

## 魚類聲音萃取與台灣魚類聲音資料庫之架構

### The Extraction of Fish Sounds and the Development of Sound Archive for Fishes in Taiwan

魏瑞昌<sup>1</sup>、劉祐麟<sup>2</sup>

#### 摘要

台灣魚類聲音資料庫的建立，可以有系統儲存聲音樣本的詳細資訊，除此之外還可提供一個資訊交流平台，增加研究人員在生物行為、物種辨認以及定位追蹤等研究的便利性。在進行資料上傳之前，資料庫會計算聲音樣本的訊雜比作為品質分析結果，提供使用者選擇聲音樣本的一項參考。由於資料來源有實際海域錄音與水族箱錄音兩大類，為了正確框選出魚類聲音訊號，本文用兩種端點偵測法進行萃取，當聲音樣本是實際海域錄音時，採用時域端點法；當聲音樣本為水族箱錄音時，採用頻率端點法。最後配合偵測結果計算魚類訊號與背景值的平均功率，得到聲音樣本的訊雜比。資料庫採用三層式架構進行資料庫建立，為了降低資料的儲存容量同時保持資料完整性，以關聯模式作為資料儲存的格式。本資料庫除了提供資料上傳的平台外，還提供生物名稱、錄音地點、錄音時間三種搜尋方式，希望增加使用者使用資料庫的便利性與效率。

**關鍵字：**魚類聲音、資料庫、時域端點法、頻率端點法、三層式架構、關聯模式

#### Abstract

The goal of development of sound archive for marine fishes in Taiwan is not only to preserve data, but also wants to provide a platform of data sharing to increase the efficiency for the study of fish behavior, automatic recognition, localization, and tracking. In order to provide the sound quality in terms of signal-to-noise ratio to users, the fish sound recording will be analyzed before uploading. Because most available data were recorded either in the field or in fish tank, the fish sounds were extracted by using two different automatic detection methods. If fish sound recordings were from the field, the Time Endpoint Detection was applied. If the recording was made in the fish tank, the Frequency Endpoint Detection was applied. From two different automatic detection methods, the sound quality was presented in the signal-to-noise ratio, which was the averaged power of signal divided by averaged power of the background noise. The archive was a 3-Tier system and in order to reduce the storage size and maintain the integrity of data, the Relational Model was applied. To improve the users' efficiency the database provides not only data uploading, but also data searching according to creature name, recording area, and recording time.

**Keywords:** Fish Sounds, Database, Time Endpoint Detection, Frequency Endpoint Detection, 3-Tier System, Relational Model.

<sup>1</sup>中山大學海下科技暨應用海洋物理研究所教授

<sup>2</sup>中山大學海下科技暨應用海洋物理研究所研究生

## 一、緒論

台灣位於亞洲大陸棚東南邊緣，地屬於亞熱帶區域，海水年均溫約 20 度，因為周遭有黑潮、大陸沿岸流與南中國海暖流經過，帶來大量的營養鹽，使得台灣的魚類資源十分豐富，加上台灣南、北兩端與離島地區屬於珊瑚礁地形，提供了魚類良好的棲息空間，因此魚類生物多樣性極為豐富。由於生活空間有限，許多珊瑚礁魚類為了防禦領域、競爭食物、生殖及交配等目的，都會發出聲音進行溝通，根據 Myrberg 於 1981 年指出 [1]，全世界至少有超過 50 科的魚類會發聲，主要是利用魚鰾、魚鰭或骨骼等進行發音，其發聲機制可能有摩擦發聲、魚鰾發聲及呼吸發聲等，且所發出聲音都含有特殊意義，會導致其他魚隻在行為上的改變。

為了有效瞭解魚類聲音所包含的資訊，研究人員常用被動式聲納，進行魚類聲音收集與資料上的分析，藉此瞭解魚類聲音與行為的相關性，其方法主要是用水下麥克風收集海洋生物的聲音後，藉由訊號處理來瞭解聲音的特徵與背後的資訊，可應用可用於定位、聲音行為分析、以及生物辨識等研究上。由於資料十分珍貴需要加以保存，為此本文應用資料庫理論進行魚類聲音資料庫的建立，此舉除了可以保存魚類聲音的多樣性外，還可提供聲音檔案的詳細資訊給使用者，增加研究人員對聲音樣本的掌握度，拓展資料使用的範圍。

聲音資料庫可以提供豐富的資訊給使用者，包含錄音人員、儀器、地點、生物名稱與聲音樣本品質等，而其中聲音樣本品質與當時錄音環境有關，能呈現生物聲音在聲音樣本中的清晰度，因此在將聲音資料輸入資料庫之前，會先對聲音樣本進行品質分析處理，應用自動化聲音萃取方法，標示出樣本中生物聲音的時間區段，再以訊雜比計算聲音樣本的品質，作為品質分析的依據，最後配合其他資訊一起輸入資料庫中儲存。

## 二、聲音品質分析

聲音品質好壞是聲音資料庫中的能提供的重要資訊之一，可用來作為使用者選擇資料的參考，當魚類聲音較小，背景雜訊較大時，其聲音樣本品質較差；反之魚類聲音越大，背景雜訊越小，其聲音樣本品質較佳。由於生物聲音屬於間歇性訊號，發聲時間無法確定，若用人工方法(例如聽取聲音樣本、或觀察時頻譜圖等方式)尋找生物聲音的時間區段，過程十分耗時，為減少人力資源上的消耗，本文以自動化萃取方法進行魚類聲音萃取，其自動化方法有兩大類：一類為依據門檻值(Threshold)來框選出聲音訊號的起始點與終點，例如端點偵測法 [2] 等；一類是用模式匹配 (Pattern-Matching) 萃取訊號，例如隱藏馬可夫 (Hidden Markov Model) 模型法萃取法 [3] 等。

本研究所使用的方法為端點偵測法，首先應用音框切割聲音樣本，接下來再對每個音框進行時域上或是頻率上的處理，最後依次比較每個音框與門檻值的大小，決定該音框是否為生物訊號。當該音框被判定為生物訊號時，框選出該音框的起始點與終點，表示為生物訊號的發聲時間與結束時間所使用的端點偵測法有二種，一種是時域端點偵測法 (Time Endpoint Detection)；一種為頻率端點偵測法 (Frequency Endpoint Detection)，其介紹如下。

### 2.1 時域端點偵測法

時域端點偵測法是藉由音框長度與重疊率的設定，呈現聲音樣本的能量變化趨勢，再透過門檻的設定進行比較，當能量值高於門檻值便被判斷為生物訊號；反之則被判斷為雜訊。首先將聲音樣本通過帶通濾波器，降低雜訊的影響，再應用音框切割聲音樣本，依式(1.1)計算每個音框下的振幅平方和，表示該音框的總能量大小，接下

來設定門檻值比較聲音能量大小，最後框選出訊號的起始與結束時間。

$$E_i = \sum_{n=1}^N S_n^2 \quad (1.1)$$

式中  $E_i$  為第  $i$  個音框的能量和， $S_n$  為第  $i$  個音框下的第  $n$  個訊號(振幅大小)

在門檻值設定中，本文使用的門檻值為聲音樣本的中位數加上一個標準差 [4]，其式如下：

$$E_{Threshold} = E_{Median} + E_{\sigma} \quad (1.2)$$

式中  $E_{Threshold}$  表示聲音樣本的門檻值

$E_{Median}$  為聲音樣本的中位數

$E_{\sigma}$  為聲音樣本的標準差

注意，本研究門檻值的參數是中位數而非平均值，主要原因平均值較容易受到高能量的生物訊號影響，當生物訊號出現頻繁時，平均值會被拉高，造成門檻值過高，導致生物聲音萃取個數變少。而中位數因為位於聲音樣本能量分佈的中間，較不受高能量生物訊號影響，因能表示環境噪音的分布趨勢。

### 2.1.2 時域端點偵測法之範例

由於目的是框選生物聲音的時間區段，且目前處理的魚類叫聲長度都不過 0.5 秒，因而將音框長度設定為 0.5 秒、重疊率 50%，在此以河口錄音而得的日本銀身為例，呈現時域端點法的偵測過程與萃取結果。首先通過帶通濾波濾除雜訊，圖 1.1(A) 為聲音樣本通過帶通濾波器(0.3~3 kHz)後，4~8 秒的放大波形圖，發現在時間 4.5 秒與 7.5 秒各有一筆魚類叫聲，圖 1.1(B)~圖 1.1(F) 為音框長度 0.5 秒、重疊率 50% 的音框切割示意圖；圖 1.1(G) 的紅點為式(1.1)所計算出的音框總能量，由於生物聲音的能量明顯高於環境背景雜訊能量，因此配合門檻值篩選後，可以將時間區段 4.5 秒與 7.5 秒的魚類聲音框選出來，如圖 1.1(G) 虛線方框的所在即為魚類聲音的時間區段。

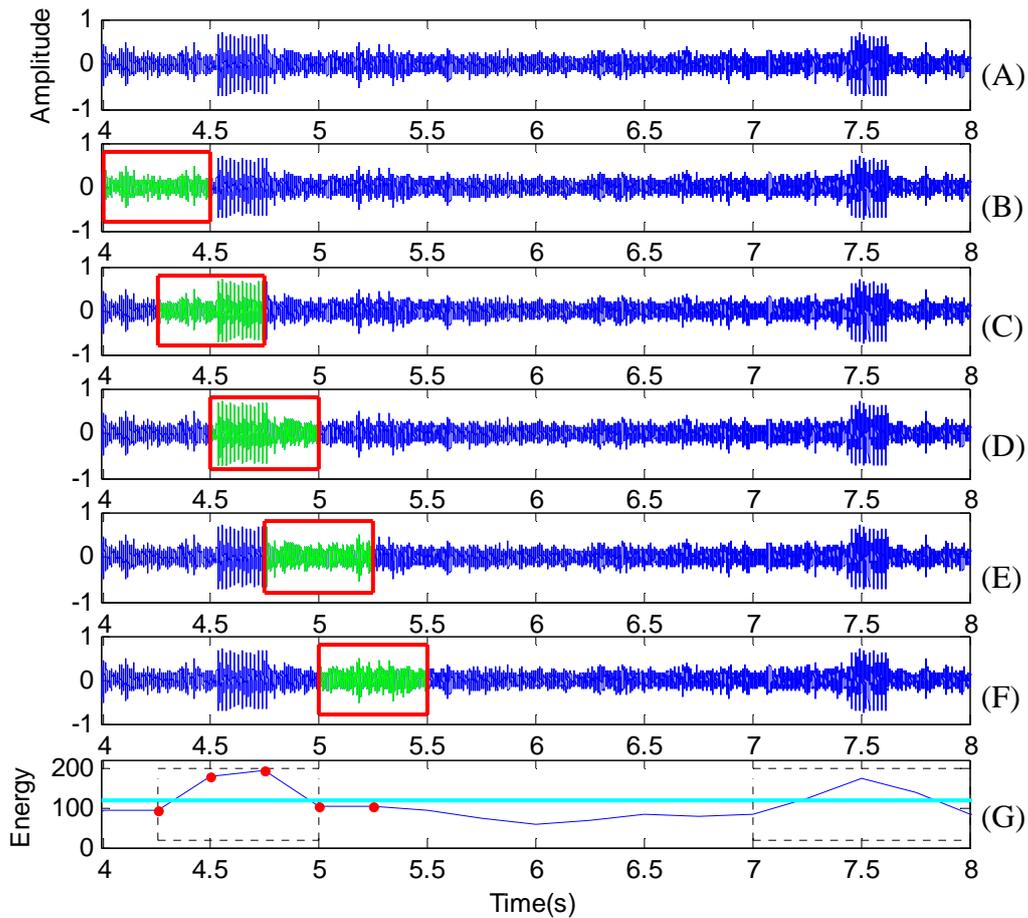


圖 1.1 日本銀身 之時域端點偵測法結果，(A)原始訊號；(B)~(F)音框切割示意圖；(G)音框能量分布與訊號萃取

## 2.2 頻率端點偵測法

頻率端點偵測法為時域端點法的進一步處理，是從頻率的角度分析聲音樣本，應用信息熵(Shannon's Entropy)探討不同音框下，因頻率組成的能量比例不同時所產生的變化。所謂的信息熵是機率的函數，其特色是在事件空間(Sample Space)中機率分布越平均時，則不確定越大；反之，當機率分布越集中在某些事件(Event)上時，不確定就越小，計算式如下：

$$H = -\sum_{i=1}^N p_i \times \log p_i \quad i = 1 \dots N \quad (1.3)$$

式中  $H$  為信息熵代表不確定性； $p_i$  為事件  $i$  的機率分布，因為  $p_i$  恆小於 1，再經過  $\log$  轉換後其值為負，因此需要取負號使信息熵變成正值。

過程是先利用音框切割聲音樣本後，將每個音框的聲音訊號進行頻譜轉換後，計算各個音框下各頻率所佔的能量比例，再用式(1.3)依次計算每個音框的信息熵，最後可得整體聲音樣本的信息熵變化。

### 2.2.1 頻率端點偵測法之範例

本文水族箱錄音所得之短吻鰻聲音樣本為例，計算聲音樣本的信息熵大小結果，圖 1.2(A)為放大聲音樣本 5~8 秒位置的示意圖，經由觀察得知在時間 7 秒附近有一筆魚類叫聲，圖 1.2(B)~圖 1.2(E)為長度 0.5 秒、重疊率 50% 的音框切割示意圖；為求詳細瞭解信息熵的計算過程，在此以單一音框圖 1.2(D)為範本，說明每個音框的信息

熵計算過程。

首先將此音框進行頻譜轉換如圖 1.2(F)，再計算該音框的各頻率所佔能量比例如圖 1.2(G)，應用式(1.3)得到該音框的信息熵大小，重複上述步驟計算每個音框的聲音訊號後可得整體聲音樣本的信息熵。觀察其分佈趨勢可知，當一個音框的組成頻率能量比例越趨向單調的話(音框中僅存在背景值噪音時)，其信息熵值就會越小，反之當一個音框的組成頻率能量比例越複雜(音框中包含魚類訊號時)，其信息熵值就會越大，最後配合門檻值設定即可萃取訊號，如圖 1.2(H)。

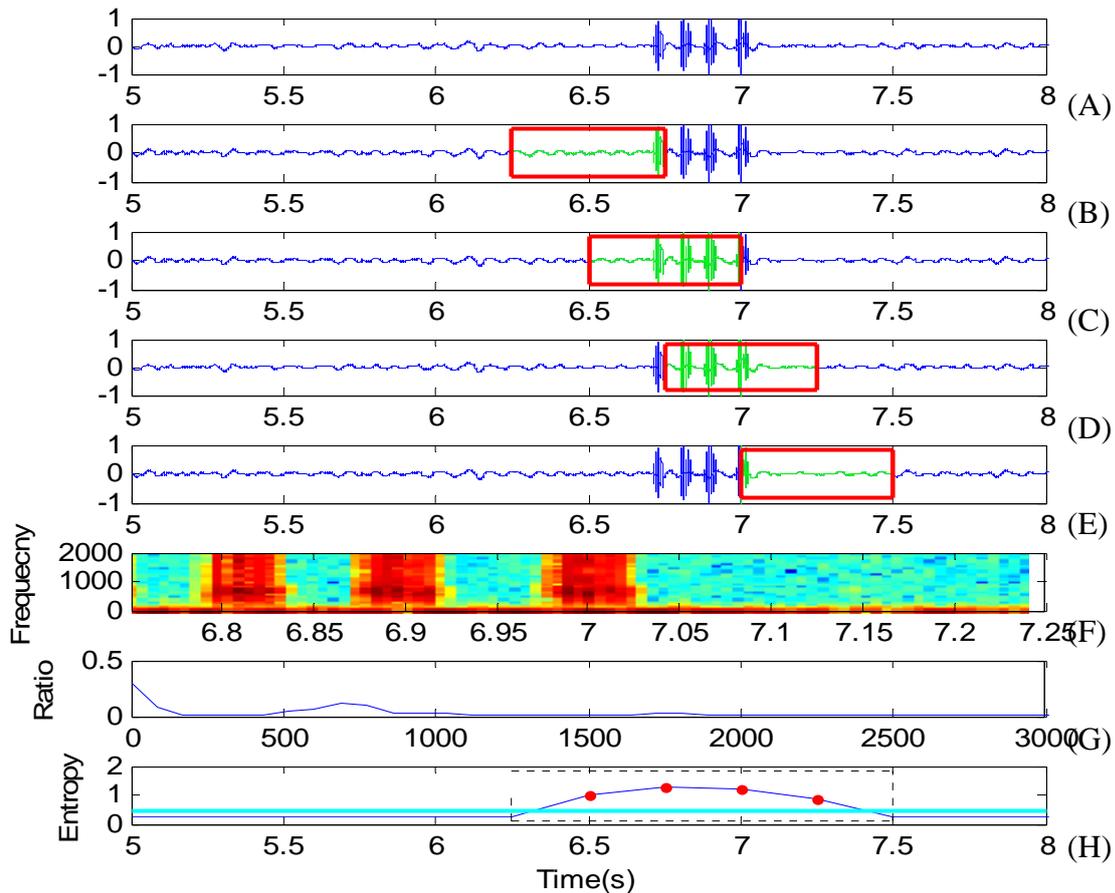


圖 1.2 短吻鰩之頻率端點偵測法的信息熵，(A)聲音樣本；(B)~(E)音框切割示意圖；(F)頻譜轉換；(G)各頻率比例；(H)信息熵變化與聲音萃取

### 2.3 聲音品質分析

本文的資料來源有水族箱錄音所得之聲音樣本與實際海域錄音之聲音樣本，水族箱聲音樣本來源有魚類聲音與突發性的雜訊(例如儀器碰撞聲)；實際海域的聲音來源除了上述兩種之外，還多了海洋環境噪音。由於兩者的聲音樣本組成不同，難以歸類出相同的特性，因此會依據資料來源的不同，選擇適合的端點偵測法。例如在水族箱錄音的聲音樣本，採用頻率端點偵測法；在實際海域的聲音樣本則採用時域端點偵測法，進行魚類聲音萃取。

而在進行訊雜比計算時，會依據聲音樣本來源選擇適當的端點偵測法，得到萃取出魚類聲音訊號時間區段，再將聲音樣本進行頻譜轉換並配合萃取結果，框選生物聲音的時間區段，最後分別計算生物訊號與背景噪音的平均功率後，代入訊雜比計算公式，可得到聲音樣本的頻值分析結果，如圖 1.3。

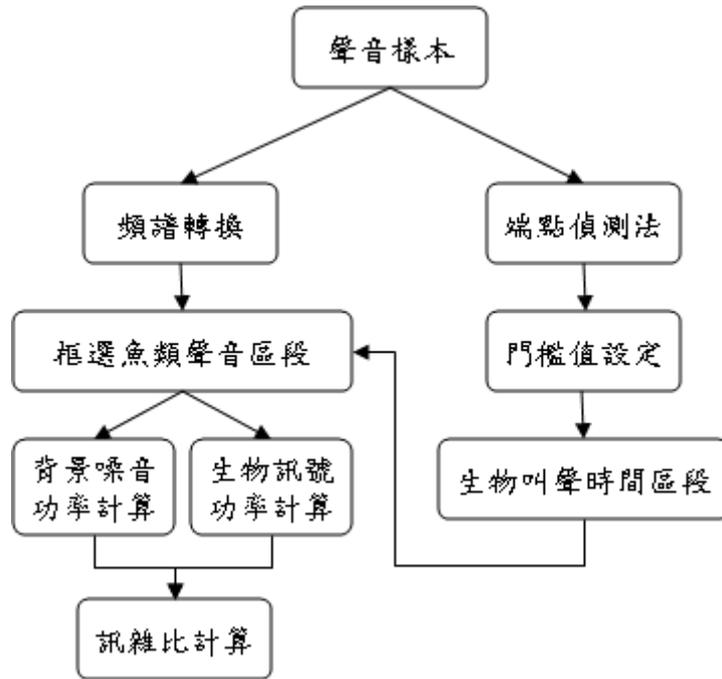


圖 1.3 訊雜比計算流程圖

### 三、資料庫結構

資料庫的建立有助於資料的儲存，本文採用三層式架構如圖1.4，將資料庫系統分成應用系統與管理系統，且將應用系統置於應用伺服器上，並依據使用者權限不同，以Hypertext Preprocessor( PHP)和SQL語言，撰寫資料上傳與搜尋的程式；將資料庫管理系統置於資料伺服器，依據來自資料庫應用系統的指令，進行資料庫的資料擷取或是資料新增與異動，以關聯模式進行資料庫綱目建立，使用的軟體為MySQL。

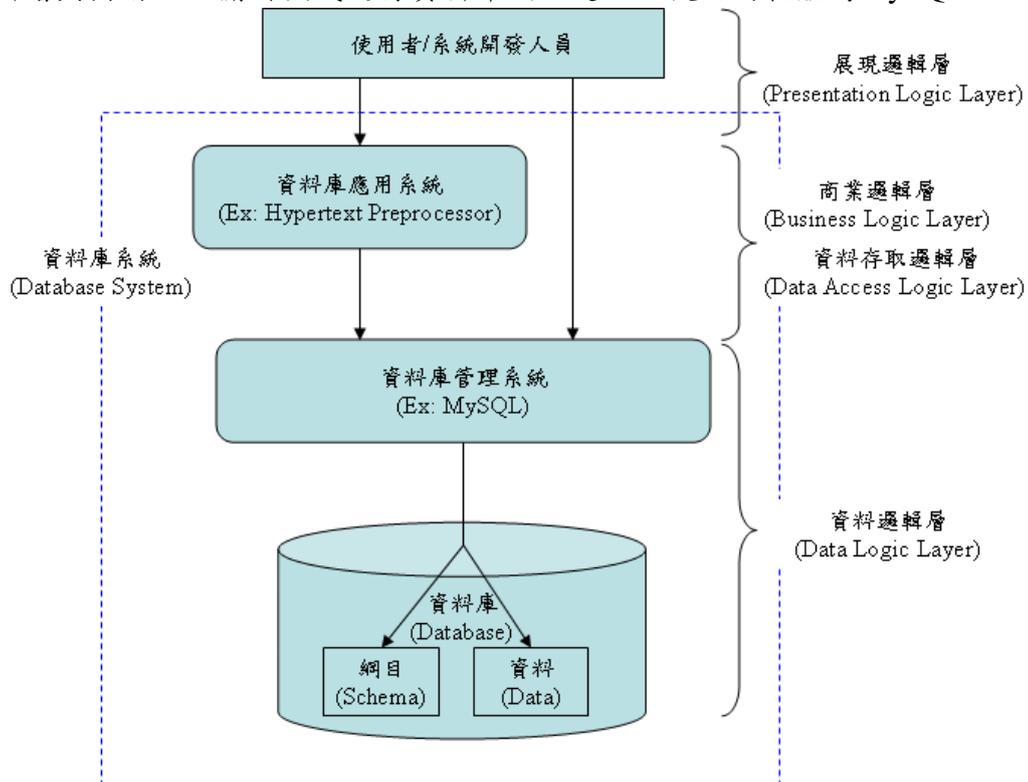


圖1.4 魚類聲音資料庫的架構[5]

本資料庫提供的聲音樣本極為詳細，依據資料間的相關性可將資料分成六大部份：錄音人員、儀器、地點、生物、聲音樣本、其他，所包含的資訊，如表1.1。由於資料十分複雜，為確保資料在新增異動時，不會發生錯誤，本研究應用關聯模式進行資料庫的建立。

表1.1 魚類聲音樣本提供之資訊

	細節介紹
錄音人員	姓名、電子信箱、電話
儀器	麥克風的型號與製造商、放大器的型號與製造商、錄音機的型號與製造商
地點	地點名稱、經緯度、錄音深度、海況、溫度、環境描述
生物	生物學名、生物俗名、生物體長
聲音樣本	儲存位置、錄音長度、錄音時間、樣本品質、增益值大小、錄音方法、時間序列圖、頻譜圖、聲音播放
其他	備註

### 3.1 關聯模式介紹

本資料庫管理系統使用的資料模式是關聯模式(Relational Model)，其特色是將資料庫的資料以表格(又稱關聯網目)來呈現，此好處是可以降低資料儲存容量，並加快資料搜尋速度，同時因為模型理論簡單易於實行，因此目前廣泛使用的資料庫模型[6]。

例如在圖 1.5 聲音資料庫中可以將原始資料分成生物名稱、地區、儀器與儲存四個關聯網目，且每個關聯網目的欄位(又稱作屬性)代表著各種不同的資料種類，而關聯網目之間的結合是透過相同的屬性來連接。例如欲查詢原始資料錄音編號 332 的紀錄時間與生物名稱時，可以從儲存關聯網目中，找到儲存編號 333 中的紀錄時間為‘2008-10-10 00:08:34’且生物編號‘001’，再透過生物名稱關聯網目查詢，可得知該生物為‘短吻鰻’。

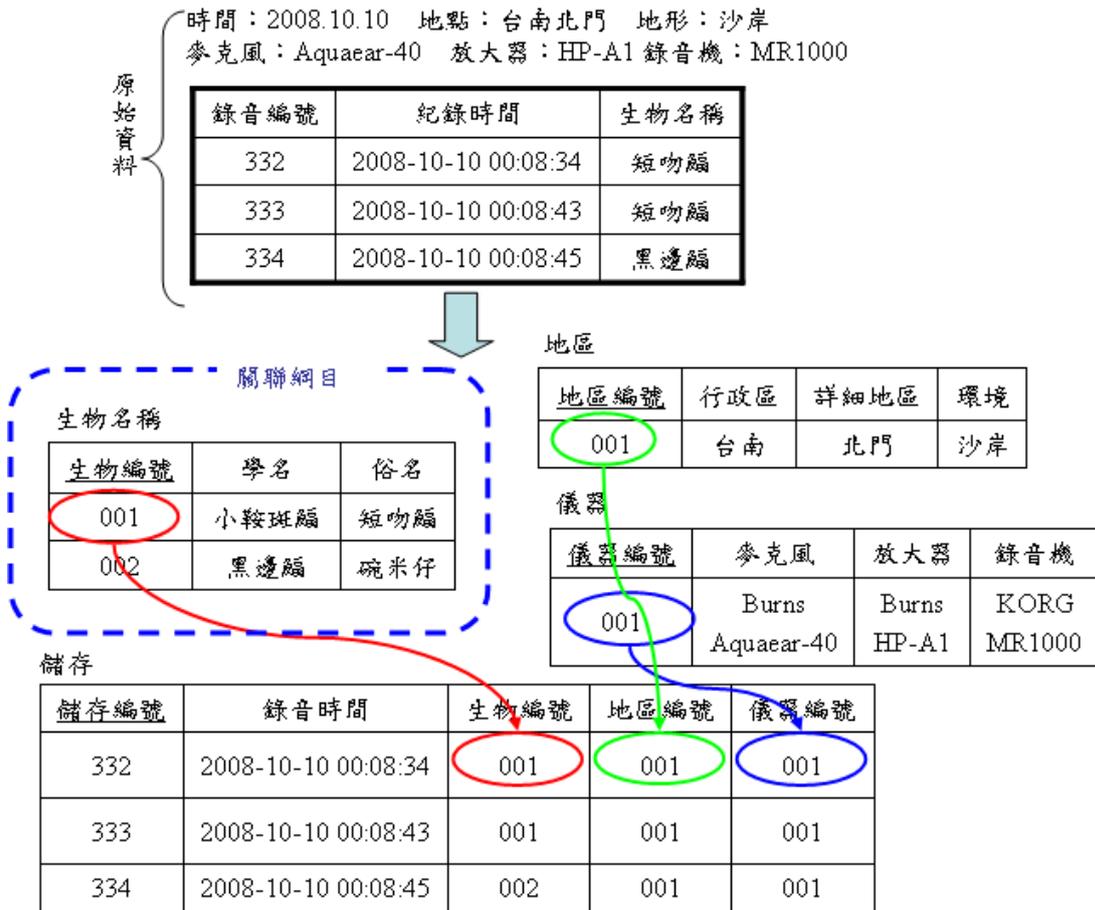


圖 1.5 聲音資料的關聯表示

#### 四、資料庫展示

台灣魚類生物多樣性極為豐富，且聲音是魚類多樣性的一環，魚類聲音資料庫的建立可以幫助保存來之不易的聲音資料，同時還提供一個生物聲學研究的交流平台 (<http://140.117.95.221/fishSoundDatabase/db-index.php>)，吸引更多人加入研究魚類聲學研究；除此之外還可向大眾推廣魚類聲音，讓更多人可以瞭解不一樣的聲音世界，以下分別介紹資料庫目前具備之功能：資料上傳、搜尋方法。

##### 4.1 資料上傳

本資料庫的資料上傳提供三種功能：錄音人員註冊與修改，以及聲音資料上傳。帳號的建立是希望錄音人員在通過適當地認證後，方能進行資料的上傳，避免不相干人事隨意輸入資料造成資料庫資料錯誤，除此之外還可以將錄音人員資料呈現在聲音樣本中，方便使用者進行聯絡，其帳號的申請內容包含中英文姓名、帳號、密碼、電話和電子信箱。

在註冊帳號後，具備權限的使用者輸入帳號與密碼後，方可執行聲音資料上傳，為求提高人員在輸入資料的方便性，並降低錯誤發生，資料庫採用選單的方式進行資料儲存，如圖1.6。選單的內容為資料庫已儲存的資料，倘若資料庫選單中無所要輸入的資訊，則需要手動輸入進資料庫中，當新增成功後即可從選單中找到以儲存之資訊。最後進行資料庫儲存時，本資料庫會出現第二次的聲音樣本確認，讓錄音人員可以再次檢查輸入資訊是否正確。



圖1.6 資料庫的資料上傳

#### 4.2 搜尋方法

本資料庫提供的搜尋方法有三種，如圖1.7，在生物搜尋中，可以直接輸入欲查詢的生物中文名稱，也可以從生物名稱清單中選擇有興趣的生物；在地點查詢中，除了直接輸入地點或是從清單選擇之外，還能直接點選地圖，得到該地點收集到的聲音樣本筆數；而除了上述的搜尋方法外，本資料庫還可以聲音樣本的錄音時間進行搜尋，讓使用者以是合得搜尋方法，得到有興趣的資料。



圖1.7 資料庫生物、地點、時間之搜尋方式

在聲音樣本呈現中，資料內容包含錄音人員、儀器、地點、時間、聲音播放、時間序列圖、頻譜圖等，除此之外還直接點選生物介紹，连接到其他生物網站(例如：台灣魚類資料庫)，讓使用者瞭解其生物的詳細資訊，如圖1.8。



圖1.8 黑邊鰻聲音樣本細節介紹

## 五、結論與建議

一個聲音資料庫的成長性，端賴基礎架構是否完善與其內容的吸引力是否足夠，當資料庫的內容越豐富，提供的服務越多樣化，便會吸引更多的人使用聲音資料庫，延長資料庫的壽命。本研究主要著重在資料庫的內容建立，除此之外也研究聲音樣本的品質分析，希望結合聲音樣本品質與其他聲音資訊，讓魚類聲音資料庫的內容呈現更加豐富，以增加使用者的便利性。

在上傳資料之前，本研究先對聲音樣本進行品質分析，以端點偵測法框選魚類訊號，計算聲音樣本的訊雜比作為品質分析結果。所使用的端點偵測法有兩種，一種是時域端點偵測法，藉由魚類聲音能量高於背景值的特性，以音框長度 0.5 秒、重疊率 50% 切割聲音樣本，計算出每個音框的能量和後，設定門檻值為聲音樣本能量的中位數加上一個標準差，框選出魚類聲音訊號；另一種方法為頻率端點偵測法，藉由魚類訊號與背景值的頻率組成比例不同，從聲音樣本的信息熵來偵測生物訊號，其方法是以音框長度 0.5 秒、重疊率 50% 進行聲音樣本切割後，將每個音框轉換成頻率訊號，找出每個音框的各頻率能量所佔比例，最後計算出音框的信息熵大小，設定門檻值為中位數加上一個標準差進行魚類聲音偵測。

本資料庫採用三層式架構，為目前現階段主流的資料結構，特色是將資料庫分成資料庫應用系統與管理系統，並將資料庫應用系統放置於應用伺服器，以 PHP 與 SQL 撰寫資料上傳、搜尋，輸出等指令；資料庫管理系統放置於資料伺服器，以關聯模式作為資料庫儲存的模式，用 MySQL 進行資料庫綱目的設定。本資料庫除了自動將聲音樣本進行品質分析外，同時還儲存了聲音樣本的詳細資訊，可充分描述聲音樣本給使用者，並連結到台灣魚類聲音資料庫，介紹更加詳細的生物資訊，最後以生物名稱、錄音地點、錄音時間三種方法進行搜尋，讓使用者可以迅速找到其需要的資料。

### 5.1 建議

聲音資料庫的建立除了可以有系統保存資料外，還可以增加資料的使用效率，便利科學家進行研究，而除了在資料保存與科學用途外，資料庫同時也可以扮演推廣大眾教育的工具，讓台灣人民對於周遭海域的魚類有進一步的認識。因此資料庫除了在內容上需要有足夠的豐富性外，還需兼顧趣味性，例如增加魚類的圖片或是影片，且依照不同的使用者層次設定網頁內容，如此才能增加資料庫的曝光度，吸引更多人來瀏覽，延長資料庫的壽命。

而在樣本品質分析研究中，本文是以訊雜比作為分析的依據，為此引用時域端點

偵測法與頻率端點偵測法，依照不同的聲音來源選擇合適的偵測法，進行魚類聲音訊號的萃取。雖然兩種偵測法可以拓展資料的來源，但是實際上端點偵測法仍有其缺點，例如門檻值設定不夠精確，導致萃取效果不佳，且當背景環境趨向複雜時，無法有效辨認魚類聲音。為此需進一步結合兩種偵測法的長處，提高辨識的正確率，或是尋求另外的偵測法，例如引用類神經網路法或是隱藏馬可夫模型法，增加萃取的可靠性，讓聲音樣本的訊雜比計算變得更加精確。

## 六、參考文獻

- [1] A. A. Myrberg, "Sound communication and interception of fishes", In: Hearing and Sound Communication in Fishes, Springer-Verlag, New York, pp. 395-452.
- [2] L. S. Huang and C. H. Yang, "A novel approach to robust speech endpoint detection in carenvironments", Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on, Volume 3, pp. 1751-1754.
- [3] P. Clemins, M. Johnson, K. Leong, and A. Savage, "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations", J. Acoust. Soc. Am. Volume 117, pp. 956-963, 2005 .
- [4] C. Erbe and A. R. King, "Automatic detection of marine mammals using information entropy", J. Acoust. Soc. Am. Volume 124, Issue 5, pp. 2833-2840, 2008.
- [5] 陳祥輝，資料庫系統理論與實務，旗標出版股份有限公司，2008。
- [6] E. F. Codd, "A Relational Model of Data for Large Shared Data Banks", Communication of the ACM, Volume 13, Issue 6, pp. 377-387, 1970.