

國立交通大學

電信工程研究所

碩士論文

母語與非母語英語之韻律模型建立

A prosody modeling approach to L1 and L2 English speech

研究生：陳韋帆

指導教授：陳信宏 博士

中華民國一百零一年七月

母語與非母語英語之韻律模型建立

A prosody modeling approach to L1 and L2 English speech

研究生：陳韋帆

Student: Wei-Fan Chen

指導教授：陳信宏 博士

Advisor: Dr. Sin-Horng Chen

國立交通大學

電信工程研究所

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 of Science

in Communication Engineering

July 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年七月

母語與非母語英語之韻律模型建立

研究生：陳韋帆

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

本論文提出一個以音節為單位的韻律模型，其中包含了音節時間模型以及音節基頻軌跡模型。本論文考慮了若干會對這兩項韻律參數影響的因子，其中包括了重音類型、詞長、鄰近的停頓類型、音節組成的因素及韻律狀態等等。本論文使用一個疊代式的訓練法來建立此模型，並使用包含了母語與非母語英文語料的 TWNAESOP 語料庫。實驗數據顯示此模型能夠有效的建立起英語的韻律模型，與許多前人的研究結果符合；並且透過探討母語與非母語語者參數間的差別，我們得以從數據上得知兩者間韻律上的差異。

A prosody modeling approach to L1 and L2 English speech

Student: Wei-Fan Chen

Advisor: Dr. Sin-Horng Chen

Institute of Communication Engineering
National Chiao Tung University

Abstract

In this thesis, a statistical prosody modeling approach for L1 and L2 English speeches is proposed. The study focuses on the modeling of two prosodic-acoustic features: syllable duration and log-pitch contour. Several major affecting factors (AFs) that influence the variations of these two features are considered. They include lexical stress, word length, nearby break type, phonemic constituent of syllable, and prosodic state. A sequential optimization procedure is adopted to automatically train the two models from the TWNAESOP corpus recorded in Taiwan. The corpus contained both L1 and L2 English speeches. Experimental results showed that most AFs estimated agreed well with our prior linguistic knowledge. The differences in the prosody of L1 and L2 speeches were explored.

致謝

本論文得以順利完成，有賴於諸位師長朋友的支持與鼓勵。首先感謝自專題生以來就在百忙之中指導我的陳信宏教授，提供了豐富的實驗室資源使得我們得以專注的從事工作；並感謝苦口婆心指導我的王逸如老師，使得我更加注重各項的細節；另外還有自大三以來就協助我的，現已高就臺北大學的江振宇學長，這數年來亦師亦友的在研究上、生活上都給我莫大的助益；同時也還感謝余秀敏老師在本研究上也提供許多英文上的專業意見。

進入這實驗室兩年的日子稍縱即逝，共同奮鬥的同窗好友亦給我諸多助力：包括有結識六年的小邱、給予我許多意見和協助的喬華、總是笑臉迎人的俊翰、埋頭工作的睿詮、理工科少見的女生雅婷、甫自他處轉來的昌佑、時常給我英文建議的靖觀以及諸多前後輩等。在諸位的支持以及笑容中，使得這兩年來的日子不至枯燥乏味。同時還感謝遠在他鄉的惠晴，是我心靈上的摯友。

本人不才，七百餘日的光陰，僅得此寥寥數筆的拙作。如有後人能從本文中得到些許啟發，即為對我最大的鼓舞。

目錄

中文摘要.....	I
Abstract.....	II
致謝.....	III
目錄.....	IV
表目錄.....	VI
圖目錄.....	VII
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 文獻回顧.....	2
1.3 研究方向.....	2
1.4 章節概要說明.....	3
第二章 資料庫介紹.....	4
2.1 語料庫簡介.....	4
2.1.1 文字資料.....	4
2.2 資料前處理.....	5
2.2.1 音節切割資訊修正.....	5
2.2.2 音節時間資訊正規化.....	8
2.2.3 基頻資訊修正.....	10
2.2.4 基頻資訊正規化.....	16
2.2.5 音節基頻軌跡參數抽取.....	17
2.2.6 停頓類型標記.....	18
2.3 語料庫統計資料.....	24
第三章 韻律模型.....	30

3.1	音節時間模型.....	30
3.1.1	影響因子.....	30
3.1.2	模型架構.....	30
3.1.3	訓練流程.....	31
3.2	音節基頻軌跡模型.....	32
3.2.1	影響因子.....	33
3.2.2	模型架構.....	33
3.2.3	訓練流程.....	34
第四章	實驗結果與分析.....	35
4.1	音節時間模型.....	35
4.2	音節基頻軌跡模型.....	40
4.3	韻律狀態結果分析.....	47
第五章	結論與未來展望.....	51
5.1	結論.....	51
5.3	未來展望.....	51
	參考文獻.....	52
	附錄 1：TWNAESOP 文本資料.....	55
	附錄 2：音節類型統計資料.....	68

表目錄

表 2.1 : wavesurfer 的基頻求取設定.....	10
表 2.2 : 刪除子音基頻前後，其投影到四維正交基底造成的誤差.....	18
表 2.3 : 起始停頓標記門檻值.....	19
表 2.4 : 決策樹所使用的問題集.....	21
表 2.5 : L1 phone 統計資料.....	25
表 2.6 : L2 phone 統計資料.....	27
表 4.1 : 語群平均值.....	35
表 4.2 : 詞長因子參數.....	36
表 4.3 : 重音因子參數.....	37
表 4.4 : 各個音素在音節組成因子中的長度(a)~(h).....	37
表 4.5 : 韻律狀態因子.....	40
表 4.6 : 雙元音的組合.....	47
表 4.7 : function word 與 content word 的時間比較.....	48
表 4.8 : target word 與 target word 邊界的時間比較.....	48
表 4.9 : 介係詞在對比句強調與非強調位置.....	49

圖目錄

圖 2.1：不良的切割位置範例 1.....	5
圖 2.2：不良的切割位置範例 2.....	6
圖 2.3：欲尋找的切割位置.....	6
圖 2.4：音節切割的狀態轉移.....	7
圖 2.5：修改過後的音節切割資訊 1 (V+sp+T)	7
圖 2.6：修改過後的音節切割資訊 2 (L+sp+T).....	8
圖 2.7(a)：短句形音節長度平均值與標準差的關係 (b)：長文形音節長度平均值與標準差的關係.....	8
圖 2.8(a)：短句形 curve fitting (b)：長文形 curve fitting	9
圖 2.9(a)：短句形 mean 與 alpha 的關係 (b)：短句形 mean 與 beta 的關係.....	9
圖 2.10(a)：長文形 mean 與 alpha 的關係 (b)：長文形 mean 與 beta 的關係.....	10
圖 2.11：利用 EM algorithm 找出的混合數三的高斯混合模型與直方圖的比較.....	12
圖 2.12：修正過後的基頻直方圖.....	12
圖 2.13：良好的基頻修正結果.....	13
圖 2.14：不良的基頻修正結果 1.....	13
圖 2.15：不良的基頻修正結果 2.....	14
圖 2.16：基頻軌跡斷點示意圖.....	14
圖 2.17：基頻軌跡修正示意圖.....	15
圖 2.18：兩階段修飾後的基頻軌跡實例 1.....	15
圖 2.19：兩階段修飾後的基頻軌跡實例 2.....	16
圖 2.20(a)：L1 語者的對數基頻平均值與標準差 (b)L2 語者的對數基頻平均值與標準差.....	17
圖 2.21：變動劇烈的基頻軌跡示意圖.....	17

圖 2.22：起始停頓標記決策樹.....	19
圖 2.23：停頓標記模型相似度上升趨勢.....	20
圖 2.24：Break syntax model 決策樹.....	20
圖 2.25：break acoustic model 決策樹.....	21
圖 2.26：四種停頓類型的聲學特性分佈.....	23
圖 4.1：音節時間模型之訓練過程相似度變化.....	35
圖 4.2：音節基頻軌跡模型之訓練過程相似度變化.....	41
圖 4.3：語群平均基頻軌跡.....	41
圖 4.4：重音因子基頻軌跡.....	42
圖 4.5：將重音因子參數轉換為 semitone.....	42
圖 4.6：文獻[22]中各種重音類型的基頻軌跡.....	43
圖 4.7：音節前停頓類型因子基頻軌跡.....	43
圖 4.8：音節後停頓類型因子基頻軌跡.....	44
圖 4.9：不同類型反問句的基頻軌跡.....	45
圖 4.10：單元音的基頻軌跡.....	45
圖 4.11：Vowel chart.....	46
圖 4.12：雙元音及兒化音的基頻軌跡.....	46
圖 4.13：雙元音在 vowel chart 上的移動圖示.....	47
圖 4.14：function word 與 content word 的基頻軌跡比較.....	48
圖 4.15：target word 與 target word 邊界的基頻軌跡比較.....	49
圖 4.16：介係詞在對比句強調與非強調位置.....	50

第一章 緒論

1.1 研究動機

英語是目前臺灣最多人學習的外語，在許多國家被列為官方語言，也是國際場合所廣泛使用的語言。因此，研究以及探討母語使用者(L1)與非母語使用者(L2)間的差異，遂成為一重要的課題。有鑑於此，在 2008 年開始了 AESOP (Asian English Speech cOrpus Project) 跨國語料收集計畫[1, 2]，此計畫收集了諸多亞洲口音的英文語音，以提供研究者分析。在 2009 年，TWNAESOP (TaiWaN Asian English Speech cOrpus Project) [3] 也在中研院的主導下隨之開始，並由此陸陸續續的有許多關於此計畫的研究 [4-6]。

受制於 L2 本身的母語，或多或少都會對其所使用的英語造成影響。僅僅從口語方面，影響就可能包括：音素的不正確、韻律的錯誤以及不當的文法。首先，英語與 L2 的母語，其所使用的音素(phone set)不盡相同。以中文母語者而言，在學習英文發音時，對於某些在中文裡不存在的音素，會感到難以學習甚至會誤用中文中存在的音素來發出該英文的音素。例如，英文中存在 IH (it)以及 IY (eat) 兩個相近的音素，後者在中文裡存在，即注音符號的一，但前者並沒有，也因此相當多中文母語者唸出前者的音時，會常以 IY 來取代。其次，英文是一種在 stress 與 stress 之間長度較為穩定的語言(stress-timed language)，而中文是一種傾向每個音節都同等長度的語言(syllable-timed language)。因此，我們常常會發現中文母語者會錯誤的將英文中的每個音節都唸得一樣長而缺乏重音與否的影響[4, 13, 14]。此一情形在句調(intonation)上，也會造成聽者難以快速獲取語者說話的重點。

有鑑於以上的現象，本研究著眼於發展一個新的韻律模型來描述英文的韻律，並希冀能藉此來觀察及探討 L1 與 L2 之間的差異。從語言學習的角度，許多差異已經被得知，但未知的是這些差異是如何表現在數據上，而不僅於感知上。

1.2 文獻回顧

研究 L1 與 L2 間的差異是一個新穎的研究，但在過去，已有許多文獻探討 L1 的韻律現象；也包括有 ToBI [7](Tone and Break Indices, 以下除了另有說明外，均指其中的 Tone 部份)，一個以簡單的符號來標記基頻軌跡的方法，以及 Fujisaki Model [8]，以數學模型來產生句調等著名的研究。

對 1.1 節所說的，本身母語對於非母語(EFL, English as Foreign Language)使用者的影響，也已經有相當豐碩的研究結果[9-14]。然而，這些研究往往受限於人工標記的困難；以 ToBI 為例，ToBI 已經被廣泛接受為標記基頻軌跡的重要方法，也有文獻[10]是基於同樣的 ToBI 標記，來探討 L1 及 L2 的表現。但是，目前依然缺乏一個能夠自動標記的有效方法，這使得研究所需的語料庫一者必須經過設計，二者必須人工標記，因此語料庫嚴重受限，也缺乏彈性。

除卻 1.1 節中提到的現象，L2 所受到的教育程度、學習英文的時間長短，也都對 L2 的表現產生影響[9]。這些影響也還會表現在重音(lexical stress)[11, 14]、句子中的突顯(prominence, sentence stress)[12]以及句調。除此之外，受制於中文語言的各種現象，也在文獻中被提出[11-14]。

綜合以上，許多學者已經進行了 L1 及 L2 之間差異的研究，但較缺乏的是透過數學模型來描述並且能比較兩者差異的方法，本研究即在此發揮。

1.3 研究方向

基於過去在中文語音上的研究成果[15, 16]，本論文提出一個以音節為單位的韻律模型。此模型包含了時間模型以及基頻軌跡模型。兩個模型皆考慮了其所可能受到的影響因子(affecting factor)影響，進而會對其產生拉長或壓縮/升高基頻或降低基頻。我們考慮的因子包括了 L1 或是 L2、語者、語句類型、重音類型、鄰近的停頓、句尾的類型以及音節內含的音素等等。

為了標記上述的影響因子，除了部份可以由字典或字面上就能輕易的獲得以

外，停頓的類型則必須採用自動標記以消除人為標記不一致的現象及節省人工，本論文也引用了[17]當中的 CART 方法來訓練一個標記停頓類型的決策樹。

本論文透過一個疊代式的訓練法，分別得出 L1 及 L2 的韻律模型及各項影響因子的參數。透過參數的分析，我們得以由數據上的觀點明確指出兩者間的差異，並透過此來有效的找出 L2 為何在韻律從感知上不如及不像 L1 的原因所在。

1.4 章節概要說明

本論文共分為下列五章

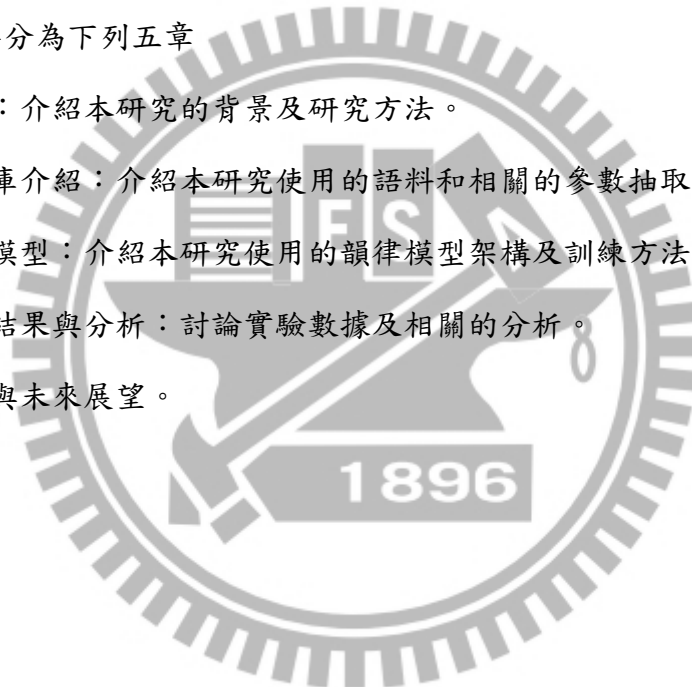
第一章 緒論：介紹本研究的背景及研究方法。

第二章 資料庫介紹：介紹本研究使用的語料和相關的參數抽取。

第三章 韻律模型：介紹本研究使用的韻律模型架構及訓練方法。

第四章 實驗結果與分析：討論實驗數據及相關的分析。

第五章 結論與未來展望。



第二章 資料庫介紹

TWNAESOP[3]是由中研院語言研究所鄭秋豫博士等在 2009 年開始建置。語者包含英文母語者 13 人(男性 6 人，女性 7 人)，中文母語者 488 人(男性 231 人，女性 257 人)，共有 501 人。每位語者依序朗讀相同的短句組、短篇文章、誘發性人機對話及自發性看圖描述等部份(詳細語句見附錄 1，設計緣由見[3])。由於本語料庫仍未處理完畢，本研究僅使用其中的 L1 13 人及 L2 83 人的語料；語料僅使用第 1 至 7 段。本章節將對此語料庫及使用的資料做介紹，在 2.1 節介紹語料庫的資料格式；2.2 節介紹本論文使用的資料前處理；2.3 節介紹語料庫的相關統計資料。

2.1 語料庫簡介

本語料庫類型為麥克風語料，使用頭戴式麥克風(Sennheiser PC155)，音檔格式為 wav 檔；取樣頻率為 16000 赫茲 (Hz)；位元率為 16 位元；聲道為單聲道。每位語者約有 1200 個音節，長約 6 分鐘(僅包含第 1 至 7 段)。錄音軟體為 TWNAESOP Recording Tool (以香港中文大學與中國科學院深圳先進技術研究院共同開發之錄音軟體 CUHK-SIAT recording tool.exe 為雛型，依此錄音實驗目的修訂而成)。錄音環境為安靜的教室、辦公室、實驗室。

2.1.1 文字資料

本語料庫所使用的時間資訊處理到音素階層，是由 HTK 先做強迫對齊後，再經過人工進行校對。使用的標音方式為 CMU 字典[18]所使用的 Arpabet，對於任何的發音錯誤均以音檔為標準進行修正。由於此切割資訊經常將爆破音(stops)多包含欲念出爆破音前的閉嘴靜音(closure silence)，此部份會在前處理階段再進

一步修正。

2.2 資料前處理

在本小節中，將會介紹擷取各個在之後的演算法中需要用到的各項參數的方法。包括有：音節切割資訊、音節基頻軌跡參數、停頓類型、標記停頓類型所需的各項參數，同時也會介紹參數正規化的方法。

2.2.1 音節切割資訊修正

雖然我們的人工切割位置已經足夠準確，但是在爆破音的地方卻會包含一部份的 silence，情況如下圖：

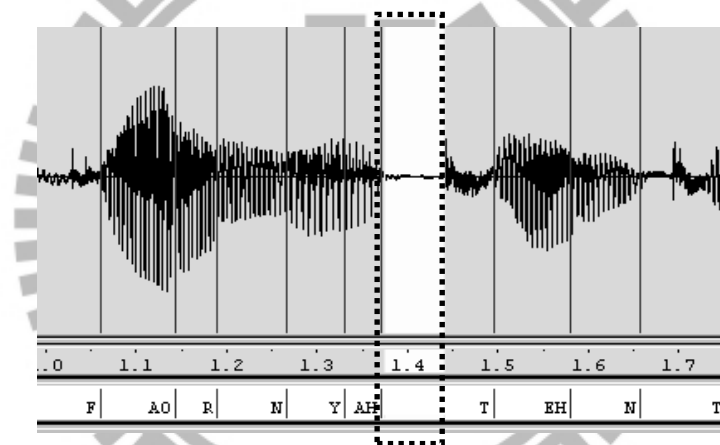


圖 2.1：不良的切割位置範例 1

此圖中為 California 及 ten 中間應有一段靜音沒有被切出。

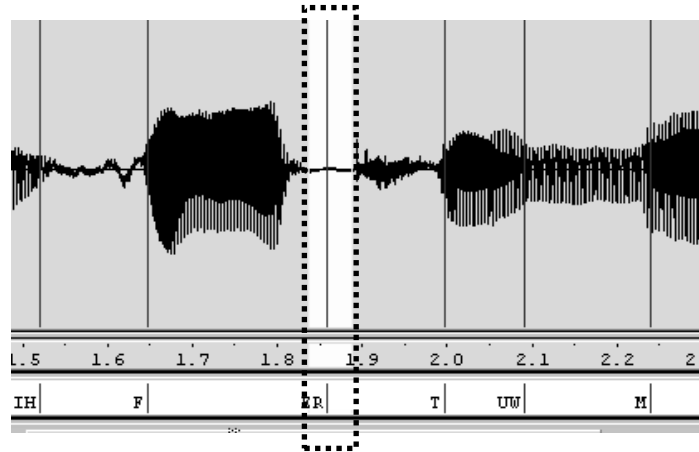


圖 2.2：不良的切割位置範例 2

此圖中為 prefer 及 to 中間應有一段靜音沒有被切出。

本論文使用 segmental k-means 來調整(voice phone)+(stops)中的切割位置，其方法如下：

對於每個句子，將句首可靠的靜音位置，抽取 13 維 MFCC(Mel-Frequency Cepstrum Coefficient)，抽取設定為：frame size：5ms, frame shift：2ms。訓練一個混合數為 1 的高斯混合模型(GMM, Gaussian Mixture Model)。再以欲尋找 sp 切割位置前段音檔的一部分(長度為原始 voice phone 的長度)，訓練 V1 model。以 sp 後段音檔一部分(30ms)，訓練 V2 model。如下圖所示，欲尋找 F-AY-V 和 T-AY-M-Z 之間的短停頓(sp)，實線框為 V1，虛線框為 V2。

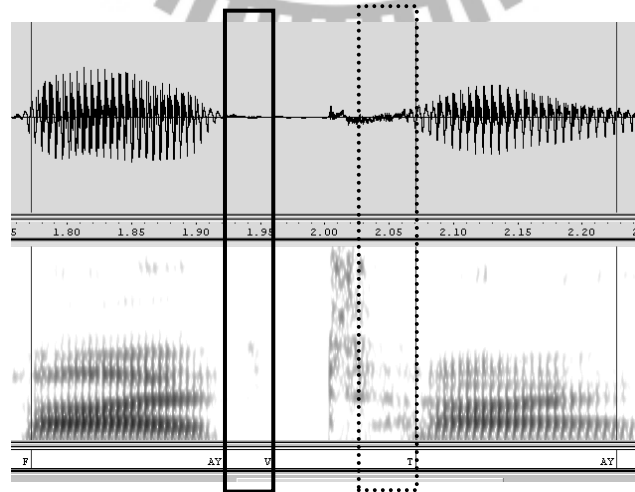


圖 2.3：欲尋找的切割位置

訓練完之後以 Viterbi algorithm 來調整 sp 的切割位置，其狀態轉移方式如下圖所示：

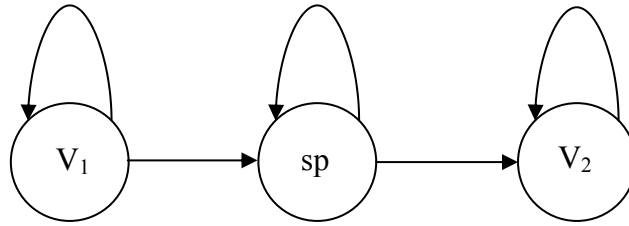


圖 2.4：音節切割的狀態轉移

走完之後再 back trace 回來就可以取得新的切割位置。處理完的結果如下：
實線為原始切割位置，虛線為調整過後的切割位置。

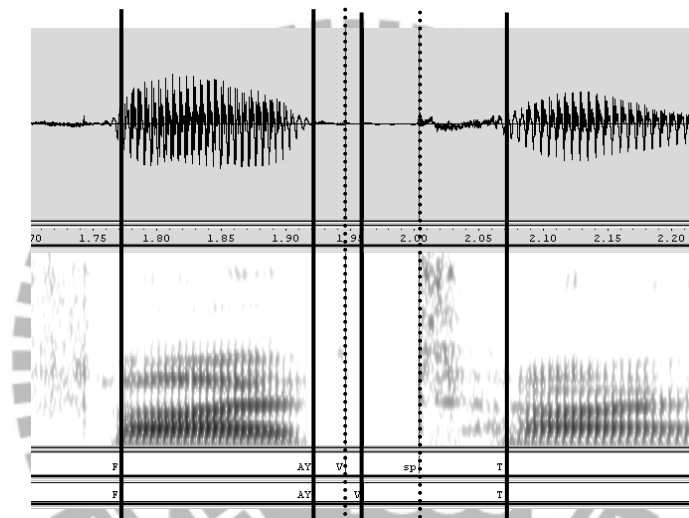


圖 2.5：修改過後的音節切割資訊 1 (V+sp+T)

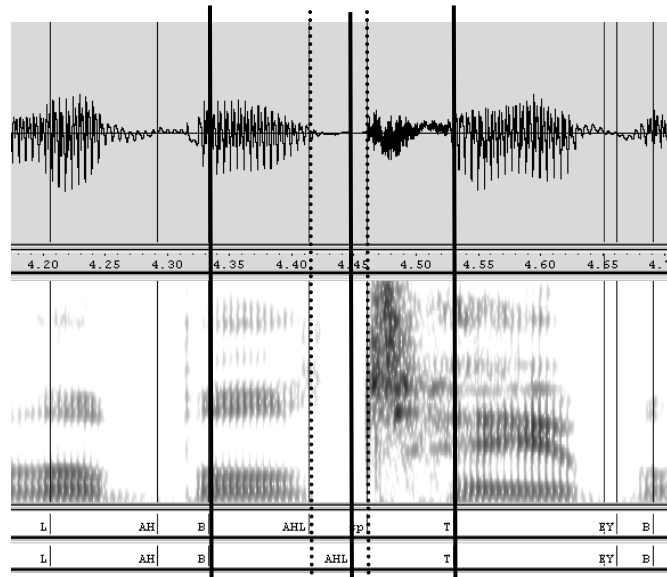


圖 2.6：修改過後的音節切割資訊 2 (L+sp+T)

2.2.2 音節時間資訊正規化

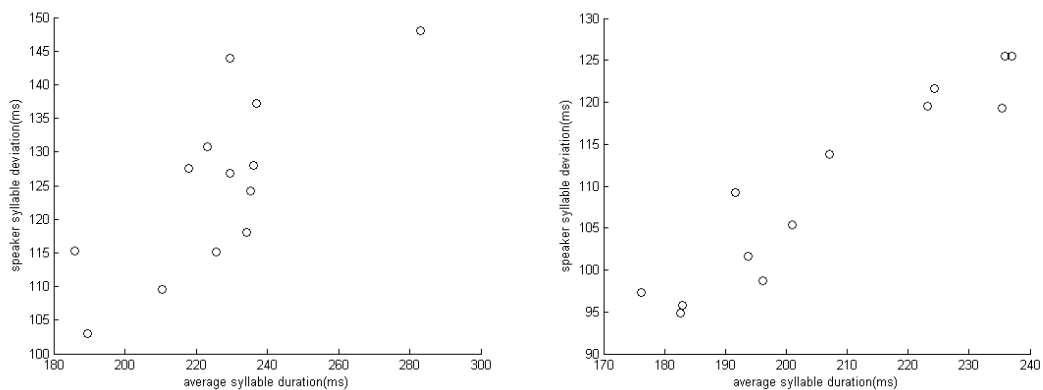


圖 2.7(a)：短句形音節長度平均值與標準差的關係 (b)：長文形音節長度平均值與標準差的關係

圖 2.7(a)與圖 2.7(b)為 L1 語者其音節長度平均值與標準差的關係，每一點是一個語者。可以看出平均值越大的，標準差也就會越大，大致上呈現線性關係，尤其長文會比短句更加明顯。

因此，本論文採用對語者及語句類型分開的正規化方法，其詳細步驟如下：

1. 求出每群的 mean 和 standard deviation，並用一條 2 維曲線做 curve fitting。

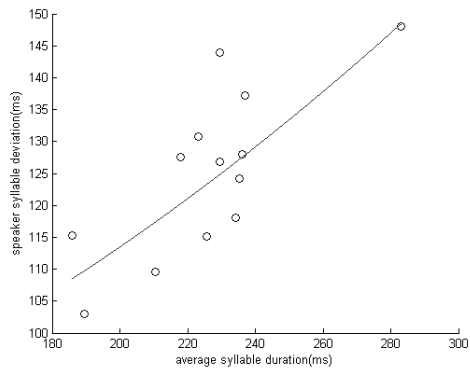
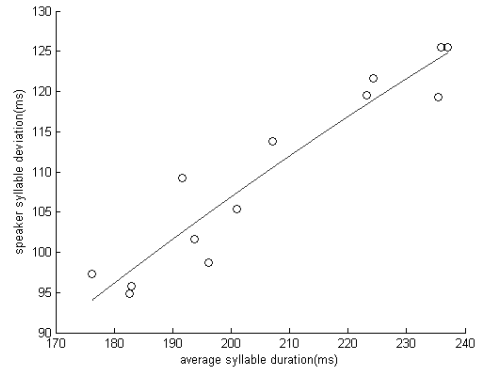


圖 2.8(a)：短句形 curve fitting



(b)：長文形 curve fitting

2. 由此曲線可以得到每個 mean 對應的 standard deviation。利用 mean 和 standard deviation 就可以算出 gamma distribution 的 alpha 和 beta 如下式：

$$\alpha = \frac{\mu^2}{\sigma^2}, \quad \beta = \frac{\sigma^2}{\mu} \quad (2-1)$$

經過此方法後，mean 與 alpha 及 mean 與 beta 的關係圖如下

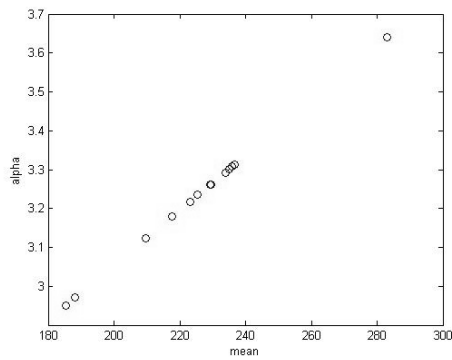
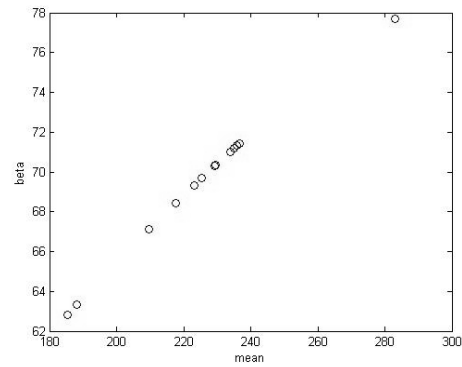


圖 2.9(a)：短句形 mean 與 alpha 的關係



(b)：短句形 mean 與 beta 的關係

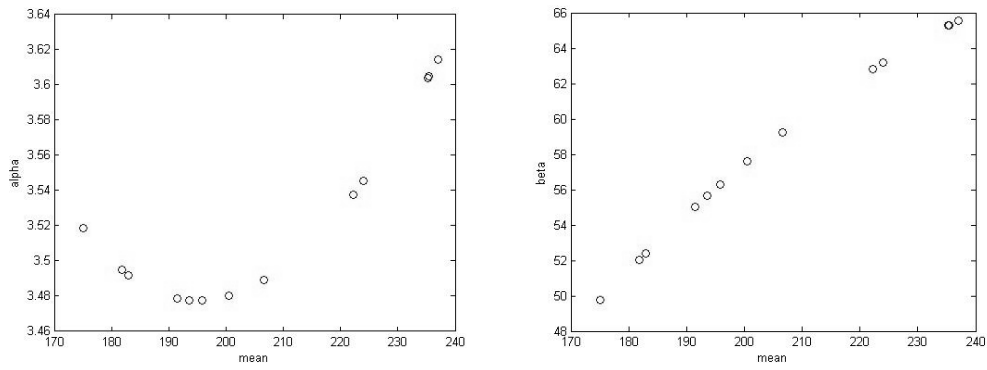


圖 2.10(a)：長文形 mean 與 alpha 的關係 (b)：長文形 mean 與 beta 的關係

3. 接著求出全部的平均 alpha 和 beta，再做 CDF 轉換如下式：

$$x'_{n,g} = G^{-1}\left(G\left(x_{n,g}, \tilde{\alpha}(SR_i), \tilde{\beta}(SR_i)\right), \bar{\alpha}, \bar{\beta}\right) \quad (2-2)$$

其中 $x'_{n,g}$ 和 $x_{n,g}$ 分別為正規化後和原先的音節長度， $\tilde{\alpha}(SR_i)$ 及 $\tilde{\beta}(SR_i)$ 是經過 2-1 求出的 alpha 及 beta， $\bar{\alpha}$ 及 $\bar{\beta}$ 是平均的 alpha 和 beta。

2.2.3 基頻資訊修正

先以 wavesurfer 用 ESPS 自動求出所有音檔的基頻，其參數設定如表 2.1，再利用兩階段修正基頻軌跡。第一階段使用[19]中提出的方法加以修飾：

表 2.1：wavesurfer 的基頻求取設定

Min pitch	50
Max pitch	600
Frame size	7.5ms
Frame shift	10ms

$$T_n = (1 + X_n)T_{n-1} \quad (2-2)$$

$$T_n = T_0 \prod_{k=1}^n (1 + X_k) \quad (2-3)$$

$$\log(T_n) \approx \log(T_0) + \sum_{k=1}^n X_k \quad (2-4)$$

$$\log(T) \propto N(\mu, \sigma^2) \quad (2-5)$$

T_n 為第 n 個音框的基頻

根據中央極限定理，取對數之後的 T 會趨近於一個高斯密度函數。考慮到求出的基頻可能會有倍頻(double pitch)或半頻(half pitch)的現象，可以表示成：

$$f(x) = \begin{cases} \frac{x}{2} & \text{with prob. } \beta \\ x & \text{with prob. } \alpha \\ 2x & \text{with prob. } 1 - \alpha - \beta \end{cases} \quad (2-6)$$

$$\hat{F}_0 = f(F_0) \quad (2-7)$$

我們可以得到：

$$\log(\hat{F}_0) \sim \beta \cdot N(\mu - \log(2), \sigma^2) + \alpha \cdot N(\mu, \sigma^2) + (1 - \alpha - \beta) \cdot N(\mu + \log(2), \sigma^2) \quad (2-8)$$

也就是求出的基頻可以表示成一個混合數為 3 的高斯混合模型，於是我們可以利用 EM algorithm 來得到這個高斯混合模型如下圖，在此例中，倍頻並不明顯。

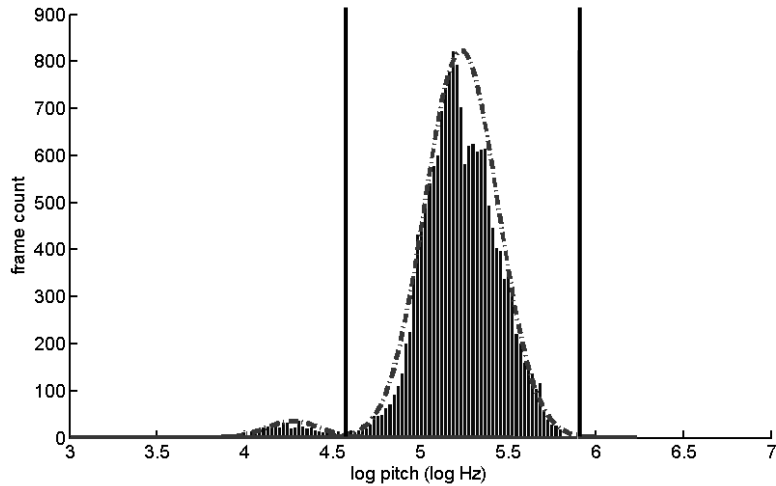


圖 2.11：利用 EM algorithm 找出的混合數三的高斯混合模型與直方圖的比較

找到三個 mixture 的參數之後，接下來利用下式來決定何者以上/以下可能是 double/half pitch:

$$N(x; \mu, \sigma^2) > N(x; \mu - \log(2), \sigma^2) \text{ left boundary} \quad (2-9-1)$$

$$N(x; \mu + \log(2), \sigma^2) > N(x; \mu, \sigma^2) \text{ right boundary} \quad (2-9-2)$$

找到門檻值(threshold)以後，如果基頻資料中有連續兩個音框落在可能是 double/half pitch 的範圍內，就將其修正。將圖 2.11 修正過後的分佈圖如下：

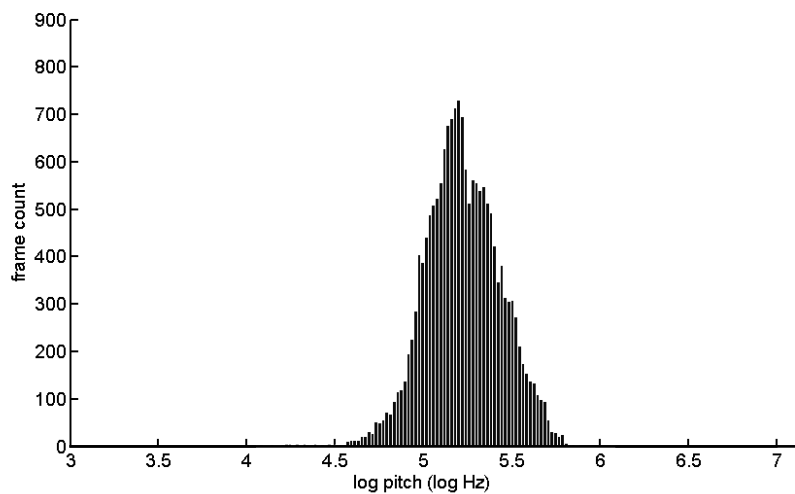


圖 2.12：修正過後的基頻直方圖

結果如下：

良好的結果：

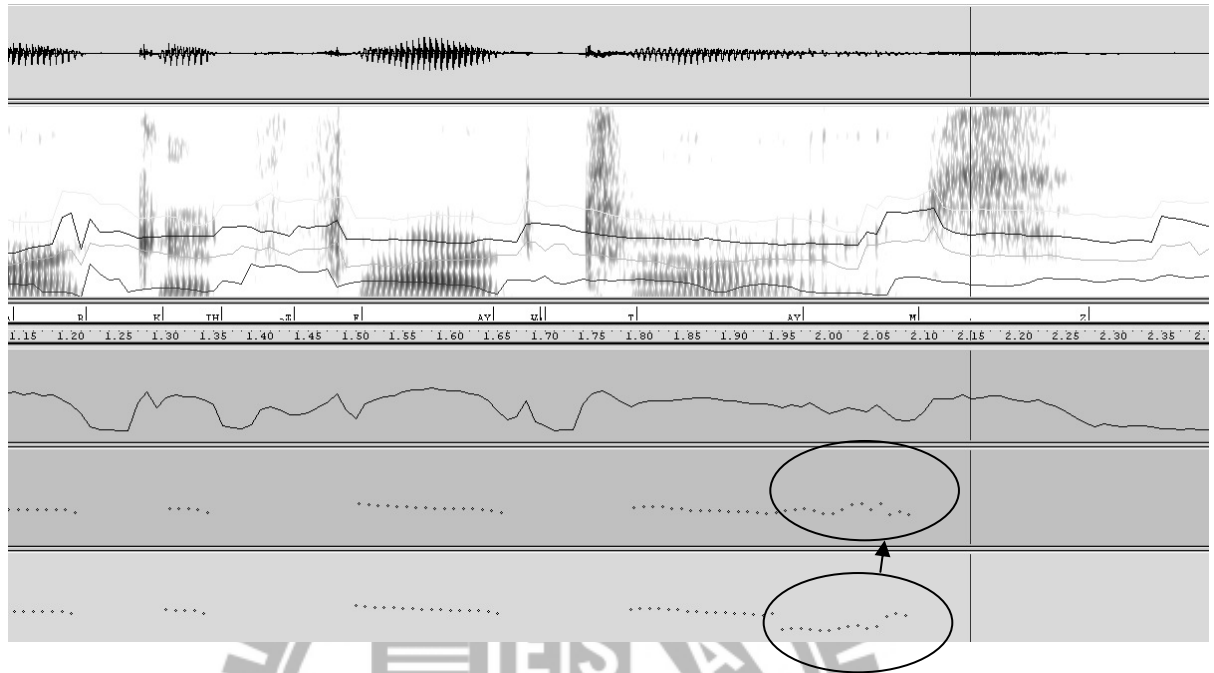


圖 2.13：良好的基頻修正結果

不好的結果：

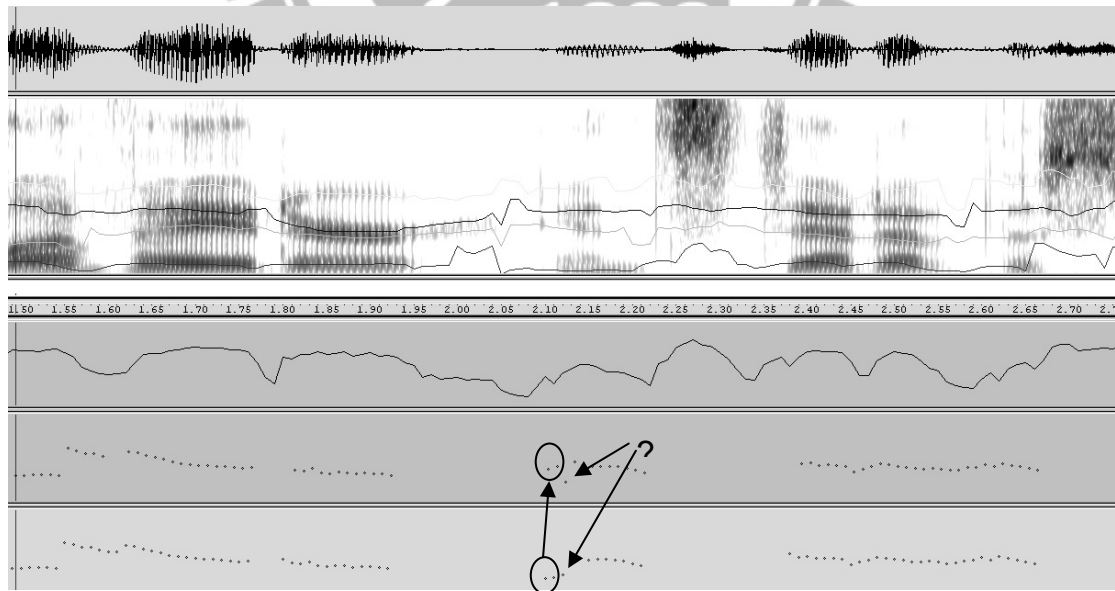


圖 2.14：不良的基頻修正結果 1

圖 2.14 是由於該點的 pitch 值恰巧高於 threshold，因此沒有被修正。

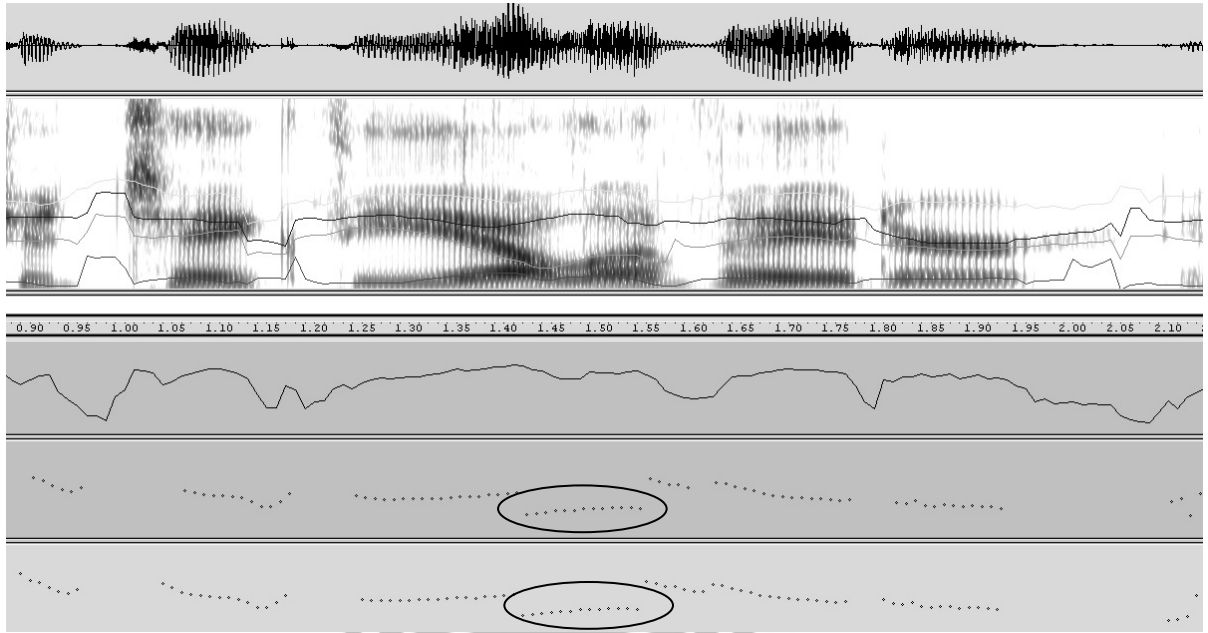


圖 2.15：不良的基頻修正結果 2

圖 2.15 也是相同的情況。

為了解決此問題，本論文提出如下的解決方法：

一段不間斷的 pitch contour，pitch 的變化不應該會有過大的落差

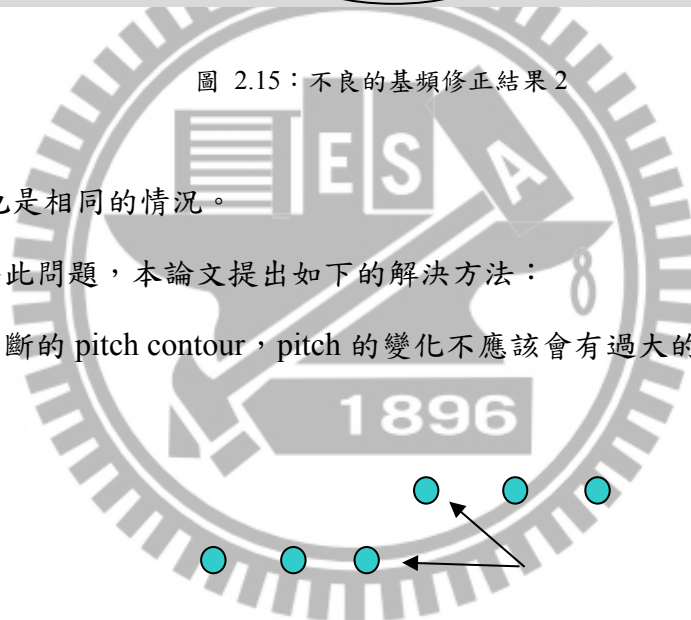


圖 2.16：基頻軌跡斷點示意圖

如圖 2.16，這兩點可能是一個斷點，也就是說：

$$\text{Break point criterion: } |2T_n - T_{n+1}| < |T_n - T_{n+1}| \quad (2-10-1)$$

$$|T_n / 2 - T_{n+1}| < |T_n - T_{n+1}| \quad (2-10-2)$$

找出斷點後，根據哪一段較長決定正確者

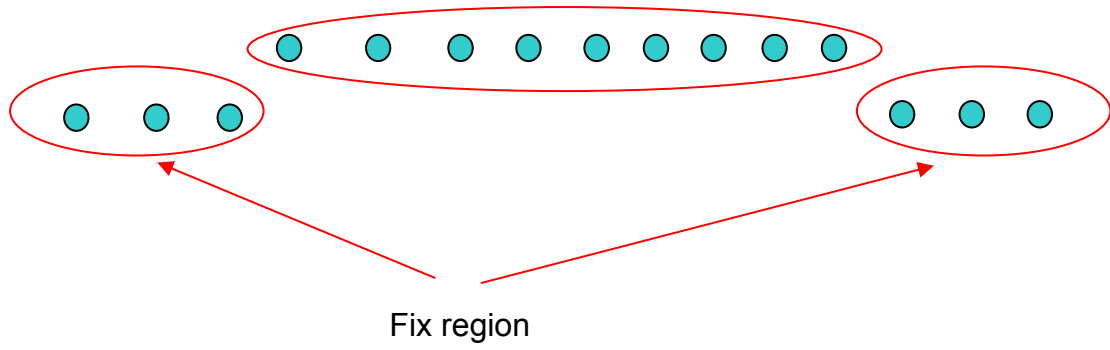


圖 2.17：基頻軌跡修正示意圖

如上圖，中間較長，因此將中間段視為正確者，而其他的往上調。

調整過後實例如下兩張圖：

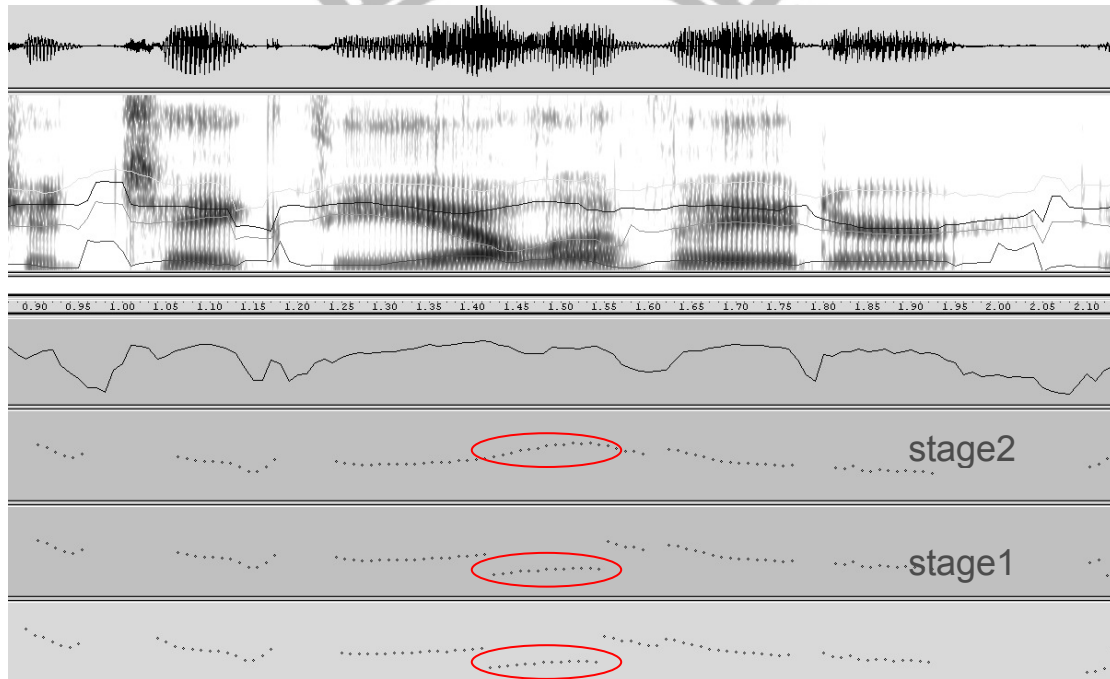


圖 2.18：兩階段修飾後的基頻軌跡實例 1

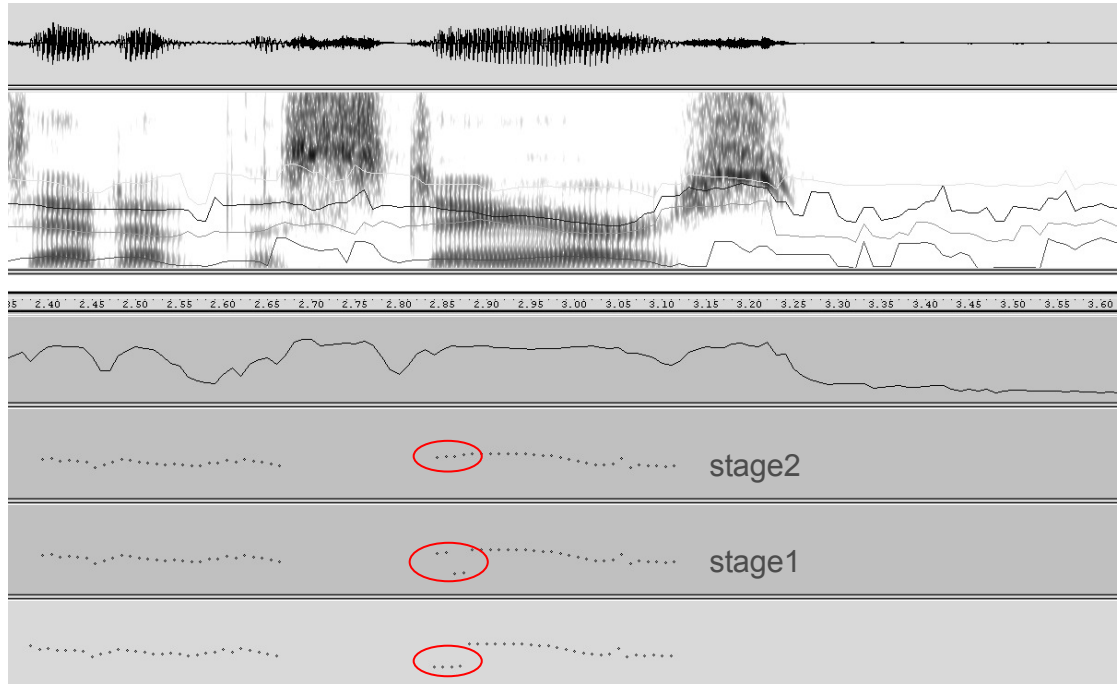


圖 2.19：兩階段修飾後的基頻軌跡實例 2

利用兩階段可以有效修正大部分錯誤。先以 stage 1 處理後，可以修正明顯的錯誤，因此可以使得 stage 2 在進行投票時，不會有錯誤者反而較多的現象。

2.2.4 基頻資訊正規化

與 2.2.2 小節類似，本論文也會對基頻進行如下的語者正規化：

$$f' = \frac{f - \mu_s}{\sigma_s} \sigma_g + \mu_g \quad (2-11)$$

其中 f 及 f' 是修正前後的對數基頻， μ_s 和 σ_s 是語者的對數基頻平均以及標準差， μ_g 和 σ_g 是全部語者的對數基頻平均和標準差。透過 2-11 式可以將全部語者的對數基頻都對應到同樣的平均和範圍。圖 2.20(a)、(b) 分別為 L1 及 L2 的對數平均基頻與標準差，其中可以看出 L1 語者無論基頻高低，其標準差變化不大，而 L2 則有較大的變化。

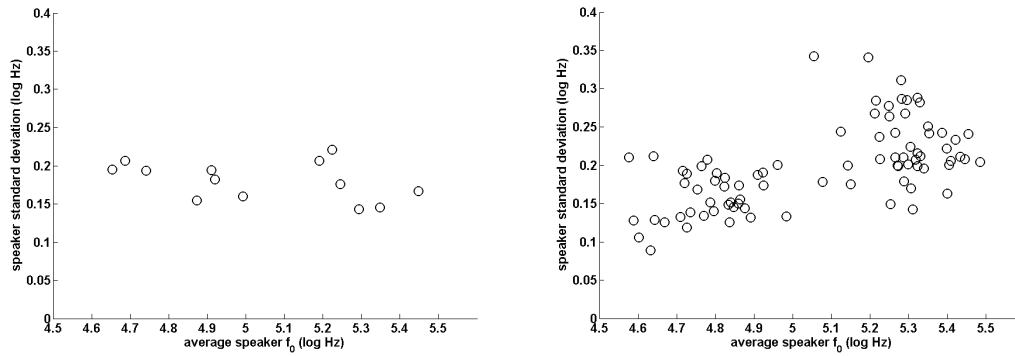


圖 2.20(a)：L1 語者的對數基頻平均值與標準差 (b)L2 語者的對數基頻平均值與標準差

2.2.5 音節基頻軌跡參數抽取

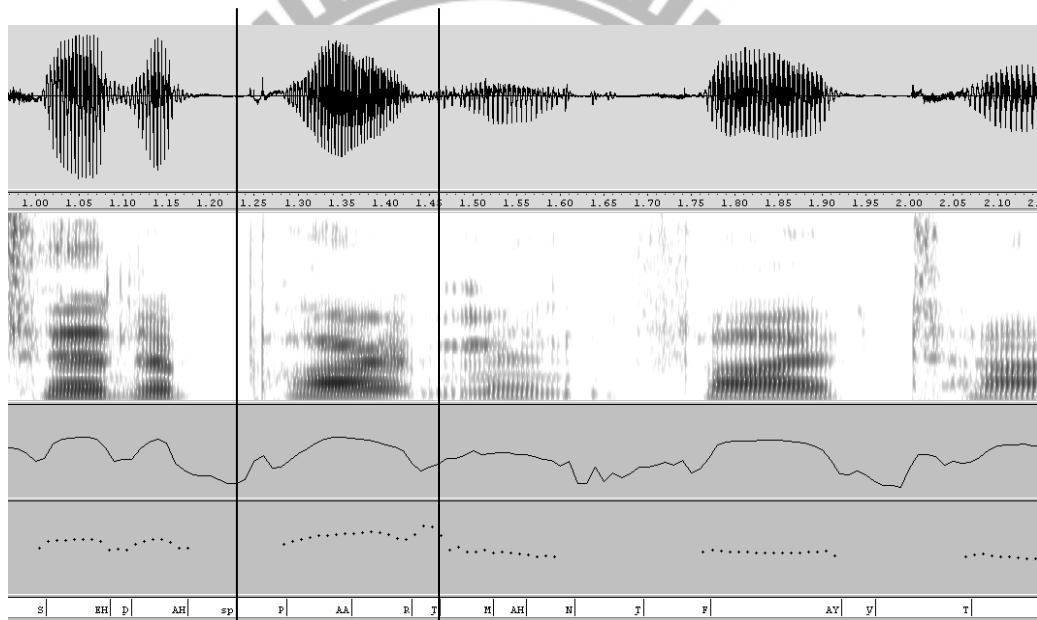


圖 2.21：變動劇烈的基頻軌跡示意圖

由圖 2.21，被標記起來的音節 P-AA-R-T 中可以看出，被修正過後的基頻軌跡，仍然會在某些區段產生劇烈的變動。在詳細觀察之後，我們可以發現這些變動較大的基頻軌跡多半發生在音節中的子音部份，如圖 2.21，在 T 的部份有較大的變動。為了修正此問題，本論文採用的基頻軌跡將只保留音節其中的元音，以及與元音相連的鼻音(有穩定的頻譜)、流音和半母音(這兩者會與元音較緊密結合)的部份。

獲得了較好的基頻區段以後，為了能夠將基頻軌跡轉換為參數，本論文採用 [16]當中的方法，將基頻軌跡投影到這四個正交基底。下表為去除字音基頻後，做四維正交展開，其誤差的比較，單位為 log Hz。

表 2.2：刪除子音基頻前後，其投影到四維正交基底造成的誤差

L1：

Order	1	2	3
RMSE(10^{-4})	25	12	7

刪除子音 pitch 之後

Order	1	2	3
RMSE(10^{-4})	25	12	7

L2：

Order	1	2	3
RMSE(10^{-4})	22	12	8

刪除子音 pitch 之後

Order	1	2	3
RMSE(10^{-4})	20	10	7

2.2.6 停頓類型標記

本論文依照[17]中的方法，訓練停頓-語言模型(break syntax model)以及停頓-聲學模型(break acoustic model)。所使用到的問題集見表 2.4，使用到的聲音參數為：停頓時間(pause duration)、基頻變動(pitch jump)及能量變動(energy jump)，三者均先經過正規化。以 L1 資料訓練此模型後，將 L2 資料正規化至與 L1 相同後，標記 L2 的停頓類型。其訓練流程簡述如下：

(1) 依[17]中的 initial 階段找出三個門檻值設定初始停頓標記，其決策樹如圖 2.22 所示。

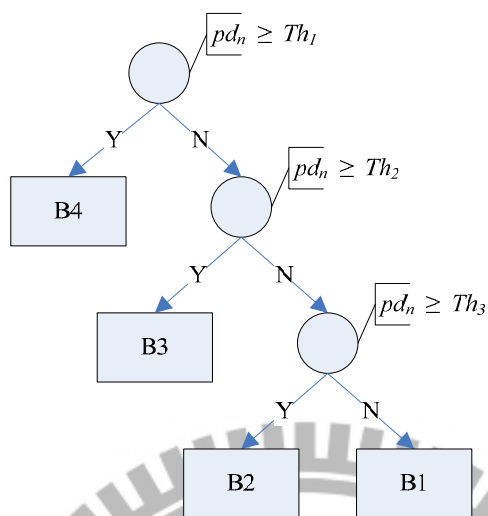


圖 2.22：起始停頓標記決策樹

表 2.3：起始停頓標記門檻值

Th1	472
Th2	199
Th3	61

- (2) 由(1)的結果依照論文中的方法訓練上述兩個起始模型。
- (3) 再由訓練出的模型標記停頓。
- (4) 由(3)的結果再次訓練兩個模型並計算其相似度(likelihood)。
- (5) 重覆(3)(4)直到相似度收斂為止。

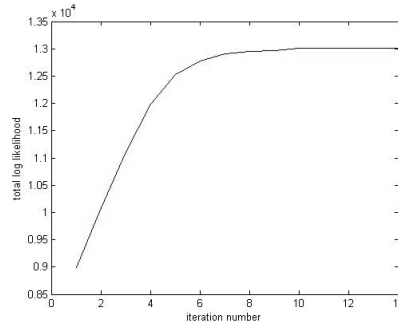


圖 2.23：停頓標記模型相似度上升趨勢

其所訓練出來的停頓標記模型，相似度上升如圖 2.23。Break syntax model 決策樹如圖 2.24。每個節點下方的直方圖代表該節點 B1~B4(由左至右)所佔的比例；直方圖下方的數字代表該節點包含的樣本數；每個節點右邊(實線)表示詢問問題為“是”，左邊(虛線)表示為“否”。

由圖 2.24 中可以看出，區分停頓長度的重要語言參數(Linguistic feature)為“是否為 PM(Q1)?”，以及“是否為 inter-word(Q3)?”。除此之外，我們也發現當右側為 function word 時，其停頓時間會較長。

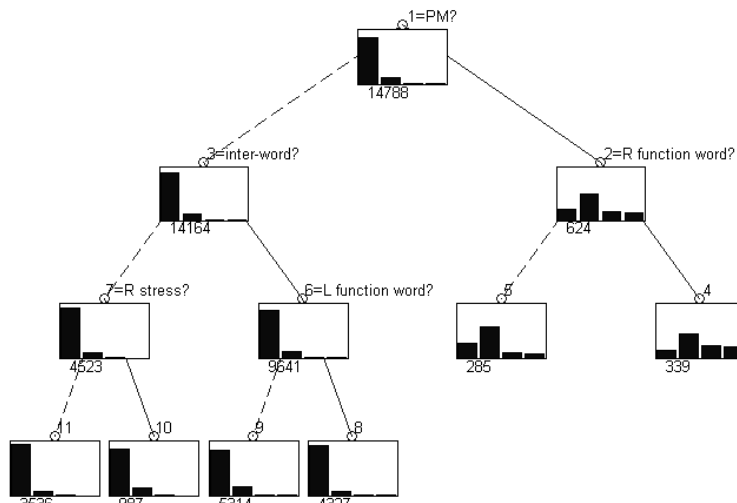


圖 2.24：Break syntax model 決策樹

表 2.4：決策樹所使用的問題集

Question number	
Q1	是否為 PM
Q2	是否為 PM 中的 . (句號)
Q3	是否為 PM 中的 , (逗號)
Q4	是否為 PM 中的 ; (分號)
Q5	是否為 PM 中的 ? (問號)
Q6	是否是 inter-word?
Q7.1	左邊是否是 stress syllable
Q7.2	右邊是否是 stress syllable
Q8.1	左邊是否是 function word
Q8.2	右邊是否是 function word

其所訓練出來的四個 Break acoustic model 如圖 2.25(a)~(d)。每個節點下方數字表示該節點平均停頓長度以及樣本數。由圖 2.25(a)中可以看出，B1 的停頓時間都較短，而最重要的問題則是“是否為 inter-word(Q1)？”。



圖 2.25 (a)：break acoustic model 決策樹 for B1(緊密結合停頓)

圖 2.25(b)中，最重要的問題與(a)相同。另外當左邊是 stress syllable，也會對停頓有影響。

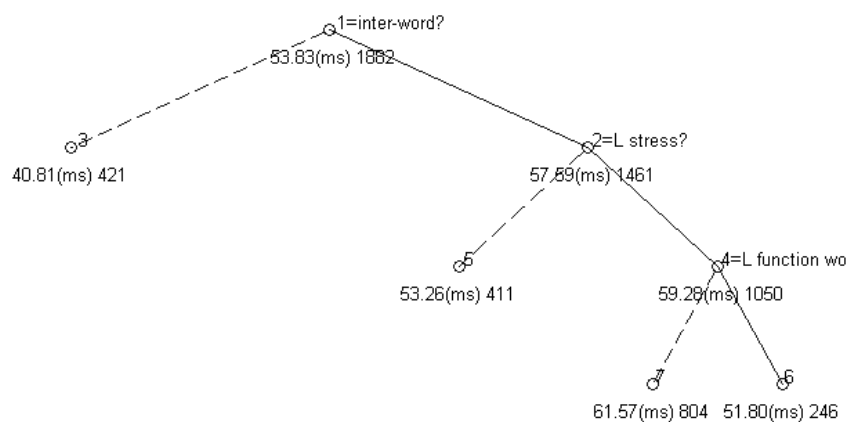


圖 2.25 (b) : break acoustic model 決策樹 for B2(音節或詞邊界停頓)

圖 2.25(c)中可以看出，B3 的決策樹較為簡單，唯一的問題為是否是逗點。

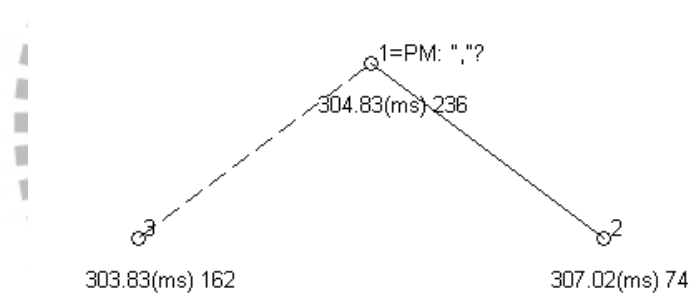


圖 2.25 (c) : break acoustic model 決策樹 for B3(較短的標點符號停頓)

圖 2.25(d)中，最重要的問題為“是否是句點(Q1)?”，文本中在一段音檔中有句點的並不多，其中有一大部分是在第 3 段有許多“NO. XXX”的語句。

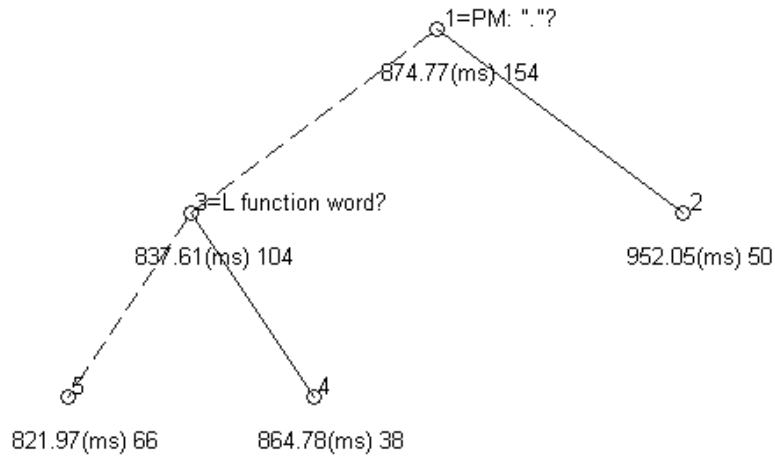


圖 2.25 (d) : break acoustic model 決策樹 for B4(較長的標點符號停頓)

四種停頓的聲音特性如圖 2.26(a)~(c)，其中(a)可以看出四種類型停頓均分的相當的清楚，僅 B1 和 B2 在停頓長度上沒有較明顯的分別；在基頻變動(b)和能量變動(c)上，大致以 B3 和 B4 較大，B1 和 B2 較小，但並不明顯。

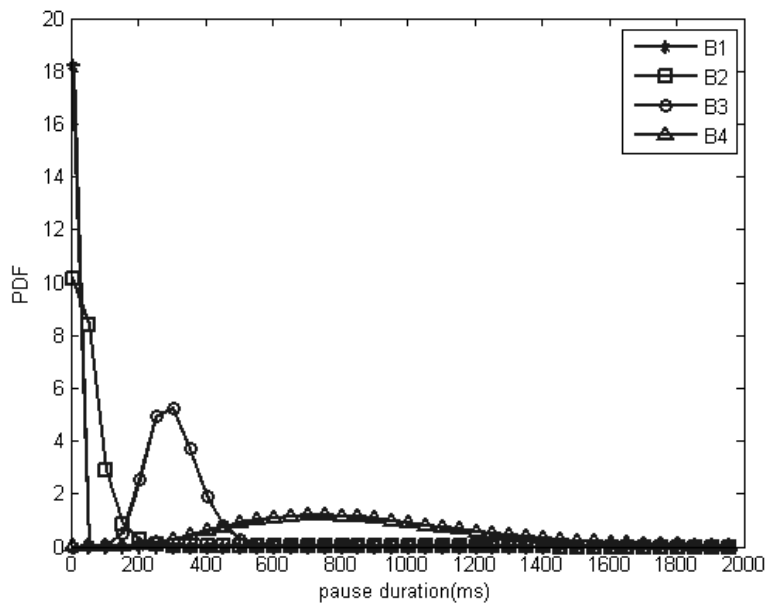


圖 2.26 (a) : 四種停頓類型的聲學特性分佈 -- 停頓長度分佈圖

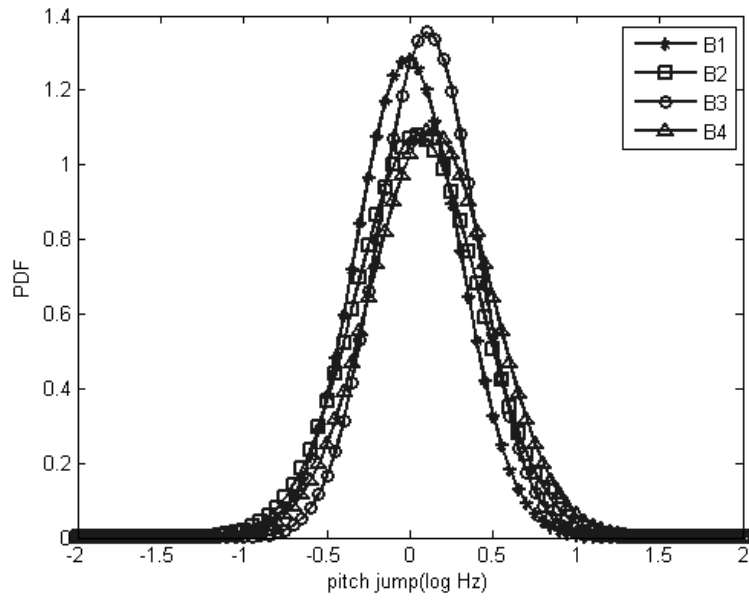


圖 2.26 (b)：四種停頓類型的聲學特性分佈 -- 基頻變動分佈圖

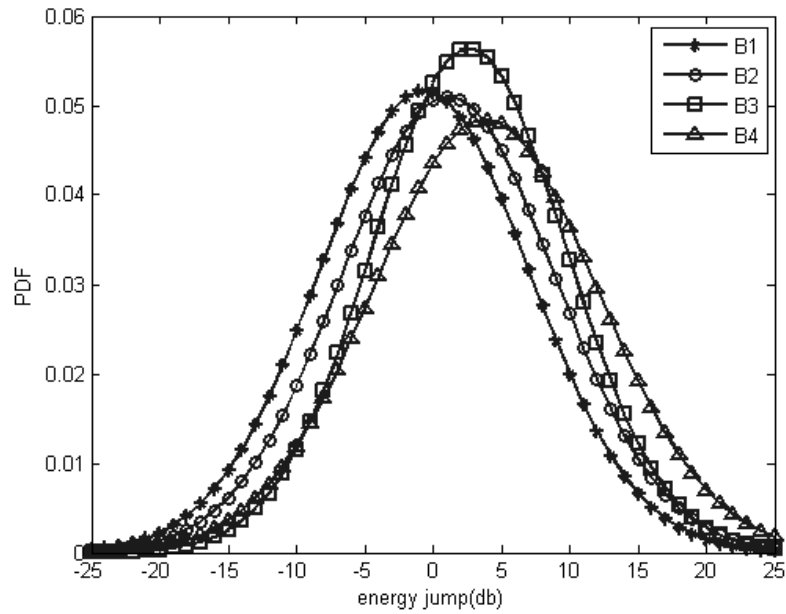


圖 2.26 (c)：四種停頓類型的聲學特性分佈 -- 能量變動分佈圖

2.3 語料庫統計資料

以下列出語料庫使用範圍內，各樣統計資料：

表 2.5 : L1 phone 統計資料

(a)單元音(monophthong)

Phone	Count	μ (ms)	σ
AO	819	76	44
AA	677	87	49
IY	1625	123	69
UW	923	95	74
EH	1790	97	47
UH	251	74	41
IH	2188	68	29
AH	2620	71	53
AE	1052	108	59
OW	776	195	103

(b)雙元音(diphthong)及兒化音(r-colored)

Phone	Count	μ (ms)	σ
EY	741	134	56
AY	1374	169	69
AW	102	178	84
OY	52	283	73
ER	783	113	66

(c)塞音(stops)

Phone	Count	μ (ms)	σ
P	638	66	26
T	3330	61	44
K	1091	73	41

B	534	44	20
D	1672	46	33
G	260	49	17

(d)塞擦音(affricates)

Phone	Count	μ (ms)	σ
CH	143	126	40
JH	207	93	32

(e)擦音(fricatives)

Phone	Count	μ (ms)	σ
F	935	116	47
V	936	76	44
TH	243	106	57
DH	910	46	19
S	1701	131	41
Z	1072	130	73
SH	286	125	23
ZH	13	83	28
HH	468	65	33

(f)鼻音(nasal)

Phone	Count	μ (ms)	σ
M	1235	95	42
N	3304	77	38
NG	338	92	55

(g)流音(liquid)

Phone	Count	μ (ms)	σ
-------	-------	------------	----------

L	1170	72	32
R	1838	79	46

(h)半母音(semivowel)

Phone	Count	μ (ms)	σ
Y	508	56	27
W	1008	75	41

(i)靜音

Phone	Count	μ (ms)	σ
sp	603	66	179
sil	2873	958	519

表 2.6 : L2 phone 統計資料

(a)單元音(monophthong)

Phone	Count	μ (ms)	σ
AO	5216	109	53
AA	4298	113	48
IY	10162	144	70
UW	5875	134	72
EH	11430	128	51
UH	1569	103	45
IH	13852	101	41
AH	16986	113	61
AE	6673	133	57

(b)雙元音(diphthong)及兒化音(r-colored)

Phone	Count	μ (ms)	σ
-------	-------	------------	----------

EY	4721	160	61
AY	8813	179	58
AW	643	209	92
OY	329	324	82
OW	4841	206	88
ER	4984	151	73

(c) 塞音(stops)

Phone	Count	μ (ms)	σ
P	4062	70	31
T	21240	75	52
K	6921	75	44
B	3395	52	24
D	10558	54	38
G	1633	55	22

(d) 塞擦音(affricates)

Phone	Count	μ (ms)	σ
CH	910	140	53
JH	1327	96	46

(e) 擦音(fricatives)

Phone	Count	μ (ms)	σ
F	5955	122	47
V	5943	87	48
TH	1551	127	53
DH	5766	68	38
S	10946	140	50

Z	6738	151	80
SH	1821	143	54
ZH	83	70	24
HH	2972	102	40

(f)鼻音(nasal)

Phone	Count	μ (ms)	σ
M	7863	100	40
N	21056	88	44
NG	2143	100	55

(g)流音(liquid)

Phone	Count	μ (ms)	σ
L	7501	85	47
R	11650	83	52

(h)半母音(semivowel)

Phone	Count	μ (ms)	σ
Y	3222	77	40
W	6504	89	47

(i)静音

Phone	Count	μ (ms)	σ
sp	8726	97	209
sil	18362	827	477

第三章 韻律模型

在本章節中，將介紹本論文所使用的兩個模型：音節時間模型以及音節基頻軌跡模型。以下將依序介紹模型設計、所使用的參數及訓練流程。

3.1 音節時間模型

時間資訊是影響韻律的重要參數之一。不像中文是傾向每個音節都有相近的長度，英文的音節長度會與各項語言參數高度相關。

3.1.1 影響因子

在本論文中，考慮了以下 7 個會影響音節長度的因子，包括了：語群(L1 或是 L2)、文章類型、語者、該音節所屬的詞有幾個音節、重音的類型、音節中內含的音素組成以及韻律狀態。其中語者和文章類型，已經在 2.2.2 小節中先做正規化處理。

3.1.2 模型架構

本論文仿照[15]中的架構，假設所有的影響因子會是加法性的延長或縮短音節長度，也就可以表示成：

$$x'_{n,g} = \hat{x}_{n,g} + \gamma_{w_n,g} + \gamma_{h_n,g} + \gamma_{s_n,g} + \gamma_{p_n,g} + \mu_g \quad (3-1)$$

其中， $x'_{n,g}$ 是正規化之後觀察到的音節時間資訊， $\hat{x}_{n,g}$ 是此模型的誤差，可以被描述成一個平均值為 0，變異數是 σ_g^2 的高斯分佈； μ_g 是該語群的平均值。 γ_a 表示影響因子 a 的參數。 $w_n \in \{1, 2, 3, 4, 5\}$ 表示該音節所屬的詞的音節數； $h_n \in \{\text{主重音、次重音、無重音、單音節}\}$ ，單音節表示該音節是所屬的詞中唯一的音節，這些標記是由 CMU 字典 0.7a[18]所標記； s_n 表示該音節中所包含的音

素； p_n 表示該音節所屬的韻律狀態，來自於上層 HPG(Hierarchy of Prosodic Phrase Group)[20]造成的影響。

3.1.3 訓練流程

- 初始階段

(1) 計算語群平均：

$$\mu_g = \sum_n x'_{n,g} / N_g \quad (3-2)$$

$$x_{n,g}^{(1)} = x'_{n,g} - \mu_g \quad (3-3)$$

其中 N_g 為語群的樣本數。

(2) 計算詞長因子初始值：

$$\gamma_{w,g} = \left(\sum_n x_{n,g}^{(1)} \delta(w_n = w) \right) / \left(\sum_n \delta(w_n = w) \right) \quad (3-4)$$

$$x_{n,g}^{(2)} = x_{n,g}^{(1)} - \gamma_{w,g} \quad (3-5)$$

(3) 計算重音因子初始值：

$$\gamma_{h,g} = \left(\sum_n x_{n,g}^{(2)} \delta(h_n = h) \right) / \left(\sum_n \delta(h_n = h) \right) \quad (3-6)$$

$$x_{n,g}^{(3)} = x_{n,g}^{(2)} - \gamma_{h,g} \quad (3-7)$$

(4) 計算音節因子初始值：

$$\gamma_{s,g} = \sum_{m \in S} \bar{d}_{m,g} - \mu_g \quad (3-10)$$

對於每個音節 s ，表示為音節中所有音素的長度相加扣除語群平均。其中 $\bar{d}_{m,g}$ 來自於語群 g 的統計資料中，音素 m 的平均長度。

$$x_{n,g}^{(4)} = x_{n,g}^{(3)} - \gamma_{s,g} \quad (3-11)$$

(5) 計算韻律狀態因子初始值：

利用 K-means 演算法將 $x_{n,g}^{(4)}$ 分成 3 群。

$$x_{n,g}^{(5)} = x_{n,g}^{(4)} - \gamma_{p,g} \quad (3-12)$$

(6) 計算變異數：

$$\sigma_g^2 = \sum_n (x_{n,g}^{(5)})^2 / N_g \quad (3-13)$$

● 疊代階段

(1) 重新計算詞長因子參數：

$$\gamma_{w,g} = \left(\sum_n (x'_{n,g} - \gamma_{h,g} - \gamma_{s,g} - \gamma_{p,g} - \mu_g) \delta(w_n = w) \right) / \sum_n \delta(w_n = w) \quad (3-14)$$

(2) 重新計算重音因子參數：

$$\gamma_{h,g} = \left(\sum_n (x'_{n,g} - \gamma_{w,g} - \gamma_{s,g} - \gamma_{p,g} - \mu_g) \delta(h_n = h) \right) / \sum_n \delta(h_n = h) \quad (3-15)$$

(3) 重新計算音節因子參數：

在 ML 的條件下，求出每個音素的長度：

$$\sum_{m \in s_n} p_{m,g} = x'_{s_n} - (\gamma_{w_n} + \gamma_{h_n} + \gamma_{p_n}) \quad (3-16)$$

之後再以類似式 3-10 求出新的參數：

$$\gamma_{s,g} = \sum_{m \in s} p_{m,g} - \mu_g \quad (3-17)$$

(4) 重新計算韻律狀態因子參數：

在第一次疊代時，使用 K-means 演算法將扣除韻律狀態以外其餘因子的殘餘值

分成 16 群；在之後的疊代，則重新標記韻律狀態。

(5) 重新計算變異數。

(6) 重複疊代階段直到收斂為止。

3.2 音節基頻軌跡模型

與中文類似，英文當中的基頻軌跡會與該音節的語言參數有關，因此，音節

基頻軌跡，也是影響韻律的重要參數。

3.2.1 影響因子

與音節時間模型類似，音節基頻軌跡模型考慮了以下 7 個影響因子：語群、語者、重音類型、音節前的停頓類型、音節後的停頓類型、音節的元音種類以及韻律狀態。其中語者因子已經在 2.2.4 小節正規化。

3.2.2 模型架構

本論文仿照[16]中的架構，假設所有的影響因子都會對 $\log F_0$ 所產生的四維正交基底增益或是減少。也就是表示成：

$$\mathbf{y}_{n,g} = \hat{\mathbf{y}}_{n,g} + \lambda_{h_n,g} + \lambda_{pb_n,g} + \lambda_{fb_n,g} + \lambda_{v_n,g} + \lambda_{p_n,g} + \boldsymbol{\mu}_g^y \quad (3-18)$$

其中 $\mathbf{y}_{n,g}$ 是正規化之後的四維基頻軌跡參數， $\hat{\mathbf{y}}_{n,g}$ 是此模型的誤差，可以被描述成一個平均值為 0，變異數是 σ_g^2 的高斯分佈； λ_b 表示影響因子 b 的參數。 h_n 表示重音類型因子，在前一節已經提過； pb_n 和 fb_n 是前後的停頓類型，由 2.2.6 小節所訓練出來的標記模型所標記； v_n 是該音節的元音類型，在音節基頻軌跡模型中使用元音而不使用音節所包含的所有音素是由於基頻軌跡會與元音類型高度相關，而與其他子音沒有太大關係； p_n 表示韻律狀態因子； $\boldsymbol{\mu}_g^y$ 表示語群的四維向量平均值。

值得注意的是停頓類型當中，B1 表示為緊密結合的停頓；B2 表示較不緊密，多為詞邊界的停頓；B3 和 B4 則分別代表較短或較長的標點符號停頓。另外還額外包含了 5 種特殊的停頓類型，分別是：1 種句首類型，4 種句尾類型。4 種句尾類型包括了：(1)直述句 (2)yes/no 疑問句 (3)反問句 (4)wh 疑問句。

3.2.3 訓練流程

音節基頻軌跡模型其訓練流程與音節時間模型大同小異，差別只在於音節基頻軌跡模型每項參數均是四維向量(vector)，而不僅僅是一個純量(scalar)。



第四章 實驗結果與分析

4.1 音節時間模型

訓練過程之相似度變化如圖 4.1，可以發現約在第七、八次左右即達到收斂。

另外 L1 會比 L2 的音節平均相似度來的高。

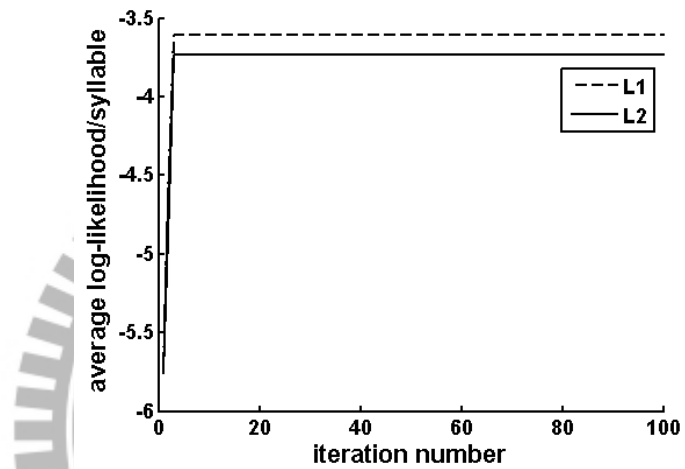


圖 4.1：音節時間模型之訓練過程相似度變化

語群平均值如表 4.1，這與我們的認知相同，母語使用者會唸得較快，以下表格單位均為 ms。

表 4.1：語群平均值

Group	L1	L2
Group mean (ms)	225	277

詞長因子參數如表 4.2，可以看出當詞長越長，每個音節也就會唸得越短。

另外，L2 的變化幅度也比 L1 來的大。

表 4.2：詞長因子參數

Word length	AP for L1	AP for L2
1	15	28
2	-10	-18
3	-29	-48
4	-20	-49
5	-71	-66

重音因子參數如表 4.3，對 L1 而言，扣除單音節詞，其重音造成的影響可排序為主重音>次重音>無重音，這與我們認知的相同；而單音節詞則與次重音類似。以 L2 而言，與 L1 最大的差異在於其次重音發的更像無重音甚至比無重音還要短，這可能來自於 L2 缺乏次重音的觀念。

同時，我們也能發現 L1 的變化幅度要比 L2 的來的大。若只討論其中主重音與無重音的關係，其結果也與中研院基於同樣資料庫的研究[6]相同(該研究只使用語料中的 target word，也就是並沒有本論文中的單音節詞重音，而次重音則歸類在無重音)，L1 將有無重音的差別拉的更開。

另外也與[21]中的結果的趨勢相同(英文母語，重音：無重音，329ms:250ms，拉長 32%；中文母語：351ms:277ms，拉長 27%)，若將本研究的語群加回去，其約為：英文母語，拉長 21%；中文母語，拉長 14%。由於[21]中並沒有考慮其他因子，因此數據有差異但趨勢不變。

表 4.3：重音因子參數

Lexical stress	AP for L1	AP for L2
Primary	24	21
Secondary	-1	-18
None	-20	-15
Mono-syllabic	0	0

音節組成因子中各個音素的數值如表 4.4(a)~(h)，這可與第二章中的統計資料做比較，由於在給予初始值時即來自於統計資料，因此在之後的疊代訓練時，也會被限制在這些值附近，故解出來的參數會與統計資料相當。

其中，顯而易見的是 L2 的參數大部分比 L1 來的更長，這與我們的認知相同。其次，也有幾組容易在長度上被 L2 所混淆的，也能從表中看出。例如，IH(it) 與 IY(eat)、EH(let)與 EY(late)，L1 發的會比 L2 發的長度差異更大((IY-IH)/IH，65%：35%；(EY-EH)/EH，32%：26%)。

表 4.4：各個音素在音節組成因子中的長度(a)~(h)

(a) 單元音(monophthong)

phone	AP for L1	AP for L2
AA	94	119
AE	111	140
AH	80	118
AO	88	118
EH	105	131
IH	79	111
IY	130	150

UH	91	112
UW	108	140
OW	192	209

(b) 雙元音(diphthong)及兒化音(r-colored)

phone	AP for L1	AP for L2
AW	171	206
AY	181	189
EY	139	165
OY	288	325
ER	116	156

(c) 塞音(stops)

phone	AP for L1	AP for L2
P	42	52
T	45	51
K	49	55
B	60	63
D	58	72
G	62	67

(d) 塞擦音(affricates)

phone	AP for L1	AP for L2
CH	126	133
JH	96	89

(e) 擦音(fricatives)

phone	AP for L1	AP for L2
F	107	113

V	73	84
TH	100	122
DH	45	63
S	129	141
Z	121	141
SH	115	137
ZH	70	55
HH	55	95

(f) 鼻音(nasal)

phone	AP for L1	AP for L2
M	92	99
N	72	84
NG	81	93

(g) 流音(liquid)

phone	AP for L1	AP for L2
L	65	82
R	73	81

(h) 半母音(semivowel)

phone	AP for L1	AP for L2
Y	44	67
W	59	76

表 4.5 為韻律狀態因子的數值及個數，由小到大排列，其中可以看見在兩端極值的個數較少，中間則較多。各個韻律狀態與語言參數無關，在以下將更進一步分析這些韻律狀態因子。

表 4.5：韻律狀態因子

AP for L1	Count	AP for L2	Count
-151	206	-178	1225
-112	711	-127	4124
-86	1213	-93	7244
-65	1703	-66	10089
-45	1805	-42	11856
-25	1951	-19	12711
-5	1894	5	12509
16	1774	29	11240
39	1427	53	9401
64	1093	81	7502
95	790	113	5412
133	567	151	3679
176	423	199	2269
229	232	264	1103
303	109	368	386
427	30	620	47

4.2 音節基頻軌跡模型

訓練過程之相似度變化如圖 4.2，與圖 4.1 類似。可以發現約在第四、五次左右即達到收斂，比音節時間模型來的快。另外 L1 會比 L2 的音節平均相似度來的高。

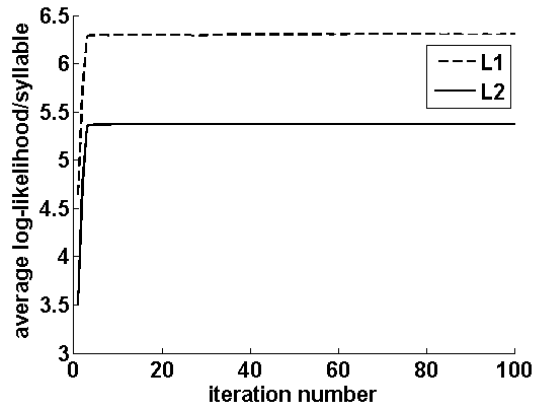


圖 4.2：音節基頻軌跡模型之訓練過程相似度變化

語群平均的基頻軌跡如圖 4.3，可以看出 L1 和 L2 表現出來的軌跡類似。

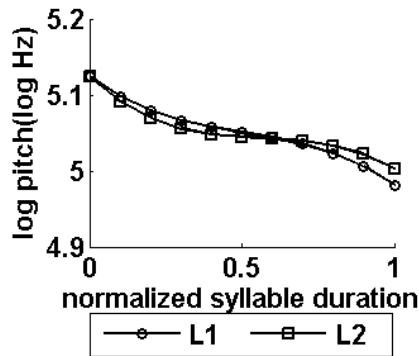


圖 4.3：語群平均基頻軌跡

重音因子的基頻軌跡如圖 4.4。以 L1 而言，最明顯的為主重音，是一個由低往高的明顯爬升曲線；其餘三種類型，皆為一緩慢下降的線條，位準由高往低排序為：單音節重音>無重音>次重音。以 L2 而言，最明顯的也是主重音，也是一由低往高的爬升曲線，但其起點較高，斜率也較低；另外三種重音類型無明顯的差別，這也與在音節時間模型中看到的結果類似，L2 並不能輕易分出次重音和無重音的差異。這也與[6]當中的結果相同，L1 更能表現出有無重音的差別。

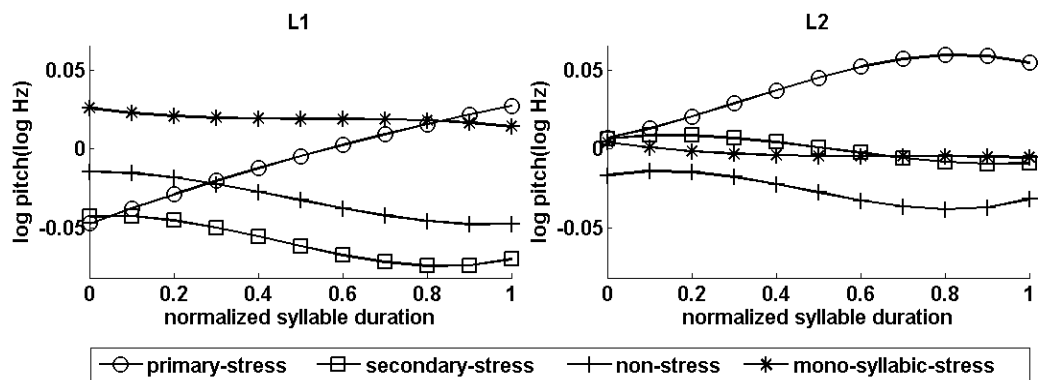


圖 4.4：重音因子基頻軌跡

為了與其他文獻做比較，圖 4.5 將 L1 的參數轉換為 semitone，並將主重音與單音節重音做加權平均。圖 4.6 為 [22] 中，各種重音音節的元音的基頻軌跡，根據其研究的演算法，其所展示的基頻軌跡可與我們的研究作為比較(該基頻軌跡是由線性回歸的殘餘值得出，可視為觀察值扣除語群平均的結果)。雖然演算法並不一樣，但仍能看出 3 種類型的重音的分佈趨勢大致相同。

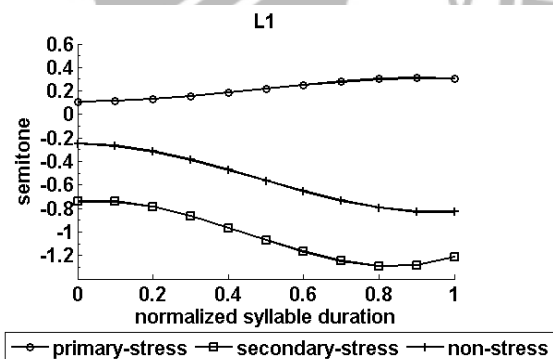


圖 4.5：將重音因子參數轉換為 semitone

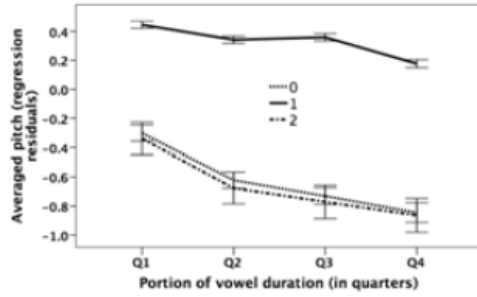


Figure 1. F0 contours of primary-stress ('1'), secondary-stress ('2') and reduced ('0') vowels.

圖 4.6：文獻[22]中各種重音類型的基頻軌跡

圖 4.7 為音節前不同類型停頓的基頻軌跡。先從 L1 討論，可以明顯看出句首是一個高而平坦的斜線；B4 與 B3 多為標點符號的停頓，只是停頓時間長度上的差異，也能看出 B4 的基頻重置(pitch reset)現象會比 B3 來的明顯(變動幅度更大)；另外 B1 和 B2，在這裡並沒有明顯的差異。

對於 L2 而言，其句首現象與 L1 類似，但在結尾會向下掉；除此之外，其 4 種類型停頓並沒有明顯的差異。

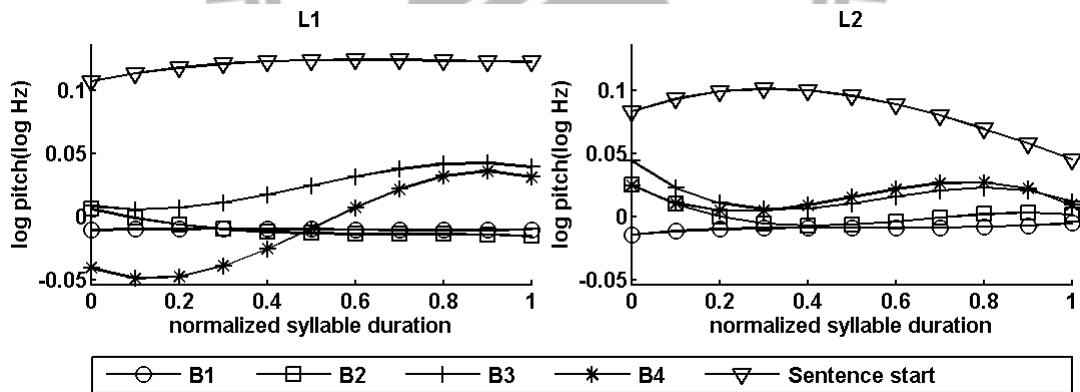


圖 4.7：音節前停頓類型因子基頻軌跡

圖 4.8 為音節後不同停頓類型的基頻軌跡。其中 4 種類型的停頓皆為一平坦直線，由高往低為 B1>B2>B3>B4，這在 L1 與 L2 中皆是如此，但 L2 的較不明顯。另外 4 種的句尾類型，L1 與 L2 的表現則較有差異：先以直述句(SE-2-1)來說，其應為一下降的斜線，這點無論在 L1 和 L2 都能看到，這是一種較為簡單

學習的句型；另一簡單的類型為 yes/no 疑問句(SE-2-1)，其應為一往上的斜線，我們也能發現 L2 表現的與 L1 十分類似；其次則是 wh 疑問句(SE-2-2)，這與 yes/no 疑問句不同，其應該類似於直述句的表現，這點我們在 L1 的圖中可以看見，不過在 L2 中，卻被念得既不像直述句，也不像 yes/no 疑問句，這是來自於有部份語者念對，卻有部份語者唸錯，兩者平均之下就會得到一個皆不像的軌跡；最後一項則是反問句，這點比較是我們少看到的，我們可以看見 L1 與 L2 的表現並不相同，不過必須進一步觀察。

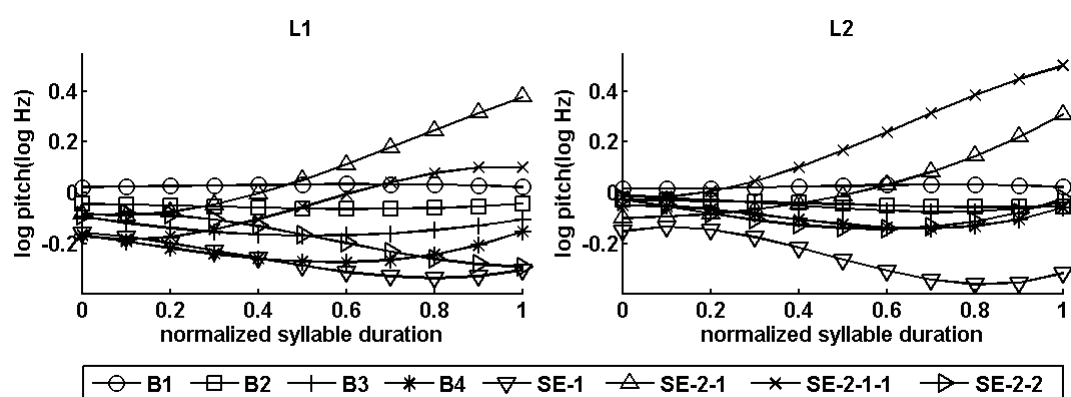


圖 4.8：音節後停頓類型因子基頻軌跡

圖 4.9 是兩種不同類型反問句的基頻軌跡，其為圖 4.8 中的 SE-2-1-1 加上不同類型反問句的平均韻律狀態。其中正面反問句(E211-positive)表示在語意上是肯定，而反問對方是否同意；反面反問句(E211-negative)則是語意上否定，反問對方是否同意。這兩者在字面上完全相同，其意思卻大異其趣(關於詳細語句請見附錄 1 中第 5 段，其中 4B 和 5A 為正面反問句；4A 和 5B 為反面反問句)。其中我們可以看見 L1 能將兩種意義的反問句表現的十分清楚，反面反問句仍帶有疑問，但正面反問句已是肯定；因此反面反問句會類似 yes/no 疑問句的上揚，而正面反問句則會像直述句一樣緩慢下降。反之在 L2 的表現，我們可以清楚的看見，L2 完全不能分辨兩者的差異，幾乎都被視為一般 yes/no 疑問句來處理。

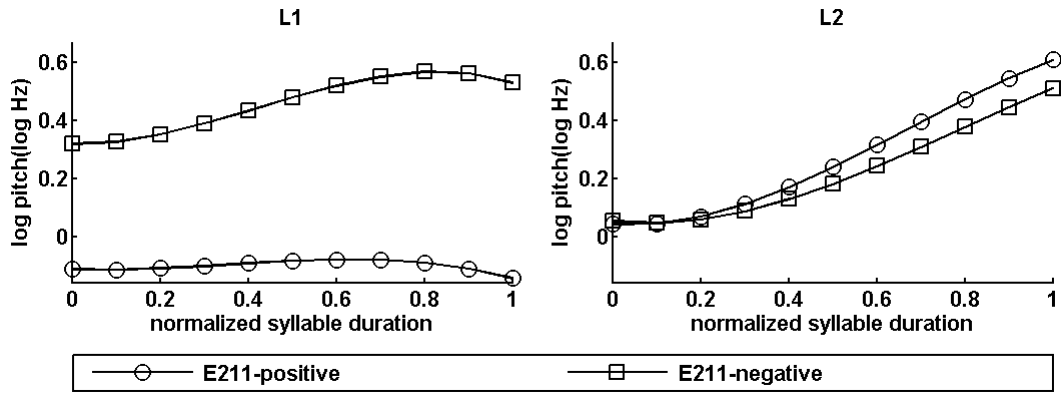


圖 4.9：不同類型反問句的基頻軌跡

圖 4.10 為不同類型單元音的基頻軌跡，其中虛線者為 high vowel，實線者為 low vowel，high/low vowel 的定義是根據 vowel chart(圖 4.11)中的高低而來。由圖 4.10 中可以看出 high vowel 均比 low vowel 來的高，這符合[23]中的觀察。

[23]中提到 UW，IY 兩個 high vowel 以及 AA 這個 low vowel，由高往低(該研究只統計 vowel 的平均基頻)可排序為： $UW \geq IY > AA$ ，這與我們的結果相同。另外該研究也有統計中文中此 3 個元音的高度，其大小順序相同，但差距較小，這也符合我們找到的現象(中文字母者可能會誤將中文中的這 3 個元音高度差異套用到英文上)。

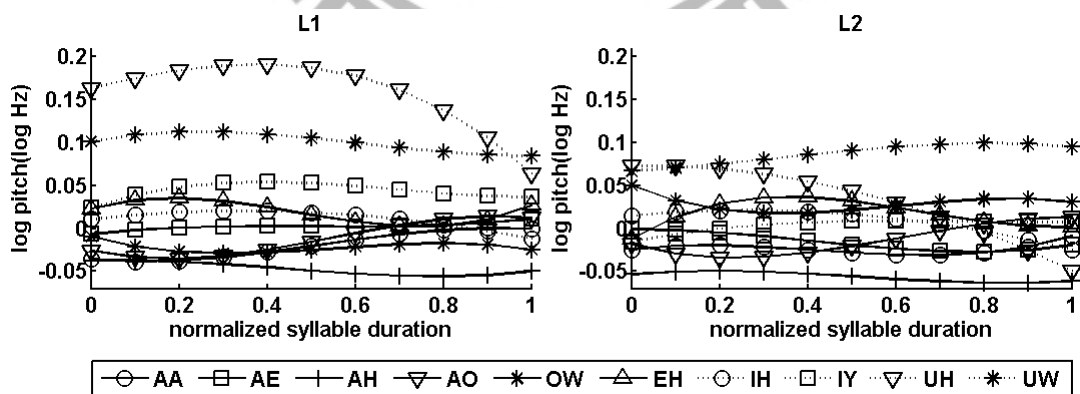


圖 4.10：單元音的基頻軌跡

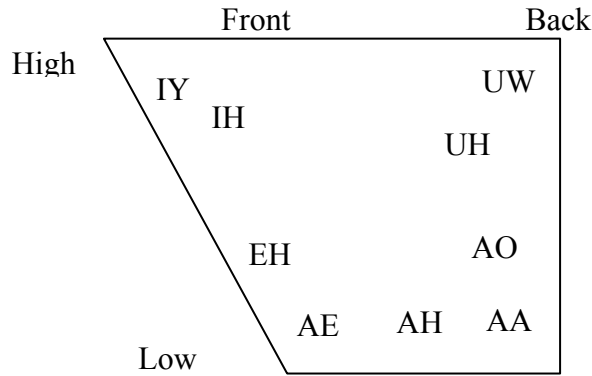


圖 4.11：Vowel chart

圖 4.12 為雙元音及兒化音的基頻軌跡，其中兒化音為 ER，其他皆為雙元音，各個雙元音的組合見表 4.5，其在 vowel chart 上的移動狀況如圖 4.13。所有的雙元音皆為由 low vowel 到 high vowel，因此可以期待看見基頻軌跡應是由低往高的曲線，但在本研究中，此現象並不明顯。另外 OY 的變化幅度與他人有較大差異，這是來自於本研究中含 OY 的音節種類過少(見附錄 2)，因此並不準確。

大部分探討雙元音的變化狀況多半在 F1 F2 上討論，較易看出不同語群、不同地區方言雙元音的移動，本研究只觀察基頻的軌跡，因此並不容易與他人的研究比較。

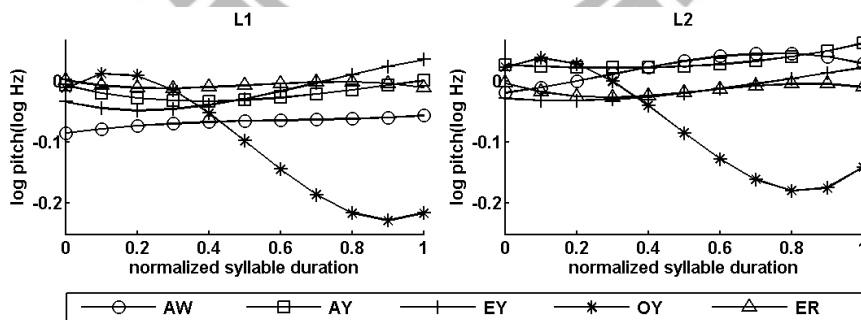


圖 4.12：雙元音及兒化音的基頻軌跡

表 4.6：雙元音的組合

EY(EH-IH)
AY(AA-IH)
AW(AA-UH)
OY(AO-IH)

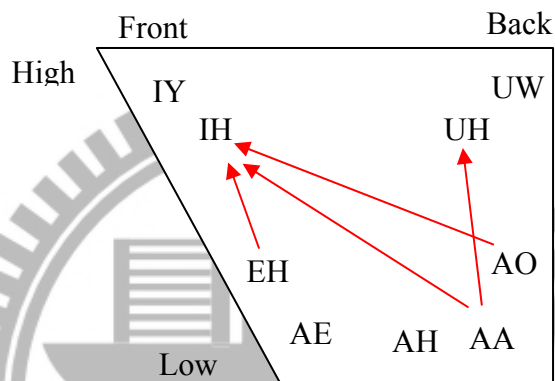


圖 4.13：雙元音在 vowel chart 上的移動圖示

4.3 韻律狀態結果分析

本節將透過各種分析比較，來探討韻律狀態與各項語言參數的關係。表 4.7 為單音節詞中，不同詞類型時，L1 與 L2 的平均時間韻律狀態，其單位為 ms，括號內為其佔 group mean 的比例。可以看出 L1 更能表現出 function word 與 content word 的差別，即意味著在一個句子之中，L1 更能強調出句子的重點何在。圖 4.14 則為在基頻軌跡上的狀況，可以看出 L2 兩者間的差異較小。

表 4.7：function word 與 content word 的時間比較

Word type	AP for L1	AP for L2
Function word	-41(-18%)	-18(-6%)
Content word	12(+5%)	10(+4%)

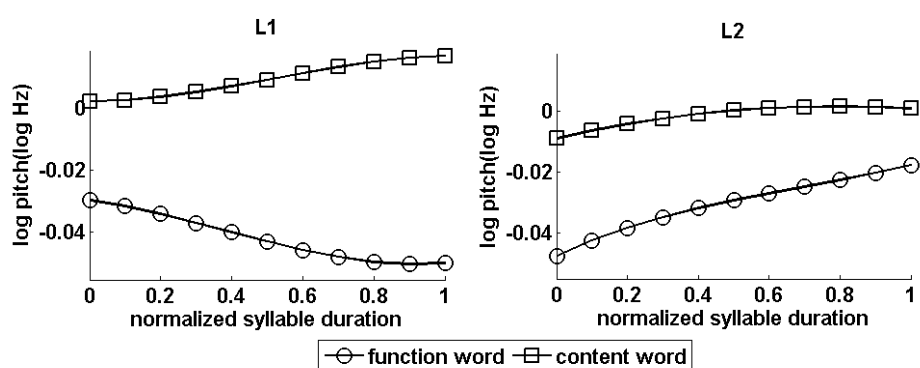


圖 4.14：function word 與 content word 的基頻軌跡比較

表 4.8 為語料中第 1 部份(簡單直述句)中 target word 與左右音節(所有音節皆是 content word)，L1 與 L2 每個音節的平均時間韻律狀態。其中可以看見，意外的是 L2 在 target word 上的表現比 L1 來的更加顯著。圖 4.15 則為在基頻軌跡上的比較，其中清楚的得知 L1 能在基頻軌跡上分出兩者的差異，而 L2 顯然不能。這可結論為中文使用者傾向使用時間差異來表現 prominence，而英文使用者則傾向使用基頻軌跡。

表 4.8：target word 與 target word 邊界的時間比較

Word type	AP for L1	AP for L2
Target word	37(+16%)	31(+11%)
Target word's neighbor	16(+7%)	-10(-4%)

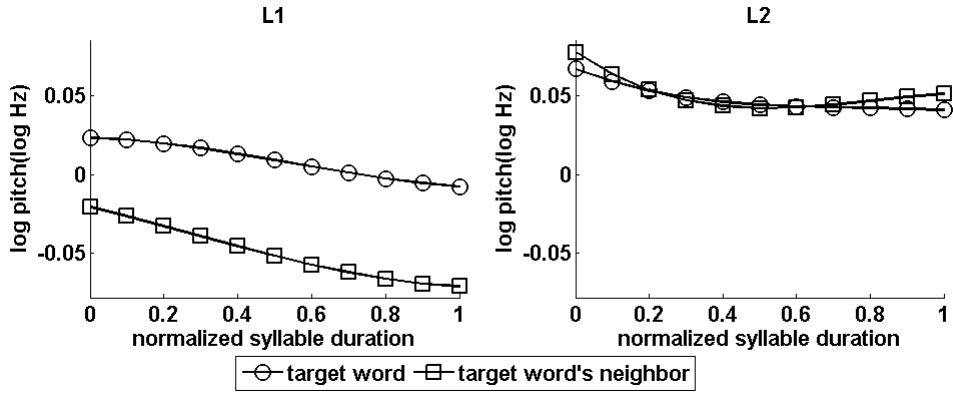


圖 4.15：target word 與 target word 邊界的基頻軌跡比較

表 4.9 為語料中第 4 段，介係詞有無被強調的平均韻律狀態比較，其中值得注意的是，時間模型中並沒有考慮停頓類型，因此這裡的 unreduced function word 會受到一部分邊界停頓的效應。不過即使如此，L1 與 L2 之間的差異仍然相當明顯，L1 更能表現出有無被強調間的差異。圖 4.16 為在基頻軌跡上的表現，其中 L1 將兩者明顯的分開，一高一低，被弱化的介係詞會較低；對於 L2 而言，其在前端沒有明顯區分，至後端拉開，但卻是 weaker 者較高。這可能來自於 L2 在念出被弱化介係詞時，仍然會受到後方的影響而拉高，反之在沒有被弱化的介係詞時，其後方為句子邊界，因此往下掉。

表 4.9：介係詞在對比句強調與非強調位置

Word type	AP for L1	AP for L2
Reduced function word	-75(-33%)	-46(-17%)
Unreduced function word	197(+88%)	156(+56%)

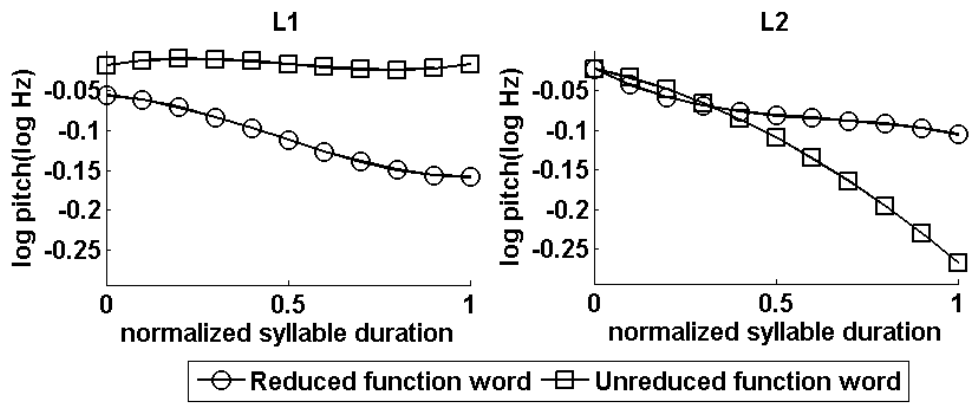


圖 4.16：介係詞在對比句強調與非強調位置



第五章 結論與未來展望

5.1 結論

本論文提出了一個以音節為單位的英文韻律模型，並考慮了許多影響因子造成的影響，同時分析了各項影響因子的參數，顯現此韻律模型能有效的模擬音節時間以及音節基頻軌跡，從中也能看出 L2 確實有許多不像 L1 的所在，有些符合我們的知識，有些則給予了我們更清楚的解釋，帶來更數據化的現象解釋。

5.3 未來展望

本研究尚有許多未盡之功，還能由後人改進。例如，考慮更多影響因子例如 prominence 使得我們可以從更多角度來觀察 L1 與 L2 的差異；其次，對於韻律狀態影響因子還有待更仔細的分析以瞭解並與 ToBI 來做比較；最後，本研究尚缺乏實現在語音辨認抑或是語音合成、語言學習之應用，期許本研究所提出的數據，能給予這些應用更為有效的助益。

參考文獻

- 【1】 Visceglia, Tanya, Tseng, Chiu-yu, Kondo, Mariki, Meng, Helen and Sagisaki, Yoshinori, “*Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project, Oriental COCOSDA)*,” Beijing, China, 2009.
- 【2】 Chiu-yu Tseng, “*Phonotactic and discourse aspects of content design in AESOP (Asian English Speech cOrpus Project)*,” Oriental-COCOSDA, Hsinchu, Taiwan, 2011.
- 【3】 C. Y. Tseng, and T. Visceglia, “*AESOP (Asian English Speech Corpus Project) and TWNAESOP*,” Invited keynote speech, Int. Conf. and Workshop on TEFL & Applied Linguistics, March, 2010.
- 【4】 Visceglia, Tanya, Tseng, Chiu-yu, Su, Zhao-yu and Huang, Chi-Feng, “*Realization of English Narrow Focus by L1 English and L1 Taiwan Mandarin Speakers*”, The 7th International Congress of Phonetic Sciences. Hong Kong, China, 2011.
- 【5】 Tseng, Chiu-yu, Su, Zhao-yu, Huang, Chi-Feng and Visceglia, Tanya, “*An Initial Investigation of L1 and L2 Discourse Speech Planning in English*,” The 7th International Symposium on Chinese Spoken Language Processing (ISCSLP) 55-59. Tainan/Sun Moon Lake, Taiwan, 2010.
- 【6】 Visceglia, Tanya, Tseng, Chiu-yu, Su, Zhao-yu and Huang, Chi-Feng, “*Interaction of Lexical and Sentence Prosody in Taiwan L2 English*,” Interspeech 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology. Tokyo, Japan, 2010.
- 【7】 Silverman, Kim, Mary Beckman, John Pitrelli, Marl Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg, “*TOBI: A*

- standard for labeling English prosody,*” In International Conference on Speech and Language Processing (ICSLP), volume 2, pages 867-870, 1992.
- 【8】 H. Fujisaki and S. Ohno, “*Analysis and modeling of fundamental frequency contours of English utterances,*” EUROSPEECH, pp. 985–988, 1995.
- 【9】 Flege, J. E., Bohn, O-S. and Jang, S., “*Effect of experience on nonnative subjects’ production and perception of English vowels,*” Journal of Phonetics, 25, 437–470, 1997.
- 【10】 Xia Wang, Aijun Li, Jia Sun and Yun Mai, “*Prosodic analysis on English mild imperatives of Chinese EFL learners,*” Speech Prosody, Chicago, USA, 2010.
- 【11】 Zhang, Y. Nissen, S. and Francis, A., “*Acoustic characteristics of English lexical stress produced by native Mandarin speakers,*” J Acoust. Soc. Am. Volume 123, Issue 9, pp.4498-4513, 2008.
- 【12】 Chen, Y., Robb. M.P., Gillbert. H. R., and Lerman, J. W, “*A study of sentence stress production in Mandarin speakers of American English,*” J Acoust. Soc. Am. 4, 1681-1690, 2001.
- 【13】 Jian, H.L. “*On the syllable timing in Taiwan English,*” Speech Prosody, Nara, Japan, 2004.
- 【14】 Jian, H.L. and Chang, V., “*A Preliminary Study of Lexical Stress in Taiwan English Homographs,*” Speech Prosody, Brazil, 2008.
- 【15】 S. H. Chen, W. H. Lai and Y. R. Wang, “*A New Duration Modeling Approach for Mandarin Speech,*” IEEE Trans. on Speech and Audio Processing, vol.11, no.4, pp. 308-320, July 2003.
- 【16】 S. H. Chen, W. H. Lai and Y. R. Wang, “*A statistics-based pitch contour model for Mandarin speech,*” J. Acoust. Soc. Am., vol.117, Issue 2, pp.908-925, 2005.
- 【17】 C. Y. Chiang, S. H. Chen, H. M. Yu and Y. R. Wang, “*Unsupervised joint*

prosody labeling and modeling for Mandarin speech," J. Acoust. Soc. Am., vol.125, Issue 2, pp. 1164-1183, 2009.

- 【18】 <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- 【19】 M. Kemal Sönmez, L. Heck, M. Weintraub and E. Shriberg, “*A Lognormal Tied Mixture Model Of Pitch For Prosody-Based Speaker Recognition,*” Eurospeech, IV-195, Sept 1997.
- 【20】 Tseng, Chiu-yu, Pin, Shao-huang, Lee, Yeh-lin, Wang, Hsin-min and Chen, Yong-cheng, “*Fluent speech prosody: framework and modeling,*” Speech Communication, Special Issue on Quantitative Prosody Modeling for Natural Speech Description and Generation, Vol. 46, issue 3-4, pp284-309, 2005.
- 【21】 Yanhong Zhang, Shawn L. Nissen, and Alexander L. Francis, “*Acoustic characteristics of English lexical stress produced by native Mandarin speakers,*” J. Acoust. Soc. Am., 2008, 4498.
- 【22】 J. Yuan, S. Isard, and M. Liberman, “*Different Roles of Pitch and Duration in Distinguishing Word Stress in English,*” Interspeech, 2008.
- 【23】 D. H. Whalen and Andrea G. Levitt, “*The universality of intrinsic F0 of vowels,*” J. Phonetics. (1995) 23, 349-366

附錄 1：TWNAESOP 文本資料

(1)：目標字X嵌於無意義簡單句句重音位置 (Target words embedded in carrier sentences)

請以自然的速度與適當的音量，讀一遍下列句子，請勿強調或加重語句中的任何特定單字或片語。

Read each sentence once at a natural speaking rate and volume. Try not to stress or emphasize any particular word or phrase.

1. I said apartment five times.

2. I said overnight five times.

3. I said misunderstand five times.

4. I said supermarket five times.

5. I said money ten times.

6. I said hospital ten times.

7. I said white wine ten times.

8. I said elevator ten times.

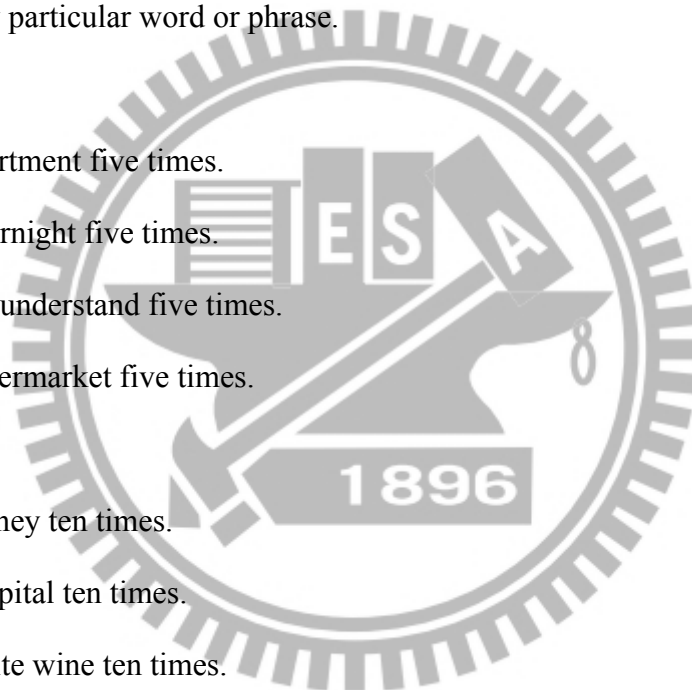
9. I said available ten times.

10. I said information ten times.

11. I said January ten times.

12. I said experience ten times.

13. I said California ten times.



14. I said Vietnamese ten times.
15. I say department store ten times.
16. I said morning ten times.

17. I said video ten times.
18. I say tomorrow ten times.
19. I said Japanese ten times.
20. I said afternoon ten times.

(2): 同(1)目標字嵌於簡單句韻律邊界位置 (Target words at prosodic boundaries)
請以自然的速度與適當的音量，讀一遍下列句子，請勿強調或加重語句中的任何
特定單字或片語。

Read each sentence once at a natural speaking rate and volume. Try not to stress or
emphasize any particular word or phrase.

1. Do you need any money?
2. Did he go to the hospital?
3. Has Jane found an apartment?
4. Can packages be shipped overnight?
5. Would you like a glass of white wine?

6. Where is the elevator?
7. When will Bill be available?
8. Who can give me the information?
9. Why do you always misunderstand?

10. Where is the nearest supermarket?
11. In December and January, the sun rises at seven in the morning.
12. Although Fred didn't have any experience, he had no trouble learning how to make a video.
13. When Sue left this evening for California, she said she would call me tomorrow.
14. If you want to learn Vietnamese, I think it will be easier than Japanese.
15. If you want to check out the new department store, we can go this afternoon

(3)：同(1)目標字嵌入強調句語句對比重音位置 (Target words in narrow focus)
請以自然的速度與適當的音量，讀一遍下列句子，並依上下文強調或加重語句中
特定單字或片語。

Read each sentence once at a natural speaking rate and volume. Emphasize the word
or phrase that seems appropriate for each context.

1. Context: Did Bill lose everything in the robbery?

No. His MONEY was taken, but they didn't take his computer.

2. Context: Can doctors give blood tests at this clinic?

No. You should go to a HOSPITAL for blood tests.

3. Context: Can we open a branch of our office in this building?

No. This is an APARTMENT building, not a commercial building.

4. Context: Will 3-day delivery be fast enough?

No. We need OVERNIGHT delivery.

5. Context: Did you order a Coke?

No. I ordered WHITE WINE, not Coke.

6. Context: How will I carry all these boxes up to the fifth floor?

You should take the ELEVATOR instead of the stairs.

7. Context: Would you like a table by the window?

Someone is already sitting there. Are there any AVAILABLE tables by the window?

8. Context: Why couldn't anyone help me at the service desk?

You should have gone to the INFORMATION desk.

9. Context: Did you misunderstand the question?

I didn't MISUNDERSTAND the question; I just chose not to answer it.

10. Context: Do you buy fruit at the farmer's market?

No. I usually buy fruit at the SUPERMARKET because they stay open later.

11. Context: Have you been trained to do this job?

No. But I think EXPERIENCE is more important than training.

12. Context: Do people speak Chinese in Vietnam?

No. They speak VIETNAMESE in Vietnam.

13. Context: Is Teresa still living in Texas?

No. Teresa lives in CALIFORNIA now.

14. Context: Is Lunar New Year in February?

No. It's in JANUARY this year.

15. Context: Does Mary's flight arrive in the evening?

No. Mary is taking a MORNING flight.

16. Context: Are we allowed to make audio and video recordings?

No. VIDEO recordings are not allowed.

17. Context: Is Anna leaving tonight for Chicago?

No. Anna is leaving TOMORROW for Chicago.

18. Context: Do you like Japanese and Korean food?

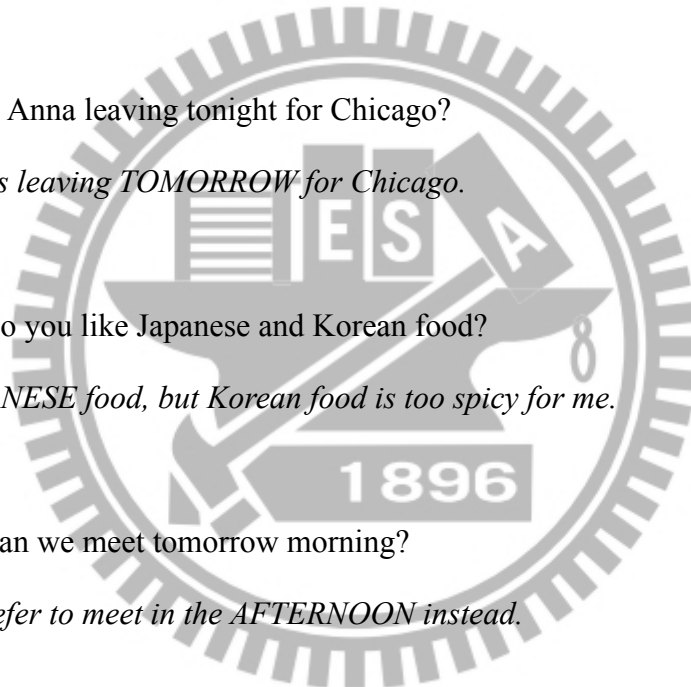
I like JAPANESE food, but Korean food is too spicy for me.

19. Context: Can we meet tomorrow morning?

I would prefer to meet in the AFTERNOON instead.

20. Context: Did you say you work at a bookstore?

No, I work at a DEPARTMENT store.

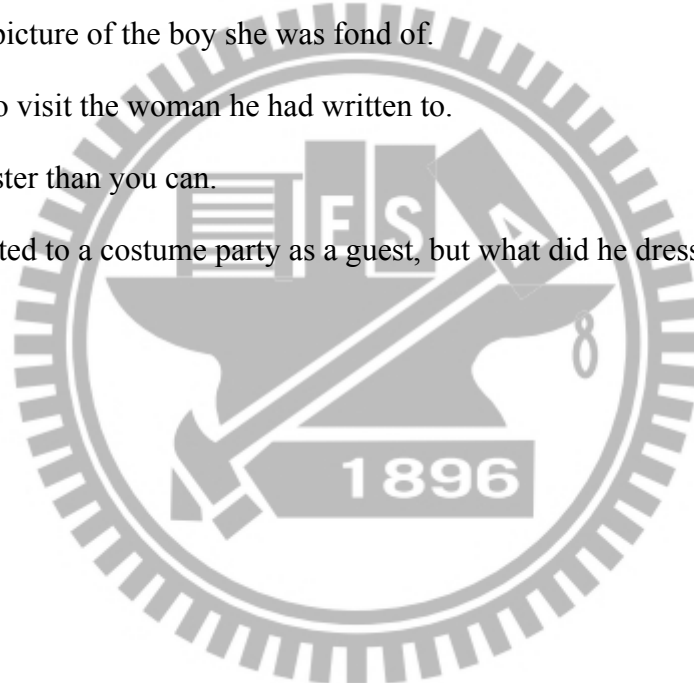


(4)：介係詞嵌入強調句語句對比重音位置 (Production of reduced and unreduced function words)

請以自然的速度與適當的音量，讀一遍下列句子，並依上下文強調或加重語句中特定單字或片語。

Read each sentence once at a natural speaking rate and volume. Emphasize the word or phrase that seems appropriate for each context.

1. If the birthday party wasn't for Mary, then who was it for?
2. Jane saw a picture of the boy she was fond of.
3. John went to visit the woman he had written to.
4. I can run faster than you can.
5. He was invited to a costume party as a guest, but what did he dress as?



(5)：字面相同但有一個意義以上的短句（Prosodic disambiguation）

下面有五組語句，每組2句，同組的2句或是字面完全相同，或是十分相似，但依據提示的情況（context），每句的意義不同，說法也不同。請依提示的情況，將每組意義不同的2句分別讀出，讀完後按Stop鍵，再按Next鍵進入下一組。

You will see 5 sets of two similar sentences appearing in two different contexts. Try to read the two versions of each sentence in such a way that you make the difference between them clear to a listener.

Context 1A: Alice is going to divorce Tom.

When Alice leaves, Tom will be upset.

Context 1B: We think Alice and Tom should stay together.

When Alice leaves Tom, we'll be upset.

Context 2A: Is anyone available to baby-sit today?

I'll look after the children until lunchtime.

Context 2B: Can you help me find the ring I lost at the kindergarten this morning?

I'll look after the children have left.

Context 3A: Fred and John are arguing. They both want Mary to be on their team.

The fight is over Mary.

Context 3B: Mary doesn't know why everyone else has already left the boxing arena.

The fight is over, Mary.

Context 4A: I'm not sure if I should let Peter into my English class.

He's a good boy, isn't he?

Context 4B: James always helps the younger children with their homework.

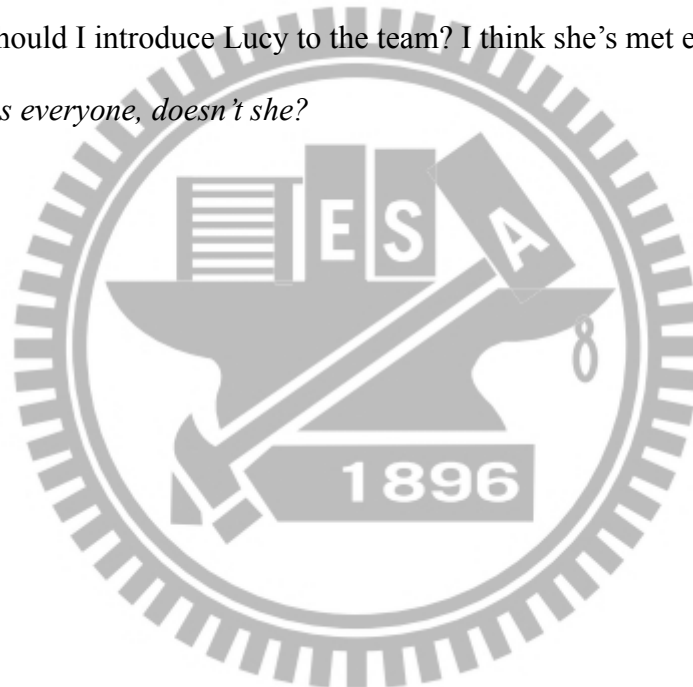
He's a good boy, isn't he?

Context 5A: Wherever Sarah goes, everyone stops and talks to her.

Say: She knows everyone, doesn't she?

Context 5B: Should I introduce Lucy to the team? I think she's met everyone before.

Say: She knows everyone, doesn't she?



(6)：朗讀短篇「北風與太陽」 (The North Wind and the Sun)

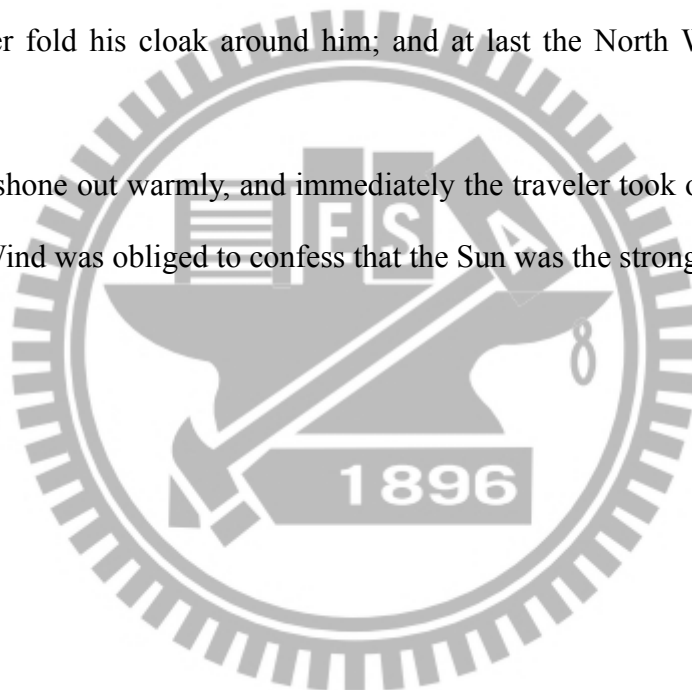
請以自然的速度與適當的音量，讀出北風與太陽的故事。

Read the following passage aloud at a natural speaking rate and volume.

The North Wind and the Sun were disputing which was the stronger when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.

Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt.

Then the Sun shone out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.



(7)：誘發性與電腦對話(Computer-prompted dialogue)

【與電腦對話】錄音說明：請你模擬長榮航空EVA公司的票務人員，接聽一位客戶訂票的電話。這位客戶打算從台北前往紐約，請你依據她的問題及需求，協助她完成訂票作業。對話時，你會從耳機裡聽到這位客戶的語音，同時螢幕也會以文字呈現她說話的內容。你回答的文字提示，也會呈現於螢幕上，協助你完成錄音工作。

In this task, you will play the role of a reservation agent at EVA airlines, helping a customer to make a reservation over the phone. She wants to book a ticket from Taipei to New York City. Please respond to her questions and requests until you have completed the reservation process. During the phone call, you will hear the customer's voice through your headset; at the same time, everything the customer says will appear on the screen. Written prompts indicating how you should respond to each question or request will also appear on the screen.

Introduction: You are a reservation agent for EVA Airlines. Help this customer reserve a flight from Taipei to New York.

Customer: Good morning. I'd like to reserve a ticket from Taipei to JFK airport in New York.

Prompt: Ask the customer: "When would you like to travel?"

When would you like to travel?

Customer: November twenty-second.

Prompt: Ask the customer: "Did you say the twenty-second or the twenty-seventh?"

Did you say the twenty-second, or the twenty-seventh?

Customer: That would be Tuesday, the twenty-second.

Prompt: Ask the customer: "Would you like a window seat or an aisle seat?"

Would you like a window seat or an aisle seat?

Customer: An aisle seat, please.

Prompt: Ask the customer: "Would you like a special dinner?"

Would you like a special dinner?

Customer: No, thank you. I don't think so.

Prompt: Ask the customer: "When would you like to reserve your returning flight?"

When would you like to reserve your returning flight?

Customer: I'm not sure when I'll be returning. So I'll call from New York to reserve the date.

Prompt: Tell the customer:

1. Your flight, BR 317, will depart from CKS airport at 11:15 AM on November 22nd.
2. You will arrive at Narita Airport at 2:50 PM.
3. You will transfer to Flight 809 to New York JFK Airport, which departs at 7:08 PM from Gate 13F.
4. You will land at JFK airport at 4:30 PM on November 22nd.

Your flight, BR 317 will depart from CKS airport on November 22nd at 11:15 AM. It will arrive at Narita Airport at 2:50 PM, where you will transfer to Flight 201 to JFK, which departs at 7:08 pm from Gate 13F. You will land at JFK airport at 4:30 PM on November 22nd.

Customer: Um... wait! Did you say the flight from Narita to New York leaves from

Gate 30 F?

Prompt: Tell the customer: “The flight leaves from Gate 13 F, not Gate 30 F.

The flight leaves from Gate 13 F, not Gate 30 F.

Customer: Oh, sorry. Got it. Gate 13 F.

Prompt: Ask the customer “May I have your name?”

May I have your name?

Customer: My name is Lucy Hasegawa-Johnson. That’s L-U-C-Y H-A-S-E-G-A-W-A
J-O-H-N-S-O-N.

Prompt: Repeat the spelling of the customer’s name L-U-C-Y H-A-S-E-G-A-W-A
J-O-H-N-S-O-N

L-U-C-Y H-A-S-E-G-A-W-A J-O-H-N-S-O-N

Customer: That’s right.

Prompt: Ask the customer “May I have your credit card number?”

May I have your credit card number?

Customer: It’s VISA number 5924-8013-6702-3516. Expiration date 09/ 2012

Prompt: Repeat the customer’s credit card number: 5924-8013-6702-3516. Expiration
date 09/ 2012

5924-8013-6702-3516. Expiration date 09/ 2012.

Customer: That’s correct.

Prompt: Ask the customer “May I have your billing address?”

May I have your billing address?

Customer: It's 1425 Lakeshore Drive, Apartment 47B, Chicago, Illinois 60195.

Prompt: Repeat the customer's address: 1425 Lakeshore Drive, Apartment 47B, Chicago, Illinois 60195

1425 Lakeshore Drive, Apartment 47B, Chicago, Illinois 60195

Customer: Yes.

Prompt: Ask the customer "May I have your contact phone number?"

May I have your contact phone number?

Customer: It's 609-472-1358.

Prompt: Repeat the customer's phone number: 609-472-1358

609-472-1358

Customer: Yes, that's right.

Prompt: Ask the customer: "Is there anything else I can help you with this morning?"

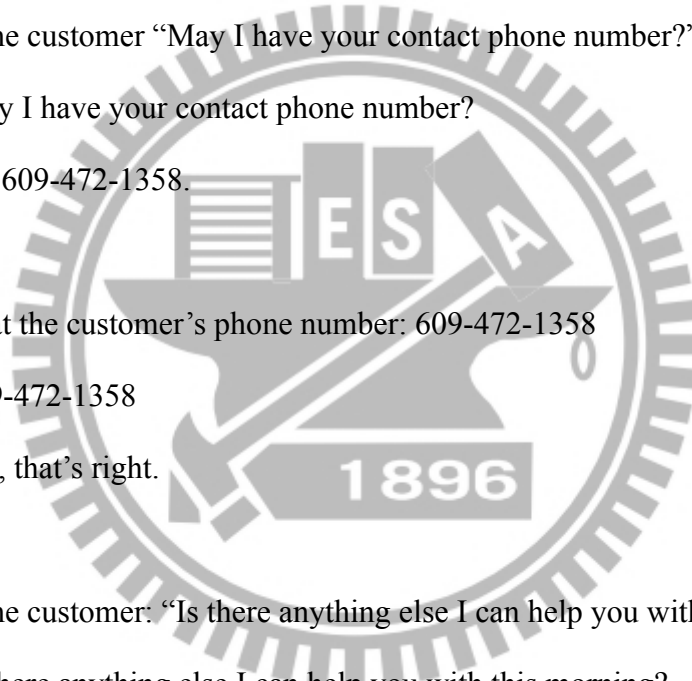
Is there anything else I can help you with this morning?

Customer: No, thank you. That's all I need today.

Prompt: Say "Goodbye and thank you for calling EVA Airlines."

Goodbye, and thank you for calling EVA Airlines.

Customer: Goodbye.



附錄 2：音節類型統計資料

Vowel	AA	AE	AH	AO	EH	IH	IY	OW	UH	UW	AW	AY	EY
Count	20	27	61	21	45	56	33	15	8	15	7	21	24
V	0	1	1	0	1	1	1	1	0	0	0	1	1
VC	2	4	4	3	8	6	1	0	0	0	1	3	2
CV	4	4	14	2	5	11	16	6	1	6	3	3	9
VCC	0	0	3	0	1	1	0	0	0	0	1	0	0
CVC	8	8	23	10	14	29	9	5	7	3	1	8	10
CCV	0	1	2	0	2	2	3	1	0	4	0	1	2
CVCC	5	6	11	4	5	6	1	1	0	0	1	2	0
CCVC	0	2	3	1	5	0	2	1	0	1	0	2	0
CCCV	0	0	0	0	0	0	0	0	0	1	0	0	0
CVCCC	1	0	0	0	2	0	0	0	0	0	0	0	0
CCVCC	0	1	0	0	2	0	0	0	0	0	0	1	0
CCCVCC	0	0	0	1	0	0	0	0	0	0	0	0	0
OY	ER	Total											
3	25	381											
0	1	9											
0	0	34											
2	15	101											
0	0	6											
1	5	141											
0	3	21											
0	1	43											

0	0	17
0	0	1
0	0	3
0	0	4
0	0	1

AA~UW：單元音

AW~OY：雙元音

ER：兒化音

AA:

class	Count	Syllable
V	0	
VC	2	AA-N, AA-R,
CV	4	HH-AA, K-AA, M-AA, N-AA,
VCC	0	
CVC	8	JH-AA-N, K-AA-N, M-AA-R, N-AA-M, N-AA-T, P-AA-R, T-AA-M, W-AA-Z,
CCV	0	
CVCC	5	F-AA-N-D, HH-AA-R-D, K-AA-R-D, P-AA-R-T, W-AA-N-T,
CCVC	0	
CCCV	0	
CVCCC	1	P-AA-R-T-S,
CCCVC	0	
CCVCC	0	

total	20	
-------	----	--

AE:

class	Count	Syllable
V	1	AE,
VC	4	AE-F, AE-N, AE-T, AE-Z,
CV	4	F-AE, JH-AE, K-AE, P-AE,
VCC	0	
CVC	8	B-AE-K, DH-AE-N, DH-AE-T, F-AE-N, HH-AE-D, HH-AE-V, HH-AE-Z, K-AE-N,
CCV	1	T-R-AE,
CVCC	6	DH-AE-T-S, L-AE-N-D, L-AE-S-T, R-AE-P-T, T-AE-K-T, TH-AE-NG-K,
CCVC	2	G-L-AE-S, T-R-AE-N,
CCCV	0	
CVCCC	0	
CCVCC	1	S-T-AE-N-D,
CCCVC	0	
total	27	

AH:

class	Count	Syllable
V	1	AH,
VC	4	AH-N, AH-P, AH-T, AH-V,
CV	14	B-AH, D-AH, DH-AH, K-AH, L-AH, M-AH, N-AH, P-AH,

		S-AH, T-AH, V-AH, W-AH, Z-AH, ZH-AH,
VCC	3	AH-N-D, AH-N-S, AH-N-T,
CVC	23	B-AH-L, B-AH-T, CH-AH-N, D-AH-Z, K-AH-M, K-AH-N, L-AH-S, M-AH-N, N-AH-M, P-AH-N, R-AH-N, S-AH-K, S-AH-M, S-AH-N, SH-AH-L, SH-AH-N, T-AH-L, T-AH-N, V-AH-L, V-AH-N, (V-L), W-AH-N, W-AH-T,
CCV	2	N-Y-AH, T-R-AH,
CVCC	11	B-AH-L-Z, D-AH-N-T, JH-AH-S-T, K-AH-N-D, L-AH-N-CH, M-AH-N-T, R-AH-S-T, T-AH-N-T, V-AH-N-TH, Z-AH-N-D, Z-AH-N-T,
CCVC	3	B-L-AH-D, D-R-AH-N, F-R-AH-M,
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	61	

AO:

class	Count	Syllable
V	0	
VC	3	AO-F, AO-L, AO-R,
CV	2	K-AO, S-AO,
VCC	0	
CVC	10	F-AO-R, G-AO-N, K-AO-L, K-AO-R, K-AO-Z, L-AO-NG, P-AO-R, M-AO-R, SH-AO-R, Y-AO-R,

CCV	0	
CVCC	4	N-AO-R-TH, P-AO-R-T, W-AO-R-M, Y-AO-R-K,
CCVC	1	S-T-AO-R,
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	1	S-T-R-AO-NG,
total	21	

AW:

class	Count	Syllable
V	0	
VC	1	AW-T,
CV	3	HH-AW, N-AW, TH-AW,
VCC	1	AW-N-D,
CVC	1	L-AW-D,
CCV	0	
CVCC	1	F-AW-N-D,
CCVC	0	
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	7	

AY:

class	Count	Syllable
V	1	AY,
VC	3	AY-D, AY-L, AY-V,
CV	3	B-AY, R-AY, W-AY,
VCC	0	
CVC	8	F-AY-T, F-AY-V, L-AY-K, N-AY-N, N-AY-T, T-AY-M, W-AY-N, W-AY-T,
CCV	1	S-P-AY,
CVCC	2	L-AY-N-Z, T-AY-M-Z,
CCVC	2	D-R-AY-V, F-L-AY-T,
CCCV	0	
CVCCC	0	
CCVCC	1	B-L-AY-JH-D,
CCCVC	0	
total	21	

EH:

class	Count	Syllable
V	1	EH,
VC	8	EH-F, EH-N, EH-K, EH-T, EH-M, EH-L, EH-R, EH-S,
CV	5	L-EH, M-EH, N-EH, R-EH, S-EH,
VCC	1	EH-L-S,
CVC	14	CH-EH-K, DH-EH-N, DH-EH-R, F-EH-S, L-EH-T, S-EH-D, S-EH-M, S-EH-P, S-EH-T, T-EH-M, T-EH-N, V-EH-M,

		W-EH-N, W-EH-R,
CCV	2	K-R-EH, S-P-EH,
CVCC	5	D-EH-S-K, G-EH-S-T, HH-EH-L-P, L-EH-F-T, W-EH-N-T,
CCVC	5	D-R-EH-S, F-R-EH-D, K-W-EH-S, S-T-EH-D, T-W-EH-N,
CCCV	0	
CVCCC	2	T-EH-M-P-T, T-EH-S-T-S,
CCVCC	2	S-T-EH-R-Z, T-W-EH-L-V,
CCCVC	0	
total	45	

ER:

class	Count	Syllable
V	1	ER,
VC	0	
CV	15	B-ER, CH-ER, D-ER, DH-ER, F-ER, G-ER, L-ER, M-ER, T-ER, TH-ER, V-ER, N-ER, P-ER, S-ER, W-ER,
VCC	0	
CVC	5	B-ER-TH, D-ER-D, L-ER-N, W-ER-K, Z-ER-V,
CCV	3	S-F-ER, S-P-ER, S-T-ER,
CVCC	1	F-ER-S-T,
CCVC	0	
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	

total	25	
-------	----	--

EY:

class	Count	Syllable
V	1	EY,
VC	2	EY-CH, EY-T,
CV	9	D-EY, DH-EY, JH-EY, K-EY, L-EY, M-EY, S-EY, T-EY, V-EY,
VCC	0	
CVC	10	D-EY-T, G-EY-T, G-EY-V, JH-EY-N, K-EY-M, L-EY-K, M-EY-K, N-EY-M, T-EY-K, W-EY-Z,
CCV	2	S-T-EY, T-R-EY,
CVCC	0	
CCVC	0	
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	24	

IH:

class	Count	Syllable
V	1	IH,
VC	6	IH-M, IH-F, IH-N, IH-Z, IH-T, IH-K,
CV	11	B-IH, D-IH, K-IH, L-IH, M-IH, N-IH, R-IH, S-IH, SH-IH,

		V-IH, Z-IH,
VCC	1	IH-T-S,
CVC	29	B-IH-L, CH-IH-L, D-IH-D, D-IH-T, D-IH-NG, DH-IH-S, F-IH-F, G-IH-V, HH-IH-M, HH-IH-Z, JH-IH-Z, K-IH-NG, K-IH-T, L-IH-NG, L-IH-S, N-IH-NG, N-IH-R, P-IH-K, T-IH-L, T-IH-NG, TH-IH-NG, V-IH-NG, W-IH-CH, W-IH-DH, W-IH-L, W-IH-N, Y-IH-R, Z-IH-Z, Z-IH-T,
CCV	2	P-R-IH, S-P-IH,
CVCC	6	D-IH-NG-Z, L-IH-V-Z, S-IH-K-S, SH-IH-P-T, TH-IH-NG-K, W-IH-N-D,
CCVC	0	
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	56	

IY:

class	Count	Syllable
V	1	IY,
VC	1	IY-V,
CV	16	B-IY, D-IY, DH-IY, HH-IY, JH-IY, L-IY, M-IY, N-IY, P-IY, R-IY, S-IY, SH-IY, T-IY, V-IY, W-IY, Z-IY,
VCC	0	
CVC	9	HH-IY-Z, M-IY-S, M-IY-T, N-IY-D, N-IY-Z, P-IY-T, S-IY-T,

		T-IY-N, W-IY-L,
CCV	3	S-L-IY, TH-R-IY, V-R-IY,
CVCC	1	L-IY-V-Z,
CCVC	2	G-R-IY-D, S-P-IY-K,
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	33	

OY:

class	Count	Syllable
V	0	
VC	0	
CV	2	B-OY, N-OY,
VCC	0	
CVC	1	N-OY-Z,
CCV	0	
CVCC	0	
CCVC	0	
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	3	

OW:

class	Count	Syllable
V	1	OW,
VC	0	
CV	6	D-OW, DH-OW, G-OW, N-OW, R-OW, S-OW,
VCC	0	
CVC	5	CH-OW-Z, F-OW-N, K-OW-K, N-OW-Z, SH-OW-N,
CCV	1	K-L-OW,
CVCC	1	F-OW-L-D,
CCVC	1	K-L-OW-K,
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	15	

UH:

class	Count	Syllable
V	0	
VC	0	
CV	1	W-UH,
VCC	0	
CVC	7	G-UH-D, K-UH-D, L-UH-K, SH-UH-D, T-UH-K, W-UH-D, Y-UH-R,
CCV	0	

CVCC	0	
CCVC	0	
CCCV	0	
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	8	

UW:

class	Count	Syllable
V	0	
VC	0	
CV	6	D-UW, HH-UW, N-UW, S-UW, T-UW, Y-UW,
VCC	0	
CVC	3	F-UW-D, N-UW-N, Y-UW-L,
CCV	4	B-L-UW, L-Y-UW, N-Y-UW, P-Y-UW,
CVCC	0	
CCVC	1	F-R-UW-T,
CCCV	1	S-P-Y-UW,
CVCCC	0	
CCVCC	0	
CCCVC	0	
total	15	