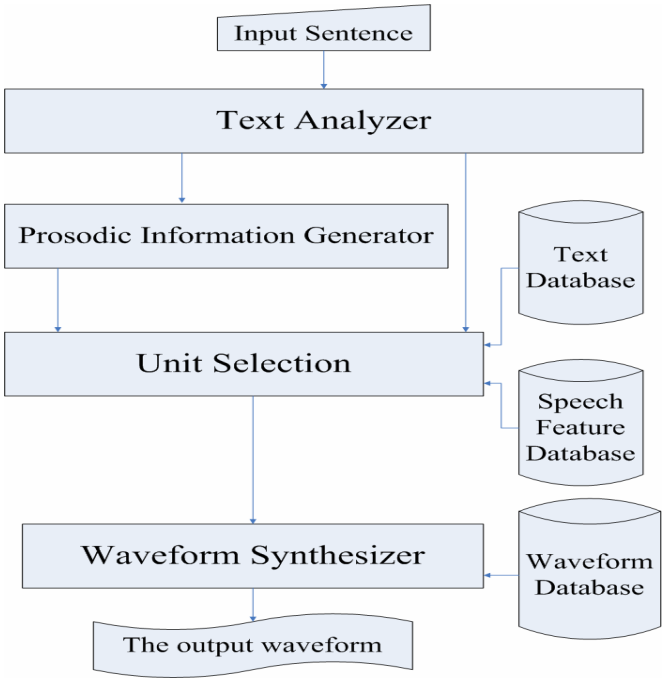



**圖 3-1-1：語音合成系統流程圖**

後，此文句首先被文字分析器(Text Analyzer)處理，文字分析器會將文字做正規化，並產生相對應的音碼串 (syllable code sequence)、詞串(word sequence)、詞類串(POS sequence)。這些資訊會分別的被送到韻律訊息產生器(Prosodic Information Generator)與合成單元選擇器(Unit Selection)中。韻律訊息產生器接收到這些資訊後，會去預測此輸入文句的韻律訊息；而合成單元選擇器之工作可分為兩階段，第一階段從語料庫的文字資料庫(Text Database)中尋找所有可能可以合成此輸入文句的文字片段。之後第二階段，藉由預測的韻律訊息與語料庫的語音參數資料庫(Speech Feature Database)，由前階段的搜尋結果中挑選出一個最佳的組合。最後，波形合成器(Waveform synthesizer)以此最佳的組合與語料庫的波形資料庫(Waveform Database)合成出最後的輸出波形

### 3.1.2 以語料庫為基礎之語音合成系統架構圖

在上一節中，我們以流程圖的方式介紹了整個系統。現在讓我們改成以系統

架構的觀點來考慮各模組間的關係。由於目前的系統中文字分析器與韻律訊息產生器皆是承襲前人的成果[9,10]，在之後的介紹中，只著重於說明與之前的效能不同之處，而專為以語料庫為基礎的語音合成系統而出現的合成單元選擇器（搜尋與挑選）部分，將會詳細描述。

合成系統分成四模組，「文字分析器」、「韻律訊息產生器」、「合成單元選擇器」、及「波形合成器」。其中「合成單元選擇器」的部分比較複雜，又分為「搜尋單元」和「挑選單元」。下表列出各模組所負責的功能及用途，在接下來的幾節中，這四模組將會一一解釋之。

**表 3-1-1：系統中各模組之功能一覽表**

| ***Module Name*** | ***Functions and purposes*** |
|---|---|
| **Text Analyzer** | Transform BIG5 string into phonetic transcription annotated with high-level linguistic description. |
| **Prosodic Information Generator** | Predict the prosody of the input text by the high-level linguistic description. |
| **Unit Selection** | Search all possible candidates of the target sentence in the corpus. And then select the best combination in the search result. |
| **Waveform Synthesizer** | Process some waveform adaptation and concatenate the specified waveforms. |

## 3.2 文字分析器說明

文字分析器的主要作用在於將使用者輸入的 BIG5 字串，轉換為以聲碼表示的串列。在轉換的同時，也會產生許多語言參數，如詞串，POS 串等，[9]中對文字分析器的實現方式，有詳細的說明，在此不再贅述。然而為了以語料庫為基礎的語音合成系統，我們將文字分析器開發了新的功能，現在的文字分析器能標記輸入文句中之詞綴與姓氏，這對之後將要提及的搜尋單元有極大的助益。

## 3.3 韻律訊息產生器說明

韻律訊息產生器的主要作用在於模擬語者的說話習慣，以猜測輸入文句的韻律訊息。韻律訊息產生器的實現方式主要可分為兩種，以規則法實現，或者是以模型實現[11]。我們採取使用遞迴式類神經網路來實現韻律訊息產生器[10]，以系統所採用的中研院中文文句結構樹資料庫以及之前所錄之音檔來訓練此韻律訊息產生器。下表為此韻律模型之訓練結果，觀察訓練後之韻律訊息產生器所重造的韻律訊息，確實貼近原始的韻律訊息。

**表 3-3-1：The RMSEs of the five synthesized prosody parameters.**

| | **Inside Test** | **Outside** |
|---|---|---|
| **F0 Contour** | 1.035ms/Frame | 1.125ms/Frame |
| **Pause Duration** | 30.29ms | 46.6ms |
| **Initial Duration** | 20.2ms | 21.1ms |
| **Final Duration** | 33.7ms | 35.9ms |
| **Energy Level** | 3.66dB | 4.2dB |

## 3.4 搜尋單元說明

### 3.4.1 搜尋單元功能說明

在討論應用於系統的搜尋法之前，有件事必須先釐清。因為我們認為「詞」是人類說話有意義的最小單元，所以我們也希望在系統中所尋找到的合成單元儘量是一個詞或是一串詞組，也就是說就算某幾個字是目標句的一部份，但如果它們並不是一個詞或是一串詞組的話，我們也不希望將它們當作是一個合成單元。

搜尋單元的目的在於將語料庫中所有可以湊成目標句的文字詞組片段找

出。然而如果我們真的以目標句的文句直接至語料庫中的文字資料中尋找，未免過於費時費力。在此我們使用前人所提出的“連續相關比對法”[1]，以下說明之。假設我們的目標句是「交通大學的語音處理實驗室」。首先，至語料庫中的文字資料庫中找出所有的「交」的所在，因為這是目標句的第一個字，所以將不是詞的開頭的「交」去除，將剩下符合條件的「交」放入稱之為「工作表格」(Working Table, WT)的暫存區中，這些表格所存放的資料格式將在後續章節詳述。如此目前「交」這個字的搜尋步驟便完成了，此時要將「工作表格」中可構成詞而且不可能再增加的項目移至最後搜尋結果的儲存區中，稱之為「詞串候選表格」(Word Sequence Candidate Table, WSCT)。接下來換成搜尋「通」這個字，至語料庫中的文字資料庫中找出所有的「通」的所在，對這些「通」的項目而言，有三種狀況：

**表 3-4-1：連續相關比對法中，搜尋過程的三種狀況及對應動作**

| 狀況 | 對應的動作 |
|---|---|
| 1.前一個字存在「工作表格」中。 | 將此字的紀錄加入「工作表格」中對應的紀錄。 |
| 2. 前一個字不存在「工作表格」中，但此字為一個詞的開始。 | 在「工作表格」新增一個項目來記錄此字。 |
| 3. 前一個字不存在「工作表格」中，且此字不是一個詞的開始。 | 忽略此記錄。 |

當所有「通」的紀錄都以這三條規則處理完後，「通」的搜尋也結束了，一樣的，將「工作表格」中可構成詞而且不可能再增加的項目移至「詞串候選表格」。然後開始下一個字「大」的搜尋，如此反覆搜尋直到目標句的最後一個字。以上就是“連續相關比對法”的基本原理，在下一節中，將會說明這些表格的資料結構，並提出一個方法來加快此一過程。

### 3.4.2 搜尋單元實作

上一節當中雖然提出了一個可行的搜尋法，然而實際運作起來不免過於費時，因此我們在這一小節中提出一些作法及資料格式來加快此一搜尋過程。

#### 3.4.2.1 字元位置表格(Character Location Table)實作

如果我們能有一個表格，輸入是一個字元，而輸出是所有此字元在此語料庫文字資料庫的位置的話，那麼對於上一章節所提出的搜尋法必定可以有所助益。因此，我們設計如此的表格，其格式如下表所示：

**表 3-4-2：字元位置表格格式**

| | |
|---|---|
| 依 | <1,1,201>，... |
| 據 | <1,2,202>，... |
| 行 | <1,3,301>，... |
| 政 | <1,4,302>，... |

其中，中文字是每項資料的索引，而<x,y,zt>代表一個位置，x 代表第幾句，y 代表第幾個字，zt 代表幾字詞的第幾個字。譬如說「政<1,4,302>」代表語料庫的第一句的第四字是「政」，而且它是在一個三字詞的第二字。如此，我們將語料庫中所有種類的中文字都建立起此表格。在建立表格的同時，也統計了整個語料庫出現的中文字種類，總共有 3,137 種類型的中文字。為了加快讀取此表格的速度，我們將做為索引的中文字做遞增排序，並依據每字所擁有的資料個數，做成一個檔頭資訊，如此我們就可以使用「二元搜尋法」來讀取此表格，而非傳統的線性搜尋。

### 3.4.2.2 工作表格(working table)實作

如上節所述，「工作表格」是為了儲存暫時的搜尋結果，以待將來放入「詞串候選表格」中。設計的表格格式如下：

**表 3-4-3：工作表格格式**

| 句子標號 | 前一個字元在目標句中的位置 | 在目標句起始位置 | 在原句中起始位置 | 已完成的詞串組合詞長 |
|---|---|---|---|---|
| 5 | 4 | 1 | 8 | 2,2 |
| 23 | 3 | 2 | 5 | 2 |
| … | … | … | … | … |

舉例說明，假設目標句是「想要知道交通大學的基本介紹」，目前的搜尋字是「交」，那麼如果語料庫中的第五句是「上｜高速高路｜前｜我｜想要｜知道｜交通｜狀況」(｜代表詞邊界)，則「工作區間」就會有一個如上表第一行一樣的紀錄。

| | | | | |
|---|---|---|---|---|
| 5 | 4 | 1 | 8 | 2,2 |

代表語料庫的第 5 句從第 8 個字開始，和目標句從第 1 個字開始相同至第 4 個字，而詞長組合是二字詞、二字詞。

### 3.4.2.3 詞串候選表格(Word Sequence Candidate Table, WSCT)實作

詞串候選表格的目的在於將搜尋結果的最長詞串儲存下來，所設計的格式如下：

**表 3-4-4：詞串候選表格格式**

| 句子標號 | 在目標句起始位置 | 在句中起始位置 | 詞串組合詞長 |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 1 | 1 | 4 | 2 |
| 2 | 5 | 1 | 2,2 |
| 2 | 7 | 10 | 2 |
| 3 | 9 | 9 | 1,2 |
| 4 | 12 | 7 | 2 |

舉例說明，表格的第一行表示在語料庫的第一句，「我｜不｜大｜想要 |結婚」（｜代表詞邊界）中的第 4 個字開始，與目標句「想要知道交通大學的基本介紹」中的第 1 個字開始的一個 2 字詞互相符合。

### 3.4.3 搜尋單元改良

在前面所提出的搜尋法中，有一個效率上的缺點，那就是當某個字出現的次數很多時，每次遇到此字就必須載入許多資料。而且如果此字又常以單字詞出現，那麼它就會留在最後的結果，然而其實這些單字詞的紀錄對合成的音檔的影響不大，其原因留待「挑選單元」的章節解釋。為了這個緣故，我們特地將這些單字詞從「字元位置表格」中獨立出來，成為「單字詞字元位置表格」(Mono-Syllabic-Word Character Location Table, Mono-SWCLT)。其使用的時機在於，直到此字的前後文句被搜尋到時，才將單字詞附加上去。而去除單字詞後的「字元位置表格」，則稱之為「多字詞字元位置表格」(Multi-Syllabic-Word Character Location Table, Multi-SWCLT)。

#### 3.4.3.1 單字詞字元位置表格(Mono-Syllabic-Word Character Location Table, Mono-SWCLT)實作

和前面所敘述的「字元位置表格」不同的是，「字元位置表格」的索引值是中文 BIG5 碼，而現在「單字詞字元位置表格」需要儲存的資訊是在語料庫中某

個位置是否是一個單字詞，如果是的話，又需要知道是什麼字。所以「單字詞字元位置表格」的索引值是語料庫的句子編號，而儲存的內容是此句中出現的單字詞與其所對應的位置。

按照這些要求，「單字詞字元位置表格」設計的格式如下：

**表 3-4-5：單字詞字元位置表格格式**

| 句子編號 | 在句中出現的單字詞 | 單字詞在句中出現的位置 |
|---|---|---|
| 1 | 的, 一, 到, … | 9, 16, 17, … |
| 2 | 一, 到, 的, 為, … | 12, 13, 17, 22, … |
| … | … | … |

以表格中的第一列為例，意思是說，語料庫的第一句的第 9 個字是單字詞「的」，第 16 個字是單字詞「一」，第 17 個字是單字詞「到」，依此類推。此表格也按照之前的方式，使用「二元搜尋法」來加快讀取速度。

### 3.4.3.2 單字詞額外比對

當我們將單字詞獨立成另一個表格時，我們也需要改變我們的搜尋法。改變兩個地方，一、當在「工作表格」新增一個項目時，必須至「單字詞字元位置表格」中檢查前面的字是否為單字詞(The backward check);二、必須在讀完「多字詞字元位置表格」後，再對未找到這個字的「工作表格」項目至「單字詞字元位置表格」中搜尋此位置是否為單字詞(The forward check)。在此，將“連續相關比對法”以流程圖表示如下。

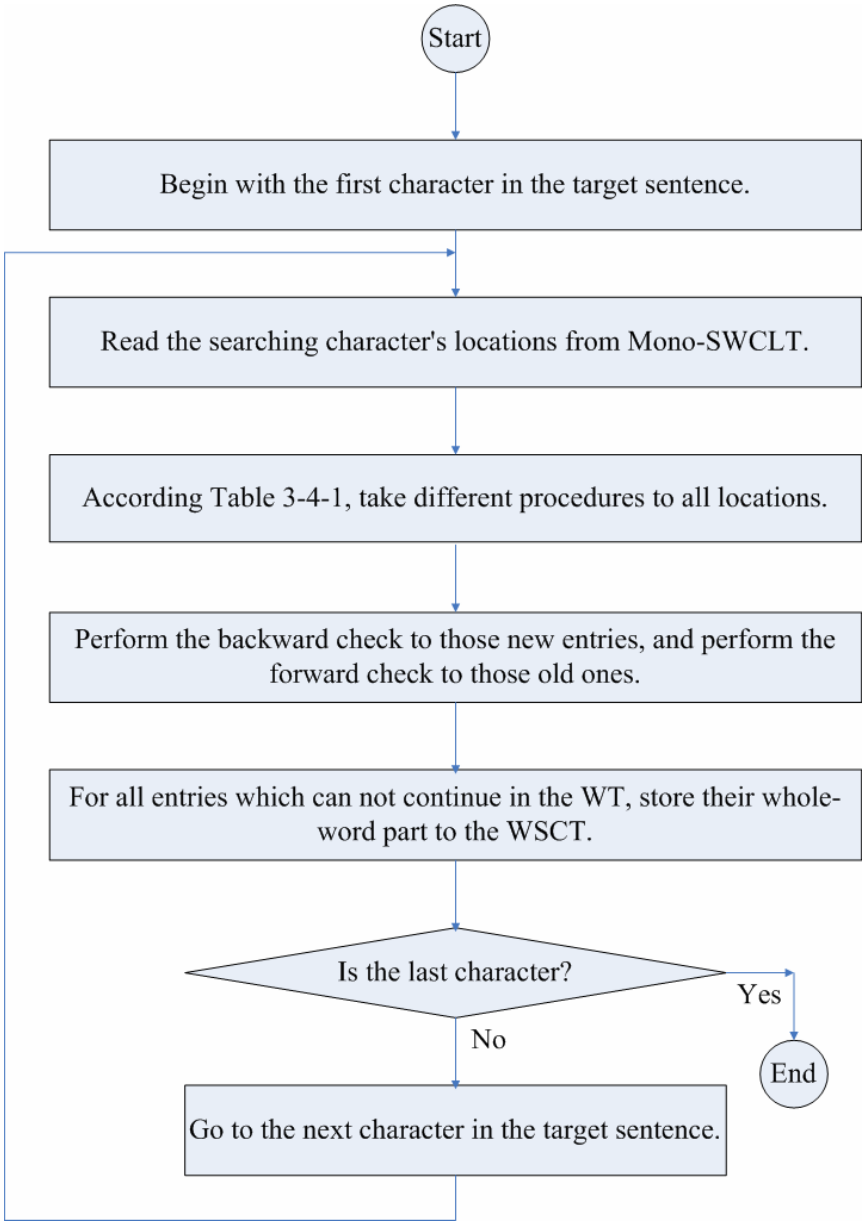



**圖 3-4-1：連續相關比對法流程圖**

### 3.4.4 彌補語料庫未出現的中文字

在上一節中，我們介紹了如何使用連續相關比對法得到以多字詞組成的搜尋結果。然而由上一小節所得到的搜尋結果是不完整的，因為目標句可能在某些部分完全沒有搜尋到目標，此節即是要討論此一問題。

### 3.4.4.1 如何判斷多字詞搜尋結果需要填補哪些位置

在開始敘述我們對於未搜尋到的目標句段落的對策之前，先解釋如何定義目標句缺少的部分。明顯的，如果目標句中的某個位置，在上述的搜尋過程中從未被找到過，其位置必為目標句缺少的。然而如果上述的搜尋結果出現詞頭尾不對齊的狀況，為了保留每個多字詞被選擇的可能性，一些額外的填補是需要的，如下例，一個方格代表一個搜尋到的部分，圖中的數字代表此部分搜尋到幾個，下例中，目標句是「檢附身分證」，搜尋到一個「附身」、一個「身分證」與一個「身分」。

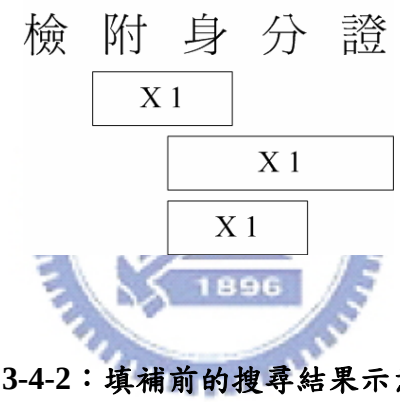



**圖 3-4-2：填補前的搜尋結果示意圖**

上例中，目標句的第一字「檢」沒有出現在任何多字詞中，可直接認定為需要填補的位子。接下來逐一檢查每個多字詞的頭尾是否有其他候選合成單元可以與之連接。首先，「附身」左方沒有任何候選合成單元，「檢」應被認為是需要填補的位置，同樣的，「附身」右方一樣沒有單元可以連接，「分」與「證」也需要填補。依照同樣的方式檢查「身分」與「身分證」，可以發現「附」也需要填補。目標句每處位置都有多字詞被搜尋到，然而有些多字詞頭尾並不相連。以「大學的」為例，其前方並無與任何多字詞相連，同樣的情況也發生在「處理」這個詞的前方與「實驗」的後方。如果我們將需要填補的目標句位置定義為完全沒有出現在多字詞的搜尋結果中，則此例僅需填補「檢」。但這樣沒有任何一個組合可

以涵蓋整個目標句。因此，為了確保每個多字詞的可能性，實行上述之檢查，發現「附」、「分」與「證」皆須填補。我們可將需要填補的目標句位置定義為「當目前的搜尋結果其前後沒有可連接的候選單元時，就採取填補動作。」如以此為原則，則上例可被填補為下圖的情況，其中紅色圓圈代表被填補的位子。至於如何進行填補的動作，由之後的數小節說明。

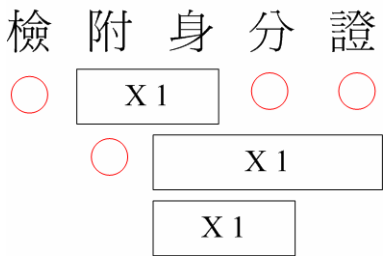



**圖 3-4-3：已標記須填補位置的搜尋結果示意圖**

### 3.4.4.2 彌補語料庫未出現的中文字，以前後詞綴填補

如之前第二章所述，有時在語料庫中我們並不能找到一個完整的詞，而是找到少一個詞綴的其他部分，譬如尋找「衛生署」，卻只能找到「衛生」。這時如果能夠藉由共用詞綴，像從「警察署」挪用後詞綴「署」，即可增加我們目標句的完整度。同樣的，在系統中也實作了「前詞綴表格」與「後詞綴表格」，當前述的搜尋方法之搜尋結果未能包括某一詞綴時(由文字分析器告知此位置是否為詞綴)，即會以記載在詞綴表格中的項目填補此一搜尋的空缺，前詞綴的空缺為前詞綴表格中相同項目填補，後詞綴的空缺為後詞綴表格中相同項目填補。

### 3.4.4.3 彌補語料庫未出現的中文字，以中文姓氏填補

當輸入文句含有中文姓名時，其姓氏的角色就類似於前詞綴。我們一樣希望

能以語料庫中的姓氏來合成此輸入文句中的姓氏。於是我們將之前第二章中所標記的中文姓氏做成「中文姓氏表格」，其實現方式類似於「多字詞字元位置表格」。如此一旦文字分析器發現輸入文句中含有中文姓名時，而此位置又被標記為需要填補的空缺，就可以以此表格中記錄的相同項目來合成。

#### 3.4.4.4 彌補語料庫未出現的中文字，以同音單字詞替代

在實現一個搜尋系統時，我們無法保證所有的搜尋目標都會存在於資料庫中。對於上兩節所提出的搜尋單元，基礎上是建立於對中文 BIG5 碼的搜尋，然而就如第二章所統計的數據顯露的訊息相同，在我們的語料庫中並未包含所有的中文字，所以以上所提出的搜尋法的搜尋結果也可能是沒有找到目標物。

對此，為了彌補語料庫中未出現的中文字，我們退而求其次試著以同音字替代。為了加快系統的效率，我們實作了一個表格來記載語料庫中同音字的位置，稱為「聲調音節位置表格」(Tonal Syllable Location Table, TSLT)。此表格的格式如同前述之「字元位置表格」，唯一不同的是其索引值為音節碼，一樣是以「二元搜尋法」來加快搜尋速度。然而，語料庫中也未包含中文聲調音節所有的可能性，語料庫中計有 1,068 種音調音節，通常認為中文音調音節有 1,300 種，此語料庫僅有 82%的涵蓋率，所以仍有可能會有缺失。此時，我們會以傳統 PSOLA 的方式來合成此字，在 3.6 節會再次說明。

### 3.4.5 搜尋單元總結

在此，為搜尋單元作一個總結。當一個目標句輸入之後，系統搜尋合成單元的順序是

1. 至多字詞字元表格尋找可用詞串。

2. 至單字詞字元表格尋找步驟 1 前後的單字詞。

3. 至前後詞綴表格填補前 2 步驟所未搜尋到的中文字。

4. 至中文姓氏表格填補前 3 步驟所未搜尋到的中文字。

5. 至音調音節位置表格填補前 4 步驟所未搜尋到的中文字。

6. 仍未填補的空缺，作一特別記號，以待之後以 PSOLA 合成。

其順序可以下圖表示之。

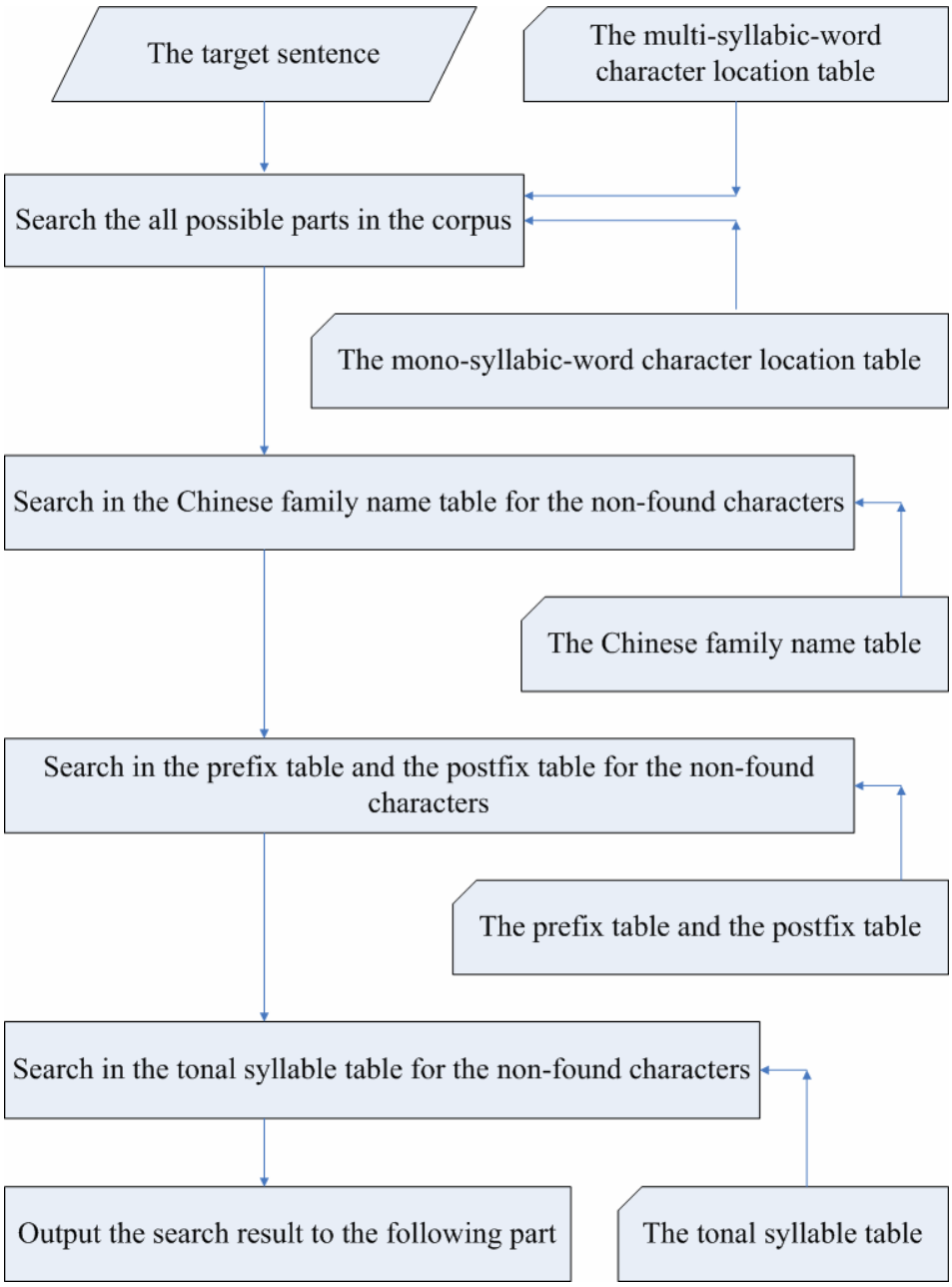



**圖 3-4-4：搜尋單元流程圖**

## 3.5 挑選單元說明

### 3.5.1 挑選單元功能說明

#### 3.5.1.1 挑選單元之目的

當我們藉由搜尋單元得到許多可以合成目標句的文句段落時，並不能隨意由其中一個組合來合成，而是應該預測每一個組合的合成品質，挑選擁有最佳合成音質的組合來合成目標句。

#### 3.5.1.2 影響合成音質的誤差因素

在[12]一書中，將影響連接型合成系統音質的因素，歸類為以下四種誤差：

1. 前後連接音歧異(Differences in phonetic contexts)

   當拿來合成的波形片段其在目標句與原本在語料庫中的前後連接音不同時，便會造成此誤差。

2. 不正確分段(Incorrect segmentation)

   即使是相同的連接關係，如果未能正確的將波形片段切割，也會造成頻譜上的不連續。

3. 聲學特性變動(Acoustic variability)

   即使去除前兩因素，聲學特性的變動亦會造成誤差，如使用不同的麥克風，說話速度快慢不一等等。

4. 韻律差異(Different prosody)

   合成單元邊緣間連接點的音高、能量等韻律參數不連續也是造成音質下降的原因。

### 3.5.2 挑選單元實作

#### 3.5.2.1 文獻回顧

從不同合成單元的組合中，挑選最恰當的一個組合的方式各家作法不同[13, 14]。不過大致上可以歸類到依據 cost function 挑選和依照決策樹(Decision Trees)決定，以下簡介此兩種方法。

1. 依照決策樹(Decision Trees)挑選

決策樹有兩種，分類樹(Classification Trees)與迴歸樹(Regression Trees)。分類樹的目的在於區分成一離散的類別數值；而迴歸樹的目的則在於分析已得到的連續型參數值。至於在挑選合成單元的應用上，決策樹主要是依據語言參數(Linguistic Features)來挑選合成單元，其中語言參數通常包括韻律參數、詞邊界、呼吸邊界等，因此較近似於迴歸樹的應用方式。

2. 依據 cost function 挑選

Cost function 在[12]一書中，被稱之為 Objective function，其目的在於建構一個數學表達式來評估連接不同音段後的聲音品質。在此方法中，把影響合成聲音品質的因素歸類到兩大項，單元誤差(Unit Cost)與轉移誤差(Transition Cost)。以下圖來解釋，

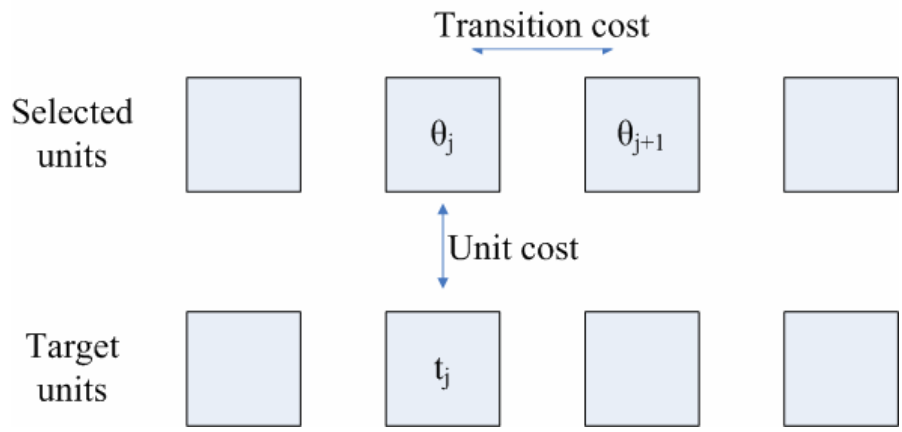



**圖 3-5-1：Tradeoff between unit and transition costs.**

當我們選定了一組合成單元組合來合成目標句時，以語料庫中的合成單元 $\theta_j$ 來取代目標句的 $t_j$，所造成的誤差為單元誤差。而將不同的合成單元如 $\theta_j$ 與 $\theta_{j+1}$ 連接起來，所造成的誤差為轉移誤差。其數學式的表達方式就會是此兩種誤差的累加，如下式，

$$d(\Theta,T)=\sum_{j=1}^{N} d_u(\Theta_j,T)+\sum_{j=1}^{N-1} d_t(\Theta_j,\Theta_{j+1})$$
$$d_u(\Theta_j,T) \text{ is the unit cost.}$$
$$d_t(\Theta_j,\Theta_{j+1}) \text{ is the transition cost.} \qquad (3\text{-}5\text{-}1)$$

而我們的目標就是要找出能將 cost function 的累加誤差降至最小的那個組合。

$$\hat{\Theta}=\arg_{\Theta}\min d(\Theta,T) \qquad (3\text{-}5\text{-}2)$$

在實現以 cost function 為挑選合成單元的系統時，最大的關鍵處在於其各個誤差的定義方式。定義這些誤差因素的方式，大致上也可以分類成兩大類，以經驗為基礎(Empirical)的定義方式，或資料處理型(Data-Driven)的定義方式。

1. 以經驗為基礎(Empirical)的定義方式

此類的系統實作者，會藉由一些聽覺上的感官實驗來瞭解各種不同類型的合成單元間，取代或者連接時，對人類聽覺的影響程度。在大量的實驗之後，決定各式各樣情況下的分數，再由其中取出分數最好的組合來合成。

2. 資料處理型(Data-Driven)的定義方式

以經驗為基礎的定義方式並不能實際表達出合成單元在語音特性上的不連續，相反地，資料處理型的定義方式藉由比較合成單元實際上的語音特性來評估合成後的音質。通常其轉移誤差會定義為前一個合成單元音段的最後一個音框之頻譜與下一個合成單元音段的第一個音框之頻譜差異的平方，數學式如下式，

$$d_t(\theta_i,\theta_j)=\left|x_i(l(\theta_i)-1)-x_j(0)\right|^2$$
$$l(\theta_i)\text{ denotes the number of frames of speech segment }\theta_i, \quad (3\text{-}5\text{-}3)$$
$$x_i(k)\text{ is the cepstrum of segment }\theta_i\text{ at frame }k.$$

但此類的定義方式只適用於合成單元處在頻譜緩慢變化的前提上，當頻譜有明顯的不連續時，應改成比較前一個合成單元音段的最後一個音框之頻譜與下一個合成單元音段的前方第一個音框之頻譜差異的平方，改良後的數學式如下式，

$$d_t(\theta_i,\theta_j)=\left|x_i(l(\theta_i)-1)-x_j(-1)\right|^2 \quad (3\text{-}5\text{-}4)$$

其單元誤差的定義方式類似於轉移誤差，只是比較的對象變成所使用的合成單元音段之頻譜與目標句預測之頻譜的差距。

### 3.5.2.2 挑選合成單元之方式

在本文章中所提及的系統其挑選合成單元的方式較類似於依據 cost function 挑選且以資料處理型(Data-Driven)的方式定義，其細節將在以下幾小節說明。

當合成系統對於合成目標句完成可用合成單元的搜尋之後，假設其中第 $i$ 種的組合為

$$S_i = U_1^i U_2^i U_3^i ... U_n^i$$

而此組合由 $U_1^i$ 、 $U_2^i$ 、…、 $U_n^i$ 等 $n$ 段候選合成單元所組成。其 cost function 數學式定義為

$$C(S_i,T) = w_{target}\sum_{k=1}^{n} C_{target}(U_k^i,T) + (1-w_{target})\sum_{k=1}^{n-1} C_{transition}(U_k^i,U_{k+1}^i) \quad (3\text{-}5\text{-}5)$$

其中 $T$ 代表預測的理想目標句，而 $C_{target}(U_k^i,T)$ 為對應於 $U_k^i$ 與目標句的合成單元目標差異(Target Cost)，意義近似於上文中提到的單元差異(Unit Cost)，而 $C_{transition}(U_k^i,U_{k+1}^i)$ 為對應於 $U_k^i$ 與 $U_{k+1}^i$ 的合成單元間轉移差異(Transition Cost)，即上文中提到的轉移差異， $w_{target}$ 則是協調此兩變數重要性的權重值。其中合成單元目標差異與合成單元間轉移差異的詳細定義，由以下章節中說明。

### 3.5.3 合成單元目標差異(Target Cost)

合成單元目標差異其主要目的在於衡量所使用的合成單元音段與相對應的目標句音段是否相似。系統實作時所考慮的因素各家不一，本系統考慮的因素為此候選合成單元之前後文相關係數與韻律參數和目標句的差異，其中韻律參數包括音節長度、音節能量、及音節平均音高(pitch mean)。合成單元目標差異之數學式為

$$C_{target}(U_k^i,T) = \sum_{j=1}^{m}\left(\frac{w_C}{Q_C} d^j_{contextual} + \frac{w_{p\_m}}{Q_{p\_m}} d^j_{pitch\_mean} + \frac{w_d}{Q_d} d^j_{duration} + \frac{w_p}{Q_p} d^j_{power}\right) \quad (3\text{-}5\text{-}6)$$

其中 $m$ 為此候選合成單元擁有之音節數， $d^j_{contextual}$ 代表此候選合成單元中第 $j$ 音節

與目標句之前後文相關係數差異，$d^{j}_{pitch\_mean}$ 代表此候選合成單元中第 $j$ 音節與目標句之平均音高差異，$d^{j}_{duration}$ 代表此候選合成單元中第 $j$ 音節與目標句之音節長度差異，$d^{j}_{power}$ 代表此候選合成單元中第 $j$ 音節與目標句之能量差異。上述變數乘上代表其影響力的權重（$w_c$ 、$w_{p\_m}$ 、$w_d$ 、$w_p$ ）與調整其數值變動範圍之正規化係數（$Q_c$ 、$Q_{p\_m}$ 、$Q_d$ 、$Q_p$ ）倒數，單元內各音節的差異總和即為整體候選合成單元與目標句之合成單元目標差異。各變數的權重值與正規化係數的設定留待第四章再予以說明，現在先說明各差異的詳細定義方式。

#### 3.5.3.1 前後文相關係數差異(Contextual Difference)

在許多關於韻律預測的研究中，都將所探討的韻律單元前後文變數視為影響此韻律單元的環境變數[15]。依據此想法，如果我們想要得到與目標句類似的韻律單元，那麼應優先選擇具有和目標句相同的前後文變數之候選單元。我們可將這些前後文變數當作分類音節的依據，這些前後文變數包括，

1. 前一音節結尾類型（Left Phonetic Context）
2. 後一音節開頭類型（Right Phonetic Context）
3. 前一音節音調（Left Tone）
4. 後一音節音調（Right Tone）
5. 位於詞中的位置（Position In Word）

由此五變數構成描述音節的「音節相關前後文變數向量」(Syllable-Dependent Contextual Vector, SDCV)，各變數的分類請參考附錄四，「音節相關前後文變數向量分類方式與統計數據」。語料庫中計有 123,128 個音節，而有 22,864 不同種類的音節相關前後文變數向量。而兩個不同音節相關前後文變數向量的差異則為

其各變數差異之權重和，如下，

$$d^{j}_{contextual} = \sum_{i=1}^{I} W_{ci} D_i$$

$d^{j}_{contextual}$: the contextual difference. (3-5-7)

$D_i$: the distance for the *i*th coordinate in the vector.

$W_{ci}$: the weight.

其中 $I = 5$，而如果比較的兩音節其第 $i$ 係數不同，則 $D_i = 1$；反之，$D_i = 0$。而各變數之權重（$W_{ci}$）設定，留待第四章說明。

### 3.5.3.2 韻律參數差異(Prosodic Information Difference)

在我們的合成系統中，韻律訊息產生器的輸出包含了音高軌跡、音節長度、與能量等三方面。於是在定義韻律參數差異時，也是由三方面構成，而各變數的詳細定義如下。平均音高差異之數學式為

$$d^{j}_{pitch\_mean} = \left| \frac{\bar{F}_j - \bar{F}_{target}}{\bar{F}_{target}} \right| \quad (3\text{-}5\text{-}8)$$

其中 $\bar{F}_j$ 為第 $i$ 種組合第 $k$ 個合成單元第 $j$ 個音節之平均音高，而 $\bar{F}_{target}$ 為目標句中相對應部分藉由前文提及之韻律訊息產生器所預測之平均音高。

音節長差異之數學式為

$$d^{j}_{duration} = \left| \frac{U_j - U_{target}}{U_{target}} \right| \quad (3\text{-}5\text{-}9)$$

其中$U_j$為第$i$種組合第$k$個合成單元第$j$個音節之音節長度，音節長度包含子音長與母音長，而$U_{target}$為目標句中相對應部分韻律訊息產生器預測之音節長度。

能量差異之數學式為

$$d_{power}^{j} = \left| \frac{E_j - E_{target}}{E_{target}} \right| \quad (3\text{-}5\text{-}10)$$

其中$E_j$為第$i$種組合第$k$個合成單元第$j$個音節之能量，這裡的能量定義與2.3.3節中之定義相同，而$E_{target}$為目標句中相對應部分韻律訊息產生器預測之能量。

### 3.5.4 合成單元間轉移差異(Transition Cost)

對於連接型的語音合成系統而言，由於合成目標句的各個部分可能由語料庫中的不同語句中取出，其連接地帶的不連續將會嚴重影響合成後的音質表現。在本系統中，合成單元間的連接差異考慮的因素有連接點個數及其位置的考量、連音效應的評估、與由中文結構樹決定的句法分數，其數學式為

$$C_{transition}(U_k^i, U_{k+1}^i) = \frac{w_{cc}}{Q_{cc}} d_{\text{concatenation\_cost}}^{k} + \frac{w_{ca}}{Q_{ca}} d_{\text{co-articulation}}^{k} \quad (3\text{-}5\text{-}11)$$

其中$C_{transition}(U_k^i, U_{k+1}^i)$代表第$i$個組合中第$k$個候選合成單元與第$k+1$個候選合成單元之合成單元間連接差異，$d_{\text{concatenation_cost}}^{k}$代表此兩候選合成單元間之連接點代價，$d_{\text{co-articulation}}^{k}$代表此兩候選合成單元間之連音效應評估代價。上述變數乘上代表其影響力的權重（$w_{cc}$、$w_{ca}$），與調整其數值變動範圍之正規化係數（$Q_{cc}$、$Q_{ca}$）

倒數。各變數的權重值與正規化係數的設定留待第四章再予以說明，現在先說明各部分的詳細定義方式。

#### 3.5.4.1 合成單元間連接代價(Concatenation Cost)

在一般連接型語音合成系統中，連接點的個數為影響合成音質的重要因數。最理想的狀況是，我們可以在語料庫中找到一個與目標句完全相同的句子，然而在實際上，只能期望合成音檔中的連接點越少越好，因為連接點的存在將會帶來語音的不流暢感。連接點個數相同的情況下，在我們以詞為優先的搜尋架構下，亦希望能夠優先選擇與系統文字分析器互相符合的詞串，因為文字分析器輸出的斷詞結果將會作為韻律訊息產生器的輸入以預測輸入文句的韻律，所以與斷詞結果相符合的詞串應當具有與預測韻律較接近的韻律訊息。在這些考量下，合成單元間連接代價的數學式為

$$d^{k}_{\text{concatenation\_cost}} = 1 + w_{\text{unmatch}} \quad \text{(3-5-12)}$$

其中 1 可以認為是對每個連接點有個基本的處罰，而 $w_{\text{unmatch}}$ 則是對那些與斷詞結果不符的連接點的加重處罰，$w_{\text{unmatch}} = 0$，如果連接點位於斷詞結果的詞邊界上；反之，$w_{\text{unmatch}}$ 介於 0 至 1 之間，視加重的程度而定，如果連接點不是位於斷詞結果的詞邊界上。

#### 3.5.4.2 合成單元間連音效應評估代價(Co-articulation Cost)

影響合成單元間的連接差異的另一個因素是連音效應。若某一合成單元具有連音效應，則其在與其他合成單元相連時，會出現雜音或者不連續的感覺，因此

連音效果越明顯的合成單元是越不好的合成單元。首先，我們應當定義如何的情況是有連音效應，評估連音效應的方法有很多，在本系統中以合成單元波形邊緣處的能量是否有明顯的下降(energy dip)來代表其連音效應。首先，對語料庫中的每一個合成單元，計算其第一音節之前三音框的平均能量，與最後一音節倒數三音框的平均能量，將此二平均能量與臨界值比較。如此平均能量小於臨界值，則視為無連音效應；反之，以平均能量與臨界值的差異作為連音效應的評估值，可以數學式表達如下，

$$\mathrm{C}_{\text{coarticalation}} = \max\{\overline{E} - E_T, 0\} \quad (3\text{-}5\text{-}13)$$

其中，$\mathrm{C}_{\text{coarticalation}}$ 代表某一音節邊緣的連音效應評估值，$\overline{E}$ 代表此邊緣 3 音框的平均能量，$E_T$ 為此臨界值，此臨界值的設定留待第四章討論。

很明顯地，當兩個合成單元連接時，其連音效應的效果是上一個合成單元的右側連音效應評估值（$C^k_{right}$）加上此合成單元的左側連音效應評估值（$C^{k+1}_{left}$），所以我們可以定義兩個合成單元間的連音效應評估代價之數學式為

$$d^k_{\text{co-articulation}} = C^k_{right} + C^{k+1}_{left} \quad (3\text{-}5\text{-}14)$$

### 3.5.5 挑選單元總結

在此將用於挑選合成單元的數學式完整說明一次，其變數與符號的定義同於以上各節，首先 cost function 由合成單元目標差異與合成單元間轉移差異組成，

$$C(S_i, T) = w_{target} \sum_{k=1}^{n} C_{target}(U^i_k, T) + (1 - w_{target}) \sum_{k=1}^{n-1} C_{transition}(U^i_k, U^i_{k+1}) \quad (3\text{-}5\text{-}5)$$

而合成單元目標差異考量前後文變數與韻律參數，而合成單元間轉移差異受連接點位置與連音效應影響，

$$C(S_i,T)=w_{target}\sum_{k=1}^{n}\{\sum_{j=1}^{m}(\frac{w_c}{Q_c}d^j_{contextual}+\frac{w_{p\_m}}{Q_{p\_m}}d^j_{pitch\_mean}+\frac{w_d}{Q_d}d^j_{duration}+\frac{w_p}{Q_p}d^j_{power})\}$$
$$+(1-w_{target})\sum_{k=1}^{n-1}\{\frac{w_{cc}}{Q_{cc}}d^k_{\text{concatenation\_cost}}+\frac{w_{ca}}{Q_{ca}}d^k_{\text{co-articulation}}\} \quad (3\text{-}5\text{-}15)$$

將上式的每一個變數皆由最原始的數據表達，則可展開為

$$C(S_i,T)=$$
$$w_{target}\sum_{k=1}^{n}\{\sum_{j=1}^{m}(\frac{w_c}{Q_c}(\sum_{i=1}^{I}W_{ci}D_i)+\frac{w_{p\_m}}{Q_{p\_m}}\left|\frac{\bar{F}_j-\bar{F}_{target}}{\bar{F}_{target}}\right|+\frac{w_d}{Q_d}\left|\frac{U_j-U_{target}}{U_{target}}\right|+\frac{w_p}{Q_p}\left|\frac{E_j-E_{target}}{E_{target}}\right|)\} \quad (3\text{-}5\text{-}16)$$
$$+(1-w_{target})\sum_{k=1}^{n-1}\{\frac{w_{cc}}{Q_{cc}}(1+w_{\text{unmatch}})+\frac{w_s}{Q_s}(C^k_{right}+C^{k+1}_{left})\}$$

即為本系統完整之挑選合成單元數學式。在上述的討論，皆假設候選合成單元由語料庫中挑出，在實際的情況下，有可能是搜尋單元未能找到而被標記為需用 PSOLA 合成的音節，在此情況下，會給予一極高的處罰，來保證由語料庫中挑出的候選合成單元其優先權。

## 3.6 波形合成器之說明

### 3.6.1 於波形間穿插靜音後連接

在保留語料庫原始音質的前提上，以大型語料庫為基礎的語音合成系統之波形合成器只要將挑選出來的合成單元連接起來即可。假設合成目標句整句都有挑選到適當的合成單元，那麼波形合成器會將那些被挑選出來的合成單元波形之間穿插長度由韻律訊息產生器所預測之靜音，一一連接起來。然而，如果任何部份的合成目標句未能在語料庫中找尋到適當的合成單元(這情況只發生在語料庫沒有此音調音節的時候)，此時波形合成器會使用之前的 PSOLA 合成技巧產生此部分，而也和其他找到的部分連接起來。

### 3.6.2 波形能量調整

就波形處理上而言，波形振幅的放大縮小並不會影響音質，而只會改變能量大小，所以就合成系統的考量上，當我們挑到一個與預測能量有些差異的候選合成單元時，應可將此候選合成單元之能量調整至所猜測的數值，而不會對音質有所損害。假設候選合成單元某音節之能量為 $E_t$，經過將此段音節波形乘以一個倍數 $a$，能量變為韻律訊息產生器所猜測的能量 $E_r$，則此倍數 $a$ 可由下式推導得知。

$$E_t = 10\log_{10}\frac{\sum(w_i x_i^2)}{N} \qquad E_r = 10\log_{10}\frac{\sum(w_i (a x_i)^2)}{N}$$

$$\frac{E_r - E_t}{10} = \log_{10}\frac{\frac{\sum(w_i (a x_i)^2)}{N}}{\frac{\sum(w_i x_i^2)}{N}} = \log_{10}\frac{\sum(w_i a^2 x_i^2)}{\sum(w_i x_i^2)} = \log_{10}(a^2) = 2\log_{10}(a) \qquad (3\text{-}6\text{-}1)$$

$$a = 10^{\frac{E_r - E_t}{20}}$$

然而問題是一個候選合成單元可能擁有數個音節，其不同音節所需調整的倍數可能不同。針對此問題，我們曾經想過兩個解決之道，一是個別音節各自按照預測能量調整，二是整段候選合成單元選擇一共同的代表能量，整段按照此代表能量與預測能量調整。最後我們採取第二種作法，因為第一種作法如果音節邊緣切割位置不準確，則此振幅調整無異形成一個波形上不連續點。我們以多音節候選合成單元的音節平均能量與目標句相同部分之平均能量作為調整的依據。

### 3.6.3 句首淡入(fade-in)與句尾漸消(fade-out)

雖然原則上盡量不調整波形以保持原始的音質，然而由於用於目標句句首和句尾的合成單元，並不保證其在語料庫中也是句首和句尾，此不一致的現象在聽覺上會造成突兀的感覺，於是波形合成器會將用於句首的合成單元實施淡入(fade-in)處理，而將用於句尾的合成單元則實施漸消(fade out)處理。實行淡入處理的部分是整段波形前 480 點樣本點，而實行漸消處理的部分是整段波形最後 480 點樣本點，兩段的實施方式都是將波形乘上一直角三角形，如下圖所示，其中 Length 為 480 samples。


**圖 3-6-1：淡入(fade in)與漸消(fade out)示意圖**

## 3.7 以語料庫為基礎之語音合成系統總結

此章中，我們介紹了所設計的以語料庫為基礎之語音合成系統。合成系統可分為四個模組，文字分析器、韻律訊息產生器、合成單元選擇器與波形合成器。合成單元選擇器又分為搜尋單元與挑選單元，搜尋單元負責尋找語料庫中可用於合成目標句的文字片段，挑選單元藉由 cost function 自搜尋結果中選擇一最佳的組合。波形合成器依照最佳組合合成目標句，除了將所挑選的波形片段連接外，如沒有適合的波形，亦可由 PSOLA 及對原始波形調整來合成。

下一章節中，將會介紹合成系統中的設定問題，及實際執行的考量，如記憶體大小及合成時間。

# 第四章　系統設定與系統效能評估

在前三章中，已說明了整體系統的架構與建構方式。然而，在實際的運用上，仍有一些系統設定需要調整，此章節即是對此方面予以說明。另一方面，也實際的測量本論文所提出的系統之效能，如合成花費時間、系統執行所需記憶體、合成音質問題分析等之類的討論。

## 4.1 系統設定

### 4.1.1 用於挑選單元之權重值與正規化參數設定

在 3.5 節所提及之挑選單元中，所使用的各類資訊差異甚大，所以其在演算法中表現的數值變動範圍亦不同。如為了調整各變數的重要性，我們需要先使用正規化參數來使得它們的數值變動範圍相同，再使用權重值賦予不同的重要性。

在 cost function 中，除了合成單元間連接代價及合成單元間連音效應評估代價可由其數學式之定義看出有明顯的數值上限外（$d^k_{\text{concatenation\_cost}} \le 1 + w_{\text{unmatch}}$，$d^k_{\text{co-articulation}} \le 2(\overline{E}_{\max} - E_T)$，$\overline{E}_{\max}$ 為語料庫中最大之 $\overline{E}$ 值），無法預測其餘變數的數值上限，為了要讓所有的變數之變動範圍有共同之基準，我們必須先藉由測試資料來獲得大量的數據，統計後可得其大略的數值上限。我們在語料庫的文章外，另外找了 10 篇文章作為測試資料，此測試資料計有 134 句目標句（以標點符號區隔），1,114 個音節，將其合成過程之數據記下。以平均音高差異為例，其數學式定義為

$$d^j_{pitch\_mean} = \left| \frac{\overline{F}_j - \overline{F}_{target}}{\overline{F}_{target}} \right| \quad (3\text{-}5\text{-}8)$$

紀錄 $d^{j}_{pitch\_mean}$ 之數值如下圖，其以 $d^{j}_{pitch\_mean}$ 值為橫軸，以此 $d^{j}_{pitch\_mean}$ 值以下的涵蓋率為縱軸，

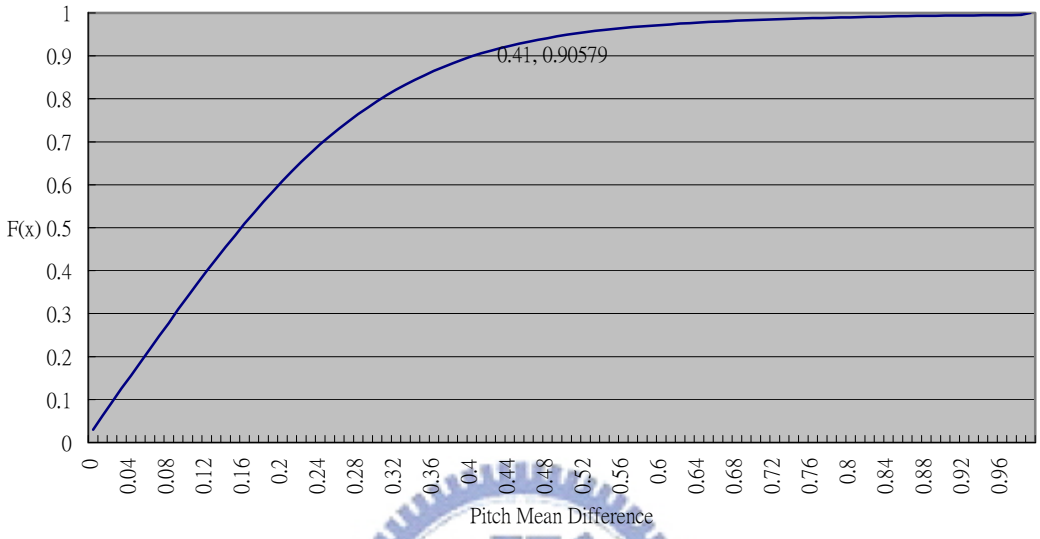



**圖 4-1-1：The cumulative distribution function of Pitch-Mean Difference**

由圖中可以看出其大部分的數值皆集中於 0.41 以下，為了排除一些極端的情況，我們取其涵蓋率達 90%的數值為其數值上限，因為 CDF 圖上對應縱軸 0.9 之橫軸數值為 0.41，所以平均音高失真度之正規化參數設為 0.41。仿照平均音高失真度的作法，我們可以得到其他變數之正規化參數，其他變數之 CDF 圖請參照附錄六、「挑選單元中各變數之統計數據」。在此將所有正規化參數列出，其符號與第三章中所定義相同。

$$Q_c = 1.02$$
$$Q_{p\_m} = 0.41$$
$$Q_d = 0.63$$
$$Q_p = 0.17$$
$$Q_{cc} = 1 + w_{\text{unmatch}}$$
$$Q_{ca} = 2(\overline{E}_{\max} - E_T)$$

至於權重值之設定，前人有些針對於此的研究[16]，由於本系統是屬於波形

連接型的系統，我們希望連接點越少越好，以減少波形連接所造成的不連續感，所以給予合成單元間連接代價此變數最大的權重值。韻律訊息也是另一個決定性的因素，依照平均音高、音節長度、音節能量的順序給予由大至小的權重，連音效應也給予一個適當的權重。在此將所有的權重值列出，其符號與第三章中所定義相同。

$$w_c = 4$$
$$w_{p\_m} = 8$$
$$w_d = 6$$
$$w_p = 4$$
$$w_{cc} = 18$$
$$w_{ca} = 4$$

### 4.1.2 用於計算前後文相關係數差異之各項係數權重值設定

在 3.5.3.1 節中提到前後文相關係數差異之定義方式，兩比較對象之前後文相關係數差異之定義為

$$d^j_{contextual} = \sum_{i=1}^{I} W_{ci} D_i \qquad (3\text{-}5\text{-}7)$$

$d^j_{contextual}$: the contextual distance.

$D_i$: the distance for the $i$th coordinate in the vector.

$W_{ci}$: the weight.

其中各變數之權重值，$W_{ci}$，亦是需要考量的因素。在前人的研究中[15,17]，這些權重值之設定可由平均鑑定分數(Mean Opinion Score, MOS)測試決定。然而在本次研究中，並未進行 MOS 測試，我們參考前人的研究結果，依照其相對重要性給予權重值，最後採用之權重值如下表所示。

表 4-1-1：前後文相關係數差異之權重值定義

| Component | Weight |
|---|---|
| Left Phonetic Context | 0.154 |
| Right Phonetic Context | 0.149 |
| Left Tone | 0.254 |
| Right Tone | 0.229 |
| Position In Word | 0.214 |

### 4.1.3 用於計算連接代價之權重值設定

在 3.5.4.1 節述及連接代價的定義方式，其中 $w_{\text{unmatch}}$ 為不在詞邊界上的連接點之加重處罰。此數值設定的目的在於期盼合成系統中的挑選機制可以優先選擇與文字分析器之斷詞結果相符的候選合成單元詞串，因為此斷詞結果將是韻律訊息產生器預測目標句韻律之依據。經過幾次測試，$w_{\text{unmatch}} = 0.5$ 是可以令挑選機制在受到斷詞結果影響下，選擇與斷詞結果相符的候選合成單元之設定值。

### 4.1.4 用於評估連音效應之能量臨界值設定

在 3.5.4.2 節述及連音效應評估代價的定義方式，其中 $E_T$ 為判斷音節波形邊緣是否受到連音效應影響之臨界值。為了決定此臨界值，首先需要知道多強的連音效應會對人類聽覺帶來可察覺的影響。我們事先統計了語料庫中各音節邊緣 3 音框的平均能量，及前文中提到之 $\overline{E}$，其分佈情況如下二圖所示，圖 4-1-2 橫軸

為此平均能量之值，縱軸為介於此範圍間之個數，統計時，並未區分左端與右端邊緣的差別，由圖中可知大部分的情況分佈於 50 至 62.5dB 之間。

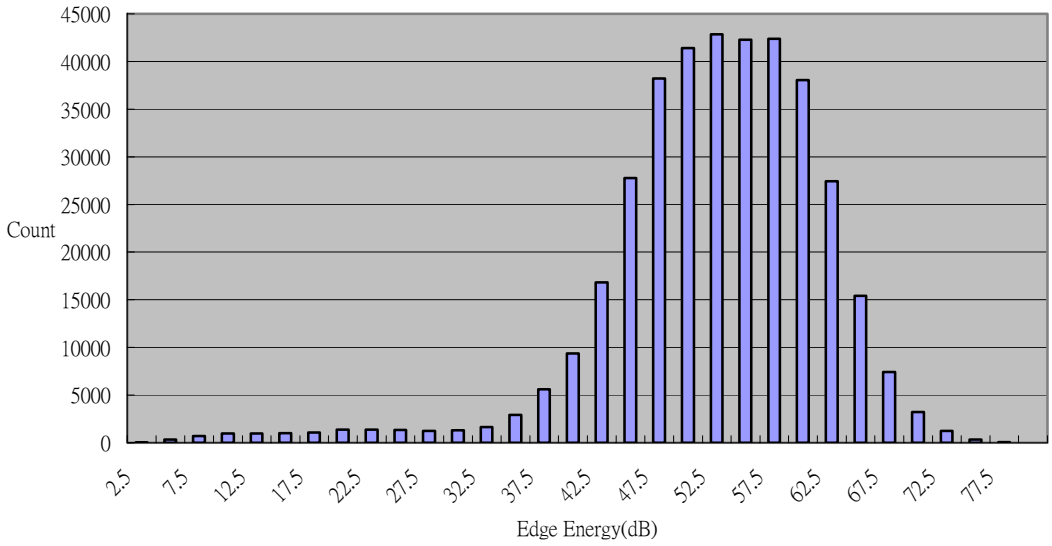



**圖 4-1-2：音節邊緣平均能量（3frames）統計圖**

而圖 4-1-3 中橫軸為此平均能量之值，縱軸為累加分佈函數數值。

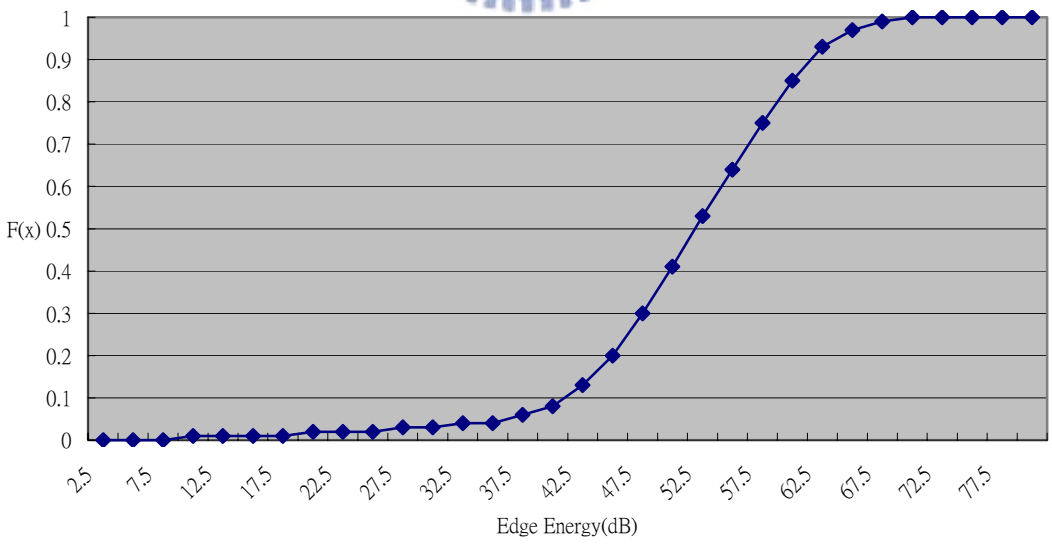



**圖 4-1-3：音節邊緣平均能量（3frames）累加分佈函數圖**

為了決定此能量臨界值，$E_T$，我們利用 4.1.1 節中所提到之測試資料。首先，我們將 $E_T$ 設為 $\overline{E}$ 之最大值 $\overline{E}_{\max}$，82.5dB，接著，我們實際合成此測試資料中之目標句，一旦發現有連音效應造成連接點波形變化明顯不順暢的情況，即將 $E_T$ 設為此連接點較大之 $\overline{E}$ 值減一，即將此候選合成單元視為有受到連音效應影響，並重新合成此句。如仍然挑選到此候選合成單元，確認挑選單元是否有其他具有較小 $\overline{E}$ 值之候選單元可選擇，如是則繼續降低 $E_T$ 值，如否則停止。經過上述之調整，最後此能量臨界值，$E_T$，設為 52.5dB，也就是說由語料中任選一音節邊緣，有 47%的機率挑到被認為是有受到連音效應影響的音節邊緣。

## 4.2 系統效能評估

### 4.2.1 系統執行時所使用之記憶體大小

而本系統所使用到的各式資料，其檔案大小列於下表中。

**表 4-2-1：合成系統各項資料列表**

| 名稱 | 用途 | 檔案大小 |
|---|---|---|
| 合成語料庫 | 語料庫音檔 | 1,045MB |
| 音節位置表格 | 記載音檔之音節位置資訊 | 1,762KB |
| 多字詞字元位置表格 | 記載多字詞之文章位置 | 2,764KB |
| 單字詞字元位置表格 | 記載單字詞之文章位置 | 209KB |
| 音調音節位置表格 | 記載同音字之文章位置 | 1,479KB |
| 前詞綴表格 | 記載前詞綴之文章位置 | 3KB |
| 後詞綴表格 | 記載後詞綴之文章位置 | 13KB |
| 音節韻律訊息表格 | 記載各音節之韻律訊息 | 2,639KB |
| 音節邊緣能量表格 | 記載各音節之邊緣平均能量 | 3,515KB |
| 中文姓氏表格 | 記載中文姓氏之文章位置 | 6KB |
| 音節相關前後文係數表格 | 記載各音節之前後文係數 | 4,391KB |
| 411 音節子母音分類表 | 記載 411 音節子母音分類 | 4KB |
| | | |
| 總大小 | | 1,062MB |

以上所列為架設系統所需之記憶體，至於系統實際執行時所佔用的記憶體，系統初始化時為 59MB，實際運作時所佔用之記憶體隨合成文句長度與搜尋結果之大小而變化，合成文句越長與搜尋結果越多則所佔用之記憶體越大。

### 4.2.2 合成目標句系統所需之時間

在一開始架構此語音合成系統之時，我們採用將所有輸入文句合成後，才播放給使用者的方式。然而由於以語料庫為基礎之合成系統的執行速度不如以往以少量語料庫配合 PSOLA 合成技術之系統快速，如此的作法會讓使用者的等待時間過久。於是我們改用一邊合成一邊播放的方式來減少使用者的等待時間，大致而言，每合成完一句目標句就立刻播放此句，利用播放此句的時間，系統同時合成下一句，一篇文章之合成過程如下圖所示。

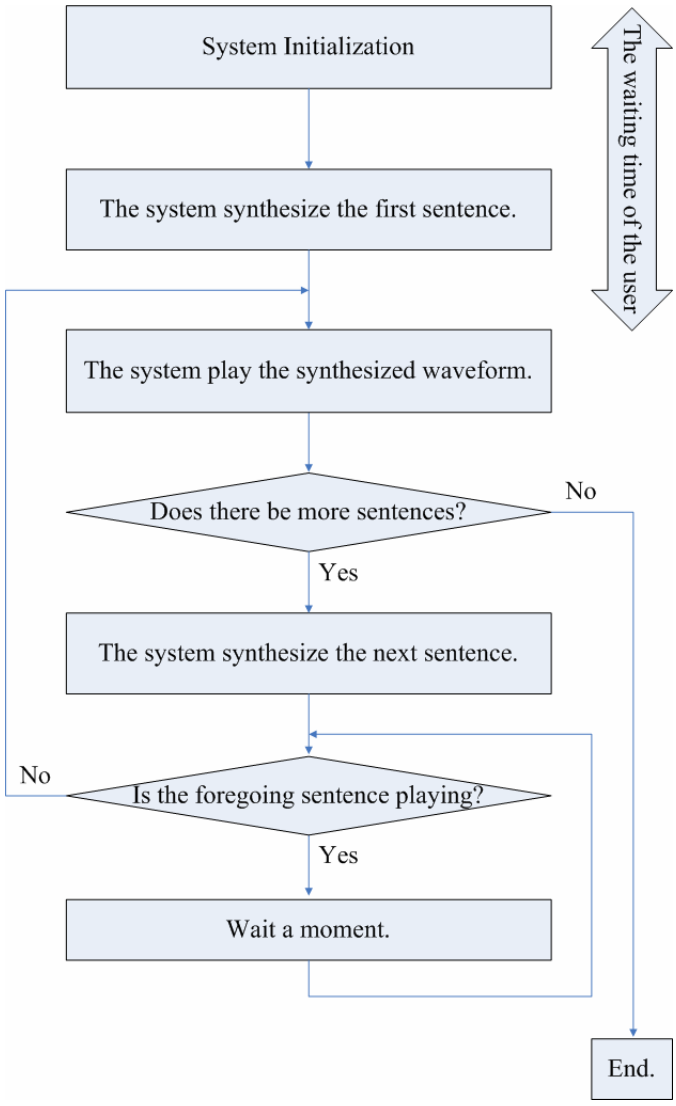



**圖 4-2-1：多句目標句之合成流程示意圖**

### 4.2.3 圖形化使用者輸出入介面

在本文中提出的語音合成系統，其輸入為 BIG5 形式的中文字，而回應系統使用者的語音輸出則為使用音效設備播放取樣率 16 kHz 之 PCM 音檔，使用者使用系統時所看到的介面如下圖所示。介面上共有四個輸出入方塊與三個按鈕，上方的方塊為「合成內容輸入方塊」，使用者可在此輸入欲合成的文句，假設在此輸入「認為有個人資料可能遭到誤用，」。按下下方中間的「TTS」按鈕，系統即會以合成內容輸入方塊內之內容來合成，並將合成波形播放。下方有三塊輸出入方塊，最左邊為「文字分析器輸出入方塊」，每當系統合成一段文句，此方塊就會顯示此文句之文字分析結果，使用者也可直接由此方塊修改文字分析結果來改變合成的聲音。下方中間的方塊為「語料庫搜尋結果輸出入方塊」，每當系統合成一段文句，此方塊就會顯示此文句在語料庫中搜尋到的多字詞清單，同樣，使用者也可直接由此方塊修改語料庫搜尋結果來改變合成的聲音。下方最右邊的方塊為「挑選結果輸出方塊」，每當系統合成一段文句，此方塊就會顯示此合成波形的組成（以空行分隔不同的合成單元）與實際挑選及預測的合成單元之韻律訊息，排列訊息依序為「（實際合成單元音節長/預測目標句音節長 | 實際合成單元能量/預測目標句能量 | 實際合成單元平均基頻/預測目標句平均基頻 | 合成單元位於語料庫中之句編號 合成單元位於語料庫中之字元編號）」。

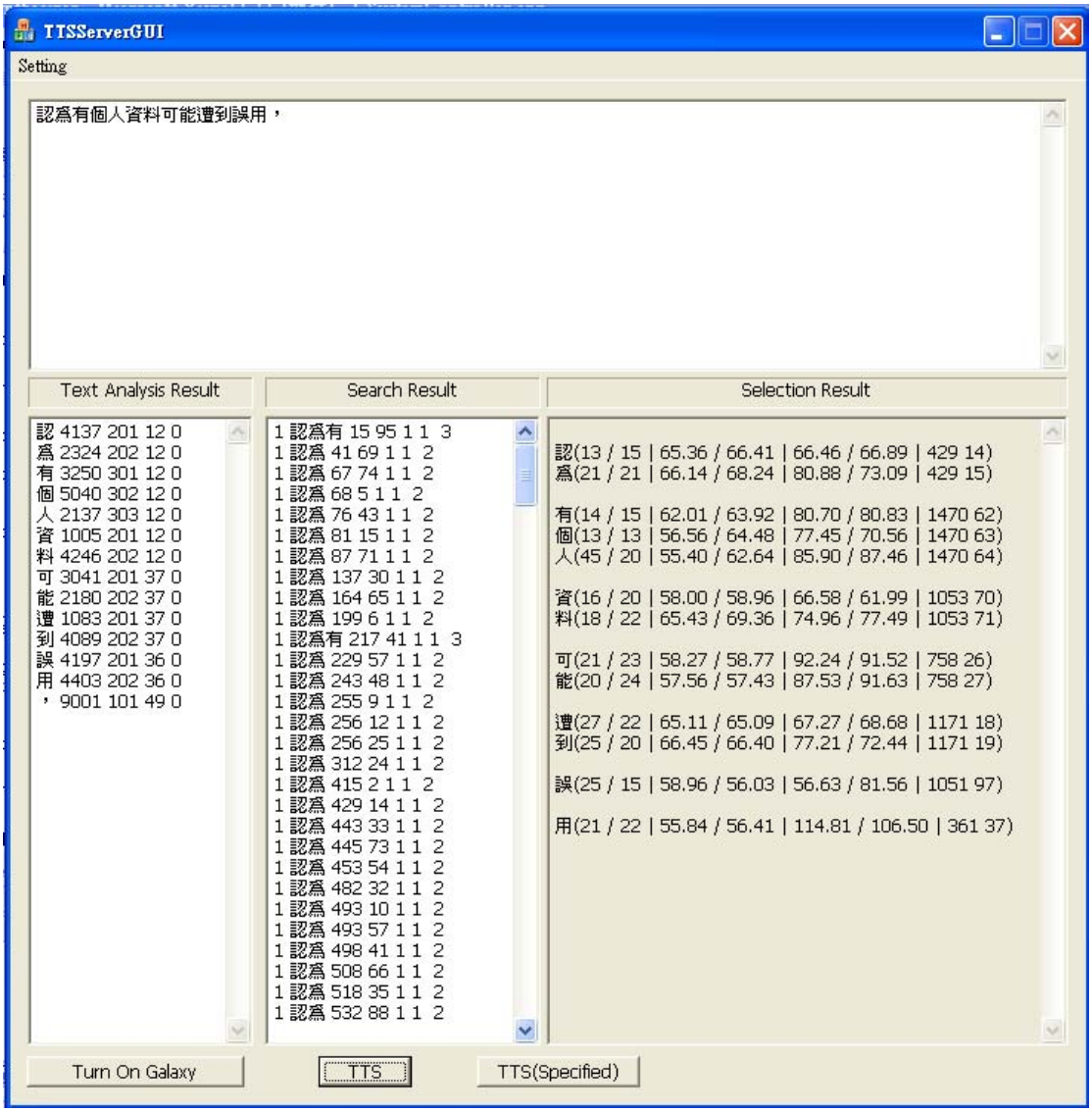



**圖 4-2-2：合成系統之使用者介面外觀**

由於當初發展系統時，希望能方便地觀察文字分析結果與多字詞搜尋結果對輸出波形的影響，我們設計了可由指定的斷詞結果與多字詞搜尋結果來合成的方式，「TTS(Specified)」按鈕即是此種合成方式。當按下「TTS(Specified)」按鈕，系統會以目前「文字分析器輸出入方塊」與「語料庫搜尋結果輸出入方塊」的內容來合成波形，也就是說，只要我們修改此兩方塊之內容，即可依照使用者的意思合成音檔。在一些情況下，文字分析器未能正確依照文句意義斷詞，如上圖中，斷詞結果為「認為 | 有個人 | 資料 | 可能 | 遭到 | 誤用」，其中「 | 」代表詞

邊界，所以挑選單元選擇了一個三字詞「有個人」。然而實際上輸入文句中並沒有「有個人」的意思，我們手動更改斷詞結果為「認為｜有｜個人｜資料｜可能｜遭到｜誤用」，如下圖，並按下「TTS(Specified)」按鈕，發現挑選單元此次選擇了較為自然的「個人」二字詞。於是，我們可以藉由此介面直接修改斷詞結果，以避免錯誤的斷詞結果影響合成文句的自然度，甚至我們可以直接修改語料庫搜尋結果輸出入方塊中的內容，來指定要由哪一個多字詞合成。

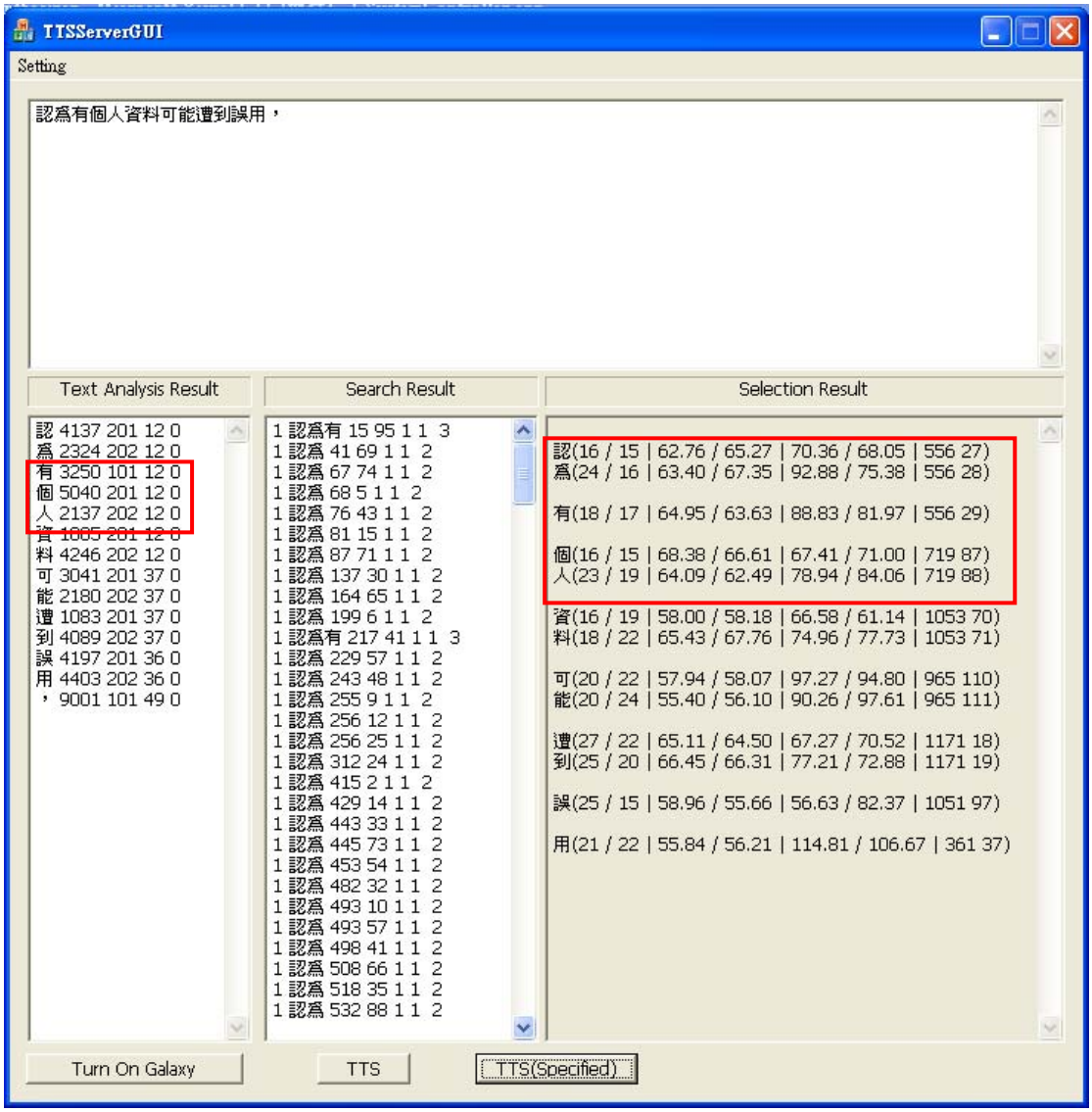



**圖 4-2-3：利用手動修改斷詞結果的範例**

## 4.3 實驗結果與分析

### 4.3.1 資料涵蓋率問題

為了減少連接點的個數，搜尋單元能否在語料庫中找到相符的多字詞至為關鍵。為了瞭解本語料庫中的詞彙豐富度，我們利用本實驗室中的詞典進行比對，然而由於詞典收錄甚多，其中不乏生冷艱澀之詞，為了實際呈現語料庫中的詞彙涵蓋率，我們依照詞典中的詞頻分別統計詞彙涵蓋率。

下圖為語料庫對於詞典的涵蓋率，在此，僅統計二字詞以上之詞，圖中橫軸為詞典之詞頻，縱軸為語料庫對於詞典中超過此詞頻之詞的涵蓋率。

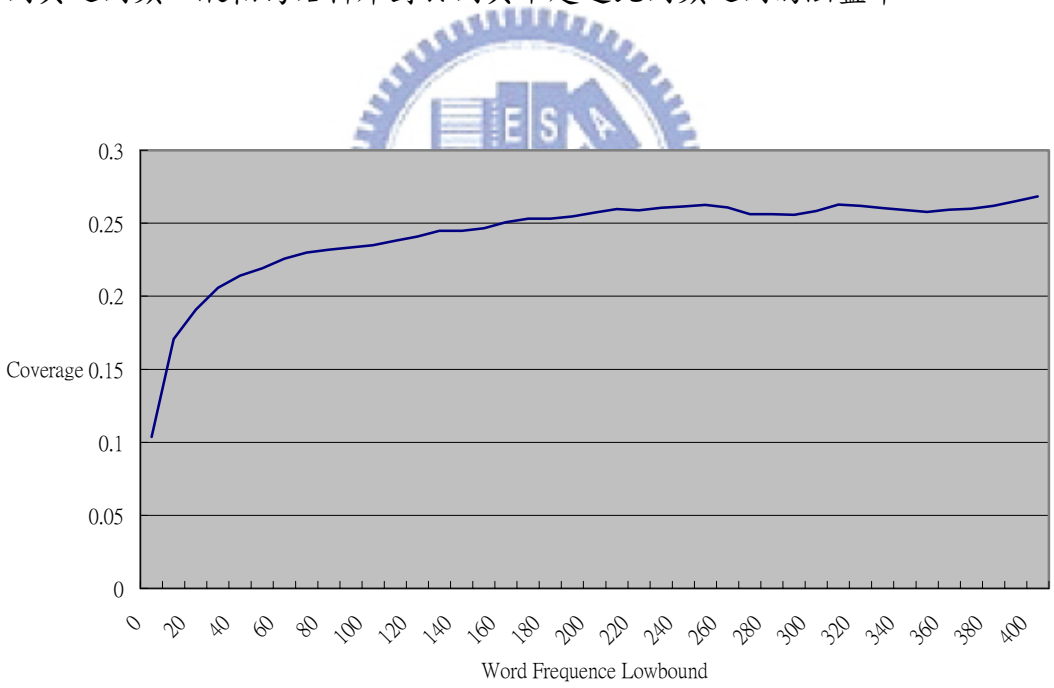



**圖 4-3-1：語料庫對詞典涵蓋率**

由圖中可看出，如不考慮詞頻，語料庫對詞典僅有 10.37%的涵蓋率。由第二章之表 2-1-1 可知，語料庫二字詞以上之詞僅有 13,462 種，而詞典中二字詞以上之詞共有 108,222 種，比例為 12.44%，然而詞典中並不收錄定量複合詞，實際

的數值必會低於此值，所以 10.37%的涵蓋率是合理的。隨著詞頻下限的上升，語料庫對於詞典的涵蓋率也隨之上升。然而，當詞頻下限上升至 200 次以上後，涵蓋率停留在 26%至 28%之間震盪。所以，如果我們任意由詞典中選擇一個二字詞，可在語料庫中找到相符合詞的機率為 10.37%，即使此詞較為常見（詞典中的詞頻達到 200 次），機率也僅有 27%左右。可見得，此語料庫之詞內容並不甚豐富，不能完全依賴多字詞的搜尋結果合成。

當搜尋單元未能在語料庫中找到與目標句相符合之詞或詞串時，假設目標句並非詞綴或中文姓氏，搜尋單元將會尋找同音字候選合成單元。關於音調音節的統計數據，在 3.4.4.4 節中提過，語料庫中共有 1,068 種音調音節，以中文音調音節有 1,300 種的概念而言，此語料庫有 82%的涵蓋率。

### 4.3.2 語料庫中切割位置不正確問題

波形連接型合成系統需要準確的音節切段位置，雖然在第二章中我們使用了 HTK 軟體訓練 HMM 模型對語料庫切割，以得到音段切割位置。然而在連續語音中，有時會發生前後音節耦合的連音現象，此時音段切割位置難以決定且通常交界處並無明顯的能量下降。在 3.5.4.2 節中，我們利用觀察音節邊緣 3 音框平均能量的方式，可以盡量不要挑選帶有連音效應的候選合成單元，但在有限的選擇性下，帶有連音效應的候選合成單元仍有可能作為最後的合成單元。

## 4.4 章節總結

本章中說明了關於系統設定以及探討系統效能的表現。在關於正規化參數設定方面，為了決定各變數之數值變動範圍，我們準備了一套小型文字資料，以統計的方式估計各變數的數值上限。在關於權重值的設定方面，我們依照設計此挑選機制的理念及參考前人的研究，給予各變數適當的權重值。在系統效能的討論上，除了條列出合成系統中各項資料所需之記憶體大小，並說明為了及時而快速的語音輸出，使用邊合成邊播音的運作方式。最後，為了使用以及掌控合成音質的便捷性，我們設計了一套圖形使用者介面。在這介面上，使用者可以自動地合成輸入文句，也可依照其意願掌控其合成方式。

# 第五章　結論與未來展望

## 5.1 結論

在本文中，我們提出一套以語料庫為基礎的中文語音合成系統之實作方式。與過去實驗室所發展的以基本音節搭配 PSOLA 技術的合成系統相比，以語料庫為基礎的合成系統其合成音質更顯得自然流利。在發展此系統時，所遇到之困難與經驗，可列於以下結論：

1. 波形連接點對合成音質有極大之影響，為了減少連接點對合成音質之傷害，語料庫應盡量提高對於常用詞之涵蓋率。
2. 合成單元其音段切割位置的不正確，將會造成合成語音中之雜音。如何自動而正確的標記音段切割位置仍是影響合成系統效能的重要因素。
3. 即使能有正確的切割位置，連音效應仍舊嚴重損害合成音質，在此實作中，我們利用能量避免選到有嚴重連音效應的音節。然而，如何有效而正確的判別連音效應亦有未盡考量之處。
4. 在本實作方式中，在一些候選合成單元有限的情況下，如語料庫中只有一個同音字，選擇機制未能發揮其作用。這時，我們應思考是否回歸到 PSOLA 的合成方式，也就是說，以語料庫為基礎的方式與 PSOLA 方式的配合，也是可以考慮的發展方向。

## 5.2 未來展望

如今，語音辨認技術已向跨語言的方向進展，相同地，語音合成也邁向多語言的研究領域[18,19]。本實驗室近年來致力於國語、閩南語、客家話三種語言的辨認與合成，三方面皆有所斬獲，但尚未發展出一套與語言無關之合成技術，將來寄望能整合三方面的技術，形成一套多語言的合成系統。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)或者其他以聽者反應為依據的評斷方式。然而此些方式不但費時費力，且標準不一，難以在不同合成系統間比較。為了明確瞭解不同作法對於合成系統效能帶來的影響，一套制式而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】 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。

【2】 Chou, F. C., C. Y. Tseng, and L. S. Lee, “A Set of Corpus-Based Tex-to-Speech Synthesis Technologies for Mandarin Chinese” in Pro. ICASSP, Vol. 10, pp.481-494, 2002.

【3】 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、中央研究院研究所。

【4】 The HTK Book (for HTK Version 3.2.1)

【5】 林立峰，“中文 TTS 系統與音合成之改進”，國立交通大學碩士論文，民國九十三年六月。

【6】 Wavesufer Homepage : http://www.speech.kth.se/wavesurfer/

【7】 Chen, S.H., S.H. Hwang and Y. R. Wang, "An RNN-based prosodic information Synthesizer for Mandarin text-to-speech", IEEE Trans. On Speech and Audio Processing, Vol. 6, NO. 3, pp.226-239, 1998.

【8】 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.

【9】 江振宇，“中文斷詞器之改進”，國立交通大學碩士論文，民國九十三年七月。

【10】黃紹華，“中文文句翻語音系統中韻律訊息產生器之研究”，國立交通大學博士論文，民國八十五年六月。

【11】Jian Yu, Jianhua Tao and Xia Wang, "Pitch Prediction for Mandarin TTS with Mutual Prosodic Constraint", ISCSLP, 2006.

【12】Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, “Spoken language processing:

a guide to theory, algorithm, and system development".

【13】Blouin C., Rosec O., Bagshaw P., D'Alessandro C., "Concatenation cost calculation and optimisation for unit selection in TTS", IEEE 2002 Workshop on Speech Synthesis. Santa Monica, USA, September 11-13, 2002.

【14】Erdem, C.; Beck, F.; Hirschfeld, D.; Hoege, H.; Hoffman R., 2002c. Robust unit selection based on syllable prosody parameters. IEEE 2002 Workshop on Speech Synthesis. Santa Monica, California USA.

【15】Chu, M., Peng, H., Yang, H. and Chang, E., "Selecting non-uniform units from a very large corpus for concatenative speech synthesizer", In Proceedings of ICASSP, Salt Lake City. 2001.

【16】Alfas, F., Llora, X., Formiga, L., Sastry, K., Goldberg, DE, "EFFICIENT INTERACTIVE WEIGHT TUNING FOR TTS SYNTHESIS: REDUCING USER FATIGUE BY IMPROVING USER CONSISTENCY", 2006 ICASSP International Conference on Acoustics, Speech and Signal Processing (ICASSP06), vol. I, pp. 865-868, Maig, Toulouse (Franca).

【17】H. Peng, Y. Zhao, and M. Chu, "Perpetually optimizing the cost function for unit selection in a TTS system with one single run of MOS evaluation", in Proc. ICSLP, (Denver, USA), 2002.

【18】R. Hoffmann et al., "A multilingual TTS system with less than 1 MByte footprint for embedded applications", Proc. ICASSP, Hong Kong, 2003.

【19】Nakamura, S. et al., "The ATR Multilingual Speech-to-Speech Translation System", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 2, MARCH 2006.

# 附錄一　國語411基本音節總音表

**表A　國語411基本音節總音表**

| 音碼 | 注音 | 漢語拼音 | 子音編號(22 類) | 母音編號（7 類） |
|---|---|---|---|---|
| 1 | ㄓ | zhi | 2 | 1 |
| 2 | ㄔ | chi | 3 | 1 |
| 3 | ㄕ | shi | 4 | 1 |
| 4 | ㄖ | ri | 5 | 1 |
| 5 | ㄗ | zi | 6 | 1 |
| 6 | ㄘ | ci | 7 | 1 |
| 7 | ㄙ | si | 8 | 1 |
| 8 | ㄚ | a | 1 | 2 |
| 9 | ㄓㄚ | zha | 2 | 2 |
| 10 | ㄔㄚ | cha | 3 | 2 |
| 11 | ㄕㄚ | sha | 4 | 2 |
| 12 | ㄗㄚ | za | 6 | 2 |
| 13 | ㄘㄚ | ca | 7 | 2 |
| 14 | ㄙㄚ | sa | 8 | 2 |
| 15 | ㄍㄚ | ga | 9 | 2 |
| 16 | ㄎㄚ | ka | 10 | 2 |
| 17 | ㄏㄚ | ha | 11 | 2 |
| 18 | ㄉㄚ | da | 15 | 2 |
| 19 | ㄊㄚ | ta | 16 | 2 |
| 20 | ㄋㄚ | na | 17 | 2 |
| 21 | ㄌㄚ | la | 18 | 2 |
| 22 | ㄅㄚ | ba | 19 | 2 |
| 23 | ㄆㄚ | pa | 20 | 2 |
| 24 | ㄇㄚ | ma | 21 | 2 |
| 25 | ㄈㄚ | fa | 22 | 2 |
| 26 | ㄛ | o | 1 | 3 |
| 27 | ㄌㄛ | lo | 18 | 3 |
| 28 | ㄅㄛ | bo | 19 | 3 |
| 29 | ㄆㄛ | po | 20 | 3 |
| 30 | ㄇㄛ | mo | 21 | 3 |
| 31 | ㄈㄛ | fo | 22 | 3 |
| 32 | ㄜ | e | 1 | 4 |
| 33 | ㄓㄜ | zhe | 2 | 4 |
| 34 | ㄔㄜ | che | 3 | 4 |
| 35 | ㄕㄜ | she | 4 | 4 |
| 36 | ㄖㄜ | re | 5 | 4 |

| | | | | |
|---|---|---|---|---|
| 37 | ㄗㄜ | ze | 6 | 4 |
| 38 | ㄘㄜ | ce | 7 | 4 |
| 39 | ㄙㄜ | se | 8 | 4 |
| 40 | ㄍㄜ | ge | 9 | 4 |
| 41 | ㄎㄜ | ke | 10 | 4 |
| 42 | ㄏㄜ | he | 11 | 4 |
| 43 | ㄉㄜ | de | 15 | 4 |
| 44 | ㄊㄜ | te | 16 | 4 |
| 45 | ㄋㄜ | ne | 17 | 4 |
| 46 | ㄌㄜ | le | 18 | 4 |
| 47 | ㄞ | ai | 1 | 4 |
| 48 | ㄓㄞ | zhai | 2 | 4 |
| 49 | ㄔㄞ | chai | 3 | 4 |
| 50 | ㄕㄞ | shai | 4 | 4 |
| 51 | ㄗㄞ | zai | 6 | 4 |
| 52 | ㄘㄞ | cai | 7 | 4 |
| 53 | ㄙㄞ | sai | 8 | 4 |
| 54 | ㄍㄞ | gai | 9 | 4 |
| 55 | ㄎㄞ | kai | 10 | 4 |
| 56 | ㄏㄞ | hai | 11 | 4 |
| 57 | ㄉㄞ | dai | 15 | 4 |
| 58 | ㄊㄞ | tai | 16 | 4 |
| 59 | ㄋㄞ | nai | 17 | 4 |
| 60 | ㄌㄞ | lai | 18 | 4 |
| 61 | ㄅㄞ | bai | 19 | 4 |
| 62 | ㄆㄞ | pai | 20 | 4 |
| 63 | ㄇㄞ | mai | 21 | 4 |
| 64 | ㄝ | eh | 1 | 2 |
| 65 | ㄓㄟ | zhei | 2 | 4 |
| 66 | ㄕㄟ | shei | 4 | 4 |
| 67 | ㄗㄟ | zei | 6 | 4 |
| 68 | ㄙㄟ | sei | 8 | 4 |
| 69 | ㄍㄟ | gei | 9 | 4 |
| 70 | ㄏㄟ | hei | 11 | 4 |
| 71 | ㄉㄟ | dei | 15 | 4 |
| 72 | ㄋㄟ | nei | 17 | 4 |
| 73 | ㄌㄟ | lei | 18 | 4 |
| 74 | ㄅㄟ | bei | 19 | 4 |
| 75 | ㄆㄟ | pei | 20 | 4 |
| 76 | ㄇㄟ | mei | 21 | 4 |
| 77 | ㄈㄟ | fei | 22 | 4 |
| 78 | ㄠ | ao | 1 | 2 |
| 79 | ㄓㄠ | zhao | 2 | 2 |

| | | | | |
|---|---|---|---|---|
| 80 | ㄔㄠ | chao | 3 | 2 |
| 81 | ㄕㄠ | shao | 4 | 2 |
| 82 | ㄖㄠ | rao | 5 | 2 |
| 83 | ㄗㄠ | zao | 6 | 2 |
| 84 | ㄘㄠ | cao | 7 | 2 |
| 85 | ㄙㄠ | sao | 8 | 2 |
| 86 | ㄍㄠ | gao | 9 | 2 |
| 87 | ㄎㄠ | kao | 10 | 2 |
| 88 | ㄏㄠ | hao | 11 | 2 |
| 89 | ㄉㄠ | dao | 15 | 2 |
| 90 | ㄊㄠ | tao | 16 | 2 |
| 91 | ㄋㄠ | nao | 17 | 2 |
| 92 | ㄌㄠ | lao | 18 | 2 |
| 93 | ㄅㄠ | bao | 19 | 2 |
| 94 | ㄆㄠ | pao | 20 | 2 |
| 95 | ㄇㄠ | mao | 21 | 2 |
| 96 | ㄡ | ou | 1 | 3 |
| 97 | ㄓㄡ | zhou | 2 | 3 |
| 98 | ㄔㄡ | chou | 3 | 3 |
| 99 | ㄕㄡ | shou | 4 | 3 |
| 100 | ㄖㄡ | rou | 5 | 3 |
| 101 | ㄗㄡ | zou | 6 | 3 |
| 102 | ㄘㄡ | cou | 7 | 3 |
| 103 | ㄙㄡ | sou | 8 | 3 |
| 104 | ㄍㄡ | gou | 9 | 3 |
| 105 | ㄎㄡ | kou | 10 | 3 |
| 106 | ㄏㄡ | hou | 11 | 3 |
| 107 | ㄉㄡ | dou | 15 | 3 |
| 108 | ㄊㄡ | tou | 16 | 3 |
| 109 | ㄋㄡ | nou | 17 | 3 |
| 110 | ㄌㄡ | lou | 18 | 3 |
| 111 | ㄆㄡ | pou | 20 | 3 |
| 112 | ㄇㄡ | mou | 21 | 3 |
| 113 | ㄈㄡ | fou | 22 | 3 |
| 114 | ㄢ | an | 1 | 2 |
| 115 | ㄓㄢ | zhan | 2 | 2 |
| 116 | ㄔㄢ | chan | 3 | 2 |
| 117 | ㄕㄢ | shan | 4 | 2 |
| 118 | ㄖㄢ | ran | 5 | 2 |
| 119 | ㄗㄢ | zan | 6 | 2 |
| 120 | ㄘㄢ | can | 7 | 2 |
| 121 | ㄙㄢ | san | 8 | 2 |
| 122 | ㄍㄢ | gan | 9 | 2 |

| | | | | |
|---|---|---|---|---|
| 123 | ㄎㄢ | kan | 10 | 2 |
| 124 | ㄏㄢ | han | 11 | 2 |
| 125 | ㄉㄢ | dan | 15 | 2 |
| 126 | ㄊㄢ | tan | 16 | 2 |
| 127 | ㄋㄢ | nan | 17 | 2 |
| 128 | ㄌㄢ | lan | 18 | 2 |
| 129 | ㄅㄢ | ban | 19 | 2 |
| 130 | ㄆㄢ | pan | 20 | 2 |
| 131 | ㄇㄢ | man | 21 | 2 |
| 132 | ㄈㄢ | fan | 22 | 2 |
| 133 | ㄣ | en | 1 | 4 |
| 134 | ㄓㄣ | zhen | 2 | 4 |
| 135 | ㄔㄣ | chen | 3 | 4 |
| 136 | ㄕㄣ | shen | 4 | 4 |
| 137 | ㄖㄣ | ren | 5 | 4 |
| 138 | ㄗㄣ | zen | 6 | 4 |
| 139 | ㄘㄣ | cen | 7 | 4 |
| 140 | ㄙㄣ | sen | 8 | 4 |
| 141 | ㄍㄣ | gen | 9 | 4 |
| 142 | ㄎㄣ | ken | 10 | 4 |
| 143 | ㄏㄣ | hen | 11 | 4 |
| 144 | ㄋㄣ | nen | 17 | 4 |
| 145 | ㄅㄣ | ben | 19 | 4 |
| 146 | ㄆㄣ | pen | 20 | 4 |
| 147 | ㄇㄣ | men | 21 | 4 |
| 148 | ㄈㄣ | fen | 22 | 4 |
| 149 | ㄤ | ang | 1 | 2 |
| 150 | ㄓㄤ | zhang | 2 | 2 |
| 151 | ㄔㄤ | chang | 3 | 2 |
| 152 | ㄕㄤ | shang | 4 | 2 |
| 153 | ㄖㄤ | rang | 5 | 2 |
| 154 | ㄗㄤ | zang | 6 | 2 |
| 155 | ㄘㄤ | cang | 7 | 2 |
| 156 | ㄙㄤ | sang | 8 | 2 |
| 157 | ㄍㄤ | gang | 9 | 2 |
| 158 | ㄎㄤ | kang | 10 | 2 |
| 159 | ㄏㄤ | hang | 11 | 2 |
| 160 | ㄉㄤ | dang | 15 | 2 |
| 161 | ㄊㄤ | tang | 16 | 2 |
| 162 | ㄋㄤ | nang | 17 | 2 |
| 163 | ㄌㄤ | lang | 18 | 2 |
| 164 | ㄅㄤ | bang | 19 | 2 |
| 165 | ㄆㄤ | pang | 20 | 2 |

| | | | | |
|---|---|---|---|---|
| 166 | ㄇㄤ | mang | 21 | 2 |
| 167 | ㄈㄤ | fang | 22 | 2 |
| 168 | ㄓㄥ | zheng | 2 | 4 |
| 169 | ㄔㄥ | cheng | 3 | 4 |
| 170 | ㄕㄥ | sheng | 4 | 4 |
| 171 | ㄖㄥ | reng | 5 | 4 |
| 172 | ㄗㄥ | zeng | 6 | 4 |
| 173 | ㄘㄥ | ceng | 7 | 4 |
| 174 | ㄙㄥ | seng | 8 | 4 |
| 175 | ㄍㄥ | geng | 9 | 4 |
| 176 | ㄎㄥ | keng | 10 | 4 |
| 177 | ㄏㄥ | heng | 11 | 4 |
| 178 | ㄉㄥ | deng | 15 | 4 |
| 179 | ㄊㄥ | teng | 16 | 4 |
| 180 | ㄋㄥ | neng | 17 | 4 |
| 181 | ㄌㄥ | leng | 18 | 4 |
| 182 | ㄅㄥ | beng | 19 | 4 |
| 183 | ㄆㄥ | peng | 20 | 4 |
| 184 | ㄇㄥ | meng | 21 | 4 |
| 185 | ㄈㄥ | feng | 22 | 4 |
| 186 | ㄧ | yi | 1 | 5 |
| 187 | ㄐㄧ | ji | 12 | 5 |
| 188 | ㄑㄧ | qi | 13 | 5 |
| 189 | ㄒㄧ | xi | 14 | 5 |
| 190 | ㄉㄧ | di | 15 | 5 |
| 191 | ㄊㄧ | ti | 16 | 5 |
| 192 | ㄋㄧ | ni | 17 | 5 |
| 193 | ㄌㄧ | li | 18 | 5 |
| 194 | ㄅㄧ | bi | 19 | 5 |
| 195 | ㄆㄧ | pi | 20 | 5 |
| 196 | ㄇㄧ | mi | 21 | 5 |
| 197 | ㄨ | wu | 1 | 6 |
| 198 | ㄓㄨ | zhu | 2 | 6 |
| 199 | ㄔㄨ | chu | 3 | 6 |
| 200 | ㄕㄨ | shu | 4 | 6 |
| 201 | ㄖㄨ | ru | 5 | 6 |
| 202 | ㄗㄨ | zu | 6 | 6 |
| 203 | ㄘㄨ | cu | 7 | 6 |
| 204 | ㄙㄨ | su | 8 | 6 |
| 205 | ㄍㄨ | gu | 9 | 6 |
| 206 | ㄎㄨ | ku | 10 | 6 |
| 207 | ㄏㄨ | hu | 11 | 6 |
| 208 | ㄉㄨ | du | 15 | 6 |

| | | | | |
|---|---|---|---|---|
| 209 | ㄊㄨ | tu | 16 | 6 |
| 210 | ㄋㄨ | nu | 17 | 6 |
| 211 | ㄌㄨ | lu | 18 | 6 |
| 212 | ㄅㄨ | bu | 19 | 6 |
| 213 | ㄆㄨ | pu | 20 | 6 |
| 214 | ㄇㄨ | mu | 21 | 6 |
| 215 | ㄈㄨ | fu | 22 | 6 |
| 216 | ㄩ | yu | 1 | 7 |
| 217 | ㄐㄩ | ju | 12 | 7 |
| 218 | ㄑㄩ | qu | 13 | 7 |
| 219 | ㄒㄩ | xu | 14 | 7 |
| 220 | ㄋㄩ | nu: | 17 | 7 |
| 221 | ㄌㄩ | lu: | 18 | 7 |
| 222 | ㄧㄚ | ya | 1 | 5 |
| 223 | ㄐㄧㄚ | jia | 12 | 5 |
| 224 | ㄑㄧㄚ | qia | 13 | 5 |
| 225 | ㄒㄧㄚ | xia | 14 | 5 |
| 226 | ㄌㄧㄚ | lia | 18 | 5 |
| 227 | ㄧㄝ | ye | 1 | 5 |
| 228 | ㄐㄧㄝ | jie | 12 | 5 |
| 229 | ㄑㄧㄝ | qie | 13 | 5 |
| 230 | ㄒㄧㄝ | xie | 14 | 5 |
| 231 | ㄉㄧㄝ | die | 15 | 5 |
| 232 | ㄊㄧㄝ | tie | 16 | 5 |
| 233 | ㄋㄧㄝ | nie | 17 | 5 |
| 234 | ㄌㄧㄝ | lie | 18 | 5 |
| 235 | ㄅㄧㄝ | bie | 19 | 5 |
| 236 | ㄆㄧㄝ | pie | 20 | 5 |
| 237 | ㄇㄧㄝ | mie | 21 | 5 |
| 238 | ㄧㄞ | yai | 1 | 5 |
| 239 | ㄧㄠ | yao | 1 | 5 |
| 240 | ㄐㄧㄠ | jiao | 12 | 5 |
| 241 | ㄑㄧㄠ | qiao | 13 | 5 |
| 242 | ㄒㄧㄠ | xiao | 14 | 5 |
| 243 | ㄉㄧㄠ | diao | 15 | 5 |
| 244 | ㄊㄧㄠ | tiao | 16 | 5 |
| 245 | ㄋㄧㄠ | niao | 17 | 5 |
| 246 | ㄌㄧㄠ | liao | 18 | 5 |
| 247 | ㄅㄧㄠ | biao | 19 | 5 |
| 248 | ㄆㄧㄠ | piao | 20 | 5 |
| 249 | ㄇㄧㄠ | miao | 21 | 5 |
| 250 | ㄧㄡ | you | 1 | 5 |
| 251 | ㄐㄧㄡ | jiu | 12 | 5 |

| | | | | |
|---|---|---|---|---|
| 252 | ㄑㄧㄡ | qiu | 13 | 5 |
| 253 | ㄒㄧㄡ | xiu | 14 | 5 |
| 254 | ㄉㄧㄡ | diu | 15 | 5 |
| 255 | ㄋㄧㄡ | niu | 17 | 5 |
| 256 | ㄌㄧㄡ | liu | 18 | 5 |
| 257 | ㄇㄧㄡ | miu | 21 | 5 |
| 258 | ㄧㄢ | yan | 1 | 5 |
| 259 | ㄐㄧㄢ | jian | 12 | 5 |
| 260 | ㄑㄧㄢ | qian | 13 | 5 |
| 261 | ㄒㄧㄢ | xian | 14 | 5 |
| 262 | ㄉㄧㄢ | dian | 15 | 5 |
| 263 | ㄊㄧㄢ | tian | 16 | 5 |
| 264 | ㄋㄧㄢ | nian | 17 | 5 |
| 265 | ㄌㄧㄢ | lian | 18 | 5 |
| 266 | ㄅㄧㄢ | bian | 19 | 5 |
| 267 | ㄆㄧㄢ | pian | 20 | 5 |
| 268 | ㄇㄧㄢ | mian | 21 | 5 |
| 269 | ㄧㄣ | yin | 1 | 5 |
| 270 | ㄐㄧㄣ | jin | 12 | 5 |
| 271 | ㄑㄧㄣ | qin | 13 | 5 |
| 272 | ㄒㄧㄣ | xin | 14 | 5 |
| 273 | ㄋㄧㄣ | nin | 17 | 5 |
| 274 | ㄌㄧㄣ | lin | 18 | 5 |
| 275 | ㄅㄧㄣ | bin | 19 | 5 |
| 276 | ㄆㄧㄣ | pin | 20 | 5 |
| 277 | ㄇㄧㄣ | min | 21 | 5 |
| 278 | ㄧㄤ | yang | 1 | 5 |
| 279 | ㄐㄧㄤ | jiang | 12 | 5 |
| 280 | ㄑㄧㄤ | qiang | 13 | 5 |
| 281 | ㄒㄧㄤ | xiang | 14 | 5 |
| 282 | ㄋㄧㄤ | niang | 17 | 5 |
| 283 | ㄌㄧㄤ | liang | 18 | 5 |
| 284 | ㄧㄥ | ying | 1 | 5 |
| 285 | ㄐㄧㄥ | jing | 12 | 5 |
| 286 | ㄑㄧㄥ | qing | 13 | 5 |
| 287 | ㄒㄧㄥ | xing | 14 | 5 |
| 288 | ㄉㄧㄥ | ding | 15 | 5 |
| 289 | ㄊㄧㄥ | ting | 16 | 5 |
| 290 | ㄋㄧㄥ | ning | 17 | 5 |
| 291 | ㄌㄧㄥ | ling | 18 | 5 |
| 292 | ㄅㄧㄥ | bing | 19 | 5 |
| 293 | ㄆㄧㄥ | ping | 20 | 5 |
| 294 | ㄇㄧㄥ | ming | 21 | 5 |

| | | | | |
|---|---|---|---|---|
| 295 | ㄨㄚ | wa | 1 | 6 |
| 296 | ㄓㄨㄚ | zhua | 2 | 6 |
| 297 | ㄔㄨㄚ | chua | 3 | 6 |
| 298 | ㄕㄨㄚ | shua | 4 | 6 |
| 299 | ㄍㄨㄚ | gua | 9 | 6 |
| 300 | ㄎㄨㄚ | kua | 10 | 6 |
| 301 | ㄏㄨㄚ | hua | 11 | 6 |
| 302 | ㄨㄛ | wo | 1 | 6 |
| 303 | ㄓㄨㄛ | zhuo | 2 | 6 |
| 304 | ㄔㄨㄛ | chuo | 3 | 6 |
| 305 | ㄕㄨㄛ | shuo | 4 | 6 |
| 306 | ㄖㄨㄛ | ruo | 5 | 6 |
| 307 | ㄗㄨㄛ | zuo | 6 | 6 |
| 308 | ㄘㄨㄛ | cuo | 7 | 6 |
| 309 | ㄙㄨㄛ | suo | 8 | 6 |
| 310 | ㄍㄨㄛ | guo | 9 | 6 |
| 311 | ㄎㄨㄛ | kuo | 10 | 6 |
| 312 | ㄏㄨㄛ | huo | 11 | 6 |
| 313 | ㄉㄨㄛ | duo | 15 | 6 |
| 314 | ㄊㄨㄛ | tuo | 16 | 6 |
| 315 | ㄋㄨㄛ | nuo | 17 | 6 |
| 316 | ㄌㄨㄛ | luo | 18 | 6 |
| 317 | ㄨㄞ | wai | 1 | 6 |
| 318 | ㄓㄨㄞ | zhuai | 2 | 6 |
| 319 | ㄔㄨㄞ | chuai | 3 | 6 |
| 320 | ㄕㄨㄞ | shuai | 4 | 6 |
| 321 | ㄍㄨㄞ | guai | 9 | 6 |
| 322 | ㄎㄨㄞ | kuai | 10 | 6 |
| 323 | ㄏㄨㄞ | huai | 11 | 6 |
| 324 | ㄨㄟ | wei | 1 | 6 |
| 325 | ㄓㄨㄟ | zhui | 2 | 6 |
| 326 | ㄔㄨㄟ | chui | 3 | 6 |
| 327 | ㄕㄨㄟ | shui | 4 | 6 |
| 328 | ㄖㄨㄟ | rui | 5 | 6 |
| 329 | ㄗㄨㄟ | zui | 6 | 6 |
| 330 | ㄘㄨㄟ | cui | 7 | 6 |
| 331 | ㄙㄨㄟ | sui | 8 | 6 |
| 332 | ㄍㄨㄟ | gui | 9 | 6 |
| 333 | ㄎㄨㄟ | kui | 10 | 6 |
| 334 | ㄏㄨㄟ | hui | 11 | 6 |
| 335 | ㄉㄨㄟ | dui | 15 | 6 |
| 336 | ㄊㄨㄟ | tui | 16 | 6 |
| 337 | ㄨㄢ | wan | 1 | 6 |

| | | | | |
|---|---|---|---|---|
| 338 | ㄓㄨㄢ | zhuan | 2 | 6 |
| 339 | ㄔㄨㄢ | chuan | 3 | 6 |
| 340 | ㄕㄨㄢ | shuan | 4 | 6 |
| 341 | ㄖㄨㄢ | ruan | 5 | 6 |
| 342 | ㄗㄨㄢ | zuan | 6 | 6 |
| 343 | ㄘㄨㄢ | cuan | 7 | 6 |
| 344 | ㄙㄨㄢ | suan | 8 | 6 |
| 345 | ㄍㄨㄢ | guan | 9 | 6 |
| 346 | ㄎㄨㄢ | kuan | 10 | 6 |
| 347 | ㄏㄨㄢ | huan | 11 | 6 |
| 348 | ㄉㄨㄢ | duan | 15 | 6 |
| 349 | ㄊㄨㄢ | tuan | 16 | 6 |
| 350 | ㄋㄨㄢ | nuan | 17 | 6 |
| 351 | ㄌㄨㄢ | luan | 18 | 6 |
| 352 | ㄨㄣ | wen | 1 | 6 |
| 353 | ㄓㄨㄣ | zhun | 2 | 6 |
| 354 | ㄔㄨㄣ | chun | 3 | 6 |
| 355 | ㄕㄨㄣ | shun | 4 | 6 |
| 356 | ㄖㄨㄣ | run | 5 | 6 |
| 357 | ㄗㄨㄣ | zun | 6 | 6 |
| 358 | ㄘㄨㄣ | cun | 7 | 6 |
| 359 | ㄙㄨㄣ | sun | 8 | 6 |
| 360 | ㄍㄨㄣ | gun | 9 | 6 |
| 361 | ㄎㄨㄣ | kun | 10 | 6 |
| 362 | ㄏㄨㄣ | hun | 11 | 6 |
| 363 | ㄉㄨㄣ | dun | 15 | 6 |
| 364 | ㄊㄨㄣ | tun | 16 | 6 |
| 365 | ㄌㄨㄣ | lun | 18 | 6 |
| 366 | ㄨㄤ | wang | 1 | 6 |
| 367 | ㄓㄨㄤ | zhuang | 2 | 6 |
| 368 | ㄔㄨㄤ | chuang | 3 | 6 |
| 369 | ㄕㄨㄤ | shuang | 4 | 6 |
| 370 | ㄍㄨㄤ | guang | 9 | 6 |
| 371 | ㄎㄨㄤ | kuang | 10 | 6 |
| 372 | ㄏㄨㄤ | huang | 11 | 6 |
| 373 | ㄨㄥ | weng | 1 | 6 |
| 374 | ㄓㄨㄥ | zhong | 2 | 6 |
| 375 | ㄔㄨㄥ | chong | 3 | 6 |
| 376 | ㄖㄨㄥ | rong | 5 | 6 |
| 377 | ㄗㄨㄥ | zong | 6 | 6 |
| 378 | ㄘㄨㄥ | cong | 7 | 6 |
| 379 | ㄙㄨㄥ | song | 8 | 6 |
| 380 | ㄍㄨㄥ | gong | 9 | 6 |

| | | | | |
|---|---|---|---|---|
| 381 | ㄎㄨㄥ | kong | 10 | 6 |
| 382 | ㄏㄨㄥ | hong | 11 | 6 |
| 383 | ㄉㄨㄥ | dong | 15 | 6 |
| 384 | ㄊㄨㄥ | tong | 16 | 6 |
| 385 | ㄋㄨㄥ | nong | 17 | 6 |
| 386 | ㄌㄨㄥ | long | 18 | 6 |
| 387 | ㄩㄝ | yue | 1 | 7 |
| 388 | ㄐㄩㄝ | jue | 12 | 7 |
| 389 | ㄑㄩㄝ | que | 13 | 7 |
| 390 | ㄒㄩㄝ | xue | 14 | 7 |
| 391 | ㄋㄩㄝ | nu:e | 17 | 7 |
| 392 | ㄌㄩㄝ | lu:e | 18 | 7 |
| 393 | ㄩㄢ | yuan | 1 | 7 |
| 394 | ㄐㄩㄢ | juan | 12 | 7 |
| 395 | ㄑㄩㄢ | quan | 13 | 7 |
| 396 | ㄒㄩㄢ | xuan | 14 | 7 |
| 397 | ㄌㄩㄢ | lu:an | 18 | 7 |
| 398 | ㄩㄣ | yun | 1 | 7 |
| 399 | ㄐㄩㄣ | jun | 12 | 7 |
| 400 | ㄑㄩㄣ | qun | 13 | 7 |
| 401 | ㄒㄩㄣ | xun | 14 | 7 |
| 402 | ㄌㄩㄣ | l:un | 18 | 7 |
| 403 | ㄩㄥ | yong | 1 | 7 |
| 404 | ㄐㄩㄥ | jiong | 12 | 7 |
| 405 | ㄑㄩㄥ | qiong | 13 | 7 |
| 406 | ㄒㄩㄥ | xiong | 14 | 7 |
| 407 | ㄦ | er | 1 | 4 |
| 408 | ㄧㄛ | yo | 1 | 5 |
| 409 | ㄥ | eng | 1 | 4 |
| 410 | ㄟ | ei | 1 | 4 |
| 411 | ㄇㄜ | me | 21 | 4 |

# 附錄二　Treebank 語料庫統計數據

各項數據列表：

| 項目 | 個數 |
|---|---|
| 中文文句結構樹 | 11,109 棵 |
| （原始）詞 | 69,062 個 |
| 中文字（音節） | 123,128 字，3,137 種類 |
| 短文 | 1,433 篇 |
| 音調音節 | 1,068 種類(82.15%) |

錄音設備與錄音環境：

| | |
|---|---|
| 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) | 20 kHz |
| 發音速度 | 每秒約 4.6 個音節 |
| 取樣大小 | 16 bits (位元) |
| 聲道 | 單聲道(mono) |
| 檔案格式 | Pcm |
| 能量（句平均） | 平均 60.81dB，最小 52.18dB，最大 66.48dB |

# 附錄三　詞綴清單與統計數據

**表 C.1　前詞綴清單**

| 前詞綴列表 | 在語料庫中的個數 |
|---|---|
| 也 | 1 |
| 大 | 36 |
| 女 | 3 |
| 不 | 42 |
| 公 | 2 |
| 反 | 2 |
| 太 | 1 |
| 半 | 2 |
| 可 | 1 |
| 另 | 7 |
| 只 | 3 |
| 正 | 2 |
| 再 | 1 |
| 回 | 1 |
| 好 | 2 |
| 有 | 4 |
| 老 | 8 |
| 自 | 2 |
| 改 | 3 |
| 更 | 2 |
| 每 | 1 |
| 男 | 1 |
| 那 | 3 |
| 金 | 4 |
| 青 | 5 |
| 活 | 2 |
| 相 | 3 |
| 紅 | 3 |
| 負 | 4 |
| 重 | 2 |

| | |
|---|---|
| 原 | 9 |
| 核 | 1 |
| 特 | 1 |
| 純 | 1 |
| 高 | 4 |
| 清 | 1 |
| 產 | 2 |
| 這 | 10 |
| 最 | 2 |
| 單 | 2 |
| 就 | 1 |
| 棉 | 1 |
| 無 | 2 |
| 發 | 2 |
| 短 | 1 |
| 硬 | 1 |
| 超 | 1 |
| 進 | 3 |
| 黃 | 2 |
| 黑 | 2 |
| 微 | 1 |
| 新 | 8 |
| 會 | 4 |
| 電 | 2 |
| 網 | 2 |
| 誤 | 1 |
| 輕 | 4 |
| 養 | 2 |
| 親 | 1 |
| 總 | 52 |
| 還 | 1 |
| 舊 | 1 |
| 轉 | 4 |

## 表 C.2　後詞綴清單

| 後詞綴列表 | 在語料庫中的個數 |
|---|---|
| 人 | 222 |
| 入 | 1 |
| 力 | 15 |
| 上 | 1 |
| 子 | 8 |
| 山 | 2 |
| 化 | 25 |
| 心 | 1 |
| 手 | 2 |
| 片 | 2 |
| 出 | 16 |
| 去 | 8 |
| 台 | 7 |
| 生 | 7 |
| 石 | 1 |
| 名 | 4 |
| 回 | 1 |
| 地 | 9 |
| 好 | 2 |
| 式 | 16 |
| 成 | 7 |
| 曲 | 2 |
| 色 | 3 |
| 衣 | 4 |
| 形 | 4 |
| 系 | 4 |
| 角 | 1 |
| 來 | 23 |
| 到 | 36 |
| 味 | 1 |
| 性 | 66 |
| 板 | 5 |
| 河 | 3 |

| | |
|---|---|
| 波 | 7 |
| 法 | 8 |
| 油 | 3 |
| 版 | 3 |
| 物 | 9 |
| 表 | 7 |
| 長 | 64 |
| 型 | 1 |
| 室 | 5 |
| 度 | 7 |
| 派 | 10 |
| 科 | 11 |
| 軍 | 3 |
| 面 | 34 |
| 島 | 2 |
| 師 | 9 |
| 書 | 10 |
| 班 | 5 |
| 級 | 6 |
| 起 | 1 |
| 區 | 24 |
| 商 | 18 |
| 國 | 57 |
| 掉 | 1 |
| 率 | 22 |
| 球 | 25 |
| 產 | 4 |
| 組 | 29 |
| 處 | 17 |
| 袋 | 1 |
| 部 | 77 |
| 期 | 8 |
| 給 | 2 |
| 腔 | 2 |
| 費 | 8 |
| 進 | 1 |
| 量 | 25 |

| | |
|---|---|
| 開 | 1 |
| 隊 | 196 |
| 感 | 14 |
| 會 | 132 |
| 業 | 18 |
| 節 | 12 |
| 群 | 7 |
| 路 | 11 |
| 團 | 12 |
| 槍 | 1 |
| 網 | 2 |
| 舞 | 2 |
| 數 | 8 |
| 熱 | 2 |
| 獎 | 5 |
| 篇 | 1 |
| 論 | 2 |
| 質 | 1 |
| 劑 | 2 |
| 學 | 4 |
| 樹 | 1 |
| 頭 | 1 |
| 聲 | 3 |
| 點 | 15 |
| 類 | 5 |
| 欄 | 3 |
| 權 | 18 |
| 觀 | 3 |

# 附錄四　音節相關前後文變數向量分類方式與統計數據

參照 3.5.3.1 節，關於音節相關前後文變數的討論。以下是各變數的分類方式，其中子音與母音的分類編號請參照附錄一，國語 411 基本音節總音表。

**表 D.1　前一音節結尾類型分類表**

| 前一音節結尾類型 |
|---|
| 母音類型編號 1 |
| 母音類型編號 2 |
| 母音類型編號 3 |
| 母音類型編號 4 |
| 母音類型編號 5 |
| 母音類型編號 6 |
| 母音類型編號 7 |
| 靜音 |

**表 D.2　後一音節開頭類型分類表**

| 後一音節開頭類型 |
|---|
| 子音類型編號 1 |
| 子音類型編號 2 |
| 子音類型編號 3 |
| 子音類型編號 4 |
| 子音類型編號 5 |
| 子音類型編號 6 |
| 子音類型編號 7 |
| 子音類型編號 8 |
| 子音類型編號 9 |
| 子音類型編號 10 |

| |
|---|
| 子音類型編號 11 |
| 子音類型編號 12 |
| 子音類型編號 13 |
| 子音類型編號 14 |
| 子音類型編號 15 |
| 子音類型編號 16 |
| 子音類型編號 17 |
| 子音類型編號 18 |
| 子音類型編號 19 |
| 子音類型編號 20 |
| 子音類型編號 21 |
| 子音類型編號 22 |
| 靜音 |

**表 D.3　前一音節音調分類表**

| 前一音節音調 | 包含音調類型 |
|---|---|
| 無前一音節(None Left) | -- |
| 高結尾(High Ending) | 一聲、二聲 |
| 低結尾(Low Ending) | 三聲、四聲、五聲 |

**表 D.4　後一音節音調分類表**

| 後一音節音調 | 包含音調類型 |
|---|---|
| 無後一音節(None Right) | -- |
| 高起頭(High Starting) | 一聲、四聲 |
| 低起頭(Low Starting) | 二聲、三聲、五聲 |

**表 D.5　位於詞中的位置分類表**

| 位於詞中的位置 |
|---|
| 詞首（Initial） |
| 詞中（Middle） |
| 詞尾（Final） |
| 單字詞（Mono） |

# 附錄五 中文姓氏清單與統計數據

**表 E　中文姓氏清單**

| 姓氏 | 出現次數 |
|---|---|
| 李 | 48 |
| 陳 | 45 |
| 吳 | 42 |
| 林 | 32 |
| 張 | 31 |
| 蔣 | 29 |
| 王 | 28 |
| 黃 | 28 |
| 羅 | 26 |
| 郝 | 25 |
| 劉 | 17 |
| 孫 | 15 |
| 金 | 14 |
| 謝 | 14 |
| 楊 | 13 |
| 康 | 12 |
| 丁 | 11 |
| 宋 | 11 |
| 蔡 | 11 |
| 江 | 10 |
| 鄭 | 10 |
| 鄧 | 10 |
| 呂 | 9 |
| 許 | 9 |
| 廖 | 9 |
| 崔 | 8 |
| 朱 | 7 |
| 周 | 7 |
| 高 | 7 |
| 郭 | 7 |
| 葉 | 7 |
| 顧 | 7 |

| | |
|---|---|
| 胡 | 6 |
| 曾 | 6 |
| 何 | 5 |
| 錢 | 5 |
| 尹 | 4 |
| 沈 | 4 |
| 唐 | 4 |
| 趙 | 4 |
| 于 | 3 |
| 邱 | 3 |
| 姜 | 3 |
| 蕭 | 3 |
| 賴 | 3 |
| 戴 | 3 |
| 汪 | 2 |
| 范 | 2 |
| 徐 | 2 |
| 馬 | 2 |
| 梁 | 2 |
| 傅 | 2 |
| 彭 | 2 |
| 薛 | 2 |
| 毛 | 1 |
| 石 | 1 |
| 余 | 1 |
| 夏 | 1 |
| 秦 | 1 |
| 袁 | 1 |
| 常 | 1 |
| 曹 | 1 |
| 陸 | 1 |
| 雷 | 1 |
| 潘 | 1 |
| 韓 | 1 |
| 魏 | 1 |
| 蘇 | 1 |

# 附錄六 挑選單元中各變數之統計數據

在此附錄中列舉在 4.1.1 節中所提到之以測試資料統計的各變數資料。這些變數分別是 $d^j_{pitch\_mean}$ 、$d^j_{duration}$ 、$d^j_{power}$ 及 $d^j_{contextual}$ ，其符號及定義同於第三章。

變數 $d^j_{pitch\_mean}$ 的統計數據：

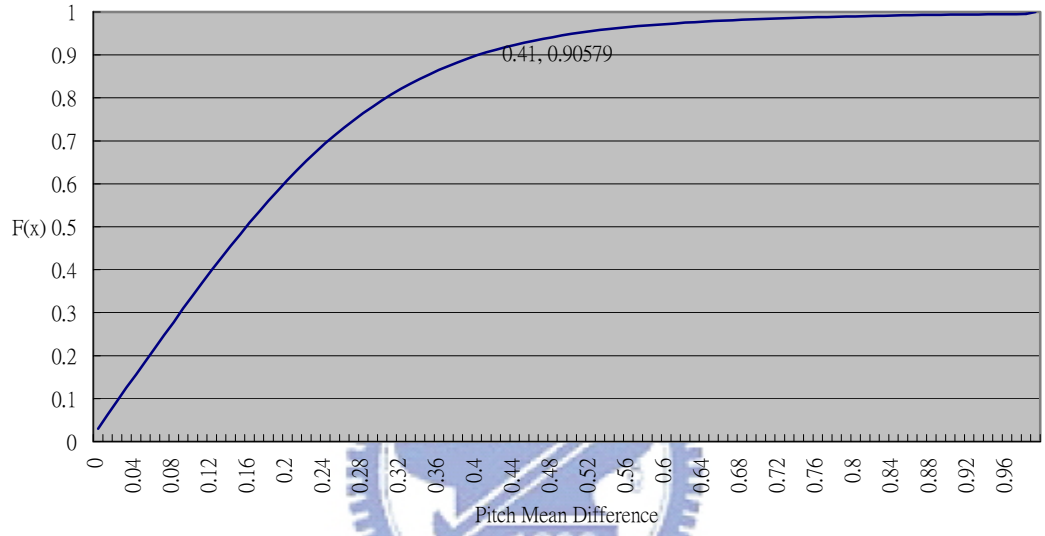



**圖 4-1-1：The cumulative distribution function of Pitch-Mean Difference**

變數 $d^j_{duration}$ 的統計數據：

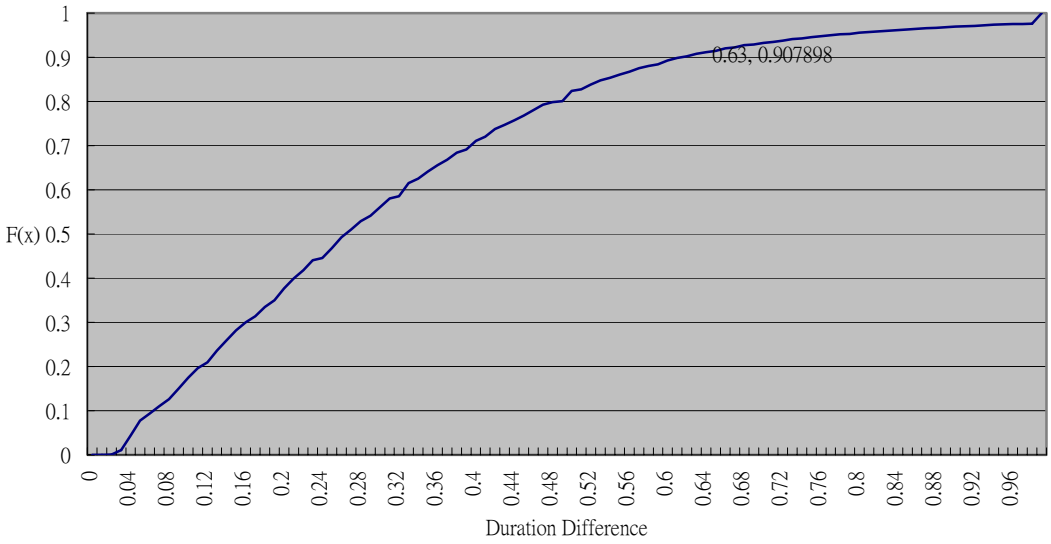



**圖 F.1: The cumulative distribution function of Duration Difference**

變數 $d^{j}_{power}$ 的統計數據：

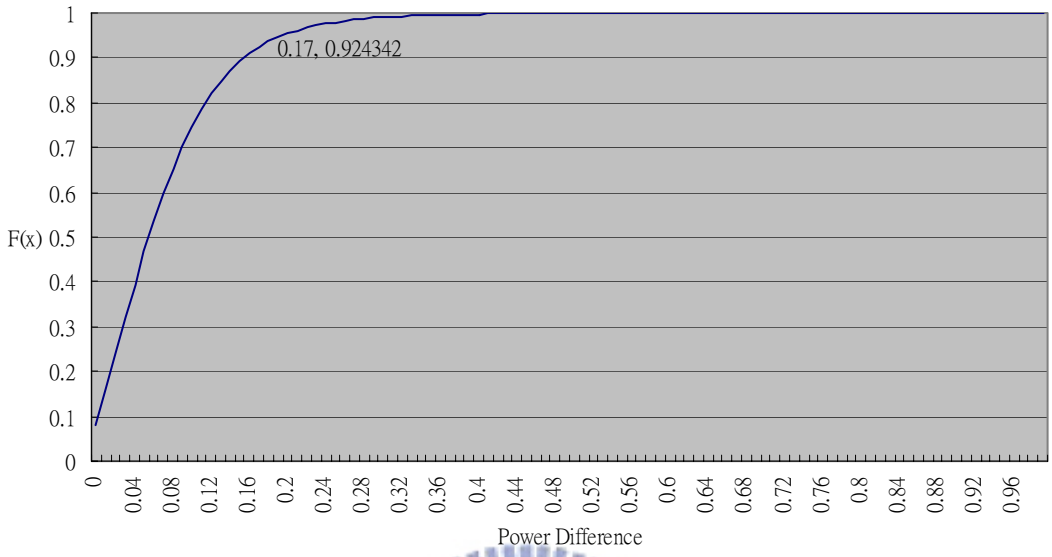



**圖 F.2: The cumulative distribution function of Power Difference**

變數 $d^{j}_{contextual}$ 的統計數據：

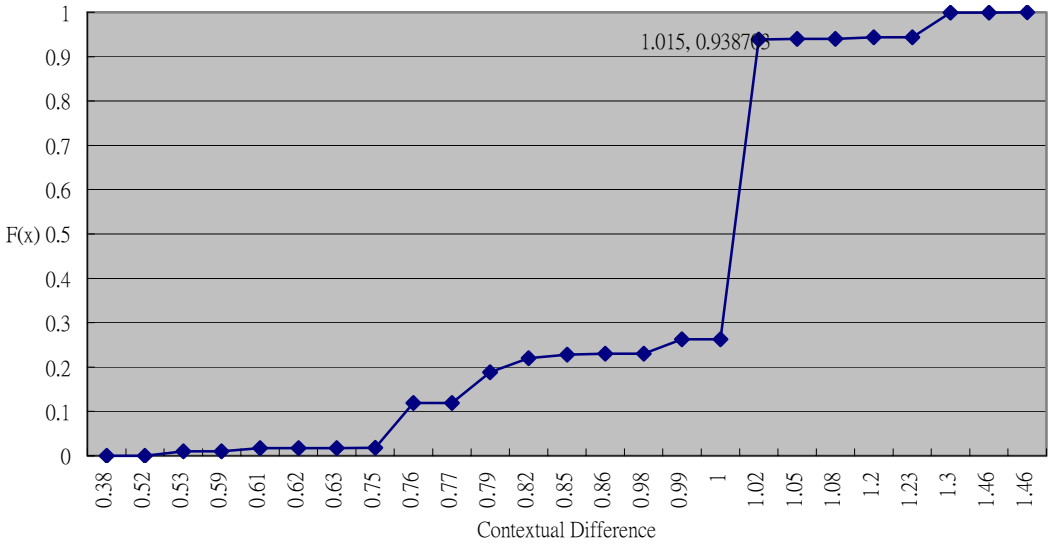



**圖 F.3: The cumulative distribution function of Contextual Difference**