

國語歌聲抖音參數之分析

Analysis of Vibrato Parameters for Mandarin Singing Voice

古鴻炎 林正甫

國立台灣科技大學資訊工程系
{guhy, M9415017}@mail.ntust.edu.tw

摘要

國語歌聲信號的合成，如果沒考慮到情感表現的因素，則合成歌聲聽起來會顯得呆板、不自然，因此本文對歌聲表情的一個重要因素“抖音”，研究它的特性參數的分析方法。抖音的三個特性參數是音位軌跡、抖動範圍、和抖動頻率，我們嘗試了基週峰谷法和瞬間頻率法，來分析一個歌聲音節中這三種參數的時變曲線。雖然這兩種分析方法是前人提出的，但是從實作與實驗的過程中，我們發現原始分析方法仍存在有實作上的幾個缺點，因此我們研究改進了這些缺點。此外，我們也考慮了抖音特性參數之曲線取樣後，再合成回來的一種作法。

1. 前言

過去我們曾經作過國語歌聲合成的研究，例如以弦波模型為基礎，分析國語音節有聲部分的諧波參數，再依據諧波參數作歌聲音節的合成[1]。由於使用弦波模型(sinusoidal model)所合成的歌聲，信號高頻部分諧波成分太強且缺少雜音，而使音質顯得不自然，因此我們後來改變以諧波加噪音模型(harmonic plus noise model, HNM)為基礎，來研究國語歌聲的合成[2]，所合成出的歌聲信號，自然度和清晰度已有大幅度的改進。

然而在先前的研究裡，我們還沒去作情感表達有關的信號合成的處理，因此所合成出的歌聲信號，聽起來感覺是機械味頗重而不自然。人類唱的歌曲所以會好聽，是因為歌聲裡含有豐富的情感表現。關於歌聲裡的表情(expression)的呈現，探究有那些聲學(acoustics)因素是和它緊密相關的，然後根據這些聲學因素去建造表情模型，以用來模仿人類歌唱的方式，可說是歌聲合成上重要的研究課題。

從文獻上發現，和歌聲表情呈現有關的因素[3]，包括：(a)抖音(Vibrato)，讓歌聲音高(pitch)顫抖，可使歌聲美化；(b)轉音(Glissando)，在兩種以上不同的音高之間順暢的轉移，可使歌聲有圓滑的表現；(c)彈性速度(Rubato)，根據音樂表現的要求，交替出現漸快和漸慢的速度。其中，以抖音為最常被表現於歌聲之中，也是最為重要的表情因素，因此我們先選擇以抖音為對象，來研究抖音有關之聲學參數的分析方法。

根據前人的研究[4, 5, 6, 7]，指出抖音現象發生時，歌聲的基頻會出現週期性的抖動，而依據基頻抖動的情形，抖音在聲學表現上可用三種參數來加以描述，也就是：(a)音位軌跡(intonation)，基頻平均位準的移動軌跡；(b)振動範圍(vibration extent)，基頻值上升或下降的幅度；(c)抖動速率(rate)，基頻抖動的速率。當分析得到抖音歌聲的參數之後，將來我們就可以利用這些參數來建造抖音有關之表情模型，用以改善合成的歌聲信號中缺乏表情展現的缺點。目前我們只研究了抖音參數分析的問題，未來將再嘗試使用分析出的抖音參數去訓練類神經網路(artificial neural network, ANN)，也就是以類神經網路來建造抖音的表情模型。

2. 抖音分析方法之回顧

過去已有一些研究者提出抖音分析的方法，這些方法可以分成兩類，分別是：(a)基週峰谷法(pitch peak and valley method)，利用抖音信號裡基頻值會隨著時間變動的特性，先偵測出各個基週波形的波峰和波谷，再藉以計算出抖音參數[8, 9, 10]；(b)瞬間頻率(Instantaneous Frequency)法，先分析抖音信號的瞬間頻率，再藉以計算出抖音參數[11, 12, 13, 14, 15]。

2.1 基週峰谷法

在此我們介紹 E. Prame 所提出的基於基週峰谷之分析方法[8, 9, 10]，他的方法的最大優點在於運算量低，所以也常被其他學者所引用。該方法的處理步驟如下：

- 基週軌跡量測，設定長度為 25ms 之音框(frame)，並且音框每次只前進 2.5ms，再將各個音框裡估計出的基頻值，串接成的基週軌跡(pitch contour)曲線，一個例子如圖 1 所示。
- 峰谷偵測，對基週軌跡曲線偵測出各個波峰與波谷的時間位置，如圖 1 裡的 t_{n-1} , t_{n+1} 表示兩個相鄰波谷的時間， t_n 表示波峰的時間，而 f_{n-1} , f_n , f_{n+1} 表示這三個時間點上的基頻值。
- 抖音參數估計，依據前一步驟所偵測到的時間值和基頻值，帶入下列公式去估計抖音參數的數值。

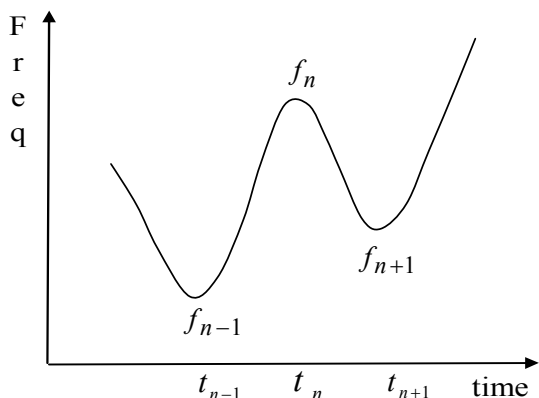


圖 1. 抖音信號的基週軌跡[8]

$$VR(n) = \frac{1}{t_{n+1} - t_{n-1}} \quad (\text{Hz}) \quad (1)$$

$$VE(n) = \frac{|f_{n-1} - 2f_n + f_{n+1}|}{|f_{n+1} + 2f_n + f_{n-1}|} \quad (\text{Hz}) \quad (2)$$

$$VD(n) = \frac{f_{n-1} + 2f_n + f_{n+1}}{F} \quad (\text{Hz}) \quad (3)$$

其中 $VR(n)$ 表示第 n 個波峰(或波谷)點上的抖動頻率(rate)， $VE(n)$ 表示第 n 個波峰點上的抖動範圍(extent)， $VD(n)$ 表示第 n 個波峰點上的音位，而 F 表示基週軌跡的幾何平均值。

2.2 瞬間頻率法

先量測抖音信號在各個時間點上的瞬間頻率值[11, 12]，再依據瞬間頻率隨著時間變化的軌跡來估計抖音參數的數值。在文獻[13]中，提到兩種計算瞬間頻率的方法，一個是基於 STFT(Short-Time Fourier Transform)的方法，另一個則是基於 Analytic Signal 的方法。

一個以 STFT 為基礎的方法是，設定音框的長度(如 1024 點)，加上 Blackman Window，讓音框每次移動 64 個取樣點，然後對每個音框作 FFT (Fast Fourier Transform)，依據 FFT 求得的頻譜曲線，尋找曲線上第一個諧波的頻率值，再將各音框求得的頻率值連合成瞬間頻率曲線。

以 Analytic Signal 為基礎的方法，首先需要對一個音符的歌聲信號作帶通(bandpass)濾波的處理，將所關心的諧波(通常是第一個諧波)以外的其它諧波成分先濾除掉。令帶通濾波後的信號是 $s(t)$ ，依據 $s(t)$ ，Gabor 提出了一個熟知的 Analytic Signal 的定義方式[16]，即以如下公式定義 $s(t)$ 的 Analytic Signal 為 $z(t)$ ，

$$z(t) = s(t) + jH[s(t)] \quad (4)$$

$$H[s(t)] = \hat{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(t')}{t-t'} dt' \quad (5)$$

$z(t)$ 是一個複數值信號，由實部(real part) $s(t)$ 與虛部(imaginary part) $H[s(t)]$ 所形成，這裡 $H[s(t)]$ 表示對 $s(t)$ 作 Hilbert 轉換[17, 18]，Hilbert 轉換能夠把信號的相位角做 ± 90 度的旋轉。例如對於一個 $q(t) = a(t)\cos(\theta(t))$

的信號， $a(t)$ 是 $q(t)$ 的瞬間振幅、而 $\theta(t)$ 是瞬間相位，當對信號 $q(t)$ 作 Hilbert 轉換之後，由於相位角旋轉了 90 度，轉換後的信號將會變成 $H[q(t)] = a(t)\sin(\theta(t))$ 。由於公式(4)得到的 $z(t)$ 是一個複數值信號，因此可將它表示成

$$z(t) = a(t)e^{j\theta(t)} \quad (6)$$

來求得瞬間振幅 $a(t)$ 和瞬間相位 $\theta(t)$ ，然後對瞬間相位 $\theta(t)$ 作微分，即可獲得瞬間頻率 $f(t)$ ，亦即 $f(t) = (1/2\pi)(d\theta(t)/dt)$ 。在得到瞬間頻率函數 $f(t)$ 之後，將它看成是具有如下公式所示的形式，

$$f(t) = VD(t) + VE(t) \cdot \cos(\phi(t)) \quad (7)$$

然後對它作分解的動作，來求得音位(intonation)軌跡 $VD(t)$ 和抖音範圍(extent)曲線 $VE(t)$ ，至於抖音頻率曲線 $VR(t)$ ，可對公式(7)裡的 $\phi(t)$ 作微分來求得。

3. 分析方法之實驗與改進

在研讀了一些相關的文獻之後，接著我們便進行抖音分析程式的製作，並且對前人提出的抖音參數分析方法作實驗驗證。從實驗的過程中，發現了前人方法在實作上的一些缺點，因此我們也研究了一些改進的作法。在進行國語歌聲抖音參數的分析之前，我們必需先作一些準備的動作，如圖 2 所示，首先是進行第 1

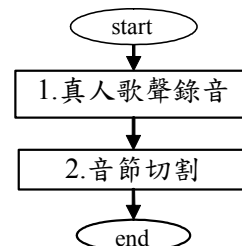


圖 2. 抖音分析的前置作業

個方塊之”真人歌聲錄音”，我們邀請了一位擅長歌唱的男性在隔音錄音室 Acoustic Systems RE-242 之內清唱歌曲，同時使用錄音軟體 Cubase SX3 來將歌聲儲存成取樣率 22,050Hz、解析度 16bits/sample 的電腦音檔，總共錄了 15 首快、慢節奏歌曲的歌聲信號。接著進行第 2 個方塊之”音節切割”的動作，對所錄製的各首歌曲音檔，依據歌譜將每一個歌詞字所對應的信號片段切割出來，並且儲存成獨立的音檔，以方便後續步驟裡的分析程式來取用，在此我們使用了 CoolEdit 軟體來作信號的編輯與切割。

3.1 基週峰谷法

對於一個歌詞字(音節)所切割出來的信號，我們一開始是使用基週峰谷法[8, 9, 10] 來作抖音參數的分析，當分析一個抖音很明顯的歌聲信號，如圖 3 裡的 /wei/ 音波形，偵測基週後得到的基週軌跡如圖 4 所示，再依公式(1)計算其抖動頻率，則得到如圖 5 所示的抖動率數值的線段序列，在此所採取的基週偵測方法是，合併使用自相關(autocorrelation)函數和

AMDF (absolute magnitude difference function)函數的一種作法[19]；另一情況，當分析一個抖音不明顯的歌聲信號，如圖 6 裡的 /yi/ 音波形，偵測基週後得到的基週軌跡如圖 7 所示，再依公式(1)計算其抖動率，則得到如圖 8 所示的單一個抖動率數值的線段。

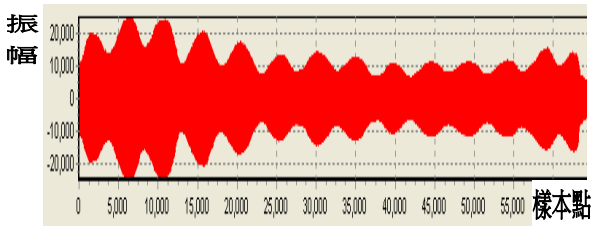


圖 3. /wei/音之波形

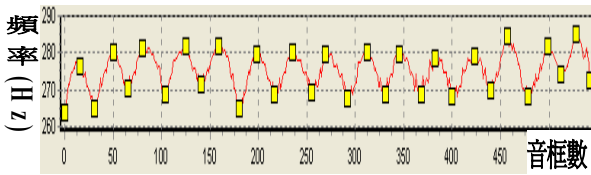


圖 4. /wei/音之基週軌跡

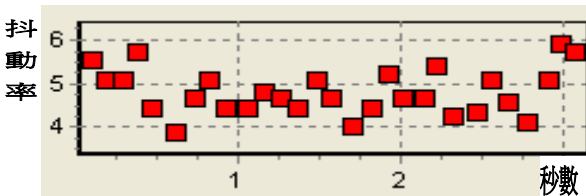


圖 5. /wei/音之抖動率

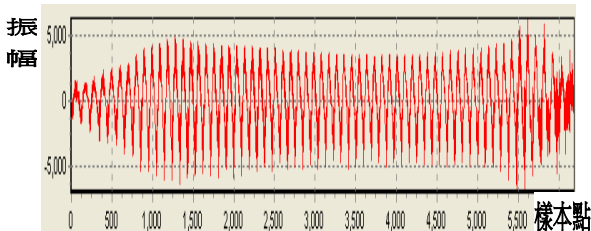


圖 6. /yi/音之波形

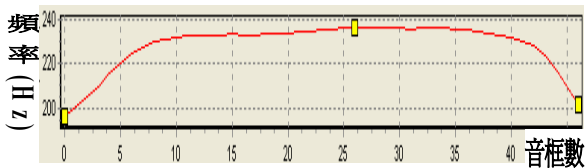


圖 7. /yi/音之基週軌跡

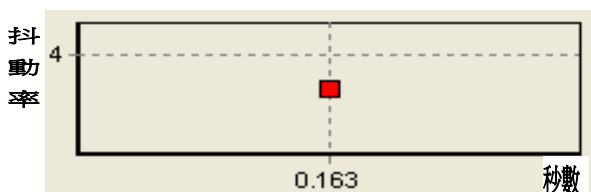


圖 8. /yi/音之抖動率

觀察圖 5 和圖 8 可發現，抖音頻率參數在時間軸上的

精確度不夠，如圖 8 只分析得到一點，這是因為圖 7 裡只有一個波峰和兩個波谷。將來若要訓練抖音參數的 ANN 模型，則輸入的資料必需要有足夠的數量，因此我們決定放棄基週峰谷法這類的作法。

3.2 瞬間頻率法

採取瞬間頻率法作抖音參數分析時，我們依照圖 9 所

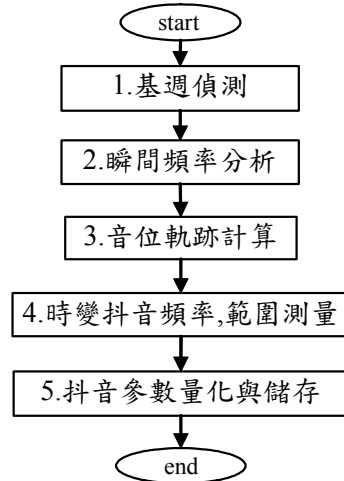


圖 9. 抖音參數分析之步驟

示的步驟來進行，首先是對一個切割出的音節信號作基週偵測，以便將音節前面可能存在的無聲(unvoiced)信號部分(如/sa/等以無聲子音開頭的音節)界定出來，在此我們仍是採取前述的合併自相關函數和 AMDF 函數的基週偵測作法[19]。偵測出音節的有聲(voiced)信號部分之後，接著就對有聲的部分進行圖 9 裡的第 2 到第 5 的步驟，較詳細的說明在以下各子節裡。

3.2.1 瞬間頻率分析

界定出有聲的信號部分之後，接著要作瞬間頻率的分析，不過在此之前我們必需先作窄頻帶的帶通(bandpass)濾波處理，只讓第一個諧波頻率附近的信號成分被保留下來。實際作法是，對整段信號作 FFT 轉換，而得到 long-term magnitude spectrum，再以程式自動偵測出頻譜曲線上第一個波峰(即諧波頻率)及此波峰左右相鄰之波谷頻率位置，設頻率分別為 f_u 和 f_v ，依據 f_u 和 f_v ，我們可將 0Hz 至 f_u 之間和 f_v 至 $F_s/2$ (即取樣率半值)之間的 spectrum amplitude 全設定為 0 值，如此就是在頻域上作窄頻帶的帶通濾波處理，然後作 IFFT (Invert FFT)轉換，來還原成時域信號。當對圖 3 的 /wei/音波形作窄頻帶通濾波之後，得到的波形如圖 10 所示；如果對圖 6 的 /yi/音波形作窄頻帶的帶通濾波處理，則得到的波形如圖 11 所示。

接著採取 2.2 節裡提到的 Analytic Signal 方法，對信號作 Hilbert 轉換，來得到另一份相位被旋轉 90 度的信號，然後依照公式(6)就可獲得信號的瞬間相位函數 $\theta(t)$ ，對 $\theta(t)$ 作微分就可得到瞬間頻率函數 $f(t)$ 。關於 Hilbert 轉換，我們由文獻上查到一個實作的方法

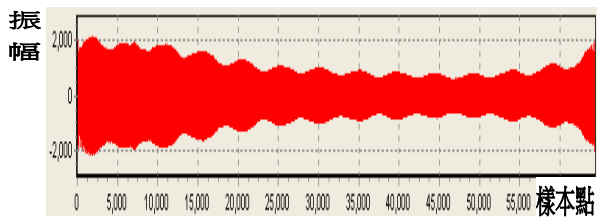


圖 10. 窄頻帶之/wei/音波形

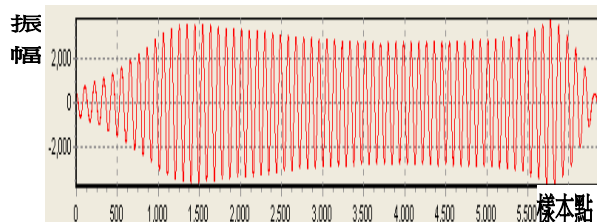


圖 11. 窄頻帶之/yi/音波形

[20, 21, 22], 其計算步驟為:

- (i) 對信號作 FFT 運算，以計算出 Long-term Spectrum；
- (ii) 調整頻譜值，將頻率值比 $F_s/2$ 小之頻率點上的振幅值強迫設定為 0；
- (iii) 作 IFFT 運算，將信號還原至時域。

依據上述的方法，對圖 10 和 11 裡的波形去計算瞬間頻率，結果分別得到如圖 12 和 13 所示的瞬間頻率之時變曲線。

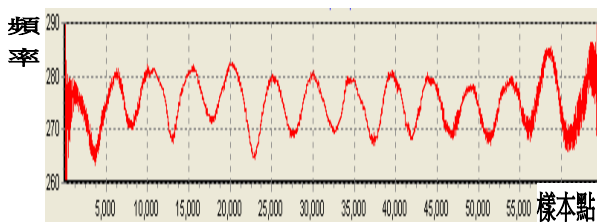


圖 12. /wei/音之瞬間頻率

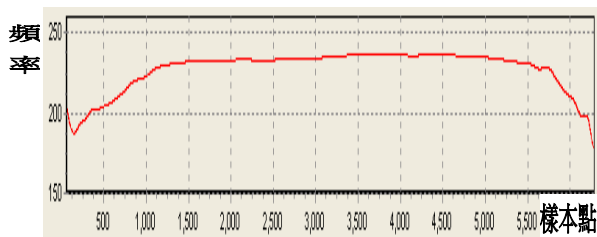


圖 13. /yi/音之瞬間頻率

3.2.2 音位軌跡計算

得到瞬間頻率曲線 $f(t)$ 之後，依據公式(7)，我們必需對它作分解的動作，以求得三個抖音參數， $VD(t)$ 、 $VE(t)$ 和 $VR(t)$ 。首先考慮音位參數曲線 $VD(t)$ 的求解，我們從文獻上查到一個低通濾波的想法[13]，因此我們使用如下的計算步驟來加以實踐:

- (i) 計算出 DC 值(即平均值)，將 $f(t)$ 減去 DC 值；
- (ii) 作 FFT 轉換，得到頻譜；
- (iii) 在頻域上作低通濾波，就是把頻譜上頻率值超過

3.5Hz 的頻率點上的振幅值強迫設定為 0；

(iv) 作 IFFT 轉換，還原成時域波形。

以此方法對圖 13 裡的瞬間頻率函數作處理，結果得到如圖 14 所示的音位軌跡曲線，所以對一個抖音不明顯的信號來說，在左右兩端會發生音位軌跡和瞬間頻率之間有著很大的曲線斜率誤差的情形，因此我們決定另外研究求解 $VD(t)$ 的方法，以緩和此種誤差情形。

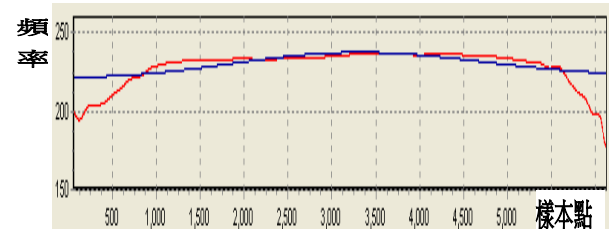


圖 14. /yi/音音位軌跡作 FFT 低通濾波

一番思考後我們決定採取的作法是，在時域上直接用 moving average 來作低通濾波處理，也就是在時刻 t 時，取 t 前後固定範圍內的瞬間頻率值，計算出平均值作為時刻 t 上的 $VD(t)$ 值。使用這樣的方法，對圖 13 的瞬間頻率作處理，結果得到如圖 15 所示的音位軌跡曲線，和圖 14 裡的比較起來，在左右兩端的曲線斜率誤差已經減少許多，不過左右兩端會失去一些時間點。

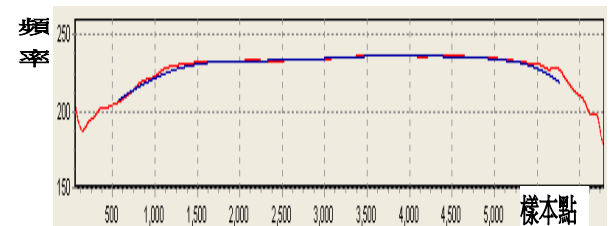


圖 15. /yi/音作 MA 濾波的音位軌跡

3.2.3 時變抖音頻率、範圍量測

依公式(7)將瞬間頻率 $f(t)$ 減去 3.2.2 節計算出的 $RD(t)$ ，就可得到 $VE(t) \cdot \cos(\phi(t))$ 的部分，不過時變的 $VE(t)$ 和 $\phi(t)$ 要如何去求取？在此，我們可再次作 Hilbert 轉換，然後依據公式(4)和(6)，來求得時變的振幅 $VE(t)$ 和 $\phi(t)$ ，接著對 $\phi(t)$ 作微分再乘上 $1/2\pi$ 來求得 $VR(t)$ 。當我們對圖 12 的瞬間頻率作上述的處理，得到的抖動頻率 $VR(t)$ 和抖動範圍 $VE(t)$ ，分別如圖 16 和 17 所示；另外當我們對圖 13 的瞬間頻率作上述的處理，得到的抖動頻率和抖動範圍，則分別如圖 18 和 19 所示。

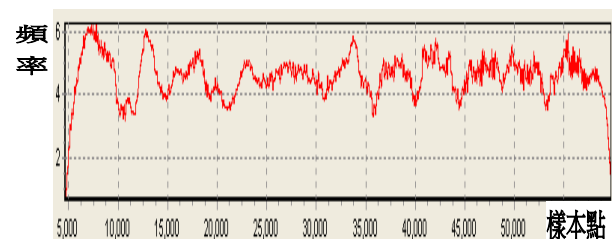


圖 16. /wei/音之抖動頻率

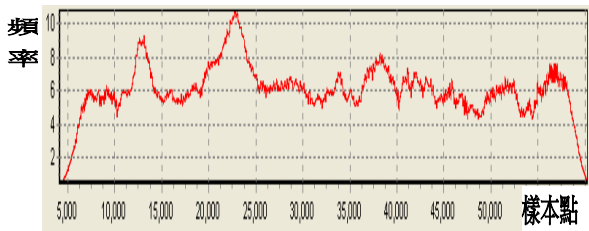


圖 17. /wei/音之抖動範圍

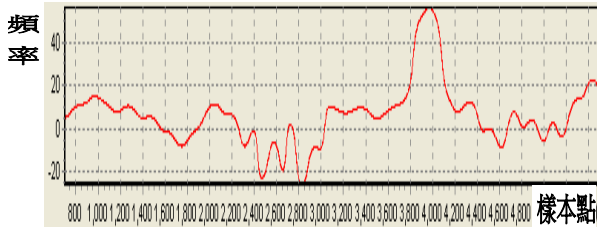


圖 18. /yi/音之抖動頻率

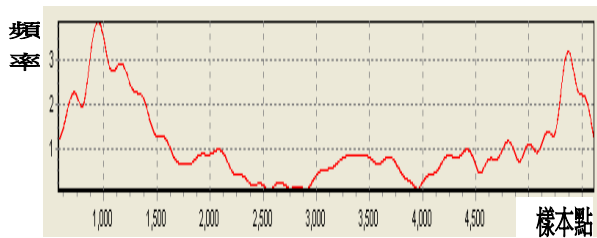


圖 19. /yi/音之抖動範圍

觀察圖 16 與 17 可發現，/wei/音的抖動頻率在 3 ~ 6 Hz 之間，而抖動範圍在 0 ~ 10 Hz 之間，這和[9]所提的抖音參數數值範圍相當吻合。不過觀察圖 18 可發現，/yi/音的抖動頻率相當大，某些時間點上已超過 50Hz，為了解決這個問題，我們研究了一個處理方法，就是在圖 13 的瞬間頻率減去圖 15 的音位軌跡之後，先讓 $VE(t) \cdot \cos(\phi(t))$ 信號通過一個 IIR 低通濾波器，然後才去作 Hilbert 轉換的處理，如此得到的抖動頻率和抖動範圍就會分別如圖 20 和 21 所顯示的情

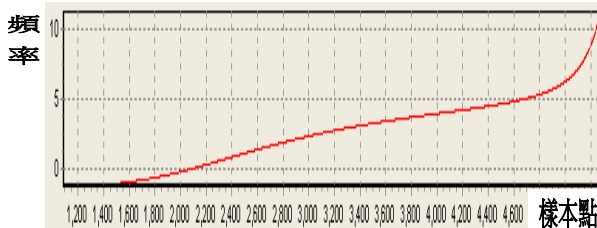


圖 20. /yi/音改進分析之抖動頻率

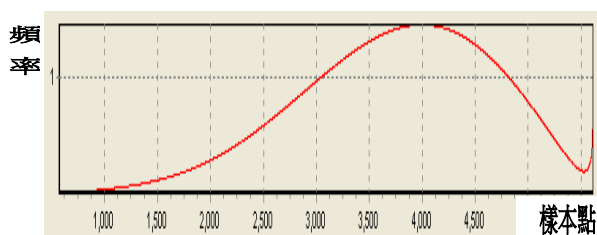


圖 21. /yi/音改進分析之抖動範圍

形，亦即抖動頻率的數值已經低於 10Hz，而抖動範圍的數值也由 3Hz 減少至 1Hz。在此所使用的 IIR 低通濾波器是 6 階 elliptic 類型，通過頻帶是 0~10Hz，15Hz 以上是截止頻帶。

3.2.4 抖音參數取樣與儲存

在分析出一個音節的三種抖音參數之後，接者要對參數數值作取樣和存檔的動作。由於各個歌聲音節的時間長度有長有短，且這些抖音參數的數值將來要拿去訓練 ANN 模型，因此必需對 $VD(t)$ 、 $VE(t)$ 、 $VR(t)$ 三個函數作適當的取樣，以便讓每個音節的資料數量保持相同，並且資料數量也不能太少而造成時間精確度不夠。在此我們決定對每一種抖音參數，在時間軸上都均勻佈放 128 個取樣點，所以三種抖音參數分別會儲存 128 個樣本值至檔案。比較特別的是抖音頻率參數 $VR(t)$ ，它的最前及最後兩點，我們存的是瞬間相位之值，一者是因為在前後邊界上不好作微分的計算，另一者是思考邊界上的瞬間相位，是作瞬間頻率曲線合成的重要資訊。

3.3 瞬間頻率之合成

在 3.2 節我們已經把分析出的抖音參數作取樣與存檔，這裡就考慮如何依據取樣過的資料去合成出瞬間頻率曲線？並且比較合成的和原始的瞬間頻率曲線之間會有多大的差異？如果能夠將瞬間頻率曲線合成出來，將來也可使用相同作法來對 ANN 模型輸出的抖音參數作瞬間頻率曲線的合成。有了瞬間頻率曲線，就可把它帶入歌聲信號合成程式，去合成出具有人聲之自然抖音特性的歌唱聲，而可呈現出人聲歌聲的重要表情。

在此我們考慮了線性內插的方式，來對兩個相鄰的抖音參數取樣值作內差，以計算此二個取樣值所管轄的瞬間頻率曲線樣本點上的抖音參數數值，實作上是使用下列的同一個公式來分別對三種抖音參數的數值作計算，

$$VQ(i \cdot N + k) = VP(i) + \frac{VP(i+1) - VP(i)}{N} \cdot k, \quad (8)$$

$$k = 0, 1, 2, \dots, N - 1$$

其中 $VP(i)$ 表示原始抖音參數曲線上的第 i 個取樣值； N 表示合成的抖音參數曲線上 $VP(i)$ 和 $VP(i+1)$ 所共同管轄的樣本點數，其數值是總樣本點數除以 127；而 $VQ(j)$ 表示合成的抖音參數曲線之第 j 個樣本點上的數值。音位軌跡 $VD(t)$ 和抖動範圍 $VE(t)$ 兩者的合成曲線，可直接使用公式(8)來計算；不過抖動頻率 $VR(t)$ 合成曲線的計算，由於原始曲線取樣時最前和最後兩點記錄的是瞬間相位 ϕ_1 和 ϕ_{128} ，因此我們以第 2 和 127 點的取樣值來分別替代第 1 和 128 點的，如此就仍然可使用公式(8)來計算 $VR(t)$ 的合成曲線。

在計算合成的瞬間頻率曲線 $f(t)$ 之前，我們先以如下公式

$$\phi(j) = \phi(j-1) + 2\pi \cdot VR(j) / f_s \quad (9)$$

計算出合成的瞬間相位曲線 $\phi(t)$ ，在此令 $\phi(0) = \phi_1$ ，然後以下列公式(10)

$$f(j) = VD(j) + VE(j) \cdot \cos(\phi(j)) \quad (10)$$

來計算出合成的瞬間頻率曲線。

當我們以前面圖 12 /wei/音和圖 13 /yi/音之瞬間頻率曲線分析出的抖音參數取樣值，來分別作合成，結果得到如圖 22 和 23 裡綠色合成的瞬間頻率曲線，而紅色表示原始的瞬間頻率曲線。分別比較圖 22 和 23 裡的兩條曲線，可以發現在大部分時間，綠、紅兩條曲線幾乎是重疊的，除了在左右兩端會有一點的差異存在。

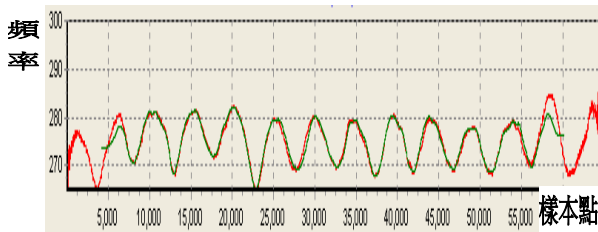


圖 22. /wei/音之合成的瞬間頻率曲線

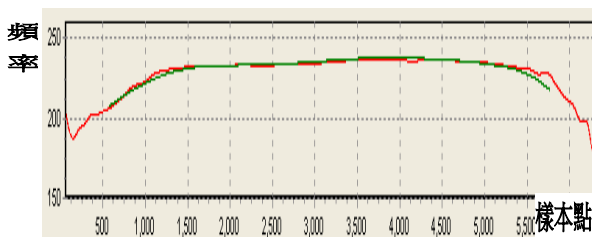


圖 23. /yi/音之合成的瞬間頻率曲線

4. 結論

由於人類歌聲中，抖音是最常被使用來表達情感的一種歌唱技巧，因此研究分析人聲歌聲中的抖音特性，並且將抖音的特性參數擷取出來，在電腦歌聲合成的研究上，是一項重要的工作。在本論文裡，我們回顧了兩類的抖音參數之分析方法，即基週峰谷法和瞬間頻率法，然後實際製作分析程式來對所錄音的歌聲音節信號作實驗驗證，並且從實驗中，發現了前人方法在施行上的缺點，因此我們也研究、提出了幾點改進的方法。

當以基週峰谷法分析抖音不明顯的音節時，所得的抖音參數數量在時間上不夠精細，而難以用於訓練抖音模型。另外，當採取瞬間頻率法作抖音分析時，我們發現音位軌跡的估計，前人的低通濾波作法，對於抖音不明顯的音節來說是有缺憾的，因此我們研究了移動平均之濾波作法，其效果不錯；再者，關於抖音頻率、範圍的估計，前人方法用於抖音不明顯的音節時，也是會出現問題(如抖音頻率太高)，因此我們研究了改進作法，就是讓瞬間頻率曲線先通過一個低通的 IIR 濾波器，如此可讓抖音頻率的數值歸於正常。分析出抖音的音位軌跡、抖動範圍、抖動頻率等參數之後，我們也研究了參數曲線的取樣、和生成的實作

方法，從實驗的結果來看，雖然我們只是使用簡單的線性內差方法，但是生成的和原始的瞬間頻率曲線，在大部分時間幾乎是重疊的，亦即十分相像。未來我們將再研究抖音模型的訓練、和產生抖音參數的問題，目前預計選擇的模型架構是 ANN，當抖音模型建造完成後，就能夠用以產生出瞬間頻率曲線，然後帶入歌聲信號合成程式，去合成出具有人聲之自然抖音特性的歌唱聲，這樣就能夠呈現出人聲歌聲的重要表情。

參考文獻

- [1] 古鴻炎、陳安璿、廖皇量，「整合MIDI伴奏之國語歌聲合成系統」，WOCMAT 2005 電腦音樂與音訊技術研討會(台北)，Session B，2005。
- [2] 古鴻炎、廖皇量，「用於國語歌聲合成之諧波加噪音模型的改進研究」，WOCMAT 2006 國際電腦音樂與音訊技術研討會(台北)，session 2 (音訊處理I)，2006。
- [3] G. Grindlay and D. Helmbold, "Modeling, analyzing, and synthesizing expressive piano performance with graphical models", Springer Netherlands, Vol. 65, pp. 361-387, Dec. 2006.
- [4] C. E. Seashore, "The natural history of the vibrato", Proc. Natl. Acad. Sci. U.S.A, Vol. 17, pp. 623-626, 1931.
- [5] J. Sundberg, "Effects of the vibrato and the singing formant on pitch", Musicologica Slovaca, Vol. 6, pp. 51-59, 1978.
- [6] J. I. Shonle and K. E. Horan, "The pitch of vibrato tones", J. Acoust. Soc. Am., Vol. 67, pp. 246-252, 1980.
- [7] J. C. Brown and K. V. Vaughn, "Pitch center of stringed instrument vibrato tones", J. Acoust. Soc. Am., Vol. 100, pp. 1728-1735, 1996.
- [8] E. Prame, "Measurements of the vibrato rate of ten singers", J. Acoust. Soc. Am., Vol. 96, pp. 1979-1984, 1994.
- [9] E. Prame, "Vibrato extent and intonation in professional western lyric singing", J. Acoust. Soc. Am., Vol. 102, pp. 616-621, 1997.
- [10] K. Kato, et al., "Blending vocal music with the sound field - the effective duration of autocorrelation function of western professional singing voices with different vowels and pitches", International Symposium on Musical Acoustics (ISMA2004), Nara, Japan, 2004.
- [11] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal, Part I: Fundamentals", Proceedings of the IEEE, Vol. 80, pp. 519-538, April 1992.
- [12] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. Part 2: Algorithms and applications", Proceedings of the IEEE, Vol. 80, pp. 539-568, April 1992.
- [13] I. Arroabarren, et al., "Measurement of vibrato in lyric singers", IEEE instrumentation and

measurement technology conference, pp. 1529-1534, 2001.

- [14] P. Howes, et al., "The relationship between measured vibrato characteristics and perception in western operatic singing", *Journal of Voice*, Vol. 18, pp. 216-230, 1997.
- [15] J. Schoukens, R. Pintelon, and H. Van Hamme, "The interpolated fast fourier transform: A comparative study", *IEEE trans. Instrum. Meas.*, Vol. 41, pp. 226-232, April 1992.
- [16] H. G. Feichtinger and T. Strohmer, *Gabor analysis and algorithms theory and applications*, Birkhauser, Boston, Dec. 1997.
- [17] D. G. Long, "Comments on Hilbert transform based signal analysis", *Microwave Earth Sensing (MERS) Laboratory*, Feb. 2004.
- [18] M. Johansson, "The Hilbert transform", Math. Dept., Växjö Universitet, Sweden, <http://w3.msi.vxu.se/exarb/>.
- [19] 古鴻炎、張小芬、吳俊欣，「仿趙氏音高尺度之基週軌跡正規化方法及其應用」，第十六屆自然語言與語音處理研討會(ROCLING XVI)，台北，第325-334頁，2004。
- [20] Hideo Suzuki, et al., "Instantaneous frequencies of signals obtained by the analytic signal method", *Acoust. Sci. & Tech*, Vol. 27, pp. 163-170, 2006.
- [21] T. Wakayama, et al., "Comparison of violin vibratos among four virtuost", *Proceedings of the International Symposium on Musical Acoustics (ISMA2004)*, Nara, Japan
- [22] C. Langton, "Hilbert transform, analytic signal and the complex envelope", *Loral Space Systems*, <http://www.complextoreal.com/tcomplex.htm>.