國語語音辨認之實用系統發展(III)

Development of Mandarin Speech Recognition Systems (III)

計畫編號：NSC-89-2213-E-009-120
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：涂家章、陳科旭、陳啟仁、武景龍、翁以哲
許亨仰　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，提出了多個有關雜訊補償及通道補償之強健式國語語音辨認法，在前兩年計畫中，完成雜訊語音切割、粗分類相似度補償法、強健式訓練法、音段式音量調適及雜訊補償法、及粗分類式訊號偏移去除法，本年度則進一步探討電話語音切割法及強健式訓練法，並提出正交轉換式訊號偏移去除法，最後完成一個國語語音辨認系統。

關鍵詞：雜訊補償，訊號偏移去除，國語語音辨認

Abstract

In this three-year project, we have proposed several noise compensation and signal bias removing methods for adverse Mandarin speech recognition. Five methods have been studied in the first two years. In this year, we further study the noisy speech segmentation method and the robust training method, and propose a new signal bias removing method. Besides, we implement a high-performance Mandarin speech recognition system.

Keywords: Noise compensation, Signal bias removing, Mandarin speech recognition.

二、緣由與目的

近年來語音辨認技術已有長足進步，一些實用系統陸續被開發出來，發展實用系統的關鍵之一在於雜訊及通道效應的去除或補償，國外對此問題已經由蒐集大量語料來廣泛地進行研究，國內在最近完成大型電話語料庫 (MAT) 之蒐集, 亦開始深入探討此問題。本計畫之目的是要使用 MAT 語料庫來進行雜訊及通道效應補償研究，以期發展實用的語音辨認系統。在前兩年計畫中，我們已提出數個方法，並經實驗證明其有效性。本年度的研究工作是對這些方法作進一步改進。以下為本年度的研究工作報告。

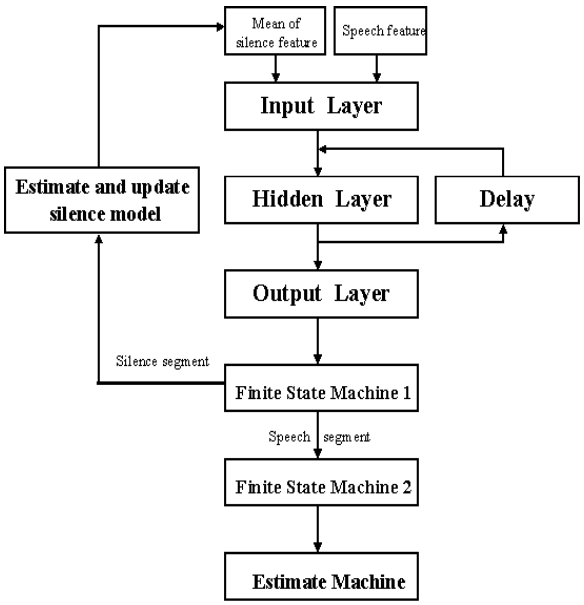
三、結果與討論：

(一) 電話語音切割法

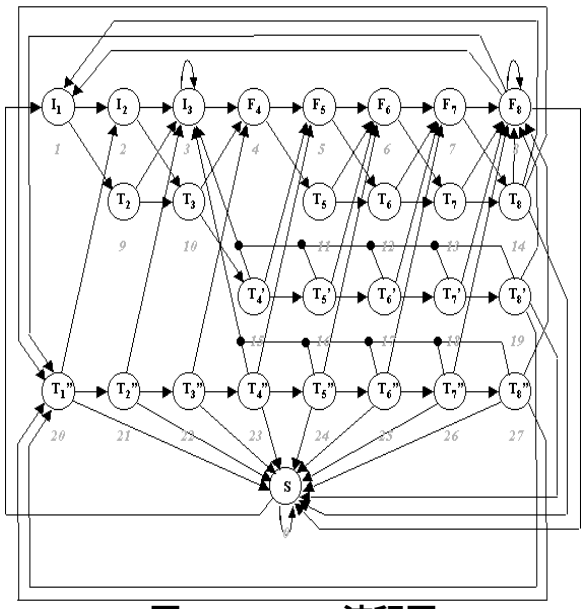
我們首先改進以前提出的 RNN-based 語音切割法，使其適用於電話語音切割。需解決的主要問題在於環境隨 channel 而改變，因此需要採用 adaptive 方式改變 RNN, 但類神經網路的 learning 一般極為緩慢，不適合採用適應性學習法則調整其參數，我們因此提出雜訊迴授方式，如圖一所示。

另外，我們改進了切割後的 finite state machine (FSM)，主要是解決以前 FSM 可能在辨認搜尋時產生沒有合法路徑的問題。它使用兩個 FSM，第一個用來做 speech/non-speech 端點切割，第二個則

用來對 speech 音段做 initial/final/breath/short-silence/transition 的精細切割，圖二為 FSM2 的流程圖，它吻合我們使用的 3-state initial 及 5-states final HMM model 音節架構。



圖一、新的雜訊語音切割法方塊圖



圖二、FSM2 流程圖

表一為使用此切割器對 MAT2000 電話語音之實驗結果，使用語料庫中的兩個子分項 DB4 （詞）及 DB5（短句），其內容包含男生 1079 人共 175305 音節，女生 1300 人共 303866 音節，由表中可看出切割之正確率很高，而混淆音框大部分發生在音段邊界處，因此對後級辨認沒有影響。表二為使用預切割之電話語音辨認實驗結果，測試語料使用 500 句連續語音，由表中可看出辨認率還稍微增加，當然速度增快許多。

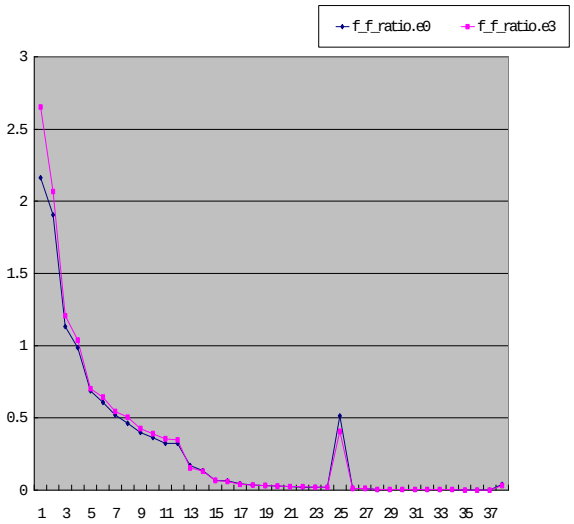
表一、使用預切割之電話語音辨認實驗結果

	Ins.	Del.	Sub.	Accu.
基本 HMM 辨認	3.26	0.74	28.9	67.1
FSM＋HMM 辨認	2.11	1.25	29.6	67.3

unit: %

(二) 電話語音強健式訓練法

我們改進以前提出用來去除雜訊的強健式訓練法，目的是要同時考慮雜訊及通道效應，以適用於電話語音辨認，圖三為其方塊圖。使用 MAT2000 來驗證此方法是否有效，圖四為 ML-trained 和本方法所得的 HMM models 的 F-ratio，由圖中可看出本方法有較高的 F-ratio，表三為它們的辨認結果比較，由表中可看出強健式訓練法有較高的音節辨認率。



圖四、ML-trained 和強健式訓練所得的 HMM models 的 F-ratio 比較

表三、ML訓練和強健式訓練的辨認率比較

	24dB	27dB	31dB
ML訓練	47.2	50	58
強健式訓練	49.6	52.2	58.3

unit: %

(三) 正交轉換式訊號偏移去除法

我們提出一個新的訊號偏移去除法以補償電話通道效應，它的基本原理是利用對辨認參數做正交轉換時，如參數中存在一個固定的偏移量，則只有zero-th order **轉換係數會受此偏移量影響，以數學式表示如下：令正交轉換以下式表示**

$$c_j(k)=\frac{1}{N+1}\sum_{i=0}^{N}\Phi_j\left(\frac{i}{N}\right)\times f_k(i)$$

其中 $f_k(i)$ **為第**k **個參數，**Φ_j **為第**j **個正交基底函數，**c_j **為第**j **個轉換係數，若參數中存在一個固定的偏移量以下式表示**

$$f_k^b(i) = f_k(i) + b_k$$

則

$$c_j^b(k)=\begin{cases} c_j(k)+b_k & \text{for } j=0 \\ c_j(k) & \text{for } j\neq 0 \end{cases}$$

基於此觀察，我們設計了以下新的訊號偏移去除法：

將每一 frame **的所有參數先轉換成正交轉換係數並構成一向量，然後以** LBG **法訓練得一** codebook**。在估計** bias **時，將測試語句的正交轉換係數對此** codebook encode**，此時只用非零階的正交轉換係數做比對；在獲得** encoding codeword **後，由** codeword **和正交轉換向量的零階係數比較而獲得** bias **估計；最後對整句話做平均而獲得** bias **估計，再由原參數中減去之。**

我們以一模擬的電話訊號做實驗來驗證此新方法的效能，表四為實驗結果，其中表(a) **第二欄表示** bias **估計的均方差，第三欄中的括號內外值分別代表** iteration 1 **次和** 10 **次的辨認率；比較**(a) (b) **兩表可看出此新方法遠較傳統** SBR **法為優。**

(四) 國語語音辨認系統

最後我們發展了一個國語語音辨認系統，使用 MAT2000 **電話語音語料庫及** TCC-300 **麥克風語音語料庫來訓練系統，其中** MAT2000 **內容包含男生** 1079 **人共** 175305 **音節，女生** 1300 **人共** 303866 **音節；**TCC-300 **則包含台大（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。此辨認系統對電話語音及麥克風語音之音節辨認率分別達到** 67.4% **及** 70.9%**，它可作為發展各種應用之核心，及未來進一步研究之基礎。**

表四、(a) **傳統** SBR method **之辨認結果**

Codeword number	Bias deviation	Syllable accuracy (%)	Relative bias estimation time
128	132.4	63.0(57.2)	0.5
256	114.6	64.6(58.7)	1.0
512	140.2	62.4(56.6)	2.0
1024	122.8	64.1(58.3)	4.0

(b)正交轉換式訊號偏移去除法之辨認結果

Window length/ shift	Bias deviation	Syllable accuracy (%)	Relative bias estimation time
4 / 1	46.4	70.1	0.21
4 / 3	46.3	70.3	0.08
6 / 1	46.0	70.0	0.21
6 / 3	45.8	69.9	0.08
8 / 1	46.7	70.1	0.21

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

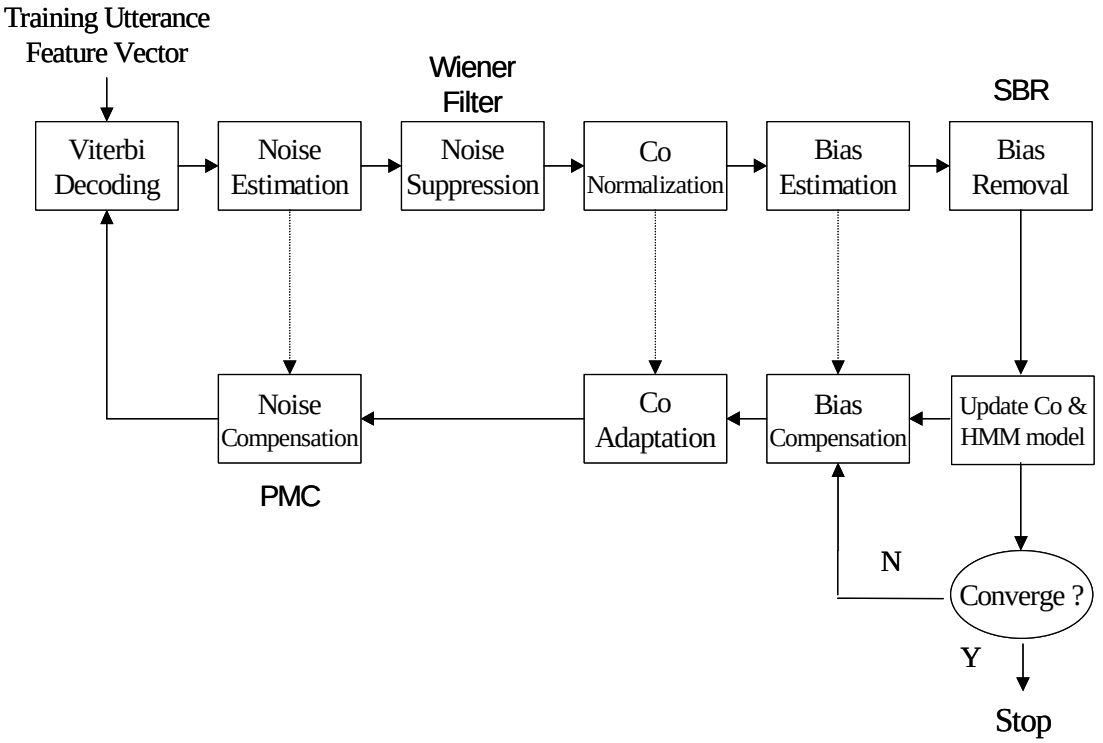
1. W. J. Wang and S. H. Chen, “Signal Bias Removal with Orthogonal Transform for Adverse Mandarin Speech Recognition,” Electronics Letters, Vol.36, No.9, pp.851-852, Apr. 2000.
2. W. T. Hong and S. H. Chen, “A Robust Training Algorithm for Adverse Mandarin Speech Recognition,” Accepted by Speech

Communications.

3. W. T. Hong and S. H. Chen, “A Segment-based C0 Adaptation Scheme for PMC-based Noisy Mandarin Speech Recognition,” ICASSP’99, Phoenix, US, March 1999.

表一、RNN-based **電話語音切割實驗結果**

上限=0.85；**下限**=0.15							
音框數	**靜音**	**聲母**	**韻母**	**呼吸聲**	**未知**	**全部**	**全部-未知**
靜音	6046119	5126	5093	340	327733	6384411	6056678
聲母	44582	2598832	172875	2109	2614066	5432464	2818398
韻母	52586	155017	6515589	506	3610782	10334480	6723698
呼吸聲	52908	5460	8321	13908	208951	289548	80597
百分比	**靜音**	**聲母**	**韻母**	**呼吸聲**	**未知**	**全部**	**全部-未知**
靜音	99.83	0.08	0.08	0.01	5.13	100.00	94.87
聲母	1.58	92.21	6.13	0.07	48.12	100.00	51.88
韻母	0.78	2.31	96.90	0.01	34.94	100.00	65.06
呼吸聲	65.65	6.77	10.32	17.26	72.16	100.00	27.84



Block diagram of the SCA-PMC method for training phase of the REST algorithm

圖三、電話語音強健式訓練法方塊圖