

行政院國家科學委員會專題研究計畫 成果報告

子計畫二：以 VoiceXML 為基礎的個人化資訊廣播站(I)

計畫類別：整合型計畫

計畫編號：NSC91-2213-E-004-012-

執行期間：91年08月01日至92年07月31日

執行單位：國立政治大學資訊科學系

計畫主持人：廖文宏

共同主持人：曾國峰

計畫參與人員：賴建安 李黛雲 孫新民

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 92 年 10 月 31 日

行政院國家科學委員會專題研究計畫成果報告

以 Web 為基礎的數位媒體服務與運作平台之研究發展

子計畫二:以 VoiceXML 為基礎的個人化資訊廣播站(I)

VoiceXML-Based Personalized Information Broadcasting System (I)

計畫編號：NSC 91-2213-E-004-012-

執行期限：2002 年 8 月 1 日至 2003 年 7 月 31 日

主持人：廖文宏 國立政治大學資訊科學系

計畫參與人員：賴建安 李黛雲 孫新民 國立政治大學資訊科學系

一、中文摘要

本篇報告描述以語音介面為設計基礎並結合網路科技、資料庫應用所發展的個人化資訊廣播服務平台及使用者語音瀏覽介面(Voice User Interface, VUI)。前者可以透過多樣化的方式(包含廣播、串流式媒體與電話線路)獲取量身訂做的資訊。主要目標包含(1) 從網路上大量的中英文新聞中，根據個人偏好設定，即時的下載與朗讀新聞 (2) 結合 E-mail 軟體，提供即時的來信通知與朗讀信件內容 (3) 透過小型 FM 發射裝置作小範圍的廣播功能，整合數位與類比的資訊傳播；後者則透過以模擬電臺點歌的「虛擬 DJ」語音服務系統之實作，探討 VoiceXML 的語音標示機制。

第一階段的計畫中已分別開發完成前述兩項子系統，未來將著重於整合其功能，以達成提供個人資訊服務、提供線上學習(e-Learning)平台、以及彌補數位落差(Digital Divide)之綜合成果。

關鍵詞：多媒體資訊系統、語音合成應用、網際網路媒體應用

Abstract

This report describes the information service platform we have built by integrating Internet technology, voice user interface and database systems. The prototype consists of two major components, namely, a personalized information broadcasting system (PIBS) and a VoiceXML-based virtual DJ system. The former is designed to deliver customized, up-to-date information collected on the Internet (including news and

e-mail) to the user via various protocols, such as streaming audio and FM broadcasting. The latter is a case study of VoiceXML specifications and voice user interface design using virtual DJ as an example.

In the second phase of the project, we will be working on the integration of PIBS and VoiceXML. Our ultimate objective is to provide a low cost, easy-to-use information service and on-line learning platform in order to address the increasingly grave 'digital divide' issue in our modern society.

Keywords: Multimedia System, Text-To-Speech, Internet Media.

二、緣由與目的

資訊技術的推廣與擴大讓知識的傳播更為平行與快速，其中網路的快速發展與資訊基礎建設(Infrastructure)的逐步完成實扮演了推波助瀾的角色，網路上的新聞媒體，討論群組等等以極快的速度交換各式各樣的訊息，國際間發生的重大消息傳播也趨近於零時差。

而就人自身的交流習慣來看，人們一向慣於利用聽和說的口頭的方式來進行互動。目前 Text-to-Speech (TTS)系統已能實現了自動的語言分析理解，而 Speech Synthesis Markup Language(SSML)規格[1]之資料檔更便於發聲引擎增加更多的韻律、音調在講話中，使合成語音更生動自然。

然而網路上資訊取得成本的差異(包含軟硬體平台的普及度以及資訊教育基本素養等因素)，直接影響了與資訊世界接軌的程度，對沒有受過足夠的資訊訓練以及

無法負擔設備費用的族群來說，數位落差因而形成[2]。為了嘗試解決此問題，我們設計了一套免費的個人化資訊廣播站(Personalized Information Broadcasting System, PIBS)，提供自動化的工具，結合 Web 技術從提供新聞的網站上擷取資訊，利用語音合成[3]的技術在本機端朗讀出來，以及利用資料庫提供新聞檢索與個人化的操作模式。

數位落差的形成主要肇因於對資訊接收成本的負擔程度之差異。分析資訊接收的成本，我們可約略將其細分為六個項目：(1)網路通訊成本(2)接收設備(硬體)之成本(3)接收軟體之成本(4)學習使用接收平台(軟硬體)之成本(5)付費內容(6)注意力(Attention)的成本。

PIBS 針對以上列舉之成本因素的因應措施如下：

- (1) 提供即時，離線以及伺服器端產生媒體串流等選項，以適應不同網路頻寬之環境。
- (2) 提供接收設備的彈性選擇，使用者可選擇的接收設備從便宜的 FM 收音機到價位較高的個人電腦。
- (3) 提供免費軟體(PIBS 本身)，甚至不需軟體(使用 FM 收音機收聽)。
- (4) 提供友善的軟體介面可降低學習成本，如以 FM 收音機收聽甚至近乎不需學習成本。
- (5) 採用有版權但無須付費之新聞資訊(如中央社)
- (6) 利用聽覺取代視覺，可以在獲取新聞資訊的同時處理其他工作，減低對注意力的要求。

此外，PIBS 中的英文新聞發音系統，如搭配適當之顯示介面(如字幕或跑馬燈標記)而可作為英文的線上學習平台，而透過 Inter-Process Communication (IPC)，PIBS 亦可整合其他應用程式之輸入，例如 Microsoft Outlook，以提供即時反應如來信告知或者重要事件提醒之功能。在語音瀏覽上配合 VoiceXML 的語法規格所設計發展的「虛擬 DJ」點歌系統，更可以透過交互式語音訪問界面的信息和服務，享受 Web 上已有的電子合成音樂(MIDI)應用及豐富的內容。

三、結果與討論

以下分別就 PIBS 及「虛擬 DJ」點歌兩系統架構進行探討。

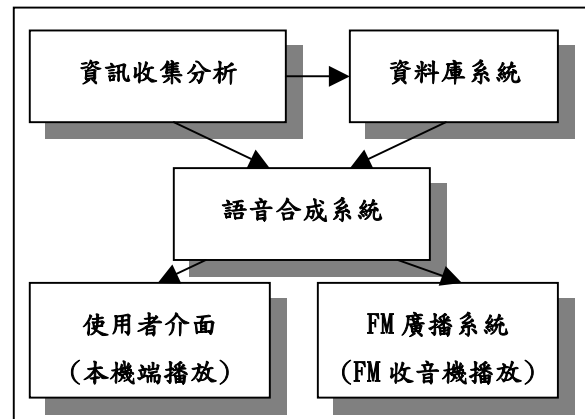


圖 1 PIBS 系統架構圖

3.1.1 PIBS 系統架構

PIBS 可分為五個部份：資訊搜集分析系統、資料庫管理系統、語音合成系統、使用者介面以及 FM 發射系統。資訊收集分析模組由 Server 端提供，負責的功能包括根據 Client 的 Request 透過網際網路取得 HTML 文件格式的新聞資訊、從取得的 HTML 文件中抽離出新聞內容，以及處理後將結果傳輸給 Client。PIBS 的資訊收集分析模組是以 Java 平台實作，以呈現跨平台、物件導向與多緒處理之特性。

資訊收集分析模組首先透過 Java 的 HTTP 功能，從各大新聞網站下載 HTML 文件，為了避免重覆取得同樣的新聞，我們僅採取一至二個網站的資料。同時系統根據資訊與網頁的更新速度有不同的下載更新頻率。例如一般新聞約兩小時更新一次，而股市資訊則是隨著使用者的要求即時取得。取得的文件是 HTML 格式，系統接下來的工作便是如何從 HTML 的 Tag 間擷取真正有意義的資訊。目前我們的方法是以新聞來源網站的 HTML 文件內容的特徵作為判斷的依據。例如 YAHOO KIMO 的新聞網頁原始碼中，在新聞內容之前會有一個標記「<!--== article begin ==-->」，新聞內容結束地方會有標記「<!--== article end ==-->」。只要找到這兩個標記，就能夠找出新聞內容的位置了。而網頁的超連結則是單純的找到<a>，取得 Attribute Href 即可。

本系統採用的新聞網頁依 Request 要求的內容不同收集不同的網頁，程式因應每個網頁的特性跟著做不同的處理，系統一共收集有「新聞」、「氣象」、「交通」、「股票」等四大類資料。中文新聞資料的來源是雅虎奇摩網站 (<http://tw.news.yahoo.com/>)，英文的則是 Yahoo.News (<http://news.yahoo.com/>)，氣象資料則是來自新浪網站 (<http://weather.sina.com.tw/>)，交通資料的來源網站是「交通千里網」 (<http://www.e-traffic.com.tw/>)，而股票資料是採用到「台灣證券交易所」 (<http://sii.tse.com.tw/>)。

圖 2 是部份的 HTML 文件，可以利用特定的 Tag 找到我們需要的資料。

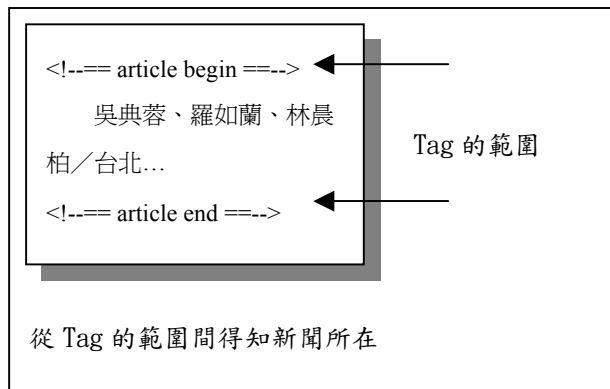


圖 2 HTML 文件片段

3.1.2 PIBS 系統實作

Server 端與 Client 端的溝通方面，我們針對不同的需求設計了三種不同的運作模式-即時模式(On Line)，離線模式(Off Line)以及串流媒體模式(Streaming Media)。

即時模式在 Client 送出 Request 之後 Server 立刻透過資訊收集分析模組取的結果送回給 Client，應用在如股市或路況資訊等即時變化的新聞。相對來說，在離線模式下，使用者可以事先設定想要聽的新聞類別與播放時間與順序，系統不需要立刻有反應，只要定時擷取資料即可。以上兩個模式在使用者介面的部份會有詳細說明。串流媒體模式下的操作則是在 Server 端進行文句轉語音之處理，並即時壓縮為 ra 或 wma 等媒體格式後傳送至 Client 端收聽。

而在資料庫系統方面，我們採用

Microsoft Access 作為資料庫的編輯軟體，Client 端利用 ODBC 連接資料庫且只提供 Query 的功能；Server 端則是由資訊收集分析模組透過 JDBC 與對資料庫有完整的操作功能。

PIBS 的 Client 端包括語音合成、E-Mail 軟體的整合和使用者的介面，實作於 Microsoft Windows 平台，開發工具採用 Microsoft Visual C++。語音合成系統方面，我們採用 Microsoft Speech API 5.0，可於微軟網站上免費下載，特色在於可以接收 XML 格式的資訊，且容易與 Visual C++ 整合開發環境結合，以節省開發時間。

不過在中文方面，目前只接受 GB 編碼的簡體中文，所以我們需要於 Server 端轉換編碼。在英文發音部份，微軟 SAPI 預設發聲引擎之效果並不十分出色，若改為使用 AT&T Natural Voices 發音引擎則有更接近人聲發音的表現。相較之下，中文語音合成的效果則仍有很大的改善空間，以本計畫所採用的 SAPI 系統而言，進行語音播放時仍須集中注意力，方能瞭解廣播之內容，與原先所設定「一心多用」的標的稍有出入，不過一旦中文語音合成技術更臻成熟之後，對於 PIBS 等以語音為基礎的系統將更為普及在日常生活的應用之上。

至於即時反應功能方面，PIBS 目前所整合的 E-Mail 軟體為 Microsoft Outlook。透過 IPC 的連結，可以即時檢查 Outlook 收信狀態，使用者可自行設定即時或定時通知、朗讀內容選擇(如主旨、寄件人或全文)，或依照 Outlook 之優先度之順序或過濾條件決定處理新信件的方式。

3.1.3 使用者介面設計

使用者介面上，我們也是利用 Visual C++ 實作，以下逐項說明主視窗的各項功能：

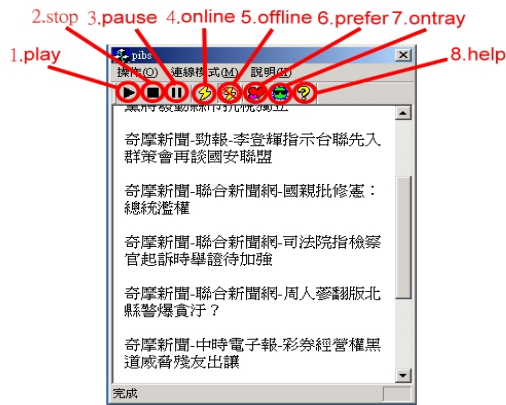


圖 3 PIBS 系統主視窗

■ **play:**

play 有兩種模式，第一種是連線(on-line)模式，當使用者在連線(on-line)模式下按下 play 按鈕，PIBS 會把使用者的個人化設定送至 server 端，並且將經過 server 端處理過後所送回的 string 以語音合成的方式讀出；第二種模式就是離線(off-line)模式，PIBS 會將最近抓下來的新聞檔案以語音合成的方式讀出。

■ **stop:**

當 PIBS 正在作語音合成的處理時，若使用者按下 stop，則語音將立即停止。若使用者還想繼續收聽語音節目，可以設定撥放的時間或者再按一次 play 按鈕，即可繼續收聽。

■ **pause(resume):**

當語音正在播放時，按下 pause 會暫時停止，再按一次就可以繼續中斷的語音。或者也可以重新勾選自己需要的新聞，再按下 play 按鍵重新播放。

■ **online:**

切換到連線模式。在連線模式下，PIBS 則是利用 Socket 直接跟 server 端連線，並從 server 端處獲得較為即時的資訊。

■ **offline:**

切換到離線模式。在離線模式下，PIBS 是利用 FTP 的機制從 server 端抓取所需要的資訊。

■ **preference:**

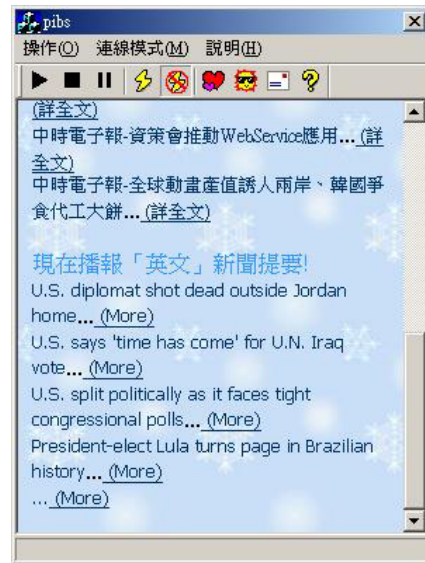
可分為兩種模式連線(on-line)與離線(off-line)：

(1) 連線(on-line mode)：使用者在此模式下，可以設定一些較具有即時性的新

聞。這個介面是 VC++ 的 property page，而按下「代碼」鍵後所跳出的畫面則是普通的 html。

(2) 離線模式(off-line mode)：使用者在此模式下僅能設定一些不具即時性的節目。另外，使用者在此模式下可以設定節目播放的時間。當使用者按下確定按鍵之後，PIBS 會依照使用者勾選的喜好設定至 server 端抓取檔案。

圖 4 至圖 7 呈現了中英文新聞標題與內容，個人喜好設定與 E-mail 即時反應的使用者介面。



圖

4

中英文新聞畫面



圖 5 新聞喜好設定

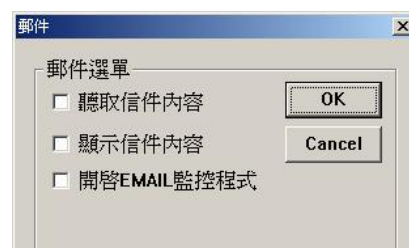


圖 6 即時反應之郵件選單



圖 7 郵件監控程式

最新版的 PIBS 安裝程式可逕至 <http://vision.cs.nccu.edu.tw/pibs> 下載。

3.1.4 FM 發射器

利用低功率的 FM 發射器，可以將電腦音效卡的輸出作短距離的廣播，如此一來使用者只需具備廉價的 FM 收音機即可收聽個人化的新聞內容。有鑑於使用收音機收聽 FM 廣播，若單純使用 TTS 技術來播報新聞稿，是不足以滿足使用者需求的，更需要加入一些環境參數，以更活潑及更具實用性的方式，來製作廣播內容，換言之，也就是以節目導向為製作原則。在 PIBS 執行過程中，所提供的開頭問候語、天氣預報服務、股市消息服務，都是在此原則之下產生。另外為避免過於單調的播報聲，也加入了音樂等特效。這是 PIBS 與一般文字轉語音的新聞系統最大的差別，此後也將由這個角度來發揚，以其個人化的特色來改進現有的被動式廣播媒體。

3.2.1 「虛擬 DJ」系統架構

以 VoiceXML 為基礎的虛擬 DJ 系統包含了：

(1) 語音辨識及合成平臺

此部份將設計並建立語音辨識及合成之運作平臺及一系列的對話流向的 VUI，考慮到 VoiceXML Gateway 的設備成本及複雜性均過高，因此在實作時變通使用 Standalone 的單機平臺模式[4]，透過麥克風及耳機模擬電話對答，而 DTMF 輸入及 TTS、ASR 等語音辨識、合成引擎則使用由 IBM 公司所釋出的 IBM WebSphere® Voice Server SDK 來對 VUI 模型進行設計與改良進行開發、偵錯、及測試的工作。

IBM 的語音辨識、合成架構[5]如圖二所示。它支援由 W3C 所制定的大部份 VoiceXML 1.0 規格及延伸下來的 2.0 規格，由 1.0 至 2.0 的改變有幾方面的比較

[6]，在文法上，2.0 新增了 Vendor - Independent XML 及 BNF 的文法定義方式（本系統是以 SDK 所支援的 JSGF[7]來寫作）。

(2) 對話內容與流程編製

設計的背景是假設在使用者開啟語音資訊系統後，會抓取主要對話文件（vxml 格式）及文法定義（JSGF 格式）的文件，兩者皆由 ASP（Active Server Page）產生；藉由與虛擬 DJ 對話的方式，讓使用者回答被詢問的問題，在限制範圍字彙（JSGF 文法定義）下，比對出符合的字詞或片語，經過確認後才進行資料庫查詢等動作。

此一元件的基本架構是將資料處理腳本及後端的聲音檔放置在提供 Web 服務的伺服器上，再使用 Voice Browser 以類似瀏覽網頁的方式開啟語音服務；在處理腳本的選擇上，我們直接採用對應於伺服器所提供的 ASP 為腳本。

(3) 後端資料庫存取

此部份應用在電子商務和全球資訊網的發展上，幾乎已經是不可獲缺的一環，相關的技術和系統架構已可依照以往類似的服務來建構。而值得一提的是，若欲結合數位音樂（mp3 或音樂 CD）於廣播內容中，恐有著作權及版權之相關規範之疑慮，因此在資料庫的內容資料搜集上，是以爭議較小的電子合成音樂檔（MIDI）做為優先採行的題材。另外系統為了仿照建立其他的語言範本，選用英文老歌作為音樂資料來源，如此在中英對照上減少許多不必要的問題。

3.2.2 「虛擬 DJ」實作成果及場景說明

以下提供數個使用者與虛擬 DJ 進行對話的場景，以說明系統整體執行過程及 help 提示的運作。其中 C 代表電腦（Computer）；H 表使用者（Human）輸入。

場景一：使用者啟動語音資訊系統與虛擬 DJ 對話問候。

C：（預錄）歡迎來到虛擬 DJ 點歌系統。有什麼想聽的西洋老歌嗎？請告訴我們的 DJ，歌名或歌手，馬上為您獻上！此外，您在任何時候都可以說，重聽、幫助及主頁，來尋求協助！

C: 您好! 我是 DJ, 凱文, 你今天心情好嗎?
H: 嗯...
C: 是很開心呢? 或者非常難過?
H: 很開心
C: 那祝福你今天整天都有好心情!

場景二：進階使用者利用快速搜尋歌名

C: (預錄) 歡迎來到虛擬 DJ 點歌系統。有什麼想聽的西洋老歌嗎? 請告訴我們的 DJ, 歌名或歌手, 馬上為您獻上! 此外, 您在任何時候都可以說, 重聽、幫助及主頁, 來尋求協助!
H: 歌名 (not allow barge-in)
C: 您好! 我是.....
H: 歌名 (barge-in)
C: 請問.....
H: 永不止息... (nomatch, barge-in)
C: 對不起, 我聽不.....
H: 歌手 (barge-in)
C: 請問歌手, 是.....

場景三：使用者利用歌手名字 (亞倫卡特) 來點閱相關歌曲

C: 接下來, 我們來找找你想要點的歌。你要先告訴我歌名? 還是歌手?
H: 歌手。
C: 請問歌手, 是合唱團嗎?
H: 不是
C: 請問歌手, 是
H: 亞倫卡特
C: 亞倫卡特, 的作品, 共有 19 首, 第 1 首, 你的寶貝, 第 2 首, 少交危險女伴, 第 3 首, 在鎮壓你, 第 4 首, 跟隨我, 第 5 首, 變得野蠻, 有你想要點的歌嗎?
H: 嗯... 沒有 (No match)
C: 不在這裡面的話, 請說上頁或下一頁
H: 下一頁
C: 第 6 首, 獲得, 第 7 首, 嗨你, 第 8 首, 我願意, 第 9 首, 你是我的全部, 第 10 首, 我將是你的, 有你想要點的歌嗎?
H: 第六 (或說“獲得”)
C: 是亞倫卡特, 的作品, 獲得, 確定嗎?
H: 對的
C: 一旦撥放, 就無法停止, 要繼續嗎?
H: 好
C: 獲得, 送給你。 (音樂聲)
.....

3.3 整合 PIBS 及「虛擬 DJ」平臺

在此次的成果之中所提出的兩個似是分開且獨立的個人化廣播系統及使用者語

音瀏覽介面, 其實有互為體用之處。在 PIBS 之中我們仍採用著傳統的圖形介面, 來讓使用者搭配控制, 但另一方面也成功的實現以 DJ 點歌方式的語音瀏覽介面, 不但驗證了以 VoiceXML 規格的确有助於降低在發展語音應用程式的複雜性, 更可以在限定用語之中輕易的達到使用者控制的效果, 而只需要一個耳機麥克風來達成 VPIBS (VoiceXML - Based Personalized Information Broadcasting System) 的目標, 但這仍不是計畫的最終目的, 之後的 PIBS 不但將結合語音控制介面成為 VPIBS, 還將結合知識推論引擎 (Inference Engine), 如此在資訊的輸入輸出上, 不在是一個被動式的程式, 更進一步主動貼近個人化需求。

五、參考文獻

- [1] W3C, “Speech Synthesis Markup Language Specification”, <http://www.w3.org/TR/speech-synthesis/2002>
- [2] 台灣國際電子商務中心, “數位落差”, http://www.nii.org.tw/cnt/info/Report/20020305_13.htm, 2002
- [3] R. Sproat (Editor), “Multilingual Text-To-Speech Synthesis: The Bell Labs Approach”, Kluwer Academic Pub, 1997.
- [4] K. R. Abbott, “Voice Enabling Web Applications: VoiceXML and Beyond”, après, 2002. p 41-47.
- [5] IBM, “IBM WebSphere® Voice Server for Windows® 2000 and AIX® Software Developers Kit VoiceXML™ Programmer’s Guide”, 2001.
- [6] C. Sharma and J. Kunins, “VoiceXML: Strategies and Techniques for Effective Voice Application Development with VoiceXML 2.0”, John Wiley & Sons, 2002.
- [7] W3C, “Jspeech Grammar Format”, <http://www.w3.org/TR/jsgf>

六、成果自評

在今年度的預定目標中是以 PIBS 個人資訊廣播平台及 VoiceXML 規格使用者語音介面 (VUI) 作為實作的重點, 之後不但分別

達成在這兩項應用程式的發展，也進一步的對兩者整合，投注更多的努力，在 PIBS 平臺的建置上，除了對於語音提供資訊服務上的實際應用，配合較高階的英文 TTS 發聲引擎(AT&T Natural Voices)，在英文教學上輔助學習效果十分顯著；另外在 VoiceXML 的語音使用者介面上，除了達成限定字彙的語音瀏覽介面；在可以預見的將來，不但可以增加 PIBS 在操作上的便利，輔以多模式語音的控制介面；更能在 PIBS 中以節目導向的廣播內容錄製方式，這在其他的語音應用程式上是尚未實現的挑戰。

同時，在 91 年末所舉辦的民生電子研討會(WCE2002)，也將當時的階段性成果分別作了探討及發表兩篇相關論文，從中獲取了寶貴的學術交流經驗以作為之後改進之參考。

可供推廣之研發成果資料表

 可申請專利 可技術移轉

日期：92年10月31日

國科會補助計畫	計畫名稱：子計畫二：以 VoiceXML 為基礎的個人化資訊廣播站(I) 計畫主持人：廖文宏 計畫編號：91-2213-E-004-012- 學門領域：Web Technology
技術/創作名稱	Personalized Information Broadcasting System
發明人/創作人	廖文宏
技術說明	中文：結合網際網路資料擷取、應用程式互動與語音合成技術，將新聞、電子郵件、重要通知等內容以語音方式呈現，並透過超低功率 FM 做小範圍廣播，使接收者使用最低廉的成本，獲取即時且量身定做的資訊。
	英文：By integrating Internet data retrieval, inter-process communication and speech synthesis (English and Mandarin) technologies, we are able to develop a personalized information service system that provides latest news, e-mail reading and event notification functionalities, all via low-power FM broadcasting. The system greatly reduces the attention cost as well as hardware cost for the potential users.
可利用之產業及可開發之產品	*即時且個人化之語音新聞廣播 *英文線上學習 *中文單字發音學習 *社區電台
技術特點	*採開放式規格(VoiceXML, HTML)及開放程式碼 *簡化使用者介面(以聽覺取代視覺) *減少注意力之要求 *接收成本低 *彌補數位落差
推廣及運用的價值	*車上電腦 *語言學習(中英文) *視障者的資訊來源 *視障者與電腦互動機制

※ 1. 每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴單位研發成果推廣單位（如技術移轉中心）。

※ 2. 本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。

※ 3. 本表若不敷使用，請自行影印使用。