行政院國家科學委員會專題研究計畫 成果報告

自動化鳥類聲紋辨識之研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-2221-E-216-022-
執 行 期 間 ：95 年 08 月 01 日至 96 年 07 月 31 日
執 行 單 位 ：中華大學資訊工程學系

計 畫 主 持 人：李建興
共 同 主 持 人：李遠坤
計畫參與人員：碩士班研究生-兼任助理：莊清乾、魏銘輝
共同主持人：李遠坤

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96 年 10 月 31 日

行政院國家科學委員會補助專題研究計畫▪成果報告　□期中進度報告

自動化鳥類聲紋辨識之研究

計畫類別：☑ 個別型計畫　　☐ 整合型計畫
計畫編號：NSC 95-2213-E-216-022-
執行期間：2006 年 08 月 01 日 至 2007 年 07 月 31 日

計畫主持人：李建興
共同主持人：李遠坤
計畫參與人員： 莊清乾、魏銘輝

成果報告類型(依經費核定清單規定繳交)：☑精簡報告　☐完整報告

本成果報告包括以下應繳交之附件：
☐赴國外出差或研習心得報告一份
☐赴大陸地區出差或研習心得報告一份
☐出席國際學術會議心得報告及發表之論文各一份
☐國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
☐涉及專利或其他智慧財產權，☐一年☐二年後可公開查詢

執行單位：中華大學資訊工程學系
中　華　民　國　96　年　10　月　30　日

摘要

本計劃提出一自動辨識鳥類鳴叫聲音之方法、對於一輸入之鳥類聲音，我們首先將此聲音之每一音節切取出來，然後以整個音節之二維倒頻譜係數為此音節之特徵向量，二維倒頻譜係數能夠表現一個音節裡聲音頻譜圖裡之靜態(static)和動態(dynamic)特性，也就是說可以表現出一個音節整體的頻率變化和細微的頻率變化；另外，二維倒頻譜係數還能夠同時解決音節長度不同的問題，因為在二維倒頻譜係數中真正有意義的是分佈於低頻的係數，所以只要取固定數目分佈在低頻的係數來辨識就可得到極佳之辨識結果。由於鳥類鳴叫聲音相當豐富多變化，即使是同一種鳥類所發出之聲音音節也有極大的差異，因此我們採用一個自動分群的方法，把屬於同一種鳥類聲音音節細分成幾個小群，因此屬於同一小群之不同音節其特徵向量會較相似，所以對於同一種鳥鳴聲，我們將採用多組特徵向量來表示其不同之音節特性。最後以線性區別分析演算法來提升辨識之正確率，此一演算法可以縮小同類特徵向量之距離而且加大不同種類特徵向量之距離，因此可以在減少特徵向量維度之情況下提高辨識率。

一. 報告內容

1. 前言

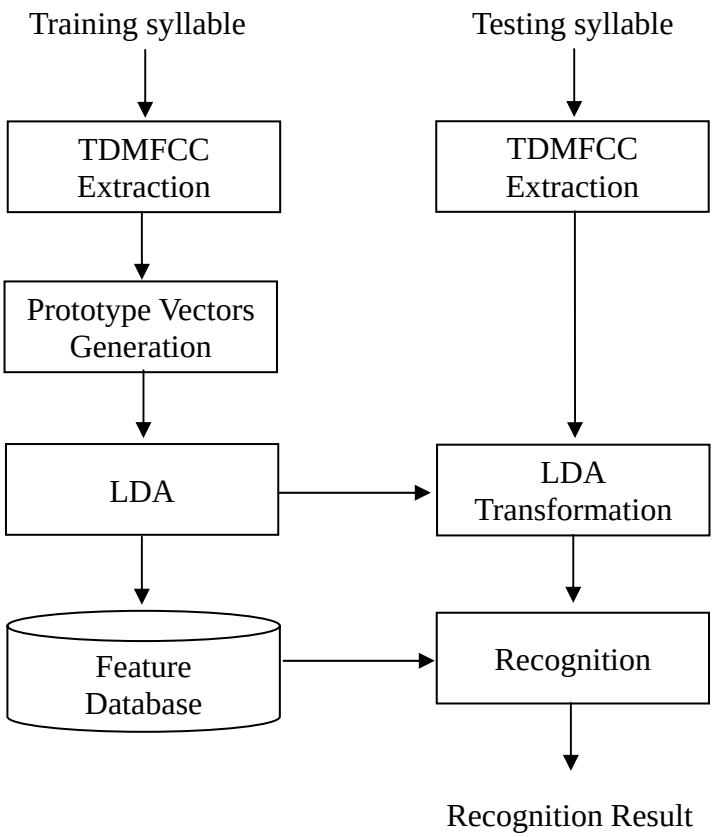
在動物叫聲的辨識中，鳥類鳴叫聲音(bird song/call)的辨識為最多人所研究。目前全世界的鳥類約有 9,200 種，臺灣已列入正式記錄的鳥類約有 450 種(中華民國野鳥學會 1995)，在分類學上分別隸屬於 18 目 68 科。由於種類繁多，不同物種間的棲息環境及生活方式也都有所差異，因此眾多研究人員投入研究生物叫聲的差異性，希望依此發現新的物種，然而目前所使用生物聲音的辨識方法，多採用人工至野外錄音，再回實驗室做人工的識別，如此相當耗時且耗力，因此本計劃擬對鳥類鳴叫聲音之自動辨識做一深入之研究。

鳥類沒有像人類般的聲帶(larynx)，其主要之發聲器官稱為鳴管(syrinx)，和人類的聲帶位置比起來，鳴管位在鳥類胸腔的更深處。鳥類的鳴管是成對的，在鳥類胸腔的深處氣管分成兩條支氣管，而鳴管有一部份在兩個支氣管內，且能發出聲音，這表示鳥類可以同時發出兩種不同的聲音，甚至可以自己鳴唱二重奏。所以不同鳥類可以不同的方式來發出成雙的鳴唱合弦。然而並不是所有的鳥類都會歌唱，也不是所有鳥類所發出的聲音都稱做歌曲(song) [1]。通常歌唱的能力僅限於燕雀目(Passeriformers)或棲息性(perching)的鳥類，所以世界上有將近一半的鳥種不會歌唱。幾乎會歌唱的鳥類都是雄鳥，而且雄鳥並不是整天或整年鳴唱，鳥兒們如何選擇歌唱之時辰及地點是觀察鳥類行為之重要指標，而其歌唱之主要目的有兩種：吸引雌鳥及宣告領土主權。就吸引雌鳥而言，雄鳥必須先確認雌鳥可以很容易的就聽見他們的歌聲，因此不同的鳥類會調整使用不同的方式來歌唱。就宣告領土主權而言，雄鳥之歌聲同樣需要被其他的雄鳥聽見，其目的有點像是在說：這個領土已經被佔領而且主人正在家裡。針對鳥類歌聲而言，其聲音結構是較複雜的，通常是將鳥類歌聲表示成一階層式之結構[2]，其中最簡單的一個鳥類聲音單元稱為音素(element)或是音調

(note)，一系列連續出現且具規律模式的 element 稱為音節(syllable)，而一連串的音節又組成了樂旨(motif)或是樂句(phrase)，一些重覆出現的 motif 的組合，就構成了歌聲的樂型(type)，最後，由一個或多個靜音區段隔開之 motif 則組成所謂的樂曲(bout)。

2. 研究目的與研究方法

關於鳥類鳴叫聲音之自動辨識之相關研究越來越多[3-12] ，本計劃是以整個音節之二維倒頻譜係數為此音節之特徵向量來區別不同種類之鳥類鳴叫聲音。本計劃之鳥類鳴聲自動辨識系統包含兩個階段，分別為訓練階段(training phase)和辨識階段(recognition phase)，訓練階段是由四個主要模組所組成:音節切割(syllable segmentation)、特徵擷取(feature extraction)、代表向量生成(prototype vector generation)和線性區別分析(linear discriminant analysis, LDA)。辨識階段是由四個主要模組所組成：音節分割、特徵擷取、線性區別分析轉換(LDA transformation)和分類(classification)。圖一為本計劃之系統架構圖。



圖一 鳥類鳴聲自動辨織系統的架構圖

2.1 音節切割

當輸入一段生物聲音訊號時，首先將一個個音節切割出來，每一個音節視為辨識系統之基本的辨識單元。而選擇以音節當成辨識單元，是因為當所輸入聲音訊號同時有許多不同種類的動物聲音時，要切出一個個的音節是比較簡單的，除此之外，利用音節所擷取出來的特徵值較為穩定。在這裡我們是依據 Harma 之方法以頻率上的資訊來完成音節切割[9]，其詳細步驟如下:

Step 1. 利用 short-time Fourier transform(STFT)建立輸入生物訊號的頻譜，我們用一個矩陣來表示此頻譜 $M(f, t)$ ，f 和 t 分別表示頻率值和時間的索引值。

Step 2. 設 n =0。

Step 3. 針對所有的 (f, t)，找出振幅強度之最大值，即 $|M(f_n, t_n)| \geq |M(f, t)|$，且將第 n 個音節的位置記錄為 (f_n, t_n)。

Step 4. 求得 $A_n(0) = 20\log_{10}|M(f_n, t_n)|$ dB，當 $A_n(0) < A_0(0) - \beta$ dB，停止切割動作，這代表第 n 個音節的強度太小了因此就不需再切割出任何音節了，β 為一門檻值(threshold)且其值設為 20。

Step 5. 從 (f_n, t_n) 的 t_n 位置找出 $t < t_n$ 的 $|M(f, t)|$ 最大值，直到 $A_n(t - t_n) < A_n(0) - \beta$ dB，同時也尋找出 $t > t_n$ 的 $|M(f, t)|$ 最大值，直到 $A_n(t - t_n) < A_n(0) - \beta$ dB。這個步驟是要找出第 n 個音節的起始時間 $(t_n - t_s)$ 和結束時間 $(t_n + t_e)$。

Step 6. 儲存第 n 個音節的時間點軌跡 $A_n(\tau)$, 其中 $\tau = t_n - t_s, ..., t_n + t_e$。

Step 7. 令 $M(f, [t_n - t_s, ..., t_n + t_s]) = 0$，目的是要將第 n 個音節的部份給清空，並令 n=n+1 且回到 **Step 3** 尋找下一個音節。

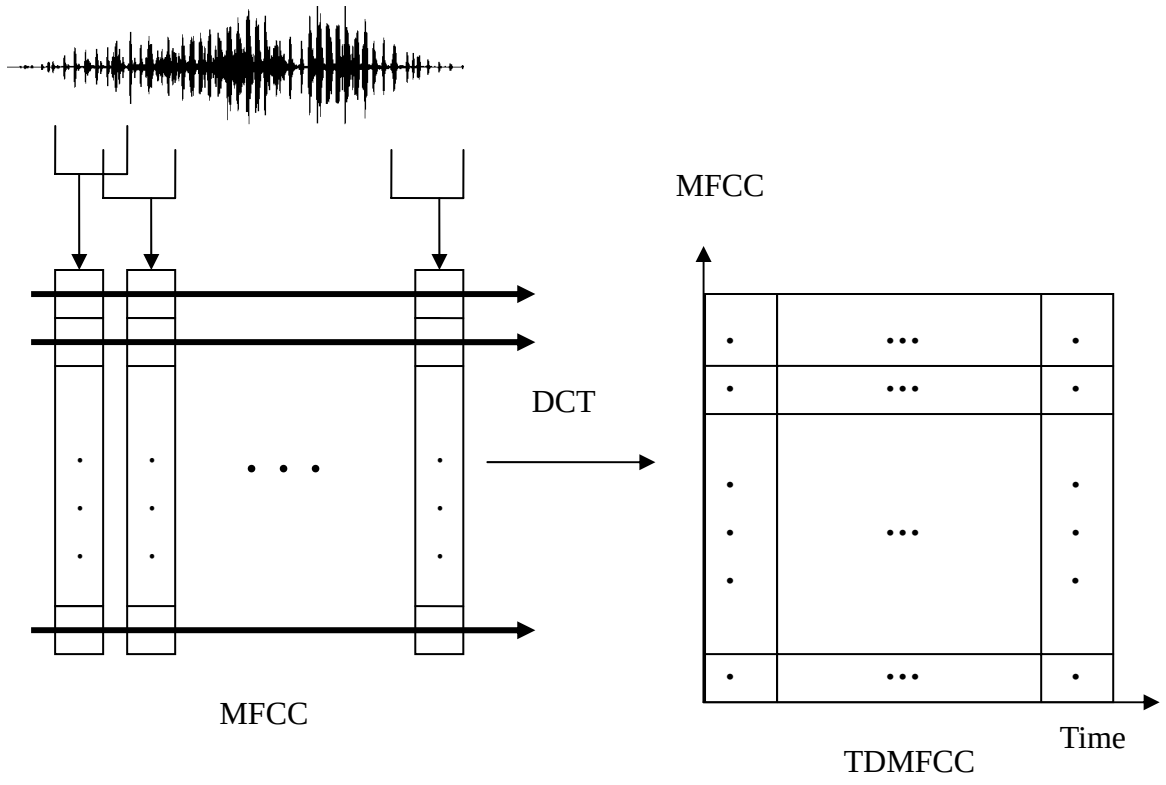
2.2 特徵擷取

對於切割出來之每一鳥類聲音之音節，我們將以二維梅爾倒頻譜係數(Two-dimensional Mel-scale Frequency Cepstral Coefficients, TDMFCCs）及動態二維梅爾倒頻譜係數(Dynamic TDMFCC, DTDMFCC)為此音節之特徵向量。

2.2.1 二維梅爾倒頻譜係數

二維倒頻譜係數(Two-dimensional cepstrum, TDC)已被用於語音辨識上[13-15]，主要原因是二維倒頻譜係數能夠表現出倒頻譜係數隨著時間的變化，對於描述相鄰音框特徵的關聯性是一個不錯的方法，另外也能表現一個音節裡聲音頻譜圖裡之靜態(static)和動態(dynamic)特性，也就是說可以表現出一個音節整體的頻率變化和細微的頻率變化；另外，

二維倒頻譜係數還能夠同時解決音節長度不同的問題，因為在二維倒頻譜係數中真正有意義的是分佈於低頻的係數，所以真正對語音辨識有幫助的是分佈在低頻的係數，而分佈在高頻的係數在語音辨識上是比較沒有意義的。因此我們擬採用二維梅爾倒頻譜係數來表示每一個隨時間改變其特性之鳥類鳴叫聲音，不只提供了梅爾倒頻譜係數的特性，也描述了梅爾倒頻譜係數隨著時間改變的特性。其做法是對各個音框之每一頻帶的對數能量頻譜值(logrithmic spectra)做二維離散餘弦轉換(discrete cosine transform, DCT)，由於二維離散餘弦轉換具有可分離特性(separability)，因此我們可以先對一音節內之每一音框計算其梅爾倒頻譜係數為此音框其特徵向量，再將這些梅爾倒頻譜係數依時間排成一矩陣之方式，針對同參數的梅爾倒頻譜係數做離散餘弦轉換，即可得到二維梅爾倒頻譜係數矩陣，其示意圖如圖二所示。



圖二 計算二維梅爾倒頻譜係數矩陣之流程圖

計算每一個音節之二維梅爾倒頻譜係數之詳細步驟如下：

步驟 1. **預強調 (Pre-emphasis)**

$$\hat{s}[n] = s[n] - \hat{a}s[n-1]$$

其中 $s[n]$為輸入訊號，$\hat{a}$ 的預設值為 0.95。

步驟 2. **取音框 (Framing)**

將每一個音節切割成一個一個的音框，大小為 512，而且為了讓每個音框的差異性不大，我們又讓每個音框重疊一半。

步驟 3. **乘上漢明視窗(Hamming Windowing)**

為了來消除每個音框與開始與結束的不連續性，每個音框都乘上一個漢明視窗，漢明視窗式子如下。

$$w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}),\ 0 \le n \le N-1$$

步驟 4. **快速傅立葉轉換(FFT)**

將音訊訊號從時域轉換成頻率域

$$X[k] = \sum_{n=0}^{N-1} \tilde{s}[n] e^{-j2\pi \frac{k}{N} n},\ 0 \le k < N,$$

其中 N 為音框大小。

步驟 5. **梅爾三角帶通濾波器(Mel-frequency Triangular band-pass filter)**

由於人耳對聲音的頻率的解析度不是呈線性關係，而是呈現對數(logarithm)變化，利用梅爾三角帶通濾波器將聲音訊號分成一個個頻帶，並算出每個頻帶的能量:

$$E_j = \sum_{k=0}^{K-1} \phi_j(k) A_k,\ 0 \le j < J,$$

J 為三角帶通濾波器之個數, A_k 為 $X[k]$的振幅:

$$A_k = |X[k]|^2,\ 0 \le k < N/2,$$

而ϕ_j 為第 j 個濾波器:

$$\phi_j[k] = \begin{cases} 0, & k \le I_l^j \text{ or } k \ge I_h^j \\ (k - I_l^j)/(I_c^j - I_l^j), & I_l^j \le k \le I_c^j \\ (I_h^j - k)/(I_h^j - I_c^j), & I_c^j \le k \le I_h^j \end{cases}$$

在這裡，I_l^j，I_c^j 和 I_h^j 分別代表第 j 個濾波器之低頻索引值，中間頻率索引值，和高頻索引值:

$$I_l^j = \frac{f_l^j}{(f_s/N)},\ I_c^j = \frac{f_c^j}{(f_s/N)},\ I_h^j = \frac{f_h^j}{(f_s/N)},$$

f_s 為取樣頻率，f_l^j，f_c^j，f_h^j 為第 j 個濾波器的低頻、中頻和高頻值，而每個濾波器的低頻、中頻和高頻值。

步驟 6. **二維離散餘弦轉換(Two-Dimensional Discrete Cosine Transform)**

我們利用二維離散餘弦轉換具有可分離特性，先對一音節內之每一音框計算其梅爾倒頻譜係數：

$$C_m^i = \sum_{j=0}^{J-1} \cos(m\frac{\pi}{J}(j+0.5))\log_{10}(E_j),\ \ 0 \le m \le L-1$$

其中 C_m^i 代表第 i 個音框之第 m 個梅爾倒頻譜係數，L 代表的是梅爾倒頻譜係數的個數。我們共用了 25 個三角濾波器，所以 J=25，而梅爾倒頻譜係數的長度為 15(L=15)。再將所有音框之梅爾倒頻譜係數依時間排成一矩陣之方式，針對同參數的梅爾倒頻譜係數再做一次離散餘弦轉換，即可得到二維梅爾倒頻譜係數矩陣(TDMFCCs)：

$$TDMFCC_m^k = \sum_{k=0}^{M-1} \cos(m\frac{\pi}{M}(k+0.5))C_m^k,\ \ 0 \le m \le L-1$$

其中 M 是一音節內之音框個數。因為在二維倒頻譜係數中真正對聲音辨識有幫助的是分佈在低頻的係數，因此我們只取前前幾個較低頻之二維梅爾倒頻譜係數為此音節之特徵向量。

2.2.2 動態二維梅爾倒頻譜係數

Furui 提出以動態特徵來辨識語音之方法[16]，其動態特徵是以迴歸係數(regression coefficient)來表現頻譜上的瞬間變化，應用在語者辨識中有著不錯的效果。其作法是對一段聲音切出數個音框，並對每個音框求出線性預估編碼(LPC)之後，將每個音框所求出線性預估編碼依時間排列，求出迴歸係數當做特徵並使用動態規畫比對演算法來辨識單詞語音，可以得到不錯的效果。令 $a_i(j)$表示在第 i 個音框之第 j 個迴歸係數，其計算方程式如下:

$$a_i(j) = \frac{\sum_{n=1}^{n_0} n(\left|E_{i+n}(j) - E_{i-n}(j)\right|)}{\sum_{n=-n_0}^{n_0} n^2},$$

$E_i(j)$表示在第 i 個音框之第 j 個線性預估編碼。

在動態二維梅爾倒頻譜係數中，我們利用迴歸係數求出在頻譜上的瞬間變化，而頻譜上的瞬間變化就像是在一張圖片中的邊線(edge)部份，也就是說如果把每一種類之鳥類鳴聲當成是一張特定的圖片，而這些圖片各自擁有獨特的邊線部份，這樣我們便能利用邊線部份進行辨識，所以我們便能利用迴歸係數來表示梅爾倒頻譜係數隨著時間變化之特性。

動態二維梅爾倒頻譜係數的做法是利用迴歸係數來當做一個高通濾波器求出頻譜中變化較大的部份，也就是說，對三角帶通濾波器之輸出值計算其迴歸係數，再去做二維離散餘弦轉換後便求得動態二維梅爾倒頻譜係數。

計算動態二維梅爾倒頻譜係數之詳細步驟如下:

步驟 1: 預強調 (Pre-emphasis)

$$\hat{s}[n] = s[n] - \hat{a}s[n-1],$$

$s[n]$ 為我們輸入訊號，$\hat{a}$ 的預設值為 0.95。

步驟 2: 取音框 (Framing)

將每一個音節切割成一個一個的音框，大小為 512，而且為了讓每個音框的差異性不大，我們又讓每個音框重疊一半。

步驟 3: 傅立葉轉換(DFT)

$$X_q[k] = \sum_{n=0}^{N-1} x_q[n]w[n]e^{-j2\pi\frac{k}{N}n},\ 0 \le k < N$$

其中 N 為音框大小，令 $x_q[n]$表示第 q 個音框之第 n 個訊號值，$X_q[k]$為第 q 個音框之第 k 個傅立葉係數，$w[n]$為漢明視窗(Hamming window)之第 n 個係數值：

$$w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}),\ 0 \le n < N$$

步驟 4: 三角帶通濾波器(Triangular band-pass filter)

利用三角帶通濾波器將聲音訊號分成一個個頻帶，並算出每個頻帶的能量:

$$E_j = \sum_{k=0}^{K-1} \phi_j(k) A_k \ ,\ 0 \le j \le J \ 。$$

步驟 5: **計算迴歸係數**

令 $E_i(j)$表示第 i 個音框之第 j 個三角帶通濾波器輸出值,將所有音框之三角帶通濾波器之輸出值依時間順序排列，計算其迴歸係數 $a_i(j)$:

$$a_i(j) = \frac{\sum_{n=1}^{n_0} n(\left|E_{i+n}(j) - E_{i-n}(j)\right|)}{\sum_{n=-n_0}^{n_0} n^2}, 0 \le j \le J \ 。$$

步驟 6: 離散餘弦轉換(Discrete Cosine Transform)

對這些迴歸係數乘上不同的餘弦值，求出動態梅爾倒頻譜係數 $C_i'(m)$:

$$C_i'(m) = \sum_{j=0}^{J-1} \cos(m\frac{\pi}{J}(j+0.5))\log_{10}(a_i(j)),\quad 0 \le m \le L-1$$

步驟 7: 對同索引值係數沿時間軸做離散餘弦轉換

令 $CC_q'(m)$ 為對所有 $C_i'(m)$ 沿著時間軸做離散餘弦轉換得到的動態二維梅爾倒頻譜係數，式子如下:

$$CC'_q(m) = \frac{1}{M-2}\sum_{i=1}^{M-2} C'_i(m)\cos(2\pi iq/M),$$

其中 q 表時間軸，$1 \le q \le M-2$，M 為音節音框總數。另外，在選取 $C'_i(m)$ 參數當作特徵時，本計劃只要取時間軸的前五個索引值，也就是動態二維梅爾倒頻譜係數區塊大小為 15×5。

對於二維梅爾倒頻譜係數或動態二維梅爾倒頻譜係數有特徵值範圍大小不同之問題，所以我利用正規化來解決這個問題，令 $F(n)$為由二維梅爾倒頻譜係數或者是動態二維梅爾倒頻譜係數組成之特徵向量，其正規化計算公式如下:

$$\hat{F}(n) = \frac{F(n) - F_{\min}(n)}{F_{\max}(n) - F_{\min}(n)},$$

其中，$\hat{F}(n)$ 為正規化後之特徵向量，$F_{\max}(n)$和 $F_{\min}(n)$ 為第 n 個特徵值之最大值和最小值。

2.3 特徵向量篩選(Feature selection)

在辨識時，因為特徵向量維度太大會影響辨識時間且辨識率不見得會較高，所以我們先對原始之特徵向量先做初步之篩選，以降低特徵向量之維度，我們採用之篩選方法是 sequential forward floating search (SFFS)演算法[17]，評估所選取之特徵向量之標準如下：

$$J(K) = \frac{\det(S_M)}{\det(S_W)},$$

其中 $J(K)$代表選取 K 個特徵值之評估值，S_M和 S_W分別代表的是混合散佈矩陣(mixture scatter matrix)和同類別之散佈矩陣(within-class scatter matrix)，而混合散佈矩陣 S_M及同類別之散佈矩陣 S_W之定義如下：

$$S_M = \sum_{j=1}^{C} N_i(\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^{\mathrm{T}},$$

$$S_W = \sum_{j=1}^{C}\sum_{i=1}^{N_i} (\mathbf{x}_i^j - \boldsymbol{\mu}_j)(\mathbf{x}_i^j - \boldsymbol{\mu}_j)^{\mathrm{T}},$$

而 $\mathbf{x}_i^j$ 代表在類別 j 中的第 i 個特徵向量，$\boldsymbol{\mu}_j$ 為第 j 類的平均向量(mean vector)，C 為類別的數目，N_j 為類別 j 裡的特徵向量個數，$\boldsymbol{\mu}_j$ 為所有特徵向量的平均向量。依據 SFFS 演算法，最後篩選之特徵值數目 d 是以下列方程式決定：

$$d = \arg\min_{1 \le k \le K} \frac{J(k+1) - J(k)}{J(k+1)} < \gamma,,$$

其中γ為大於 0 之常數值。

2.4 代表向量生成

由於鳥類鳴叫聲音相當豐富多變化，因此就算有兩個音節是從同一種鳥類聲音中所切割出來的，所擷取出來的特徵向量也可能會有明顯的不同，所以對於每一種鳥類聲音，我們將使用一個編碼簿(codebook)包含好幾個特徵向量來表示一種鳥類的鳴叫聲，因此屬於同一種鳥類聲音之不同音節可以分成幾個小群，而屬於同一小群之不同音節其特徵向量會較相似，所以同一種鳥類聲音必須以好幾個特徵向量來表示。

在本計畫中，我們將採用一個向量量化法(Vector Quantization, VQ)的分群方法，稱為 progressive constructive clustering (PCC) algorithm [18]，自動把屬於同一種鳥類聲音細分成幾個小群，令 $S^j = \{\mathbf{x}_1^j, \mathbf{x}_2^j, \cdots, \mathbf{x}_{N_j}^j\}$ 為所有第 j 種鳥類的特徵向量之集合，N_j 為 S^j 的大小，而 PCC 演算法的敘述如下：

步驟 **1.** 取 $\mathbf{x}_1^j$ 為群別 1 的中心點 $(\mathbf{c}_1)$。令 $i = 2$ 且 $nc = 1$。

步驟 **2.** 取特徵向量 $\mathbf{x}_i^j$ 和集合 $\{\mathbf{c}_1, \ldots, \mathbf{c}_{nc}\}$ 中每一個向量做比較以便決定其屬於那一群，比較方法為：令 $\mathbf{c}_k$ 和 $\mathbf{x}_i^j$ 的距離最短且表示為 $d(\mathbf{x}_i^j, \mathbf{c}_k)$，當 $d(\mathbf{x}_i^j, \mathbf{c}_k) \le T_d$ 時將 $\mathbf{x}_i^j$ 歸類為群別 k，如果不是，就跳到步驟 **4**。其中 T_d 之定義為：

$$T_d = \alpha \times 0.001 \times d,$$

d 為特徵向量之長度，α為正整數。

步驟 **3.** 當群別 k 有加入新的成員時，便重新計算代表群別 k 的中心點，之後跳到步驟 **5**。

步驟 **4.** 令 $nc = nc + 1$ 並得到一個代表新群別的中心點 $\mathbf{c}_{nc} = \mathbf{x}_i^j$。

步驟 **5.** 如果 $i = N_j$ 便結束；否則令 $i = i+1$ 並回到步驟 **2**。

2.5 線性區別分析演算法(Linear Discriminant Analysis, LDA)

線性區別分析演算法之目的是將一個高維度的特徵向量轉換成一個低維度的向量，並且增加辨識的準確率[19]，線性區別分析演算法主要處理不同類別間的區別程度而不是用於不同類別之表示方式。線性區別分析演算法的主要精神是要把同類之間的距離最小化，並且把不同類別之間的距離給最大化，所以，必需決定一個轉換矩陣(transformation matrix)來將維度 n 的特徵向量轉換成維度 d 的向量，在這裡 $d \le n$，透過這樣的轉換我們能夠增強

不同類別之間的差異性。最常使用的轉換矩陣主要依據 Fisher criterion J_F 來求得:

$$J_F(A) = tr((A^{\mathrm{T}} S_W A)^{-1}(A^{\mathrm{T}} S_B A)),$$

其中，S_W 和 S_B 分別代表的是同類別之散佈矩陣(within-class scatter matrix)和不同類別之散佈矩陣(between-class scatter matrix)，而同類別之散佈矩陣的公式如下:

$$S_W = \sum_{j=1}^{C} \sum_{i=1}^{N_i} (\mathbf{x}_i^j - \boldsymbol{\mu}_j)(\mathbf{x}_i^j - \boldsymbol{\mu}_j)^{\mathrm{T}},$$

而 $\mathbf{x}_i^j$ 代表在類別 j 中的第 i 個特徵向量，$\boldsymbol{\mu}_j$ 為第 j 類的平均向量(mean vector)，C 為類別的數目，N_j 為類別 j 裡的特徵向量個數。而不同類別之散佈矩陣公式如下:

$$S_B = \sum_{j=1}^{C} (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^{\mathrm{T}},$$

$\boldsymbol{\mu}$ 為所有類別的平均向量。線性區別分析演算法的目的是要去求出能夠使不同類別之散佈矩陣和同類別之散佈矩陣的比值為最大值轉換矩陣(transformation matrix) A_{opt}，而其維度大小為 $n \times d$:

$$A_{opt} = \arg\max_{A} \frac{tr(A^{\mathrm{T}} S_B A)}{tr(A^{\mathrm{T}} S_W A)}.$$

此一轉換矩陣，可經由求出 $S_W^{-1} S_B$ 的 eigenvectors 來得到，而 A_{opt} 之 d 個行向量為前 d 個最大 eigenvalue 值所對應之 eigenvector。假設 eigenvalues 的順序為非遞增的，則所保留之 eigenvector 個數可由以下公式求得:

$$\sum_{i=1}^{d} \lambda_i \geq 0.95 \sum_{i=1}^{n} \lambda_i,$$

λ_i 為第 i 個 eigenvalue。

在我們決定出最佳的轉換矩陣 A_{opt} 後，我們以 A_{opt} 將每一 n 維的特徵向量轉換為 d 維之向量。令 $\mathbf{f}_j$ 為類別 j 裡維度為 n 的特徵向量，轉換成維度為 d 的向量之公式如下:

$$\mathbf{x}_j = A_{opt}^{\mathrm{T}} \mathbf{f}_j.$$

2.6 辨識階段

在辨識的部份中，在輸入每個鳥類鳴叫聲音後，首先將每一個音節切割出來，並求出每個音節的二維倒頻譜係數為此音節之特徵向量，我們同時利用轉換矩陣 A_{opt} 來將此特徵向量轉換成轉換成較低維度的特徵向量 $\mathbf{x}$，接著，計算此一向量和代表每一個代表向量($\mathbf{x}^k$, $k = 1, 2, \ldots, N$, 其中 N 為所有代表向量之個數，N 可能遠大於資料庫中鳥類之種類數目)之間的距離，令 $\mathbf{x}^r$ 與 $\mathbf{x}$ 之距離最小：

$$d(\mathbf{x}, \mathbf{x}^r) \le d(\mathbf{x}, \mathbf{x}^k), \quad 1 \le k \le N, \ k \ne r.$$

在這裡的距離公式是歐基里德距離(Euclidean distance)：

$$d(\mathbf{x}, \mathbf{x}^k) = \sum_{m=1}^{d} (x_m - x_m^k)^2$$

所辨認之鳥鳴聲種類代表編碼 s 即由 $\mathbf{x}^r$ 所屬之鳥類種類之特徵向量所形成之集合來決定：

$$s = i \ \text{ if } \ x^r \in G^i, \ 1 \le i \le N_s,$$

其中 G^i 表示所有用以代表第 i 種鳥類之特徵向量所形成之集合，N_s 代表資料庫中鳥類之種類數目。

3. 實驗結果與討論

在實驗所用鳥類聲音資料檔案，取樣頻率為 44100 Hz，音訊範圍大小為 16 bits，主要來源是聲音光碟及網際網路中，總共分成兩個資料庫。第一個資料庫有 420 種日本之鳥類鳴聲[20]，總共切割之音節數目為 29,754 個音節，在對資料庫中所有鳥類聲音擷取特徵進行辨識前，我們以頻譜能量之資訊對每個鳥類聲音檔案切出音節，隨機取一半之聲音音節做為訓練音節，而另一半之聲音音節為測試音節。第二個資料庫有 28 種在台灣錄音得到之鳥類鳴聲[21-23]，此 28 鳥類之詳細資料請參考表一，其中訓練及測試音節分別自不同聲音檔案中擷取出來，總計有 3,550 個訓練音節及 685 個測試音節。

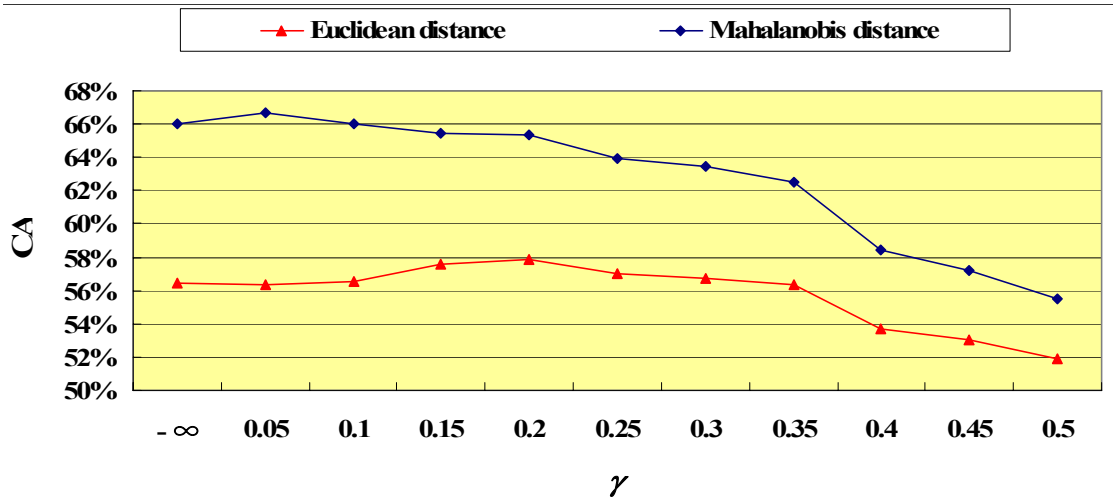
在實驗中，我們比較我們所提出的方法(STDMFCC 及 SDTDMFCC)與其他已知之方法(ALPCC 及 AMFCC)。首先，為了評估特徵向量篩選演算法對於辨識正確率之影響，我們對於 SFFS 演算法中所用到的參數值(γ)代入不同數值以評估其辨識正確率(如圖二)，由圖二可以看出當以 Mahalanobis distance 來計算兩特徵向量之距離時，其辨識正確率會比用 Euclidean distance 來得高，而且當 γ = 0.05 時可以得到最高之辨識正確率，然後隨著 γ 值變大而遞減。另外，以 Mahalanobis distance 來計算兩特徵向量之距離時，當 γ = 0.2 時可以得到最高之辨識正確率，然後一樣隨著 γ 值變大而遞減。表二顯示針對不同 γ 值所篩選保留之特徵向量維度，由此表可以得知利用 SFFS 特徵向量篩選演算法可以將特徵向量維度降低並且提高辨識正確率。

表一 第二個資料庫之鳥類鳴聲資料

Common name	Latin name	Training syllables	Test syllables
Crested Serpent Eagle	Spilornis cheela	12	5
Bronzed Drongo	Dicrurus aeneus	266	60
Gray-headed Pygmy Woodpecker	Dendrocopos canicapillus	16	17
Blue Shortwing	Brachypteryx montana	307	25
Streak-breasted Scimitar Babbler	Pomatorhinus ruficollis	105	24
Taiwan Firecrest	Regulus goodfellowi	213	56
Taiwan Sibia	Heterophasia auricularis	84	12
White-throated Laughing Thrush	Garrulax albogularis	233	77
White-breasted Water Hen	Amaurornis phoenicurus	155	14
Beavan's Bullfinch	Pyrrhula erythaca	73	9
Gray sided Laughing Thrush	Garrulax caerulatus	42	41
Alpine Accentor	Prunella collaris	112	31
Green-backed Tit	Parus monticolus	140	14
Taiwan Yuhina	Yuhina brunneiceps	62	14
Red-headed Tit	Aegithalos concinnus	139	30
Collared Bush Robin	Erithacus johnstoniae	278	15
Taiwan Bulbul	Pycnonotus taivanus Styan	78	24
Taiwan Hill Partridge	Arborophila crudigularis	190	33
Verreaux's Bush Warbler	Cettia acanthizoides	60	7
Oriental Cuckoo	Cuculus saturatus	323	39
Taiwan Tit	Parus holsti	214	26
Vivid Niltava	Niltava vivida	92	15
Coal Tit	Parus ater	173	33
Crested Goshawk	Accipiter trivirgatus	34	16
Gould's Fulvetta	Alcippe brunnea	32	20
Collared Pigmy Owlet	Glaucidium brodiei	73	14
Swinhoe's Pheasant	Lophura swinhoii	24	6
Steere's Liocichla	Liocichla steerii	20	8
Total syllable number		3550	685

表二 對於不同γ值所篩選保留之特徵向量維度

γ	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
feature number	45-46	28-30	22	19	16	15	14	10-11	10	9



圖二 在 SFFS 演算法中，參數值(γ)與辨識正確率之關係

另外，為了評估 PCC 分群演算法對於辨識正確率之影響，我們對於 PCC 演算法中所用到的參數值(α)代入不同數值以評估其辨識正確率(如表三)，由表三及圖二可以看出當以較多組特徵向量來代表一種鳥類鳴聲時，其辨識正確率可以提高約 20%。

為了評估 LDA 演算法對於辨識正確率之影響，我們對於 PCC 演算法後再加上 LDA 演算法中，此一實驗結果顯示於表四中，由表四可以看出以 Mahalanobis distance 或 Euclidean distance 來計算兩特徵向量之距離時，其辨識正確率是相同，主要的原因是在 LDA 演算法中我們加入了 whitening 之步驟，使得每一特徵值之變異數值是一樣的。

最後，表五比較這幾種方法對於第一個資料庫之辨識正確率，由此一表中，我們可看出 STDMFCC 之正確率比 ALPCC 及 AMFCC 高，而且當加上 DTDMFCC 時(即 SDTDMFCC)，其辨識正確率最高可達 89.63%。表六則比較這幾種方法對於第二個資料庫之辨識正確率，由此一表中，我們可看出 STDMFCC 之正確率比 ALPCC 及 AMFCC 高，而且當加上 DTDMFCC 時(即 SDTDMFCC)，其辨識正確率最高可達 79.01%。

表三 分群參數值(α)及特徵向量篩選參數值(γ)對辨識正確率之影響

Distance metric	γ	α				
		1	2	3	4	5
Euclidean distance	0.05	80.23	80.77	80.81	80.33	79.47
	0.10	84.42	84.62	84.02	83.61	82.33
	0.15	85.83	85.78	84.88	84.44	83.30
	0.20	86.41	86.30	85.54	84.77	83.56
	0.25	86.46	86.38	85.59	84.29	83.15
Mahalanobis distance	0.05	84.68	86.06	86.15	85.57	84.84
	0.10	87.63	88.43	88.45	87.87	86.93
	0.15	88.42	89.12	88.78	88.47	87.81
	0.20	88.61	89.18	88.85	88.49	87.28
	0.25	88.50	88.86	88.54	87.71	86.73

表四 加上LDA演算法對辨識正確率之影響

Distance metric	γ	α				
		1	2	3	4	5
Euclidean distance	0.05	85.34	86.93	86.88	86.16	85.52
	0.10	87.48	88.33	88.28	88.04	86.85
	0.15	87.84	88.55	88.47	88.08	87.29
	0.20	87.85	88.46	88.73	87.64	86.58
	0.25	87.07	87.88	87.39	86.57	85.54
Mahalanobis distance	0.05	85.34	86.93	86.88	86.16	85.52
	0.10	87.48	88.33	88.28	88.04	86.85
	0.15	87.84	88.55	88.47	88.08	87.29
	0.20	87.85	88.46	88.73	87.64	86.58
	0.25	87.07	87.88	87.39	86.57	85.54

表五 對於第一個資料庫鳥類鳴聲檔案之辨識正確率

Algorithm	α				
	1	2	3	4	5
ALPCC	70.72	66.84	62.83	60.66	58.82
AMFCC	86.98	87.27	86.70	86.46	85.16
STDMFCC	87.13	88.07	88.09	87.48	86.56
DTDMFCC	84.69	85.11	84.43	83.36	82.10
SDTDMFCC	88.79	89.56	89.63	89.17	88.39

表六 對於第二個資料庫鳥類鳴聲檔案之辨識正確率

Algorithm	α							
	25	30	35	40	45	50	55	60
ALPCC	22.34	21.17	20.73	20.44	20.73	20.58	19.42	20.44
AMFCC	60.29	59.12	59.71	60.58	58.69	58.83	59.42	57.37
STDMFCC	77.23	77.81	78.03	77.63	76.72	76.35	77.59	76.50
DTDMFCC	78.28	77.59	77.92	76.50	77.45	77.63	77.63	77.04
SDTDMFCC	77.41	76.57	76.31	77.45	79.01	79.27	78.69	77.99

二. 參考文獻

[1] http://www.earthlife.net/birds/song.html

[2] E. A. Brenowitz, D. Margoliash, and K. M. Nordeen, “An introduction to birdsong and the avian song system”, *Journal of Neurobiology*, Vol. 33, Issue 5, pp. 495-500, Nov. 1997.

[3] 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.

[4] J. Kogan and D. Margoliash, “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.

[5] A. L. McIlraith and H. C. Card, “Birdsong recognition with DSP and neural networks”, in

Proc. of IEEE Conf. on Communications, Power, and Computing, Vol. 2, pp. 409-414, May 1995.

[6] A. L. McIlraith and H. C. Card, "A comparison of backpropagation and statistical classifiers for bird identification", in *Proc. of IEEE Int. Conf. on Neural Networks*, Vol. 1, pp. 100-104, June 1997.

[7] 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.

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

[9] A. Harma, "Automatic identification of bird species based on sinusoidal modeling of syllables", in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 545-548, 2003.

[10] A. Harma and P. Somervuo, "Classification of the harmonic structure in bird vocalization", in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 701-704, 2004.

[11] P. Somervuo and A. Harma, "Bird song recognition based on syllable pair histograms", in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 825-828, 2004.

[12] P. Somervuo, A. Harma, and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition", *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 14, No. 6, Nov. 2006, pp. 2252-2263.

[13] Y. Ariki, S. Mizuta, M. Magata, and T. Sakai, "Spoken-word recognition using dynamic features analysed by two-dimensional cepstrum", *IEE Proceedings, Pt. I*, No. 2, Apr. 1998, pp. 133-140.

[14] H. F. Pai and H. C. Wang, "A study of the two-dimensional cepstrum approach for speech recognition", *Computer Speech and Language*, No. 6, 1992, pp. 361-375.

[15] C. T. Lin, H. W. Nein, and J. Y. Hwu, "GA-based noisy speech recognition using two-dimensional cepstrum", *IEEE Trans. on Speech and Audio Processing*, Vol. 8, No. 6, Nov. 2000, pp. 664-675.

[16] S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 1, Feb. 1986, pp. 52-59.

[17] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection",

Pattern Recognition Letters, Vol. 15, 1994, pp. 1119-1125.

[18] N. Akrout, C. Diab, R. Prost, and R. Goutte, "A fast algorithm for vector quantization: application to codebook generation in image subband coding", *Signal Processing VI*: *Theories and Application*, Vol. 3, 1992, pp. 1227-1230.

[19] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. New York:Wiley, 2000.

[20] T. Kabaya and M. Matsuda, *The Songs and Calls of 420 Birds in Japan*. Tokyo: Shogakukan, 2001.

[21] Yushan National Park, *CD Sound of the Mountain IV: The songs of Wild Birds*. Taiwan, 1995.

[22] Yushan National Park, *CD Sound of the Mountain V: The songs of Wild Birds*. Taiwan, 1996.

[23] http://www.fhk.gov.tw

三. 計畫成果自評

建立生物多樣性資料庫是推動生物保育、教育及研究的重要基礎工作。於「生物多樣性公約」之第十七條即要求各國需成立生物多樣性資訊之交換中心，積極蒐集整理本土生物多樣性之資料，並與其他國家分享，以促進生物多樣性之保育、利用、管理、研究及教育，同時也可提振各國分類學的能力建設。台灣的土地面積雖不大，卻擁有異常豐富的生物多樣性資源，特有生物種類繁多，臺灣已列入正式紀錄的鳥類約有 450 種(中華民國野鳥學會 1995)，青蛙種類約有 31 種，蟬類有 59 種(4 種新種)，台灣蟋蟀種類約有八十幾種，以聲音自動辨識系統來記錄生物的棲息環境，不僅有助於了解這些生物的生態變化，並能減少對生態的影響。

利用生物聲音去辨識物種做生態評估是近來常用的方法，其可應用在生物種類統計(ecological censusing)、環境監控(environment monitoring)和生物多樣性評估(biodiversity assessment)等方面。數位內容及數位博物館為政府近幾年來發展之重點產業，其內容包括圖像、字元、影像、語音等資料加以數位化並整合運用。在數位博物館中，本土生物之聲音資料庫的建立成了其中重要而基本的工作。在生物叫聲的辨識中，鳥類鳴叫聲音(bird song/call)的辨識又為最多人所研究。目前全世界的鳥類約有 9,200 種，臺灣已列入正式記錄的鳥類約有 450 種，在分類學上分別隸屬於 18 目 68 科。由於種類繁多，在做生態調查時，若以人工的方式來進行，相當耗時且耗力，因此本計劃擬對鳥類鳴叫聲音之自動辨識做一深入之研究以輔助鳥類族群之生態、棲地之變化，並能減少對生態的影響。自動生物聲音辨識的研究，尤其是國內，進行的仍舊很少，因此我們希望透過自動辨識系統配合適當的硬

體設備，能發現更多未曾記載的生物物種，建立更完善的台灣生物聲音資料庫，配合政府發展數位內容及數位博物館的計畫。本計畫已完成可自動辨識鳥類鳴聲之辨識系統，但如何適應各種不同之錄音環境及錄音器材以提高辨識率是未來希望能進一步改善之方向。

目前我們已發表之相關論文如下：

期刊論文 (Journal Papers)：

[1] C. H. Lee, C. H. Chou, C. H. Han, and R. Z. Huang, "Automatic Recognition of Animal Vocalizations Using Averaged MFCC and Linear Discriminant Analysis", *Pattern Recognition Letters*, Vol. 27, Issue 2, Jan. 2006, pp. 93-101. (SCI, EI)

[2] C. H. Lee, Y. K. Lee and R. Z. Huang, "Automatic recognition of bird songs using cepstral coefficients", *Journal of Information Technology and Applications*, Vol. 1, No. 1, May 2006, pp. 17-23.

[3] C. H. Lee, C. C. Han, and S. F. Lin, "Automatic Identification of Bird Species by Their Sounds Using Two Dimensional Cepstral Coefficients", prepare for submission to *Pattern Recognition*.

研討會論文 (Conference Papers)：

[1] C. H. Lee, C. H. Chou, and R. Z. Huang, "Automatic Recognition of Bioacoustic Sounds: an Experiment on the Frog Vocalizations", in *Proceedings of the 17th IPPR Conference on Computer Vision, Graphics, and Image Processing*, Hualien, Aug. 15-17, 2004.

[2] C. H. Lee, C. H. Chou, C. C. Han, and R. Z. Huang, "Automatic Recognition of Frog Calls Using Averaged MFCC and Linear Discriminant Analysis", in *Proceedings of the 9th Conference on Artificial Intelligence and Applications*, Taipei, Nov. 5-6, 2004.

[3] C. H. Lee, C. C. Lien and R. Z. Huang, "Automatic Recognition of Birdsongs Using Mel-frequency Cepstral Coefficients and Vector Quantization", in *Proceedings of International MultiConference of Engineering and Computer Scientists*, Hong Kong, 2006, pp. 331-335.

[4] C. H. Chou, C. H. Lee and H. W. Ni, "Bird Species Recognition by Comparing the HMMs of the Syllables", in *Proceedings of Second International Conference on Innovative Computing, Information and Control*, Kumamoto, Japan, Sep. 5-7, 2007.