

國語語音辨認之實用系統發展(I)

Development of Mandarin Speech Recognition Systems (I)

計畫編號：NSC87-2213-E-009-056

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：國語連續音節辨認、雜訊語音切割、雜訊補償、偏移補償)

本計畫的目標為建立一個電話語音、語者獨立、國語大字彙連續語音辨認系統，以作為資料查詢之用。本報告提出若干強健式語音辨認法，包括：雜訊語音切割、雜訊補償、語者/通道偏移補償、相似值補償、及強健式訓練等，以提升系統辨認率。

英文摘要(關鍵詞：continuous Mandarin syllable recognition noisy speech segmentation, noise compensation, bias compensation)

The goal of the project is to develop a prototype system of speaker independent, continuous Mandarin telephone speech recognition system for information inquiry applications. Several robust speech recognition methods are discussed in this report. They include noisy speech segmentation, noise compensation, speaker/channel bias compensation, likelihood compensation, and robust training.

計畫源由及目的：

近年來，國內在語音辨認之研究上已有很大進展，許多系統被開發出來，包含臺大的金聲系列，成大的音

中仙，TL 的大字彙及電話 keyword spotting 系統以及其他許多中量或大量詞彙辨認系統。另外國內目前正錄製一套電話語音語料庫，稱為 MAT，可供國內研究使用，而中研院也提供一個詞庫及一個已斷好詞之大型 text corpus。而我們經多年的研究，也有許多研究成果，本計畫之目的是要以過去之研究成果為基礎，以三年的時間發展先進的語音辨認雛型系統：研究過程將以系統發展的觀點，對所有的關鍵問題做深入探討，以期最後能完成 on-line、快速、對環境及語者能快速調適及高辨認率的雛型系統。而第一年度的目標是在整理 MAT 語料庫與建立基本辨認模型，完成一個以 Win95 之 TAPI 為介面、on-line、結合預切割和 beam search 之電話單詞辨認系統。

研究方法及步驟：

1. MAT 語料之整理與基本辨認系統

MAT 語料是經由電話線在全台收集的語音資料，目前仍然繼續在收錄當中。在此先期實驗中，我們先採用第一期收集的語料，其中語者編號為 5 的倍數者為測試資料，其餘則為訓練資料。基本系統之語音模型為 100 個右相關 initial 與 39 個 final HMM 模型，男女各建立一組模型，辨認時以音節為搜尋單位，以一階式動態搜尋法做連續音節辨認。表一為

此語料庫之特徵參數資料。表二是各小套語料的辨認結果，其中 **Baseline** 和 **SBR** 分別代表傳統 ML 訓練方法及 **Signal Bias Removal** [1] 方法；**MAT2** 之字彙為數字再加數個特定字；**MAT3** 為單音節；**MAT4** 為短詞；**MAT5** 為連續音(句子)。

由於電話線通道變化，語者偏移性及雜訊干擾，加上第一期資料庫的錄音狀況不是很好（錄音設定不一致、level saturation、笑聲、誤唸等等），故而辨認率偏低。**SBR** 方法大約提高 4-5 個百分點的音節辨認率，但仍然未達實用系統的要求。因此本年度計畫對環境效應及語者特性的整合辨認及訓練法則加以探討，以期提高辨認率。

表一. 取樣方式與特徵參數

Sampling frequency	8kHz
Frame shift	10ms
Frame length	30ms (Hamming window)
Feature parameters	12-order mel-cepstrum + 12-order Δ mel-cepstrum + Δ log energy

表二. MAT 各子語料庫辨認結果

	Baseline	SBR
MAT2	75.3	79.1
MAT3,4	34.9	41.0
MAT5	36.2	41.4

2. 語音切割與 PMC 強健式辨認

我們在以往實驗結果的分析上發現，雜訊語音的辨認率大幅下降的原

因，常是因為子音、低能量的韻尾部分和雜訊三者之間互相混淆所導致(例如出、ㄥ、ㄌ和雜訊)。因此我們認為在雜訊環境下的語音粗分類切割，不但可以據以估算環境雜訊模型，更可以直接和主辨認器的搜尋結合，用來輔助雜訊語音的辨認。圖一為我們所提出的整合切割、PMC 雜訊補償、及相似值補償的國語連續音節辨認系統 [2]，輸入語音先經 RNN 切割，再由 non-speech 部份去估計雜訊模型，用於 PMC 雜訊補償；同時使用 RNN 的輸出，分別對 HMM 的 *initial*、*final* 和 *noise* 部分進行相似值補償 (Likelihood compensation) 來增進辨認率。在一組兩男兩女 **Multi-speakers** 的基本音節辨認實驗中，測試時使用受白雜訊 (white noise) 污染之語音在 9dB、18dB 和 30dB SNR 環境下來辨認，其辨認率如表三所示。可以發現加入相似值補償之方法 (PMC/LC) 能大幅增加辨認率。

表三. 相似度補償之辨認實驗結果

SNR(dB)	PMC	PMC/LC
9	26.9	33.0
18	48.3	52.0
30	65.4	71.8

3. 強健式訓練法 (Robust Training)

在此研究主題中，主要探討在雜訊污染下，如何整合強健式辨認於模型訓練中，以直接由雜訊語料中訓練出適合辨認時所用的雜訊補償的語音模型。因為電話語音在收集時就隱含了各種雜訊在其中，使用傳統的方法

訓練出的模型會參雜了雜訊效應。我們提出一個 **Robust training algorithm** [3] 來克服此問題。在一組兩男兩女 **Multi-speakers** 的實驗中，訓練語料是由受 12dB、24dB、36dB 的直昇機雜訊 (Lyn) 污染源所構成的雜訊語音。測試時則由受直昇機雜訊污染之語音在 9dB、18dB、30dB SNR 環境下來辨認。採用標準的 PMC 雜訊補償法 [4] 來辨認。表四為實驗結果，其中 HMM1 為使用乾淨語音模型，HMM2 則使用我們提出的 **Robust training** 方法所得到的語音模型，來做 PMC 雜訊補償辨認。可以看出 **Robust training** 方法求出的模型可以得到較好的結果。雖然此實驗只考慮雜訊污染源，但相信可以於 **robust training** 方法中同時考慮語者偏移及通道效應，而在電話語音辨認上得到更好的結果。

表四.強健式訓練法之辨認實驗。

SNR(dB)	HMM1	HMM2
9	39.1	43.6
18	58.6	62.8
30	71.2	73.6

4. 離型展示系統

我們使用經 SBR 做通道偏移處理後，訓練出來的 HMM 模型，在 windows 95 作業系統下，以 TAPI 為電話電腦介面，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，再以一階式動態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，以完成所需的語音輸入與語音輸出介面，在測試台北市 1922 個金融公司電話號碼查詢下，可以得到 **Top1=93.6%**，**Top5=97.9%** 的單詞辨認率[5]。

結論：

在本計畫中，我們初步探討電話語音、語者獨立之國語連續音節辨認，除了提出雜訊語音切割及相似度補償法，以改進 PMC 強健是辨認法，並對雜訊效應提出強健式訓練法，在模擬的雜訊語料中得到不錯的結果，未來將擴展強健式訓練法使其同時能克服雜訊與通道偏移效應。

計畫成果自評：

本報告內容與原計畫內容相符，

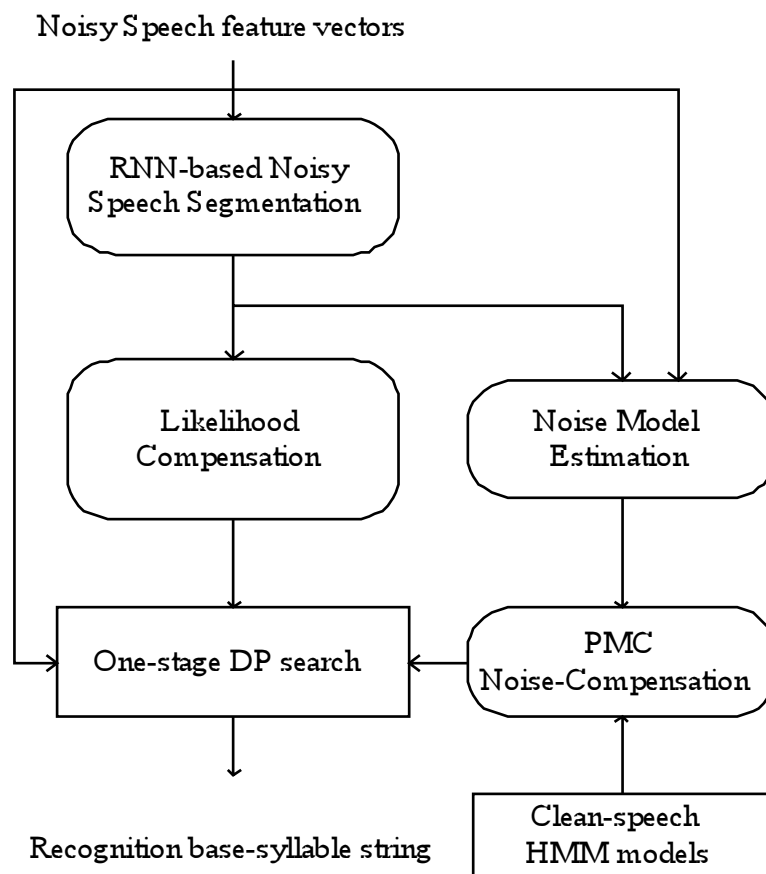
發表之論文：

1. 蕭志祥，強健式國語連續音節辨認，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, 1996.
- [2] Wei-Tyng Hong and Sin-Hong Chen, "A robust RNN-based pre-classification for Noisy

- Mandarin speech recognition,” *EuroSpeech 97*, Vol. 3, pp. 1083 - 1086, 1997.
- [3] Wei-Tyng Hong and Sin-Hong Chen, “A robust training algorithm for noisy Mandarin speech,” to appear in *ISCSLP98*.
- [4] M. J. F. Gales and S. J. Young, “Robust continuous speech recognition using parallel model combination,” *IEEE Trans. Speech and Audio Processing*. Vol 5, 352-359, 1996.
- [5] Y. R. Wang and Sin-Hong Chen, “Mandarin telephone speech recognition for automatic telephone number directory service,” *Proc. 1998 IEEE Conf. Acoust. Speech Sig. Process.* Vol 2, pp. 841-844, 1998.



圖(一) 切割與PMC補償整合式辨認系統