



文化事業與管理研究
第 五 期
2 0 1 0 年 6 月
頁 1 ~ 2 2 頁

Journal of Cultural Enterprise
and Management No.5
June, 2010
P. 1~22

資訊站語音使用介面研究

楊聰仁* 蔡明勳**

摘 要

本文主要探討語音介面的設計，我們將互動式資訊站加入語音辨識技術來設計一個使用者介面，並將此語音介面系統運用在博物館的導覽上面。系統透過微軟 SAPI 語法的撰寫與 XML 文法的設計，讓使用者透過簡單的語句來取得博物館的導覽資訊系統，系統會先判斷使用者說出的語句並引導使用者到所需要的服務內容。研究者在系統完成之後，透過放聲思考法的運用，讓使用者一邊試驗系統一邊回答問題，針對系統的優缺點等相關問題提出見解，設計者收集這些資料更改系統並將語音輸入與觸控式輸入做相關比較，並使用 Z 檢定的驗證來證明我們所設計的系統有更佳的效率。

關鍵詞：語音辨識、多媒體互動系統、語音使用者介面、博物館導覽

*南華大學出版與文化事業管理研究所 所長

**鴻海科技集團 研發工程師



Speech Interface on Kiosk Navigation

Tsung-Jen, Yang Ming-Shun, Tsai

Abstract

We design a Chinese voice user interface for Kiosk of museum in this paper. We use the Microsoft Speech SDK as engine and XML Grammar to get higher recognition rate. We also improve the voice control and design new XML Grammar to let user to speak natural Language. It showed that the voice user interface navigation system provides a more user friendly and enhances the effectiveness and efficiency for museum navigation. We also conducted an evaluation on this system. It turned out that the voice user interface was better than the touch user interface in the museum's Kiosk navigation.

Keywords: Speech Recognition, Voice User Interface, Museum Navigation, XML Grammar



壹、緒論

自然語言是人類最習慣的溝通方式，也是與機器最自然的使用介面，提供資訊的資訊站(Kiosk)如能接受語音的使用操作，將可提供使用者更佳的使用經驗。

語音辨識的目的在於讓電腦可以聽懂人們的語句輸入，當電腦可以聽懂人們說的話之後就可以命令電腦執行相對應的工作。聲音藉由類比訊號轉變到數位訊號裝置輸入電腦，並以數值方式儲存後，語音辨識程式便開始以事先儲存好的聲音樣本與輸入的測試聲音樣本進行比對。比對完成後電腦得出一個它認為最“像”的聲音樣本序號，我們就可以知道使用者剛剛輸入進去的聲音代表何意，進而命令電腦做事。

博物館導覽系統的開發已有多多年，從早期的紙本 DM 發放、志工協助導覽、耳機配合錄音帶導覽，轉變成以電腦或是互動式多媒體資訊站(Kiosk)為主的數位導覽方式，提供使用者資訊的查詢與服務、館區的介紹與導覽服務，作品的說明簡介等。黃光男(1998)認為博物館是一個獨立完整的運作空間，博物館的參觀族群，涵蓋了各種年齡、性別、族群與職業，因此在此環境下提供的一種自由性的、自願性的學習活動。

隨著人工智慧技術演進，語音辨識技術的應用也越趨成熟，使用者經由語音辨識技術就能以口語與電腦進行互動。近年來隨著資訊化程度越來越高，很多政府機構都導入資訊化的服務，如建置服務網頁、資料處理電腦化、建置 Kiosk 導覽系統(互動式多媒體資訊站)等，讓一般使用者對於想知道的資訊取得更容易，也可以節省機構對於人事費用的開銷，本篇研究主要是探討將語音辨識技術應用在博物館的資訊導覽，設計一個新的使用者介面(UI, User Interface)導覽系統，並與現有的觸控式 Kiosk 系統作效能比較，以瞭解語音介面的效能。

本研究以現有博物館 Kiosk 系統—「科技台灣驚嘆號」為平台，加入語音辨識之技術，來探討使用者對於現有 Kiosk 系統與加入語音辨識之 Kiosk 系統的滿意度調查，與加入語音辨識後，觸控式系統與語音辨識系統完成一個任務所使用的時間比較，並透過問卷方式調查使用者對於語音使用者介面的接受度如何，與觸控式的介面相互比較，並利用統計方式取得使用者對於語音使用者介面的滿意度。



貳、文獻探討

語音的使用者介面主要以語音辨識(Speech Recognition)、XML 文法設計(XML Grammar)、語音使用者介面(Voice User Interface)、語音合成發音(Text To Speech)與互動式多媒體資訊站(Kiosk)等相關技術所組成。

一、語音辨識

利用電腦技術來處理語言問題，是現代科技中一項重要的發展，其最終目的就是希望人類可以與機器利用自然語言(Natural language)來做溝通，在許多科幻電影中常常可以看到機器人可以知道人們所表達的語句，進而去完成人們所要機器人去完成的目標，這其中牽扯到許多科技技術的應用，語音辨識技術也是其中一個重要的指標。利用裝置將類比的聲音轉變成數位訊號讓電腦可以讀取並做出指令，每隔一段距離取得一組樣本(如圖 1)，並將此類比樣本資料經過二進制轉換，變成數位訊號以便電腦讀取，經過轉換之後再將訊號送到辨識引擎裡面作機率分析，引擎會根據訊號的傳遞選出機率最高的辨識字元。

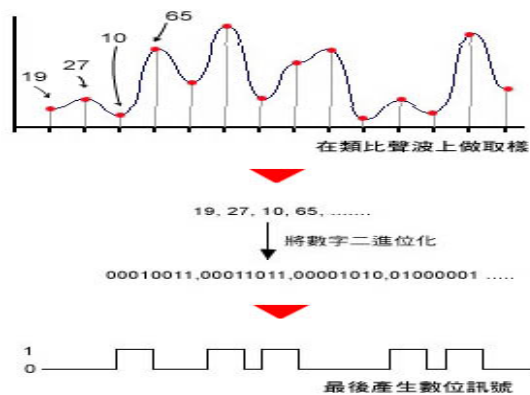


圖 1 類比訊號轉換成數位訊號圖

以目前有的語音辨識引擎而言，開發系統時候比較常用到的有 Microsoft Speech SDK、IBM ViaVoice、VoiceXML 與飛利浦的辨識引擎，通常會視需求來選擇應用的辨識引擎，因為採用 Visual Basic 來開發系統，將語音辨識之技術加入觸控式 Kiosk 導覽系統，實做一個新的使用者介面，讓使用者透過聲音來控制並取得所需要的服務資訊，但由於系統處理使用者的自然語言發音(NLP, Natural Language Processing)所需要的語料庫過於龐大，會造成辨識率(Recognition Rates)大幅下降，因此需定義其辨識的文法(XML Grammar)，將可能出現的關鍵字(Key words)輸入 XML Grammar 中，提升語音辨識過程中的辨識率。

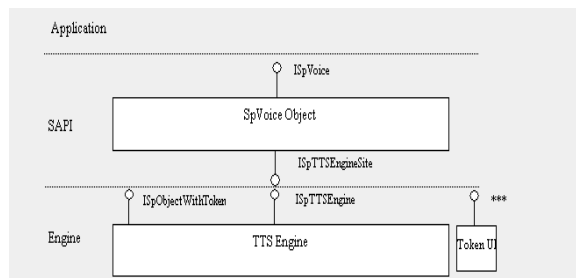


圖 2 語音辨識引擎示意圖

語音辨識(ASR)的處理（廖峻峰，2003）包含：1.使用者輸入(User Input)、2.訊號處理(Digitization)、3.訊號轉換(Phonetic Breakdown)、4.統計模型(Statistical Modeling)、5.對應程式(Matching)。所以，根據以上列表並利用 SAPI 進行系統開發，如圖 2 所示，當使用者說出關鍵字之後，Application 端會將關鍵字等傳送到 SAPI 端作處理，如果需要電腦合成語音的話，TTS Engine 就會啟動，將合成之後的聲音傳送到 SAPI 等待被使用，依照這個循環，程式與 SDK 引擎互相配合而完成本研究的導覽系統。

二、語音合成發音

語音合成又名文句轉語音，是指將輸入的文字或儲存於電腦中的文件模擬人聲發出語音的技術。研究顯示語音使用者介面回覆使用者答覆的時候可以有兩種方法(Dirk, Fernando & Tao, 2005)，一是使用預錄的聲音(Prerecorded Sound)、另一個是利用語音合成技術(TTS, Text To Speech)。使用預錄聲音可以預先錄製真人的發音，讓使用者介面回應時不會有呆板的機器聲，不過缺點是不能靈活應用，因為只要服務內容一改變，錄製的聲音檔案便不能使用，如果要再錄製的話，所花費的時間、人力、物力都會相當的多，



所以本系統採用讓電腦自己合成聲音檔的方式，利用語音合成方法來發聲，雖然聲音聽起來沒有預錄的聲音優美，但是在符合經濟成本的考量下，可以隨時更改的語音合成比較適合。另外，語音合成技術還有以下優點：1.發音的自然度（清晰、流暢）、2.破音字的處理、3.即時處理的能力。透過語音合成技術的支援，可以減低系統開發的複雜度，而語音合成發音的做法有以下（王小川，2005）：

1. 頻譜參數合成方法(Articulatory Synthesis)

2. 波形拼接法(Formant Synthesis)

3. 串接合成法(Concatenated Synthesis)

對於 TTS 系統而言，無論接受的是一段文字的輸入或是一篇文章，這些文字本身並沒有包含任何聲學特性（說話的聲調，停頓方式，發音長短等韻律），只有語言學的特性，所以必須透過自動預測的機制來產生這些文字可能的聲學特性(acoustic feature)。而所謂自動預測的機制，一般有 rule-based 跟 knowledge based 兩種方法，但是這兩種方法合成的聲音不但平淡又缺乏吸引力且遇到連續發音或要保留語者音色時表現都不好，因此近來串接合成法很受歡迎。另外，目前在語音科技技術的應用上，語音合成技術已廣泛被應用在以下各領域，如：個人電腦輔助、數位內容聲音快速生產、無障礙網頁發音、手機、PDA 朗讀即時資訊、電話查詢系統、電話通報、語音叫號、生產操作提示、收銀盤點、家電、玩具、導覽系統。所以，本系統將語音合成技術運用在語音導覽上面，讓使用者在瀏覽系統時候，透過 Kiosk 平台操作，系統會透過發聲裝置將語音合成的聲音表現出來，之後有需要修改瀏覽內容時候，只需要改變 TTS 的文字內容即可，不需要再大費周章錄製聲音。

三、XML 文法設計

因為語音辨識之語料庫過於龐大，因此我們透過撰寫 XML 語法來提升語音辨識的辨識能力，如下圖 3。將導覽過程所需用到的字彙統整並加以處理寫成 XML 可執行的語法，並利用 SAPI 去呼叫此 XML 文法，XML 被呼叫後會產生一個“.cfg”暫存檔案，經過定義過的編碼，可以讓辨識率增加。



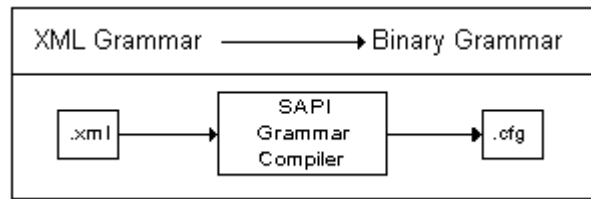


圖 3 XML 文法與主程式之間的關係

根據 Hirohiko Sagawa et al. (2004)所提如何規劃一個完善的 VoiceXML Grammar，如下圖 4，當使用者說 “I said”，之後可以說出 Tokyo、tomorrow 與 Tokyo tomorrow 等，也就是說衍生出來的文法有：

1. I said Tokyo.
2. I said tomorrow.
3. I said Tokyo tomorrow.

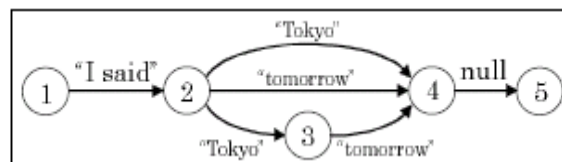


圖 4 Grammar 設計圖

使用者只要說出上列三種之一的片語，VoiceXML 語言皆可判斷成想問東京明天的天氣，透過上述文法的組合，即可變化成多種組合以供系統呈現。而 XML 的設計裡面與 VoiceXML 大同小異，彼此之間的差異只在定義上面不同，還有 XML 文法本身並沒有執行的能力，必須配合主程式去呼叫才能有所作用。

四、語音使用者介面

聲音有著以下的限制(Dirk Schnelle et al., 2005)：1.聲音是短暫的，加上使用者的短暫記憶(STM, Sort Term Memory)，當你聽到的時候，它已經消失了。2.聲音是一種看不見的



東西，並沒有畫面的呈現。3.聲音有著不勻稱的特性，說話比打字還快速，但是聆聽又比閱讀慢。因此，為了解決聲音在先天的缺失，我們在系統中加入一個使用者介面，除了聲音之外，還有一個可以供使用者觀看的畫面，改善聲音的限制與讓使用者可以依循畫面的指引完成所需要的服務。文獻指出，一個語音使用者介面(Voice User Interface)需要有以下幾個條件，才可以稱作好的語音使用者介面(Dirk Schnelle et al., 2005)：1.提供選單(Menu)，讓使用者可以透過畫面來改善短暫記憶的問題。2.提供求救功能(Help)，使用者在瀏覽系統期間，可能會迷失方向，所以提供功能讓使用者可以隨時隨地說出 Help 指令，提醒使用者在這麼階段可以做什麼事，可以說出什麼指令來與語音介面互動。如下圖 5，使用者在讀取 Info1 與 Info2 時候可以隨時使用 Help 指令。

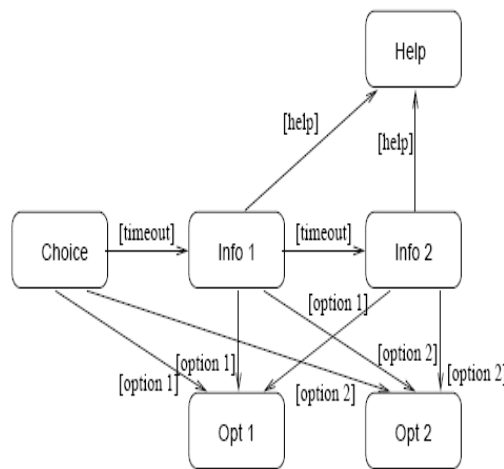


圖 5 使用者使用 Help 指令圖

3.線性選項(Linear Choose)，當使用者對於選取所需要的服務選項時，可以快速的取得需要的資訊，也可以透過下一步(Next)或是上一步(Previous)來控制。還有，在設計一個使用者介面時候，也需要考慮到一些例外狀況，因為靠聲音來控制系統，所以使用者在使用系統途中可能會有迷失方向、語音辨識錯誤、或是使用到一半使用者中途離開等狀況發生，所以需要加入一些例外狀況來防止系統長時間等待或是出現錯誤狀況，分別為以下五點(Dirk Schnelle et al., 2005)：1.當使用者沒有注意聆聽系統說明或是不知道該發出什麼聲音時候，等待時間(timeout)的指令這時候就可以發揮效用，經過一段時間系統沒有收到任何指令或是聲音時候，系統會參照等待時間的控制程序進行反覆說明(repeat)或是執行



下一個動作功能。2.為了防止系統出現錯誤情況(error)發生，有經驗的介面設計者通常會強迫使用者聽完冗長的說明，當使用者聽完說明之後會比沒有聽過說明的使用者更能熟悉系統的操作過程。3.使用者在使用系統期間，可能會發生不知道該如何使用或是不知道該說些什麼關鍵字來控制系統，這時候系統提示詳細的說明(more detail explanations)就可以幫助使用者在這個階段完成取得需要的服務。4.不同的錯誤情況與不同的等待時間有著不同的重要性。5.過低的辨識成效(Low recognition performance) 與背景噪音(Background noise)容易造成錯誤的發生。另外，文獻也有提到，在語音使用者介面下需要有以下步驟，可以幫助使用者瀏覽系統(Daniela Oria et al., 2005)：幫助(Help)、列表(List commands/options)、主選單(Main menu)、退回上一頁(Go back)、重覆(Repeat)、待命(Standby)、運算(Operator)。所以在設計語音導覽系統時候，我們將「幫助」、「主選單」、「退回上一頁」等功能加入，將以上條件考慮進去，進而完成現在的語音 Kiosk 導覽系統，系統剛完成時候一定會有欠缺考慮的細項，所以完成系統後先進行測試，透過使用者的角度來改善系統的流暢度與親切度。

五、互動式多媒體資訊站

Kiosk 的原義是指售票亭，在本研究內可以解釋為資訊服務站，它通常是由工業級的電腦組成，加上可以隨使用環境不同而變化的軟體程式，將可以提供在不同環境下達到多元的服務，本實驗將 Kiosk 加入語音辨識功能並運用在博物館導覽系統。由於 Kiosk 的應用方面很廣，這本篇研究裡面我們將 Kiosk 應用在博物館資訊的導覽，當使用者進入博物館時候，可以透過我們的 Kiosk 系統做預先取得基本導覽資訊的服務，例如，使用者想知道半導體資訊的展區，在 Kiosk 導覽裡面我們提供使用者樓層資訊，或是，使用者可以在 Kiosk 平台上預先取得接下來要參觀的展區的資訊，透過 Kiosk 可以知道展區特展的內容，因為 Kiosk 建置簡單、可以節省許多人事成本，加上可以因應使用環境來做細部調整，故我們把它當作實驗平台。一個直覺式反映的 Kiosk 使用者介面，需要以下功能(L.Lamel et al., 1998)：

1.容易在任務(task)間作切換。2.圖片的切換，讓使用者知道系統目前狀況。3.兩階段的協助(Help)指令，當系統待命超過一段時間就發出指令，確保系統不會因為待機而當機。所以我們在設計語音 Kiosk 系統時候，也加入 Help 指令，不過系統的 Help 指令是引導使用者在這麼階段可以做些什麼事情，例如：回到上一頁、往下一頁等資訊，當系統確定收到使用者發出的指令之後，會將瀏覽資訊以網頁型式呈現出來，讓使用者可以知道系統確實收到指定並作出動作。



參、系統設計

本研究主要是設計一個新的語音使用者介面來瀏覽博物館資訊，透過此語音辨識系統，使用者可以簡單的說出需要的服務項目、關鍵字，即可將資訊等服務呈現在使用者面前。

一、系統架構：

本研究著重於系統的設計，在原有的觸控式 Kiosk 導覽系統上再增加一個語音使用者介面，如下圖 6，使用者透過 Kiosk 裝置說出需要的服務，系統會將使用者說出的語句傳送到語音辨識引擎(Speech Engine)，之後透過自動語音辨識來(ASR, Automatic Speech Recognition)處理，ASR 會尋找 XML 的語料庫文法(XML Grammar)比對語句，從語料庫裡面抓到對的關鍵字(Keyword)傳回系統，作處理，進而將使用者需要的資料以網頁方式傳回到 Kiosk 平台來呈現。

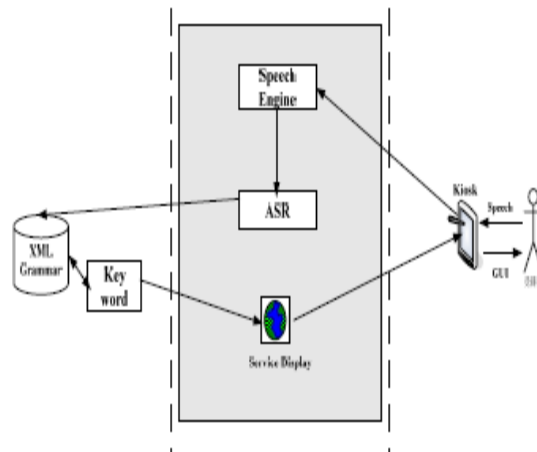


圖6 使用者資料流向圖

加入新的語音使用者介面可以讓使用者利用自然語音輸入取代傳統觸控，使用者也可以透過螢幕的畫面呈現、網頁的提示、語音提示等，可以知道自己在這麼階段可以做些什麼，在系統中，使用者透過 Kiosk 平台觀看需要的資訊時候，網頁下面會有跑馬燈(紅線標示)提示使用者可以說出什麼關鍵字來驅動下一個服務，來完成需要的使用者的資訊取得。如下圖 7，跑馬燈出現「在這邊你可以由左上角的選項說出你要的服務，如：



半導體等」，表示使用者可以透過以上提示說出需要的服務，避免使用者在瀏覽系統時不知道該說出什麼關鍵字而發生卡住的狀況。

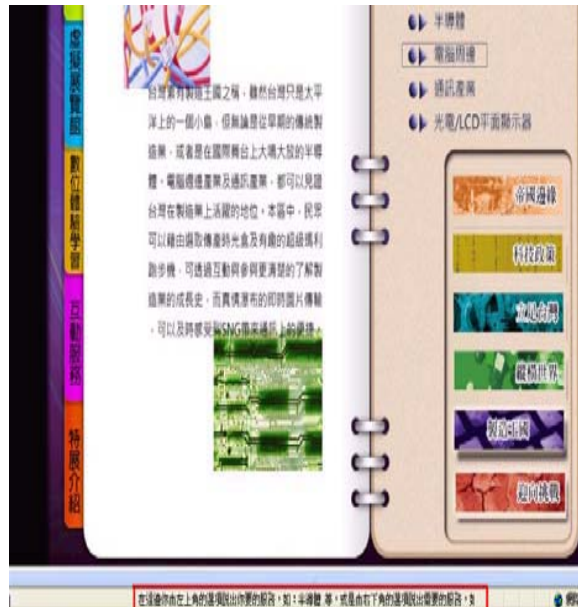


圖 7 網頁使用者提示圖

當使用者使用 Kiosk 系統導覽時候，其目的是為了取得資訊內容的服務，所以下圖 8 表示使用者使用系統時候可以取得到的資訊內容，例如使用者想知道半導體的內容，如：晶圓製造等資訊內容的話，利用傳統觸控式需要從製造王國=>半導體才可以取得相關資訊，但是使用語音使用者介面的話，直接在開始畫面說出需要的服務內容，即可以取得資訊。



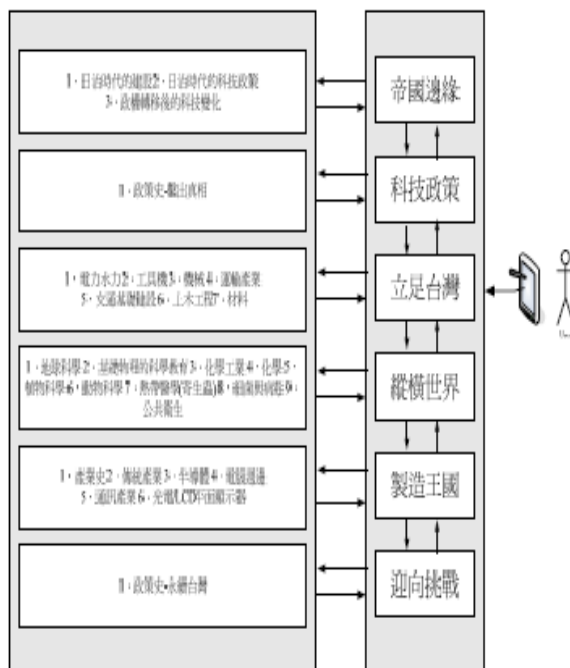


圖 8 博物館資訊圖

二 可延伸標記語言設計：

在 XML 的文法設計中，辨識過程其實就是求出音框對狀態關係的過程，最後求出一個最大機率的路徑（王小川，2004），如下圖 9。果只是單純的辨識關鍵詞的話，那麼只要遵循狀態與音框關係的時間規則，就能求出一條具有最大機率值的路徑。

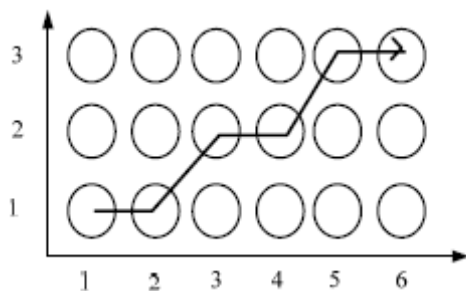


圖 9 語音路徑選擇圖之一



在單純關鍵詞的辨識過程中，狀態轉移的過程也比較單純，如下圖 10，從“*I would like to drive from Seattle to New York*” 文法中可以看出眾多選擇，例如：Method 的子選項中可以有 Fly 與 Drive 兩種選擇，Direction 則有 From 與 To 選項，City 的子選項有 Seattle、New York、Los Angeles 與 Albuquerque 四種選擇，依照機率值來選擇最佳路徑，所以系統辨識出 *I would like to drive from Seattle to New York* 的選項機率值為 128 分之 1。而 XML 文法設計範圍越大，則有越多的選項以供選擇。

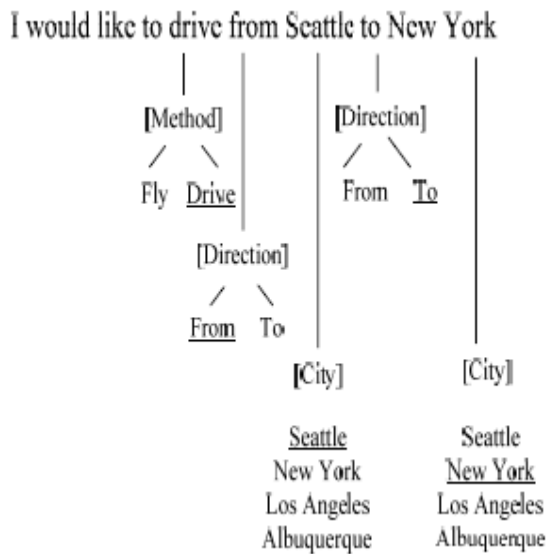


圖 10 語音路徑選擇圖之二

如圖 11，是系統 XML 設計圖，它是參考上圖 10 來設計此系統的 XML 文法，設計者透過 XML 的設計可以讓語音導覽系統更加多元化，也讓使用者可以利用自然語言來瀏覽系統。



```
- <L propname="name0">
  <!-- 產業史 keywords -->
  <p val="1">產業歷史</p>
  <p val="2">製造歷史</p>
  <p val="3">發展史</p>
  <!-- 傳統產業 keywords -->
  <p val="4">傳統</p>
  <p val="5">鞋業</p>
  <p val="6">食品</p>
  <p val="7">紡織</p>
  <!-- 半導體 keywords -->
  <p val="8">晶圓</p>
- <p val="9">
  <o>微體</o>
  電路
</p>
<p val="10">矽</p>
<!-- 電腦週邊 keywords -->
<p val="11">印表機</p>
<p val="12">鍵盤</p>
<p val="13">滑鼠</p>
<!-- 通訊產業 keywords -->
<p val="14">手機</p>
<p val="15">大哥大</p>
<p val="16">黑金剛</p>
<p val="17">數據機</p>
<p val="18">電纜</p>
<!-- 光電 keywords -->
<p val="19">映像管</p>
<p val="20">顯示器</p>
<p val="21">液晶螢幕</p>
```

圖 11 XML 文法設計圖

三 開發工具：

1. 硬體：

- (1) 電腦主機：3H 筆記型電腦一台
- (2) 測試平台：HP 平板電腦
- (3) 高感應式麥克風

2. 軟體：

- (1) Microsoft Speech SDK 6.1
- (2) Microsoft Visual Studio 2005
- (3) XML Grammar Compliance Tool



- (4) SQL Server 2005 Express
- (5) Dreamweaver 2004

肆、效能評估

本實驗總共分為兩個階段來測試，第一階段是語音瀏覽程式剛完成時候，讓使用者使用語音 Kiosk 系統來做測試，透過使用者實驗過程中，設計者利用放聲思考法來針對系統本身流程與優缺點等問題提問，讓使用者說出自己對系統的看法，設計者再透過收集使用者想法的資料來當做參考，進而更改語音程式內容或是流程，以符合貼近容易使用與親切的使用者介面系統。

1. 第一階段實驗：

第一階段實驗採用放聲思考法來測試，實驗一開始時候，邀請幾個研究生來幫忙測試系統，透過研究生親自的流程實驗，並利用放聲思考法來實驗，在測試系統的同時並講出系統的優缺點，以下是針對第一階段實驗者的說明與對話紀錄，圖 12 表示第一階段實驗之流程規劃。

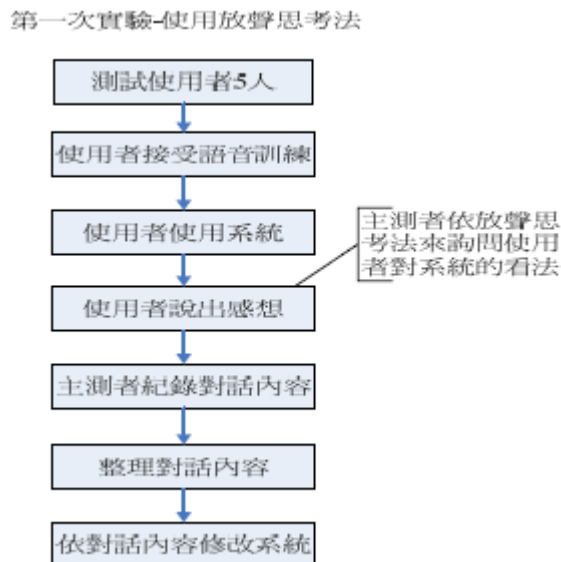


圖 12 第一階段實驗流程圖



透過使用者對系統的建議，我們整理成下表 1，根據以下建議我們來修改系統的細項，希望可以更符合使用者的需求。

表 1 使用者意見整理表

使用者一：希望加入可以隨時插話的機制。
使用者二：希望回應速度與辨識能力再加強。
使用者三：希望服務開啓時候可以更快。
使用者四：希望服務開啓可以快一點、辨識率有點不高。
使用者五：希望服務可以多一點。

2. 第二階段實驗：

在時間效能檢定方面，我們進行了第二階段實驗，第二階段實驗內我們邀請二十位受測者來進行改良後系統的實驗，一開始我們將受測者採取隨機抽樣的方式來分組，分別為語音使用者介面組與觸控式使用者介面組，而每一組人數為 10 人，分組之後，第一組語音使用者介面組 10 位受測者分別測試語音線性任務與語音非線性任務；第二組觸控式使用者介面組 10 人分別測試觸控式線性任務與觸控式非線性任務，所收集到的時間花費表如表 2。



表 2 使用者在各個任務底下花費時間表 單位(秒)

編號	線性觸控任務	非線性觸控任務	觸控式平均	線性語音任務	非線性語音任務	語音平均
1	35	95	65	36	33	34.5
2	42	112	77	34	31	32.5
3	38	127	82.5	23	36	29.5
4	36	103	69.5	28	28	28
5	42	116	79	22	31	26.5
6	43	124	83.5	43	49	46
7	41	101	71	24	27	25.5
8	31	132	81.5	31	39	35
9	42	119	80.5	22	36	29
10	36	98	67	24	23	23.5
平均	38.6	112.7	75.65	28.7	33.3	31

如表 2 所顯示，在觸控式使用者介面輸入的平均為 75.65 秒，而語音使用者介面輸入的平均為 31 秒，這表示語音的控制介面可能比觸控式使用者介面在操控介面以取得資訊的時間上較少。我們進行 Z 檢定來分析觸控式使用者介面所花費的時間與語音使用者介面所花費的時間是否達到顯著差異，我們訂定 μ_1 為觸控式任務輸入時間平均數，也就是（線性觸控式輸入時間+非線性觸控式輸入時間）÷2，而 μ_2 為語音任務輸入時間平均數，也就是（線性語音輸入時間+非線性語音輸入時間）÷2，另外我們設定虛無假設與對立假設為以下：

$$H_0 : \mu_2 - \mu_1 \geq 0$$

$$H_1 : \mu_2 - \mu_1 < 0$$



表 3 95%信賴區間下 Z 檢定分析

兩個母體平均數 差異檢定	語音介面組	觸控介面組
平均數	31	75.65
已知的變異數	47.44722222	41.94444444
觀察值個數	10	10
假設的均數差	0	
z	-14.93388993	
P(Z<=z) 單尾	0	
臨界值：單尾	1.644853627	
P(Z<=z) 雙尾	0	
臨界值：雙尾	1.959963985	

表 4 99%信賴區間下 Z 檢定分析

兩個母體平均數 差異檢定	語音介面組	觸控介面組
平均數	31	75.65
已知的變異數	47.44722222	41.94444444
觀察值個數	10	10
假設的均數差	0	
z	-14.93388993	
P(Z<=z) 單尾	0	
臨界值：單尾	2.326347874	
P(Z<=z) 雙尾	0	
臨界值：雙尾	2.575829304	

由表 3 與表 4 的資料顯示，Z 值為-14.93388993，小於臨界值，所以拒絕 H0 接受 H1，也就是說觸控式輸入所花費時間顯著大於語音輸入所花費時間，利用 Z 檢定分析顯示透



過語句來控制瀏覽流程，語音的使用者介面在操控 Kiosk 平台的資訊取得任務所需的時間，顯著少於觸控式使用者介面。也就是說語音的使用者介面在 Kiosk 平台上比觸控式的使用者介面更具時間的效益。

根據使用者對系統的意見作為參考並修改系統服務內容，希冀可以讓語音辨識的服務真正加入 Kiosk 系統之中，這樣使用者可以透過新的服務方式來認識並取得需要的資訊。透過此系統的呈現，讓使用者利用語音來控制系統，並在使用者測試後參照使用者意見加以修改系統，以便更能符合需求。在實作完系統之後發現，其實語音辨識技術之應用在台灣其實沒有很普遍，希望這篇文章的完成可以對語音辨識技術之應用能有一些的貢獻，也希望能幫助後來研究的人能有一些幫助。

參考文獻

- Adam S., Kandarp M. (2003). *Design and Evaluation of Voice User Interface*.
http://facweb.cs.depaul.edu/astele/research/Bibliography/Steele_MSEC2003a.pdf, 資料參考時間：2007 年 8 月。
- Dirk S., Fernando L., Tao W. (2005). *Audio Navigation Patterns*.
<http://elara.tk.informatik.tu-darmstadt.de/publications/2005/AudioNavigationPatterns.pdf>, 資料參照時間：2007 年 7 月。
- Hataoka, N., Y. Obuchi, T. Mitamura, E. Nyberg (2004). *Robust Speech Dialog Interface for Car Telematics Service*. Consumer Communications and Networking Conference, 331-335.
- James A. Rodger, Parag C. Pendharkar (2004). A field study of the impact of gender and user's technical experience on the performance of voice-activated medical tracking application. *HCI Issues in Mobile Computing, Volume 60, Issues 5-6, 529-544*.
- L. Lamel, S. Bennacef, J.L. Gauvain, H. Dartigues, J.N. Temem (2002). User evaluation of the



MASK Kiosk. *Speech Communication*, Volume 38, Issues 1-2, 131-139.

M.J. T. (2006). VoiceXML dialog system of the multimodal IP-Telephone – The application for voice ordering service. *Expert Systems with Applications*, Volume: 31, Issue: 4, November, 684-696.

Microsoft Speech SDK. <http://www.microsoft.com/seppch/download/sdk51>, 資料參照時間：2007年10月。

Oria, D. & Vetek, A. (2005). Automatic Generation of Speech Interfaces for Web-based Application. *Automatic Speech Recognition and Understanding*, Issue 27, Nov.-1, 39-402.

Sagawa, H., T. Mitamura, E. Nyberg (2004). *Correction Grammars for Error Handling in a Speech Dialog System*. Proceedings of HLT/NAACL 2004, Boston.

W3CVoiceXML. <http://www.voicexml.org>, 資料參照時間：2007年11月。

Yandong F., Elizabeth A. K. (2005). A Hybrid Dialogue Strategy for Speech-enabled Mobile Commerce. *e-Business Engineering*, Issue 18-21, 110-117.

Zhang W., Duffy V.G., Linn R (1999). Voice Recognition Based Human-Computer Interface Design. *Computers and Industrial Engineering*, Volume 37, Number 1, 305-308.

王小川 (2004)。語音訊號處理。初版。台北：全華科技圖書股份有限公司。

朱孝國。朱孝國的筆記本。取自 <http://irw.ncit.edu.tw/peterju/>，資料參照時間：2007年12月13日。

朱育民。應用語音技術建置一個國小英語教學之學習護照系統。取自 http://thesis.lib.ncu.edu.tw/ETD-db/ETD-searchc/view_etd?URN=9152400，資料參照時間：2007年12月。

林士翔，陳柏林 (2007)。現階段強健性語音辨識技術研究之簡介，Web2.0 技術與應用研討會。國立台灣師範大學資訊工程學系，計算語言學學會通訊(ACLCLP Newsletter)專



文，第十六卷第四期，頁 8-15。

科技台灣驚嘆號—2007 科學季刊。取自 <http://www.2007science.com.tw/vm/Default.html>，資料參照時間：2007 年 10 月。

曾鈺涓（2005）。**經驗無線-博物館數位導覽系統案例研究**。數位設計研討會。台中：國立臺中技術學院。

黃光男（1998）。**博物館行銷策略**。再版。台北市：藝術家出版社。

廖峻峰（2003）。**語音技術與 VoiceXML 應用**。取自 http://www.try.idv.tw/try/talks/VoiceXML_Report.pdf，資料參照時間：2007 年 11 月。

賴鼎陞、高淑惠、黃雅惠（2004）。**博物館數位導覽系統建置與使用者評估**。取自 <http://datf.iis.sinica.edu.tw/Papers/2004datfpapers/post/7.pdf>，資料參照時間：2007 年 11 月。

謝金順、張武成（2003）。**線上遊戲軟體設計因素與使用者滿意度關聯之研究**。**資訊管理展望**，第五卷，第 1 期。

