

行政院國家科學委員會補助
大專學生參與專題研究計畫研究成果報告

* ***** *
* 計 畫 *
* : 以 MPEG-7 之 ART 轉換應用於鳥類聲紋辨識之研究 *
* 名 稱 *
* ***** *

執行計畫學生： 羅永淇
學生計畫編號： NSC 97-2815-C-216-008-E
研究期間： 97年07月01日至98年02月28日止，計8個月
指導教授： 周智勳

處理方式： 本計畫涉及專利或其他智慧財產權，2年後可公開查詢

執行單位： 中華大學資訊工程學系

中華民國

98年03月27日

行政院國家科學委員會補助

大專學生參與專題研究計畫研究成果報告

* 計畫 *
* 名稱 : 以 MPEG-7 之 ART 轉換應用於鳥類聲紋辨識之研究 *
* *****

執行計畫學生： 羅永淇

學生計畫編號： NSC97-2815-C-216 -008 -E

研究期間： 97年7月1日至 98年2月底止，計8個月

指導教授： 周智勳 博士

執行單位：中華大學資訊工程學系

中華民國 98 年 3 月 25 日

目錄

| | |
|------------------------------|----|
| 一、摘要..... | 4 |
| 二、研究動機與研究問題 | 4 |
| 三、文獻回顧與探討 | 5 |
| 四、研究方法及步驟 | 7 |
| (一)聲音訊號切割(Segmentation)..... | 8 |
| (二)DFT | 9 |
| (三)產生聲譜圖 | 10 |
| (四)ART 轉換..... | 15 |
| (五)比對..... | 17 |
| (六)實驗結果..... | 18 |
| 五、評估與展望 | 22 |
| 六、結語..... | 22 |
| 七、參考文獻 | 23 |

圖形目錄

| | |
|---|----|
| <圖一>鳥聲辨識系統之流程圖 | 7 |
| <圖二>利用 DFT 由 time domain 轉為 frequency domain..... | 9 |
| <圖三> (a)原始鳥聲譜圖 (b) 切割完成後每一音節之聲譜圖 | 10 |
| <圖四>將音框訊號拉寬之例子 | 11 |
| <圖五>將音框訊號拉寬之圖形 | 12 |
| <圖六>將音框訊號刪減之例子 | 13 |
| <圖七>將音框訊號刪減之圖形 | 14 |
| <圖八> ART 轉換後(a)實部 (b)虛部 | 16 |
| <圖九> 系統介面圖 | 17 |
| <圖十> ART 基底 $m = 12$, $n = 3$ 比對之結果..... | 18 |
| <圖十一> ART 基底 $m = 12$, $n = 12$ 比對之結果..... | 19 |
| <表一> ART 利用 KNN 比對之成功率 | 20 |
| <表二> ART 利用 NN 比對之成功率 | 21 |

一、摘要

本研究計畫將應用 MPEG-7 之 ART (Angular Radial Transform) 轉換以辨識鳥類聲紋，並且以此區別各個鳥類之種類。首先，我們將鳥聲訊號切割 (Segmentation) 為一個個的音節，並且以音節為辨識單元，然後以傅立葉轉換來產生每一音節之聲音訊號之聲譜圖 (Spectrogram)，而且將此聲譜圖視為一張影像，因此我們再以 MPEG-7 之 ART 轉換來擷取聲譜圖中隨聲音變化之形狀特徵，用來比對鳥聲聲紋之差異性以分辨出每一種鳥類的聲音。

二、研究動機與研究問題

動物叫聲辨識中，鳥類鳴叫聲音之辨識研究是最普遍的，全世界鳥類約有九千多種，由於鳥類的種類相當多，而且不同的鳥種，其鳴叫聲通常也會有所不同，生物學家為取得鳥聲樣本，大多採用野外錄音，再回實驗室做人工辨識區別，這樣的研究方式既費力又費時，若能夠有更有效的方法，減少人力及時間的花費該有多好呀！過去傳統的聲音訊號處理都是將聲音訊號以傅立葉轉換為頻譜，再以頻譜特徵來辨識，本研究是將聲音訊號轉換為聲譜圖，並將其視為一張影像，再利用 MPEG-7 之 ART 轉換來擷取聲譜圖之形狀變化特徵來辨識，與傳統不同的地方是前者是以聲音訊號之特性做為切入點，而我們的方法是以影像之形狀變化做為切入點，希望以不同的角度切入來建立鳥類聲紋辨識系統。本研究計畫主要的目的在建構一個擁有各種常見鳥類聲音的資料庫，並可用此資料庫記錄不同鳥類鳴叫聲音之特徵，每當我們錄製一段鳥類聲音訊號時，便可以此自動比對、分辨每種鳥類聲音的訊號。

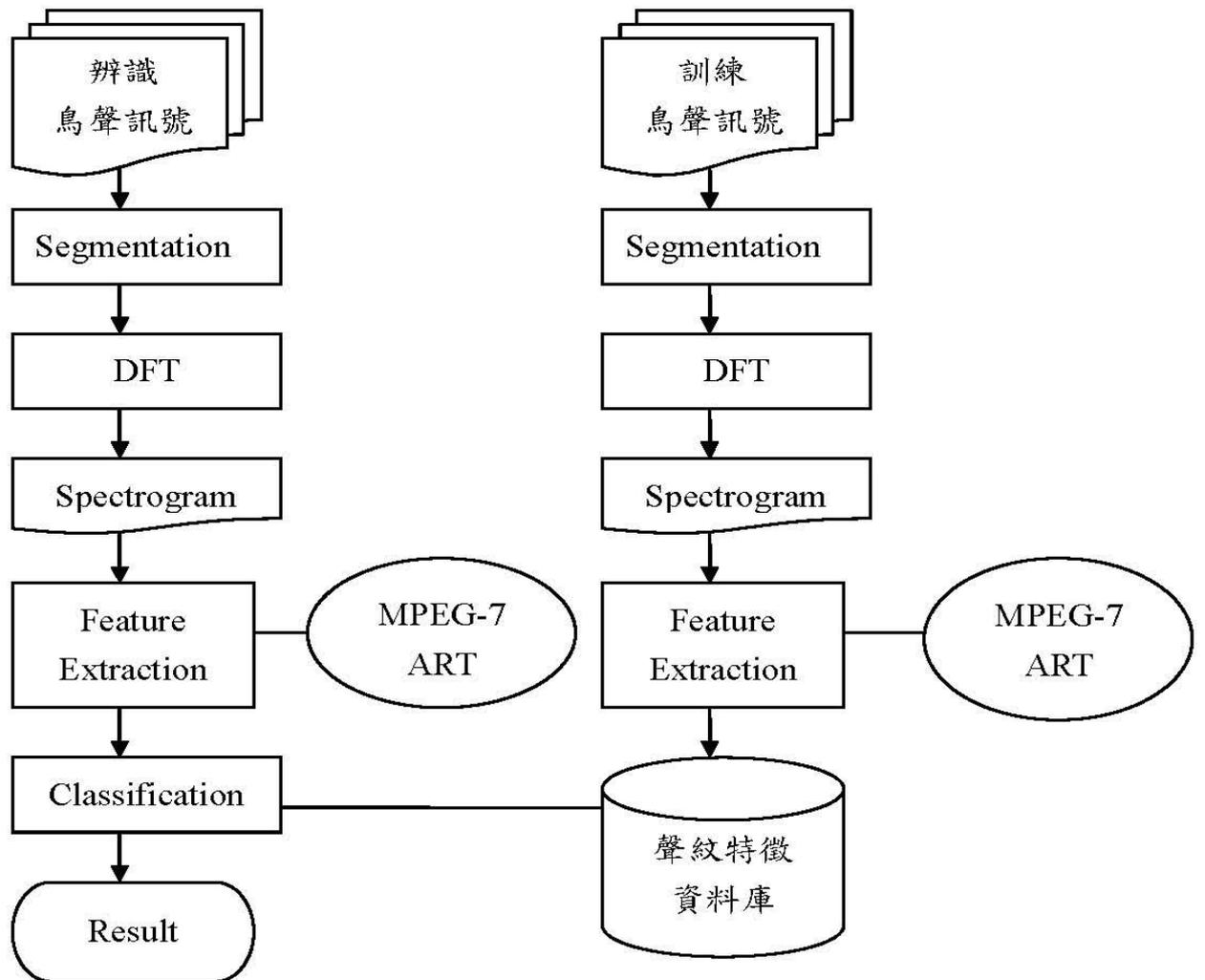
三、文獻回顧與探討

東華大學張勇富碩士論文中提出以語料分析為主之鳥聲辨識系統[4]。其系統首先以能量資訊(energy information)來切割出鳥鳴聲中的音節(syllable)，而每一音節中會包含多個音框(frame)，然後擷取每個音框的頻譜中發生振幅最大值之頻率當基本頻率，最後以所有音框的基本頻率之中間值(median frequency)來辨識鳥類種類。Anderson 等人利用 DTW(dynamic time warpping)來分析連續錄音中鳥類歌聲中的每一音節[5]。其方法是直接比較這些聲音訊號的聲譜圖，並找出聲譜圖上的詞組單元(constituent)和邊界(constituent boundaries)，然後對聲譜圖上的振幅取對數當做特徵向量，而擷取之頻率範圍為 0.5~10KHz。他們以這套方法來辨識靛青(indigobunting)和錦花雀(zebra finch)這兩種鳥類，其測誦聲音檔是在低噪音的環境下收集而來的，而且每一鳥種具代表性之音節採用人工的方式切出其聲音訊號樣本。其實驗結果顯示當音節變化不大時辨識結果的準確率可以達到 97%，但是當音節結構變化大時，準確率會下降到 84%。Kogan 和 Margoliash 比較動態時間校正和隱藏馬可夫模型(hidden Markov model)在辨識鳥類聲音上的效能[6]。而 DTW 和前述方法相同，在以 HMM 來辨識時取六類不同的特徵及參數來當做 HMM 的輸入包括 linear predictive coding(LPC)、LPC cepstral coefficients(LPCC)、LPC reflection、mel-frequency cepstral coefficients(MFCC)、log mel-filter bank channel、linear mel-filter bank channel。其中以 MFCC 效能最好，而選用之 MFCC 參數包括其能量值(energy)以及一次和二次導函數值(firat and second derivatives)。其實驗結果顯示 DTW 的辨識效能不錯，對雜訊干擾較嚴重的輸入聲音訊號或易混淆之短叫聲，使用 DTW 時需要一些較專業背景知識來選擇恰當且具代表性之聲音訊號樣本以輔助 DTW 之運作。而 HMM 則需要對聲音訊號更好的切割和歸類的動作，其缺點為對發聲時間較短和結構複雜之鳥鳴聲常有判斷錯誤的情形發生。

McIlraith 和 Card 提出利用類神經網路和統計方法來分辨六種鳥鳴聲, (song sparrow、fox sparrow、marsh wren、sedge wren、yellow warbler 和 red-winged blackbird) [7-10], 其擷取特徵包括時域及頻譜之資訊。時域採用 song element 之個數, song element 長度之平均值及標準差和靜音(silence)時段之長度平均值及標準差。頻譜則採用 LPC 倒頻譜係數, 並將訊號分成九個頻帶(subband)後, 每個頻帶之頻譜能量(power spectral density)之平均值與標準差, 再利用倒傳遞類神經路(backpropagation neural networks)來做分類, 最後的準確率為 82%, 而利用二次區別分析演算法(quadratic discriminant analysis), 可以將準確率提升到 93%。

四、研究方法及步驟

本計畫之鳥類鳴叫聲音辨識系統之流程圖如圖一所示。



圖一 鳥聲辨識系統之流程圖

首先，將一段連續鳥聲訊號切成數個音節，接著以音節做為辨識單元，再以傅立葉轉換產生其聲譜圖，然後將此聲譜圖視為一張影像，以 MPEG-7 之 ART 轉換來擷取其聲譜圖中形狀變化之特徵，用來比對聲紋之差異性以分辨出每一種鳥類的聲音。

(一)聲音訊號切割(Segmentation)

在此步驟中，我們是採用 Somervuo 及 Harma 兩人所提出之方法[1]，其詳細步驟如下。

步驟 1:首先，將取到的鳥聲訊號切為一個一個的音框，且相鄰兩個音框間重疊一半。

步驟 2:對每一個切割出來的音框，計算其能量總和：

$$E(n) = \sum_{m=0}^{N-1} 20 \log_{10} (w[m]x_n[m])^2$$

，其中 $x_n[m]$ 代表第 n 個音框訊號， $w[m]$ 為 Hanning 視窗函數，然後以所有音框的對數能量總和最小值者做為初始之背景環境噪音之能量值 N_{dB} ，取臨界值 $T = \frac{N}{2} dB$ 。

步驟 3:篩選對數能量總和低於目前環境噪音 N_{dB} 之音框並加以標記。

步驟 4:將前一步驟所標記之所有音框的對數能量總和加總，並取臨界值

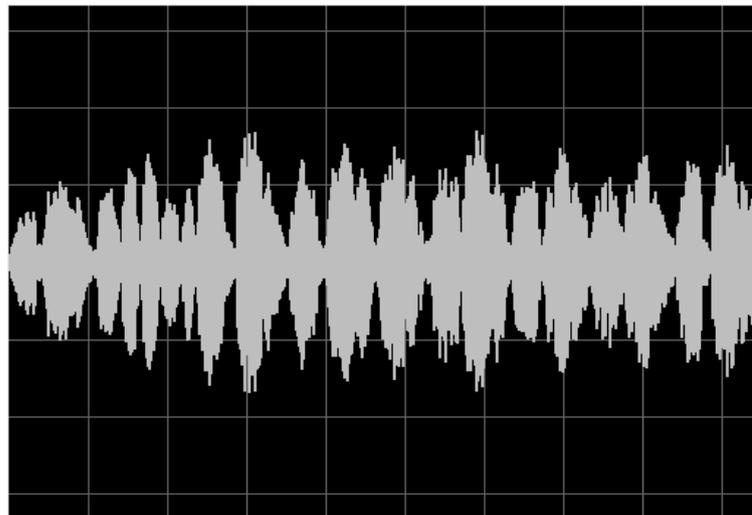
$$T = \frac{N}{2} dB。$$

步驟 5:重覆步驟 3 及步驟 4，直到篩選之標記音框不再改變為止。

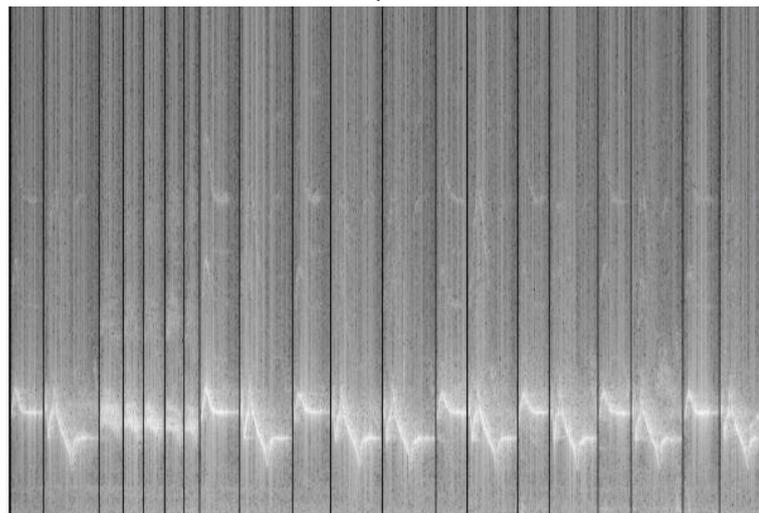
(二)DFT

將上述切好的一節一節鳥聲音訊，利用下列公式，將 time domain 轉換至 frequency domain(如圖二)。

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{k}{N} n} \quad k = 0, 1, \dots, N-1.$$



time domain



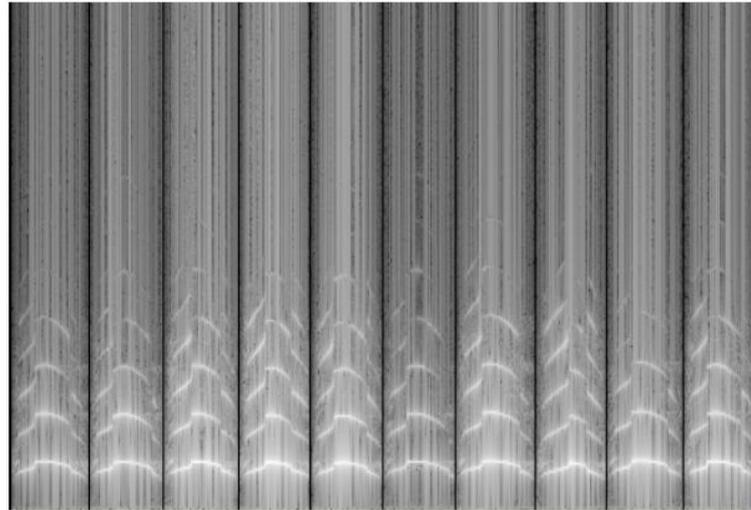
frequency domain

圖二 利用 DFT 由 time domain 轉為 frequency domain

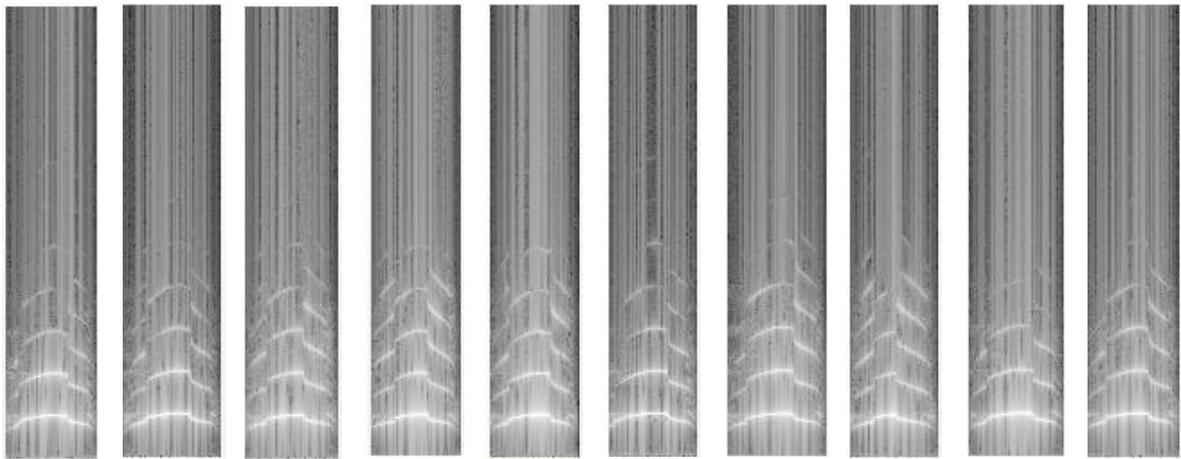
(三)產生聲譜圖

我們再利用上述所產生的音節，切割成一個個音框，再以傅立葉轉換成聲譜圖。

(如下圖)



(a)

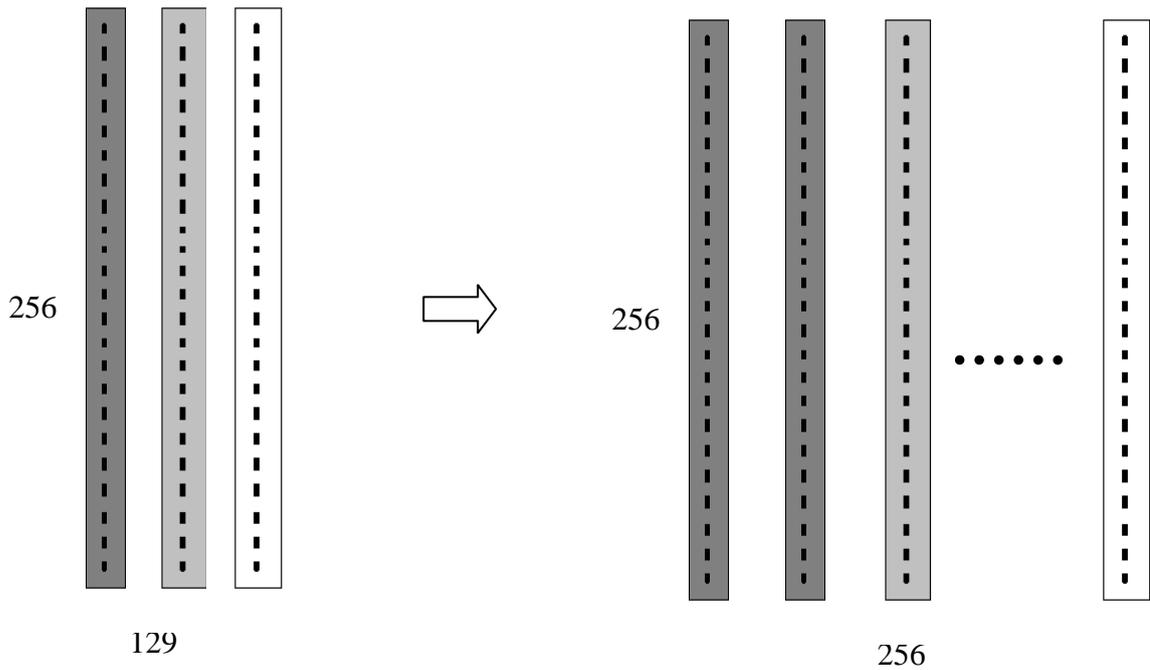


(b)

圖三 (a)原始鳥聲譜圖 (b) 切割完成後每一音節之聲譜圖

由於，每一段鳥聲訊號所產生聲譜圖的音框個數不一，而 MPET-7 之 ART 的 MASK 是一張圓餅形的圖形，因此我們將聲譜圖的寬度(Frame Number)固定與高度(Frequency Coefficient)相同大小，其演算法如下：

(1) Frequency Coefficient > Frame Number



圖四

我們要將音框訊號由 n 個 frame 補成 256 個 frame，步驟如下。

步驟 1:
$$\frac{\text{Frequency Coefficient}(256)}{\text{Frame Number}(N)} = Q \dots R_1$$

商數 Q: 代表每一個音框訊號要補成 Q 個音框訊號。

餘數 R_1 : 代表餘數

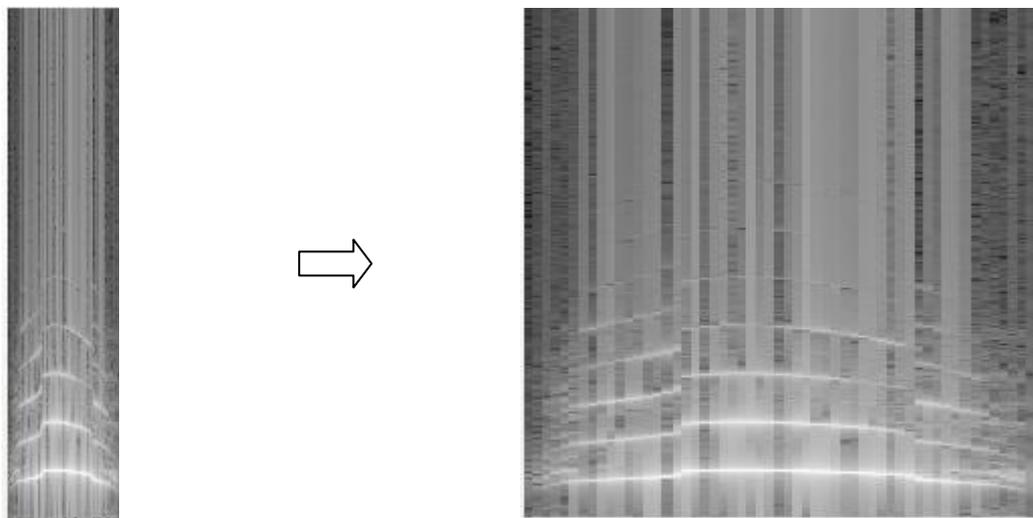
我們將 256 除以 N 以決定每個 frame 要複製的個數，如果不能整除就會有不能補滿 256 個 frame 的狀況發生，這時要進到步驟 2。

步驟 2: $\frac{\text{Frequency Coefficient}(256)}{R_1} = T_0 \dots R_2$

商數 T_0 : 代表每 T_0 個 frame 就要多內插一筆資料進入

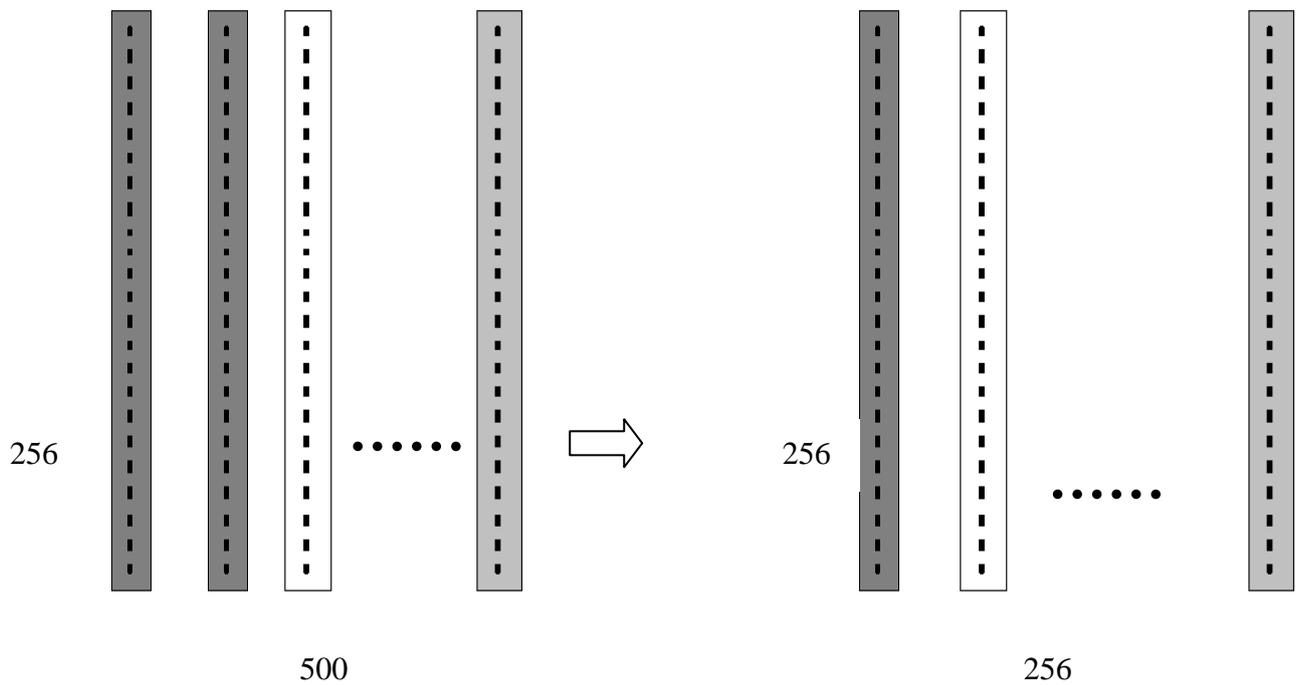
餘數 R_2 : 代表是否還有空隙

由於 frame 沒辦法補滿的關係，因次要選擇適當的時機將資料內插進去，也就是每 T_0 要多加一筆資料進去，而資料的選擇就是以最接近的音框訊號為代表(如下圖)。



圖五

(2) Frame Number > Frequency Coefficient



圖六

我們要將音框訊號由 N 個 frame 刪成 256 個 frame，步驟如下。

步驟 1:
$$\frac{\text{Frame Number}(N)}{\text{Frequency Coefficient}(256)} = Q \dots R_1$$

商數 Q: 代表每幾個音框資訊留一個音框資訊，其餘資訊刪除。

我們將 N 除以 256 以決定每幾個 frame 要留下 1 個，如果不能整除就會有多餘的 frame 的狀況發生，這時要進到步驟 2。

步驟 2: $\frac{256 + R_n}{R_1} = T_0 \dots R_2$

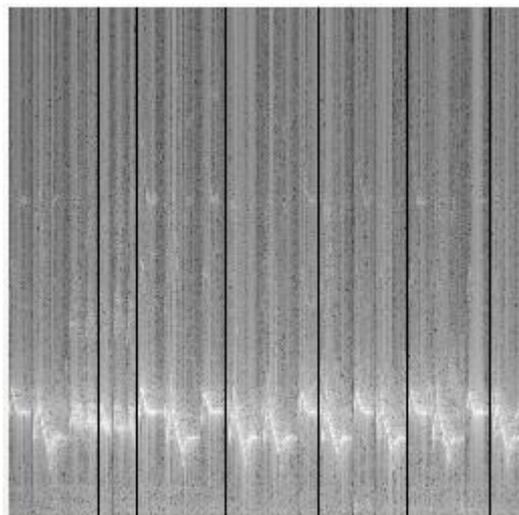
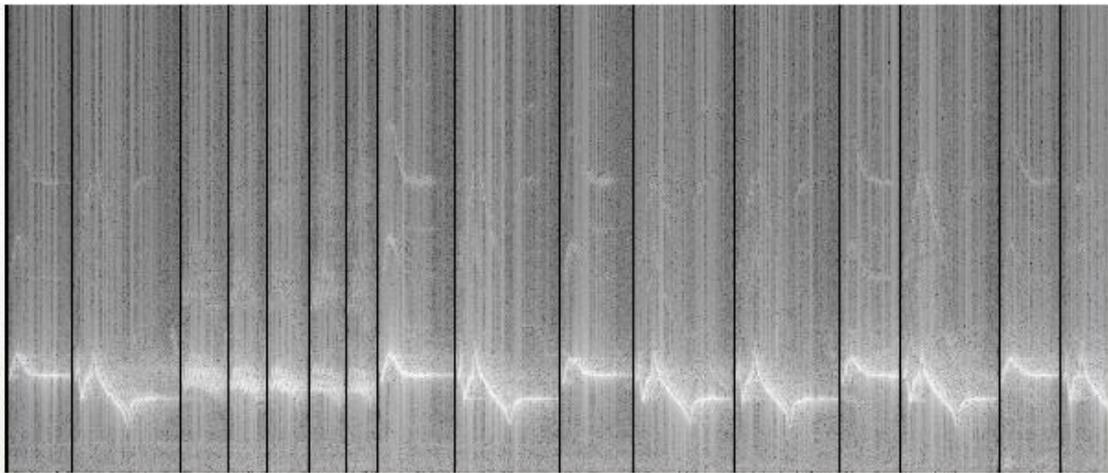
商數 T_0 : 以剩下未刪的 frame 為主，每隔 T_0 個刪一個訊號。

餘數 R_2 : 代表是否還有多餘的音框訊號。

$256 + R_1$: 代表剩下的 frame 個數。

如上述，為了使寬度(Frame Number)與高度(Frequency Coefficient)相同，

我們需要將剩下的 R_2 個音框刪除。(如下圖)。



圖七

(四)ART 轉換

在 MPEG-7 之標準中，採用一組 ART 的係數來描述特體之內部形狀 (Region Shape Descriptor) [2,3]。而 ART 係數之定義如下：

$$\begin{aligned} f_{nm} &= \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle \\ &= \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta) \rho \, d\rho \, d\theta \\ &= \int_0^{2\pi} \int_0^1 A_m(\theta) R_n(\rho) f(\rho, \theta) \rho \, d\rho \, d\theta \end{aligned}$$

其中 $f(\rho, \theta)$ 為影像以極座標表示之灰階值；而 $V_{nm}(\rho, \theta)$ 為 ART 之基底函數，而且可分為角度 (angular) 及徑度 (radial) 變化之兩函式 $V_{nm}(\rho, \theta) = A_m(\theta) R_n(\rho)$ ， $A_m(\theta)$ 和 $R_n(\rho)$ 可分別以指令數函數及餘弦函數來表示：

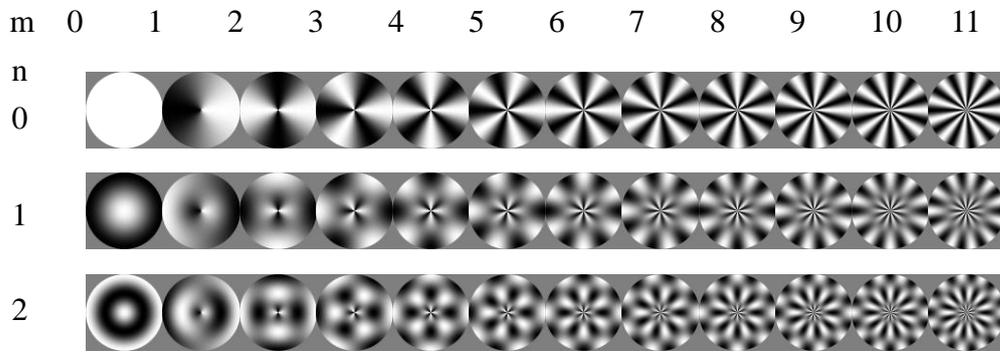
$$\begin{aligned} A_m(\theta) &= \frac{1}{2\pi} \exp(jm\theta) \quad , \\ R_n(\rho) &= \begin{cases} 1 & n=0 \\ 2 \cos(\pi n \rho) & n \neq 0 \end{cases} \quad . \end{aligned}$$

m 、 n 值分別代表 ART 的角度係數和半徑係數。圖三顯示 ART 基底函數之實部及虛部值，由圖三可以發現當角度係數 m 值增大時，ART 基底實(虛)部圖形的黑白扇形區域之圓心角角度變化頻率隨 m 值增大而遞增；當半徑係數 n 值增大時，ART 基底實(虛)部圖形的黑白同心圓數目隨 n 值增大而增加(漣漪現象)，為了不受形狀縮放之影響，我們將每個 ART 係數正規化，將每個 ART 係數除以 $m=0$ 及 $n=0$ 的 ART 係數，即為 ART 描述子：

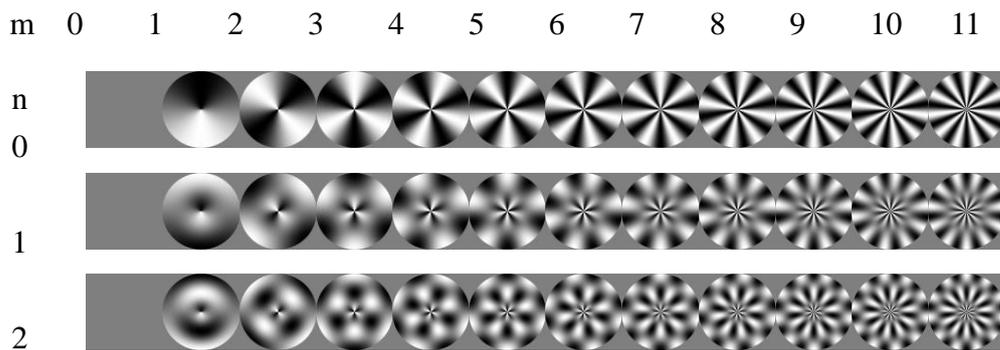
$$ART[i] = \frac{f_{nm}}{f_{00}} \quad , \quad i = 12 \times n + m - 1 \quad .$$

在 MPEG-7 標準中，採用 35 個 ART 係數 ($m=12$ 、 $n=3$)來表達物體內部區域的形狀特徵。對於兩張影像 q 及 s ，其 ART 特徵的差異 $d_{ART}(q,s)$ 之定義為

$$d_{ART}(q,s) = \sqrt{\sum_{i=0}^{34} (ART_q[i] - ART_s[i])^2}$$



(a)

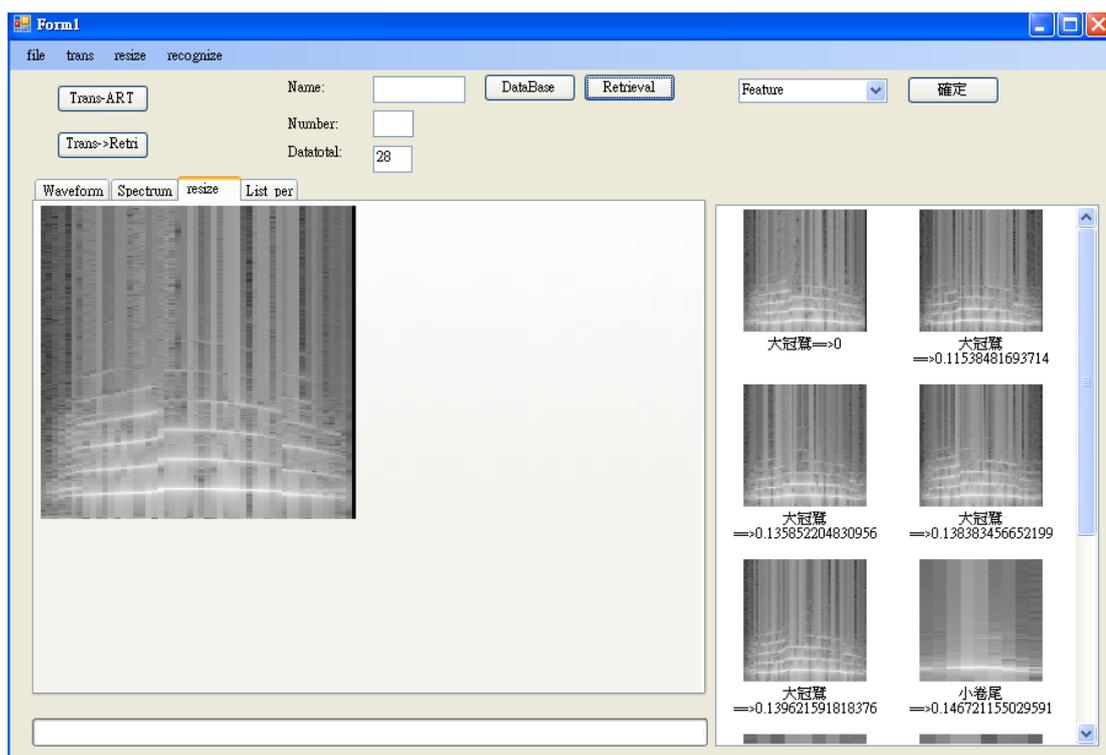


(b)

圖八 ART 轉換後(a)實部 (b)虛部

(五)比對

在辨識時，我們將上述鳥聲特徵以 ART 轉換比對後的結果取差異性最小之鳥類，做為辨識之鳥種，我們算出 testing data 及 training data 兩張圖的特徵距離，並選擇差異性最小的前十名。



圖九

在計算成功率的部分，我利用 Nerest Neighbor Classifier 取前十名中的第一名視為該種鳥聲及 K-nearest neighbor method，在兩種鳥聲比對時，我們將前十名出現最多的鳥聲種類，將 testing data 視為該種鳥聲，然後利用下列公式計算出成功率。

$$\frac{S}{T} = Q$$

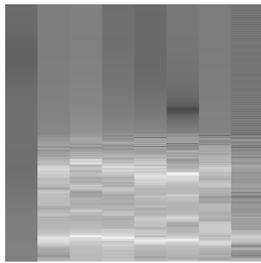
S:表示比對成功的鳥聲數目。

T:表示該鳥聲總數。

(六)實驗結果

將轉為聲譜圖之鳥聲訊號將其寬及高調整為正方形後，以 MPEG-7 之 ART 套用計算其特徵值，並且將 ART 之基底修改 ($m = 12, n = 3$ 或 $m = 12, n = 12$) 用以辨識及比較，並算出其平均成功率。

比對之鳥聲--筒鳥



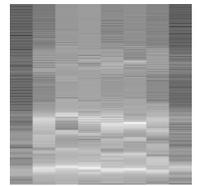
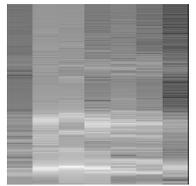
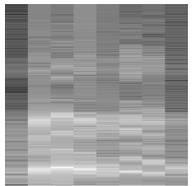
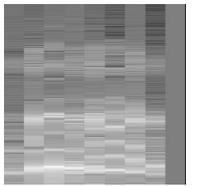
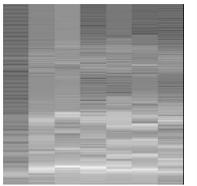
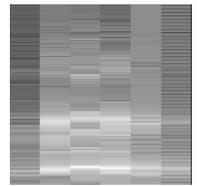
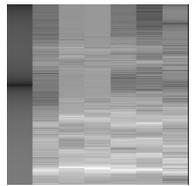
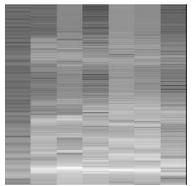
(a)將 ART 之基底設為 $m = 12, n = 3$:

以下為比對後前十名之鳥聲，利用上述之定義 $d_{ART}(q, s)$ 比較其差異性，由左至右，越往右邊差異性越大。

| | | | | | |
|-----------------|------|------|------|------|------|
| | | | | | |
| 鳥類: | 筒鳥 | 筒鳥 | 筒鳥 | 白喉笑鵝 | 鳳頭蒼鷹 |
| $d_{ART}(q, s)$ | 0.16 | 0.24 | 0.27 | 0.28 | 0.28 |
| | | | | | |
| 鳥類: | 鳳頭蒼鷹 | 鳳頭蒼鷹 | 鳳頭蒼鷹 | 白喉笑鵝 | 白喉笑鵝 |
| $d_{ART}(q, s)$ | 0.28 | 0.31 | 0.31 | 0.31 | 0.32 |

圖十

(b)將 ART 之基底設為 $m = 12, n = 12$:

| | | | | | |
|----------------|---|---|---|---|---|
| |  |  |  |  |  |
| 鳥類: | 筒鳥 | 筒鳥 | 筒鳥 | 筒鳥 | 筒鳥 |
| $d_{ART}(q,s)$ | 0.19 | 0.37 | 0.41 | 0.50 | 0.55 |
| |  |  |  |  |  |
| 鳥類: | 筒鳥 | 筒鳥 | 筒鳥 | 小啄木 | 小啄木 |
| $d_{ART}(q,s)$ | 0.55 | 0.58 | 0.61 | 0.61 | 0.62 |

圖十一

由上例顯示兩種基底比對後，可看出其差異性，上例(a)為取 ART[12][3]之特徵值比對之結果，而(b)為取 ART[12][12]之特徵值比對之結果，其顯示 ART[12][3]之特徵值比對之成功率較 ART[12][12]之特徵值比對之結果低。

以下為比對後的鳥聲機率:

| | Training data | Testing data | Art[12] [3] KNN 百分比 | Art[12] [12] KNN 百分比 |
|-------|---------------|--------------|------------------------|-------------------------|
| 大冠鷲 | 10 筆 | 4 筆 | 100% | 100% |
| 小卷尾 | 229 筆 | 37 筆 | 62.16% | 70.27% |
| 小啄木 | 17 筆 | 25 筆 | 8% | 0% |
| 小翼鵝 | 296 筆 | 29 筆 | 55.17% | 48.28% |
| 小彎嘴畫眉 | 120 筆 | 22 筆 | 77.27% | 77.27% |
| 火冠戴菊鳥 | 194 筆 | 57 筆 | 96.49% | 98.25% |
| 白耳畫眉 | 98 筆 | 14 筆 | 71.43% | 71.43% |
| 白喉笑鵝 | 100 筆 | 37 筆 | 83.78% | 86.49% |
| 白腹秧雞 | 172 筆 | 15 筆 | 26.67% | 40% |
| 灰鷲 | 70 筆 | 8 筆 | 87.50% | 87.50% |
| 竹鳥 | 31 筆 | 31 筆 | 77.42% | 74.19% |
| 岩鷲 | 122 筆 | 53 筆 | 22.64% | 24.53% |
| 青背山雀 | 140 筆 | 14 筆 | 0% | 7.14% |
| 冠羽畫眉 | 49 筆 | 12 筆 | 33.33% | 50% |
| 紅頭山雀 | 61 筆 | 24 筆 | 75% | 100% |
| 栗背林鴿 | 230 筆 | 18 筆 | 88.89% | 83.33% |
| 烏頭翁 | 131 筆 | 30 筆 | 93.33% | 76.67% |
| 深山竹雞 | 123 筆 | 27 筆 | 51.85% | 62.96% |
| 深山鷲 | 51 筆 | 8 筆 | 75% | 100% |
| 筒鳥 | 284 筆 | 45 筆 | 62.22% | 86.67% |
| 黃山雀 | 222 筆 | 27 筆 | 88.89% | 96.30% |
| 黃腹琉璃 | 76 筆 | 12 筆 | 91.67% | 100% |
| 煤山雀 | 149 筆 | 34 筆 | 91.18% | 85.29% |
| 鳳頭蒼鷹 | 32 筆 | 16 筆 | 81.25% | 100% |
| 頭烏線 | 32 筆 | 18 筆 | 11.11% | 11.11% |
| 鴿鷲 | 61 筆 | 14 筆 | 0% | 0% |
| 藍腹鷲 | 23 筆 | 10 筆 | 90% | 100% |
| 藪鳥 | 20 筆 | 5 筆 | 0% | 20% |
| 平均 | | | 61% | 66% |

表一

ART[12][3]:取 $m = 12$, $n=3$ 的特徵值, 用以辨識其成功率。

ART[12][12]:取 $m = 12$, $n=12$ 的特徵值, 用以辨識其成功率。

| | Training data | Testing data | Art[12] [3] NN 百分比 | Art[12] [12] NN 百分比 |
|-------|---------------|--------------|-----------------------|------------------------|
| 大冠鷲 | 10 筆 | 4 筆 | 100% | 100% |
| 小卷尾 | 229 筆 | 37 筆 | 83.78% | 91.89% |
| 小啄木 | 17 筆 | 25 筆 | 0% | 0% |
| 小翼鵝 | 296 筆 | 29 筆 | 58.62% | 68.97% |
| 小彎嘴畫眉 | 120 筆 | 22 筆 | 68.18% | 50% |
| 火冠戴菊鳥 | 194 筆 | 57 筆 | 98.25% | 94.74% |
| 白耳畫眉 | 98 筆 | 14 筆 | 64.29% | 78.57% |
| 白喉笑鵝 | 100 筆 | 37 筆 | 91.89% | 91.89% |
| 白腹秧雞 | 172 筆 | 15 筆 | 33.33% | 46.67% |
| 灰鷲 | 70 筆 | 8 筆 | 87.50% | 100% |
| 竹鳥 | 31 筆 | 31 筆 | 100% | 100% |
| 岩鷲 | 122 筆 | 53 筆 | 43.40% | 75.47% |
| 青背山雀 | 140 筆 | 14 筆 | 0% | 0% |
| 冠羽畫眉 | 49 筆 | 12 筆 | 50% | 58.33% |
| 紅頭山雀 | 61 筆 | 24 筆 | 87.50% | 95.83% |
| 栗背林鴉 | 230 筆 | 18 筆 | 83.33% | 94.44% |
| 烏頭翁 | 131 筆 | 30 筆 | 83.33% | 86.67% |
| 深山竹雞 | 123 筆 | 27 筆 | 33.33% | 59.26% |
| 深山鷲 | 51 筆 | 8 筆 | 62.50% | 100% |
| 筒鳥 | 284 筆 | 45 筆 | 68.89% | 95.56% |
| 黃山雀 | 222 筆 | 27 筆 | 88.89% | 96.30% |
| 黃腹琉璃 | 76 筆 | 12 筆 | 100% | 100% |
| 煤山雀 | 149 筆 | 34 筆 | 91.18% | 82.35% |
| 鳳頭蒼鷹 | 32 筆 | 16 筆 | 100% | 100% |
| 頭烏線 | 32 筆 | 18 筆 | 61.11% | 72.22% |
| 鴿鷲 | 61 筆 | 14 筆 | 0% | 0% |
| 藍腹鵝 | 23 筆 | 10 筆 | 100% | 100% |
| 藪鳥 | 20 筆 | 5 筆 | 40% | 40% |
| 平均 | | | 67% | 74% |

表二

五、評估與展望

本研究計畫種過 ART 特徵比對後的成功機率平均約為 70%，但某些鳥種的辨識率有待加強，在未來期望能運用一些音訊處理技巧將辨識率提高，本實驗目前在嘗試利用 MFCC 及 NASE 兩種方法，由於人類所能聽到的訊號有限，我們嘗試利用以上兩種方法模擬人類聽覺及音樂的十二律，將轉為 frequency domain 之聲譜圖中的一個個 frame 抽出，將訊號能量切為一個一個的 subband，省去一些人類聽不到的訊號，用以辨識。

六、結語

鳥聲聲紋辨識普遍運用在生物學上，若能有一個好的工具用以幫助鳥種的分辨，就能免去花費人力及時間去辨識鳥類的困擾，因此多花心思建置鳥聲辨識系統能夠省去許多人力，既方便又省時。

本實驗完全是針對將音訊切割為一個個音節後的訊號，希望能將音節的辨識率做得更精確，剛開始為方便後續做實驗，需寫出一個前置程式將一個一個音節切出，之後就進入正題，將訊號由 time domain 轉為 frequency domain 存取聲譜資料，之後針對這些聲譜資料做一些特徵擷取或套用一些音訊處理技巧，而這部分在實驗中是最難的，由於需要找出不同的特徵值，並計算其成功率，還得在一個一個陌生的數據中判斷程式是否有出錯。

從這次實驗中，學到不少有關音訊的技巧，並學會了一個人執行專案，將一個程式從零建起，也學會了當遇到問題時，如何有效的解決困難，這樣的經驗，有助於在未來問題處理上。

七、參考文獻

- [1] P. Somervuo, A. Harma, “Parametric representations of bird sounds for automatic species recognition”, IEEE Trans. on Audio, Speech, and Language Processing, Vol. 14, No. 6, pp.2252-2263, Nov. 2006.
- [2] B. S. Manjunath, P. Salembier, T. Sikora, “Introduction to MPEG-7: Multimedia Content Description Interface“, Wiley, pp. 231-241, June 2002.
- [3] ISO/IEC 15938-3: 2001, “Multimedia content description interface – Part 3 Visual”, Version 1.
- [4] 張勇富, ”以語料分析為主的鳥音辨識系統研究”, 國立東華大學碩士論文, 中華民國九十二年七月.
- [5] S. E. Anderson, A. S. Dave, and D. Margoliash, “Template-based automatic recognition of birdsong syllables from continuous recordings”, Journal of the Acoustical Society of America, Vol. 100, No. 2, pp. 1209-1219, Aug. 1996.
- [6] J. Kogan and D. Margliash, “Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study”, Journal of the Acoustical Society of America, Vol. 103, No. 4, pp. 2187-2196, Apr. 1998.
- [7] A. L. McIlraith and H. C. Card, ”Birdsong recognition with DSP and neural networks”, in Proceedings of IEEE Conference on Communications, Power, and Computing, Vol. 2, pp. 409-414, May 1995.
- [8] A. L. McIlraith and H. C. Card, ”A comparison of backpropagation and statistical classifiers for bird identification”, in Proceedings of IEEE International Conference on Neural Networks, Vol. 1, pp. 100-104, June 1997.

[9] A. L. McIlraith and H. C. Card, "Birdsong recognition using backpropagation and multivariate statistics", IEEE Trans. on Signal Processing, Vol. 45, No. 11, pp. 2740-2748, Nov. 1997.

[10] A. L. McIlraith and H. C. Card, "Bird song identification using artificial neural networks and statistical analysis", in Proceedings of Canadian Conference on Electrical and Computer Engineering, Vol. 1, pp. 63-66, May 1997.