南華大學科技學院永續綠色科技碩士學位學程

碩士論文

Master Program of Green Technology for Sustainability
College of Science and Technology
Nanhua University

Master Thesis

臺灣特有種鳥類聲音辨識系統之研究

The Research of Audio Recognition System for Taiwanese Endemic Birds

林財生

Cai-Sheng Lin

指導教授：陳萌智 博士

Advisor: Meng-Zhi Chen, Ph.D.

中華民國 111 年 6 月

June 2022

南 華 大 學

永續綠色科技碩士學位學程

碩士學位論文

臺灣特有種鳥類聲音辨識系統之研究

The Research of AudioRecognition System for Taiwanese Endemic Birds

研究生：林財生

經考試合格特此證明

口試委員：翁富美

陳前智

陸海文

指導教授：陳前智

系主任(所長)：洪耀明(甲)

口試日期：中華民國 111 年 6 月 28 日

致謝

在這三年的研究所的時光中，非常感謝我的指導老師，不斷的給我許多的建議以及研究的方向，也謝謝你的指導與督促，讓我的論文能如期產出，辛苦您了。

感謝擔任口試委員的翁富美教授、陸海文教授、陳萌智教授，謝謝你們在百忙之中抽空來擔任口試委員，也感謝你們在口試上提出的寶貴建議。

感謝在召會生活中的弟兄姊妹，謝謝你們讓我在這三年間過的非常充實，謝謝父老與聖徒的關心與愛宴，謝謝服侍者的督促與鼓勵，謝謝同伴的陪伴與幫助，謝謝你們。

另外感謝我的朋友們，謝謝你們的關心與陪伴，你們的問候與關心雖然有時會讓我有壓力，但也告訴我，需要快點完成我的論文。

最後謝謝我的媽媽與姊姊，在這段時間的支持與關心，讓我在可以沒有壓力的就讀研究所，謝謝你們各方面的支持。

林財生謹誌

中華民國 一一一年 七月

摘要

人工智慧可加速達成聯合國永續發展目標（SDGs），例如：SDG 15 的陸地生態(Life on land)利用物種識別和智慧物聯網的廣泛運用，追蹤陸地動物的遷徙、族群數量水準等活動，進而增強永續的陸地生態系統。據報導知，臺灣的鳥類占了全球鳥種的二十分之一，因此有許多賞鳥人士慕名而來，所以本研究將快速發展的人工智能應用於聲音辨識技術，先擷取鳥類聲音的樣本之特徵資料，並將其以 AI 深度學習中的卷積神經網路建立模組，將模組建置在 APP 期望能滿足眾多賞鳥愛好者的使用需求。為了探討臺灣特有種鳥類聲音辨識系統的服務需求，我們以有使用過 APP 應用程式經驗的年輕族群為測驗對象，引用服務體驗工程法為理論基礎，訪談與觀察探討使用者行為中的隱藏的意義，從歸納出臺灣特有種鳥類聲音辨識系統的服務需求。

根據服務體驗工程法中的五大構面進行訪談，並將訪談的資料匯整到五大模型中，分析在使用臺灣特有種鳥類聲音辨識系統的潛在的問題與需求。根據研究訪談的結果，發現使用者的需求：

(1)目前辨識率為 77%，需要再提升預測的正確率。

(2)系統設計上需要加強美工部分以及，資訊反饋的豐富度。

(3)改善聲音易容易被干擾的因素。

以上三項可做為未來後續服務的主要依據。

關鍵詞：鳥類聲音辨識、深度學習、服務體驗工程

Abstract

Artificial intelligence can accelerate the achievement of the United Nations Sustainable Development Goals (SDGs), such as: SDG 15's Life on land uses species identification and the widespread use of the AIoT(Artificial Intelligence of Things) to track the migration of terrestrial animals, population levels and other activities, and then Enhance sustainable terrestrial ecosystems.

According to reports, Taiwan's birds account for one-twentieth of the world's bird species, so many bird watchers come here. Therefore, this research applies the rapidly developing artificial intelligence to sound recognition technology. The characteristic data of the sample is used to build a module with the convolutional neural network in AI deep learning, and the module is installed in the APP to meet the needs of many bird watching enthusiasts. In order to explore the service needs of Taiwan's endemic bird sound recognition system, we took young people with experience in using APP applications as the test objects, cited the service experience engineering method as the theoretical basis, and discussed the hidden meaning in user behavior through interviews and observations. , summed up the service requirements of Taiwan's endemic bird sound recognition system.

Interviews were conducted according to the five aspects of the service experience engineering method, and the interview data were compiled into five models to analyze the potential problems and needs of using the sound recognition system for endemic species of birds in Taiwan. Based on the results of the research interviews, the needs of users were identified:

(1) The current recognition rate is 77%, and the accuracy of prediction needs to be improved.

(2) The system design needs to strengthen the art part and the richness of information feedback.

(3) Improve the factors that the sound is easily disturbed.

The above three items can be used as the main basis for future follow-up services.

Keywords: Bird sound recognition, Deep learning, Service experience engineering

目錄

表目錄

圖目錄

公式目錄

第 1 章　緒論

1.1研究動機

臺灣地形豐富，既有海拔 4000 公尺的高山，也有低於海拔 500 公尺平原，因為這樣的高低落差，讓臺灣在氣候方面不是單一型氣候，而是包括熱帶型、亞熱帶型、溫帶型..等的多種氣候。這些豐富的氣候，也造就了臺灣生物的多樣性，而鳥類更是在這其中最豐富的，占了全球鳥種的二十分之一。不只如此，在這特別的氣候以及種類繁多的鳥類的環境影響之下，自西元 2000 年開始，在臺灣掀起了一股賞鳥的熱潮，在 2020 年發表臺灣國家鳥類報告(The State of Taiwan’s Birds 2020）的助理研究員林大利就在採訪中指出國人在 eBird Taiwan 的賞鳥紀錄資料庫中，有近 3900 人貢獻鳥類觀察紀錄，並累計超過 47 萬 3 千份賞鳥紀錄，為東亞排名第一，全球第排名七。不只如此在 2021 年 5 月 14 日由 eBird 所舉辦的全球觀鳥大日(Global Big Day)，據官方統計臺灣上傳了 987 份紀錄，紀錄了 266 種鳥類，排行全球第 15 名。讓國人如此著迷於賞鳥的原因，不單是因臺灣鳥類眾多，更是因為臺灣擁有多達 30 種的特有種鳥類，例如：黑長尾雉、小彎嘴、藍腹鷴、繡眼畫眉、深山竹雞、白耳畫眉...。

在網路尚不發達的年代，賞鳥所需的裝備相當繁瑣，除了基本的

望遠鏡腳架等，還需要攜帶紙本圖鑑跟紀錄本；而在現今受惠於發達的科技，只需要攜帶手機，即可滿足圖鑑以及記錄本的功能。不只如此也因著科技的日新月異，人工智慧、機器學習、深度學習、神經網路...等技術漸趨成熟而其帶來的聲音辨識也不斷被提及，在生活、醫療、娛樂、運動中，都有這些技術的蹤跡。本研究期許藉由聲音辨識技術的應用，以滿足國人的賞鳥需求。

1.2研究目的

現今是一個人工智慧盛行的時代，AlphaGo 展示了強大的運算處理能力，而生活中小至影像辨識應用的結帳，大至自駕車等，都是人工智慧應用的結果。鑒於研究動機所提到的需求與背景，加上市場產品的情形，本研究希望結合 AI 發展出一套款鳥類語音辨識系統，以增加賞鳥時辨識鳥類的選項。因此，本研究之目的在於從使用者的觀點進行體驗觀察、訪談及分析，並歸納臺灣特有種鳥類聲音辨識系統的服務需求。

本研究之目的如下：

一、使用者實際使用系統的辨識成功率。

二、臺灣特有種鳥類聲音辨識系統是否能滿足使用者的需求。

三、利用本研究所歸納出的結論，進一完提升系統品質，以滿足使用者的需求。

1.3研究流程

本研究的研究流程如下，首先分析問題以及找出需求，緊接著找尋相關文獻，確定研究方法，再根據研究方法蒐集資料，建立並訓練模組，確認其準確率，使用其模組建置網頁程式，並以 APP 的形式呈現，另一方面會建置資料庫來與 APP 連結。完成後進行使用者訪談，找出使用者對本系統的問題與需求，並以此做出本研究的結論與建議。以下是本研究的研究流程圖，如圖 1- 1 所示：

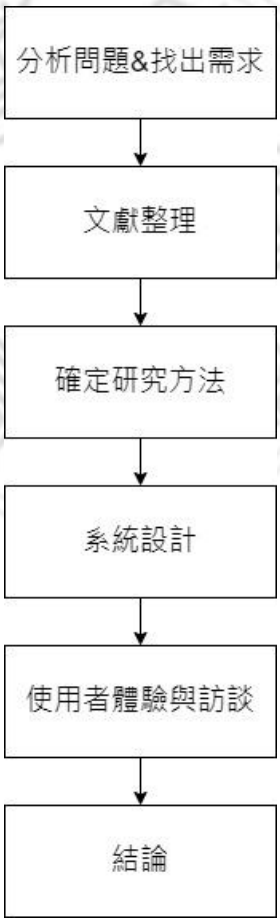


圖 1- 1 研究流程圖

第 2 章　文獻回顧與探討

2.1聲音辨識的相關研究

聲音與人類生活有著密不可分的關係，過去人們只能用耳朵去分辨諸多不同的聲音並做出判斷，伴隨著人類對聲音辨識的需求與日俱增，許多聲音辨識的科技也隨之問世，且日漸強大，現今聲音辨識不僅能區分人聲、物聲、動物聲，更能分辨不同的環境聲(黃俊卿,2017)。聲音辨識在我們如今的生活中也隨處可見。例如：翻譯與人機互動上有分辨語句的語音辨識技術(Avutu, S. R.,2017)，藉著聲音變化來控制的智能家電(Erić, T., 2017)，更有出現能擷取聲音特徵的身分驗證系統(Yan, Z.,2016)；在醫療上也有針對呼吸道疾病(Bajaj, V. ,2020)、肺音(Bongo, L. A. ,2017)、心音(Kuan, K. L. ,2010)的應用；而在環境上也有針對動物、自然現象的應用。

本研究參考了各個學者對於鳥類的聲音辨識的所提出的研究與技術。McIlraith 和 Card 是首先在較大量鳥聲辨識中使用自動分類，在 1995~1997 年又陸續提出許多辨識鳥類的方法，他們運用類神經網路、統計方法、時間參數、短時間光譜訊息、二次判別分析..等(Card, H. C. ,1997)。Ander sonetal.使用了 Dynamic time warping(DTW)，來直接比較光譜圖(Anderson, S. E.,1996)。Kogan 和 Margo 比較了 Hidden

Markov models(HMMs)和上述所說的 DTW 並藉接發展出可以分辨兩個物種的聲音系統。在他們的研究中 HMMs 和梅爾頻率倒譜係數(MFCC，Mel-frequen cycepstral coefficients)是在有背景雜訊的情況下還能有效辨識的方法。

直至現今 HMMs 和 MFCC 仍佔有重要的一席之地，而 MFCC 更是成為了語音辨識最普遍使用的方法，近年來許多的聲音辨識的研究也都把梅爾頻率倒譜係數當作特徵提取方法，在醫療上有聽診器(Wang, G.,2019)、肺部的聲音檢測(Chambres, G.,2018)，也都是使用MFCC。

2.2梅爾頻率倒譜係數

梅爾頻率倒譜係數(MFCC,Mel-frequen cycepstral coefficients)，為近年來使用最普遍的聲音辨識系統上的特徵函數，此技術於 1980 年由 S.B.Davis 和 Paul Mermelstein 所提出來(Chen, Y. J. ,2009)。主要功能為能將聲音的頻譜包絡係數化成倒頻譜(Huang, X.,2001)。梅爾倒頻譜(Mel-Frequency Cepstrum，MFC)最大的特色在於梅爾倒頻譜上的頻帶是均勻分布於梅爾刻度(melscale)上，讓此頻帶相較一般所看到線性的倒頻譜之表示方法，更接近人類的聽覺系統(audio system)。當實際頻率小於 1kHZ 時，梅爾頻率與實際頻率呈現線性關係；當頻率大於 1KHZ 時，兩者則呈現對數關係，其關係表示如圖 2- 1 所示，以下介

紹少 MFCC 之步驟，實行流程如圖 2- 2 所示。

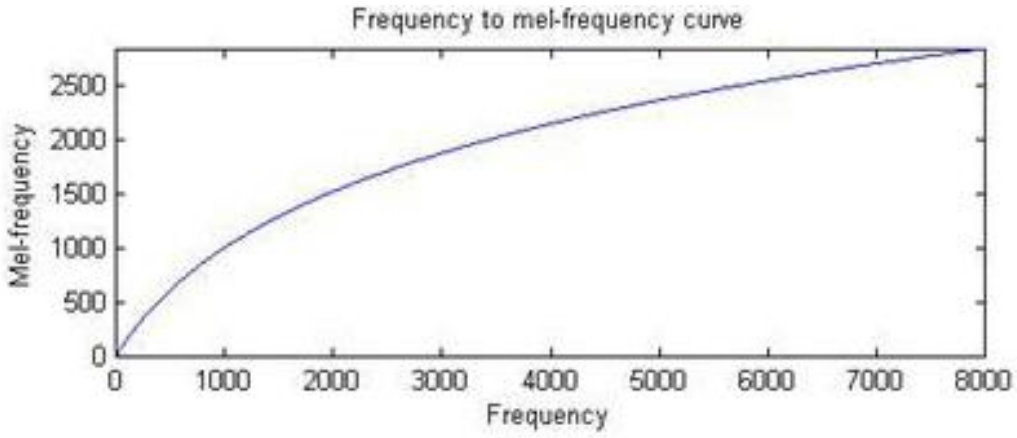


圖 2- 1 實際頻率和梅爾頻率的關係圖

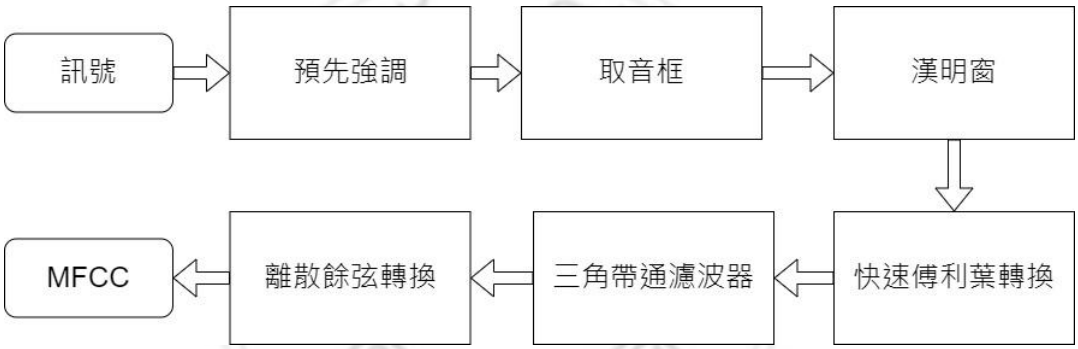


圖 2- 2 取梅爾頻率倒譜係數的流程圖

2.2.1 預先強調(Pre-mphasis)

預先強調主要是用來降低量化失真(Quantizing Distortion),量化失真是一種在數字模擬過程中當數字信號失真達到一定程度所而導致原始音頻不規則失真的失真類型，在處理聲音訊號時在高頻率段振幅就會比較小，且隨著頻率越來越高，振幅也越來越小。這會導致有效振幅的數字變小，與原本的波型的誤差也越來越大，所以才需要使用預先處理來增加高頻率的振幅，降低取樣高頻的量化失真。

2.2.2 取音框(Frame Blocking)

語音或非語音訊號通常屬於非穩態，也就是隨時間一直改變因此分析聲音訊時音訊必須將聲音訊沆分成許多短暫的音框 (fratne) 來處理，一般取 20~40ms 為適合範圍，同時避免相鄰的兩個音框間的變化太大，在取音框時會讓相鄰兩音框間有段重疊的區域，重疊區域通常是音框長度的一半或三分之一。

2.2.3 漢明窗(Hamming window)

將取樣好的每個音框都乘上漢明窗(王小川,2004)以加強音框中訊號左、右端的連續性，避免在進行傅立葉轉換時，因為音框中訊號左右端的太明顯的不連續變化而產生不存在原聲音訊號的成分能量，導致產生分析誤差。而且漢明窗的頻譜顯示最大的旁瓣(side lobe)會比主瓣(main lobe)小 42.76dB，能夠有效減少訊號在頻域產生的額外頻率成分。漢明窗的形式如式 2-1：

$$w[n] = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right], \quad 0 \le n \le N-1$$

公式 2- 1 漢明窗的形式

其中 N 為音框的取樣點數。取樣好的每個音框乘上漢明窗後形式如公式 2-2：

$$x_w[n] = x_{PE}[n] \cdot w[n]$$

公式 2- 2 每個音框乘上漢明窗後形式

2.2.4 快速傅利葉轉換（Fast Fourier Transform,FFT）

由於訊號在時域（Time domain）上的變化通常很難看出訊號的特性，所以通常每個音框還必需再經過 FFT，將它轉換成頻域（Frequency domain）並取絕對值後平方得到功率頻譜(powers pectrum)，得到在頻譜上的能量分佈(Hubel, D. H., 1960)。所得到的能量分佈即聲音的特性。

2.2.5 三角帶通濾波器（Triangular Bandpass Filters）

將所得的聲音特性乘以一組 20 個三角帶通濾波器(王小川,2004)，求得每一個濾波器輸出的對數能量（Log Energy）。必須注意的是：這 20 個三角帶通濾波器在「梅爾頻率」（Mel Frequency）上是平均分佈的，而梅爾頻率和一般頻率 f 的關係式如公式 2-3：

$$Mel(f) = 2595 * log(1 + \frac{f}{700}) = 1125 * ln(1 + \frac{f}{700})$$

公式 2- 3 梅爾頻率和一般頻率 f 的關係

梅爾頻率代表一般人耳對於頻率的感受度，由此也可以看出人耳對於頻率 f 的感受是呈對數變化的。在低頻部分，人耳感受是比較敏銳；在高頻部分，人耳的感受就會越來越粗糙。

三角帶通濾波器有兩個主要目的，第一是降低資料量，第二是對頻譜進行平滑化，並消除諧波作用，突顯原先語音的共振峰，也就是音調或音高，不呈現在 MFCC 參數內，以 MFCC 為特徵的語音辨識系統，並不會受到輸入語音之音調影響。

2.2.6 離散餘弦轉換（Discrete cosine transform,DCT）

離散餘弦轉換是一個無損的，可逆的數學過程，它是一種時域到頻域的轉換。離散餘弦轉換公式如公式 2-4：

$$C_m = \sum_{K=1}^{N} \cos\left[m^{(k-0.5)\frac{\pi}{N}}\right]^{E_k}$$

公式 2- 4 離散餘弦轉換

2.3深度學習

機器學習中的深度學習(Deep Learning)，以神經網路為架構式，對大量資料進行的演算法，達到訓練模組與模組建構。神經網路是由多個資訊處理單位集合所組成，彼此團隊合作並互相傳送資訊，就像大腦中的神經元一樣，但無法像人類一樣串聯所有細節之間的脈絡關係。不過類神經網路可以找出模式，藉著設計者指派神經元特定的工作，再藉著神經網路的合作，理解各個元素間的關係與相關性，並找出元素通常如何互相搭配以及影響。

在 1943 年，Warren McCulloch 和 Walter Pitts 使用數學和算法創建了一個複制神經網絡的計算系統，是深度學習的起源。這項技術一直到 1980 年代都得到了一些進展。但在 1999 年，因著當時硬體發展迅速，計算機處理速度和圖形處理單元都得到發展，使其在接下來的十年裡，讓笨拙和低效的系統變得快了 1000 倍。

此後，Google 在 2012 年通過一種可以識別貓的算法 The Cat Experiment，將深度學習提升到一個新的水平，此算法使用無監督學習和 10,000,000 張貓的圖像來訓練它識別貓，但那時它識別的貓的比例依然不到所有貓的圖像的 16%。

2016 年，Google Deep Mind 的算法 AlphaGo 通過強大的學習能力創造了歷史，在韓國首爾擊敗了職業圍棋手。

現今大數據分析突飛猛進，神經網絡也越來越複雜。導致計算機在觀察、學習和對複雜情況做出反應方面加快了不只一個檔次，在某些情況下甚至比人類思維還快。模型使用了大量標記數據和層數較多的神經網絡進行訓練，並借助圖像分類、翻譯能力和語音識別技術，甚至不需要人類的幫助進行解碼模式識別。常見的神經網路建構有卷積神經網路(CNN)、遞歸神經網路(RNN)、長短期記憶神經網路(LSTM)。

2.4卷積神經網絡

2.4.1 發展

卷積神經網路(Convolutional Neural Network,CNN)是深度學習裡極為重要的一門分支，在 1950~1960，神經生理學家 Hubel 和 Wiesel 基於對猴子和貓視覺的研究，提出了腦中兩種基本型態的視覺細胞，分別稱為 simple cells 和 complex cells(Hubel, D. H.,1960)。1980 年日本電腦科學家福島邦彥提出，CNN 架構的源頭 Neocognitron 神經感知機(福島邦彦,1997)1987 年 1987 年 Alex Waibel 和 Geoffrey Hinton 等人在日本的 ATR(Advanced Telecommunication ResearchInstitute)發表了第一個使用 back propagation 做梯度下降訓練的卷積神經網路，時間延遲神經網路 TDNN(Lang, K. J.,1989)。

1989~1998 卷積網路(convolutional nets)之父 Yann Le Cunn 表了許多關於 CNN 的研究，為現代 CNN 架構打下了確實的基礎(Freund, Y.,1996)。1998 年 LeNet-5 提出與現今幾乎一樣的 CNN 架構，並在文字圖像辨識上無人出其右，但當時硬體發展跟不上，資料也不大，導致在其年代神經網路並不能發揮其優勢(Brunot, A.,1995)。

2006 年硬體發展迅速 Kumar Chellapilla，利用 GPU 平行運算的能力，將 CNN 模型 forward propagation 和 back-propagation 的速度提昇 3~4 倍，使 CNN 模型不論是在訓練或測試時，時間都大幅縮短，這也

是第一個用 GPU 進行訓練的 CNN(Chellapilla, K.,2006)。同年史丹佛大學電腦科學教授李飛飛為擴大及增進訓練 AI 所能使用的資料開創了 Image Netdataset。

2012 年 ILSVRC 競賽上 Alex Krizhevsky 使用了 CNN 架構的 Alex Net 在圖像分類項目上取得優勝，且遙遙領先其他 CNN 架構的隊伍，這也影響往後幾年的比賽都被 CNN 架構的模型所統治(Hinton, G. E. ,2012)。

2014 年牛津大學 VGG 以 19 層的 CNN 將 Top5 errorrate 從原本 Alex Net 的 16.4%下降到 7.3%(He, K.,2015)，而 Google Net 再以 22 層的 CNN 將 Top5 errorrate 下降到 6.7%，此時距離達到人類的 Top 5 errorrate:5.1%已經不遠。隨著 Alex Net,VGG,Google Net 在層數上的加深模型表現也越來越好，開始出現一個問題：The deeper,the better?，意即層數越多模組越好，但事實卻是層數到達臨界後，效能反而下降了。

2015 年微軟研究院的何凱明他設計了帶有殘差結構的神經網路，並發表了深達 152 層的 ResNet. ◦並在該年的 ILSVRC 競賽中拿下第一。其 Top5 errorrate 是超越人類 5.1%的 3.57%，ResNet 的出現也為 ILSVRC 的該項分類競賽畫上了句點(He, K.,2016)。

2.4.2 原理以及架構

卷積神經網路由卷積層、全連通層、池化層組成，搭配反向傳播演算法的運算，能夠利用輸入資料的二維結構擷取特徵並適當的收斂與學習，在圖像和語音辨識方面有出色的表現，其架構如圖 2- 3 所示(陳家安,2019)。

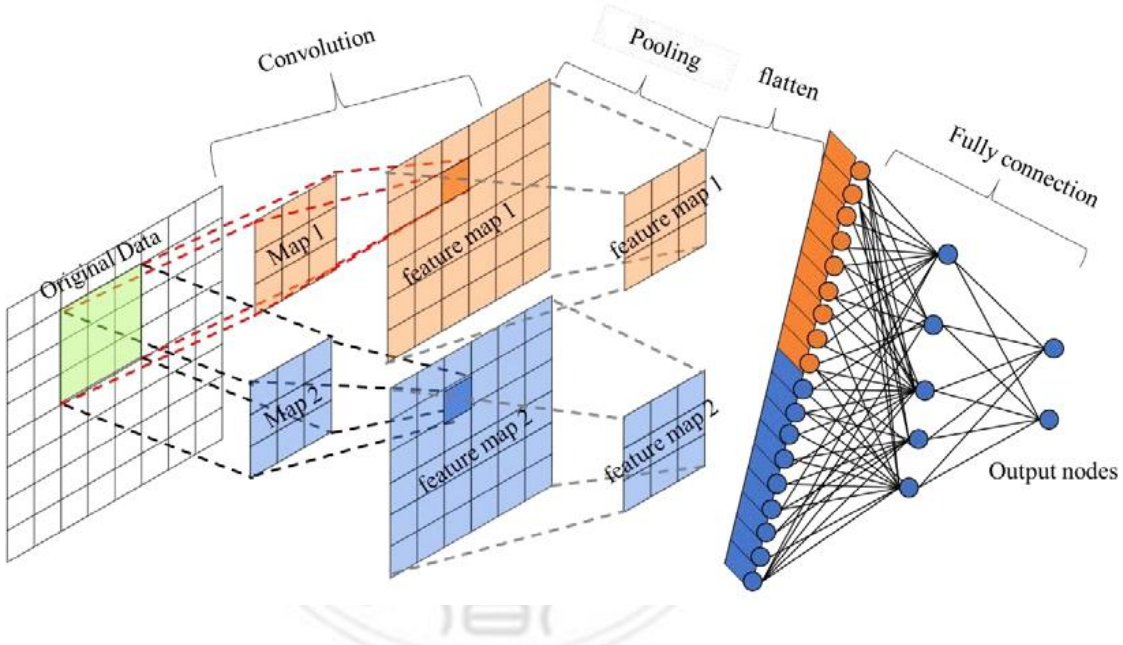


圖 2- 3 卷積神經網路的架構

2.5服務體驗工程法

「服務體驗工程法」(Service experience engineering,S.E.)經由資訊工業策進會提出的方法論(資訊工業策進會,2008)，期因緣乃是在 2008 年，科技業快速發展的情況下，臺灣的服務產值 GDP 居然佔了七成，進而顯示其相關的產業是臺灣未來開發的重點之一，但策進會發現國

內在創新服務發展上缺乏系統化的服務，於是資策會整合服務、應用、平臺三大中心和 12 團隊研發的輔導經驗，採擷德國工研院 IAO 發展 41 種方法的服務工程方法論及美國 IDEO 設計公司在顧客體驗洞察中所研發 51 種洞察方法之各方專長，和義大利學者 Roberta 在服務設計 40 個工具，結合辛辛那提大學教授、美國智能維修系統研究中心(IMS Center)主任與上海交通大學產業技術研究院院長李傑博士開發的創新矩陣服務發想工具，創造適合臺灣企業能使用的服務設計理論，服務體驗工程方法論，也可簡稱為 SEE 方法(Service Experience)。

服務體驗工程為服務與研發工作的流程模式方法論，整體模式主要分為三大階段(Phase)，分別是趨勢研究(FIND)、服務價值鏈研究和服務實驗(Design Lab)。而服務體驗工程方法論在此模式架構下把創新服務發展所需要研究的之面向及可使用的工具與方法進行系統性統整。從三大個階段還衍生出六大程序，包括趨勢研究、產業價值鏈研究、服務塑模、概念驗證、服務驗證、商業驗證。研究者可以根據服務或產品的特性，選擇合適的方法以及程序(王熙哲,2012)，如圖 2- 4 所示(服務體驗工程方法流程圖,2009)。

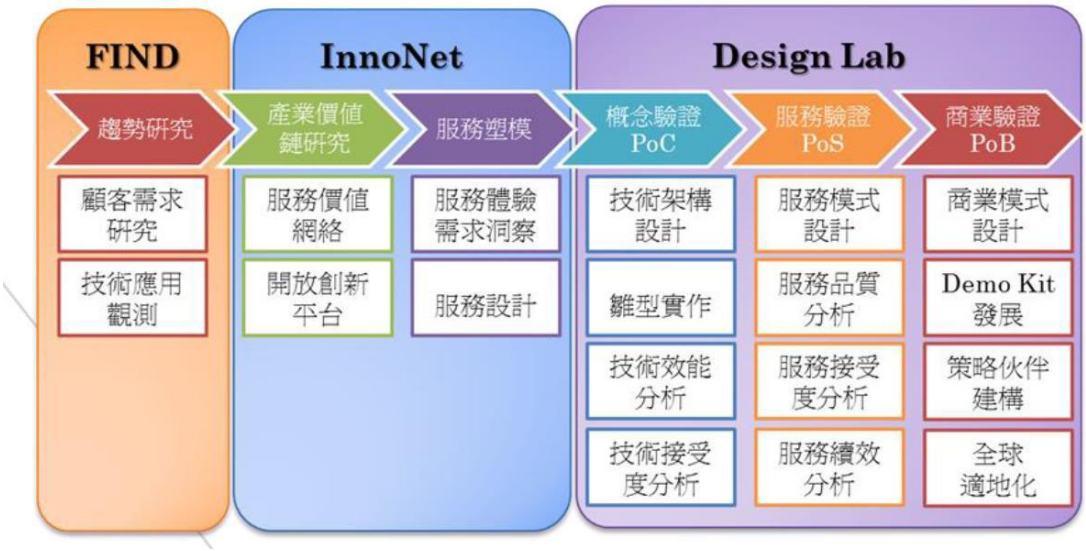


圖 2- 4 服務體驗工程方法流程圖

趨勢研究(FIND)：服務體驗工程方法論中第一個階段為 FIND 階段，研究大環境的發展趨勢，找出消費者的需求或潛藏的商機。此階段主要的工作，利用研究資訊技術的發展趨勢或舊有研究資料等大環境趨勢的變化，蒐集新穎服務的創意，並具體化評估過濾的過程。目的是決定一個新服務的創新可行性。節省新服務的研發成本，並確認新服務的市場之接受度，須要針對新創意可行性以及市場潛力提早做出評估與調查。 FIND 階段研究結果為產出成功率高、可行性高的新服務創意。然後就是進入創新服務的研發工作(林曉琪,2010)。

服務價值鏈研究：主要分為兩個程序，產業價值鏈研究和服務塑模，產業價值鏈的研究主要目的是為了企業在界定新的創新服務產業所涵蓋的內容，同時完成企業的創新服務的雛型，作服務塑模階段時參考的依據。服務塑模部分也有兩個主要工作，經服務體驗洞察，尋

找使用者隱藏的真實需求，並利用系統化的服務設計將服務模型與藍圖具體呈現出來。

服務實驗(Design Lab)：服務正式建置或產品上市前，進行服務可行性驗證，透過與顧客實際的參與，實測結果作為上市前的最後調整，以最小風險讓產品或服務上市。在「生活實驗室」創新系統，用實際生活環境作為場景，研究受試者的參與、使用狀況，測試服務設計成果，進行必需的調整及改良。服務實驗的工作，可分成概念驗證（Proof of Concept，POC）、服務驗證（Proof of Service，POS）及商業驗證（Proof of Business，POB），驗證某服務概念的可行性與商業價值。

2.5.1 服務體驗脈絡洞察法

服務體驗脈絡洞察法在研究進行上可分成四個階段，其中四個階段包含界定議題和洞察目標、規劃觀察計畫、執行現場觀察、訪談資料蒐集與分析(資訊工業策進會,2008)，。透過上述四個階段蒐集資料，以五大模型作為分析工具。以五個不同構面的觀察重點去訪談使用者，將體驗訪談的資料彙整到五大塑模中，歸類出互動模型、序列模型、文化模型、工具模型、實體模型等。

服務體驗工程方法論重視研究者到實際場域與使用者互動，這樣才能探討出使用者需求以及服務價值核心。方法上主要以 A.E.I.O.U 五大構面作為觀察重點的依據，五大構面分別為活動(Activities)、環境

(Environments)、互動(Interactions)、物件(Objects)、使用者(Users)，而且訪談過程以參與式現場觀察，採用體驗旅遊框架讓受測者能在邊測試的過程中接受系統設計者訪談。

體驗旅遊框架:此觀察訪談法主要先界定使用本系統的過程中每個重要階段活動，事先分為好多個特定框架，並預先設想被觀察者在這些活動接觸點中會有哪些行為或活動。後在這些重要的接觸點裡，尋找出使用者對本系統的隱藏需求或服務潛在失效點，然後記錄使用者在體驗過程中行為模式。

行為塑模:行為塑模主要是利用圖形的方式表達，分別有互動模型(Flow Model)、文化模型(Cultural Model)、序列模型(Sequence Model)、工具模型(Artifact Model)、實體模型(Physical Model)(林義倫,2010)。

2.6小結

本研究根據所蒐集的文獻在模組的建構與訓練上選擇以梅爾頻率倒譜係數為聲音特徵擷取方法，並以深度學習中的卷積神經網路為架構，構建成本系統。再進入服務體驗工程法的 InnoNet 階段，探討出使用者使用系統的問題與真實需求。使用方法分為非參與式觀察法以及參與式觀察法。參與式觀察法則會依據體驗旅遊框架的方式，讓使用者依照本研究所設定的方式進行，並以觀察、訪談的方式找出使用者的需求，彙整成行為塑模中的五大模型，互動模型(Flow Model)、

文化模型(Cultural Model)、序列模型(Sequence Model)、工具模型(Artifact Model)、實體模型(Physical Model)。非參與式觀察法則是讓，使用者自由使用系統，再就由 A.E.I.O.U 五個構面，活動(Activities)、環境(Environments)、互動(Interactions)、物件(Objects)、使用者(Users)來進行訪談。

第 3 章　研究方法

3.1研究設計

本研究設計特有種鳥類辨識系統，利用服務體驗工程方法論中的體驗觀察與訪談，訪談 10 位測試者在體驗本系統之後的觀感和需求，進一步找出系統的優劣點，並將訪談的內容彙整到五大行為模型作分析，五大行為是透過五種不同面向去觀察，包括互動模型、序列模型、文化模型、工具器物模型和實體模型，透過行為統整後利用圖形的方式表達，從中找尋使用者的需求點及問題點，並提出對本系統之功能及服務上實際的情境脈絡。

3.1.1　系統設計

(一)系統架構

系統的設計上為了規劃邊緣運算的功能故採用以 TensorflowJS 架構運行聲音辨識，由於目前手機系統須要使用 SSL 才可以運行收音程式，然團隊架設的 WEB SERVER 並無 SSL 架構，故系統設計上分為儲存資料的資訊系統(IIS Web Server)與聲音辨識系統(Github Web Server)。如圖 3- 1 所示：

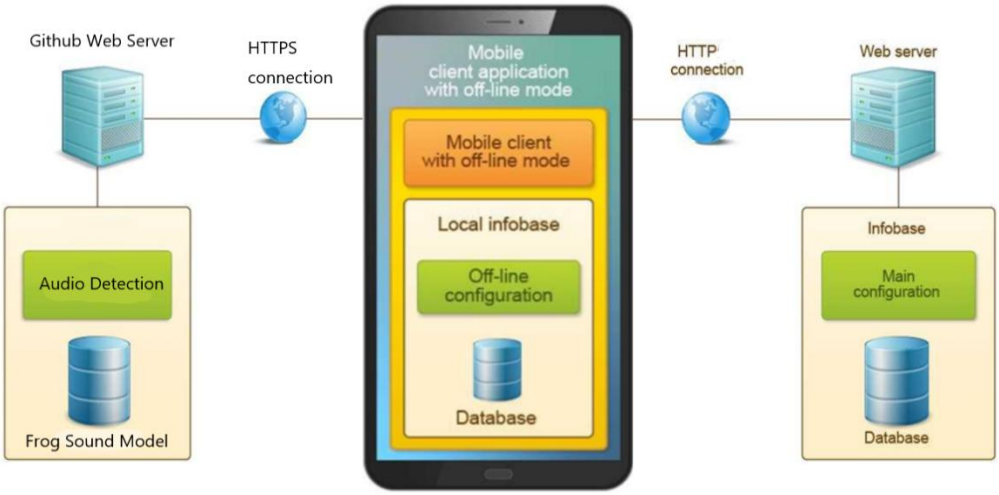


圖 3- 1 系統架構

(二)資料預處理

在開始建置系統前需要先進行模組的訓練，首先必須蒐集本研究所需要的音訊資料，資料來源則是 xeno-canto 網站，樣本的選擇則是臺灣特有種鳥類的 10 種，分別是台灣山鷓鴣、台灣竹雞、白耳畫眉、台灣叢樹鶯、台灣藍鵲、臺灣紫嘯鶇、台灣灰鶯、黑長尾雉、栗背林鴝、藍腹鷴，表 3- 1 是本研究所蒐集的音訊資料。將資料匯入後，首先需要將類別型資料轉成二進制的矩陣，然後是特徵資料需根據音訊的長短，本研究會將所有樣本的長度統一修改為 5 秒，總數為 1882 筆，之後進行提取聲音特徵 MFCC；MFCC 的參數為 40，所以每筆特徵值矩陣都會有 40 個數值，這時再求出 MFCC 的平均值與標準差讓特徵值能夠達到歸一化，以下是資料遇處理的流程，如圖 3- 2 所示。

表 3- 1 樣本資料

資料名稱	台灣山鷓鴣	台灣竹雞	白耳畫眉	台灣叢樹鶯	台灣藍鵲	臺灣紫嘯鶇	台灣灰鶯	黑長尾雉	栗背林鴝	藍腹鷴
資料長度	15分43秒	18分54秒	32分36秒	18分34秒	14分48秒	18分50秒	03分57秒	07分42秒	13分21秒	07分15秒

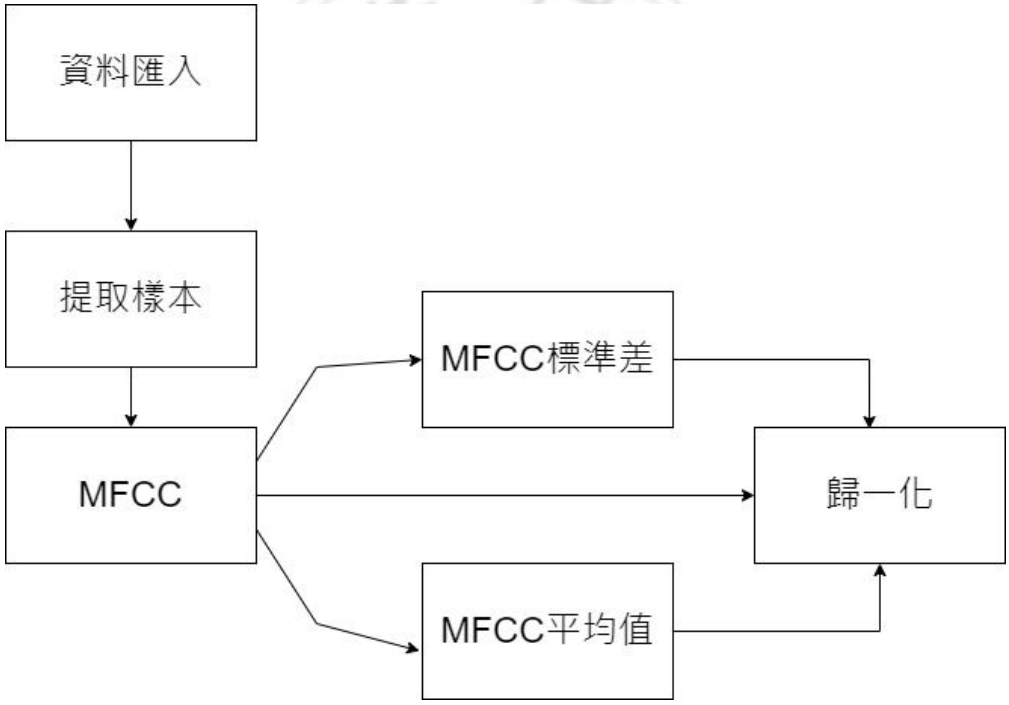


圖 3- 2 資料預處理流程

(二)模組建立

在完成資料預處理後，必須對資料進行取樣，將資料分為訓練資料與測試資料，訓練資料為全部資料的 70%剩下的 30%則是測試的資料，之後進入模組的建立與訓練，本研究的卷積神經網路架構參考 Inception 的架構，會有 3 個輸入的卷積層，且為了防止資料過度擬合化增加了許多 Dropout 層，而全連接層則是有四層，又因是分類的模型第四層的的函數選擇 Softmax，最後損失函數選擇 categorical_crossentropy，優化器則是 adam，以下是本研究所使用的卷積神經網路架構，如圖 3- 3 所示。

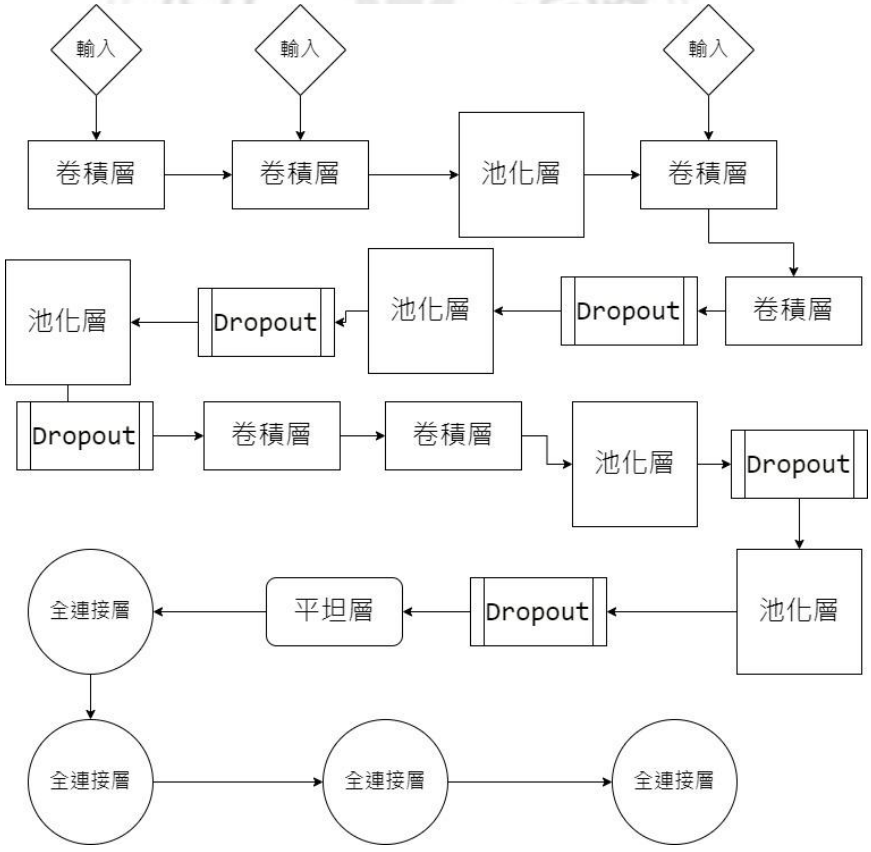
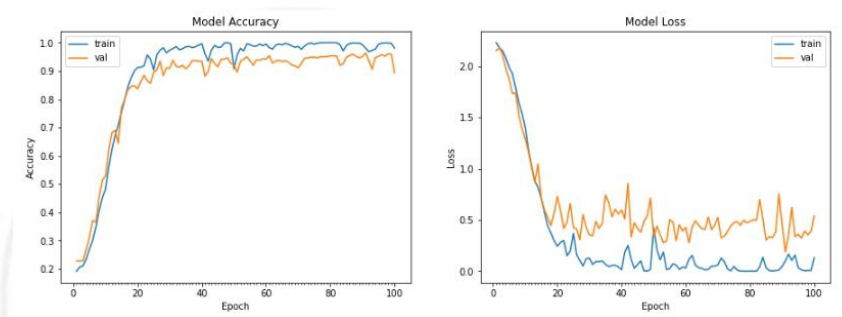
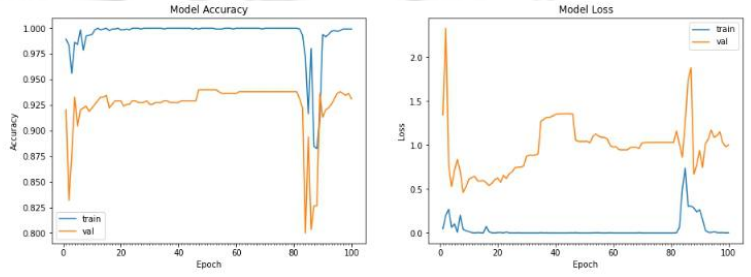
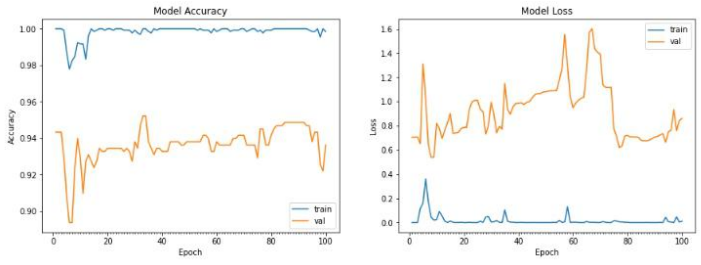
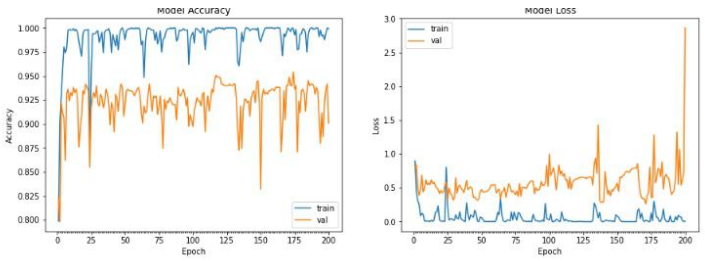
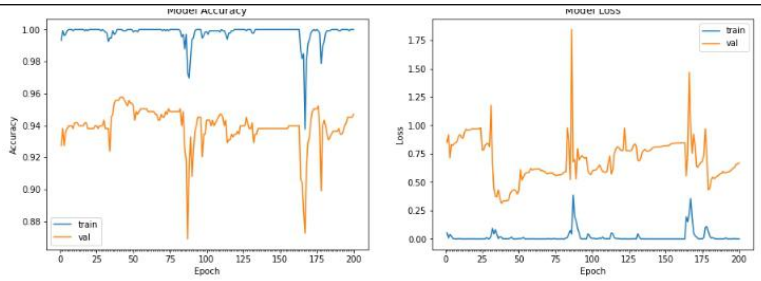
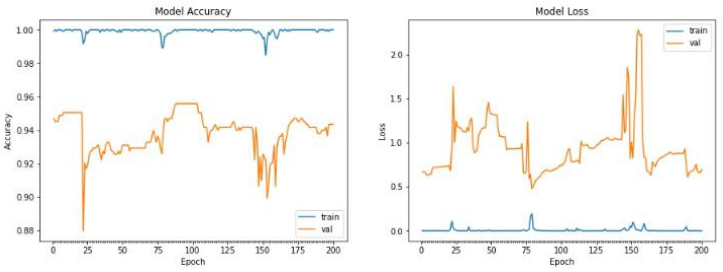


圖 3- 3 卷積神經網路架構

模型訓練的情況則是有以 bach_size 與 epochs 有所不同。以下是各個 epochs 以及 bach_size 的訓練情況如所示。可以看見 epochs=100 以及 bach_size=16 的模型相較於其他模型的正確率呈現穩定成長以及損失函數則是穩定下降，選擇 epochs=100，bach_size=16 作為本研究的模型訓練參數。

表 3- 2 模組訓練情況

Epochs=100 bach_size=16	
Epochs=100 bach_size=32	
Epochs=100 bach_size=64	
Epochs=200 bach_size=16	

Epochs=200 bach_size=32	
Epochs=200 bach_size=64	

(三)模組驗證

在機器學習的分類領域中，常使用混淆矩陣(confusion matrix)的元素計算精確率(precision)、回應率(recall)及 F1 度量，以判斷該模型的表現。以下是模型預測的混淆矩陣，如圖 3-4 所示：

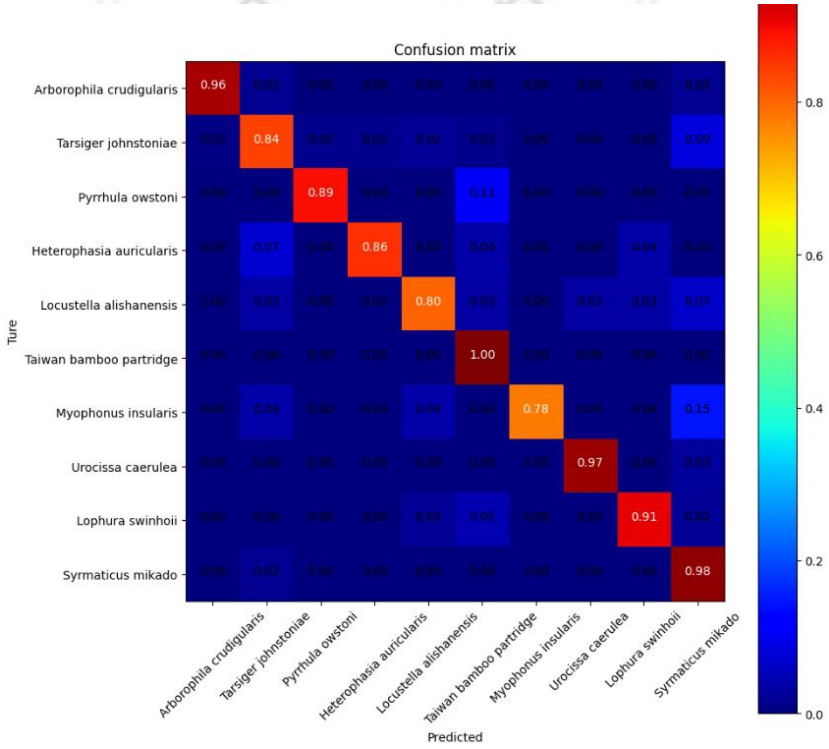


圖 3-4 模組混淆矩陣

緊接著我們要使用混淆矩陣得到的數值來計算模組的精確率(precision)、召回率(recall)及 F1-source。

(1)回應率(recall)：

回應率公式為 TP/(TP+FN)

(2)精確率(precision)：

精確率得公式為 TP/(TP+FP)，

(3)F1-度量：

F1-度量是「precision」和「recall」的調和平均數（harmonic mean），可看作是該二指標的綜合指標，能較全面地評斷模型的表現。公式如公式 3-1 下：

$$F_\beta = (1+\beta^2)\times\frac{precision\times recall}{(\beta^2\times precision)+recall}$$

公式 3- 1 F1-度量

(4)正確率

交叉驗證法(crossvalidation)，它是一種統計學上將資料樣本切割成較小子集的實證方法。主要用於模組的建立與訓練，在樣本空間中，將一部份樣本作為樣本資料來訓練模組，剩下的樣本則使用，剛剛訓練完的模組進行預測，在來看此次的預測結果，與實際結果是否相符，並得出其訓練的正確率(accuracyrate)。

以下是藉由 python 所產生的所有資料如圖 3- 5 所示，可以看到整個模組的精確率為 88.49%，回應率為 89.94%，F1 則是 89.21，正確率為 89.38%。表 3- 3 是整理的各個鳥類辨識結果的精確率與回應率還有 F1-source。

```
All_prediction 89.38 %
0       Lophura swinhoii        prediction= 98.04 %,    recall 96.15 %, f1 97.09
1       Urocissa caerulea       prediction= 93.91 %,    recall 83.72 %, f1 88.52
2       Heterophasia auricularis        prediction= 97.1 %,     recall 89.33 %, f1 93.05
3       Locustella alishanensis         prediction= 92.31 %,    recall 85.71 %, f1 88.89
4       Arborophila crudigularis        prediction= 90.74 %,    recall 80.33 %, f1 85.22
5       Taiwan bamboo partridge         prediction= 51.61 %,    recall 100.0 %, f1 68.08
6       Myophonus insularis     prediction= 100.0 %,    recall 77.78 %, f1 87.5
7       Pyrrhula owstoni        prediction= 97.4 %,     recall 97.4 %,  f1 97.4
8       Tarsiger johnstoniae    prediction= 92.86 %,    recall 90.7 %,  f1 91.77
9       Syrmaticus mikado       prediction= 70.89 %,    recall 98.25 %, f1 82.36
prediction= 88.49 recall= 89.94 F1= 89.21
```

圖 3- 5 精確率、召回率、F1 的總表

表 3- 3 鳥類辨識數值整理

中文名稱	英文名稱	precision	recall	F1
深山竹雞	Arborophila crudigularis	90.74%	80.33%	85.22
白耳畫眉	Heterophasia auricularis	97.1%	89.33%	93.5
臺灣叢樹鶯	Locustella alishanensis	92.31%	85.71%	88.89
藍腹鷴	Lophura swinhoii	98.04%	96.15%	97.09
臺灣紫嘯鶇	Myophonus insularis	100%	77.78%	87.5
台灣灰鶯	Pyrrhula owstoni	97.4%	97.4%	97.4
黑長尾雉	Syrmaticus mikado	70.89%	98.25%	82.36
臺灣竹雞	Taiwan bamboo partridge	51.61%	100%	68.8

栗背林鴝	Tarsiger johnstoniae	92.86%	90.7%	91.77
臺灣藍鵲	Urocissa caerulea	93.91%	83.72%	88.52

(四)臺灣特有種鳥類聲音辨識 APP

以下為臺灣特有種鳥類聲音辨識的頁面如圖 3- 10 所示，分別有辨識功能、圖鑑功能以及查詢功能。

圖 3- 6 臺灣特有種鳥類聲音辨識 APP

本 APP 之辨識功能，如圖 3- 11 所示，當使用者點擊開始按鈕後進入開始辨識如圖 3- 12 所示，辨識成功後可儲存至資料庫裡，如圖 3- 13 所示。

圖 3- 7 開啟鳥類聲音辨識功能

圖 3- 8 系統將所得聲音進行辨識

圖 3- 9 鳥類聲音辨識結果

本 APP 之圖鑑功能，由列表形式來表示，如圖 3- 14 所示，當點選要查詢的鳥種，會顯示鳥種的名稱、圖片、簡介，如圖 3- 15 所示。

圖 3- 10 圖鑑頁面

圖 3- 11 鳥類圖鑑查詢結果

本 APP 之查詢功能是利用資料庫，以列表的形示呈現過往的辨識紀錄，如圖 3- 16 所示，當點選其中一筆紀錄會在地圖標住當初所辨識的地點，如圖 3- 17 所示。

圖 3- 12 查詢頁面

圖 3- 13 地圖標註所辨識的地點之示意圖

3.2研究對象與實施過程

研究樣本通常是根據想要解決的問題來決定的，當問題範圍較為

狹窄或是研究樣本的工作及環境性質較為接近相似時，研究樣本的數量可以不用太多(林義倫,2010)，且根據學者拜爾以及霍茲布萊建議，在執行實務上的研究樣本數量建議以 10 到 15 名較為合適(Beyer, H. ,1997)。因此本研究徵求了 10 位 18 歲到 25 歲年輕族群為本研究的研究對象，這些族群所屬工作性質以及環境較為一致。

本研究的實施過程由使用者操作臺灣特有種鳥類辨識系統，模擬在戶外賞鳥的情況，每個使用者以體驗本研究的三個功能作為任務並在正式進行研究前，事前徵詢受測者的意願，同意讓研究者在旁邊進行觀察、訪談並說明其詳細研究進行方式。

3.3脈絡洞察法

脈絡洞察法是藉由體驗觀察、體驗訪談以及體驗分析直接進入使用者實際體驗場域中，其中體驗觀察採用非參與和參與式方法，可以了解使用者在真實情境最直接行為和使用者在本系統服務模式中的行為與動作。

而體驗訪談則是透過使用者的操作過程中藉著，訪談與觀察以了解每個環節使用者的真實狀況，並加上前面訪談以及觀察資料蒐集，結合互動模型、序列模型、工具器物模型、文化模型及實體模型五大行為塑模方式 ,說明使用者在進行特定步驟或活動的行為模式及需求。

為了掌握每個使用者共同的行為模式，因此集結全部研究對象的

行為模型整合成「彙整行為模型」，以完整呈現本系統整體使用者行為模式和需求。

3.4資料蒐集－服務體驗觀察與訪談

3.4.1 服務體驗觀察法

本研究採用非參與式現場觀察法來了解使用者使用本系統的行為模式，非參與式現場觀察法為在不破壞和影響觀察對象的原有結構和內部關係，進行使用者在操作階段的觀察與紀錄，藉此能夠獲得有關較深層的結構和關係的材料；同時，更容易貼近使用者真正行為。本研究也使用參與式觀察法，觀察每個使用流程，使用者可能會有的狀況，藉此了解使用者的需求，做為資料的蒐集方法。因上述兩種觀察都是以人類生活方式進行各個面項的研究，取此此兩種方法可以描述發生了什麼、所涉及的人或物、事發的時間和地點、發生的過程和原因等研究者所關注的問題，即回答何時（When）、在什麼地方(Where）、對哪些對象(Who）、採取哪一種或幾種方式（What）以及如何發生(How）、為什麼發生(Why）等問題。所以本研究觀察重點為活動(Activities)、環境(Environments)、互動(Interactions)、物件(Objects)、使用者(Users)，為 A.E.I.O.U 五種構面(林義倫,2010)。以下將為 A.E.I.O.U 五種構面進行說明。

1. A 活動(Activities)：在特定的活動中，人們的行為模式為何會有哪些流程。
2. E 環境(Environments)：空間使用上是個人空間還是共享空間空間的特色為何。
3. I 互動(Interactions)：在人與人或人與物件間有哪些互動行為特別、特殊的互動行為。
4. O 物件(Objects)：使用者的活動環境中有哪些物品和設備這些物品和設備是跟哪些活動相關。
5. U 使用者(Users)：使用者的價值觀及或偏見為何。

藉由 A.E.I.O.U 五種構面，本研究設計了關於使用者在使用臺灣特有種鳥類聲音辨識系統的各種問題，為表 3- 4 所示：

表 3- 4 A.E.I.O.U 五種構面

觀察構面	觀察重點
A-活動	使用者在操作系統時所遇到的問題與困難
E-環境	使用者在戶外或吵雜環境使用時的方便性?
I-互動	1.使用者對於系統回饋的動作與資訊，是否需要改善 2.辨識結果的正確率
O-物件	系統介面是否會影響使用者的操作
U-使用者	1.使用者對於系統使用的前後看法 2.使用者認為本系統可否實際應用在賞鳥...等情況

3.4.2 體驗訪談

體驗訪談有別於傳統的訪談，體驗訪談在過程會結合「參與式現場觀察法」與「體驗經驗框架」，讓使用者在操作的當下，同時接受研究者立即的詢問與觀察，讓研究者可以了解使用者的操作及動機。

(一)參與式現場觀察法：

在使用者同意的前提下，進行研究者制定的特定活動，從中觀察使用者當時的行為，並在使用者進行每個動作當下，馬上進行訪談，藉此了解每個環節使用者對於系統的真實狀況。

(二)體驗旅途框架：

在操作系統時，每個系統之服務關鍵接觸點，將其區分出來，用以了解使用者在每個接觸點中使用感受以及其重要行為。以下為本研究以使用者體驗本系統的體驗經驗框架如圖 3- 19 所示。

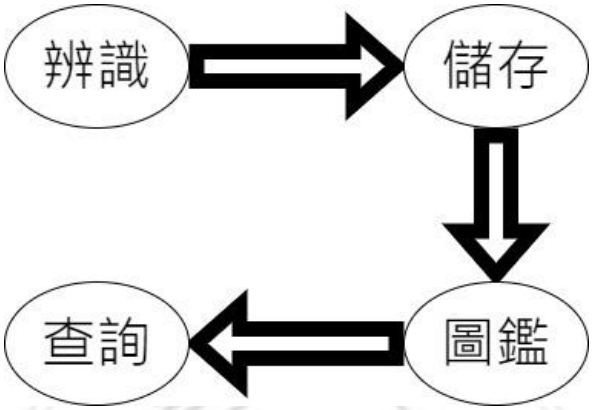


圖 3- 14 臺灣特有種鳥類辨識系統體驗經驗框架

(三)行為塑模：

在資策會提出「服務體驗工程法研究篇」和「顧客洞察者的田野手冊」中行為塑模為脈絡觀察法延續的活動，主要將訪談蒐集而來的資料，透過圖形的方式表達出來，以達到最後資料分析的目的。藉著五大行為模型，分析並瞭解使用主需求(林義倫,2010)。

五大行為模型分別有為互動模型(Flow Model)、文化模型(Cultural Model)、序列模型(Sequence Model)、工具器物模型(Artifact Model)、實體模型(Physical Model)。

(1)互動模型：

互動模型主要指使用者在執行某種任務時，會與哪些人、事、物接觸與溝通，藉此協助研究者了解使用者從事活動時，與誰溝通、如何互動等有關內容。藉由互動模型，可以將本系統使用者進行「鳥類聲音辨識」任務時，了解使用者在事件過程中，與其他人、事、物接觸與互動之過程，如圖 3- 20 所示。

圖 3- 15 臺灣特有種鳥類辨識系統互動模型

(2)序列模型：

將整個活動流程以及步驟依照先後順序彙整到序列模型中，觀察每一個階段流程，將觀察到的行為模式彙整到序列模型中，從中發現有哪些流程中環節有問題或是缺少哪些流程步驟如圖 3- 21 所示。

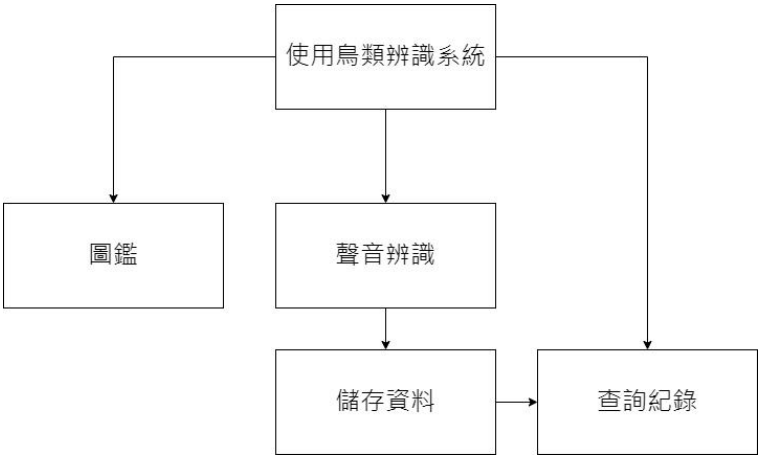


圖 3- 16 臺灣特有種鳥類辨識系統序列模型

(3)工作模型：

記錄使用者在某些特定的任務或活動中所使用的工具和物品，可透過照片或拍攝記錄當時所使用的物品並加以說明，可繪製成使用工具需求表。

(4)文化模型：

指使用者受到內、外因素影響(如:法律、產業規定、工會、情感、態度)後，所受到影響使用者的行為為何。因影響的因素眾多，且每個人所受的影響比重也不同，故文化模型較難具體化的呈現出使用者受到哪些因素而影響其行為與表現。

(5)實體模型：

實體模型主要是紀錄與完成某個任務時的相關場域，透過實體環境場域與設備擺放位置，並從實體環境下知道有哪些阻礙，再從這些阻礙下讓使用者知道有哪些可使用的工具器物或調動的工作人員等。

第 4 章　研究結果

4.1受測者基本資料

本研究以服務體驗之需求脈絡洞察的模式，針對年輕族群使用臺灣特有種鳥類辨識系統進行聲音辨識，並以此進行洞察研究。為了找出使用者對於本系統服務產生的需求以及問題，研究過程中採以參與式觀察法以及訪談法來進行，透過上述方式進行，藉此了解現有服務設計缺口，以及對於使用者會產生的困難點與需求。在研究樣本的部分，採取隨機徵求有使用過 APP 應用程式的 10 名年輕族群，以下是本研究在服務體驗過程中所蒐集到的受測者資料表如表 4- 1 所示。

表 4- 1 受測者基本資料

編號	年齡	職業	使用過動物聲音辨識 APP 的經驗	使用語音辨識 APP 的經驗
受測者一	19	學生	無	有
受測者二	21	學生	無	有
受測者三	22	學生	無	無
受測者四	19	學生	無	有
受測者五	20	學生	無	無
受測者六	20	學生	無	有

受測者七	22	學生	有	有
受測者八	23	學生	無	無
受測者九	24	學生	有	有
受測者十	22	學生	無	有

根據表 4- 1 的資料所示，本研究的受測者大部分都有接觸過語音辨識系統，但幾乎沒有使用過動物類聲音辨識。本研究讓受測者以現場使用臺灣特有種鳥類辨識服務體驗的方式並藉由現場參與式觀察與訪談，從中瞭解使用者體驗後的觀感以及所遇到的問題和不足之處，作為本研究後續的改善，以期許本研究的台灣特有種鳥類辨識系統能更貼近使用者需求。

4.2受測者服務體驗與訪談

經 10 名測試者體驗後，多次訪談分析，並以活動(Activities)、環境(Environments)、互動(Interaction)、物件(Objects)、使用者(Users)五個構面來呈現出使用者對於台灣特有種鳥類辨識系統的真實狀況與觀感。以下內容為受測者接受的問題，並以編號來區別受測者：

(一)活動(Activities)

訪談問題：使用者在操作系統時所遇到的問題與困難?

回應內容：

受測者一：在使用系統上沒有甚麼困難和問題，在很多地方好像

是怕我們按錯一樣，都不給我們按。

受測者二：操作簡單明瞭，語音辨識感覺很厲害。

受測者三：總體來說沒什麼問題，但真心覺得功能可以多一點。

受測者四：操作沒有問題，但辨識有時候會出錯。

受測者五：圖鑑功能太單調了。

受測者六：很新鮮，之前都沒有使用過動物類的語音辨識 APP。

受測者七：查詢功能我覺得很棒可以在地圖上看到我在哪個時候看到哪個種類的鳥。

受測者八：操作沒問題，但我覺得用不太到，因為我不賞鳥。

受測者九：語音辨識很酷，其他的功能覺得還好。

受測者十：操作很簡單，操作上也沒有問題。

(二)環境(Environments)

訪談問題：使用者在戶外或吵雜環境使用時的方便性?

回應內容：

受測者一：如果我是賞鳥者，只靠手機就能知道是什麼鳥，我覺得很方便，在戶外的辨識也能正常完成。

受測者二：辨識在戶外沒有問題但在吵雜的環境就會被干擾，有時候會辨識錯誤，不過還是覺得蠻方便的畢竟只需要手機就可以了。

受測者三：很方便但吵雜的環境辨識常常錯誤。

受測者四：覺得不是很方便，太吵雜的環境就無法使用。

受測者五：戶外可以使用蠻方便的。

受測者六：我覺得蠻方便的，但如果要到深山裡，可能沒有那麼方便，因為可能會收不到網路，也要考慮山里的其他聲音。

受測者七：覺得還蠻方便的，吵雜的環境只要聲音不太大聲也還是可以辨識。

受測者八：手機就能操作很方便。

受測者九：不方便，環境聲音太大就會辨識錯誤。

受測者十：帶入賞鳥者的腳色感覺很方便，什麼都不用帶只要帶手機。

(三)互動(Interaction)

訪談問題 1：使用者對於系統回饋的動作與資訊，有哪部分需要改善？

回應內容：

受測者一：各方面都蠻好的。

受測者二：辨識跟查詢沒問題，但是圖鑑可以豐富一點

受測者三：查詢功能或許可以看到更多人發現哪隻鳥並呈現在同

一張地圖上

受測者四：感覺還好，沒甚麼需要改善的

受測者五：圖鑑功能太單調了

受測者六：查詢功能可以用其的方式搜尋，例：鳥類、日期..等

受測者七：圖鑑功能可能可以加聲音或是分布圖甚麼的。

受測者八：圖鑑太單薄了，其他的還可以。

受測者九：可以讓查詢功能的地圖顯示多一點標記。

受測者十：整體還不錯，但是圖鑑要加強。

訪談問題 2：辨識結果的的正確次數?

回應內容：

將回應結果整理如表 4- 2 所示。

表 4- 2 辨識正確的次數

編號	測試次數	正確次數
受測者一	10	8
受測者二	10	6
受測者三	10	9
受測者四	10	7
受測者五	10	8

受測者六	10	8
受測者七	10	8
受測者八	10	7
受測者九	10	7
受測者十	10	9

(四)物件(Objects)

訪談問題：系統介面是否會影響使用者的操作?

回應內容：

全體受測者的回應大致相同，操作介面非常友善，介面美觀可以再多有琢磨，但不會影響操作。

(五)使用者(Users)

訪談問題 1：使用者對於系統使用前後的看法?

回應內容：

受測者一：使用前覺得有點好奇，使用後覺得還不錯，但還是可以加強，例如：畫面美觀方面。

受測者二：使用前有點期待，覺得語音辨識厲害，使用後也覺得還可以。

受測者三：使用前覺得很新鮮，也很期待，但使用後有點失望，功能有點少。

受測者四：使用前沒什麼感覺，使用後覺得語音辨識還不錯，美術可能要再加強。

受測者五：使用前沒什麼感覺，使用後還可以，但圖鑑可以加強。

受測者六：使用前很期待沒有看過這種類型的 APP，使用後覺得辨識的功能還不錯，但其他功能就還好，介面耶需要做漂亮一點。

受測者七：使用前不覺得怎麼樣，使用後有點改觀，雖然介面有點陽春。

受測者八：使用前覺得應該很厲害，使用後較為失望，功能偏少，且在深山應該用不了。

受測者九：使用前沒有很期待，使用後覺得語音辨識蠻厲害的。

受測者十：使用前沒有很看好，使用後覺得還不錯，不過有些功能要加強，查詢和圖鑑。

訪談問題 2：使用者認為本系統若實際應用在賞鳥上，是否可行？

回應內容：

受測者一：覺得可行，也感覺可以辨識其他動物的聲音。

受測者二：應該可以吧，不過各個功能可能要再加強。

受測者三：應該不行，聲音太容易被環境所影響。

受測者四：不行，先不說環境問題，在山上可能還收不到訊號。

受測者五：可以，不過功能要再加強，環境問題也需要克服。

受測者六：雖然感覺還不錯，但環境因素影響太大，感覺不太可行。

受測者七：應該可以吧，不過部分功能跟介面上可能要調整。

受測者八：感覺可以，但感覺比起鳥我覺得其他的動物會更加有趣。

受測者九：可以，但感覺應用在別的地方也很有趣。

受測者十：感覺可以，很方便，不過介面可能要改一下。

蒐集完受測者的回答後，將問題整理並彙整，以下是針對A.E.L.O.U 五構面所提出的問題，以及受測者的回答所整理的問題與需求，由表 4- 3 所示：

表 4- 3 A.E.I.O.U 五構面問題彙整

A.E.L.O.U	問題敘述	需求點
活動(Activities)	圖鑑功能單薄。	圖鑑功能上增加其他資訊，例如：叫聲、分布位置..等。
	介面資訊單調。	提高系統內容豐富度。
環境(Environments)	吵雜環境導致辨識功能辨識錯誤或無法辨識。	聲音容易受環境影響。
互動(Interaction)	圖鑑跟查詢功能的反饋太少。	查詢功能可以增加其他人的賞鳥紀錄，以及在地圖上顯示過往的賞鳥紀錄。
	辨識正確率為 77%。	提高辨識率。
物件(Objects)	畫面不好看。	加強畫面美工。
使用者(Users)	應用在其他情況? 使用後的感覺很不錯。	或許可應用在其他方面，對語音辨識感到新奇。

4.3五大行為模型

透過服務體驗脈絡洞察法中的五大行為模型，透過非參與式以及參與式觀察法，將過程中觀察以及訪談並蒐集受測者在期間內、外在行為資料加以統整，了解受測者真實的需求，以下是透過五大行為模型，所得出的問題與需求，由表 4-4 所示：

表 4-4 五大行為模型問題彙整

模型	問題、狀況&需求
互動	圖鑑功能以及查詢功能，反饋資訊不豐富。
序列	系統介面功能單調，需要增加系統的豐富度。
文化	在使用語音辨識上並無困難，且覺得新奇，但因大部分人不常去賞鳥，會減少使用者的興趣。
工具	工具為手機，受網路限制。
實體	山上的環境對於系統的使用有疑慮，考慮因素為網路，聲音吵雜程度。

第 5 章　結論與建議

隨著國內賞鳥者的需求與日劇增，加上科技的日新月異，人工智慧跟深度學習等技術也日漸成熟，本研究使用了卷積神經網絡開發出臺灣特有種鳥類辨識系統，輔以服務體驗工程法，在活動、環境、互動、物件、使用者五大不同的構面再加上觀察訪談法與五大模型彙整，分析出臺灣特有種鳥類辨識系統的潛在問題與需求。其中又以系統的辨識成功率為重中之重，根據上述的實驗結果可以看出本系統測試的成功率在 60%~90%，雖然正確率達一半以上，但仍有可繼續提升的空間，期望讓正確率能有穩定的發揮，而在系統介面的美工和功能的豐富性上，也仍有可以進一步改善的空間，最後本研究目前仍受限於吵雜環境下之聲音的正確辨識率，是為未來可以進一步發展的課題。目前市面上的賞鳥 APP 的語音辨識功能尚未成熟，本研究受此啟發，結合深度學習開發出台灣特有種鳥類聲音辨識系統，期望能滿足市場上的需求，也提供給需求者另一種選擇。本研究受限於時間及可行性因素，研究樣本數僅有 10 位受測者，因此實驗的深度以及服務需求程度尚嫌不足。期許後續研究者可以增加研究樣本，加深訪談研究內容，並探討更多不同使用者的不同需求，讓鳥類語音辨識系統能夠更加完善。

參考文獻

王小川（2004）。語音訊號處理。全華圖書，台北市

王熙哲、林曉琪（2010）。應用服務體驗工程方法於銀髮族家事服務系統設計,產業與管理論壇，36-53。

王熙哲、許惠詮（2012）。應用服務體驗工程方法於銀髮族家事服務系統設計。龍華科技大學資訊管理學系碩士論文。

林承億（2017）。獨立成分分析法應用於青蛙聲音辨識。碩士論文，國立臺灣科技大學

林義倫、吳明珊、陳以玲、張呈瑋、黃宣龍、蕭淑玲（2010）。顧客洞察者的田野手冊,台北:經濟部技術處資策會創新應用服務研究所

服務體驗工程方法流程圖（2009）。策會創新應用服務研究所

陳家安（2019）。以深度學習方法實作簡單語音辨識模型,清華大學

黃俊卿（2017）。基於物聯網之環境聲音辨識偵測平臺。 碩士論文，東華大學

資訊工業策進會（2008）。服務體驗工程方法指引-研究篇,臺北資策會創新應用服務研究所。

庄野逸, 永原健一, 岡田真人,&福島邦彦.（1997）. ネオコグニトロンの実用化: 大規模データベースによる評価. 電子情報通信

学会技術研究報告. NC, ニューロコンピューティング, 97(116), 65-71.

Avutu, S. R., Bhatia, D., & Reddy, B. V. (2017, January). Voice control module for low cost local-map navigation based intelligent wheelchair. In 2017 IEEE 7th International Advance Computing Conference (IACC) (pp. 609-613). IEEE.

Anderson, S. E., Dave, A. S., & Margoliash, D. (1996). Template‐based automatic recognition of birdsong syllables from continuous recordings. The Journal of the Acoustical Society of America, 100(2), 1209-1219.

Bajaj, V. ,Demir, F., & Sengur, A. (2020). Convolutional neural networks based efficient approach for classification of lung diseases. Health information science and systems, 8(1), 1-8.

Bongo, L. A. ,Grønnesby, M., Holsbø, E., Melbye, H., &,Solis, J. C. A. (2017). Feature extraction for machine learning based crackle detection in lung sounds from a health survey. arXiv preprint arXiv:1706.00005.

Brunot, A.,Bottou, L., Cortes, C.,Denker, J.,Jackel, L.,LeCun, Y.& Vapnik, V. ... (1995, November). Comparison of learning algorithms for

handwritten digit recognition. In International conference on artificial neural networks (Vol. 60, No. 1, pp. 53-60).

Beyer, H. & Holtzblatt, K. (1997). Contextual design: defining customer-centered systems. Elsevier.

Card, H. C. , & McIlraith, A. L. (1997, May). Bird song identification using artificial neural networks and statistical analysis. In CCECE'97. Canadian Conference on Electrical and Computer Engineering. Engineering Innovation: Voyage of Discovery. Conference Proceedings (Vol. 1, pp. 63-66). IEEE.

Card, H. C. , & McIlraith, A. L. (1997). Birdsong recognition using backpropagation and multivariate statistics. IEEE Transactions on Signal Processing, 45(11), 2740-2748.

Chellapilla, K., Puri, S., & Simard, P. (2006, October). High performance convolutional neural networks for document processing. In Tenth international workshop on frontiers in handwriting recognition. Suvisoft.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE transactions on acoustics,

speech, and signal processing, 28(4), 357-366.

Erić, T., Ivanović, S., Milivojša, S., Matić, M., & Smiljković, N. (2017, September). Voice control for smart home automation: Evaluation of approaches and possible architectures. In 2017 IEEE 7th International Conference on Consumer Electronics-Berlin (ICCE-Berlin) (pp. 140-142). IEEE.

Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In icml (Vol. 96, pp. 148-156).

Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). Spoken language processing: A guide to theory, algorithm, and system development. Prentice hall PTR.

Hubel, D. H., & Wiesel, T. N. (1960). Receptive fields of optic nerve fibres in the spider monkey. The Journal of physiology, 154(3), 572.

Hinton, G. E. ,Krizhevsky, A., &Sutskever, I.(2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

He, K.,Sun, J.,Zou, J.& Zhang, X.(2015). Accelerating very deep convolutional networks for classification and detection. IEEE

transactions on pattern analysis and machine intelligence, 38(10), 1943-1955.

He, K., Ren, S., Sun, J. & Zhang, X.(2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Kuan, K. L. (2010). A framework for automated heart and lung sound analysis using a mobile telemedicine platform (Doctoral dissertation, Massachusetts Institute of Technology).

Lang, K. J.,Hinton, G., Hanazawa, T., Shikano, K., & Waibel, A.(1989). Phoneme recognition using time-delay neural networks. IEEE transactions on acoustics, speech, and signal processing, 37(3), 328-339.

Wang, G. ,Ma, Y., Xu, X., Yu, Q., Zhang, Y., Li, Y., ,& Zhao, J. (2019, October). LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm. In 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1-4). IEEE.

Yan, Z., & Zhao, S. (2016, August). A usable authentication system based on personal voice challenge. In 2016 International Conference on

Advanced Cloud and Big Data (CBD) (pp. 194-199). IEEE.

附錄

程式碼

裁切聲音程式碼：

```
mp3 = AudioSegment.from_file(
    '/content/drive/My Drive/2020MAI/voice/文件/鳥類/黑長尾雉/{}'
    .format(each), "mp3") # 開啟mp3檔案
  # # mp3[17*1000+500:].export(filename[0], format="mp3") #
size = 5000 # 切割的毫秒數 10s=10000

chunks = make_chunks(mp3, size) # 將檔案切割為15s一塊

for i, chunk in enumerate(chunks):

        chunk_name = "{}{}.wav".format(each.split(".")[0],i)
        print(chunk_name)
        chunk.export(
                '/content/drive/My Drive/2020MAI/voice/B文件/{}'
                .format(chunk_name), format="wav")
```

提取聲音特徵程式碼：

```
file_name = os.path.join(data_dir,str(row.ID)+'.wav')
print(file_name)
try:
        X, sample_rate = librosa.load(file_name, res_type='kaiser_fast')
        mfccs = np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40).T, axis = 0)

except Exception as e:
        print("Error encountered while parsing file: ", e)
        return None, None
feature = mfccs
data_id = row.ID

return feature
```

資料歸一化：

```
import numpy as np
X = np.array(df.feature.tolist()).astype(np.float32)
y = np.array(df.Class.tolist())
y

mfcc_mean = np.mean(X, axis=0)
mfcc_std = np.std(X, axis=0)

X = [(X - mfcc_mean) / (mfcc_std + 1e-14) for X in X]
X = np.array(X).astype(np.float32)

X
```

CNN 模組：

```
def categorical_classifier():
    model = Sequential()

    model.add(Conv1D(100, 10, input_shape=(40,1), activation = 'relu',padding='same') )
    model.add(Conv1D(120, 10,input_shape=(40,1) ,activation = 'relu',padding='same') )
    model.add(MaxPooling1D(pool_size =2,padding='same'))
    model.add(Conv1D(160, 10,input_shape=(40,1), activation = 'relu',padding='same') )
    model.add(Conv1D(320, 10, activation = 'relu',padding='same') )
    model.add(Dropout(0.1))
    model.add(MaxPooling1D(pool_size =2,padding='same'))
    model.add(Dropout(0.3))
    model.add(Conv1D(200, 10, activation = 'relu',padding='same') )
    model.add(Conv1D(100, 10, activation = 'relu',padding='same') )
    model.add(MaxPooling1D(pool_size =2,padding='same'))
    model.add(Dropout(0.2))

    model.add(GlobalAveragePooling1D())
    model.add(Dropout(0.5))
    model.add(Flatten())

    model.add(Dense(units=400, activation = 'relu'))
    model.add(Dense(units=200, activation = 'relu'))
    model.add(Dense(units=100, activation = 'relu'))

    model.add(Dense(units=10, activation = 'Softmax'))

    model.compile(loss='categorical_crossentropy', metrics=['accuracy'], optimizer='adam')
    return model
```