

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(I) - 總計劃 Spoken Language Processing for Mandarin Chinese under Network Environment

計畫編號：NSC 87-2213-E-002-017

執行期限：86 年 8 月 1 日至 87 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了。

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)

近年來在語音辨認技術逐漸成熟及實際應用需求快速成長之下，各種口語對話系統（Spoken dialogue systems）被廣泛地探討，關鍵詞擷取（Keyword Spotting）技術咸認為處理這類的語音辨認問題最有效的方式。關鍵詞擷取並不需要把語者所說的每一個字詞辨認出來，只需要知道特定的若干個關鍵詞是否在句子中出現，在處理可能包含遲疑（Hesitations）、重複（Repetitions）、詞典外詞彙（Out-of-vocabulary）等口語化的自發性語音（Spontaneous speech），或是考慮不同應用的可攜性（Portability）時，關鍵詞擷取可能比大字彙語音辨認（Large-vocabulary speech recognition）更有實用價值，技術上也更為可行。本年度在關鍵詞擷取技術的研究方面，主要探討填充模型（Filler models）及語段確認（Utterance verification）技術，已完成國語關鍵詞擷取核心架構，並進一步與其他應用系統結合，包括國語語音資料檢索系統、自動語音查號對話系統等。

在語音辨認系統由實驗室走向實際應用的過程中，不同語者和環境產生的變異性（speaker and environmental variations）往往造成辨認的正確率大幅下降，因此系統的強健性（Robustness）成為不可或缺的一環。其中包括了語者不特定（Speaker-independent）的處理、背景雜訊的消除（noise cancellation）以及通道雜訊的補償（channel noise compensation）。在過去一年的研究中，我們首先利用三連音素（triphone）來建立聲學模型，以增加處理自發性

口語語音（spontaneous speech）的能力，同時利用決策樹（decision tree）的建立，使得聲學模型的參數能共享，進一步改善模型的強健性。此外提出一個相關性平行模型整合（correlated parallel model combination）的演算法來提升系統在有背景雜訊下的正確率。另外，利用倒頻譜偏移消除（cepstral bias removal）的技術來補償通道雜訊造成的影響。接著，我們進一步提出一個時域特徵參數（temporal features）的改良性演算法，來增強語音特徵參數在雜訊環境下的鑑別力（discriminability）及一種基於亂度（Entropy）的語音信號端點偵測（End-point Detection）法。此外，為了能更有效的結合聲學處理及語言學處理，我們首先在聲學處理的輸出改為非對齊（non-aligned）的音節格構（syllable lattice）以取代以前所採用的排列音節格構，其目的為提高聲學處理輸出的涵蓋率（inclusion rate），以降低語言學處理的複雜度，而為了處理此非對齊的音節格構，我們在搜尋上也作了若干的改進，以增快搜尋的速度。在語者調適技術方面，完成幾種 State-of-the-art 技術（包括：maximum a posteriori (MAP)，maximum likelihood linear regression (MLLR)，vector field smoothing (VFS)）的實驗探討，找到具加成性的結合方式，達到快速且精確的模型調適效果。利用以上所提出來的技術，我們成功的整合並完成一個在電話環境（telephone environments）下的語音辨認系統，進一步將與其他應用系統結合，包括國語語音資料檢索系統、自動語音查號對話系統等。以上成果均有論文發表在語音處理領域重要期刊及會議。

2. 語言處理技術(子計劃二)

目前第一年計劃已經完成，我們利用派樹（PAT-tree）為主要資料結構發展前述各項技術。目前為止我們已改進 PAT-tree 結構使得大量語料得以便利存取。對 1GB3 的文字資料一般只需約兩倍資料量大小（即約 2GB）索引空間但卻可存取任一字串的頻率、相鄰字串等訊息，使得包括字串間的 Mutual Information 或 Association，甚至馬可夫機率參數皆可容易獲取以及隨時更新。利用此基本資料結構，配合資訊蜘蛛技術

(Information Spider)已可由網路大量且自動抓取資源，配合子字畫 III 分類技術，可以建立分類語料庫，建立分類語言模型。從網路資源收集角度，可改進傳統語料收集方法有語料種類少，成本高，不太容易具備分類訊息並且時效差等缺失。另外，我們還進一步發展出高階語言模型 (High-Order Language Model)，由於傳統馬可夫語言模型仍有許多問題，如受限於個人使用環境以及語料取得都是小型 (Small-Scale) 模型，且只使用 Bigram 技術因此如此無法掌握 (Modeling) 長距離 (Long Distance) 現象以及會有許多 Unseen Parameters。進而利用此結構我們也研究動態語言模型 (Dynamic Language Model)。動態語言模型的構想是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。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。此外，上述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)

本計劃在整合性計劃中扮演的是技援性質的角色，文件自動分類可以將網路資源分類，做為檢索之用，同時用來建立不同領域的語言模型及詞彙庫，以增進語音處理及語言處理的能力。而不同領域詞彙庫的建立是透過關鍵詞自動辨識得到的有關文件自動分類的研究，多以文件所含詞彙為特徵向量比較不同類型之間特徵向量的夾角做為區分的方法。本計劃則進一步擴大文件的特徵選取，以找到更精確的自動分類方法，在關鍵詞多為該領域的特殊用語，通常不為一般詞典所收錄，譬如專有名詞中的人名、公司名、地名、物品名等，這些關鍵詞無法用現有的中文分詞技術判斷出來，目前有一些研究成果，多用統計形式的作法，來辨別人名等少數特殊專有名詞，對一般性的關鍵詞欠缺全面性的考慮及有效的辨識方法。本計劃則針對所有可能不同類別的關鍵詞，以離線及線上作業兩種方式辨認可能的關鍵詞。這個工作除了對資訊檢索極為重要，而且對不同類文件的語言分析也是必要的步驟之一。

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¹

- [1] Jia-lin Shen, Wen-liang Hwang, "New Temporal Features for Robust Speech Recognition with Emphasis on Microphone Variations", *to appear in Computer Speech and Language*, 1998.
- [2] Jieh-weih Hung Jia-lin Shen, Lin-shan Lee, "Improved Robustness for Speech Recognition Under Noisy Conditions Using Correlated Parallel Model Combination", *ICASSP'98*, pp. 553-556.
- [3] Tai-Hsuan Ho, Kae-Cherng Yang, Kuo-Hsun Huang and Lin-Shan Lee, "Improved Search Strategy for Large Vocabulary Continuous Mandarin Speech Recognition", *ICASSP'98*, pp. 825-828.
- [4] Jia-lin Shen, Lin-shan Lee, "Improved Robust Speech Recognition Considering Signal Correlation Approximated by Taylor Series", *ICSLP'98*.
- [5] Jia-lin Shen, Wen-liang Hwang, "Statistical Integration of Temporal Filter Banks for Robust Speech Recognition Using Linear Discriminant Analysis (LDA)", *ICSLP'98*.
- [6] Jia-lin Shen, Jieh-weih Hung and Lin-shan Lee, "Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments", *ICSLP'98*.
- [7] Jieh-weih Hung, Jia-lin Shen, and Lin-shan Lee, "Improved Robustness for Speech Recognition Under Noisy Conditions Using Modified Parallel Model Combination", *ICSLP'98*.
- [8] Jia-lin Shen, In-Jeih To, Bo-Yu Liang, and Lin-shan Lee, "Speaker-Independent Continuous Mandarin Speech Recognition Under Telephone Environments", *ROCLING-XI*, 1998.
- [9] Jia-lin Shen, Bo-Yu Liang, and Lin-shan Lee, "Decision Tree Clustering for Acoustic Modeling in Speaker-Independent

¹ 參考文獻均為執行本計畫的相關著作。

- Continuous Mandarin Speech Recognition”, *ISCSLP’98*.
- [10] Berlin Chen, Hsin-min Wang, Lee-feng Chien, and Lin-shan Lee, “A*-Admissible Key-Phrase Spotting with Sub-Syllable Level Utterance Verification”, *ISCSLP’98*.
- [11] Berlin Chen and Hsin-min Wang, “A Vocabulary-Flexible Chinese Keyword Spotting System”, *RAMS’98*, pp. 176-180.
- [12] Bo-Ren Bai, Chun-Liang Chen, Lee-Feng Chien, Lin-shan Lee, Intelligent Retrieval of Dynamic Networked Information from Mobil terminals Using Spoken Natural Language Queries, *IEEE Transactions on Consumer Electronics*, NO. 1, Vol. 44, pp. 62-72, 1998 .
- [13] Sung-Chien Lin and Lee-Feng Chien , Robust Retrieval of Chinese Document Databases with Unconstrained Speech Queries, accepted and to appear on special issue on “Information Retrieval on Oriental Languages”, *Computer Processing of Chinese and Oriental Languages*, World Scientific Press, 1998.
- [14] Chun-Liang Chen ,Bo-Ren Bai, Lee-Feng Chien, Lin-Shan Lee,PAT-tree-based Language Modeling with Initial Application of Chinese Speech Recognition Output Verification, *International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Singapore, 1998.
- [15] Bo-Ren Bai, Berlin Chen, Hsin-Min Wang, Lee-Feng Chien, Lin-Shan Lee, Large-Vocabulary Chinese Text/Speech Information Retrieval Using Mandarin Speech Queries, *International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Singapore, 1998.
- [16] Chun-Liang Chen and Lee-Feng Chien, OCR Text Verification and Retrieval Using a PAT-tree-based Ngram Indexing Technique, The First Asia Digital Library Workshop, Hong Kong ,168-175, Aug. 1998.
- [17] Chun-Liang Chen, Bo-Ren Bai, Lee-Feng Chien and Lin-Shan Lee’ CPAT-Tree-Based Language Models with an Application for Text Verification in Chinese, *ROCLING’98*.
- [18] Lee-Feng Chien, Min-Chan Chen, Chun-Liang Chen, Bo-Ren Bai, Internet-based Chinese Text Corpus Classification and Domain-specific Keyterm Extraction, to appear on *First workshop on Computational Terminology (ACL/COLING-98 COMPUTERM Workshop)*, Montreal, Canada, Aug., 1998.
- [19] Lee-Feng Chien, Min-Jer Lee, Hsiao-Tiech Pu and Lin-Shan Lee, Recent Results on Internet-based Chinese Text Corpus Classification Using Automatically Extracted Keywords, *First Oriental COCOSDA Workshop*, Japan, 1998.
- [20] Kai-Chen Yang, T. Ho, Lee-Feng Chien, L. S. Lee, Statistics-based Segmentation Patterns-A New Direction for Chinese Language Modeling, *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, USA (ICASSP’98).
- [21] Sung-Chien Lin, Lee-Feng Chien, Lin-Shan Lee,Multi-language Model for Mandarin Speech Recognition, *Proceedings of 5Th European Conference on Speech Communication and Technology* , Greece (EuroSpeech’97).
- [22] Keh-Jiann Chen, WenTsuei, and Lee-Feng Chien, “PAT-Trees with the Deletion Function as the Learning Devices for Linguistic Patterns”, *COLING-ACL ’98*.