

2018 Arm Design Contest 設計競賽

初 賽 報 告

作品名稱：iHands 中文手語辨識系統

隊名：我們超厲害

中華民國 107 年 9 月 13 日

注意：

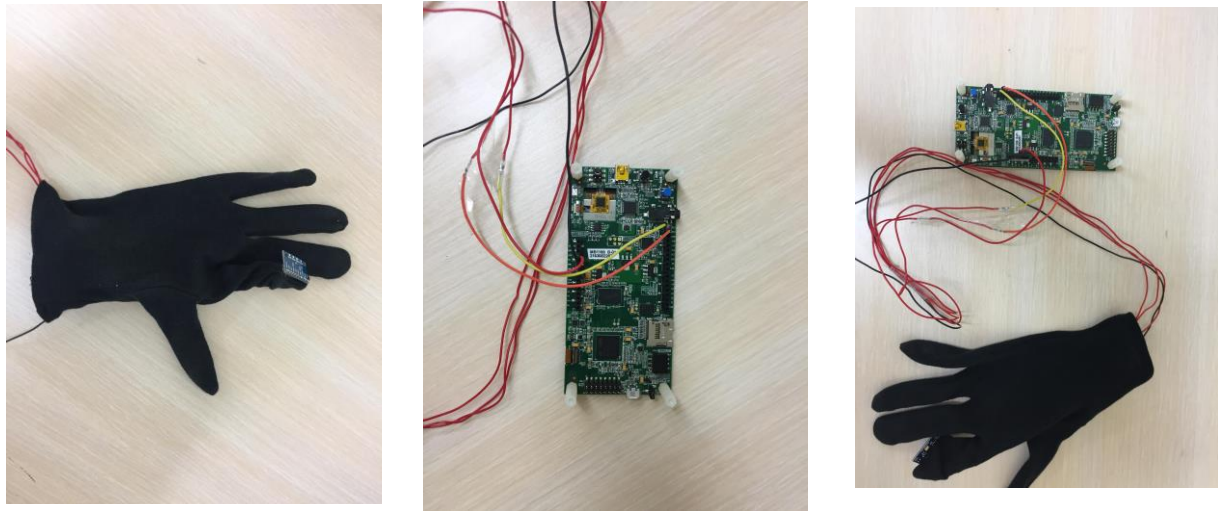
1. 為維護比賽公平性，請勿於報告中洩露比賽隊伍之學校與教授名稱。
2. 敬請依此格式撰寫報告，若有不適用之項目，該項目請留空白，切勿刪除。
3. 至多請勿超過十五頁，內文字體大小以12字體，行距以1.5行距為原則，並加上頁碼。
4. 報告得以中文或英文撰寫。
5. 請以PDF為檔案格式提交。

1、摘要

生活周遭一切逐漸智慧化，自動開關、調節亮度的電燈、自動連網請商家送菜的冰箱、自動播放的天氣資訊與新聞、無須動手就能安全地開上路的汽車...等。智慧生活為人們的生活帶來更多便利，然而弱勢族群的生活卻沒有什麼改變，我們體認到此一現實後，決定以人道為出發點，以 STM32F469 開發板整合現代機器辨識能力、訊號處理基本演算法，完成一中文手語辨識系統，並且搭載在穿戴式裝置上，只需要一雙手套，便可以即時將手語轉成語音，讓普通人馬上知道聾啞人士所表達的話語。

2、作品功能與實用性

作品如何改善智慧生活與結合未來應用趨勢之功能與實用性



如圖所示，此手套可以量測指尖點的相對加速度、角速度、磁場，我們可以藉由手指指尖的變化，進行中文手語的辨識，瞬間即時口譯成語音播放出來。

本系統主要功能為：

1. 量測手指之彎曲度變化：利用手套，量測出手指的彎曲程度。
2. 資料的分析統計特性：藉由手套讀取手指變化，用以分析指尖的動作模式。
3. 辨識手語：利用我們找出的手指之彎曲度變化及三度空間指尖之移動軌跡，進行機器辨識之運算，辨識出正確的手語。
4. 辨識結果發聲：利用我們辨識出的中文手語，進行辨識結果的發聲。

世界上的聾啞人口眾多，他們因為天生的身體殘疾而造成與人溝通很大的障礙，他們只能使用手語溝通，但懂得手語的人又佔據極少數，他們大多都是使用唇語來和普通人溝通，但這溝通方式常常會讓正常人看不懂，還需要配合比手劃腳，從正確性到溝通的速度都會造成他們很大的不便。此不便利不僅僅造成人際關係的問題，更會影響到工作層面上，使他們擁有較少的工作機會，大多數還只能擔任臨時工，領取較低的薪資並且還沒有受到政府保障。

有了這雙手套，聾啞人士將可以獲得和普通人一樣的溝通能力，能與人面對面的交談，清楚得表達自己想說的話。他們不會再因為溝通上的問題，失去他們應得的

工作機會，可以好好地發惠他們的才能，不會因為天生的缺陷而被忽略，被埋沒。在未來，因為此開發版是使用 arm 的 CPU core，如果可以把此功能搭載到智慧型手機上，將可以使這功能普及並且輕便化，只要戴上這雙手套，再把它利用藍芽連上手機，便可以達成即時口譯的功能。

3、設計創意性

作品的創新概念，凡市面上未曾發表過的功能設計或改良市場上已開發成品之創意構想

手語辨識對聾啞人士而言一直是非常重要、急迫的需求，然而可能是因為需求量不多，根據我們的搜尋，中文市場上一直沒有出現針對中文的手語翻譯，因此為了廣大的中文使用者著想，我們體認到此一創作的必需與緊急。

近年來隨著機器學習、人工智慧風潮盛行，在英語市場上有些商品以影像辨識的技術來偵測手語，然而我們認為影像辨識在手語辨識上存在著極大的缺陷：

其一，手語人士需要雙手完成手語，攝像鏡頭必須是穿戴裝置，然而手語涵蓋的範圍很廣，許多手語甚至需要比到頭上，見下圖，因此攝像鏡頭是無法成功辨識的。

其二，影像辨識非常容易受到背景的影響，例如，只要背景是皮膚色淺色，就很容易辨識錯手部的位置，因此辨識率一直是我們所擔憂的。



圖：狗

因此我們採取穿戴裝置，只使用一雙手套讓配戴更方便，並且配合運動感測器，測量手指路徑變化，如語音辨識一般，準確地分割手語字和字中間的斷點，期望達到穩定的偵測效果。

4、作品應用技術

試以軟體、硬體，與系統整合使用的技術分別論述之，其中包括現有技術使用或自行開發之創新技術

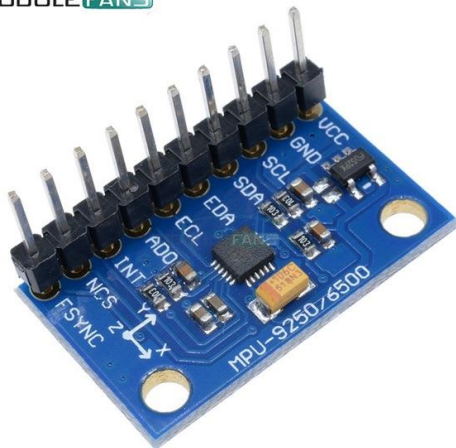
本作品共整合了一個硬體模組、一個傳輸介面、三種演算法、，其功用如下：

硬體模組一：MPU 9250 九軸運動傳感器

MPU9250 是一顆九軸運動傳感器，這裡的九軸與三維空間所講的軸不同，其實是三顆三軸傳感器的合體，分別是加速度計、陀螺儀、磁力計。總之，利用運動時產生的加速度、角速度、磁場強度、就可以反向推算物體實際運動的軌跡。

操作 MPU-9250 晶片時，感應器主要透過 I2C，進行高效能的九軸動態感應。我們透過模組預設的连接介面 I2C 進行操作，透過 STM32F469 控制此模組。

MODULE **FANS**



傳輸介面一：I²C

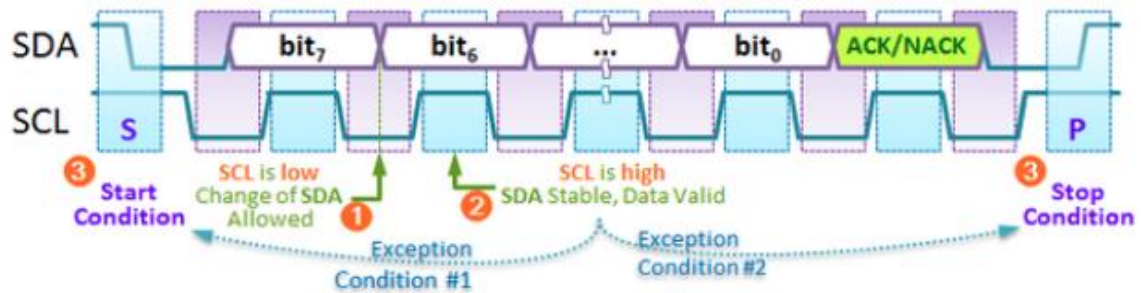
I²C（Inter-Integrated Circuit）是 STM32F469 與 Mpu9250 溝通的介面，I²C 協定設計得十分的精巧。

首先，先介紹實體層，也就是電路訊號的部份：

SCL 為 Low 時，SDA 可以改變資料。如下圖標示 ① 的位置。

SCL 為 High 時，SDA 必需保持訊號穩定，不可以改變，以方便對方讀取(栓鎖)資料。如下圖標示 ② 的位置。

SCL 為 High 時，如果 SDA 有變動則視為特殊狀況：Start (啟始, SDA 由 High 轉為 Low) 或 Stop (結束, SDA 由 Low 轉為 High)。如下圖兩個標示 ③ 的位置。



其次，一般 I2C 晶片常用的傳輸格式有下列二種：

(啟始)-[控制]-[指令]-[資料]-(結束)

(啟始)-[控制 0]-[指令]-(r 啟始)-[控制 1]-[資料]-(結束)

上述格式中，(啟始)和(結束)是前段 "I2C 實體層" 第三點所說的 Start (啟始)及 Stop (結束) 訊號狀態。[控制]，[指令]，[資料] 則都是附有 ACK bit 之 1 組或數組 8 bits 訊號。

總之，我們運用 I2C 的介面以讀取 MPU9250 的資訊。

演算法一：Kalman 濾波器

卡爾曼濾波 (Kalman filter) 是一種高效率的遞歸濾波器，它能夠從一系列的不完全及包含雜訊的測量中，估計動態系統的狀態。藉由卡爾曼濾波，我們可以使九軸運動傳感器得到的資料更加的準確，因為卡爾曼濾波會根據各測量量在不同時間下的值，考慮各時間下的聯合分布，再產生對未知變數的估計。

其含有五大公式

$$X(k|k-1) = F X(k-1|k-1) + B U(k) \dots(1)$$

$$P(k|k-1) = F P(k-1|k-1) F^T + Q \dots(2)$$

$$X(k|k) = X(k|k-1) + K_g(k) (Z(k) - H X(k|k-1)) \dots(3)$$

$$K_g(k) = P(k|k-1) H^T / (H P(k|k-1) H^T + R) \dots(4)$$

$$P(k|k) = (I - K_g(k) H) P(k|k-1) \dots(5)$$

演算法二：資料庫辨識

運用大量的統計資料，推估指尖的運動模式，我們先錄取很多手指運動中、靜止狀態像的資料，利用機器辨識的方法，使 STM32F469 可以推估指尖目前的運動狀態。

演算法三：Dynamic Time Warping

Dynamic Time Warping 稱為動態時間歸準。應用很廣，主要功能是比较不同時間長短資訊的相似性。可能大家學過這些類似的課程都看到過這個演算法。

最經常被使用來衡量距離的方式應該是歐式距離，但是對於一些特殊情況，歐式距離有著極大的不足，尤其是在比較時間序列的差異，比如

A : 1,1,1,10,2,3

B : 1,1,1,2,10,3 ,

如果使用歐式距離，則總和距離應該是 128，差異可說是非常大，然而這兩著的相似度是極高的。就是在這樣的情況下，有人提出了 DTW 演算法，用來比較時間序列上的差異性。

總而言之，我們認為科技研發的重點在於生活應用，為了促進便利、安全與舒適的生活，我們結合 stm32f469 開發板及跨時代機器辨識技術，以弱勢族群的生活為基點，打造智慧化的友善科技。

5、實作設計考量

將作品設計理念轉化為成果之細節規劃與創作過程

我們決定好這個題目後，剛開始我們想嘗試使用影像辨識的方法來實作手語辨識，在使用者的肩膀上架設攝影機。但後來經過討論，因為影像的資料量太大，所拍到的背景不會一致，要在板上實作需要很大的記憶體量和運算量，所以讓我們一度陷入困難，並且認真去思考這個题目的可行性。

經過了一些研究和網路上的資料查詢，我們最終決定使用九軸感應器作為預測手語的器具，接收感應器的訊號。之後我們發現，九軸感應器的訊號沒有我們預期的那麼好取得，並且會有很多的雜訊影響我們的準確度，所以我們使用了卡曼 (kalman) 濾波器，經過實驗發現它所呈現的效果可以顯著，可以把儀器自己產生的誤差給過濾掉。

然而，還有一個很大的困難點是處理斷字的問題，因為我們在比手語的時候，手沒辦法在比完一個字後，真正的完全停下來，又因為儀器非常精準的關係，我們又更難去判斷字和字中間分段的時間點，判斷出之間的斷點對我們整個辨識系統來說是極其重要的。我們第一個想到的方法是把每個手語比的時間固定，例如都預設在一秒，我們就一秒比一個字，一秒過後就送進板子去做手語判斷。此方法非常簡單，但是沒辦法運用在大眾的聾啞人士身上，手語中比每一個字的時間都不一樣，每個人習慣不同比出來的語速也都不盡相同，用這方法做出來的辨識系統一定會限制很多而且不夠強健，和我們想要讓大眾都可以輕鬆使用的初衷不大相同。最後此問題我們使用的是有點像機器學習方法，錄了很多動和停的資料，得出一個判斷模型，用來找出正確手語表達的斷點，進而去分出每一段手語的意思。

並且我們發現，九軸裡面的九種 (三維加速度、三維角速度、三圍磁場) 不同資料也有些會影響我們在判斷停跟動時的準確度。一開始我們使用了六種 (三維加速度、三維角速度) 資料來做，但實驗結果發現非常得不準，前三維就算我們不動也還是會因為每一次停下的角度不同而有所不同。最後我們只使用了三種 (三維角速度) 資料來做判斷，準確率可以達到一百趴。

再來，因為每一次比的手語速度有快有慢，同一個人也不可能每次比完都是取道一樣的資料數量，如果第一次比比較快，取到了 5 筆 data，第二次我們比的比較慢，

取到了 15 筆 data，那這樣就算是比出同一個中文字，把資料拿進去判斷一定會是不準確的。所以我們討論後，使用了我們在學校數位語音辨識課程之中學到的 DTW (Dynamic Time Warping) 演算法，它可以針對不同時間長度下的資料去做比對，並且不會失真。經過我們的測試，不管我們比得慢比得快，它都可以很精準地幫我們做出預測，並且整個演算法是很快速的，不會使我們預測所花的時間變長，一樣可以瞬間翻譯出中文手語。

6、實作結果

作品實際操作之成果，且是否與設計理念相符，又能否完整應用於智慧生活

此項作品總共有兩項計劃

計劃一

1. 以 STM32F469 開發板完成 I2C 通訊介面，與 MPU9250 傳遞資訊，取得第一手九軸資料。
2. Kalman filter: 為了追求更加穩定的資訊，需要對資料做前處理，
3. Movement predict: 先蒐集大量移動與靜止資料，並且偵測手指當下的狀態，確認其是否移動。
4. DTW: 利用動態時間歸準比較當下動作，與資料庫中資訊的相似度，藉此來完成手語偵測。

計劃二

1. 以 STM32F469 開發板藍牙通訊介面，將處理過的資料藉由藍牙傳遞給予任意行動裝置。

目前為止，已經完成計劃一的實作，在此針對目前實作結果分析。

a、手語偵測結果

如下圖所示：兩手皆伸出大拇指與食指，並且從胸前由中間往兩旁移動，就可以聽見儀器發出『大』的聲音



如下圖所示：兩手在胸前，指尖互相輕觸，比出家庭的形狀，就可以聽見儀器發出『家』的聲音



b、字句延遲時間

目前為止，手語使用者完成手勢後，儀器大約會延遲 0.1 至 0.5 秒的時間處理資料，才可以發出聲音，但是用於日常生活中的基本溝通，已經非常充足了。

c、資料庫大小

除了影片呈現的結果以外，我們的詞庫還有許多日常生活常見的字詞，像是：『我』、『的』、『名字』...其餘字句，除了上述兩字以外，我們的詞庫尚有更多的語句，我們手語辨識系統的辭庫還在擴充當中，目前大概是 30-40 字左右，已達可以正常和人溝通的水準。

7、結論

我們根據自身調查，發現對於聾啞人士擁有的問題並加以分析，了解到「手語翻譯」對於他們的重要性。為此，我們結合自身所學的知識（機器辨識、訊號處理）並與 STM32F469 進行整合，開發出一套「iHands 中文手語辨識系統」協助普通人了解聾啞人士欲表達的想法，幫助聾啞人士們走出差別待遇的陰霾下，讓他們得以在社會立足。我們利用九軸感測器接取手部數據，對此進行分析過後，預測出正確的手語意義並發聲出來，而對 STM32F469 的優化還能持續進行，提供更即時、準確的中文手語辨識結果

8、參考文獻

- [1] 台灣手語線上辭典 (<http://140.123.46.77/TSL/>)
- [2] 卡爾曼濾波 (<https://goo.gl/GRJuxM>)
- [3] 簡介 I2C (<https://goo.gl/dQxoD9>)
- [4] 九軸圖片 (<https://goo.gl/h1yb3A>)