

## 本期要目

- |   |         |
|---|---------|
| 壹. 第十一屆理監事名單                                      | 第三~四頁   |
| 貳. 出席 2009 年國際學術會議感想 (鄭秋豫)                        | 第五~十頁   |
| 參. 專文- 語音辨識中於語音特徵時間序列域與調變頻譜域之<br>強健性處理技術之介紹 (洪志偉) | 第十一~二十頁 |

### 第九屆博碩士論文獎得獎名單

#### 博士論文獎

優等獎一名：獲獎金二萬元及獎狀

得獎者：古倫維 (臺灣大學資訊工程所)

論文：意見分析之研究與應用

A Study on Opinion Analysis and its Applications

指導教授：陳信希 教授

佳作獎一名：獲獎金一萬元及獎狀

得獎者：吳孟淞 (成功大學資訊工程學系)

論文：貝氏鑑別性資訊檢索模型

Bayesian Discriminative Models for Information Retrieval

指導教授：簡仁宗 教授

#### 碩士論文獎

優等獎一名：從缺

佳作獎三名：獲獎金伍仟元及獎狀

1. 得獎者：李政德 (臺灣大學資訊網路與多媒體研究所)

論文：異質性社群網路探勘：中心度、分群、資訊摘要

Mining Heterogeneous Social Networks: Centrality, Clustering, and Abstraction

指導教授：林守德 教授

2. 得獎者：關松堅 (成功大學資訊工程學系)  
論文：應用階層懲罰機率之 FOIL 及多維語言模型於英語錯誤句之分類與訂正  
Hierarchical Penalized Probabilistic FOIL and Multidimensional Language Model for ESL Error Classification and Correction

指導教授：吳宗憲 教授

3. 得獎者：周裕倫 (交通大學電信工程學系)  
論文：中文自發性語音之韻律標記及韻律模式

Joint Prosody Labeling and Modeling for Mandarin Spontaneous Speech

指導教授：王逸如 教授

### ROCLING-2009 會議圓滿結束

由國立中興大學資訊科學與工程學系與本會共同主辦的「第二十一屆自然語言與語音處理研討會」已於 98/9/2 在台中中興大學理學大樓演講廳順利圓滿結束，參與此次盛會的人士分別來自美國及台灣，與會人數多達 200 人次。本次會議共收錄了 20 篇口頭報告論文及 10 篇海報論文。李佳蓉小姐、林怡君小姐、陳瑞呈先生、劉培森博士及鄭卜壬教授共同著作之「Query Formulation by Selecting Good Terms」獲得最佳論文獎，於會議閉幕式中，分別獲頒獎狀乙紙，並共同獲頒獎金伍千元。會議論文已建置於本會網站：  
[http://www.aclclp.org.tw/pub\\_proce\\_c.php](http://www.aclclp.org.tw/pub_proce_c.php)。

## 廣義知網中文詞知識庫 開放申請

廣義知網(E-HowNet)為詞彙語義表達系統，由兩個主要成分組成：

- (一) 詞彙知識表達：其內容包含於中文詞知識庫，中文詞知識庫為一包含九萬目詞的電子辭典。本詞庫收的詞包含一般用詞、常用專有名詞、成語、慣用語、常用派生詞、異體詞、合併詞以及少數特殊領域用語和古漢語詞語。每個詞項包含的訊息如下：1. 詞條(中文詞) 2. 詞頻 3. 發音(包括注音符號及漢語拼音) 4. 詞類 5. 英文翻譯 6. 廣義知網概念式及其展開式。

註：中文詞庫新增加了第 5 和第 6 項的語義資訊，以”毛筆”這一個詞條為例，新增加的部分如下：

writing brush

def: {筆:material={毛}}

def: {PenInk|筆墨:telic={write|

寫:instrument={~}},material={hair|毛}}

- (二) E-HowNet 義元(primitive concepts)知識架構(ontology)。

詳細說明請參閱「廣義知網知識本體架構」網頁：<http://ckip.iis.sinica.edu.tw/taxonomy/>。

「廣義知網中文詞知識庫」由中央研究院中文詞知識庫小組執行及研究，已授權本會發行。申購語料庫工本費二萬元，已申購「中文詞庫」者，只需支付工本費一萬元。欲申請者請至本會網站查詢相關資訊([http://www.aclclp.org.tw/use\\_ckip\\_c.php](http://www.aclclp.org.tw/use_ckip_c.php))。

## 中研院資訊所特聘研究員許聞廉的 研究團隊在「國際分子生物競賽」： 蛋白質名詞搜尋項目中， 取得第一名之佳績

由中央研究院資訊所許聞廉教授所帶領的智慧型代理人實驗室(IASL)研究團隊(此實驗室同時也發展「自然輸入法」)，與元智大學資訊系蔡宗翰教授的團隊(IISR)合作，於 2009 年西班牙馬德里舉行的 BioCreAtIvE II.5 蛋白質名詞搜尋之「國際分子生物競賽」項目中，取得第一名的佳績。該競賽的參賽者包括來自亞利桑那州立大學(2007 年上一屆競賽第一名團隊)、日本東京大學、科羅拉多丹佛大學以及歐洲等世界各地的頂尖團隊。

該競賽在全文的語料上發展現今最先進之蛋白質基因名詞搜尋技術。蛋白質之間的相互作用對生物學家極為重要。這項技術能夠大量簡化生物資料庫建構者在進行全文資料庫建構時所要進行的繁雜流程。跟其它的競賽相比，本屆的 BioCreAtIvE II.5 採用了一種全新的評分方式：利用在線伺服器(online server)立即驗證的方法，來評估參賽團隊處理全文文獻的能耐。在該競賽中，IASL-IISR 針對全文處理蛋白質名稱，提出最佳化的方法，並搭配以支持向量機(SVM)為基礎的評分方式，在競賽中獲取此佳績。

IASL-IISR 團隊成員：戴鴻傑、賴柏廷、黃吉心、張晏菁、鮑岳洋、吳欣達、蔡宗翰教授與許聞廉教授。

# 中華民國計算語言學學會

## 第十一屆理監事名冊

職稱	姓名	性別	現職及學歷	兼任職務
理事 (常務理事)	王新民	男	中央研究院資訊所副研究員	
			臺灣大學電機工程博士	
	古鴻炎	男	臺灣科技大學資訊工程學系副教授	
			臺灣大學資訊工程博士	
	余明興	男	中興大學資訊工程系副教授	
			臺灣大學資訊工程博士	
理事長	吳宗憲	男	成功大學資訊工程所教授	
			成功大學電機工程博士	
	高照明	男	臺灣大學外文系助理教授	電腦輔助語言教學組 召集人
			英國曼徹斯特科技大學語言工程博士	
	張景新	男	暨南國際大學資訊工程系助理教授	
			清華大學電機工程博士	
	張智星	男	清華大學資訊工程系教授	
			美國加州大學柏克萊校區電機博士	
副理事長	許聞廉	男	中央研究院資訊所研究員	
			美國康乃爾大學作業研究博士	
	陳光華	男	臺灣大學圖書資訊系副教授	期刊總編輯
			臺灣大學資訊工程博士	
	陳柏琳	男	臺灣師範大學資訊工程系副教授	學術出版委員會 主任委員
			臺灣大學資訊工程博士	
	曾元顯	男	臺灣師範大學資訊中心研究員	期刊總編輯
			臺灣大學資訊工程博士	
	曾淑娟	女	中央研究院語言所副研究員	
			德國畢勒費德大學語言學博士	

續次頁

承前頁

職稱	姓名	性別	現職及學歷	兼任職務
	廖元甫	男	台北科技大學電腦與通訊所副教授 交通大學電信博士	口語處理組召集人
	鄭秋豫	女	中央研究院語言所研究員 美國布朗大學語言學博士	
	簡立峰	男	臺灣 Google 研究所所長 臺灣大學資訊工程博士	
候補理事	蔡素娟	女	中正大學語言學研究所教授 美國亞歷桑那大學語言學博士	
	張森嘉	男	工研院資通所前瞻技術中心研究副組長 交通大學電子研究所博士	
監事	王小川	男	清華大學電機系教授 美國堪薩斯大學電機工程博士	
	張俊盛	男	清華大學資訊工程系教授 美國紐約大學計算機科學博士	
	陳克健	男	中央研究院資訊所研究員 紐約州立大學水牛城分校計算機博士	
	陳信宏	男	交通大學電信工程系教授 美國德州理工大學電子工程博士	
召集人	陳信希	男	臺灣大學資訊工程系教授 臺灣大學電機工程博士	

附註：

1. 理事、候補理事及監事任期自 98 年 9 月 1 日起至 100 年 9 月 1 日止。
2. 理事及監事排序依姓氏筆畫排序。

## 出席 2009 年幾個國際學術會議感想

### 中央研究院語言學研究所鄭秋豫

今（2009）年的 8 月和 9 月出席了好幾個與語音研究相關的國際學術會議，首先我以 Oriental-COCOSDA (Asian Region of International Committee for Co-ordination and Standardization of Speech Databases) 召集人及 2009 年第十二屆 Oriental-COCOSDA 國際會議 (International Conference on Speech Databases and Assessments, 8 月 10-12 日) 副主席的身份，到北京清華大學出席會議，發表論文，並主持會員大會。也同時出席 AESOP(Asian English Speech cOrpus Project, 8 月 8-9 日) 研討會及順道應邀訪問中國科學院自動化研究所及中國社科院語言學研究所。然後於 8 月 13 日前往甘肅省省會蘭州，出席第十屆人機語音通訊學術會議 NCMMSC (National Conference on Man-Machine Speech Communication) 2009 會議 (8 月 14-16 日)，發表論文。隨即於 9 月前往英國，出席國際口語溝通學會 ISCA (International Speech Communication Association) 的年度會議 Interspeech2009 (9 月 7-10 日)，發表論文。並以 ISCA 下 SIG-CSLP(Special Interest Group-Chinese Spoken Language Processing)現任主席的身份，召開 SIG-CSLP 的會員大會，並參加 ISCA 所有 SIG 的會議。

在北京和蘭州出席的三項會議，原本是配合國際學術會議第十二屆 Oriental-COCOSDA2009 的會期與地點一同安排的 AESOP2009 年度研討會及每二年舉行的人機語音通訊學術會議 NCMMSC2009。AESOP2009 研討會由日本早稻田大學主辦，Oriental-COCOSDA 2009 國際會議由新疆大學信息學院籌組，NCMMSC2009 由新疆師範大學籌組，原訂在新疆的烏魯木齊舉行，未料 7 月 5 日烏魯木齊發生回民暴動事件，7 月 6 日 Oriental-COCOSDA 各國代表及論文作者均紛紛來電子郵件詢問會議是否改期或移地舉行。Oriental-COCOSDA 的中國代表北京清華大學 FIT(Future Internet Technology) 學院副院長鄭方教授也立刻來電詢問應變行動。由於台北與新疆的聯絡因網路與國際電話全部斷絕，北京與烏魯木齊電話可通，於是由鄭方教授擔任聯繫，展開台北、北京與烏魯木齊三地間的討論與協調。此時距離會期僅一個月，烏魯木齊何時完全恢復正常尚無法預期，經多次電話討論多方協調的結果，最後在 7 月 10 日（星期五）定案：會議日期不變，移地舉行。AESOP 與 Oriental-COCOSDA 2009 在北京清華大學舉行，NCMMSC2009 在蘭州西北民族大學舉行。在不到一個月的時間裡，會議的場地、參會者的交通、住宿都需重新安排，而三項會議均能如期在原訂時間順利舉行，主事者的應變及配合，獲得國際友人一致的稱許。

本次 Oriental-COCOSDA2009 會議共收到 37 篇論文，連同主題演講全部經三人匿名審查，共錄取口頭論文 25 篇，此外並請四位主題講員，分別為：日本早稻田大學 Yoshinori Sagisaka 教授，日本 NICT (National Institute of Communication Technology) Satoshi Nakamura 博士，中國科學院自動化研究所陶建華博士與新疆大學 Wushour Silamu 教授。他們也都提供了全文並且也送審。未進入檢索的論文，我們也讓作者們以壁報論文方式發表。這次會議在不到一個月的時間裡更換地點，成功舉行，共有 51 位來自 11 個國家的學者註冊參加會議，中國 17 人、日本 15 人、台灣 7 人、新加坡 2 人，印度、愛爾蘭、

韓國、馬來西亞、菲律賓、泰國、越南各 1 人。台灣共有 4 篇論文。我自 2006 年起擔任國際學術組織 Oriental-COCOSDA(Oriental chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques) <http://mastarpj.nict.go.jp/o-cocosda/>召集人，這一次的 Oriental COCOSDA2009 國際會議，是該組織自 1998 年以來第十二屆年度國際會議。此一組織係亞洲地區區域性國際學術組織，成立於 1998 年，設召集人一人，國家及地區代表二至三人，至 2009 年會議前共有十三個國家及地區參加，按英文字母順序分別為：中國、香港、印度、印尼、日本、韓國、馬來西亞、蒙古、尼泊爾、台灣、泰國、越南、巴基斯坦。該組織不收任何會員費，自 1998 年由日本舉辦第一次國際會議開始，每年舉行國際會議一次，由參加各國自籌經費輪流主辦。會議除安排主題報告及論文發表外，並舉行會員大會，由各國家代表報告一年內收集語音資料庫的情況及相關活動。台灣地區代表為清華大學電機系王小川教授及台大電機系李琳山教授。我因此擔任 Oriental-COCOSDA 2009 大會副主席，每年參與所有籌備、邀請主題報告講員、論文送審、會議安排工作，也因此對此項會議有深刻的瞭解。

Oriental-COCOSDA 會議由 1997 年開始籌組，1998 年以來在會員國間輪流，每年舉辦會議，集結論文成集，今年的最大突破，是會議論文集第一次得以進入國際工程師學會 IEEE Xplore 檢索。這是 Oriental-COCOSDA 國際會議舉行以來第一次進入檢索，對參與會議學者是一大學術品質的提升及鼓勵。會議的另一項重要活動是一年一度的會員大會及國家區域報告，今年由於菲律賓派出代表出席會議，會員國增至 14 國。共收到報告 13 篇，中國、香港、印度、印尼、日本、韓國、馬來西亞、菲律賓、尼泊爾、新加坡、台灣、泰國、越南。不克出席的有香港、印尼、馬來西亞與尼泊爾，由召集人鄭秋豫代為報告。我自 2006 年起，要求所有的報告都需事先繳交，以一頁為限，可以是 6 張 PPT 做成一張壁報，也可以是一頁的 Word 檔案，如此則留下記錄。在會後我也將會議記錄寄送所有的參會人員及代表。負責整個學會運作、確定每年都舉辦國際會議的只有召集人一位，日常業務以校長兼打鐘的方式，全權處理。

Oriental-COCOSDA 的網址設在日本，由 NICT (National Institute of Information and Communication Technology) 支援，這次會議後，我在更新網站時也做了較重大的更動，除了 1997 年以來歷年舉辦會議的記錄和籌組會議的手冊外，我增加了歷年來會員大會的會議記錄，也將 2008、2009 的國家區域報告按照國別列於網站上，方便同行參閱。

配合 Oriental-COCOSDA2009 會期，在會前二天舉行的還有 AESOP (Asian English Speech cOrpus Project)的年度國際研討會。這項研討會的重點有二：(1) 8 月 8 日舉行 AESOP 跨國研究計畫決策委員會(Steering Committee)會議，(2) 8 月 9 日舉行的 2009 年 AESOP 年度國際會議。AESOP 跨國研究計畫決策委員會(Steering Committee)會議由 AESOP 發起人日本早稻田大學 Yoshinori Sagisaka 教授主持，委員除 Yoshinori Sagisaka 教授外，還包括早稻田大學 Mariko Kondo 教授、香港中文大學 Helen Meng 教授(由 Dr. Wai-Kit Lo 代表)、印尼的 Dr. Sakriani Sakti 及台灣中研院的鄭秋豫及銘傳大學的助理教授 Tanya Visceglia。主旨是討論並決定下一年收集語料的方向、原則、內容，並決定除原有成員外，邀請新成員參加。本計畫原有成員為：日本、台灣、香港、泰國、印尼、

蒙古。擬邀請成員為：中國、韓國、越南、菲律賓、印度。第二天（8月9日）舉行2009年AESOP年度國際會議。由我開始以TWNAESOP國際合作計畫主持人身份報告計畫內容。參加人數約三十五人，原有成員即擬邀請成員全部出席。會議由上午9點至下午6時。上午由原有成員報告各國與AESOP相關的研究計畫，其他報告包括香港中文大學、泰國NECTEC及日本早稻田大學的相關計畫，並進行討論。下午則由AESOP原有成員及擬邀請或參加成員報告各國與L2英語相關的研究近況、進度及規劃，共計有11份口頭報告，所有受邀的國家均同意參與。這項合作將以收集開發英語為第二外語（L2）的語音資料庫為目標，所有收到的語音資料庫都將包含共同的核心語料，並以公開資源的方式，無償提供語音學界使用。AESOP的目的請參看收入Interspeech2009, (Sep. 6-10, 2009), Brighton, U.K.會議論文集的Meng, H., Tseng, C., Kondo, M., Harrison, A. and Visceglia, T. “Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper.” pp. 1715-1718。核心語料的設計原則及內容請參看收入Oriental COCOSDA 2009論文集的Visceglia, T, Tseng, C., Kondo, M., Meng, H. and Sagisaka, Y., “Phonetic Aspects of Content Design in AESOP (Asian English Speech cOrpus Project).”及收入The 10th National Conference on Man-Machine Speech Communication (NCMMSC 2009, 第十屆全國人機語音通訊會議)論文集的Visceglia, T and Tseng, C. 的“Spontaneous Speech Database Design for Suprasegmental Characteristics in Asian L2 English.” pp. 549-554。以上活動顯示，如果積極參與國際學術合作計畫，是一個提升面對更大挑戰、研究水準的正向方式。

不過，更大的震撼是8月14-16日在蘭州舉行的第十屆人機語音通訊學術會議NCMMSC2009會議。我因為擔任這個會議的常設機構委員，多年來總是盡量提交論文出席會議，很深切的體會到這個二年一度會議的水準越來越高。2007年我曾出席第9屆會議，當時就覺得整體水準很高，會後由北京清華學報（EI檢索）的專刊審稿與編輯作業都是國際水準，這次會中也得知該專刊的引用率為這兩年清華學報之最高。本次2009年的會議共發表103篇論文，全部雙盲審查。錄取論文中將選出30篇論文（不及30%）由清華學報出版專刊，不但將列入EI，也將列入中國的國家檢索，更具體的提升了論文的水準。本次會議共有四場主題演講，講員為：美國國家科學院院士、中央研究院院士美國喬治亞理工大學莊炳煌教授，中國科學院院士張鈞教授，日本JAIST(Japan Advanced Institute of Science and Technology)Masato Akagi教授，日本JAIST黨建武教授，均為一時之選。由於論文篇數甚多，會議採平行場次，議題包括TTS, Emotion, ASR, Phonetics, Audio, VPR & LID, NLP&SD, APP、Corpora及其他。此外有四場國際性特別議題討論，邀請國際學者集稿組織，論文全為英文，口頭報告也全以英語舉行，分別為由國立新加坡大學的李海洲、馬賓教授組織的Asian Speech Resource，日本JAIST教授Masato Akagi教授組織的Speech and Language Processing，新疆師範大學組織的Under Resourced and Minority Speech Processing，及日本東京大學Nobuaki Minematsu教授組織的Infant Language Development。這四場國際性的特別討論吸引了中國以外的學者投稿、參與，把會議的國際性大幅提昇，也提供了當地研究生的參與，是非常可借鏡的方式。此外，本次會議設研究生最佳論文獎也非常精彩，在會前論文經審查內容評比後獲選的論文，必須再由評審委員至發表演場評比口頭報告的內容與應對，最後才投票選出獲獎者。將現場報告的準備與應對也納入評比，是鼓勵研究生加強表達能力最直接的途徑，也非常

值得借鏡。

由於烏魯木齊事件而臨時移師到蘭州的會議，不但如期舉行，臨時被徵召的主辦單位西北民族大學，也展現了令人欽佩的應變能力，將會議辦理的有聲有色。我們應邀參觀了西北民族大學的信息工程學院，發現他們多年來因為地緣，積極的開發了少數民族的文字輸入系統，最有成的是藏文輸入，為西藏語言的電子典藏開創了先例。

隨後於9月初在英國舉行的 ISCA 年會 Interspeech2009 (2009.9.7-11)，也有許多發人深省之處。這是一個大型會議，除主會議的會期5天外，一向有各種以研究課題為主旨的課程性研討 (tutorial)，也有各種衛星會議。本次會議主旨為 Speech and Intelligence，會期5天，包括12個次領域，會前一日(2009.9.6)有六場 tutorial，還有六個衛星工作坊同時舉行。大會共收到論文1,302篇，由645位同行學者審查人，每篇論文均經三位審查人雙盲審查，共錄取762篇，錄取率為58%。大會期間共有39場平行口頭報告論文及30場壁報論文。參會人數及全球分佈。註冊參加會議的有1,164人，來自全球49個國家，50%參會者來自歐洲，20%來自美國，20%來自日本。參會人數最多的前五個國家依序為美國、日本、英國、德國、法國。我們若以日本與美國的總體研究人口相比，日本絕對不及美國，但日本可以在這樣一個大型國際會議中與美國各佔出席人數的20%，表示日本在全球學術的總體影響力不輸美國，這需要投注多少資源、提出多少配套措施、經過多少年的經營才能達到。明年的 Interspeech2010 會議將由日本主辦，我們這次在會場裡也看到日本學者不分資格，輪流在攤位前值班，致送海報集紀念品，邀請大夥參加明年的會議。足見日本學術界採用了各種積極的方式，在世界舞台上扮演舉足輕重的角色，這一點絕對值得我們借鏡。此外，多年來日本也積極培養國際學生，他們訓練出來的博士，離開日本後在全世界都可以找到工作，成為重量級學者。隨便舉個例，多年來在美國及瑞士任職，研究語音識別有成的捷克裔 Hynek Hermansky 教授多年來在美國、瑞士工作，研究成績斐然，他的博士學位獲自日本東京大學。東京工業大學、京都大學在這方面的成績也不遑多讓。

特別發人深省的還有二點：一為主題報告的人選，二為頒發獎項的啟發。主題報告方面，本次會議共邀請4位主題講員，其中2為資深學者：一為本次 ISCA Medalist 日本東京工業大學教授 Sadaoki Furui，講題為“Selected Topics from 40 Years of Research on Speech and Speaker Recognition”。一為美國華盛頓大學教授 Mari Ostendorf 講題為“Transcribing Human Directed Speech for Spoken Language Processing”。另2位較為資淺，一位是 MIT 副教授 Deb Roy (1999年獲 PhD)，講題為“New Horizons in the Study of Child Language Acquisition”。另一位是 UC Berkeley 助理教授 Thomas Griffiths (2005年獲 PhD)，講題為“Connecting Human and Machine Learning via Probabilistic Models of Cognition”。這樣的安排非常有趣，資淺的學者因為累積有限，提出的看法狹隘許多，爭議性也較大，但會議的用心很清楚，有創新跟有結果都值得同行的特別關注，這一點也是西方學術界較具包容性的實例。

頒發獎項方面，本次會議共頒發：(1) ISCA Medalist，得獎人為日本東京工業大學教授 Sadaoki Furui。(2) ISCA Fellow 共有六位學者獲獎。(3) 最佳論文獎，共有6篇論文獲獎。最引人深思的是 ISCA Medalist，這是 ISCA 學會頒發的最高榮譽，每年只有

一位得獎人。去年(2008)的得獎人 Hiroya Fujisaki 教授與這次得獎的 Furui 教授都是日本本土培植出的學者，他們個人的學術成就與國際影響力由此不需贅言，日本學術界的整體國際影響力，也由此可見一斑。Furui 教授現年 64 歲，曾任 ISCA 理事長，已發表的論文達八百餘篇。這八百餘篇論文代表不但的是他三十餘年來帶領團隊的規模、訓練出來的研究人才、提出的研究課題與解答與建立的語音系統，一方面代表個別學者在國際學術界的影響力；另一方面提供一個讓我們重新思考獨立研究能力、團隊研究的力量及學術影響力等問題的新角度，及兼容並蓄面對學術的態度。

最後談一談 ISCA 學會提升水準的具體措施。我是現任的 SIG-CSLP 主席，在會期間先召開了 SIG-CSLP 會議，又參加了 ISCA 所有 SIG 的會議。這次 ISCA 理事選舉，台大電機系的李琳山教授是國內唯一當選 ISCA 理事的學者，他在二期八年的任期中，積極的為 ISCA 開展了全球性的學者與研究生的聯繫網絡，在這次會議中卸任。新任的理事共有 9 位，分別是 Jean-Francois BONASTRE (法國)、Nick CAMPBELL (愛爾蘭)、Keikichi HIROSE (日本)、David HOUSE (瑞典)、Haizhou LI (新加坡)、Douglas O'SHAUGHNESSY (加拿大)、Michael PICHENY (美國)、Yannis STYLIANOU (希臘) 及 Isabel TRANCOSO (葡萄牙)，加上任期未滿的 7 位理事 Jean-Francois BONASTRE (法國)、Bernd MOBIUS (德國)、Allan BLACK (美國)、Eva HAJICOVA(捷克)、Helen MENG(香港)、Yoshinori Sagisaka (日本)、Tanya Schultz (美國)，一共有 16 位理事。值得注意的是在 16 位理事中，華人理事佔了二席，日本學者也佔了二席。若以華人地區與日本總研究人口來看，根本不成比例，而日本學者的整體優勢不言而喻。除了研究成績優異外，他們努力的目標，並非僅著眼於獲得個人學術肯定與榮譽，而是積極參與國際社群活動、為這些服務，因此日本學界的全球影響力也是我們應該參考的。日本學者也積極帶領跨國研究團隊，在亞洲地區就語音科技開發方面，便有 NICT 資深學者帶領的機器翻譯團隊 A-STAR (Asian Speech Translation Advanced Research Consortium) 及 Yoshinori Sagisaka 帶領的跨國 L2 亞洲英語資料庫 AESOP 研究團隊，反觀我們國內，致力學術研究的同仁不少，參與國際學術社群活動的卻有限。這與日本學者無論會議規模大小，均可見積極參與，並以優異的研究成績，積極的在國際社群耕耘。

在 ISCA 的會員大會及 SIG 會議中，國際主席葡萄牙學者 Isabel Trancoso 教授都提出了該學會對提升論文水準的配套規劃重點，就是加強論文進入檢索。具體措施包括：(1) 編輯出版 Speech Communication 的 special issue。ISCA 學會的期刊是 Speech Communication，一年四期，每期發表論文六至十篇，早已進入 SCI 檢索。現在鼓勵每二年定期舉辦較小型國際會議的 SIG 如 SpeePro (Speech Prosody)、CSLP (Chinese Spoken Language Processing) 在舉行會議後，向 Speech Communication 提出特刊 (special issue) 的具體規劃，擔任客座編輯，推廣 SIG 活動。(2) 將 Interspeech 論文集 Proceedings 推進現有的檢索系統。Interspeech 是 ISCA 的年度會議，一向是全文投稿並出版會議論文集，主席宣布 Interspeech2007 及 2008 的論文集都已進入 IEEE Xplore 檢索，現有鑑於審稿作業日趨完善、論文錄取率逐年下降，將所有 Interspeech 及 SIG 會議的論文集以包裹方式進入檢索指日可待，這對提升 ISCA 的國際學術水準將更有大幅的協助。

而與國內語音學界最相關的 SIG-CSLP 會務與 ISCSLP2010 (International Symposium

on Chinese Language Processing) 籌畫、會議論文集進入檢索及出版一期 *Speech Communication* 的特刊。ISCSLP2010 是每二年舉辦一次的中文口語分析盛會，也是 SIG-CSLP 最重要的學術活動，在台灣、香港、中國、新加坡四地輪流舉行，2010 年是第二次輪回台灣，由成功大學主辦，王駿發教授擔任大會主席，我是現任 SIG-CSLP 主席，因此應邀擔任大會副主席，並在 Brighton 召開國際組織會議，決定會議時間為 2010 年 12 月初，會議地點為成大與日月潭，所有論文循 Interspeech 體例，都是全文論文，四頁八點字，三人雙盲審查，根據過去的會議經驗，預計錄取 120 篇左右論文。ISCSLP2008 會議論文集已進入 IEEE Xplore 檢索，ISCSLP2010 不但將維持此例建立傳統，並與 ISCA 取得協議，會後將向 *Speech Communication* 期刊提出規劃案，自錄取的 120 篇論文中，由作者改寫為期刊論文投稿，選擇六至十篇，編纂一期特刊。

在不到兩個月的時間裡，我一共出席了四項會議。這些會議不少共同特點，發人省思。一是參與國際學術不僅是提交優異的論文出席國際會議，而是要積極的參與國際學術社團的服務工作，爭取主辦國際學術會議，參與國際合作研究計畫，培養優秀的研究生，才能展現參與能力及影響力。二是國際學術團體如何提出提升整體學術水準的具體措施，無論是 Oriental-COCOSDA 從 2009 年起會議論文集進入 IEEE Xplore 檢索，或是 ISCA 學會的具體規劃，都值得國內參考。請看 ISCA 較 IEEE 年輕許多，但在歷屆理事長的努力下，除發行有相當水準的期刊 *Speech Communication* 外，也累積有足夠的學術能量連同 Interspeech 會議論文集與各 SIG 的會議，拉提學會內的 SIG，以編輯 *Speech Communication* 特刊的方式，向更大的學術社群爭取優良的論文，並以包裹方式向各種學術資料庫敲門。三是由 ISCA 理事會 16 位理事的地理區域分佈，可以見到 ISCA 的全球性及其國際影響力，也更見日本學術界的整體競爭力與影響力，值得我們效法。我希望國內語音學界的同仁與研究生除了積極的從中文語音研究出發，加強研究成績，更以日本學者為典範，積極參與並爭取亞洲區域性與全球國際性國際學術團體的公職，立足台灣放眼世界，與更廣大的學術社群在第一線互動。最後，由於我們研究的是中文口語分析，如何提升這個學術社群在全球的整體影響力，當然不能自外於與大陸學術界，因此像人機語音通訊學術會議 NCMMSC (National Conference on Man-Machine Speech Communication) 這樣的學術會議，應直接面對，提交高品質的論文積極參與，需要更多的規劃。對於彼岸學術研究的提升，也應深思我們未來的方向與發展。

# 語音辨識中於語音特徵時間序列域與調變頻譜域之 強健性處理技術之介紹

## Introduction to the approaches in improving the robustness of speech features in speech recognition

洪志偉 Jaih-weih Hung

國立暨南國際大學電機系

Dept of Electrical Engineering, National Chi Nan University

Taiwan, Republic of China

jwhung@ncnu.edu.tw

### 摘要

本專文主要內容是介紹在語音辨識中，歷年來許多擷取語音特徵隨時間變化特性之訊息之技術，這些技術之主要目的在於使處理後的新語音特徵，能夠更具強健性（使語音特徵能較不受到外在與語音資訊無關之干擾），進而提昇雜訊環境下語音辨識的精確度，使語音辨識能更具應用之價值。文中我們將簡要提及歷年來學者以各種不同的角度，分析語音特徵隨時間變化的特性，進而藉由這些觀察到的特性來強化語音本身的訊息，降低或消弭非語音之干擾的效應。我們也將提及這些根據不同角度發展的技術，多數彼此之間存在著良好的加成性，意即當兩種（甚至多種）技術同時使用時，相對於單一技術而言，對語音特徵的精確度會有更佳的提昇表現。

### 一、緒論

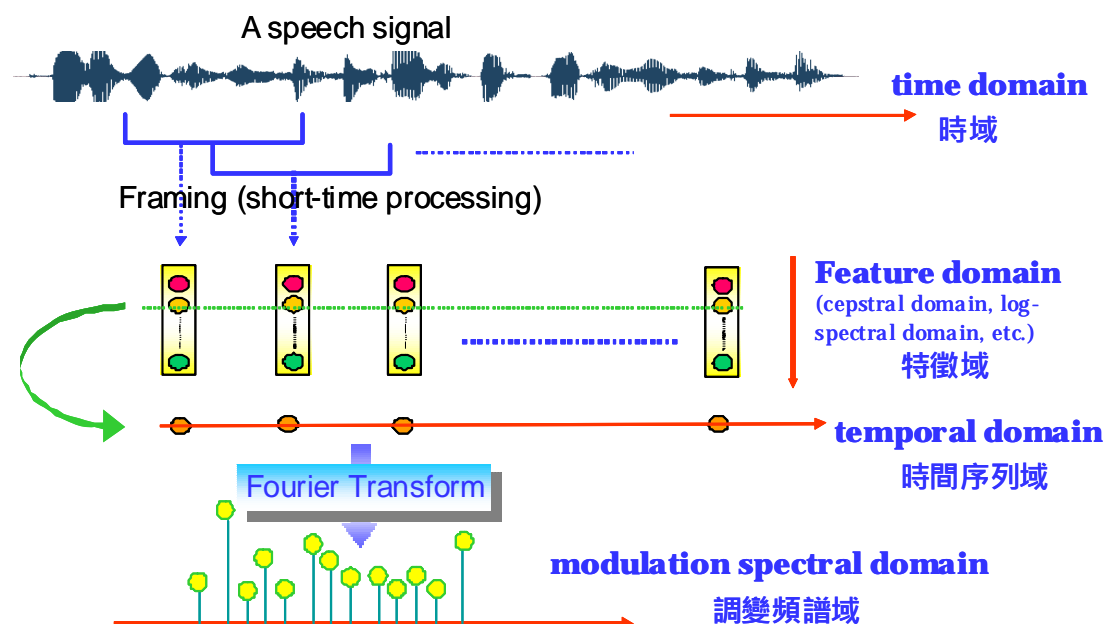
近年來，語音處理之領域的學者持續地開發研究，使語音處理相關理論與技術不斷精進成熟，逐漸趨於實際應用的目的，雖然語音處理技術逐步成熟至應用性的階段，然而仍存有許多需要突破的問題與開發的技術，就語音辨識(speech recognition)而言，一方面，我們希望能找尋更有鑑別力(discriminating capability)的語音特徵表示式或語音模型訓練技術，使不同聲學單位之語音其特徵或聲學模型彼此的相異度能夠更大，進而降低辨識時不同聲學單位的混淆度。另一方面，由於一語音辨識系統常因訓練環境與應用環境上的諸多變異性(如訓練環境與應用環境的不同語者差異、同一語者之發音方式差異、與語音無關之周遭雜訊干擾、語音傳遞之通道差異等)，而使辨識效能受到明顯影響，而針對以上所述的變異性所發展之各種演算法，通常統稱為強健性(robustness)技術。

以上我們所論及的提昇語音辨識之強健性技術，一直是語音處理領域中熱門的研究主題，由於語音訊號本身資訊十分豐富與多樣化，許多語音相關領域之學者從語音訊號不同的特性切入，而研發出各種五花八門的提昇語音辨識強健性技術。在本篇專文裡，我們將單一介紹，從語音特徵的『時間序列域』(temporal domain)與其相對的『調變頻譜

域』(modulation frequency domain)之角度加以切入，所研發出的各類強健性方法。

在逐一介紹這些技術之前，我們先利用圖一，簡要說明何謂語音特徵之時間序列與調變頻譜，如圖一所示，在語音辨識時，由於語音訊號本身是一時變(time-varying, non-stationary)的訊號（意即其統計特性隨著時間而改變更新），一般的處理方式皆是把它作分段式短時間的切割，將其分解成一段一段（通常前後部分重疊）的短時間語音訊號，然後個別針對每一小段的語音訊號作處理或轉換，得到其對應的語音特徵向量，圖一上半部顯示了語音訊號從時域(time domain)轉換至特徵域(feature domain)的程序，語音特徵向量本身可能是此小段語音訊號的頻譜(spectrum)、對數頻譜(logarithmic spectrum)、線性預測係數(linear prediction coefficients, LPC)或目前最廣為使用的倒頻譜(cepstrum)特徵等。

由於每個特徵向量是由不同的短時間訊號所擷取出來的，若依照原對應時間的先後排列，我們就相當於可以在時間序列域(temporal domain)中觀察這些語音特徵向量值隨著時間而產生的變化。一般而言，許多在時間序列域所發展的特徵改進技術，是針對特徵向量的每一維度個別去作處理的，這類似於在處理二維空間的訊號時，特意簡化成處理其各個一維空間的訊號一般。雖然這樣的處理方式可能忽略了特徵向量各維度間彼此的相關性(correlation)，但是由於在分析個別之一維特徵序列的特性上較為簡易，因此其也廣為學者使用。



圖一、語音辨識中，語音在各種不同領域(domain)之呈現示意圖

前面提到，我們希望去分析探討語音特徵隨著時間而產生的變化，依照一般訊號處理的方式，將一訊號利用傅立葉(Fourier Analysis)時，就可以藉由其所得到的頻譜觀測出此訊號中所包含變化快慢（即頻率）成份的量值，循此觀念，我們若將一連串具有時間

先後的語音特徵（即後面所提的語音特徵時間序列）利用傅立葉轉換，得到其對應的頻譜，即是如圖一所顯示的，把特徵序列從時間序列域轉換成調變頻譜域(modulation spectrum domain)。爲了更釐清語音之頻譜成份與調變頻譜域之成份之不同，我們用一個簡單的例子來說明：若一語音訊號其取樣頻率爲 8 kHz，其經上述之分段程序與短時間傅立葉轉換後，其頻譜的頻率範圍爲 0 至 4 kHz，此頻率的高低意謂了聲音本身的高昂或低沈，例如女生聲音較尖銳高昂，其對應的高頻率成份相對於男生而言就會比較多。相對地，假設上述分段程序中，是每隔 10 ms 取一段，進而求取這段語音訊號的特徵，此代表特徵的取樣頻率爲  $1/10\text{ms} = 100 \text{ Hz}$ ，我們將一連串之特徵時間序列取傅立葉轉換得到其調變頻譜時，此頻譜的範圍即爲 0 至 50 Hz，此頻率的高低是隱含了聲音特性變化的快慢（如人說話速度的快慢），而非前面所提到的聲音音調的高低。如一個女生特意持續尖叫時，在一般的頻域上，其對應的頻率可能很高（因爲聲音高昂），但在調變頻譜域上，其頻率可能趨近於 0（因爲聲音幾乎都沒有變化）。

在時間序列域或調變頻譜域上分析語音特性的國際學者中，Hynek Hermansky 教授及其研究團隊應算是最廣爲人知，且相關研究成果卓越，首先，在其文獻[1]中，提到了人在聽音時是著重於聽語音相對變化的成份，藉此辨識語音的內容，而一般人發音時，其語音變化的速度（嚴格來說，是音節單位變化的速度）不會太快或太慢，而人耳在聽音時，對於某語音變化速度（即調變頻率）範圍的敏感度相對其他頻率而言相對較高，根據這些觀察，即可以在分析一段訊號時，將一些認定並不太可能是語音變化速度的成份（過快或過慢）過濾掉，藉此降低或消弭非語音的干擾成份，並將認定爲人耳較敏感的語音調變頻率成份將以強調，藉此強化語音本身成份，使後續的語音辨識精準度能提高。

在文獻[2]中，作者藉由語音辨識實驗，說明了無論是對人耳辨識或機器辨識而言，各個語音調變頻率成份都有相對不同的重要性（此實驗是主要藉由一可調之帶通濾波器，過濾出語音訊號特定範圍的調變頻率成份來作辨識，藉由辨識精確度高低來衡量各調變頻率成份的重要性），文獻提及 1 Hz 至 16 Hz 的調變頻譜範圍，對語音辨識格外重要，即便過濾掉此範圍以外的調變頻率成份，對語音辨識結果也不會有明顯影響。而在 1 Hz 至 16 Hz 的範圍中，大致又以 2 Hz 至 8 Hz 範圍內的頻率成份對語音辨識有關鍵性的影響，若刻意過濾掉此範圍之頻率成份，將嚴重影響語音辨識的精確度。以上這些早期文獻的試驗結果，都在其他許多相關或後續的特徵時間序列處理之文獻中直接或間接被證實。在下一章中，我們將介紹一些有名的特徵時間序列處理演算法，這些演算法都有助於增加原始語音特徵的強健性。

## 二、語音特徵時間序列之處理演算法相關介紹

在本章中，我們將簡要介紹一系列語音特徵時間序列相關處理技術，大致上，我們根據其不同的發展背景加以分類，然而，不同類別之間的技術並非是彼此完全無關的，

有些技術可能同時屬於兩個類別以上，在此做技術上的分類是希望能較清楚有次序地介紹各類技術。

## 1. 特徵時間序列濾波器設計：

這一類的方法主要是為特徵時間序列設計一濾波器，目的在使濾波器處理後的新特徵能更具有鑑別力或強健性。最早期的濾波器設計，或許應該首推特徵序列的一階差量與二階差量的求取[3]，差量的求取在於取得一段特徵時間序列裡，相鄰特徵之間的變化程度，其基本原理即是我們所熟悉的線性回歸(linear regression)，而一階差量與二階差量的求取，本身即是運用了帶通濾波器(band-pass filter)，大部分的語音辨識實驗驗證了，在原始特徵附加上其一階差量與二階差量特徵後，語音辨識精確度皆有明顯的提升，此意謂了差量特徵的確包含了語音辨識鑑別的資訊。

其次，根據上一章的介紹，學者發現在語音資訊主要包含於相對低頻（調變頻率）的成分，因此設計出了著名的相對頻譜(Relative Spectral, RASTA)法[1]，其對應的濾波器是上述之一階差量濾波器串連一個低通濾波器而得，其更著重於將屬於語音成分的調變頻率成分過濾出來，並且降低非語音的成分。此方法目前亦廣為大家所使用。

由於原始 RASTA 濾波器之設計只是根據一個較小型的語音特徵資料庫所設計出的『固定型式』濾波器，其並不保證在不同的語音特徵下皆能有最佳的特性，因此後來有學者陸續提出了『資料導向』(data-driven)的概念，即根據特定的語音特徵或辨識環境，設計其專屬的時間序列濾波器[4]，這些濾波器的求取常是伴隨著某個最佳化準則，例如文獻[5,6]提及了使用三種不同最佳化準則：主軸成分分析、線性鑑別分析與最小分類誤差等，而文獻[7]則使用了著名的韋納濾波器(Wiener filter)設計準則。另外，這些資料導向的濾波器設計，所用的特徵資料，有時是語音特徵時間序列本身[4,5]，有時則是時間序列經傅立葉轉換而得的調變頻譜[6]或功率頻譜密度[7]，這些濾波器根據其不同的最佳化準則，獲得的形式也各自都不同，對於增強語音特徵的鑑別力或強健性（特別著重於強健性）也各稍有程度上的差別，但有趣的是，這些濾波器其頻率響應(frequency response)其主要的通帶(passband)仍偏於中低頻，大概都在 1 Hz 至 20 Hz 之間，此現象也呼應了我們前面開始所提的，語音成分主要都是偏重在中低頻的調變頻率，強調這一部份的頻譜有助於提升語音辨識的效能。

## 2. 特徵統計值正規化技術：

這一類的方法之發展背景，主要是語音特徵在受到雜訊干擾或通道效應後，其值會產生明顯的偏差，若將語音特徵視為一隨機變數(random variable)，則意謂了此隨機變數因為干擾，導致其諸多統計量（例如平均值、變異數、中央動差、動態範圍或機率分佈函數等）皆產生偏差。因此這一類的方法，主要程序是同時對訓練聲學模型的語音特徵與辨識的語音特徵做處理（正規化），使處理後之兩方特徵的統計量能趨於一致，進而降低因為干擾所造成雙方不匹配的現象。

如前所述，此時語音特徵被視為一隨機變數，則在處理其需考量其統計量的估測方法。一般而言，我們假設此隨機變數具有遍歷性(ergodicity)，簡言之就是其統計量可以藉由特徵在時間上的平均趨近而得。換句話說，一特徵時間序列的每點特徵都視為此隨機變數的樣本(sample)，藉由此樣本集合來估測隨機變數的統計值。以下，我們簡單介紹幾種著名的特徵統計值正規化技術：

(1) 平均值消去法(mean subtraction, MS)[8]：

此技術在於使訓練與辨識雙方的特徵其平均值趨於一致。此法最早是運用於倒頻譜(cepstral coefficients)上，且其正規化後的平均值設為 0，其原始假設是在語料平衡的前提下，乾淨語音之倒頻譜的平均值會趨近於 0，因此若發現倒頻譜本身平均值不為 0，則將此平均值扣除即可。此方法原先主要是運用於降低通道效應(channel effect)的干擾上，但許多語音辨識的實驗也顯示了，其對於雜訊干擾的降低也有一定的效用。

(2) 平均值與變異數正規化法(mean and variance normalization, MVN)[9]：

如其名字所述，此技術在於使訓練與辨識雙方的特徵其平均值與變異數皆趨於一致。相對於前面的 MS 法而言，MVN 額外正規化了變異數，因此更能降低訓練與辨識雙方之特徵的不匹配，一般實驗也都顯示了，MVN 效果比 MS 來的更好。

(3) 統計圖等化法(histogram equalization, HEQ)[10]：

此技術在於使訓練與辨識雙方之特徵的統計圖(相當於估測的機率分佈)趨於一致，使訓練與辨識雙方的特徵處理過後，在統計特性上幾乎有最大的匹配度，這是因為兩隨機變數之機率分佈一致時，其所有統計值(如 MS 所用的平均值，MVN 所用的平均值與變異數，還有每一階動差等)都會相同，因此從統計的角度上，HEQ 效用明顯優於 MS 與 MVN。

(4) 高階動差正規化法(HOMN)[11]：

此技術相對於 HEQ 而言，並非針對所有階層的動差，而是挑選部分較高階層的動差加以正規化，其主要可能的理由在於，當 HEQ 正規化所有階層的動差時，相對也可能降低了不同聲學特徵彼此間的差異性(此關係到特徵本身的鑑別力)，另一方面，雜訊干擾對於高階動差的影響相對較大，因此 HOMN 選擇性地正規化少數的高階動差，且在文獻[11]中顯示了，HOMN 所帶來的效果確實優於 HEQ，足以支持前述的論點。

以上所介紹之各種語音特徵統計正規化法，傳統上是直接執行於語音特徵的時間序列上，且其統計量的求得通常是利用整段語音之特徵，然而近年來有許多研究，根據類似的觀念，而有不同的執行方式，卻也同樣獲得良好甚至更加的效果，例如：

- (1) 文獻[12]首先提出，在語音特徵序列之調變頻譜(強度頻譜)上執行 HEQ 法，發現其對於語音特徵的強健性也有明顯的改善，而文獻[13]提及在調變頻譜上執行 MS 與 MVN 法，同樣具有明顯的效果，而這些處理在調變頻譜上的技術若與原先處理於特徵時間序列的技術相結合，也發現能有良好的加成性。

- (2) 文獻[14,15]提及，利用事先訓練的碼簿來估測語音特徵的統計值，藉此取代原先利用整段語音特徵的估測方式，此碼簿式估測法在 MS 與 MVN 技術上比原先整段式估測法表現更佳，且具有線上運算(online implementation)的優點；文獻[16]則更進一步提及，適當地組合碼簿與整段特徵(或組合碼簿與片段特徵)的資訊來估測語音特徵的統計量，無論在上述之 MS、MVN、HEQ 與 HOMN 技術上都明顯優於原先碼簿式或整句式的估測法。

### 3. 語音特徵序列濾波法與特徵統計正規化之組合

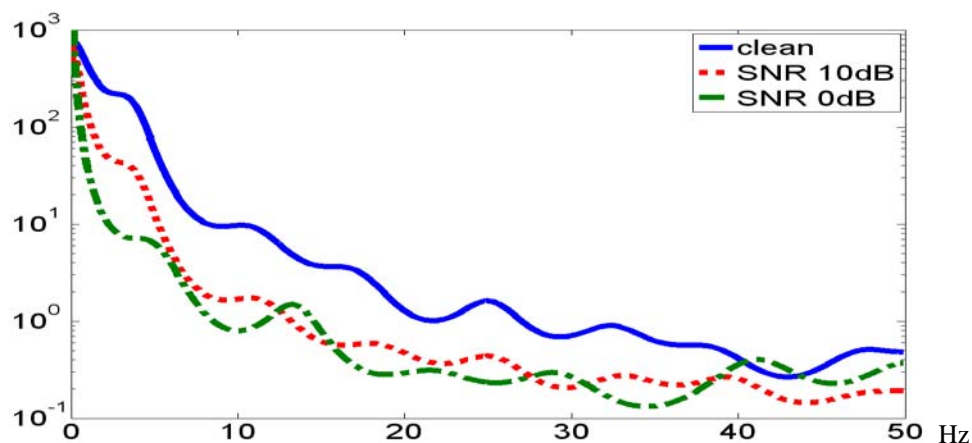
前述之第二類的特徵統計正規化法，基礎目的是拉近訓練與應用兩環境之特徵的統計特性，並未如第一類的濾波處理技術，考慮到語音特徵時間序列之不同調變頻帶對語音辨識之不同的重要性，因此近來許多文獻所發展的技术，同時考慮到以上這兩類方法的觀點。以下我們簡要介紹幾個此類的方法：

- (1) 文獻[17]所提出之平均與變異數正規化組合 ARMA 濾波器法(mean and variance normalization plus ARMA filtering, MVA)，是將一低通濾波器作用於 MVN 處理後的特徵時間序列上，藉此進一步凸顯位於中低頻之重要的語音成分，而得到十分優異的強健性效果。
- (2) 文獻[12]所提及之調變頻譜上的 HEQ 法，當其適當地調整低頻成分與高頻成分的比值時，可以得到更佳的效能。
- (3) 文獻[18]所使用的分頻式 MVN 與 HEQ 法，是使用離散小波轉換(discrete wavelet transform)，首先將語音特徵序列分成數個子調變頻帶的特徵序列，分別對每個子頻帶序列做 MVN 或 HEQ 處理後，在使用反離散小波轉換將子頻帶特徵合成為原全頻帶特徵，藉此對不同的調變頻率成分分別處理，相對原先的全頻帶式的處理方式，可以得到更具強健性的特徵。
- (4) 文獻[7]中所提出的時間序列結構正規化法(temporal structure normalization, TSN)，是先將語音特徵序列經過 MVN 法處理後，再設計其對應的韋納濾波器(Wiener filter)，使所得之新特徵序列其功率頻譜之形狀逼近一預設之功率頻譜，此處理相對於單一使用 MVN 法與單一使用韋納濾波器法處理，都能使新特徵得到更好的辨識精確度。

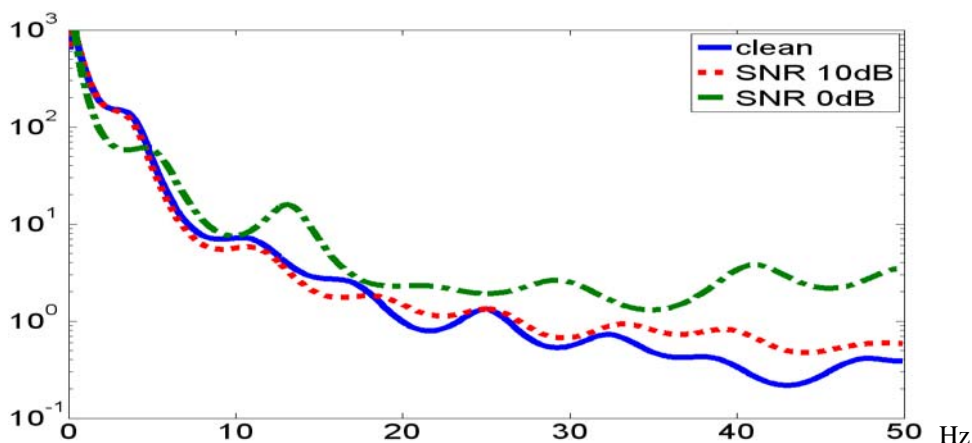
## 三、語音特徵時間序列之處理演算法效能之簡要展示

為了較清楚地顯示這裡所提及之許多特徵時間序列處理法的雜訊強健性效果，在此簡要地使用一系列的功率頻譜密度(power spectral density, PSD)圖來加以說明，我們利用了 AURORA-2 資料庫[19]裡 MIP\_28826Z4A 語音檔，加入不同訊雜比(signal-to-noise ratio, SNR)的人聲(babble)雜訊(0dB 與 10dB)，再經各種技術加以處理，最後求取其功率頻譜

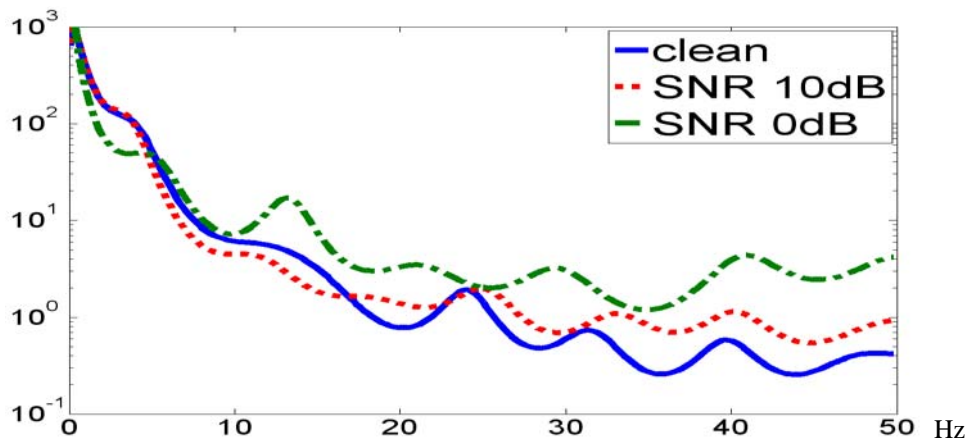
密度。圖二為原始未處理之功率頻譜密度曲線圖，從此圖我們可以看出，雜訊的存在使乾淨語音與雜訊語音產生明顯的 PSD 失真，此失真極可能即為影響語音辨識精確度的主要因素之一。圖三、圖四與圖五則分別為時間序列之 MVN 法、HEQ 法與調變頻譜之 HEQ 法處理後之第一維倒頻譜特徵  $c_1$  之功率頻譜密度曲線圖。將這三個圖與圖二相比較，我們可以看出，PSD 失真現象明顯改善許多，尤其是在中低頻（約 20 Hz 以下）的區域。藉由降低 PSD 之失真，語音辨識的精確度也因此明顯上升。



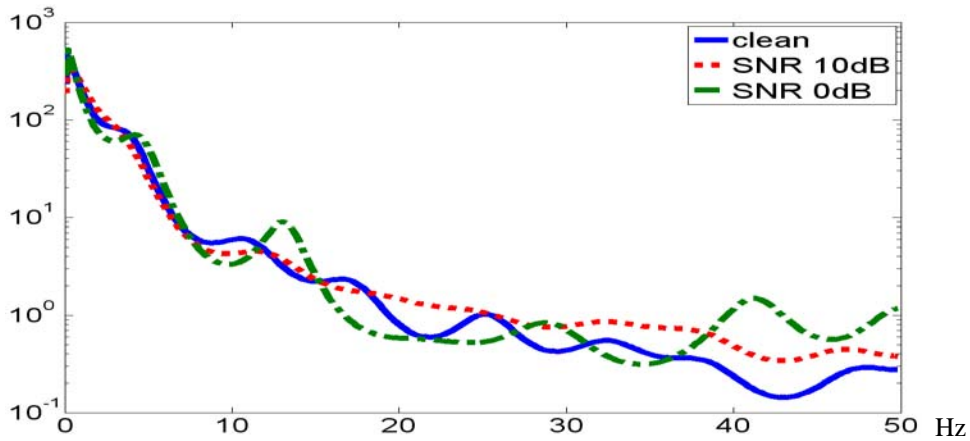
圖二、在人聲雜訊干擾下，未處理之第一維倒頻譜序列之 PSD 圖



圖三、在人聲雜訊干擾下，時間序列 MVN 法處理後之第一維倒頻譜序列之 PSD 圖



圖四、在人聲雜訊干擾下，時間序列 HEQ 法處理後之第一維倒頻譜序列之 PSD 圖



圖五、在人聲雜訊干擾下，調變頻譜 HEQ 法處理後之第一維倒頻譜序列之 PSD 圖

#### 四、結論

在此專文中，我們簡要介紹了幾類作用於語音特徵時間序列或調變頻譜上的強健性技術，概括說來，相較於其他許多強健性技術而言，這些技術中的優點在於其執行複雜度相對較低，但卻又能明顯提升語音特徵之雜訊強健性，同時，這些技術大部分彼此之間可以互相加成，或是與其他類型的強健性技術整合，而得到更理想的強健性效能。在過去文獻中，這些技術多數是運用於處理加成性雜訊環境或較簡單的通道干擾環境中，我們期待未來這些技術能進一步拓展延伸，使其能成功運用於其他種類的干擾環境，例如回音雜訊環境、無線通訊環境或水下通訊環境中，進而使語音辨認與處理的技術具有更廣泛的實用價值。

## 五、參考文獻

- [1] H. Hermansky and N. Morgan, "RASTA processing of speech," IEEE Trans. on Speech and Audio Processing, 1994
- [2] N. Kanedera, T. Arai, H. Hermansky and M. Pavel, "On the importance of various modulation frequencies for speech recognition", Proceedings of Eurospeech, 1997
- [3] S. Furui, "Speaker-independent isolated word Recognition based on emphasized spectral dynamics", Proceedings of ICASSP 1986, pp. 1991-1994, April, 1986
- [4] S. van Vuuren and H. Hermansky, "Data-driven design of RASTA-like filters", Proceedings of EUROSPEECH 1997, pp. 409-412, 1997
- [5] J-W. Hung and L-S. Lee, "Optimization of temporal filters for constructing robust features in speech recognition", IEEE Trans. on Audio, Speech and Language Processing, 2006
- [6] J-W. Hung and W-Y. Tsai, "Constructing modulation frequency domain based features for robust speech recognition", IEEE Transactions on Audio, Speech and Language Processing, Mar 2008
- [7] X. Xiao, E-S. Chng, and Haizhou Li, "Temporal structure normalization of speech feature for robust speech recognition", IEEE Signal Processing Letters, 2007
- [8] Atal, B.S, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," Journal of the Acoustical Society of America vol. 55, 1304-1312, 1974
- [9] R. Haeb-Umbach, "Investigations on inter-speaker variability in the feature space," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 1999
- [10] F. Hilger and H. Ney, "Quantile based histogram equalization for noise robust large vocabulary speech recognition," IEEE Trans. on Audio, Speech and Language Processing, 2006
- [11] C-W. Hsu and L-S. Lee "Higher order cepstral moment normalization (HO CMN) for robust speech recognition", Proceedings of ICASSP, 2004
- [12] L-C. Sun, C-W. Hsu and L-S. Lee, "Modulation spectrum equalization for robust speech recognition", in Proc. IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), pp.81-86, 2007
- [13] W-H. Tu, S-Y. Huang and J-W. Hung, "Sub-band modulation spectrum compensation for robust speech recognition", accepted by IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), 2009
- [14] J-W. Hung, "Cepstral statistics compensation and normalization using online pseudo stereo codebooks for robust speech recognition in additive noise environments", IEICE Transactions on Information and Systems, Feb 2008
- [15] J-W. Hung, "Cepstral statistics compensation using online pseudo stereo codebooks for robust speech recognition in additive noise environments", 2006 International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), May 2006

- [16] J-W. Hung and W-H. Tu, "Incorporating codebook and utterance information in cepstral statistics normalization techniques for robust speech recognition in additive noise environments", IEEE Signal Processing Letters, June 2009
- [17] C-P. Chen and J-A. Bilmes, "MVA processing of speech features", IEEE Trans. on Audio, Speech, and Language Processing, pp.257-270, 2006
- [18] J-W. Hung and H. T. Fan, "Sub-band feature statistics normalization techniques based on a discrete wavelet transform for robust speech recognition", IEEE Signal Processing Letters, September 2009
- [19] H. G. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", Proceedings of ISCA IJWR ASR2000, Paris, France, pp.181-188, 2000