適合視障者使用之電腦界面技術與系統設計(I)ü子計畫二：盲用電腦之國語單詞輸入及語音輸出系統之發展

Speech Input and Output Interface of Computer for Blind Users

計畫編號：NSC-89-2614-E-009-001
執行期限：88 年 8 月 1 日至 89 年 7 月 31 日

主持人：陳信宏　國立交通大學電信工程學系
schen@cc.nctu.edu.tw
計畫參與人員：郭威志、何鎮仲、倪景滌、蔡偉和
國立交通大學電信工程研究所

一、中文摘要

本計畫歷經三年的研究，完成電話語音及麥克風語音國語音節辨認核心，使用 MAT-2000 及 TCC-300 語料庫訓練右文相關聲韻母 HMM 模型，其音節辨認率分別達到 67.1% 及 70.9%，其效能已接近 state-of-art 水準。使用此二國語語音辨認核心及以前完成的文句翻語音 (TTS) 系統，我們發展了以下三個盲用輔具：(1) 電話語音電子郵件系統，以聲控方式操作，使用TTS讀取 email，並以壓縮語音回信；(2) 電子書閱讀系統，以聲控方式操作，使用 TTS 將內容讀出；(3) 電子辭典查詢系統，以聲音輸入詞，經辨認後，將詞之解釋以 TTS 讀出。本計畫已完成雛形展示系統，可作為未來發展實用系統之基礎。

關鍵詞：關鍵詞：盲用輔具，國語語音辨認，文句翻語音。

Abstract

In this three-year project, we have developed a telephone-speech and a microphone-speech Mandarin syllable recognizers. Both of them use the HMM approach to train right-context-dependent initial and final models using MAT-2000 and TCC-300 databases. Syllable accuracy rates of 67.1% and 70.9% were obtained. Using these two Mandarin speech recognizers and a previously developed text-to-speech (TTS) system, we have implemented three blind-aid systems. One is a telephone-speech email read/reply system. Another is a dictionary access system. The other is an electronic book read system. They are all voice-controlled with input speech recognized by these two speech recognizers and voice-responded with output speech generated by the TTS system.

Keywords: Blind-aid system, Mandarin speech recognition, Text-to-speech.

二、緣由與目的

視障者是社會中弱勢的團體，其生活空間受到許多限制而較常人狹窄許多，雖然今日電腦網路的蓬勃發展，但卻未為視障者帶來任何的便利，主要的障礙在於缺乏適合視障者使用的電腦，尤其是中文化的系統，因此本三年計畫主要目標是開發適合盲人使用的電腦語音輸入及輸出介面，以發展盲用網路資訊擷取雛形系統，作為未來發展實用系統之基礎。

三、結果與討論：

(一) 電話及麥克風國語音節辨認核心

我們使用 MAT-2000 電話語音語料庫及 TCC-300 麥克風語料庫製作了國語音節辨認核心，系統之設計參數說明如下：

* **辨認參數抽取**：首先將輸入訊號之直流偏移值 （DC bias）先去除，然後由每一 frame 求出 38 維特徵參數，包括 12 維

的倒頻譜參數（MFCC）、12 維的一階差量倒頻譜參數（delta MFCC）、12 維的二階差量倒頻譜參數（delta-delta MFCC）、1 維的一階差量對數能量（delta-log-energy）以及 1 維的二階差量對數能量（delta-delta-log-energy）；frame 長度為 30 ms，移動間隔為 10 ms。

* **通道效應補償**：採用 SBR (Signal Bias Removal) 法，去除訊號中的通道偏移，它先建立一個 codebook，將每一 frame 之辨認參數 encode，由 encoding error vector 之平均來求出通道偏移量，再將其移除。

* **訓練方法**：模型的訓練是使用「切割－K 均值訓練程序」(segmental k-means training procedure)，它是一個包含兩個步驟的反覆疊代程序，一為使用現有 model 對 training utterances 進行切割，另一為 model 的 updating。

* **辨認方法**：辨認時，我們是使用「一階動態規劃演算法」(one stage DP algorithm)來求取最佳的辨認音節串。

* **音長模型**：使用之音長模型為對狀態長度分布假設為加瑪（Gamma）分布。

* **HMM 模型**：我們使用聲母 (initial) 及韻母 (final) 次音節為 acoustic modeling units，建立 HMM 模型，首先建立 100 個右相關（right-final-dependent, RFD）聲母模型及 40 個前後文不相關 (context-independent, CI) 韻母模型，其中聲、韻母模型分別使用 3 及 5 個狀態，而模型為 gender dependent，亦即男女分別建立模型；接著將 CI 韻母模型再細分為右文相關 (right-context-dependent, RCD) 之韻母模型，細分之方法採用決策樹法 (decision tree method)，它的原理是建立在語音學 (phonetic) 及聲學 (acoustic) 上，利用聲音的特性來分類，找出最佳的類別，其做法是由所有語音資料中，把含有相同的不相關模型的語料放在根節點 (root node)，再根據我們所給的問題，看有那些符合那些不符合，而據此分裂成兩個節點，每一節點為一群；再由所有的問題分群中，依一分群效能量測來決定最佳的問題及分群；循序重覆分群，最後完成一個 tree，由每一末梢節點訓練一個 RCD HMM 初始模型。完成 RCD 韻母模型分群後，再使用 segmental k-means 訓練程序來獲得最後之 HMM 模型。

* **實驗結果**：電話語音的辨認實驗是採用 MAT2000 語料庫中的兩個子分項 DB4（詞）及 DB5（短句）為訓練語料，其內容包含男生 1079 人共 175305 音節，女生 1300 人共 303866 音節，建立了男女兩套的 100 個聲母模型與 40 個 CI 及 290 個 RCD 韻母模型，測試語料使用 500 句連續語音，結果列於表一。

表一、電話語音的辨認實驗結果

	Ins.	Del.	Sub.	Accuracy
CI	0.95%	1.73%	31.28%	66.0%
RCD	3.26%	0.74%	28.94%	67.4%

麥克風語音的辨認實驗是採用 TCC-300 語料庫，它包含台大（短句）、成大（長文）及交大（長文）錄製的 300 人語料，男生 150 人共 165727 音節，女生 150 人共 166981 音節；用此語料分別建立了男女兩套的 100 個聲母模型與 40 個 CI 及 290 個 RCD 韻母模型，我們使用 4/5 的語者語料來訓練模型，其餘 1/5 的語者語料來測試，結果列於表二。

表二、麥克風語音的辨認實驗結果

	Ins.	Del.	Sub.	Accuracy
CI	1.35%	0.63%	28.63%	69.3%

RCD	1.43%	0.48%	27.13%	70.9%

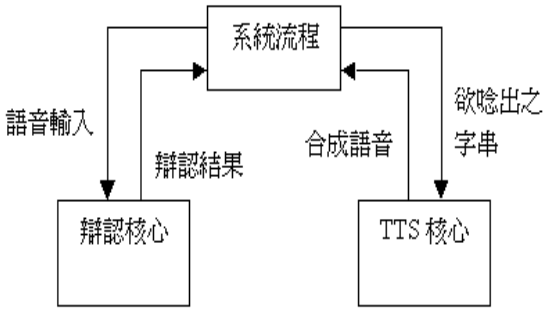
* **實驗結果分析：上述兩實驗結果均較去年之結果有大幅進步，主要改進在於：**(1)**使用大的語料庫來訓練** HMM **模型；**(2)**嚴謹的控制整個訓練過程**;(3)**使用** decision tree **法來訓練** RCD final **模型。**

(二) 盲用輔具雛形展示系統

使用上述的兩個國語語音辨認核心及以前完成的文句翻語音 (TTS) **系統，我們發展了以下三個盲用輔具：**(1) **電話語音電子郵件系統，以聲控方式操作，使用** TTS **讀取** email**，並以壓縮語音回信；**(2) **電子書閱讀系統，以聲控方式操作，使用** TTS **將內容讀出；**(3) **電子辭典查詢系統，以聲音輸入詞，經辨認後，將詞之解釋以** TTS **讀出。它們均採用圖一之系統架構，以下簡略說明它們的功能：**

A. 電話讀取電子郵件系統

當使用者撥電話進入系統後，它會利用 TTS **來和使用者進行溝通，引導其依序輸入姓名和密碼以進入郵件伺服器，**



圖一、盲用輔具系統架構圖

並可進一步的執行電子郵件功能，其動作如同由電腦經網路進入電子郵件系統一般。而在讀信方面，乃是以 TTS **將信件內容讀出，進而利用系統辯認的核心來辯認出使用者口語輸入的功能項目，由於為小字彙辯認，因此有高的辯認率。而回信方面是採用語音壓縮的方式，將壓縮的語音回覆給寄件者。而在程式設計方面，採用了** FSM (Finite State Machine)**的觀念，將不同的動作放置在不同的狀態並完成，此設計方式是可輕易的檢測並修正系統。**

B. 電子書閱讀系統

此系統設計之主要差別乃是改變系統流程的內容，針對讀電子書的目的而規劃出適合的功能。另外，在此系統中亦提供記憶的功能，能紀錄使用者以前已讀段落，如此一來，當下次在使用此系統時便可接續之前已讀之段落，以 TTS **唸出使用者欲聽取的文章內容。在辯認率的表現方面，由於也是使用小字彙詞庫，所以依舊有很好的辯認率。**

圖二即是 A、B **兩系統的觀測介面，經由介面所提供的訊息，便可清楚的掌握系統的狀態及表現。**

C. 電子詞典查詢系統

此系統乃是使用麥克風將欲查詢的單詞輸入，經由辯認找出對應的解釋，再利用 TTS **將解釋唸出。**C **與** A、B **系統的主要差異在於此為一大詞彙辯認系統。在辯認率方面，經測試的結果，約為** 35**（正確句數）**/40**（總句數）**＝87.5%。

四、計畫成果自評：

本報告內容與原計畫內容相符。

五、參考文獻

1. Y. F. Liao and S. H. Chen, “A Modular RNN-based Method for Continuous Mandarin Speech Recognition,” to appear in IEEE Trans. Speech and Audio Processing.

2. W. J. Wang, Y. F. Liao and S. H. Chen, “Prosodic Modeling of Mandarin Speech and Its Application to Lexical Decoding,” Eurospeech’99, Budapest, Hungary, Sept. 1999.

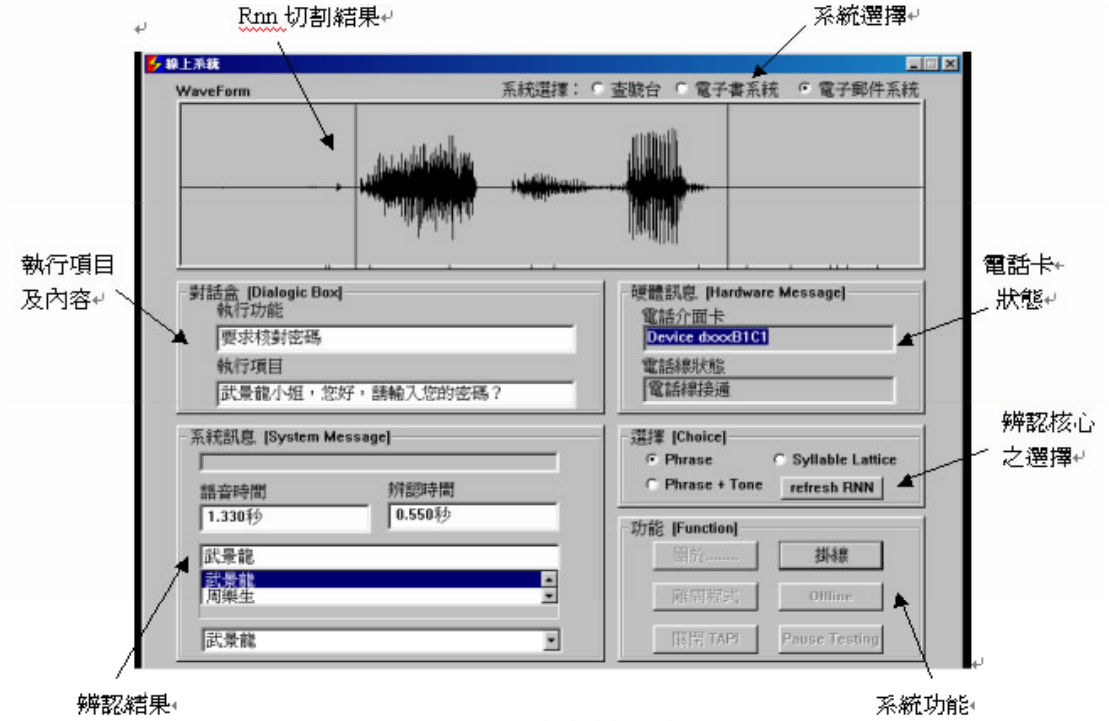
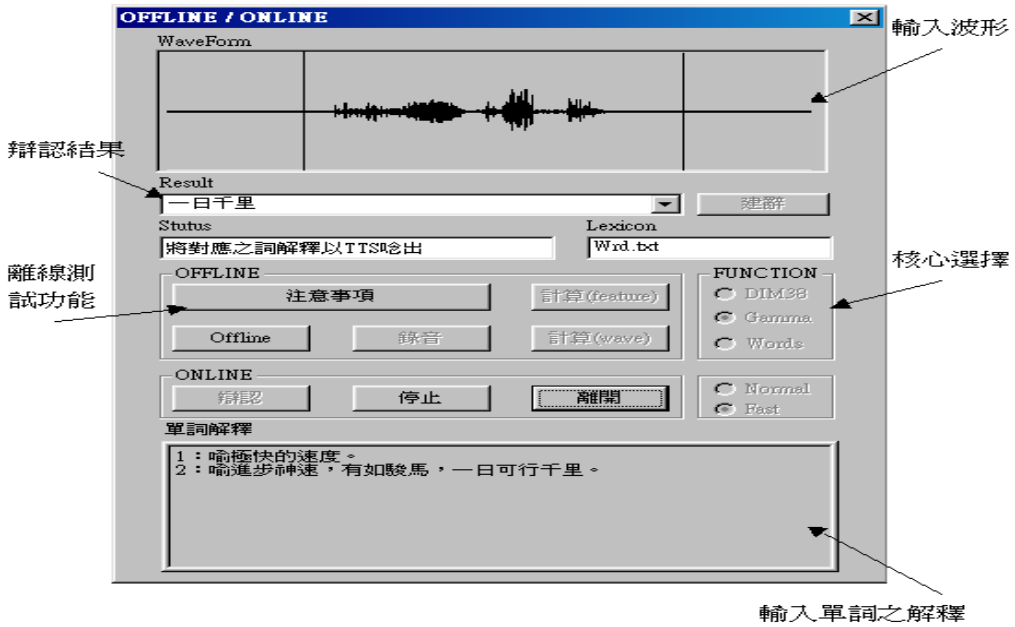


Figure. 系統界面者明圖

圖二、電話語音電子郵件系統及電子書閱讀系統



圖三、電子辭典查詢系統