

# 美和科技大學

## 100 年度教師專題研究計畫 結案報告

計畫名稱：聲控網頁系統

計畫編號：BMH-100-DIM -001

計畫期間：100.01.01.~100.12.31.

計畫主持人：鄭元飛

共同主持人：鍾德炤、鍾沛穎

研究助理：

經費總額：伍萬元

經費來源：100 年度教育部獎補助款

## 一、題目：

聲控網頁系統 (Sound Controlled Web Page)

## 二、中文摘要：

聲控系統已應用到許多的地方,例如聲控撥號、聲控開機、聲控相框、聲控電燈等,本計劃主要是將聲控系統應用在網頁設計上面,特別是在網頁的超連結上面。聲控本身有它自己的困難度,造成辨識上的困難,例如雜音、相似音等,往往會造成辨識上的錯誤,所以有很多不同的聲控方式,本研究主要是將已有的某一種聲控系統,應用在網頁超連結上面,其聲控方法和傳統上的聲控方法不同,比較能夠應用到網頁的設計上面,希望該設計能提供另一種網頁切換的方式,有別於以往的滑鼠切換、鍵盤切換或觸碰式螢幕的切換方法,提供給身障者另一種的選擇方式。

## 三、英文摘要：

Sound controlled system has been used in many areas, for example, sound dialing, sound switch, sound photo-frame and sound lamp. This project will use the sound controlled system in the web page design, especially in the hyper-link of the web page. The sound controlled method has many difficulties, causes recognizing problems, for example, noises and similar sound. A lot of different methods have been used. This research is to use one kind of sound controlled system in the hyper-link of the web page. This sound controlled system is different from the traditional method, which can be more easily used in the web page design. Hopefully, this design can provide another method to do the hyper-link in the web page, which is different from the traditional mouse click, keyboard and touchable screen, to provide another choice for the disabled people.

## 四、前言：

聲控系統已經出現在很多的地方了,例如聲控撥號、聲控開機、聲控電燈等,但是在傳統的網頁瀏覽切換的超連結上面並未出現。本文主要是介紹一種特別的聲控超連結切換的方式,聲控切換的方法,主要是提供除了滑鼠及觸碰以外的第三種方式,此方法比較麻煩,但對於某些無法使用滑鼠及觸碰的殘障人士,或許可利用此種方式瀏覽網頁。

傳統聲控雖在不受干擾的環境下可獲得不錯的辨識效果，但在實際日常生活的環境下，往往辨識的準確度大幅降低，聲控本身有它自己的困難度，造成辨識上的困難，例如雜音、相似音等，往往會造成辨識上的錯誤。長久以來的研究方向，主要是強化語音訊號及語音參數，以降低雜訊干擾的影響。

有三種語音強化的方向，如語音強化技術、強健性語音特徵及聲學模型調適等[1]。語音強化技術主要目的在於提升語音訊號本身的品質，如頻譜消去法[2]或維爾濾波器[3]等，而強健性語音特徵主要是擷取不易受到環境變化干擾而失真的參數，例如倒頻譜平均消去法[4]或倒頻譜正規化法[5]等。聲學模型調適主要是對乾淨語音所訓練而成的模型，加上特殊音調以調整聲音的機率分佈參數，例如最大事後機率法則[6]及最大相似度線性回歸法[7]等。

本論文所採用的聲控方式和傳統方式不太相同，傳統方式主要是針對聲音的意義進行直接的解讀，而本文所採用的方法，主要是讀取聲音的變動，有如電位的高低不同而產生的二元訊號。利用聲音變動所產生的二元訊號來發出命令。其優點是聲音判讀方式較為單純，主要是利用聲音的大小，來進行二元訊號的傳遞。其缺點則是方式較為繁瑣，所需時間較長。

此種方法並無法取代傳統聲控的模式，主要是提供另一種聲音控制的方式，在傳統聲控無法進行的情況之下，提供了另一種的控制方法，以便在特殊的情況之下可以使用，例如它不限於某種語系，不限於語言的清晰於否，只要有聲音的變動，例如聲音大小或音調的不同，即可進行控制。

在網頁瀏覽方面，傳統上各個網頁的切換，主要是依據各個文字超連結的設定，利用滑鼠或觸碰式螢幕來做超連結的切換，以達到切換到不同網頁的目的。

本文將提供第三種的方法，也就是聲控的方法，來完成超連結切換的動作。此種方式所用的網頁技術是原本就存在的，所以目前所有電腦的瀏覽器，例如 Internet Explorer 或 FireFox 都是可以支援此種聲控的切換方式。

此種聲控網頁的切換方法，其所使用到的現存技術包含了 Java Applet 技術及 Java Script 技術，也就是說，該網頁的超連結切換，只不過是利用 Java Applet 來驅動 Java Script 來完成超連結的動作，而聲控系統的讀取是存在於 Java Applet 內。詳細的系統說明，將在以下論文中介紹。

## 五、研究方法、步驟：

- 1.聲音讀取:聲音轉換為數位資料，電腦讀取聲音有不同的頻率，不同的頻道，不同的儲存大小。可能都需要進一步的比較和測試，以尋求最好的方式。
- 2.聲音辨別:聲音大小及意義的判定。各種不同的讀取設備可能會出現不同的結果，主要的困難度則是要如何分隔不同設備的影響所造成的差異。
- 3.參數測試:各種參數設定對聲音讀取的影響。設備有不同的參數設定，那種參數設定可以得

到最好的結果，可能需要多方面的嘗試。

4.系統架構:實際聲控網頁的設計，可能結合了 html 語法，Java Applet 技術，JavaScript 技術等。

5.Java Applet 程式設計:利用 Java Applet 設計聲音讀取系統，以便嵌入於網頁之中讀取聲音。

6. JavaScript 網頁程式設計:利用 JavaScript 網頁程式來連結 applet 及網頁，以便執行超連結的功能。

7.實際測試:實機測試。可能遭遇的困難則是在不同的電腦上執行，可能其穩定性或正確性會不同。該網頁切換的方式，主要是給身障者使用，然而該種方式是否真的適用於身障者，尚需仔細的評估。

## 六、研究結果與討論:

### 系統介紹

此聲控網頁系統主要分成二大部份:(1)網頁內容及教示部份(2)Java Applet 聲音讀取及控制部份。

#### 網頁內容及教示部份

此系統主要是以聲控完成網頁的超連結動作，所以其主體還是網頁部份。然而為了配合聲控的產生，其網頁設計可能需要符合下列二個條件:(1)最好分割一個框架放置內含 Java Applet 的網頁，而這個網頁最好不要隨著網頁的切換而消失，也就是說利用別的框架作為網頁的切換，而不要動到這個框架的內容。(2)超連結部份需要有一個教示的顯示，這個顯示主要有二個目的，第一，告訴使用者這個地方有一個超連結。第二，引導使用者發出聲音以便驅動這個超連結。

#### Java Applet 聲音讀取及控制部份

這個 Java Applet 是一個聲音讀取及指令控制的一個程式，它內嵌在一個網頁上面以便隨時接收使用者經由電腦麥克風傳來的訊息。使用者可依照網頁上超連結的教示顯示而發出聲音，利用變動聲音的大小產生二元訊號，再利用該二元訊號判定是那一個地方的超連結，而後利用 Java Script 來驅動該超連結以達成網頁切換的目的。使用者只需依據指示發聲或不發聲，就能夠驅動該處的超連結。

#### 網頁超連結控制

網頁的切換主要是依據瀏覽器所提供的超連結功能，該超連結功能主要是利用滑鼠來驅動，

若是觸碰式螢幕，也可利用觸碰的方式來驅動。另外一種的驅動方式則是利用 JavaScript(或 VBScript)這個瀏覽器所支援的網頁程式來驅動。經由寫一段 JavaScript 的程式，瀏覽器可以判讀它並產生超連結的功能。

這個聲控網頁系統，主要是利用 Java Script 的程式來驅動超連結。然而要執行那一種的超連結呢?那就是要依據 Java Applet 所接收到的聲控訊息來判定該執行那一段的 Java Script 指令。

Java Applet 是嵌在網頁上的一個物件，該物件具有硬體操控的能力，也可以驅動 Java Script。而整個的超連結流程，就是使用者發出聲音，Java Applet 控制麥克風接收聲音並判別聲音的訊息以便驅動那一段 Java Script 的超連結。

### **超連結的教示顯示**

超連結中的教示顯示所呈現的二元訊號主要是供使用者遵循並依其頻率發聲或不發聲。各該項目訊號係由一組項目二元碼組成，而該顯示可利用光度、顏色、文字或圖形之差異產生二種相異的顯示型態，以分別表示二元碼中之位元 1 及 0。該二相異的顯示型態係例如光度之「亮」及「暗」、顏色之「深」及「淺」或「白色」及「黑色」、文字「正」及「反」、圖形「○」及「□」等。

由於該顯示需要二種相異的顯示型態分別表示二元碼中之位元 1 及 0，則可藉由多媒體 Flash 軟體、Gif 圖檔、java applet、網頁程式 VBScript 或 JavaScript 撰寫編輯各該選項訊號，各該選項訊號即可透過該電腦螢幕顯示出，以供使用者觀察並由其中選擇一項目訊號發聲。為了隨時教導使用者何時發聲，因此該二元訊號顯示是一直重覆出現的。假如你知道每個選項的頻率及二元號碼，你可以自己獨立的發出二元訊號。例如自己看著手錶依照所知的頻率發出聲音來，理論上就不見得需要眼睛盯著螢幕，但是因為有時選項很多，實際上大部份的人是記不住的，因此還是需要看著螢幕上的顯示。

有時為了讓使用者更加明白其二元訊號，有時這個教示顯示可設計的更加清楚，將所有的二元碼都顯示出來，例如設計成一個類似於時鐘的形態，指針轉向某一個位置時就發聲或不發聲，總之只要能輔助使用者發聲的顯示都是可以的，所以有很多的方式可以設計。

這些教示顯示一般都是放在超連結的前面或本身就是一個超連結的物件，這樣使用者就能很清楚的知道每個超連結所屬的命令為何，並依據這個命令發聲以進行超連結頁面的轉換。

### **螢幕顯示與聲音接收的同步與非同步**

經由該教示顯示所呈現的二元訊號項目，為供使用者下達命令時所遵循並配發給每個需要超連結的地方所使用。假如聲音讀取的 Java Applet 可同時控制教示顯示及處理聲音讀取裝置

所接收的訊息，則可進行同步控制，二元訊號的設計則可以不需要那麼複雜，用簡單的二元碼，例如 10、01、11、00 等，即可進行四個選項的二元碼表示。若是同步的情況之下，對於使用者而言，進行聲控的超連結相對之下會簡單許多，因為所需要的二元碼變化不必那麼多。然而要設計一個顯示與接收同步的網頁系統會使得程式的攢寫變得很複雜，網頁超連結的教示顯示都必須以 Java Script 或 Java Applet 來完成，以便及時的接收指令，而聲音讀取的 Java Applet 也會變得複雜，它除了接收聲音以外，還是隨時對外發出命令改變每個超連結的顯示狀態，這將使得設計網頁變得很難，假如網頁內容要時常更換的話，這並不符合經濟效益。假如電腦效能不好還會使顯示延遲，整個時間控制將會亂掉，造成無法接收的狀況，所以最適當的方式，還是以非同步的控制方式較為適當，不但方便，也較不容易出錯。

在非同步的情況之下，教示顯示的二元訊號項目呈現獨立的狀態，聲音讀取的 Java Applet 則無法了解教示顯示的訊號狀況，所以無法及時的和所接收的聲音訊息互相比較，這時二元碼的設計就會比較複雜了。因為該二元碼會一直重複的循環出現，所以不知目前讀取的位置為何，將無法進行比較。

因此在非同步的情況之下，唯一的辦法就是設計一組可判別循環的二元碼了。這有點像是設計一組一維條碼，必須定義條碼何時開始，何時結束。這樣才能了解讀取的位置，並從開始的位置進行分析比較。所以非同步的二元碼會比較長也比較複雜。表 1 中舉例介紹了供非同步狀況時所使用的二元碼。

以項目 1 為例，其教示顯示狀態為...11011...，當位元 1 出現時則發聲，當位元 0 出現時則不發聲。在程式判讀時，位元 11 則為二元碼比較時的起始或結束位置。為了避免和其它不同位元長度的二元碼搞混，位元 11 的二側可以補以適當長度的 0。除了避免搞混以外，還可以拉長使用者的反應時間，不致造成使用者太大的壓力。

在表 1 的二元碼中，可能會發現中間的位元 1 並沒有連續的狀況，主要是因為在這個例子中，程式是利用連續位元 11 的長度來判定二元碼的開始及結束位置，假如中間出現了連續 11 的情形，程式將無法判別二元碼的開始及結束位置，也就無法判定正確的二元訊號了。

看了表 1 的非同步二元訊號表，可知從 3 位元的長度到了 7 位元的長度，總共只有 12 個項目可以使用。假如在同步的情況之下，理論上 7 個位元總共有 2 的 7 次方減 1 的選項，也就是 127 個選項。所以若是採用同步的方式，二元訊號的設計就不會那麼複雜了，超連結相對就比較容易，選項也比較多。但是在網頁的程式設計就比較困難，因為 Java Applet 除了要控制聲音的讀取之外，還要控制每一個超連結的教示顯示，這種設計可以做到，但並不容易。

表 1 非同步二元訊號表

項 目	二元碼	位元長 度
1	101	3
2	1001	4
3	10001	5
4	10101	5
5	100001	6
6	101001	6
7	100101	6
8	1000001	7
9	1000101	7
10	1001001	7
11	1010001	7
12	1010101	7

## 二元訊號的頻率與設計

二元訊號的頻率關係著使用者發聲的長度，若發聲較長，對使用者來說可能比較累，但相對之下程式的判讀會比較容易，也比較正確。但若是頻率較短，發聲較不累，但節奏可能較快，相對之下程式的判讀會比較難，也比較會不正確。何者是比較適當的頻率呢？目前還未進行分析測試。那一種訊號的設計是最好的呢？也還未知。因為有不同的頻率與不同的二元碼設計，目前還未進行更進一步的統計分析與測試，所以那一種最好目前並不知道。

二元訊號的設計除了表 1 的方式之外，我們也可以使它設計的更為複雜，將自動更正的概念放在訊號內，多增加幾個位元，以便其中一個位元錯誤時可以自動更正。或將位元增加以減輕發聲時的壓力。然而將位元碼長度變長，則頻率必須要加快，不然要完成一個命令必須要等太久也不好，這其中的取捨尚需更進一步的測試與分析。

## 聲音讀取時間延遲

螢幕顯示二元訊號時，因為資料量少，程式簡單或根本只是 GIF 簡易的動畫圖檔而已，比較不會有時間延遲的問題。也就是每個位元時間的顯示設定為 1 秒，則它的顯示就會剛好一秒。但是聲音讀取時，因為資料量較大。假設使用者發聲 1 秒鐘，聲音讀取時間也是 1 秒鐘，但是經 Java Applet 接收後，可能處理的資料量會超過 1 秒鐘，這樣會造成比對上的錯誤，假

如電腦是屬於舊的等級，在讀取這種網頁時，最好關掉其他的工作，以便有最好的效能接收聲音來瀏覽網頁。

### 聲音大小選定

目前二元碼的辨別，主要是依照聲音的大小來做為判別的標準。當聲音大小超過某一個標準時，則位元判定為1，否則為0。就像是高電壓及低電壓的原理所造成的位元處理一樣。然而聲音大小的分辨標準在那裏，目前只是依照背景聲音的大小來做調整，當背景聲音大時，標準就提高，當然使用者所發出的聲音就要更大。當背景聲音小時，標準就可以降低，使用者就可以不用發出那麼大聲了。

依據實際的測試結果發現，現在電腦的聲音接收控制好像有最佳化的設計，例如連續聲音很大時，它會將靈敏度降低，當連續聲音很小時，它會將靈敏度提高，這對於歌曲或語言的錄製或許有幫助，但對於聲控系统而言卻是不好的。因為當你發出連續聲音時，接收器會有鈍化的現象，也就是你要愈來愈大聲，才有相同大小的讀取資料，這對於使用者發聲而言是一個很大的挑戰，因此聲音讀取的頻率不宜過長。但以聲音判讀而言，頻率太短則不容易判讀且容易有錯，要如何取得一個平衡點需要更進一步的測試。

### 傳統聲控比較

和傳統的聲控系统比較起來，使用上較為繁瑣。然而傳統的聲控，可能需要聲音的學習與記錄，或者是需要龐大的語音資料庫來針對不同的語系及不同的語言及音調來做為處理。而目前的聲控方式可以同時應付不同語系及不同人的音調，不需再做聲音學習的動作。

目前的聲控方式無法取代傳統的聲控方式，但可以輔助傳統聲控方式的不足，例如在跨語言的平台上或者是共用軟體上面，可用在大家都可以相同使用的介面上。

### 傳統超連結比較

以一般人的使用而言，目前的超連結方式已經足夠。滑鼠或觸碰式螢幕所驅動的超連結比目前的聲控方式更於方便使用。以目前聲控方式所驅動的超連結不但速度慢需要等待，而且很繁瑣，所以目前的潛在使用者，應定位於無法使用滑鼠或觸碰式螢幕的情況。例如雙手受傷的人或躺在病床上雙手只能看到螢幕而無法觸碰到螢幕的人。

雖然目前的使用看似受限，但卻提供了第三種的操控方式，它並不是取代，而是增加另一種的選擇，它可以和另外兩種方式並存，而不需要更改太多的東西。雖以目前所看到的正常情況之下並無任何的優點，但某些特殊的情況之下，或許可以發揮它的功能。



## 七、結論

聲控系統已經出現在很多的地方了，例如聲控撥號、聲控開機、聲控電燈等，但是在傳統的網頁瀏覽切換的超連結上面並未出現。本文主要是介紹一種特別的聲控超連結切換的方式，聲控切換的方法，主要是提供除了滑鼠及觸碰以外的第三種方式，此方法比較麻煩，但對於某些無法使用滑鼠及觸碰的殘障人士，或許可利用此種方式瀏覽網頁。

雖然目前的使用看似受限，但卻提供了第三種的操控方式，它並不是取代，而是增加另一種的選擇，它可以和另外兩種方式並存，而不需要更改太多的東西。雖以目前所看到的正常情況之下並無任何的優點，但以特殊的情況之下，或許可以發揮它的功能。

此種方式所用的網頁技術是原本就存在的，所以目前所有電腦的瀏覽器，例如 Internet Explorer 或 FireFox 都是可以支援此種聲控的切換方式。只是目前的聲控網頁主要是應用在一般的個人電腦上面，然而目前智慧型手機及平板電腦的普及，該網頁系統要應用在不同的平台上面應該也是可行的。

## 八、中英文參考文獻

1. A. Viikki and K. Laurila, "Cepstral Domain Segmental Feature Vector Normalization for Noise Robust Speech Recognition," *Speech Communication*, Vol. 25, 1998.
2. S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. on ASSP*, Vol.27, No. 2, pp.133-120, 1979.
3. X. Huang, A. Acero and H. Hon, "Spoken Language Processing: A Guide to Theory, Algorithm and System Development," Prentice Hall PTR Upper Saddle River, NJ, USA, 2001.
4. S. Furui, "Cepstral Analysis Techniques for Automatic Speaker Verification," *IEEE Trans. on ASSP*, 1981.
5. A. Viikki and K. Laurila, "Cepstral Domain Segmental Feature Vector Normalization for Noise Robust Speech Recognition," *Speech Communication*, Vol. 25, 1998.

6. J.L. Gauian and C.H. Lee, "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," IEEE Trans. on Speech and Audio Processing, 1994.

7. C.J. Leggetter and P.C. Woodland, "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," Computer Speech and Language, 1995.