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-87-2213-E-009-027
執行期限：86/8/1 - 87/7/31
主持人：陳信宏 交通大學電信系教授
schen@cc.nctu.edu.tw

中文摘要(關鍵詞：盲用電腦,國語單詞辨認，文字翻語音。)

在本計畫中，我們嘗試整合單詞辨認系統與文字翻語音系統，在windows 環境，實做一語者獨立之國語單詞輸入及語音輸出雛型展示系統，研究內容包括錄製長文語料庫，建立基本 HMM 模型，並以 SBR、CMN 去除語音信號偏移量者效應，及語者分群與 MLLR、LS 語者調適法，進一步提升辨認率，最後建立一雛型展示系統。

英文摘要(關鍵詞：computer for blind user, isolated word recognition, text-to-speech.)

In this report, we combine the isolated word recognition and text-to-speech sub-systems to form a speaker-independent, voice-command and voice-output system. Several signal bias removing and speaker normalization/ adaptation algorithms including SBR, CMN, MLLR, and LS methods are intensively studied. A prototype system is also developed.

計畫源由及目的：

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，為人們帶來更寬闊的視野，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，更是缺乏，只能自行發展，因此本三年計畫計畫發展適合盲人使用的電腦語音輸入及輸出介面，以發展實用的盲用網路資訊擷取系統。而第一年度的目標是在windows 環境下，發展一個小字彙語者獨立國語單詞 on-line 辨認系統，以作為盲用電腦之語音輸入，並結合文字翻語音系統以語音輸出作為回應。

研究方法及步驟：

1. 國語長文語料庫製作

我們首先規劃錄製兩百人的語音資料庫—國語長文語料庫，其語料內容包括 411 個國語基本單音節、26 個英文字母、數字串與選自中研院平衡語料庫中的各類文章（長文語料），以 20 人(10 男 10 女)為一組，共 10 組，分批錄製，前 5 組使用相同的頭帶式麥克風與聲霸卡錄音，以製作基本辨認系統，後 5 組，則以組別為單位使用不同的麥克風與聲霸卡錄製，以測試通道效應，其錄音環境設定見表格一。

表格一. 國語長文語料庫錄音環境

	說明
錄音文章	選自於中央研究院平衡語料庫
錄音環境	交大及成大語音實驗室
麥克風	VR3560 頭帶式麥克風（台灣樓式電子工業製造）
聲霸卡	16-bit sound blaster
取樣頻率	16 kHz
錄音時間	720.55 分鐘
總音節數	149,682 個音節

2. 基本HMM語者獨立辨認系統

前 5 組語料錄製完成後，經前置處理更正錯誤與注音後，求取特徵參數（見表格二），以前 4 組共 80 人當訓練語料，最後 1 組共 20 人當測試語料，先以 411 個國語單音節之語料，建立 411 的國語單音節 HMM 模型（以 100 個右相關 initial 與 39 個final HMM 模型組合而成，男女各建立一組模型），再加上靜音與呼吸聲之 HMM 模型，當作連續語音 HMM 模型之起始模型，以長文語料重新訓練，得到最後之語者獨立國語連續語音 HMM 模型，此基本系統之辨認率見表格三。

表格二. 取樣方式與特徵參數

Sampling frequency	16kHz
Frame shift	10ms
Frame length	30ms (Hamming window)
Feature parameters	12-order mel-cepstrum + 12-order Δmel-cepstrum + Δlog energy

表格三. 基本系統之辨認率

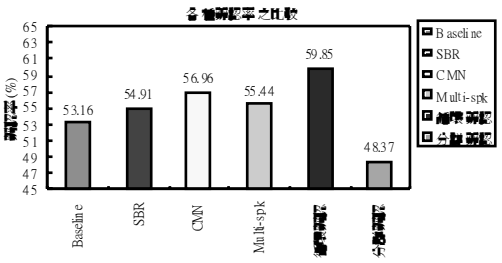
	Ins.	Del.	Sub.	Acc. rate
Mix=10	543	174	10610	52.50%
Mix=15	549	182	10438	53.16%
Mix=20	517	170	9925	55.50%

3. 語者正規化

我們針對基本系統中 mix=15（見表格三）的情況（接下來的實驗，皆以

表格三中 mix=15 的情形為基準），以兩種語者正規化的方法— signal bias removing (SBR)與Cepstrum mean normalization (CMN)，嘗試消除不同語者間的偏移量 (bias)，以得到一正規化的特徵參數，重新估算出較好的HMM 模型，其中 SBR 以每一句子，CMN 以每一人為單位，做偏移量計算，結果請見圖表一，以 SBR 與 CMN 方法各可以有效地，將辨認率從原本的 53.16% 提升到 54.91% 與 56.96%。

圖表一. SBR 與 CMN 之辨認結果



4. 語者調適

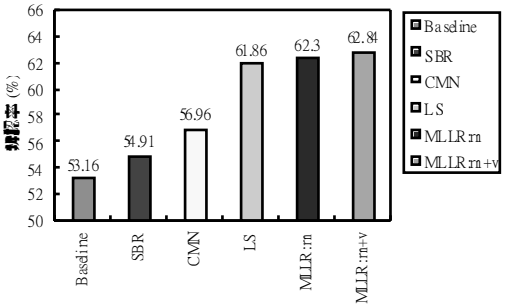
我們亦嘗試以 maximum likelihood linear regression (MLLR) 演算法，與 MLLR 的一個特殊情形 least square regression (LS)，在正式辨認某一語者的語音前，以下列兩種方式做語者調適:

a. 利用少數此語者的語料，調整 HMM 模型以適合辨認此語者，再做辨認。

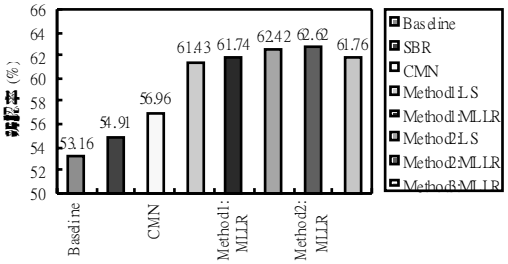
b. 將每一語者的所有語料，相對於一初始 HMM 模型，求取每一語者的正規化函數，將所有語者的語料依此正規化函數做正規化後，重新求取新的 HMM 模型，而在實際辨認某一語者時，則需利用少數此語者的語料，求取此語者的正規化函數，將此語者的語音正規化後，再做辨認。

方式 a 的辨認結果見圖表二，方式 b 的結果見圖表三，可將辨認率提升到 62.84% 與 62.62%。

圖表二. MLLR 與 LS 語者調適之辨認率



圖表三. MLLR 與 LS 語者正規化之辨認率



5. 雛形展示系統

我們使用經 SBR 做語者正規化後，訓練出來的 HMM 模型，在 windows 作業系統下，建立一語者獨立之國語獨立詞辨認核心，先將所有的獨立詞以程式自動建成詞樹，再以 one-stage 演算法找出最可能的獨立詞，並加上 beam search 與 RNN 預切割網路，大量刪除較不可能的路徑，以加快辨認速度，此語音辨認核心再與 text-to-speech 結合，即可完成所需的語音輸入與語音輸出界面，初期測試台北市 1900 個金融公司電話號碼查詢，與電信局台中清水局一萬兩千筆公司行號電話號碼查詢，目前則正進行對話式網路即時新聞閱讀展示系統。

結論：

在本計畫中，我們初步嘗試整合單詞辨認系統與文字翻語音系統，在 windows 系統上，實做一語者獨立之國語單詞輸入及語音輸出雛形展示系統，並研究以 SBR、CMN 去除語音信號偏移量者效應，及語者分群與 MLLR、LS 語者調適法，未來將據此建立一對話式網際網路資訊查詢系統。

計畫成果自評：

本報告內容與原計畫內容相符，

發表之論文：

1. 蔡忠安，語者調適和正規化技術在語音辨認之初步研究，交通大學碩士論文，中華民國八十七年六月。

參考文獻：

[1] M. Rahim and B-H. Juang, “Signal bias removal by maximum likelihood estimation for robust telephone speech recognition,” *IEEE Trans. Speech and Audio Processing*, Vol. 4, no. 1, pp. 19-30, January (1996)

[2] Gales, M. J. F. & Woodland, P. C. “Maximum likelihood linear transformations for HMM-based

speech recognition," *Technical Report CUED/F-INFENG/TR291*, Cambridge University, Cambridge, UK (1997)

[3] M. Tonomura, T. Kosaka and S. Matsunaga, "Speaker adaptation based on transfer vector field smoothing using maximum a posteriori probability estimation," *Computer Speech and Language* 10, pp. 117-132 (1996)

[4] J. Takahashi & S. Sagayama, "Vector field smoothed Bayesian learning for fast and incremental speaker/telephone-channel adaptation," *Computer Speech and Language* 11, pp.127-146 (1997)

圖表四. 展示系統架構