

中文單詞之韻律模式研究

A Study on Prosodic Modeling for Isolated Mandarin Words

陳啓風 Chi-Feng Chen
國立交通大學電信工程學系
Department of Communication Engineering
National Chiao Tung University
linuxe.cm94g@nctu.edu.tw

江振宇 Chen-Yu Chiang
國立交通大學電信工程學系
Department of Communication Engineering
National Chiao Tung University
gene.cm91g@nctu.edu.tw

王逸如 Yih-Ru Wang
國立交通大學電信工程學系
Department of Communication Engineering
National Chiao Tung University
yrwang@cc.nctu.edu.tw

陳信宏 Sin-Horng Chen
國立交通大學電信工程學系
Department of Communication Engineering
National Chiao Tung University
schen@mail.nctu.edu.tw

摘要

在本文中，我們對中文單詞提出了以音節為基本單位的基頻軌跡及音節長度的韻律模型。在基頻軌跡模型中，我們考慮了聲調、音節在詞的位置以及前後音節連音的三種影響因素，並假設這些影響因素彼此獨立相加而組成音節基頻軌跡。在音節長度模型中，我們考慮了聲調、音節在詞的位置、基本音節以及前後音節連音的四種影響因素，我們同樣假設這些影響因素為彼此獨立且具加成性。我們使用一個含 107,936 個詞的單一女性語者的語料庫來評估所提方法是否有效，我們並用決策樹來分析音節長度如何受音節音素結構的影響，也用決策樹來分析音節間 pause 的長度和前後音節音素結構的關係，實驗結果顯示訓練後此兩韻律模型的影響因素都符合我們對中文韻律的知識。

Abstract

In this paper, syllable-based prosody modelings of pitch contour and syllable duration for

isolated Mandarin words are proposed. In the syllable pitch contour model, three main affecting factors of tone, syllable position in word, and inter-syllable coarticulation are considered. These three affecting factors are assumed to be independent and additive. Similarly, in the syllable duration model, four affecting factors of tone, syllable position in word, base-syllable, and inter-syllable coarticulation are considered. We also assume that these affecting factors are independent and additive. A large single female-speaker speech database containing 107,936 words was used to evaluate the performance of the proposed methods. After well-training, the decision tree method was used to analyze the 411 affecting factors of base-syllable and to explore the relationship between inter-syllable pause duration and the nearby linguistic features. Experimental results showed that all these affecting factors conformed to our knowledge about Mandarin prosody.

關鍵詞：韻律模式，基頻軌跡，影響因素，連音

Keywords: Prosody modeling, Pitch contour, Affecting factor, Coarticulation

一、緒論

文句轉語音系統要能合成出自然流利的語音，關鍵在於韻律的變化是否自然順暢。韻律的變化包括音調的高低起伏、音量的強弱、發音的長短及停頓的時機、長度等。目前韻律的合成方法大致分為規則法 [1,2]、類神經網路 [3,4]和統計法。規則法是以語言學的方法，歸納出一些發音的規則，利用這些規則來產生合成語音的韻律。但是人類說話的方式變化複雜，不容易掌握。類神經網路是利用一組複雜的網路來模擬人腦的記憶與學習功能，其學習方法是採用漸進式的修正錯誤與更新記憶的方式，需經由長時間的學習訓練，雖有不錯的效果，但無法分析影響韻律的因素。本文以統計法的方式，可從大量的語音資料中統計出韻律變化，利用所考慮影響韻律的因素加總後，控制韻律變化，並分析各個影響因素對韻律訊息的影響程度。

本文著重於以中文單詞語料庫為基礎之韻律模式的研究，探討音節的基頻軌跡及長度模式，考慮幾個主要的影響因數，希望藉此了解中文單詞的音節基頻軌跡及長度如何變化，以作為未來中文語音合成系統產生韻律信息之用，期望合成出自然流暢的中文單詞聲音。本論文在接下來的第二部份會介紹我們所提出的韻律模型，第三部份介紹模型的訓練方法，實驗結果在第四部份討論，最後於第五部份對於本研究給予一個結論。

二、韻律模式

韻律模式以音節為單位，在給予特定的語言資訊後，使之預測音節的基頻軌跡及長度，做為韻律訊息，並分析音節基頻軌跡及長度在各個因素的影響程度。考慮主要影響因素分別為：聲調(tone)、音節在詞的位置(word-position)、基本音節(base syllable)、音節間的連音狀態(inter-syllable coarticulation state)。

(一)、基頻軌跡之韻律模型

假設所有影響因素可用累加的方式來表示音節的基頻軌跡，如式子(1)：

$$\mathbf{sp}_n = \mathbf{sp}_n^r + \beta_{t_n} + \beta_{w_n} + \beta_{c_{n-1}, p_{n-1}}^f + \beta_{c_n, p_n}^b + \mu^p \quad (1)$$

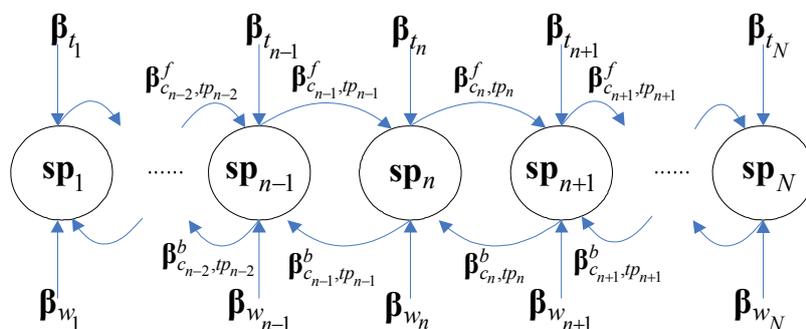
其中 \mathbf{sp}_n 、 \mathbf{sp}_n^r 、 β_{t_n} 、 β_{w_n} 分別為 pitch 模型中第 n 個音節的基頻軌跡參數向量、基頻軌跡參數向量殘餘值(residual)、聲調及詞中位置影響因素； \mathbf{sp}_n 為由一段音節基頻軌跡轉化為四個正交參數表示的向量，轉換方法參見 [5]； $w_n \in \{(2,1), (2,2), \dots, (j,k), \dots, (8,8)\}$ 代表音節在詞的位置，其中 (j,k) 代表 j 字詞中的第 k 個音節； c_n 、 $tp_n=(t_n, t_{n+1})$ 分別為在第 n 個音節與第 $n+1$ 個音節間的連音狀態、聲調組合(tone pair)，在這裡音節間的連音狀態 $c_n \in \{c1, c2, c3\}$ 代表音節間的連音程度， $c1$ 、 $c2$ 、 $c3$ 分別為強連音(tight)、正常連音(normal)、弱連音(loose)； β_{c_n, tp_n}^b 為第 n 個音節受第 $n+1$ 個音節的後向影響因素(backward affecting factor)； $\beta_{c_{n-1}, tp_{n-1}}^f$ 為第 n 個音節受第 $n-1$ 個音節的前向影響因素(forward affecting factor)； μ^p 為基頻軌跡參數的整體平均(global mean)。

(二)、音節長度模型

假設所有影響因素可用累加的方式來表示音節的長度，如式子(2)：

$$sd_n = sd_n^r + \gamma_{t_n} + \gamma_{w_n} + \gamma_{sy_n} + \gamma_{c_{n-1}, fi_{in_{n-1}}}^f + \gamma_{c_n, fi_{in_n}}^b + \mu^d \quad (2)$$

其中 sd_n 、 sd_n^r 、 γ_{t_n} 、 γ_{w_n} 、 γ_{sy_n} 分別為 duration 模型中第 n 個音節的長度、長度殘餘值、聲調、詞中位置及基本音節影響因素； c_n 、 fi_{in_n} 分別為在第 n 個音節與第 $n+1$ 個音節間的連音狀態及第 n 個音節韻母類別與第 $n+1$ 個音節聲母類別之組合(final-initial class pair)； $\gamma_{c_n, fi_{in_n}}^b$ 為第 n 個音節長度受第 $n+1$ 個音節的後向影響因素(backward affecting factor)； $\gamma_{c_{n-1}, fi_{in_{n-1}}}^f$ 為第 n 個音節長度受第 $n-1$ 個音節的前向影響因素(forward affecting factor)； μ^d 為音節長度的整體平均(global mean)。音節韻律與影響因素的關係示意圖以 pitch 模型為例，如圖一：



圖一、音節基頻軌跡參數向量與影響因素關係圖

我們分別假設 \mathbf{sp}_n^r 及 sd_n^r 呈 $N(\mathbf{sp}_n^r; \mathbf{0}, \mathbf{R}^p)$ 及 $N(sd_n^r; 0, R^d)$ 的高斯分佈(Gaussian distribution)，因此 \mathbf{sp}_n 與 sd_n 可表示成數學式如式(3)及(4)：

$$P(\mathbf{sp}_n | t_n, w_n, c_{n-1}, c_n, tp_{n-1}, tp_n) = N(\mathbf{sp}_n; \beta_{t_n} + \beta_{w_n} + \beta_{c_{n-1}, tp_{n-1}}^f + \beta_{c_n, tp_n}^b + \mu^p, \mathbf{R}^p) \quad (3)$$

$$P(sd_n | t_n, w_n, sy_n, c_{n-1}, c_n, fi_{in_{n-1}}, fi_{in_n}) = N(sd_n; \gamma_{t_n} + \gamma_{w_n} + \gamma_{sy_n} + \gamma_{c_{n-1}, fi_{in_{n-1}}}^f + \gamma_{c_n, fi_{in_n}}^b + \mu^d, R^d) \quad (4)$$

三、模型的訓練

爲了求取韻律模式的各個參數，我們採用 sequential optimization 的方法以及最大相似度法則(Maximum likelihood criterion)的條件來訓練模型，我們首先定義相似度函數(likelihood function)如下式：

$$L^p = \sum_{n=1}^N \log N(\mathbf{sp}_n; \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{w_n} + \boldsymbol{\beta}_{c_{n-1}, tp_{n-1}}^f + \boldsymbol{\beta}_{c_n, tp_n}^b + \boldsymbol{\mu}^p, \mathbf{R}^p) \quad (5)$$

$$L^d = \sum_{n=1}^N \log N(sd_n; \gamma_{t_n} + \gamma_{w_n} + \gamma_{sy_n} + \gamma_{c_{n-1}, fi_{in_{n-1}}}^f + \gamma_{c_n, fi_{in_n}}^b + \boldsymbol{\mu}^d, R^d) \quad (6)$$

pitch 以及 duration 模型獨立訓練各自的參數，且訓練方法類似，訓練的過程可分爲兩大部分，第一部分爲參數的初始化，第二部份爲以疊代法的 sequential optimization。以下以訓練 pitch 模型爲例：

(一)、參數的初始化(Initialization)

(a) 直接平均所有音節的 \mathbf{sp}_n ，求出整體 pitch 平均(global pitch mean) $\boldsymbol{\mu}^p$

(b) 以下式求取聲調影響因素的初始值：

$$\boldsymbol{\beta}_t = \frac{\sum_n ((\mathbf{sp}_n - \boldsymbol{\mu}^p) \delta(t_n = t))}{\sum_n \delta(t_n = t)}, \text{ for } t = 1, 2, \dots, 5 \quad (7)$$

(c) 以下式求取詞位置影響因素的初始值：

$$\boldsymbol{\beta}_w = \frac{\sum_n ((\mathbf{sp}_n - \boldsymbol{\beta}_{t_n} - \boldsymbol{\mu}^p) \delta(w_n = w))}{\sum_n \delta(w_n = w)}, \text{ for } w = (2, 1), (2, 2), \dots, (8, 8) \quad (8)$$

(d) 以下列的條件，標記音節間的連音狀態 c_n

- I. 若兩音節間基頻軌跡相連接，表示兩連音互相影響程度強，連音狀態標記爲強連音 c1。
- II. 兩音節間的基頻軌跡不相連接，但音節間的時間區間內最低能量較大(大於一個臨界值)，連音狀態標記爲正常連音 c2。
- III. 不滿足以上條件者，則連音狀態標記爲弱連音 c3。

(e) 求取前後音節影響因素的初始值，如下式：

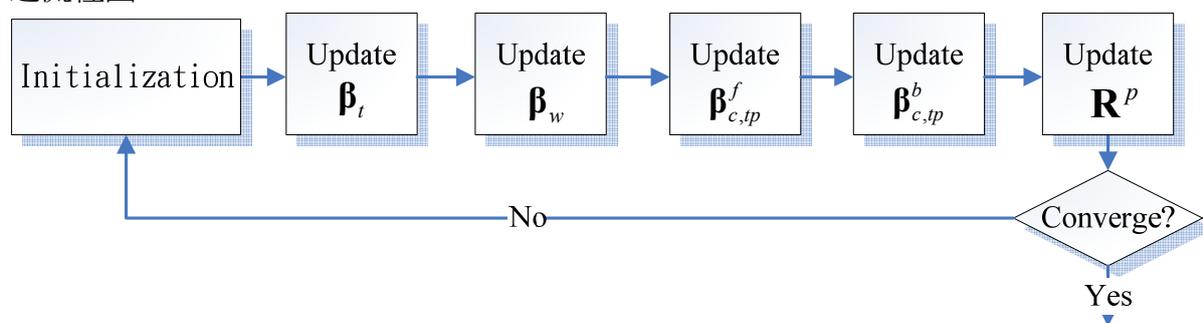
$$\boldsymbol{\beta}_{c, tp}^f = \frac{\sum_{n=1}^N \mathbf{sp}_n \delta(c_{n-1} = c) \delta(tp_{n-1} = tp)}{\sum_{n=1}^N \delta(c_{n-1} = c) \delta(tp_{n-1} = tp)} - \frac{\sum_{n=1}^N \mathbf{sp}_n \delta(c_{n-1} = c) \delta(t_n = j)}{\sum_{n=1}^N \delta(c_{n-1} = c) \delta(t_n = j)} \quad (9)$$

$$\boldsymbol{\beta}_{c, tp}^b = \frac{\sum_{n=1}^N \mathbf{sp}_n \delta(c_n = c) \delta(tp_n = tp)}{\sum_{n=1}^N \delta(c_n = c) \delta(tp_n = tp)} - \frac{\sum_{n=1}^N \mathbf{sp}_n \delta(c_n = c) \delta(t_n = i)}{\sum_{n=1}^N \delta(c_n = c) \delta(t_n = i)} \quad (10)$$

For $c = 1 \sim 3$ and $tp = (i, j)$

(二)、以疊代法的 sequential optimization

各個影響因素初始化後依序將聲調(β_t)、word-position(β_w)、受前後音節等影響因素($\beta_{c,tp}^f, \beta_{c,tp}^b$)及 covariance matrix(R^p)的參數值更新，然後使用更新後的參數值，算出整個訓練語料的目標函數值，一直重覆更新參數值及目標函數值，直到含數值收斂，如圖二之流程圖：



圖二、訓練流程圖

而 duration model 各個參數更新方法與 pitch model 類似，而參數更新的順序為聲調(γ_t)、word-position(γ_w)、基本音節(γ_{sy})、受前後音節等影響因素($\gamma_{c,fi_in}^f, \gamma_{c,fi_in}^b$)及 covariance matrix(R^d)。

四、實驗結果與分析

實驗語料庫之單詞來自於『NCTU 文句分析器』的詞典選擇而來，以聲調平衡為主要的選擇條件，總共有 107936 個詞，277218 個字，其中詞長最短為二字詞、最長八字詞，詞長統計如表一，聲調統計如表二，語料庫是由專業的女性廣播人員以流利的方式唸出錄製，錄音場所為一般安靜房間。

表一、詞長數量之統計

詞長	二字詞	三字詞	四字詞	五字詞	六字詞	七字詞	八字詞
數量	64872	26026	16062	797	124	49	6

表二、聲調數量之統計

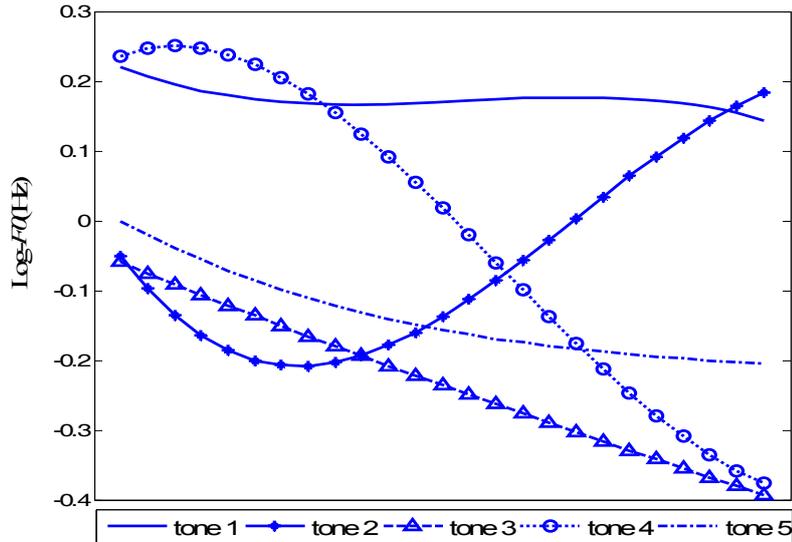
聲調	一聲	二聲	三聲	四聲	五聲
數量	62349	69278	48904	94786	1901

接下來我們依序分析基頻軌跡、音節長度韻律模型以及預測 pause 長度。

(一)、基頻軌跡韻律模型

1、聲調影響因素(Tone affecting factor) β_t

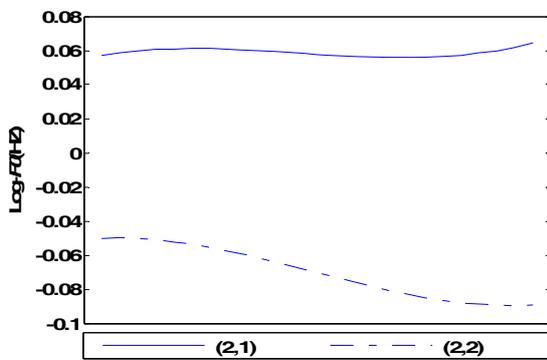
聲調影響因素的基頻軌跡如圖三所示，觀察得知，由模型所訓練出來的聲調影響因素符合我們所認知的聲調基頻軌跡。



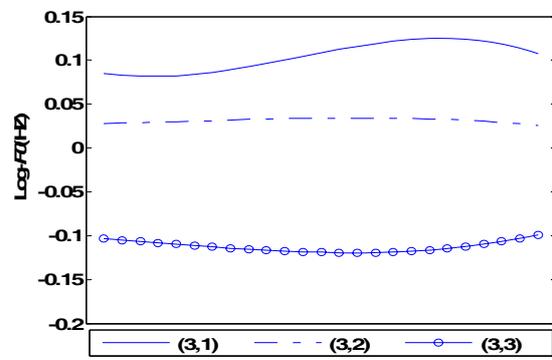
圖三、聲調影響因素基頻軌跡

2、音節在詞的位置影響因素(Word position affecting factor) β_w

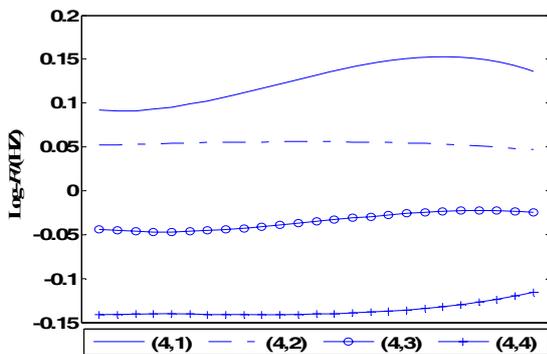
Word position affecting factor 的基頻軌跡影響如下圖所示，分別有二字詞至五字詞位置影響因素(圖四~七)，同時將詞首與詞末的基頻軌跡整合比較(圖八、九)，可觀察得，音節在詞的位置愈接近詞首，基頻愈高，愈接近詞末愈低，且詞首有上仰趨勢(圖八)，在詞尾均有微幅的上揚(圖九)，同時我們也發現到詞長越長，則基頻軌跡變化的動態範圍愈大，如二字詞的動態範圍在 0.06 ~ -0.08，而五字詞的動態範圍在 0.15 ~ -0.17 之間。



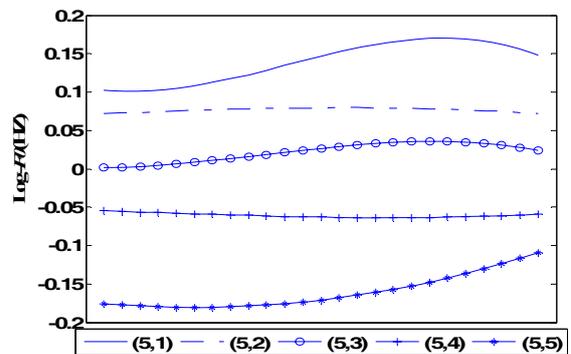
圖四、二字詞影響因素基頻軌跡



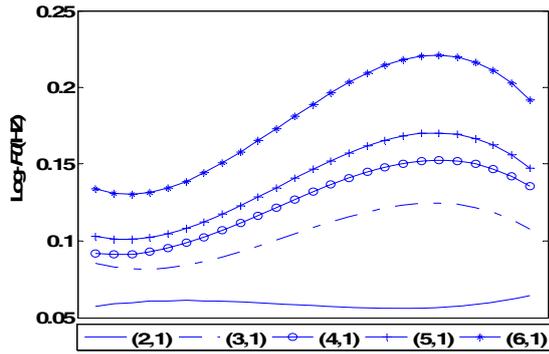
圖五、三字詞影響因素基頻軌跡



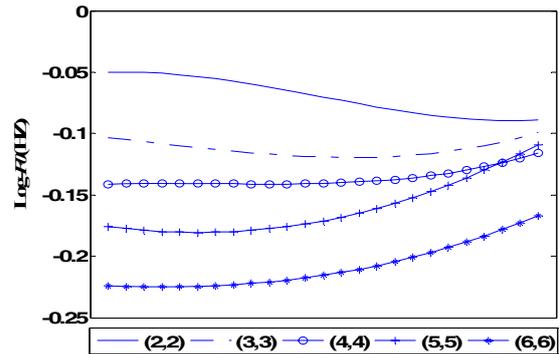
圖六、四字詞影響因素基頻軌跡



圖七、五字詞影響因素基頻軌跡



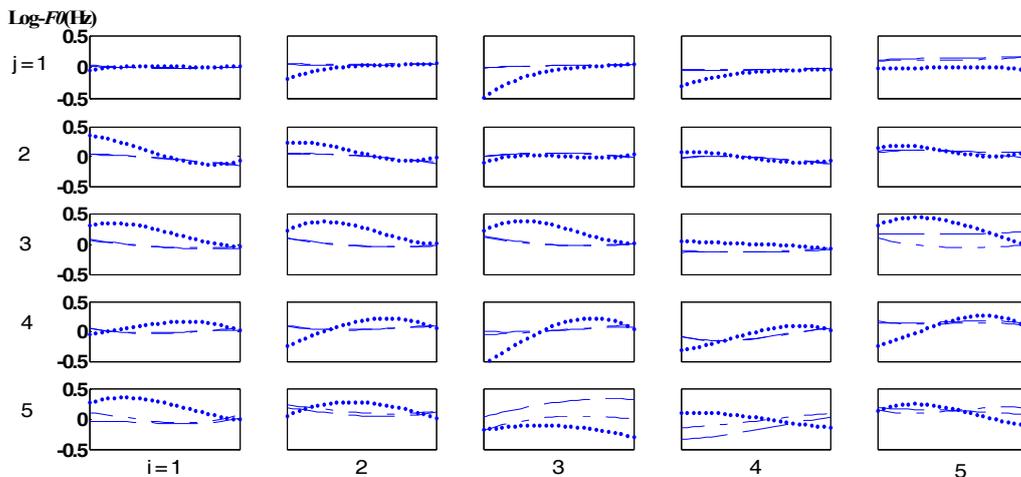
圖八、整合詞首基頻軌跡



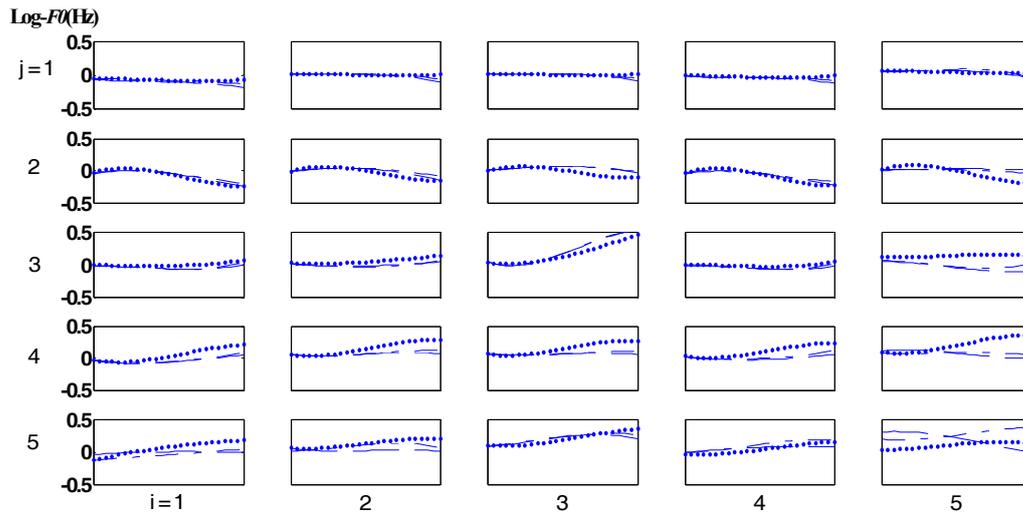
圖九、整合詞末基頻軌跡

3、受前後音節影響因素(forward and backward affecting factor)

圖十為五個聲調受前音節中各個聲調在各個連音狀態影響的基頻軌跡。可觀察得c1明顯比c2及c3易受影響。以c1為觀察對象，目前音節的基頻軌跡前端高度，受前一個音節基頻軌跡後端高度影響，若前音節基頻軌跡後端比受影響的前端聲調基頻軌跡還高，則受影響的前端基頻軌跡會往上改變，反之往下改變。圖十一為受後音節的影響，其影響原理類似於受前音節的影響因素，可觀察得受後音節較受前音節的基頻軌跡影響相對較小。



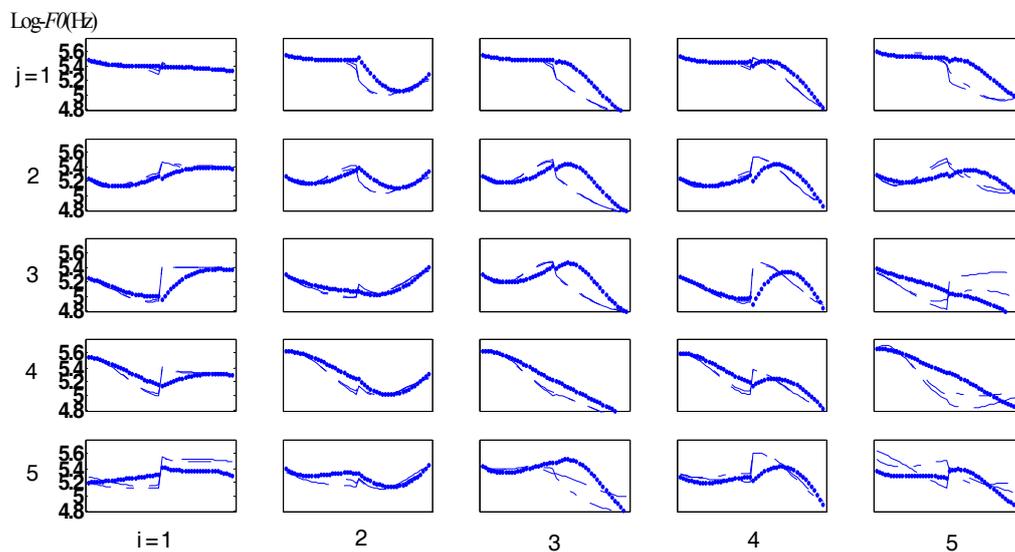
圖十、Forward affecting factor $\beta_{c,tp}^f$, $c = \{c1,c2,c3\}$, $tp=(i,j)$, 其中點線(...)為c1、點虛線(-·-)為c2、虛線(--)為c3



圖十一、Backward affecting factor $\beta_{c,tp}^b$, $c = \{c1,c2,c3\}$, $tp=(j,i)$, 其中點線(...)為c1、點虛線(--·)為c2、虛線(--)為c3

4、二字詞基頻軌跡預測

圖十二為各種二字詞聲調組合(tone pair)的預測結果，假設二字詞音節基頻軌跡相連且相互影響，可觀察得c1明顯易受影響，而且c1在受前後音節影響下，兩音節基頻軌跡相連續且平滑。而三聲接三聲變二聲接三聲的語言特性也可由下圖證實。

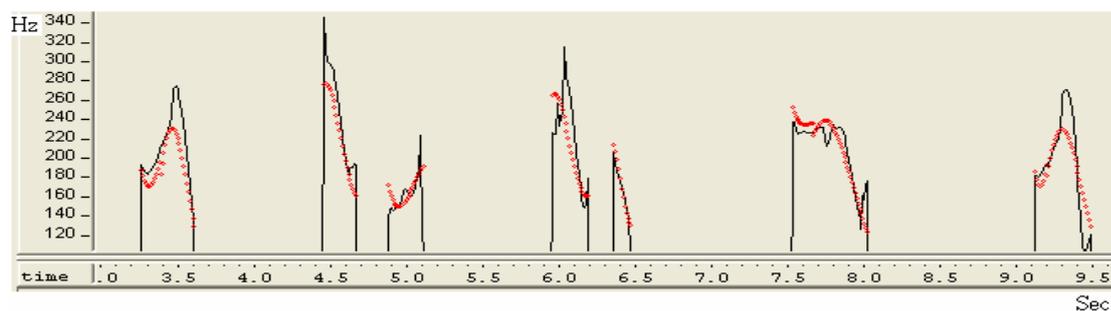


圖十二、二字詞基頻軌跡變化圖，其中點線(...)為c1、點虛線(--·)為c2、虛線(--)為c3，其j為第一音節，i為第二個音節

5、預估之基頻軌跡範例

訓練模型後，pitch 模型預測音節基頻軌跡結果如下圖所示，其中黑色線(實線)為原始音

節的基頻軌跡，紅色線(點線)為 pitch 模型所預測的基頻軌跡，可觀察得，不僅個別音節的基頻軌跡走勢相似，且強連音音節基頻軌跡也預測的不差。



圖十三、Pitch 模型預測基頻軌跡

6、Covariance matrix 比較

Covariance matrices 在訓練前(covariance matrices of the original syllable F0)與訓練後(covariance matrices of the normalized syllable F0)分別為 $\mathbf{R}_{original}^P$ 及 \mathbf{R}^P 。可觀察得訓練模型後 covariance 明顯有意義下降。

$$\mathbf{R}_{original}^P = \begin{bmatrix} 0.040124 & 0.0053695 & -0.0020751 & -0.00075677 \\ 0.0053695 & 0.018669 & 0.0020441 & -0.0016243 \\ -0.0020751 & 0.0020441 & 0.0037581 & 8.3657 \times 10^{-5} \\ -0.00075677 & -0.0016243 & 8.3657 \times 10^{-5} & 0.0011703 \end{bmatrix}$$

$$\mathbf{R}^P = \begin{bmatrix} 0.011843 & 0.0021328 & 9.2796 \times 10^{-5} & -0.00035788 \\ 0.0021328 & 0.0052484 & 0.0011326 & -0.00030867 \\ 9.2796 \times 10^{-5} & 0.0011326 & 0.0022113 & 0.00033545 \\ -0.00035788 & -0.00030867 & 0.00033545 & 0.00090293 \end{bmatrix}$$

(二)、Duration 模型

1、聲調影響因素(Tone affecting factor)

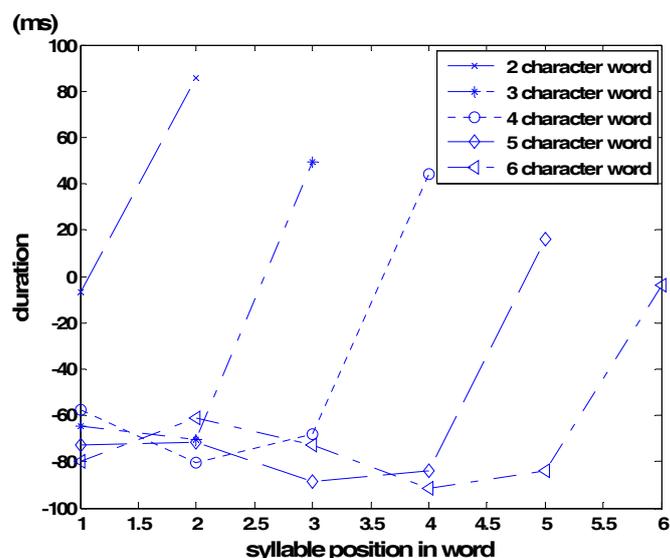
聲調對音節長度影響程度如表三，觀察得五聲音節最短，二聲音節最長。

表三、聲調對音節長度影響

聲調	一聲	二聲	三聲	四聲	五聲
長度	6.9ms	35.0ms	-22.5ms	-17.2ms	-82.2ms

2、詞的位置影響因素(Word-position affecting factor)

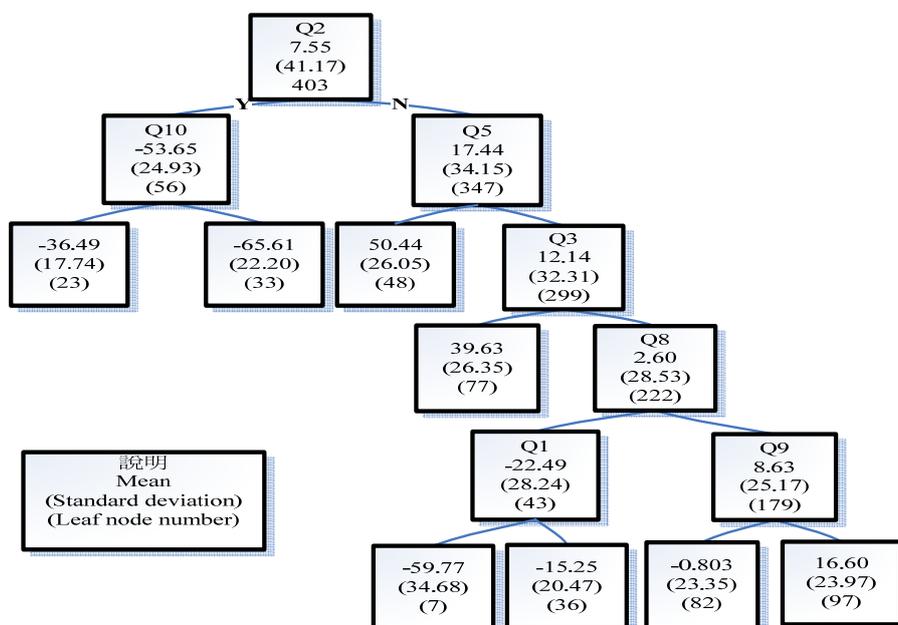
音節在詞的位置對長度影響程度如圖十四，可觀察在詞的字末位置音節長度較長，觀察五字詞、六字詞可得知詞愈長愈容易產生較短的音節。



圖十四、詞的位置影響因素

3、音節影響因素(syllable affecting factor)

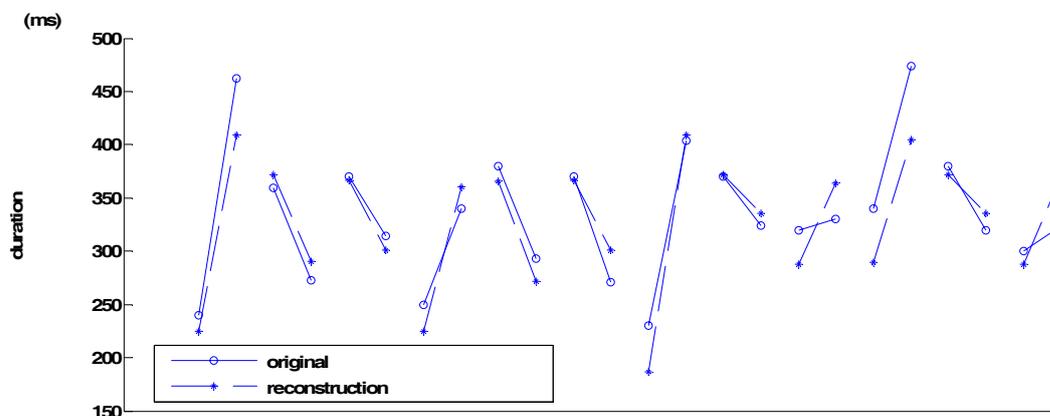
我們使用決策樹觀察音節影響因素對音節長度影響程度，如圖十五。由決策樹可觀察出聲母為空聲母(Q1)、{b、d、g}(Q2)或韻母類別為單母音(Q8)這類音節長度較短，而鼻音結束(Q10)、{f,s,sh,x,h}(Q3)及{c,ch,q}(Q5)這類摩擦音的音節長度會較長。



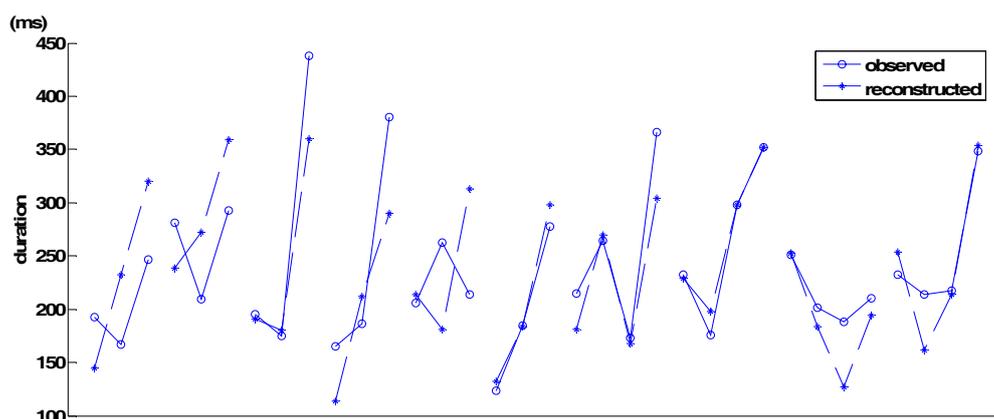
圖十五、以決策樹分析音節影響因素分類結果

4、音節長度預估之範例

Duration 模型預測音節長度與實際長度比較，如圖十六、十七分別為二字詞及三、四字詞的比較圖。其中 **observed** 為音節實際長度，**reconstructed** 為音節預測長度，其每線段為一單詞，大致有不錯的預測結果。



圖十六、二字詞音節實際長度與模型預測長度之比較圖



圖十七、三字詞及四字詞音節實際長度與模型預測長度之比較圖

5、Variance 比較

訓練模型後，以訓練前 variance 為 9304.5 與訓練後 variance 為 2494.7。可觀察得使用模型後變異量明顯有意義下降。

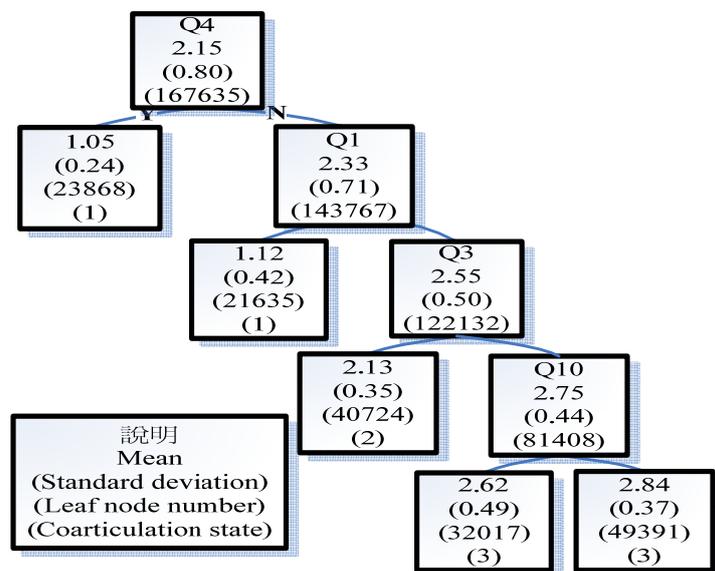
(三)、以決策樹預測 **pause** 長度

在下面小節中，利用決策樹預測音節間的連音狀態及 **pause** 長度，其應用於 question set 中的聲母類別和韻母類別分別為兩音節的間隔區間相鄰的聲母類別及韻母類別。

1、預測音節間之連音狀態

以連音狀態做為決策樹分類的目標，由 **c1** 設值為 1，**c2** 設值為 2，**c3** 設值為 3。由分類

結果觀察得聲母為 NULL (Q1)及{m,n,l,r}(Q4)分別為 mean=1.1173 及 mean = 1.047，意指分佈較集中於 c1；而聲母為{f,s,sh,x,h}(Q3)，此類 mean=2.1341，是指分佈較集中於 c2，其他類別假設為 c3。如圖十八所示。



圖十八、以決策樹分析語料庫連音狀態的分類結果

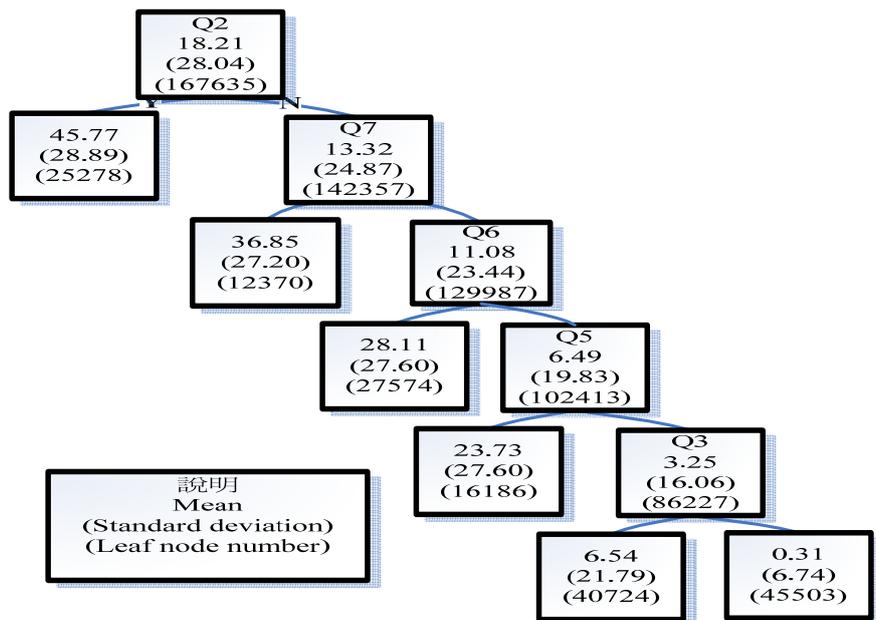
2、預測音節間之 pause 長度

由語料庫中 pause 的長度視為目標值，其單位為毫秒(ms)。由圖十九觀察得 Pause 長度分類明顯與 pause 相鄰的聲母類別有關。將 Pause 長度與聲母類別關係整理如下：
 爆破音_不送氣 > 爆破音_送氣 > 塞擦音_不送氣 > 塞擦音_送氣 > 摩擦音_清音 > m,n,l,r 及空聲母。

假設預測 pause 長度可直接利用相鄰的聲母類別判斷，由圖十九可觀察得 pause 長度分類的 mean 設為預測的 pause 長度，整理於表四，其中聲母類別為 NULL 及 {m,n,l,r} 的 mean 為 0.31028，可假設 pause 長度為 0ms，滿足由圖十八預測此種類別為連音狀態最明顯的 c1。

表四、pause 長度選取表

類別	聲母	Pause 長度
1	ㄇ、ㄋ、ㄌ、ㄍ、空聲母 (鼻音_濁音)	0ms
2	ㄈ、ㄊ、ㄑ、ㄒ、ㄓ (摩擦音_清音)	7ms
3	ㄅ、ㄆ、ㄇ (爆破音_不送氣)	46ms
4	ㄆ、ㄑ、ㄒ (塞擦音_不送氣)	28ms
5	ㄆ、ㄑ、ㄒ (爆破音_送氣)	37ms
6	ㄆ、ㄑ、ㄒ (塞擦音_送氣)	24ms



圖十九、以決策樹分析語料庫 pause 長度的分類結果

再假設預測 pause 長度與 pause 在詞的位置有相關性，利用 pause 在詞的位置以決策樹分析整理於表五及表六：

表五、二字詞第一個 pause 位置的 pause 長度選取表

類別	1	2	3	4	5	6
(2,1)	0ms	8ms	56ms	34ms	44ms	29ms

表六、三字詞至四字詞 pause 在詞位置的 pause 長度選取表其中(詞長，詞中第幾個 pause)。

類別 位置	1	2	3	其他類別
(3,1)	0ms	5ms	36ms	23ms
(3,2)	0ms	7ms	45ms	28ms
(4,1)	0ms	6ms	33ms	20ms
(4,2)	0ms	7ms	41ms	26ms
(4,3)	0ms	8ms	47ms	28ms

由表六可觀察得主要影響 pause 長度主要為聲母類別有 1、2、3 類，且在詞中較後面的 pause 有較長的停頓時間。

五、結論

使用各種影響因素相加，來預測各種韻律訊息，及討論各種影響因素對韻律訊息的分析，由實驗結果證實各種影響因素相加便可預測各種韻律訊息，其各種影響因素分析也

符合語言特性。相信未來應用於語音合成，可明顯自然流暢許多。最後，本文所提出的方法，不因語言不同而有所改變，所以未來可朝向建立一套整合國、台、客語的韻律產生器及韻律分析邁進。

參考文獻

- [1] L.-S. Lee, C.-Y. Tseng, and M. Ouh-Young, "The Synthesis Rule in a Chinese Text-to-Speech System," *IEEE Trans. Acoust, Speech, Signal Processing*, vol.37, no.9, pp.1309-1319, Sep. 1989.
- [2] L.-S. Lee, C.-Y. Tseng, and C.-J. Hsieh, "Improved Tone Concatenation Rules in a Formant-based Chinese Text-to-Speech System," *IEEE Trans. Speech and Audio Processing*, vol.1, no.3, pp.287-294, July 1993.
- [3] S.-H. Chen, S.-H. Hwang, and Y.-R. Wang, "An RNN-based Prosodic Information Synthesizer for Mandarin Text-to-Speech," *IEEE Trans. Speech and Audio Processing*, vol.6, no.3, pp.226-239, May 1998.
- [4] 黃紹華，"中文文句翻語音系統中韻律訊息產生器之研究"，國立交通大學博士論文，民國八十五年六月。
- [5] S.-H. Chen, W.-H. Lai, and Y.-R. Wang, "A Statistics-base Pitch Contour Model for Mandarin Speech," *J. Acoust. Soc. AM.* 117(2), Feb. 2005, pp. 908-925
- [6] C.-Y. Chiang, Y.-R. Wang, and S.-H. Chen, "On the Inter-syllable Coarticulation Effect of Pitch Modeling for Mandarin Speech," *Proc. of Interspeech 2005*, Lisbon, Portugal, pp. 3269-3272
- [7] 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, July