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(3/3)-
子計畫一：適合網路環境下之國語聲學處理技術之研究(3/3)
Acoustic Processing Technology for Mandarin Chinese under Network Environment

計畫編號：NSC 89-2213-E-002-030
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

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 Compatible to Network Environment" is to face the new era of network information world. The role of this project is then focusing 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: 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，之後並透過國科會的技術移轉程序，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；大眾化資訊環境以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(或人與網路)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便得十分複雜，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，系統設計未必再受限於個人電腦的條件，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，第一年計畫已於 87 年 7 月結束，並有豐碩的成果產出，今年邁入第二年，亦有非常不錯的成果。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

本計畫今年度主要成果如下：

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)。其基本原理為現場即時抽取一小段雜訊，製作成雜訊模型，再和在安靜環境下的語音所訓練出來的語音模型加成，就可至做出在當時現場雜訊環境下的語音模型，用來作當時雜訊環境下的語音辨識。

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，我們發現在轉換部份及加成部份 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升。這裡面包括轉換過程的不對稱高斯法及混合分裂法及加成部份的交叉估測法。我們進一步證明這些方法並可以和目前國際上發表的一些其他的雜訊處理方式如向量泰勒展開法等互相加成，獲得更好的結果。

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，就能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。當使用者的訓練語料非常少時（例如只有幾秒鐘的語料，亦即在使用者開口的幾秒鐘之內就馬上可以相當精確的辨識他的聲音），這個問題的難度就變得相當高。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)，是利用少量訓練語料找一個線性轉換的公式。另一種作法是所謂「時徵語音法」(Eigen Voice)，是找出大量訓練語者的訓練語料所構成的向量空間的特徵向量。

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語構想，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，如何找出兩種技術確實可以獲得更好的結果，而不同的結合方法也可得到不同的好處，仍在進一步仔細分析中。

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。於是電腦可以在軟體上作大量的實驗，即使根本沒有對話系統製作出來。也可測試例如一百萬人使用所獲得的正確率或觀念傳送效率等，並詳細分析諸如不同的對話流程控制、對話策略設計，不同的語音辨認效果或語音瞭解機制各會獲得怎樣的效果等等。這一套軟體模擬的方法因此可以用來進行各種對話系統的設計分析。設計人員可以在設計分析完成後再去製作真的對話系統，大幅節省人力時間。

以上許多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

本計畫原預期完成之工作項目包括:

1. 前端訊號處理及特徵擷取
2. 聲學模型研究
3. 改進模型之學習功能
4. 改進音韻片段的聲學處理技術
5. 連續語音之進一步辨認技術
6. 不同語者特性變化之研究
7. 背景雜訊及干擾消除技術
8. 電話線之語音辨認技術
9. 關鍵詞偵測(Keyword Spotting)技術
10. 自發性語音之處理技術
11. 聲學和語言學處理之整合與分離研究
12. 語者驗證技術

皆已如期順利推展，詳見過去三年報告中的「結果與討論」。整體而言，研究成果相當豐碩，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，本年度共計完成學術論文17篇，詳見參考文獻[1-17]。在人才培育方面共訓練出1名博士及4名碩士。

五、參考文獻[1]

[1] 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] Jeih-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 Recognition", Sixth European Conference on Speech Communication and Technology, Budapest, Hungary, Sept.

[1] 參考文獻[1-17]均為執行本計畫的相關著作。

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", 2nd 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 Liau, 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] Jeih-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.