

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(1)

子計畫一：適合網路環境下之國語聲學處理技術之研究(1)

Acoustic Processing Technology for Mandarin Chinese under Network Environment

計畫編號：NSC 87-2213-E-002-018

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

發展「適合網路環境下之中文口語處理技術」的目標是為迎向未來的網路資訊世界，開拓口語處理技術的新天地，本子計畫的任務則是在聲學處理技術方面推動前瞻性的學術研究，並考慮在新的網路環境下聲學處理技術所面臨的新挑戰。在未來的世界裡，網路將成為全球性的最大資訊系統，使用者的終端設備將多元化，同時語音介面需求將大增，各種應用環境亦使聲學問題複雜化，在這樣的環境下口語處理將極具挑戰性，聲學處理的難度也極高，故需此一計畫來作長期深入的探討。

關鍵詞：口語處理、語音辨認、網路

Abstract

To develop the "Chinese Spoken Language Processing Technology under Network Environment" is to face the new era of network information world. The role of this project is then focused on the acoustic processing technology in the group project, considering the constraints and challenges under network environment. In future network information world, the user terminals may not be a PC and 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. Acoustic processing in such situations will be highly challenging and very difficult,

therefore require in-depth long-term scientific research.

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。

在聲學處理方面，由於人機(人與網路)的互動大幅增加，文字輸入處理已經不再是最主要的電腦應用，故語音聽寫輸入可能不再是最主要的應用方向；諸如網路瀏覽搜尋、資訊檢索、電話業務及應用、對話系統等都可能和聽寫機有相當不同的環境；一方面網路應用繁多，使用者及使用狀況均使得十分複雜，語音、發聲型態、背景雜訊、干擾等聲學條件將大幅改變並不易掌握，不可能再像聽寫機那樣可以單純化，又一方面在網路環境下可在 Server 提供龐大的計算資源，系統設計未必在受限於個人電腦條件，因此整個技術發展有了新的空間。為了充分利用網路資源提高聲學處理技術能力，本計畫預計以三年時間擴充聲學處理能力使其適合在網路環境發展種各應用。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。

三、結果與討論

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）技術咸認是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。

四、計畫成果自評

本計畫原預期進行之工作項目包括：

1. 前端訊號處理及特徵擷取
 2. 聲學模型研究
 3. 連續語音之進一步辨認技術
 4. 不同語者特性變化之研究
 5. 背景雜訊及干擾消除技術
 6. 電話線之語音辨認技術
 7. 關鍵詞偵測(Keyword Spotting)技術
 8. 自發性語音之處理技術
- 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，共計完成學術論文 11 篇，詳見參考文獻[1-11]。

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] Jieih-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, Jieih-weih Hung and Lin-shan Lee, "Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments", *ICSLP'98*.
- [7] Jieih-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 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", *ICSLP'98*.
- [11] Berlin Chen and Hsin-min Wang, "A Vocabulary-Flexible Chinese Keyword Spotting System", *RAMS' 98*, pp. 176-180.

¹ 參考文獻[1-11]均為執行本計畫的相關著作。