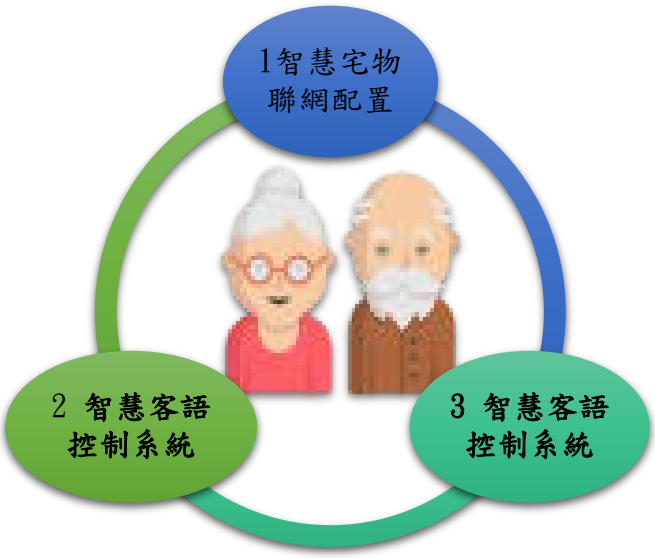
108 年度客家知識體系發展補助計畫

結案報告書



計畫名稱：運用 AI 技術於臺灣客語語音辨識之研究 － 以建置客家智慧宅為例

Research and Implementation for Hakka Speech Recognition Based on AI Techniques -An Example on Hakka Smart Home

主　持　人：黃豐隆博士　職稱：副教授

機構及科系所：國立聯合大學資工系

E-MAIL ：ncat70@gmail.com

民國 109 年 9 月 10 日

本計劃得以完成

感謝

客家委員會的補助、支持

計畫摘要：

為因應當前臺灣社會老人化趨勢，近年來政府大力推動長照 2.0。本計畫的研究主題主要在於，建置完成臺灣客語智慧的居家應用系統-「**臺灣客語智慧宅**」。目前臺灣的客語語言亦流失的嚴重問題，本計劃運用人工智慧(A.I.)技術作為客語語音辨識的基礎，結合當前熱門最新 ICT 相關技術，含物聯網(Internet of Things, IoT)、大數據(Big data)與雲端計算(cloud computing)資料庫，期望為客家族群眾多長輩們打造一個具智慧、便利與安全樂利的居家環境。

目前臺灣學術界與民間業界投入研發智慧宅裡的許多應用，常見的功能多在手機 APP 端結合居家設備的相關控制。本計劃提出一項創新的智慧宅環境構想，建置臺灣客家語言環境，使用者可以經由客語的語音來操作控制居家環境裡的各項設備，如:大門開關與電燈開關等，可為長者處理日常簡易的事務。

主持人多年來從事客語語言處理與歷史文化的相關研究，可以運用既有專業強項來結合目前社會的需求。本計劃研究核心在於客語語音辨識的智慧系統，可以建構一個全臺唯一結合物聯網與大數據的「**臺灣客家智慧宅**」，逐漸發展具客家特色的亮點成果。

關鍵字: 人工智慧、深度學習，客語語音辨識，智慧宅，物聯網。

Currently, Taiwan's academic group and industry have invested in the research and development of smart homes, in which the common functions are combined with the control of home devices in the mobile APP. Our project proposes an innovative smart home environment concept, and implement a Taiwanese Hakka language environment which is a first smart home with Hakka speech recognition. Users can control various devices in the home environment through the voice of the Hakka language, including: gate access control, 3C home appliances, and robotic companion robot, accompany the Hakka elderly to talk, or assist them to handle simple matters in the living environment.

The project host has been engaged in the study of Hakka language processing, history and culture for many years, and can use the existing professional strengths to combine the needs of the current society. The core of this project is the intelligent system of speech recognition of the Hakka language. It can construct a “Taiwan Hakka Smart House” that combines the Internet of Things and big data with the whole Taiwan, and gradually develops the highlights with Hakka characteristics.

Keywords: AI, Deeping Learning, Hakka Speech Recognition, Smart home, Internet of Things.

壹、計劃概述

隨著年紀的增加、父母與長輩的記憶力和行動能力會慢慢地退化或老化，許多說明操作或是多功能介面對年長者來說都是非常吃力的行為。臺灣社會已出現老年化與少子化趨勢，年長者的居家生活與照護已成為一項社會的重要議題。

近年來，隨著資訊技術與居家安全觀念的蓬勃發展，智慧宅(Smart home)逐漸成形、落實。在智慧宅環境中，以網路連線使得居家的相關物品相連技術，各項無線感測器材、網路傳送生活環境裡的相關資訊，整合於雲端平台，提供便利、舒適與智慧功能。例如，整合大門的電子鎖、燈光、電動窗簾和影像攝影機等，使用者可以經由操作手機上 APP，不在家裡也能了解居家的即時現況，主人返回到家後可以享受無比便利舒適的生活環境 !屋主可使用手機 APP 遠端為家人開門或上鎖關門解除保全時連動打開空調、燈光、電視、音樂…等各種家電。如有任何狀況，可立刻聯絡家屬或相關單位，即時協助處理，確保屋主享受安心舒適的智慧生活。

1.1 智慧宅簡介

因應臺灣社會老人化的趨勢，政府正大力推行長照 2.0 政策。本計畫的研究議題，係建置完成一個臺灣客語智慧的居家應用系統-「**臺灣客語智慧宅**」。目前臺灣的客語語言亦流失的嚴重問題，本計劃運用人工智慧(A.I.)技術作為客語語音辨識的基礎，結合當前熱門最新 ICT 相關技術，含物聯網(Internet of Things, IoT)、大數據(Big data)與雲端計算(cloud computing)，希望可以為客家族群眾多長輩們打造一個具智慧、便利與安全樂利的居家環境。

目前臺灣學術界與民間業界已投入研發智慧宅裡的生活應用，常見的功能多限於手機 APP 端結合居家設備的相關控制。本計劃提出一項創新的智慧宅環境構想，建置臺灣客家語言環境，使用者可以經由客語的語音來操作控制居家環境裡的各項設備，包含:大門門禁、3C 家電等設備，進一步可以指揮聽得懂客語的居家陪伴機器人 (Robotics)，陪伴老人談天，或者協助客家長者處理簡易的事務。

「**科技始於人性**」，本計畫設計理念是從客家長輩的角度出發，聚焦於客家長者老人生活之所需。可以透過人工智慧、物聯網、雲端運算、機器學習與大數據等技術，來提升老人的居家照護與居家生活品質。主持人所屬的研究單位-聯合大學資工系，已有一完整的空間，面積約有 110 平方公尺(約有 30 坪空間)，並已初步完成基本配置，本計劃構想的場域與功能可以在這裡完成實踐。

由於目前客家語言流失十分嚴重，正面臨消失的可能，本計劃的成果有助於推展使用臺灣客語，進而促進客家文化的傳承與發揚。本計劃中的語音包含中文與四縣腔客語，整個研究架構未來可以延伸至其它腔調的客語，甚至臺灣的閩南

語與原住民語言，擴充性、實用性與應用性高。本計畫所包含的重要技術，如圖 1 所示。

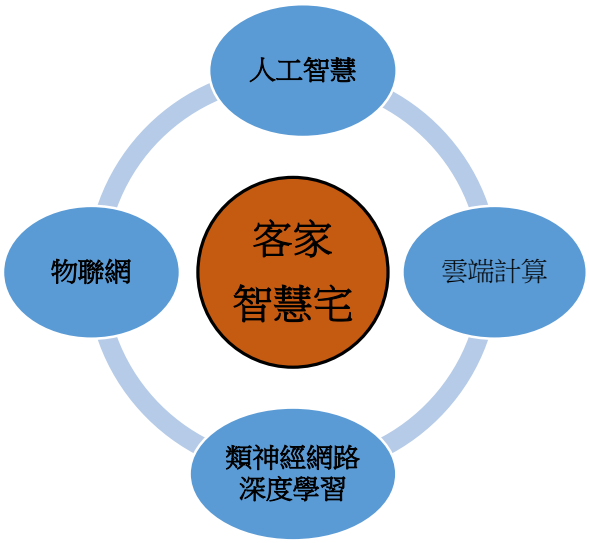


圖 1：本計畫客家智慧宅系統相關技術。

人工智慧與深度學習簡介

人工智慧(Artificial Intelligence, AI)，發展可分為三個階段，起步期、專家系統期、深度學習期。起步期主要透過簡單的邏輯來判斷；到了 1980 年，開始利用知識模型，建立專家系統(Expert Systems)，運用在特定領域中，後續則往統計模型的方法去模擬人類決策。2006 年，多倫多大學教授 Geoffrey Hinton 提出「深度學習」(Deep Learning) 的類神經網路(Neural Network)，才正式爆發新一次的人工智慧浪潮。事實上，在早期就有類神經網路的學問，但是受限於當時計算機硬體的速度，尚無大進展。深度學習的類神經網路興起後，伴隨計算機運算力指數性提升，尤其是 GPU(圖形處理單元)，價格亦降低至一般可負擔的水平，重新帶來人工智慧契機。

人工智慧的運用非常廣泛，其中語音識別(speech Recognition)和圖像識別為最多的兩個主軸。語音辨識，可辨別語音的內容與其聲調等，再進行自然語言處理後，分析其語音的語意與情感等，將用於操作聊天機器人與翻譯機器人等。

物聯網發展簡介

物聯網(Internet of Things，IoT)，是指經由網際網路、電信網路，使所有獨立功能的物體彼此互聯互通的環境，簡而言之，生活週遭的設備可以物物相聯，並加以應用。IoT 一般為無線網路環境，可以將設備聯結在一起，對各種裝置作控制與操作，也可以遙控家庭中的各種 3C 家電裝置。採集

各項資料後，可聚集成為大數據資料，作進一步的分析與加值應用，提供生活環境所需的資訊，進而提升生活品質。

IoT 使分散的資訊得以整合，應用範圍十分廣泛，包括：居家住宅、運輸和物流領域、工業製造、健康醫療領域範圍與辦公室等智慧型環境，依國內外研究機構預測，全世界的產值高達數十兆美元。

1.2 臺灣客語現況與語言特性

客家委員會的 2004 年客語使用狀況調查報告明確指出，阻礙種族文化傳承的最大原因是「語言的失傳」，因此，保存客家文化的首要工作，就是推廣客語的學習。客家委員會在 2010 年至 2011 年的全臺調查結果，臺灣客家族群占 18%，約有 420 萬人，人口數量僅次於閩南族群。再依 2016 年的調查，客家人者占 16%，推估人口總數減為 382 萬人，還是臺灣第二大主要族群。另外，根據 2013 年客語使用狀況調查，目前臺灣的客家人口中，能與其他人溝通的腔調，以「四縣腔」比例為最高，約為 56%，其次為「海陸腔」42%，四縣腔與海陸腔語言為臺灣客語之二大人口。

長久以來, 臺灣民眾一般使用中文(國語)與閩南語(福老話)為主要語言，很高比例的客家人口因生活環境因而可聽得懂閩南語，成為「福佬客」，指:會講閩南話的客家人。近幾十年來整個臺灣社會大環境的變化，導致客語與文化面臨大量流失的問題，客語斷層的危機愈來愈嚴重了。近年來，政府開始意識到保護客家文化和客家語的重要性，行政院已制定《客家基本法》，規定客家語為大眾交通工具播報用語言之一，然而語言傳承與發揚工作已刻不容緩了。

四縣腔、海陸腔客家腔聲調

表1: 四縣與海陸腔客語聲調與調形符號。

R1. 由兩個陰平(調號↙)字構成的字彙，讀時前字變調讀**陽平**(∨) 陰平 + 陰平 ➔ **陽平** + 陰平 **例：** 「新衫」 sin↙ sam↙ ➔ sin∨ sam↙ 「買新衫」 mai↙ sin↙ sam↙ ➔ mai∨ sin∨ sam↙
R2. 陰平與陰去（ ）構成的詞彙，讀時前字變調讀**陽平**（∨） 陰平 + 陰去 ➔ **陽平** + 陰平 **例：** 「針線」 ziim↙ sien ➔ ziim∨ sien 「拿針線」 na↙ ziim↙ sien ➔ na∨ ziim∨ sien
R3. 陰平字與陽入字構成的詞彙，讀音時前字變調讀**陽平**（∨） 陰平+ 陽入➔ 陽入 + 陰平 **例：** 「音樂」 im↙ ngok˚ ➔ im∨ ngok˚ 「聽音樂」 tang↙ im↙ ngok˚ ➔ tang∨ im∨ ngok˚

在語音聲調(Tone)方面，國語有四種，閩南語八種，海陸腔客家話的聲調則有七個聲調(Tone)，即:陰平(調號1)、陰上(調號2)、陰去(調號3)、陰入(調號4)、陽平(調號5)、陽去(調號7)與陽入(調號8)。表1 是四縣與海陸腔語音中七個聲調調名與調形符號表示，可以看出，相較於四縣腔六種聲調，海陸腔語音多一種「陽去」，共有七個聲調。**四縣腔之連音變調規則(Tone Sandhi):**四縣客語的變調規則可規納出三個情況，如表2所示:

表2: 四縣客語的變調規則

調類	陰平	陽平	上聲	陰去	陰入 (b, d, g)	陽入 (b, d, g)	陽去
四縣調號	24	11	31	55	21(↖)	5(高入)	
四縣調形	fu↗	fu∨	fu↖	fu	fug•	fug˚	
例字	夫	扶	虎	富	福	服	
國語(近似)	2 聲↗	3 聲∨	4 聲↖	1 聲			
	↕	↕	↕	↕	↕	↕	
海陸調號	31	55	24	11	5	21(↖)	33
海陸調形	fu↖	fu	fu↗	fu∨	fug˚	fug•	fu⊢
例字	夫	扶	虎	富	福	服	護
國語(近似)	4 聲↖	1 聲	2 聲↗	3 聲∨			

註: 1)表中最右邊二種(低入調、高入調)為入聲韻。
2)海陸腔較四縣腔多一種「陽去」之語音，故有 7 種聲調。

本計劃研究的客家語音辨識的語言，以四縣腔為主，未來完成的資訊技術可延伸至其它客語腔調，甚至臺灣其它的語言，如閩南語或原住民語言等。

1.3 本計劃研究議題

主持人多年來從事客語語言處理與歷史文化的相關研究，可以運用既有專業強項來結合目前社會的需求。本計劃研究核心在於客語語音辨識的智慧系統，可以建構一個全臺唯一結合物聯網與大數據的「**臺灣客家智慧宅**」，逐漸發展具客家特色的亮點成果。本計劃研究的客家語音辨識的語言以四縣腔為主，資訊技術可延伸至其它腔調，或臺灣其它的本土語言，如閩南語或原住民語言，使用的對象亦不限於客家的長者，一般民眾均可輕易使用智慧宅的相關功能。

本計劃包含三項，如下所示。

1. 智慧宅物聯網配置
2. 人工智慧語音辨識技術
3. 智慧客語控制系統

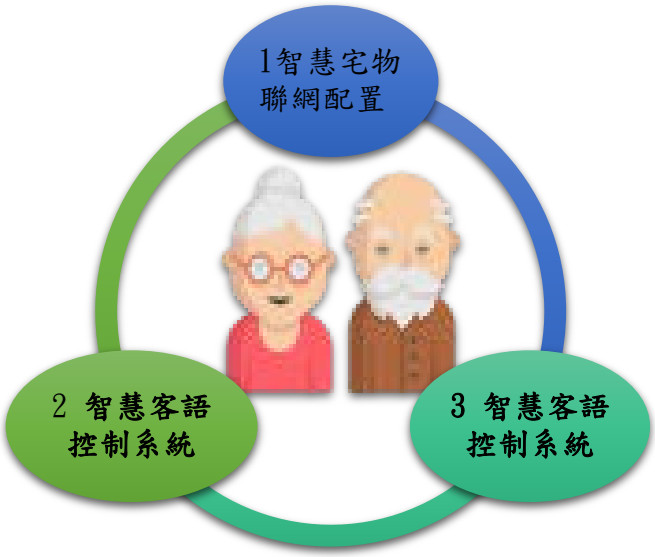


圖 2：本計劃智慧宅的主要功能。

貳、研究方法

2.1 客語語言模型處理

語音辨識的目的在於，以電腦自動將人類的自然語音內容轉換為相應的正確文字，一般常用於語音撥號、語音導航、室內裝置控制、語音文件檢索。如上所述，本計劃處理的客語屬於具有聲調的語言，我們以客語四縣腔為主。經由人工智慧技術與深度學習方法，將客語語音轉換成正確詞彙，以利後續物聯網系統的操作與控制。

語音辨識技術所涉及的領域包括：人工智慧、訊號處理、統計式語言模型、機率論和資訊理論、發聲機理和聽覺機理等。首先，說明統計式語言模型的原理如下。

語言模型(Language Model)

語言模型(Language Model)是斷詞演算法中，用來選擇出最佳斷詞詞序列的重要元件。目前語言模型的設計，可分為：1)以文法取向 2)以統計取向，兩種設計方法。

1)文法取向：

這類語言模型的作法，是根據文法及語意規則來制定條件及規則。再經由文句剖析器(Parser)對句子做剖析，得到文法結構樹、建置成文法取向的語言模型。其優點是易於擷取詞彙或文句中的意義，可用於語言處理，如：機器翻譯、斷詞處理…等。但缺點是應用在語音辨識時，無法處理語法不合句法規則的句子。這類語言模型能應用的範圍有限，因此目前的語言模型，都還是以統計取向為主。

2)統計取向：

這類語言模型的作法，是透過大量的訓練語料，統計出該語料的機率分佈模型。若使用在多個可能斷詞序列的選擇上，是將長度為 N 的詞串(N-Gram)機率，從語言模型中查出，並且找出可組成最大機率的詞串序列。而應用於語音辨識上，該方法的優點是一定能辨識出一個最可能的句子，對於語法錯誤的容錯率高，但缺點是必須收集大量的訓練語料進行訓練。

如圖 3 所示，一個語言模型被應用在決定最佳斷詞序列(找出一個機率最大的目標詞串 W)的例子。原始詞串 S 中有多個可能詞串序列，透過語言模型計算機率後，將得到最佳的目標詞串 W ，註記成 $w_1, w_2, w_3, \ldots, w_m$ 或是 w_1^m ，$w_i \in V$ ，代表 m 個詞所組成的詞串，V 則是詞典為所有詞的集合。S 可以為任何所有可能的詞串。

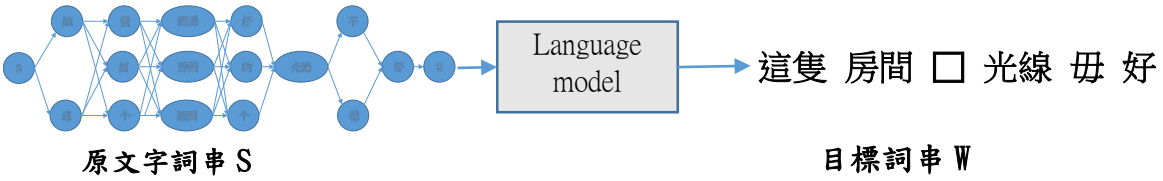


圖 3：語言模型在客語斷詞處理的應用

一個目標詞串 $W = w_1 \ldots w_m$ 的機率 $P(W)$ ，W 為 $w_1, w_2, \cdots, w_m$ 欲轉換的字串 S ，表示如下：

$$P(w_1 w_2 w_3 \ldots w_m) \qquad (1)$$

上式(1)可寫成：

$$\begin{aligned} P(w_1^m) &= P(w_1)P(w_2 \mid w_1^1)P(w_3 \mid w_1^2)\ldots P(w_m \mid w_1^{m-1}) \\ &= P(w_1)\prod_{i=2}^{m} P(w_i \mid w_1^{i-1}) \end{aligned} \qquad (2)$$

其中 $P(w_i \mid w_i^{i-1})$ 是詞 w_i 在特定歷史詞串 $h_i = w_1, w_2, w_3, w_4, \cdots, w_{i-1}$ 的情況下，出現的條件機率。實際上在建立語言模型時，並不會把所有可能的參數 $P(w_i \mid w_1 w_2 w_3 \ldots w_{i-1})$ 都儲存起來。

因為針對長度 m ，歷史詞串長度為 $n-1$ 時，所有可能的組合個數為 $|V|^n$ 。這樣的情況下，即使詞典所收錄的詞彙量不大，但只要詞串長度稍長，參數就會有驚人的成長，因此必須對參數加以簡化。譬如歷史詞串長度 $n-1=0$ 時，可以表示如下：

$$P(w_1^m) = \prod_{i=1}^{m} P(w_i) \qquad \text{(uni-gram)} \qquad (3)$$

歷史詞串長度 $n-1=1$ 時，可以表示如下：

$$P(w_1^m) = P(w_1)\prod_{i=2}^{m} P(w_i \mid w_{i-1}) \qquad \text{(bigram)} \qquad (4)$$

歷史長度$n-1=2$時，可以表示如下：

$$P(w_1^m) = P(w_1)P(w_2 \mid w_1)\prod_{i=3}^{m} P(w_i \mid w_{i-2}^{i-1}) \quad \text{(trigram)} \qquad (5)$$

通式則可以表示如下：

$$P(w_1^m) = \prod_{i=1}^{m} P(w_i \mid w_{i-n+1}^{i-1}) \quad \text{(ngram)} \qquad (6)$$

從統計觀點估測$P(w_i \mid w_{i-n+1}, w_{i-n+2}, ..., w_{i-1})$的方式，是根據最大相似度估測法(Maximum likelihood estimation, MLE)，得到下式:

$$P(w_i \mid w_{i-n+1}, w_{i-n+2}, ...w_{i-1}) = \frac{C(w_{i-n+1}, w_{i-n+2}, ..., w_i)}{C(w_{i-n+1}, w_{i-n+2}, ..., w_{i-1})} \qquad (7)$$

其中$C(\bullet)$表示詞串出現的次數。至於，因資料稀疏(Data Sparsity)問題，將採用平滑方法解決之，參見下一小節進一步說明。

語言模型評量(Evaluation)

本計畫將使用交叉熵(Cross Entropy)和混淆度(Perplexity)，作為評量語言模型效能的工具，並透過實驗觀察出最佳的語言模型的平滑方法(smoothing method)。交叉熵和混淆度是相當重要且通用的標準，混淆度是根據資訊理論(Information Theory)而來。對一組測試資料T，其中$e_1, e_2, \cdots, e_m$為m個測試事件，則測試語句T的機率可以表示如下:

$$P(T) = \prod_{i=1}^{m} P(e_i) \qquad (8)$$

其中$P(e_i)$為每個 events 的機率值，在測試資料中$H(T)$可視為這m個 events 需要編碼的長度位元數，表示如下:

$$H(T) = -\sum_{i=1}^{m} P(e_i)\log_2 P(e_i) \qquad (9)$$

$$PP(T) = 2^{H(T)} \qquad (10)$$

整體而言，較低的$H(T)$可以推導出較低的$PP(T)$，意即擁有較低$PP(T)$的語言模型，會有較好的性能。因此混淆度可以解釋成語言模型估測一個詞串後面平均可能的可接詞數。混淆度低，表示一個詞串後面有較少的選擇，辨認時就愈能找到正確的答案。

另外，交叉熵CH亦是一種量測語言模型的方式，若是語言模型能夠精準的預測出接下來的 events，則交叉熵勢必較低。在一般的情況下，$CH >= H$，H表示使用相同的模型進行訓練、測試。實際上，我們對測試模型的機率分佈並不清楚，所以必須依靠訓練模型M進行預估。M的交叉熵表示如下:

$$CH(P, M) = -\sum P(e)\log_2 M(e) \qquad (11)$$

根據 Shannon-McMillan-Breiman theorem，上式可以化簡為:

$$CH(P, M) = \lim_{n\to\infty} \frac{1}{n}\log_2 M(w_1 w_2 w_3 ... w_n) \qquad (12)$$

交叉熵 CH 是熵 $H(p)$ 的上限值，換句話說，$H(P) <= CH(P,M)$ ，其意義指的是可以用訓練模型 M 來估計每個 events 的機率。

效能評量方法:

本計畫的實驗方法，「輸入中文未斷詞的句子」經客語斷詞演算法處理後，將可得到該句「客語斷詞及標詞性」的輸出，再評估輸出的斷詞結果與正確答案，即可算出正確率，以作為各種實驗方法之改善依據。

在斷詞效能評估的方法，我們使用正確率(Precision)、召回率(Recall)、以及 F-分數(F-score)來評估系統的效能，這三種方法的定義如下所示：

$$\text{正確率}=\frac{\text{正確斷出的詞數}}{\text{全部總詞數}} \quad (13)$$

$$\text{召回率}=\frac{\text{正確斷出的詞數}}{\text{標準答案之總詞數}} \quad (14)$$

$$F-\text{score}=\frac{2\times\text{正確率}\times\text{召回率}}{\text{正確率}+\text{召回率}} \quad (15)$$

建置客語語言模型 (Hakka Language Models)

本研究中，有關計算詞串機率值客語語言模型，係依據客語語料訓練建置而成的，如前面所述我們採用 N-gram。由於客語語料之採集較不易，因此我們想到幾種取得語料的方式，

1)目前所收集客委會客詞典，其中客語例句作為訓練之客語語料來源。

2)由中文語料部份也許先做出一個國客語詞彙對應表，在將中文語料轉成客語語料，再作為訓練語言模型之用。

由於客語語料明顯少於華語語料，因此，我們初步先建置客語單字之 N-gram，將建置客語字(character)之 unigram, bi-gram 與 tri-gram 等三個語言模型。進一步以客詞詞典作為斷詞之依據，前述之語料經斷詞之後，用為訓練客語詞(word)語言模型。同樣，我們將建置客語字詞(word)之 unigram, bi-gram 與 tri-gram 等三個語言模型。

因由客語語料缺少，將面臨資料稀疏(data sparseness)所造成之未知詞(unknown event)問題，因此需使用滑化(Smoothing methods)方面加以克服。我們將採用 Back-off 法，在 N-gram 出現未知詞時，退回(N-1)-gram 求其未知詞之機率。如果在客語詞 N-gram，將退回至客語字 N-gram，以 Back-off 平滑方法解決語言模型之資料稀疏問題。

2.2 AI 類神經網路與深度學習

深度學習(Deep Learning)與類神經網路(Neural Networks)

人工智慧領域中，**深度學習**(Deep Learning)是機器學習(machine Learning)的一種方法，它以類神經網路(Neural Net.)為基本架構，可對大量資料進行其特徵(Features)學習的演算法。深度學習的好處是用非監督式或半監督式的特徵學習和分層特徵擷取獲得演算法。特徵學習的目標是尋求更好的表示方法並建立更好的模型來從大規模未標記資料中學習這些表示方法，因此可以代替人工取得之特徵，因而實現人工智慧的目標。

目前 google 平台對於大部份的強勢語言，如英文、日本與中文等，均可提供良好的語音辨識能力，正確率亦很高，已有許多的商業應用的例子。然而，對於本研究之臺灣客語語言而言，目前 google 就無能為力了，因此有賴我們進一步之研發。

TensorFlow 功能簡介

TensorFlow 是由 Google 所開發的一款深度學習模型框架，早期是供 Google 自己內部使用，例如: Gmail 過濾垃圾信、Google 翻譯、Google 相簿、Google 圖片辨識、Google 語音助理、YouTube 內容識別等等，直到 2015 年 11 月才正式對外開放發布。TensorFlow 內建有許多的程式庫可供使用者進行模型訓練使用，其中 TensorFlow 實作的例子如下:

TensorFlow 中建立線性模型 (y = W.x + b)

上式 x,y 是變數值，W, b 是變量，例如: 輸入大量實際數據訓練模型,做梯度下降,使用線性回歸來找到最佳的 W 和 b，這樣對於給定的任意特徵值 x，我們就可以通過將 W、b 和 x 的值代入到模型中得到預測 y。

TensorFlow 利用資料流圖(Data Flow Graphs)來表達數值運算的開放式原始碼函式庫。資料流圖中的節點(Nodes)被用來表示數學運算，而邊(Edges)則用來表示在節點之間互相聯繫的多維資料陣列，即張量(Tensors)。它靈活的架構能夠在不同平台上執行運算。

TensorFlow 一詞，Tensor 與 Flow 的意義如下：

- **Tensor**：是中文翻譯是張量，其實就是一個 *n* 維度的陣列或列表。如一維 Tensor 即向量，如果是二維 Tensor 相當於一個矩陣。
- **Flow**：指 Graph 運算過程中的資料流。

Data Flow Graphs

資料流圖(Data Flow Graphs)是一種使用有向圖節點(Node)與邊(Edge)用以描述電腦計算的過程。節點表示數學操作，亦表示資料 I/O 端點; 而邊則表示節點之間的關係，用來傳遞操作之間互相使用的多維陣列(Tensors)，Tensor 在圖中流動的資料表示。當節點相連的邊傳來資料流，這時節點就會被分配到運算裝置上異步(節點之間)或同步(節點之內)的執行。

在人工智慧的學習型態中，訓練階段的數學模型的建立，往往使用機器學習方式，經過不斷的調整效能，先將特徵參數取得，再進一步經由類神經網路加以訓練，完成最後的模型。然而深度學習的方式則有所不同，主要差異在於，特徵參數取得與類神經網路訓練的過程，特徵的選取與類神經網訓練均在同時進行，參見圖4。經過演化後，目前深度學習已成為 AI 的學習主流方式。

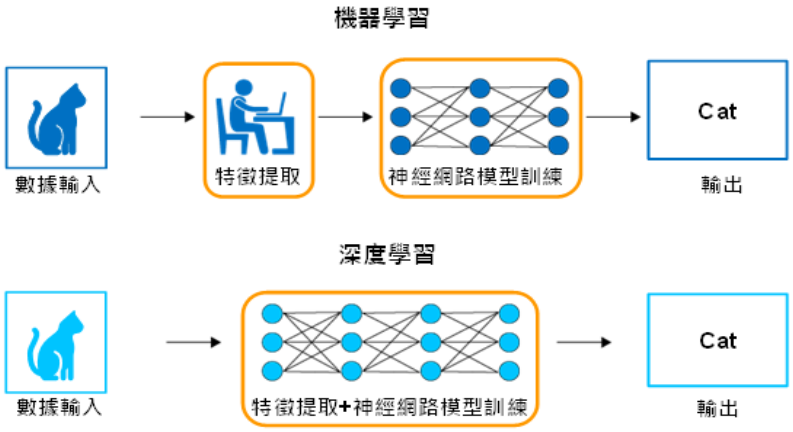


圖 4: 人工智慧之學習型態-深度學習。

深度類神經網路（Deep Neural Networks, DNN）是一種模型，可以使用反向傳播(Backward Propagation)演算法進行訓練。有關權重(Weight)之更新(Update)，可使用下式進行隨機梯度下降法求得：

$$\Delta w_{ij}(t+1) = \Delta w_{ij}(t) + \eta \frac{\partial C}{\partial w_{ij}} \tag{16}$$

其中，η為學習率，C為成本函式(Cost function)。

這一函式的選擇與學習的類型，以及啟動功能相關。為了在一個多分類問題上進行監督學習，通常的選擇是使用 ReLU 作為啟用功能，而使用交叉熵(Cross Entropy)作為代價函式。

Softmax 函式定義如下:

$$p_j = \frac{\exp(x_j)}{\sum_k \exp(x_k)} \tag{17}$$

其中，x_j 和 x_k 分別代表對單元 j 和 k 的輸入，而 p_j代表類別 j 的機率。

上述深度學習的類神經網路即是具備至少有一層隱藏層(Hidden Layer)之類神經網路，作為複雜的非線性系統，有較多的層次將可提供更高的學習能力。深度神經網路通常都是往前式(Forward)之類神經網路，在語言之模型建置方面，亦有拓展到遞迴式類神經網路。卷積深度神經網路（Convolutional Neural Networks, CNN）在電腦視覺領域得到了成功的應用，可作為聽覺模型被使用在自動語音辨識的領域，相較之下，比以往方法獲得較好辨識效果。

本計劃將採用深度學習方式，錄製後的客語語音經過語音前處理完成後，再經由此類神經網路訓練完成，將其辨識效益提升至最佳狀態，作為測試階段使用。

切割出來的客語詞彙則是訓練的語音資料，在 TensorFlow 卷積式類神經網路架構上，經過深度學習過程、不斷修正訓練參數，以達到最佳的辨識效果，語音辨識訓練過程如圖 5 所示。

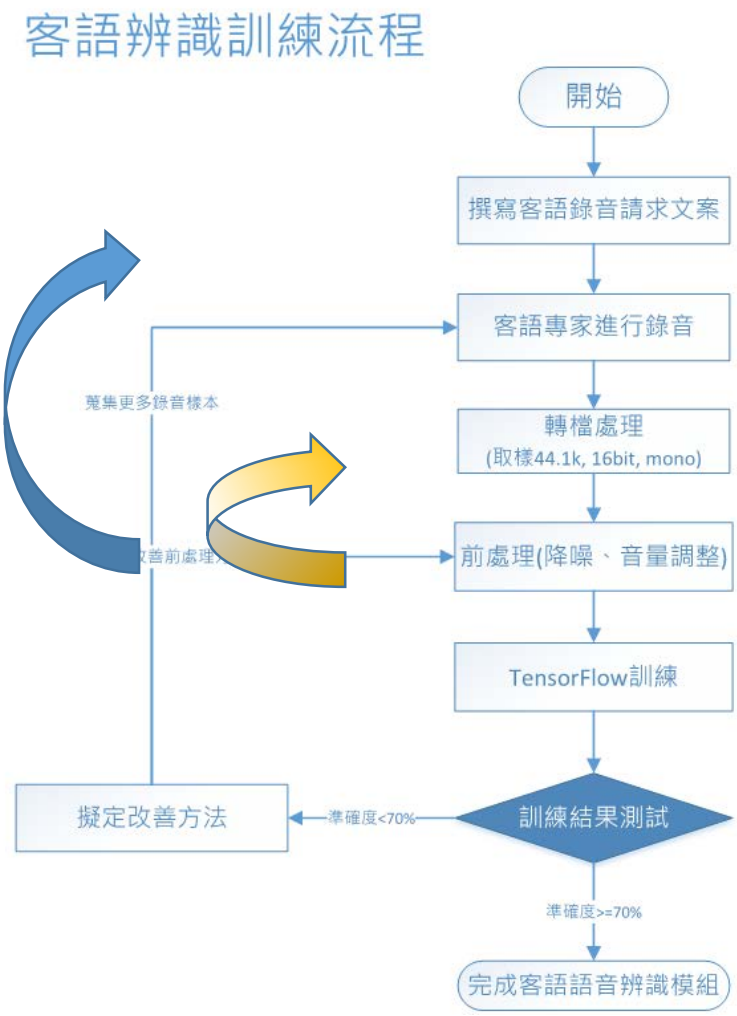


圖 5: 客語語音訓練的流程。

2.3 智慧客語控制系統

如前所述，我們以人工智慧的語言處理技術建置具客語的語言模組，包含中文、客語雙語，期望建置具有客語語音辨識功能、聽得懂客語(含中文)的「臺

灣客語智慧宅」。經過人工智慧之深度學習階段，完成客語語音辨識模組，因此客家長者經由客語語音指令，進而控制物聯網裡的設備，凡居家門禁、客廳廚房大小電燈，以及電風扇還有 webcam 監視系統與音響電視等配備。

本計劃中，我們採用 Respeaker 智慧音箱輸入語音、擷取音檔，經過前處理(含音量與雜訊處理)，將辨識語音送至 google 模組判斷是否為其可以辨識的中文語音，或者再送給客語語音辨識模組，確定語音辨識的詞彙內容，接著再進行後續物聯網中相關指令的控制。Respeaker 智慧音箱處理流程，如圖 6 所示。

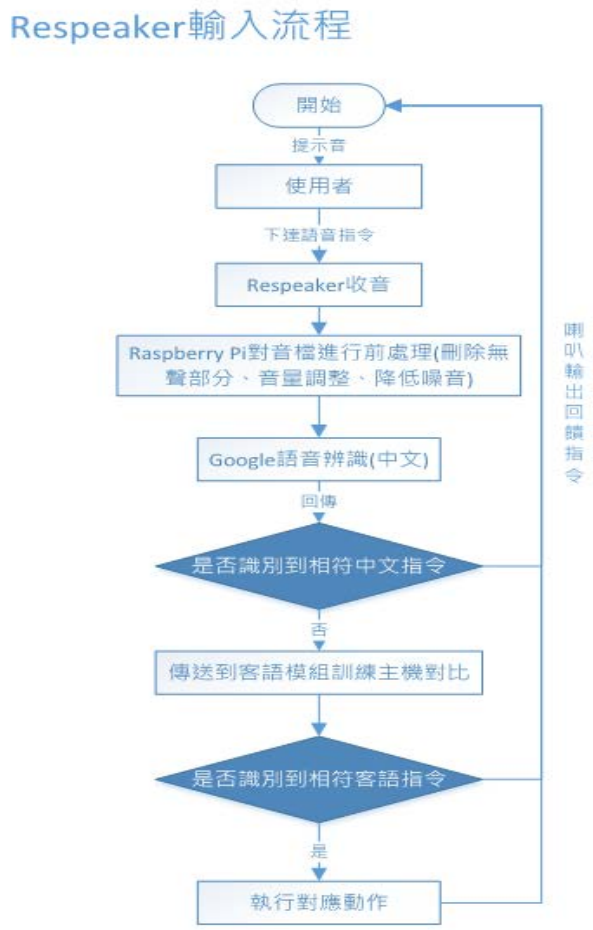


圖 6: Respeaker 智慧音箱處理流程

整合人工智慧客語-中文語音辨識技術與物聯網的智慧居家環境，架圖參見圖 11，系統特點說明如下：

運用人工智慧(AI) 客語-中文語音辨識(Speech Recognition)，長者可以用自然的語言操控居家環境中的家電設備，如門禁、電燈、冷氣、電視及各種 3C 產品。運用語音合成功能，提供多語音輸出的對話，包含中文、客語。在智慧的居家環境中，未來可以運用資料探勘方法，從家電操作的習慣，以機器學習技術學習使用的習慣，結合大數據(Big Data)方法，學習使用者的行為模式，分析並在預測使用者的下一步的動作，經由智慧的判斷，可以達到智慧開關與節能減碳的目的。

智慧門禁系統(智慧避卡鎖，不必攜帶卡片或磁釦等)，不需任何卡片，運用手機與客語語音即可打開大門順利進出。

一個可以讓讓老人以本土化的客家語進行居家控制的關鍵字語音識別系統架構，如圖 7 所示。該架構的辨識伺服器上的設計與其他雲端服務相似，建立一個可供其他裝置傳送請求的 API 服務介面提供連線。

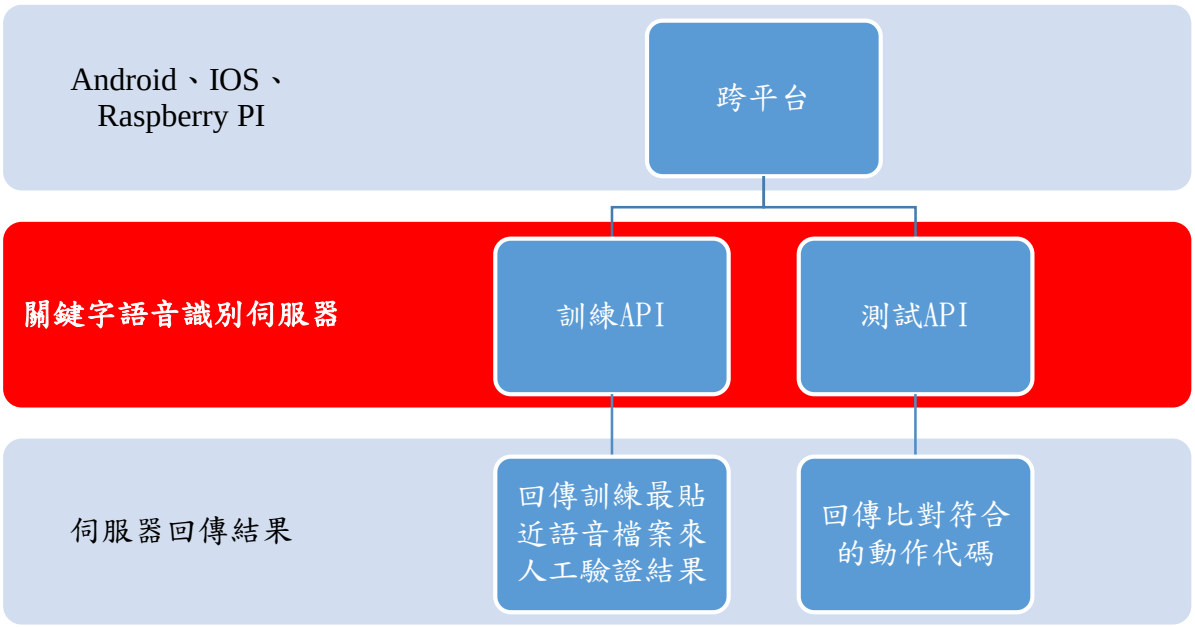


圖 7：互動機器人系統架構。

參、實驗結果說明分析

3.1 智慧宅物聯網配置與整合

本計劃的研究與實踐場域，在主持人所屬的研究單位-聯合大學資工系，已完成一完整的空間規畫，面積約 110 平方公尺(約 34 坪左右)，已初步完成室內的基本設計，本計劃構想的場域與功能可以在此完成與實踐。

目前，這個空間已有初步的基礎建置，如完整的空間、無線網路環境，以及一般的居家家電設備，如門禁、客廳廚房之大小電燈，電風扇、還有 webcam 監視系統，以及音響與電視等配備。本研究將在此基礎之上，研發客語語音辨識的智慧系統，結合人工智慧與物聯網技術，進一步整合於客語智慧宅場域。本計劃應可建置全國第一個客家語言與文化元素之智慧宅環境。

圖 8:聯大資工系智慧宅場域，已初步完成基本配置，將持續擴充建置客語智慧宅系統。

智慧客語控制系統簡述

本土化的語音在台灣為台語以及客家語，苗栗地區是台灣客家族群的重要地區，客委會調查顯示全國客家人口約有 450 萬，佔臺灣人口 2 成以上，臺灣民眾常使用中文(國語)或者臺語(閩南語)溝通，因此我們以人工智慧的語言處理技術建置具多語言環境的系統，可以含客語與中文，此為全國第一個具客語語音辨識功能、聽得懂客語的「**臺灣客語智慧宅**」。

可以經由客語語音指令控制物聯網系統中的相關設備，舉凡居家的門禁、客廳廚房大小電燈，電風扇與 webcam 監視系統，含音響電視等配備。

在智慧宅裡，有關居家環境中的安全(Home Security)因素，如門窗門禁、室內溫濕度與有毒氣體自動偵測等功能，以及異常狀況提醒與通知功能，亦將列入本計劃的的建置項目之中。

3.2 實驗進行說明

以下，分階段說明計劃實驗進行的內容。

階段一：語音庫資料蒐集

■ 開發「**手機行動客語錄音系統**」，如下圖所示，可向熟悉客家人士蒐集語料。

- 機動性高且能夠連續錄製多次。
- 依順序自動分類音檔到對應資料夾。

■ 再開發「**WEB 版線上客語錄音系統**」，如下圖所示，客家人可以直接在家透過電腦或手機錄音，並上傳音檔。

- 錄音時間更彈性、簡單。
- 可以預先聆聽錄音結果，允許只上傳清晰與正確的錄音。

階段二：AI 辨識模組訓練

■ 將蒐集的語料進行篩選，排除錄不清楚、雜訊過多或 NG 的音檔，將之進行前處理，轉換成統一格式(16KHz, Mono)，相同的開頭空白間格秒數(200ms)，並確定已分類到正確的資料夾中。

- 錄音正確、無雜訊
- 統一格式
- 開頭靜音 200 毫秒

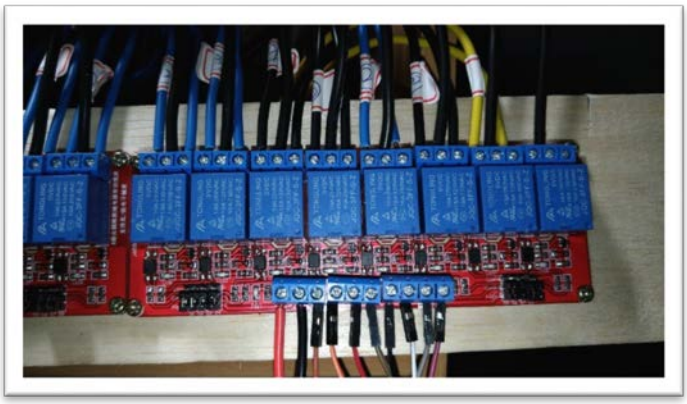
■ 至 2020 年 5 月，蒐集的客語語音音檔 10 個指令共約近 3000 個，排除錄製錯誤、失敗、背景吵雜等不佳的檔案，目前估算**有效音檔約為 2309 個**，對應各項指令如下:

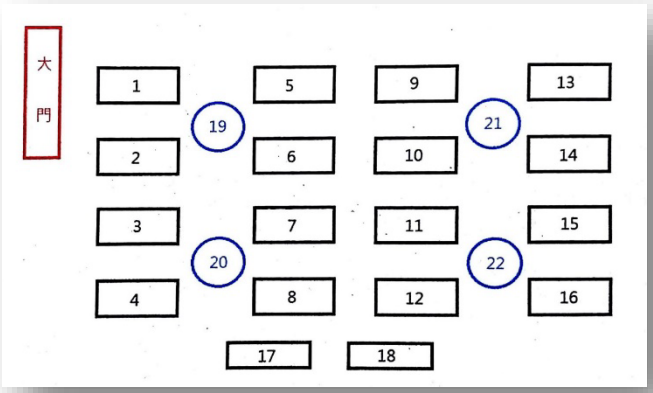
指令	音檔數量	指令	音檔數量
03 請開門	238	09 開大燈	243
04 請關門	241	10 關大燈	236
05 打開冷氣	235	18 開小燈	227
06 關掉冷氣	204	19 關小燈	236
16 關警示	230	20 救命	219

- 使用 Google TensorFlow 中的 audio_recognitiond 開放原始碼建立客語語音辨識模組訓練。

訓練流程如下：

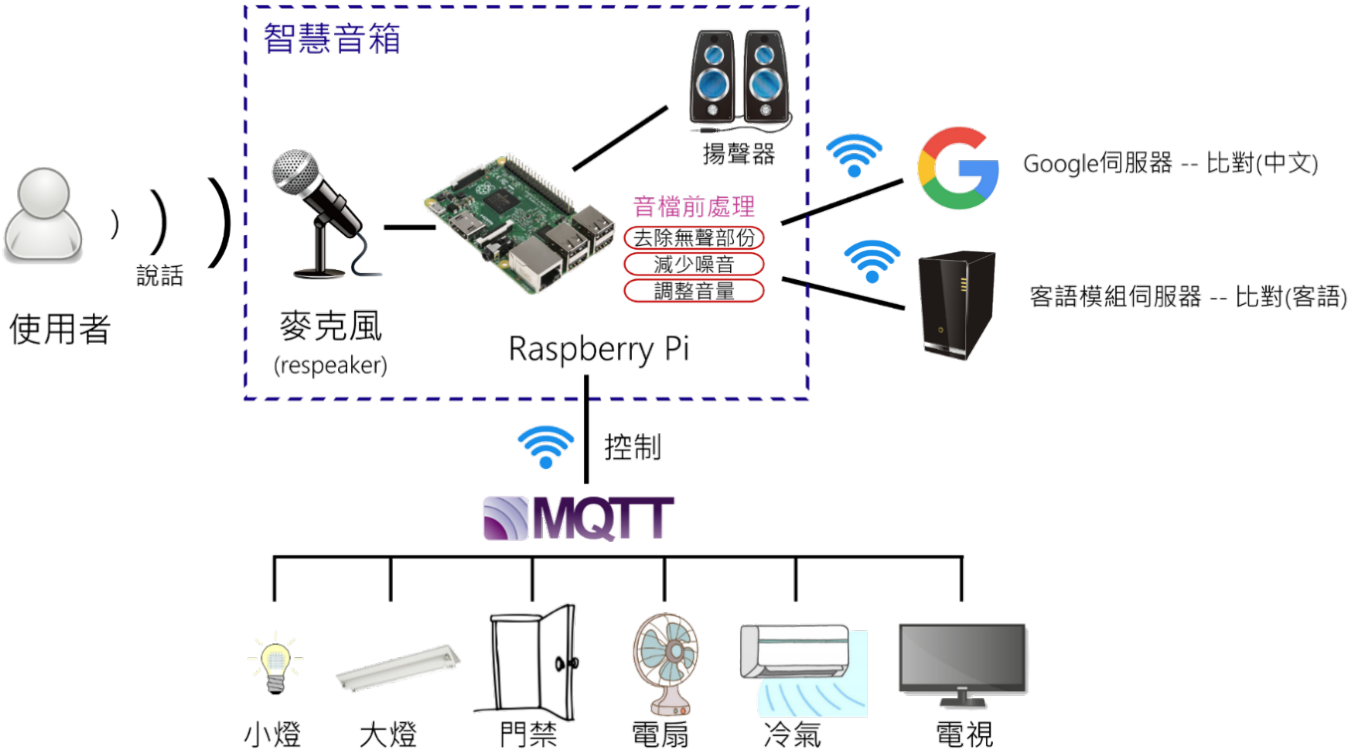
- 請電機工程師完成智慧宅場域配電。
- 測試繼電器編號對應的控制項目。





階段三：後端居家相關設備控制

- 當使用者說出客家語指令，會透過 Wi-Fi 傳輸與伺服器比對辨識結果。
- 透過 MQTT 通訊協定來控制對應的設備，如開/關門、大燈、小燈之關開之動作等。



階段四：實際測試結果與優化

- 測試環境：安靜環境，距離麥克風大約 1 公尺。
- 初步測試結果：新訓練的辨識模組可以達到不錯的正確率，
  唯有「**06 關掉冷氣**」和「**16 關警示**」的正確率較低，分別為 62%和 54%。
- 平均正確率約 80%
- 正確率較低可能原因：收錄語料時產生的偏差。
  - 例如「關掉冷氣」有些客家人會說「關掉」，有的則說「關閉」。
  - 「關警示」的「警」字在客語中也有人讀「ㄍㄧㄥˇ」的音和「ㄐㄧㄥˇ」的音。

指令	正確率	指令	正確率
03請開門	13/13 (100%)	09開大燈	11/13 (85%)
04請關門	11/13 (85%)	10關大燈	12/13 (92%)
05打開冷氣	10/13 (77%)	18開小燈	12/13 (92%)
06關掉冷氣	8/13 (62%)	19關小燈	12/13 (92%)
16關警示	7/13 (54%)	20救命	13/13 (100%)

3.3 研究成果效益:

本計畫主題有關具客語語音辨識的臺灣客家智慧宅之相關研究，達到的成果效益如下:

- 因應社會趨勢，建置**全國唯一**客語(四縣腔)語音辨識的智慧宅物聯網系統雛型。
- 建構一個具特色、支援雙種語言，如中文、客語語音。
- 客家長者可以進行居家各種家電設備之語音控制。
- 使用者不限於客家長者，一般客家人均使用。
- 研發成果有助於客家語言與文化之推展與傳承。
- 研發技術可擴展至客語其它腔調，以及其它臺灣本土語言。

四、結語

本計畫研究主題在於建置完成臺灣客語智慧的居家應用系統-「**臺灣客語智慧宅**」。我們運用人工智慧(A.I.)技術作為客語語音辨識的方法，結合當前熱門最新 ICT 相關技術，含物聯網、大數據與雲端計算資料庫，期望為客家族群眾多長輩們打造一個具智慧、便利與安全樂利的居家環境。

目前臺灣學術界與民間業界投入研發智慧宅裡的許多應用，常見的功能多在手機 APP 端結合居家設備的相關控制。本計劃提出一項創新的智慧宅環境構想，建置臺灣客家語言環境，使用者可以經由客語的語音來操作控制居家環境裡的各項設備，如:大門開關與電燈開關等，可為長者處理日常簡易的事務。

依實驗結果顯示，在 10 個日常居家指令中，客語語音辨識正確率達 80%左右，可以運用客語語音作為操作室內 3C 設備，未來我們將持續改善係統辨識的效益。

一)未來改善方向:

實際應用下，經過我們的反覆討論、分析，發現目前有三個問題需要解決的問題：

1. **吵雜環境問題**:

   如果在錄音時周邊有其他人說話或是有其他的聲響(如電視機)，系統收錄到這類外界干擾的雜音就容易辨識錯誤，我們可能需要調整麥克風的靈敏度以及指向性，並且優化降噪的程式處理。

2. **喚醒時機問題**:

   現行的機制是手動執行語音辨識程式，但是真實的環境下不會是要說客語時才去「按」執行，我們需要有一個好的啟動方式，比如透過特定的字詞喚醒辨識系統等等。

3. **一詞多種說法問題**:

   一種詞彙並不是只有一種講法，例如:「關掉冷氣」，可能只說「**關冷氣**」、「**關掉冷氣**」或「**冷氣關掉**」 ，但現在組只適用於設定好的講法「**關掉冷氣**」 。

致 謝:

本計劃得以完成，感謝客家委員會補助支持

謝謝聯合大學資工系王能中教授、客家研究學院張陳基教授與葉昌玉老師，以及提供錄音的客家朋友、同學們的協助。

恁仔細!