盲胞有聲書語音查詢系統

林政源、謝明峰、張智星
國立清華大學資訊工程學系
{gavins, pacific, jang}@wayne.cs.nthu.edu.tw

摘要：

我們設計一套採用語音輸入和輸出的有聲書查詢系統，目的在讓視力障礙的盲胞能方便查詢並收聽清大盲友會有聲書。盲胞不但可以查詢資料庫的書籍，也可以下載有聲書直接收聽。這個系統採用了兩大語音處理的技術：語音辨識與語音合成。前者是利用 HMM-based 的原理，而後者是採用 concatenation-based PSOLA 的合成技術。在系統設計方面則運用了 Microsoft .NET 架構下的 Web Service 來進行所有功能的整合。最後，在系統辨識評比方面，我們也得到了不錯的成果。

一．系統介紹：

本系統是架構在 Microsoft .NET Framework 之下，利用 Web Service 的功能，來進行各種網路資料的傳輸。使用者利用簡單的按鍵來選擇使用書名、作者或出版社來查詢，再經過麥克風的語音輸入，系統會將語音檔傳到 Web Service 中，然後採用梅爾刻度式倒頻譜（MFCC）[2]的方法進行語音特徵參數粹取。再將粹取後的參數和資料庫中訓練過的所有語句進行比對，找出分數最高的來當作辨認結果。利用辨認結果去查詢資料庫，取得該筆資料的所有相關資訊，例如書籍編

號、書名等等。

在輸出方面，我們採用文字顯示與語音合成兩種並行的設計。使用者可以從顯示器得知查詢結果，或者利用喇叭聆聽查詢結果。倘若使用者有興趣試聽找到的書籍，可以直接按鍵下載，系統將會從網路上抓取該本有聲書，直接撥放給使用者試聽。

以下是整個系統以語音輸入與文字和語音輸出的流程圖：

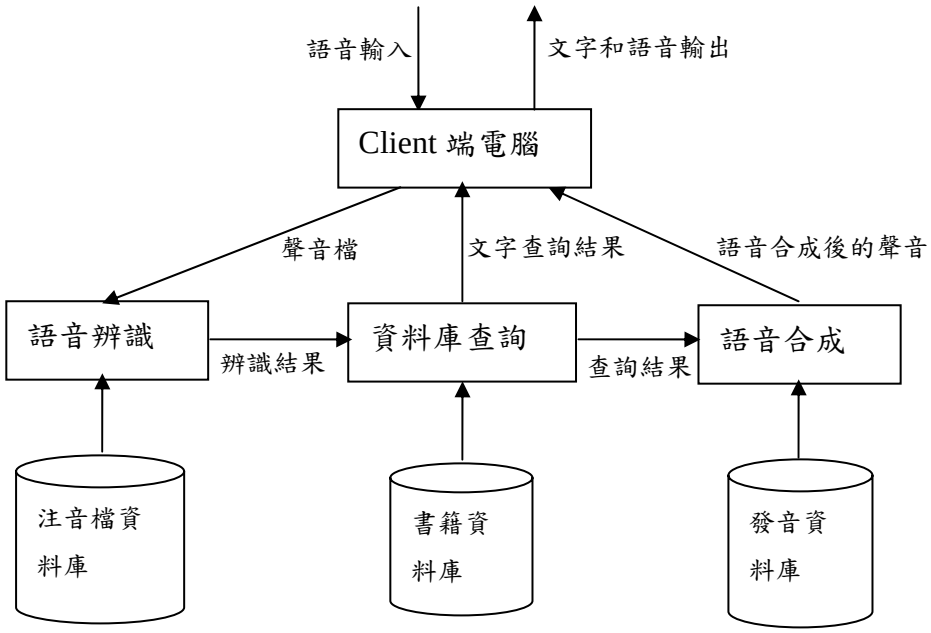

圖 1.系統架構圖

二・採用技術：

1. Microsoft .NET 平台：

.NET 平台提供了 Web Service 功能，在網路迅速普及，且寬頻蔚為風潮的現在，更能顯現其便利性。我們可將 Web Service 看成一個可使用的遠端函式。程式設計師只要連上網路，並了解函式的輸入參數和輸出結果，就可以直接使用該函式而不必花時間重新撰寫。充份用到了資源再利用的好

處。因此，在設計本系統的時候，將所需要的語音辨識、資料庫查詢、文字轉語音都以 Web Service 方式實作，如此可以減輕使用者端電腦的負擔，達到快速搜尋和方便使用的目的。

2. 語音辨識核心：

語音辨識的最大好處就是使用者可以不用依傳統文字輸入方式作查詢，而改以更人性化的語音輸入方式查詢，以期能拉近人與機器間的距離。而語音辨識主要分為二個步驟，MFCC (Mel-scale Frequency Cepstral Coefficient) 特徵參數粹取與 HMM (Hidden Markov Model)比對辨識[2]。分述如下：

I. 語音特徵參數粹取：

特徵參數粹取的目的在於將一段語音檔的聲波(Waveform 形式)轉為另一個參數表示(通常資料量會明顯降低)，以便將來辨識之用。而這裡使用梅爾倒頻譜參數(MFCC)，其考慮到人耳對頻率的特性，較其他方法為佳。

II. 比對辨識：

早期以 DTW (Dynamic Time Warping) [2]的技術來實作語音辨識，然而效果並不甚好。我們的系統則是採用隱藏式馬可夫模型(HMM, Hidden Markov Model) 為其辨識核心，其具有語音統計特性，經證實能夠有效模擬語音的細微變化，為近年來語音辨識最為廣泛的使用方法。

以上這二個步驟可利用 HTK (HMM Tool Kit) [3]加以完成。HTK 是一套功能

強大的語音辨識軟體，可以將大量的語音用 HMM 訓練之後，加以辨識。所以本系統採用 HTK 為辨識核心。並將所有的書名、作者、出版社從資料庫中粹取出來，進行標注音的動作。再將要被辨識的語音檔做特微參數粹取，利用 Viterbi Search 演算法，將粹取出來的參數和之前標好的注音檔比對，找出一個最相似句子，當做語音辨識的結果。

3. 語音合成 (TTS, Text to Speech) :

為了能讓盲胞也能知道查詢的結果，整個系統需要能將文字轉成語音(TTS, Text to Speech)的功能，才能有聲音的輸出。首先先將文字輸入到 TTS 系統中，TTS 系統在收到文字後，根據原有在資料庫中的語音檔進行連音，調整長度、大小及聲調的動作。這裡採用的方法是基週同步疊加法，PSOLA (Pitch Synchronous Overlap and Add) [1]。另外處理中文時，必須考慮到聲調的轉換 (例如，李總統這三個字的聲調為：3 聲、2 聲、3 聲)，所以我們必須另外建立一些常用的辭庫，來作注音修正。最後，將所得的語音檔播放出來即可。

三・操作介面說明：

本系統的操作說明如下：

1. 開始執行程式時，會以語音提示使用者使用語音查詢的方法，按數字鍵 1 進行書名查詢，按 2 進行作者查詢，按 3 進行出版社查詢

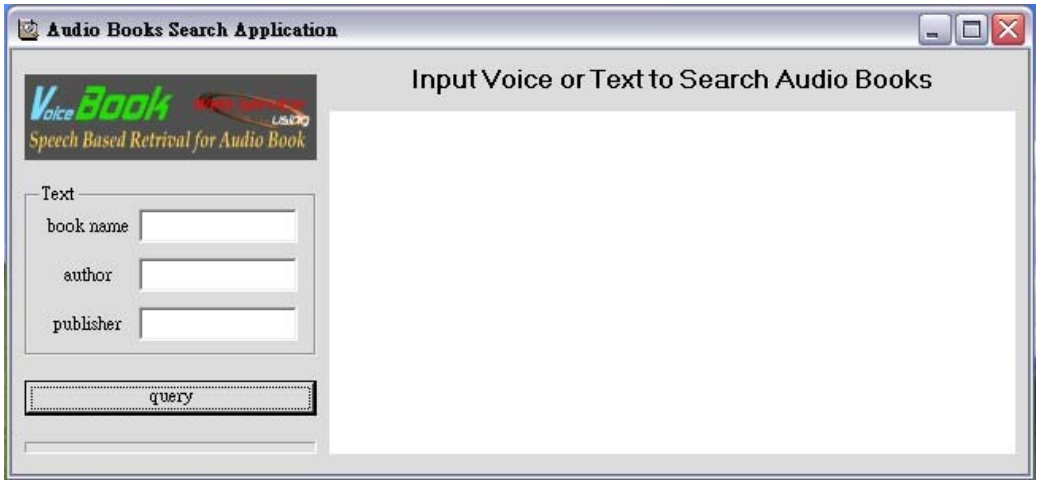

圖 2. 開始執行時的使用者介面

2.在按下數字鍵之後，語音會提示在嗶聲後開始 3 秒錄音，而嗶聲響起，左下角的 progress bar 會開始進行,在此時對著麥克風說出要查詢的關鍵字。

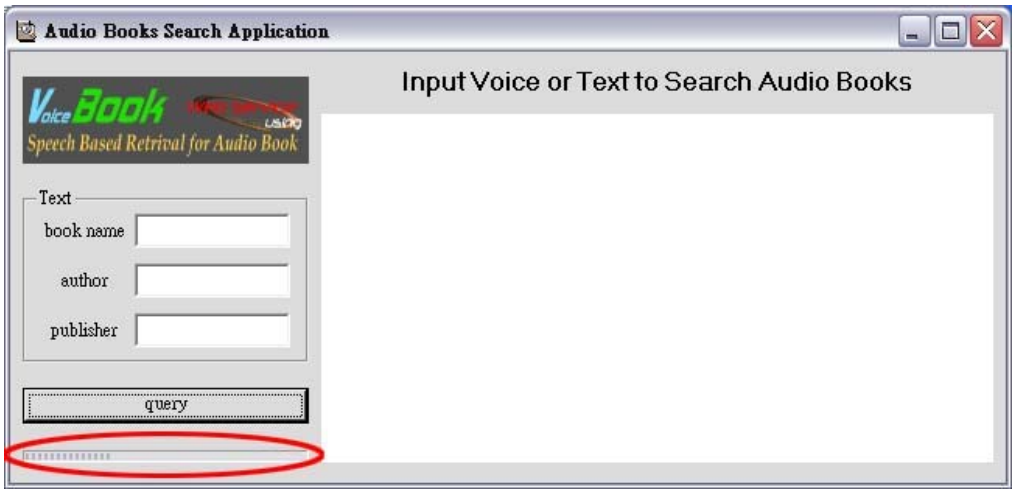

圖 3. 錄音時, progress bar 正在移動

3.錄完音之後，將會傳回辨識結果，不但將結果顯示出來，也利用語音合成的技術，將答案唸出來。若想聽該筆有聲書，可以按下 play 鈕或是按下指定的數字鍵，系統將會下載有聲書並播放出來。

圖 4. 所得的輸出結果，除了印出來外也會唸出來

4. 本系統也可以使用文字查詢，可直接在欄位上填入想查詢的資料

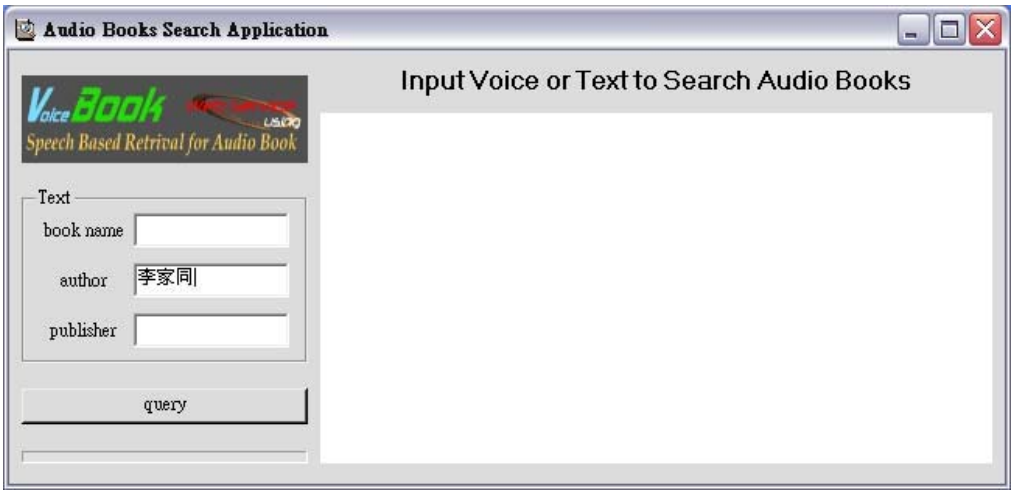

圖 5. 將關鍵字填入欲查詢的欄位中

四・實驗成果：

本系統的目的在於發展出語音和文字二種輸入和輸出方式，更優於一般資料庫只有文字輸入的方式。如此將使得盲胞和不會中文輸入法的一般民眾能方便的使用。因此，最重要的就是語音輸入的辨識率和系統操作的方便性。就辨識率而言，由於本語音辨識系統是採用最接近的句子當做辨識結果。被辨識系統資料的

多寡，平均每句的字數，都會影響正確率。下表是我們測試的結果：

資料庫	總共筆數	平均每筆字數	測試次數	正確次數	辨識率
書名	12147	7.24	113	95	80.53%
作者	5277	4.33	105	79	75.28%
出版社	1091	3.96	127	104	81.89%

從上表可得知，本系統的辨識率已達到大約 80%的水準，已可達到方便使用的地步。而方便性方面，由於完全是以簡單的數字鍵和聲音來做輸入，所以對視障者和一般人來說，都能夠輕易地操作。經由實際全盲的人測試，印證了我們的想法。

五．結論：

在本論文中，我們已經實際設計出一套完整的有聲書查詢系統，這是基於兩種語音技術下的整合系統，利用.NET 架構的 Web Service 來結合系統所應用的技術，經過多人的測試，其結果堪稱理想且實用價值極高。未來可以更進一步提高辨識率或加速比對時間以求系統更加完善。

六．參考書目：

1.Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon. "Spoken Language Processing." Prentice Hall

2. Fundamentals of Speech Recognition, Prentice Hall.

3. The HTK book, 2000, copyright for Microsoft Corporation.