

# 應用語音辨識技術於鳥鳴聲辨識

廖偉恩

國立台北科技大學

電腦與通訊研究所

t8418044@ntut.edu.tw

黎欣捷

國立台北科技大學

電腦與通訊研究所

t9419004@ntut.edu.tw

蔡偉和

國立台北科技大學

電腦與通訊研究所

whtsai@ntut.edu.tw

## 摘要

野外賞鳥已成為大眾休閒的新趨勢，但一般民眾常只能看見鳥或聽見鳥鳴聲，卻不知其種類為何。為了協助大眾識別鳥類，本論文探討鳥鳴聲自動辨識問題，透過語音辨識相關技術，設計鳥鳴聲辨識系統。我們分別從音色及音高兩個層面進行分析，利用梅爾刻度倒頻譜係數表示鳥鳴聲的音色特徵，並搭配高斯混合模型進行特徵的參數模型化與比對；而音高層面分析則試圖求取鳥鳴聲所對應的音符，再利用雙連文模型捕捉音符的動態變化資訊，並據以比對未知鳥鳴聲。我們挑選出大台北地區常見的十種鳥類，並從商業 CD 及鳥類相關網站上收集鳥鳴聲資料，使系統訓練和測試音檔分別屬不同的來源。實驗結果發現，採用音色、音高、與結合兩者的系統辨識正確率分別為 71.1%、72.1%、與 75.04%。

**關鍵詞：**音色、音高、高斯混合模型、雙連文模型。

## 1. 前言

目前全世界大約有九千七百多種鳥類，而台灣這樣一個小島就佔了約二十分之一的種類，雖然我們周遭住有許多這些可愛的鄰居，但往往都只聽到它們的叫聲，卻不知它們是誰。鳥類的鳴聲豐富且多變，我們期望藉由物種之間的鳴聲差異性，發展出一套鳥鳴聲辨識系統，讓不是鳥類專家的一般民眾，也可以從自己隨意錄製的一段鳥鳴聲音檔中，讓系統判斷所屬鳥種並獲得之相關訊息。

目前鳥鳴聲自動辨識的相關研究仍十分有限。文獻[1]中使用動態時間校正(Dynamic Time Warping, DTW)演算法，將測試聲音檔的頻譜圖(spectrogram)與事先經過專家挑選的樣板音檔作比對。文獻[2]中分別比較 DTW 和隱藏式馬可夫模型(Hidden Markov Model, HMM)在辨識鳥類聲音上的效能，其中使用 6 種特徵參數：線性預估編碼參數(Linear Predict Coding, LPC)、線性預估倒頻譜係數、LPC reflection、梅爾倒頻譜係數(Mel-frequency Cepstral Coefficients, MFCC) [3]、log mel-filter bank channel 和 linear mel-filter bank channel。實驗結果顯示，使用 DTW 的辨識效能不錯，但是對於雜訊較大的輸入聲音檔或是容易混淆的短促鳴叫聲，則需要挑選更適合的頻譜圖樣本，這道手續通常需要經驗豐富的專家反覆地試驗。對於使用 HMM，辨識效果則取決於輸入參數的鑑別度，但效能不一定比 DTW 好。而不論是 DTW 或 HMM，所使用的辨識線索皆來自於音色(timbre)特徵。本論文所提出之辨識系統，除了考慮音色特徵參數外，更加上音高(pitch)特徵參數。實驗結果發現將這兩種特徵參數進行結合後可有效提升鳥鳴聲辨識正確率。

本論文的章節組織如下：第二章描述辨識系統架構；第三章說明音色特徵參數擷取及統計模型；第四章描述音高特徵參數擷取及統計模型；第五章結合音色與音高特徵來設計辨識系統；第六章將說明本研究所使用的鳥鳴聲資料庫並呈現實驗結果；第六章進行本研究之總結。

## 2. 辨識系統架構

本論文所提出的辨識系統如圖 1。我們可將其分成三個單元，分別是「音色分析」、「音高分析」、與「整合決策」。各單元皆包含兩種操作模式，一為訓練、另一為測試，簡介如下：

- (1) 音色分析  
此單元目標是擷取各種鳥之鳴聲的音色特徵，並表示為統計模型，以便識別未知鳥鳴聲。
- (2) 音高分析  
此單元目標是擷取各種鳥之鳴聲的音高特徵，並表示為統計模型，以便識別未知鳥鳴聲。
- (3) 整合決策  
此單元整合音色與音高的判斷資訊，進行最後決策以告知使用者辨識結果。

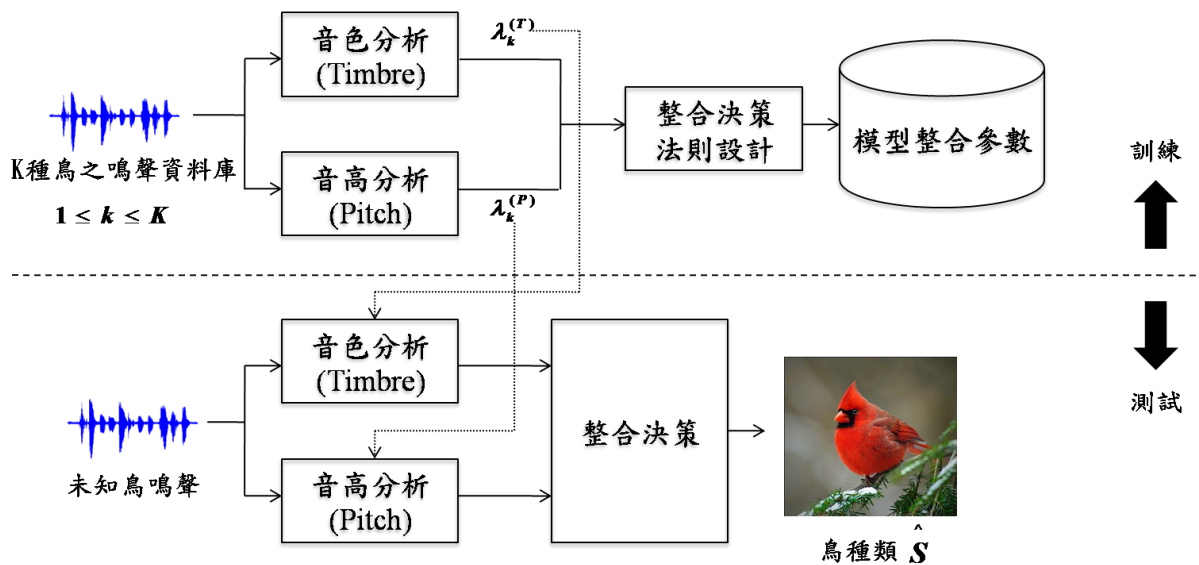


圖 1 本論文之鳥鳴聲辨識系統架構圖。

## 3. 音色分析

音色分析過程如圖 2 所示。主要包括預處理、特徵參數擷取、與統計模型建立及匹配。

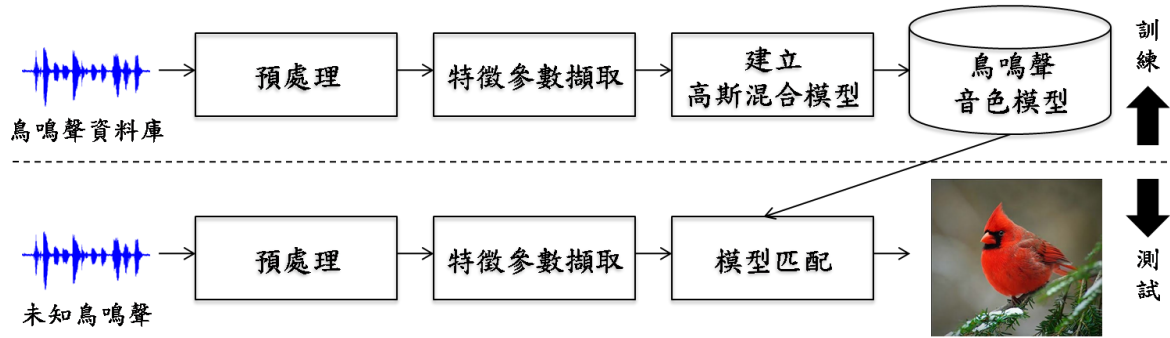


圖 2 音色分析過程。

### 3.1 特徵參數擷取

音色之差異如同鋼琴與吉他都彈奏相同音符，卻聽起來有不同的聲音。不同的鳥種，可視為不同的樂器一樣，音色也不相同。

首先，我們先將鳥鳴聲訊號經由短時間（short-term）快速傅利葉轉換（Fast Fourier Transform, FFT）成為一串音框頻譜序列。令  $x_{t,j}$  為第  $t$  個音框訊號在第  $j$  個 FFT 頻率刻度上的絕對振幅頻譜（magnitude spectrum），其中  $1 \leq j \leq J$ ，而  $J$  為頻率刻度的總數。

再將各音框的絕對振幅頻譜送入一個「三角頻帶組」（triangular filter banks），其中頻帶中心即是梅爾刻度值，該刻度值 mel 與頻率  $f$  Hz 的轉換方式為  $\text{mel}(f) = 2595 \cdot \log_{10}(1 + f/700)$ ，而頻帶的寬度為兩相鄰梅爾刻度差。接著計算各頻帶的能量值  $SG_{t,b} = \sum_{j=l_b}^{u_b} |x_{t,j}|^2$ ，其中  $l_b$  為第  $b$  個頻帶中最低的 FFT 頻率刻度， $u_b$  為第  $b$  個頻帶中最高的 FFT 頻率刻度，然後再將所有頻帶的對數能量值進行離散餘旋轉換（Discrete Cosine Transform, DCT）[4]，以取得倒頻譜係數（cepstral coefficients）[4]，詳細計算方式如下：

$$\mathbf{X}_t = \frac{1}{B} \sum_{b=1}^B \log \left( \sum_{j=l_b}^{u_b} |x_{t,j}|^2 T_b(j) \right) \cos \left( \frac{\pi i}{B} (b-0.5) \right) \quad (1)$$

其中  $B$  為頻帶總數， $T_b(j)$  為第  $b$  個頻帶的三角濾波器。

### 3.2 高斯混合模型(GMM)

為了凝聚同種鳥之不同叫聲的共有音色特徵，我們利用「高斯混合模型」[4][5]來進行 MFCC 參數統計。高斯混合模型是聲學訊號分類中最常見且最成功的模型之一。一個高斯混合模型包含若干高斯機率密度函式，每一高斯機率密度函式以期望值  $\mu_g$  與變異量  $\Sigma_g$  所描述，其中  $g$  代表  $G$  個高斯機率密度函式中之第  $g$  個，另外含一加權數  $w_g$  將各高斯機率密度函式加總成為一機率密度。我們將模型參數記為： $\lambda_k^{(T)} = \{w_g, \mu_g, \Sigma_g | 1 \leq g \leq G\}$ 。這些參數可經由最大化期望值法（Expectation-Maximization, EM）[4][5] 估算出，此即訓練階段所須執行之工作。若資料庫中有  $K$  隻鳥，則我們產生  $K$  個模型  $\lambda_1^{(T)}, \lambda_2^{(T)}, \dots, \lambda_K^{(T)}$ ，代表這些鳥鳴聲的音色模型。

在測試階段，若有一未知鳥鳴聲片段之特徵向量序列為  $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ ，其中每一向量維度為  $D$ ，則我們對各模型  $\lambda_k^{(T)}$ ， $1 \leq k \leq K$  分別計算似然率：

$$P(\mathbf{X}|\lambda_k^{(T)}) = \prod_{k=1}^K P(\mathbf{X}_t|\lambda_k^{(T)}) \quad (2)$$

其中

$$P(\mathbf{X}_t|\lambda_k^{(T)}) = \sum_{g=1}^G w_g \frac{1}{\sqrt{(2\pi)^D |\Sigma_g|}} \exp\left(-\frac{1}{2}(\mathbf{X}_t - \mu_g)' \Sigma_g^{-1} (\mathbf{X}_t - \mu_g)\right) \quad (3)$$

則根據最大似然率決策法(maximum likelihood decision)[4]， $\mathbf{X}$  應判斷為

$$S = \underset{k}{\operatorname{arg\,max}} P(\mathbf{X}|\lambda_k^{(T)}) \quad (4)$$

在實作上，考慮鳥鳴聲資料庫可能並無包含大量的聲音樣本可供精確的模型訓練，因此我們藉由模型調適技術[6]來產生個個鳥種的模型。該技術先根據EM演算法將所有訓練用之鳥鳴聲見立一個高斯混合模型當作通用模型(Universal Model)，再經由最大事後機率(Maximum A Posterior, MAP)調適法進行通用模型之調整，以產生各鳥種的高斯混合模型。

#### 4. 音高分析

音高參數與聲音的基頻(fundamental frequency, F0)有關，一連串基頻高低不同的聲音串在一起就如同不同音符被演奏出一般。本單元的基本概念是假設每種鳥都有其各自的歌聲或歌唱語言，像是音符高低相連有其獨特的規則。若我們能捕捉每種鳥的音符相連接資訊，則可據此識別未知的鳥鳴聲所屬鳥種類。如圖3所示為音高分析過程，主要包括音高特徵擷取及統計模型建立與匹配。

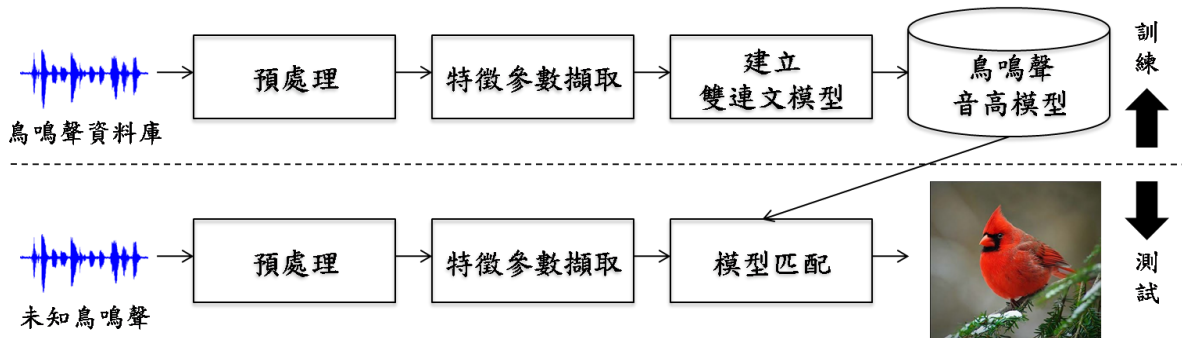


圖3 音高分析過程。

##### 4.1 特徵參數擷取

我們採用次諧波總和法(Sub-Harmonic Summation, SHS)[7]進行音高求取。SHS的原理是根據基頻除了本身的能量較高外，其倍頻諧波(Harmonic)的能量也通常較高，若某一基頻與其倍頻諧波的能量總和明顯較高於其他頻率時，該頻率極有可能為基頻。其求取音高流程如圖4。

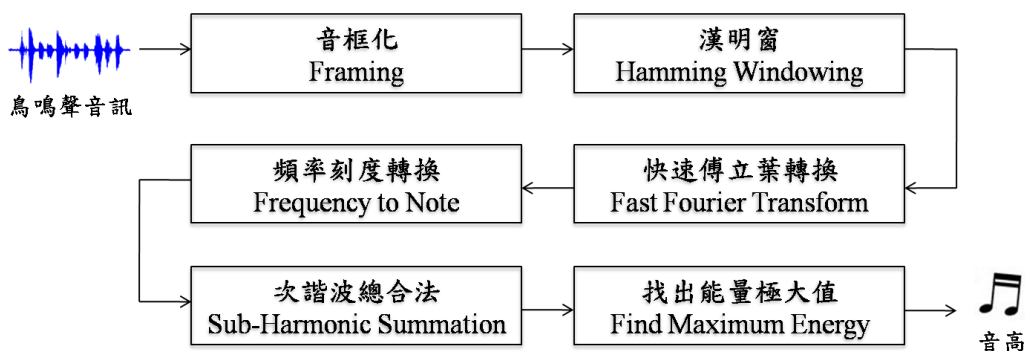


圖 4 音高求取流程圖

令  $x_{t,j}$  為第  $t$  個音框訊號在第  $j$  個 FFT 頻率刻度上的振幅，其中  $1 \leq j \leq J$ ，而  $J$  為頻率刻度總數。透過方程式 5 將 FFT 刻度轉為 MIDI 音符刻度  $e_1, e_2, \dots, e_N$

$$N(j) = 12 * \log_2 \left( \frac{freq(j)}{440} \right) + 69.5 \quad (5)$$

$$y_{t,n} = \max_{\forall j, N(j)=e_n} x_{t,j} \quad (6)$$

則我們可求取訊號在  $t$  時間屬於音符  $e_n$  的能量。理想上，若在時間  $t$  的鳴唱音符為  $e_n$ ，則屬於音符刻度  $e_n$  的能量  $y_{t,n}$  應是所有  $y_{t,1}, y_{t,2}, \dots, y_{t,N}$  中最大者，如(6)所示。但由於有諧波的存在，能量  $y_{t,n}$  可能並非最大。因此為了避免誤判，可利用 SHS 的觀念檢查各音符及其若干個八度音符的加權能量和 (Strength)：

$$z_{t,n} = \sum_{c=0}^C h^c y_{t, n+12c} \quad (7)$$

其中  $C$  是欲列入考慮的八度音符數，而  $h$  為小於 1 的權重。據此判定演唱音符應為

$$o_t = \arg \max_{1 \leq n \leq N} z_{t,n} \quad (8)$$

## 4.2 鳥鳴聲之模型建立

為了捕捉音符間相連的動態資訊，我們採用雙連文模型 (Bi-gram Model) 技術來訓練屬於鳥鳴聲的模型。

當給定一串音符序列  $o_1, o_2, \dots, o_t, \dots$ ，雙連文模型可用以描述該序列中各音符前後相連接的關係，其做法是統計所有可能之兩音符組合，例如  $w_1$  與  $w_2$  的發生頻率或機率： $P(o_t = w_1 | o_{t-1} = w_2)$ 。若兩個音符  $w_1$  與  $w_2$  常出現於某一種鳥的鳴聲中，則機率  $P(o_t = w_1 | o_{t-1} = w_2)$  值會較大；反之，若兩個音符幾乎不會出現於某一種鳥的鳴聲中，則機率  $P(o_t = w_1 | o_{t-1} = w_2)$  值會較小。

在測試階段，若有一未知鳥鳴聲片段之音符序列為  $\mathbf{O} = \mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_T$ ，則我們對各模型  $\lambda_k^{(P)}$ ， $1 \leq k \leq K$  分別計算似然率：

$$P(\mathbf{O}|\lambda_k^{(P)}) = \prod_{t=1}^{t=T} P(O_t|\lambda_k^{(P)}) \quad (9)$$

則根據最大似然率決策法 (maximum likelihood decision) [4]， $\mathbf{O}$  應判斷為

$$S = \arg \max_k P(\mathbf{O}|\lambda_k^{(P)}) \quad (10)$$

## 5. 結合音色與音高分析之辨識系統

經實驗發現，使用音色分析與音高分析所獲得的鳥鳴聲辨識結果有許多差異與互補之處，因此嘗試結合音色與音高之辨識系統，如圖 6 所示。

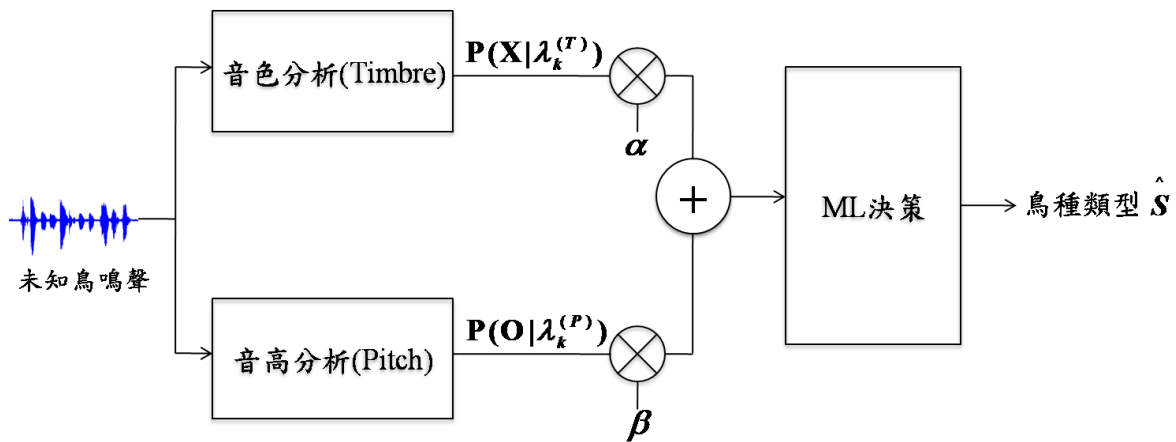


圖 6 整合音色與音高之辨識系統架構圖。

將一未知鳥種的鳴聲訊號，經由音色分析與音高分析後產生兩似然率，再將這兩似然率做加權總和，最後挑選加總後的最大似然率。據此判斷該鳥鳴聲為何種鳥，即：

$$\hat{S} = \arg \max_{1 \leq k \leq K} [\alpha \cdot P(\mathbf{X}|\lambda_k^{(T)}) + \beta \cdot P(\mathbf{O}|\lambda_k^{(P)})] \quad (11)$$

在本篇論文中， $\alpha$  與  $\beta$  分別設定成 0.6 與 0.4。

## 6. 鳥鳴聲辨識系統

### 6.1 鳥鳴聲資料庫

本論文所使用的鳥鳴聲音檔有兩個來源，分別是市面上販售的商業 CD 及網路上收集來的鳥鳴聲音檔，整理如表 1、2 所示：

表 1 商業 CD 之來源。

專輯名稱	出版商
鳥-野鳥鳴唱圖鑑	風潮有聲出版有限公司
台北鳥視界	台北市政府新聞處

八仙山國家森林遊樂區常見鳥類鳴聲	行政院農業委員會林務局
大雪山國家森林遊樂區常見鳥類鳴聲	行政院農業委員會林務局

表 2 網路上收集鳥鳴聲之來源。

網站名稱	網站位置
台灣大學動物博物館	<a href="http://archive.zo.ntu.edu.tw/">http://archive.zo.ntu.edu.tw/</a>
國立鳳凰谷鳥園	<a href="http://www.fhk.gov.tw/">http://www.fhk.gov.tw/</a>
Bird Call Recordings	<a href="http://www.geocities.com/RainForest/9003/birdcall.htm">http://www.geocities.com/RainForest/9003/birdcall.htm</a>
Macaulay Library	<a href="http://macaulaylibrary.org/index.do">http://macaulaylibrary.org/index.do</a>

這些商業 CD 和網站包含各種不同鳥類的鳴聲，在考慮到訓練語料量的多寡、地域的合理性、科目差異和鳴聲類型上，本篇論文共挑選了 10 種大台北地區常見的鳥種來進行實驗，分別為小卷尾、小啄木、小彎嘴畫眉、山紅頭、五色鳥、白耳畫眉、紅嘴黑鶉、紫嘯鶉、黃嘴角鴉、樹鵲。

一我們將商業 CD 音檔及網路上音檔兩種來源合併後各取一半，使得訓練音檔與測試音檔之檔案數不會差異太大。

使用商業 CD 上的音檔，原始音檔格式一律為雙聲道 44.1KHz 的 PCM WAV 檔。至於網路上收集而來的音檔，原始格式有些是經過 MP3 壓縮過的音檔，也有 PCM 格式的 WAV 檔，而取樣頻率從 8KHz 至 48KHz 不等。為了實驗的一致性及節省運算時間，我們把兩者來源的音檔皆調整為 22.05KHz 單聲道的 PCM WAV 檔。

## 6.2 實驗結果

### 6.2.1 基於音色分析之鳥聲辨識結果

首先進行不同高斯混合數的鳥鳴聲辨識實驗，結果列於表 3。從表中得知，當混合數為 64 時，其辨識率優於其他混合數。因此在本篇論文中，GMM 之混合數皆為 64。其辨識結果如表 4，總辨識率為 71.08%，其辨識率之計算方式如(12)。若僅觀察表中鳥鳴聲，僅有小彎嘴畫眉的辨識率較高，其餘的鑑別度都頗低。

$$\text{辨識率正確率} = \frac{\text{被正確辨識出的音檔數目}}{\text{總測試音檔數目}} \times 100\% \quad (12)$$

表 3 不同高斯混合數之辨識率。

高斯混合數	4	8	16	32	64	128
小卷尾	55.84%	58.44%	58.44%	59.74%	64.93%	62.33%
小啄木	59.8%	59.8%	60.78%	62.74%	62.74%	62.74%
小彎嘴	80%	81.93%	83.22%	83.22%	82.58%	81.93%
山紅頭	69.13%	69.13%	70.37%	71.6%	70.37%	69.13%
五色鳥	72.14%	73.05%	73.97%	74.42%	74.88%	74.88%
白耳畫眉	74.41%	75.58%	77.9%	77.9%	76.744%	76.74%
紅嘴黑鶉	62.42%	63.69%	64.96%	66.24%	68.78%	67.51%
紫嘯鶉	68%	72%	76%	76%	76%	76%
黃嘴角鴉	63.96%	63.96%	63.96%	63.96%	67.56%	65.76%
樹鵲	50.68%	52.05%	56.16%	57.53%	56.16%	54.79%
總辨識率	67.12%	68.23%	69.52%	70.25%	71.08%	70.25%

表 4 基於音色之鳥鳴聲辨識混淆矩陣。

	小卷尾	小啄木	小彎嘴	山紅頭	五色鳥	白耳畫眉	紅嘴黑鶇	紫嘯鶇	黃嘴角鴉	樹鵲
小卷尾	64.93%	10.38%	0%	9.09%	0%	6.49%	6.49%	0%	2.59%	0%
小啄木	9.8%	62.74%	14.7%	2.94%	3.92%	2.94%	2.94%	0%	0%	0%
小彎嘴	0%	6.45%	82.58%	0%	0%	0%	0%	6.45%	0%	4.51%
山紅頭	7.4%	3.7%	6.17%	70.37%	6.17%	3.7%	0%	2.46%	0%	0%
五色鳥	0%	0.91%	0%	0%	74.88%	0%	13.69%	0%	6.84%	3.65%
白耳畫眉	3.48%	0%	8.13%	0%	0%	76.74%	5.81%	5.81%	0%	0%
紅嘴黑鶇	0%	6.39%	0%	4.45%	8.28%	0%	68.78%	2.54%	15.92%	0%
紫嘯鶇	0%	0%	16%	0%	0%	8%	0%	76%	0%	0%
黃嘴角鴉	2.7%	13.51%	0%	9%	0%	7.2%	0%	0%	67.56%	0%
樹鵲	2.73%	0%	10.95%	0%	0%	0%	21.91%	0%	8.21%	56.16%

### 6.2.2 基於音高方法分析之鳥聲辨識結果

觀察現有鳥鳴聲資料庫，其基頻範圍介於 366 Hz 至 8591 Hz，對應音符為 66-120。基於音高分析之辨識結果如表 5，總辨識率為 72.09%。若僅觀察表中鳥鳴聲，僅有小彎嘴畫眉及五色鳥的辨識率高於 80%，其餘的鑑別度都頗低。

表 5 基於音高之鳥鳴聲辨識混淆矩陣。

	小卷尾	小啄木	小彎嘴	山紅頭	五色鳥	白耳畫眉	紅嘴黑鶇	紫嘯鶇	黃嘴角鴉	樹鵲
小卷尾	61.03%	19.48%	0%	10.38%	0%	5.19%	3.89%	0%	0%	0%
小啄木	2.94%	71.56%	12.74%	0%	1.96%	0%	3.92%	0%	2.94%	3.92%
小彎嘴	0%	7.74%	82.58%	0%	0%	0%	0%	7.74%	0%	1.93%
山紅頭	1.23%	0%	7.4%	75.3%	2.46%	1.23%	0%	0%	0%	0%
五色鳥	0%	1.36%	0%	1.82%	82.19%	0%	11.41%	0%	2.73%	0.45%
白耳畫眉	0%	0%	11.62%	0%	0%	76.74%	5.81%	5.81%	0%	0%
紅嘴黑鶇	0%	6.36%	0%	2.54%	9.55%	0%	63.05%	2.54%	15.92%	0%
紫嘯鶇	0%	0%	4%	0%	12%	28%	0%	56%	0%	0%
黃嘴角鴉	7.2%	21.62%	5.4%	0%	0%	0.09%	0%	0%	64.86%	0%
樹鵲	5.4%	0%	8.21%	0%	0%	0%	24.65%	0%	2.73%	58.9%

### 6.2.3 將音色及音高方法結合後之鳥鳴聲辨識結果

如第 5 章介紹，我們將音色及音高兩種方法結合。在(11)式中，我們將 GMM 計算後的似然率  $P(\mathbf{X}|\lambda_k^{(T)})$  及 Bigram 計算出的  $P(\mathbf{O}|\lambda_k^{(P)})$  分別乘上  $\alpha$  與  $\beta$  之權重，在本篇論文裡， $\alpha=0.6$ 、 $\beta=0.4$ 。

觀察表 6，結合後之鳴聲總辨識率達 75.04%。觀察表中鳥鳴聲之辨識率，小彎嘴畫眉、五色鳥、白耳畫眉及紫嘯鶇的辨識率皆高於 80%，且最低的辨識率也高於 60%。

表 6 整合音色、音高方法後之鳥鳴聲辨識混淆矩陣。

	小卷尾	小啄木	小彎嘴	山紅頭	五色鳥	白耳畫眉	紅嘴黑鶇	紫嘯鶇	黃嘴角鴉	樹鵲
小卷尾	67.53%	12.98%	0%	9.09%	0%	5.19%	2.59%	0%	2.59%	0%
小啄木	2.94%	75.49%	9.8%	0%	0.09%	0%	3.92%	0%	2.94%	3.92%
小彎嘴	0%	5.8%	85.16%	0%	0%	0%	0%	5.8%	0%	3.2%
山紅頭	1.23%	0%	6.17%	75.3%	2.46%	1.23%	0%	1.23%	0%	0%
五色鳥	0%	1.36%	0%	1.82%	83.1%	0%	9.13%	0%	3.19%	1.36%
白耳畫眉	0%	0%	10.46%	0%	0%	80.23%	4.65%	3.48%	0%	0%



紅嘴黑鵝	0%	6.36%	0%	1.27%	9.55%	0%	65.6%	1.27%	15.92%	0%
紫嘯鵝	0%	0%	4%	0%	4%	12%	0%	80%	0%	0%
黃嘴角鵝	7.2%	21.62%	5.4%	0%	0%	1.8%	0%	0%	63.96%	0%
樹鵝	4.1%	0%	5.47%	0%	0%	0%	21.91%	0%	2.73%	65.75%

## 7. 結論與未來展望

本論文提出結合音色與音高線索來辨識鳥類鳴聲。實驗發現，當我們單獨用音色或音高線索來辨識十種鳥的鳴聲時，僅有 1 至 2 種鳥的鳴聲辨識率高於 80%；但若將兩種特徵結合後，將有 4 種鳥辨識率超過 80%，驗證結合使用音色及音高之方法能提升鳥鳴聲辨識效能。

由於受限於鳥鳴聲音檔的收集數量有限，所以本篇論文只挑選大台北地區常見的十種鳥種，對於系統的強健性和實用性還需要加入更多鳥種來驗證。相較於目前人聲辨識的成果來說，鳥鳴聲辨識系統還有極大的改進空間。

## 參考文獻

- [1] 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.
- [2] 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.
- [3] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. on Acoustic, Speech and Signal Processing.*, vol. 28, no. 4, pp. 357–366, Aug. 1980.
- [4] X. Huang, A. Acero, and H. W. Hon, *Spoken Language Processing*, Prentice Hall, 2001.
- [5] A. Ramalingam and S. Krishnan, "Gaussian Mixture Modeling of Short-Time Fourier Transform Features for Audio Fingerprinting," *IEEE Transactions on Information Forensics and Security*, 2006.
- [6] D. Reynolds and T. Quatieri, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing 10*, PP. 19-41, 2000.
- [7] M. Piszczalski and B. A. Galler, "Predicting musical pitch from component frequency ratios", *Journal of the Acoustical Society of America*, 66(3), pp.710—720, 1979.