國立交通大學

電信工程研究所

碩士論文

台語語音合成技術之研究

The Research of Speech Synthesis Technology for Taiwanese

研究生：趙良基

指導教授：陳信宏博士

中華民國一百零二年七月

台語語音合成技術之研究

The Research of Speech Synthesis Technology for Taiwanese

研究生：趙良基
指導教授：陳信宏博士

Student：Liang-Ji Chao
Advisor：Dr. Sin-Horng Chen

國立交通大學
電信工程研究所
碩士論文

A Thesis
Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
In

Communication Engineering

June 2013
Hsinchu, Taiwan, Republic of China

中華民國一百零二年七月

台語語音合成技術之研究

研究生：趙良基　　指導教授:陳信宏　博士

國立交通大學電信工程研究所碩士班

中文摘要

本論文利用基於隱藏式馬可夫模型實作台語語音合成系統，並合併國語字典與台語字典建立出一國語對台語翻譯字典，我們利用此國對台翻譯字典將中文文章翻譯為台語文章，再將翻譯後的文章，輸入至台語語音合成系統，實作出一中文文章輸入台語語音合成輸出系統。

最後考慮到台語與中文語法結構上的相似性，我們訓練一中文韻律階層式模型，並透過此模型之訓練結果，來對台語語料庫分析聲調影響台語的音節長度、音節能量及台語停頓分佈情形等特性。

The Research of Speech Synthesis Technology for Taiwanese

Student：Liang-ji Chao Advisor:Dr. Sin-Horng Chen

Institute of Communication Engineering

National Chiao Tung University

Abstract

In this thesis, we propose a Taiwanese speech synthesis system based on Hidden Markov Model and combine Mandarin dictionary and Taiwanese dictionary to build up a Mandarin-to-Taiwanese translation dictionary. With this dictionary, Mandarin articles can be translated into Taiwanese articles as input for Taiwanese speech synthesis system. Hence, a Taiwanese speech synthesis system with Mandarin input article is implemented.

Consider the similarities on the grammar structure of Mandarin and Taiwanese, we train a Mandarin Hierarchical prosody model and use the training result to analysis the Taiwanese tone effect on Taiwanese syllable duration,syllable energy and pause distribution of Taiwanese speech database.

致謝

終於畢業了!!感謝陳信宏和王逸如老師，感謝陳老師每次不管再怎麼忙，總是會抽空找大家咪聽以及來實驗室督促我們的進度，每次和老師討論時總是會看到老師相當的興奮並激動，讓我也不禁對研究充滿了許多抱負;感謝王老師讓我真的體會到研究應該要怎麼做，如果只是想一步登天，到時候怎麼掛掉的都不知道，碩一時如果沒有王老師的教導，我想我的基礎離研究還有一大步吧，是你，王老師，若不是你的諄諄教誨不會有今天的我。

接著要感謝的當然是 707 神人，性獸大大，雖然現在你已不在交大，但每次我有問題時，你總還是會跟我討論並教我該如何改善，這研究如果沒有了你，我想我不會這麼順利，實在是太感謝了!!感謝將我帶入語音合成殿堂的喬華學長，你真的很厲害，我的程式能力如果能有你一半就好了;感謝熱情活潑的靜觀,實驗室有你真的相當開心順利，異常熱鬧;感謝人很好的小蝦，常常陪我一起健身，雖然我不像你有人魚線，但該有的線我可沒少;感謝內在外在都很優的子睿，你井然有序分析事情的能力相當吸引人，碩一的課如果沒了你，真不知道我現在會在哪裡;感謝胸肌扛壩子奕勳，雖然感覺到你的反社會，不過不知道為什麼就是很想找你出去玩 哈哈;感謝平頭阿龐，從碩一開始就當鄰居,雖然有了秀秀的加入讓我跟你相處的時間少了許多 QQ,不過跟你聊天真的很開心;感謝婉君，你直率不做作的個性，一直是我很喜歡的個性，我真的對你很好，不要把我當壞人好嗎...;也感謝在我碩二時陪我打拼的學弟妹，聲鋒你的程式能力不錯，相信大老闆已經看上你了;曾一起為了 prosody 奮鬥的王柏，讓我這個 prosody 小弱弱學到了不少;用遍了所有交友 APP 的茂隆，打球不是打架，好險葛格夠厚不然就重傷了;707 吳宗憲丫璋，你實在是很有趣，常常會因為你的一個梗讓我笑很久;比我高很多的阿俊，籃球這麼強，還不跟我打球!!分享很多東西給我的仲毛，你真是我心靈上的好朋友阿;來去無蹤影的佩樺，哥不能再幫你擋了，你要加油，台語很有趣的。感謝陪我經歷過研究所生涯的各位，這段日子很開心，讓研究所一點都不無聊。

另外，也感謝大學的朋友，雖然大家都到了不同地方，不過還是常聚在一起，讓我青春熱血的感覺還沒消失，你們是我最快樂的存在，不管未來會怎樣，我們感情永遠不變。

最後，感謝我的父母，雖然很常回台北，但總沒待在家過，不過我知道你們永遠在背後支持著我做的任何決定，感謝你們從小對我的栽培，在此謹以此論文獻給你們。

目錄

表目錄

圖目錄

第一章 緒論

1.1 研究背景、動機

台語在台灣是常被使用的一種方言，或許有些人的祖先並不是閩南人，但在台灣生活久了之後，也漸漸學會說個一兩句台語，可見台語在台灣的影響力有多麼的深遠。

隨著現今科技的發展，文字轉語音技術可以應用在很多地方，例如:導航系統、語言學習機或者是幫助視障朋友使用電腦等相關電子設備的軟體等，可以改善傳統的鍵盤輸入，螢幕輸出人機溝通方式，改用更直覺的語音人機介面，透過語音辨識使語音當作輸入，語音合成則可用來作為語音輸出，使得與機器的溝通上更加容易。但台灣目前的語音合成大部分皆以中文為主流研究對象，關於台語語音合成相關研究就相對較少，由於台語在台灣文化中有其不可或缺的重要性，因此若能在中文的語音合成系統加上台語語音合成系統，就能讓中文文章的語音合成有更多的選擇性。

近期語音合成系統廣為使用的合成方式主要有兩種，分別是大型語料庫(Corpus-based)(Chou et al.,2002)及隱藏式馬可夫模型(HMM-based approach)(Tokuda et al.,2000)的語音合成方法;大型語料庫合成語音方法為由已錄製好的語料庫中，挑選出適當的語音信號片段串接合成，因此是以原音來做呈現，擁有極佳的的合成品質，但是如果要合成出各種不同組合的語音，則需要錄製大量的語料來作為挑選單元的基礎，要收集如此眾多不同組合之語料並不是一件容易的事，因此，對於擁有豐富語言特性的台語來說，單元選取並不是一個適合的方法。

基於隱藏式馬可夫模型語音合成器是一種統計式的參數語音合成方法，也是目前最被廣泛使用的合成方法，它以文脈相關隱藏式馬可夫模型(Context-dependent HMMs,CDHMMs)來模擬不同語言參數下的聲學信號，從語料庫訓練得到頻譜模型(spectral parameter model)、基頻模型(F0 parameter model)及音長模型(duration model)。欲

合成語音時，利用上述已經訓練好的三種模型，依據輸入文本的語言參數找到適當的 CDHMM 模型並串接之，再以特殊的演算法由串接之 CDHMM 參數產生 frame spectrum 及 frame F0 參數，最後將 spectral 和 f0 參數輸入 MLSA 濾波器(Mel Log Spectrum Approximation filter)(Imai,1983)輸出合成訊號。

當想要以已經訓練好的現有模型去合成出不同特性的語音訊號時，則可利用調整參數的方式達到目的，如內插法(interpolation methods)(Yoshimura et al.,2000)、調適(adaptation methods)(Tamura et al.,21)。

跟單元選取方法是不同的，使用隱藏式馬可夫模型合成方法，並不需要太大的語料庫，只需要足夠的語料就能利用現有的隱藏式馬可夫模型去合成出不同特性的語音訊號，因此考慮到實驗室所收集到的台語語料庫有限以及台語語音豐富的語言特性等因素，本論文以隱藏式馬可夫模型合成來作為本論文之方法。

1.2 研究方向

本論文考慮到台語與中文的語法架構相似性，以及台語文字並沒有統一的情形下，造成的文字歧異性，因此希望能夠利用文字較統一的中文文章來當為台語語音合成系統的文字輸入，並利用中文的文字分析器先將中文文章做斷詞動作，接著在不變動此斷詞結果的情況下，將中文詞一個個依據本論文所做出之國台語對照字典對照出相對應之台語詞，因此斷詞好的中文文章會轉換成與中文斷詞結構相同的台語文章，接著將轉換後的文章輸入到 HTS(HMM-based speech synthesis system toolkit)系統中合成出台語語音，實作出一台語語音合成系統，論文最後利用中文韻律模型的訓練結果，來分析台語的音調、停頓邊界等語言現象。

1.3 章節概要說明

本論文一共分為五個章節，各章節內容分配如下:

第一章:緒論:介紹本論文之研究背景、動機與研究方向

第二章:台語語料庫介紹:介紹台語語音特性、變調規則、國台語字典的建立及本論文所使用之語料庫

第三章:台語語音合成系統實作:介紹基於 HMM 之語音合成系統原理及實作出一個中文文字轉台語語音合成系統

第四章:利用階層式韻律模型分析台語語料庫:本章節利用中文階層式韻律模型對本論文所使用之語料庫做分析

第五章:結論與未來展望:對本篇論文提出的語音合成系統做結論，並說明未來的改進方向

第二章　台語語料庫介紹

2.1 台語語音特性

台語和國語一樣皆為聲調語言(tonal language)，每個音節都由聲母、韻母和聲調三個因素所組成。

表 2-1 音節組成因素

<table>
<tr><td colspan="4">聲調</td></tr>
<tr><td rowspan="2">聲母</td><td colspan="3">韻母</td></tr>
<tr><td>韻首</td><td>韻腹</td><td>韻尾</td></tr>
</table>

聲母出現在音節前端，所以也叫做首音，台語共有18個聲母，其中「零聲母」並非沒有聲母，而是喉塞音，根據發音唇齒相關位置、清音濁音、送氣不送氣，分類列表如下，包括漢語拼音及[範例文字]表示法:

表 2-2 台語語音中的聲母分類表

發音方法 / 發音部位	清音				擦音	鼻音	濁音	
	塞音		塞擦音				邊音及零聲母	塞音及塞擦音
	不送氣	送氣	不送氣	送氣				
雙唇	p[兵]	ph[偏]				m[滿]		b[文]
舌尖中	t[斗]	th[天]				n[卵]	l[來]	
舌尖前			ch[走]	chh[春]	s[沙]			j[認]
舌根	k[該]	kh[開]				ng[硬]		g[眼]
喉					h[喜]		0[安]	

去掉聲母之後，剩下的部分稱為韻母，韻母可細分為韻頭、韻腹、韻尾三個部分，其中韻腹是每個韻母都有的，韻首或韻尾則不一定。韻腹和韻尾都屬於元音，聲帶震動，音強較大，在波形上可看到較大的振幅，呈現週期性。台語一共有68個韻母，分別為開尾韻、鼻聲韻、普通入聲韻、音節性輔音、喉塞入聲韻五類，其中四類再依據開口、齊齒及合口細分，分類列表如下，包括漢語拼音及[範例文字]表示法:

表 2-3 台語語音中的韻母分類表

一、開尾韻

開口		a[佳]	o[島]	oo[姑]	e[洗]	ai[泰]	au[草]	
齊齒	i[止]	ia[爺]	io[搖]		ue[瓜]		iau[少]	iu[樹]
合口	u[舊]	ua[華]				uai[怪]		ui[肥]

二、鼻聲韻

	m 韻尾		n 韻尾		ng 韻尾		
開口	am[堪]	om[蓼]	an[班]		ang[紅]	ong[王]	
齊齒	iam[險]	im[金]	ian[電]	in[真]	iang[雙]	iong[良]	ing[永]
合口			uan[灣]	un[溫]			

三、普通入聲韻

	p 韻尾		t 韻尾		k 韻尾		
開口	ap[合]		at[賊]		ak[六]	ok[國]	
齊齒	iap[業]	ip[入]	iat[列]	it[七]	iak[約]	iok[菊]	ik[竹]
合口			uat[雪]	ut[術]			

四、音節性輔音

m[毋]	ng[光]

五、喉塞入聲韻

開口	ah[鴨]	eh[厄]	oh[學]	aih	auh

齊齒	ih[舌]	iah[役]		ioh[藥]	
合口	uh[突]	uah[活]	ueh[劃]		

與國語不同的是台語基本音節(base-syllable)為877個，較國語多出許多；另外，台語的聲調分為陰平、陰上、陰去、陰入、陽平、陽上、陽去、陽入等八個聲調，但其中二(陰上)、六聲(陽上)已合併，故實際上只剩下七種聲調，也較國語的聲調多，各聲調之特徵及例字如表2-4所示，其典型基頻軌跡(pitch contour)如圖2.1所示。

表 2-4 台語八聲例表

聲調	台文字	羅馬拼音
一聲(陰平)	衫	saN
二聲(陰上)	短	te2
三聲(陰去)	褲	khou3
四聲(陰入)	闊	Khoah4
五聲(陽平)	人	lang5
六聲(陽上)	矮	e2
七聲(陽去)	鼻	phiN7
八聲(陽入)	直	tit8

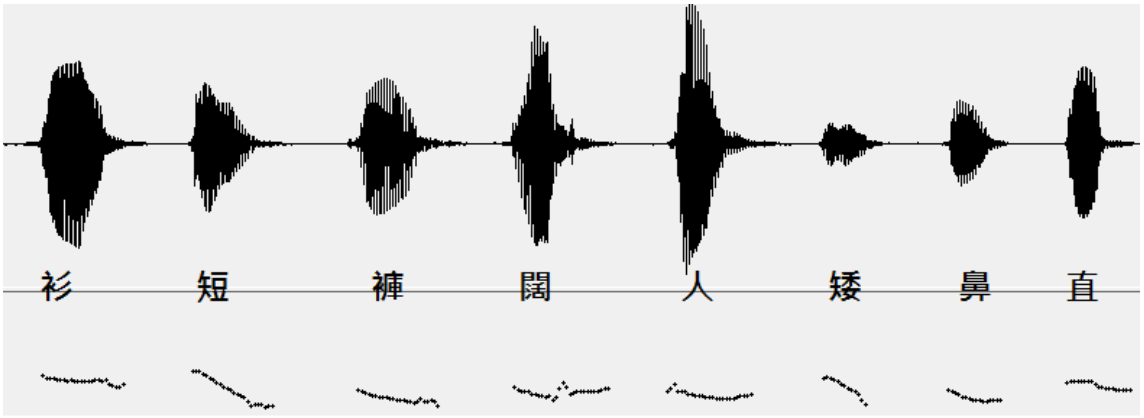


圖 2.1 台語八聲調之波形及基頻軌跡

2.2 台語變調規則

要合成出自然的台語語音，其中台語聲調的變調是一個重要因素，聲調是台語一項很重要的特色，很多情形下不同的聲調，會造成語詞的意義或是詞性的不同，例如:擔 taN-1 聲<挑>為動詞、taN-3 聲<擔子>變成名詞;分 hun-1 聲<分開>為動詞 hun-7 聲<份數>變成量詞。

在詞層次，大多數情形是最後一音節讀本調，其餘讀變調。然而在句子的層次，大部分情形是在詞組或是標點符號的分界處最後一音節讀本調，其餘讀變調(包含詞的最後一音節也讀變調)。

變調的部分，除了規則變調(圖 2.2)外，變調又分為以下幾種情形:

1. 隨前變調：一般為代名詞或人名的後綴，前面一音節讀本調，此音節的聲調視前面聲調而定，為 1 或 3 或 7 聲，例如:A-eng-a [阿瑛啊](7-1-1)(第二個"a"是後綴，所以隨前變為 1 聲)。
2. 輕聲:輕聲前讀本調，輕聲的部分讀 3 聲或 4 聲(入聲)，例如:chau-chhut-lai [跑出來](2-4-3)(「出」念回本調 4 聲，「來」唸成 3 聲輕聲，原本「出來」一詞聲調為 8-5)。
3. 再變調：多半出現在喉塞音(-h)4 聲，規則變調兩次(4→2→1)，例如:beh-thak-chu [要讀書](1-4-1)(beh4 聲應變 2 聲，實際變 1 聲)。
4. a[仔]前變調:a 前的音節，只有 1、2 聲同規則變調，其餘不同，例如:sun-a[孫仔](7-2)(「孫」本調為 1 聲)。
5. 三連音變調:三連音疊詞的第 1 音節，2、3、4 聲同規則變調，其餘不同，例如:chheng- chheng- chheng[清清清](5-7-1)(第一個「清」因為三連音的緣故由 1 聲變 5 聲，第二個「清」則是照基本規則變調由 1 聲變 7 聲)。
6. 升調:通常發生在日語借詞，詞的第一個聲調變調為 5 聲，例如:han-to-lu[方向盤](5-1-3)。

由於考慮到平常對話的台語口語中，並非完全依照上述規則來進行變調，以及目前偵測台語變調位置的困難，因此在本篇論文中，所使用的變調規則為詞層次的規則變調。

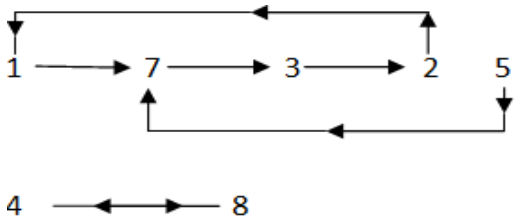


圖 2.2 規則變調示意圖(圖中數字代表台語七種聲調)

2.3 語料庫簡介

在進行語音合成之前，首先必須收集足夠的語料來進行合成模型的訓練，且這些語料都必須要有其相對應的台語標音及切割位置，但並不是人人都能標準的說台語或是標示台語的拼音。以下將介紹本論文所使用的語料庫以及在訓練合成模型前對語料庫做了那些前處理。

本論文所使用的音檔由一位專業男性台語錄音員所錄製，其文章內容為阿瑛的故事，文章內容以漢羅拼音來標記，音檔的錄製方式是將文章分成許多段落錄製而成，全部音檔時間共109分鐘，總音節數23631個，最長段落的音檔字數約為282個字，音檔均為20kHz的取樣頻率及16-bit之PCM格式。

此語料庫經由本實驗室將文本內容作斷詞及詞性標記，斷詞之後的文本，最大詞單元字數為 7 字詞，各字數統計資料如下表:

表 2-5 文本斷詞長度數目統計

斷詞單元字數統計	
1 字詞	5300
2 字詞	4683
3 字詞	2221
4 字詞	1215
5 字詞	162
6 字詞	22
7 字詞	1

接著此語料庫藉由潘荷仙老師實驗室所做的聲調標記及音節時間切割資訊，使得此語料庫所含資訊更加完整豐富。聲調標記的方法直接使用人工聽音檔，藉由聽到的語者聲調，來標記文本中相對音節的聲調，因此在此所標記的聲調為已經經過語者變調後之聲調。

由於在語音合成中，聲調是一個很重要的影響因素，因此統計了此文本中的聲調分布狀況，由表 2-6 可看出 1、2、3、7 聲出現數量較多，不過也可發現到其他三個聲調的數量並不會有過少到造成訓練量不夠的問題。

表 2-6 聲調分佈統計

聲調	數量
第 1 聲	4062
第 2 聲	3432
第 3 聲	3782
第 4 聲	1775

第 5 聲	1282
第 7 聲	6434
第 8 聲	2018

同時利用已經切割好的音節位置(切割方法使用人工切割),找出音節相對應的 initial final 分為 3:7 等分，接著再利用人工手動調整到較適當的 initial final boundary，圖 2.3 為音檔的切割狀況。

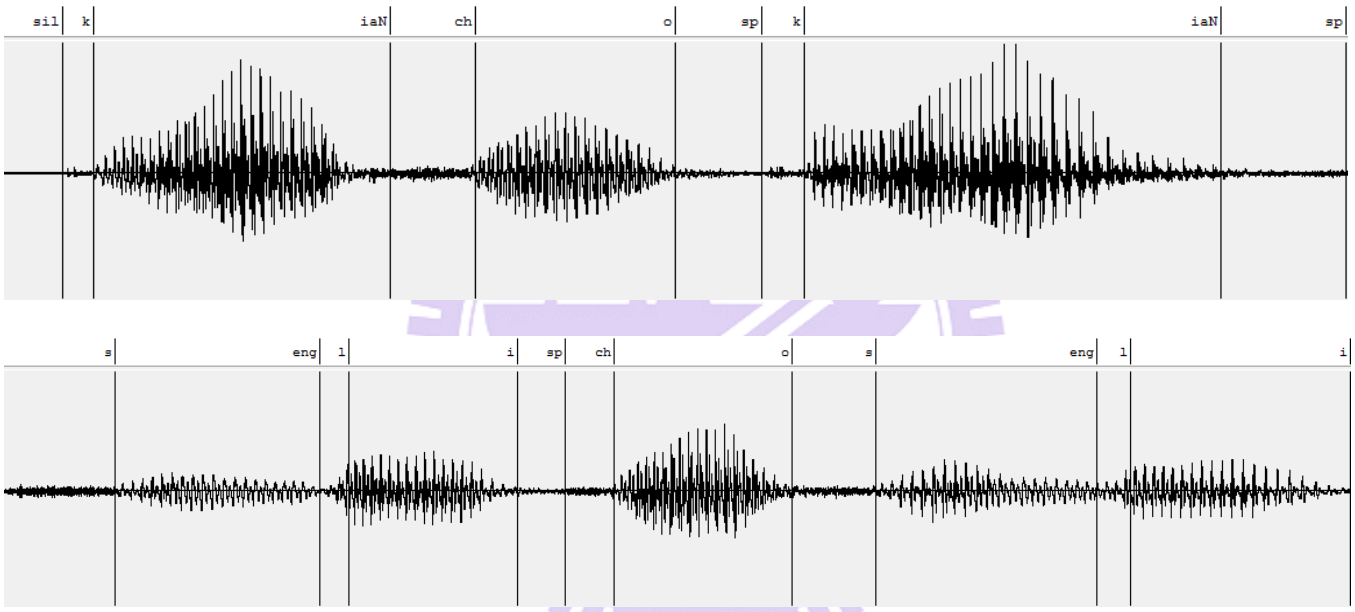


圖 2.3 音檔切割狀況

2.4 國台語字典的合併

為了建構國語轉台語語音合成系統，就必須具備國語字詞轉成台語字詞的翻譯字典，因此我們將利用實驗室已有的國語轉台語字典及國語字典做合併的動作，首先把國語字典中缺少台語詞翻譯(在此所稱的翻譯是指將國語詞對應到台語拼音)的國語詞補上台語翻譯，並收錄一些新詞條擴充到此字典。

目前實驗室所擁有的字典資料如下: (1)國台翻譯字典共 126831 筆詞條; (2)國語字典

121624 筆詞條。這兩個字典所使用的拼音系統皆使用教會羅馬拼音系統，國台翻譯字典的各字詞資料統計如下表 2-7:

表 2-7 國台翻譯字典各字詞數量統計

一字詞	13038
二字詞	75914
三字詞	26678
四字詞	9208
五字詞	1577
六字詞	416
總和	126831

國語字典各字詞資料統計如下表 2-8:

表 2- 8 國語字典各字詞數量統計

一字詞	13459
二字詞	64827
三字詞	26044
四字詞	16067
五字詞	999
六字詞	155
七字詞	65
八字詞	8
總和	121624

2.4.1 國語轉台語一詞多音現象

由於大部分人在學台語時，都是聽長輩或者其他人的發音學來的，因此雖然是同個字，但卻可能因為發音方法的不同，造成同個字不同發音，例如:「日」就有 Chit、git、jit、lit 四種發音，韻母都相同，只差在聲母的發音方法不同，再加上有些詞的字不管是文讀或白讀大家都聽得懂，因此大家也就不在意哪種讀音比較正確了，所以在國台翻譯字典中國語字詞翻譯到台語字詞時可能會發生同一個國語詞對應到很多不同的台語發音，例如:日光中的「日」就有 Chit、git、jit、lit 四種念法再加上「光」也有 Kng、kong 兩種念法,所以組合起來就有高達8種的念法,因此我們針對此現象對國語轉台語字典，做了詞重複出現次數的統計如下:

表 2-9 詞重複出現次數統計

重複出現次數 (次)	數量
2	7448
3	2216
4	1677
5	66
6	44
7	32
8	7
11	1
13	1
總數	11492

以下再針對各字詞分別做重複出現次數統計，統計資料如下:

表 2-10 各字詞重複出現次數統計

(1) 一字詞

重複出現次數(次)	2 次	3 次	4 次	5 次	6 次
數量	376	65	56	2	1

(2) 二字詞

重複出現次數(次)	2 次	3 次	4 次	5 次	6 次	7 次	8 次	11 次	13 次
數量	6115	1052	464	53	29	29	7	1	1

(3) 三字詞

重複出現次數(次)	2 次	3 次	4 次	5 次	6 次	7 次
數量	935	1099	1157	11	14	3

(4) 四字詞

重複出現次數(次)	2 次
數量	14

(5) 五字詞

重複出現次數(次)	2 次
數量	8

針對一字多音的部分，我們利用計算 entropy，來評估一字多音的情況如何， entropy 使用公式為:

$$Entropy = \sum_{i} P * \ln(P) \qquad (2\text{-}1)$$

i:當前一字詞的總發音數量

P:固定一種發音情形下，現有國台語字典中出現的機率值

Entropy 為一大於等於零的數值，數值越小代表越容易傾向於某一種讀音，當數值等於零時，代表此一字詞只會有一種發音情況。

字典中所有的一字詞共有 13038 個，在不考慮 tone 的相異性(同一讀音，但在不同字詞中有不同聲調先不考慮，在此都算同一種讀音)，只考慮讀音相同與否，共計有 3520 個字 entropy 大於零，大約佔一字詞的 27%，而其他 73%的一字詞不管在哪個詞彙中都只會有一個讀音。

表 2-11 未考慮 tone 的情況下，一字詞 entropy 的分布情形

Entropy 值	數量
>2	2
2>Entropy>1	482
1>Entropy>0	3036
Entropy=0	9518

在加入了 tone 的考慮後，entropy 大於零的一字詞共 3957 個，較未考慮 tone 的影響下多了 437 個字，但普遍來說 entropy 都不算太高。

表 2-12 考慮 tone 的情況下，一字詞的 entropy 分布情形

Entropy 值	數量
>2	7
2>Entropy>1	958
1>Entropy>0	2992
Entropy=0	9081

entropy 較高的一字詞，原因大多為這個字在不同詞彙下的讀音都不太相同，且各讀音中會有幾個讀音的出現機率很平均，造成 entropy 被拉大

例如:「券」: kng3-票「券」 機率:0.506

koan3-債「券」機率: 0.291

khng3-寶「券」機率:0.012

khoan3-契「券」機率:0.189

因此如果只有一字詞的時候，很難只給定一個發音，但一字多音這個問題，在考慮到多字詞，經由詞彙的意義去做判斷後，都可以很容易給定一個恰當的讀音。

另外由表 2-9 可看出其中二字詞及三字詞所佔的重複次數比例較其他字詞為高，二字詞及三字詞重複的這些詞其原因大概可分為以下情形:

1. 不同地區會有不同的讀音，例如:海口腔、南、北部腔。
2. 不同情景造成不同讀音，例如:「丟掉」根據場景可能是:a.我把某樣東西「丟掉」了，此時的情況為丟掉某樣東西了，讀音為 tan3-tiau7 或者是 b.我好像有某樣東西「丟掉」了，此時的情況又變成是我的某樣東西不見了，讀音卻變了成 pang3-kiN3。
3. 文讀白讀兩者皆可，例如:「台東」有人念白讀 tai5-tang1，也有人念文讀 tai5-tong1，這兩種讀音都有人使用，且不會造成聽者的誤解，因此兩種讀音都是正確的讀音。

在一詞多音方面，我們的做法是把一個詞的多種發音，選出較常用且出現機率較高的讀音，且考慮不會造成意義上的不同為原則來選，不過這種做法，並無法有效解決第二點的情形，因此之後希望能夠以較大單元做斷詞，利用前後文的情境加入考慮，希望能大幅改善這種情形。

2.4.2 台語的文、白異讀現象

漢語有著文白異讀的現象，台灣的閩南語也是如此，在所有方言之中屬閩南語的文讀與白讀之間差異最為顯著，由表面觀察可以大致分別為因使用場合的不同而有不同的讀音，然而，深入探討其中的原因，則牽涉到語言的歷史。

「文讀音」又叫做「讀書音」、「讀音」、「孔子白」，而「白讀音」又叫做「口語音」、「語音」。文讀音應屬閩南人仿官話讀書之音，再經過語音的自然演變而逐漸形成。口語音則為各個不同時期古音殘留的演變結果。讀書比較講究規範，因此讀書音的音韻系統較為單純而穩定，並且比較接近國語體系。相對地，口語音就比較複雜，有些口語音甚至保留了上古漢語音韻特徵，成為漢語方言之中碩果僅存的活化石。

這些不同讀音的字詞在約定俗成下，什麼情況下哪些字詞該文該白，閩南人基本上已有相當共識，任意變換，雖然未必會造成誤解，卻很難有一套系統可以完全定義出文讀白讀的使用規則。

台語的「文讀」通常是使用在配合文字而讀的場合; 「白讀」則是在日常口語會話中的場合使用，例如: 「知」:知人知面不知心-「知」讀 ti(文讀)，你知我知-「知」讀 chai(白讀); 「眉」:眉開眼笑-「眉」讀 bi5(文讀)，目眉-「眉」讀 bai5(白讀)，上面兩個例子都是一個字有兩種讀音，分別適用於兩種語用場合這就是「文白異讀」。
但是很多時候「文白異讀」的分別，並不是那麼容易就可以區分出來的，如下表所示:

表 2-13 常用字的文白異讀

	文讀音	白讀音
大	大丈夫-tai7	大小-toa7
學	學校-hak8	學開車-oh8
娘	娘子-niu5	阿娘-nia5
香	香港-hiang	香味-phang
山	孫中山-san	山頂-soaN

樂	音樂-gak8	快樂-lok8
家	家庭-ka	人家-ke
雪	雪白-soat4	白雪-seh4
光	天光-kng	光景-kong
人	人生-jin5	人-lang5
不	不由己-put4	不捨-m7
仔	水筆仔-chu2	歌仔戲-a2
林	林口-na5	林桑-lim5
水	水啦-sui2	放水流-chui2
高	高雄-ko	高低-koan5
較	比較-kau3	較好-kah8
果	果然-ko2	果菜-ke2
書	讀書-chheh	讀書-chu

以上這些例子是文讀與白讀複雜的混在日常生活中一起用，且很難區分其中的規則。在表 2-9 更可以看出有些字有五、六種以上的讀音，如:

「香」:hiang-(香油、沈香、香芹菜)

phang-(香水、香粉、香味)

hiuN-(香火、香客、香爐)等。

有時候一個字不同的文白發音也會造成意義上的不同，如下表:

表 2-14 文白發音造成意義不同

詞	文讀	白讀
加工	ka kang「加工」	ke kang「多此一舉」
中央	tiong iong「中央」	tiong ng「中間」

起初文讀與白讀分別應用在不同的應用場合，二者可以共存，隨後文讀用法漸為普及，進入了日常生活的口語世界，發生文白混雜現象，相互拉鋸下的結果，或文讀取代白話，或文讀消失，也或者二者融合產生新的形式，文白夾雜著使用，原來文白區分的體系逐漸消失(即因場合不同而使用不同語音)，文白不再只是讀書與口語形式的不同，而是可能代表著不同字詞意義或有不同的語法功能。

台語的文白異讀現象相當繁複，目前還無法依據其用法，建立一個完整的系統，不過現在我們是要用到台語的語音合成系統上，因此不管是文讀或白讀，目的是要能夠讓使用者明白要表達的意思，所以我們在選擇一個字詞應該使用哪一種讀音時，使用的方法是:

1. 成語及專有名詞幾乎都使用文讀，例如: 三緘其口、大放厥詞、薩克森安哈爾特州等，除非是其它有習慣用法則會另外使用白讀。
2. 人的名多以文讀，而姓多用白讀，但主要還是以習慣或好說好聽為準則，例如: 馬英九 Ma ing-ku 姓和名都為文讀，陳水扁 Tan Tsui-pinn 則都為白讀。
3. 扣除 1、2 兩種詞後，剩下的詞彙則依據日常生活中大家習慣的念法來選擇恰當的讀音，若是遇到國語字典中出現，但在台語中幾乎不常用到的詞彙則直接選擇以文讀拼音，例如: 拮据、春霖，其它日常生活中常出現的詞彙，則選擇以大家都較為熟知的讀音來選擇讀音。

2.4.3 建立國語字典詞彙的台語相對應拚音

為了使國語詞轉台語發音系統更強健，因此我們將合併數個詞典，使得中文翻譯台語文的詞條數更為豐富，在這邊利用:(1)實驗室建立的中文對台語拼音自典，此字典共包含 126831 筆詞條，內容包含國字及相對台語拼音，以及(2)國語字典，此字典包含 121624 個詞條，包含內容為國字與國語拼音，因為我們要做的是將國語文字轉成台語向對應拼音，因此國語字典中的詞條都應該有其相對應的台語拼音，因此我們先統計國語字典中的詞條與現有的中文對台語拼音字典對應情況，統計如下表:

表 2-15 國語字典對應到台語拼音統計表

	國語字典各字詞數量	台語詞典有對應的數量	缺少對應的詞條數量(有對應的詞條所占比例)
一字詞	13459	13038	421(96.87%)
二字詞	64827	33705	31122(51.99%)
三字詞	26044	4627	21417(17.77%)
四字詞	16067	2177	13890(13.55%)
五字詞	998	49	949(4.91%)
六字詞	155	9	146(5.81%)
七字詞	65	0	65(0%)
八字詞	8	0	8(0%)
總和	121624	53605	68019(44.07%)

由上表可看出現有國語字典的對應情形，在一字詞部分可以看到對應比例已經高達97%，至於未對應到的一字詞除了少數未收錄到的之外，其他則是日常生活中已經很少使用甚至根本完全沒在使用的古字。其他多字詞沒對應到的詞條有些是屬於專有名詞或人名，例如:摩納哥、鳶尾科、劉永福、劉銘傳，在四字詞的部分則是因為有成語的部分，在台語沒有直接對應到的讀音，例如:曇花一現、積弱不振、簞食壺漿，五字詞以上為國語俚語或公司組織比較難以有台語對應，例如:船到橋頭自然直、賠了夫人又折兵、醉翁之意不在酒、聯合國教科文組織。

不過為了使國語文字翻台語文字系統更為通用，我們還是想辦法去補齊這些缺少的詞條。首先判斷字典中缺少的詞條是否有需要補到字典中，若需要補齊，考慮查詢實驗室樟樹出版社出版的新編華台語對照典，若字典中沒收錄的考慮是否直接用文讀或白讀來翻譯，或是由一字詞中逐字找出拼音填上。

以下針對各字詞缺少對應詞條的對應方法做介紹:

- **一字詞**

利用新編華台語對照典，找出相對應一字詞拼音，其餘國語字典中找不到對應的一字詞，考慮到日常生活中已很少使用，所以這邊暫時不考慮加入對應的台語拼音。

- **二字詞**

將二字詞拆解成兩個單獨的一字詞，再利用現有台語字典，搜尋所有詞條，找出這兩個一字詞出現過的所有拼音組合，並藉由人工挑選出適當的拼音組合，例如:「一陣」的「一」有(1)it、(2)chit 兩種讀音，「陣」有(1)chun、(2)tin 兩種讀音，因此我們可以選出讀音組合為 chit8-chun7。

- **三字詞**

將三字詞拆解成:1.二字詞+一字詞(考慮到若是後詞綴則直接補上，例如:觀音山的「山」、普洱茶的「茶」) 2.一字詞+二字詞(考慮到若為前詞綴則直接補上，例如:下個月的「下」、大主顧的「大」) 3.三個一字詞，依序利用上面三種拆解規則，找出在現有台語詞典中有收錄的詞條，是否在三字詞拆解成小單元詞後有所對應，拆解的順序利用上面所列的次序為先後，找出適當的台語對應拼音組合，再利用人工挑選出適當的拼音組合，例如:研究生、研究所、研究院、研究費，上述詞條的「研究」在第一拆解方法中即可拆解成「研究」+後綴，「研究」一詞在台語詞典中有收錄台語讀音，因此經過人工確認，直接沿用收錄在台語詞典中的讀音為適當之後，即可把「研究」的發音 gian2-kiu3，直接補到此三字詞中的「研究」兩個字當中，接著再利用人工挑選出適當的第三個字讀音。

- 四字詞及五字詞

四字詞及五字詞類似三字詞作法，將四字詞依序拆解成:

1. 三字詞+一字詞(考慮到若是後詞綴則直接補上)
2. 三字詞+一字詞(考慮到若為前詞綴則直接補上)
3. 二字詞+二字詞
4. 二字詞+一字詞+一字詞
5. 一字詞+一字詞+二字詞
6. 四個一字詞

五字詞則拆解成:

1. 一字詞+四字詞
2. 四字詞+一字詞
3. 二字詞+三字詞
4. 三字詞+二字詞
5. 一字詞+二字詞+二字詞
6. 二字詞+二字詞+一字詞
7. 一字詞+一字詞+三字詞
8. 三字詞+一字詞+一字詞
9. 一字詞+一字詞+一字詞+二字詞
10. 二字詞+一字詞+一字詞+一字詞
11. 五個一字詞

四字詞及五字詞依序利用上面幾種拆解規則，找出在現有台語詞典中有收錄的詞條，是否在多字詞拆解成相對小單元後有所對應，拆解的順序利用上面所列的次序為先後，找出每種組合相對應的台語發音對應的拼音組合後，再利用人工挑選出適當的拼音組合。

- 六、七、八字詞

考慮到六字詞以上缺少的詞條只有 219 個，若依照上面的方法做拆解，會拆解得太凌亂且過於沒有效率，因此六字詞以上的詞，除了可以看出由那些小單元的詞構成的以外，只拆解成一字詞的組合，再利用人工挑選出適當的拼音組合，例如:杜思妥也夫斯基、布宜諾斯艾利斯、凱薩琳丹妮芙。

另外可以分辨出由較小單元詞條所構成的詞如: 海軍 - 軍官 - 學校、工業 - 技術 - 研究院、第一次 - 世界大戰，都是很容易可以利用比較小單元的詞條找出相對應的台語讀音。

利用以上方法把國語字典中的所有詞條都有了相對應的台語發音，再加上原本的國對台字典收錄的詞做統計之後，合併後新的字典統計結果如下:

表 2-16 統整後字典統計表

	國語字典詞條數(已做完國台對應)	國台翻譯字典未對應到國語字典的詞條數	合併後的字典總詞條數
二字詞	64827	42209	107036
三字詞	26044	22051	48095
四字詞	16067	7031	23098
五字詞	999	1528	2527
六字詞	155	407	562
七字詞	65	0	65
八字詞	8	0	8
總和	108165	73226	181391

再加上一字詞部分利用新編華台語對照典找出原本未對應到的一字詞共 399 個:

	國語詞典中包含數量	國台翻譯字典中與國語詞典有對應到詞條數量	統整及找字典後有對應的總詞條數
一字詞	13459	13038	13437

如此一來便有了一個具有國語字、國語音節碼及台語音節碼的字典，為了使此字典不只是國語字翻台語拼音，也希望能夠在此字典中加入台語文字表示。

因此接著利用鄭良偉教授所提供的詞典共 76977 筆詞條，此詞典資訊包含台語漢羅文字、台語拼音、相對華文，由於鄭良偉老師所提供的字典收錄詞條皆已收錄在實驗室已有的台語詞典中，因此在此只利用鄭良偉老師所提供的對應台語文字，方法為利用已

建立好的國語字對台語拼音字典中的台語拼音，與鄭良偉教授所提供詞典的台語拼音做比對，找出相同台語拼音的台語文字。

利用拼音去找出相對應的台文，會碰到一個問題，就是在相同拼音下的台文選擇會很多，例如:「大去」對應到的詞有-1.大去 2.大氣 3.大器，上面三個選項都是同一發音tai7-khi3，但是卻都代表不同意思，因此必須選擇第一個選項「大去」，才是正確的文字對應。因此若單靠拼音去做選擇，是無法達到很好的對應，但若經由華文意義來挑選適當對應台文，就會容易許多，因此我們利用此一方法，進一步建立具有國語字、國語音節碼、台語字及台語音節碼，達到更完整的國台對應。

表 2-17 有對應到台語文字的詞條數統計

字數	國對台拼音總詞條數目	有對應到台語文字的詞條數目
二字詞	107036	40901
三字詞	48095	6392
四字詞	23098	1723
五字詞	2527	34
六字詞	562	3
七字詞	65	0
八字詞	8	0
總和	181391	49053(27%)

第三章　台語語音合成系統實作

3.1 系統環境、及工具簡介

本論文的台語語音合成系統建構在 Linux 作業系統之下，使用的核心版本為 2.6.32.26-175.fc12，Linux 作業系統具有高穩定性，多使用者、多工等特點，且所要求的硬體門檻較低，其中的軟體大部分原始碼公開(open source)，並且允許使用者依個人要求而修改。

此語音合成系統的主要部分，是使用日本名古屋大學資工研究所(Nagoya Institute of Technology Department of Computer Science)開發的 HTS 2.1 [1](HMM-based speech synthesis system,version 2.1)。HTS 2.1 本身並不是一個獨立運作的系統，此系統是基於 HTK 3.4 (Hidden Markov model toolkit, version 3.4)[2]的修改版本。HTK 是由英國劍橋大學電機系所開發出來的隱藏式馬可夫模型開發工具，提供了相當豐富的指令，方便使用者實作隱藏式馬可夫模型之建構，以及使用隱藏式馬可夫模型作為語音辨認之用;HTS 保留了大部分 HTK 的指令，只針對語音合成上的一些需要做更動。

3.2HTS (HMM-based speech synthesis system toolkit)

HTS[3,4]的開發已有十幾年的歷史，主要是應用 HTK 建構隱藏式馬可夫模型的方便性，修改為適合語音合成的版本。

在 HTS 的訓練部分可以單純視為 HTK 訓練工具的修改版本，主要的修改部分如下:

(一) 增加串流相關文本分類法(stream-dependent context clustering)。

(二) 針對基頻參數的模型建立，以多空間機率分布(MSD,Multi-space probability distribution) [5]作為狀態輸出的機率密度函數。

(三) 狀態持續時間的模型建立和分類(State duration modeling and clustering)[6]。

3.3HTS 系統流程

HTS 的系統架構大致可分為訓練與合成兩大部分，詳細系統架構為圖 3.1 所示，在 HTS 訓練部分我們將台語聲母、韻母、長靜音(SIL)以及短靜音(SP)模擬成五個狀態的 HMM 模型，也就是將他們模擬成最小的 HMM 訓練單元，再對於每個最小單元給予文本標記紀錄其文脈相關資訊，及利用語料求取好的語音聲學參數和文本標示，訓練出文脈相關的頻譜及音高 CDHMM(context dependent HMM)及 state duration 模型，以下各小節將針對各部分做解釋。

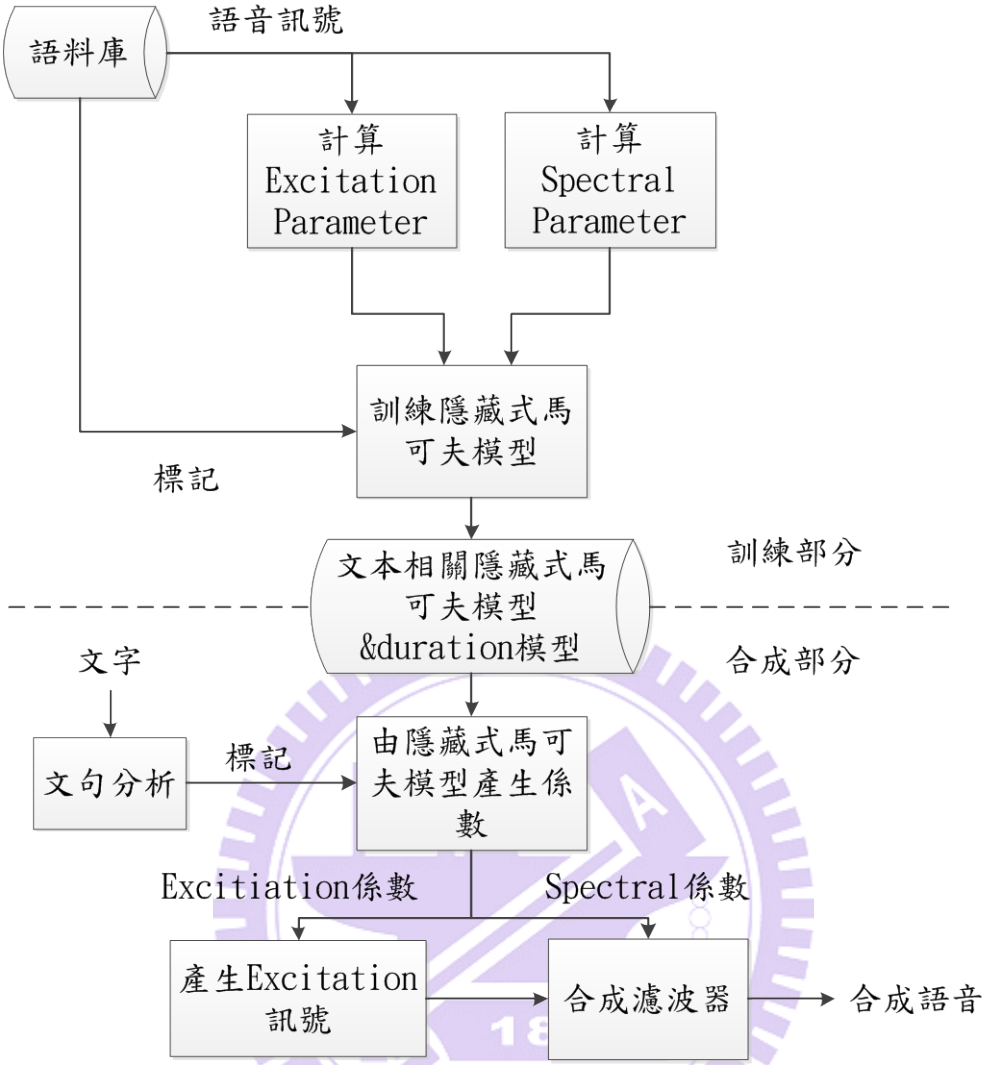


圖 3.1 HTS 系統架構

3.3.1 文本標記(label)

文本標記提供訓練 CDHMM 以及 state duration 所需的文脈相關語言參數，以及在合成時挑選適當的 CDHMM 及 state duration 模型時也會需要用到文脈相關語言參數。訓練 CDHMM 時依照文本標示提供的文脈相關資訊對聲學參數做訓練，文本標示的文脈相關參數會影響 HMM 單元本身的頻譜及韻律變化，也會影響 HMM 單元之間連接的狀況，如連音現象、詞首詞尾和句首句尾明顯的音高差異及音節伸長縮短。本系統所使用的文脈相關資訊如表 3-18:

表 3-18 文脈相關語言參數

P_{n-1}，P_n，P_{n+1}	Previous(PRE)/Current(CUR)/Following(FOL) Initial/Final/SP
ST_{n-1}，ST_n，ST_{n+1}	Lexical tones of PRE/CUR/FOL syllable
PW_1/PW_2	Syllable position in a lexical word(LW)(forward/backward)
PS_1/PS_2	Syllable position in a sentence(forward/backward)
PM	Punctuation mark after the current syllable
WL_{n-2}，WL_{n-1}，WL_n，WL_{n+1}，WL_{n+2}	Lengths of PRE-PRE/PRE/CUR/FOL/FOL-FOL LWs in syllable
WP_{n-2}，WP_{n-1}，WP_n，WP_{n+1}，WP_{n+2}	POSs of PRE-PRE/PRE/CUR/FOL/FOL-FOL LWs
SL_{n-1}，SL_n，SL_{n+1}	Lengths of PRE/CUR/FOL sentences in syllable

由於我們也將長靜音(SIL)以及短靜音(SP)視為 HMM 的訓練單元，長靜音就是在音檔開始和結束的靜音部分，而短靜音則定義為語句中音節間大於 25ms 的靜音停頓，所以在文本標示中，對於短靜音也給予文脈相關資訊，在訓練時也會學習到不同文脈相關資訊下的停頓長度。

3.3.2 聲學參數(Spectral and excitation parameter extraction)

將語料庫經過 3.3.1 的文本標記後，即完成語料庫文本標記部分，接著再從語料庫中的語音訊號抽取聲學參數。本研究中 CDHMM 模擬的聲學參數為廣義梅爾倒頻譜係數(Mel-generalized cepstrum,MGC)(Tokuda et al.,1994) [7]及基頻(F0)。廣義梅爾倒頻譜係數可藉由調整其 γ 參數，將語音信號頻譜以 all pole(γ=-1)、Cepstrum(γ=0)或是以廣義的 pole 和 zeros 一起表示(γ≠-1,0)，亦可調整 α 參數以代表不同的 frequency wrapping，以方便考量人耳的聽覺效應。在本研究中，我們使用 SPTK(SPTK Working Group,2009) [8]工具抽取 24 階廣義梅爾倒頻譜係數，設定 γ=0 以及 α=0.5，音檔取樣頻率為 20kHz，所使用的分析音框大小為 25ms(500 個資料點)的漢明窗(Hamming window)，音框位移為 5ms(100 個資料點)。另外，抽取基頻參數則使用 Wavesurfer 工具中的 ESPS 方法求取

(Sjlander and Beskow,2000)，分析音框大小(window size)為 7.5ms，而音框位移(window size)為 5ms。

3.3.3 隱藏式馬可夫模型之訓練

經過文本標記及聲學參數的抽取後，接著就可以進行隱藏式馬可夫模型的訓練階段，由於文本標示的文脈相關資訊組合相當多，每一種組合都是個別的 CDHMM，在訓練語料不夠充足的情況下，多數的 CDHMM 訓練資料量會有過少的情形，使得訓練出來的模型會不夠準確造成過度訓練(overfitting)，因此本研究使用標準的 Tree-based CDHMM 訓練方法(Zen et al.,2007;Yoshimura,2002)，以決策樹搭配適當的問題集來做分群訓練，以語言學的知識為基礎設計出合理的問題集，問題集架構為表 3-19 所示，對於某些資料量較少的模型可以合併在一起訓練，以增加訓練時的資料量，如此可訓練出較強健的模型。

表 3-19 問題集架構

level	ID Description
Sil,Sp	Previous sp/sil Current sp/sil Following sp/sil
Syllable level	Previous initial/final Current initial/final Following initial/final Syllable position in Subword Syllable position in Word Syllable position in Sentance
Initial/Final level	According to the pronunciation characteristics Category
Tone	Previous Tone Current Tone Following Tone
POS	LLL & LL & L POS Current POS RRR & RR & R POS
Word length	LLL & LL & L Word/SubWord length

	Current Word/SubWord length RRR & RR & R Word/SubWord length
PM	Pre_PM& Fol_PM
Current Sentence Length In Current Syllable Follow Sentence Length In Current Syllable	

問題集大致分為幾個層次:

- 判斷前一個、現在、後一個所接的語音單元是否為長靜音(SIL)或短靜音(SP)。
- 考慮前一個、現在、後一個所接的聲母及韻母，並依據聲母或韻母的發音方法、發音位置、送氣或不送氣以及清音濁音來設定問題集。
- 依據現在音節聲調以及前一個、後一個所接的音節聲調做分類。
- 考慮前後及現在詞的詞性，將中研院 46 類詞類依實詞虛詞、八大詞類及其他特殊詞類集合合併，產生問題集。
- 考慮現在音節所在的詞長和詞的位置，如現在音節是否在詞首或詞尾。
- 考慮當前音節所在的句子前後所接標點符號，設定問題集。
- 考慮現在音節所在的句長和句子的位置，如現在音節是否在句首或句尾。

由上列問題集概述的考量，本研究所設定的問題集共約 2200 個左右。

3.3.4 隱藏式馬可夫模型之語音合成

利用前一節所訓練出來的文脈相關頻譜、音高 CDHMM 及 state duration 模型，來進行語音的合成。合成時，先輸入一段文字經由文字分析後產生文本標記，接著利用此文本標記，再依據決策樹上每個節點的問題，找出適當的 CDHMM 做串接，進而產生 Excitiation(Pitch)以及 Spectral(MGC)參數。

接著將上一步得到的每個音框之 logF0(Pitch)和 MGC 頻譜參數輸入至 MSLA filter (Mel-Log Spectrum Approximation filter)(Imai,1983)產生出合成語音。

3.4 中文文字轉台語語音合成系統

此研究所做的中文轉台語語音合成系統，是利用 3.3 的 HTS 方法，來訓練台語語料庫的文脈相關頻譜、音高 CDHMM 及 state duration 模型，接著在合成部分的文本標記之前，做了一些前處理，使得合成時所用到的文本標記資訊為台語的資訊，詳細的系統架構如圖 3.2 所示:

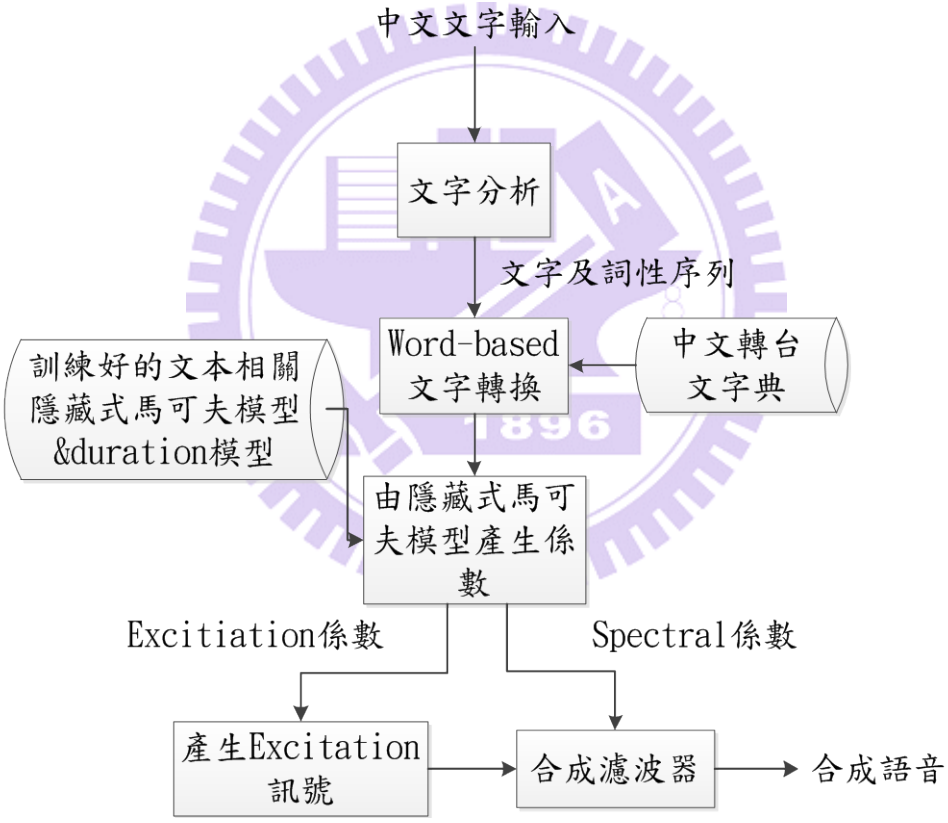


圖 3.2 中文文字轉台語語音合成系統架構圖

以下各小節將針對各方塊逐一說明。

3.4.1 文字分析(Text Analysis)

文字分析(Text analysis)是文字轉語音合成系統的第一級，由於在此所輸入的文字還是國語文字，因此在這邊是利用國語的斷詞器做文字分析，傳統國語斷詞器使用的是長詞優先斷詞規則，最著名的是中央研究院的中文斷詞系統。但自 2000 年起，由於 CRF (Conditional random field)方法[9]被提出，並有效的使用在自然語言處理中的各個問題，都被證實較傳統規則法或其他統計方式為佳。

因此，本系統的文字分析的斷詞及詞類標記部分，便是採用 CRF 的方法做為核心。

3.4.2 Word-based 中文文字轉台語文字轉換

這一部分是將國語文字經由文字分析後的斷詞結果，所產生的文字序列，利用第二章所建立好的國台語對應字典，依據斷詞的結果找出字典中相對應的國語字詞條，接著依照字典中的對應把國語文字及發音轉換成台語的文字及發音，做 Word-based 的中文轉台語文字的轉換。

3.4.3 台語語音合成

中文文字在經過文字分析以及 Word-based 的文字轉換後，產生一個新的台語文本，再利用此轉換過後的文本產生新的文本標示，接著由已經訓練好的 CDHMM 模型，產生 logF0 以及 MGC 頻譜參數，輸入至 MSLA filter 合成語音。

3.5 利用語音合成系統檢查國台字典的正確性

由於 2.4 節所利用的字典合併方法，並不能完全保證每個詞條都能正確對應到台語的聲調及音節，因此我們依照本章節所實做出來的台語語音合成系統，實作出一套只念單詞的單詞系統，並利用此系統做人工的檢查，來確認現有國台轉換字典的正確性，此合併後的國台轉換字典共 194828 個詞，扣除掉原本已收錄的國台對照字典共 126831 個詞後，剩下 67997 個詞，為我們所新增的對應，因此檢查的重點為這 67997 個詞，以下列出檢查這些詞所採取的方法:

1. 每個詞的音節碼是否正確。
2. 單詞中的聲調是否變調正確。
3. 若詞條為台語中不會出現或使用的詞，則利用台語一字詞直接翻譯的方法來檢查其音節碼，聲調部分則依所對應一字詞的聲調取代。

在利用此系統檢查字典的過程中可以發現，由 2.4 節所利用的長詞優先拆詞對應法則，所做出來的詞由於其台語拼音皆是經由人工所挑選，因此在對應的台語拼音上沒有什麼太大的問題，但在聲調的部分，則是利用每個小單元詞的聲調，做為大單元詞聲調的對應，因此聲調在這邊是一個檢查的重點。

我們在此定義一個詞為一個變調群組，依據台語變調規則，群組中除了最後一個字不變調外，其餘前面字都做變調的動作。我們按照此定義來對每個詞確認是否每個詞的聲調，都是以詞為變調邊界做變調之後的變調結果，因此就可以確認字典中的詞拼音及聲調對應的正確性。

第四章 利用階層式韻律模型分析台語語料庫

由於考慮到台語與中文在語法結構上的相似性，因此在本章節中利用江振宇博士所提出之中文韻律模型為基礎[10]，訓練一階層式韻律模型，並透過此模型訓練結果，來對台語語料庫做分析。

4.1 漢語語音階層式韻律架構

依據語言學家的研究[11]，語音的韻律結構呈現階層式架構。[12]提出了韻律標記的概念並定義了階層式多短語韻律句群(Hierarchical Prosodic Phrase Grouping,HPG)架構，其架構如圖 4.1 所示，最底層為音節層次(SYLlable layer,SYL)，其中聲調為最強烈的影響因素，聲調不單只影響音節基頻軌跡之走向，也影響了音節長度及能量位階;往上發展依序為韻律詞層次(Prosodic Word layer,PW)，由雙音節或多音節所構成的詞組，通常在句法和語意上關係緊密;韻律短語層次(Prosodic Phrase layer,PPh)，由一或多個韻律詞組成，結尾常會帶有不明顯但可察覺之停頓;呼吸組層次(Breath Group,BG)，由單一或數個韻律短語所組成的句子，其結尾通常帶有明顯停頓;最上層為韻律組句(Prosodic phrase Group,PG)，由一個或數個呼吸組構成。

停頓標記是用來區分韻律組成份子的邊界，B0 和 B1 區分了 SYL 的邊界，其中 B0 表示 reduced syllabic boundary，而 B1 表示 normal syllabic boundary，這兩種停頓類別通常不具明顯停頓;B2 和 B3 分別是韻律詞和韻律短語的邊界;B4 則代表了呼吸組的邊界，和 B2、B3 比較起來會有較明顯的停頓;至於 B5 定義了韻律句組邊界，代表一個完整的段落結束，通常句尾會有音節長度拉長(final lengthening)及能量減弱等現象。

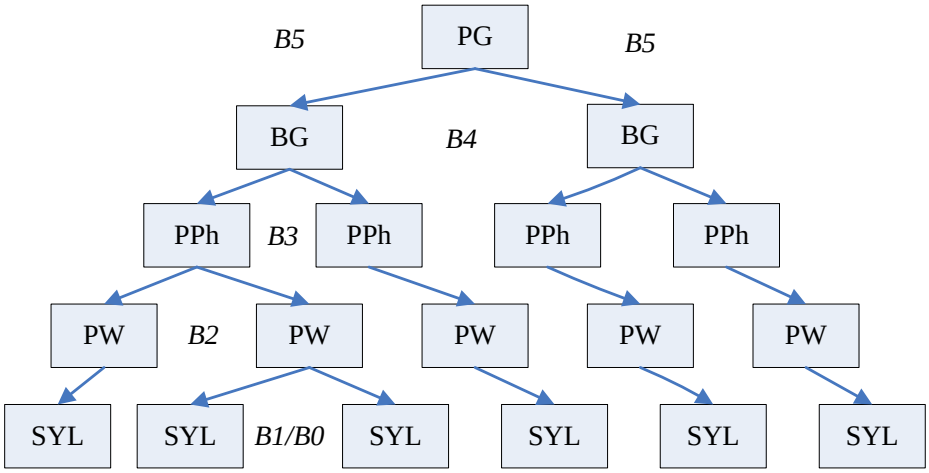


圖 4.1:中文語音韻律階層式架構概念[12]

本研究使用之語料庫為大段落的語音，因此就以 HPG 架構為基礎，經過進一步修改後，利用修改後的韻律階層架構作為本研究所使用的韻律階層式架構。首先將 B2 再細分為 B2-1、B2-2、B2-3，分別代表明顯音高重置(pitch reset)、短停頓(short pause)及含有音節拉長效應(duration lengthening)之韻律詞邊界等不同現象。接著將 BG 和 PG 合併為同一層，因為這兩層所描述的韻律特性相近，B4 則和 B5 合成為 B4。整個架構從 5 層變為 4 層，如圖 4.2 所示。最後採用的 7 種韻律邊界停頓(break type)為 **B**={B0,B1,B2-1,B2-2,B2-3,B3,B4}，以此來標記四種韻律單元:音節(SYL)、韻律詞(PW)、韻律短語(PPh)、呼吸組/韻律句組(BG/PG)。

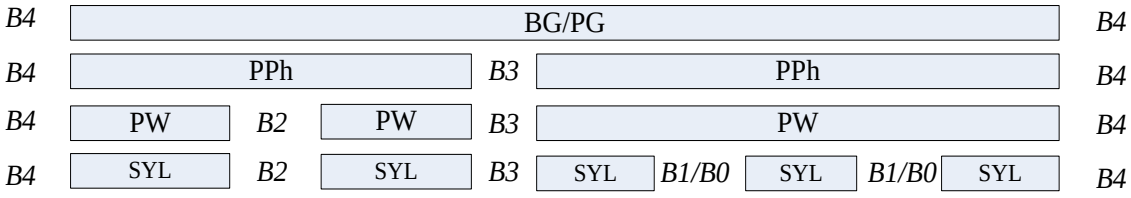


圖 4.2:本研究所採用之階層式韻律架構[10]

4.2 韻律模型設計

本研究所採用的韻律模型為江振宇博士所提出之中文韻律模型，其依據圖 4.2 表示之中文階層式韻律架構,可針對一未經人工事先標記的語料,利用語言參數和聲學參數，自動標記出停頓標記及韻律狀態。此演算法具備兩大優點:

1. 可自動標記，解決了傳統上韻律標記多為人工標記，既耗時耗力，又容易有不一致的問題。
2. 透過此模型可清楚分析韻律詞層次以上的韻律變化趨勢。

韻律標記問題可視為，在給定語料庫之語音聲學參數集合 A，和對應的語言參數集合 L 下，求取最佳韻律標記集合 T 的過程，即

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

韻律標記集合 T={B,PS}包括兩類語音韻律資訊，第一類為音節邊界停頓標記(Break Type)，在本論文所使用的音節邊界停頓標記集合為 B={B0,B1,B2-1,B2-2,B2-3,B3,B4}; 另一類的韻律標記為音節的韻律狀態，音節韻律狀態有 3 種 PS={p,q,r}，各別代表的意義分別為經過量化和正規化音節基頻韻律狀態 p、音節長度韻律狀態 q 和音節能量韻律狀態 r。

將正規化後的基頻韻律狀態扣掉音節層次對基頻的貢獻，即扣除聲調和連音的影響因素，此時音節基頻的韻律狀態代表的是韻律詞、韻律短語、呼吸組/韻律句組對基頻的貢獻;至於音長或能量強度則分別扣除語句、聲調、基本音節類型或韻母類型的影響因素。

聲學參數也可分為兩類，其中一類的聲學參數和韻律狀態的標記有很大相關性，而與音節邊界停頓標記的相關性則非常小，屬於這類的聲學參數有音節基頻軌跡、音長和音節能量;另一類的聲學參數則特性相反，和音節邊界停頓標記有很大的相關性，而與韻律狀態標記的相關性小，屬於這類的聲學參數有音節邊界的停頓時長(pause duration)、

音節邊界的 energy-dip level、正規化的能量差、正規化的基頻差(normalized pitch jump)以及正規化的音節長度拉長因子(normalized duration lengthening facor)等。

在此我們定義 A 包含音節基頻軌跡序列 sp、停頓時長序列 pd、音節能量低點(energy-dip level)序列 ed、音節長度序列 sd、音節能量序列 se、正規化的音節內基頻差序列 pj 及正規化的音節長度拉長因子序列 dl 和 df，其中 pj 定義為:

$$pj_n = (\mathbf{sp}_{n+1}(1) - \boldsymbol{\beta}_{t_{n-1}}(1)) - (\mathbf{sp}_n(1) - \boldsymbol{\beta}_{t_n}(1)) \quad (4\text{-}2)$$

4-2 式括號中的 1 代表參數的第一維，下標 n 表示此為第 n 個音節，β_{t_n} 為聲調影響因素 t_n 的 affecting patterns(APs)，而正規化的音節長度拉長因子序列 dl 和 df 定義為:

$$dl_n = (sd_n - \gamma_{t_n} - \gamma_{s_n}) - (sd_{n-1} - \gamma_{t_{n-1}} - \gamma_{s_{n-1}}) \quad (4\text{-}3)$$

和

$$df_n = (sd_n - \gamma_{t_n} - \gamma_{s_n}) - (sd_{n+1} - \gamma_{t_{n+1}} - \gamma_{s_{n+1}}) \quad (4\text{-}4)$$

其中 γ_t 和 γ_s 分別表示聲調與基本音節類型影響因素的 APs，因此聲學參數集合 A={sp,sd,se,pd,ed,pj,dl,df}。

為了更清楚的說明這些聲學參數，將 A 進一步細分為三個類別:音節韻律參數(Syllable Prosodic Feature) X={sp,sd,se}，音節內韻律參數(Inter-syllabic Prosodic Feature) Y={pd,ed}以及音節差韻律參數(Differential Prosodic Feature) Z={pj,dl,df}。

至於語言參數，則用 L 來表示所有的語言參數集合。其中特別將音節聲調、基本音節類型與韻母類型從 L 獨立出來，因為這三個參數分別對音節基頻軌跡、音長與音節能量有顯著的影響;此外考慮到不同語句時，說話速度的變動會造成音長的變化以及音量變動會造成能量的變化，因此再把語句層次的正規化因子獨立出來;扣除從 L 中獨立出來的語言參數後，剩餘的語言參數，則統一定義為 I(reduced linquistic feature set)。這些符號的定義整理在下表 4-1。

表 4-1:韻律標記、韻律參數和語言參數的表示法

T : prosodic tag	**B** : break type={*B0*, *B1* ,*B2-1*, *B2-2*, *B2-3*, *B3*, *B4*}	
	PS : prosodic state	**p** : pitch prosodic state
		q : duration prosodic state
		r : energy prosodic state
A : prosodic feature	**X** : syllable prosodic feature	**sp** : syllable pitch contour
		sd : syllable duration
		se : syllable energy level
	Y : inter-syllabic prosodic feature	**pd** : pause duration
		ed : energy-dip level
	Z : differential prosodic features	**pj** : normalized pitch jump
		dl : normalized duration lengthening factor 1
		df :normalized duration lengthening factor 2
L : linguistic feature	**l**: reduced linguistic feature set	
	t: syllable tone sequence	
	s: base-syllable type sequence	
	f: final type sequence	
	u:utterance sequence	

綜合上述之討論，可將 4-1 式改寫為

$$P(\mathbf{T},\mathbf{A}\,|\,\mathbf{L}) = P(\mathbf{A}\,|\,\mathbf{T},\mathbf{L})P(\mathbf{T}\,|\,\mathbf{L}) = P(\mathbf{X},\mathbf{Y},\mathbf{Z}\,|\,\mathbf{B},\mathbf{PS},\mathbf{L})P(\mathbf{B},\mathbf{PS}\,|\,\mathbf{L})$$
$$\approx P(\mathbf{X}\,|\,\mathbf{B},\mathbf{PS},\mathbf{L})P(\mathbf{Y},\mathbf{Z}\,|\,\mathbf{B},\mathbf{L})P(\mathbf{PS}\,|\,\mathbf{B})P(\mathbf{B}\,|\,\mathbf{L}) \qquad (4\text{-}5)$$

其中 $P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})$ 稱為音節韻律模型，用來敘述音節韻律參數受到停頓標記 **B**、韻律狀態 **PS** 和語言參數 **L** 之間的影響而產生的變化； $P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$ 稱為停頓聲學模型，用來描述在各個不同停頓標記 B 和語言參數 L 下，其韻律邊界的聲學特性; $P(\mathbf{PS}|\mathbf{B})$ 稱為韻律狀態模型，描述韻律狀態在不同停頓標記 B 下的轉移變化; $P(\mathbf{B}|\mathbf{L})$ 稱為停頓標記語言模型，描述在不同的語言參數 L 下，各種停頓標記出現的頻率。以下分成四小節針對這四種韻律模型做詳細探討。

4.2.1 音節韻律模型

音節韻律模型 $P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})$ 可進一步分解成三個子模型，分別模擬音節基頻軌跡序列 sp、音長序列 sd 和音節能量序列 se，並假設 sp、sd 和 se 的變化在此只受到以下幾個影響因素:音節聲調 t、基本音節 s、韻母類型 f、語句 u、韻律狀態 PS={p,q,r}和韻律邊界停頓 B，數學式表示如下:

$$\begin{aligned} &p(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L}) \approx p(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{t})\, p(\mathbf{sd}|\mathbf{q},\mathbf{t},\mathbf{s},\mathbf{u})\, p(\mathbf{se}|\mathbf{r},\mathbf{t},\mathbf{f},\mathbf{u}) \\ &\approx \prod_{n=1}^{N} p(\mathbf{sp}_n|B_{n-1}^{n}, p_n, t_{n-1}^{n+1}) \prod_{n=1}^{N} p(sd_n|q_n,t_n,s_n,u_n) \prod_{n=1}^{N} p(se_n|r_n,t_n,f_n,u_n) \end{aligned} \tag{4-6}$$

其中 $\prod_{n=1}^{N} p(\mathbf{sp}_n|B_{n-1}^{n}, p_n, t_{n-1}^{n+1})$ 在模擬音節基頻軌跡受的各種影響，在此假設第 n 個音節之基頻軌跡 $\mathrm{sp_n}$ 會受到的 AP 為基頻韻律狀態 $\mathrm{p_n}$、目前聲調 $\mathrm{t_n}$ 以及給定韻律邊界停頓 $\mathrm{B_{n-1}}$ 和 $\mathrm{B_n}$ 情況下，前後一個音節聲調 $\mathrm{t_{n-1}}$ 和 $\mathrm{t_{n+1}}$ 造成的連音影響，此處 $B_{n-1}^{n}=(B_{n-1},B_n)$，$t_{n-1}^{n+1}=(\mathrm{t}_{n-1},\mathrm{t}_n,\mathrm{t}_{n+1})$，而 $\mathrm{sp_n}$ 則代表第 n 個音節基頻軌跡進行正交展開，投影到四個 Legendre 多項式基底所得到的四維正交參數[13]，依以上描述可將 $\mathrm{sp_n}$ 表示成:

$$\mathbf{sp}_n = \mathbf{sp}_n^r + \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^{f}_{B_{n-1}t_{n-1}^{n}} + \boldsymbol{\beta}^{b}_{B_n t_n^{n+1}} + \mu \tag{4-7}$$

(4-7)式的每項 $\boldsymbol{\beta}_x$ 表示音節基頻軌跡影響因素為 x 時的 AP，$\boldsymbol{\beta}^{f}_{B_{n-1}t_{n-1}^{n}}$ 和 $\boldsymbol{\beta}^{b}_{B_n t_n^{n+1}}$ 分別是第 n-1 個和第 n+1 個音節所貢獻的前後音節影響效應的 APs，μ 則是總體平均值(global mean);另外為了限制韻律狀態只對目前音節的 LogF0 level 有影響，故我們將 $\boldsymbol{\beta}_n$ 定義為四維正交係數的第一維且為非零值;$\mathbf{sp}_n^r$ 是正規化後的 $\mathbf{sp}_n$，為 $\mathbf{sp}_n$ 扣除 $\boldsymbol{\beta}_{t_n}$、$\boldsymbol{\beta}_{p_n}$、$\boldsymbol{\beta}^{f}_{B_{n-1},t_{n-1}^{n}}$、$\boldsymbol{\beta}^{b}_{B_n,t_n^{n+1}}$ 和 μ 的殘餘值(residual)。圖 4.3 為 $\mathbf{sp}_n$ 與影響因子之間的關係圖，在此假設 $\mathbf{sp}_n^r$ 為一平均值為零的高斯分佈隨機變數，即 $N(\mathbf{sp}_n^r;\mathbf{0},\mathbf{R})$，因此得到

$$p(\mathbf{sp}_n \mid p_n, B_{n-1}^{n}, t_{t-1}^{t+1}) = N(\mathbf{sp}_n; \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^{f}_{B_{n-1},t_{n-1}^{n}} + \boldsymbol{\beta}^{b}_{B_n,t_n^{n+1}} + \boldsymbol{\mu}, \mathbf{R}) \tag{4-8}$$

其中 R 定義為 $\mathbf{sp}_n^r$ 的共變數矩陣(covariance matrix)。

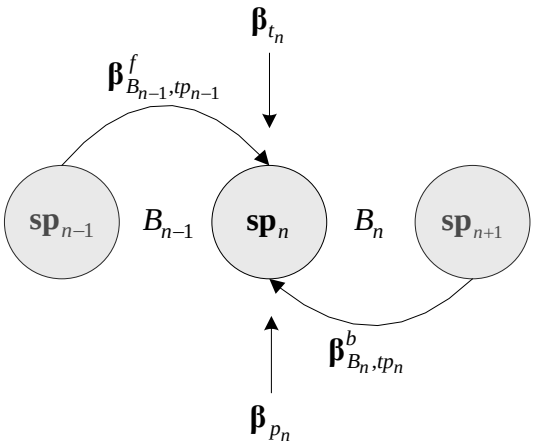


圖 4.3:音節基頻軌跡與其影響因素關係圖

依此類推，第二個模型和第三個模型可表示成:

$$p(sd_n \mid q_n, s_n, t_n) = N(sd_n; \gamma_{t_n} + \gamma_{s_n} + \gamma_{q_n} + \mu_d, R_d) \tag{4-9}$$

$$p(se_n \mid r_n, f_n, t_n) = N(se_n; \omega_{t_n} + \omega_{s_n} + \omega_{q_n} + \mu_e, R_e) \tag{4-10}$$

(4-9)式模擬了音節時長 sd_n，其中 γ_{t_n}、γ_{s_n} 和 γ_{q_n} 分別為聲調、基本音節類型和韻律狀態對 sd_n 的 APs，μ_d 和 R_d 分別為 sd_n 總體平均值及殘餘值之共變異數矩陣;(4.10)式模擬了音節能量位階 se_n，其中 ω_{t_n}、ω_{s_n} 和 ω_{q_n} 分別為聲調、聲母類型和韻律狀態對 se_n 的 APs，μ_{se} 和 R_{se} 則分別為 se_n 總體平均值及其殘餘值之變異數。

4.2.2 停頓聲學模型

將停頓聲學模型 $P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$ 做進一步分解

$$P(\mathbf{Y},\mathbf{Z}\,|\,\mathbf{B},\mathbf{L}) \approx P(\mathbf{Y},\mathbf{Z}\,|\,\mathbf{B},\mathbf{l}) \approx \prod_{n=1}^{N} p(pd_n, ed_n, pj_n, dl_n, df_n \,|\, B_n, \mathbf{l}_n) \tag{4-11}$$

在此用來描述韻律邊界的聲學特性參數為音節內及差分韻律參數{Y, Z}={**pd**, **ed**, **pj**, **dl**, **df**}，pd_n 為第 n 個音節所跟隨的接合點(juncture n，以第 n 個接合點表示)停頓長度；ed_n 為第 n 個接合點的能量下降程度；pj_n 為跨越第 n 個接合點的正規化基頻差，其定義如(4-2)、(4-3)、(4-4)所示。

由於對韻律停頓而言 I_n 的限制仍太大，故藉由分類樹與決策樹(Classification and Regression Tree,CART)演算法來估計 $p(pd_n, ed_n, pj_n, dl_n, df_n \,|\, B_n, \mathbf{l}_n)$，其節點分類標準依據最大概似函數增益(maximum likelihood gain)搭配一個事先設計好的問題集去實施 CART 演算法，依據不同的韻律邊界停頓將所有音節邊界的 pd_n、ed_n、pj_n、dl_n、df_n 做好分類，並於決策樹的每個終止節點(leaf node)統計參數分佈。在此我們將 pd_n 以伽瑪分佈(Gamma distribution)來模擬，而 ed_n、pj_n、dl_n、df_n 以高斯分佈模擬，假設五種聲學間彼此互相獨立，因此 $p(pd_n, ed_n, pj_n, dl_n, df_n \,|\, B_n, \mathbf{l}_n)$ 會是一個伽瑪分佈和四個高斯分佈的乘積，其數學式如下：

$$\begin{aligned}\prod_{n=1}^{N} p(pd_n, ed_n, pj_n, dl_n, df_n \,|\, B_n, \mathbf{l}_n) \approx \prod_{n=1}^{N} \Big\{ & g(pd_n; \alpha_{B_n,L_n}, \beta_{B_n,L_n}) N(ed_n; \mu_{ed,B_n,L_n}, \sigma^2_{ed,B_n,L_n}) \\ & N(pj_n; \mu_{pj,B_n,L_n}, \sigma^2_{pj,B_n,L_n}) N(dl_n; \mu_{dl,B_n,L_n}, \sigma^2_{dl,B_n,L_n}) \\ & N(df_n; \mu_{df,B_n,L_n}, \sigma^2_{df,B_n,L_n}) \Big\}\end{aligned} \tag{4-12}$$

4.2.3 韻律狀態模型

韻律狀態模型可依據三種韻律狀態拆解成三個子模型，分別用來模擬音節基頻、長度及能量三種韻律狀態，如 4-13 式所示:

$$P(\mathbf{PS}\,|\,\mathbf{B}) \approx P(\mathbf{p}\,|\,\mathbf{B})P(\mathbf{q}\,|\,\mathbf{B})P(\mathbf{r}\,|\,\mathbf{B}) \tag{4-13}$$

而 $P(\mathbf{p}\,|\,\mathbf{B})$、$P(\mathbf{q}\,|\,\mathbf{B})$ 和 $P(\mathbf{r}\,|\,\mathbf{B})$ 可以用雙連文模型(Bigram Models)分別表示為

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

$$P(\mathbf{q}\,|\,\mathbf{B}) \approx p(q_1)\left[\prod_{n=2}^{N} p(q_n \,|\, q_{n-1}, B_{n-1})\right] \tag{4-15}$$

和

$$P(\mathbf{r}\,|\,\mathbf{B}) \approx p(r_1)\left[\prod_{n=2}^{N} p(r_n \,|\, r_{n-1}, B_{n-1})\right] \tag{4-16}$$

其中 $p(p_1)$ 、 $p(\mathrm{q}_1)$ 和 $p(\mathrm{r}_1)$ 分別表示各個不同韻律狀態的初始機率， $p(p_n \,|\, p_{n-1}, B_{n-1})$ 、 $p(\mathrm{q}_n \,|\, \mathrm{q}_{n-1}, B_{n-1})$ 和 $p(\mathrm{r}_n \,|\, \mathrm{r}_{n-1}, B_{n-1})$ 則分別代表三種韻律狀態，給定 B_{n-1} 的情況下，從第 n-1 個音節的韻律狀態轉移到第 n 個音節韻律狀態的轉移機率。

4.2.4 停頓語法模型

最後停頓語法模型 break-syntax 模型 $P(\mathbf{B}|\mathbf{L})$ 可先簡化為 $P(\mathbf{B}|\mathbf{I})$ ，並假設每個音節邊界都可分開模擬，因此可表示成

$$P(\mathbf{B}\,|\,\mathbf{l}) = \prod_{n=1}^{N-1} p(B_n \,|\, \mathbf{l}_n) \tag{4-17}$$

其中 $p(B_n \,|\, \mathbf{l}_n)$ 使用 CART 演算法依據最大概似函數增益為分裂準則訓練得到。

4.3 韻律標記及模型訓練方法

A-PLM 法依據最大概似法則(maximum likelihood,ML)，同時預估 8 個韻律模型的參數並對所有語句做韻律標記，經一連串的最佳化程序直到收斂。整個演算過程可分為兩部分:初始化和疊代，初始化過程會對所有語句做初始的韻律標記，及預估前面所討論的 8 個子模型韻律參數的初始值;疊代的過程先對所有語句定義概似函數(Likelihood Function)

$$Q=\left(\prod_{n=1}^{N} p(\mathbf{sp}_n \mid p_n, B_{n-1}^{n}, t_{n-1}^{n+1})\, p(sd_n \mid q_n, t_n, s_n, \mathrm{u}_n)\, p(se_n \mid r_n, t_n, f_n, \mathrm{u}_n)\right)$$
$$\left(p(p_1)\, p(q_1)\, p(r_1) \prod_{n=2}^{N} p(p_n \mid p_{n-1}, B_{n-1})\, p(q_n \mid q_{n-1}, B_{n-1})\, p(r_n \mid r_{n-1}, B_{n-1})\right) \qquad (4\text{-}18)$$
$$\left(\prod_{n=1}^{N-1} \left(p(pd_n, ed_n, pj_n, dl_n, df_n \mid B_n, \mathbf{l}_n)\, p(B_n \mid \mathbf{l}_n)\right)\right)$$

接著利用一個多重步驟的疊代程序，反覆更新所有韻律標記和 8 個韻律子模型的參數，詳細說明可參考[1]。

4.4 韻律模型訓練結果與分析

訓練韻律模型所使用的訓練語料與訓練 HTS 所用語料相同皆為阿瑛的故事，共 255 句，總音節數共 23631 個。接著採取修正型 PLM 演算法疊代訓練至第 79 次達到收斂，其對應的目標總概似度(total likelihood of objective function)如圖 4.3 所示。接下來將針對模型訓練結果進行台語語料的分析。

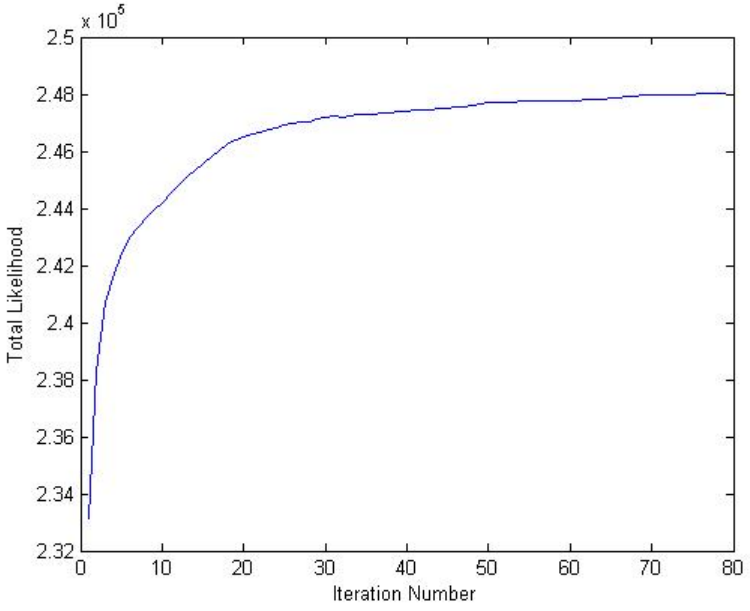


圖 4.4:疊代次數與目標總概似度

圖 4.5 顯示基頻軌跡的聲調 APs，由圖 4.5 可看出與圖 2.1 台語聲調基頻軌跡圖相符合，由此可確認此語料庫語者的聲調分佈相當明顯。

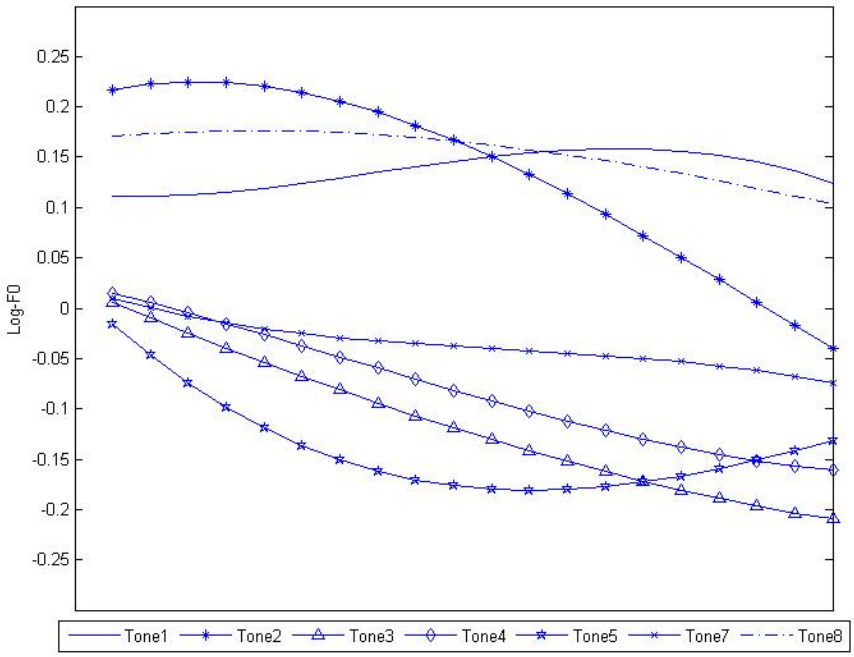


圖 4.5:基頻軌跡聲調 APs

接下來分析音節長度韻律模型，影響因子包含聲調、基本音節類型和韻律狀態。圖 4.6 顯示音節長度的聲調 APs，發現到其中台語五聲的音節長度明顯拉長許多，圖 4.7 顯示音節長度的基本音節類型 APs，此基本音節類型是把台語 877 音節類型依發音特性分成 80 類，由圖 4.7 可看出其中第 80 類的音節發音最長，此類對應到的 877 音節類型包括” iun”、”jun”、”chhun”、”sun”。

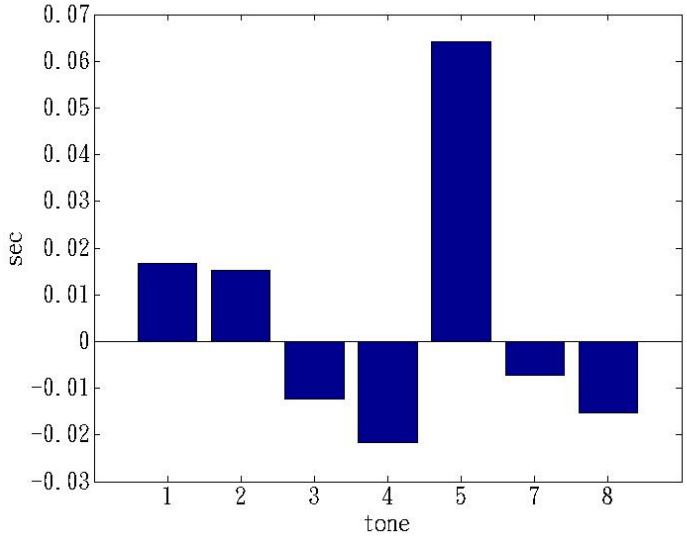


圖 4.6:音節長度之聲調 APs

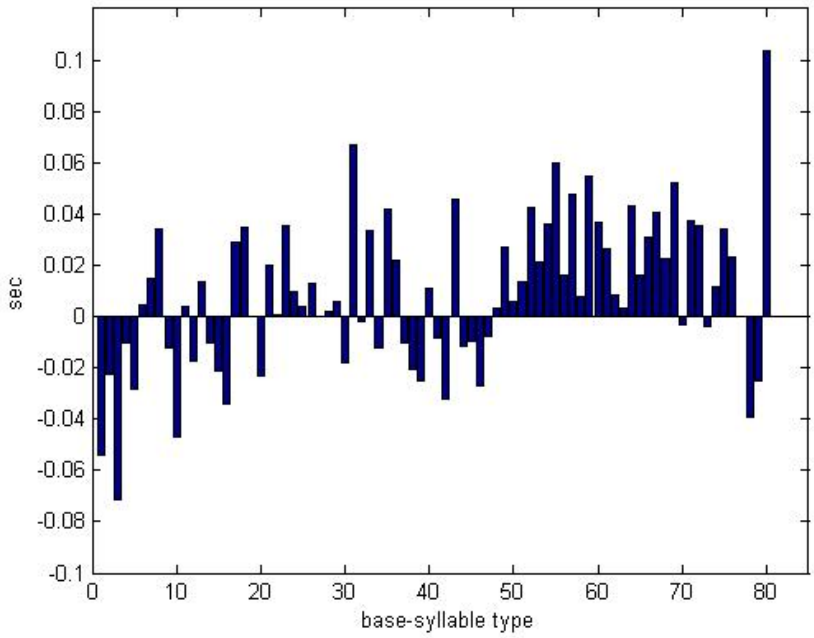


圖 4.7:音節長度之基本音節類型 APs

音節能量位階韻律模型，影響因子包含聲調、韻母類型及韻律狀態。圖 4.7 顯示音節能量位階的聲調 APs，其中台語以一、二、八聲的音節能量位階較大，其他聲調則較小，圖 4.8 顯示音節能量位階的韻母類型 APs，在此韻母類型有 63 類，其中以第 33 類的”iong”音節能量位階最小，此類韻母類型對應到的 877 音節類型如”chhiong”、”hiong”、”siong”等;第 10 類的”at”音節能量位階最大，此韻母類型對應到

877 音節類型如”tat”、”bat”等。

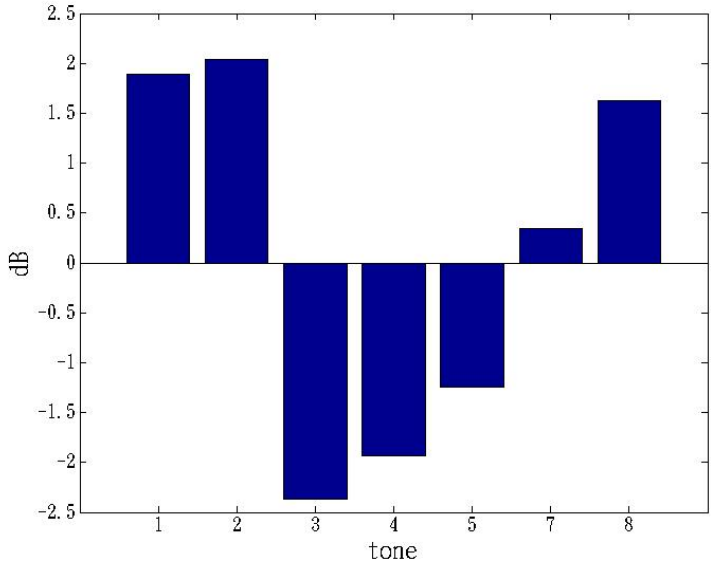


圖 4.8:音節能量位階之聲調 APs

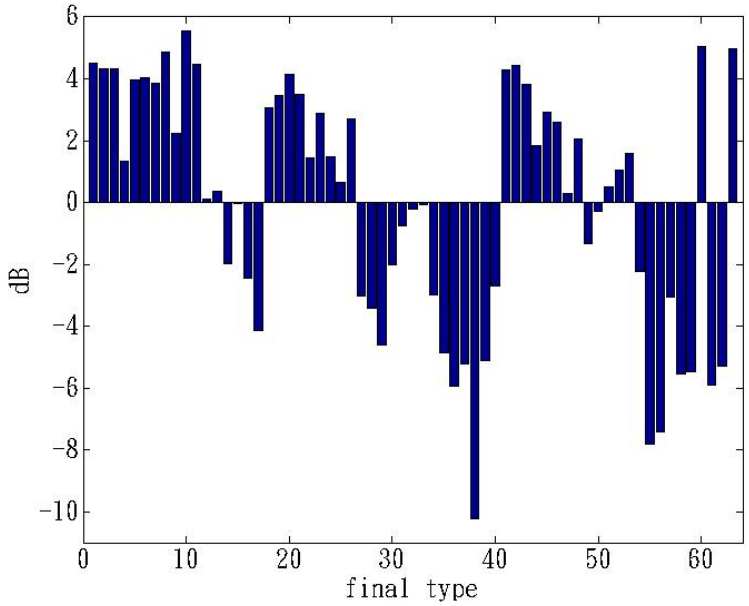
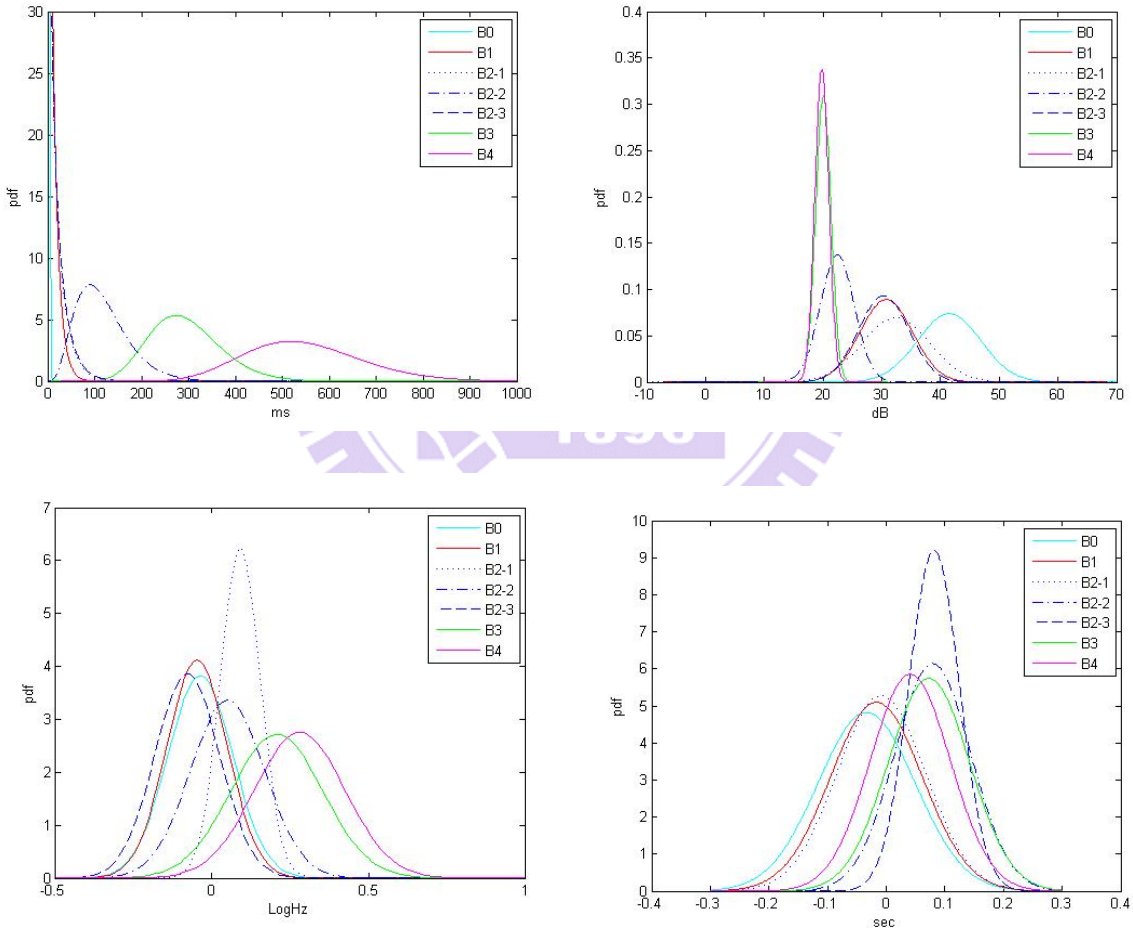


圖 4.9:音節能量位階之韻母類型 APs

停頓聲學模型由 CART 演算法建構而成，用以描述七種停頓標記 B、語言參數 I 以及音節間韻律參數{Y}={pd,ed}和音節差韻律參數{Z}={pj,dl,df}之間的關係。

圖 4.10 顯示在不同停頓標記下，決策樹根節點(root node)五種韻律參數的機率分佈。由圖可發現越上層韻律架構的停頓標記如 B3、B4，擁有較長的停頓時長、較低的能量低點、較明顯的基頻跳躍及音節拉長因子;而 B0、B1 的停頓時長都非常的短，但 B0 的能量低點較大，表示 B0 為兩音節緊密連接的邊界;B2-2 則有中等的停頓時長;B2-1 和 B2-3 的能量低點與停頓時長分佈與 B1 相似，但 B2-1 擁有較明顯的基頻跳躍，B2-3 則是音節拉長因子較為明顯。



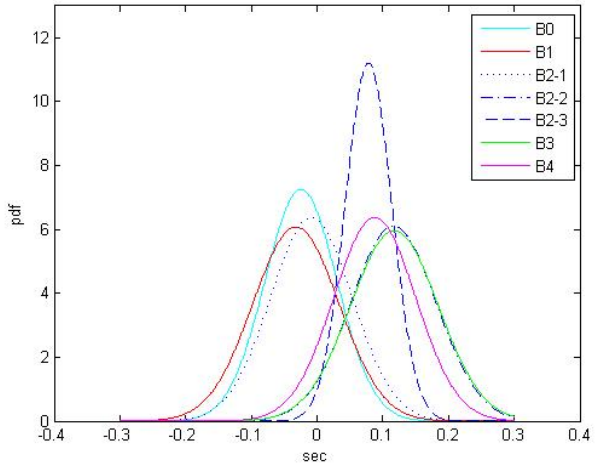


圖 4.10: (a)停頓音節長度，(b)音節能量低點，(c)正規化基頻跳躍值，(d)正規化音節拉長因子 l，(e)正規化

修正型停頓語法模型之建構是以 CART 演算法訓練一顆決策樹，根據語言參數對不同類型的停頓標記做分類，再對所有終止節點估計 $P(\mathbf{B}|\mathbf{I})$。

本論文在 CART 演算法的兩個分裂停止條件如下:

1. 決策樹分裂出之子節點，其最小樣本數必須大於 100。
2. 決策樹訓練過程中，其相對相似度增益(relative likelihood gain)必須大於 0.01。

決策樹訓練出的結果如圖 4.11 所示，每個節點都有其對應編號及問題，編號 1 為根節點，節點中的直方圖為該節點各停頓標記之機率分佈，由左至右分別為 B0、B1、B2-1、B2-2、B2-3、B3、B4，節點中的數值為該節點的總樣本數，另外，實線表示父節點的問題為”是”，虛線則為”否”。

在根節點第一個問題為 PM，由節點 2 之直方圖可知大部分樣本數都集中於 B3 和 B4，顯示大部分的 B3 和 B4 都產生於標點符號處;由節點 4 直方圖分佈可確信在標點符號為逗號時，其對應的停頓標記幾乎皆為 B3 或 B4，代表此語者經常以逗號來當作一個韻律短語(PPh)和呼吸組/韻律句組(BG/PG)結尾，圖中節點 5 的 B4 機率分佈也很高，推測其原因，除了逗號之外，很有可能是句號;從節點 2 往下長，問題集偏向句子層次的語言參數居多，例如 LPS>=6(前一個句子的長度是否大於等於 6)，表示這邊的韻律組成分子邊界大多屬於 major break。接下來分析韻律邊界為 non-PM 的部分，節點 3 以是否為

inter-word 邊界來分裂節點，若為”否”，即 intra-word 邊界，如節點 7 所示，幾乎都屬於 B0、B1 等 non-break 類別;若為 inter-word 邊界，則如節點 6 所示，此類之停頓時長 B1 到 B2-3 都有可能出現，從節點 6 開始往下長，發現 break 分佈範圍很廣，因此容易造成此類型邊界不容易得到一致性的標記結果。

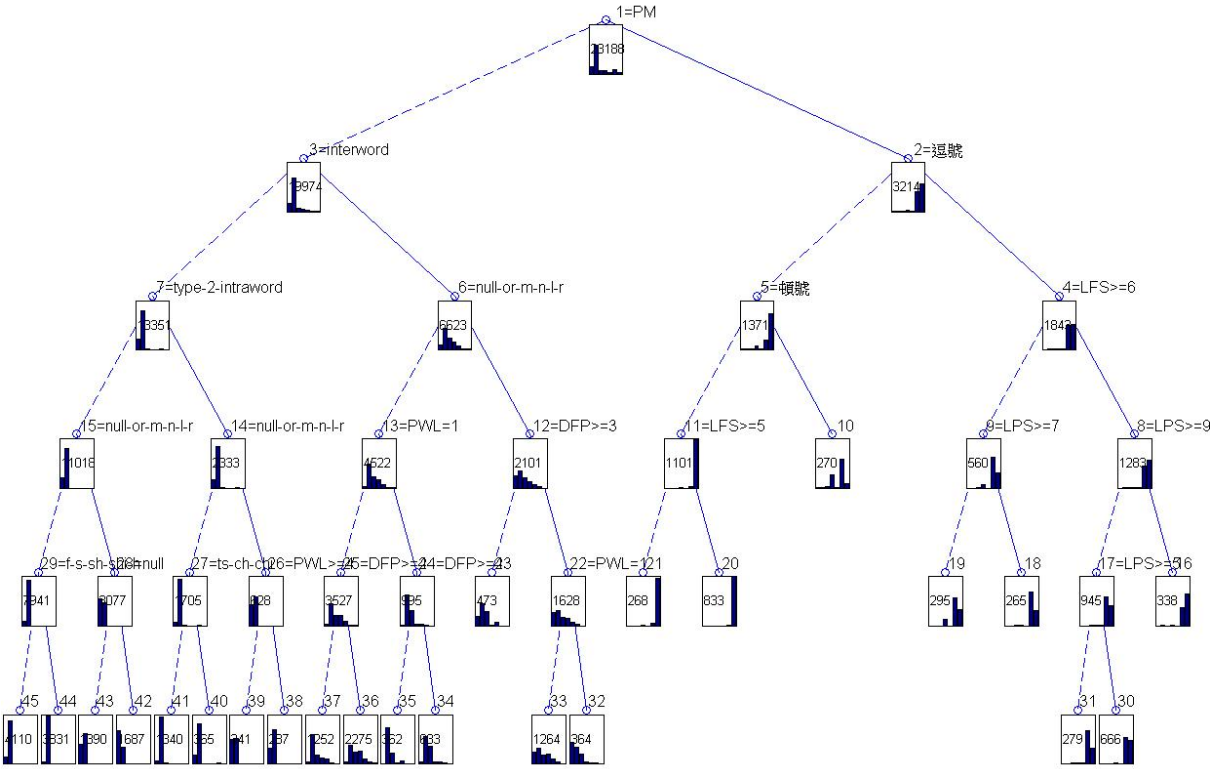


圖 4.11:停頓語法模型決策樹，節點中直方圖為各停頓標記的發生機率，由左至右分別是 B0,B1,B2-1,B2-2,B2-3,B3,B4，數值為該節點總樣本數

第五章　結論與未來展望

5.1 結論

本論文中，我們根據中文文法與台語文法的相似性，先對中文文章做斷詞，並在不變動此斷詞結果情況下，依據本論文方法所做出的國台對照字典，將斷詞結果一一做對應，轉換成為台語拼音的文章，再將此文章輸入到 HMM-based 語音合成器中，實現中文文章轉台語語音合成系統，此系統包含斷詞與詞性標記單元、國台對照字典及基於 HMM 之語音合成器。

最後利用中文韻律模型為基礎，分析了台語語料庫的語音特性，得到了以下幾項結論:

1. 台語第五聲調，音節長度明顯比其他聲調拉長了許多。
2. 第一、二及八聲調，音節能量則較大。
3. 由停頓語法模型決策樹可發現到 intra-word 的停頓較小，inter-word 的停頓範圍則較廣從 B0 到 B2-3 皆有可能。

5.2 未來展望

針對本論文所使用的方法及技術，有以下幾點可提供未來繼續探討並改進:

1. 目前台語詞典的收錄雖然已有一定數量，但在一些特殊詞，例如:俚語或外來語，數量還是非常少，未來若能收錄更多詞條，將能使得此合成系統更加豐富。
2. 系統中的台語變調規目前僅利用到台語中的規則變調，並未加上其它特殊變調規則，若能再加上其它的變調規則，將可使此台語語音合成系統的聲調變化更加自然。

參考文獻

【1】 HMM-based Speech Synthesis System (HTS),http://hts.sp.nitech.ac.jp/

【2】 Hidden Markov Model Toolkit (HTK),http://htk.eng.cam.ac.uk/

【3】 K.Tokuda, T. Kobayashi and S. Imai, *Speech Parameter Generation from HMM using Dynamic Features*, Proc. ICASSP, 1995.

【4】 Tokuda, Yoshimura, Masuko, Kobayashi, Kitamura, *Speech Parameter Generation Algortihms for HMM-based Speech Synthesis, ICASSP, 2000.*

【5】 Tokuda, Masuko, Miyazaki, Kobayashi, *Multi-Space Probability Distribution HMM*, IEICE Trans. Inf. & Syst., 2000.

【6】 Yoshimura, Tokuda, Masuko, Kobayashi, Kitamura, *Simultaneous modeling of Spectrum, Pitch and Duration in HMM-based Speech Synthesis*, Proc EU-ROSPEECH, 1999.

【7】 Tokuda ,*Spectral Estimation of Speech by Mel-Generalized Cepstral Analysis*,IEEE,1994.

【8】 Speech Signal Processing Toolkit (SPTK),http://sp-tk.sourceforge.net

【9】 Lafferty .Conditional Random Field:Probabilistic Models for Segmenting and Labeling Sequence Data,2001.

【10】 C. Y. Chiang, "Unsupervised Joint Prosody Labeling and Modeling for Mandarin Speech," Department of Communication Engineering, NCTU, Dissertation for Doctor of Philosophy, March 2009.

【11】 Z. Sheng, J.-H. Tao, and D.-L. Jiang, "Chinese prosody phrasing with extended features," Proceedings of the IEEE ICASSP 2003, Vol. 1,pp.492-495.

【12】 C.-Y. Tseng, S.-H. Pin, Y.-L. Lee, H.-M. Wang, and Y.-C. Chen, "Fluent speech prosody:Framework and modeling," Speech Commun. Special issue on quantitative prosody modeling for natural speech description and generation,46,284-309(2005).

【13】 S.-H. Chen and Y.-R. Wang, "Vector quantization of pitch information in Mandarin speech,"I EEE Trans. Commun., vol.38, no.9,pp. 1317-1320, Sept. 1990.