

用動態時間校正做特定種類 鳥鳴偵測的方法

An efficient approach to birdsong detection with
Dynamic Time Warping

TAA2019 何佩蓁

錄音技術應用於野生動物調查



▲圖1、林務局森林資源調查隊使用之長時間錄音調查設備與架設方式

姜博仁, 蔡哲民, 蔡世超, 吳禎祺, 鄭蕙如 (2015)
錄音技術應用於野生動物調查之應用與評估, 臺灣林業 41:4 2015.08

取樣監聽

- 1:1 全時監聽不符成本效益
- 日出後與日落後各 15分鐘加上日出後一小時開始每小時取樣2分鐘, 約5~6%的音檔可記錄到 73.5%(春夏)和 55.2%(秋冬)的物種。
- 夜間頻譜全時觀察, 增加監聽量至 7%, 物種涵蓋率可提升至80.7%和70.8%。
- 鳥類多日行性, 夜間頻譜全時觀察監聽, 對鳥類涵蓋率提升較少。

台灣藍鵲

為臺灣特有種鳥類，主要分布在海拔 1,000公尺以下的地區，而 1,800公尺以下的中、低海拔闊葉林或次生林也可以看見牠的蹤跡。臺灣藍鵲性情兇悍，領域性強，會攻擊入侵者，也會集體禦敵。中華民國行政院農業委員會則將臺灣藍鵲列為 **保育類動物**。

維基百科，自由的百科全書

保育類台灣藍鵲出沒北市 民眾嫌煩「BB彈攻擊」恐觸法

2017/05/04 中央社

台灣藍鵲寶寶蹲「粉紅飼養箱」 羅東夜市擺攤賣遭移送

2018/08/03 ETToday

台灣藍鵲現蹤北市 8幼鳥遭獵捕幸追回

2017/06/14 中時

台灣藍鵲



保護狀況



科學分類

界：動物界 Animalia

門：脊索動物門 Chordata

綱：鳥綱 Aves

目：雀形目 Passeriformes

科：鴉科 Corvidae

屬：藍鵲屬 *Urocissa*

種：台灣藍鵲 *U. caerulea*

維基百科，自由的百科全書

鳥鳴偵測方法

特徵向量抽取

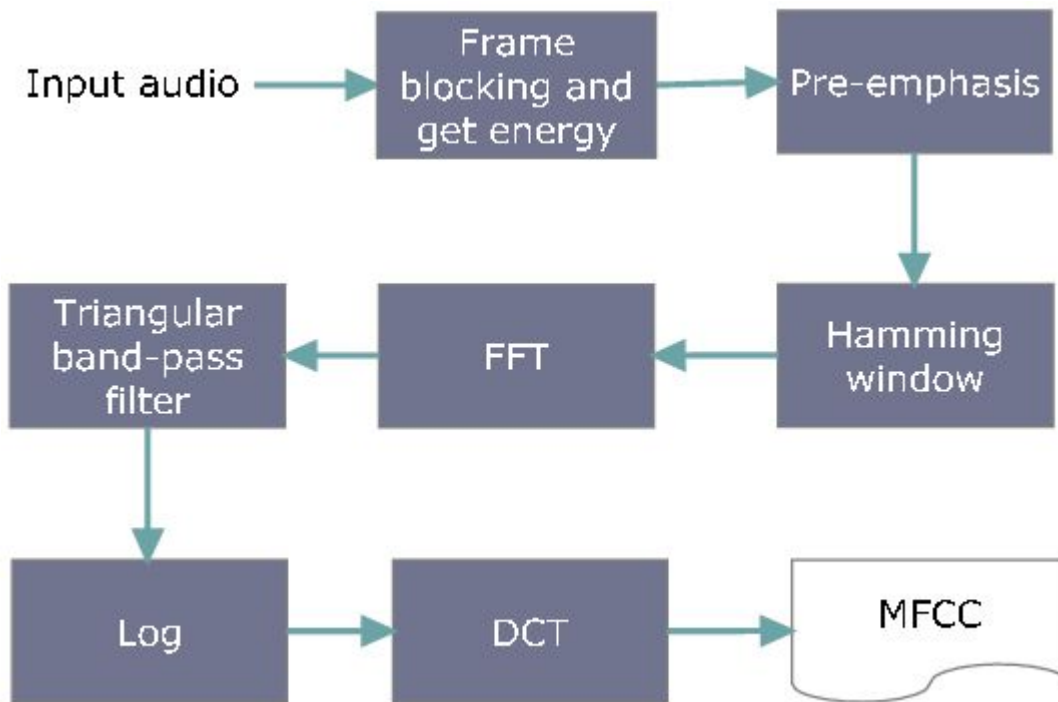
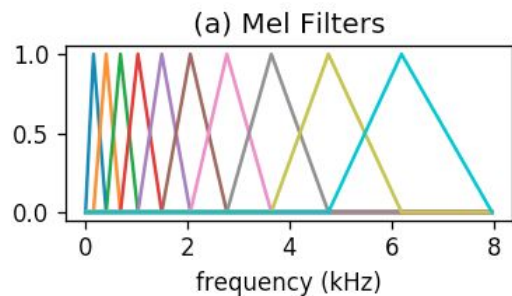


動態時間校正

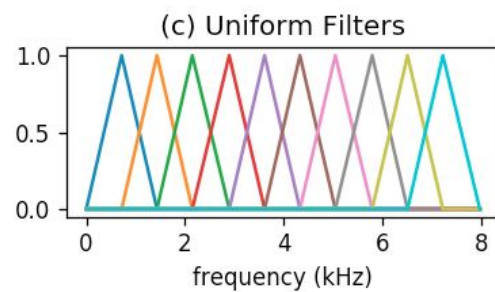
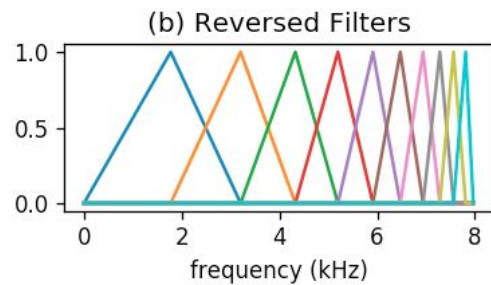
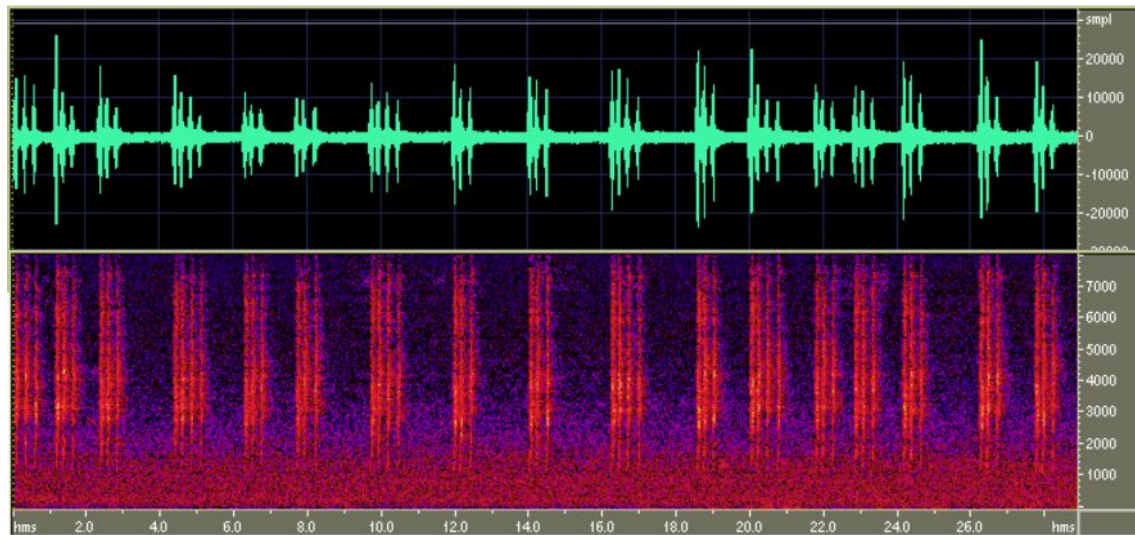


鳥鳴時點偵測

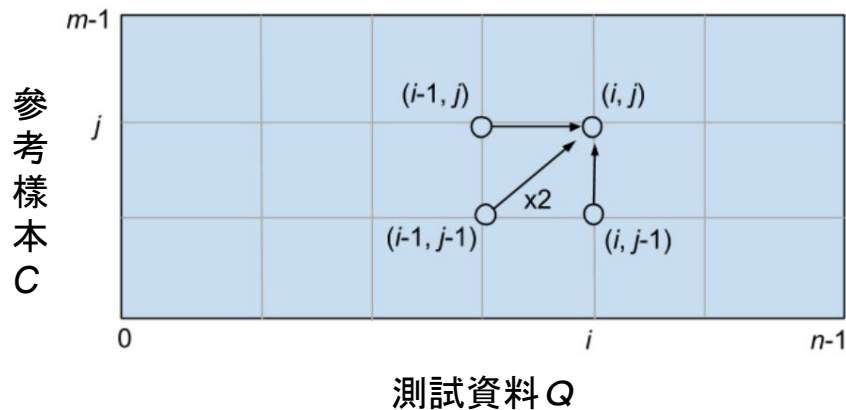
梅爾濾波器和倒頻譜係數



台灣藍鵲頻譜



動態時間校正(Dynamic Time Warping, DTW)



$$a(i, j) = \min \begin{cases} a(i, j-1) + d(q_i, c_j) + penalty \\ a(i-1, j) + d(q_i, c_j) + penalty \\ a(i-1, j-1) + 2 \times d(q_i, c_j) \end{cases}$$

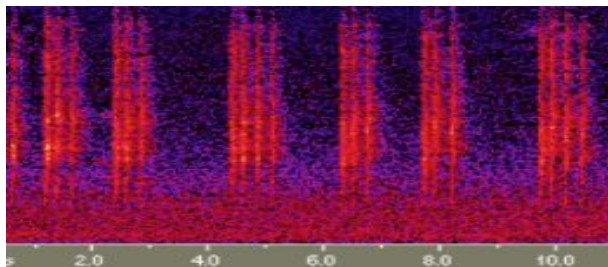
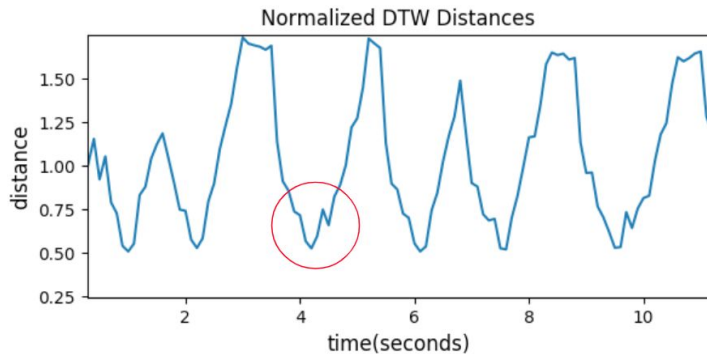
鳥鳴長度不一，DTW較不受長度不同而影響

$DTW(C, Q) \leftarrow a(n-1, m-1) / (m+n-1)$

從 (0,0) 到 (n-1, m-1) 所有路徑的最小距離

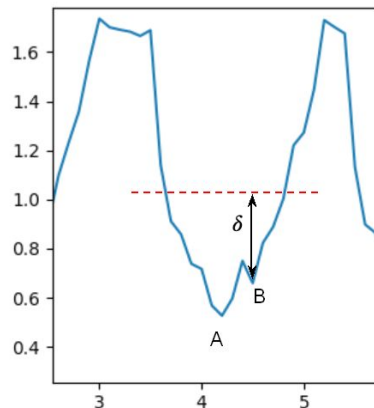
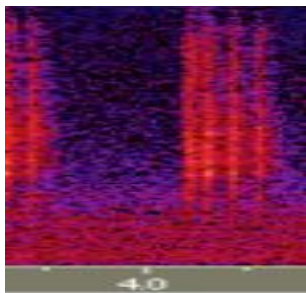
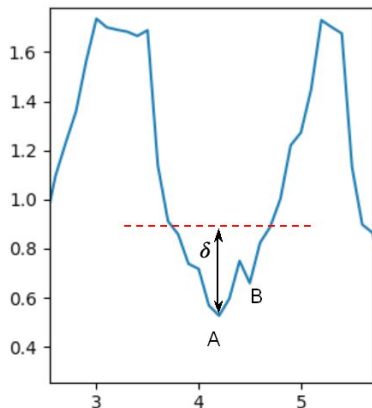
penalty: 使路徑傾向往對角線方向前進

鳥鳴時點偵測



```
m = len(C) # 參考樣本 C 的長度
n = 1.2 * m # 測試樣本Q稍長於 C
d = []
t = 0
while t < len(data):
    Q = data[t:t+n]
    distance = DTW(Q,C)
    d.append(distance)
    t += 0.1 second
plot(d)
```


鳥鳴時點偵測



檢測 A 或 B 是否為鳥鳴：

$threshold = local_min + \delta$

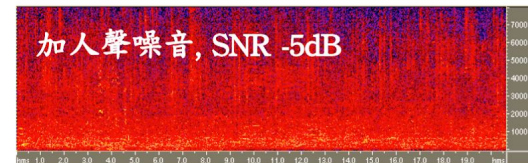
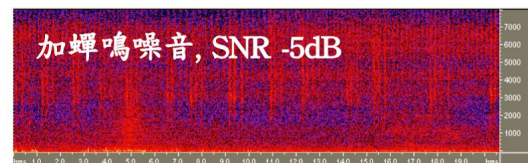
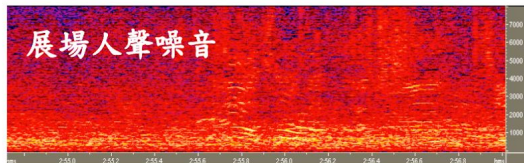
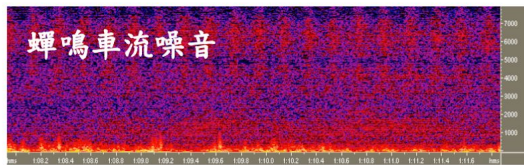
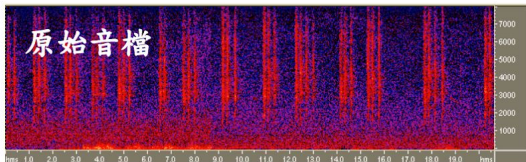
切下倒錐型曲線最低點才是鳥鳴發生的起始位置

結果：

B 非鳥鳴發生起始位置

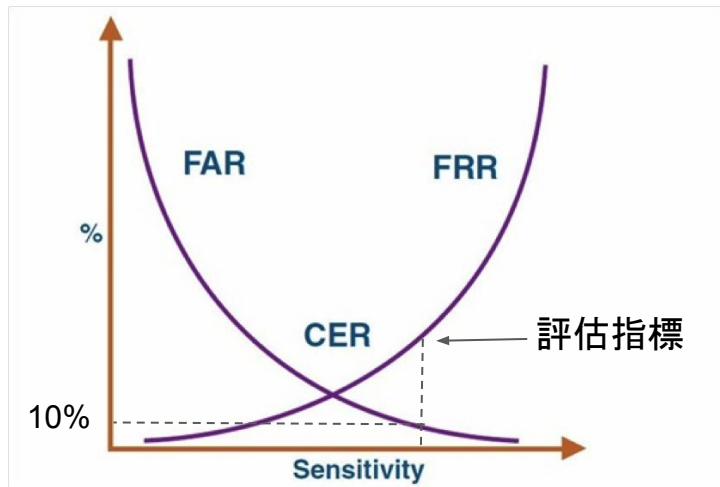
實驗資料

- 鳥鳴: xeno-canto (<https://www.xeno-canto.org>), 共 3 分 22 秒, 109 次鳥鳴, 每次持續約 0.5 至 1.2 秒
- 噪音: 1) 路旁行道樹蟬鳴加車流, 2) 世貿展場人聲噪音。



實驗條件

- 偵測到的鳥鳴發生時間與人工標註答案必須在 0.5 秒內才算正確。
- 誤接受率(FAR)須 10% 以下, 以誤拒絕率(FRR, 未偵測到的比率)為評估指標, 越低越好。



實驗1 - 時域特徵

蟬鳴加車噪	原始音檔	SNR 10dB	SNR 5dB	SNR 0dB	SNR -5dB
過零率	41.3%	55.0%	23.9%	17.4%	28.4%
短時能量	20.2%	40.4%	63.3%	85.3%	88.1%

人聲噪音	原始音檔	SNR 10dB	SNR 5dB	SNR 0dB	SNR -5dB
過零率	41.3%	52.3%	60.6%	57.8%	53.2%
短時能量	20.2%	30.3%	51.4%	86.2%	89.9%

過零率效果不甚穩定。

短時能量隨噪音增加偵測效果大幅降低。

整體而言，時域特徵在鳥鳴偵測上的表現不佳。

實驗2 - 頻域特徵在蟬鳴噪音環境的偵測效果

蟬鳴加車噪	原始音檔	SNR 10dB	SNR 5dB	SNR 0dB	SNR -5dB
梅爾濾波器	8.3%	12.8%	12.8%	26.6%	35.8%
集中高頻濾波器	11.0%	11.0%	11.9%	41.3%	38.5%
平均分佈濾波器	11.9%	10.1%	12.8%	19.3%	16.5%
梅爾倒頻譜	21.1%	25.7%	24.8%	37.6%	50.5%
倒頻譜－集中高頻	33.9%	26.6%	19.3%	16.5%	22.9%
倒頻譜－平均分佈	23.9%	24.8%	22.9%	26.6%	27.5%

梅爾濾波器對原始音檔效果最佳，錯誤率僅 8.3%。

整體而言在蟬鳴噪音下，濾波器組效果優於倒頻譜係數。

平均分佈濾波器較不受蟬鳴噪音所影響。

實驗3 - 頻域特徵在人聲噪音環境的偵測效果

人聲噪音	原始音檔	SNR 10dB	SNR 5dB	SNR 0dB	SNR -5dB
梅爾濾波器	8.3%	9.2%	16.5%	33.0%	63.3%
集中高頻濾波器	11.0%	11.0%	12.8%	18.3%	59.6%
平均分佈濾波器	11.9%	11.0%	11.9%	16.5%	64.2%
梅爾倒頻譜	21.1%	18.3%	14.7%	23.9%	36.7%
倒頻譜－集中高頻	33.9%	20.2%	17.4%	15.6%	23.9%
倒頻譜－平均分佈	23.9%	20.2%	23.9%	25.7%	41.3%

濾波器組在 SNR 5dB 以上的中、低度人聲噪音表現較佳。

在 -5dB 高強度人聲噪音下倒頻譜係數普遍表現較好。

集中高頻的濾波器所產生的倒頻譜係數最能適應 -5dB 高強度人聲噪音。

結論

- 本方法不論在安靜或噪音環境下，確實能有效地偵測台灣藍鵲鳥鳴。
- 實驗測試資料量不夠大，仍需更多大量資料、更多實驗來佐證本方法可靠性。
- 我們相信本方法是合理且有效的，繼續深入研究，應能協助生態保育監聽人員發揮更大的工作效率。