第 19 屆行動計算研討會

基於語音辨識之智慧型輪椅控制

Shun-Hung Tsai 蔡舜宏
國立台北科技大學自動化科技研究所
shtsai@mail.ntut.edu.tw

Yi-Min Liao 廖亦珉
國立台北科技大學自動化科技研究所
weining6911@gmail.com

Liang-Chia Chen 陳亮嘉
國立台灣大學機械工程學系暨研究所
lchen@ntu.edu.tw

摘要

本研究將利用 Raspberry Pi 嵌入式平台開發一套語音辨識系統，搭配外部的立體聲麥克風以及外部音效裝置構成系統之硬體架構應用於電動輪椅上。首先，利用語音辨識演算法，將聲音訊號進行訊號前處理再求取訊號之特徵參數。除此之外，利用 Fuzzy c-means 模糊 C 平均值演算法進行語音分群比對，使其能準確辨識出特定語者之語意來決定輸出之電壓以控制電動輪椅。最後經由實驗與結果分析，實際驗證本語音系統辨識率的精確性及可靠性。

關鍵詞：智慧型輪椅、語音辨識、嵌入式系統.

Abstract

A voice recognition system employing Raspberry Pi embedded platform for intelligent wheelchair control is developed. By utilizing the voice pre-processing algorithm, the characteristic parameters of voice signals can be obtained. In addition, using the fuzzy c-means clustering algorithm, the voice signals can be classified to recognize the specific language semantics accurately and then to provide the input voltage for robust control of the intelligent wheelchair. To verify its feasibility, an experiment was performed to prove the control accuracy and reliability of the developed voice identification system.

Keywords: Intelligent Wheelchair, Voice Recognition, Embedded Systems.

一、研究動機與目的

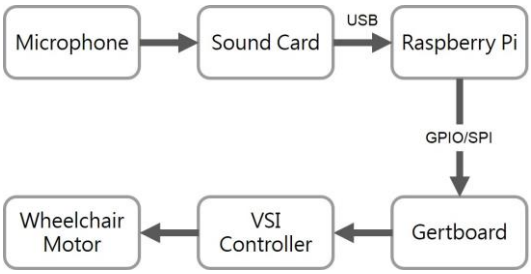
101 年底領有身心障礙證明者近 112 萬人，續創新高[2]。人口老化以及身心障礙問題，一直以來都是開發中或已開發國家面臨的社會問題。行動遲緩以及行動不便的人在高齡化的社會中顯而易見，如何利用現今的科技幫助行動不方便的身心障礙人士成為重要議題。對這些人而言，輪椅可以幫助他們抵達目的地，並且減輕照顧者的負擔，在日常生活中可以照顧自己的起居生活，且進行短距離的移動，不會讓行動不便對他們的生活造成嚴重影響，即使如此，一般型的輪椅對於年齡較高以及行動不便的人來說，操控上仍有一定難度，因為控制輪椅需要運用手臂力量讓輪椅前進或是後退，因此智慧型輪椅的發展儼然是未來趨勢[3]。使用者只需透過輪椅上面的搖桿就可以順利操控前進、轉彎、後退的動作。

近年來台灣的人口結構中，老年人口及行動不便之人士一直佔有相當高的比例，許多高齡者及行動不便之人士仰賴輪椅代步，在這些人口中，常因為功能退化或疾病的關係，自行操控輪椅有相當高的難度，往往無法藉由輪椅自主行動。因此本研究提出利用嵌入式系統，結合語音聲控，幫助行動不便之人士及高齡者經由語音聲控方式自行操控輪椅。希望能提升輪椅使用者的便利性，減輕照顧者的負擔。

二、硬體介紹

電動輪椅的操縱桿所產生的問題在於不斷重複固定動作與姿勢，長時間使用容易造成肌肉疲勞或局部傷害，評估下來聲音的利用變成了最理想最自然的輸入方式，對於大多數無法使用搖桿控制的身障者而言，聲控是一個較理想的控制概念，結合 Raspberry Pi 嵌入式系統可提高電動輪椅操控之便利性也易於後續之系統整合，其系統架構如圖(1)所示。

HyWeb



圖(1) 硬體架構圖

(1) Raspberry Pi 介紹

Raspberry Pi 是一款基於 Linux 系統的開發平台，如圖(2)所示，它由英國的 Raspberry Pi 基金會所開發，模組採用 BCM2835 的系統單晶片（SoC）的運算核心，配備一枚 700MHz 的 ARM 架構處理器，512MB 記憶體，使用 SD 卡當作系統儲存媒體，且擁有一個網路裝置(Ethernet)，兩個 USB 2.0 介面，以及 HDMI（支援聲音輸出）和 RCA 端子輸出支援[4]。

圖(2) Raspberry Pi Model-B 示意圖[4]

(2) Gertboard 擴充板

語音訊號經過處理後須透過 Gertboard 將數位訊號轉換為類比訊號輸出到 VSI 控制器，以達成控制電動輪椅馬達的目的，圖(3)為 Raspberry Pi 與 Gertboard 連接圖。Gertboard 包含了強大的擴充功能，它可以補足 Raspberry Pi 在其它功能上面的不足。利用 Gertboard 的 D/A 模組將訊號轉換為類比電壓控制 VSI 控制器，並不會與原本的線路造成衝突。

圖(3) Gertboard 示意圖[4]

(3) USB 音效卡

本研究使用外部音效卡，目的將麥克風接收到的語音類比訊號轉換成數位訊號，並設置音量及增益(Gain)大小，音效卡採用 2 聲道模擬 7.1 聲道，支援喇叭及耳機，含麥克風孔、支援 USB 介面隨插即用。

(4) 麥克風

本研究因考慮到行動不便者可能無法利用手部操控電動輪椅控制器，因此在聲控系統硬體中選用領夾式麥克風，如圖(4)所示。

圖(4) AT-9902 實體圖[8]

(5) 電動輪椅

本研究使用之電動輪椅車為必翔公司所生產，如圖(5)所示，輪椅包含一個控制方向的 VSI 控制器，屬於一個完全可程式的控制器，可以利用它操控輪椅的方向和速度，可依據手指施力大小決定輪椅移動的速度，控制器已預先將程式寫入，以滿足一般使用者的需求。

圖(5) 必翔電動輪椅車[1]

HyWeb

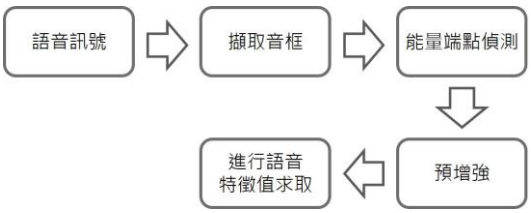
(6) VIS 控制器

電動輪椅控制器，如圖(6)所示，此裝置具電路檢查功能，在不破壞原電路設計的前提下，保留 VSI 控制器的檢查電路，從外部連接控制電路達到其控制目的。

圖(6) VSI 控制器[1]

三、語音前處理

語音前處理如圖(7)所示，其主要目的是希望能從中找到一個代表語音產生的數學模式，然後藉著此模式將語音訊號數位化，以利於數位訊號處理[7]。將需要的語音訊號切割出來，進行能量端點偵測計算起始語音及結束語音位置，並做音量補償，最後進行語音訊號的特徵擷取。



圖(7) 語音前處理流程圖

(1) 語音取樣頻率

本研究對於語音訊號的處理格式是採用 16bit，取樣頻率為 8000Hz，根據取樣定理，取樣頻率需大於原始聲音訊號的兩倍，將取樣頻率大於原始訊號兩倍以上就能確保訊號不失真[10]。若以8000Hz 作為語音的取樣頻率，代表1 秒鐘會有8000 個取樣點，如果將這些資料點直接取樣進來，其實很難達成即時辨識的效果，所以將語音做前置處理與找到語音特徵值是必要的步驟。取樣頻率越高，不僅佔用的記憶體越大，也會導致辨識的時間過長，因此本研究以取樣頻率 8000Hz 作為標準。

(2) 擷取音框

由於語音訊號的資料量非常龐大，為了瞭解其訊號的細部變化，會在整段語音訊號中選擇一個固定的資料長度[11]，將語音切割成小段個別處理，代表聲音訊號同時分成數個音框，一般音框使用的資料長度為 64、128 以及 256，設定好的取樣點集合成一個語音處理的最小單位。

(3) 能量端點偵測

能量偵測是在一段語音訊號中判斷有聲與無聲的一個方法，訊號在靜音部份的能量一定比有聲訊號的能量低[12]，因此只需設立一個門檻值來區分有聲與無聲，然後判斷平均能量從第幾個音框開始超過預設門檻，就能判斷有聲部分是從第幾個音框開始，利用一個短時間內(即一個音框)所有取樣值的能量和除以取樣值個數所得的平均能量。根據能量偵測法刪除一些明顯的雜訊，找出一個預設門檻，當音框的能量大於預設門檻時，此音框即為有聲區域，以此作為有聲訊號之端點。

(4) 預增強

語音訊號經由嘴唇發出後會造成高頻的損失，而系統無法像人類耳朵敏銳到能夠將損失的高頻部分補回[12]，為補償語音訊號受到發音過程所壓抑的高頻部分，故先將語音作預增強，將語音訊號先經過一個高通濾波器，進行高頻率訊號的提升，其中 α 介於 0.9 與 1 之間，本論文取中間值 0.95，預增強後的語音訊號會變得較為清晰，其訊號表示為：

$$S_p[n] = S[n] - \alpha \cdot S[n-1] \qquad (1)$$

四、語音特徵值求取

將語音訊號前處理後得到的結果進行語音訊號特徵值的求取，找出具有代表性的特徵是語音辨識重要的一環，若能有效地擷取語音訊號之特徵，在辨識上將會獲得很大的助益。

(1) 相關係數法

一般來說，即便是相同語句，每一筆語音訊號的資料長度也會不同，所以很容易造成兩訊號間無法進行比對辨識，為了將兩個語音訊號長度變成一致，利用相關係數法透過不斷的計算語音訊號之相關特徵分析變數之間相互關係的方法，分析各變數間的相關程度大小與方向，一般是利用兩變數間的離差；假設有兩組樣本 $x_1, x_2, ..., x_n$ 與 $y_1, y_2, ..., y_n$，其樣本平均數分別為 $\bar{x}, \bar{y}$，樣本標準差分別為 S_x, S_y，且兩組樣本之共異變數(Covariance) S_{xy} 定義為：

$$S_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right) \quad (2)$$

則相關係數 r 定義為：

$$r=\frac{S_{xy}}{S_x S_y}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}} \quad (3)$$

兩組樣本之間的相關程度，其 r 值介於 -1 與 1 之間。透過不斷的計算相關係數值 r，得到的 r 值越大則代表兩兩訊號符合度最佳，此時將再訊號切割致同樣的資料長度，便是我們要的結果。

(2) 多項式曲線擬合

利用最小平方法求出最能表示出資料趨勢的函數，以多項式模型進行曲線擬合，稱為多項式擬合[6]，多項式擬合曲線時，未必階數越高擬合效果越佳，且越高的階數需要較常的計算時間，較大的記憶空間，因此本研究使用的方法是先繪出所有的資料點，再依資料分布情形選擇階數，使用多項式數學模型進行曲線擬合，公式如下所示：

$$y=f\left(x\right)=a_0x+a_1x+\cdots+a_nx^n \quad (4)$$

五、語音辨識方法

本研究利用相關係數法(Correlation Coefficient)計算語音訊號彼此間的相關係數，再運用模糊 C 平均值演算法作為語音辨識準則。控制電動輪椅的方向共有五種，首先經由資料輸入與初始化設定所需參數，接著決定分群的數目，然後不停地透過分群計算公式更新群集歸屬度，直到執行至小於預設之收斂值後停止；以下將介紹模糊分群演算法。

(1) 模糊 C 平均值演算法

模糊 C 平均值演算法(Fuzzy c-means clustering method, 簡稱 FCM)是 Bezdek[9]在 1973 年所提出的 c-means 演算法而衍生出來的，其概念是經由模糊處理後，不再將每一筆資料絕對的歸屬於某一個特定群集，而是賦予每個資料點屬於每一群的歸屬程度，再利用這些歸屬程度決定此資料點是屬於那一個群集，整個目標函數所代表的意義為各樣本資料到各集群中心的加權平方和，用這樣的方式解決最佳集群的問題使分群的效果能夠提升，以 FCM 的概念使分群結果更為合理化並增加分群之精確度，其目標函數(Objective function)公式定義如下：

$$P=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(U_{ij}\right)^m dist\left(x_j\ ,v_i\right)^2 \quad (5)$$

歸屬矩陣 U 表示為：

$$U=\begin{bmatrix}U_{11} & U_{12} & \cdots & U_{1n}\\ U_{21} & U_{22} & \cdots & U_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ U_{c1} & U_{c2} & \cdots & U_{cn}\end{bmatrix} \quad (6)$$

第 1 群到第 c 群的歸屬值總合為 1。

$$\sum_{i=1}^{c}U_{ij}=1\ ,\ \forall j=1,2,...,n \quad (7)$$

為了要滿足目前條件，可以利用 Lagrange function 得到一個新的目標函數 P_{new}：

$$P_{new}=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(U_{ij}\right)^m dist\left(x_j,v_i\right)^2+\sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{c}U_{ij}-1\right) \quad (8)$$

其中 λ_j 為 Lagrange Multipliers。為了要最佳化目標函數 P_{new}，針對所導入的參數進行微分，得到以下兩個方程式。

$$v_i=\frac{\sum_{j=1}^{n}\left(U_{ij}\right)^m x_j}{\sum_{j=1}^{n}\left(U_{ij}\right)^m} \quad (9)$$

與

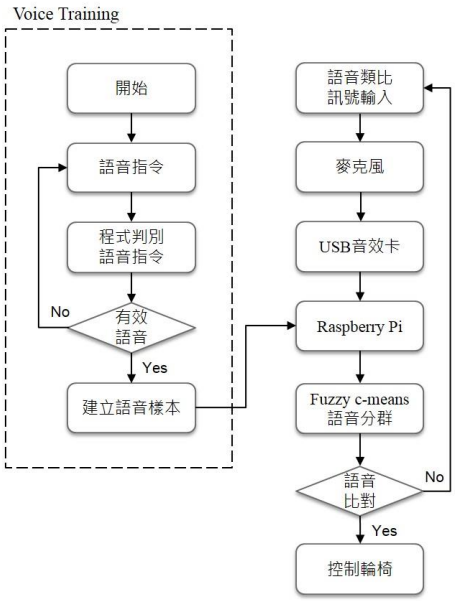
$$U_{ij}=\frac{1}{\sum_{k=1}^{n}\left[\frac{dist\left(x_j,v_i\right)}{dist\left(x_j,v_k\right)}\right]^{\frac{2}{m-1}}} \quad (10)$$

FCM 之目的是為了分析已知數據的分佈情形，以便掌握數據的性質，由分佈的中心點作為設定初始歸屬函數的中心點[5]。而理想的分群法則是希望將 n 個點或 c 個族群完全分離，使得同一個族群中的特徵高於其他族群；而所謂的特徵或相似性，代表資料點到群中心點的歸屬度大小，故歸屬函數與群中心點位置為 FCM 所要求的，由於分群中心 v_i 與歸屬函數 U_{ij} 互為因果關係，因此 FCM 是一種反覆疊代的計算。

Hy Web

六、實驗流程

首先，如圖(8)所示，將特定語者之語音指令，個別求取出不同的特徵值，並將所有的指令特徵值參數儲存至 Raspberry Pi 系統的記憶體作為比對樣本，此時，使用者可透過麥克風說出想要執行的命令，系統會即時的把接收到的語音訊號擷取出特徵值，並代入 FCM 做語音特徵值分群，辨識出對應的指令，回傳訊號給系統下達命令控制輪椅動作。



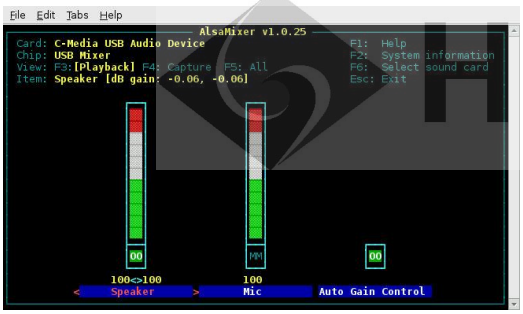
圖(8) 語音系統流程圖

(1) 錄音規格

本系統錄製語音取樣率為 8000Hz，並採用單聲道，共設置五個控制指令進行輪椅操控，分別為：前進、後退、左轉、右轉及停止。

(2) 音效系統設置

為了讓語音能透過 Raspberry Pi 進行處理，首先我們必須在 Raspberry Pi 系統下進行麥克風、喇叭音量、增益大小(Gain)以及外部硬體設置，如圖(9)所示，並在 Raspberry Pi 建置的 Linux 系統裡進行錄音與播放測試。



圖(9) 音量與增益設置介面

(3) 語音樣本建立

語音辨識前會先建立語音樣本以進行語音訊號比對，所謂的語音樣本，即是對語者進行錄製代表性的語音作為辨識比對的標準，本實驗語音資料採樣對象為特定語者，語音內容為操控輪椅的指令分別為前進、後退、左轉、右轉以及停止，錄音環境為實驗室，可能影響錄音品質的雜訊為周邊儀器設備運作聲音。在 Raspberry Pi 系統上進行語音樣本錄製並將接收到的來源語音指令進行前處理，最後擷取到的語音特徵值儲會存至 Raspberry Pi 記憶體裡，作為語音比對樣本。

(4) 語音指令辨識結果

將原始聲音計算其每個音框的音量大小，每個音框大小為 128 筆資料，每筆資料皆代表音框之音量大小。接著，將語音訊號進行多項式曲線擬合，取得該音量之趨勢線，由曲線擬合預估音量，當作該命令之音量特徵。每執行一個指令，系統便會產生一組屬於該指令的特徵值，在得到指令特徵值之後，利用模糊 C 平均值演算法將指令特徵值代入，透過不斷的疊代可得到各群中心點與歸屬度參數，依據此方法能將每一次輸入的語音指令進行語音特徵值分群比對，使其能準確辨識出特定語者之語意決定輸出之電壓以控制電動輪椅。

七、結論

本研究在語音辨識方面採用模糊 C 平均值演算法將得到的語音特徵值加以分群辨識，使其能準確辨識出特定語者之語意來決定輸出之電壓以控制電動輪椅，此外，語音系統在硬體架構上，採用 Raspberry Pi 當作開發平台，體積小穩定性佳，功能整合與擴充性強，最後經由實驗結果驗證本語音系統辨識率的精確性。

八、致謝

本文感謝科技部計畫編號 MOST 103-2221-E-027-089-

參考文獻

[1] 必翔實業，http://www.pihsiang.com.tw/

[2] 內政部統計處，http://www.moi.gov.tw/stat/news_content.aspx?sn=7516

[3] 元智大學老人福祉研究中心，

http://grc.yzu.edu.tw/IRW/index.aspx

[4] 台灣樹莓派，http://www.raspberrypi.com.tw/

[5] 李允中、王小璠、蘇木春，模糊理論及其應用，全華圖書，2012。

[6] 張智星，MATLAB 程式設計與應用，清蔚出版社，2004。

[7] 楊鎮光，VISUAL BASIC 與語音辨識，松崗出版社，2002。

[8] Audio-technica http://www.audio-technica.com.tw/

[9] J. C. Bezdec, "Pattern recognition with fuzzy objective function algorithms," New York: Plenum Press, 1981.

[10] J. R. Deller, *Discrete Time Processing of Speech Signals*, Macmillan, 1993.

[11] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall Co. Ltd , pp 200-232, 1993.

[12] L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances," *The Bell System Technique Journal*, Vol. 54, February 1975, pp. 297-315.

HyWeb