

# 行政院國家科學委員會專題研究計畫 期末報告

## 應用影像形狀特徵於鳥類聲紋辨識之研究

計畫類別：個別型  
計畫編號：NSC 100-2221-E-216-035-  
執行期間：100年08月01日至101年10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、陳建宏  
計畫參與人員：碩士班研究生-兼任助理人員：黃百祥  
碩士班研究生-兼任助理人員：陳昕宏  
博士班研究生-兼任助理人員：林懷三

報告附件：出席國際會議研究心得報告及發表論文

公開資訊：本計畫可公開查詢

中華民國 102 年 01 月 29 日

中文摘要：傳統上用來辨識鳥類鳴聲之聲音特徵主要是源自語音/語者辨識或音訊/音樂之分類，本計劃中我們以不同之角度切入來擷取辨識特徵，主要是以影像處理之方法來辨識鳥類鳴聲。首先我們將鳥類鳴聲之聲譜圖(spectrogram)視為一張影像，而鳥類鳴聲中較顯著之頻率變化片段在聲譜圖影像中則顯示出明顯之邊緣特徵，因此我們以 MPEG-7 之徑角轉換(angular radial transform, ART)來擷取聲譜圖影像中之外形變化。然而聲譜圖影像中之外形變化主要是沿著垂直之頻率軸及水平之時間軸之方向，而徑角轉換是描述二維影像形狀中角度及徑度之變化趨勢，為了有效利用徑角轉換之特性，我們提出扇形展開演算法將聲譜圖影像轉換成扇形影像，因此可以利用徑角轉換來擷取聲譜圖之頻譜及時間變化情況，最後再以高斯混合模型來辨識鳥類鳴聲。實驗結果顯示此一方法比傳統之聲音特徵，包括線性預測倒頻譜係數、梅爾倒頻譜係數、及二維梅爾倒頻譜係數等，可得到更高的正確率。

中文關鍵詞：鳥類鳴聲，聲譜圖，徑角轉換，高斯混合模型

英文摘要：

英文關鍵詞：

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

應用影像形狀特徵於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 100-2221-E-216-035-  
執行期間：2011年08月01日至2012年10月31日

計畫主持人：李建興  
共同主持人：連振昌、陳建宏  
計畫參與人員：林懷三、黃百祥、陳昕宏

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 102 年 01 月 30 日

## 摘要

傳統上用來辨識鳥類鳴聲之聲音特徵主要是源自語音/語者辨識或音訊/音樂之分類，本計劃中我們以不同之角度切入來擷取辨識特徵，主要是以影像處理之方法來辨識鳥類鳴聲。首先我們將鳥類鳴聲之聲譜圖(spectrogram)視為一張影像，而鳥類鳴聲中較顯著之頻率變化片段在聲譜圖影像中則顯示出明顯之邊緣特徵，因此我們以 MPEG-7 之徑角轉換(angular radial transform, ART)來擷取聲譜圖影像中之外形變化。然而聲譜圖影像中之外形變化主要是沿著垂直之頻率軸及水平之時間軸之方向，而徑角轉換是描述二維影像形狀中角度及徑度之變化趨勢，為了有效利用徑角轉換之特性，我們提出扇形展開演算法將聲譜圖影像轉換成扇形影像，因此可以利用徑角轉換來擷取聲譜圖之頻譜及時間變化情況，最後再以高斯混合模型來辨識鳥類鳴聲。實驗結果顯示此一方法比傳統之聲音特徵，包括線性預測倒頻譜係數、梅爾倒頻譜係數、及二維梅爾倒頻譜係數等，可得到更高的正確率。

## 一. 報告內容

### 1. 前言

生態保育及生物多樣性資訊調查研究越來越受到全世界各國之重視，而評估一地區生態保育環境之好壞，一項重要之指標是調查此一地區之動植物分佈群相及其變化狀況，因此在 2001 年成立之全球生物多樣性資訊機構，目的是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。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，通常我們在野外比較容易聽到鳥類鳴聲而不易見其形體，此外鳥類鳴聲早已進化成與特定之種類相關(species-specific)，也就是說不同之鳥類種類之鳴聲會有所不同，因此利用鳥類的鳴聲來辨識其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鳥類之種類及其分佈定位。目前全世界的鳥類約有 9,200 多種，根據 2012 年台灣鳥類名錄，台灣野生鳥類有 601 種，包括 24 種特有種和 59 種特有亞種[1]，由於鳥類種類相當豐富，而目前鳥類鳴聲的辨識方法，多採用人工至野外錄音，再回實驗室做人工的

識別。若能利用鳥類鳴聲來自動辨識其種類，可以節省相當多的人力與時間。

一般而言，鳥類鳴聲結構相當複雜，通常可表示成階層式之結構[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音素稱為音節(syllable)，而一連串的音節又組成了樂旨(motif)或是樂句(phrase)，一些重覆出現的樂旨的組合，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樂旨則組成所謂的樂曲(bout)。

傳統上，透過專家以視覺觀察聲譜圖/聲紋圖(spectrogram/sonogram)之做法來分析鳥類鳴聲是常見之方式，然而此種方式相當耗時費力，而且相當主觀，因此並不適用於大量且長期之調查分析。最近圖型識別之方法已廣為應用於鳥類鳴聲之自動分類辨識，首先，自鳥類鳴聲中擷取聲音特徵，然後以分類辨識演算法來區別鳥種。一般而言，用於鳥類鳴聲辨識之聲音特徵可以簡單分成四類：頻譜特徵(spectral feature)、時間特徵(temporal feature)、倒頻譜特徵(cepstral feature)、及調變頻譜特徵(modulation spectral feature)。頻譜特徵通常描述聲音訊號之音色特性，可以用來分辨不同來源之聲音訊號，常見之頻譜特徵有頻譜密度(spectral density)[3-5]、線性預測編碼(Linear Predictive Coding, LPC)係數[6-8]、LPC反射係數[6]、個別頻帶(subband)之頻譜密度[6, 9-11]、頻率/振幅軌跡[12-15]、小波轉換係數[16]、頻譜質心(spectral centroid)[17, 18]、頻譜寬度(spectral bandwidth)[5, 17-19]、頻譜滑動頻率(spectral roll-off frequency)[17, 18]、頻譜變遷度(spectral flux)[17, 18]、頻譜平滑度(spectral flatness)[17, 18]、頻譜範圍(frequency range)[17, 18]、短時距能量(short time energy)[17-19]、最大頻率[17, 20]、最小頻率[17, 20]、及平均頻率[5, 19]等。

時間特徵主要是描述聲音訊號中與時間相關之特性，包括音素數目[7]、音素長度之平均值及標準差[7]、靜音長度之平均值及標準差[7]、越零率(zero-crossing rate)[17, 18]、音節長度[17, 18, 20]等。

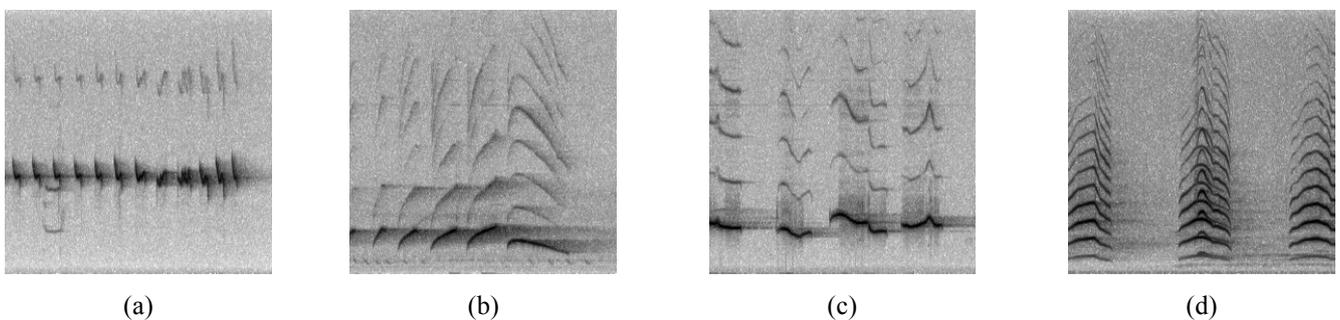
倒頻譜特徵，包括LPC倒頻譜係數(LPCC)及梅爾倒頻譜係數(MFCC)，已廣泛應用於語音辨識、音訊分類、及鳥類鳴聲辨識上，事實上，將梅爾倒頻譜係數應用於語音辨識[24]、音訊分類[25]、及鳥類鳴聲辨識[5, 6, 17-19, 21-23]上都可以得到不錯的效果。

調變頻譜特徵曾經應用於語音/語者辨識[26-28]及音訊/音樂辨識上[29-32]，其主要用來描述聲音訊號長期之頻譜變化，因此可以呈現每一特定頻率(或頻帶)隨時間變化之過程。調變頻譜是組合聲音頻率及調變頻率之二維表示方式，若對一特定聲音頻率追蹤其隨時間變化之軌跡，稱之為流動頻譜，而全部頻率之流動頻譜則構成二維調變頻譜。

一般而言，以音節為辨識單元較以音框為辨識單元可以得到較高之正確率，而整首歌曲之辨識率又比單一音節之辨識率來的高，其實一首歌曲中所包含之訊息不止是短時間之頻率音色特性，還包含有隨時間而變化之結構，為了描述此一性質，有些研究者嘗試將短時間之音框特徵整合成單一之長時間之音節特徵[5, 19]，或是使用歌曲中相鄰音節之關連性[13, 17, 20]，或是採用二維頻譜/倒頻譜分析[23]，或是採用調變頻譜分析[17, 28, 33, 34]等。

一旦將聲音特徵擷取出來，就可以設計一分類器來辨識鳥類鳴聲，曾經用來辨識鳥類鳴聲之監督式/非監督式之分類方法有動態時間較正(dynamic time warpping, DTW)[3, 4, 6, 13]、隱藏馬可夫模型(hidden Markov model, HMM)[4, 6, 15, 17]、多層認知網路(supervised multilayer perceptron, MLP)[7, 9-11, 16]、非監督式自我組織圖(unsupervised self-organizing maps, SOM)[16, 20]、單值型態遞迴類神經模糊網路架構(singleton-type recurrent neural fuzzy networks, SRNFN)[8]、二次區別分析演算法(quadratic discriminant analysis, QDA)[9-11]、線性區別分析演算法(linear discriminant analysis, LDA)[21-23]、高斯混合模型(Gaussian mixture model, GMM)[17]、向量支持機(support vector machine, SVM)[18]、及貝氏推理法[5, 19]等。

如前所述，用來辨識鳥類鳴聲之聲音特徵主要是源自語音/語者辨識或音訊/音樂之分類，然而鳥類專家可以從觀察聲譜圖/聲紋圖來辨識鳥類之種類，因此本計劃中我們嘗試以影像處理之方法來辨識鳥類鳴聲。首先我們將鳥類鳴聲之聲譜圖視為一張影像，而鳥類鳴聲中較顯著之頻率變化片段在聲譜圖影像中則顯示出明顯之邊緣，如圖一所示，因此我們以MPEG-7之徑角轉換(angular radial transform, ART)[35]來擷取聲譜圖影像中之外形變化特徵，用以辨識鳥類鳴聲。

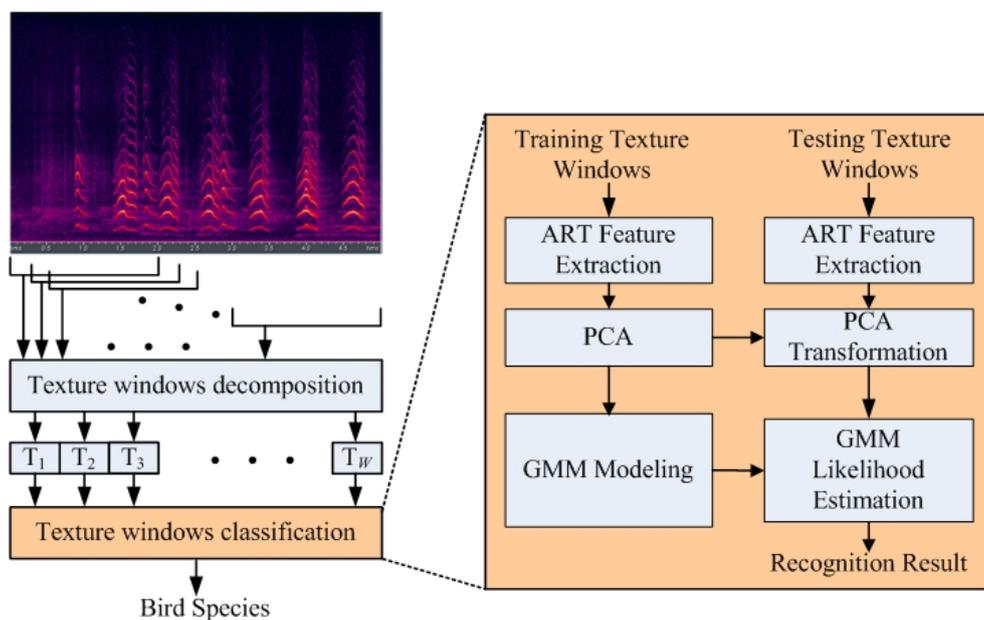


圖一、不同鳥類鳴聲之聲譜圖 (a) 火冠戴菊鳥(Taiwan Firecrest) (b) 白耳畫眉(Taiwan Sibia) (c) 黃腹琉璃(Vivid Niltava) (d) 鳳頭蒼鷹(Crested Goshawk)

## 2. 研究目的與研究方法

本計劃中我們採用影像之形狀特徵於聲音訊號之辨識特徵擷取，再應用高斯混合模型於連續鳥類鳴聲之辨識。

首先我們假設欲辨識之鳥類鳴聲是一固定長度之聲音片段(實驗之鳥類鳴聲長度分別為 3 秒鐘及 5 秒鐘)，而非事先以人工方式或自動切割之音節，以避免過多之人為介入。此一自動化連續鳥類鳴聲辨識系統包含兩個階段，分別為訓練階段和辨識階段，訓練階段是由四個主要模組所組成：分析視窗切割、ART 特徵擷取、PCA 分析、及 GMM 模型建立。辨識階段一樣是由四個主要模組所組成：分析視窗切割、ART 特徵擷取、PCA 轉換、及 GMM 分類，圖二為本辨識系統之架構圖。



圖二、連續鳥類鳴聲辨識系統架構圖

### 一. 分析視窗切割

在本計劃之連續鳥類鳴聲辨識系統中，我們先將輸入之 3 秒鐘或 5 秒鐘之連續鳥類鳴聲切割為數個固定長度(2 秒鐘)之分析視窗，相鄰之分析視窗間重疊約 1.75 秒，因此每一 3 秒鐘或 5 秒鐘之聲音片段可以分別切割 5 個或 13 個分析視窗，為了避免受到雜音或背景聲音之影響，我們計算每一分析視窗之能量值，如果其能量值比起最大能量值之分析視窗還低 40 dB，我們就將此分析視窗視為靜音視窗，其餘之分析視窗則視為有

聲視窗，對於靜音視窗我們就不做辨識動作，最後我們將全部有聲視窗以 GMM 來決定此一 3 秒鐘或 5 秒鐘之連續鳥類鳴聲之辨識類別。

## 二. ART 特徵擷取

對大部分鳥類鳴聲，特別是具有明顯泛音之鳥類鳴聲，我們可以在其聲譜圖之頻率軸發現明顯週期性之性質，同樣的在其聲譜圖之時間軸上也會呈現或多或少之規律性(請參考圖一)，為了描述這些性質，我們採用 MPEG-7 之 ART 轉換自其聲譜圖中擷取辨識特徵，用來辨識鳥類鳴聲。

### (一) MPEG-7 ART 轉換

在 MPEG-7 標準中，ART 轉換係數是描述二維影像形狀特徵之方法[35]，而 ART 係數  $f_{nm}$  之定義如下：

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

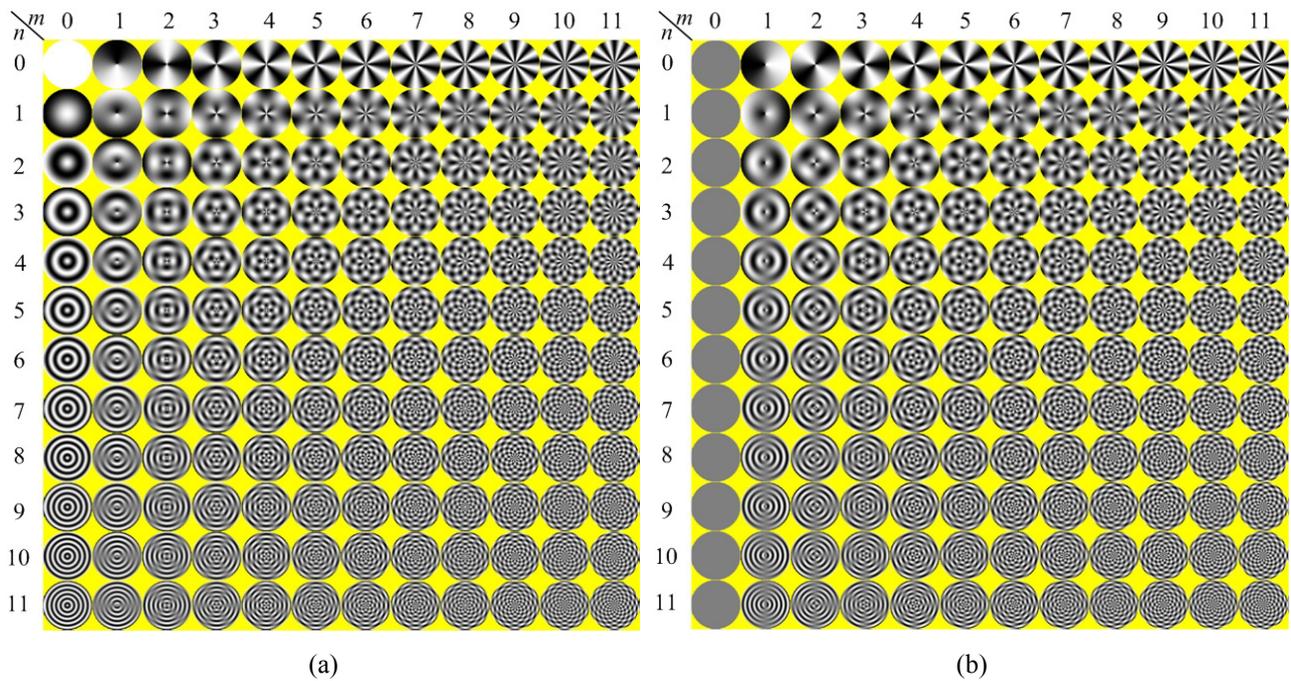
其中  $f(\rho, \theta)$  為極座標平面上之影像灰階值， $V_{nm}(\rho, \theta)$  為 ART 之基底函數，而且可分離為角度(angular)及徑度(radial)變化之函數：

$$V_{nm}(\rho, \theta) = A_m(\theta) R_n(\rho)$$

$A_m(\theta)$  和  $R_n(\rho)$  可分別以指數函數及餘弦函數來表示：

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

其中  $m$  及  $n$  分別代表 ART 的角度係數和半徑係數。圖三顯示 ART 基底函數之實部與虛部值，我們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)。

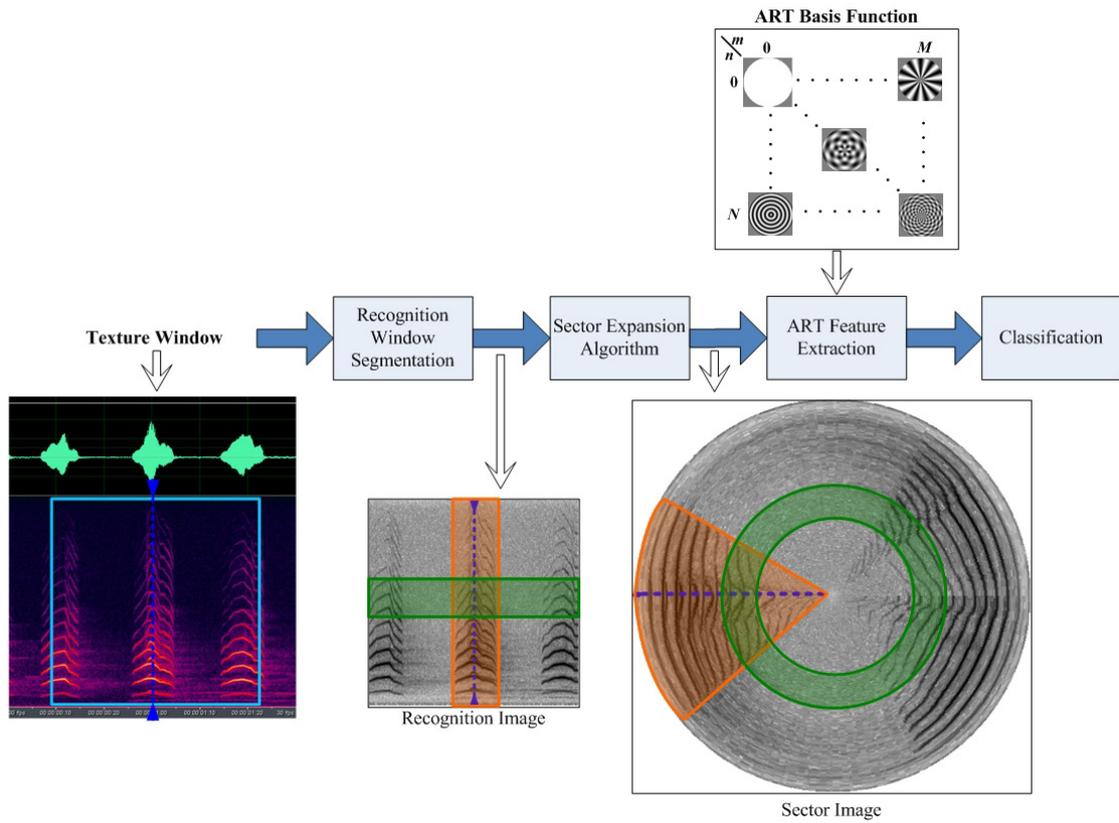


圖三、 ART 基底函數實虛部變化情形( $0 \leq m < 12, 0 \leq n < 12$ ) (a)實部 (b)虛部。

## (二) ART 特徵擷取

對於每一個有聲視窗，我們將其分割為多個部分重疊之音框，音框長度為 512 個訊號樣本，相鄰音框間重疊 256 個訊號樣本，然後我們找出具有最大能量之音框，加上其前面之 127 個音框及後面之 128 個音框，構成具有 256 個音框之辨識視窗，每一音框取其前 256 個 FFT 係數值，即可將此一辨識視窗視為  $256 \times 256$  之影像，然後我們將每一個 ART 基底函數與此  $256 \times 256$  之辨識視窗影像做旋積(convolution)運算即可得到一組 ART 係數。

然而，我們觀察發現視窗影像中之外形變化主要是沿著垂直之頻率軸及水平之時間軸之方向，而 ART 是描述二維影像形狀中角度及徑度之變化趨勢，為了有效利用 ART 之特性，我們提出一演算法將辨識視窗影像之每一列轉換為一個圓，每一行轉換為一條輻形線，因此相鄰之幾列(橫條)可轉換為一環形區域，而相鄰之幾行(直條)可轉換為一扇形區域(請參考圖四)，我們將此轉換稱為扇形展開演算法(sector expansion algorithm)且將轉換後之影像稱為扇形影像(sector image)，之後我們將每一個 ART 基底函數與此扇形影像(不是辨識視窗影像)做旋積運算即可得到一組 ART 係數。



圖四、ART 特徵擷取示意圖

假設  $I(f, t)$  為辨識影像，其中  $f (0 \leq f \leq 255)$  代表頻率變數而  $t (0 \leq t \leq 255)$  代表時間變數， $I_S(u, v) (0 \leq u, v \leq 511)$  為轉換後之扇形影像，其中頻率變數  $f$  將與扇形影像之徑度變數  $\rho (0 \leq \rho \leq 255)$  重疊，時間變數  $t$  將與扇形影像之角度變數  $\theta (0 \leq \theta \leq 2\pi)$  重疊，將辨識影像映射至扇形影像之關係式如下(請參考圖五)：

$$u = 256 - \Delta u$$

$$v = 256 + \Delta v$$

其中

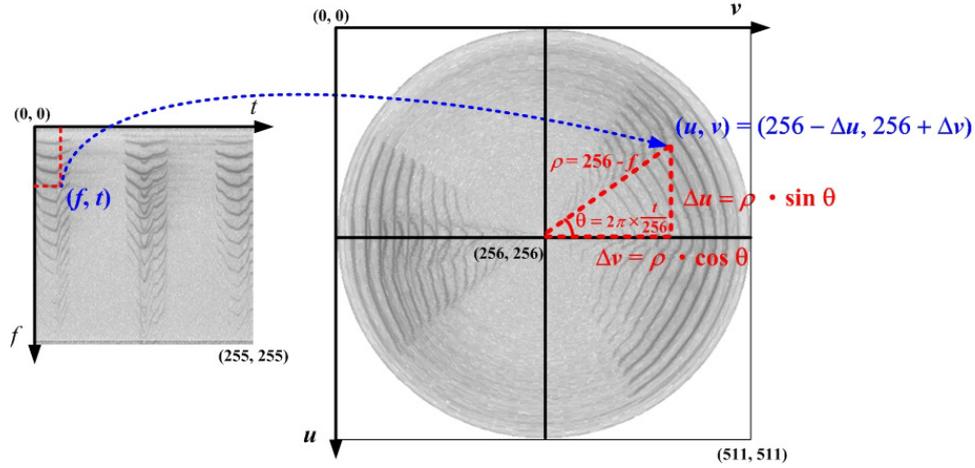
$$\Delta u = \rho \sin \theta$$

$$\Delta v = \rho \cos \theta$$

且

$$\theta = t \times \frac{2\pi}{256}$$

$$\rho = f$$



圖五、辨識影像映射至扇形影像之對應關係

我們將每一個 ART 基底函數與此扇形影像做旋積運算即可得到一組 ART 係數  $f_{nm}$ ：

$$f_{nm} = \langle V_{nm}(\rho, \theta), I_S(u, v) \rangle$$

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 ( $m=0, n=0$ ) 的 ART 係數，即為 ART 特徵描述子：

$$\text{ART}_{nm} = \frac{f_{nm}}{f_{00}}$$

因此，ART 特徵向量可以表示為

$$\begin{aligned} \mathbf{f}_{ART} &= [f_{ART}(1), f_{ART}(2), \dots, f_{ART}(N \times M - 1)]^T \\ &= [f(0, 1), \dots, f(0, M - 1), f(1, 0), \dots, f(1, M - 1), \\ &\quad \dots, f(N - 1, 0), \dots, f(N - 1, M - 1)]^T \end{aligned}$$

此外，為了避免某些特別大或特別小之 ART 特徵值影響辨識結果，我們以下列方程式將每一個 ART 特徵值再做一次正規化：

$$x_{ART}(i) = \begin{cases} 1, & f_{ART}(i) > P_{95} \\ \frac{f_{ART}(i) - P_5}{P_{95} - P_5}, & P_5 \leq f_{ART}(i) \leq P_{95}, 0 \leq i \leq N \times M - 1 \\ 0, & f_{ART}(i) < P_5 \end{cases}$$

其中  $P_{95}$  (或  $P_5$ ) 分別代表於所有訓練之特徵向量中有 95% (或 5%) 之特徵值小於或等於此一數值，因此，最後之 ART 特徵向量可以表示如下：

$$\mathbf{x}_{ART} = [x_{ART}(1), x_{ART}(2), \dots, x_{ART}(N \times M - 1)]^T$$

### 三. PCA 轉換

PCA 是先計算所有訓練資料之特徵向量的平均變異數矩陣  $E[\mathbf{X}\mathbf{X}^T]$  之 eigenvalue 及 eigenvector [36]，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了分佈的寬廣，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 eigenvalue 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

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

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

步驟 2：計算平均變異數矩陣， $\mathbf{C}$

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

步驟 3：求取變異數矩陣  $\mathbf{C}$  的 eigenvalue 及 eigenvector 並將其依 eigenvalue 值由大至小重新排序

步驟 4：設定臨界值  $\alpha$  (表示所要保留的資訊量程度)，以計算轉換後維度  $d$

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

其中  $\lambda_i$  表示第  $i$  大之 eigenvalue， $D$  為轉換前之維度

步驟 5：以所保留之  $d$  個 eigenvector 對所有資料作線性轉換

$$\mathbf{x}_{\text{PCA}} = \mathbf{A}_{\text{PCA}}^T \mathbf{x}'_i$$

其中  $\mathbf{A}_{\text{PCA}}$  為此  $d$  個較大 eigenvector 構成之 PCA 轉換矩陣。

### 四. GMM 模型建立

由於鳥類鳴聲之變化相當豐富，因此僅以一組特徵向量來代表一種鳥類聲音必然是不夠，所以我們將使用高斯混合模型(Gaussian mixture model, GMM)來描述多變化之鳥類鳴聲。傳統上，使用 GMM 模型於分類辨識時，對每一類別之鳥類鳴聲需分別建立其

GMM 模型，而 GMM 模型之參數通常是使用 EM(Expectation Maximization)演算法[37] 來預測，此演算法主要是用在預測 GMM 模型中多變數機率分佈函數之參數值，其目的是找到最佳之參數集合  $\Theta$  使得  $p(\mathbf{X}|\Theta)$  最大，其中  $\mathbf{X} = \{\mathbf{x}_t, t = 1, 2, \dots, T\}$  為訓練資料之集合， $T$  為訓練資料之數目； $\Theta = \{\pi_g, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g | g = 1, 2, \dots, G\}$ ， $\pi_g$  為在 GMM 模型中第  $g$  個高斯分佈之事前機率(prior probability)， $\boldsymbol{\mu}_g$  為平均值向量， $\boldsymbol{\Sigma}_g$  為共變異數矩陣(covariance matrix)， $G$  為 GMM 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 GMM 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

計算所有資料屬於 GMM 模型中每一高斯分佈之機率比值作為預測值：

$$p(\theta_g | \mathbf{x}_t) = \frac{\pi_g p(\mathbf{x}_t | \theta_g)}{\sum_{r=1}^G \pi_r p(\mathbf{x}_t | \theta_r)}$$

其中

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

### 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、平均值向量、共變異數矩陣)：

$$\begin{aligned} \bar{\pi}_r &= \frac{1}{T} \sum_{t=1}^T p(\theta_r | \mathbf{x}_t), \\ \bar{\boldsymbol{\mu}}_r &= \frac{\sum_{t=1}^T p(\theta_r | \mathbf{x}_t) \mathbf{x}_t}{\sum_{t=1}^T p(\theta_r | \mathbf{x}_t)}, \\ \bar{\boldsymbol{\Sigma}}_r &= \frac{\sum_{t=1}^T p(\theta_r | \mathbf{x}_t) (\mathbf{x}_t - \bar{\boldsymbol{\mu}}_r) (\mathbf{x}_t - \bar{\boldsymbol{\mu}}_r)^T}{\sum_{t=1}^T p(\theta_r | \mathbf{x}_t)} \end{aligned}$$

步驟 4: 重覆執行步驟 2-3，直到收斂為止。

## 五. 辨識階段

在辨識階段，我們先將輸入之鳥類鳴聲切割為數個分析視窗，相鄰之分析視窗間重疊 1.75 秒，我們計算每一分析視窗之能量值，如果其能量值比起最大能量值之分析視窗還低 40 dB，我們就將此分析視窗視為靜音視窗，否則將其視為有聲視窗，我們將靜音視窗刪除不做辨識動作，然後對全部有聲視窗中擷取其 ART 特徵向量，假設此輸入之鳥類鳴聲中有  $W$  個有聲視窗(以  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_W$  表示)，其 ART 特徵向量以  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_W$  表示，接著以 PCA 轉換矩陣  $\mathbf{A}_{\text{PCA}}$  將每一特徵向量  $\mathbf{x}_k$  轉換為較短之特徵向量：

$$\mathbf{f}_k = \mathbf{A}_{\text{PCA}}^T \mathbf{x}_k$$

假設鳥類鳴聲之辨識類別有  $S$  種，其 GMM 參數集合為  $\Theta^s = \{\pi_g^s, \boldsymbol{\mu}_g^s, \boldsymbol{\Sigma}_g^s \mid g = 1, \dots, G\}$ ,  $s = 1, 2, \dots, S$ ，以最大對數相似度(maximum log-likelihood)值來決定辨識之類別：

$$\hat{s} = \arg \max_{1 \leq s \leq S} \sum_{k=1}^W \log p(\mathbf{f}_k \mid \Theta^s)$$

## 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，訓練資料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，表一顯示實驗之鳥類名稱及每種鳥類之辨識鳴聲片段數目。在此實驗中我們對每一鳴聲片段擷取其 ART 特徵，表二比較各種特徵向量對於 28 種鳥類之辨識率，我們比較之特徵向量有線性預測倒頻譜係數(LPCC)、梅爾倒頻譜係數(MFCC)、及各類二維梅爾倒頻譜係數(TDMFCC、DTDMFCC 及 SDTDMFCC)等，其中線性預測倒頻譜係數及梅爾倒頻譜係數之特徵擷取是以短時間之音框(約 0.1 秒)為單位，ART 特徵及各類二維梅爾倒頻譜係數之特徵擷取是以較長時間之分析視窗(約 2 秒)為單位。實驗結果顯示二維梅爾倒頻譜係數之辨識率比起以音框為單位線性預測倒頻譜係數及梅爾倒頻譜係數還要高，而 ART 特徵有最佳之辨識正確率。當鳴聲片段長度為 3 秒及 5 秒時，最佳辨識率分別為 86.30%及 94.62%。

表一、28 種鳥類之不同長度( $D$ )之訓練鳴聲片段數目( $N^{Tr}$ )與測試鳴聲片段數目( $N^{Te}$ )

中文名稱	英文名稱	拉丁文學名	$D = 3 \text{ seconds}$		$D = 5 \text{ seconds}$	
			$N^{Tr}$	$N^{Te}$	$N^{Tr}$	$N^{Te}$
大冠鷲	Crested Serpent Eagle	Spilornis cheela	107	5	105	3
小卷尾	Bronzed Drongo	Dicrurus aeneus	128	10	126	8
小啄木	Gray-headed Pygmy Woodpecker	Dendrocopos canicapillus	50	9	48	7
小翼鶇	Blue Shortwing	Brachypteryx montana	172	6	170	4
小彎嘴畫眉	Streak-breasted Scimitar Babbler	Pomatorhinus ruficollis	147	16	145	4
火冠戴菊鳥	Taiwan Firecrest	Regulus goodfellowin	92	10	90	8
白耳畫眉	Taiwan Siberia	Heterophasia auricularis	97	5	95	3
白喉笑鶇	White-throated Laughing Thrush	Garrulax albogularis	61	8	59	6
白腹秧雞	White-breasted Water Hen	Amauromis phoenicurus	83	6	81	4
灰鶯	Beavan's Bullfinch	Pyrrhula erythaca	104	3	102	1
竹鳥	Gray-sided Laughing Thrush	Garrulax caeruleus	77	79	75	77
岩鶇	Alpine Accentor	Prunella collaris	62	9	60	7
青背山雀	Green-backed Tit	Parus monticolus	127	4	125	2
冠羽畫眉	Taiwan Yuhina	Yuhina brunneiceps	62	6	60	4
紅頭山雀	Red-headed Tit	Aegithalos concinnus	98	9	96	7
栗背林鴉	Collared Bush Robin	Erithacus johnstoniae	147	5	145	3
烏頭翁	Taiwan Bulbul	Pycnonotus taivanus Styan	58	8	56	6
深山竹雞	Taiwan Hill Partridge	Arborophila crudigularis	141	10	139	8
深山鶯	Verreaux's Bush Warbler	Cettia acanthizoides	72	8	70	6
筒鳥	Oriental Cuckoo	Cuculus saturatus	124	10	122	8
黃山雀	Taiwan Tit	Parus holsti	116	7	114	5
黃腹琉璃	Vivid Niltava	Niltava vivida	91	8	89	6
煤山雀	Colal Tit	Parus ater	105	10	103	8
鳳頭蒼鷹	Crested Goshawk	Accipiter trivirgatus	113	11	111	9
頭烏線	Gould's Fulvetta	Alcippe brunnea	41	7	39	5
鸚鵡	Collared Pigmy Owlet	Glaucidium brodiei	59	16	57	9
藍腹鵝	Swinhoe's Pheasant	Lophura swinhoii	92	5	90	3
藪鳥	Steere's Liocichla	Liocichla steerii	57	6	55	4
Total number of birdsong segments			2683	296	2627	225

表二、比較各種特徵向量對於 28 種鳥類之辨識率( $CA$ )，參數  $G$  為 GMM 模型中高斯分佈之群數， $\alpha$  為 PCA 門檻值

Descriptor	$D = 3$		$D = 5$	
	$CA$ (%)	$(G, \alpha)$	$CA$ (%)	$(G, \alpha)$
LPCC	30.41	(50, 0.98/0.99)	40.00	(30, 0.99)
MFCC	46.62	(35, 0.98/0.99)	56.89	(45, 0.95/0.96/0.97)
TDMFCC	69.86	(10, 0.96)	77.13	(5, 0.95)
DTDMFCC	76.03	(5, 0.99)	83.86	(10, 0.99)
SDTDMFCC	73.63	(10, 0.95)	79.82	(10, 0.95/0.96)
ART	86.30	(5, 0.97/0.98)	94.62	(5, 0.95/0.97)

表三呈現鳴聲片段長度為5秒之每種鳥類鳴聲之個別辨識率，其中ART基底函數為 $6 \times 24$ ，GMM模型中高斯分佈之群數 $G = 5$ ，PCA門檻值 $\alpha = 0.97$ 。由此表我們發現除了小啄木(Gray-headed Pygmy Woodpecker)及青背山雀(Green-backed Tit)外，大部份鳥類鳴聲之辨識率都可以接受，仔細觀察辨識率較差之原因可歸納有二：訓練樣本數目較少及背景

雜音之影響，因此未來將收集更多之訓練樣本以及提出可減少背景雜音之演算法以提升辨識率。

表三、鳴聲片段長度為 5 秒之每種鳥類鳴聲之個別辨識率(CA)

類別代碼	中文名稱	英文名稱	$N^{Tr}$	$N^{Te}$	CA (%)
1	大冠鷲	Crested Serpent Eagle	105	3	100.00
2	小卷尾	Bronzed Drongo	126	8	100.00
3	小啄木	Gray-headed Pygmy Woodpecker	48	7	0.00
4	小翼鶇	Blue Shortwing	170	4	100.00
5	小彎嘴畫眉	Streak-breasted Scimitar Babbler	145	4	100.00
6	火冠戴菊鳥	Taiwan Firecrest	90	8	100.00
7	白耳畫眉	Taiwan Sibia	95	3	100.00
8	白喉笑鶇	White-throated Laughing Thrush	59	6	50.00
9	白腹秧雞	White-breasted Water Hen	81	4	100.00
10	灰鷲	Beavan's Bullfinch	102	1	100.00
11	竹鳥	Gray-sided Laughing Thrush	75	77	100.00
12	岩鶇	Alpine Accentor	60	7	100.00
13	青背山雀	Green-backed Tit	125	2	0.00
14	冠羽畫眉	Taiwan Yuhina	60	4	100.00
15	紅頭山雀	Red-headed Tit	96	7	100.00
16	栗背林鶇	Collared Bush Robin	145	3	100.00
17	烏頭翁	Taiwan Bulbul	56	6	100.00
18	深山竹雞	Taiwan Hill Partridge	139	8	100.00
19	深山鶲	Verreaux's Bush Warbler	70	6	100.00
20	筒鳥	Oriental Cuckoo	122	8	100.00
21	黃山雀	Taiwan Tit	114	5	100.00
22	黃腹琉璃	Vivid Niltava	89	6	100.00
23	煤山雀	Coal Tit	103	8	100.00
24	鳳頭蒼鷹	Crested Goshawk	111	9	100.00
25	頭烏線	Gould's Fulvetta	39	5	100.00
26	鶇鶇	Collared Pigmy Owlet	57	9	100.00
27	藍腹鶇	Swinhoe's Pheasant	90	3	100.00
28	藪鳥	Steere's Liocichla	55	4	75.00

## 二. 參考文獻

- [1] <http://www.bird.org.tw>
- [2] E. A. Brenowitz, D. Margoliash, and K. M. Nordeen, "An introduction to birdsong and the avian song system," *J. Neurobiology*, vol. 33, no. 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," *J. Acoust. Soc. Amer.*, vol. 100, no. 2, pp.1209-1219, Aug. 1996.
- [4] H. Tyagi, R. M. Hegde, H. A. Murthy, and A. Prabhakar, "Automatic identification of bird calls using spectral ensemble average voice prints," in *Proc. 13th European Signal Process. Conf. (EUSIPCO '06)*, Florence, Italy, Sep. 2006.

- [5] F. Briggs, R. Raich, and X. Z. Fern, "Audio classification of bird species: a statistical manifold approach," in *Proc. 9th IEEE Int. Conf. Data Mining*, 2009, pp. 51-60.
- [6] 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," *J. Acoust. Soc. Amer.*, 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 *Proc. IEEE Conf. Commun., Power, Computing*, 1995, vol. 2, pp. 409-414.
- [8] 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, Dec. 2007.
- [9] A. L. McIlraith and H. C. Card, "A comparison of backpropagation and statistical classifiers for bird identification," in *Proc. IEEE Int. Conf. Neural Networks*, 1997, vol. 1, pp. 100-104.
- [10] A. L. McIlraith and H. C. Card, "Birdsong recognition using backpropagation and multivariate statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2740-2748, Nov. 1997.
- [11] A. L. McIlraith and H. C. Card, "Bird song identification using artificial neural networks and statistical analysis," in *Proc. Canadian Conf. Elect. Comput. Eng.*, 1997, vol. 1, pp. 63-66.
- [12] A. Härmä, "Automatic identification of bird species based on sinusoidal modeling of syllables," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 5, pp. 545-548.
- [13] P. Somervuo and A. Härmä, "Bird song recognition based on syllable pair histograms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2004, vol. 5, pp. 825-828.
- [14] Z. Chan and R. C. Maher, "Semi-automatic classification of bird vocalizations using spectral peak tracks," *J. Acoust. Soc. Amer.*, vol. 120, no. 5, pp. 2974-2984, Nov. 2006.
- [15] C. H. Chou, C. H. Lee, and H. W. Ni, "Bird species recognition by comparing the HMMs of the syllables," in *Proc. 2nd Int. Conf. Innovative Comput, Inf. Control*, Kumamoto, Japan, Sep. 2007.
- [16] A. Selin, J. Turunen, and J. T. Tantt, "Wavelets in recognition of bird sounds," *EURASIP J. Adv. Signal Process.*, vol. 2007, Article ID 51806, 9 pages.
- [17] P. Somervuo, A. Härmä, and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 2252-2263, Nov. 2006.
- [18] A. Fagerlund, "Bird species recognition using support vector machines," *EURASIP J. Adv. Signal Process.*, vol. 2007, Article ID 38637, 8 pages.
- [19] B. Lakshminarayanan, R. Raich, and X. Fern, "A syllable-level probabilistic framework for bird species identification," in *Proc. IEEE Int. Conf. Mach. Learning Appl.*, 2009, pp. 53-59.
- [20] E. E. Vallejo, M. L. Cody, and C. E. Taylor, "Unsupervised acoustic classification of bird species using hierarchical self-organizing maps," in *Proc. 3rd Australian conf. Progress Artificial Life*, 2007, pp. 212-221.
- [21] C. H. Lee, Y. K. Lee and R. Z. Huang, "Automatic recognition of bird songs using cepstral

- coefficients,” *J. Inf. Technol. Appl.*, vol. 1, no. 1, pp. 17-23, May 2006.
- [22] C. H. Lee, C. C. Lien and R. Z. Huang, “Automatic recognition of birdsongs using Mel-frequency cepstral coefficients and vector quantization,” in *Proc. Int. MultiConference Eng. Comput. Scientists*, Hong Kong, 2006, pp. 331-335.
- [23] C. H. Lee, C. C. Han, and C. C. Chuang, “Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1541-1550, Nov. 2008.
- [24] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Prentice-Hall, 1993.
- [25] J. J. Aucouturier and F. Pachet, “Representing music genres: a state of the art,” *J. New Music Research*, vol. 32, no. 1, pp. 83-93, 2003.
- [26] B. Kingsbury, N.Morgan, and S. Greenberg, “Robust speech recognition using the modulation spectrogram,” *Speech Commun.*, vol. 25, no. 1, pp. 117-132, 1998.
- [27] V. Tyagi, I. McCowan, H. Misra, and H. Bourlard, “Mel-cepstrum modulation spectrum (MCMS) features for robust ASR,” in *Proc. Workshop Automat. Speech Recognition and Understanding*, 2003.
- [28] T. Kinnunen, “Joint acoustic-modulation frequency for speaker recognition,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. 1, pp. 14-19.
- [29] F. Mörchen, A. Ultsch, M. Thies and I. Löhken, “Modeling timbre distance with temporal statistics from polyphonic music,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 81-90, Jan. 2006.
- [30] C. H. Lee, J. L. Shih, K. M. Yu and J. M. Su, “Automatic music genre classification using modulation spectral contrast feature,” in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2007, pp. 204-207.
- [31] S. Sukittanon, L. E. Atlas, and J. W. Pitton, “Modulation-scale analysis for content identification,” *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 3023-3035, Oct. 2004.
- [32] C. H. Lee, J. L. Shih, K. M. Yu, and H. S. Lin, “Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features,” *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 670-682, June 2009.
- [33] N. C. Singh and F. E. Theunissen, “Modulation spectra of natural sounds and ethological theories of auditory processing,” *J. Acoust. Soc. Amer.*, vol. 114, no. 6, pp. 3394-3411, Dec. 2003.
- [34] H. Wang, S. Saar, O. Tchernichovski, and P. P. Mitra, “Characterization of birdsong using spectral correlations,” presented at *37th Annu. Meeting of the Soc. for Neuroscience*, 2007.
- [35] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: multimedia content description interface*, Wiley, 2002.
- [36] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. New York: Wiley, 2000.
- [37] D. Jurafsky and J. H. Martin, *Speech and Language Processing, 2nd Ed.* NJ: Pearson Education Inc., 2009.

### 三. 計畫成果自評

生態保育及生物多樣性資訊調查研究越來越受到全世界各國之重視，對於鳥類之生態調查已廣泛用於評估關於氣候變遷、農業開發及人類干擾等因素對於生態保育及生物多樣性之影響。鳥類的群聚數量及分佈通常被視為生態環境健康與否之指標，因此鳥類生態調查結果可以提供政府相關單位許多資訊，包含棲地保留及瀕臨絕種物種之保護計畫與行動等。台灣野生鳥類有 601 種，包括 24 種特有種和 59 種特有亞種，由於種類繁多，在做生態調查時，若以人工的方式來進行，相當費力耗時，若能利用鳥類的鳴聲來自動辨識其種類，可以節省相當多的人力與時間，因此我們希望對鳥類鳴聲之自動辨識做一深入之研究，來調查各地區之鳥類族群分佈狀況及其棲地之變化情形，建立更完善的生態保育體系。

本計畫中我們以影像處理之方法來辨識鳥類鳴聲。首先我們將鳥類鳴聲之聲譜圖視為一張影像，再以扇形展開演算法將聲譜圖影像轉換成扇形影像，如此可以有效利用 MPEG-7 之徑角轉換來擷取影像中之外形變化特徵，提高辨識率。完成之工作項目如下：

1. 以 MPEG-7 徑角轉換(angular radial transform, ART)自聲音訊號之聲譜圖擷取辨識特徵之演算法。
2. 將聲譜圖影像轉換成扇形影像之演算法。
3. 降低特徵向量維度之 PCA 演算法。
4. GMM 辨識演算法。

歷年來我們已經發表與聲音訊號辨識分類相關之論文如下：

期刊論文 (Journal Papers)：

- [1] C. H. Lee, S. B. Hsu, J. L. Shih, and C. H. Chou, "Continuous Birdsong Recognition Using Gaussian Mixture Modeling of Image Shape Features", *IEEE Trans. on Multimedia*, Vol. 15, No. 2, Feb. 2013, pp. 454-463.
- [2] C. H. Lee, C. H. Chou, and J. C. Fang, "Automatic Music Genre Classification Using Modulation Spectral Features and Nonparametric Discriminant Analysis", *Journal of Information Technology and Applications*, Vol. 5, No. 2, June 2011, pp. 75-82.
- [3] C. H. Lee, J. L. Shih, K. M. Yu, and H. S. Lin, "Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features", *IEEE Trans. on Multimedia*, Vol. 11, No. 4, June 2009, pp. 670-682. (SCI, EI)
- [4] 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)
- [5] **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)
- [6] **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.
- [7] J. L. Shih, **C. H. Lee**, and S. W. Lin, “Automatic classification of musical audio signals”, *Journal of Information Technology and Applications*, Vol. 1, No. 2, Sep. 2006, pp. 95-105.

研討會論文 (Conference Papers) :

- [1] **C. H. Lee**, C. H. Chou, C. C. Lien, and J. C. Fang, “Music Genre Classification Using Modulation Spectral Features and Multiple Prototype Vectors Representation”, in *Proc. of the 4th International Congress on Image and Signal Processing (CISP'11)*, Oct. 15-17, 2011, Shanghai, China, pp. 2762-2766. (NSC-99-2221-E-216-048, EI)
- [2] **C. H. Lee**, H. S. Lin, C. H. Chou, and J. L. Shih, “Modulation Spectral Analysis of Static and Transitional Information of Cepstral and Spectral Features for Music Genre Classification”, in *Proc. of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'2009)*, Sep. 12-14, 2009, Kyoto, Japan, pp. 1030-1033. (EI)
- [3] **C. H. Lee**, J. L. Shih, K. M. Yu, H. S. Lin, and M. H. Wei, “Fusion of Static and Transitional Information of Cepstral and Spectral Features for Music Genre Classification”, in *Proc. of the IEEE Asia-Pacific Services Computing Conference*, Dec. 9-12, 2008, Yilan, Taiwan. (EI)
- [4] **C. H. Lee**, J. L. Shih, K. M. Yu and H. S. Lin, “Modulation Spectral Analysis of Audio Features for Music Genre Classification”, in *Proc. of the 21th IPPR Conference on Computer Vision, Graphics, and Image Processing*, Yilan, Aug. 24-26, 2008.
- [5] 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. (EI)
- [6] **C. H. Lee**, J. L. Shih, K. M. Yu and J. M. Su, “Automatic Music Genre Classification Using Modulation Spectral Contrast Feature”, in *Proceedings of IEEE International Conference on Multimedia and Expo*, Beijing China, July 2007, pp. 204-207. (EI)
- [7] **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.
- [8] **C. H. Lee**, J. L. Shih, and S. W. Lin, “A novel approach to music genre classification”, in *Proceedings of the 18th IPPR Conference on Computer Vision, Graphics, and Image Processing*, Taipei, Aug. 20-22, 2005.

# 行政院國家科學委員會補助國內專家學者出席國際學術會議報告

102 年 1 月 30 日

附件三

報告人姓名	李建興	服務機構 及職稱	中華大學資訊工程系
會議 時間 地點	2012/3/28~2012/4/1 日本，京都	本會核定 補助文號	計劃編號： NSC 100-2221-E-216-035-
會議 名稱	(中文) (英文) The 37th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)		
發表 論文 題目	(中文) (英文) 3D MODEL RETRIEVAL USING 2D CEPSTRAL FEATURES		

## 一、參加會議經過

3/28 搭乘 08:30 長榮航空班機前往日本關西國際機場，隨後轉搭關西日本 JR 國鐵抵達京都，下榻於新都飯店(New Miyako Hotel)。

3/29 早上從京都車站搭乘地鐵前往會議地點—京都國際會議中心(Kyoto International Conference Center)，車程約30分鐘，先註冊報到，我們的論文則是安排於下午之poster session (MMSP-P2: Multimedia Recognition, Search, and Retrieval)，因此會場中有許多語音、影像、多媒體及訊號處理相關論文發表，幾乎世界各國人士都有，會場擠滿人潮，和百貨公司週年慶不相上下，討論相當熱絡。

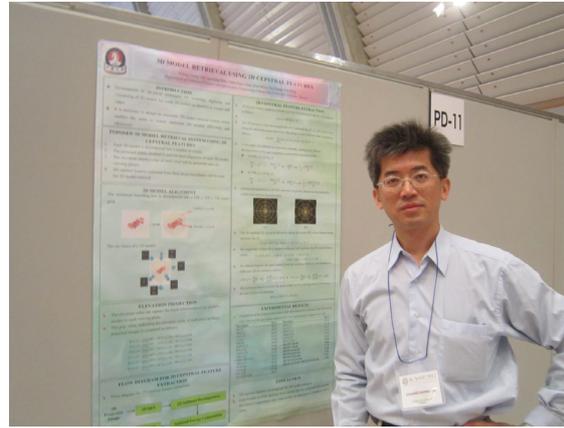
3/30 早上參與 Session AASP-P7: Music: Classification and Recognition 的發表，涵蓋許多音樂訊號分析辨識、樂器辨識、及演唱者辨識等議題。下午則參與 Session AASP-P8: Content Analysis for Music, Multimedia, and Medicine，有許多和音樂、多媒體、及醫療資訊內容分析之相關論文發表，包含 audio event detection, automatic music tagging, speech/music feature extraction, 及 sparse representation 等本人感興趣之議題。

3/31 本日於京都市區觀光，參訪許多京都著名觀光景點，如三十三間堂、清水寺、祇園、二條城等，原本這段時間是京都櫻花盛開之時節，但因今年天氣較往年寒冷，因此只有少數櫻花樹有開花，顯得相當稀疏，有些失望。

4/1 搭乘長榮航空班機返回台灣。



ICASSP 2012 會場-京都國際會議中心



poster session 發表之論文海報

## 二、與會心得

1. 本次會議 ICASSP'2012 是由 IEEE Signal Society 主辦，已經舉辦 37 屆，是每年訊號處理學界之年度盛會，與會之世界各國人士都有，但是相對的很少看到台灣的學界人士，殊為可惜。
2. 主辦地點京都國際會議中心(Kyoto International Conference Center)，除了有一可容納 2000 人之大型會議廳外，尚有十幾個中小型會議廳，軟硬體設施相當完備，充分呈現一流國際城市該有之氣勢。

## 三、考察參觀活動(無是項活動者省略)

## 四、建議

國內也常舉辦國際性研討會，但是會議場所因陋就簡，無法容納大型研討會議。因此國內各縣市該思考如何提供一個可以辦較大型會議及晚宴之場所，是可加強的地方。

## 五、攜回資料名稱及內容

論文行動碟及論文摘要輯

## 六、其他

非常感謝國科會之補助得以參加該研討會。

# 國科會補助計畫衍生研發成果推廣資料表

日期:2013/01/29

國科會補助計畫	計畫名稱: 應用影像形狀特徵於鳥類聲紋辨識之研究
	計畫主持人: 李建興
	計畫編號: 100-2221-E-216-035- 學門領域: 自然語言與語音處理
無研發成果推廣資料	

100 年度專題研究計畫研究成果彙整表

計畫主持人：李建興		計畫編號：100-2221-E-216-035-					
計畫名稱：應用影像形狀特徵於鳥類聲紋辨識之研究							
成果項目		量化			單位	備註(質化說明： 如數個計畫共同 成果、成果列為該 期刊之封面故 事...等)	
		實際已達成 數(被接受 或已發表)	預期總達成 數(含實際已 達成數)	本計畫實 際貢獻百 分比			
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	1	1	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 (本國籍)	碩士生	2	2	100%	人次	
		博士生	1	1	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	1	1	100%	篇	C. H. Lee, S. B. Hsu, J. L. Shih, and C. H. Chou, 'Continuous Birdsong Recognition Using Gaussian Mixture Modeling of Image Shape Features', IEEE Trans. on Multimedia, Vol. 15, No. 2, Feb. 2013, pp. 454-463.
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		章/本
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 (外國籍)	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		

		專任助理	0	0	100%	
其他成果 (無法以量化表達之 成果如辦理學術活 動、獲得獎項、重要 國際合作、研究成果 國際影響力及其他協 助產業技術發展之具 體效益事項等，請以 文字敘述填列。)		無				

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

# 國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

## 1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

## 2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表  未發表之文稿  撰寫中  無

專利： 已獲得  申請中  無

技轉： 已技轉  洽談中  無

其他：（以 100 字為限）

已經發表之論文如下：

C. H. Lee, S. B. Hsu, J. L. Shih, and C. H. Chou, ' Continuous Birdsong Recognition Using Gaussian Mixture Modeling of Image Shape Features' , IEEE Trans. on Multimedia, Vol. 15, No. 2, Feb. 2013, pp. 454-463.

## 3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

生態保育及生物多樣性資訊調查研究越來越受到全世界各國之重視，對於鳥類之生態調查已廣泛用於評估關於氣候變遷、農業開發及人類干擾等因素對於生態保育及生物多樣性之影響。鳥類的群聚數量及分佈通常被視為生態環境健康與否之指標，因此鳥類生態調查結果可以提供政府相關單位許多資訊，包含棲地保留及瀕臨絕種物種之保護計畫與行動等。台灣野生鳥類有 601 種，包括 24 種特有種和 59 種特有亞種，由於種類繁多，在做生態調查時，若以人工的方式來進行，相當費力耗時，若能利用鳥類的鳴聲來自動辨識其種類，可以節省相當多的人力與時間，因此我們希望對鳥類鳴聲之自動辨識做一深入之研究，來調查各地區之鳥類族群分佈狀況及其棲地之變化情形，建立更完善的生態保育體系。

本計畫之研究成果可用以輔助生態調查，節省人力與時間，並能撰寫 APP 應用程式，提供人們戶外休閒賞鳥時，辨別鳥類之種類，豐富賞鳥之樂趣並提供相關之知識與保育觀念。