

# 國語雙字語詞聲調評分系統

## A Scoring System for Mandarin Tones Uttered in Disyllabic Words

古鴻炎#      孫世諺#      張小芬\*  
Hung-Yan Gu#      Shih-Yan Sun#      Hsiao-Fen Chang\*

#國立台灣科技大學資工系      \*國立台灣海洋大學  
#National Taiwan University of Science and Technology, \*National Taiwan Ocean University  
e-mail: { guhy@mail.ntust.edu.tw, joanne@ntou.edu.tw }  
<http://www.csie.ntust.edu.tw/>

### 摘要

本文對國語雙字詞的聲調發音，研究了一種可行的、語者無關的聲調評分方法，並完成可實際操作的評分系統。評分的處理分為四個步驟，首先作基週偵測；接著作基週軌跡預處理，以更正錯誤的軌跡點，及對基週軌跡作時間與音高的正規化；再者是以距離量測作樣式比對；最後是聲調評分決策，輸出之分數以五等第計分，由高至低為 5 至 1 分。經由測試實驗發現，程式評分和人工評分之間的分數誤差的平均值，已可小於 0.5 分。

關鍵詞：基週偵測，基週軌跡，音高正規化，聲調評分

### 1. 前言

國語是一個聲調語言，會因為聲調的不同而影響語意。不過許多的外國語言，並未以聲調來分辨語意，因此一般外國人在學習國語時，常會因自己母語的說話習慣，而忽略了國語的聲調。因此，外國人和聽障人士學習國語時，正確的聲調發音的學習，是非常重要的[1, 2, 3]。

學習國語的聲調，一般是在課堂上由老師發音帶著學生唸，然後老師再根據學生的發音情況去指導修正。現在隨著資訊科技的進步，語言學習可以透過更多樣的形式來進行，我們期望有一個互動的機制，使國語聲調的學習，可以更為方便，因此我們就著手研究國語聲調的評分方法，及製作可實際操作的電腦聲調評分系統。

關於國語或閩南語聲調之電腦評量的研究，過去已有一些研究報告被提出[4, 5, 6, 7]。有些研究是以單字的聲調發音來作評量 [4, 5]，但單字發音之方式，和實際溝通時之連續語音的發音方式不同，聲調的特性已經改變很多，例如在唸單字的第三聲時，會發出全上聲(先降再升)，而在語詞或語句裡，第三聲大多只發

前半上，再者兩個都是第四聲的語詞(如”重要”)，第一個四聲的音調下降幅度會明顯地減少。另外，有些研究是以語句整體的聲調發音來作綜合評量[6, 7]，而不管各個組成字的細部聲調發音，這樣的評量方式，我們又覺得太過於籠統。在本研究裡，除了要兼顧語句裡的聲調發音特性，也希望針對各個字作較細部的聲調評量，因此我們選擇以雙字語詞作為聲調發音練習的單位，並且對兩個組成字分別去作聲調發音的評分。

本研究的聲調評分方法，主要的處理流程如圖 1 所示，第一個方塊是”基週偵測”，把輸入的語音信號(取樣率 22,050Hz)切成一序列的音框(frame)，音框長度為 25ms，且每次前進 8ms，然後對各個音框去計算自相關係數和 AMDF 係數 [8]，再依這兩種係數的比值去決定出基週的週期長度，較詳細的決定方法可參考我們先前的論文[9]或 Kim, H. Y.等人的論文[10]。圖 1 裡的第二個方塊是”基週軌跡預處理”，預處理的主要工作包含了：(a)倍頻與半頻之頻率更正、(b)連音之分段、(c)時間與音高之正規化等，較詳細的處理方法將在第二節裡說明。圖 1 裡的第三個方塊是”樣式比對”，把事先準備的雙字詞參考發音的基週軌跡拿出，和新輸入的發音的基週軌跡作比對，由於已經作過時間的正規化，所以這裡的比對只是依據特定的距離量測方法來計算距離，較詳細的說明在第三節裡。圖 1 裡的第四個方塊是”評分決策”，這裡所使用的決策方法，並不是語音辨識裡常用的 NN 或 KNN[11]方法，因為分數是可以作加減計算的，所以我們的決策方法都牽涉到”平均”的觀念，詳細的說明在第四節裡。

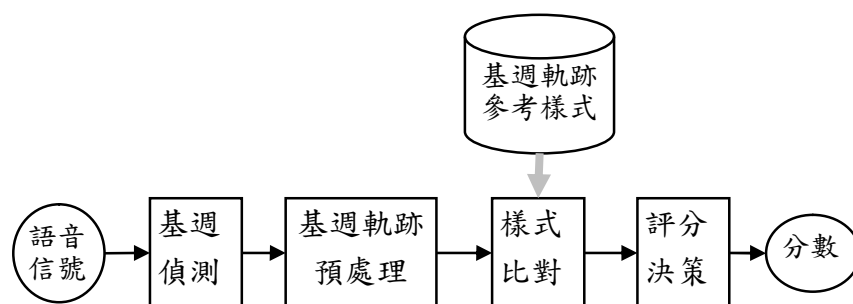


圖 1 聲調評分之主要處理流程

除了研究、探討評分的方法，我們也已經把圖 1 所示的處理流程，製作成一個可實際操作的國語雙字詞聲調的評分系統，系統的畫面如圖 2 所示。使用者可以從圖 2 右上角方塊內點選欲作練習的課程及聲調組合，然後預錄好的老師示範發音之基週軌跡就會被立即偵測出來，並顯示於圖 2 左上方塊內；此時使用者可以按錄音按鈕來練習發音，而其基週軌跡也會被立即偵測並顯示於圖 2 左邊中間的方塊內，如此就可以比較自己的和老師的基週軌跡，在趨勢(向上或向下)及高低的位置是否有差別，也就是經由視覺回饋來幫助學習。進一步使用者也可按圖

2 右下之評分按鈕，讓電腦來作評分，圖 2 裡顯示的是，老師示範的”葡萄”之發音，經電腦評分後分別都得到 5 分，電腦的評分範圍是 1 至 5 分。

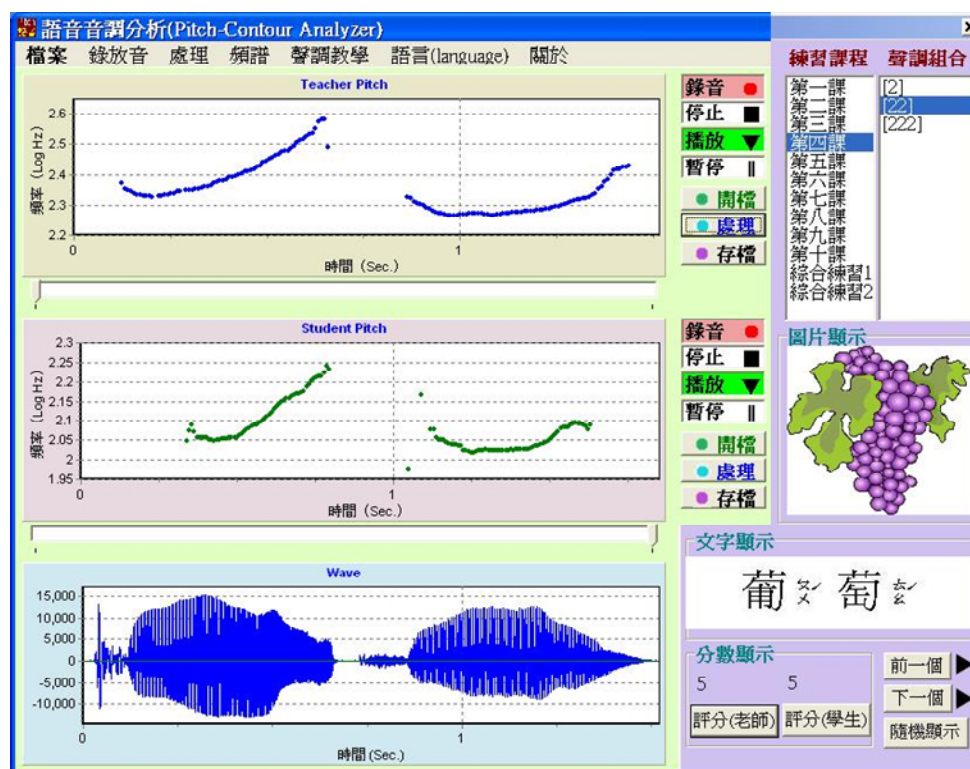


圖 2 國語雙字詞聲調評分系統之畫面

## 2. 基週軌跡預處理

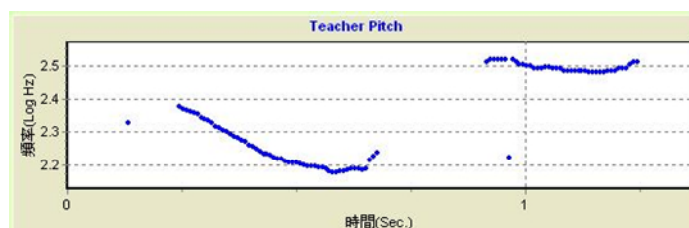
在圖 1 的基週偵測方塊，並不保證對每個音框都可以偵測到正確的基頻值，例如偵測到真正基頻的倍頻、半頻，或未偵測到，這裡基頻值是以 10 為底之 log Hz 尺度來表示。再者，有些雙字語詞的發音會因為 coarticulation 而發生連音的情況，而得到一長條連續的基週軌跡，這就需要作軌跡切割分段的處理。此外，我們希望讓男生、女生之不同使用者，都能來操作我們的評分系統，因此來自不同人、不同音高的基週軌跡，必需先作音高的正規化處理。詳細來看，我們的基週軌跡預處理會依序執行如下的處理項目：(a)音高範圍決定，(b)倍頻、半頻更正，(c)中值法平滑處理，(d)軌跡段落尋找與分段，(e)時間正規化，(f)音高正規化。以下就對各個處理項目作介紹，不過中值法平滑處理[8]，由於是常見的處理方法，我們就省略了，在此使用的是 5 點的中值平滑處理(medium smoothing)。

### 2.1 音高範圍決定

由於系統並不知道使用者的特性(音調高低的範圍)，所以一開始會把音高值的上、下限設定得較保守，如此畫出來的軌跡圖(以”草蝦”發音為例)，就會如圖 3(a)所示，顯得平平扁扁的，因此我們設計了音高範圍的動態估計方法，當依據估計的結果來設定音高範圍的上、下限值，就可得到如圖 3(b)所示的對比較強烈、較有視覺效果的軌跡圖。除了視覺效果的增進，動態方式決定音高範圍，還可用來濾除一些基週偵測錯誤的雜點，如圖 3(a)左邊散佈的點。



3(a) 固定的音高範圍



3(b) 動態決定音高範圍

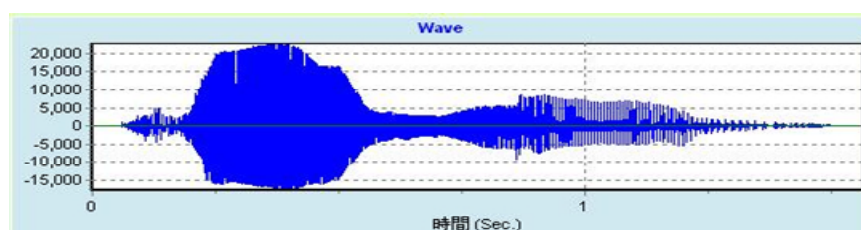
圖 3 音高範圍設定對基週軌跡圖的影響

本文的音高範圍動態估計方法是，基於 histogram 的觀念，把縱軸(log10 Hz 頻率)平均切成寬度為 0.05 之一序列的區間，然後對各個音框依其基頻值決定所應掉落的區間，接著統計各個區間所收到的音框個數。假設第  $n$  個區間收到最多的音框數，則接著從  $n$  往上( $n+1, n+2, \dots$ )檢查，直到發現有連續三個區間都沒有收到音框，就將此時的三個區間之中間區間所對應的頻率值，設定為音高值的上限；依此程序從  $n$  往下( $n-1, n-2, \dots$ )檢查，也可得到音高值的下限。

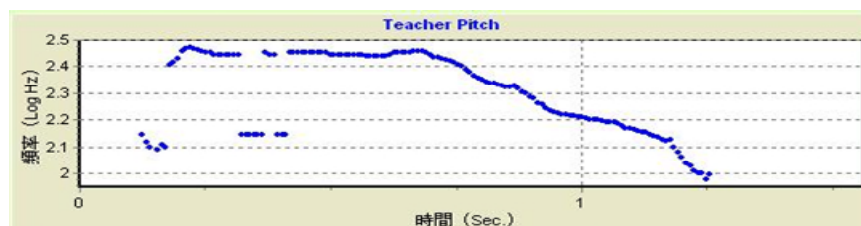
## 2.2 倍頻、半頻更正

圖 1 裡的基週偵測方塊，可能在某一時刻附近的連續數個音框，都偵測到真正頻率的半頻值，如圖 4(a)是”充滿”發音的信號波形，而對應的基週軌跡曲線如 4(b)所示，軌跡曲線有兩個片段發生半頻值之錯誤，前一片段裡有連續 6 點，而後一片段裡有連續 3 點，這樣的錯誤已經無法由下一個處理項目”中值法平滑化”來作更正，並且”充滿”的基週軌跡發生了連音現象，若不謹慎更正半頻值錯誤，則很難正確完成之後第二個處理項目”基週軌跡的分段”，因此我們設計了倍頻、半頻的更正處理方法，執行該方法後，就會得到如圖 4(c)所示的軌跡曲線，即前

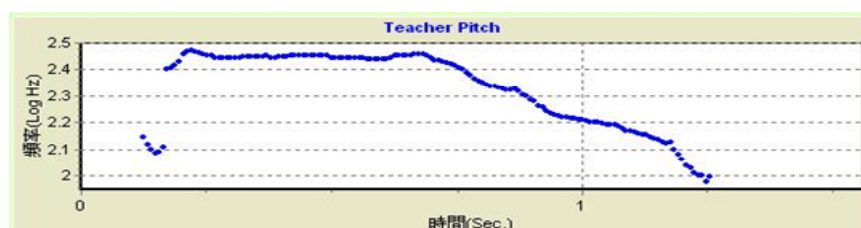
述的兩處半頻值錯誤已被更正了。



4(a) “充滿”的信號波形



4(b) 有半頻錯誤的軌跡曲線



4(c) 更正半頻錯誤後的軌跡曲線

圖 4 有、無半頻錯誤之基週軌跡比較

本文的倍頻、半頻更正方法是，以多數的正確點的力量迫使少數的錯誤點移動到正確位置，實作上則是使用動態規劃的方法。在此令  $p(t,1)$  存第  $t$  個音框的基週偵測結果(log10 Hz 尺度)， $p(t,0)$  存它的半頻值且  $p(t, 2)$  存它的倍頻值，也就是  $p(t,0) = p(t,1) - \log_{10}2$ ， $p(t,2) = p(t,1) + \log_{10}2$ ，而更正方法就是，先計算遞迴公式(1)，

$$D(t, i) = \min_{0 \leq j \leq 2} \left[ D(t-1, j) + (p(t, i) - p(t-1, j))^2 \cdot 100 \right] + c(i) \quad (1)$$

$$c(i) = \begin{cases} 0, & \text{if } i = 1 \\ 1, & \text{if } i = 0 \text{ or } i = 2 \end{cases}$$

依音框編號  $t$  由小至大計算，到達最後之音框編號  $T$  時，再從三種  $i$  值的  $D(T, i)$  中選出最小者，作 back-tracking 來找出最佳路徑。此公式裡， $D(t, i)$  表示從起始點走到音框  $t$ 、音高代號  $i$  之最小累積距離， $c(i)$  表示行走半頻、或倍頻點的懲罰距離(penalty)。

## 2.3 軌跡片段尋找及分段

一個雙字詞的基週軌跡，理想上應該有兩個主要的片段(segment)，再加上一些長度較短的小片段，例如圖 5 是”回去”的波形及其基週軌跡，除了兩個主要片段之外，還存在許多較短的片段，其中被圈起來的是較長者。一個片落的認定是，片段中時間上相鄰的三個軌跡點，設它們的頻率值為  $x, y, z$ ，則  $|(z-y) - (y-x)| = |z-2y+x|$  要小於一個門檻值，即斜率不可突然大幅度改變，在此設定的門檻值是 0.05。要找出主要的片段，我們首先依時間順序檢查各音框的基頻值，找出各個片段的起點及它的長度，再將所找到的片段按照時間長度作排序，然後就可把兩個最長的片段拿出，作為雙字詞的兩個組成字的基週軌跡。

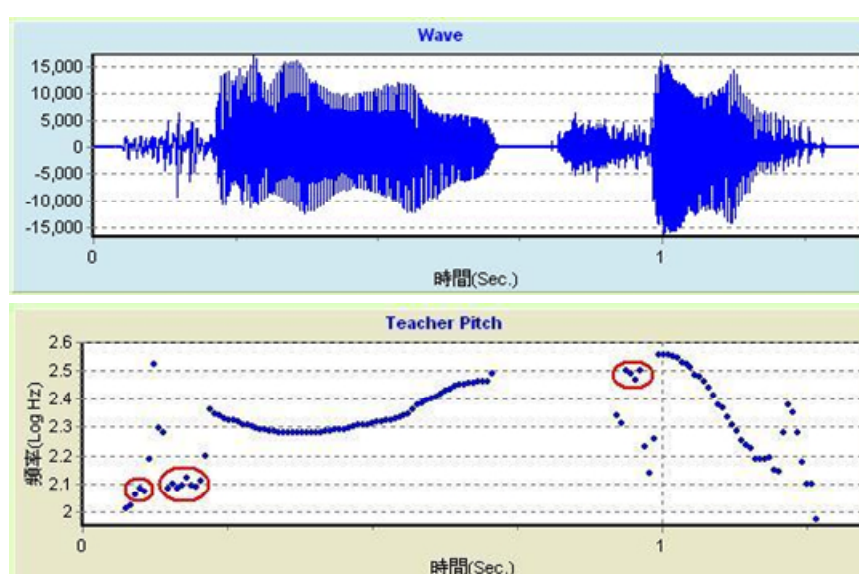


圖 5 “回去”的信號波形、及其基週軌跡

當一個雙字詞的兩個字之間發生連音現象時，如圖 4 顯示的波形和基週軌跡，則時間長度第一與第二的兩個片段，它們的長度差異會很大，例如第一長的超過 400ms 而第二長的卻小於 80ms，此時我們就要捨棄第二長的，而將第一長的軌跡片段作分段的處理。我們的分段方法是，先計算各音框的短時能量值，令其中最大能量值之 rms 值為  $E_x$ ，接著設定  $E_x/6$  為 rms 能量值的門檻，去檢查能量曲線，找出能量值低於門檻、並且落於第一長基週軌跡片段之時間範圍內的一個最長的時間片段，若找到了，則依此低能量的片段去對基週軌跡曲線作分段，若不能找到低能量的片段(表示能量曲線起伏不大)，就把門檻升高到  $E_x/5$ 、 $E_x/4$ ，再重複作前述的動作。經由這裡的分段處理後，圖 4(c)裡一長條的基週軌跡曲線，就可切成如圖 6 所示的兩段的紅點片段了。

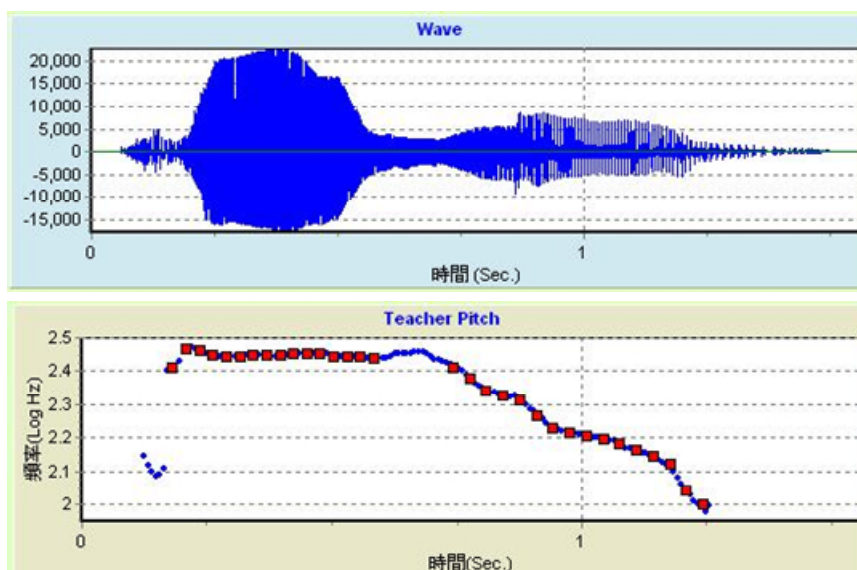


圖 6 基週軌跡分段處理

## 2.4 時間、音高正規化

一個音節的時間長度，會因為說話者說話速度的不同，或同一個人不同次說話之速度的不同，而使得音長有長有短。為了方便後續的處理，在此我們先作時間正規化之處理，將時間長短不一的音節基週軌跡，一律轉換成 16 維度的頻率向量來表示。作法是在一個音節的時間範圍裡均勻放置 16 個音高取樣點，然後以 Lagrange 內插法來求出各取樣點上的音高值[9]。一個時間正規化的例子，如圖 6 所示。

求出一個雙字詞兩音節共 32 點的時間正規化之基週軌跡  $g_0, g_1, \dots, g_{15}, g_{16}, g_{17}, \dots, g_{31}$  後，接著需考慮音高正規化的問題，因為不同使用者之間的音高位準差異是很大的，尤其是男、女生之間的音高差異。在此我們考慮了兩種正規化方法，第一種稱為音節關聯式作法，先求出一個雙字詞發音的平均音高值，公式如下：

$$\mu = \frac{1}{32} \sum_{j=0}^{31} g_j \quad (2)$$

再將基週軌跡每一點的音高值減去平均值，即令  $\alpha_j = g_j - \mu$ ，而求得音高正規化後的基週軌跡  $\alpha_0, \alpha_1, \dots, \alpha_{15}, \alpha_{16}, \alpha_{17}, \dots, \alpha_{31}$ 。另外一種音高正規化的方法，稱為音節獨立式作法，先計算前後音節各自的平均音高值  $\mu_1$  和  $\mu_2$ ，接著前後音節的音高值分別減去各自的平均值，即  $\beta_j = g_j - \mu_1, 0 \leq j \leq 15; \beta_j = g_j - \mu_2, 16 \leq j \leq 31$ ，如此可求得另一種音高正規化的基週軌跡。

### 3. 樣式比對

#### 3.1 語料準備

在基週軌跡的參考樣式的準備上，需要有標準發音與不標準的發音，這樣才能夠應付各種可能的使用者。因此我們收集了以國語為母語者所唸的標準音，與外國人、聽障學生唸的非標準發音。參與錄音的語者共 28 人，其中 19 個人的發音作為參考樣式，包括標準音樣本 4 位，外國人 13 位，聽障者 2 位。剩下的 9 個人的發音則作為測試語料，包含標準音樣本 3 位，外國人 6 位。標準音樣本是請研究室裡發音標準的研究生，在實驗室裡錄製，外國人的語音樣本是在國際教會錄得，而聽障生的發音是在國小的啓聰班錄得。

雙字詞發音所用的詞彙包含：冰棒、熨斗、帆船、肥皂、小熊、草地、崖谷、愛心、卡車、燈塔、西瓜、圍巾、照片、拼圖、氣球等共 15 個，含蓋了 15 種雙字詞的聲調組合，每個詞彙中，第二個字都含有子音聲母，以避免發生連音現象，而造成參考音的基週軌跡樣式的求取錯誤。

#### 3.2 人工評分

對於這些錄到的雙字詞發音，首先要對各個雙字詞的兩個組成字分別作人工評分的工作，有了參考分數之後，才能夠對不同的評分模式作測試及評估。這裡的評分方式是，分成五個等第，最高為 5 分，最低為 1 分，其於是 4、3、2 分。我們希望每一個字音所得到的分數，都能夠符合一般人耳的聽覺，因此將一般人所唸的聲調發音定為 5 分，而將唸的比較差的聽障學生的聲調發音定為 1 分，以此為標準，將外國人所唸的發音拿去作人工評分，分數則介於 1 分到 5 分之間。

至於評分者的人選，係由五位實驗室的研究生來擔任，在評分前先聽過正常人與聽障學生的發音，依此為標準才開始作評分。五位評分者對於同一個字音所給予的分數，我們以多數決來確定最終分數，遇到分數不能決定時，則取五個人評分的中間值作為客觀分數。

#### 3.3 距離量測

由於雙字詞不論是參考音或是測試音的基週軌跡，都已經先作了時間正規化和音高正規化的處理，使得一個音節的基週軌跡都是固定地表示成 16 維度的頻率向量，因此圖 1 裡的樣式比對方塊，其實只是作頻率向量之間的距離量測。

距離量測的方法中，最常被使用的是幾何距離，此外還有一個常見的是曼哈



頓距離，令  $X = \langle x_0, x_1, \dots, x_{15} \rangle$ ,  $Y = \langle y_0, y_1, \dots, y_{15} \rangle$  是兩個頻率向量，則幾何距離與曼哈頓距離的定義分別為：

$$d_e(X, Y) = \sqrt{\sum_{i=0}^{15} (x_i - y_i)^2} \quad (3)$$

$$d_m(X, Y) = \sum_{i=0}^{15} |x_i - y_i| \quad (4)$$

本文研究了這兩種距離量測，觀察它們對於評分正確性的影響。

#### 4. 評分決策

在語音辨識的應用裡，一般的觀念是直接把”距離”和”分數”連結在一起，直覺認為兩者是成正比的關係。但是在本研究裡，距離和分數並不是直接相關的，距離值只是作為對參考樣式進行排序的依據，而參考樣式的人工評分數值才是作為評分決策的直接依據。

由於每一種聲調組合的雙字詞的參考發音(19 個)，都已經過人工工作評分，所以各個參考發音的兩個組成音節已有參考的分數。我們的評分方法是，將測試的雙字詞發音的前後音節之基週軌跡，和參考發音前後音節的基週軌跡分別作距離量測，再依距離值由小到大作排序，距離越小表示相似的程度越高，然後依據排名在前面的參考發音音節對應的人工評分，使用某一種評分決策方法來決定分數值。

關於評分決策的方法，本文嘗試了二種作法，第一種稱為”直接平均法”，就是直接取前 M 名的分數來計算出平均值，第二種稱為”中間平均法”，先從前 M+2 名的分數中排除最大與最小的分數值，再計算剩下的分數的平均值。

以如下的例子來說，假設測試的雙字詞發音是”熨斗”，它的前後音節和各個參考發音的前後音節分別作距離量測後，再依距離值對前、後音節分別作排序而得到如圖 7 所示的排名次序。如果評分決策方法是**直接平均法**，則”熨”的分數值，當 M 為 3 時是  $(5+4+3)/3 = 4$ ；而”斗”的分數值，當 M 為 2 時是  $(3+5)/2 = 4$ ，依此可類推其它 M 值時的分數。如果分數決策方法是**中間平均法**，則”熨”的評分方式是，當 M 為 3 時取距離排名前五名的分數值 5、4、3、5、3，然後刪去最大分數值 5 與最小分數值 3，再取剩下三個分數的平均而得到 4；”斗”的評分方式是，當 M 為 2 時，取距離排名前四名的分數值 3、5、3、3，然後刪去最大分數值 5 與最小分數值 3，再取剩下二個分數的平均而得到 3，依此可類推其它

M 值時的作法。

語者	前音	距離	分數
no2.	熨	0.013	5 分
no1.	熨	0.014	4 分
no3.	熨	0.016	3 分
no5.	熨	0.025	5 分
no4.	熨	0.028	3 分
no6.	熨	0.034	3 分
⋮			

語者	後音	距離	分數
no7.	斗	0.011	3 分
no5.	斗	0.015	5 分
no1.	斗	0.026	3 分
no3.	斗	0.031	3 分
no2.	斗	0.037	2 分
no8.	斗	0.042	4 分
⋮			

圖 7 前、後音節分別依距離值作排序

## 5. 評分實驗

### 5.1 音高正規化方法之比較

首先我們就 2.4 節提到的關聯式與獨立式之音高正規化方法，作評分效果的實驗，評分效果好壞的比較，我們以誤差分數的平均值 AVG，誤差分數的標準差 STD，和最大誤差分數 MAX 等三個統計項目來觀察。關於其它條件，距離量測先採取幾何距離，而評分決策方法先採取 M 值為 3 之中間平均法。

實驗後，我們得到如表 1 和 2 所示的數值。觀察誤差平均值、誤差標準差、和最大誤差分數等項目在 ALL (四種聲調一起計算)欄下，可發現關聯式的都比獨立式的要好很多，此外就各個聲調分別來看，也是關聯式的都比獨立式的要好，因此以後的評分實驗裡，我們就都採取關聯式的音高正規化方法。

表 1. 關聯式音高正規化方法之評分誤差

	一聲	二聲	三聲	四聲	ALL
AVG	0.462	0.550	0.771	0.462	0.548
STD	0.383	0.462	0.539	0.442	0.469
MAX	1.333	2.000	2.333	1.666	2.333

表 2. 獨立式音高正規化方法之評分誤差

	一聲	二聲	三聲	四聲	ALL
AVG	0.699	0.638	0.993	0.583	0.711
STD	0.709	0.563	0.807	0.443	0.650
MAX	3.000	2.666	3.666	1.666	3.666

## 5.2 距離量測之比較

除了幾何距離，在 3.3 節裡還提到了曼哈頓距離，因此這裡就對這兩種距離量測作評分效果的比較，關於評分決策的方法，我們先採取 M 值為 3 之中間平均法。實驗後，我們得到幾何距離的評分誤差情況如表 1 所示，而曼哈頓距離的評分誤差情況則如表 3 所示。

表 3. 曼哈頓距離之評分誤差

	一聲	二聲	三聲	四聲	ALL
AVG	0.444	0.555	0.641	0.486	0.524
STD	0.376	0.503	0.479	0.457	0.460
MAX	1.666	2.000	2.000	1.666	2.000

比較表 1 和表 3 的誤差分數值，可發現第二、四聲時，幾何距離的誤差比曼哈頓距離的小一些，而第一、三聲時則顛倒過來；若就整體來看，即看 ALL 欄，曼哈頓距離在 AVG, STD, MAX 等三個項目，則都比幾何距離的好一些。

## 5.3 評分決策方法之比較

在第四節裡提到了兩種評分決策方法，即直接平均法和中間平均法，因此這裡就對這兩種決策方法來作比較。每一種決策方法的實驗裡，我們分別對兩種距離量測作實驗，然後把四個聲調的評分誤差放在一起計算出誤差的兩種統計值 AVG 和 STD。此外當使用不同的決策人數、不同的 M 值時，也會得到不同的評分誤差統計數值，所以 M 也是一個需考慮的因素。

實驗後，對於兩種決策方法我們分別得到如表 4 和 5 的數值。在表 4 之直接平均法裡，最小的 AVG 值是 0.514，它是使用曼哈頓距離和決策人數設為 1 時得到的，次小的是 0.518，則是使用幾何距離和決策人數設為 2 時得到的，所以使用直接平均法時，決策人數 M 的選擇，會和所使用的距離量測有關係，在目前參考發音為 19 個的情況下，M 值超過 2 後的評分誤差 AVG 值會持續變大、變差。在表 5 之中間平均法裡，最小的 AVG 值是 0.487，它是使用曼哈頓距離和決策人數設為 2 時得到的，次小的是 0.490，則是使用幾何距離和決策人數設為 2 時得到的，所以決策人數 M 的選擇，對於兩種距離量測來說，都是設定為 2 是

最好的，M 值小於 2 或多於 2 都會使評分誤差 AVG 值變大、變差。如果就這兩種決策方法作比較，中間平均法的評分誤差 AVG 值 0.487，要比直接平均法的 0.514 好 5.3%；此外曼哈頓距離得到的評分誤差 AVG 值，都可以比幾何距離的稍小一些；至於 STD 值，一般來說會隨 M 值的增加而減小。

表 4. 直接平均法之評分誤差

		決策人數				
		1	2	3	4	5
幾何距離	AVG	0.548	0.518	0.543	0.566	0.634
	STD	0.707	0.542	0.474	0.454	0.436
曼哈頓距離	AVG	0.514	0.537	0.535	0.578	0.623
	STD	0.681	0.527	0.457	0.458	0.426

表 5. 中間平均法之評分誤差

		決策人數				
		1	2	3	4	5
幾何距離	AVG	0.529	0.490	0.548	0.640	0.698
	STD	0.623	0.518	0.469	0.426	0.437
曼哈頓距離	AVG	0.522	0.487	0.528	0.649	0.712
	STD	0.585	0.528	0.460	0.439	0.456

## 6. 結論

國語聲調的評量，尤其是多字語詞之組成字的聲調評量，電腦的評分必需接近人耳的評分，如此，外國人及聽障人士才能有效的使用電腦來學習國語。本文研究、製作了一個國語雙字詞聲調之評分系統，在基週軌跡預處理方面，研究了幾個改進基週軌跡正確性的方法，例如研究了基於 histogram 之音高範圍決定方法，以增進視覺效果；提出了基於動態規劃之倍頻、半頻的更正方法；研究了基於短時能量之基週軌跡分段方法。此外，我們也研究、實驗了不同的音高正規化方法，不同的距離量測方法，不同的評分決策方法，來觀察這些因素對於評分誤差大小的影響，實驗結果顯示，當使用關聯式音高正規化、曼哈頓距離量測、中間平均法之決策方法時，可以得到最小的評分誤差平均值 0.487 分，這相當於滿分 5 分的 9.7%。

目前所使用的測試語料只有 9 個人，還不是很充分，未來可以再錄製更多的語料來作進一步測試。此外，還有一些相關的問題，也可在未來作進一步考慮，例如音域範圍的問題，每個人的音域寬窄不一，並不是本文的音高正規化處理可以直接解決；還有每個人的講話方式(style)，也會有不小的差異，在國語雙字詞的基週軌跡上，就是兩字之間的相對的軌跡高度、斜率的變異問題。雖然本文並未直接去解決這樣的問題，但是我們採取的是間接的方式，來緩和此類問題的影響，那就是收集很多人的雙字詞發音來作為參考樣式，因此目前收集的 19 人的參考發音，也許尚不足夠，還需收集更多人的。

## 致謝

感謝國科會計畫的支援，計畫編號 NSC 94-2614-S-019-001

## 參考文獻

- [1] 金東垠，韓籍學生華語聲調錯誤分析與教學研究，碩士論文，國立臺灣師範大學華語文教學研究所，2003。
- [2] 張可家、陳麗美，“日本學生學習華語的聲調偏誤分析：以二字調為例”，第十七屆自然語言與語音處理研討會(台南)，第 125-139 頁，2005。
- [3] 鍾玉梅，「聽障兒童的說話問題」，聽語會刊，第 10 期，第 72-79 頁，1994。
- [4] 梅永人，國語聲調電腦評量模式之研究，碩士論文，國立台中師範學院 教育測驗統計研究所，2000。
- [5] 黃重光，以自組織特徵映射建立國語聲調電腦評量模式之研究，碩士論文，國立台中師範學院 教育測驗統計研究所，2001。
- [6] 李俊毅，語音評分，碩士論文，國立清華大學 資訊工程研究所，2000。
- [7] 蔡岳廷、廖嘉新、呂道誠、呂仁園，“台灣閩南語聲調評分系統評估與研究”，第十七屆自然語言與語音處理研討會(台南)，第 227-237 頁，2005。
- [8] O'Shaughnessy, D., *Speech Communication: Human and Machine*, 2<sup>nd</sup> ed., IEEE Press, 2000.
- [9] 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(台北)，第 325-334 頁，2004。
- [10] Kim, H. Y., *et al.*, “Pitch Detection with Average Magnitude Difference Function Using Adaptive Threshold Algorithm for Estimating Shimmer and Jitter”, Proc. of the 20th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, Vol. 6, pp. 3162 -3164, 1998.
- [11] Rabiner, L. and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, 1993.