

語音合成初探

彭 玉 峰

摘 要

羅馬拼音可以拼造出所有的語言，我們發現當中的五個元音具有週期性，因此嘗試使用微積分中的一階導數與二階導數來計算週期，並求出元音的發音頻率範圍，進而嘗試利用最小週期內的訊號來進行元音的合成。

關鍵詞：元音、導數、頻率、週期、合成。

The First Concept of Synthesis

Yu-Fong Peng

Abstract

A 「speech voice」 is usually defined as an oscillating function of time-domain. It has been known that every time-domain function may be represented by one-dimension numeric vector and each speech voice may be represented by Romanization. This paper presents a numerical method for synthesis the vocal of Romanization. The first derivation and second derivation are used to analyze these speech signals. The Turbo C language is used to code these programs. The numerical method for the synthesis is solved by computing these periods which between a local maximum point and the next. These signals of period are used to synthesis the vocal of Romanization. In this research we known these frequencies of vocal are variation which between 100 Hz and 500Hz.

Keywords : vocal 、 derivation 、 frequency 、 period 、 synthesis.

壹、導 論

我們發現羅馬拼音當中的五個元音在編碼的過程中，編碼資料的重覆性很高，因此我們研判元音的變化具有週期性質（彭玉峰，民國 88 年）。基於「週期函數」具有「連續而且可微分」的數學性質；因此我們利用微積分中的一階導數及二階導數求極值的方法，判定極大值發生的位置，在根據兩兩極大值發生的位置來決定週期，進而計算出五個元音的頻率範圍，最後嘗試利用適當的週期訊號合成出元音。

貳、數學分析方法

一、傅立葉轉換

給一個定義在整個實數域 $\mathbf{R} = (-\infty, \infty)$ 上的函數 $f(x)$ ，定義傅立葉轉換為 $F(w) = F[f(x)] = \int_{-\infty}^{\infty} f(x)e^{-sx} dx$ (2.1)，其中 $s = i\omega$ ； i 為虛數； w 為 $(2\pi \times \text{頻率})$ 同時定義傅立葉反轉換為 $F^{-1}[F[f(x)]] = f(x)$ 。

定理一：(傅立葉轉換基本性質)

設 $|f(x)| \in L(\mathbf{R})$ ； $|g(x)| \in L(\mathbf{R})$ ； a, b 為常數

$$(1) \text{ 線性性質：} F[af(x) + bg(x)] = a F[f(x)] + b F[g(x)]； \quad (2.2)$$

$$(2) \text{ 第一移位定理：} F[f(x+a)] = e^{as} F[f(x)] \quad (2.3)$$

二、捲積

捲積是一種把兩個函數結合在一起產生新函數的合成法；兩個函數 $f(x)$ 與 $g(x)$ 的捲積用 $f(x) * g(x)$ 表示，其數學定義為 $f(x) * g(x) = \int_{\mathbf{R}} f(y)g(x-y)dy$ (2.4)

引理一：設 $f(x)$ 、 $g(x)$ 、 $h(x)$ 為三個函數，則

$$(1) f(x) * g(x) = g(x) * f(x) \quad (2.5)$$

$$(2) [f(x) * g(x)] * h(x) = f(x) * [g(x) * h(x)] \quad (2.6)$$

定理二：(捲積定理) 設 $|f(x)| \in L(R)$ ， $|G(w)| \in L(R)$ ，則

(1) $f(x)$ 有一個有界的傅立葉轉換 $F(w)$ ，

$$G(w) \text{ 有一個有界的傅立葉反轉換 } g(x)； \quad (2.7)$$

$$(2) F(w) G(w) \in L(R)； \quad (2.8)$$

$$(3) F[f(x) * g(x)] = F(w) G(w) \quad (2.9)$$

三、微積分

定理三：設函數 $f(x)$ 在 $[a, b]$ 上連續，在 (a, b) 內可微。

$$(1) \text{ 若在 } (a, b) \text{ 內處處有 } f'(x) > 0, \text{ 則 } f(x) \text{ 在 } [a, b] \text{ 上是嚴格增函數。} \quad (2.10)$$

$$(2) \text{ 若在 } (a, b) \text{ 內處處有 } f'(x) < 0, \text{ 則 } f(x) \text{ 在 } [a, b] \text{ 上是嚴格減函數。} \quad (2.11)$$

$$(3) \text{ 若在 } (a, b) \text{ 內處處有 } f'(x) = 0, \text{ 則 } f(x) \text{ 在 } [a, b] \text{ 上是常數。} \quad (2.12)$$

定理四：設函數 $f(x)$ 在 $[a, b]$ 上連續，在 (a, b) 內可微，且 $a < c < b$ ：

$$(1) \text{ 當 } x < c \text{ 時有 } f'(x) > 0 \text{ 且 } x > c \text{ 時有 } f'(x) < 0, \\ \text{則 } f(x) \text{ 在點 } c \text{ 上取極大值。} \quad (2.13)$$

$$(2) \text{ 當 } x < c \text{ 時有 } f'(x) < 0 \text{ 且 } x > c \text{ 時有 } f'(x) > 0, \\ \text{則 } f(x) \text{ 在點 } c \text{ 上取極小值。} \quad (2.14)$$

定理五：若 $a < c < b$ 且 $f'(c) = 0$ ；設函數 $f(x)$ 在 (a, b) 內具有二階導數。

$$\text{則 (1) 當 } f''(x) \text{ 在 } (a, b) \text{ 內為負值時, } f(x) \text{ 在點 } c \text{ 處取極大值。} \quad (2.15)$$

$$(2) \text{ 當 } f''(x) \text{ 在 } (a, b) \text{ 內為正值時, } f(x) \text{ 在點 } c \text{ 處取極小值。} \quad (2.16)$$

從定理四及定理五中，我們發現在極大值發生前的斜率為正值，且極大值發生的附近有向下凹的性質，所以我們得到以下推論：

推論一：設函數 $f(x)$ 在點 c 具有二階導數，且當 $f'(c) = 0$ ；

若 $f''(c) < 0$ ，函數 $f(x)$ 在點 c 處有極大值。 (2.17)

四、導數與數位濾波

語音訊號本身具有連續的性質，羅馬拼音的元音又判定具有週期性（彭玉峰,1999），所以我們使用一階導數與二階導數來分析這些元音訊號。假設有一連續函數 $f(x)$ 具有二階導數，我們定義以下兩個函數分別是 $P(x) = f(x) - f(x-1)$ (2.18) 以及 $Q(x) = f(x-1) - 2f(x) + f(x+1)$ (2.19)

根據捲積的定義，我們可以將 $P(x)$ 及 $Q(x)$ 重新寫成 $f(x) * \{-1, 1\}$ 及 $f(x) * \{1, -2, 1\}$ 。再從數位濾波的角度而言， $\{-1, 1\}$ 和 $\{1, -2, 1\}$ 的作用可以視為高通濾波器。根據 Gilbert Strang 在 1989 年的論文中提及高通數位濾波器的正交性質：定理六：設 $g(x)$ 為一組高通濾波器且 $G(w)$ 為其傅立葉轉換，若 $|G(w)|^2 + |G(w + \pi)|^2 = 1$ 成立，則 $g(x)$ 為正交。 (2.20)

以下我們將針對 $g_1 = \{-1, 1\}$ 和 $g_2 = \{1, -2, 1\}$ 進行驗算。首先我們利用定理一的式(2.2)和式(2.3)將 g_1 和 g_2 進行傅立葉轉換，

$$\begin{aligned} G_1(w) &= \mathcal{F}\{-1, 1\} = 1 - e^{-s} \\ &= 1 - [\cos(w) - i \times \sin(w)] \\ &= [1 - \cos(w)] + i \times \sin(w) \\ &= \mathcal{F}\{1, -2, 1\} = e^{-s} - 2 + e^s \\ &= [\cos(w) - i \times \sin(w)] - 2 + [\cos(w) + i \times \sin(w)] \\ &= 2\cos(w) - 2 \end{aligned}$$

我們計算 $|G_1(w)|^2 + |G_1(w + \pi)|^2$

$$= [\cos(w) - 1]^2 + [\sin(w)]^2 + [\cos(w + \pi) - 1]^2 + [\sin(w + \pi)]^2$$

$$\begin{aligned}
 &= [\cos^2(w) - 2\cos(w) + 1 + \sin^2(w)] + [\cos(w) + 1]^2 + \sin^2(w) \\
 &= [2 - 2\cos(w)] + [\cos^2(w) + 2\cos(w) + 1 + \sin^2(w)] \\
 &= [2 - 2\cos(w)] + [2 + 2\cos(w)] \\
 &= 4
 \end{aligned}$$

以及

$$\begin{aligned}
 &|G_2(w)|^2 + |G_2(w + \pi)|^2 \\
 &= [2\cos(w) - 2]^2 + [2\cos(w + \pi) - 2]^2 \\
 &= [2\cos(w) - 2]^2 + [-2\cos(w) - 2]^2 \\
 &= [2\cos(w) - 2]^2 + [2\cos(w) + 2]^2 \\
 &= 4\cos^2(w) - 8\cos(w) + 4 + 4\cos^2(w) + 8\cos(w) + 4 \\
 &= 8\cos^2(w) + 8
 \end{aligned}$$

我們可以輕易的修正 $g_1 = \{-0.5, 0.5\}$ 來滿足正交性質，雖然 g_2 為一實數，但是 g_2 仍無法輕易修正來滿足正交性質。幸運的是 $\{1, -2, 1\} = \{-1, 1\} * \{-1, 1\}$ ，從定理二的式(2.9)，可以將 g_2 重寫成 $g_1 * g_1$ 。換言之，引用引理一的式(2.6)，可以將二階導數重新表示為

$$\begin{aligned}
 Q(x) &= [f(x) * \{-1, 1\}] * \{-1, 1\} \\
 &= P(x) * \{-1, 1\}
 \end{aligned}$$

因為 g_2 可以從 $g_1 * g_1$ 得到；也就是說 $f(x-1) - 2f(x) + f(x+1)$ 和 $P(x) - P(x-1)$ 是等效的，可以毋須顧慮 g_2 高通濾波器的正交性質。

參、研究方法

一、使用軟體

(一) 用 ARACK32 的錄音軟體以 8bits 的樣本大小，取樣率為 22050Hz，錄製語音。語音資料為 0 至 255 的字元型態，128 代表聲音的訊號為零。檔案格式為 wav 檔。

(二) 程式語言採用 TURBO C。

二、演算法

(一) 演算法 3·1 (計算一階導數及二階導數)

輸入：羅馬拼音之元音字元檔

輸出：元音一階導數字元檔及元音二階導數字元檔

1、取語音資料的總長度；記為 N

2、令計算二階導數的判定旗標為零；記為 flag=0

3、令二階導數的暫存值為零；記為 Temp=0

4、讀取第一個語音字元；記為 r1

5、For i=2 to N Do

5-1 讀取下一個語音字元；記為 r2

5-2 一階導數=r2-r1

5-3 存入一階導數輸出檔

5-4 If (flag=1) then

5-4-1 二階導數=Temp+r2

5-4-2 存入二階導數輸出檔

5-5 Temp=r1-2xr2

5-6 r1=r2

5-7 flag=1

6、結束

(二) 演算法 3·2 (計算元音週期)

輸入：羅馬拼音之元音字元檔、元音一階導數字元檔及元音二階導數字元檔

輸出：週期元音字元檔及元音頻率值

- 1、讀取元音資料總長度；記為 N
- 2、計算元音資料中振幅為正的平均值與標準差；分別記為 μ 及 σ
- 3、設定選取尺度為 $\mu + \sigma$ ；記為 Scale
- 4、設定計數值為零；記為 $ct=0$
- 5、讀取第一個元音字元資料
- 6、For $i=2$ To $N-1$ Do
 - 6-1 再讀取元音字元資料；記為 r
 - 6-2 讀取一階導數資料；記為 D_1
 - 6-3 讀取二階導數資料；記為 D_2
 - 6-4 IF ($r > \text{Scale}$ 且 $D_1 = 0$ 且 $D_2 < 0$) /*運用推論一*/
 - then 6-4-1 計數值 $ct+1$
 - 6-4-2 記錄相對極大值發生位置；記為 $\text{locate}[ct]=i$
- 7、 $i=0$
- 8、For $j=1$ To $ct-1$ Do
 - 8-1 計算週期 $\text{locate}[j+1]-\text{locate}[j]$ ；記為 p_d
 - 8-2 $i=i+1$
 - 8-3 記錄週期值；記為 $\text{period}[i]=p_d$
 - 8-4 元音頻率值 = $22050/p_d$
 - 8-5 列印「元音頻率值」

9、將適當週期內的元音資料輸出至「週期元音字元檔」

10、結束

(四) 演算法 3·3 (合成元音)

輸入：週期元音字元檔

輸出：合成元音語音字元檔

1、讀取週期元音字元資料總長度；記為 N

2、設定複製倍數；記為 T

3、For i=1 To N Do

 3-1 讀取週期元音資料

4、For j=1 To T Do

 4-1 輸出週期元音資料至「合成元音語音字元檔」

5、結束

肆、研究結果

表一所列的結果，除了原始語音波形及合成波形以外，還有用語音最大值 255 來標示極值發生位置的波形圖。表二至表六為各個語音樣本的頻率統計值。

表 1：五個元音之各類波形圖

| 元音名稱 | 原始波形 | 標示極值之波形 | 合成圖 |
|------|------|---------|------|
| y | 圖 1 | 圖 2 | 圖 3 |
| i | 圖 4 | 圖 5 | 圖 6 |
| u | 圖 7 | 圖 8 | 圖 9 |
| e | 圖 10 | 圖 11 | 圖 12 |
| o | 圖 13 | 圖 14 | 圖 15 |

語音合成初探

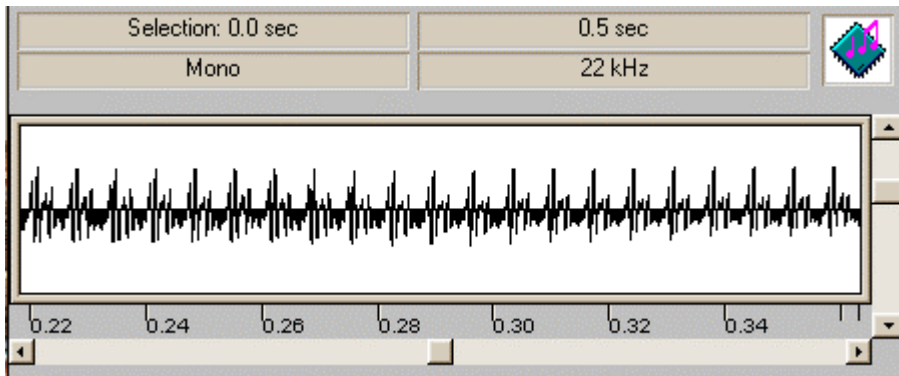


圖 1：「元音 Y」原始波形顯微圖

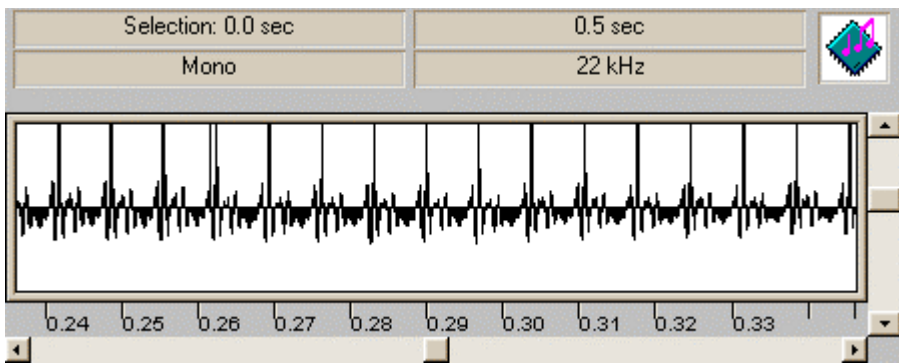


圖 2：「元音 Y」標示極值的波形顯微圖

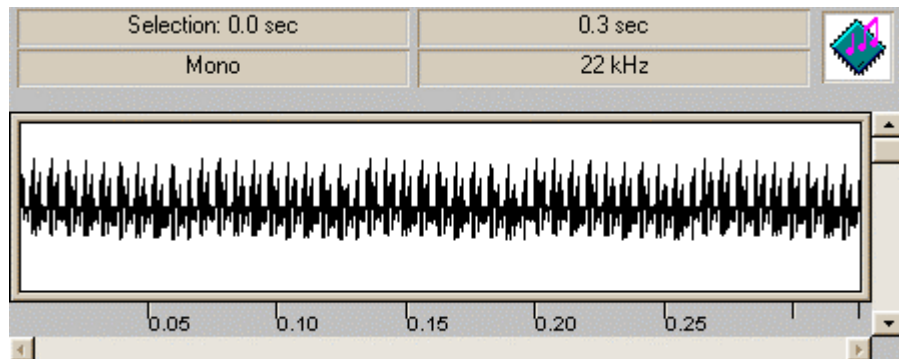


圖 3：「元音 Y」的週期合成波形圖

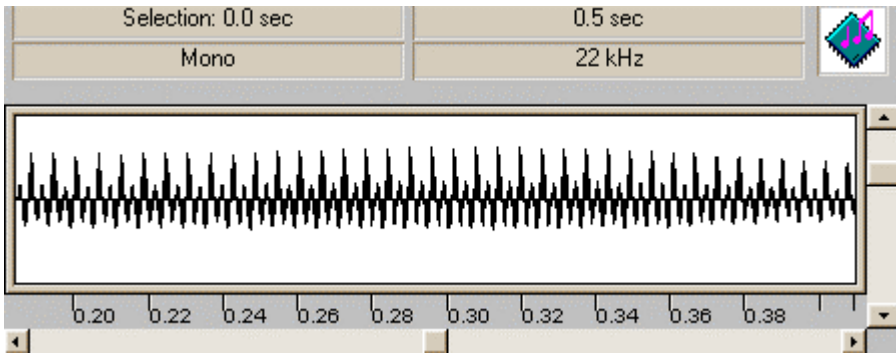


圖 4：「元音 i」原始波形顯微圖

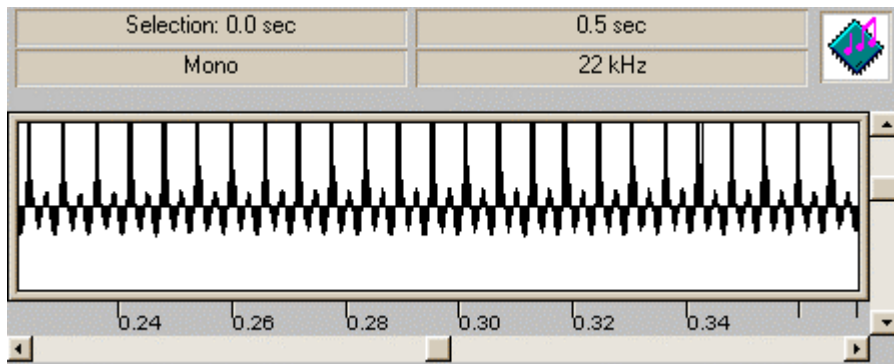


圖 5：「元音 i」標示極值的波形顯微圖

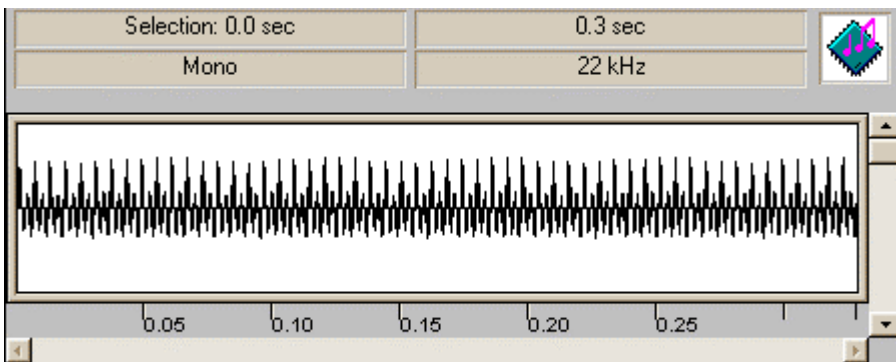


圖 6：「元音 i」的週期合成波形圖

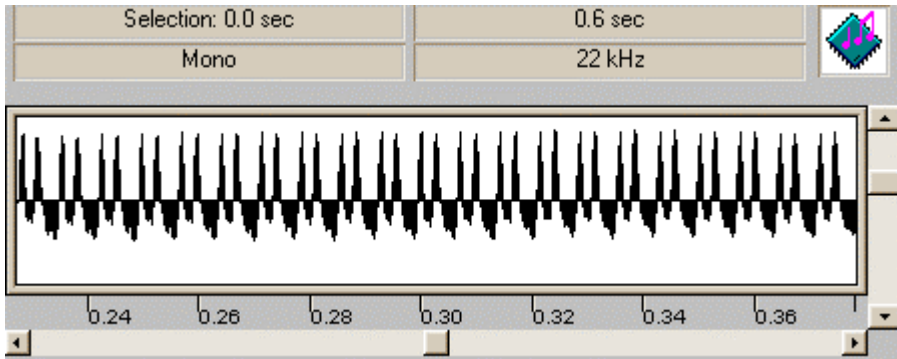


圖 7：「元音 u」原始波形顯微圖

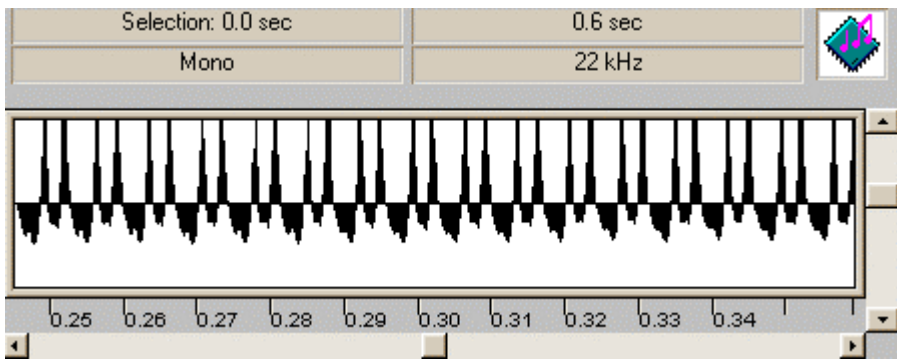


圖 8：「元音 u」標示極值的波形顯微圖

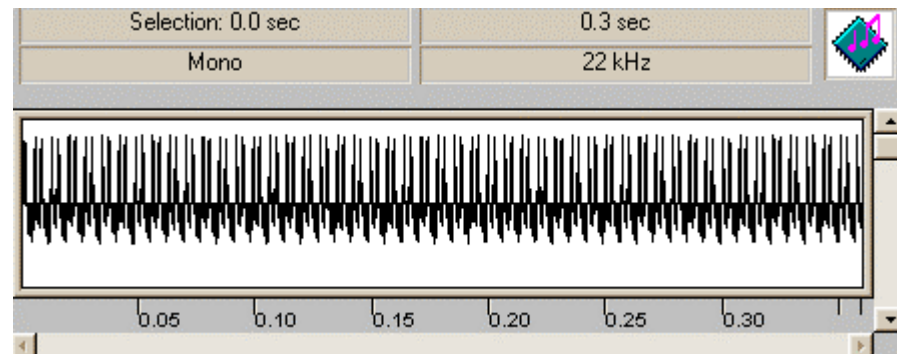


圖 9：「元音 u」的週期合成波形圖

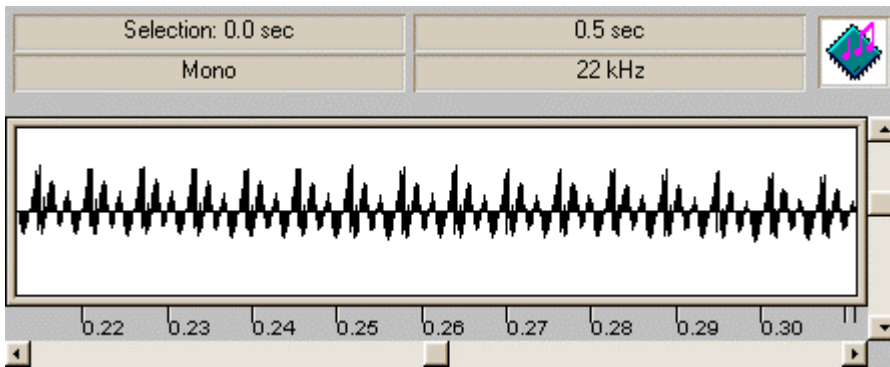


圖 10：「元音 E」的原始波形顯微圖

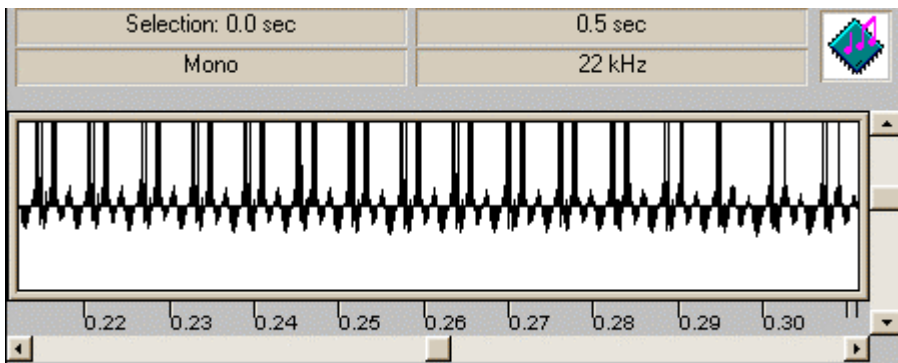


圖 11：「元音 E」標示極值的波形顯微圖

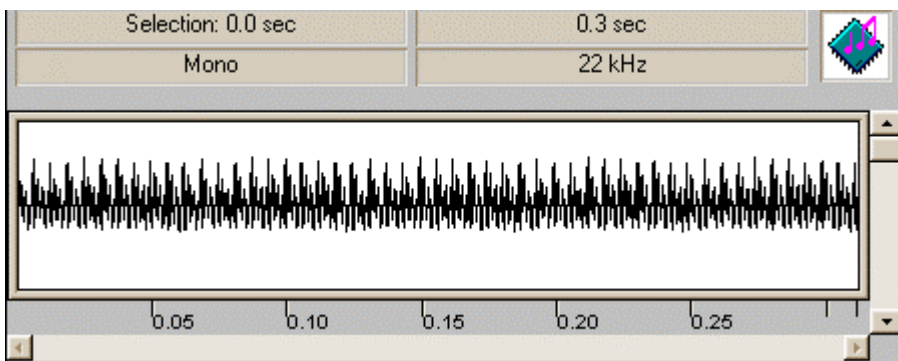


圖 12：「元音 E」的週期合成波形圖

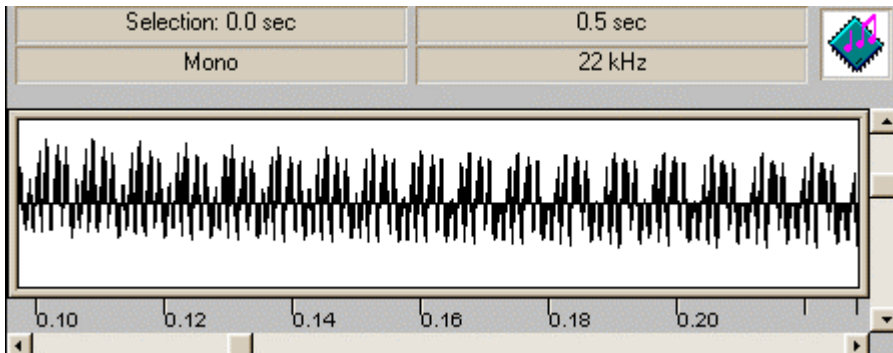


圖 13：「元音 o」的原始波形顯微圖

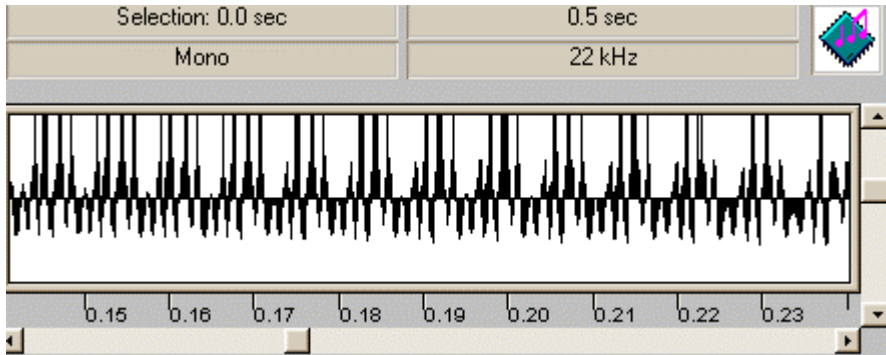


圖 14：「元音 o」標示極值的波形顯微圖

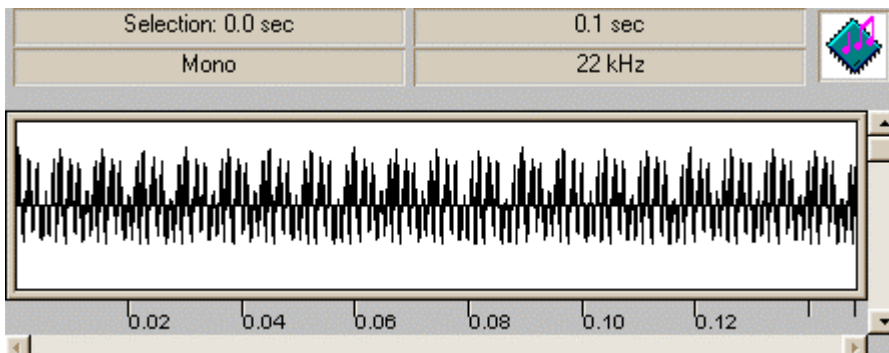


圖 15：「元音 o」的週期合成波形圖

表 2：七個「元音 y」的頻率統計值

| 編號 | 週期 個數 | 最大值 (H z) | 最小值 (H z) | 眾數 (H z) | 中位數 (H z) | 平均值 (H z) | 標準差 (H z) |
|----|----------|--------------|--------------|-------------|--------------|--------------|--------------|
| 1 | 37 | 269 | 98 | 114 | 114 | 122 | 32 |
| 2 | 66 | 401 | 34 | 345 | 339 | 321 | 65 |
| 3 | 47 | 302 | 35 | 251 | 251 | 242 | 38 |
| 4 | 41 | 311 | 130 | 263 | 259 | 253 | 33 |
| 5 | 60 | 286 | 79 | 232 | 232 | 231 | 22 |
| 6 | 70 | 266 | 130 | 256 | 256 | 255 | 17 |
| 7 | 88 | 416 | 135 | 408 | 245 | 265 | 107 |

表 3：七個「元音 i」的頻率統計值

| 編號 | 週期 個數 | 最大值 (H z) | 最小值 (H z) | 眾數 (H z) | 中位數 (H z) | 平均值 (H z) | 標準差 (H z) |
|----|----------|--------------|--------------|-------------|--------------|--------------|--------------|
| 1 | 57 | 180 | 74 | 171 | 167 | 164 | 14 |
| 2 | 77 | 380 | 86 | 345 | 345 | 327 | 59 |
| 3 | 67 | 311 | 87 | 279 | 279 | 276 | 26 |
| 4 | 33 | 272 | 115 | 122 | 123 | 139 | 43 |
| 5 | 38 | 279 | 111 | 129 | 128 | 132 | 30 |
| 6 | 40 | 266 | 113 | 123 | 122 | 130 | 33 |
| 7 | 29 | 237 | 97 | 105 | 106 | 129 | 47 |

表 4：七個「元音 u」的頻率統計值

| 編號 | 週期 個數 | 最大值 (H z) | 最小值 (H z) | 眾數 (H z) | 中位數 (H z) | 平均值 (H z) | 標準差 (H z) |
|----|----------|--------------|--------------|-------------|--------------|--------------|--------------|
| 1 | 73 | 324 | 38 | 315 | 302 | 277 | 66 |
| 2 | 85 | 387 | 339 | 361 | 356 | 356 | 10 |
| 3 | 35 | 356 | 117 | 119 | 119 | 144 | 67 |
| 4 | 35 | 119 | 110 | 114 | 114 | 114 | 2 |
| 5 | 105 | 311 | 124 | 306 | 267 | 247 | 61 |
| 6 | 70 | 324 | 115 | 259 | 251 | 218 | 78 |
| 7 | 79 | 315 | 48 | 298 | 259 | 236 | 76 |

表 5：七個「元音 e」的頻率統計值

| 編號 | 週期 個數 | 最大值 (H z) | 最小值 (H z) | 眾數 (H z) | 中位數 (H z) | 平均值 (H z) | 標準差 (H z) |
|----|----------|--------------|--------------|-------------|--------------|--------------|--------------|
| 1 | 40 | 266 | 103 | 104 | 158 | 158 | 55 |
| 2 | 74 | 735 | 117 | 565 | 485 | 389 | 196 |
| 3 | 45 | 565 | 53 | 565 | 127 | 264 | 204 |
| 4 | 81 | 334 | 96 | 290 | 283 | 276 | 30 |
| 5 | 65 | 565 | 122 | 551 | 176 | 290 | 180 |
| 6 | 92 | 788 | 148 | 298 | 513 | 485 | 184 |
| 7 | 57 | 334 | 225 | 269 | 269 | 273 | 27 |

表 6：七個「元音 o」的頻率統計值

| 編號 | 週期 個數 | 最大值 (H z) | 最小值 (H z) | 眾數 (H z) | 中位數 (H z) | 平均值 (H z) | 標準差 (H z) |
|----|----------|--------------|--------------|-------------|--------------|--------------|--------------|
| 1 | 53 | 817 | 62 | 596 | 311 | 367 | 204 |
| 2 | 59 | 668 | 90 | 551 | 230 | 275 | 144 |
| 3 | 83 | 735 | 144 | 320 | 311 | 343 | 128 |
| 4 | 68 | 408 | 88 | 269 | 269 | 270 | 42 |
| 5 | 98 | 735 | 63 | 324 | 324 | 378 | 152 |
| 6 | 66 | 380 | 112 | 259 | 259 | 259 | 34 |
| 7 | 83 | 711 | 44 | 689 | 283 | 318 | 197 |

伍、討 論

從圖二、圖五、圖八、圖十一及圖十四，我們確知利用一階導數與二階導數的數學方法可以找出極值發生的位置。可是，倘若我們將所有兩兩極值發生的間隔，全部當成週期；從圖中不難發現有一些的極值發生的位置太過於接近，並不適合當作週期；為此我們將所得到的週期值，全數相加求平均，利用此平均週期值做為篩選標準，揀選那些高於平均週期的值，如此便可以得到比較合理的週期值。

假設目前得到 N 個發生極值的位置座標，根據元音的壓縮率大約五分之一（彭玉峰，民國 88 年），因此決定只取 $\frac{N}{5}$ 個發生極值的位置座標來合成元音。我們取座標從 $\frac{2N}{5}$ 到 $\frac{3N}{5}$ 所對應實際位置的語音訊號來合成元音，得到圖三、圖六、圖九、圖十二及圖十五。合成的波形雖然看起來和原始波形有一些出入，但就聽覺而言，已經十分接近了。

因為我們採用取樣頻率為 22050Hz，所以我們利用公式：頻率值 = 22050/週期值，將所求得的週期值換算出實際的元音頻率，得到表二至表六。從表二至表六中，我們發現元音的頻率並非單一定值，而是一個特定的範圍值。如果我們以統計學中常用的眾數、中位數及平均數，可以得到表七：五個元音的發音頻率範圍

表 7：五個元音的發音頻率範圍

| 元音 | 眾數 (H z) | 中位數 (H z) | 平均值 (H z) |
|----|-------------|--------------|--------------|
| y | 114~408 | 114~339 | 122~321 |
| i | 105~345 | 106~345 | 129~327 |
| u | 114~361 | 114~356 | 114~356 |
| e | 104~565 | 127~513 | 158~485 |
| o | 259~689 | 230~324 | 270~378 |

陸、結 論

我們從以上的討論可以歸納出以下兩點：

- (一) 如果用來合成的「週期訊號」太少，會使得聲音聽起來很單調，不夠自然，但根據元音的壓縮率大約五分之一（彭玉峰，民國 88 年），我們取 $\frac{N}{5}$ 個發生極值座標所對應實際位置的語音訊號來合成元音，就不會有這個問題了。（其中 N 為發生極值位置的座標總數目）。
- (二) 元音訊號的頻率並非單一定值，而是一特定範圍；若我們單看頻率的平均值，可以推論元音頻率大多是介於 100Hz 到 500Hz 之間的低頻。

對於元音週期性質的探討，並且利用該週期訊號來合成，我們有了初步的結果。但如何利用更少的週期訊號來合成，而可以避免單調的現象是一個課題。我想下一步應該是找出輔音的特性，畢竟羅馬拼音的基礎，就建立在這兩種發音狀態上。不過更有趣的是如何利用元音的週期性質及輔音的特質做語音切割，進行最有挑戰的語音辨認。

參考文獻

中文部分

- 【1】 彭玉峰，「羅馬拼音的同構數學模型」，康寧學報，第二卷，第一期
- 【2】 劉斌豪，「拉普拉斯變換原理與解題」，中央圖書出版社，(1982)
- 【3】 蕭世文，「聲霸卡程式設計實務音效發展與應用」，旗標出版社，(1995)

英文部分

- 【1】 JHING-FA WANG AND SHI-HUANG CHEN AND JYH-SHING SHYUU ,
「 WAVELET TRANSFORMS FOR SPEECH SIGNAL PROCESSING 」 ,
JOURNAL OF THE CHINESE INSTITUTE OF ENGINEERS」 , VOL.22 , 1999 ,
NO5 , pp549-560
- 【2】 Gilbert Strang and Truong Nguyen , 「 Wavelet and Filter Banks 」 , Wellesley
Cambridge ,(1996)
- 【3】 L.R.RABINER AND R.W.SCHAFFER , 「 DIGITAL PROCESSING OF SPEECH
SIGNALS 」 PRENTICE HALL , (1978)
- 【4】 RONALD N.BRACEWELL , 「 The Fourier Transform and Its Application 」 ,
McGRAW-Hill (1986)
- 【5】 Strang G. , 「 Wavelet and Dilation Equation:A Brief Introduction 」 , SIAM
Review,Vol.31,1989,No.4, 614-627
- 【6】 TOM M.APOSTOL , 「 CALCULUS 」 , WILEY (1969)