



研究計畫

整合雲端語者辨識及語音對話之銀髮族智慧家電服務系統設計與實現

*林博川

東方設計學院遊戲與動畫設計系

摘要

本計畫提供使用者可以在不使用麥克風的情況下，自由地說出需求或感受，說話者以自然語言的方式輸入，可以包含綴語、綴詞，系統便能自動判斷使用者身分及使用意圖後，直接對暖氣、燈光與咖啡機進行適當的控制。使用者不需要知道每項家電產品其複雜的控制方法與相關功能，藉由多為空間收音技術、使用者的語音意圖辨識、語者辨識，即可進行耦合服務，自動控制家電，並藉由系統回饋的語句對話得到家電操作之確認。本計畫預計以二年的時間完成下列工作項目：

- **完成多聲道聲音捕捉之無線傳輸模組：**包含嵌入式聲音捕捉之無線傳輸模組開發、Android 作業系統移植與系統 Bootloader 移植、多重感測器之介面製作。
- **完成雲端計算之認知多維空間收音技術：**將配置多麥克風陣列，測試相關收音設備，搭配指向性麥克風於室內空間，進行多軌即時錄音並利用演算法進行環境雜訊消除。
- **完成雲端銀髮族家電使用意圖辨識：**探討老年人的聲音特徵、語音特性，蒐集各國有關老人語音辨識的相關文獻，改良傳統特徵方式與改良 HMM 模型以適合在雲端與裝置端使用。
- **完成雲端語者身份辨識：**測試聲音經過語音信號前處理後，利用 SVM 支援向量機之一對一或一對多投票樣本分類機制進行家庭成員身分辨識，預計 10 人使用情況下辨識率達 90% 以上。
- **完成雲端家電操作之語音回饋對話系統：**根據家電使用意圖辨識結果進行語句結構分析、對話決策，找出最佳回應對話文字利用語音合成為一語句，傳送回使用者進行家電操作確認。
- **完成 3D 互動虛擬服務生與場景設計：**以虛擬動畫方式呈現一位虛擬服務生，與銀髮族使用者進行語音互動，以照護以及陪伴為考量，設計讓人感到溫馨、溫暖、快樂為目的。

關鍵字：資訊家電、雲端運算、虛擬服務生、3D 動畫設計

1. 研究目的

在銀髮族數位生活空間中，有多台家電裝置位於不同的區域，當使用者說出需求指令時，本計畫提出之系統會將使用者之語音透過雲端計算，藉此辨識使用者想要對於那一區域的家電裝置，例如客廳的電視或是廚房的咖啡機，使得適當的家電裝置能夠接收控制指令。此外，亦使用語者

辨識，確定使用者的身份，對於其家電控制之喜好作統計並作出最適當的設定，例如電視的最佳音量、冷氣的溫度或是咖啡是否加糖或奶精等。

本計畫提出的使用情境如下(如圖 1)：爺爺回到家裡，說：「我剛回到家怎麼感覺好冷喔！來杯熱咖啡吧！」，3D 虛擬服務生便能自動判斷使用者身份、及其使用意圖後，直接對暖氣、燈光與咖啡機進行適當的控制，並回答：「爺爺您好，我已經幫您開啟暖器到您最喜歡的溫度，燈光也調整到暖色系，熱咖啡也幫您泡好了！」。

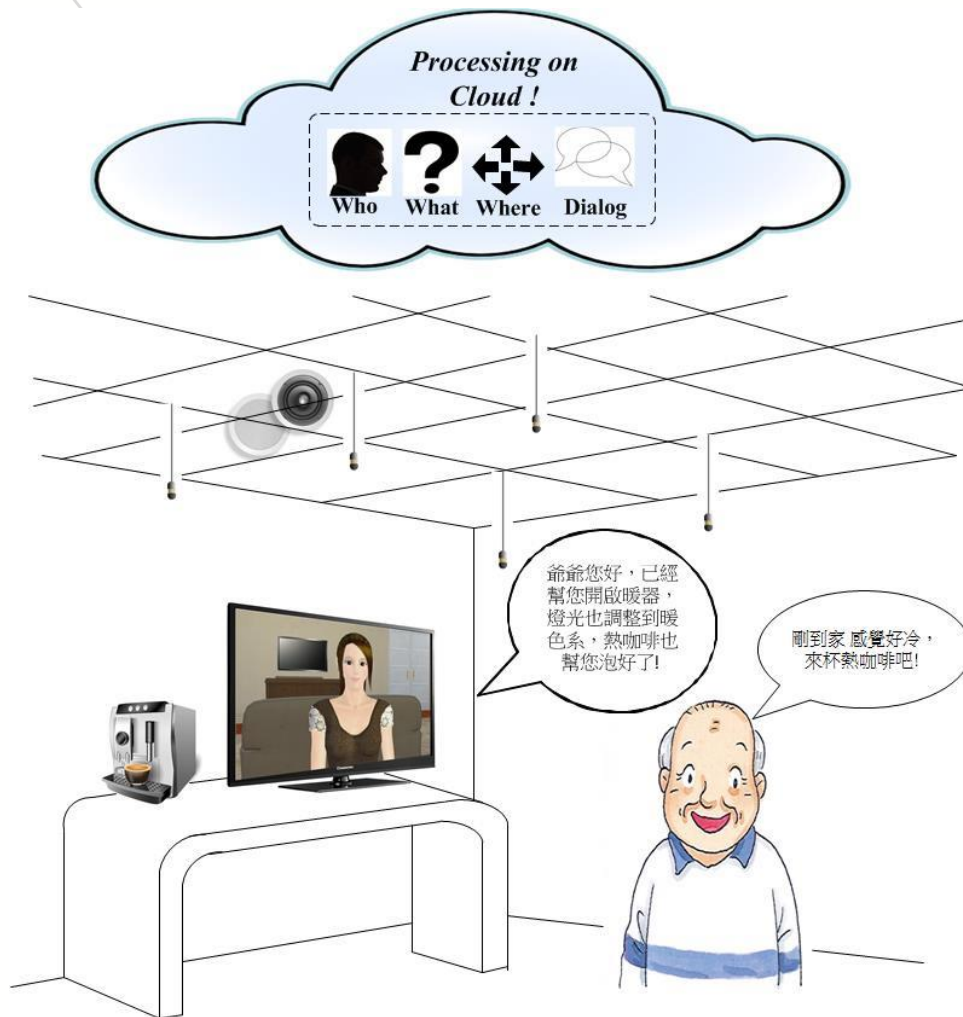


圖 1. 語者/語音意圖辨識應用於家電控制情境圖

欲達成上述的智慧家電使用情境，計畫將實現一個多聲道聲音捕捉之無線傳輸模組捕捉使用者的語音，且該模組的麥克風與相關感測元件能監控環境變化，並執行多軌錄音後藉由無線網路上傳至雲端，將需要高運算量的語音/語者辨識系統透過雲端執行，提升辨識準確率。此外一個回饋式對話系統也將提供語句回饋給使用者進行家電操作之確認工作，將合成出來的語句以及家電控制命令由雲端回傳到數位電視，並以一位 3D 虛擬服務生進行視覺呈現，使得家電裝置能夠與使用者自然的互動，猶如一位貼身服務生。本計畫歸納出以下六大研究主題：

- (1) 多聲道聲音捕捉之無線傳輸模組之設計與實現
- (2) 雲端計算之認知多維空間收音技術
- (3) 雲端計算之家電使用意圖辨識
- (4) 雲端計算之語者身份辨識
- (5) 雲端家電操作回饋對話系統
- (6) 3D 互動虛擬服務生與場景設計

2. 研究計畫之背景

在經濟不景氣下，許多人相信資訊家電、智慧型手機與 PDA 等產品將是挽救台灣 PC 王國的最後一張王牌。智慧化家庭是其中重要的一項應用，例如近年 CES 展 LG 所展示應用在智慧家電控制的 THINQ 技術，便採用 Wi-Fi 作為路由器或是閘道器，透過智慧手機和平板裝置，進一步控制洗衣機和烘乾機、電冰箱和微波爐等家電用品，並結合智慧電表，形成一套居家智慧化的互連強化管理架構。三星也推出結合 Wi-Fi 和圖形化螢幕選單的聯網冰箱，Panasonic 則是展示以聯網電視為核心平台、結合 Wi-Fi 家電產品的解決方案。居家智慧化的物聯網應用，涵蓋面越來越廣泛，包括 GE 和英特爾合推家庭遠距照護解決方案，A-Lu 和 Tantalus 合作的智慧電網方案，正文科技展示的家庭資訊裝置等等，Control 4 也展示以單一介面操控家電的解決方案。IEK 工研院產業情報網預估，智慧化居家操控應用將會帶動更多的新應用和操作體驗，家電無線化的腳步已經起動。

微軟公司於 2012 年針對居家智慧化操控應用，設計出結合語音辨識和體感偵測的操作介面，並進一步提出以遊戲機 XBOX 作為串聯智慧家庭操控核心平台的架構。另外 Samsung 公司發表電視能透過電視內建的鏡頭及麥克風，感應動作控制、語音控制及手勢辨識功能，還能透過 20 種語言以上的語音來操控電視，使用者只需以一種語言說出指令便可啟動任一應用程式及網路瀏覽器，或揮動雙手便可進行網頁瀏覽及選擇連結，取代傳統對遙控器的依賴。然而目前的相關展示接以固定單一的家電為主，並無實際與雲端資料庫連結或使用者身份、使用習慣進行最佳耦合，因此使用者必須記憶並學習相關語音指令，對於電子產品不熟悉之高齡使用者而言功能過於複雜，學習過於困難。本計畫提出之使用情境即為解決上述使用困難的問題，提供一套完整的銀髮族人機介面，打破以複雜下拉式選單方式呈現功能，取而代之是以一位 3D 虛擬服務生進行互動，以豐富的表情、情緒回饋方式給銀髮族使用者，讓使用者更加感到貼心與幸福感，並力求簡單、好用、免學習，提出完整解決方案。

3. 本計畫之重要性與應用範圍

本計畫主要包含 6 大研究主題，包含多聲道聲音捕捉之無線傳輸模組設計與實現、雲端計算之認知多維空間收音技術、雲端計算之家電使用意圖辨識、雲端計算之使用者身份辨識、雲端家電操作回饋對話系統與 3D 虛擬服務生與場景設計。本計畫之運作方法與流程如圖 2 所示：使用者自身感受說出後，語音將透過聲音捕捉，進行語音接收與 AD 轉換後將聲學訊號傳送至雲端伺服

器，而伺服器上提供五項主要功能，包含：

- (1) 雲端語音辨識：用於判定使用意圖。
- (2) 雲端語者辨識：用於判定使用者身份。
- (3) 雲端耦合服務：用於確定需要控制的家電種類以及控制細部項目。
- (4) 雲端對話回應決策：用於發話給使用者進行確認，將傳送控制信號給特定家電產品，進行精準控制與設定。
- (5) 雲端語音合成：用於將對話結果輸出給 3D 虛擬服務生，以動畫回饋給使用者。

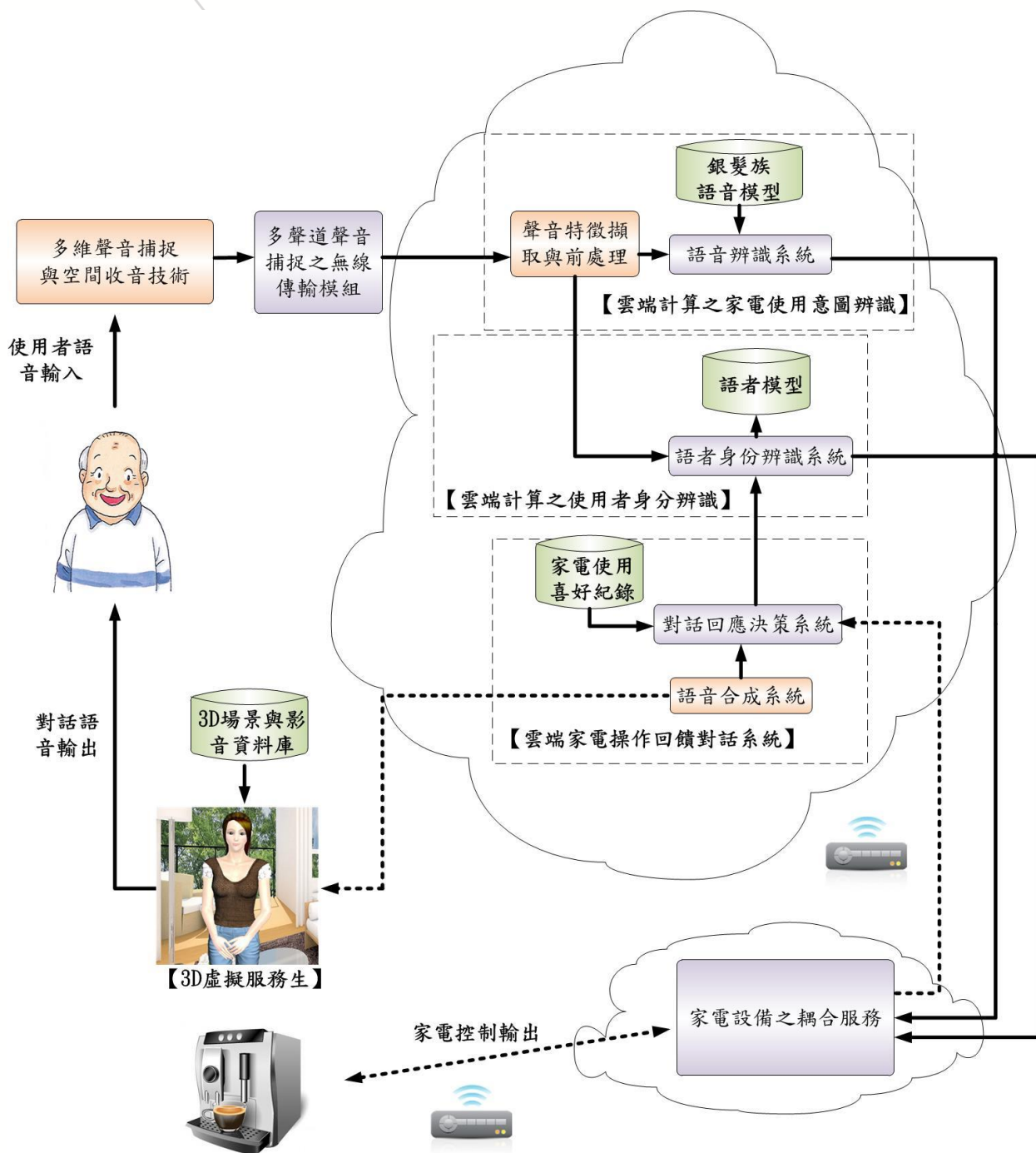


圖 2. 以情境說明本計畫技術關聯性

本計畫研究創新點說明如下：

- 具有多情境的對話方式：
 - (1) 開啟與關閉家電的對話
 - (2) 一般操作家電的對話
 - (3) 特殊操作家電的對話
 - (4) 具記憶型態家電的對話

- 具強健性的對話系統架構，提高對話意圖辨識準確率
 - (1) 分析以及將家電對話方式分類，提升意圖辨識率
 - (2) 利用關鍵詞擷取技術達到使用者直覺式的家電對話操作效能
 - (3) 多模式對話策略選擇，提升意圖辨識率根據使用者的互動方式給予適當的回應

- 探討研發強健對話意圖辨識前瞻技術
 - (1) Two-pass based keyword spotting strategy 技術
 - (2) Wake-up-word recognition 技術
 - (3) BLSTM (Bidirectional Long Short-Term Memory)技術

Design and implementation of an intelligent service cloud system for appliance control based on integration of senior user identification and speech dialog

Lin, P.-C.

Tung Fang Design University, Department of Digital Game and Animation Design

This project aims to develop a system which allows users to speak without using microphone to control home appliances. Users can use natural language sentences (may contains suffix words) as input for the proposed system. In the above scenario, the user does not need to know the complex control methods and related functions of each household electrical appliance. Using multidimensional space audio recording technology, results of intent recognition and speech recognition can work with coupling services. Moreover, a dialogue system can feedback a related sentence for the user to confirm the final operations of appliances. Research issues include:

- Complete the wireless multi-channel sound catching module;
- Complete the multidimensional audio recording with cloud computing;
- Complete user intent recognition for cloud information appliances;
- Complete cloud speaker recognition on cloud;
- Complete cloud appliances operating feedback dialogue system;
- Complete the 3D interactive virtual waiter and scene design.

Keywords: information appliance, cloud computing, virtual waiter, 3D animation design