行政院國家科學委員會專題研究計畫 成果報告

強韌性連續鳥類鳴聲自動辨識之研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2221-E-216-043-
執 行 期 間 ：96 年 08 月 01 日至 97 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：連振昌
計畫參與人員：碩士班研究生-兼任助理人員：林懷三
碩士班研究生-兼任助理人員：魏銘輝
碩士班研究生-兼任助理人員：李筱萱

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

中 華 民 國 97 年 10 月 27 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

強韌性連續鳥類鳴聲自動辨識之研究

計畫類別：☑ 個別型計畫　　☐ 整合型計畫
計畫編號：NSC 96-2221-E-216-043-
執行期間：2007 年 08 月 01 日 至 2008 年 07 月 31 日

計畫主持人：李建興
共同主持人：連振昌
計畫參與人員： 林懷三、魏銘輝、李筱萱

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　97　年　10　月　30　日

摘要

本計劃對鳥類鳴叫聲音之自動辨識做一深入之研究，以輔助調查鳥類族群之生態、棲地之變化，並能減少對生態的影響。首先將輸入之鳥類聲音中之每一音節切取出來，然後我們提出結合倒頻譜濾波法(cepstral filtering)及調變頻譜過濾演算法(modulation spectrum filtering)來濾除聲音訊號之雜訊。接著以整個音節之二維倒頻譜係數、動態二維梅爾倒頻譜係數及 MPEG-7 之聲音頻譜封包(Normalized Audio Spectrum Envelope, NASE)之調變係數為此音節之特徵向量，然後以主軸分析演算法(Principal Component Analysis, PCA)來降低特徵向量之維度，對於每一種鳥類聲音，我們使用高斯混合模型(Gaussian mixture model, GMM)來描述並得到同一種鳥類聲音之代表特徵向量群組，再以線性區別分析演算法(Linear Discriminant Analysis, LDA)來提升辨識之正確率，最後將各種特徵向量整合在一起，進一步提高辨識率，當結合三種特徵向量時可得到之最佳辨識正確率為 83.90%。

一. 報告內容

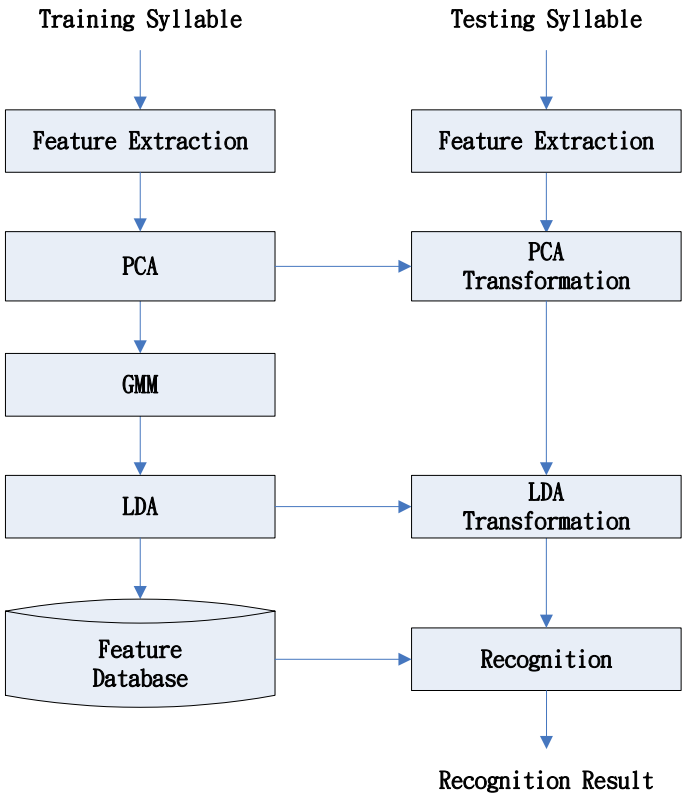
1. 前言

全球生物多樣性資訊機構(GBIF)已在 2001 年正式成立，其目的在配合生物多樣性國際公約之要求，推動全球各國成立生物多樣性資訊交換中心，以進行生物多樣性資料之蒐集、整理與保存，使各國生物多樣性之資訊可與全球其他國家分享，促進生物多樣性之保育、利用、管理、研究及教育。國科會於 2001 年起推動建立「台灣生物多樣性國家資訊網」計畫 (Taiwan Biodiversity National Information Network, TaiBNET)，目前已完成本地生物多樣性專家名錄及台灣物種名錄兩個資料庫。此外國科會正在推動之「數位典藏國家型科技計畫」，其重點在將典藏之生物標本或文獻解說等基本資料予以數位化，農委會多年來也持續在推動全省分年分區生態之調查與時間空間分佈資料之數位化工作。從事生態調查之工作通常是由專家或具備豐富野外調查經驗之人員來執行，一般而言，都是依據視覺上(外形、顏色等)及聽覺上(聲紋)之特徵來辨識動物種類，但是對於某些動物，其習性隱匿不易觀察，若想見其行蹤，則是難上加難，例如八色鳥是國際鳥類聯盟公佈的全球遭受威脅鳥種之一，估計全球數量可能不到一萬隻，在台灣也十分稀少，而鳴叫聲是其互相溝通聯繫的重要工具，通常我們在野外比較容易聽到其叫聲而不易見其形體，此外生物的叫聲早已進化成與特定之物種相關，也就是不同之物種之聲音會有所不同，因此利用生物的鳴叫聲來辨識生物種類是相當自然且有效可行的方法，

可以幫助生態調查者確認生物之種類及其分佈定位。因此本計劃對鳥類鳴叫聲音之自動辨識做一深入之研究，以輔助調查鳥類族群之生態、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，本計畫希望透過此自動辨識系統配合適當的硬體設備，建立更完善的台灣鳥類聲音資料庫。

2. 研究目的與研究方法

近幾年來，自動辨識鳥類鳴叫聲音之相關研究越來越多[1-17] ，本計劃提出一套強韌性之鳥類音節辨識系統。此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)，訓練階段是由五個主要模組所組成:音節切割(syllable segmentation)、特徵擷取(feature extraction)、主軸分析演算法(Principal Component Analysis, PCA)、代表向量生成(prototype vector generation)和線性區別分析演算法(linear discriminant analysis, LDA)。辨識階段是由五個主要模組所組成：音節分割、特徵擷取、主軸分析轉換(PCA transformation)、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 鳥類鳴聲自動辨識系統的架構圖

2.1 聲音訊號強化

本計劃擬提出之聲音訊號強化演算法是結合倒頻譜濾波法(cepstral filtering)及調變頻譜演算法(modulation spectrum filtering)來濾除聲音訊號之雜訊。首先，我們將時間域之鳥類聲音訊號以傅立葉轉換得到其頻譜，接著對頻譜係數取對數可得到對數頻譜(log magnitude spectrum)，然後將對數頻譜做餘弦轉換可得到其倒頻譜係數，之後對倒頻譜係數做低通濾波計算，再做反餘弦轉換得到回復之對數頻譜，此一步驟之目的是對聲音頻譜做平滑化(smoothing)以避免通道失真之影響，然後我們將以調變頻譜濾波法來消除環境雜訊，首先，對回復之對數頻譜之同一頻率係數沿著時間軸做餘弦轉換成調變頻譜(modulation spectrum)，保留調變頻譜 1~16 赫茲(Hertz, Hz)範圍內的係數，再做反餘弦轉換可得到一個乾淨的頻譜，將此頻譜之係數正規化至[0, 1]，再與原始之頻譜相乘，即可對原始聲音訊號作強化以去除雜訊，圖二顯示原始有雜訊之鳥類鳴聲之頻譜及強化後之頻譜。其詳細的步驟如下：

步驟 1. 取音框 (frameing)大小為 512，重疊一半。

步驟 2. 傅立葉轉換

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

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 3. 取對數頻譜

$$M_q[k]=\log_{10}\left|X_q[k]\right|$$

其中 $M_q[k]$為對數頻譜係數

步驟 4. 餘弦轉換

對所有對數頻譜係數做餘弦轉換以得到倒頻譜係數：

$$C_q[n]=\sqrt{\frac{2}{N}}\mu[n]\sum_{k=0}^{N-1}M_q[k]\cos\left(\frac{\pi k(2n+1)}{2N}\right),\ 0\le n\le N\text{-}1$$

其中 $C_q[n]$為倒頻譜係數，$\mu[n]$之定義如下：

$$\mu[n]=\begin{cases}\dfrac{1}{\sqrt{2}}, & n=0\\ 1, & 1\le n<N\end{cases}$$

步驟 5. 將 DC 值及高頻之倒頻譜係數濾除，僅保留一半低頻之倒頻譜係數值：

$$C_q^{'}[n]=\begin{cases}C_q[n], & 1\le n\le N/2\\ 0\ , & \text{otherwise}\end{cases}$$

步驟 6. 反餘弦轉換

對倒頻譜係數做反餘弦轉換以得到更平滑之對數頻譜：

$$M_q^{'}[k]=\sqrt{\frac{2}{N}}\sum_{n=0}^{N-1}\mu[n]C_q^{'}[n]\cos\left(\frac{\pi k(2n+1)}{2N}\right),\ 0\le k\le N-1$$

步驟 7. 調變頻譜轉換

對 $M_q^{'}[k]$ 沿著時間軸取 P 個音框($P=86$ 約 1 秒之時間)之相同頻率之對數頻譜係數做餘弦轉換：

$$Y_n[k]=\sqrt{\frac{2}{P}}\mu[p]\sum_{p=0}^{P-1}M_{q+p}^{'}[k]\cos\left(\frac{\pi n(2p+1)}{2P}\right),\ n\in[0,P-1]$$

其中 $Y_n[k]$表示第 k 個頻率之第 n 個調變頻譜係數(調變頻率 $=n$ Hz)。

步驟 8. 將 DC 值及高頻之調變頻譜係數濾除，保留 1~16Hz 範圍之調變頻譜係數值，

步驟 9. 轉換回復至對數頻譜

對保留之調變頻譜係數做反餘弦轉換以得到乾淨之對數頻譜：

$$M_{q+n}^{''}[k]=\sqrt{\frac{2}{P}}\sum_{p=0}^{P-1}\mu[p]Y_p[k]\cos\left(\frac{\pi n(2p+1)}{2P}\right),\ n\in[0,P-1]$$

步驟 10. 正規化

對數頻譜係數正規化：

$$M_q^{nor}[k]=\frac{M_q^{''}[k]-s_{\min}}{s_{\max}-s_{\min}}$$

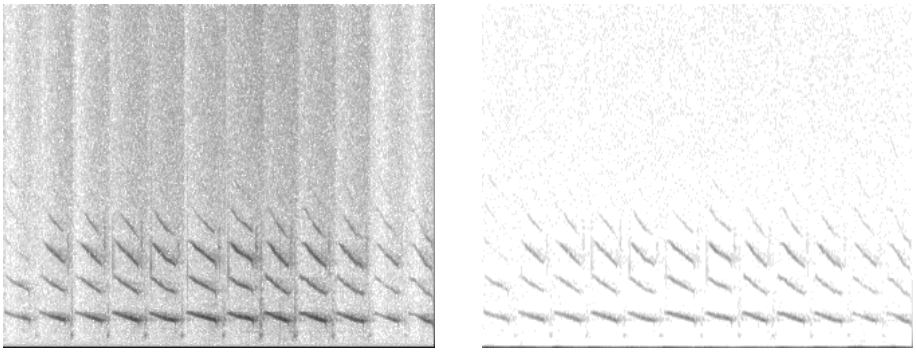
其中 $s_{\max}=\max_{1\le q\le P,0\le k<N}(M_q^{''}[k])$，$s_{\min}=\min_{1\le q\le P,0\le k<N}(M_q^{''}[k])$。

步驟 11. 強化原始訊號

以正規化之對數頻譜係數為權重與原始頻譜相乘可得到濾除雜訊後之強化訊號

$$X_q^{enh}[k]=\alpha X_q[k]M_q^{nor}[k],\ 1\le q\le P,0\le k<N$$

其中α為強化係數。



(a) (b)

(c) (d)

圖二 (a)原始有雜訊之鳥類鳴聲之頻譜 (b)做倒頻譜運算後之頻譜 (c)做調變頻譜處理後之頻譜 (d)強化後之頻譜。

2.2 特徵擷取

對於切割出來之每一鳥類聲音之音節，我們將自倒頻譜域(cepstral domain)中擷取二維梅爾倒頻譜係數(Two-dimensional Mel-scale Frequency Cepstral Coefficients, TDMFCC）及動態二維梅爾倒頻譜係數(Dynamic TDMFCC, DTDMFCC)，以及自頻譜域(spectral domain)擷取 MPEG-7 之聲音頻譜封包(Normalized Audio Spectrum Envelope, NASE)為此音節之特徵向量。

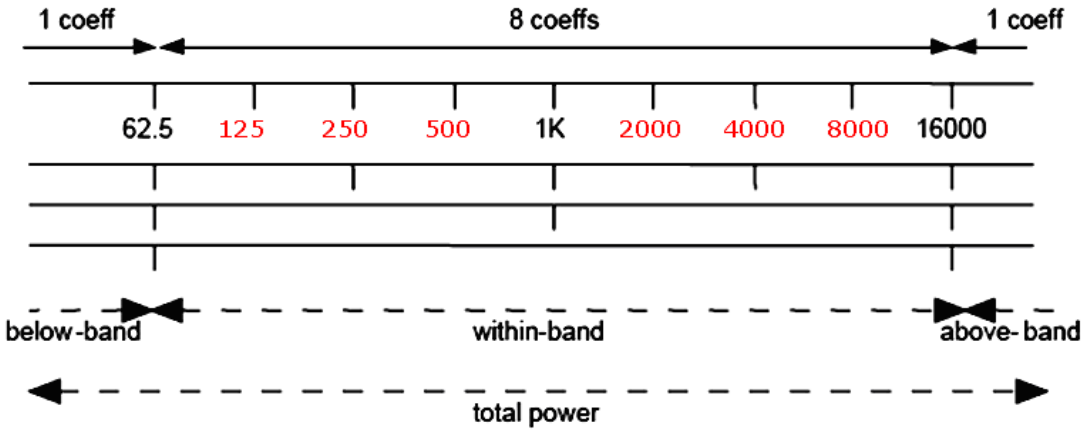
2.2.1 MPEG-7 之聲音頻譜封包之調變係數

在MPEG-7標準中[18]，是以對數之頻率間格來描述音訊訊號的頻譜圖。由於人類對於頻率的敏感度是成對數對應關係，所以我們使用對數頻率來取頻率間距，如此可以兼顧簡潔性與描述性。聲音頻譜封包(audio spectrum envelope, ASE)在MPEG-7標準裡普遍用於表示原始聲音訊號中每一頻帶之功率頻譜，主要是描述介於loEdge (預設62.5Hz)與hiEdge (預設為16000Hz)間的頻譜圖資訊，將介於loEdge與hiEdge間的頻率再分成多個頻

帶。而每一頻帶的頻寬解析度是以八度音(Octave)解析度為基準，以1000Hz為中心上下區分。其中每一頻帶之邊緣頻率(Edge)的公式如下

$$Edge = 2^{rm} \times 1000$$

其中r是八度音的解析度，m是整數。此外又加上兩個額外的頻帶，一個為0Hz到loEdge的頻帶能量總合，一個為hiEdge到取樣頻率一半的頻帶能量合，圖三為一個八度音解析度之Edge分隔圖，表一是每個臨界頻帶濾波器的頻帶範圍。



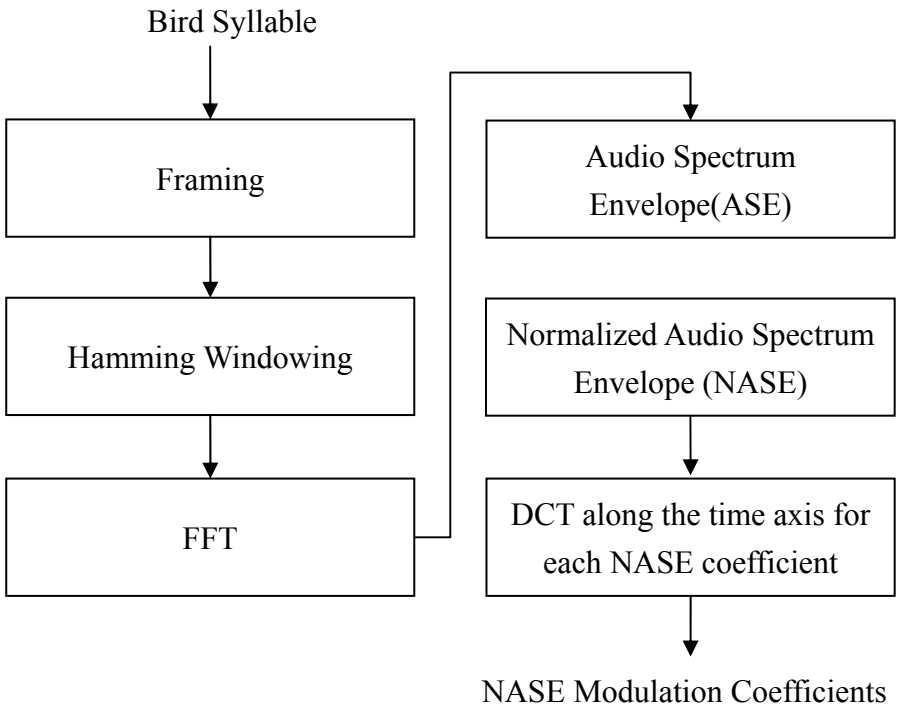
圖三 八度音頻帶濾波器

表一 八度音頻帶濾波器中每個臨界頻帶濾波器的頻帶範圍(取樣頻率為 44100Hz)

index	loEdge	hiEdge
ϕ_0	0 Hz	62.5 Hz
ϕ_1	62.5 Hz	125 Hz
ϕ_2	125 Hz	250 Hz
ϕ_3	250 Hz	500 Hz
ϕ_4	500 Hz	1000 Hz
ϕ_5	1000 Hz	2000 Hz
ϕ_6	2000 Hz	4000 Hz
ϕ_7	4000 Hz	8000 Hz
ϕ_8	8000 Hz	16000 Hz
ϕ_9	16000 Hz	22050 Hz

NASE在MPEG-7標準中是針對每一個音框之ASE係數轉換至分貝之刻度後做正規化之動作，然而對一個鳥聲音節而言，可能包含了許多音框，所以我們希望能夠把NASE係數隨著時間的變化性和相鄰音框之間的相關性考慮進來，所以我們將每一個音框中相

同頻帶之NASE係數沿著時間軸之方向做離散餘弦轉換(DCT)，便取得到此一鳥類音節之NASE調變係數，但我們只取低頻部份作為特徵值，圖四為求取NASE調變係數特徵之流程圖。



圖四 計算正規化之聲音頻譜封包調變係數之流程圖

此一特徵之產生步驟如下：

步驟 1: 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 2: 乘上漢明視窗(Hamming Windowing)

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 3: 快速傅立葉轉換(FFT)

將音訊訊號從時域轉換成頻率域

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

N：音框大小, $\tilde{s}[n]$：離散訊號, l：音框索引值

步驟 4: 計算 ASE

首先根據 Parseval's 定理，計算功率頻譜之平均值

$$\overline{P_l} = \frac{1}{NE_w}\sum_{k=0}^{N-1}\left|X_l[k]\right|^2$$

其中

$$E_w = \sum_{n=0}^{N-1}\left|w[n]\right|^2$$

接著計算功率頻譜之係數值並做正規化

$$P_l(k) = \frac{1}{NE_w}\left|X_l[k]\right|^2 \quad \text{for } k = 0 \text{ and } k = \frac{N}{2}$$

$$P_l(k) = 2\frac{1}{NE_w}\left|X_l[k]\right|^2 \quad \text{for } 0 < k < \frac{N}{2}$$

使用八度音頻帶濾波器(如表 2.2)將聲音訊號分成一個個頻帶，並算出每個頻帶的能量:

$$ASE(l,f) = \sum_{k=0}^{N-1}\phi_f(k)P_l(k) \text{ ， } 0 \le f < F$$

其中 f 為八度音頻帶範圍之索引值，F 為八度音頻帶之數目，而 $\phi_f(k)$ 為表示第 j 個濾波器：

$$\phi_f(k) = \begin{cases} 1 & loEdge_j \le k < hiEdge_j \\ 0 & otherwise \end{cases}$$

最後將其轉換至分貝單位

$$ASE_{dB}(l,f) = 10\log_{10}(ASE(l,f))$$

步驟 5: 計算 NASE

對每一個 ASE_{dB} 做正規化之動作，首先，先計算每一個音框之 RMS 值，在這裡以 R_l 來表示為第 l 個音框之 RMS 值：

$$R_l = \sqrt{\sum_{f=0}^{F-1}\left(ASE_{dB}(l,f)\right)^2} \text{ ， } 0 \le f < F$$

接著計算 NASE 值：

$$NASE(l,f) = \frac{ASE_{dB}(l,f)}{R_l}, \ 0 \le l < L$$

其中 L 為表示音框之總數目。

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

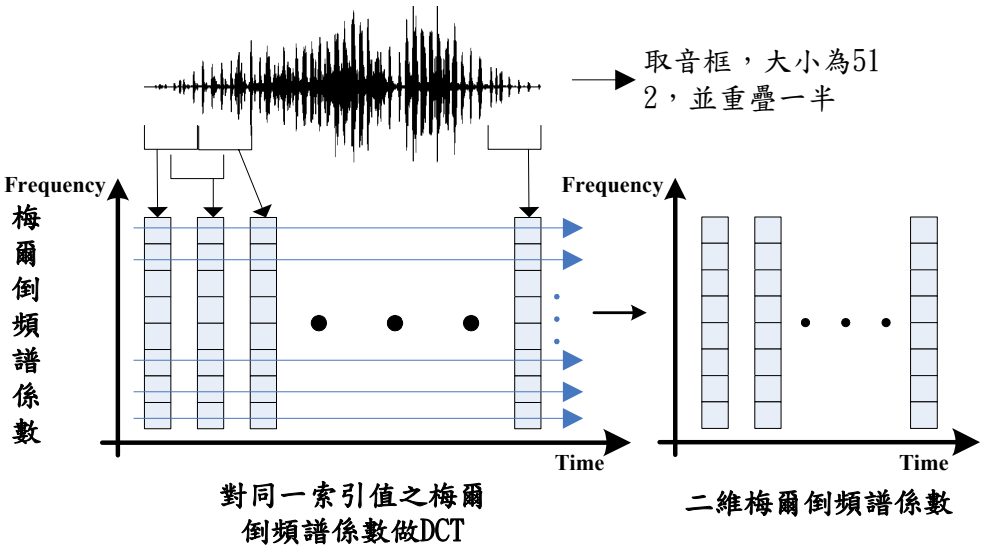
令 $M(l,f)$ 為對所有 $NASE(l,f)$ 沿著時間軸之方向做離散餘弦轉換所得到之調變頻譜係數，式子如下：

$$M(l,f)=\frac{1}{L}\sum_{i=0}^{L-1}NASE(i,f)\cos(2\pi il/L),\ 0\le l<L,\ 0\le f<F$$

另外，在選取 $M(l,f)$ 參數當作特徵時，我們只取時間軸的前五個索引值，也就是 NASE 調變頻譜係數區塊大小為 8×5；此外，令 R_l' 為對 R_l 作離散餘弦轉換，亦只取前五個係數作為特徵，所以取 9×5 共 45 個特徵作為 NASE 之調變頻譜特徵。

2.2.2 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[19-21]，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和動態特性，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，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，其示意圖如圖五所示。



圖五 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：

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

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)：

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

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。

2.2.3 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[22]，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 $C_i'(m)$:

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

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 主軸分析演算法(Principal Component Analysis, PCA)

主軸分析演算法 [23]之主要目的是降低特徵向量之維度，但是降低特徵向量之維度會損失部分資訊，所以我們要如何在降低維度後還能保持最大之資訊量，因而不影響辨識之結果，甚至是刪除那些降低辨識率的特徵，而使得辨識率上升，這個問題是 PCA 所要解決的主要課題。

PCA 是先計算所有訓練資料之特徵向量的平均變異數矩陣之 eigenvalue 及 eigenvector，並以 eigenvector 當作基底來做線性轉換，而 eigenvalue 的大小可以決定其對應之 eigenvector 轉換後之特徵所保留之資訊量大小，eigenvalue 越大表示資料作線性轉換後，特徵的變異數值會越大，而變異數的大小又表示分佈的寬廣，資料分佈越廣表示所保留之資訊量越大，也就是說，以 eigenvalue 值較大之 eigenvector 做為線性轉換之基底，轉換後的特徵分佈範圍會比以 eigenvector 較小的 eigenvector 轉換後的分佈範圍來得大。PCA 之進行步驟如下：

步驟 1：計算平均向量

$$\mathbf{m} = E[\boldsymbol{X}]$$

其中 $\boldsymbol{X}$是所有訓練資料之集合，$X = \{\mathbf{x}_i \mid i = 0 \ldots N\}$，$\mathbf{m}$ 是所有訓練資料的平均向量，N是訓練資料的數量。

步驟 2：令平均向量為 **0**

$$\mathbf{x}_i^{'} = \mathbf{x}_i - \mathbf{m}$$

步驟 3：求取平均變異數矩陣，**C**

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i^{'} (\mathbf{x}_i^{'})^{\mathrm{T}}$$

步驟 4：求取變異數矩陣 **C** 的 eigenvalue 及 eigenvector 並將其依 eigenvalue 值由大至小

重新排序

步驟 5：設定臨界值 *T*(表示所要保留的資訊量程度)，以計算轉換後維度 *d*

$$\sum_{i=1}^{d}\lambda_i \geq T \times \sum_{i=1}^{D}\lambda_i$$

其中λ_i表示第 *i* 大之 eigenvalue，*D* 為轉換前之維度

步驟 6：以所保留之 *d* 個 eigenvector 對所有資料作線性轉換

$$\mathbf{y} = \mathbf{E}^{\mathrm{T}}\mathbf{x}_i'$$

其中 **E** 為此 *d* 個較大 eigenvector 構成之轉換矩陣。

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，我們將使用高斯混合模型(Gaussian mixture model, GMM)來描述，也就是說屬於同一種鳥類聲音之不同音節可以分成幾個小群(高斯分佈)，而屬於同一小群之不同音節其特徵向量會較相似。

傳統上，使用高斯混合模型於分類辨識，對不同類別之資料需分別建立其高斯混合模型，一般傳統上高斯混合模型之參數預測是使用 EM (Expectationn Maximization)演算法[24]。此演算法主要是用來預測高斯混合模型中多變數機率分佈函數之參數值，其目的是要找到最佳之參數Θ使得 $p(\mathbf{X}\,|\,\Theta)$ 最大，其中 $\mathbf{X} = \{\mathbf{x}(t), t = 1, 2, \ldots, N\}$ 為訓練資料之特徵向量集合，*N* 為訓練資料之數目；$\Theta \equiv \{p(\theta_r), \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \mid r = 1, 2, ..., M\}$，$p(\theta_r)$ 為在高斯混合模型中第 *r* 個高斯分佈之事前機率(Prior Probability)，$\boldsymbol{\mu}_r$ 為平均值向量，$\boldsymbol{\Sigma}_r$ 為共變異數矩陣(Covariance matrix)，*M* 為高斯混合模型中高斯分佈之群數。EM 演算法之詳細步驟如下：

步驟 1: 執行 K-Means 演算法

首先，依據高斯混合模型中所指定之高斯分佈群數執行 K-Means 演算法分群，以每群之平均值向量作為每個高斯分佈之平均值向量之初始值，且將共變異數矩陣之初始值設為單位矩陣。

步驟 2: Expectation-Step

對所有資料計算其屬於高斯混合模型中每一高斯分佈之機率比值作為預測值，其公式如下：

$$p(\theta_r \mid \mathbf{x}(t)) = \frac{p(\theta_r)p(\mathbf{x}(t) \mid \theta_r)}{\sum_{r=1}^{M} p(\theta_r)p(\mathbf{x}(t) \mid \theta_r)}$$

其中

$$p(\mathbf{x}(t) \mid \theta_r) = \frac{1}{\sqrt{(2\pi)^d \boldsymbol{\Sigma}_r}} \exp\left(-\frac{(\mathbf{x}(t) - \boldsymbol{\mu}_r)^T \boldsymbol{\Sigma}_r^{-1} (\mathbf{x}(t) - \boldsymbol{\mu}_r)}{2} \right)$$

d 為特徵向量之維度。

步驟 3: Maximization-Step

利用**步驟 2** 所計算之預測值，更新預測之參數值：

權重值：

$$p(\overline{\theta}_r) = \frac{1}{N} \sum_{t=1}^{N} p(\theta_r \mid \mathbf{x}(t))$$

平均值向量：

$$\overline{\boldsymbol{\mu}}_r = \frac{\sum_{t=1}^{N} p(\theta_r \mid \mathbf{x}(t))\mathbf{x}(t)}{\sum_{t=1}^{N} p(\theta_r \mid \mathbf{x}(t))}$$

共變異數矩陣：

$$\overline{\boldsymbol{\Sigma}}_r = \frac{\sum_{t=1}^{N} p(\theta_r \mid \mathbf{x}(t))(\mathbf{x}(t) - \overline{\boldsymbol{\mu}}_r)(\mathbf{x}(t) - \overline{\boldsymbol{\mu}}_r)^T}{\sum_{t=1}^{N} p(\theta_r \mid \mathbf{x}(t))}$$

步驟 4: 重覆執行**步驟 2～3**，直到收斂為止。

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

線性區別分析演算法[23]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，線性區別分析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 $d \leq n$，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:

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

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:

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

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:

$$S_B = \sum_{j=1}^{C} n_j(\mu_j - \boldsymbol{\mu})(\mu_j - \boldsymbol{\mu})^T$$

$\boldsymbol{\mu}$ 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 $n \times d$:

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

此一轉換矩陣，可經由求出 $S_W^{-1}S_B$ 的特徵向量(eigenvectors)來得到，而 A_{opt} 之 d 個行向量為前 d 個最大特徵值(eigenvalue)值所對應之特徵向量。在此我們是取(鳥鳴聲種類數目-1)作為 d 值。

在我們決定出最佳的轉換矩陣 A_{opt} 後，我們以 A_{opt} 將每一正規化(normalized)後之 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:

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

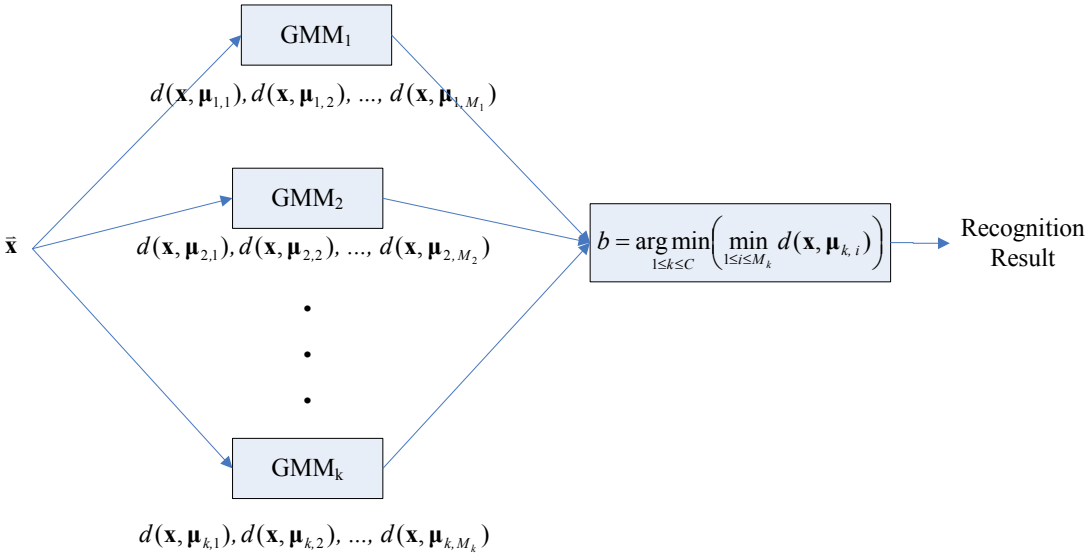
2.6 辨識階段

在輸入一個測試音節之後，我們先計算其 MPEG-7 之聲音頻譜封包之調變係數(MNASE)、二維梅爾倒頻譜係數(TDMFCC)和動態二維梅爾倒頻譜係數(DTDMFCC)等特徵向量，然後對各個特徵值作正規化，接著以主軸分析演算法來降低降低特徵向量維度，最後利用線性區別分析演算法再進一步降低特徵向量維度且提高不同類別間特徵向量之距離，辨識時以歐基里德距離(Euclidean distance)來計算測試特徵向量和每一鳥類鳴叫聲所建立之高斯混合模型中高斯分佈的向量平均值之間的距離，取最小距離作為代表

距離該種鳥類之距離，最後我們取測試特徵向量與不同鳥類之高斯混合模型所計算出之距離最小者，即代表辨識結果為該鳥種，其辨識架構圖如圖六所示，令 C 為鳥種數目，b 代表辨識出來之鳥類種類，式子如下：

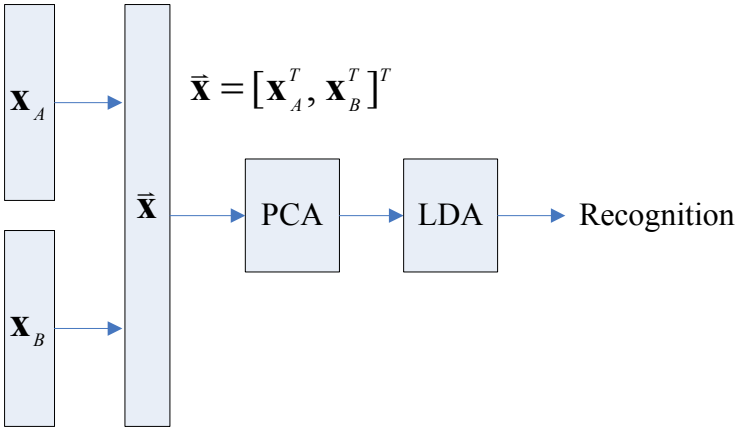
$$b = \arg\min_{1\le k\le C}\left(\min_{1\le i\le M_k} d(\mathbf{x}, \boldsymbol{\mu}_{k,i})\right)$$

其中 $\mathbf{x}$ 為測試特徵向量；M_k 為第 k 個高斯混合模型中高斯分佈群數；$\boldsymbol{\mu}_{k,i}$ 為第 k 個高斯混合模型中第 i 個高斯分佈的向量平均值。

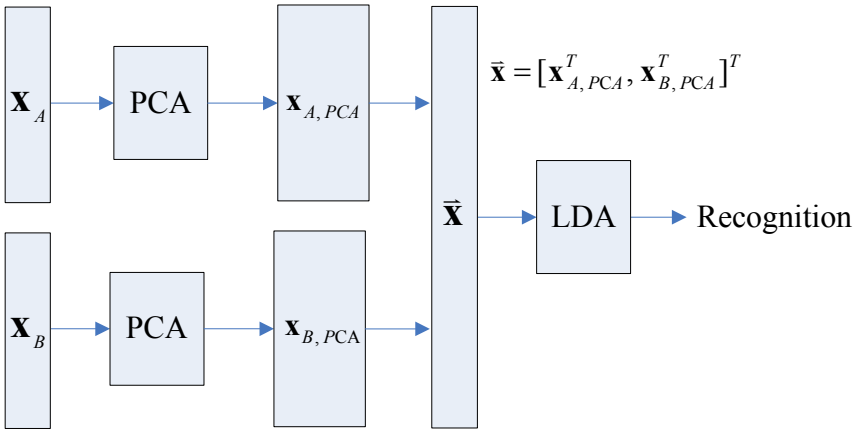


圖六 應用高斯混合模型於鳥類聲音辨識之系統架構圖

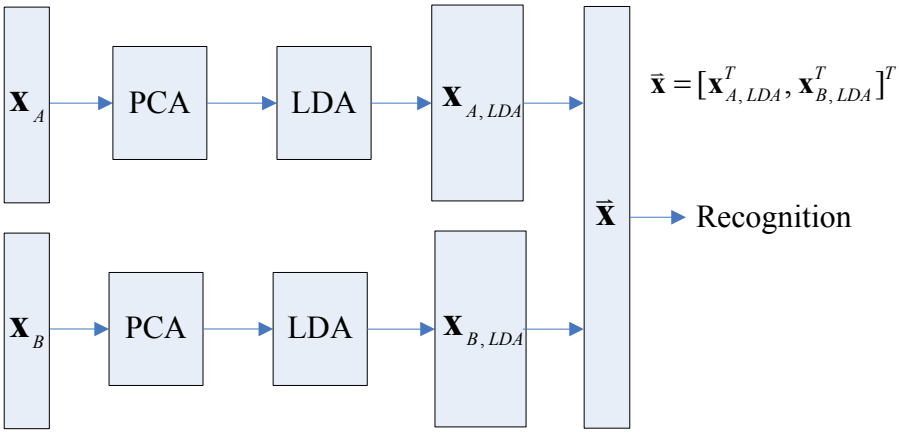
此外，我們藉由結合不同性質之特徵向量作辨識以提高辨識率，結合的方式有三種，第一種為先將兩種不同特徵向量串成一組新的特徵向量同上述之辨識流程來判斷測試音節為何種鳥類(如圖七)。第二種方法則是先將兩種特徵各自經由 PCA 之運算降低特徵維度後再串成一組新的特徵向量同上述之辨識流程來判斷測試音節為何種鳥類(如圖八)。最後一種方法則是先分別依上述之辨識流程對個別執行線性區別分析後，才串接成一組新的特徵向量，再計算特徵向量之間的距離，取距離最小者來代表該鳥類(如圖九)。



圖七 先結合特徵向量再進行辨識之流程



圖八 先將個別特徵向量維度以 PCA 降低後再結合進行辨識



圖九 先將個別特徵向量維度以 LDA 降低後再結合進行辨識

3. 實驗結果與討論

在實驗中所使用之鳥類鳴聲資料，總共有 28 種台灣鳥類，其訓練資料之鳥類錄音與測試資料之鳥類錄音皆為在不同環境下使用不同錄音設備之錄音。首先，將所有鳥類

錄音重新取樣調整為 44100 Hz，音訊範圍大小為 16 bits，對資料庫中所有鳥類擷取特徵進行辨識前，我們以人工的方式對每個聲音檔案切出鳥類音節，在 28 種鳥類共有 3143 個訓練音節和 646 個測試音節，在表二中顯示每種鳥類聲音所切割之音節數目。接著對每一音節取其二維梅爾倒頻譜係數(TDMFCC)、動態二維梅爾倒頻譜係數(DTDMFCC)和正規化之聲音頻譜封包調變係數(MNASE)作為特徵(每一特徵向量之維度請參考表三)，首先利用主軸分析演算法降低特徵維度，在表格四中表示特徵維度在 PCA 門檻值(0.9~1)間之變化。

我們比較二維梅爾倒頻譜係數、動態二維梅爾倒頻譜係數、NASE 調變係數、結合二維梅爾倒頻譜係數和動態二維梅爾倒頻譜係數以及最後結合此三種特徵向量一起辨識之正確率，表格五顯示每一個別特徵向量對於 28 種鳥類之辨識率。其實驗結果顯示二維梅爾倒頻譜係數之特徵最佳之辨識正確率為取 PCA 門檻值 0.93～0.94 其辨識正確率為 80.5%；動態二維梅爾倒頻譜係數之特徵最佳之辨識正確率為取 PCA 門檻值 0.93 其辨識正確率為 75.85%；NASE 調變係數之特徵最佳辨識正確率為取 PCA 門檻值 0.97 其辨識正確率為 63.47%。

在結合多種特徵向量一起辨識之部份，我們比較四種不同之結合方法：結合方法一(Fusion-1) 先將特徵向量串接結合，再加上 PCA 及 LDA 分析後計算距離(如圖七)；結合方法二(Fusion-2)先將個別特徵向量以 PCA 降低維度後再串接結合，再加上 LDA 分析後計算距離(如圖八)；結合方法三(Fusion-3)先將個別特徵向量以 PCA 及 LDA 分析後計算個別特徵向量距離，再將個別距離加總以得到整體特徵向量之距離(參考圖九)；結合方法四(Fusion-4)先將個別特徵向量以 PCA 及 LDA 分析後計算個別特徵向量距離，再將個別距離相乘以得到整體特徵向量之距離(參考圖九)。

表六比較各種結合方法將二維梅爾倒頻譜係數和動態二維梅爾倒頻譜係數之特徵向量結合之辨識正確率，使用結合方法一(Fusion-1)之最佳辨識正確率中為取 PCA 門檻值 0.95～0.96 其辨識正確率為 82.66%，而使用結合方法二(Fusion-2)之最佳辨識正確率中為取 PCA 門檻值 0.90 其辨識正確率為 82.51%，使用結合方法三(Fusion-3)之最佳辨識正確率中為取 PCA 門檻值 0.99 其辨識正確率為 81.42%；最後使用結合方法四(Fusion-4)之最佳辨識正確率中為取 PCA 門檻值 0.95 其辨識正確率為 82.51%。

表七比較各種結合方法將三種特徵向量(TDMFCC, DTDMFCC, MNASE)結合之辨識正確率，使用結合方法一(Fusion-1)之最佳辨識正確率中為取 PCA 門檻值 0.90 其辨識正確率為 78.79%，而使用結合方法二(Fusion-2)之最佳辨識正確率中為取 PCA 門檻值

0.90 其辨識正確率為 76.01%，使用結合方法三(Fusion-3)之最佳辨識正確率中為取 PCA 門檻值 0.93, 0.95 或 0.96 其辨識正確率為 81.11%;最後使用結合方法四(Fusion-4)之最佳辨識正確率中為取 PCA 門檻值 0.95 其辨識正確率為 83.90%，綜合以上實驗數據來看以結合三種特徵向量並使用結合方法四(Fusion-4)對於鳥類鳴聲之辨識有較佳之結果。

表二 28 種鳥類訓練音節與測試音節數目

Bird Name	Training Syllable	Test Syllable
大冠鷲	10	4
小卷尾	229	37
小啄木	17	25
小翼鶇	296	29
小彎嘴畫眉	120	22
火冠戴菊鳥	194	57
白耳畫眉	98	14
白喉笑鶇	100	37
白腹秧雞	172	15
灰鶯	70	8
竹鳥	31	31
岩鷚	122	53
青背山雀	140	14
冠羽畫眉	49	12
紅頭山雀	61	24
栗背林鴝	230	18
烏頭翁	131	30
深山竹雞	123	27
深山鶯	51	8
筒鳥	284	45
黃山雀	222	27
黃腹琉璃	76	12
煤山雀	149	34
鳳頭蒼鷹	32	16
頭烏線	32	18
鵂鶹	61	14
藍腹鷴	23	10
藪鳥	20	5

表三 擷取之特徵向量維度

Feature	Dimension
TDMFCC	75
DTDMFCC	75
MNASE	45

表四 經由 PCA 後各特徵向量在不同門檻值下所保留之特徵維度

Threshold	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1
TDMFCC	36	38	40	43	45	48	51	54	58	64	74
DTDMFCC	31	33	35	37	38	42	44	48	52	58	74
MNASE	21	22	23	24	25	27	45	49	53	59	74
TDMFCC+ DTDMFCC	50	53	56	60	64	69	74	82	92	107	148
TDMFCC+ DTDMFCC +MNASE	60	64	68	73	78	84	92	102	115	135	193

表五 個別特徵向量之辨識率(%)

Feature	PCA Threshold										
	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1
TDMFCC	77.86	78.48	78.02	80.50	80.50	80.19	78.79	79.57	80.34	80.34	75.23
DTDMFCC	73.84	74.61	74.46	75.85	72.76	74.77	73.37	72.60	71.21	71.83	74.15
MNASE	57.28	58.05	58.05	60.53	61.46	62.38	62.38	63.47	58.67	57.12	56.04

表六 比較各種結合方法將 TDMFCC 和 DTDMFCC 結合之辨識正確率(%)

Fusion method	PCA Threshold										
	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1
Fusion-1	81.27	82.35	80.50	80.96	80.19	82.66	82.66	78.95	78.48	77.55	75.39
Fusion-2	82.51	78.02	78.64	82.20	82.04	80.80	78.64	80.96	80.50	76.78	75.08
Fusion-3	80.19	79.88	79.41	80.96	80.50	80.19	80.65	80.34	80.65	81.42	78.95
Fusion-4	80.34	78.79	78.95	81.42	80.65	82.51	80.96	79.41	78.64	76.63	76.16

表七將三種特徵向量(TDMFCC, DTDMFCC, MNASE)結合之辨識正確率(%)

Fusion method	PCA Threshold										
	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1
Fusion-1	78.79	76.63	77.24	75.85	75.54	76.63	75.70	71.36	75.39	74.15	70.28
Fusion-2	76.01	75.08	75.08	75.08	74.61	73.84	74.15	73.22	73.07	74.15	72.45
Fusion-3	79.57	80.96	79.41	81.11	80.96	81.11	81.11	80.96	79.41	78.17	73.68
Fusion-4	80.50	80.03	79.57	82.04	82.20	83.90	82.04	80.81	77.40	75.54	73.37

二. 參考文獻

[1] 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.

[2] 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.

[3] A. L. McIlraith and H. C. Card, “A comparison of backpropagation and dtatistical classifiers for bird identification”, in *Proceedings of IEEE International Conference on Neural Networks* , Vol. 1, pp. 100-104, June 1997.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] A. Harma, “Automatic identification of bird species based on sinusoidal modeling of syllables”, in *Proceedings of IEEE International Conference on Acoustics, Speech, and*

Signal Processing, Vol. 5, pp. 545-548, 2003.

[9] A. Harma and P. Somervuo, “Classification of the harmonic structure in bird vocalization”, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 701-704, 2004.

[10] P. Somervuo and A. Harma, “Bird song recognition based on syllable pair histograms”, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 825-828, 2004.

[11] S. Fagerlund and A. Harma, “Parametrization of inharmonic bird sounds for automatic recognition”, in Proceedings of the *13th European Signal Processing Conference* (*EUSIPCO 2005*), Antalya, Turkey, Sep. 2005.

[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, pp. 2252-2263, Nov. 2006.

[13] M. B. Trawicki and M. T. Johnson, “Automatic song-type classification and speaker identification of Norwegian Ortolan Bunting (Emberiza Hortulana) vocalizations”, in *Proc. of IEEE Workshop on Machine Learning for Signal Processing*, pp. 277-282, Sep. 2005.

[14] A. Selin, J. Turunen, and J. T. Tanttu, “Wavelets in Recognition of Bird Sounds”, *EURASIP Journal on Advances in Signal Processing,* Vol. 2007, Article ID 51806, 9 pages.

[15] S. A. Selouani, M. Kardouchi, E. Hervet, and D. Roy, “Automatic Birdsong Recognition Based on Autoregressive Time-Delay Neural Networks”, in *Proc. of 2005 ICSC Congress on Computational Intelligence Methods and Applications*, Dec. 2005.

[16] C. F. Juang and T. M. Chen, ”Birdsong recognition using prediction-based recurrent neural fuzzy networks,” Neurocomputing, vol.71, no.1-3, pp.121-130, 12 2007. (SCI、

EI).

[17] Edgar E. Vallejo, Martin L. Cody, and Charles E. Taylor, “Unsupervised Acoustic Classification of Bird Species Using Hierarchical Self-organizing Maps”, *Springer-Verlag Berlin Heidelberg* 2007, ACAL 2007, LNAI 4828, pp. 212 – 221, 2007.

[18] H. G. Kim, N. Moreau, and T. Sikora, “Audio classification based on MPEG-7 spectral basis representation”, *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 716-725, May 2004

[19] Y. Ariki, S. Mizuta, M. Nagata, and T. Sakai “Spoken-word recognition using dynamic features snalysed by two-dimensional cepstrum”, *IEE Proceedings on Communications, Speech and Vision*, Vol. 136, No. 2, pp. 133-140, April, 1989.

[20] H. F. Pai and H. C. Wang, “A study of the two-dimensional cepstrum approach for speech recognition”, *Computer Speech and Language*, Vol. 6, pp. 361-375, 1992.

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

[22] S. Furui, “Speaker-independent isolated word recognition using dynamic features of speech spectrum”, *IEEE Trans. Speech and Audio Processing*, Vol. ASSP-34, No. 1, pp. 52-59, February, 1986.

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

[24] D. A. Reynolds and R. C. Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” *IEEE Trans. Speech Audio Processing*, vol. 3, no. 1, pp. 72–83, Jan. 1995.

三. 計畫成果自評

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、環境監控和生物多樣性評估等方面。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，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。

此外，建立包括圖像、字元、影像、語音等數位內容之數位博物館為政府近幾年來發展之重點產業，希望將各種多媒體資料加以數位化並整合運用。其中本土生物之聲音資料庫的建立在數位博物館中是重要而基本的工作。在生物叫聲的辨識中，鳥類鳴叫聲音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，因此本計劃對鳥類鳴叫聲音之自動辨識做一深入之研究以輔助調查鳥類族群之生態、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。本計畫已完成可自動辨識鳥類鳴聲之辨識系統，可以應用於各種不同之錄音環境及錄音器材，最佳辨識率可達 83.90%。

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. C. Han, and C. C. Chuang, “Automatic CLassification of Bird Species by Their Sounds Using Two Dimensional Cepstral Coefficients”, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 16, No. 8, Nov. 2008, pp. 1541-1550. (SCI, EI)

[2] 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)

[3] 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.

研討會論文 (Conference Papers)：

[1] 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.

[2] 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.

[3] 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.

[4] 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.