中華民國計算語言學學會
通訊 ACLCLP
發行人：許聞廉
主　編：王新民
執行編輯：黃　琪
The Association for Computational Linguistics and Chinese Language Processing 第二十四卷第一期
E-Mail：aclclp@hp.iis.sinica.edu.tw
Website : http://www.aclclp.org.tw
地 址：台北市研究院路二段128號中研院資訊所
電 話：(02)2788-3799 ext:1502
傳 眞：(02)2788-1638
劃 撥：19166251

本 期 要 目

ROCLING 2013

「第二十五屆自然語言與語音處理研討會」由國立中山大學與本會共同主辦，謹訂於十月四日(星期五)~五日(星期六)假高雄中山大學國際會議廳舉行。本次會議將廣邀學界與產業界投稿，並經嚴謹同行評審審稿程序，由學者發表高水準的論文，同時並舉行國科會研究計畫執行成果發表會。本次很榮幸邀請到在語音信號處理研究領域有卓越貢獻的日本豐田工業大學芝加哥分校(Toyota Technological Institute at Chicago)校長 Sadaoki Furui 教授作專題演講；會中亦將安排一場座談會，讓與會學者以及從事相關研究之企業廠商等進行討論與交流，透過產、學、研各界分享交流此領域最新的研發成果，藉此帶動自然語言及語音相關資訊技術的研究與發展。議程及徵稿啓事請參閱本刊第 5~8 頁。

Important Dates

- Paper submission: 8 August, 2013
- Notification of paper acceptance: 5 September, 2013
- Submission of camera-ready papers: 19 September, 2013
- Conference date: 4-5 October, 2013

線上會議管理系統

「線上會議管理系統」爲中研院資訊所研發成果，本系統爲一完整之全方位會議管理系統，包含線上投稿、審稿、報名註冊、流程管理、線上刷卡繳費，權限管理功能。經由此系統，可接受會議線上報名、論文投稿、審查、電郵自動發送等各項會議舉辦所需要的自動化管理功能。本系統已由中央研究院授權本會對外提供服務，服務對象限國內學術會議，系統詳細說明及申請方式請至本會網站查詢：http://www.aclclp.org.tw/sys_confer.php。

「語料庫語言學的新方向」研習營

「語料庫語言學的新方向」研習營謹訂於 2013 年 6 月 10 日假中研院人文館舉行，會中除了邀請香港理工大學人文學院院長黃居仁教授作專題演講外，並將舉行兩場專題討論。本活動免費參加，第一階段報名僅開放「台灣語言學學會」會員及本會會員優先報名，**報名期限至 5/10 日止**，名額有限！有興趣之會員請儘速報名。研習營詳細內容請參閱本刊第 9 頁。

The 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013)

第二十五屆自然語言與語音處理研討會

https://sites.google.com/site/rocling2013/

一、時間：2013 年 10 月 4 日(五)~10 月 5 日(六)

二、地點：高雄市國立中山大學國際會議廳

三、主辦單位：國立中山大學、中華民國計算語言學學會

四、大會組織

Conference Chairs

Hung-Duen Yang	楊弘敦	National Sun Yat-sen University
Chia-Ping Chen	陳嘉平	National Sun Yat-sen University
Wen-Lian Hsu	許聞廉	Academia Sinica

Program Chairs

Chia-Hui Chang	張嘉惠	National Central University
Chia-Ching Wang	王家慶	National Central University

Local Organization Chairs

Shi-Huang Chen	陳璽煌	Su-Te University
Shing-Tai Pan	潘欣泰	National University of Kaohsiung
Yu-Hsien Chiu	邱毓賢	Kaohsiung Medical University
Jiun-Ching Chen	陳俊卿	Su-Te University

Publication Chair

Jui-Feng Yeh	葉瑞峰	National Chia-Yi University

Publicity Chairs

Shih-Hung Wu	吳世弘	Chaoyang University of Technology
Wen-Hsing Lai	賴玟杏	National Kaohsiung First university of Science and Technology

Advisory Committee

Jason S. Chang	張俊盛	National Tsing Hua University
Berlin Chen	陳柏琳	National Taiwan Normal University
Hsin-Hsi Chen	陳信希	National Taiwan University
Keh-Jiann Chen	陳克健	Academia Sinica
Sin-Horng Chen	陳信宏	National Chiao Tung University
Jen-Tzung Chien	簡仁宗	National Chiao Tung University
Chu-Ren Huang	黃居仁	Hong Kong Polytechnic University
Jyh-Shing Jang	張智星	National Taiwan University
Chih-Chung Kuo	郭志忠	Industrial Technology Research Institute
Chin-Hui Lee	李錦輝	Georgia Institute of Technology
Lin-shan Lee	李琳山	National Taiwan University
Hai-zhou Li	李海洲	Institute of Infocomm Research, Singapore
Chin-Yew Lin	林欽佑	Microsoft Research Asia
Helen Meng	蒙美玲	Chinese University of Hong Kong
Hsiao-Chuan Wang	王小川	National Tsing Hua University
Hsin-Min Wang	王新民	Academia Sinica
Jhing-Fa Wang	王駿發	Tajen University
Chung-Hsien Wu	吳宗憲	National Chen Kung University

Steering Committee

Jing-Shin Chang	張景新	National Chi Nan University
Kuang-hua Chen	陳光華	National Taiwan University
Zhao-Ming Gao	高照明	National Taiwan University
Hung-Yan Gu	古鴻炎	National Taiwan University of Science and Technology
Jeih-weih Hung	洪志偉	National Chi Nan University
Jeng-Neng Hwang	黃正能	University of Washington, USA
Yuan-Fu Liao	廖元甫	National Taipei university of Technology
Chao-Lin Liu	劉昭麟	National Chengchi University
Jyi-Shane Liu	劉吉軒	National Chengchi University
Feng Zhu Luo	羅鳳珠	Yuan Ze University
Chin-Chin Tseng	曾金金	National Taiwan Normal University
Yuen-Hsien Tseng	曾元顯	National Taiwan Normal University
Ming-Shing Yu	余明興	National Chung Hsing University

Program Committee

Guo-Wei Bian	邊國維	Huafan University
Ivan Bulyko		BBN, USA
Tao-Hsing Chang	張道行	National Kaohsiung University of Applied Sciences
Yu-Yun Chang	張瑜芸	National Taiwan University
Yi-Hsiang Chao	趙怡翔	Chien-Hsin University of Science and Technology
Chien Chin Chen	陳建錦	National Taiwan University
Kuang-Hua Chen	陳光華	National Taiwan University
Pu-Jen Cheng	鄭卜壬	National Taiwan University
Tai-Shih Chi	冀泰石	National Chiao Tung University
Chih-Yi Chiu	邱志義	National ChiaYi University
Hung-Yan Gu	古鴻炎	National Taiwan University of Science and Technology
Shu-Kai Hsieh	謝舒凱	National Taiwan University
Jen-Wei Huang	黃仁暐	National Cheng Kung University
Jeng-Neng Hwang	黃正能	University of Washington, USA
Wei-Tyng Hong	洪維廷	Yuan Ze University
Shigeru Katagiri		Doshisha University, Japan
ChangIck Kim		KAIST, Korea
Lun-Wei Ku	古倫維	Academia Sinica
Chih-Chung Kuo	郭志忠	Industrial Technology Research Institute
Wen-Hsing Lai	賴玟杏	National Kaohsiung First university of Science and Technology
Bor-Shen Lin	林伯慎	National Taiwan University of Science and Technology
Chuan-Jie Lin	林川傑	National Taiwan Ocean University
Shou-De Lin	林守德	National Taiwan University
Shu-Yen Lin	林淑晏	National Taiwan Normal University
Chao-Hong Liu	劉昭宏	National Cheng Kung University
Chao-Lin Liu	劉昭麟	National Chengchi University
Jyi-Shane Liu	劉吉軒	National Chengchi University
Wen-Hsiang Lu	盧文祥	National Cheng Kung University
Yuen-Hsien Tseng	曾元顯	National Taiwan Normal University
Ming-Feng Tsai	蔡銘峰	National Chengchi University
Richard Tzong-Han Tsai	蔡宗翰	Yuan Ze University
Hsu Wang	王旭	Yuan Ze University
Wei-Ho Tsai	蔡偉和	National Taipei University of Technology

Chin-Chin Tseng	曾金金	National Taiwan Normal University
Jiun-Shiung Wu	吳俊雄	National Chung Cheng University
Shih-Hung Wu	吳世弘	Chaoyang University of Technology
Cheng-Zen Yang	楊正仁	Yuan Ze University
Jui-Feng Yeh	葉瑞峰	National ChiaYi University
Liang-Chih Yu	禹良治	Yuan Ze University
Ming-Shing Yu	余明興	National Chung Hsing University

五、暫訂議程

Friday, 4 October		
9:00-9:50	Registration	
9:50-10:00	Opening Ceremony	Hung-Duen Yang President of National Sun Yat-sen University
10:00-11:00	Keynote Speech I	Speaker: Inviting
11:00-11:20	Coffee Break	
11:20-12:40	Oral Session I	
12:40-13:40	Lunch/ACLCLP Assembly	
13:40-15:00	Oral Session II	
15:00-15:20	Coffee Break / IJCLCLP editors meeting	
15:20-16:40	Oral Session III	
16:40-17:40	Panel Discussion	
17:40-18:20	Break (NSYSU – Banquet place)	
18:20-21:00	Banquet	
Saturday, 5 October		
9:30-10:30	Keynote Speech II	Speaker: Sadaoki Furui Chair: Hsin-Min Wang
10:30-11:00	Coffee Break	
11:00-12:20	Oral Session IV	
12:20-13:20	Lunch	
13:20-14:20	Poster Session : Poster papers	
14:20-15:40	Oral Session V / Poster Session	
15:40-16:00	Coffee Break	
16:00-17:00	Oral Session VI	
17:00-17:20	Closing Ceremony and Best Paper Award	

六、 專題演講

Prof. Sadaoki Furui

Tokyo Institute of Technology/Toyota Technological Institute at Chicago

Sadaoki Furui received the B.S., M.S., and Ph.D. degrees from the University of Tokyo, Japan in 1968, 1970, and 1978, respectively. After joining the Nippon Telegraph and Telephone Corporation (NTT) Labs in 1970, he has worked on speech analysis, speech recognition, speaker recognition, speech synthesis, speech perception, and multimodal human-computer interaction. From 1978 to 1979, he was a visiting researcher at AT&T Bell Laboratories, Murray Hill, New Jersey. He was a Research Fellow and the Director of Furui Research Laboratory at NTT Labs. He became a Professor at Tokyo Institute of Technology in 1997, and was given the title of Professor Emeritus in 2011. He is now serving as President of Toyota Technological Institute at Chicago (TTI-C). He has authored or coauthored over 900 published papers and books including "Digital Speech Processing, Synthesis and Recognition." He was elected a Fellow of the IEEE (1993), the Acoustical Society of America (ASA) (1996), the Institute of Electronics, Information and Communication Engineers of Japan (IEICE) (2001) and the International Speech Communication Association (ISCA) (2008). He received the Paper Award and the Achievement Award from the IEICE (1975, 88, 93, 2003, 2003, 2008), and the Paper Award from the Acoustical Society of Japan (ASJ) (1985, 87). He received the Senior Award and Society Award from the IEEE SP Society (1989, 2006), the ISCA Medal for Scientific Achievement (2009), and the IEEE James L. Flanagan Speech and Audio Processing Award (2010). He received the NHK (Nippon Hoso Kyokai: Japan Broadcasting Corporation) Broadcast Cultural Award (2012) and the Okawa Prize (2013). He also received the Achievement Award from the Minister of Science and Technology and the Minister of Education, Japan (1989, 2006), and the Purple Ribbon Medal from Japanese Emperor (2006).

Data-intensive Automatic Speech Recognition Based on Machine Learning

Abstract

Since speech is highly variable, even if we have a fairly large-scale database, we cannot avoid the data sparseness problem in constructing automatic speech recognition (ASR) systems. How to train and adapt statistical models using limited amounts of data is one of the most important research issues in ASR. This talk summarizes major techniques that have been proposed to solve the generalization problem in acoustic model training and adaptation, that is, how to achieve high recognition accuracy for new utterances. One of the common approaches is controlling the degree of freedom in model training and adaptation. The techniques can be classified by whether a priori knowledge of speech obtained from a speech database such as those recorded using many speakers is used or not. Another approach is maximizing "margins" between training samples and the decision boundaries. Many of these techniques have also been combined and extended to further improve performance.

Although many useful techniques have been developed, we still do not have a golden standard that can be applied to any kind of speech variation and any condition of the speech data available for training and adaptation. We need to focus on collecting rich and effective speech databases covering a wide range of variations, active learning for automatically selecting data for annotation, cheap, fast and good-enough transcription, and efficient supervised, semi-supervised, or unsupervised training/adaptation, based on advanced machine learning techniques. We also need to extend current efforts to understand more about human speech processing and the mechanism of natural speech variation.

The 25th Conference on Computational Linguistics and Speech Processing
4-5 October 2013, National Sun Yat-sen University, Kaohsiung, Taiwan

Call for Papers

Welcome to ROCLING 2013, the flagship conference on computational linguistics, natural language processing, and speech processing in Taiwan. The conference will be held in National Sun Yat-sen University located in Kaohsiung. Kaohsiung is a rich and dynamic city with airport, high-speed railway, and mass-transportation subway. National Sun Yat-sen University is right next to the harbor area with world-class views. ROCLING 2013 is an academic event you do not want to pass! We invite paper submission, including-but-not-limited-to, in the following research areas:

- Natural Language Processing
  - Information retrieval
  - Machine translation
  - Named entity recognition
  - Question answering and information extraction
- Computational Linguistics
  - Computational morphology
  - Computational phonology
  - Parsing and part-of-speech tagging
- Speech Processing
  - Speech recognition
  - Speech synthesis
  - Voice conversion
  - Multi-lingual systems
  - Speaker recognition
  - Language identification
- Affective Computing
  - Affect recognition
  - Affect synthesis
  - Affective learning

Important Dates

Paper Submission: 8 August 2013
Notification of paper acceptance: 5 September 2013
Submission of camera-ready papers: 19 September 2013
Conference Dates: 4-5 October 2013

Paper Publications

Each submission will be reviewed based on originality, significance, technical soundness, and relevance to the conference. The conference proceedings will be included in the ACL Anthology. Selected papers will be invited to be extended and published in a special issue of the International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP).

Keynote Speakers

Sadaoki Furui
Tokyo Institute of Technology, Japen

Jianfeng Gao (uncertain)
Microsoft Research

Committee

Conference Co-Chairs
Hung-Duen Yang, Taiwan
Chia-Ping Chen, Taiwan
Wen-Lian Hsu, Taiwan

Advisory Committee
Jason S. Chang, Taiwan
Berlin Chen, Taiwan
Hsin-Hsi Chen, Taiwan
Keh-Jiann Chen, Taiwan
Sin-Horng Chen, Taiwan
Jen-Tzung Chien, Taiwan
Chu-Ren Huang, China
Jyh-Shing Jang, Taiwan
Chih-Chung Kuo, Taiwan
Chin-Hui Lee, USA
Chin-Yew Lin, China
Lin-shan Lee, Taiwan
Hai-zhou Li, Taiwan
Helen Meng, China
Hsiao-Chuan Wang, Taiwan
Hsin-Min Wang, Taiwan
Jhing-Fa Wang, Taiwan
Chung-Hsien Wu, Taiwan

「語料庫語言學的新方向」研習營

會議日期：2013 年 6 月 10 日
會議地點：中央研究院人文館三樓第二會議室
主辦單位：中央研究院語言學研究所、中華民國計算語言學學會、台灣語言學學會

議程：

時　間	議　程
8:30~9:00	報　到
9:00~10:30	主講人：黃居仁（香港理工大學人文學院） 題目：Expressive Meaning and Information Credibility: New Directions in Corpus and Computational Linguistics
10:30~10:50	茶敘時間
10:50~11:50	專題討論：自然語音語料庫：單位、邊界與結構 引言／討論人：曾淑娟（中研院語言所）
11:50-13:30	午餐時間
13:30~14:30	專題討論：巨量資料下的中文語料庫語言學：方法與省思 引言／討論人：謝舒凱（台灣大學語言所）
14:30-14:50	茶敘時間
14:50~15:50 （開放研究生、博士後及助理教授參加）	學術生涯規劃諮詢 主持人：黃居仁（香港理工大學人文學院）

備註：

1. 本活動免費參加，請線上報名；
   報名網址：http://www.aclclp.org.tw/register/form.php。
2. 本活動優先開放「中華民國計算語言學學會」及「台灣語言學學會」會員報名，報名日期至 5/10 止。
3. 參加人數限額 100 名。
4. 報名後擬修改報名資料或取消報名，請至線上系統修改，或 Email:crlu@gate.sinica.edu.tw(盧秋蓉 小姐)。
5. 聯絡電話：02-26525000 分機 6119。

Organizing Committee

General Chairs:
Thomas Fang ZHENG (Tsinghua Univ.)
Zhi DING (Univ. California–Davis, USA)

Technical Program Chairs:
Jiangtao Gene WEN (Tsinghua Univ.)
Xiaodong WANG (Columbia Univ., USA)

Industry Forum Chairs:
Eric CHANG (Microsoft Research Asia)
Jun WU (Tencent Co.)

Finance Chair:
Qiang ZHOU (Tsinghua Univ.)

Local Arrangement Chair:
Xiaojun WU (Tsinghua Univ.)

Publicity Chairs:
Lei XIE (Northwest. Polytech. Univ.)
Ming ZHOU (Microsoft Research Asia)

Publication Chair:
Changchun BAO (Beijing Univ. of Technology)
Kehu YANG (Xidian Univ.)

Registration Chairs:
Jia JIA (Tsinghua Univ.)
Askar HAMDULLA (Xinjiang Univ.)

Summer School Chairs:
Changhui ZHANG (Tsinghua Univ.)
Kai YU (Shanghai Jiaotong Univ.)

Special/Invited Session Chairs:
Jingdong CHEN (Northwest. Polytech. Univ.)
Chunyang YANG (Beihang Univ.)

International Liaisons:
Waleed ABDULLA (Univ. of Auckland, New Zealand)
Kenneth LAM (Hong Kong Polytech. Univ., Hong Kong)
Hitoshi KIYA (Tokyo Metropolitan Univ., Japan)
C.-C. Jay KUO (Univ. of Southern California, USA)
Yo-sung HO (Gwangju Institute of Sci. & Technol., Korea)
Haizhou LI (Institute for Infocomm Research, Singapore)
Chung-Hsien WU (National Cheng-Kung Univ., Taiwan)

Steering committee

Chair: Min WU (Univ. Maryland–College Park, USA)

Members-at-Large:
Yun HE (Tsinghua Univ.)
Jiangno HUANG (Northwest. Polytech. Univ.)
Jiwu HUANG (Sun-Yat-Sen Univ.)
Xinghao LIU (Shanghai Jiaotong Univ.)

SPS VP Conference: Wan-Chi SIU

SPS VP Technical Direction: Ahmed TEWFIK

SPS Region-10 Director: Ta-sung LEE

SPS Conference Manager: Lisa SCHWARZBEK

IEEE China Operation Director: Ning HUA

Exhibit and Demos

ChinaSIP 2013 welcomes exhibitions of products and demonstrations of R&D systems within the areas relevant to the conference.

Supporter Opportunities

Numerous opportunities are available to industrial, academia, and non-profit organizations to support the conference.

Please refer to the conference website for details on exhibit, demo, and supporter opportunities.

http://www.chinasip2013.org

The first IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP 2013) will be held 6-10 July 2013 at the China National Convention Center (CNCC), Beijing, China.

Sponsored by the IEEE Signal Processing Society (SPS), ChinaSIP® is a new annual summit and international conference held in China for domestic and international scientists, researchers, and practitioners to network and discuss the latest progress in theoretical, technological, and educational aspects of signal and information processing. ChinaSIP is a unique platform developed by IEEE SPS to help colleagues in China engage with the global community, and offer global colleagues opportunities to network and develop international collaborations.

As the inaugural summit and conference, ChinaSIP 2013's features include:

- **Technical tracks and industry forum.** Papers and presentations along the regular technical tracks as listed below focus on novel and significant research contributions. An industry forum provides a platform for exchange and networking among SIP industries as well as between academia and industries.
- **Invited papers and open-call papers.** Special invitations will be extended to major influential research groups in China to submit their latest contributions. Invited papers will be peer reviewed, and only papers with sufficient quality and significance will be accepted. In parallel, papers are also accepted through an open call from the community at large on a competitive basis.
- **Journal poster sessions.** Journal poster sessions provide a venue for overview and showcase of recent publications accepted by SPS journals. These already published journal papers will not be re-published with the ChinaSIP proceedings onto the IEEE Xplore®; a weblink and/or a reprint copy will be made available to attendees to facilitate at-conference exchanges.
- **Professional development program.** Several professional development activities will be organized, such as townhall meetings with the SPS leadership, trends/overview sessions, publication (EIC/AE) panels, and Fellow development sessions.
- **Summer schools.** The conference will set up summer schools before the regular sessions begin for students, researchers and practitioners to learn the state-of-the-art technologies and tools.

The regular technical program tracks and topics include (but not limited to):

- Signal/Information Processing Theory and Methods
- Speech, Language, and Audio
- Image, Video, and Multimedia
- Signal Processing for Communications and Networking
- Signal Sensing, Radar, Sonar, and Sensor Networks
- SIP Hardware/Software Designs and Systems
- Information Forensics and Security
- Pattern Recognition and Machine Learning
- Signal/Info Processing for Bioinformatics & Bio/Medicine

Submission of Papers

The official language of the conference is English. Prospective authors are invited to submit up to 4 pages in length (with an optional 5th page containing only references). The conference proceedings will be published at the IEEE Xplore®, and will be indexed by both IEEE Xplore® and EI Compendex.

The IEEE Signal Processing Society enforces a "no-show" policy. Any accepted paper included in the final program is expected to have at least one author or qualified proxy attend and present the paper at the conference. Authors of the accepted papers included in the final program who do not attend and present at the conference will be added to a "No-Show List", compiled by the Society. The "no-show" papers will not be published by IEEE on IEEE Xplore® or other public access forums, but these papers will be distributed as part of the on-site electronic proceedings and the copyright of these papers will belong to the IEEE.

Important Dates

Submission of Regular Full Papers:	8 Feb 2013
Submission of Special Sessions & Invited Papers:	28 Feb 2013
Notification of Paper Acceptance:	20 Apr 2013
Authors Registration Deadline:	10 May 2013
Attendees Advanced Registration Deadline:	1 Jun 2013
Summer School Dates:	6-7 Jul 2013
Summit and Conference Dates:	8-10 Jul 2013

IEEE Signal Processing Society

傅立葉頻譜圖之聯合時頻域分析及其語音強化之應用

徐忠謙、冀泰石
交通大學電信工程研究所

1. 簡介

語音強化演算法常在各種的應用中被廣泛的使用，以提升受損語音的語音品質，常見的方法有維納濾波器、統計模型架構法、子空間法等，然而這些方法在高訊雜比的情況下效果都不錯，但是在低訊雜比的情況下所得到的結果常常無法令人滿意[1]。這些演算法不是在時間維度就是在頻率維度使用分析-調整-合成(AMS)語音強化架構來消除雜訊，然而人的聽覺感知通常是針對時間與頻率兩維度同時來處理聲音，所以人們在只有單純的時域或頻域干擾下並不會受到很嚴重的影響。

心理聲學的研究發現，較慢的時域調變(小於 16 Hz)與語音的理解度有高度相關，此類的研究多藉由模糊時域的封包後再量測語音的理解度而完成[2][3]，這些研究成果啓發了許多工程上的應用。舉例來說萃取每個時間區間中的時域調變所建立的雙頻率模型(音頻頻率以及時域調變頻率) 在聲音的編碼[4]以及語音分離[5]的研究，除此之外，有關時域調變的特徵參數也被使用來建立強健型語音辨識系統[6]以及語者識別系統[7]，另一個例子是將常見的頻譜相減法被推廣至時域調變空間的語音強化技術[8]。

神經生理學上的證據也顯示了大腦聽覺皮層上(A1)的神經會根據輸入聲音的不同頻域與時域調變而有相對應的反應。從工程的觀點來看，A1 上的神經元可視爲根據不同的聯合時頻域調變解析度對輸入聲音進行拆解分析，根據這些發現，我們建立了一個運算式聽覺模型[9]，此模型也已經使用在許多的研究上，例如用以評估語音的理解度[10]、進行語音非語音的識別[11]以及樂器辨識[12]等。除神經生理學的研究外，心裡聲學的研究也藉由人類聽覺實驗驗證了與語音理解度相關的重要聯合時頻域調變範圍[13]。近幾年來，在語音工程上的應用都能發現使用聯合時頻域調變的概念，例如擷取強健型時頻域特徵參數用於語音辨識[14]以及語者識別系統[15]。

類似我們之前所提出的聽覺模型[9]，在此專文中我們將介紹一個針對傅立葉頻譜的聯合時頻域分析合成框架[16]，並將 AMS 架構推廣至聯合時頻域調變空間來進行語音強化，我們這樣的做法是基於雜訊在聯合時頻域調變空間中並不是均勻的影響語音，而這也正是人類在噪音環境下依然有不錯表現的主要原因。本專文的其餘部分如下，在第二段中我們將概述如何針對傅立葉頻譜進行聯合時頻域的分析與合成，並在第三段展現語音與雜訊在時頻域調變空間的不同的能量分布情形，接著在第四段中我們提出了一套語音強化技術，其將傳統的 AMS 架構下的維納濾波器推廣至時頻域調變空間，以及此技術之主客觀評比結果，第五段爲簡單的討論以及結論。

2. 針對傅立葉頻譜之聯合時頻域的分析與合成

在先前的研究中[16]，我們已將聽覺皮層的聯合時頻域分析概念實作在傅立葉頻譜上，藉由該論文所提出的一系列聯合時頻域調變濾波器，我們可以在聯合時頻域的分析過程中擷取到各種語音結構特性的資訊，例如音高(pitch)、諧波成分(hamonicity)、共振峰(formant)、振幅調變(AM)、頻率調變(FM)以及語音的起始與結束點(onset/offset)等。

2.1 聯合時頻域分析

首先，使用短時傅立葉轉換對輸入的聲音訊號求得其傅立葉頻譜圖，接著將頻譜輸入一組具零相位響應的二維聯合時頻域帶通濾波器($STMF_s$)進行分析。其中具有向下特徵(下標符號'+')以及向上特徵(下標符號'−')的二維聯合時頻域調變濾波器的頻率響應如下所示：

$$\mathrm{STMF}_{+}(\omega,\Omega)=\begin{cases}|\mathcal{F}\{\mathrm{h_{rate}}(t)\}\otimes\mathcal{F}\{\mathrm{h_{scale}}(f)\}|, 0\le\omega;\Omega\le\pi\\ 0, \text{otherwise}\end{cases}, \quad (式\ 1)$$

$$\mathrm{STMF}_{-}(\omega,\Omega)=\begin{cases}|\mathcal{F}\{\mathrm{h_{rate}}(t)\}\otimes\mathcal{F}\{\mathrm{h_{scale}}(f)\}|, -\pi\le\omega\le 0;0\le\Omega\le\pi\\ 0, \text{otherwise}\end{cases}, \quad (式\ 2)$$

上式中，$\mathcal{F}$代表一維的傅立葉轉換; ⊗代表外積; π表示傅立葉頻譜圖上時間與頻率軸的取樣頻率之一半。Rate(符號爲 ω，單位赫茲(Hz)，爲頻率(frequency))和 scale(符號爲 Ω，單位毫秒(ms)，爲倒頻率(quefrency))定義爲時間與頻率之傅立葉空間，$\mathrm{h_{rate}}(t)$以及$\mathrm{h_{scale}}(f)$爲 1 維常數-Q(constant Q，$Q_{3dB}=2$)之時域與頻域濾波器之脈衝響應，詳細的方程式細節可以在[16]中找到。如(式 1)以及(式 2)中所示，向下特徵及向上特徵的聯合時頻域調變濾波器($STMF_s$)分別落在$\omega-\Omega$空間中之第一象限及第二象限。最後輸入訊號的傅立葉頻譜圖經由聯合時頻域調變濾波器組($STMF_s$)分析後之四維輸出結果爲：

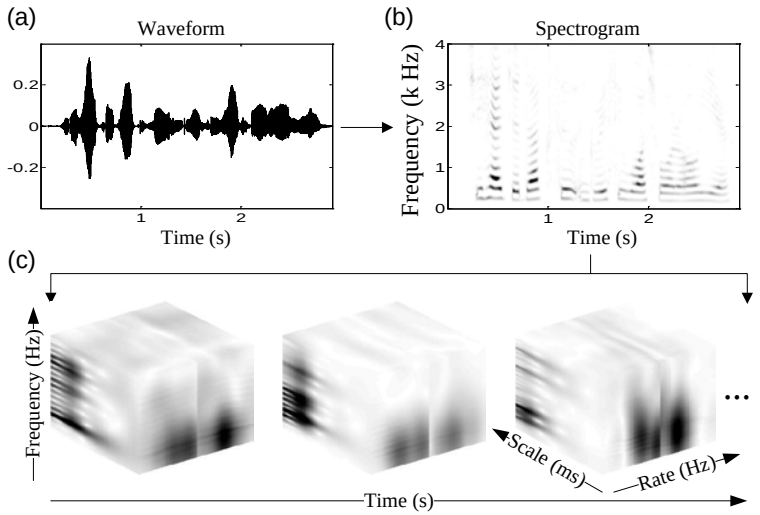
$$C(t,f,\omega,\Omega)=\mathcal{F}_{2D}^{-1}\{\mathcal{F}_{2D}\{X(t,f)\}\cdot \mathrm{STMF}_{\pm}(\omega,\Omega)\}, \quad (式\ 3)$$

上式中，$\mathcal{F}_{2D}$與$\mathcal{F}_{2D}^{-1}$定義爲二維傅立葉轉換以及逆轉換。圖一展現的是傅立葉頻譜圖的聯合時頻分析及其四維輸出結果。

2.2 聯合時頻域合成

(式 3)中的聯合時頻分析爲線性運算，所以其四維的輸出$C(t,f,\omega,\Omega)$可重建回傅立葉頻譜圖$X'(t,f)$，如以下所示：

$$X'(t,f)=\Re\left\{\mathcal{F}_{2D}^{-1}\left\{\frac{\sum_{\omega,\Omega}\mathcal{F}_{2D}\{C(t,f,\omega,\Omega)\}}{\sum_{\omega,\Omega}\mathrm{STMF}_{\pm}(\omega,\Omega)}\right\}\right\}, \quad (式\ 4)$$



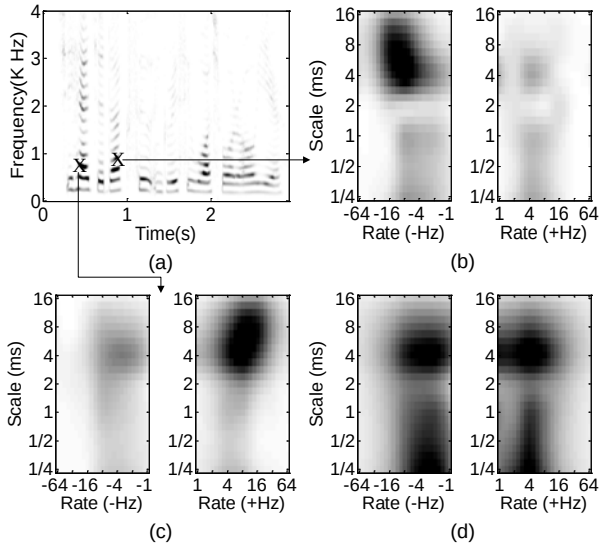
圖一：傅立葉頻譜圖之聯合時頻分析及其四維輸出結果；(a)語音之時間訊號；(b)語音訊號之傅立葉頻譜圖；(c)聯合時頻分析之四維輸出(scale-rate-frequency-time)。

重建回來的傅立葉頻譜圖可以藉由重疊相加法(OLA)重建回聲音訊號，在[16]中，我們已經展示針對傅立葉頻譜圖的分析與合成流程，不僅可以提供類似於聽覺模型[9]中所提出的聯合時頻域分析，其合成回聲音的步驟更可提供較好的聲音品質與較低的運算複雜度。

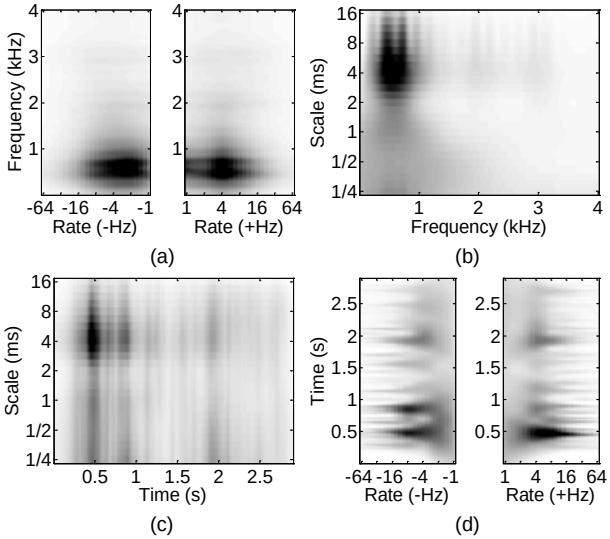
3. 語音與雜訊的聯合時頻域特徵

在分析的過程中，我們選擇了 1~64 Hz 的時間動態(rate)調變，以及 0.5~16 ms 頻率動態(scale)調變範圍來針對傅立葉頻譜圖進行二維解析。圖二(a)展示一從 TIMIT 語料庫中女性語者之傅立葉頻譜圖，圖二(b)(c)爲圖二(a)傅立葉頻譜圖上標記爲‘x’的時頻點的 rate-scale 表示圖，其定義爲$|C(\omega, \Omega; t_0, f_0)|$。在每個局部時頻點的向上及向下頻譜特徵可被此二維分析解析出來。舉例來說，圖二(b)顯示出在負 rate 上有個突出的峰點，這表示局部時頻點‘x’上具有向上、且約 16 Hz 的時間動態以及 4~8 ms 的頻譜動態調變。圖二(d)顯示的是整個傅立葉頻譜圖的平均 rate-scale 表示圖。在 4 Hz 的峰值顯示的是一般人的講話速度以及 4 ms 顯示的是此女性語者的基頻約爲 250 Hz。

進一步推廣，四維的 time-frequency-rate-scale 表示式$|C(t, f, \omega, \Omega)|$可以針對任兩個維度來做平均進而解析語音的特性。如圖三(a)所示，rate-frequency 表示圖顯示大部分的語音能量會集中在 frequency=1k Hz 以下以及 rate=16 Hz 以下並且有個在 rate=4 Hz 的峰值。在圖三(b)中，frequency-scale 表示圖顯示在 frequency=1K 以下以及在 scale=4 ms 附近有較高的能量，這也描述了語者的基頻以及共振峰的頻率間隔。至於圖三(c)(d)的 rate-time、time-scale 顯示的是在每個時間點上的時間動態的調變結果與其頻率結構。

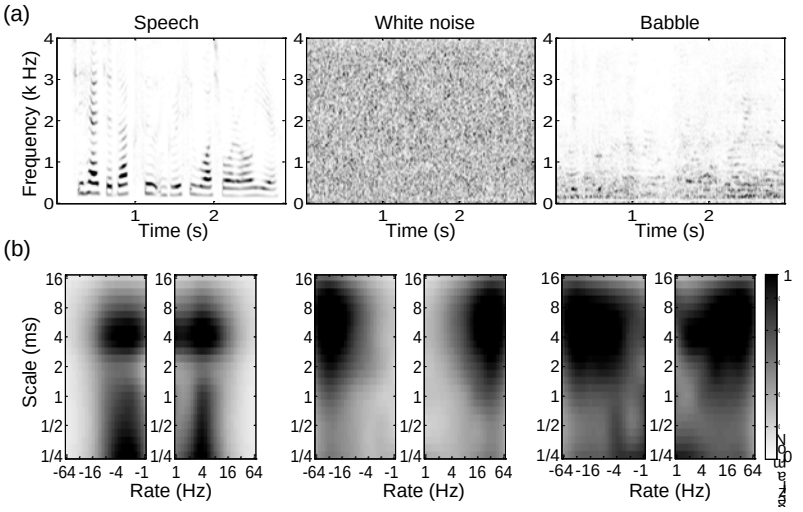


圖二：乾淨語音的 Rate-scale 表示圖。



圖三：聯合時頻域多維度解析之不同表示面向。

接著，我們比較輸入語音與雜訊在 rate-scale 空間中調變能量的分布。圖四(a)由左至右分別表示乾淨語音、白雜訊、人聲雜訊的傅立葉頻譜圖（雜訊由 NOISEX-92 資料庫中所取得），圖四(b)則表示語音與雜訊分別所對應的 rate-scale 空間上的調變能量分布。這些圖顯示出，雜訊 rate-scale 的調變能量較集中於高 rate 以及高 scale 區域，這表示，與雜訊相比，語音訊號具有較爲平滑的時頻域調變。



圖四：乾淨語音、白雜訊以及聲音雜訊的聯合時頻域調變能量分布。

4. 語音強化演算法及其驗證

圖四(b)中的調變能量分布展現了雜訊在 rate-scale 空間中不是均勻的影響語音，根據這項觀察，我們在此提出一套作用於 rate-scale 調變空間上的維納濾波器，以用來降低雜訊對語音的影響。在我們的實驗中，二維的時頻調變濾波器組，涵蓋了所有的 rate-scale 成份，我們並沒有事先濾除任何的 rate-scale 成份，這些二維時頻域調變濾波器組將傅立葉頻譜圖根據不同的 rate-scale 參數組合進行二維分解，我們將標準的維納濾波器[1]推廣到 rate-scale 調變子空間中，其增益函數如下式：

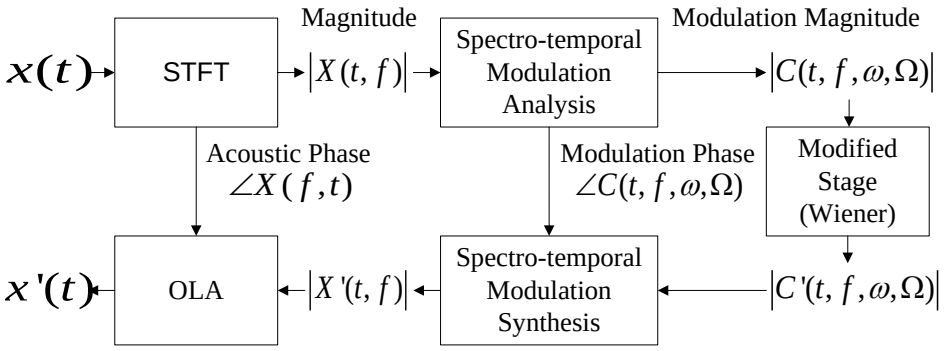
$$W(f;t_n,\omega_i,\Omega_j)=\left(\frac{P_S(f;t_n,\omega_i,\Omega_j)}{P_S(f;t_n,\omega_i,\Omega_j)+\alpha P_N(f;\omega_i,\Omega_j)}\right), \qquad (式\ 5)$$

在此，P_S、P_N表示在每一個 rate-scale 調變子空間中所估計的乾淨語音與雜訊能量，$P_N(f;\omega_i,\Omega_j)$可以由$P_N(t,f;\omega_i,\Omega_j)$對時間軸積分而得，而$P_N(t,f;\omega_i,\Omega_j)$可由輸入的雜訊語音前端的雜訊所估計而得，值得注意的是，爲了求得低 rate(<1 Hz)子空間中的$P_N(f)$，需要較長的估計區間(>1 s)才行。在此，我們假設至少有 1 秒以上的雜訊出現在輸入訊號的前端，則$P_S(f;t_n,\omega_i,\Omega_j)$可以由輸入訊號的能量頻譜$P_{S+N}(f;t_n,\omega_i,\Omega_j)$減去估計的$P_N(f;\omega_i,\Omega_j)$而得，α參數定義爲雜訊消除因子，它會無可避免地影響到高訊雜比以及低訊雜比下的雜訊消除結果[1]。

因此，調變子空間的語音強化可簡單表示爲，分別考慮每個 rate-scale 子空間(ω, Ω)中所算出的增益函數，如下所示：

$$C'(f; t_n, \omega_i, \Omega_j) = W(f; t_n, \omega_i, \Omega_j) \cdot C(f; t_n, \omega_i, \Omega_j) \ , \qquad (式\ 6)$$

調整過後的頻譜圖可根據(式 4)重建而得，接著使用重疊相加法可得到強化後的語音。圖五展示了我們所提出的聯合時頻域分析-調整-合成(ST AMS)的演算法架構。



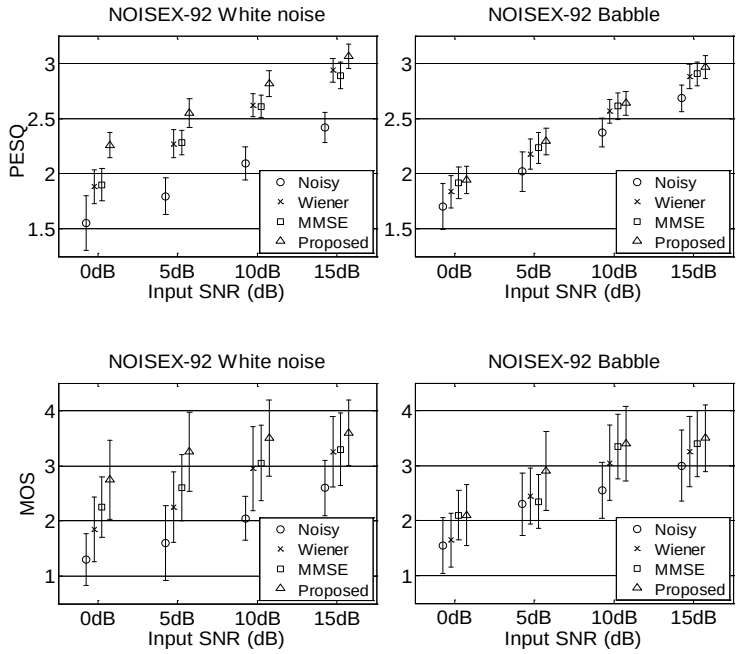
圖五：聯合時頻域分析-調整-合成的語音強化演算法架構圖。

在實驗中，我們使用 NOIZEUS 的語料庫[1]，此語料庫包含 30 句音素平衡的句子，在我們的模擬中，具有雜訊的語句是由加入 NOISEX-92 中的白雜訊以及人聲雜訊而得，並調整爲具備四種不同的訊雜比(15dB、10dB、5dB、0dB)。我們以主、客觀評比來衡量所提出的聯合時頻域語音強化演算法，並與隨著時間更新估計事前訊雜比的維納濾波器[17]，以及根據最小方均根誤差法設計的能進一步消除樂聲雜訊的先進方法[18]相比較。客觀的語音品質量測分數 PESQ[19]已被証明是一個可靠且與主觀量測結果具高度相關性的量測標準[20]，而在主觀的量測實驗中，各種訊雜比結合三種語音強化方法後的聲音以隨機順序經由 Sennheiser HD 380 Pro 耳機播放給十位受試者聽(22~26 歲且聽力正常)，受試者被要求以 5 個等級分數（1:很差;2:差;3:中等;4:好;5:很好）來衡量所聽到的聲音品質，將所有受試者的分數平均後的結果視爲聲音品質的平均意見分數(MOS)。

爲了決定α值，我們考慮了所有白雜訊訊雜比下的平均 PESQ 分數，平均分數在α=1 時爲 2.45 分，並往上攀升直到α=7 時達到最高的 2.67 分，並在α=10 後分數開始下降，因此，在我們的演算法中每個 rate-scale 子空間中的α值被統一設定爲 7。我們將不同訊雜比、雜訊類型以及各個語音強化演算法處理過後語音的平均 PESQ 和 MOS 分數與標準差展示於圖六中，結果顯示，我們所提出的語音強化演算法在白雜訊的情況下對語音品質有顯著的提升，但在人聲雜訊下僅呈現較少的提升，這是因爲如同圖四所示，語音與人聲雜訊在 rate-scale 空間中有較高的重疊性。

5. 結論

在本專文中，我們討論了一個針對傅立葉頻譜圖的聯合時頻域分析與合成程序，而聯合時頻域的調變濾波器組能對傅立葉頻譜圖進行解析，並展現每個時頻點的局部調變能量的分布，用以分析語音特徵。在這樣的聯合時頻域分析框架下，我們將一般的分析-調整-合成架構推廣至聯合時頻調變空間，提出單聲道的語音強化技術。模擬的結果顯示出我們所提出的時頻域調變維納濾波器法在主觀與客觀的語音品質量測上都有較好的表現。



圖六：各種語音強化演算法、
訊雜比及雜訊類型的主觀 PESQ 與客觀 MOS 分數比較圖。

如圖五所示，在我們提出的方法中，只有調變子空間的振幅被調整，而調變子空間的頻譜圖相位以及聲學相位在重建的過程中皆使用原本雜訊語句的相位，若能藉由相位恢復演算法[21]估計更適合的相位，相信能進一步的提升重建後的語音品質。另一方面，我們目前將所有子空間的α值設定爲一定值，我們認爲應該需要根據背景雜訊在不同的 rate-scale 子空間中的能量分布來設定各別子空間的α參數值，我們最終的目標是要在目前的演算法中加入針對雜訊時頻調變內容的時變預估器，進一步的提升語音強化效能。

6. 參考文獻

[1] P. C. Loizou, *Speech Enhancement: Theory and Practice* (CRC, New York, 2007).

[2] R. Drullman, J. M. Festen and R. Plomp, "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1053-1064, 1994.

[3] R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 5234, pp. 303-304, 1995.

[4] J. K. Thompson and L. E. Atlas, "A non-uniform modulation transform for audio coding with increased time resolution," in *Proc. ICASSP*, vol. 5, pp. 397-400, 2003.

[5] S. M. Schimmel, L. E. Atlas and K. Nie, "Feasibility of single channel speaker separation based on modulation frequency analysis," in *Proc. ICASSP*, vol. 4, pp. 605-608, 2007.

[6] J. Tchorz and B. Kollmeier , "A model of auditory perception as front end for automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 2040-2050, 1999.

[7] T. H. Falk and W.-Y. Chan, "Modulation spectral features for robust far-field speaker identification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 90-100, 2009.

[8] K. Paliwal, K. Wójciciki and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Comm.* , vol. 52, no. 5, pp. 450-475, 2010.

[9] T. Chi, P. Ru and S. A. Shamma, "Multi-resolution spectro-temporal analysis of complex sounds," *J. Acoust. Soc. Am.*, vol. 118, no. 2, pp. 887-906, 2005.

[10] M. Elhilali, T. Chi, and S. A. Shamma, "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," *Speech Comm.* , vol. 41, pp. 331–348, 2003.

[11] N. Mesgarani, M. Slaney, and S. A. Shamma, "Discrimination speech from nonspeech based on multiscale spectro-temporal modulations," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no.3, pp. 920–930, 2006.

[12] K. Patil, D. Pressnizer, S. A. Shamma, and M. Elhilali, "Music in our ears: the Biological bases of musical perception," *PLOS Comp. Bio.*, vol. 8, no. 11, e1002759, 2012.

[13] T. M. Elliott and F. E. Theunissen, "The modulation transfer function for speech intelligibility," *PLoS Comp. Bio.*, vol. 5, no. 3, e1000302, 2009.

[14] R. M. Stern and N. Norgan, "Hearing is believing: biologically inspired methods for robust automatic speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 34–43, 2012.

[15] H. Lei, B. T. Meyer, and N. Mirghafori, "Spectro-temporal Gabor features for speaker recognition," in *Proc. ICASSP*, pp. 4241–4244, 2012.

[16] T.-S. Chi and C.-C. Hsu, "Multiband analysis and synthesis of spectro-temporal modulations of Fourier spectrogram," *J. Acoust. Soc. Am.*, vol. 129, no. 5, pp. EL190-EL196, 2011.

[17] P. Scalart and J. V. Filho, “Speech enhancement based on a priori signal to noise estimation,” in *Proc. ICASSP*, vol. 2, pp. 629-632, 1996.

[18] O. Cappé, “Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor,” *IEEE Trans. Speech and Audio Process.*, vol. 2, no. 2, pp. 345-349, 1994.

[19] “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” ITU-T Recommendation, ITU-T, Geneva, Switzerland, p. 862 (2001).

[20] Y. Hu and P.C. Loizou, “Evaluation of objective quality measures for speech enhancement,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no.1, pp. 229-238, 2008.

[21] D. W. Griffin and J. S. Lim, “Signal estimation from modified short-time Fourier transform,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236-243, 1984.