

# 行政院國家科學委員會專題研究計畫成果報告

## 適合網路環境下之中文口語處理技術之研究- 總計劃(3/3)

### Spoken Language Processing for Mandarin Chinese under Network Environment

計畫編號：NSC 89-2213-E-002-029

執行期限：88年8月1日至89年7月31日

主持人：李琳山 國立台灣大學資訊工程研究所

E-mail: lsl@speech.ee.ntu.edu.tw

#### 一、中文摘要

由於大眾化資訊環境已逐步由個人電腦轉移至網路，網際網路(Internet)已經成為全球性最大的資訊系統，未來使用者的終端設備將多元化，包括雙向電視(Interactive TV)、PDA等均有可能，而個人電腦只是其中之一，反而成為此一網路的視窗。由於龐大的資訊與運算能力都可能由網路隨時取得，使用者的 Client 端在記憶容量及計算速度要求上因而大為降低，而整個網路可能成為軟體系統發展的重要平台。在這樣新的網路環境下，口語語言處理的需求大為增加，而相關配合技術，包括聲學處理技術、語言模型技術、語言分析技術等也都需要有全新的面貌，以面對新的挑戰。因此，本計劃將充分利用網路資源提高語音及語言處理技術能力，以期在網路環境發展多元的口語處理應用。

關鍵詞：口語處理、語音辨認、網路

#### Abstract

Since in future network world, the user terminals may not be a PC, but the network provides plenty and plurality of resources which are also dynamic. There will be large number of users under completely uncontrollable environments and conditions. Spoken language processing in such situations will be highly challenging and very difficult, therefore require in-depth long-term scientific research. For this purpose, the role of this project is focusing on the improvement of acoustic and language processing technology with the utilization of network resources, and expected to create diverse language processing applications on the Internet.

Keywords: Chinese Spoken Language

Processing, Speech Recognition, Internet

#### 二、計畫緣由與目的

語音處理在人類發展智慧型人機介面的過程中，一直佔有相當重要的地位；而近年在電腦技術飛躍進步的環境下及大力追求自然化、多媒體、多模式化(Multi-modality)、生活化的目標下，語音在人機介面的角色更加重要。世界各主要國家為研究其本國語言之語音介面，幾乎都投入大量的人力物力。在中文社會中，由於中文鍵盤輸入難度特別高，一直阻礙中文社會全面資訊化的進展，所以發展一套便捷的口語處理技術，使中文得以方便適應新的資訊環境，尤其是十分迫切需要。台大的「語音實驗室」多年來和中研院資訊所的中文語言研究結合，在國語聽寫機的研究已經有相當成果，在過去多項研究計畫及最近三年的「實用智慧型國語聽寫機」產學合作研究計畫的大力推動下，已有第一、第二、第三代國語聽寫機「金聲一號、二號、三號」分別在80年、82年及84年完成，並於85年完成視窗九五版的「金聲三號」，至此國語聽寫機的構想已經相當接近具實用性產品，之後並透過國科會之技術移轉程序，已有業者推出初步產品，面對市場的考驗。

但在另一方面，資訊科技也在飛躍進步；大眾化資訊環境已逐步由個人電腦轉移至網路，網際網路(Internet)已經成為全球性最大的資訊系統，未來使用者的終端設備將多元化，包括雙向電視(Interactive TV)、電話、PDA等均有可能，而個人電腦只是其中之一，反而成為此一網路的視窗。由於龐大的資訊與運算能力都可能由網路隨時取得，使用者的 Client 端在記憶容量及計算速度要求上因而大為降低，而整個網路可能成為軟體系統發展的重要平台。在這樣新的網路環境下，口語語言處理

的需求大為增加，而相關配合技術，包括聲學處理技術，語言模型技術，語言分析技術等也都需要有全新的面貌，以面對新的挑戰。

### 三、結果與討論

#### 1. 聲學處理技術(子計劃一)

本年度成果如下：

##### (1) 雜訊環境語音辨認

現階段國際上最常用也有一定成效的方法之一，是所謂「平行模型加成法(Parallel Model Combination, PMC)」。我們發現 PMC 的方法其效果常常不如預期的好。它確可提高辨認的正確率，但所獲得的正確率比起真的用現場雜訊環境下的語音所訓練出來的模型的辨認率有相當的差距。進一步的研究分析，我們認為是 PMC 為了作即時轉換加成，在減少計算量的目標下用了不少的假設及近似法，那些地方事實上引進了不少誤差。經過詳細的理論分析，我們發現在轉換部分及加成部分 PMC 的方法都引進誤差，這些誤差我們也用實際語音訊號的分析來加以證實。我們進一步提出了新的轉換及加成的演算法，計算量並沒有顯著增加，卻可以大幅減少這些誤差，使得正確率有相當幅度的提昇。

##### (2) 少量語料語者調適技術

所謂語者調適，是指使用者只要使用較少量的訓練語料，就可以將系統調適至可以相當精確的辨識該使用者的聲音的程度。這個問題目前國際上最常見到的兩種作法，一是所謂「最大相似度線性回歸法」(Maximum Likelihood Linear Regression, MLLR)，另一種作法是所謂「特徵語音法」(Eigen Voice)。我們的研究發現，上述兩種方法事實上有相當的可以互相結合互補的空間，因為它們是基於不同的原理與構想，因此我們的研究發展出一系列結合這兩種技術的方法。第一個構想是用 MLLR 技術所發展出來的線性轉換公式中的參數建構一個不同的向量空間，再在這個新的向量空間中找出特徵向量及建構出特徵空間。這時又有不少可行的作法，包括究竟如何建構 MLLR 的向量空間，如何找出新語者的對應點，及如何減少所需的記憶體容量等。

##### (3) 口語對話系統之電腦分析及設計

口語對話系統涉及相當多的技術，包括語音辨認、語音瞭解、對話流程控制、對話策略設計、回覆語句建構、語音合成等部份，各部份間的關係十分複雜，始終不曾有較有效的分

析模型。我們的研究則想出了一整套用電腦模擬來進行口語對話系統的設計分析的方法，對話系統好比是一個互動式的機制，使用者藉口語對話將一系列的觀念(Concept)傳送給系統。系統一方面要一一接收到使用者的觀念，一方面要確定所收到的是對的，一方面逐步完成使用者的指示。語音辨認及語音瞭解的正確性或錯誤發生都可視為隨機程序，由電腦在模擬中預先設定自動產生，各種可能的設計目標，不論正確性，對話效率等，均可設定為參數由電腦在模擬中求出來。

#### 2. 語言處理技術(子計劃二)

為了充分利用網路資源提高語言模型技術能力，本計畫的目的在擴充語言模型能力使其適合在網路環境發展各種應用。有關這項研究主題不僅在中文世界有其迫切需求，即使在國際上也極具前瞻性與競爭力。本子計畫第三年工作項目主要為利用網路動態語料，包括文件與搜尋引擎查詢記錄等，發展不同主題領域的關鍵詞擷取技術，關聯詞抽取技術，與語音辨認結合的互動檢索技術等。計畫執行迄今上述目標已經初步達成，除了發展出一些重要技術，也完成學術論文多篇。

網路資源源源不斷，造成語言知識獲取方式的改變，而在網路環境發展應用系統，有關 CPU 與 Memory Power 將可無限供應，還有網路環境發展語言處理系統需要更加國際化，不僅須同時考慮繁體與簡體文字現象，也需要考慮英語語言現象，此外網路環境語言處理系統須適合多元應用，包括電話介面、PDA 介面以及雙向電視等。因此在網路上發展語言模型技術，語音聽寫機並非唯一應用，其他如手寫，OCR，資料庫檢索查詢也同樣有高度需求。這樣網路環境所造成語言處理技術需求正是本計畫的基礎。

本子計畫第一年工作項目包括完成語料自動收集技術(Automatic Corpus Collection)，高階語言模型(High-Order Language Model)，動態儲存管理(Dynamic Access and Storage)與快速訓練(Fast Training)。第二年的研究也已經發展成功網路動態語料自動分類與儲存技術，動態關鍵詞擷取技術，主題類似雙語語料自動收集與關鍵詞翻譯抽取技術，這些技術對發展網路環境的語言處理技術需求都是相當重要的基礎。第三年計劃主要為利用網路動態語料，包括文件與搜尋引擎查詢記錄等，發展不同主題領域的關鍵詞擷取技術，關聯詞抽取技術，與語音辨認結合的互動檢索技術等。

在我們沿用第一年發展出的派樹(PAT-tree)為主要資料結構發展各項技術。為了有效收集網路資源我們發展分類技術，建立分類語料庫，由於PAT-tree適合表達n-gram語言模型，因此所建立的分類索引即可建立成分類語言模型。同時考慮提升語音辨認或其他語言收集領域詞彙的不易，我們還發展動態關鍵術語抽取技術，建立所謂動態辭典(Live Dictionary)。這種技術是基於動態語言模型的構想，希望語言模型的參數訓練是開放式的可以不斷隨訓練語料改變而更新，同時抽取出的術語也可隨資料更新而改變。進而為了網路雙語應用需求我們也研究術語翻譯抽取技術。

第三年起我們的研究著重利用子計畫第三年工作項目主要為利用網路動態語料，包括文件與搜尋引擎查詢記錄等，發展不同主題領域的關鍵詞擷取技術，關聯詞抽取技術，與語音辨認結合的互動檢索技術等。以上技術也成功應用在語音資訊檢索。

### 3. 網路文件之語言分析(子計劃三)

語言分析，語音辨認等高階語文處理技術，都需要深度的語言知識及語言規律的支持，否則電腦不容易處理複雜的語言現象，如果電腦對處理的輸入語文類型一無所知，用一般的語言規律做語句分析，用一般的語言模型做語音辨識，效果一定很差，因為不同領域，不同語式的文件句型用詞有相當大的差異，因此語文處理時對文件領域做調適的步驟，非常重要。由於網際網路的興起，網路資源的檢索、分類、應用成為一門非常重要的學問，如何以電腦輔助使網路資源發揮最大的效用，是本計畫的一個主要目的。希望以總計畫發展出的語音處理介面，使用者可以透過語音，以自然語言的方式，找到他所需要的文件或資訊。本計畫前兩年的工作是利用自動化方式收集網路上的文件，研究文件的自動分類及不同類別文件的關鍵詞自動抽取技術。從中研院平衡語料庫估算，未知詞在實際的文章中約佔3.51%，由於這些詞無法直接從辭典中獲得詞類及語意訊息。在[Chen et al. 97]中曾經以詞首字及詞尾字和詞類的關係來猜測未知詞的詞類。本年度計畫針對未知詞中名詞性複合詞的語意做猜測研究。我們提出一種近似模型來表達複合詞的構詞律，這個模型可以用來一、作為構詞律，二、評估結構好壞的工具，三、解決複合詞語意的歧義。

近似模型以語法語意及統計機率關係表達開放型未知詞組成分的構詞律模型，這個

模型用來表達及測度複合詞的結構合理程度。我們以複合詞為例，從語料庫中，自動抽取各類複合詞中的詞幹意類，以作為訓練規則，透過比對一未知詞詞幹意類與規則意類間的相似度，判斷其是否為合理的複合詞。對於一含有高頻詞尾(-Y)的未知詞XY，我們可以利用詞幹語意統計，從其詞幹(X)詞尾(-Y)間的語意關係，判斷其作為一個複合詞的合理程度(即該未知詞是一個複合詞的可能性)。我們的方法是，以300萬目詞的語料庫(中研院平衡語料庫)作訓練資料，用程式從中自動抽取該未知詞詞尾所衍生的複合名詞，統計其詞幹(X)之語意類分佈(Semi)，然後比較該未知詞詞幹之語意類Sem和構詞規則內各語意類Semi的近似程度。

我們以語料中任意抽取兩百個未知複合名詞作測試，結果發現電腦對複合詞的語意偵測，正確率可以達到82%。對歧義詞的判定正確率也可以達到81%。

### 四、計畫成果自評

本計畫原預期進行之工作項目皆已達到一定成果，詳見前述「結果與討論」，且大部份的成果都已在(或即將在)相關國際研討會及重要期刊發表。

### 五、參考文獻<sup>1</sup>

- [1] Jia-lin Shen, Wen-liang Hwang, "New Lin-shan Lee, Yumin Lee, "Voice Access of Global Information for Broadband Wireless: Technologies of Today and Challenges of Tomorrow" (invited paper), to appear on Proceedings of the IEEE, Feb. 2001.
- [2] Jieh-weih Hung, Jia-lin Shen and Lin-shan Lee, "New Approaches for Domain Transformation and Parameter Combination for Improved Accuracy in Parallel Model Combination (PMC) Techniques", paper accepted by IEEE Transactions on Speech and Audio Processing.
- [3] Bor-shen Lin, Lin-shan Lee, "Computer-aided Analysis and Design for Spoken Dialogue Systems Based on Quantitative Simulations", paper accepted by IEEE Transactions on Speech and Audio Processing.
- [4] Tai-Hsuan Ho, Chin-Jung Liu, Herman Sun, Ming-Yi Tsai, Lin-shan Lee, "Phonetic State Tied-Mixture Tone Modeling for Large Vocabulary Continuous Mandarin Speech

<sup>1</sup> 參考文獻均為執行本計畫的相關著作。

- Recognition”, Sixth European Conference on Speech Communication and Technology, Budapest, Hungary, Sept. 1999, pp. 883-886.
- [5] Bor-Shen Lin, Hsin-min Wang, Lin-shan Lee, “Consistent Dialogue across Concurrent Topics Based on An Expert System Model”, Sixth European Conference on Speech Communication and Technology, Budapest, Hungary, Sept. 1999, pp. 1427-1430.
- [6] Fu-Chiang Chou, Chiu-Yu Tseng, Lin-shan Lee, “Selection of Waveform Units for Corpus-based Mandarin Speech Synthesis Based on Decision Trees and Prosodic Modification Costs”, Sixth European Conference on Speech Communication and Technology, Budapest, Hungary, Sept. 1999, pp. 2295-2298.
- [7] Bor-shen Lin, Hsin-ming Wang, Lin-shan Lee, “A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History”, IEEE Automatic Speech Recognition and Understanding Workshop, Keystone, Colorado, USA, Dec. 1999.
- [8] Lin-shan Lee, “Programs and Activities on Chinese Spoken Language Processing”, IEEE Automatic Speech Recognition and Understanding Workshop, Keystone, Colorado, USA, Dec. 1999.
- [9] Lin-shan Lee, Lee-Feng Chien, “Live Lexicons and Dynamic Corpora Adapted to the Network Resources for Chinese Spoken Language Processing Applications in an Internet Era”, 2<sup>nd</sup> International Conference on Language Resources and Evaluation, Athens, Greece, May–June 2000, pp. 931-936.
- [10] Berlin Chen, Hsin-min Wang, Lin-shan Lee, “Retrieval of Broadcast News Speech in Mandarin Chinese Collected in Taiwan Using Syllable-level Statistical Characteristics”, IEEE International Conference on Acoustics, Speech and Signal Processing, Istanbul, Turkey, June 2000, SP-P9.14, pp. III-1771-1774.
- [11] Bor-shen Lin, Lin-shan Lee, “Fundamental Performance Analysis for Spoken Dialogue Systems Based on A Quantitative Simulation Approach”, IEEE International Conference on Acoustics, Speech and Signal Processing, Istanbul, Turkey, June 2000, SP-L9.2, pp. II-1221-1224.
- [12] Lin-shan Lee, Lee-Feng Chien, Yumin Lee, “Global Information Access by Chinese Spoken Language in A Wireless Era — Overview with Some Recent Results”, International Symposium on Chinese Spoken Language Processing, Oct. 2000, Beijing, China.
- [13] Bor-shen Lin, Lin-shan Lee, “Computer-Aided Design/Analysis for Chinese Spoken Dialogue System”, International Symposium on Chinese Spoken Language Processing, Oct. 2000, Beijing, China.
- [14] Berlin Chen, Hsin-ming Wang, Lin-shan Lee, “Retrieval of Mandarin Broadcast News Using Spoken Queries”, International Conference on Spoken Language Processing, Oct. 2000, Beijing, China.
- [15] Kuan-ting Chen, Wen-wei Liao, Hsin-ming Wang, Lin-shan Lee, “Fast Speaker Adaptation Using Eigenspace-based Maximum Likelihood Linear Regression”, International Conference on Spoken Language Processing, Oct. 2000, Beijing, China.
- [16] Jieh-wei Hung, Hsin-ming Wang, Lin-shan Lee, “Automatic Metric-based Speech Segmentation for Broadcast News via Principal Component Analysis”, International Conference on Spoken Language Processing, Oct. 2000, Beijing, China.
- [17] Hsiao-Chuan Wang, Frank Seide, Chiu-yu Tseng, Lin-Shan Lee, “MAT-2000: Design, Collection, and Validation of a Mandarin 2000-Speaker Telephone Speech Database”, 6th International Conference on Spoken Language Processing, Oct. 2000, Beijing, China, Vol.IV, 460-463.
- [18] Lee-Feng Chien, et al., Incremental Extraction of Domain-Specific Terms from Online Text Collections, *Recent Advances in Computational Terminology*, ed. By D. Bourigault et al., Book in preparation, 2000.
- [19] Chien-Kang Huang, Lee-Feng Chien, Yen-Jen Oyang, Clustering of Similar Query Sessions Towards Interactive Web Search, ROCLING-2000.
- [20] S. L. Chuang, Hsiao-Tieh Pu, W. H. Lu, Lee-Feng Chien, Auto-construction of a Live Thesaurus from Search Term Logs for Interactive Web Search (Poser Paper), ACM SIGIR’2000, Greece.
- [21] W. H. Lu, S. L. Chuang, Hsiao-Tieh Pu, W. H. Lu, Lee-Feng Chien, Live Thesaurus Construction for Interactive Voice-based Web Search, ICSLP’2000.
- [22] Chen, K.J. & Chao-jan Chen, 2000, “Automati Semantic Classification for Chinese Unknown Compound Nouns,” *Coling 2000*.